Finding missing links in interaction networks

J. Christopher D. Terry (1) james.terry@zoo.ox.ac.uk Corresponding author

Owen T. Lewis (1) owen.lewis@zoo.ox.ac.uk

1: Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK.

Running title: Finding missing links in interaction networks

Keywords: ecological network, bipartite network, missing links, sampling completeness, network metrics

Data accessibility statement

R functions to run the predictive models are publicly available in the R package *cassandRa* on CRAN. All code and necessary data to generate simulated datasets and carry out analyses are available on GitHub (github.com/jcdterry/PredictingMissingLinks). Should the manuscript be accepted, this will also be archived in a public repository and the data doi included.

Statement of Authorship

JCDT initiated the research, in discussion with OTL. JCDT conducted all analyses and wrote the first draft of the manuscript. Both authors contributed to revisions.

Abstract

Documenting which species interact within ecological communities is challenging and labour-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, but that the predictive power of structural models varies with the underlying network structure. Predictions can be improved by ensembling multiple models. Sample-coverage estimators of the number of missed interactions are highly correlated with the number of missed interactions, but strongly biased towards underestimating the true number of missing links. Augmenting observed networks with most-likely missing links improves estimates of qualitative network metrics. Simple to implement tools to identify likely missing links can enable prioritisation of research effort and more robust assessment of network properties.

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 & 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 & 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 ecological mechanisms (Lee & Guénard 2019). Furthermore, under-sampling will skew the apparent prevalence of interactions towards 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 stabilise (Rivera-Hutinel *et al.* 2012; Costa *et al.* 2016), explicit observation models (Weinstein & 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 & Bascompte 2007; Blüthgen *et al.* 2008; Poisot *et al.* 2012b; Fründ *et al.* 2016; de Aguiar 2018; 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 not static, 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 dynamically important. 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 & McCann 2016; Jacquet *et al.* 2016; Kadoya *et al.* 2018).

Ecological networks are manifestly non-random (Dunne & Pascual 2006; Montoya *et al.* 2006; Vázquez *et al.* 2009; Bascompte & 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 & Sales-Pardo 2009) and *k*-nearest neighbour 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 part of benchmarking tests, often the inference scenarios are not orientated towards ecological needs and to date there has been limited uptake of these methods from the ecological community.

Furthermore, it is common in empirical ecological networks for species to have been sampled to markedly different extents (although 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 wide range of ecological bipartite networks with a range of models, including a new species-level coverage-deficit approach which incorporates information on the completeness of sampling for individual species. We show that inferential models can perform well, especially when used in ensemble, and demonstrate how they could inform empirical work to document ecological networks and help evaluate hypotheses.

Methods

Predictive Models

We seek to infer the true interaction frequency matrix composed of elements a_{ij} starting from an observed bipartite adjacency matrix **O**. This has n rows representing each focal layer species and m columns detailing each potential interaction partner. Statistical models can be used to assign to each

unobserved interaction in a network a relative probability $P(a_{ij} > 0)$ of being a 'missing link' (false negative), rather than a true negative.

Here, we introduce and test a new coverage deficit model and four structural models described in the literature, singularly and in combination. All models were fit using the *cassandRa* R package (Terry 2019). Initially we focus on identifying the qualitative network (i.e. where $a_{ij} \neq 0$), but we later test the potential to augment networks with likely missed links and the quantitative relationship between $P(a_{ij} > 0)$ and a_{ij}. We assume throughout that the species observed (the dimensions of **O**) are a complete list – we do not attempt here to predict interactions with unobserved species.

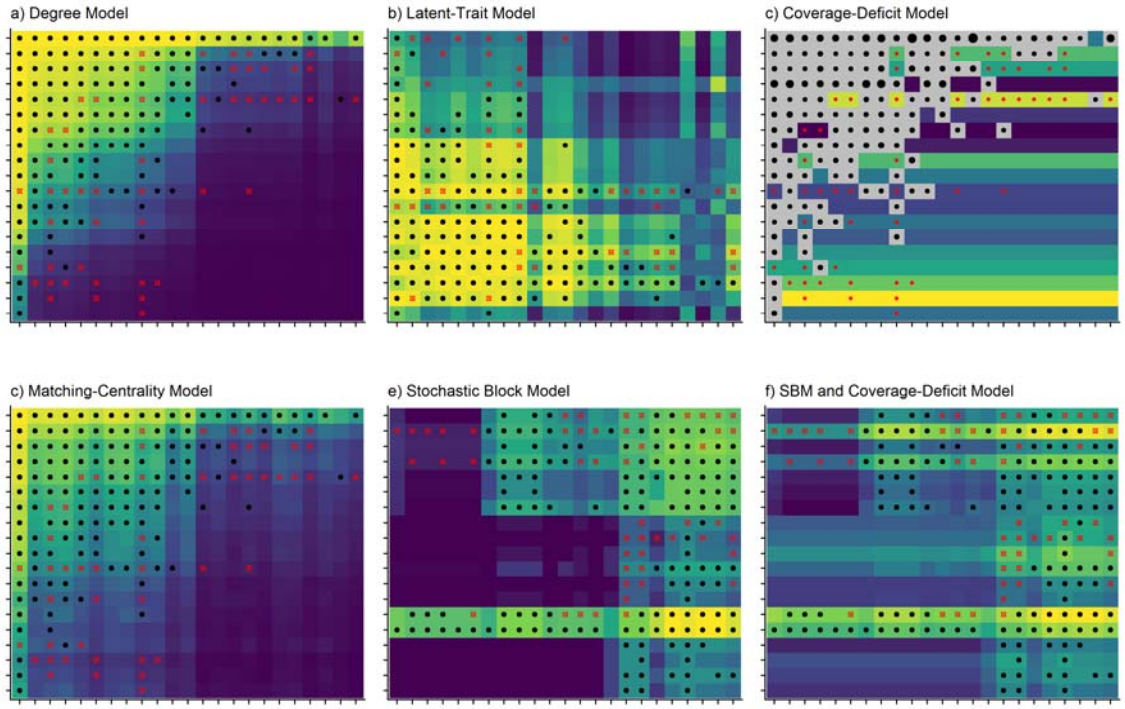

Figure 1. Demonstration of the fits of predictive models to a subsampled simulated bipartite network. Black dots (⍰) indicate observed interactions, red targets (⍰) indicate unobserved missing links that we wish to identify. Brighter colours indicate a higher probability of that interaction being present according to that model. Species are sorted to highlight the features of each model. In the coverage deficit model, observed interactions (grey) are not assigned a likelihood and the size of the dot is proportional to the number of observations. AUC values for this example were as follows: Degree: 0.80,

Latent-Trait: 0.76, Matching-Centrality: 0.81, SBM: 0.88, Coverage-Deficit: 0.7, Averaging the SBM and the Coverage-Deficit model 0.92.

Coverage-Deficit Model

When ecological bipartite networks are constructed, it is common that one level (the focal level) of a network is directly sampled and the interaction partners in the other level counted. Examples include gathering insect hosts and rearing parasitoids from them, or observing pollinators that visit a set of plant species. This sampling may be opportunistic, matched to the local abundance, or standardised in some other manner. The result is a set of observed interaction counts with each focal layer species. Although these interaction counts may be later rescaled to better represent local abundances, here we assume the raw observation data is available. The method relies on discretisable relationships where individual interaction events can be observed. This includes many ecological interactions such as parasitism, pollination visits and feeding observations, but excludes diffuse facilitatory or competitive interactions.

An extensive literature has developed to assess the sample completeness of species inventories (Chao & Colwell 2017), based on the observation that the majority of information about the number of missing observations is contained in the most infrequent observations. Due to its deep roots in information theory 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). This estimate of *coverage deficit* ($\hat{C}_{def}$), is a lower bound estimator of the proportion of interactions that occur between as-yet-unobserved partners, or alternatively the probability that the next interaction partner observed would not yet have been observed with that species. It is defined as:

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

where f_1 is the number of interactions observed only once, f_2 is the number of interactions observed exactly twice, and n is the total observed interactions involving that focal species.

Unfortunately, ecological interaction networks typically contain a number of focal layer species that are extremely poorly observed, resulting in situations where the Chao1 estimator is undefined or inappropriate. For very poorly sampled species, the distribution of interactions cannot itself provide much information (Colwell & Coddington 1994). Where observations are exclusively singletons 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 not having yet observed any of the unobserved missing links is: $L = \left(1 - \hat{C}_{def}\right)^n$. While the maximum likelihood estimate of $\hat{C}_{def}$ will be 0, with a Bayesian approach and assuming a flat prior the posterior mean can be found directly as: $\mathbb{E}(\hat{C}_{def}|n) = 1 - \frac{1}{n+2}$. We use this alternative estimate when $n \leq 5$, all the observations are singletons ($f_1 = n$), or if there are no singletons or doubletons ($f_1 + f_2 = 0$). We use the $\hat{C}_{def_i}$ estimate for each focal layer species i to assign a relative probability to each potential missing interaction as:

$$P\left(a_{ij} > 0\right) \sim \frac{\hat{C}_{def_i}}{\sum_j(\boldsymbol{O}_{ij}=0)}.$$

Since coverage deficit is strongly correlated with sample size, we also tested a simpler model that assigns the relative probability of missed interactions based solely on the number of observations of each focal layer-species: $P\left(a_{ij} > 0\right) \sim \frac{1}{\sum_j(\boldsymbol{O}_{ij})}$.

Latent-Trait Model

Species do not interact at random with each other – similar species tend to interact with a similar sets 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 in directly measured traits, such as corolla and tongue length (Vizentin-Bugoni *et al.* 2014; Klumpers *et al.* 2019) 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 & Traveset 2016). Nonetheless, such approaches have had successes and a family of approaches have been developed including matching models (Rohr *et al.* 2016), Dirichlet-multinomial regression (Crea *et al.* 2016), matrix factorisation (Seo & Hutchinson 2018) and fourth-corner analyses (Spitz *et al.* 2014). While it is possible to fit any number of latent traits, here we examine using a simple single latent trait model (detailed in SI 3) to reduce the potential of overfitting (Rohr *et al.* 2016).

Degree 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 that a missed interaction involving a 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). Simple degree-models can fit 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. For our degree model (detailed in SI 3), we assign each species a

connectance term. When both species have high observed degrees, 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 degree model and is fit in a similar manner (SI 3). This dual approach has been shown to have a high capacity to tightly to diverse network structures (Rohr *et al.* 2016).

Stochastic Block Model

Ecological networks, especially antagonistic networks, frequently show comparatively discrete compartments or modularisation, where subsets of species interact more strongly within their group than with outsiders (Olesen *et al.* 2007, 2008; Allesina & 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). These have been shown to perform well on ecological data in comparison to other clustering algorithms (Leger *et al.* 2015) and are seeing increased use in the ecological literature (Sander *et al.* 2015; Kéfi *et al.* 2016). In an SBM, each species is assigned to a group in a defined set. The probability of interaction between two species is defined based on their group membership. We find optimal group assignations and fit the model using a degree-corrected bipartite-SBM specific algorithm (Larremore *et al.* 2014, SI 3).

Combinations of Models

The aforementioned models can be combined into ensembles to harness multiple sources of information and structure. For example, if a generalist species was also comparatively poorly sampled, this could indicate even more strongly the location of likely missing links. Probabilities can be combined by either averaging or multiplying with distinct consequences. Multiplying emphasises the extreme probabilities each model finds, while averaging highlights interactions that are

consistently identified by multiple models. Here we test ensembling the Matching-Centrality model with an SBM, and each of the structural models with the coverage-deficit model, either averaged or multiplied. Before combining models we standardise each set of probabilities assigned to unobserved interactions to sum to 1.

Testing Predictions

Datasets

Testing the efficacy of extrapolations requires knowledge of the 'true' network. Since the efficacy of the method will depend on the network structure 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 parameterised by the 'true' interaction frequency. This replicates ecological sampling processes where the most frequent interactions are highly likely to have been observed and the rarest missed.

We initially generated 2000 simulated networks from a probabilistic two-trait niche model (modified from Fründ *et al.* 2016) described in full in SI 1. Key network properties were calculated using the *bipartite* R package (Dormann *et al.* 2008): connectance, weighted nestedness (Galeano *et al.* 2009), cluster coefficient (Watts & 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 and the (low) correlations between them are depicted in SI 1.

Our objective was to generate interaction matrices that represent a wide range of ecologically 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 parsimoniously represent ecological networks. To reduce

223 circularity, our generative model is considerably more complex than our predictive models and we
224 examine how predictive model performance changes in response to network properties.

225 Using the 'true' networks we took 300-2000 samples per network to generate our 'observed'
226 networks. We discarded occasional cases where all true interactions were observed. This resulted in
227 a low mean true sample coverage deficit of 2.4%, corresponding to an average network
228 completeness of 64% due to skewed interaction strength distributions. Our systems therefore
229 represent reasonably 'well-sampled' systems that are nonetheless missing a large proportion of the
230 interactions present.

231 We collated a diverse set of 113 empirical networks representing antagonistic, mutualistic and
232 commensalistic interaction types (SI 2). We collated quantitative single-class bipartite networks from
233 the *Web of Life* (www.web-of-life.es) repository with 30-200 species and 40-431 observed
234 interactions. This included 48 plant-pollinator (19 sources), 23 mammal-flea (one source, but across
235 a wide geographic range), 2 plant-ant mutualism (2 sources) and 10 seed disperser (from 7 sources)
236 networks. We supplemented this with 25 mammal-dung beetle interaction networks (Frank *et al.*
237 2018, selecting those with ≥30 total species, ≥40 observed interactions, and ≥5 mammal species) and
238 5 host-parasitoid networks (Tylianakis *et al.* 2007).

239 These networks are not exhaustively sampled (SI 2). Nonetheless, we took the empirical interaction
240 frequency to be the 'true' network to define a multinomial distribution. We subsampled each
241 empirical network 20 times, each time taking a different fraction between 0.2 and 1 of the original
242 sample size, taken to be the sum of the observations. We excluded cases where the observed web
243 included more than 95% of the 'true' interactions.

244 *Testing model capacity*

245 Using each model or combination of models we assign each unobserved interaction a relative
246 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 re-evaluation. 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 performance at identifying missing links with the area under receiver operating characteristic curve (AUC) metric to assess the information content of a signal. 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's rank-correlation with the 'true' relative interaction strengths of the unobserved interactions for each model.

We used the *EstimateR* function in the *vegan* R package (Oksanen *et al.* 2018) to estimate the number of missing links based on the total distribution of observed interaction counts using the bias-corrected Chao1 index. We then augmented the observed interaction networks by adding the estimated number of the most probable missing links to the network based on the average of the SBM, matching-centrality and coverage-deficit models. Each selected missing link was taken to be 'observed' once and network-level metrics calculated for the observed, augmented and true networks. We tested three qualitative network metrics that had been identified (Fründ *et al.* 2016) as being susceptible to misestimation in under-sampled networks (connectance, nestedness and niche overlap), alongside two metrics considered to be comparatively resistant to under-sampling (H2' and weighted NODF nestedness).

Results

Locating missing links in simulated networks

In general, the network models showed impressive performance at identifying likely missing links. The overall best performing model was the stochastic block model, closely followed by the other three network models (Figure 2a). The coverage deficit model lagged behind structural models, but was more informative than simply ranking focal species by sample size. Combinations of models tended to perform better than individual models on the simulated data (Table 1), with the overall best performing model averaging the stochastic block model, the matching-centrality model and the coverage deficit model. However, the marginal improvements from combining models were often small.

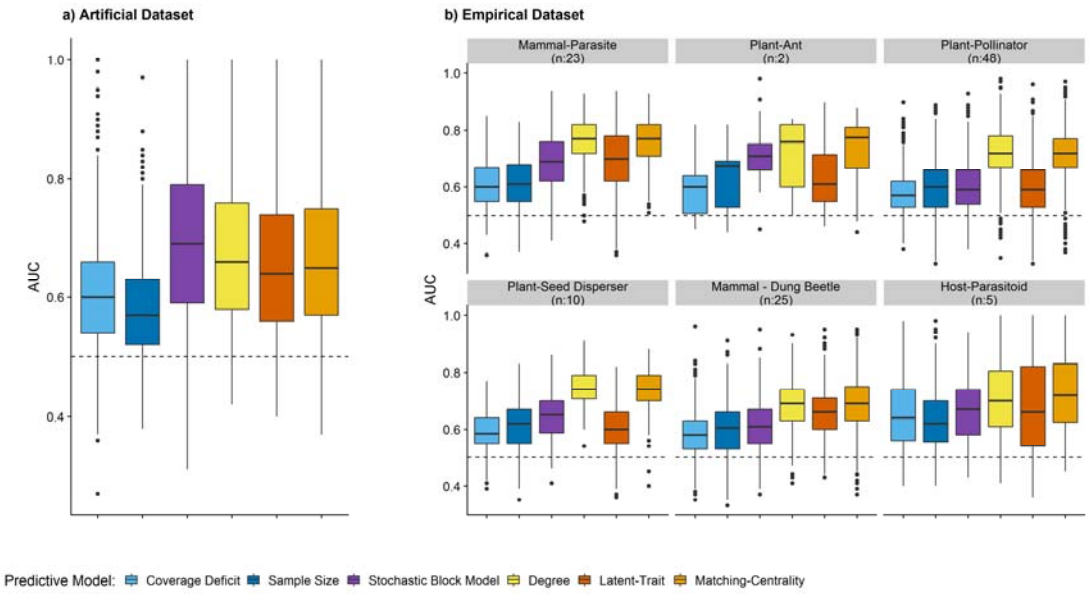

Figure 2. a) Model performance across 1985 simulated networks. Mean differences between the statistical models were small, although given the large sample size all pairwise differences were statistically significant (Holm-corrected t-tests, paired by network ID, p<0.0001). **b)** Model performance on the empirical data sets, subdivided by network class. Each empirical network was subsampled 20 times to different fractions of the original sample size. Models perform differently on different classes of data (Model comparison, accounting for dataset ID as a random effect, shows model:class interaction term is significant: p< 0.0001, n_{obs} = 12672). In all cases boxes show first and third quartiles.

	Simulated Dataset			Empirical Dataset		
Model		$+C_{def}$ (Multiplied)	$+C_{def}$ (Averaged)		$+C_{def}$ (Multiplied)	$+C_{def}$ (Averaged)
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]
Degree	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
SBM	0.691 [bcd]	0.689 [bc]	0.695 [cd]	0.630[f]	0.584[a]	0.590[c]
Combined Models						
SBM and Matching-Centrality (Multiplied)	0.702 [de]	0.705[ef]	0.711[f]	0.647 [gh]	0.594[c]	0.604[de]
SBM and Matching-Centrality (Averaged)	0.705[ef]	0.709[f]	0.713 [g]	0.648 [gh]	0.600[d]	0.603[d]

Table 1 Mean AUC values using different predictive models or combinations of models across the whole simulated (n= 1985) and empirical (n=113, 2112 subsamples) network sets. Incorporating coverage deficit tends to lead to modest improvements in the simulated networks but not in the empirical. Sets of results not found to be significantly different (Holm-correct t-tests, paired by network ID) are indicated with grouping letters, separate for simulated and empirical results.

Model performance and relative hierarchy was affected by the underlying network structure (Figure 3). On average, missing links were identified best in nested, unclustered networks where interactions were evenly distributed. An SBM was generally either as-good-as, or better than, the other models across the networks, but particularly at large matrix size, low connectance, low (or negative) nestedness. The coverage deficit and focal layer 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 appreciably change with increased sample size, overall interaction coverage deficit or samples per host (Figure S4). However, with an increased fraction of observed interactions, the structural models increased their performance, while the coverage deficit model performance remained constant (Figure S4).

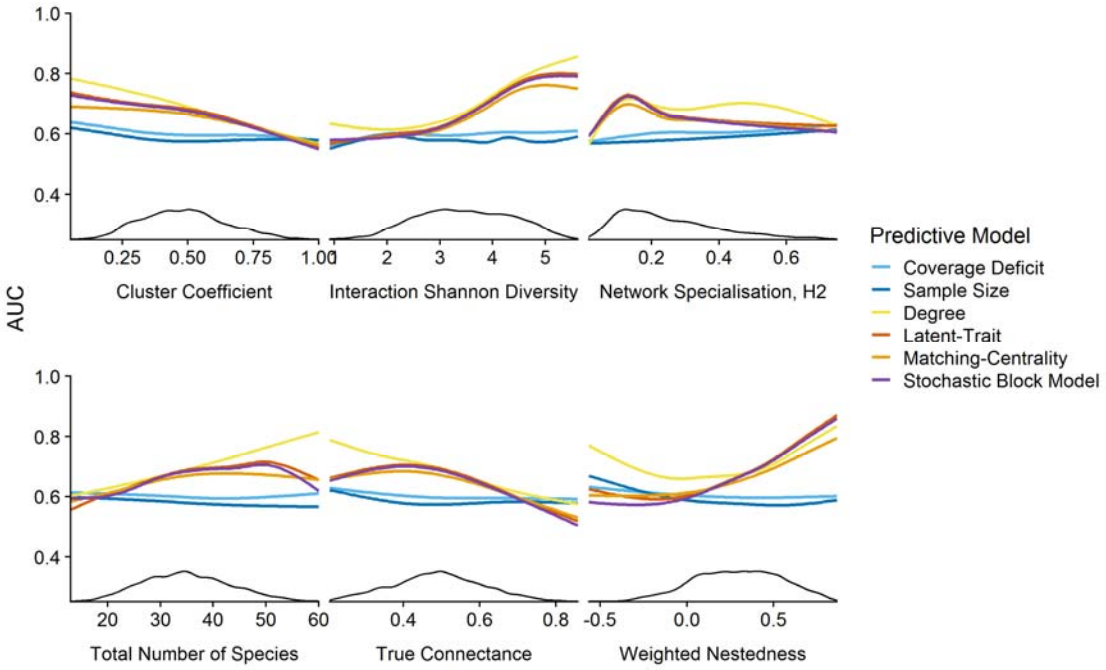

Figure 3. Response of model performance to 'true' network structure. Smoothed line shows a GAM fit for each predictive model. Lower black line shows the local density distribution of that property across the simulated test dataset.

Locating missing links in empirical networks

The performance of the different models varied considerably between the different classes of networks (Figure 2b). In line with our expectations, the degree model performed best in the mutualistic networks, while the latent-trait model showed its best performance on antagonistic networks. In all six network classes the matching-centrality model performed the best on average. The 'true network' properties (Fig S3) or the degree of sampling rarefaction (Fig S4) had relatively little effect on the predictive performance. An exception was networks that show 'anti-nestedness' -

in these cases the predictive capacity of the centrality model was low, although the matching-centrality model maintained its advantage.

The coverage-deficit model was very poor at predicting missing links in the empirical data and combining it with the network-structure models lead to a *reduction* in overall predictive power (Table 1). Ranking by the focal species observation count performed better overall.

Identifying frequency and number of missing interactions

Correlations between probabilities assigned to missing links ($P(a_{ij} > 0)$) and their 'true' interaction frequency were stronger in the simulated dataset than the empirical (Table S2). Of the structural models, the degree-model was on average the most strongly positively correlated with interaction strength (ρ_{Sim} = 0.51, ρ_{Emp} = 0.15), while the SBM was the least (ρ_{Sim} = 0.28, ρ_{Emp} = 0.06). The coverage deficit model was negatively correlated with true frequency (ρ_{Sim} = -0.35, ρ_{Emp} = -0.05) i.e. missing links successfully identified due to poor sampling tended to be infrequent.

The Chao1 estimate of the number of missing interactions appeared to perform well on the empirical networks but consistently underestimated the true interaction count and coverage deficit in the simulated data (Figure 4).

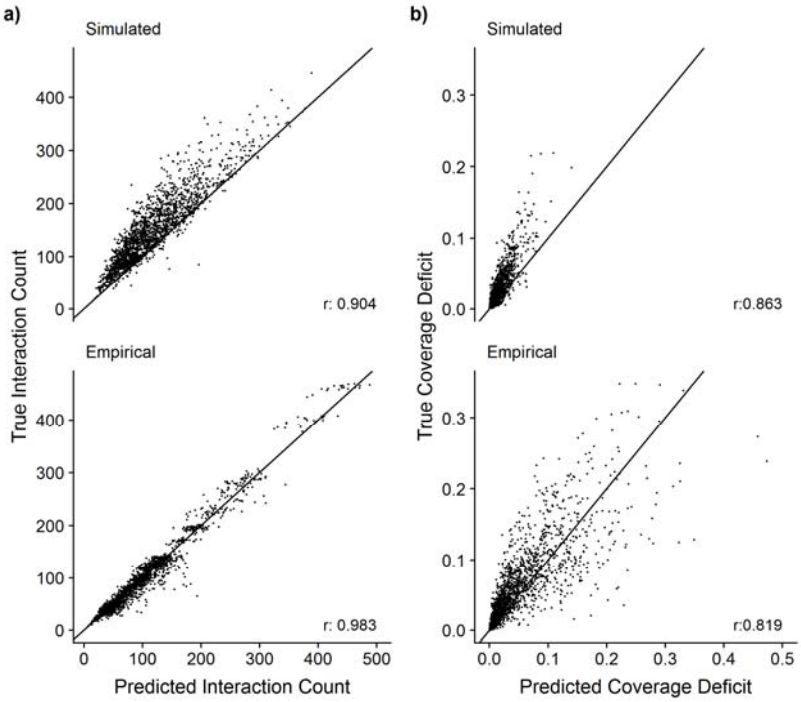

Figure 4. Relationships between **a)** estimated sample coverage and 'true' sample coverage and **b)** estimated number of missing interactions and the 'true' number of missed interactions. Results are shown using the bias-corrected Chao1 estimator, the ACE estimator had very similar results. Line shows a 1:1 relationship that an unbiased metric would follow.

Improving estimates of network metrics

The 'augmented' networks showed increased correlation, reduced bias and reduced error in qualitative metrics, compared to the original observed networks (Figure 5a). The quantitative network metrics (H2'and weighted NODF, Figure 5b) were already well estimated by the under-sampled network, and the inclusion of additional interactions made little difference.

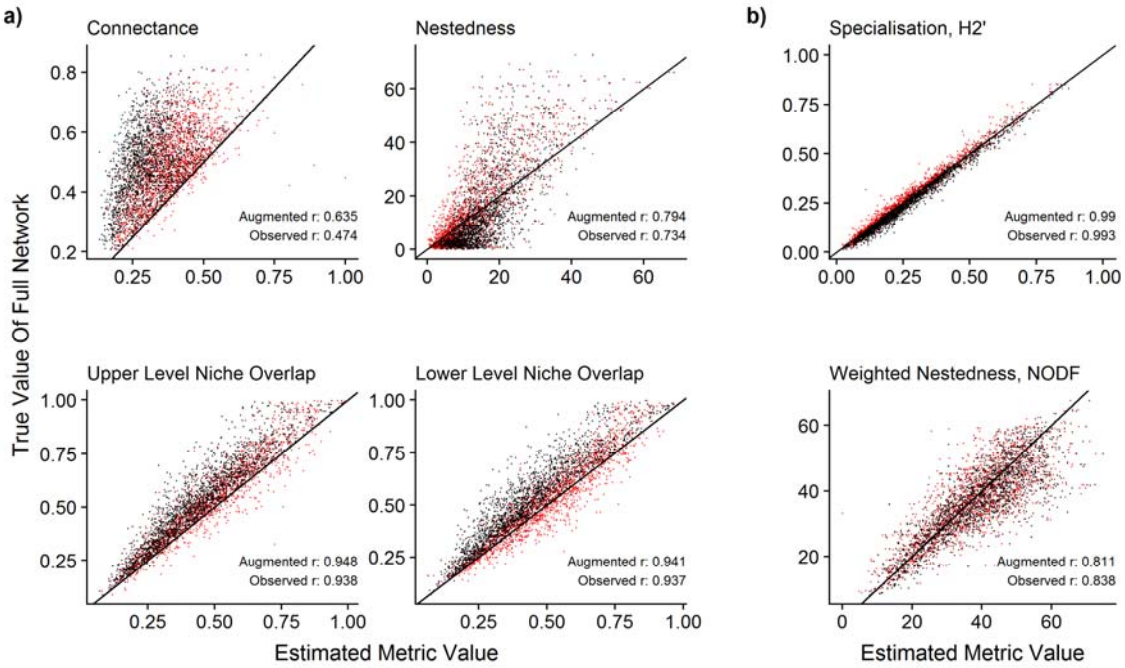

Figure 5. Relationship between network metrics calculated on the observed (black) and 'augmented' (red) interaction networks to the true network value in the simulated networks. Line shows 1:1 relationship, Pearson's correlation coefficient (r) shown for each metric. Qualitative metrics (a) show notable improvements in correlation with the true values while quantitative (b) networks are largely unchanged.

Discussion

We have shown that is possible to infer missing links in range of ecological 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 ecological insight.

Choosing between predictive models

Ideally, a model used to infer missed interactions would capture the true ecological drivers, 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 be dependent on the data set. In our simulated data it would appear that a stochastic block model had the best individual performance. In our empirical datasets, the matching-centrality model performed consistently well (as found by Rohr *et*

357 *al.* 2016). However, we have reason to be distrustful of the empirical datasets (discussed below). The
358 pronounced, likely artefactual, abundance-generality relationship will favour degree-models.

359 Nonetheless, we would argue that splitting hairs over the best performing model is not necessarily
360 the most productive route. Networks will be structured by multiple processes and we note that in
361 our simulated data sets that the very best performance comes from the combination of different
362 models. These identify different missing interactions – structure-based models pick out the more
363 frequent interactions while coverage deficit models highlight comparatively infrequent interactions
364 that would be the hardest to determine through further undirected sampling. Future progress will
365 come from operationalising estimated missing links, rather than from further incremental model
366 refinements.

367 *Distinctions between empirical and simulated networks*

368 The empirical and simulated datasets substantially overlapped in key network metrics. The main
369 relevant difference between our datasets is increased correlation in empirical networks between
370 species marginal totals and generality (mean Spearman's ρ_{Sim} = 0.43, ρ_{Emp} = 0.76). This can account
371 for the poor performance of the coverage-deficit model in these cases. Disentangling the extent to
372 which apparent specialism of rarely observed species is a sampling artefact is challenging (Blüthgen
373 *et al.* 2006), since there are ecological explanations for a relationship (Fort *et al.* 2016; Dormann *et*
374 *al.* 2017; Simmons *et al.* 2019). For our purposes here, this likely bias in the structure of empirical
375 ecological networks has the consequence that predictive models may identify both missing links
376 introduced by our subsampling procedure and missing links due to gaps in the original dataset. We
377 would therefore place more weight on the results on the simulated data, while noting that, despite
378 the obstacles, the structural models are still able to perform reasonably well on the sparse empirical
379 data.

380 *Estimating the quality of sampling and the number of missing interactions*

381 Sample coverage estimators offer promise for estimating the number of missed interactions (Chacoff
382 *et al.* 2012; Traveset *et al.* 2015; Jordano 2016; MacGregor *et al.* 2017). The distinction between
383 sample-coverage and network completeness is worth emphasising. Apparently high values of sample
384 coverage alone should not be used as a simple proxy for 'network completeness', although they are
385 related to the reliability of certain quantitative network metrics (Henriksen *et al.* 2019). We found
386 the whole-network-level Chao1 estimate of missed interactions to be strongly correlated with the
387 'true' number, but notably underestimated the true number of missing links in our simulated
388 networks, a pattern also observed by Fründ *et al.* (2016). Although there appeared to be no
389 systematic bias in estimated network completeness in the empirical datasets, these 'true' networks
390 are themselves estimated to be significantly incomplete, suggesting an overall underestimation here
391 also. Our focus here was on the location, not the number of missing links. Future work should
392 investigate how long tails of infrequent interactions relate to this underestimation and consider
393 whether species rank abundance distribution estimators (Chao *et al.* 2014) can be usefully
394 generalised to assigning a frequency distribution to estimated missing links.

395 *The way forward*

396 Most of the models used here, and indeed our fitting approaches, likely have further room for
397 refinement, and there are many further classes of predictive models that we did not test, such as
398 nearest neighbour algorithms (Desjardins-Proulx *et al.* 2017). Furthermore, we based our models
399 entirely on the information contained in the network itself. In most cases, ecologists constructing a
400 network have access to (sometimes considerable) additional information and expertise on the study
401 system. Introducing pre-existing knowledge about likely or forbidden interactions, or pooled
402 knowledge from multiple sample sites offer considerable opportunities to maximise the information
403 in a network (Gray *et al.* 2015; Pomeranz *et al.* 2018). External information would be particularly
404 useful in separating 'forbidden links' (Olesen *et al.* 2011) that never occur due to fundamental
405 incompatibilities from those that occur elsewhere but would not within those spatio-temporal

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 was only gathered in the day.

Because the naïve 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. Due to 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, rather than developing new models to gain additional marginal improvements in predictive capacity, focus instead needs to switch to developing frameworks to make use of 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 deliminated 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 well 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 (Mercer 1966) or may have significant consequences (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 principle uses for inferred missing links:

Firstly, directing further sampling where the goal is a descriptive network. In many cases the topography of the network is of principle interest, given the potential for interaction strengths to vary through time and in response to perturbation. For example, when tracing the relationships of

431 species that may be the target for eradication measures, such as disease-vector mosquitoes (Collins
432 *et al.* 2019), a critical first step is an inventory of interactions in the community. Since in most cases
433 investment in sampling is a finite resource, directing sampling towards high-likelihood interactions
434 could more efficiently increase the overall network completeness. However, we caution that
435 directed sampling will introduce biases in the generated network (Gibson *et al.* 2011; Falcão *et al.*
436 2016) and may hinder cross-community comparability (Poisot *et al.* 2012a). In certain cases
437 independently sampled networks, even if less complete, may be more useful than comparisons
438 between 'augmented' networks. The use of information from other networks may lead them to
439 appear more homogenous than they really are and lead to underestimations of interaction β-
440 diversity through time and space (Poisot *et al.* 2015; Graham & Weinstein 2018).

441 Secondly, the identification of probable missing links can allow better testing of the reliability of
442 conclusions drawn from networks, potentially through explicitly probabilistic networks (Poisot et al.
443 2016a). Statistical resampling and rarefaction procedures that can only work backwards or adding
444 random interactions may both introduce biases. Whilst it is established that certain quantitative
445 metrics are relatively robust to under-sampling (Blüthgen *et al.* 2008; Fründ *et al.* 2016; Henriksen *et*
446 *al.* 2019), these metrics may not align with the goals of the study. If the conclusions are robust to the
447 inclusion of probable missing links then confidence in the results would be increased. Conversely, if
448 the inclusion of most-likely missing links erased the differences this would signal that the conclusions
449 are less robust than hoped. In these cases, care should be taken when selecting a predictive model
450 to avoid a clash with the properties under investigation.

451 *Conclusion*

452 The identification of likely missing links is a powerful tool to include in the arsenal of network
453 ecologists. We have shown that they can perform well across a wide range of network types.
454 Importantly, inferential models need not be technically challenging to implement – for example the
455 *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

Thanks to the CERO group for comments on this project. Funding was provided by NERC grant NE/N010221/1.

References

1.de Aguiar, M.A.M. (2018). Revealing biases in the sampling of ecological interaction networks. *bioRxiv*

2.Allesina, S. & Pascual, M. (2009). Food web models: a plea for groups. *Ecol. Lett.*, 12, 652–662

3.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

4.Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Funct. Ecol.*, 30, 1894–1903

5.Bascompte, J. & Jordano, P. (2013). *Mutualistic Networks*. Princeton University Press, Princeton

6.Bascompte, J., Jordano, P., Melian, C.J. & Olesen, J.M. (2003). The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci.*, 100, 9383–9387

7.Berlow, E.L. (1999). Strong effects of weak interactions in ecological communities. *Nature*, 398, 330–334

8.Bersier, L.-F., Banašek-Richter, C. & Cattin, M.-F. (2002). Quantitative descriptors of food web matrices. *Ecology*, 83, 2917–2926

9.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

10.Blüthgen, N., Fründ, J., Vázquez, D.P. & Menzel, F. (2008). What Do Interaction Network Metrics Tell Us about Specialization and Biological Traits. *Ecology*, 89, 3387–3399

11.Blüthgen, N., Menzel, F. & Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC Ecol.*, 6, 9

12.CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017). Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecol. Lett.*, 20, 385–394

13.Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012). Evaluating sampling completeness in a desert plant-pollinator network. *J. Anim. Ecol.*, 81, 190–200

14.Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.*, 11, 265–270

15.Chao, A. & Colwell, R.K. (2017). Thirty years of progeny from Chao's inequality: Estimating and comparing richness with incidence data and incomplete sampling. *Stat. Oper. Res. Trans.*, 41, 3–54

16.Chao, A., Hsieh, T.C., Chazdon, R.L. & Colwell, R.K. (2014). Unveiling the Species-Rank Abundance Distribution by Generalizing Good-Turing Sample Coverage Theory. *Ecology*, 96, 1189–1201

17.Clauset, A., Moore, C. & Newman, M.E.J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, 453, 98–101

18.Collins, C.M., Bonds, J.A.S., Quinlan, M.M. & Mumford, J.D. (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. *Med. Vet. Entomol.*, 33, 1–15

19.Colwell, R.K. & Coddington, J.A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. B-Biological Sci.*, 345, 101–118

20.Costa, J.M., da Silva, L.P., Ramos, J.A. & Heleno, R.H. (2016). Sampling completeness in seed dispersal networks: When enough is enough. *Basic Appl. Ecol.*, 17, 155–164

21.Crea, C., Ali, R.A. & Rader, R. (2016). A new model for ecological networks using species-level traits. *Methods Ecol. Evol.*, 7, 232–241

22.Delmas, E., Besson, M., Brice, M.-H., Burkle, L.A., Dalla Riva, G. V., Fortin, M.-J., *et al.* (2019). Analysing ecological networks of species interactions. *Biol. Rev.*, 94, 16–36

23.Desjardins-Proulx, P., Laigle, I., Poisot, T. & Gravel, D. (2017). Ecological interactions and the Netflix problem. *PeerJ*, 5, e3644

24.Dormann, C.F., Frund, J. & Martin Schaefer, H. (2017). Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. *Annu. Rev. Ecol. Evol. Syst*, 48, 559–84

25.Dormann, C.F., Gruber, B. & Frund, J. (2008). Introducing the bipartite Package: Analysing Ecological Networks. *R News*, 8, 8–11

26.Dudley, J.P., Hang'Ombe, B.M., Leendertz, F.H., Dorward, L.J., de Castro, J., Subalusky, A.L., *et al.* (2016). Carnivory in the common hippopotamus Hippopotamus amphibius: implications for the ecology and epidemiology of anthrax in African landscapes. *Mamm. Rev.*, 46, 191–203

27.Dunne, J.A. & Pascual, M. (2006). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford University Press, Oxford

28.Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N.P., *et al.* (2013). The dimensionality of ecological networks. *Ecol. Lett.*, 16, 577–583

29.Falcão, J.C.F., Dáttilo, W. & Rico-Gray, V. (2016). Sampling effort differences can lead to biased conclusions on the architecture of ant-plant interaction networks. *Ecol. Complex.*, 25, 44–52

30.Fort, H., Vázquez, D.P. & Lan, B.L. (2016). Abundance and generalisation in mutualistic networks: Solving the chicken-and-egg dilemma. *Ecol. Lett.*, 19, 4–11

31.Frank, K., Krell, F.-T., Slade, E.M., Raine, E.H., Chiew, L.Y., Schmitt, T., *et al.* (2018). Global dung webs: high trophic generalism of dung beetles along the latitudinal diversity gradient. *Ecol. Lett.*, 21, 1229–1236

32.Fründ, J., Mccann, K.S. & Williams, N.M. (2016). Sampling bias is a challenge for quantifying specialization and network structure: Lessons from a quantitative niche model. *Oikos*, 125, 502–513

33.Galeano, J., Pastor, J.M. & Iriondo, J.M. (2009). Weighted-Interaction Nestedness Estimator (WINE): A new estimator to calculate over frequency matrices. *Environ. Model. Softw.*, 24, 1342–1346

34.Gellner, G. & McCann, K.S. (2016). Consistent role of weak and strong interactions in high- and low-diversity trophic food webs. *Nat. Commun.*, 7, 11180

35.Gibson, R.H., Knott, B., Eberlein, T. & Memmott, J. (2011). Sampling method influences the structure of plant-pollinator networks. *Oikos*, 120, 822–831

36.Goldwasser, L. & Roughgarden, J. (1997). Sampling Effects and the Estimation of Food-Web Properties. *Ecology*, 78, 41–54

37.González-Varo, J.P. & Traveset, A. (2016). The Labile Limits of Forbidden Interactions. *Trends Ecol. Evol.*, 31, 700–710

38.Gotelli, N.J. & Colwell, R.K. (2010). Estimating species richness. In: *Biological Diversity: Frontiers in Measurement and Assessment* (eds. Magurran, A. & McGill, B.J.). Open University Press, pp. 416–422

39.Graham, C.H. & Weinstein, B.G. (2018). Towards a predictive model of species interaction beta diversity. *Ecol. Lett.*, 21, 1299–1310

40.Gray, C., Baird, D.J., Baumgartner, S., Jacob, U., Jenkins, G.B., O'Gorman, E.J., *et al.* (2014). Ecological networks: the missing links in biomonitoring science. *J. Appl. Ecol.*, 51, 1444–1449

41.Gray, C., Figueroa, D.H., Hudson, L.N., Ma, A., Perkins, D. & Woodward, G. (2015). Joining the dots: An automated method for constructing food webs from compendia of published interactions. *Food Webs*, 5, 11–20

42.Guimera, R. & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci.*, 106, 22073–22078

43.Henriksen, M. V., Chapple, D.G., Chown, S.L. & McGeoch, M.A. (2019). The effect of network size and sampling completeness in depauperate networks. *J. Anim. Ecol.*, 88, 211–222

44.Jacquet, C., Moritz, C., Morissette, L., Legagneux, P., Massol, F., Archambault, P., *et al.* (2016). No complexity–stability relationship in empirical ecosystems. *Nat. Commun.*, 7, 12573

45.Jordano, P. (2016). Sampling networks of ecological interactions. *Funct. Ecol.*, 30, 1883–1893

46.Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of plant-animal interactions. *Ecol. Lett.*, 6, 69–81

47.Kadoya, T., Gellner, G. & McCann, K.S. (2018). Potential oscillators and keystone modules in food webs.

Ecol. Lett., 21, 1330–1340

48.Kéfi, S., Miele, V., Wieters, E.A., Navarrete, S.A. & Berlow, E.L. (2016). How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience. *PLOS Biol.*, 14, e1002527

49.Klumpers, S.G.T., Stang, M. & Klinkhamer, P.G.L. (2019). Foraging efficiency and size matching in a plant-pollinator community: the importance of sugar content and tongue length. *Ecol. Lett.*, 469–479

50.Lafferty, K.D., Allesina, S., Arim, M., Briggs, C.J., De Leo, G., Dobson, A.P., *et al.* (2008). Parasites in food webs: the ultimate missing links. *Ecol. Lett.*, 11, 533–546

51.Laigle, I., Aubin, I., Digel, C., Brose, U., Boulangeat, I. & Gravel, D. (2018). Species traits as drivers of food web structure. *Oikos*, 127, 316–326

52.Larremore, D.B., Clauset, A. & Jacobs, A.Z. (2014). Efficiently inferring community structure in bipartite networks. *Phys. Rev. E*, 90, 012805

53.Lee, R.H. & Guénard, B. (2019). Choices of sampling method bias functional components estimation and ability to discriminate assembly mechanisms. *Methods Ecol. Evol.*, 2019, 1–12

54.Leger, J.-B., Daudin, J.-J. & Vacher, C. (2015). Clustering methods differ in their ability to detect patterns in ecological networks. *Methods Ecol. Evol.*, 6, 474–481

55.MacGregor, C.J., Evans, D.M. & Pocock, M.J.O. (2017). Estimating sampling completeness of interactions in quantitative bipartite ecological networks: incorporating variation in species specialisation. *bioRxiv*, 195917

56.Martinez, N.D., Hawkins, B.A., Dawah, H.A. & Feifarek, B.P. (1999). Effects of Sampling Effort on Characterization of Food-Web Structure. *Ecology*, 80, 1044–1055

57.McCann, K., Hastings, A. & Huxel, G.R. (1998). Weak trophic interactions and the balance of nature. *Nature*, 395, 794–798

58.Memmott, J. (2009). Food webs: a ladder for picking strawberries or a practical tool for practical problems? *Philos. Trans. R. Soc. B Biol. Sci.*, 364, 1693–1699

59.Mercer, A.J. (1966). Turnstones feeding on human corpse. *Br. Birds*, 59, 307

60.Montoya, J.M., Pimm, S.L. & Solé, R. V. (2006). Ecological networks and their fragility. *Nature*, 442, 259–64

61.Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from proxies. *Trends Ecol. Evol.*, 30, 347–356

62.Morris, R.J., Gripenberg, S., Lewis, O.T. & Roslin, T. (2014). Antagonistic interaction networks are structured independently of latitude and host guild. *Ecol. Lett.*, 17, 340–349

63.Nielsen, A. & Bascompte, J. (2007). Ecological networks, nestedness and sampling effort. *J. Ecol.*, 95, 1134–1141

64.Novotny, V., Miller, S.E., Baje, L., Balagawi, S., Basset, Y., Cizek, L., *et al.* (2010). Guild-specific patterns of species richness and host specialization in plant-herbivore food webs from a tropical forest. *J. Anim. Ecol.*, 79, 1193–203

65.Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al.* (2018). vegan: Community Ecology Package

66.Olesen, J.M., Bascompte, J., Dupont, Y.L., Elberling, H., Rasmussen, C. & Jordano, P. (2011). Missing and forbidden links in mutualistic networks. *Proc. R. Soc. B Biol. Sci.*, 278, 725–732

67.Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. (2007). The modularity of pollination networks. *Proc. Natl. Acad. Sci.*, 104, 19891–19896

68.Olesen, J.M., Bascompte, J., Elberling, H. & Jordano, P. (2008). Temporal dynamics in a pollination network. *Ecology*, 89, 1573–1582

69.Paine, R.T. (1980). Food Webs: Linkage, Interaction Strength and Comunity Infrastructure. *J. Anim. Ecol.*, 49, 666–685

70.Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. (2012a). The dissimilarity of species interaction networks. *Ecol. Lett.*, 15, 1353–1361

71.Poisot, T., Canard, E., Mouquet, N. & Hochberg, M.E. (2012b). A comparative study of ecological specialization estimators. *Methods Ecol. Evol.*, 3, 537–544

72.Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016a). The structure of probabilistic networks. *Methods Ecol. Evol.*, 7, 303–312

73.Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124, 243–251

74.Poisot, T., Stouffer, D.B. & Kéfi, S. (2016b). Describe, understand and predict: why do we need networks in ecology? *Funct. Ecol.*, 30, 1878–1882

75.Pomeranz, J.P.F., Thompson, R.M., Poisot, T. & Harding, J.S. (2018). Inferring predator–prey interactions in food webs. *Methods Ecol. Evol.*, 2018, 1–12

76.Rasmussen, C., Dupont, Y.L., Mosbacher, J.B., Trøjelsgaard, K. & Olesen, J.M. (2013). Strong Impact of

Temporal Resolution on the Structure of an Ecological Network. *PLoS One*, 8, e81694

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

78.Rohr, R.P., Naisbit, R.E., Mazza, C. & Bersier, L.-F. (2016). Matching–centrality decomposition and the forecasting of new links in networks. *Proc. R. Soc. B Biol. Sci.*, 283, 20152702

79.Rohr, R.P., Naisbit, R.E., Mazza, C. & Bersier, L.-F. (2018). Statistical Approaches for Inferring and Predicting Food-Web Architecture. In: *Adaptive Food Webs, Stability and Transistions of Real and Model Ecosystems* (eds. Moore, J.C., de Ruiter, P.C., McCann, K.S. & Wolters, V.). Cambridge University Press, pp. 178–192

80.Sander, E.L., Wootton, J.T. & Allesina, S. (2015). What Can Interaction Webs Tell Us About Species Roles? *PLoS Comput. Biol.*, 11, e1004330

81.Schleuning, M., Ingmann, L., Strauß, R., Fritz, S.A., Dalsgaard, B., Matthias Dehling, D., *et al.* (2014). Ecological, historical and evolutionary determinants of modularity in weighted seed-dispersal networks. *Ecol. Lett.*, 17, 454–463

82.Seo, E. & Hutchinson, R.A. (2018). Predicting Links in Plant-Pollinator Interaction Networks Using Latent Factor Models With Implicit Feedback. *AAAI*

83.Simmons, B.I., Vizentin-Bugoni, J., Maruyama, P.K., Cotton, P.A., Marín-Gómez, O.H., Lara, C., *et al.* (2019). Abundance drives broad patterns of generalisation in plant–hummingbird pollination networks. *Oikos*, oik.06104

84.Song, C., Rohr, R.P. & Saavedra, S. (2017). Why are some plant–pollinator networks more nested than others? *J. Anim. Ecol.*, 86, 1417–1424

85.Spitz, J., Ridoux, V. & Brind'Amour, A. (2014). Let's go beyond taxonomy in diet description: Testing a trait-based approach to prey-predator relationships. *J. Anim. Ecol.*, 83, 1137–1148

86.Staniczenko, P.P.A., Kopp, J.C. & Allesina, S. (2013). The ghost of nestedness in ecological networks. *Nat. Commun.*, 4, 1391

87.Staniczenko, P.P.A., Lewis, O.T., Tylianakis, J.M., Albrecht, M., Coudrain, V., Klein, A.-M., *et al.* (2017). Predicting the effect of habitat modification on networks of interacting species. *Nat. Commun.*, 8, 792

88.Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). Linear filtering reveals false negatives in species interaction data. *Sci. Rep.*, 7, 45908

89.Terry, J.C.D. (2019). cassandRa v.0.1.0, *https://cran.r-project.org/package=cassandRa*

90.Terry, J.C.D., Morris, R.J. & Bonsall, M.B. (2017). Trophic interaction modifications: an empirical and theoretical framework. *Ecol. Lett.*, 20, 1219–1230

91.Traveset, A., Olesen, J.M., Nogales, M., Vargas, P., Jaramillo, P., Antolín, E., *et al.* (2015). Bird–flower visitation networks in the Galápagos unveil a widespread interaction release. *Nat. Commun.*, 6, 6376

92.Tylianakis, J.M., Laliberté, E., Nielsen, A. & Bascompte, J. (2010). Conservation of species interaction networks. *Biol. Conserv.*, 143, 2270–2279

93.Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature*, 445, 202–5

94.Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009). A consumer's guide to nestedness analysis. *Oikos*, 118, 3–17

95.Valdovinos, F.S. (2019). Mutualistic networks: moving closer to a predictive theory. *Ecol. Lett.*

96.Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A., Galetti, M., *et al.* (2015). Beyond species loss: The extinction of ecological interactions in a changing world. *Funct. Ecol.*, 29, 299–307

97.Vázquez, D.P., Bluthgen, N., Cagnolo, L. & Chacoff, N.P. (2009). Uniting pattern and process in plant-animal mutualistic networks: A review. *Ann. Bot.*, 103, 1445–1457

98.Vizentin-Bugoni, J., Maruyama, P.K., Debastiani, V.J., Duarte, L. da S., Dalsgaard, B. & Sazima, M. (2016). Influences of sampling effort on detected patterns and structuring processes of a Neotropical plant-hummingbird network. *J. Anim. Ecol.*, 85, 262–272

99.Vizentin-Bugoni, J., Maruyama, P.K. & Sazima, M. (2014). Processes entangling interactions in communities: forbidden links
are more important than abundance in a hummingbird-plant network. *Proc. R. Soc. B Biol. Sci.*, 281, 20132397

100.Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440–442

101.Weinstein, B.G. & Graham, C.H. (2017). On comparing traits and abundance for predicting species interactions with imperfect detection. *Food Webs*, 11, 17–25