

The missing link: discerning true from false negatives when sampling species interaction networks

[Michael D. Catchen](#)^{1,2} [Timothée Poisot](#)^{3,2} [Laura Pollock](#)^{1,2} [Andrew Gonzalez](#)^{1,2}

¹ McGill University ² Québec Centre for Biodiversity Sciences ³ Université de Montréal

Correspondance to:

Michael D. Catchen — michael.catchen@mcgill.ca

Abstract: Ecosystems are composed of networks of interacting species. These interactions allow communities of species to persist through time through both neutral and adaptive processes. Still a robust understanding of (and ability to predict and forecast) interactions among species remains elusive. This knowledge-gap is largely driven by a shortfall of data—although species occurrence data has rapidly increased in the last decade, species interaction data has not kept pace, largely due to the intrinsic difficulty and effort required to sample interactions. These sampling challenges bias data and hinder inferences about the structure and dynamics of interactions networks. Here, we demonstrate the realized false-negative rate (the percentage of species that actually interact but for which we do not yet have a record) can be quite high, even in thoroughly sampled systems, due to the intrinsic variation in abundances across species in a community. We illustrate how a null model of occurrence detection can be used to estimate the false-negative rate in a given dataset. One hypothesis is that interactions between “rare” species are themselves rare because these species are less likely to encounter one-another than species of higher relative abundance. However, we demonstrate that across several datasets of spatial or temporally replicated networks, there are positive associations that suggest these interactions actually exist but just are not observed. Finally, we assess how false negatives influence various models of network prediction, and recommend directly accounting for observation error in predictive models. We conclude by discussing how the understanding of false-negatives can inform how we design monitoring schemes for species interactions.

1 Introduction

2 Species interactions drive many processes in evolutionary biology and community ecology. A better
3 understanding of interactions among species is an imperative to both mitigate the potentially harmful
4 impacts of anthropogenic change on Earth's biodiversity (Makiola *et al.* 2020) and to predict zoonotic
5 spillover of disease to prevent future pandemics (Becker *et al.* 2021). However, meeting these challenges is
6 difficult because interactions are intrinsically hard to sample (Jordano 2016). Over the past few decades
7 biodiversity data has become increasingly available—for example, remote-sensing has enabled collection
8 of data on spatial scales and resolutions previously unimaginable (Stephenson 2020), while the adoption
9 of open data practices (Kenall *et al.* 2014) have substantially increased the amount of data available to
10 ecologists. Still, widespread data about species interactions remains elusive (Poisot *et al.* 2021). Observing
11 an interaction between two species often requires human observation because remote sampling methods
12 can primarily detect co-occurrence (but see Niedballa *et al.* (2019)), and this itself is not necessarily
13 indicative of an interaction (Blanchet *et al.* 2020). This constraint induces biases on species interaction
14 data subject to the spatial and temporal scales that current observation methods can feasibly sample. This
15 is further compounded by semantic confusion around the word “interaction”—for example one might
16 consider competition a type of species interaction, even though it is marked by a lack of co-occurrence
17 between species, unlike other types of interactions, like trophism or pollination, which require both
18 species to be together at the same place and time. We define interaction in the latter sense, where two
19 species have fitness consequences on one-another if they are in the sample place at the same time. In
20 addition, here we only consider direct (not higher-order) interactions.

21 The importance of *sampling effort* and its impact on resulting ecological data has produced a rich body of
22 literature. The recorded number of species in a dataset or sample depends on the total number of
23 observations (Walther *et al.* 1995; Willott 2001)—as do estimates of population abundance (Griffiths
24 1998)—in addition to spatial coverage and species detectability. This has motivated more quantitatively
25 robust approaches to account for error in sampling data in many contexts: to determine if a given species
26 is extinct (Boakes *et al.* 2015), to determine sampling design (Moore & McCarthy 2016), and to measure
27 species richness across large scales (Carlson *et al.* 2020). In the context of interactions, an initial concern
28 was the compounding effects of limited sampling effort combined with the amalgamation of data (across
29 both study sites, time of year, and taxonomic scales) could lead any empirical set of observations to

30 inadequately reflect the reality of how species interact (Paine 1988) or the structure of the network as a
31 whole (Martinez *et al.* 1999; McLeod *et al.* 2021). Martinez *et al.* (1999) showed that in a plant-endophyte
32 trophic network, network connectance is robust to sampling effort, but this was done in the context of a
33 system for which observation of 62,000 total interactions derived from 164,000 plant-stems was feasible. In
34 some systems (e.g. megafauna food-webs) this many observations is either impractical or infeasible due to
35 the absolute abundance of the species in question.

36 We cannot feasibly observe all (or even most) of the interactions that occur in an ecosystem. This means
37 we can be confident two species actually interact if we have a record of it (given an estimate of species
38 misidentification probability), but not at all confident that a pair of species *do not* interact if we have *no*
39 *record* of those species observed together. In other words, it is difficult to distinguish *true-negatives* (two
40 species never interact) from *false-negatives* (two species interact sometimes, but we do not have a record of
41 it). For a concrete example of a false-negative in a food web, see fig. 1. Because even the most highly
42 sampled systems will still contain missing interactions, there is increasing interest in combining
43 species-level data (e.g. traits, abundance, range, phylogenetic relatedness, etc.) to build models to predict
44 interactions between species we haven't observed together before (Strydom *et al.* 2021). However, the
45 noise of false-negatives could impact the efficacy of our predictive models and have practical
46 consequences for answering questions about interactions (de Aguiar *et al.* 2019). This data constraint is
47 amplified as the interaction data we have is geographically biased toward the usual suspects (Poisot *et al.*
48 2021). We therefore need a statistical approach to assessing these biases in the observation process and
49 their consequences for our understanding of interaction networks.

50 The intrinsic properties of ecological communities create several challenges for sampling: first, species are
51 not observed with equal probability—we are much more likely to observe a species of high abundance
52 than one of very low abundance (Poisot *et al.* 2015). Canard *et al.* (2012) presents a null model of food-web
53 structure where species encounter one-another in proportion to each species' relative-abundance. This
54 assumes that there are no associations in species co-occurrence due to an interaction (perhaps because
55 this interaction is “important” for both species; Cazelles *et al.* (2016)), but in this paper we later show
56 increasing strength of associations leads to increasing probability of false-negatives in interaction data,
57 and that these positive associations are rampant in existing network data. Second, observed co-occurrence
58 is often equated with meaningful interaction strength, but this is not necessarily the case (Blanchet *et al.*
59 2020)—a true “non-interaction” would require that neither of two species, regardless of whether they

60 co-occur, ever exhibit any meaningful effect on the fitness of the other. So, although co-occurrence is not
61 directly indicative of an interaction, it is a precondition for an interaction.

62 Here, we illustrate how our confidence that a pair of species never interacts highly depends on sampling
63 effort, and suggest that surveys of species interactions can benefit from simulation modeling of the
64 sampling process. We demonstrate that the realized false-negative rate of interactions is directly related to
65 the relative abundance of the species pool, and demonstrate how simulation can be used to produce a null
66 estimate of the false-negative rate given total sampling effort (the total count of all individuals of all
67 species seen). We show that positive associations in co-occurrence data can increase the realized number
68 of false-negatives, and demonstrate these positive associations are rampant in network datasets. We
69 conclude by recommending that the simulation of sampling effort and species occurrence can and should
70 be used to help design surveys of species interaction diversity (Moore & McCarthy 2016), and by
71 advocating use of null models like those presented here as a tool for both guiding design of surveys of
72 species interactions and for modeling detection error in predictive models.

73 [Figure 1 about here.]

74 **How many observations of a non-interaction do we need to be confident** 75 **it's a true negative?**

76 We start with a naive model of interaction detection: we assume that every interacting pair of species is
77 incorrectly observed as not-interacting with an independent and fixed probability, which we denote p_{fn}
78 and subsequently refer to as the False-Negative Rate (FNR). If we observe the same species not-interacting
79 N times, then the probability of a true-negative (denoted p_{tn}) is given by $p_{tn} = 1 - (p_{fn})^N$. This relation
80 (the probability-mass-function of geometric distribution, a special case of the negative-binomial
81 distribution) is shown in fig. 2(A) for varying values of p_{fn} and illustrates a fundamental link between our
82 ability to reliably say an interaction doesn't exist— p_{tn} —and the number of times N we have observed a
83 given species. In addition, note that there is no non-zero p_{fn} for which we can ever *prove* that an
84 interaction does not exist—no matter how many observations of non-interactions N we have, $p_{tn} < 1$.

85 From fig. 2(A) it is clear that the more often we see two species co-occurring, but *not interacting*, the more
86 likely the interaction is a true-negative. This has several practical consequences: first it means negatives

87 taken outside the overlap of the range of each species aren't informative because co-occurrence was not
 88 possible, and therefore neither was an interaction. Second, we can use this relation to compute the
 89 expected number of total observations needed to obtain a “goal” number of observations of a particular
 90 pair of species (fig. 2(B)). As an example, if we hypothesize that A and B do not interact, and we want to see
 91 species A and B both co-occurring and *not interacting* 10 times to be confident this is a true negative, then
 92 we need an expected 1000 observations of all species if the relative abundances of A and B are both 0.1.
 93 Because the true FNR is latent, we can never actually be sure what the actual number of false negatives in
 94 our data—however, we can use simulation to estimate it for datasets of a given size using neutral models
 95 of observation. If some of the “worst-case” FNRs presented in fig. 2(A) seem unrealistically high, consider
 96 that species are observed in proportion to their relative abundance. In the next section we demonstrate
 97 that the distribution of abundance in ecosystems can lead to very high realized values of FNR (p_{fn}) simply
 98 as an artifact of sampling effort.

99 [Figure 2 about here.]

100 **False-negatives as a product of relative abundance**

101 We now show that the realized FNR changes drastically with sampling effort due to the intrinsic variation
 102 of the abundance of individuals of each species within a community. We do this by simulating the process
 103 of observation of species interactions, applied both to 243 empirical food webs from the Mangal database
 104 (Banville *et al.* 2021) and random food-webs generated using the niche model, a simple generative model
 105 of food-web structure that accounts for allometric scaling (Williams & Martinez 2000). Our neutral model
 106 of observation assumes each observed species is drawn in proportion to each species' abundance at that
 107 place and time. The abundance distribution of a community can be reasonably well described by a
 108 log-normal distribution (Volkov *et al.* 2003). In addition to the log-normal distribution, we also tested the
 109 case where the abundance distribution is derived from power-law scaling $Z^{(\log(T_i)-1)}$ where T_i is the
 110 trophic level of species i and Z is a scaling coefficient (Savage *et al.* 2004), which yields the same
 111 qualitative behavior. The practical consequence of abundance distributions spanning many orders of
 112 magnitude of abundance is that observing two “rare” species interacting requires two low probability
 113 events: observing two rare species *at the same time*.

114 To simulate the process of observation, for an ecological network M with S species, we sample abundances
 115 for each species from a standard-log-normal distribution. For each true interaction in the adjacency
 116 matrix M (i.e. $M_{ij} = 1$) we estimate the probability of observing both species i and j at a given place and
 117 time by simulating n observations of all individuals of any a species, where the species of the individual
 118 observed at the $\{1, 2, \dots, n\}$ -th observation is drawn from the generated log-normal distribution of
 119 abundances. For each pair of species (i, j) , if both i and j are observed within the n -observations, the
 120 interaction is tallied as a true positive if $M_{ij} = 1$. If only one of i or j are observed—but not both—in these
 121 n observations, but $M_{ij} = 1$, this is counted as a false-negative, and a true-negative otherwise. For each
 122 pair of species (i, j) , if both i and j are observed within the n -observations, the interaction is tallied as a
 123 true positive if $M_{ij} = 1$. If only one of i or j are observed—but not both—in these n observations, but
 124 $M_{ij} = 1$, this is counted as a false-negative, and a true-negative otherwise ($M_{ij} = 0$).
 125 In fig. 2(C) we see this model of observation applied to niche model networks across varying levels of
 126 species richness, and in fig. 2(D) the observation model applied to Mangal food webs. For all niche model
 127 simulations in this manuscript, for a given number of species S the number of interactions is drawn from
 128 the flexible-links model fit to Mangal data (MacDonald *et al.* 2020), effectively drawing the number of
 129 interactions L for a random niche model food-web as

$$L \sim \text{BetaBinomial}(S^2 - S + 1, \mu\phi, 1 - \mu\phi)$$

130 where the MAP estimate of (μ, ϕ) applied to Mangal data from (MacDonald *et al.* 2020) is
 131 ($\mu = 0.086, \phi = 24.3$). All simulations were done with 500 independent replicates of unique niche model
 132 networks per unique number of observations n . All analyses presented here are done in Julia v1.8
 133 (Bezanson *et al.* 2015) using both EcologicalNetworks.jl v0.5 and Mangal.jl v0.4 (Banville *et al.* 2021) and
 134 are hosted at (GITHUB_LINK_TODO). Note that the empirical data, for the reasons described above, very
 135 likely already contains many false negatives, we'll revisit this issue in the final section.

136 From fig. 2(C) it is evident that the number of species considered in a study is inseparable from the
 137 false-negative rate in that study, and this effect should be taken into account when designing samples of
 138 ecological networks in the future. We see a similar qualitative pattern in fig. 2(D) where the FNR drops off
 139 quickly as a function of observation effort, mediated by total richness. The practical consequence of the
 140 bottom row of fig. 2 is whether the total number of observations of all species (the x-axis) for the threshold

141 FNR we deem acceptable (the y-axis) is feasible. This raises two points: first, empirical data on
142 interactions are subject to the practical limitations of funding and human-work hours, and therefore
143 existing data tend to fall on the order of hundreds or thousands observations of individuals per site. Clear
144 aggregation of data on sampling effort has proven difficult to find and a meta-analysis of network data and
145 sampling effort seems both pertinent and necessary, in addition to the effects of aggregation of interactions
146 across taxonomic scales (Gauzens *et al.* 2013; Giacomuzzo & Jordán 2021). This inherent limitation on
147 in-situ sampling means we should optimize where we sample across space so that for a given number of
148 samples, we obtain the maximum information possible. Second, what is meant by “acceptable” FNR? This
149 raises the question: does a shifting FNR lead to rapid transitions in our ability inference and predictions
150 about the structure and dynamics of networks, or does it produce a roughly linear decay in model efficacy?
151 We explore this in the next section.

152 We conclude this section by advocating for the use of neutral models similar to above to generate
153 expectations about the number of false-negatives in a data set of a given size. This could prove fruitful
154 both for designing surveys of interactions but also because we may want to incorporate models of
155 imperfect detection error into predictive interactions models, as Joseph (2020) does for species occurrence
156 modeling. Additionally, we emphasize that one must consider the context for sampling—is the goal to
157 detect a particular species (as in fig. 2(C)), or to get a representative sample of interactions across the
158 species pool? These arguments are well-considered when sampling individual species (Willott 2001), but
159 have not yet been adopted for designing samples of communities.

160 **Positive associations increase the false-negative rate**

161 This model above doesn’t consider the possibility that there are positive or negative associations which
162 shift the probability of observing two species together due to their interaction (Cazelles *et al.* 2016).
163 However, here we demonstrate that the probability of observing a false negative can be higher if there is
164 some positive association in the occurrence of species *A* and *B*. If we denote the probability that we
165 observe the co-occurrence of two species *A* and *B* as $P(AB)$ and if there is no association between the
166 marginal probabilities of observing *A* and observing *B*, denoted $P(A)$ and $P(B)$ respectively, then the
167 probability of observing their co-occurrence is the product of the marginal probabilities for each species,
168 $P(AB) = P(A)P(B)$. In the other case where there is some positive strength of association between

169 observing both A and B because this interaction is “important” for each species, then the probability of
 170 observation both A and B , $P(AB)$, is greater than $P(A)P(B)$ as $P(A)$ and $P(B)$ are not independent and
 171 instead are positively correlated, i.e. $P(AB) > P(A)P(B)$. In this case, the probability of observing a single
 172 false-negative in our naive model from fig. 2(A) is $p_{fn} = 1 - P(AB)$, which due to the above inequality
 173 implies $p_{fn} > 1 - P(A)P(B)$. This indicates an increasingly greater probability of a false negative as the
 174 strength of association gets stronger, $P(AB) \rightarrow P(AB) \gg P(A)P(B)$. However, this still does not consider
 175 variation in species abundance in space and time (Poisot *et al.* 2015). If positive or negative associations
 176 between species structure variation in the distribution of $P(AB)$ across space/time, then the
 177 spatial/temporal biases induced by data collection would further impact the realized false negative rate, as
 178 the probability of false negative would not be constant for each pair of species across sites.

179 To test for these positive associations in data we scoured Mangal for datasets with many spatial or
 180 temporal replicates of the same system. For each dataset, we compute the marginal probability $P(A)$ of
 181 occurrence of each species A across all networks in the dataset. For each pair of interacting species A and
 182 B , we then compute and compare the probability of co-occurrence if each species occurs independently,
 183 $P(A)P(B)$, to the empirical joint probability of co-occurrence, $P(AB)$. Following our analysis above, if
 184 $P(AB)$ is greater than $P(A)P(B)$, then we expect our neutral estimates of the FNR above to underestimate
 185 the realized FNR. In fig. 3, we see the difference between $P(AB)$ and $P(A)P(B)$ for the seven suitable
 186 datasets with enough spatio-temporal replicates and a shared taxonomic backbone (meaning all
 187 individual networks use common species identifiers) found on Mangal to perform this analysis. Further
 188 details about each dataset are reported in tbl. 1.

189 In each of these datasets, the joint probability of co-occurrence $P(AB)$ is decisively greater than our
 190 expectation if species co-occur in proportion to their relative abundance $P(A)P(B)$. This suggests that
 191 there may not be as many “neutrally forbidden links” (Canard *et al.* 2012) as we might think, and that the
 192 reason we do not have records of interactions between rare species is probably due to observation error.
 193 This has serious ramifications for the widely observed property of nestedness seen in bipartite networks
 194 (Bascompte & Jordano 2007)—perhaps the reason we have lots of observations between generalists is
 195 because they are more abundant, and this is particularly relevant as we have strong evidence that
 196 generalism drives abundance (Song *et al.* 2022), not vice-versa.

[Figure 3 about here.]

Table 1: This table describes the datasets used in the above analysis (Fig 2). The table reports the type of each dataset, the total number of networks in each dataset (N), the total species richness in each dataset (S), the connectance of each metaweb (all interactions across the entire spatial-temporal extent) (C), the mean species richness across each local network \bar{S} , the mean connectance of each local network \bar{C} , the mean β -diversity among overlapping species across all pairs of network species ($\bar{\beta}_{OS}$), and the mean β -diversity among all species in the metaweb ($\bar{\beta}_{WN}$). Both metrics are computed using KGL β -diversity (Koleff *et al.* 2003)

Network	Type	N	S	C	\bar{S}	\bar{C}	$\bar{\beta}_{OS}$	$\bar{\beta}_{WN}$
Kopelke <i>et al.</i> (2017)	Food Web	100	98	0.037	7.87	0.142	1.383	1.972
Thompson & Townsend (2000)	Food Web	18	566	0.014	80.67	0.049	1.617	1.594
Havens (1992)	Food Web	50	188	0.065	33.58	0.099	1.468	1.881
Ponisio <i>et al.</i> (2017)	Pollinator	100	226	0.079	23.0	0.056	1.436	1.870
Hadfield <i>et al.</i> (2014)	Host-Parasite	51	327	0.085	32.71	0.337	1.477	1.952
Closs & Lake (1994)	Food Web	12	61	0.14	29.09	0.080	1.736	1.864
CaraDonna <i>et al.</i> (2017)	Pollinator	86	122	0.18	21.42	0.312	1.527	1.907

The impact of false-negatives on network properties and prediction

Here, we assess the effect of false negatives on our ability to make predictions about interactions, as well as their effect on network structure. The prevalence of false-negatives in data is the catalyst for interaction prediction in the first place, and as a result methods have been proposed to counteract this bias (Stock *et al.* 2017; Poisot *et al.* 2022). However, it is feasible that the FNR in a given dataset is so high that it could induce too much noise for an interaction prediction model to detect the signal of possible interaction between species.

To test this we use the dataset from Hadfield *et al.* (2014) that describes host-parasite interaction networks sampled across 51 sites, and the same method as Strydom *et al.* (2021) to extract latent features for each species in this dataset based on applying PCA to the co-occurrence matrix. We then predict a metaweb (equivalent to predicting true or false for an interaction between each species pair, effectively a binary classification problem) from these species-level features using four candidate models for binary classification—three often used machine-learning (ML) methods (Boosted Regression Tree (BRT), Random Forest (RF), Decision Tree (DT)), and one naive model from classic statistics (Logistic Regression

(LR)). Each of the ML models are bootstrap aggregated (or bagged) with 100 replicates each. We partition the data into 80-20 training-test split, and then seed the training data with false negatives at varying rates, but crucially do nothing to the test data. We fit all of these models using MLJ.jl, a high-level Julia framework for a wide-variety of ML models (Blaom *et al.* 2020). We evaluate the efficacy of these models using two common measures of binary classifier performance: the area under the receiver-operator curve (ROC-AUC) and the area under the precision-recall curve (PR-AUC), for more details see Poisot (2022). Here, PR-AUC is slightly more relevant as it is a better indicator of prediction of false-negatives. The results of these simulations are shown in fig. 4(A&B).

[Figure 4 about here.]

One interesting result seen in fig. 4(A&B) is that the ROC-AUC value does not approach random in the same way the PR-AUC curve does as we increase the added FNR. The reason for this is that ROC-AUC is fundamentally not as useful a metric in assessing predictive capacity as PR-AUC. As we keep adding more false-negatives, the network eventually becomes a zeros matrix, and these models can still learn to predict “no-interaction” for all possible species pairs, which does far better than random guessing (ROC-AUC = 0.5) in terms the false positive rate (one of the components of ROC-AUC). This highlights a more broad issue of label class imbalance, meaning there are far more non-interactions than interactions in data. A full treatment of the importance of class-balance is outside the scope of this paper, but is explored in-depth in Poisot (2022).

Although these ML models are surprisingly performant at link prediction given their simplicity, there have been several major developments in applying deep-learning methods to many tasks in network inference and prediction—namely graph-representation learning (GRL, Khoshraftar & An (2022)) and graph convolutional networks (Zhang *et al.* 2019). At this time, these advances can not yet be applied to ecological networks because they require far more data than we currently have. We already have lots of features that could be used as inputs into these models (i.e. species level data about occurrence, genomes, abundance, etc.), but our network datasets barely get into the hundreds of local networks sampled across space and time (tbl. 1). Once we start to get into the thousands, these models will become more useful, but this can only be done with systematic monitoring of interactions. This again highlights the need to optimize our sampling effort to maximize the amount of information contained in our data given the expensive nature of sampling interactions.

We also consider how the FNR affects network properties. In fig. 4(C) we see the mean trophic level across networks simulated using the niche model (as above), across a spectrum of FNR values. In addition to the clear dependence on richness, we see that mean trophic level, despite varying widely between niche model simulations, tends to be relatively robust to false negatives and does not deviate widely from the true value until very large FNRs, i.e. $p_{fn} > 0.7$. This is not entirely unsurprising. Removing links randomly from a food-web is effectively the inverse problem of the emergence of a giant component (more than half of the nodes are in a connected network) in random graphs (see Li *et al.* (2021) for a thorough review). The primary difference being that we are removing edges, not adding them, and thus we are witnessing the dissolution of a giant component, rather than the emergence of one. Further applications of percolation theory to the topology of ecological networks could improve our understanding of how false-negatives impact the inferences about the structure and dynamics on these networks.

Discussion

Species interactions enable the persistence and functioning of ecosystems, but our understanding of interactions is limited due to the intrinsic difficulty of sampling. Here we have provided a null model for the expected number of false-negatives in an interaction dataset. We demonstrated that we expect many false-negatives in species interaction datasets purely due to the intrinsic variation of abundances within a community. We also, for the first time to our knowledge, measured the strength of association between co-occurrence and interactions (Cazelles *et al.* 2016) across many empirical systems, and found that these positive associations are both very common, and showed algebraically that they increase the realized FNR. We have also shown that false-negatives could further impact our ability to both predict interactions and infer properties of the networks, which highlights the need for further research into methods for correcting this bias in existing data.

A better understanding of how false-negatives impact species interaction data is a practical necessity—both for inference of network structure and dynamics, but also for prediction of interactions by using species level information. False-negatives could pose a problem for many forms of inference in network ecology. For example, inferring the dynamic stability of a network could be prone to error if the observed network is not sampled “enough.” What exactly “enough” means is then specific to the application, and should be assessed via methods like those here when designing samples. Further,

269 predictions about network rewiring (Thompson & Gonzalez 2017) due to range shifts in response to
270 climate change could be error-prone without accounting for interactions that have not been observed but
271 that still may become climatically infeasible. As is evident from fig. 2(A), we can never guarantee there are
272 no false-negatives in data. In recent years, there has been interest toward explicitly accounting for
273 false-negatives in models (Stock *et al.* 2017; Young *et al.* 2021), and a predictive approach to
274 networks—rather than expecting our samples to fully capture all interactions (Strydom *et al.* 2021). As a
275 result, better models for predicting interactions are needed for interaction networks. This includes
276 explicitly accounting for observation error (Johnson & Larremore 2021)—certain classes of models have
277 been used to reflect hidden states which account for detection error in occupancy modeling (Joseph 2020),
278 and could be integrated in the predictive models of interactions in the future.

279 A brief caveat here is that we do not consider false-positives—in large part false-positives can be explained
280 by misidentification of species, although this could be a relevant consideration in some cases. The same
281 logic that we apply to false-negatives could easily be applied to false-positives, e.g. that we can be much
282 more confident that an interaction is a true positive if we have observed it 50 times rather than only once,
283 and we could similarly model this using the geometric distribution as in fig. 2(A). However, because
284 ecological networks are so sparse, there are far more negatives than positives in the dataset, and therefore
285 likely to be far more false-negatives than false-positives in absolute terms.

286 This work has several practical consequences for the design of interaction samples. Simulating the process
287 of observation could be a powerful tool for estimating the sampling effort required by a study that takes
288 relative abundance into account, and provides a null baseline for expected FNR. It is necessary to take the
289 size of the species pool into account when deciding how many total samples is sufficient for an
290 “acceptable” FNR (fig. 2(C & D)). Further the spatial and temporal turnover of interactions means any
291 approach to sampling prioritization must be spatiotemporal. We demonstrated earlier that observed
292 negatives outside of the range of both species aren’t informative, and therefore using species distribution
293 models could aid in this spatial prioritization of sampling sites.

294 Our work highlights the need for a quantitatively robust approach to sampling design, both for
295 interactions (Jordano 2016) and all other aspects of biodiversity (Carlson *et al.* 2020). As anthropogenic
296 forces create rapid shifts in our planet’s climate and biosphere, this is an imperative to maximize the
297 amount of ecological information we get in our finite samples, and make our inferences and decisions
298 based on this data as robust as possible. Where we choose to sample, and how often we choose to sample

299 there, has strong impacts on the inferences we make from data. Incorporating a better understanding of
300 sampling effort and bias to the design of biodiversity monitoring systems, and the inference and predictive
301 models we apply to this data, is imperative in understanding how biodiversity is changing, and making
302 actionable forecasts about the future of ecological interactions on our planet.

303 References

- 304 Banville, F., Vissault, S. & Poisot, T. (2021). Mangal.jl and EcologicalNetworks.jl: Two complementary
305 packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6, 2721.
- 306 Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.
307 *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.
- 308 Becker, D.J., Albery, G.F., Sjodin, A.R., Poisot, T., Bergner, L.M., Dallas, T.A., *et al.* (2021). Optimizing
309 predictive models to prioritize viral discovery in zoonotic reservoirs.
- 310 Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B. (2015). Julia: A Fresh Approach to Numerical
311 Computing.
- 312 Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.
313 *Ecology Letters*, 23, 1050–1063.
- 314 Blaom, A.D., Kiraly, F., Lienart, T., Simillides, Y., Arenas, D. & Vollmer, S.J. (2020). MLJ: A Julia package
315 for composable machine learning. *Journal of Open Source Software*, 5, 2704.
- 316 Boakes, E.H., Rout, T.M. & Collen, B. (2015). Inferring species extinction: The use of sighting records.
317 *Methods in Ecology and Evolution*, 6, 678–687.
- 318 Canard, E., Mouquet, N., Marescot, L., Gaston, K.J., Gravel, D. & Mouillot, D. (2012). Emergence of
319 Structural Patterns in Neutral Trophic Networks. *PLOS ONE*, 7, e38295.
- 320 CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M., *et al.* (2017).
321 Interaction rewiring and the rapid turnover of plantpollinator networks. *Ecology Letters*, 20, 385–394.
- 322 Carlson, C.J., Dallas, T.A., Alexander, L.W., Phelan, A.L. & Phillips, A.J. (2020). What would it take to
323 describe the global diversity of parasites? *Proceedings of the Royal Society B: Biological Sciences*, 287,
324 20201841.

325 Cazelles, K., Araújo, M.B., Mouquet, N. & Gravel, D. (2016). A theory for species co-occurrence in
326 interaction networks. *Theoretical Ecology*, 9, 39–48.

327 Closs, G.P. & Lake, P.S. (1994). Spatial and Temporal Variation in the Structure of an Intermittent-Stream
328 Food Web. *Ecological Monographs*, 64, 1–21.

329 de Aguiar, M.A.M., Newman, E.A., Pires, M.M., Yeakel, J.D., Boettiger, C., Burkle, L.A., *et al.* (2019).
330 Revealing biases in the sampling of ecological interaction networks. *PeerJ*, 7, e7566.

331 Gauzens, B., Legendre, S., Lazzaro, X. & Lacroix, G. (2013). Food-web aggregation, methodological and
332 functional issues. *Oikos*, 122, 1606–1615.

333 Giacomuzzo, E. & Jordán, F. (2021). Food web aggregation: Effects on key positions. *Oikos*, 130,
334 2170–2181.

335 Griffiths, D. (1998). Sampling effort, regression method, and the shape and slope of sizeabundance
336 relations. *Journal of Animal Ecology*, 67, 795–804.

337 Hadfield, J.D., Krasnov, B.R., Poulin, R. & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative
338 Analyses of Ecological Interactions. *The American Naturalist*, 183, 174–187.

339 Havens, K. (1992). Scale and Structure in Natural Food Webs. *Science*, 257, 1107–1109.

340 Johnson, E.K. & Larremore, D.B. (2021). Bayesian estimation of population size and overlap from random
341 subsamples.

342 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.

343 Joseph, M.B. (2020). Neural hierarchical models of ecological populations. *Ecology Letters*, 23, 734–747.

344 Kenall, A., Harold, S. & Foote, C. (2014). An open future for ecological and evolutionary data? *BMC*
345 *Evolutionary Biology*, 14, 66.

346 Khoshraftar, S. & An, A. (2022). A Survey on Graph Representation Learning Methods.

347 Koleff, P., Gaston, K.J. & Lennon, J.J. (2003). Measuring beta diversity for presenceabsence data. *Journal*
348 *of Animal Ecology*, 72, 367–382.

349 Kopelke, J.-P., Nyman, T., Cazelles, K., Gravel, D., Vissault, S. & Roslin, T. (2017). Food-web structure of
350 willow-galling sawflies and their natural enemies across Europe. *Ecology*, 98, 1730–1730.

Li, M., Liu, R.-R., Lü, L., Hu, M.-B., Xu, S. & Zhang, Y.-C. (2021). Percolation on complex networks: Theory and application. *Physics Reports*, Percolation on complex networks: Theory and application, 907, 1–68.

MacDonald, A.A.M., Banville, F. & Poisot, T. (2020). Revisiting the Links-Species Scaling Relationship in Food Webs. *Patterns*, 1.

Makiola, A., Compson, Z.G., Baird, D.J., Barnes, M.A., Boerlijst, S.P., Bouchez, A., *et al.* (2020). Key Questions for Next-Generation Biomonitoring. *Frontiers in Environmental Science*, 7.

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.

McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*, 130, 2250–2259.

Moore, A.L. & McCarthy, M.A. (2016). Optimizing ecological survey effort over space and time. *Methods in Ecology and Evolution*, 7, 891–899.

Niedballa, J., Wilting, A., Sollmann, R., Hofer, H. & Courtiol, A. (2019). Assessing analytical methods for detecting spatiotemporal interactions between species from camera trapping data. *Remote Sensing in Ecology and Conservation*, 5, 272–285.

Paine, R.T. (1988). Road Maps of Interactions or Grist for Theoretical Development? *Ecology*, 69, 1648–1654.

Poisot, T. (2022). Guidelines for the prediction of species interactions through binary classification.

Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., *et al.* (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography*, 48, 1552–1563.

Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M.J., Becker, D.J., Brierly, L., *et al.* (2022). Network embedding unveils the hidden interactions in the mammalian virome.

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

Ponisio, L.C., Gaiarsa, M.P. & Kremen, C. (2017). Opportunistic attachment assembles plantpollinator networks. *Ecology Letters*, 20, 1261–1272.

378 Savage, V.M., Gillooly, J.F., Brown, J.H., West, G.B. & Charnov, E.L. (2004). Effects of Body Size and
 379 Temperature on Population Growth. *The American Naturalist*, 163, 429–441.

380 Song, C., Simmons, B.I., Fortin, M.-J. & Gonzalez, A. (2022). Generalism drives abundance: A
 381 computational causal discovery approach. *PLOS Computational Biology*, 18, e1010302.

382 Stephenson, P. (2020). Technological advances in biodiversity monitoring: Applicability, opportunities
 383 and challenges. *Current Opinion in Environmental Sustainability*, Open issue 2020 part A: Technology
 384 Innovations and Environmental Sustainability in the Anthropocene, 45, 36–41.

385 Stock, M., Poisot, T., Waegeman, W. & De Baets, B. (2017). Linear filtering reveals false negatives in
 386 species interaction data. *Scientific Reports*, 7, 45908.

387 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). A
 388 roadmap towards predicting species interaction networks (across space and time). *Philosophical
 389 Transactions of the Royal Society B: Biological Sciences*, 376, 20210063.

390 Thompson, P.L. & Gonzalez, A. (2017). Dispersal governs the reorganization of ecological networks under
 391 environmental change. *Nature Ecology & Evolution*, 1, 1–8.

392 Thompson, R.M. & Townsend, C.R. (2000). Is resolution the solution?: The effect of taxonomic resolution
 393 on the calculated properties of three stream food webs. *Freshwater Biology*, 44, 413–422.

394 Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003). Neutral theory and relative species abundance
 395 in ecology. *Nature*, 424, 1035–1037.

396 Walther, B.A., Cotgreave, P., Price, R.D., Gregory, R.D. & Clayton, D.H. (1995). Sampling Effort and
 397 Parasite Species Richness. *Parasitology Today*, 11, 306–310.

398 Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

399 Willott, S.j. (2001). Species accumulation curves and the measure of sampling effort. *Journal of Applied
 400 Ecology*, 38, 484–486.

401 Young, J.-G., Valdovinos, F.S. & Newman, M.E.J. (2021). Reconstruction of plantpollinator networks from
 402 observational data. *Nature Communications*, 12, 3911.

403 Zhang, S., Tong, H., Xu, J. & Maciejewski, R. (2019). Graph convolutional networks: A comprehensive
 404 review. *Computational Social Networks*, 6, 11.

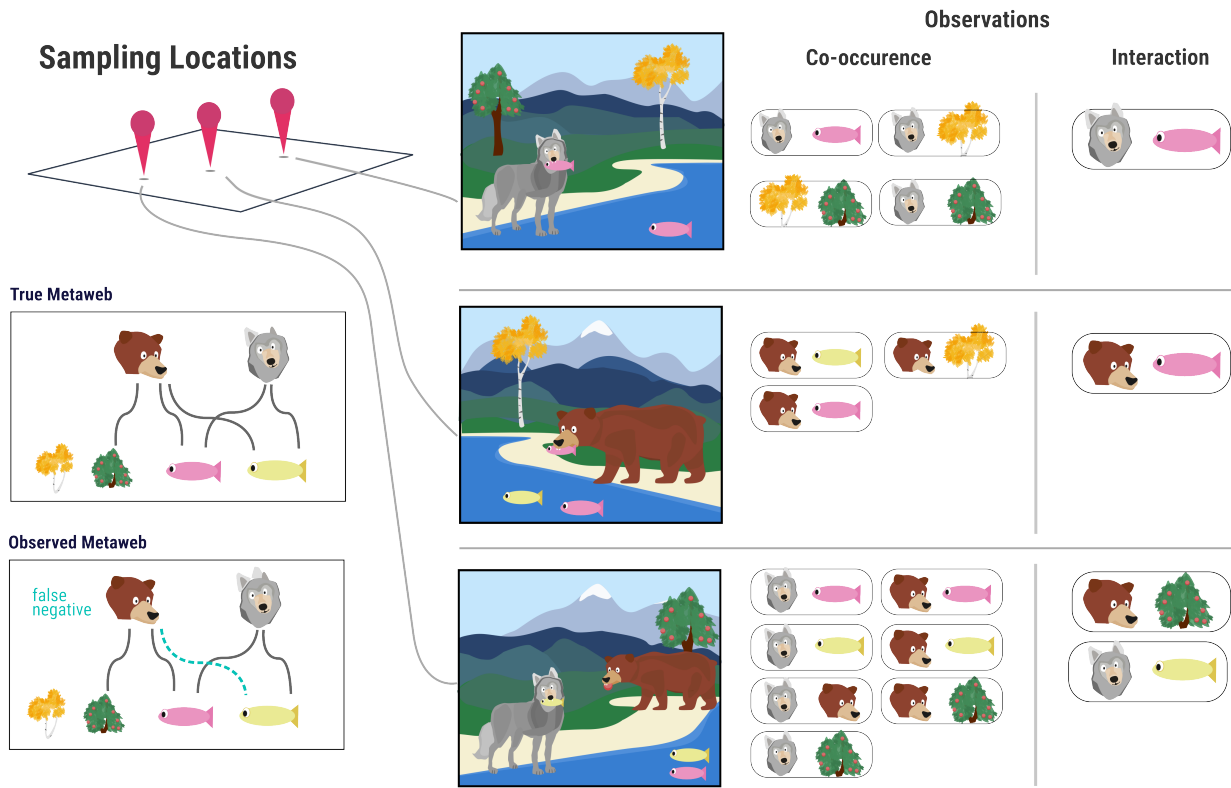


Figure 1: This conceptual example considers a sample of the trophic community of bears, wolves, salmon (pink fish), pike (yellow fish), berry trees, and aspen trees. The true metaweb (all realized interactions across the entire spatial extent) is shown on the left. In the center is what a hypothetical ecologist samples at each site. Notice that although bears are observed co-occurring with both salmon and pike, there was never a direct observation of bears eating pike, even though they actually do. Therefore, this interaction between bears and pike is a false negative.

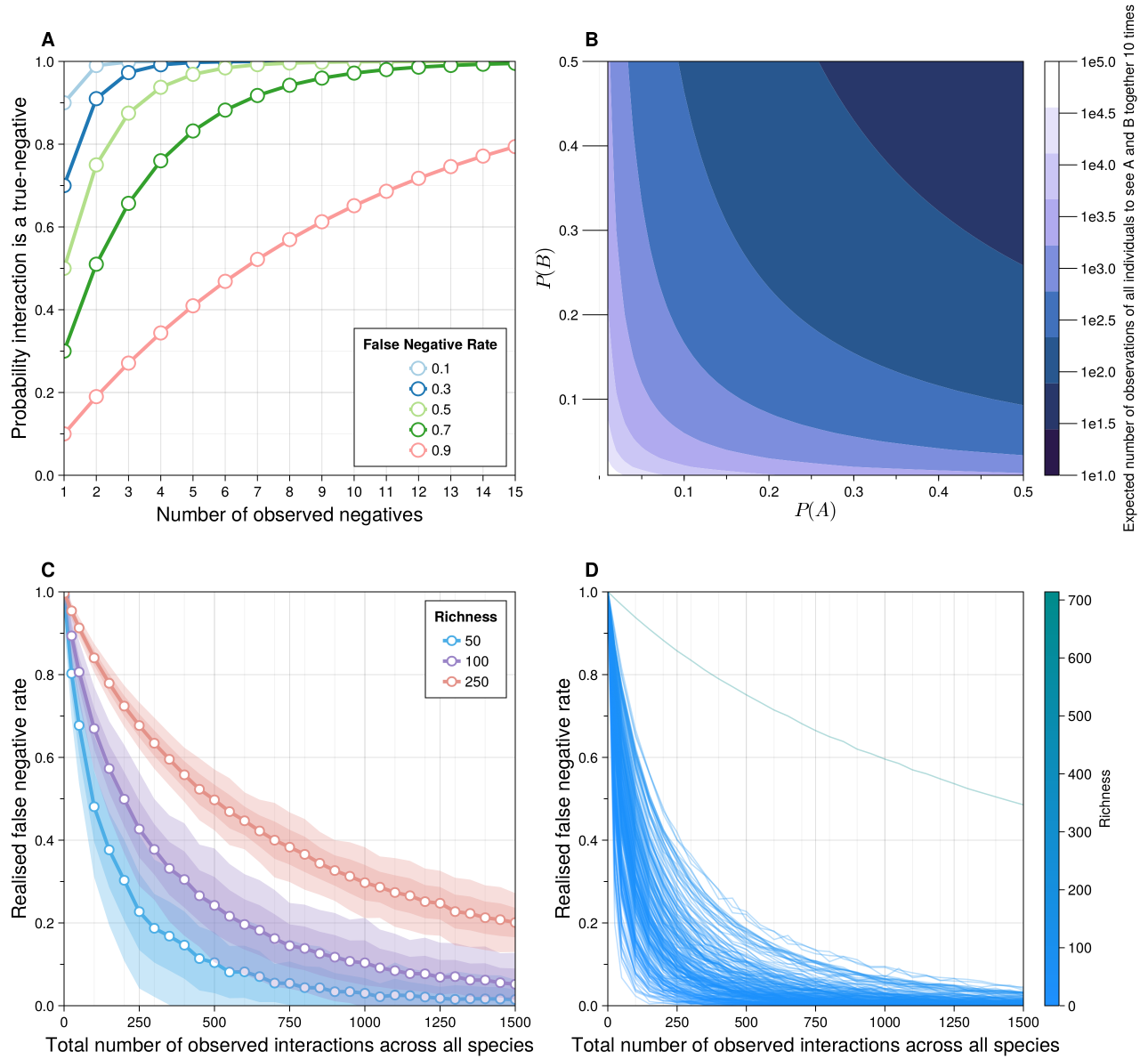


Figure 2: **(A)** The probability that an observed interaction is a true negative (y-axis) given how many times it has been sampled as a non-interaction (x-axis). Each color reflects a different value of p_{fn} , the false-negative rate (FNR)—this is effectively the cdf of the geometric distribution. **(B)** The expected number of total observations needed (colors) to observe 10 co-occurrences between a species with relative abundance $P(A)$ (x-axis) and a second species with relative abundance $P(Y)$. **(C)**: False negative rate (y-axis) as a function of total sampling effort (x-axis) and network size, computed using the method described above. For 500 independent draws from the niche model (Williams & Martinez (2000)) at varying levels of species richness (colors) with connectance drawn according to the flexible-links model (MacDonald *et al.* (2020)) as described in the main text. For each draw from the niche model, 200 sets of 1500 observations are simulated, for which each the mean false negative rate at each observation-step is computed. Means denoted with points, with 1 in the first shade and 2 in the second. **(D)**: Same as **(C)**, except using empirical food webs from Mangal database, where richness. The outlier on **(D)** is a 714 species food-web.

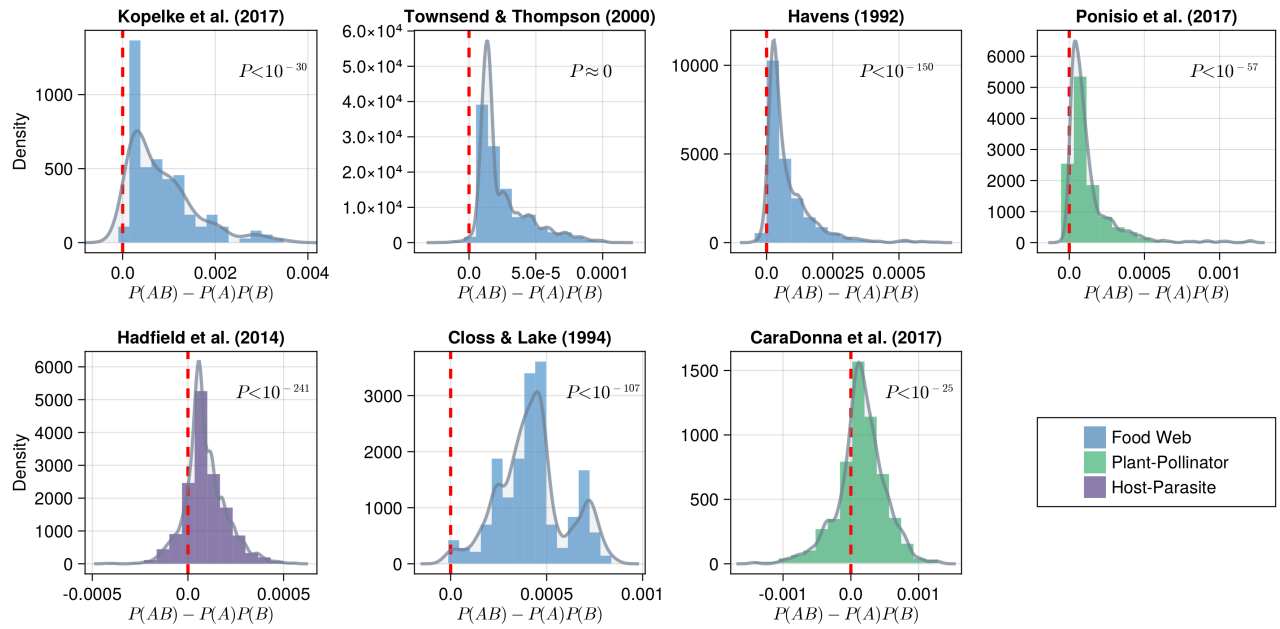


Figure 3: The difference between joint-probability of co-occurrence ($P(AB)$) and expected probability of co-occurrence under independence ($P(A)P(B)$) for interacting species for each dataset. The red-dashed line indicates 0 (no association). Each histogram represents a density, meaning the area of the entire curve sums to 1. The continuous density estimate (computed using local smoothing) is shown in grey. The p-value on each plot is the result of a one-sided t-test comparing the mean of each distribution to 0.

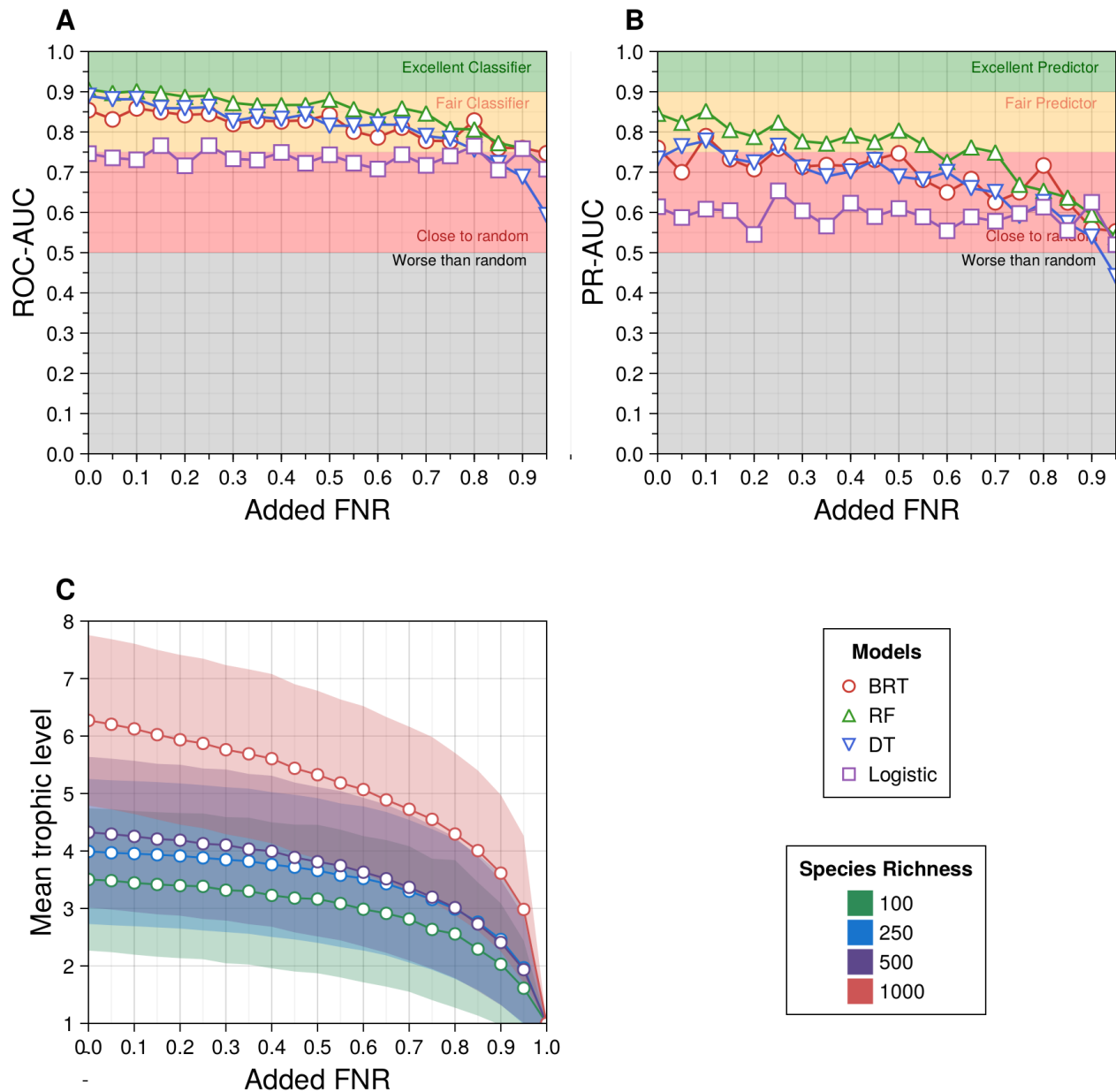


Figure 4: **(A)** The area-under the receiver-operator curve (ROC-AUC) and **(B)** The area-under the precision-recall curve (PR-AUC; right) for each different predictive model (colors/shapes) across a spectrum of the proportion of added false negatives (x-axis). **(C)** The mean trophic-level of all species in a network generated with the niche model across different species richnesses (colors). For each value of the FNR, the mean trophic level was computed across 50 replicates. The shaded region for each line is one standard-deviation across those replicates.