Contextual evidence for categorising interactions and guiding their discovery

Kesem Abramov [1], Shir Miryam Nehoray [1], Rami Puzis [2], Peter J. Mucha [3], and Shai Pilosof [1,4,*]

[1] *Department of Life Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel*
[2] *Department of Systems Information and Software Engineering, Ben-Gurion University of the Negev, Be'er Sheva, Israel*
[3] *Department of Mathematics, Dartmouth College, Hanover, NH, USA*
[4] *The Goldman Sonnenfeldt School of Sustainability and Climate Change, Ben-Gurion University of the Negev, Be'er Sheva, Israel*

August 19, 2026

*Corresponding author. Email: pilos@bgu.ac.il

Abstract

A surge of recent methods tackles incomplete ecological interaction data through link prediction. However, validating predictions means field-sampling an immense number of candidate interactions, possibly with a poorly suited method. This costly endeavour lacks operative guides grounded in strong ecological and statistical theory. We present a guided-sampling framework combining link prediction with within-system variability across replicated networks as contextual evidence, generating a fine-resolution, ecologically-aware link taxonomy. It resolves categories ecologists have long struggled to separate, such as forbidden interactions versus those merely missing from the local sample, pinpoints the most probable unknowns, and directs targeted, cost-efficient discovery. We formalise how confidence in each assignment is quantified and grows as evidence accumulates. Treating validation as evidence accumulation makes it a tractable sampling problem rather than a permanent limit on knowledge. We provide a worked analysis of an empirical data set and an online guide at http://lpguide.ecomplab.com.

Link prediction is incomplete without evidence

Ecological interactions are the backbone of ecosystem function [1]. However, our knowledge of interaction data, typically represented as links in ecological networks, is incomplete, constrained by finite sampling effort, methodological biases, and the cryptic nature of many species [2–5]. These gaps, which are collectively called the "Eltonian Shortfall", bias our representation of ecological networks and hamper our ability to predict how communities will respond to perturbations [6]. Given the central role of interactions for biodiversity and human well-being [7–10], there is a surge of attempts to predict unobserved links using an expanding array of link prediction approaches [11–17]. However, models return candidate interactions, not evidence that they exist. Ranking candidates by their plausibility within a network [18, 19], or interpreting which features drove a prediction [20], might narrow the search, but a reliability score is a model-internal verdict rather than independent evidence: its plausibility is assessed against the same network and method that produced it. The

result is typically a set of candidates far too large to verify exhaustively in the field, with no independent basis for deciding which to trust. This leaves modellers uncertain about whether a prediction is valid, and field ecologists with no operational guidance on how to act on it.

In this paper we propose that validation and interpretation of ecological link prediction should be framed as a process of **evidence accumulation**, in which confidence emerges through repeated, independent lines of evidence rather than from a single evaluation exercise. Evidence accumulation operates along two dimensions: contextual evidence, drawn from replicated networks sharing the same species pool and ecological context [21, 22], and methodological evidence, drawn from independent sampling methods with different detection abilities and biases [23, 24]. Building on these two lines of evidence, we develop a taxonomy for categorising predicted links and offer guidelines for directing sampling efficiently toward the evidence each category requires. We further show how confidence in a link's categorisation can be quantified and how it grows as evidence accumulates, and demonstrate the framework's application to a real ecological system. To facilitate the presentation and intuition we provide an interactive guide at `http://lpguide.ecomplab.com`.

Why current validation schemes fall short

The root of the problem of validating link predictions lies in how predictive performance is evaluated. Validation ideally requires ground truth (Box 1). Such ground truth is almost never available, because confirming an interaction's presence, and even more so its absence, is prohibitively costly. Standard practice therefore evaluates predictions against the observed interactions in the focal network alone, treating them as ground truth — a process we refer to as 'within-network evaluation' (Box 1). This substitution is the crux of the problem. A predicted interaction that is unobserved in the data is classified as a false positive and read as a candidate missing link the model has recovered. However, because many interactions go undetected even under intensive sampling [2, 4], such a prediction is ambiguous: it may be a real interaction the sampling missed, or a model error. The ambiguity runs deeper still: when a model correctly predicts an unobserved interaction not to exist, it flags it as a true negative. However, this single label conflates three ecologically distinct realities: a forbidden interaction that is biologically impossible [25], a feasible interaction unrealised in the focal network due to local conditions [26–28], or a real interaction missed during sampling. Within-network evaluation cannot distinguish between these scenarios.

For these reasons, when predictions are made, researchers occasionally attempt to validate them by searching for evidence in sources external to the system at hand, querying interaction databases such as GloBI [29] or the Web of Life [30], conducting literature searches [31, 32], or consulting experts [33]. However, such external evidence carries important limitations. Because interaction patterns vary across space, time, and ecological contexts [34–36], external records may confirm predicted links that do not occur in the focal system, simply because the ecological context has been stripped away. Taxonomic inconsistencies between datasets further complicate interpretation, and external evidence is often infeasible for systems with high endemicity, such as island communities [37]. Moreover, interaction similarity between networks decays with both geographical and compositional distance [35], driving a corresponding decay in the transferability of interaction information [37].

Box 1: Basics of within-network evaluation

The most common approach for assessing performance of link prediction is to compare model output to observed interactions, treated as ground truth, to quantify how well the model captures a network's interaction structure using appropriate metrics.

Evaluation therefore begins by splitting the data into training and testing sets. The standard way to do this is link withholding: a subset of observed links is removed, the model is trained on the remainder, and the withheld links are predicted and compared to their observed values. Both observed links and unobserved pairs can be withheld, so that predictive performance can be scored against both classes (see,

e.g., [38]). To use the data efficiently and obtain a stable estimate of performance, withholding is typically organised as k-fold cross-validation: the data are partitioned into k equally-sized folds, each withheld in turn while the model is trained on the remaining $k-1$, and performance is averaged across the k rounds [39]. Its extreme case is the leave-one-out design, where every link is withheld and predicted individually [39]. A specific use of the same machinery is not to estimate performance but to generate candidates by flagging the highest-scoring unobserved pairs as candidate missing links [19].

The predicted links are then evaluated in one of three ways. Threshold-based evaluation converts predicted probabilities into binary outcomes using a cutoff (often 0.5, or chosen by sensitivity analysis [37]): predictions above it are interactions, those below non-interactions. Performance is summarised in a confusion matrix of true positives, true negatives, false positives, and false negatives, from which metrics such as precision are derived [40]. Threshold-free evaluation instead avoids a single cutoff and summarises the model's ability to separate presences from absences across all thresholds, using measures such as the area under the precision–recall curve (PR-AUC) or the receiver operating characteristic curve (ROC-AUC) [17, 40]. When the target is predicting interaction strength (such as visitation frequency) rather than occurrence, observed and predicted values are compared directly through correlation or weighted performance metrics (NSE, NNSE) [37, 41].

Ecological networks are sparse, and this class imbalance complicates evaluation: a model can achieve high accuracy by correctly predicting non-interactions while still failing to recover true interactions [17, 40]. The imbalance can be addressed during the training phase, for example by assigning greater weight to the rarer interaction class, and during the evaluation phase, by choosing metrics that remain informative under imbalance rather than overall accuracy [40].

Critically, all of these approaches treat unobserved interactions as true absences against which predictions are evaluated, and assume observations to be a true realization. These are central assumptions that the evidence accumulation framework is designed to address.

A framework for evidence accumulation and guided sampling

We present a framework rooted in two complementary modes of accumulating additional evidence: obtaining contextual evidence and increasing methodological evidence.

Obtaining contextual evidence

A preferable, more ecologically informed alternative to validation using external sources is obtaining contextual evidence: leveraging replicated networks, sampled across space or time within the same system for guiding interaction discovery. Because replicates share the same species pool and ecological context as the focal network, they come from the same distribution as the focal network and therefore preserve the conditions under which interactions are or are not realised, which is the information that external evidence discards. Therefore, within-system evidence is less likely to produce false confirmations. What constitutes a valid replicate depends on the researcher's knowledge of the system [36, 42], with ecological similarity (e.g., species composition and environmental context) serving as the guiding principle [43]. Crucially, this is not a new practice: many studies already sample interaction networks repeatedly across sites or time points as part of their core research design [34, 36, 44].

Contextual evidence connects to two established lines of work: structure-aware (block) cross-validation [45], which concerns how a model's reliability is tested when the data have natural structure, and metawebs [15, 22, 28], regional catalogues of all interactions possible for a given species pool. From different directions, both show that a system's partition structure carries information: block cross-validation reveals that ignoring it in evaluation makes models look more reliable than they are, while the metaweb literature finds that pooling it away over-predicts local interactions and discards their variability. Prior to the actual application of a model, its performance is typically assessed by cross-validation, in which the data are split so that the model is trained

on certain parts and tested on others (Box 1). Roberts et al. [45] show that because ecological observations are non-independent in space, time, or phylogeny, random splitting leaks correlated observations across the train–test boundary, so predictive error is underestimated. Their solution is to split along the structure itself. In our case, that would be separating the information in the focal network from that in the replicates (see below). Banville et al. [28] formalised the distinction between an interaction's regional feasibility (its metaweb probability) and its local realisation, which additionally requires local co-occurrence and realisation where the partners meet; local probability is thus bounded above by the regional one, and equating the two over-predicts links. Downscaling a metaweb this way yields only an upper bound on local interactions [22, 28]. The variation between local networks reflects the gap between regional feasibility and local realisation, which Banville et al. [28] frame as a primary open challenge that must ultimately be validated in the field. We leverage this gap to generate contextual evidence and to provide operational guidance for targeted sampling.

A taxonomy of predicted links

How, then, can we raise our confidence in local-scale predictions? A predictive model can generate far more candidate links than any field programme can verify. Ranking them by predicted probability is a useful first move, but it orders the candidate set along a single, model-internal axis. It identifies which links the model favours, not what evidence any of them is missing, and therefore not what a field programme should do next. Two links with the same score can require incompatible actions: one has already been recorded in a neighbouring replicate and needs only further sampling locally, whereas the other has been seen nowhere in the system and will not be resolved by any amount of the same method. Researchers' knowledge of the ecosystem and methodology informs the actions that can be taken.

We perform link prediction on each network separately, preserving its local context. We then use between-network variation as evidence. Concretely, we add a third dimension to the classical confusion matrix (Box 1): whether a predicted interaction is observed in at least one replicate within the same system, yielding a taxonomy of eight link categories (Fig. 1). This approach resolves ambiguities invisible to within-network evaluation; for instance, whether an interaction absent from the focal system reflects a truly forbidden link or is simply locally absent (present in another replicate). It also refines the classical division of unobserved interactions into forbidden and missing links [4, 25, 26], a distinction that predates predictive methods and is too coarse for them. For example, a link classified as a true negative (unobserved and predicted absent) is *possibly forbidden* if it is seen nowhere in the system, but *locally absent* if it appears in another replicate. For simplicity, we first discuss a deterministic (error-free) categorisation; however, each of the three components of evidence carries some error. We then provide a quantitative Bayesian framework that propagates the uncertainty of the evidence into a posterior confidence for each categorisation, and show how confidence grows with the number of replicates in which the prediction is consistently supported, whether through repeated observation or repeated absence.

The taxonomy we propose helps guide sampling and modelling decisions (Fig 2). Among predicted links, *recurrent links* — predicted and observed locally and in replicates — likely require no further action: consistent observation across independent networks indicate compelling evidence. Recurrent links may involve abundant or prominent species, but can also reflect strong partner fidelity in specialists that consistently interact with the same partners across environments [35, 46]. *Locally unique* links are predicted and observed only in the focal network. They may arise through several non-mutually exclusive mechanisms: species rarity reducing encounter probability [4, 47], high endemicity driving species turnover [35], or interaction rewiring under local conditions such as reduced availability of alternative partners [48]. Local uniqueness might be, in many cases, an ecological finding rather than a failure to obtain evidence. *Possibly missing* links are predicted but unobserved locally, yet observed in replicates. This is arguably the most actionable category, as the predicted interaction is demonstrably observed elsewhere, making insufficient local sampling the most parsimonious explanation; however, phenological mismatches or other local biological constraints can also lead to a locally unobserved link [49]. Targeted detection efforts using the same or complementary sampling methods can discriminate between these alternatives [5, 50]. *Phantom* links are predicted but never observed anywhere, requiring independent

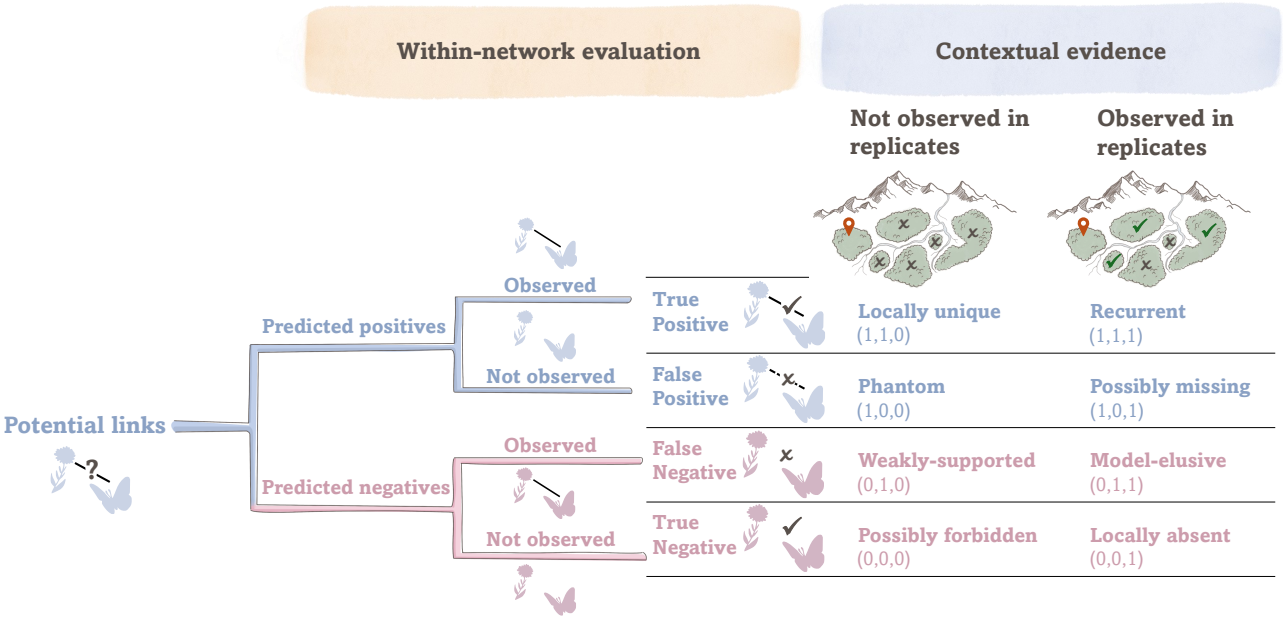

Fig. 1: A taxonomy of predicted links. Predicted ecological interactions are assigned to one of eight link categories using three sources of evidence E: whether a link is predicted by the model (Y), observed in the focal network (O_l), and observed in at least one within-system replicate (O_r). In the simple deterministic version, each source is a single bit, taking the value 1 when it supports the link and 0 when it does not. For example, a *locally unique* link is predicted and observed in the focal network, but recorded in no replicate, giving $E = (Y, O_l, O_r) = (1, 1, 0)$. Within-network evaluation alone (left) uses only Y and O_l, yielding the four classical confusion matrix classes: true positives, false positives, true negatives, and false negatives (Box 1). Adding contextual evidence from within-system replicates (right) resolves these into eight categories, each carrying a distinct ecological interpretation. Section Bayesian Inference attaches a confidence to the assignment by allowing each source of evidence to carry an error. The Supplementary Information states and relaxes the assumptions behind this simple version, grounds the error rates in realisation and detection, and lets confidence accumulate over replicates. We illustrate the framework using a spatial system, though it can also apply to temporal or other forms of replication.

methodological evidence to distinguish model artefacts from systematically missed interactions. This distinction is particularly important when the predictive models are based solely on network structure [14, 18, 37], lacking explicit ecological features such as phylogeny or species traits [11, 13, 21].

Among unpredicted links, the taxonomy gives empirical meaning to the theoretical distinction between metawebs and realised networks [28, 51]. *Locally absent* links are unobserved locally but detected in replicates, confirming biological feasibility while attributing local absence to ecological context. This is consistent with documented spatial variability in interaction structure [35, 36, 52]. *Possibly forbidden* links remain absent everywhere and are unsupported by prediction. Links in this category are strong candidates for genuinely impossible interactions constrained by morphological, physiological, or phenological mismatches [25]. However, 'forbiddenness' is better viewed as a continuum than a binary fact [26]: intraspecific trait variability means that an interaction predicted to be forbidden from mean trait values may occur in some individuals, populations, or years. Consistent absence across replicates spanning varied local conditions therefore provides stronger (though never definitive [27]) evidence of true biological constraints than any single-network assessment alone. Nevertheless, absence can still result from a method that cannot detect the interaction, which a complementary method might reveal. *Weakly-supported* links are observed locally but are neither predicted (locally) nor detected in replicates. They may therefore reflect observation or identification errors [53, 54], extremely rare or sporadic interactions, or genuine context-specific realisations that both the model and the broader system fail to capture. Finally, *model-elusive* links are observed locally and across replicates but not predicted. When rare, they reveal specific blind spots. When common, they likely signal a more fundamental model failure requiring revision rather than further sampling.

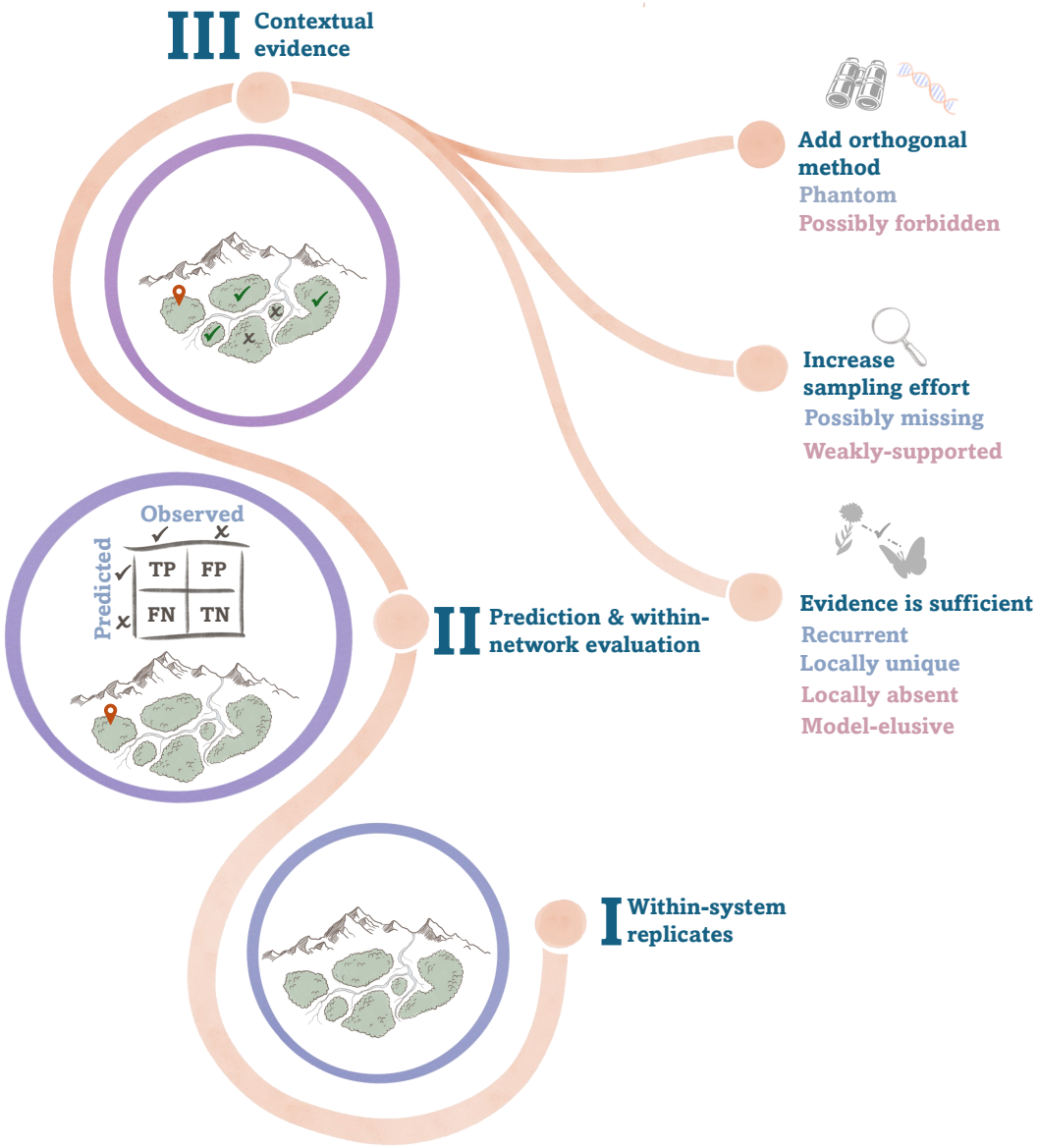

Fig. 2: A framework for evidence accumulation. Evidence accumulation is a sequential process. First, contextual evidence is drawn from within-system replicates, categorising each link. Then, methodological evidence can be added, either by increasing sampling effort with the same method or by adding methods with different detection biases. For each link category, the framework suggests an action: whether no further action is needed, sampling effort should be increased, or independent methods should be used. Link categories are colour-coded consistently across this figure and Fig. 1.

Accumulating additional methodological evidence

Field ecologists must make the most of limited resources, which means narrowing the set of interactions worth sampling. Link prediction is a first step: it produces a set of candidate links that can be ranked by their probabilities [19]. Contextual evidence narrows that set further, setting specific, ecologically interpretable detection goals — for example, targeting locally absent or possibly forbidden links. Contextual evidence can be applied without predictions, but the ecological interpretation is then less resolved. For instance, without a prediction, possibly missing and locally absent links become indistinguishable, as both are unobserved locally yet present in replicates.

Contextual evidence, though necessary, is not always sufficient. Replicates sampled with the same method share its biases: e.g., an interaction consistently missed by visual observation goes undetected across every replicate, and a prediction model trained on structurally similar networks reproduces the same blind spots [23, 24, 55]. Evidence accumulation therefore requires a second stage: improving the methodological evidence (Fig. 2), comprising two complementary possible actions. The first is to increase sampling effort with the same method, appropriate mainly when that method has already detected the interaction in replicates. The second is to add orthogonal methods with complementary detection biases [24, 56], to overcome the limits of any single method [23, 57]. The two kinds of evidence are applied in sequence, considering contextual evidence first because only by categorising a link through within-system replication does the appropriate methodological action become clear (Fig. 1). We demonstrate the application of the framework to a real plant–pollinator system below (Section Empirical demonstration).

Together, these two kinds of evidence make the taxonomy actionable rather than merely descriptive. Each category points to a specific follow-up action, or, for well-supported links, to accepting the current evidence as sufficient (Fig. 2). As a general guideline, we recommend the following logic. For *recurrent*, *locally unique*, *locally absent*, and *model-elusive* links, no further sampling is likely needed. In those cases, the categorisation is already interpretable (though frequent model-elusive links point to a need for model revision rather than more data). For *possibly missing* and *weakly-supported* links, increased sampling effort at the focal site is likely sufficient and targeted. For *phantom* and *possibly forbidden* links, independent methods are necessary, because intensifying the same method will likely not help. Nevertheless, the course of action rests ultimately with the researchers, who know the boundaries of their system and can weigh the cost of adding methods against that of intensifying sampling. Familiarity with a system also lets researchers translate contextual evidence into mechanism, understanding for instance why a predicted interaction is unique to one network yet absent from others.

Three additional considerations inform how the framework is applied in practice. First, the logic of contextual evidence can in principle be transposed to the methodological dimension, yielding an analogous taxonomy. However, this is rarely equivalent, and is often less feasible. Researchers typically have access to far fewer independent methods than replicates, and methods are not ecologically exchangeable in the way replicates are. A link detected by DNA barcoding but not by visual observation carries a different ecological meaning than a link observed in one location but not another. In our view, contextual evidence therefore remains the more accessible foundation for this taxonomic framework. Second, within the methodological dimension, the framework prescribes which action is appropriate for each link category, but not which is feasible.

Third, changing the predictive model is another lever available to researchers, but it operates differently from the other two: model revision changes the predicted dimension of the categorisation, which cascades into reassigning link categories across the contextual dimension. It is therefore not a validation action but a reset of the categorisation itself, and its effects on link category distributions can also be diagnostic, revealing which categories shrink or grow when model assumptions change. This is particularly informative for model-elusive links: when a link is consistently observed but never predicted, a conceptual change in model architecture may be necessary to resolve the (apparent) model failures, such as using a different kind of model (e.g., tree-based vs. neural networks) or changing the set of features used for prediction [11, 13, 16, 58].

A Bayesian reading of the link taxonomy

Until now, we deterministically assigned each link to one of the eight categories from Fig. 1, based on three pieces of evidence: the model's prediction (Y), whether it was observed locally (O_l), and whether it was observed in the replicates (O_r). For example, a *possibly missing* link is predicted as positive ($Y = 1$), not observed locally ($O_l = 0$), and observed in replicates ($O_r = 1$). We collect these evidence as $E = (Y, O_l, O_r)$.

However, every evidence source $d \in \{Y, l, r\}$ is error-prone. The prediction can be wrong about the truth and not merely about the data it was fitted to; local sampling may miss interactions; and a realisable link may go unrecorded across the replicates. Therefore, the deterministic assignment, which we use for simplicity, effectively contains uncertainty. For instance, a *possibly missing* link may actually be a *recurrent* one that was missed locally due to error in local detection. We thus expand the deterministic approach by attaching a confidence to each category assignment: instead of one label per link, the evidence supports several categories as plausible alternatives among which probability is divided.

We treat the true category $c \in C$ as unknown and ask how probable each category is given the evidence E. Given the error rate of each source of evidence (ϵ_d), Bayes' theorem inverts the observed evidence into posterior probability:

$$\underbrace{P(C \mid E)}_{\text{posterior}} = \frac{\overbrace{P(E \mid C)}^{\text{likelihood}} \overbrace{P(C)}^{\text{prior}}}{\underbrace{P(E)}_{\text{evidence}}}.$$

The error rates can be used in calculating the likelihood. We write $z_{c,d} \in \{0, 1\}$ for the value that component $d \in \{Y, l, r\}$ takes in category $c \in C$, that is, the evidence that category c would produce if no source erred. We denote the sources of error as $\epsilon_d \in \{\epsilon_Y, \epsilon_l, \epsilon_r\}$. Assuming the error sources are independent given the category, each contributes $1 - \epsilon_d$ when the observed value E_d matches $z_{c,d}$, and ϵ_d when it differs:

$$P(E \mid C = c) = \prod_{d \in \{Y,\, l,\, r\}} \epsilon_d^{\mathbf{1}[E_d \neq z_{c,d}]} \, (1 - \epsilon_d)^{\mathbf{1}[E_d = z_{c,d}]},$$

where $\mathbf{1}[\cdot]$ is the indicator function equal to 1 when its input is true and 0 when false.

This framing acknowledges the idea of probabilistic interactions [28], but flips the point of view. Instead of calculating link probability under given conditions, which are in many cases unknown, we place a posterior on link categorisation. The factors which affect link probability (e.g., co-occurrence, abundance, traits, environment, sampling) enter in two places: they govern realisation and detectability, which underlie the error rates, and they inform the prior on the categories. This view is actionable: the posterior identifies which category the uncertainty favours, and thus the most appropriate action to resolve it.

The simplest derivation uses binary evidence ($E_d \in \{0, 1\}$), fixed error rates, a uniform prior, and independent errors. We present here two extensions (Fig. 3), and in the Supplementary Information we state the full set of assumptions explicitly, discuss how they can be relaxed, and provide worked calculations. In addition, we provide a rich interactive visualisation of how the posterior responds to the error rates, following the same development, available at `http://lpguide.ecomplab.com`.

One straightforward extension of the binary framework is to consider cumulative O_r (see Supplementary Information for the mathematical development). As replicates accumulate consistent evidence (repeated observation or repeated absence) the posterior concentrates on one category and confidence grows (Fig. 3a). However, confidence does not grow without limit. Once the replicate evidence is resolved, the posterior cannot exceed a maximum contextual confidence $\kappa = (1 - \epsilon_Y)(1 - \epsilon_l)$, set by the model and the local sampling alone. Replicates therefore increase the rate to reach that maximum rather than raising it. Raising κ requires a better model or a complementary method, which is the methodological dimension of the framework. The rate of accumulation

is ecologically informative because, under repeated detection, a posterior that sharpens quickly marks a link consistent across contexts, whereas slow saturation signals context-dependence.

Confidence can also be aggregated across categories. We denote by L_l and L_r the truths that O_l and O_r observe with error: whether the link is realised where we sampled, and whether it is realisable in the replicates. The signature of a category, $z_{c,l}$ and $z_{c,r}$, is the value each of these truths takes in it (Supplementary Information). Then, the *feasibility confidence* $\phi = P(L_l = 1 \text{ or } L_r = 1 \mid E)$ is then the posterior probability that at least one of them holds. Ecologically, it is a lower bound on the probability that the link is feasible, because a feasible link may go unrealised everywhere we looked, and unlike confidence in any single category, it is not bounded by κ (Fig. 3a; Supplementary Information).

The prior is the second place where ecology enters. A uniform prior gives all eight categories equal weight before any evidence, which is rarely what we believe because interactions are sparse, and a pair of generalists is far more likely to interact than a pair of specialists [59]. Researchers can implement such knowledge by changing the priors (Fig. 3b). For example, suppose the evidence we have at hand is $E = (1, 0, 1)$, the signature of a *possibly missing* link. Without changing that evidence, the prior shifts the categorisation towards *phantom* for a rare, specialist pair (reading the replicate record as a false detection, $L_r = 0$ although $O_r = 1$), or towards *recurrent* for a pair of strong generalists, which we believe will consistently interact (reading the local non-detection as a miss, $L_l = 1$ although $O_l = 0$) (Fig. 3b).

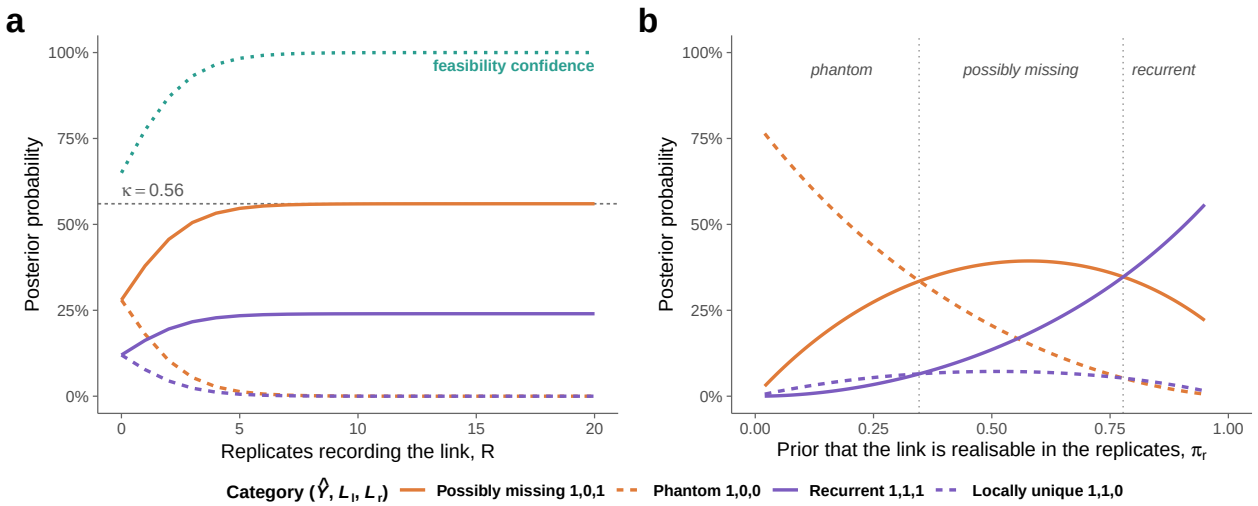

Fig. 3: Using the Bayesian framework to improve guided sampling. In this example the link is predicted ($Y = 1$), unobserved locally ($O_l = 0$), and searched for in every replicate. The legend gives each category's error-free signature. The lines differ in whether the link was realised where we sampled (distinguished by colour), and whether it is realisable in the replicates (solid where it is and dashed where it is not). **(a)** Posterior probability of four categories as within-system replicates accumulate. Confidence in *possibly missing* is limited to the maximum contextual confidence $\kappa = (1 - \epsilon_Y)(1 - \epsilon_l)$ (horizontal dotted line). The aggregate *feasibility confidence* ϕ is not bounded by κ. **(b)** The same four categories under an informed (non-uniform) prior, with the binary replicate evidence after five replicates ($O_r = 1$). Both priors (on local and replicate truths) rise with the pair's degree, so the axis runs from two interacting specialists on the left to two interacting generalists on the right. Dotted vertical lines mark where the leading category changes, named in italics. The same evidence and the same error rates therefore support three different readings, depending only on what is ecologically known about the pair. The illustrative rates are $\epsilon_Y = 0.2$ and $\epsilon_l = 0.3$, giving $\kappa = 0.56$, with per-replicate probabilities of recording the link of 0.105 where it is realisable in the replicates and 0.05 where it is not; they show the behaviour of the framework rather than describing any particular system. Together the panels turn the posterior into a sampling decision: (a) how many replicates is it worth sampling, and (b) how sampling would be affected by what we know about the species. We provide interactive versions of these plots in https://lpguide.ecomplab.com

Empirical demonstration of the evidence accumulation framework

To show how the framework guides gathering evidence in a real system, we applied it to plant-pollinator interactions on Cabrera island (Balearic archipelago, western Mediterranean Sea). Serra-Marin et al. [56] monitored pollination at six sites on the island by direct field observation and, critically, ran automated cameras in parallel at every site. This second, independent method lets us test the framework's predictions for the link categories that call for the use of additional methodology (Fig. 2).

We predicted all possible interactions among co-occurring species at each site using network embedding, which draws on latent structural patterns in each site's own pollination network, following the pipeline of Abramov et al. [37] under a leave-one-out approach (each potential interaction, observed or not, was withheld and predicted). Comparing predictions to field observations yields a within-site confusion matrix (Fig. 4A, first column). For every prediction, we then asked whether the same interaction appeared at any other site (second column), using this contextual evidence to assign each prediction to a taxonomy category (third column). At the final step, we tested targeted categories against the independent camera records (fourth column).

Across the six sites, field observation recorded 1624 potential interactions, of which 1232 (75.9%) were unobserved. Contextual evidence flagged 51 of these unobserved interactions as possibly missing while 103 observed links were weakly supported. The 51 possibly missing links were the higher priority of the two for validation, and targeting them alone reduced detection effort by roughly 24-fold from the original 1232 unobserved links.

Two other categories show why our guiding principles are useful: 941 links were possibly forbidden and 104 were phantom links with no support in the data at all, both signalling that increasing sampling effort of direct observations would be an inefficient course of action. Instead, the recommended course of action is to add an orthogonal method (Fig. 2). Cameras confirmed this reasoning directly: 8.5% of the possibly forbidden and 16.3% of the phantom links were captured on camera. Cameras also provided support for the categorisation system by recovering 51% of the links flagged as possibly missing, supporting the high priority that the framework assigns this category. Since direct observation is capable of detecting these links at other sites, the framework's prescribed action, more of the same effort, would have recovered them without the cost of a second method.

This guidance helps researchers make decisions at fine resolutions, such as whether to inspect specific sites or species of interest, and where to focus observation efforts (Fig. 4B). For instance, *Teucrium capitatum*, with multiple possibly missing links, rewards more intensive observation with the same method; *Helichrysum stoechas*, whose interactions largely escaped both observation and prediction, calls for the complementary sampling method instead; and *Daucus carota*, a generalist rich in locally unique links, shows how replication across sites turns single-network noise into ecological signal.

Assumptions of evidence accumulation

Like any other inferential framework, contextual evidence rests on several assumptions. The deterministic categorisation reads the evidence at face value, taking an unobserved link as evidence for absence; that is not an assumption, but rather the starting point to which the probabilistic version (Section Bayesian Inference) injects confidence by attaching an error rate to each source of evidence. The assumptions underlying the confidence concern the ecology, the sampling, and the predictive model. We briefly discuss the main ones here. In the Supplementary Information we expand on the assumptions and show how to relax them.

Ecology. Link feasibility is assumed to be a regional property (sensu Banville et al. [28]). Realisation itself may vary freely across replicates, which is the signal the framework exploits, and is what separates *locally absent* (feasible but unrealised in a patch) from *possibly forbidden* (infeasible anywhere) links. The simplest probabilistic version further assumes a uniform prior, with all eight categories equally probable before any evidence is seen. However, interactions are sparse and degree distributions are skewed, so this is rarely true: co-occurrence, abundance, generalism, trait matching and phylogeny all bear on how probable a link is before it is sampled, and can inform the prior instead.

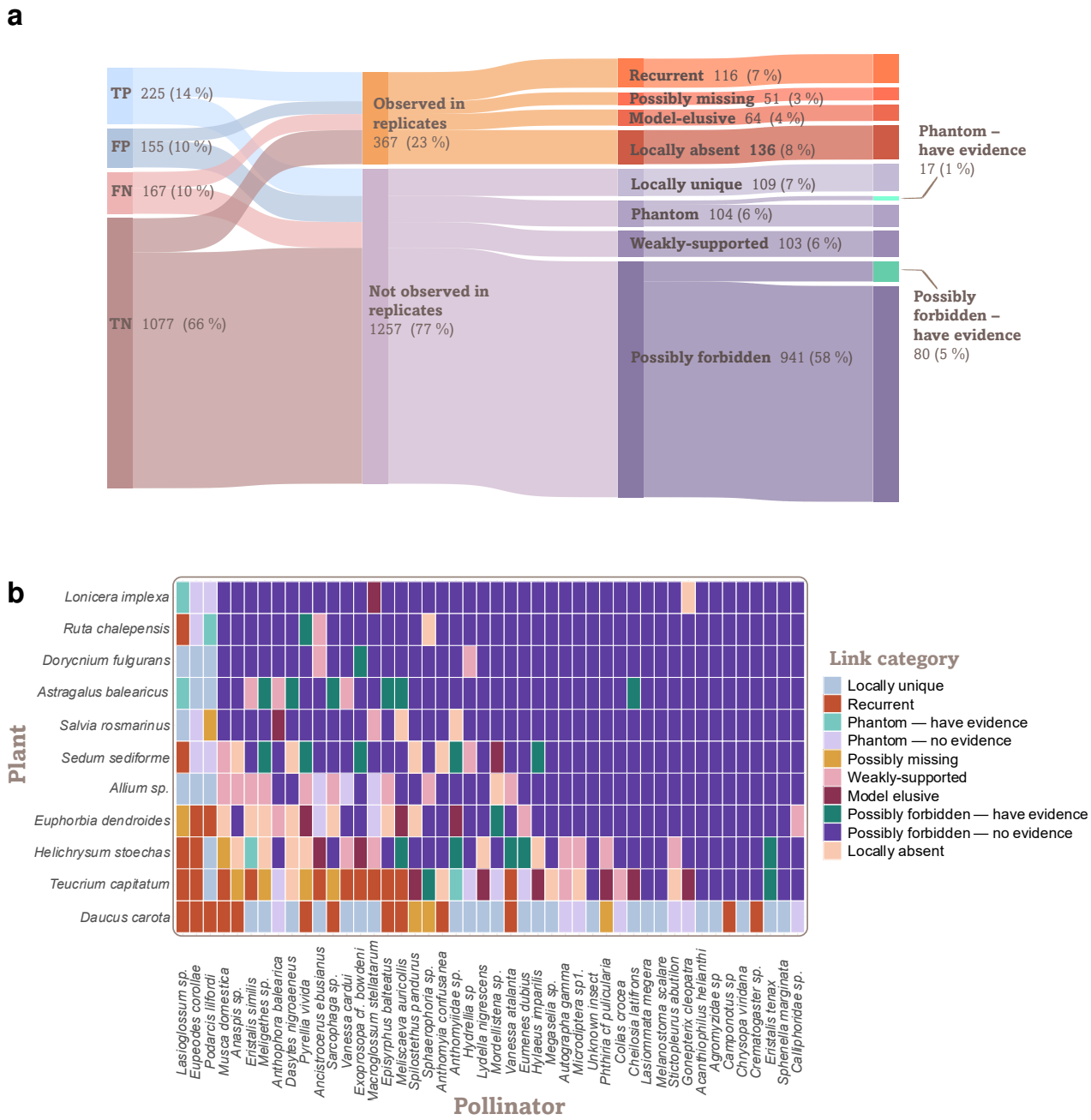

Fig. 4: Distribution of evidence-based categories for plant–pollinator interactions in Cabrera island. (A) Flow of predicted links (for direct observations data set) through within-network evaluation classes, contextual evidence from other sites, which yields the taxonomy categories, and complementary-method (camera) evidence for targeted categories. Counts represent unique potential interactions summed across sites; percentages are out of the total number of potential links. An interactive version of the figure, with controls for sampling method and classification threshold, is available at https://lpguide.ecomplab.com/. **(B)** A detailed map of predictions between co-occurring species at the most species-rich site in the system, coloured by link category. Each cell represents an interaction between a plant species (row) and a pollinator species (column).

Sampling. We assume that replicates are sampled with comparable effort, methods, and taxonomic resolution; otherwise apparent evidence reflects sampling differences rather than feasibility. The likelihood further treats the local and replicate observations as erring independently. This assumption rarely holds because replicates typically share one sampling method and its blind spots carry across all of them. A feasible link the method cannot detect stays absent locally and in every replicate, masquerading as *possibly forbidden*. This is why contextual evidence is necessary but not sufficient, and why a complementary method helps only if its detection biases differ from those of the first [24, 57]. As we demonstrate in the analysis, automated cameras, whose biases differ from direct observation, recovered *possibly forbidden* and *phantom* links, which were not detected via human visual sampling (Section Empirical demonstration). Finally, the detection error rates are assumed known and common to all links. They are not arbitrary, since the false-negative rate of interaction sampling can be estimated from sampling effort and species relative abundances [27], but links differ in how readily they are seen.

The predictive model. The assumptions here are essentially those of any link prediction model applied in ecology, because all of them train the model on observed data rather than on the true data. The model is nevertheless assumed to err independently of the observations, even though its error rate normally comes from cross-validation against those same observations (Box 1), which returns a network-level average that is then applied to the single link in question. Using that rate assumes that the pairs the model was trained on, the pairs it was tested on, and the pairs the framework asks it about are drawn from similar distributions. However, predictive models in ecology are typically biased by the data: training data are dominated by abundant, well-covered species, whereas the links at issue lie in the sparse tail where the model has least support and taxonomic bias is strongest [60]. Blocking the withholding by site or clade rather than at random [45], and keeping the replicate networks out of the training data, are design choices that make those assumptions more defensible.

Future challenges and research directions

Establishing that variation between networks is evidence opens a set of questions about what counts as a replicate and what its evidence is worth. Replicates differ in how closely they resemble the focal network, so relating ecological dissimilarity to evidential weight is a natural next step [37]. Contextual evidence can be drawn across space or time, yet these sources of variation are not ecologically equivalent. It remains open whether they should carry the same evidential weight, and if specific adaptations to temporal networks are necessary. Beyond space and time, other axes of variability may serve as well, and identifying them would widen the framework's reach. A special case are communities affected by perturbations, because range shifts and invasions generate assemblages for which no sufficiently similar reference network exists. Finding contextual evidence for predicted interactions in these emerging communities is therefore a critical research route.

A second set of questions concerns estimating the error rates and priors that the probabilistic version takes as input. The sampling rates are within reach of established designs, through repeated visits, effort- and abundance-based estimators, or negative controls [27, 61], and the realisation rate can be estimated from the replicates themselves, though only as a lower bound [28]. The model error is the hardest, because cross-validation measures a model against the observations it was fitted to rather than against the truth, so recovering it calls for calibration against an orthogonal method or latent-class estimation from several imperfect ones. The prior raises a different question: traits, phylogeny and independently surveyed abundances are external to the interaction data, whereas degree and generalism are read off the network and are already part of the evidence. Establishing which sources are external, and how to weight their contribution, will increase our confidence in the guided sampling.

A third set concerns turning confidence into design. Because replication buys speed towards a maximum set by the model and the sampling method, but cannot raise it, the practical question is when to add replicates and when to raise that maximum instead, which is a decision about cost that the accumulation curves can inform. The maximum is raised either by a complementary sampling method or by a better model, and the

latter invites treating model ensembles as evidence [40, 62]: much as a replicate joins the local observation, a second model can enter as a further evidence source with its own error rate. In the same spirit, matching sampling and prediction methods to the specific needs of each link category would maximise coverage while minimising redundancy with evidence already in hand. Finally, the feasibility confidence offers an entry point to tasks beyond validation, such as assembling metawebs or deciding which unobserved pairs should be treated as true absences when training the next model, closing the loop between evidence accumulation and the models that generate the predictions.

Concluding remarks

Ecological link prediction is maturing rapidly, yet two problems persist: interaction data are incomplete and biased (e.g., by species abundance) [6], and validating predictions is usually prohibitively costly. We address the second, and in doing so offer a route to the first. The remedy is not more sophisticated models alone, but a shift in how predictions are interpreted to guide data collection. Reframing link prediction as evidence accumulation provides an ecologically grounded reading of what a predicted link means. Our taxonomy of eight link categories and the two dimensions of evidence accumulation, contextual and methodological, turn that shift into a practical roadmap for the field. Guided sampling then closes a loop the field leaves open: validation feeds back into the models, so each round of targeted fieldwork trains a better model that sharpens the guidance for the next, forming an active-learning cycle [63, 64]. In practice, evidence accumulation offers concrete guidelines for validating link predictions and should be built into study design from the outset. Crucially, it is testable with data that many research groups already gather, since study designs often include replicated networks across sites or time and the capacity to employ more than one method. The challenges ahead are therefore not only about building better models, but about better ways of knowing what our models mean.

Data availability

The data used for the empirical demonstration are available in the repository set up in the original publication by Serra-Marin et al. [56] at https://doi.org/10.5281/zenodo.17130777.

Acknowledgements

We thank Barry Biton for insightful discussions.

Funding

The study was funded by the Israel Science Foundation (grant 491/25) to SP and the Human Frontiers Science Program (grant RGY0064/2022) to SP. SP was also supported by the Dartmouth Kalaniyot in the Sciences Visiting Scholar Program. PJM was supported by the NSF (grant DEB-2308460). KA was supported by the Kreitman School of Advanced Graduate Studies, Ben-Gurion University of the Negev, and by the Planning and Budgeting Committee (VATAT) of the Council for Higher Education, through the Levtzion Scholarship for Outstanding Doctoral Students.

References

1. Jordano, P. Chasing ecological interactions. en. *PLoS Biology* **14,** e1002559. doi:10.1371/journal.pbio.1002559 (2016).
2. Chacoff, N. P., Vázquez, D. P., Lomáscolo, S. B., Stevani, E. L., Dorado, J. & Padrón, B. Evaluating sampling completeness in a desert plant–pollinator network. *Journal of Animal Ecology* **81,** 190–200. doi:10.1111/j.1365-2656.2011.01883.x (2012).

3. Souza, C. S., Oliveira, P. E., Rosa, B. B. & Maruyama, P. K. Integrating nocturnal and diurnal interactions in a Neotropical pollination network. en. *The Journal of ecology* **110,** 2145–2155. doi:10.1111/1365-2745.13937 (2022).

4. Jordano, P. Sampling networks of ecological interactions. en. *Functional ecology* **30,** 1883–1893. doi:10.1111/1365-2435.12763 (2016).

5. Llopis-Belenguer, C., Balbuena, J. A., Blasco-Costa, I., Karvonen, A., Sarabeev, V. & Jokela, J. Sensitivity of bipartite network analyses to incomplete sampling and taxonomic uncertainty. en. *Ecology* **104,** e3974. doi:10.1002/ecy.3974 (2023).

6. Peralta, G., CaraDonna, P. J., Rakosy, D., Fründ, J., Pascual Tudanca, M. P., Dormann, C. F., Burkle, L. A., Kaiser-Bunbury, C. N., Knight, T. M., Resasco, J., Winfree, R., Blüthgen, N., Castillo, W. J. & Vázquez, D. P. Predicting plant-pollinator interactions: concepts, methods, and challenges. en. *Trends in ecology & evolution* **39,** 494–505. doi:10.1016/j.tree.2023.12.005 (2024).

7. Derocles, S. A. P., Lunt, D. H., Berthe, S. C. F., Nichols, P. C., Moss, E. D. & Evans, D. M. Climate warming alters the structure of farmland tritrophic ecological networks and reduces crop yield. en. *Molecular ecology* **27,** 4931–4946. doi:10.1111/mec.14903 (2018).

8. Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Brierley, L., Albery, G. F., Gibb, R. J., Seifert, S. N. & Carlson, C. J. Network embedding unveils the hidden interactions in the mammalian virome. en. *Patterns (New York, N.Y.)* **4,** 100738. doi:10.1016/j.patter.2023.100738 (2023).

9. Tylianakis, J. M., Didham, R. K., Bascompte, J. & Wardle, D. A. Global change and species interactions in terrestrial ecosystems. en. *Ecology letters* **11,** 1351–1363. doi:10.1111/j.1461-0248.2008.01250.x (2008).

10. Pearse, I. S. & Altermatt, F. Predicting novel trophic interactions in a non-native world. en. *Ecology letters* **16,** 1088–1094. doi:10.1111/ele.12143 (2013).

11. Dallas, T., Park, A. W. & Drake, J. M. Predicting cryptic links in host-parasite networks. en. *PLoS computational biology* **13,** e1005557. doi:10.1371/journal.pcbi.1005557 (2017).

12. Desjardins-Proulx, P., Laigle, I., Poisot, T. & Gravel, D. Ecological interactions and the Netflix problem. en. *PeerJ* **5,** e3644. doi:10.7717/peerj.3644 (2017).

13. Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M. & Hartig, F. Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. en. *Methods in ecology and evolution / British Ecological Society* **11,** 281–293. doi:10.1111/2041-210x.13329 (2020).

14. Terry, J. C. D. & Lewis, O. T. Finding missing links in interaction networks. en. *Ecology* **101,** e03047. doi:10.1002/ecy.3047 (2020).

15. Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz, N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L. & Poisot, T. A roadmap towards predicting species interaction networks (across space and time). en. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **376,** 20210063. doi:10.1098/rstb.2021.0063 (2021).

16. Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V. & Poisot, T. Food web reconstruction through phylogenetic transfer of low-rank network representation. en. *Methods in ecology and evolution* **13,** 2838–2849. doi:10.1111/2041-210x.13835 (2022).

17. Biton, B., Puzis, R. & Pilosof, S. Inductive link prediction facilitates the discovery of missing links and enables cross-community inference in ecological networks. en. *Nature ecology & evolution.* doi:10.1038/s41559-025-02715-6 (2025).

18. Guimerà, R. & Sales-Pardo, M. Missing and spurious interactions and the reconstruction of complex networks. en. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 22073–22078. doi:10.1073/pnas.0908366106 (2009).

19. Stock, M., Poisot, T., Waegeman, W. & De Baets, B. Linear filtering reveals false negatives in species interaction data. en. *Scientific reports* **7,** 45908. doi:10.1038/srep45908 (2017).

20. Rossi, A., Firmani, D., Merialdo, P. & Teofili, T. *Explaining link prediction systems based on knowledge graph embeddings* in *Proceedings of the 2022 International Conference on Management of Data* (ACM, New York, NY, USA, 2022). doi:10.1145/3514221.3517887.

21. Strydom, T, Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M, Fortin, M., Mercier, B, Pollock, L. J., Runghen, R, Dalla Riva, G. V. & Poisot, T. Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations. *Methods in ecology and evolution / British Ecological Society* **14,** 2917–2930. doi:10.1111/2041-210X.14228 (2023).

22. Dansereau, G., Barros, C. & Poisot, T. Spatially explicit predictions of food web structure from regional-level data. en. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **379,** 20230166. doi:10.1098/rstb.2023.0166 (2024).

23. Quintero, E., Isla, J. & Jordano, P. Methodological overview and data-merging approaches in the study of plant–frugivore interactions. en. *Oikos (Copenhagen, Denmark)* **2022.** doi:10.1111/oik.08379 (2022).

24. Cuff, J. P., Deivarajan Suresh, M., Dopson, M. E. G., Hawthorne, B. S. J., Howells, T., Kitson, J. J. N., Miller, K. A., Xin, T. & Evans, D. M. en. in *Advances in Ecological Research* 1–34 (Elsevier, 2023). doi:10.1016/bs.aecr.2023.09.002.

25. Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C. & Jordano, P. Missing and forbidden links in mutualistic networks. en. *Proceedings. Biological sciences / The Royal Society* **278,** 725–732. doi:10.1098/rspb.2010.1371 (2011).

26. González-Varo, J. P. & Traveset, A. The Labile Limits of Forbidden Interactions. *Trends in Ecology & Evolution. Multiplicity in Unity: Plant Subindividual Variation and Interactions with AnimalsHandbook of the Birds of the World (Vol. 2)The Complete Birds of the Western Paleartic (CD-ROM)Genetics and Analysis of Quantitative Traits* **31,** 700–710. doi:10.1016/j.tree.2016.06.009 (2016).

27. Catchen, M. D., Poisot, T., Pollock, L. J. & Gonzalez, A. The missing link: Discerning true from false negatives when sampling species Interaction networks. *EcoEvoRxiv*. doi:10.32942/x2dw22 (2023).

28. Banville, F., Strydom, T., Blyth, P. S. A., Brimacombe, C., Catchen, M. D., Dansereau, G., Higino, G., Malpas, T., Mayall, H., Norman, K., Gravel, D. & Poisot, T. Deciphering probabilistic species interaction networks. en. *Ecology Letters* **28,** e70161. doi:10.1111/ele.70161 (2025).

29. Poelen, J. H., Simons, J. D. & Mungall, C. J. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological informatics* **24,** 148–159. doi:10.1016/j.ecoinf.2014.08.005 (2014).

30. Fortuna, M. A., Ortega, R. & Bascompte, J. The web of life. *arXiv [q-bio.PE]* (2014).

31. Farrell, M. J., Elmasri, M., Stephens, D. A. & Davies, T. J. Predicting missing links in global host-parasite networks. en. *The Journal of animal ecology* **91,** 715–726. doi:10.1111/1365-2656.13666 (2022).

32. Stock, M., Piot, N., Vanbesien, S., Meys, J., Smagghe, G. & De Baets, B. Pairwise learning for predicting pollination interactions based on traits and phylogeny. *Ecological modelling* **451,** 109508. doi:10.1016/j.ecolmodel.2021.109508 (2021).

33. Fu, X., Seo, E., Clarke, J. & Hutchinson, R. A. Link prediction under imperfect detection: Collaborative filtering for ecological networks. *IEEE Transactions on Knowledge and Data Engineering* **33,** 3117–3128. doi:10.1109/tkde.2019.2962031 (2021).

34. CaraDonna, P. J., Petry, W. K., Brennan, R. M., Cunningham, J. L., Bronstein, J. L., Waser, N. M. & Sanders, N. J. Interaction rewiring and the rapid turnover of plant-pollinator networks. en. *Ecology letters* **20,** 385–394. doi:10.1111/ele.12740 (2017).

35. Trøjelsgaard, K., Jordano, P., Carstensen, D. W. & Olesen, J. M. Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. en. *Proceedings. Biological sciences / The Royal Society* **282,** 20142925. doi:10.1098/rspb.2014.2925 (2015).

36. Poisot, T., Stouffer, D. B. & Gravel, D. Beyond species: why ecological interaction networks vary through space and time. en. *Oikos* **124,** 243–251. doi:10.1111/oik.01719 (2015).

37. Abramov, K., Galai, G., Biton, B., Puzis, R. & Pilosof, S. Predicting ecological interactions across space through pairwise integration of latent network patterns. en. *Methods in Ecology and Evolution.* doi:10.1111/2041-210x.70357 (2026).

38. He, X., Ghasemian, A., Lee, E., Schwarze, A. C., Clauset, A. & Mucha, P. J. Link prediction accuracy on real-world networks under non-uniform missing-edge patterns. en. *PloS one* **19,** e0306883. doi:10.1371/journal.pone.0306883 (2024).

39. Yates, L. A., Aandahl, Z., Richards, S. A. & Brook, B. W. Cross validation for model selection: A review with examples from ecology. en. *Ecological monographs* **93.** doi:10.1002/ecm.1557 (2023).

40. Poisot, T. Guidelines for the prediction of species interactions through binary classification. en. *Methods in ecology and evolution / British Ecological Society* **14,** 1333–1345. doi:10.1111/2041-210x.14071 (2023).

41. Dormann, C. F., Castillo, W. J., Tudanca, M. P. P., CaraDonna, P., Burkle, L. A., Rakosy, D., Winfree, R., Zoller, L., Benadi, G., Kaiser-Bunbury, C. N. & Others. Predicting interaction frequency in plant-pollinator networks (2025).

42. Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., Maglianesi, M. A., Melián, C. J., Pitteloud, C., Roslin, T., Rohr, R., Saavedra, S., Thuiller, W., Woodward, G., Zimmermann, N. E. & Gravel, D. Comparing species interaction networks along environmental gradients: Networks along environmental gradients. en. *Biological Reviews of the Cambridge Philosophical Society* **93,** 785–800. doi:10.1111/brv.12366 (2018).

43. Poisot, T., Canard, E., Mouillot, D., Mouquet, N. & Gravel, D. The dissimilarity of species interaction networks. en. *Ecology letters* **15,** 1353–1361. doi:10.1111/ele.12002 (2012).

44. Hervías-Parejo, S., Colom, P., Beltran Mas, R., E. Serra, P., Pons, S., Mesquida, V. & Traveset, A. Spatio-temporal variation in plant–pollinator interactions: a multilayer network approach. en. *Oikos (Copenhagen, Denmark)* **2023.** doi:10.1111/oik.09818 (2023).

45. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. & Dormann, C. F. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. en. *Ecography* **40,** 913–929. doi:10.1111/ecog.02881 (2017).

46. Fortuna, M. A., Nagavci, A., Barbour, M. A. & Bascompte, J. Partner fidelity and asymmetric specialization in ecological networks. en. *The American naturalist* **196,** 382–389. doi:10.1086/709961 (2020).

47. Dorado, J., Vázquez, D. P., Stevani, E. L. & Chacoff, N. P. Rareness and specialization in plant-pollinator networks. en. *Ecology* **92,** 19–25. doi:10.1890/10-0794.1 (2011).

48. Ward, C. A., Tunney, T. D., Hale, K. R. S., O'Connor, R. F. & McCann, K. S. The rewiring of ecological networks in a variable world. en. *Nature Reviews Biodiversity* **2,** 355–369. doi:10.1038/s44358-026-00159-9 (2026).

49. Peralta, G., Vázquez, D. P., Chacoff, N. P., Lomáscolo, S. B., Perry, G. L. W. & Tylianakis, J. M. Trait matching and phenological overlap increase the spatio-temporal stability and functionality of plant-pollinator interactions. en. *Ecology letters* **23,** 1107–1116. doi:10.1111/ele.13510 (2020).

50. García, Y., Giménez-Benavides, L., Iriondo, J. M., Lara-Romero, C., Méndez, M., Morente-López, J. & Santamaría, S. Addition of nocturnal pollinators modifies the structure of pollination networks. en. *Scientific reports* **14,** 1226. doi:10.1038/s41598-023-49944-y (2024).

51. Strydom, T., Dunhill, A. M., Dunne, J. A., Poisot, T. & Beckerman, A. P. Scaling from Metawebs to Realised Webs: A Hierarchical Approach to Network Ecology (2026).

52. Trøjelsgaard, K. & Olesen, J. M. Ecological networks in motion: micro- and macroscopic variability across scales. en. *Functional ecology* **30,** 1926–1935. doi:10.1111/1365-2435.12710 (2016).

53. Miller, D. A., Nichols, J. D., Mcclintock, B. T., Grant, E. H. C., Bailey, L. L. L. & Weir, L. A. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. en. *Ecology* **92,** 1422–1428. doi:10.1890/10-1396.1 (2011).

54. Chambert, T., Miller, D. A. W. & Nichols, J. D. Modeling false positive detections in species occurrence data under different study designs. en. *Ecology* **96,** 332–339. doi:10.1890/14-1507.1 (2015).

55. Brimacombe, C., Bodner, K., Michalska-Smith, M., Poisot, T. & Fortin, M.-J. Shortcomings of reusing species interaction networks created by different sets of researchers. en. *PLoS Biology* **21,** e3002068. doi:10.1371/journal.pbio.3002068 (2023).

56. Serra-Marin, P. E., Solé-Ribalta, A., Lana, A., Borge-Holthoefer, J., Hervías-Parejo, S. & Traveset, A. Comparative assessment of automated and manual monitoring in comprehensive plant–pollinator communities. en. *Methods in Ecology and Evolution* **16,** 2960–2978. doi:10.1111/2041-210x.70165 (2025).

57. Oreskes, N. Evaluation (not validation) of quantitative models. en. *Environmental health perspectives* **106 Suppl 6,** 1453–1460. doi:10.1289/ehp.98106s61453 (1998).

58. Benadi, G., Dormann, C. F., Fründ, J., Stephan, R. & Vázquez, D. P. Quantitative Prediction of Interactions in Bipartite Networks Based on Traits, Abundances, and Phylogeny. en. *The American naturalist* **199,** 841–854. doi:10.1086/714420 (2022).

59. Vázquez, D. P., Melián, C. J., Williams, N. M., Blüthgen, N., Krasnov, B. R. & Poulin, R. Species abundance and asymmetric interaction strength in ecological networks. en. *Oikos (Copenhagen, Denmark)* **116,** 1120–1127. doi:10.1111/j.0030-1299.2007.15828.x (2007).

60. Kampe, J. N., DeSisto, C. M. M. & Dunson, D. B. Link prediction in ecological meta-networks under extreme taxonomic bias. en. *Methods in Ecology and Evolution.* doi:10.1111/2041-210x.70368 (2026).

61. Royle, J. A. N-mixture models for estimating population size from spatially replicated counts. en. *Biometrics* **60,** 108–115. doi:10.1111/j.0006-341X.2004.00142.x (2004).

62. Adhurya, S. & Park, Y.-S. A novel method for predicting ecological interactions with an unsupervised machine learning algorithm. en. *Methods in ecology and evolution* **15,** 1247–1260. doi:10.1111/2041-210x.14358 (2024).

63. Settles, B. *From theories to queries: Active learning in practice* in *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010* (eds Guyon, I., Cawley, G., Dror, G., Lemaire, V. & Statnikov, A.) **16** (PMLR, Sardinia, Italy, 2011), 1–18.

64. Henrys, P. A., Mondain-Monval, T. O. & Jarvis, S. G. Adaptive sampling in ecology: Key challenges and future opportunities. en. *Methods in Ecology and Evolution* **15,** 1483–1496. doi:10.1111/2041-210x.14393 (2024).