

# Graph embedding and transfer learning can help predict potential species interaction networks despite data limitations

Tanya Strydom<sup>1,2</sup>; Salomé Bouskila<sup>1</sup>; Francis Banville<sup>1,2,3</sup>; Ceres Barros<sup>4</sup>; Dominique Caron<sup>5,2</sup>; Maxwell J. Farrell<sup>6</sup>; Marie-Josée Fortin<sup>6</sup>; Benjamin Mercier<sup>3,2</sup>; Laura J. Pollock<sup>5,2</sup>; Rogini Runghen<sup>7</sup>; Giulio V. Dalla Riva<sup>7</sup>; Timothée Poisot<sup>1,2</sup>

## Abstract:

- Metawebs (networks of potential interactions within a species pool) are a powerful abstraction to understand how large-scale species interaction networks are structured.
- Because metawebs are typically expressed at large spatial and taxonomic scales, assembling them is a tedious and costly process; predictive methods can help circumvent the limitations in data deficiencies, by providing a first approximation of metawebs.”
- One way to improve our ability to predict metawebs is to maximize available information by using graph embeddings, as opposed to an exhaustive list of species interactions. Graph embedding is an emerging field in machine learning that holds great potential for ecological problems.
- Here, we outline how the challenges associated with inferring metawebs line-up with the advantages of graph embeddings; followed by a discussion as to how the choice of the species pool has consequences on the reconstructed network, specifically as to the role of human-made (or arbitrarily assigned) boundaries and how these may influence ecological hypotheses.

# 1 Introduction

2 The ability to infer potential biotic interactions could serve as a significant breakthrough in our ability  
3 to conceptualize networks over large spatial scales (Hortal et al., 2015). Reliable inferences would  
4 not only boost our understanding of the structure of species interaction networks, but also increase  
5 the amount of information that can be used for biodiversity management. In a recent overview of  
6 the field of ecological network prediction, Strydom, Catchen, et al. (2021) identified two challenges of  
7 interest to the prediction of interactions at large scales. First, there is a relative scarcity of relevant  
8 data in most places globally – which, due to the limitations in most predictive methods, restricts the  
9 ability to infer interactions to locations where it is least required (*i.e.* regions where we already have  
10 interaction data) leaving us unable to make inference in data scarce regions (where we most need  
11 it); second, accurate predictors are important for accurate predictions, and the lack of methods that  
12 can leverage a small amount of *accurate* data is a serious impediment to our predictive ability. In this  
13 contribution, we (i) highlight the power of viewing (and constructing) metawebs as *probabilistic* objects  
14 in the context of low-probability interactions, (ii) discuss how a family of machine learning tools (graph  
15 embeddings and transfer learning) can be used to overcome data limitations to metaweb inference,  
16 and (iii) highlight how the use of metawebs introduces important questions for the field of network  
17 ecology.

18 In most places, our most reliable biodiversity knowledge is that of a species pool where a set of po-  
19 tentially interacting species in a given area could occur: through the analysis of databases like the  
20 Global Biodiversity Information Facility (GBIF) or the International Union for the Conservation of Na-  
21 ture (IUCN), it is possible to construct a list of species for a region of interest. Following the definition  
22 of Dunne (2006), a metaweb is the ecological network analogue to the species pool; specifically, it in-  
23 ventories all *potential* interactions between species for a spatially delimited area (and so captures the  
24  $\gamma$  diversity of interactions as per Poisot et al. (2012)). However, inferring the potential interactions  
25 between these species still remains a challenge. And yet, the metaweb holds valuable ecological in-  
26 formation: it represents the joint effect of functional, phylogenetic, and macroecological processes  
27 (Carlson et al., 2022; Morales-Castilla et al., 2015; Morales-Castilla et al., 2021). Specifically, it repre-  
28 sents the “upper bounds” on what the composition of the local networks, given a local species pool,  
29 can be (see *e.g.* McLeod et al., 2021); this information can help evaluate the ability of ecological as-  
30 semblages to withstand the effects of, for example, climate change (Fricke et al., 2022). These local  
31 networks may be reconstructed given an appropriate knowledge of local species composition and pro-  
32 vide information on the structure of networks at finer spatial scales. This has been done for example

1 for tree-galler-parasitoid systems (Gravel et al., 2018), fish trophic interactions (Albouy et al., 2019),  
2 terrestrial tetrapod trophic interactions (J. Braga et al., 2019; O'Connor et al., 2020), and crop-pest  
3 networks (Grünig et al., 2020).

4 The metaweb itself is not a prediction of local networks at specific locations within the spatial area it  
5 covers: it will have a different structure, notably by having a larger connectance (see *e.g.* Wood et al.,  
6 2015) and complexity (see *e.g.* Galiana et al., 2022), than any of these local networks. Local networks  
7 (which capture the  $\alpha$  diversity of interactions) are a subset of the metaweb's species and its realized  
8 interactions, and have been called "metaweb realizations" (Poisot et al., 2015). Differences between  
9 local networks and their metawebs are due to chance, species abundance and co-occurrence, local  
10 environmental conditions, and local distribution of functional traits, among others. Specifically, al-  
11 though co-occurrence can be driven by interactions (Cazelles et al., 2016), co-occurrence alone is not  
12 a predictor of interactions (Blanchet et al., 2020; Thurman et al., 2019), and therefore the lack of co-  
13 occurrence cannot be used to infer the lack of a feasible interaction. Yet, recent results by Saravia et  
14 al. (2021) strongly suggested that local (metaweb) realizations only respond weakly to local conditions:  
15 instead, they reflect constraints inherited by the structure of their metaweb. This sets up the core goal  
16 of predictive network ecology as the prediction of metaweb structure, as it is required to accurately  
17 produce downscaled, local predictions.

## 18 **A metaweb is an inherently probabilistic object**

19 Treating interactions as probabilistic (as opposed to binary) events is a more nuanced and realistic way  
20 to represent them. Dallas et al. (2017) suggested that most interactions (links) in ecological networks  
21 are cryptic, *i.e.* uncommon or hard to observe. This argument echoes Jordano (2016): sampling ecolog-  
22 ical interactions is difficult because it requires first the joint observation of two species, and then the  
23 observation of their interaction. In addition, it is generally expected that weak or rare interactions will  
24 be more prevalent in networks than common or strong interactions (Csermely, 2004); this is notably  
25 the case in food chains, wherein many weaker interactions are key to the stability of a system (Neutel et  
26 al., 2002). In the light of these observations, we expect to see an over-representation of low-probability  
27 (hereafter rare) interactions under a model that accurately predicts interaction probabilities.

28 Yet, the original metaweb definition, and indeed most past uses of metawebs, was based on the pres-  
29 ence/absence of interactions. Moving towards *probabilistic* metawebs, by representing interactions as  
30 Bernoulli events (see *e.g.* Poisot et al., 2016), offers the opportunity to weigh these rare interactions

1 appropriately. The inherent plasticity of interactions is important to capture: there have been docu-  
2 mented instances of food webs undergoing rapid collapse/recovery cycles over short periods of time  
3 (*e.g.* Pedersen et al., 2017). Furthermore, because the structure of the metaweb cannot be known in ad-  
4 vance, it is important to rely on predictive tools that do not assume a specific network topology for link  
5 prediction (Gaucher et al., 2021), but are able to work on generalizations of the network that capture  
6 statistical processes giving it its structure. These considerations emphasize why metaweb predictions  
7 should focus on quantitative (preferentially probabilistic) predictions, and this should constrain the  
8 suite of models that are appropriate for prediction. Binary classifiers based on probabilities have an  
9 extremely robust methodology to validate them, and this applies naturally to the prediction of interac-  
10 tions (Poisot, 2023).

11 It is important to recall that a metaweb is intended as a catalogue of all potential (feasible) interactions,  
12 which is then filtered for a given application (Morales-Castilla et al., 2015). It is therefore important  
13 to separate the interactions that happen “almost surely” (repeated observational data), “almost never”  
14 (repeated lack of evidence *or* evidence that the link is forbidden through *e.g.* trait mis-match), and inter-  
15 actions with a probability that lays somewhere in between (Catchen et al., 2023). Although metawebs  
16 can (and in practice likely do) include false positives, these are statistically negligible compared to the  
17 false negatives. Furthermore, Strydom et al. (2022) shows that t-SVD embedding is extremely robust  
18 to (and able to detect) the presence of false positives. In a sense, because most ecological interactions  
19 are elusive, we should consider the direct consequences this has on sampling: once the common in-  
20 teractions are documented, the effort required in documenting each rare interaction will increase ex-  
21 ponentially (Jordano, 2016). Recent proposals in other fields relying on machine learning approaches  
22 emphasize the idea that algorithms meant to predict, through the assumption that they approximate  
23 the process generating the data, can also act as data generators (Hoffmann et al., 2019). High quality  
24 observational data can be used to infer core rules underpinning network structure, and be supple-  
25 mented with synthetic data coming from predictive models trained on them, thereby increasing the  
26 volume of information available for analysis. Indeed, Strydom, Catchen, et al. (2021) suggested that  
27 knowing the metaweb may render the prediction of local networks easier, because it fixes an “upper  
28 bound” on which interactions can exist. In this context, a probabilistic metaweb represents an aggre-  
29 gation of informative priors on the biological feasibility of interactions, which is usually hard to obtain  
30 yet has possibly the most potential to boost our predictive ability of local networks (Bartomeus, 2013;  
31 Bartomeus et al., 2016). This would represent a departure from simple rules expressed at the network  
32 scale (*e.g.* Williams & Martinez, 2000) to a view of network prediction based on learning the rules that

1 underpin interactions *and* their variability (Gupta et al., 2022).

2 [Figure 1 about here.]

### 3 **Graph embedding offers promises for the inference of potential in-** 4 **teractions**

5 Graph (or network) embedding (Figure 1) is a family of machine learning techniques, whose main task  
6 is to learn a mapping function from a discrete graph to a continuous domain (Arsov & Mirceva, 2019;  
7 Chami et al., 2022). Their main goal is to learn a low dimensional vector representation of the graph  
8 (embeddings), such that its key properties (*e.g.* local or global structures) are retained in the embed-  
9 ding space (Yan et al., 2005). The embedding space may, but will not necessarily, have lower dimen-  
10 sionality than the graph. Ecological networks are promising candidates for the routine application of  
11 embeddings, as they tend to possess a shared structural backbone (see *e.g.* Bramon Mora et al., 2018),  
12 which hints at structural invariants in empirical data. Assuming that these structural invariants are  
13 common enough, they would dominate the structure of networks, and therefore be adequately cap-  
14 tured by the first (lower) dimensions of an embedding, without the need to measure derived aspects  
15 of their structure (*e.g.* motifs, paths, modularity, ...).

### 16 **Graph embedding produces latent variables (but not traits)**

17 Before moving further, it is important to clarify the epistemic status of node values derived from em-  
18 beddings: specifically, they are *not* functional traits, and therefore should not be interpreted in terms  
19 of effects or responses. As per the framework of Malaterre et al. (2019), these values neither derive  
20 from, nor result in, changes in organismal performance, and should therefore not be used to quantify  
21 *e.g.* functional diversity. This holds true even when there are correlations between latent values and  
22 functional traits: although these enable an ecological discussion of how traits condition the structure  
23 of the network, the existence of a statistical relationship does not elevate the latent values to the status  
24 of functional traits.

25 Rather than directly predicting biological rules (see *e.g.* Pichler et al., 2020 for an overview), which may  
26 be confounded by the sparse nature of graph data, learning embeddings works in the low-dimensional  
27 space that maximizes information about the network structure. This approach is further justified by  
28 the observation, for example, that the macro-evolutionary history of a network is adequately repre-

1 sented by some graph embeddings [Random dot product graphs (RDPG); see Dalla Riva & Stouffer  
2 (2016)]. In a recent publication, Strydom et al. (2022) have used an embedding (based on RDPG) to  
3 project a metaweb of trophic interactions between European mammals, and transferred this infor-  
4 mation to mammals of Canada, using the phylogenetic distance between related clades to infer the  
5 values in the latent subspace into which the European metaweb was projected. By performing the  
6 RDPG step on re-constructed values, this approach yields a probabilistic trophic metaweb for mam-  
7 mals of Canada based on knowledge of European species, despite a limited ( $\approx 5\%$ ) taxonomic overlap,  
8 and illustrates how the values derived from an embedding can be used for prediction without being  
9 “traits” of the species they represent.

## 10 **Ecological networks are good candidates for embedding**

11 Ecological networks are inherently low-dimensional objects, and can be adequately represented with  
12 less than ten dimensions (J. Braga et al., 2019; M. P. Braga et al., 2021; Eklöf et al., 2013). Simulation  
13 results by Botella et al. (2022) suggested that there is no dominant method to identify architectural  
14 similarities between networks: multiple approaches need to be tested and compared to the network  
15 descriptor of interest on a problem-specific basis. This matches previous results on graph embedding,  
16 wherein different embedding algorithms yield different network embeddings (Goyal & Ferrara, 2018),  
17 calling for a careful selection of the problem-specific approach to use. Additionally, Ghasemian et  
18 al. (2020) suggest that in some cases, nodes embeddings can be outperformed by other methods, re-  
19 inforcing the need to thoroughly select the appropriate data analysis technique. In Table 1, we present  
20 a selection of common graph and node embedding methods, alongside examples of their use to predict  
21 interactions or statistical associations between species. These methods rely largely on linear algebra  
22 or pseudo-random walks on graphs. All forms of embeddings presented in Table 1 share the common  
23 property of summarizing their objects into (sets of) dense feature vectors, that capture the overall net-  
24 work structure, pairwise information on nodes, and emergent aspects of the network, in a compressed  
25 way (*i.e.* with some information loss, as we later discuss in the illustration). Node embeddings tend to  
26 focus on maintaining pairwise relationships (*i.e.* species interactions), while graph embeddings focus  
27 on maintaining the network structure (*i.e.* emergent properties). Nevertheless, some graph embed-  
28 ding techniques (like RDPG, see *e.g.* Wu et al., 2021) will provide high-quality node-level embeddings  
29 while also preserving network structure.

### **i** Box 1 - Graph Neural Networks

One prominent family of approaches we do not discuss in the present manuscript is Graph Neural Networks [GNN; Zhou et al. (2020)]. GNN are, in a sense, a method to embed a graph into a dense subspace, but belong to the family of deep learning methods, which has its own set of practices (see *e.g.* Goodfellow et al., 2016). An important issue with methods based on deep learning is that, because their parameter space is immense, the sample size of the data fed into them must be similarly large (typically thousands of instances). This is a requirement for the model to converge correctly during training, but this assumption is unlikely to be met given the size of datasets currently available for metawebs (or single time/location species interaction networks). This data volume requirement is mostly absent from the techniques we list below. Furthermore, GNN still have some challenges related to their shallow structure, and concerns related to scalability (see Gupta et al., 2021 for a review), which are mostly absent from the methods listed in Table 1. Assuming that the uptake of next-generation biomonitoring techniques does indeed deliver larger datasets on species interactions (Bohan et al., 2017), there is nevertheless the potential for GNN to become an applicable embedding/predictive technique in the coming years.

1

2 Graph embeddings *can* serve as a dimensionality reduction method. For example, RDPG (Strydom et  
3 al., 2022) and t-SVD [truncated Singular Value Decomposition; Poisot et al. (2021)] typically embed  
4 networks using fewer dimensions than the original network [the original network has as many dimen-  
5 sions as species, and as many informative dimensions as trophically unique species; Strydom, Dalla  
6 Riva, et al. (2021)]. However, this is not necessarily the case – indeed, one may perform a PCA (a special  
7 case of SVD) to project the raw data into a subspace that improves the efficacy of t-SNE [t-distributed  
8 stochastic neighbor embedding; Maaten (2009)]. There are many dimensionality reductions (Anowar  
9 et al., 2021) that can be applied to an embedded network should the need for dimensionality reduction  
10 (for example for data visualization) arise. In brief, many graph embeddings *can* serve as dimensional-  
11 ity reduction steps, but not all do, neither do all dimensionality reduction methods provide adequate  
12 graph embedding capacities. In the next section (and Figure 1), we show how the amount of dimen-  
13 sionality reduction can affect the quality of the embedding.

Table 1: Overview of some common graph embedding approaches, by type of embedded objects, alongside examples of their use in the prediction of species interactions. These methods have not yet been routinely used to predict species interactions; most examples that we identified were either statistical associations, or analogues to joint species distribution models. <sup>a</sup>: application is concerned with *statistical* interactions, which are not necessarily direct biotic interactions; <sup>b</sup>: application is concerned with joint-SDM-like approach, which is also very close to statistical associations as opposed to direct biotic interactions. Given the need to evaluate different methods on a problem-specific basis, the fact that many methods have not been used on network problems is an opportunity for benchmarking and method development. Note that the row for PCA also applies to kernel/probabilistic PCA, which are variations on the more general method of SVD. Note further that t-SNE has been included because it is frequently used to embed graphs, including of species associations/interactions, despite not being strictly speaking, a graph embedding technique (see *e.g.* Chami et al., 2022).

Method	Object	Technique	Reference	Application
t-SNE	nodes	statistical divergence	Hinton & Roweis (2002)	(Cieslak et al., 2020, species- environment responses <sup>a</sup> ) (Gibb et al., 2021, host-virus network representation)
LINE	nodes	stochastic gradient descent	Tang et al. (2015)	
SDNE	nodes	gradient descent	D. Wang et al. (2016)	
node2vec	nodes	stochastic gradient descent	Grover & Leskovec (2016)	
HARP	nodes	meta-strategy	H. Chen et al. (2017)	
DMSE	joint nodes	deep neural network	D. Chen et al. (2017)	(D. Chen et al., 2017, species- environment interactions <sup>b</sup> )
graph2vec	sub-graph	skipgram network	Narayanan et al. (2017)	



Method	Object	Technique	Reference	Application
RDPG	graph	SVD	Young & Scheinerman (2007)	(Dalla Riva & Stouffer, 2016, trophic interactions) (Poisot et al., 2021, host-virus network prediction)
GLEE	graph	Laplacian eigenmap	Torres et al. (2020)	
DeepWalk	graph	stochastic gradient descent	Perozzi et al. (2014)	(Wardeh et al., 2021, host-virus interactions)
GraphKKE	graph	stochastic differential equation	Melnyk et al. (2020)	(Melnyk et al., 2020, microbiome species associations <sup>a)</sup> )
FastEmbed	graph	eigen decomposition	Ramasamy & Madhow (2015)	
PCA	graph	eigen decomposition	Surendran (2013)	(Strydom, Catchen, et al., 2021, host-parasite interactions)
Joint methods	multiple graphs	multiple strategies	S. Wang et al. (2021)	

<sup>1</sup> The popularity of graph embedding techniques in machine learning is more than the search for struc-  
<sup>2</sup> tural invariants: graphs are discrete objects, and machine learning techniques tend to handle contin-  
<sup>3</sup> uous data better. Bringing a sparse graph into a continuous, dense vector space (Xu, 2021) opens up

1 a broader variety of predictive algorithms, notably of the sort that are able to predict events as prob-  
2 abilities (Murphy, 2022). Furthermore, the projection of the graph itself is a representation that can  
3 be learned; Runghen et al. (2021), for example, used a neural network to learn the embedding of a  
4 network in which not all interactions were known, based on the nodes' metadata. This example has  
5 many parallels in ecology (see Figure 1 C), in which node metadata can be represented by phylogeny,  
6 abundance, or functional traits. Using phylogeny as a source of information assumes (or strives to cap-  
7 ture) the action of evolutionary processes on network structure, which at least for networks have been  
8 well documented (M. P. Braga et al., 2021; Dalla Riva & Stouffer, 2016; Eklöf & Stouffer, 2016; Stouf-  
9 fer et al., 2007; Stouffer et al., 2012); similarly, the use of functional traits assumes that interactions  
10 can be inferred from the knowledge of trait-matching rules, which is similarly well supported in the  
11 empirical literature (Bartomeus, 2013; Bartomeus et al., 2016; Goebel et al., 2023; Gravel et al., 2013).  
12 Relating this information to an embedding rather than a list of network measures would allow to cap-  
13 ture their effect on the more fundamental aspects of network structure; conversely, the absence of a  
14 phylogenetic or functional signal may suggest that evolutionary/trait processes are not strong drivers  
15 of network structure, therefore opening a new way to perform hypothesis testing.

## 16 **An illustration of metaweb embedding**

17 In this section, we illustrate the embedding of a collection of bipartite networks collected by Hadfield et  
18 al. (2014), using t-SVD and RDPG. Briefly, an RDPG decomposes a network into two subspaces (left and  
19 right), which are matrices that when multiplied give an approximation of the original network. RDPG  
20 has the particularly desirable properties of being a graph embedding technique that produces rele-  
21 vant node-level feature vectors, and provides good approximations of graphs with varied structures  
22 (Athreya et al., 2017). The code to reproduce this example is available as supplementary material (note,  
23 for the sake of comparison, that Strydom, Catchen, et al., 2021 have an example using embedding  
24 through PCA followed by prediction using a deep neural network on the same dataset). The resulting  
25 (binary) metaweb  $\mathcal{M}$  has 2131 interactions between 206 parasites and 121 hosts, and its adjacency  
26 matrix has full rank (*i.e.* it represents a space with 121 dimensions). All analyses were done using Julia  
27 (Bezanson et al., 2017) version 1.7.2, *Makie.jl* (Danisch & Krumbiegel, 2021), and *EcologicalNetworks.jl*  
28 (Poisot et al., 2019).

29 [Figure 2 about here.]

30 In Figure 2, we focus on some statistical checks of the embedding. In panel A, we show that the av-

1 eraged  $L_2$  loss (*i.e.* the mean of squared errors) between the empirical and reconstructed metaweb  
2 decreases when the number of dimensions (rank) of the subspace increases, with an inflection at 39  
3 dimensions (out of 120 initially) according to the finite differences method. As discussed by Runghen  
4 et al. (2021), there is often a trade-off between the number of dimensions to use (more dimensions are  
5 more computationally demanding) and the quality of the representation. In panel **B**, we show the in-  
6 crease in cumulative variance explained at each rank, and visualize that using 39 ranks explains about  
7 70% of the variance in the empirical metaweb. This provides different information from the  $L_2$  loss  
8 (which is averaged across interactions), as it works on the eigenvalues of the embedding, and there-  
9 fore captures higher-level features of the network. In panel **C**, we show positions of hosts and parasites  
10 on the first two dimensions of the left and right subspaces. Note that these values largely skew nega-  
11 tive, because the first dimensions capture the coarse structure of the network: most pairs of species  
12 do not interact, and therefore have negative values. Finally in panel **D**, we show the predicted weight  
13 (*i.e.* the result of the multiplication of the RDGP subspaces at a rank of 39) as a function of whether  
14 the interactions are observed, not-observed, or unknown due to lack of co-occurrence in the original  
15 dataset. This reveals that the observed interactions have higher predicted weights, although there is  
16 some overlap; the usual approach to identify potential interactions based on this information would  
17 be a thresholding analysis, which is outside the scope of this manuscript (and is done in the papers  
18 cited in this illustration). Because the values returned from RDGP are not bound to the unit interval, we  
19 performed a clamping of the weights to the unit space, showing a one-inflation in documented inter-  
20 actions, and a zero-inflation in other species pairs. Panel **D** specifically shows that species pairs with  
21 no documented co-occurrence have weights that are not distinguishable from species pairs with no  
22 documented interactions; in other words, looking at the embedding, species that do not co-occur are  
23 not easily distinguished from species that do not interact. This suggests that (as befits a host-parasite  
24 model) the ability to interact is a strong predictor of co-occurrence.

25 [Figure 3 about here.]

26 In Figure 3, we relate the values of latent variables for hosts to different ecologically-relevant data; we  
27 can perform this additional step, because the results presented in Figure 2 show that we can extract  
28 an embedding of the metaweb that captures enough variance to be relevant. Importantly, this is true  
29 for both  $L_2$  loss (indicating that RDGP is able to capture pairwise processes) and the cumulative vari-  
30 ance explained (indicating that RDGP is able to capture network-level structure), which suggests that  
31 these approaches may allow to predict interactions *and* network structure. In panel **A**, we show that  
32 host with a higher value on the first dimension have fewer parasites. This relates to the body size of

1 hosts in the *PanTHERIA* database (Jones et al., 2009), as shown in panel **B**: interestingly, the position  
2 on the first axis is only weakly correlated to body mass of the host; this matches well established re-  
3 sults showing that body size/mass is not always a direct predictor of parasite richness in terrestrial  
4 mammals (Morand & Poulin, 1998), a result we observe in panel **C**. Finally, in panel **D**, we can see  
5 how different taxonomic families occupy different positions on the first axis, with *e.g.* Sciuridae be-  
6 ing biased towards higher values. These results show how we can look for ecological informations in  
7 the output of network embeddings, which can further be refined into the selection of predictors for  
8 transfer learning.

## 9 **The metaweb merges ecological hypotheses and practices**

10 Metaweb inference seeks to provide information about the interactions between species at a large spa-  
11 tial scale, typically a scale large enough to be considered of biogeographic relevance (indeed, many of  
12 the examples covered in the introduction span areas larger than a country, some of them global). But  
13 as Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”;  
14 any inference of a metaweb must therefore contend with several novel, interwoven, families of prob-  
15 lems. In this section, we outline three that we think are particularly important, and discuss how they  
16 may be addressed with subsequent data analysis or simulations, and how they emerge in the specific  
17 context of using embeddings; some of these issues are related to the application of these methods at  
18 the science-policy interface. Addressing these considerations as part of the methodological discussion  
19 is particularly important, as the construction of metawebs can perpetuate legacies of biases in data  
20 (Box 2).

## 21 **Identifying the properties of the network to embed**

22 If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the chances  
23 decrease of finding another area with enough related species (through phylogenetic relatedness or  
24 similarity of functional traits) to make a reliable inference. This is because transfer requires similar-  
25 ity (Figure 1). A diagnostic for the lack of similar species would likely be large confidence intervals  
26 during estimation of the values in the low-rank space. In other words, the representation of the orig-  
27 inal graph is difficult to transfer to the new problem. Alternatively, if the initial metaweb is too large  
28 (taxonomically), then the resulting embeddings would need to represent interactions between taxo-  
29 nomic groups that are not present in the new location. This would lead to a much higher variance

1 in the starting dataset, and to under-dispersion in the target dataset, resulting in the potential un-  
2 der or over estimation of the strength of new predicted interactions. Llewelyn et al. (2022) provided  
3 compelling evidence for these situations by showing that, even at small spatial scales, the transfer of  
4 information about interactions becomes more challenging when areas rich with endemic species are  
5 considered. The lack of well documented metawebs is currently preventing the development of more  
6 concrete guidelines. The question of phylogenetic relatedness and distribution is notably relevant if  
7 the metaweb is assembled in an area with mostly endemic species (*e.g.* a system that has undergone re-  
8 cent radiation or that has remained in isolation for a long period of time might not have an analogous  
9 system with which to draw knowledge from), and as with every predictive algorithm, there is room  
10 for the application of our best ecological judgement. Because this problem relates to distribution of  
11 species in the geographic or phylogenetic space, it can certainly be approached through assessing the  
12 performance of embedding transfer in simulated starting/target species pools.

### 13 **Identifying the scope of the prediction to perform**

14 The area for which we seek to predict the metaweb should determine the species pool on which the  
15 embedding is performed. Metawebs can be constructed by assigning interactions in a list of species  
16 within specific regions. The upside of this approach is that information relevant for the construction  
17 of this dataset is likely to exist, as countries usually set conservation goals at the national level (Buxton  
18 et al., 2021), and as quantitative instruments are consequently designed to work at these scales (Turak  
19 et al., 2017); specific strategies are often enacted at smaller scales, nested within a specific country  
20 (Ray et al., 2021). However, there is no guarantee that these arbitrary boundaries are meaningful. In  
21 fact, we do not have a satisfying answer to the question of “where does an ecological network stop?”,  
22 the answer to which would dictate the spatial span to embed/predict. Recent results by Martins et  
23 al. (2022) suggested that networks are shaped within eco-regions, with abrupt structural transitions  
24 from an eco-region to the next. Should this trend hold generally, this would provide an ecologically-  
25 relevant scale at which metawebs can be downscaled and predicted. Other solutions could leverage  
26 network-area relationships to identify areas in which networks are structurally similar (see *e.g.* Fortin  
27 et al., 2021; Galiana et al., 2018, 2022). Both of these solutions require ample pre-existing information  
28 about the network in space. Nevertheless, the inclusion of species for which we have data but that are  
29 not in the right spatial extent *may* improve the performance of approaches based on embedding and  
30 transfer, *if* they increase the similarity between the target and destination network. This proposal can  
31 specifically be evaluated by adding nodes to the network to embed, and assessing the performance of

1 predictive models (see *e.g.* Llewelyn et al., 2022).

## 2 **Putting models in their context**

3 Predictive approaches in ecology, regardless of the scale at which they are deployed and the intent of  
4 their deployment, originate in the framework that contributed to the ongoing biodiversity crisis (Adam,  
5 2014) and reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020). The risk  
6 of embedding this legacy in our models is real, especially when the impact of this legacy on species  
7 pools is being increasingly documented. This problem can be addressed by re-framing the way we  
8 interact with models, especially when models are intended to support conservation actions. Particu-  
9 larly on territories that were traditionally stewarded by Indigenous people, we must interrogate how  
10 predictive approaches and the biases that underpin them can be put to task in accompanying Indige-  
11 nous principles of land management (Eichhorn et al., 2019; No'kmaq et al., 2021). The discussion  
12 of “algorithm-in-the-loop” approaches that is now pervasive in the machine learning community pro-  
13 vides examples of why this is important. Human-algorithm interactions are notoriously difficult and  
14 can yield adverse effects (Green & Chen, 2019; Stevenson & Doleac, 2021), suggesting the need to  
15 systematically study them for the specific purpose of, here, biodiversity governance. Improving the al-  
16 gorithmic literacy of decision makers is part of the solution (*e.g.* Lamba et al., 2019; Mosebo Fernandes  
17 et al., 2020), as we can reasonably expect that model outputs will be increasingly used to drive policy  
18 decisions (Weiskopf et al., 2022). Our discussion of these approaches need to go beyond the techni-  
19 cal and statistical, and into the governance consequences they can have. To embed data also embeds  
20 historical and contemporary biases that acted on these data, both because they shaped the ecolog-  
21 ical processes generating them (see Box 2), and the global processes leading to their measurement  
22 and publication. For a domain as vast as species interaction networks, these biases exist at multiple  
23 scales along the way, and a challenge for prediction is not only to develop (or adopt) new quantitative  
24 tools, but to assess the behavior of these tools in the proper context.

## 25 **Conclusion**

26 Although promising, the application of embeddings to metaweb prediction still involved several chal-  
27 lenges. First, there is a need to understand how to define a metaweb as a single, cohesive, unit of  
28 ecological organisation. This is likely to have very different answers based on the specific taxonomic  
29 group, temporal and spatial resolution, and question being investigated. Second, there is a need to un-

1 derstand the scale at which these predictions are relevant. Although we have documented many cases  
2 of using embedding to fill gaps in the metaweb, these techniques can likely be brought into a spatial  
3 (and possibly temporal) context. The validation of these predictions will have to proceed jointly with  
4 empirical sampling of interactions, but also with the design of downsampling methods. Finally, there  
5 is a need for a greater understanding of how biases in the data propagate to the predictions. Because  
6 the volume of metawebs is currently low, and because graph embeddings have not been commonly  
7 applied, we anticipate that this discussion will take place organically in the coming years.

### **i** Box 2 - Minding legacies shaping ecological datasets

In large parts of the world, boundaries that delineate geographic regions are a legacy of settler colonialism, which drives global disparity in capacity to collect and publish ecological data. Applying any embedding to biased data does not debias them, but rather embeds these biases, propagating them to the models using embeddings to make predictions. Furthermore, the use of ecological data itself is not an apolitical act (Nost & Goldstein, 2021): data infrastructures tend to be designed to answer questions within national boundaries (therefore placing contingencies on what is available to be embedded), their use often drawing upon, and reinforcing, territorial statecraft (see *e.g.* Barrett, 2005). As per Machen & Nost (2021), these biases are particularly important to consider when knowledge generated algorithmically is used to supplement or replace human decision-making, especially for governance (*e.g.* enacting conservation decisions on the basis of model prediction). As information on networks is increasingly leveraged for conservation actions (see *e.g.* Eero et al., 2021; Naman et al., 2022; Stier et al., 2017), the need to appraise and correct biases that are unwittingly propagated to algorithms when embedded from the original data is immense. These considerations are even more urgent in the specific context of biodiversity data. Long-term colonial legacies still shape taxonomic composition to this day (Lenzner et al., 2022; Raja, 2022), and much shorter-term changes in taxonomic and genetic richness of wildlife emerged through environmental racism (Schmidt & Garroway, 2022). Thus, the set of species found at a specific location is not only as the result of a response to ecological processes separate from human influence, but also the result of human-environment interaction as well as the results of legislative/political histories.

8

9 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional  
10 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and  
11 Omàmiwininiwak nations. We thank Victoria Hemming for feedback on this manuscript. TP, TS, DC,

1 and LP received funding from the Canadian Institute for Ecology & Evolution. FB is funded by the Insti-  
2 tute for Data Valorization (IVADO). TS, SB, and TP are funded by a donation from the Courtois Founda-  
3 tion. M-JF acknowledges funding from NSERC Discovery Grant and NSERC CRC. RR is funded by New  
4 Zealand's Biological Heritage Ngā Koiora Tuku Iho National Science Challenge, administered by New  
5 Zealand Ministry of Business, Innovation, and Employment. BM is funded by the NSERC Alexander  
6 Graham Bell Canada Graduate Scholarship and the FRQNT master's scholarship. LP acknowledges  
7 funding from NSERC Discovery Grant (NSERC RGPIN-2019-05771). TP acknowledges financial sup-  
8 port from the Fondation Courtois, and NSERC through the Discovery Grants and Discovery Accelerator  
9 Supplement programs. MJF is supported by an NSERC PDF and an RBC Post-Doctoral Fellowship.

10 **Conflict of interest:** The authors have no conflict of interests to disclose

11 **Authors' contributions:** TS and TP conceived the ideas discussed in the manuscript. All authors con-  
12 tributed to writing and editing the manuscript.

13 **Data availability:** The code to collect the data use in figures is distributed as part of the Supplementary  
14 Material.

## 15 **References**

- 16 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 17 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,  
18 Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel,  
19 D. (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8, 8), 1153–  
20 1161. <https://doi.org/10.1038/s41559-019-0950-y>
- 21 Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality  
22 reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science*  
23 *Review*, 40, 100378. <https://doi.org/10.1016/j.cosrev.2021.100378>
- 24 Arsov, N., & Mirceva, G. (2019). *Network Embedding: An Overview*. <http://arxiv.org/abs/1911.11726>
- 25 Athreya, A., Fishkind, D. E., Levin, K., Lyzinski, V., Park, Y., Qin, Y., Sussman, D. L., Tang, M., Vo-  
26 gelstein, J. T., & Priebe, C. E. (2017). *Statistical inference on random dot product graphs: A survey*  
27 (arXiv:1709.05454). arXiv. <http://arxiv.org/abs/1709.05454>
- 28 Barrett, S. (2005). *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (1st ed.). Ox-  
29 ford University PressOxford. <https://doi.org/10.1093/0199286094.001.0001>
- 30 Bartomeus, I. (2013). Understanding linkage rules in plant-pollinator networks by using hierarchical



1 models that incorporate pollinator detectability and plant traits. *PloS One*, 8(7), e69200. [http://](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069200)  
2 [journals.plos.org/plosone/article?id=10.1371/journal.pone.0069200](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069200)

3 Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A  
4 common framework for identifying linkage rules across different types of interactions. *Functional*  
5 *Ecology*, 30(12), 1894–1903. <http://onlinelibrary.wiley.com/doi/10.1111/1365-2435.12666/full>

6 Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical  
7 Computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>

8 Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interac-  
9 tions. *Ecology Letters*.

10 Bohan, D. A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017).  
11 Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Net-  
12 works. *Trends in Ecology & Evolution*. <https://doi.org/10.1016/j.tree.2017.03.001>

13 Botella, C., Dray, S., Matias, C., Miele, V., & Thuiller, W. (2022). An appraisal of graph embeddings for  
14 comparing trophic network architectures. *Methods in Ecology and Evolution*, 13(1), 203–216. <https://doi.org/10.1111/2041-210X.13738>

15

16 Braga, J., Pollock, L. J., Barros, C., Galiana, N., Montoya, J. M., Gravel, D., Maiorano, L., Montemag-  
17 giori, A., Ficetola, G. F., Dray, S., & Thuiller, W. (2019). Spatial analyses of multi-trophic terres-  
18 trial vertebrate assemblages in Europe. *Global Ecology and Biogeography*, 28(11), 1636–1648. <https://doi.org/10.1111/geb.12981>

19

20 Braga, M. P., Janz, N., Nylin, S., Ronquist, F., & Landis, M. J. (2021). Phylogenetic reconstruction of an-  
21 cestral ecological networks through time for pierid butterflies and their host plants. *Ecology Letters*,  
22 *n/a(n/a)*. <https://doi.org/10.1111/ele.13842>

23 Bramon Mora, B., Gravel, D., Gilarranz, L. J., Poisot, T., & Stouffer, D. B. (2018). Identifying a common  
24 backbone of interactions underlying food webs from different ecosystems. *Nature Communications*,  
25 9(1), 2603. <https://doi.org/10.1038/s41467-018-05056-0>

26 Buxton, R. T., Bennett, J. R., Reid, A. J., Shulman, C., Cooke, S. J., Francis, C. M., Nyboer, E. A., Pritchard,  
27 G., Binley, A. D., Avery-Gomm, S., Ban, N. C., Beazley, K. F., Bennett, E., Blight, L. K., Bortolotti, L.  
28 E., Camfield, A. F., Gadallah, F., Jacob, A. L., Naujokaitis-Lewis, I., ... Smith, P. A. (2021). Key infor-  
29 mation needs to move from knowledge to action for biodiversity conservation in Canada. *Biological*  
30 *Conservation*, 256, 108983. <https://doi.org/10.1016/j.biocon.2021.108983>

31 Carlson, C. J., Albery, G. F., Merow, C., Trisos, C. H., Zipfel, C. M., Eskew, E. A., Olival, K. J., Ross,  
32 N., & Bansal, S. (2022). Climate change increases cross-species viral transmission risk. *Nature*,  
33 607(7919), 555–562. <https://doi.org/10.1038/s41586-022-04788-w>

- 1 Catchen, M., Poisot, T., Pollock, L., & Gonzalez, A. (2023). *The missing link: Discerning true from false*  
2 *negatives when sampling species interaction networks* (Preprint 4929). EcoEvoRXiv. [https://doi.org/10.](https://doi.org/10.32942/X2DW22)  
3 [32942/X2DW22](https://doi.org/10.32942/X2DW22)
- 4 Cazelles, K., Araújo, M. B., Mouquet, N., & Gravel, D. (2016). A theory for species co-occurrence in  
5 interaction networks. *Theoretical Ecology*, 9(1), 39–48. <https://doi.org/10.1007/s12080-015-0281-9>
- 6 Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., & Murphy, K. (2022). Machine Learning on Graphs: A  
7 Model and Comprehensive Taxonomy. *Journal of Machine Learning Research*, 23(89), 1–64. [http://](http://jmlr.org/papers/v23/20-852.html)  
8 [jmlr.org/papers/v23/20-852.html](http://jmlr.org/papers/v23/20-852.html)
- 9 Chen, D., Xue, Y., Fink, D., Chen, S., & Gomes, C. P. (2017). *Deep Multi-species Embedding*. 3639–3646.  
10 <https://www.ijcai.org/proceedings/2017/509>
- 11 Chen, H., Perozzi, B., Hu, Y., & Skiena, S. (2017). *HARP: Hierarchical Representation Learning for Networks*.  
12 <http://arxiv.org/abs/1706.07845>
- 13 Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity, colo-  
14 nialism and conquest in an era of capitalist globalization. In *NGOization: Complicity, contradictions*  
15 *and prospects* (pp. 24–44). Bloomsbury Publishing.
- 16 Cieslak, M. C., Castelfranco, A. M., Roncalli, V., Lenz, P. H., & Hartline, D. K. (2020). T-Distributed  
17 Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Ma-*  
18 *rine Genomics*, 51, 100723. <https://doi.org/10.1016/j.margen.2019.100723>
- 19 Csermely, P. (2004). Strong links are important, but weak links stabilize them. *Trends in Biochemical*  
20 *Sciences*, 29(7), 331–334. <https://doi.org/10.1016/j.tibs.2004.05.004>
- 21 Dalla Riva, G. V., & Stouffer, D. B. (2016). Exploring the evolutionary signature of food webs' backbones  
22 using functional traits. *Oikos*, 125(4), 446–456. <https://doi.org/10.1111/oik.02305>
- 23 Dallas, T., Park, A. W., & Drake, J. M. (2017). Predicting cryptic links in host-parasite networks. *PLOS*  
24 *Computational Biology*, 13(5), e1005557. <https://doi.org/10.1371/journal.pcbi.1005557>
- 25 Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia.  
26 *Journal of Open Source Software*, 6(65), 3349. <https://doi.org/10.21105/joss.03349>
- 27 Domínguez, L., & Luoma, C. (2020). Decolonising Conservation Policy: How Colonial Land and Conser-  
28 vation Ideologies Persist and Perpetuate Indigenous Injustices at the Expense of the Environment.  
29 *Land*, 9(3, 3), 65. <https://doi.org/10.3390/land9030065>
- 30 Dunne, J. A. (2006). The Network Structure of Food Webs. In J. A. Dunne & M. Pascual (Eds.), *Ecological*  
31 *networks: Linking structure and dynamics* (pp. 27–86). Oxford University Press.
- 32 Eero, M., Dierking, J., Humborg, C., Undeman, E., MacKenzie, B. R., Ojaveer, H., Salo, T., & Köster, F.  
33 W. (2021). Use of food web knowledge in environmental conservation and management of living

1 resources in the Baltic Sea. *ICES Journal of Marine Science*, 78(8), 2645–2663. [https://doi.org/10.](https://doi.org/10.1093/icesjms/fsab145)  
2 [1093/icesjms/fsab145](https://doi.org/10.1093/icesjms/fsab145)

3 Eichhorn, M. P., Baker, K., & Griffiths, M. (2019). Steps towards decolonising biogeography. *Frontiers of*  
4 *Biogeography*, 12(1), 1–7. <https://doi.org/10.21425/F5FBG44795>

5 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti,  
6 M., Guimarães, P. R., Lomáscolo, S. B., Martín González, A. M., Pizo, M. A., Rader, R., Rodrigo, A.,  
7 Tylianakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks.  
8 *Ecology Letters*, 16(5), 577–583. <https://doi.org/10.1111/ele.12081>

9 Eklöf, A., & Stouffer, D. B. (2016). The phylogenetic component of food web structure and intervality.  
10 *Theoretical Ecology*, 9(1), 107–115. <https://doi.org/10.1007/s12080-015-0273-9>

11 Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes. *Pro-*  
12 *ceedings of the Royal Society B: Biological Sciences*, 288(1949), rspb.2020.1889, 20201889. [https://doi.](https://doi.org/10.1098/rspb.2020.1889)  
13 [org/10.1098/rspb.2020.1889](https://doi.org/10.1098/rspb.2020.1889)

14 Fricke, E. C., Ordonez, A., Rogers, H. S., & Svenning, J.-C. (2022). The effects of defaunation on plants'  
15 capacity to track climate change. *Science*. [https://www.science.](https://www.science.org/doi/abs/10.1126/science.abk3510)  
16 [org/doi/abs/10.1126/science.](https://www.science.org/doi/abs/10.1126/science.abk3510)  
[abk3510](https://www.science.org/doi/abs/10.1126/science.abk3510)

17 Galiana, N., Lurgi, M., Bastazini, V. A. G., Bosch, J., Cagnolo, L., Cazelles, K., Claramunt-López, B., Emer,  
18 C., Fortin, M.-J., Grass, I., Hernández-Castellano, C., Jauker, F., Leroux, S. J., McCann, K., McLeod, A.  
19 M., Montoya, D., Mulder, C., Osorio-Canadas, S., Reverté, S., ... Montoya, J. M. (2022). Ecological net-  
20 work complexity scales with area. *Nature Ecology & Evolution*, 1–8. [https://doi.org/10.1038/s41559-](https://doi.org/10.1038/s41559-021-01644-4)  
21 [021-01644-4](https://doi.org/10.1038/s41559-021-01644-4)

22 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,  
23 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),  
24 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

25 Gaucher, S., Klopp, O., & Robin, G. (2021). Outlier detection in networks with missing links. *Computa-*  
26 *tional Statistics & Data Analysis*, 164, 107308. <https://doi.org/10.1016/j.csda.2021.107308>

27 Ghasemian, A., Hosseinmardi, H., Galstyan, A., Airoidi, E. M., & Clauset, A. (2020). Stacking models for  
28 nearly optimal link prediction in complex networks. *Proceedings of the National Academy of Sciences*,  
29 117(38), 23393–23400. <https://doi.org/10.1073/pnas.1914950117>

30 Gibb, R., Albery, G. F., Becker, D. J., Brierley, L., Connor, R., Dallas, T. A., Eskew, E. A., Farrell, M. J.,  
31 Rasmussen, A. L., Ryan, S. J., Sweeny, A., Carlson, C. J., & Poisot, T. (2021). Data Proliferation, Rec-  
32 onciliation, and Synthesis in Viral Ecology. *BioScience*, 71(11), 1148–1156. [https://doi.org/10.1093/](https://doi.org/10.1093/biosci/biab080)  
33 [biosci/biab080](https://doi.org/10.1093/biosci/biab080)

- 1 Goebel, L. G. A., Vitorino, B. D., Frota, A. V. B., & Santos-Filho, M. dos. (2023). Body mass determines  
2 the role of mammal species in a frugivore-large fruit interaction network in a Neotropical savanna.  
3 *Journal of Tropical Ecology*, 39, e12. <https://doi.org/10.1017/S0266467422000505>
- 4 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- 5 Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey.  
6 *Knowledge-Based Systems*, 151, 78–94. <https://doi.org/10.1016/j.knosys.2018.03.022>
- 7 Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B.,  
8 Tylianakis, J. M., Wood, S. A., & Roslin, T. (2018). Bringing Elton and Grinnell together: A quanti-  
9 tative framework to represent the biogeography of ecological interaction networks. *Ecography*, 0(0).  
10 <https://doi.org/10.1111/ecog.04006>
- 11 Gravel, D., Poisot, T., Albouy, C., Velez, L., & Mouillot, D. (2013). Inferring food web structure from  
12 predator-prey body size relationships. *Methods in Ecology and Evolution*, 4(11), 1083–1090. <https://doi.org/10.1111/2041-210X.12103>
- 13
- 14 Green, B., & Chen, Y. (2019). Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness  
15 in Risk Assessments. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 90–99.  
16 <https://doi.org/10.1145/3287560.3287563>
- 17 Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*  
18 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864. <https://doi.org/10.1145/2939672.2939754>
- 19
- 20 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs  
21 shift towards increased linkage and suitability overlap under climate change. *Communications Biol-*  
22 *ogy*, 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>
- 23 Gupta, A., Furrer, R., & Petchey, O. L. (2022). Simultaneously estimating food web connectance and  
24 structure with uncertainty. *Ecology and Evolution*, 12(3), e8643. <https://doi.org/10.1002/ece3.8643>
- 25 Gupta, A., Matta, P., & Pant, B. (2021). Graph neural network: Current state of Art, challenges and  
26 applications. *Materials Today: Proceedings*, 46, 10927–10932. [https://doi.org/10.1016/j.matpr.2021.](https://doi.org/10.1016/j.matpr.2021.01.950)  
27 [01.950](https://doi.org/10.1016/j.matpr.2021.01.950)
- 28 Hadfield, J. D., Krasnov, B. R., Poulin, R., & Nakagawa, S. (2014). A Tale of Two Phylogenies: Compar-  
29 ative Analyses of Ecological Interactions. *The American Naturalist*, 183(2), 174–187. [https://doi.org/](https://doi.org/10.1086/674445)  
30 [10.1086/674445](https://doi.org/10.1086/674445)
- 31 Herbert, F. (1965). *Dune* (1st ed.). Chilton Book Company.
- 32 Hinton, G., & Roweis, S. T. (2002). Stochastic neighbor embedding. *NIPS*, 15, 833–840.
- 33 Hoffmann, J., Bar-Sinai, Y., Lee, L. M., Andrejevic, J., Mishra, S., Rubinstein, S. M., & Rycroft,

- 1 C. H. (2019). Machine learning in a data-limited regime: Augmenting experiments with  
2 synthetic data uncovers order in crumpled sheets. *Science Advances*, 5(4), eaau6792. <https://doi.org/10.1126/sciadv.aau6792>
- 3
- 4 Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven  
5 Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and*  
6 *Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- 7 Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes,  
8 E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price,  
9 S. A., Rigby, E. A., Rist, J., ... Purvis, A. (2009). PanTHERIA: A species-level database of life history,  
10 ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184.  
11 *Ecology*, 90(9), 2648–2648. <https://doi.org/10.1890/08-1494.1>
- 12 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30(12), 1883–  
13 1893. <https://doi.org/10.1111/1365-2435.12763>
- 14 Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation.  
15 *Current Biology*, 29(19), R977–R982. <https://doi.org/10.1016/j.cub.2019.08.016>
- 16 Lenzner, B., Latombe, G., Schertler, A., Seebens, H., Yang, Q., Winter, M., Weigelt, P., van Kleunen, M.,  
17 Pyšek, P., Pergl, J., Kreft, H., Dawson, W., Dullinger, S., & Essl, F. (2022). Naturalized alien floras still  
18 carry the legacy of European colonialism. *Nature Ecology & Evolution*, 1–10. <https://doi.org/10.1038/s41559-022-01865-1>
- 19
- 20 Llewelyn, J., Strona, G., Dickman, C. R., Greenville, A. C., Wardle, G. M., Lee, M. S. Y., Doherty, S., Shabani,  
21 F., Saltré, F., & Bradshaw, C. J. A. (2022). *Predicting predator-prey interactions in terrestrial endotherms*  
22 *using random forest* [Preprint]. Ecology. <https://doi.org/10.1101/2022.09.02.506446>
- 23 Maaten, L. van der. (2009). Learning a Parametric Embedding by Preserving Local Structure. *Pro-*  
24 *ceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, 384–391. <https://proceedings.mlr.press/v5/maaten09a.html>
- 25
- 26 Machen, R., & Nost, E. (2021). Thinking algorithmically: The making of hegemonic knowledge in cli-  
27 mate governance. *Transactions of the Institute of British Geographers*, 46(3), 555–569. <https://doi.org/10.1111/tran.12441>
- 28
- 29 Malaterre, C., Dussault, A. C., Mermans, E., Barker, G., Beisner, B. E., Bouchard, F., Desjardins, E.,  
30 Handa, I. T., Kembel, S. W., Lajoie, G., Maris, V., Munson, A. D., Odenbaugh, J., Poisot, T., Shapiro, B.  
31 J., & Suttle, C. A. (2019). Functional Diversity: An Epistemic Roadmap. *BioScience*, 69(10), 800–811.  
32 <https://doi.org/10.1093/biosci/biz089>
- 33 Martins, L. P., Stouffer, D. B., Blendinger, P. G., Böhning-Gaese, K., Buitrón-Jurado, G., Correia,

1 M., Costa, J. M., Dehling, D. M., Donatti, C. I., Emer, C., Galetti, M., Heleno, R., Jordano, P.,  
2 Menezes, Í., Morante-Filho, J. C., Muñoz, M. C., Neuschulz, E. L., Pizo, M. A., Quitián, M.,  
3 ... Tylianakis, J. M. (2022). Global and regional ecological boundaries explain abrupt spa-  
4 tial discontinuities in avian frugivory interactions. *Nature Communications*, 13(1, 1), 6943.  
5 <https://doi.org/10.1038/s41467-022-34355-w>

6 McLeod, A., Leroux, S. J., Gravel, D., Chu, C., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Poisot, T., & Wood,  
7 S. A. (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*,  
8 *n/a(n/a)*. <https://doi.org/10.1111/oik.08650>

9 Melnyk, K., Klus, S., Montavon, G., & Conrad, T. O. F. (2020). GraphKKE: Graph Kernel Koopman em-  
10 bedding for human microbiome analysis. *Applied Network Science*, 5(1), 96. <https://doi.org/10.1007/s41109-020-00339-2>

11

12 Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from  
13 proxies. *Trends in Ecology & Evolution*, 30(6), 347–356. <https://doi.org/10.1016/j.tree.2015.03.014>

14 Morales-Castilla, I., Pappalardo, P., Farrell, M. J., Aguirre, A. A., Huang, S., Gehman, A.-L. M., Dallas, T.,  
15 Gravel, D., & Davies, T. J. (2021). Forecasting parasite sharing under climate change. *Philosophical*  
16 *Transactions of the Royal Society B: Biological Sciences*, 376(1837), 20200360. <https://doi.org/10.1098/rstb.2020.0360>

17

18 Morand, S., & Poulin, R. (1998). Density, body mass and parasite species richness of terrestrial mam-  
19 mals. *Evolutionary Ecology*, 12(6), 717–727. <https://doi.org/10.1023/A:1006537600093>

20 Mosebo Fernandes, A. C., Quintero Gonzalez, R., Lenihan-Clarke, M. A., Leslie Trotter, E. F., & Jokar  
21 Arsanjani, J. (2020). Machine Learning for Conservation Planning in a Changing Climate. *Sustain-*  
22 *ability*, 12(18, 18), 7657. <https://doi.org/10.3390/su12187657>

23 Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press. [probml.ai](https://probml.ai)

24 Naman, S. M., White, S. M., Bellmore, J. R., McHugh, P. A., Kaylor, M. J., Baxter, C. V., Danehy, R. J.,  
25 Naiman, R. J., & Puls, A. L. (2022). Food web perspectives and methods for riverine fish conserva-  
26 tion. *WIREs Water*, *n/a(n/a)*, e1590. <https://doi.org/10.1002/wat2.1590>

27 Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., & Jaiswal, S. (2017). *Graph2vec:*  
28 *Learning Distributed Representations of Graphs*. <http://arxiv.org/abs/1707.05005>

29 Neutel, A.-M., Heesterbeek, J. A. P., & de Ruiter, P. C. (2002). Stability in Real Food Webs: Weak Links  
30 in Long Loops. *Science*, 296(5570), 1120–1123. <https://doi.org/10.1126/science.1068326>

31 No'kmaq, M., Marshall, A., Beazley, K. F., Hum, J., Joudry, shalan, Papadopoulos, A., Pictou, S., Rabesca,  
32 J., Young, L., & Zurba, M. (2021). “Awakening the sleeping giant”: Re-Indigenization principles for  
33 transforming biodiversity conservation in Canada and beyond. *FACETS*, 6(1), 839–869.

- 1 Nost, E., & Goldstein, J. E. (2021). A political ecology of data. *Environment and Planning E: Nature and*  
2 *Space*, 25148486211043503. <https://doi.org/10.1177/25148486211043503>
- 3 O'Connor, L. M. J., Pollock, L. J., Braga, J., Ficetola, G. F., Maiorano, L., Martinez-Almoyna, C., Mon-  
4 temaggiore, A., Ohlmann, M., & Thuiller, W. (2020). Unveiling the food webs of tetrapods across  
5 Europe through the prism of the Eltonian niche. *Journal of Biogeography*, 47(1), 181–192. <https://doi.org/10.1111/jbi.13773>
- 6
- 7 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen,  
8 H., Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the  
9 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*,  
10 4(7), 170215. <https://doi.org/10.1098/rsos.170215>
- 11 Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. *Pro-*  
12 *ceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–  
13 710. <https://doi.org/10.1145/2623330.2623732>
- 14 Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M., & Hartig, F. (2020). Machine learning algorithms  
15 to infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology*  
16 *and Evolution*, 11(2), 281–293. <https://doi.org/10.1111/2041-210X.13329>
- 17 Poisot, T. (2023). Guidelines for the prediction of species interactions through binary classification.  
18 *Methods in Ecology and Evolution*, 14(5), 1333–1345. <https://doi.org/10.1111/2041-210x.14071>
- 19 Poisot, T., Belisle, Z., Hoebeke, L., Stock, M., & Szefer, P. (2019). EcologicalNetworks.jl - analysing  
20 ecological networks. *Ecography*. <https://doi.org/10.1111/ecog.04310>
- 21 Poisot, T., Canard, E., Mouillot, D., Mouquet, N., & Gravel, D. (2012). The dissimilarity of species inter-  
22 action networks. *Ecology Letters*, 15(12), 1353–1361. <https://doi.org/10.1111/ele.12002>
- 23 Poisot, T., Cirtwill, A. R., Cazelles, K., Gravel, D., Fortin, M.-J., & Stouffer, D. B. (2016). The structure  
24 of probabilistic networks. *Methods in Ecology and Evolution*, 7(3), 303–312. <https://doi.org/10.1111/2041-210X.12468>
- 25
- 26 Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Albery, G. F., Gibb, R. J., Seifert,  
27 S. N., & Carlson, C. J. (2021). *Imputing the mammalian virome with linear filtering and singular value*  
28 *decomposition*. <http://arxiv.org/abs/2105.14973>
- 29 Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks  
30 vary through space and time. *Oikos*, 124(3), 243–251. <https://doi.org/10.1111/oik.01719>
- 31 Raja, N. B. (2022). Colonialism shaped today's biodiversity. *Nature Ecology & Evolution*, 1–2. <https://doi.org/10.1038/s41559-022-01903-y>
- 32
- 33 Ramasamy, D., & Madhow, U. (2015). Compressive spectral embedding: Sidestepping the SVD. In C.

- 1 Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information pro-*  
2 *cessing systems* (Vol. 28). Curran Associates, Inc. [https://proceedings.neurips.cc/paper/2015/file/](https://proceedings.neurips.cc/paper/2015/file/4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf)  
3 [4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf](https://proceedings.neurips.cc/paper/2015/file/4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf)
- 4 Ray, J. C., Grimm, J., & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of  
5 federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068. [https://doi.org/](https://doi.org/10.1139/facets-2020-0075)  
6 [10.1139/facets-2020-0075](https://doi.org/10.1139/facets-2020-0075)
- 7 Runghen, R., Stouffer, D. B., & Dalla Riva, G. V. (2021). *Exploiting node metadata to predict interactions*  
8 *in large networks using graph embedding and neural networks*. [https://doi.org/10.1101/2021.06.10.](https://doi.org/10.1101/2021.06.10.447991)  
9 [447991](https://doi.org/10.1101/2021.06.10.447991)
- 10 Saravia, L. A., Marina, T. I., Kristensen, N. P., De Troch, M., & Momo, F. R. (2021). Ecological network  
11 assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, n/a(n/a).  
12 <https://doi.org/10.1111/1365-2656.13652>
- 13 Schmidt, C., & Garroway, C. J. (2022). Systemic racism alters wildlife genetic diversity. *Proceedings of*  
14 *the National Academy of Sciences*, 119(43), e2102860119. <https://doi.org/10.1073/pnas.2102860119>
- 15 Stevenson, M. T., & Doleac, J. L. (2021). *Algorithmic Risk Assessment in the Hands of Humans* (SSRN Schol-  
16 arly Paper 3489440). <https://doi.org/10.2139/ssrn.3489440>
- 17 Stier, A. C., Samhouri, J. F., Gray, S., Martone, R. G., Mach, M. E., Halpern, B. S., Kappel, C. V., Scar-  
18 borough, C., & Levin, P. S. (2017). Integrating Expert Perceptions into Food Web Conservation and  
19 Management. *Conservation Letters*, 10(1), 67–76. <https://doi.org/10.1111/conl.12245>
- 20 Stouffer, D. B., Camacho, J., Jiang, W., & Nunes Amaral, L. A. (2007). Evidence for the existence of a  
21 robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*,  
22 274(1621), 1931–1940. <https://doi.org/10.1098/rspb.2007.0571>
- 23 Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of  
24 Species' Roles in Food Webs. *Science*, 335(6075), 1489–1492. [https://doi.org/10.1126/science.](https://doi.org/10.1126/science.1216556)  
25 [1216556](https://doi.org/10.1126/science.1216556)
- 26 Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V.,  
27 Mercier, B., Pollock, L. J., Runghen, R., Dalla Riva, G. V., & Poisot, T. (2022). Food web reconstruction  
28 through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution*,  
29 n/a(n/a). <https://doi.org/10.1111/2041-210X.13835>
- 30 Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz,  
31 N. R., Higinio, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap  
32 towards predicting species interaction networks (across space and time). *Philosophical Transactions*  
33 *of the Royal Society B: Biological Sciences*, 376(1837), 20210063. <https://doi.org/10.1098/rstb.2021.>



1 0063

2 Strydom, T., Dalla Riva, G. V., & Poisot, T. (2021). SVD Entropy Reveals the High Complexity of Ecolog-  
3 ical Networks. *Frontiers in Ecology and Evolution*, 9. <https://doi.org/10.3389/fevo.2021.623141>

4 Surendran, S. (2013). Graph Embedding and Dimensionality Reduction - A Survey. *International*  
5 *Journal of Computer Science & Engineering Technology*, 4(1). [https://www.semanticscholar.org/paper/](https://www.semanticscholar.org/paper/Graph-Embedding-and-Dimensionality-Reduction-A-Surendran/3f413d591e4b2b876e033eeb9390e232ad482)

6 [Graph-Embedding-and-Dimensionality-Reduction-A-Surendran/3f413d591e4b2b876e033eeb9390e232ad482](https://www.semanticscholar.org/paper/Graph-Embedding-and-Dimensionality-Reduction-A-Surendran/3f413d591e4b2b876e033eeb9390e232ad482)

7 Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). LINE: Large-scale Information Network  
8 Embedding. *Proceedings of the 24th International Conference on World Wide Web*, 1067–1077. [https:](https://doi.org/10.1145/2736277.2741093)

9 [//doi.org/10.1145/2736277.2741093](https://doi.org/10.1145/2736277.2741093)

10 Thurman, L. L., Barner, A. K., Garcia, T. S., & Chestnut, T. (2019). Testing the link between species  
11 interactions and co-occurrence in a trophic network. *Ecography*, 0. [https://doi.org/10.1111/ecog.](https://doi.org/10.1111/ecog.04360)

12 [04360](https://doi.org/10.1111/ecog.04360)

13 Torres, L., Chan, K. S., & Eliassi-Rad, T. (2020). GLEE: Geometric Laplacian Eigenmap Embedding.  
14 *Journal of Complex Networks*, 8(2), cnaa007. <https://doi.org/10.1093/comnet/cnaa007>

15 Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., Delacruz, J., Dunkerley, G., Fernandez, M., Ferrier,  
16 S., Gill, M., Jones, H., Koen, T., Leys, J., McGeoch, M., Mihoub, J.-B., Scanes, P., Schmeller, D., &

17 Williams, K. (2017). Using the essential biodiversity variables framework to measure biodiversity  
18 change at national scale. *Biological Conservation*, 213, 264–271. [https://doi.org/10.1016/j.biocon.](https://doi.org/10.1016/j.biocon.2016.08.019)

19 [2016.08.019](https://doi.org/10.1016/j.biocon.2016.08.019)

20 Wang, D., Cui, P., & Zhu, W. (2016). Structural Deep Network Embedding. *Proceedings of the 22nd ACM*  
21 *SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1225–1234. [https://doi.org/](https://doi.org/10.1145/2939672.2939753)

22 [10.1145/2939672.2939753](https://doi.org/10.1145/2939672.2939753)

23 Wang, S., Arroyo, J., Vogelstein, J. T., & Priebe, C. E. (2021). Joint Embedding of Graphs. *IEEE Transac-*  
24 *tions on Pattern Analysis and Machine Intelligence*, 43(4), 1324–1336. [https://doi.org/10.1109/TPAMI.](https://doi.org/10.1109/TPAMI.2019.2948619)

25 [2019.2948619](https://doi.org/10.1109/TPAMI.2019.2948619)

26 Wardeh, M., Baylis, M., & Blagrove, M. S. C. (2021). Predicting mammalian hosts in which novel coron-  
27 aviruses can be generated. *Nature Communications*, 12(1, 1), 780. [https://doi.org/10.1038/s41467-](https://doi.org/10.1038/s41467-021-21034-5)

28 [021-21034-5](https://doi.org/10.1038/s41467-021-21034-5)

29 Weiskopf, S. R., Harmáčková, Z. V., Johnson, C. G., Londoño-Murcia, M. C., Miller, B. W., Myers, B. J. E.,  
30 Pereira, L., Arce-Plata, M. I., Blanchard, J. L., Ferrier, S., Fulton, E. A., Harfoot, M., Isbell, F., Johnson,

31 J. A., Mori, A. S., Weng, E., & Rosa, I. M. D. (2022). Increasing the uptake of ecological model results in  
32 policy decisions to improve biodiversity outcomes. *Environmental Modelling & Software*, 149, 105318.

33 <https://doi.org/10.1016/j.envsoft.2022.105318>

- 1 Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, 404(6774),  
2 180–183. <https://doi.org/10.1038/35004572>
- 3 Wood, S. A., Russell, R., Hanson, D., Williams, R. J., & Dunne, J. A. (2015). Effects of spatial scale of  
4 sampling on food web structure. *Ecology and Evolution*, 5(17), 3769–3782. [https://doi.org/10.1002/  
5 ece3.1640](https://doi.org/10.1002/ece3.1640)
- 6 Wu, D., Palmer, D. R., & Deford, D. R. (2021). *Maximum a Posteriori Inference of Random Dot Product Graphs  
7 via Conic Programming* (arXiv:2101.02180). arXiv. <http://arxiv.org/abs/2101.02180>
- 8 Xu, M. (2021). Understanding Graph Embedding Methods and Their Applications. *SIAM Review*, 63(4),  
9 825–853. <https://doi.org/10.1137/20M1386062>
- 10 Yan, S., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A general framework for dimen-  
11 sionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition  
12 (CVPR'05)*, 2, 830–837 vol. 2. <https://doi.org/10.1109/CVPR.2005.170>
- 13 Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In  
14 A. Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.  
15 [https://doi.org/10.1007/978-3-540-77004-6\\_11](https://doi.org/10.1007/978-3-540-77004-6_11)
- 16 Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural net-  
17 works: A review of methods and applications. *AI Open*, 1, 57–81. [https://doi.org/10.1016/j.aiopen.  
18 2021.01.001](https://doi.org/10.1016/j.aiopen.2021.01.001)

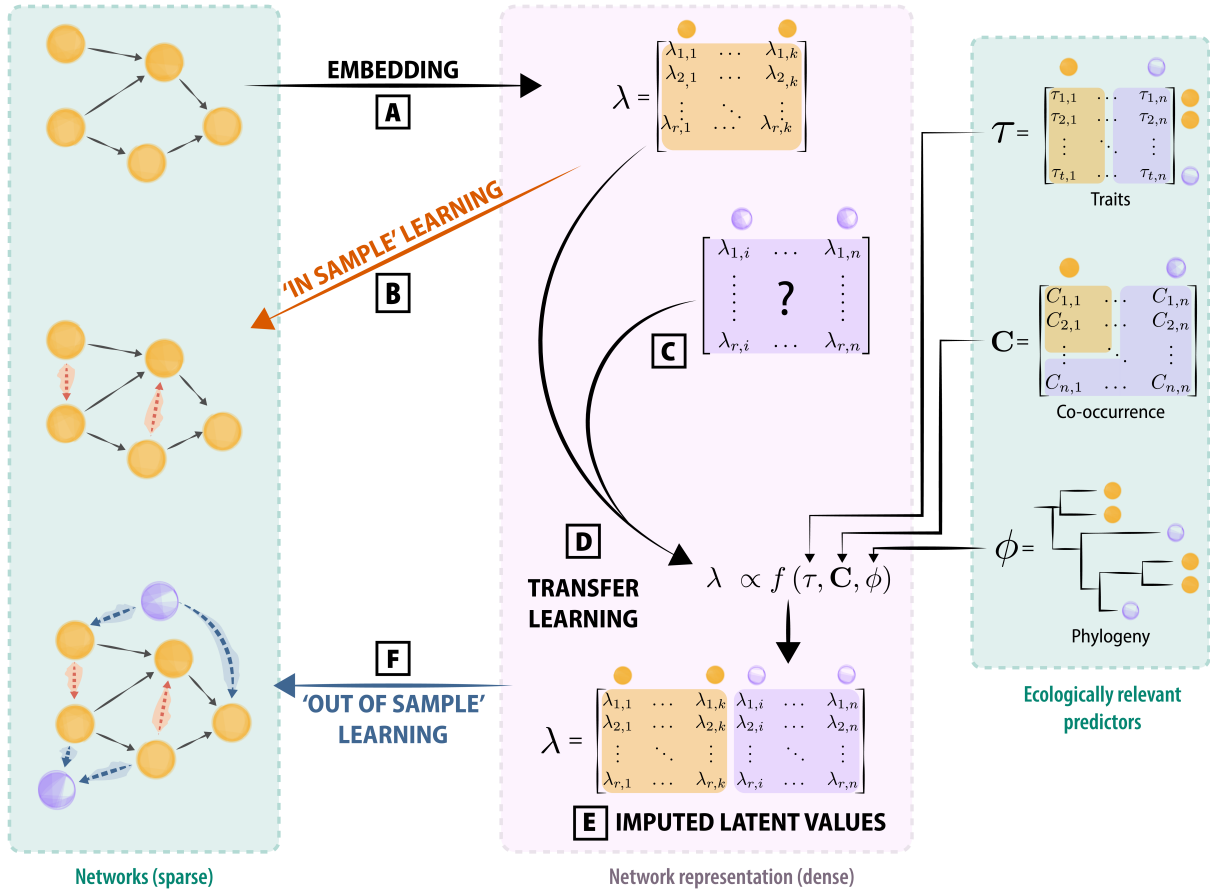


Figure 1: The embedding process (A) can help to identify links (interactions) that may have been missed within the original community (represented by the orange dashed arrows, B). Transfer learning (D) allows for the prediction of links (interactions) even when novel species (C) are included alongside the original community. This is achieved with other ecologically relevant predictors (e.g. traits) in conjunction with the known interactions to infer latent values (E). Ultimately this allows us to predict links (interactions) for species external to the original sample (blue dashed arrows) as well as missing within sample links (F). Within this context the predicted (and original) networks as well as the ecological predictors used (green boxes) are products that can be quantified through measurements in the field, whereas the embedded as well as imputed matrices (purple box) are representative of a decomposition of the interaction matrices onto the embedding space

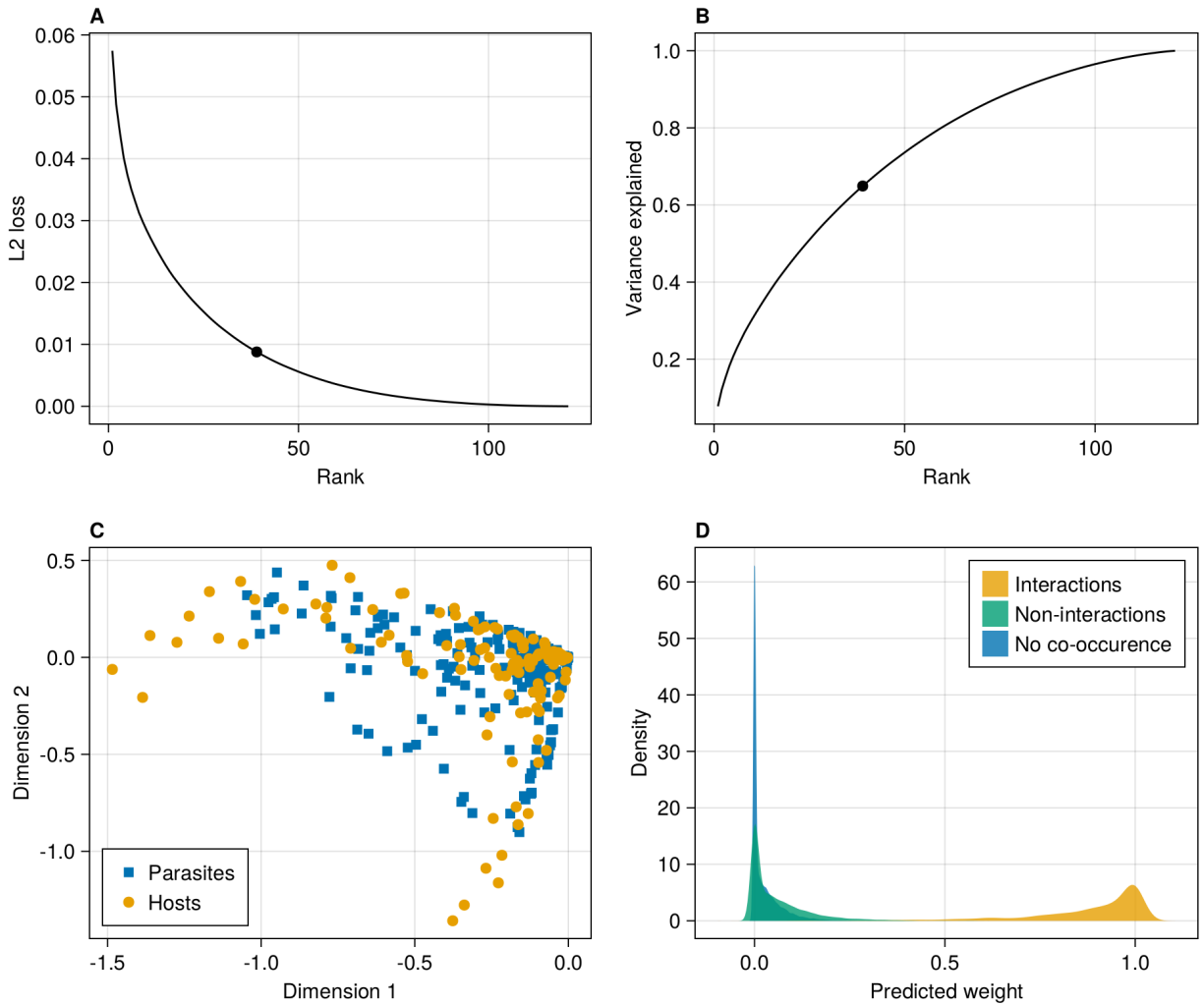


Figure 2: Validation of an embedding for a host-parasite metaweb, using Random Dot Product Graphs. **A**, decrease in approximation error as the number of dimensions in the subspaces increases. **B**, increase in cumulative variance explained as the number of ranks considered increases; in **A** and **B**, the dot represents the point of inflexion in the curve (at rank 39) estimated using the finite differences method. **C**, position of hosts and parasites in the space of latent variables on the first and second dimensions of their respective subspaces (the results have been clamped to the unit interval). **D**, predicted interaction weight from the RDPG based on the status of the species pair in the metaweb. Source: [Demonstration of metaweb embedding using RDPG](#)

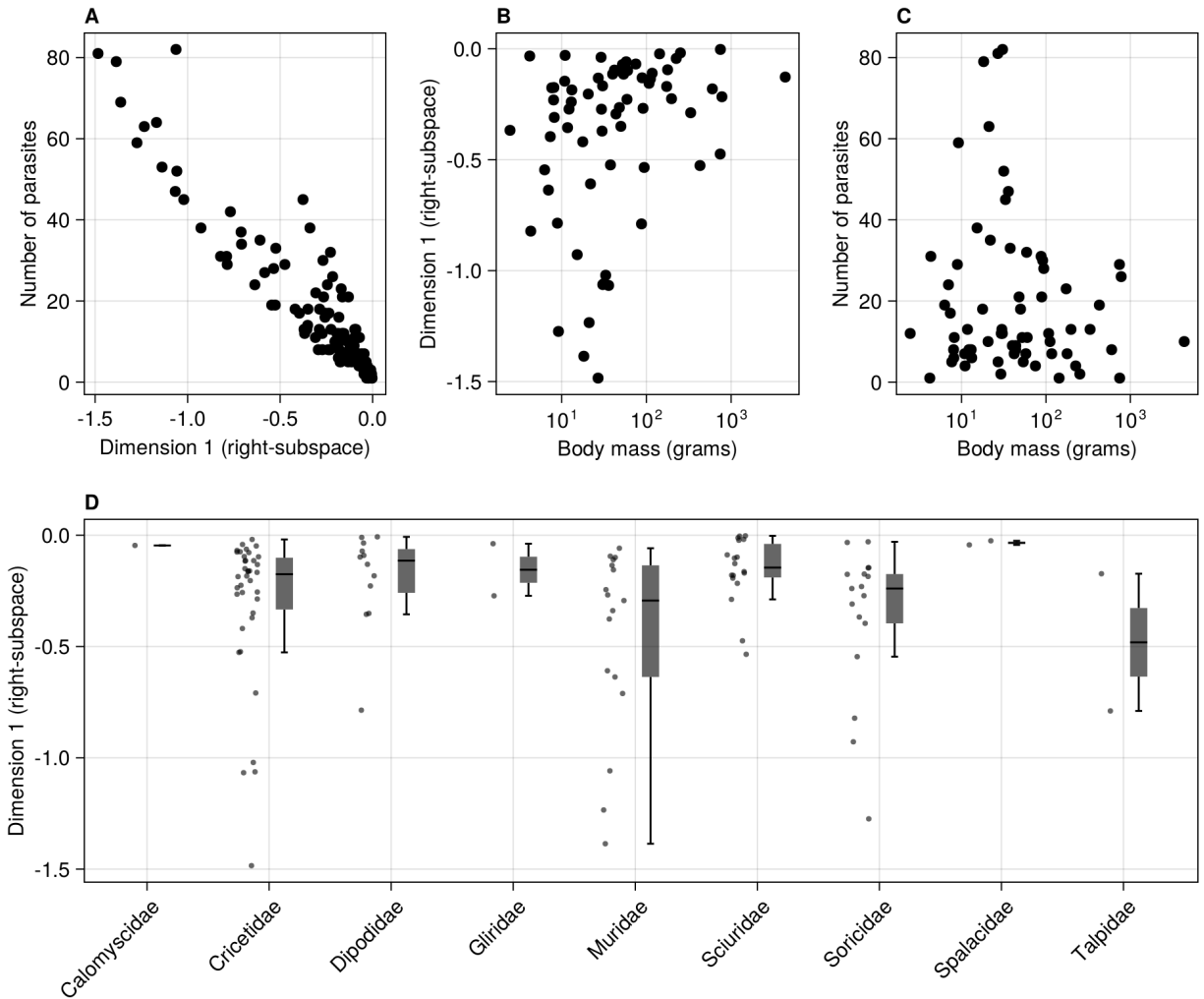


Figure 3: Ecological analysis of an embedding for a host-parasite metaweb, using Random Dot Product Graphs. **A**, relationship between the number of parasites and position along the first axis of the right-subspace for all hosts, showing that the embedding captures elements of network structure at the species scale. **B**, weak relationship between the body mass of hosts (in grams) and the position alongside the same dimension. **C**, weak relationship between body mass of hosts and parasite richness. **D**, distribution of positions alongside the same axis for hosts grouped by taxonomic family. Source: [Demonstration of metaweb embedding using RDPG](#)