

# Graph embedding and transfer learning can help predict 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>2,5</sup>  
Maxwell J Farrell<sup>6</sup> Marie-Josée Fortin<sup>6</sup> Victoria Hemming<sup>7</sup> Benjamin Mercier<sup>2,3</sup> Laura  
J. Pollock<sup>2,5</sup> Rogini Runghen<sup>8</sup> Giulio V. Dalla Riva<sup>9</sup> Timothée Poisot<sup>1,2,‡</sup>

<sup>1</sup> Département de Sciences Biologiques, Université de Montréal, Montréal, Canada <sup>2</sup> Quebec Centre for Biodiversity Science, Montréal, Canada <sup>3</sup> Département de Biologie, Université de Sherbrooke, Sherbrooke, Canada <sup>4</sup> Department of Forest Resources Management, University of British Columbia, Vancouver, B.C., Canada <sup>5</sup> Department of Biology, McGill University, Montréal, Canada <sup>6</sup> Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada <sup>7</sup> Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, Canada <sup>8</sup> Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Canterbury, New Zealand <sup>9</sup> School of Mathematics and Statistics, University of Canterbury, Canterbury, New Zealand

‡ Equal contributions

## Correspondance to:

Timothée Poisot — [timothee.poisot@umontreal.ca](mailto:timothee.poisot@umontreal.ca)

, (networks of potential interactions within a species pool) are a powerful abstraction to understand how large-scale species interaction networks are structured. 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 ‘draft’ metawebs. way to improve our ability to predict metawebs is to maximize available information by using graph embeddings, instead of the list of species interactions. Graph embedding is an emerging field in machine learning that holds great potential for ecological problems. this perspective, we outline how the challenges associated with inferring metawebs line-up with the advantages of graph embeddings; as well as discuss how the choice of the species pool has consequences on the reconstructed network, but also embeds hypotheses about which human-made boundaries are ecologically meaningful.

This work is released by its authors under a CC-BY 4.0 license



Last revision: *January 24, 2023*

1 Being able to infer *potential* interactions could serve as a significant breakthrough in our ability to start  
2 thinking about species interaction networks over large spatial scales (Hortal et al., 2015). Understanding  
3 species interactions holds enormous potential to not only understand and more rapidly learn about  
4 species interactions and metawebs, but also how changes in management of a single species may impact  
5 non-target species. In a recent overview of the field of ecological network prediction, Strydom, Catchen, et  
6 al. (2021) identified two challenges of interest to the prediction of interactions at large scales. First, there  
7 is a relative scarcity of relevant data in most places globally – paradoxically, this restricts our ability to infer  
8 interactions for locations where inference is perhaps the least required (and leaves us unable to make  
9 inference in regions without interaction data); second, accurate predictors are important for accurate  
10 predictions, and the lack of methods that can leverage a small amount of *accurate* data is a serious  
11 impediment to our predictive ability. In most places, our most reliable biodiversity knowledge is that of a  
12 species pool (*i.e.* a set of potentially interacting species in a given area): through the analysis of databases  
13 like the Global Biodiversity Information Facility (GBIF) or the International Union for the Conservation of  
14 Nature (IUCN), it is possible to construct a list of species in a region of interest; but inferring the potential  
15 interactions between these species is difficult.

16 Following the definition of Dunne (2006), a metaweb is the ecological network analogue to the species  
17 pool; specifically, it inventories all *potential* interactions between species for a spatially delimited area (and  
18 so captures the  $\gamma$  diversity of interactions). The metaweb is not a prediction of the network at a specific  
19 point within the spatial area it covers: it will have a different structure, notably by having a larger  
20 connectance (see *e.g.* Wood et al., 2015) and complexity (see *e.g.* Galiana et al., 2022), from any of these  
21 local networks. These local networks (which capture the  $\alpha$  diversity of interactions) are a subset of the  
22 metaweb's species and realized interactions, and have been called “metaweb realizations” (Poisot et al.,  
23 2015). Differences between local networks and their metawebs are due to chance, species abundance and  
24 co-occurrence, local environmental conditions, and local distribution of functional traits, among others.  
25 Specifically, although co-occurrence can be driven by interactions (Cazelles et al., 2016), co-occurrence  
26 alone is not a predictor of interactions (Blanchet et al., 2020; Thurman et al., 2019), and therefore lack of  
27 co-occurrence cannot be used to rule out lack of a feasible interaction. Yet, recent results by Saravia et al.  
28 (2021) strongly suggest that the local (metaweb) realizations only respond weakly to local conditions:  
29 instead, they reflect constraints inherited by the structure of their metaweb. This establishes the metaweb  
30 structure as the core goal of predictive network ecology, as it is a required information to accurately

31 produce downscaled, local predictions.

32 Because the metaweb represents the joint effect of functional, phylogenetic, and macroecological  
33 processes (Morales-Castilla et al., 2015), it holds valuable ecological information. Specifically, it represents  
34 the “upper bounds” on what the composition of the local networks, given a local species pool, can be (see  
35 e.g. McLeod et al., 2021); this information can help evaluate the ability of ecological assemblages to  
36 withstand the effects of, for example, climate change (Fricke et al., 2022). These local networks may be  
37 reconstructed given an appropriate knowledge of local species composition and provide information on  
38 the structure of food webs at finer spatial scales. This has been done for example for tree-galler-parasitoid  
39 systems (Gravel et al., 2018), fish trophic interactions (Albouy et al., 2019), tetrapod trophic interactions  
40 (Braga et al., 2019; O’Connor et al., 2020), and crop-pest networks (Grünig et al., 2020). In this  
41 contribution, we highlight the power in viewing (and constructing) metawebs as *probabilistic* objects in  
42 the context of rare interactions, discuss how a family of machine learning tools (graph embeddings and  
43 transfer learning) can be used to overcome data limitations to metaweb inference, and highlight how the  
44 use of metawebs introduces important questions for the field of network ecology.

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

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

55 Yet the original metaweb definition, and indeed most past uses of metawebs, was based on the  
56 presence/absence of interactions. Moving towards *probabilistic* metawebs, by representing interactions as  
57 Bernoulli events (see e.g. Poisot et al., 2016), offers the opportunity to weigh these rare interactions  
58 appropriately. The inherent plasticity of interactions is important to capture: there have been documented

instances of food webs undergoing rapid collapse/recovery cycles over short periods of time (*e.g.* Pedersen et al., 2017). Furthermore, because the structure of the metaweb cannot be known in advance, it is important to rely on predictive tools that do not assume a specific network topology for link prediction (Gaucher et al., 2021), but are able to work on generalizations of the network. These considerations emphasize why metaweb predictions should focus on quantitative (preferentially probabilistic) predictions, and this should constrain the suite of appropriate models used to predict them.

It is important to recall that a metaweb is intended as a catalogue of all potential (feasible) interactions, which is then filtered for a given application (Morales-Castilla et al., 2015). It is therefore important to separate the interactions that happen “almost surely” (repeated observational data), “almost never” (repeated lack of evidence *or* evidence that the link is forbidden through *e.g.* trait mis-match), and interactions with a probability that lays somewhere in between. In a sense, that most ecological interactions are elusive can call for a slightly different approach to sampling: once the common interactions are documented, the effort required in documenting each rare interaction will increase exponentially (Jordano, 2016). Recent proposals in other fields emphasize the idea that machine learning algorithms can also act as data generators (Hoffmann et al., 2019): high quality observational data can be used to infer core rules underpinning network structure, and be supplemented with synthetic data coming from predictive models trained on them, thereby increasing the volume of information available for analysis. Indeed, Strydom, Catchen, et al. (2021) suggested that knowing the metaweb may render the prediction of local networks easier, because it fixes an “upper bound” on which interactions can exist. In this context, a probabilistic metaweb represents an aggregation of informative priors on the biological feasibility of interactions, which is usually hard to obtain yet has possibly the most potential to boost our predictive ability (Bartomeus et al., 2016).

[Figure 1 about here.]

## **Graph embedding offers promises for the inference of potential interactions**

Graph (or Network) embedding (fig. ??) is a family of machine learning techniques, whose main task is to learn a mapping function from a discrete graph to a continuous domain (Arsov & Mirceva, 2019; Chami et

86 al., 2022). Their main goal is to learn a low dimensional vector representations for the nodes of the graph  
87 (embeddings), such that key properties of the graph (e.g. local or global structures) are retained in the  
88 embedding space (Yan et al., 2005). Ecological networks are an interesting candidate for the widespread  
89 application of embeddings, as they tend to possess a shared structural backbone (see *e.g.* Bramon Mora et  
90 al., 2018 for food webs), which hints at structural invariants in empirical data; assuming that these  
91 structural invariants are indeed widespread, they would dominate the structure of networks, and therefore  
92 be adequately captured by the first (lower) dimensions of an embedding, without the need to measure  
93 derived aspects of their structure (*e.g.* motifs, paths, modularity, ...).

94 Indeed, food webs are inherently low-dimensional objects, and can be adequately represented with less  
95 than ten dimensions (Braga et al., 2019; Eklöf et al., 2013). Simulation results by Botella et al. (2022)  
96 suggest that there is no best method to identify architectural similarities between networks, and that  
97 multiple approaches need to be tested and compared to the network descriptor of interest. This matches  
98 previous, more general results on graph embedding, which suggest that different embedding algorithms  
99 yield different network embeddings (Goyal & Ferrara, 2018), calling for a careful selection of the  
100 problem-specific approach to use. In tbl. ??, we present a selection of common graph and node embedding  
101 methods, alongside examples of their use to predict species interactions; most of these methods rely either  
102 on linear algebra, or on pseudo-random walks on graphs. All forms of embeddings presented in the table  
103 share the common property of summarizing their objects into (sets of) dense feature vectors, that capture  
104 the overall network structure, pairwise information on nodes, and emergent aspects of the network, in a  
105 compressed way (*i.e.* with some information loss, as we later discuss in the illustration). Node embeddings  
106 tend to focus on maintaining pairwise relationships, while graph embeddings focus on maintaining the  
107 network structure. Nevertheless, some graph embedding techniques (like RDPG, see *e.g.* Wu et al., 2021)  
108 will provide usable node-level embeddings.

109 One prominent family of approaches we do not discuss in the present manuscript is Graph Neural  
110 Networks [GNN; Zhou et al. (2020)]. GNN are, in a sense, a method to embed a graph into a dense  
111 subspace, but belong to the family of deep learning methods, which has its own set of practices (see *e.g.*  
112 Goodfellow et al., 2016). An important issue with methods based on deep learning is that, because their  
113 parameter space is immense, the sample size of the data fed into them must be similarly large (typically  
114 thousands of instances). This is a requirement for the model to converge correctly during training, but this  
115 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 tbl. ???. Assuming that the uptake of next-generation biomonitoring techniques does indeed deliver larger datasets on species interactions (Bohan et al., 2017), there is a potential for GNN to become an applicable embedding/predictive technique in the coming years.

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 a lot of 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 tSNE 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
tSNE	nodes	statistical divergence	Hinton & Roweis (2002)	Gibb et al. (2021); Cieslak et al. (2020) <sup>a</sup>
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) <sup>b</sup>

Method	Object	Technique	Reference	Application
graph2vec	sub-graph	skipgram network	Narayanan et al. (2017)	
RDPG	graph	SVD	Young & Scheinerman (2007)	Poisot et al. (2021); Dalla Riva & Stouffer (2016)
GLEE	graph	Laplacian eigenmap	Torres et al. (2020)	
DeepWalk	graph	stochastic gradient descent	Perozzi et al. (2014)	Wardeh et al. (2021)
GraphKKE	graph	stochastic differential equation	Melnyk et al. (2020)	Melnyk et al. (2020) <sup>a</sup>
FastEmbed	graph	eigen decomposition	Ramasamy & Madhow (2015)	
PCA	graph	eigen decomposition	S & Surendran (2013)	Strydom, Catchen, et al. (2021)
Joint methods	multiple graphs	multiple strategies	S. Wang et al. (2021)	

The popularity of graph embedding techniques in machine learning is more than the search for structural invariants: graphs are discrete objects, and machine learning techniques tend to handle continuous data better. Bringing a sparse graph into a continuous, dense vector space (Xu, 2021) opens up a broader variety of predictive algorithms, notably of the sort that are able to predict events as probabilities (Murphy, 2022). Furthermore, the projection of the graph itself is a representation that can be learned; Runghen et al. (2021), for example, used a neural network to learn the embedding of a network in which not all interactions were known, based on the nodes' metadata. This example has many parallels in ecology (see fig. ?? C), in which node metadata can be represented by phylogeny, abundance, or functional traits. Using phylogeny as a source of information assumes (or strives to capture) the action of evolutionary processes on network structure, which at least for food webs have been well documented (Dalla Riva & Stouffer, 2016; Stouffer et al., 2012; **Stouffer2007EviExi?**); similarly, the use of functional traits assumes



133 that interactions can be inferred from the knowledge of trait-matching rules, which is similarly well  
 134 supported in the empirical literature (Bartomeus, 2013; Bartomeus et al., 2016; **Gravel2013InfFoo?**).  
 135 Relating this information to an embedding rather than a list of networks measures would allow to capture  
 136 their effect on the more fundamental aspects of network structure; conversely, the absence of a  
 137 phylogenetic or functional signal may suggest that evolutionary/trait processes are not strong drivers of  
 138 network structure, therefore opening a new way to perform hypothesis testing.

139 Before moving further, it is important to clarify the epistemic status of node values derived from  
 140 embeddings: specifically, they are *not* functional traits, and therefore should not be discussed in terms of  
 141 effects or responses. As per the framework of Malaterre et al. (2019), these values neither derive from, nor  
 142 result in, changes in organismal fitness, and should therefore not be used to quantify *e.g.* functional  
 143 diversity. This holds true even when there are correlations between latent values and functional traits:  
 144 although these enable an ecological discussion of how traits condition the structure of the network, the  
 145 existence of a statistical relationship does not elevate the latent values to the status of functional traits.

146 Rather than directly predicting biological rules (see *e.g.* Pichler et al., 2020 for an overview), which may be  
 147 confounded by the sparse nature of graph data, learning embeddings works in the low-dimensional space  
 148 that maximizes information about the network structure. This approach is further justified by the  
 149 observation, for example, that the macro-evolutionary history of a network is adequately represented by  
 150 some graph embeddings [Random dot product graphs (RDPG); see Dalla Riva & Stouffer (2016)]. In a  
 151 recent publication, Strydom et al. (2022) have used an embedding (based on RDPG) to project a metaweb  
 152 of trophic interactions between European mammals, and transferred this information to mammals of  
 153 Canada, using the phylogenetic distance between related clades to infer the values in the latent sub-space  
 154 into which the European metaweb was projected. By performing the RDPG step on re-constructed values,  
 155 this approach yields a probabilistic trophic metaweb for mammals of Canada based on knowledge of  
 156 European species, despite a limited ( $\approx 5\%$ ) taxonomic overlap.

157 Graph embeddings *can* serve as a dimensionality reduction method. For example, RDPG (Strydom et al.,  
 158 2022) and t-SVD [truncated Singular Value Decomposition; Poisot et al. (2021)] typically embed networks  
 159 using fewer dimensions than the original network [the original network has as many dimensions as  
 160 species, and as many informative dimensions as trophically unique species; Strydom, Dalla Riva, et al.  
 161 (2021)]. But this is not necessarily the case – indeed, one may perform a PCA (a special case of SVD) to  
 162 project the raw data into a subspace that improves the efficacy of t-SNE [t-distributed stochastic neighbor

embedding; Maaten (2009)]. There are many dimensionality reductions (Anowar et al., 2021) that can be applied to an embedded network should the need for dimensionality reduction (for example for data visualisation) arise. In brief, many graph embeddings *can* serve as dimensionality reduction steps, but not all do, neither do all dimensionality reduction methods provide adequate graph embedding capacities. In the next section (and fig. ??), we show how the amount of dimensionality reduction can affect the quality of the embedding.

## **An illustration of metaweb embedding**

In this section, we illustrate the embedding of a collection of bipartite networks collected by Hadfield et al. (2014), using t-SVD and RDPG (see Strydom et al., 2022 for the full details). Briefly, an RDPG decomposes a network into two subspaces (left and right), which are matrices that when multiplied give an approximation of the original network. RDPG has the particularly desirable properties of being a graph embedding technique that produces relevant node-level feature vectors, and provides good approximations of graphs with varied structures (Athreya et al., 2017). The code to reproduce this example is available as supplementary material (note, for the sake of comparison, that Strydom, Catchen, et al., 2021 have an example using embedding through PCA followed by prediction using a deep neural network on the same dataset). The resulting (binary) metaweb  $\mathcal{M}$  has 2131 interactions between 206 parasites and 121 hosts, and its adjacency matrix has full rank (*i.e.* it represents a space with 121 dimensions). All analyses were done using Julia (Bezanson et al., 2017) version 1.7.2, *Makie.jl* (Danisch & Krumbiegel, 2021), and *EcologicalNetworks.jl* (Poisot et al., 2019).

[Figure 2 about here.]

In fig. ??, we focus on some statistical checks of the embedding. In panel **A**, we show that the averaged  $L_2$  loss (*i.e.* the sum of squared errors) between the empirical and reconstructed metaweb decreases when the number of dimensions (rank) of the subspace increases, with an inflection at 39 dimensions (out of 120 initially) according to the finite differences method. As discussed by Runghen et al. (2021), there is often a trade-off between the number of dimensions to use (more dimensions are more computationally demanding) and the quality of the representation. In panel **B**, we show the increase in cumulative variance explained at each rank, and visualize that using 39 ranks explains about 70% of the variance in

the empirical metaweb. This is a different information from the  $L_2$  loss (which is averaged across interactions), as it works on the eigenvalues of the embedding, and therefore captures higher-level features of the network. In panel **C**, we show positions of hosts and parasites on the first two dimensions of the left and right subspaces. Note that these values largely skew negative, because the first dimensions capture the coarse structure of the network: most pairs of species do not interact, and therefore have negative values. Finally in panel **D**, we show the predicted weight (*i.e.* the result of the multiplication of the RDGP subspaces at a rank of 25) as a function of whether the interactions are observed, not-observed, or unknown due to lack of co-occurrence. This reveals that the observed interactions have higher predicted weights, although there is some overlap; the usual approach to identify potential interactions based on this information would be a thresholding analysis, which is outside the scope of this manuscript (and is done in the papers cited in this illustration). Because the values returned from RDGP are not bound to the unit interval, we performed a clamping of the weights to the unit space, showing a one-inflation in documented interaction, and a zero-inflation in other species pairs. This last figure crosses from the statistical into the ecological, by showing that species pairs with no documented co-occurrence have weights that are not distinguishable from species pairs with no documented interactions, suggesting that (as befits a host-parasite model) the ability to interact is a strong predictor of co-occurrence.

[Figure 3 about here.]

The results of fig. ?? show that we can extract an embedding of the metaweb that captures enough variance to be relevant; specifically, this is true both of  $L_2$  loss (indicating that RDGP is able to capture pairwise processes) and the cumulative variance explained (indicating that RDGP is able to capture network-level structure). Therefore, in fig. ??, we relate the values of latent variables for hosts to different ecologically-relevant data. In panel **A**, we show that host with a higher value on the first dimension have fewer parasites. This relates to the body size of hosts in the *PanTHERIA* database (Jones et al., 2009), as shown in panel **B**: interestingly, the position on the first axis is only weakly correlated to body mass of the host; this matches well established results showing that body size/mass is not always a direct predictor of parasite richness in terrestrial mammals (Morand & Poulin, 1998), a result we observe in panel **C**. Finally, in panel **D**, we can see how different taxonomic families occupy different positions on the first axis, with *e.g.* Sciuridae being biased towards higher values. These results show how we can look for ecological informations in the output of network embeddings, which can further be refined into the selection of

219 predictors for transfer learning.

## 220 **The metaweb embeds both ecological hypotheses and practices**

221 The goal of metaweb inference is to provide information about the interactions between species at a large  
222 spatial scale. But as Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide  
223 problems”; any inference of a metaweb at large scales must contend with several novel, and interwoven,  
224 families of problems. In this section, we list some of the most pressing research priorities (*i.e.* problems  
225 that can be addressed with subsequent data analysis or simulations), as well as issues related to the  
226 application of these methods at the science-policy interface.

227 The first open research problem is the taxonomic and spatial limit of the metaweb to embed and transfer.  
228 If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the chances of  
229 finding another area with enough related species (through phylogenetic relatedness or similarity of  
230 functional traits) to make a reliable inference decreases; this would likely be indicated by large confidence  
231 intervals during estimation of the values in the low-rank space, meaning that the representation of the  
232 original graph is difficult to transfer to the new problem. Alternatively, if the initial metaweb is too large  
233 (taxonomically), then the resulting embeddings would need to represent interactions between taxonomic  
234 groups that are not present in the new location. This would lead to a much higher variance in the starting  
235 dataset, and to under-dispersion in the target dataset, resulting in the potential under or over estimation of  
236 the strength of new predicted interactions. The lack of well documented metawebs is currently preventing  
237 the development of more concrete guidelines. The question of phylogenetic relatedness and distribution is  
238 notably relevant if the metaweb is assembled in an area with mostly endemic species (*e.g.* a system that  
239 has undergone recent radiation or that has remained in isolation for a long period of time might not have  
240 an analogous system with which to draw knowledge from), and as with every predictive algorithm, there  
241 is room for the application of our best ecological judgement. Because this problem relates to distribution  
242 of species in the geographic or phylogenetic space, it can certainly be approached through assessing the  
243 performance of embedding transfer in simulated starting/target species pools.

244 The second series of problems relate to determining which area should be used to infer the new metaweb  
245 in, as this determines the species pool that must be used. Metawebs can be constructed by assigning  
246 interactions in a list of species within geographic boundaries. The upside of this approach is that

247 information at the country level is likely to be required for biodiversity assessments, as countries set  
248 conservation goals at the national level (Buxton et al., 2021), and as quantitative instruments are designed  
249 to work at these scales (Turak et al., 2017); specific strategies are often enacted at smaller scales, nested  
250 within a specific country (Ray et al., 2021). But there is no guarantee that these boundaries are  
251 meaningful. In fact, we do not have a satisfying answer to the question of “where does a food web stop?”;  
252 the most promising solutions involve examining the spatial consistency of network area relationships  
253 (Fortin et al., 2021; see e.g. Galiana et al., 2018, 2019, 2021), which is impossible in the absence of enough  
254 information about the network itself. This suggests that inferred metawebs should be further downscaled  
255 to allow for *a posteriori* analyses. The methodology for metaweb downscaling is currently limited, and it is  
256 likely that the sustained effort to characterize the spatial dependency of food web structure will lead to  
257 more prescriptive guidelines about the need for prediction downscaling. Recent results by Martins et al.  
258 (2022) suggest that networks are shaped within eco-regions, with abrupt structural transitions from an  
259 eco-region to the next. Should this trend hold generally, this would provide an ecologically-relevant scale  
260 at which metawebs can be downscaled and predicted.

261 The final family of problems relates less to ecological methods than to the praxis of ecological research.  
262 Operating under the context of national divisions, in large parts of the world, reflects nothing more than  
263 the legacy of settler colonialism, which drives a disparity in available ecological data. Applying any  
264 embedding to biased data does not debias them, but instead embeds these very same biases, propagating  
265 them to the machine learning models using embeddings to make predictions. Indeed, the use of ecological  
266 data is not an apolitical act (Nost & Goldstein, 2021), as data infrastructures tend to be designed to answer  
267 questions within national boundaries (therefore placing contingencies on what is available to be  
268 embedded), and their use often draws upon and reinforces territorial statecraft. As per Machen & Nost  
269 (2021), this is particularly true when the output of “algorithmic thinking” (e.g. relying on machine  
270 learning to generate knowledge) can be re-used for governance (e.g. enacting conservation decisions at the  
271 national scale). As information on species interaction networks structure is increasingly leveraged as a  
272 tool to guide conservation actions [see e.g. recent discussions for food-web based conservation; Eero et al.  
273 (2021); Naman et al. (2022); Stier et al. (2017)], the need to appraise and correct biases that are unwittingly  
274 propagated to algorithms when embedded from the original data is paramount. These considerations are  
275 even more urgent in the specific context of biodiversity data, as long-term colonial legacies still shape  
276 taxonomic composition to this day (Lenzner et al., 2022; Raja, 2022), and where much shorter-term

277 changes in taxonomic and genetic richness emerged as a consequence of environmental racism. (Schmidt  
278 & Garroway, 2022).

279 Predictive approaches, regardless of the scale at which they are deployed and the intent of their  
280 deployment, originate in the framework that contributed to the ongoing biodiversity crisis (Adam, 2014)  
281 and reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020). Particularly on  
282 Turtle Island and other territories that were traditionally stewarded by Indigenous people, these  
283 approaches should be replaced (or at least guided and framed) by Indigenous principles of land  
284 management (Eichhorn et al., 2019; No'kmaq et al., 2021), as part of an “algorithm-in-the-loop” approach.  
285 Human-algorithm interactions are notoriously difficult and can yield adverse effect (Green & Chen, 2019;  
286 Stevenson & Doleac, 2021), suggesting the need to systematically study them for the specific purpose of  
287 biodiversity governance, as well as to improve the algorithmic literacy of decision makers. As we see  
288 artificial intelligence/machine learning being increasingly mobilized to generate knowledge that is  
289 lacking for conservation decisions (*e.g.* Lamba et al., 2019; Mosebo Fernandes et al., 2020) and drive policy  
290 decisions (Weiskopf et al., 2022), our discussion of these tools need to go beyond the technical and  
291 statistical, and into the governance consequences they can have.

292 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional  
293 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and  
294 Omàmiwininiwak nations. TP, TS, DC, and LP received funding from the Canadian Institute for Ecology  
295 & Evolution. FB is funded by the Institute for Data Valorization (IVADO). TS, SB, and TP are funded by a  
296 donation from the Courtois Foundation. CB was awarded a Mitacs Elevate Fellowship no. IT12391, in  
297 partnership with fRI Research, and also acknowledges funding from Alberta Innovates and the Forest  
298 Resources Improvement Association of Alberta. M-JF acknowledges funding from NSERC Discovery  
299 Grant and NSERC CRC. RR is funded by New Zealand’s Biological Heritage Ngā Koiora Tuku Iho  
300 National Science Challenge, administered by New Zealand Ministry of Business, Innovation, and  
301 Employment. BM is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the  
302 FRQNT master’s scholarship. LP acknowledges funding from NSERC Discovery Grant (NSERC  
303 RGPIN-2019-05771). TP acknowledges financial support from the Fondation Courtois, and NSERC  
304 through the Discovery Grants and Discovery Accelerator Supplement programs. MJF is supported by an  
305 NSERC PDF and an RBC Post-Doctoral Fellowship.

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

307 **Authors' contributions:** TS, and TP conceived the ideas discussed in the manuscript. All authors  
308 contributed to writing and editing the manuscript.

309 **Data availability:** There is no data associated with this manuscript.

## 310 References

311 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.

312 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,  
313 Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel, D.  
314 (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8), 1153–1161.  
315 <https://doi.org/10.1038/s41559-019-0950-y>

316 Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality  
317 reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer*  
318 *Science Review*, 40, 100378. <https://doi.org/10.1016/j.cosrev.2021.100378>

319 Arsov, N., & Mirceva, G. (2019). Network Embedding: An Overview. *arXiv:1911.11726 [Cs, Stat]*.  
320 <https://arxiv.org/abs/1911.11726>

321 Athreya, A., Fishkind, D. E., Levin, K., Lyzinski, V., Park, Y., Qin, Y., Sussman, D. L., Tang, M., Vogelstein,  
322 J. T., & Priebe, C. E. (2017). *Statistical inference on random dot product graphs: A survey* (No.  
323 arXiv:1709.05454). arXiv. <https://arxiv.org/abs/1709.05454>

324 Bartomeus, I. (2013). Understanding linkage rules in plant-pollinator networks by using hierarchical  
325 models that incorporate pollinator detectability and plant traits. *PloS One*, 8(7), e69200.

326 Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A  
327 common framework for identifying linkage rules across different types of interactions. *Functional*  
328 *Ecology*, 30(12), 1894–1903.

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

331 Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions.  
332 *Ecology Letters*.

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

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

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

343 Bramon Mora, B., Gravel, D., Gilarranz, L. J., Poisot, T., & Stouffer, D. B. (2018). Identifying a common  
 344 backbone of interactions underlying food webs from different ecosystems. *Nature Communications*,  
 345 9(1), 2603. <https://doi.org/10.1038/s41467-018-05056-0>

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

351 Cazelles, K., Araújo, M. B., Mouquet, N., & Gravel, D. (2016). A theory for species co-occurrence in  
 352 interaction networks. *Theoretical Ecology*, 9(1), 39–48.  
 353 <https://doi.org/10.1007/s12080-015-0281-9>

354 Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., & Murphy, K. (2022). Machine Learning on Graphs: A  
 355 Model and Comprehensive Taxonomy. *Journal of Machine Learning Research*, 23(89), 1–64.

356 Chen, D., Xue, Y., Fink, D., Chen, S., & Gomes, C. P. (2017). *Deep Multi-species Embedding*. 3639–3646.

357 Chen, H., Perozzi, B., Hu, Y., & Skiena, S. (2017). HARP: Hierarchical Representation Learning for  
 358 Networks. *arXiv:1706.07845 [Cs]*. <https://arxiv.org/abs/1706.07845>

359 Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity,  
 360 colonialism and conquest in an era of capitalist globalization. In *NGOization: Complicity*,



361 *contradictions and prospects* (pp. 24–44). Bloomsbury Publishing.

362 Cieslak, M. C., Castelfranco, A. M., Roncalli, V., Lenz, P. H., & Hartline, D. K. (2020). T-Distributed  
 363 Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Marine*  
 364 *Genomics*, 51, 100723. <https://doi.org/10.1016/j.margen.2019.100723>

365 Csermely, P. (2004). Strong links are important, but weak links stabilize them. *Trends in Biochemical*  
 366 *Sciences*, 29(7), 331–334. <https://doi.org/10.1016/j.tibs.2004.05.004>

367 Dalla Riva, G. V., & Stouffer, D. B. (2016). Exploring the evolutionary signature of food webs' backbones  
 368 using functional traits. *Oikos*, 125(4), 446–456. <https://doi.org/10.1111/oik.02305>

369 Dallas, T., Park, A. W., & Drake, J. M. (2017). Predicting cryptic links in host-parasite networks. *PLOS*  
 370 *Computational Biology*, 13(5), e1005557. <https://doi.org/10.1371/journal.pcbi.1005557>

371 Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia.  
 372 *Journal of Open Source Software*, 6(65), 3349. <https://doi.org/10.21105/joss.03349>

373 Domínguez, L., & Luoma, C. (2020). Decolonising Conservation Policy: How Colonial Land and  
 374 Conservation Ideologies Persist and Perpetuate Indigenous Injustices at the Expense of the  
 375 Environment. *Land*, 9(3), 65. <https://doi.org/10.3390/land9030065>

376 Dunne, J. A. (2006). The Network Structure of Food Webs. In J. A. Dunne & M. Pascual (Eds.), *Ecological*  
 377 *networks: Linking structure and dynamics* (pp. 27–86). Oxford University Press.

378 Eero, M., Dierking, J., Humborg, C., Undeman, E., MacKenzie, B. R., Ojaveer, H., Salo, T., & Köster, F. W.  
 379 (2021). Use of food web knowledge in environmental conservation and management of living  
 380 resources in the Baltic Sea. *ICES Journal of Marine Science*, 78(8), 2645–2663.  
 381 <https://doi.org/10.1093/icesjms/fsab145>

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

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

388 Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes.

389 *Proceedings of the Royal Society B: Biological Sciences*, 288(1949), rspb.2020.1889, 20201889.  
 390 <https://doi.org/10.1098/rspb.2020.1889>

391 Fricke, E. C., Ordonez, A., Rogers, H. S., & Svenning, J.-C. (2022). The effects of defaunation on plants'  
 392 capacity to track climate change. *Science*.

393 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.  
 394 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,  
 395 n/a(n/a). <https://doi.org/10.1111/ecog.05229>

396 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is  
 397 scale dependent: Understanding the biotic specialization of hostparasitoid networks. *Ecography*, 42(6),  
 398 1175–1187. <https://doi.org/10.1111/ecog.03684>

399 Galiana, N., Lurgi, M., Bastazini, V. A. G., Bosch, J., Cagnolo, L., Cazelles, K., Claramunt-López, B., Emer,  
 400 C., Fortin, M.-J., Grass, I., Hernández-Castellano, C., Jauker, F., Leroux, S. J., McCann, K., McLeod, A.  
 401 M., Montoya, D., Mulder, C., Osorio-Canadas, S., Reverté, S., ... Montoya, J. M. (2022). Ecological  
 402 network complexity scales with area. *Nature Ecology & Evolution*, 1–8.  
 403 <https://doi.org/10.1038/s41559-021-01644-4>

404 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,  
 405 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),  
 406 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

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

410 Gibb, R., Albery, G. F., Becker, D. J., Brierley, L., Connor, R., Dallas, T. A., Eskew, E. A., Farrell, M. J.,  
 411 Rasmussen, A. L., Ryan, S. J., Sweeny, A., Carlson, C. J., & Poisot, T. (2021). Data Proliferation,  
 412 Reconciliation, and Synthesis in Viral Ecology. *BioScience*, 71(11), 1148–1156.  
 413 <https://doi.org/10.1093/biosci/biab080>

414 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

415 Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey.  
 416 *Knowledge-Based Systems*, 151, 78–94. <https://doi.org/10.1016/j.knosys.2018.03.022>

417 Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B.,  
 418 Tylianakis, J. M., Wood, S. A., & Roslin, T. (2018). Bringing Elton and Grinnell together: A quantitative  
 419 framework to represent the biogeography of ecological interaction networks. *Ecography*, 0(0).  
 420 <https://doi.org/10.1111/ecog.04006>

421 Green, B., & Chen, Y. (2019). Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in  
 422 Risk Assessments. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 90–99.  
 423 <https://doi.org/10.1145/3287560.3287563>

424 Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*  
 425 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.  
 426 <https://doi.org/10.1145/2939672.2939754>

427 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs  
 428 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,  
 429 3(1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

430 Gupta, A., Matta, P., & Pant, B. (2021). Graph neural network: Current state of Art, challenges and  
 431 applications. *Materials Today: Proceedings*, 46, 10927–10932.  
 432 <https://doi.org/10.1016/j.matpr.2021.01.950>

433 Hadfield, J. D., Krasnov, B. R., Poulin, R., & Nakagawa, S. (2014). A Tale of Two Phylogenies: Comparative  
 434 Analyses of Ecological Interactions. *The American Naturalist*, 183(2), 174–187.  
 435 <https://doi.org/10.1086/674445>

436 Herbert, F. (1965). *Dune* (First). Chilton Book Company.

437 Hinton, G., & Roweis, S. T. (2002). Stochastic neighbor embedding. *NIPS*, 15, 833–840.

438 Hoffmann, J., Bar-Sinai, Y., Lee, L. M., Andrejevic, J., Mishra, S., Rubinstein, S. M., & Rycroft, C. H. (2019).  
 439 Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers  
 440 order in crumpled sheets. *Science Advances*, 5(4), eaau6792.  
 441 <https://doi.org/10.1126/sciadv.aau6792>

442 Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven  
 443 Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and*  
 444 *Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>

445 Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E.  
 446 H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S.  
 447 A., Rigby, E. A., Rist, J., ... Purvis, A. (2009). PanTHERIA: A species-level database of life history,  
 448 ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184.  
 449 *Ecology*, 90(9), 2648–2648. <https://doi.org/10.1890/08-1494.1>

450 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30(12), 1883–1893.  
 451 <https://doi.org/10.1111/1365-2435.12763>

452 Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation.  
 453 *Current Biology*, 29(19), R977–R982. <https://doi.org/10.1016/j.cub.2019.08.016>

454 Lenzner, B., Latombe, G., Schertler, A., Seebens, H., Yang, Q., Winter, M., Weigelt, P., van Kleunen, M.,  
 455 Pyšek, P., Pergl, J., Kreft, H., Dawson, W., Dullinger, S., & Essl, F. (2022). Naturalized alien floras still  
 456 carry the legacy of European colonialism. *Nature Ecology & Evolution*, 1–10.  
 457 <https://doi.org/10.1038/s41559-022-01865-1>

458 Maaten, L. van der. (2009). Learning a Parametric Embedding by Preserving Local Structure. *Proceedings*  
 459 *of the Twelfth International Conference on Artificial Intelligence and Statistics*, 384–391.

460 Machen, R., & Nost, E. (2021). Thinking algorithmically: The making of hegemonic knowledge in climate  
 461 governance. *Transactions of the Institute of British Geographers*, 46(3), 555–569.  
 462 <https://doi.org/10.1111/tran.12441>

463 Malaterre, C., Dussault, A. C., Mermans, E., Barker, G., Beisner, B. E., Bouchard, F., Desjardins, E., Handa,  
 464 I. T., Kembel, S. W., Lajoie, G., Maris, V., Munson, A. D., Odenbaugh, J., Poisot, T., Shapiro, B. J., &  
 465 Suttle, C. A. (2019). Functional Diversity: An Epistemic Roadmap. *BioScience*, 69(10), 800–811.  
 466 <https://doi.org/10.1093/biosci/biz089>

467 Martins, L. P., Stouffer, D. B., Blendinger, P. G., Böhning-Gaese, K., Buitrón-Jurado, G., Correia, M., Costa,  
 468 J. M., Dehling, D. M., Donatti, C. I., Emer, C., Galetti, M., Heleno, R., Jordano, P., Menezes, Í.,  
 469 Morante-Filho, J. C., Muñoz, M. C., Neuschulz, E. L., Pizo, M. A., Quitián, M., ... Tylianakis, J. M.  
 470 (2022). Global and regional ecological boundaries explain abrupt spatial discontinuities in avian  
 471 frugivory interactions. *Nature Communications*, 13(1), 6943.  
 472 <https://doi.org/10.1038/s41467-022-34355-w>

473 McLeod, A., Leroux, S. J., Gravel, D., Chu, C., Cirtwill, A. R., Fortin, M.-J., Galiana, N., Poisot, T., & Wood,

474 S. A. (2021). Sampling and asymptotic network properties of spatial multi-trophic networks. *Oikos*,  
475 *n/a(n/a)*. <https://doi.org/10.1111/oik.08650>

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

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

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

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

487 Murphy, K. P. (2022). *Probabilistic machine learning: An introduction*. MIT Press.

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

491 Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., & Jaiswal, S. (2017). Graph2vec:  
492 Learning Distributed Representations of Graphs. *arXiv:1707.05005 [Cs]*.  
493 <https://arxiv.org/abs/1707.05005>

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

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

499 Nost, E., & Goldstein, J. E. (2021). A political ecology of data. *Environment and Planning E: Nature and*  
500 *Space*, 25148486211043503. <https://doi.org/10.1177/25148486211043503>

501 O'Connor, L. M. J., Pollock, L. J., Braga, J., Ficetola, G. F., Maiorano, L., Martinez-Almoyna, C.,

502 Montemaggiori, A., Ohlmann, M., & Thuiller, W. (2020). Unveiling the food webs of tetrapods across  
 503 Europe through the prism of the Eltonian niche. *Journal of Biogeography*, 47(1), 181–192.  
 504 <https://doi.org/10.1111/jbi.13773>

505 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,  
 506 Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the  
 507 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*, 4(7),  
 508 170215. <https://doi.org/10.1098/rsos.170215>

509 Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations.  
 510 *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data*  
 511 *Mining*, 701–710. <https://doi.org/10.1145/2623330.2623732>

512 Pichler, M., Boreux, V., Klein, A.-M., Schleuning, M., & Hartig, F. (2020). Machine learning algorithms to  
 513 infer trait-matching and predict species interactions in ecological networks. *Methods in Ecology and*  
 514 *Evolution*, 11(2), 281–293. <https://doi.org/10.1111/2041-210X.13329>

515 Poisot, T., Belisle, Z., Hoebeke, L., Stock, M., & Szefer, P. (2019). EcologicalNetworks.jl - analysing  
 516 ecological networks. *Ecography*. <https://doi.org/10.1111/ecog.04310>

517 Poisot, T., Cirtwill, A. R., Cazelles, K., Gravel, D., Fortin, M.-J., & Stouffer, D. B. (2016). The structure of  
 518 probabilistic networks. *Methods in Ecology and Evolution*, 7(3), 303–312.  
 519 <https://doi.org/10.1111/2041-210X.12468>

520 Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Albery, G. F., Gibb, R. J., Seifert, S. N.,  
 521 & Carlson, C. J. (2021). Imputing the mammalian virome with linear filtering and singular value  
 522 decomposition. *arXiv:2105.14973 [q-Bio]*. <https://arxiv.org/abs/2105.14973>

523 Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary  
 524 through space and time. *Oikos*, 124(3), 243–251. <https://doi.org/10.1111/oik.01719>

525 Raja, N. B. (2022). Colonialism shaped today’s biodiversity. *Nature Ecology & Evolution*, 1–2.  
 526 <https://doi.org/10.1038/s41559-022-01903-y>

527 Ramasamy, D., & Madhow, U. (2015). Compressive spectral embedding: Sidestepping the SVD. In C.  
 528 Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information*  
 529 *processing systems* (Vol. 28). Curran Associates, Inc.

Ray, J. C., Grimm, J., & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068.  
<https://doi.org/10.1139/facets-2020-0075>

Runghen, R., Stouffer, D. B., & Dalla Riva, G. V. (2021). *Exploiting node metadata to predict interactions in large networks using graph embedding and neural networks*.  
<https://doi.org/10.1101/2021.06.10.447991>

S, N. S., & Surendran, S. (2013). Graph Embedding and Dimensionality Reduction - A Survey. *International Journal of Computer Science & Engineering Technology*, 4(1).

Saravia, L. A., Marina, T. I., Kristensen, N. P., De Troch, M., & Momo, F. R. (2021). Ecological network assembly: How the regional metaweb influences local food webs. *Journal of Animal Ecology*, n/a(n/a).  
<https://doi.org/10.1111/1365-2656.13652>

Schmidt, C., & Garroway, C. J. (2022). Systemic racism alters wildlife genetic diversity. *Proceedings of the National Academy of Sciences*, 119(43), e2102860119. <https://doi.org/10.1073/pnas.2102860119>

Stevenson, M. T., & Doleac, J. L. (2021). *Algorithmic Risk Assessment in the Hands of Humans* ({{SSRN Scholarly Paper}} No. 3489440). <https://doi.org/10.2139/ssrn.3489440>

Stier, A. C., Samhouri, J. F., Gray, S., Martone, R. G., Mach, M. E., Halpern, B. S., Kappel, C. V., Scarborough, C., & Levin, P. S. (2017). Integrating Expert Perceptions into Food Web Conservation and Management. *Conservation Letters*, 10(1), 67–76. <https://doi.org/10.1111/conl.12245>

Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science*, 335(6075), 1489–1492. <https://doi.org/10.1126/science.1216556>

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. (2022). Food web reconstruction through phylogenetic transfer of low-rank network representation. *Methods in Ecology and Evolution*, n/a(n/a). <https://doi.org/10.1111/2041-210X.13835>

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. (2021). A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1837), 20210063.



558 <https://doi.org/10.1098/rstb.2021.0063>

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

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

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

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

569 Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., DelaCruz, J., Dunkerley, G., Fernandez, M., Ferrier,  
 570 S., Gill, M., Jones, H., Koen, T., Leys, J., McGeoch, M., Mihoub, J.-B., Scanes, P., Schmeller, D., &  
 571 Williams, K. (2017). Using the essential biodiversity variables framework to measure biodiversity  
 572 change at national scale. *Biological Conservation*, 213, 264–271.  
 573 <https://doi.org/10.1016/j.biocon.2016.08.019>

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

577 Wang, S., Arroyo, J., Vogelstein, J. T., & Priebe, C. E. (2021). Joint Embedding of Graphs. *IEEE*  
 578 *Transactions on Pattern Analysis and Machine Intelligence*, 43(4), 1324–1336.  
 579 <https://doi.org/10.1109/TPAMI.2019.2948619>

580 Wardeh, M., Baylis, M., & Blagrove, M. S. C. (2021). Predicting mammalian hosts in which novel  
 581 coronaviruses can be generated. *Nature Communications*, 12(1), 780.  
 582 <https://doi.org/10.1038/s41467-021-21034-5>

583 Weiskopf, S. R., Harmáčková, Z. V., Johnson, C. G., Londoño-Murcia, M. C., Miller, B. W., Myers, B. J. E.,  
 584 Pereira, L., Arce-Plata, M. I., Blanchard, J. L., Ferrier, S., Fulton, E. A., Harfoot, M., Isbell, F., Johnson,  
 585 J. A., Mori, A. S., Weng, E., & Rosa, I. M. D. (2022). Increasing the uptake of ecological model results in



586 policy decisions to improve biodiversity outcomes. *Environmental Modelling & Software*, 149, 105318.  
 587 <https://doi.org/10.1016/j.envsoft.2022.105318>

588 Wood, S. A., Russell, R., Hanson, D., Williams, R. J., & Dunne, J. A. (2015). Effects of spatial scale of  
 589 sampling on food web structure. *Ecology and Evolution*, 5(17), 3769–3782.  
 590 <https://doi.org/10.1002/ece3.1640>

591 Wu, D., Palmer, D. R., & Deford, D. R. (2021). *Maximum a Posteriori Inference of Random Dot Product*  
 592 *Graphs via Conic Programming* (No. arXiv:2101.02180). arXiv. <https://arxiv.org/abs/2101.02180>

593 Xu, M. (2021). Understanding Graph Embedding Methods and Their Applications. *SIAM Review*, 63(4),  
 594 825–853. <https://doi.org/10.1137/20M1386062>

595 Yan, S., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A general framework for  
 596 dimensionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern*  
 597 *Recognition (CVPR'05)*, 2, 830–837 vol. 2. <https://doi.org/10.1109/CVPR.2005.170>

598 Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In A.  
 599 Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.  
 600 [https://doi.org/10.1007/978-3-540-77004-6\\_11](https://doi.org/10.1007/978-3-540-77004-6_11)

601 Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural  
 602 networks: A review of methods and applications. *AI Open*, 1, 57–81.  
 603 <https://doi.org/10.1016/j.aiopen.2021.01.001>

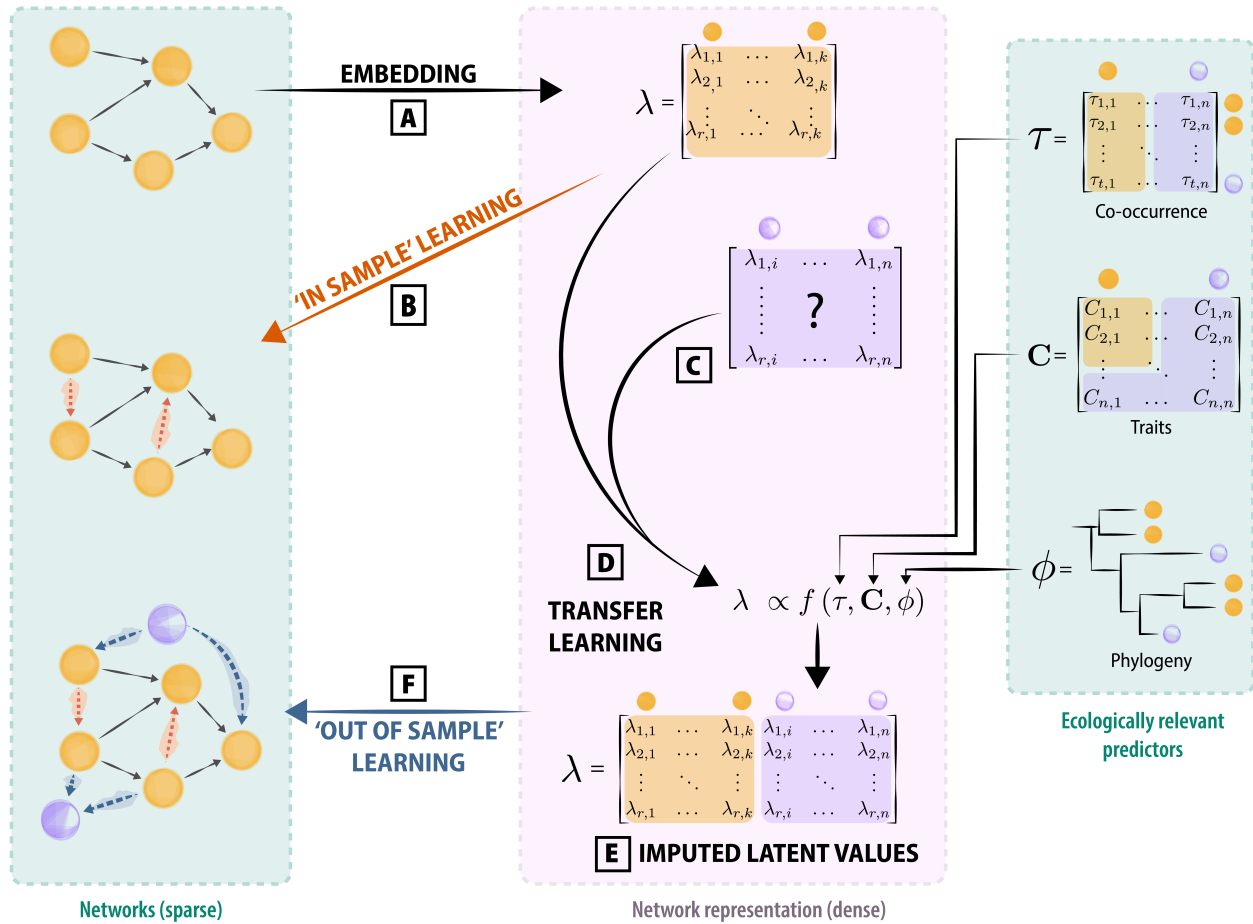


Figure 1: The embedding process (A) can help to identify links (interactions) that may have been missed within the current sample (community, B), while transfer learning (D) allows for predicting links (interactions) for species external to the original sample as well as within sample links (F). Transfer learning allows for the prediction links (interactions) when novel species (C) are included in the sample by learning using other relevant predictors (e.g. traits) in conjunction with the known interactions to infer latent values (E). Within this context the predicted networks as well as the ecological predictors used (green boxes) are products that can be quantified through field measurements, whereas the embedded as well as imputed matrices are representative of **TK**

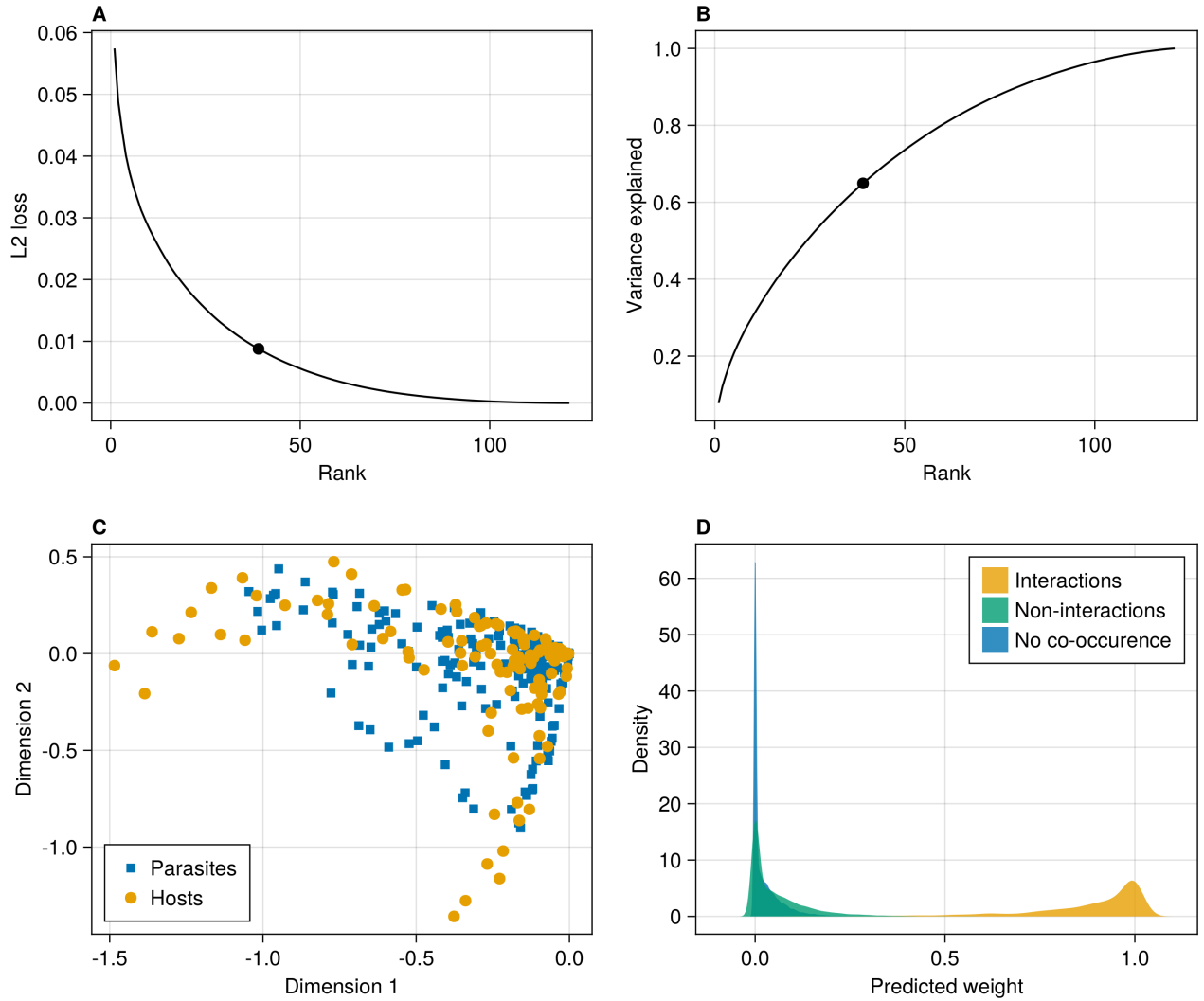


Figure 2: Validation of an embedding for a host-parasite metaweb, using Random Dot Product Graphs. **A**, decrease in approximation error as the number of dimensions in the subspaces increases. **B**, increase in cumulative variance explained as the number of ranks considered increases; in **A** and **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.

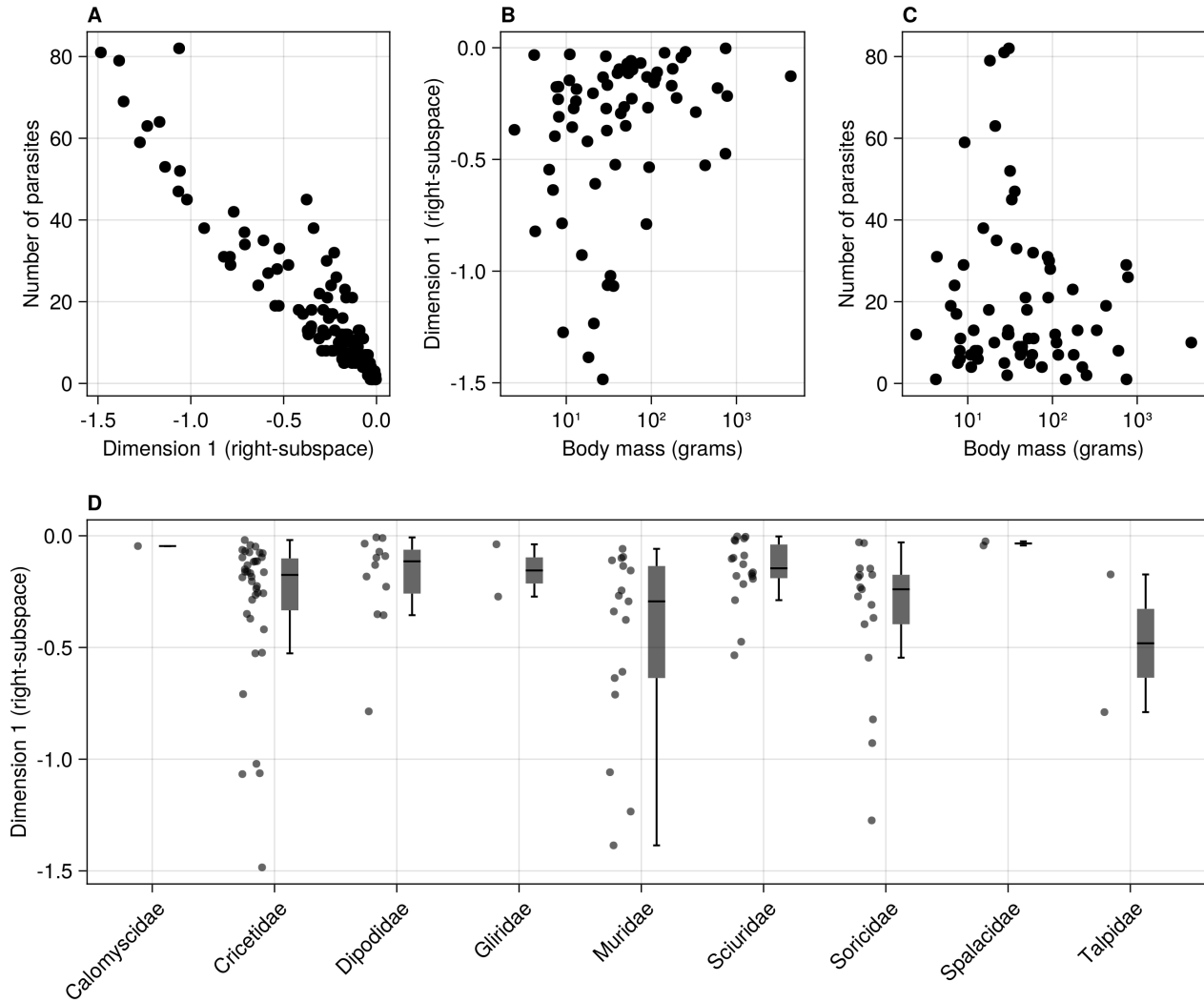


Figure 3: Ecological analysis of an embedding for a host-parasite metaweb, using Random Dot Product Graphs. **A**, relationship between the number of parasites and position along the first axis of the right-subspace for all hosts, showing that the embedding captures elements of network structure at the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