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

Tanya Strydom [1,2,‡] Salomé Bouskila [1] Francis Banville [1,2,3] Ceres Barros [4] Dominique Caron [2,5] Maxwell J Farrell [6] Marie-Josée Fortin [6] Victoria Hemming [7] Benjamin Mercier [2,3] Laura J. Pollock [2,5] Rogini Runghen [8] Giulio V. Dalla Riva [9] Timothée Poisot [1,2,‡]

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

[‡] Equal contributions

Correspondance to:

Timothée Poisot — `timothee.poisot@umontreal.ca`

1. Metawebs, (networks of potential interactions within a species pool) are a powerful abstraction to understand how large-scale species interaction networks are structured.

2. 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 'draft' metawebs.

3. One 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.

4. In 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.

Last revision: *January 25, 2023*

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

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

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

A metaweb is an inherently probabilistic object

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

Yet the original metaweb definition, and indeed most past uses of metawebs, was based on the presence/absence of interactions. Moving towards *probabilistic* metawebs, by representing interactions as Bernoulli events (see *e.g.* Poisot et al., 2016), offers the opportunity to weigh these rare interactions 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 (Catchen et al., 2023). 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 relying on machine learning approaches emphasize the idea that algorithms meant to predict, through the assumption that they approximate the process generating the data, 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, 2013; Bartomeus et al., 2016). This would represent a departure from simple rules expressed at the network scale (*e.g.* Williams & Martinez, 2000) to a view of network prediction based on learning the rules that underpin interactions *and* their variability (Gupta et al., 2022).

[Figure 1 about here.]

Graph embedding offers promises for the inference of potential interactions

Graph (or Network) embedding (fig. 1) is a family of machine learning techniques, whose main task is to learn a mapping function from a discrete graph to a continuous domain (Chami et al., 2022; **arsov_network_2019?**). Their main goal is to learn a low dimensional vector representations of the graph (embeddings), such that its key properties (*e.g.* local or global structures) are retained in the embedding space (Yan et al., 2005). The embedding space may, but will not necessarily, have lower dimensionality than the graph. Ecological networks are promising candidates for the routine application of embeddings, as they tend to possess a shared structural backbone (see *e.g.* Bramon Mora et al., 2018), which hints at structural invariants in empirical data. Assuming that these structural invariants are common enough, they would dominate the structure of networks, and therefore be adequately captured by the first (lower) dimensions of an embedding, without the need to measure derived aspects of their structure (*e.g.* motifs, paths, modularity, …).

Indeed, food webs are inherently low-dimensional objects, and can be adequately represented with less than ten dimensions (J. Braga et al., 2019; M. P. Braga et al., 2021; Eklöf et al., 2013). Simulation results by Botella et al. (2022) suggests that there is no dominant method to identify architectural similarities between networks: multiple approaches need to be tested and compared to the network descriptor of interest on a problem-specific basis. This matches previous results on graph embedding, wherein different embedding algorithms yield different network embeddings (Goyal & Ferrara, 2018), calling for a careful selection of the problem-specific approach to use. In tbl. 1, we present a selection of common graph and node embedding methods, alongside examples of their use to predict interactions or statistical associations between species. These methods rely largely on linear algebra or pseudo-random walks on graphs. All forms of embeddings presented in the table share the common property of summarizing their objects into (sets of) dense feature vectors, that capture the overall network structure, pairwise information on nodes, and emergent aspects of the network, in a compressed way (*i.e.* with some information loss, as we later discuss in the illustration). Node embeddings tend to focus on maintaining pairwise relationships (*i.e.* species interactions), while graph embeddings focus on maintaining the network structure (*i.e.* emergent properties). Nevertheless, some graph embedding techniques (like RDPG, see *e.g.* Wu et al., 2021) will provide high-quality node-level embeddings while also preserving network structure.

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 tbl. 1. 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. [a]: application is concerned with *statistical* interactions, which are not necessarilly direct biotic interactions; [b]: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)	Cieslak et al. (2020) [a] Gibb et al. (2021)
LINE	nodes	stochastic gradient descent	Tang et al. (2015)	
SDNE	nodes	gradient descent	D. Wang et al. (2016)	

Method	Object	Technique	Reference	Application
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) [b]
graph2vec	sub-graph	skipgram network	Narayanan et al. (2017)	
RDPG	graph	SVD	Young & Scheinerman (2007)	Dalla Riva & Stouffer (2016); Poisot et al. (2021)
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) [a]
FastEmbed	graph	eigen decomposition	Ramasamy & Madhow (2015)	
PCA	graph	eigen decomposition	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. 1 **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 (M. P. Braga et al., 2021; Dalla Riva & Stouffer, 2016; Eklöf & Stouffer, 2016; Stouffer et al., 2012; Stouffer et al., 2007); similarly, the use of functional traits assumes that interactions can be inferred from the knowledge of trait-matching rules, which is similarly well supported in the empirical literature (Bartomeus, 2013; Bartomeus et al., 2016; Goebel et al., 2023; Gravel et al., 2013). Relating this information to an embedding rather than a list of networks measures would allow to capture their effect on the more fundamental aspects of network structure; conversely, the absence of a phylogenetic or functional signal may suggest that evolutionary/trait processes are not strong drivers of network structure, therefore opening a new way to perform hypothesis testing.

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

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

Graph embeddings *can* serve as a dimensionality reduction method. For example, RDPG (Strydom et al., 2022) and t-SVD [truncated Singular Value Decomposition; Poisot et al. (2021)] typically embed networks using fewer dimensions than the original network [the original network has as many dimensions as species, and as many informative dimensions as trophically unique species; Strydom, Dalla Riva, et al. (2021)]. But this is not necessarily the case – indeed, one may perform a PCA (a special case of SVD) to 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. 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. 2, 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 39) 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 RDPG are not bound to the unit interval, we performed a clamping of the weights to the unit space, showing a one-inflation in documented interactions, 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. 2 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 RDPG is able to capture pairwise processes) and the cumulative variance explained (indicating that RDPG is able to capture network-level structure). Therefore, in fig. 3, 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 establihed 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 predictors for transfer learning.

The metaweb embeds both ecological hypotheses and practices

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

Identifying the properties of the network to embed

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

Identifying the scope of the prediction to perform

The area used to infer the new metaweb in determines the species pool that must be used to perform this embedding. Metawebs can be constructed by assigning interactions in a list of species within geographic boundaries. The upside of this approach is that information at the country level is likely to be required for biodiversity assessments, as countries set conservation goals at the national level (Buxton et al., 2021), and as quantitative instruments are designed to work at these scales (**turak_measuring_2017?**); specific strategies are often enacted at smaller scales, nested within a specific country (Ray et al., 2021). But there is no guarantee that these boundaries are meaningful. In fact, we do not have a satisfying answer to the question of "where does an ecological network stop?". Recent results by Martins et al. (2022) suggest that networks are shaped within eco-regions, with abrupt structural transitions from an eco-region to the next. Should this trend hold generally, this would provide an ecologically-relevant scale at which metawebs can be downscaled and predicted. Other solutions leverage network-area relationships to identify areas in which networks are structurally similar (see *e.g.* Fortin et al., 2021; Galiana et al., 2022, 2018), which requires ample pre-existing information about the network. This suggests that inferred metawebs should be further downscaled to allow for *a posteriori* analyses; Llewelyn et al. (2022) provide compelling evidence for the fact that when a place is rich in endemic species, even at smaller spatial scales, the transfer of information about interactions becomes more challenging, re-emphasizing the need to identify guidelines for when and where network predictions can be mapped.

Minding legacies shaping ecological datasets

Operating under the context of national divisions, in large parts of the world, reflects nothing more than the legacy of settler colonialism, which drives a disparity in available ecological data. Applying any embedding to biased data does not debias them, but instead embeds these very same biases, propagating them to the machine learning models using embeddings to make predictions. Indeed, the use of ecological data itself is not an apolitical act (Nost & Goldstein, 2021), as data infrastructures tend to be designed to answer questions within national boundaries (therefore placing contingencies on what is available to be embedded). Furthermore, their use often draws upon, and reinforces, territorial statecraft (see *e.g.* Barrett, 2005). As per Machen & Nost (2021), these biases are particularly important to consider when the output of "algorithmic thinking" (*i.e.* relying on machine learning to generate knowledge or as a substitute to human decision-making) can be re-purposed for governance (*e.g.* enacting conservation decisions on the basis of model prediction). As information on species interaction networks structure is increasingly leveraged as a tool to guide 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 paramount. These considerations are even more urgent in the specific context of biodiversity data, as long-term colonial legacies still shape taxonomic composition to this day (Lenzner et al., 2022; Raja, 2022), and where much shorter-term changes in taxonomic and genetic richness of wildlife emerged through environmental racism (Schmidt & Garroway, 2022).

Conclusion: metaweb prediction in context

Predictive approaches, regardless of the scale at which they are deployed and the intent of their deployment, originate in the framework that contributed to the ongoing biodiversity crisis (Adam, 2014) and reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020). The risk of embedding this legacy in our models is great, especially when the impact of this legacy on species pools is being increasingly documented. This problem can be addressed by re-framing the way we interact with models, especially in the context where these models are intended to support conservation actions. Particularly on territories that were traditionally stewarded by Indigenous people, these approaches should be replaced (or at least guided and framed) by Indigenous principles of land management (Eichhorn et al., 2019; No'kmaq et al., 2021), as part of an "algorithm-in-the-loop" approach. Human-algorithm interactions are notoriously difficult and can yield adverse effect (Green & Chen, 2019; Stevenson & Doleac, 2021), suggesting the need to systematically study them for the specific purpose of biodiversity governance, as well as to improve the algorithmic literacy of decision makers. As we see artificial intelligence/machine learning being increasingly mobilized to generate knowledge that is lacking for conservation decisions (*e.g.* Lamba et al., 2019; Mosebo Fernandes et al., 2020) and drive policy decisions (Weiskopf et al., 2022), our discussion of these tools need to go beyond the technical and statistical, and into the governance consequences they can have.

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

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

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

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

References

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

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

Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer*

Science Review, *40*, 100378. https://doi.org/10.1016/j.cosrev.2021.100378

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

Barrett, S. (2005). *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/0199286094.001.0001

Bartomeus, I. (2013). Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. *PloS One*, *8*(7), e69200. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069200

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

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

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

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

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

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

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

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

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

Catchen, M., Poisot, T., Pollock, L., & Gonzalez, A. (2023). *The missing link: Discerning true from false negatives when sampling species interaction networks* (Preprint No. 4929). EcoEvoRXiV. https://doi.org/10.32942/X2DW22

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

Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., & Murphy, K. (2022). Machine Learning on Graphs: A Model and Comprehensive Taxonomy. *Journal of Machine Learning Research*, *23*(89), 1–64. http://jmlr.org/papers/v23/20-852.html

Chen, D., Xue, Y., Fink, D., Chen, S., & Gomes, C. P. (2017). *Deep Multi-species Embedding*. 3639–3646. https://www.ijcai.org/proceedings/2017/509

Chen, H., Perozzi, B., Hu, Y., & Skiena, S. (2017). *HARP: Hierarchical Representation Learning for Networks*. http://arxiv.org/abs/1706.07845

Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity, colonialism and conquest in an era of capitalist globalization. In *NGOization: Complicity, contradictions and prospects* (pp. 24–44). Bloomsbury Publishing.

Cieslak, M. C., Castelfranco, A. M., Roncalli, V., Lenz, P. H., & Hartline, D. K. (2020). T-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Marine*

Genomics, *51*, 100723. https://doi.org/10.1016/j.margen.2019.100723

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

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

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

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

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

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

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

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

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

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

Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1949), rspb.2020.1889, 20201889.

https://doi.org/10.1098/rspb.2020.1889

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

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

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

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

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

Goebel, L. G. A., Vitorino, B. D., Frota, A. V. B., & Santos-Filho, M. dos. (2023). Body mass determines the role of mammal species in a frugivore-large fruit interaction network in a Neotropical savanna. *Journal of Tropical Ecology*, *39*, e12. https://doi.org/10.1017/S0266467422000505

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

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

Gravel, D., Baiser, B., Dunne, J. A., Kopelke, J.-P., Martinez, N. D., Nyman, T., Poisot, T., Stouffer, D. B., Tylianakis, J. M., Wood, S. A., & Roslin, T. (2018). Bringing Elton and Grinnell together: A quantitative framework to represent the biogeography of ecological interaction networks. *Ecography*, *0*(0).

https://doi.org/10.1111/ecog.04006

Gravel, D., Poisot, T., Albouy, C., Velez, L., & Mouillot, D. (2013). Inferring food web structure from predator-prey body size relationships. *Methods in Ecology and Evolution*, *4*(11), 1083–1090. https://doi.org/10.1111/2041-210X.12103

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

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

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

Gupta, A., Furrer, R., & Petchey, O. L. (2022). Simultaneously estimating food web connectance and structure with uncertainty. *Ecology and Evolution*, *12*(3), e8643. https://doi.org/10.1002/ece3.8643

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

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

Herbert, F. (1965). *Dune* (1st ed.). Chilton Book Company.

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

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

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

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

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

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

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

Llewelyn, J., Strona, G., Dickman, C. R., Greenville, A. C., Wardle, G. M., Lee, M. S. Y., Doherty, S., Shabani, F., Saltré, F., & Bradshaw, C. J. A. (2022). *Predicting predator-prey interactions in terrestrial endotherms using random forest* [Preprint]. Ecology. https://doi.org/10.1101/2022.09.02.506446

Maaten, L. van der. (2009). Learning a Parametric Embedding by Preserving Local Structure. *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics*, 384–391. https://proceedings.mlr.press/v5/maaten09a.html

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

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

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

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

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

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

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

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

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

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

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

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

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

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

O'Connor, L. M. J., Pollock, L. J., Braga, J., Ficetola, G. F., Maiorano, L., Martinez-Almoyna, C., Montemaggiori, A., Ohlmann, M., & Thuiller, W. (2020). Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *Journal of Biogeography*, *47*(1), 181–192. https://doi.org/10.1111/jbi.13773

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

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

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

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

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

Poisot, T., Ouellet, M.-A., Mollentze, N., Farrell, M. J., Becker, D. J., Albery, G. F., Gibb, R. J., Seifert, S. N., & Carlson, C. J. (2021). *Imputing the mammalian virome with linear filtering and singular value decomposition*. http://arxiv.org/abs/2105.14973

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

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

Ramasamy, D., & Madhow, U. (2015). Compressive spectral embedding: Sidestepping the SVD. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 28). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/file/4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf

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

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., Camacho, J., Jiang, W., & Nunes Amaral, L. A. (2007). Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1621), 1931–1940. https://doi.org/10.1098/rspb.2007.0571

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. https://doi.org/10.1098/rstb.2021.0063

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

Surendran, S. (2013). Graph Embedding and Dimensionality Reduction - A Survey. *International Journal of Computer Science & Engineering Technology*, *4*(1). https://www.semanticscholar.org/paper/Graph-Embedding-and-Dimensionality-Reduction-A-Surendran/3f413d591e4b2b876e033eeb9390e232ad4826ca

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

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

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

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

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

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

Weiskopf, S. R., Harmáčková, Z. V., Johnson, C. G., Londoño-Murcia, M. C., Miller, B. W., Myers, B. J. E., Pereira, L., Arce-Plata, M. I., Blanchard, J. L., Ferrier, S., Fulton, E. A., Harfoot, M., Isbell, F., Johnson, J. A., Mori, A. S., Weng, E., & Rosa, I. M. D. (2022). Increasing the uptake of ecological model results in policy decisions to improve biodiversity outcomes. *Environmental Modelling & Software*, *149*, 105318. https://doi.org/10.1016/j.envsoft.2022.105318

Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, *404*(6774), 180–183. https://doi.org/10.1038/35004572

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

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

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

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

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

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

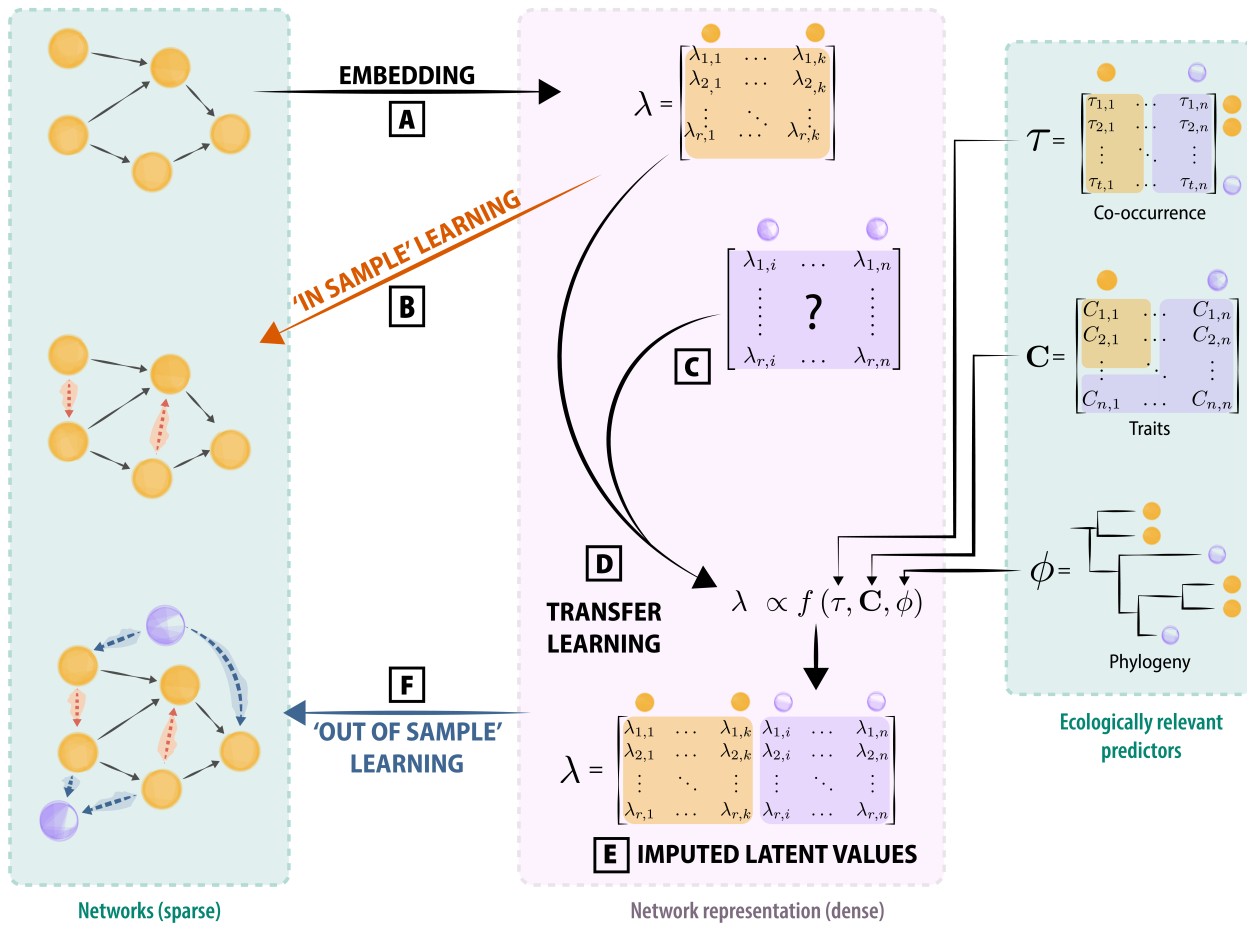

Figure 1: The embedding process (**A**) can help to identify links (interactions) that may have been missed within the original community (represented by the orange dashed lines, **B**). Transfer learning allows for the prediction links (interactions) when novel species (**C**) are included alongside the original community (**D**). This is achieved by learning using other 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 as well as 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 field measurements, whereas the embedded as well as imputed matrices (purple box) are representative of a decomposition of the interaction networks

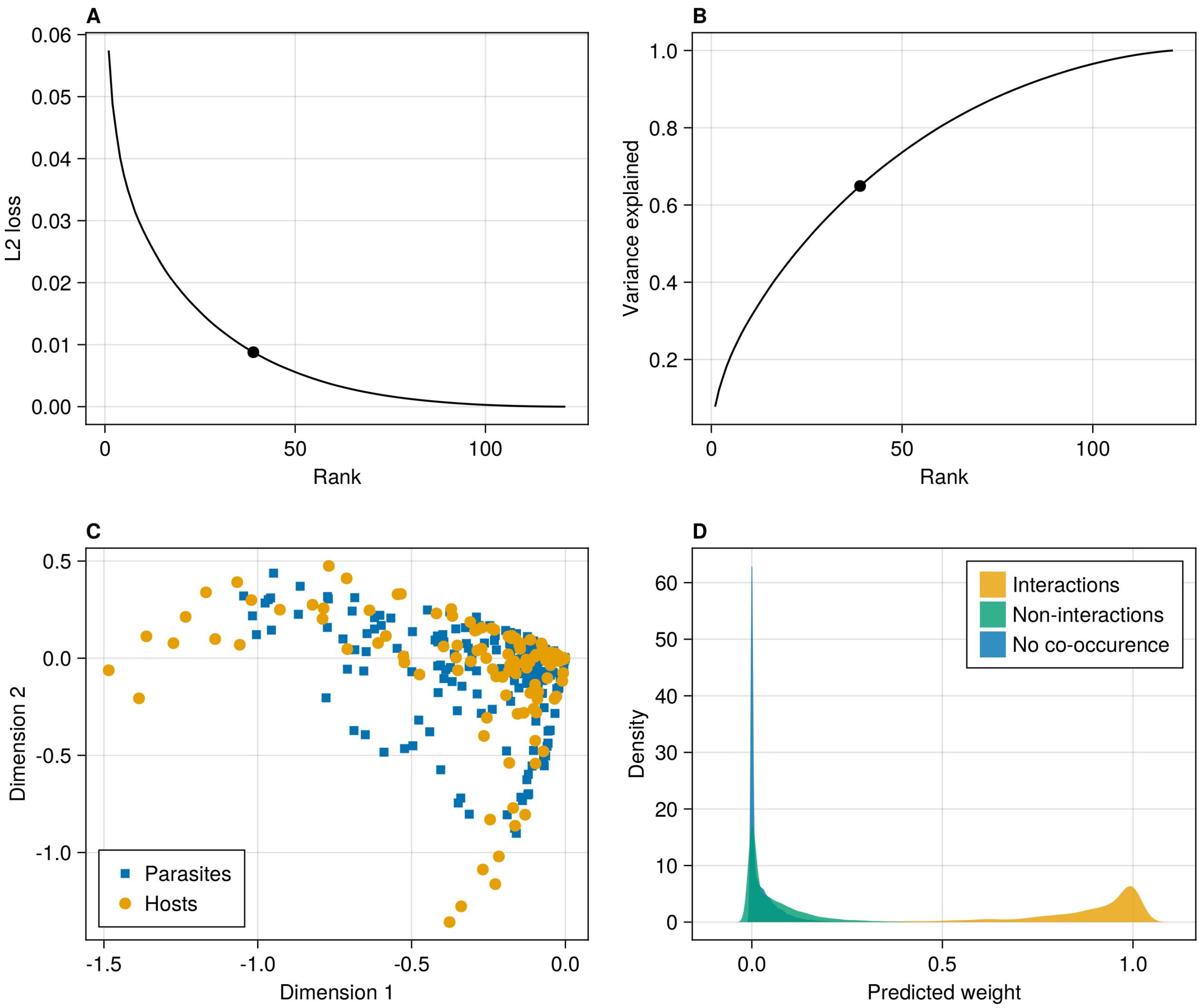

Figure 2: Validation of an embedding for a host-parasite metaweb, using Random Dot Product Graphs. **A**, decrease in approximation error as the number of dimensions in the subspaces increases. **B**, increase in cumulative variance explained as the number of ranks considered increases; in **A** and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

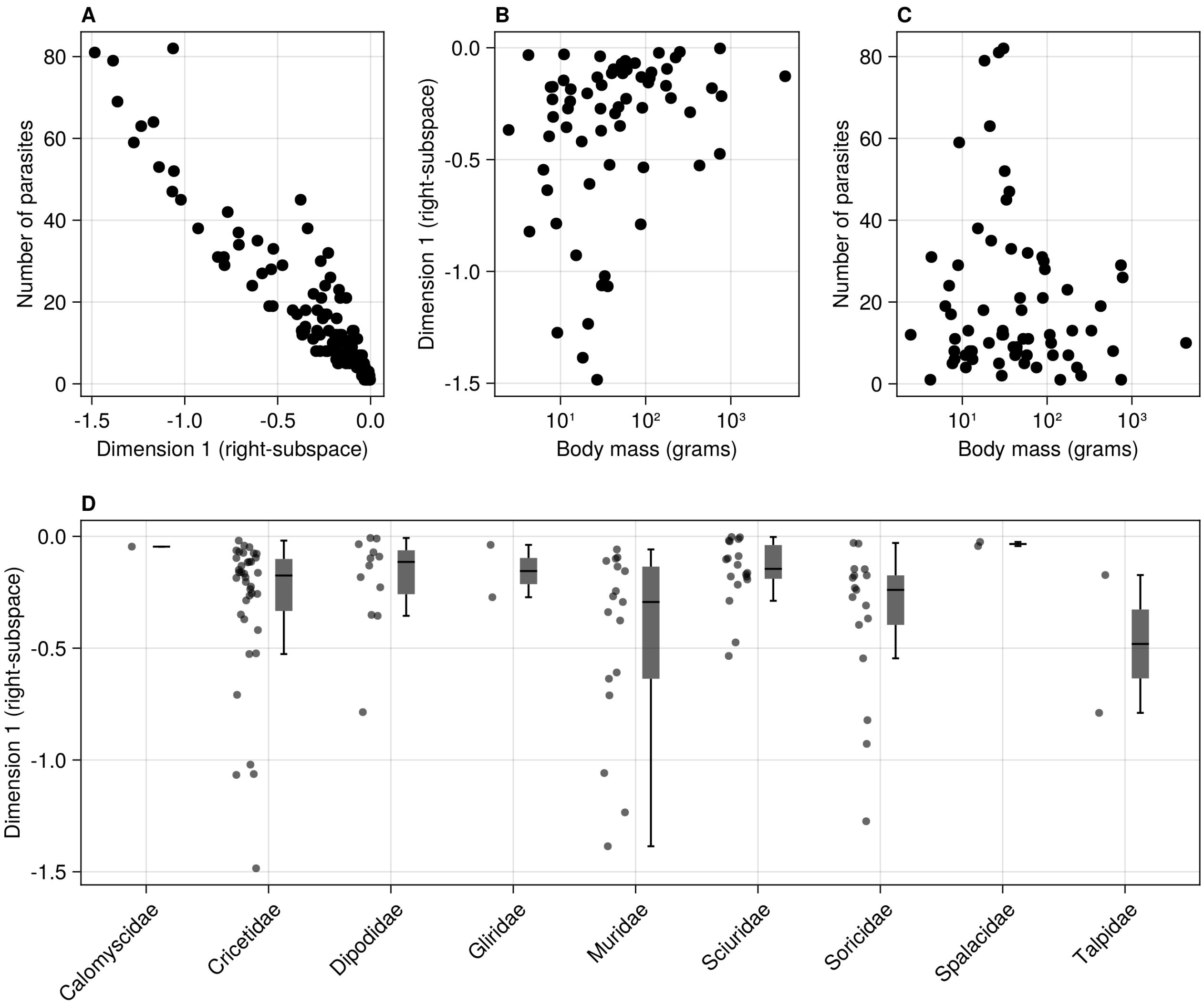

Figure 3: Ecological analysis of an embedding for a host-parasite metaweb, using Random Dot Product Graphs. **A**, relationship between the number of parasites and position along the first axis of the right-subspace for all hosts, showing that the embedding captures elements of network structure at the 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.