

# 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>1,3,2</sup> Ceres Barros<sup>4</sup> Dominique Caron<sup>5,2</sup>  
Maxwell J Farrell<sup>6</sup> Marie-Josée Fortin<sup>6</sup> Victoria Hemming<sup>7</sup> Benjamin Mercier<sup>3,2</sup> Laura  
J. Pollock<sup>5,2</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

‡ These authors contributed equally to the work

## Correspondance to:

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

1. Metawebs, i.e. networks of potential interactions within a species pool, are a powerful abstraction to understand how large-scales 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 the predictive ability is to maximize the information used for prediction, by using graph embeddings rather than 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; furthermore, because metawebs are inherently spatial objects, we 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.

1 Being able to infer *potential* interactions could be the catalyst for significant breakthroughs in our ability  
2 to start thinking about species interaction networks over large spatial scales (Hortal et al., 2015).

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

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

28 Because the metaweb represents the joint effect of functional, phylogenetic, and macroecological  
29 processes (Morales-Castilla et al., 2015), it holds valuable ecological information. Specifically, it is the  
30 “upper bounds” on what the composition of the local networks, given the local species pool, can be (see

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

## 41 **The metaweb is an inherently probabilistic object**

42 Treating interactions as probabilistic (as opposed to binary) events is a more nuanced and realistic way to  
43 represent them. Dallas et al. (2017) suggested that most links in ecological networks are cryptic, *i.e.*  
44 uncommon or hard to observe. This argument echoes Jordano (2016): sampling ecological interactions is  
45 difficult because it requires first the joint observation of two species, and then the observation of their  
46 interaction. In addition, it is generally expected that weak or rare links to be more prevalent in networks  
47 than common or rare links (Csermely, 2004), compared to strong, persistent links; this is notably the case  
48 in food chains, wherein many weaker links are key to the stability of a system (Neutel et al., 2002). In the  
49 light of these observations, we expect to see an over-representation of low-probability interactions under a  
50 model that accurately predicts interaction probabilities. Yet the original metaweb definition, and indeed  
51 most past uses of metawebs, was based on the presence/absence of interactions. Moving towards  
52 *probabilistic* metawebs, by representing interactions as Bernoulli events (see e.g. Poisot et al., 2016), offers  
53 the opportunity to weigh these rare interactions appropriately. The inherent plasticity of interactions is  
54 important to capture: there have been documented instances of food webs undergoing rapid  
55 collapse/recovery cycles over short periods of time (e.g. Pedersen et al., 2017). These considerations  
56 emphasize why metaweb predictions should focus on quantitative (preferentially probabilistic)  
57 predictions; this should constrain the suite of appropriate models.

58 Yet it is important to recall that a metaweb is intended as a catalogue of all potential interactions, which is

59 then filtered (Morales-Castilla et al., 2015). In a sense, that most ecological interactions are elusive can call  
60 for a slightly different approach to sampling: once the common interactions are documented, the effort  
61 required in documenting each rare interaction will increase exponentially. Recent proposals suggest that  
62 machine learning algorithms, in these situations, can act as data generators (Hoffmann et al., 2019): high  
63 quality observational data can generate the core rules underpinning the network structure, and be  
64 supplemented with synthetic data coming from predictive models, increasing the volume of information  
65 available for inference. Indeed, Strydom, Catchen, et al. (2021) suggested that knowing the metaweb may  
66 render the prediction of local networks easier, because it fixes an “upper bound” on which interactions  
67 can exist. In this context a probabilistic metaweb represents an aggregation of informative priors on the  
68 interactions, elusive information with the potential to boost our predictive ability (Bartomeus et al., 2016).

69 [Figure 1 about here.]

## 70 **Graph embedding offers promises for the inference of potential** 71 **interactions**

72 Graph embedding (fig. 1) is a varied family of machine learning techniques aiming to transform nodes and  
73 edges into a vector space (Arsov & Mirceva, 2019), usually of a lower dimension, whilst maximally  
74 retaining key properties of the graph (Yan et al., 2005). Ecological networks are an interesting candidate  
75 for the widespread application of embeddings, as they tend to possess a shared structural backbone  
76 (Bramon Mora et al., 2018), which hints at structural invariants that can be revealed at lower dimensions.  
77 Indeed, food webs are inherently low-dimensional objects, and can be adequately represented with less  
78 than ten dimensions (Braga et al., 2019; Eklöf et al., 2013). Simulation results by Botella et al. (2022)  
79 suggest that there is no best method to identify architectural similarities between networks, and that  
80 multiple approaches need to be tested and compared to the network descriptor of interest. This matches  
81 previous, more general results on graph embedding, which suggest that the choice of embedding  
82 algorithm matters for the results (Goyal & Ferrara, 2018). In tbl. 1, we present a selection of common  
83 graph embedding methods, alongside examples of their use to predict species interactions; most of these  
84 methods rely either on linear algebra, or on pseudo-random walks on graphs.

85 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 about currently available 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 about here.]

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, 2020) 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 nodes metadata. This example has many parallels in ecology (see fig. 1 C), in which node metadata can be given by phylogeny or 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 (RDPG; see Dalla Riva & Stouffer, 2016). In a recent publication, (Strydom2022FooWeb?) 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 by using the phylogenetic distance between related clades to infer the values in the latent sub-space into which the metaweb was projected.

115 By performing the RDPG step on re-constructed value, this approach yields a probabilistic trophic  
 116 metaweb for mammals of Canada based on knowledge of European species, despite a limited ( $\approx 5\%$ )  
 117 taxonomic overlap.

118 Graph embeddings *can* serve as a dimensionality reduction method. For example, RDPG  
 119 (**Strydom2022FooWeb?**) and t-SVD (Poisot et al., 2021) typically embed networks using fewer  
 120 dimensions than the original network (the original network has as many dimensions as species, and as  
 121 many informative dimensions as trophically unique species; Strydom, Dalla Riva, et al., 2021). But this is  
 122 not necessarily the case – indeed, one may perform a PCA (a special case of SVD) to project the raw data  
 123 into a subspace that improves the efficacy of t-SNE (Maaten, 2009). There are many dimensionality  
 124 reductions (Anowar et al., 2021) that can be applied to an embedded network should the need for  
 125 dimensionality reduction (for example for data visualisation) arise. In brief, many graph embeddings *can*  
 126 serve as dimensionality reduction steps, but not all do, neither do all dimensionality reduction methods  
 127 provide adequate graph embedding capacities. In the next section (and fig. 2), we show how the amount of  
 128 dimensionality reduction can affect the quality of the embedding.

## 129 **An illustration of metaweb embedding**

130 In this section, we illustrate the embedding of a collection of bipartite networks collected by Hadfield et al.  
 131 (2014), using truncated Singular Value Decomposition (t-SVD) and RDPG (see **Strydom2022FooWeb?**  
 132 for the full details). Briefly, an RDPG decomposes a network into two subspaces (left and right), which are  
 133 matrices that when multiplied give an approximation of the original network. The code to reproduce this  
 134 example is available as supplementary material (note, for the sake of comparison, that Strydom, Catchen,  
 135 et al., 2021 have an example using embedding through PCA followed by prediction using a deep neural  
 136 network on the same dataset). The resulting (binary) metaweb  $\mathcal{M}$  has 2131 interactions between 206  
 137 parasites and 121 hosts, and its adjacency matrix has full rank (*i.e.* it represents a space with 121  
 138 dimensions). All analyses were done using Julia (Bezanson et al., 2017) version 1.7.2, *Makie.jl* (Danisch &  
 139 Krumbiegel, 2021), and *EcologicalNetworks.jl* (Poisot et al., 2019).

140 [Figure 2 about here.]

141 The embedding of the metaweb holds several pieces of information (fig. 2). In panel **A**, we show that the

$L_2$  loss (*i.e.* the sum of squared errors) between the empirical and reconstructed metaweb decreases when number of dimensions (rank) of the subspace increases, with a point of inflection around 25 dimensions. 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 this instance, accepting  $L_2 = 500$  as an approximation of the network means that the error for every position in the metaweb is  $\approx (500/(206 \times 121))^{1/2}$ . In **B**, we show the 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. In **C**, 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 possible 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). Note that the values are not bound to the unit interval, which emphasizes the need for either scaling or clamping (although thresholding analyses are insensitive to this choice). Finally, in **D**, we show that the embedding, as it captures structural information about the network, holds ecological information; indeed, the position of the parasite on the first dimension of the left sub-space is a linear predictor of its number of hosts.

## **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 addressed with subsequent data analysis or simulations), as well as issues related to the application of these methods at the science-policy interface.

The first open research problem is the taxonomic and spatial limit of the metaweb to embed and transfer. 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. In addition, other problems can arise due to non-overlapping trait distributions in the known and unknown species. Alternatively a 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-dispersal 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 dispersal is notably true if the metaweb is assembled in an area with mostly endemic species (*e.g.* a system that has undergone recent radiation and 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 dispersal 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.

The second series of problems relate to determining which area should be used to infer the new metaweb in, as this determines the species pool that must be used. 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 goals at the national level (Buxton et al., 2021), and as quantitative instruments are designed to work at these scales (Turak et al., 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 a food web stop?”; the most promising solutions involve examining the spatial consistency of network area relationships (Fortin et al., 2021; see *e.g.* Galiana et al., 2018, 2019, 2021), which is impossible in the absence of enough information about the network itself. This suggests that inferred metawebs should be further downscaled to allow for *a posteriori* analyses. The methodology for metaweb downscaling is currently limited, and it is likely that the sustained effort to characterize the spatial dependency of food web structure will lead to more prescriptive guidelines about the need for prediction downscaling.

The final family of problems relates less to ecological methods than to the praxis of ecological research.

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 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), and their use often draws upon and reinforces territorial statecraft. As per Machen & Nost (2021), this is particularly true when the output of “algorithmic thinking” (e.g. relying on machine learning to generate knowledge) can be re-used for governance (e.g. enacting conservation decisions at the national scale). As information on species interaction networks structure is increasingly leveraged as a tool to guide conservation actions (see e.g. recent discussions for food-web based conservation; 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. Predictive approaches deployed at the continental scale, no matter their intent, originate in the framework that contributed to the ongoing biodiversity crisis (Adam, 2014) and reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020). Particularly on Turtle Island and other 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

230 partnership with fRI Research, and also acknowledges funding from Alberta Innovates and the Forest  
231 Resources Improvement Association of Alberta. M-JF acknowledges funding from NSERC Discovery  
232 Grant and NSERC CRC. RR is funded by New Zealand's Biological Heritage Ngā Koiora Tuku Iho  
233 National Science Challenge, administered by New Zealand Ministry of Business, Innovation, and  
234 Employment. BM is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the  
235 FRQNT master's scholarship. LP acknowledges funding from NSERC Discovery Grant (NSERC  
236 RGPIN-2019-05771). TP acknowledges financial support from the Fondation Courtois, and NSERC  
237 through the Discovery Grants and Discovery Accelerator Supplement programs. MJF is supported by an  
238 NSERC PDF and an RBC Post-Doctoral Fellowship.

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

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

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

## 243 References

- 244 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 245 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,  
246 Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel, D.  
247 (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8, 8),  
248 1153–1161. <https://doi.org/10.1038/s41559-019-0950-y>
- 249 Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality  
250 reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer*  
251 *Science Review*, 40, 100378. <https://doi.org/10.1016/j.cosrev.2021.100378>
- 252 Arsov, N., & Mirceva, G. (2019). *Network Embedding: An Overview*. <http://arxiv.org/abs/1911.11726>
- 253 Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A  
254 common framework for identifying linkage rules across different types of interactions. *Functional*  
255 *Ecology*, 30(12), 1894–1903.  
256 <http://onlinelibrary.wiley.com/doi/10.1111/1365-2435.12666/full>

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

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

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

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

269 Bramon Mora, B., Gravel, D., Gilarranz, L. J., Poisot, T., & Stouffer, D. B. (2018). Identifying a common  
 270 backbone of interactions underlying food webs from different ecosystems. *Nature Communications*,  
 271 9(1), 2603. <https://doi.org/10.1038/s41467-018-05056-0>

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

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

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

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

284 Cieslak, M. C., Castelfranco, A. M., Roncalli, V., Lenz, P. H., & Hartline, D. K. (2020). T-Distributed

285 Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. *Marine*  
 286 *Genomics*, 51, 100723. <https://doi.org/10.1016/j.margen.2019.100723>

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

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

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

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

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

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

300 Eero, M., Dierking, J., Humborg, C., Undeman, E., MacKenzie, B. R., Ojaveer, H., Salo, T., & Köster, F. W.  
 301 (2021). Use of food web knowledge in environmental conservation and management of living  
 302 resources in the Baltic Sea. *ICES Journal of Marine Science*, 78(8), 2645–2663.  
 303 <https://doi.org/10.1093/icesjms/fsab145>

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

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

310 Fortin, M.-J., Dale, M. R. T., & Brimacombe, C. (2021). Network ecology in dynamic landscapes.  
 311 *Proceedings of the Royal Society B: Biological Sciences*, 288(1949), rspb.2020.1889, 20201889.  
 312 <https://doi.org/10.1098/rspb.2020.1889>

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

316 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.  
 317 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,  
 318 *n/a(n/a)*. <https://doi.org/10.1111/ecog.05229>

319 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is  
 320 scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*,  
 321 *42*(6), 1175–1187. <https://doi.org/10.1111/ecog.03684>

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

327 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,  
 328 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, *2*(5),  
 329 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

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

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

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

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

340 Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*

341 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864.  
 342 <https://doi.org/10.1145/2939672.2939754>

343 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs  
 344 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,  
 345 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

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

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

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

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

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

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

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

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

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

368 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>

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>

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](https://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.,



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

400 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,  
 401 Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the  
 402 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*, 4(7),  
 403 170215. <https://doi.org/10.1098/rsos.170215>

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

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

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

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

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

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

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

423 <https://proceedings.neurips.cc/paper/2015/file/4f6ffe13a5d75b2d6a3923922b3922e5-Paper.pdf>

424 Ray, J. C., Grimm, J., & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of

425 federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068.  
 426 <https://doi.org/10.1139/facets-2020-0075>

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

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

433 Stevenson, M. T., & Doleac, J. L. (2021). *Algorithmic Risk Assessment in the Hands of Humans* [SSRN  
 434 Scholarly Paper]. <https://doi.org/10.2139/ssrn.3489440>

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

438 Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz,  
 439 N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap  
 440 towards predicting species interaction networks (across space and time). *Philosophical Transactions of*  
 441 *the Royal Society B: Biological Sciences*, 376(1837), 20210063.  
 442 <https://doi.org/10.1098/rstb.2021.0063>

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

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

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

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

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

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

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

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

470 Xu, M. (2020). *Understanding graph embedding methods and their applications*.  
 471 <http://arxiv.org/abs/2012.08019>

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

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

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

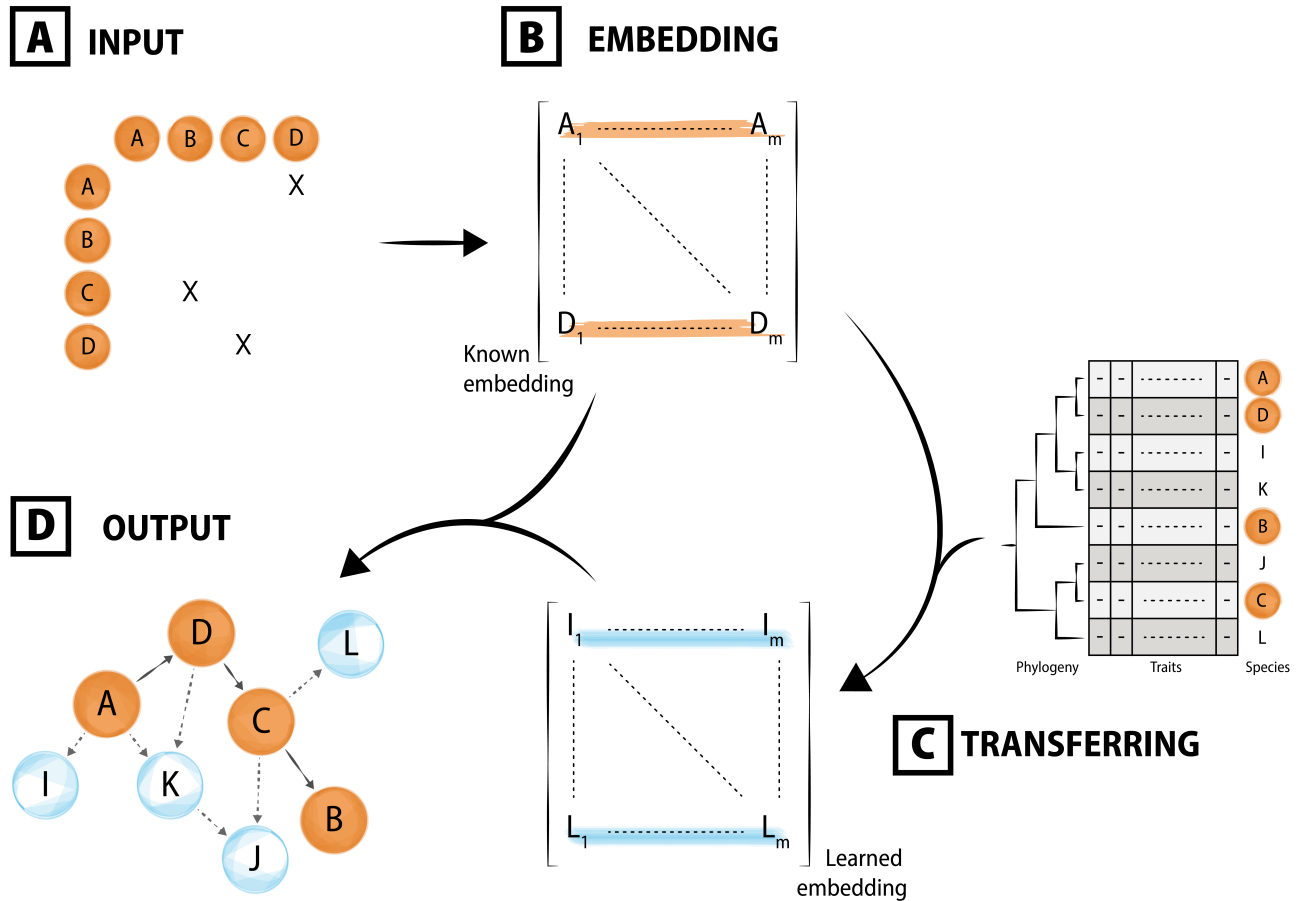


Figure 1: Overview of the embedding process. A network (A), represented here as its adjacency matrix, is converted into a lower-dimensional object (B) where nodes, subgraphs, or edges have specific values (see tbl. 1 for an overview of methods and their use for species interactions). For the purposes of prediction, this low-dimensional object encodes feature vectors for *e.g.* the nodes. Embedding also allows to visualize the structure in the data differently (see fig. 2), much like with a principal component analysis. From a low-dimensional feature vector, it is possible to develop predictive approaches. Nodes in an ecological network are usually species (C), for which we can leverage phylogenetic relatedness (*e.g.* **Strydom2022FooWeb?**) or functional traits to fill the values of additional species we would like to project in this space (here for nodes I, J, K, and L) from the embedding of known species (here, nodes A, B, C, and D). Because embeddings can be projected back to a graph, this allows us to reconstruct a network with these new species (D). This entire cycle constitutes an instance of transfer learning, where the transferred information is the representation of graph A through its embedding.

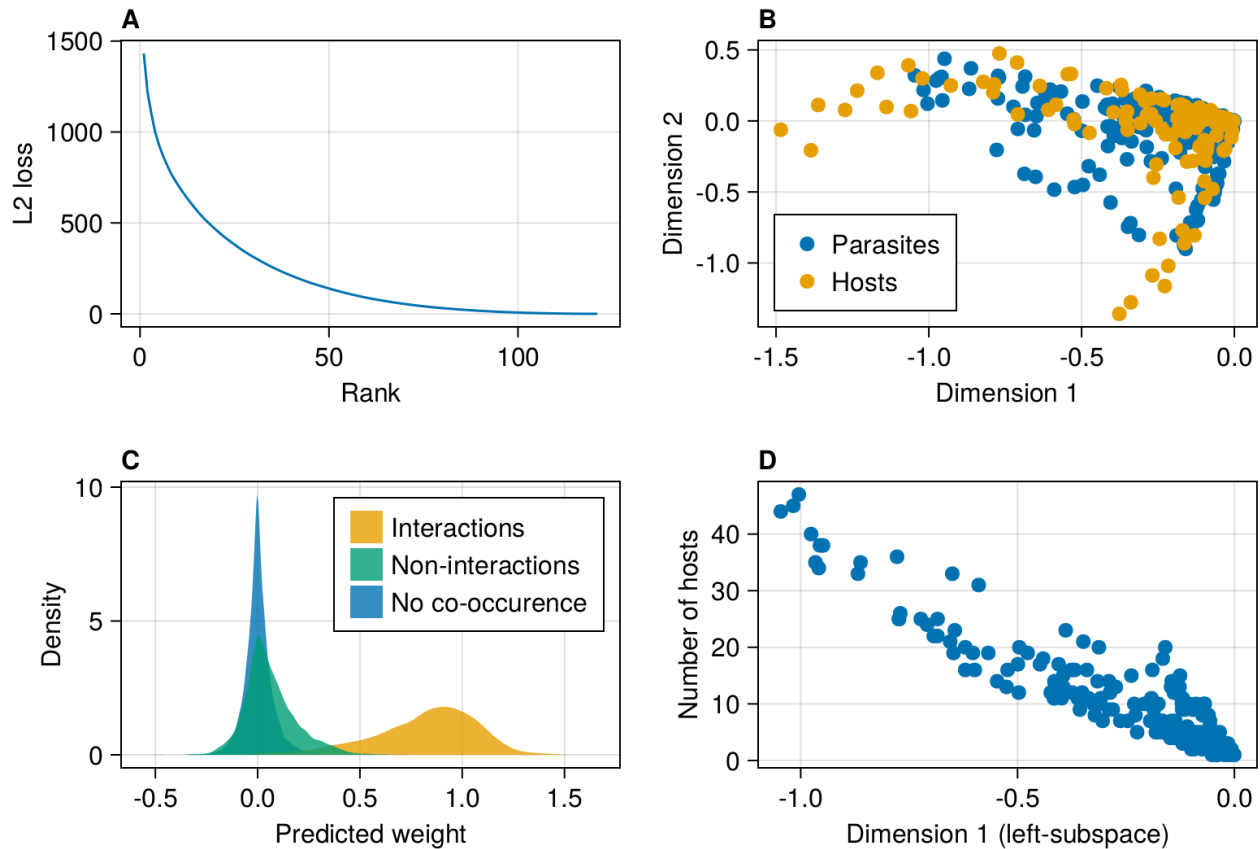


Figure 2: Illustration of an embedding for an host-parasite metaweb, using Random Dot Product Graphs. **A**, decrease in approximation error as the number of dimensions in the subspaces increases. **B**, position of hosts and parasites in the first two dimensions of their respective subspaces. **C**, predicted interaction weight from the RDPG based on the status of the species pair in the metaweb. **D**, relationship between the position on the first dimension and parasite generalism.

Table 1: Overview of some common graph embedding approaches, by time of publication, 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>: statistical interactions; <sup>b</sup>: joint-SDM-like approach.

Method	Embedding approach	Reference	Application in species interactions
tSNE	nodes through statistical divergence	Hinton & Roweis (2002)	Cieslak et al. (2020) <sup>a</sup>
RDPG	graph through SVD	Young & Scheinerman (2007)	Poisot et al. (2021); Dalla Riva & Stouffer (2016)
DeepWalk	graph walk	Perozzi et al. (2014)	Wardeh et al. (2021)
FastEmbed	graph through PCA/SVD analogue	Ramasamy & Madhow (2015)	
LINE	nodes through statistical divergence	Tang et al. (2015)	
SDNE	nodes through auto-encoding	D. Wang et al. (2016)	
node2vec	nodes embedding	Grover & Leskovec (2016)	
graph2vec	sub-graph embedding	Narayanan et al. (2017)	
DMSE	joint nodes embedding	D. Chen et al. (2017)	D. Chen et al. (2017) <sup>b</sup>
HARP	nodes through a meta-strategy	H. Chen et al. (2017)	
GraphKKE	graph embedding	Melnyk et al. (2020)	Melnyk et al. (2020) <sup>a</sup>
Joint methods	multiple graphs	S. Wang et al. (2021)	