

# The prediction of species interaction networks is facilitated by graph embedding and transfer learning 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

86 Networks (GNN; Zhou et al., 2020); GNN are, in a sense, a method to embed a graph into a dense  
87 subspace, but belong to the family of deep learning methods, which has its own set of practices (see *e.g.*  
88 Goodfellow et al., 2016). An important issue with methods based on deep learning is that because their  
89 parameter space is immense, the sample size of the data fed into them must be similarly large (typically  
90 thousands of instances). This is a requirement for the model to converge correctly during training, but this  
91 assumption is unlikely to be met given the size of datasets about currently available metawebs (or single  
92 time/location species interaction networks). This data volume requirement is mostly absent from the  
93 techniques we list below. Furthermore, GNN still have some challenges related to their shallow structure,  
94 and concerns related to scalability (see Gupta et al., 2021 for a review), which are mostly absent from the  
95 methods listed in tbl. 1. Assuming that the uptake of next-generation biomonitoring techniques does  
96 indeed deliver larger datasets on species interactions (Bohan et al., 2017), there is a potential for GNN to  
97 become an applicable embedding/predictive technique in the coming years.

98 [Table 1 about here.]

99 The popularity of graph embedding techniques in machine learning is more than the search for structural  
100 invariants: graphs are discrete objects, and machine learning techniques tend to handle continuous data  
101 better. Bringing a sparse graph into a continuous, dense vector space (Xu, 2020) opens up a broader variety  
102 of predictive algorithms, notably of the sort that are able to predict events as probabilities (Murphy, 2022).  
103 Furthermore, the projection of the graph itself is a representation that can be learned; Runghen et al.  
104 (2021), for example, used a neural network to learn the embedding of a network in which not all  
105 interactions were known, based on nodes metadata. This example has many parallels in ecology (see  
106 fig. 2), in which node metadata can be given by phylogeny or functional traits. Rather than directly  
107 predicting biological rules (see *e.g.* Pichler et al., 2020 for an overview), which may be confounded by the  
108 sparse nature of graph data, learning embeddings works in the low-dimensional space that maximizes  
109 information about the network structure. This approach is further justified by the observation, for  
110 example, that the macro-evolutionary history of a network is adequately represented by some graph  
111 embeddings (RDPG; see Dalla Riva & Stouffer, 2016). In a recent publication, Strydom, Bouskila, et al.  
112 (2021) have used an embedding (based on RDPG) to project a metaweb of trophic interactions between  
113 European mammals, and transferred this information to mammals of Canada by using the phylogenetic  
114 distance between related clades to infer the values in the latent sub-space into which the metaweb was

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

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

128 [Figure 2 about here.]

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

130 In this section, we illustrate the embedding of a collection of bipartite networks collected by **REF**, using  
131 truncated Singular Value Decomposition (t-SVD). The code to reproduce this example (as a Jupyter  
132 notebook) is available as supplementary material. The resulting (binary) metaweb  $\mathcal{M}$  has 2131 interactions  
133 between 206 parasites and 121 hosts (the same metaweb was used for the illustration of interaction  
134 inference through deep learning in Strydom, Catchen, et al., 2021), and its adjacency matrix has full rank  
135 (*i.e.* it represents a space with 121 dimensions). All analyses were done using Julia (Bezanson et al., 2017)  
136 version 1.7.2, *Makie.jl* (Danisch & Krumbiegel, 2021), and *EcologicalNetworks.jl* (Poisot et al., 2019).

137 [fig. 3](#)

138 [Figure 3 about here.]

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

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

146 The first open research problem is the taxonomic and spatial limit of the metaweb to embed and transfer.  
147 If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the chances of  
148 finding another area with enough related species (through phylogenetic relatedness or similarity of  
149 functional traits) to make a reliable inference decreases; this would likely be indicated by large confidence  
150 intervals during estimation of the values in the low-rank space, meaning that the representation of the  
151 original graph is difficult to transfer to the new problem. In addition, other problems can arise due to  
152 non-overlapping trait distributions in the known and unknown species. Alternatively a metaweb is too  
153 large (taxonomically), then the resulting embeddings would need to represent interactions between  
154 taxonomic groups that are not present in the new location. This would lead to a much higher variance in  
155 the starting dataset, and to under-dispersal in the target dataset, resulting in the potential under or over  
156 estimation of the strength of new predicted interactions. The lack of well documented metawebs is  
157 currently preventing the development of more concrete guidelines. The question of phylogenetic  
158 relatedness and dispersal is notably true if the metaweb is assembled in an area with mostly endemic  
159 species (*e.g.* a system that has undergone recent radiation and might not have an analogous system with  
160 which to draw knowledge from), and as with every predictive algorithm, there is room for the application  
161 of our best ecological judgement. Because this problem relates to dispersal of species in the geographic or  
162 phylogenetic space, it can certainly be approached through assessing the performance of embedding  
163 transfer in simulated starting/target species pools.

164 The second series of problems relate to determining which area should be used to infer the new metaweb  
165 in, as this determines the species pool that must be used. Metawebs can be constructed by assigning  
166 interactions in a list of species within geographic boundaries. The upside of this approach is that  
167 information at the country level is likely to be required for biodiversity assessments, as countries set goals



168 at the national level (Buxton et al., 2021), and as quantitative instruments are designed to work at these  
169 scales (Turak et al., 2017); specific strategies are often enacted at smaller scales, nested within a specific  
170 country (Ray et al., 2021). But there is no guarantee that these boundaries are meaningful. In fact, we do  
171 not have a satisfying answer to the question of “where does a food web stop?”; the most promising  
172 solutions involve examining the spatial consistency of network area relationships (Fortin et al., 2021; see  
173 e.g. Galiana et al., 2018, 2019, 2021), which is impossible in the absence of enough information about the  
174 network itself. This suggests that inferred metawebs should be further downscaled to allow for *a posteriori*  
175 analyses. The methodology for metaweb downscaling is currently limited, and it is likely that the  
176 sustained effort to characterize the spatial dependency of food web structure will lead to more prescriptive  
177 guidelines about the need for prediction downscaling.

178 The final family of problems relates less to ecological methods than to the praxis of ecological research.  
179 Operating under the context of national divisions, in large parts of the world, reflects nothing more than  
180 the legacy of settler colonialism, which drives a disparity in available ecological data. Applying any  
181 embedding to biased data does not debias them, but instead embeds these very same biases, propagating  
182 them to the machine learning models using embeddings to make predictions. Indeed, the use of ecological  
183 data is not an apolitical act (Nost & Goldstein, 2021), as data infrastructures tend to be designed to answer  
184 questions within national boundaries (therefore placing contingencies on what is available to be  
185 embedded), and their use often draws upon and reinforces territorial statecraft. As per Machen & Nost  
186 (2021), this is particularly true when the output of “algorithmic thinking” (e.g. relying on machine  
187 learning to generate knowledge) can be re-used for governance (e.g. enacting conservation decisions at the  
188 national scale). As information on species interaction networks structure is increasingly leveraged as a  
189 tool to guide conservation actions (see e.g. recent discussions for food-web based conservation; Eero et al.,  
190 2021; Naman et al., 2022; Stier et al., 2017), the need to appraise and correct biases that are unwittingly  
191 propagated to algorithms when embedded from the original data is paramount. Predictive approaches  
192 deployed at the continental scale, no matter their intent, originate in the framework that contributed to  
193 the ongoing biodiversity crisis (Adam, 2014) and reinforced environmental injustice (Choudry, 2013;  
194 Domínguez & Luoma, 2020). Particularly on Turtle Island and other territories that were traditionally  
195 stewarded by Indigenous people, these approaches should be replaced (or at least guided and framed) by  
196 Indigenous principles of land management (Eichhorn et al., 2019; No’kmaq et al., 2021), as part of an  
197 “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 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>

Arsov, N., & Mirceva, G. (2019). *Network Embedding: An Overview*. <http://arxiv.org/abs/1911.11726>

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. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>

Bohan, D. A., Vacher, C., Tamaddon-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., Montemaggiore, 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>

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>

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>

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

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

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>

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

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

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

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

318 Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*  
319 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.  
320 <https://doi.org/10.1145/2939672.2939754>

321 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs  
322 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,  
323 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

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

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

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

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

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

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>

Maaten, L. van der. (2009). Learning a Parametric Embedding by Preserving Local Structure. *Proceedings of the Twelfth 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>

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., 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>

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>

Stevenson, M. T., & Doleac, J. L. (2021). *Algorithmic Risk Assessment in the Hands of Humans* [SSRN Scholarly Paper]. <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>

Strydom, T., Bouskila, S., Banville, F., Barros, C., Caron, D., Farrell, M. J., Fortin, M.-J., Hemming, V., Mercier, B., Pollock, L., Runghen, R., Riva, G. V. D., & Poisot, T. (2021). *Food web reconstruction through phylogenetic transfer of low-rank network representation*. <https://doi.org/10.32942/osf.io/y7sdz>

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.

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

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

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

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

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

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

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

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

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

- 448 Xu, M. (2020). *Understanding graph embedding methods and their applications*.  
449 <http://arxiv.org/abs/2012.08019>
- 450 Yan, S., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A general framework for  
451 dimensionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern*  
452 *Recognition (CVPR'05)*, 2, 830–837 vol. 2. <https://doi.org/10.1109/CVPR.2005.170>
- 453 Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In A.  
454 Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.  
455 [https://doi.org/10.1007/978-3-540-77004-6\\_11](https://doi.org/10.1007/978-3-540-77004-6_11)
- 456 Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural  
457 networks: A review of methods and applications. *AI Open*, 1, 57–81.  
458 <https://doi.org/10.1016/j.aiopen.2021.01.001>

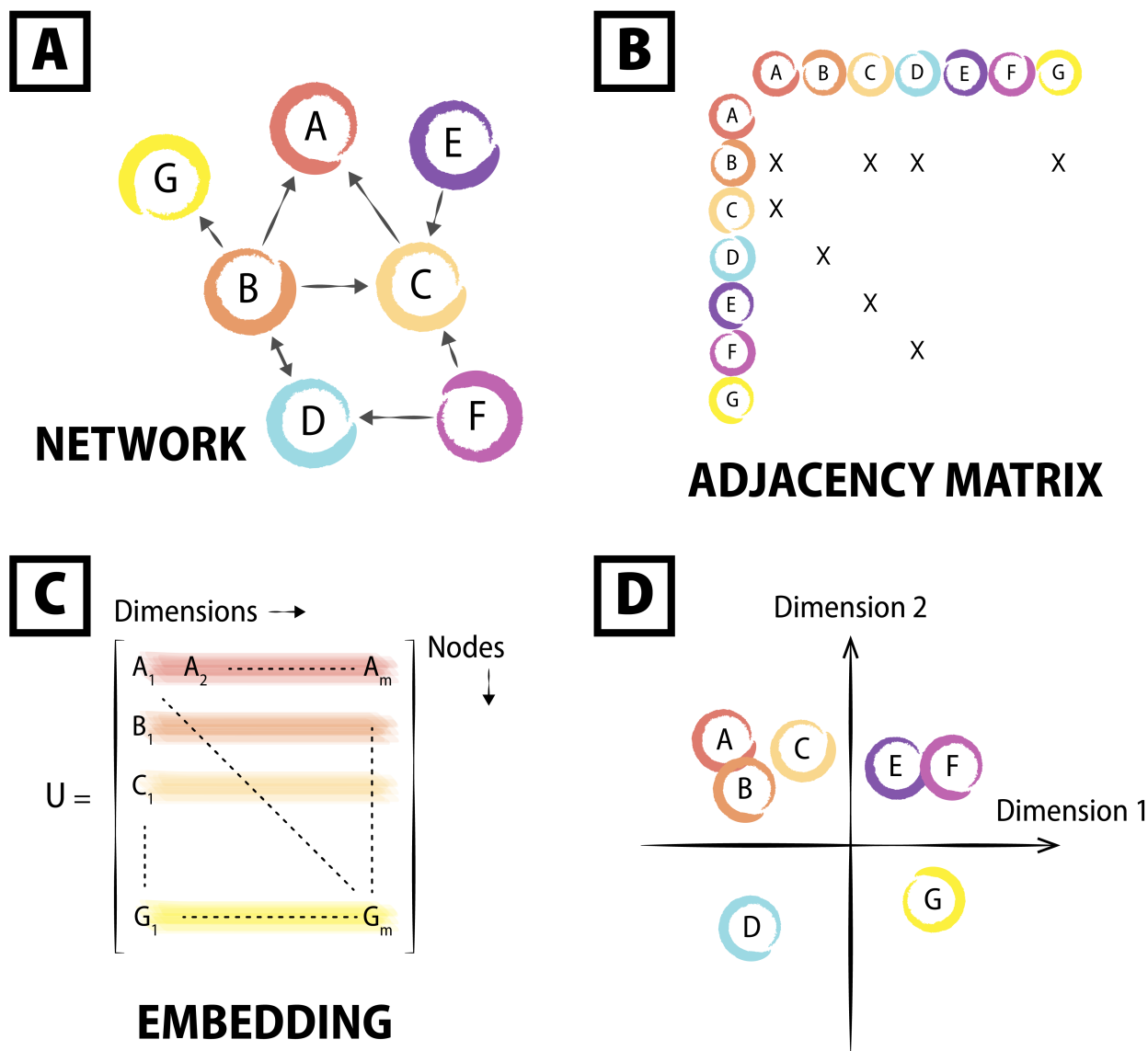


Figure 1: Overview of the embedding process. A network (A), represented as its adjacency matrix (B), is converted into a lower-dimensional object (C) where nodes, subgraphs, or edges have specific values (see tbl. 1). For the purposes of prediction, this low-dimensional object encodes feature vectors for *e.g.* the nodes. Embedding also allows to visualize the structure in the data differently (D), much like with a principal component analysis.

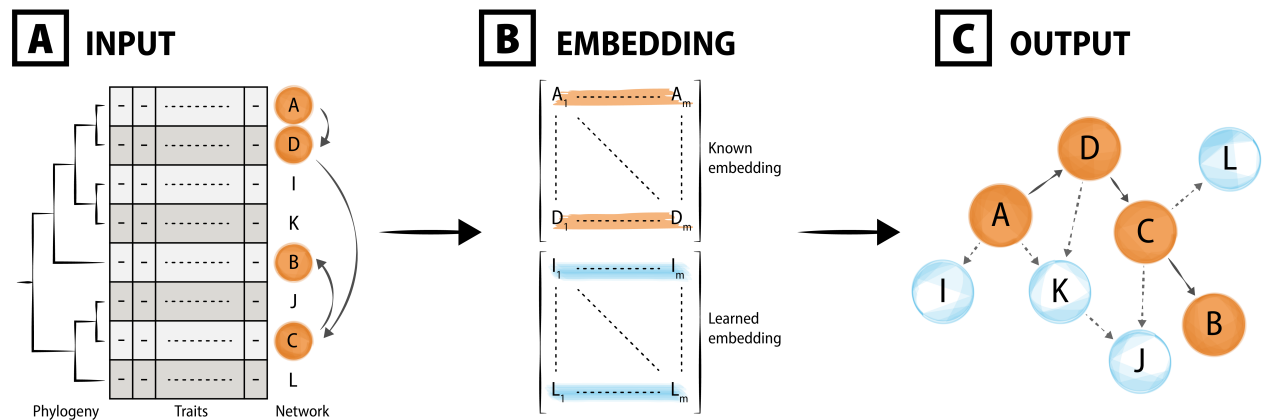


Figure 2: From a low-dimensional feature vector (see fig. 1), it is possible to develop predictive approaches. Nodes in an ecological network are species, for which we can leverage phylogenetic relatedness (*e.g.* Strydom, Bouskila, et al., 2021) or functional traits to fill the values of additional species we would like to project in this space (here, I, J, K, and L) from the embedding of known species (here, 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. This approach constitutes an instance of transfer learning.

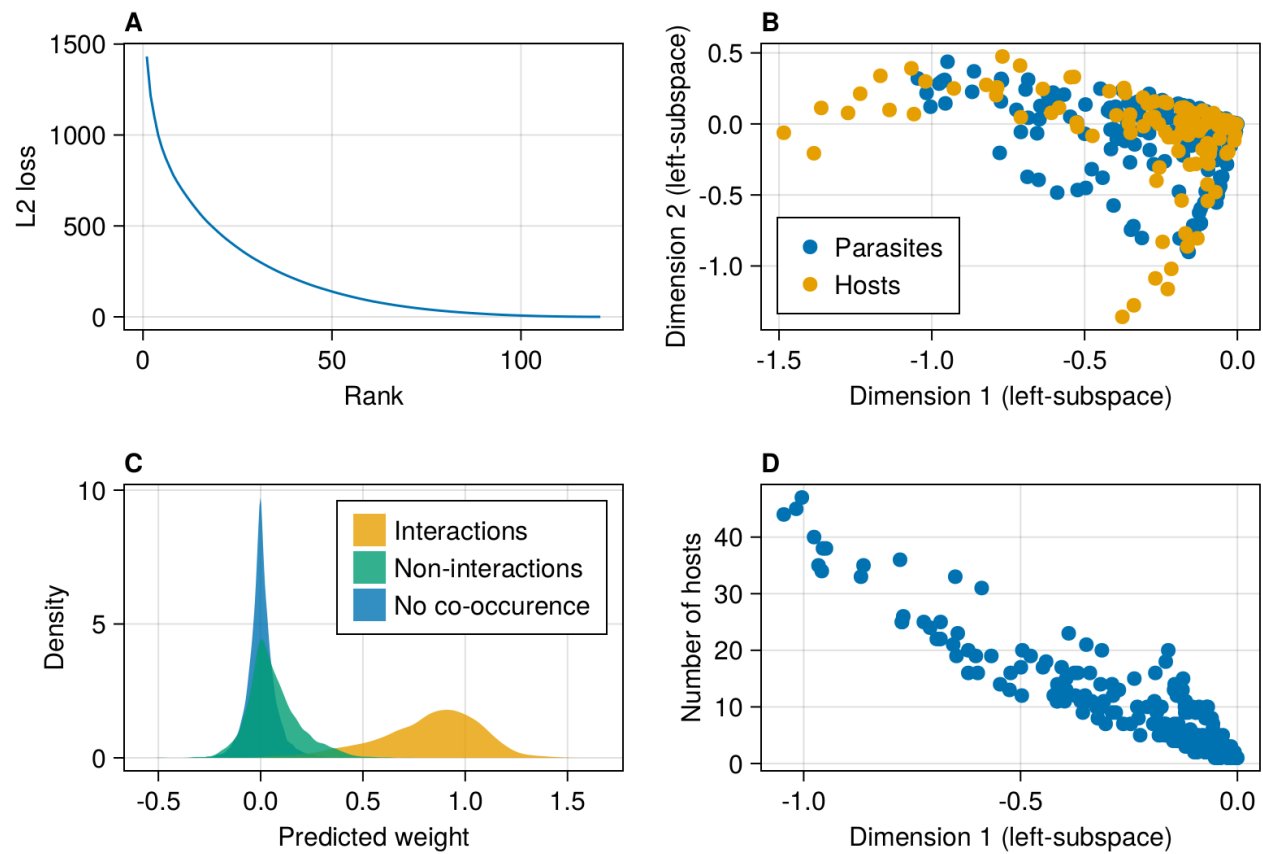


Figure 3: Need to add text here

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)	