

# Predicting metawebs: transfer of graph embeddings can help alleviate spatial data deficiencies

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 to locations where inference is perhaps the least required (and still leaves us  
9 unable to address regions without data); second, accurate predictions often demand accurate predictors,  
10 and the lack of methods that can leverage small amount of data is a serious impediment to our global  
11 predictive ability. In most places, our most reliable biodiversity knowledge is that of a species pool:  
12 through the analysis of databases like GBIF or IUCN, it is possible to establish a list of species in a region  
13 of interest; but establishing the interactions between these species is difficult.

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

27 Because the metaweb represents the joint effect of functional, phylogenetic, and macroecological  
28 processes (Morales-Castilla et al., 2015), it holds valuable ecological information. Specifically, it is the  
29 “upper bounds” on what the composition of the local networks, given the local species pool, can be (see  
30 e.g. McLeod et al., 2021); this information can help evaluate the ability of ecological assemblages to

31 withstand the effects of, for example, climate change (Fricke et al., 2022). These local networks may be  
32 reconstructed given an appropriate knowledge of local species composition and provide information on  
33 the structure of food webs at finer spatial scales. This has been done for example for tree-galler-parasitoid  
34 systems (Gravel et al., 2018), fish trophic interactions (Albouy et al., 2019), tetrapod trophic interactions  
35 (O'Connor et al., 2020; **Braga2019SpaAna?**), and crop-pest networks (Grünig et al., 2020). In this  
36 contribution, we highlight the power in viewing (and constructing) metawebs as *probabilistic* objects in  
37 the context of rare interactions, discuss how a family of machine learning tools (graph embeddings and  
38 transfer learning) can be used to overcome data limitations to metaweb inference, and highlight how the  
39 use of metawebs introduces important questions for the field of network ecology.

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

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

57 Yet it is important to recall that a metaweb is intended as a catalogue of all potential interactions, which is  
58 then filtered (Morales-Castilla et al., 2015). In a sense, that most ecological interactions are elusive can call

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

[Figure 1 about here.]

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

Graph embedding (fig. 1) is a varied family of machine learning techniques aiming to transform nodes and edges into a vector space (Arsov & Mirceva, 2019), usually of a lower dimension, whilst maximally retaining key properties of the graph (Yan et al., 2005). Ecological networks are an interesting candidate for the widespread application of embeddings, as they tend to possess a shared structural backbone (Mora et al., 2018), which hints at structural invariants that can be revealed at lower dimensions. Indeed, food webs are inherently low-dimensional objects, and can be adequately represented with less than ten dimensions (Eklöf et al., 2013; Braga2019SpaAna?). Simulation results by Botella et al. (2022) suggest that there is no best method to identify architectural similarities between networks, and that multiple approaches need to be tested and compared to the network descriptor of interest. This matches previous, more general results on graph embedding, which suggest that the choice of embedding algorithm matters for the results (Goyal & Ferrara, 2018). In tbl. 1, we present a selection of common graph embedding methods, alongside examples of their use to predict species interactions.

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)
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)	

83 The popularity of graph embedding techniques in machine learning is more prosaic than the search for  
84 structural invariants: graphs are discrete objects, and machine learning techniques tend to handle  
85 continuous data better. Bringing a sparse graph into a continuous, dense vector space (Xu, 2020) opens up  
86 a broader variety of predictive algorithms, notably of the sort that are able to predict events as probabilities  
87 (Murphy, 2022). Furthermore, the projection of the graph itself is a representation that can be learned;  
88 Runghen et al. (2021), for example, used a neural network to learn the embedding of a network in which  
89 not all interactions were known, based on nodes metadata. This example has many parallels in ecology  
90 (see fig. 2), in which node metadata can be given by phylogeny or functional traits. Rather than directly  
91 predicting biological rules (see *e.g.* Pichler et al., 2020 for an overview), which may be confounded by the  
92 sparse nature of graph data, learning embeddings works in the low-dimensional space that maximizes  
93 information about the network structure. This approach is further justified by the observation, for  
94 example, that the macro-evolutionary history of a network is adequately represented by some graph  
95 embeddings (RDPG; see Dalla Riva & Stouffer, 2016).

96 [Figure 2 about here.]

## 97 **The metaweb embeds ecological hypotheses and practices**

98 The goal of metaweb inference is to provide information about the interactions between species at a large  
99 spatial scale. But as Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide  
100 problems”; any inference of a metaweb at large scales must contend with several novel, and interwoven,  
101 families of problems.

102 The first is the taxonomic and spatial limit of the metaweb to embed and transfer. If the initial metaweb is  
103 too narrow in scope, notably from a taxonomic point of view, the chances of finding another area with  
104 enough related species (through phylogenetic relatedness or similarity of functional traits) to make a  
105 reliable inference decreases; this would likely be indicated by large confidence intervals during estimation  
106 of the values in the low-rank space, or by non-overlapping trait distributions in the known and unknown  
107 species. The lack of well documented metawebs is currently preventing the development of more concrete  
108 guidelines. The question of phylogenetic relatedness and dispersal is notably true if the metaweb is  
109 assembled in an area with mostly endemic species (*i.e.* potentially limited phylogenetic/species overlap),  
110 and as with every predictive algorithm, there is room for the application of our best ecological judgement.

111 The second series of problems relate to determining which area should be used to infer the new metaweb  
112 in, as this determines the species pool that must be used. Metawebs can be constructed by assigning  
113 interactions in a list of species within geographic boundaries. The upside of this approach is that  
114 information at the country level is likely to be required for biodiversity assessments, as countries set goals  
115 at the national level (Buxton et al., 2021), and as quantitative instruments are designed to work at these  
116 scales (Turak et al., 2017); specific strategies are often enacted at smaller scales, nested within a specific  
117 country (Ray et al., 2021). But there is no guarantee that these boundaries are meaningful. In fact, we do  
118 not have a satisfying answer to the question of “where does a food web stop?”; the most promising  
119 solutions involve examining the spatial consistency of network area relationships (Fortin et al., 2021; see  
120 e.g. Galiana et al., 2018, 2019, 2021), which is impossible in the absence of enough information about the  
121 network itself. This suggests that inferred metawebs should be further downscaled to allow for *a posteriori*  
122 analyses.

123 The final family of problems relates less to ecological concepts and more to the praxis of ecological  
124 research. Operating under the context of national divisions, in large parts of the world, reflects nothing  
125 more than the legacy of settler colonialism. Indeed, the use of ecological data is not an apolitical act (Nost  
126 & Goldstein, 2021), as data infrastructures tend to be designed to answer questions within national  
127 boundaries, and their use often draws upon and reinforces territorial statecraft. As per Machen & Nost  
128 (2021), this is particularly true when the output of “algorithmic thinking” (e.g. relying on machine  
129 learning to generate knowledge) can be re-used for governance (e.g. enacting conservation decisions at the  
130 national scale). We therefore recognize that predictive approaches deployed at the continental scale, no  
131 matter their intent, originate in the framework that contributed to the ongoing biodiversity crisis (Adam,  
132 2014), reinforced environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020), and e.g. as on  
133 Turtle Island, should be replaced by Indigenous principles of land management (Eichhorn et al., 2019;  
134 No’kmaq et al., 2021). As we see artificial intelligence/machine learning being increasingly mobilized to  
135 generate knowledge that is lacking for conservation decisions (e.g. Lamba et al., 2019; Mosebo Fernandes  
136 et al., 2020), our discussion of these tools need to go beyond the technical, and into the governance  
137 consequences they can have.

138 **Acknowledgements:** We acknowledge that this study was conducted on land within the traditional  
139 unceded territory of the Saint Lawrence Iroquoian, Anishinabewaki, Mohawk, Huron-Wendat, and  
140 Omàmiwininiwak nations. TP, TS, DC, and LP received funding from the Canadian Institute for Ecology



141 & Evolution. FB is funded by the Institute for Data Valorization (IVADO). TS, SB, and TP are funded by a  
142 donation from the Courtois Foundation. CB was awarded a Mitacs Elevate Fellowship no. IT12391, in  
143 partnership with fRI Research, and also acknowledges funding from Alberta Innovates and the Forest  
144 Resources Improvement Association of Alberta. M-JF acknowledges funding from NSERC Discovery  
145 Grant and NSERC CRC. RR is funded by New Zealand's Biological Heritage Ngā Koiora Tuku Iho  
146 National Science Challenge, administered by New Zealand Ministry of Business, Innovation, and  
147 Employment. BM is funded by the NSERC Alexander Graham Bell Canada Graduate Scholarship and the  
148 FRQNT master's scholarship. LP acknowledges funding from NSERC Discovery Grant (NSERC  
149 RGPIN-2019-05771). TP acknowledges financial support from NSERC through the Discovery Grants and  
150 Discovery Accelerator Supplement programs.

## 151 References

- 152 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 153 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,  
154 Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel, D.  
155 (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8, 8),  
156 1153–1161. <https://doi.org/10.1038/s41559-019-0950-y>
- 157 Arsov, N., & Mirceva, G. (2019, November 26). *Network Embedding: An Overview*.  
158 <http://arxiv.org/abs/1911.11726>
- 159 Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A  
160 common framework for identifying linkage rules across different types of interactions. *Functional*  
161 *Ecology*, 30(12), 1894–1903.  
162 <http://onlinelibrary.wiley.com/doi/10.1111/1365-2435.12666/full>
- 163 Botella, C., Dray, S., Matias, C., Miele, V., & Thuiller, W. (2022). An appraisal of graph embeddings for  
164 comparing trophic network architectures. *Methods in Ecology and Evolution*, 13(1), 203–216.  
165 <https://doi.org/10.1111/2041-210X.13738>
- 166 Buxton, R. T., Bennett, J. R., Reid, A. J., Shulman, C., Cooke, S. J., Francis, C. M., Nyboer, E. A., Pritchard,  
167 G., Binley, A. D., Avery-Gomm, S., Ban, N. C., Beazley, K. F., Bennett, E., Blight, L. K., Bortolotti, L. E.,

168 Camfield, A. F., Gadallah, F., Jacob, A. L., Naujokaitis-Lewis, I., ... Smith, P. A. (2021). Key  
 169 information needs to move from knowledge to action for biodiversity conservation in Canada.  
 170 *Biological Conservation*, 256, 108983. <https://doi.org/10.1016/j.biocon.2021.108983>

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

173 Chen, H., Perozzi, B., Hu, Y., & Skiena, S. (2017, November 16). *HARP: Hierarchical Representation*  
 174 *Learning for Networks*. <http://arxiv.org/abs/1706.07845>

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

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

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

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

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

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

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

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

194 Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti,  
 195 M., Guimarães, P. R., Lomáscolo, S. B., Martín González, A. M., Pizo, M. A., Rader, R., Rodrigo, A.,

196 Tylanakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks.  
 197 *Ecology Letters*, 16(5), 577–583. <https://doi.org/10.1111/ele.12081>

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

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

204 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.  
 205 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,  
 206 *n/a(n/a)*. <https://doi.org/10.1111/ecog.05229>

207 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is  
 208 scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*,  
 209 42(6), 1175–1187. <https://doi.org/10.1111/ecog.03684>

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

215 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,  
 216 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),  
 217 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

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

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

224 Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning for Networks. *Proceedings of the*  
225 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.  
226 <https://doi.org/10.1145/2939672.2939754>

227 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs  
228 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,  
229 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

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

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

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

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

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

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

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

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

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

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

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

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

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

262 Narayanan, A., Chandramohan, M., Venkatesan, R., Chen, L., Liu, Y., & Jaiswal, S. (2017, July 17).  
 263 *Graph2vec: Learning Distributed Representations of Graphs*. <http://arxiv.org/abs/1707.05005>

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

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

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

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

275 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,  
 276 Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the  
 277 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*, 4(7),  
 278 170215. <https://doi.org/10.1098/rsos.170215>

279 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., 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, May 31). *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>

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*

308 phylogenetic transfer of low-rank network representation. <https://doi.org/10.32942/osf.io/y7sdz>

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

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

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

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

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

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

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

334 Xu, M. (2020, December 14). *Understanding graph embedding methods and their applications*.  
 335 <http://arxiv.org/abs/2012.08019>

- 336 Yan, S., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A general framework for  
337 dimensionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern*  
338 *Recognition (CVPR'05)*, 2, 830–837 vol. 2. <https://doi.org/10.1109/CVPR.2005.170>
- 339 Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In A.  
340 Bonato & F. R. K. Chung (Eds.), *Algorithms and Models for the Web-Graph* (pp. 138–149). Springer.  
341 [https://doi.org/10.1007/978-3-540-77004-6\\_11](https://doi.org/10.1007/978-3-540-77004-6_11)



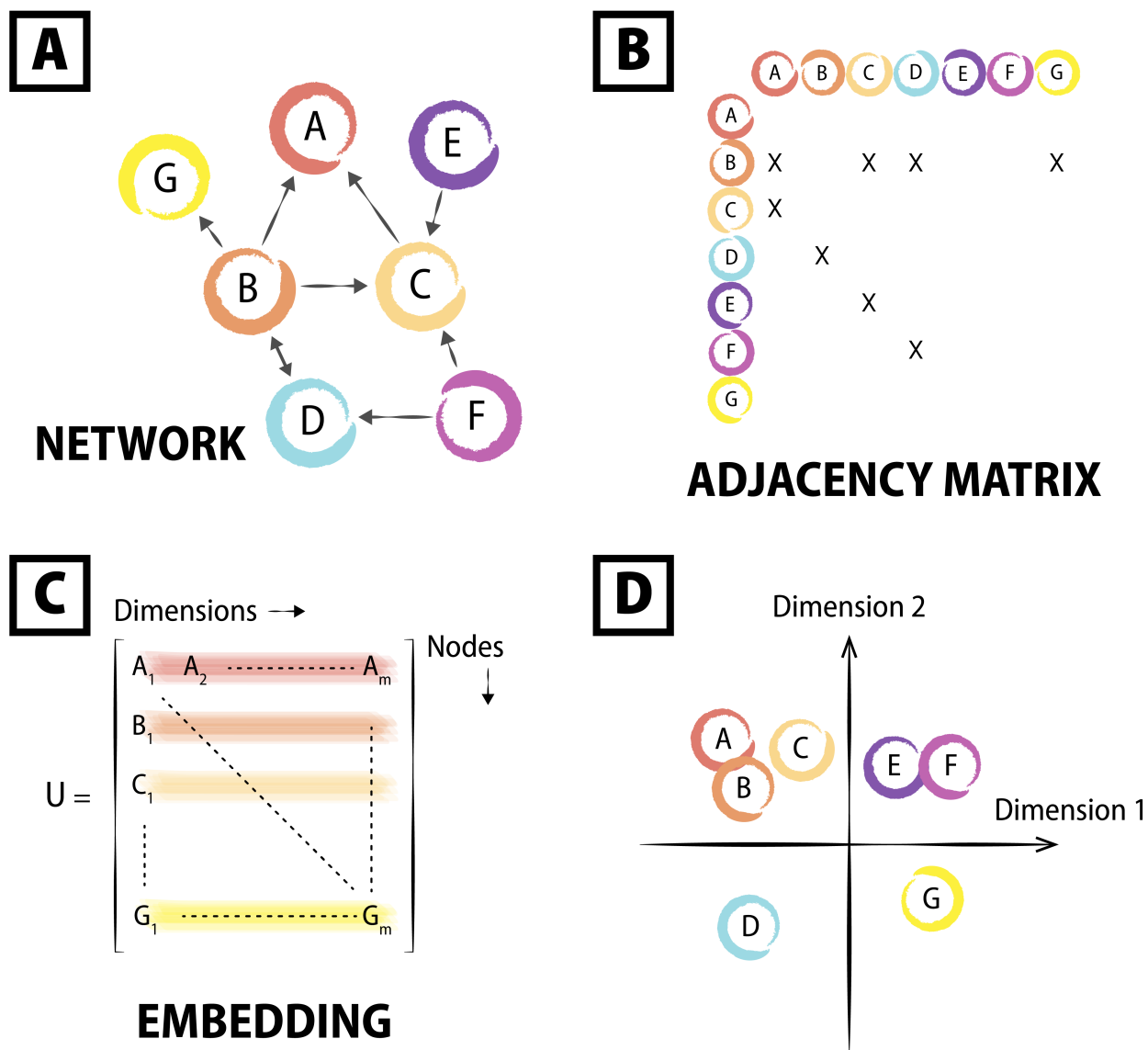


Figure 1: Overview of the embedding process. A network (A), represented as its adjacency matrix (B), is converted into a lower-dimensional object (C) where nodes, subgraphs, or edges have specific values (see tbl. 1). For the purposes of prediction, this low-dimensional object encodes feature vectors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

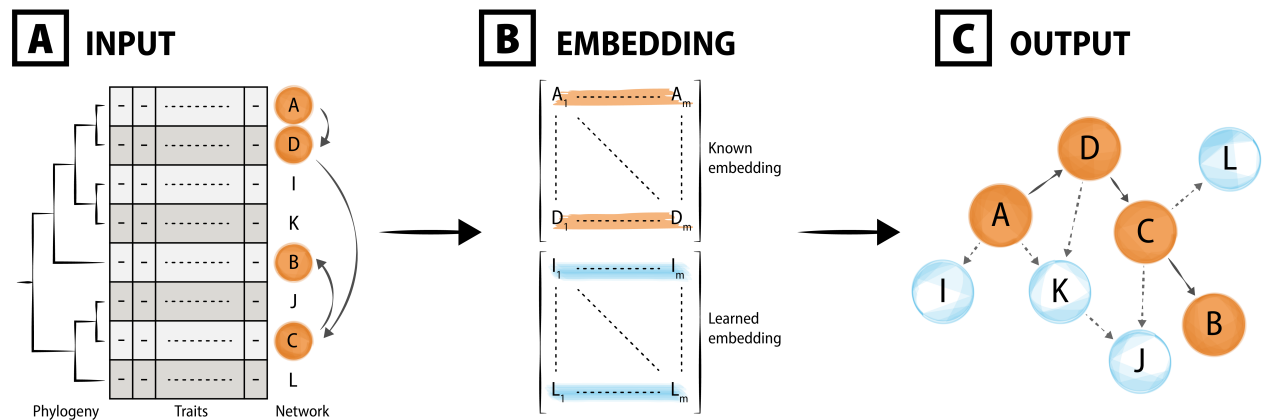


Figure 2: From a low-dimensional feature vector (see fig. 1), it is possible to develop predictive approaches. Nodes in an ecological network are species, for which we can leverage phylogenetic relatedness (*e.g.* Strydom, Bouskila, et al., 2021) or functional traits to fill the values of additional 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.