

Predicting metawebs: graph embeddings can help alleviate spatial data deficiencies

Tanya Strydom^{1,2,‡} Timothée Poisot^{1,2,‡}

¹ Département de Sciences Biologiques, Université de Montréal, Montréal, Canada ² Quebec Centre for Biodiversity Science, Montréal, Canada

[‡] These authors contributed equally to the work

Correspondance to:

Timothée Poisot — 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 Having a general solution for inferring *potential* interactions (despite the unavailability of interaction data)
2 could be the catalyst for significant breakthroughs in our ability to start thinking about species interaction
3 networks over large spatial scales. In a recent overview of the field of ecological network prediction,
4 Strydom et al. (2021) identified two challenges of interest to the prediction of interactions at large scales.
5 First, there is a relative scarcity of relevant data in most places globally – paradoxically, this restricts our
6 ability to infer interactions to locations where inference is perhaps the least required; second, accurate
7 predictions often demand accurate predictors, and the lack of methods that can leverage small amount of
8 data is a serious impediment to our predictive ability globally.

9 Following the definition of Dunne (2006), a metaweb is a network analogue to the regional species pool;
10 specifically, it is an inventory of all *potential* interactions between species from a spatially delimited area
11 (and so captures the γ diversity of interactions). The metaweb is, therefore, *not* a prediction of the food
12 web at a specific locale within the spatial area it covers, and will have a different structure (notably by
13 having a larger connectance; see e.g. Wood et al., 2015). These local food webs (which captures the α
14 diversity of interactions) are a subset of the metaweb’s species and interactions, and have been called
15 “metaweb realizations” (Poisot et al., 2015). Differences between local food web and their metaweb are
16 due to chance, species abundance and co-occurrence, local environmental conditions, and local
17 distribution of functional traits, among others.

18 Because the metaweb represents the joint effect of functional, phylogenetic, and macroecological
19 processes (Morales-Castilla et al., 2015), it holds valuable ecological information. Specifically, it is the
20 “upper bounds” on what the composition of the local networks can be (see e.g. McLeod et al., 2021). These
21 local networks, in turn, can be reconstructed given appropriate knowledge of local species composition,
22 providing information on structure of food webs at finer spatial scales. This has been done for example for
23 tree-galler-parasitoid systems (Gravel et al., 2018), fish trophic interactions (Albouy et al., 2019), tetrapod
24 trophic interactions (O’Connor et al., 2020), and crop-pest networks (Grünig et al., 2020). Whereas the
25 original metaweb definition, and indeed most past uses of metawebs, was based on the presence/absence
26 of interactions, we focus on *probabilistic* metawebs where interactions are represented as the chance of
27 success of a Bernoulli trial (see e.g. Poisot et al., 2016); therefore, not only does our method recommend
28 interactions that may exist, it gives each interaction a score, allowing us to properly weigh them.

29 **The metaweb is an inherently probabilistic object**

30 Yet, owing to the inherent plasticity of interactions, there have been documented instances of food webs
31 undergoing rapid collapse/recovery cycles over short periods of time (Pedersen et al., 2017). The
32 embedding of a network, in a sense, embeds its macro-evolutionary history, especially as RDPG captures
33 ecological signal (Dalla Riva & Stouffer, 2016); at this point, it is important to recall that a metaweb is
34 intended as a catalogue of all potential interactions, which should then be filtered (Morales-Castilla et al.,
35 2015). In practice (and in this instance) the reconstructed metaweb will predict interactions that are
36 plausible based on the species' evolutionary history, however some interactions would/would not be
37 realized due to human impact.

38 Dallas et al. (2017) suggested that most links in ecological networks may be cryptic, *i.e.* uncommon or
39 otherwise hard to observe. This argument essentially echoes Jordano (2016): the sampling of ecological
40 interactions is difficult because it requires first the joint observation of two species, and then the
41 observation of their interaction. In addition, it is generally expected that weak or rare links would be more
42 common in networks (Csermely, 2004), compared to strong, persistent links; this is notably the case in
43 food chains, wherein many weaker links are key to the stability of a system (Neutel et al., 2002). In the
44 light of these observations, the results in fig. ?? are not particularly surprising: we expect to see a surge in
45 these low-probability interactions under a model that has a good predictive accuracy. Because the
46 predictions we generate are by design probabilistic, then one can weigh these rare links appropriately. In a
47 sense, that most ecological interactions are elusive can call for a slightly different approach to sampling:
48 once the common interactions are documented, the effort required in documenting each rare interaction
49 may increase exponentially. Recent proposals suggest that machine learning algorithms, in these
50 situations, can act as data generators (Hoffmann et al., 2019): in this perspective, high quality
51 observational data can be supplemented with synthetic data coming from predictive models, which
52 increases the volume of information available for inference. Indeed, Strydom et al. (2021) suggested that
53 knowing the metaweb may render the prediction of local networks easier, because it fixes an “upper
54 bound” on which interactions can exist; indeed, with a probabilistic metaweb, we can consider that the
55 metaweb represents an aggregation of informative priors on the interactions.

56 **Graph embedding offers promises the inference of potential interactions**

57 Graph embedding is a varied family of machine learning techniques aiming to transform nodes and edges
58 into a vector space, usually of a lower dimension, whilst maximally retaining key properties of the graph.
59 Ecological networks are an interesting candidate for the widespread application of embeddings, as they
60 tend to possess a shared structural backbone (Mora et al., 2018), which hints at structural invariants that
61 can be revealed at lower dimensions. Indeed, previous work by (Eklof2013DimEco?) suggests that food
62 webs are inherently low-dimensional objects, and can be adequately represented with less than ten
63 dimensions.

64 The popularity of graph embedding techniques in machine learning is rather intuitive: while graphs are
65 discrete objects, machine learning techniques tend to handle continuous data better. Therefore, bringing a
66 discrete graph into a continuous vector space opens up a broader variety of predictive algorithms.

Method	Embedded object	Reference
DeepWalk	graph walk	Perozzi et al. (2014)
node2vec	node embedding	Grover & Leskovec (2016)
graph2vec	sub-graph embedding	Narayanan et al. (2017)

67 **The metaweb embeds hypotheses about which spatial boundaries are**
68 **meaningful**

69 As Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”; in
70 this regard, our approach (and indeed, any inference of a metaweb at large scales) must contend with
71 several interesting and interwoven families of problems. The first is the limit of the metaweb to embed
72 and transfer. If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the
73 chances of finding another area with enough related species to make a reliable inference decreases; this
74 would likely be indicated by large confidence intervals during ancestral character estimation, but the lack
75 of well documented metawebs is currently preventing the development of more concrete guidelines. The
76 question of phylogenetic relatedness and dispersal is notably true if the metaweb is assembled in an area
77 with mostly endemic species, and as with every predictive algorithm, there is room for the application of

78 our best ecological judgement. Conversely, the metaweb should be reliably filled, which assumes that the
 79 S^2 interactions in a pool of S species have been examined, either through literature surveys or expert
 80 elicitation. Supp. Mat. 1 provides some guidance as to the type of sampling effort that should be
 81 prioritized. While RDPG was able to maintain very high predictive power when interactions were missing,
 82 the addition of false positive interactions was immediately detected; this suggests that it may be
 83 appropriate to err on the side of “too many” interactions when constructing the initial metaweb to be
 84 transferred. The second series of problems are related to determining which area should be used to infer
 85 the new metaweb in, as this determines the species pool that must be used.

86 In our application, we focused on the mammals of Canada. The upside of this approach is that
 87 information at the country level is likely to be required by policy makers and stakeholders for their
 88 biodiversity assessment, as each country tends to set goals at the national level (Buxton et al., 2021) for
 89 which quantitative instruments are designed (Turak et al., 2017), with specific strategies often enacted at
 90 smaller scales (Ray et al., 2021). And yet, we do not really have a satisfying answer to the question of
 91 “where does a food web stop?”; the current most satisfying solutions involve examining the spatial
 92 consistency of network area relationships (see *e.g.* Galiana et al., 2018, 2019, 2021; **Fortin2021NetEco?**),
 93 which is of course impossible in the absence of enough information about the network itself. This
 94 suggests that an *a posteriori* refinement of the results may be required, based on a downscaling of the
 95 metaweb. The final family of problems relates less to the availability of data or quantitative tools, and
 96 more to the praxis of spatial ecology. Operating under the context of national divisions, in large parts of
 97 the world, reflects nothing more than the legacy of settler colonialism. Indeed, the use of ecological data is
 98 not an apolitical act (Nost & Goldstein, 2021), as data infrastructures tend to be designed to answer
 99 questions within national boundaries, and their use both draws upon and reinforces territorial statecraft;
 100 as per Machen & Nost (2021), this is particularly true when the output of “algorithmic thinking” (*e.g.*
 101 relying on machine learning to generate knowledge) can be re-used for governance (*e.g.* enacting
 102 conservation decisions at the national scale). We therefore recognize that methods such as we propose
 103 operate under the framework that contributed to the ongoing biodiversity crisis (Adam, 2014), reinforced
 104 environmental injustice (Choudry, 2013; Domínguez & Luoma, 2020), and on Turtle Island especially,
 105 should be replaced by Indigenous principles of land management (Eichhorn et al., 2019; No’kmaq et al.,
 106 2021). As we see AI/ML being increasingly mobilized to generate knowledge that is lacking for
 107 conservation decisions (*e.g.* Lamba et al., 2019; Mosebo Fernandes et al., 2020), our discussion of these

108 tools need to go beyond the technical, and into the governance consequences they can have.

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

122 References

- 123 Adam, R. (2014). *Elephant treaties: The Colonial legacy of the biodiversity crisis*. UPNE.
- 124 Albouy, C., Archambault, P., Appeltans, W., Araújo, M. B., Beauchesne, D., Cazelles, K., Cirtwill, A. R.,
125 Fortin, M.-J., Galiana, N., Leroux, S. J., Pellissier, L., Poisot, T., Stouffer, D. B., Wood, S. A., & Gravel, D.
126 (2019). The marine fish food web is globally connected. *Nature Ecology & Evolution*, 3(8, 8),
127 1153–1161. <https://doi.org/10.1038/s41559-019-0950-y>
- 128 Buxton, R. T., Bennett, J. R., Reid, A. J., Shulman, C., Cooke, S. J., Francis, C. M., Nyboer, E. A., Pritchard,
129 G., Binley, A. D., Avery-Gomm, S., Ban, N. C., Beazley, K. F., Bennett, E., Blight, L. K., Bortolotti, L. E.,
130 Camfield, A. F., Gadallah, F., Jacob, A. L., Naujokaitis-Lewis, I., ... Smith, P. A. (2021). Key
131 information needs to move from knowledge to action for biodiversity conservation in Canada.
132 *Biological Conservation*, 256, 108983. <https://doi.org/10.1016/j.biocon.2021.108983>
- 133 Choudry, A. (2013). Saving biodiversity, for whom and for what? Conservation NGOs, complicity,
134 colonialism and conquest in an era of capitalist globalization. In *NGOization: Complicity*,

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

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

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

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

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

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

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

149 Galiana, N., Barros, C., Braga, J., Ficetola, G. F., Maiorano, L., Thuiller, W., Montoya, J. M., & Lurgi, M.
 150 (2021). The spatial scaling of food web structure across European biogeographical regions. *Ecography*,
 151 *n/a(n/a)*. <https://doi.org/10.1111/ecog.05229>

152 Galiana, N., Hawkins, B. A., & Montoya, J. M. (2019). The geographical variation of network structure is
 153 scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*,
 154 42(6), 1175–1187. <https://doi.org/10.1111/ecog.03684>

155 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., Gravel, D., & Montoya,
 156 J. M. (2018). The spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2(5),
 157 782–790. <https://doi.org/10.1038/s41559-018-0517-3>

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

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

163 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 855–864.
 164 <https://doi.org/10.1145/2939672.2939754>

165 Grünig, M., Mazzi, D., Calanca, P., Karger, D. N., & Pellissier, L. (2020). Crop and forest pest metawebs
 166 shift towards increased linkage and suitability overlap under climate change. *Communications Biology*,
 167 3(1, 1), 1–10. <https://doi.org/10.1038/s42003-020-0962-9>

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

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

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

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

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

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

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

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

189 Mosebo Fernandes, A. C., Quintero Gonzalez, R., Lenihan-Clarke, M. A., Leslie Trotter, E. F., & Jokar
 190 Arsanjani, J. (2020). Machine Learning for Conservation Planning in a Changing Climate.

191 Sustainability, 12(18, 18), 7657. <https://doi.org/10.3390/su12187657>

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

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

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

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

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

205 Pedersen, E. J., Thompson, P. L., Ball, R. A., Fortin, M.-J., Gouhier, T. C., Link, H., Moritz, C., Nenzen, H.,
 206 Stanley, R. R. E., Taranu, Z. E., Gonzalez, A., Guichard, F., & Pepin, P. (2017). Signatures of the
 207 collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*, 4(7),
 208 170215. <https://doi.org/10.1098/rsos.170215>

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

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

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

217 Ray, J. C., Grimm, J., & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of
 218 federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068.

219 <https://doi.org/10.1139/facets-2020-0075>

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

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

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