

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

, i.e. networks of potential interactions within a species pool, are a powerful abstraction to understand how large-scales species interaction networks are structured. 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. 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. 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 due
16 to chance, species abundance and co-occurrence, local environmental conditions, and local distribution of
17 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 Yet, owing to the inherent plasticity of interactions, there have been documented instances of food webs
30 undergoing rapid collapse/recovery cycles over short periods of time (Pedersen *et al.* 2017). The

31 embedding of a network, in a sense, embeds its macro-evolutionary history, especially as RDPG captures
32 ecological signal (Dalla Riva & Stouffer 2016); at this point, it is important to recall that a metaweb is
33 intended as a catalogue of all potential interactions, which should then be filtered (Morales-Castilla *et al.*
34 2015). In practice (and in this instance) the reconstructed metaweb will predict interactions that are
35 plausible based on the species' evolutionary history, however some interactions would/would not be
36 realized due to human impact.

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

55 As Herbert (1965) rightfully pointed out, “[y]ou can’t draw neat lines around planet-wide problems”; in
56 this regard, our approach (and indeed, any inference of a metaweb at large scales) must contend with
57 several interesting and interwoven families of problems. The first is the limit of the metaweb to embed
58 and transfer. If the initial metaweb is too narrow in scope, notably from a taxonomic point of view, the
59 chances of finding another area with enough related species to make a reliable inference decreases; this
60 would likely be indicated by large confidence intervals during ancestral character estimation, but the lack

61 of well documented metawebs is currently preventing the development of more concrete guidelines. The
62 question of phylogenetic relatedness and dispersal is notably true if the metaweb is assembled in an area
63 with mostly endemic species, and as with every predictive algorithm, there is room for the application of
64 our best ecological judgement. Conversely, the metaweb should be reliably filled, which assumes that the
65 S^2 interactions in a pool of S species have been examined, either through literature surveys or expert
66 elicitation. Supp. Mat. 1 provides some guidance as to the type of sampling effort that should be
67 prioritized. While RDPG was able to maintain very high predictive power when interactions were missing,
68 the addition of false positive interactions was immediately detected; this suggests that it may be
69 appropriate to err on the side of “too many” interactions when constructing the initial metaweb to be
70 transferred. The second series of problems are related to determining which area should be used to infer
71 the new metaweb in, as this determines the species pool that must be used.

72 In our application, we focused on the mammals of Canada. The upside of this approach is that
73 information at the country level is likely to be required by policy makers and stakeholders for their
74 biodiversity assessment, as each country tends to set goals at the national level (Buxton *et al.* 2021) for
75 which quantitative instruments are designed (Turak *et al.* 2017), with specific strategies often enacted at
76 smaller scales (Ray *et al.* 2021). And yet, we do not really have a satisfying answer to the question of
77 “where does a food web stop?”; the current most satisfying solutions involve examining the spatial
78 consistency of network area relationships (see *e.g.* Galiana *et al.* 2018, 2019, 2021; **Fortin2021NetEco?**),
79 which is of course impossible in the absence of enough information about the network itself. This
80 suggests that an *a posteriori* refinement of the results may be required, based on a downscaling of the
81 metaweb. The final family of problems relates less to the availability of data or quantitative tools, and
82 more to the praxis of spatial ecology. Operating under the context of national divisions, in large parts of
83 the world, reflects nothing more than the legacy of settler colonialism. Indeed, the use of ecological data is
84 not an apolitical act (Nost & Goldstein 2021), as data infrastructures tend to be designed to answer
85 questions within national boundaries, and their use both draws upon and reinforces territorial statecraft;
86 as per Machen & Nost (2021), this is particularly true when the output of “algorithmic thinking” (*e.g.*
87 relying on machine learning to generate knowledge) can be re-used for governance (*e.g.* enacting
88 conservation decisions at the national scale). We therefore recognize that methods such as we propose
89 operate under the framework that contributed to the ongoing biodiversity crisis (Adam 2014), reinforced
90 environmental injustice (Choudry 2013; Domínguez & Luoma 2020), and on Turtle Island especially,

91 should be replaced by Indigenous principles of land management (Eichhorn *et al.* 2019; No'kmaq *et al.*
92 2021). As we see AI/ML being increasingly mobilized to generate knowledge that is lacking for
93 conservation decisions (*e.g.* Lamba *et al.* 2019; Mosebo Fernandes *et al.* 2020), our discussion of these tools
94 need to go beyond the technical, and into the governance consequences they can have.

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

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

109 Albouy, C., Archambault, P., Appeltans, W., Araújo, M.B., Beauchesne, D., Cazelles, K., *et al.* (2019). The
110 marine fish food web is globally connected. *Nature Ecology & Evolution*, 3, 1153–1161.

111 Buxton, R.T., Bennett, J.R., Reid, A.J., Shulman, C., Cooke, S.J., Francis, C.M., *et al.* (2021). Key
112 information needs to move from knowledge to action for biodiversity conservation in Canada.
113 *Biological Conservation*, 256, 108983.

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

117 Csermely, P. (2004). Strong links are important, but weak links stabilize them. *Trends in Biochemical*
118 *Sciences*, 29, 331–334.

119 Dalla Riva, G.V. & Stouffer, D.B. (2016). Exploring the evolutionary signature of food webs' backbones

120 using functional traits. *Oikos*, 125, 446–456.

121 Dallas, T., Park, A.W. & Drake, J.M. (2017). Predicting cryptic links in host-parasite networks. *PLOS*
 122 *Computational Biology*, 13, e1005557.

123 Domínguez, L. & Luoma, C. (2020). Decolonising Conservation Policy: How Colonial Land and
 124 Conservation Ideologies Persist and Perpetuate Indigenous Injustices at the Expense of the
 125 Environment. *Land*, 9, 65.

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

128 Eichhorn, M.P., Baker, K. & Griffiths, M. (2019). Steps towards decolonising biogeography. *Frontiers of*
 129 *Biogeography*, 12, 1–7.

130 Galiana, N., Barros, C., Braga, J., Ficetola, G.F., Maiorano, L., Thuiller, W., *et al.* (2021). The spatial scaling
 131 of food web structure across European biogeographical regions. *Ecography*, n/a.

132 Galiana, N., Hawkins, B.A. & Montoya, J.M. (2019). The geographical variation of network structure is
 133 scale dependent: Understanding the biotic specialization of host–parasitoid networks. *Ecography*, 42,
 134 1175–1187.

135 Galiana, N., Lurgi, M., Claramunt-López, B., Fortin, M.-J., Leroux, S., Cazelles, K., *et al.* (2018). The
 136 spatial scaling of species interaction networks. *Nature Ecology & Evolution*, 2, 782–790.

137 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2018). Bringing Elton
 138 and Grinnell together: A quantitative framework to represent the biogeography of ecological
 139 interaction networks. *Ecography*, 0.

140 Grünig, M., Mazzi, D., Calanca, P., Karger, D.N. & Pellissier, L. (2020). Crop and forest pest metawebs shift
 141 towards increased linkage and suitability overlap under climate change. *Communications Biology*, 3,
 142 1–10.

143 Herbert, F. (1965). *Dune*. 1st edn. Chilton Book Company, Philadelphia.

144 Hoffmann, J., Bar-Sinai, Y., Lee, L.M., Andrejevic, J., Mishra, S., Rubinstein, S.M., *et al.* (2019). Machine
 145 learning in a data-limited regime: Augmenting experiments with synthetic data uncovers order in
 146 crumpled sheets. *Science Advances*, 5, eaau6792.

147 Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, 30, 1883–1893.

148 Lamba, A., Cassey, P., Segaran, R.R. & Koh, L.P. (2019). Deep learning for environmental conservation.
 149 *Current Biology*, 29, R977–R982.

150 Machen, R. & Nost, E. (2021). Thinking algorithmically: The making of hegemonic knowledge in climate
 151 governance. *Transactions of the Institute of British Geographers*, 46, 555–569.

152 McLeod, A., Leroux, S.J., Gravel, D., Chu, C., Cirtwill, A.R., Fortin, M.-J., *et al.* (2021). Sampling and
 153 asymptotic network properties of spatial multi-trophic networks. *Oikos*, n/a.

154 Morales-Castilla, I., Matias, M.G., Gravel, D. & Araújo, M.B. (2015). Inferring biotic interactions from
 155 proxies. *Trends in Ecology & Evolution*, 30, 347–356.

156 Mosebo Fernandes, A.C., Quintero Gonzalez, R., Lenihan-Clarke, M.A., Leslie Trotter, E.F. & Jokar
 157 Arsanjani, J. (2020). Machine Learning for Conservation Planning in a Changing Climate.
 158 *Sustainability*, 12, 7657.

159 Neutel, A.-M., Heesterbeek, J.A.P. & de Ruiter, P.C. (2002). Stability in Real Food Webs: Weak Links in
 160 Long Loops. *Science*, 296, 1120–1123.

161 No'kmaq, M., Marshall, A., Beazley, K.F., Hum, J., Joudry, Shalan, Papadopoulos, A., *et al.* (2021).
 162 “Awakening the sleeping giant”: Re-Indigenization principles for transforming biodiversity
 163 conservation in Canada and beyond. *FACETS*, 6, 839–869.

164 Nost, E. & Goldstein, J.E. (2021). A political ecology of data. *Environment and Planning E: Nature and*
 165 *Space*, 25148486211043503.

166 O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., *et al.* (2020).
 167 Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. *Journal of*
 168 *Biogeography*, 47, 181–192.

169 Pedersen, E.J., Thompson, P.L., Ball, R.A., Fortin, M.-J., Gouhier, T.C., Link, H., *et al.* (2017). Signatures of
 170 the collapse and incipient recovery of an overexploited marine ecosystem. *Royal Society Open Science*,
 171 4, 170215.

172 Poisot, T., Cirtwill, A.R., Cazelles, K., Gravel, D., Fortin, M.-J. & Stouffer, D.B. (2016). The structure of
 173 probabilistic networks. *Methods in Ecology and Evolution*, 7, 303–312.

174 Poisot, T., Stouffer, D.B. & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary
 175 through space and time. *Oikos*, 124, 243–251.

- 176 Ray, J.C., Grimm, J. & Olive, A. (2021). The biodiversity crisis in Canada: Failures and challenges of
177 federal and sub-national strategic and legal frameworks. *FACETS*, 6, 1044–1068.
- 178 Strydom, T., Catchen, M.D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., *et al.* (2021). A
179 roadmap towards predicting species interaction networks (across space and time). *Philosophical*
180 *Transactions of the Royal Society B: Biological Sciences*, 376, 20210063.
- 181 Turak, E., Brazill-Boast, J., Cooney, T., Drielsma, M., DelaCruz, J., Dunkerley, G., *et al.* (2017). Using the
182 essential biodiversity variables framework to measure biodiversity change at national scale. *Biological*
183 *Conservation*, SI:Measures of biodiversity, 213, 264–271.
- 184 Wood, S.A., Russell, R., Hanson, D., Williams, R.J. & Dunne, J.A. (2015). Effects of spatial scale of
185 sampling on food web structure. *Ecology and Evolution*, 5, 3769–3782.