

Building a better metaweb: predicting spatiotemporally explicit plant-pollinator networks

[Michael D. Catchen](#)^{1,2} [Paul CaraDonna](#)^{3,4} [Jane E. Ogilvie](#)³ [Francis Banville](#)³ [Dominique Caron](#)^{1,2}
[Philippe Desjardins-Proulx](#)^{5,2} [Norma R. Forero-Muñoz](#)^{5,2} [Dominique Gravel](#)^{6,2} [Laura Pollock](#)^{1,2}
[Tanya Strydom](#)^{5,2} [Timothée Poisot](#)^{5,2} [Julian Resasco](#)⁷ [Andrew Gonzalez](#)^{1,2}

¹ McGill University ² Québec Centre for Biodiversity Sciences ³ Rocky Mountain Biological Laboratory
⁴ Chicago Botanic Garden ⁵ Université de Montréal ⁶ Université de Sherbrooke ⁷ University of
Colorado Boulder

Correspondance to:

Michael D. Catchen — michael.catchen@mail.mcgill.ca

Purpose: This template provides a series of scripts to render a markdown document into an interactive website and a series of PDFs.

Motivation: It makes collaborating on text with GitHub easier, and means that we never need to think about the output.

Internals: GitHub actions and a series of python scripts. The markdown is handled with pandoc.

1 Abstract

2 Using a data set of [DESCRIBE EACH DATASET IN A NICE WAY], we predict a spatiotemporally explicit
3 metaweb of interactions between bumblebees (*Bombus*) and wildflowers (within *find clade*). We integrate
4 this data with crowdsourced occurrence data and climate data to [best paint the picture of the Colorado
5 bumblebee-plant metaweb]. Using temporal climate data, we forecast how the spatiotemporal overlap of
6 interacting species will change under proposed climate scenarios. We use this to estimate what
7 interactions between bees and plants need the most attention to prevent the spatiotemporal decoupling of
8 an interactions from threatening ecosystem functioning or the persistence of a species.

9 Introduction

10 Species interactions are important. It is ultimately interactions between individuals of different species
11 that drive the structure, dynamics, and persistence of ecosystems, and the abundance and diversity of the
12 species within them. Plant-pollinator interactions specifically drive the function and persistence of
13 “architecture of biodiversity” (Bascompte & Jordano 2007). However, we are far from a robust
14 understanding of plant-pollinator networks. This is because sampling interactions is costly. Interactions
15 vary in space and time (Poisot *et al.* 2015)—particularly relevant in this system (CaraDonna *et al.* 2014).
16 This is why there is interest in using models to predict interactions from sparse data (Strydom2021?). In
17 this paper, we combine several datasets, each spanning several years, to produce spatially and temporally
18 explicit predictions of the bumblebee (genus *Bombus*) and wildflower pollination network across the state
19 of Colorado.

20 We do this in two parts: (1) metaweb prediction and (2) conditioning our metaweb prediction on
21 co-occurrence probability. First, we build a model to predict the metaweb—the network of *all*
22 interactions, aggregated across all times and spatial locations—of *Bombus* and wildflower species across
23 Colorado. (Why do this? The metaweb is more predictable than local interactions.) We do this using
24 network embedding (cite?). Network embedding takes each node in the network (either a bumblebee or a
25 wildflower) and represents it in a latent n dimensional space. Combination of running models on
26 Temporal niche (T), Phylogenetic niche (P), Environmental niche (E), and relative abundance in
27 community (RA).

28 Second, we then use this metaweb to predict the structure of networks at specific locations and times of
29 year (Gravel *et al.* 2019). Finally we suggest a map of sampling priority, which suggests the locations to
30 sample that will best improve our understanding of the Colorado *Bombus* pollination metaweb.
31 Why is this good for science, what does this contribute to our understanding of plant-pollinator ints,
32 networks, *Bombus*, predictive models, etc., and how can these results be useful.

33 **Data**

34 We use three separate field datasets to estimate the Colorado *Bombus* metaweb.

35 **Methods**

36 [Figure 1 about here.]

37 **Metaweb Model**

38 **Feature Embedding**

39 **Environmental niche features**

40 We take the 19 BioClim layers from CHELSA (cite; 1km resolution) and a map of elevation and PCA them.
41 A resulting 4 layers cover 99.5% of the variance. We use species occurrence data from GBIF, and consider
42 each occurrence record as a point in environment space. Then we fit a multivariate normal distribution to
43 these points in environmental space.

44 **Temporal niche features**

45 We take the mean and variance of the distribution of number of observations per week of year in the
46 interaction field data.

47 **Phylogenetic features**

48 **Phylogeny Construction** We construct phylogenies for both *Bombus* and wildflower species using
49 barcode markers, mitochondrial COI and chloroplast rbcL, respectively. These sequences were obtained
50 from NCBI GenBank for all species. For species for which no sequence was available (only a handful of
51 plants), their was substituted with a barcode from a member of the same genus. Justify why this is fine
52 here.

53 These sequences were aligned using ClustalOmega v???, and then a posterior distribution of phylogenies
54 and consensus tree was obtained via MrBayes v??., using XX substitution model with gamma-distributed
55 rates. Run until convergence, which here we define as the standard-deviation of splits falling below 0.1.

56 **Creating an embedding from phylogenies** We simulate traits. Relationship between number of
57 traits, num output PCA dimensions, and number of used dimensions in the model matter. Describe how
58 that works.

59 **Relative Abundance**

60 This embedding is the simplest.

61 **Metaweb Model Fitting and Validation**

62 We fit a bunch of models using MLJ.jl.

63 Some of them are bagged, some are not bagged.

64 AUC-ROC and AUC-PR values below in fig. [2](#)

65 [Figure 2 about here.]

66 What does this tell us? The ensemble model is regularly the best for ROC, but not for PR. This illustrates
67 an inherent trade-off between models being as “right” as possible versus a model being useful for
68 discovering false-negatives.

69 **Spatiotemporally Explicit Networks**

70 Now that we have a metaweb, we can extend this to predict interactions at particular places and times by
71 decomposing the probability of interaction at particular place and time into probability of interaction
72 multiplied by probability of co-occurrence via properties of conditional probability (Gravel *et al.* 2019).

73 How do we define? $P(i \leftrightarrow j)$ Is it $P(A_{ij})P(O_{ij})$ or $P(A_{ij})P(O_i)P(O_j)$

74 **Figure 3: Maps over time figure and Prob(Connectance) vs. Month figure**

75 **Sampling Prioritization**

76 How do we improve our understanding of this pollination network, or determine if it is changing over
77 time?

78 **Figure 4: Uncertainty and sampling priority map**

79 **Discussion**

80 Bascompte, J. & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity.
81 *Annual Review of Ecology, Evolution, and Systematics*, 38, 567–593.

82 CaraDonna, P.J., Iler, A.M. & Inouye, D.W. (2014). Shifts in flowering phenology reshape a subalpine plant
83 community. *Proceedings of the National Academy of Sciences*, 111, 4916–4921.

84 Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., *et al.* (2019). Bringing Elton
85 and Grinnell together: A quantitative framework to represent the biogeography of ecological
86 interaction networks. *Ecography*, 42, 401–415.

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

Metaweb Prediction

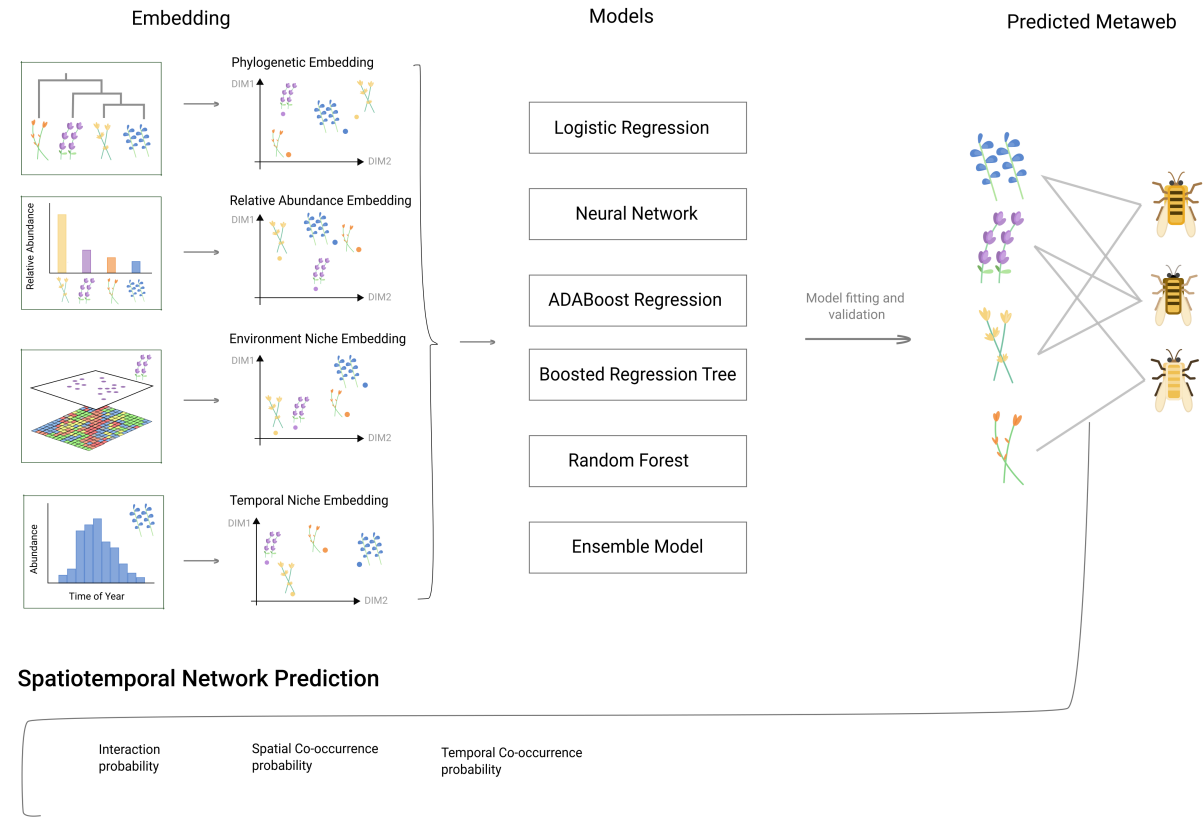


Figure 1: todo

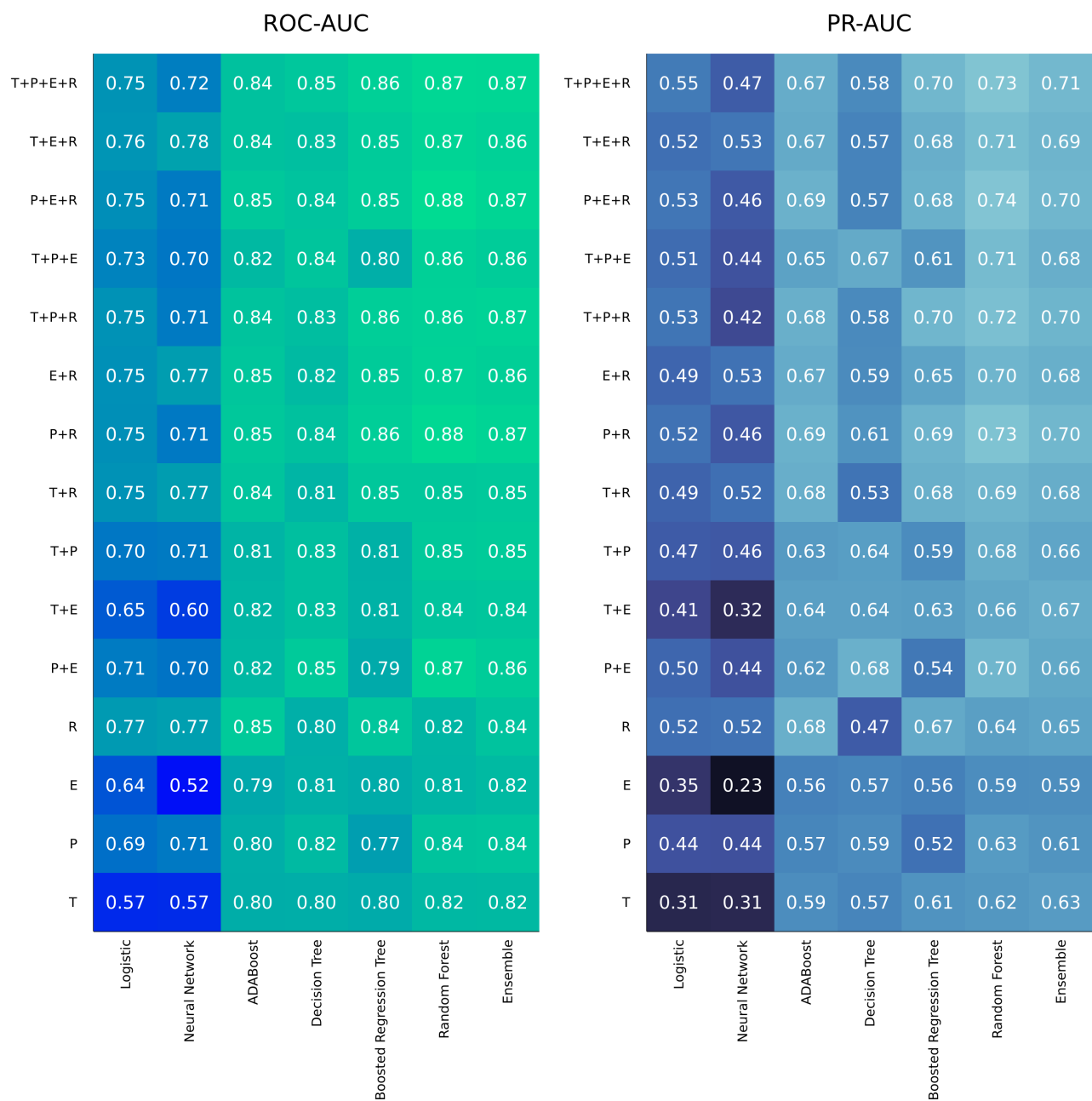


Figure 2: todo