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 Bontanic 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 scritpts. The markdown is handled with pandoc.

0.1. Abstract

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

1 _____

Introduction

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

We do this in two parts: (1) metaweb prediction and (2) conditioning our metaweb prediction on co-occurrence probability. First, we build a model to predict the metaweb—the network of *all* interactions, aggregated across all times and spatial locations—of *Bombus* and wildflower species across Colorado. (Why do this? The metaweb is more predictable than local interactions.) We do this using network
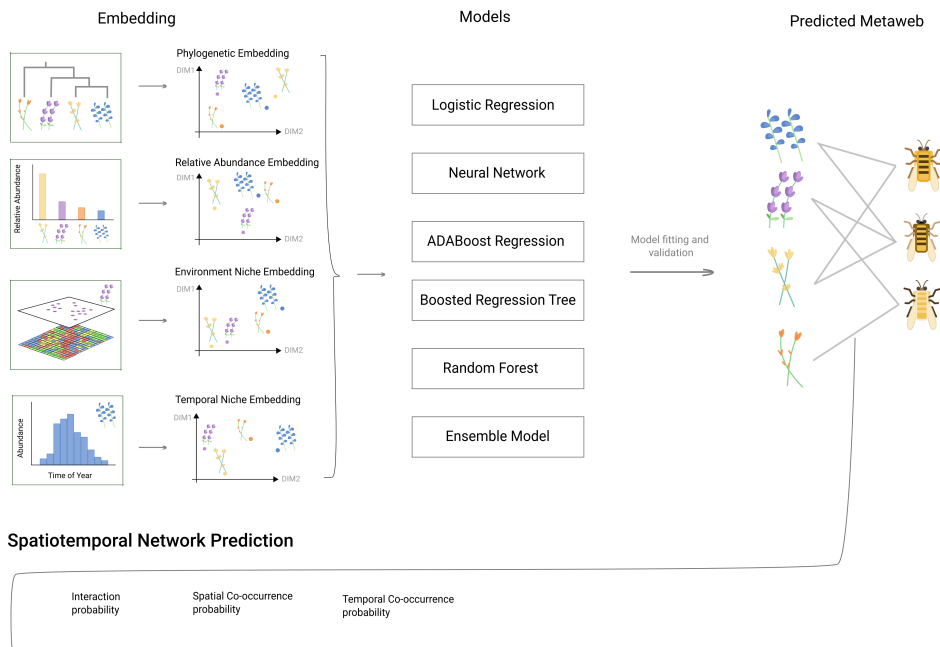
Figure 1 todo

embedding (**cite?**). Network embedding takes each node in the network (either a bumblebee or a wildflower) and represents it in a latent n dimensional space. Combination of running models on Temporal niche (T), Phylogenetic niche (P), Environmental niche (E), and relative abundance in community (RA).

Second, we then use this metaweb to predict the structure of networks at specific locations and times of year (Gravel *et al.* 2019). Finally we suggest a map of sampling priority, which suggests the locations to sample that will best improve our understanding of the Colorado *Bombus* pollination metaweb.

Why is this good for science, what does this contribute to our understanding of plant-pollinator ints, networks, Bombus, predictive models, etc., and how can these results be useful.

2
Data

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

3
Methods

4
Metaweb Model

4.1. Feature Embedding

4.1.1 *Relative Abundance*

4.1.2 Phylogenetic features

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

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

4.1.2.2 Creating an embedding from phylogenies We simulate traits.

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

4.1.4 Temporal niche features
We take the mean and variance of the distribution of number of observations per week of year in the interaction field data.

4.2. Metaweb Model Fitting and Validation
We fit a bunch of models using MLJ.jl.

Some of them are bagged, some are not bagged.

AUC-ROC and AUC-PR values below in fig. 2

What does this tell us? The ensemble model is regularlly the best for ROC, but not for PR. This illustrates a tradeoff between models being as "right" as possible versus a model being useful for discovering false-negatives.

5
Spatiotemporally Explicit Networks

Now that we have a metaweb, we can extend this to predict interactions at partilar places and times by decomposing the probability of interaction at particular place and time into probability of interaction times probability of cooccurrence via properties of conditional prob. (Gravel *et al.* 2019)

$P(A_{ij})P(O_{ij})$ vs $P(A_{ij})P(O_i)P(O_j)$

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

6
Sampling Prioiritization

Figure 4: Uncertainty and sampling priority map

Discussion

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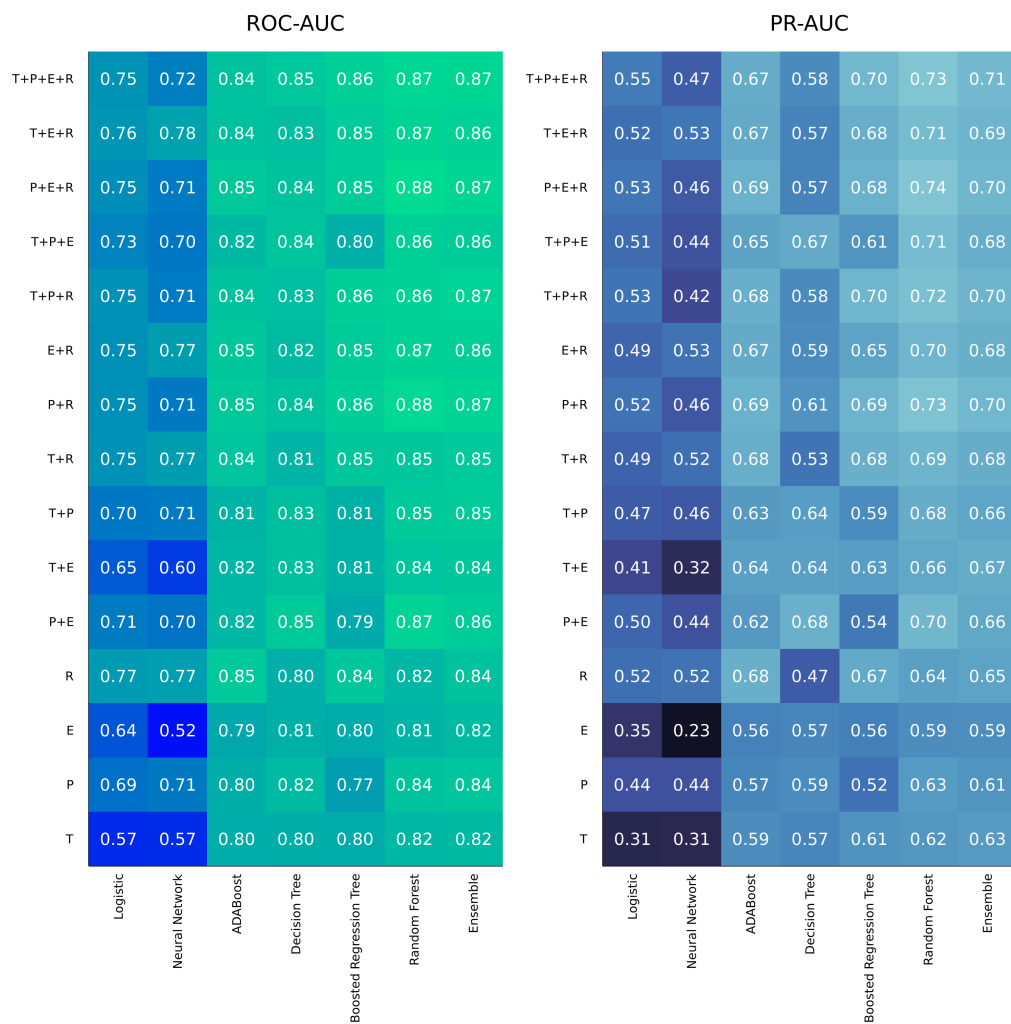
ROC-AUC

	Logistic	Neural Network	ADABoost	Decision Tree	Boosted Regression Tree	Random Forest	Ensemble
T+P+E+R	0.75	0.72	0.84	0.85	0.86	0.87	0.87
T+E+R	0.76	0.78	0.84	0.83	0.85	0.87	0.86
P+E+R	0.75	0.71	0.85	0.84	0.85	0.88	0.87
T+P+E	0.73	0.70	0.82	0.84	0.80	0.86	0.86
T+P+R	0.75	0.71	0.84	0.83	0.86	0.86	0.87
E+R	0.75	0.77	0.85	0.82	0.85	0.87	0.86
P+R	0.75	0.71	0.85	0.84	0.86	0.88	0.87
T+R	0.75	0.77	0.84	0.81	0.85	0.85	0.85
T+P	0.70	0.71	0.81	0.83	0.81	0.85	0.85
T+E	0.65	0.60	0.82	0.83	0.81	0.84	0.84
P+E	0.71	0.70	0.82	0.85	0.79	0.87	0.86
R	0.77	0.77	0.85	0.80	0.84	0.82	0.84
E	0.64	0.52	0.79	0.81	0.80	0.81	0.82
P	0.69	0.71	0.80	0.82	0.77	0.84	0.84
T	0.57	0.57	0.80	0.80	0.80	0.82	0.82

PR-AUC

	Logistic	Neural Network	ADABoost	Decision Tree	Boosted Regression Tree	Random Forest	Ensemble
T+P+E+R	0.55	0.47	0.67	0.58	0.70	0.73	0.71
T+E+R	0.52	0.53	0.67	0.57	0.68	0.71	0.69
P+E+R	0.53	0.46	0.69	0.57	0.68	0.74	0.70
T+P+E	0.51	0.44	0.65	0.67	0.61	0.71	0.68
T+P+R	0.53	0.42	0.68	0.58	0.70	0.72	0.70
E+R	0.49	0.53	0.67	0.59	0.65	0.70	0.68
P+R	0.52	0.46	0.69	0.61	0.69	0.73	0.70
T+R	0.49	0.52	0.68	0.53	0.68	0.69	0.68
T+P	0.47	0.46	0.63	0.64	0.59	0.68	0.66
T+E	0.41	0.32	0.64	0.64	0.63	0.66	0.67
P+E	0.50	0.44	0.62	0.68	0.54	0.70	0.66
R	0.52	0.52	0.68	0.47	0.67	0.64	0.65
E	0.35	0.23	0.56	0.57	0.56	0.59	0.59
P	0.44	0.44	0.57	0.59	0.52	0.63	0.61
T	0.31	0.31	0.59	0.57	0.61	0.62	0.63

Figure 2 todo

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

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

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