

Describing Sub-community Structure In a Plant-Pollinator Network With Physical Traits

Lauren Beatty

Eco-Informatics Summer Institute

H.J. Andrews Experimental Forest and Oregon State University
Macalester College, St. Paul, MN

Ryan DeMuse

Eco-Informatics Summer Institute

H.J. Andrews Experimental Forest and Oregon State University
University of San Diego, San Diego, CA

19 August, 2014

Abstract

Ecological plant-pollinator networks often contain underlying community structure. This structure often presents itself in the form of network modules. These modules are subsets of the greater network and can be looked at as smaller networks in their own right or single components in a greater network. We look for underlying network structure in plant-pollinator interactions data collected between 2011 and 2013 in the H.J. Andrews Experimental Forest through the Eco-Informatics Summer Institute at Oregon State University. We analyze this data and compute modules to group species of plants and pollinators to more easily study this large and complex ecological network. We then construct binary trees from a matrix of given covariates to determine whether insect and plant traits provide any indication of module membership.

Introduction

In ecological literature concerning mutualistic plant-pollinator networks, there are often too many species and too many interactions to easily understand the structure of the network. It is common for these networks to be visualized with bipartite plots, where two distinct sets of nodes represents plants and pollinators and lines between them represent interactions. These visualizations are too noisy to convey much information to the reader because every species and interaction is represented in an already large, complex network (Fig. 1). It would be much easier to interpret meaningful results from these plots if there were fewer species present in the network or to try to analyze a subset of the network and scale those methods and results to the larger network as a whole. Another approach would be to group the species and analyze connections between groups of species. This would make analysis of species interactions much more manageable.

In this paper, we will analyze a plant-pollinator network observed in a montane meadow in the H.J. Andrews Experimental Forest. We apply a method to partition the graph into modules. A module is defined as a set of plant-pollinator interactions that maximizes connections within modules and minimizes interactions between modules (Dormann 2013). Modules reveal underlying sub community structure. Given these modules, we use physical plant and insect traits to fit regression and classification trees. These trees use our physical covariates to predict which module the species will belong.

The objectives of this work are to (1) describe modules and identify which physical traits explain or characterize sub community structure within the network, and (2) predict modules based on plant and pollinator traits. The overall objective is to create a simpler

network, in which nodes will be groups of plants and groups of pollinators characterized by a set of physical traits, which still accurately represents the meadow interactions.

Study area and data

Our data was collected from 2011 to 2013 in a subalpine xeric meadow (Carpenter Ridge) in the Carpenter Mountain Complex at H.J. Andrews Experimental Forest. Two parallel transects were laid out in the meadow. Five 3 x 3m plots were laid out at 15m intervals in each. At each plot, a flower survey was taken before each watch. The number of stems was counted for each flowering plant in the plot. Flowers per stem was counted for up to ten stems for each species. Stems were multiplied by the average flowers per stem to create an estimate of abundance within the plot. Each plot was watched for 15 minutes, with minutes recorded discretely. Once an interaction occurred, if the pollinator could not be identified by sight, it was caught, pinned, and sent to Prof. Andrew Moldenke, an entomologist at Oregon State University, for identification. Since minutes were recorded discretely, at the end of each minute the observer would record all interactions that had happened.

Methods

We implemented a two-step approach to analyzing our data on the 2011 to 2013 interactions data from the Carpenter Ridge Meadow. We hope that the methods that we used to analyze the Carpenter Ridge meadow can be applied to the network as a whole or any subset of the network. We chose to analyze a single meadow so that we could complete a more thorough and complete analysis that can be applied to other meadows.

Defining modules

The first step was to use the function *computeModules*, part of the R programming language package *bipartite* (Dormann and Rouven), to compute modules based on the binary matrix of the presence or absence of plant pollinator interactions. Creating modules is a way to more simply describe an ecological system. It does not attempt to represent every single species, but rather it uses interactions data to place the data into a simpler set of modules (Dormann 2013). The simplest characteristic of a module is that interactions within modules are more prevalent than interactions between modules. Modules attempt to show any identifiable substructure in a network (Dormann 2013). The algorithm that is implemented in the *computeModules* function in the *bipartite* package is outlined in Dormann (2013). The function takes as an input a weighted bipartite graph and computes modules from the data set using Newman's modularity measure. A high modularity measure implies a high number of connections within modules and a low number of connections between different modules. The algorithm is designed to build a graph so that nearby species are more likely to interact; then the algorithm swaps branches at random levels to determine if the newly constructed graph is better than the old graph. If the new graph is better than the previous, the new one is saved and the process is repeated and continues until no better graph is found after a predefined number of swaps (Dormann 2013).

The algorithm for the *computeModules* function divides the graph of interactions between plant and pollinator species into modules such that three conditions are satisfied. Each module must be a connected subgraph of the network. For our network, this means that each plant must have a corresponding pollinator and vice versa. Secondly each vertex belongs to

exactly one module. This ensures that the modules are disjoint and each module contains unique individuals in relation to other modules. Lastly the edge weights within a module are higher than the edge weights outside of modules.

Predicting modules: The Classification Tree Algorithm

Once the data was partitioned into modules, we used classification trees to predict module outcome based on physical trait covariates. Classification trees consist of nodes which split species into increasingly smaller groups. At the bottom of the tree a prediction of which module the species will belong to is given. Unlike traditional modelling techniques, classification trees have the ability to handle categorical and ordinal variables (De'ath et. al. 2010).

Furthermore, trees elegantly deal with distinct interactions of traits. Trees were appropriate for our work because it was immediately apparent that no single physical trait described the modules well, and because many of the covariates we had at our disposal were categorical. We used a dataset of covariates created by Prof. Moldenke. The covariates are listed in Figure 2.

In making trees we had two goals in mind. Firstly, we wanted to find trees with the lowest residuals possible. That is, we wanted to find the tree that most accurately placed a given individual into the correct module based on known covariate information. We also wanted to determine which single covariate is the best predictor of module membership.

We used the CART algorithm developed by Breiman et al. (1984) to create our trees. This algorithm has been incorporated into R by Therneau et al. (2014) in a package called *rpart*. For categorical variables, the tree will split cases on a true/false basis. For quantitative variables, the tree will split cases if they are greater than or less than a given value. The ultimate

goal of the algorithm is to take the covariates of a species and predict a class of outcome (module).

The algorithm makes splits which maximize the decrease in impurity. At node T , impurity is measured with the Gini index (Breiman et al., 1984).

$$gini(T) = 1 - \sum_{i=1}^n p_i^2$$

Where the set of modules is $\{1, 2, \dots, n\}$ and p_i is the relative frequency of species in module i in node T . The algorithm will split node T into nodes T_1 and T_2 such that the weighted average of $gini(T_1)$ and $gini(T_2)$ is minimized. The user defines a complexity parameter (cp) so that the algorithm stops splitting if it cannot decrease impurity by a factor of cp, or if every node contains species of only one module (Therneau et al., 2014).

Results

The function *computeModules* identified 8 modules in the data collected from 2011 to 2013 in the Carpenter Ridge meadow in the Carpenter Mountain Complex of the H.J. Andrews Experimental forest (Fig. 3), with a Modularity value of 0.47.

For insects, individual biomass, taxon group, taxon guild, trophic guild, and tongue length all produced trees with some predictive power (Fig. 4,5,6,7,8). Insect biomass correctly placed 64% of insect species into their module. The other covariates correctly classified anywhere from 35% to 51% of the species.

For plants, biomass, lifeform, and tube type produced somewhat useful trees. Our plant biomass tree correctly placed 74% of flower species, while tube type and lifeform correctly

placed 37% and 47% of species respectively (Fig. 9,10,11). Plant microhabitat produced nothing but trivial trees with a single node grouping all species into module 6.

When multiple covariates were combined, better trees could be found. Fig. 12 shows an insect tree produced using individual biomass, tongue length, and taxon guild. This tree correctly placed 64% of individuals. For plants, using multiple covariates could actually increase the accuracy of the tree. Fig. 13 shows a tree produced using lifeform, tube type, and biomass. This tree correctly placed 84% of species.

Discussion

The modularity value of 0.47 was the highest Modularity constant that we could manage to obtain in our data analysis. Dormann 2013 suggests that the *computeModules* program be run several times under the same conditions to produce the best possible modules. Furthermore, due to the nature of the algorithm, insofar as it chooses a random starting position with low modularity and randomly swaps vertices to produce modules with higher modularity, with more computation time and power it is likely that better modules could be produced.

We also compare trees for plants made with single covariates and trees for pollinators with single covariates. In this analysis, we are not specifically interested in trees that have really good percentages of correct module placement, but rather trees which have better percentages compared to one another. For example, a tree that correctly places 60% of individuals is not a good tree for predicting module membership, but it is much better than a tree that only places 30% of individuals in the correct module. Approaching our analysis from this perspective provides more meaningful results because it tells us which covariates are better predictors of which module a given plant or pollinator belongs. From this perspective we determine that the

best covariate for predicting which module a plant or pollinator belongs to would be biomass for both. Most of the trees produced using a single covariate did not have high accuracy. This means that based on our results, traits are not an overwhelmingly good predictor of module membership.

In Fig. 9, plant biomass correctly placed 74% of individuals whereas in Fig. 13, a combination of lifeform, biomass, and tube type result in 84% correctly placed individuals. Similarly for insects, Fig. 4 correctly places 64% of individuals and Fig. 12 also correctly places 64% of individuals. Nevertheless, it is clear that Fig. 12 is a much simpler tree than Fig. 4. Fig. 12 has a height of 5 whereas Fig. 4 has a height of 9 making the tree in Fig. 4 much more complex. Thus, trees produced with multiple covariates are better than trees produced with single covariates.

Using the trees from Fig. 12 and Fig. 13 we constructed Fig. 14, a table listing sets of plant and pollinator traits by module. This allowed us to see if any patterns could be noticed when plant traits and pollinator traits for a given module were examined simultaneously. While most of the modules did not show obvious patterns, there were a few interesting results. Module 6 is characterized by plants of relatively small biomass and pollinators of the taxon guilds long-tongued bee fly, conopid fly, and megachilid bees of biomass less than 77. Module 2 is characterized by moderately-sized plants with bowl shaped tubes, and a wide variety of pollinator traits. Since physical traits of pollinators are not well defined in this module, it likely means that moderately-sized flowers with bowl tubes can interact with and are attractive to a diverse range of insect species.

Conclusions and Further Work

Our study attempted to determine which physical covariates best predicted module membership. We found that while biomass was the best predictor for both plants and pollinators, we could create better trees with multiple covariates. When trees were created with multiple covariates they could improve accuracy and decrease complexity over trees created with one trait. This suggests that sub community structure in our plant-pollinator network is best characterized by sets of traits, instead of single covariates alone. We believe that this method of looking at traits as an indicator of module membership is a valid method to explore and could be a powerful indicator in other mutualistic networks of different scales and complexity.

One possible direction for further work would be to use a more detailed and comprehensive list of covariates for the plants and pollinators for making the tree diagrams. Biomass was the best predictor of module membership for both plants and pollinators. Biomass was also the most precise covariate we had available since it is a continuous variable. Therefore the indication from the current data that we have seems to be that the more precise the covariate value, the easier it is to produce a more accurate tree for module membership.

Another further direction of this research would be to use this method on other meadows in the H.J. Andrews Forest as well as analysis on the complex and entire network scales. This would take a significant amount of computation time to construct the modules due to the sheer size and complexity of the network, but it would be interesting to see the module structure of such a large network compared to that of a single meadow.

References

- Breiman, Leo, Jerome Friedman, Charles J. Stone, and Richard A. Olshen. *Classification and regression trees*. CRC press, 1984.
- De'ath, G., & Fabricius, K.D. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *The Ecological Society of America*, 81(11). Retrieved from <http://www.esajournals.org/doi/pdf/10.1890/0012-9658%282000%29081%5B3178%3ACARTAP%5D2.0.CO%3B2>
- Dormann, Carsten F., and Rouven Strauss. "Detecting Modules in Quantitative Bipartite Networks: The QuaBiMo Algorithm." (12 April 2013): n. pag. *ArXiv.org*. Web. 15 Aug. 2014. <<http://arxiv.org/pdf/1304.3218.pdf>>.
- Therneau, Terry., Atkinson, Beth., & Ripley, Brian. (2014). rpart: Recursive Partitioning and Regression Trees. R package

Appendix

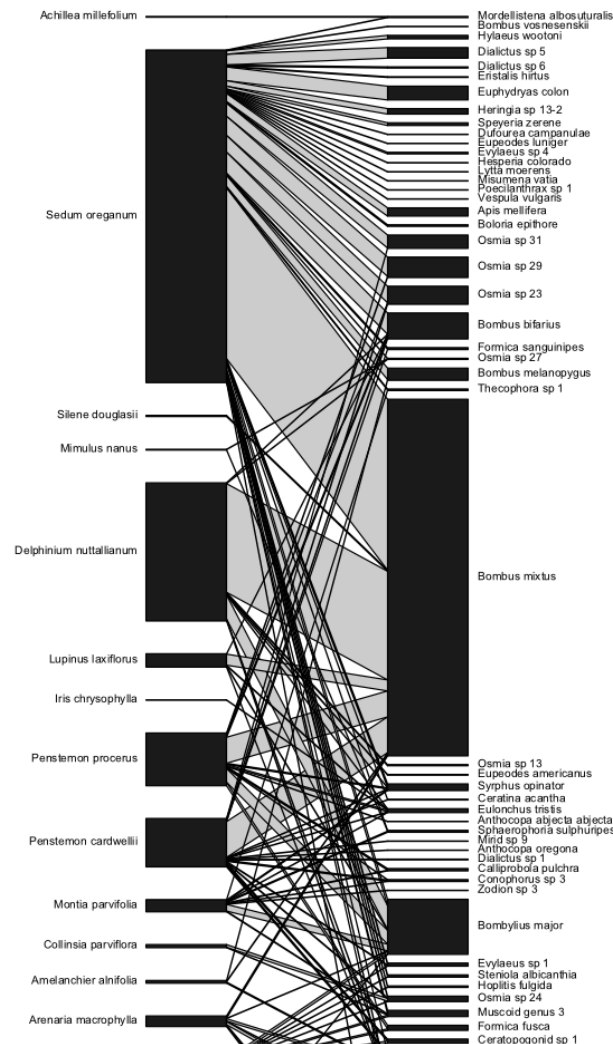


Figure 1. A bipartite graph of the observed plant-pollinator interactions from 2011 to 2013 in the Carpenter Ridge Meadow

Figure 3. Module diagram produced using the *computeModules* function from the bipartite R package. Plant species along the vertical axis and Pollinator species lie along the horizontal axis. The red boxes indicate the distinct modules.

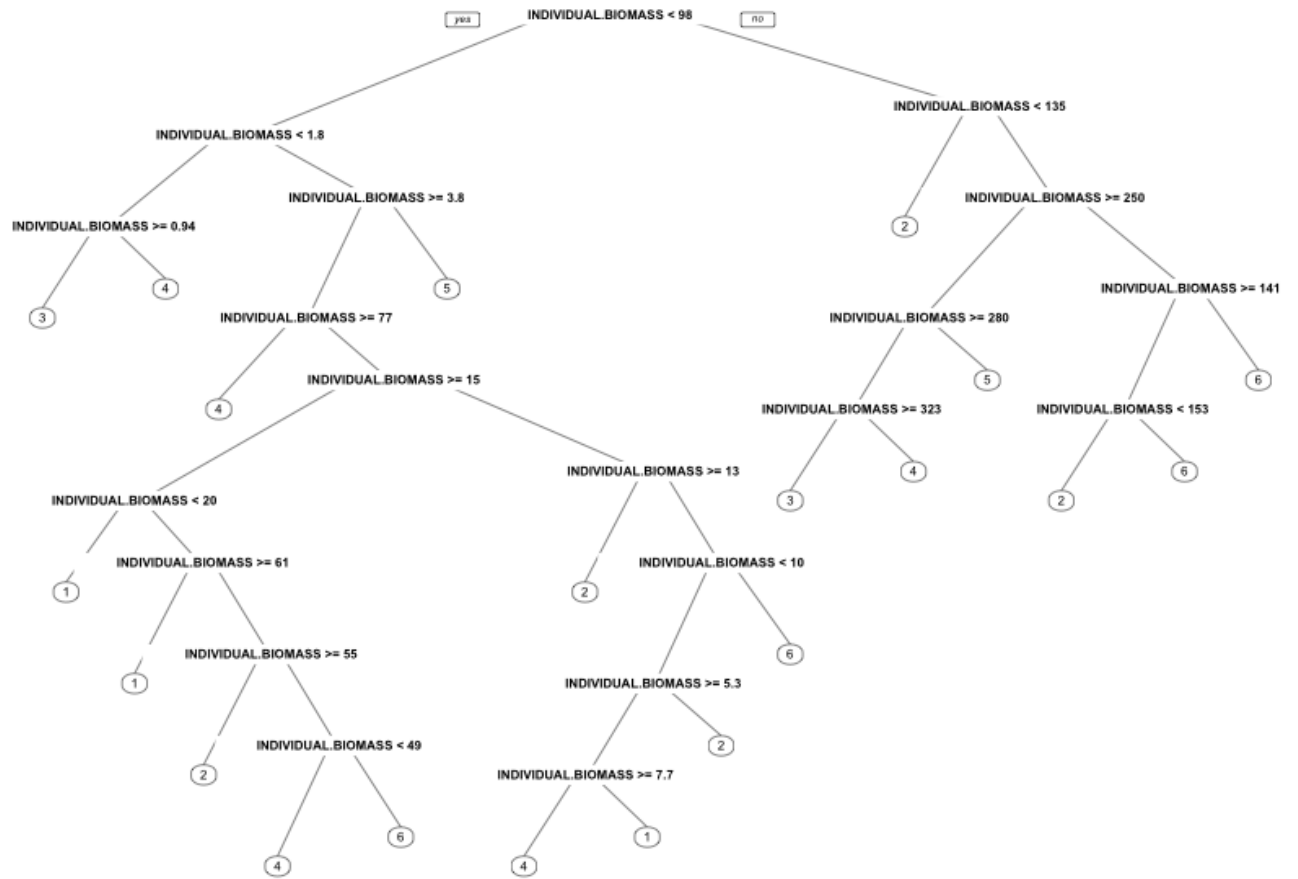


Figure 4. Tree produced using the insect biomass covariate with the *rpart* function. This tree correctly places 64% of distinct species into their correct modules.

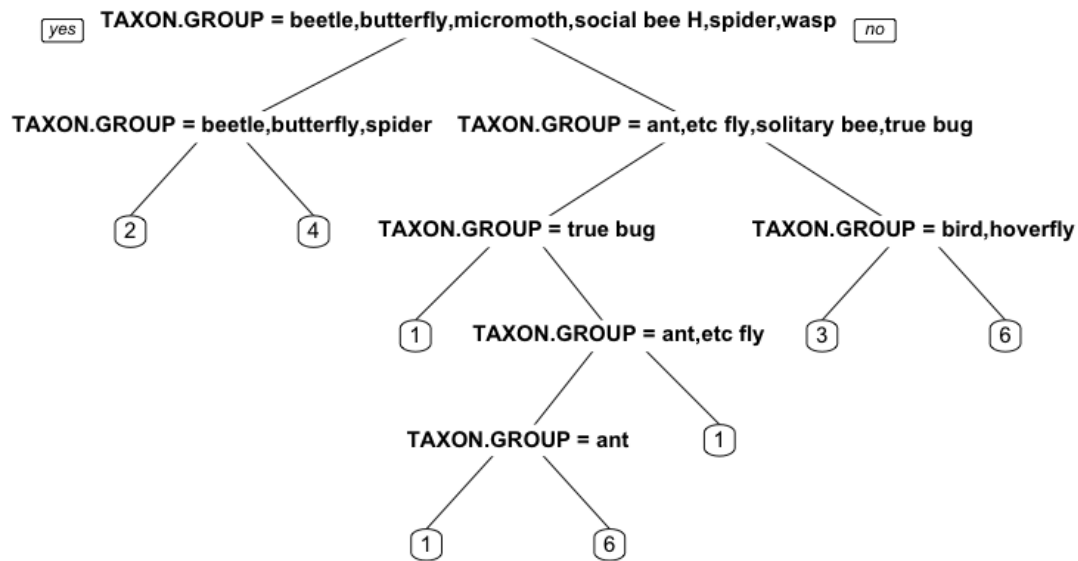


Figure 5. Tree produced using the insect taxon group covariate. This tree correctly placed 45% of distinct species.

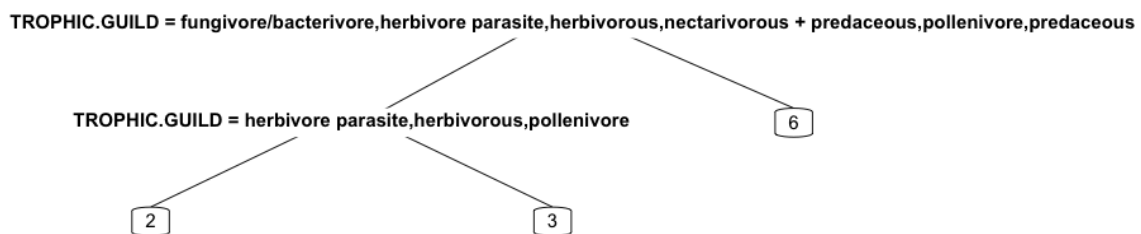


Figure 6. Tree produced using the insect trophic guild covariate. The tree correctly placed 35% of individuals.

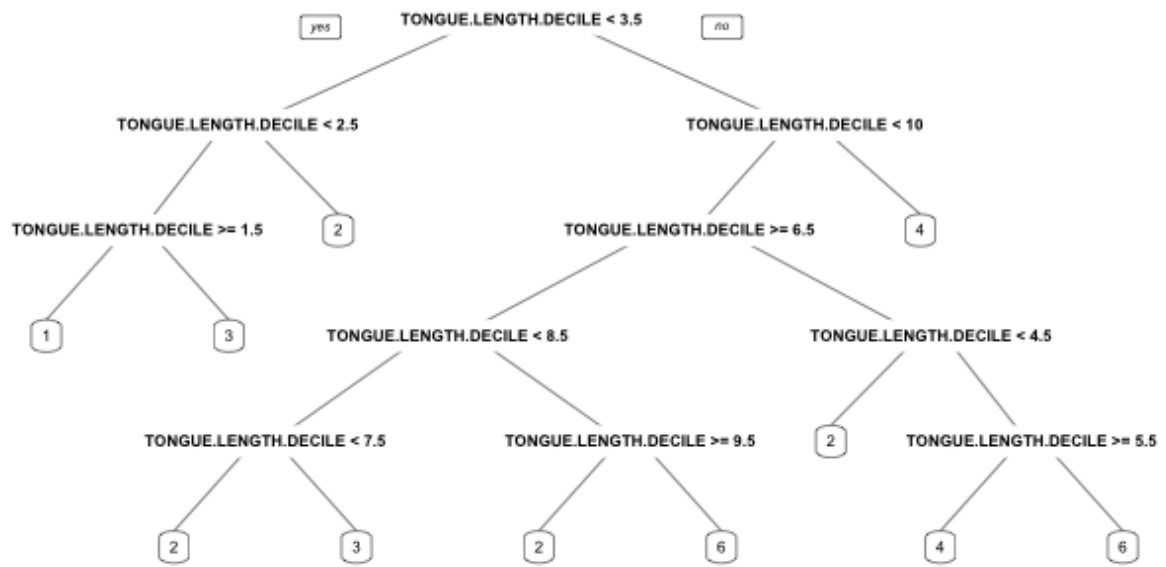


Figure 7. Insect tree produced using insect tongue length. Correctly classified 43% of species.

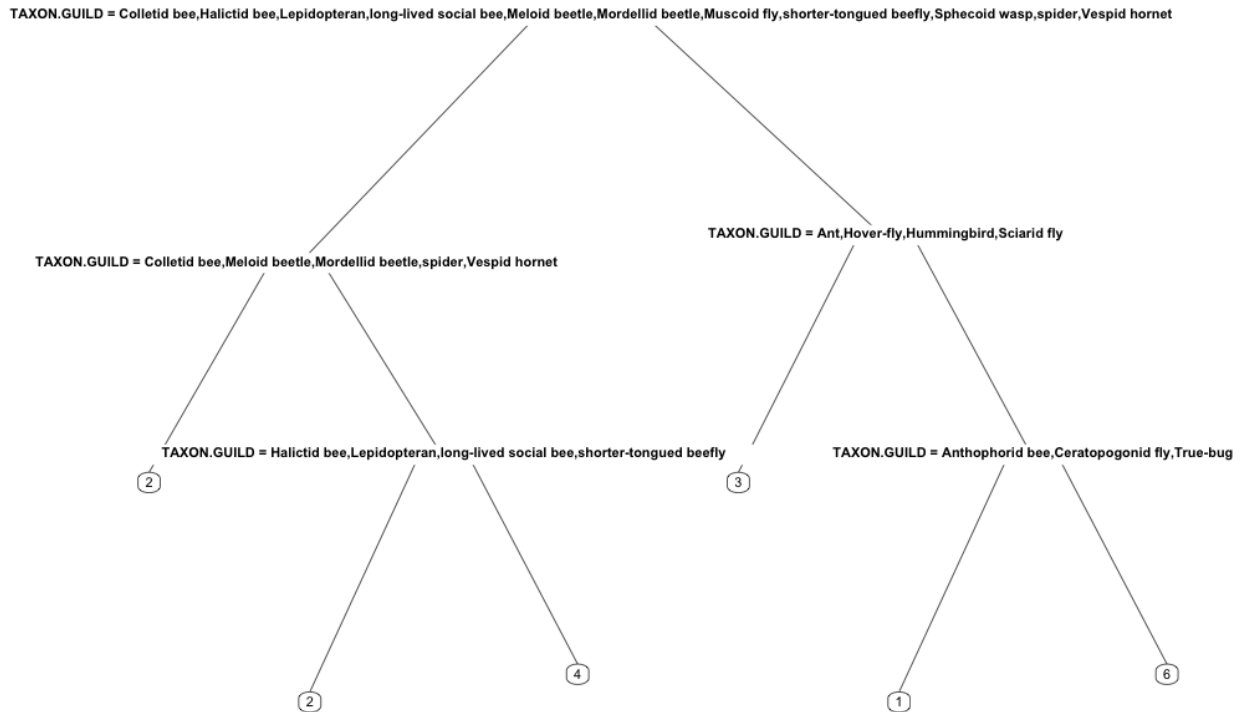


Figure 8. Insect tree produced with insect taxon guild. Correctly places 51% of species.

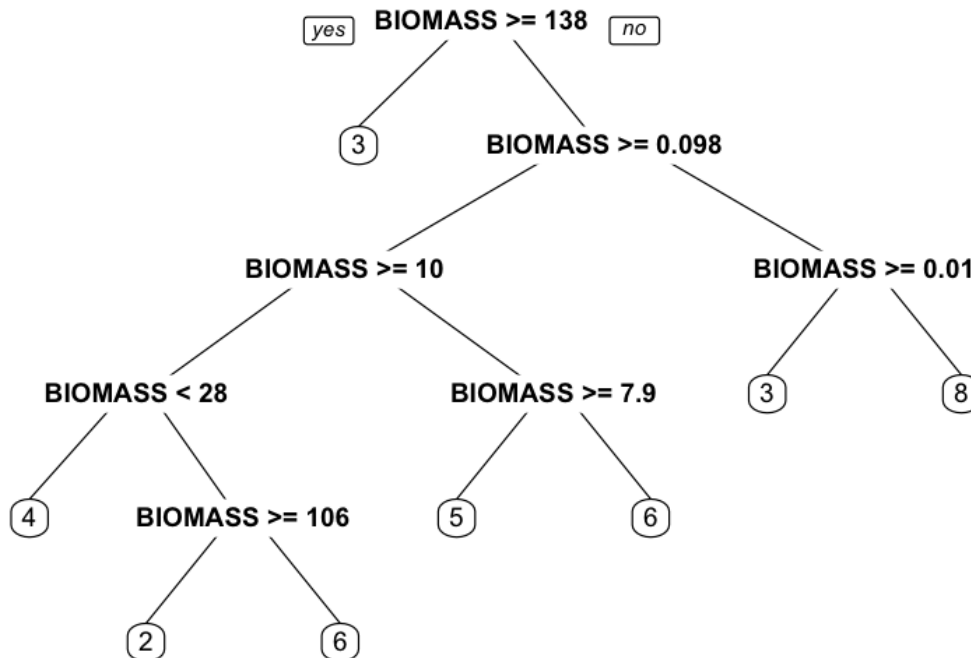


Figure 9. Tree produced using the plant biomass covariate. This tree correctly placed 74% of individuals.

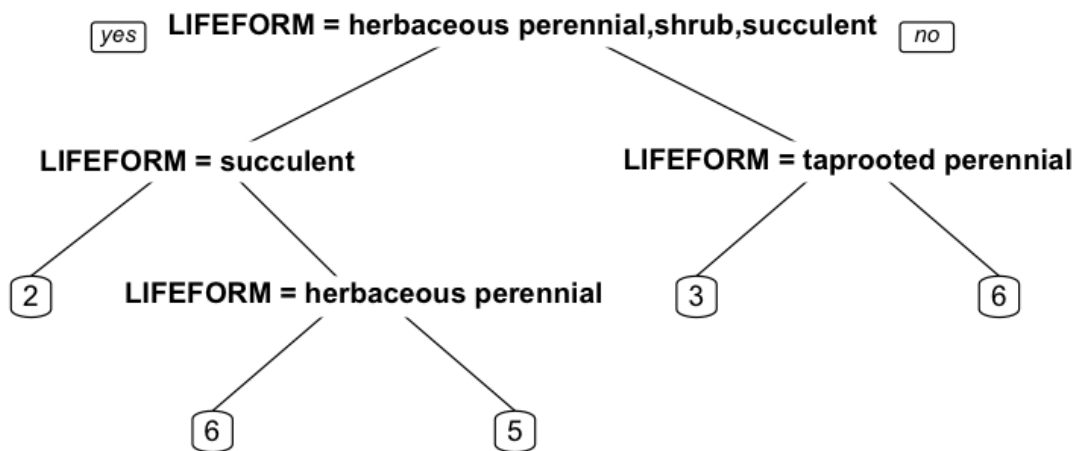


Figure 10. Tree produced using the plant lifeform covariate. Correctly places 47% of individuals.

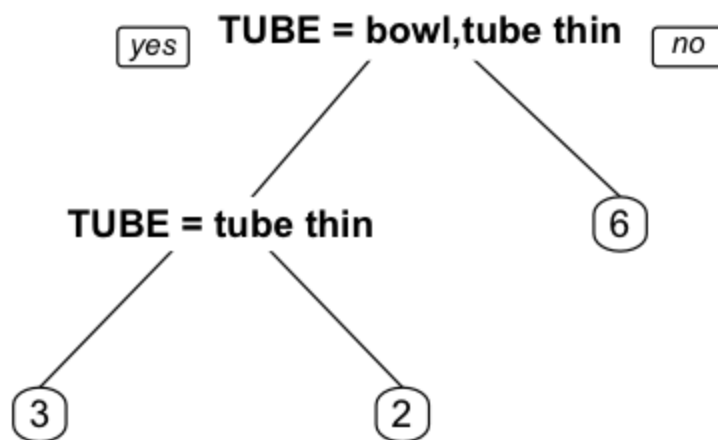


Figure 11. Plant tree produced using tube type. Correctly classifies 37% of species.

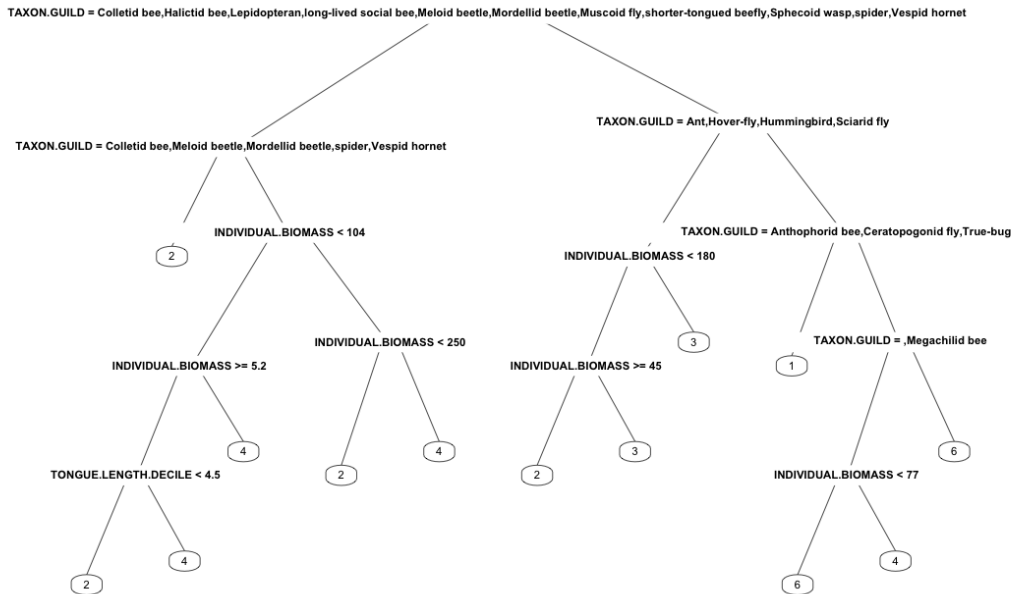


Figure 12. Insect tree produced with taxon guild, biomass, and tongue length. Correctly places 64% of species and is

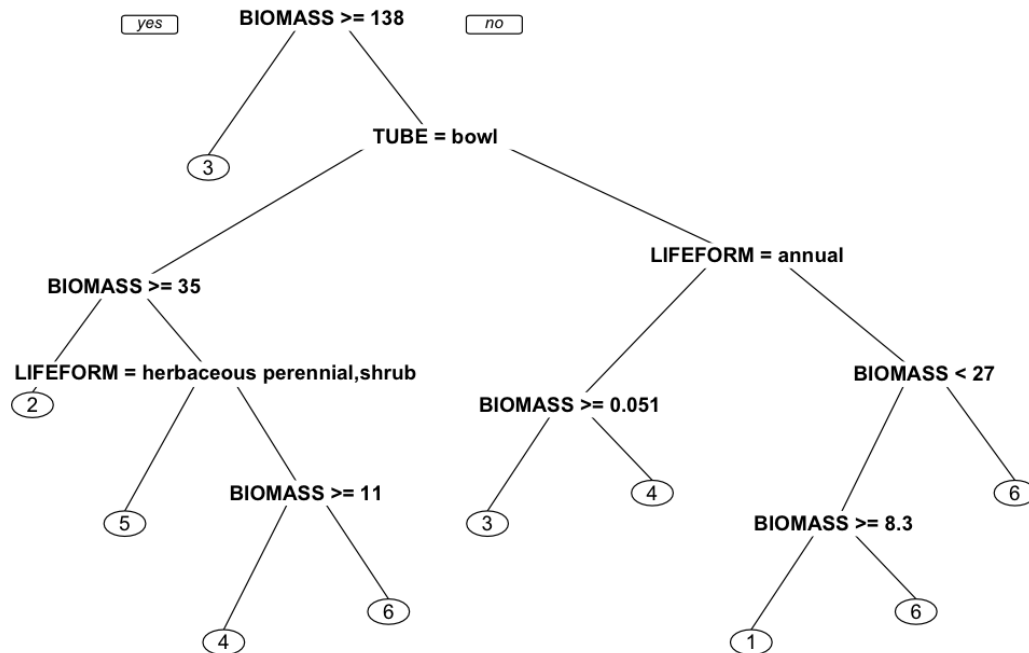


Figure 13. Plant tree produced by lifeform, biomass, and tube type. Correctly places 84% of species.

Module	Plants	Pollinators
1	<ul style="list-style-type: none"> • 27>Biomass≥8.3 • Lifeform = Bulb, herbaceous perennial, Taprooted perennial, Succulent, Shrub 	<ul style="list-style-type: none"> • NA
2	<ul style="list-style-type: none"> • 138>Biomass≥35 • Tube = bowl 	<ul style="list-style-type: none"> • Taxon guild = Colletid bee, Meloid beetle, Mordellid beetle, spider, Vespidae hornet • Taxon guild = Halictid bee, Lepidopteran, long-lived social bee, Muscoid fly, shorter-tongued bee fly, Sphecoid wasp • 104>Biomass≥5.2 • 4.5>Tongue length decile • Taxon guild = Halictid bee, Lepidopteran, long-lived social bee, Muscoid fly, shorter-tongued bee fly, Sphecoid wasp • 250>Biomass≥104 • Taxon guild = Ant, hover-fly, hummingbird, sciarid fly • 180>Biomass≥45
3	<ul style="list-style-type: none"> • Biomass≥138 	<ul style="list-style-type: none"> • Taxon guild = Ant, hover-fly, hummingbird, sciarid fly • 45>Biomass
	<ul style="list-style-type: none"> • 138>Biomass≥0.051 • Tube = closed, thin, wide, very wide • Lifeform = annual 	<ul style="list-style-type: none"> • Taxon guild = Ant, hover-fly, hummingbird, sciarid fly • 180≤Biomass
4	<ul style="list-style-type: none"> • 35>Biomass≥11 • Lifeform = annual, bulb, taprooted perennial, succulent • Tube = bowl 	<ul style="list-style-type: none"> • Taxon guild = Halictid bee, Lepidopteran, long-lived social bee, Muscoid fly, shorter-tongued bee fly, Sphecoid wasp • 5.2>Biomass • Taxon guild = Halictid bee, Lepidopteran, long-lived social bee, Muscoid fly, shorter-tongued bee fly, Sphecoid wasp • 104>Biomass≥5.2 • Tongue length decile≥4.5
	<ul style="list-style-type: none"> • 0.051>Biomass • Lifeform = annual • Tube = closed, thin, wide, very wide 	<ul style="list-style-type: none"> • Taxon guild = Megachilid bee • 77>Biomass • Taxon guild = Halictid bee, Lepidopteran, long-lived social bee, Muscoid fly, shorter-tongued bee fly, Sphecoid wasp • 250≥Biomass
5	<ul style="list-style-type: none"> • 35>Biomass • Tube=Bowl • Lifeform = herbaceous perennial, shrub 	<ul style="list-style-type: none"> • NA
6	<ul style="list-style-type: none"> • 11>Biomass • Lifeform = annual, bulb, taprooted perennial, succulent • Tube = bowl 	<ul style="list-style-type: none"> • Taxon guild = Long-tongued bee fly, conopid fly
	<ul style="list-style-type: none"> • 8.3>Biomass • Lifeform = bulb, herbaceous perennial, taprooted perennial, succulent, shrub • Tube = closed, thin, wide, very wide 	<ul style="list-style-type: none"> • Taxon guild = Megachilid bee • 77>Biomass
	<ul style="list-style-type: none"> • 27≥Biomass • Lifeform = bulb, herbaceous perennial, taprooted perennial, succulent, shrub • Tube = closed, thin, wide, very wide 	

Figure 14. Table which shows sets of plant and pollinator traits by module. Sets of plant and pollinator traits were taken from the trees in Fig. 12 and Fig. 13.