

coNet: Create network models in R using co-occurrence data

24 April 2019

Introduction

- Network models are one way to represent systems (???).
- It's not always possible to directly observe relationships (i.e. ecological interactions) (???).
- In ecology, space can be used as a substitute for time (???).
- This assumes that variance of conditions and processes are minimally impactful (???).
- This method implements an adaptation of the method developed by Araujo et al. 2011 expanded with the application of Bayesian probability.

Methods

Overview

1. Calculate pairwise conditional probabilities $P(S_i|S_j) = \frac{P(S_i, S_j)}{P(S_j)}$
2. Calculate variances $V(S_i S_j) = N * E[P(S_i, S_j)] * (1 - E[P(S_i, S_j)])$
3. Confidence interval $CI_{95\%} = E[S_i S_j] * Z_{95\%} * \sqrt{V(S_i S_j)}$
4. Edge removal $If CI_l \leq [S_i S_j] \leq CI_u, then P(S_i|S_j) = P(S_i, S_j)$
5. Re-scaling $D_{ij} = P(S_i|S_j) - P(S_i, S_j)$

Pairwise Conditional Probabilities

- Uses the conditional probabilities of each species pair, i.e. the probability of observing one species given an observation of another species $P(S_i|S_j)$, based on the method developed by (???).
- To calculate conditional probabilities, we quantified the individual probabilities of species occurrences (i.e. $P(S_i)$ or $P(S_j)$) and the joint probability of co-occurrences $P(S_i, S_j)$ using the frequencies of each species and their co-occurrences. We were then able to calculate the conditional probabilities of each species pair as $P(S_i|S_j) = \frac{P(S_i, S_j)}{P(S_j)}$, based on the axioms of probability.
- This yielded a matrix that could possibly be asymmetric, i.e. $P(S_i|S_j)$ does not have to be equal to $P(S_j|S_i)$.
- Another property of this matrix is that the diagonal (S_{ii}) is equal to one for all species present and zero for species that were not observed in any cell.

Edge Removal and Re-scaling

- We then applied an analytical procedure to remove non-significant links between species. This procedure determines if the joint probability of a species pair (i.e. $P(S_i, S_j)$) is different from zero (Fig.-1).
- Here, a confidence interval $CI_{95\%}$ is calculated as $CI_{95\%} = E[S_i S_j] * Z_{95\%} * \sqrt{V(S_i S_j)}$, where the expected frequency of co-occurrences $E(S_i S_j)$ is the total number of cells surveyed (N) times the independent probabilities of each species $P(S_i) * P(S_j)$, $Z_{95\%}$ is the Z-score for 95% from a Z-distribution and the expected variance of $E(S_i S_j)$ is the total number of cells times the expected probability of $S_i S_j$ and its compliment (i.e. $V(S_i S_j) = N * E[P(S_i, S_j)] * (1 - E[P(S_i, S_j)])$).

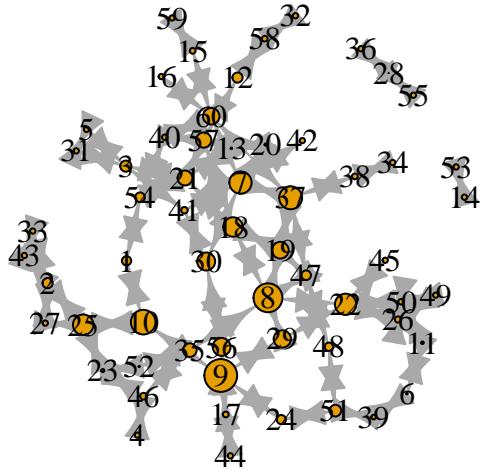
- If the observed number of co-occurrence falls outside of the confidence interval, the joint probability $P(S_i, S_j)$ is determined to be equal to the product of the individual probabilities (i.e. $P(S_i)\hat{P}(S_j)$), and the conditional probability reduces to the individual probability of that species $P(S_i)$.
- Therefore, unless the co-occurrence of a species pair falls outside the confidence interval, the probability that the observation of one species given the other is no different than simply observing that species alone.
- This enables us to remove links from a given network by re-scaling the resulting conditional probabilities by subtracting the individual probabilities from the conditional probabilities (i.e. how different the conditional probability is from the independent probability), which makes any species with a non-significant conditional probability zero.
- The resulting matrix ($\mathbf{D} = D_{ij}$) can be interpreted as how one species impacts another with zero being no effect and values less than or greater than zero interpreted as negative and positive effects, respectively.
- Here, we will refer to this matrix (\mathbf{D}) as an interaction matrix with the properties that it can be asymmetric (i.e. P_{ij} does not necessarily equal P_{ji}), and the diagonal (P_{ii}) is zero (i.e. a species does not influence its own probability of being observed).

Results

A simple example using data from Keith et al. 2010.

- Arthropods sampled on individual trees in a common garden
- Import data

```
dat <- read.csv("../data/arth09.csv")
##
dat <- dat[, -1:-2]
dat[is.na(dat)] <- 0
dat[dat != 0] <- 1
##
pb.cn <- coNet(dat, ci.p = 99)
pb.cn <- pb.cn[apply(abs(pb.cn), 1, sum) > 0, apply(abs(pb.cn), 2, sum) > 0]
spp <- rownames(pb.cn)
rownames(pb.cn) <- colnames(pb.cn) <- 1:nrow(pb.cn)
pb.ig <- igraph::graph_from_adjacency_matrix(abs(pb.cn), mode = "directed", weighted = TRUE)
pb.btw <- betweenness(pb.ig, normalized = TRUE)
names(pb.btw) <- spp
plot(pb.ig, vertex.size = (pb.btw*100)^(0.85),
      vertex.label.color = "black",
      vertex.boarder.color = "white",
      arrow.size = 0.1)
```



Discussion

Citations

Acknowledgments

References

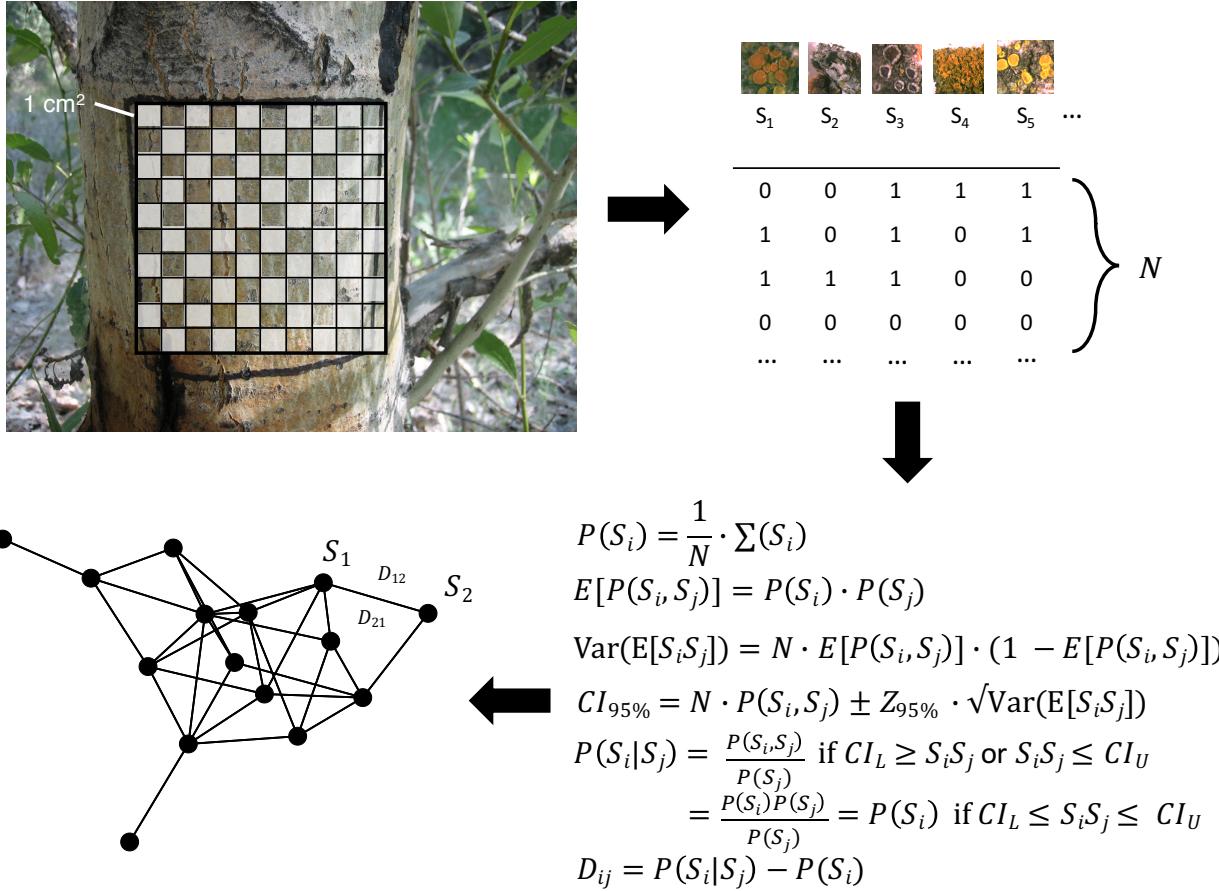


Figure 1: Lichen interaction networks were constructed by conducting field observations in 1 cm² cells within a 10 cm² grid on each tree using a checkerboard pattern (grey cells). Thus, a set of N total cell observations were recorded for each tree with the presence or absence of each species recorded for each cell. Applying the probability-based network modeling method adapted from [Araujo2011], we calculated the conditional probabilities, $P(S_i | S_j)$, for all species pairs and removed (i.e. set equal to zero) species pairs whose joint probabilities, $P(S_i, S_j)$, were not significant using a confidence interval based comparison of their observed co-occurrence frequency, $S_i S_j$, to that expected due to chance alone, $E[P(S_i, S_j)] = P(S_i)P(S_j)$, and $P(S_i | S_j)$ reduces to $P(S_i)$, the observed individual probability of species S_i .