

# Co-occurrence Network Modeling

M.K. Lau

May 23, 2012

## Contents

<b>1</b>	<b>Summary</b>	<b>1</b>
<b>2</b>	<b>Code</b>	<b>3</b>
<b>3</b>	<b>Example</b>	<b>5</b>

## 1 Summary

The method presented in Araújo et al. 2011 [1] is aimed at resolving networks of co-occurrence patterns from community datasets. Although it is framed within a biogeographic context, it has broader applicability to datasets with repeated samples of species presence-absences. The method uses a probability framework where the observed joint probabilities of species are tested against a null model for their joint probabilities, which is the product of their individual probabilities. The process for two species (a and b) given the number of occurrences ( $n$ ) and a total number of observations ( $N$ ) is based on calculating:

1. the individual probabilities:

- $P(a) = \frac{n_a}{N}$
- $P(b) = \frac{n_b}{N}$

2. the expected joint probability (i.e. the null expectation):

- $E[P(a, b)] = P(a) \cdot P(b)$

3. the observed and expected number of co-occurrences:

- $n_{ab} = \frac{n_{ab}}{N}$
- $E[n_{ab}] = N \cdot E[P(a, b)]$

4. the variance of the expected number of co-occurrences:

- If  $E[n_{ab}] \sim \text{bin}(N, E[P(a, b)])$ , then:
- $s^2 = N \cdot E[P(a, b)] \cdot (1 - E[P(a, b)])$  [3]

5. a confidence interval for the expected number of co-occurrences:

- $CI = E[n_{ab}] \pm 2 \cdot \sqrt{s^2}$
- Note: this is a parametric approach, using either a  $t$  or  $\chi^2$  distribution for the statistic and relying on the central limit theorem

6. And if  $n_{ab}$  falls outside of  $CI_{E[n_{ab}]}$ , then the co-occurrence pattern is significant (i.e. non-zero)

7. Inference uses the Bray-Curtis dissimilarity of all pairs of species with significant co-occurrence patterns for network inference.

8. The network is further pruned using a percolation threshold procedure. Note: the justification for this is weak.

The paper goes on to detail several statistics that can be calculated from the empirical co-occurrence networks, including:

- Symmetry
- Degree distribution
- Strength distribution
- Species degree
- Species strength (in and out)

## 2 Code

```
> araujoNet <- function(x,method='bray',min.abundance=1){  
+                               #get the absent species names  
+   y <- colnames(x)[apply(x,2,sum) >= min.abundance]  
+                               #remove low abundance species  
+   x <- x[,apply(x,2,sum) >= min.abundance]  
+                               #assure presence-absence matrix  
+   x[x!=0] <- 1  
+                               #warn if matrix is empty  
+   if (ncol(x) <= 1){warning('Community matrix is empty',quote=FALSE)}else{}  
+                               #caluclate the Bray-Curtis dis.  
+   d <- as.matrix(vegdist(t(x),method=method))
```

```

+                                     #calculate individual prob.
+   Pa <- apply(x,2,function(x) length(x[x!=0])/length(x))
+                                     #calculate null joint probability
+   Pab <- array(NA,dim=c(length(Pa),length(Pa)))
+   rownames(Pab) <- colnames(Pab) <- names(Pa)
+   for (i in 1:nrow(Pab)){
+     for (j in 1:ncol(Pab)){
+       Pab[i,j] <- Pa[i]*Pa[j]
+     }
+   }
+                                     #Calculate confidence limits
+                                     #number of co-occurrences
+   N1 <- nrow(x) * Pab
+                                     #calculate the variance
+   V1 <- nrow(x) * Pab * (1-Pab)
+                                     #+/- 2 SD limits (~95% confidence)
+   cl.u <- N1 + 2*sqrt(V1)
+   cl.l <- N1 - 2*sqrt(V1)
+                                     #observed number of co-occurrences
+   Nab <- N1 * 0
+   for (i in 1:nrow(Nab)){
+     for (j in 1:ncol(Nab)){
+       Nab[i,j] <- length(x[x[,i] != 0 & x[,j] != 0,i])
+     }
+   }
+                                     #prune within confidence limits

```

```

+   dp <- d * 0
+   dp[Nab > cl.u] <- d[Nab > cl.u]
+   dp[Nab < cl.l] <- d[Nab < cl.l]
+
+                                     #add back absent species
+   x.names <- c(colnames(x),y)
+   d <- cbind(d,array(0,dim=c(nrow(d),length(y))))
+   d <- rbind(d,array(0,dim=c(length(y),ncol(d))))
+   rownames(d) <- colnames(d) <- x.names
+   dp <- cbind(dp,array(0,dim=c(nrow(dp),length(y))))
+   dp <- rbind(dp,array(0,dim=c(length(y),ncol(dp))))
+   rownames(dp) <- colnames(dp) <- x.names
+
+                                     #pack for export
+   out <- list(x=x,d=d,dp=dp)
+   return(out)
+ }
>

```

### 3 Example

```

> library(vegan)
> library(sna)

```

Tools for Social Network Analysis

Version 2.2-0 created on 2010-11-21.

copyright (c) 2005, Carter T. Butts, University of California-Irvine

Type `help(package="sna")` to get started.

```

> data(dune)
> test <- araujoNet(dune)
> deg <- degree(test$dp[apply(test$dp,1,sum)!=0,apply(test$dp,2,sum)!=0]
+           ,rescale=FALSE)
> gplot(test$dp[apply(test$dp,1,sum)!=0,apply(test$dp,2,sum)!=0],
+       displaylabels=TRUE,gmode='graph',label.cex=0.65,
+       vertex.sizes=50,vertex.col='black',edge.col='darkgrey',
+       vertex.cex=deg,edge.lwd=(test$dp[apply(test$dp,1,sum)!=0,
+                                           apply(test$dp,2,sum)!=0]+1)^2)
>

```

The output from this function is currently a list of three matrices. The first (x) is the presence-absence matrix used to generate the co-occurrence network. The second (d) is the full Bray-Curtis dissimilarity matrix. The third (dp) is the Bray-Curtis dissimilarity matrix with the non-significant dissimilarities equal to zero. This last matrix can easily be plotted using the *gplot* function in the **sna** package [2] (Fig. 1).

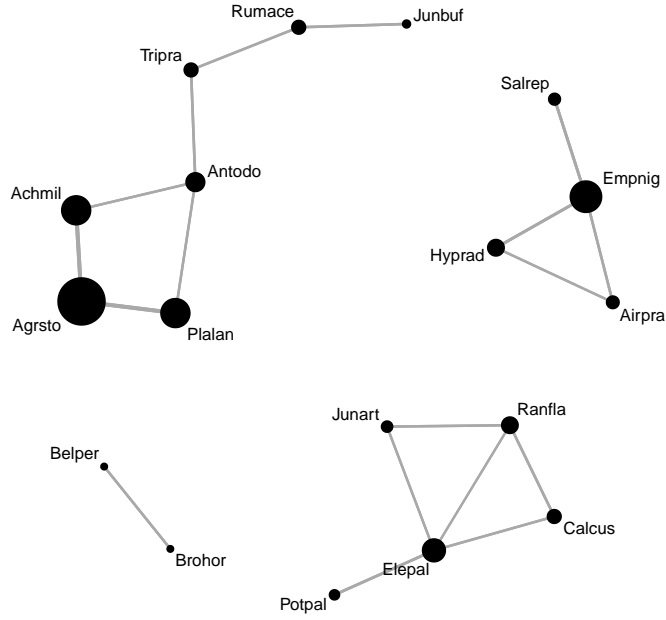


Figure 1: Co-occurrence network for the *dune* plant dataset in the `vegan` package [4]. Species that have no significant co-occurrence patterns are not shown. Lines are the Bray-Curtis dissimilarities for pairs of species with significant joint probabilities. Species nodes and edges are scaled by their degrees and squared dissimilarities, respectively. Network figure generated using the `gplot` function in the `sna` package [2].

## References

- [1] M.B. Araújo, A. Rozenfeld, C. Rahbek, and P.A. Marquet. Using species co-occurrence networks to assess the impacts of climate change. *Ecography*, 34:897–908, 2011.
- [2] Carter T. Butts. *sna: Tools for Social Network Analysis*, 2010. R package version 2.2-0.
- [3] M.A. McCarthy. *Bayesian Methods for Ecology*. Cambridge University Press, Cambridge, U.K., 2007.
- [4] Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, and Helene Wagner. *vegan: Community Ecology Package*, 2012. R package version 2.0-3.