Letter to the Editor

Newly proposed node segregation index is not suitable for analyzing unimode food webs

CrossMark

ABSTRACT

Strona and Veech (2015) developed a new node segregation (or node overlap) index for analysing ecological network structure based on the Veech (2013)'s species co-occurrence probabilistic model, which was originally applied to species-site matrices. However, a species-site matrix for analysing species co-occurrence patterns and an adjacency matrix for characterising unimode network structures are different. Directly applying Veech's species co-occurrence probabilistic model to adjacency matrices in unimode food webs is problematic. The central critical problem is related to the number of free species (or nodes/vertices) in the unimode network that can be the neighbors (have links to connect) of a focused species or a focused pair of species. This number is typically less than the total number of species in real food webs. That is, species are not independent from each other in unimode networks. For a simple undirected unimode network without self-loops, based on the criterion whether there is a link between two species for a focused pair, a correct probabilistic model is developed to accurately compute the probability of observing some shared neighbors for a pair of species in the network. Numerical simulation show that the node overlap calculated using the correct and original probabilistic models present remarkable differences, especially when a unimode network is nested and contains generalists. In summary, The correct probabilistic model should be used if ones want Strona and Veech (2015)'s node segregation index to work for unimode food webs.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Based on Veech's species co-occurrence probabilistic model (Veech, 2013), a recent study (Strona and Veech, 2015) developed a node segregation index for quantifying the structure in ecological networks. They claimed that their model was suitable for both bipartite and unimode networks. Although the model is appropriate for plant-pollinator bipartite networks, it is actually not applicable to unimode food web structures. Here a bipartite network is a graph whose vertices can be divided into two independent disjoint sets such that one vertex from one disjoint set is allowed to have edges connecting other vertices in another disjoint set (but no edges are allowed to link other vertices of the same set). By contrast, a unimode network is a much free graph without the above constraints imposed on the bipartite networks.

The studied objects in species-site matrices, bipartite networks and unimode networks are different. The studied objects of the original Veech's co-occurrence probabilistic model are species and sites, which are totally independent from each other. As such, the species-site matrix could be rectangular and it is not necessary to have equal column and row numbers. The adjacency matrix characterizing the structure of bipartite networks is similar: plants (or hosts) and pollinators (or parasites) are independent functional groups and the row and column numbers can be different. Here, an

adjacency matrix is defined as a square matrix, the elements of which indicate whether pairs of vertices are connected (i.e., have an edge linking them together, typically recorded as 1) or not (typically recorded as 0) in the network. However, the adjacency matrix characterizing the structure of unimode food webs is quite different: it contains the same set (or at least some) of species in both rows and columns. That is, the adjacency matrix is typically (or can be transformed back to) a square matrix. As such, a direct application of Veech's co-occurrence probabilistic model in analyzing unimode network structure in Strona and Veech (2015) deserves further investigation.

Later I will point out problems of directly applying the original Veech's co-occurrence probabilistic model to analyze undirected unimode networks. The same problems are applied to directed unimode food webs because they can be divided to indegree and outdegree sub-networks (each is actually an undirected network). Here, indegree and outdegree are defined as the number of head endpoints and tail endpoints adjacent to a focused node respectively. In graph theory, a simple undirected graph contains no self-loops. That is, the diagonal elements of the adjacency matrix must be always zero. For ecological food webs, self-loops are possible for primary producers or species in other trophic levels because they don't need to feed on other species or have interactions with conspecifics. However, even for non-simple undirected networks with self-loops, the probability of observing some common neighbors (here a neighbor is defined as a species

that has a link or edge to connect the focused species) for a pair of species is still different from the one computed using Veech's co-occurrence probabilistic model. Details are presented in the following section.

2. Problems of directly applying Veech's model

For the original Veech's species co-occurrence probabilistic model, the calculation formula for the co-occurrence probability of a pair of species over n sites is given by (Strona and Veech, 2015; Veech, 2013),

$$p_{\text{Veech}}(k|n,d_1,d_2) = \frac{\binom{n}{k}\binom{n-k}{d_1-k}\binom{n-d_1}{d_2-k}}{\binom{n}{d_1}\binom{n}{d_2}} \tag{1}$$

where k is the number of shared sites for both species; d_1 and d_2 are the number of sites where species 1 and 2 are observed to occur respectively.

When this model is applied to networks, the definitions of corresponding parameters are needed to change: n should redefined as the total number of species in the network, d_1 and d_2 become the number of neighbors that species 1 and 2 are observed to have in the network, and k is the number of shared common neighbors for both species.

Now I show the errors of directly applying the original Veech's probabilistic model as did by Strona and Veech (2015) in the unimode network setting: in a simple undirected two-species food web ($n=2$), each species is assumed to have one neighbor ($d_1 = 1, d_2 = 1$). Among possible configurations in this simplest network (Fig. 1), for the cases without (Fig. 1A) and with the allowance of self-loops (Fig. 1B), there are no common neighbors for both species: the possible neighbors of species 1 are species 2 only for the case without self-loops (Fig. 1A) or species 2 and itself for the case with self-loops (Fig. 1B). Analogously, the possible neighbors of species 2 can be species 1 only for the case without self-loops (Fig. 1A) or species 1 and itself for the case with the allowance of self-loops (Fig. 1B). No matter whether self-loops are allowed, the correct probability for observing some shared

neighbors ($k > 0$) for both species should be always 0 and the probability of observing no neighbors ($k = 0$) should be always 1.

However, if one directly applies the original Veech's Eq. (1) as did by Strona and Veech (2015), the probability that both species in Fig. 1 share a common neighbor, given that both are observed to have one neighbour, is $p_{\text{Veech}}(k=1|n=2,d_1=1,d_2=1)=0.5$. Moreover, the probability that both species share no neighbors using Veech's model is $p_{\text{Veech}}(k=0|n=2,d_1=1,d_2=1)=0.5$. These two probabilities are wrong, because it is not possible to share a common neighbor in the two-species network (Fig. 1). The analyses on the outdegree sub-network (the same to indegree sub-network) of two-species directed network (self-loops can be allowed or not allowed; Fig. 2) were identical: the probability that both species in Fig. 2 share no neighbors or a common neighbor, given that both are observed to have an outdegree, should be always 0, not 0.5.

So why is Veech's species co-occurrence model wrong? The key reason is that two species of the focused pair are not independent from each other and other species in the unimode network. Using Fig. 1 as an example, to fulfill the requirement of $n = 2, d_1 = 1$, and $d_2 = 1$, if self-loops are not allowed, species 1 must have a link connecting species 2. If self-loops are allowed, species 1 can be a neighbor of itself (the last configuration in the lower right panel of Fig. 1), but in this situation, species 2 is not allowed to connect species 1 anymore (otherwise species 1 will have two neighbors: itself and species 2, leading to $d_1 = 2$). In comparison, when applying Veech's model to a two species-two site matrix, under the constraint of $n = 2, d_1 = 1$, and $d_2 = 1$, four scenarios are possible in which each species is allowed to occupy a site independently (Fig. 3). The corresponding probability of sharing one site for both species is 0.5, being equal to the value computed using Veech's Eq. (1).

Based on the above discussion, it is misleading to directly apply Veech's model (1) in food webs with complex trophic levels. Veech's model was originally developed for species-site matrices in which species and sites are totally independent from each other. It is appropriate to be applied to bipartite networks, but for unimode food webs, the model becomes problematic. Therefore, It is necessary to propose a correct probabilistic model for such a situation and use it to replace Veech's Eq. (1) if ones want the
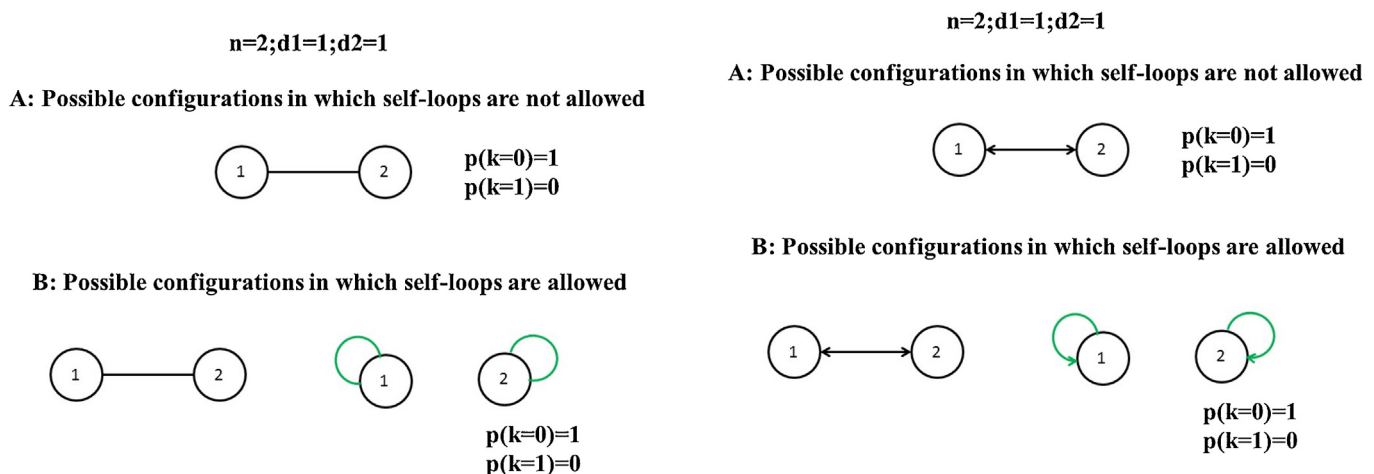
n=2;d1=1;d2=1

A: Possible configurations in which self-loops are not allowed

B: Possible configurations in which self-loops are allowed

Fig. 1. Possible structures for a two-species simple undirected food web (in which each species has a neighbor: $d_1 = 1; d_2 = 1$) for the cases that self-loops are inhibited (A) and permitted (B), respectively. k indicates the number of shared neighbors. An undirected self-loop (denoted by a green curved line) for a node in the last configuration (lower right panel) could be recognized as a neighbor of itself.

n=2;d1=1;d2=1

A: Possible configurations in which self-loops are not allowed

B: Possible configurations in which self-loops are allowed

Fig. 2. Possible structures for a two-species directed outdegree food web (in which each species has an outdegree: $d_1 = 1; d_2 = 1$) for the cases that self-loops are inhibited (A) and permitted (B), respectively. k indicates the number of shared neighbors that both species have outward links to connect. A directed self-loop (denoted by a green curved arrow) for a node (1 or 2) in the last configuration (lower right panel) could be recognized as an outdegree for the node (or an indegree for indegree network).