



# Integrating species and interactions into similarity metrics: a graph theory-based approach to understanding community similarity

Daniela N. López<sup>1,2</sup>, Patricio A. Camus<sup>3,4</sup>, Nelson Valdivia<sup>5,6</sup> and Sergio A. Estay<sup>1,7</sup>

<sup>1</sup>Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

<sup>2</sup>Programa de Doctorado en Ciencias mención Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

<sup>3</sup>Departamento de Ecología, Facultad de Ciencias, Universidad Católica de la Santísima Concepción, Concepción, Chile

<sup>4</sup>Centro de Investigación en Biodiversidad y Ambientes Sustentables (CIBAS), Universidad Católica de la Santísima Concepción, Concepción, Chile

<sup>5</sup>Instituto de Ciencias Marinas y Limnológicas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

<sup>6</sup>Centro FONDAP de Investigación en Dinámica de Ecosistemas Marinos de Altas Latitudes (IDEAL), Universidad Austral de Chile, Valdivia, Chile

<sup>7</sup>Center of Applied Ecology and Sustainability (CAPES), Pontificia Universidad Católica de Chile, Santiago, Chile

## ABSTRACT

Community similarity is often assessed through similarities in species occurrences and abundances (i.e., compositional similarity) or through the distribution of species interactions (i.e., interaction similarity). Unfortunately, the joint empirical evaluation of both is still a challenge. Here, we analyze community similarity in ecological systems in order to evaluate the extent to which indices based exclusively on species composition differ from those that incorporate species interactions. Borrowing tools from graph theory, we compared the classic Jaccard index with the graph edit distance (GED), a metric that allowed us to combine species composition and interactions. We found that similarity measures computed using only taxonomic composition could differ strongly from those that include composition and interactions. We conclude that new indices that incorporate community features beyond composition will be more robust for assessing similitude between natural systems than those purely based on species occurrences. Our results have therefore important conceptual and practical consequences for the analysis of ecological communities.

**Subjects** Biodiversity, Ecology

**Keywords** Community similarity, Graph edit distance (GED), Ecological communities

## INTRODUCTION

Characterizing the degree of similarity between ecological communities has been one of the central topics in ecology (*Devictor et al., 2010; Morlon, Kefi & Martinez, 2014; Petchey & Gaston, 2006*). In most studies, ecological similarity has mainly been evaluated by

Submitted 10 August 2018

Accepted 25 April 2019

Published 31 May 2019

Corresponding author

Daniela N. López,  
danyelalopez@gmail.com

Academic editor

Mark Costello

Additional Information and  
Declarations can be found on  
page 15

DOI 10.7717/peerj.7013

© Copyright  
2019 López et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

comparing species compositions, quantifying the spatial turnover of species through cluster membership of communities and by examining the position and distance between communities in reduced species-ordination spaces (Jaillard *et al.*, 2018; Olden *et al.*, 2004). In this vein,  $\beta$ -diversity portrays the variation in species composition among spatially or temporally separated communities (Anderson, Ellingsen & McArdle, 2006; Whittaker, 1972; Whittaker, 1960), and thus reflects two different phenomena, spatial species turnover and nestedness of assemblages (Baselga, 2012; Baselga, 2010).  $\beta$ -diversity can be easily measured from species presence-absence data (Koleff, Gaston & Lennon, 2003; Wilson & Shmida, 1984) or abundances (Barwell, Isaac & Kunin, 2015), and both approaches include some widely studied indices such as Euclidean distances, Bray-Curtis, and Jaccard (Anderson, Ellingsen & McArdle, 2006). Despite the widespread use of  $\beta$ -diversity and other indices to characterize communities, the quantification of similarity among ecological communities, beyond species compositions remains a challenge.

In an analogous way, when ecological communities are represented by interaction networks, the similarity between spatially or temporally separated communities has been evaluated on the basis of the average number of shared nodes (species) and the degree of node overlapping (i.e., node overlapping index; Strona & Veech, 2015; Strona *et al.*, 2018; Zhang *et al.*, 2016). These measures assume that a high compositional similarity implies a high similarity in species interactions (i.e., a high “interaction similarity”). On the other hand, it has been described that species composition can be a driver of interaction turnover in mutualistic networks (Bezemer *et al.*, 2010; Trøjelsgaard *et al.*, 2015), but there is also evidence based on traditional methods ( $\beta$ -diversity) that shows a complete lack of correlation between composition similarity and interaction similarity (Poisot *et al.*, 2012). Therefore, the relationship between composition and interaction similarity remains unclear.

Graph theory provides us with conceptual and practical tools that allow us to integrate the composition and interaction measurements of similarities of systems (Ibragimov *et al.*, 2013; Riesen, 2015). These similarity measures include global metrics based on isomorphic relations or on graph transformations (Bunke & Allermann, 1983; Dehmer, 2010; Emmert-Streib, Dehmer & Shi, 2016; Solé-Ribalta, Serratos & Sanfeliu, 2012). Similarity measures based on isomorphic relationships, quantify the exact match of nodes and links between two graphs, i.e., the “exact graph matching”. Graph transformation, on the other hand, uses the concept of error-tolerant graph matching to calculate a measure of similarity based on the minimum cost (i.e., a unit of dissimilarity) of transforming one graph into another (Dehmer, 2010; Riesen, 2015). A widely used graph transformation method is the graph edit distance (GED), in which each transformation has a cost, so that a greater number of changes mirrors higher dissimilarity between the analyzed networks (Bunke & Allermann, 1983; Emmert-Streib, Dehmer & Shi, 2016). In biological sciences, GED has been used to compare protein-protein interaction (PPI) networks in human, yeast, and fruit fly, allowing researchers to infer the biological function of proteins and genes (Ibragimov *et al.*, 2013; Neyshabur *et al.*, 2013). By means of applying these tools to ecological data we can obtain indices that include both species composition and species interactions, providing a more complete view of community similarity.

In this study we apply conceptual tools borrowed from graph analysis to quantify community similarity, taking into account both species composition and species interactions. In addition, we compare the inferences obtained from this approach with those obtained using the classical compositional approach in order to evaluate the degree to which interaction similarity can be inferred from compositional similarity. These analyses were applied to an extensive consumer-resource interaction network.

## MATERIALS AND METHODS

### Biological data

We analyzed a dataset ([https://www.nceas.ucsb.edu/interactionweb/html/thomps\\_towns.html](https://www.nceas.ucsb.edu/interactionweb/html/thomps_towns.html)) of consumer-resource interactions obtained from the National Center for Ecological Analysis and Synthesis (NCEAS). The dataset includes 16 communities (Akatore A, Berwick, Blackrock, Broad, Canton, Catlins, DempSp, German, Kyeburn, Narrowdale, NorthCol, Powder, Stony, SuttonSp, and Venlaw), covering ca. 200 km of the Taieri River tributaries in Otago, New Zealand. These sites include pine forest, broadleaf forest, pasture grassland, and tussock grassland with recorded taxonomic identities of aquatic insects, algae, and fish species and their trophic interactions (Jaarsma *et al.*, 1998; Thompson & Townsend, 2005; Thompson & Townsend, 2003; Thompson & Townsend, 2000; Thompson & Townsend, 1999). The size of these networks varies from 48 to 113 taxonomic identities and from 110 to 832 consumer-resource interactions. We assign an identification code to taxonomic identities to facilitate comparison among different networks. We made some modifications to the network dataset prior to the analyses (see [Data S1](#)), checking scientific names and correcting typographical errors. Finally, similarity measures were calculated for paired combinations of sites in Taieri River (120 paired comparisons for each scenario).

### Graph edit distance (GED): concept and application

We used the graph edit distance (GED) as a metric that includes both compositional and interaction similarities. GED is a widely used graph transformation method in which each transformation (edition) necessary to pass from one network to another has a cost, so that a greater number of changes implies a higher cost, and this mirrors higher dissimilarity between the compared communities (Bunke, 1983; Emmert-Streib, Dehmer & Shi, 2016). This feature of GED represents an advantage over the composition similarity analysis, because it allows the inclusion of interaction similarity in the metrics and to assign different degrees of importance to species or interactions in the network through differential costs for each type of edition (Bunke & Allermann, 1983; Emmert-Streib, Dehmer & Shi, 2016). Despite the fact that our study only deals with trophic interactions, GED could be applied to networks containing different kinds of links, including other ecological interactions like facilitation, competition, and parasitism.

Let us consider two networks represented by the graphs  $g_1 = (V_1, E_1)$  and  $g_2 = (V_2, E_2)$ , where  $V$  is a set of nodes,  $E$  a set of links  $(u, v)$ , where  $u \in V$  is the source node and  $v \in V$  the target node of a directed link (Dehmer, 2010; Riesen, 2015). The idea behind GED includes transforming one graph into another using edit operations ( $e_i$ ) such as deletions, substitutions, and insertions of nodes and links (Bunke & Allermann, 1983; Dehmer, 2010;

*Emmert-Streib, Dehmer & Shi, 2016; Riesen, 2015*). A given transformation is represented by an edit path ( $\lambda$ ), which is a set of edit operations that transform  $g_1$  to  $g_2$ . The set of all possible  $\lambda$  is called  $\nu(g_1, g_2)$  (*Riesen, 2015*). Assuming each edit operation  $e_i$  has an associated cost  $c(e_i)$ , we can assign a relative cost to the  $k^{\text{th}}$  edit path  $\lambda_k$ :

$$C(\lambda_k) = \sum c(e_i), \forall e_i \in \lambda_k$$

With this information we can define GED as the edit path with the minimum cost (*Dehmer, 2010; Ibragimov et al., 2013*),

$$GED = \min[C(\lambda_k)], \forall \lambda_k \in Y(g_1, g_2).$$

The application of GED to the study of trophic networks allows us to evaluate differences in species compositions and those due to the absence of interactions despite species co-occurrences.

### Cost operations

Given the nature of food webs, we considered that the lowest cost was that of deleting or adding an interaction, followed by deleting or adding a species, and finally the most costly edition was that of flipping an interaction, this is a change to the hierarchy of the consumer-resource interactions (*Table 1*, scenarios 30, 35, 36). We considered 49 scenarios of editing costs (8,820 paired comparisons of our networks), which differed from each other in the relative magnitude of the costs of each edition operation  $c(e_i)$ , as shown on *Table 1*. We include two cost scenarios of flipping an interaction (0.25 and 0.75) represent two contrasting scenarios (low and high cost), and 5 levels of costs for deleting/adding an interaction or node. Also, eight scenarios were included specifically to assess how much GED changed by minimizing the importance of trophic interactions (reducing the costs to deleting or adding interactions, *Table 1*).

GEDs were calculated using the software Cytoscape (*Shannon et al., 2003*) with the GEDEVO plugin (*Ibragimov et al., 2013*), which implements an evolutionary algorithm for GED calculation. Due to GEDEVO implementing an evolutionary algorithm for estimating GED, there is no straightforward rule for stopping the iterative process. In this case we used 1,000 iterations without improvements as termination criterion for the minimization of GED. In the case of GEDEVO, GED scores ranged from 0 to 1, where 0 represented maximum similarity and 1 represented perfect dissimilarity. According to *Malek (2015)*, the method implemented in GEDEVO is that the GED of a given edit path is the result of the sum of the GED estimated for each node. The GED forming each node is transformed into a score based on the highest possible GED for a single node in the whole network.

### Compositional similarity and comparison with GED scores

In order to compare the GED results with those from a traditional compositional approach, we first computed the Jaccard index from the presence/absence species data of each network. Jaccard dissimilarity index was computed as  $(a + b)/(a + b + c)$ , where  $a$  is the number of species only present in the first network,  $b$  is the number of species only present in the second network, and  $c$  is the number of species shared between both networks.

**Table 1** Scenarios of editing costs, AMI values of relationship between the similitude in species composition (1-Jaccard) and similarity of interactions (1-GED).

Scenarios	Flipping an interaction	Deleting or adding a species	Deleting or adding an interaction	AMI <sub>(1-GED, 1-Jaccard)</sub>
1	1	1	1	0.032
2	0.25	0	0.25	<b>0.134</b>
3	0.25	0	0.5	0.071
4	0.25	0	0.75	0.064
5	0.25	0	1	0.078
6	0.25	0.25	0	0.047
7	0.25	0.25	0.25	0.074
8	0.25	0.25	0.5	0.062
9	0.25	0.25	0.75	0.073
10	0.25	0.25	1	0.065
11	0.25	0.5	0	0.039
12	0.25	0.5	0.25	0.051
13	0.25	0.5	0.5	0.035
14	0.25	0.5	0.75	0.05
15	0.25	0.5	1	0.062
16	0.25	0.75	0	<b>0.029</b>
17	0.25	0.75	0.25	0.047
18	0.25	0.75	0.5	0.05
19	0.25	0.75	0.75	0.037
20	0.25	0.75	1	0.05
21	0.25	1	0	0.029
22	0.25	1	0.25	0.054
23	0.25	1	0.5	0.055
24	0.25	1	0.75	0.035
25	0.25	1	1	0.041
26	0.75	0	0.25	<b>0.084</b>
27	0.75	0	0.5	0.051
28	0.75	0	0.75	0.024
29	0.75	0	1	0.024
30	0.75	0.25	0	0.018
31	0.75	0.25	0.25	0.074
32	0.75	0.25	0.5	0.058
33	0.75	0.25	0.75	0.044
34	0.75	0.25	1	0.031
35	0.75	0.5	0	<b>0.014</b>
36	0.75	0.5	0.25	0.06
37	0.75	0.5	0.5	0.05
38	0.75	0.5	0.75	0.027
39	0.75	0.5	1	0.037
40	0.75	0.75	0	0.03

*(continued on next page)*

Table 1 (continued)

Scenarios	Flipping an interaction	Deleting or adding a species	Deleting or adding an interaction	AMI <sub>(1-GED, 1-Jaccard)</sub>
41	0.75	0.75	0.25	0.047
42	0.75	0.75	0.5	0.068
43	0.75	0.75	0.75	0.039
44	0.75	0.75	1	0.068
45	0.75	1	0	0.036
46	0.75	1	0.25	0.026
47	0.75	1	0.5	0.062
48	0.75	1	0.75	0.056
49	0.75	1	1	0.04

In our results, the GED and Jaccard scores were expressed in terms of similarity (i.e., 1-GED and 1-Jaccard, respectively). We used the Adjusted Mutual Information metric (AMI) to assess the amount of information shared between GED and Jaccard.

Mutual Information (MI) is a measure from information theory and based on entropy estimations. MI quantifies the mutual dependence between the variables or, in other words, how much information about one variable is possible to obtain by knowing the second variable. MI can theoretically range from 0 to  $\infty$ . So, to standardize the values and make them comparable we used the method from [Vinh, Epps & Bailey \(2010\)](#). In this case AMI is:

$$AMI_{max} = I(U, V) - E\{I(U, V)\} / \max\{H(U), H(V)\} - E\{I(U, V)\}$$

where,  $U$  and  $V$  are two datasets,  $I(U, V)$  represents mutual information,  $E\{I(U, V)\}$  the expected mutual information, and finally  $H(U, V)$  joint entropy, thus:

$$I(U, V) = \sum_{u \in U} \sum_{v \in V} p(u, v) \log p(u, v) / p(u)p(v),$$

$$H(U, V) = - \sum_{u \in U} \sum_{v \in V} p(u, v) \log p(u, v), \text{ where the entropy of a random variable is defined by:}$$

$$H(U) = - \sum p(u) \log p(u) \text{ or}$$

$$H(V) = - \sum p(v) \log p(v),$$

with probability functions  $p(u)$  and  $p(v)$ .

The advantage of AMI is that it allows us to quantify linear and non-linear relationship between variables. AMI scores ranged from 0 to 1, where 0 represented perfect independence (no information is shared between indices) and 1 means both variables contain exactly the same information ([Vinh, Epps & Bailey, 2010](#)). Jaccard index and AMI calculations were performed in the R environment ([R Development Core Team, 2016, Data S2](#)). Jaccard index values were calculated using vegan package for R ([Oksanen et al., 2013](#)) and AMI values were calculated using the function discretize from the R package infotheo ([Meyer, 2014](#)).

We used the same approach to compare dissimilarity of interactions ( $B_{WN}$ ) and the dissimilarity of interactions due to species turnover ( $B_{ST}$ ) as there are defined in

**Table 2** Scenarios of editing cost, AMI values of relationship between the dissimilarity of interactions ( $B_{WN}$ ), dissimilarity of interactions due to species turnover ( $B_{ST}$ ) and similarity of interactions (1-GED).

Scenarios	Flipping an interaction	Deleting or adding a species	Deleting or adding an interaction	AMI (1-GED, $B_{ST}$ )	AMI (1-GED, $B_{WN}$ )
1	1	1	1	~0	0.017
2	0.25	0	0.25	0.014	<b>0.156</b>
3	0.25	0	0.5	0.025	0.099
4	0.25	0	0.75	0.027	0.083
5	0.25	0	1	0.007	0.066
6	0.25	0.25	0	0.001	0.034
7	0.25	0.25	0.25	0.002	0.059
8	0.25	0.25	0.5	<b>0.029</b>	0.102
9	0.25	0.25	0.75	0.019	0.086
10	0.25	0.25	1	0.023	0.071
11	0.25	0.5	0	~0	0.025
12	0.25	0.5	0.25	~0	0.025
13	0.25	0.5	0.5	0.004	0.013
14	0.25	0.5	0.75	0.002	0.056
15	0.25	0.5	1	0.024	0.05
16	0.25	0.75	0	~0	0.018
17	0.25	0.75	0.25	~0	0.019
18	0.25	0.75	0.5	~0	0.020
19	0.25	0.75	0.75	~0	0.021
20	0.25	0.75	1	0.015	0.049
21	0.25	1	0	0.002	0.021
22	0.25	1	0.25	~0	0.014
23	0.25	1	0.5	~0	0.017
24	0.25	1	0.75	~0	0.009
25	0.25	1	1	~0	0.03
26	0.75	0	0.25	~0	0.088
27	0.75	0	0.5	0.006	0.045
28	0.75	0	0.75	0.007	0.032
29	0.75	0	1	0.01	0.009
30	0.75	0.25	0	0.005	0.015
31	0.75	0.25	0.25	0.015	0.068
32	0.75	0.25	0.5	0.009	0.04
33	0.75	0.25	0.75	~0	0.025
34	0.75	0.25	1	0.007	0.037
35	0.75	0.5	0	~0	0.019
36	0.75	0.5	0.25	~0	0.055
37	0.75	0.5	0.5	0.006	0.035
38	0.75	0.5	0.75	~0	0.015
39	0.75	0.5	1	~0	0.023
40	0.75	0.75	0	0.011	0.029

*(continued on next page)*

Table 2 (continued)

Scenarios	Flipping an interaction	Deleting or adding a species	Deleting or adding an interaction	AMI (1-GED, $B_{ST}$ )	AMI (1-GED, $B_{WN}$ )
41	0.75	0.75	0.25	0.015	0.031
42	0.75	0.75	0.5	~0	0.038
43	0.75	0.75	0.75	~0	0.022
44	0.75	0.75	1	~0	0.037
45	0.75	1	0	~0	0.024
46	0.75	1	0.25	~0	0.011
47	0.75	1	0.50	~0	0.032
48	0.75	1	0.75	0.006	0.036
49	0.75	1	1	~0	0.008

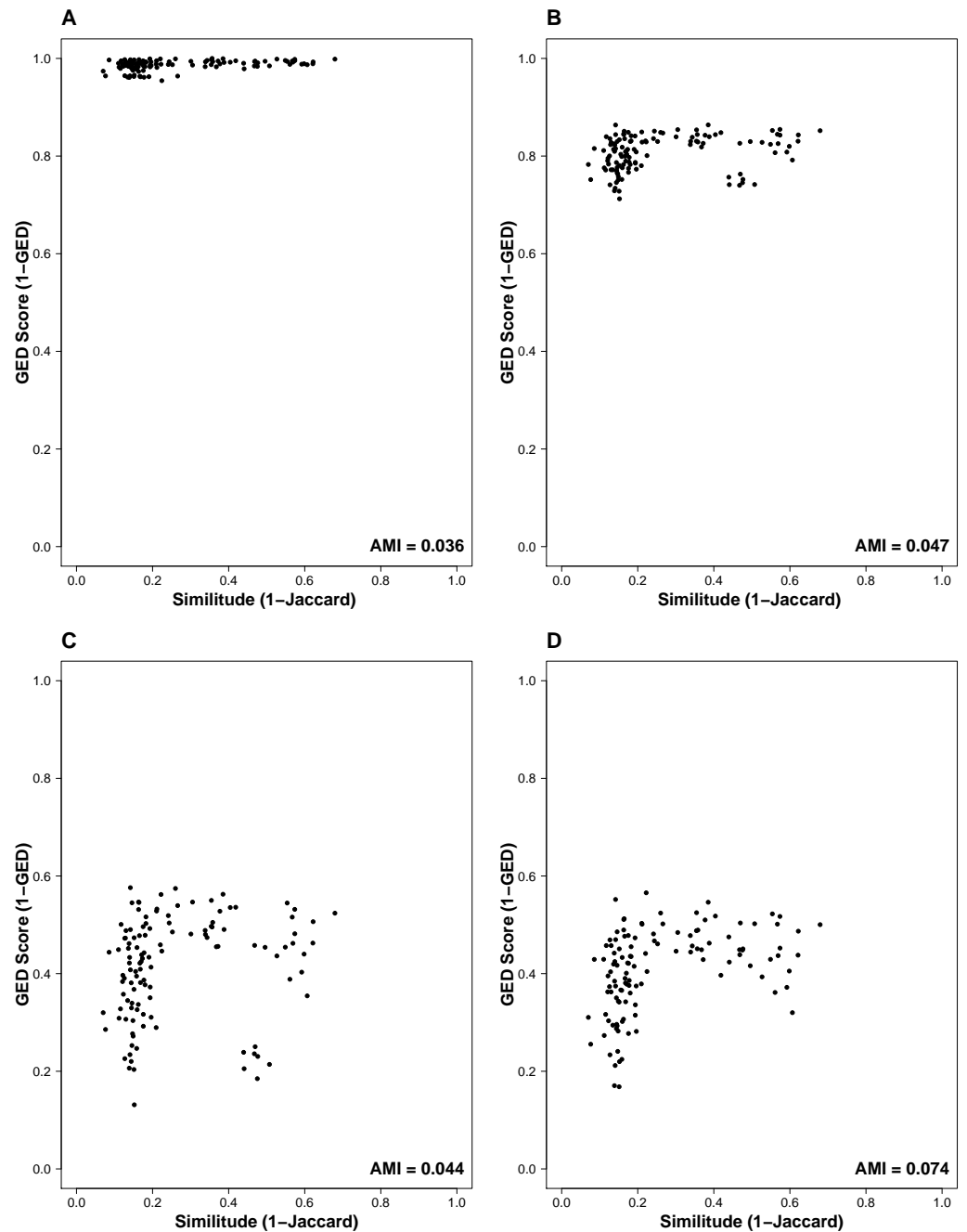
*Poisot et al. (2012)* with the estimated GED values in the 49 scenarios (Table 2).  $B_{WN}$  and  $B_{ST}$  values were calculated using betalink package for R (*Poisot et al., 2012*). Finally, we used the AMI to compare the amount of information shared between GED and  $B_{WN}$ , and between GED and  $B_{ST}$ . The objective of this comparison is to evaluate the relationship of GED with another recent methodological approach to the problem, and also to evaluate the relationship of GED under different cost scenarios with the result obtained by partitioning network similarity into species and interaction  $\beta$ -diversity as is proposed by *Canard (2011)* and *Poisot et al. (2012)*.

## RESULTS

The interaction similarity scores based on the GED ranged from 0.099 to 0.99 for the Taieri River networks. On the other hand, compositional similarity (Jaccard) ranged from 0.08 to 0.67. The highest values of GED were found when the interactions had zero cost (Fig. 1A). We found lower values of GED when the adding/deleting nodes had zero cost and adding/deleting interactions had a low cost (Figs. 1C and 1D, Table 1). A general pattern in the relationship between Jaccard and GED can be observed in Fig. S1. Through almost all scenarios, at low levels of the Jaccard index (<0.2), estimated GED values tend to be highly variable, but at moderate-high values (>0.2) the relationship turns asymptotic (Figs. 1C and 1D, Fig. S1). Due to this relationship, AMI values were low, regardless of the scenario. Estimated values were far from 1, with a maximum of 0.134 for scenario 2 and minimum of 0.014 for scenario 35 (Fig. 2, Table 1). The interpretation of these markedly low AMI values irrespective of the scenario means both metrics seem to be unrelated and do not share significant amounts of information (see Fig. 2). In practical terms, network similarity (nodes and links) cannot be inferred or predicted from similitude at the species level.

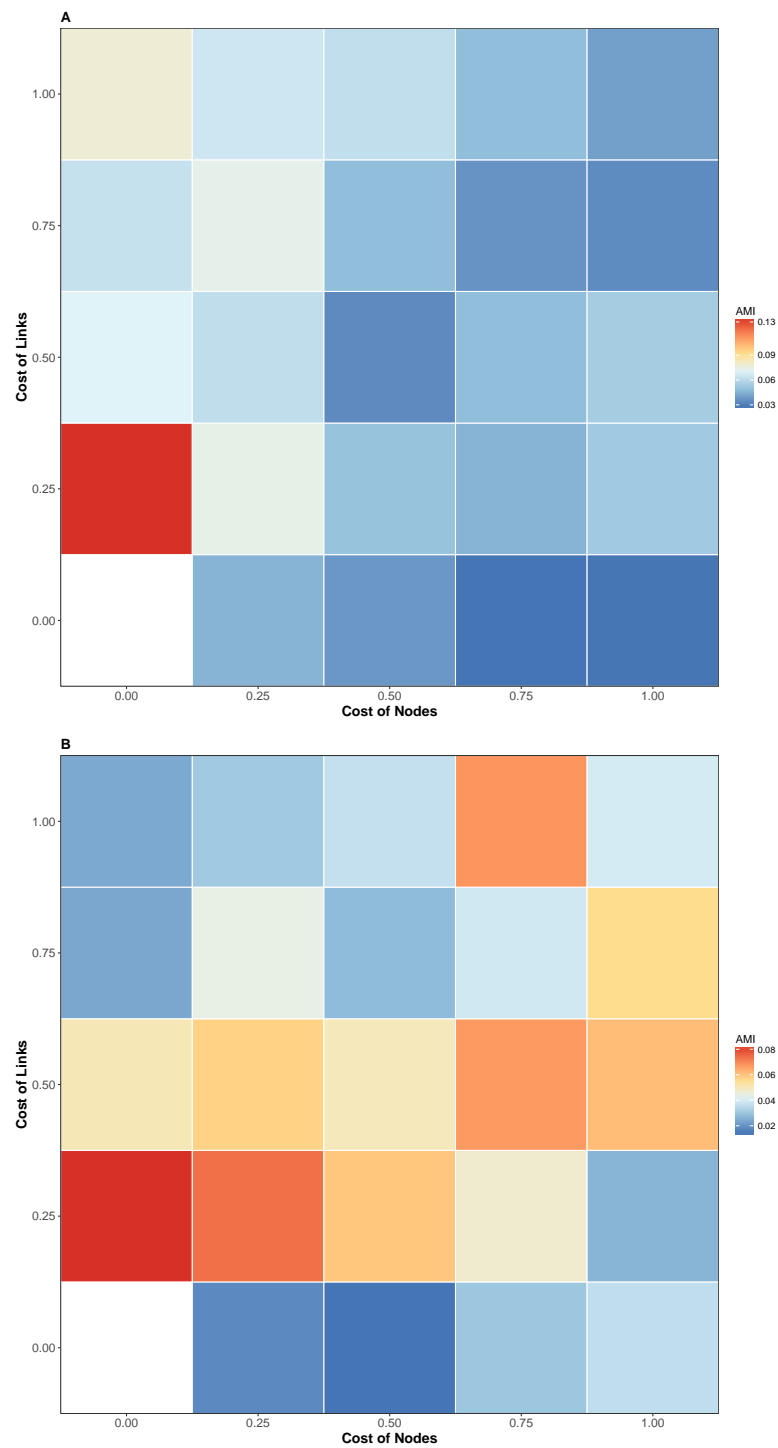
We also found a weak relationship between  $B_{ST}/B_{WN}$  and GED across all scenarios (Fig. 3, Table 2). Despite of this, AMI values were consistently higher for the relationship between  $B_{WN}$  and GED (AMI from 0.008 to 0.156) than between  $B_{ST}$  and GED (AMI from ~0 to 0.029). The maximum AMI value was observed when the cost of adding/deleting nodes was lower than adding/deleting links (Table 2).





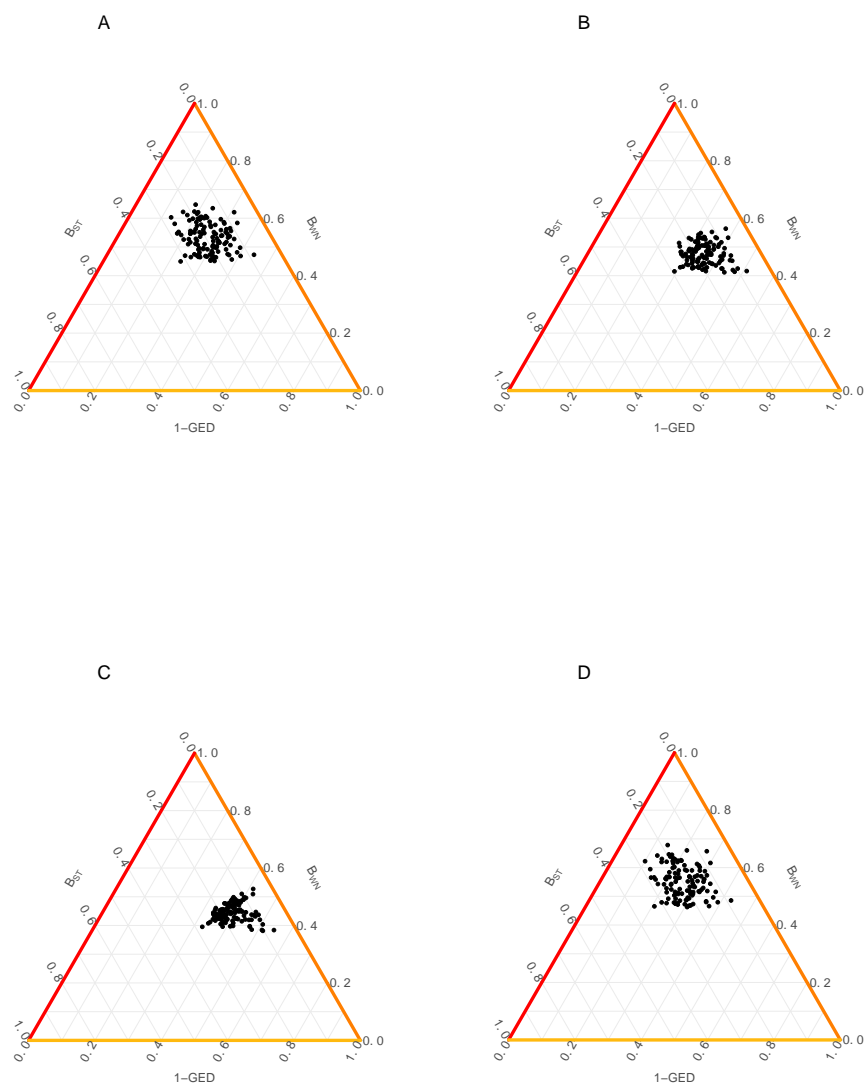
**Figure 1** Relationship between the similitude in species composition (1-Jaccard) and similarity of interactions (1-GED) in grassland systems. (A) Flipping: 0.75, Add or delete nodes: 1, Add or delete links: 0, (B) Flipping: 0.75, Add or delete nodes: 0.75, Add or delete links: 0.25, (C) Flipping: 0.75, Add or delete nodes: 0.25, Add or delete links: 0.75, (D) Flipping: 0.25, Add or delete nodes: 0, Add or delete links: 1. AMI values are included in each plot. Each point represents a pairwise comparison between food webs.

Full-size [DOI: 10.7717/peerj.7013/fig-1](https://doi.org/10.7717/peerj.7013/fig-1)



**Figure 2** AMI values of relationship between the similitude in species composition (1-Jaccard) and similarity of interactions (1-GED) in grassland systems. Color codes shows the AMI values. (A) Flipping 0.25, (B) Flipping 0.75.

Full-size  DOI: [10.7717/peerj.7013/fig-2](https://doi.org/10.7717/peerj.7013/fig-2)



**Figure 3** Relationship among the dissimilarity of interactions due to species turnover ( $B_{ST}$ ), dissimilarity of interactions ( $B_{WN}$ ) and Graph Edit Distance (1-GED) in grassland systems. (A) Flipping: 0.75, Add or delete nodes: 1, Add or delete links: 0,25, (B) Flipping: 0.75, Add or delete nodes: 1, Add or delete links: 0.50, (C) Flipping: 0.75, Add or delete nodes: 0.25, Add or delete links: 0.25, (D) Flipping: 0.75, Add or delete nodes: 0.25, Add or delete links: 1.

Full-size  DOI: 10.7717/peerj.7013/fig-3

## DISCUSSION

We found a weak relationship between community similarities measured using the Jaccard and GED indices; the AMI values were far from those expected for both metrics equally describing the communities. Similar results have been described by other studies in which differences between similarity metrics (composition similarity—interaction similarity) has been explained by phenomena such as changes in phenology ([Edwards & Richardson, 2004](#)), trophic interactions mediated by the presence of a third species, and changes in abundance, behavior, or physiology ([Burkle, Marlin & Knight, 2013](#); [Poiso et al., 2012](#)). Undoubtedly, these phenomena confer variation to communities in different ways, and thus affect the occurrence of both species and trophic interactions. In the same vein, changes in climatic conditions can decouple the synchrony of ecological processes, such as predator–prey interactions ([Harrington, Woiwod & Sparks, 1999](#); [Thackeray et al., 2010](#)). This implies that short-term changes in species' seasonal phenology could result in the loss of some predator–prey interactions, thus impacting interaction similarities. Likewise, mechanisms such as changes in species abundances and predator behavior can jointly affect the trophic similarity of communities ([Abrams, 1982](#); [Arditi & Ginzburg, 2012](#)). Examples of these mechanisms include changes in abundances that affect predator–prey encounter probabilities ([Poiso, Stouffer & Gravel, 2015](#)), and prey behavioral changes causing prey to be less vulnerable given high predator densities ([Charnov, 1976](#); [Skalski & Gilliam, 2001](#)). Thus, all these pieces of evidence suggest that species co-occurrence is necessary but not the only factor that influences the occurrence of ecological interactions ([Lopez et al., 2017](#); [Poiso et al., 2012](#); [Thompson & Townsend, 1999](#)).

[Canard \(2011\)](#) and [Poiso et al. \(2012\)](#) developed a similar approach based on  $\beta$ -diversity measure of dissimilarity, traditionally used to assess community similarity. When comparing these methods with the approach presented here, results showed that both approaches do not seem to share information (as reflected in AMI values). This could suggest that, because of their different origin, both approaches are not measuring exactly the same dimension of community similarity. However, GED values were significantly more related to  $B_{WN}$  than to  $B_{ST}$  despite the cost scenario. This is not surprising because, as it is calculated,  $B_{WN}$  contains to  $B_{ST}$ , so this apparent pattern is just a consequence of  $B_{ST}$  being a subset of  $B_{WN}$ . In this context, we think that a valuable advantage of using GED is the flexibility of assigning different degrees of importance to species or interactions in the network through differential costs for each type of edition ([Bunke & Allermann, 1983](#); [Emmert-Streib, Dehmer & Shi, 2016](#)). This characteristic of GED allows researchers to focus the analysis into the specific components of their interest.

On the other hand, GED is a measure of the similarity between pairwise networks that incorporates both species composition and the structure of interactions, this implies that GED includes compositional metrics. In this vein, two communities with high values of Jaccard would be also similar in functionality (high GED or similar paths of energy flows) if and only if the occurrence of the species is strongly related to the occurrence of the interactions.

Recently, other authors have proposed a different approach to integrate compositional and “interactional” community similarities (*Schmidt, Rodrigues & Von Mering, 2017*). In this approach, species co-occurrences are used to infer biotic interactions, through similarity-based network inference. This kind of inference includes constructing a network from pairwise co-occurrences and/or mutual exclusion, and some metric is used to quantify and determine the significance of the similarity of the pairwise distributions (*Faust & Raes, 2012; Schmidt, Rodrigues & Von Mering, 2017*). Despite the fact that these methods have been successfully applied in microbial ecology, and their use has complemented the information provided by standard analytical approaches (*Barberán et al., 2012; Cazelles et al., 2016; Faust & Raes, 2012; Freilich et al., 2018; Morales-Castilla et al., 2015; Schmidt, Rodrigues & Von Mering, 2017*), the utility of this method to analyze interactions at macro-scales is questionable, mainly due to the independence between co-occurrence measures and interactions (*Akin & Winemiller, 2006; Lopez et al., 2017; Poisot et al., 2012; Saavedra et al., 2016; Thompson & Townsend, 1999*). Some key restrictions of this theoretical framework have been associated to its inability to exactly reproduce the interaction networks (*Freilich et al., 2018; Morales-Castilla et al., 2015*), and that co-occurrences alone are not sufficient to provide insight about the biotic interactions in these communities. As such, several have noted the importance of directly collecting information of trophic interactions when analyzing natural communities (*Cazelles et al., 2016; Morales-Castilla et al., 2015*). This being said, collecting information about trophic interactions is expensive and time-consuming, so in many cases the implementation of this kind of sampling is not possible. Overall, given that metrics used to describe communities do not necessarily share significant levels of information, researchers should clearly identify not only their research question, but also the metrics (richness, composition, evenness, abundance, interactions) that best describe the dimension of the community to be analyzed.

The method used in this study reveals the effect that trophic interactions have on community similarity and highlights the constraints at analyzing community similarity in one or a few ecological dimensions. Similarly, others have developed a new approach to separate the effects of interactions in ecosystem functioning from those of species composition (*Jaillard et al., 2018*). These authors have found that the effects of interactions and composition are independent, but both contribute significantly to ecosystem functioning. Thus, the direct interpretation of species interactions from co-occurrence data remains controversial (*Cazelles et al., 2016*). In our case, and unlike microbial ecology studies, it was not necessary to infer trophic interactions because they were directly measured. Furthermore, here the AMI values confirmed that species co-occurrence and interactions did not equally portray the similarity of the communities regardless of the different cost scenarios assigned to the species and the structure of the interaction network.

The data analysis used in this study could be useful when interactions between species should be inferred from co-occurrence data. In these cases, potential bias from assumed relationships between composition and interactions could be avoided by attributing different costs when calculating GED. For example, the impact of interaction similarities on the whole similarity could be approximated by assigning different costs to link editions. If the whole similarity is notoriously affected by link edition costs then co-occurrences

and interactions are likely independent. Because the selection of a particular cost scenario could be considered subjective, and so affect the global result of the analysis, we think that following some simple criteria can reduce this subjectiveness. In our case and following the point of view of an ecologist, the deletion of a node in a network can reflect the temporal extinction of a species, and an addition reflects the opposite. The same can be applied to links that represent interactions. We considered that deleting or adding links to a network has a lower cost due to the high variability of the analyzed trophic interactions over time ([Lopez et al., 2017](#)), and because the presence of an interaction depends hierarchically on species composition (a given interaction occurs if and only if both interacting species co-occur). The flipping of an interaction was attributed a higher cost, because this edition changes the hierarchy of the consumer-resource interactions and the direction of energy transfer within the network, which is an ecologically costly and rare phenomenon.

In the ecological literature some level of association between state variables like richness, abundances, species composition, or biomass is usually assumed; however, many of these variables does not show robust association patterns ([Edwards & Richardson, 2004](#); [Poisot et al., 2012](#); [Pool et al., 2016](#)). The poor association between the similarity indices shown in this study reinforces the idea that each state variable represents a different dimension of natural systems. Future studies could take advantages of methods like those used in this study to ensure the robust assessment of the similitude of natural systems.

## CONCLUSIONS

In summary, our results show: (a) a weak relationship between measures of similarity using only the species composition and that those include composition and interactions, evidence the need of adding structural relationships to the similarity measure and (b) GED can be advantageous in the analysis of networks and ecological communities due to its flexibility in assigning different cost schemes depending of the researcher interest. A more practical consequence of our results is a cautionary note on community-level interpretation of similarity. In the analysis of ecological communities, it is commonly assumed that different sites, assemblages, or communities can be considered replicates. However, the question that arises is how similar must communities be in terms of any state variable to represent an adequate replicate of the system?

Finally, our results show that graph edit distance (GED) might be a valuable metric for the analysis of ecological communities. The use of an integrated measure that allows us to incorporate information of the composition and interaction structure of communities is useful for effectively establishing whether two communities, rather than operational units, could be considered as equivalent ecological systems. Our results further give warning about the need to account for the particularities of the multiple state variables that represent dimensions of ecological systems.

## ACKNOWLEDGEMENTS

We thank the Interaction Web Database of National Center for Ecological Analysis and Synthesis (<https://www.nceas.ucsb.edu/interactionweb/>) at the University of California, Santa Barbara, USA.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

Financial support was provided by CONICYT Grant N° 21140959 (DNL), FONDECYT N° 1040425 (PAC), 1190529 (Nelson Valdivia), 1160370 (SAE) and to the Center of Applied Ecology and Sustainability (CAPES) CONICYT PIA/BASAL FB0002 (SAE). Nelson Valdivia was supported by FONDAP IDEAL grant N° 15150003. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

CONICYT: N° 21140959.

FONDECYT: N° 1040425, 1190529, 1160370.

Center of Applied Ecology and Sustainability (CAPES) CONICYT PIA/BASAL: FB0002.

FONDAP IDEAL: N° 15150003.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Daniela N. Lopez analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Patricio A. Camus and Nelson Valdivia contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Sergio A. Estay analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The datasets are available in [Dataset S1](#). The original data set is available at the Web Database of the National Center for Ecological Analysis and Synthesis (NCEAS), at the University of California, Santa Barbara, USA.

([https://www.nceas.ucsb.edu/interactionweb/html/thomps\\_towns.html](https://www.nceas.ucsb.edu/interactionweb/html/thomps_towns.html)).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7013#supplemental-information>.

## REFERENCES

- Abrams PA. 1982. Functional responses of optimal foragers. *American Naturalist* 120(3):382–390 DOI 10.1086/283996.
- Akin S, Winemiller KO. 2006. Seasonal variation in food web composition and structure in a temperate tidal estuary. *Estuaries and Coasts* 29(4):552–567 DOI 10.1007/BF02784282.
- Anderson MJ, Ellingsen KE, McArdle BH. 2006. Multivariate dispersion as a measure of beta diversity. *Ecology Letters* 9(6):683–693 DOI 10.1111/j.1461-0248.2006.00926.x.
- Arditi R, Ginzburg LR. 2012. *How species interact: altering the standard view on trophic ecology*. New York: Oxford University Press DOI 10.1086/669300.
- Barberán A, Bates ST, Casamayor E, Fierer N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *Multidisciplinary Journal of Microbial Ecology* 6(2):343–351 DOI 10.1038/ismej.2011.119.
- Barwell LJ, Isaac NJB, Kunin WE. 2015. Measuring  $\beta$ -diversity with species abundance data. *Journal of Animal Ecology* 84(4):1112–1122 DOI 10.1111/1365-2656.12362.
- Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* 19(1):134–143 DOI 10.1111/j.1466-8238.2009.00490.x.
- Baselga A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography* 21(12):1223–1232 DOI 10.1111/j.1466-8238.2011.00756.x.
- Bezemer TM, Fountain MT, Barea JM, Christensen S, Dekker SC, Duyts H, Van Hal R, Harvey JA, Hedlund K, Maraun M. 2010. Divergent composition but similar function of soil food webs of individual plants: plant species and community effects. *Ecology* 91(10):3027–3036 DOI 10.1890/09-2198.1.
- Bunke H. 1983. What is the distance between graphs. *Bulletin of the European Association for Theoretical Computer Science* 20:35–39.
- Bunke H, Allermann G. 1983. Inexact graph matching for structural pattern recognition. *Pattern Recognition Letters* 1(4):245–253 DOI 10.1016/0167-8655(83)90033-8.
- Burkle LA, Marlin JC, Knight TM. 2013. Plant–pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science* 339(6127):1611–1615 DOI 10.1126/science.1232728.
- Canard E. 2011. Espace et neutralité dans les réseaux d'interactions écologiques. PhD thesis, Université Montpellier.
- Cazelles K, Araújo MB, Mouquet N, Gravel D. 2016. A theory for species co-occurrence in interaction networks. *Theoretical Ecology* 9(1):39–48 DOI 10.1007/s12080-015-0281-9.
- Charnov EL. 1976. Optimal foraging, the marginal value theorem. *Theoretical Population Biology* 9:129–136 DOI 10.1016/0040-5809(76)90040-X.
- Dehmer M. 2010. *Structural analysis of complex networks*. New York: Springer Science & Business Media DOI 10.1007/978-0-8176-4789-6.



- Devictor V, Mouillot D, Meynard C, Jiguet F, Thuiller W, Mouquet N. 2010.** Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecology Letters* **13**(8):1030–1040 DOI [10.1111/j.1461-0248.2010.01493](https://doi.org/10.1111/j.1461-0248.2010.01493).
- Edwards M, Richardson AJ. 2004.** Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* **430**(7002):881–884 DOI [10.1038/nature02808](https://doi.org/10.1038/nature02808).
- Emmert-Streib F, Dehmer M, Shi Y. 2016.** Fifty years of graph matching, network alignment and network comparison. *Information Sciences* **346**:180–197 DOI [10.1016/j.ins.2016.01.074](https://doi.org/10.1016/j.ins.2016.01.074).
- Faust K, Raes J. 2012.** Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**(8):538–550 DOI [10.1038/nrmicro2832](https://doi.org/10.1038/nrmicro2832).
- Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA. 2018.** Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* **99**(3):690–699 DOI [10.1002/ecy.2142](https://doi.org/10.1002/ecy.2142).
- Harrington R, Woiwod I, Sparks T. 1999.** Climate change and trophic interactions. *Trends in Ecology & Evolution* **14**(4):146–150 DOI [10.1016/S0169-5347\(99\)01604-3](https://doi.org/10.1016/S0169-5347(99)01604-3).
- Ibragimov R, Malek M, Guo J, Baumbach J. 2013.** GEDEVO: an evolutionary graph edit distance algorithm for biological network alignment. In: *OASISs-OpenAccess Series in Informatics*. Germany: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik DOI [10.4230/OASISs.GCB.2013.68](https://doi.org/10.4230/OASISs.GCB.2013.68).
- Jaarsma NG, De Boer SM, Townsend CR, Thompson RM, Edwards ED. 1998.** Characterising food-webs in two New Zealand streams. *New Zealand Journal of Marine and Freshwater Research* **32**(2):271–286 DOI [10.1080/00288330.1998.9516825](https://doi.org/10.1080/00288330.1998.9516825).
- Jaillard B, Richon C, Deleporte P, Loreau M, Violle C. 2018.** An a posteriori species clustering for quantifying the effects of species interactions on ecosystem functioning. *Methods in Ecology and Evolution* **9**(3):704–715 DOI [10.1111/2041-210X.12920](https://doi.org/10.1111/2041-210X.12920).
- Koleff P, Gaston KJ, Lennon JJ. 2003.** Measuring beta diversity for presence–absence data. *Journal of Animal Ecology* **72**(3):367–382 DOI [10.1046/j.1365-2656.2003.00710.x](https://doi.org/10.1046/j.1365-2656.2003.00710.x).
- Lopez DN, Camus PA, Valdivia N, Estay SA. 2017.** High temporal variability in the occurrence of consumer-resource interactions in ecological networks. *Oikos* **126**(12):1699–1707 DOI [10.1111/oik.04285](https://doi.org/10.1111/oik.04285).
- Malek M. 2015.** CytoGEDEVO: A Cytoscape app for fast and interactive network alignment. Master Thesis, Saarland University. Germany 82.
- Meyer PE. 2014.** Infotheo: Information-Theoretic Measures. R package version, 1(0). Available at <http://CRAN.R-project.org/package=infotheo>.
- Morales-Castilla I, Matias MG, Gravel D, Araújo MB. 2015.** Inferring biotic interactions from proxies. *Trends in Ecology & Evolution* **30**(6):347–356 DOI [10.1016/j.tree.2015.03.014](https://doi.org/10.1016/j.tree.2015.03.014).
- Morlon H, Kefi S, Martinez ND. 2014.** Effects of trophic similarity on community composition. *Ecology Letters* **17**(12):1495–1506 DOI [10.1111/ele.12356](https://doi.org/10.1111/ele.12356).
- Neyshabur B, Khadem A, Hashemifar S, Arab SS. 2013.** NETAL: a new graph-based method for global alignment of protein–protein interaction networks. *Bioinformatics* **29**(13):1654–1662 DOI [10.1093/bioinformatics/btt202](https://doi.org/10.1093/bioinformatics/btt202).

- Oksanen J, Blanchet FG, Michael F, Kindt R, Legendre P, McGlinn DP, Minchin PR, O'hara RB, Simpson G, Solymos P, Stevens MH, Szoecs E, Wagner H. 2013. Package 'vegan' Community ecology package. 2(9).
- Olden JD, Poff NL, Douglas MR, Douglas ME, Fausch KD. 2004. Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology and Evolution* 19(1):18–24 DOI 10.1016/j.tree.2003.09.010.
- Petchey OL, Gaston KJ. 2006. Functional diversity: back to basics and looking forward. *Ecology Letters* 9(6):741–758 DOI 10.1111/j.1461-0248.2006.00924.x.
- Poisot T, Canard E, Mouillot D, Mouquet N, Gravel D, Jordan F. 2012. The dissimilarity of species interaction networks. *Ecology Letters* 15(12):1353–1361 DOI 10.1111/ele.12002.
- Poisot T, Stouffer DB, Gravel D. 2015. Beyond species: why ecological interaction networks vary through space and time. *Oikos* 124(3):243–251 DOI 10.1111/oik.01719.
- Pool TK, Cucherousset J, Boulêtreau S, Villéger S, Strecker AL, Grenouillet G. 2016. Increased taxonomic and functional similarity does not increase the trophic similarity of communities. *Global Ecology and Biogeography* 25(1):46–54 DOI 10.1111/geb.12384.
- R Development Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Riesen K. 2015. *Structural pattern recognition with graph edit distance: approximation algorithms and applications*. Switzerland: Springer DOI 10.1007/978-3-319-27252-8.
- Saavedra S, Rohr RP, Fortuna MA, Selva N, Bascompte J. 2016. Seasonal species interactions minimize the impact of species turnover on the likelihood of community persistence. *Ecology* 97(4):865–873 DOI 10.1890/15-1013.1.
- Schmidt TSB, Rodrigues JFM, Von Mering C. 2017. A family of interaction-adjusted indices of community similarity. *Multidisciplinary Journal of Microbial Ecology* 11(3):791–807 DOI 10.1890/15-1013.1.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11):2498–2504 DOI 10.1101/gr.1239303.
- Skalski GT, Gilliam JF. 2001. Functional responses with predator interference: viable alternatives to the holling type II model. *Ecology* 82(11):3083–3092 DOI 10.1890/0012-9658(2001)082[3083:FRWPIV]2.0.CO;2.
- Solé-Ribalta A, Serratoso F, Sanfeliu A. 2012. On the graph edit distance cost: properties and applications. *International Journal of Pattern Recognition and Artificial Intelligence* 26(05):1260004 DOI 10.1142/S021800141260004X.
- Strona G, Matthews TJ, Kortsch S, Veech JA. 2018. NOS: a software suite to compute node overlap and segregation in ecological networks. *Ecography* 41(3):558–566 DOI 10.1111/ecog.03447.

- Strona G, Veech JA. 2015.** A new measure of ecological network structure based on node overlap and segregation. *Methods in Ecology and Evolution* **6(8)**:907–915 DOI [10.1111/2041-210X.12395](https://doi.org/10.1111/2041-210X.12395).
- Thackeray SJ, Sparks TH, Frederiksen M, Burthe S, Bacon PJ, Bell JR, Botham MS, Brereton T, Bright PW, Carvalho L, Clutton-Brock TIM, Dawson A, Edwards M, Elliott JM, Harrington R, Johns D, Jones I, Jones JT, Leech DI, Roy DB, Scott WA, Smith M, Smithers RJ, Winfield I, Wanless S. 2010.** Trophic level asynchrony in rates of phenological change for marine, freshwater and terrestrial environments. *Global Change Biology* **16(12)**:3304–3313 DOI [10.1111/j.1365-2486.2010.02165.x](https://doi.org/10.1111/j.1365-2486.2010.02165.x).
- Thompson RM, Townsend CR. 1999.** The effect of seasonal variation on the community structure and food-web attributes of two streams: implications for food-web science. *Oikos* **87**:75–88 DOI [10.2307/3546998](https://doi.org/10.2307/3546998).
- Thompson RM, Townsend CR. 2000.** Is resolution the solution? the effect of taxonomic resolution of three stream food webs on the calculated properties. *Freshwater Biology* **44(3)**:413–422 DOI [10.1046/j.1365-2427.2000.00579.x](https://doi.org/10.1046/j.1365-2427.2000.00579.x).
- Thompson RM, Townsend CR. 2003.** Impacts on stream food webs of native and exotic forest: an intercontinental comparison. *Ecology* **84(1)**:145–161 DOI [10.1890/0012-9658\(2003\)084\[0145:IOSFWO\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0145:IOSFWO]2.0.CO;2).
- Thompson RM, Townsend CR. 2005.** Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in streams. *Oikos* **108(1)**:137–148 DOI [10.1111/j.0030-1299.2005.11600.x](https://doi.org/10.1111/j.0030-1299.2005.11600.x).
- Trøjelsgaard K, Jordano P, Carstensen DW, Olesen JM. 2015.** Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. *Proceedings of the Royal Society B: Biological Sciences* **282(1802)**:1–9 DOI [10.1098/rspb.2014.2925](https://doi.org/10.1098/rspb.2014.2925).
- Vinh NX, Epps J, Bailey J. 2010.** Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *Journal of Machine Learning Research* **11**:2837–2854 DOI [10.1098/rspb.2014.2925/10.1145/1553374.1553511](https://doi.org/10.1098/rspb.2014.2925/10.1145/1553374.1553511).
- Whittaker RH. 1960.** Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* **30(3)**:279–338 DOI [10.2307/1943563](https://doi.org/10.2307/1943563).
- Whittaker RH. 1972.** Evolution and measurement of species diversity. *Taxon* **21(2/3)**:213–251 DOI [10.2307/1218190](https://doi.org/10.2307/1218190).
- Wilson MV, Shmida A. 1984.** Measuring beta diversity with presence-absence data. *The Journal of Ecology* **72(3)**:1055–1064 DOI [10.2307/2259551](https://doi.org/10.2307/2259551).
- Zhang S, Liu H, Ning X, Zhang X, Xu Y, Liu P, Li X, Ren W, Van Mieghem P, Omic J, Poulin R, Montoya JM, Sole RV, Lu X, Gray C, Brown LE, Ledger ME, Milner AM, Mondragón RJ, Woodward G, Ma A, Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T, Lo CM, Morand S, Galzin R, Estudio P, Arneberg P, Skorpning A, Grenfell B, Read AF. 2016.** Vital nodes identification in complex networks. *Physics Reports* **650**:1–63 DOI [10.1155/2014/407639](https://doi.org/10.1155/2014/407639).