



Combining centrality indices: Maximizing the predictability of keystone species in food webs

Catarina Gouveia^{a,b}, Ágnes Mór h^{b,c}, Ferenc Jord n^{b,c,d,*}

^a Informatics Department, Lisbon University Science Faculty, Lisbon, Portugal

^b Centre for Ecological Research, Budapest, Hungary

^c Evolutionary Systems Research Group, Centre for Ecological Research, Tihany, Hungary

^d Stazione Zoologica, Napoli, Italy

ARTICLE INFO

Keywords:

Food web
Keystone species
Centrality
Multiple indices
Simulation
Machine learning

ABSTRACT

Network analysis offers a rich toolkit to study various graph models in biology. In ecology, centrality indices have been suggested to indicate keystone species in interaction networks and to quantify their importance in an ecosystem. There is a large number of centrality indices, however, and it is often unclear what is their precise biological meaning, how are they related to each other and which one is the “best” predicting the functioning of the modelled biological system. It is a major challenge to use simple structural indicators in order to predict the outcome of much more complicated dynamical simulations. The question is which one is the most predictive one and what is the meaning of particular structural indices. Here we use machine learning techniques to combine k centrality indices out of n in such a way that the gained combined index (a “cocktail” of single indices) correlates better with simulated dynamics. In particular, we are interested in rank correlations between single-node and multi-node centrality and simulated node importance. We identify index families based on similarity. The best single-index correlations (weighted degree centrality) can predict simulated food web dynamics with an accuracy up to 70.06%. This accuracy can be raised reasonably, using the best cocktail, up to 78.42%. This is a combination of node degree (D) and 5-step-long weighted importance index (WI^5). Since they have completely different properties (the former is local and binary, the latter is *meso*-scale and weighted), we can demonstrate that a good cocktail has to combine indices from different families in order to best improve predictions. If one needs to predict dynamics from structure, there is a way to use wise proxies of simple topological indices – instead of performing complicated simulations.

1. Introduction

In ecological communities, keystone species may be of high importance because they are richly connected in the interaction network. Thus, the structural analysis of inter-specific interaction networks (e.g. food webs) may provide indicators for keystone species: centrality indices quantify importance. In food web simulation models, also dynamical indices of importance can be obtained. It is a major challenge to understand how can structure predict dynamics, i.e. how is the importance rank of species, based on particular structural node centrality indices correlated to the importance rank of species based on simulated, dynamical importance indices. The similarity of centrality indices (Jord n et al., 2007) and the correlation between structural and dynamical indices (Jord n et al., 2008) have already been studied. It is

an old question to what extent structure can predict the behaviour of individual nodes and network dynamics in both food webs (Allesina et al., 2009) and other networks (Zotenko et al., 2008).

Instead of focusing on the single most appropriate index (that is not easy to determine), one may use a combination of a few indices. The simplest and oldest dilemma is whether to use binary or weighted networks – with the corresponding network analytical techniques (Scotti et al., 2007). In the search for predictive ecosystem models, targeting conservation problems, several methods are being used and compared (Ortiz et al., 2013, 2017). This challenge has been investigated for both trophic networks (Jord n et al., 2006; Bauer et al., 2010; Lai et al., 2015) and habitat networks (Baranyi et al., 2011; Pereira and Jord n, 2017; Pereira et al., 2017). Since a few indices can be chosen in many ways out of many (Estrada, 2007; Pocock et al., 2011; Borrett, 2013; Lau et al.,

* Corresponding author at: Centre for Ecological Research, 1113, Karolina 29, Budapest, Hungary.

E-mail address: jordan.ferenc@ecolres.hu (F. Jord n).

<https://doi.org/10.1016/j.ecolind.2021.107617>

Received 8 June 2020; Received in revised form 24 January 2021; Accepted 13 March 2021

Available online 31 March 2021

1470-160X/  2021 The Author(s).

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2017), and these can be combined in many ways, the problem is calling for advanced computational tools.

Here we use machine learning techniques to see how certain combinations of k centrality indices out of n can correlate to dynamical importance. The questions are: (1) which indices to use, (2) how to combine them and (3) to what extent the correlation can be improved.

2. Materials and methods

2.1. Data

We used 1000 randomly generated networks composed of 15 species: 12 consumers and 3 basal species. The maximum number of top predators was set to 4 and 36 links were randomly generated between species. The basal species #1, #2 and #3 were unchanged for the community response simulation. Therefore, we did not study the dynamical effects of these nodes in this analysis. Links in each network were considered undirected. In the case of weighted topological indices, weights for the arcs were generated as $\frac{out-degree}{degree}$.

We decided to fix these parameters as we wanted to focus only on the combination of indices, excluding other effects, like the influence of network size. Keeping the number of nodes constant, the importance of pure topology can be demonstrated. Further, a network of 15 nodes is large enough to be already interesting from a topological point of view but also small enough for understanding and interpretation.

With these data, a matrix $M_{12n \times k}$ was constructed: being n each network (composed of 12 nodes) and k the respective 18 topological indices and the community response simulation's results. This matrix was then processed in two different ones: $M_{F_{12n \times k}}$ and $M_{R_{12n \times k}}$. $M_{F_{12n \times k}}$ containing the original values derived by each index and the community response simulation in each column, compressed in the range [0,1]. $M_{R_{12n \times k}}$ containing the rank order values, i.e., in this matrix the real values were replaced by the respective node rank order (from 1 to 12). To nodes in the same network, with the same index value, it was assigned a random order.

2.2. Topological indices

To access the positional importance of nodes in a network one can use a range of different network indices (local or global, weighted or unweighted, hierarchical). For this work we analysed 18 different indices and the results provided by a simulation to evaluate the dynamics of a network in case of a single species perturbation – community response. These indices are briefly presented in Table 1. The distributions of index values, for each index, are shown in Supplementary Material E.

2.3. Networks dynamics

To model the dynamics of the hypothetical food web, we used a simulation model of coupled differential equations. The dynamics of species interactions were described by the equations given in Table 2 (Eq. 6). After balancing the simulation models, we performed perturbations of each species and measured the response of each other species. Community response (CR_j) for the perturbation of species j was defined according to the response function given in Table 2. We did not consider self effects ($i \neq j$ in Eq. 7, see Table 2). The simulated community importance of species j is measured as the community response generated by the perturbation of species j .

2.4. Statistical analyses

Two simple analyses were made to compare each structural index with the simulation results: a single correlation and a cluster analysis – Spearman Correlation was applied, in both cases, to both matrices.

Table 1
Topological indices used in food web analysis.

Index name	Description
Degree (D)	Number of other nodes connected directly to the considered node (in a food web represents the sum of its preys and predators).
Weighted Degree (wD)	Sum of weights of links adjacent to the considered node (Wasserman and Faust, 1994).
Betweenness centrality (BC)	Frequency of the considered node on the shortest paths connecting all pairs of other nodes (Wasserman and Faust, 1994). $BC_i = \frac{2 \sum_{j < k} g_{jk}(i)}{(N-1)(N-2)}, i \neq j \text{ and } k(1)$ g_{jk} is the number of equally shortest paths between nodes j and k , $g_{jk}(i)$ is the number of these shortest paths to which node i is incident in the length of the shortest path between nodes i and j in the network.
Closeness centrality (CC)	Quantifies how short are the minimal paths from a given node to all others. $CC_i = \frac{N-1}{\sum_{j=1}^N d_{ij}}, i \neq j(2)$ d_{ij} is the length of the shortest path between nodes i and j in the network (Wasserman and Faust, 1994).
Positional importance based on indirect chain effects (TI ¹ , TI ³ , TI ⁵ , WI ¹ , WI ³ , WI ⁵)	The topological importance of species i when effects “up to” n steps are considered is the sum of effects originated from species i up to n steps averaged over by the maximum number of steps considered (i.e. n): $TI_i^n = \frac{\sum_{m=1}^n \sigma_{m,i}}{n} = \frac{\sum_{m=1}^n \sum_{j=1}^N \alpha_{m,ji}}{n}(3)$ $\alpha_{m,ji}$ is the effect of j on i when i can be reached from j in n steps (Jordán et al., 2003a). WI_i^n is exactly the same but with weighted links. We analyzed indirect effects of different maximum length ($n = 1, 3, 5$).
Status index and its components (s, s' and Δs)	In a directed strong hierarchy, the status is the sum of distances to each other nodes. Reversing the hierarchy (reverting the direction of the links), the same calculation will give the contrastatus of each node (s') (Harary, 1959). $\Delta s_i = s_i - s'_i(4)$ Δs_i is called the net status of node i .
Keystone index and its components (K, K _{bu} , K _{td} , K _{dir} , K _{indir})	The keystone index of a species i is defined as (Jordan et al., 1999): $K_i = K_{bu,i} + K_{td,i} = K_{dir,i} + K_{indir,i} = \sum_{c=1}^n \frac{1}{d_c} (1 + K_{bc}) + \sum_{e=1}^m \frac{1}{f_e} (1 + K_{te})(5)$ n is the number of predators eating species i , d_c is the number of prey species of its c^{th} predator, K_{bc} is the bottom-up keystone index of the c^{th} predator, m is the number of prey eaten by species i , f_e is the number of predators of its e^{th} prey, K_{te} is the top-down keystone index of the e^{th} prey, $K_{bu,i}$ is the bottom-up keystone index, $K_{td,i}$ is the top-down keystone index, $K_{dir,i}$ represents the direct effects for node i , $K_{indir,i}$ represents the indirect effects for node i .

The cluster analysis was performed using an UPGMA classification, to compare all topological indices used. Community response values were not considered since our aim was to understand which were the most related topological indices. Distances between indices i and j were calculated using $d_{ij} = 1 - |\rho_{ij}|$, for each network. The next step was to perform a consensus dendrogram using the majority rule – only groups with 50% appearance in the original set of 1000 networks appeared in the final dendrogram.

Table 2
Brief description of the models used to model network dynamics and evaluate community response.

	Description
Network dynamics ($\frac{dN_i}{dt}$)	<p>The dynamics of each species can be described by the following differential equation (Mórhé et al., 2009):</p> $\frac{dN_i}{dt} = r_i N_i \left(1 - \frac{N_i}{K_i} \right) + \sum_{\rho=\text{resources}} N_i \varepsilon_{i\rho} \frac{N_i^\rho \omega_{i\rho}}{N_0^\rho + \omega_{i\rho} Q_{i\rho}} - \sum_{c=\text{resources}} N_i \varepsilon_{ci} \frac{N_i^\rho \omega_{ci}}{N_0^\rho + \omega_{ci} Q_{ci}} - d_i N_i \quad (6)$ <p> N_i means the abundance of species i, r_i is the rate of increase, K_i is the carrying capacity of the logistic model, d_i is the mortality rate of consumer species i, $\omega_{i\rho}$ is species i's relative consumption rate when consuming ρ, N_0 is the half-saturation density, $Q_{i\rho}$ is the sum of the abundances of the resources i can consume. The relative consumption rates are inversely proportional to the number of resources: $\omega_i = 1/n$ (Mórhé et al., 2018). </p>
Community response (CR_j)	$CR_j = \sum_{i=1}^n \left \frac{N_{i(j)}^t}{N_i^t} - 1 \right / 14, (i \neq j) \quad (7)$ <p> $N_{i(j)}^t$ is the population size of species i, at time t in a simulation where species j was perturbed, N_i^t is the population size of species i at time t without perturbation, i.e., in a reference simulation. </p>

2.5. Combination of indices

Using some of the topological indices to try to predict the same results obtained throughout dynamical approaches would be of added value inasmuch as static food webs and the embedded species importance are easier to assess.

We address the question how to best predict network dynamics by structure. Instead of applying dynamical models that are data-intensive and complicated, we make the possibly best approximation by combining a certain number of structural indices in some appropriate

way. We use machine learning in order to find the best “cocktail” of topological indices: how many of them are needed, which ones and how to combine them exactly.

2.6. Program and parameters used

In order to find mathematical expressions that could predict the simulation response values, based in the used topological indices, we used a Genetic Programming (GP) with symbolic regression.

This algorithm was trained using the values from both matrixes

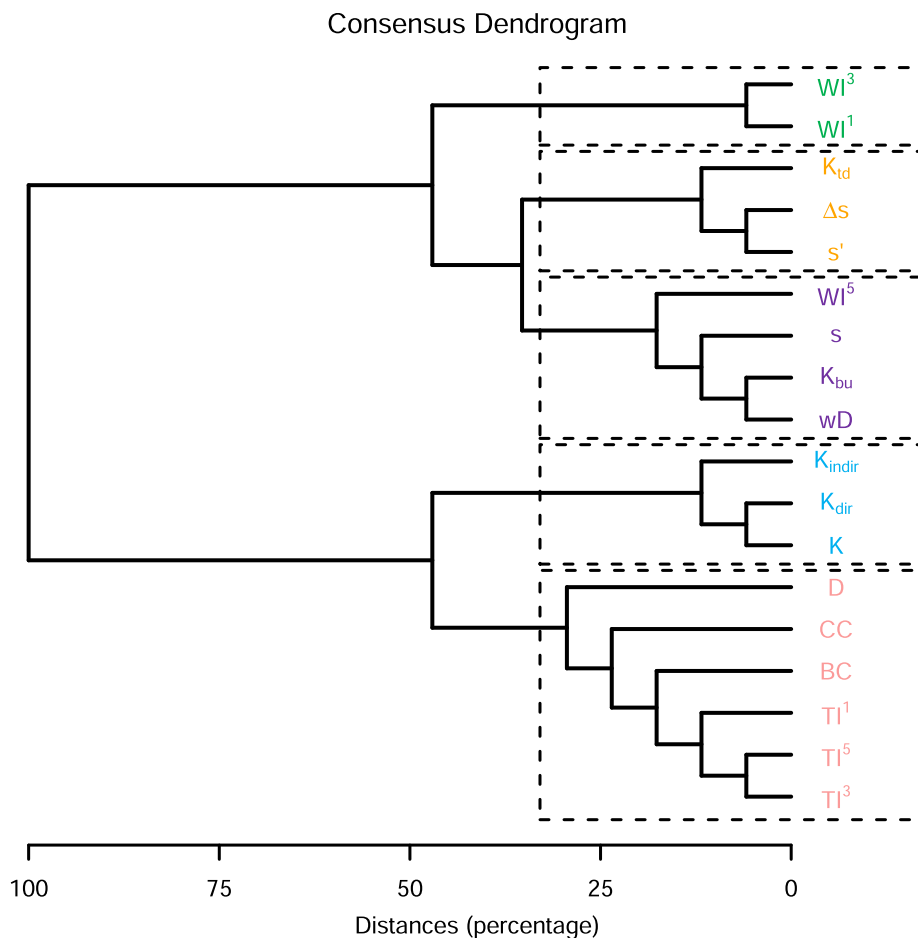


Fig. 1. Consensus dendrogram between node ranks of different topological indices. The consensus dendrogram was the same for the ordinal and metric values (distance $d_{ij} = 1 - |\rho_{ij}|$). For a distance of 30%, 5 groups can be formed: $\{WI^1, WI^3\}$, $\{\Delta s, s', K_{td}\}$, $\{K_{bu}, wD, s, WI^5\}$, $\{K, K_{dir}, K_{indir}\}$, $\{TI^3, TI^5, TI^1, BC, CC, D\}$.

($M_{r_{012n \times k}}$ and $M_{r_{12n \times k}}$). An attempt of performing a grid search was made but due to computational and time constraints, specific parameters were applied instead.

We performed 10,000 iterations in each of which we assigned random values based in a uniform distribution to the parameters related to mutations. Each final mathematical formula obtained was the best result of an “evolutionary” process of 100 generations, chosen from initial population sizes of 1000, 5000 and 10000. Spearman correlation, using a cross-validation test, was used to access the quality of the results. The data was partitioned into ten subsets of equal size. One of these ten subsets was used as the test data set, while the other nine were used as the training data set. The cross-validation process was then repeated ten times, with each of the ten subsets used once as the test set. The results of all iterations were then averaged to produce a single estimation for the accuracy.

3. Results and discussion

3.1. Cluster analysis

A consensus dendrogram between topological indices’ results was performed using either ordinal and metric values (the similarity distance used was $d_{ij} = 1 - |\rho_{ij}|$). The result (Fig. 1) was the same, for both matrices, so only one figure is showed. We decided to focus our analysis in the 5 groups formed considering a similarity distance of 30%: $\{WI^3, WI^1\}$, $\{\Delta s, s', K_{td}\}$, $\{K_{bu}, wD, s, WI^5\}$, $\{K, K_{dir}, K_{indir}\}$, $\{TI^3, TI^5, TI^1, BC, CC, D\}$. However, an alternative analysis with a similarity distance of 40% (identifying 4 groups) can be consulted in Fig. C1.

These groups show redundant information: indices within same groups are highly correlated to each other, due to their mathematical construction or to their final attributed results regarding nodes’ importance in networks. Consequently, if we want to simply evaluate a food web in a static way, we can use one index from each group: we can reduce the complexity of the analysis, without losing much information.

3.2. Single-node correlation

Table 3 shows the Spearman correlation for each of the 18 topological indices related to the community response. These correlations were calculated using metric data ($M_{r_{12n \times k}}$). Results for ordinal data can be seen in Table A1. Spearman correlation averages presented in these tables were calculated using the absolute values of the obtained correlations.

Indices $\{wD, K_{bu}, WI^5, WI^3, K_{td}, s, \Delta s, WI^1, s'\}$ correlated, in both cases, with community response values ($\rho \geq 0.5, p\text{-value} \leq 0.05$) while $\{K_{indir}, K_{dir}, K, BC, TI^1, TI^3, TI^5, D, CC\}$ were not correlated. Only two of the correlated indices were positively correlated: the keystone index based on top-down approach and the contrastatus $\{K_{td}, s'\}$. All indices correlated better with the rank order data than with the metric data. However, $\{wD, WI^5, K_{bu}, WI^3, WI^1\}$ performed better when applied to metric data – increasing the correlation in about 0.85%.

It is also noticeable that the order of the indices (from the most correlated one to the least correlated one) was slightly different whether we used ordinal and metric values (Table A1). Furthermore, if we take a closer look, the first two branches formed in Fig. 1 depict exactly these two groups of indices.

3.3. Combination of k indices

The obtained cocktails can be consulted in Tables 4–6. For better results’ visualization, indices were coloured as in the five groups present in Fig. 1. Only the top 20 results are shown. Table 4 shows best spearman correlation cocktails’ values, Table 5 displays the most frequent results (out of the 30 000 generated in total) and Table 6 shows the indices’ groups (or “families”, within the top 20 most frequent). Results for

Table 3

Spearman correlation and respective p -values related to “community response” - metric values. Symbols represent each of the families found in the dendrogram: \cdot , purple family; \circ , green family; \blacksquare , yellow family; \circ , blue family; $*$, pink family. Spearman correlation (ρ) was calculated for each of the 18 topological indices versus the community response results. p -values were derived for each Spearman correlation; Average Spearman correlation was obtained using absolute values; the shaded zone shows the correlated indices.

	Spearman correlation (ρ)	p -value
wD .	-0.7006	< 0.05
WI^5 .	-0.6773	< 0.05
K_{bu} .	-0.6755	< 0.05
WI^3 \diamond	-0.6645	< 0.05
WI^1 \diamond	-0.6002	< 0.05
K_{td} \blacksquare	0.5986	< 0.05
s .	-0.5866	< 0.05
Δs \blacksquare	-0.5806	< 0.05
s' \blacksquare	0.5438	< 0.05
K_{indir} \circ	0.3226	< 0.05
K°	0.2665	< 0.05
K_{dir} \circ	0.1524	< 0.05
BC *	-0.1401	< 0.05
TI^1 *	-0.1381	< 0.05
TI^3 *	-0.1059	< 0.05
TI^5 *	-0.1009	< 0.05
D *	-0.0548	< 0.05
CC *	-0.0330	< 0.05

ordinal data can be consulted in Tables B1, B2 and B3.

The top 20 mathematical combinations used a range of 4 to 14 different indices in the same expression (Table 4). These results showed significant increases in correlation values when compared to single correlations.

The mathematical expression with the best Spearman correlation showed an increase of about 8.36% when compared to the best correlation using a single index (78.42% versus 70.06%).

We also analysed the most frequent cocktails of indices derived from the algorithm used. In this scenario, only two or three indices were combined, and Spearman correlation was significantly better than single-index correlations. For instance, Spearman correlation obtained was 77.02% (for two different mathematical expressions – one relying in two indices, and the other one in three). The loss of information here was not significant since the previously best correlation obtained was 78.42% (and 70.06% for the best single correlation), which implies that these approximations are still valid to predict our “target values” more than using only one topological index. We can conclude that, in general, we can use these simpler mathematical expressions (Table 5) to get good predictions without much information loss and with a decrease in effort and computational time.

Table 6 shows the most common groups of indices. We can see that the “purple family” was present in almost every expression (in Table 5). This family was more frequently married with the pink, green and blue. Note that green, yellow and purple families represent community response correlated indices while blue and pink represent not correlated indices. Consequently, when using metric data one can use correlated and uncorrelated indices together to have strong and fast predictions (regarding community response values).

For these predictions, we can rely on 13 “families” that mix 9 indices (see Table 6). This means that in order to simplify our approach and analysis we can rely in some combination of these nine indices. The most frequent index was the degree (D) – it occurred in 10 out of 13 families – followed by weighted topological importance 5-step-long (WI^5), present in 9. Topological indices $TI^5, WI^3, wD, TI^3, TI^1, BC$ and K_{dir} were also relevant. It is also interesting to highlight that the combination of

Table 4

Top-20 mathematical expressions, regarding the $M_{r_{12n \times k}}$ matrix. Symbols represent each of the families found in the dendrogram: \cdot , purple family; \diamond , green family; \blacksquare , yellow family; \circ , blue family; $*$, pink family. Spearman correlation (ρ) derived directly from the algorithm used. Relative frequency in percentage is the number of times each result appeared among all the results (i.e., the 30 000 iterations executed), in percentage. Symbols represent each of the families found in the dendrogram: \cdot , purple family; \diamond , green family; \blacksquare , yellow family; \circ , blue family; $*$, pink family. Spearman correlation (ρ) derived directly from the algorithm used. Relative frequency in percentage is the number of times each result appeared among all results (i.e., among the 30 000 iterations executed), in percentage.

Expression	Spearman correlation (ρ)	Relative frequency (percentage)
$K_{indir}^{2\circ} \times WI^{\cdot} \times s \cdot \times (CC^* - K_{dir}^{\circ})(K_{dir}^{\circ} - 0.028) + \left(\frac{WI^{\cdot}}{CC^*} - \frac{D^*}{WI^{\cdot}}\right) \times \frac{K_{bu} \cdot \times WI^{\cdot}}{wD^{\cdot}} - (D - \Delta s) \times (\Delta s - K^{\circ})(wD \cdot \times TI^{1*}) \times$	(8) 0.7842	0.003
$\times \left(\frac{BC^* \times wD^{\cdot}}{D^*} - WI^{\cdot} \times \Delta s + TI^{3*}\right)$		
$\left(\frac{TI^{5*}}{WI^{\cdot}} - WI^3 - BC^*\right) \left(\frac{s \cdot}{wD \cdot} + K_{id} - BC^*\right) \times \left(TI^{5*} \times WI^{\cdot} \diamond + K_{id} - TI^{1*} + \frac{WI^{\cdot} \diamond \times D^*}{WI^{5.2 \cdot}}\right)$	(9) 0.7818	0.003
$wD \cdot + \frac{WI^{\cdot}}{D^* + CC^* \times s \cdot}$	(10) 0.7801	0.003
$\frac{\Delta s}{s \cdot} \times wD \cdot \times TI^{5*} + \frac{WI^{\cdot}}{TI^{5*}}$	(11) 0.7801	0.003
$(CC^* + 1)WI^{\cdot} + wD \cdot - \frac{TI^{5*}}{WI^{\cdot}} - \left(\frac{K_{bu} \cdot}{wD \cdot} + \Delta s - BC^*\right) \left(WI^3 \diamond \times WI^{\cdot} + \frac{D^*}{WI^3 \diamond}\right)$	(12) 0.7799	0.003
$\frac{WI^{\cdot}}{TI^{3*}} + TI^{1*} + BC^* - \frac{D^*}{WI^{\cdot}} (\Delta s + K_{id})$	(13) 0.7791	0.003
$TI^{5*} + wD \cdot - 1 - (K^{\circ} - BC^*)(TI^{3*} + K_{dir}^{\circ}) - \left(\frac{wD \cdot}{K_{bu} \cdot} - \frac{D^*}{WI^3 \diamond} + \frac{WI^{\cdot}}{TI^{3*} \times TI^{5*}} \times CC^*\right)$	(14) 0.7789	0.003
$\frac{(s \cdot + TI^{5*})(wD \cdot - D^*)}{BC^* \times wD \cdot + TI^{1*} \times WI^{\cdot}} + (WI^{\cdot} + \Delta s)(0.833TI^{1*}) \times (CC^* - WI^3 \diamond + K_{id} \times D^*)$	(15) 0.7788	0.003
$WI^{\cdot} \times 0.547 - CC^* \times K_{id} - \left(\frac{D^*}{0.804}\right)(\Delta s - D^*) - 1 - wD \cdot \times K^{\circ} - 0.099TI^{5*} - \frac{WI^3 \diamond}{D^*}$	(16) 0.7788	0.003
$\frac{D^*}{WI^{\cdot}} (\Delta s + K_{indir}^{\circ}) - (TI^{3*} + CC^*) \frac{wD \cdot}{WI^{\cdot}}$	(17) 0.7786	0.003
$\frac{WI^{\cdot}}{D^*} - (K_{bu} \cdot \times TI^{1*}) + \frac{BC^*}{\Delta s} \times \frac{wD \cdot}{TI^{5*}}$	(18) 0.7786	0.003
$\frac{(TI^{3*} + CC^*)wD \cdot}{K_{bu} \cdot} - D^* \left(\frac{1}{WI^{\cdot}} + CC^*\right)$	(19) 0.7784	0.003
$\left(s \cdot + wD \cdot - \frac{D^*}{WI^{\cdot}}\right) \left(TI^{1*} \times CC^* - K_{bu} \cdot - 0.916\right)$	(20) 0.7782	0.003
$\left(\frac{WI^{\cdot}}{D^*} + K_{bu} \cdot\right) \frac{wD \cdot}{K_{bu} \cdot}$	(21) 0.7781	0.003
$\left(\frac{WI^{\cdot}}{D^*} \times K_{bu} \cdot + 1\right) wD \cdot$	(22) 0.7781	0.003
$(TI^{1*} + WI^1 \diamond)(s \cdot - 0.314) + wD \cdot \left(1 + \frac{1}{D^*}\right) - s \cdot$	(23) 0.7781	0.003
$\frac{0.990D^*}{0.593WI^{\cdot}} (\Delta s - wD \cdot + s \cdot \times K^{\circ})$	(24) 0.7770	0.003
$(1 + 0.661K_{id} - TI^{3*} \times D^*)(WI^{\cdot} \times s \cdot \times K^{\circ} \times TI^{1*} - \frac{D^*}{WI^{\cdot}} + s \cdot + K_{id})$	(25) 0.7770	0.003
$(BC^* - s \cdot + WI^{\cdot} \times wD \cdot) - (TI^{3*} + WI^3 \diamond)(WI^1 \diamond - CC^*) - \left(\frac{D^*}{WI^1 \diamond} + \Delta s + K_{id}\right) \left(\frac{WI^1 \diamond}{WI^{\cdot}} - s \cdot \times K^{\circ}\right)$	(26) 0.7770	0.003
$(\Delta s - K_{dir}^{\circ})(\Delta s - TI^{5*}) - (CC^* \times wD \cdot) - \frac{WI^{\cdot}}{D^*}$	(27) 0.7770	0.007

indices $\{WI^n, TI^n\}$ was present in half of these mathematical expressions and the combination of $\{WI^n, D\}$ in 13 out of the most frequent top-20 expressions showed.

We also checked which were the most frequent results in all the generated mathematical formulae (Tables C1, C2). Results showed what we could preview from Table 6: the degree (D) and the 5-step-long weighted topological importance index (WI^5) were the ones that appeared the most in all results, followed by TI^5 and WI^3 . The same results for ordinal data are present in Tables D1, D2.

4. Conclusions

In network research, a number of centrality indices are used. Their relationships are quite well known but combining them in order to derive more powerful indices has been already done (in systems biology, see del Rio et al. 2009; in landscape ecology, see Jordán et al. 2003b) but not in a systematic way as presented here.

Our results clustered topological indices into their families of similarity. We found that the weighted degree index is good enough to predict simulated food web dynamics with an accuracy up to 70.06%. Using the cocktails of indices obtained, however, we can raise this accuracy up to 78.42%. In addition, we concluded that the 5-step-long weighted topological index (WI^5), which considers long, weighted, indirect interactions, along with the degree index (D) that considers short, binary, direct interactions – almost opposite indices – when combined,

complemented each other, and enabled us to predict easily and accurately the simulation results.

It is noted that these results provide new ways to achieve the order of importance of each species obtained through complex biological simulations. We can either use simple structural indices and choose the most appropriate one(s), or apply some combinations of these indices. The best combination (D and WI^5) contains an index for direct, binary effects and another one for indirect, weighted effects. These are quite different indices, and this is why it is interesting and convenient to combine them: they provide complementary and adequate information (Gouveia, 2019).

In future studies, one can focus on 1) using more diverse food webs of a larger range of parameters, in order to increase generality, analyse potential scaling issues (effects of network size) and get more accurate combinations, 2) the application of the present results to real food webs in order to identify which are the keystone species and check their biological role in those networks and 3) the application of the same approach to other network types (e.g. protein–protein interaction networks). A further research direction is the improved combination of topological centrality measures and thermodynamic indicators (Ludovisi 2014, Ludovisi and Scharler 2017). In order to test the generality and predictability of the results, both in silico (e.g. EcoSim) and in vivo (e.g. mesocosms) experiments can be conducted (the latter being in progress). We believe that more concise and efficient ways to identify keystone species in ecological networks will be essential for the future of

Table 5
Top-20 frequent mathematical expressions (relative frequency), regarding $M_{r12n \times k}$ matrix.

indices	Spearman correlation (ρ)	Relative frequency (percentage)
$WF^{\circ} \cdot$	(28) 0.7616	15.027
D^*	(29) 0.7616	6.733
$WF^{\circ} \cdot$	(30) 0.7586	2.897
$TF^{\circ} *$	(31) 0.7586	2.157
$WF^{\circ} \cdot$	(32) 0.7546	1.257
$TF^{\circ} *$	(33) 0.7546	0.983
$WF^{\circ} \cdot$	(34) 0.7540	0.733
D^*	(35) 0.7702	0.593
$WF^{\circ} \diamond$	(36) 0.7702	0.573
$wD \cdot + \frac{WF^{\circ} \cdot}{D^*}$	(37) 0.7694	0.557
$\frac{WF^{\circ} \diamond}{D^*} + wD \cdot$	(38) 0.7539	0.445
$TF^{\circ} *$	(39) 0.7538	0.363
$\frac{WF^{\circ} \diamond}{D^*}$	(40) 0.7535	0.330
$wD \cdot$	(41) 0.7529	0.290
$BC^* - \frac{D^*}{WF^{\circ} \cdot}$	(42) 0.7686	0.276
$D^* - TF^{\circ} *$	(43) 0.7686	0.270
$\frac{WF^{\circ} \cdot}{D^*} - BC^*$	(44) 0.7677	0.267
$TF^{\circ} * - \frac{D^*}{WF^{\circ} \cdot}$	(45) 0.7676	0.260
$\frac{WF^{\circ} \cdot}{D^*} - TF^{\circ} *$	(46) 0.7660	0.237
$\frac{WF^{\circ} \cdot}{D^*} - K_{dir}^{\circ}$	(47) 0.7677	0.233
$TF^{\circ} * - \frac{D^*}{WF^{\circ} \cdot}$		

Table 6
Most frequent mathematical “families” of indices derived from Table 5. Symbols represent each of the families found in the dendrogram: \cdot , purple family; \diamond , green family; \blacksquare , yellow family; \circ , blue family; $*$, pink family. Relative frequency in percentage is the number of times each result appeared among all the results (i.e., the 30 000 iterations executed), in percentage.

Results	Relative frequency (percentage)
$WF^{\circ} \cdot, D^*$	21.760
$WF^{\circ} \cdot, TF^{\circ} *$	5.054
$WF^{\circ} \cdot, TF^{\circ} *$	2.240
$WF^{\circ} \diamond, D^*$	1.326
$WF^{\circ} \diamond, TF^{\circ} *$	0.808
$wD \cdot, WF^{\circ} \cdot, D^*$	0.573
$WF^{\circ} \cdot, BC^*, D^*$	0.560
$WF^{\circ} \diamond, wD \cdot, D^*$	0.557
$WF^{\circ} \cdot, D^*, TF^{\circ} *$	0.543
$wD \cdot, D^*$	0.330
$WF^{\circ} \cdot, D^*, TF^{\circ} *$	0.260
$WF^{\circ} \cdot, K_{dir}^{\circ}, D^*$	0.237
$WF^{\circ} \cdot, D^*, TF^{\circ} *$	0.233

systems-based ecology that aims to achieve objective conservation priorities or regulations to manage ecosystems. We suggest that our machine learning-based approach to maximize the predictive power of structural analysis can be a major step towards simple and fast, yet quite realistic research on food webs.

CRedit authorship contribution statement

CG: Data curation, Formal analysis, Methodology, Writing - original draft. **AM:** Formal analysis, Methodology. **FJ:** Conceptualization, Funding acquisition, Supervision, Writing - review and editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Research of AM and FJ was funded by the National Research, Development and Innovation Office – NKFIH grants OTKA K 116071 and GINOP-2.3.2-15-2016-00057. CG was funded by an Erasmus scholarship. We thank Drs Francisco Pinto, András Horváth, Attila Csikász-Nagy and two anonymous Reviewers for their helpful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2021.107617>.

References

Allesina, S., Pascual, M., Bourne, P.E., 2009. Googling food webs: Can an eigenvector measure species' importance for coextinctions? *PLoS Comput. Biol.* 5 (9), e1000494.

Baranyi, G., Saura, S., Podani, J., Jordán, F., 2011. Contribution of habitat patches to network connectivity: redundancy and uniqueness of topological indices. *Ecol. Indic.* 11, 1301–1310.

Bauer, B., Jordán, F., Podani, J., 2010. Node centrality indices in food webs: rank orders versus distributions. *Ecol. Complex.* 7, 471–477.

Borrett, S.R., 2013. Throughflow centrality is a global indicator of the functional importance of species in ecosystems. *Ecol. Indic.* 32, 182–196.

del Rio, G., Koschutzki, D., Coello, G., 2009. How to identify essential genes from molecular networks? *BMC Syst. Biol.* 3, 102.

Estrada, E., 2007. Characterization of topological keystone species: local, global and meso-scale centralities in food webs. *Ecol. Complex.* 4 (1–2), 48–57.

Gouveia, C., 2019. Combination of Topological Indices in Network Analysis: a Computational Approach. PhD dissertation. Universidade de Lisboa.

Harary, F., 1959. Status and contrastatus. *Am. Sociol. Assoc.* 22 (1), 23. <https://doi.org/10.2307/2785610>.

Jordan, F., Takacs-Santa, A., Molnar, I., 1999. A reliability theoretical quest for keystones. *Oikos* 86 (3), 453. <https://doi.org/10.2307/3546650>.

Jordán, F., Liu, W.-C., van Veen, J.F., 2003a. Quantifying the importance of species and their interactions in a host-parasitoid community. *Commun. Ecol.* 4 (1), 79–88.

Jordán, F., Báldi, A., Orci, K.M., Rácz, I., Varga, Z., 2003b. Characterizing the importance of habitat patches and corridors in maintaining the landscape connectivity of a Pholidoptera transylvanica (Orthoptera) metapopulation. *Landscape Ecol.* 18, 83–92.

Jordán, F., Liu, W. C., Davis, A. J., 2006. Topological keystone species: measures of positional importance in food webs. *Oikos* 112, 535–546.

Jordán, F., Benedek, Z., Podani, J., 2007. Quantifying positional importance in food webs: A comparison of centrality indices. *Ecol. Model.* 205 (1–2), 270–275.

Jordán, F., Okey, T.A., Bauer, B., Libralato, S., 2008. Identifying important species: Linking structure and function in ecological networks. *Ecol. Model.* 216 (1), 75–80.

Lai, S.-M., Liu, W.-C., Jordán, F., 2015. A trophic overlap-based measure for species uniqueness in ecological networks. *Ecol. Model.* 299, 95–101.

Lau, M.K., Borrett, S.R., Baiser, B., Gotelli, N.J., Ellison, A.M., 2017. Ecological network metrics: opportunities for synthesis. *Ecosphere* 8 (8), e01900. <https://doi.org/10.1002/ecs2.1900>.

Ludovisi, A., 2014. Effectiveness of entropy-based functions in the analysis of ecosystem state and development. *Ecol. Indic.* 36, 617–623.

Ludovisi, A., Scharler, U.M., 2017. Towards a sounder interpretation of entropy-based indicators in ecological network analysis. *Ecol. Indic.* 72, 726–737.

Mórhé, Á., Jordán, F., Szilágyi, A., Scheuring, I., 2009. Overfishing and regime shifts in minimal food web models. *Commun. Ecol.* 10 (2), 236–243.

Mórhé, Á., Endrédi, A., Jordán, F., 2018. Additivity of pairwise perturbations in food webs: Topological effects. *J. Theor. Biol.* 448, 112–121.

Ortiz, M., Levins, R., Campos, L., Berríos, F., Campos, F., Jordán, F., Hermsillo, B., Gonzalez, J., Rodriguez, F., 2013. Identifying keystone trophic groups in benthic ecosystems: implications for fisheries management. *Ecol. Indic.* 25, 133–140.

Ortiz, M., Hermsillo-Núñez, B., González, J., Rodríguez-Zaragoza, F., Gómez, I., Jordán, F., 2017. Quantifying keystone species complexes: ecosystem-based conservation management in the King George Island (Antarctic Peninsula). *Ecol. Indic.* 81, 453–460.

- Pereira, J., Jordán, F., 2017. Multi-node selection of patches for protecting habitat connectivity: fragmentation versus reachability. *Ecol. Indic.* 81, 192–200.
- Pereira, J., Saura, S., Jordán, F., 2017. Single-node versus multi-node centrality in landscape graph analysis: key habitat patches and their protection for twenty birds in NE Spain. *Meth. Ecol. Evol.* doi: 10.1111/2041-210X.12783.
- Pocock, M.J.O., Johnson, O., Wasiuk, D., 2011. Succinctly assessing the topological importance of species in flower-pollinator networks. *Ecol. Complex.* 8 (3), 265–272.
- Scotti, M., Podani, J., Jordán, F., 2007. Weighting, scale dependence and indirect effects in ecological networks: a comparative study. *Ecol. Complex.* 4 (3), 148–159.
- Wasserman, S., Faust, K., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.
- Zotenko, E., Mestre, J., O’Leary, D.P., Przytycka, T.M., Rost, B., 2008. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput. Biol.* 4 (8), e1000140.