Important species may be in critically central network positions in ecological interaction networks. Beyond quantifying which one is the most central species in a food web, a multimode approach can identify the key sets of the most central $n$ species as well. However, for sets of different size $n$, these structural keystone species complexes may differ in their composition. If larger sets contain smaller sets, higher nestedness may be a proxy for predictive ecology and efficient management of ecosystems. On the contrary, lower nestedness makes the identification of keystones more complicated. Our question here is how the topology of a network can influence nestedness as an architectural constraint. Here, we study the role of keystone species complexes in 27 real food webs and quantify their nestedness. After quantifying their topology properties, we determine their keystone species complexes, calculate their nestedness, and statistically analyze the relationship between topological indices and nestedness. A better understanding of the cores of ecosystems is crucial for efficient conservation efforts, and to know which networks will have more nested keystone species complexes would be a great help for prioritizing species that could preserve the ecosystem’s structural integrity.

1. Introduction

Understanding and predicting the robustness and vulnerability of complex ecological networks is a topic of increasing relevance. There is a general agreement that nodes in certain critical network positions may have disproportionately large effects on network functioning. The loss of these key nodes may easily generate cascading effects in the network, so their management is important. These cascading interactions are hard to predict, since secondary effects depend on the particular architecture of the network. Thus, the question of how network topology influences the systemic importance of critical nodes emerges. Focusing research on these key nodes can be one way on how to tame and handle complexity [1] and assess the relative importance of species in ecological communities [2–4].

Various network centrality measures can quantify and identify important network positions [5, 6], and structural analyses [7–9] are increasingly supported by dynamical studies [10, 11]. The latter suggest that key positions may not be identified only by local indices (e.g., node degree). Instead, network measures considering the indirect neighbourhood (e.g., betweenness centrality) of nodes are needed. A number of experimental [12] and modelling [13] works support the importance of indirect effects in biological systems. There is growing interest in nonlocal, mesoscale network indices [5].

Apart from expanding the neighbourhood of focal nodes (increasing the distance for network effects), it has also been suggested that the number of local nodes may also be expanded from 1 to $n$. The centrality of node sets has been discussed [14, 15] and applied in other fields of science (e.g., landscape ecology [16, 17]). This approach suggests that the positional importance of network nodes may not be characterized independently, one by one, but rather simultaneously. Support for the relevance of multispecies vulnerability analyses comes from both empirical (e.g., keystone species complexes [18]) and modelling (multispecies fisheries [19])
directions. Recent attempts have been made to model and determine the identity of keystone species complexes in real ecosystems by network analysis [20–22].

Although the predominant view on network robustness is focused on local and single-node analyses (i.e., degree distribution [8, 23, 24]), here, we take a nonlocal, multinode approach to the problem. In this paper, (1) we quantify the macroscopic (network-level) topological properties of 27 real food webs, (2) we calculate the centrality of their node sets, (3) we quantify the nestedness of the highest centrality sets, and (3) we study the correlation between nestedness and topological network properties. We argue that large nestedness makes the network more predictable and manageable [25], so our results may have implications to the efficiency of conservation efforts.

2. Materials and Methods

2.1. Food Webs. We used 27 food webs freely available from the NCEAS database (http://www.nceas.ucsb.edu/interactionweb). These describe various, mostly terrestrial ecosystems. For the complete species lists and more biological information, see the original source. Before the analyses, we deleted isolated nodes and small components from the networks and focused only on the giant component (this typically means the deletion of only 0–5% of the original nodes). Furthermore, nodes were recoded, so numbering starts with zero.

The food webs are coded as follows: aka a (Akatore A, pine forest, Otago, New Zealand), aka b (Akatore B, pine forest, Otago, New Zealand), ber (Berwick, pine forest, Otago, New Zealand), black (Blackrock, pasture grassland, Otago, New Zealand), broad (Broad, pasture grassland, Otago, New Zealand), cant (Canton, pasture grassland, Otago, New Zealand), carpinteria (Carpinteria salt marsh, California, USA), cat (Catlins, pine forest, Otago, New Zealand), cowl (Coweeta1, pine forest, North Carolina, USA), cow17 (Coweeta17, pine forest, North Carolina, USA), demp au (Dempsters tussock grassland in autumn, Otago, New Zealand), demp sp (Dempsters tussock grassland in spring, Otago, New Zealand), demp su (Dempsters tussock grassland in summer, Otago, New Zealand), german (German, tussock grassland, Otago, New Zealand), healy (Healy tussock grassland, Otago, New Zealand), kyeb (Kyeburn, tussock grassland, Otago, New Zealand), likyke (LilKyeburn, tussock grassland, Otago, New Zealand), martins (Martins, pine forest, Maine, USA), narr (Narrowdale, pine forest, Otago, New Zealand), north (NorthCol, broadleaf forest, Otago, New Zealand), powder (Powder, broadleaf forest, Otago, New Zealand), stony (Stony, tussock grassland, Otago, New Zealand), sutton au (Sutton tussock grassland in autumn, Otago, New Zealand), sutton sp (Sutton tussock grassland in spring, Otago, New Zealand), sutton su (Sutton tussock grassland in summer, Otago, New Zealand), troy (Troy, pine forest, Maine, USA), and ven (Venlaw, pine forest, Otago, New Zealand). Geographic distribution is thus quite narrow, but this does not seem to have any known effect on the results.

2.2. Network Analysis. We calculated nine global (macroscopic) topological properties for each network. The number of nodes (N) and the number of interactions (L) are trivial properties of every network. Their combination provides the connectance (C) (or density) of the network:

\[ C = \frac{2 * L}{N(N-1)}, \]

where undirected interactions are considered with no self-loop. Based on individual node degree values, we can compute a macroscopic network measure, the average degree (avD), calculated for all nodes in the network.

The clustering coefficient (CCi) of node i equals the density of the subnetwork composed of the neighbours of node i. This is the probability that its two neighbours j and k will be directly linked to each other. It can be defined as

\[ CC_i = \frac{2 * |E(G_i)|}{D_i * (D_i - 1)}, \]

where Gi is the subgraph composed of the nodes that are directly linked to node i, |E(Gi)| is the number of edges in this subgraph, and Di is the degree of node i. The whole network can be characterized by the average clustering coefficient calculated for all nodes (avCC), and this can be also weighted by the degree value of particular nodes (weighted clustering coefficient: wCC). The latter gives larger emphasis on clusters around more connected nodes.

The distance between two nodes i and j in a network (dij) is the minimal number of links connecting them (i.e., the length of the shortest path length between i and j). The whole network can be characterized by the average of shortest path lengths (avSPL) and their maximum value (diameter, d). When a network is composed of more than one component, some distance values will be infinite (for nodes m and n belonging to different components). This makes it impossible to calculate distance-based network metrics. In these cases, the reciprocal distance between nodes i and j can be given as

\[ d_{ij}^* = \frac{1}{d_{ij}}, \]

and this measure can be used also when a network consists of more than one component (since the reciprocal of infinity equals, by definition, zero). The distance-weighted fragmentation (DF) of the network can be calculated as

\[ DF = 1 - \frac{\sum_{i,j}^N 2 * d_{ij}^*}{i * j}, \]

which is the average reciprocal distance for each pair of nodes in the network.

We selected these macroscopic network properties because they are simple, yet, they reflect several local
(degree-related), mesoscale (clustering-related) and global (distance-related) properties of the networks.

2.3. Multinode Centrality. Apart from computing the centrality of individual graph nodes, one can define and quantify also the centrality of sets of nodes (see Figure 1). Multinode centrality analyses have already been performed for different types of ecological networks including food webs [26] and habitat networks [27].

The most central multinode sets of \( n = 1 \) to 4 nodes were identified for the 27 food webs, according to two different aspects of key player selection. First, how to best fragment (disrupt) the network by removing \( n \) key nodes (the “negative” version of the key player problem; KPP-Neg) and second, how to best send a message out from \( n \) nodes of the network to others (the “positive” version; KPP-Pos, see [15]). For KPP-Neg, we determined the most central node sets considering binary (\( F \)) and distance-weighted (\( FR \)) fragmentation centrality. For KPP-Pos, we determined the most central node sets considering binary \( m \)-reach centrality (\( Mm \)) and distance-weighted (\( DR \)) reachability with \( m = 1, 2 \) and 3 steps (\( M1, M2, \) and \( M3 \), respectively).

Each of the four multinode centrality measures were computed for \( n = 1 \) to 4 nodes (\( n = 1 \) is clearly single-node). Multinode key sets were calculated using Pyntacle, our high-performance network analysis tool.

2.4. Nestedness. The nestedness of presence-absence ecological data [28] has a rich literature with well-developed methods ([29, 30]; for software, see [31]). The nestedness approach has also been extended to ecological interactions in binary networks [32, 33]. Here, we study the nestedness of ecological interaction networks in a very different way (see [15, 20, 25]), quantifying the set–subset relationships of central nodes in a network.

We calculated the nestedness of central node sets (i.e., the overlap among the sets of size \( n = 1 \) to 4) using the Nrow metric [34]. Nrow is the average percentage of nodes from smaller sets that are contained in larger sets, taking all possible pairs of sets. For example, for the food web \( \text{demp au} \), the \( M2 \) key player sets for \( n = 1 \) to 4 nodes were \{0\} for \( n = 1 \), \{1 \} for \( n = 2 \), \{0 08 76 \} for \( n = 3 \), and \{76 18 37 66 \} for \( n = 4 \). For \( n = 1 \) and \( n = 2 \), there is perfect overlap. For \( n = 1 \) and \( n = 3 \), there is partial overlap, since the smaller set (\( n = 1 \)) is a subset of the larger one (\( n = 3 \)). For \( n = 2 \) and \( n = 4 \), there is no overlap, since the two sets have no common elements. Averaging all the 6 overlaps, we have \( \text{Nrow} = 47.22 \), which is the nestedness value for \( M2 \) in the \( \text{demp au} \) food web (see the species identities for this food web in Discussion). The same was done for the remaining centralities (\( F \), \( FR \), \( M2 \), \( M3 \), and \( DR \)) and for all food webs.

2.5. Statistical Analysis. We compared the 9 topological properties of the 27 food webs with their 6 nestedness metrics by Spearman correlation, because most topological properties were not normally distributed. We considered only correlations of 0.60 and above (as well as −0.60 and below). Correlations were calculated in R 3.3.0 [35].

3. Results

3.1. Network Metrics. The studied macroscopic network parameters are presented in Table 1. The smallest and the largest networks, in terms of the number of nodes, were the \( \text{cat} (N = 48) \) and the \( \text{carpinteria} \) food webs \( (N = 128) \), respectively. Depending on the various actual numbers of links (\( L \)), connectance ranged from \( C = 0.06 \) (\( \text{aka a, cow17, martins, narr, and troy} \)) to \( C = 0.16 \) (\( \text{demp su} \)). Average degree ranged from \( \text{avD} = 4 \) (\( \text{aka b, cow17, and narr} \)) to \( \text{avD} = 18.72 \) (\( \text{carpinteria} \)). Diameter ranged from \( d = 4 \) (\( \text{black, cow17, german, healy, and stony} \)) to \( d = 7 \) (\( \text{cow1} \)), and the average shortest path length ranged from \( \text{avSPL} = 2.19 \) (\( \text{carpinteria} \)) to \( \text{avSPL} = 2.9 \) (\( \text{cow1} \)). The average clustering coefficient ranged from \( \text{avCC} = 0.02 \) (\( \text{cat, kyeb, sutton sp, and sutton su} \)) to \( \text{avCC} = 0.25 \) (\( \text{carpinteria} \)), and the weighted clustering coefficient ranged from \( \text{wCC} = 0 \) (\( \text{broad, sutton sp, and sutton su} \)) to \( \text{wCC} = 0.25 \) (\( \text{carpinteria} \)). Finally, distance-based fragmentation ranged from \( \text{DF} = 0.48 \) (\( \text{carpinteria and demp su} \)) to \( \text{DF} = 0.6 \) (\( \text{troy} \)).

3.2. Nestedness. Our question was if topology has any significant effect on the nestedness of keystone species complexes in the studied 27 food webs. Between 9 topological properties and 6 nestedness metrics for each food web, we analysed 54 correlations. Only 4 of them were significant (shown in Figure 2), and in each of these \( M2 \) was the nestedness index (\( F \), \( FR \), \( DR \), \( M1 \), and \( M3 \) did not show any significant correlation). \( M2 \) correlated positively with \( \text{DF} \) and \( \text{avSPL} \) and
The four significant correlations are between M2 and DF (rho = 0.681; p = 0.0009), M2 and C (rho = −0.678; p = 0.001), M2 and avD (rho = −0.637; p = 0.00035), and M2 and avSPL (rho = 0.605; p = 0.00084). All of them are strongly significant.

Only a few topological features can be used as a proxy for assessing the nestedness of central node sets, but most of these show quite strong correlations. Our results suggest that in networks where shortest paths are shorter and density is higher, nestedness is lower, so systems-based conservation can be less predictive and efficient. One example is the Sutton tussock grassland in springtime (Figure 3(a), Supplementary Material) (available here). Here, the single most central organism in the network is unidentifiable detritus (#0, black in Figure 3(a)). The most central pair is the diatom Cocconeis spp. and the larvae of the riffle beetle Hydrota nitida (#10 and #61, blue). The group of the three most central network positions is the red alga Audouinella spp., the diatom Navicula avenacea, and the caddisfly Pycnocentrodes spp. (#9, #30, and #70, red). The four most central organisms are the alga Epithemia zebra, the diatom Eunotia spp., the fishfly Archi- cauloides diversus, and Chironomid type "Diamesid blond" (#18, #19, #49, and #52, orange). Hence, the increasing core of key organisms is perfectly unnested (M2 = 0, up to 4 groups). Accordingly, DF is low (0.51), C is high (0.14), avD is high (10.49), and avSPL is small (2.39). Apart from the single-node core (n = 1), the larger cores (n > 1) are always composed of both plants (e.g., diatoms) and animals (e.g., caddisfly).

On the contrary, in less connected and less compact networks, nestedness is higher, so a multispecies view fairly reinforce the results of single-species analyses. One example is the Dempsters tussock grassland in autumn (Figure 3(b), Supplementary Material). Here, the single most central organism in the network is unidentifiable detritus (#0, black). The most central pair is unidentifiable detritus and terrestrial invertebrates (#2, blue). The group of the three most central network positions is unidentifiable detritus,
and the caddisflies *Olinga feredayi*, and *Tiphobiosis sp.* (#68 in orange and #76 in red). The four most central organisms are *Tiphobiosis sp.*, alga *Epithemia zebra* (#18, yellow), another alga *Spirogyra sp.* (#37, yellow), and a mayfly *Nesameletus ornatus* (#66 yellow). Here, the composition of the core is a little bit more nested (M2 = 47.22) and, accordingly, DF is somewhat higher (0.53), C is lower (0.12), avD is a little lower (9.88), and avSPL is longer (2.47).

Supplementary Material show the nestedness patterns for each food web. The numbers are the codes for species, and these are generally not comparable for different networks. However, node #0 is almost always *unidentifiable detritus* (or some similarly large aggregated group, e.g., *terrestrial invertebrate remains*). In many networks, this is part of the key player complexes. Biologically speaking, this is an artefact: the detritus is clearly a well-connected component of food webs. Only other species in the key player complexes can be biologically interpreted. It is also noted that *Unidentifiable detritus*, even if it is frequently the key group for \( n = 1 \), is frequently missing from larger key player sets (e.g., for \( n = 4 \) in the *demp au* food web). So, even if it dominates the network structure in itself, its position is not significant anymore if we think in terms of a larger network core.

Apart from the large aggregated groups typically being in the centre of the network, the four organisms that can be in the key position also in single-species cores (\( n = 1 \)) are the diatom *Fragilaria vaucheriae* (#19 in the broad food web), the shore crab *Hemigrapsus oregonensis* (#45 in the carpinteria food web), the mayfly *Deleatidium* spp. (#34 in the north food web), and the diatom *Rhoicosphenia curvata* (#16 in the powder food web). *Hemigrapsus* appears in all of the four studied key player sets in the carpinteria food web \( (n = 1, 2, 3, 4) \).

Some communities are described by several versions of the food web (e.g., seasonal versions like *demp au*, *demp sp*, and *demp su*). In some cases, these versions differ a lot in nestedness (*demp* and *sutton*), while in other cases, there is only a small difference between the versions (*aka* and *cow*).

### 4. Discussion

The dynamical behaviour of complex ecological systems can be dominated by a few critically important components. Finding these could dramatically increase our understanding, the predictability of models, and the efficiency of management efforts. We studied a comparable set of empirical food webs and identified the structurally most important nodes in them. Whether these small sets were nested was correlated to some topological properties of these networks.

Network features influencing nestedness can be regarded as topological constraints on the predictability and efficiency of management and systems-based conservation. It remains unclear to us how M2 and M3 can be negatively and positively correlated with avD, respectively.

We need to much better understand the biology of the key groups and the ecology of nested vs. nonnested
Figure 3: The food webs of the Sutton tussock grassland in spring (a; sutton sp) and the Dempster tussock grassland in autumn (b; demp au). The coloured species are explained in the text.
communities. If certain groups (e.g., zooplankton and diatoms) appear frequently in the core of food webs, these can be thought to be real keystone species. This is especially important if the core is nested: this means that the particular community is really dominated by a single species. We still know nothing about the kinds of communities (or the set of abiotic factors) that can be associated with nested patterns. Biologically speaking, this is the most promising future research line.

All of our results are based on a set of 27 empirical food webs in the size range between 48 and 128 trophic groups. This is the typical size scale for food webs in the literature. All the webs were described by the same methodological standards, so they are comparable to each other. In order to see if these results are generalizable, research is needed in at least two directions.

First, one wants to see if topological properties scale with network size. For this, much larger networks should be studied—and the topological properties studied here can be more and more relevant and interesting for larger graphs. The limitation here is that empirical networks are not larger. Much larger networks ($N > 500$) could be constructed by dramatically increasing the resolution of trophic groups (e.g., by adding bacteria and replacing trophic groups by biological species), but these networks would not be biologically comparable to the present ones (even if being mathematically more interesting).

Second, the toy network of the same size range can be generated by various algorithms (already in progress), and empirical topologies could be compared to the theoretical distributions. This kind of randomization analysis is fairly straightforward in community ecology; however, it is not easy to see which generative algorithms give the most realistic results (e.g., [36] but see [37]). These studies could reveal if the reported relationships are universal properties of networks, or they are specific to only food webs for some biological (ecological) reasons (Capocéfalo et al. unpublished). If the results are food web-specific, we need to understand the biological reasons. If the results will be shown to be of general nature, conclusions can be drawn also in other fields of research. For example, terrorist network webs have been shown to have large average shortest paths and low density [38], properties suggesting that their efficient “management” is possible—in a security and defence sense.

This paper is of mostly conceptual and methodological nature. We suggest that the search for the cores of ecosystem networks opens several research lines that could massively contribute to systems-based conservation biology and management, with applications ranging from marine fisheries to pollination systems.

Data Availability

Data are available at the personal website of the corresponding author: https://ferencjordan.webnode.hu/.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

FJ was funded by the grant GINOP-2.3.2-15-2016-00057. FJ and JP were funded by the grant OTKA K 116071.

Acknowledgments

We are grateful to Zsófia Benedek and Marco Scotti for earlier discussions.

Supplementary Materials

Supplementary Material: the M2 nestedness values for each network ($m$-reach for 2 steps) and the identity of nodes for key player sets (of different sizes $n$) are presented. (Supplementary Materials)

References


