

Article

Aggregating a Plankton Food Web: Mathematical versus Biological Approaches

Ferenc Jordán ^{1,2,3,*} , Anett Endrédi ¹ , Wei-chung Liu ⁴ and Domenico D'Alelio ³

¹ Danube Research Institute, MTA Centre for Ecological Research, 1113 Budapest, Hungary; anett.endredi@gmail.com

² Evolutionary Systems Research Group, MTA Centre for Ecological Research, 8237 Tihany, Hungary

³ Stazione Zoologica Anton Dohrn, Department of Integrative Marine Ecology, 80121 Naples, Italy; domenico.dalelio@szn.it

⁴ Academia Sinica, Taipei 11529, Taiwan; wliu56@gate.sinica.edu.tw

* Correspondence: jordan.ferenc@gmail.com; Tel.: +36-20-428-51-62

Received: 29 October 2018; Accepted: 14 December 2018; Published: 19 December 2018



Abstract: Species are embedded in a web of intricate trophic interactions. Understanding the functional role of species in food webs is of fundamental interests. This is related to food web position, so positional similarity may provide information about functional overlap. Defining and quantifying similar trophic functioning can be addressed in different ways. We consider two approaches. One is of mathematical nature involving network analysis where unique species can be defined as those whose topological position is very different to others in the same food web. A species is unique if it has very different connection pattern compared to others. The second approach is of biological nature, based on trait-based aggregations. Unique species are not easy to aggregate with others because their traits are not in common with the ones of most others. Our goal here is to illustrate how mathematics can provide an alternative perspective on species aggregation, and how this is related to its biological counterpart. We illustrate these approaches using a toy food web and a real food web and demonstrate the sensitive relationships between those approaches. The trait-based aggregation focusing on the trait values of size (*sv*) can be best predicted by the mathematical aggregation algorithms.

Keywords: food web; plankton; redundancy; keystone species; trophic roles

1. Introduction

Community ecology, a major research area of ecology, focuses on the coexistence of multiple species. A rich network of several interaction types glues together complex, multispecies ecosystems [1,2]. These interaction types include, for example, competition, predator-prey interactions, mutualism, facilitation, and others. In this kind of community network, the ecological functioning and evolutionary success of the species also depend on how the others perform, e.g., extinctions can cascade through the system and trigger secondary extinctions. The disappearance of some species results in significant community-level changes, while the consequences of many others' dysfunction are hard to realize. The concept of keystone species emerged out of this observation: Disturbing these species will generate a disproportionately large community response, mediated by several possible mechanisms [3].

Since it is recognized, there is a great interest to quantify the importance of species and to identify keystones, which can lead us closer to understand better the roles and functions species play in the ecosystem and, ultimately, the level of redundancy in the system [4]. Network analysis can be an appropriate approach to estimate the impact of keystone species on the community which they belong to, since inter-specific interactions can be comfortably modelled by graphs. While many ecologists

agree that predator-prey effects dominate in ecological dynamics, in many cases, food webs (i.e., the network of who eats whom) can be used as a representation of the ecosystem's interaction structure. Related to the trophic dimensions of the ecological niche concept [5], food webs are a good start to quantify the trophic roles of species in ecosystems [6,7].

An issue strongly related to how to quantify roles and positions in food webs is aggregation. Raw data are always aggregated (to some extent) into functional groups. The resolution of these are typically lower than the species-level (multi-species groups, e.g., benthic invertebrates), but sometimes can even be higher (separating different developmental stages of the same species, e.g., adult/juvenile herring). One of the key questions is how the structural and dynamical properties of the food web change by aggregation. For instance, some functional groups contain dozens of species: Ideally, although these are taxonomically different, they can be considered ecologically quasi-equivalent. If the number of species is large in a functional group, then each of its species can be considered quite redundant (e.g., all of them are detritivores).

Research on aggregation had already been focused on its effects on network properties from an empirical perspective [1,8,9] and from a modelling perspective [10–14]. According to the result of these studies, aggregation can have a huge impact on food web properties, thus it would be of great value to standardize methodology. A key question is how to set the standards to aggregation algorithms: Should the latter be biological considerations or mathematical rules? Trait-based approaches suggest that large databases of biological properties can effectively help categorize all species into coherent functional groups [15]. Conversely, more precise and objective—but blind to expertise—indicate the opportunity to adopt mathematical criteria in pursuing aggregation operations.

Since many research results show the importance of some key traits in ecology (e.g., body size, [16]), ecologists always used traits or trait combinations for aggregation while constructing food webs: E.g., a functional group like “large benthic invertebrates” is a combination of three traits (size, habitat, taxonomy). The most pressing question is now how to minimize the number of traits which still efficiently categorizes organisms into functionally meaningful ecological units (and which traits should be used).

A simple, but elegant, definition of keystone species is that they are single-species functional groups [17]. This means that they are really unique, being responsible for a particular combination of ecological processes and there is no one like them. They are not easily replaceable, so if they go extinct, an ecological function disappears with them. To some extent, such a definition is coded in the interaction network, so that topological analyses may help reveal keystones and indicate how to quantify their roles and importance. It is noted here that the latest trend is to search for a small number of species in a keystone species complex [18], functioning as the “core” of the ecosystem [19].

In this paper, we discuss how mathematics can help us define similar groups (for aggregation). We (1) present some methods for quantifying the positional similarity of nodes in networks, (2) we illustrate these methods on a simple toy network and then (3) we perform the same analyses on a real food web. Finally, (4) we discuss the limits and challenges of this approach.

2. Materials and Methods

2.1. Data

For illustration, we use a toy network (Figure 1) and a real food web (Figure 2). The real food web describes the oligotrophic state of Gulf of Naples, as sampled at the Long-Term Ecological Research station MareChiara (LTER-MC, [20]). This network describes the planktonic food web composed of phytoplankton (autotrophic unicellular organisms), micro-zooplankton (phago-trophic unicellular organisms), and meso-zooplankton (small invertebrates) [21,22]. The level of trophic resolution in the network is much higher than the average since a large amount of biological information is available for all planktonic organisms in the ecosystem.

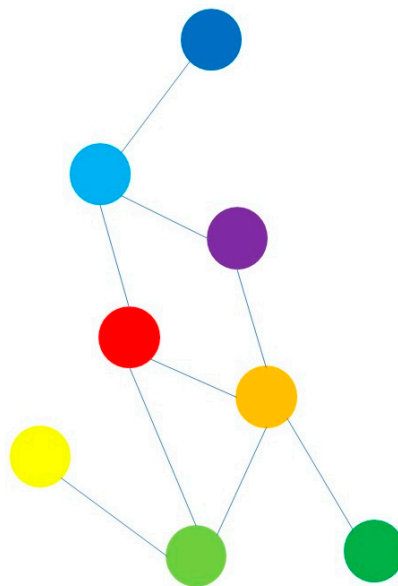


Figure 1. A toy food web for illustration. The dark blue and the yellow nodes are top-predators and the light green and dark green nodes are the producers. The direction of carbon flows is not shown (from the bottom upwards) for simplicity.

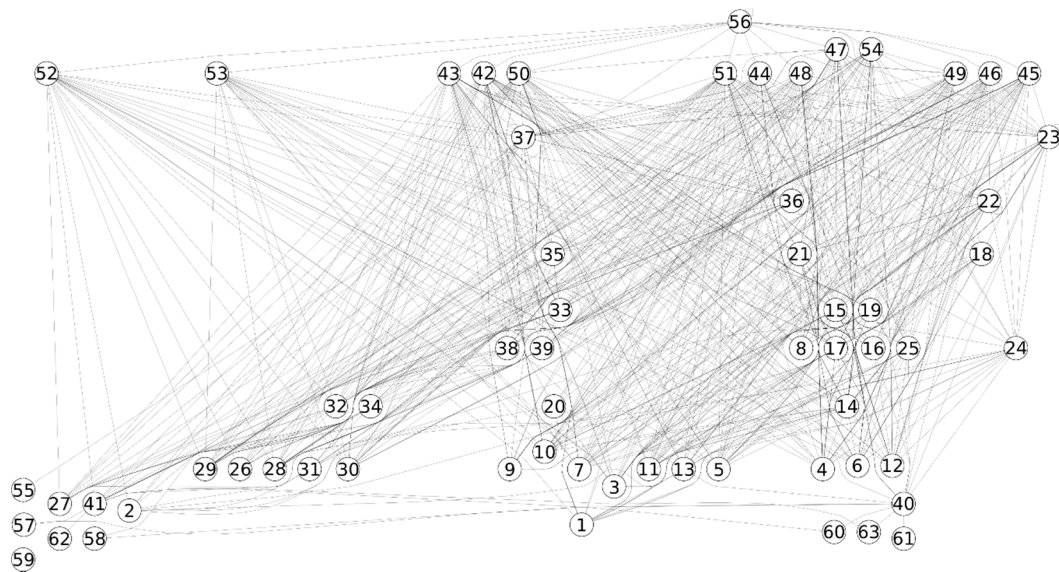


Figure 2. The food web of the Gulf of Naples at LTER-MC.

The main property of the plankton food web investigated therein is the low specialization of organisms resulting in a marked, though partial, trophic redundancy in most functional nodes. The effort made in producing a network-model for the Gulf of Naples plankton [21] was directed toward integrating trophic diversity and community functioning: Therefore, taxonomic aggregation was limited to the aggregation of abundances of sibling species (i.e., belonging to the same biological genus) that formed a unique functional node in the resulting network. By taking the latter option, functional nodes included organisms having the same trophic properties.

2.2. Methods

In both networks, we analyzed the similarity of network nodes based on the overlap in direct and indirect interactions (TO, STO) and regular equivalence (REGE). While TO and STO consider undirected networks, the REGE approach is based on directed predator-prey interactions. The former

method expresses the number of shared preys and predators and quantifies unique vs. redundant food web positions in this sense. The latter method quantifies feeding types and guild identity. These approaches are described below.

2.2.1. Topological Overlap (TO, STO)

Müller et al. [23] developed a methodology for quantifying the interaction strength between species in a host-parasitoid community and Jordán et al. [24] later then generalized such an approach for food webs. The principle here is that one species can affect the other via direct interaction and via indirect pathways.

Let us consider a food web of N species, and if two species i and j are connected then the direct effect of i on j (or the one-step effect) is:

$$a_{ij} = \frac{1}{D_j}, \quad (1)$$

where D_j is the number of nodes directly connected to j (i.e., its degree). Here, i strongly affects j if i is j 's only neighbor; whereas i can only affect j weakly if j has many neighbors. We then construct a square matrix $\mathbf{A}_{(1)}$ where the ij th element, $A_{(1)ij}$, is the one step effect of i on j . Two-step effects between species can be quantified by using matrix multiplication:

$$\mathbf{A}_{(1)} \times \mathbf{A}_{(1)} = \mathbf{A}_{(1)}^2 = \mathbf{A}_{(2)}, \quad (2)$$

where the ij th element of matrix $\mathbf{A}_{(2)}$, namely $A_{(2)ij}$, is the 2-step effect of i on j . Three-step effects can be obtained by calculating $\mathbf{A}_{(1)}^3$ (i.e., resulting in matrix $\mathbf{A}_{(3)}$), and in general we only need to calculate $\mathbf{A}_{(1)}^n$ for n -step effects (i.e., resulting in matrix $\mathbf{A}_{(n)}$).

The matrix $\mathbf{A}_{(n)}$ contains some interesting information. It can be partitioned into two matrices:

$$\mathbf{A}_{(n)} = \mathbf{Q}_{(n)} + \mathbf{R}_{(n)}. \quad (3)$$

In Equation (3), $\mathbf{Q}_{(n)}$ is a diagonal matrix where its ii th element, $Q_{(n)ii}$, represents the n -step long self effect of species i on itself. $Q_{(n)ii}$ is non-zero if there exists at least one loop of length n linking species i to itself. $\mathbf{R}_{(n)}$ is a hollow matrix where its ij th element, $R_{(n)ij}$, is the n -step long effect of species i on species j . $R_{(n)ij}$ is non-zero if there exists at least one pathway of length n linking species i and j . Both $Q_{(n)ii}$ and $R_{(n)ij}$ can be partitioned into effects associated with individual pathways. For instance, consider a simple food web with 4 species where species i consumes species k and h , and both species k and h in turn consume species j . For $n = 2$, the 2-step long self-effect of species i on itself is:

$$Q_{(2)ii} = a_{ik}a_{ki} + a_{ih}a_{hi}. \quad (4)$$

In Equation (4), $a_{ik}a_{ki}$ is the effect associated with the pathway “node i –node k –node i ”, which is the product of two direct effects (i.e., the direct effect of i on k and the direct effect of k on i); and $a_{ih}a_{hi}$ is the effect associated with the pathway “node i –node h –node i ”, which is also the product of two direct effects (i.e., the direct effect of i on h and the direct effect of h on i). Similarly, the 2-step long effect of species i on species j is:

$$R_{(2)ij} = a_{ik}a_{kj} + a_{ih}a_{hj}, \quad (5)$$

where $a_{ik}a_{kj}$ is the effect associated with the pathway “node i –node k –node j ” and $a_{ih}a_{hj}$ is the effect associated with the pathway “node i –node h –node j ”.

Since there may be several paths of various lengths between two given species i and j in a food web, we can quantify the average effects of species i on j up to path length n :

$$E_{(n)ij} = \frac{1}{n} \left(A_{(1)ij} + A_{(2)ij} + A_{(3)ij} + \dots A_{(n)ij} \right) \quad (6)$$

We then construct an interaction matrix $E_{(n)}$, where the ij th element is $E_{(n)ij}$. Given the way how $E_{(n)ij}$ is defined, its value therefore falls within the interval $[0, 1]$. Taking the advantage of the interaction matrix $E_{(n)}$, we can quantify the positional uniqueness of individual species as follows [25]. First, for a threshold effect size T , we construct an interactor matrix M_T . The ij th element of this matrix, M_{Tij} , takes the label S if $M_{Tij} > T$ (i.e., j is i 's strong interactor); if not then M_{Tij} takes the label W (i.e., j is i 's weak interactor). The i th row of matrix M_T can be considered as the interactor profile of species i , which indicates what type of interactor (strong or weak) another species is to species i . Second, for a species pair ij , we compare the i th row with the j th row of the matrix M_T ; and the number of “ S ” matches, namely TO_{Tij} , is a measure of trophic overlap between species i and j . A large TO_{Tij} value indicates species i and j share many strong interactors, whereas a small value says that they share few strong interactors. After all TO_T values have been computed for all species pairs, we then put them in a square matrix TO_T where the ij th element is TO_{Tij} . Finally, the extent of trophic overlap between species i and all other species in the same food web, namely TO_{Ti} , can be quantified by summing up the i th row of the TO_T matrix. A given species i is truly unique if it has a small TO_{Ti} value, because it shares fewer strong interactors with all other species in the same food web (e.g., an aggregated trophic guild); whereas a large TO_{Ti} indicates the redundancy of species i 's role in the interaction structure of the food web (e.g., many generalist species).

This method can be improved by taking into account all information in species' interactor profiles. Here, TO_{Tij} is now the number of “ S ” matches plus the number of “ W ” matches between the i th and the j th rows of the matrix M_T . In other words, we now define the extent of trophic overlap between species i and j by counting the number of strong and weak interactors they have in common. As before, the sum of the i th row (TO_{Ti}) of matrix TO_T is the unique value of species i for threshold effect size T . A different T value results in a different TO_T matrix and therefore gives a different uniqueness value for species i (i.e., TO_{Ti}). Thus, a better uniqueness measure should take into consideration information derived from different T values. Since the minimum and the maximum values of an element of the interaction matrix $E_{(n)}$ are 0 and 1 respectively, we only need to explore T within the interval $[0, 1]$ by using a suitable increment value. A suitable increment value of T is chosen such that any smaller increment values don't change the outcome of our analysis. Here, we systematically vary T from 0 to 1 in increments of, for example, 0.001. When $T = 0$, all species will have the same strong interactors (as they have no weak interactors) and they will have the same species uniqueness values (i.e., TO_{Ti}). As we move T away from 0, TO_{Ti} values will start to diverge. Those TO_{Ti} values will be most heterogeneous when T reaches a certain value, and beyond which TO_{Ti} values will then converge. When $T = 1$, all species will have the same species uniqueness values as they all have the same weak interactors (and they have no strong interactors). This extreme situation would mean that each species in an ecological community is largely independent of the others (or only very weakly influenced by them).

For each species i , a unique TO profile can be obtained when TO_{Ti} is plotted against different values of T ; and the sum of all TO_{Ti} values across the entire range of T values is now the new measure for species i 's uniqueness (STO_i). A unique species tend to have a small STO_i value as it has few strong and weak interactors in common with all other species across the entire range of T values [26]. This describes a situation when, for example, a herbivore feeds on rare plants and do not consume abundant, dominant plants consumed by many others.

2.2.2. Regular Equivalence (REGE)

The positional similarity of network nodes can be calculated by using the regular equivalence measure (REGE, [6,27,28]). This measure quantifies the similarity between the positions of network nodes i and j based on their network neighborhood. Briefly, two nodes are said to be regular equivalent if they are connected with the same types of nodes. In ecology terms, two species are more regularly equivalent if they have similar (but not necessarily the same) predators and preys: For example, two canopy insects feeding on leaves of different trees and consumed by different bird species. A REGE

matrix **S** is the output of such an analysis, where the ij th element, S_{ij} , expresses the extent of similarity between nodes i and j . An iterative algorithm is used to determine **S** [6]. First, we define a $N \times N$ matrix $\mathbf{R}_{(t)}$ whose ij th element, $R_{(t)ij}$, is the extent of regular equivalence between i and j at iteration t . Second, we carry out the following procedures:

1. At iteration 0, all node pairs are perfectly equivalent (i.e., $R_{(0)ij} = 1$).
2. At iteration $t + 1$, the extent of regular equivalence between i and j , $R_{(t+1)ij}$, is determined as follows.
 - (a) For outgoing links only, for each neighbor k of i (i.e., species i and its predator k), we determine which neighbor m of j (i.e., species j and its predator m) that is most equivalent to k according to $\mathbf{R}_{(t)}$ (i.e., the largest $R_{(t)km}$); and then we define a quantity $X_{i,k,j}$ which takes the value of $R_{(t)km}$. Likewise, from the perspective of j , we determine the value of $X_{j,m,i}$.
 - (b) For incoming links only, for each neighbor h of i (i.e., species i and its prey h), we determine which neighbor n of j is most equivalent to h according to $\mathbf{R}_{(t)}$ (i.e., the largest $R_{(t)hn}$); and we then define a quantity $Y_{i,h,j}$ which takes the value of $R_{(t)hn}$. Similarly, from the perspective of j , one determines the value of $Y_{j,n,i}$.
 - (c) The extent of regular equivalence between i and j at iteration $t + 1$ is defined as:

$$R_{(t+1)ij} = \frac{\sum_k X_{i,k,j} + \sum_m X_{j,m,i} + \sum_h Y_{i,h,j} + \sum_n Y_{j,n,i}}{\text{MAX}(\sum_k X_{i,k,j} + \sum_m X_{j,m,i} + \sum_h Y_{i,h,j} + \sum_n Y_{j,n,i})},$$

where the denominator is the maximum possible value of the numerator if i and j are perfectly equivalent. $R_{(t+1)ij}$ is bounded between 0 and 1, with 1 indicating i and j are perfectly equivalent.

3. We repeat procedure 2 after a predefined number of iterations and let matrix **S** be the final $\mathbf{R}_{(t+1)}$ matrix.

The sum of the i th row of matrix **S** provides a measure of redundancy. Beyond numerical results, we illustrate the similarities for all nodes by using a dendrogram. This dendrogram can be cut at any threshold level in order to define and create aggregated functional groups. An analysis was carried out by using the UCINET software [29]. Similar approaches have already been suggested in ecology (cf. tropho-species [7,30]).

2.2.3. Trait-Based Similarity Measures

For real networks, biological information is also available for characterizing graph nodes. These are (recently) called “traits”; large trait databases are meant to create the chance for the future “big data ecology”. We used 3 traits (b for biomass, s for size and c for carbon content) and used each of them for aggregation in 3 ways. Carbon content and biomass are generally correlated, yet we used both traits as carbon content is an individual-level property (i.e., quantifying the body mass of a single individual belonging to a functional node), while biomass is a population-level property (i.e., quantifying the overall mass of a functional node). We (1) aggregated species with equal trait values (e for equal), (2) aggregated species belonging to 30 evenly defined range of traits values (v for value) and (3) aggregated species based on their trait value ranks in groups of 3 (r for rank). Out of the 9 possible combinations only 6 seemed to be relevant for further analysis (sv , ce , cr , cv , be , bv): Three other possibilities were not considered, because of measurement problems (se and sr) and providing irrelevant information (br). In each case, aggregated groups were named by letters (while numbers refer to the original trophic groups in References [21,22]). Finally, we aggregated the species also by their trophic status (ts , with categories like autotrophs (coded as A), mixotrophs (M), heterotrophs (H)

and detritivores (D). Altogether, the above-mentioned procedures provided 7 ways of aggregation and 7 low-resolution, aggregated food webs.

3. Results

In the toy network (Figure 1), the computation of the REGE half-matrix (Table 1B) is based on the adjacency matrix (Table 1A) and the calculations of the $TO^{2;0.15}$ half-matrix ($n = 2$, $T = 0.15$; Table 1D) is based on the TI^2 -matrix (Table 1C). For both REGE and TO, a ranking of graph nodes can be given (Table 1B,D, on the left). Nodes r (red) and o (orange) are in the most redundant positions according to both approaches. The rank of node p (purple) shows the largest difference: REGE suggests that its topological position is more redundant, while TO suggest it to be more unique. Based on REGE (Figure 3), its position is quite similar to that of node r (red), but TO considers the difference larger, because of the link between r (red) and lg (light green).

Table 1. The adjacency matrix (A), the regular equivalence (REGE) half matrix (B), the TI matrix (C) and the TO half matrix for the toy network. For REGE and TO we also present the ranking of the nodes. Their color code is explained in Figure 3.

A									B										
	db	dg	lb	lg	o	p	r	y	db	dg	lb	lg	o	p	r	y	db	sum	rank
db	0	0	0	0	0	0	0	0	0	67	0	28	49	50	56	56	db	250	r
dg	0	0	0	0	1	0	0	0	0	18	89	55	49	36	0	0	dg	247	o
lb	1	0	0	0	0	0	0	0	0	39	68	85	83	53	53	53	lb	346	p
lg	0	0	0	0	1	0	1	1	0	65	61	51	0	77	89	58	lg	305	lb
o	0	0	0	0	0	1	1	0	0	77	89	58	0	94	45	67	o	412	lg
p	0	0	1	0	0	0	0	0	0	94	45	67	0	0	0	0	p	411	db
r	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	r	420	dg
y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	y	223	y

C									D										
	db	dg	lb	lg	o	p	r	y	db	dg	lb	lg	o	p	r	y	db	sum	rank
db	0.17	0	0.17	0	0	0.08	0.06	0	0	2	0	0	0	2	2	0	db	6	r
dg	0	0.13	0	0.04	0.13	0.06	0.04	0	0	0	0	0	0	0	0	0	dg	0	o
lb	0.5	0	0.64	0.06	0.1	0.25	0.17	0	0	1	2	3	4	0	4	0	lb	12	lb
lg	0	0.13	0.06	0.26	0.17	0.06	0.21	0.5	0	4	0	4	2	1	5	2	lg	11	lg
o	0	0.5	0.14	0.22	0.27	0.25	0.22	0.17	0	1	5	2	0	3	0	0	o	14	p
p	0.17	0.13	0.17	0.04	0.13	0.15	0.1	0	0	0	3	0	0	0	2	0	p	9	db
r	0.17	0.13	0.17	0.21	0.17	0.15	0.15	0.17	0	0	0	0	0	0	2	0	r	20	y
y	0	0	0	0.17	0.04	0	0.06	0.17	0	0	0	0	0	0	0	0	y	6	dg

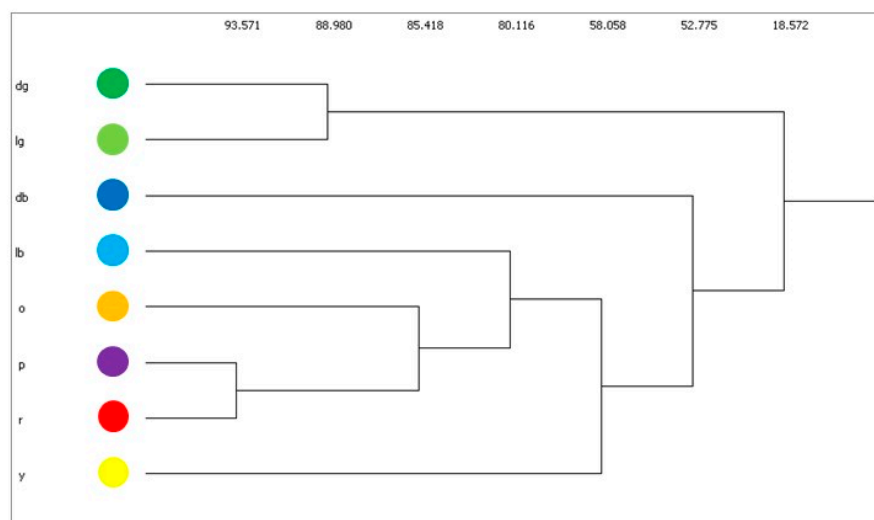


Figure 3. The REGE-based similarity dendrogram for the toy network, shown in Figure 1. The x-axis shows the similarity of network positions: The path is shorter between similar nodes. Nodes are marked by both colors and abbreviations (dg, dark green; lg, light green; db, dark blue; lb, light blue; o, orange; p, purple; r, red; y, yellow).

For the food web of Gulf of Naples (Figure 2), the computation of the REGE half-matrix (Table S1b) is based on the adjacency matrix (Table S1a), and the calculation of the $TO^{2;0.02}$ half-matrix (Table S1d) is based on the TI^2 -matrix (Table S1c). For both REGE and TO, a ranking of graph nodes can be given (Table S1b,d, on the left). Nodes #31 #16 and #13 are in the most redundant positions according to REGE (see also Figure 4), and nodes #51, #42 and #46 are the most redundant ones according to TO. Interestingly, while the two approaches show quite a coincidence for the small toy network, the ranks are very different for this much larger web (see Figure 5). Yet, for the end of the ranks (the less redundant, more unique topological positions), the coincidence is much stronger (#56, #61 and #60 for REGE and #61, #55 and #20 for TO, see Table S1b,d, from the bottom up).

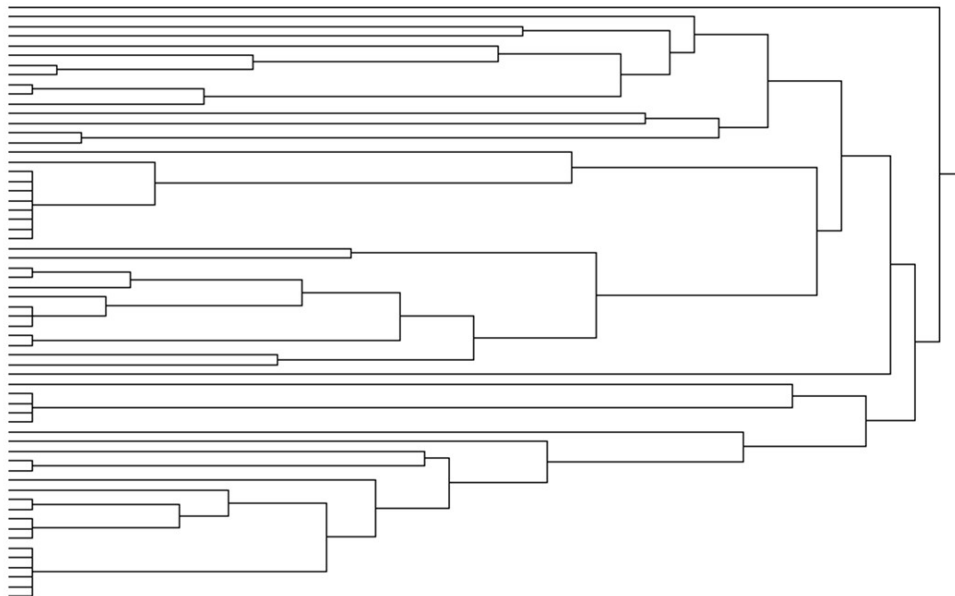


Figure 4. The REGE-based similarity dendrogram for the Gulf on Naples food web, shown in Figure 2. The sum of horizontal distances shows similarity of network positions: The path is shorter between similar nodes. Node identity is not shown (for technical reasons).

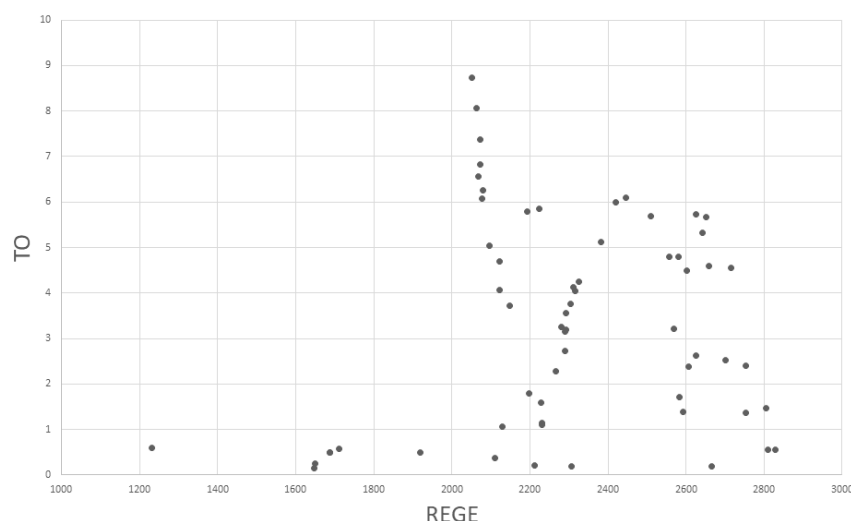


Figure 5. The relationship between the REGE and TO values for each node in the Gulf of Naples food web (Figure 2).

For this latter network, we studied two additional features of the aggregation process. One potential reason for the difference is the inappropriate use of the threshold value for calculating TO^n (for both networks, so far, the average of all TI -values was used, rounded up to second decimal).

TO-values for different thresholds are shown in Table S2 and visualized in Figure 6a. Another potential reason is that STO may provide different results. Table S3 shows the nodal STO-values visualized in Figure 6b. Table S4 shows the biological organisms represented by the graph nodes of the Gulf of Naples food web.

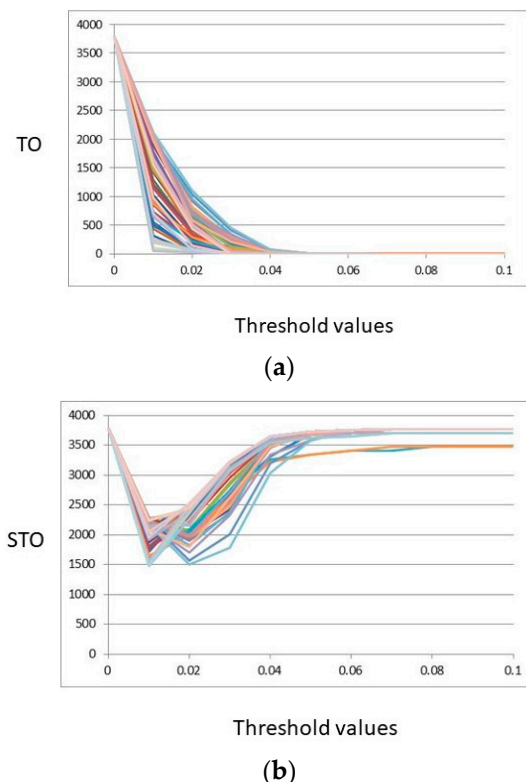


Figure 6. The TO (a) and STO (b) profile of nodes in the Gulf of Naples food web for different thresholds ($T = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1$). Every node is represented by a colorful curve. Note that for $T = 0$ we see the maximal value of the indices.

Following the mathematical analysis of node similarity, we take a trait-based view and use biological knowledge for defining similar nodes and aggregate the network accordingly. In Figure 7, we show all aggregated versions of the original food web, shown in Figure 2. Aggregating according to trophic status (ts), the results are quite trivial: Heterotrophs and mixotrophs consume detritus and autotrophs. According to aggregation based on carbon values (cv), we get 5 nodes like A: #49 (juvenile Calanoids), B: #54 (*Oithona* spp), C: #56 (carnivores), D: #52 (salps) and E: everything else. According to biomass values (bv), we have A: #62 (generic particulate detritus), B: #41 (heterotrophic bacteria), C: #29 (coccolithophorids) and #56 (carnivores) and D: everything else. According to size values (sv), we have A: #56 (carnivora), B: #52 (salps), C: #50 (Appendicularia) and #57 (Appendicularia houses), D: #45 (*Acartia clausii*) and #46 (*Temora stylifera*) and #47 (*Centropages typicus*) and #48 (other calanoids) and #51 (doliolids), E: Everything else. According to carbon equality (ce), carbon ranks (cr) and biomass equality (be), most groups remain un-aggregated. The most unique nodes can be seen in Table 2.

When comparing the most unique network positions based on REGE (e.g., #56), TO (e.g., #60), STO (e.g., #51) and trait-based aggregations (e.g., #56), we derived the following conclusions: (i) The relationship between TO and STO was quite sensitive to the threshold used—i.e., when the latter is set between 0.01 and 0.02, this relationship changes sign and it becomes continuously weaker (Figure S5); (ii) REGE correlates a little better with TO (Figure S6) than with STO (Figure S7); and (iii) in both cases, changing the threshold results in quantitative effects, but not in qualitative ones.

The organisms that were unaffected by the bv type of trait-based aggregations could be characterized by any specific combination of TO and REGE (Figure S8). Separation on the TO/REGE

plain was better for *cv* (Figure S9) and especially *sv* (Figure S10). Similar results hold for the STO/REGE plain for *bv* (Figure S11), *cv* (Figure S12) and *sv* (Figure S13).

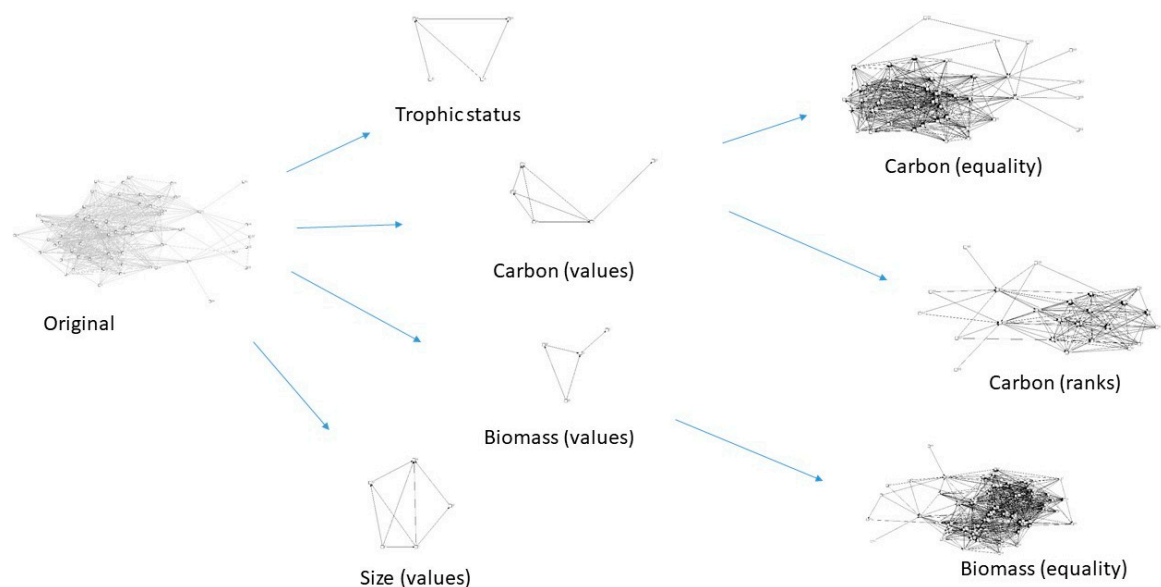


Figure 7. On the left, the original food web, shown in Figure 2. In the middle and on the right, its aggregated versions corresponding to 7 different ways of aggregation.

Table 2. The identity of nodes in the not aggregated or only weakly aggregated trophic groups (A, B, C, D) according to 3 different trait-based aggregation procedures (*cv*, *bv*, *sv*; see explanation in the text).

Scenarios	<i>cv</i>	<i>bv</i>	<i>sv</i>
A	49	62	56
B	54	41	52
C	56	29, 56	50, 57
D	52		45, 46, 47, 48, 51
E	others	others	others

4. Discussion

Aggregating taxa to manage ecological complexity and, at the same time, reduce computational complexity is a key issue of ecology studies on plankton food-webs, like the one investigated herein. This need stems from the huge diversity of microscopic organisms present in aquatic environments, even in small water volumes [31]. Based on our results, aggregating planktonic organisms by size values appears as the best option among the available ones, since it allows partially matching the need to produce a reliable food-web including ecologically relevant planktonic consumers. Yet, a sufficient coverage for trophic behavior is not fully accomplished by aggregating by trophic status or biomass, while carbon-aggregation can be useful to some extent.

Most plankton models published so far tended to compress plankton diversity within few trophic groups, owing to a longstanding tradition based on a very simple representation of a plankton trophic chain including inorganic nutrients as an input for a single phytoplankton group and the latter being food for a single zooplankton group (i.e., the so-called NPZ modelling scheme, [32]). Conceptual advancements indicate the importance of expanding the biological resolution in plankton models by separating between micro- (size < 200 μm) and meso- (200 μm < size < 2 cm, approximately) -plankton and also by distinguishing between a number of meso-zooplankton sub-groups characterized by distinct trophic behaviors [33].

Our mathematical study suggests that size-aggregation can give rise to functional nodes representing, although partially, all the major trophic levels in the plankton food web of the Gulf of

Naples, with an opportune expansion in the meso-zooplankton assemblage that results aggregated in 4 nodes, as follows: (A) Carnivores (#56); (B) salps (#52); (C) Appendicularia and Appendicularia houses (#50,57), (D) *Acartia clausii*, *Temora stylifera*, *Centropages typicus*, other calanoids, doliolids (#45–48,51) (Figure 7). A partial reliability of size aggregation in the plankton food-web stems from the comparison between the results of the present mathematical effort and the ecological properties of the plankton food-web analyzed, which are synthesized in previous works [21,22]. Such a comparison is summarized below.

Firstly, carnivores in the Gulf of Naples food-web (A) are represented mainly by arrow-worms (Chaetognata), which set at the highest trophic level and have a considerably different trophic behavior from all the other nodes represented in the aggregated web. Salps and Appendicularia (B,C) mainly feed on microbes and they set at a lower trophic rank in respect to carnivores; they both are filter-feeders of small plankton particles, but, by being of different sizes, they fall into different nodes and this discrimination captures correctly their slightly different affinities for microbes of different sizes [34]. Appendicularia houses (#57) are opportunely considered in the (C) node, since they constitute a particular detritus-form derived from living Appendicularia individuals. Calanoid copepods (Crustacea, D) set well all together since they are reciprocally closely related, show very plastic diets, in comparison with animals in B and C, which are more focused on small particles. Nonetheless, doliolids, which are more closely related with salps and Appendicularia, are aggregated together with copepods and this can represent a limit to size aggregation.

For the reasons shown above, the analytical approaches presented herein can set an opportune threshold for the aggregation of meso-zooplanktonic animals. When coupled with food-web topology, size-based aggregation is an apparently effective criterion for isolating some of the main meso-zooplankton nodes. In addition to size, carbon-based aggregation can also help identify some other important groups with particular trophic characteristics, like *Oithona* spp. (#54), which have a relatively small carbon value, but a relatively high trophic position—i.e., they eat on other meso-zooplankton. Herein, we evaluated various ways of aggregation focusing on single particular traits—but combining different traits seems to be a further opportune direction.

One might by the way notice in the analyses shown herein the general aggregation of microbial plankton, from phyto- to micro-zooplankton, in one and a single functional node: This is an important limit of all the aggregation criteria followed herein. Previous researches remark the fundamental role of heterotrophic and mixotrophic (i.e., contemporarily photosynthetic and phagotrophic) micro-zooplankton in driving plankton food-webs [35]. Therefore, it may be helpful to repeat the aggregation exercise herein presented but limiting the analyses to the sole functional node E, i.e., the one aggregating everything but some meso-zooplankton nodes according to the size-criterion. Furthermore, one should investigate further the opportunity to aggregate cladocerans (Crustacea #42,43) all in the E group: At first glance, this option may be justified by the relatively small size and carbon content of these animals in comparison with the other meso-zooplankton and they also have a diet similar to micro-zooplankton. In future studies, more detailed demographic information of each species should be considered in the aggregation process. These include physiological characteristics, such as growth rate, reproductive strategy and generation time. Aggregation based on these criteria is important in studying the behavior of dynamic models as different aggregation strategies surely result in different functional groups being modelled and consequently resulting in different model dynamics. Finally, future analyses may be considering weighted networks produced based on trophic flows, which can provide more realistic results.

Our systemic view provides a quantitatively holistic view on the structure of ecosystems. Earlier studies have shown that REGE can result in taxonomically both homogeneous and heterogeneous groups [14]: Heterogeneous groups can reveal ecological functional similarity not reflected in taxonomic closeness. We are suggesting that comparing biological and mathematical definitions of similarity and the consequent aggregation methods can provide standards and, at the same time, consider biological knowledge for better understanding of ecological functionality. Based

on a single case study, we prefer not to make major statements. It is still a long way to go if we really want to understand (1) how are mathematical and trait-based aggregations related to each other and (2) how to use mathematics in order to replace (predict) biological aggregation if big databases are lacking.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2227-7390/6/12/336/s1>. Table S1: The adjacency matrix (A), the REGE half matrix (B), the TI matrix (C) and the TO half matrix for the Gulf of Naples network. For REGE and TO we also present the ranking of the nodes. Table S2: The ranking of nodes based on their TO-values for different thresholds ($T = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06$). Table S3: The ranking of nodes based on their STO-values for different thresholds ($T = 0.01, 0.02, 0.03, 0.04, 0.05, 0.06$). Table S4: The identity of graph nodes in the Gulf of Napoli food web. Figure S1: Correlations between TO and STO for several threshold values. Figure S2: Correlations between TO and REGE for several threshold values. Figure S3: Correlations between STO and REGE for several threshold values. Figure S4: The position of the not aggregated (A: #62; B: #41) or only weakly aggregated (C: #29 and #56) trophic groups in the TO/REGE plain. All organisms aggregated into E (everything else) are shown. Based on the “bv” aggregation (see Table 2). The plots are shown for 6 threshold values. Figure S5: The position of the not aggregated (A: #49; B: #54; C: #56; D: #52) trophic groups in the TO/REGE plain. All organisms aggregated into E (everything else) are shown. Based on the “cv” aggregation (see Table 2). The plots are shown for 6 threshold values. Figure S6: The position of the not aggregated (A: #56; B: #52) or only weakly aggregated (C: #50 and #57; D: #45, #46, #47, #48 and #51) trophic groups in the TO/REGE plain. All organisms aggregated into E (everything else) are shown. Based on the “sv” aggregation (see Table 2). The plots are shown for 6 threshold values. Figure S7: The position of the not aggregated (A: #62; B: #41) or only weakly aggregated (C: #29 and #56) trophic groups in the STO/REGE plain. All organisms aggregated into E (everything else) are shown. Based on the “bv” aggregation (see Table 2). The plots are shown for 6 threshold values. Figure S8: The position of the not aggregated (A: #49; B: #54; C: #56; D: #52) trophic groups in the STO/REGE plain. All organisms aggregated into E (everything else) are shown. Based on the “cv” aggregation (see Table 2). The plots are shown for 6 threshold values. Figure S9: The position of the not aggregated (A: #56; B: #52) or only weakly aggregated (C: #50 and #57; D: #45, #46, #47, #48 and #51) trophic groups in the STO/REGE plain. All organisms aggregated into E (everything else) are shown. Based on the “sv” aggregation (see Table 2). The plots are shown for 6 threshold values.

Author Contributions: Conceptualization, F.J.; Data curation, D.D.A.; Investigation, A.E.; Methodology, F.J. and W.-c.L.; Software, W.-c.L.; Supervision, F.J.; Visualization, D.D.A.; Writing—original draft, F.J.; Writing—review and editing, F.J., A.E., W.-c.L. and D.D.A.

Funding: The work of A.E. and F.J. was funded by National Research, Development and Innovation Office—NKFIH, grant number OTKA K 116071. The work of F.J. was funded by the grant GINOP-2.3.2-15-2016-00057. The work of D.D.A. was funded by the Flagship Project RITMARE—The Italian Research for the Sea—funded by the Italian Ministry of Education, University, and Research within the National Research Program 2011–2013.

Acknowledgments: We are grateful to the participants of the DynaTrait conference in Hannover (October 2017) for some initial discussions. Three anonymous reviewers are acknowledged for helpful and constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martinez, N.D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **1991**, *61*, 367–392. [[CrossRef](#)]
2. Thompson, R.M.; Brose, U.; Dunne, J.A.; Hall, R.O.; Hladysz, S.; Kitching, R.L.; Martinez, N.D.; Rantala, H.; Romanuk, T.N.; Stouffer, D.B.; et al. Food webs: Reconciling the structure and function of biodiversity. *Trends Ecol. Evol.* **2012**, *27*, 689–697. [[CrossRef](#)] [[PubMed](#)]
3. Power, M.E.; Tilman, D.; Estes, J.A.; Menge, B.A.; Bond, W.J.; Mills, L.S.; Daily, G.; Castilla, J.C.; Lubchenco, J.; Paine, R.T. Challenges in the quest for keystones. *BioScience* **1996**, *46*, 609–620. [[CrossRef](#)]
4. Lawton, J.H.; Brown, V.K. Redundancy in ecosystems. In *Biodiversity and Ecosystem Function*; Schulze, E.D., Mooney, H.A., Eds.; Springer: Berlin, Germany, 1994; pp. 255–270.
5. Cohen, J.E. *Food Webs and Niche Space*; Princeton University Press: Princeton, NJ, USA, 1978.
6. Luczkovich, J.J.; Borgatti, S.P.; Johnson, J.C.; Everett, M.G. Defining and measuring trophic role similarity in food webs using regular equivalence. *J. Theor. Biol.* **2003**, *220*, 303–321. [[CrossRef](#)] [[PubMed](#)]
7. Cirtwill, A.R.; Dalla Riva, G.V.; Gaiarsa, M.P.; Bimler, M.D.; Cagua, E.F.; Coux, C.; Dehling, D.M. A review of species role concepts in food webs. *Food Webs* **2018**. [[CrossRef](#)]
8. Hall, S.J.; Raffaelli, D. Food-web patterns: Lessons from a species-rich web. *J. Anim. Ecol.* **1991**, *60*, 823–841. [[CrossRef](#)]

9. Goldwasser, L.; Roughgarden, J. Construction and analysis of a large caribbean food web. *Ecology* **1993**, *74*, 1216–1233. [[CrossRef](#)]
10. Hirata, H.; Ulanowicz, R.E. Information theoretical analysis of the aggregation and hierarchical structure of ecological networks. *J. Theor. Biol.* **1985**, *116*, 321–341. [[CrossRef](#)]
11. Pahl-Wostl, C. The possible effects of aggregation on the quantitative interpretation of flow patterns in ecological networks. *Math. Biosci.* **1992**, *112*, 177–183. [[CrossRef](#)]
12. Solow, A.R.; Beet, A.R. On lumping species in food webs. *Ecology* **1998**, *79*, 2013–2018. [[CrossRef](#)]
13. Jordán, F.; Osváth, G. The sensitivity of food web topology to temporal data aggregation. *Ecol. Model.* **2009**, *220*, 3141–3146. [[CrossRef](#)]
14. Patonai, K.; Jordán, F. Aggregation of incomplete food web data may help to suggest sampling strategies. *Ecol. Model.* **2017**, *352*, 77–89. [[CrossRef](#)]
15. Scotti, M.; Hartvig, M.; Winemiller, K.O.; Li, Y.; Jauker, F.; Jordán, F.; Dormann, C.F. Trait-based and process-oriented modeling in ecological network dynamics. In *Adaptive Food Webs*; Moore, J.C., de Ruiter, P.C., McCann, K.S., Wolters, V., Eds.; Cambridge University Press: Cambridge, UK, 2017; pp. 228–256.
16. Sommer, U.; Charalampous, E.; Scotti, M.; Moustaka-Gouni, M. Big fish eat small fish: Implications for food chain length? *Comm. Ecol.* **2018**, *19*, 107–115. [[CrossRef](#)]
17. Bond, W.J. Keystone species. In *Biodiversity and Ecosystem Function*; Schulze, E.D., Mooney, H.A., Eds.; Springer: Berlin, Germany, 1994.
18. Daily, G.C.; Ehrlich, P.R.; Haddad, N.M. Double keystone bird in a keystone species complex. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 592–594. [[CrossRef](#)] [[PubMed](#)]
19. Ortiz, M.; Hermosillo-Núñez, B.; González, J.; Rodríguez-Zaragoza, F.; Gómez, I.; Jordán, F. Quantifying keystone species complexes: Ecosystem-based conservation management in the King George Island (Antarctic Peninsula). *Ecol. Indic.* **2017**, *81*, 453–460. [[CrossRef](#)]
20. d’Alcalà, M.R.; Conversano, F.; Corato, F.; Licandro, P.; Mangoni, O.; Marino, D.; Mazzocchi, M.G.; Modigh, M.; Montresor, M.; Nardella, M.; et al. Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): An attempt to discern recurrences and trends. *Sci. Mar.* **2004**, *68*, 65–83.
21. D’Alelio, D.; Libralato, S.; Wyatt, T.; Ribera d’Alcalà, M. Ecological-network models link diversity, structure and function in the plankton food-web. *Sci. Rep.* **2016**, *6*, 21806. [[CrossRef](#)]
22. D’Alelio, D.; Montresor, M.; Mazzocchi, M.G.; Margiotta, F.; Sarno, D.; Ribeira d’Alcalà, M. Plankton food-webs: To what extent can they be simplified? *Adv. Oceanogr. Limnol.* **2016**, *7*. [[CrossRef](#)]
23. Müller, C.B.; Adriaanse, I.C.T.; Belshaw, R.; Godfray, H.C.J. The structure of an aphid–parasitoid community. *J. Anim. Ecol.* **1999**, *68*, 346–370. [[CrossRef](#)]
24. Jordán, F.; Liu, W.-C.; van Veen, F.J.F. Quantifying the importance of species and their interactions in a host-parasitoid community. *Commun. Ecol.* **2003**, *4*, 79–88. [[CrossRef](#)]
25. Jordán, F.; Liu, W.-C.; Mike, Á. Trophic field overlap: A new approach to quantify keystone species. *Ecol. Model.* **2009**, *220*, 2899–2907. [[CrossRef](#)]
26. Lai, S.M.; Liu, W.C.; Jordán, F. A trophic overlap-based measure for species uniqueness in ecological networks. *Ecol. Model.* **2015**, *299*, 95–101. [[CrossRef](#)]
27. Lorrain, F.; White, H.C. Structural equivalence of individuals in social networks. *J. Math. Sociol.* **1971**, *1*, 49–80. [[CrossRef](#)]
28. Everett, M.G.; Borgatti, S. Role colouring a graph. *Math. Soc. Sci.* **1991**, *21*, 183–188. [[CrossRef](#)]
29. Borgatti, S.P.; Everett, M.G.; Freeman, L.C. *Ucinet for Windows: Software for Social Network Analysis*; Analytic Technologies: Harvard, MA, USA, 2002.
30. Yodzis, P.; Winemiller, K.O. In search of operational trophospecies in a tropical aquatic food web. *Oikos* **1999**, *87*, 327–340. [[CrossRef](#)]
31. Carradec, Q.; Coordinators, T.O.; Pelletier, E.; Da Silva, C.; Alberti, A.; Seeleuthner, Y.; Blanc-Mathieu, R.; Lima-Mendez, G.; Rocha, F.; Tirichine, L.; et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **2018**, *9*, 373. [[CrossRef](#)] [[PubMed](#)]
32. Franks, P.J.S. NPZ models of plankton dynamics: Their construction, coupling to physics, and application. *J. Oceanogr.* **2002**, *58*, 379–387. [[CrossRef](#)]

33. Mitra, A.; Castellani, C.; Gentleman, W.C.; Jonasdottir, S.H.; Flynn, K.J.; Bode, A.; Halsband, C.; Kuhn, P.; Licandro, P.; Agersted, M.D.; et al. Bridging the gap between marine biogeochemical and fisheries sciences; configuring the zooplankton link. *Progr. Oceanogr.* **2014**, *129*, 176–199. [[CrossRef](#)]
34. Conley, K.R.; Lombard, F.; Sutherland, K.S. Mammoth grazers on the ocean's minuteness: A review of selective feeding using mucous meshes. *Proc. Biol. Soc.* **2018**, *285*, 20180056. [[CrossRef](#)]
35. Flynn, K.J.; Stoecker, D.K.; Mitra, A.; Raven, J.A.; Glibert, P.M.; Hansen, P.J.; Granéli, E.; Burkholder, J.M. Misuse of the phytoplankton–zooplankton dichotomy: The need to assign organisms as mixotrophs within plankton functional types. *J. Plankton Res.* **2012**, *35*, 3–11. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).