Author for correspondence:
Benoit Gauzens
e-mail: benoit.gauzens@univ-rennes1.fr

THE ROYAL SOCIETY
PUBLISHING

Trophic groups and modules: two levels of group detection in food webs

Benoit Gauzens[1,2], Elisa Thébault[1], Gérard Lacroix[1,3] and Stéphane Legendre[4]

[1]UMR 7618-iEES Paris (CNRS, UPMC, UPEC, Paris Diderot, IRD, INRA), Université Pierre et Marie Curie, Bâtiment A, 7 quai St Bernard, 75252 Paris cedex 05, France
[2]UMR 6553 Ecobio, Université de Rennes 1, Avenue du Général Leclerc, Campus de Beaulieu, 35042 RENNES Cedex, France
[3]UMS 3194 (CNRS, ENS), CEREEP – Ecotron Ile De France, Ecole Normale Supérieure, 78 rue du Château, 77140 St-Pierre-lès-Nemours, France
[4]UMR 8197 IBENS (CNRS, ENS), École Normale Supérieure, 46, rue d'Ulm, 75230 Paris cedex 05, France

Within food webs, species can be partitioned into groups according to various criteria. Two notions have received particular attention: trophic groups (TGs), which have been used for decades in the ecological literature, and more recently, modules. The relationship between these two group concepts remains unknown in empirical food webs. While recent developments in network theory have led to efficient methods for detecting modules in food webs, the determination of TGs (groups of species that are functionally similar) is largely based on subjective expert knowledge. We develop a novel algorithm for TG detection. We apply this method to empirical food webs and show that aggregation into TGs allows for the simplification of food webs while preserving their information content. Furthermore, we reveal a two-level hierarchical structure where modules partition food webs into large bottom–top trophic pathways, whereas TGs further partition these pathways into groups of species with similar trophic connections. This provides new perspectives for the study of dynamical and functional consequences of food-web structure, bridging topological and dynamical analysis. TGs have a clear ecological meaning and are found to provide a trade-off between network complexity and information loss.

1. Introduction

In nature, species in communities are connected by their predation links, and these complex interactions can be represented by a network. The topology of these food webs is non-random and can have a considerable influence on their functionality [1,2], including their ability to persist. As for many complex networks [3], the notion of a group (a collection of nodes with specific characteristics) is a major topological feature of food webs [4–6], with important functional implications [7,8]. However, this notion of group covers a large set of definitions (trophic groups (TGs), modules, regular equivalence groups, structural role groups) and methods (modularity maximization, Markov chain clustering, statistical block modelling, spectral approaches), giving different insights on network structure (see [9,10] for reviews on these notions). In food-web ecology, groups have been identified mainly according to two distinct definitions: modules and TGs (figure 1a,b), but we still do not know how these two notions are related.

The notion of modularity (or community structure) refers to groups of nodes interacting more frequently between themselves than with other nodes. Modularity detection is challenging in view of its relation with network functionality [11]. For example, a modular structure can buffer the propagation of perturbations, determining the stability or resilience of ecological networks [8]. Mechanisms that give rise to modularity in food webs are not totally understood. Modules have been related to a variety of attributes, from niche organization of species and their diet [12] to phylogeny [13] or spatial segregation between species [14]. For example, in the food web of Chesapeake Bay, the split found between two
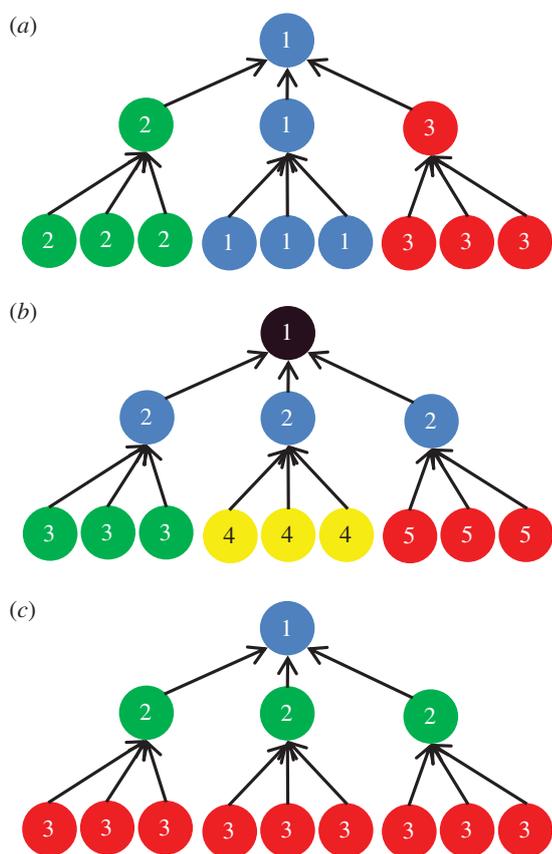
Figure 1. Representation of different group detection methods for a hypothetical food web: (a) modularity (three modules), (b) TGs method (five TGs) and (c) AP method (three AP groups). Nodes of the same colour and with the same numbers belong to the same group. This hypothetical food web has the topology of an N-levels tree where each non-basal species has exactly d prey. Different partitions of this example of food web ($N = 3$, $d = 3$, $S = 13$ species) are shown: three modules, five TGs and three AP groups. In the general case of a regular N-levels directed tree with in-degree d, the number of species is $S = 1 + d + \cdots + d^{N-1}$. The number of modules, TGs and AP groups are, respectively, d, $1 + 1 + d + \cdots + d^{N-2}$ and N. These numbers differ in general, with more TGs than modules or AP groups. We can observe here that AP groups correspond in this case to regular groups, based on the regular equivalence definition.

large modules corresponds closely to the division between pelagic and benthic species [15].

The study of food-web modularity is only recent, and historically, food webs have been mainly described in terms of TGs, in relation to the notion of trophic relationships introduced by Elton [16]. TGs are comprised of species that share similar sets of prey and predators. Aggregation into TGs has been used to simplify the representation of food webs, circumventing methodological difficulties induced by the complexity of trophic relationships in empirical data [4,17], and allowing the comparison of datasets and models of similar resolution [18]. In fact, food webs were for a long time described at the TG level rather than at the species level [19,20]. The simplification of food webs into TGs is also central to the study of ecosystem dynamical and functional properties [21].

Several methods have been developed to detect TGs in food webs, based on two different notions. First, a set of methods inherited from the notion of *structural equivalence* [22]. Two nodes in a graph are said to be structurally equivalent if they relate to the same group of nodes. This

assumption was then relaxed to allow nodes with similar but not identical relations to be said to be structurally equivalent. A classical method is to measure interaction similarity between nodes and then use a hierarchical clustering method (a stepwise classification process) to define structurally equivalent groups. In ecology, the Jaccard index has been used to define the amount of trophic overlap between taxa [23,24]. The main limit of the use of hierarchical clustering methods is that the number of groups does not appear as an emergent property, a threshold value for trophic similarity delimiting the groups or for the number of groups itself has to be preset.

A second method of detecting TGs in food webs is based on the notion of *regular equivalence*, inherited from the concept of its role in social sciences [25]. A group of regularly equivalent nodes contains species that are connected to the same set of groups containing regularly equivalent nodes. Regular equivalence was introduced not to detect groups of nodes with similar interaction patterns but to aggregate entities with the same role. Regular equivalence is classically illustrated with the example of interactions in a hospital: two nurses do not necessarily interact with the same people (they can have different patients or interact with different doctors) but they interact with similar types of people (patients, doctors, etc.). Thus, nurses have the same role in the hospital. The method of Luczkovich *et al.* [26] uses the notion of regular equivalence in ecology to group species, but the number of groups used for model selection has to be predefined and it potentially creates groups of species that do not share any trophic interactions. Block modelling approaches introduce an objective criterion for model selection. In their seminal paper, Allesina & Pascual [5] use AIC to select among models. In subsequent articles, Bayes factors [14,27] or normalized maximum likelihoods [28] were used. The main advantage of block modelling is the use of objective criteria for model selection, implying that the number of groups is not predefined. It however shares the same limit as all methods using the notion of regular equivalence by potentially aggregating nodes without any common connection (figure 1c).

We propose here a new method of TG detection based on structural equivalence in order to avoid the limits of regular equivalence (lumping species without any common prey or predator in the same group), but with the ability to determine the number of groups as an emergent property of the system.

We then use the different notions of groups used in ecology to understand whether food webs are better described when grouped according to TGs or to modules, and whether modules and TGs give opposite, similar or complementary descriptions of food-web topology. While modularity is gaining increasing interest in food-web studies [6,12,29], its relationship with TG arrangements is unknown as both network patterns have been studied independently. Detecting how different network decompositions are combined in food webs is important for understanding their structure and can reveal new network properties. It is also critical to assess the relevant and redundant features of network structure and to move beyond a disconnected view of food-web patterns. It has been shown that species aggregation into TGs did not affect the perception of food-web response to top-predator manipulation in an experiment [17]. Such result suggests that food webs might be mostly structured in TGs.

We therefore address here two different questions. First, we propose an efficient method to detect TGs in food webs. Second, using nine aquatic food webs of different resolutions, we compare these TGs to groups obtained by modularity detection [15] and groups obtained by the model of Allesina & Pascual [5], thereafter referred to as the AP model. The AP model is a block modelling approach that achieves the best compromise between the number of groups (network complexity) and information loss, using AIC for model selection. Depending on the structure of the considered network, the AP model will detect modules (i.e. groups of nodes interacting more frequently between themselves) or groups of regular equivalent species. The point is that classical methods for role detection create groups of regularly equivalent species (species in different groups are connected exactly to the same set of groups), whereas the AP method creates groups with group-specific connections to other groups. We show that TGs give a reliable picture of food webs in regard to information theory while preserving ecological significance, as we obtain close correspondences between the TG model and the AP model. This close matching does not hold when the methods are applied to two social networks, the Zachary's karate club [30] and the social prison inmate [31]. By comparing the trophic position of species in module and TG arrangements, we reveal a previously undetected link between TGs and modules: modules decompose the food web into disjoined vertical pathways of energy flow, and, within modules, TGs are composed of species of similar trophic levels.

2. Material and methods

2.1. A model for the detection of trophic groups

A TG is usually defined as a group of species that interact with similar prey and predators. We mathematically translate this definition using the notion of trophic similarity [23] and the conceptual framework of modularity detection [32]. The notion of trophic similarity is related to the notion of structural equivalence. It allows to avoid the drawback of regular equivalence where species without any common interactions can be grouped in the same TG. Using comparison to a random model, modularity detection allows to obtain the number of groups as an emergent property, which is not possible when hierarchical classification methods are used to detect groups of structural equivalence.

The modularity of a given partition E (a particular arrangement of the species in non-intersecting groups) in a network is given by the difference between the within-groups link density and its random expectation [33]:

$$M(E) = \sum_{s=1}^{|E|} \left(\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right), \tag{2.1}$$

where $|E|$ is the number of elements in the partition (the number of modules), l_s is the number of links between nodes in the s module, L is the total number of links of the food web, and d_s is the sum of degrees of species belonging to module s. The parameter l_s/L is the fraction of links inside module s (within-group link density), and $(d_s/2L)^2$ is an approximation of this expected quantity by chance alone.

For TG detection, we keep the comparison with a random null model, but instead of using the proportion of within-group links, our index is based on trophic similarity. The trophic similarity of two species is their number of common prey and predators divided by their total number of prey and predators. We transpose this definition using an analogy with the modularity index, by comparing the observed trophic similarity between all pairs of species in the same group to its expected value in a random graph. For a given partition E, our index is defined as

$$G(E) = \sum_{g=1}^{|E|} \frac{1}{|g|} \sum_{\substack{i,j \in G \\ i<j}} (T(i,j) - E(T(i,j))), \tag{2.2}$$

where $|g|$ is the number of nodes in group g, $|E|$ is the number of groups in the partition E. $T(i, j)$ (and its expected value in a random graph $E(T(i, j))$) is the ratio between the number of prey and predators interacting with species i and j, and the number of prey and predators interacting with species i or species j:

$$T(i,j) = \frac{|P_i \cap P_j| + |p_i \cap p_j|}{|P_i \cup P_j| + |p_i \cup p_j|}, \tag{2.3}$$

where P_i and p_i represent, respectively, the set of predators and prey of species i, $|P_i \cap P_j|$ is the cardinality of the intersection of P_i and P_j (i.e. the number of prey and predators common to species i and j). The value of $T(i, j)$ is directly obtained from the in- and out-degrees of species i and j in the food web. The computation of $E(T(i, j))$ is described in the electronic supplementary material, S1.

Group detection is performed by maximizing the TG index $G(E)$ using a simulated annealing algorithm for each of the considered networks (table 3). The N_W computer program was used to perform the computations [34].

2.2. Networks studied

Analyses were made on a dataset of nine food webs and two social networks. Food webs were chosen for their low level of aggregation (i.e. most trophic interactions are described at the species and genus level and not at the level of large TGs). The nine food webs are Benguela [35], Bridge Brooke Lake [36], Carribean reef [37], Chesapeake Bay [38], Créteil Lake (electronic supplementary material, S3), Tuesday Lake [39], Carpinteria [40], DempsterSu [41] and Ythan estuary [42]. The two social networks, the prison inmate [31] and Zachary's karate club [30] graphs, are classical examples in social science studies. They were used to assess whether the specific results we found for food webs were also relevant for other kinds of networks. A specific focus was put on the Lake Créteil food web to investigate the characteristics of the TGs found by our method. The Lake Créteil food web was created on the basis of a summer mesocosm study [4] conducted by G. Lacroix and co-workers; we thus have a good knowledge of the ecology of this food web.

2.3. Comparison between group arrangements of the different detection methods

In order to assess whether food webs are better described when grouped according to TGs or to modules, we compare the TGs obtained with our method and the modules to the groups obtained by the AP model. For both modularity and the AP model, we used a simulated annealing algorithm to detect groups in the considered food webs. To assess the correspondence between the different group detection methods, we used a mutual information criteria [33]. The normalized mutual information I_{EF} between two partitions is defined as the ratio between the mutual information of the partitions and the mean of their respective entropy [43]

$$I_{EF} = \frac{-\sum_{i=1}^{|E|} \sum_{j=1}^{|F|} n_{ij}^{EF} \log (n_{ij}^{EF} S / N_i^E N_j^F)}{\frac{1}{2} \times \left(\sum_{i=1}^{|E|} N_i^E \log (N_i^E / S) + \sum_{j=1}^{|F|} N_j^F \log (N_j^F / S) \right)}. \tag{2.4}$$

Here, S is the number of species, $|E|$ and $|F|$ are the number of groups in partitions E and F, respectively, N_i^E and N_j^F are the number of nodes in group i of partition E and group j of partition F. Finally, n_{ij}^{EF} is the number of nodes that are both in group i of

partition E and in group j of partition F. The mutual information between partitions E and F is equal to 1 if both partitions are identical, and 0 if there is no matching.

2.4. Relationships between trophic groups and modules

We investigated the links between TGs and modules in three ways: first by comparing the distribution of species' trophic level between these two types of groups, second by measuring whether TGs were embedded in modules and third by characterizing the contribution to modularity of species belonging to TGs that were split across different modules.

2.4.1. Distribution of species' trophic level in trophic groups and modules

The trophic level of a species is defined as 1 plus the mean trophic level of its prey, with the trophic level of basal species set to 0. For all food webs, we calculated the variance in species trophic level either within modules or within TGs. To test whether variance of species trophic levels within modules differed from random expectation, we used a null model approach. This null model distributes species randomly in different modules, while keeping the number of modules and their respective sizes as in the original network (100 000 replications, p-value is the probability to obtain a higher variance of trophic levels within the food web modules than expected from the null model). To test whether variance of species trophic levels within TGs differed from random expectation, we used the same null model as described above, but with a random attribution of species to TGs instead of modules (100 000 replications, in this case the p-value is the probability to obtain a lower variance of trophic levels within the TGs than expected from the null model).

2.4.2. Module diversity of trophic groups

To assess whether species affiliated to the same TG also belong to the same module, we measured an index of module diversity for TGs

$$D_g = 1 - \sum_{s=1}^{m} \left(\frac{g_s}{|g|}\right)^2, \tag{2.5}$$

where g_s is the number of species of group g that belong to module s (i.e. the cardinality of the intersection of g and s), and $|g|$ is the number of species in g (the underlying partition is implicit in this notation). D_g is 0 if all species of a TG belong to the same module and is $1 - 1/|g|$ when all species in the group belong to different modules. These values are compared to a null model where the partition into TGs is identical to that obtained with our model, but where species are randomly distributed among modules while keeping the same number of modules and their respective sizes as in the original food web. Comparisons are made with 100 000 values of diversity obtained with the null model.

2.4.3. Participation coefficient of species to modules

We observed that each TG was in general embedded into a single module. We tested whether species of TGs that were split across different modules occupied a particular position within the modular structure. In order to determine the species' contribution to network modularity, we computed the participation coefficient [44]. Based on the Simpson diversity index, the participation coefficient measures the diversity of connections of species i to the different modules of the network:

$$PC_i = 1 - \sum_{S=1}^{m} \left(\frac{l_{is}}{d_i}\right)^2. \tag{2.6}$$

Here, m is the number of modules, l_{is} is the number of links between species i and the species of module s, and d_i is

Table 1. The number of pairs of species belonging to the same groups but without any common interactions is non-zero for the AP method and almost zero for the TG method.

networks	AP	TGs
Creteil	0	0
DempsterSU	73	0
Tuesday Lake	11	0
Cheasapeake Bay	62	0
Ythan Estuary	62	0
Bridge Brook Lake	7	0
Caribbean reef	27	1
Carpinteria	39	1
Tuesday Lake	11	0

the degree (number of prey and predators) of species i. PC_iP_i equals 0 when all links of i are in its own module and is $1 - 1/m$ when links are uniformly distributed among modules. Student's t-tests are then used to compare indices found for species in TGs belonging to different modules and species in TGs belonging to only one module.

3. Results

3.1. The different aggregation methods are expected to return different groups

This is shown using a simple network, a directed tree in which all species except the basal species have the same number of prey (figure 1). We can notice in figure 1 a major difference between AP groups and TGs: in the case of AP groups, all basal species are lumped together while this is not the case for TGs. With AP groups, species can belong to the same group even if they do not share any common predators (table 1). In this particular topology, AP groups are equivalent to groups found using a regular equivalence method [26].

3.2. Example of functional divisions in the food web of Lake Créteil

In the food web of Lake Créteil, the TG method identifies 13 TGs. They tend to discriminate species according to trophic level (either phytoplankton, zooplankton, carnivorous or omnivorous) as well as body size, taxonomy and habitat (table 2). Note that it is difficult to assess the relevance of the group comprised of the trophospecies 'Bacteria', 'DOM and POM' (dissolved and particulate organic matter) and 'Biofilm', as the ecological role of these constituents can be different. This part of the network, which groups together detrital and littoral components of the food web, is not well known. Considering more precisely bacterial diversity and biofilm composition could lead to a different result.

Using module detection [33], we observe that most species within a TG belong to the same module (i.e. TGs are a subpartition of modules; figure 2 and table 3). Thus, within a module, TGs interact mostly between themselves. Moreover, we can appreciate in figure 2b that modules assemble TGs along energetic pathways in the food web (vertical

Table 2. Groups obtained by our TG detection method in relation to group characteristics for the Lake Créteil food web. These groups are represented by the corresponding colours in figure 2.

TGs	group characteristics
Abramis brama, Rutilus rutilus, Acanthocyclops robustus	omnivorous fish and large cyclopoids (blue-green)
Asplanchna girodi, Asplanchna priodonta, Thermocyclops crassus, Thermocyclops oithonoides	carnivorous Rotifers and small cyclopoids (white)
Eudiaptomus gracilis, Eutytemora velox	omnivorous calanoids (green)
Cephallodella sp., *Chydorus sphaericus, Lecane bulla, Lecane luna, Lecane stichaea, Lepadella* sp., *Testidunella patina*, Chironomidae	benthic or littoral species and detritivorous or bactivorous organisms (brown)
Hexarthra mira, Filinia longiseta	rotifers consuming small algal cells and bacteria (pink)
bdelloid species, *Bosmina coregoni, Bosmina longirostris, Brachionus angularis, Brachionus calyciflorus, Brachionus quadridentatus, Keratella cochlearis, Keratella quadrata*, nauplii of Calanoida, nauplii of Cyclopidae, *Polyarthra dolichoptera-vulgaris, Polyarthra major, Pompholyx sulcata, Trichocerca* sp.	small herbivorous zooplankton (dark green)
Ceriodaphnia dubia, Ceriodaphnia pulchella, Daphnia cucullata, Daphnia galeata, Daphnia galeata × *D. cucullata, Diaphanosoma brachyurum, Synchaeta pectinata*	large herbivorous Cladocera (purple)
DOM and POM, Bacteria, biofilm	components of the detrital and littoral pathway (orange)
Ceratium hirundinella, Nitzschia sp., *Pediastrum boryanum, Synedra ulna*	large or protected, poorly edible, algae (light purple)
Dictyosphaerium pulchellum, Navicula sp., *Pediastrum duplex, Schroederia indica, Staurastrum* sp., *Trachelomonas* sp.	algae mainly consumed by graspers within zooplankton (light blue)
Coelastrum spp., *Colacium* sp., *Cosmarium* sp., *Cryptomonas* sp., *Desmodesmus quadricauda, Oocystis lacustris, Scenedesmus acuminatus*	edible algae consumed by herbivorous and omnivorous zooplankton (dark blue)
Chroomonas sp., *Crucigenia* spp., *Cyclotella ocellata, Monoraphidium contortum, Tetraedron minimum*	edible algae consumed by herbivorous zooplankton (yellow)
Quadricoccus ellipticus, small undetermined unicells	small phytoplanktonic species, highly edible algae for filter feeders (red)

Table 3. Number of groups obtained using TG, modularity (M) and the Allesina & Pascual (AP) detection methods, with the degree of overlap between the different partitions. P sets the *p*-value of the difference of participation coefficients between species in TGs belonging to different modules and species in TGs belonging to only one module. D is the *p*-value of the difference in diversity of modules for TGs compared with a null model. The asterisk (*) symbol corresponds to food webs for which all TGs are in a single module. Hence, statistical analyses on P were not relevant in this case.

	species (links)	TG	AP	M	TG-AP overlap	module-AP overlap	P	D
Benguala [35]	29 (203)	7	7	3	0.841	0.397	0.0459	$<10^{-4}$
Bridge Brooke Lake [36]	75 (553)	12	9	3	0.92	0.631	*	$<10^{-4}$
Carribean reef [37]	249 (3313)	46	28	3	0.775	0.365	$<10^{-4}$	$<10^{-4}$
Chesapeake Bay [38]	33 (72)	13	7	3	0.745	0.428	0.4793	$<10^{-4}$
Créteil Lake SI3	67 (718)	13	12	3	0.922	0.4738	0.0194	$<10^{-4}$
Tuesday Lake [45]	73 (410)	17	11	2	0.834	0.449	*	$<10^{-4}$
Carpinteria [40]	128 (2290)	37	28	3	0.872	0.379	0.289	$<10^{-4}$
DempsterSu [41]	107 (966)	25	12	3	0.7129	0.410	$<10^{-4}$	$<10^{-4}$
Ythan Estuary [42]	92 (409)	26	13	3	0.755	0.317	$<10^{-4}$	$<10^{-4}$

component). The first module (left part of figure 2*b*) brings together food chains involving small herbivorous zooplankton and calanoids. The second module (middle part of figure 2*b*) brings together food chains involving large filter feeders (Cladocera). These two modules are mainly pelagic and separate energetic pathways according to body size and behaviour of
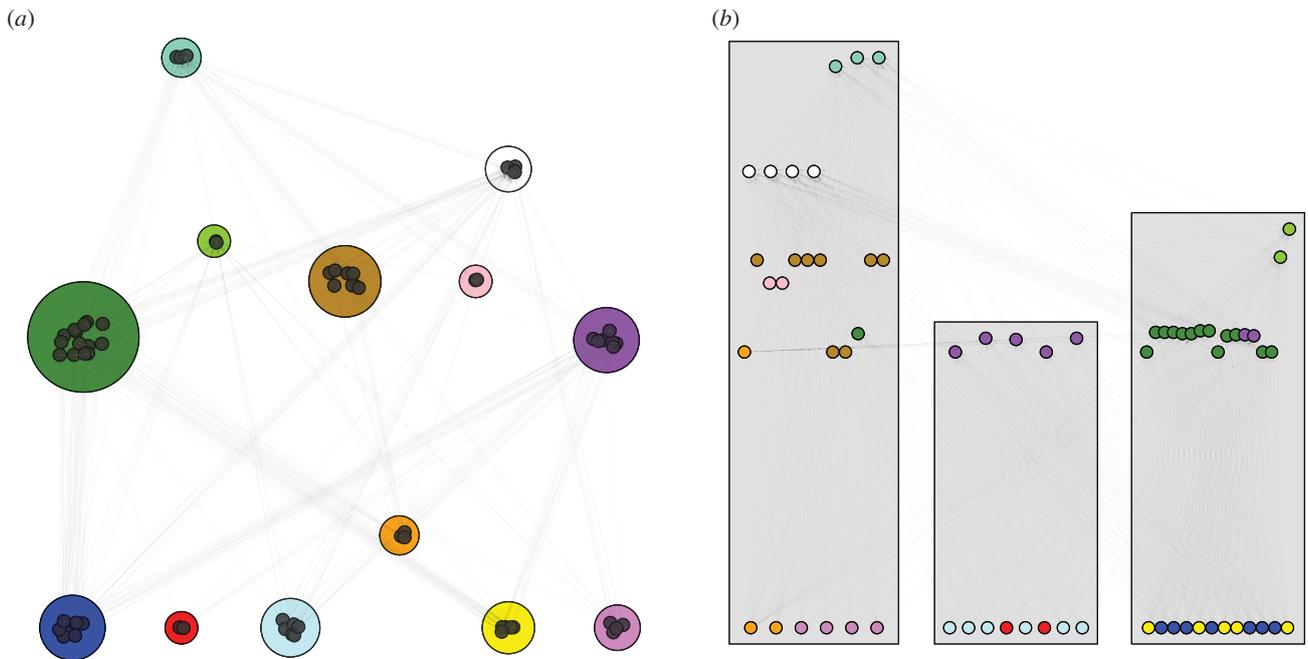
Figure 2. Representation of the Lake Créteil food web partitioned with the TG method (a,b), and module detection (b). (a) TGs are delimited by coloured discs whose sizes are proportional to the number of species in each TG, and species are represented by small grey circles. (b) Modules are delimited by grey rectangles, and species are represented by small circles whose colour corresponds to their TG in (a). The vertical dimension corresponds to the species' trophic levels (b) and the average trophic level of TGs (a). The compositions and characteristics of the TGs for the Lake Créteil food web are described in table 2.

herbivores (small versus large graspers and filter feeders). The third module (right part of figure 2b) brings together trophic pathways dominated by organisms that are mainly omnivorous and are able to feed on littoral and benthic organisms. Hence, in the food web of Lake Créteil, modules appear as assemblages of trophic chains that link TGs with common major characteristics (size, behaviour, edibility, spatial niche).

3.3. Comparison between group arrangements of the different detection methods

In the nine empirical food webs considered, TG always leads to partitions with a higher number of groups than modularity (table 3). Indeed, modularity leads to partitions with a very low number of modules, suggesting that the number of independent subnets is limited (table 3). The number of groups obtained with the AP method is always higher than with modularity and lower than with TG (with one exception for the Benguela food web; table 3).

Correspondence indices between groups obtained by TG and AP are significantly higher than correspondence indices between modularity and AP (paired Student's t-test, $p < 0.001$). The high degree of overlap between TG and AP (table 3, correspondence close to 1) suggests that an important part of the information carried by food-web structure can be attributed to TGs. Strikingly, and despite totally different goals, the AP method (looking for the most informative partitions) and the TG method lead to similar results (table 3) even if the AP method still groups species without any common interaction, whereas the TG method does not (table 1). This close match between the two methods seems to be specific to food webs. Indeed, when comparisons are made on the two social networks, the Zachary's karate club and the prison inmate, correspondence indices are much lower with values of 0.531 and 0.478, respectively.

3.4. Relationships between trophic groups and modules
3.4.1. Distribution of species' trophic level in trophic groups and modules

Food-web representations combining trophic levels of species and their affiliation to modules and TGs (figures 2 and 3; electronic supplementary material, S2) suggest that, whereas species in the same TG tend to occupy the same trophic level, species in the same module often belong to different trophic levels. We computed the variance of species' trophic levels within either modules or TGs. In all the food webs studied, the average variance of species' trophic levels in modules was always higher than in TGs ($p < 10^{-4}$ for all networks). Furthermore, the variance of trophic levels of species belonging to the same module was higher than what was expected by chance alone ($p < 10^{-4}$ for all food webs). The opposite pattern was found when considering the variance of trophic levels of species sharing the same TGs ($p < 10^{-4}$ for all networks). By definition, species in a module are highly connected. As most trophic relations occur between species of different trophic levels, this could explain why species in the same module tend to belong to different trophic levels. Therefore, modules reflect particular energetic pathways, representing parallel trophic chains.

3.4.2. Modules' diversity of trophic groups and participation coefficient of species to modules

We observe that species in a TG tend to belong to a same module (figures 2 and 3; electronic supplementary material, S2). Thus, TGs tend to be embedded in modules. For all food webs, the average module diversity D_g of TGs was close to 0 and belonged to the 5% lowest values generated from the null model. This highlights a hierarchical two-level structure of food webs, where a partition into modules is further partitioned into TGs.
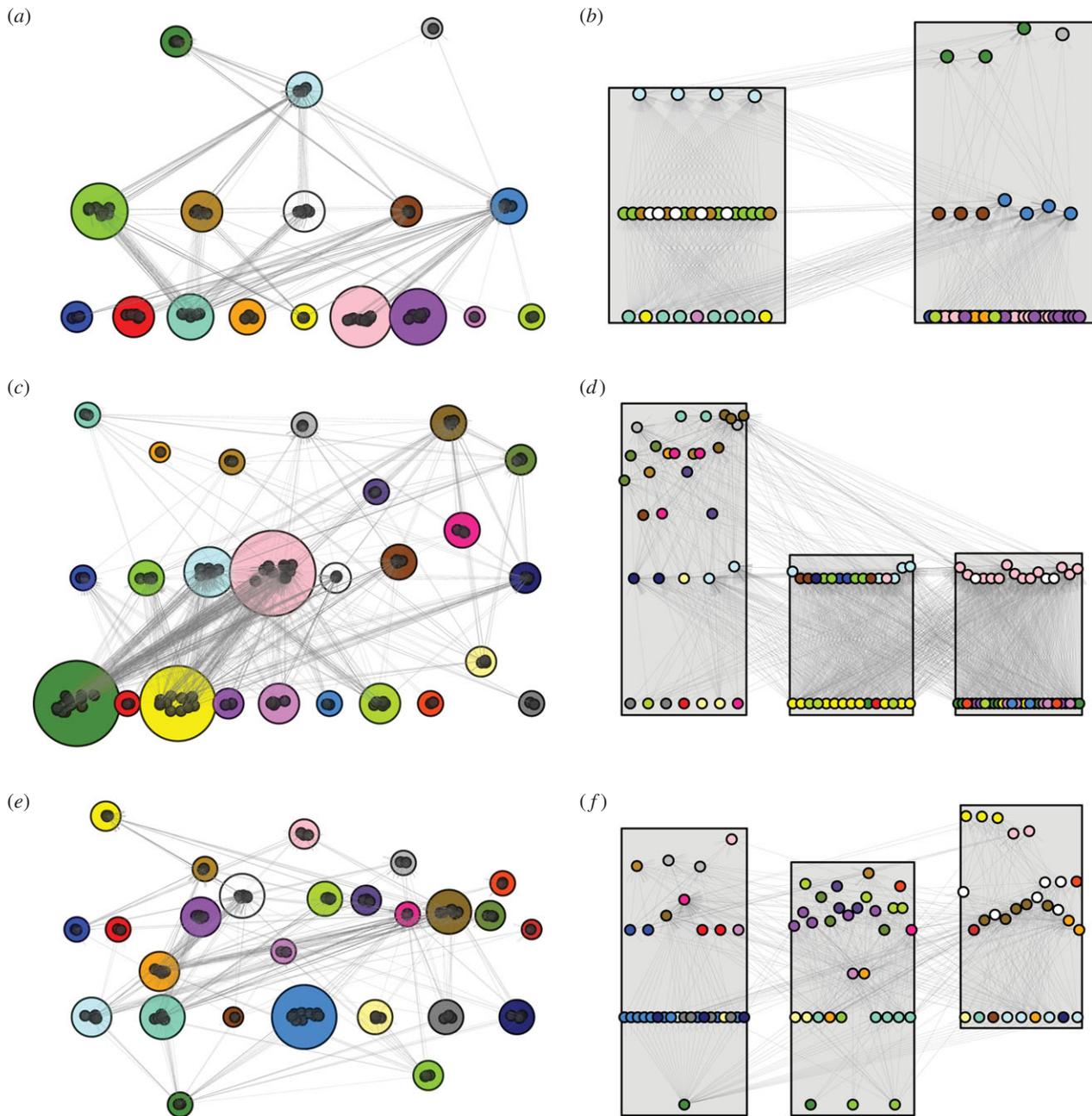
Figure 3. Representation of the Tuesday Lake (*a,b*), DempsterSu (*c,d*) and Ythan Estuary (*e,f*) food webs, with species sorted according to their TGs (*a,c,e*) and their modules (*b,d,f*). Same conventions as in figure 2.

Although striking, this arrangement of TGs into modules is not perfect. Species of a given TG are in some instances dispatched in different modules. The mean participation coefficients to modules of species in TGs dispatched in different modules are in most cases significantly lower than species in groups that belong to a single module. Indeed, in most food webs, the species of TGs that are split in several modules are those that contribute the least to the modular structure of the food webs (table 3).

4. Discussion

Thanks to the development of a new algorithm to identify TGs in food webs, our study reveals two important features of the structure of empirical food webs. First, we show that lumping species according to TGs allows the simplification of food webs while preserving the information carried by the initial network structure. Second, by considering TGs

and modules together, we put forward a previously unnoticed pattern of organization of food webs: modules are composed of species from different trophic levels and are further partitioned into TGs; they represent energetic pathways linking TGs from the bottom to the top of the food web.

4.1. An algorithm to identify trophic groups

Whereas the concept of TGs is widely used in the ecological literature since Elton [16] and Lindeman [46], the characterization of TGs is usually based on (subjective) expert knowledge. In the existing methods of food-web aggregation into TGs [23,24], the number of TGs is defined by the user and is not an emergent property of the network. Using the methodology developed for modularity indices, our method of TG detection circumvents previous limitations [5,47] where the ecological meaning of the partitions returned does not come from the method itself. By contrast, our method is based on the

ecological notion of trophic similarity, and by extension on the notion of nodes with similar patterns of connections.

4.2. Trophic groups: main underlying structure of food webs?

The TG method and the AP method detect groups according to totally different criteria. The AP method aims to find partitions corresponding to the best trade-off between information loss and reduction of complexity using the AIC, without any notion of ecology. The TG method finds clusters of species with similar sets of prey and predators. The match found between the partitions returned by the two methods shows that TGs support a large part of the information carried by the underlying structure of the food web, as given by the AP method. The relevance of species' aggregation into TGs has already been suggested to reflect functional properties [4,17,21,48] or to identify structural patterns [14]. We highlight here that food-web decomposition into TGs aggregates species with minimal loss of information while keeping a clear ecological meaning, and with the potential to reflect the functioning of the network. The relevance of such aggregation criteria (groups of nodes interacting with similar groups of nodes) seems very general for food webs. On the other hand, the aggregation process did not prevent information loss when it was applied to the two social networks. An intuitive explanation might be that species with similar prey and predators do not predate on each other while in social networks, actors with similar relationships tend to know each other and are often not precluded from interacting.

4.3. Trophic groups and modules: complementary views of food-web structure

Though we show that the notion of TG prevails in food webs, our study also confirms that modules are an important feature. Previous studies have already shown that food webs are more modular than random networks [12]. This suggests patterns of organization similar to those observed in other biological networks (gene–protein, plant–pollinator, neuronal) and in some small-world networks [49]. While modular patterns still need to be explained in food webs, we observe that modules represent parallel pathways of energy from producers to consumers, delimiting distinct food chains (figures 2 and 3; electronic supplementary material, S2). This is in accordance with previous results [12] showing that species in the same module (according to the notion of directed modularity) are globally located on trophic chains coming from similar basal species. We reveal that the variance of species trophic levels within modules is higher than expected by chance. The opposite result is found when groups are determined only accordingly to prey or predator similarity [12].

Despite having intuitively nearly opposite definitions (modules represent groups of species interacting mostly with one another, whereas TGs correspond to groups of species interacting with other well-defined groups of species), modules and TGs are linked and provide complementary pictures of food-web structure. It appears that food webs present a two-level hierarchical structure, with each TG belonging globally to a single module. The existence of network hierarchical structure has already been described for social networks [50]. Some TGs are, however, sometimes split across several modules. Species of such TGs share the same neighbourhood, as they are in the same TG, but belong to different communities (modules). These species are connected more diversely to modules than other species, therefore, they potentially bridge different modules. As the modular structure limits the propagation of perturbations [8], species bridging different modules could play a key role by interconnecting distinct subnets of energetic pathways, and allowing different ecological processes (perturbations, trophic cascades, etc.) to shift from one module to another.

4.4. Implications for future research

The functional implications of modularity are currently widely explored [8,51], but little is known about the functional implications of the TG structure. Indeed, while modules are characterized by a high density of within links, the implications of the architecture defined by TGs (few links within TGs and a large number of links between some TGs) have not been addressed. TGs are often used as a simplification, making the system more readable, sometimes as a consequence of external constraints (spatial segregation [14]), but the functional implications of TG patterns are worth exploring. For example, we still do not know how the dynamics of TGs is related to the individual dynamics of their component species.

Species richness within TGs could be considered as functional redundancy. The deletion of a whole group might lead to the loss of an entire set of specific connections, which could potentially have dramatic effects on system properties. As many topological studies [52–55] focus on the detection of key species in networks, the determination of the aggregated network of TGs addresses the question in a new way by considering potential key species as elements of TGs characterized by a low diversity.

As food-web descriptions are becoming more and more precise—recent published food webs contain several thousand links the reduction in complexity will become a critical issue. Our approach has the advantage of delineating TGs in such a way that complexity is reduced while keeping a clear ecological meaning. However, we need to know the entire network to simplify it. The next step will be to consider the correspondences between the biological traits of species within and between TGs, in order to develop methods able to reconstruct TGs and their links using species attributes. Addressing this question may improve our comprehension of the parameters involved in the trophic niche space (set of ecological parameters determining the trophic relationships of species). Several parameters, such as size [56], phylogenetic relationships [13,27] or behaviour [57] have been already considered. Even if they are limited to trophic relationships, these studies might provide a useful tool for the generic classification of species.

Improving our comprehension of network simplification is essential to address the structure–function relationship in food webs. As modelling approaches cannot encompass the entire complexity of food webs, food-web simplification via TG detection provides a trade-off between consistency and mathematical tractability, relating structural properties and functional issues.

References

1. May RM. 1972 *Stability and complexity in model ecosystems*. Princeton, NJ: Princeton University Press.

2. Rooney N, McCann KS. 2012 Integrating food web diversity, structure and stability. *Trends Ecol. Evol.* **27**, 40–46. (doi:10.1016/j.tree.2011.09.001)

3. Palla G, Derényi I, Farkas I, Vicsek T. 2005 Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818. (doi:10.1038/nature03607)

4. Hulot FD, Lacroix G, Loreau M. 2014 Differential responses of size-based functional groups to bottom–up and top–down perturbations in pelagic food webs: a meta-analysis. *Oikos* **123**, 1291–1300. (doi:10.1111/oik.01116)

5. Allesina S, Pascual M. 2009 Food web models: a plea for groups. *Ecol. Lett.* **12**, 652–662. (doi:10.1111/j.1461-0248.2009.01321.x)

6. Tinker TM, Guimarães PR, Novak M, Marquitti FMD, Bodkin JL, Staedler M, Bentall G, Estes JA. 2012 Structure and mechanism of diet specialisation: testing models of individual variation in resource use with sea otters. *Ecol. Lett.* **15**, 475–483. (doi:10.1111/j.1461-0248.2012.01760.x)

7. Thébault E, Fontaine C. 2010 Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**, 853–856. (doi:10.1126/science.1188321)

8. Stouffer DB, Bascompte J. 2011 Compartmentalization increases food-web persistence. *Proc. Natl Acad. Sci. USA* **108**, 3648–3652. (doi:10.1073/pnas.1014353108)

9. Fortunato S. 2010 Community detection in graphs. *Phys. Rep.* **486**, 75–174. (doi:10.1016/j.physrep.2009.11.002)

10. Leger J-B, Vacher C, Daudin J-J. 2013 Detection of structurally homogeneous subsets in graphs. *Stat. Comput.* **24**, 675–692. (doi:10.1007/s11222-013-9395-3)

11. Watts DJ, Strogatz SH. 1998 Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)

12. Guimerà R, Stouffer DB, Sales-Pardo M, Leicht EA, Newman MEJ, Nunes Amaral L. 2010 Origin of compartmentalization in food webs. *Ecology* **91**, 2941–2951. (doi:10.1890/09-1175.1)

13. Rezende EL, Albert EM, Fortuna MA, Bascompte J. 2009 Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecol. Lett.* **12**, 779–788. (doi:10.1111/j.1461-0248.2009.01327.x)

14. Baskerville EB, Dobson AP, Bedford T, Allesina S, Anderson TM, Pascual M. 2011 Spatial guilds in the Serengeti food web revealed by a Bayesian group model. *PLoS Comput. Biol.* **7**, e1002321. (doi:10.1371/journal.pcbi.1002321)

15. Girvan M, Newman MEJ. 2002 Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826. (doi:10.1073/pnas.122653799)

16. Elton C. 1927 *Animal ecology*. Chicago, IL: University of Chicago Press.

17. Gauzens B, Legendre S, Lazzaro X, Lacroix G. 2013 Food-web aggregation, methodological and functional issues. *Oikos* **122**, 1606–1615. (doi:10.1111/j.1600-0706.2013.00266.x)

18. Dunne JA et al. 2013 Parasites affect food web structure primarily through increased diversity and complexity. *PLoS Biol.* **11**, e1001579. (doi:10.1371/journal.pbio.1001579)

19. Martinez N. 1993 Effect of scale on food web structure. *Science* **260**, 242–243. (doi:10.1126/science.260.5105.242)

20. Sugihara G, Bersier LF, Schoenly K. 1997 Effects of taxonomic and trophic aggregation on food web properties. *Oecologia* **112**, 272–284. (doi:10.1007/s004420050310)

21. Boit A, Martinez ND, Williams RJ, Gaedke U. 2012 Mechanistic theory and modelling of complex food-web dynamics in Lake Constance. *Ecol. Lett.* **15**, 594–602. (doi:10.1111/j.1461-0248.2012.01777.x)

22. Lorrain F, White HC. 1971 Structural equivalence of individuals in social networks. *J. Math. Sociol.* **1**, 49–80. (doi:10.1080/0022250X.1971.9989788)

23. Martinez ND. 1991 Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392. (doi:10.2307/2937047)

24. Yodzis P, Winemiller KO. 1999 In search of operational trophospecies in a tropical aquatic food web. *Oikos* **87**, 327–340. (doi:10.2307/3546748)

25. White DR, Reitz KP. 1983 Graph and semigroup homomorphisms. *Soc. Networks* **5**, 193–234. (doi:10.1016/0378-8733(83)90025-4)

26. Luczkovich J, Borgatti SP, Johnson JC, Everett MG. 2003 Defining and measuring trophic role similarity in food webs using regular equivalence. *J. Theor. Biol.* **220**, 303–321. (doi:10.1006/jtbi.2003.3147)

27. Eklöf A, Helmus MR, Moore M, Allesina S. 2011 Relevance of evolutionary history for food web structure. *Proc. R. Soc. B* **279**, 1588–1596. (doi:10.1098/rspb.2011.2149)

28. Staniczenko PPA, Lewis OT, Jones NS, Reed-Tsochas F. 2010 Structural dynamics and robustness of food webs. *Ecol. Lett.* **13**, 891–899. (doi:10.1111/j.1461-0248.2010.01485.x)

29. Olesen JM, Bascompte J, Dupont YL, Jordano P. 2007 The modularity of pollination networks. *Proc. Natl Acad. Sci. USA* **104**, 19 891–19 896. (doi:10.1073/pnas.0706375104)

30. Zachary W. 1977 An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **33**, 452–473.

31. Wright B, Evitts M. 1961 Direct factor analysis in sociometry. *Sociometry* **24**, 82–98. (doi:10.2307/2785931)

32. Newman M, Girvan M. 2004 Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113. (doi:10.1103/PhysRevE.69.026113)

33. Guimerà R, Sales-Pardo M, Amaral L. 2007 Module identification in bipartite and directed networks. *Phys. Rev. E* **76**, 036102. (doi:10.1103/PhysRevE.76.036102)

34. Legendre S, Gauzens B, Lacroix G. 2013 The N_W program: trophic networks analysis. See http://www.biologie.ens.fr/~legendre/n_w/n_w.html.

35. Yodzis P. 1998 Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *J. Anim. Ecol.* **67**, 635–658. (doi:10.1046/j.1365-2656.1998.00224.x)

36. Havens K. 1992 Scale and structure in natural food webs. *Science* **257**, 1107–1109. (doi:10.1126/science.257.5073.1107)

37. Opitz S. 1996 Trophic interactions in Caribbean coral reefs. ICLARM Tech. Rep., vol. 43. Manila, Philippines: ICLARM.

38. Baird D, Ulanowicz RE. 1989 The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* **59**, 329–364. (doi:10.2307/1943071)

39. Jonsson T, Cohen JE, Carpenter SR. 2005 Food webs, body size, and species abundance in ecological community description. *Adv. Ecol. Res.* **36**, 1–84. (doi:10.1016/S0065-2504(05)36001-6)

40. Lafferty KD, Dobson AP, Kuris AM. 2006 Parasites dominate food web links. *Proc. Natl Acad. Sci. USA* **103**, 11 211–11 216. (doi:10.1073/pnas.0604755103)

41. Thompson RM, Townsend CR. 2000 Is resolution the solution?: the effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshw. Biol.* **44**, 413–422. (doi:10.1046/j.1365-2427.2000.00579.x)

42. Hall SJ, Raffaelli D. 1991 Food-web patterns: lessons from a species-rich web. *J. Anim. Ecol.* **60**, 823–842. (doi:10.2307/5416)

43. Fred ALN, Jain AK. 2003 Robust data clustering. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2**, II-128–II-133. (doi:10.1109/CVPR.2003.1211462)

44. Guimerà R, Nunes Amaral LA. 2005 Functional cartography of complex metabolic networks. *Nature* **433**, 895–900. (doi:10.1038/nature03288)

45. Jonsson T, Cohen J, Carpenter S. 2005 Food webs, body size, and species abundance in ecological community description. *Adv. Ecol. Res.* **36**, 1–84. (doi:10.1016/S0065-2504(05)36001-6)

46. Lindeman RL. 1942 The trophic-dynamic aspect of ecology. *Ecology* **23**, 399–417. (doi:10.2307/1930126)

47. Newman MEJ, Leicht EA. 2007 Mixture models and exploratory analysis in networks. *Proc. Natl Acad.*

Sci. USA **104**, 9564–9569. (doi:10.1073/pnas.0610537104)

48. Madigan DJ, Carlisle AB, Dewar H, Snodgrass OE, Litvin SY, Micheli F, Block BA. 2012 Stable isotope analysis challenges wasp-waist food web assumptions in an upwelling pelagic ecosystem. *Sci. Rep.* **2**, 654. (doi:10.1038/srep00654)

49. Pan R, Sinha S. 2009 Modularity produces small-world networks with dynamical time-scale separation. *Europhys. Lett.* **85**, 68006. (doi:10.1209/0295-5075/85/68006)

50. Lancichinetti A, Fortunato S, Kertész J. 2009 Detecting the overlapping and hierarchical community structure in complex networks. *New*

J. Phys. **11**, 033015. (doi:10.1088/1367-2630/11/3/033015)

51. Variano EA, Lipson H. 2004 Networks, dynamics, and modularity. *Phys. Rev. Lett.* **92**, 1–4. (doi:10.1103/PhysRevLett.92.188701)

52. Allesina S, Bodini A. 2004 Who dominates whom in the ecosystem? Energy flow bottlenecks and cascading extinctions. *J. Theor. Biol.* **230**, 351–358. (doi:10.1016/j.jtbi.2004.05.009)

53. Jordán F, Okey TA, Bauer B, Libralato S. 2008 Identifying important species: linking structure and function in ecological networks. *Ecol. Model.* **216**, 75–80. (doi:10.1016/j.ecolmodel.2008.04.009)

54. Allesina S, Pascual M. 2009 Googling food webs: can an eigenvector measure species' importance for coextinctions? *PLoS Comput. Biol.* **5**, e1000494. (doi:10.1371/journal.pcbi.1000494)

55. Jordán F, Osváth G. 2009 The sensitivity of food web topology to temporal data aggregation. *Ecol. Model.* **220**, 3141–3146. (doi:10.1016/j.ecolmodel.2009.05.002)

56. Woodward G, Speirs D, Hildrew A. 2005 Quantification and resolution of a complex, size-structured food web. *Adv. Ecol. Res.* **36**, 85–135. (doi:10.1016/S0065-2504(05)36002-8)

57. Petchey OL, Beckerman AP, Riede JO, Warren PH. 2008 Size, foraging, and food web structure. *Proc. Natl Acad. Sci. USA* **105**, 4191–4196. (doi:10.1073/pnas.0710672105)