

functionInk: An efficient method to detect functional groups in multidimensional networks reveals the hidden structure of ecological communities

Alberto Pascual-García  | Thomas Bell 

Department of Life Sciences, Imperial College London, Ascot, UK

Correspondence

Alberto Pascual-García
Email: alberto.pascual.garcia@gmail.com

Present address

Alberto Pascual-García, Institute of Integrative Biology, ETH-Zürich, Zürich, Switzerland

Funding information

H2020 European Research Council, Grant/Award Number: 311399-Redundancy; Simons Foundation; Royal Society University Research Fellowship

Handling Editor: Luisa Carvalheiro

Abstract

1. Complex networks have been useful to link experimental data with mechanistic models, and have become widely used across many scientific disciplines. Recently, the increasing amount and complexity of data, particularly in biology, has prompted the development of multidimensional networks, where dimensions reflect the multiple qualitative properties of nodes, links or both. As a consequence, traditional quantities computed in single dimensional networks should be adapted to incorporate this new information. A particularly important problem is the detection of communities, namely sets of nodes sharing certain properties, which reduces the complexity of the networks, hence facilitating its interpretation.
2. In this work, we propose an operative definition of ‘function’ for the nodes in multidimensional networks. We exploit this definition to show that it is possible to detect two types of communities: (a) modules, which are communities more densely connected within their members than with nodes belonging to other communities, and (b) guilds, which are sets of nodes connected with the same neighbours, even if they are not connected themselves. We provide two quantities to optimally detect both types of communities, whose relative values reflect their importance in the network.
3. The flexibility of the method allowed us to analyse different ecological examples encompassing mutualistic, trophic and microbial networks. We showed that by considering both metrics we were able to obtain deeper ecological insights about how these different ecological communities were structured. The method mapped pools of species with properties that were known in advance, such as plants and pollinators. Other types of communities found, when contrasted with external data, turned out to be ecologically meaningful, allowing us to identify species with important functional roles or the influence of environmental variables. Furthermore, we found that the method was sensitive to community-level topological properties like nestedness.
4. In ecology there is often a need to identify groupings including trophic levels, guilds, functional groups or ecotypes. The method is therefore important in providing an objective means of distinguishing modules and guilds. The method we developed, *functionInk* (functional linkage), is computationally efficient at handling

large multidimensional networks since it does not require optimization procedures or tests of robustness. The method is available at: <https://github.com/apascualgarcia/functionlnk>.

KEYWORDS

community detection, functional groups, guilds, microbial networks, modules, multiplex networks, mutualistic networks, trophic networks

1 | INTRODUCTION

Networks have played an important role in the development of ideas in ecology, particularly in understanding food webs (Cohen & Stephens, 1978), and flows of energy and matter in ecosystems (MacArthur, 1955). However, modern ecological datasets are becoming increasingly complex, notably within microbial ecology, where multiple types of information (taxonomy, behaviour, metabolic capacity, traits) on thousands of taxa can be gathered. A single network might therefore need to integrate different sources of information, leading to connections between nodes representing relationships of different types, and hence with different meanings. Advances in network theory have attempted to develop tools to analyse these more sophisticated networks, encompassing multiplex, multilayer or multivariate networks, reviewed in (Kivelä et al., 2014). There could therefore be much value in extending complex networks tools to ecology in order to embrace these new concepts.

In this paper we aim to address a particularly relevant problem in complex networks theory, namely the detection of ‘communities’ (Fortunato, 2010) when the network contains different types of links. In addition, we are interested in finding a method with the flexibility to identify different types of communities. This is motivated by the fact that, in ecology, it is recognized that communities may have different topologies (Allesina & Pascual, 2009) and often an intrinsic multilayer structure (Pilosof, Porter, Pascual, & Kéfi, 2017). We aim to detect two main types of communities. Firstly, the most widely adopted definition of community is the one considering sets of nodes more densely connected within the community than with respect to other communities, often called *modules* (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006). An example in which modules are expected is in networks representing significant co-occurrences or segregations between microbial species, when these relationships are driven by environmental conditions. Since large sets of species may simultaneously change their abundances in response to certain environmental variables (Carr, Diener, Baliga, & Gibbons, 2019; Pascual-García, Tamames, & Bastolla, 2014), this results in large groups of all-against-all co-occurring species, and between-group segregations. Secondly, we are interested in finding nodes sharing a similar connectivity pattern even if they are not connected themselves. An example comes from networks connecting consumers and their resources, when communities are determined looking for consumers sharing similar resource preferences. This idea is aligned with the classic Eltonian definition of niche, which emphasizes the functions of a species rather than their

habitat (Elton, 1927). We call this second class of communities *guilds*, inspired by the ecological meaning in which species may share similar ways of exploiting resources (i.e. similar links) without necessarily sharing the same niche (not being connected themselves), emphasizing the functional role of the species (Simberloff & Dayan, 1991). Consequently, guilds may be quite different to modules, in which members of the same module are tightly connected by definition. The situation in which guilds are prevalent is known as disassortative mixing (Newman & Leicht, 2007), and its detection has received comparatively less attention than the ‘assortative’ situation (which results in modules) perhaps with the exception of bipartite networks (Estrada & Rodríguez-Velázquez, 2005; Newman, 2006).

There are many different approximations for community detection in networks, as summarized in (Schaub, Delvenne, Rosvall, & Lambiotte, 2017). However, despite numerous advances in recent years, it is difficult to find a method that can efficiently find both modules and guilds in multidimensional networks, and that is able to identify which is the more relevant type of community in the network of interest. This might be because there is no algorithm that can perform optimally for any network (Peel, Larremore, & Clauset, 2017), and because each type of approximation may be suited for some networks or to address some problems but not for others, as we illustrate below.

Traditional strategies to detect modules explore trade-offs in quantities like the betweenness and the clustering coefficient (Boccaletti et al., 2006), as in the celebrated Newman–Girvan algorithm (Girvan & Newman, 2002). Generalizing the determination of modules to multidimensional networks is challenging. Consider, for instance, that a node A is linked with a node B and this is, in turn, linked with a node C, and both links are of a certain type. If A is then linked with node C with a different type of link, should the triangle ABC be considered in the computation of the clustering coefficient? One solution proposed comes from the consideration of stochastic Laplacian dynamics running in the network (Lambiotte, Delvenne, & Barahona, 2008), where the permanence of the informational fluxes in certain regions of the network reflects the existence of communities. This approximation has been extended to consider multilayer networks (Mucha, Richardson, Macon, Porter, & Onnela, 2010), even if there are modules defined in different layers that highly overlap, hence defining communities (combination of modules) across layers (De Domenico, Lancichinetti, Arenas, & Rosvall, 2015). A fundamental caveat for these methods is that the links must have a clear interpretation for how their presence affects informational fluxes. Returning to the above example, if the links AB and BC represent

mutualistic interactions and the link AC represents a competitive one, can a random walk follow the link AC when this interaction does not describe a flux of biomass between the species but rather a disruption in the flux of biomass of AB and BC?

A related approach searches for modules using an optimization function that looks for a partitioning in a multilayer network that maximizes the difference between the observed model and a null model which considers the absence of modules (Newman, 2006). This strategy can be applied to multidimensional networks, but raises questions such as which is the appropriate null model, and how to determine the coupling between the different layers defining the different modes of interaction (Bazzi et al., 2016). In addition, since these approximations focus on the detection of modules, they neglect the existence of guilds or other network structures.

Regarding the search of guilds, this problem has received notable attention in social sciences following the notion of structural equivalence. Two nodes are said to be structurally equivalent if they have the same connectivity in the network (Wasserman & Faust, 1994). The connectivity may be defined either analysing if two nodes share the same neighbours, if two nodes are connected with neighbours of the same type even if they are not necessarily the same (following some preassigned roles for the nodes, e.g. prey are structurally equivalent because are connected to predators) or a combination of both. Social agents often have an assigned role, which is why structural equivalence is particularly important in social networks.

An approximation that has exploited the idea of structural equivalence is stochastic blockmodelling (Holland, Laskey, & Leinhardt, 1983), which considers generative models with parameters fitted to the observed network. The approach brings greater flexibility because different models can accommodate different types of communities (De Bacco, Power, Larremore, & Moore, 2017). Therefore, this approximation could be used to search for both modules and guilds (Newman, 2016). There are, however, also caveats to the approach, since it is a challenge to determine whether the underlying assumptions of a particular block model is appropriate for the data being used (Karrer & Newman, 2011). Moreover, even when the model brings an analytically closed form, the estimation of the parameters may be computationally intractable (Valles-Catala, Massucci, Guimera, & Sales-Pardo, 2016), hence requiring costly optimality procedures or tests for robustness (Ganji et al., 2018).

In this work, we build on the idea of structural equivalence noting that a node belonging to either a guild or a module is, in both cases, structurally equivalent to the other nodes in its community. This observation was acknowledged in social sciences in the definition of λ -communities (Borgatti, Everett, & Shirey, 1990), which are types of communities encompassing both modules and guilds, whose relevance has also been previously recognized in the ecological literature (Allesina & Pascual, 2009; Luczkovich, Borgatti, Johnson, & Everett, 2003). From this observation, we wondered whether it is possible to find a similarity measure between nodes that quantifies their structural equivalence, even when different types of links are considered. We could then join nodes according to this similarity measure while monitoring whether the communities that are formed

are guilds or modules. A similar approach was investigated by Yodzis et al. to measure trophic ecological similarity (Yodzis & Winemiller, 1999), but they did not identify an appropriate threshold for determining community membership (which they call 'trophospecies').

We have developed an approach that builds on these results and develops a method to determine objective thresholds for identifying modules and guilds in ecological networks. We show that a modification of the community detection method developed by Ahn, Bagrow, and Lehmann (2010), leads to the identification of two quantities we call internal and external partition densities. For a set of nodes joined within a community by means of their structural equivalence similarity, the partition densities quantify whether their similarities come from connections linking them with nodes outside the community (external density) or within the community (internal density). Notably, our method generates maximum values for the two partition densities along the clustering, allowing us to objectively determine thresholds for the similarity measure in which the communities correspond to the definition of modules (for the internal density) and guilds (external density). Since the elements within both types of communities are structurally equivalent, modules and guilds can be understood as different kinds of *functional groups*—in the Eltonian sense—and this is the name we adopt here. We reserve the term 'community' for a more generic use, because other types of communities beyond functional groups may exist, such as core-periphery structures (Guimera & Amaral, 2005).

We call our method *functionlnk* (functional linkage), emphasizing how the number and types of links of a node determine its functional role in the network. We illustrate its use by considering complex ecological examples, for which we believe the notion of functional role is particularly relevant. We show in the examples that, by combining the external and internal partition densities, we are able to identify the underlying dominant structures of the network (either towards modules or towards guilds). Moreover, selecting the most appropriate community definition in each situation provides results that are comparable to state-of-the-art methods. This versatility in a single algorithm, together with its low computational cost to handle large networks, makes our method suitable for any type of complex, multidimensional network.

2 | MATERIALS AND METHODS

2.1 | Structural equivalence similarity in multidimensional networks

Our method starts by considering a similarity measure between all pairs of nodes that quantifies the fraction of neighbours connected with links of the same type that they share (Figure 1). This is a natural definition of structural equivalence for multidimensional networks, which is agnostic to the specific information that the interaction carries. For simplicity, we present a derivation for a network that contains two types of links. We use undirected positive (+) interactions (e.g. a mutualism) and negative (-) interactions (e.g. competition) to illustrate the method, but these could be replaced by any two link types.

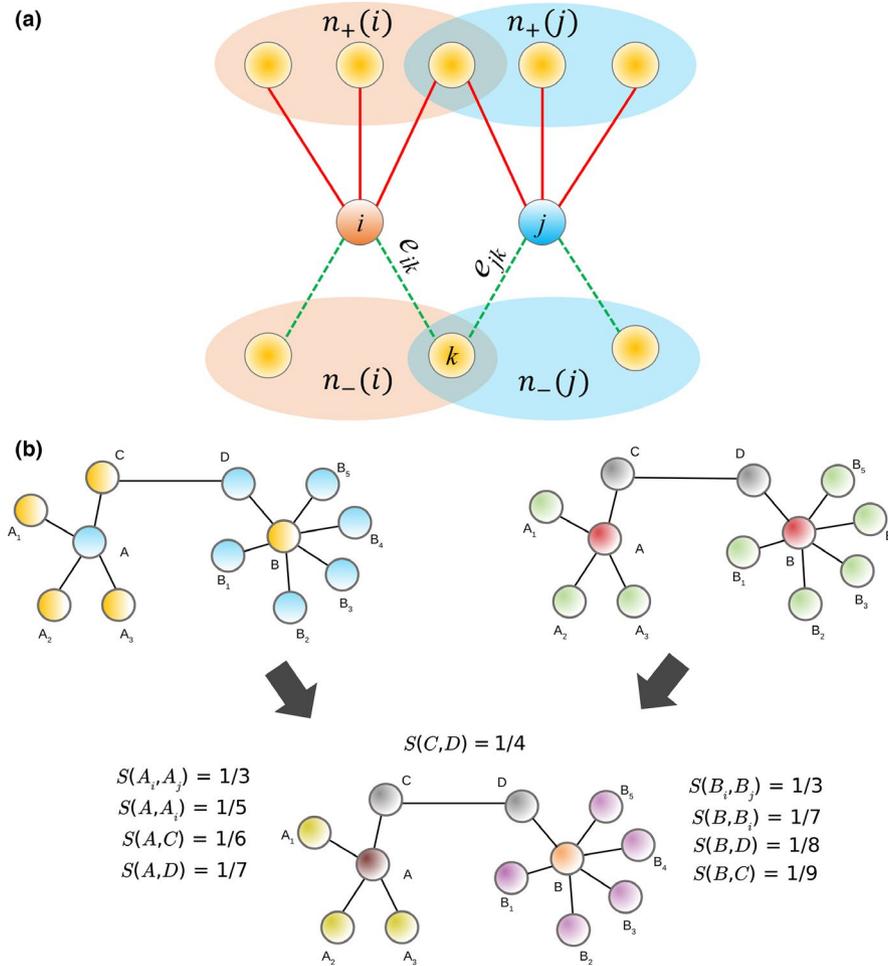


FIGURE 1 Illustration of the method. (a) The similarity between nodes i and j is computed considering the neighbours of each node and the types of interactions that link them. In this example, two types of link are shown: positive (+) interactions are solid links connecting the sets of neighbours $n_+(i)$ and $n_+(j)$. Negative (-) links are shown as dotted links connecting the sets of neighbours $n_-(i)$ and $n_-(j)$. Following Equation 2, $|n(i) \cap n(j)| = 2$ and $|n(i) \cup n(j)| = 8$, which yields $S^{(j)}(i, j) = 2/8$. If, for instance, e_{ik} changes from being - to +, the node k would no longer belong to the set $n(i) \cap n(j)$, being the new similarity: $S^{(j)}(i, j) = 1/8$. In Ahn et al. (2010) the similarity computed in this way is assigned to the links e_{ik} and e_{jk} . (b) Structural equivalence can be defined in different ways. In the top-left network we considered that blue and yellow colours encode a priori information describing the roles of the nodes. Identifying sets of nodes connected similarly to nodes with equivalent roles (i.e. the emphasis is on the roles and not on the specific neighbours, a situation called regular equivalence; Borgatti & Everett, 1989) leads to two communities (the yellow and blue sets of nodes themselves), because every blue node is connected to a yellow one. The method of Guimera and Amaral (2005), determines communities focusing on their topological role (top-right network) by identifying central (A and B), peripheral (A₁-A₃ and B₁-B₅) and connector nodes (C and D). functionInk (bottom network) defines communities by joining nodes with approximately the same neighbours and, if there are roles for the nodes, these can be incorporated defining link types (one type for each pair of roles connected, in the example only one type is needed). All non-zero Jaccard similarities of the example are shown. Clustering these similarities will lead to different partitions and, stopping at $S^{(j)} = 1/4$, communities being the intersection of those found in the above networks are obtained, highlighting the potential to identify communities considering both the roles and topological features. Figure adapted from Guimera and Amaral (2005)

Extending the method to an arbitrary number of link types is presented in the Supporting Information. We call $\{i\}$ the set of N nodes and $\{e_{ij}\}$ the set of M links in a network. We call $n(i)$ the set of neighbours of i , that can be split into different subsets according to the types of links present in the network.

For two types of links, we split the set of neighbours linked with the node i into those linked through positive relationships, $n_+(i)$, or through negative relationships, $n_-(i)$; we follow a notation similar to the one presented in (Ahn et al., 2010), but note that $n(i)$ there denotes neighbours irrespective of the type of links. Distinguishing link

types induces a division in the set of neighbours of a given node into subsets sharing the same link type, as shown in Figure 1a. More specifically, in the absence of link types we define the Jaccard similarity between two nodes i and j as:

$$S^{(j)}(i, j) = \frac{|n(i) \cap n(j)|}{|n(i) \cup n(j)|}, \quad (1)$$

where $|\cdot|$ is the cardinality of the set (the number of elements it contains). This metric was shown to lead to clusters of species that are

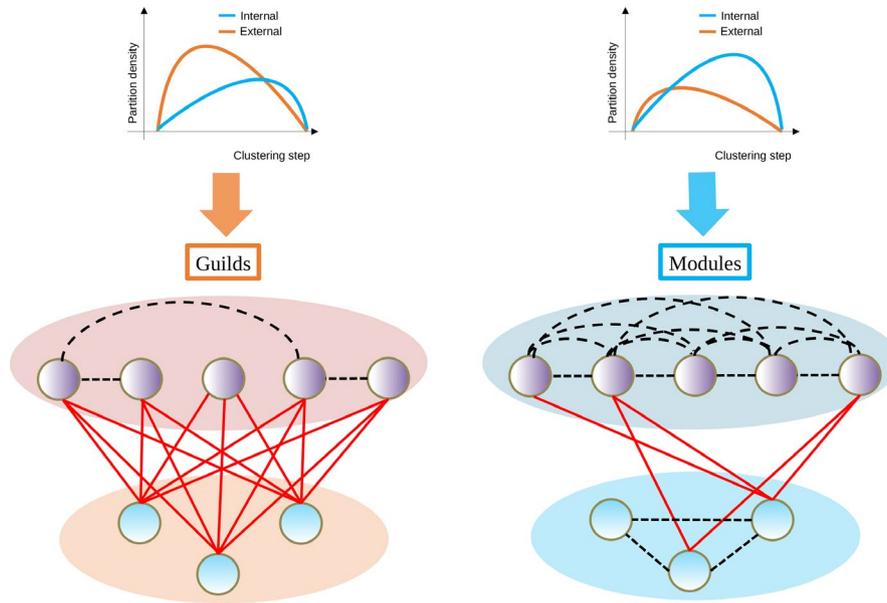


FIGURE 2 Definition of guilds and modules. For each set of nodes n_c^{int} belonging to the same community c (nodes within the same shaded area) we consider the number of links within the community (black dashed links, called m_c^{int} in the main text) out of the total number of possible internal links, to compute the internal partition density (see upper curves). We also computed the external partition density, which is the density of links connecting nodes external to the community (m_c^{ext} , solid red lines linking nodes belonging to different communities) out of the total number of possible external links. We call guilds the communities determined at the maximum of the external partition density, and modules those found at the maximum of the internal partition density. The relative value of the external and internal partition densities allow us to estimate which kind of community dominates the network. In the example, guilds dominate the network on the left, and modules dominate the network on the right

more consistent with cophenetic clustering than other alternatives (Yodzis & Winemiller, 1999), and generalizing this expression to multiple attributes is achieved simply by differentiating the type of neighbours depending on the types of connections. For two attributes (see Supporting Information for an arbitrary number of attributes) this leads to

$$S^{(j)}(i, j) = \frac{|n_+(i) \cap n_+(j)| + |n_-(i) \cap n_-(j)|}{|n_+(i) \cup n_+(j) \cup n_-(i) \cup n_-(j)|} \quad (2)$$

Accounting for the weight of the links can be made with the generalization of the Jaccard index provided by the Tanimoto coefficient (Tanimoto, 1958), $S^{(T)}(i, j)$, presented in Supporting Information.

Finally, we introduce a modification to the above definition of $S^{(j)}$ to account for the particular case in which i and j are only connected between themselves, i.e. they do not share any neighbours according to the above definition. This is problematic because we want to distinguish this situation from the one in which they do not share any nodes, for which we get $S(i, j) = 0$. We resolve this situation by considering that a node is its own neighbour, in which case two nodes only connected between themselves would yield $S(i, j) = 1$. However, we note that this would also be the value between two nodes that are connected and that also share all neighbours (a motif known as a clique), irrespective of the number of neighbours they share, because the similarity measure saturates. We argue that this situation is unsatisfactory because there is stronger evidence that two nodes are structurally equivalent when they share connections

and creating transitive motifs, since transitivity is a key property in the definition of equivalence classes (Pascual-García, Abia, Ortiz, & Bastolla, 2009). The situation can be resolved by using the convention that, for two connected nodes, the intersection set is reduced by one, i.e. $|n(i) \cap n(j)| \rightarrow |n(i) \cap n(j)| - 1$. This convention has the interesting property that, for cliques, increasing the number of nodes involved also increases the similarity between its members, resulting in an upper bound of one and a lower bound of 1/2 (a 2-node clique). In addition, two connected nodes that share neighbours but are not connected themselves have a smaller difference in the similarity compared to nodes within cliques that share the same number of neighbours, thus facilitating the identification of guilds. In Figure 1 we illustrate the computation of this similarity with a simple example.

2.2 | Identification of communities through clustering and similarity cut-offs

Once the similarity between nodes is computed, the next objective is to define and identify structurally equivalent communities. As explained in the Introduction, there are different possible definitions of structural equivalence, illustrated in Figure 1b. In the figure is shown how the similarity metric proposed together with an agglomerative clustering to join nodes in communities, encapsulates these different notions of structural equivalence. A critical question, however, is how to objectively determine the threshold to stop the clustering (Harrer & Schmidt, 2013)?

This question is often addressed by iteratively ‘partitioning’ the network into the distinct communities, and monitoring each partition with a function having a well-defined maximum or minimum that determines the threshold of the optimal partition. In Ahn et al. (2010), the authors proposed to join links of a network according to a similarity measure between the links with an agglomerative clustering, and to monitor the clustering with a quantity called the partition density. The partition density is the weighted average across communities of the number of links within a community out of the total possible number of links (which depends on the number of nodes in the community). We re-considered the method of Ahn et al. (2010; which was originally defined over partitions of links, see Supporting Information), to work over partitions of nodes, and we developed two partition densities, with two distinct meanings. To develop these measures we noted that, when joining nodes into a cluster, we are concluding that these nodes share (approximately) the same neighbours connected with the same type of links, but the nodes joined may or may not be connected between them. We therefore redefined the partition density so that it distinguishes between the contribution to the link density arising from the connections *within* a community from connections shared with external nodes *between* communities.

Formally, given a node i , we differentiate neighbours that are within the same community ($n^{\text{int}}(i)$, where int stands for ‘interior’) from neighbours that are in different communities ($n^{\text{ext}}(i)$), hence $n(i) = n^{\text{int}}(i) \cup n^{\text{ext}}(i)$ (Figure 2). For a singleton (a community of size one) $n^{\text{int}}(i) = \{i\}$ and $n^{\text{ext}}(i) = \emptyset$. Similarly, the set of links $m(i)$ connecting the node i with other nodes can also be split into two sets: the set connecting the node with neighbours within its community $m^{\text{int}}(i)$, and those connecting it with external nodes $m^{\text{ext}}(i)$. This distinction was also considered in the problem of coloring nodes (Everett & Borgatti, 1996).

Therefore, for each partition of nodes into T communities our method identifies, for each community c , the total number of nodes it contains, n_c^{int} , and the total number of links connecting these nodes m_c^{int} . In addition, it computes the total number of nodes in other communities that have connections to the nodes in the community, n_c^{ext} , through a number of links m_c^{ext} . Clearly, to identify n_c^{ext} neighbours, at least n_c^{ext} links are required and thus an increasing number of links in excess, $m_c^{\text{ext}} - n_c^{\text{ext}}$ are necessary to obtain an increasing contribution to the similarity of the nodes in the community through external links (however, this is not a sufficient condition, see Supporting Information). In this way, a relevant quantity to characterize a community is the fraction of links in excess out of the total possible number ($m_c^{\text{ext}} - n_c^{\text{ext}}/n_c^{\text{ext}}(n_c^{\text{int}} - 1)$). We note that this calculation does not take into account multiple link types. The weighted average of this quantity through all communities leads to the definition of external partition density:

$$D^{\text{ext}} = \frac{1}{M} \sum_c \frac{m_c^{\text{ext}}}{2} \frac{(m_c^{\text{ext}} - n_c^{\text{ext}})}{n_c^{\text{ext}}(n_c^{\text{int}} - 1)}, \quad (3)$$

where M is the total number of links. We now follow a similar reasoning to consider a necessary condition to obtain an increasing contribution

to the similarity of the nodes through the internal links (see Supporting Information). We acknowledge that in a community created by joining nodes through the similarity measure we propose, it may happen that $n_c^{\text{int}} > 0$ even if $m_c^{\text{int}} = 0$. Therefore, any link is considered a link in excess, leading to the following expression for the internal partition density, which quantifies the fraction of internal links in excess out of the total:

$$D^{\text{int}} = \frac{1}{M} \sum_c m_c^{\text{int}} \frac{2m_c^{\text{int}}}{n_c^{\text{int}}(n_c^{\text{int}} - 1)}. \quad (4)$$

Finally, we define the total partition density as the sum of both internal an external partition densities:

$$D^{\text{total}} = D^{\text{int}} + D^{\text{ext}},$$

and hence, if all the fractions in D^{int} and D^{ext} are equal to one, i.e. all possible links in excess are realized, D^{total} equals to one. Since at the beginning of the clustering the communities have a low number of members, most of the contribution towards D^{total} comes from D^{ext} while, in the final steps, where the communities become large, D^{int} will dominate. All three quantities will reach a maximum value along the clustering (for the internal it could be at the last step) and, if one of them clearly achieves a higher value, it will be indicative that one type of functional group is dominant in the network. If that is the case, the maximum of D^{total} —which is always larger or equal to $\max(\max(D^{\text{int}}), \max(D^{\text{ext}}))$ —, will be at a clustering step close to the step in which the dominating quantity peaks. If neither D^{ext} nor D^{int} clearly dominates, D^{total} will peak at an intermediate step between the two partial partition densities maxima, suggesting that this intermediate step is the best candidate of the optimal partition for the network. Communities determined at this intermediate point where they can be both guilds and modules will be called, generically, functional groups.

3 | RESULTS

3.1 | Plant-pollinator networks

To illustrate the use of the method we start analysing a synthetic example. In ecological systems, species are often classified into communities according to their ecological interactions, such as in mutualistic networks of flowering plants and their animal pollinators. These networks are characterized by intraspecific and interspecific competition within both the pool of plants and the pool of animals, and by mutualistic relationships between plants and animals, leading to a bipartite network.

To investigate the performance of our method and, in particular, the influence of the topological properties into the partition density measures, we generated a set of artificial mutualistic networks with diverse topological properties, following the method presented in (Pascual-García & Bastolla, 2017). For the mutualistic interactions,

we focused on two properties: the connectance κ_{mut} , which is the fraction of observed interactions out of the total number of possible interactions, and the nestedness ν as defined in (Bastolla et al., 2009; see Supporting Information), which codifies the fraction of interactions that are shared between two species of the same pool, averaged over all pairs of species. We selected these measures for their importance in the stability-complexity debate in mutualistic systems (Pascual-García & Bastolla, 2017), and the similarity between the nestedness (which, in the definition we adopt here, represents the mean ecological overlap between species) and the notion of structural equivalence we considered (see Supporting Information). For the competition matrices, we considered random matrices with different connectances, κ_{comp} , since it is difficult to estimate direct pairwise competitive interactions experimentally, and they are frequently modelled with a mean field competition matrix.

We verified that in all networks the set of plants and animals are joined in the very last step of the clustering irrespective of the clustering method used, a result that must follow by construction. As expected, the curves monitoring the external and internal partition densities depend on the properties of the networks. We illustrate this finding in Figure 3, where we have selected two networks with contrasting topological properties. One of the networks has high connectance within the pools and low connectance and nestedness between the pools. The internal partition density peaks at the last step minus one (i.e. where the two pools are perfectly separated) consistent with the definition of modules, where the intra-modules link density is higher than the inter-modules link density. On the other hand, the second network has intra-pool connectance equal to zero, and very high connectance and nestedness between the pools (see Figure 3). We selected $\kappa_{\text{comp}} = 0$ for simplicity in the network representation, but similar results are obtained for low values of κ_{comp} , see for instance Figure S4. In this second network (see Figure 3, right panel), only the external partition density peaks and, at the maximum, the communities that we identified clearly reflect the structural equivalence of the node members in terms of their connectance with nodes external to the group, as we expect for the definition of guilds. The ecological information retrieved for guilds is clearly distinct from the information retrieved for the modules, the former being related to the topology of the network connecting plants and animals. We observe that guilds identify specialist species clustered together, which are then linked to generalist species of the other pool: a structure typical of networks with high nestedness.

The method identified several interesting guilds and connections between them. For instance, generalists Plant 1, Animal 1 and Animal 2 (and to a lesser extent Plant 2) have a low connectivity between them but, being connected to many specialists, determine a region of high vulnerability, in the sense that a directed perturbation over these species would have consequences for many other species. This is confirmed by the high betweenness of these nodes (proportional to the size of the node in the network). In addition, the algorithm is able to identify more complex partitions of nodes into communities. As an example of this, Animal 16 (turquoise) is split from Animals 10 and 11 (cyan), which form a second community, and

from Animals 15, 18 and 19 (light pink) that are joined into a third community, despite of the subtle connectivity differences between these six nodes. Finally, it also detects communities of three or more species that have complex connectivity patterns which, in this context, may be indicative of functionally redundant species (e.g. red and blue communities).

Examples with other intermediate properties are analysed in the Figures S2 and S3. Broadly speaking, either the internal or the total partition density maximum peaks at the last step minus one, allowing for detection of the two pools of species. Nevertheless, the method fails to find these pools in situations in which the similarity between members of distinct pools is comparable to the similarity of members belonging to the same pool. This may be the case if the connectances are small (see Figure S4). The relative magnitude of the external versus internal partition density depends on the connectance between the pools of plants and animals and on the connectance within the pools respectively (see Figure S2). Interestingly, networks for which the nestedness is increased keeping the remaining properties the same generated an increase in the external partition density (see Figure S3). These examples illustrate how the external partition density is sensitive to complex topological properties, in particular to an increase in the disassortativity of the network, as expected when guilds are dominant.

3.2 | Trophic networks

We tested our method in a comprehensive multidimensional ecological network of 106 species distributed in trophic layers with approximately 4,500 interactions, comprising trophic and non-trophic interactions (approximately 1/3 of the interactions are trophic; Kéfi, Miele, Wieters, Navarrete, & Berlow, 2016). This network was analysed looking for communities extending a stochastic blockmodelling method (Newman & Leicht, 2007) to deal with different types of interactions (Kéfi et al., 2016). The estimation of the parameters of the model through an Expectation-Maximization algorithm requires controlling the influence of random starting conditions since each initial condition may lead to a different result, and hence is needed to test the robustness of the results. Here we show that, in this example, our method is comparable with this approximation, and it has the advantage of being deterministic. Moreover, the simplicity of the method allows us to handle large networks with arbitrary number of types of links and to evaluate the consistency of the communities found, as we show in the following.

Our method finds a maximum for the internal density when there are only three communities. Previous descriptions of the network identified three trophic levels in the network (Predators, Herbivores and Basal species). The latter are further subdivided into subgroups (e.g. Kelps, Filter feeders), and there are some isolated groups like one Omnivore and Plankton. To match these subgroups we observed that the total partition density reaches a maximum close to the maximum of the external partition density (step 69) and maintains this value along a plateau until step 95 (see Figure S5). We analysed results at both clustering thresholds finding that, at step 95, we obtain

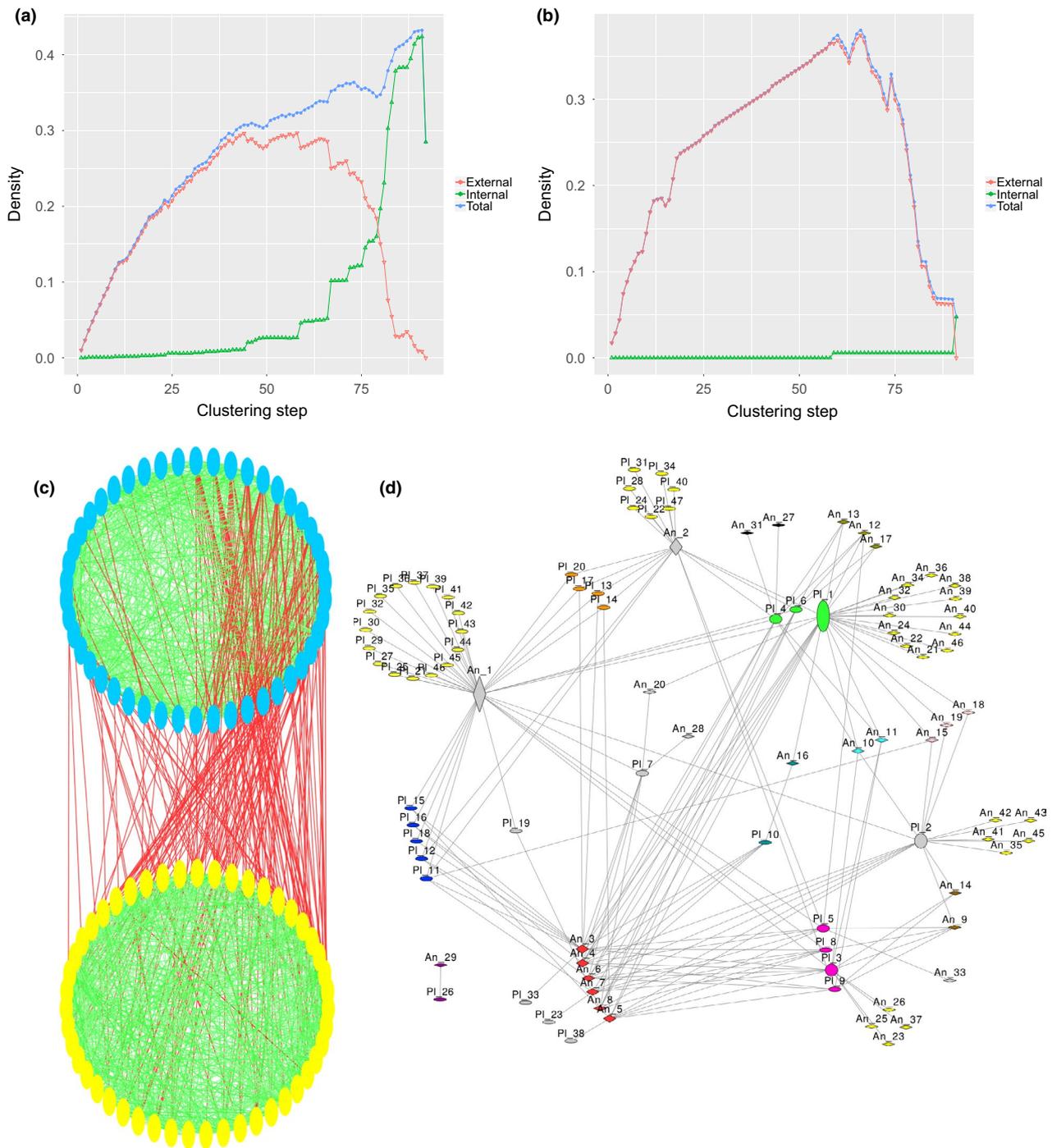


FIGURE 3 Analysis of synthetic mutualistic networks. (Top left) Partition densities for a network with $\kappa_{\text{comp}} = 0.5$, nestedness $\nu = 0.15$ and $\kappa_{\text{mut}} = 0.08$ and (top right) for a network with $\kappa_{\text{comp}} = 0$, nestedness $\nu = 0.6$ and $\kappa_{\text{mut}} = 0.08$. The high density of competitive links in the first network makes the internal partition density dominate, leading to two modules representing the plant-pollinator pools (bottom left network), while reducing the density of competitive links to zero in the second network makes the external partition density to dominate, finding guilds (bottom right, with plants labelled 'PI' and animals labelled 'An'). The small increase in the internal partition density for this network at step 59 is due to two specialist species joined at that step (animal 29 and plant 56, shown at the bottom left of the network). Nodes are coloured according to their functional group in both networks although, in the network finding guilds (bottom right), specialist species are yellow, single species communities are grey and the size of the nodes is proportional to their betweenness

modules with a good agreement with the trophic levels, as shown in Figure 4. On the other hand, at step 69 we find a larger number of communities, some of which fit the definition of modules and others the definition of guilds (see Figure 4).

To shed some light on the information obtained from this second network, we compared the classification obtained by Kéfi et al. (2016; in the following reference classification) and our method. We computed several similarity metrics comparing the classification we

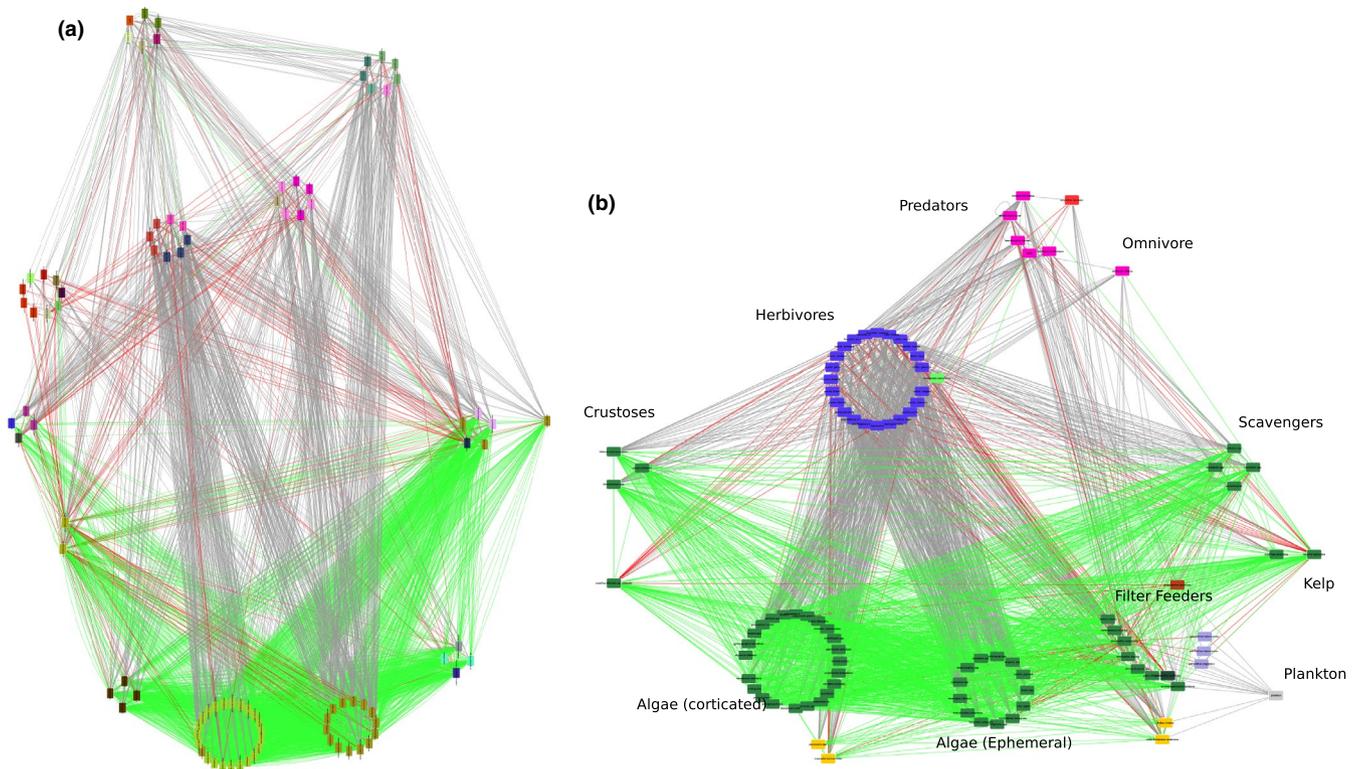


FIGURE 4 Determination of guilds and modules in a large trophic network. Trophic networks with links representing trophic (grey), non-trophic positive (red) and negative (green) interactions. (Left) Nodes are grouped according to the classification found in Kéfi et al. (2016; reference classification), and coloured by the guilds found with functionInk at the maximum of the external partition density. (Right) Nodes are grouped according to the trophic levels and coloured by the modules found by functionInk (see main text for details). The modules separate the three main trophic levels: predators, herbivores and basal species, further separating some of them into subgroups, such as filter feeders and plankton, which is an orphan module

obtained at each step of the agglomerative clustering with functionInk and the reference classification (see Section 2). In Figure 5, we show that the similarity between both classifications is highly significant (Z -score > 2.5) and is maximized when the external partition density is also maximized, i.e. at step 69. This is particularly apparent for the Wallace 01, Wallace 10 and Rand indexes (see Figure 6; Figure S6). Notably, communities in the reference classification were also interpreted as functional groups in the same sense proposed here (Kéfi et al., 2016).

Nevertheless, there are some discrepancies between both classifications. In particular, although there is a complete correspondence between the two largest communities in both classifications, there are a number of intermediate communities in the reference classification whose members are classified differently in our method. To illustrate these discrepancies, we plotted a heatmap of the Tanimoto coefficients of members of four communities of intermediate sizes containing discrepancies, showing their membership in both the reference and the functionInk classification with different colours (see Figure 5). The dendrograms cluster rows and columns computing the Euclidean distance between their values. Therefore, these dendrograms are very similar to the method encoded in functionInk, and the communities must be consistent, representing a powerful way to visually inspect results. Indeed, the dendrograms are in correspondence with both functionInk and reference communities,

but we observe some discrepancies. For instance, the community found by the reference classification containing several *Petrolishtes* species, joins species that have low similarity regarding the number and type of interactions as measured by the Tanimoto coefficients, while functionInk joins together the three species with high similarity, leaving aside the remainder species. Therefore, despite the methodological differences between both methods, the different classifications produce similar outcomes, but result in different sized clusters (a different cut-off in the dendrograms, with functionInk finding finer clusters). The advantage of functionInk is then apparent in the simplicity of the method, which permits validation through visual inspection of the consistency of the classification.

3.3 | Microbial networks

We discuss a last example of increasing importance in current ecological research, which is the inference of interactions among microbes sampled from natural environments. We considered a large matrix with more than 700 samples of 16S rRNA operative taxonomic units (OTUs) collected from rain pools (water-filled tree-holes) in the United Kingdom (Pascual-García & Bell, 2019a; Rivett & Bell, 2018; see Supporting Information). We analysed β -diversity similarity of the samples contained in the matrix with the Jensen–Shannon

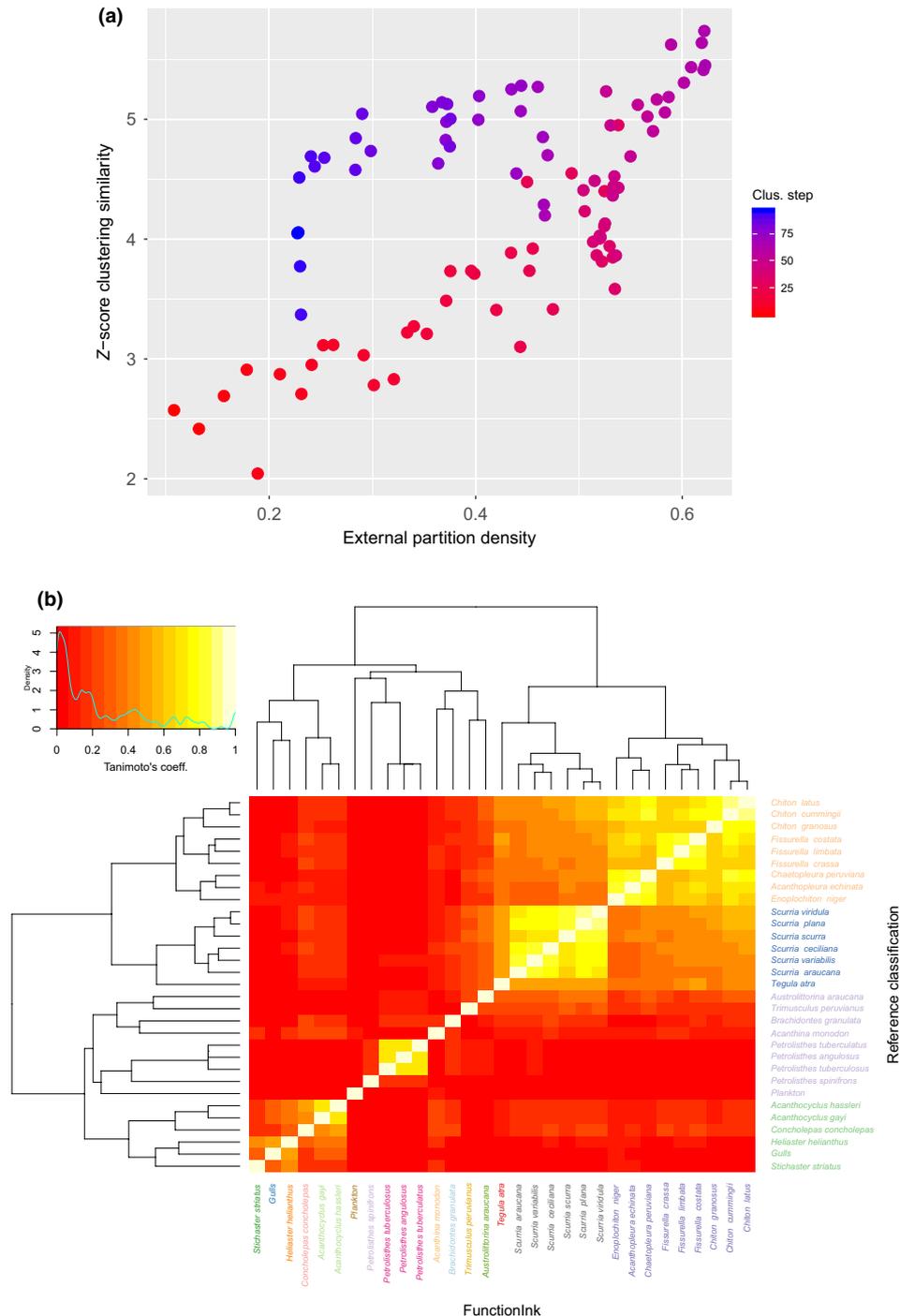


FIGURE 5 Comparison between the reference classification in the trophic network and functionInk. (Top) Z-score of the Wallace 10 index (Shotwell, 2013), measuring the similarity between the reference classification and the functionInk method at each clustering step. The similarity with the reference classification (see main text) is maximized around the maximum of the external partition density. (Bottom) Comparison of communities 1, 4, 7 and 9 in the reference classification, whose members were classified differently by functionInk. Colours in the names of species in rows (columns) represent community membership in the reference (functionInk) classifications. The heatmap represents the values of the Tanimoto coefficients, and the dendrograms are computed using Euclidean distance and clustered with complete linkage. Both classifications are generally consistent with the dendrograms, but with functionInk finding finer clusters

divergence metric (Endres & Schindelin, 2003), further classifying the samples automatically, leading to six disjoint clusters we call β -diversity classes (i.e. clusters of samples, see Section 2). Next, we inferred a network of significant positive (co-occurrences) or negative (segregations) correlations between OTUs using SparCC (Friedman

& Alm, 2012; see Section 2), represented in Figure S7. Applying functionInk to the network of inferred correlations, we aimed to understand the consistency between the results of functionInk (modules and guilds) and the β -diversity classes. The rationale is that, by symmetry, communities determined from significant co-occurrences and

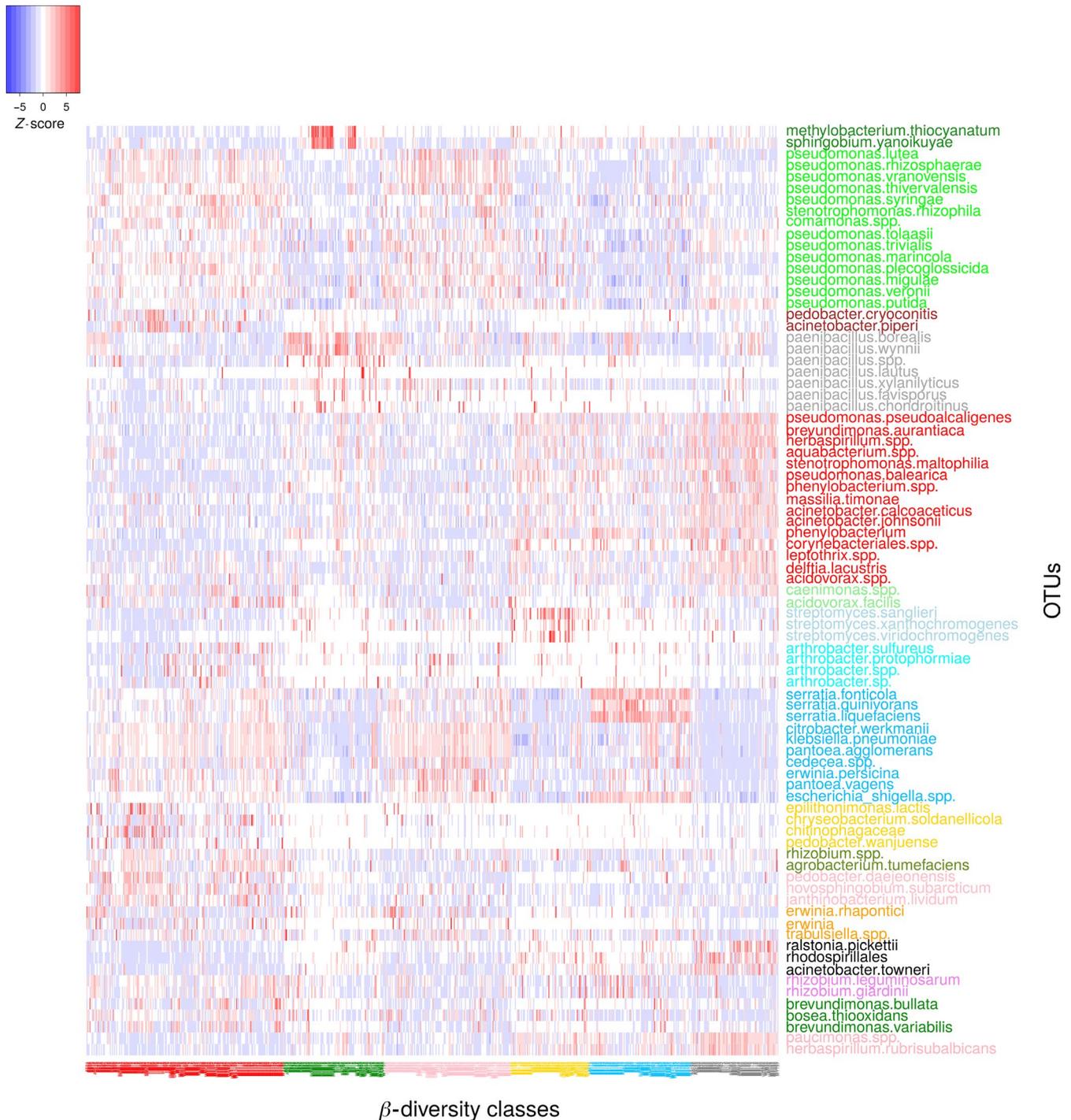


FIGURE 6 Comparison between β -diversity classes and functional groups in a microbial network. Heatmap representing the Z-score of the log-transformed abundances of the OTUs (see Section 2). Species are coloured according to their functional group membership obtained at the maximum of the total partition density. Samples are coloured according to one of the six community classes found in Pascual-García and Bell (2019a) after optimal clustering with a β -diversity distance. Orphan clusters were excluded except for five *Paenibacillus* species (characteristic of the green class) that were added to the functional group formed by *Paenibacillus borealis* and *Paenibacillus wynnii*. The heatmap blocks show segregation and co-occurrence between modules, further mapping the β -diversity classes

segregations between OTUs should reflect the similarity and dissimilarity between the samples, hence validating the method.

Contrasting with the trophic network analysed in the previous example, the external partition density brings a poor reduction of the complexity of the network (peaking after only 22 clustering

steps), and the internal partition density is higher, hence suggesting a more relevant role for modules (see Figure S8). Differences in the three stopping criteria are shown in Figure S7, where two large modules are apparent, with a large number of intra-cluster co-occurrences (continuous links) and intercluster segregations

(dotted links). Note that this is quite different to what is found in macroscopic trophic networks, where pools of species (e.g. prey) have within module competitive (segregating) interactions, while between-modules interactions can be positive (for predators) or negative (for prey).

There is reasonable agreement between the functional groups found at the maximum of the total partition density and the β -diversity classes, as shown in Figure 6. Moreover, the detection of networks complements the information that β -diversity classes provides, since it is possible to individuate the key players of these classes (see Supporting Information). Notably, it was shown in Pascual-García and Bell (2019a) that the β -diversity classes might be related to a process of ecological succession driven by environmental variation, the functional groups are likely driven by environmental preferences rather than by ecological interactions, likely explaining the large number of positive co-occurrences. This speaks against a naive interpretation of correlation networks in microbial samples as ecological interactions unless environmental preferences are under control (Carr et al., 2019; Pascual-García et al., 2014).

4 | DISCUSSION

We presented a novel method for the analysis of multidimensional networks, with nodes with an arbitrary number of link types. We implemented the method adopting the definition of structural equivalence, which underlies both the similarity measure definition and the rationale behind both the clustering and our definition of partition densities. We selected a set-theoretic similarity measure quantifying the number of nodes that are shared with the same type of interaction, which we believe is a natural definition of structural equivalence for multidimensional networks, and that has the advantage that it does not make assumptions on how the information flows in the network, typical of approximations based on Laplacian dynamics (see e.g. Lambiotte et al., 2008; Mucha et al., 2010). This allows us to join nodes simply by their similarity, with no need for specific assumptions about the network structure. Moreover, this similarity can also be naturally linked to two measures of nodes' partitioning that allowed us to propose a clear differentiation between modules (determined by the maximum of the internal partition density) and guilds (determined by the maximum of the external partition density).

Beyond these technical advantages, we illustrated the versatility of functionInk using several ecological examples. The relative value between the internal and external partition density immediately yields information on whether the network is dominated by modules, guilds or intermediate structures. This allows for increasing flexibility in the analysis of the networks, and for a more nuanced interpretation of network structure and species' roles in the ecosystem. For both mutualistic and trophic networks, the internal partition density correctly finds the trophic layers, justifying the success of the original method (Ahn et al., 2010). Our extension recovered the functional groups as determined by Kéfi et al. (2016) through the external partition density, and the visual inspection reflects a good

consistency with the definition we proposed for functional groups in terms of structural equivalence. Moreover, in the mutualistic networks, we showed that the functional groups discovered in this way were sensitive to changes to high-order topological properties such as the nestedness.

The analysis of the microbial network was dominated by modules rather than guilds. Interestingly, these modules had intra-cluster positive correlations, contrary to what would be expected in a macroscopic trophic network, where competitive interactions would be dominant between members of the same trophic layer. We selected in this example for further exploration the functional communities found at the maximum of the total partition density, with some groups having properties closer to those of guilds and others closer to modules. The communities that we identified were in good agreement with the functional communities found using β -diversity similarity (Pascual-García & Bell, 2019a), supporting the consistency of the method.

To finish, we highlight some limitations of the method. Firstly, it may have problems if the communities are highly overlapping (Ahn et al., 2010; De Domenico et al., 2015). In these cases, it would be convenient to inspect the partition at the three classifications given by the different partition densities, since it is likely that overlapping communities are split in an earlier classification and then joined at later steps of the clustering. Another possibility is to combine it with the approximation proposed in (Ahn et al., 2010), that has both compatible and, at the same time, complementary results (see Supporting Information). To continue with, our approximation does not consider yet the case in which there are multiedges in the network, although real networks are typically very sparse and the probability of finding multiedges is small (Karrer & Newman, 2011). Finally, although the method might not be able to achieve the generality of other approximations aiming to find any arbitrary structure in the network (De Bacco et al., 2017; Ganji et al., 2018; Newman & Leicht, 2007), such approximations require either heuristics to find a solution for the parameters—and hence a unique optimal solution is not guaranteed—or a computationally costly sampling of the parameter space. Our method relies on a deterministic method whose results are easily inspected, and its computational cost for a network with N node scales as N^2 for the similarity metric computation plus the clustering, which is order N . The method is freely available in the address (<https://github.com/apascualgarcia/functionInk>) and, importantly, although we developed it with ecological networks in mind, it can be applied to any kind of network.

ACKNOWLEDGEMENTS

A.P.-G. thanks Sebastian Bonhoeffer for his support. We thank Michael Schaub for critical comments on an earlier version of the manuscript and to two anonymous referees whose comments greatly improved the manuscript. The research was funded by a European Research Council starting grant (311399-Redundancy) awarded to T.B. T.B. was also funded by a Royal Society University Research Fellowship. A.P.-G. was also funded by the Simons

Collaboration: Principles of Microbial Ecosystems (PriME), award number 542381.

AUTHORS' CONTRIBUTIONS

A.P.-G. conceived the project and designed the methodology; A.P.-G. performed the analysis; T.B. contributed data; A.P.-G. and T.B. analysed the data; A.P.-G. led the writing of the manuscript. Both authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

The method is available at <https://github.com/apascualgarcia/functionlnk> together with the synthetic mutualistic networks analysed in the first part of the manuscript. The first release is deposited in Zenodo with <https://doi.org/10.5281/zenodo.3653947> (Pascual-García, 2020). Data published in Pascual-García and Bell (2019b) were retrieved from the URL <https://doi.org/10.5281/zenodo.3539537> and data in Kéfi et al. (2016) from URL <https://doi.org/10.5061/dryad.b4vg0>.

ORCID

Alberto Pascual-García  <https://orcid.org/0000-0002-8444-3196>

Thomas Bell  <https://orcid.org/0000-0002-2615-3932>

REFERENCES

- Ahn, Y.-Y., Bagrow, J. P., & Lehmann, S. (2010). Link communities reveal multi-scale complexity in networks. *Nature*, *466*(7307), 761–764. <https://doi.org/10.1038/nature09182>
- Allesina, S., & Pascual, M. (2009). Food web models: A plea for groups. *Ecology Letters*, *12*(7), 652–662.
- Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B., & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, *458*, 1018–1020. <https://doi.org/10.1038/nature07950>
- Bazzi, M., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., & Howison, S. D. (2016). Community detection in temporal multilayer networks, with an application to correlation networks. *Multiscale Modeling & Simulation*, *14*(1), 1–41. <https://doi.org/10.1137/15M1009615>
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4), 175–308.
- Borgatti, S. P., & Everett, M. G. (1989). The class of all regular equivalences: Algebraic structure and computation. *Social Networks*, *11*(1), 65–88. [https://doi.org/10.1016/0378-8733\(89\)90018-X](https://doi.org/10.1016/0378-8733(89)90018-X)
- Borgatti, S. P., Everett, M. G., & Shirey, P. R. (1990). LS sets, lambda sets and other cohesive subsets. *Social Networks*, *12*(4), 337–357. [https://doi.org/10.1016/0378-8733\(90\)90014-Z](https://doi.org/10.1016/0378-8733(90)90014-Z)
- Carr, A., Diener, C., Baliga, N. S., & Gibbons, S. M. (2019). Use and abuse of correlation analyses in microbial ecology. *The ISME Journal*, *13*(11), 2647–2655. <https://doi.org/10.1038/s41396-019-0459-z>
- Cohen, J. E., & Stephens, D. W. (1978). *Food webs and niche space*. Princeton, NJ: Princeton University Press.
- De Bacco, C., Power, E. A., Larremore, D. B., & Moore, C. (2017). Community detection, link prediction, and layer interdependence in multilayer networks. *Physical Review E*, *95*(4), 042317. <https://doi.org/10.1103/PhysRevE.95.042317>
- De Domenico, M., Lancichinetti, A., Arenas, A., & Rosvall, M. (2015). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Physical Review X*, *5*(1), 011027. <https://doi.org/10.1103/PhysRevX.5.011027>
- Elton, C. S. (1927). *Animal ecology*. New York, NY: The Macmillan Company.
- Endres, D. M., & Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information Theory*, *49*, 1858–1860. <https://doi.org/10.1109/TIT.2003.813506>
- Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Spectral measures of bipartivity in complex networks. *Physical Review E*, *72*(4), 046105. <https://doi.org/10.1103/PhysRevE.72.046105>
- Everett, M. G., & Borgatti, S. P. (1996). Exact colorations of graphs and digraphs. *Social Networks*, *18*(4), 319–331. [https://doi.org/10.1016/0378-8733\(95\)00286-3](https://doi.org/10.1016/0378-8733(95)00286-3)
- Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3–5), 75–174.
- Friedman, J., & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Computational Biology*, *8*(9), e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>
- Ganji, M., Chan, J., Stuckey, P. J., Bailey, J., Leckie, C., Ramamohanarao, K., & Davidson, I. (2018). Image constrained blockmodelling: A constraint programming approach. In M. Ester & D. Pedreschi (Eds), *Proceedings of the 2018 SIAM international conference on Data Mining* (pp. 19–27). Philadelphia, PA: SIAM.
- Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(12), 7821–7826. <https://doi.org/10.1073/pnas.122653799>
- Guimera, R., & Amaral, L. A. N. (2005). Cartography of complex networks: Modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, *2005*(02), P02001.
- Harrer, A., & Schmidt, A. (2013). Blockmodelling and role analysis in multi-relational networks. *Social Network Analysis and Mining*, *3*(3), 701–719. <https://doi.org/10.1007/s13278-013-0116-x>
- Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic block-models: First steps. *Social Networks*, *5*(2), 109–137. [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, *83*(1), 016107. <https://doi.org/10.1103/PhysRevE.83.016107>
- Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., & Berlow, E. L. (2016). Data from: How structured is the entangled bank? The surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience. *Dryad Digital Repository*, <https://doi.org/10.5061/dryad.b4vg0>
- Kivelä, M., Arenas, A., Barthélemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, *2*(3), 203–271.
- Lambiotte, R., Delvenne, J.-C., & Barahona, M. (2008). Laplacian dynamics and multiscale modular structure in networks. *arXiv Preprint arXiv:0812.1770*.
- Luczakovich, J. J., Borgatti, S. P., Johnson, J. C., & Everett, M. G. (2003). Defining and measuring trophic role similarity in food webs using regular equivalence. *Journal of Theoretical Biology*, *220*(3), 303–321. <https://doi.org/10.1006/jtbi.2003.3147>
- MacArthur, R. (1955). Fluctuations of animal populations and a measure of community stability. *Ecology*, *36*, 533. <https://doi.org/10.2307/1929601>
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time dependent, multiscale, and multiplex networks. *Science*, *328*(5980), 876–878. <https://doi.org/10.1126/science.1184819>
- Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, *74*(3), 036104. <https://doi.org/10.1103/PhysRevE.74.036104>
- Newman, M. E. (2016). Equivalence between modularity optimization and maximum likelihood methods for community detection. *Physical Review E*, *94*(5), 052315. <https://doi.org/10.1103/PhysRevE.94.052315>

- Newman, M. E., & Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23), 9564–9569. <https://doi.org/10.1073/pnas.0610537104>
- Pascual-García, A. (2020). Data from: apascualgarcia/functionInk: Deposited zenodo version (Version v1.0.0-zenodo). *Zenodo*, <https://doi.org/10.5281/zenodo.3653947>
- Pascual-García, A., Abia, D., Ortiz, Á. R., & Bastolla, U. (2009). Cross-over between discrete and continuous protein structure space: Insights into automatic classification and networks of protein structures. *PLoS Computational Biology*, 5(3), e1000331. <https://doi.org/10.1371/journal.pcbi.1000331>
- Pascual-García, A., & Bastolla, U. (2017). Mutualism supports biodiversity when the direct competition is weak. *Nature Communications*, 8, 14326. <https://doi.org/10.1038/ncomms14326>
- Pascual-García, A., & Bell, T. (2019a). Community-level signatures of ecological succession in natural bacterial communities. *bioRxiv*, 636233.
- Pascual-García, A. & Bell, T. (2019b). Data from: Input data and Supplementary Results for "Community-level signatures of ecological succession in natural bacterial communities". *Zenodo*, <https://doi.org/10.5281/zenodo.3539537>
- Pascual-García, A., Tamames, J., & Bastolla, U. (2014). Bacteria dialog with Santa Rosalia: Are aggregations of cosmopolitan bacteria mainly explained by habitat filtering or by ecological interactions? *BMC Microbiology*, 14(1), 284. <https://doi.org/10.1186/s12866-014-0284-5>
- Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances*, 3(5), e1602548. <https://doi.org/10.1126/sciadv.1602548>
- Pilosof, S., Porter, M. A., Pascual, M., & Kéfi, S. (2017). The multilayer nature of ecological networks. *Nature Ecology & Evolution*, 1(4), 0101. <https://doi.org/10.1038/s41559-017-0101>
- Rivett, D. W., & Bell, T. (2018). Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature Microbiology*, 3(7), 767–772. <https://doi.org/10.1038/s41564-018-0180-0>
- Schaub, M. T., Delvenne, J.-C., Rosvall, M., & Lambiotte, R. (2017). The many facets of community detection in complex networks. *Applied Network Science*, 2(1), 4. <https://doi.org/10.1007/s41109-017-0023-6>
- Shotwell, M. S. (2013). profdpm: An r package for map estimation in a class of conjugate product partition models. *Journal of Statistical Software*, 53(8), 1–18.
- Simberloff, D., & Dayan, T. (1991). The guild concept and the structure of ecological communities. *Annual Review of Ecology and Systematics*, 22(1), 115–143. <https://doi.org/10.1146/annurev.es.22.110191.000555>
- Tanimoto, T. T. (1958). *Elementary mathematical theory of classification and prediction*. IBM Internal Report.
- Valles-Catala, T., Massucci, F. A., Guimera, R., & Sales-Pardo, M. (2016). Multilayer stochastic block models reveal the multilayer structure of complex networks. *Physical Review X*, 6(1), 011036. <https://doi.org/10.1103/PhysRevX.6.011036>
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge, UK: Cambridge University Press.
- Yodzis, P., & Winemiller, K. O. (1999). In search of operational trophospecies in a tropical aquatic food web. *Oikos*, 87(2), 327–340. <https://doi.org/10.2307/3546748>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Pascual-García A, Bell T. functionInk: An efficient method to detect functional groups in multidimensional networks reveals the hidden structure of ecological communities. *Methods Ecol Evol*. 2020;00:1–14. <https://doi.org/10.1111/2041-210X.13377>