

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

May 31, 2019

Alberto Pascual-García^{1, 2,*} and Thomas Bell¹

(1) Department of Life Sciences, Silwood Park Campus, Imperial College London, Ascot, United Kingdom

(2) Current address: Institute of Integrative Biology, ETH-Zürich, Zürich, Switzerland

(*) Correspondence: alberto.pascual.garcia@gmail.com.

Abstract

Complex networks have been useful to link experimental data with mechanistic models, becoming widely used in modern science. More recently, the increasing amount and complexity of data, in particular in biology, prompted the development of multidimensional networks, where dimensions reflect the multiple qualitative properties of nodes, links, or both, classifying them into types. 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 reduced the complexity of the networks, hence facilitating its interpretation. The two traditional approximations to this problem come either through the identification of *modules* (communities more densely connected within their members than with nodes belonging to other communities) or of structurally equivalent communities (sets of nodes connected with the same neighbours, even if they are not connected themselves), that we call *guilds*. The relevance of this distinction is notable in biology, where we aim to differentiate between trophic levels, guilds, or concepts such as functional groups or ecotypes. In this work, we argue that structural equivalence represents an appropriate definition of the function of a node in the network, and we exploit this fact to show that it is possible to detect modules and guilds that, in this way, can be understood as different kinds of functional groups. We call our method functionInk (functional linkage), a method capable of objectively finding simultaneously both modules and guilds, and to determine which is the most relevant kind of functional group for a given network. Notably, it is computationally efficient handling large multidimensional networks, since it does not require an optimization procedure nor tests for robustness. The method is freely available at: [HTTPS://GITHUB.COM/APASCUALGARCIA/FUNCTIONINK](https://github.com/APASCUALGARCIA/FUNCTIONINK).

1 Introduction

1 Networks have played an important role in the development of ideas in ecology, particularly in understanding food
2 webs [1], and flows of energy and matter in ecosystems [2]. However, modern ecological datasets are becoming
3 increasingly complex, notably within microbial ecology, where multiple types of information (taxonomy, behaviour,
4 metabolic capacity, traits) on thousands of taxa can be gathered. A single network might therefore need to
5 integrate different sources of information, leading to connections between nodes representing relationships of
6 different types, and hence with different meanings. Advances in network theory have attempted to develop tools
7 to analyses these more sophisticated networks, encompassing ideas such as multiplex, multilayer, multivariate
8 networks, reviewed in [3]. There could therefore be much value in extending complex networks tools to ecology
9 in order to embrace these new concepts.

10 Broadly speaking, a network represents how a large set of entities *share* or *transmit* information. This definition
11 is intentionally empty-of-content to illustrate the challenges we face in network analysis. For instance, a network
12 in which information is shared may be built relating genes connected if their sequence similarity is higher than
13 certain threshold. In that case, we may capture how their similarity diverged after an evolutionary event such as
14 a gene duplication. On the other hand, networks may describe how information is transmitted, as in an ecosystem
15 in which we represent how biomass flows through the trophic levels or how behavioural signals are transmitted
16 among individuals. We aim to illustrate with these examples that, when building networks that consider links of
17 different nature (e.g. shared vs. transmitted information) or different physical units (e.g. biomass vs. bits) care
18 should be taken in extrapolating methods from single-dimensional to multidimensional complex networks.

19 A particularly relevant problem in mutiplex networks is the detection of “communities”, which are defined in
20 network theory as being sets of nodes sharing similar topological properties. Perhaps the most widely adopted
21 definition of community is the one considering sets of nodes more densely connected within the community than
22 with respect to other communities, often called *modules* [4]. Strategies to detect modules explore trade-offs in
23 quantities like the betweenness and the clustering coefficient [4], as in the celebrated Newman-Girvan algorithm
24 [5]. Generalizing traditional quantities like the clustering coefficient to multidimensional networks is difficult.
25 Consider, for instance, that a node A is linked with a node B and this is, in turn, linked with a node C, being
26 both links of a certain type. If A is then linked with node C with a different type of link, should the triangle ABC
27 considered in the computation of the clustering coefficient?

28 An approximation to such generalization was the derivation of the Newman-Girvan modularity definition,
29 which considers the dynamics of the flux of information in the network (in a statistical sense) through a Laplacian
30 dynamics [6]. This method was extended to consider multilayer networks [7] but it is unclear whether it can be
31 extended to networks considering *any* type of edges. It would likely be possible if, for the flux of information of
32 interest, the different types of edges have a clear meaning on how their presence affect informational fluxes. If the
33 links types are, however, qualitatively very distinct and have no explicit relationship with the fluxes dynamics,
34 the application of a dynamical model would likely be meaningless.

35 Another alternative is given by the notion of structural equivalence. Two nodes are said to be structurally
36 equivalent if they share the same number (and type, if the network is multidimensional) of links. This framework
37 was originally developed for social systems [8], where the role of the nodes (social agents) is important, and it
38 is encoded in the nodes’ interactions. This notion also goes hand in hand with the Eltonian classic definition of
39 niche, which emphasizes species function rather than species habitat [9]. We note that, since two nodes can be
40 structurally equivalent even if they are not connected themselves, communities determined under this definition
41 may be quite different to modules, in which members of the same module are tightly connected by definition. This
42 latter community structure is known as disassortative mixing [10], and has received comparatively less attention
43 than the “assortative” situation, leading to modules, perhaps with the exception of bipartite networks [11, 12].
44 We call to this second class of communities guilds, inspired in the ecological meaning in which species may share
45 similar ways of exploiting resources (i.e. similar links) without necessarily sharing the same niche (not being
46 connected themselves), emphasizing the functional role of the species [13]. Nevertheless, we note that the nodes
47 within modules are also structurally equivalent. Therefore, the notion of structural equivalence seems to open an
48 avenue to identify and distinguish both guilds and modules. Both types of communities can then be understood
49 as *functional groups* —in the eltonian sense— and this is the name we adopt here. We reserve in this way the term
50 community for a more generic use, since other types of communities beyond functional groups may exist.

51 In this work, we show that a modification of the community detection method developed by Ahn et al. [14],
52 leads to the identification of two quantities we call internal and external partition densities, which allow the
53 identification of modules and guilds, respectively. For a set of nodes joined within a community by means of their

54 structural equivalence similarity, the partition densities quantify whether their similarities come from connections
55 linking them with nodes outside the community (external density) or within the community (internal density).
56 One of the main challenges in this work is to determine the structural equivalence similarity threshold above
57 which nodes are considered to belong to the same community, a problem that, in the literature, led to definitions
58 such as regular equivalence [8], which proves to be problematic, as illustrated in Fig. 1. Notably, our method
59 brings absolute maximum values for internal and external partition densities, allowing us to objectively determine
60 optimal cut-offs for the structural equivalence similarities. In addition, the method has several advantages that
61 make it particularly suitable for analysing large networks. First, we generalized the method to consider an
62 arbitrary number of link types, which makes it suitable for the analysis of multidimensional networks. Second,
63 the method is deterministic, and hence it does not require costly optimality or tests for robustness [10], whose
64 improvement has attracted much attention in recent years [15].

65 We call our method `functionInk` (functional linkage), emphasizing the fact that the number and types of links
66 of a node determine its functional role in the network. We illustrate its use by considering complex biological
67 examples, for which we believe the notion of functional role is particularly relevant. For instance, modules may be
68 of interest if we aim to detect sets of species within the same trophic level, with competitive interactions within the
69 set and other types of interactions like prey-predator or mutualistic, between sets. In this case, the `functionInk`
70 method is able to accommodate the different types of interactions (predator-prey, competitive, mutualisms) in these
71 networks. In other circumstances, we may be interested in identifying biological entities having a similar function
72 despite not interacting themselves. This kind of community would be closer to the notion of guild than to the
73 concept of a trophic level [13]. We show in the examples that, combining the external and internal partition
74 densities, we are able to identify the underlying type of dominant structures of the network (either towards
75 modules or towards guilds). Moreover, selecting the most appropriate community definition in each situation
76 provides results that are comparable to state-of-the-art methods. This versatility in a single algorithm, together
77 with its low computational cost to handle large networks and its ability to work in multidimensional networks,
78 makes our method suitable for any type of complex, multidimensional network.

79 Results

80 Structural equivalence similarity in multidimensional networks

81 We modified and extended the method presented in Ahn et al. [14] to consider i) different types of links, where
82 types are classified according to their qualitative attributes and ii) the different expressions defined between nodes
83 instead of between links. The latter modification has several technical advantages. Most notably it allows us to
84 propose two quantities, the external and internal partition densities, which we use to identify guilds and modules,
85 respectively.

86 The method starts by considering a similarity measure between all pairs of nodes, that quantifies the number
87 and type of interactions they share, shown in Fig. 1. For simplicity, we present the derivation for two types of
88 undirected interactions (for instance positive, +, and negative, -), and its extension for an arbitrary number of
89 types is presented in Methods. We call $\{i\}$ the set of N nodes and $\{e_{ij}\}$ the set of M edges in a network. We call
90 $n(i)$ the set of neighbours of i , that can be split into different subsets according to the types of links present in the
91 network. We split the set of neighbours linked with the node i into those linked through positive relationships,
92 $n_+(i)$, or through negative relationships, $n_-(i)$ (see Fig. 1, we follow here a notation similar to the one presented
93 in [14], but note that $n(i)$ stands there for neighbours irrespective of the kind of edges).

94 This splitting of neighbours into different subsets according to their identity, is one of the modifications of the
95 original method. Distinguishing link types induces a division in the set of neighbours of a given node into subsets
96 sharing the same link type, shown in Fig. 2A. More specifically, in the absence of link types we define the Jaccard
97 similarity between to nodes i and j as:

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

98 where $|\cdot|$ is the cardinality of the set (the number of elements it contains). Generalizing this expression to two
99 attributes (see Suppl. Methods for an arbitrary number of attributes) leads to

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

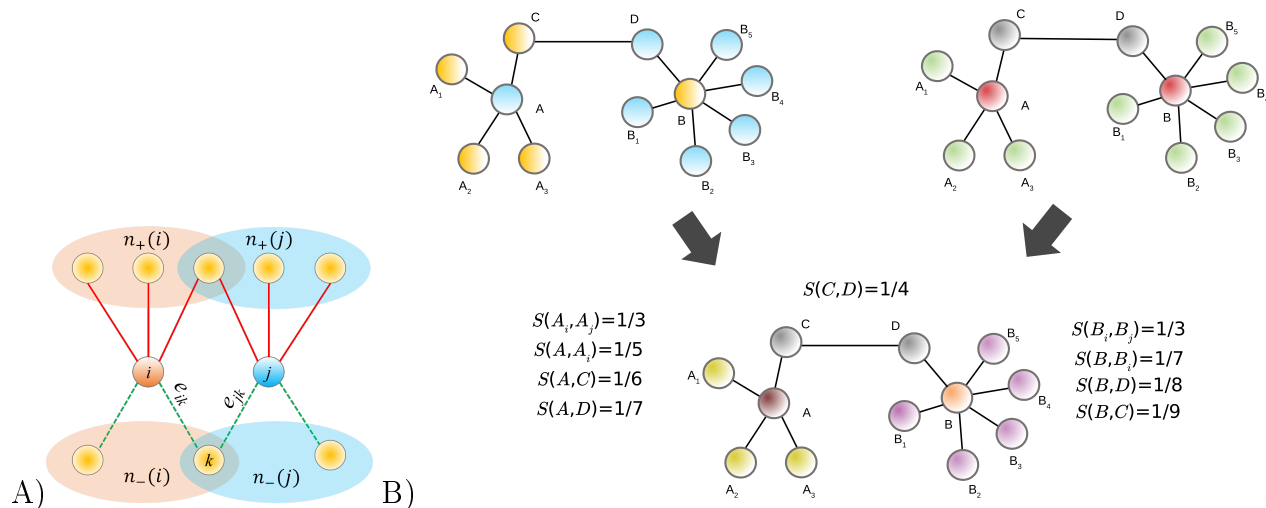


Figure 1: Illustration of the method. (A) The similarity between nodes i and j is computed considering the neighbours that each node has and the types of interactions linking them. In this example, continuous links stand for positive interactions, determining the set of neighbours $n_+(i)$ and $n_+(j)$, and negative ones are shown with dotted links connecting the sets $n_-(i)$ and $n_-(j)$. In Ref. [14] the similarity computed in this way would be assigned to the links e_{ik} and e_{jk} . (B) Under the blockmodelling approach, nodes that are regularly structurally equivalent are classified in two communities (blue and yellow, top-left network). The method of Guimerá and Amaral determines communities through their topological role (top-right network) having central networks (A and B nodes), peripheral (A₁-A₃ and B₁-B₅) and connectors (C and D). functionInk (bottom network) defines communities joining nodes sharing approximately the same number and type of links. All non-zero Jaccard similarities $S^{(j)}(n(i), n(j))$ of the example are shown. Clustering these similarities will lead to different partitions and, stopping at $S^{(j)} = 1/4$, would lead to communities being the intersection of those found in the above networks, highlighting the potential to identify communities whose nodes are joined through a local notion of function that encapsulates more global topological features. Figure adapted from [16].

Accounting for the weight of the edges can be made with the generalization of the Jaccard index provided by the Tanimoto coefficient [17], $S^{(T)}(i, j)$, presented in Methods.

We note the particular case in which i and j are only connected between them which, with the above definition, means that they do not share any neighbours. 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 $S(i, j) = 0$. On the other hand, if they share a connection with respect to a third node, we want to distinguish the situation in which all three nodes are connected (a perfect transitive motif, the archetype of a module) from a situation in which they only share a neighbour but they are not themselves connected (the archetype of a guild). If we consider that a node is always a neighbour of itself, i.e. $\{i\} \in n(i)$, in both situations $S(i, j) = 1$ and we cannot distinguish these cases. Therefore, we take the convention that, for a connection between two nodes $|n(i) \cap n(j)| = 1$ and $|n(i) \cup n(j)| = 2$. In Fig. 1 we illustrate the computation of this similarity with a simple example.

Identification of optimal similarity cut-offs

Once the similarity between nodes is computed, the next objective is to cluster the nodes using a similarity measure in order to identify communities (see Methods). Clustering is performed in a stepwise manner, where nodes that are increasingly dissimilar in their links are iteratively partitioned into communities. A critical question in clustering procedures is to identify the clustering step for which the optimal partition is achieved [18]. This question is often addressed by proposing a measure that monitors the clustering and that has a well defined maximum or minimum determining a threshold to stop the clustering. In [14], they proposed a quantity called the partition density, whose maxima determines the optimal clustering threshold (that we recover for completeness in Methods). We reconsider the definition of partition density because it was originally defined over edge partitions. We develop a similar measure that defines partition densities across nodes, and which adds a new dimension to the investigation of node partitioning. To develop this measure, we note that, when we join nodes into a partition, we are concluding that these nodes approximately share the same number and type of connections, but we actually do not know whether they are connected between themselves or not. We therefore redefine the partition density differentiating between the contribution of links density arising from the connections *within* the community from connections shared with respect to external nodes, i.e. *between* partitions.

Given a node i , we differentiate between those neighbours that are within the same community to which the node belongs, that we call $n^{\text{int}}(i)$ (where int stands for “interior”), and those in the exterior of the community, $n^{\text{ext}}(i)$, hence $n(i) = n^{\text{int}}(i) \cup n^{\text{ext}}(i)$ (See Fig. 2). For a singleton (a cluster of size one) $n^{\text{int}}(i) = \{i\}$ and $n^{\text{ext}}(i) = \emptyset$. Similarly, the set of edges $m(i)$ linking the node i with other nodes can be also split into two sets: the set linking the node with neighbours within its community $m^{\text{int}}(i)$, and those linking it with external nodes $m^{\text{ext}}(i)$.

Given a partition of nodes into T communities, our method identifies, for each community, the total number of nodes within it, 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 with connections to the nodes in the community C (n_c^{ext}) through a number of links m_c^{ext} . Clearly, to identify n_c^{ext} neighbours requires at least n_c^{ext} links, and the number of links in excess, $m_c^{\text{ext}} - n_c^{\text{ext}}$, contributes to the similarity of the nodes in the community through external links. Therefore, we quantified the fraction of links in excess out of the total possible number $(m_c^{\text{ext}} - n_c^{\text{ext}})/n_c^{\text{ext}} (n_c^{\text{int}} - 1)$. 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 edges. We follow now a similar reasoning now considering the internal links, but we should acknowledge that in a community created linking 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:

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

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

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

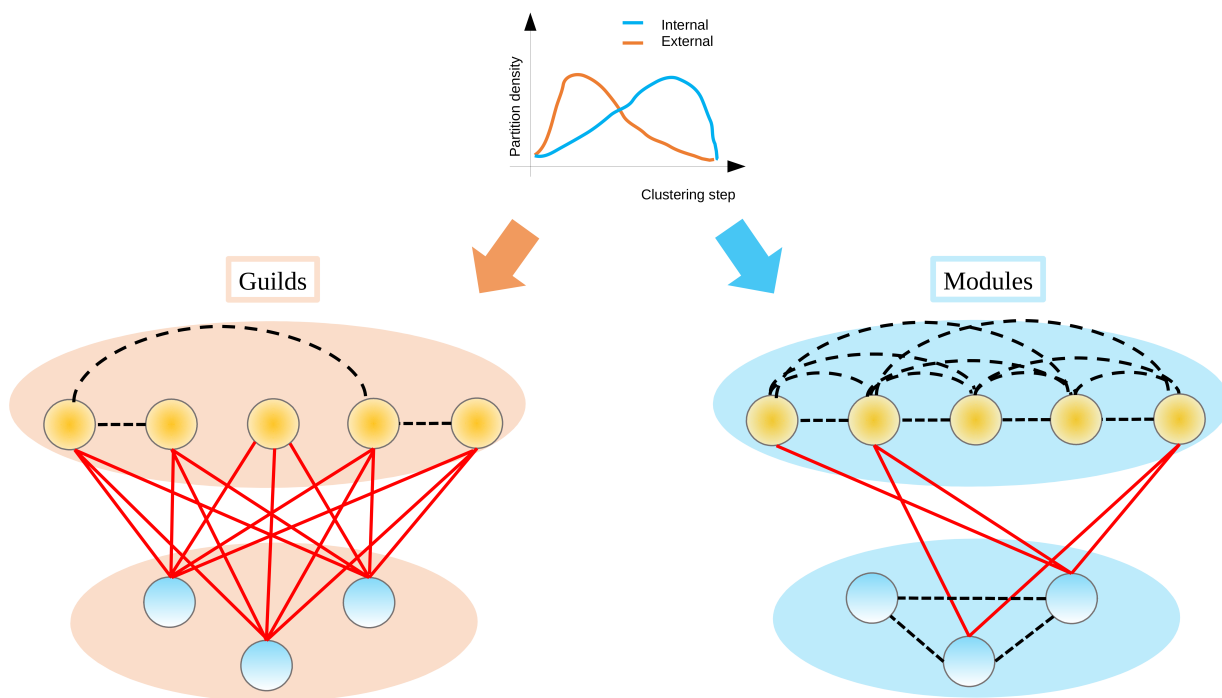


Figure 2: **Definition of guilds and modules.** For each set of nodes belonging to the same community c (n_c^{int} , nodes within the same shaded area) we consider the number of links within the community (black dashed links, called m_c^{int} in the main text) out of the total number of links of the community, to compute the internal partition density. We identify modules as those communities that maximize the internal partition density (right figure). We then compute the density of links connecting nodes external to the community (m_c^{ext} , solid red lines linking nodes in other partitions) out of the total to estimate the external partition density. We call guilds to those communities obtained maximizing this quantity.

144 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,
145 D^{total} equals to one. Since at the beginning of the clustering the communities have a low number of members,
146 most of the contribution towards D^{total} comes from D^{ext} while, in the last steps where the communities become
147 large, D^{int} will dominate. All three quantities will reach an absolute value along the clustering and, if one of them
148 clearly achieves a higher value, it will be indicative that one kind of functional group is dominant in the network.
149 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
150 close to one of these. If neither D^{ext} nor D^{int} clearly dominates, D^{total} will peak at an intermediate step between
151 the two partial partition densities, suggesting that this intermediate step is the best candidate of the optimal
152 partition for the network, and communities determined at this intermediate point will be called, generically,
153 functional groups.

154 Plant-pollinator networks

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

160 To investigate the performance of our method and, in particular, the influence of the topological properties
161 into the partition density measures, we generated a set of artificial mutualistic networks with diverse topological
162 properties, following the method presented in [19]. For the mutualistic interactions, we focused on two properties:
163 the connectance κ_{mut} , which is the fraction of observed interactions out of the total number of possible interactions,
164 and the nestedness ν as defined in [20] (see Methods), which codifies the fraction of interactions that are
165 shared between two species, averaged over all pairs of species. We selected these measures for their importance
166 in the stability-complexity debate in mutualistic systems [19], and the similarity between thenestedness (which,
167 in the definition we adopt here, represents the mean ecological overlap between species) and the notion of struc-
168 tural equivalence we considered. For the competition matrices, we considered random matrices with different
169 connectances, κ_{comp} , since it is difficult to estimate direct pairwise competitive interactions experimentally, and
170 it is frequently modelled with a mean field competition matrix.

171 In all networks, the set of plants and animals are joined in the very last step of the clustering irrespective of
172 the clustering method used, indicating that our similarity measure is appropriate and that the method is robust
173 with respect to the clustering method selected. As expected, the curves monitoring the external and internal
174 partition densities depends on the properties of the networks. We illustrate this finding in Fig. 3, where we have
175 selected two networks with contrasting topological properties. One of the networks has high connectance within
176 the pools and low connectance and nestedness between the pools. The internal partition density peaks at the
177 last step minus one (i.e. where the two pools are perfectly separated) consistent with the definition of modules,
178 where the intra-modules link density is higher than the inter-modules link density. On the other hand, the second
179 network has intra-pool connectance equals to zero, and very high connectance and nestedness between the pools
180 (see Fig. 3). We selected a $\kappa_{\text{comp}} = 0$ for simplicity in the network representation, but similar results are obtained
181 for low values of κ_{comp} , see for instance Suppl. Fig. 10. In this second network (see Fig. 3, right panel), only
182 the external partition density peaks and, at the maximum, the communities that we identified clearly reflect the
183 structural equivalence of the nodes members in terms of their connectance with nodes external to the group,
184 as we expect for the definition of guilds. The ecological information retrieved for guilds is clearly distinct from
185 the information retrieved for the modules, being the former related to the topology of the network connecting
186 plants and animals. We observe that guilds identify specialist species clustered together, which are then linked
187 to generalists species of the other pool: a structure typical of networks with high nestedness.

188 The method identified several interesting guilds and connections between them. For instance, generalists
189 Plant 1, Animal 1 and Animal 2 (and to a lesser extent Plant 2) have a low connectivity between them but, being
190 connected to many specialists, determine a region of high vulnerability, in the sense that a directed perturbation
191 over these species would have consequences for many other species. This is confirmed by the high betweenness of
192 these nodes (proportional to the size of the node in the network). In addition, the algorithm is able to identify
193 more complex partitions of nodes into clusters. As an example of this, Animal 16 (torquoise) is split from
194 Animals 10 and 11 (cyan), which form a second cluster, and from Animals 15, 18 and 19 (light pink) that are
195 joined into a third cluster, despite of the subtle connectivity differences between these six nodes. Finally, it also
196 detects clusters of three or more species that have complex connectivity patterns which, in this context, may be

197 indicative of functionally redundant species (e.g. red and blue clusters).

198 Examples with other intermediate properties are analyzed in the Suppl. Figs. 8 and 9. Broadly speaking,
 199 either the internal or the total partition density maximum peaks at the last step minus one, allowing for detection
 200 of the two pools of species. Nevertheless, the method fails to find these pools if the within-pools connectance
 201 is very low, since the network becomes highly disassortative (see Suppl. Fig. 10). The relative magnitude of
 202 the external vs. internal partition density depends on the connectance between the pools of plants and animals
 203 and on the connectance within the pools, respectively (see Suppl. Fig. 8). Interestingly, networks for which the
 204 nestedness is increased being the remaining properties the same, generated an increase in the external partition
 205 density (see Suppl. Fig. 9). These examples illustrate how the external partition density is sensitive to complex
 206 topological properties, in particular to an increase in the disassortativity of the network.

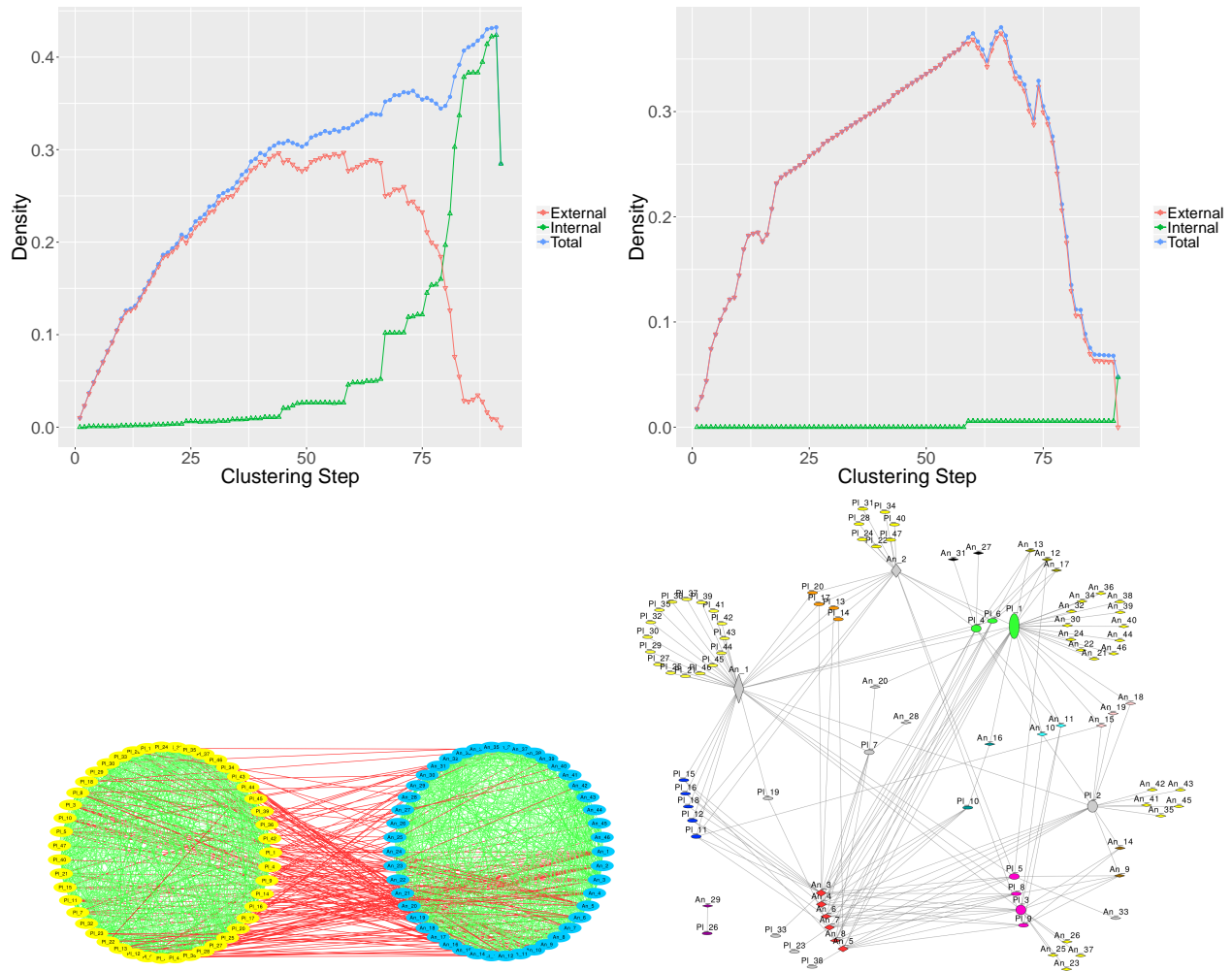


Figure 3: **Synthetic mutualistic networks.** (Top left) Partition densities for a network with $\kappa_{\text{comp}} = 0.5$, nestedness $\nu = 0.15$ and $\kappa_{\text{mut}} = 0.08$ and (top right) for a network with $\kappa_{\text{comp}} = 0$, nestedness $\nu = 0.6$ and $\kappa_{\text{mut}} = 0.08$. 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). In the networks, plants are labelled “PI” and animals are labelled “An”. Nodes are coloured according to their functional group. In the network finding guilds (bottom right), specialist species are yellow, single species clusters are grey, and the size of the nodes is proportional to their betweenness.

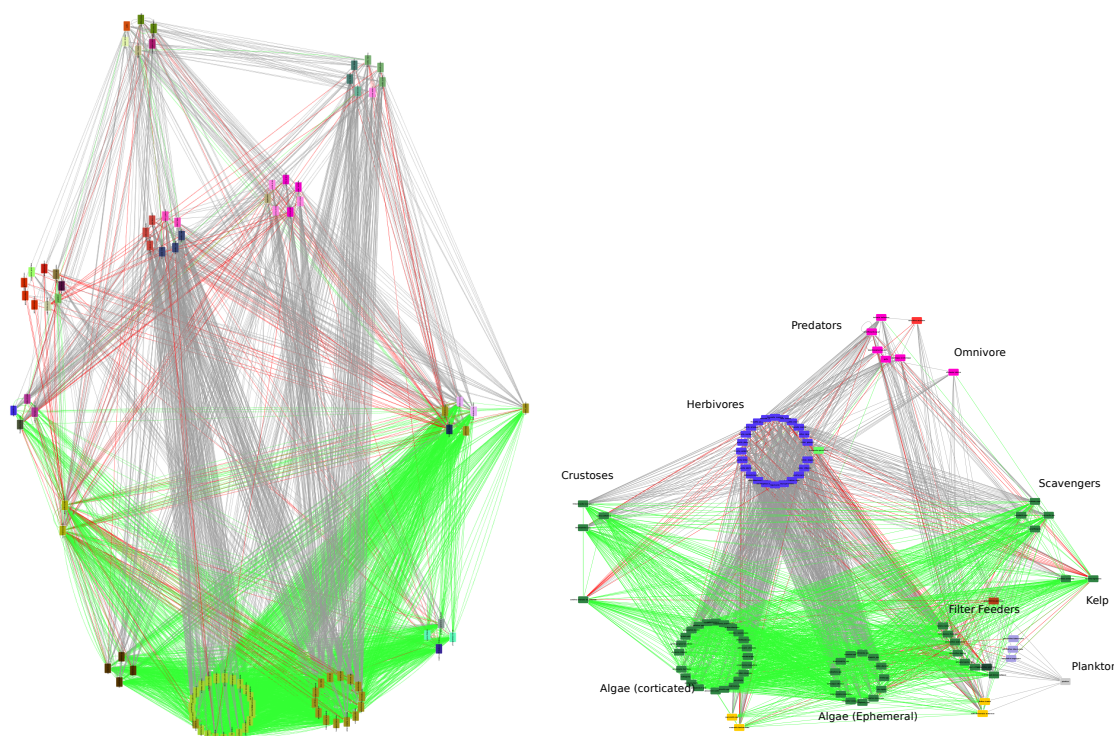


Figure 4: **Analysis of a trophic network.** Trophic networks with links representing trophic (grey), non-trophic positive (red), and negative (green) interactions. (Left) Nodes are grouped according to the classification found in [21] (reference classification), and coloured by the guilds found with 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 levels, predators, herbivores and basal species, although it separates these into finer subgroups (filter feeders) and plankton, which is an orphan module.

207 1.1 Trophic networks

208 We tested our method in a comprehensive multidimensional ecological network of 106 species distributed in trophic
209 layers with approximately 4500 interactions, comprising trophic and non-trophic interactions (approximately 1/3
210 of the interactions are trophic) [21]. This network was analysed looking for communities extending a stochastic
211 blockmodelling method [10] to deal with different types of interactions [21]. The estimation of the parameters of
212 the model through an Expectation-Maximization algorithm requires a heuristic approximation, and hence it is
213 needed to test the robustness of the results found. Here we show that, in this example, our method is comparable
214 with this approximation, and it has the advantage of being deterministic. Moreover, the simplicity of the method
215 allows us to handle large networks with arbitrary number of types of links and to evaluate and interpret the
216 results, as we show in the following.

217 Our method finds a maximum for the internal density when there are only three clusters. Previous descriptions
218 of the network identified three trophic levels in the network (Predators, Herbivores and Basal species). The latter
219 are further subdivided into subgroups like (e.g. Kelps, Filter feeders), and there are some isolated groups like
220 one Omnivore and Plankton. To match these subgroups we observed that the total partition density reaches
221 a maximum close to the maximum of the external partition density (step 69) and maintains this value along a
222 plateau until step 95 (see Suppl. Fig. 11). We analysed results at both clustering thresholds finding that, at step
223 95, we obtain modules with a good agreement with the trophic levels, shown in Fig. 4. On the other hand, at step
224 69 we find a larger number of communities some of which fit the definition of modules and others the definition
225 of guilds.

226 To shed some light on the information obtained from this second network, we compared the classification
227 obtained by Kefi et al. [21] (in the following reference classification) and our method, shown in Fig. 4. We
228 computed a number of similarity metrics between the classification we obtained at each step of the agglomerative

229 clustering with functionInk and the reference classification (see Methods). In Fig. 5, we show that the similarity
230 between both classifications is highly significant (Z-score > 2.5) and is maximized when the external partition
231 density is also maximized, i.e. at step 69. This is particularly apparent for the Wallace 01, Wallace 10 and Rand
232 indexes (see Fig. 7 and Suppl. Fig. 12). Clusters in the reference classification were also interpreted as functional
233 groups in the same sense proposed here [21], supporting our arguments to use the external partition density as
234 a quantity to detect guilds.

235 Nevertheless, there are some discrepancies between both classifications. In particular, although there is a com-
236 plete correspondence between the two largest clusters in both classifications, there are a number of intermediate
237 clusters in the reference classification whose members are classified differently in our method. To illustrate these
238 discrepancies, we plotted a heatmap of the Tanimoto coefficients of members of four clusters of intermediate size
239 containing discrepancies, showing their membership in both the reference and the functionInk classification with
240 different colours (see Fig. 4). The dendrograms cluster rows and columns computing the Euclidean distance be-
241 tween their values. Therefore, this illustration is very similar to functionInk, and the clusters must be consistent,
242 representing a powerful way to visually inspect results. Indeed, the blocks found in the heatmap are in correspon-
243 dence with functionInk clusters, as expected, but we observe some discrepancies with the reference classification.
244 For instance, the cluster found by the reference classification containing several *Petrolishtes* species, joins species
245 that have low similarity regarding the number and type of interactions as measured by the Tanimoto coefficients,
246 while functionInk joins together the three species with high similarity, leaving aside the remainder species. Of
247 course, we cannot discard that the method used in the reference classification captures other properties justifying
248 the differences. But it is immediately apparent the advantages provided by the simplicity our method, which
249 permits validation through visual inspection of the consistency of the classification.

250 1.2 Microbial networks

251 We discuss a last example of increasing importance in current ecological research, which is the inference of
252 interactions among microbes sampled from natural environments. We considered a large matrix with more than
253 700 samples of 16S rRNA operative taxonomic units (OTUs) collected from rain pools (water-filled tree-holes)
254 in the UK [23, 24] (see Suppl. Methods). We analysed β -diversity similarity of the communities contained in
255 the matrix with the Jensen-Shannon divergence metric [25], further classifying the communities automatically,
256 leading to 6 disjoint clusters (see Methods). Next, we inferred a network of significant positive (co-occurrences)
257 or negative (segregations) correlations between OTUs using SparCC [26], represented in Fig. 6 (see Methods).
258 Applying functionInk to the network of inferred correlations, we aimed to understand the consistency between
259 the results of functionInk (modules and guilds) and the β -diversity-classes. The rationale is that, by symmetry,
260 significant co-occurrences and segregations between OTUs should reflect the similarity and dissimilarity between
261 the communities, hence validating the method.

262 Contrasting with the trophic network analysed in the previous example, the external partition density achieves
263 a low relative value and brings a poor reduction of the complexity of the network, suggesting optimal clustering
264 after just 22 clustering steps (see Suppl. Fig. 13). The internal partition density achieves a higher value, hence
265 suggesting that modules are more relevant than functional groups in this network. Two large modules are apparent,
266 see Fig. 6, with a large number of intracluster co-occurrences (continuous links) and interclusters segregations
267 (dotted links). Note that this is quite different to what is found in macroscopic trophic networks, where pools of
268 species (e.g. prey) have within module competitive (segregating) interactions, while between-modules interactions
269 can be positive (for predators) or negative (for prey). In addition, the total partition density peaks at a much
270 higher value and it seems to be able to split some of the large modules into smaller motifs, some of which were
271 identified as guilds and have clearly distinctive connectivities, that we analyse in further detail. Since some of
272 these motifs have characteristics closer to those of guilds while others are closer to modules, we refer to them
273 generically as functional groups.

274 There is reasonable agreement between the functional groups found at the maximum of the total partition
275 density and the β -diversity-classes shown in Fig. 6. Since it was shown in [23] that the β -diversity-classes might
276 be related to a process of ecological succession driven by environmental variation, the functional groups are likely
277 driven by environmental preferences rather than by ecological interactions, likely explaining the large number of
278 positive co-occurrences. This speaks against a naïve interpretation of correlation networks in microbial samples as
279 ecological interactions, unless environmental preferences are controlled [27]. However, the detection of networks
280 complements the information that β -diversity-classes provides, since it is possible to individuate the key players
281 of these classes. For instance, only two OTUs from the green functional group have an important number of

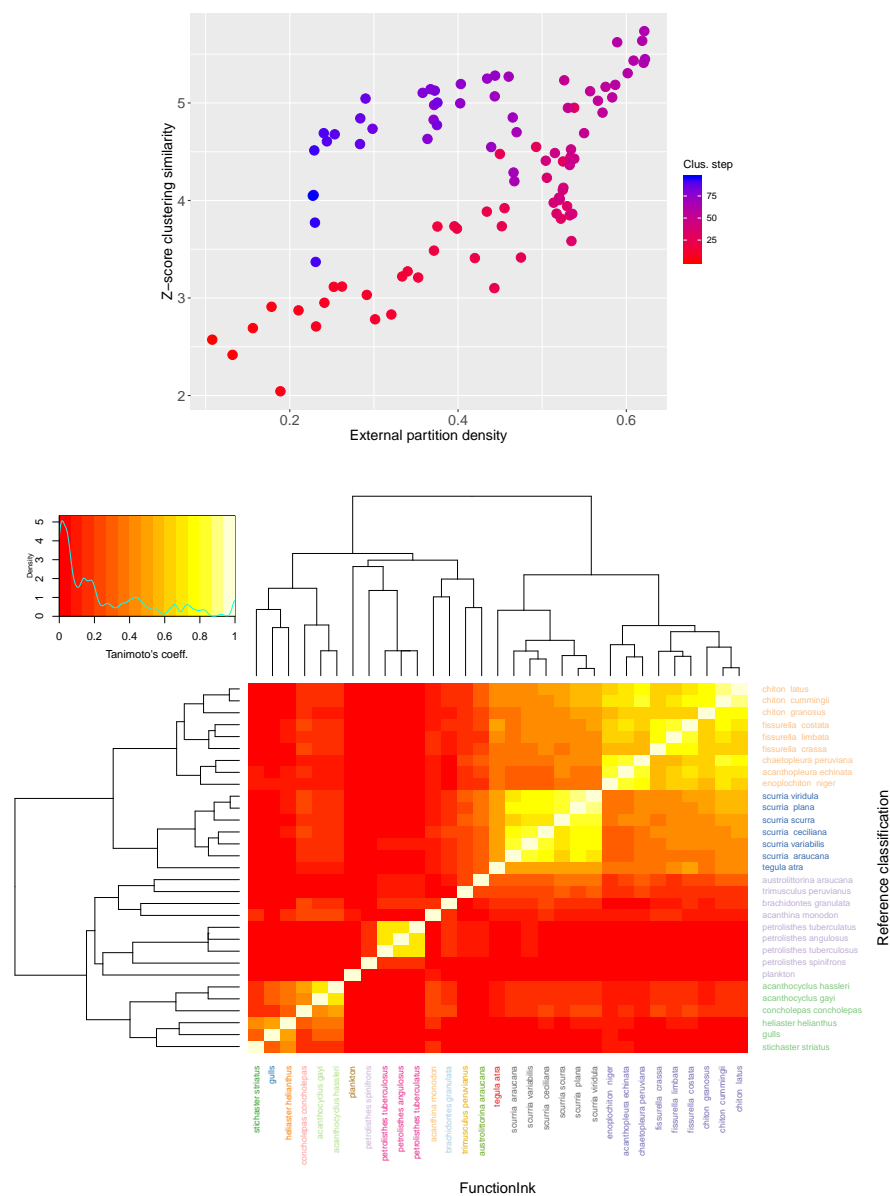


Figure 5: **Analysis of guilds.** (Top) Z-score of the Wallace 10 index [22], measuring the similarity between the reference classification and the functionInk method at each clustering step. The similarity with the reference classification is maximized around the maximum of the external partition density. (Bottom) Comparison of clusters 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 cluster membership in the functionInk (reference) classifications. The heatmap represents the values of the Tanimoto coefficients, and the dendrograms are computed using Euclidean distance and clustered with complete linkage. The heatmap blocks of high similarity are in some cases inconsistent with the reference classification. Given the simplicity of the interpretation of these coefficients, it is difficult to explain the clustering of distant members by the reference classification.

282 co-occurrences with members of the red functional group, and only one of them has a significant segregation
283 with respect to a member of the blue functional groups. These two highly abundant segregating OTUs are
284 *Pseudomonas putida* and *Serratia fonticola*, both of which were shown to dominate two of the β -diversity-classes
285 [23]. The functional groups hence allow us to easily identify their most important partners and, more in general,
286 to analyse in detail how clusters of communities are structured.

287 Discussion

288 We presented a novel method for the analysis of multidimensional networks, with nodes containing an arbitrary
289 number of link types. The method extends and generalizes the method proposed by Ahn et al. [14] and presents
290 a number of advantages with respect to other approximations. First, we developed the method to work with
291 nodes instead of with links (which was the case for the original method [14]), which we find more intuitive,
292 and allows the interpretation of the communities and its analysis with current visualization software. From an
293 ecological perspective, we were also interested in the functional role of the nodes. In this sense, the definition of
294 a species function in the network is straightforward for nodes, adopting the definition of structural equivalence
295 used in social networks. This notion underlies both the similarity measure definition and the rationale behind
296 both the clustering and our definition of external partition density. Working with this definition we defined two
297 measures of nodes partitioning. While the internal partition density is very similar to the definition provided in
298 [14] (see Suppl. Methods), the external partition density brings a new dimension, being similar in spirit to the
299 search of structural equivalent clusters in social networks [8]. This allowed us to propose a clear differentiation
300 between modules (determined by the maximum of the internal partition density) and guilds (determined by the
301 maximum of the external partition density). Although the method might not be able to achieve the generality
302 of approximations such as the mixture models proposed in [10], which aims to find any arbitrary structure in
303 the network, such approximations are far from being generalized to an arbitrary number of link types, as we
304 presented here. In addition, these are statistical approximations requiring either heuristics to find a solution for
305 the parameters –and hence a unique optimal solution is not guaranteed–, or a computationally costly sampling
306 of the parameter space. Our method relies on a deterministic method whose results are easily inspected, given
307 the simplicity of the similarity metric used and the partition density functions proposed to monitor the optimal
308 clustering.

309 Beyond these technical advantages, we illustrated the versatility of functionInk using several ecological ex-
310 amples. The relative value between the internal and external partition density, immediately yields information
311 on whether the network is dominated by modules, guilds, or intermediate structures. This allows for increasing
312 flexibility in the analysis of the networks, and for a more nuanced interpretation of network structure and species'
313 roles in the ecosystem. For both mutualistic and trophic networks, the internal partition density correctly finds
314 the trophic layers, justifying the success of the original method [14]. Our extension recovered the functional groups
315 as determined by Kefi et al. [21] through the external partition density, and the visual inspection reflects a good
316 consistency with the definition we proposed for functional groups in terms of structural equivalence. Moreover,
317 in the mutualistic networks, we showed that the functional groups discovered in this way was sensitive to changes
318 to high-order topological properties such as the nestedness.

319 The analysis of the microbial network was dominated by modules rather than guilds. Interestingly, these mod-
320 ules had intra-cluster positive correlations, contrary to what would be expected in a macroscopic trophic network,
321 where competitive interactions would be dominant between members of the same trophic layer. We selected in this
322 example for further exploration the functional communities found at the maximum of the total partition density,
323 with some groups having properties closer to those of guilds and others closer to modules. The communities that
324 we identified were in good agreement with the functional communities found using β -diversity similarity [23],
325 supporting the consistency of the method. Interestingly, it was found in [23] that similar β -diversity-classes were
326 driven by environmental conditions. Although co-occurring more often in the same environment may be indicative
327 of a higher probability of interaction [28], the most economical hypothesis is that they co-occur because they share
328 similar environmental preferences, and hence it cannot be disentangled the type of interaction (if any) unless the
329 environmental variables are under control.

330 functionInk requires the computation of Jaccard or Tanimoto coefficients, whose computational cost scales as
331 N^2 , being N the number of nodes. However, the similarity coefficients only need to be computed once, and then
332 the clustering with different methods and posterior analysis are at most order N , making the method suitable for
333 large networks. The method is available in the address ([HTTPS://GITHUB.COM/APASCUALGARCIA/FUNCTIONINK](https://github.com/apascualgarcia/functionink))
334 and, importantly, although we developed it with ecological networks in mind, it can be applied to any kind of

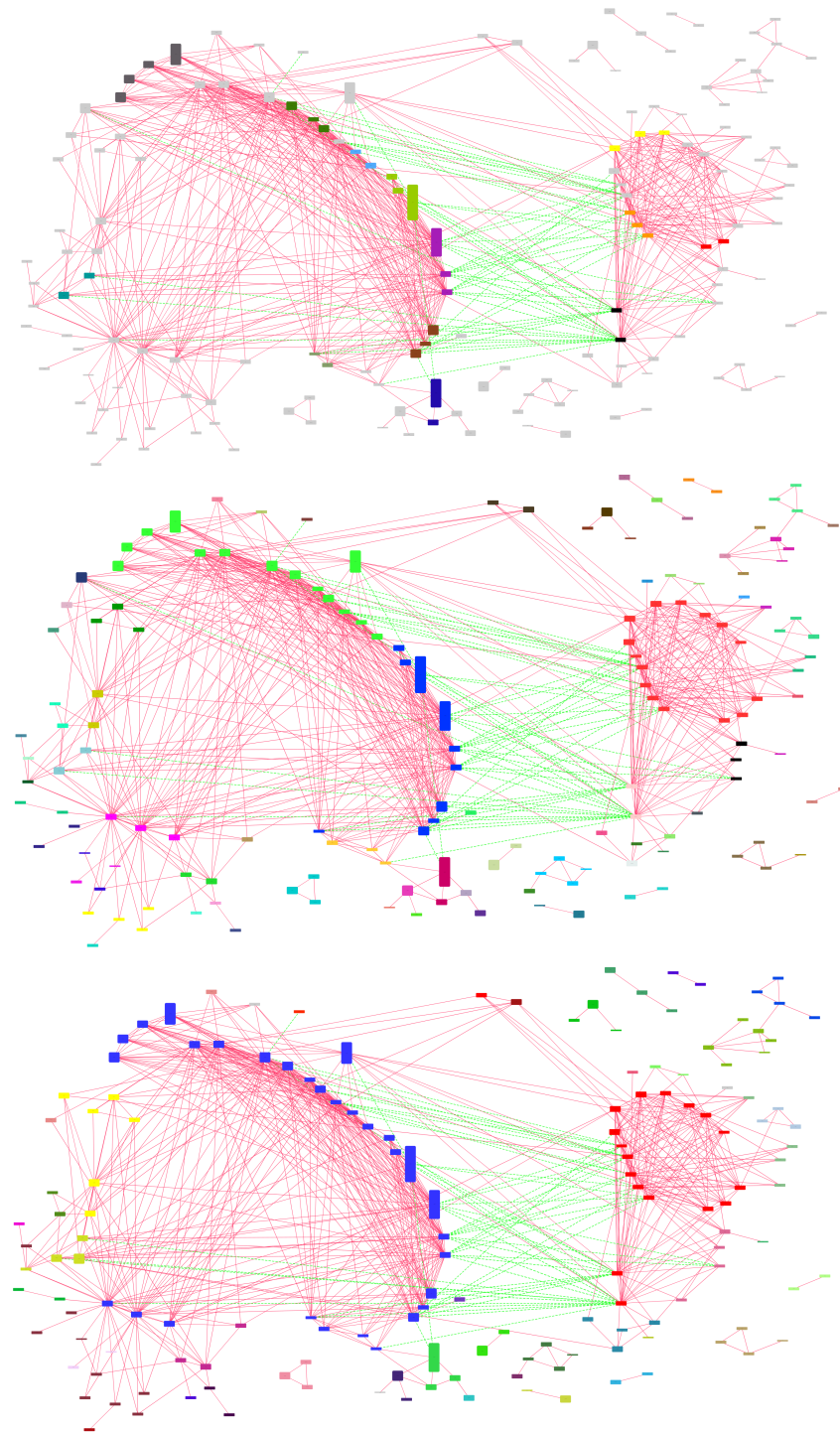


Figure 6: **Comparison of functional groups in the microbial network.** Network of significant co-occurrences (continuous links) and segregations (dotted links) at the species level (nodes). Colours indicate functional group membership, which was determined by the maximum of the external (top), total (middle) and internal partition densities (bottom). Orphan nodes are coloured grey in the top figure for clarity. The higher value of the internal partition density (see Suppl. Fig. 13) suggests that a modular structure is the more appropriate to describes the functional groups. This is confirmed by the low number of guilds (top figure) and the good agreement between the global topological structure and the modules (bottom figure).

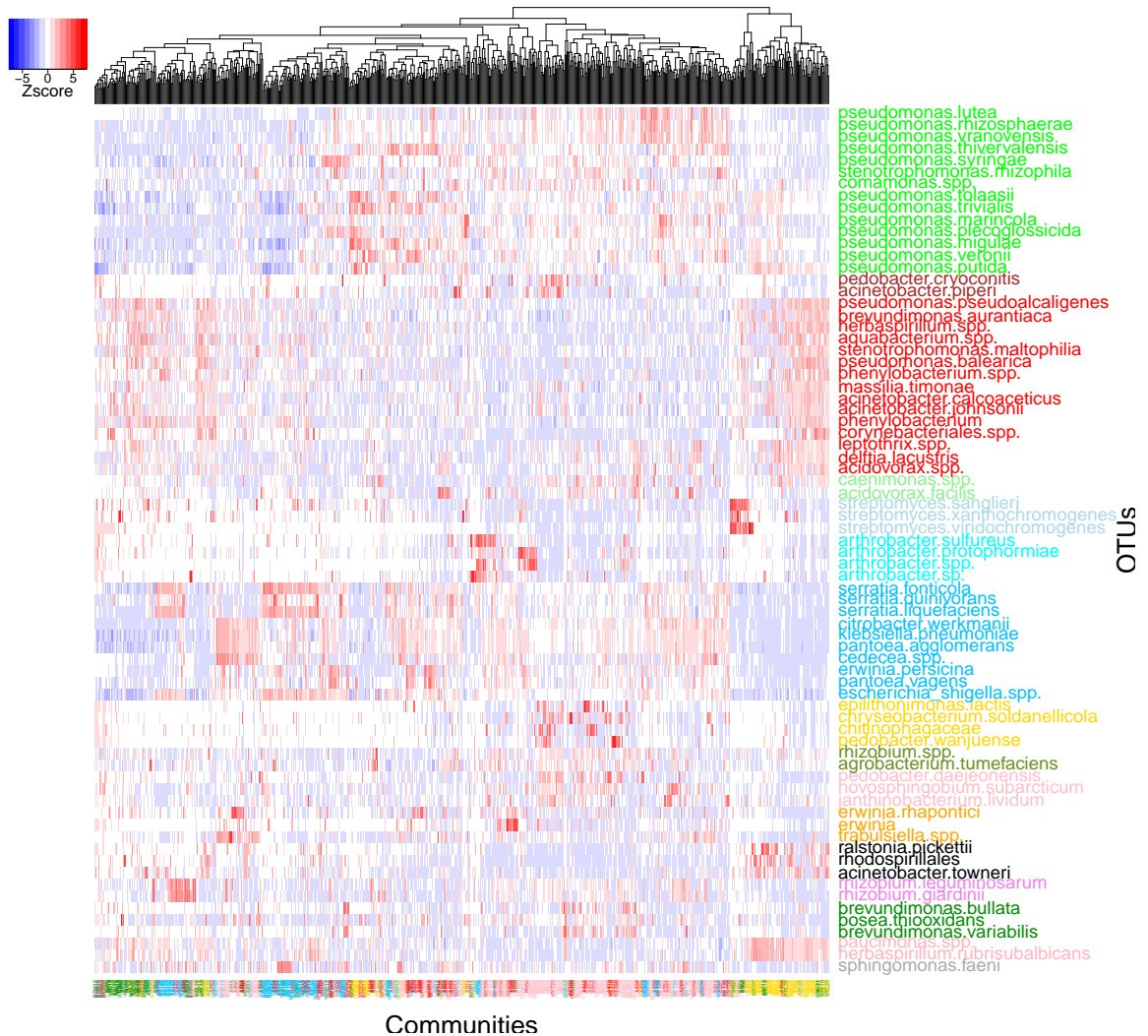


Figure 7: **Analysis of a microbial network.** Heatmap representing the z-score of the log-transformed abundances of the OTUs (see Methods). Species are coloured according to their functional group membership obtained at the maximum of the total partition density. Samples are coloured according to one of the six community classes found in [23]. Heatmap blocks show segregation and co-occurrence between modules, further mapping the β -diversity classes.

335 network.

336 Methods

337 Generalization of the Jaccard and Tanimoto coefficients to an arbitrary number of 338 link types

339 Consider a network with a set $\{i\}$ of N nodes and a set $\{e_{ij}\}$ of M links. These links are classified into Ω types
340 labelled with the index $\alpha = (1, \dots, \Omega)$. These types would typically account for differential qualitative responses
341 of the nodes properties due to the interactions. For example, if we consider that the nodes are species and
342 the property of interest is the species abundances, the effect of cooperative or competitive interactions on the
343 abundances can be codified using two different types of links: positive and negative. If these relations are inferred
344 through correlations between abundances, we could use a quantitative threshold (for instance a correlation equal
345 to zero) to split the links into positive and negative correlations. In general, we may use a number of qualitative
346 attributes or quantitative thresholds in the weights of the links to determine different types of links.

347 We call $n(i)$ the set of neighbours of i , and we split these neighbours into (at most) Ω different subsets
348 according to the types of links present in the network. The Jaccard coefficient defined in 2 can be extended
349 (already considering similarities between nodes) as:

$$S^{(J)}(n_i, n_j) = \frac{\bigcup_{\alpha=1}^{\Omega} |n_{\alpha}(i) \cap n_{\alpha}(j)|}{\bigcup_{\alpha=1}^{\Omega} |n_{\alpha}(i) \cup n_{\alpha}(j)|} \quad (5)$$

350 Accounting for the weight of the edges can be made with the generalization of the Jaccard index provided by
351 the Tanimoto coefficient [17]. We first introduce the method without differentiating between different types of
352 neighbours. Consider the vector $\mathbf{a}_i = (\tilde{A}_{i1}, \dots, \tilde{A}_{iN})$ with

$$\tilde{A}_{ij} = \frac{1}{k_i} \sum_{i' \in n(i)} w_{ii'} \delta_{ij} + w_{ij} \quad (6)$$

353 where w_{ij} is the weight of the edge linking the nodes i and j , $k_i = |n(i)|$ and δ_{ij} is the Kronecker's delta
354 ($\delta_{ij} = 1$ if $i \neq j$ and zero otherwise). Determining the quantity $W_{ij} = \mathbf{a}_i \mathbf{a}_j = \sum_k \tilde{A}_{ik} \tilde{A}_{kj}$, the Tanimoto similarity
355 is defined as

$$S^{(T)}(e_{ik}, e_{jk}) = \frac{W_{ij}}{W_{ii} + W_{jj} - W_{ij}} \quad (7)$$

356 Working with link types requires a generalization of the above expression. Consider for the moment two types
357 related with a positive $w_{ij} > 0$ or a negative $w_{ij} < 0$ weight of the links. The term $\tilde{A}_{ii} = 1/k_i \sum_{i'} w_{ii'}$ is the
358 average of the strengths of the edges connected with node i , and it is desirable to keep this meaning when
359 considering two types to properly normalize the Tanimoto similarity. This is simply achieved redefining \tilde{A}_{ij} as

$$\tilde{A}_{ij} = \frac{1}{k_i} \sum_{i' \in n(i)} \text{abs}(w_{ii'}) \delta_{ij} + w_{ij} \quad (8)$$

360 On the other hand, the similarity is essentially codified in the term W_{ij} that we now want to redefine to
361 account for two types of interactions in such a way that only products $\tilde{A}_{ik} \tilde{A}_{kj}$ between terms with the same
362 sign contribute to the similarity. This is achieved with the following definition, which generalizes the Tanimoto
363 coefficient

$$W_{ij} = \sum_k \tilde{A}_{ik} \tilde{A}_{kj} \delta(\text{sgn}(\tilde{A}_{ik}) - \text{sgn}(\tilde{A}_{kj})) \quad (9)$$

364 where $\text{sgn}(\cdot)$ is the sign function and $\delta(a - b)$ is the Dirac delta function ($\delta(a - b) = 0$ if $a \neq b$). Generalizing
365 to an arbitrary number of types can be achieved by defining a variable μ_{ij} that returns the type of the link, i.e.
366 $\mu_{ij} = \alpha$ with α being a factor variable which, for the example of positive and negative links, is codified by the
367 sign of the links' weight. We finally generalize the expression 9 as follows

$$W_{ij} = \sum_k \tilde{A}_{ik} \tilde{A}_{kj} \delta(\mu_{ik} - \mu_{kj}). \quad (10)$$

368 Finally, the generalization of the external and internal partition densities to consider multiple types of links
 369 simply requires us to correctly classify the neighbours of each node accounting for the different types $n(i) =$
 370 $\bigcup_{\alpha=1}^{\Omega} n_{\alpha}^{int}(i) \cup \bigcup_{\alpha=1}^{\Omega} n_{\alpha}^{ext}(i)$. Similarly, the set of edges $m(i)$ linking the node i with other nodes must be also split
 371 into sets according to the different types $m(i) = \bigcup_{\alpha=1}^{\Omega} m_{\alpha}^{int}(i) \cup \bigcup_{\alpha=1}^{\Omega} m_{\alpha}^{ext}(i)$. The expressions for the internal
 372 and external partition densities remain otherwise the same.

373 Original definition of partition density

374 For completeness, we present the definition of partition density presented in [14]. In short, the method starts
 375 building a similarity measure between any pair of links sharing one node in common. Two links will be similar if
 376 the nodes that these two links do not share have, in turn, similar relationships with any other node, shown in Fig.
 377 1. From this similarity measure, edges are clustered and an optimal cut-off for the clustering is found monitoring a
 378 measure called *partition density* (which in this paper we call *internal partition density*). The optimal classification
 379 found at the cut-off, determines groups of links that are similar because they connect nodes that are themselves
 380 similar in terms of their connectivity. Therefore, the nodes are classified indirectly, according to the groups that
 381 their respective links belong, and a node may not belong to a single community but to several communities if its
 382 links belong to different clusters. This is claimed to be an advantage with respect to other methods (in particular
 383 for high density networks) as membership to a single cluster is not enforced. At every step of the clustering it is
 384 obtained a partition $P = P_1, \dots, P_C$ of the links into C subsets. For every subset, the number of links is $m_c = |P_c|$
 385 and the number of nodes that these edges are linking is $n_c = |\cup_{e_{ij} \in P_c} \{i, j\}|$. The density of links for the cluster
 386 C is then

$$D_c = \frac{m_c - (n_c - 1)}{n_c(n_c - 1)/2 - (n_c - 1)} \quad (11)$$

387 where the normalization considers the minimum $(n_c - 1)$ and maximum $(n_c(n_c - 1)/2)$ number of links that
 388 can be found in the partition. The difference with respect to Eq. 4, is that a term $(n_c^{int} - 1)$ is now subtracted.
 389 The reason is that, in Ahn *et al.* method, clustering with links implies that two nodes in the same cluster must
 390 share links. But, according to our definition of function, two nodes may be structurally equivalent even if there
 391 is no interaction between them.

392 The partition density D is then given by the average of the density of links for all the partitions, weighted by
 393 the number of links

$$D = \frac{2}{M} \sum_c m_c \frac{m_c - (n_c - 1)}{(n_c - 2)(n_c - 1)} \quad (12)$$

394 where M is the total number of links. It was shown that when using agglomerative clustering, this function
 395 achieves a maximum which determines the optimal partition [14].

396 Clustering algorithm

397 After computing the similarity between nodes with the method presented in the Results, the algorithm clusters
 398 nodes using one of three hierarchical clustering algorithms: average linkage [29], single linkage and complete
 399 linkage. Starting from each node being a separate cluster, at each step t all algorithms join the two most similar
 400 clusters A and B , and compute the similarity between the new combined cluster and all other clusters C in a way
 401 that depends on the clustering algorithm.

402 Single linkage is the most permissive algorithm, because the similarity it assigns to the new cluster is the
 403 maximum similarity between the two clusters joined and clusters C :

$$S^{t+1}(AB, C) = \max(S(A, C), S(B, C)).$$

404 where t labels the step of the algorithm, A and B are the clusters that are joined, AB denotes the new
 405 composite cluster, and C is any other cluster. On the other hand, complete linkage is the most restrictive,
 406 assigning the minimum similarity

| ν | κ_{mut} | κ_{comp} |
|-------|-----------------------|------------------------|
| 0.15 | 0.08 | 0 |
| 0.35 | 0.16 | 0.15 |
| 0.6 | 0.28 | 0.5 |

Table 1: **Topological properties of the bipartite networks analysed.** Different combinations of nestedness (ν), intra-pools connectance κ_{comp} and inter-pools connectance κ_{mut} were analysed.

$$S^{t+1}(AB, C) = \min(S(A, C), S(B, C)).$$

407 Finally, average linkage assigns an intermediate value computed as the weighted average similarity with the
408 two joined clusters

$$S^{t+1}(AB, C) = \frac{n_A S^t(A, C) + n_B S^t(B, C)}{n_A + n_B}$$

409 being n_A and n_B the number of elements that A and B contain, respectively. Identification of the two pools
410 of plants and animals is independent of the clustering method used, but the maximum of the external partition
411 density is achieved earlier for single linkage and later for complete linkage; we found a good compromise between
412 the number and the size of the clusters working with average linkage, but the clustering method could be selected
413 according to information known from the links. In our experience, single linkage is easily dominated by the giant
414 cluster in high density networks in which modules are prevalent (rather than for guilds). The appropriate clustering
415 method should be guided by the research question. For instance, if gene homology is explored, it is probably
416 more appropriate to use single linkage (as a relative of one gene's relative is also its relative, i.e. transitivity is
417 automatically fulfilled [30]). On the other hand, if we analyse well-differentiated functional similarity, it might be
418 more appropriate to be conservative and use complete linkage.

419 Plant-pollinator networks and topological properties

420 We selected six plant-pollinator networks artificially generated in [19] with known topological properties, sum-
421 marized in Table. We consider as topological properties the connectance (fraction of links) of the mutualistic
422 matrix, the connectance of the competition matrices, and the definition of nestedness provided in [20]. Given a
423 mutualistic matrix $A_{ik}^{(P)}$ representing presence-absence of interaction between the set of plants, indexed by i , and
424 the set of animal species, indexed by k , we compute the degree of a species as $n_i^{(P)} = \sum_k A_{ik}^{(P)}$ (see Ref. [20]
425 in Supplementary Material). A similar definition would apply for animals $n_k^{(A)} = \sum_i A_{ik}^{(P)}$. Next we define the
426 ecological overlap between two species of plants i and j as the number of insects that pollinate both plants:

$$n_{ij}^{(P)} = \sum_k A_{ik}^{(P)} A_{jk}^{(P)},$$

427 a definition that is equivalent to the Jaccard similarity used in this work. Summing over every pair of plants
428 and normalizing leads to the definition of nestedness:

$$\nu^{(P)} = \frac{\sum_{i < j} n_{ij}^{(P)}}{\sum_{i < j} \min(n_i^{(P)}, n_j^{(P)})}.$$

429 A symmetric definition applies for animals, so we take as final definition of nestedness $\nu = \max(\nu^{(P)}, \nu^{(A)})$.

430 Trophic networks

431 We downloaded the network and metadata provided in [21] and compared the clusters found with those obtained
432 by functionInk. After computing the Tanimoto coefficients as explained above, we cluster the nodes and retrieve
433 the classification found at each step. We then computed five indexes (Rand, Fowlkes and Mallows, Wallace 10,
434 Wallace 01 and Jaccard), implemented in the R PCI function of the PROFDPM package [22]. In order to assign a

435 significance value for the different indexes we obtained, for each index x , a bootstrapped distribution with mean
436 $\bar{x}_{(B)}$ and standard deviation $\sigma_{(B)}$, resampling with replacement the samples and recomputing the indexes 10^3
437 times. Next we computed 10^3 completely random classifications, obtained by shuffling the identifiers relating each
438 sample with one of the classifications, and retrieving the maximum $x_{(R)}$. We finally verified that the random
439 value was significantly different from the bootstrapped distribution by computing the z-scores:

$$z = \frac{\text{abs}(x_{(R)} - \bar{x}_{(B)})}{\sigma_{(B)}},$$

441 which we considered significant if it was higher than 2.5. Heatmaps were generated with the HEATMAP.2 function
442 in R package GPLOTS.

443 Bacterial networks

444 We considered a public dataset of 753 bacterial communities sampled from rainwater-filled beech tree-holes (*Fagus*
445 spp.) [24], leading to 2874 Operative Taxonomic Units (OTUs) at the 97% of 16 rRNA sequence similarity. These
446 communities were compared with Jensen-Shannon divergence [25], and automatically clustered following the
447 method proposed in Ref. [31] to identify enterotypes. The clusters found with this method in [23] were used to
448 colour the community labels in Fig. 6.

449 The inference of the OTU network started quantifying correlations between OTUs abundances with SparCC
450 [26]. To perform this computation, from the original OTUs we reduced the data set removing rare taxa with
451 less than 100 reads or occurring in less than 10 samples, leading to 619 OTUs. Then, the significance of the
452 correlations was evaluated bootstrapping the samples 100 times the data and estimating pseudo p-values for each
453 of the $N(N - 1)/2$ pairs. A relationship between two OTUs was considered significant and represented as a link
454 in the network if the correlation was larger than 0.2 in absolute value and the pseudo p-value lower than 0.01.
455 The network obtained in this way was analysed using Cytoscape [32].

456 Acknowledgements

457 We thank Michael Schaub for critical comments on an earlier version of the manuscript. The research was funded
458 by a European Research Council starting grant (311399-Redundancy) awarded to T.B. T.B. was also funded by
459 a Royal Society University Research Fellowship. APG was also funded by the Simons Collaboration: Principles
460 of Microbial Ecosystems (PriME), award number 542381.

461 **Supplementary Figures**

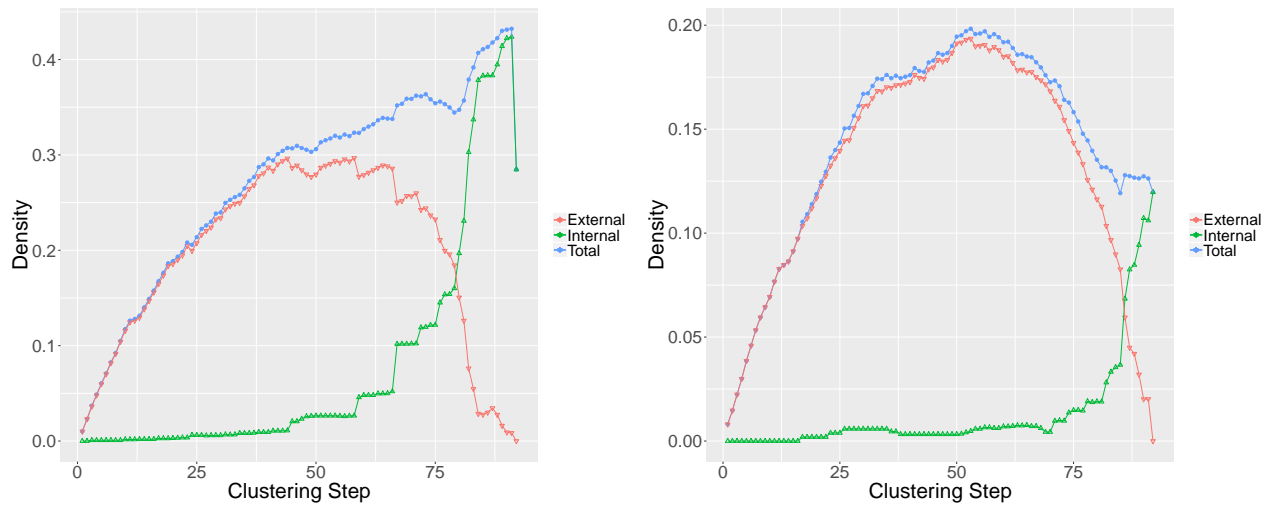


Figure 8: **Partition densities of synthetic mutualistic networks.** Networks with nestedness $\nu = 0.15$, $\kappa_{\text{mut}} = 0.08$, and $\kappa_{\text{comp}} = 0.5$ (left) or $\kappa_{\text{comp}} = 0.15$ (right). Changing the connectance change the relative value between the external and internal partition densities.

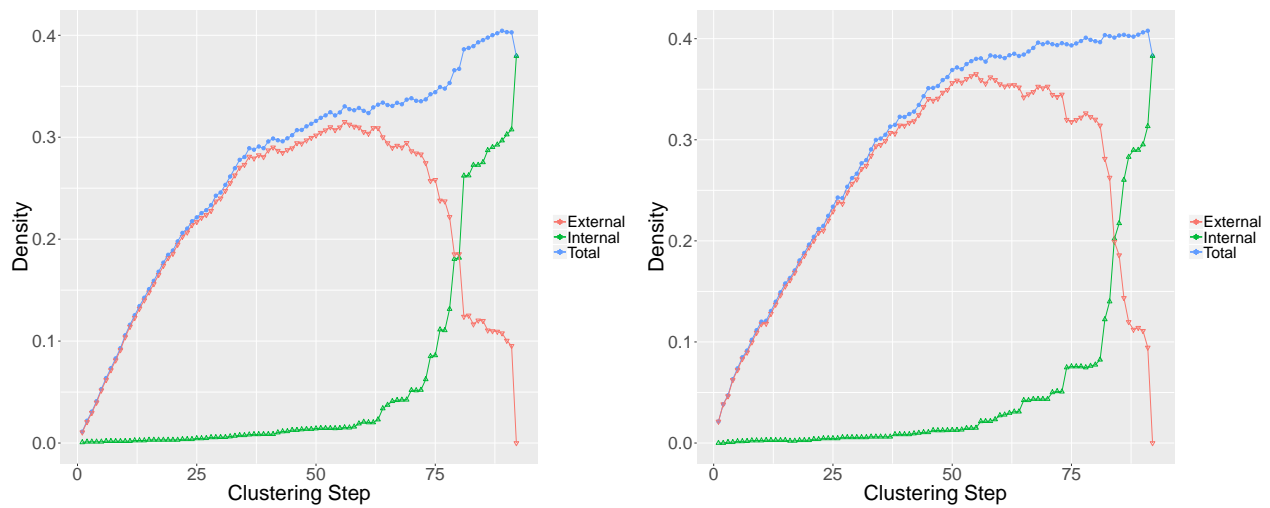


Figure 9: **Partition densities for synthetic mutualistic networks.** Networks with $\kappa_{\text{comp}} = 0.5$, $\kappa_{\text{mut}} = 0.28$ and $\nu = 0.35$ (left) or $\nu = 0.6$ (right). The high connectance of both networks make the internal partition density dominant, and two pools are detected through the total partition density. Nevertheless, the increase of the nestedness is detected through an increase in the internal partition density, which makes the second network more disassortative.

462

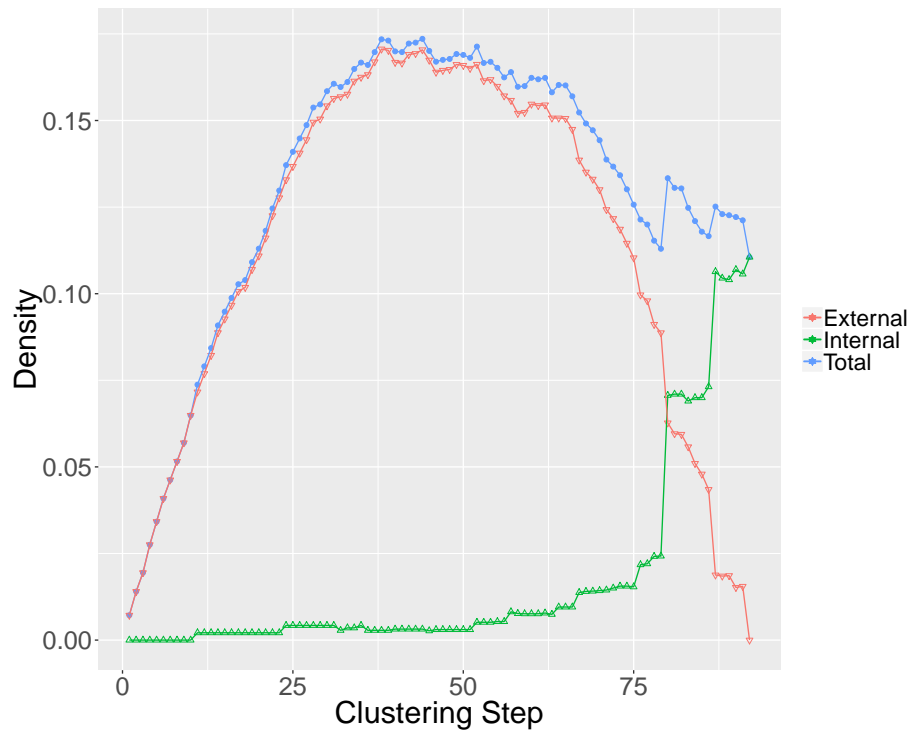


Figure 10: **Partition densities of synthetic mutualistic networks.** Network with nestedness $\nu = 0.05$, $\kappa_{\text{mut}} = 0.065$ and $\kappa_{\text{comp}} = 0.15$. The low connectance hinders the detection of the two pools of plants and pollinators.

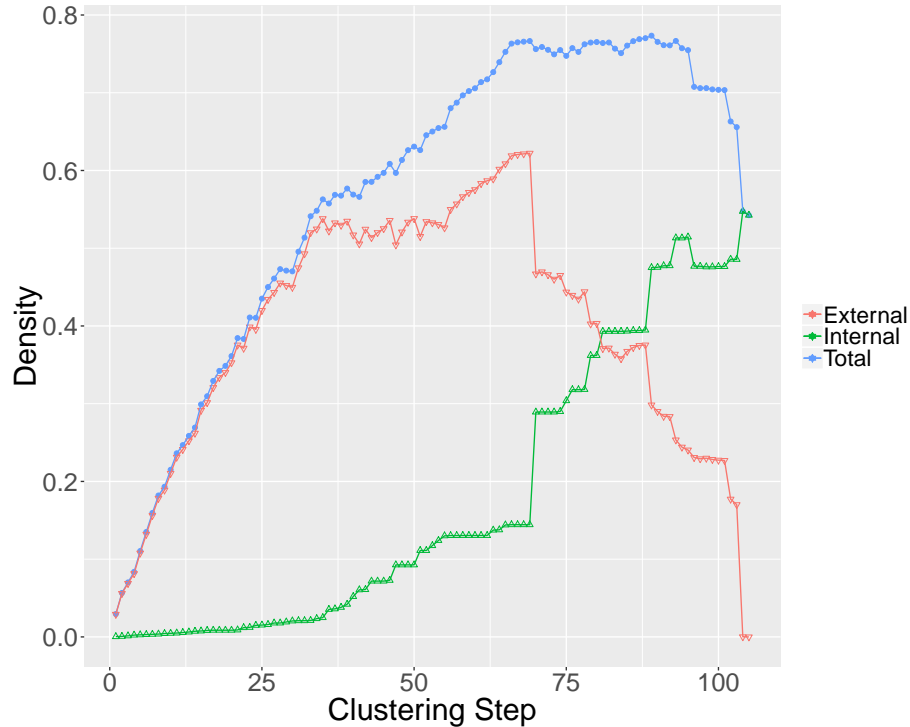


Figure 11: **Partition density of the trophic network.** The internal partition density peaks when there are three clusters, consistent with the existence of three trophic layers. The external partition density has a maximum at step 69, which is analysed in detail with respect to the reference classification found in [21].

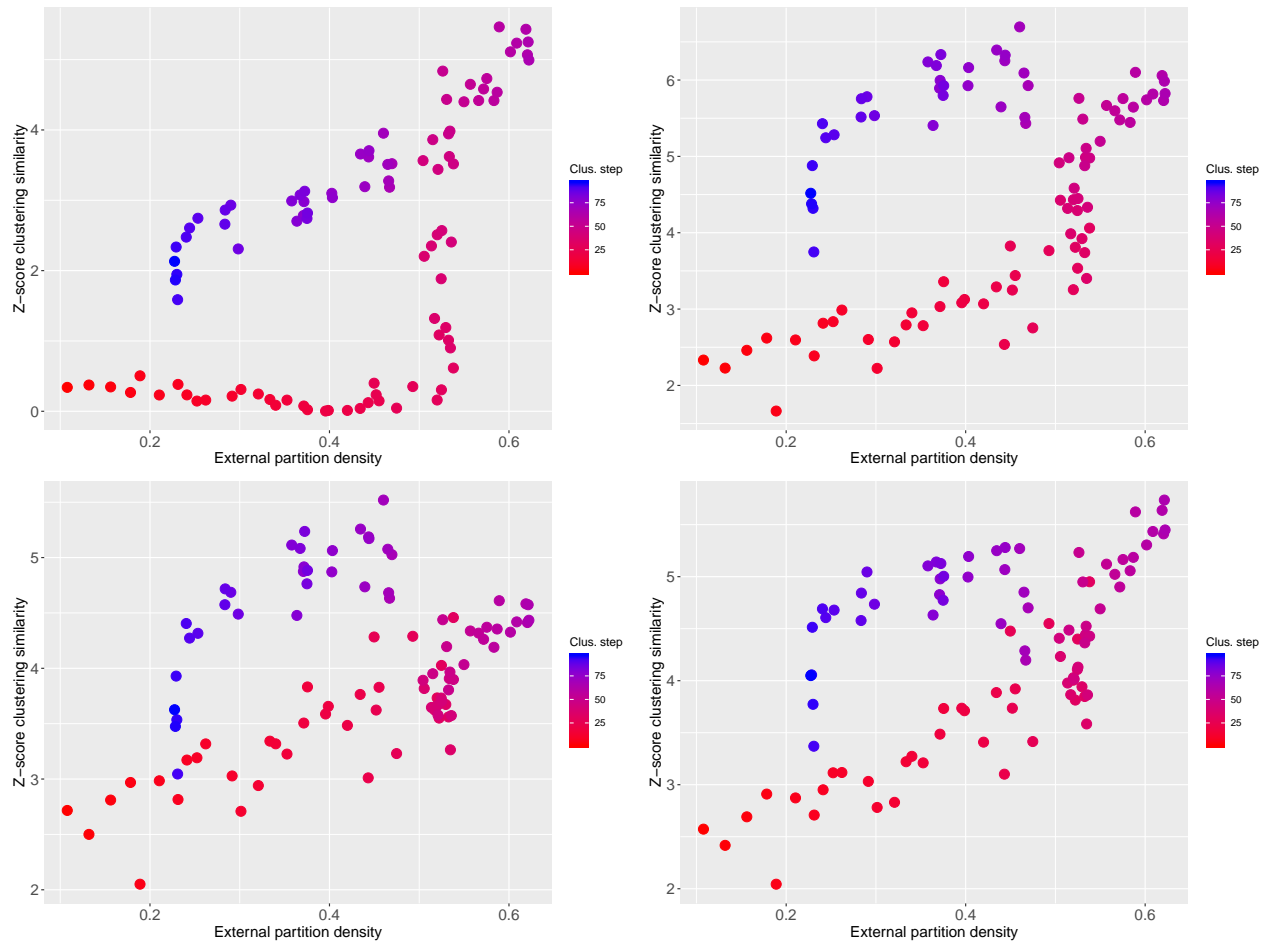


Figure 12: **Comparison between classifications of the trophic network.** Similarity between the reference classification found in [21] and the one found with functionInk is performed with the Z-score of a different indexes: Wallace 01 (Top left), Fowlkes and Mallows (Top right), Jaccard (Bottom left) and Rand (Bottom right). All indexes bring significant values and the maximum similarity is close to the maximum of the external partition density.

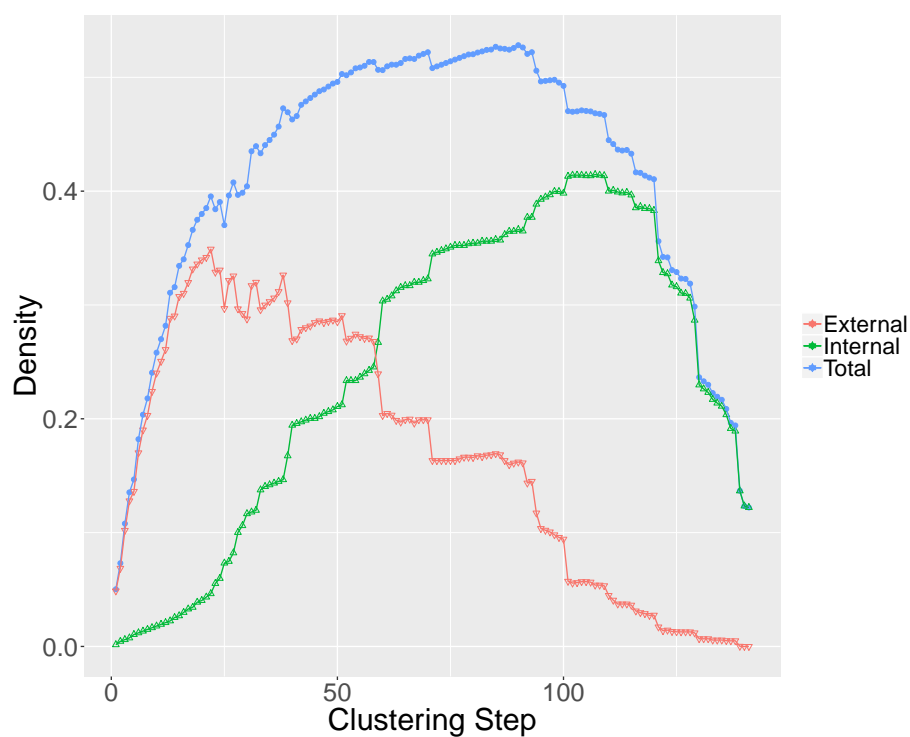


Figure 13: **Partition density of the microbial network.** The external partition density brings a poor reduction in the complexity of the network, with only 22 elements joined, while the internal partition density achieves a higher value and still a good number of clusters. Results suggest that modules are more relevant in this network given the high number of intra-cluster co-occurrences, later confirmed by visual inspection in the Main Text.

References

- 463
- 464 [1] J. E. Cohen and D. W. Stephens, *Food webs and niche space*. No. 11, Princeton University Press, 1978.
- 465 [2] R. MacArthur, “Fluctuations of animal populations and a measure of community stability,” *Ecology*, vol. 36,
466 p. 533, July 1955.
- 467 [3] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,”
468 *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014.
- 469 [4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and
470 dynamics,” *Physics reports*, vol. 424, no. 4, pp. 175–308, 2006.
- 471 [5] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” *Proceedings of the
472 national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.
- 473 [6] R. Lambiotte, J.-C. Delvenne, and M. Barahona, “Laplacian dynamics and multiscale modular structure in
474 networks,” *arXiv preprint arXiv:0812.1770*, 2008.
- 475 [7] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, “Community structure in time-
476 dependent, multiscale, and multiplex networks,” *science*, vol. 328, no. 5980, pp. 876–878, 2010.
- 477 [8] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*, vol. 8. Cambridge university
478 press, 1994.
- 479 [9] C. S. Elton, *Animal ecology*. New York: The Macmillan Company, 1927.
- 480 [10] M. E. Newman and E. A. Leicht, “Mixture models and exploratory analysis in networks,” *Proceedings of the
481 National Academy of Sciences*, vol. 104, no. 23, pp. 9564–9569, 2007.
- 482 [11] M. E. Newman, “Finding community structure in networks using the eigenvectors of matrices,” *Physical
483 review E*, vol. 74, no. 3, p. 036104, 2006.
- 484 [12] E. Estrada and J. A. Rodríguez-Velázquez, “Spectral measures of bipartivity in complex networks,” *Physical
485 Review E*, vol. 72, no. 4, p. 046105, 2005.
- 486 [13] D. Simberloff and T. Dayan, “The guild concept and the structure of ecological communities,” *Annual review
487 of ecology and systematics*, vol. 22, no. 1, pp. 115–143, 1991.
- 488 [14] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, “Link communities reveal multiscale complexity in networks,”
489 *Nature*, vol. 466, no. 7307, pp. 761–764, 2010.
- 490 [15] M. Ganji, J. Chan, P. J. Stuckey, J. Bailey, C. Leckie, K. Ramamohanarao, and I. Davidson, “Image con-
491 strained blockmodelling: a constraint programming approach,” in *Proceedings of the 2018 SIAM International
492 Conference on Data Mining*, pp. 19–27, SIAM, 2018.
- 493 [16] R. Guimera and L. A. N. Amaral, “Cartography of complex networks: modules and universal roles,” *Journal
494 of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 02, p. P02001, 2005.
- 495 [17] T. T. Tanimoto, “elementary mathematical theory of classification and prediction,” 1958.
- 496 [18] A. Harrer and A. Schmidt, “Blockmodelling and role analysis in multi-relational networks,” *Social Network
497 Analysis and Mining*, vol. 3, no. 3, pp. 701–719, 2013.
- 498 [19] A. Pascual-García and U. Bastolla, “Mutualism supports biodiversity when the direct competition is weak,”
499 *Nature Communications*, vol. 8, p. 14326, Feb. 2017.
- 500 [20] U. Bastolla, M. A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, and J. Bascompte, “The architecture
501 of mutualistic networks minimizes competition and increases biodiversity,” *Nature*, vol. 458, pp. 1018–1020,
502 Apr. 2009.

- 503 [21] S. Kéfi, V. Miele, E. A. Wieters, S. A. Navarrete, and E. L. Berlow, “How structured is the entangled
504 bank? the surprisingly simple organization of multiplex ecological networks leads to increased persistence
505 and resilience,” *PLoS biology*, vol. 14, no. 8, p. e1002527, 2016.
- 506 [22] M. S. Shotwell *et al.*, “profdpm: An r package for map estimation in a class of conjugate product partition
507 models,” *J Stat Softw*, vol. 53, no. 8, pp. 1–18, 2013.
- 508 [23] A. Pascual-García and T. Bell, “Community-level signatures of ecological succession in natural bacterial
509 communities,” *bioRxiv*, p. 636233, 2019.
- 510 [24] D. W. Rivett and T. Bell, “Abundance determines the functional role of bacterial phylotypes in complex
511 communities,” *Nature microbiology*, p. 1, 2018.
- 512 [25] D. M. Endres and J. E. Schindelin, “A new metric for probability distributions,” *IEEE Transactions on
513 Information Theory*, vol. 49, pp. 1858–1860, July 2003.
- 514 [26] J. Friedman and E. J. Alm, “Inferring correlation networks from genomic survey data,” *PLoS computational
515 biology*, vol. 8, no. 9, p. e1002687, 2012.
- 516 [27] A. Pascual-García, J. Tamames, and U. Bastolla, “Bacteria dialog with santa rosalia: Are aggregations of
517 cosmopolitan bacteria mainly explained by habitat filtering or by ecological interactions?,” *BMC microbiology*,
518 vol. 14, no. 1, p. 284, 2014.
- 519 [28] M. T. Agler, J. Ruhe, S. Kroll, C. Morhenn, S.-T. Kim, D. Weigel, and E. M. Kemen, “Microbial hub taxa
520 link host and abiotic factors to plant microbiome variation,” *PLoS Biology*, vol. 14, no. 1, p. e1002352, 2016.
- 521 [29] R. R. Sokal, “A statistical method for evaluating systematic relationships,” *Univ Kans Sci Bull*, vol. 38,
522 pp. 1409–1438, 1958.
- 523 [30] A. Pascual-García, D. Abia, Á. R. Ortiz, and U. Bastolla, “Cross-over between discrete and continuous
524 protein structure space: insights into automatic classification and networks of protein structures,” *PLoS
525 computational biology*, vol. 5, no. 3, p. e1000331, 2009.
- 526 [31] M. Arumugam, J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap,
527 T. Bruls, J.-M. Batto, *et al.*, “Enterotypes of the human gut microbiome,” *Nature*, vol. 473, no. 7346,
528 pp. 174–180, 2011.
- 529 [32] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and
530 T. Ideker, “Cytoscape: a software environment for integrated models of biomolecular interaction networks,”
531 *Genome research*, vol. 13, no. 11, pp. 2498–2504, 2003.