APPLICATION

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Identifying flow modules in ecological networks using Infomap

Carmel Farage[1] | **Daniel Edler**[2,3,4] | **Anna Eklöf**[5] | **Martin Rosvall**[2] |
Shai Pilosof[1]

[1]Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

[2]Integrated Science Lab, Department of Physics, Umeå University, Umeå, Sweden

[3]Gothenburg Global Biodiversity Centre, Gothenburg, Sweden

[4]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden

[5]Division of Theoretical Biology, Department of Physics, Chemistry and Biology, Linköping University, Linköping, Sweden

Correspondence
Shai Pilosof
Email: pilos@post.bgu.ac.il

Abstract

1. Analysing how species interact in modules is a fundamental problem in network ecology. Theory shows that a modular network structure can reveal underlying dynamic ecological and evolutionary processes, influence dynamics that operate on the network and affect the stability of the ecological system.

2. Although many ecological networks describe flows, such as biomass flows in food webs or disease transmission, most modularity analyses have ignored network flows, which can hinder our understanding of the interplay between structure and dynamics.

3. Here we present Infomap, an established method based on network flows to the field of ecological networks. Infomap is a flexible tool that can identify modules in virtually any type of ecological network and is particularly useful for directed, weighted and multilayer networks. We illustrate how Infomap works on all these network types. We also provide a fully documented repository with additional ecological examples. Finally, to help researchers to analyse their networks with Infomap, we introduce the open-source R package infomapecology.

4. Analysing flow-based modularity is useful across ecology and transcends to other biological and non-biological disciplines. A dynamic approach for detecting modular structure has strong potential to provide new insights into the organisation of ecological networks.

KEYWORDS
community detection, dynamics, ecological networks, flow, Infomap, modularity, multilayer

1 | INTRODUCTION

Understanding the interplay between the structure and dynamics of complex ecological systems is at the heart of network ecology. Partitioning a network into modules composed of nodes more tightly connected to each other than to other nodes is a leading example. Modules are a topological description of realised interaction patterns. It has been shown that a modular structure can make ecological communities locally stable (Grilli et al., 2016), increase species persistence (Stouffer & Bascompte, 2011), serve as a signature for evolutionary processes (Pilosof et al., 2019) and slow down the spread of perturbations (see Gilarranz et al. 2017 for experimental evidence).

There are three main ways to detect modules in networks (Rosvall et al., 2018): (a) By maximising the internal density of links within groups of nodes (Newman & Girvan, 2004; Olesen et al., 2007; Thébault, 2013); (b) by identifying structurally equivalent groups in which nodes connect to others with equal probability, typically studied using stochastic-block models (Holland et al., 1983), known as the 'group model' in ecology (Allesina & Pascual, 2009) and (c) by optimally describing modular flows on networks (Rosvall et al., 2010; Rosvall & Bergstrom, 2008) (Supporting Information Text 1). These approaches have been developed for different purposes, with different mathematical functions and algorithms to detect an 'optimal' partition of a network. Therefore, there is no single 'true' network partition (Peel et al., 2017). Instead, the method applied should match the question

(Ghasemian et al., 2019; Rosvall et al., 2018). For example, many ecological systems describe flows on networks, including biomass flow in food webs (Baird & Ulanowicz, 1989), movement of individuals between patches (Hanski & Gilpin, 1991) and gene flow among individuals and populations (Fletcher Jr et al., 2013). In such cases, understanding how network flows organise in modules can be more relevant to the system at hand than maximising internal interaction density.

To date, maximising variants of Newman–Girvan's combinatorial modularity score Q is the dominant approach in ecology (reviewed in Thébault (2013)). While this method undoubtedly has provided many insights, it is not designed to capture network flows. Also, modularity maximisation methods for various applications are scattered in different software implementations. For example, the R package bipartite (Dormann et al., 2009) has an implementation for modularity maximisation in bipartite weighted and unweighted networks, while Netcarto (Guimerà & Nunes Amaral, 2005) is an implementation for unipartite, undirected networks. To fill these conceptual and technical gaps, we present an established method for detecting flow-based modules called Infomap.

Infomap has several advantages for ecological research. First, it can be applied to many types of networks, including directed/undirected, weighted/unweighted, unipartite/bipartite and multilayer networks. Second, it is computationally effective, supporting studies of large networks or comparing observed networks with many randomised networks. Third, it can incorporate node attributes by explicitly considering information such as taxonomy for the partitioning into modules. Fourth, it can detect hierarchical structures of modules within modules. Finally, Infomap has online documentation and an active development team that has made it user-friendly and flexible. These advantages make Infomap a highly accessible tool that can be applied to virtually any kind of ecological system. Moreover, Infomap has been thoroughly described mathematically and computationally (Rosvall et al., 2010, 2014; Rosvall & Bergstrom, 2008, 2011) and has already been benchmarked against other methods (Aldecoa & Marín, 2013; Lancichinetti & Fortunato, 2009), providing a sound theoretical and applied understanding of the method.

Despite these advantages, Infomap has only been used in a handful of ecological studies (Bernardo-Madrid et al., 2019; Pilosof et al., 2019, 2020). Therefore, our goal here is twofold: (a) introduce Infomap to ecologists with guidelines on how to apply it to particular problems and (b) help users analyse their networks with the dedicated R package infomapecology we have developed—a one-stop-shop that also integrates with other R packages commonly used by ecologists such as bipartite and igraph.

2 | INFOMAP AND THE MAP EQUATION OBJECTIVE FUNCTION

2.1 | General approach to network partitioning

To understand how Infomap works, it is helpful first to understand the general approach for modularity analysis (Supporting Information Text 2). A particular assignment of the nodes into modules is called a *network partition*. As even small networks can have an enormous number of possible partitions, search algorithms measure the quality of a given partition with an objective function. The algorithms then make a small change in the partition, such as moving a node from one module to another, and test whether the value of the objective function improves. Modularity analysis algorithms differ in the search algorithms and objective functions they apply.

Infomap optimises the objective function known as the *map equation* using a modified and extended Louvain search algorithm (Blondel et al., 2008). Specifically, the algorithm finds the partition that best compresses a description of flows on the network. The network flows are modelled by a random walker or observed empirical flows if available (Supporting Information Text 3). The random walker moves across nodes in a way that depends on the direction and weight of the links, and tends to stay longer in dense areas that then represent modules. For a given partition of the network, there is an associated information cost, measured in bits, for describing the movements of the random walker. The map equation converts the flow rates within and between the modules to an information-theoretic modular description measure of the random walker's movements on the network. *Minimising* the map equation over possible network partitions corresponds to detecting the most modular structure possible in the dynamics on the network.

2.2 | The map equation: Linking structure and information

To calculate the map equation, Infomap uses *node* and *link rates*, which are calculated based on link direction and weights. For example, in the schematic network in Figure 1a, there are 14 directed links of weight of 1, resulting in total incoming link weight of 14. Therefore, each directed link carries flows of link visit rate 1/14. These can also be viewed as seven undirected links (flow equals link weights in undirected networks). Nodes with two incoming links have a node visit rate of 2/14, and nodes with three links have a node visit rate of 3/14. These rates are included in the so-called 'module codebook'. In the one-module solution, all the nodes belong to a single module and, therefore, to a single module codebook (Figure 1c). In the two-module solution (Figure 1b), there are two module codebooks (Figure 1d). To describe a random walk in the latter case, it is also necessary to consider the rates of entering and exiting each module using the *module entry rate* and the *module exit rate*, respectively (which are equal for undirected networks). Module entry rates are encoded in an 'index codebook'. In the two-module solution, these events are 'enter green' and 'enter orange', which both occur at rate 1/14. The rates of exiting modules are encoded within the module codebooks (Figure 1d).

The map equation uses Shannon's source coding theorem (Shannon, 1948) to convert the rates encoded in the codebooks to information measured in bits. Specifically, given a network partition M, we can calculate the minimum amount of information needed
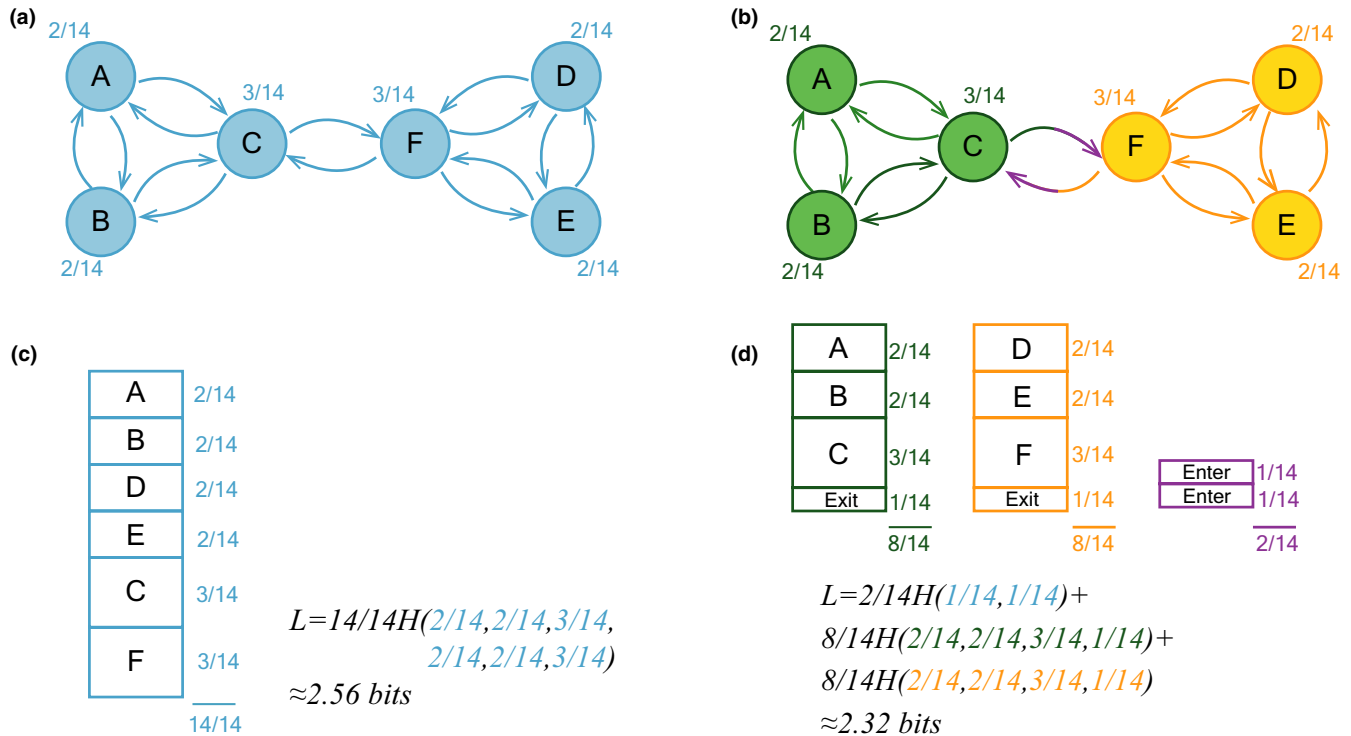
(a)

(b)

(c)

$$L=14/14H(2/14,2/14,3/14, \\ 2/14,2/14,3/14) \\ \approx 2.56 \text{ bits}$$

(d)

$$L=2/14H(1/14,1/14)+ \\ 8/14H(2/14,2/14,3/14,1/14)+ \\ 8/14H(2/14,2/14,3/14,1/14) \\ \approx 2.32 \text{ bits}$$

FIGURE 1 Basics of the map equation. The example is for a schematic network with 14 directional links, in which the total rate of flow is 14. (a) A one-module solution. The node visit rates written besides each node are the number of incoming links. (b) A two-module solution. Each module is represented by a different colour. The purple arrow heads illustrate that these links are considered as module entry links. (c) A single module codebook for a one-module solution. Each block represents a node, with width proportional to the node's visit rate. (d) Three codebooks for a two-module solution. An index codebook to encode module enter rates (purple codebook), and two module codebooks (green and orange) for node visit rates of each module. The total rate of use of a codebook is the sum of its rates (indicated below the codebook). In (c) and (d), for each solution, we calculate L, the value of the map equation, which is the entropy of the use rates within each codebook, weighted by the use rates of the codebooks. An expanded explanation that includes the relationship between flow rates and information theory are found in Figure S1

to describe an average movement length of a random walker. This quantity L is the entropy H of the events encoded in the codebooks, weighted by the use rate of each codebook (equations in Figure 1c,d). Summing the terms for the index codebook and the module codebooks, we obtain the map equation (Rosvall et al., 2010; Rosvall & Bergstrom, 2008),

$$L(\mathrm{M}) = q_\curvearrowright H(\mathcal{Q}) + \sum_{i=1}^{m} p^i_\circlearrowright H\left(\mathcal{P}^i\right), \qquad (1)$$

where $H(\mathcal{Q})$ and $H\left(\mathcal{P}^i\right)$ are the entropy values of the index codebook and the codebook of module i, respectively. These entropy terms are weighted by the rate at which the codebooks are used. The index codebook is weighted by the rate of entering any module, $q_\curvearrowright$, and each module codebook i is weighted by its within-module flow, $p^i_\circlearrowright$, which includes the node-visit rates and the exit rate in module i. For the examples in Figure 1, $L(\mathrm{M}_1) \approx 2.56$ for the one-module solution and $L(\mathrm{M}_2) \approx 2.32$ for the two-module solution. The two-module solution requires fewer bits and hence better captures the modular structure of the network.

In practice, Infomap can use either real measured flows or estimates of flows (Supporting Information Text 3.2). In the latter and

more typical case, Infomap derives link and node visit rates using an iterative process akin to the PageRank algorithm (Brin & Page, 1998). First, each node receives an equal amount of flow volume. Then, iteratively until all node visit rates are stable, each node distributes all its flow volume to its neighbours proportionally to the outgoing link weights. We note that PageRank is only used for directed networks because it is superfluous for undirected networks. A comprehensive description on flow models are found in the Supplementary Information (Supporting Information Text 3.2) and in Rosvall and Bergstrom (2008), Rosvall et al. (2010), Bohlin et al. (2014) and De Domenico et al. (2015).

2.3 | Extension to multilayer networks

In multilayer networks, nodes representing observable entities such as species are called *physical nodes*. Realisations of physical nodes in a given layer—for example, in different time points, patches or interaction types—are called *state nodes*. The random walker moves from state node to state node within and across the layers. However, the encoded position always refers to the physical node (see dynamic visualisation: https://www.mapequation.org/apps/multilayer-network/index.html). This approach provides two advantages. First, it enables

a physical node to be assigned to different modules in different layers. From an ecological perspective, this is crucial as a certain species can have different functions in different layers. For example, there is a strong spatial and temporal variation in plant–pollinator interactions (Olesen et al., 2008; Trøjelsgaard et al., 2015) Second, it enables to model the coupling between layers without interlayer links. This feature is particularly useful in ecology because interlayer links are often challenging to measure empirically (Hutchinson et al., 2018). If interlayer links are not provided, the random walker 'relaxes' to the current physical node in a random layer at a 'relax rate' r, without recording this movement. By gradually tuning the relax rate, it is possible to explore the relative contribution of intra- and interlayer links to the structure (Figure 5 and Supporting Information Text 3.4).

3 | IMPLEMENTATION, AVAILABILITY AND CODE

Full documentation of Infomap, including tutorials, instructions and visualisation tools, is available at https://www.mapequation.org/infomap/. Detailed installation instructions for infomap and infomapecology, detailed descriptions of input/output formats, source code of infomapecology and the code used to produce the analyses in this paper are available at https://ecological-complexity-lab.github.io/infomap_ecology_package/. In addition, each function in infomapecology has examples in its description, accessible via R's help (e.g. ?create_monolayer_object).

3.1 | General approach

When using infomapecology, the first step is to convert the input data to an object of class monolayer or multilayer. The monolayer class is an R list with information about the network (e.g. bipartite, directed), a list of nodes and their attributes, and network representations as a matrix, an edge list and an igraph object. With multiple data structures, it is easy to streamline and standardise the workflow with other R packages. As ecological networks are typically relatively small, using multiple data structures have limited computational consequences. If the network is large, it is straightforward to extract only a single data structure or use sparse matrices. A monolayer object is created using the function create_monolayer_object, which as input can take matrices, edge lists and igraph objects, and can also incorporate node attributes. With a created monolayer object, Infomap is ready to run. A basic example:

```
# Use the memmott1999 bipartite network represented as a matrix
from package bipartite
monolayer_network <- create_monolayer_object (memmott1999,
    bipartite = T, directed = F, group_names = c('Animals', 'Plants'))

# Run Infomap
modularity_results <- run_infomap_monolayer (monolayer_network,
    infomap_executable = 'Infomap', flow_model = 'undirected', silent = T, trials = 20, two_level = T, seed = 123)
```

For multilayer networks, the input must be in the form of an edge list. The exact format depends on the existence of interlayer edges. A data frame with nodes is also necessary. It is also possible to provide information on each layer (e.g. coordinates). Infomapecology will standardise the input and produce a multilayer object with intralayer and interlayer edges, and information on nodes and layers. A multilayer network example:

```
# Create a multilayer object with the Siberia data set provided with
the package
NEE2017 <- create_multilayer_object (extended = siberia1982_7_
    links, nodes = siberia1982_7_nodes, intra_output_extended = T,
    inter_output_extended = T)

# Run infomap
NEE2017_modules <- run_infomap_multilayer (M = NEE2017,
    relax = F, flow_model = 'directed', silent = T, trials = 100,
    seed = 497294, temporal_network = T)
```

For monolayer and multilayer networks, the results are stored in objects of class infomap_monolayer and infomap_multilayer, respectively, which contain the call for Infomap, the value of L, the number of modules and a data frame with the module affiliation of nodes.

3.2 | Use cases

Thanks to its flexibility, Infomap can find modules in many types of networks. Here we exemplify with directed, weighted networks, which are adequate for representing flows, and multilayer networks for analysing modular flows over time. We present other types of networks, including bipartite networks, and hierarchical modularity in the Extended Use Cases (Supporting Information Text 4). The goal of all these use cases is to demonstrate the capacity and flexibility of the framework and to provide general guidelines. We aim to help users analyse their networks rather than to provide full interpretations of the analysed networks.

3.3 | Weighted and directed networks

To demonstrate the usefulness of Infomap in identifying flows in weighted networks, we use data from Gilarranz et al. (2017), who built an experimental network of 20 cups (patches) connected by tubes and partitioned into four modules (Figure 2). Gilarranz et al. (2017) allowed springtails to disperse freely between the patches and showed that the effects of perturbation to a particular node in the network—leading to local extinction of springtails in the patch—are primarily contained within the cup's module. Flow modules can provide an adequate description of this dispersal system.

When we applied Infomap and Newman–Girvan's modularity score Q to the original, unweighted and undirected network (springtails can move in both directions with uniform constraints on movement), both methods partitioned the network into the same four
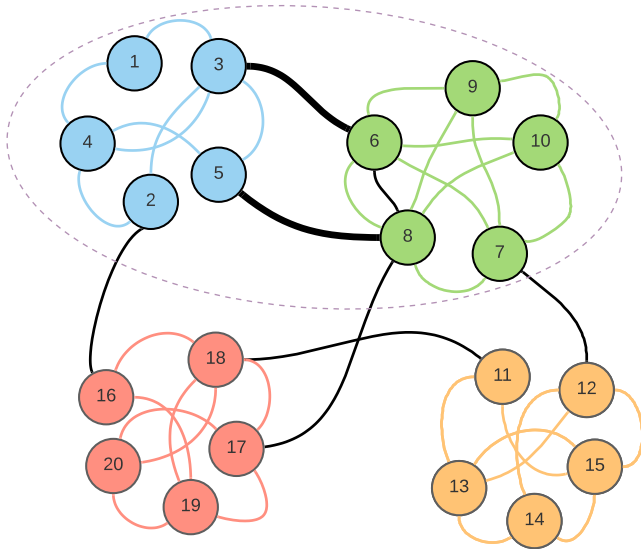

FIGURE 2 Comparing Infomap to Q in weighted network. The network is the one designed by Gilarranz et al. (2017) to have four modules, depicted by node colours. Edges within and between modules are coloured by either module colour or black, respectively, and their weight is 1. We computationally increased the weight of two edges between the green and light blue modules (thick black lines) from 1 to 4 in increments of 0.1. This analysis showed that Infomap has a threshold (2.1) above which the two strongly connected blue and green modules merge into a single module (depicted by a dashed ellipse), while Q considers them as four modules consistently as it only uses the unweighted information

experimentally designed modules. However, when we computationally increased the connectivity between two of the designed modules in the network, Infomap identified three modules by merging the two original modules as expected. In comparison, Q still found the same four modules (Figure 2). If we were to repeat the experiment with increased link weights by using wider tubes, we would expect local extinctions to be confined to the 10 nodes within the new module. This ecological example with network flows indicates that Infomap is more sensitive to changes in flows than Q (Table S1). Lancichinetti and Fortunato (2009) and Aldecoa and Marín (2013) show quantitative comparisons of Infomap and Newman–Girvan's modularity score optimised with the Louvain method, and Rosvall et al. (2018) illustrate how the flow-based map equation and the combinatorial modularity score highlight different aspects of networks.

As an example of a directed network, we use data from Tur et al. (2016), who measured directed flows of pollen grains (links) in south Andean communities, at three elevations. In their networks, nodes are plant species and links are directed from species i to j when pollen of species i was detected on stigmas of species j (i is the donor species and j is the receptor). The weights of the links are the number of pollen grains identified. Links between nodes represent pollen movement between species (heterospecific pollination) while self-links represent conspecific pollination. Heterospecific pollination occurs when pollinators visit plants of different species and is a cost on reproductive success (see more in Tur et al. (2016)). Because the relative flow of self and non-self pollen (con- vs. hetero-specific pollination) has ecological and evolutionary consequences, identifying higher-level modules of pollen flow and the roles of particular species in dominating this flow can provide a new perspective into the functioning of this community.

We mapped the pollen movement with and without self-links and found that the structure was considerably different. With self-links, Infomap identified 13 modules, and without self-links 7 (Figure 3a,b).
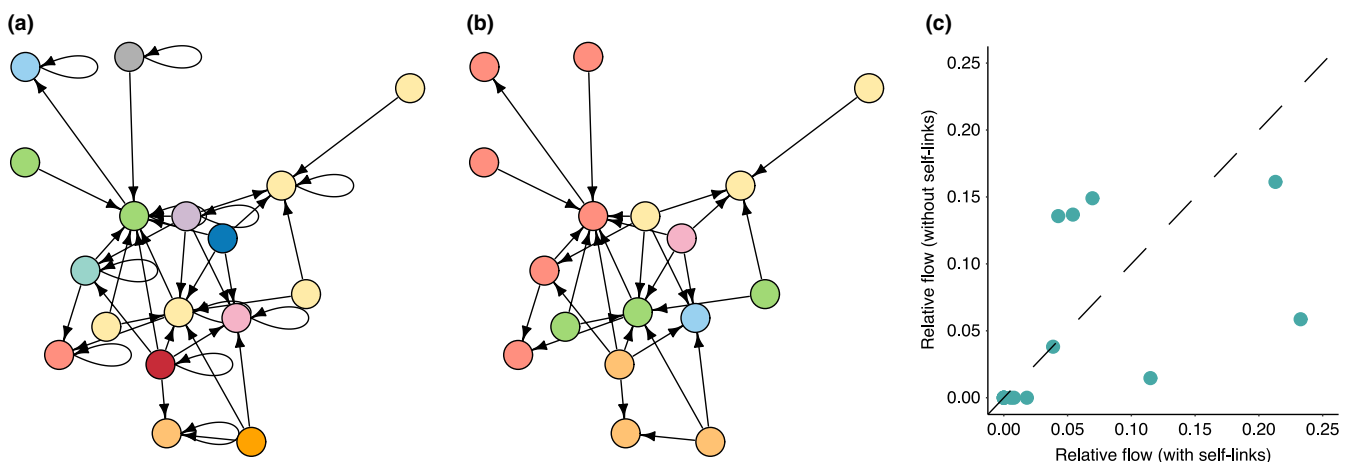

FIGURE 3 Modularity of a pollen movement network with and without self links. (a) and (b) The network is that of elevation 2000 from Tur et al. (2016). Link width (log-transformed) depicts the mean number of grains of a donor found on a recipient (arrow direction). Node background depicts the affiliation of nodes to modules when self-links are not included or included. The position of the nodes in the two networks is identical. It is clear that nodes that are grouped in the same module without self-loops are not grouped together when loops are included. For example, the six red nodes in panel (b) are grouped to six different groups in panel (a). (c) Comparison of the relative outgoing flow of nodes with and without self-links. The diagonal dashed line represents an equal contribution to flow when self-links are included or not. Plants that are above or below this line have a discrepancy in their contribution to flow, and therefore, structure

The increased number of modules with self-loops results from high conspecific pollen flows compared with heterospecific pollination. Because Infomap also quantifies the relative amount of flow at each node, this comparison allows us to look into the roles of individual species. For example, plants that have a large flow of conspecific pollination, but its pollen is also found on many other plants (outgoing flow) likely effect pollination success of other plants via generalist pollinators that visit them (Figure 3c).

3.4 | Temporal multilayer network

There are many types of multilayer networks in ecological systems and the ability of Infomap to integrate layers of different kinds opens up a range of possibilities for their analysis. Per our goal in this paper, we present an example of a temporal multilayer network, which represents flows over time. We use a host–parasite network recorded over 6 years, in which both interlayer and intralayer links are quantified (Pilosof et al. 2017). The dataset is included in infomapecology and we analyse it in two ways: First, we analysed the network using the existing interlayer links. We found that 47.4% of the modules persisted for all six layers while 7.89% appeared in only two layers. No module appeared in only a single layer (Figure 4a). This indicates that the grouping of species has a strong temporal component (although we cannot rule out biases due to uneven sampling across time). A second finding is that affiliation of species

to modules is flexible: Infomap assigned 21.8% of the species to at least two different modules during the 6 years. Infomap can assign a species to one module at one time-point (layer), and a to different module in the next layer because different state nodes represent the same species in different layers (Figure 4b). Biologically, flexibility in module affiliation in this system may capture interannual variation in host and parasite population dynamics.

To illustrate Infomap's capabilities to model interlayer links, in a second analysis, we ignored the interlayer links and used global relax rates to mimic the typical situation in which interlayer links have not been measured. We limited the relaxation of the random walker between layers to one layer forward, with no backwards relaxation because time has a direction. By systematically changing the value of r, we effectively examined the effect of increasing interlayer connectivity on the structure. The higher the relax rate, the more frequent the movement of the random walker between layers, tightening the connection between layers and potentially affecting structure (e.g. creating modules that persist for longer times). While we do detect variation in the number of modules, module composition and persistence, this variation is not considerable (Figure 5). Nevertheless, these results are specific for this network, and we recommend this kind of sensitivity analysis to choose the appropriate relax rate that best expresses the dynamics of the network. Moreover, the precise definition of interlayer links or the use of relax rates should be one of the primary considerations when analysing multilayer networks (Hutchinson et al., 2018; Pilosof et al., 2017).
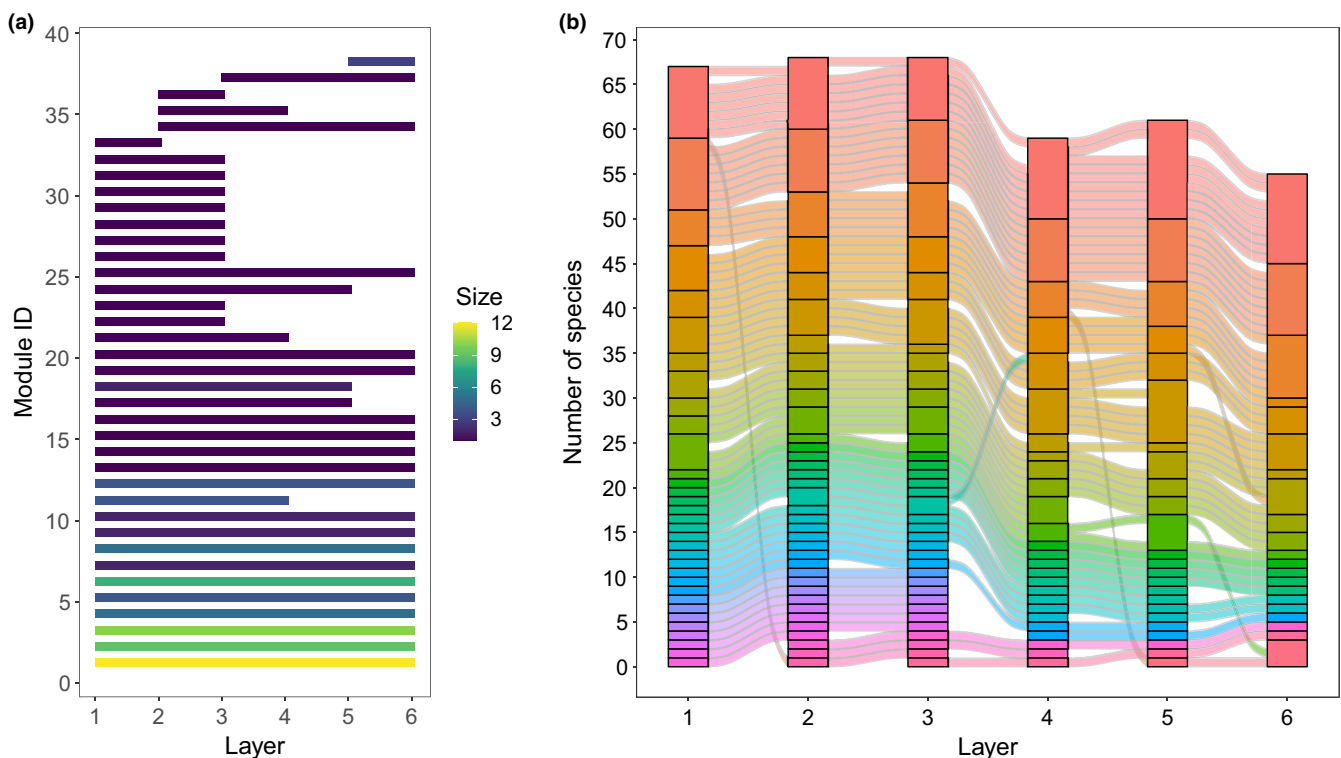

FIGURE 4 Modular structure of a temporal network. (a) The persistence of each module in time. Colours depict module size: the total number of species in each module. (b) An alluvial diagram for the flow of species from different modules among layers. Species are clustered in modules, presented in coloured blocks. Each line represents a species, and line colours correspond to the module in which the species originates
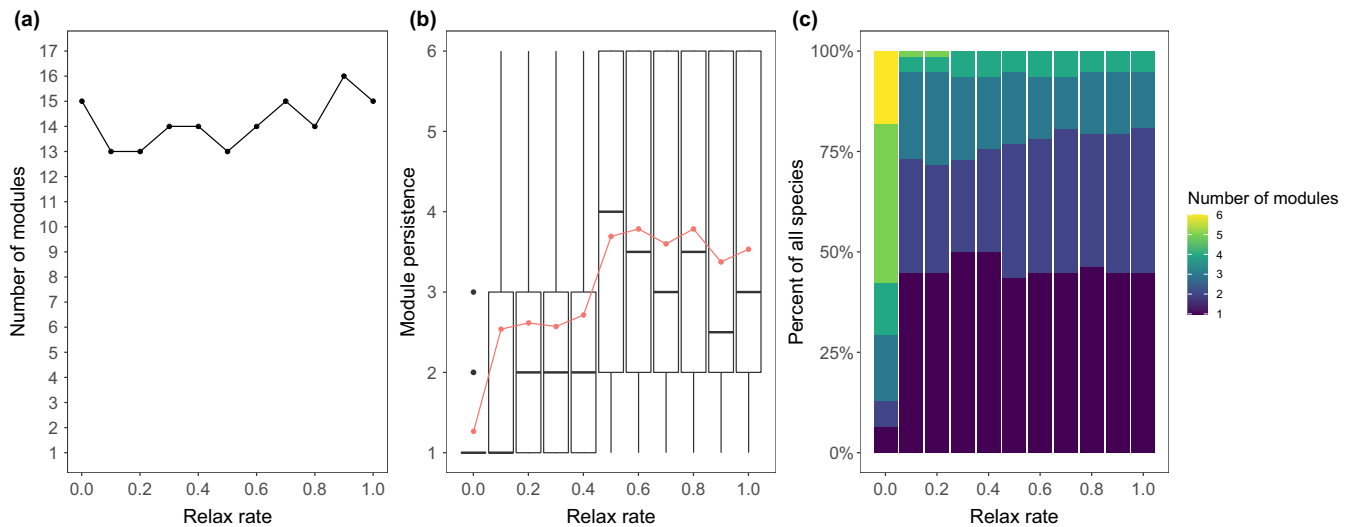
FIGURE 5 Structure as a function of increasing global relax rate. (a) Variation in the number of modules. (b) Distribution of module persistence (i.e. the number of layers throughout which a module exists). Boxplots represent the range, 95% quantiles and median (black line). The average is marked with red points and line. As expected, module persistence increases with increasing relax rate. (c) Species flexibility: The bars depict the percentage of species appearing in different numbers of modules

4 | CONCLUSIONS

Modularity is a cornerstone in ecological network analysis because it provides a higher-level simplification of complex ecological systems. Other community detection methods have also shown to be highly relevant for ecological networks, such as stochastic block models which can identify species that are performing unique roles in ecological communities (Sander et al., 2015). Another core concept in research on ecological networks is analyses of the dynamic processes taking place on the network (e.g. Otto et al. (2007)). Nevertheless, the algorithms commonly used in ecology focus on network topology and do not specifically view modules as dynamical building blocks. Here, we aimed to fill this gap by introducing Infomap to ecological research. Modules revealed by different methods (e.g. Infomap or Q) will highlight different aspects of networks (Rosvall et al., 2018; Table S1). Infomap, which seeks to coarse-grain the system's dynamics, will identify flow modules, which will likely better capture structural patterns important for the dynamics of the system than other methods.

Like any other method for detecting modules, Infomap cannot find a 'true' partitioning of a network (Peel et al., 2017) because such partitioning does not exist. We advocate the application of a method appropriate for the question (Table S1). For example, if the goal is to detect species that consume, or are consumed by, similar species, then stochastic block models (e.g. the group model (Allesina & Pascual, 2009)) are adequate (Table S1). When applied to undirected networks, Infomap provides accurate solutions according to benchmark tests. Nevertheless, Newman–Girvan modularity may be more appropriate if the goal is to detect topological groups by comparing to a random expectation.

The performance and flexibility of Infomap offer several advantages. It is an efficient and fast algorithm, which is particularly useful when analysing a large number of networks (e.g. during hypothesis testing) or large and dense networks. It is also flexible

and handles many network types. The possibility of using node attributes to inform the analysis is another advantage (Supporting Information Text 3.5), highly relevant for ecological data, in particular as all interactions rarely are captured in the data (Jordano, 2016). Additional information from other systems, such as information on the role of species traits (Eklöf et al., 2013) and taxonomic classification for interactions (Eklöf et al., 2012), or expert knowledge can then be valuable information for detecting modules.

Modularity has mainly been a theoretical construct in network ecology and empirical work is needed to complement the many generated hypotheses, including the effects on system stability (Dormann et al., 2017; Grilli et al., 2016). As an algorithm specifically designed for coarse-graining the dynamics and identifying flow modules, Infomap is highly relevant for analysing ecological networks (Calatayud et al., 2019; Edler et al., 2017; Pilosof et al., 2019). The incentives, guidelines and examples presented in this application paper provide a springboard to take maximum advantage of empirical work in network ecology.

AUTHORS' CONTRIBUTIONS

M.R. and S.P. conceived the study; C.F., A.E. and S.P. collected the data; C.F., D.E., A.E. and S.P. analysed the data; C.F., M.R. and S.P.

led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ORCID

Carmel Farage https://orcid.org/0000-0001-8646-8111
Daniel Edler https://orcid.org/0000-0001-5420-0591
Anna Eklöf https://orcid.org/0000-0003-2721-0051
Martin Rosvall https://orcid.org/0000-0002-7181-9940
Shai Pilosof https://orcid.org/0000-0003-0430-5568

REFERENCES

Aldecoa, R., & Marín, I. (2013). Exploring the limits of community detection strategies in complex networks. *Scientific Reports*, *3*, 2216.

Allesina, S., & Pascual, M. (2009). Food web models: A plea for groups. *Ecology Letters*, *12*, 652–662. https://doi.org/10.1111/j.1461-0248.2009.01321.x

Baird, D., & Ulanowicz, R. E. (1989). The seasonal dynamics of the chesapeake bay ecosystem. *Ecological Monographs*, *59*, 329–364. https://doi.org/10.2307/1943071

Bernardo-Madrid, R., Calatayud, J., González-Suárez, M., Rosvall, M., Lucas, P. M., Rueda, M., Antonelli, A., & Revilla, E. (2019). Human activity is altering the world's zoogeographical regions. *Ecology Letters*, *22*, 1297–1305.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, *2008*, P10008.

Bohlin, L., Edler, D., Lancichinetti, A., & Rosvall, M. (2014). Community detection and visualization of networks with the map equation framework. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 3–34). Springer International Publishing.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*, 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Calatayud, J., Bernardo-Madrid, R., Neuman, M., Rojas, A., & Rosvall, M. (2019). Exploring the solution landscape enables more reliable network community detection. *Physical Review E*, *100*(5), 052308. https://doi.org/10.1103/PhysRevE.100.052308

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*, 011027.

Dormann, C. F., Fründ, J., Blüthgen, N., & Gruber, B. (2009). Indices, graphs and null models: Analyzing bipartite ecological networks. *The Open Ecology Journal*, *2*, 7–24. https://doi.org/10.2174/1874213000902010007

Dormann, C. F., Fründ, J., & Schaefer, H. M. (2017). Identifying causes of patterns in ecological networks: Opportunities and limitations. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 559–584. https://doi.org/10.1146/annurev-ecolsys-110316-022928

Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2017). Infomap bioregions: Interactive mapping of biogeographical regions from species distributions. *Systematic Biology*, *66*, 197–204.

Eklöf, A., Helmus, M. R., Moore, M., & Allesina, S. (2012). Relevance of evolutionary history for food web structure. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1588–1596.

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti, M., Guimarães, P. R., Lomáscolo, S. B., Martn González, A. M., Pizo, M. A., Rader, R., Rodrigo, A., Tylianakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks. *Ecology Letters*, *16*, 577–583.

Fletcher Jr, R. J., Revell, A., Reichert, B. E., Kitchens, W. M., Dixon, J. D., & Austin, J. D. (2013). Network modularity reveals critical scales for connectivity in ecology and evolution. *Nature Communications*, *4*, 2572.

Ghasemian, A., Hosseinmardi, H., & Clauset, A. (2019). Evaluating overfit and underfit in models of network community structure. *IEEE Transactions on Knowledge and Data Engineering*, *32*, 1722–1735. https://doi.org/10.1109/TKDE.2019.2911585

Gilarranz, L. J., Rayfield, B., Liñán-Cembrano, G., Bascompte, J., & Gonzalez, A. (2017). Effects of network modularity on the spread of perturbation impact in experimental metapopulations. *Science*, *357*, 199–201. https://doi.org/10.1126/science.aal4122

Grilli, J., Rogers, T., & Allesina, S. (2016). Modularity and stability in ecological communities. *Nature Communications*, *7*, 12031.

Guimerà, R., & Nunes Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature*, *433*, 895–900. https://doi.org/10.1038/nature03288

Hanski, I., & Gilpin, M. (1991). Metapopulation dynamics: Brief history and conceptual domain. *Biological Journal of the Linnaean Society*, *42*, 3–16.

Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, *5*, 109–137. https://doi.org/10.1016/0378-8733(83)90021-7

Hutchinson, M. C., Bramon Mora, B., Pilosof, S., Barner, A. K., Kéfi, S., Thébault, E., Jordano, P., & Stouffer, D. B. (2018). Seeing the forest for the trees: Putting multilayer networks to work for community ecology. *Functional Ecology*, *1*, 55.

Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, *30*, 1883–1893.

Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *80*, 056117.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *69*, 026113.

Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 19891–198916.

Olesen, J. M., Bascompte, J., Elberling, H., & Jordano, P. (2008). Temporal dynamics in a pollination network. *Ecology*, *89*, 1573–1582. https://doi.org/10.1890/07-0451.1

Otto, S. B., Rall, B. C., & Brose, U. (2007). Allometric degree distributions facilitate food-web stability. *Nature*, *450*, 1226–1229. https://doi.org/10.1038/nature06359

Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances*, *3*, e1602548. https://doi.org/10.1126/sciadv.1602548

Pilosof, S., Alcalá-Corona, S. A., Wang, T., Kim, T., Maslov, S., Whitaker, R., & Pascual M. (2020). The network structure and eco-evolutionary dynamics of CRISPR-induced immune diversification. *Nature Ecology & Evolution*, *4*(12), 1650–1660. https://doi.org/10.1038/s41559-020-01312-z

Pilosof, S., Farage, C., & Edler, D. (2021). Ecological-Complexity-Lab/infomap_ecology_package: Release along with submission to CRAN (Version V1.0.0b). *Zenodo*, https://doi.org/10.5281/zenodo.4535342

Pilosof, S., He, Q., Tiedje, K. E., Ruybal-Pesántez, S., Day, K. P., & Pascual, M. (2019). Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in plasmodium falciparum. *PLoS Biology*, *17*, e3000336.

Pilosof, S., Porter, M. A., Pascual, M., & Kéfi, S. (2017). The multilayer nature of ecological networks. *Nature Ecology & Evolution*, *1*, 0101.

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2010). The map equation. *The European Physical Journal Special Topics*, *178*, 13–23.

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 1118–1123.

Rosvall, M., & Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE*, *6*, e18209.

Rosvall, M., Delvenne, J. C., Schaub, M. T., & Lambiotte, R. (2018). Different approaches to community detection. In P. Doreian, V. Batagelj, & A. Ferligoj (Eds.), *Advances in network clustering and block-modeling* (pp. 71–87). Wiley.

Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D., & Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, *5*, 4630.

Sander, E. L., Wootton, J. T., & Allesina, S. (2015). What can interaction webs tell us about species roles? *PLoS Computational Biology*, *11*, e1004330.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Labs Technical Journal*, *27*, 379–423.

Stouffer, D. B., & Bascompte, J. (2011). Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 3648–3652.

Thébault, E. (2013). Identifying compartments in presence-absence matrices and bipartite networks: Insights into modularity measures. *Journal of Biogeography*, *40*, 759–768. https://doi.org/10.1111/jbi.12015

Trøjelsgaard, K., Jordano, P., Carstensen, D. W., & Olesen, J. M. (2015). Geographical variation in mutualistic networks: Similarity, turnover and partner fidelity. *Proceedings of the Royal Society B*, *282*, 20142925. https://doi.org/10.1098/rspb.2014.2925

Tur, C., Sáez, A., Traveset, A., & Aizen, M. A. (2016). Evaluating the effects of pollinator-mediated interactions using pollen transfer networks: Evidence of widespread facilitation in south andean plant communities. *Ecology Letters*, *19*, 576–586. https://doi.org/10.1111/ele.12594

SUPPORTING INFORMATION

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