HAL
open science

A methodological framework to analyse determinants of host–microbiota networks, with an application to the relationships between Daphnia magna' s gut microbiota and bacterioplankton

François Massol, Emilie Macke, Martijn Callens, Ellen Decaestecker

▶ **To cite this version:**

François Massol, Emilie Macke, Martijn Callens, Ellen Decaestecker. A methodological framework to analyse determinants of host–microbiota networks, with an application to the relationships between Daphnia magna' s gut microbiota and bacterioplankton. Journal of Animal Ecology, In press, 10.1111/1365-2656.13297 . hal-02942286

HAL Id: hal-02942286

https://hal.science/hal-02942286v1

Submitted on 6 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A methodological framework to analyse determinants of host-microbiota networks, with an application to the relationships between *Daphnia magna*'s gut microbiota and bacterioplankton

François Massol[1,2], Emilie Macke[3], Martijn Callens[3,4], Ellen Decaestecker[3]

[1] UMR 8198 Evo-Eco-Paleo, SPICI group, Univ. Lille, F-59000 Lille, France

[2] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France

[3] Laboratory of Aquatic Biology, KU Leuven (Kulak), Dept of Biology, E. Sabbelaan 53, BE-8500, Kortrijk, Belgium

[4] Centre d'Ecologie Fonctionnelle et Evolutive, UMR CNRS 5175, 34293 Montpellier, France

Corresponding author: F. Massol, Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Center for Infection and Immunity of Lille, F-59000 Lille, France; email: francois.massol@univ-lille.fr

Running headline: Analyzing host-microbiota networks

Abstract

1. The past thirty years have seen both a surge of interest in assessing ecological interactions using tools borrowed from network theory and an explosion of data on the occurrence of microbial symbionts thanks to next-generation sequencing. Given that classic network methods cannot currently measure the respective effects of different environmental and biological drivers on network structure, we here present two methods to elucidate the determinants of bipartite interaction networks.
2. The first method is based on classifications and compares communities within networks to the grouping of nodes by treatment or similar controlling groups. The second method assesses the link between multivariate explanatory variables and network structure using redundancy analyses after singular value decomposition. In both methods, the significance of effects can be gauged through two randomizations.
3. Our methods were applied to experimental data on *Daphnia magna* and its interactions with gut microbiota and bacterioplankton. The whole network was affected by *Daphnia*'s diet (algae and/or cyanobacteria) and sample type, but not by *Daphnia* genotype. At coarse grains, bacterioplankton and gut microbiota communities were different. At this scale, the structure of the gut microbiota-based network was not linked to any explanatory factors, while the bacterioplankton-based network was related to both *Daphnia*'s diet and genotype. At finer grains, *Daphnia*'s diet and genotype affected both microbial networks, but the effect of diet on gut microbiota network structure was mediated solely by differences in microbial richness. While no reciprocal effect between the microbial communities could be found, fine-grained analyses presented a more nuanced picture, with bacterioplankton likely affecting the composition of the gut microbiota.
4. Our methods are widely applicable to bipartite networks, can elucidate both controlled and environmental effects in experimental setting using a large amount of sequencing data, and can tease apart reciprocal effects of networks on one another. The two-fold approach we propose has the advantage of being able to tease apart effects at different scales of network structure, thus

allowing for detailed assessment of reciprocal effects of linked networks on one another. As such, our network methods can help ecologists understand huge datasets reporting microbial co-occurrences within different hosts.

Keywords

bipartite interaction networks ; *Daphnia magna* ; gut microbiota ; modular networks ; random dot-product graph model

Data archiving

The dataset and the R codes will be archived on Zenodo.

Introduction

The past thirty years have seen the rapid development of ecological interaction network research, with the parallel growth of datasets (e.g. the interaction web database, https://www.nceas.ucsb.edu/interactionweb/ or the mangal repository, https://mangal.io; see also Bohan *et al.*, 2016; Poisot *et al.*, 2016) and methods to analyze them (Jordano, 1987; Memmott, 1999; Stouffer *et al.*, 2007; Bascompte & Stouffer, 2009; Kissling *et al.*, 2012; Stouffer *et al.*, 2012; Weitz *et al.*, 2013; Nogales *et al.*, 2016; García-Callejas, Molowny-Horas & Araújo, 2018; Joffard *et al.*, 2019) with a view to describe regularities in species interactions (e.g. degree distributions, modules, motifs…) and ultimately to explain why some species interact and others do not. Many early analyses focused on network connectance (i.e. the density of the graph) and the distribution of species degrees within networks (e.g. Jordano, 1987; Dunne, Williams & Martinez, 2002), spurred by the long debate between the 'constant degree' and 'constant connectance' predictions from the cascade and niche food web models, respectively (Cohen & Briand, 1984; Williams & Martinez, 2000). Following the pioneering work of Bascompte et al. (2003), other network metrics such as nestedness and modularity have become the subject of many ecological studies (Lewinsohn *et al.*, 2006; Olesen *et al.*, 2007; Fortuna *et al.*, 2010; Thébault & Fontaine, 2010). Recently, assessments of ecological networks have turned towards more sophisticated metrics and models encompassing e.g. motif counts, block models, degree equitability and abundance-corrected measures of specialization (Stouffer *et al.*, 2005; Blüthgen, Menzel & Blüthgen, 2006; Leger, Daudin & Vacher, 2015). However, despite a few notable exceptions (Bartomeus, 2013; Bartomeus *et al.*, 2016; CaraDonna *et al.*, 2017; Joffard *et al.*, 2019; de Manincor *et al.*, in press), ecological network analyses are still not assessing the amount of network variation driven by different environmental and biological factors.

In parallel with the increasing interest in ecological networks, the development of next-generation molecular ecology methods has set up the stage for an explosion of the number of datasets describing microbial interaction networks, from planktonic networks, e.g. informed by the Tara scientific cruise (Lima-Mendez *et al.*, 2015; Guidi *et al.*, 2016), to plant-fungus antagonistic (Vacher, Piou & Desprez-

Loustau, 2008) or mutualistic (Encinas-Viso *et al.*, 2016) interaction networks, phage-bacteria infection networks (Weitz *et al.*, 2013), or mammal species-gut microbiota associations (Ley *et al.*, 2008). Although inferring true interactions from co-occurrence is a difficult endeavour (Vacher *et al.*, 2016; Bohan *et al.*, 2017; Derocles *et al.*, 2018) – indeed, neither do co-occurrences necessarily imply interaction nor does the absence of co-occurrence imply the absence of interaction –, the analysis of host-microbe association networks can still benefit from the use of ecological network methods. In particular, since host-microbe associations are much more amenable to controlled experiments than marine food webs or plant-pollinator networks, they can provide a good starting point to test methods aimed at elucidating the drivers of network structure because drivers can be varied independently, thus removing the possibility of confounding effects.

As stated above, classic network methods do not measure the respective effects of different environmental and biological drivers on network structure. A few methods have been proposed (see e.g. Kamenova *et al.*, 2017 for a short review of existing models) to assess network structure as the result of latent or explicit traits (such as organism size), sometimes combining the information provided by traits with that provided by phylogenies or geographical distributions of species (Rohr *et al.*, 2010; Gravel *et al.*, 2013; Ovaskainen *et al.*, 2016; Rohr *et al.*, 2016), but not in an integrative framework allowing all types of external factors to be tested. However, one very promising method (Dalla Riva & Stouffer, 2016) based on low-dimension embedding of adjacency matrices through singular value decomposition allows the partitioning of network ‘inertia’ through the use of classic multivariate redundancy analyses (Sabatier, Lebreton & Chessel, 1989; Borcard, Legendre & Drapeau, 1992; Dray, Legendre & Peres-Neto, 2006; Peres-Neto *et al.*, 2006) and has been successfully applied to the study of orchid-pollinator interactions across Europe (Joffard *et al.*, 2019). In parallel, the study of multi-layer network structure, and notably the search for congruence between ‘mesoscale’ structures (i.e. modules or blocks) in two paired networks, has led to the development of a comparison method based on classifications obtained by modularity optimization and the use of classification congruence indices (Astegiano, Altermatt & Massol, 2017).

Here we propose two methods to assess the effects of different drivers on the structure of host-microbiota interaction networks. The first method compares communities within networks (*i.e.* groups of nodes which interact more between them than with nodes from other communities) to the grouping of nodes by external factors (e.g. treatments) in order to assess whether a single factor explains a significant part of the network structure. When two or more factors are considered, a similar approach is developed to adapt canonical correspondence analysis to the exploration of communities. The second method transforms host-microbiota networks into datasets amenable to redundancy analyses. In both cases, gauging the significance of external factors can be performed through two different randomizations, which can help tease apart richness effects (modalities of the factors control the number of links per node, but not the specificity of the links) from affinity effects. We illustrate the potential of both methods using the results of an experiment on *Daphnia magna* and its interactions with gut microbiota and bacterioplankton under controlled diets.

Materials & Methods

Host-microbiota data as interaction networks

In the following, we will consider bipartite networks, i.e. networks involving two disjoint sets of nodes, called levels, with edges only connecting nodes from two different levels. A host-microbiota association network is a bipartite network, with host populations or individuals forming one level, and microbial species, the other one. A bipartite network, with n nodes in the first level and p nodes in the second level, can be mathematically represented using an incidence matrix $\mathbf{B}$ of dimensions n by p. We will generally assume that the networks under study are not weighted, i.e. an existing link between two nodes is coded as "1" and an absent link is noted "0". Indeed, since microbial sequence counts can vary widely between samples (e.g. due to PCR amplification heterogeneity), only relative microbial abundances can be obtained per host, which prevents the absolute quantification of links between host and microbial nodes (Amend, Seifert & Bruns, 2010; McMurdie & Holmes, 2014; Thomas *et al.*, 2016). For datasets able to more accurately quantify link weights, most of the analyses presented below could be adapted to weighted networks, e.g. by using block search instead of community search (Leger, Daudin & Vacher, 2015), and are further tackled in the Discussion.

Community detection

A community within a network is a group of nodes which interact more between them than with nodes from other communities. Different algorithms exist to find communities (Fortunato, 2010). Of late, bipartite network studies in ecology have narrowed down their interest to two procedures (Leger, Daudin & Vacher, 2015): modularity optimization algorithms and latent block models. Modularity is a network metric based on the amount of interactions within communities compared to what would be expected from the number of interactions per node, and modularity increases when more interactions occur within communities than between them (Newman, 2006b). Block models apply goodness-of-fit procedures to find the best sets of nodes so that the probability of finding an edge (or the value of an edge in weighted networks) between two random nodes is determined by the sets the two nodes belong to (Govaert & Nadif, 2008). Leger et al. (2015) have shown that some modularity-optimizing

algorithms are best suited to discover communities in non-weighted bipartite networks (the leading-eigenvector method of Newman, 2006a in particular), while block models reign undisputed when the task is to find communities within weighted bipartite networks. Given our focus on non-weighted networks, we will assume that community detection is performed using the leading-eigenvector modularity-optimizing algorithm. In practice, we will use the 'cluster_leading_eigen' function in the R package 'igraph' (Csardi & Nepusz, 2006).

Null models for bipartite networks

A striking feature of most statistics computed on networks is that they have no expected asymptotic distribution. To cope with this, network statistics are tested against their distribution after proper randomization of the initial network, i.e. using a null model. Null models assume that certain features of the network must be kept in all randomizations, but are designed so that the distribution of networks obeying these constraints is sampled as uniformly as possible. It is possible to imagine a wide array of null models for networks (Orsini *et al.*, 2015). In the context of host-microbiota network, the most simple choice is to assume that host and microbial nodes are given their degrees (i.e. the number of nodes they interact with), but may interact randomly. Producing this null model, called the configuration model, can be done easily for bipartite networks using the recently published 'curveball' algorithm (Strona *et al.*, 2014), with functions 'simulate' and 'nullmodel' of R package 'vegan' (Oksanen *et al.*, 2018). In the following, each 'curveball'-based test was performed using 10,000 simulated networks.

Approximating networks

To understand the effects of external variables on network structure, a first step can be to approximate the incidence matrix using some simple equation, preferably related to well-known statistical procedures. The two methods presented below are based on two such approximations (Fig. 1). As the two approximations do not focus on the same underlying network structures, their respective analyses

can lead to complementary, and sometimes seemingly discrepant, results. However, such discrepancies can be easily explained by differences in the approximation method.

The first approximation consists in realizing that an $n \times p$ incidence matrix **B** can be decomposed into a matrix product involving the first-level and second-level community-membership matrices, respectively $\mathbf{M}_1$ and $\mathbf{M}_2$ (filled with 1's and 0's to indicate in which community each node belongs) and the reduced matrix defining interaction density between communities **B**':

$$\mathbf{B} \approx \mathbf{M}_1.\mathbf{B}'.\mathbf{M}_2 \qquad (1)$$

where dots represent matrix products. Binary matrix $\mathbf{M}_1$ is $n \times b_1$, where $b_1 \leq n$ is the number of communities found among the n nodes from the first level, and we assume that each node belongs to a single community and each community has at least one node; binary matrix $\mathbf{M}_2$ is $b_2 \times p$, where $b_2 \leq p$ is the number of communities found among the p nodes from the second level; the reduced matrix **B**' is $b_1 \times b_2$. Its element b'_{ij} yields the probability that a node within community i of the first level interacts with a node within community j of the second level. The approximation given by equation (1) becomes an equality when every node has its own community and becomes quite a poor fit when there is only one parameter to define probabilities of interactions. Given equation (1), an option to study the effects of external factors on network structure is to study their effects on node memberships, i.e. on matrices $\mathbf{M}_1$ and $\mathbf{M}_2$. Because nodes within each community can have different degrees, this approximation can partially help tease apart effects due to node degrees from those due to node membership in a given community, to the extent that the nodes belonging to the different communities have similar distributions of degrees.

The second approximation we propose follows the logic of Dalla Riva and Stouffer (2016) based on the random dot-product graph model (RDPG, Young & Scheinerman, 2007). The singular value decomposition of an $n \times p$ incidence matrix **B** can be written as:

$$\mathbf{B} = \mathbf{U}.\mathbf{D}.\mathbf{V}^T \qquad (2)$$

where **U** and **V** are orthogonal matrices, $\mathbf{V}^T$ denotes the transposed version of **V**, and the square matrix **D** is diagonal and its values are the singular values of matrix **B** (all non-negative), usually sorted in

decreasing order. As Dalla Riva and Stouffer (2016) note, a useful approximation of the network can be obtained by finding the square-root of **D**, noted **S**, and define a number of 'latent traits' q less than the number of singular values, a matrix **L** as the first q columns of **U.S** and a matrix **R** as the first q columns of **V.S**, so that:

$$\mathbf{B} \approx \mathbf{L.R}^T \tag{3}$$

The approximation given by equation (3) becomes an equality when q equals the number of singular values. When q = 1, each node at both levels is exactly defined by a single value, which mimics the effect of heterogeneity of degrees among nodes (and leads to quite a poor fit for approximation [3]).

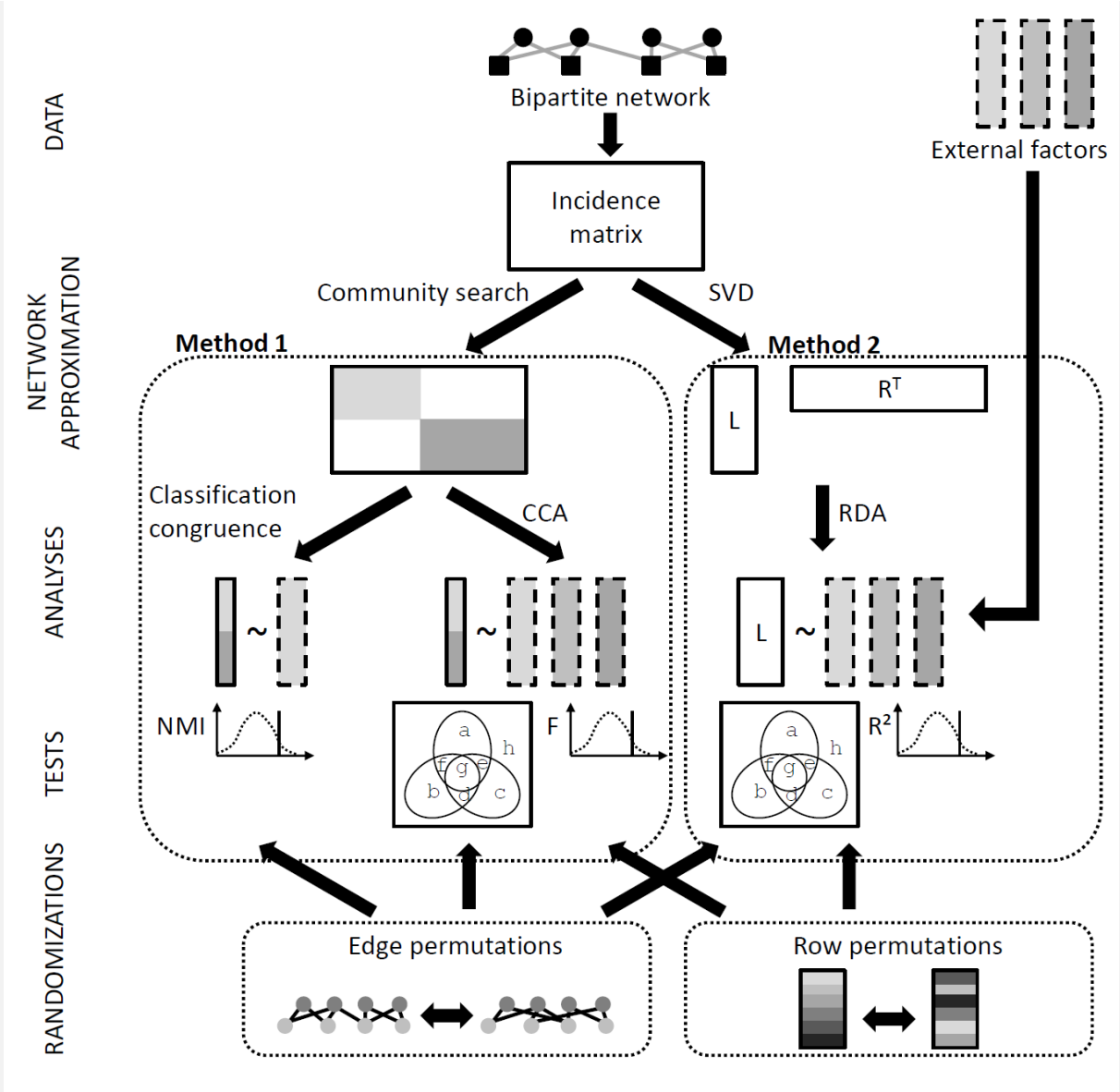


Fig. 1 – Summary of the proposed methods. Starting from an incidence matrix describing a bipartite network, one can either (i) perform a community search, then work on node communities by assessing their congruence with other classifications or perform a canonical correspondence analysis (CCA) with respect to several external factors (method 1), or (ii) approximate the incidence matrix as the product of two reduced ones through a singular value decomposition (SVD) and then analyze these reduced matrices using a redundancy analysis (RDA; method 2). In both cases, the results of analyses can be tested twice: (i) through row permutations, one can assess whether effects would have been expected from the imbalance and correlations between external factors (RDA and CCA); (ii) through edge permutations on the initial network, one can test whether an effect significant for the first test is only due to differences in numbers of interactions between factor levels, if the second test is not significant (a richness effect, e.g. a diet effect because *Daphnia* fed with *Scenedesmus* interact with more bacteria species) or, if the second is also significant, is due to an affinity effect (i.e. factor levels selectively associated with certain interactions).

Method 1: node classification-based tests

To assess the effects of categorical drivers on network organization, a first method is to test the congruence of node classifications obtained through community-search algorithms with those associated with external categorical variables, i.e. to study the links between matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ from the previous section with matrices describing external categorical factors. Such a method is useful for analyzing the results of controlled experiments since external categorical variables then amount to treatments. Any grouping of the nodes (host, microbiota, or both) is effectively a classification in the statistical sense, which can be compared with other classifications of the same data (Danon *et al.*, 2005). Following Astegiano et al. (2017), we propose to use the Normalized Mutual Information index (NMI) to gauge the congruence of two classifications. The NMI takes values between 0 and 1, 0 indicating no congruence and 1, perfect congruence. To test for the significance of a given NMI between two classifications, at least one of which being the classification of network nodes in communities, the network is randomized using the 'curveball' algorithm (Strona *et al.*, 2014). This computation of NMI, implemented using the function 'compare' in the R package 'igraph' (Csardi & Nepusz, 2006), can help assess the effect of a single external factor on network structure.

To extend the same logic to multiple factors, we propose to use Canonical Correspondence Analysis (CCA, ter Braak, 1986) on the **M** matrices. CCA decomposes the variation of the explained factor (here, the classification of nodes) through projections into the eigen-spaces induced by the external factors. It can classically test the significance of a given 'fraction' (e.g. chi square explained by factors X or Y once the effect of Z has been removed) by comparing the obtained F statistic to those yielded by randomizations of data rows (Peres-Neto *et al.*, 2006). This first randomization tests whether an effect, e.g. host diet, is more related to network structure than expected by randomly assigning its values. However, we also test whether an effect that is deemed significant following the first test is purely due to heterogeneity in node degrees between communities (i.e. not significantly different F from edge-permuted expectation; richness effect) or not (affinity effect) following the configuration model. This second randomization tallies up the probability that randomizing the network, keeping the number of links per node constant, would produce effects as strong as those obtained with the real

network. To do so, we use the 'curveball' algorithm and compare the F-statistics obtained when performing CCA on the observed vs. simulated datasets. F-statistics larger than 95% of the simulated F's for the same fraction indicate an affinity effect. Performing CCA can be done using the function 'cca' in the R package 'vegan' (Oksanen *et al.*, 2018). Using CCA to assess the covariation of network communities with external factors is but one of the many existing multivariate methods (Legendre & Gallagher, 2001; Blanchet *et al.*, 2014) – other potentially useful approaches are tackled in the Discussion. However, with all those approaches, the same underlying process (approximation by communities, quantification of explained fractions, two randomization-based tests) should be applied.

Method 2: singular value decomposition-based tests

As mentioned above, the approximation of a network by community memberships overlooks differences in degrees among nodes belonging to the same community. Since these differences can also hold some of the underlying network structure, we propose to also model the effects of external variables on network structure using the RDPG decomposition proposed by Dalla Riva and Stouffer (2016; equation [3]). Following equation (3), a given $n \times p$ bipartite network can be approximated as two matrices (**L** and **R**) with a low number of columns and as many rows as nodes (n in **L**, p in **R**). Matrices **L** and **R** can be analysed through a Redundancy Analysis (RDA) to gauge how much variation among rows is explained by external variables, similarly to what is performed for CCA (i.e. quantifying the variation explained by each fraction such as that explained by factors X or Y once the effect of Z has been removed, Joffard *et al.*, 2019). In RDA, variation is understood in the classic sum-of-square sense and can be quantified using adjusted R^2 (Peres-Neto *et al.*, 2006). Because the information stored in matrices **L** and **R** is represented by real numbers which do not correspond to presences or absences, multivariate analyses such as RDA or distance-based RDA (Legendre & Gallagher, 2001) applied to these matrices do not suffer from the 'double zero' problem (Legendre & Legendre, 2012). This means that different multivariate approaches similar to classic RDA could be applied instead (Blanchet *et al.*, 2014), but lacking a clear rationale for favouring one over the others, we will tackle these other approaches in the Discussion.

As for CCA, the classic test of significance of a 'fraction' is based on the randomization of dataset rows. Again, as in the previous method, we complement this first test by randomizing edges and gauging whether the adjusted R^2 obtained using the true data is higher than 95% of the simulated ones. A fraction that would be doubly significant, i.e. with adjusted R^2 higher than those expected from both the row and edge permutations, would indicate an affinity effect, which cannot be solely interpreted as stemming from heterogeneity in node degrees; by contrast, an effect deemed significant on the first test but not on the second one would signal a richness effect, i.e. the differences in connections among nodes explained by this effect could be simply understood as differences in the numbers of connections per node, not the identity of the nodes they are connected to (Joffard *et al.*, 2019).

One issue arises in the case of the method presented here: how can one choose the number of vectors to keep after SVD? For instance, assuming that **L** is the focus of an RDA, leaving **L** with the first 10 or 100 columns will lead to different sums of squares to explain and, hence, to different R^2 statistics. Another similar problem arises when e.g. one wants to explain a SVD-based matrix $\mathbf{L}_1$ using another SVD-based matrix $\mathbf{L}_2$ (e.g. explaining plant-pollinator associations using plant-herbivore associations): how many vectors should one keep in $\mathbf{L}_2$? These two dimensionality problems can be solved, but in different ways.

The number of vectors to retain in the explained table is really a choice of object to model – with more vectors retained, one obtains a finer approximation of the network. In the *Daphnia*-microbiota example given below, as we focus on mesoscale network structure (i.e. communities), we select the number of vectors to retain by looking at the congruence between communities of approximated networks with those of the original network using the NMI. Approximated networks, in this case, need to be binary, so we resort to transforming the **L.R** product from equation (3) into a binary incidence matrix using a threshold value (**L.R** values over threshold yield network edges). The threshold value can be obtained by maximizing the sum of sensitivity and specificity of the approximation, using function 'optim.thresh' in the R package 'SDMTools' (Van Der Wal *et al.*, 2014). It should be repeated that our choice of criterion is here arbitrary – if this method were to be used to qualify other network structures, such as e.g. motif relative frequencies, one could come up with other criteria to

optimize to find the "best vectors" (e.g. minimize Mahalanobis distance between motif relative frequency vectors of observed vs. approximated networks). Dealing with all possible focal structures and the way to best represent them using SVD vectors is, however, beyond the point of this study.

Regarding the number of vectors/columns in the explanatory tables (obtained from networks or otherwise extracted in a way that allows choosing which vectors to use or not, e.g. like Moran Eigenvector Maps, Dray et al. 2006), we use forward selection with double stopping criteria (p-value <0.05 and adjusted R^2 less than that found with all vectors), as described by Blanchet et al. (2008) and advocated by Bauman et al. (2018). This was practically implemented borrowing from the code proposed by Bauman et al. (2018), using functions 'RsquareAdj' and 'forward.sel' from the R package 'adespatial' (Dray *et al.*, 2019).

Application: Daphnia magna gut microbiota and bacterioplankton data

The methods presented here were applied to experimental data obtained for another study (Macke *et al.*, 2020).

***Daphnia* genotypes**. Nine *Daphnia magna* genotypes (G1 to G9) were used in the experiment. G1, G4 and G9 were obtained from resting eggs sampled in three sediment core sections in a 8.7 ha shallow man-made pond located in Oud Heverlee, Belgium (Stoks *et al.*, 2016). G2 was isolated from Bysjön lake in Sweden. G3 was hatched from sediment of a small, fishless and mesotrophic pond located near Knokke, Belgium (51°20'05.62" N, 03°20'53.63" E). G5-G8 were hatched from sediment of a eutrophic pond containing fish and located in Heverlee, Belgium (50°51'47.82" N, 04°43'05.16" E).

Preparation of diets. The unicellular green alga *Scenedesmus obliquus* (hereafter called *Scenedesmus* or abbreviated as "S"; strain CCAP 276/3A, provided by the Culture Collection of Algae and Protozoa, UK) and the unicellular cyanobacteria *Microcystis aeruginosa* (hereafter called *Microcystis* or abbreviated as "M"; strain PCC 7806, provided by the Pasteur Culture Collection, Institut Pasteur, Paris, France) were used as food for *Daphnia*. The *Microcystis* strain used in the present study produces toxins and bioactive compounds such as microcystins (Rohrlack *et al.*, 2001). *Scenedesmus* and *Microcystis* were cultivated under sterile conditions at 20±2°C and a light:dark cycle of 16:8 h, in

2L glass bottles with constant stirring and aeration. Filters (0.22 µm) were placed at the input and the output of the aeration system to avoid bacterial contamination. Algae were harvested weekly in early stationary phase. Axenity was checked on LB medium agar plates.

Experiment. For each of the nine *Daphnia* genotypes (G1-G9), three maternal lines were cultured under standardized conditions (2 L jars, 19±1°C; 16:8 h light:dark cycle). They were fed daily with saturating amounts of *Scenedesmus*. Medium was refreshed once a week. When a sufficient number of individuals was reached, 120 juveniles were sampled from each maternal line and divided into two 2L experimental jars (each containing 60 individuals, split-brood design). The first jar was fed a *Scenedesmus* diet (100% *Scenedesmus*), while the second was fed a *Microcystis* diet, composed of a mixture of *Microcystis* and *Scenedesmus* in a proportion adjusted so as to avoid too high mortality in *Daphnia*, but always ranging between 50 and 80% of *Microcystis* (same ratio in all jars). In total there were 54 populations (9 genotypes × 2 diets × 3 replicates). Food was provided every other day with a final carbon concentration of approximately 1.5 mg $C.L^{-1}$. Medium was refreshed every other week. Water from a pond on the campus (Kortrijk, Belgium, 50°48'30.3"N, 3°17'38.0"E) was added to the ADaM medium (15% of the final volume) every other week in order to provide a large diversity of bacteria and optimal growth conditions for the *Daphnia*.

Sampling microbiotas. After 1.5 years (*circa* 58 generations) of exposure to the two types of diet, bacterioplankton and gut microbiota compositions were assessed through next-generation sequencing of 16S rRNA. To obtain gut microbiota samples, 20 adult *Daphnia* were collected from each population and placed in autoclaved ADaM medium for 24h to reduce the amount of contaminating food particles within the gut (Callens *et al.*, 2016). *Daphnia* guts were subsequently extracted using sterilized dissecting needles under a stereomicroscope and placed in 1.5 ml Eppendorf tubes containing 10 µl of deionized sterile water. For bacterioplankton characterization, 100 ml of medium was sampled from each population and filtered with a 0.22 µm syringe filter. The filter was subsequently placed in a 1.5 ml Eppendorf tube. Gut microbiota and bacterioplankton samples were immediately placed at -20°C until further processing.

Determination of microbiota composition. DNA was extracted using a PowerSoil DNA isolation kit (MO BIO laboratories) and dissolved in 20 µL MilliQ water. The full length 16S rRNA gene was

amplified with primers 27F and 1492R on 10 ng of template (94°C - 30s; 50°C - 45s; 68°C - 90s; 30 cycles) using a high-fidelity Pfx polymerase (Life technologies). PCR products were subsequently purified using the QIAquick PCR purification kit (Qiagen). To obtain dual-index amplicons of the V4 region, a second amplification was performed on 5 μL PCR product using primers 515F and 806R for 30 cycles (94°C - 30s; 55°C - 30s; 68°C - 60s). Both primers contained an Illumina adapter and an 8-nt barcode at the 5'-end. For each sample, PCRs were performed in triplicate, pooled and gel-purified using the QIAquick gel extraction kit (Qiagen). An equimolar library was prepared by normalizing amplicon concentrations with a SequalPrep Normalization Plate (Applied Biosystems) and subsequent pooling. Amplicons were sequenced using a v2 PE500 kit with custom primers on the Illumina Miseq platform (KU Leuven Genomics Core) producing 2 × 250-nt paired-end reads. Sequence reads were processed using R package 'phyloseq', following Callahan *et al.* (2016b). Sequences were trimmed (the first 10 nucleotides and from position 190 onwards were removed) and filtered (maximum of 2 expected errors per read) on paired ends jointly. Sequence variants were inferred using the high-resolution DADA2 method (Callahan *et al.*, 2016a), and chimeras were removed. Taxonomy was assigned with a naive Bayesian classifier using the RDP v14 training set. Amplicon Sequence Variants (ASV) with no taxonomic assignment at phylum level or which were assigned as "Chloroplast" or "Cyanobacteria" were removed from the dataset. The final dataset contained 1,500,800 reads, on average 29,427 reads per sample (min. = 5,804 reads, max. = 78,154 reads).

Results

Communities in the bipartite networks

Applying the leading-eigenvector community search algorithm to the whole network (gut microbiota and bacterioplankton samples together) led to three communities (Fig. 2), with a relatively high and significant modularity score ($Q = 0.303$, $p < 10^{-4}$). Visual inspection evinces that found communities perfectly match the classification of nodes by type of microbial sample (Fig. 2).

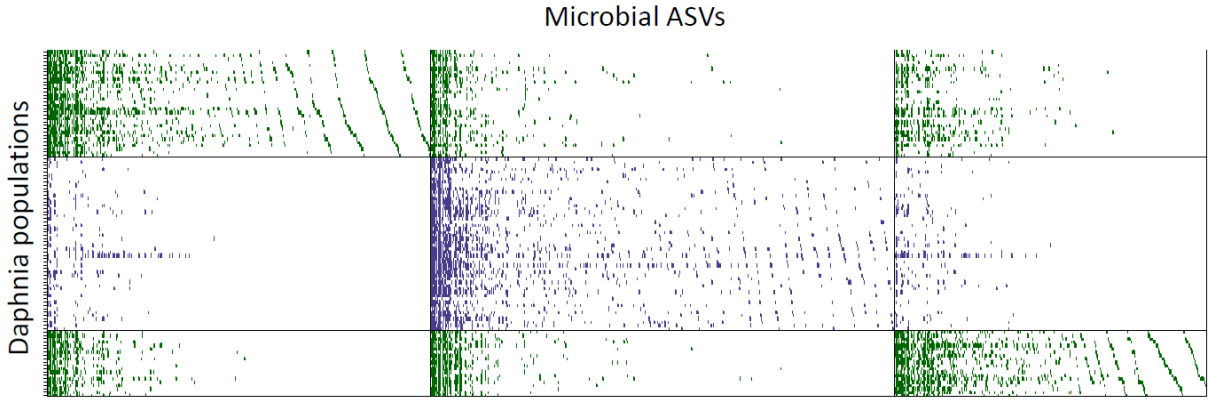


Fig. 2 – Result of the community search within the whole network (all 104 *Daphnia* samples × 1656 microbial ASV). The leading-eigenvector modularity optimization algorithm evinced three communities, here represented by the gray lines dividing columns and rows of the incidence matrix, with dots representing existing interactions (green dots for bacterioplankton, blue dots for gut microbiota). Communities found by the algorithms perfectly correspond to sample types, with two groups within bacterioplankton and one fitting all gut microbiota interactions.

The same community-search algorithm was also applied to the two sub-networks obtained by taking only gut microbiota or bacterioplankton samples (Supp. Figs S1 and S2). In the gut microbiota network, 16 communities were found (Supp. Fig. S1), with a moderate and not significant modularity score ($Q = 0.242$, $p = 0.1739$). In the bacterioplankton network, 6 communities were found (Supp. Fig. S2), with a moderate but significant modularity ($Q = 0.216$, $p = 0.0021$). These 6 communities were poorly related to the two-community division of the bacterioplankton network obtained by running the community-search algorithm on the whole network (results not shown).

Congruence of classifications

The communities found in the whole network were highly congruent with sample type (bacterioplankton vs. gut microbiota; NMI = 0.806, $p < 10^{-4}$; Figs. 1 and 3a), moderately congruent with diet (*Scenedesmus* vs. *Microcystis* diets; NMI = 0.096, p = 0.0163; Fig. 3b) and not congruent with *Daphnia* genotype (NMI = 0.025, p = 0.9377; Fig. 3b).

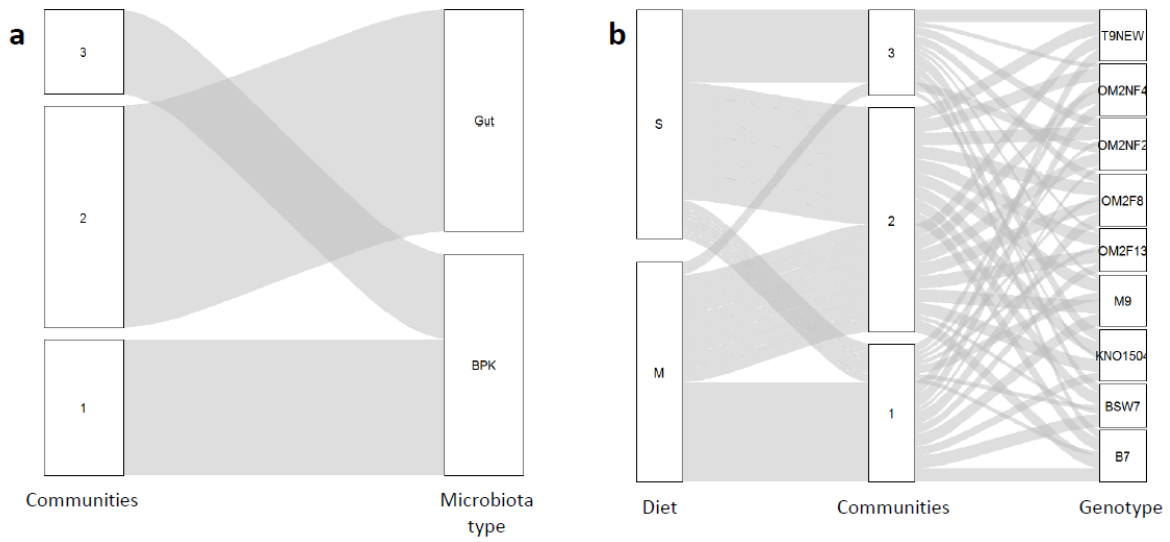


Fig. 3 – Alluvial plots representing the congruence of community-based classification of the whole network (comprising both gut microbiota and bacterioplankton samples) found by the leading-eigenvector algorithm and other classifications based on treatments. Vertically stacked white solid boxes represent groups of nodes following a given classification; grey flows represent the correspondence of nodes between classifications, with larger flows indicating more nodes shared by the two groups linked by the flow. (**a**) Congruence between communities and the type of microbiota sample (gut microbiota, "Gut", and bacterioplankton, "BPK"). (**b**) Congruence between communities and the two treatment factors, diet (on the left-hand side; *Scenedesmus* diet, S, and mixed *Microcystis* and *Scenedesmus* diet, M) and *Daphnia* genotype (on the right-hand side; nine different genotypes indicated by different codes).

Communities found in the two sub-networks based on different types of samples were moderately, but not significantly, congruent with one another (NMI = 0.335, p = 0.0551; Fig. 4a). The communities of both sub-networks were significantly congruent with diet (gut microbiota: NMI = 0.212, p = 0.0009; bacterioplankton: NMI = 0.439, p = 0.0004; Fig. 4b-c), but not with *Daphnia* genotype (gut microbiota: NMI = 0.462, p = 0.0543; bacterioplankton: NMI = 0.288, p = 0.2025; Fig. 4b-c), although visual inspection of Fig. 4c suggests a weak association between bacterioplankton-based

communities and genotype (communities 2 and 6 could be associated with some genotypes). Because we intuitively expected *Daphnia* genotypes to affect associations only after sample type and diet, we checked the congruence between communities found in the four sub-networks corresponding to treatments (diet [*Scenedesmus* vs. *Microcystis*] crossed with sample type [gut microbiota vs. bacterioplankton]). The *Microcystis* × gut microbiota sub-network was the only one to be significantly modular and for which communities were significantly congruent with *Daphnia* genotypes (Supp. Table S1).

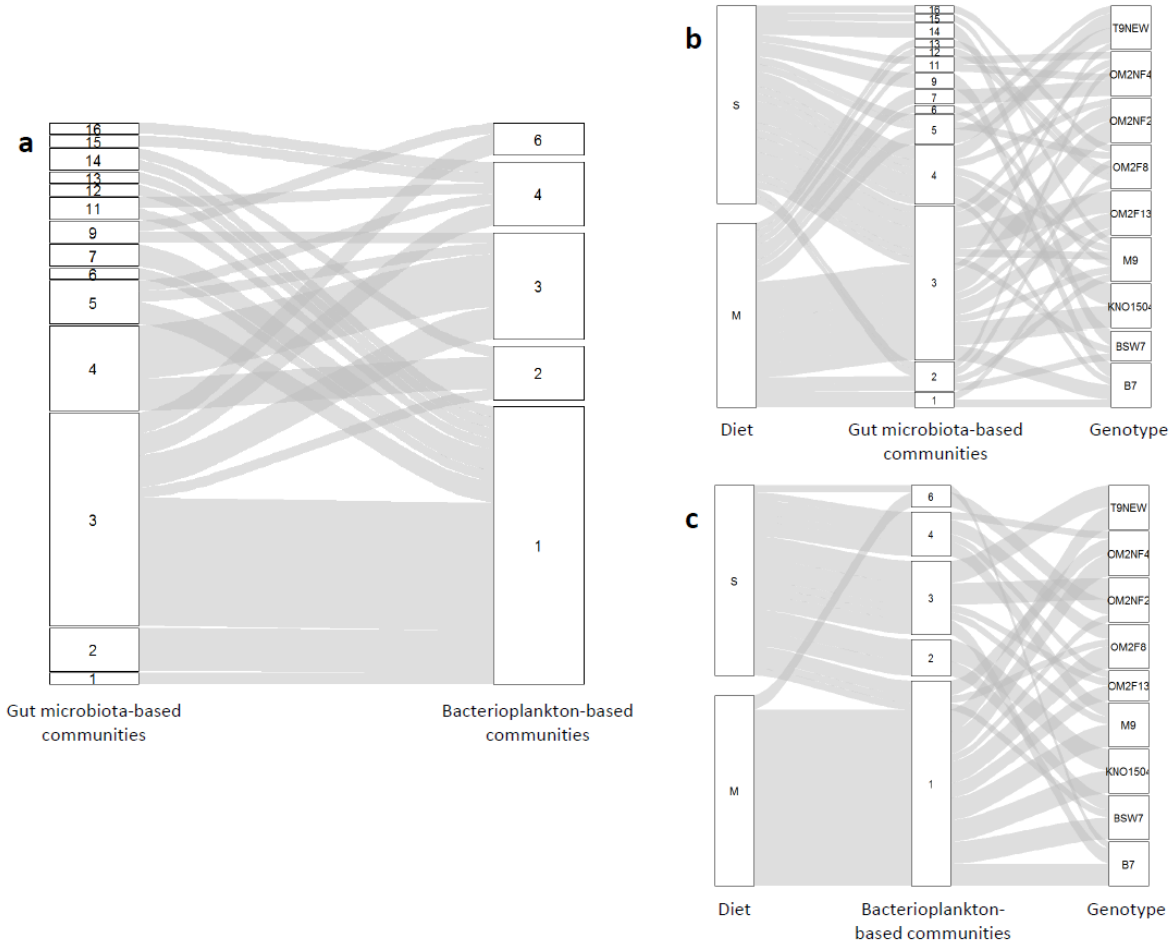


Fig. 4 – Alluvial plots representing the congruence of community-based classifications of the two different microbiota networks, based on gut microbiota or bacterioplankton samples, found by the leading-eigenvector algorithm and other classifications based on treatments. (**a**) Congruence between communities among gut microbiota samples (left-hand side) and those found among bacterioplankton samples (right-hand side). The numbers of communities do not exactly correspond to those found by the algorithm (respectively, 16 and 6) because some communities only comprise microbial ASV nodes, not *Daphnia* population nodes. (**b**) Congruence between gut microbiota-based communities and the two treatment factors, diet (on the left-hand side; *Scenedesmus* diet, S, and mixed *Microcystis* and *Scenedesmus* diet, M) and *Daphnia* genotype (on the right-hand side; nine different genotypes

indicated by different codes).(**c**) Congruence between bacterioplankton-based communities and the two treatment factors, diet and genotype (as in panel **b**).

Canonical correspondence analyses

The results of CCA applied to the communities found in the whole network confirm some of the results found by congruence comparisons (Table 1). Sample type significantly explained communities, irrespectively of whether the effects of diet, genotype, or both, were removed first (Table 1). In all cases, both types of randomization led to significant effects, thus indicating an affinity effect of sample type on network structure – remember that the three communities represented in Fig. 2 perfectly matched sample type, with two communities for bacterioplankton and one for gut microbiota samples. All effects linked to diet, although weaker than sample type effects, were found significant using both types of randomization procedures (Table 1), thus indicating a significant affinity effect of diet on communities, which thus should correspond to the division between the two bacterioplankton communities. By contrast, all assessments of *Daphnia* genotype effects on network structure resulted in low and not significant F-values (Table 1).

Effect	**df**	**χ^2**	**F**	**Row perm. p-value**	**Edge perm. p-value**
type	1	1.000	103.0	0.0001	0.0008
diet	1	0.154	8.482	0.0004	0.0006
genotype	8	0.077	0.513	0.9581	-
type+diet	2	1.153	70.14	0.0001	0.0007
type+genotype	9	1.070	13.16	0.0001	0.0008
diet+genotype	9	0.233	1.494	0.1459	-
type+diet+genotype	10	1.225	16.27	0.0001	0.0007
type | diet+genotype	1	0.992	120.3	0.0001	0.0007
diet | type+genotype	1	0.155	18.81	0.0002	0.0004
genotype | type+diet	8	0.072	1.164	0.3937	-
type | genotype	1	0.992	101.3	0.0001	0.0008
type | diet	1	1.000	120.4	0.0001	0.0007
diet | genotype	1	0.156	8.277	0.0008	0.0007
diet | type	1	0.153	18.46	0.0001	0.0004
genotype | type	8	0.070	0.954	0.5256	-
genotype | diet	8	0.079	0.567	0.9309	-

Table 1 – Results of the CCA applied to the whole network to explain network communities using sample type (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. *Effect*: the explanatory effects and the conditioning ones (figured after the vertical line); *df*: degrees of freedom (= number of categories - 1); χ^2: values of the corresponding chi squared statistic; *F*: values of the F-statistic; *row perm. p-value*: probability that a

randomized version of the explained community table, once removed the effect of conditioning variables, obtains a F-statistic equal or larger to the one obtained with real data; *edge perm. p-value*: probability that a randomized version of the *Daphnia*-microbial ASV network, keeping node degrees constant, obtains a F-statistic equal or larger to the one obtained with real data. This second probability was only computed for effects that were significant for the first test.

Focusing on the gut microbiota network, CCA confirmed that bacterioplankton-based communities did not significantly explain gut microbiota network structure (Table 2). The congruence between gut microbiota communities and diet was partially refuted by the CCA: diet effects were not significant once the effect of bacterioplankton-based communities was accounted for (effects diet | bpk and diet | bpk + genotype), and the potential effect of diet on the two other rows (effects diet and diet | genotype) was only significant for the first test, hence suggesting a weak richness effect, confounded with potential effects of the bacterioplankton.

Effect	**df**	**χ^2**	**F**	**Row perm. p-value**	**Edge perm. p-value**
bpk	4	1.443	0.842	0.2586	-
diet	1	0.537	1.180	0.0052	0.1619
genotype	8	3.262	1.088	0.0033	0.7980
bpk+diet	5	1.829	0.878	0.2083	-
bpk+genotype	12	4.475	1.100	0.0245	0.7888
diet+genotype	9	3.793	1.174	0.0006	0.7699
bpk+diet+genotype	13	4.768	1.110	0.0362	0.8182
bpk | diet+genotype	4	0.975	0.542	0.8924	-
diet | bpk+genotype	1	0.293	0.573	0.5367	-
genotype | bpk+diet	8	2.939	0.949	0.0575	-
bpk | genotype	4	1.213	0.684	0.5519	-
bpk | diet	4	1.292	0.752	0.4722	-
diet | genotype	1	0.531	1.131	0.0074	0.3906
diet | bpk	1	0.386	0.818	0.2954	-
genotype | bpk	8	3.033	0.987	0.0304	0.8112
genotype | diet	8	3.256	1.100	0.0031	0.7983

Table 2 – Results of the CCA applied to the gut microbiota-based network, trying to explain network communities using communities found in the bacterioplankton-based network ("bpk"), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. See Ttable 1 caption for further details.

The absence of congruence between gut microbiota communities and *Daphnia* genotypes, clear from NMI and alluvial plots (Fig. 4b), was somehow moderated by the CCA results: genotype had a significant effect for the first randomization procedure in all cases but the one conditioning by both

diet and bacterioplankton-based communities (effect genotype | bpk+diet in Table 2). However, when the first test was significant, the second never was (Table 2), thus suggesting that some genotypes could be more likely associated with certain communities because of the number of microbial ASVs they are associated with.

In the bacterioplankton network, communities were significantly explained by diet, thus confirming earlier insights from NMI and Fig. 4c (Table 3). In all cases, the effect was significant for both randomization procedures, thus suggesting an affinity effect. Although weaker (lower F statistics), the same conclusion could be reached for the effect of *Daphnia* genotype (Table 3), thus contradicting NMI comparisons and corroborating the hint given by Fig. 4c. In this case, an affinity effect (both tests significant) was also reported. The absence of correspondence between gut microbiota- and bacterioplankton-based communities (Fig. 4a) was confirmed by non-significant effects in CCA (effects gut, gut | diet, gut | genotype and gut | diet + genotype in Table 3).

Effect	**df**	**χ^2**	**F**	**Row perm. p-value**	**Edge perm. p-value**
gut	13	1.601	1.232	0.4527	-
diet	1	0.573	4.643	0.0001	0.0034
genotype	8	1.565	1.941	0.0023	0.0257
gut+diet	14	1.993	1.563	0.1476	-
gut+genotype	21	3.102	2.243	0.0324	0.0797
diet+genotype	9	2.122	2.676	0.0001	0.0042
gut+diet+genotype	22	3.485	2.772	0.0025	0.0372
gut | diet+genotype	13	1.363	1.459	0.2845	-
diet | gut+genotype	1	0.383	3.500	0.0006	0.0071
genotype | gut+diet	8	1.493	2.226	0.0038	0.0152
gut | genotype	13	1.538	1.470	0.2532	-
gut | diet	13	1.420	1.172	0.5490	-
diet | genotype	1	0.557	5.174	0.0001	0.0023
diet | gut	1	0.391	3.028	0.0011	0.0083
genotype | gut	8	1.501	2.008	0.0067	0.0253
genotype | diet	8	1.550	2.148	0.0006	0.0150

Table 3 – Results of the CCA applied to the bacterioplankton-based network, trying to explain network communities using communities found in the gut microbiota-based network ("gut"), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. See table 1 caption for further details.

Singular value decomposition and redundancy analyses

When modelling the whole network, visual inspection of how congruent the modules obtained from the approximated networks were with those of the observed network indicated that congruence had several local maxima (Supp. Fig. S3a), while mean absolute error (MAE) declined steadily with number of vectors (Supp. Fig. S3b). The adjusted R^2 of individual fractions attributable to sample type or diet also declined with the number of vectors, while the fraction attributable to genotype had a maximum at 13 vectors (Supp. Fig. S4). Because the first maximum of NMI was found with three vectors, we present all RDA results on the whole network using three vectors only. However, given the patterns reported in Supp. Fig. S4, we also checked robustness with different number of vectors (5, 13, and 100; Supp. Tables S2-S4).

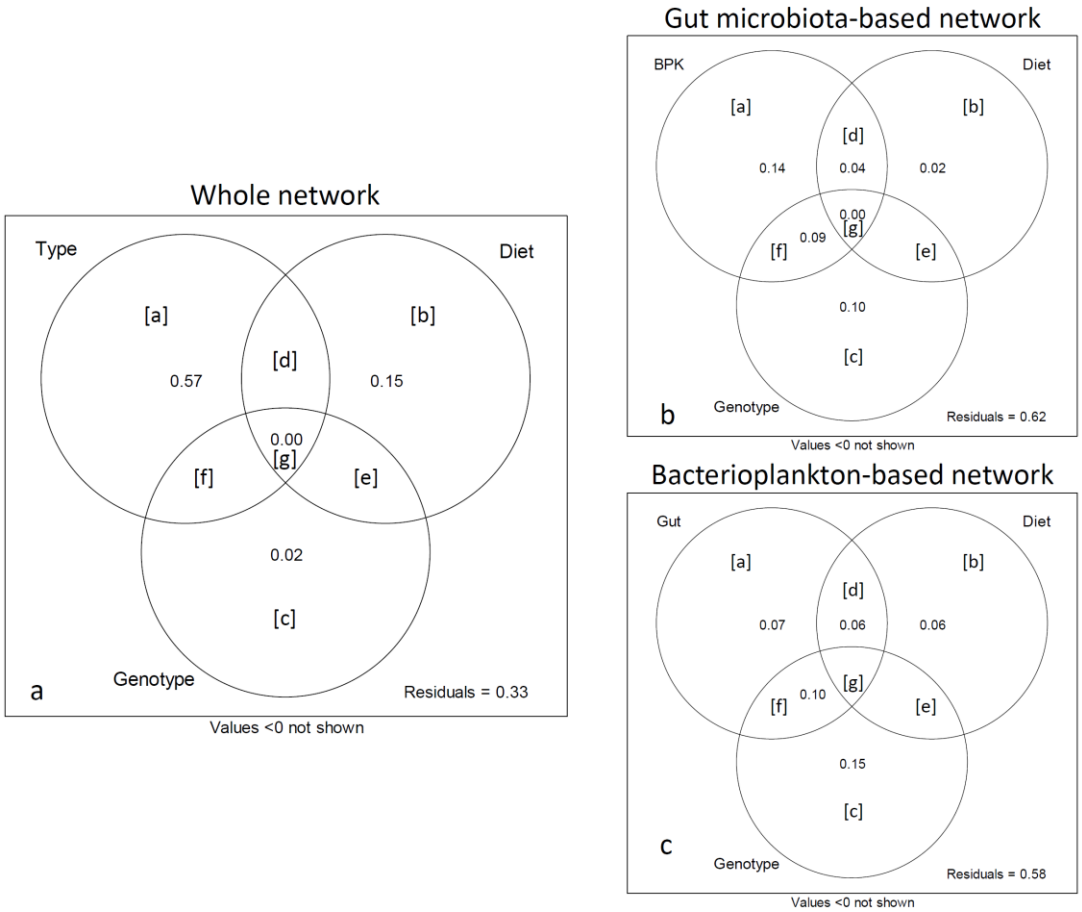


Fig. 5 – Venn diagrams representing the partition of variation (redundancy analysis, RDA) within the reduced matrices obtained by singular value decomposition (SVD) of incidence matrices (SVD-RDA method in Fig. 1). Fractions [a], [b], … [g] indicate the same fractions as those found in Tables 4-6. Given values are non-negative adjusted R^2 (negative adjusted R^2 are omitted for clarity) of individual

fractions (i.e. [a] for Type, not [adfg]), all adjusted R^2 are given in Tables 4-6. Tested factors comprise: "type", the type of microbiota sample (bacterioplankton vs. gut microbiota); "diet", the diet treatment (*Scenedesmus* vs. *Microcystis*); "genotype", the genotype of *Daphnia* populations; "BPK", reduced matrix obtained from applying the SVD on the bacterioplankton-based incidence matrix; "Gut", reduced matrix obtained from applying the SVD on the gut microbiota-based incidence matrix. SVD-reduced matrices comprise different numbers of vectors (see main text for details). (**a**) Results of the RDA applied to the whole network comprising both gut microbiota and bacterioplankton samples. (**b**) Results of the RDA applied to the gut microbiota-based network. (**c**) Results of the RDA applied to the bacterioplankton-based network.

Fraction	**Effect**	**df**	**adjusted R^2**	**Row perm. p-value**	**Edge perm. p-value**
[adfg]	type	1	0.52	0.0001	0.0002
[bdeg]	diet	1	0.13	0.0001	0.0001
[cefg]	genotype	8	-0.03	0.8997	-
[abdefg]	type+diet	2	0.66	0.0001	0.0001
[acdefg]	type+genotype	9	0.53	0.0001	0.0002
[bcdefg]	diet+genotype	9	0.11	0.0011	0.0002
[abcdefg]	type+diet+genotype	10	0.67	0.0001	0.0001
[a]	type | diet+genotype	1	0.57	0.0001	0.0002
[b]	diet | type+genotype	1	0.15	0.0001	0.0001
[c]	genotype | type+diet	8	0.02	0.0822	-
[d]	-	-	-0.01	-	0.9997
[e]	-	-	-0.01	-	1.0000
[f]	-	-	-0.04	-	0.9960
[g]	-	-	0.00	-	0.0060
[h]	residuals	-	0.33	-	-
[ad]	type | genotype	1	0.56	0.0001	0.0002
[af]	type | diet	1	0.53	0.0001	0.0002
[bd]	diet | genotype	1	0.14	0.0001	0.0001
[be]	diet | type	1	0.14	0.0001	0.0001
[ce]	genotype | type	8	0.01	0.3185	-
[cf]	genotype | diet	8	-0.02	0.8218	-

Table 4 – Results of the RDA-SVD applied to the whole network with the first three vectors retained, to explain network communities using sample type (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. *Fraction*: symbolic representation of the components of variations explained by the different factors as represented in Fig. 5; *Effect*: the explanatory effects and the conditioning ones (figured after the vertical line); *df*: degrees of freedom of the explanatory variables; *R^2*: values of the corresponding coefficient of determination (expressed in percentage); *adjusted R^2*: R^2 values corrected for the number of degrees of freedom; *row perm. p-value*: probability that a randomized version of the explained community table, once removed the effect of conditioning variables, obtains an adjusted R^2 equal or larger to the one obtained with real data; *edge perm. p-value*: probability that a randomized version of the *Daphnia*-microbial ASV network, keeping node degrees constant, obtains an adjusted R^2 equal or larger than the one obtained with real data. Dashes indicate values that cannot be computed and/or that cannot be tested.

Sample type (gut microbiota vs. bacterioplankton) explained 57% of the sum of squares on its own (Fig. 5, Table 4). Diet explained 15% and genotype, 2% (Fig. 5). Type and diet had doubly significant effects, whereas genotype effect was not significant at all (Table 4). However, testing with different numbers of vectors yielded contrasted results for the effect of genotype, since all factors (diet, type and genotype) had doubly significant effects when considering approximations of 5, 13 or 100 vectors (Supp. Tables S2-S4). This suggests that all tested factors play a role in shaping the network, with different affinities between *Daphnia* and their microbes depending on their diet, their genotype and where the microbes are actually sampled, but the effect of genotype could only be detected by using sufficiently detailed approximations. In other words, *Daphnia* genotype determines whether a given *Daphnia* population is likely to associate with certain microbial species rather than others within the same community (Supp. Tables S2-S4), but not whether the same *Daphnia* population is more likely to associate with microbes from a given community rather than from another one (Table 1).

In the gut microbiota network, the MAE of the approximated network steadily decreased with the number of retained vectors (Supp. Fig. S5b) while congruence between its communities and those of the original network displayed several local maxima between *ca*. 5 and 30 vectors (Supp. Fig. S5a). As the first important local NMI maximum was found for 10 retained vectors, we present all RDA results on gut microbiota using the first 10 columns of approximation (3). However, robustness checks were also performed using 30 vectors (Supp. Table S5).

Fraction	**Effect**	**df**	**adjusted R^2**	**Row perm. p-value**	**Edge perm. p-value**
[adfg]	bpk	9	0.27	0.0001	0.0001
[bdeg]	diet	1	0.05	0.0001	0.0004
[cefg]	genotype	8	0.18	0.0001	0.0001
[abdefg]	bpk+diet	10	0.27	0.0001	0.0001
[acdefg]	bpk+genotype	17	0.35	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.24	0.0001	0.0001
[abcdefg]	bpk+diet+genotype	18	0.38	0.0001	0.0001
[a]	bpk | diet+genotype	9	0.14	0.0009	0.0188
[b]	diet | bpk+genotype	1	0.02	0.0299	0.1160
[c]	genotype | bpk+diet	8	0.10	0.0002	0.0084
[d]	-	-	0.04	-	0.0006
[e]	-	-	-0.01	-	0.7075
[f]	-	-	0.09	-	0.0007
[g]	-	-	0.00	-	0.7983
[h]	residuals	-	0.62	-	-
[ad]	bpk | genotype	9	0.18	0.0001	0.0002

[af]	bpk | diet	9	0.22	0.0001	0.0001
[bd]	diet | genotype	1	0.06	0.0001	0.0002
[be]	diet | bpk	1	0.01	0.1569	-
[ce]	genotype | bpk	8	0.09	0.0015	0.0106
[cf]	genotype | diet	8	0.19	0.0001	0.0000

Table 5 – Results of the redundancy analysis / singular value decomposition (RDA-SVD) analysis applied to the gut microbiota-based network with the first 10 vectors retained, trying to explain network communities using the SVD-reduced matrix for the bacterioplankton network ("bpk", with 9 vectors chosen using the forward selection procedure of Blanchet et al. 2008), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. See table 4 caption for further details.

Both bacterioplankton network structure (here summarized using 9 vectors among the first 30, following the procedure of Blanchet et al. 2008) and *Daphnia* genotype had doubly significant effects in all tested models (Table 5), indicating affinity effects of bacterioplankton and genotype. The effect of diet was weaker (adj. R^2 = 2%), significant against the other two effects together (effect: diet | bpk+genotype, Table 5), and yet failed to reach significance when conditioning for bacterioplankton only (effect: diet | bpk, Table 5). Because the diet | bpk+genotype effect was not significant for the second tests suggests that the weak effect of diet on gut microbiota network structure is only a richness effect. Using the first 30 vectors (instead of 10) of the approximation of the gut microbiota incidence matrix yielded relatively similar results, with the exception of a consistently significant richness effect of diet (Supp. Table S5).

In the bacterioplankton network, the approximation of the network by SVD displayed steadily decreasing MAE with the number of retained vectors (Supp. Fig. S6b) and multiple local maxima for the NMI between communities of the original and approximated networks were found (Supp. Fig. 6a). A first local maximum NMI was obtained for 11 retained vectors, which we used in the ensuing RDA. We also checked the robustness of RDA results using the first 30 vectors (Supp. Table S6).

All effects could be interpreted as affinity effects since both tests proved significant (Table 6). Thus, communities of the gut microbiota network, diet and genotype had an affinity effect on the structure of bacterioplankton communities. When taking into accounts 30 vectors of the approximation of the bacterioplankton network, the effect of the gut microbiota network was only significant when tested

against one other factor, and not when conditioned by both diet and genotype (Supp. Table S6), but this was obtained with a forward selection procedure opting for retaining only one vector from the gut microbiota matrix. This suggests that the correspondence between both network structures might be restricted to relatively fine scales.

Fraction	**Effect**	**df**	**adjusted R^2**	**Row perm. p-value**	**Edge perm. p-value**
[adfg]	gut	7	0.22	0.0001	0.0001
[bdeg]	diet	1	0.10	0.0001	0.0001
[cefg]	genotype	8	0.23	0.0001	0.0001
[abdefg]	gut+diet	8	0.27	0.0001	0.0001
[acdefg]	gut+genotype	15	0.36	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.36	0.0001	0.0001
[abcdefg]	gut+diet+genotype	16	0.42	0.0001	0.0001
[a]	gut | diet+genotype	7	0.07	0.0084	0.0025
[b]	diet | gut+genotype	1	0.06	0.0001	0.0001
[c]	genotype | gut+diet	8	0.15	0.0001	0.0001
[d]	-	-	0.06	-	0.0006
[e]	-	-	-0.01	-	0.9491
[f]	-	-	0.10	-	0.0001
[g]	-	-	-0.01	-	0.9795
[h]	residuals	-	0.58	-	-
[ad]	gut | genotype	7	0.13	0.0001	0.0001
[af]	gut | diet	7	0.17	0.0001	0.0001
[bd]	diet | genotype	1	0.12	0.0001	0.0001
[be]	diet | gut	1	0.05	0.0002	0.0001
[ce]	genotype | gut	8	0.14	0.0001	0.0001
[cf]	genotype | diet	8	0.26	0.0001	0.0001

Table 6 – Results of the redundancy analysis / singular value decomposition (RDA-SVD) analysis applied to the bacterioplankton-based network with the first 11 vectors retained, trying to explain network communities using the SVD-reduced matrix for the gut microbiota network (“gut” ”, with 7 vectors chosen using the forward selection procedure of Blanchet et al. 2008), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. See table 4 caption for further details.

Discussion

The methodological framework we propose was successfully applied to results of an experiment aimed at uncovering the potential reciprocal effects of *Daphnia* gut microbiota and bacterioplankton in the face of diets of heterogeneous edibility. In an earlier study (Macke *et al.*, 2020), we suggested that the gut microbiome was different from the surrounding bacterioplankton and that both microbial pools showed a dependency on *Daphnia* genotype and diet, on the basis of statistical analyses (GLM, PERMANOVA, Mantel tests) performed directly on the taxonomic composition and relative abundances of microbial samples. The present study confirms the difference between the two sample types (Figs. 2-3, Tables 1 and 4), but clarifies their dependency on diet and genotype.

The gut microbiota evinced a very low non-significant modularity, with communities only mildly matching diet and genotype (Fig. 4, Table 2). However, delving more into the details of the gut microbiota structure using the SVD approximation highlighted affinity effects of all tested components (bacterioplankton network structure, *Daphnia* genotype and diet; Table 5), thus suggesting that the dependencies found in our earlier study depend on relatively fine-grained structures (non-systematic associations of very few bacterial ASVs with certain *Daphnia* populations), and not on mesoscale structures such as network communities. This was confirmed by analyzing communities within the sub-networks corresponding to diet and sample type, with the *Microcystis* × gut microbiota sub-network being significantly modular and congruent with the classification by *Daphnia* genotypes (Supp. Table S1). The bacterioplankton communities matched diet well, with an affinity of certain bacterioplankton species for one diet over another (Fig. 4). The CCA also suggested an affinity effect due to *Daphnia* genotype (Table 3), and both effects were further confirmed by the RDA-SVD method (Table 6). The absence of correspondence between gut microbiota and bacterioplankton communities (Fig. 4a) was confirmed by CCAs (Tables 2 and 3), but a reciprocal affinity effect was evinced at finer scales through RDA-SVD (Tables 5 and 6), with a stronger effect of bacterioplankton on gut microbiota than the reverse, thus suggesting that the exact association of *Daphnia* populations with their gut microbes within certain communities (selected by diet and genotype) might be partially

determined by the surrounding bacterioplankton, itself partially shaped by *Daphnia* genotype and diet. Overall, present results are coherent with those of our earlier studies, but could pinpoint differences in effect size due to differences in considered grain, especially when varying the number of vectors retained for the RDA-SVD approach (Table 4 vs. Supp. Tables S2-S4, Table 5 vs. Supp. Table S5, Table 6 vs. Supp. Table S6).

In methodological terms, our approach has several advantages over existing methods. First, working on approximations of the network rather than on the network directly circumvented the problem of node degree dependency on one another. This is a serious issue with methods based e.g. on generalized linear model directly explaining interaction or interaction strength based on node properties (e.g. Gravel *et al.*, 2019). Such limitations can also be partially removed by using node-wise random effects to account for heterogeneity in node degrees (e.g. de Manincor *et al.*, in press). Second, the approximation-based nature of the methods we propose allows an assessment of effects acting at different network scales, from community scale down to finer ones. Although communities do represent informative structures to understand networks, there are indeed limits to what they can capture, in particular due to theoretical resolution limits (Fortunato & Barthélemy, 2007) and to nodes belonging to multiple communities (Palla *et al.*, 2005). Current block model approaches also allow the incorporation of external variables (Leger, 2016), and could thus theoretically be used to decipher effects acting within and among communities. However, these approaches only consider external variables assigned to dyads (i.e. pairs of nodes), thus preventing the assessment of effects linked to e.g. *Daphnia* genotype in the present study. Finally, one major advantage of our method is that it is not computationally as extensive as the other approaches able to both measure the effect of external variables on networks and account for intrinsic dependencies within networks (e.g. exponential random graph models, latent block models with covariates, or Bayesian structural equation models).

The approaches we advocated in the present study can possibly be extended in various ways. One major extension is to allow the use of weighted incidence matrices. To do so, at least two issues need

to be dealt with. First, the congruence and CCA approach is based on network communities. Following Leger et al. (2015), communities should be discovered using latent block modelling (LBM), which takes us back to the computation time problem – LBMs are known to be notoriously long to obtain (see computation times given by Leger et al. 2015), and thus might be temporarily unsuitable for large datasets such as host-microbial ASVs association networks. Second, the 'curveball' algorithm used for the second test does not have an equivalent for weighted networks. There, the challenge lies in finding a null model randomizing edge among nodes while both keeping total weights and number of non-zero weights per node constant. R method 'quasiswap_count' in the 'vegan' function 'commsim' and the 'swap.web' function in package 'bipartite' propose potential algorithms for this, but do not guarantee uniformity of the space of sampled matrices.

The present methodological framework can allow other types of network approximations. For instance, using normalized role vectors obtained from the decomposition of node positions within motifs, as advocated by Simmons et al. (2019), might provide another entry point into the structure of networks. However, contrary to community memberships or SVD vectors, role vectors are not orthogonal (i.e. there are correlations between positions obtained from motifs of different sizes), which has to be accounted for in order to develop a useful statistical approach based on these vectors, probably by filtering role vector correlations through principal component analysis or similar approaches. Another possibility is to modify the SVD approximation by working on the Laplacian of the adjacency matrix (Griffith & Li, 2017), or a simple transformation of the Laplacian such as the one used for Moran Eigenvector Maps [MEM] (Dray, Legendre & Peres-Neto, 2006). Eigenvectors obtained by such methods have more direct interpretations than those obtained by SVD of the incidence matrix – the value of MEM eigenvectors, for instance, change more or less rapidly from one node to the next depending on their associated eigenvalues (Thioulouse, Chessel & Champely, 1995; but this is not true of the eigenvectors of Laplacian matrices, see Griffith & Li, 2017). However, Laplacian matrices assume that each node's degree is known *a priori* and MEM make heavy use of weights associated to edges in the network – two assumptions that the SVD does not make.

The “analysis” step of Fig. 1 can also be modified. We chose to use classic multivariate analyses (RDA and CCA); RDA and CCA, however, are implicitly based on Euclidean and Chi-square distances between data points, respectively. Other distances can be used following the distance-based RDA (db-RDA) framework established by Legendre & Gallagher (2001) and extended by Blanchet et al. (2014). Because the SVD approach provides an approximation of the network as a matrix of real values (the **L** matrix), it does not suffer from the ‘double zero’ issue which classically affects distances computed on species presence/absence or abundance tables (Legendre & Legendre, 2012), and thus can safely be analyzed using most distance functions. By contrast, the community-detection approach provides a binary membership matrix suffering from the ‘double zero’ issue, which calls for a careful choice of distance to perform db-RDA. Ideally, a consensus approach based on a variety of db-RDA analyses (as advocated by Blanchet *et al.*, 2014) could lead to a more complete picture for both the community-detection and SVD approximations, provided one focuses more on checking the agreement of fraction tests than on finding which db-RDA provides the highest values for explained sum-of-squared distances. We compared the CCA approach performed on the whole network with a Jaccard distance-based-RDA. This analysis recovered results qualitatively similar to those obtained in Table 1 (results not shown). However, the current implementation of db-RDA in the R package ‘vegan’ is quite slower than that of CCA, which prevents the use of the consensus approach when combined with the double randomization tests. An efficient possibility for using such an alternative approach is given by performing RDA on pre-transformed data, using function ‘decostand’ in the R package ‘vegan’. This does not allow all the variety offered by db-RDA but nonetheless offers a few alternatives to the two paths presented here.

Finally, both the RDA and CCA approaches can be improved in the context of observational studies (i.e. when external variables are not controlled) by embedding these analyses into a Structural Equation Modelling (SEM) framework. Such an extension is way beyond the scope of this study, as no implementation of such a hybrid model exists yet, but this could arguably help decipher the complex

interactions between considered variables. For instance, when some variables have been measured together with assessments of one or more networks, embedding an RDA into a SEM might allow assessments of common causal pathways (X affecting network A and network B), indirect causal relationships (X affecting network A, in turn affecting network B) and other complex causal pathways within a large subset of potential causal models. Combined with the variable-scale property of the SVD-RDA approach proposed here, this embedding of the RDA into an SEM might ultimately lead to a finer assessment of possible causal relationships among networks and external variables at different network scales.

Acknowledgements

We wish to thank L. De Meester for his constructive comments and C. Sueur and S. Sosa for the invitation to submit to this special feature. Development of the methods benefited from discussions with N. de Manincor, A. Fisogni, A. Berquer, A. Tasiemski, B. Schatz, M. Grenié, N. Joffard, S. Robin, S. Donnet, S. Ouadah and M. Thomas. We are grateful to S. Donnet and S. Ouadah for providing us with the R script for drawing alluvial diagrams. We also wish to thank I. Kaygorodova, Y. Sapozhnikova, P. Brodin, and C. Daniel for the opportunity to present and discuss these methods at various venues. We thank the editor, associate editor and reviewers, whose comments greatly improved the clarity of the paper. Funding for the experiment was provided by the KU Leuven C16/2017/02 project, and an FWO postdoctoral fellowship (N°12R4917N, to EM) and FWO G092619N to EDC. FM was funded by the CNRS and ANR projects ARSENIC (grant no. 14-CE02-0012), NGB (grant no. 17-CE32-0011), and ECONET (grant no. 18-CE02-0010). This research project was performed within the framework of the FWO EVENET network.

Author contributions

EM and ED conceived the ideas, designed the approach to the experiments, performed the experiments and collected the data. FM developed the methods, analyzed the data and led the writing of the first

version of the manuscript. All authors contributed to revised versions and approved the final version of the manuscript.

Data accessibility

Data and R scripts are available on Zenodo at https://doi.org/10.5281/zenodo.3904387.

References

Amend, A.S., Seifert, K.A. & Bruns, T.D. (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology,* **19,** 5555-5565.

Astegiano, J., Altermatt, F. & Massol, F. (2017) Disentangling the co-structure of multilayer interaction networks: degree distribution and module composition in two-layer bipartite networks. *Scientific Reports,* **7,** 15465.

Bartomeus, I. (2013) Understanding linkage rules in plant-pollinator networks by using hierarchical models that incorporate pollinator detectability and plant traits. *PLOS ONE,* **8,** e69200.

Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016) A common framework for identifying linkage rules across different types of interactions. *Functional Ecology,* **30,** 1894-1903.

Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003) The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences of the U.S.A.,* **100,** 9383-9387.

Bascompte, J. & Stouffer, D.B. (2009) The assembly and disassembly of ecological networks. *Philosophical Transactions of the Royal Society B-Biological Sciences,* **364,** 1781-1787.

Bauman, D., Drouet, T., Dray, S. & Vleminckx, J. (2018) Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. *Ecography,* **41,** 1638-1649.

Blanchet, F.G., Legendre, P., Bergeron, J.A.C. & He, F. (2014) Consensus RDA across dissimilarity coefficients for canonical ordination of community composition data. *Ecological Monographs,* **84,** 491-511.

Blanchet, F.G., Legendre, P. & Borcard, D. (2008) Forward selection of explanatory variables. *Ecology,* **89,** 2623-2632.

Blüthgen, N., Menzel, F. & Blüthgen, N. (2006) Measuring specialization in species interaction networks. *BMC Ecology,* **6,** 9.

Bohan, D.A., Landuyt, D., Ma, A., Macfadyen, S., Martinet, V., Massol, F., …Woodward, G. (2016) Networking our way to better ecosystem service provision. *Trends in Ecology & Evolution,* **31,** 105-115.

Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G. (2017) Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution,* **32,** 477-487.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology,* **73,** 1045-1055.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. (2016a) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods,* **13,** 581.

Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J. & Holmes, S.P. (2016b) Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research,* **5,** 1492-1492.

Callens, M., Macke, E., Muylaert, K., Bossier, P., Lievens, B., Waud, M. & Decaestecker, E. (2016) Food availability affects the strength of mutualistic host-microbiota interactions in *Daphnia magna*. *The ISME Journal,* **10,** 911-920.

CaraDonna, P.J., Petry, W.K., Brennan, R.M., Cunningham, J.L., Bronstein, J.L., Waser, N.M. & Sanders, N.J. (2017) Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology Letters,* **20,** 385-394.

Cohen, J.E. & Briand, F. (1984) Trophic links of community food webs. *Proceedings of the National Academy of Sciences of the United States of America,* **81,** 4105-4109.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems,* **1695,** 1-9.

Dalla Riva, G.V. & Stouffer, D.B. (2016) Exploring the evolutionary signature of food webs' backbones using functional traits. *Oikos,* **125,** 446-456.

Danon, L., Díaz-Guilera, A., Duch, J. & Arenas, A. (2005) Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, **2005,** P09008.

de Manincor, N., Hautekeete, N., Piquot, Y., Schatz, B., Vanappelghem, C. & Massol, F. (in press) Does phenology explain plant-pollinator interactions at different latitudes? An assessment of its explanatory power in plant-hoverfly networks in French calcareous grasslands. *Oikos*.

Derocles, S.A.P., Bohan, D.A., Dumbrell, A.J., Kitson, J.J.N., Massol, F., Pauvert, C., …Evans, D.M. (2018) Biomonitoring for the 21st Century: Integrating Next-Generation Sequencing Into Ecological Network Analysis. *Advances in Ecological Research,* pp. 1-62. Academic Press.

Dray, S., Bauman, D., Blanchet, F.G., Borcard, D., Clappe, S., Guenard, G., …Wagner, H.H. (2019) adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-4. https://CRAN.R-project.org/package=adespatial

Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling,* **196,** 483-493.

Dunne, J.A., Williams, R.J. & Martinez, N.D. (2002) Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences,* **99,** 12917-12922.

Encinas-Viso, F., Alonso, D., Klironomos, J.N., Etienne, R.S. & Chang, E.R. (2016) Plant–mycorrhizal fungus co-occurrence network lacks substantial structure. *Oikos,* **125,** 457-467.

Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., …Bascompte, J. (2010) Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology,* **79,** 811-817.

Fortunato, S. (2010) Community detection in graphs. *Physics Reports,* **486,** 75-174.

Fortunato, S. & Barthélemy, M. (2007) Resolution limit in community detection. *Proceedings of the National Academy of Sciences,* **104,** 36-41.

García-Callejas, D., Molowny-Horas, R. & Araújo, M.B. (2018) Multiple interactions networks: towards more realistic descriptions of the web of life. *Oikos,* **127,** 5-22.

Govaert, G. & Nadif, M. (2008) Block clustering with Bernoulli mixture models: Comparison of different approaches. *Computational Statistics & Data Analysis,* **52,** 3233-3245.

Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., …Roslin, T. (2019) Bringing Elton and Grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. *Ecography,* **42,** 401-415.

Gravel, D., Poisot, T., Albouy, C., Velez, L. & Mouillot, D. (2013) Inferring food web structure from predator–prey body size relationships. *Methods in Ecology and Evolution,* **4,** 1083-1090.

Griffith, D.A. & Li, B. (2017) A geocomputation and geovisualization comparison of Moran and Geary eigenvector spatial filtering. *2017 25th International Conference on Geoinformatics,* pp. 1-4.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., …Gorsky, G. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature,* **532,** 465-470.

Joffard, N., Massol, F., Grenié, M., Montgelard, C. & Schatz, B. (2019) Effect of pollination strategy, phylogeny and distribution on pollination niches of Euro-Mediterranean orchids. *Journal of Ecology,* **107,** 478-490.

Jordano, P. (1987) Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. *American Naturalist,* **129,** 657-677.

Kamenova, S., Bartley, T., Bohan, D., Boutain, J.R., Colautti, R.I., Domaizon, I., …Massol, F. (2017) Invasions toolkit: current methods for tracking the spread and impact of invasive species. *Advances in Ecological Research,* **56,** 85-182.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., …O'Hara, R.B. (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography,* **39,** 2163-2178.

Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia,* **129,** 271-280.

Legendre, P. & Legendre, L.F. (2012) *Numerical ecology*. Elsevier.

Leger, J.-B. (2016) Blockmodels: A R-package for estimating in Latent Block Model and Stochastic Block Model, with various probability functions, with or without covariates. *arXiv preprint arXiv:1602.07587*.

Leger, J.-B., Daudin, J.-J. & Vacher, C. (2015) Clustering methods differ in their ability to detect patterns in ecological networks. *Methods in Ecology and Evolution,* **6,** 474-481.

Lewinsohn, T.M., Prado, P.I., Jordano, P., Bascompte, J. & Olesen, J.M. (2006) Structure in plant-animal interaction assemblages. *Oikos,* **113,** 174-184.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., …Gordon, J.I. (2008) Evolution of mammals and their gut microbes. *Science,* **320,** 1647-1651.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., …Raes, J. (2015) Determinants of community structure in the global plankton interactome. *Science,* **348,** 1262073.

Macke, E., Callens, M., Massol, F., Vanoverberghe, I., De Meester, L. & Decaestecker, E. (2020) Diet and Genotype of an Aquatic Invertebrate Affect the Composition of Free-Living Microbial Communities. *Frontiers in Microbiology,* **11**.

McMurdie, P.J. & Holmes, S. (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology,* **10,** e1003531.

Memmott, J. (1999) The structure of a plant-pollinator food web. *Ecology Letters,* **2,** 276-280.

Newman, M.E.J. (2006a) Finding community structure in networks using the eigenvectors of matrices. *Physical Review E,* **74,** 036104.

Newman, M.E.J. (2006b) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences,* **103,** 8577-8582.

Nogales, M., Heleno, R., Rumeu, B., González-Castro, A., Traveset, A., Vargas, P. & Olesen, J.M. (2016) Seed-dispersal networks on the Canaries and the Galápagos archipelagos: interaction modules as biogeographical entities. *Global Ecology and Biogeography,* **25,** 912-922.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., …Wagner, H. (2018) vegan: Community Ecology Package. R package version 2.4-6. https://CRAN.R-project.org/package=vegan

Olesen, J.M., Bascompte, J., Dupont, Y.L. & Jordano, P. (2007) The modularity of pollination networks. *Proceedings of the National Academy of Sciences,* **104,** 19891-19896.

Orsini, C., Dankulov, M.M., Colomer-de-Simon, P., Jamakovic, A., Mahadevan, P., Vahdat, A., …Krioukov, D. (2015) Quantifying randomness in real networks. *Nature Communications,* **6,** 8627.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016) Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution,* **7,** 549-555.

Palla, G., Derenyi, I., Farkas, I. & Vicsek, T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature,* **435,** 814-818.

Peres-Neto, P.R., Legendre, P., Dray, S. & Borcard, D. (2006) Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology,* **87,** 2614-2625.

Poisot, T., Baiser, B., Dunne, J.A., Kéfi, S., Massol, F., Mouquet, N., …Gravel, D. (2016) mangal – making ecological network analysis simple. *Ecography,* **39,** 384-390.

Rohr, R.P., Naisbit, R.E., Mazza, C. & Bersier, L.-F. (2016) Matching–centrality decomposition and the forecasting of new links in networks. *Proceedings of the Royal Society of London B: Biological Sciences,* **283**.

Rohr, R.P., Scherer, H., Kehrli, P., Mazza, C. & Bersier, L.F. (2010) Modeling food webs: exploring unexplained structure using latent traits. *American Naturalist*, **176,** 170-177.

Rohrlack, T., Dittmann, E., Börner, T. & Christoffersen, K. (2001) Effects of Cell-Bound Microcystins on Survival and Feeding of *Daphnia* spp. *Applied and Environmental Microbiology*, **67,** 3523-3529.

Sabatier, R., Lebreton, J.-D. & Chessel, D. (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. *Multiway data analysis* (eds R. Coppi & S. Bolasco), pp. 341-352. Elsevier Science Publishers, B.-V.

Simmons, B.I., Cirtwill, A.R., Baker, N.J., Wauchope, H.S., Dicks, L.V., Stouffer, D.B. & Sutherland, W.J. (2019) Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, **128,** 154-170.

Stoks, R., Govaert, L., Pauwels, K., Jansen, B. & De Meester, L. (2016) Resurrecting complexity: the interplay of plasticity and rapid evolution in the multiple trait response to strong changes in predation pressure in the water flea Daphnia magna. *Ecology Letters*, **19,** 180-190.

Stouffer, D.B., Camacho, J., Guimera, R., Ng, C.A. & Amaral, L.A.N. (2005) Quantitative patterns in the structure of model and empirical food webs. *Ecology*, **86,** 1301-1311.

Stouffer, D.B., Camacho, J., Jiang, W. & Amaral, L.A.N. (2007) Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B-Biological Sciences*, **274,** 1931-1940.

Stouffer, D.B., Sales-Pardo, M., Sirer, M.I. & Bascompte, J. (2012) Evolutionary conservation of species' roles in food webs. *Science*, **335,** 1489-1492.

Strona, G., Nappo, D., Boccacci, F., Fattorini, S. & San-Miguel-Ayanz, J. (2014) A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature Communications*, **5**.

ter Braak, C.J.F. (1986) Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology*, **67,** 1167-1179.

Thébault, E. & Fontaine, C. (2010) Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, **329,** 853-856.

Thioulouse, J., Chessel, D. & Champely, S.p. (1995) Multivariate analysis of spatial patterns: a unified approach to local and global structures. *Environmental and Ecological Statistics*, **2,** 1-14.

Thomas, A.C., Deagle, B.E., Eveson, J.P., Harsch, C.H. & Trites, A.W. (2016) Quantitative DNA metabarcoding: improved estimates of species proportional biomass using correction factors derived from control material. *Molecular Ecology Resources*, **16,** 714-726.

Vacher, C., Piou, D. & Desprez-Loustau, M.-L. (2008) Architecture of an Antagonistic Tree/Fungus Network: The Asymmetric Influence of Past Evolutionary History. *PLOS ONE*, **3,** e1740.

Vacher, C., Tamaddoni-Nezhad, A., Kamenova, S., Peyrard, N., Moalic, Y., Sabbadin, R., …Bohan, D.A. (2016) Learning ecological networks from next-generation sequencing data. *Advances in Ecological Research* (eds G. Woodward & D.A. Bohan), pp. 1-39. Academic Press.

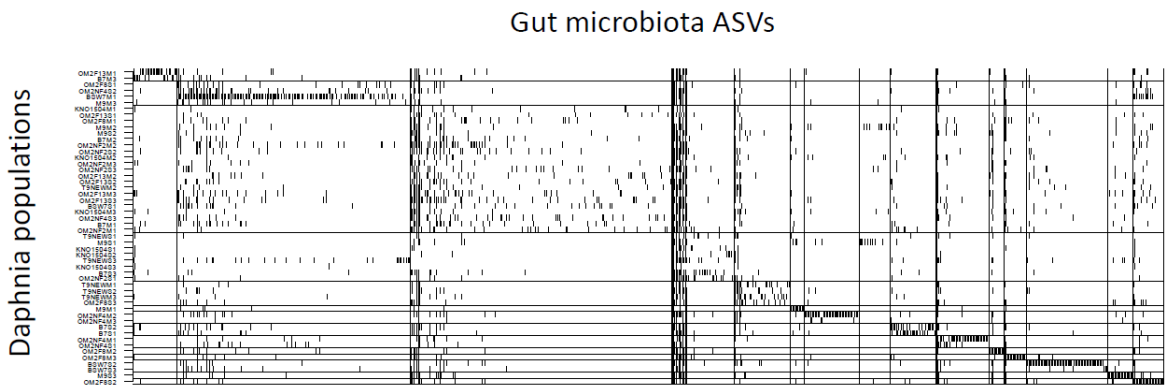
Van Der Wal, J., Falconi, L., Januchowski, S., Shoo, L. & Storlie, C. (2014) SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises. R package version 1.1-221. https://CRAN.R-project.org/package=SDMTools

Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B. & Hochberg, M.E. (2013) Phage–bacteria infection networks. *Trends in Microbiology*, **21,** 82-91.

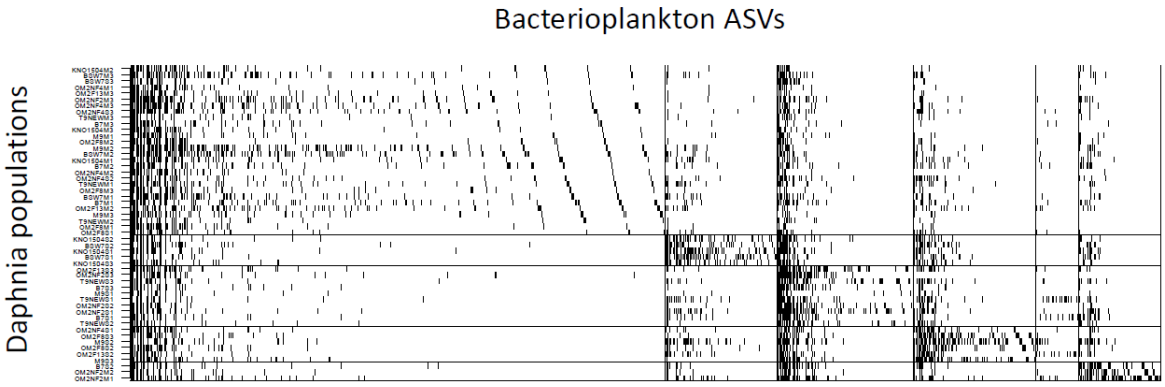
Williams, R.J. & Martinez, N.D. (2000) Simple rules yield complex food webs. *Nature*, **404,** 180-183.

Young, S.J. & Scheinerman, E.R. (2007) Random dot product graph models for social networks. *International Workshop on Algorithms and Models for the Web-Graph*, pp. 138-149. Springer.

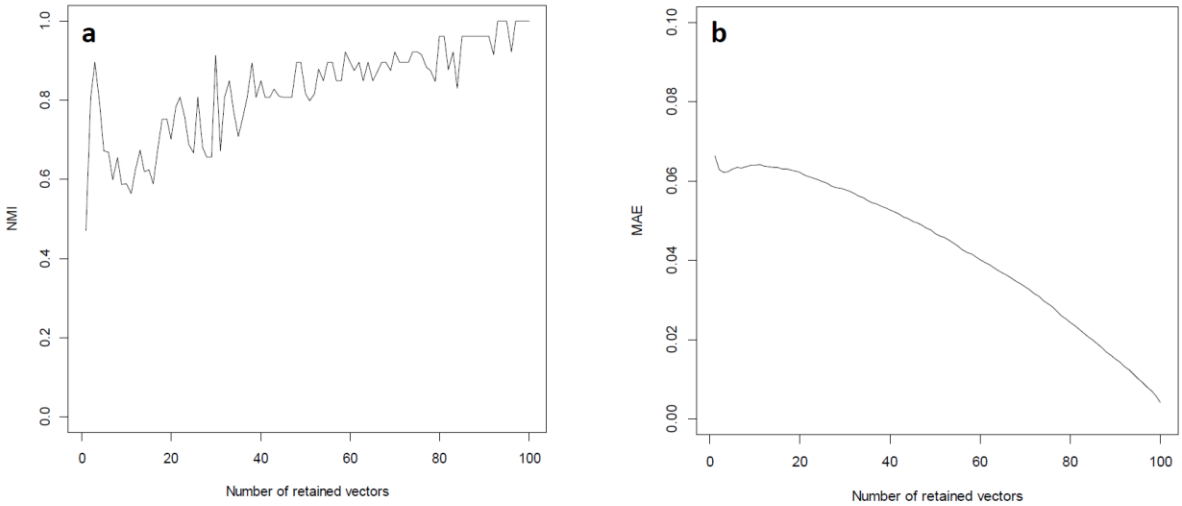
Supplementary Figures and Tables



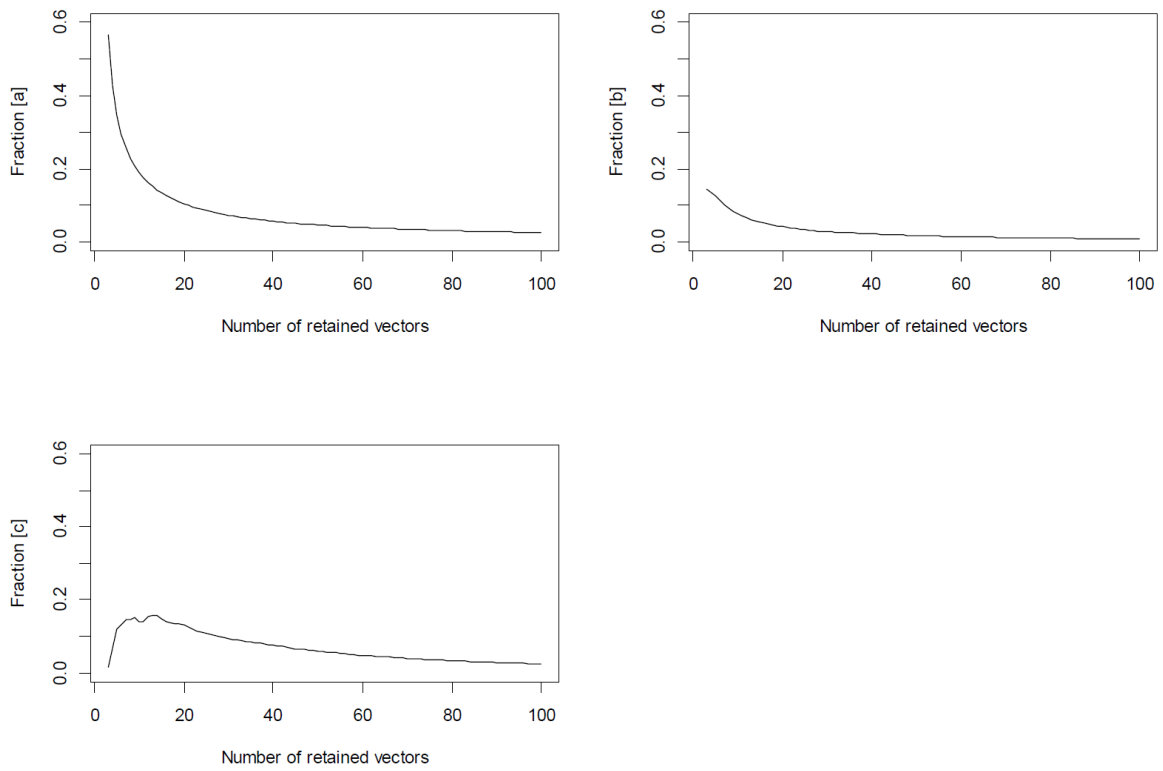
Supp. Fig. S1 – Result of the community search within the "gut microbiota network" (52 gut microbiota samples × 768 microbial ASVs). The leading-eigenvector modularity optimization algorithm evinced 16 communities, here represented by the gray lines dividing columns and rows of the incidence matrix, with dots representing existing interactions.



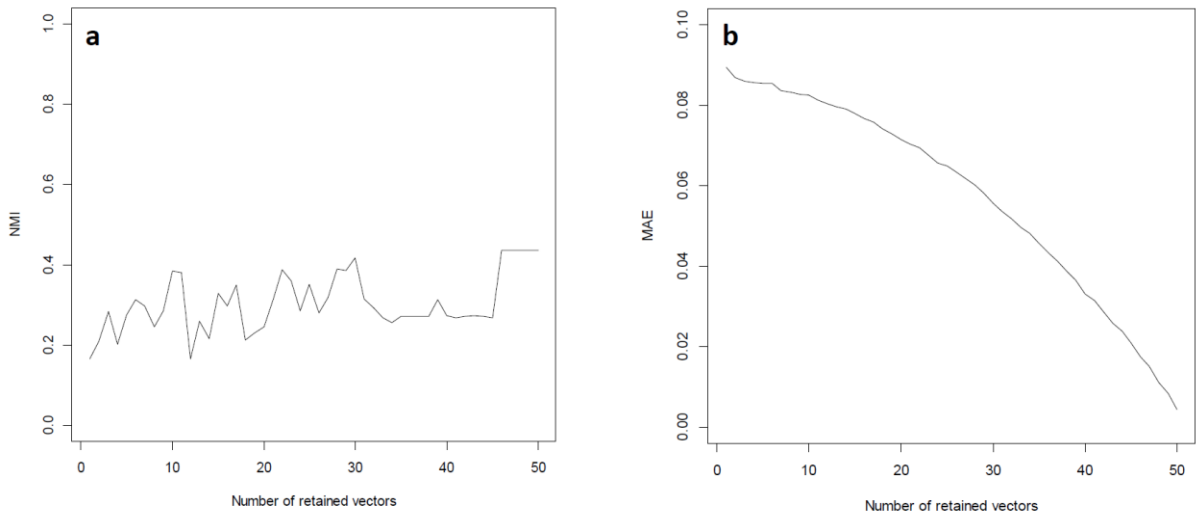
Supp. Fig. S2 – Result of the community search within the "bacterioplankton network" (52 bacterioplankton samples × 1061 microbial ASVs). The leading-eigenvector modularity optimization algorithm evinced 6 communities, here represented by the gray lines dividing columns and rows of the incidence matrix, with dots representing existing interactions.



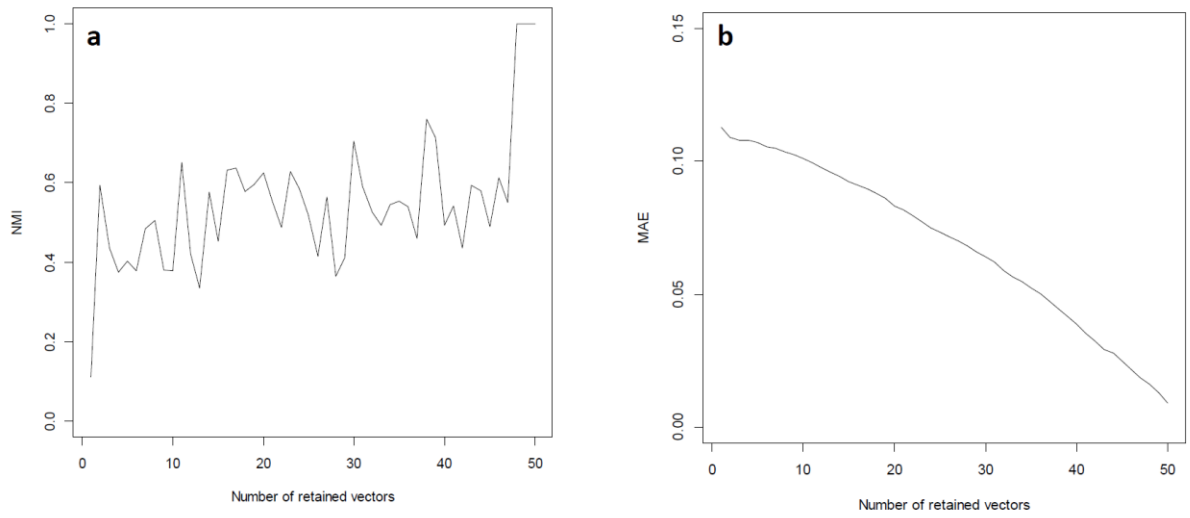
Supp. Fig. S3 – Measures of the fit of the approximation given by equation (3) for the whole network. (**a**) Normalized Mutual Information index (NMI) measuring the congruence between the communities given to *Daphnia* populations (the host nodes) in the original network and the approximated network, as a function of the number of vectors retained in the approximation of equation (3). Approximated networks were obtained using a threshold on **L.R** values optimizing the sum of sensitivity and specificity, as described in the Materials & Methods. (**b**) Mean absolute error (MAE) for approximation (3) as a function of the number of vectors retained in the approximation of equation (3). The MAE was obtained as $\frac{1}{np}\sum_{i,j}\left|b_{ij} - \sum_{k} l_{ik} r_{jk}\right|$, where **B** is the original $n \times p$ incidence matrix, **L** and **R** are the components of equation (3) and the index k is only allowed to vary up to the number of vectors retained.



Supp. Fig. S4 – Variation of the adjusted R^2 attributable to individual fractions [a] (effect: type | diet + genotype), [b] (effect: diet | type + genotype) and [c] (effect: genotype | type + diet) of the RDA applied to the whole network, as functions of the number of SVD vectors retained when approximating the incidence matrix of the network (here, varied between 3 and 100 vectors). Fractions [d]-[g] are omitted because they should be negative by definition, since all explanatory factors have been varied independently (and hence non-adjusted R^2 attributable to two factors at the same time should all be zero, see Table 4).



Supp. Fig. S5 – Measures of the fit of the approximation given by equation (3) for the gut microbiota network. (**a**) Normalized Mutual Information index (NMI) measuring the congruence between the communities given to *Daphnia* populations (the host nodes) in the original network and the approximated network, as a function of the number of vectors retained in the approximation of equation (3). Approximated networks were obtained using a threshold on **L.R** values optimizing the sum of sensitivity and specificity, as described in the Materials & Methods. (**b**) Mean absolute error (MAE) for approximation (3) as a function of the number of vectors retained in the approximation of equation (3). Further details in the caption of Supp. Fig. S3.



Supp. Fig. S6 – Measures of the fit of the approximation given by equation (3) for the bacterioplankton network. (**a**) Normalized Mutual Information index (NMI) measuring the congruence between the communities given to *Daphnia* populations (the host nodes) in the original network and the approximated network, as a function of the number of vectors retained in the approximation of equation (3). Approximated networks were obtained using a threshold on **L.R** values optimizing the sum of sensitivity and specificity, as described in the Materials & Methods. (**b**) Mean absolute error (MAE) for approximation (3) as a function of the number of vectors retained in the approximation of equation (3). Further details in the caption of Supp. Fig. S3.

Sub-network	Modularity	# modules	Mod. p-value	NMI	NMI p-value
gut microbiota × M	0.333	16	0.0481	0.807	0.0001
gut microbiota × S	0.299	13	0.3621	0.610	0.2964
bacterioplankton × M	0.238	7	0.5327	0.502	0.5003
bacterioplankton × S	0.184	2	0.3892	0.172	0.9324

Supplementary Table S1 – Community search and congruence of communities with *Daphnia* genotypes within each of the four sub-networks defined by treatment, i.e. for each of the combination of sample type (gut microbiota vs. bacterioplankton) and each of the diet (*Scenedesmus* [S] vs. *Microcystis* [M]). Modularity: modularity score obtained for this sub-network; # modules: number of communities maximizing modularity; mod. p-value: p-value of the edge randomization test for modularity; NMI: normalized mutual information index obtained by comparing the classification by communities and the classification by genotypes; NMI p-value: p-value of the edge randomization test for NMI.

Fraction	Effect	df	adjusted R^2	Row perm. p-value	Edge perm. p-value
[adfg]	type	1	0.32	0.0001	0.0002
[bdeg]	diet	1	0.11	0.0001	0.0001
[cefg]	genotype	8	0.08	0.0004	0.0003
[abdefg]	type+diet	2	0.43	0.0001	0.0001
[acdefg]	type+genotype	9	0.42	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.21	0.0001	0.0001
[abcdefg]	type+diet+genotype	10	0.55	0.0001	0.0001
[a]	type | diet+genotype	1	0.35	0.0001	0.0002
[b]	diet | type+genotype	1	0.13	0.0001	0.0001
[c]	genotype | type+diet	8	0.12	0.0001	0.0003
[d]	-	-	-0.01	-	0.9991
[e]	-	-	-0.01	-	1.0000
[f]	-	-	-0.03	-	0.9984
[g]	-	-	0.00	-	0.0010
[h]	residuals	-	0.45	-	-
[ad]	type | genotype	1	0.34	0.0001	0.0002
[af]	type | diet	1	0.32	0.0001	0.0002
[bd]	diet | genotype	1	0.12	0.0001	0.0001
[be]	diet | type	1	0.12	0.0001	0.0001
[ce]	genotype | type	8	0.11	0.0001	0.0004
[cf]	genotype | diet	8	0.09	0.0003	0.0003

Supplementary Table S2 – Results of the redundancy analysis / singular value decomposition (RDA-SVD, with the first five vectors retained) analysis applied to the whole network (comprising both gut microbiota and bacterioplankton samples), trying to explain network communities using sample type (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. All else as in Table 4.

Fraction	Effect	df	adjusted R^2	Row perm. p-value	Edge perm. p-value
[adfg]	type	1	0.14	0.0001	0.0002
[bdeg]	diet	1	0.05	0.0001	0.0001
[cefg]	genotype	8	0.14	0.0001	0.0001
[abdefg]	type+diet	2	0.19	0.0001	0.0001
[acdefg]	type+genotype	9	0.29	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.20	0.0001	0.0001
[abcdefg]	type+diet+genotype	10	0.35	0.0001	0.0001
[a]	type | diet+genotype	1	0.15	0.0001	0.0002
[b]	diet | type+genotype	1	0.06	0.0001	0.0001
[c]	genotype | type+diet	8	0.16	0.0001	0.0001
[d]	-	-	0.00	-	0.9993
[e]	-	-	-0.01	-	1.0000
[f]	-	-	-0.01	-	0.9999
[g]	-	-	0.00	-	0.0007
[h]	residuals	-	0.65	-	-
[ad]	type | genotype	1	0.15	0.0001	0.0002
[af]	type | diet	1	0.14	0.0001	0.0002
[bd]	diet | genotype	1	0.06	0.0001	0.0001
[be]	diet | type	1	0.06	0.0001	0.0001
[ce]	genotype | type	8	0.15	0.0001	0.0001
[cf]	genotype | diet	8	0.14	0.0001	0.0001

Supplementary Table S3 – Results of the redundancy analysis / singular value decomposition (RDA-SVD, with the first 13 vectors retained) analysis applied to the whole network (comprising both gut microbiota and bacterioplankton samples), trying to explain network communities using sample type (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. All else as in Table 4.

Fraction	Effect	df	adjusted R^2	Row perm. p-value	Edge perm. p-value
[adfg]	type	1	0.02	0.0001	0.0001
[bdeg]	diet	1	0.01	0.0001	0.0001
[cefg]	genotype	8	0.02	0.0001	0.0001
[abdefg]	type+diet	2	0.03	0.0001	0.0001
[acdefg]	type+genotype	9	0.05	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.03	0.0001	0.0001
[abcdefg]	type+diet+genotype	10	0.06	0.0001	0.0001
[a]	type | diet+genotype	1	0.03	0.0001	0.0001
[b]	diet | type+genotype	1	0.01	0.0001	0.0001
[c]	genotype | type+diet	8	0.02	0.0001	0.0001
[d]	-	-	0.00	-	1.0000
[e]	-	-	0.00	-	1.0000
[f]	-	-	0.00	-	1.0000
[g]	-	-	0.00	-	0.0007
[h]	residuals	-	0.94	-	-
[ad]	type | genotype	1	0.03	0.0001	0.0001
[af]	type | diet	1	0.02	0.0001	0.0001
[bd]	diet | genotype	1	0.01	0.0001	0.0001
[be]	diet | type	1	0.01	0.0001	0.0001
[ce]	genotype | type	8	0.02	0.0001	0.0001
[cf]	genotype | diet	8	0.02	0.0001	0.0001

Supplementary Table S4 – Results of the redundancy analysis / singular value decomposition (RDA-SVD, with the first 100 vectors retained) analysis applied to the whole network (comprising both gut microbiota and bacterioplankton samples), trying to explain network communities using sample type (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. All else as in Table 4.

Fraction	Effect	df	adjusted R^2	Row perm. p-value	Edge perm. p-value
[adfg]	bpk	6	0.06	0.0001	0.0001
[bdeg]	diet	1	0.02	0.0001	0.0001
[cefg]	genotype	8	0.08	0.0001	0.0001
[abdefg]	bpk+diet	7	0.07	0.0001	0.0001
[acdefg]	bpk+genotype	14	0.14	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.10	0.0001	0.0001
[abcdefg]	bpk+diet+genotype	15	0.15	0.0001	0.0001
[a]	bpk | diet+genotype	6	0.05	0.0020	0.0054
[b]	diet | bpk+genotype	1	0.01	0.0391	0.0751
[c]	genotype | bpk+diet	8	0.08	0.0001	0.0001
[d]	-	-	0.01	-	0.0032
[e]	-	-	0.00	-	0.1652
[f]	-	-	0.01	-	0.0165
[g]	-	-	0.00	-	0.9819
[h]	residuals	-	0.85	-	-
[ad]	bpk | genotype	6	0.05	0.0001	0.0001
[af]	bpk | diet	6	0.05	0.0002	0.0001
[bd]	diet | genotype	1	0.02	0.0002	0.0001
[be]	diet | bpk	1	0.01	0.0105	0.0021
[ce]	genotype | bpk	8	0.08	0.0001	0.0001
[cf]	genotype | diet	8	0.09	0.0001	0.0001

Supplementary Table S5 – Results of the redundancy analysis / singular value decomposition (RDA-SVD, with the first 30 vectors retained) analysis applied to the gut microbiota-based network, trying to explain network communities using the SVD-reduced matrix for the bacterioplankton network ("bpk", forward selection of 6 vectors), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. All else as in Table 5.

Fraction	Effect	df	adjusted R^2	Row perm. p-value	Edge perm. p-value
[adfg]	gut	1	0.02	0.0001	0.0001
[bdeg]	diet	1	0.03	0.0001	0.0001
[cefg]	genotype	8	0.09	0.0001	0.0001
[abdefg]	gut+diet	2	0.05	0.0001	0.0001
[acdefg]	gut+genotype	9	0.11	0.0001	0.0001
[bcdefg]	diet+genotype	9	0.13	0.0001	0.0001
[abcdefg]	gut+diet+genotype	10	0.13	0.0001	0.0001
[a]	gut | diet+genotype	1	0.00	0.2973	-
[b]	diet | gut+genotype	1	0.03	0.0001	0.0001
[c]	genotype | gut+diet	8	0.09	0.0001	0.0001
[d]	-	-	0.01	-	0.0001
[e]	-	-	0.00	-	0.9175
[f]	-	-	0.01	-	0.0016
[g]	-	-	-0.01	-	0.9988
[h]	residuals	-	0.87	-	-
[ad]	gut | genotype	1	0.01	0.0038	0.0002
[af]	gut | diet	1	0.01	0.0001	0.0024
[bd]	diet | genotype	1	0.04	0.0001	0.0001
[be]	diet | gut	1	0.03	0.0001	0.0001
[ce]	genotype | gut	8	0.09	0.0001	0.0001
[cf]	genotype | diet	8	0.10	0.0001	0.0001

Supplementary Table S6 – Results of the redundancy analysis / singular value decomposition (RDA-SVD, with the first 30 vectors retained) analysis applied to the bacterioplankton-based network, trying to explain network communities using the SVD-reduced matrix for the gut microbiota network ("gut" forward selection of 1 vector), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. All else as in Table 6.