



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-02942286>

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.

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

4

5 François Massol^{1,2}, Emilie Macke³, Martijn Callens^{3,4}, Ellen Decaestecker³

6

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

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

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

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

13

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

17

18 **Running headline:** Analyzing host-microbiota networks

19

20

21 **Abstract**

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

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

51

52 **Keywords**

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

55 **Data archiving**

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

57

58 **Introduction**

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

80

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

85 Loustau, 2008) or mutualistic (Encinas-Viso *et al.*, 2016) interaction networks, phage-bacteria
86 infection networks (Weitz *et al.*, 2013), or mammal species-gut microbiota associations (Ley *et al.*,
87 2008). Although inferring true interactions from co-occurrence is a difficult endeavour (Vacher *et al.*,
88 2016; Bohan *et al.*, 2017; Derocles *et al.*, 2018) – indeed, neither do co-occurrences necessarily imply
89 interaction nor does the absence of co-occurrence imply the absence of interaction –, the analysis of
90 host-microbe association networks can still benefit from the use of ecological network methods. In
91 particular, since host-microbe associations are much more amenable to controlled experiments than
92 marine food webs or plant-pollinator networks, they can provide a good starting point to test methods
93 aimed at elucidating the drivers of network structure because drivers can be varied independently, thus
94 removing the possibility of confounding effects.

95

96 As stated above, classic network methods do not measure the respective effects of different
97 environmental and biological drivers on network structure. A few methods have been proposed (see
98 e.g. Kamenova *et al.*, 2017 for a short review of existing models) to assess network structure as the
99 result of latent or explicit traits (such as organism size), sometimes combining the information
100 provided by traits with that provided by phylogenies or geographical distributions of species (Rohr *et*
101 *al.*, 2010; Gravel *et al.*, 2013; Ovaskainen *et al.*, 2016; Rohr *et al.*, 2016), but not in an integrative
102 framework allowing all types of external factors to be tested. However, one very promising method
103 (Dalla Riva & Stouffer, 2016) based on low-dimension embedding of adjacency matrices through
104 singular value decomposition allows the partitioning of network ‘inertia’ through the use of classic
105 multivariate redundancy analyses (Sabatier, Lebreton & Chessel, 1989; Borcard, Legendre & Drapeau,
106 1992; Dray, Legendre & Peres-Neto, 2006; Peres-Neto *et al.*, 2006) and has been successfully applied
107 to the study of orchid-pollinator interactions across Europe (Joffard *et al.*, 2019). In parallel, the study
108 of multi-layer network structure, and notably the search for congruence between ‘mesoscale’
109 structures (i.e. modules or blocks) in two paired networks, has led to the development of a comparison
110 method based on classifications obtained by modularity optimization and the use of classification
111 congruence indices (Astegiano, Altermatt & Massol, 2017).

112

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

125

126

127 **Materials & Methods**

128 *Host-microbiota data as interaction networks*

129 In the following, we will consider bipartite networks, i.e. networks involving two disjoint sets of
130 nodes, called levels, with edges only connecting nodes from two different levels. A host-microbiota
131 association network is a bipartite network, with host populations or individuals forming one level, and
132 microbial species, the other one. A bipartite network, with n nodes in the first level and p nodes in the
133 second level, can be mathematically represented using an incidence matrix \mathbf{B} of dimensions n by p .
134 We will generally assume that the networks under study are not weighted, i.e. an existing link between
135 two nodes is coded as “1” and an absent link is noted “0”. Indeed, since microbial sequence counts can
136 vary widely between samples (e.g. due to PCR amplification heterogeneity), only relative microbial
137 abundances can be obtained per host, which prevents the absolute quantification of links between host
138 and microbial nodes (Amend, Seifert & Bruns, 2010; McMurdie & Holmes, 2014; Thomas *et al.*,
139 2016). For datasets able to more accurately quantify link weights, most of the analyses presented
140 below could be adapted to weighted networks, e.g. by using block search instead of community search
141 (Leger, Daudin & Vacher, 2015), and are further tackled in the Discussion.

142

143 *Community detection*

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

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

160

161 *Null models for bipartite networks*

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

174

175 *Approximating networks*

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

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

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

$$186 \quad \mathbf{B} \approx \mathbf{M}_1 \cdot \mathbf{B}' \cdot \mathbf{M}_2 \quad (1)$$

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

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

$$203 \quad \mathbf{B} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V}^T \quad (2)$$

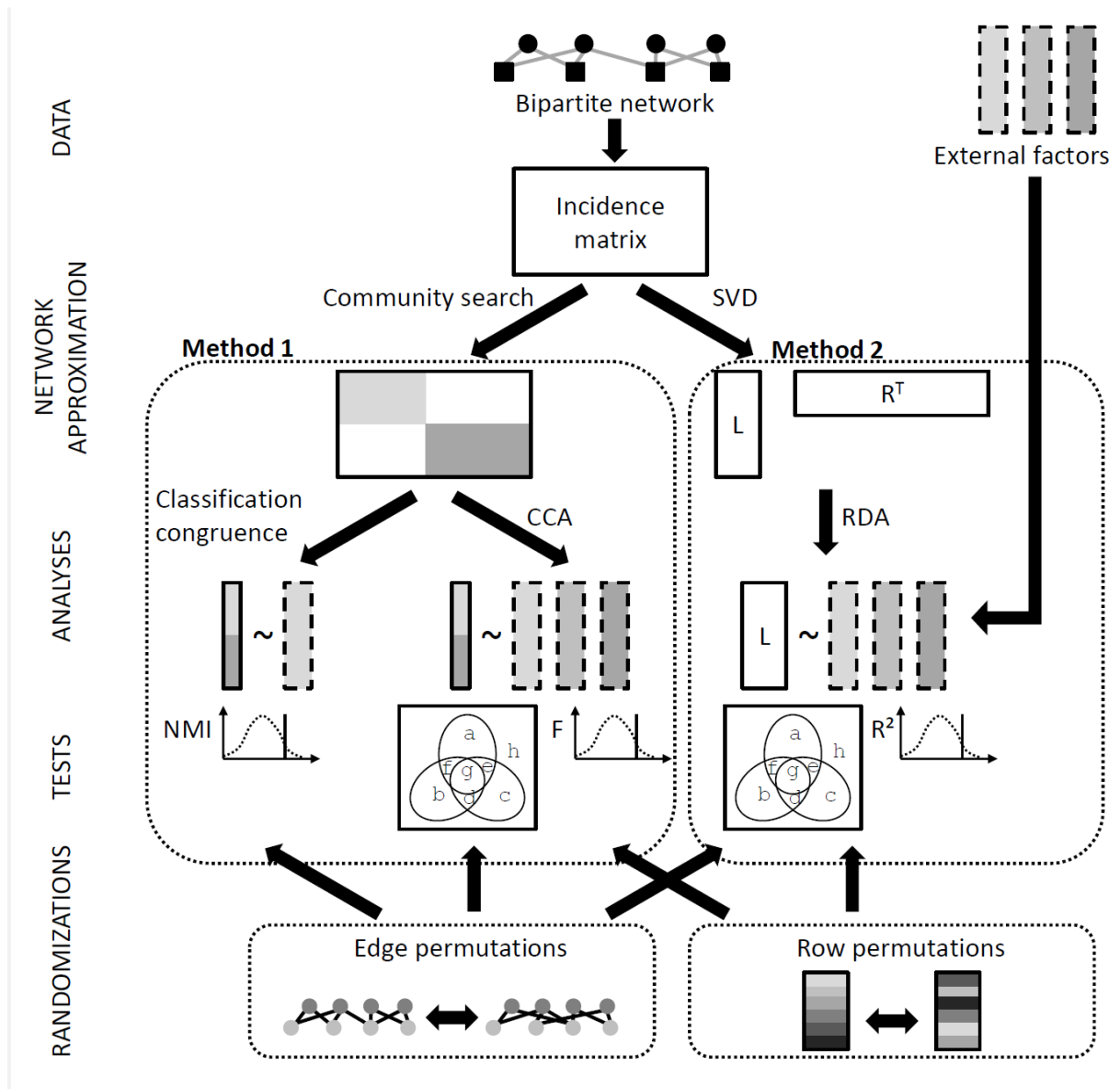
204 where \mathbf{U} and \mathbf{V} are orthogonal matrices, \mathbf{V}^T denotes the transposed version of \mathbf{V} , and the square matrix
205 \mathbf{D} is diagonal and its values are the singular values of matrix \mathbf{B} (all non-negative), usually sorted in

206 decreasing order. As Dalla Riva and Stouffer (2016) note, a useful approximation of the network can
207 be obtained by finding the square-root of \mathbf{D} , noted \mathbf{S} , and define a number of ‘latent traits’ q less than
208 the number of singular values, a matrix \mathbf{L} as the first q columns of $\mathbf{U}\mathbf{S}$ and a matrix \mathbf{R} as the first q
209 columns of $\mathbf{V}\mathbf{S}$, so that:

$$210 \quad \mathbf{B} \approx \mathbf{L}\mathbf{R}^T \quad (3)$$

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

214



215
 216 **Fig. 1** – Summary of the proposed methods. Starting from an incidence matrix describing a bipartite
 217 network, one can either (i) perform a community search, then work on node communities by assessing
 218 their congruence with other classifications or perform a canonical correspondence analysis (CCA)
 219 with respect to several external factors (method 1), or (ii) approximate the incidence matrix as the
 220 product of two reduced ones through a singular value decomposition (SVD) and then analyze these
 221 reduced matrices using a redundancy analysis (RDA; method 2). In both cases, the results of analyses
 222 can be tested twice: (i) through row permutations, one can assess whether effects would have been
 223 expected from the imbalance and correlations between external factors (RDA and CCA); (ii) through
 224 edge permutations on the initial network, one can test whether an effect significant for the first test is
 225 only due to differences in numbers of interactions between factor levels, if the second test is not
 226 significant (a richness effect, e.g. a diet effect because *Daphnia* fed with *Scenedesmus* interact with
 227 more bacteria species) or, if the second is also significant, is due to an affinity effect (i.e. factor levels
 228 selectively associated with certain interactions).

230 ***Method 1: node classification-based tests***

231 To assess the effects of categorical drivers on network organization, a first method is to test the
232 congruence of node classifications obtained through community-search algorithms with those
233 associated with external categorical variables, i.e. to study the links between matrices \mathbf{M}_1 and \mathbf{M}_2 from
234 the previous section with matrices describing external categorical factors. Such a method is useful for
235 analyzing the results of controlled experiments since external categorical variables then amount to
236 treatments. Any grouping of the nodes (host, microbiota, or both) is effectively a classification in the
237 statistical sense, which can be compared with other classifications of the same data (Danon *et al.*,
238 2005). Following Astegiano *et al.* (2017), we propose to use the Normalized Mutual Information index
239 (NMI) to gauge the congruence of two classifications. The NMI takes values between 0 and 1, 0
240 indicating no congruence and 1, perfect congruence. To test for the significance of a given NMI
241 between two classifications, at least one of which being the classification of network nodes in
242 communities, the network is randomized using the ‘curveball’ algorithm (Strona *et al.*, 2014). This
243 computation of NMI, implemented using the function ‘compare’ in the R package ‘igraph’ (Csardi &
244 Nepusz, 2006), can help assess the effect of a single external factor on network structure.

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

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

265

266 ***Method 2: singular value decomposition-based tests***

267 As mentioned above, the approximation of a network by community memberships overlooks
268 differences in degrees among nodes belonging to the same community. Since these differences can
269 also hold some of the underlying network structure, we propose to also model the effects of external
270 variables on network structure using the RDPG decomposition proposed by Dalla Riva and Stouffer
271 (2016; equation [3]). Following equation (3), a given $n \times p$ bipartite network can be approximated as
272 two matrices (**L** and **R**) with a low number of columns and as many rows as nodes (n in **L**, p in **R**).
273 Matrices **L** and **R** can be analysed through a Redundancy Analysis (RDA) to gauge how much
274 variation among rows is explained by external variables, similarly to what is performed for CCA (i.e.
275 quantifying the variation explained by each fraction such as that explained by factors X or Y once the
276 effect of Z has been removed, Joffard *et al.*, 2019). In RDA, variation is understood in the classic sum-
277 of-square sense and can be quantified using adjusted R^2 (Peres-Neto *et al.*, 2006). Because the
278 information stored in matrices **L** and **R** is represented by real numbers which do not correspond to
279 presences or absences, multivariate analyses such as RDA or distance-based RDA (Legendre &
280 Gallagher, 2001) applied to these matrices do not suffer from the ‘double zero’ problem (Legendre &
281 Legendre, 2012). This means that different multivariate approaches similar to classic RDA could be
282 applied instead (Blanchet *et al.*, 2014), but lacking a clear rationale for favouring one over the others,
283 we will tackle these other approaches in the Discussion.

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

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

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

311 optimize to find the “best vectors” (e.g. minimize Mahalanobis distance between motif relative
312 frequency vectors of observed vs. approximated networks). Dealing with all possible focal structures
313 and the way to best represent them using SVD vectors is, however, beyond the point of this study.
314 Regarding the number of vectors/columns in the explanatory tables (obtained from networks or
315 otherwise extracted in a way that allows choosing which vectors to use or not, e.g. like Moran
316 Eigenvector Maps, Dray et al. 2006), we use forward selection with double stopping criteria (p-value
317 <0.05 and adjusted R² less than that found with all vectors), as described by Blanchet et al. (2008) and
318 advocated by Bauman et al. (2018). This was practically implemented borrowing from the code
319 proposed by Bauman et al. (2018), using functions ‘RsquareAdj’ and ‘forward.sel’ from the R package
320 ‘adespatial’ (Dray *et al.*, 2019).

321

322 ***Application: Daphnia magna gut microbiota and bacterioplankton data***

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

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

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

338 2L glass bottles with constant stirring and aeration. Filters (0.22 μm) were placed at the input and the
339 output of the aeration system to avoid bacterial contamination. Algae were harvested weekly in early
340 stationary phase. Axenity was checked on LB medium agar plates.

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

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

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

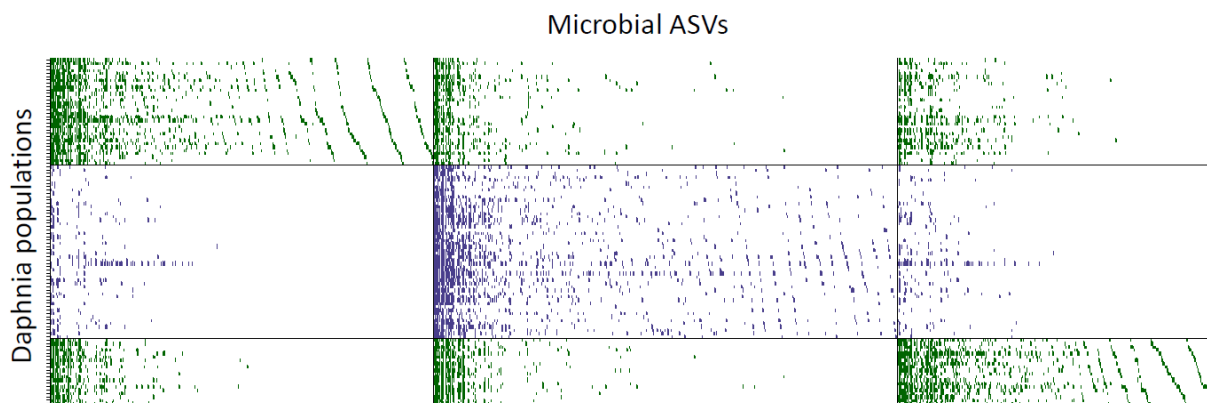
366 amplified with primers 27F and 1492R on 10 ng of template (94°C - 30s; 50°C - 45s; 68°C - 90s; 30
367 cycles) using a high-fidelity Pfx polymerase (Life technologies). PCR products were subsequently
368 purified using the QIAquick PCR purification kit (Qiagen). To obtain dual-index amplicons of the V4
369 region, a second amplification was performed on 5 µL PCR product using primers 515F and 806R for
370 30 cycles (94°C - 30s; 55°C - 30s; 68°C - 60s). Both primers contained an Illumina adapter and an 8-
371 nt barcode at the 5'-end. For each sample, PCRs were performed in triplicate, pooled and gel-purified
372 using the QIAquick gel extraction kit (Qiagen). An equimolar library was prepared by normalizing
373 amplicon concentrations with a SequalPrep Normalization Plate (Applied Biosystems) and subsequent
374 pooling. Amplicons were sequenced using a v2 PE500 kit with custom primers on the Illumina Miseq
375 platform (KU Leuven Genomics Core) producing 2 × 250-nt paired-end reads. Sequence reads were
376 processed using R package 'phyloseq', following Callahan *et al.* (2016b). Sequences were trimmed
377 (the first 10 nucleotides and from position 190 onwards were removed) and filtered (maximum of 2
378 expected errors per read) on paired ends jointly. Sequence variants were inferred using the high-
379 resolution DADA2 method (Callahan *et al.*, 2016a), and chimeras were removed. Taxonomy was
380 assigned with a naive Bayesian classifier using the RDP v14 training set. Amplicon Sequence Variants
381 (ASV) with no taxonomic assignment at phylum level or which were assigned as "Chloroplast" or
382 "Cyanobacteria" were removed from the dataset. The final dataset contained 1,500,800 reads, on
383 average 29,427 reads per sample (min. = 5,804 reads, max. = 78,154 reads).

384

385 **Results**

386 *Communities in the bipartite networks*

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



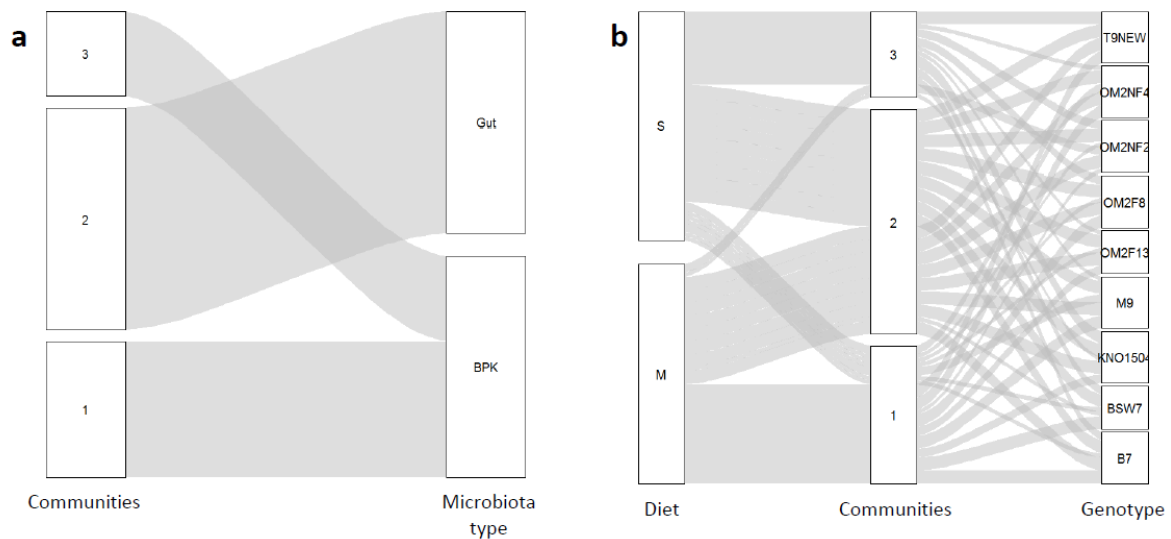
391
392 **Fig. 2** – Result of the community search within the whole network (all 104 *Daphnia* samples \times 1656
393 microbial ASV). The leading-eigenvector modularity optimization algorithm evinced three
394 communities, here represented by the gray lines dividing columns and rows of the incidence matrix,
395 with dots representing existing interactions (green dots for bacterioplankton, blue dots for gut
396 microbiota). Communities found by the algorithms perfectly correspond to sample types, with two
397 groups within bacterioplankton and one fitting all gut microbiota interactions.

398
399 The same community-search algorithm was also applied to the two sub-networks obtained by taking
400 only gut microbiota or bacterioplankton samples (Supp. Figs S1 and S2). In the gut microbiota
401 network, 16 communities were found (Supp. Fig. S1), with a moderate and not significant modularity
402 score ($Q = 0.242$, $p = 0.1739$). In the bacterioplankton network, 6 communities were found (Supp. Fig.
403 S2), with a moderate but significant modularity ($Q = 0.216$, $p = 0.0021$). These 6 communities were
404 poorly related to the two-community division of the bacterioplankton network obtained by running the
405 community-search algorithm on the whole network (results not shown).

406

407 **Congruence of classifications**

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



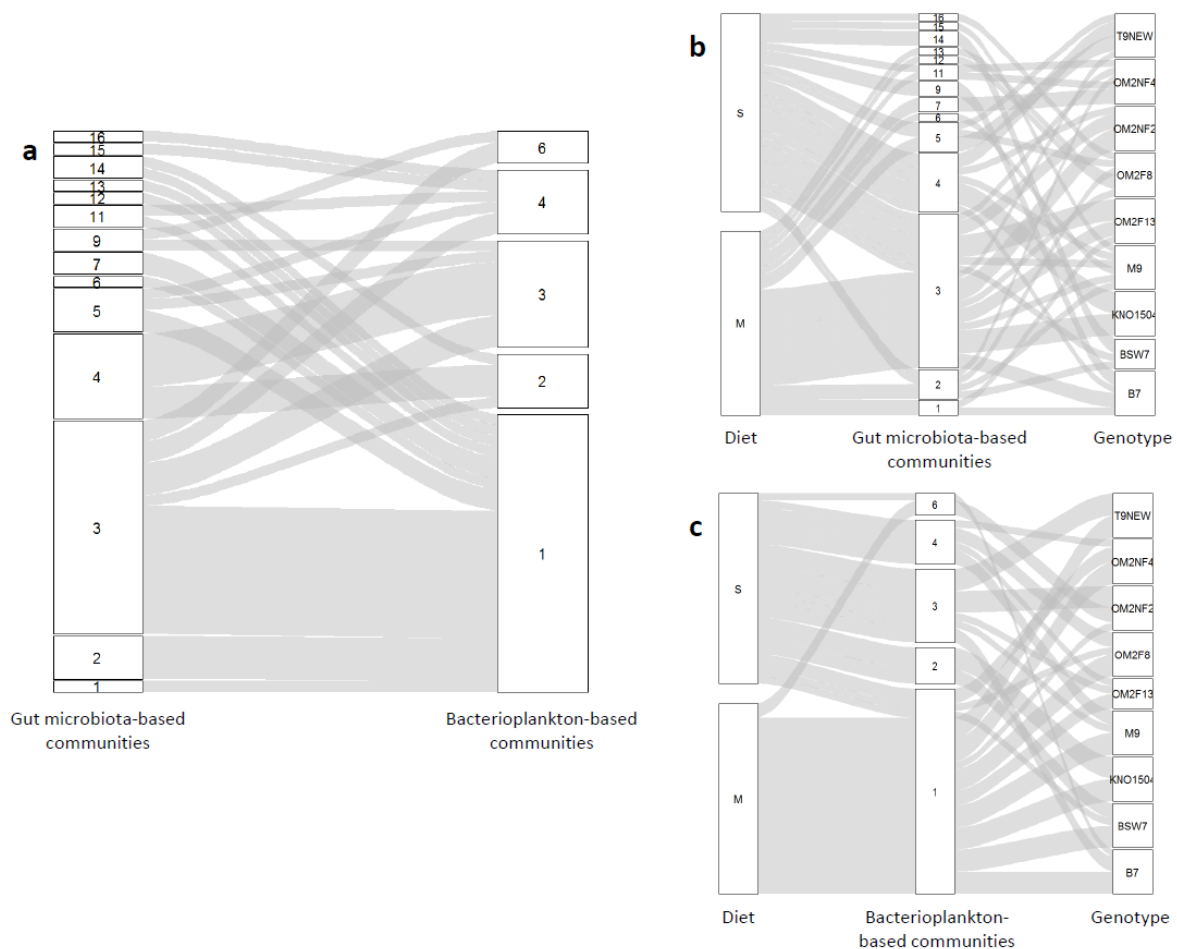
412

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

423

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

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



437

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

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

449

450 **Canonical correspondence analyses**

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

462

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	-

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

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

473

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

481

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

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

486

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

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

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

502

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

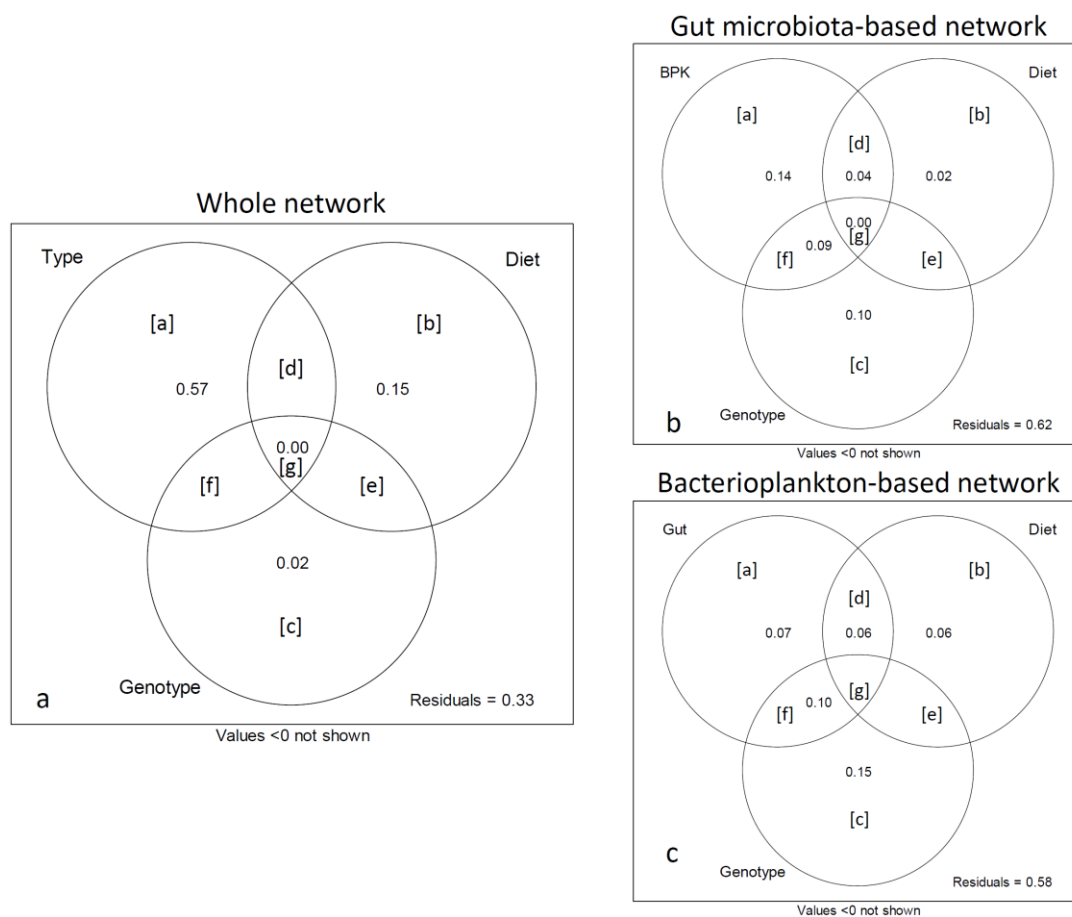
503 **Table 3** – Results of the CCA applied to the bacterioplankton-based network, trying to explain
 504 network communities using communities found in the gut microbiota-based network (“gut”), diet
 505 (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. See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

507

508 ***Singular value decomposition and redundancy analyses***

509 When modelling the whole network, visual inspection of how congruent the modules obtained from
 510 the approximated networks were with those of the observed network indicated that congruence had
 511 several local maxima (Supp. Fig. S3a), while mean absolute error (MAE) declined steadily with
 512 number of vectors (Supp. Fig. S3b). The adjusted R^2 of individual fractions attributable to sample type
 513 or diet also declined with the number of vectors, while the fraction attributable to genotype had a
 514 maximum at 13 vectors (Supp. Fig. S4). Because the first maximum of NMI was found with three
 515 vectors, we present all RDA results on the whole network using three vectors only. However, given
 516 the patterns reported in Supp. Fig. S4, we also checked robustness with different number of vectors (5,
 517 13, and 100; Supp. Tables S2-S4).

518



519

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

524 fractions (i.e. [a] for Type, not [adfg]), all adjusted R^2 are given in Tables 4-6. Tested factors
525 comprise: “type”, the type of microbiota sample (bacterioplankton vs. gut microbiota); “diet”, the diet
526 treatment (*Scenedesmus* vs. *Microcystis*); “genotype”, the genotype of *Daphnia* populations; “BPK”,
527 reduced matrix obtained from applying the SVD on the bacterioplankton-based incidence matrix;
528 “Gut”, reduced matrix obtained from applying the SVD on the gut microbiota-based incidence matrix.
529 SVD-reduced matrices comprise different numbers of vectors (see main text for details). (a) Results of
530 the RDA applied to the whole network comprising both gut microbiota and bacterioplankton samples.
531 (b) Results of the RDA applied to the gut microbiota-based network. (c) Results of the RDA applied to
532 the bacterioplankton-based network.
533

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	-

534 **Table 4** – Results of the RDA-SVD applied to the whole network with the first three vectors retained,
535 to explain network communities using sample type (gut microbiota vs. bacterioplankton), diet
536 (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet), and *Daphnia* genotype. *Fraction*:
537 symbolic representation of the components of variations explained by the different factors as
538 represented in Fig. 5; *Effect*: the explanatory effects and the conditioning ones (figured after the
539 vertical line); *df*: degrees of freedom of the explanatory variables; *R²*: values of the corresponding
540 coefficient of determination (expressed in percentage); *adjusted R²*: R^2 values corrected for the number
541 of degrees of freedom; *row perm. p-value*: probability that a randomized version of the explained
542 community table, once removed the effect of conditioning variables, obtains an adjusted R^2 equal or
543 larger to the one obtained with real data; *edge perm. p-value*: probability that a randomized version of
544 the *Daphnia*-microbial ASV network, keeping node degrees constant, obtains an adjusted R^2 equal or
545 larger than the one obtained with real data. Dashes indicate values that cannot be computed and/or that
546 cannot be tested.

547

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

560 In the gut microbiota network, the MAE of the approximated network steadily decreased with the
561 number of retained vectors (Supp. Fig. S5b) while congruence between its communities and those of
562 the original network displayed several local maxima between *ca.* 5 and 30 vectors (Supp. Fig. S5a). As
563 the first important local NMI maximum was found for 10 retained vectors, we present all RDA results
564 on gut microbiota using the first 10 columns of approximation (3). However, robustness checks were
565 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

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

572

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

583

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

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

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

597

598

599

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

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

606

607

608 Discussion

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

617

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

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

640

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

659

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

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

672

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

689

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

711

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

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

724

725 **Acknowledgements**

726 We wish to thank L. De Meester for his constructive comments and C. Sueur and S. Sosa for the
727 invitation to submit to this special feature. Development of the methods benefited from discussions
728 with N. de Manincor, A. Fisogni, A. Berquer, A. Tasiemski, B. Schatz, M. Grenié, N. Joffard, S.
729 Robin, S. Donnet, S. Ouadah and M. Thomas. We are grateful to S. Donnet and S. Ouadah for
730 providing us with the R script for drawing alluvial diagrams. We also wish to thank I. Kaygorodova,
731 Y. Sapozhnikova, P. Brodin, and C. Daniel for the opportunity to present and discuss these methods at
732 various venues. We thank the editor, associate editor and reviewers, whose comments greatly
733 improved the clarity of the paper. Funding for the experiment was provided by the KU Leuven
734 C16/2017/02 project, and an FWO postdoctoral fellowship (N°12R4917N, to EM) and FWO
735 G092619N to EDC. FM was funded by the CNRS and ANR projects ARSENIC (grant no. 14-CE02-
736 0012), NGB (grant no. 17-CE32-0011), and ECONET (grant no. 18-CE02-0010). This research
737 project was performed within the framework of the FWO EVENET network.

738

739 **Author contributions**

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

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

744

745 **Data accessibility**

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

747

748 **References**

- 749 Amend, A.S., Seifert, K.A. & Bruns, T.D. (2010) Quantifying microbial communities with 454
750 pyrosequencing: does read abundance count? *Molecular Ecology*, **19**, 5555-5565.
- 751 Astegiano, J., Altermatt, F. & Massol, F. (2017) Disentangling the co-structure of multilayer
752 interaction networks: degree distribution and module composition in two-layer bipartite
753 networks. *Scientific Reports*, **7**, 15465.
- 754 Bartomeus, I. (2013) Understanding linkage rules in plant-pollinator networks by using hierarchical
755 models that incorporate pollinator detectability and plant traits. *PLOS ONE*, **8**, e69200.
- 756 Bartomeus, I., Gravel, D., Tylianakis, J.M., Aizen, M.A., Dickie, I.A. & Bernard-Verdier, M. (2016)
757 A common framework for identifying linkage rules across different types of interactions.
758 *Functional Ecology*, **30**, 1894-1903.
- 759 Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003) The nested assembly of plant-animal
760 mutualistic networks. *Proceedings of the National Academy of Sciences of the U.S.A.*, **100**,
761 9383-9387.
- 762 Bascompte, J. & Stouffer, D.B. (2009) The assembly and disassembly of ecological networks.
763 *Philosophical Transactions of the Royal Society B-Biological Sciences*, **364**, 1781-1787.
- 764 Bauman, D., Drouet, T., Dray, S. & Vlemingckx, J. (2018) Disentangling good from bad practices in
765 the selection of spatial or phylogenetic eigenvectors. *Ecography*, **41**, 1638-1649.
- 766 Blanchet, F.G., Legendre, P., Bergeron, J.A.C. & He, F. (2014) Consensus RDA across dissimilarity
767 coefficients for canonical ordination of community composition data. *Ecological Monographs*,
768 **84**, 491-511.
- 769 Blanchet, F.G., Legendre, P. & Borcard, D. (2008) Forward selection of explanatory variables.
770 *Ecology*, **89**, 2623-2632.
- 771 Blüthgen, N., Menzel, F. & Blüthgen, N. (2006) Measuring specialization in species interaction
772 networks. *BMC Ecology*, **6**, 9.
- 773 Bohan, D.A., Landuyt, D., Ma, A., Macfadyen, S., Martinet, V., Massol, F., ... Woodward, G. (2016)
774 Networking our way to better ecosystem service provision. *Trends in Ecology & Evolution*,
775 **31**, 105-115.
- 776 Bohan, D.A., Vacher, C., Tamaddoni-Nezhad, A., Raybould, A., Dumbrell, A.J. & Woodward, G.
777 (2017) Next-generation global biomonitoring: Large-scale, automated reconstruction of
778 ecological networks. *Trends in Ecology & Evolution*, **32**, 477-487.
- 779 Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological
780 variation. *Ecology*, **73**, 1045-1055.
- 781 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. (2016a)
782 DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13**,
783 581.
- 784 Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J. & Holmes, S.P. (2016b) Bioconductor
785 Workflow for Microbiome Data Analysis: from raw reads to community analyses.
786 *F1000Research*, **5**, 1492-1492.

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

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

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

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

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

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

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

804 Derocles, S.A.P., Bohan, D.A., Dumbrell, A.J., Kitson, J.J.N., Massol, F., Pauvert, C., ...Evans, D.M.
805 (2018) Biomonitoring for the 21st Century: Integrating Next-Generation Sequencing Into
806 Ecological Network Analysis. *Advances in Ecological Research*, pp. 1-62. Academic Press.

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

810 Dray, S., Legendre, P. & Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for
811 principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, **196**, 483-
812 493.

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

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

817 Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D., Krasnov, B.R., ...Bascompte, J.
818 (2010) Nestedness versus modularity in ecological networks: two sides of the same coin?
819 *Journal of Animal Ecology*, **79**, 811-817.

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

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

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

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

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

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

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

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

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

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

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

845 Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McNerny, G.J., ...O'Hara, R.B.
846 (2012) Towards novel approaches to modelling biotic interactions in multispecies assemblages
847 at large spatial extents. *Journal of Biogeography*, **39**, 2163-2178.

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

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

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

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

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

858 Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., ...Gordon, J.I.
859 (2008) Evolution of mammals and their gut microbes. *Science*, **320**, 1647-1651.

860 Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., ...Raes, J. (2015)
861 Determinants of community structure in the global plankton interactome. *Science*, **348**,
862 1262073.

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

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

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

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

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

873 Nogales, M., Heleno, R., Rumeu, B., González-Castro, A., Traveset, A., Vargas, P. & Olesen, J.M.
874 (2016) Seed-dispersal networks on the Canaries and the Galápagos archipelagos: interaction
875 modules as biogeographical entities. *Global Ecology and Biogeography*, **25**, 912-922.

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

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

881 Orsini, C., Dankulov, M.M., Colomer-de-Simon, P., Jamakovic, A., Mahadevan, P., Vahdat, A.,
882 ...Krioukov, D. (2015) Quantifying randomness in real networks. *Nature Communications*, **6**,
883 8627.

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

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

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

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

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

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

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

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

904 Simmons, B.I., Cirtwill, A.R., Baker, N.J., Wauchope, H.S., Dicks, L.V., Stouffer, D.B. & Sutherland,
905 W.J. (2019) Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*,
906 **128**, 154-170.

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

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

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

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

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

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

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

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

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

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

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

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

938 Weitz, J.S., Poisot, T., Meyer, J.R., Flores, C.O., Valverde, S., Sullivan, M.B. & Hochberg, M.E.
939 (2013) Phage–bacteria infection networks. *Trends in Microbiology*, **21**, 82-91.

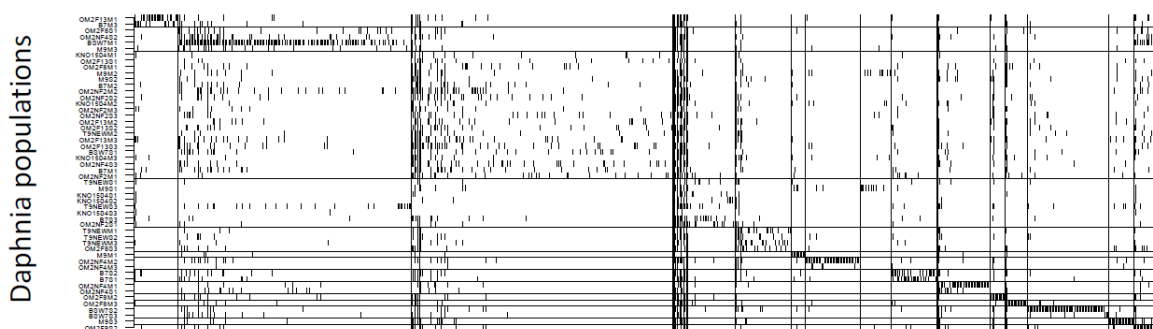
940 Williams, R.J. & Martinez, N.D. (2000) Simple rules yield complex food webs. *Nature*, **404**, 180-183.

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

943

944

Gut microbiota ASVs

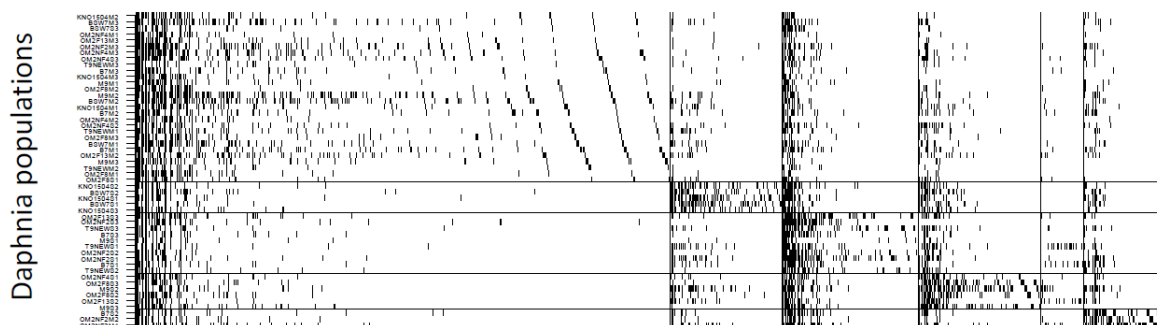


946

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

951

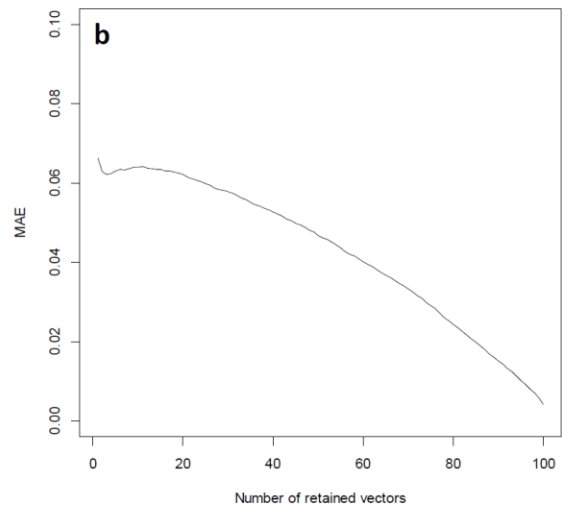
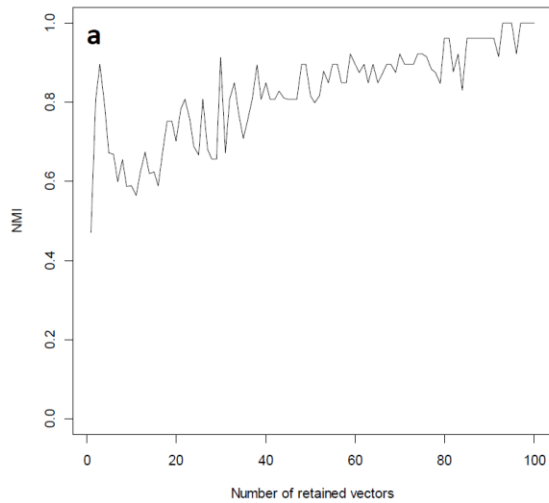
Bacterioplankton ASVs



952

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

957

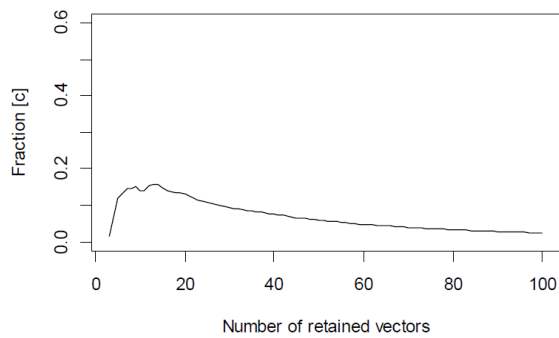
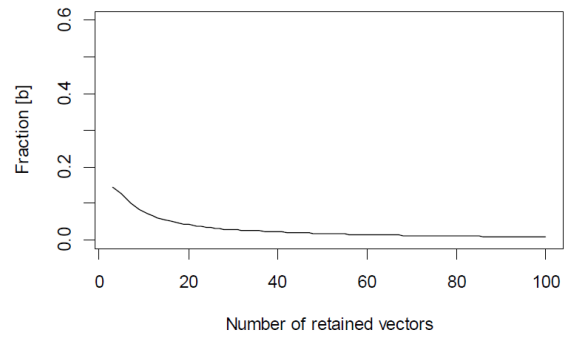
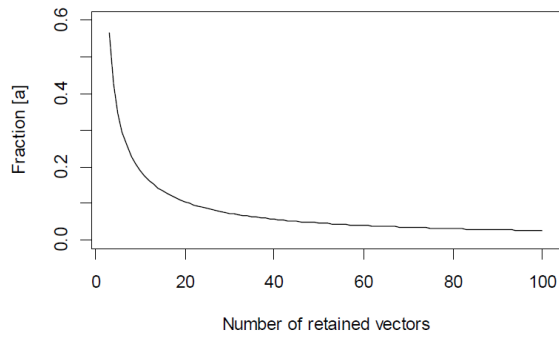


958

959 **Supp. Fig. S3** – Measures of the fit of the approximation given by equation (3) for the whole network.
 960 (a) Normalized Mutual Information index (NMI) measuring the congruence between the communities
 961 given to *Daphnia* populations (the host nodes) in the original network and the approximated network,
 962 as a function of the number of vectors retained in the approximation of equation (3). Approximated
 963 networks were obtained using a threshold on $\mathbf{L}\cdot\mathbf{R}$ values optimizing the sum of sensitivity and
 964 specificity, as described in the Materials & Methods. (b) Mean absolute error (MAE) for
 965 approximation (3) as a function of the number of vectors retained in the approximation of equation (3).

966 The MAE was obtained as $\frac{1}{np} \sum_{i,j} \left| b_{ij} - \sum_k l_{ik} r_{jk} \right|$, where \mathbf{B} is the original $n \times p$ incidence matrix, \mathbf{L} and
 967 \mathbf{R} are the components of equation (3) and the index k is only allowed to vary up to the number of
 968 vectors retained.

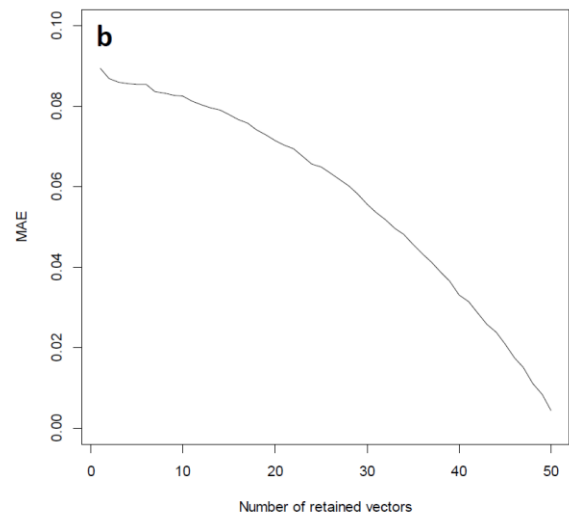
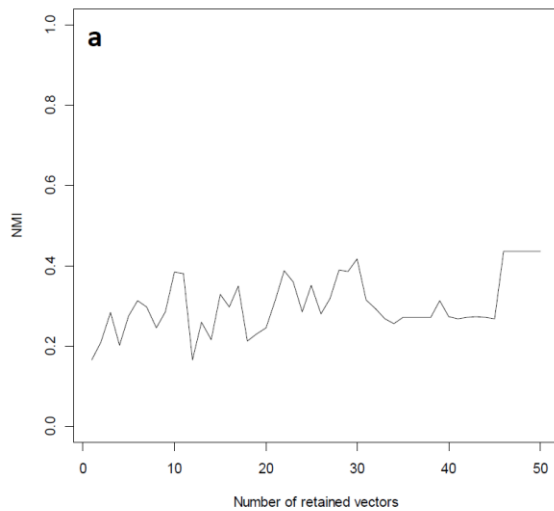
969



970

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

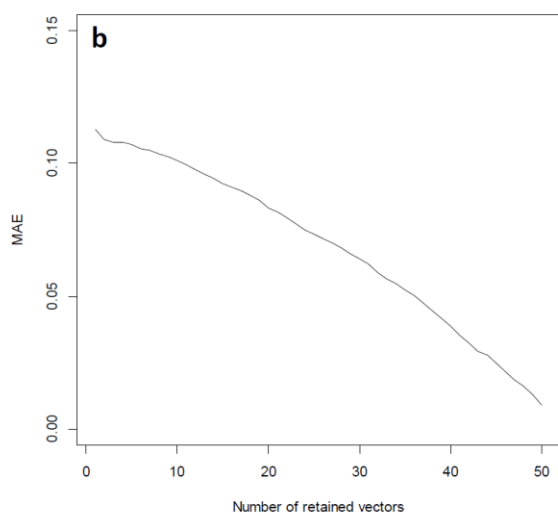
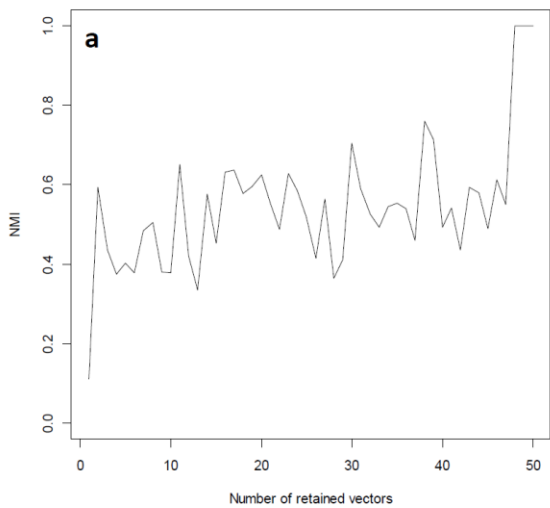
978



979

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

988



989

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

998

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

1000 **Supplementary Table S1** – Community search and congruence of communities with *Daphnia*
1001 genotypes within each of the four sub-networks defined by treatment, i.e. for each of the combination
1002 of sample type (gut microbiota vs. bacterioplankton) and each of the diet (*Scenedesmus* [S] vs.
1003 *Microcystis* [M]). Modularity: modularity score obtained for this sub-network; # modules: number of
1004 communities maximizing modularity; mod. p-value: p-value of the edge randomization test for
1005 modularity; NMI: normalized mutual information index obtained by comparing the classification by
1006 communities and the classification by genotypes; NMI p-value: p-value of the edge randomization test
1007 for NMI.

1008

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

1010 **Supplementary Table S2** – Results of the redundancy analysis / singular value decomposition (RDA-
1011 SVD, with the first five vectors retained) analysis applied to the whole network (comprising both gut
1012 microbiota and bacterioplankton samples), trying to explain network communities using sample type
1013 (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus*
1014 diet), and *Daphnia* genotype. All else as in Table 4.

1015

Fraction	Effect	df	adjusted R ²	Row perm. value	p- value	Edge perm. value	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	

1017 **Supplementary Table S3** – Results of the redundancy analysis / singular value decomposition (RDA-
1018 SVD, with the first 13 vectors retained) analysis applied to the whole network (comprising both gut
1019 microbiota and bacterioplankton samples), trying to explain network communities using sample type
1020 (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus*
1021 diet), and *Daphnia* genotype. All else as in Table 4.

1022

1023

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

1025 **Supplementary Table S4** – Results of the redundancy analysis / singular value decomposition (RDA-
1026 SVD, with the first 100 vectors retained) analysis applied to the whole network (comprising both gut
1027 microbiota and bacterioplankton samples), trying to explain network communities using sample type
1028 (gut microbiota vs. bacterioplankton), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus*
1029 diet), and *Daphnia* genotype. All else as in Table 4.

1030

1031

Fraction	Effect	df	adjusted R ²	Row perm. value	p- value	Edge perm. value	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	

1033 **Supplementary Table S5** – Results of the redundancy analysis / singular value decomposition (RDA-
1034 SVD, with the first 30 vectors retained) analysis applied to the gut microbiota-based network, trying to
1035 explain network communities using the SVD-reduced matrix for the bacterioplankton network (“bpk”,
1036 forward selection of 6 vectors), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet),
1037 and *Daphnia* genotype. All else as in Table 5.

1038

1039

1040

Fraction	Effect	df	adjusted R ²	Row perm. value	p-	Edge perm. value	p-
[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	

1042 **Supplementary Table S6** – Results of the redundancy analysis / singular value decomposition (RDA-
1043 SVD, with the first 30 vectors retained) analysis applied to the bacterioplankton-based network, trying
1044 to explain network communities using the SVD-reduced matrix for the gut microbiota network (“gut”
1045 forward selection of 1 vector), diet (*Scenedesmus* diet vs. mixed *Microcystis* and *Scenedesmus* diet),
1046 and *Daphnia* genotype. All else as in Table 6.

1047