

## Using causality and correlation analysis to decipher microbial interactions in activated sludge — [Source link](#)

[Weiwei Cai](#), [Xiangyu Han](#), [Hong Yao](#)





**Institutions:** [Beijing Jiaotong University](#)

**Published on:** 26 Sep 2021 - [bioRxiv](#) (Cold Spring Harbor Laboratory)

**Topics:** [Nitrospira](#) and [Activated sludge](#)

Related papers:

- [Changes in the Structure of the Microbial Community Associated with Nannochloropsis salina following Treatments with Antibiotics and Bioactive Compounds.](#)
- [Pathway-based and phylogenetically adjusted quantification of metabolic interaction between microbial species](#)
- [Long-term successional dynamics of microbial association networks in anaerobic digestion processes.](#)
- [Functionally relevant diversity of closely related Nitrospira in activated sludge](#)
- [Network topology reveals high connectance levels and few key microbial genera within soils](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/using-causality-and-correlation-analysis-to-decipher-xdba6vmfcg>

1           **Using causality and correlation analysis to decipher**  
2                   **microbial interactions in activated sludge**

3

4   Weiwei Cai<sup>a</sup>, Xiangyu Han<sup>a</sup>, Hong Yao<sup>a\*</sup>

5   <sup>a</sup> School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China

6

7

8

9 **ABSTRACT**

10 Network theory is widely used to understand microbial interactions in activated sludge  
11 and numerous other artificial and natural environments. However, when using  
12 correlation-based methods, it is not possible to identify the directionality of interactions  
13 within microbiota. Based on the classic Granger test of sequencing-based time-series  
14 data, a new Microbial Causal Correlation Network (MCCN) was constructed with  
15 distributed ecological interaction on the directed, associated links. As a result of  
16 applying MCCN to a time series of activated sludge data, we found that the hub species  
17 OTU56, classified as belonging the genus *Nitrospira*, was responsible for completing  
18 nitrification in activated sludge, and mainly interacted with *Proteobacteria* and  
19 *Bacteroidetes* in the form of amensal and commensal relationships, respectively.  
20 Phylogenetic tree suggested a mutualistic relationship between *Nitrospira* and  
21 denitrifiers. *Zoogloea* displayed the highest *ncf* value within the classified OTUs of the  
22 MCCN, indicating that it could be a foundation for activated sludge through forming  
23 the characteristic cell aggregate matrices into which other organisms embed during floc  
24 formation. Overall, the introduction of causality analysis greatly expands the ability of  
25 a network to shed a light on understanding the interactions between members of a  
26 microbial community.

27

28

## 29 INTRODUCTION

30 Ecological interactions, such as those involved in the exchange of resources or space,  
31 within microbial communities have been a topic of intense interest in microbial ecology,  
32 (Hibbing et al., 2010). The interactions of species are considered a driving force  
33 promoting ecological function of the microbial community, and due to its importance,  
34 the structure of communities have been described by species interaction networks for  
35 over a century (Berlow et al., 2009; Poisot et al., 2015). Although networks were  
36 initially was applied to the study of food webs, the concept has been expanded to  
37 microbial ecology to unravel ecological interactions (Ings et al., 2009; Kéfi et al., 2012).  
38 Therefore, microbial interactions within a community are more likely to be reflected by  
39 network theory, which can be established through a set of methodologies by  
40 mathematical correlation. Recently, network theory has been commonly used to explore  
41 the microbiomes of natural and artificial environments, such as soil (Barberan et al.,  
42 2012), sediments (Ji et al., 2016), bioreactors (Liang et al., 2018), and wastewater  
43 treatment plants (Global Water Microbiome Consortium et al., 2019).

44 In wastewater treatment plants, activated sludge has served as the core unit for  
45 wastewater treatment for over a century (Jenkins and Wanner, 2014). The highly  
46 diverse microorganisms in activated sludge thrive on organic compounds that are  
47 enriched in carbon(C), nitrogen (N), sulfur (S), phosphorus (P), and various trace  
48 elements, forming a complex web of ecological interactions based the competition for  
49 resources and space (Liébana et al., 2016; Xia et al., 2018). A series of graphical

50 methods have been developed for constructing correlation or co-occurrence networks,  
51 to visualize and elucidate the complex microbial interactions of species in activated  
52 sludge, gut microbiome, natural environment (Weiss et al., 2016). Previous studies on  
53 co-occurrence or correlation networks have defined multiple relationships between  
54 species with a pairwise similarity matrix or sparse multiple regression analysis  
55 respectively (Faust and Raes, 2012). Generally, nodes and links in a network,  
56 respectively, represented species and interactions, yet, these interactions were only  
57 defined by positive or negative association, which limited further understanding of  
58 ecological interactions between species. As an intrinsic property of correlation analysis,  
59 previous networks were commonly undirected, demonstrating specific interactions  
60 among species, such as competition and symbiosis. Although a few studies have  
61 attempted directed networks, provided according to the time lag, to show a direction  
62 between nodes (Deng et al., 2016; Ju and Zhang, 2015), most studies rarely explore the  
63 possibility of causality analysis from time series data, which could enhance our  
64 understanding of ecological interactions.

65 Therefore, through a combination of correlation and causality analyses this study  
66 focused on constructing a directed network to discern the sophisticated interactions  
67 between members of an activated sludge microbiome. A previously published 259 day  
68 high-through sequencing data set was employed for correlation analysis and Granger  
69 test (Jiang et al., 2018). Coupling the correlation and causality analyses allowed  
70 construction of a microbial causal correlation network (MCCN), which demonstrated

71 that the microbial interactions in activated sludge could be classified as mutualism,  
72 synergism, commensalism, neutralism, predation (parasitism), amensalism, and  
73 competition (antagonism). Hub-species OTU56 belonged to *Nitrospira* and showed  
74 more diverse interactions with *Proteobacteria* as compared to *Bacteroidetes*. Moreover,  
75 the *Zoogloea* were potentially the key genus that induced changes in many of the  
76 activated sludge bacteria due to its role in scaffold construction during sludge floc  
77 formation. The application of MCCN will provide information on the ecological  
78 interactions between different species in both natural and artificial ecosystems.

79

80

## 81 **RESULTS AND DISCUSSIONS**

### 82 *Applicability of Granger causality*

83 The assembly of the microbial community is commonly recognized as the result of  
84 deterministic and stochastic processes. The role of deterministic processes is considered  
85 to be limited in stable environments, therefore stochasticity could play an important  
86 role in gradually shifting community structure (Zhou and Ning, 2017). Due to the  
87 mutual influence of both processes, the abundance of a specific species is assumed to  
88 be the sum of a baseline and random variation. The variation of species over long  
89 periods of time should appear to be random. Within a steady state microbial community,  
90 the variation in abundance of specific species could be subjected to a joint distribution  
91 over time, as the present microbial community evolves from the previous state, while  
92 time should have limited influence on the variation of the microbial community.  
93 Although past observations are important to forecast future trends, these predictions do  
94 not completely depend on them. Therefore, there could be an autocorrelation process,  
95 which produces a time lag representing only finite past values that is applied to the  
96 forecasting. Deng et al. (2016) used the time lag to construct the correlation network  
97 with time-series data unravel microbial succession within an uranium bioremediation  
98 site (Deng et al., 2016). Additionally, David et al. (2014), when analyzing the effect of  
99 host lifestyle on human microbiota, relied on the autocorrelated process of time series  
100 (David et al., 2014), which demonstrated that OTUs variation complied with the time  
101 series model. We applied the data of 98 key OTUs obtained over the course of 259 days

102 to fit into the ADF (augmented Dickey–Fuller test) test to verify whether microbial data  
103 is irrelevant to time or not. If the data is not stationary, which defined as the time series  
104 data is independent time, the difference between adjacent values will be applied to all  
105 data. The result of the stationary check is shown in Supplementary S2. All OTUs  
106 fulfilled the requirement of stationary after difference, 51 OTUs required difference  
107 treatment while the rest were stationary without difference treatment.

### 108 ***Overall topological indexes of the causal network***

109 The visualized causal network is shown in Fig 1. 98 OTUs were used for Microbial  
110 Granger Causal Network (MGCN) construction, which created 1865 links between the  
111 nodes at a significant threshold of  $p < 0.05$ . Granger causality is commonly not  
112 symmetric, network building had to be directed. The bidirectional links were defined  
113 as a feedback from the source to the target OTU, indicating that either node could  
114 improve the forecasting accuracy of the other. A unidirectional link indicates the source  
115 OTU significantly improved the forecasting accuracy of the target OTU but not vice  
116 versa. The outdegree and indegree directed links, defined by the direction of links in or  
117 out of the specific node, were counted separately. As shown in Table 2, the distribution  
118 of nodes degree tended to be normal rather than following power-law, regardless of  
119 whether indegree or outdegree, implying that the causal network was not scale-free  
120 (Deng et al., 2012).

121 The average clustering coefficient, which reflected the clustering degree of the overall  
122 network, was defined as the average of clustering coefficient over all nodes. The

123 clustering coefficient of MGCN (0.449) was higher than previously described  
124 undirected networks, including grassland soils (0.1~0.22), lake sediment (0.09), and  
125 groundwater condition (0.17-0.29), and was comparable with the value of 0.466  
126 observed in a previous activated sludge study (Ju and Zhang, 2015). Watts and Strogatz  
127 (1998) introduced the random rewiring procedure to interpolate regular and random  
128 networks, in which the regular lattice is highly clustered while the random network is  
129 poorly clustered (Watts and Strogatz, 1998). Therefore, the higher relative clustering  
130 exhibited by the causality indicated that the network was defined rather than random.  
131 The average shortest average path was 2.149, which was smaller than within the  
132 undirected network. Hence, we derived a relatively clustered network connected by  
133 shorter paths, demonstrating that neighbouring nodes were closely connected. To  
134 confirm the small-world property, randomized networks with the same nodes and  
135 degrees as the original network were constructed. The average clustering coefficient  
136 and shortest paths were  $\sim 0.196$  and  $\sim 1.823$  respectively, whereas the ratio of Granger  
137 network to the random network of clustering coefficient and shortest path can be  
138 determined (Liao et al., 2011). As the ratio was equal to  $\sim 1.943$ , this indicated the  
139 network possessed small-world properties.

#### 140 *Indexes of nodes*

141 According to the definition of *cs*, its magnitude represents the ability of a specific OTU  
142 to cause the variation among its neighbours. A value of 1 indicates that an OTU can  
143 affect its neighbours without being affected by them, while zero indicates the opposite.

144 The  $C_{recip}$  reflected counts of reciprocating links, which exhibited feedback behaviour  
145 of each OTU, thereby higher values indicated that an OTU is likely to interact with  
146 others. Therefore, as shown in Fig. 2, as  $C_{recip}$  increases, the  $cs$  will tended to approach  
147 0.5, displaying an equilibrium of indegree and outdegree links. All nodes displayed a  
148  $C_{recip}$  value of less than 0.5, suggesting bidirectional links were not dominant in the  
149 relationship of all nodes. However, it was interesting that more interactions could be  
150 positively related to the equilibrium trend of  $cs$ . The  $cs$  and  $C_{recip}$  were both relatively  
151 quantified as the proportion excluded the magnitude of degrees, node size in Fig. 2 is  
152 proportional to degree of connection with neighbouring nodes. The average number of  
153 neighbours for a node was  $\sim 27.08$ . The majority of nodes with a large number of  
154 neighbours had higher  $C_{recip}$  and neutral position of  $cs$ . Nodes with lower  $C_{recip}$  and lower  
155  $cs$  indicated that more links were indegree, with the reverse, higher  $C_{recip}$  and  $cs$   
156 indicating more links were outdegree. Integration of the relative proportion and  
157 neighbour number that was considered as an absolute quantity was beneficial for  
158 inferring the central output nodes in the network, which should possess lower  $C_{recip}$ ,  
159 higher  $cs$ , within fairly large size of neighbours. The average of  $cs$  was  $\sim 0.491$ , showing  
160 that the number of outdegree and indegree links were nearly identical. The average of  
161  $C_{recip}$  was 0.24, implying mutual cause is not predominant due to the lower proportion  
162 in total links. Additionally,  $ncf$ , the difference between net outdegree and net indegree,  
163 of nodes ranged from -20 to 21, as shown in fig. S3. The average of individual  
164 outdegree was 8.14, the average net indegree was the same. Moreover, the number of

165 OTUs with positive *ncf* were greater than that of negative *ncf*, indicating more than 50%  
166 of the relationships in the system displayed Granger causality in the activated sludge  
167 system.

### 168 ***Bonferroni-correction***

169 The Bonferroni-corrected MGCN (BoMGCN) was produced from the original  
170 significant network (Fig. 3). The corrected network was sparser in comparison with the  
171 causal network, containing only 81 nodes and 730 links, and a lower clustering  
172 coefficient (0.373). The reduced network was highly conservative as the Bonferroni-  
173 correction excludes all potential type I error (false link was accepted) and displayed a  
174 slightly improved stability as revealed by the R square of power-law. The value of  
175 outdegree R square was 0.064, close to zero, yet the value for indegrees was 0.468,  
176 indicating a significant increase. Although the values were too small to be wholly fitted  
177 into power law, they indicated that some nodes in the BoMGCN had a greater or lesser  
178 effect on other nodes. An improvement of scale-free property was also observed, as  
179 well as an increase in the small world index, as represented by an increase in the ratio  
180 of  $\sigma$  ( $\sim 2.617$ ), caused by a decrease in clustering coefficient and increase in the average  
181 shortest path, showing BOGCN is more likely to fall in the rules of a small world.  
182 Within the random network derived from BOGCN there was a clear decrease in  
183 clustering coefficient (0.119). Additionally, the properties of total nodes were slightly  
184 distinguished from the original MGCN network as a clear decline of *cd* value. The  
185 average *cs* increased from 0.491 to 0.504. Overall, the BoMGCN reduced the size of

186 the network while keeping its basic properties. According to the classification of OTUs  
187 in BoMGCN, *Proteobacteria* was the predominant nodes, the hub species was  
188 *Nitrospira*, indicating the nitrogen-associated species has a broader social connection  
189 with other microbes.

#### 190 ***Correlation-network supplemented to causal network***

191 The MGCN showed the casual effect within the microbial community, however,  
192 information about positive or negative correlations between nodes was missing,  
193 therefore a Bonferroni-corrected microbial correlation network (BoMCN) based on  
194 Spearman's correlation (shown in Fig. S4) was applied to supplement the MGCN,  
195 constructing a Microbial Causal Correlation Network (MCCN). The multiple  
196 relationships between two OTUs could be revealed more explicitly according to this  
197 combination of causality and correlation. Previously, correlation analysis was generally  
198 used to discern the negative and positive relations within a microbial network,  
199 indicating the ecological interactions between members of the community (Faust and  
200 Raes, 2012). As shown in Figure 4a, a combination of correlation and Granger causality  
201 could construct a new relationship, which shows the directional connection among  
202 nodes including the positive or negative effect they have on each other. As shown in  
203 Figure 4b, the MCCN was composed of 73 nodes and 441 links. Although the causality  
204 is at a higher level compared with correlation, i.e., all nodes with causal links should  
205 show strong mutual interaction, the missing nodes and links could be ascribed to the  
206 Granger causality, that is not a real causal relationship, due to the limitations of the

207 method. Technically, the Granger test has been widely used for predicting the causal  
208 effect, in the context of the current study, the Granger test was utilized to forecast the  
209 relations between OTUs, which may lead to a better understanding of microbial  
210 behaviours and relationships within a community. The network can be used as an  
211 essential tool to predict microbial interaction when the real community could be too  
212 complex for accurate study, as additional efforts would be required to verify  
213 interactions between species.

214 The combination of correlation and Granger test allows observation of more specific  
215 interactions between two species, therefore a MCCN network could be applied to  
216 predict ecological relationships for community analysis. As shown in Figure 4a, there  
217 are seven patterns of species interactions, including mutualism, synergism,  
218 commensalism, neutralism, predation (parasitism), amensalism, and competition  
219 (antagonism) (Pepper et al., 2015). According to the results of MCCN, both mutualism  
220 and synergism should be a bidirectional edge with a positive effect on both species, as  
221 each species would derive benefits from the other, such that it would be difficult to  
222 distinguish them apart. Commensalism can be reflected by a unidirectional link with  
223 positive effect as species A can obtain a metabolite produced by species B. Although  
224 species B would be irrelevant to the growth of species A, as there is no feedback from  
225 A to B, the sequencing data of two species would be positively correlated as more  
226 species B would secrete more metabolites for species A. Oppositely, a unidirectional  
227 connection with negative effect is classified as amensalism due to the general release

228 of inhibitors from species A to species B. Here the quantity of species A will be relevant  
229 to the production of inhibitors, such as antibiotics, which can reduce the number of  
230 species B, thereby the contrary growth of species A and B will lead to a negative  
231 correlation. Although the predation (parasitism) can be implied by the negative  
232 bidirectional edge, the sequencing data used in this study contained only information  
233 from the 16S rRNA gene of bacteria, with no information about protozoa or phages,  
234 resulting in the exclusion of predation (parasitism) from the MCCN of the microbial  
235 community (Deng et al., 2016). Finally, a negative bidirectional link could also indicate  
236 competition between species. The MCCN is a powerful tool to recognize multiple  
237 interactions of microbes by specifying the endogeneity of correlation, which has been  
238 widely used as a statistic proof of microbial interaction within a network (Weiss et al.,  
239 2016).

#### 240 *Core species in MCCN*

241 The nodes with amounts of links would be considered as “hubs” in the MCCN. OTU56  
242 was the hub species with the greatest number of indegrees (31) and second highest  
243 number of outdegrees (16). It was classified as belonging to the genus *Nitrospira*, a  
244 globally distributed group of nitrite oxidizers, which are capable of completing  
245 nitrification from ammonia to nitrate by one step (van Kessel et al., 2015). As shown  
246 in Figures 5, S5, and S6, OTU56 closely interacted with 24 OTUs from the phylum  
247 *Proteobacteria*, 8 OTUs from the phylum Bacteroidetes, and the 6 remaining OTUs  
248 interacted with 5 additional phyla. 21 OTUs displayed negative interactions with  
249 *Nitrospira*, 14 should be amensalism and 7 were competition relationship. Interestingly,

250 all competition interactions originated from *Proteobacteria* to *Nitrospira*, showing a  
251 number of *Proteobacteria* may depress the growth of *Nitrospira*. This could be ascribed  
252 to the fact that most bacteria related to nitrogen cycle were *Proteobacteria* (Costa et al.,  
253 2006). Additionally, OTU56 unidirectionally interacted with OTUs from *Bacteroidetes*,  
254 for which there were only two types of interactions, commensalism and amensalism,  
255 with 3 and 5 links, respectively. According to a global diversity and biogeography study  
256 of over 300 wastewater treatment plants, only 28 out of 61448 OTUs, accounting for  
257 12.4% of the 16S rRNA gene sequences, were defined as core OTUs, and these mainly  
258 consisted of *Proteobacteria*, *Bacteroidetes*, and *Nitrospira* in activated sludge (Global  
259 Water Microbiome Consortium et al., 2019). Therefore, the results of MCCN in this  
260 study are consistent, as *Proteobacteria* and *Bacteroidetes* actively interacted with the  
261 core species of *Nitrospira*, a group which plays a crucial role of nitrification in activated  
262 sludge. At the genus level, the majority of species that interacted with OTU56 were  
263 unclassified, and of those that could be identified, *Azospira*, which possesses  
264 denitrification activity, exhibited a mutualistic relationship with *Nitrospira*, as well as  
265 with OTU176167 and OTU92689, which were most closely related to the genus  
266 *Dechloromonas*, members of which are capable of reducing nitrate or chloride. The  
267 above mutualistic relationships could be achieved in nitrogen cycling processes, with  
268 denitrification removing nitrate as a product inhibitor to *Nitrospira*, meanwhile,  
269 *Nitrospira* could supply nitrate as a substrate for denitrifies.

270 OTU180929, which had the most outdegree links (17), was classified as  
271 *Sinobacteraceae* at the family level. Members of this family are known to play a role  
272 in the degradation of aliphatic, aromatic hydrocarbon compounds and small organic  
273 acids (Gutierrez et al., 2013; Zhang et al., 2018). The number of net outdegrees and net  
274 indegree indicated the trending of nodes to cause a change of others or be affected by  
275 others. OTU180929, belonging to the genus *Zoogloea*, possessed 13 net outdegree and  
276 13 net indegree links separately. *Zoogloea* has previously been demonstrated to be a  
277 bacterial genus important in the process of floc formation (Shao et al., 2009), and in  
278 this study is represented by OTU180929 and OTU178488. As shown in Figure S7, the  
279 *ncf* of *Zoogloea* was the highest value within the sum of classified OTUs, indicating  
280 that *Zoogloea* could enhance the growth of most species, i.e., it could be the foundation  
281 for the formation of activated sludge. However, the *ncf* of unclassified OTUs was still  
282 higher, reaching 18. The culture-dependent methods build the basics of microbiology  
283 research, which investigate the role of specific species (mostly are filamentous) in  
284 sludge flocculation and foaming (Nielsen et al., 2009). The unclassified nodes in  
285 MCCN showed there is still a massive microbial dark matter in activated sludge wait  
286 to be cultured. The network approach has been used to elucidate and prioritize the  
287 microbial dark matter in microbial community (Zamkovaya et al., 2021). Although  
288 activated sludge has been a widely employed strategy in wastewater treatment plants  
289 for over 100 years (Nielsen and McMahon, 2014), its microbiome still contains many

290 mysteries, and is abundant with unknown species that are only gradually being  
291 elucidated by recent progress in culture-dependent and independent technologies.

292 In conclusion, the coupling of correlation and causality was crucial to understand  
293 ecological interactions within the microbial community. The Microbial Causal  
294 Correlation Network (MCCN) showed a sophisticated causal network in activated  
295 sludge and identified the fundamental species, with highest *ncf* value, as *Zoogloea*. The  
296 Microbial Causal Correlation Network (MCCN) and phylogenetic analysis together  
297 pointed out the core-species of *Nitrospira* (OTU56) could have mutualistic interactions  
298 with denitrifiers in activated sludge. However, most species that interacted with OTU56  
299 were still unclassified, implying a greater sequencing depth would be the key to  
300 improve the understanding of activated sludge.

301

## 302 **MATERIALS AND METHODS**

### 303 ***Sequencing data derivation***

304 The sequencing data were acquired from NCBI (accession number: PRJNA324303),  
305 which has been published previously (Jiang et al., 2018). The time-series data set  
306 included sequencing data for 259 days taken from a long-term operational wastewater  
307 treatment plant. The primers were F515 and R806, which covered mostly bacteria and  
308 archaea. The achieved fastq files were combined and processed online using a galaxy  
309 platform (Feng et al., 2017). OTUs were created with 97% cut-off through Uparse  
310 clustering method. RDP classifier assigned one representative sequence from each

311 OTU to bacteria or archaeal taxonomy according to the 16S rRNA Greengene Database.  
312 The final OTUs table was prepared for the subsequent process. The phylogenetic tree  
313 was created with Mega software with N-J method, the visualization was completed  
314 online (<https://itol.embl.de>) (Letunic and Bork, 2019).

### 315 ***Stationarity***

316 Stationarity is an important concept to time series analysis and is a precondition to  
317 Granger Causality. The properties of stationarity were defined by the three main factors  
318 in terms of mean, variance, and covariance. The stationarity indicates that there was no  
319 change of trend in the time data, and it is known as a changeless process of the joint  
320 distribution within a specific displacement. The stationary implies that the expectation  
321 value of OTUs will fluctuate around the mean value of their neighbourhood rather than  
322 depend on time. This allows an estimation of the significant interval for the variation.  
323 Therefore, the stationarity analysis should be performed before analyzing time series  
324 data. It can be tested by detecting the presence or absence of unit root. The ADF-test  
325 was employed to verify if the time series data conformed to the stationary. If the original  
326 data is not subjected to stationarity, we used the difference, one minus another one to  
327 calculate the difference, to obtain the stationarity data. All abundance data of OTUs  
328 were filtered with the stationary test, while data that failed to go through ADF test after  
329 two rounds of using difference would be summed in the separate file as nonstationary  
330 data. Although the abundances of OTUs may vary on a large scale, even seemingly  
331 without a mean value, the difference would be stationary in most situations. The

332 operating reactors could be affected by many factors, which would shift the microbial  
333 community via stimulation the metabolism of specific species.

### 334 *Granger Causality*

335 The Granger causality test is a statistical hypothesis test that determined the role of one  
336 time series in forecasting another one (Granger, 1969). Herein, the Granger causality is  
337 limited within interpreting the interaction of two OTUs which were subjected to the  
338 ARMA (Autoregressive–moving-average model) model. To  $i$ th OTU, the ARMA  
339 model is shown as the equation:

$$340 \quad z_{ti} = \delta + u_{ti} + \theta_1 z_{(t-1)i} + \theta_2 z_{(t-2)i} - \dots - \theta_q z_{(t-q)i} + -\varphi_1 u_{ti} - \varphi_2 u_{(t-1)i} - \dots$$
$$341 \quad - \varphi_q u_{(t-q)i}$$

342 We simplified the equation for OTUs to the following format.

$$343 \quad z_{ti} = \delta + \sum_{l=1}^q \theta_l z_{(t-l)i} - \sum_{l=1}^q \varphi_l u_{(t-l)i} = \delta + \sum_{l=1}^q \theta_l z_{(t-l)i} + v_i$$

344  $v_i$  is the random variation (white-noise series). Thus, we assumed the model for  $i$ th  
345 OTU is X, the model for  $j$ th OTU is Y. Both equations are as follows:

$$346 \quad X_t = \delta + \sum_{l=1}^q \theta_l z_{(t-l)i} + v_i$$
$$347 \quad Y_t = \delta + \sum_{l=1}^q \theta_l z_{(t-l)j} + v_i$$

348 To know the interaction of X and Y, we assumed X and Y are interplays in their  
349 respective model predictions. The new models are derived as:

$$350 \quad X = \delta + \sum_{l=1}^q \theta_l X_{(t-l)i} + \sum_{l=1}^q \theta_l Y_{(t-l)j} + v_{ij}$$

351 
$$Y = \delta + \sum_{l=1}^q \theta_l X_{(t-l)i} + \sum_{l=1}^q \theta_l Y_{(t-l)j} + v_{ji}$$

352 If the time series data followed the above equation, meaning the past values of X will  
353 contribute to predict current Y, and vice versa. However, real data could be applied to  
354 the following equation:

355 
$$X_t + b_j Y_t = \delta + \sum_{l=1}^q \theta_l X_{(t-l)i} + \sum_{l=1}^q \theta_l Y_{(t-l)j} + v_{ij}$$

356 
$$Y_t + c_i X_t = \delta + \sum_{l=1}^q \theta_l X_{(t-l)i} + \sum_{l=1}^q \theta_l Y_{(t-l)j} + v_{ji}$$

357 If  $b_j$  and  $c_i$  are not equal to 0 at the same time, this will be a model with  
358 instantaneous causality. In other words, the  $v_{ij}$  and  $v_{ji}$  would be the key to determine  
359 the Granger causality, if the variation could be decreased when applying the  $b_j \neq 0$ ,  
360 representing the  $j$ th OTU can contribute to the prediction of  $i$ th OTU, otherwise, there  
361 was no improvement of predicting  $i$ th OTU with  $j$ th OTU information. Therefore, the  
362 Granger causality can be tested by the ANOVA analysis to obtain a p-value. This  
363 relation between X and Y was termed as Granger causality by which implied X or Y  
364 can cause each other. Herein, the causal effects were attributed to the property of edges  
365 in the network, while OTUs would be the nodes.

### 366 ***Network construction***

367 All OTUs were filtered with two specific conditions such that OTUs with more than  
368 80% non-zero values would be preserved, and the residuals should comply with that of  
369 at least one abundance of individual OTU reached more than 0.01% in all samples. The  
370 total number of OTUs was 98. The ADF test was applied to verify the stationarity of

371 time series data and provide a proper lag for the next modelling process. The difference  
372 is calculated once OTUs fail in the ADF test, the results of difference will track the  
373 ADF test again. All OTUs were reserved by twice difference treatment. The time series  
374 matrix successfully inspected by the ADF test was used for the Granger test in pair.  
375 Before the operation of the Granger test, the order was been determined by VAR (R  
376 package) (Pfaff, 2008). Subsequently, the lag was transferred to the Granger test. The  
377 p-value threshold of Granger test was restricted by the following two methods. Due to  
378 the massively paired results, the links confirmed by the significance value could still  
379 cause statistical type I error, hence we introduced Bonferroni multiple-comparisons  
380 procedure and false discovery rate (FDR) to correct the threshold. Bonferroni multiple-  
381 comparisons procedure was determined by the following equation.

$$382 \quad \alpha^* = \frac{0.05}{\binom{k}{2}}$$

383 In the FDR test, all links that were selected by 0.05 significant cut-off are reordered  
384 according to the magnitude of the p-value. FDR values were calculated by the following  
385 equation.

$$386 \quad q_i = k \frac{p_i}{i}$$

$$387 \quad FDR_i = \min(q_i, \dots, q_k)$$

388 Where,  $i$  is the rank of the p-value in  $k$  links, which is the total links preserved by  
389 the previous threshold. The critical FDR value is normally 0.05. FDR has a great power  
390 to detect genuine positive effects, while the Bonferroni adjustment is more conservative  
391 and considers all comparisons to be statistically independent. The final file was

392 transferred to Cytoscape software for further visualization and analysis. All analysis  
393 processes were completed with R, and several shiny apps had been built for this study  
394 (Stationary check: <https://caiweiwei.shinyapps.io/stationarycheck/>. Granger Causality  
395 network website: <https://caiweiwei.shinyapps.io/causalnetwork/>; Correlation network:  
396 <https://caiweiwei.shinyapps.io/Cornetwork/>; MCCN:  
397 <https://caiweiwei.shinyapps.io/combinenetwork/>). The specific instruction for each app  
398 is provided in the supplementary (S1).

### 399 *Network indexes*

400 The several properties of the causal network were referenced from the literature of Anil  
401 Seth (Seth, 2005), and termed as the Causal score ( $cs$ ), Causal density ( $cd$ ), Net Causal  
402 flow ( $ncf$ ), and Causal reciprocity ( $c_{recip}$ ). Table 1 shows equations for all corresponding  
403 properties.

404 As the network had been directed, outdegree and indegree represented the direction  
405 of edges within two nodes. A causal score ( $cs$ ) was determined by the ratio of outdegree  
406 to total degrees of a specific node, reflecting the OTU influenced other OTUs rather  
407 than being influences. The causal score is defined as  $cs = \frac{\text{number of outdegrees}}{\text{number of indegrees}}$  in unweighted graphs (graphs in which all links are  
408 equivalent). If  $cs > 1$ , the corresponding OTU has active output, otherwise it is being  
409 passively influenced. The causal density is also termed as causal efficiency of the  
410 network, which, to some extent, represents the connectivity of the network. The net  
411 causal flow is the difference between outdegree and indegree of each node, indicating  
412

413 the contribution of the individual node would be either active or passive. Herein, the  
414 active state represents the species intentionally affect others, while the passive indicates  
415 it is affected by others. Although causal flow is like causal score, the former is intended  
416 to be independent of the quantity of balanced efferent and afferent connections. The  
417 causal reciprocity is the fraction of links with a directly reciprocal edge. Overall, the  
418 causal score and flow are applied to evaluate the role of each node, while the rest  
419 describes the whole network. Additionally, the supplemented indexes, including  
420 connectivity, centrality, stress centrality etc., were analyzed with the Cytoscape  
421 software tool (Feng et al., 2017).

#### 422 **Data availability**

423 All raw data were acquired from NCBI (accession number: PRJNA324303) (Jiang et  
424 al., 2018).

#### 425 **ACKNOWLEDGEMENTS**

426 This study is supported by Beijing Outstanding Young Scientist Program  
427 (BJJWZYJH01201910004016) and National Natural Science Foundation of China (NSFC,  
428 No. 51908030).

#### 429 **Competing Financial Interests**

430 The authors declare no competing financial interests.

#### 431 **Supporting Information**

432 Figure S1. Description of MGCN apps family.; Figure S2. Stationary check results for 98  
433 OTUs.; FigureS3. *ncf* values of all OTUs in MGCN.; Figure S4. Heatmap of MCN and the

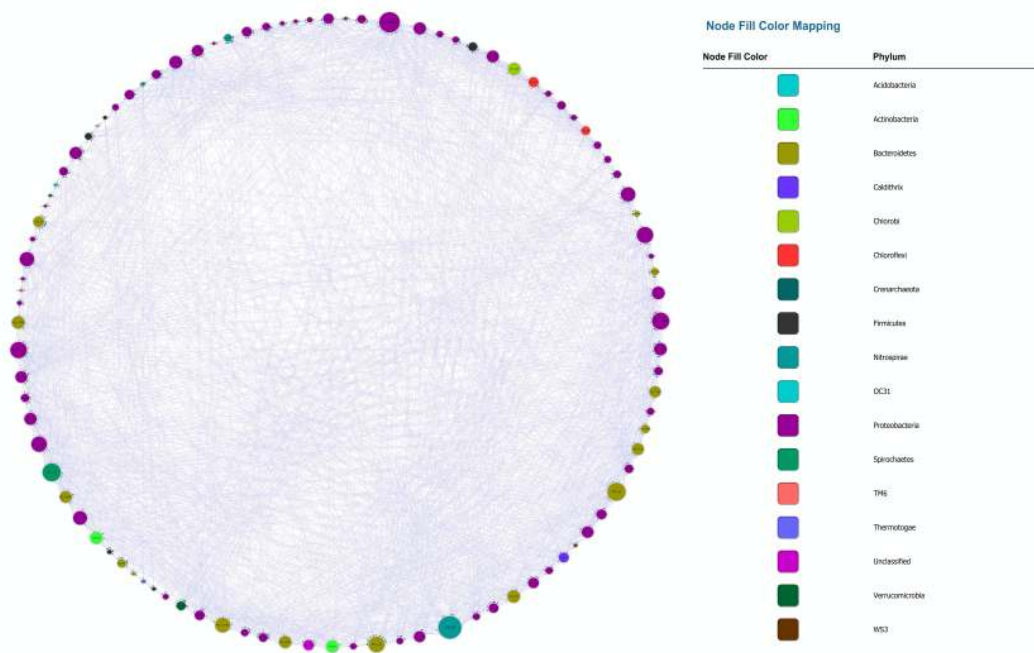
434 network of BoMCCN.; Table S1. Table of the summary of species interaction around the hub  
435 OTU56 (S5), Figure S6. MCCN of core-species; Figure S7 Node index of the specific genera  
436 from MCCN.

437 **REFERENCE**

- 438 Barberan, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore  
439 co-occurrence patterns in soil microbial communities. *ISME J* 6, 343–51.  
440 <https://doi.org/10.1038/ismej.2011.119>
- 441 Berlow, E.L., Dunne, J.A., Martinez, N.D., Stark, P.B., Williams, R.J., Brose, U., 2009. Simple  
442 prediction of interaction strengths in complex food webs. *Proceedings of the National*  
443 *Academy of Sciences* 106, 187–191. <https://doi.org/10.1073/pnas.0806823106>
- 444 Costa, E., Pérez, J., Kreft, J.-U., 2006. Why is metabolic labour divided in nitrification? *Trends*  
445 *in Microbiology* 14, 213–219. <https://doi.org/10.1016/j.tim.2006.03.006>
- 446 David, L.A. *et al.* (2014) Host lifestyle affects human microbiota on daily timescales. *Genome*  
447 *Biol.*, **15**, R89.
- 448 Deng, Y., Jiang, Y.H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological network  
449 analyses. *BMC Bioinformatics* 13, 113. <https://doi.org/10.1186/1471-2105-13-113>
- 450 Deng, Y., Zhang, P., Qin, Y., Tu, Q., Yang, Y., He, Z., Schadt, C.W., Zhou, J., 2016. Network  
451 succession reveals the importance of competition in response to emulsified vegetable  
452 oil amendment for uranium bioremediation: Competition in bioremediation system.  
453 *Environmental Microbiology* 18, 205–218. <https://doi.org/10.1111/1462-2920.12981>
- 454 Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nat Rev Microbiol*  
455 10, 538–50. <https://doi.org/10.1038/nrmicro2832>
- 456 Feng, K., Zhang, Z., Cai, W., Liu, W., Xu, M., Yin, H., Wang, A., He, Z., Deng, Y., 2017.  
457 Biodiversity and species competition regulate the resilience of microbial biofilm  
458 community. *Mol Ecol* 26, 6170–6182. <https://doi.org/10.1111/mec.14356>
- 459 Global Water Microbiome Consortium, Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X.,  
460 Zhang, Q., Brown, M.R., Li, Z., Van Nostrand, J.D., Ling, F., Xiao, N., Zhang, Y.,  
461 Vierheilig, J., Wells, G.F., Yang, Y., Deng, Y., Tu, Q., Wang, A., Zhang, T., He, Z., Keller,  
462 J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J.M., He, Q., Curtis,  
463 T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., 2019. Global  
464 diversity and biogeography of bacterial communities in wastewater treatment plants.  
465 *Nat Microbiol* 4, 1183–1195. <https://doi.org/10.1038/s41564-019-0426-5>
- 466 Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-  
467 spectral Methods. *Econometrica* 37, 424–438. <https://doi.org/10.2307/1912791>
- 468 Gutierrez, T., Green, D.H., Nichols, P.D., Whitman, W.B., Semple, K.T., Aitken, M.D., 2013.  
469 Polycyclovorans *algicola* gen. nov., sp. nov., an Aromatic-Hydrocarbon-Degrading  
470 Marine Bacterium Found Associated with Laboratory Cultures of Marine  
471 Phytoplankton. *Appl. Environ. Microbiol.* 79, 205–214.  
472 <https://doi.org/10.1128/AEM.02833-12>
- 473 Hibbing, M.E., Fuqua, C., Parsek, M.R., Peterson, S.B., 2010. Bacterial competition: surviving  
474 and thriving in the microbial jungle. *Nature Reviews Microbiology* 8, 15–25.  
475 <https://doi.org/10.1038/nrmicro2259>
- 476 Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards,  
477 F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., Lewis, H.M.,  
478 Olesen, J.M., Veen, F.J.F.V., Warren, P.H., Woodward, G., 2009. Review: Ecological

- 479 networks – beyond food webs. *Journal of Animal Ecology* 78, 253–269.  
480 <https://doi.org/10.1111/j.1365-2656.2008.01460.x>
- 481 Jenkins, D., Wanner, J., 2014. *Activated Sludge - 100 Years and Counting*. IWA Publishing.
- 482 Ji, Y., Angel, R., Klose, M., Claus, P., Marotta, H., Pinho, L., Enrich-Prast, A., Conrad, R., 2016.  
483 Structure and function of methanogenic microbial communities in sediments of  
484 Amazonian lakes with different water types. *Environmental Microbiology* 18, 5082–  
485 5100. <https://doi.org/10.1111/1462-2920.13491>
- 486 Jiang, X.T., Ye, L., Ju, F., Wang, Y.L., Zhang, T., 2018. Toward an Intensive Longitudinal  
487 Understanding of Activated Sludge Bacterial Assembly and Dynamics. *Environ Sci*  
488 *Technol* 52, 8224–8232. <https://doi.org/10.1021/acs.est.7b05579>
- 489 Ju, F., Zhang, T., 2015. Bacterial assembly and temporal dynamics in activated sludge of a full-  
490 scale municipal wastewater treatment plant. *The ISME Journal* 9, 683–695.  
491 <https://doi.org/10.1038/ismej.2014.162>
- 492 Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., Boit, A.,  
493 Joppa, L.N., Lafferty, K.D., Williams, R.J., Martinez, N.D., Menge, B.A., Blanchette,  
494 C.A., Iles, A.C., Brose, U., 2012. More than a meal... integrating non-feeding  
495 interactions into food webs. *Ecology Letters* 15, 291–300.  
496 <https://doi.org/10.1111/j.1461-0248.2011.01732.x>
- 497 Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new  
498 developments. *Nucleic Acids Res* 47, W256–W259.  
499 <https://doi.org/10.1093/nar/gkz239>
- 500 Liang, B., Ma, J., Cai, W., Li, Z., Liu, W., Qi, M., Zhao, Y., Ma, X., Deng, Y., Wang, A., Zhou,  
501 J., 2018. Response of chloramphenicol-reducing biocathode resistome to continuous  
502 electrical stimulation. *Water Res* 148, 398–406.  
503 <https://doi.org/10.1016/j.watres.2018.10.073>
- 504 Liao, W., Ding, J., Marinazzo, D., Xu, Q., Wang, Z., Yuan, C., Zhang, Z., Lu, G., Chen, H.,  
505 2011. Small-world directed networks in the human brain: Multivariate Granger  
506 causality analysis of resting-state fMRI. *NeuroImage* 54, 2683–2694.  
507 <https://doi.org/10.1016/j.neuroimage.2010.11.007>
- 508 Liébana, R., Arregui, L., Santos, A., Murciano, A., Marquina, D., Serrano, S., 2016. Unravelling  
509 the interactions among microbial populations found in activated sludge during biofilm  
510 formation. *FEMS Microbiology Ecology* 92, fiw134.  
511 <https://doi.org/10.1093/femsec/fiw134>
- 512 Nielsen, P.H., Kragelund, C., Seviour, R.J., Nielsen, J.L., 2009. Identity and ecophysiology of  
513 filamentous bacteria in activated sludge. *FEMS Microbiology Reviews* 33, 969–998.  
514 <https://doi.org/10.1111/j.1574-6976.2009.00186.x>
- 515 Nielsen, P.H., McMahon, K.D., 2014. Microbiology and microbial ecology of the activated  
516 sludge process. *Activated sludge–100 years and counting* 53–75.
- 517 Pepper, I.L., Gerba, C.P., Gentry, T.J. (Eds.), 2015. *Environmental microbiology*, Third edition.  
518 ed. Elsevier/AP, Academic Press is an imprint of Elsevier, Amsterdam.
- 519 Pfaff, B., 2008. VAR, SVAR and SVEC Models: Implementation Within R Package **vars**. *J.*  
520 *Stat. Soft.* 27. <https://doi.org/10.18637/jss.v027.i04>

- 521 Poisot, T., Stouffer, D.B., Gravel, D., 2015. Beyond species: why ecological interaction  
522 networks vary through space and time. *Oikos* 124, 243–251.  
523 <https://doi.org/10.1111/oik.01719>
- 524 Seth, A.K., 2005. Causal connectivity of evolved neural networks during behavior. *Network:  
525 Computation in Neural Systems* 16, 35–54.  
526 <https://doi.org/10.1080/09548980500238756>
- 527 Shao, Y., Chung, B.S., Lee, S.S., Park, W., Lee, S.-S., Jeon, C.O., 2009. *Zoogloea caeni* sp. nov.,  
528 a floc-forming bacterium isolated from activated sludge. *International Journal of  
529 Systematic and Evolutionary Microbiology*, 59, 526–530.  
530 <https://doi.org/10.1099/ijs.0.65670-0>
- 531 van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op den Camp, H.J.M., Kartal,  
532 B., Jetten, M.S.M., Lücker, S., 2015. Complete nitrification by a single microorganism.  
533 *Nature* 528, 555–559. <https://doi.org/10.1038/nature16459>
- 534 Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of ‘small-world’ networks. *Nature* 393,  
535 440. <https://doi.org/10.1038/30918>
- 536 Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z.,  
537 Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F.,  
538 Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary  
539 widely in sensitivity and precision. *The ISME Journal* 10, 1669–1681.  
540 <https://doi.org/10.1038/ismej.2015.235>
- 541 Xia, Y., Wen, X., Zhang, B., Yang, Y., 2018. Diversity and assembly patterns of activated sludge  
542 microbial communities: A review. *Biotechnology Advances* 36, 1038–1047.  
543 <https://doi.org/10.1016/j.biotechadv.2018.03.005>
- 544 Zamkovaya, T., Foster, J.S., de Crécy-Lagard, V., Conesa, A., 2021. A network approach to  
545 elucidate and prioritize microbial dark matter in microbial communities. *The ISME  
546 Journal* 15, 228–244. <https://doi.org/10.1038/s41396-020-00777-x>
- 547 Zhang, B., Xu, X., Zhu, L., 2018. Activated sludge bacterial communities of typical wastewater  
548 treatment plants: distinct genera identification and metabolic potential differential  
549 analysis. *AMB Expr* 8, 184. <https://doi.org/10.1186/s13568-018-0714-0>
- 550 Zhou, J.Z., Ning, D.L., 2017. Stochastic Community Assembly: Does It Matter in Microbial  
551 Ecology? *Microbiol Mol Biol R* 81. <https://doi.org/10.1128/MMBR.00002-17>  
552  
553



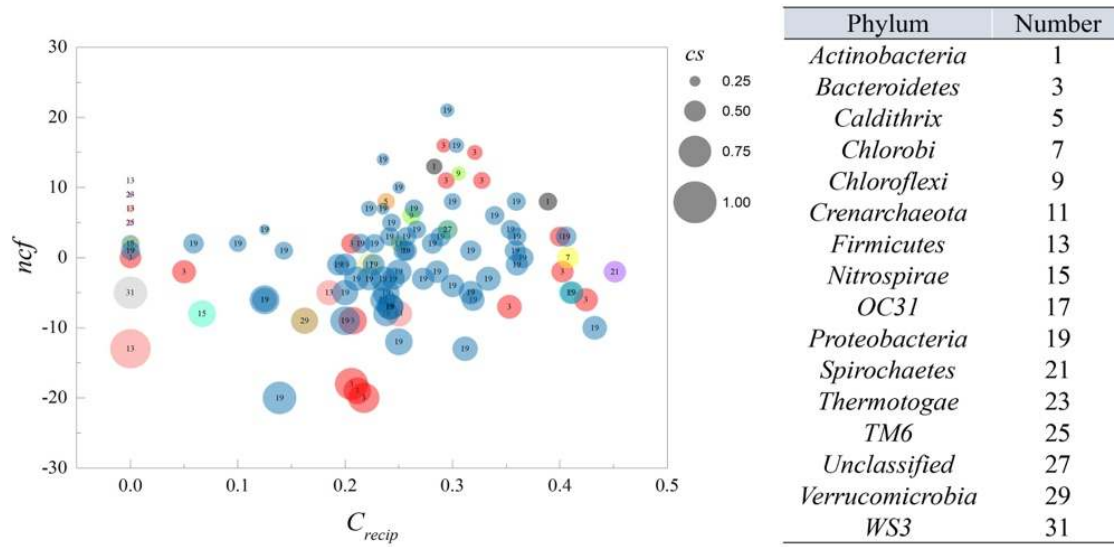
554

555 Figure 1. MGCN, each colour represents a separate phylum. The size of the node and

556 node label is proportionate to the edge number of each node from 0 to 110. The arrows

557 represent the direction of Granger causality.

558

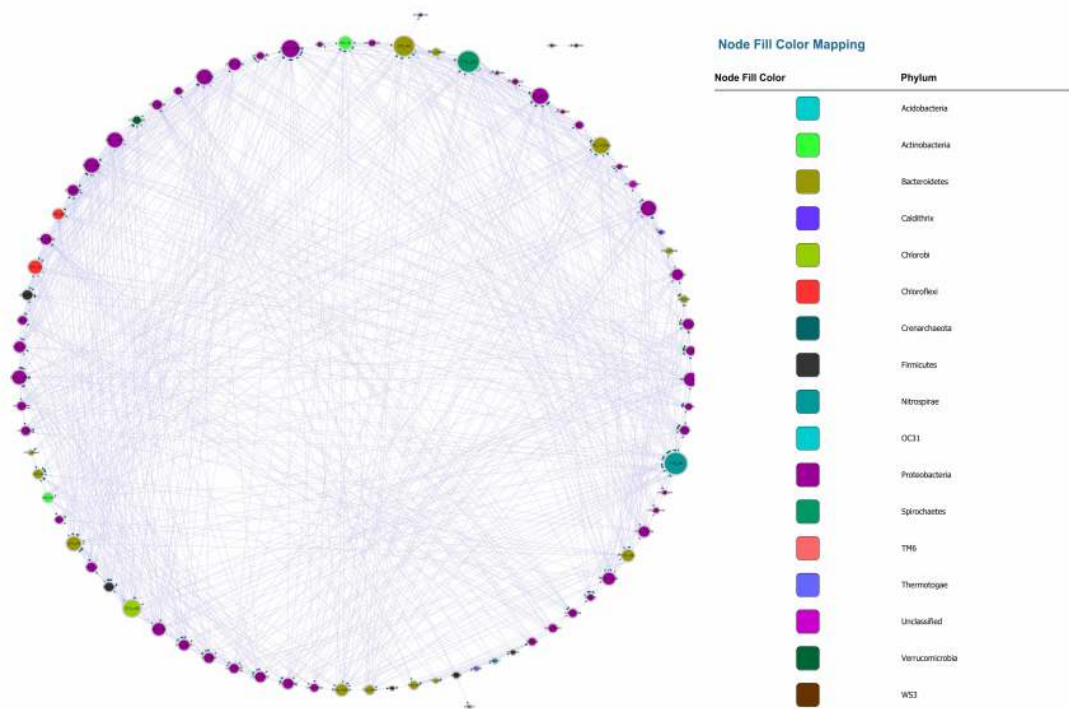


559

560

561 Figure 2.  $ncf$  and  $C_{recip}$  from MGCN. Each circle represents an individual node from  
 562 MGCN with size representing the  $cs$  value. The number within the circle corresponds  
 563 to the classification of OTU at the phylum level.

564



565

566 Figure 3. BoMGCN. Each colour represents a separate phylum. The size of the node  
567 and node label are proportionate to the edge number of each node from 0 to 110. The  
568 arrows represent the direction of Granger causality.

569





588 Table 1 Network indexes

Name	Equation	Description
Causal score ( <i>cs</i> )	$cs = \frac{n_o}{n_o + n_i}$	$n_o$ is the outdegree, $n_i$ is the indegree.
Causal density ( <i>cd</i> )	$cd = \frac{n}{2N(N-1)}$	$n$ is the total number of significant
Graph efficiency ( <i>ge</i> )	$ge = 1 -$	causal links preserved in the network
	$\frac{n - \binom{N-1}{2}}{\binom{N}{2}}$	file. $N$ is the size of the network.
Net causal flow ( <i>ncf</i> )	$ncf$	$n_{no}$ is the net outdegree of a specific
	$= n_{no} - n_{ni}$	node. $n_i$ is the net indegree of the
	$n_{no} = N_v - n_i$	individual node. $N_v$ is the sum of all
	$n_{ni} = N_v - n_o$	neighbours of specific OTU.
Causal reciprocity ( <i>C<sub>recip</sub></i> )	$C_{recip} = \frac{n_{recip}}{n}$	$n_{recip}$ represents the number of reciprocal links in the total network. $n$ is the value of total links.

589

590

591

Table 2 Properties of different networks

	Granger		Random		Spearman		Random		MCCN	Random MCCN
	MGCN	BoMGCN	MGCN	BoMGCN	MCN	BoMCN	MCN	BoMCN		
network size	98	81	98	81	98	98	98	97	73	73
network density					0.811	0.568	0.811	0.568		
links	1856	730	1856	730	3856	2644	3856	2644	441	441
power law (in)	0.041	0.468	0	0.284	0.211	0.101 (in total)	0.292	0.02	0.552	0.016
power law (out)	0.03	0.064	0.043	0					0.595	0.111
average clustering coefficient	0.449	0.373	0.196	0.119	0.868	0.753	0.812	0.57	0.352	0.084
network diameter	6	8	3	4	2	4	2	2	7	5
network radius	3	1	2	3	2	2	2	2	1	3
average shortest paths	2.149	2.647	1.823	2.21	1.189	1.463	1.189	1.432	2.866	2.571
average number of neighbors	27.082	14.716	34.143	17.21	78.694	54.515	78.694	54.515	9.753	11.507
<i>cd</i>	0.098	0.056							0.042	
<i>ge</i>	0.815	0.9							0.93	
average <i>cs</i>	0.491	0.504	NA		NA		NA		0.509	NA
average $C_{recip}$	0.238	0.148							0.157	

NA represents there is no data for this property.