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

Weiwei Cai[a], Xiangyu Han[a], Hong Yao[a*]

[a] School of Civil Engineering, Beijing Jiaotong University, Beijing 100044, China

ABSTRACT

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

INTRODUCTION

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

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

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

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

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

RESULTS AND DISCUSSIONS

Applicability of Granger causality

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

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

Overall topological indexes of the causal network

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

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

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

Indexes of nodes

According to the definition of *cs*, its magnitude represents the ability of a specific OTU to cause the variation among its neighbours. A value of 1 indicates that an OTU can 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 ~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 ~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

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

Bonferroni-correction

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

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

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

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

Core species in MCCN

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

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

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

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

Stationarity

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

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

Granger Causality

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

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

We simplified the equation for OTUs to the following format.

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

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

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

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

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

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

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

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

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

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

$$q_i = \mathrm{k}\frac{p_i}{i}$$

$$FDR_i = \min(q_i, \ldots, q_k)$$

Where, i is the rank of the p-value in k links, which is the total links preserved by the previous threshold. The critical FDR value is normally 0.05. FDR has a great power to detect genuine positive effects, while the Bonferroni adjustment is more conservative 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* = the number of outdegrees
408 divided by the number of indegrees in unweighted graphs (graphs in which all links are
409 equivalent). If $cs > 1$, the corresponding OTU has active output, otherwise it is being
410 passively influenced. The causal density is also termed as causal efficiency of the
411 network, which, to some extent, represents the connectivity of the network. The net
412 causal flow is the difference between outdegree and indegree of each node, indicating

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.

REFERENCE

Barberan, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J 6, 343–51. https://doi.org/10.1038/ismej.2011.119

Berlow, E.L., Dunne, J.A., Martinez, N.D., Stark, P.B., Williams, R.J., Brose, U., 2009. Simple prediction of interaction strengths in complex food webs. Proceedings of the National Academy of Sciences 106, 187–191. https://doi.org/10.1073/pnas.0806823106

Costa, E., Pérez, J., Kreft, J.-U., 2006. Why is metabolic labour divided in nitrification? Trends in Microbiology 14, 213–219. https://doi.org/10.1016/j.tim.2006.03.006

David,L.A. *et al.* (2014) Host lifestyle affects human microbiota on daily timescales. *Genome Biol.*, **15**, R89.

Deng, Y., Jiang, Y.H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological network analyses. BMC Bioinformatics 13, 113. https://doi.org/10.1186/1471-2105-13-113

Deng, Y., Zhang, P., Qin, Y., Tu, Q., Yang, Y., He, Z., Schadt, C.W., Zhou, J., 2016. Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation: Competition in bioremediation system. Environmental Microbiology 18, 205–218. https://doi.org/10.1111/1462-2920.12981

Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nat Rev Microbiol 10, 538–50. https://doi.org/10.1038/nrmicro2832

Feng, K., Zhang, Z., Cai, W., Liu, W., Xu, M., Yin, H., Wang, A., He, Z., Deng, Y., 2017. Biodiversity and species competition regulate the resilience of microbial biofilm community. Mol Ecol 26, 6170–6182. https://doi.org/10.1111/mec.14356

Global Water Microbiome Consortium, Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M.R., Li, Z., Van Nostrand, J.D., Ling, F., Xiao, N., Zhang, Y., Vierheilig, J., Wells, G.F., Yang, Y., Deng, Y., Tu, Q., Wang, A., Zhang, T., He, Z., Keller, J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J.M., He, Q., Curtis, T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat Microbiol 4, 1183–1195. https://doi.org/10.1038/s41564-019-0426-5

Granger, C.W.J., 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica 37, 424–438. https://doi.org/10.2307/1912791

Gutierrez, T., Green, D.H., Nichols, P.D., Whitman, W.B., Semple, K.T., Aitken, M.D., 2013. Polycyclovorans algicola gen. nov., sp. nov., an Aromatic-Hydrocarbon-Degrading Marine Bacterium Found Associated with Laboratory Cultures of Marine Phytoplankton. Appl. Environ. Microbiol. 79, 205–214. https://doi.org/10.1128/AEM.02833-12

Hibbing, M.E., Fuqua, C., Parsek, M.R., Peterson, S.B., 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nature Reviews Microbiology 8, 15–25. https://doi.org/10.1038/nrmicro2259

Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., Lewis, H.M., Olesen, J.M., Veen, F.J.F.V., Warren, P.H., Woodward, G., 2009. Review: Ecological

networks – beyond food webs. Journal of Animal Ecology 78, 253–269. https://doi.org/10.1111/j.1365-2656.2008.01460.x

Jenkins, D., Wanner, J., 2014. Activated Sludge - 100 Years and Counting. IWA Publishing.

Ji, Y., Angel, R., Klose, M., Claus, P., Marotta, H., Pinho, L., Enrich-Prast, A., Conrad, R., 2016. Structure and function of methanogenic microbial communities in sediments of Amazonian lakes with different water types. Environmental Microbiology 18, 5082–5100. https://doi.org/10.1111/1462-2920.13491

Jiang, X.T., Ye, L., Ju, F., Wang, Y.L., Zhang, T., 2018. Toward an Intensive Longitudinal Understanding of Activated Sludge Bacterial Assembly and Dynamics. Environ Sci Technol 52, 8224–8232. https://doi.org/10.1021/acs.est.7b05579

Ju, F., Zhang, T., 2015. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. The ISME Journal 9, 683–695. https://doi.org/10.1038/ismej.2014.162

Kéfi, S., Berlow, E.L., Wieters, E.A., Navarrete, S.A., Petchey, O.L., Wood, S.A., Boit, A., Joppa, L.N., Lafferty, K.D., Williams, R.J., Martinez, N.D., Menge, B.A., Blanchette, C.A., Iles, A.C., Brose, U., 2012. More than a meal… integrating non-feeding interactions into food webs. Ecology Letters 15, 291–300. https://doi.org/10.1111/j.1461-0248.2011.01732.x

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47, W256–W259. https://doi.org/10.1093/nar/gkz239

Liang, B., Ma, J., Cai, W., Li, Z., Liu, W., Qi, M., Zhao, Y., Ma, X., Deng, Y., Wang, A., Zhou, J., 2018. Response of chloramphenicol-reducing biocathode resistome to continuous electrical stimulation. Water Res 148, 398–406. https://doi.org/10.1016/j.watres.2018.10.073

Liao, W., Ding, J., Marinazzo, D., Xu, Q., Wang, Z., Yuan, C., Zhang, Z., Lu, G., Chen, H., 2011. Small-world directed networks in the human brain: Multivariate Granger causality analysis of resting-state fMRI. NeuroImage 54, 2683–2694. https://doi.org/10.1016/j.neuroimage.2010.11.007

Liébana, R., Arregui, L., Santos, A., Murciano, A., Marquina, D., Serrano, S., 2016. Unravelling the interactions among microbial populations found in activated sludge during biofilm formation. FEMS Microbiology Ecology 92, fiw134. https://doi.org/10.1093/femsec/fiw134

Nielsen, P.H., Kragelund, C., Seviour, R.J., Nielsen, J.L., 2009. Identity and ecophysiology of filamentous bacteria in activated sludge. FEMS Microbiology Reviews 33, 969–998. https://doi.org/10.1111/j.1574-6976.2009.00186.x

Nielsen, P.H., McMahon, K.D., 2014. Microbiology and microbial ecology of the activated sludge process. Activated sludge–100 years and counting 53–75.

Pepper, I.L., Gerba, C.P., Gentry, T.J. (Eds.), 2015. Environmental microbiology, Third edition. ed. Elsevier/AP, Academic Press is an imprint of Elsevier, Amsterdam.

Pfaff, B., 2008. VAR, SVAR and SVEC Models: Implementation Within *R* Package **vars**. J. Stat. Soft. 27. https://doi.org/10.18637/jss.v027.i04

Poisot, T., Stouffer, D.B., Gravel, D., 2015. Beyond species: why ecological interaction networks vary through space and time. Oikos 124, 243–251. https://doi.org/10.1111/oik.01719

Seth, A.K., 2005. Causal connectivity of evolved neural networks during behavior. Network: Computation in Neural Systems 16, 35–54. https://doi.org/10.1080/09548980500238756

Shao, Y., Chung, B.S., Lee, S.S., Park, W., Lee, S.-S., Jeon, C.O., 2009. Zoogloea caeni sp. nov., a floc-forming bacterium isolated from activated sludge. International Journal of Systematic and Evolutionary Microbiology, 59, 526–530. https://doi.org/10.1099/ijs.0.65670-0

van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op den Camp, H.J.M., Kartal, B., Jetten, M.S.M., Lücker, S., 2015. Complete nitrification by a single microorganism. Nature 528, 555–559. https://doi.org/10.1038/nature16459

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440. https://doi.org/10.1038/30918

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. The ISME Journal 10, 1669–1681. https://doi.org/10.1038/ismej.2015.235

Xia, Y., Wen, X., Zhang, B., Yang, Y., 2018. Diversity and assembly patterns of activated sludge microbial communities: A review. Biotechnology Advances 36, 1038–1047. https://doi.org/10.1016/j.biotechadv.2018.03.005

Zamkovaya, T., Foster, J.S., de Crécy-Lagard, V., Conesa, A., 2021. A network approach to elucidate and prioritize microbial dark matter in microbial communities. The ISME Journal 15, 228–244. https://doi.org/10.1038/s41396-020-00777-x

Zhang, B., Xu, X., Zhu, L., 2018. Activated sludge bacterial communities of typical wastewater treatment plants: distinct genera identification and metabolic potential differential analysis. AMB Expr 8, 184. https://doi.org/10.1186/s13568-018-0714-0

Zhou, J.Z., Ning, D.L., 2017. Stochastic Community Assembly: Does It Matter in Microbial Ecology? Microbiol Mol Biol R 81. https://doi.org/10.1128/MMBR.00002-17

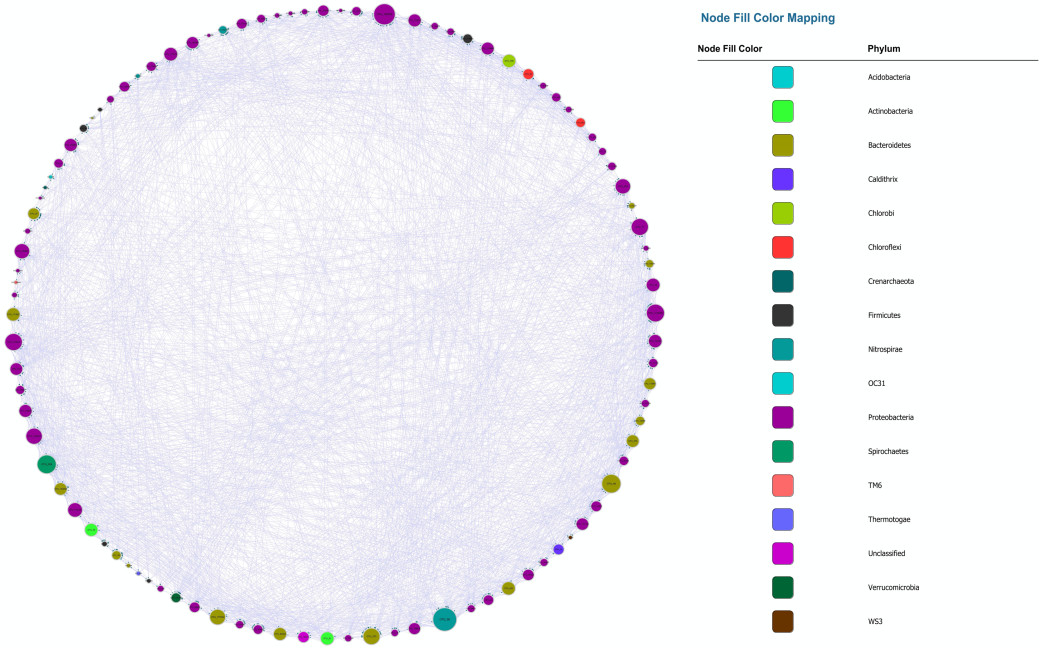
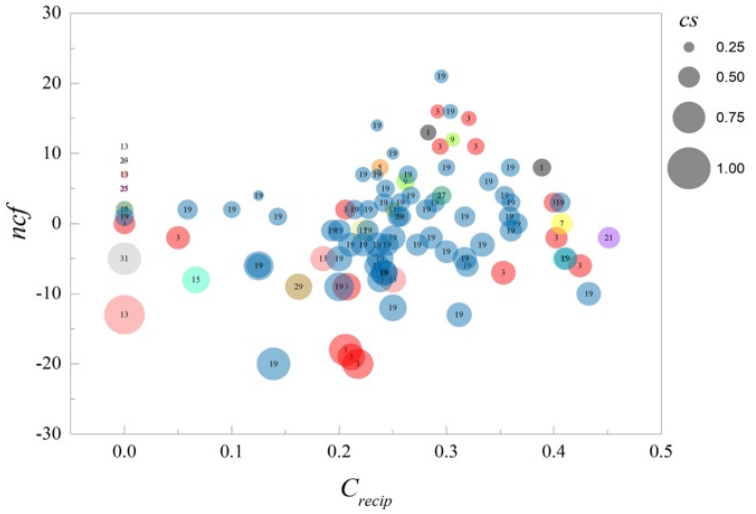


Figure 1. MGCN, each colour represents a separate phylum. The size of the node and node label is proportionate to the edge number of each node from 0 to 110. The arrows represent the direction of Granger causality.



Phylum	Number
Actinobacteria	1
Bacteroidetes	3
Caldithrix	5
Chlorobi	7
Chloroflexi	9
Crenarchaeota	11
Firmicutes	13
Nitrospirae	15
OC31	17
Proteobacteria	19
Spirochaetes	21
Thermotogae	23
TM6	25
Unclassified	27
Verrucomicrobia	29
WS3	31

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

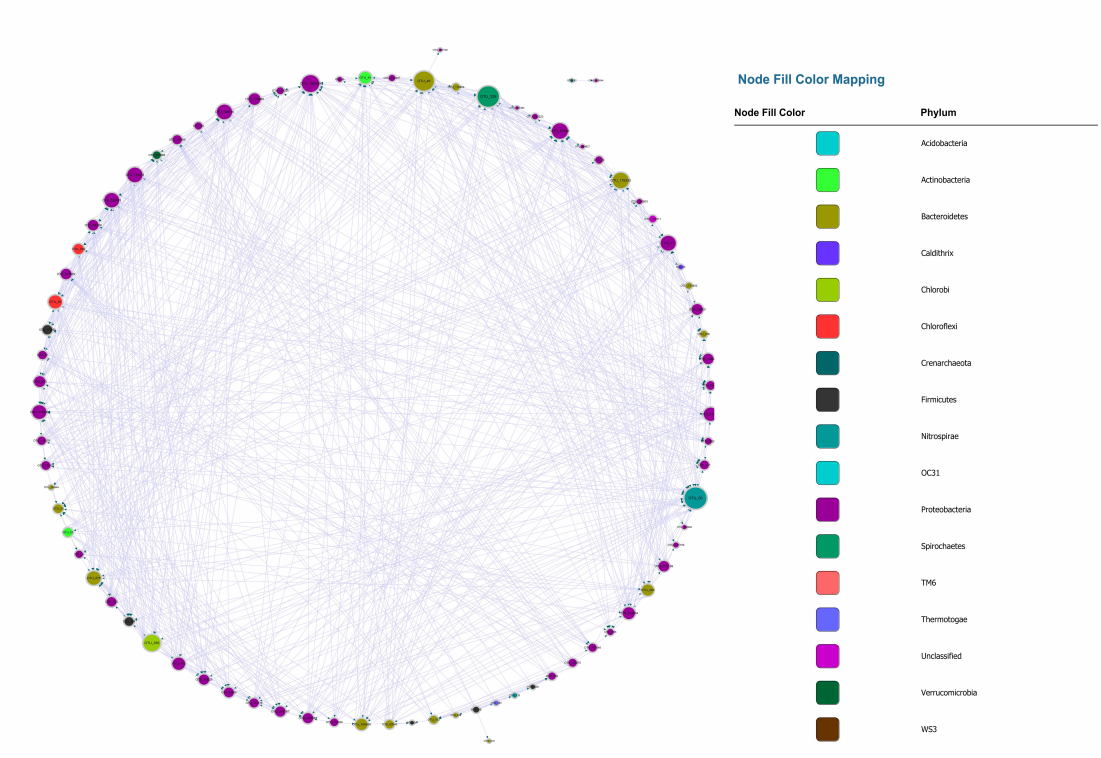
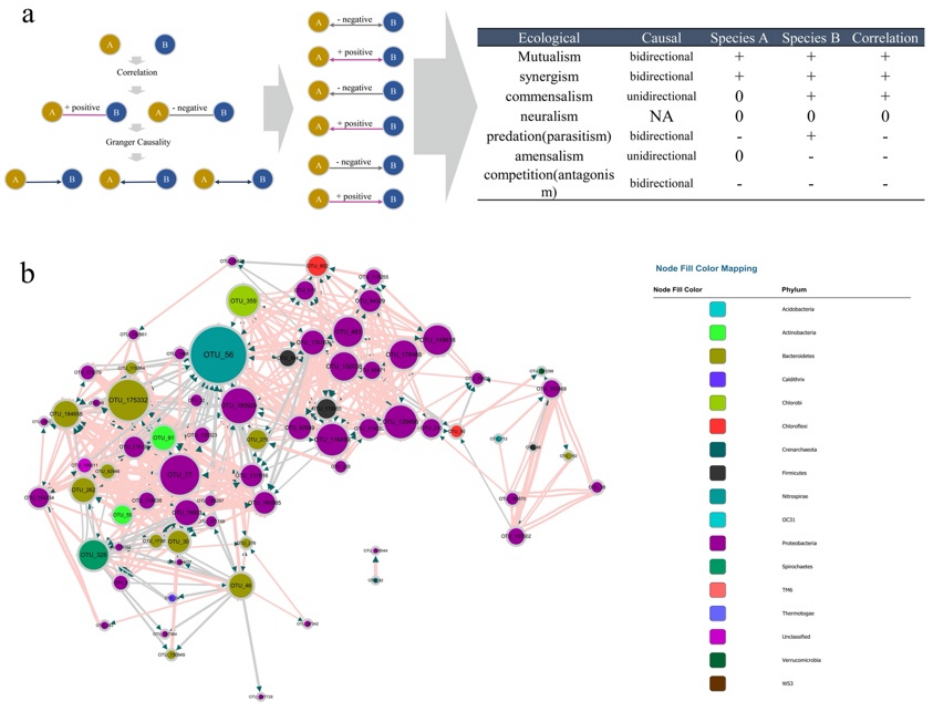


Figure 3. BoMGCN. Each colour represents a separate phylum. The size of the node and node label are proportionate to the edge number of each node from 0 to 110. The arrows represent the direction of Granger causality.

570



Ecological	Causal	Species A	Species B	Correlation
Mutualism	bidirectional	+	+	+
synergism	bidirectional	+	+	+
commensalism	unidirectional	0	+	+
neuralism	NA	0	0	0
predation(parasitism)	bidirectional	-	+	-
amensalism	unidirectional	0	-	-
competition(antagonism)	bidirectional	-	-	-

571

572 Figure 4. (A) Principle of the MCCN inference from the combination of causality and
573 correlation, the detail of correspondence from MCCN link to ecological interaction on
574 the right table. (B) MCCN, each colour represents an individual phylum. The size of
575 the node and node label are linearly proportionate to the edge number of each node
576 from 0 to 50. The arrows represent the direction of Granger causality. Pink and grey
577 link colours represent positive and negative associations, respectively. The size of the
578 link is proportionate to the correlation absolute value from 0 to 1.

579

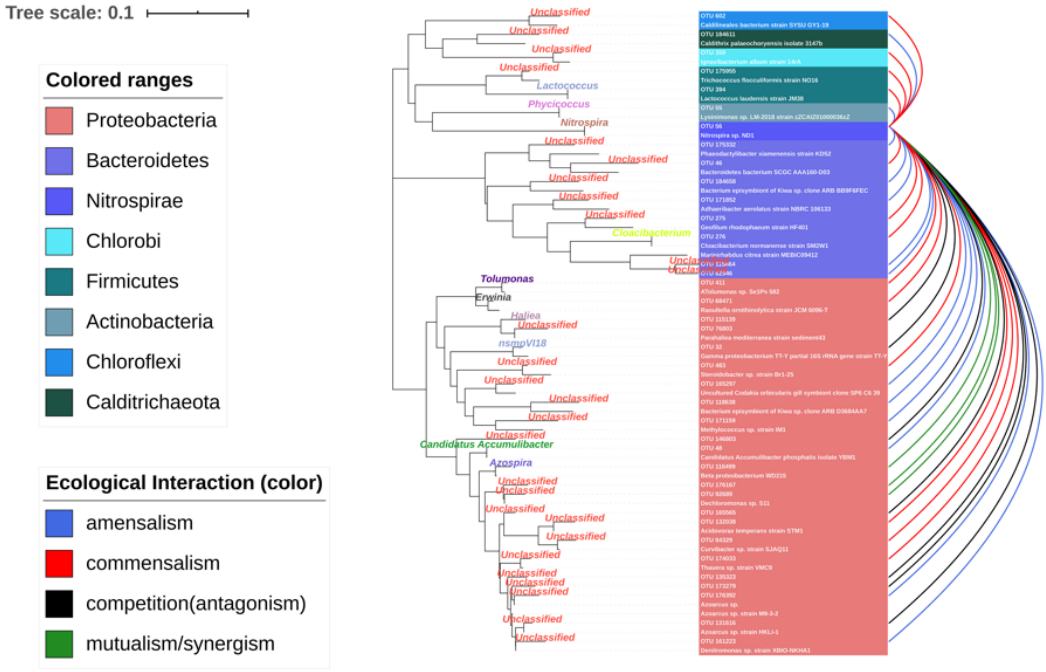


Figure 5. Ecological interaction of OTU56 with others at the OTU level. The colour represents the type of interaction. The phylogenetic tree shows the closest species according to the results of NCBI blast.

588 Table 1 Network indexes

Name	Equation	Description
Causal score (*cs*)	$cs = \frac{n_o}{n_o+n_i}$	n_o is the outdegree, n_i is the indegree.
Causal density (*cd*)	$cd = \frac{n}{2N(N-1)}$	n is the total number of significant causal links preserved in the network file. N is the size of the network.
Graph efficiency (*ge*)	$ge = 1 - \frac{n-(N-1)}{\binom{N}{2}}$	
Net causal flow (*ncf*)	$ncf = n_{no} - n_{ni}$ $n_{no} = N_v - n_i$ $n_{ni} = N_v - n_o$	n_{no} is the net outdegree of a specific node. n_i is the net indegree of the individual node. N_v is the sum of all neighbours of specific OTU.
Causal reciprocity (c_{recip})	$c_{recip} = \frac{n_{recip}}{n}$	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			Random
	MGCN	BoMGCN	MGCN	BoMGCN	MCN	BoMCN	MCN	BoMCN	MCCN	MCCN
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
cd	0.098	0.056							0.042	
ge	0.815	0.9	NA		NA		NA		0.93	NA
average *cs*	0.491	0.504							0.509	
average c_{recip}	0.238	0.148							0.157	

NA represents there is no data for this property.