*Published online 25 June 2020*

*NAR Genomics and Bioinformatics, 2020, Vol. 2, No. 2* **1**
*doi: 10.1093/nargab/lqaa042*

# Correlation-Centric Network (CCN) representation for microbial co-occurrence patterns: new insights for microbial ecology

**Pengshuo Yang [1], Chongyang Tan[1], Maozhen Han[1], Lin Cheng[2], Xuefeng Cui[3] and Kang Ning[1,*]**

[1]Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China, [2]Department of Engineering, Trinity College, 300 Summit Street, Hartford, CT 06106, USA and [3]School of Computer Science and Technology, Shandong University, Qingdao, Shandong 250100, China

Received November 25, 2019; Revised May 29, 2020; Editorial Decision June 04, 2020; Accepted June 05, 2020

## ABSTRACT

**Mainstream studies of microbial community focused on critical organisms and their physiology. Recent advances in large-scale metagenome analysis projects initiated new researches in the complex correlations between large microbial communities. Specifically, previous studies focused on the nodes (i.e. species) of the Species-Centric Networks (SCNs). However, little was understood about the change of correlation between network members (i.e. edges of the SCNs) when the network was disturbed. Here, we introduced a Correlation-Centric Network (CCN) to the microbial research based on the concept of edge networks. In CCN, each node represented a species–species correlation, and edge represented the species shared by two correlations. In this research, we investigated the CCNs and their corresponding SCNs on two large cohorts of microbiome. The results showed that CCNs not only retained the characteristics of SCNs, but also contained information that cannot be detected by SCNs. In addition, when the members of microbial communities were decreased (i.e. environmental disturbance), the CCNs fluctuated within a small range in terms of network connectivity. Therefore, by highlighting the important species correlations, CCNs could unveil new insights when studying not only the functions of target species, but also the stabilities of their residing microbial communities.**

## INTRODUCTION

As the quantity of microbiome data is growing exponentially, the diversity and complexity of the microbial community have made the integration and characterizing of microbial communities even more pressing issue (1,2). Thus, applying microbial network analysis to identify alternative community states and niches becomes an increasingly popular tool to investigate microbial community structures (3,4). Previous research has shown that microbial communities often exhibit non-random species–species co-occurrence patterns, and these observations suggest that a community structure is imprinted by species and their interactions (5,6). These interactions could provide a system-level view of a microbial community and a deep insight into the functional distribution in a microbial community (7,8).

The widely used Species-Centric Networks (SCNs, or species–species co-occurrence networks) are powerful tools to detect the co-occurrence correlation between two species in the microbial community. Specifically, the SCNs are constructed based on the relative abundance calculated from all the samples, and they provide an effective way to initiate species correlations in a microbial community (9,10). Hence, the SCNs achieve great success in detecting key species and key subnetworks (11,12), investigating the response of microbial community to external disturbances (13,14) and characterizing the microbial species' distribution and ecology (12,15). Such foundational works on species co-occurrences have led to the development of many methodological tools (16).

Despite the successful applications of the SCNs, one major limitation of the SCNs is that the importance of edges (i.e. species–species correlations) received insufficient attention (17,18). Previous research has demonstrated the importance that correlations of the edge distribution and properties could depict complex biological networks from another point of view and be of great significance in biological network researches (19,20). Hence, systematic analysis of species–species correlations is becoming one of the emerging fronts for microbiome research. This is mainly be-

*To whom correspondence should be addressed. Tel: +86 027 87793041; Fax: +86 027 87793041; Email: ningkang@hust.edu.cn

© The Author(s) 2020. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

cause of their important roles in facilitating the formation or the adoption of dynamic changes of a microbial community (21,22).

To exploit species–species correlations systematically, we apply the concept of edge networks by introducing the Correlation-Centric Networks (CCNs) to analyze microbial networks(16). Formally, each node of a CCN represents a species–species co-occurrence, and each edge of CCN represents a species shared between two co-occurrences. Comparing to the classic SCN, each node of a CCN is equivalent to an edge of its corresponding SCN and each edge of a CCN can be mapped to a node of its corresponding SCN. Thus, a CCN is an edge network representation of its corresponding SCN. Moreover, one important property of a CCN is its linear time complexity to be transformed from its corresponding SCN, and vice versa (20,23). This means that a CCN contains all the information in its corresponding SCN, and vice versa. Previously, the CCNs have been applied in another biological network analysis: the CCN is applied as predictive biomarkers in the gene co-expression network to predict diseases (24) and also as an indicator for personalized characterization of diseases (25). Moreover, the CCN also provides a unique perspective for systems biology with omics data (26,27). However, to the best of our knowledge, it has never been applied to analyze microbial networks.

In order to study the advantages of CCNs, we constructed them to analyze two well-studied cohorts of microbial communities: human gut microbiome time-series (collected in 1 year, during two dry and wet phases for Hadza hunter-gatherers) and ocean microbiome spatial series (collected in five major oceans of the world). In both microbial communities, the environmental factors were recorded. Based on these records, the relationship between the community structure and the external changes was investigated. The results of investigation of global and local properties indicated that the CCN retained all the information from the SCN and magnified the differences (i.e. two datasets were constructed with similar SCNs but identified with different CCNs). This observation indicated that the CCN was a better network from which we could detect critical species and correlations (both can serve as biomarkers) that drive differences in microbial communities. Moreover, based on network stability analysis, the CCNs were more robust than SCNs because CCNs possess more redundant information to cope with missing information. Specifically, when removing network members, the CCNs could better reflect the dynamic changes in the species–species correlation in order to cope with the course of environmental disturbance. Therefore, our results suggested that the CCN is a more powerful tool for investigating how species correlations would adapt to environmental disturbance in microbial communities.

## MATERIALS AND METHODS

### Microbial community cohorts

In order to demonstrate the advantages of CCNs over SCNs, we used two datasets in the following experiments: one human gut microbiome and one marine microbiome. For the human gut microbiome, the Hadza hunter-gatherers human gut microbiome data were obtained from the NCBI under accession number SRP110665 (28). All the 40 samples were divided into four phases as in the original research (28): dry_13 (dry phase in year 2013), wet_early_2014 (early wet phase in year 2014), wet_late_2014 (late wet phase in year 2014) and dry_14 (dry phase in year 2014). Previous studies found that their diet and their microbial communities' structure underwent a great change corresponding to the phase's change (28). After the samples were processed by MetaPhlAn (version 2.0) (29), taxonomical annotations for all the 40 samples were identified and calculated on genus level. Only genera with an average relative abundance over 0.1% in all samples were selected to filter out the most abundant genera which could play important roles in the microbial community (30,31). Finally, 64 genera (dry_13), 84 genera (wet_early_2014), 89 genera (wet_late_2014) and 62 genera (dry_14) were selected for network construction.

The marine microbiome was collected from the Tara Oceans Project (PRJEB1787 on EBI Metagenomics Portal) (32). We analyzed the prokaryotic part of this project, including 245 runs of more than 1.3 TB public sequencing data. Based on the physical and chemical information, we divided the 245 runs into five categories from their sampling regions: the Pacific Ocean (PO), the Atlantic Ocean (ATO), the Indian Ocean (IO), the Arctic Ocean (ARO) and the Mediterranean Sea (MS). To obtain the taxonomical annotations of the processed reads and to calculate the taxonomical abundance distribution on the genus level, we used Parallel-Meta (version 3.0) (33) to select only genera with an average relative abundance of over 0.1% in all samples, including 48 genera from PO, 32 genera from ATO, 38 genera from IO, 20 genera from ARO and 40 genera from MS.

### Network construction

To quantitatively identify the correlation between species, the SCN was constructed. Furthermore, based on the correlation distribution in SCN, the CCN was constructed. First, based on the relative abundance distribution on genus level, we constructed the species–species co-occurrence networks using Pearson correlation coefficient as follows:

$$\begin{aligned} \mathrm{PCC}_{x_i,x_j} &= \frac{C(x_i,x_j)}{\sqrt{(V(x_i)V(x_j))}} \\ &= \frac{E((x_i-\mu_i)(x_j-\mu_j))}{\sqrt{(E(x_i-\mu_i)^2E(x_j-\mu_j)^2)}} \end{aligned} \quad (1)$$

where $x_i$ and $x_j$ represented the relative abundances of two species detected in different samples, and $\mu$ represent the average abundance of the corresponding species across all samples.

Subsequently, we need to select a suitable threshold to define whether two species have a correlation or not. If a correlation was detected, a node should be added in the CCN and an edge should be added in the SCN. Thus, this step determined the topology of the network (as shown in Figure 1). To reduce the false positive for detected correlations, the biological significance of the correlation was needed. Previous studies reported that the drastic reduction in the number of edges may lead to the omission of important biological interactions (34). In this research, we found that the number of nodes and edges for both the CCN and the SCN would

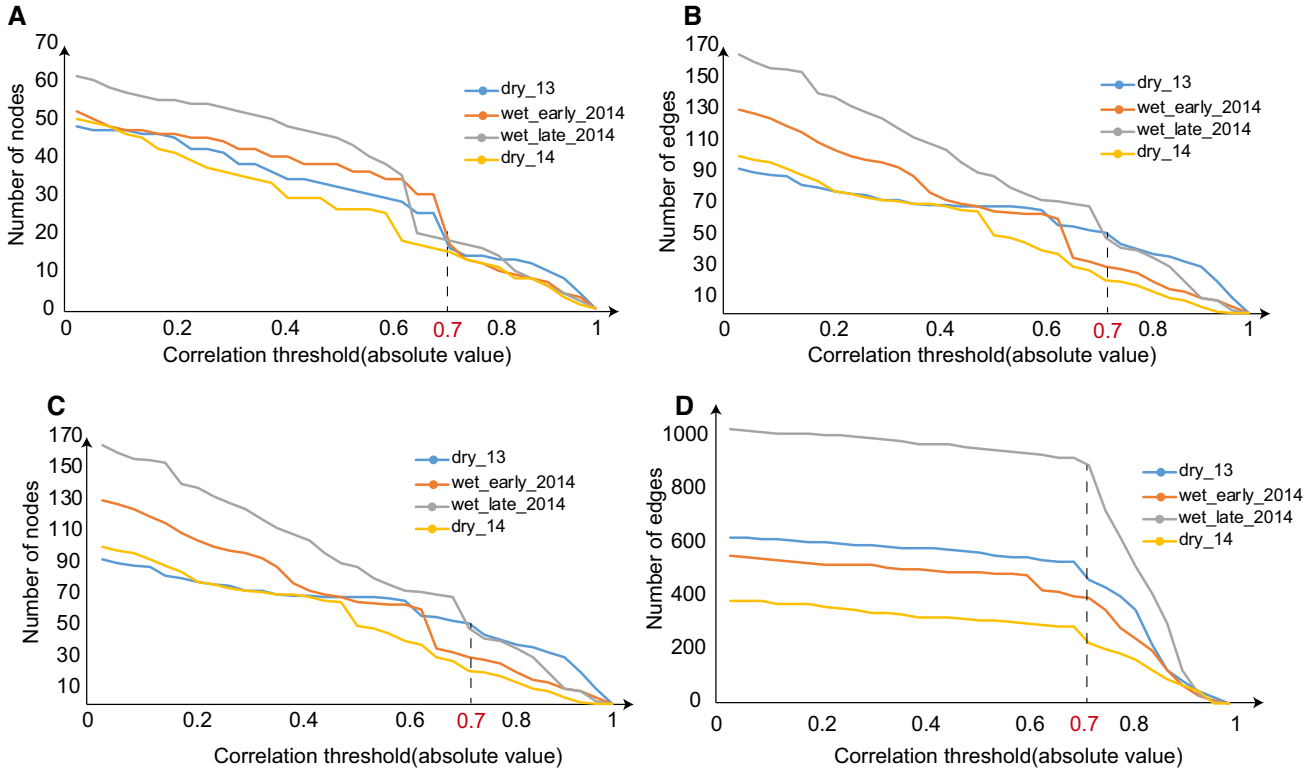


**Figure 1.** Correlation threshold analysis on the Hadza gut microbiome community. Here, the horizontal axis represents the threshold for the absolute species–species correlation, and the vertical axis represents the number of nodes or edges for the constructed network. Specifically, the numbers of nodes and edges of the constructed SCN are shown in figures (**A**) and (**B**), respectively. Similarly, the numbers of nodes and edges of the constructed the CCN are show in figures (**C**) and (**D**), respectively.

drop drastically when setting ±0.7 as the threshold. Hence, ±0.7 was selected to avoid such negative effects.

A SCN could be represented by an N × L incidence matrix B, where N represented the number of species and L represented the number of species–species co-occurrences (as shown in Figure 2A and B). In matrix B, an incidence (i.e. an element) was set to one if and only if there is a correlation between a pair of network members. Otherwise, the incidence is set to zero (Figure 2B). Thus, the network graph can be drawn given the incidence matrix (Figure 2C), and the incidence matrix can be calculated given the network graph (Figure 2D). The degree $k_i$ of node i and the number of nodes $k_\alpha$ attached to edge α (which always equals 2 since the SCN and CCN are both undirected graph) can be calculated as follows:

$$k_i = \sum_{\alpha} B_{i\alpha} \qquad k_\alpha = \sum_{i} B_{i\alpha} \tag{2}$$

where $B_{i\alpha}$ is an element in incidence matrix B.

Similarly, a CCN can be represented by an L × L incidence matrix C, and it can be calculated as follows:

$$C_{\alpha\beta} = \sum_{i} B_{i\alpha} B_{i\beta}(1 - \partial_{\alpha\beta}) \tag{3}$$

where the α and β represents two different edges of the SCN, and $\delta_{\alpha\beta}$ represents the isomorphism mapping (a mapping between two structures of the same type that can be reversed by an inverse mapping) of correlation between α and β in matrix C. Accordingly, the network graph can be drawn given the incidence matrix, and the incidence matrix can be calculated given the network graph.

It is important to note that the CCN retained all the information from the corresponding SCN while magnifying the differences. Whitney's uniqueness theorem (35) states that the original SCN can be fully recovered from its corresponding CCN. That means converting SCN into a CCN will not suffer information loss. Yet, the connectivity of a CCN is significantly higher than that of a SCN. Specifically, the connectivity in CCN can be measured by the number of edges $k_c$, which can be calculated from the degree $k_i$ of the original SCN as follows:

$$k_c = \sum_{i} k_i(k_i - 1)/2 \tag{4}$$

This equation illustrates how connectivity is significantly boosted by converting a SCN to a CCN. High-degree species are more frequently associated with other members in the community, and play the role as a potential 'hub' in the microbial community (36). Higher connectivity indicates that the CCN can characterize the high degree species of the original SCN very well, providing a perspective to further understand the role of these important members in microbial networks (36).

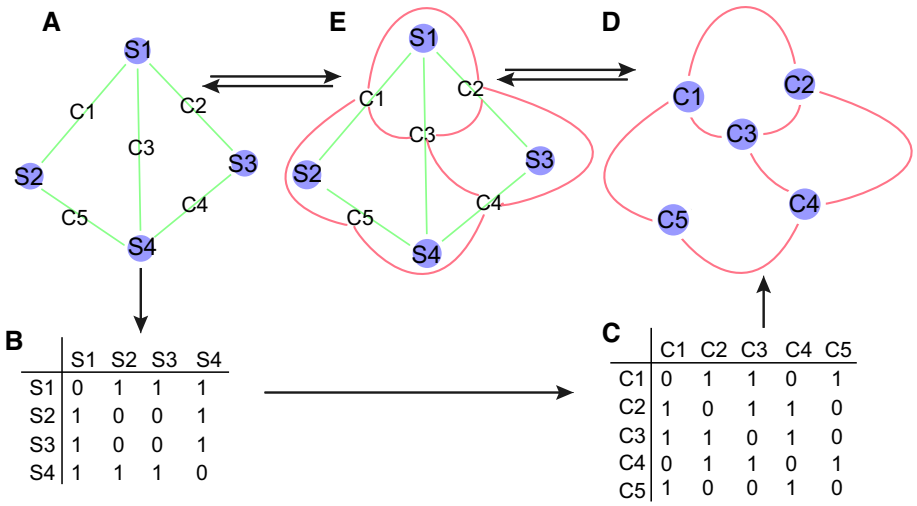


**Figure 2.** Converting a SCN to a CCN. (**A**) A simple SCN to be converted. (**B**) The incident matrix for the SCN constructed directly from the SCN. (**C**) The incidence matrix for the CCN calculated from (Equation 3). (**D**) The CCN constructed directly from the incident matrix for the CCN. (**E**) The relationship between the SCN and the CCN illustrated by merging them into one graph.

The workflow of SCN construction is shown in Algorithm 1, as below:

**Algorithm 1. Algorithm for SCN construction**

i. Input: relative abundance table of microbial community (matrix A). $x_i$ and $x_j$: two species in different samples. $\mu$: the average abundance of the corresponding species across all samples.
ii. Construct the SCN: for each pair of species, $x_i$ and $x_j$, calculate Pearson correlation coefficient $PCC_{xi,xj}$ as per (Equation 1).
iii. Convert the network graph of SCN into incidence matrix B, as exemplified in Figure 2A and B.
iv. Compute the incidence matrix C for CCN.as in (Equation 3).
v. The corresponding network graph for incidence matrix C is calculated, as exemplified in Figure 2C and D.

## Network assessment

After the SCN and the corresponding CCN were constructed, a comprehensive comparison of the two networks was conducted. First, to deduce and quantify the characteristics of the SCNs and CCNs, global network properties which capture topological features of large complex networks were evaluated and compared, including degree distribution, robustness, network diameter, etc. Second, to find the network members that mediate the differences, local network properties were investigated. Finally, the impact of missing information in the networks was explored.

First, the global properties were calculated and the nodes with high degrees were chosen as hub nodes (19). Based on these global properties, networks were compared based on the Jaccard index, which measured the overlap of network members for any two networks (37). The distance measured by Jaccard index was proportional to the magnitude of difference between two compared networks.

Second, to investigate the local properties of networks, we identified network biomarkers based on the K-nearest neighbor algorithm (KNN) classifier evaluated by the jackknife test (38). Network biomarkers detected the species with statistically different degree distribution in different networks. Network biomarkers identified network members which potentially drove the difference of networks, reflecting the adaptation of microbial community to environmental disturbance. These biomarkers commonly play an important role in the microbial community, such as regulating species balance or maintaining community robustness. Further investigation of these biomarkers could provide us an in-depth understanding of the microbial community changes. Especially in a CCN, the network biomarkers identified could potentially reflect the important microbial correlations that adapt to environmental disturbance.

Finally, to investigate the network stability, we applied R package SpiecEasi (version 1.0.4) for a CCN and the corresponding SCN to simulate 'environmental disturbance' (39). Biological robustness refers to the characteristic that the biological system keeps its structure and function stable when it is disturbed by environmental disturbance. Biological robustness is ubiquitous in biological systems and it is an important network character worthy of careful investigation. This simulation removed network members from high to low in the order of degree distribution across many replicates. For every degree of 'environmental disturbance', the network connectivity (the connection of various parts of a network to one another) was calculated to measure the stability of network. In the simulation experiment, network member removal implied species removal in the SCN, while edge removal implied interaction removal. To compare the SCN and CCN in a fair manner, a CCN network was reconstructed after nodes were removed from the corresponding SCNs. It was apparent that when the number of removed nodes reached a certain amount, the network would collapse, leading to very low connectivity of the network, which meant network breakdown. Moreover, the networks were illustrated for SCN (removing 5, 10, 20 and 35% nodes) and the corresponding CCN (removing 10, 25, 45 and 65% nodes) for one phase (wet_late_2014) as an example to explore the dynamic changes in the course of environmental disturbance.

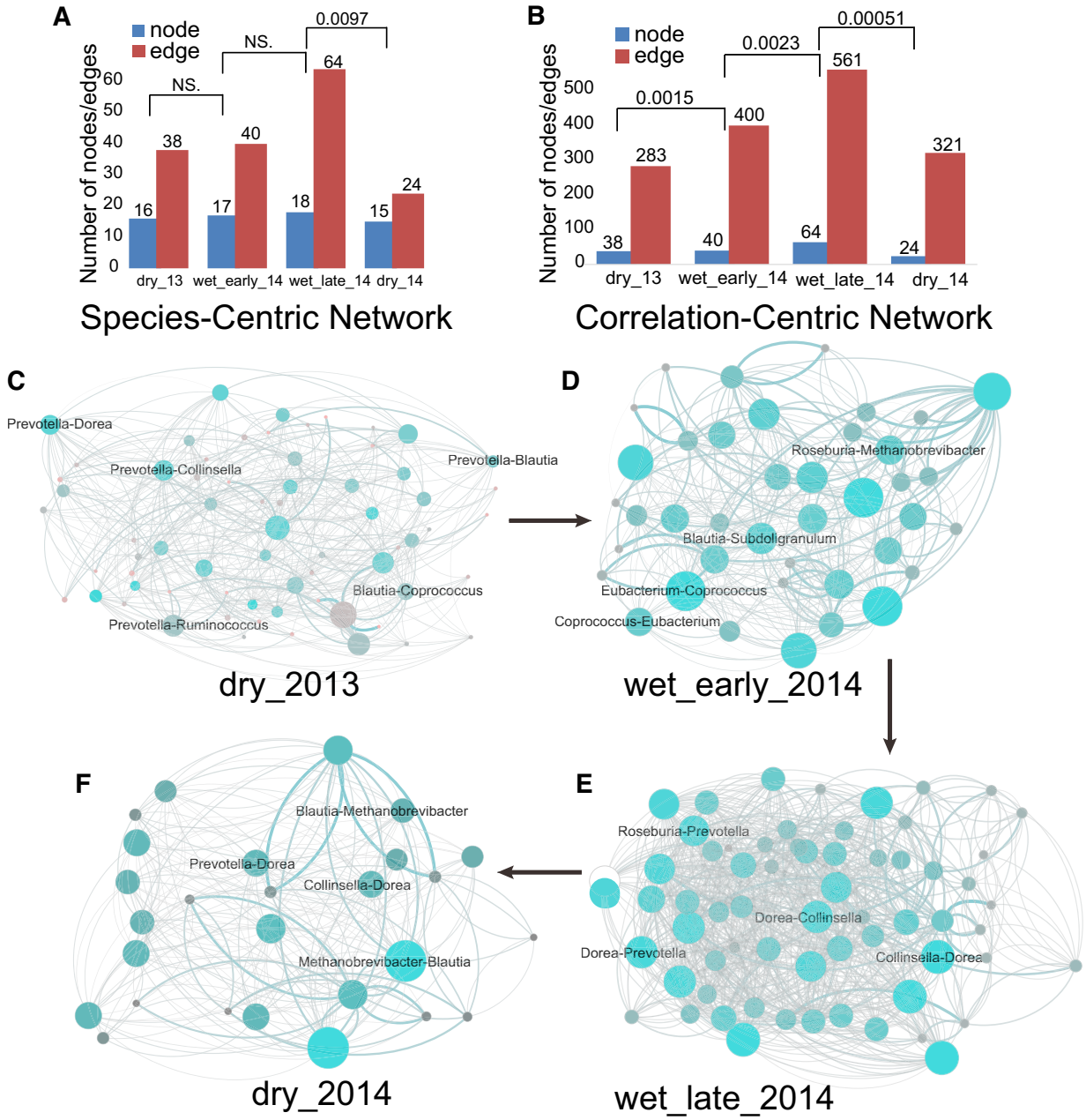


**Figure 3.** The SCNs and the CCNs for the Hadza gut microbiome community. (**A**) The number of nodes and edges for SCNs. To test whether two networks are significantly different, the Kruskal–Wallis test was applied for adjacent phases. NS denotes 'Not Significant' with *P*-values greater than 0.05. (**B**) The number of nodes and edges of CCNs. Similarly, the Kruskal–Wallis test was applied. (**C**–**F**), The CCNs of different phases. Here, the size of a node is proportional to its degree, and the correlations with high degrees are highlighted in bold.

## RESULTS AND DISCUSSIONS

### Network construction and observations during the phase transition

Based on the time-series human gut microbiome data, we promoted the SCNs for four phases (as shown in Figure 3A). Statistical results demonstrated that the SCNs provided insufficient resolution to distinguish these networks. During the phase transition, the topology converted as follows. For the dry phase of 2013, the network was composed of 16 genera and 38 correlations. For the wet phase of 2014, the network was composed of 17 and 18 genera, 40 and 64 correlations in the early and late wet phases, respectively. For the dry phase in 2014, the network contained 15 genera and 24 correlations, returning to a low complexity state. This conversion suggested that networks in wet phases were more complex than those in dry phases as reported previously (28). The topology conversion exhibited a cyclical

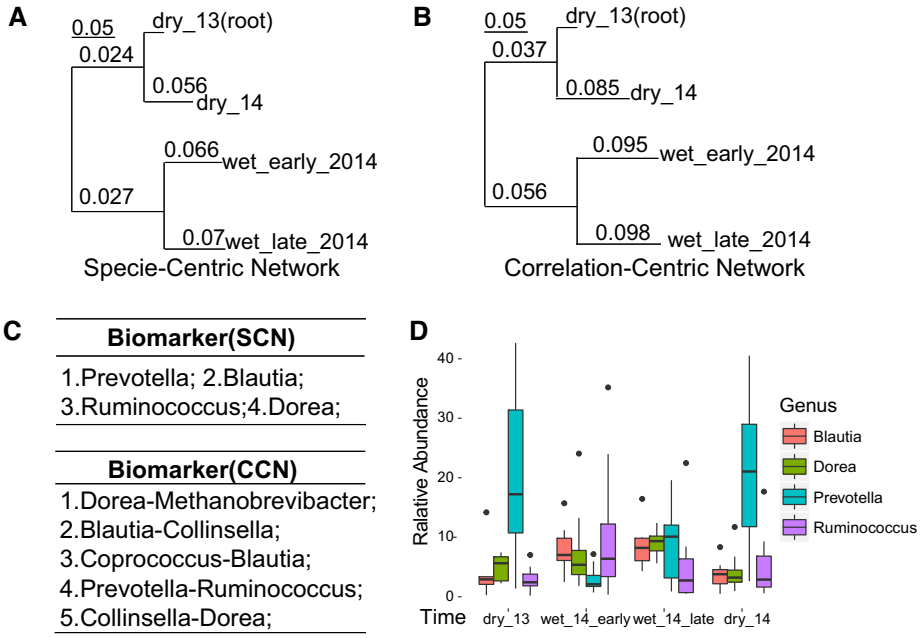


**Figure 4.** Network prosperity and stability analysis for SCNs and CCNs. (**A**) The similarity tree constructed by the Jaccard index for the SCNs. The Jaccard indexes are labeled in the corresponding branch, and the difference between compared networks is proportional to the Jaccard index. Similarly, (**B**) the similarity tree constructed by the Jaccard index for CCNs. (**C**) the biomarkers identified by SCNs and CCNs. (**D**) The box plots showing the abundance distributions in different phases for the biomarkers.

change in the gut microbial that adapted to environment change. However, the Kruskal–Wallis test results showed that the significant difference was only detected between the networks constructed by the dry phase in 2014 and the late wet phase ($P$-value = 0.0097), indicating the SCNs provide insufficient details to reduce the microbial communities.

The corresponding CCNs were constructed for four phases (as shown in Figure 3B). Based on the species correlation perspective, the differences between the microbial communities can be reflected. For the dry phase of year 2013, the CCN was consisted of 38 nodes and 283 edges (as shown in Figure 3C). For the wet phase in 2014, the networks were consisted of 40 and 64 nodes, 400 and 561 edges in early and late wet phases, respectively (as shown in Figure 3D and E). For the dry phase in 2014, the network was composed of 24 genera and 321 correlations (as shown in Figure 3F). The Kruskal–Wallis test results indicated that networks between dry_13 and wet_early_2014 ($P$-value = 0.0015), wet_early_2014 and wet_late_2014 ($P$-value = 0.0023), wet_late_2014 and dry_14 ($P$-value = 0.00051) were statistically different. According to (Equation 4), a species with a large degree yields more nodes in the corresponding CCNs, and this species led to the expansion of differences in four phases. Hence, we speculated that most of the important constituent species of gut microbiome could be preserved during the changing phases, and the changes of species correlations played a role in adapting to the changes in the environment, as reported previously (40,41). This result indicates that the species–species correlations possess a greater discriminative power than species for distinguishing samples, which is consistent with previous studies (24).

After investigating the SCNs and the corresponding CCNs on the system level, the comparison results indicated that the CCNs could maintain all the information and magnify the differences from the SCNs. During phase transitions, the CCNs possessed a higher network density (0.523 and 0.235 on average of four phases for CCN and SCN, respectively) and cluster coefficient (0.615 and 0.311 on average of four phases for CCN and SCN, respectively). Based on their global properties, similarity trees were constructed for the SCNs (as shown in Figure 4A) and its corresponding CCNs (as shown in Figure 4B). Both the SCNs and the CCNs constructed for dry phases were clustered in the same branch and wet phases in the other branch. This observation suggests that both the SCNs and the CCNs could detect this annual cyclic reconfiguration of the microbiome, as reported by previous research (28). However, the increase of Jaccard index in the CCN for the same branch showed that the differences among networks were magnified by the species–species correlations. Based on (Equation 4), species with a large degree (usually considered important members in the network) corresponded to more nodes than species with a low degree in the CCN, supporting the hypothesis that the magnified difference was potentially reflected by these important species.

To find the network members that potentially reflect this difference, the network biomarkers among the four phases were identified as shown in Figure 4C. For the SCNs, genera *Prevotella* ($P$-value = 0.0013), *Ruminococcus* ($P$-value = 0.0025), *Blautia* ($P$-value = 0.0092) and *Dorea* ($P$-value = 0.01) were identified as biomarkers. Genera *Prevotella*, *Ruminococcus*, *Blautia* are all dominant genera in Hadza gut

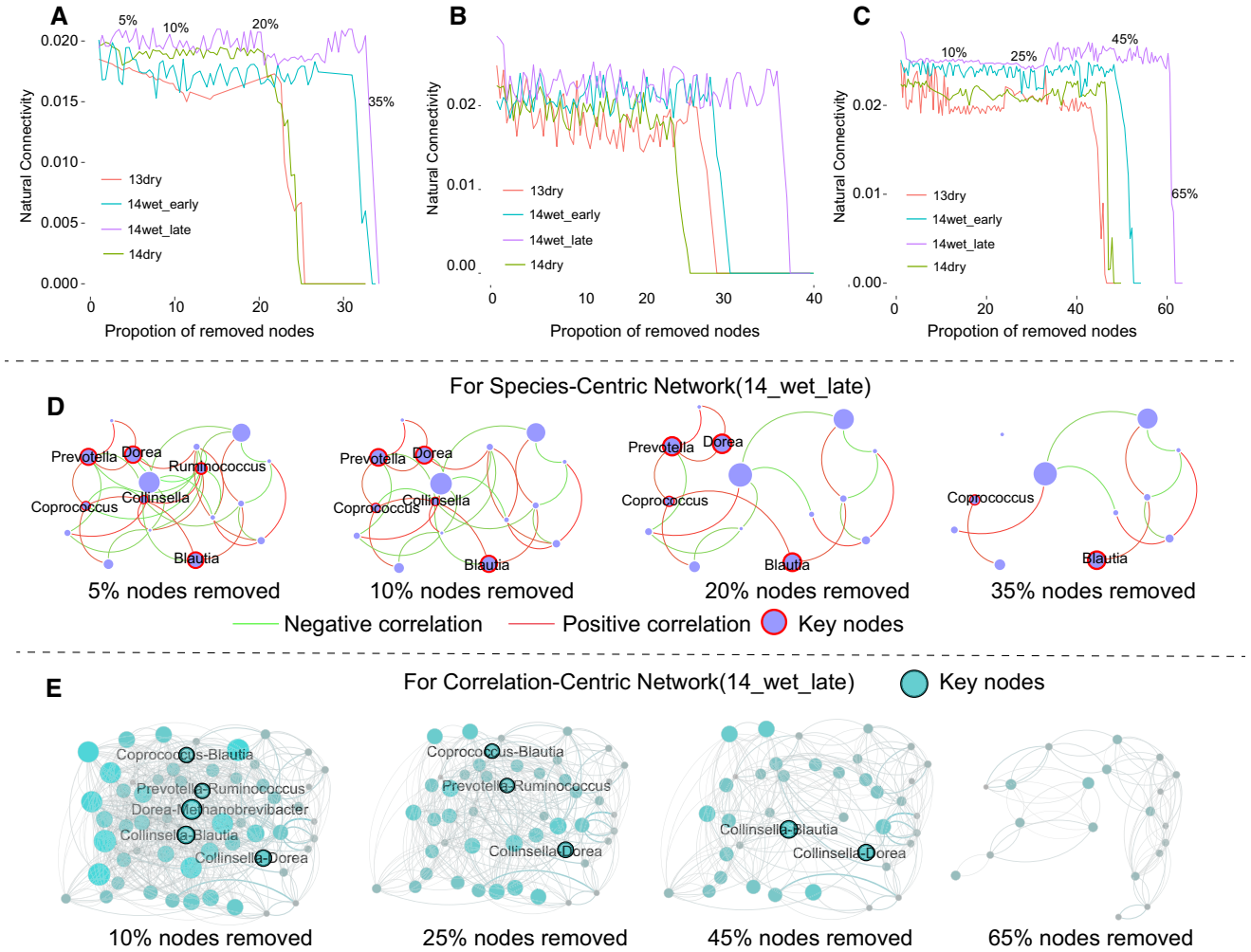


**Figure 5.** Network stability analysis on time-series human gut microbiome. (**A**–**C**) The stability analysis for SCNs and CCNs. Panel A is constructed based on the SCNs and Panel C is constructed based on the CCNs. In panel B, note that the statistical result is based on the CCNs that are reconstructed after nodes are removed from the corresponding SCNs. (**D**) Case studies of SCNs after removing 5, 10, 20 and 35% of nodes from the SCNs. Here, the positive correlations are labeled in red and the negative correlations are labeled in green. (**E**) Case studies of CCNs after removing 10, 25, 45 and 65% of nodes from the SCNs. Here, the key correlations are labeled as black circles.

microbiome (12.94 ± 12.41%, 3.09 ± 2.25%, 5.91 ± 4.05%, respectively) and their relative abundance varies at different phases as shown in Figure 4D. Exemplified by genera *Prevotella* and *Ruminococcus*, in the wet phases, the abundance of *Prevotella* decreased significantly. This might be driven by diet change (28), which has an inverse abundance distribution for genus *Ruminococcus*. Furthermore, for SCNs, the correlation between *Prevotella* and *Ruminococcus* was identified as biomarker, which indicated that their trade-off correlation can reflect the changes of the networks. Previous work has also supported the hypothesis that the two genera cooperated to respond to dietary changes (42).

Another biomarker that is only detected in the CCN, the correlation between *Blautia* and *Collinsella*, was correlated with improving host immunity and resisting various diseases (43,44). In the wet phase, the increase of relative abundance for two genera reflected that dietary changes may cause a disturbance in the microbial community structure. These two genera remained co-occurred even in the transition between dry and wet phases, which indicated that the cooperation of the two genera was important to maintain the stability of the microbial community. Compared to the SCN, the biomarkers found by the CCN could interpret how microbial communities respond to environmental disturbance better than SCNs. Thus, in the case of changes in the host diet during the seasonal variations, the CCN approach was shown to provide a deeper microbial network perspective of the change of microbial community to adapt to these external changes.

### Network stability analysis under external interference

After investigating the network properties, the network stability was explored by removing network members progressively (as shown in Figure 5A and C). The analysis for CCNs demonstrates that species correlation had redundancies to maintain the stability of microbial community in the event that network members were removed. To achieve a fair comparison, after the members of the SCN were removed, the corresponding CCNs were reconstructed. As shown in Figure 5A, the SCNs constructed in the early and late wet phases broke down when 31 and 32% of nodes were

removed, respectively. As shown in Figure 5B, the corresponding CCNs broke down when 33 and 38% of nodes were removed. Similarly, Figure 5C shows that the corresponding CCNs in early and late wet phases kept in a stable state until 51 and 60% of nodes were removed, respectively. Moreover, to compare the stability under different noise distributions (please see Supplementary Figure S1), various correlation thresholds were set to introduce different degrees of noise, and network stabilities of a SCN and a CCN were explored when different percentages of nodes were removed (i.e. 5, 10, 15 and 20% members were removed). It can be observed that a CCN exhibits higher stability than that of the corresponding SCN because the network connectivity only fluctuated within a small range before breaking down. This is mainly because the species–species correlations had redundancies to maintain the network properties, and they provided a perspective to interpret the stability of microbial community.

Even under external interference, a CCN could deal with dynamic change of the relationship between species when such changes occur. Analysis results indicate that community members tend to be associated with multiple members under environmental stresses, so that they can maintain backup relationships (a network member tends to be associated with multiple members with similar functions or taxonomic status) and avoid the collapse of the whole community. For example, based on the microbial composition on genus level in late wet phase, four SCNs were constructed by removing a different percentage of nodes (5, 10, 20 and 35%) (Figure 5D) and its corresponding CCNs by removing a different percentage of nodes (10, 25, 45 and 65%) (Figure 5E). In SCNs, genus *Coprococcus* which was identified as the dominant genus in all four phases ($3.09 \pm 2.25\%$), still existed in the broken-down network. Specifically, previous studies have illustrated the influence of relationship between *Catenibacterium* and *Coprococcus* to their hosts' health (45). This correlation disappeared when 20% of the nodes were removed in the SCN, but still exists when 65% of the nodes were removed in the CCN. Furthermore, a previous research has reported that the relationship between two probiotics (*Blautia* and genus *Collinsella*) protects the stability of the microbial community during environmental disturbances (46). This important correlation was identified as network biomarker in our study where it was detected in all four disrupted CCNs, but not in their corresponding SCNs. Simply put, the CCNs provide deeper insights to detect the dynamic process of network collapse: SCNs only detect changes in species composition, while CCNs have the ability to detect changes in species correlations. Again, these results indicate that community members tend to be associated with multiple members under environmental stresses. This redundancy of species correlation gives them the advantage of maintaining backup relationships, which could avoid the collapse of the whole community under environmental stresses, which could be supported by previous researches (2,47).

### Analysis based on CCLasso algorithm instead of Pearson correlation

All preceding analyses in this paper were conducted with the Pearson correlation. In order to show that our conclusions still hold for other correlations, we repeated the analysis using the CCLasso algorithm. The application of CCLasso algorithm on network analysis could eliminate the effect of sparse distribution on network construction, which could accurately reconstruct the microbial networks (48). We reconstructed the networks based on CCLasso algorithm instead of Pearson correlation, and found that the results can again confirm the previous results. At first, based on results in Supplementary Figures S2 and 3, a CCN has the ability to retain the characteristics of the species–species co-occurrence network, while at the same time possessing a greater discriminative power when comparing samples. Second, based on network stability analysis (as shown in Supplementary Figure S4), only small fluctuations occurred (rather than breaking down) when a large number of network members were removed. Although different network construction methods were performed, similar conclusions were drawn, indicating the conclusions we deduced in previous sections are insensitive to methods of calculation.

### Analysis based on marine microbiome instead of gut microbiome

All preceding analyses in this paper were conducted on the human gut microbial communities. In order to show that our conclusions still hold for other datasets, we repeated the analysis on the marine microbial communities. Specifically, we reconstructed the network with the same workflow using Pearson correlation coefficient. As shown in Figure 6, the marine microbial dataset from the Tara Oceans Project is divided into five regions: the PO, the ATO, the IO, the ARO and the MS. The properties of network constructed with different thresholds were calculated for constructing the network (as shown in Supplementary Figure S5). The analysis results lead to similar results based on the human gut microbiome.

We reanalyzed the network based on marine microbiome instead of human gut microbiome, and it led to similar conclusions: first, the similarity tree results (as shown in Figure 6A and B) indicated that the CCN retained the characteristics of the corresponding SCN and reflected a greater difference. Second, the network stability was investigated for the SCNs in Figure 6C and the CCNs in Figure 6E. To compare the stability of the SCNs and their corresponding CCN fairly, after members of the SCN were removed, these new networks were then recalculated to derive the corresponding CCN (Figure 6D). Based on different microbial communities (different taxonomical distribution and environmental factor), similar conclusions were drawn, indicating the conclusions we deduced in previous sections are insensitive to microbial communities.

## CONCLUSION

In this work, we introduced the CCN based on the concept of edge networks in a microbial community. Based on two cohorts of microbial communities, three features were highlighted. At first, a CCN has the ability to retain the characteristics of the corresponding SCN, while possessing a greater discriminative power when comparing samples. Second, only small fluctuations occurred, rather than breaking

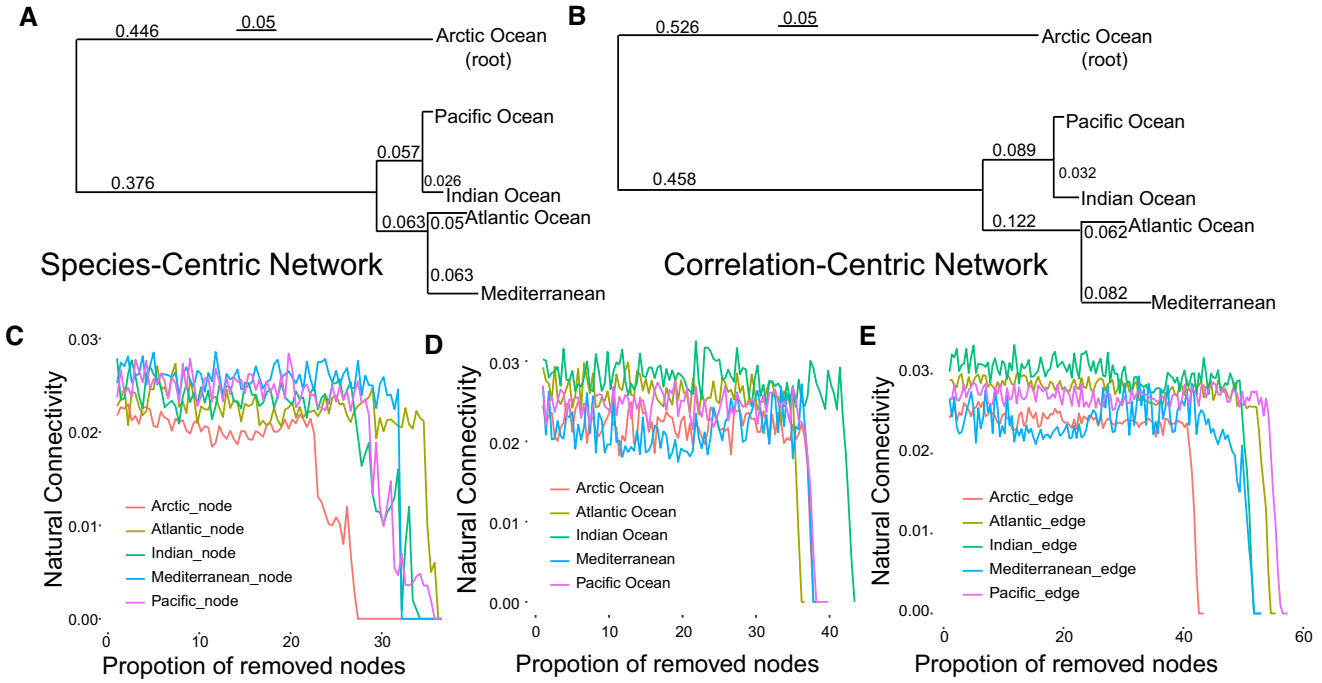


**Figure 6.** Experiments on the SCNs and the CCNs on ocean microbiomes with five regions. (**A**) The similarity tree constructed by the Jaccard index for the SCNs. The Jaccard indexes are labeled in the corresponding branch, and the difference between compared networks is proportional to the Jaccard index. Similarly, (**B**) the similarity tree constructed by the Jaccard index for CCNs. (**C**) The stability analysis for SCNs. (**D**) Stability analysis based on a CCN calculated from SCN after node removed. (**E**) The stability analysis for CCNs. Note that the CCNs are always reconstructed after nodes are removed from the corresponding SCNs.

down, when a large number of network members were removed from the CCNs. Third, the CCNs provide a microbial network perspective for the dynamic change of network correlations responding to different external factors.

These results highlight the importance of species–species correlations for maintaining the functions as well as stabilities of their residing microbial communities. And the network stability results indicate that community members tend to be associated with multiple members under environmental stresses, so that they can maintain backup relationships and prevent the collapse of the whole community. Therefore, the CCNs could reveal important ecological patterns, and are especially suitable for longitudinal studies to predict ecological patterns along the time series.

While our edge-based networks provide a flexible and valuable tool to gain a deeper understanding of complex microbial systems, we acknowledge a few limitations. Most importantly, microbiome datasets are being produced faster than ever, but we are only beginning to understand the structure and functioning of microbial communities. Hence, there exist no appropriate protocols to investigate the correlations among the species in microbial communities, nor is there a benchmark dataset to measure research tools' accuracy. One feasible way to detect the accuracy of the constructed network is to integrate multiple data sources and to perform multiple tests, which is applied in our research. Networks do not explain causality or even mechanisms. Networks are best considered as a generator of new hypotheses rather than anything to draw solid conclusions from. The integration of external data provides additional support for the hypotheses based on microbial networks.

## DATA AVAILABILITY

The source code is available at https://github.com/HUST-NingKang-Lab/Correlation-Centric-Network.

## SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

## ACKNOWLEDGEMENTS

We thank Francis Ng of Silverlight Inc. for proofreading of this work.

*Authors' contributions*: This study was designed by K.N. L.C. and X.C. contributed to the construction of the edge network. P.Y. and C.T. collected samples. P.Y. and M.H. analyzed the data. P.Y. and K.N. wrote the initial draft of the manuscript. All authors revised the manuscript.

## FUNDING

National Key Research and Development Program of China [2018YFC0910502]; National Natural Science Foundation of China [31671374, 31871334].

*Conflict of interest statement.* None declared.

## REFERENCES


1. Duran,P., Thiergart,T., Garrido-Oter,R., Agler,M., Kemen,E., Schulze-Lefert,P. and Hacquard,S. (2018) Microbial interkingdom interactions in roots promote arabidopsis survival. *Cell*, **175**, 973–983.
2. Rottjers,L. and Faust,K. (2018) From hairballs to hypotheses-biological insights from microbial networks. *FEMS Microbiol. Rev.*, **42**, 761–780.
3. Kumar,M., Ji,B., Zengler,K. and Nielsen,J. (2019) Modelling approaches for studying the microbiome. *Nat. Microbiol.*, **4**, 1253–1267.
4. Xiao,Y., Angulo,M.T., Friedman,J., Waldor,M.K., Weiss,S.T. and Liu,Y.Y. (2017) Mapping the ecological networks of microbial communities. *Nat. Commun.*, **8**, 1258–1270.
5. Ellegaard,K.M. and Engel,P. (2016) Beyond 16S rRNA community profiling: intra-species diversity in the gut microbiota. *Front. Microbiol.*, **7**, 1475–1459.
6. de Vries,F.T., Griffiths,R.I., Bailey,M., Craig,H., Girlanda,M., Gweon,H.S., Hallin,S., Kaisermann,A., Keith,A.M., Kretzschmar,M. *et al.* (2018) Soil bacterial networks are less stable under drought than fungal networks. *Nat. Commun.*, **9**, 913–920.
7. Raman,A.S., Gehrig,J.L., Venkatesh,S., Chang,H.W., Hibberd,M.C., Subramanian,S., Kang,G., Bessong,P.O., Lima,A.A.M., Kosek,M.N. *et al.* (2019) A sparse covarying unit that describes healthy and impaired human gut microbiota development. *Science*, **365**, eaau4735.
8. Surana,N.K. and Kasper,D.L. (2017) Moving beyond microbiome-wide associations to causal microbe identification. *Nature*, **552**, 244–247.
9. Sung,J., Kim,S., Cabatbat,J.J.T., Jang,S., Jin,Y.S., Jung,G.Y., Chia,N. and Kim,P.J. (2017) Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nat. Commun.*, **8**, 15393.
10. Lax,S., Cardona,C., Zhao,D., Winton,V.J., Goodney,G., Gao,P., Gottel,N., Hartmann,E.M., Henry,C., Thomas,P.M. *et al.* (2019) Microbial and metabolic succession on common building materials under high humidity conditions. *Nat. Commun.*, **10**, 1767.
11. Chen,L., Jiang,Y., Liang,C., Luo,Y., Xu,Q., Han,C., Zhao,Q. and Sun,B. (2019) Competitive interaction with keystone taxa induced negative priming under biochar amendments. *Microbiome*, **7**, 77–94.
12. Xue,Y., Chen,H., Yang,J.R., Liu,M., Huang,B. and Yang,J. (2018) Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. *ISME J.*, **12**, 2263–2277.
13. Kerdraon,L., Barret,M., Laval,V. and Suffert,F. (2019) Differential dynamics of microbial community networks help identify microorganisms interacting with residue-borne pathogens: the case of Zymoseptoria tritici in wheat. *Microbiome*, **7**, 125.
14. Yilmaz,B., Juillerat,P., Oyas,O., Ramon,C., Bravo,F.D., Franc,Y., Fournier,N., Michetti,P., Mueller,C., Geuking,M. *et al.* (2019) Microbial network disturbances in relapsing refractory Crohn's disease. *Nat. Med.*, **25**, 323–336.
15. Marasco,R., Rolli,E., Fusi,M., Michoud,G. and Daffonchio,D. (2018) Grapevine rootstocks shape underground bacterial microbiome and networking but not potential functionality. *Microbiome*, **6**, 3–20.
16. Evans,T.S. and Lambiotte,R. (2010) Line graphs of weighted networks for overlapping communities. *Eur. Phys. J. B*, **77**, 265–272.
17. De Smet,R. and Marchal,K. (2010) Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.*, **8**, 717–729.
18. Huang,J.K., Carlin,D.E., Yu,M.K., Zhang,W., Kreisberg,J.F., Tamayo,P. and Ideker,T. (2018) Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.*, **6**, 484–495.
19. Yang,B., Li,M., Tang,W., Liu,W., Zhang,S., Chen,L. and Xia,J. (2018) Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nat. Commun.*, **9**, 87–108.
20. Yu,X., Zhang,J., Sun,S., Zhou,X., Zeng,T. and Chen,L. (2017) Individual-specific edge-network analysis for disease prediction. *Nucleic Acids Res.*, **45**, e170.
21. Layeghifard,M., Hwang,D.M. and Guttman,D.S. (2017) Disentangling Interactions in the Microbiome: a network perspective. *Trends Microbiol.*, **25**, 217–228.
22. Yan,Y., Qiu,S., Jin,Z., Gong,S., Bai,Y., Lu,J. and Yu,T. (2017) Detecting subnetwork-level dynamic correlations. *Bioinformatics*, **33**, 256–265.
23. Zeng,T., Zhang,W., Yu,X., Liu,X., Li,M., Liu,R. and Chen,L. (2014) Edge biomarkers for classification and prediction of phenotypes. *Sci. China Life Sci.*, **57**, 1103–1114.
24. Yu,X., Li,G. and Chen,L. (2014) Prediction and early diagnosis of complex diseases by edge-network. *Bioinformatics*, **30**, 852–859.
25. Liu,X., Wang,Y., Ji,H., Aihara,K. and Chen,L. (2016) Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res.*, **44**, e164.
26. Chen,L. (2014) Systems biology with omics data. *Methods*, **67**, 267–268.
27. Wang,Y., Zhang,X.S. and Chen,L. (2018) Integrating data- and model-driven strategies in systems biology. *BMC Syst. Biol.*, **12**, 38–41.
28. Smits,S.A., Leach,J., Sonnenburg,E.D., Gonzalez,C.G., Lichtman,J.S., Reid,G., Knight,R., Manjurano,A., Changalucha,J., Elias,J.E. *et al.* (2017) Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science*, **357**, 802–806.
29. Segata,N., Waldron,L., Ballarini,A., Narasimhan,V., Jousson,O. and Huttenhower,C. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods*, **9**, 811–814.
30. Hegde,S., Khanipov,K., Albayrak,L., Golovko,G., Pimenova,M., Saldana,M.A., Rojas,M.M., Hornett,E.A., Motl,G.C., Fredregill,C.L. *et al.* (2018) Microbiome interaction networks and community structure from Laboratory-Reared and Field-Collected aedes aegypti, Aedes albopictus, and Culex quinquefasciatus mosquito vectors. *Front. Microbiol.*, **9**, 2160–2176.
31. Steele,J.A., Countway,P.D., Xia,L., Vigil,P.D., Beman,J.M., Kim,D.Y., Chow,C.E., Sachdeva,R., Jones,A.C., Schwalbach,M.S. *et al.* (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.*, **5**, 1414–1425.
32. Pesant,S., Not,F., Picheral,M., Kandels-Lewis,S., Le Bescot,N., Gorsky,G., Iudicone,D., Karsenti,E., Speich,S., Trouble,R. *et al.* (2015) Open science resources for the discovery and analysis of Tara Oceans data. *Sci.Data*, **2**, 150023–150039.
33. Jing,G., Sun,Z., Wang,H., Gong,Y., Huang,S., Ning,K., Xu,J. and Su,X. (2017) Parallel-META 3: comprehensive taxonomical and functional analysis platform for efficient comparison of microbial communities. *Sci. Rep.*, **7**, 666–682.
34. Smita,S., Katiyar,A., Chinnusamy,V., Pandey,D.M. and Bansal,K.C. (2015) Transcriptional Regulatory Network Analysis of MYB transcription factor family genes in rice. *Front. Plant Sci.*, **6**, 1157–1176.
35. Whitney,H. (1932) Congruent graphs and the connectivity of graphs. *Am. J. Math.*, **54**, 150–168.
36. Evans,T.S. and Lambiotte,R. (2009) Line graphs, link partitions, and overlapping communities. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **80**, 016105–016113.
37. Vorontsov,I.E., Kulakovskiy,I.V. and Makeev,V.J. (2013) Jaccard index based similarity measure to compare transcription factor binding site models. *Algorith.Mol. Biol.*, **8**, 16–23.
38. Li,P., Hu,Y., Yi,J., Li,J., Yang,J. and Wang,J. (2015) Identification of potential biomarkers to differentially diagnose solid pseudopapillary tumors and pancreatic malignancies via a gene regulatory network. *J. Transl. Med.*, **13**, 30–40.
39. von Mering,C., Kurtz,Z.D., Müller,C.L., Miraldi,E.R., Littman,D.R., Blaser,M.J. and Bonneau,R.A. (2015) Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.*, **11**, e1004226.
40. Angulo,M.T., Moog,C.H. and Liu,Y.Y. (2019) A theoretical framework for controlling complex microbial communities. *Nat. Commun.*, **10**, 381–384.
41. Kelder,T., Stroeve,J.H., Bijlsma,S., Radonjic,M. and Roeselers,G. (2014) Correlation network analysis reveals relationships between diet-induced changes in human gut microbiota and metabolic health. *Nutr. Diabetes*, **4**, e122.
42. Wu,G.D., Chen,J., Hoffmann,C., Bittinger,K., Chen,Y.Y., Keilbaugh,S.A., Bewtra,M., Knights,D., Walters,W.A., Knight,R. *et al.* (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science*, **334**, 105–108.

43. Jost,T., Lacroix,C., Braegger,C.P., Rochat,F. and Chassard,C. (2014) Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ. Microbiol.*, **16**, 2891–2904.
44. Hu,J., Nie,Y., Chen,J., Zhang,Y., Wang,Z., Fan,Q. and Yan,X. (2016) Gradual changes of gut microbiota in weaned miniature piglets. *Front. Microbiol.*, **7**, 1727–1742.
45. Duncan,S.H., Holtrop,G., Lobley,G.E., Calder,A.G., Stewart,C.S. and Flint,H.J. (2004) Contribution of acetate to butyrate formation by human faecal bacteria. *Br. J. Nutr.*, **91**, 915–923.
46. Fujio-Vejar,S., Vasquez,Y., Morales,P., Magne,F., Vera-Wolf,P., Ugalde,J.A., Navarrete,P. and Gotteland,M. (2017) The gut microbiota of healthy chilean subjects reveals a high abundance of the Phylum Verrucomicrobia. *Front. Microbiol.*, **8**, 1210–1221.
47. Cornforth,D.M. and Foster,K.R. (2013) Competition sensing: the social side of bacterial stress responses. *Nat. Rev. Microbiol.*, **11**, 285–293.
48. Fang,H., Huang,C., Zhao,H. and Deng,M. (2015) CCLasso: correlation inference for compositional data through Lasso. *Bioinformatics*, **31**, 3172–3180.