BMC Bioinformatics

SOFTWARE

Open Access

MicroSSNet: an R package for microbial network construction and analysis at the single-sample and aggregated levels

Zecheng Tang[1†], Daohua Zhuang[1†], Xinmin Duan[3†], Qingqing Gong[1], Chen Tian[1], Peicheng Jiang[1], Jiangkun Yu[1], Fei Li[1], Fangfang Zhao[4], Guolin Shi[5], Hang Yang[1], Qinghang Du[1], Tong Li[1], Zhiqiang Ye[2*] and Zhigang Zhang[1*]

[†]Zecheng Tang, Daohua Zhuang and Xinmin Duan contributed equally to this work.

*Correspondence:
Zhiqiang Ye
zhiqiang.yes@gmail.com
Zhigang Zhang
zhangzhigang@ynu.edu.cn
[1]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, School of Life Sciences, Yunnan University, Kunming 650091, Yunnan, China
[2]Key Laboratory of Pesticide and Chemical Biology of Ministry of Education, School of Life Sciences, Central China Normal University, Wuhan 430079, Hubei Province, China
[3]Department of Orthopedic Surgery, Inner Mongolia International Mongolian Medical Hospital, Hohhot City 010020, Inner Mongolia, China
[4]College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China
[5]Neurosurgery, The Second Affiliated Hospital of Kunming Medical University, Kunming 650101, China

Abstract

Background Network analysis is a fundamental tool for elucidating microbial interactions, which are crucial for understanding the mechanisms that shape ecosystem structure and function. However, aggregated co-abundance/co-occurrence network approaches that infer pairwise relationships among biological entities from large sample collections often overlook sample-specific interaction patterns. To address this limitation, we developed MicroSSNet, an R package designed for analyzing microbial networks, including both aggregated and single-sample networks.

Results We designed MicroSSNet primarily to fill the current gap in bioinformatics tools for constructing single-sample networks (SSNs) from microbiome data, and we evaluated both the performance and limitations of ssPCC-based SSNs using simulated and real datasets. Through Monte Carlo simulations, we assessed the statistical behavior of ssPCC and highlighted scenarios in which ssPCC is less powerful. We then applied MicroSSNet to two distinct datasets: a human gut metagenomic dataset and a soil 16S rRNA gene dataset. In the human gut dataset, SSNs revealed unique edges not detected in the aggregated network. In the soil dataset, SSN features showed some predictive value for group classification. However, SSN-derived patterns should be interpreted cautiously, as they may not exclusively reflect true interaction changes. MicroSSNet additionally implements a full aggregated-network workflow, including bipartite networks and extensive topological property analysis.

Conclusions Together, MicroSSNet offers a framework for constructing and analyzing both single-sample and aggregated microbial networks. In this work, we also highlight the potential and limitations of single-sample network approaches, supporting their application as exploratory tools in microbiome research across individual and population levels. The package is freely available on GitHub (https://github.com/TangZecheng622/MicroSSNet).

Keywords Microbiome, Network analysis, R package, Single-sample network

Background

The rapid advancement of high-throughput sequencing technologies has propelled microbial community research to unprecedented levels, driving continuous innovation in bioinformatics tools [1]. Large-scale analyses based on 16S rRNA sequencing and metagenomic datasets have enabled a deeper exploration of complex microbial interactions. Recently, complex network analysis [2], a method widely utilized in sociology [3], computer science [4], and physics [5], has been increasingly applied in microbiology [6], offering new insights into ecological structure, stability, and function [7]. Consequently, network modeling has emerged as a critical strategy for interpreting omics data and understanding microbial ecosystem dynamics.

Microbial network analyses provide ecological insights from both local and global perspectives. Locally, identifying key nodes reveals species that are critical for maintaining community stability [8]. Globally, network structures can reflect community responses to surrounding changes. For instance, Yuan et al. [9] demonstrated that soil microbial networks presented increased complexity and stability under warming scenarios, suggesting intrinsic self-regulatory capabilities within microbial communities.

Given the inherent sparsity and compositionality of microbiome datasets, sparse modeling methods such as SparCC [10] and SPIEC-EASI [11] have been widely adopted to infer robust microbial associations. However, these aggregated network approaches that infer pairwise relationships among biological entities from large sample collections often overlook sample-specific features crucial for understanding microbial heterogeneity and fail to capture the dynamic, sample-specific interaction patterns that are critical for personalized microbiome research.

Single-sample network (SSN) analysis offers a complementary perspective by estimating individualized networks that capture potential variation in pairwise associations across samples. There is growing interest in constructing SSNs from interaction data, and several methodological approaches have been proposed to achieve this. The single-sample Pearson correlation coefficient (ssPCC) [12] and linear interpolation to obtain network estimates for single samples (LIONESS) [13] are two main computational approaches. While ssPCC excels in differential network analysis with reference controls, LIONESS provides a flexible framework for inferring networks without predefined groups [14].

Despite their demonstrated potential across multiple research areas [15–17], the application of SSN methods in microbiome research remains limited. On the tooling side, microbiome-ready SSN tools—especially ssPCC-based modules—are largely absent; most existing packages target aggregated networks. At the application level, evaluations demonstrating the feasibility and value of ssPCC-based SSNs across microbiome datasets are lacking.

To address these gaps, we developed MicroSSNet, an open-source R package designed for microbial network analysis at both single-sample and aggregated levels—explicitly filling the tooling gap for ssPCC-based SSN construction. Importantly, SSN features are statistical quantities derived from the assumptions of ssPCC or LIONESS, and they may be influenced by factors such as abundance variability, noise, and baseline correlation strength. As such, these estimates do not necessarily represent pure interaction-level changes. Therefore, SSN-based findings should be interpreted cautiously and viewed primarily as exploratory signals rather than definitive biological conclusions.

Implementation

MicroSSNet is designed to analyze microbiome datasets from diverse sources (e.g., the environment, gut, and skin) at both the single-sample and aggregated levels. As with aggregated networks, SSNs are applicable to microbiome datasets from diverse sources. However, SSNs are edge-centric—they emphasize changes in specific pairwise associations, whereas aggregated networks are topology-centric, highlighting structures such as hubs (Fig. 1a).

MicroSSNet provides microbial network construction, analysis, and visualization through streamlined inputs and extensive analytical functionality (Supplementary Fig. S1). It requires a feature-abundance table as the mandatory input and can optionally be supplemented with taxonomy and metadata tables.

MicroSSNet supports two primary analytical models: SSN analysis, which employs either ssPCC or LIONESS to infer SSNs, and aggregated network analysis, which constructs networks via established methods such as SparCC, SPIEC-EASI, Pearson, or Spearman. In practice, CLR-transformed abundances are recommended for Pearson/Spearman-based aggregated networks and for ssPCC/LIONESS-derived SSNs.

Following network construction, MicroSSNet can compute network properties for both single-sample and aggregated networks. The integrated visualization module enables intuitive exploration through node coloring on the basis of taxonomic annotation or modularity classification and introduces bipartite networks to effectively illustrate cross-domain microbial associations. Additionally, the tool incorporates PCA, PCoA, and limma-based analyses to identify and compare differences among SSNs (Fig. 1b).

Finally, all network data and associated statistical results are systematically exported in both textual and graphical formats (Supplementary Fig. S2), providing comprehensive support for downstream analysis and interpretation.

Results

Statistical performance of the ssPCC test

We evaluated the performance of the ssPCC module using both simulated and real microbiome datasets. Unless otherwise specified, SSN hereafter denotes single-sample networks constructed using ssPCC.

The purpose of this test was to evaluate two properties of the ssPCC module on simulated data: sensitivity (the ability to detect changes in correlation) and false-positive control (Type I error). We generated correlations between reference samples (ρ0) and single-query samples (ρq) using bivariate log-normal distributions, with systematically varied correlation structures. In each simulation, a set of n reference samples was generated with a predefined correlation coefficient (ρ_0), and a single query sample was generated with a potentially different correlation (ρq). By systematically varying ρ_0, ρq, mean abundances, and coefficients of variation (CV), we evaluated the sensitivity of ssPCC under different conditions.

The results for a sample size of $n = 1{,}000$ demonstrated that the sensitivity of the ssPCC test increased with the baseline correlation strength. When the baseline (reference) correlation satisfied $|\rho_0| \geq 0.6$, the maximum sensitivity exceeded 40%; at $|\rho_0| = 0.9$, the maximum sensitivity was > 60%, indicating strong sensitivity under highly correlated baselines (Fig. 2a). However, this also highlights an inherent limitation: ssPCC is most

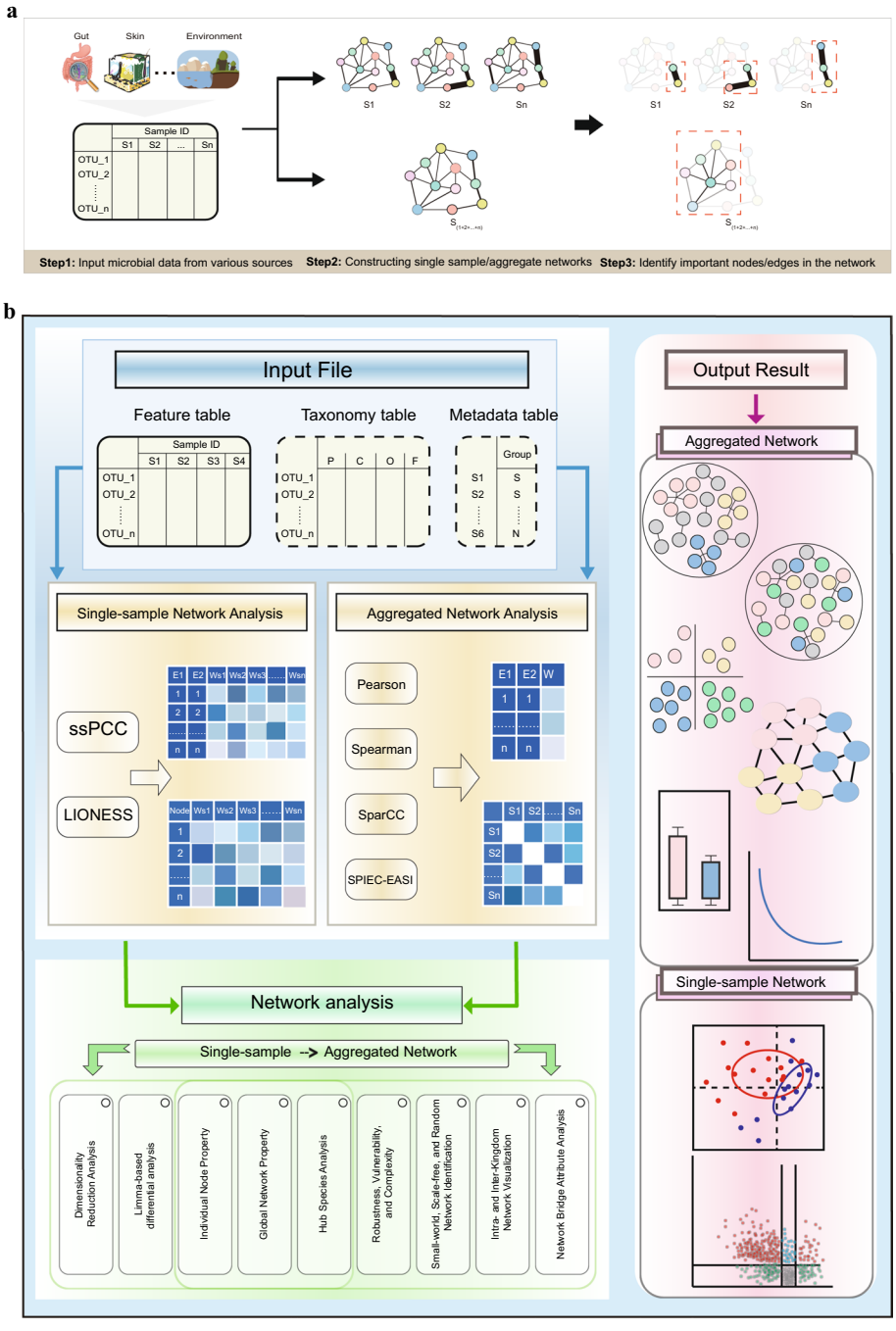

Fig. 1 **a** Overview of MicroSSNet's Working Principles. Both aggregated and single-sample network analyses apply to microbiome datasets from diverse sources (e.g., environmental, gut, and skin). However, single-sample networks are edge-centric—they emphasize changes in specific pairwise associations, whereas aggregated networks are topology-centric, highlighting the global structure. **b** Workflow and outputs of MicroSSNet

effective when strong pairwise correlations are present and does not capture all types of network changes.

The type I error rates were well controlled across the reference sample sizes. Once the baseline included ≥ 20 samples, the empirical error rate remained close to the nominal 5% (± 2.5%), with notable inflation observed at n = 5 (Fig. 2b). Consistent with this pattern, the sensitivity was essentially stable across moderate-to-large sample sizes, and the

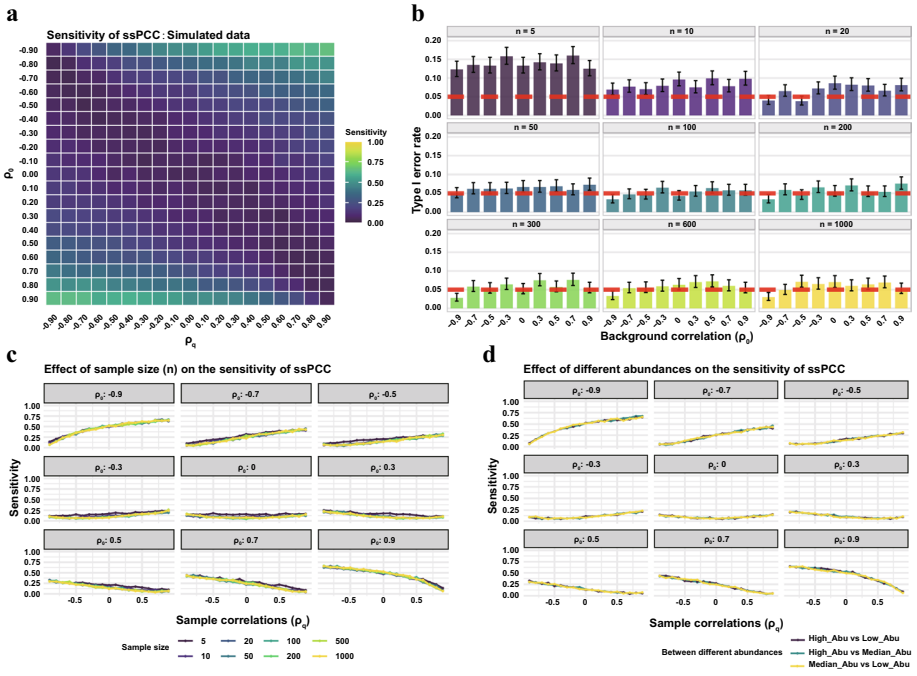

Fig. 2 Evaluation of the Statistical Performance and Limitations of the Single-sample Pearson Correlation Coefficient (ssPCC). where ρ_0 is the correlation of the reference network and ρq is the correlation of the query sample. **a** Actual statistical power of the ssPCC test based on a Z test as a function of the effect size, estimated from simulated data. **b** Actual Type I error rate (α) of the ssPCC test based on a Z test as a function of the reference sample size. The dashed red line denotes the nominal 5% significance level. Error bars represent 95% confidence intervals calculated using the Wilson score method based on the number of false positives and total tests at each sample size. **c** Effect of sample size (n) on the sensitivity of the ssPCC. **d** Effects of different abundances on the sensitivity of the ssPCC

apparently higher sensitivity at n = 5 reflects inflated false positives rather than improved sensitivity (Fig. 2c). These findings suggest that ssPCC becomes statistically reliable once the reference group reaches approximately 20 samples, although substantially larger reference sets do not yield proportionally higher sensitivity.

Furthermore, we assessed the sensitivity of the ssPCC test across varying abundance levels. Our analyses confirmed that the test maintained consistent sensitivity across abundance levels of 1%, 0.1%, and 0.01%, representing high, medium, and low abundances, respectively, indicating robust performance regardless of rare species abundance levels (Fig. 2d). This abundance-independent behavior indicates that the test is robust to the rarity of taxa.

To assess the effect of abundance variability on ssPCC performance, we simulated four CV settings—low variance (CV = 0.2), medium variance (CV = 0.4), high variance (CV = 0.8), and abnormally high variance (CV = 1.2)—representing dispersion in microbiome data. Sensitivity decreased monotonically with increasing CV (Supplementary Fig. S3), reflecting reduced sensitivity under conditions of high abundance variability.

These results highlight several dimensions of ssPCC's sensitivity limitations: even under controlled and idealized simulation settings, ssPCC does not detect all correlation changes and cannot achieve uniformly low false-positive rates. Real microbiome datasets may exhibit additional sources of variability—such as compositional effects, overdispersion, and heterogeneous baseline correlations—that further challenge ssPCC performance.

Human gut metagenomic dataset validation for SSN analysis

To evaluate the applicability of ssPCC-based single-sample networks to real microbiome data, we applied MicroSSNet to a gut metagenomic dataset from individuals who relocated from low altitude to high altitude and later returned [18]. This exploratory analysis aimed to detect individual-level interaction changes that might be masked by correlation averaging in aggregated networks, rather than to infer mechanistic causality.

At LS2 (day 24 of moving to high altitude), using CQ1 (10 days before moving to high altitude) as the reference, we computed the SSN edge weights for each of 39 samples. Given prior reports linking *Blautia_A wexlerae* to altitude-associated microbiome responses, we focused on edges connected to this taxon to characterize potential sample-level patterns of interaction change.

In this heatmap, each row represents the SSN edge-weight between *Blautia_A wexlerae* and another taxon. Positive edge weights (shown in red) indicate a shift toward stronger positive correlation relative to the baseline network, while negative weights (blue) indicate a shift toward negative correlation (Fig. 3a).

Cosine dissimilarities between samples were calculated based on the SSN edge-weight matrix, and column clustering was performed using partitioning around medoids (PAM) for K = 2–30. The optimal number of clusters (K = 14) was selected by maximizing the average silhouette width (Supplementary Fig. S4a). The resulting 14 clusters (C1–C14) represent groups of samples with potentially similar interaction-change patterns (Fig. 3a, Supplementary Fig. S4b). Clusters C5–C14 (≤ 3 samples each) were considered rare or individualized, whereas C1–C4 (≥ 5 samples) were classified as reproducible patterns.

To characterize recurrent patterns within each cluster, we defined an edge as "reproducible" if it showed significant SSN weights in ≥ 80% of samples in that cluster. Four reproducible patterns were identified, each containing a set of edges exhibiting consistent changes among its member samples (Fig. 3b). Differences across clusters reflect overall patterns in the direction and composition of interaction changes rather than dependence on any single edge.

For comparison, we constructed an aggregated Pearson correlation network for LS2 (Supplementary Fig. S4c). Aggregated networks capture stable, high-prevalence associations, while SSN-specific edges reflect individual-level changes that averaging may mask. For example, the *B. wexlerae–A. hallii* edge appears positive in clusters C1 and C2 but negative in C4, and the *B. wexlerae–M. torques* edge shows a strong shift only in C13 (n = 1), illustrating how SSNs can potentially reveal heterogeneous or low-prevalence deviations that do not appear in the aggregated analysis.

However, such SSN edges must be interpreted with caution. Even when an edge exhibits consistent changes across multiple samples, the SSN signal may reflect noise or other confounding factors rather than true biological interaction changes. Therefore, SSNs provide exploratory insights that complement aggregated networks, but the inferred edge changes should not be taken as direct evidence of altered ecological interactions.

Soil 16S rRNA dataset validation for SSN analysis

To explore the classification and predictive performance of SSN features, we applied MicroSSNet to an environmental microbiome dataset—a long-term soil warming 16S rRNA dataset. A reference network was constructed using untreated 2009 samples, and

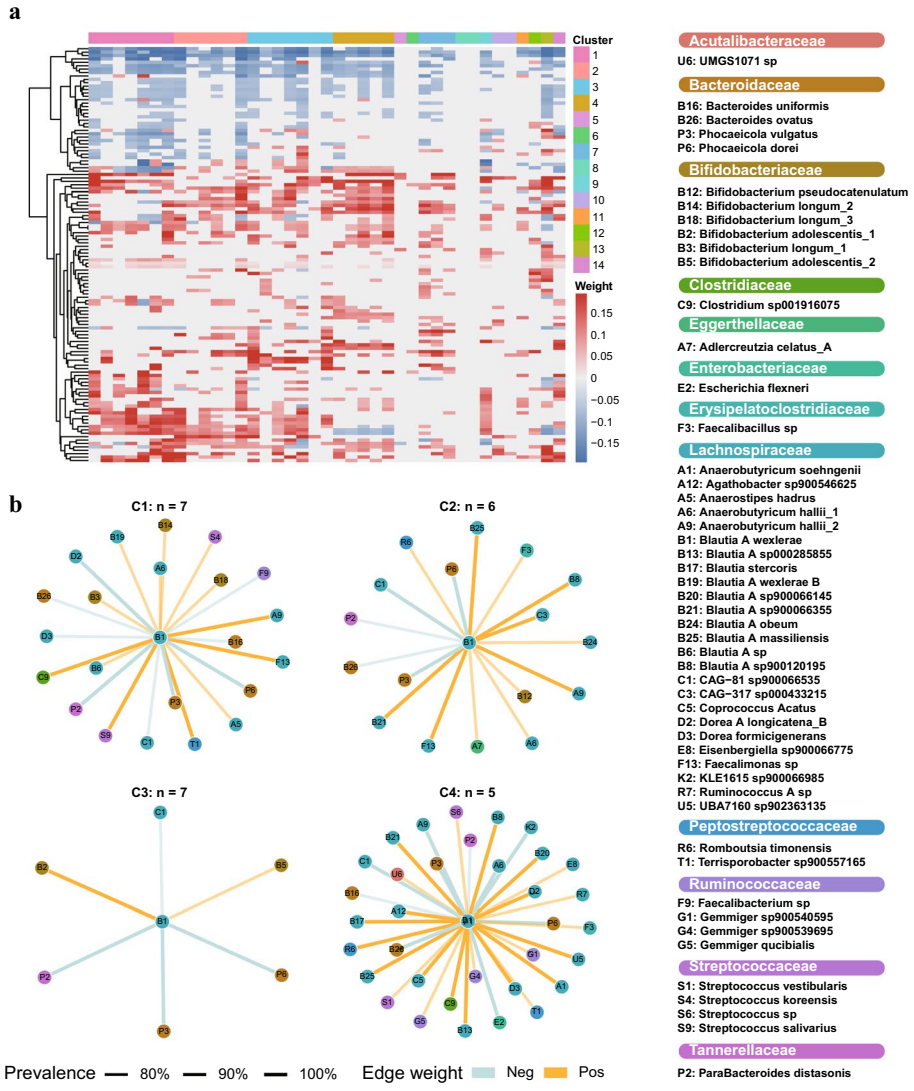

Fig. 3 Validation results on a human gut metagenomic dataset for single-sample network analysis. **a** Each column represents one sample (LS2), and each row represents a microbial interaction change (edge) between *Blautia_A wexlerae* and another taxon, as predicted by SSN. The color indicates the SSN edge weight, with red indicating a shift toward stronger positive correlation, and blue indicating a shift toward negative correlation. Column clustering was performed using partitioning around medoids (PAM) based on cosine dissimilarities computed from the SSN edge-weight matrix. Row clustering was performed using hierarchical clustering. **b** The four reproducible SSN patterns (C1–C4; ≥ 5 samples each) are visualized as cluster-specific single-sample interaction patterns centered on *Blautia_A wexlerae*; the edges shown are those present in ≥ 80% of samples within the respective cluster (prevalence ≥ 0.8). The labels (e.g., A1, F9) represent the taxa connected to *Blautia_A wexlerae*. The initial letter indicates the genus abbreviation, and the number corresponds to the taxon index shown in the accompanying legend

SSNs were subsequently generated for 2014 samples from both warming and control treatments.

Both abundance profiles and SSN features showed significant separation between warming and control samples (Supplementary Fig. S5a–c). When the sample size was reduced (e.g., n = 8), the PERMANOVA p-value for abundance-based separation increased to 0.848 (Fig. 4a), whereas SSN node-weight features still provided significant separation (Fig. 4b). This suggests that SSN features may capture aspects of between-sample variability that are not fully reflected in abundance alone, especially in small sample sets where abundance-based metrics often lack statistical power.

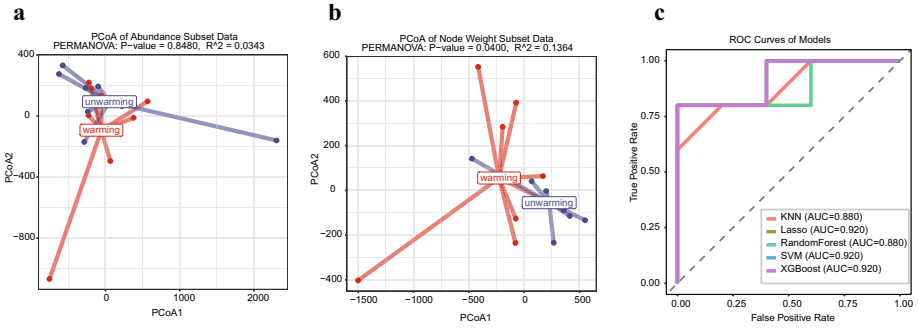

Fig. 4 Validation Results on a Soil 16S rRNA Dataset for Single-Sample Network Analysis. PCoA plots were generated using Euclidean distances computed from **a** relative abundance profiles and **b** SSN node-weight features, each derived from 8 samples. Statistical significance of group separation was assessed using PERMANOVA. PCoA in this study refers to metric Multidimensional Scaling (MDS) computed from Euclidean distances. **c** Prediction results of various machine learning models based on single-sample network node-weight features

In addition, using SSN features for classification yielded consistently strong performance across multiple models (k-nearest neighbors, AUC = 0.88; lasso, AUC = 0.92; random forest, AUC = 0.88; support vector machine, AUC = 0.92; XGBoost, AUC = 0.92), highlighting the predictive potential of SSN features (Fig. 4c). However, compared to abundance-based features, SSN-based models did not show significantly improved classification performance (Supplementary Fig. S5d).

Notably, there was substantial overlap between the differentially abundant taxa identified in SSN-based and abundance-based analyses (Supplementary Fig. S5e). To reduce or exclude classification signals driven by abundance differences, we removed all nodes with significant between-group abundance differences. We then classified samples using SSN and abundance features with XGBoost under a nested 5×5 outer cross-validation. Paired AUCs (SSN vs. abundance) were compared using the Wilcoxon signed-rank test. As expected, the AUC of the abundance-based model decreased (mean AUC = 0.628), whereas the SSN-based features retained relatively strong classification performance (mean AUC = 0.83) (Supplementary Fig. S5f). These results indicate that SSN features retain predictive value independent of abundance-driven signals.

Overall, these results explored the predictive and classification potential of SSN features, as they may capture structural changes in single-sample networks. SSN features enable computation of pairwise sample distances and embedding (e.g., PCoA/UMAP) to characterize cohort structure [19], and they serve as informative features for machine learning–based classification and prediction. They also support differential analysis to localize perturbed edges and nodes. Nonetheless, it is critical to recognize that a portion of the classification and predictive performance of SSN features may be attributable to underlying abundance variation. Accordingly, this approach is particularly appropriate for settings in which changes in microbial interactions occur without concurrent alterations in taxonomic abundance.

Aggregated network analysis on the soil 16S rRNA dataset

To complement the single-sample analyses, we further applied MicroSSNet to aggregated network construction and visualization via 2014 warming and ambient (control) soil samples reported by Yuan et al. (2021). Aggregated networks were constructed with aggregation_netpipeline (Pearson; r.threshold = 0.8, p.threshold = 0.01) and visualized for the warming treatment (Supplementary Fig. S6a). To illustrate cross-kingdom

(bacteria–fungi) associations, we used bipartite_netpipeline (Pearson; r.threshold = 0.6, p.threshold = 0.05) and visualized the warming treatment (Supplementary Fig. S6b).

In the analysis of global and local network properties, MicroSSNet not only computes metrics such as average degree, diameter, average local efficiency, and relative modularity (RM) but also provides visualizations of Zi–Pi [20] (within-module degree Zi; participation coefficient Pi) (Supplementary Fig. S6c) as well as vulnerability [21], robustness [22], and complexity [23, 24] (Supplementary Fig. S6d–f)). Across these analyses, the network patterns and treatment differences were directionally consistent with prior results [9] for this dataset (Supplementary Table 1).

Discussion

MicroSSNet is a user-friendly and flexible R package that offers pipeline functions for streamlined analysis, as well as modular functions for advanced users to customize each step of the workflow. The software is intended for researchers studying microbial communities who aim to explore population-level trends or individual-specific interaction patterns.

Compared with existing microbial network analysis tools such as iNAP [25], NetCoMi [26] and ggClusterNet [27], which focus solely on aggregated networks, MicroSSNet fills a gap in existing bioinformatics tools by providing an integrated pipeline for single-sample network construction based on the ssPCC and LIONESS frameworks. This allows users to explore interaction patterns specific to individual samples, which are masked in the aggregated network or difficult to resolve when sample sizes are insufficient for reliable cohort-level network construction.

In addition to SSN analysis, MicroSSNet also supports the construction and analysis of aggregated networks and bipartite networks, and includes modules for modularity analysis, as well as the evaluation of global network properties such as robustness, vulnerability, and complexity. The current implementation of cross-domain interactions as bipartite networks in MicroSSNet was selected for its simplicity and interpretability, making it easier for users to visualize and analyze the relationships between different microbial domains. While this approach maintains computational stability and usability, it is recognized that a more integrated framework for cross-domain interactions could provide deeper biological insights. Therefore, we plan to explore extending cross-domain analysis to broader, integrated SSN or aggregated network frameworks in future versions of the package.

Simulation tests demonstrated that the ssPCC-based SSN method achieves reasonable control of type I error rates (≈ 5%) when the reference group includes at least 20 samples. When reference sample size falls below this threshold, particularly below 10, users should interpret SSN results with increased caution, as the likelihood of false positives increases and statistical reliability decreases. Sensitivity is also affected by baseline correlation strength and noise. This limitation indicates that in settings with weak baseline interactions or high data variability, ssPCC is unlikely to detect stable network changes. Therefore, results in such cases should be interpreted more cautiously, as it may be difficult to capture stable SSN features. It is recommended to appropriately retain low-prevalence SSN features rather than directly filtering them.

In real datasets, we used SSN features to identify multiple clusters of samples and detect many sample-specific edges that were not present in the aggregated network.

These additional edges may reflect individual-level interaction signals that are masked in cohort-averaged analyses. SSN features also showed potential for sample classification, yet this effect might not be solely due to interaction-level differences. Therefore, such features should be interpreted as exploratory indicators rather than direct evidence of biological interaction changes. Any results obtained from SSNs should be validated through longitudinal studies, independent cohorts, and follow-up mechanistic investigations to ensure robustness and biological relevance.

The application of MicroSSNet can be extended to a wide range of microbial ecosystems. Potential use cases include disease-associated human microbiomes (e.g., gut [28], oral, and skin), as well as plant rhizosphere and aquatic microbial communities. Future development will focus on expanding input compatibility to support multi-omics integration, including metatranscriptomics, metabolomics, and host-derived omics data. Additionally, we plan to incorporate more recent single-sample network construction methods—such as IENA [29], CSN [30], SSPGI [31], and SWEET [32]—to further enhance the analytical flexibility and methodological diversity of the framework.

Conclusions

We developed MicroSSNet, an R package that implements a ssPCC module and integrates the LIONESS method within a single-sample microbiome network analysis framework, alongside an aggregated-network workflow, addressing the current lack of accessible single-sample network tools—particularly for ssPCC-based single-sample analysis. We then used simulations to systematically evaluate its efficacy and limitations in microbiome network inference, followed by exploratory applications to real datasets to illustrate its potential applications in microbiome research.

In human gut metagenomes, SSNs suggested potential individual-specific changes in microbial associations that may be reproducible within small subgroups and attributable to inter-individual heterogeneity. Because SSNs quantify per-sample ΔPCC (reference-relative Pearson correlation change) rather than relying on cohort-averaged network topology, they can resolve interaction patterns that disappear in aggregated analyses—such as low-prevalence or opposing-direction changes—and localize which samples changed and in what way. The differences from aggregated network results reflect distinct estimation targets, rather than inconsistencies.

In the soil microbial 16S rRNA data, SSN features effectively distinguished samples under different environmental conditions and showed strong potential for classification and prediction. However, this discriminatory power may be partially driven by underlying abundance information. Therefore, when interaction patterns change without substantial shifts in abundance—or when confounding taxa with large abundance changes are removed—SSN-based distinctions may more accurately reflect true interaction-level differences. This finding suggests that microbial interaction changes hold potential as informative indicators when analyzing microbiome data.

Compared with existing microbial network analysis tools, the SSN framework implemented in MicroSSNet enables individualized network characterization. This offers exploratory potential for personalized microbiome research and for investigating dynamic ecosystem processes at higher resolution. However, SSN features should be viewed as preliminary indicators requiring further validation.

Methods

Data simulations

We designed a Monte Carlo simulation to evaluate the statistical power of the single-sample Pearson correlation change (ssPCC) method when the query sample and the reference group differ in (i) the correlation between two operational taxonomic units (OTUs), (ii) their mean abundances, and (iii) the reference group sample size.

The reference group (n samples) was generated from a bivariate lognormal distribution with correlation ρ_0, mean abundances μ_{ref}, and coefficients of variation (CV):

$$\ln\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$$

$$\mu_{\log} = \ln(\mu_{abu}) - \frac{1}{2}\ln\left(1 + C\mathrm{V}^2\right)$$

$$\sigma_{\log} = \sqrt{\ln\left(1 + C\mathrm{V}^2\right)}$$

A query sample (1 sample) was generated from a bivariate lognormal distribution with correlation ρ_q, mean abundance μ_q, and CV values potentially different from those of the reference group.

Datasets

We evaluated MicroSSNet via two publicly available datasets; no new sequencing data were generated for this study.

Soil 16S rRNA dataset. Amplicon sequences were obtained from the National Center for Biotechnology Information (NCBI) under BioProject PRJNA331185.

Human gut metagenomic dataset. Human metagenomic sequencing data were obtained from the Genome Sequence Archive for Human (GSA-Human) at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (NGDC/BI-G, CAS), under accession HRA003616 (publicly accessible at https://ngdc.cncb.ac.cn/gsa-human).

All the data were de-identified and publicly available; therefore, no IRB approval was required for this analysis.

Network construction

Aggregated network construction (correlation-based networks)

MicroSSNet supports multiple microbial correlation-based network inference methods, including SparCC, SPIEC-EASI, Pearson, and Spearman correlations, to address diverse research scenarios and data characteristics.

Data filtering: Prior to network construction, feature abundance matrices can be filtered according to experimental needs. Users can apply the top parameter to select high-abundance species on the basis of absolute or relative abundance thresholds and the pvl. threshold parameter to remove low-prevalence taxa.

Algorithm settings:

SparCC: This method employs sparse covariance estimation and multiple permutations (default 100) to infer robust sparse correlation matrices among microbial taxa.

SPIEC-EASI: This method utilizes sparse inverse covariance estimation (graphical lasso, glasso) to construct microbial co-occurrence networks, integrating the stability approach to regularization selection (StARS) for network robustness assessment.
Pearson/Spearman correlation: This method is suitable for assessing linear or nonlinear correlations among microbial taxa.
Matrix filtering: This tool provides a padj parameter for multiple-testing correction. It further employs thresholds such as r.threshold (default $r \geq 0.6$) and p.threshold (default $p \leq 0.05$) to retain significant and robust associations for the final network.

Aggregated network inference methods	
Pearson	Measures the linear relationship between two variables $r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \sum (y_i - \overline{y})^2}}$
Spearman	Measures the rank correlation between two variables $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$
SparCC	Methods based on sparsity assumption and simulation $\rho_{ij} \approx \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i) \cdot \mathrm{Var}(X_j)}}$
SPIEC-EASI	A method for inferring ecological association networks via sparse inverse covariance estimation $\hat{\Theta} = \arg\min_{\Theta} \left[\mathrm{tr}\,(S\Theta) - \log\det(\Theta) + \lambda \|\Theta\|_1 \right]$

Single-sample network construction

Single-sample networks can be inferred via two supported methods—LIONESS and ssPCC.

LIONESS: Linear interpolation theory is used to consider individual samples as changes in the aggregated network, thereby deriving single-sample network structures.
ssPCC: Employs a leave-one-out strategy, iteratively adds or removes each sample from the reference network, calculates the changes (ΔPCC), and uses Z tests to select significantly changed edges ($p \leq 0.05$), thus forming a single-sample network.

Single-sample network inference methods	
SSN	Differential correlation between reference network and reference network + sample of interest $\Delta PCC = PCC_{r+1} - PCC_r$ *With* $PCC(x_i, x_j) = \frac{Covariance(x_i, x_j)}{\sqrt{Variance(x_i) Variance(x_j)}}$
LIONESS	Linear interpolation $e_{i,j}^q = N\left(e_{i,j}^a - e_{i,j}^{a-q}\right) + e_{i,j}^{a-q}$ With the edge weight in, respectively the all-samples network (α) or the all-samples network without the sample of interest q (α-q) and N scaling factor inversely proportional to the total number of samples

Visualization and random network validation

Network visualization

Nodes can be colored on the basis of taxonomic annotations or modular results. Additionally, bipartite networks are provided to illustrate interactions across multiple domains (e.g., bacteria–fungi). Output files compatible with Cytoscape and Gephi are also generated to facilitate further editing and visualization.

Random network validation

A total of 100 Erdos–Renyi random networks serve as null models to statistically evaluate topological properties. Wilcoxon rank-sum tests were used to analyze network

robustness, vulnerability, and complexity. The tool supports random node removal (ranging from 5 to 95%) to simulate network responses under varying change scenarios.

Network property analysis

Aggregated network properties

Aggregated analyses assess global network properties, including node count, edge count, connectivity, average local clustering coefficient, and network diameter. Furthermore, it evaluates relative modularity, robustness, vulnerability and complexity.

Relative modularity (RM), which measures how modular a network is compared with the mean expected modularity, is more meaningful for comparing modular structures across different networks. In this study, RM was calculated as the ratio of the difference between the modularity of an empirical network and the mean modularity from the random networks over the mean modularity from the random networks.

$$RM = \frac{Q_{observed} - Q_{random}}{Q_{random}}$$

where $Q_{observed}$ is the modularity of the empirical network and where Q_{random} is the mean modularity from random networks.

Network robustness refers to the percentage of remaining species after a certain species becomes extinct. Extinction was simulated through randomly eliminating a certain percentage of nodes or a certain number of hubs:

$$\text{Robustness} = \frac{\sum_{i \neq j} b_j p_{ij}}{\sum_{i \neq j} b_j}$$

where b_j is the relative abundance of species j and P_{ij} is the Pearson correlation coefficient between species i and j. In our study, stability = 0 means that all the associations of species have been eliminated, and < 0 means that the symbiotic relationships between species i and other species are not enough to sustain their survival. In these cases, node i is thought to be extinct and was eliminated. This calculation continued until the stability value of all the species was > 0.

Vulnerability refers to how quickly biological or ecological consequences propagate across parts or entire networks and is characterized by the maximal vulnerability of nodes in the network:

$$\text{Vulnerability} = \max\left(\frac{E - E_i}{E}\right)$$

where E is the global network efficiency and E_i is the efficiency after eliminating node i and all its edges (Deng et al., 2012). Global network efficiency is determined by averaging the efficiency of all paired nodes and is calculated via the following equation, where $d_{(i,j)}$ is the number of edges between nodes i and j.

$$E = \frac{1}{n\,(n-1)} \sum_{i \neq j} \frac{1}{d\,(i,j)}$$

where $d_{(i,j)}$ is the number of edges in the shortest path between nodes i and j. Efficiency describes how fast information spreads within the network. In ecological networks, it

can provide information on how fast the consequences of biological/ecological events traverse parts or the entire network.

A comprehensive index was established to reflect the complexity of the microbial network and was calculated by averaging the standardized scores of the topological properties, including the number of nodes, number of edges, connectance, average degree, global clustering coefficient, average clustering coefficient, average neighborhood, centralization degree, eigenvector centralization, and cohesion, as follows:

$$\text{Complexity} = \frac{X_{\text{raw}} - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

where X_{raw}, X_{min}, and X_{max} represent the raw topological properties and the minimum and maximum values across all plots, respectively.

MicroSSNet also implements comparisons with random null models to reveal topological characteristics and environmental responsiveness.

Single-sample network properties

Single-sample network analysis focuses on individual-specific network characteristics:

Global properties: Connectivity, degree distribution, and diameter for individual sample networks.

Limma-based differential analysis of SSN node- and edge-weight features

For groupwise comparisons of single-sample network (SSN) features, we tested node weight and edge weight matrices via the empirical Bayes linear modeling framework in limma. The samples were aligned to metadata, and a no-intercept design was specified as ~ 0 + group, followed by all pairwise contrasts between group levels. For each contrast, we fitted linear models (lmFit), applied the contrast matrix (contrasts.fit), and obtained moderated statistics via empirical Bayes shrinkage (eBayes). P values across all features (all nodes or all edges) within a contrast were adjusted via the Benjamini–Hochberg procedure to control the within-contrast false discovery rate (FDR) at q = 0.05. Features were called up or down when they met both criteria: FDR-adjusted q < 0.05 and an effect-size threshold |log2FC| ≥ 1; otherwise, they were considered NS (not significant). When an optional log transform is used (i.e., log2(x + 1) on feature values prior to modeling), log2FC is interpretable as a fold-change (a threshold of 1 corresponds to a ≥ twofold). Without the log transform, logFC represents a difference in the original feature scale.

Supervised classification and ROC–AUC evaluation

We assessed the discriminative power of the SSN node weight features via five classifiers: random forest, SVM with RBF kernel, k-nearest neighbors, logistic regression, and Lasso-regularized logistic regression. The data were split into a stratified training set and a held-out test set. Within the training set, we performed stratified fivefold cross-validation for hyperparameter tuning, optimizing the ROC–AUC; any required preprocessing (centering/scaling) was fit within folds to prevent leakage. After tuning, each model was refitted on the full training set and evaluated once on the held-out test set. The primary endpoint was the test ROC–AUC computed from class probabilities (with the positive class prespecified); accuracy and ROC curves are reported for completeness. All models followed the same protocol.

Clustering of SSN interaction modes

We clustered samples based on the SSN edge-weight matrix. A cosine dissimilarity matrix was computed on the column vectors of the analysis matrix via proxy:dist (method = "cosine"). We then applied the PAM over K = 2–30. For each K, we fitted the PAM and computed the average silhouette width (ASW); the final number of clusters (best K) was chosen as the K that maximized the ASW.

Availability and requirements

Project name: MicroSSNet

Project home page: https://github.com/TangZecheng622/MicroSSNet

Operating system(s): Platform independent (tested on Windows and Linux; should work on any system supporting R)

Programming language: R

Other requirements: R (version 4.4.0 or higher); R packages data.table, dplyr, ggdist, ggplot2, ggpubr, ggsci, Hmisc, igraph, limma, mgm, networktools, qgraph, scales, SpiecEasi, tidyr, vegan, magrittr, methods, purrr, rlang, tibble; optional: lionessR (for selected functions)

License: MIT license (MIT + file LICENSE)

Any restrictions to use by non-academics: None

Abbreviations

SSN	Single-sample networks or individual-specific network
ssPCC	Single-sample network constructed based on single-sample Pearson correlation coefficient
LIONESS	Linear interpolation to obtain network estimates for single samples
ΔPCC	Reference-relative Pearson correlation change
SSN feaures	Features derived from SSNs, including SSN edge-weights and SSN node-weights.
SSN edge-weight	The change in Pearson correlation for a taxon pair after adding one query sample to the reference set; SSN edge-weight = ΔPCC
SSN node-weight	The sum of SSN edge-weights incident to a node

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-026-06444-w.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Acknowledgements
Not applicable.

Author contributions
All authors contributed to the study and analytical workflow. The initial idea and framework were conceived by ZZ and DZ. The pipeline was built and maintained by ZT. During pipeline development, XD, QG, PJ, and CT contributed to pipeline testing and its application in network analyses. JY, FZ, GS, and HY participated in discussions. FL, QD, and TL participated in the revision of Figure 3. The manuscript was written by ZT and revised by ZZ and ZY. All authors reviewed and approved the final manuscript.

Funding
This study was supported by the Second Tibetan Plateau Scientific Expedition and Research (STEP) Program (2019QZKK0503, 2021QZKK0103), the Yunnan Province Xing Dian Talent Support Program (Yunling Scholar), the Central Guidance on Local Science and Technology Development Fund of Yunnan Province (202407AA110009),

the Xizang Autonomous Region Science and Technology Program Project (XZ202401YD0012), the National Natural Science Foundation of China (U2002206, 31970571) and the Yunnan University Medical Research Foundation (YDYXJJ-2024-0009).

Data availability

The soil 16S rRNA amplicon sequences analysed in this study are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under BioProject accession PRJNA331185 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA331185).The human gut metagenomic sequencing data used in this study have been deposited in the Genome Sequence Archive for Human (GSA-Human) at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under accession HRA003616 (https://ngdc.cncb.ac.cn/gsa-human). These human data are under controlled access for ethical and privacy reasons and were used under license; they are available from the repository and from the corresponding author on reasonable request and subject to approval by GSA-Human. All other data analysed during the current study are included in this published article and its supplementary information files. The R package MicroSSNet used for network construction and analysis is freely available on GitHub (https://github.com/TangZecheng622/MicroSSNet).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 19 November 2025 / Accepted: 1 April 2026

Published online: 13 April 2026

References

1. Liu YX, Qin Y, Chen T, Lu M, Qian X, Guo X, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell. 2021;12(5):315–30.
2. Strogatz SH. Exploring complex networks. Nature. 2001;410(6825):268–76.
3. Moody J. Race, school integration, and friendship segregation in America. Am J Sociol. 2001;107(3):679–716.
4. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett. 2001;86(14):3200–3.
5. Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A, Caldarelli G. The statistical physics of real-world networks. Nat Rev Phys. 2019;1(1):58–71.
6. Corel E, Lopez P, Méheust R, Bapteste E. Network-thinking: graphs to analyze microbial complexity and evolution. Trends Microbiol. 2016;24(3):224–37.
7. Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: a network perspective. Trends Microbiol. 2017;25(3):217–28.
8. Du Y, Yang Y, Wu S, Gao X, He X, Dong S. Core microbes regulate plant-soil resilience by maintaining network resilience during long-term restoration of alpine grasslands. Nat Commun. 2025;16(1):3116.
9. Yuan MM, Guo X, Wu L, Zhang Y, Xiao N, Ning D, et al. Climate warming enhances microbial network complexity and stability. Nat Clim Change. 2021;11(4):343–8.
10. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687.
11. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol. 2015;11(5):e1004226.
12. Liu X, Wang Y, Ji H, Aihara K, Chen L. Personalized characterization of diseases using sample-specific networks. Nucleic Acids Res. 2016;44(22):e164–e164.
13. Kuijjer ML, Tung MG, Yuan G, Quackenbush J, Glass K. Estimating sample-specific regulatory networks. iScience. 2019;14:226–40.
14. De Marzio M, Glass K, Kuijjer ML. Single-sample network modeling on omics data. BMC Biol. 2023;21(1):296.
15. Ponce-Cusi R, López-Sánchez P, Maracaja-Coutinho V, Espinal-Enríquez J. Single-sample networks reveal intra-cytoband co-expression hotspots in breast cancer subtypes. IJMS. 2024;25(22):12163.
16. Wei L, Xin Y, Pu M, Zhang Y. Patient-specific analysis of co-expression to measure biological network rewiring in individuals. Life Sci Alliance. 2024;7(2):e202302253.
17. Li S, Schmid KT, De Vries DH, Korshevniuk M, Losert C, Oelen R, et al. Identification of genetic variants that impact gene co-expression relationships using large-scale single-cell data. Genome Biol. 2023;24(1):80.
18. Su Q, Zhuang DH, Li YC, Chen Y, Wang XY, Ge MX, et al. Gut microbiota contributes to high-altitude hypoxia acclimatization of human populations. Genome Biol. 2024;25(1):232.
19. Jahagirdar S, Saccenti E. Evaluation of single sample network inference methods for metabolomics-based systems medicine. J Proteome Res. 2021;20(1):932–49.
20. Guimerà R, Nunes Amaral LA. Functional cartography of complex metabolic networks. Nature. 2005;433(7028):895–900.
21. Banerjee S, Schlaeppi K, Van Der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. Nat Rev Microbiol. 2018;16(9):567–76.
22. Montesinos-Navarro A, Hiraldo F, Tella JL, Blanco G. Network structure embracing mutualism–antagonism continuums increases community robustness. Nat Ecol Evol. 2017;1(11):1661–9.

23. Herren CM, McMahon KD. Cohesion: a method for quantifying the connectivity of microbial communities. ISME J. 2017;11(11):2426–38.
24. Zhang C, Lei S, Wu H, Liao L, Wang X, Zhang L, et al. Simplified microbial network reduced microbial structure stability and soil functionality in alpine grassland along a natural aridity gradient. Soil Biol Biochem. 2024;191:109366.
25. Feng K, Peng X, Zhang Z, Gu S, He Q, Shen W, et al. iNAP: an integrated network analysis pipeline for microbiome studies. iMeta. 2022;1(2):e13.
26. Peschel S, Müller CL, Von Mutius E, Boulesteix AL, Depner M. NetCoMi: network construction and comparison for microbiome data in R. Brief Bioinform. 2021;22(4):bbaa290.
27. Wen T, Xie P, Yang S, Niu G, Liu X, Ding Z, et al. ggClusterNet: an R package for microbiome network analysis and modularity-based multiple network layouts. iMeta. 2022;1(3):e32.
28. Cheng R, Wang L, Le S, Yang Y, Zhao C, Zhang X, et al. A randomized controlled trial for response of microbiome network to exercise and diet intervention in patients with nonalcoholic fatty liver disease. Nat Commun. 2022;13(1):2555.
29. Yu X, Zhang J, Sun S, Zhou X, Zeng T, Chen L. Individual-specific edge-network analysis for disease prediction. Nucleic Acids Res. 2017;45(20):e170.
30. Dai H, Li L, Zeng T, Chen L. Cell-specific network constructed by single-cell RNA sequencing data. Nucleic Acids Res. 2019;47(11):e62.
31. Chen Y, Gu Y, Hu Z, Sun X. Sample-specific perturbation of gene interactions identifies breast cancer subtypes. Brief Bioinform. 2021;22(4):bbaa268.
32. Chen HH, Hsueh CW, Lee CH, Hao TY, Tu TY, Chang LY, et al. SWEET: a single-sample network inference method for deciphering individual features in disease. Brief Bioinform. 2023;24(2):bbad032.

Publisher's Note