

REVIEW

Open Access



Networks as tools for defining emergent properties of microbiomes and their stability

Kacie T. Kajihara^{1*} and Nicole A. Hynson¹

Abstract

The potential promise of the microbiome to ameliorate a wide range of societal and ecological challenges, from disease prevention and treatment to the restoration of entire ecosystems, hinges not only on microbiome engineering but also on the stability of beneficial microbiomes. Yet the properties of microbiome stability remain elusive and challenging to discern due to the complexity of interactions and often intractable diversity within these communities of bacteria, archaea, fungi, and other microeukaryotes. Networks are powerful tools for the study of complex microbiomes, with the potential to elucidate structural patterns of stable communities and generate testable hypotheses for experimental validation. However, the implementation of these analyses introduces a cascade of dichotomies and decision trees due to the lack of consensus on best practices. Here, we provide a road map for network-based microbiome studies with an emphasis on discerning properties of stability. We identify important considerations for data preparation, network construction, and interpretation of network properties. We also highlight remaining limitations and outstanding needs for this field. This review also serves to clarify the varying schools of thought on the application of network theory for microbiome studies and to identify practices that enhance the reproducibility and validity of future work.

Keywords Co-occurrence networks, Microbiome, Stability, Fungi, Bacteria, Data science, Multi-omics

Introduction

With the advancement of new technologies applied to the field of biology, a new perspective has emerged of our place in the biosphere. We now recognize that microbes underpin the function of all sectors of the planet, and we, in fact, live in a microbial world [1]. From mountains to oceans and from the bottom to the top of all food webs, ecosystem and organismal health rely upon the microbiome (communities of bacteria, archaea, fungi, and other microeukaryotes). Thus, harnessing the beneficial properties of microbiomes that support the health of hosts and habitats is an increasingly important pursuit as we seek

more personalized medicine, as global change continues to disrupt ecosystems, and as access to nutritious food and clean water is challenging even among some of the wealthiest nations. However, fundamental to the success of efforts such as microbiome engineering is understanding microbiome stability, the ability of a microbiome to resist or recover from disturbances. Stable microbiomes are commonly assessed via DNA sequencing to identify compositional or functional traits that persist in the face of disturbances, ranging from oral antibiotics in humans affecting the gut microbiome [2] to the ability of corals to withstand thermal stress [3]. However, these methods do not necessarily consider the diversity and complexity of interactions that characterize microbial life and potentially foster community stability [4]. To address this, co-occurrence networks have emerged by way of graph theory as a way to model communities in the context of their potential interactions [5]. Co-occurrence analyses have long been used in community ecology to study

*Correspondence:

Kacie T. Kajihara
kaciekaj@hawaii.edu

¹ Pacific Biosciences Research Center, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA



the architecture of food webs, and many of the concepts used to study microbiome co-occurrence networks originate from these works [6–9]. Given that no microbe acts in isolation, network stability research offers pertinent insights to guide future research for microbiome design and engineering [10, 11].

The co-occurrence networks discussed here are composed of nodes and edges, where nodes can represent microbial taxa, genes, metabolites, or other compositional properties of the microbiome and edges indicate statistically significant relationships between them. In brief, networks are constructed via pairwise comparisons to determine whether there is a significant likelihood that the given microbes tend to co-occur or trend towards mutual exclusion. Edges can be classified as either positive or negative and by the strength of the predicted association. Negative edges potentially represent relationships such as competition or predation, and positive edges could indicate mutualism or commensalism, and experimental co-cultivation has previously supported the fidelity of some edges inferred in silico [12–14]. Each resulting network represents a snapshot of the microbiome that can be used to characterize potential interaction patterns and to predict stability and be comparable to other networks similarly constructed [15]. However, the path to implementing network construction and downstream analyses like stability is littered with considerations, and few standards currently exist among the research community. Beyond this, the network topological metrics used as indicators of stability (e.g., degree, connectivity, or clustering coefficient) often co-correlate [16] or are associated with multiple plausible but conflicting interpretations, further challenging the interpretation of these analyses.

The following comprises a user's guide for generating co-occurrence networks and implementing downstream analyses such as the assessment of stability (Fig. 1). From raw DNA sequencing data to network stability analysis, we outline considerations for data curation and software selection, as well as collate network topological metrics and network properties used to study stability. We highlight areas of congruence and incongruence on which metrics and properties indicate stability, review the remaining challenges in this area of study, and share suggestions for future work.

Stage 1: Data preparation

Networks can be used to ask a range of biological and ecological questions, offering a window, for example, into the predicted importance of certain taxa in an ecosystem [12], potential drivers of community assembly [17], or the robustness of the entire system across disturbance gradients [18, 19]. To ask such questions, the

data must first be curated to ensure that the inferred network is biologically and ecologically relevant to the study system and questions. This involves meeting the statistical expectations for inferring the network, minimizing the likelihood of spurious connections, and constructing networks in a manner that ensures reliable comparisons between them.

Taxonomic agglomeration

Many microbiome network studies cluster sequences into operational taxonomic units (OTUs) at 97% sequence similarity (e.g., [18–20]), or amplicon sequence variants (ASVs) delineated by single-nucleotide differences [21–25], while some examine networks among higher taxonomic groupings such as genera or classes [26, 27]. The level at which microbial sequencing reads are binned and the taxonomic grouping affects what each node represents and what is indicated by edges between nodes. For example, cross-domain networking among classes of bacteria and fungi may represent higher-level ecological and biological interactions, whereas networks built from barcode loci ASVs may indicate distinct patterns among more closely related microbes. Researchers should consider what level of taxonomic resolution and sequence agglomeration is appropriate for their questions of interest [28–30]. It is also possible to cluster ASVs into OTUs while maintaining ASV seed sequences, which enables conversions between groupings as desired (as in [31]).

Binning taxa into 97% similarity OTUs or higher taxonomic groupings instead of ASVs also brings the benefit of reduction in dataset size and zero inflation (discussed in the next section). Feasible richness levels for network construction generally range in the order of hundreds to thousands of taxa (e.g., [24, 32]). The larger the dataset, in terms of taxa and/or samples, the more computational resources and time will be needed to build a network. Certain software have workarounds for batching out iterations or using lossy processes to speed up runtimes [33], though generally the pairwise nature of co-occurrence calculations makes network construction less amenable to parallelization.

Data filtering

Microbiome data are often zero-inflated, which can cause erroneous predictions and lowered precision in correlation-based network methods, such as Spearman's and Pearson's correlations, *SparCC*, and the maximal information coefficient (MIC) [34]. Previous recommendations include taxa filtering to at least a 20% prevalence threshold to ensure that the interactions represented in the network are most likely to be biologically real and meaningful, though these cutoffs remain arbitrary and a subject of debate [35, 36], ranging anywhere from 10%

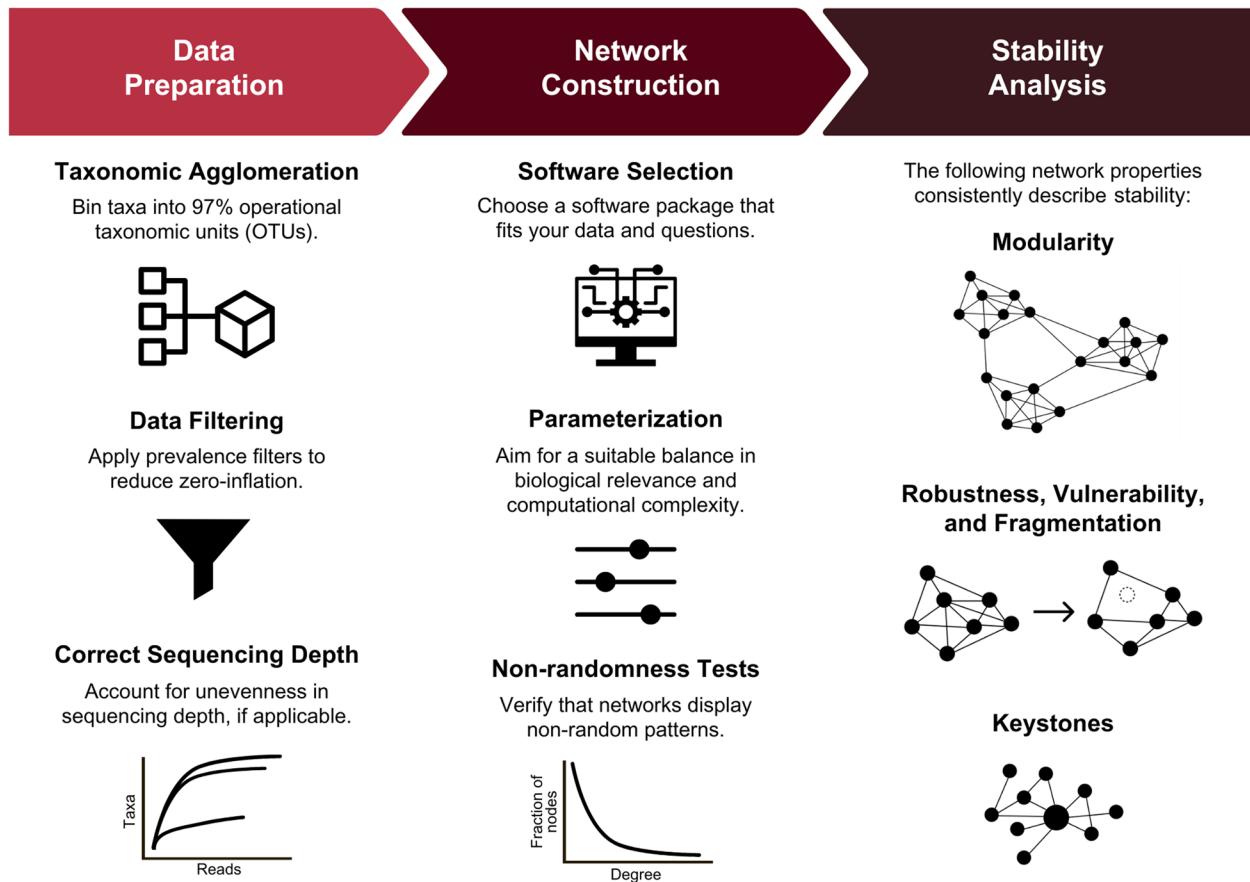


Fig. 1 Example workflow for microbial network stability analyses. To prepare amplicon sequencing data for network assembly, data are commonly center-log ratio transformed, clustered into 97% operational taxonomic units (OTUs; or kept at 100% amplicon sequence variants, ASVs), prevalence filtered, and controlled for uneven sequencing and sampling bias (via rarefaction or other normalization). General data preparation steps apply to other data types as well. To construct networks, researchers should spend time comparing the strengths and weaknesses of various co-occurrence network software and then fine-tune their parameters. After networks are built, we suggest that they be tested for nonrandomness. Analyses should center on network modularity, robustness, vulnerability, fragmentation, and the presence of keystone nodes, as these have the most consistent interpretations with network stability

[19, 21] to >60% [37]. Depending on the question, more stringent filters may be appropriate, for example, network comparisons made within-host on human skin and lung microbiomes discarded taxa present in less than 33% of samples [13], while a comparison of soil microbiomes across different environments applied a 10% prevalence filter [19]. Data may also be pre-filtered to exclude pairings that cannot be reliably tested due to zero inflation [36], as a high incidence of zeroes not only complicates the detection of negative associations but can also artificially inflate positive correlations [35]. These filtering steps inevitably cull members of the rare biosphere, which has been touted in various systems for its ecological importance [38]. The spectrum of prevalence filtering thresholds from low to high generally represents a trade-off between inclusivity and accuracy. Data filtering will

also reduce the dataset and subsequent computational burden of network construction. Here, as elsewhere in this guide, researchers are urged to consider these trade-offs in deciding what parameters make the most sense for their questions and systems.

Sequencing or sampling bias

Rarefaction, the process of randomly subsampling to a set number of sequencing reads across samples, is commonly used in microbial network studies to address uneven sequencing depth [22, 24, 39]. There is strong debate on the appropriateness of rarefaction with regard to diversity analyses [40, 41], but the effects of rarefying in networks vary by data association algorithms used during network construction [42]. A comparison of correlation-based methods found that rarefying caused tools

like *CoNet* and Spearman's and Pearson's correlations to suffer a decrease in precision, but that others such as *SparCC*, Bray–Curtis dissimilarity, and MIC could still reliably infer interactions [34]. The robustness of other network inference methods to rarefaction, such as graphical probabilistic models, has yet to be benchmarked.

To address sampling bias, sampling intensity may be held constant across treatments. For example, researchers may standardize samples to the lowest common replicate number across plots or individuals or use frequency distributions of reads by taxon or sample to maintain a standard proportion of the sampling effort [24]. As with prevalence filtering, these normalization methods will likely remove rare taxa.

Compositional data bias

Microbiome data are compositional, in that counts represent proportions of taxa relative to the total number of sequencing reads in a sample, and not absolute abundances [43]. Observations in microbiome data are thus not independent, which breaks the assumptions of traditional correlation analysis and can result in a network with many false-positive signals [44, 45]. A common solution is to use the center-log ratio transformation to remove dependencies between proportions [45], either applied to the entire data table (as in *SPIEC-EASI*, a graphical method [33]) or to pairs of taxa (as in *SparCC*, a correlation method [44]). Alternatively, software using Dirichlet multinomial models aims to directly account for compositional data [46, 47].

Inter-kingdom data

Specific data pre-processing steps are needed to create networks involving multiple domains of life, such as bacteria, archaea, fungi, and other microeukaryotes [13, 48–52], as their compositions are specific to each dataset and not the concatenated whole [53]. The *SPIEC-EASI* package automatically accommodates inter-kingdom data by independently transforming datasets with the center-log ratio transformation, which satisfies the equations to generate the inverse covariance matrix [13, 33]. Datasets used to make correlation-based networks should also be transformed independently before concatenation to avoid introducing bias and spurious edges [53].

Stage 2: Network construction

Software selection

At the time of writing, a broad suite of software is available for network generation, with each option offering different underlying models suitable for different needs [54]. Correlation-based methods are commonly used to infer links in microbial network construction [21, 24] but can be prone to issues involving compositionality

and arbitrary significance cutoffs [54]. To promote statistical integrity in a correlation-based network, users may involve null models [55] or random matrix theory-imposed correlation cutoffs to separate organized information from noise [56]. The package *CoNet* employs an ensemble approach to reduce false positives, preserving only those edges supported across multiple correlation, similarity, or dissimilarity methods [57]. When using correlation-based methods, a false discovery rate should be used to correct for multiple comparisons (e.g., [58]).

Network software using probabilistic graphical models have also emerged to mediate several of the aforementioned challenges, along with the issue of indirect edges, or edges that arise due to shared responses to other taxa instead of a direct interaction. Graphical methods infer edges based on conditional dependencies, where an association is drawn between two nodes when there is a linear relationship between them, given all other nodes in the network [33], and an edge will not be drawn if that relationship can be explained by external taxa. Calculating networks in this manner often requires additional computational power [54]. One well-known algorithm in this family, *SPIEC-EASI*, was also designed to address the high dimensionality of microbiome data, where samples are often far outnumbered by taxa, which can otherwise result in overfitting [33].

Indirect edges may also result from common responses to environmental effects, such as pH, water availability, or mineral levels [42, 59]. Some packages are able to accommodate environmental data [57, 60], while other studies include environmental variables in their correlation analyses, either to observe how the environment structures co-occurrence patterns or to cancel them out [61–63]. Other suggestions to account for environmental effects include generating separate networks in instances where environmental data are expected to vary, for example, across water depths, as indirect edges should be less prevalent within a given environment [64]. It may also be possible to discern certain indirect edges resulting from geographic or environmental variation using phylogenetic distances to assess whether inferred edges are consistent with habitat filtering, dispersal limitation, or biotic interactions [65].

For an in-depth review on contemporary co-occurrence network software and their advantages and pitfalls, including packages suitable for metagenomic data, see [54]. Overall, difficulties remain across the gamut of network software in accounting for computational complexity and zero-inflated data [34, 54].

Software parameterization

The parameters of a set of successfully constructed networks should have been fine-tuned so that the networks

achieve a suitable balance in interpretability, biological relevance, and computational complexity, and are comparable among treatments. This may include choosing between Spearman's or Pearson's correlations, defining cutoffs for what is deemed a significant interaction, regulating network sparsity, or setting how many iterations of the calculation to perform [33, 56]. Different datasets will likely call for different parameters, but the selected values should always be biologically justifiable. For example, microbiome studies often favor Spearman's correlations over Pearson's, as the latter assumes normally distributed data with linear relationships, to which microbial data may not conform [55], and Pearson's correlations may be less sensitive to detecting negative associations [36]. Networks resulting from the same type of model can also display a great deal of variation due to the freedom of choice inherent in parameterization [35]; thus, if different networks are to be compared, they should be constructed using identical parameters.

Nonrandomness in networks

Finally, it is useful to confirm that the resulting networks differ from random networks, which have Poisson degree distributions [66, 67]. Networks are sometimes considered non-random if their topology differs from a distribution of topological values derived from random networks with the same node and edge numbers [56]. Often, studies have called networks non-random if they follow a power-law degree distribution (i.e., they are scale-free) and are considered small-world, or formed of more tightly knit clusters connected by relatively short paths [68]. However, nonrandom networks need only be non-Poisson and can take many different forms. Biological networks may also organize themselves similarly to

random networks [69, 70], but are generally expected to be non-random [68]. Confirmatory studies with empirical and synthetic datasets are warranted to better define statistical expectations for non-random network structure in microbiomes especially, as these may differ from the macro-organismal, gene, and protein networks from which many of these tests are derived [68, 71].

Stage 3: Network topology implies properties related to stability

After networks are constructed, their topology and other properties can be used to assess factors such as stability, though their interpretations can be clouded by different, even contradictory explanations for what values imply a stable network. The following sections describe various network metrics (Fig. 2) and how microbiome studies have positioned them in the context of stability (Table 1).

Network topological metrics and properties

Modularity

Modularity is a measurement of network partitioning into distinct and highly connected subcommunities or modules. Members of modules are thought to associate with one another due to shared functions, strategies, or environmental preferences [19, 61] and are only weakly connected to nodes in other modules. For example, in plant-pollinator networks, modules circumscribe coevolutionary units of taxonomically related species with convergent traits [80]. In microbiomes, modules may imply niche partitioning [64] and have been used to study habitat preferences [81]. Modules are desirable in terms of network stability because local disturbances are more likely to be contained within the module and not propagate. This idea of compartmentalization imparting

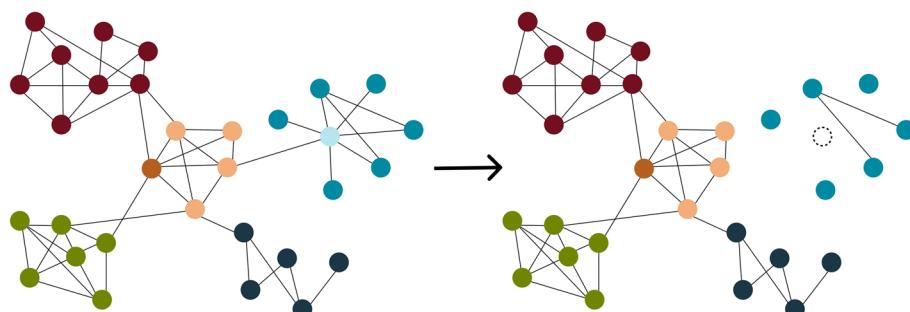


Fig. 2 Visualization of network metrics commonly used to assess stability. The hypothetical network is partitioned into five modules, each distinguished by color. Nodes are much more connected within-module than among modules. The dark red module has a higher incidence of edges, indicating a relatively higher connectivity and linkage density compared to the navy blue module. Modules in dark red, orange, and green have higher clustering coefficients than modules in turquoise and navy blue, as they cluster into multiple "triangles." The node in dark orange is an example keystone, as it has a high degree (connections to other nodes) and high betweenness centrality (it frequently lies on the shortest path between all pairs of nodes) and low relative prevalence (not shown). The node in light blue has relatively high vulnerability, in that its removal has a large effect on network efficiency or how traversable the network is. The loss of the vulnerable light blue node also leads to fragmentation of the turquoise module and a decrease in overall network robustness

Table 1 Network topological metrics and network properties used in microbial network stability studies, with references. The majority of the listed metrics can be calculated in the igraph package [72]. For calculating network robustness, we suggest the packages brainGraph [73] or NetSwan [<https://cran.r-project.org/web/packages/NetSwan/index.html>]

Metric	Definition	High value = stable	Low value = stable
Average path length	The average shortest path (number of edges) between all pairs of nodes	[23, 26]	[39]
Betweenness centrality	Of a node, frequently occurring on the shortest path between all pairs of nodes	[23]	[20, 24, 74]
Connectivity	A measurement of how many connections are formed between nodes in a network	[18, 26]	[20, 25, 39, 58, 75]
Clustering coefficient/transitivity	The degree to which nodes in a network cluster together	[26, 39, 71, 76]	[20, 58]
Degree	The number of edges connected to a node	[76, 77]	[20, 58, 75]
Linkage density	The ratio of realized network edges to the total possible number of edges	[18, 24]	[58, 75]
Modularity	The partitioning within network communities into distinct and highly connected subcommunities	[18–20, 23–25, 39, 58, 77–79]	[75]
Network size	The total numbers of nodes in a network	[24, 39, 76]	
Vulnerability	The maximum impact of a node's removal on network efficiency		[39, 77]
Cohesion	A measurement of the abundance-weighted positive and negative correlations in a network, standardized by a null model. Or more simply, the ratio of negative to positive edges	[18, 19, 23, 24, 77, 79]	[20, 21, 26, 39, 48, 58, 75]
Complexity	Various definitions (see Table 2)	[18, 19, 21, 26, 39, 77]	[25, 58, 75]
Fragmentation	The number of disconnected subgraphs network divided by the total number of nodes, upon stepwise node removal	[22, 24]	
Inter-kingdom associations	Networks involving interactions between microorganisms from different kingdoms (e.g., fungi with bacteria)	[13, 23]	[48]
Keystone taxa	A taxon with disproportionate importance on ecosystem function and stability in relation to its biomass. In networks, nodes with high betweenness centrality and degree or high connectivity within and among modules	[24, 39, 39, 76, 77, 79]	
Robustness	Networks able to resist collapse upon targeted node removal	[13, 22, 23, 25, 39, 48, 58, 75–77]	

stability is echoed in food web theory [82–85]. Other than one study that found stable networks to possess fewer modules [75], the literature is in strong agreement that high modularity indicates increased stability [18–20, 23–25, 39, 58, 77–79], though all of these studies were system-specific, with the majority focused on soil.

Modularity is considered a component of network complexity [18, 77, 77] and is sometimes itself a criterion for measuring stability [19, 58, 79]. There are many ways to delineate modules, and community detection remains an active area of research in network science [86, 87]. Commonly used algorithms in microbiome analysis include fast greedy clustering [88], random walks [88], and simulated annealing [89], ordered approximately by speed (high to low) and potential precision (low to high). See [11] and [90] for more in-depth comparisons of modularity algorithms in microbiome networks.

Betweenness centrality

Centrality is a metric that positions the contributions of individual nodes to network structure. Along with degree centrality (often reported simply as degree, see below), betweenness centrality is one of the most commonly used metrics in microbiome network studies. A node has high betweenness centrality if it frequently occurs on the shortest path between all pairs of nodes. Betweenness centrality has been used to computationally identify putative microbial keystone taxa [13, 91]. Nodes with high betweenness centrality have been termed “gatekeepers,” as their loss disproportionately affects the fragmentation of the network [24, 74]. These nodes may serve important roles in bridging network modules or distinct subcommunities [52].

Betweenness centrality can also be assessed as a global network property, for example, as the average or maximum betweenness centrality across all nodes [20, 39].

Low global betweenness centrality is often associated with higher stability [20, 24], although one study comparing networks across anthropogenic and natural disturbances found the opposite [23]. Thus, the relationship between betweenness centrality and network stability tends to be negative but is yet unresolved.

Connectivity

Network connectivity refers to the degree of connections formed between nodes in a network. Although this is a specific metric that can be calculated for individual vertices [72], studies often use this term while referring to other metrics such as degree (the number of connections to a node), linkage density (the proportion of realized links relative to the total possible number of links), or the total number of links (edge number). These metrics are described in further detail below. A more connected network may have a higher incidence of generalist species with broader niche preferences and potentially broader interaction breadths; for example, in plant-animal interaction networks, generalists are more densely linked [92].

To date, studies have not reached a consensus on the role of connectivity in stability. Some have postulated that high connectivity is stabilizing because it offers redundancy, where a higher density of connections between nodes and network compartments compensates for the loss of certain edges in the event of disturbance [26]. Connectivity has been described as increasing stability because a higher density of links may increase network complexity, which has been tied to stability [18]. Others argue that a high level of connectivity makes a system less stable since it is more vulnerable to cascade effects, where disturbances may propagate more easily through the network [39]. Connectivity may also reduce stability more indirectly, by decreasing the specificity of links (i.e., a highly connected node exhibits more “generalist” characteristics), which in turn reduces modularity [25, 92].

Degree

Node degree, or degree centrality, measures the number of edges connected to a node and is either taken as is or normalized by the total number of connections in the network. When the average node degree is used as a proxy for connectivity, networks with low degrees are usually associated with stability [20, 58, 75]. Degree can also be used to identify important nodes, in which a high degree indicates a hub and potential keystone taxon, and a higher incidence of keystones may indicate increased global network stability [76, 77]. High average degree may also indicate high network complexity [25], but various interpretations of stability with complexity abound (see “Complexity” section).

Network size

The size of a network is measured by the total number of nodes. A large network may increase its resistance to perturbation [76], and in general, this metric is seen in association with stable networks [24, 39]. However, the level of network stability is predetermined by other metrics such as robustness (i.e., network size itself does not suggest stability).

Clustering coefficient or transitivity

Clustering coefficient, also known as transitivity, measures the degree to which nodes’ neighbors are connected with one another, which indicates the communicability of the network. In microbiomes, higher clustering coefficients have been speculated to suggest cross-feeding relationships [92], especially when predicted associations are positive [62]. A high clustering coefficient coupled with low path length suggests a small-world structured network [77], which is efficient but potentially less stable due to vulnerability to node loss [39]. Indeed, increased efficiency may imply that these networks are less error-tolerant, making them more vulnerable to extinction events [68, 93]. High clustering coefficients have been associated with increasing levels of network degradation and the reduction of modularity, which by extension reduces stability [58].

High clustering coefficients also suggest a more compact/aggregated co-occurrence pattern, which has been seen as stabilizing due to more efficient resource transfer [76]. Similarly, [26] found compactness to be a feature of their stable networks.

Average path length

The average path length is another measure of network efficiency and is measured as the average shortest path between all pairs of nodes. The interpretation of path length with stability depends on whether studies position efficiency as enhancing or diminishing stability. A smaller average path length indicates a more compact network, which has been associated with more efficient resource sharing and thus higher stability [26], as well as the opposite, with increased vulnerability to node losses [39]. Higher average path lengths have also been associated with stable sites compared to disturbed sites [23].

Linkage density

Linkage density, also known as graph density or network density, is the ratio of realized edges to the total possible number of edges. Higher linkage density may increase the complexity of a network, increasing its stability [18]. However, May’s 1972 theorem indicates the opposite [83], where systems become less stable as complexity increases; a microbiome network study across

a permafrost degradation gradient supports this [58]. Both schools of thought originate from food web theory [9], and both are found in microbiome networks, where lower [75] and higher [24] levels of density are said to describe stable networks. When linkage density is weighted by association strength, it is referred to as connectance (not to be confused with connectivity), and increased connectance has been associated with increased stability [18, 39].

Vulnerability

Vulnerability is calculated as how strongly a node contributes to the global efficiency of the network, where a node's vulnerability is the loss in network efficiency when the node and all of its edges are removed [16]. The maximum node vulnerability represents the vulnerability of the entire network. Low vulnerability indicates a more stable network [39, 77]. Given that vulnerability is calculated on the basis of network efficiency (i.e., the compactness of the network, see “Average path length” section), this viewpoint also assumes a positive relationship between efficiency and stability.

Complexity

The complexity-stability debate stems from decades earlier in the food web literature, where MacArthur [94] purported that an increased number of species and links increased stability, while May [83] found the opposite. Beyond degree and connectivity, network complexity has been described in many ways over the years, from combinations of topological metrics [18, 25, 39, 75, 77] to the diversity in interaction strength [18, 77] or interaction sign [19, 23, 77, 79].

As a result, there is no single, accepted way to define a complex network (Table 2), but other than a few studies that found stable networks to be less complex [25, 58, 75], the literature has more or less converged on the viewpoint that increased complexity confers stability in microbiome networks [18, 19, 21, 26, 39, 77], with support in macroecological food webs [84, 95–97]. See [9] for an in-depth review of complexity and stability in ecological networks.

Robustness

Robust networks are those that are able to resist rapid collapse in the face of disturbance. It is one of the few metrics with strong agreement in the literature, where greater network robustness equates to greater stability [13, 22, 23, 25, 39, 48, 58, 75–77].

Disturbances are imposed computationally via targeted node removal (usually in order of decreasing node importance, or randomly), and the remaining largest structure in the network is divided by its starting size

Table 2 Network metrics and properties used to indicate complexity

Metric	References
Cohesion	[19, 77]
Connectance	[18, 39]
Connectivity	[18, 39, 75]
Degree	[21, 25]
Evenness of association strength	[18]
Linkage density	[18, 58]
Modularity	[18, 25]
Network size	[25, 39, 75]
Number of kestones	[39]
Ratio of edges to nodes	[26]

after each node removal [98]. Larger proportions are considered more robust, as the network has maintained a greater extent of its connections despite the loss of highly connected nodes. This method is likened to the idea of extinction cascades, where the loss of one species leads to the loss of others dependent on it [97].

Robustness is alternatively measured by “natural connectivity,” a method also involving targeted node removal [99]. Both forms of robustness are generally plotted as a curve tracking values of network “wholeness” (e.g., the fractional size of the largest remaining network component; *y*-axis) over the number of nodes removed, ranging from zero to the total number of nodes in the network (*x*-axis). A larger area under the curve [13, 48] or lower absolute value of the slope [23] indicates greater robustness. To derive a single robustness value for a network, some studies have used the number of remaining connections after removing a certain proportion of nodes, for example, 50% of nodes [77] or five module hubs [22, 39].

Fragmentation

Fragmentation is calculated by dividing the number of disconnected subgraphs in a network (i.e., a network whose nodes and edges are subsets of a larger graph) by the total number of nodes following stepwise node removal [100]. Lower fragmentation values are indicative of greater stability [22, 24]. The loss of potential “gatekeeper” nodes with high betweenness centrality can lead to greater fragmentation of networks [100].

Cohesion (negative-to-positive interaction ratio)

Though many of the aforementioned metrics rely on non-negative edges to calculate, positive and negative edge weights can be incorporated when assessing cohesion. Cohesion measures the abundance-weighted positive and negative correlations in a network, standardized by

a null model [101], though sometimes studies will calculate the ratio of negative-to-positive interactions alone to assess stability [23, 24]. A higher ratio of predicted negative interactions is thought to impart stability to a network because such a network is better able to contain the negative effects of disturbances [102]. Positive interactions can lead to strong positive feedback loops, in which the loss of one species leads to the mutual downfall of all others with which it is linked. Approximately half of surveyed studies using this metric found a higher rate of negative interactions to be indicative of stable networks [18, 19, 23, 24, 77, 79]. This is also consistent with the stress-gradient hypothesis [103], where positive interactions increase in adverse conditions thought to destabilize communities. In microbiomes, this effect may occur via a proliferation of stress-tolerant species and/or facilitative associations [19].

Other studies find a greater incidence of positive interactions to either underpin [21, 26, 39, 48, 75] or describe [20, 58] more stable networks. Under this paradigm, greater levels of cooperation may increase the overall efficiency of resource transfer and imply stable coexistence across species [21, 48].

Keystone taxa

Keystone taxa were those originally hypothesized by Paine [104] to be disproportionately important for ecosystem function and stability in relation to their prevalence or biomass. Unlike in macro-organismal systems, where a potential keystone organism may be physically removed or naturally absent, the keystone phenomenon is challenging to directly apply and observe in microbial systems due to their often diverse and ephemeral nature. Instead, networks have been adopted to identify potential keystone taxa, or highly important nodes, which are then removed computationally as a test of their importance for maintaining network structure. To date, the literature is in agreement that keystones have a positive relationship with stability [24, 26, 39, 75, 77, 79].

Keystones are thought to be important for network stability because they uphold important structures in the network. Various methods exist to identify putative keystones. Some are based on network topology metrics, such as using degree to pinpoint hubs, while others use a combination of degree and betweenness centrality [13, 91]. These nodes are then further filtered by relative abundance to maintain low-abundance candidates as a parallel to macro-organismal keystone species properties. Another method ascribes topological roles to nodes based on within- and among-module connectivity [80, 89]. Keystones identified with this method are not necessarily screened for low abundance.

As with some other ecological “rules” primarily derived from the study of macro-organisms [105], it may be that the traditional definition of a keystone species does not directly apply to microorganisms, or at least warrants flexibility [106]. Microbial metacommunities do not seem to be defined by habitable patches but rather stepping stones that facilitate dispersal across distance and disjunct suitable habitats [107], and as such, microbes may not be restricted to the contexts (i.e., hosts, habitats, or environments) in which keystone-like behavior is likely to be observed [106]. Experimental validation is necessary to determine the plausibility of microbial keystones [12, 14], as network-based simulations of putative keystone removal rely on the assumption that network edges represent biotic associations, which may not be the case in co-occurrence networks [35].

Inter-kingdom interactions

When inter-kingdom interactions including both prokaryotic and eukaryotic microbes are included in co-occurrence networks, it is often found that stability increases, generally as a result of the addition of fungi [13, 23], which act as connectors between network modules [79]. Multiple hypotheses may explain this phenomenon: fungi may produce metabolites that bacteria may exploit when nutrients are limited [108], provide physical space for bacterial colonization and dispersal [109], or support bacterial resistance to hydric stress [23]. One study of host evolution in wild and domesticated rice found a different effect, where the addition of fungi to bacterial networks decreased network robustness but increased transitivity (i.e., clustering coefficient) and modularity [48]. In this study, stability was indicated by network robustness, and the reduction of stability in the presence of fungi was attributed to the loss of fungal-bacterial modules [48]. More integrative microbiome research is needed, as many studies focus on bacteria alone [18, 24, 39, 76, 77, 79], but the addition of multiple guilds and/or kingdoms may change stability assessments [52].

Remaining challenges

One of the largest shortcomings of contemporary network analyses is the inability to make robust statistical comparisons between networks. Networks require a sufficient sample size to reliably infer interactions [33], but this often means every sample is put towards meeting a suitable sample-to-feature ratio and not towards generating replicate networks that would be needed to create distributions of topological metrics or other properties. Thus, we often cannot say that one network is significantly more modular than another, for example, because only one modularity value can be calculated

per network. Some studies have used subnetworks corresponding to individual plots from a single empirical network to make statistical comparisons across sites [18]. Researchers must take care to not pseudoreplicate if taking this approach, ensuring that spatial autocorrelation and proper replication are considered with respect to their questions.

When interpreting co-occurrence networks, it is important to consider the biological or ecological relevance of a given pattern, or the degree to which such relevance is attainable [64]. Data preprocessing can help to reduce spurious and indirect edges (see above), but even then, networks have limitations in representing biotic signals, as correlations may not capture asymmetric, directional associations such as trophic interactions [110]. Network structure may also be driven by various factors, including environmental filtering, dispersal limitation, stochastic processes, or biotic interaction, and the interpretations of a given network metric or property should consider the role of these factors. For example, when most links are driven by abiotic variation, calculating network robustness via simulated species extinction may not be biologically relevant. However, it remains difficult to determine which edges and network structures are governed by specific ecological processes (but see [65]); thus, future network interpretations will likely be constrained by this caveat.

Co-occurrence networks should be used to generate, and not validate hypotheses [35], especially those that assume biotic interactivity (e.g., ratios of cooperation to competition). Regardless, networks have proven utility in studying ecological theory (e.g., [106]), allow complex, species-rich microbial communities to be more interpretable, and will likely remain useful tools to guide the direction of resources in more intensive pursuits, such as co-culturing or whole genome sequencing.

The development of comprehensive tools for network generation will help ensure that resulting networks represent “interactomes” to the greatest possible extent [61]. Current tools offer some, or most, but not all of the functionality needed to do this, including the ability to handle rare species, the compositionality and high dimensionality of microbiome data, indirect edges, environmental effects, multiple domains of life, and large amounts of taxa, while minimizing the use of arbitrary thresholds. Added functionality will require more computational power, so options to compute networks in batches or in a lossy manner may be necessary. As methods evolve and converge, there will hopefully be less variability among software in what edges are inferred [34].

Conclusions

Research into microbiome networks is challenged by a forked path of decision-making and the existence of multiple plausible interpretations and few accepted standards. For example, of the network properties described in this review, it is generally agreed that stable networks are those partitioned into many distinct modules (high modularity, [18–20, 23, 25, 39, 48, 58, 75–77]) that are resilient in the face of targeted node removal (high robustness [13, 22, 23, 25, 39, 48, 58, 75–77]), have low vulnerability [39, 77] and low fragmentation [22, 24], and are held together by keystone taxa [24, 26, 39, 75–77, 79]. Although certain network metrics correlate with these properties (e.g., degree with robustness or keystone designation), a suite of topological metrics should be considered when describing stable networks.

One potential reason for the lack of convergence between network metrics and their interpretations across studies is that the majority of studies looked at single sample types (mostly soil; [18–21, 23, 25, 26, 39, 75, 78, 79]) and single domains of life (mostly bacteria; [18, 24, 39, 76, 77, 79]). A strong need for the field is to assess networks representing the diversity of ecosystems *in situ* to differentiate system-specific stability quirks from universal traits.

With further confirmatory experiments, benchmarking work on microbiome data, and computationally scalable tools able to handle the multidimensionality of the data [42, 54], co-occurrence networks can transform from being an incredibly informative tool to one capable of robust hypothesis generation and testing.

Acknowledgements

We are very grateful for the expertise provided by Maggie Mengting Yuan, Nicola Kriefall, and Elizabeth Gross in the development and revision of this review. We also thank the Hau’oli Mau Loa Foundation for fellowship support to K.T.K.

Authors’ contributions

K.T.K. researched for and wrote the manuscript. N.A.H. substantially contributed to formative discussions, and wrote, reviewed, and edited the manuscript.

Funding

Support for the writing of this review was provided by the W. M. Keck Foundation, NSF award 2124922, and the Hau’oli Mau Loa Foundation.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 3 May 2024 Accepted: 4 July 2024

Published online: 28 September 2024

References

- McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci.* 2013;110:3229–36.
- Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci.* 2011;108(supplement_1):4554–61.
- Doering T, Wall M, Putchim L, Rattanawongwan T, Schroeder R, Hentschel U, et al. Towards enhancing coral heat tolerance: a “microbiome transplantation” treatment using inoculations of homogenized coral tissues. *Microbiome.* 2021;9:102.
- Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol.* 2012;10:538–50.
- Barberan A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 2012;6:343–51.
- DeBach P. The competitive displacement and coexistence principles. *Annu Rev Entomol.* 1966;11:183–212.
- Pimm SL. The structure of food webs. *Theor Popul Biol.* 1979;16:144–58.
- Bascompte J, Jordano P. Plant-animal mutualistic networks: the architecture of biodiversity. *Annu Rev Ecol Evol Syst.* 2007;38:567–93.
- Landi P, Minoarivelo HO, Bränström Å, Hui C, Dieckmann U. Complexity and stability of ecological networks: a review of the theory. *Popul Ecol.* 2018;60:319–45.
- Jones EM, Marken JP, Silver PA. Synthetic microbiology in sustainability applications. *Nat Rev Microbiol.* 2024. <https://doi.org/10.1038/s41579-023-01007-9>.
- Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: a network perspective. *Trends Microbiol.* 2017;25:217–28.
- Agler MT, Ruhe J, Kroll S, Morhenn C, Kim S-T, Weigel D, et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol.* 2016;14:e1002352.
- Tipton L, Müller CL, Kurtz ZD, Huang L, Kleerup E, Morris A, et al. Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome.* 2018;6:1–14.
- Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, et al. Microbial interkingdom interactions in roots promote *Arabidopsis* survival. *Cell.* 2018;175(973–983):e14.
- Shi S, Nuccio EE, Shi ZJ, He Z, Zhou J, Firestone MK. The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecol Lett.* 2016;19:926–36.
- Costa L da F, Rodrigues FA, Travieso G, Villas Boas PR. Characterization of complex networks: a survey of measurements. *Adv Phys.* 2007;56:167–242.
- Morueta-Holme N, Blonder B, Sandel B, McGill BJ, Peet RK, Ott JE, et al. A network approach for inferring species associations from co-occurrence data. *Ecography.* 2016;39:1139–50.
- Luo S, Png GK, Ostle NJ, Zhou H, Hou X, Luo C, et al. Grassland degradation-induced declines in soil fungal complexity reduce fungal community stability and ecosystem multifunctionality. *Soil Biol Biochem.* 2023;176:108865.
- Hernandez DJ, David AS, Menges ES, Searcy CA, Afkhami ME. Environmental stress destabilizes microbial networks. *ISME J.* 2021;15:1722–34.
- de Vries FT, Griffiths RI, Bailey M, Craig H, Girlanda M, Gweon HS, et al. Soil bacterial networks are less stable under drought than fungal networks. *Nat Commun.* 2018;9:3033.
- Jiao S, Chu H, Zhang B, Wei X, Chen W, Wei G. Linking soil fungi to bacterial community assembly in arid ecosystems. *IMeta.* 2022;1:e2.
- Li Y, Kan J, Liu F, Lian K, Liang Y, Shao H, et al. Depth shapes microbiome assembly and network stability in the Mariana Trench. *Microbiology Spectrum.* 2023;12:e02110–e2123.
- Maurice K, Bourceret A, Youssef S, Boivin S, Laurent-Webb L, Damasio C, et al. Anthropic disturbances impact the soil microbial network structure and stability to a greater extent than natural disturbances in an arid ecosystem. *Sci Total Environ.* 2024;907: 167969.
- Wang H, Chen F, Zhang C, Wang M, Kan J. Estuarine gradients dictate spatiotemporal variations of microbiome networks in the Chesapeake Bay. *Environmental Microbiome.* 2021;16:22.
- Wu L, Ren C, Jiang H, Zhang W, Chen N, Zhao X, et al. Land abandonment transforms soil microbiome stability and functional profiles in apple orchards of the Chinese Loess Plateau. *Sci Total Environ.* 2024;906: 167556.
- Guo Y, Song B, Li A, Wu Q, Huang H, Li N, et al. Higher pH is associated with enhanced co-occurrence network complexity, stability and nutrient cycling functions in the rice rhizosphere microbiome. *Environ Microbiol.* 2022;24:6200–19.
- Lima LFO, Weissman M, Reed M, Papudesi B, Alker AT, Morris MM, et al. Modeling of the coral microbiome: the influence of temperature and microbial network. *mBio.* 2020;11:e02691–19.
- Chiarello M, McCauley M, Villéger S, Jackson CR. Ranking the biases: the choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLOS ONE.* 2022;17:e0264443.
- Glassman SI, Martiny JBH. Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere.* 2018;3:<https://doi.org/10.1128/mSphere.00148-18>.
- Joos L, Beirinckx S, Haegeman A, Debeke J, Vandecasteele B, Baeyen S, et al. Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genomics.* 2020;21:733.
- Tipton L, Zahn GL, Darcy JL, Amend AS, Hynson NA. Hawaiian fungal amplicon sequence variants reveal otherwise hidden biogeography. *Microb Ecol.* 2021. <https://doi.org/10.1007/s00248-021-01730-x>.
- Tu Q, Yan Q, Deng Y, Michaletz ST, Buzzard V, Weiser MD, et al. Biogeographic patterns of microbial co-occurrence ecological networks in six American forests. *Soil Biol Biochem.* 2020;148: 107897.
- 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: e1004226.
- Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 2016;10:1669–81.
- Guseva K, Darcy S, Simon E, Alteio LV, Montesinos-Navarro A, Kaiser C. From diversity to complexity: microbial networks in soils. *Soil Biol Biochem.* 2022;169: 108604.
- Cougoul A, Baily X, Vourc'h G, Gasqui P. Rarity of microbial species: In search of reliable associations. *PLoS One.* 2019;14:e0200458.
- Shi Y, Pan Y, Xiang L, Zhu Z, Fu W, Hao G, et al. Assembly of rhizosphere microbial communities in *Artemisia annua*: recruitment of plant growth-promoting microorganisms and inter-kingdom interactions between bacteria and fungi. *Plant Soil.* 2022;470:127–39.
- Jousset A, Bienhold C, Chatzinotas A, Gallien L, Gobet A, Kurm V, et al. Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* 2017;11:853–62.
- Yuan MM, Guo X, Wu L, Zhang YA, Xiao N, Ning D, et al. Climate warming enhances microbial network complexity and stability. *Nat Clim Chang.* 2021;11:343–8.
- McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol.* 2014;10: e1003531.
- Schloss PD. Rarefaction is currently the best approach to control for uneven sequencing effort in amplicon sequence analyses. *msphere.* 2024;9(2):e0035423.
- Faust K. Open challenges for microbial network construction and analysis. *ISME J.* 2021;15:3111–8.
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egoscue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol.* 2017;8:2224.
- Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol.* 2012;8(9):e1002687.
- Aitchison J. A new approach to null correlations of proportions. *J Int Assoc Math Geol.* 1981;13:175–89.

46. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS ONE*. 2012;7:e30126.

47. Jiang S, Xiao G, Koh AY, Chen Y, Yao B, Li Q, et al. HARMONIES: a hybrid approach for microbiome networks inference via exploiting sparsity. *Front Genet*. 2020;11:445.

48. Lee KK, Kim H, Lee Y-H. Cross-kingdom co-occurrence networks in the plant microbiome: importance and ecological interpretations. *Front Microbiol*. 2022;13:953300.

49. Kurth JK, Albrecht M, Glaser K, Karsten U, Vestergaard G, Armbruster M, et al. Biological soil crusts on agricultural soils of mesic regions promote microbial cross-kingdom co-occurrences and nutrient retention. *Front Microbiol*. 2023;14:1169958.

50. Banerjee S, Zhao C, Kirkby CA, Coggins S, Zhao S, Bissett A, et al. Microbial interkingdom associations across soil depths reveal network connectivity and keystone taxa linked to soil fine-fraction carbon content. *Agr Ecosyst Environ*. 2021;320:107559.

51. Li Z, Sun L, Liu S, Lei P, Wang R, Li S, et al. Interkingdom network analyses reveal microalgae and protostomes as keystone taxa involved in nutrient cycling in large freshwater lake sediment. *FEMS Microbiol Ecol*. 2023;99:fiad111.

52. Yang T, Tedersoo L, Liu X, Gao G-F, Dong K, Adams JM, et al. Fungi stabilize multi-kingdom community in a high elevation timberline ecosystem. *iMeta*. 2022;1:e49.

53. Brunner JD, Robinson AJ, Chain PS. Combining compositional data sets introduces error in covariance network reconstruction. *ISME Commun*. 2024;4:ycae057. <https://doi.org/10.1093/ismeco/ycae057>.

54. Matchado MS, Lauber M, Reitmeyer S, Kacprowski T, Baumbach J, Haller D, et al. Network analysis methods for studying microbial communities: a mini review. *Comput Struct Biotechnol J*. 2021;19:2687–98.

55. Connor N, Barberán A, Claußen A. Using null models to infer microbial co-occurrence networks. *PLoS ONE*. 2017;12:e0176751.

56. Deng Y, Jiang Y-H, Yang Y, He Z, Luo F, Zhou J. Molecular ecological network analyses. *BMC Bioinformatics*. 2012;13:113.

57. Faust K, Raes J. oNet app: inference of biological association networks using Cytoscap. *F1000Res*. 2016;5:1519.

58. Wu M-H, Chen S-Y, Chen J-W, Xue K, Chen S-L, Wang X-M, et al. Reduced microbial stability in the active layer is associated with carbon loss under alpine permafrost degradation. *Proc Natl Acad Sci*. 2021;118:e2025321118.

59. Goberna M, Verdú M. Cautionary notes on the use of co-occurrence networks in soil ecology. *Soil Biol Biochem*. 2022;166:108534.

60. Deutschmann IM, Lima-Mendez G, Krabberød AK, Raes J, Vallina SM, Faust K, et al. Disentangling environmental effects in microbial association networks. *Microbiome*. 2021;9:1–18.

61. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, et al. Determinants of community structure in the global plankton interactome. *Science*. 2015;348:1262073.

62. Faust K, Lima-Mendez G, Lerat J-S, Sathirapongsasuti JF, Knight R, Huttunen C, et al. Cross-biome comparison of microbial association networks. *Front Microbiol*. 2015;6:1200.

63. Ritter CD, Forster D, Azevedo JA, Antonelli A, Nilsson RH, Trujillo ME, et al. Assessing biotic and abiotic interactions of microorganisms in Amazonia through co-occurrence networks and DNA metabarcoding. *Microb Ecol*. 2021;82:746–60.

64. Röttjers L, Faust K. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol Rev*. 2018;42:761–80.

65. Goberna M, Montesinos-Navarro A, Valiente-Banuet A, Colin Y, Gómez-Fernández A, Donat S, et al. Incorporating phylogenetic metrics to microbial co-occurrence networks based on amplicon sequences to discern community assembly processes. *Mol Ecol Resour*. 2019;19:1552–64.

66. Barabasi A-L. Chapter 3: Random Networks. In: Barabasi A-L, editor. *Network Science*; 2016. <http://networksciencebook.com>. Accessed 1 July 2024.

67. Brodo AD, Claußen A. Scale-free networks are rare *Nat Commun*. 2019;10:1017.

68. Barabasi A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004;5:101–13.

69. Ma B, Wang H, Dsouza M, Lou J, He Y, Dai Z, et al. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J*. 2016;10:1891–901.

70. Scholz M. Node similarity as a basic principle behind connectivity in complex networks. *J Data Min Digit Humanit*. 2015;2015. <https://doi.org/10.46298/jdmh.33>.

71. Liu C, Ma Y, Zhao J, Nussinov R, Zhang Y-C, Cheng F, et al. Computational network biology: data, models, and applications. *Phys Rep*. 2020;846:1–66.

72. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, complex systems*. 2006;1695:1–9.

73. Watson CG. brainGraph: graph theory analysis of brain MRI data. *R package version*. 2020;3.

74. Dai W, Chen J, Xiong J. Concept of microbial gatekeepers: positive guys? *Appl Microbiol Biotechnol*. 2019;103:633–41.

75. Fan K, Weisenhorn P, Gilbert JA, Chu H. Wheat rhizosphere harbors a less complex and more stable microbial co-occurrence pattern than bulk soil. *Soil Biol Biochem*. 2018;125:251–60.

76. Wang J, Wang C, Zhang J, Wu X, Hou Y, Zhao G, et al. Decreased precipitation reduced the complexity and stability of bacterial co-occurrence patterns in a semiarid grassland. *Front Microbiol*. 2022;13:1031496.

77. Liu S, Yu H, Yu Y, Huang J, Zhou Z, Zeng J, et al. Ecological stability of microbial communities in Lake Donghu regulated by keystone taxa. *Ecol Ind*. 2022;136:108695.

78. Gao E, Ma H, Yang T, Kaiser-Bunbury CN, Zhao Z. Meadow transformations alter above and below-ground ecological networks and ecosystem multifunctionality. *Funct Ecol*. 2023;37:1703–16.

79. Yang L, Sun R, Li J, Zhai L, Cui H, Fan B, et al. Combined organic-inorganic fertilization builds higher stability of soil and root microbial networks than exclusive mineral or organic fertilization. *Soil Ecology Letters*. 2023;5:220142.

80. Olesen JM, Bascompte J, Dupont YL, Jordano P. The modularity of pollination networks. *Proc Natl Acad Sci*. 2007;104:19891–6.

81. Chaffron S, Rehrauer H, Pernthaler J, von Mering C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res*. 2010;20:947–59.

82. Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW. Compartments revealed in food-web structure. *Nature*. 2003;426:282–5.

83. May RM. Will a large complex system be stable? *Nature*. 1972;238:413–4.

84. Stouffer DB, Bascompte J. Compartmentalization increases food-web persistence. *Proc Natl Acad Sci*. 2011;108:3648–52.

85. Thébaud E, Fontaine C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*. 2010;329:853–6.

86. Schaub MT, Delvenne J-C, Rosvall M, Lambiotte R. The many facets of community detection in complex networks. *Applied network science*. 2017;2:1–13.

87. Fortunato S, Hric D. Community detection in networks: a user guide. *Phys Rep*. 2016;659:1–44.

88. Pons P, Latapy M. Computing communities in large networks using random walks. In: *Computer and Information Sciences-ISCIS 2005: 20th International Symposium, Istanbul, Turkey, October 26–28, 2005. Proceedings* 20. Springer; 2005. 284–93.

89. Guimera R, Nunes Amaral LA. Functional cartography of complex metabolic networks. *Nature*. 2005;433:895–900.

90. Mitra K, Carvunis A-R, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet*. 2013;14:719–32.

91. Wall CB, Egan CP, Swift SIO, Hynson NA. Three decades post-forestation has not led to the reassembly of arbuscular mycorrhizal fungal communities associated with remnant primary forests. *Mol Ecol*. 2020;29:4234–47.

92. Lewinsohn TM, Inácio Prado P, Jordano P, Bascompte J, M. Olesen J. Structure in plant–animal interaction assemblages. *Oikos*. 2006;113:174–84.

93. Baldassano SN, Bassett DS. Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Sci Rep*. 2016;6:26087.

94. MacArthur R. Fluctuations of animal populations and a measure of community stability. *Ecology*. 1955;36:533–6.

95. Allesina S, Tang S. Stability criteria for complex ecosystems. *Nature*. 2012;483:205–8.

96. Rooney N, McCann K, Gellner G, Moore JC. Structural asymmetry and the stability of diverse food webs. *Nature*. 2006;442:265–9.
97. Dunne JA, Williams RJ. Cascading extinctions and community collapse in model food webs. *Philos Trans R Soc Lond B Biol Sci*. 2009;364:1711–23.
98. Iyer S, Killingback T, Sundaram B, Wang Z. Attack robustness and centrality of complex networks. *PLoS ONE*. 2013;8:e59613.
99. Peng G, Wu J. Optimal network topology for structural robustness based on natural connectivity. *Physica A*. 2016;443:212–20.
100. Widder S, Besemer K, Singer GA, Ceola S, Bertuzzo E, Quince C, et al. Fluvial network organization imprints on microbial co-occurrence networks. *Proc Natl Acad Sci*. 2014;111:12799–804.
101. Herren CM, McMahon KD. Cohesion: a method for quantifying the connectivity of microbial communities. *ISME J*. 2017;11:2426–38.
102. Coyte KZ, Schlüter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science*. 2015;350:663–6.
103. Bertness MD, Callaway R. Positive interactions in communities. *Trends Ecol Evol*. 1994;9:191–3.
104. Paine RT. A note on trophic complexity and community stability. *Am Nat*. 1969;103:91–3.
105. Dickey JR, Swenson RA, Turner SC, Winfrey CC, Yaffar D, Padukone A, et al. The utility of macroecological rules for microbial biogeography. *Front Ecol Evol*. 2021;9:633155.
106. Banerjee S, Schlaepi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol*. 2018;16:567–76.
107. Amend AS, Swift SO, Darcy JL, Belcaid M, Nelson CE, Buchanan J, et al. A ridge-to-reef ecosystem microbial census reveals environmental reservoirs for animal and plant microbiomes. *Proc Natl Acad Sci*. 2022;119:e2204146119.
108. Warmink JA, Nazir R, Van Elsas JD. Universal and species-specific bacterial 'fungiphiles' in the mycospheres of different basidiomycetous fungi. *Environ Microbiol*. 2009;11:300–12.
109. Kohlmeier S, Smits TH, Ford RM, Keel C, Harms H, Wick LY. Taking the fungal highway: mobilization of pollutant-degrading bacteria by fungi. *Environ Sci Technol*. 2005;39:4640–6.
110. Freilich MA, Wieters E, Broitman BR, Marquet PA, Navarrete SA. Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? *Ecology*. 2018;99:690–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.