



## Perspectives Paper

## From diversity to complexity: Microbial networks in soils

Ksenia Guseva<sup>a,\*,\*\*</sup>, Sean Darcy<sup>a</sup>, Eva Simon<sup>a,c</sup>, Lauren V. Alteio<sup>a</sup>, Alicia Montesinos-Navarro<sup>b</sup>, Christina Kaiser<sup>a,\*</sup>

<sup>a</sup> Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria

<sup>b</sup> Centro de Investigaciones sobre Desertificación (CIDE, CSIC-UV-GV), Carretera de Moncada-Náquera Km 4.5, 46113, Moncada, Valencia, Spain

<sup>c</sup> Doctoral School in Microbiology and Environmental Science, University of Vienna, Vienna, Austria

## ARTICLE INFO

## Keywords:

Microbial network analysis  
Co-occurrence networks  
Ecological networks  
Microbial community structure  
Soil microbial ecology

## ABSTRACT

Network analysis has been used for many years in ecological research to analyze organismal associations, for example in food webs, plant-plant or plant-animal interactions. Although network analysis is widely applied in microbial ecology, only recently has it entered the realms of soil microbial ecology, shown by a rapid rise in studies applying co-occurrence analysis to soil microbial communities. While this application offers great potential for deeper insights into the ecological structure of soil microbial ecosystems, it also brings new challenges related to the specific characteristics of soil datasets and the type of ecological questions that can be addressed. In this Perspectives Paper we assess the challenges of applying network analysis to soil microbial ecology due to the small-scale heterogeneity of the soil environment and the nature of soil microbial datasets. We review the different approaches of network construction that are commonly applied to soil microbial datasets and discuss their features and limitations. Using a test dataset of microbial communities from two depths of a forest soil, we demonstrate how different experimental designs and network constructing algorithms affect the structure of the resulting networks, and how this in turn may influence ecological conclusions. We will also reveal how assumptions of the construction method, methods of preparing the dataset, and definitions of thresholds affect the network structure. Finally, we discuss the particular questions in soil microbial ecology that can be approached by analyzing and interpreting specific network properties. Targeting these network properties in a meaningful way will allow applying this technique not in merely descriptive, but in hypothesis-driven research. Analysing microbial networks in soils opens a window to a better understanding of the complexity of microbial communities. However, this approach is unfortunately often used to draw conclusions which are far beyond the scientific evidence it can provide, which has damaged its reputation for soil microbial analysis. In this Perspectives Paper, we would like to sharpen the view for the real potential of microbial co-occurrence analysis in soils, and at the same time raise awareness regarding its limitations and the many ways how it can be misused or misinterpreted.

## 1. Introduction

Networks present a powerful mathematical framework to explore the complexity of relationships within an ecological community. Their potential lies in a formal but intuitive representation of the organization of complex systems, where entities (e.g. species) are displayed as nodes, and interactions or associations among them as edges (links). Since the beginning of the century, network analysis has become an established tool for investigating species relationships in ecosystems (May 1974;

Bascompte, 2009; Poisot et al., 2016) and has experienced an upswing in soil ecological research over the past ten years (Fig. S1).

One of the original aims of the field is to understand the outcome of the combination of species interactions present in nature, which allows to gain valuable information on ecosystem properties. Data for ecological network analysis of species interactions is ideally obtained by direct observation of physical contact, e.g. pollination of plants by insects or birds, or mycorrhizal fungi colonizing plant roots. However even in such cases identifying species interactions within an ecosystem can be

DOI of original article: <https://doi.org/10.1016/j.soilbio.2021.108534>.

\* Corresponding author.

\*\* Corresponding author.

E-mail addresses: [ksenia.guseva@univie.ac.at](mailto:ksenia.guseva@univie.ac.at) (K. Guseva), [christina.kaiser@univie.ac.at](mailto:christina.kaiser@univie.ac.at) (C. Kaiser).

<https://doi.org/10.1016/j.soilbio.2022.108604>

Received 24 August 2021; Received in revised form 8 February 2022; Accepted 9 February 2022

Available online 18 March 2022

0038-0717/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

challenging, as it requires repeated observations or a mechanistic understanding of them. An alternative approach for investigating networks of species relationships has been the analysis of spatial co-occurrence patterns of species with so-called co-occurrence networks. Although the use of co-occurrence data often limits possible conclusions on species interactions, it contains valuable information regarding community structure and assembly mechanisms which can be explored by the use of network analysis (Blanchet et al., 2020).

In microbial ecosystems, associations between microbial taxa can usually only be assessed via co-occurrence data, as interactions between microbes in their natural environment are much more difficult to observe than in macro-ecosystems. Their assessment would require complicated experiments, for example, involving records of events of feeding or cross-feeding, mechanisms behind biofilm formation or other types of interactions. All of which are impossible to carry out in situ, particularly in complex environments such as the soil. Network analysis in microbial ecology therefore needs to rely on the observation of species co-occurrence or co-exclusion patterns within molecular microbial datasets (e.g. derived from 16S amplicon sequencing) across a sufficiently high number of environmental samples (Röttgers and Faust, 2018). Networks generated from this indirect information are known as 'microbial co-occurrence networks' or 'microbial association networks', and have gained much popularity in microbial ecology in the last decade (Fig. S1). Analogous to their macro-ecological counterparts, the observed structure of microbial co-occurrences is the result of community assembly, a complex ecological mechanism, which is influenced by various processes such as environmental filtering (i.e. species that respond in the same way to an environmental factor tend to co-occur in samples with variation of that environmental factor), species interactions (i.e. two microbial taxa need to exchange specific metabolites for increasing their fitness), dispersal dynamics and stochastic processes (Faust, 2021). Network analysis of in situ distributions of microbial taxa in soil reveal the final result of these community assembly processes, and can therefore be a great starting point to better understand them. To interpret these networks it is important to distinguish microbial associations from microbial interactions (Barner et al., 2018). While the first refers to the observed statistical signal obtained from co-occurrence (or abundance) patterns, microbial interactions constitute the relationships between species present in natural systems.

The advent of high-throughput sequencing techniques has led to widespread use of co-occurrence analysis of microbial communities, which went along with the development of significant methodological advances. It is currently used in a rapidly growing number of soil microbial ecology studies. While not more than 10% of microbial network analysis studies were conducted on soil datasets before 2010, this number has steadily increased since then to 35% in 2020, see (Fig. S1) of Supplementary Material. While this approach has great potential to enrich soil studies, the intrinsic complexity of soil adds new challenges to the construction and interpretation of network models. The difficulty consists in taking into account the inter- and intra-variability of samples, which are a result of soil heterogeneity (Carr et al., 2019). The first leads to a high number of environmentally driven association patterns, which can be mistaken as species interactions (Armitage and Jones, 2019). The second, refers to the use of large sample volumes, where naturally segregated and chemically diverse micro-habitats are mixed (usually 0.25 cm<sup>3</sup> soil for DNA analysis, derived from a homogenized sampling volume of 250–500 cm<sup>3</sup> soil). The resulting loss of information about the small-scale physical, chemical and biological diversity of soil may also confound the signal of biological interactions (Berry and Widder, 2014).

Overall, network analysis relies on a clear understanding of what the constructed network model represents and the data used for it (Poisot et al., 2016). Unfortunately, due to misuse, networks have unfairly acquired a mixed reputation in soil microbial ecology. On one hand they are considered essential for understanding the structure and properties of microbial communities (Faust and Raes, 2012). On the other hand, theoretical ecologists warn against inappropriate conclusions derived

from inaccurate data handling, unsuitable inference methods and more often an inappropriate interpretation of the obtained network model (Berry and Widder, 2014; Carr et al., 2019; Blanchet et al., 2020). As the use of networks in studies of soil microbial communities continues to increase, we see an urgency to discuss the challenges the approach inherits from other fields and new challenges the soil system poses.

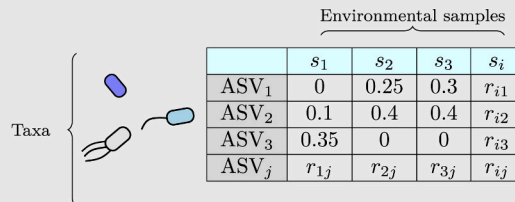
In this Perspective we describe the main ideas behind network inference from soil microbial co-occurrence datasets (Sec. 2), the network construction itself (Sec. 3), and the mathematical framework for the analysis of such networks (Sec. 4). After this overview, we discuss the challenges of application of these methods to soil microbial communities (Sec. 5). The network construction is illustrated using a dataset of microbial communities from different depths of a forest soil. We discuss how differences in the experimental design, data preparation and filtering, as well as network construction algorithms affect the structure of the resulting networks. We also discuss how these differences affect our interpretation of the edges in the network and influence our ecological conclusions. We argue that to exploit the full potential of network analysis for soil ecological studies it takes a combined effort from both experimental and theoretical sides. Applying network properties in a meaningful way enables the application of network analysis not only in merely descriptive, but also in hypothesis-driven research.

## 2. Networks as models of microbial communities in soil

Microbial association networks are constructed based on co-occurrences of microbial taxa across a sufficiently high number of environmental samples (BOX 1). Each node in the constructed network represents one microbial taxonomic unit while an edge between two nodes represents a significant association between these microbial taxa. There are different ways how network construction algorithms determine whether two taxa are significantly associated, and thus connected by an edge in the network. The most common approach is based on pairwise correlations among taxa across all samples. (for details see Sec. 3 and BOX 1). One of the key prerequisites for interpreting the structure of the resulting network is to understand what an edge, i.e. a significant association, represents.

To understand what an edge (observed association) could potentially mean we need to understand the mechanisms that lead to the observed co-occurrence pattern, i.e. the mechanisms by which microbial communities assemble. Microbial community assembly is driven by multiple processes, such as environmental filtering, inter-specific interactions, reproduction/mortality events, dispersal limitation and sporadic mutations (Nemergut et al., 2013). Besides, other factors such as legacy effects, the importance of the order by which species are introduced into a system, show how community composition can depend on elusive elements (Fukami, 2015).

The combined outcome of all these hidden ecological processes leads to the spatial distribution of microbial taxa in soil. The relevance of each process for co-occurrence patterns at a given scale, however, remains controversial. Soil microbial network analysis faces the challenge to use observations of microbial distributions which are the result of multiple processes for constructing reliable models of microbial communities as schematically depicted in (Fig. 1). It is important to keep in mind that network analysis targets co-occurrence pattern of microbial taxa, while the spatial distribution of microbes in the soil is regulated by the dynamics and interactions of individual microbial cells. Dynamics and interactions of individual soil microbes may not only be taxon-specific, but also depend on the current state of their local environment, in particular when considering the high physical, chemical and biological heterogeneity at the soil microscale. Constructed networks thus in general reflect co-occurrences caused by diverse and complex dynamics of individual microbes, but may only be able to capture a signal of interactions between species, if these interactions are linked to general species-specific traits and are not strongly modulated by environmental variability within and between samples, see also Sec. 5.2.

**Box 1****1. Environmental dataset**


	Environmental samples			
	$s_1$	$s_2$	$s_3$	$s_i$
ASV <sub>1</sub>	0	0.25	0.3	$r_{i1}$
ASV <sub>2</sub>	0.1	0.4	0.4	$r_{i2}$
ASV <sub>3</sub>	0.35	0	0	$r_{i3}$
ASV <sub>j</sub>	$r_{1j}$	$r_{2j}$	$r_{3j}$	$r_{ij}$

$r_{ij}$  – measure of ASV<sub>j</sub> (taxa j) in sample  $s_i$ , often available as relative abundance data.

Each taxa will be represented as a node in the network.

**Calculating pairwise associations****2. Pairwise association matrix**

	ASV <sub>1</sub>	ASV <sub>2</sub>	ASV <sub>3</sub>	...	ASV <sub>i</sub>
ASV <sub>1</sub>	—	0.1	0.3	...	$m_{i1}$
ASV <sub>2</sub>	0.1	—	0.23	...	$m_{i2}$
ASV <sub>3</sub>	0.3	0.23	—	...	$m_{i3}$
ASV <sub>j</sub>	$m_{1j}$	$m_{2j}$	$m_{3j}$	...	$m_{ij}$

$m_{ij}$  – measure of association between two taxa. This can be:

- co-occurrence (co-absence) of the two taxa;
- positive (negative) correlation between taxa;
- other measures of associations.

**Selecting significant associations**

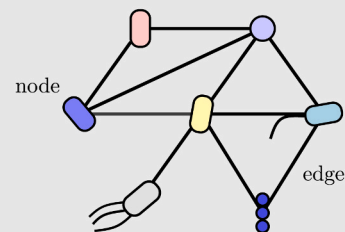
“Edge” thresholds determine which associations are significant and are transformed into edges. Examples:

- Percentage of samples in which two taxa need to co-occur;
- Significance level for correlations.

**3. Adjacency matrix**

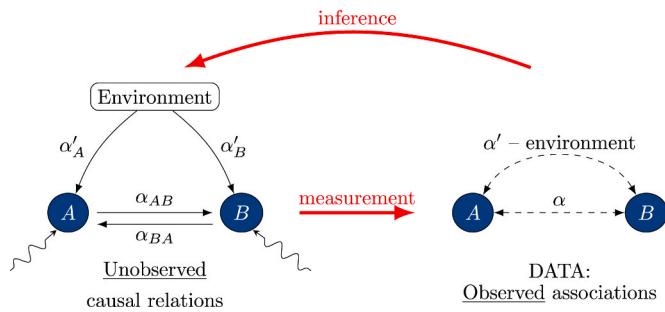
	ASV <sub>1</sub>	ASV <sub>2</sub>	ASV <sub>3</sub>	...	ASV <sub>i</sub>
ASV <sub>1</sub>	0	0	1	...	$e_{i1}$
ASV <sub>2</sub>	0	0	1	...	$e_{i2}$
ASV <sub>3</sub>	1	1	0	...	$e_{i3}$
ASV <sub>j</sub>	$e_{1j}$	$e_{2j}$	$e_{3j}$	...	$e_{ij}$

$e_{ij}$  – edge between significantly associated ASVs (nodes), the presence or absence of links is represented by 0 or 1 respectively.



Further options:

- distinguish positive and negative associations;
- weighted edges according to the strength of associations.



**Fig. 1.** The hidden relationships (left diagram) within an ecosystem are reconstructed (inferred) from limited observations and measurements, in this case of associations between microorganisms (right diagram).

The process of reconstructing an underlying ecological reality from observed data by means of an abstract mathematical model is called inference. Network inference, which is also often termed network reconstruction, correspondingly looks for mathematical models that represent relationships in a system as a network with the aim to reconstruct the hidden ecological relationships. Despite the importance

of building networks on a suitable mathematical model, many state-of-the-art network reconstruction methods don't go beyond a mere identification of patterns, such as significant correlations, and thereby only provide a naive representation of the hidden reality without a clear interpretation. This is often the case for microbial co-occurrence networks in soil which are most often constructed based on pairwise correlations. To understand what an edge, that connects two taxa in a network, represents, we need to understand why these two taxa significantly co-occur across a high number of samples. Despite the wealth of underlying causes for the spatial distribution of microbial taxa in soil, significant species co-occurrences, which are the basis for network edges, are often interpreted to be caused mainly by interactions between the taxa. To avoid such misleading conclusions from network analyses, we suggest that thinking about the underlying biological processes and statistical assumptions should be at the core of network model construction of soil microbial communities. Nevertheless, we concede that the detection of associations can be a valuable intermediate step, which can help to generate hypotheses about a given system and its taxa. For a safe interpretation, it is essential to (i) consider the details of the experimental design, as well as the assumptions of the used statistical methods, as both may influence the potential meanings of network edges, and (ii) (ideally) develop network reconstruction methods that go

## 1. Original Dataset

	sample 1	sample 2	sample 3	sample $i$	total sum for ASV
ASV <sub>1</sub>	0	25	3	$a_{i1}$	$A_1 = \sum_i^n a_{i1}$
ASV <sub>2</sub>	1	4	4	$a_{i2}$	$A_2$
ASV <sub>3</sub>	35	0	0	$a_{i3}$	$A_3$
ASV <sub><math>j</math></sub>	$a_{1j}$	$a_{2j}$	$a_{3j}$	$a_{ij}$	$A_j$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
total read sum for sample	$S_1 = \sum_j^N a_{1j}$	$S_2$	$S_3$	$S_i$	

- $a_{ij}$  – count reads of the sequence  $j$  (ASV <sub>$j$</sub> ) in sample  $s_i$ .
- $N$  – total number of ASVs.
- $n$  – total number of samples.

## 2. Data Filtering

- Total sum threshold  
eliminate all ASVs with  $A_i < A_{min}$
- Prevalence threshold  
eliminate all ASVs with prevalence  $< P_{min}$

## 3. Data transformation

### Scaling by the library size

$r_{ij} = a_{ij}/S_i$  (relative abundance)

$$T_i = \sum_j^N r_{ij} = 1$$

### Rarefaction

$n_{ij}$  – subsampled from original data

$$T_i = \sum_j^N n_{ij} = cte$$

### Log ratios (example clr)

$$l_{ij} = \log \left( \frac{a_{ij}}{(\prod_j^n a_{ij})^{1/n}} \right)$$

$$T_i = \sum_j^N l_{ij} = 0$$

### Additional data

qPCR or ddPCR to normalize by gene copy number

## 4. Prepared Dataset

(Example with scaling by the library size)

	sample 1	sample 2	sample 3	sample $i$
ASV <sub>1</sub>	0	0.25	0.3	$r_{i1}$
ASV <sub>5</sub>	0.105	0.10	0.05	$r_{i5}$
ASV <sub>10</sub>	0	0.45	0.21	$r_{i10}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
ASV <sub><math>N'</math></sub>	$r_{1N'}$	$r_{2N'}$	$r_{3N'}$	$r_{iN'}$

- $N'$  – total number of ASVs in the trimmed data-set ( $N' < N$ )

**Fig. 2.** Workflow for the preparation of a dataset which precedes the network construction. It involves filtering out some ASVs and a data transformation, such as scaling the number of reads  $a_{ij}$  by the total read sum in a sample  $S_i$ , or taking a centered log ratio (clr). The transformed dataset contains relative abundance values constrained by a constant (cte) sum  $T_i$ , or even inferred absolute abundance values for each ASV in each individual sample.



towards a more accurate representation of the underlying ecological relationships (see Sec. 5).

### 3. Network construction

Network model construction aims to reconstruct a network of ecological relationships existing in nature from observational data. This aim is inherently challenging and can only result in a model that in the best case reflects some aspects of the existing ecological network. While there is an agreement that a graphical representation of ecological relationships in a system is useful, there is no universal framework to infer these relationships from environmental datasets. In this section we describe the most popular statistical approaches used to assess associations between microbes and discuss their potential and drawbacks.

We illustrate the process of network construction and the effect of considering different experimental designs on the network structure by examining the microbial communities from forest soil at two depths, based on 18 samples taken from a lower soil layer (15–20 cm depth) and 20 samples from the upper soil layer (0–5 cm depth). The community composition was evaluated by sequencing 16S SSU rRNA marker gene. The description of sampling, DNA extraction and sequencing processes are given in the Supplementary material. The purpose of this analysis is to elucidate what effect i) the inclusion of a strong environmental factor which may influence the spatial distribution of microbial taxa in the soil (soil depth) and ii) the choice of a certain statistical network inference method can have on the resulting network structure and interpretation.

#### 3.1. Preparation of the dataset

In most network construction approaches, associations between microorganisms are identified either from probabilities of co-presence (co-absence) or from pairwise correlations of species abundance in a number of environmental samples. Abundances of microbial taxa in natural microbial communities are usually assessed by amplicon sequencing of highly preserved, phylogenetically informative marker genes, such as the 16S ribosomal RNA for bacteria and the internal

transcribed spacer (ITS) for fungi. The sequenced reads are grouped by sequence similarity into operational taxonomic units (OTUs) or recovered from sequenced reads as amplicon sequence variants (ASVs) (Callahan et al., 2017), which can then be classified into taxonomic categories.

Two main characteristics of amplicon datasets are: (i) they are compositional, which means that abundances of taxa can only be interpreted in relative terms (Gloor et al., 2016; Morton et al., 2019), and, especially for soil microbial communities, (ii) taxa are sparsely distributed, which means that datasets contain a high amount ASVs or OTUs that only occur in a fraction of the samples (Alteio et al., 2021). Both of these characteristics pose a major challenge to network construction approaches, as correlations carried out on relative abundance and sparsely distributed datasets can lead to spurious results (Morton et al., 2019; Alteio et al., 2021). In the next subsections we discuss how to reduce the potential bias introduced by these features of the dataset, by applying appropriate data filtering and transforming steps before the actual network construction (Fig. 2).

##### 3.1.1. Interpreting reads: estimations of abundances of microorganisms from high-throughput sequencing data

Amplicon sequencing data is compositional, which means that it contains information only on relative abundances. There are several methods to extract this relative information out of measured count reads, which is most often done by: (1) rescaling the count reads by the library size ( $a_{ij}/S_i$ ) or (2) rarefaction, see (Fig. 2). The constant sum constraint ties all values together and consequently the obtained values in each sample are not independent. The computation of correlations on these non independent values is unreliable and thus gives rise to spurious values (Lovell et al., 2015; Gloor et al., 2016). The introduced bias is stronger if the reads of some taxa dominate the data, while it plays a minor role for data sets of highly diverse communities. Constructing correlation-based networks from compositional datasets is common practice, and particularly often seen in studies of soil microbial communities. It is important to keep in mind that, depending on the dataset, these networks may be dominated by spurious associations. To overcome the problem of spurious correlations caused by compositional data, log ratios instead of transformations (1 and 2) were proposed (Fig. 2). However, they require an adequate exchange of zeros by pseudo-counts, which implicitly assumes that all unfiltered taxa are present in all samples, but were not detected due to shortcomings of sequencing.

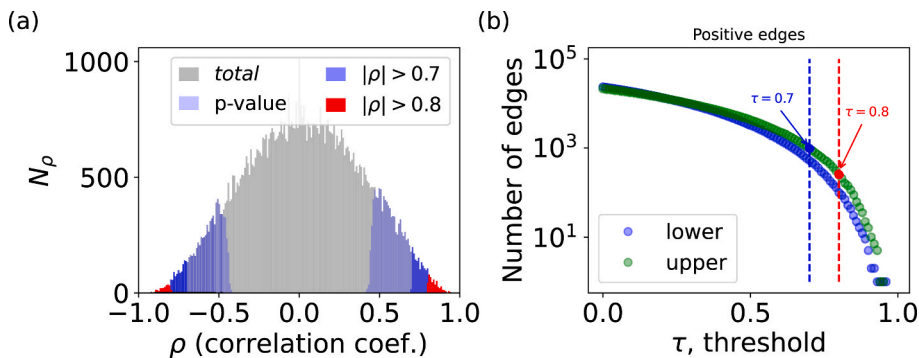
As it is not possible to directly recover absolute abundances from sequencing data, complementing amplicon sequencing data with additional data can improve quantitative insights into microbial communities. One possibility is to use qPCR or droplet digital PCR (ddPCR) to quantify the number of marker gene copies in the sample (Perisin et al., 2016; Jian et al., 2020). Another option is to infer marker gene copy numbers based on the phylogeny of organisms (Kembel et al., 2012). However this step relies heavily on a good classification of the taxonomy and while the approach generally works well for gut bacteria, it can be more challenging for fungi (Lavrinenko et al., 2021).

##### 3.1.2. Interpreting zeros: sparsity in amplicon sequencing datasets

Another important aspect of amplicon sequencing data sets is their sparsity, which results from biological complexity as well as technical challenges associated with nucleic acid extraction and sequencing. Particularly in soil, which is characterized by high species diversity and heterogeneity, most detected taxa are rare and only appear in a small fraction of the total number of samples. Consequently, these datasets are often highly zero-inflated, for example for our datasets close to 50% of all read values across samples are zeros. Observed zeros result from several very different processes (Kaul et al., 2017). On one hand, there are biological zeros that show the real absence of a microorganism in the community and result from a complex interplay of community assembly and dynamics, see Sec. 2. On the other hand, the data can display

**Table 1**  
Definitions used in this Perspective Paper.

	Useful definition
Co-occurrence or association networks	Networks constructed from co-occurrence or co-exclusion (abundance) data.
Interaction networks	Network of causal relations that exist between organisms in nature.
Network reconstruction	The aim to reconstruct the underlying ecological reality, for example the network of interactions in nature, from observed data.
Network model construction or network construction	Construction of a model network from observed data, which may or may not capture the underlying ecological reality.
Prevalence	Number of samples where a species is present across the complete set of samples.
Compositional data	Data where its components represent proportions (parts of some-thing), in other words their sum has a constant value.
Subcompositional coherence	Results obtained for a whole data set should not contradict the results obtained from any of its parts. Here in particular if we take into account all taxa or filter some of them out.
Prevalence threshold	A threshold, which cuts off taxa with low prevalence in environmental samples. Done previous to network construction.
Total sum threshold	A threshold, which filters taxa with low total number of reads, in other words low average read numbers in all environmental samples. Done before network construction.
Edge threshold	Threshold for pairwise association measure (e. g. correlation value) defining the establishment of an edge between taxa. Done for network construction.



**Fig. 3.** The choice of the threshold, to establish an association edge, changes the sparsity of the constructed networks. (a) Frequency,  $N_\rho$  of the Spearman's rank correlation coefficients  $\rho$  obtained for all ASV pairs in the upper soil core (in gray). The selected  $\rho$  values by the use of different edge thresholds  $\tau$  are marked with different colors (red and dark blue). In light blue we show  $\rho$  selected as significant by comparison with a null model (for that we use 1000 shuffled versions of our dataset, and p-value = 0.05). (b) Number of positive edges ( $\rho > 0$ ) in networks constructed with different edge thresholds. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

technical zeros that result from upstream sample processing, or limitations of sequencing data. Upstream challenges are related to DNA extraction bias and amplicon primer design, which may not capture particular subsets of organisms. In addition, zeros can be caused by undersampling of a soil sample. As a result of the high diversity of soil, high sequencing depth must be achieved to capture the scope of diversity. The sparsity of the data and the low number of sequenced samples typical for most experiments make it difficult to establish meaningful associations among species with low prevalence. First of all, the appearance of taxa in only a few samples makes it statistically impossible to spot any negative associations among them (Cougoul et al., 2019). Second, the observation of co-absence of taxa in samples can result in spurious positive associations. False positive associations may arise since correlations, such as the widely used Spearman's rank, can overestimate obtained values due to many repeated zero values (tied ranks) (Connor et al., 2017). Since we do not know how to correctly interpret zeros in our data, we should not attribute significance to these correlation values inflated by species absences. Therefore, it is common to establish a prevalence threshold, requiring e.g. a presence in 20% of all samples, or set a total sum threshold requiring a minimum number of reads, see (Table 1). The arbitrariness of thresholds was pointed out in Cougoul et al. (2019), and remains controversial in sequencing data analysis. Alternatively randomly chosen pseudocounts can be added to the dataset (Connor et al., 2017). For our dataset of forest soil we have used a total sum threshold of 100 reads. Due to the structure of our dataset, in which taxa with low reads also were of low prevalence, this criteria eliminated the taxa of low read numbers, as well as low prevalence, leaving 595 ASVs for the combined dataset of upper and lower soil layers, as well as 283 and 304 ASVs for the separate analysis of upper and lower soil, respectively. We have checked that zeros do not have influence for the established pairwise associations in this trimmed dataset, see details in the Supplementary material.

### 3.2. Establishing associations

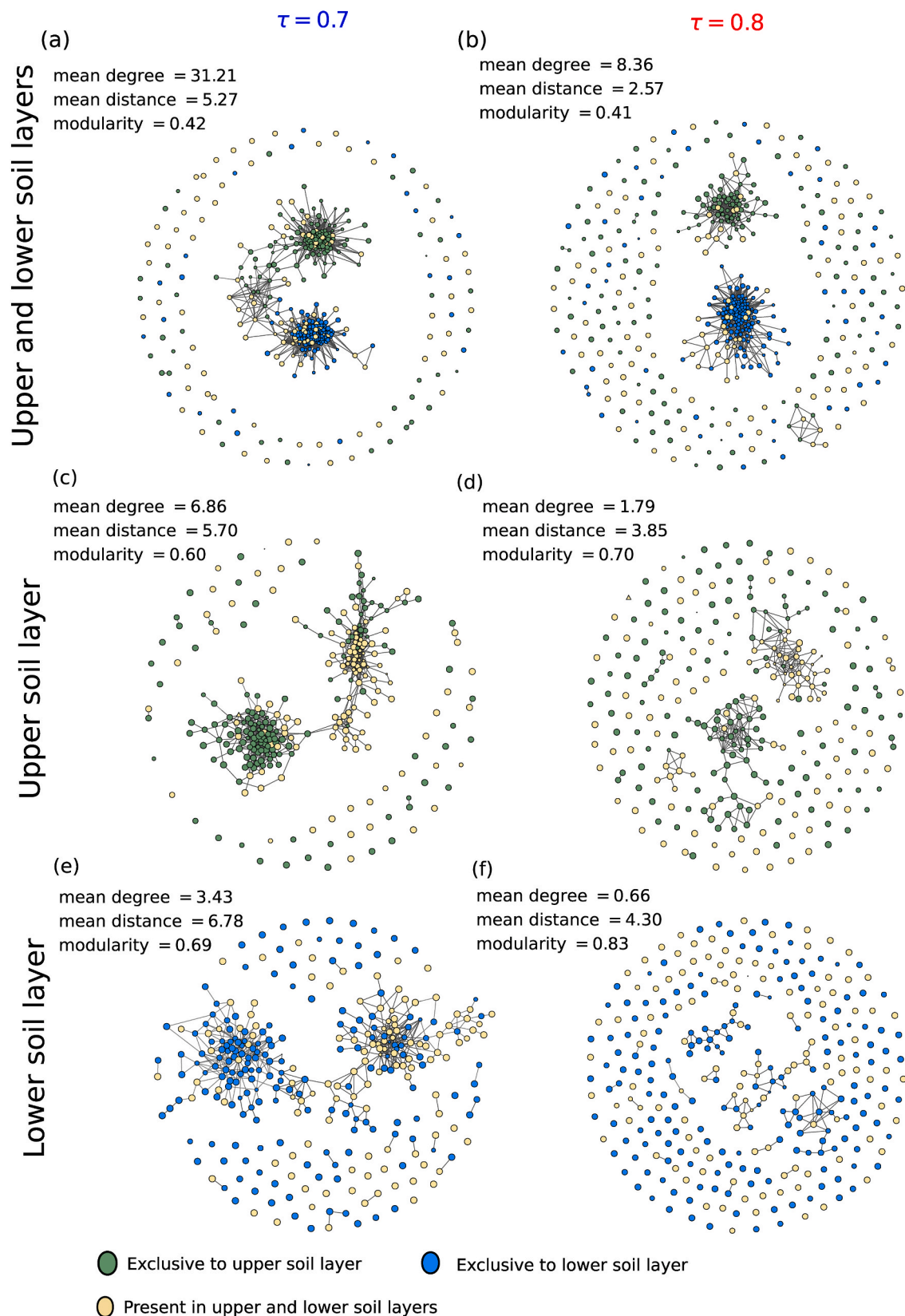
#### 3.2.1. Choice of metric to measure associations and choice of an edge threshold

The most common methods of network construction rely on establishing edges between taxa with significant pairwise associations. In this case the network depends strongly on two key decisions: (1) the choice of a metric to quantify the strength of association's signal; and (2) the choice of a significance threshold - referred to as edge threshold (see BOX1), which serves as a criterium to differentiate a true signal from noise. The strength of associations (1) is most often measured using parametric Pearson or non-parametric Spearman or Kendall- $\tau$  correlation coefficients. The latter two are more convenient since they can be used on non-normally distributed datasets, as is the case of microbiomes. However these are not the only possibilities for (1) and measures such as Bray Curtis, Kullback Leibler dissimilarities or even mutual/maximal information can also be used to quantify associations. After the metric is chosen the next step is (2) — the choice of an edge threshold,  $\tau$ . Before

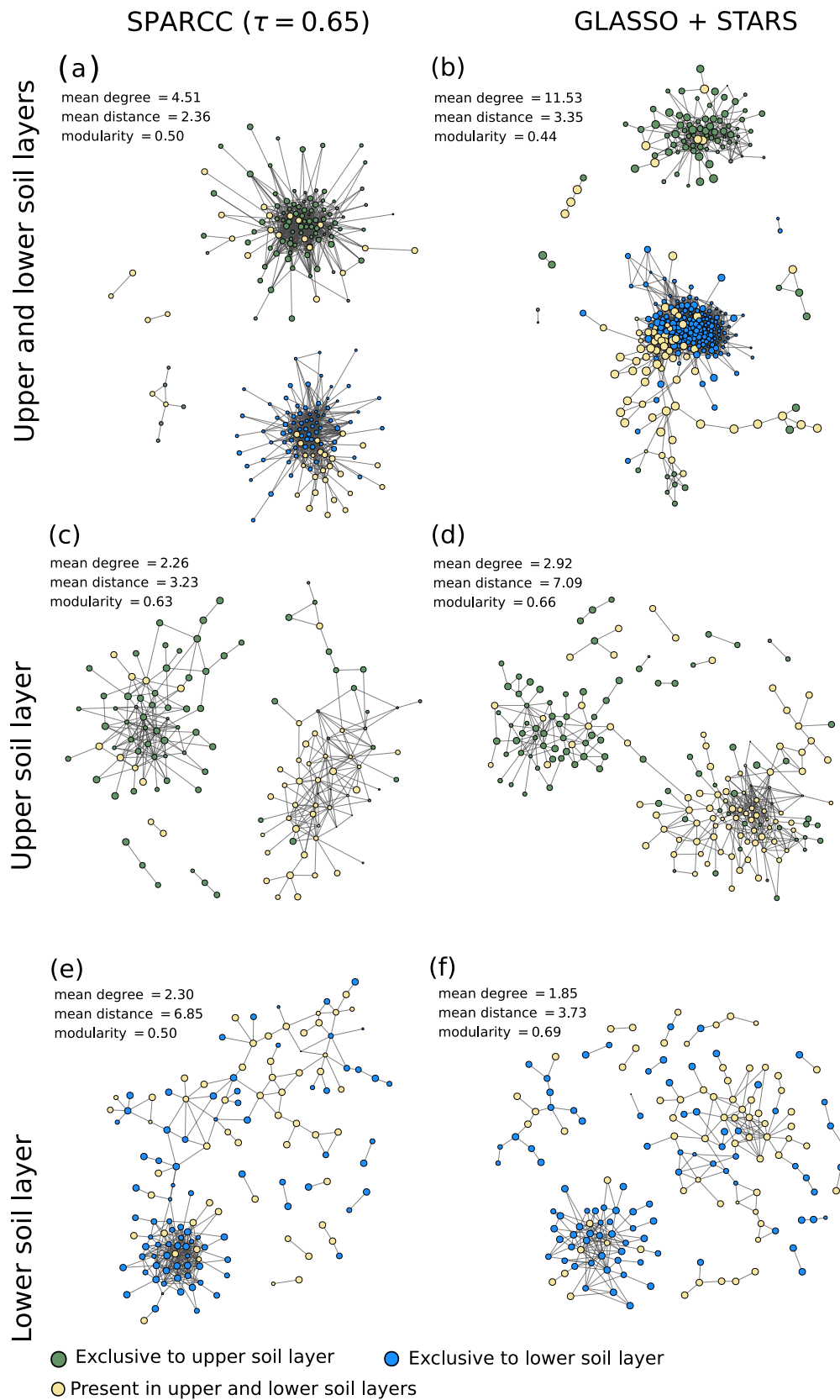
this threshold is applied, the networks are fully connected, with each node (taxa) connected to all other nodes. In these networks the weight of each edge is given by the corresponding pairwise correlation (association) value  $\rho$ . To illustrate this, we provide the histogram of Spearman correlation coefficients for our dataset of the upper soil core, which shows that the weights of most of such edges are close to zero and cannot be interpreted as a strong signal of an association (Fig. 3a). The solution therefore, is to eliminate edges with weights below a given edge significance threshold  $\tau$ . In this way, the threshold is related to the sparsity of the network, where lower values result in very densely connected networks, while high values in networks with fewer edges, see (Fig. 3b). The choice of the threshold value can be either totally arbitrary using a single value for all edges, or it can be evaluated for each edge by comparisons with null models. In the second case, the null model is obtained by shuffling read values among taxa and samples. In other words, a new data set is obtained by randomly selecting values from the original one. The shuffling can be followed by a possible rescaling, due to new total read values in each sample. This procedure is done to break any non-random pairwise association patterns while keeping some of the original characteristics of the data (e.g. sparsity, compositionality and marginal count distributions of taxa). Each pairwise correlation value in the original data set is then ranked against the correlation value of the same pair in the shuffled dataset (Faust and Raes, 2016). This allows one to check the probability of observing such correlation by mere chance (i. e. in similar random data). The p-value determines if the edge should be kept or eliminated. (Fig. 3a) marks with light blue the pairwise correlation values which are left after such a procedure (using 1000 shuffled datasets), and demonstrates that the procedure also works as a sort of edge threshold. Another popular method for edge threshold selection is using ideas from random matrix theory (RMT, mathematical research area that studies matrices with random elements) (Deng et al., 2012). The approach determines at which threshold the generated adjacency matrix no longer has properties compatible to that of a fully random network and uses that as edge threshold. Although the procedure takes away the arbitrariness of the choice, it also selects a hard threshold, and relies on the assumption that biological networks have a completely random structure at the low  $\tau$  limit.

We illustrate the networks obtained by the use of two thresholds  $\tau = 0.7$  and  $0.8$  in (Fig. 4a, c, e) and (Fig. 4b, d, f) respectively. For both thresholds, networks from upper and lower soil layers, as well as from both soil layers combined present two highly interconnected sets of nodes (modules or clusters): While the two sets are completely disconnected for  $\tau = 0.8$ , we can detect candidate taxa that connect the two potential communities for the lower threshold. The choice of threshold also changes important network properties, such as mean degree, mean distance and modularity. This demonstrates that the choice of edge thresholds can influence the interpretation of the network structure in terms of its connectivity and the relevance of individual taxa in the system.

The advantage of using the strategies described so far is that they involve simple computations. However, as described before they suffer

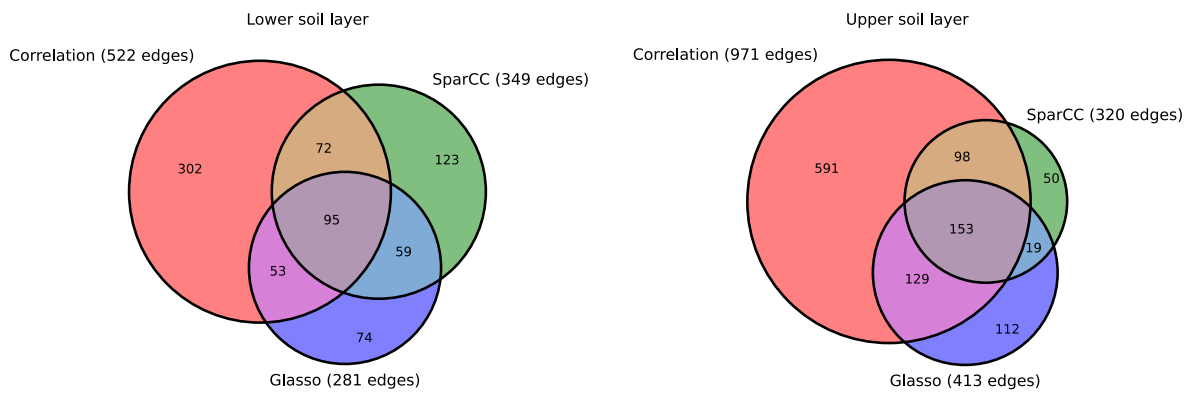


**Fig. 4.** Association networks of microbial taxa in a Beech forest soil at different soil depths constructed from Spearman's rank correlation coefficient at different edge thresholds. Shown are networks constructed from upper and lower soil layers combined (a, b), and from each soil layer separately (c–f), for edge thresholds  $\tau = 0.7$  (a, c, e) and  $0.8$  (b, d, f). Notice how the selected value for the edge threshold changes the number of edges in the network as well as the network structure (for details on preparation of the data set see Supplemental material). The network of the combined dataset (a, b) shows not the total 595, but only 461 ASVs shared with either one of the two other datasets (lower panels). The network with the phylogenetic classifications of nodes is also provided in the Supplement. Network visualization was done with graph-tool (Peixoto, 2014).



**Fig. 5.** Networks of positive associations for microorganisms in a Beech forest soil at different soil depths constructed by different network inference methods. Networks constructed from upper and lower soil layers combined are shown in (a, b), and from each soil layer separately in (c–f). Networks were constructed by (left panel) SparCC or by (right panel) SPIEC-EASI-like approach (transformed with clr, and then using GLASSO combined with Stars). To improve visibility nodes with no edges are not shown.





**Fig. 6.** The edges detected by different network construction methods, using Spearman correlation (edge threshold 0.7), SPARCC (threshold 0.65) and GLASSO (STARS for model selection) with and the overlap among them.

from several limitations: they do not take into account the compositional nature of the dataset and they cannot clearly separate association signals from noise. We can see that the choice of a threshold constitutes a compromise between the detection of spurious links (false positives) and neglecting existent associations (false negatives). The problem of threshold selection is not unique to species co-occurrence networks, but also exist in other areas from inference of gene co-expression to social networks (Couto et al., 2017). The disadvantage of any hard threshold, independent of the selection criteria, is that the thresholding procedure itself may distort the underlying data, strongly affecting the observed network patterns (Cantwell et al., 2020). In particular, as we explain in the next section, this occurs because there is no possible threshold that can distinguish direct from indirect associations. For microbial networks what we often see is that thresholding leads to very dense network structures ('hairballs' (Faust, 2021)). All these reasons make it difficult to derive an ecological interpretation from the pattern in the constructed networks. Several network construction tools were developed to tackle the listed issues, we describe some of them in the next section.

### 3.2.2. Statistical methods that improve reconstruction of ecological networks

In this subsection we mention some methods developed to improve the reconstruction of ecological relationships. However all of these methods depend on the quality of the data and one should not expect signal of interactions in the data if the scale of sampling is inadequate (see Sec. 5.3). In the last decade several ready-to-use tools were made available for network construction. Here we highlight: CoNet, which combines several association measures described in Sec. 3.2.1 (Faust and Raes, 2016); and MENA, which uses random matrix theory to decide on the threshold choice (Zhou et al.; Deng et al., 2012); both are widely used for network construction for microbial soil communities. Another tool which stands out is SparCC due to its ability to treat possible compositionality bias in the data (Friedman and Alm, 2012). It uses log-ratio variance (vlr) to estimate associations, a transformation that presents advantages as it is subcompositionally coherent (see definition in Table 1). More precisely this transformation allows a consistent analysis of any subset of the data, without any contradictions in the results. The SparCC algorithm works under the assumption of few associations between taxa. The final adjacency matrix is evaluated again either by establishing an edge threshold or a comparison with null models.

To improve our ecological interpretation, it is also of interest to distinguish direct from indirect dependencies between taxa. For example, it is well known that two taxa may show a significant correlation because both are under the influence of a third taxon, although they do not directly interact with each other. The difficulty is that no choice of edge threshold for correlations is capable of distinguishing between direct and indirect associations. To solve this issue a new set of

algorithms for network construction was proposed based on probabilistic graphical models (Kurtz et al., 2015; Fang et al., 2017; Yang et al., 2017; Yoon et al., 2019; Jiang et al., 2021). These models represent complex probabilistic relations in the form of diagrams (graphs), where nodes represent taxa and edges represent conditional dependencies between them. Such a graph can be easily constructed for Gaussian multivariate data and for this case it is known as Gaussian graphical model (other names such as partial correlation network or concentration graph are also used). The conditional dependencies are evaluated from non-zero elements of the inverse covariance matrix (Bishop, 2007). It is important to note that strictly speaking for non-Gaussian settings, as is the case for microbial data sets, the relationship between the inverse covariance and conditional dependencies is not known. However, the approach is still used as it is expected to be informative beyond its standard domain (Liu et al., 2009). The inverse covariance is computed with a popular statistical method known as glasso, lasso regularized maximum likelihood estimator. In this case, the sparsity of the network is tuned with a regularization parameter,  $\lambda$ , with small values of  $\lambda$  corresponding to densely connected networks and large values to fully disconnected networks. The method in general selects a range of  $\lambda$  values and generates several networks. The next step is to select the best model, considering its complexity and data fitting ability. The most well known model selection algorithms are BIC, EBIC, Stars. SPIEC-EASI (Kurtz et al., 2015) was one of the first tool packages based on graphical models and therefore is the most well-known. However it is not the only package based on a graphical model, there is a series of other recently developed tools among them: gCoda (Fang et al., 2017), SPRING (Yoon et al., 2019) and BC-GLASSO (Jiang et al., 2021). We also illustrate the outcome of network construction using SPARCC and a SPIEC-EASI-like (clr transform, glasso and Stars) algorithm in (Fig. 5). For simplicity, we do not show nodes which have zero edges in this image. Again, as in (Fig. 4) it is possible to identify the formation of two clusters, one with the taxa unique to the respective core depth (nodes shown in blue and green), and another with taxa shared by the two cores (nodes shown in red). Despite the differences between methods, they capture similar patterns, although only half of the edges are shared among them, see (Fig. 6). Nevertheless, the ecological processes behind the observed structure remain unclear. Note that these methods perform rather poorly when used on artificial data sets produced even from simple models of population dynamics such as gLV (generalized Lotka-Volterra) when different environmental conditions are used (Berry and Widder, 2014; Hirano and Takemoto, 2019). One of the reasons is that indirect dependencies between taxa can also appear due to environmental factors, for example due to niche overlap. It is possible to introduce environmental parameters as additional nodes in the network (Faust, 2021). Another form to deal with environmental confounders is to use a model based approach as done by mLDM (Yang et al., 2017), a tool introduced recently based on an hierarchical Bayesian statistical framework.



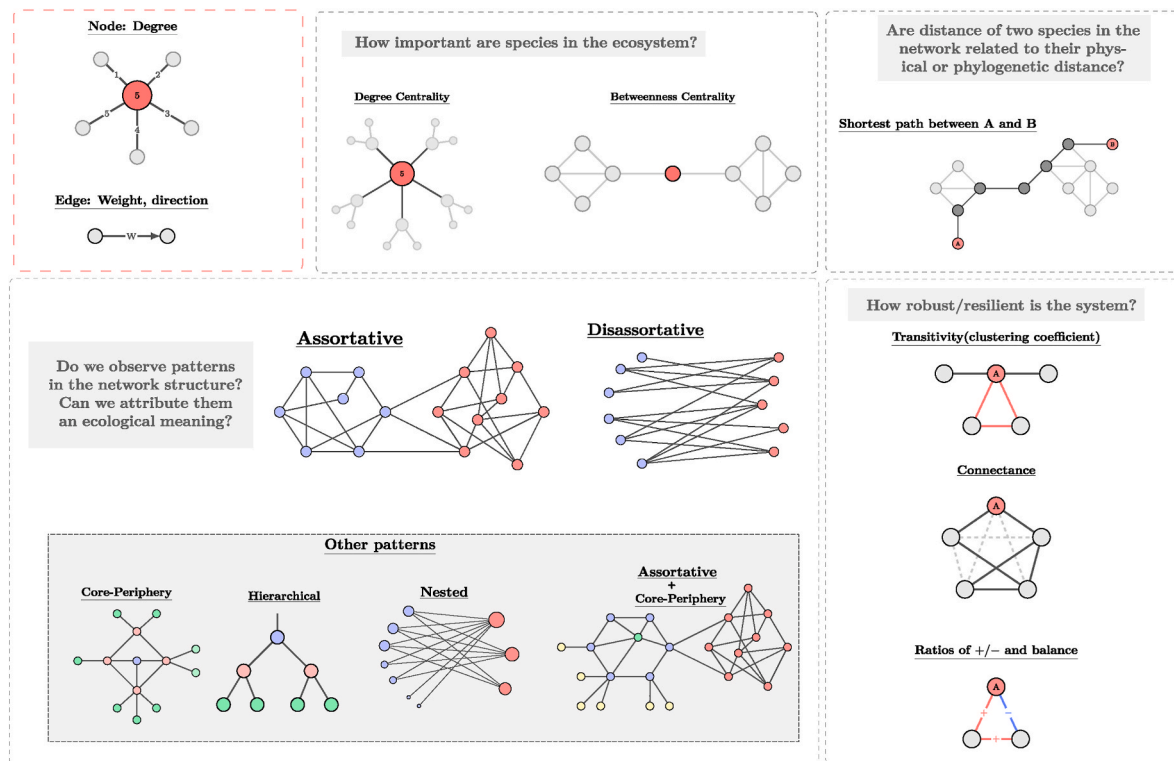


Fig. 7. Some network properties that can be used to characterise microbial communities, with the respective examples of questions that can be addressed by their use.

### 3.3. Summary of test dataset analysis

We used microbial co-occurrence data from upper and lower soil layers of a European Beech forest to test whether the choices for a specific network inference method (i.e. correlation-based methods with different edge thresholds, (Fig. 4), a method using log-ratio transformations to address a possible compositionality bias, SPARCC, and a method considering indirect associations, GLASSO and Stars; (Fig. 5), or the inclusion of an environmental factor (soil depth) would influence the structure of the constructed network, and the conclusions that can be drawn from it. The datasets of upper and lower soil layers together consisted of 466 ASVs, from which 162 were exclusively occurring only in the upper soil samples, and 183 only in the lower soil sample, and 121 were shared between the two (Fig. S5). We used this taxon-specific information to investigate if and to what extent the resulting network structure was influenced by taxon preference for upper or lower soil, two habitats which differed in many environmental conditions, like temperature, moisture, or substrate input. To visualize this, we categorized the nodes into those which exclusively occur in one of the soil layers after filtering ASVs with low read numbers (Fig. S5), which could be interpreted as ‘specialists’ and those which occurred in both soil layers, which could be interpreted as ‘generalists’. Irrespective of the method used, networks which were constructed including the environmental gradient, i.e. based on the combined dataset from both soil layers, always exhibited two highly connected sets of nodes (clusters), one containing all taxa dominant in the upper and the other those from the lower soil layer. This demonstrates that taxa which co-occur due to similar environmental preferences strongly dominate the network structure, potentially obscuring other ecological relationships.

Ruling out this environmental factor, i.e. constructing networks for each layer separately, revealed an interesting pattern. Here again we see in all methods that two clusters (modules) formed, of which one was dominated by the specialists and the other one dominated by the generalists. This is an interesting finding, and the next step would be to

think about its interpretation, or to formulate a testable hypothesis for future research based on this result. Both, however, go beyond the scope of this paper. Here, we just would like to point out that this pattern was conserved, in principle, in all the methods (Figs. 4 and 5). The only observable limitation is that in the most sparsely connected network, i.e. the correlation network with a high edge threshold in the lower soil layer (Fig. 4f), this finding may not be visible anymore owing to the small number of connections. While the coarse pattern seems to be robust and independent of the method, however, we also observed that the majority of links were different between the different methods (Fig. 6). For a more detailed analysis, where pairwise associations between individual taxa are of interest, one should thus consider that methodological choices could influence conclusions at that level. For a display of networks with nodes coloured according to their phylogeny see (Fig. S7 and S8). Interestingly, however, both correlation ( $\tau = 0.7$ ) and SPIEC-EASI like methods identify the same central taxa that mediated the connection between the modules of the upper soil layer (ASV\_32a\_pom, a *Proteobacteria Rhizobiales*). From this small test we conclude that strong patterns in the dataset will be captured by most methods, while the difference in the detailed connection pattern may be quite different, and conclusions on individual connections should thus be done with care.

One interesting question here would be if it was possible to reveal these clustering patterns also using traditional ordination approaches. In principle, ordination-based analyses, such as nonmetric multidimensional scaling (NDMS) or correspondence analysis (CA) can explore and visualize community dissimilarities between samples (i.e. ‘beta-diversity’) or, if turned around, dissimilarities between taxa in their occurrence pattern across samples. One could argue that the latter approach may reveal a potential clustering of taxa in a similar way as co-occurrence network analysis. The reduction of the high dimensional nature of the dataset realized by ordination methods, however, may discard valuable information, affecting the conclusions. Network models, on the contrary, work in the original multidimensional space

**Table 2**

Examples of ecological questions that can be assessed with particular network properties depending on the meaning of the links in the network.

Network property	If network edges can be interpreted as interactions	If network edges can be interpreted as co-occurrences
<b>Importance of species</b>		
Node degree	Are there species more generalist in their interactions than others? How important are species for ecosystem functioning? (Röttgers and Faust, 2018)	Are there species with broader niche preference than others?
Degree centrality	Do highly connected taxa (i.e. hubs) support higher levels of ecosystem functions?	
Betweenness centrality	Are there taxa which act as brokers, transmitting the effect of multiple interactions, as many paths between taxa pass through them?	
<b>Cluster (community) structure and its ecological interpretation</b>		
Assortative patterns (Modularity)	Do we find groups of species that tend to interact more among each other than with other species (specificity of interactions)? (Torrecillas et al., 2014) What characterizes these groups?	Do some species tend to co-occur with each other more often than with other species? (Barberán et al., 2012; Williams et al., 2014; Toju et al., 2016; Shi et al., 2020) How are environmental factors reflected in the co-occurrence patterns? (Deng et al., 2012; Karimi et al., 2016; Röttgers and Faust, 2018; Eldridge et al., 2020; Zhu et al., 2020; Hernandez et al., 2021) Do some factors (i.e. invasion processes) enhance a cluster pattern of co-occurrence? (Zhang et al., 2021) (invasion can be seen as a perturbation) Do species co-occurring with few species tend to co-occur with species that co-occur with many others? (Co-occurrence patterns mirror confounders, e.g. relative abundance of species or broad/narrow niche preference)
Disassortative patterns (Nestedness)	Do interaction specialists interact more with generalists than with each other? (Wei et al., 2015)	Do species co-occurring with few species tend to co-occur with species that co-occur with many others? (Co-occurrence patterns mirror confounders, e.g. relative abundance of species or broad/narrow niche preference)
<b>Distances in the network</b>		
Path distance (of pairs or average)	Are the interactions overall tight (direct) or loose (indirect)? (Song et al., 2021; Zhan et al., 2021)	Can distances in the network help us to understand the community assembly process? (Morueta-Holme et al., 2016)
<b>Network stability (Robustness and resilience)</b>		
Connectance	Which proportion of all potential interactions are actually realized (complexity/stability of the network)? (May 1972; Thébault and Fontaine, 2010) Are there redundant pathways in the interaction network?	
Transitivity	Do we find positive (negative) feedback loops? (Coyte et al., 2015)	
Ratio of +/–	What is the ratio of cooperation to competition? (Saiz et al., 2017; Hernandez et al., 2021)	

and therefore are more appropriate to detect such patterns. To test this, we applied taxa-based NMDS to the same dataset we used for the network analysis. While NMDS was also able to separate lower and upper soil specialists in the combined dataset of both layers, the method could not detect a generalist and a specialist-dominated cluster in the dataset of each separate layer as observed by network analysis (see Fig. S9).

#### 4. Network analysis

Network analyses can be applied to different types of data, e.g. data of soil samples collected in the field, root samples collected in different experimental settings, or even datasets generated by simulations (Berry and Widder, 2014; Ofaim et al., 2017; Hewavitharana et al., 2019). These datasets can create a variety of different networks such as resource competition networks (species and resources as nodes), food-webs (antagonistic interactions), mutualistic networks (positive interactions) or general co-occurrence networks. In this section we will illustrate the properties that can be central to formulate testable hypotheses for the biological system, see Fig. 7.

##### 4.1. Properties of the nodes

The importance of a species in an ecosystem can be classified according to how many other species it can directly or indirectly influence. In a network of interactions the direct influence of a node is given by its degree (number of edges), see Fig. 7. Node with large degree, or high connectivity, are known as hubs and represent species generalists or 'key taxa'. While nodes at the network border, classified as peripherals, represent specialists (Olesen et al., 2007). This definition can be extended to encompass more complex network structures and the nodes

can be classified in four simple categories: local or global-hubs, connectors or peripherals (Guimerà and Nunes Amaral, 2005; Olesen et al., 2007).

Although this concept to interpret nodes with a high connectivity as 'key taxa' stems from plant-animal interaction networks, it was later adopted in a large and still growing number of soil ecological studies to identify 'keystone taxa' in soil microbial datasets based on co-occurrence networks (e.g. (Zhou et al., 2011; Lu et al., 2013; Lupatini et al., 2014; Jiang et al., 2016; Ling et al., 2016; Shi et al., 2016; Chen et al., 2018; Qi et al., 2019; Varsadiya et al., 2021; Yuan et al., 2021)). Other attempts to classify keystone taxa in soil microbial co-occurrence networks select nodes with high degree and high closeness centrality, but low betweenness centrality (Banerjee et al., 2018, 2019; Lin et al., 2019), guided by a theoretical analysis of co-occurrence networks constructed from generalized Lotka-Volterra equations (Berry and Widder, 2014). It must be noted that one of the selection criteria in this approach, low betweenness centrality, differs fundamentally from the selection criteria of high connectivity defined in (Olesen et al., 2007), so that the two approaches would likely lead to the identification of different key-note species when applied to the same dataset.

Most important, however, both (and other) approaches to identify 'keystone species' from soil microbial datasets currently suffer from a lack of conceptual justification of why a node with a certain topology, for example a highly connected ('hub') node, should indicate a 'keystone' role of the respective taxa in soil microbial co-occurrence networks (Röttgers and Faust, 2019). A 'keystone species' is defined as 'a species whose impact on its community or ecosystem is large, and disproportionately large relative to its abundance' (Power et al., 1996). This means that the removal of such a species will have a detrimental effect on the community or ecosystem, a behavior which has been computationally shown when removing hub species in ecological

networks (Dunne et al., 2002; Memmott et al., 2004). A requirement for transferring this conclusion from the constructed network behavior to ecosystem behavior, however, is that the edges in the constructed network represent vital and essential links between species in nature. While an edge in an interaction network, such as the plant-pollinator networks described in Olesen et al. (2007), reflects with a certain likelihood an essential interaction between two species, so that the removal of a hub species may indeed lead to the loss of other species, this cannot be assumed for soil microbial co-occurrence networks. Even if the simulated removal of a node in a microbial co-occurrence network affects the computed network structure, it would not need to have the same effect in nature, as species who co-occur do not necessarily depend on each other.

Taken together, we suggest to stay away from classifying taxa as keystone species based on soil microbial co-occurrence analysis, without further conceptual and experimental validation (Banerjee et al., 2018, 2019; Röttgers and Faust, 2019). It is clear that the need for a different interpretation of edges in a co-occurrence network compared to edges in an interaction network (Table 2) must inflict on the interpretation of the potential role of nodes. We therefore suggest to redefine the potential ecological roles of hub or connector species for soil microbial co-occurrence networks, rather than adopting the keystone species approach from plant-animal interaction networks. Based on a concept that suits the characteristics of soil microbial datasets such a redefinition (for example a highly connected node can be interpreted as a taxon with a broad niche preference, Table 2) may be helpful for identifying and exploring microbial taxa with potentially interesting ecological roles in soil ecosystems.

#### 4.2. Average path length of the network and degree distribution

The architecture of biological networks is often far from random. In general, interaction networks, such as food webs and mutualistic networks are known to have broad degree distributions, with few very connected nodes (hubs) and the majority of species that interact only with few others (Camacho et al., 2002; Montoya et al., 2006). This degree distribution also leads to short path length between nodes, meaning that randomly chosen species have only few edges between them (Williams et al., 2002). The questions are how and why such types of architecture develop for microbial communities and their relation to stability or community assembly mechanisms. In the case of co-occurrence networks there is need to disentangle possible reasons behind the observed patterns, such as evolutionary history, dispersal dynamics or negative/positive biotic interactions (Weiher et al., 2011; Morueta-Holme et al., 2016; Goberna et al., 2019). Relevant questions that still have to be approached are for example: how relative distances between taxa in the network are related to either phylogenetic similarity or even to physical distance if environmental gradient is present.

#### 4.3. Cluster (network community) detection

Ecological networks can be composed of hundreds, or sometimes even thousands, of nodes. It is natural, therefore, to wish for identification of representative high-level subsets or groups to organise this complex information. However, we do not want any partition, but the one that can inform us of underlying functional, evolutionary or environmental units within these ecosystems. This is an easy procedure if we have metadata on such groups beforehand, e.g. division of microorganisms in groups predominant in certain environments (Eldridge et al., 2020). However, in most cases we lack this information and the underlying partition is precisely what we would like to learn about the system. The problem therefore is to determine node groups based on network topology, for instance by checking the density of connections among them (Newman, 2010). This is known as cluster detection or community detection. Note that this term is used for 'network communities', which consist of several similar nodes, not to be confused with

microbial communities. This task consists of: (1) finding a natural division in the network, without being able to look at its structure; and more importantly (2) evaluating if these patterns are statistically/ecologically meaningful or if they can appear by mere chance.

Certainly the most popular algorithm developed for cluster detection (1) relies on modularity maximization, which formulates the task of community detection as an optimization problem. This is done by using a mathematical function known as modularity to assign a score to each division of the network (Newman and Girvan, 2004; Newman, 2004, 2006). Although the method has attracted substantial popularity due to its simplicity, it has drawbacks which are often ignored in applications. While the method works well when a clear division exists, it is also known to find spurious partitions when there are none to be found, e.g. in completely random networks (Guimerà et al., 2004). This conflicts with our objective (2) of selecting only statistically meaningful partitions. Moreover, a comparison with a null model cannot solve issue (2) since it can only rule out or accept the null model (random network structure), but cannot assess if the found partition is in fact statistically significant. This is because finding enough evidence to reject a null model is a much weaker requirement than having enough evidence to accept a particular network partition.<sup>1</sup> Due to these problems several other methods have been proposed to extract meaningful communities from networks (Fortunato, 2010; Fortunato and Hric, 2016). We highlight the inference methods based on stochastic block models (SBMs) (Karrer and Newman, 2011; Abbe, 2018; Peixoto, 2019). Their main advantages is their ability to distinguish signal from noise, taking care of (2) in a principled manner, and possess also the ability to detect not only patterns based on assortative structures (Newman, 2004), but others, such as core periphery, or self-similar hierarchical patterns (Fig. 7).

The kinds of patterns identified by cluster detection have been proved useful for understanding ecological networks: Structural assortative as well as nested patterns have been linked to stability in interaction networks (Thébault and Fontaine, 2010; Grilli et al., 2016). In association networks clusters have mainly been used to identify and separate different niches occupied by microorganisms.

#### 4.4. Network stability

Over the past decades an increasing number of studies have approached the question of ecosystem stability by analyzing the architecture of biological interaction networks (Montoya et al., 2006; Okuyama and Holland, 2008; Bascompte, 2009; Landi et al., 2018). The stability is characterized by how systems respond to disturbances, and whether or not they can return back to their original function after being perturbed. It can be evaluated with the use of different criteria, measuring the recovery process (resilience), endurance (robustness) or the flexibility (persistence) of an ecological system (Hodgson et al., 2015; Landi et al., 2018). Although sometimes these terms are used interchangeably, what is important to note is that the type of perturbation and the type of analysis may depend on the research question. For example one can investigate how a community endures species extinctions (robustness). Alternatively, one can observe how quickly the multi-species dynamics recovers to a decrease of species abundances or to invasion of a new taxa (resilience).

A significant fraction of the studies on stability are theoretical and use numerically generated datasets (e.g. gLV, generalized Lotka-Volterra model). They address a long standing debate in ecology on the relationship between the ecosystem's complexity (e.g. number of connections in an ecological network) and its stability. The controversy in the field started around the 1970s when theory showed a conflict between the two aspects (May, 1972, 1974; Sales-Pardo, 2017). Currently we

<sup>1</sup> For example, a network partition where only one of hundreds of detected communities is statistically significant will be enough to reject a null model, but it would be mostly statistical noise.

know the network properties that foster stability, but only for some types of networks. For example, nestedness and connectivity seem to promote stability in mutualistic (Thébault and Fontaine, 2010) and resource competition networks (Wei et al., 2015), while for food-webs a compartmentalized structure, low connectance seem to bring stability (Neutel et al., 2007; Gross et al., 2009). In the presence of both positive and negative interaction evaluating the stability is more difficult. It is speculated that the presence of mutualistic interactions may destabilize ecosystems, since they generate co-dependencies (Coyte et al., 2015). Trophic interactions, on the contrary, were shown to promote stability due to negative feedbacks (Coyte et al., 2015). Due to the difficulties presented above, claims about stability are even more unreliable for microbial co-occurrence networks, where there is not only the possibility of different interaction types but the role of links is less clear.

On the experimental side, studies have focused on characterizing how the ecological networks and ecosystem function change in reaction to a perturbation, or along gradients of environmental conditions. Some of these studies showed that connectivity, density of the links and transitivity decrease in disturbed environments (Karimi et al., 2016, 2017; de Vries et al., 2018). Going one step further, some studies used soils that were exposed to different treatments or perturbations to investigate a possible link between the network structure and the functionality of microbial communities, or even ecosystem multifunctionality (Wagg et al., 2019; Yuan et al., 2021; Qiu et al., 2021; Jiao et al., 2022). However these studies based their conclusions primarily on correlations between network complexity and functionality measures, and were thereby unable to provide any mechanistic foundations for their claims. Again, we recommend to be cautious to draw causal relationships from correlative pattern, which we think is especially problematic for correlations between two highly aggregated measures, such as network complexity and soil or ecosystem multifunctionality, as they naturally may share a high number of co-correlating factors and interdependencies. Therefore, instead of trying to correlate given functions with network patterns, we suggest to use studies encompassing different treatments or perturbations to test hypotheses that for example link biodiversity and complexity. Such hypotheses can be formulated based for example by randomly (or systematically) removing nodes from an unperturbed network and recording how the connectance of the network changes. These hypothesis can then be validated empirically by comparing it to networks constructed from observational data of a different treatment (where a perturbation or stress is applied), checking if the loss of diversity affects the network in a similar way. Such comparisons can also allow to analyze if perturbations or stress leads to loss of edges without a loss of corresponding taxa (Valiente-Banuet et al., 2015) or there is a formation of new patterns. Formulation of hypotheses on the resilience of a system can be an important way to validate network models inferred by the current methods. For example, tracking the presence of known keystone species and how the network structure is affected by them can be of interest. Finally, we would like to mention the appeal of using co-occurrence networks as indicators of ecosystem's quality and function as proposed by Karimi et al. (2017).

## 5. Challenges and way forward

### 5.1. Specific challenges in soil microbial ecology

Although there are several possible drivers behind the formation of microbial communities, species-species interactions have taken the spotlight of current research on taxa co-occurrence patterns. Despite its appeal, the detection of interactions from this type of data has proven to be extremely challenging. Moreover, it is still unclear whether our limitation in spotting species-species interactions can be attributed to an inefficiency of current inference methods, or to the weakness of interaction signal in abundance (co-occurrence) data (Blanchet et al., 2020).

Clearly, besides the potential signal of species interactions, association networks contain other valuable information about relationships

within an ecosystem and can help us to understand ecological processes behind the high microbial diversity characteristic to soil. However it is important to emphasize that any network model and its analysis is only as good as our interpretation of it. Therefore, it is desirable to include an interpretation of all associations represented as edges to the most possible extent. This is particularly hard for soils, since they are characterized by their immense biological diversity and heterogeneous physical structure. Here we summarize the important soil characteristics that may impact network construction, interpretation and analysis, and subsequently discuss how we can use experimental design and statistical methods to improve our ability to reconstruct ecological relationships.

#### 5.1.1. Soil microbial datasets, what to include as nodes?

Soil harbours one of the highest biodiversity of all microbial environments on Earth (Thompson et al., 2017; Walters and Martiny, 2020), raising the concern of how many species can we include in the network model. The characterization of this diversity is not easy, and depends on the number of samples used and the design of the sequencing experiment (Hermans et al., 2019). Increasing the sampling effort by including more replicates results in discovery of larger numbers of rare taxa. Although it is considered that these taxa significantly contribute to ecosystem processes (Jousset et al., 2017; Hermans et al., 2019), it is still unclear how they can be included in the network construction, as taxa with low prevalence lead to spurious associations, and are thus often completely discarded in the data preparation step (see Sec. 3.1). In addition, one has to be aware that the amplified DNA may also belong to relic (exogenous) DNA, which persist in soil for several years (Alteio et al., 2021).

Fungal-bacterial associations are another important component of the soil ecosystem (Peay et al., 2016). The understanding of fungal diversity across biomes remains more limited due to challenges in sequencing the ITS region of fungal DNA. Integrating the community sequence data from fungi and bacteria to construct combined networks will certainly expand our ability to explain the observed community patterns in the soil environment. However, the construction of such cross kingdom networks faces new challenges, since the analysis would have to properly combine two compositional datasets. Despite this, cross-kingdom networks have been constructed by applying correlation-based approaches on merged 16S rRNA gene and ITS relative abundance datasets (Delgado-Baquerizo et al., 2020; Jiao et al., 2020). We would however argue against such an approach, as combining correlations within and across compositional datasets in the same network may lead to a spurious connection pattern. Possible alternatives propose the use of centered log ratios individually on each dataset before merging them together (Tipton et al., 2018), transform the relative abundance data into a presence/absence dataset (which implies a loss of information) or normalize using qPCR data (Alteio et al., 2021). In addition, one can, depending on the research question, consider constructing bipartite networks, which only contain edges between the kingdoms (Montesinos-Navarro et al., 2012; Toju et al., 2015), or multilayer networks, which contain different types of edges (Pilosof et al., 2017).

#### 5.2. Soil microbial networks: what type of associations should we expect?

Soil provides a unique environment for microbes, because it is spatially structured and chemically and physically heterogeneous across several scales. Within the complex soil matrix, numerous edaphic properties impact the structure of the microbial communities therein, including pH, salinity, temperature and moisture (Frindte et al., 2019; Zheng et al., 2019). Additionally, temporal dynamics such as weather patterns, root exudation and other seasonal inputs of organic material may influence the structure and activity of a microbial community (Kuznyakov and Blagodatskaya, 2015; Chernov and Zhelezova, 2020). As a result of soil heterogeneity and temporal fluctuations, it is critical to consider that sequence-based approaches reveal only a snapshot of the microbial community present at a given time. Together, these properties



of soil as a habitat for microbes influence the possible scope for interpretation of microbial association networks.

Heterogeneity of soil samples is a major source of concern in the development of reliable inference methods to uncover ecological relationships. We see the heterogeneity impacting the sampling and inference in two different ways. First, it can lead to hidden physical or chemical differences between apparently similar samples (Armitage and Jones, 2019). Since environmental drivers lead to strong differences in the community structure, such drivers may dominate the detected association signal. In this case, additional environmental measurements can help to interpret the observed patterns. The second concerns the large volume of each sample used in the analysis. Nucleic acids extracted for sequencing studies are derived from amounts of soil ranging from 250 mg to 2 g of material. Very often, they are aliquots from an even larger soil volume, for example several soil cores that have been mixed and homogenized to be representative for a certain field site or plot. Soil microbes, however, interact with each other at the scale of tens of  $\mu\text{m}$  (Raynaud and Nunan, 2014). It is obvious that co-occurrence of microbial taxa across soil samples taken at a scale which is by orders of magnitude larger than the scale at which they interact, does not necessarily mean that they interact with each other. In line with this, community assembly theory suggests that the larger the sampling units scale the more environmental filtering effects dominate species co-occurrences (Kraft et al., 2015). Although the mix of ecological factors operating at larger sampling scales obscures causal distinctions, studies have shown that networks generated with this type of data hold predictive value for central ecosystem functioning (Shi et al., 2016; Wagg et al., 2019). We should, however, lower our expectations to capture a clear signal of microbial interactions, since even small soil samples may contain numerous metacommunities - small consortia linked by dispersion and diffusion (Armitage and Jones, 2019). Two microbial taxa can be separated by unconquerable distances even in a small piece of soil given its microscale physical structure and the spatial heterogeneity of microbial niches. In view of this complexity it is controversial if in such samples biological interactions can leave any signal at all. For example, models suggest that negative interactions, such as competition, can be completely hidden at scales much larger than the scales of the interactions themselves (Araújo and Rozenfeld, 2014). In summary, the associations described in co-occurrence networks from standard soil samples have to be carefully interpreted. Such networks are expected to be dominated by associations driven by the environment, however the obtained network structure may be a result of several unknown factors. In relation to interactions among organisms, especially the signal of competitive interactions will be obscured. In the next section we give suggestions on how to disentangle the different factors behind a given co-occurrence dataset and suggest alternative or complementary experiments.

### 5.3. Reconstructing networks of ecological relationships

#### 5.3.1. Network interpretation is influenced by experimental design

In general, edges in networks based on the characterization of variability of microbial communities in multiple soil samples can inform about species co-occurrence patterns across these samples. If the objective is to infer microbial interactions, we advise to work with datasets extracted from specific micro-environments where interactions are prone to occur. Such environments can be, for example, mycorrhizal fungi within plant species root tips (Montesinos-Navarro et al., 2012), soil aggregates (Wilpiszeski et al., 2019; Szoboszlay and Tebbe, 2021) or synthetic systems (Cordero and Datta, 2016). In the case of soil aggregates, it is possible that niche partitioning also leads to spatial segregation with their pore structure. However, even in this case, aggregates are considered to provide enough physical proximity and a traversable matrix necessary for competitive and facilitatory exchange. This is critical as aggregates contain microbes' primary source of nutrients and their occupancy enables the establishment of stable source populations.

Alternatively, to improve interpretation one can amplify and sequence specific functional genes in place of taxonomic marker genes, to target microbial groups with known traits, as was done in Jones and Hallin (2019) to find relationships between nitrifiers in soil. Certainly, in networks based on functional information, such as metabolic profiles associated with distinct groups of microorganisms that transform metabolites, edges more likely represent potential interactions among taxa (Ofaim et al., 2017; Hewavitharana et al., 2019).

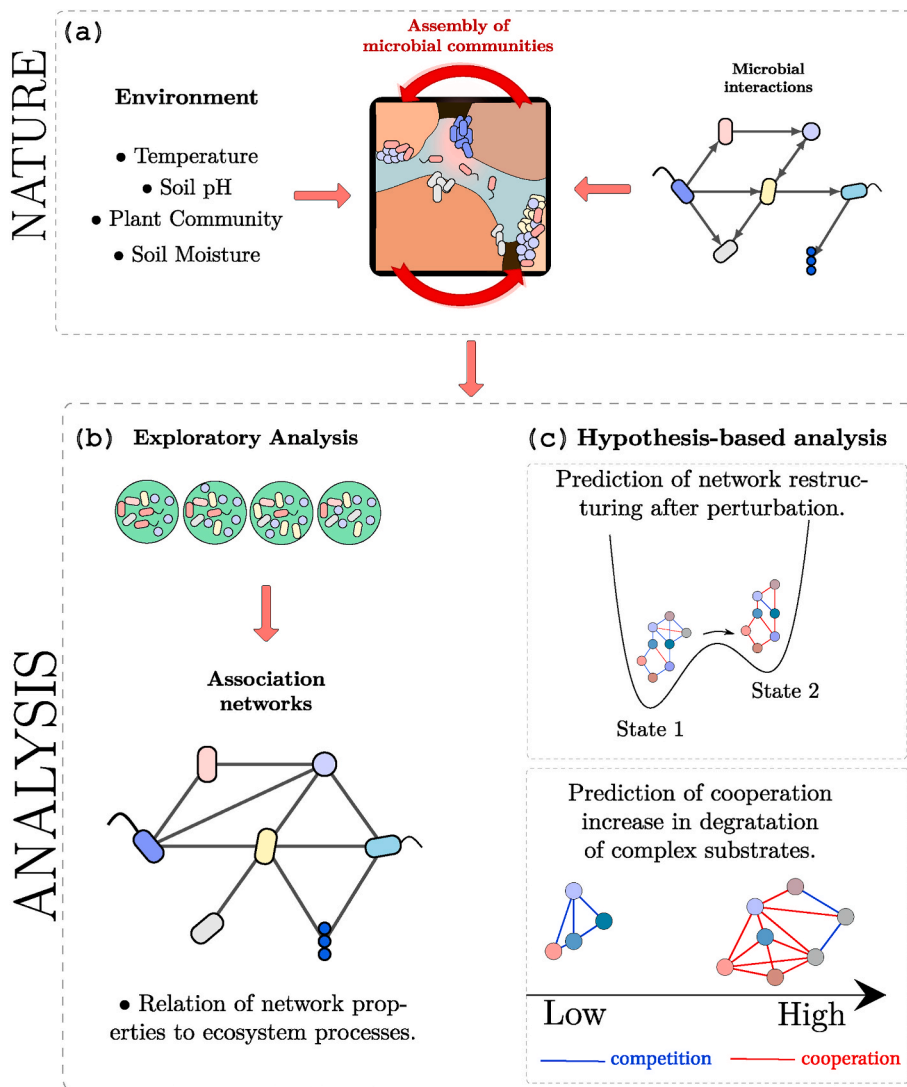
Another important factor that influences our interpretation of edges is the distribution of the heterogeneity across samples. Microbial datasets collected along large environmental gradients can reveal differences in environmental niches occupied by microorganisms. Indirect associations through environmental factors will also be predominant if more than one treatment is considered within the same network. For example, datasets encompassing different seasons, depths, temperatures and biomes will be prone to generate co-occurrence patterns derived from the different levels of the treatment, but will also blur the patterns of co-occurrence among taxa within the levels of the treatment, as is also demonstrated in our example of microbial networks across different soil depths (Figs. 4 and 5, Sec. 3.3). Therefore, the inclusion (or not) of treatments in network construction will impact which ecological processes can be assessed. The focus of the hypothesis to be tested will define whether networks should be constructed from combined data of different treatments, or whether separate networks should be built for each homogeneous environment.

Several strategies were proposed to understand which one of the established edges are in fact due to the indirect influence of environmental factors (Faust, 2021). A strong imprint of environmental factors or treatments often leads to clusters or modules being assorted by these factors. This can be tested for by categorizing taxa according to the environment or treatment where they predominantly occur and check if the network structure is shaped by these categories (Jiao et al., 2020), (Figs. 4 and 5). If there is an obvious difference between treatments or environmentally different sites, depending on the research question, these could be compared and studied separately. In that way, the influence of the environment on the network structure would be reduced, allowing other potential ecological factors, such as interactions, to become more visible in the network structure. Alternatively environmental factors can be directly integrated into the network construction, for example by representing them as nodes in the network, enabling one to examine their influence on the other nodes (taxa) (Faust, 2021). In the same way, data on the distance between samples can also be used to check if the links result indirectly due to spatial dispersion patterns (D'Amen et al., 2018; Goberna et al., 2019).

#### 5.3.2. Network interpretation can be improved by statistical methods

In summary edges in a network construction can appear due to: (1) species-species interactions, (2) association mediated by the environmental factors (niche overlap), (3) spatial variability due to dispersion dynamics, (4) due to influence of a third interfering species present in the network (common to methods based on correlations), (6) spurious associations due to inappropriate data handling (errors in dealing with sparsity and compositionality) and (7) noisy measurement. It is clear that the possible interpretations of an ecological network rely not only on the data but also on the statistical methods used for its construction, especially with respect to items from (4–7) listed above. As we have described in Sec. 3 these methods can significantly differ in their assumptions and strategies, and therefore a single dataset can produce networks with very distinct sets of edges, see (Fig. 6). Part of the difficulty current approaches meet is addressing the nature of amplicon-sequence data itself (item 6), we advise the use of qPCR to reliably infer the abundances of taxa from sequence read numbers (Alteio et al., 2021). We emphasize again here that correlations are unreliable to establish interpretable edges, since they do not offer a clear criteria for distinguishing signal from noise and to distinguish direct from indirect associations, see Sec. 3. This can be improved with model-based





**Fig. 8.** Use of network analysis in exploratory vs hypothesis driven research. (a) Taking into account that community composition is shaped by underlying ecological mechanisms driven by the physical environment or by interactions among taxa. (b) In general, the measured co-occurrence patterns are used to construct association networks, which are different from the interaction networks. (c) Hypothesis-driven research is key to design comparative experiments aiming to isolate and contrast the community assembly mechanism of interest. In particular, established ecological theories can complement hypothesis generation.

inference approaches and incorporation of additional information (environmental factors, distance between samples) into the network construction, facilitating the edge interpretation (items 1–3). At the moment, the most promising direction is biologically relevant generative models, such as those based on generalized-Lotka-Volterra equations (gLTV) (Bucci et al., 2016; Gibson and Gerber, 2018). Although these methods can be computationally demanding and are still not sufficiently efficient for large data sets, this is a very rapidly advancing area where alternatives constantly appear.

### 5.3.3. Questions that can be addressed in different networks

The nature of the edges in a network affects the type of questions that can be addressed with network analysis. Ecological questions can be assessed with network analyses when particular network properties match the expected ecological patterns. Thus, when it is expected that groups of taxa tend to interact more among them than with other taxa, network properties such as network community structure (modularity) can be useful to test that specific hypothesis. In some cases, a given network property can be informative, independent of whether the edges represent interactions or only have a more broad meaning such as co-occurrence (including niche overlap). In that sense, it might be interesting to test whether there are groups of taxa that tend to interact or that tend to co-occur. However, other network properties can only meaningfully be evaluated when we know that edges describe

interactions among taxa. For instance, that might be the case of the average path length among nodes, which can reflect the contribution of complex indirect (i.e. high average paths length) vs. direct interactions among taxa (i.e. short paths), but has a less useful interpretation in terms of co-occurrence. As explained in the previous subsections the experimental design from which the data is obtained and the statistical methods used will constrain the meaning of the edges in a network, which in turn will condition the network properties that would be useful to assess specific ecological questions (Table 2).

### 5.3.4. Descriptive, exploratory or hypothesis-based approaches?

Recently we have observed an increase in the number of articles which use microbial networks as a form of data visualization for co-occurrence patterns in soil. These studies define the edges of networks as ‘co-occurrences of microorganisms’ (or even incorrectly as microbial interactions), and have a mainly descriptive character. Without presenting hypotheses of biological mechanisms behind the observed statistical patterns in co-occurrence data, such studies are of little interest to soil microbial researchers. In the previous section we presented several research questions which can be addressed with network analysis. Addressing these questions using the constructed network can be a simple form to extend and improve such descriptive studies. Networks should work as a tool to help us understand the processes that structure microbial communities, and therefore the interpretation of observed

patterns and creation of hypotheses from these patterns are not optional and should be included.

Finally, we shortly discuss possible ecological hypotheses that might be tested or used in microbial network analysis. As described in Sec. 4.1, identification of potential keystone species or functional components of the network can help us to forecast structural shifts in the network in response to environmental perturbations, see (Fig. 8c). Another possible research direction is to compare the network architecture shaped by particular conditions, e.g. the available substrate, see (Fig. 8c). In this context, networks can identify taxa with potential similar functionality, since those are expected to compete and be negatively associated across communities (Brown and Wilson, 1956). In the same way we can formulate hypotheses on functionally complementary species, expected to be present in the degradation of complex substrates (Lindemann, 2020). This can be done using networks constructed from datasets extracted from different environmental conditions, expecting a change in predominance of facilitation to competition with increasing nutrient availability (Hoek et al., 2016). These nutrient-based community responses can be combined with a stress response (stress imposed by an environmental factor). What is expected then, is that high stress would reduce investment into competition traits, while facilitation would enable stress mediation, equivalent to what we know from plant ecology (Bertness and Callaway, 1994; Hammarlund and Harcombe, 2019; Piccardi et al., 2019).

## 6. Conclusion

As our understanding of soil biological diversity has increased with the use of high-throughput sequencing techniques, the new challenge for the field of soil microbial ecology has become to understand the complex web of relationships among microorganisms. This remarkable shift in the field is reflected in a rising number of papers using network analysis. The central feature of network models is that they naturally bring microbial associations to the center of our attention. The approach forces us to see the taxa not as isolated units, but to build an integrated picture of microbial communities and also of their surrounding environment. Although these network models open a whole new dimension of possible research, at the moment we have to proceed with caution. The reason is that we are still lacking reliable inference methods and consistent experimental protocols that can successfully reconstruct the hidden network of ecological relationships from environmental samples. Therefore we can only grasp part of the potential of network analysis with a careful interpretation of results obtained from co-occurrence analysis. Finally, we emphasize that the potential of networks goes far beyond being a visualization tool and the information gained from networks constructed from soil microbial community data can - despite limitations - deliver valuable insight into ecosystem organization when applied in hypothesis-driven research. While the traditional question addressed by DNA sequencing analyses of soil microbial communities is “who is there”, the novel question which can be asked by network analysis, when successfully applied, is “who co-occurs with whom, and why?”.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

KS, LA, ES and CK have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (consolidator grant agreement No 819446, granted to CK).

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.soilbio.2022.108604>.

Sequence data has been deposited in the NCBI database and is available under BioProject ID PRJNA763823.

## References

- Abbe, E., 2018. Community detection and stochastic block models. *Foundations and Trends™ in Communications and Information Theory* 14, 1–162. <https://doi.org/10.1561/01000000067>.
- Alteio, L.V., Séneca, J., Canarini, A., Angel, R., Jansa, J., Guseva, K., Kaiser, C., Richter, A., Schmidt, H., 2021. A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry* 160, 108357. <https://doi.org/10.1016/j.soilbio.2021.108357>.
- Araújo, M.B., Rozenfeld, A., 2014. The geographic scaling of biotic interactions. *Ecography* 37, 406–415. <https://doi.org/10.1111/j.1600-0587.2013.00643.x>.
- Armitage, D.W., Jones, S.E., 2019. How sample heterogeneity can obscure the signal of microbial interactions. *The ISME Journal* 13, 2639–2646. <https://doi.org/10.1038/s41396-019-0463-3>.
- Banerjee, S., Schlaeppli, K., van der Heijden, M.G.A., 2018. Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology* 16, 567–576. <https://doi.org/10.1038/s41579-018-0024-1>.
- Banerjee, S., Schlaeppli, K., van der Heijden, M.G.A., 2019. Reply to ‘Can we predict microbial keystones?’. *Nature Reviews Microbiology* 17. <https://doi.org/10.1038/s41579-018-0133-x>, 194–194.
- Barberán, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal* 6, 343–351. <https://doi.org/10.1038/ismej.2011.119>.
- Barner, A.K., Coblenz, K.E., Hacker, S.D., Menge, B.A., 2018. Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology* 99, 557–566. <https://doi.org/10.1002/ecy.2133>.
- Bascompte, J., 2009. Disentangling the web of life. *Science* 325, 416–419. <https://doi.org/10.1126/science.1170749>.
- Berry, D., Widder, S., 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology* 5, 219. <https://doi.org/10.3389/fmicb.2014.00219>.
- Bertness, M.D., Callaway, R., 1994. Positive interactions in communities. *Trends in Ecology & Evolution* 9, 191–193. [https://doi.org/10.1016/0169-5347\(94\)90088-4](https://doi.org/10.1016/0169-5347(94)90088-4).
- Bishop, C.M., 2007. *Pattern Recognition and Machine Learning*, first ed. Springer, New York. 2006. corr. 2nd printing 2011 edition.
- Blanchet, F.G., Cazelles, K., Gravel, D., 2020. Co-occurrence is not evidence of ecological interactions. *Ecology Letters* 23, 1050–1063. <https://doi.org/10.1111/ele.13525>.
- Brown Jr., W.L., Wilson, E.O., 1956. Character displacement. *Systematic Biology* 5, 49–64. <https://doi.org/10.2307/2411924>.
- Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., Deng, L., Yeliseyev, V., Delaney, M.L., Liu, Q., Olle, B., Stein, R.R., Honda, K., Bry, L., Gerber, G.K., 2016. MDSINE: microbial dynamical systems Inference engine for microbiome time-series analyses. *Genome Biology* 17, 121. <https://doi.org/10.1186/s13059-016-0980-6>.
- Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* 11, 2639–2643. <https://doi.org/10.1038/ismej.2017.119>.
- Camacho, J., Guimerà, R., Nunes Amaral, L.A., 2002. Robust patterns in food web structure. *Physical Review Letters* 88, 228102. <https://doi.org/10.1103/PhysRevLett.88.228102>.
- Cantwell, G.T., Liu, Y., Maier, B.F., Schwarze, A.C., Serván, C.A., Snyder, J., St-Onge, G., 2020. Thresholding normally distributed data creates complex networks. *Physical Review E* 101, 062302. <https://doi.org/10.1103/PhysRevE.101.062302>.
- Carr, A., Diener, C., Baliga, N.S., Gibbons, S.M., 2019. Use and abuse of correlation analyses in microbial ecology. *The ISME Journal* 13, 2647–2655. <https://doi.org/10.1038/s41396-019-0459-z>.
- Chen, S., Qi, G., Luo, T., Zhang, H., Jiang, Q., Wang, R., Zhao, X., 2018. Continuous-cropping tobacco caused variance of chemical properties and structure of bacterial network in soils. *Land Degradation & Development* 29, 4106–4120. <https://doi.org/10.1002/ldr.3167>.
- Chernov, T.I., Zhelezova, A.D., 2020. The dynamics of soil microbial communities on different timescales: a review. *Eurasian Soil Science* 53, 643–652. <https://doi.org/10.1134/S106422932005004X>.
- Connor, N., Barberán, A., Clauset, A., 2017. Using null models to infer microbial co-occurrence networks. *PLoS One* 12, e0176751. <https://doi.org/10.1371/journal.pone.0176751>.
- Cordero, O.X., Datta, M.S., 2016. Microbial interactions and community assembly at micro-scales. *Current Opinion in Microbiology* 31, 227–234. <https://doi.org/10.1016/j.mib.2016.03.015>.
- Cougoul, A., Bailly, X., Vourc'h, G., Gasqui, P., 2019. Rarity of microbial species: in search of reliable associations. *PLoS One* 14, e0200458. <https://doi.org/10.1371/journal.pone.0200458>.
- Couto, C.M.V., Comin, C.H., Costa, L.d.F., 2017. Effects of threshold on the topology of gene co-expression networks. *Molecular BioSystems* 13, 2024–2035. <https://doi.org/10.1039/c7mb00101k>.

- Coyte, K.Z., Schluter, J., Foster, K.R., 2015. The ecology of the microbiome: networks, competition, and stability. *Science* 350, 663–666. <https://doi.org/10.1126/science.aad2602>.
- D'Amen, M., Mod, H.K., Gotelli, N.J., Guisan, A., 2018. Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. *Ecography* 41, 1233–1244. <https://doi.org/10.1111/ecog.03148>.
- Delgado-Baquerizo, M., Reich, P.B., Trivedi, C., Eldridge, D.J., Abades, S., Alfaro, F.D., Bastida, F., Berhe, A.A., Cutler, N.A., Gallardo, A., García-Velázquez, L., Hart, S.C., Hayes, P.E., He, J.Z., Hseu, Z.Y., Hu, H.W., Kirchmair, M., Neuhauser, S., Pérez, C.A., Reed, S.C., Santos, F., Sullivan, B.W., Trivedi, P., Wang, J.T., Weber-Grullon, L., Williams, M.A., Singh, B.K., 2020. Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology & Evolution* 4, 210–220. <https://doi.org/10.1038/s41559-019-1084-y>.
- Deng, Y., Jiang, Y.H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular ecological network analyses. *BMC Bioinformatics* 13, 113. <https://doi.org/10.1186/1471-2105-13-113>.
- Dunne, J.A., Williams, R.J., Martinez, N.D., 2002. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters* 5, 558–567. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>.
- Eldridge, D.J., Travers, S.K., Val, J., Wang, J.T., Liu, H., Singh, B.K., Delgado-Baquerizo, M., 2020. Grazing regulates the spatial heterogeneity of soil microbial communities within ecological networks. *Ecosystems* 23, 932–942. <https://doi.org/10.1007/s10021-019-00448-9>.
- Fang, H., Huang, C., Zhao, H., Deng, M., 2017. gCoda: conditional dependence network inference for compositional data. *Journal of Computational Biology* 24, 699–708. <https://doi.org/10.1089/cmb.2017.0054>.
- Faust, K., 2021. Open challenges for microbial network construction and analysis. *The ISME Journal* 15, 3111–3118. <https://doi.org/10.1038/s41396-021-01027-4>.
- Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10, 538–550. <https://doi.org/10.1038/nrmicro2832>.
- Faust, K., Raes, J., 2016. CoNet app: inference of biological association networks using Cytoscape. *F1000Research* 5, 1519. <https://doi.org/10.12688/f1000research.9050.2>.
- Fortunato, S., 2010. Community detection in graphs. *Physics Reports* 486, 75–174. <https://doi.org/10.1016/j.physrep.2009.11.002>.
- Fortunato, S., Hric, D., 2016. Community detection in networks: a user guide. *Physics Reports* 659, 1–44. <https://doi.org/10.1016/j.physrep.2016.09.002>.
- Friedman, J., Alm, E.J., 2012. Inferring correlation networks from genomic survey data. *PLoS Computational Biology* 8, e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>.
- Frindt, K., Pape, R., Werner, K., Löffler, J., Knief, C., 2019. Temperature and soil moisture control microbial community composition in an arctic-alpine ecosystem along elevational and micro-topographic gradients. *The ISME Journal* 13, 2031–2043. <https://doi.org/10.1038/s41396-019-0409-9>.
- Fukami, T., 2015. Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution and Systematics* 46, 1–23. <https://doi.org/10.1146/annurev-ecolsys-110411-160340>.
- Gibson, T., Gerber, G., 2018. Robust and Scalable Models of Microbiome Dynamics for Bacteriotherapy Design.
- Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It's all relative: analyzing microbiome data as compositions. *Annals of Epidemiology* 26, 322–329. <https://doi.org/10.1016/j.annepidem.2016.03.003>.
- Goberna, M., Montesinos-Navarro, A., Valiente-Banuet, A., Colin, Y., Gómez-Fernández, A., Donat, S., Navarro-Cano, J.A., Verdú, M., 2019. Incorporating phylogenetic metrics to microbial co-occurrence networks based on amplicon sequences to discern community assembly processes. *Molecular Ecology Resources* 19, 1552–1564. <https://doi.org/10.1111/1755-0998.13079>.
- Grilli, J., Rogers, T., Allesina, S., 2016. Modularity and stability in ecological communities. *Nature Communications* 7, 12031. <https://doi.org/10.1038/ncomms12031>.
- Gross, T., Rudolf, L., Levin, S.A., Dieckmann, U., 2009. Generalized models reveal stabilizing factors in food webs. *Science* 325, 747–750. <https://doi.org/10.1126/science.1173536>.
- Guimerà, R., Nunes Amaral, L.A., 2005. Functional cartography of complex metabolic networks. *Nature* 433, 895–900. <https://doi.org/10.1038/nature03288>.
- Guimerà, R., Sales-Pardo, M., Amaral, L.A.N., 2004. Modularity from fluctuations in random graphs and complex networks. *Physical Review E* 70, 025101. <https://doi.org/10.1103/PhysRevE.70.025101>.
- Hammarlund, S.P., Harcombe, W.R., 2019. Refining the stress gradient hypothesis in a microbial community. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1910420116>.
- Hermans, S.M., Buckley, H.L., Lear, G., 2019. Perspectives on the impact of sampling design and intensity on soil microbial diversity estimates. *Frontiers in Microbiology* 10, 1820. <https://doi.org/10.3389/fmicb.2019.01820>.
- Hernandez, D.J., David, A.S., Menges, E.S., Searcy, C.A., Afkhami, M.E., 2021. Environmental stress destabilizes microbial networks. *The ISME Journal* 15, 1722–1734. <https://doi.org/10.1038/s41396-020-00882-x>.
- Hewavitharana, S.S., Klarer, E., Reed, A.J., Leisso, R., Poirier, B., Honaas, L., Rudell, D. R., Mazzola, M., 2019. Temporal dynamics of the soil metabolome and microbiome during simulated anaerobic soil disinfection. *Frontiers in Microbiology* 10, 2365. <https://doi.org/10.3389/fmicb.2019.02365>.
- Hirano, H., Takemoto, K., 2019. Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformatics* 20, 329. <https://doi.org/10.1186/s12859-019-2915-1>.
- Hodgson, D., McDonald, J.L., Hosken, D.J., 2015. What do you mean, 'resilient'. *Trends in Ecology & Evolution* 30, 503–506. <https://doi.org/10.1016/j.tree.2015.06.010>.
- Hoek, T.A., Axelrod, K., Biancalani, T., Yurtsev, E.A., Liu, J., Gore, J., 2016. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biology* 14, e1002540. <https://doi.org/10.1371/journal.pbio.1002540>.
- Jian, C., Luukkainen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS One* 15, e0227285. <https://doi.org/10.1371/journal.pone.0227285>.
- Jiang, C.Y., Dong, L., Zhao, J.K., Hu, X., Shen, C., Qiao, Y., Zhang, X., Wang, Y., Ismagilov, R.F., Liu, S.J., Du, W., 2016. High-throughput single-cell cultivation on microfluidic streak plates. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.03588-15>.
- Jiang, D., Sharpton, T., Jiang, Y., 2021. Microbial interaction network estimation via bias-corrected graphical lasso. *Statistics in Biosciences* 13, 329–350. <https://doi.org/10.1007/s12561-020-09279-y>.
- Jiao, S., Lu, Y., Wei, G., 2022. Soil multitrophic network complexity enhances the link between biodiversity and multifunctionality in agricultural systems. *Global Change Biology* 28, 140–153. <https://doi.org/10.1111/gcb.15917>.
- Jiao, S., Yang, Y., Xu, Y., Zhang, J., Lu, Y., 2020. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *The ISME Journal* 14, 202–216. <https://doi.org/10.1038/s41396-019-0522-9>.
- Jones, C.M., Hallin, S., 2019. Geospatial variation in co-occurrence networks of nitrifying microbial guilds. *Molecular Ecology* 28, 293–306. <https://doi.org/10.1111/mec.14893>.
- Jousset, A., Biehnd, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Küsel, K., Rillig, M.C., Rivett, D.W., Salles, J.F., van der Heijden, M.G.A., Youssef, N.H., Zhang, X., Wei, Z., Hol, W.H.G., 2017. Where less may be more: how the rare biosphere mediates ecosystems strings. *The ISME Journal* 11, 853–862. <https://doi.org/10.1038/ismej.2016.174>.
- Karimi, B., Maron, P.A., Chemidlin-Prevost Boure, N., Bernard, N., Gilbert, D., Ranjard, L., 2017. Microbial diversity and ecological networks as indicators of environmental quality. *Environmental Chemistry Letters* 15, 265–281. <https://doi.org/10.1007/s10311-017-0614-6>.
- Karimi, B., Meyer, C., Gilbert, D., Bernard, N., 2016. Air pollution below WHO levels decreases by 40 % the links of terrestrial microbial networks. *Environmental Chemistry Letters* 14, 467–475. <https://doi.org/10.1007/s10311-016-0589-8>.
- Karrer, B., Newman, M.E.J., 2011. Stochastic blockmodels and community structure in networks. *Physical Review E* 83, 016107. <https://doi.org/10.1103/PhysRevE.83.016107>.
- Kaul, A., Mandal, S., Davidov, O., Peddada, S.D., 2017. Analysis of microbiome data in the presence of excess zeros. *Frontiers in Microbiology* 8, 2114. <https://doi.org/10.3389/fmicb.2017.02114>.
- Kemmel, S.W., Wu, M., Eisen, J.A., Green, J.L., 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology* 8, e1002743. <https://doi.org/10.1371/journal.pcbi.1002743>.
- Kraft, N.J.B., Adler, P.B., Godoy, O., James, E.C., Fuller, S., Levine, J.M., 2015. Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology* 29, 592–599. <https://doi.org/10.1111/1365-2435.12345>.
- Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A., 2015. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computational Biology* 11, e1004226. <https://doi.org/10.1371/journal.pcbi.1004226>.
- Kuzaykov, Y., Blagodatkaya, E., 2015. Microbial hotspots and hot moments in soil: concept & review. *Soil Biology and Biochemistry* 83, 184–199. <https://doi.org/10.1016/j.soilbio.2015.01.025>.
- Landi, P., Minoarivelo, H.O., Braennstroem, A., Hui, C., Dieckmann, U., 2018. Complexity and stability of ecological networks: a review of the theory. *Population Ecology* 60, 319–345. <https://doi.org/10.1007/s10144-018-0628-3>.
- Lavrinienko, A., Jernfors, T., Koskimäki, J.J., Pirttilä, A.M., Watts, P.C., 2021. Does intraspecific variation in rDNA copy number affect analysis of microbial communities? *Trends in Microbiology* 29, 19–27. <https://doi.org/10.1016/j.tim.2020.05.019>.
- Lin, Y., Ye, G., Kuzaykov, Y., Liu, D., Fan, J., Ding, W., 2019. Long-term manure application increases soil organic matter and aggregation, and alters microbial community structure and keystone taxa. *Soil Biology and Biochemistry* 134, 187–196. <https://doi.org/10.1016/j.soilbio.2019.03.030>.
- Lindemann, S.R., 2020. A piece of the pie: engineering microbiomes by exploiting division of labor in complex polysaccharide consumption. *Current Opinion in Chemical Engineering* 30, 96–102. <https://doi.org/10.1016/j.coche.2020.08.004>.
- Ling, N., Zhu, C., Xue, C., Chen, H., Duan, Y., Peng, C., Guo, S., Shen, Q., 2016. Insight into how organic amendments can shape the soil microbiome in long-term field experiments as revealed by network analysis. *Soil Biology and Biochemistry* 99, 137–149. <https://doi.org/10.1016/j.soilbio.2016.05.005>.
- Liu, H., Lafferty, J., Wasserman, L., 2009. The nonparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research* 10, 2295–2328. <https://doi.org/10.1184/r1/6610712>.
- Lovell, D., Pawlowsky-Glahn, V., Egozcue, J.J., Marguerat, S., Bähler, J., 2015. Proportionality: a valid alternative to correlation for relative data. *PLoS Computational Biology* 11, e1004075. <https://doi.org/10.1371/journal.pcbi.1004075>.
- Lu, L., Yin, S., Liu, X., Zhang, W., Gu, T., Shen, Q., Qiu, H., 2013. Fungal networks in yield-involving and -debilitating soils induced by prolonged potato monoculture. *Soil Biology and Biochemistry* 65, 186–194. <https://doi.org/10.1016/j.soilbio.2013.05.025>.



- Lupatini, M., Suleiman, A.K.A., Jacques, R.J.S., Antoniolli, Z.I., de Siqueira Ferreira, A., Kuramae, E.E., Roesch, L.F.W., 2014. Network topology reveals high connectance levels and few key microbial genera within soils. *Frontiers in Environmental Science* 2, 10. <https://doi.org/10.3389/fenvs.2014.00010>.
- May, R.M., 1972. Will a large complex system be stable? *Nature* 238, 413–414. <https://doi.org/10.1038/238413a0>.
- May, R.M., 1974. *Stability and Complexity in Model Ecosystems*, vol. 1. Princeton University Press.
- Memmott, J., Waser, N.M., Price, M.V., 2004. Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London Series B Biological Sciences* 271, 2605–2611. <https://doi.org/10.1098/rspb.2004.2909>.
- Montesinos-Navarro, A., Segarra-Moragues, J.G., Valiente-Banuet, A., Verdú, M., 2012. The network structure of plant-arbuscular mycorrhizal fungi. *New Phytologist* 194, 536–547. <https://doi.org/10.1111/j.1469-8137.2011.04045.x>.
- Montoya, J.M., Pimm, S.L., Solé, R.V., 2006. Ecological networks and their fragility. *Nature* 442, 259–264. <https://doi.org/10.1038/nature04927>.
- Morton, J.T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L.S., Edlund, A., Zengler, K., Knight, R., 2019. Establishing microbial composition measurement standards with reference frames. *Nature Communications* 10, 2719. <https://doi.org/10.1038/s41467-019-10656-5>.
- Moruea-Holme, N., Blonder, B., Sandel, B., McGill, B.J., Peet, R.K., Ott, J.E., Violle, C., Enquist, B.J., Jørgensen, P.M., Svenning, J.C., 2016. A network approach for inferring species associations from co-occurrence data. *Ecography* 39, 1139–1150. <https://doi.org/10.1111/ecog.01892>.
- Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F., Knelman, J.E., Darcy, J.L., Lynch, R.C., Wickey, P., Ferrenberg, S., 2013. Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews* : Microbiology and Molecular Biology Reviews 77, 342–356. <https://doi.org/10.1128/MMBR.00051-12>.
- Neutel, A.M., Heesterbeek, J.A.P., van de Koppel, J., Hoenderboom, G., Vos, A., Kaldewey, C., Berendse, F., de Ruiter, P.C., 2007. Reconciling complexity with stability in naturally assembling food webs. *Nature* 449, 599–602. <https://doi.org/10.1038/nature06154>.
- Newman, M., 2010. *Networks: an Introduction*. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>.
- Newman, M.E.J., 2004. Fast algorithm for detecting community structure in networks. *Physical Review E* 69, 066133. <https://doi.org/10.1103/PhysRevE.69.066133>.
- Newman, M.E.J., 2006. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 8577–8582. <https://doi.org/10.1073/pnas.0601602103>.
- Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. *Physical Review E* 69, 026113. <https://doi.org/10.1103/PhysRevE.69.026113>.
- Ofaim, S., Ofek-Lalzar, M., Sela, N., Jinag, J., Kashi, Y., Minz, D., Freilich, S., 2017. Analysis of microbial functions in the rhizosphere using a metabolic-network based framework for metagenomics interpretation. *Frontiers in Microbiology* 8, 1606. <https://doi.org/10.3389/fmicb.2017.01606>.
- Okuyama, T., Holland, J.N., 2008. Network structural properties mediate the stability of mutualistic communities. *Ecology Letters* 11, 208–216. <https://doi.org/10.1111/j.1461-0248.2007.01137.x>.
- Olesen, J.M., Bascompte, J., Dupont, Y.L., Jordano, P., 2007. The modularity of pollination networks. *Proceedings of the National Academy of Sciences* 104, 19891–19896. <https://doi.org/10.1073/pnas.0706375104>.
- Peay, K.G., Kennedy, P.G., Talbot, J.M., 2016. Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology* 14, 434–447. <https://doi.org/10.1038/nrmicro.2016.59>.
- Peixoto, T.P., 2014. The graph-tool python library. <https://doi.org/10.6084/m9.figshare.1164194>.
- Peixoto, T.P., 2019. Bayesian stochastic blockmodeling. In: *Advances in Network Clustering and Blockmodeling*. John Wiley & Sons, Ltd, pp. 289–332. <https://doi.org/10.1002/9781119483298.ch11>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119483298.ch11>.
- Perisin, M., Vetter, M., Gilbert, J.A., Bergelson, J., 2016. 16Stimator: statistical estimation of ribosomal gene copy numbers from draft genome assemblies. *The ISME Journal* 10, 1020–1024. <https://doi.org/10.1038/ismej.2015.161>.
- Piccardi, P., Vessman, B., Mitri, S., 2019. Toxicity drives facilitation between 4 bacterial species. *Proceedings of the National Academy of Sciences* 116, 15979–15984. <https://doi.org/10.1073/pnas.1906172116>.
- Pilosof, S., Porter, M.A., Pascual, M., Kéfi, S., 2017. The multilayer nature of ecological networks. *Nature Ecology & Evolution* 1, 1–9. <https://doi.org/10.1038/s41559-017-0101>.
- Poisot, T., Stouffer, D.B., Kéfi, S., 2016. Describe, understand and predict: why do we need networks in ecology? *Functional Ecology* 30, 1878–1882. <https://doi.org/10.1111/1365-2435.12799>.
- Power, M.E., Tilman, D., Estes, J.A., Menge, B.A., Bond, W.J., Mills, L.S., Daily, G., Castilla, J.C., Lubchenco, J., Paine, R.T., 1996. Challenges in the Quest for Keystone: identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems. *BioScience* 46, 609–620. <https://doi.org/10.2307/1312990>.
- Qi, G., Ma, G., Chen, S., Lin, C., Zhao, X., 2019. Microbial network and soil properties are changed in bacterial wilt-susceptible soil. *Applied and Environmental Microbiology* 85. <https://doi.org/10.1128/AEM.00162-19> e00162-19.
- Qiu, L., Zhang, Q., Zhu, H., Reich, P.B., Banerjee, S., van der Heijden, M.G.A., Sadowsky, M.J., Ishii, S., Jia, X., Shao, M., Liu, B., Jiao, H., Li, H., Wei, X., 2021. Erosion reduces soil microbial diversity, network complexity and multifunctionality. *The ISME Journal* 15, 2474–2489. <https://doi.org/10.1038/s41396-021-00913-1>.
- Raynaud, X., Nunan, N., 2014. Spatial ecology of bacteria at the microscale in soil. *PLoS One* 9, e87217. <https://doi.org/10.1371/journal.pone.0087217>.
- Röttgers, L., Faust, K., 2018. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiology Reviews* 42, 761–780. <https://doi.org/10.1093/femsre/fuy030>.
- Röttgers, L., Faust, K., 2019. Can we predict keystones? *Nature Reviews Microbiology* 17. <https://doi.org/10.1038/s41579-018-0132-y>, 193–193.
- Saiz, H., Gómez-Gardeñes, J., Nuche, P., Girón, A., Pueyo, Y., Alados, C., 2017. Evidence of structural balance in spatial ecological networks. *Ecography* 40, 733–741. <https://doi.org/10.1111/ecog.02561>.
- Sales-Pardo, M., 2017. The importance of being modular. *Science* 357, 128–129. <https://doi.org/10.1126/science.aan8075>.
- Shi, S., Nuccio, E.E., Shi, Z.J., He, Z., Zhou, J., Firestone, M.K., 2016. The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecology Letters* 19, 926–936. <https://doi.org/10.1111/ele.12630>.
- Shi, Y., Delgado-Baquerizo, M., Li, Y., Yang, Y., Zhu, Y.G., Peñaflor, J., Chu, H., 2020. Abundance of kinless hubs within soil microbial networks are associated with high functional potential in agricultural ecosystems. *Environment International* 142, 105869. <https://doi.org/10.1016/j.envint.2020.105869>.
- Song, P., Xiao, Y., Ren, Z.J., Brooks, J.P., Lu, L., Zhou, B., Zhou, Y., Freguia, S., Liu, Z., Zhang, N., Li, Y., 2021. Electrochemical biofilm control by reconstructing microbial community in agricultural water distribution systems. *Journal of Hazardous Materials* 403, 123616. <https://doi.org/10.1016/j.jhazmat.2020.123616>.
- Szoboszlai, M., Tébbe, C.C., 2021. Hidden heterogeneity and co-occurrence networks of soil prokaryotic communities revealed at the scale of individual soil aggregates. *MicrobiologyOpen* 10, e1144. <https://doi.org/10.1002/mbo3.1144>.
- Thébault, E., Fontaine, C., 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329, 853–856. <https://doi.org/10.1126/science.1188321>.
- Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Loccey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., Navas-Molina, J.A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J.T., Mirarab, S., Zech Xu, Z., Jiang, L., Haroon, M.F., Kanbar, J., Zhu, Q., Jin Song, S., Kosciolk, T., Bokulich, N.A., Lefler, J., Brislawn, C.J., Humphrey, G., Owens, S.M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J.A., Clauset, A., Stevens, R.L., Shade, A., Pollard, K.S., Goodwin, K.D., Jansson, J.K., Gilbert, J.A., Knight, R., 2017. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551, 457–463. <https://doi.org/10.1038/nature24621>.
- Tipton, L., Müller, C.L., Kurtz, Z.D., Huang, L., Kleerup, E., Morris, A., Bonneau, R., Ghedin, E., 2018. Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome* 6, 12. <https://doi.org/10.1186/s40168-017-0393-0>.
- Toju, H., Guimarães, P.R., Olesen, J.M., Thompson, J.N., 2015. Below-ground plant-fungus network topology is not congruent with above-ground plant-animal network topology. *Science Advances* 1, e1500291. <https://doi.org/10.1126/sciadv.1500291>.
- Toju, H., Yamamoto, S., Tanabe, A.S., Hayakawa, T., Ishii, H.S., 2016. Network modules and hubs in plant-root fungal biomes. *Journal of The Royal Society Interface* 13, 20151097. <https://doi.org/10.1098/rsif.2015.1097>.
- Torreillas, E., del Mar Alguacil, M., Roldán, A., Díaz, G., Montesinos-Navarro, A., Torres, M.P., 2014. Modularity reveals the tendency of arbuscular mycorrhizal fungi to interact differently with generalist and specialist plant species in gypsum soils. *Applied and Environmental Microbiology* 80, 5457–5466. <https://doi.org/10.1128/AEM.01358-14>.
- Valiente-Banuet, A., Aizen, M.A., Alcántara, J.M., Arroyo, J., Cocucci, A., Galetti, M., García, M.B., García, D., Gómez, J.M., Jordano, P., Medel, R., Navarro, L., Obeso, J.R., Oviedo, R., Ramírez, N., Rey, P.J., Traveset, A., Verdú, M., Zamora, R., 2015. Beyond species loss: the extinction of ecological interactions in a changing world. *Functional Ecology* 29, 299–307. <https://doi.org/10.1111/1365-2435.12356>.
- Varsadiya, M., Urich, T., Hugelius, G., Bärta, J., 2021. Fungi in permafrost-affected soils of the Canadian arctic: Horizon- and site-specific keystone taxa revealed by Co-occurrence network. *Microorganisms* 9, 1943. <https://doi.org/10.3390/microorganisms9091943>.
- de Vries, F.T., Griffiths, R.I., Bailey, M., Craig, H., Giralanda, M., Gweon, H.S., Hallin, S., Kaisermann, A., Keith, A.M., Kretzschmar, M., Lemanceau, P., Lumini, E., Mason, K.E., Oliver, A., Ostle, N., Prosser, J.I., Thion, C., Thomson, B., Bardgett, R.D., 2018. Soil bacterial networks are less stable under drought than fungal networks. *Nature Communications* 9, 3033. <https://doi.org/10.1038/s41467-018-05516-7>.
- Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E.E., van der Heijden, M.G.A., 2019. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications* 10, 4841. <https://doi.org/10.1038/s41467-019-12798-y>.
- Walters, K.E., Martiny, J.B.H., 2020. Alpha-, beta-, and gamma-diversity of bacteria varies across habitats. *PLoS One* 15, e0233872. <https://doi.org/10.1371/journal.pone.0233872>.
- Wei, Z., Yang, T., Friman, V.P., Xu, Y., Shen, Q., Jousset, A., 2015. Trophic network architecture of root-associated bacterial communities determines pathogen invasion and plant health. *Nature Communications* 6, 8413. <https://doi.org/10.1038/ncomms9413>.
- Weier, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., Bentivenga, S., 2011. Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, 2403–2413. <https://doi.org/10.1098/rstb.2011.0056>.
- Williams, R.J., Berlow, E.L., Dunne, J.A., Barabási, A.L., Martinez, N.D., 2002. Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences* 99, 12913–12916. <https://doi.org/10.1073/pnas.192448799>.

- Williams, R.J., Howe, A., Hofmockel, K.S., 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology* 5, 358. <https://doi.org/10.3389/fmicb.2014.00358>.
- Wilpiszeski, R.L., Aufrecht, J.A., Retterer, S.T., Sullivan, M.B., Graham, D.E., Pierce, E. M., Zablocki, O.D., Palumbo, A.V., Elias, D.A., 2019. Soil aggregate microbial communities: towards understanding microbiome interactions at biologically relevant scales. *Applied and Environmental Microbiology* 85. <https://doi.org/10.1128/AEM.00324-19> e00324-19.
- Yang, Y., Chen, N., Chen, T., 2017. Inference of environmental factor-microbe and microbe-microbe associations from metagenomic data using a hierarchical bayesian statistical model. *Cell Systems* 4, 129–137. <https://doi.org/10.1016/j.cels.2016.12.012> e5.
- Yoon, G., Gaynanova, I., Müller, C.L., 2019. Microbial networks in SPRING - semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data. *Frontiers in Genetics* 10, 516. <https://doi.org/10.3389/fgene.2019.00516>.
- Yuan, M.M., Guo, X., Wu, L., Zhang, Y., Xiao, N., Ning, D., Shi, Z., Zhou, X., Wu, L., Yang, Y., Tiedje, J.M., Zhou, J., 2021. Climate warming enhances microbial network complexity and stability. *Nature Climate Change* 11, 343–348. <https://doi.org/10.1038/s41558-021-00989-9>.
- Zhan, P., Liu, Y., Wang, H., Wang, C., Xia, M., Wang, N., Cui, W., Xiao, D., Wang, H., 2021. Plant litter decomposition in wetlands is closely associated with phyllospheric fungi as revealed by microbial community dynamics and co-occurrence network. *The Science of the Total Environment* 753, 142194. <https://doi.org/10.1016/j.scitotenv.2020.142194>.
- Zhang, G., Bai, J., Tebbe, C.C., Huang, L., Jia, J., Wang, W., Wang, X., Yu, L., Zhao, Q., 2021. *Spartina alterniflora* invasions reduce soil fungal diversity and simplify co-occurrence networks in a salt marsh ecosystem. *The Science of the Total Environment* 758, 143667. <https://doi.org/10.1016/j.scitotenv.2020.143667>.
- Zheng, Q., Hu, Y., Zhang, S., Noll, L., Böckle, T., Dietrich, M., Herbold, C.W., Eichorst, S. A., Woeckel, D., Richter, A., Wanek, W., 2019. Soil multifunctionality is affected by the soil environment and by microbial community composition and diversity. *Soil Biology and Biochemistry* 136, 107521. <https://doi.org/10.1016/j.soilbio.2019.107521>.
- Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., Zhi, X., . Functional molecular ecological networks. *mBio* 1, e00169-10. doi:10.1128/mBio.00169-10.
- Zhou, J., Deng, Y., Luo, F., He, Z., Yang, Y., 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO<sub>2</sub>. *mBio*. <https://doi.org/10.1128/mBio.00122-11>.
- Zhu, R., Liu, J., Wang, J., Han, W., Shen, Z., Muraina, T.O., Chen, J., Sun, D., 2020. Comparison of soil microbial community between reseeded grassland and natural grassland in Songnen Meadow. *Scientific Reports* 10, 16884. <https://doi.org/10.1038/s41598-020-74023-x>.