Environment Agency

Understanding ecosystem dynamics using ecological network analysis

Chief Scientist's Group report

December 2024

We are the Environment Agency. We protect and improve the environment.

We help people and wildlife adapt to climate change and reduce its impacts, including flooding, drought, sea level rise and coastal erosion.

We improve the quality of our water, land and air by tackling pollution. We work with businesses to help them comply with environmental regulations. A healthy and diverse environment enhances people's lives and contributes to economic growth.

We can't do this alone. We work as part of the Defra group (Department for Environment, Food & Rural Affairs), with the rest of government, local councils, businesses, civil society groups and local communities to create a better place for people and wildlife.

Published by:

Environment Agency
Horizon House, Deanery Road,
Bristol BS1 5AH

www.gov.uk/environment-agency

Further copies of this report are available from our publications catalogue: www.gov.uk/government/publications or our National Customer Contact Centre: 03708 506 506

Email: research@environment-agency.gov.uk

Author(s):
Nick Aldred, Alex J. Dumbrell, Darren Evans, Laura Hunt, Hannah M. Lewis, Fredric M. Windsor

Keywords:
ecological, network analysis, microbial, ecology, ecosystem, dynamics, function, metabarcoding, metagenomics, metrics

Environment Agency's Project Managers:
Kerry Walsh and Jonathan Warren

Project numbers:
SC220040, SC220041 and SC220046

Citation:
Environment Agency (2024) Understanding ecosystem dynamics using ecological network analysis. Environment Agency, Bristol.

Research at the Environment Agency

Scientific research and analysis underpins everything the Environment Agency does. It helps us to understand and manage the environment effectively. Our own experts work with leading scientific organisations, universities and other parts of the Defra group to bring the best knowledge to bear on the environmental problems that we face now and in the future. Our scientific work is published as summaries and reports, freely available to all.

This report comprises three Think Pieces authored by academic experts, which were commissioned by the Environment Agency's Chief Scientist's Group. The Think Pieces are published as submitted, except for reference lists and glossaries, which were synthesized for the purpose of this report. These views and opinions expressed in these think pieces are those of the authors. Other elements of the report were authored by scientists in the Environment Agency's Chief Scientist's Group.

You can find out more about our current science programmes at https://www.gov.uk/government/organisations/environment-agency/about/research

If you have any comments or questions about this report or the Environment Agency's other scientific work, please contact research@environment-agency.gov.uk.

Dr Robert Bradburne
Chief Scientist

Contents

Executive summary

Microbes (e.g. bacteria, fungi, algae, protozoa) are ecologically important components of ecosystems that are critical to the function and health of ecosystems. However, Microbes have been less extensively included in biomonitoring programmes compared to other taxa. Their sensitivity to biotic and abiotic drivers (e.g., species introductions and climate change) means that this group represents an untapped source of information to enhance our understanding of ecosystem health.

Ecological network science (ENS) is an emerging and developing field in ecology that captures ecological interactions within and across ecosystems and the resultant networks that form. Interactions between organisms and the ecological networks they form are fundamental to ecosystem processes and therefore the delivery of ecosystem services. Approaches in ecological network science are able to identify interactions, construct networks, and analyse properties that describe the fundamental structure of ecosystems, such as network robustness and resilience, from ecological datasets. Combining these techniques with the wealth of data generated by next generation sequencing (NGS) presents new opportunities to unravel the complexities of ecosystem function and develop new measures and metrics of ecosystem health.

Through three think piece papers authored by experts in the fields of ecological network science and microbial ecology, this report explores the state of ecological network science, including its potential application to understand impacts of environmental change on ecosystems as well as its limitations. It sets out recommendations to explore a comprehensive spatial and temporal molecular microbial dataset, generated by the Environment Agency using biofilm samples (aggregate of microbes within a biological matrix found on moist surfaces) collected as part of a new national monitoring programme, designed to assess the state of English rivers and their ecosystems.

The think piece papers outlined several advantages that ENS has over other data analysis methods, the most pertinent of which is its ability to integrate multiple taxa and multidimensional datasets. Trait data can also be incorporated to help capture functional diversity. However, there are challenges in the application of ENS and the interpretation of outputs, particularly where interactions have to be inferred through patterns of co-occurrence. One think piece author described microbial network science as being ‘in its infancy’ and advised caution in its use, with another advising that the limitations of ENS and its application to microbial data need to be carefully considered.

All think piece papers established that interactions and therefore networks would be inferred from the microbial DNA data, because microbial interactions cannot be observed. Inferring interactions can be challenging. Multiple methods for inferring ecological interactions and networks were identified in the think piece papers including graphical inference methods (specifically SParse InversE Covariance Estimation for Ecological Association Inference (SPIEC-EASI) a method and software package designed for the analysis of microbial interaction networks based on DNA sequencing data), methods based on maximum entropy and matrix autoregression, joint species distribution models

and machine learning. While think piece authors recommended different approaches to inference, there was agreement in recommending the application of multiple different inference methods, which should be tested and evaluated to identify the most appropriate method(s) to apply to the microbial dataset.

It was also recommended that timeseries analysis techniques such as Local Similarity Analysis be applied to longitudinal data generated through the project over time, because this could provide insight to how microbial communities respond to environmental perturbations through time and allow us to explore causal relationships. Complementary analysis techniques including multivariate general linear modelling and structural equation modelling were also recommended, as this may indicate potential causal relationships between environmental change, changes in the microbial ecological network, and wider ecosystem properties.

Several metrics were identified to allow for microbial networks to be compared between sites and through time. Topological metrics involve the geometry and connectedness of a network; however, it is uncertain how these metrics relate to ecosystem properties. There was therefore disagreement between think piece papers as to the usefulness of these metrics. Identifying key nodes or hubs in the network may identify keystone taxa, which authors proposed could be the focus of future research. Robustness analysis was also proposed as a way to measure the tolerance of an ecosystem to species extinction. Response diversity, which is not a form of network analysis, was also proposed as a way to capture and explore the potential responses of an ecosystem to a stressor (or stressors). It was also recommended that new metrics specifically linked to stressors or elements of ecosystem function could be developed using the microbial dataset.

Informed by the think piece papers, we intend to apply ENS approaches to the analysis of the NGS data generated through the RSN. As advised, we will experiment with different network inference methods and data analysis pipelines as well as other, more traditional data analysis techniques and methods.

Introduction

1. Microbial molecular ecology and the potential for new metrics and indicators of environmental change

In 2018, Defra (Department for Environment, Food and Rural Affair) published its 25 Year Environment Plan - a commitment to enhance natural capital to ensure the continued provision of the benefits (ecosystem services ES)) of natural capital to society. ES are not only dependent on biodiversity but are influenced by both structure and functioning of the entire ecosystem and understanding both are essential for effective management and conservation. Ecosystems comprise of intricate networks of interacting organisms with abiotic aspects of the environment. By understanding their dynamics, including the flow of energy, nutrient cycling, trophic interactions, we can start to gain better insights into the mechanisms driving ecosystem function and thus ecosystem services (Jax, 2005). Current biomonitoring tools tend to focus on the assessment of community structure, but do not capture information on ecosystem function. However, recent advances in next generation sequencing (NGS, also known as high-throughput sequencing) technologies alongside development of big data analytics, are likely to play an important role in improving our understanding of ecosystem responses to stressors and enabling mechanistic insights into those responses (Derocles et al. 2018; Cordier et al 2019).

Microbes (which include bacteria, algae, fungi, and other protists) are a ubiquitous and critical component of freshwater ecosystems but compared to macroscopic components, microbes are poorly understood. Microbes attach to surfaces and develop biofilms. Microbial life in freshwater ecosystems biofilms dominate microbial life in streams and rivers and drive crucial ecosystem processes and thus the delivery of ES such as nutrient and carbon cycling (Falowski et al., 2008; Battin et al., 2016; Lehtovirta-Morley, 2018; Liu et al., 2021a). Biofilms are comprised of a diverse aggregate of microbial communities within an extracellular polymeric substance (EPS) matrix, which promotes the growth and survival of the overall community (Watnick and Kolter, 2000; Flemming et al., 2016; Penesyan et al., 2021). As well as being important sites for ecosystem processes, biofilms are also hotspots for microbial interactions such as horizontal gene transfer due to the diversity of microbes that are in close proximity (Flemming et al., 2016).

Diatoms, a type of algae commonly found in microbial biofilms are sensitive to changes in environmental conditions and are used in Europe as biological indicators to assess water quality to support the implementation of the Water Framework Directive (Kelly et al., 1998). While other microbial components of biofilms have been shown to respond, in terms of their diversity and function, to a range of pressures and stressors including eutrophication and metal and organic pollution, the full suite of taxa are not readily represented in biomonitoring programmes and are considered the missing link to improving our understanding of and developing new bioindicators and metrics for assessing impacts on aquatic ecosystems (Sagova-Mareckova et al., 2021). As part of the Environment Agency’s research programme into the development of eDNA-based methods for

environmental monitoring and assessment, a comprehensive eDNA microbiome dataset is being developed from river biofilm samples collected as part of the Environment Agency's River Surveillance Network (RSN) to explore potential microbial indicators of environmental change and for measuring ecosystem health.

2. Molecular data generated through the River Surveillance Network

The RSN is a national river monitoring programme developed and run by the Environment Agency as part of the Natural Capital Ecosystem Assessment programme. The RSN is designed to be representative of the English river network, to give a national-scale picture of rivers and their ecosystems and identify where and how they are changing. Other sampling points that make up existing or legacy monitoring programmes tend to be located to monitor human or point-source pressures and on larger rivers, which gives a biased picture of the state of English rivers. However, the RSN is designed in such a way, using a 'Generalised, Randomised, Tessellation, Stratified' (GRTS) approach, that the resulting sample should be representative and unbiased (Brown et al., 2015). GRTS is a spatially balanced, probabilistic sampling design (Kermovant et al., 2016), that was developed for application to large-scale monitoring and river systems (Steven and Olsen, 1999, 2004). Sample sites are located with a stochastic component rather than a fixed interval (Brown et al., 2015).

A range of biotic and abiotic facets of the river environment are measured at sites monitored as part of the RSN (i.e., co-located monitoring), including a suite of water quality parameters (e.g., pH, conductivity etc.), water chemistry, and invertebrate and macrophyte survey data. This, coupled with data about the sampling location (e.g., land use), allows for the RSN to provide insight into a range of pressures, such as habitat, pollution etc. Biofilm samples have also been collected at the RSN sites as part of routine sampling, some of which have been analysed through this project using eDNA techniques to gain insight into microbial communities.

A comprehensive spatial and temporal microbial dataset has been generated from biofilm samples collected during 2021, 2022, and 2023 using NGS sequencing. The dataset includes metabarcoding data for bacteria (16S), fungi (ITS), diatoms and other phytobenthic algae (rbcL) and other microeukaryotes (18S) from nearly 700 sites (sampling strategy is summarised in Table 1. In addition to the metabarcoding data, metagenomic data has been generated for 450 biofilm samples that span a nitrate gradient and represent samples collected from the 72 sites sampled over the three-year period.

Table 1 – Sampling strategy for biofilm samples collected by the RSN 2021-2023.

	2021		
72 same river sites sampled twice a year* for 3 years	72	72	72
13 same sites sampled twice a year* in 2 years	11	13	2
612 unique sites sampled twice a year* for one year	266	202	144
Total sites monitored per year	349	287	218
Total biofilm samples analysed per year	698	574	436

*Spring & Autumn

3. Ecological Network Science

Ecological network science (ENS), or network ecology, is a field of ecology that concerns ecological interactions within an ecosystem and the resultant networks that form (Borrett et al., 2014). ENS approaches aim to identify and characterise interactions or relationships between species (or occasionally individuals) and construct networks based on these interactions. However, ENS has largely been applied to the study of macro-organisms in terrestrial environments, particularly to study food web dynamics.

Combining the NGS data generated from the analysis of RSN biofilm samples with ENS approaches may offer a unique opportunity to understand and develop metrics that better capture the function and dynamics of freshwater microbial communities and their wider ecosystems. However, ENS has not been widely applied to microbial molecular studies (Deng et al., 2012) or to the study of freshwater ecosystems (Windsor, 2023). Therefore, the benefits of taking an ENS-based approach for exploring the NGS data generated through the RSN, particularly for exploring changes in community function, are not yet fully understood.

4. This report

Authored by experts in the fields of ecological network science and microbial ecology, this report explored the state of ENS, including its potential application to understand impacts of environmental change on ecosystems, as well as its limitations. Its sets out recommendations to explore a comprehensive spatial and temporal molecular microbial dataset generated from river biofilms across England.

Three UK-based academics were commissioned by the Environment Agency to each write a think piece paper to explore the state of ENS. This report brings together their views on

the applications and limitations of ENS when applied to microbial molecular data as well as recommendations about how the river biofilm microbial dataset described above could be explored to bolster understanding of ecosystem function, with a view to generating new microbial metrics of ecosystem health.

Authors were asked to specifically address the following questions in their think piece paper.

- Where has ENS been applied and what can it tell us? This will include advantages and limitations and will provide an overview of the research landscape.

- Can ENS be applied to measure ecosystem resilience using microbial molecular data and how established is this as an approach?

- What network properties can be used to measure a community's response to stressors and resilience? Provide a critical evaluation and evidence-based opinion on why and how they have potential.

- What are the gaps in knowledge that need addressing for network approaches to be used?

- What recommendation and approaches would you put forward to explore our multi-taxa microbial molecular dataset?

Think piece paper lead authors were:

Think piece 1 - Dr Fredric Windsor, a lecturer and academic at Cardiff University (School of Biosciences) with an interest and expertise in network ecology, specifically inter-specific interactions, and the response of ecosystems to change. Windsor has led and co-authored empirical research articles on ecological network science, including on its application to freshwater ecology, which have been published in peer-reviewed journals including the Journal of Biogeography, the Journal of Applied Ecology, Agriculture, Ecosystems, and Environment, Perspectives in Ecology and Conservation, and Methods in Ecology and Evolution. Despite a focus on the ecology of macro-organisms, Windsor has co-authored a study addressing the inclusion of (dietary) metabarcoding data into ecology network analysis.

Think piece 2 - Professor Darren Evans is a Professor of Ecology and Evolution at the University of Newcastle (School of Natural and Environmental Sciences). His research expertise lies in the application of network theory and DNA-metabarcoding data to understand species interaction and ecosystem function. Evans has published extensively on network ecology and specifically on the integration of eDNA-metabarcoding and other molecular data into network analysis in peer-reviewed journals including Science, Philosophical transactions of the Royal Society of London B, PLoS One, Functional Ecology, Molecular Ecology and Methods in Ecology and Evolution and has co-authored

several book chapters. Evans currently sits on the Biodiversity Expert Committee for the Department of Environment, Food, and Rural Affairs and the Expert Committee on Forest Science for the Forestry Commission and is an editor of the Journal of Animal Ecology.

Think piece 3 - Professor Alex Dumbrell is a Professor of Molecular Ecology in the School of Life Sciences at the University of Essex. His research expertise combines ecological theory, with advance informatics and molecular techniques to Reunderstand the mechanisms regulating biodiversity (mainly mirobial) and ecosystem functioning, and how this changes in the presence of multiple stressors (e.g., chemicals, warming, land-use change, urbanisation etc). He has published extensively on a range of topics including the application of NGS to freshwater biomonitoring and network ecology. Research and review articles led and co-authored by Dumbrell on molecular ecology and network science have been published in peer-reviewed journals including including Nature Food, Nature Communications, Nature Climate Change, Proceedings of the National Academy of Sciences, Molecular Ecology, Journal of Ecology, Journal of Applied Ecology, Global Change Biology and Trends in Ecology and Evolution. Dumbrell is the current editor-in-chief of Advances in Ecological Research. He also sits on the Peer Review College of the Natural Environment Research Council (NERC), the UK Research and Innovation Panel of Experts, and on the NERC Environmental Omics Facility steering committee.

Introduction, recommendation synthesis, and next steps regarding the application of ENS to the NGS microbial data generated through the Environment Agency's RSN are authored by Dr Laura Hunt, who is an environmental scientist in the Environment Agency's Chief Scientist's Group.

Think piece 1: Developing an understanding of ecological network science for river ecosystems

Fredric M. Windsor[1]

[1]School of Biosciences, Cardiff University

Sir Martin Evans Building, Museum Avenue, Cardiff, CF10 3AX

Executive summary

Ecological Network Science (ENS), otherwise known as network ecology, is the study of ecological interactions and the networks they form. ENS has commonly been applied to networks of different species (e.g., plants and pollinators, predators and prey, seeds and seed dispersers), but there are other scales and levels of biological organisation at which networks can be used. Individual, spatial, social-ecological networks all provide promise.

Extra effort, beyond that of population and community sampling, is required to collect data on ecological interactions. As such, it is necessary that the benefits and drawbacks of ENS are understood and evaluated on a case-by-case basis. The major advantages of ENS are: (i) flexibility; (ii) scale independence; (iii) aid in the visualisation and understanding of complex problems; and (iv) can link ecological structure to ecosystem functions and services. However, some limitations of ENS are: (i) labour and financial expense of data collection; (ii) limited spatial and temporal coverage; (iii) restricted spatial and temporal resolution; and (iv) potential to create unnecessary complexity (depending on the question).

Freshwater ecosystems have received little research surrounding ecological networks, mainly due to the difficulties surrounding collecting interaction data under water – especially in lotic systems. Currently a range of methods are being developed to open up this field, including; automated image and video footage analysis, stable isotopes and molecular methods. By combining data from different methods, it will be possible to construct networks with positive and negative ecological interactions, and better understand the structure of ecosystems.

Molecular methods, such as metabarcoding and metagenomics, are a flexible and powerful tool. These techniques offer an opportunity to gain information on species identities, phylogenies and traits. A significant limitation, however, is that interactions between organisms often have to be inferred, especially for microbial communities.

Methods of network inference have been developing over the past decade, and range in data requirements and accuracy. Although previously criticised for their simplicity and inaccuracy, recent advances have looked to incorporate extra data on biological traits and

phylogenies to enhance accuracy and realism. Recent methodological frameworks have been developed to account for all of the variation in the co-occurrence and abundance of organisms related to these factors, with the remaining unexplained variation ascribed to biological interactions.

Ecosystem scale properties can be assessed using ENS. Resilience, in particular, can be assessed as ENS offers an opportunity to account for both direct and indirect effects of environmental change and other drivers of ecosystem structure and function. Many methods have been proposed for monitoring stability and resilience of ecosystems in the face of change, however, a range of recent developments allow for an understanding of these properties across different scales (local to global, short- to long-term).

Environment Agency microbial community data, generated using metabarcoding and metagenomics, can be analysed using a range of different methods. Ecological interactions and networks can be constructed from network inference, based on a framework that can account for spatial and environmental factors (e.g., water quality and quantity), as well as potential biotic interactions (i.e., facilitation or competition). Integrating this information in network models at the river catchment scale would allow for an understanding of site and catchment resilience to further environmental change.

1. Introduction

Ecological Network Science (ENS), otherwise known as network ecology, involves the study of ecological interactions and the resultant networks they form (Borrett et al., 2014). These networks of interactions are incredibly diverse, ranging from mutualistic through to antagonistic and everything in between (Jordano, 2016a), as well as spanning spatial and temporal scales. These individual interactions, and their combinations, convey a range of ecosystem functions (e.g., decomposition, pollination, herbivory), are critical for the resilience of ecosystems to environmental change (Evans et al., 2016) and can provide early warning signals for biodiversity loss (Dirzo et al., 2014; Valiente-Banuet et al., 2015). Monitoring and conserving ecological interactions and networks is therefore critically important (Gray et al., 2014), and requires further research (Windsor et al., 2023).

Ecological interactions take many forms in the natural world. At one end of the spectrum, mutualisms occur, where both interacting nodes benefit from the interaction, for example pollination and seed dispersal (May, 1982). At the other, antagonisms, where one node benefits and the other is detrimentally affected, such as predation, decomposition and herbivory (Jordano, 2016b). There are also a variety of interaction types in between, for example commensalistic and amensalistic interactions (see Table 1). As well as being diverse in nature, interactions can form at different levels of biological organisation. As an example, networks of interactions form between cells, individual organisms, species, habitats, ecosystems and continents. I will talk more about this in Applications of ENS.

The following sections of the report aim to cover a range of distinct, yet inherently linked aspects of ENS. Section 2 describes a range of ecological networks that are commonly constructed across different scales and levels of biological organisation and their

applications for monitoring and management. Section 3 identifies the major gaps in ENS, focusing on its applications for decision-making and monitoring. Section 4 focuses in on where ENS has been applied to freshwater ecosystems, which is a far smaller body of research in comparison to terrestrial and marine ecosystems. Section 5 covers the nuances and considerations around the use of molecular methods for constructing ecological networks and how the data generated can be used for ENS. Section 6 runs through the methods that can be used for inferring ecological interactions from monitoring data (molecular or otherwise). Section 7 details the current state of play in understanding the resilience and stability of ecosystems using an ENS framework. Finally, Section 8 synthesises the above sections to provide some specific recommendations for the Environment Agency metagenomic dataset.

2. Applications of ENS

Ecological networks can be constructed and analysed at lots of different scales, from cells to ecosystems. These analyses can be used to understand the structure and function of ecosystems and in turn inform decision-making with regards to management. Below is a brief summary of ENS across a range of scales of particular interest for monitoring.

2.1 Biological networks

Individual-based networks focus on individual elements, where the nodes are elements with their own agency – for example individual cells or organisms (Guimarães, 2020). These can be incredibly useful when investigating fundamental ecological questions, such as individual behaviours (e.g., prey preferences) and how they influence populations, communities, ecosystems and ecological processes/functions (Arroyo-Correa et al., 2021; Gómez and Perfectti, 2011). Typically, these networks have been constructed through direct, repeated observations of individual behaviour. In aquatic systems, these methods have been applied in a few unique circumstances, for invertebrates and fish. For example, studies have looked at the movement of individual brook charr (*Salvelinus fontinalis*) throughout river courses over their migration cycle to understand the genetic flow upstream and downstream (Morrissey and Ferguson, 2011). When multiple datasets are investigated in tandem individual-based network analyses also can provide greater resolution and more information on seasonal and ontogenetic shifts, and a better understanding of temporal variation in network structure in comparison to aggregated, inter-specific networks (Woodward et al., 2010). In terms of applied research, however, there has been limited use of individual-based methods in decision-making, conservation or restoration (but see literature on landscape connectivity and individual-based spatial networks; Kanagaraj et al., 2013). Into the future, nevertheless, these techniques will undoubtedly prove useful for understanding ecosystem response to environmental change.

Inter-specific networks are the most commonly used ecological networks due to the fact that species are common currency in population and community ecology. Studies typically assess different antagonistic and mutualistic interactions, including but not limited to: plant-pollinator, host-parasite, plant-herbivore, predator-prey, plant-frugivore and plant-seed disperser (Mougi and Kondoh, 2012). These networks are constructed using lots of different types of methods, including: direct observation, molecular analyses, statistical inference, amongst others (Jordano, 2016b). Ultimately, however, all methods identify individual taxa and seek to identify the interactions between them, to compile complex networks of interactions. Such networks provide a large amount of additional information in excess of that achieved from ecological communities alone – for example details of decomposition, predation and/or pollination (Felipe-Lucia et al., 2022). As a result of this, and due to their inherent structure incorporating species richness and biomass flow, ecological networks provide an suitable option for understanding the links between biodiversity and ecological function (Poisot et al., 2013). Something which is uncommon in the field of ecology. Inter-specific networks have been proposed as widely applicable for decision-making (Windsor et al., 2023, 2022), however, there still remains limited uptake of these methods on the ground.

2.2 Social networks

Social networks can also be constructed, including humans and their actions. Commonly these are individual-based networks, but they can also represent groups of people (e.g., land managers or regulators). Although these types of networks are not directly relevant to ENS from a strict ecological perspective, they are useful for thinking about the implementation and effectives of management activities. In particular, a new paradigm has formed around the use of social-ecological networks (SEN), and the ability for these frameworks to directly incorporate management, ecosystems and ecosystem services (Bodin et al., 2019; Felipe-Lucia et al., 2022; Sayles et al., 2019). This allows for an integrated understanding of human actions on ecological systems, moving beyond studies that include the effects of anthropogenic stressors implicitly in their design (e.g., through measuring ecological structure or function along environmental gradients). It also provides significant promise for making changes based on the ecological data collected. For example, avenues of remediation, restoration, regulation or other management decisions surrounding ecological systems can immediately be identified within the social network. Further to this, we can assess the direct and indirect effects of management on society and biodiversity, with the aim of preventing unintentional effects.

2.3 Spatial networks

Spatial networks are a key component of modern decision-making in ecology and environmental science, and are one of few examples where networks have been completely integrated into landscape planning and management. For example, the “Making Space for

Nature" report by Lawton in 2010 references connectivity and isolation of habitats, and the limited potential for movement of organisms at the landscape scale (Lawton et al., 2010).

Spatial networks can resemble all of the above network typologies (individual, inter-specific and multilayer), however, they are different in the sense that space is the primary focus of the networks of interactions. The links/edges in these networks can represent many different processes, but most often dispersal, foraging and gene flow between populations or habitats are investigated (Pilosof et al., 2017). Scale, as with other types of networks, is important and space can be incorporated at local scales (e.g., movement between habitat patches) or global scales (e.g., transcontinental migrations).

In river systems, spatial networks are confined to the river channels (for completely aquatic organisms), and can represent unidirectional flows for water and nutrients, or bidirectional flows for mobile organisms such as insects, fish, birds and mammals. Physically, river systems have been treated as dendritic networks previously (Peterson et al., 2013), and such methods have been used to estimate nutrient flows, water temperature and other processes that operate in a single direction (e.g., effects of climate change and wildfire on stream temperatures and salmonid thermal habitats; Isaak et al., 2010). Indeed, the use of spatial networks in freshwater ecosystem assessments has been identified as a significant area of future research and innovation (Erős and Lowe, 2019).

2.4 Merged networks

Some methods have been relatively recently developed to combine different networks into one overarching framework. Such methods, often termed multiplex, multitrophic or multilayer networks (see Table 1), consolidate lots of different types of data across various scales (e.g., individuals through to ecosystems).

Multilayer networks, are situated at the forefront of network science, being applied to emerging complex problems, such as disease spread, migration, and meta-populations or meta-communities (Pilosof et al., 2017; Silk et al., 2018). The layers in a multilayer network represent discrete units in space or time, for example, habitat patches, ecosystems or islands but also sample units (e.g., weeks, months, seasons or years). There are interactions within the layers (intra-layer interactions), but also interactions between the layers (inter-layer interactions) which can represent a range of different processes – e.g., the dispersal of different taxa between habitats (Hutchinson et al., 2019; Pilosof et al., 2017). Multilayer network analyses are not widely applied, mainly due to the limited datasets available at suitable scales. However, they present great promise as they are able to integrate ecological interactions at different spatial and temporal scales, and have the potential to understand scale variant properties of networks within one analysis. For example, one could describe an entire landscape using a multilayer approach: inter-specific interaction networks within habitats and the dispersal networks for different species across habitats. This would allow for an understanding of the dynamics of ecosystems incorporating effects at multiple scales (Box 1).

Box 1. Examples of multilayer network analyses for integrating data across scales.

In a seminal review by Pilosof et al. (2017), several different examples were given involving multilayer networks (either temporal, spatial, or multiple different interaction types) to demonstrate their utility. A particularly interesting example is provided from an analysis of data from Norwood farm, where the robustness of different interaction networks was calculated based on changes to the structure of linked networks. This is an interesting example as network robustness is a commonly used metric of resilience and/or stability. Two layers of interactions, plant-pollinator and plant-leaf miner parasitoids, with shared plant species, were analysed. Pollinators were removed from the network and the robustness (Table 1) of plants and leaf miner parasitoids were assessed. Plants would become extinct if they lost their pollinators, and parasitoids would become extinct if they lost their host plants, secondary and tertiary extinctions, respectively. It was found that the robustness of parasitoids was significantly different in the multilayer analysis than when analysing the individual layers, showing that the interplay of different interaction types has a significant impact on the whole system structure and dynamics.

An modelling-based example, combining different levels of biological organisation (population, community and meta-community) is presented by Scotti et al. (2013). This modelling paper showed that the probability of different interactions forming (e.g., social interactions or migration) affects meta-population sizes and spatial heterogeneity across the wider food web. Altering parameters across the three different levels of organisation influenced the population dynamics all individuals in the food web. However, unexpectedly community dynamics (e.g., food web interactions) were not the overriding structuring force on populations, instead social and landscape processes had a greater effect.

Both examples show the importance of taking an integrated approach, and incorporating associated and interlinked interactions across different levels of biological organisation.

2.5 Applied ENS

Network science, and by extension ENS, is often used in applied research. Recent reviews have acknowledged the diversity of potential applications of network thinking to ecology and

biogeography (Harvey et al., 2017; Sayles et al., 2019; Windsor et al., 2023): (i) conservation; (ii) invasion biology; (iii) restoration; (iv) biomonitoring; (v) species distribution modelling; (vi) risk governance; and (vii) landscape planning (i.e., corridors and habitat connectivity). Below I provide a series of examples highlighting the potential applications of ENS and their benefit.

i) Conservation has a long history with network ecology (Harvey et al., 2017). A specific example from terrestrial ecology, is that of planning ecological conservation and restoration of plant-pollinator networks in forest ecosystems (Devoto et al., 2012). Individuals were prioritised for conservation based on restoration targets from plant-pollinator networks across a landscape, such as functional redundancy or complementarity. The framework, prioritising species conservation based on potential endpoints informed by ecological networks, was identified as a powerful tool for both conservation and restoration ecology (Devoto et al., 2012). There are many more instances where ENS can be applied to select conservation priorities or be used as a framework for assessing the effectiveness of conservation activities for ecosystem-scale interventions.

ii) Ecological networks can aid in the identification of potential invaders (e.g., those with a high level of generalised interactions; Traveset and Richardson, 2014), their potential role in naïve systems (Romanuk et al., 2009), and the subsequent effects (e.g., competitive exclusion, population decline and reorganising communities; Hui and Richardson, 2022). Certainly, studies have shown that invasive species (both plants and seed dispersers) may dominate interaction networks and play key roles influencing network topology – the rules affecting how these invaders affect network structure, however, are the same rules as those governing the role of native organisms (Vizentin-Bugoni et al., 2021).

iii) For restoration ecology, networks provide an opportunity to integrate lots of different types of information on organisms and their interactions, including the evolution or organisms and interactions (see Segar et al., 2020), and dynamical responses of organisms to species introductions and other changes (Raimundo et al., 2018). In restoration, a lot of emphasis is placed on the unknown indirect effects of introducing species, and thus these methods appear particularly useful within this subfield of ecology. Furthermore, restoration has primarily focused on a subset of organisms or individual species, whereas ENS provides an opportunity to understand ecosystem-scale effects of such management strategies.

iv) Biomonitoring benefits from an ecological network perspective, in the sense that networks can provide additional information on ecosystem degradation (Gray et al., 2014). For example, significant changes in network structure can occur with little to no changes in the species richness or composition of the community (Valiente-Banuet et al., 2015). This is also important as there is potential for changes or a loss of interactions to be early warning signals of species loss or ecosystem scale changes (Windsor et al., 2023). It is important, however, not to see these methods as a replacement for standard biomonitoring, as the inverse of the above example can also be the case, where species identities change without significant fluctuations in network structure (Petanidou et al., 2008). Furthermore, there remains a series of developments in this field that are required prior to the widespread role out of ENS in standardised biomonitoring (Derocles et al., 2018).

v) The distribution of a species is inherently influenced by ecological interactions (Wisz et al., 2013). Recently, research has shown that data on ecological interactions can be used to truth and sense-check distribution data (Higino et al., 2023) allowing for an improved understanding of where organisms reside. Accounting for interactions, or their absence, in macroecology and biogeography is a promising area of future research – with a wide range of fundamental and applied uses (Windsor et al., 2023).

vi) Decision-making is usually based on data provided by ecological studies, however, ENS through SEN analyses provides an opportunity to directly incorporate ecological data in human decisions surrounding the environment (Windsor et al., 2022). Risk governance, for example, has been shown to benefit from a SEN approach. For large wildfires (>100 km), alignment between the interactions of different actors (e.g., Forest rangers, land owners and residents) and the connectivity of spatial ecological networks (e.g., habitat connectivity), is critically important for the success of wildfire management (Hamilton et al., 2019). This example of applied ENS, although specific to risk governance, is more widely applicable – demonstrating the ability for ENS to include social, economic and any other interaction data from the surrounding systems.

vii) Networks can be spatially organised (see 2.3 Spatial networks) and as such, ENS provides an opportunity for landscape planning and optimisation. Certainly, this has been where most applied ENS has been focused. For example, ENS can be used to quantify the levels of connectivity and the potential benefit provided by protected areas considering their location in spatial networks (Rayfield et al., 2011). Identifying areas that are good locations for conservation or restoration considering their position within the landscape is important as it can maximise the efficacy of site-based management strategies (Isaac et al., 2018).

Although ENS have been widely applied to different fields of ecology and biogeography, there are a variety of technological and methodological developments that are required prior to the widespread uptake of such methods. Nevertheless, ENS has shown to be a useful tool in applied ecology and only through ground-truthing and continued application will these methods become socialised to a wider audience.

2.6 Advantages of ENS

Constructing and analysing networks is not always the most straightforward exercise in comparison to commonly used methods in population and community ecology (e.g., community sampling and taxonomic identification). But there are a number of advantages of ENS, from both a fundamental and applied perspective. ENS are:

- **Intuitive.** Scientists have an innate appreciation for the interconnectedness of the natural worlds. Furthermore, many ecologists have been using networks, in the form of relational databases, to store all kinds of ecological data. This familiarity means that uptake and understanding of the results of network analyses is often high.
- **Able to integrate data across scales**. As there is no prerequisite on what a node or a link can be (unlike other subfields of ecology where the elements are prescriptively

defined), networks can be used to integrate data across scales. This is uncommon in ecology, where scale-dependence, if often a significant issue and means that the results of many studies are not generalisable.

- **Useful aids in the visualisation of complex problems**. Network diagrams provide a brilliant tool for visualising complex problems and the interconnectedness. As long as this is carefully executed (i.e., not confusing or overwhelming in terms of showing complexity) it is incredibly effective (Pocock et al., 2016).
- **Suitable for linking ecological structure to function**. Ecological interactions describe the structure of ecosystems; however, they also represent ecological functions. For example, the interactions between predators and prey represent energy flux, plant and pollinators represent pollination, etc. Thus, by studying the structure and strength of ecological interactions, and their variation in space and time therefore affords information on the mechanisms underlying ecological function.
- **Direct links to ecosystem services**. Beyond ecological functions, some interactions are ecosystem services. There are several examples of where direct and indirect ecological interactions are also ecosystem services, for example: plant-pollinator (pollination) and host-parasitoid (biological control). Furthermore, indirect effects on ecosystem service provision can be identified, for example predator-prey interactions between sport fish (e.g., salmonids) and insects, indirectly affect the recreational and provisioning services provided by fish by affecting population growth rates, biomass accrual, etc.
- **Flexible and can include different types of ecological and social interactions**. As highlighted in above sections, lots of different information can be summarised in a network format. This provides significant analytical power when investigating problems that span different systems, or scales, as different types of data can be integrated with relative ease (Guimarães, 2020).

The extra effort taken to capture data on the interactions between organisms therefore appears worth the effort. Nevertheless, there are challenges and limitations of ENS, and its use should be considered on a case-by-case basis.

2.6 Disadvantages of ENS

There are a number of limitations surrounding the use of ENS, which caveat some of the advantages described above. There are several fundamental challenges of using ecological networks, especially in an applied context:

- **Expense.** Compared to population and community ecology, collecting information on ecological interactions requires significant financial and labour costs. Large numbers of ecologists are required to identify organisms in the field at different spatial and temporal scales, or samples can be collected and significant laboratory hours and consumables costs is incurred, for example, for molecular methods (Evans et al., 2016).
- **Complexity.** The aim of networks is to convey and understand the complex interactions present in the natural world. This, however, can be off-putting and overwhelming, and difficult to make sense of. Care therefore needs to be taken in presenting the important aspects of complexity without present a ‘ball of wool”. The

added value from visualising and appreciating the complexity of ecological networks should always be communicated effectively (Pocock et al., 2016).

There are also significant technological and methodological challenges associated with collecting data on ecological networks:

- **Spatial coverage.** Historically the coverage of ecological network data has been spatially restricted. Often studies sacrifice spatial coverage for temporal resolution, or vice versa (Windsor et al., 2023). Most studies to date have been completed at the site scale, for example across a single farm (e.g., Norwood farm; Evans et al., 2013; Pocock et al., 2012), or in a single river reach (Windsor et al., 2019). Studies at larger spatial scales have typically focused on a subset of interactions or taxa (e.g., only focusing on seed dispersing birds). This is changing, with studies able to make use of emerging technologies and methods (Besson et al., 2022). However, for the most part, spatial coverage of primary data studies remains restricted to landscape scales at best. This limits the extent to which findings can be translated across different sites or systems.
- **Temporal resolution and data aggregation.** Ecological interactions do not occur all the time, or everywhere. Some interactions are common and may be frequently observed, whereas others may occur very infrequently, requiring a large sampling effort in order to detect them (Jordano, 2016b). As such, building complex ecological networks, incorporating large numbers of interactions, a certain level of temporal aggregation is often required. As ecological systems are extremely dynamic, this incurs some trade-offs in our ability to understand how interactions vary in time. As above, new methods for collecting data on ecological interactions are reducing this issue, allowing for comprehensive analyses of interactions, such that more information on interactions can be collected per timepoint – leading to a lesser need to aggregate, or aggregation over smaller timescales. This all allows for data to be analysed at higher temporal resolutions, understanding the dynamics of interactions and moving beyond the “snapshot” provided by previous network studies.
- **Sampling completeness, missing interactions and forbidden links.** Linked to the above points on spatial coverage and temporal aggregation, making sure all possible interactions are sampled, is a large challenge. Some interactions occur commonly, and are easy to observe, others occur less frequently, and require a greater level of sampling effort to observe (Olesen et al., 2010). It is important that we understand how many of the potential or viable ecological interactions we have detected as network size (i.e., the number of nodes and links in a network) has a large bearing on many network properties. Indeed, in the past, large-scale analyses of network patterns can simply be an artefact of network size, rather than actual patterns in the topologies of the networks (Nielsen and Bascompte, 2007). It is difficult to assess sampling completeness, as we have an imperfect knowledge around which interactions are expected or not expected to occur, and this will vary in space or time. Nevertheless, it is generally well accepted that some level of data aggregation from field sampling is required to attain an acceptable probability that all ecological interactions have been sufficiently sampled.

As ENS is an emerging field of research, many of these drawbacks are actively being addressed, for example, automated sampling of ecological interaction data (Besson et al., 2022) and distributed, rapid next generation sequencing

3. Gaps in knowledge

As with other fields of research, ENS has gaps. However, unlike many other fields, these gaps often relate to our ability to collect interaction data at suitable spatial and temporal resolutions.

Classical methods for constructing ecological networks (e.g., gut content analysis, timed walks, rearing parasitoids) are time consuming, costly and requires significant labour (see 2.7 Disadvantages of ENS). All of which restrict the number of time-points and locations that can be sampled to a degree that represents reality. Although new methods provide significant promise (see 5 Molecular methods and considerations for ENS) there remains a limited number of studies that operate at temporal and spatial scales useful for decision-making, conservation, restoration, or other management activities.

Studies on ecological networks are limited to site and landscape scales, although this is changing (e.g., Galiana et al., 2021, 2018). This means that we do not well understand biogeographic patterns in ecological interactions, as well as ecological networks (Windsor et al., 2023). Expanding the focus of studies to operate across broader scales (i.e., across multiple habitats or catchments) will provide potentially significant advances in our fundamental understanding. Furthermore, studies at this scale will be more suitable for informing management activities (Hutchinson et al., 2019). Expanding our research to understand spatial patterns in network structure and function, and applying this information to conservation, restoration, and other activities, is vitally important.

Due to the limitations surrounding temporal resolution and the regular need to aggregate data across multiple time steps, we have a poor understanding of the dynamics of ecological networks over ecological meaningful timescales (e.g., days to weeks). However, we know that interactions are likely to be changing in their identity and frequency over a range of temporal scales (e.g., hourly to multi-decadal), based on a variety of different factors. These factors depend somewhat on the interaction of interest (e.g., predation, pollination or competition), but can be anything from resource availability (e.g., phenology of food plants or prey) through to individual-level foraging behaviour (i.e., accessing different habitats throughout the day). Understanding this variability and dynamism in ecological interactions is crucial for predicting the effects of environmental change as it has a range of connotations for processing such as adaptive rewiring (e.g., the ability of an organism to persist by changing who it interacts with; Raimundo et al., 2018; Thierry et al., 2011) and cascading extinctions (Dunne and Williams, 2009; Vieira and Almeida-Neto, 2015). Thus, being able to collect data on interactions at high temporal resolutions is a current gap, but a focus of much research.

There remains a significant gulf in our understanding of ecological interactions and networks between different ecosystems. Terrestrial and marine systems have long received a greater amount of attention in the field of network ecology, with freshwaters receiving relatively little attention. I focus on this in more detail below (4 ENS in freshwater ecosystems), as it is such a significant and contemporary gap in our understanding.

4. ENS in freshwater ecosystems

The focus of much ENS research to date has been in terrestrial and marine ecosystems. There are, however, a range of ecological interactions in freshwaters (Figure 1) that could be critically important to their function and overall health – as is the case for other ecosystems (Dirzo et al., 2014; Valiente-Banuet et al., 2015). We currently only understand a small number of the potential interactions that could be present, and the vast majority of research focuses on trophic interactions and food webs (Silknetter et al., 2020; Windsor, In Review). Classical work, by pioneers like Charles Elton, did investigate competition and other non-trophic interactions (Elton, 1929). Indeed, there is a strong legacy of behavioural ecology in freshwaters that would lend itself to collecting more information on the wider suite of mutualistic and antagonistic interactions in these understudied environments. Despite classical research, the vast majority of research has focused on trophic interactions and the food webs that they generate. The absence of data on non-trophic interactions in freshwaters is problematic.

There is a high potential diversity of ecological interactions in freshwaters (see Figure 1 for some examples). They range from mutualistic through to antagonistic, and can involve both direct and indirect interactions between a range of taxa. These interactions can be summarised across scales, and also at the system scale (i.e., across dendritic river networks). Unlike in terrestrial and marine systems, our understanding of ecological networks in freshwater ecosystems is restricted to food webs (Ings et al., 2009). Yet, recent reviews have highlighted the importance of positive interactions, i.e., mutualisms and commensalisms, and their diversity in freshwaters (Silknetter et al., 2020). There is, however, significant potential to use ENS in freshwater ecosystems, particularly river systems which have a spatial network structure.

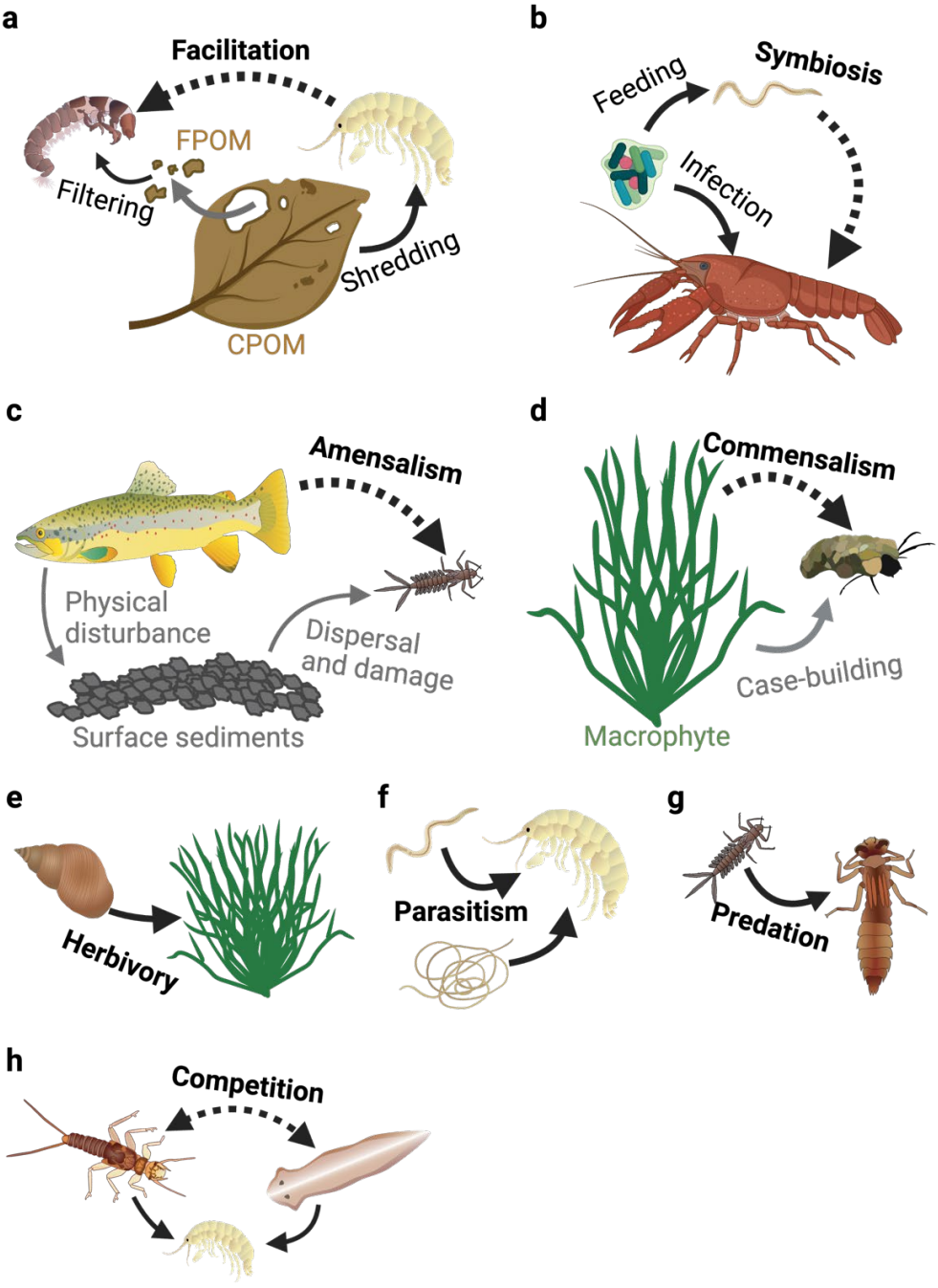


Figure 1 - Examples of ecological interactions in freshwaters. Produced with permission from Windsor et al. (In Review).

5. Molecular methods and considerations for ENS

DNA and RNA-based methods for sampling ecological networks is an area of emerging promise for ENS applications (Derocles et al., 2018; Evans et al., 2016; Evans and Kitson, 2020). These methods present great promise as they offer a way of detecting previously hidden interactions (i.e., from microorganisms) and thus fleshing out networks (Clare et al., 2019; Miller et al., 2021; Vacher et al., 2016), but also allow for collecting data at greater

spatial and temporal resolutions than possible with classical methods (Bohan et al., 2017). Below I discuss a range of specific advantages of this group of methods for constructing ecological networks.

Where other sampling methods fail to non-invasively sample a wide range of different interactions, molecular methods succeed. For example, faecal samples (Drake et al., 2022), mouthpart and skin swabs (Evans and Kitson, 2020), water and air sampling (Clare et al., 2022), all provide the opportunity to detect interactions ranging from mutualisms through to antagonisms (e.g., symbioses, parasites, and parasitoids). Collecting this array of interactions using a single method means that it is possible to reduce the biases associated with merging together interaction data from multiple types of sampling method (see Cuff et al., 2022). Whilst minimising sampling costs and maximising processing efficiency (i.e., all of those samples, after some sample specific extractions, can be processed in exactly the same manner).

Molecular methods also provide an opportunity to detect previously difficult to observe interactions. These interactions may be challenging to detect and/or quantify for a variety of reasons, but these methods in one way or another can help:

- **Cryptic organisms (i.e., where taxonomic identification is extremely difficult).** As molecular methods are not completely reliant on taxonomy (e.g., species can be identified based on phylogenetic dissimilarity), we can distinguish between different operational taxonomic units in cases where taxonomy might struggle and generate estimates of species richness (Helmus et al., 2007). Identification of unknown organisms and those taxa from poorly classified groups of organisms, such as microorganisms and microbes (Feng et al., 2019), is possible using these methods, unlike classic taxonomic methods which are reliant on expert taxonomists.
- **Rare organisms or interactions.** Molecular methods are not contingent on the organisms and interactions being present at the exact moment in time that sampling takes place, as is the case for most classic methods of network construction. In fact, DNA-based methods can be used to detect interactions that have occurred sometime in the past, or from carrion prey (e.g., Neidel et al., 2022), depending on the DNA degradation rates in the given environment.
- **Soft or amorphous organisms/tissues.** A limitation of visual analysis of gut contents or faecal samples is the inability to detect and quantify the contributions of organisms without hard parts, e.g., chitin or bone. Molecular methods are not subject to these same limitations and can detect organisms irrespective of their tissue composition (Symondson, 2002).
- **Symbionts and parasites.** By sequencing the tissues of organisms, it is possible to concomitantly detect endosymbionts and endoparasites – interaction organisms that are often difficult to detect using standard visual methods (Miller et al., 2021). This is especially useful for organisms without distinctive features, or those groups which have poorly described taxonomies.

A further major benefit of molecular approaches, touched upon above, is the relatively low cost of such methods. Sample collection requires simply the sampling of individuals, populations or communities using standard monitoring methods (e.g., kick sampling, Surber sampling or malaise traps), which is streamlined in comparison to the methods used for collecting information on interactions in the field (e.g., timed walks, transects). Samples are

then transported to the laboratory, where the majority of the time and financial expense is incurred. Per unit sample costs, however, for barcoding and metabarcoding have reduced drastically over the past decade – mainly due to advances in throughput of next generation sequencing platforms (Shokralla et al., 2012; Srivathsan et al., 2021). There is the significant caveat, however, that this is the cost for researchers operating in Europe, North America and Asia, and the costs of consumables (e.g., plastics, reagents) increases considerably in other regions of the globe where import taxes are levied against such items (e.g., South and Central America).

Using molecular methods to identify interactions also provides an additional layer of information that can be used to understand the structure of ecosystems – phylogenies (Evans et al., 2016). Phylogenetic relatedness of organisms can be derived from meta-barcoding data using de novo methods or constructed by matching taxonomic assignments from metabarcoding and bioinformatics (e.g., using GenBank and BOLD) to existing databases containing phylogenetic information on different organisms (e.g., rotl; Michonneau et al., 2016). Phylogenetically-structured networks have the ability to improve predictions of network structure in the face of environmental change or management. For example, predictions of species introductions can be enhanced by assuming that the introduced species will interact in a similar way to the most phylogenetically similar species in the existing network (Raimundo et al., 2018).

As with any other sampling method, however, molecular methods also have limitations. One of the major drawbacks of molecular-based ecological networks is the necessity that data be binary (Cuff et al., 2022). Although read counts have been used to produce quantitative interaction data (e.g., trophic interactions; Deagle et al., 2013), with significant assumptions and caveats. In recent work methods of semi-quantitative networks based on molecular methods, but also combinations of different methods, have been presented, somewhat circumventing this limitation (Cuff et al., 2022). However, in comparison to other methods, the accuracy and precision of interaction quantification derived from molecular techniques remains restricted. This is problematic as information on the strength of interactions between organisms is critical for understanding the structure and function of ecosystems (Berlow et al., 2004).

Another limitation for non-sample-specific methods (i.e., molecular methods applied to whole community samples), is that ecological interactions have to be inferred. This has many caveats which are discussed below (6 Inferring ecological networks from monitoring data), but in essence it generates data on ecological interactions and networks in which we have less confidence.

Finally, there are also some fundamental ecological caveats or limitations of molecular methods and interpreting the data that they produce. Problems around detecting cannibalism and intra-guild predation (Traugott et al., 2013), identifying complex diets in omnivores (Tercel et al., 2021) and swamping of dietary data by predator DNA (Cuff et al., 2023), all limit the degree to which we can construct accurate and representative ecological networks.

6. Inferring ecological networks from monitoring data

Ecological interactions, and the ecological networks they form, can be assembled from alternative data sources, such as population estimates across communities (Morales-Castilla et al., 2015) or trait matching based on a priori knowledge of which traits dictate ecological interactions (i.e., body size). This is a dynamic area of research with significant interest (as directly collecting information on ecological interactions is challenging). There are a variety of different methods that require different levels of data input.

The simplest methods for inferring ecological interactions from community data are those that use associations and correlations for presence/absence and abundance data, respectively (e.g., trophic and non-trophic networks; Sander et al., 2017). These methods require the least amount of data types, but still require a large amount of data across either space or time, to be accurate. There have been significant criticism around these methods (Blanchet et al., 2020), as they provide no indication of an actual biotic interaction, and could simply result from organisms displaying similar habitat preferences. For example, it has been shown in stream fishes that species-habitat relationships were the major determinant of species co-occurrences, as opposed to species interactions (Peres-Neto, 2004). The direction of interactions, i.e., positive or negative, is also difficult to extract for these data. For example, it has been shown that abundance correlations for microbes convey very limited amounts of information on the networks of interactions in modelled microbial communities, and the direction of relationships did not always match the direction of interactions (Pinto et al., 2022) – i.e., negative co-occurrence patterns did not indicate antagonistic interactions. There are more complicated methods of analysis, i.e., conditional dependence-based methods (Feng et al., 2022), and for certain taxonomic groups, without good information on biological traits or other information, this may be the only option.

Going a step further, it is possible to combine co-occurrence data (either presence/absence or abundances), with information on life history, morphology, bioenergetics or other extra biological data for organisms (Pichler et al., 2020). This additional step can range from simple (i.e., rules around size-structure in food webs; Pomeranz et al., 2019) through to complicated (i.e., calculating risk-reward for different seeds based on nutrients and handling time; Pocock et al., 2021). An alternative is using databases of previously observed interactions, for example Database of Insects and their Food Plants, or WebBuilder for trophic interactions in river organisms (Gray et al., 2015). Adding in ecological rules to what are solely statistical relationships increases the accuracy of estimates, but at the cost of collecting or collating those data. For well understood organisms, these data may be easy to come by, however, for organisms where these data are difficult to procure (e.g., micro-organisms) this method may not be suitable.

The most complicated, yet arguably accurate, methods that currently exist aim to explain all potential variation in the presence/absence or abundances of different organisms. The methods use spatially and temporally structured variables, such as site-specific environmental conditions (temperature, water quantity and quality, pH), regional species pools and evolutionary histories (e.g., phylogenies), before then ascribing any remaining, unexplained co-occurrence to biotic interactions (Tikhonov et al., 2020). There are a variety

of different methods that have been used, most of which have been derived from attempts to improve Joint-Species Distribution Models (J-SDMs; Zurell et al., 2018). Incorporating eco-evolutionary information, as well as how these processes manifest themselves in the distribution of organisms, appears to be the most robust method for inferring interactions. Nevertheless, these methods are contingent upon high quality data collected at appropriate spatial and temporal scales, and a large amount of information on the target organisms, all of which restrict the widescale application of methods across taxonomic groups, such as microbes. Nevertheless, this is where data from metabarcoding and metagenomics provides an opportunity to collect data that conforms to the above specifications (i.e., highly replicated spatially and temporally, as well as being able to collect data on traits and phylogenies).

Inferring networks remains a growing area of interest in network ecology, considering the potential for using data that has not been specifically collected to construct networks. It, however, is clear that there are many caveats and assumptions that must be made, limiting the applicability of these methods and necessitating a case-by-case appraisal of the suitability of different techniques. However, it is also the case that for some groups of organisms, inferring interactions is the only feasible method of network construction currently, without a paradigm shift in how we collect data on interactions for microorganisms.

7. Analysing ecosystem-scale properties (e.g., resilience and stability)

There are a variety of options for understanding ecosystem-scale resilience using ENS, as well as using simple monitoring data (i.e., presence/absence or abundances). Resilience, however, is a poorly defined concept (Pimm et al., 2019). As such, it is important to understand the component of resilience or stability that pertains to any given research question, as this will dictate the appropriateness of different metrics or measurements. Below, I detail some of the options for understanding the ability of an ecosystem to respond and be resistant/resilient to environmental change.

7.1 Topological properties of networks

Many topological properties of networks have been suggested to represent the stability or resilience of ecological networks to change. These are often used as proxies, as these metrics do not require dynamic models or other more complicated analyses to understand whether a network is dynamically stable. Below I list some commonly used properties used to describe the complexity (and thus stability), and examples from the literature, as well as information as to the degree to which these metrics can be used to understand stability or resilience.

- **Network size** – the number of nodes (e.g., species richness). This is the oldest measure of complexity (see MacArthur, 1955), but it does not account for a range of important structures within ecosystems. This measure is reliant on the assumption that higher biodiversity generates higher stability, which is known to not necessarily be the case (McCann, 2000).

- **Connectance** – the ratio of observed links to potential links or the overall density of links within a network. It was researchers using connectance or link density as a proxy for the stability of a system that originally kick-started the complexity-stability debate (Landi et al., 2018). Experimentalists showed that simple systems were less stable (Pimm, 1984), whereas some modellers showed that the more interconnected and complex a network is, the lower the system's stability (May, 1974) and others showed the opposite (De Angelis, 1975). As this is a contentious measure of stability, it is not the necessarily the best metric to use, however, it does convey the complexity of a system, if not the systems' stability.

- **Interaction evenness** – the distribution of links across nodes. An uneven distribution of links across a network can mean that it if that node is lost, there is likely to be a large number of unconnected species – which can be problematic for both mutualistic and antagonistic networks. However, this metric does not provide any information on how likely it is that a system changes state or that the well-connected species will be lost. This means that inferring stability from this metric is challenging.

- **Modularity** – the number of modules (well-connected subnetworks) in a network. There are mixed results of the effects of modularity on stability, but generally modularity has a moderate stabilising effect, but the inverse of modularity, i.e., uniform connectance, is destabilising (Grilli et al., 2016). Again, this is not the case in all systems, and information on the ecological meaning of modules is not conveyed in this simple metric.

- **Nestedness** – the degree to which interactions from less well-connected nodes are subsets of the interactions of well-connected nodes. Generally, it is shown that nested networks are less stable (Staniczenko et al., 2013) as the loss of nodes or interactions would result in a wider impact as nodes are jointly connected. Nestedness and modularity are often talked about as being two sides of the same coin (Fortuna et al., 2010), therefore they offer complementary, if not slightly different information on the system.

As described above, many of the metrics have complicated and inconsistent relationships with the stability of ecosystems, thus they do not provide useful proxies, unless under specific circumstances or within highly controlled experiment systems where the behaviour of the organisms or communities are well characterised.

7.2 Network robustness

Robustness is a classic method of assessing the tolerance of an ecological network to the removal of nodes (Delmas et al., 2019). It is a useful technique as robustness can be calculated through a logical set of rules, and it can be used to understand whether certain

network properties (e.g., nestedness, connectance, modularity) convey enhanced or decreased resilience to node loss. In ecological networks, most examples of robustness analyses are completed for inter-specific interaction networks, where the nodes represent species. Therefore, in these situations we are investigating the ability of an ecosystem to tolerate extinctions.

There are many different methods for assessing the robustness of a network to extinctions, but the general framework is consistent:

1. Generate an extinction list (e.g., a sequence of nodes to remove from the network, $node_i$ … $node_{i+n}$)
2. Remove node i from the list (primary extinctions)
3. Assess whether any other nodes in the network have become isolated (e.g., no longer connected to any others in the network)
4. Remove nodes that have become isolated (secondary extinctions)
5. Repeat steps 2-4 until all nodes are lost from the network
6. Calculate the robustness metric (e.g., 50 % loss of species; Dunne et al., 2002)

There are lots of options for customising each step of this procedure, depending on the question being asked and also the data available. Firstly, the extinction list can be based on data collected from the field which convey some indication of extinction risk (e.g., body size or abundances), or some topological property of the nodes (e.g., degree) and therefore increase the ecological realism (Bane et al., 2018; Lu et al., 2016) – i.e., making the correct species extinct in the right order. Secondly, rather than a node becoming extinct once all of its links have been removed, a threshold of loss can be set, such that a node becomes extinct if it loses 50% of its links, or in the case of weighted networks, 50% of the value of its links (Bane et al., 2018). This means that you can adjust the sensitivity of different nodes to primary extinctions and potentially increase the realism of the robustness analyses, as it is not necessarily a suitable assumption to expect a species or individual to persist as long as it remains connected within the network. Thirdly, after a primary extinction, different ecological processes can be simulated, such as interaction rewiring – where a node can create new interactions with a different set of nodes, based on a predefined series of rules. These rules can be: (i) evolutionary – rewiring can occur with other nodes that interact with phylogenetically similar species (Rezende et al., 2007); (ii) morphological – organisms with similar traits can rewire to interact with other nodes. For example, size-matching insects and plants (see Pichler et al., 2020); (iii) phenological – organisms can only interact if they are present in the same place at the same time. For example, plants only flower at specific times of year, so pollinators will only interact with those plant species during defined time periods (Peralta et al., 2020); or (iv) behavioural – some organisms are more or less specialised and therefore can switch to different resources more or less easily (behavioural plasticity). For example, pollinators will make use of resources available to them in a non-selective way – i.e., there is an element of choice when multiple resources are available – and this affects robustness (Kaiser-Bunbury et al., 2010).

It is also possible to enhance the ecological realism of robustness analyses by incorporating dynamic processes, such as short-term changes in population sizes alterations in biological

traits (i.e., changing body size), but also longer-term processes such as evolution and co-evolution (Graham et al., 2023). These methods are known as dynamic or adaptive network models (see Segar et al., 2020), and have been applied to terrestrial ecosystems (Maia et al., 2021). They allow for feedback between ecological and evolutionary drivers of network structure, as would be the case in the real world.

In comparison to other topological measures, network robustness provides a more realistic estimate of resilience or stability, as it involves a logical and ecologically sensible method of understanding changes within existing ecosystems. Other metrics don't take this extra step and instead implicitly suggest stability or resilience.

7.3 Engineering resilience

First proposed by Holling (1973), the idea of using resilience in ecology is interesting, but difficult to define property of an ecosystem. Holling later went on to define two different types of resilience: engineering and ecological (Holling, 1996). Engineering resilience (Table 1) is a more straightforward property of a system to assess as it assumes a single stable equilibrium and one can measure how long it takes to return to this equilibrium. Whereas, more realistic, but more difficult to measure, ecological resilience assumes multiple potential equilibria that a system can occupy.

Methods have been developed to calculate engineering resilience for ecological networks, including both positive and negative interactions (Sauve et al., 2016). Termed 'stability', and calculated as the minimum intra-specific competition (i.e., the smallest level of self-dampening) to achieve stability within a community (Neutel et al., 2007, 2002). We can use these methods to determine how 'quickly' a system can return to its previous state based on a static snapshot of an ecological network. This essentially generates similar data to the methods described in 7.5 Persistence, however, it only requires static data (i.e., data from a snapshot in time), rather than a time series of values.

7.4 Response diversity

Response diversity is the range of potential responses or reactions of organisms to a given stressor (e.g., environmental change) or set of stressors (Elmqvist et al., 2003). The theory is rooted in the insurance hypothesis – that greater biodiversity guarantees that some organisms will maintain functioning even when others are lost (Yachi and Loreau, 1999). It is seen as a crucial mechanism underlying ecological resilience to change, especially during periods when there is reorganisation. Although response diversity is not a new concept, it has seen a recent resurgence in ecology, as a result of new conceptual and empirical thinking and an appetite for consolidating ideas around the empirical aspects of this concept (Ross et al., 2022). This has also led to a push in calculating metrics of response diversity, which until recently has been out of reach.

Ross et al. (2022) outlines a simple, yet robust, framework to calculate response diversity by measuring environment-dependent ecological responses to biotic or abiotic environmental variables. It is based on performance-environment relationships (i.e., the

relationship between biological traits and gradients of environmental conditions; Figure 2). The method is suitable for application to field-based community data, including that collected over time and across multiple sites. It is, however, contingent upon measuring trait information on the different species within the community. Saying this, these data can take different forms, and the authors in fact encourage the selection of the correct data for the questions being asked. As such the framework is flexible and can be coupled with standard biomonitoring (i.e., species abundances) and secondary data on traits (e.g., cell size for microbes). It is feasible to use ecological interactions, associated with defined ecosystem functions, i.e., decomposition or resource production, as the traits in these analyses. There is also potential, for datasets that have multiple environmental variables operating simultaneously, to use environmental condition surfaces, to account for interactive effects of stressors on ecological communities (Ross et al., 2022). This would allow for a more complete understanding of the interactive (additive, antagonistic or synergistic) effects of multiple different stressors.

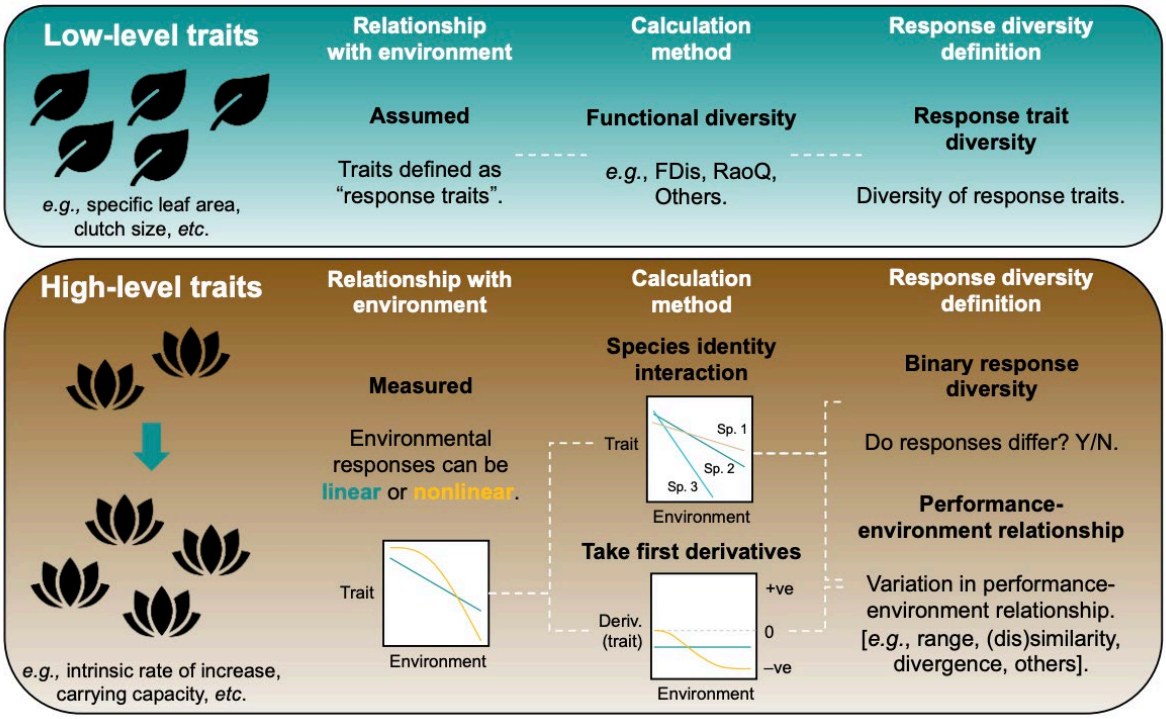


Figure 2. Measuring response diversity using low- and high-level traits. From Ross et al (2022) CC-BY 2.0.

7.5 Persistence

Core to our understanding of stability, is the concept of persistence – the relative continuity in ecological structure or function over time. This is a classic form of analysis in community ecology, where the persistence of community structure is analysed over time to understand different aspects of the system (e.g., environmental variability; Milner et al., 2006). Applying

this same thought process to networks, it is possible to gain information on the temporal or spatial stability or resilience of ecological networks by analysing network structure. This could take the form of assessing specific interactions (i.e., interaction fidelity; Parra et al., 2022), subnetworks (e.g., persistent homology), or the overall network topology.

Persistence, as with other stability metrics, can be measured in a variety of ways. A simple method would be to calculate a similarity/dissimilarity metric for matrices (e.g., Jaccard's index), this would indicate relative turnover in interactions and their strengths, and highlight where changes are occurring most rapidly or significantly. Other more recent methods have also been developed for monitoring persistence at both the subnetwork and network scale, including techniques that use subnetwork persistence to infer wider ecological persistence (see Song et al., 2022).

7.6 Stability

There are many different methods for monitoring the stability of an ecosystem, ranging from data on abundances in highly resolved time series, through to energy landscape analysis (Ross et al., 2021). Irrespective of the method of calculating stability and the data used, inter-specific interactions emerge as dominant drivers (Mougi and Kondoh, 2012). As such, using ecological networks to understand stability is vital.

Options for metricising stability range from simple to complex, and there has been a wide range of different metrics used to describe stability (Donohue et al., 2013):

- **Variability (i.e., coefficient of variation).** Calculated for structural or functional measurement (e.g., biomass in ciliates; Leary and Petchey, 2009). This could be variability in a network topology metric (i.e., connectance or nestedness).
- **Compositional turnover.** Jaccard similarity of communities between consecutive time points. This can also be calculated for adjacency matrices of networks to look at interaction turnover.
- **Number of extinctions (see 7.2 Robustness).** Fundamentally based on interactions, the direct and indirect links between species control the levels of extinction.
- **Number of invasions.** The number of taxa observed in plots at the end of sampling that were absent at the start of sampling. This can also be used to assess interactions that were not present at the start of sampling – i.e., invasion or assembly of new interactions in response to a stressor.
- **Resistance.** Also referred to as 'inertia', it is the extent of change in community structure in response to a disturbance/perturbation. Measured as the inverse of the Euclidean distance of each sample from the centroid of a control treatment. The Euclidean coordinates for sites could be constructed based on the ecological networks they form, rather than the usual community-based approach.

Undoubtedly, there are also other metrics that could be calculated to describe the resistance or resilience of networks to change. However, simple, well-understood and widely used

metrics will provide the most easily understandable results, and can be linked to the underlying mechanisms responsible for stability.

7.7 Functional diversity/stability

To add an extra level of detail, information on the traits of organisms can be included in analyses. For example, comprehensive trait databases exist for macroinvertebrates and fish (https://www.freshwaterecology.info/), but also diatoms and algae (e.g., Lange et al., 2016) and other microbial organisms, such as bacteria (e.g., BactoTraits; Cébron et al., 2021). These data can be leveraged to understand different aspects of the system – i.e., the loss of traits and functional diversity (Escalas et al., 2019) – which is not always related to taxonomic diversity (Mouton et al., 2020). As indicated in above sections (e.g., 7.2 Network Robustness), these data can be integrated into other analysis to provide enhanced ecological realism. They can also be used in their own right, as ultimately the trait profile of a population or community is an artefact of biotic (e.g., ecological interactions) and abiotic conditions.

8. Recommendations for Environment Agency data

The metabarcoding and metagenomic data collected by the Environment Agency can be used in its own right, as well as to infer ecological interactions, to understand the resilience of river ecosystems to environmental change. The following sections draw on the information and methods discussed in more detail in the above sections. Below I provide recommendations along these two lines. First, I run through the data processing and metric calculation that could be completed on the data and then detail the specific analyses that could be completed. This includes an extra section detailing how ENS could be used in conjunction with network models on the river networks themselves to monitor resilience at the landscape scale (accounting for variable downstream and upstream dispersal of organisms).

8.1 Data processing and metrics

Prior to detailed analyses, a range of metrics and data wrangling procedures will need to be completed on the described dataset to generate informative outputs. Below I describe the manipulation and metrics that could be used to get the data into a suitable format for the subsequent analyses described in 8.2 Data analysis.

The metagenomic data can be used to generate community data (i.e., species lists), as well as information on gene presence and function. The first step for these data will be to create a sample unit (site, season and year) by species (either Operational Taxonomic Units or Amplicon Sequence Variants; OTUs or ASVs, respectively) matrix, populated by presence or absence, as well as read numbers (see above discussion of the challenges of such data).

Secondly, data can also be summarised with a sample unit by gene matrix, detailing the levels of gene presence across the sample units.

The matrices can then be analysed in their entirety using multivariate analyses, or a priori information on important species or genes can be used to create univariate metrics for further analysis. For example, from the community data, BMWP or ASPT could be calculated, or from the gene presence data, the diversity of genes associated with antimicrobial resistance could be recorded for each sample.

8.2 Interrogating metabarcoding/metagenomic data

Gene presence data could be summarised in a co-occurrence network, where it would be possible to understand those genes that are more or less associated with one another – this would provide information on the suite of genes present across sites and therefore provide a more comprehensive understanding of potentially important functions, such as antimicrobial resistance (AMR), present across the different sites. As with all methods of network inference, or association networks, caution is required when interpreting the results. However, the comprehensive spatial and temporal coverage of the proposed data will help to ameliorate many concerns commonly levelled at these methods through sheer observational power.

Depending on the gene regions targeted, or if shotgun metagenomics is used to analyse samples (suggested for ~250 sites selected to monitor specific stressors), it might be possible to estimate other parameters of interest. Firstly, AMR could be detected within the samples (Grenni, 2022). Secondly, maladaptation to future pressures could be assessed (Lind et al., in press). Also, the response of organisms to different stressors could be compared by using the combined metabarcoding and metagenomic data which would provide information on species identities and genes present across a gradient of environmental conditions.

8.3 Inferring ecological interactions and networks

A Bayesian analytical framework for Joint-Species distribution modelling, called Hierarchical Modelling of Species Communities (HMSC), provides a brilliant opportunity to analyse the data collected but also construct ecological networks, based on best estimates of biotic interactions. I provide a full description of the entire analysis pipeline in 8.2 Data analysis. For the following description, it is simply important to know that in using this method, variation in the occurrence and co-occurrence of different organisms can be ascribed to environmental conditions (e.g., water quantity and quality, temperature, sediment conditions), traits and phylogeny. Once this variation is accounted for and partitioned, the remaining variance can be ascribed to biotic interactions. Remaining negative associations between organisms could indicate antagonistic interactions (e.g., competition or predation)

and remaining positive associations could indicate mutualistic interactions (e.g., facilitation) or niche partitioning.

The temporal and spatial replication within the EA data provides a brilliant opportunity to implement these analyses. Furthermore, it is possible to extract information on the phylogenies of organisms, either from existing databases (e.g., rotl; Michonneau et al., 2016) by matching species taxonomic identities from metabarcoding data or by constructing phylogenetic trees based on sequenced gene regions (e.g., using BEAST; Drummond and Rambaut, 2007; Elias et al., 2013). Doing so would allow for the most accurate estimations of ecological interactions as all possible mechanistic drivers (environmental, temporal, functional and phylogenetic variation) are included in the analytical framework. It important to note that there has been significant criticism of abundance correlations for constructing networks for microbial communities (described above in 6 Inferring ecological networks from monitoring data). Nevertheless, HMSC goes beyond the methods commonly criticised (i.e., correlating abundances without accounting for the influence of other abiotic and biotic factors), and therefore appears more robust.

Using HMSC it would be possible to create merged networks of positive and negative associations between OTUs from the metabarcoding data. Using these networks, there are a range of other analyses that could be used to determine response diversity, or other metrics that indicate the potential resilience of ecosystems to further environmental change. Some network-based suggestions for further analysis include:

- Interaction diversity
- Network robustness
- Engineering resilience
- Phylogenetically structured networks

All methods above could make use of additional information on biological traits or other information that is available for the taxa assessed.

8.4 Response diversity

Using the large number of sites along different environmental gradients it will be possible to calculate response diversity for different ecosystems using the framework laid out in Ross et al. (2022). Briefly, along the gradients of different conditions (e.g., urbanisation, water quality or nutrient status) a performance-environment curve can be generated for each taxon (i.e., abundance, or proportion of samples detected for temporal data). Summaries of the performance curves, provided by first derivatives of generalised additive models, can then be used to calculate diversity using hill numbers. The diversity values could then be compared across catchments or regions to understand the resilience of those areas to further change.

Response diversity could be calculated for sites based on temporal data collected over the duration of the study. Coupled with the water chemistry data collected at the same time as biological samples, it would be possible to calculate performance-environment relationships for different drivers across sites (e.g., turbidity, water temperature, flow rate) using the framework described in Ross et al., (2022). Data could then be interrogated in a number of ways to allow for a site-specific and catchment-level understanding of the potential ecological resilience to further changes.

Outcomes of these analyses could be linked into the ENS framework, by looking at similarity in responses between interlinked species (i.e., is the reason for similar responses actually a result of interactions between those organisms), and also using these data to inform robustness simulations (i.e., in the extinction sequence organisms with similar responses are lost in close succession to one another).

8.5 Spatial multilayer network (meta-community approach)

The dendritic nature of river systems means it is possible to look at resilience at the river catchment scale. A meta-community approach (e.g., Mougi and Kondoh, 2012), including ecological interactions, could facilitate an understanding of the resilience of individual sites within the context of immigration and emigration of organisms, and thus recolonisation post-disturbance. This would allow for an understanding of resilience (including recolonisation processes from other sites) beyond that of most studies that focus on the resistance of an ecosystem to change.

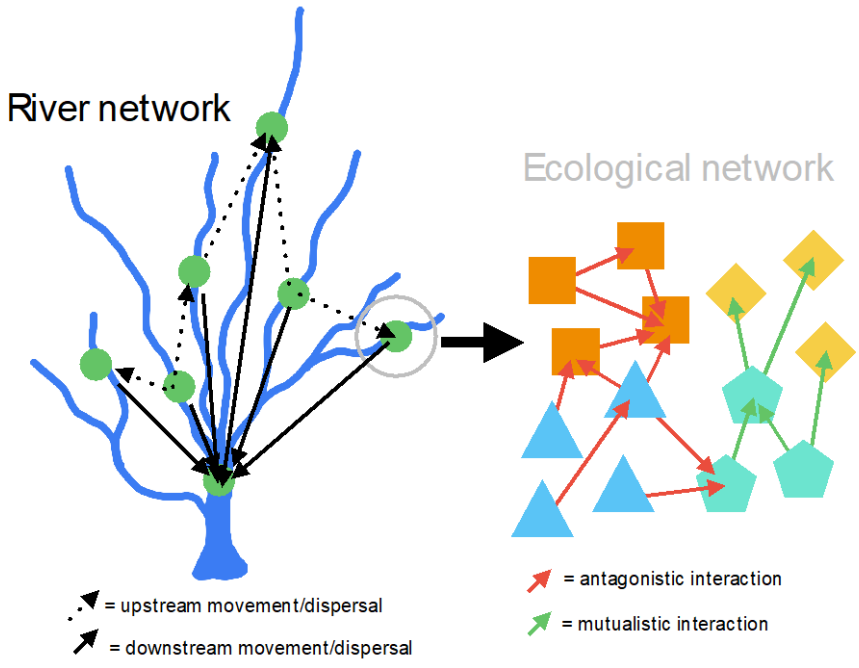


Figure 3. Diagram of a meta-community approach for the EA dataset. Sample locations can be connected using a spatial network, with downstream and upstream dispersal for different types of organisms (distance decay functions could be used depending on the dispersal ability of the organism). Ecological networks at each site constructed using inference.

Any potential model could be structured around a multi-layer network framework (Figure 3) with the following properties and parameters:

- Nodes
    - Organisms (OTUs or ASVs for the metabarcoding data)
- Intra-layer links (site level)
    - Ecological interactions between different organisms (OTUs or ASVs for the metabarcoding data)
    - Positive and negative interactions between organisms
- Inter-edges (river network level)
    - Organism dispersal links sample locations together into a meta-community
        - Aquatic organisms can only disperse through the river network (both passively downstream and actively upstream)
        - Terrestrial life stages of some organisms can disperse over land
        - Some organisms cannot disperse upstream depending on the strength of their dispersal
        - The distance of dispersal for organisms is dependent on their traits (i.e., dispersal distance over land or swimming strength, or distance covered per unit time)

Different stressors could then be simulated at the catchment scale. For example, the upstream links for aquatic organisms could be removed between certain sections of rivers to simulate the installation of a weir or a dam preventing upstream dispersal. As a further example, species could be removed across all sites to simulate species loss due to non-specific stressors, such as water temperature or flow restrictions. This modelling approach could also be tied into stability and response diversity calculations. For example, the probability of recolonisation of an organism from a proximal site, derived from this model, could be used to adjust the extinction probabilities in robustness simulations. This would help to understand the role of the river system in affecting site-specific conditions.

9. Conclusions

ENS provides a potential opportunity to understand how the structure, function and resilience of ecosystems change in space and time. It allows for the integration of many types of data and therefore provides a flexible framework that can make use of all available data, without being overly reliant on any one data source. Gaining information beyond that of just presence/absence and abundances also offers lots of opportunities for predicting

changes to ecosystems, and the potential to identify early warning signals of future biodiversity loss.

Combining ENS and molecular methods appears to be a particularly effective way of constructing ecological networks at large spatial and temporal scales, and is also appropriate for understanding the ‘unseen’ biodiversity (i.e., microbial communities).

Although ENS is methodologically and theoretically challenging, it presents a unique suite of techniques to understand the natural world, and how it responds to environmental change and anthropogenic pressures.

Think piece 2: Developing an understanding of ecological network science (ENS) for river ecosystems

Darren Evans[1]

[1]School of Natural and Environmental Sciences, **Newcastle University**
Agriculture Building, King's Road, Newcastle upon Tyne, NE1 7RU

Summary

The Environment Agency (EA) plans to generate a comprehensive DNA dataset from river biofilm samples collected over a three-year period as part of their newly designed River Surveillance Network (RSN) programme. This will generate microbial community data using high throughput sequencing, including metabarcoding data for bacteria (16S), fungi (ITS), diatoms and other photosynthetic microbes (rbcL), protists and other eukaryotes (18S).

By using data generated from ~250 sites (500 samples) linked to a specific pressure (e.g., sewage treatment works or nutrient gradient), there is an unprecedented opportunity to examine how biodiversity and ecosystem functioning responds in the most holistic way to date. A key EA aim is to explore the potential to define and characterise the resilience of ecosystems using microbial networks in aquatic freshwater systems. However, because microbial networks are not yet well developed in terms of their inference and interpretability, there are numerous challenges to consider in the construction and analysis of freshwater microbial networks. But there are some very promising developments in network ecology for assessing the vulnerability of terrestrial and aquatic ecosystems based on modelling perturbations and species extinctions.

This Think Piece will address EA's technical requirements by providing: 1) a short overview/tutorial of network theory, including a description of species- and network-level properties (both qualitative and quantitative) that can be studied; 2) an overview of the pros and cons of eDNA based methods for generating network-specific data; 3) an introduction to network construction methods, with a detailed overview of the pros and cons of interaction inference methods that would be applicable to eDNA generated data (e.g. co-occurring organisms found within biofilm samples); 4) an overview of the ENS research landscape and how this can be applied in the context of biomonitoring (esp. the study of freshwater systems; 5) applications of network theory to study the resilience of interacting communities to perturbations and species loss (e.g. robustness and adaptive network models), and how this can be applied to microbial networks; 6) a critical evaluation of how and why ENS has potential for examining the resilience of freshwater systems, whilst identifying the knowledge gaps; 7) recommendations to explore multi-taxa microbial datasets, with a particular focus on multi-layer networks.

Given there are multiple steps involved in sample collection, laboratory processing, DNA/RNA extraction and metabarcoding, network inference and analysis, the Think Piece will create a workflow that identifies the pros and cons of each (including knowledge gaps), ultimately resulting in the construction of highly-resolved ecological networks. Finally, it will provide recommendations for how the specific data can be analysed, with potential for better understanding biodiversity and ecosystem functioning relationships in rivers, but with a focus on new, network-derived metrics that could be used for biomonitoring.

It should be noted that although much of the seminal food-web studies focussed on aquatic systems (e.g. Cohen et al., 2009), most of the recent advances in network ecology have tended to focus on above-ground, bipartite terrestrial networks, especially mutualisms (e.g. plant-pollinator networks). Thus, in building the workflow for the construction and analysis of microbial networks, I will mostly draw from such examples, but acknowledge that microbial systems can operate quite differently, have different architectures (e.g. Toju et al., 2014) and potentially high variation (Barroso-Bergadà et al., 2021). In other words, the study of microbial networks, generated using advanced molecular methods, is still in its infancy.

1. What are ecological networks and how can they be put to use in river ecosystems?

First, it is important to define what we mean by an ecological network as they can represent different concepts in ecology, for example a foodweb, host-parasite interactions, or even a network of river systems. Here, I will refer to **interspecific species-interaction networks**, as these will be the main types derived from the type of data generated using molecular methods. However, before we can understand how they can be used to study aquatic freshwater systems, it is important to understand some basic concepts of network science.

1.1 The Basics

Network ecology is rooted in the mathematical field of graph theory where a ‘graph’ is a representation of a set of entities and where some pairs of these entities are connected. In this context a graph is made up of ‘vertices’ (singular: a ‘vertex’) or ‘nodes’ which are connected by ‘edges’ or ‘links’ (Fig.1.1). In essence, that is it: a network comprises pairs of nodes connected by links.

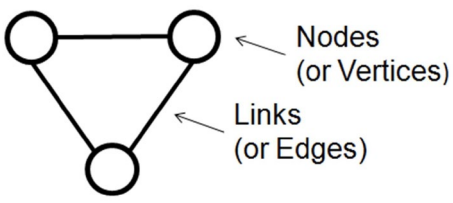


Figure 1.1 - A network comprises pairs of nodes connected by links

In many ecological networks, the links represent 'interactions' which have a specific context and meaning. The different terms are used interchangeably within the network ecology literature, partly because ecologists draw quite heavily from (and use tools developed by) scientists working in other fields, especially computer science. The terminology can sometimes cause confusion so it is worth taking note what the difference is between, 'vertices' and 'links' (see Table 1.1). In general, I will use the terms 'nodes' and 'links' throughout this Think Piece because these are widely used and easier to understand. A comprehensive overview of networks is provided by Newman (2018), and Jensen (2022) examines their emerging properties, some of which we examine here.

Table 1.1. A simplified explanation of the link between terms in network ecology. Although in this Think Piece I will mostly use the terminology associated with complexity science, some of the analytical approaches use the terms from graph theory, so it is important to be able to translate from one to the other.

Descriptor	Complexity science	Mathematics (Graph Theory)
The whole system	Network	Graph
The individual entities	Node	Vertex (plural: Vertices)
Their connections	Link	Edge

1.1.1 Defining nodes and links from eDNA generated data

Within ecological networks, nodes can represent a range of entities from individual species to populations to habitat patches or sites. Here, I consider them as individual 'species', or organisms mainly derived from eDNA microbial community data (i.e. bacteria, fungi, diatoms and other photosynthetic microbes, protists and other eukaryotes) but with all of the caveats of applying the species concept to microbial systems.

The links between nodes will vary incredibly according to the network being studied, but can include (but are not restricted to):

- A trophic relationship (antagonism) that is feeding upon something.
- A mutualism, such as pollination or seed-dispersal.
- Movement of individuals or genetic material.
- Energy or nutrient flow.
- Passing on information or disease.

- Associations and co-occurrence.

Links usually join two nodes, but it is conceivable that links can go from one node back to the same node. These are called self-links and the most obvious example of this is cannibalism in food webs.

A major challenge with microbial community data derived from eDNA is to know how these organisms are interacting. This will be discussed later, but for now we will assume that node and link information can be derived to construct highly-resolved species-interaction networks.

1.1.2 Different types of networks

A graph can represent a network of ecological interactions that may be undirected, meaning that there is no distinction between the two nodes associated with each link (Fig. 1.1), or its links may be directed from one node to another (Fig. 1.2a). In the latter case, this might show which organisms are consuming which within a food-web,for example. The degree of a node is the number of links that connect to it, where self-links are counted twice.

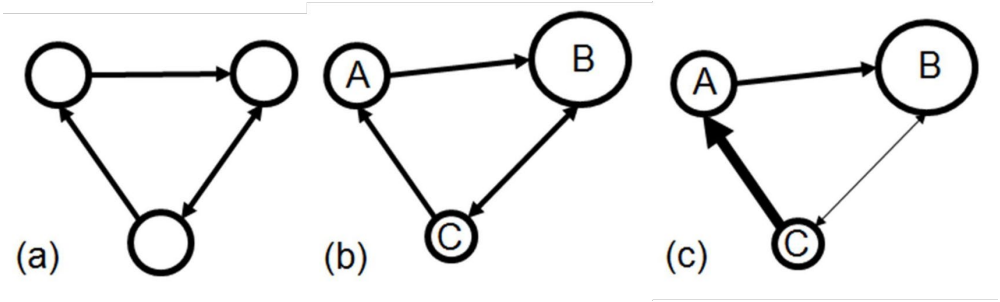


Figure 1.2. Adding detail to a network including (b) individual species and abundances and (c) strength of interactions.

The number of nodes (or species) and links (interactions between species) provides the basic information required to begin to calculate important **'emergent properties'** of the network such as complexity (e.g. the average number of links per node), connectance (i.e. the proportion of possible links between nodes that are realized: links/nodes^2) and measures of the importance of the nodes, such as degree centrality (covered later). However, Figure 1.2a tells us rather little about the network, for example whether the nodes are the same species or different, whether they represent individuals, populations or species, and whether the entity represented by the node is abundant or rare, large or small. Ecological networks can include such attributes of the nodes (although it is worth noting that most network metrics do not incorporate node attributes, although node attributes are extremely useful when constructing ecologically meaningful null models for analysis) The example in Figure 1.1b is weighted by species abundance and we can see that there are more individuals of species B (shown by a largest circle size), with species C having the least individuals (shown by the smallest circle).

Likewise, although the existence of a link provides some information about the presence of pairwise relations between the nodes (i.e. the 'topology' of the network), it does not show

the strength of the interaction and whether one pair of nodes is interacting more relative to another pair. Graphs with only information about the presence of links are described as qualitative or unweighted or binary networks. However, in addition to assigning specific information about the nodes, graphs can be weighted to include information about the interaction ‘strength’, such as the frequency of interaction between nodes. These networks are then called weighted or quantitative networks. Figure 1.1c shows the same network as Fig 1.1b, but here the interactions are weighted showing that interaction strength is highest between species C and A and is lowest between species C and B. Such information is important when going on to calculate some of the structural properties and metrics of the network.

One other distinction between different types of networks, which is especially important in ecology, is when there are two distinct types of node. These networks are called **bipartite networks**, in contrast to the unipartite networks described above. The key aspect to bipartite networks is that links go from one type of node to another type of node; nodes of the same type are not linked. Examples of these include plants and pollinators, plants and seed dispersers, species and sites (or habitats). In the bipartite network shown in Fig. 1.3 (Memmott, 1999), a pollinator could visit any of the plants (in theory, at least, because if a link is not observed then it is relevant to ask why not), whereas it is meaningless to consider a pollinator visiting another pollinator, or a flower being visited by a different flower.

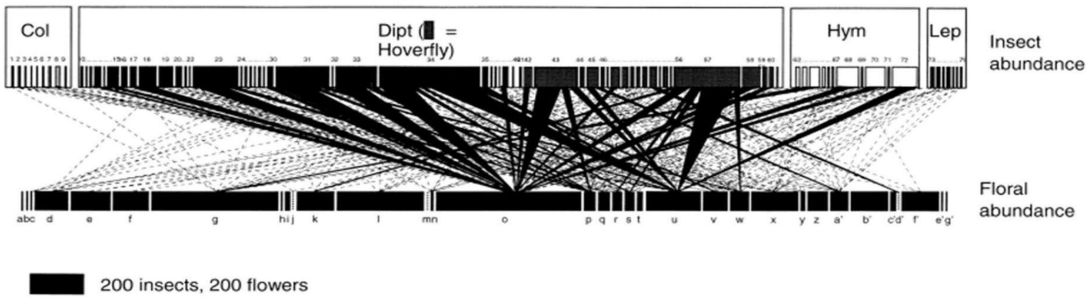


Figure 1.3. Seminal work by Memmott (1999) showed the structure of a quantitative, bipartite plant-pollinator network. Rectangles show the abundance of different plant and pollinator species, with triangles showing the frequency of interactions between them. It is notable that even in this example, not all insects are identified to species-level, but some attributes of the nodes (e.g. family) are included.

Given the prevalence of bipartite networks, especially in ecology but also in other areas of complexity science, there are particular analytical methods and visualisation tools which take account of the bipartite structure of the network. This could be important for the study of river plant-microbe interactions, for example. Further layers can be added to create tripartite and increasingly multipartite networks (e.g. plants-fish-parasites). In terms of analyses, these multipartite networks can either be considered as bipartite networks layered on top of each other, or bespoke methods can be used to analyse them. For many

multipartite networks (e.g. in food-webs) it is not really meaningful to distinguish between the different types of node and so they are almost always considered as unipartite networks. This could well be the case for microbial networks derived from biofilm. Of course, bipartite networks can also be undirected or directed. Traditionally the networks of animals visiting plants are usually considered as undirected bipartite networks (because the link represents a mutualism). However, they could also be considered as directed bipartite networks with the links from flower to the animal representing trophic interactions (gaining energy and nutrients by feeding on nectar or pollen) and the links from the animal to the flower representing, ultimately, fertilisation (and although pollination biology is a fascinating topic we shall not explore its complexities furthermore). Examples of different types of ecological networks are described in Table 1.2.

Table 1.2. Examples of the different types of ecological networks based on link direction and weighting.

Network type	Link direction	Weighting	Examples
unipartite	undirected	unweighted	simple social network
unipartite	undirected	weighted	habitat patch networks, weighted by a function of distance; some social networks
unipartite	directed	unweighted	habitat patch network with thresholds; many food-webs
unipartite	directed	weighted	food-webs with weighted links (e.g. feeding rates or nutrient flow)
bipartite	un/directed*	unweighted	simple two-level food-webs or mutualistic networks
bipartite	un/directed*	weighted	as above, but with weighted links

* although many bipartite networks are treated as if they are undirected, the implication is that either the direction is obvious (parasites parasitise their hosts, not the other way round) or it is a mutualism and so the implication is that the benefits gained by both parties are symmetrical (e.g. flowers are pollinated and insects gain food through nectar, which technically means that the network is a multiplex network, with different types of links). For convenience I will usually refer to bipartite networks as undirected.

Figure 1.4 illustrates the value of a network approach at different scales with empirical data, showing how lots of information can be contained in a fairly intuitive visualisation. For example, the complexity of the whole system is demonstrated in the top figure, while detail of the context of a few different species is shown at the bottom. The bottom figure enables us to see something of the diversity (and, for those who know it, the phylogenetic diversity) of the birds and the fruits. The important links are emphasised by their width and the asymmetry of the interaction is shown by the relative width of the green arrows (birds feeding on fruits) compared to the yellow arrows (seeds being dispersed by birds). Crucially, **the network shows how it is possible to study biodiversity (across two trophic levels) and ecosystem processes (i.e. seed dispersal) in a single conceptual framework**, something I will return to later

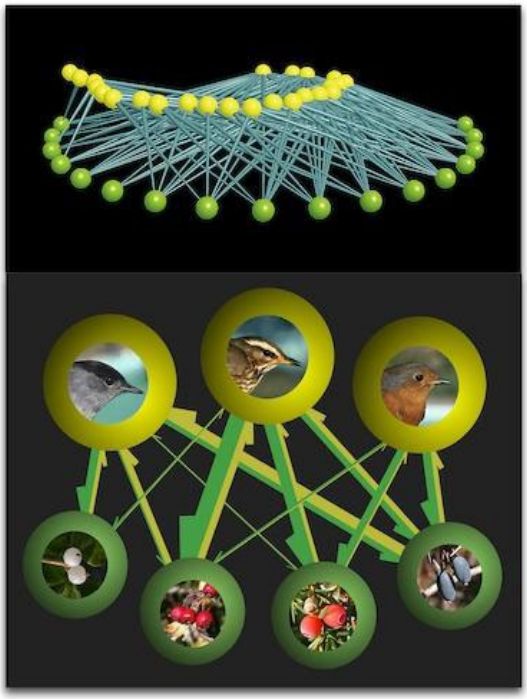


Figure 1.4. An example of a bipartite network showing the complexity of qualitative, bipartite interactions and quantitative information of the bird seed-dispersal network. Image courtesy of Pedro Jordano.

1.1.3 Networks and scales

So bringing it all together, networks are valuable conceptually because they provide a scientific way of approaching data on apparently complicated systems at different scales (Fig. 1.5). They can also incorporate lots of community data (including multiple microbial components) as whole networks, or networks of networks.

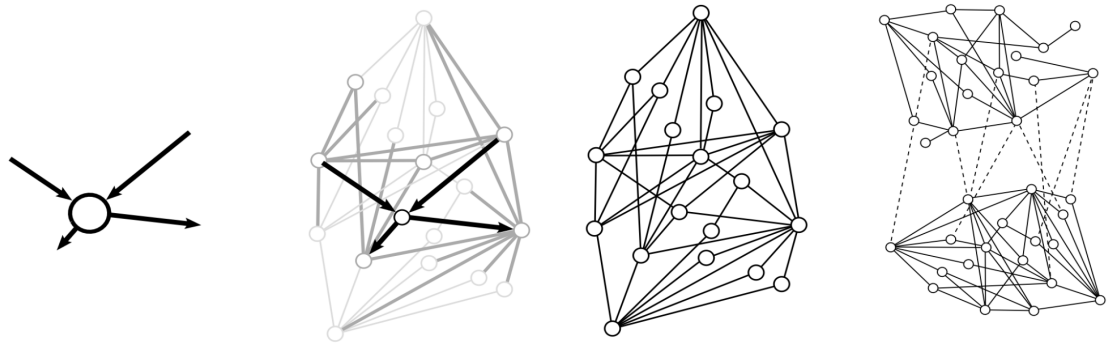


Figure 1.5. Node (a) through to network-level attributes (c) and networks of networks (d) can be studied using advances in complexity science.

At the scale of the whole network (Fig. 1.5c), a complexity science approach provides a way of simplifying the description of the arrangement of links between nodes, and so describing the **'emergent properties'** of the whole system. There may be many links or few, they may be clustered or evenly distributed, the system may be nested or modular (see below). Some of these descriptions are incredibly simple, others are much more mathematically complex. I will discuss these descriptions and metrics later, but the point is that they can be mathematically described.

Secondly, at the very local scale, nodes can be described (Fig. 1.5a). This is, in some senses, a trivial property because a full network is not needed in order to describe the individual nodes. However, to put simply, a network allows, for all the nodes in the network, a complete description of the number of links to or from a node for (i.e. the **'degree'** of the node, which for directed networks can be separated in the in-degree and the out-degree). This helps to define how generalist or specialist a species is (with in- and out-degree defining its degree of generalism as a predator or prey), with other measures of closeness centrality and betweenness centrality etc. depending on specific questions.

Going up a step in the scale, networks allow a simple way of assessing individual nodes in their context in the system – this is what can be referred to as the meso-scale (Fig. 1.5b). For example, the vulnerability of a specialist insect which depends on a specialist plant (e.g. a close co-evolutionary relationship) is very different to the vulnerability of a specialist insect which depends on a generalist plant. Another example is that the 'importance' of a habitat patch acting as a stepping stone between two regions is very different to the importance of a habitat patch acting as a network hub. Whether the importance is greater or less depends on how you define 'importance', which is a topic covered in depth by Jensen (2022).

Finally, there is the scale greater than the individual network: meta-networks (Fig. 1.5d). So far, relatively few studies have considered these more complicated systems. One example would be networks which share the same nodes but the nodes are linked by different types of interaction. An example of this could be trophic interactions and seed

dispersal (Fig. 1.4), or maybe disease transmission and association in a social network. These types of network are technically called multiplex or multilayer networks (although mutualistic networks such as seed-dispersal or plant-pollinator networks are, in a strict sense, directed multiplex networks, they are usually considered as undirected networks). Another type of network is the network of networks, where individual networks are linked by shared nodes (e.g. the shared plants in Pocock, Evans and Memmott, 2012) or nodes in the different networks are linked together (e.g. power stations and internet hubs linked by their physical proximity in Buldyrev et al., 2010). Conceptually there is also the potential for networks to be hierarchical, so individual networks nested within networks. Examples of these could include metacommunities, where communities of interacting species exist in a network of habitat patches linked by dispersal of individual species. In general these meta-networks have not been well studied and so their importance and usefulness in ecology is an important research question.

1.1.4 Examining the structure, complexity and dynamics of ecological networks

Ecological networks describe the interactions between species, the underlying structure of communities and the function and stability of ecosystems (Montoya, Pimm and Solé, 2006). They have the potential to quantify the effects of environmental change on a wide range of complex ecological interactions (Tylianakis et al., 2008).

We have already seen in 1.1.2 that once we have information about nodes and links in a network, we can begin to describe it using qualitative and quantitative metrics in order to look for universal patterns, or to model the **extinction dynamics** using simulations.

For example, Fig. 1.6 shows some potential ways in which bipartite networks could be structured, from discrete groups of interacting species (**compartments**, Fig 1.6B) or a highly nested structure (Fig. 1.6C). Many network studies examine nestedness and modularity. **Nestedness** measures the number of interactions shared between nodes, and **modularity**, the number of interactions shared between a subset (module) of nodes (reviewed by Bascompte and Jordano, 2013). Bersier et al. (2002) used the Chesapeake Bay ecosystem to calculate food web properties to demonstrate differences between species, links, chains and even consumer-prey asymmetries, paving the way for networks to be described both qualitatively and quantitatively.

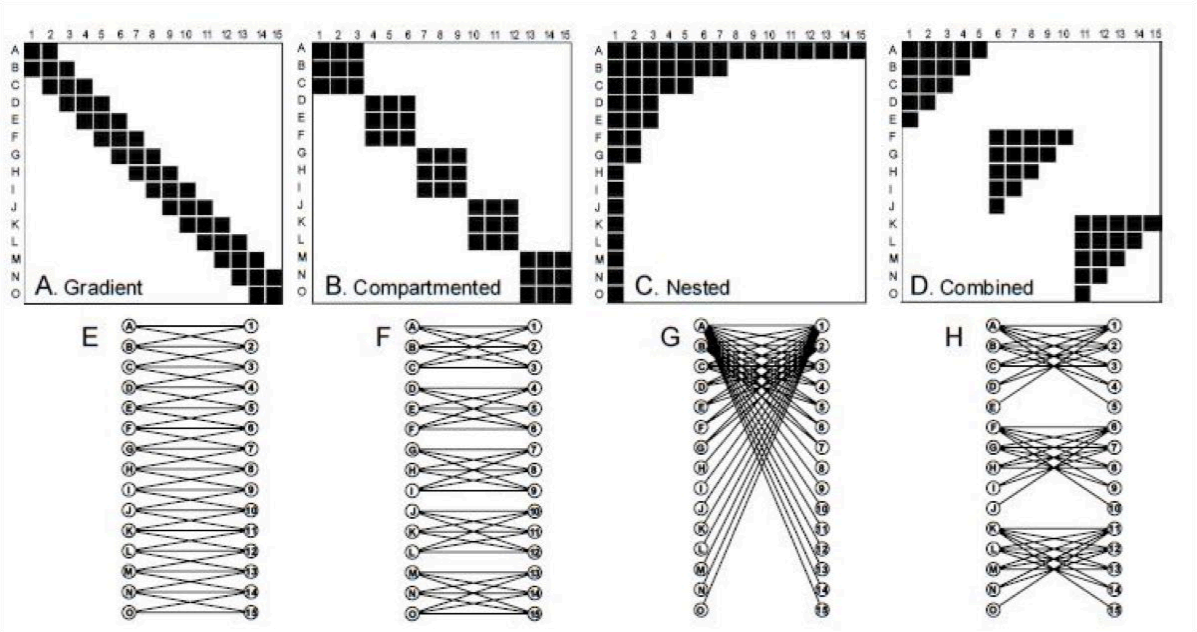


Figure 1.6. The interactions between species can be examined in matrices (A-D) and visualised (E-H). In this example of bipartite networks, letters can represent plants and numbers represent different microbes interacting with them.

A network level approach allows one to account for the whole community scale (Dunne, Williams and Martinez, 2002), thus integrating all **direct** and i**ndirect interactions** (Berlow et al., 2009). Indirect effects (e.g. shown in Fig. 1.5c) have been shown to drive coevolution in mutualistic networks (Guimarães et al., 2017) and need accounting for when predicting how extinctions affect the integrity of ecological networks (Pires et al., 2020). Thus, because networks are multi-faceted objects with a rich range of structure, ecologists have been looking for **emerging properties** that can be easily measured and analysed, and that relate to **ecological properties and processes**. A number of challenges exist in applying these approaches to microbial networks (discussed later).

There are a burgeoning number of freely available software packages to calculate qualitative and quantitative network metrics in ecology (e.g. Dormann et al., 2009; Blonder et al., 2012; Vaughan et al., 2018) as well as calls for syntheses and standardisation (Lau et al., 2017).

1.1.5 The assembly of complex plant-fungus networks

Drawing this together, I present a bipartite plant-fungus network by Toju et al. (2014), generated less than 10 years ago, to demonstrate the state-of-the-art in terms of network construction methods for poorly described systems, and how the the network ‘architecture’, ‘topology’ or structure can be studied (Fig. 1.7). The study used next-generation (454) sequencing technology, to uncover the network architecture of below-ground plant–fungus symbioses. They examined the symbiotic network of a temperate forest in Japan, which included 33 plant species and 387 functionally and phylogenetically diverse fungal taxa. In the absence of species-specific data, molecular analyses provide fungal operational taxonomic units (OTUs), which can be used as an alternative identifier within the ecological network.

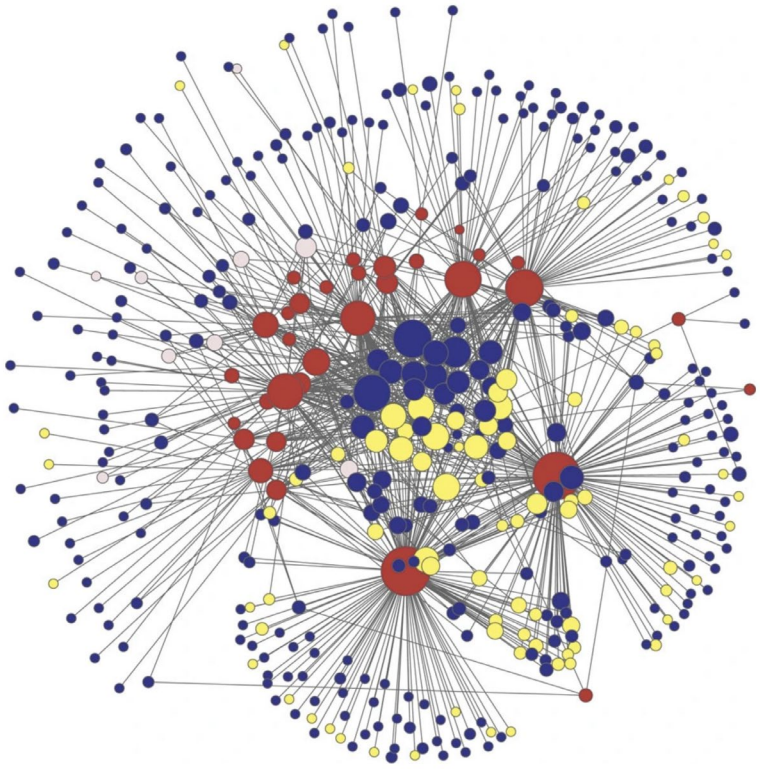


Figure. 1.7. The architecture of the below-ground plant–fungus network in a temperate forest in Japan. Source: Toju et al. 2014.

In the bipartite network, plant species (red) interact with ectomycorrhizal (yellow) and arbuscular mycorrhizal (pink) fungal OTUs as well as OTUs with unknown ecological functions (blue). The size of nodes represents the relative abundance of plant species or fungal OTUs in their dataset. The authors compared measures of modularity, nestedness and specialization (H'_2) with null models and found that overall network architecture differed fundamentally from that of other ecological networks. Although a relatively simple study, it did show that incorporating microbial data into species-rich ecological networks meant that they are more architecturally diverse than previously recognized. It also provided a roadmap for the construction and analysis of microbial networks using molecular methods.

1.1.6 Robustness

Of the numerous ecological network properties, network ‘robustness’ [a measure of the tolerance of the network to species extinctions (Dunne, Williams and Martinez, 2002; Memmott, Waser and Price, 2004)] has received particular attention, partly driven by advances in computational modelling (Kaiser-Bunbury et al., 2010; Staniczenko et al., 2010), but mostly by the desire to understand the real threat of biodiversity loss to ecosystem services and functioning (Evans, Pocock and Memmott, 2013). Figure 1.8 shows how robustness is calculated as the area under the curve by plotting the number of secondary extinctions that occur when, for example, the prey (bottom level) for fish species (top level) go extinct one-by-one. Secondary extinctions occur when the fish are no longer connected to the network due to the loss of their food sources. Simulations can be run thousands of times on a computer until the network completely collapses.

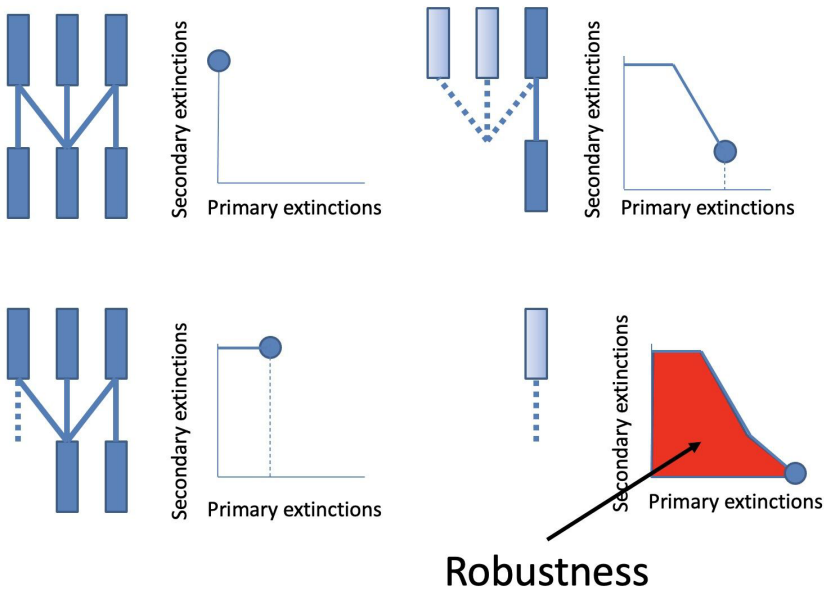


Figure 1.8. A simple demonstration of how the sequential loss of species in the lower trophic level (primary extinctions) affect the number of secondary extinctions in the higher trophic level. Eventually, the network collapses and robustness is calculated as the area under the curve.

Our understanding of network robustness to species loss has advanced from studies of simple qualitative, bipartite mutualistic networks (Memmott et al. 2004), to investigations of patterns across ecosystems (Srinivasan et al., 2007) and to quantitative approaches that take into account species abundance (Kaiser-Bunbury et al., 2010). It is worth noting here that the models used can identify species (or hubs) that are disproportionately important to network integrity, which could be used for targeted restoration in order to build **resilience** in ecosystems, although these ideas are still theoretical (Pocock, Evans and Memmott, 2012; Raimundo, Guimarães and Evans, 2018). It is the focus on primary and secondary extinctions that I will return to later in the context of assessing river ecosystem **vulnerability**.

In summary, networks encode the interactions between the components in complex systems and play an essential role in understanding their structure, complexity and stability. Microbial ecological networks can provide a system-level insight for comprehensively understanding complex microbial interactions (Lv et al., 2019), which play important roles in microbial community assembly and ecosystem functioning.

2. The pros and cons of eDNA based methods for generating network-specific data

Most of the examples provided so far assume that all species-interactions within an ecosystem can be readily determined to create highly-resolved ecological networks. In

reality, network ecology is beset with sampling issues regarding sampling completeness (Macgregor, Evans and Pocock, In press), problems regarding specimen processing/identification (Derocles et al., 2015) and a range of other biases. However, there has been progress in these areas and it is generally recognised that a combination of sampling methods will result in the best resolved networks (Evans and Kitson, 2020).

Advances in DNA sequencing technologies are resolving previously intractable questions in functional and taxonomic biodiversity and provide enormous potential to determine hitherto difficult to observe species interactions. Combining DNA metabarcoding approaches with ecological network analysis presents important new opportunities for understanding large-scale ecological and evolutionary processes, as well as providing powerful tools for building ecosystems that are resilient to environmental change (Evans et al., 2016). Several challenges, however, surround the use of metabarcoding, especially when metabarcoding-based interactions are merged with observation-based species interaction data. These include difficulties surrounding the **quantification of species interactions**, **sampling perspective discrepancy** (i.e. zoocentric vs. phytocentric sampling), experimental biases, **reference database omissions** (i.e. most organisms still lack a DNA barcode) and assumptions regarding direct and indirect consumption events, for example (Cuff et al., 2022). But these problems are not insurmountable and considerable effort is currently underway to resolve such issues (e.g. Cuff et al., 2023).

The major advantage of eDNA approaches for generating network-specific data is the ability to scale-up in space and time. Metabarcoding allows more efficient processing of samples, and therefore the analysis of larger numbers, compared to conventional methods (e.g. microscopy). Constructing multiple replicated networks across a range of treatments, sites or time points, and testing for structural differences, comprises a powerful alternative, although it can be hampered by the difficulty of generating sufficient data for multiple, well-sampled networks. For metabarcoding, investment mainly scales per plate (≤96 samples) rather than per sample (Derocles, Bohan and Dumbrell, 2018), whereas for microscopy, investment of materials and especially time increases linearly for every sample.

2.1. Considerations for microbial networks

While powerful, DNA-based approaches do suffer from limitations important to consider when interpreting results of subsequent network analyses. Toju (2015) reviewed the use of DNA barcoding in microbial ecological network analyses, using the term to represent taxon identification based on specific DNA sequences and consistent with DNA metabarcoding approaches. That review addressed issues related to **sequencing-based approaches**, including a list of common **genetic markers** for different taxa, an approach for inferring interaction frequencies from association networks and host frequencies, and tools for data processing and analysis. As such, it is a valuable guide for researchers embarking on surveys of ecological associations. However, there are further issues not immediately apparent which may represent substantial sources of bias and inaccuracy when characterising microbe association networks, especially biases associated with **isolation**,

amplification, and **sequencing of DNA** and interpreting **weights of interactions** (see Bennett, Evans and Powell, 2019 for a review of plant-microbial networks).

It is important to acknowledge that, to date, our knowledge of ecological network architecture largely stems from empirical studies on macro-organismal systems such as those of plant–pollinator, plant–seed disperser, and prey–predator interactions described earlier. Whilst there has been a rapid growth in plant-microbe studies (that take advantage of DNA-based methods), microbial ecological networks are still in their infancy of both network inference and biological interpretation (Lv et al., 2019).

2.1.1 Interaction data is missing from eDNA studies

Environmental DNA (eDNA) has seen a significant increase in application in freshwater systems with a concurrent growth in protocol developments and a drive to gain a better understanding of ecological interactions and ecosystem processes (Schenekar, 2023). Whilst these approaches appear to be developing well for assessing the ecological status of freshwater systems (e.g. Vasselon et al., 2017), they generally **do not provide information on important ecological relationships**, ranging from mutualism to competition, that in addition to other factors (such as niche preferences and random processes) are known to shape microbial abundances (Faust and Raes, 2012).

2.1.2 Influence of abiotic and biotic factors on aquatic eDNA studies

How abiotic factors influence the transport, persistence, and fate of eDNA in ecosystems remains a substantial challenge. In aquatic systems, numerous abiotic factors, including stream flow (Curtis et al., 2021), substrate type , temperature (Jo et al., 2019), UV light (Kessler et al., 2020), and pH (Strickler et al., (2015); reviewed by Harrison et al., (2019)) impact the availability and longevity of eDNA. Furthermore, microbial activity can play a significant role in the degradation of eDNA, ostensibly impacting botanical eDNA from all media (Zulkefli, Kim and Hwang, 2019). In short, while some studies have examined the abiotic impacts on the availability, longevity, and transport of botanical eDNA (Zhu, 2006; Poté, Ackermann and Wildi, 2009; Yoccoz et al., 2012), considerable research is required on a per system basis to understand the full intricacies and interplay of abiotic factors. The key point is that **these factors can have significant implications when trying to construct accurate, highly-resolved microbial networks**.

2.1.3 Contamination

Across eDNA studies, the utilisation of field, extraction, and amplification blanks, sterilised equipment, and bleach solutions to control for contamination control have been established as best practices to account for contamination at all stages of sampling and sample processing (see Johnson et al., 2023 for a review). It is vital to ensure that biotic

contamination is both understood and proper control procedures are in place, as this too can be an important consideration for network construction methods and analyses if 'false' organisms are detected.

3. Building microbial networks from community presence-absence data generated from eDNA

Microorganisms form complex ecological interactions, including mutualisims such as cross-feeding and cooperation interactions, antagonisms such as predator-prey and host-parasite interactions, and loss-loss relationships such as competitive exclusion interactions (Faust and Raes, 2012). These microbial interactions are known to be critical properties of microbial communities and play important roles in microbial community assembly, although they are poorly understood in freshwater systems.

3.1 Network inference

To date, microbial interaction networks (especially those generated from eDNA samples) are mostly created using **network inference** methods, as most (if not all) of the organism interactions are not known. Network inference is "the process of reconstructing the wiring diagram of a complex system from the behaviour of its components" (Faust and Raes, 2012). For microbial communities, the goal of network inference is to predict ecological relationships between microorganisms from abundance data, which can be problematic when based on eDNA generated community data. Nevertheless, co-occurrence and correlation patterns found in these datasets are increasingly used for the prediction of species interactions. It should be noted, however, that there are different methods for inferring the presence of interactions **but these are poorly studied and rarely applied to microbial networks**, where testing can be harder than in conventional, macro-scale networks (but see below).

Nevertheless, there have been some studies that have analysed microbial co-occurance networks generated from biofilm. Widder et al. (2014) combined co-occurrence analyses of biofilms based on next-generation sequencing with a probabilistic hydrological model, and showed how fragmentation of microbial co-occurrence networks (with a focus on Betweenness centrality and the random removal of nodes (similar to robustness) change across stream networks. But real-world studies such as this are few.

The EA's RSN microbial community data is well suited to **testing different network inference methods** as it is collected over spatial and temporal replicates, which is necessary to infer ecological networks. Specifically, inference methods based on MaxEnt (Volkov et al., 2009) and Matrix Autoregression (MAR) approaches (with either single (Hampton et al., 2013) or multiple delays (Barraquand et al., 2021)), supplemented with trait and phylogenies/taxonomic information (Ovaskainen et al., 2017), can be compared and validated. Faust and Reas (2012) suggest that, in addition to predicting links between taxa, the analysis of microbial association networks "reveals niches, points out keystone species and indicates alternative community configurations." But they also caveat that

testing and evaluation is needed to determine the best-performing network inference technique. To address this they point to the cultivation of unknown microorganisms, combinatorial labelling and parallel cultivation as methods that could allow systematic co-culturing and perturbation experiments, the latter lending itself to robustness analyses (described earlier) where species are removed.

Matchado et al., (2021) provide a review of state-of-the-art methods to infer intra-kingdom interactions in microbial communities derived from DNA/RNA-based methods, ranging from simple correlation- to complex conditional dependence-based methods.

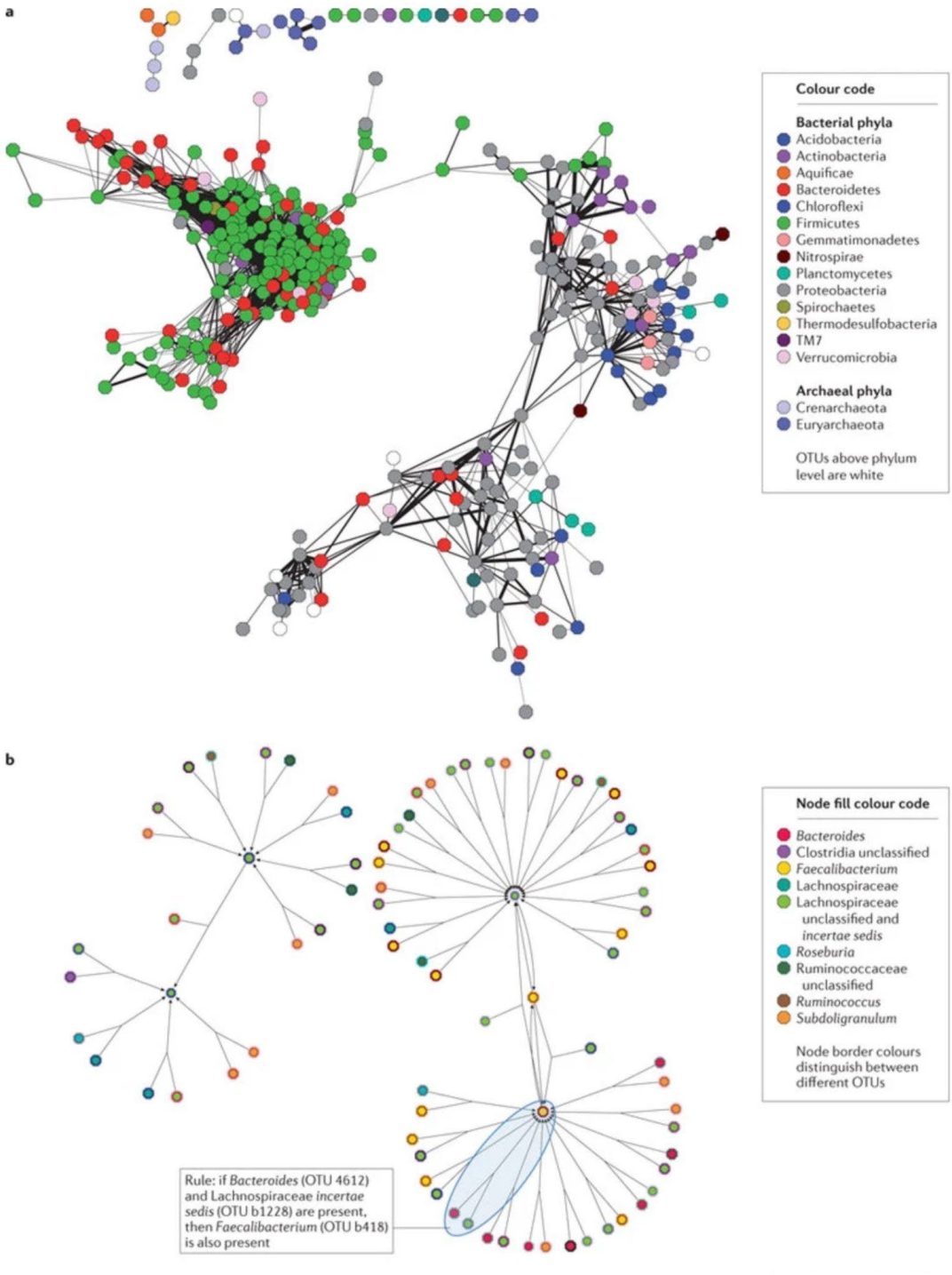
3.1.1 Caveats for microbial networks

Network inference techniques are frequently applied to microbial presence–absence or abundance data to detect significant patterns of co-presence and mutual exclusion between taxa and to represent them as a network. Barroso-Bergadà et al., (2021) recently examined the pros and cons of crop microbial networks inferred from eDNA data but found (i) very high variability of network replicates within each system; (ii) a low number of network replicates per system, due to the large number of samples needed to build each network; and (iii) difficulty in interpreting links of inferred networks. So caution is needed before embarking on network analyses of the RSN microbial networks.

3.2 Putting microbial data into multi-layer networks

Microbial community data generated from high throughput sequencing of biofilm provides the opportunity to move away from the study of bipartite networks to include multiple interaction types by merging bacteria, fungi, diatoms, protists and other eukaryotes into '**multilayer networks**' (see 1.5d). The emergent field of multilayer networks provides a natural framework for extending analyses of ecological systems to include multiple layers of complexity, as it specifically allows one to differentiate and model 'intralayer' and 'interlayer' connectivity (Pilosof et al., 2017).

Faust and Reas (2012) demonstrated how complex networks, generated through inference models, can result in a link in a network that connects more than two nodes in a directed way to point from the independent taxa to the dependent taxon (termed here as a directed hypergraph). Figure 3.1a displays a microbial network inferred from a similarity-based approach, in which pairwise relationships are represented by edges connecting two nodes, whereas Fig. 3.1b gives an example of a directed hypergraph that results from '**association rule mining**' in a global microbial presence–absence data set (see 3.3).



.

Figure 3.1. Pairwise (a) and complex relationships (b) were inferred from a global microbial operational taxonomic unit (OTU) presence–absence data set. a | Each node represents an OTU, and each edge represents a significant pairwise association between them. b | This network summarizes association rules mined with an a priori algorithm and filtered with multiple testing correction. The text box provides an example for such a rule. As the data set is extremely sparse, rules are restricted to positive associations involving up to three OTUs. Each node in network b represents an OTU, whereas each edge corresponds to a rule. In contrast to network a, an edge can connect three OTUs if they are all involved in the same rule. For ease of interpretation, the same OTU (with the same node fill and border colour) may occur multiple times in network b.

3.3 Using machine learning to construct networks

There is considerable interest in building whole ecological networks of interactions from data using statistical or logic-based machine learning. The idea behind these machine learning methods is that **embedded in a dataset is the imprint of the recent processes and interactions** that created the data and this information can be recovered to reconstruct networks. The underlying hypothesis of machine learning for network reconstruction is therefore that ecological interactions produce correlations and relational patterns in the abundance of species that can be recovered. In statistical machine learning, the variation in the sample is treated statistically, typically using Bayesian approaches (Jakuschkin et al., 2016; Vacher et al., 2016). Significant correlations between any given pair of species within the data are then considered as potential network links. Logic-based machine learning treats relational patterns rather like the structure of grammar in a language (Muggleton and de Raedt, 1994; Tamaddoni-Nezhad et al., 2006).

In both statistical and logic-based machine learning, the challenge is to sort, from the list of interactions hypothesised by the learning algorithms, those l**inks that are ecologically meaningful from those that are artefacts**. This selection approach is done differently in the two approaches. In logic-based machine learning, the grammar for an interaction can be coded as background information. In an agroecological network learnt by Bohan et al. (2011) and Tamaddoni-Nezhad et al. (2013), trophic interactions were selected by background information that was a set of grammar rules (a model) for a trophic interaction whereby the predator and prey species must co-occur in the same samples and predators should be larger than their prey. A trophic interaction between two species was only identified if this grammar rule was realised. In statistical approaches, links are selected using environmental factors or species functional trait covariates integrated into the modelling (Cazelles et al., 2016; Jakuschkin et al., 2016; Ovaskainen et al., 2016; Vacher et al., 2016).

Learning networks is currently limited by our background information rules. Mechanistic rules for trophic interactions, based upon body or gape size, allow the reconstruction of food webs, but challenges for microbial networks remain. Where ecological networks are structured by processes for which we have no general mechanistic explanation, there is no background information that can be employed, and machine learning is of little value for reconstructing networks. However, recent developments in logical machine learning are now allowing background information rules to be discovered from data. Tamaddoni- Nezhad et al. (2015) showed using simple subnetworks that the trophic interaction rule ‘big things eat small things’ can be recovered. Developments of this work are now extending this possibility of rule learning to larger and noisier data sets, and with applications for biomonitoring (Ghannam and Techtmann, 2021).

4. An overview of the network ecology research landscape and how this can be applied in the context of biomonitoring

For the past decade, ecologists have made compelling arguments that **reconciling biodiversity and ecosystems function in a single conceptual framework is best achieved through the application of a network approach** (Thompson et al., 2012). Whilst theoretical ecologists have considered universal patterns (e.g. Bascompte and Jordano, 2013), issues of scale (Guimarães, 2020), environmental change (Memmott et al., 2007) and co-evolution (Thompson, 2014), applied ecologists have advocated new network approaches for biomonitoring (Gray et al., 2014; Derocles, Bohan and Dumbrell, 2018; Raimundo, Guimarães and Evans, 2018). Indeed, Tylianakis (2010) proposed a number of network descriptors (e.g. connectance and nestedness) that can readily be incorporated into biomonitoring schemes.

Bohan et al. (2017) imagined a world where a global array of autonomous samplers, capable of in situ DNA-sequencing, would upload data to the cloud for network construction using machine learning (Fig 4.1). The overall aim would be an analysis across highly-resolved replicated networks (similar in scope to the RSN) to detect change in network structure across the sample array. Although ahead of its time, there has been rapid development and validation of eDNA for biomonitoring aquatic systems (Schenekar, 2023) and field-based sequencing is now happening, thanks to new platforms (i.e. Oxford Nanopore). However, **in-cloud network reconstruction methods are a long way behind** and, even if changes in network structures could be detected, it is unclear what this would actually mean in terms of ecosystem functioning and resilience (it is usually assumed that some sort of ecosystem degradation can be detected).

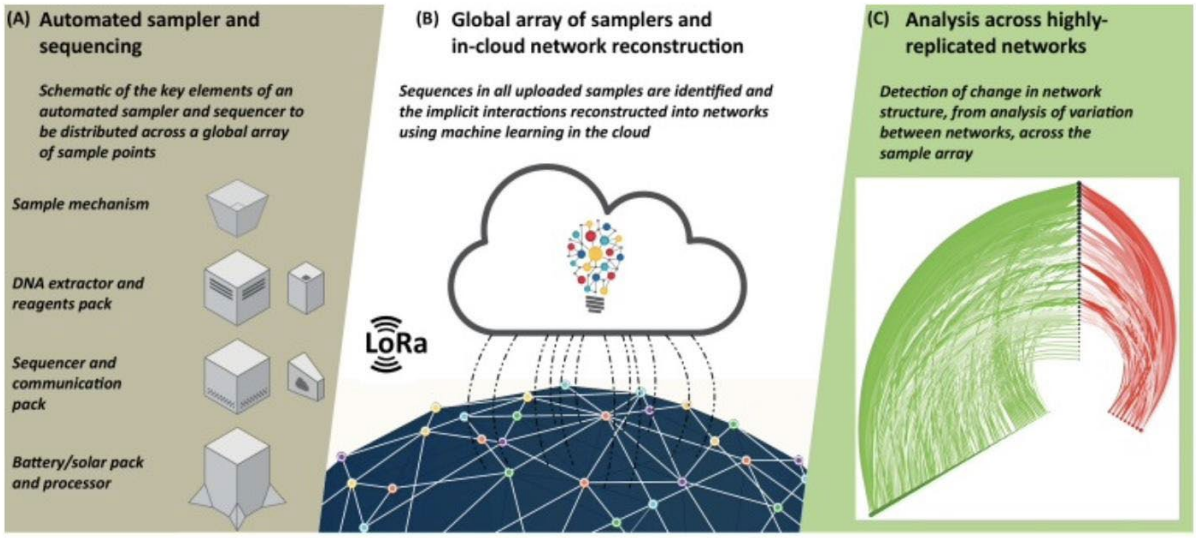


Figure 4.1. Large-Scale Biomonitoring using next-generation sequencing. (A) Schematic illustration of the key components of an autonomous sampler. (B) Diagram of an array of

sample points, each with a sampler, and the upload of sequence data to the cloud. Management, identification and reconstruction of network structure is done in the cloud. (C) Detection and analysis of change in the structure of the monitored networks. Source Bohan et al. (2017).

Nevertheless, there are some positive developments in this area. Derocles et al. (2018) provide a framework for measuring the robustness of multilayer networks (derived form next-generation sequencing) in the long term as one way of integrating ecological metrics more generally into biomonitoring schemes to better measure biodiversity and ecosystem functioning. An advantage using next-generation sequencing in this way is the ability to construct **phylogenetically-structured networks**, which allow the use of more sophisticated adaptive network models (see 5.3), that **allow both ecological and evolutionary questions to be investigated**. For example, DNA sequences have been used to explore phylogenetic signals in a network context (Elias, Fontaine and van Veen, 2013; Rafferty and Ives, 2013). This approach would certainly bring added value to biomonitoring programs as ecosystem condition and coevolutionary processes could be monitored together. Moreover, this approach is very well suited to the study of microbial networks, as phylogenetic placements of OTUs (in this case undescribed microbes) can be determined using the sequencing data (see Czech et al., 2022 for a review).

In summary, the application of network metrics for biomonitoring has been advocated for over a decade, but has lagged behind the rapid development of bulk-sample metabarcoding and eDNA methods. Whilst the latter is revolutionary, in its basic form it still only provides species inventories and does not take full advantage of the molecular data generated. With testing and validation, this information can be used to construct large-scale, multi-layer ecological networks, which can be used to monitor both ecological processes and ecosystem resilience.

5. Applications of network theory to study the resilience of interacting communities to perturbations and species loss

5.1 Robustness

Conceptually, we have already seen how the extinction dynamics of an ecosystem can be studied using network robustness measures, and how this has been applied to the study of plant-microbe interactions (Bennett, Evans and Powell, 2019). Studies have progressed from simple qualitative, bipartite mutualistic networks, to investigations of patterns across ecosystems. There are also more sophisticated ways to study secondary extinctions, for example using Bayesian network approaches (Eklöf, Tang and Allesina, 2013) and/or taking into account interaction strengths (Bane, Pocock and James, 2018). Eng and Borenstein (2018) examined taxa-function robustness in microbial communities and suggested that community functional robustness to taxonomic perturbations could vary

widely across communities with different compositions, and concluded that a systematic study of the inherent link between community composition and robustness is lacking.

Pocock et al. (2012) constructed and analysed the first 'network of ecological networks' (i.e. 11 groups of animals interacting with shared plants on farmland), providing **new analytical tools** for understanding both the consequences of species extinctions across multiple animal groups, as well as the potential for ecological restoration. The Norwood Farm study provided a method to calculate the relative **importance** of plants (and habitats, Evans et al. (2013)), and thus identified **key species and hubs** that were disproportionately important in the network of networks. Although yet to be tested empirically, one application of this approach is that important plants could be **targets for conservation and restoration** that would benefit multiple animal groups. By examining the robustness of the joined networks, the study found that animal groups varied in their robustness to sequences of plant extinction, with the plant–pollinator network exhibiting much lower robustness than the seed-feeding bird network. Therefore, using a network approach, the authors argue that it should be possible to use robustness analysis to identify more **sensitive groups for targeted conservation effort** and/or assessment for biomonitoring rather than spending limited funds on charismatic species.

5.2 Socio-ecological networks

I have already shown how multilayer networks can link multiple groups of taxa for analyses. However, such networks can incorporate a diverse range of data layers, including information about people and ecosystem services relevant to policy makers and resource managers. For example, Keyes et al. (2021) compared the robustness of salt marsh food webs and the ecosystem services. Simulating twelve extinction scenarios for estuarine food webs with seven services, they find that food web and service robustness are highly correlated, but that robustness varies across services depending on their trophic level and redundancy. They used robustness to identify species that provide ecosystem services but do not play a critical role in stabilising food webs – whereas they found species playing supporting roles in ecosystem services through interactions are critical to the robustness of both food webs and services. Thus, there is potential for creating socio-ecological networks from microbial networks if the services provided by the taxa within them can be determined. Thus, there is considerable scope for work in this area as part of the RSN.

5.3 Adaptive networks

Adaptive network models allow us to better understand and predict how both **ecological and evolutionary processes** shape biodiversity and ecosystem functioning. In adaptive networks, the feedback between the macroscopic dynamics of interaction structure and the microscopic dynamics of population-level processes shapes interactions, abundances, and traits, hence influencing resilience and functional diversity.

According to Raimundo et al. (2018), the increasing availability of **phylogenetically-structured network** data generated through next-generation sequencing techniques (see 4), alongside the standardisation of biomonitoring protocols, can integrate evolutionary principles into adaptive network models for biomonitoring and/or ecological restoration, providing highly-resolved information for model parameterization and assessment across temporal and spatial scales (Fig. 5.1)

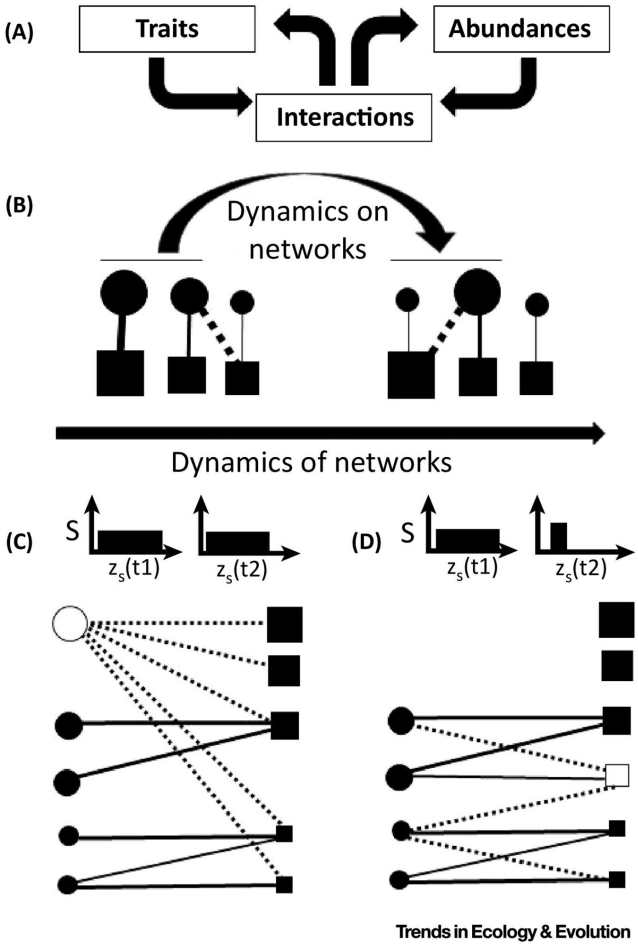


Figure 5.1. Adaptive networks capture the interplay of processes occurring at different levels of biological organization and can be used to predict how restoration practices will affect community-level properties. Source: Raimundo et al. (2018).

Adaptive Network Models (ANMs) account for the feedback loop between: (i) the dynamics of networks, which refers to temporal variation in the network structure due to **interaction rewiring**; and (ii) the dynamics on networks, which refers to changes in population-level properties of the species that form the network, such as mean traits and abundances (Figure 5.1B). ANMs can explore the relative roles of candidate mechanisms that produce biodiversity patterns, such as neutral and niche-based processes which can influence patterns of interaction among species (Vázquez et al., 2009). They can also provide testable predictions for changes in biodiversity arising from management practices that add or remove species from communities and refer to: (i) network structure (which could be used as a proxy for resilience); (ii) the distribution of species abundances, which relates to stability (Allesina and Tang, 2015); and (iii) the community-level distribution of traits, which relates to both robustness and functional diversity (e.g. Pillar et al., 2013).

Unlike the Norwood Farm example, which is essentially a **static** snap-shot of interacting species in time, ANMs allow an understanding of interaction wiring, i.e. the reconfiguration of an ecological network arising from the establishment or cessation of pairwise

interactions as a consequence of **adaptive or stochastic processes**. The incorporation of realistic rewiring mechanisms (Kaiser-Bunbury et al., 2010; Ramos-Jiliberto et al., 2012) into ANMs can help to **predict to what extent an ecosystem will be able to absorb perturbations** by the reconfiguration of its interaction patterns without changes to ecosystem functioning. A roadmap for operationalizing this approach for restoration and biomonitoring purposes is available, but has only been considered in the context of conventional plant-animal species-interaction networks, and not specifically for microbial networks (see Raimundo, Guimarães and Evans, 2018).

5.4 Complexity-stability modelling

Understanding what happens to ecological networks when they lose a significant fraction of species is essential for assessing, and potentially mediating, the current impacts of biodiversity loss. Traditional theoretical approaches to ecosystem stability, well-known from the complexity–stability debate (May, 1972; Allesina and Tang, 2015; Donohue et al., 2016; Landi et al., 2018), are unable to answer such questions as they concern **the response of populations to small perturbations**. Alternatives have been developed, including network robustness approaches described above, and relevant population-dynamics simulations that have typically considered the effect of the loss of a single species (Pimm, 1984). Such simulations have highlighted factors such as complexity (Borrvall, Ebenman and Tomas Jonsson, 2000; Ebenman, Law and Borrvall, 2004; Lundberg, Ranta and Kaitala, 2008; Kaneryd et al., 2012), trophic level (Quince, Higgs and McKane, 2005; Borrvall and Ebenman, 2006), interaction distribution (Fowler, 2010; Kadoya and McCann, 2015) and network structure (Thebault, Huber and Loreau, 2007) as playing a role in the response of the ecosystem to species loss. The results, however, are sometimes conflicting (Thebault, Huber and Loreau, 2007; Sanders et al., 2018). Moreover, **simulation approaches are numerical and typically restricted to small model ecosystems**, both of which make it difficult to derive firm conclusions connecting network features and the vulnerability of species to secondary extinction (Rohr, Saavedra and Bascompte, 2014).

Emary and Evans (2021) overcame many of these issues by describing a random-matrix theory of ecological species loss. Their model assumes a large network of interacting species with dynamics described by a generalised Lotka–Volterra model with random coefficients. They assume that an ecosystem starts in equilibrium and is then subject to the **rapid loss of a significant fraction of the initial species** (as might occur in e.g. a large pollution event). Using an adaptation of the dynamical cavity method, they obtained exact results for the distribution of species abundances following this primary extinction event. From this, they describe how **the severity of secondary extinctions depends not only on the size of the primary extinction, but also on the nature of inter-species interactions**. They found that the capacity of an ecosystem to survive the loss of a significant fraction of its species depends on the relative importance of **competitive, mutualistic and antagonistic interactions.**

One of the main predictions of this work is that, for small primary extinctions when ‘only’ a few species have gone extinct, **the ecosystem responds by changing its equilibrium**

abundances to the new values that are normally distributed around the old values, without secondary extinctions taking place. Although still theoretical, empirical observation of this effect would not only lend some support, but the size and direction of the response would allow outstanding interaction parameters to be inferred. Once in possession of these parameters, **predictions** could be made concerning the response to large-scale primary extinctions. This could potentially provide **early warning** of the effects of such disturbances and facilitate the identification of the most vulnerable ecosystems, which would allow management and conservation efforts to be targeted.

5.5 How and why network ecology has potential for examining the resilience of freshwater systems

Drawing the network construction and analysis components together, I have shown that freshwater microbial community data generated using high throughput sequencing (including metabarcoding data for bacteria, fungi, diatoms and other photosynthetic microbes, protists and other eukaryotes) can be merged to construct multilayer networks, based on network inference methods (with further testing and validation work required). Assuming competitive, mutualistic, antagonistic and other interactions can be determined (for example using machine learning), the resilience of these networks to a range of perturbations can be examined using the theoretical advances described above - although they are yet to be tested on empirical data. Clearly more work is needed to bridge the gaps between theoretical and empirical network ecology, but there is certainly **scope for developing freshwater 'vulnerability' indices** using adaptive network and random-matrix theory.

6. Recommendations to explore multi-taxa microbial datasets, with a particular focus on multi-layer networks

Biofilms form complex networks of interactions: a surprising level of multi-cellular behaviour and extensive three-dimensional structures act in concert to create characteristic taxonomic and functional diversity (Besemer, 2015). Yet the causes and consequences of biofilm biodiversity remain insufficiently understood. Ecological networks provide a framework by which the functioning of aquatic environments and the ecosystem services they provide can be studied.

One of the main constraints to operationalizing adaptive network ecology for biomonitoring, restoration and management is the lack of spatial and temporal data (Raimundo, Guimarães and Evans, 2018). This is necessary for model parameterisation, determining rewiring rules and testing theoretical predictions. The DNA dataset from river biofilm samples collected over a three-year period as part of the EA's newly designed River Surveillance Network (RSN) programme provides **an unprecedented opportunity to apply spatio-temporal data to a number of fundamental questions in network ecology**, from methodological validation through to biomonitoring and long-term

forecasting. Below I recommend a number of key questions and/or recommendations that could be beneficial for freshwater biomonitoring.

6.1 Linking ecological network structure to measures of water quality

The RSN collects both biotic and abiotic data. This provides an unprecedented opportunity to explore relationships between network structure and complexity (both merged and non-merged, e.g. diatom networks), abiotic factors and ultimately measures of water quality, thus **linking biodiversity and ecosystem functioning** in a novel way.

6.2 Overcoming primer, DNA extraction, sequencing and bioinformatic biases

Many studies of soil microbial communities tend to analyse bacteria and fungi separately due to known amplification biases. **A concerted focus on biofilm samples is recommended, as biases in organism detection and abundance will undoubtedly affect network construction methods and analyses**. Repeat samples in space and time will allow for sampling completeness to be tested (for example through standard species accumulation curves), and with similar approaches available for ecological networks (e.g. Macgregor, Evans and Pocock, In Press). Around 80% sampling completeness is the general rule of thumb before cross-network comparisons can be made. However, the systematic structure of the RSN can still be used to explore relationships in space and time, as any biases will be standard across the network. Quantification of microbe abundances will remain a challenge.

6.3 Multilayer networks created using muti-taxa microbial datasets

Microbial networks generated from next-generation sequencing will, in their most basic form, be undirected and unweighted. While examination of network complexity and robustness is possible, I have already shown how the capacity of an ecosystem to survive the loss of a significant fraction of its species depends on the relative importance of competitive, mutualistic and antagonistic interactions. Thus **a key challenge is to determine the direction and strength of relationships between microbes**. While some of this information can be brought a priori into network models (e.g. for well studied groups of diatoms), a programme of work (both computer and lab-based) is required to better understand such interactions. The problem becomes manifest when multi-taxa microbial datasets are merged into multilayer networks, where the number of potential direct and indirect interactions can grow rapidly. However, state-of-the-art mathematical and machine learning approaches are helping to resolve such issues, although to my knowledge) not specifically in freshwater microbial systems.

6.4 Identifying key hubs

Multilayer network analysis should be able to **identify key taxa, or groups of interacting organisms (hubs or modules), within the networks** that could be involved in key ecological processes. Thus a preliminary analysis of the RSN data should be seen as hypothesis forming, allowing more targeted microbial research to confirm or otherwise some of the emerging properties associated with ecosystem functioning. If this is indeed the case, then these could be candidates for future biomonitoring work using next-generation sequencing, as opposed to generating entire networks, which may continue to be costly (although sequencing costs are scalable). Or the latter might be preferable for the early detection of pests and pathogens as well as a more holistic understanding of freshwater microbial ecology.

6.5 Measuring the vulnerability of freshwater ecosystems - new metrics

I have shown how examining the extinction dynamics of ecological networks provides new measures of ecosystem resilience to perturbation. With work, **it is not inconceivable that the vulnerability of a river ecosystem could be assessed** by taking a water sample, extracting DNA and RNA, and using sequencing data to construct networks that can then be perturbed using the approaches described to provide an overall resilience metric. **One immediate way to operationalise and test this is to use the RSN data from samples taken up- and down-stream from sewage outlets.** The resilience of freshwater microbial networks to this perturbation event can be compared. Furthermore, I have already described how advancing network approaches for biomonitoring is hampered by the lack of spatially and temporally resolved data required for model parameterisation. By using data generated from ~250 sites (500 samples), there is an unprecedented opportunity to test and validate models empirically.

6.6 Big data and processing power

A final consideration is the storage and sharing of microbial switch to this approach, especially if the RSN will continue to analyse biofilm samples into the future. Running robustness analyses (e.g. based on 10,000 random simulation events) takes large computer processing power, even more so for multilayer networks. Here there is an opportunity to construct multilayer networks for ~250 sites, which is unprecedented, and simultaneous extinction modelling will require access to High Power Computers.

Acknowledgements

I thank members of the Network Ecology Group at Newcastle University for helpful early discussion on the topic, including potential ways of generating abundance data from microbial samples. I drew on a draft chapter of an unpublished book co-authored with

Michael Pocock, UKCEH, when explaining the basics of network ecology at the beginning of the Think Piece.

Think piece 3: Next Generation Biomonitoring of Microbial Communities: Integrating High-Throughput Sequencing with Ecological Network Analysis

Alex J. Dumbrell, Nick Aldred and Hannah M. Lewis[1]

[1]School of Life Sciences, **University of Essex**
Wivenhoe Park, Colchester CO4 3SQ

Summary

Ecosystems globally are under increasing pressures from multiple and often interacting stressors, ranging from inputs of novel chemicals to the persistent pressures associated with climate change. The need to understand the ecological effects of these stressors, how they shape the biodiversity and functionality of ecosystems, and what management solutions may mitigate against negative effects, has generated renewed interest in environmental monitoring. To date, most environmental biomonitoring programs have calculated diversity indices/metrics from a targeted group of indicator taxa, relying on phenomenological pattern fitting, with little mechanistic insight or predictive capacity. Moreover, their narrow taxonomic focus, excludes capturing the full diversity of the functionally most important components of natural ecosystems, the microbes. Thus, there is need for a new biomonitoring approach able to capture the diversity and functions of microbial communities, and provide a mechanistic, predictive framework linking this to ecosystem health, stability, and functionality.

Microbial communities underpin all biogeochemical cycles, including major transformations of carbon and nitrogen, as well as providing the base of most food webs. Thus changes in the diversity, composition and ecological interactions within these communities have direct links to ecosystem functioning. The development of molecular-based methods for profiling microbial communities, notably next generation (high-throughput or second generation) DNA/RNA sequencing (NGS) has provided the tools required to survey the diversity and composition of microbial communities, and certain functional attributes (e.g. genes associated with carbon and nitrogen transformations), but it does not directly quantify ecological interactions. Yet, it is this network of ecological interactions that ultimately determines the stability of ecological communities and their capacity to deliver ecosystem functions. To address this, various computational methods have been proposed to infer microbial ecological networks, which describe aspects of ecological interactions, from NGS data.

Microbial ecological network analysis uses associations within spatial or temporal species count data generated from NGS surveys to infer ecological interactions between constituent microbial species. Multiple algorithms exist to achieve this that range in their

computational complexity, applicability to different datasets, ease of use, sample size requirements, precision and sensitivity. Some of these algorithms simply describe which microbial species are more or less likely to co-occur, while others are able to resolve the form (e.g. positive, negative, and neutral) and magnitude of interactions, with many algorithms delivering outputs between these two extremes. Once microbial ecological networks have been inferred, metrics of these networks (e.g. number of nodes, edges, betweenness centrality, connectivity etc) can be calculated and used as a proxy for various ecosystem properties. Multiple studies now exist that correlate properties of microbial ecological networks with ecosystem attributes, processes and functions. However, a lack of consistency in both findings, and approaches, including naming definitions of network properties, exists across studies, making generalisations limited.

While microbial ecological network analysis offers a promising approach for use in contemporary biomonitoring programs, it is not without limitations. Multiple methodological considerations remain without satisfactory solutions; for example, best practice approaches for data pre-processing, dealing with rarer taxa and environmental factors, capturing higher order interactions, and understanding what various network metrics and emergent properties mean in an ecological context. Perhaps the biggest limitation is that microbial ecological network analysis captures association networks, and doesn't necessarily represent underlying biological interactions. Currently our understanding of ecological interactions within microbial communities is limited, and it remains relatively unresolved as to exactly what ecological interactions (and underlying community and population processes) are represented by inferred interactions from microbial ecological network analysis. This leaves an important missing link between mechanisms grounded in ecological theory that explain how species interactions support the stability and functioning of ecosystems, and how this relates to inferred microbial ecological networks.

New biomonitoring programs (e.g. River Surveillance Network (RSN) programme) need to carefully consider the limits of microbial ecological network analysis within the context of the data they collect. Users need to remain abreast of methodological developments within this rapidly evolving field of research, and be aware that as biomonitoring programs continue, a shift from network inferences based on spatial (cross-sectional) data to those based on time-series (longitudinal) data may yield more robust insights. Moreover, biomonitoring programs should not rely solely on microbial ecological network analysis to assess the health, stability and functioning of ecosystems. New developments in related analyses, which use functional and trait based approaches, or those that rely on functional as opposed to phylogenetic marker genes, can complement microbial ecological network analysis to provide a more holistic understanding of ecosystems.

1. Introduction

1.1 Integrating Microbes into Next Generation Biomonitoring

The integration of DNA (or eDNA) metabarcoding (and metagenomic) data with ecological network analysis is increasingly viewed as a leading contender to provide a Next Generation Biomonitoring (NGB) solution to track, understand, and predict the health of ecosystems (Bohan et al. 2017; Derocles et al. 2018; Dubart et al. 2021). At its core, NGB is deceptively simple. Nucleic acids (commonly DNA, but can also be RNA) are extracted from environmental samples, and either phylogenetic marker genes (e.g. 16S, 18S, COI, etc) are amplified and sequenced (metabarcoding), or the entire extract is shotgun sequenced (metagenomics). The metabarcoding or metagenomic data are then used to reconstruct ecological networks, and the properties of these networks are used to make inferences about the health, stability and functionality (among other properties) of the ecosystem. This broad NGB approach has only recently (over the last ~10yrs) become a viable option for monitoring ecosystems due to the simultaneous maturation of two separate methodologies (Bohan et al. 2017): (1) massively parallel (next or second generation) sequencing (NGS) to provide the depth and coverage of require DNA/RNA sequence data, and (2) ecological network analysis, specifically developments in machine learning, statistical inference, and inductive logic approaches to infer the ecological networks. Consequently, NGB provides two (among many other) significant potential benefits to biomonitoring systems; it relies on a single common unit of measure to capture biodiversity (DNA/RNA sequence data), and returns an ecological network analysis with properties that reflect multiple levels of biological organisation, spanning populations, communities, and ecosystem stability and functionality as well as the potential to better predict the status of ecosystem services delivery.

Focusing on quantifying biodiversity via DNA and/or RNA based methods has clear benefits over traditional morphology based approaches with regards to time, expertise, consistency and comparability (both across taxa and sites) of the data (Cordier et al. 2020). Importantly, and in the context of this think piece, it also provides a method to capture microbial diversity (Clark et al. 2018). Since the advent of modern molecular microbial ecology, microbiologists have relied on DNA/RNA based methods to quantify microbial diversity and overcome the inherent limitations of culture-based techniques that often struggle to capture the majority of microbes present in a sample (Clark et al. 2018). It is now the rule and not the exception that environmental microbial ecology researchers use 16S, 18S, ITS (or other phylogenic marker gene) metabarcoding to quantify microbial diversity, with examples covering all ecosystems (e.g. Cordier et al. 2022; Thompson et al. 2017; Tedersoo et al. 2014). Microbes are ubiquitous, with >1030 microbes (Whitman et al. 1998) from >1012 different species (Locey & Lennon 2016), and comprising >109 tons of carbon (Kallmeyer et al. 2012) globally. They support a range of ecosystem processes and functions, alongside driving all major biogeochemical cycles, and thus, having a disproportionately large impact on ecosystem services (Falkowski et al. 2008). Yet, microbial community analysis have generally been excluded from routine biomonitoring, despite the tools now being readily available to do this and the role of microbial

communities in ecological food webs having been highlighted long before the advent of next generation sequencing (e.g. Woodward et al. 2005). Moreover, microbial populations and communities can act as sentinels to environmental change, responding rapidly to novel abiotic stressors (e.g. Thompson et al. 2016) and also changes in the structure of the wider food web (e.g. Ferguson et al. 2021).

Effective environmental biomonitoring programs capture information about the state, health and functionality of ecosystems, and the progress of any management practices applied to these systems. The ability of DNA and/or RNA based methods to sample local species diversity across all domains of life provides an ideal foundation to achieve this. This is because a positive relationship between local diversity and ecosystem functions – the processes controlling the flows of energy, nutrients and organic matter through the environment – exists across ecosystems and has been confirmed by multiple studies (Cardinale et al. 2012). For example, multiple plant biodiversity experiments have shown a positive asymptotic response of community productivity to increasing plant species richness (Hector et al. 1999; Cardinale et al. 2007), and similar relationships have been observed in microbial communities (Bell et al. 2005). Furthermore, biodiversity-ecosystem functioning (BEF) relationships have been observed across multiple ecosystem functions, and it is apparent that within a multifunctionality context (i.e. considering all functions within an ecosystem), that different species contribute differentially to different functions, and that interactions between species further supports ecosystem multifunctionality (Hector & Bagchi 2007; Slade et al.2017). Similarly, the relationship between biodiversity and stability may be viewed in a multistability context. Ecological stability is inherently multidimensional, comprising, for example, asymptotic stability, variability, persistence, resistance and resilience (Pimm 1984). As with ecosystem functions, different species may contribute differentially to supporting different aspects of stability (Donohue et al. 2013). Thus from a biomonitoring perspective, a single metric of alpha-diversity (e.g. species richness) is a poor predictor across multiple functions and stabilities, as it ignores species-specific contributions, and the ecological interactions that promote multifunctionality and stability. A framework that explicitly considers the network of ecological interactions within natural ecosystems, and the functional contributions of different species members, is a more desirable biomonitoring option.

However, despite the huge potential for NGB to provide the global solution to the routine monitoring of the health, stability and functionality of ecosystems there are still many nuances and considerations that require exploration. This is particularly true if we focus solely on applying the NGB philosophy to the study of microbial biodiversity. In this think piece, we will explore combining microbial DNA-based metabarcoding and shotgun metagenomics with ecological network analysis to fully integrate this into NGB approaches. We will outline (1) the available methods for reconstructing microbial ecological networks before (2) exploring the metrics used to examine network properties and what these may reveal about microbial communities. Building on this, we (3) examine examples of where network analysis has been applied, and what (if anything) it has revealed about stressor responses and higher level ecosystem properties, such as ecological stability and functionality. Next, we will explore (4) the current methodological limitations to microbial ecological network analysis, including inherent limitations in the

underlying molecular data, sampling scales (spatial and temporal) and sampling designs, and what can robustly be inferred from microbial ecological network analysis. Complementing this, we will examine (5) the most significant knowledge gap that needs addressing to make microbial ecological network analysis a routine component of NGB. We will examine (6) complementary analyses to microbial ecological network analysis, and what combination of network metrics and non-network analyses based on the same dataset may be most appropriate for routine biomonitoring. Finally, we will (7) focus on the biofilm samples collected over a three-year period as part of the Environment Agency's newly designed River Surveillance Network (RSN) programme, and the types of analyses most suited to that specific dataset.

2. Building Microbial Ecological Networks

2.1 Microbial Ecological Network Inference

The majority of well resolved ecological networks (especially food webs) have been constructed for macro-organisms (e.g. Woodward et al. 2005). These typically use direct observations of biotic interactions to determine links between species (nodes) in the network, although for ecosystems with poorly defined interactions (e.g. agricultural systems), logic-based machine learning methods have managed to reconstruct ecological networks as accurately as methods based on direct observations of biotic interactions (Tamaddoni-Nezhad et al. 2013). These macro-organism ecological networks range from describing food webs, host-parasitoid webs and mutualistic webs (among others) and importantly are underpinned by a substantial body of supporting theory that explains what is happening in these networks, and thus what network properties mean in an ecological context (Ings et al. 2009). For example, within food webs there is a clear understanding of what a predator-prey interaction is, and supporting theory that describes how changes in its interaction strength may alter networks properties, such as the role weak interactions play in stabilizing webs (Kadoya & McCann 2015). Therefore, within a biomonitoring context, changes in the properties of macro-organism ecological networks in response to environmental perturbations, has a direct, known, and understandable connection to the underlying biology of the system. However, in contrast, microbial ecological networks are almost exclusively inferred, with no direct observations of biotic interactions included in their construction, and in general, construction is entirely based on patterns of species co-occurrence in DNA/RNA metabarcoding/metagenomic data (Faust 2021). Consequently, a substantial body of research into developing the most robust microbial ecological network inference algorithms exist, with multiple new algorithms published yearly (see Vacher et al. 2016; Cappellato et al. 2021; Matchado et al. 2021 for reviews).

The overall aim of inferring microbial ecological networks from species co-occurrence data, is to identify taxa (species, Operational Taxonomic Units (OTUs) or Amplicon Sequence Variants (ASVs)) that co-occur in a non-random manner and are thus more likely to be interacting with each other while avoiding erroneously assigning interactions

between species that are independently responding to the same local environmental conditions but do not interact. Within the inferred microbial ecological networks, nodes represent species, and the edges (links between nodes) represent the potential interactions inferred from the co-occurrence data. Typically, these observed microbial ecological networks are compared against null model networks inferred from randomly generated data and this is used to evaluate the strength and significance of inferred interactions and evaluate overall statistical confidence (Cardona et al. 2016). Building on this the strength of the pairwise interactions can be evaluated (often denoted by the thickness of the edge) and more recent algorithms also try to evaluate the direction of these interactions (e.g. positive, negative, neutral etc.). Once microbial ecological networks have been constructed, multiple network properties can be examined and used to evaluate ecological properties of the system being studied (see below and Liu et al. 2021a). There are multiple methods available for inferring microbial ecological networks, ranging from those based on correlation/regression approaches, graphical model inference, Bayesian and other statistical inference approaches, to logic-based and Machine Learning algorithms (Cardona et al. 2016; Vacher et al. 2016).

2.2 Correlation or Regression Methods for Network Inference

Correlation/regression (association-based) approaches are one of the most straightforward and popular approaches for inferring microbial ecological networks (Cardona et al. 2016), and include tools such as: **ReBoot** (Faust et al. 2012); **CCREPE** (The Huttenhower Lab); **WGCNA** (Langfelder & Horvath 2008); **SparCC** (Friedman & Alm 2012); **REBACCA** (Ban et al. 2015); **CCLasso** (Fang et al. 2015); **CoNet** (Faust et al. 2012); **Meta-Network** (Yang et al. 2019); **CCN** (Yang et al. 2020); and **MENAP** (Deng et al. 2012). The range of different correlation/regression-based approaches represent developments in network science and specifically various methods and approaches that attempt to overcome some of the major limitations of inferring microbial ecological networks (see **Limitations to Microbial Ecological Networks Analysis**).

Briefly, these methods first identify pairwise associations within the species co-occurrence data, using methods like Spearman or Pearson correlation coefficients, multiple linear regressions (typically sparse multiple regression to avoid issues of overfitting; Cardona et al. 2016), and quantitative (e.g. Bray Curtis) or qualitative (e.g. Jaccard's) similarity indices, or they use an ensemble of these different methods (e.g CoNet; Faust et al. 2012; CCREPE; The Huttenhower Lab), or other variations on this theme such as association rule mining (Meta-Network; Yang et al. 2019). Typically, associations are then compared to a null model, which provides the probability distribution of the association metric (e.g. Spearman correlations coefficient, or Bray Curtis similarity) produced via a randomisation method. The simplest method is often to randomly shuffle the rows of the OTU/ASV input data and calculate association metric for a given number of permutations. P-values are calculated as the probability the observed association metric between two taxa (OTUs/ASV) is greater than the one calculated by a chance resampling of the initial data. Significant associations are commonly defined as those where $P < 0.05$ (Weiss et al. 2016), and various corrections

such as Bonferroni or False Discovery Rate, are used to overcome issues of multiple testing. Modifications to this class of network inference methods, includes overcoming issues with non-independence within compositional data (see **Limitations to Microbial Ecological Networks Analysis**), such as permutation and bootstrap methods that resample the OTU/ASV dataset (e.g. ReBoot; Faust et al. 2012; CCREPE The Huttenhower Lab).

Of all these methods, SparCC (Friedman & Alm 2012) has gained notable traction, popularity and widespread use, this provides the benefit of multiple other studies using a consistent method against which to contextualise findings. SparCC was an early method to tackle issues with the compositional nature of microbial NGS datasets (see **Limitations to Microbial Ecological Networks Analysis**), whereby an increase in the absolute abundance of a single taxa can cause decreases in the relative abundance of all other taxa in the absence of any changes in their absolute abundance. SparCC uses an iterative approximation approach and log-ratio transformed data on which to base pairwise associations (Friedman & Alm 2012), with more recent methods building on this and claiming improved accuracy (REBACCA; Ban et al. 2015).

Another method attempting to overcome various methodological limitations is to construct molecular ecological networks (MENs) using random matrix theory (RMT). Here, interaction thresholds are detected automatically from data rather than being set arbitrarily or inferred from biological information (which is unlikely to be available for the majority of microbes) (Deng et al. 2012). RMT was developed in the area of theoretical physics and shown to be applicable to many biological networks; for example, protein interaction networks, metabolic networks in yeast (Luo et al. 2006) and gene co-expression networks (Luo et al. 2007). It has been applied to microbial ecological networks using both functional (creating fMENs; Zhou et al. 2010) and phylogenetic (creating pMENs; Zhou et al. 2011) marker genes. The aim of the RMT is to create the adjacency matrix for the system. This is a matrix of all pairwise interactions of all species/functional genes included in the analysis. From a phylogenetic perspective, the network inferred from the adjacency matrix is the map of which species interact, and if a weighted network is used, how strongly the species interact. OTU/ASV count data are standardised for each species across samples and pairwise Pearson correlation coefficients are calculated to measure the abundance-similarity between species. The absolute values of these correlation coefficients are the basis for the adjacency matrix, bound between zero if that value is below the threshold, and the absolute value of the correlation coefficient otherwise (Deng et al. 2012). The appropriate threshold is determined using an iterative process where the nearest neighbour spacing distribution (NNSD) of eigen values of the adjacency matrix switches from Gaussian orthogonal ensemble to Poisson (Deng et al. 2012). At each iteration, rows and columns of the matrix are removed, if all of their interactions fall to zero after the threshold is adjusted in the search, then the new eigen values and their new NNSD is then calculated (Deng et al. 2012). Freely available pipelines (MENAP) have been created to carry out these processes (Deng et al. 2012).

2.3 Influence of Data Properties on Correlation or Regression Based Network Inference

It is worth noting that none of the aforementioned methods are perfect, and of those that have previously been directly compared a wide range of precision and sensitivity is observed (Weiss et al. 2016). Importantly, it appears that different methods are better suited (in terms of sensitivity and precision) to dealing with data with different attributes. For example, SparCC performs particularly well on data with high compositionality properties, whereas an ensemble approach (e.g. combining CoNET, SparCC, Pearson and Spearman) is preferred when dealing with data that are <50% sparse (Weiss et al. 2016). Generally speaking, the correlation/regression (association-based) class of network inference methods, have decent false-positive rates (but in cases where very low false positives are required an ensemble of CoNet and Pearson works best), but all benefit from increasing the critical value of association metrics from $P < 0.05$ to $P < 0.001$ (Weiss et al. 2016). When tested on modelled data, designed to include a range of interactions (amensalism, commensalism, competition, mutualism, parasitic, syntrophic), most methods detect interactions that were from mutualism and commensalism, but those originating from amensalism and syntrophy where never detected (Weiss et al. 2016). This raises a number of questions, including: are there methods to infer edges within microbial networks that arise from all forms of biological interactions, can we robustly estimate what form that interaction originally took in the environment (e.g. competition vs predation), and do interactions whose definitions largely arise from observations of macro-organisms apply to microbial communities in the first place?

2.4 Overcoming Limitations of Correlation or Regression Based Network Inference

Correlation/regression (association-based) approaches often still fail to differentiate between direct and indirect associations (Matchado et al. 2021) and many still don’t fully account for issues of non-independence within the OTU/ASV co-occurrence data (compositionality) and the far greater number of OTUs than samples in most microbial surveys (Cardona et al. 2016). A number of different approaches (including, but not limited to, graphical model inferences, Bayesian/statistical inference, and logic-based and ML algorithms) have been developed to overcome some or all of these limitations. However, it is worth noting that many of these more sophisticated algorithms come with considerable computational costs. Thus a trade off between analytical speed, computational power, and resources (both computational and human expertise), and the precision, sensitivity and rigour of the analysis is likely to exist for the time being. This may become a core consideration in a NGB context, if a decent first-order approximation of the microbial ecological networks can be generated quickly and be used to detect rapid or transient changes in the environment rather than waiting longer for the inferred network to be better resolved.

2.5 Graphical Methods for Network Inference

As with correlation/regression (association-based) approaches, multiple methods based on graphical model inference exist, including: **gCoda** (Fang et al. 2017); **MDiNE** (McGregor et al. 2020); **MixMPLN** (Tavakoli & Yooseph 2019); **MPLasso** (Lo & Marculescu 2017a); **SPRING** (Yoon et al. 2019); and **SPIEC-EASI** (Kurtz et al. 2015). Arguably the most well known is SPIEC-EASI, and similar to SparCC has provided multiple studies against which to contextualise findings. Also, many of the other methods have been developed from SPIEC-EASI, and now offer increased speed and performance (e.g. gCoda; Fang et al. 2017; and to some extent SPRING Yoon et al. 2019). There are also tools to implement many of these in combination with correlation/regression (association-based) approaches via user friendly packages (e.g. NetCoMi; Peschel et al. 2021, microeco; Liu et al. 2021b), making them an appealing option. Broadly, these graphical model inference methods take an approach that begins with the consideration of conditional independence between species (OTUs/ASV) in the network. Conditional independence suggests that for two species they are conditionally independent when the abundance value of one species (relative to all species withing the dataset) does not provide any information on the probability of the other species occurring. This results in an undirected weighted graph where the edges imply this conditional dependency between nodes. Underpinning this is often some form of partial correlation (or similar approaches) that helps to differentiate between direct and indirect interactions. The majority of approaches also consider the influences of biological covariates and other methodological biases (e.g. NGS read depth). For example, SPIEC-EASI (Kurtz et al. 2015) avoids erroneously inferred interactions driven by abiotic factors and indirect effects via using sparse inverse covariance to capture associations or neighbourhood selection. It avoids issues of compositionality by using a center-log transformation of the OTU/ASV co-occurrence (abundance) data and deals with issues of data sparsity via random subsampling.

2.6 Bayesian Methods for Network Inference

Bayesian methods have the potential to provide a greater predictive capability when capturing microbial interactions. Several packages have been developed that rely on Bayesian approaches, including: **BAnOCC** (Schwager et al. 2017); **BioMiCo** (Shafiei et al. 2015); **BiomeNet** (Shafiei et al. 2014), which initially focused on inferring metabolic interaction networks within microbial communities; and **MDSINE** (Bucci et al. 2016) and **CGBayesNets** (Lugo-Martinez et al. 2019), which use Dynamic Bayesian networks and are applied to time-series data. Other methods mentioned previously (e.g. MDiNE; McGregor et al. 2020), include some aspect of Bayesian inference within their algorithms. Bayesian fitting methods provide a platform to simultaneously capture multiple forms of relationships between species within a community (e.g. linear vs non-linear) through considering the joint multivariate probability distributions of multiple species simultaneously (Cardona et al. 2016). This has the potential to reveal interactions within complex communities and various interdependencies that may be missed using other

statistical inference methods. However, while Bayesian approaches are increasing being used, far fewer studies have deployed these methods compared to correlation or graphical methods (e.g. SparCC and SPIEC-EASI). Whilst this is not an inherent limitation, it does mean contextualising results of Bayesian network inferences is harder. Moreover, very few benchmarking studies that examine the performance (e.g. sensitivity, precision, etc.) of microbial ecological network algorithms have considered Bayesian methods. However, one major criticism of inferring microbial ecological networks, is that a comprehensive and quantitative comparison and validation of **all** approaches, and how these link to different data properties does not exist.

2.7 Logic-Based and Machine Learning (ML) Algorithms for Network Inference

Logic-based and ML algorithms offer a very promising approach to inferring microbial ecological networks. However, these have yet to be applied widely to the analysis of microbial communities surveyed using NGS approaches (Vacher et al. 2016). Nonetheless, these methods have been used to successfully learn ecological interaction networks from species co-occurrence data when it was combined with background knowledge of the ecosystem when applied to macro-organisms (Bohan et al. 2011; Tamaddoni-Nezhad et al. 2013). Including reconstructing ecological networks as accurately as methods based on direct observations of biotic interactions (Tamaddoni-Nezhad et al. 2013). Barroso Bergadà (2022) made recent progress in applying these approaches to microbial NGS data, and developed (with collaborators) “InfIntE: a generic, logic-based inference tool for learning networks in R” for conducting the analysis. The Abductive/Inductive Logic Programming behind InfIntE performed well at identifying ecological interactions in both trials using computational simulations, and when applied to microbial NGS data collected from vineyards (Barroso Bergadà 2022). Impressively, it was able to capture known interactions, and their forms (e.g. negative, positive, neutral) between species (Barroso Bergadà 2022). As these methods develop, they may provide the ideal platform for inferring microbial ecological networks in a NGB context.

2.8 Inferring Microbial Ecological Networks from Longitudinal (time series) Data

The majority of approaches already described are most appropriately applicable to cross-sectional data. This is intentional, as all new NGB programmes will initially only have spatial (cross-sectional) data available, until they have operated for sufficient time to build a time-series and provide longitudinal data. However, inference of microbial ecological networks from longitudinal as opposed to cross-sectional data may be more rigorous. While the same issues of compositionality and sparsity of DNA/RNA-based NGS data exist and need to be dealt with in longitudinal approaches, by taking repeated samples from the same system, the likelihood of introducing spurious association from indirect abiotic effect

may be greatly reduced. Currently, there are a number of tools for inferring microbial ecological networks from longitudinal data (Faust et al. 2015), but given the rapid speed of development of new approaches, novel NGB programmes should explore newer options at the point it becomes clear that sufficient longitudinal data will be generated.

Available methods for dealing with longitudinal/time-series data include: statistical approaches based on autoregressive integrated moving average (**ARIMA**) with Poisson errors and fit with elastic-net regularisation (Ridenhour et al. 2017); **RMN** (Tsai et al. 2015); **eLSA** (Xia et al. 2013); **FASTLSA** (Durno et al. 2013); and a suite of methods based around generalised Lotka-Volterra (**gLV**) models which includes, **MTPLasso** (Lo & Marculescu 2017b); **TIME** (Baksi et al. 2018); **metaMIS** (Shaw et al. 2016); and **LIMITS** (Fisher & Mehta 2014). The use of gLV is intuitive from an ecological perspective as vast amounts of ecological theory, including food web theory (as the dominant form of ecological network analysis), is built on examining Lotka-Volterra dynamics (e.g. Lewis & Law 2007). The gLV approach models the community's dynamics by considering the growth rate of each species within the network, their influence on all other species' growth rates, and additional stochastic influences from sampling errors and environmental factors (Cappellato et al. 2021). The model interaction coefficients are then estimated from the data using a range of regression approaches and these become edges within the network. The statistical approaches behind the inference of these model parameters is the major difference between the various algorithms behind these analyses. Potentially using gLV approaches has a great benefit to NGB, as it could provide a historically better developed mechanistic underpinning of the observed microbial ecological networks. However, there remains questions over the logistical feasibility of this approach to deliver within a NGB context. For example, while longer-term trends in the behaviour of the ecosystem will be described it may not be the most suitable approach for detecting rapid or more transient phenomena. It is also unclear as to what an appropriate samples size is (i.e. number of samples through time) to produce a rigorous analysis of microbial ecological networks using these methods.

One other longitudinal/time-series data option is based around Local Similarity Analysis (LSA; e.g. eLSA; Xia et al. 2013 and FASTLSA; Durno et al. 2013). LSA infers co-occurrence networks via a dynamic programming approach that identifies lags and association between two times series by minimising similarity scores between then. This makes it particularly good at identifying species interactions that occur in short periods over time, or those where the cause-effect relationship involves lags (Cappellato et al. 2021). They are also particularly suited to accurately identifying associations within data where interactions are sparse (Weiss et al. 2016).

3. Analysing Microbial Ecological Networks

3.1 Network Properties and Metrics

Multiple properties of microbial ecological networks can be measured, ranging from those of individual network nodes to higher level properties of the network. Even the simplest of these measures should be considered carefully in the context of the microbial ecological network. The number of nodes in the network is the simplest network measure, and is the number of OTUs/ASVs/taxa/species/functional genes that are connected. Assuming for simplicity, that the nodes are species, even this simple measure will only be equivalent to the ecological concept of species richness if all species sampled are present in the network, i.e. there are no species that exist in isolation (or these species are not removed from analysis). Number of nodes is commonly used as measure of ecosystem response to a perturbation (Favila et al. 2022). For ease of understanding, each node will be considered as a microbial species while defining the network measures. The edges between nodes represent the associations between the species, with the weighting (in a weighted network) indicating a measure of the strength of the interaction/association. All measures that follow have been used to infer something about the ecological role of the species (e.g. a keystone species), the health or functioning of the ecosystem or its response to some form of perturbation. It is important to be cautious in these interpretations however, as most are still based on extrapolating from other ecological networks where the interaction is better understood, without experimental validation in microbial networks (Guseva et al. 2022).

The **degree** (or occasionally connectivity e.g. Deng et al. 2012) of a node is the number of edges connected to that node (i.e. the number of other species that a species is associated with) and its **degree centrality** is its degree divided by the maximum possible degree were the node is directly connected to all other nodes (i.e. degree standardised between 0 and 1). Measures of degree have been used in identifying hub species and keystone species, although the validity of this is questioned by Guseva et al. (2022), who suggest it is more reflective of niche preference. The **betweenness** of a particular node is the number of shortest paths between any two other nodes that pass through that node. **Betweenness centrality** of a node is its betweenness divided by its maximum possible betweenness (Liu et al. 2021a) (i.e. betweenness standardised between 0 and 1 as it otherwise increases with the size of the network). Confusingly, some studies do not standardise betweenness in this way, but still refer to it as betweenness centrality (e.g. Ma et al. 2016; Xue et al. 2018) and others do standardise it, but refer to it only as betweenness (e.g. Deng et al. 2012). To add additional confusion, betweenness is also sometimes referred to as stress or stress centrality (Deng et al. 2012; Costa et al. 2019). Alongside having high degree, the betweenness of a node is often used to identify keystone species, however sometimes low betweenness centrality is required (Berry & Widder 2014, Ma et al. 2016; Xue et al. 2018; Liu et al. 2021a,c) and sometimes high betweenness centrality is required (Tipton et al. 2018; Ishimoto et al. 2021). The **clustering coefficient** of a node is the number of nodes that it is directly connected to, that are also directly connected to each other divided by the total number of edges

possible if all of its neighbours were directly connected to each other (Watts & Strogatz, 1998). The **closeness** or **closeness centrality** of a node is the reciprocal of the sum of the shortest path between it and all other nodes in the network (Liu et al. 2021a). Again, this is sometimes, but not always, standardised by multiplying by the number of nodes in the network minus one (i.e. the minimum possible distance in an undirected graph; Wasserman & Faust 1994). It can identify nodes that can influence the network fastest and therefore has a role in identifying keystone species (Favila et al. 2022).

Some measures of centrality have already been mentioned, and although less frequently used, there are other measures of centrality that are considered for microbial communities. In particular, **eigenvector centrality** of a node is its associated value in the standardised eigenvector that corresponds to the maximum eigenvalue of the adjacency matrix (Liu et al. 2021). It reflects a node's importance based on the importance of its neighbours, where importance is measured by degree centrality i.e. neighbours with a high degree are worth more (Golbeck 2013). It is not obvious what eigenvector centrality means for a microbial co-occurrence network, although it has been measured for several (e.g. Deng et al. 2012; Yuan et al. 2021; Mercado et al. 2022; Fang et al. 2023) and Baldassano & Basset (2016) found that it was a useful measure to identify species previously implicated in, and possibly implicated in, Inflammatory Bowel Disease.

Next, we consider metrics at the level of the network. The **average degree** is simply the average of all node degrees (Liu et al. 2021) and reflects the complexity of the network and hence might have a role in stability, particularly resilience (Favila et al. 2022). The **density** or **connectance** is the total number of edges in the network as a proportion of all possible edges if the network were fully connected, and is also possibly related to ecosystem resilience (Favila et al. 2022). **Connectivity** is sometimes used instead of connectance, but is simply the total number of edges in the network (Favila et al. 2022). The **degree distribution** is the probability distribution that a randomly chosen node has a particular degree. This can be used to infer non-random assembly processes (Barabási & Albert 1999). **Average shortest path length**/**characteristic path length**, **average geodesic path length** in unweighted networks are all terms for the length of the shortest path between two nodes in the network, averaged over all pairs of nodes (Watts & Strogatz, 1998; Favila et al. 2022). One issue to be aware of is that in weighted networks the shortest path is actually the path with the smallest sum of weights (this is true in all networks, but in unweighted networks all non-zero interaction weights are 1) (Liu et al. 2021a), which is meaningless in networks where interaction strengths reflect interaction frequency (Costa et al. 2019). This path length is usually small in microbial networks and has implications for ecological resilience through the speed of the network's response to perturbations, and also community cohesion (Favila et al. 2022). The **clustering coefficient** of a network is the clustering coefficient of each node averaged over all nodes in the network (Watts & Strogatz 1998). This means that it is the average probability that two species which are both directly associated with a third species are also directly associated with each other. High clustering coefficient has implications for redundancy (Favila et al. 2022), and can indicate modules within a network (Liu et al. 2021a). **Modularity** is a measure of the compartmentalisation of the network into subgroups that are highly connected within themselves but sparsely connected with each other (Newman

2006), and the presence of modules can be due to shared ecological functions of the clustered species, spatial compartmentalization or similar habitat preferences (Favila et al. 2022), or niches (Faust and Raes, 2012, Favilla et al. 2022). The presence of modules has also been implicated in community resilience and resistance (aspects of stability) (Favila et al. 2022).

Hub nodes/modular hub nodes are nodes that are highly connected to other nodes (i.e. have a high degree) in the network or within a module. The presence of a few highly connected nodes is a property of scale-free networks (Barabási & Albert 1999), which microbial association networks have been shown to be (see Faust & Raes 2012). Such networks are robust to random node removal but not to the removal of hubs (Albert et al. 2000), although the importance of the hubs in microbial networks is questioned (Faust & Raes 2012). In a co-occurrence network, hub nodes/modular hub nodes represent species that are directly associated with many other species (across the whole network or within a module). Consequently, species identified as occupying hub nodes are often considered as possible keystone species within the network. Sometimes keystones are identified as those nodes with a high degree alongside high closeness centrality (Berry & Widder, 2014), playing an important role in stability (Liu et al. 2022). The role of betweenness in identifying keystone species is less clear. **Connector nodes (bridging nodes)** are nodes that connect modules within networks, thus have high betweenness centrality (Costa et al. 2019; Favila et al. 2022) (or other centrality measure; Costa et al. 2019). These nodes play an important role in communication (Favila et al. 2022) and the spread of perturbations (Costa et al. 2019). Although connector nodes do not necessarily have a high degree, they are also sometimes considered as keystone species (Favila et al. 2022; Liu et al. 2022) as their removal could cause networks to become disconnected.

4. Current applications of microbial ecological networks analysis

Network theory and analysis have been applied to many systems from social networks, computer networks, and power networks, to biological networks such as neural networks, epidemiological networks and molecular interaction networks (for a review, see Strogatz 2001). In particular, a large body of work taking a network approach to the classification and modelling of macro-organisms in food webs, mutualistic webs and host-parasitoid webs exists (for a review, see Ings et al. 2009). This approach proved particularly useful in understanding the role of complexity in the stability of food webs (see Ings et al. 2009). After applying network theory to interaction networks of macro-organisms it is logical to extend this to the networks of microbial interactions. Network theory has been well developed in these aforementioned areas and hence potentially useful insights may be learned from these other disciplines.

The simplest use of the network approach applied to microbial networks is to measure network topology (e.g. clustering coefficient, average degree, degree distribution, mean

shortest path) as a way to classify or compare between networks (Steele et al. 2011). This approach has been used to describe how microbial networks change under changing conditions, for example elevated carbon dioxide levels (Zhou et al. 2011), warming (Deng et al. 2012; Yuan et al. 2021), salinity (Zheng et al. 2017) seasonality (Lin et al. 2019; Fang et al. 2023), elevation (Chen et al. 2022) and water and nutrient availability (Hernandez et al. 2021). It has also been used to compare between undisturbed networks and those subject to industrial waste (Zapellini et al. 2015). Network measures have also been used to investigate the effects of and recovery from perturbations such as changes in pH (Feng et al. 2017) and drought (de Vries et al. 2018).

Network approaches applied to microbial communities predominantly investigate four main aspects of interest (and these are not mutually exclusive within studies). In the majority of studies, understanding how the community network topology changes, is a precursor to understanding the **stability** of the community, although this is not entirely straightforward in co-occurrence networks. The identification of **keystone species** through the properties of the nodes also makes up a significant amount of research in this area, but is also not without controversy. The inclusion of **environmental variables** as network nodes has enabled some interesting findings. Finally, there is some research that directly considers **ecosystem functioning**. It is worth noting, that the response of microbial ecological networks to stressors (especially multiple interacting stressors) and how this links to issues of stability and functioning is poorly explored in freshwater systems (Codello et al. 2022).

4.1 Ecological Stability Inferred from Microbial Ecological Network Analysis

To a large extent, community stability is inferred from our understanding of how network properties relate to stability developed from other areas. For example, Wang et al. (2018) showed that microbial (bacterial and archaeal) interactions in empirical semi-arid grassland soil networks strengthened with higher precipitation, generating more complex networks, with higher clustering and connectance, more negative interactions, more module hubs and connectors. From this they inferred (rather than tested) that higher precipitation would lead to higher stability networks, or conversely that desertification would lead to less stable networks. In an attempt to provide some empirical support for a link between certain network properties and stability, de Vries et al. (2018) considered bacterial and fungal networks before, during and after drought conditions, measuring network properties associated with stability and also community measures of stability (species richness, species evenness and Bray-Curtis similarity between drought and control communities) through time. They found that bacterial networks showed network properties associated with low stability (e.g. high connectivity and centrality, and low modularity), while fungal networks showed properties associated with higher stability. They also found that bacterial communities were more strongly impacted by drought than fungal networks (i.e. were less resistant) and bacterial communities did not return to pre-drought communities (i.e. were less resilient), indicating agreement with the low-stability indicators of high connectance and centrality and low modularity. Interestingly, connectedness and

centrality of nodes in bacterial networks increased further after drought conditions, indicating even lower stability. The opposite was true of fungal networks. In their study system, there is considerably more going on that could affect stability than just network properties. The plants present were strongly affected by drought and had a much stronger link with the bacterial community than with the fungal community. However, the study is one of only a few to explicitly investigate rather than assume a link between network properties and stability in microbial networks. It would be particularly useful to consider network topological properties (connectance etc.) and measures of community stability in response to perturbations only within bacterial networks, but this would probably require a large range of habitats and/or bacterial taxa present to generate sufficient differences in network measures to be insightful.

The concept of network modularity has also been explored to assess microbial networks. For example, commonalities between taxa/OTUs in a module can be discovered once modules (groups of closely interacting taxa/OTUs) have been identified (e.g. Zhou et al. 2011; Xiong et al. 2018). High modularity is an indicator of stability in many ecological networks (while not without controversy and conflicting results, this seems to be generally accepted – see Teng & McCann 2004; Stouffer & Bascompte 2011; Grilli et al. 2016). Recently, Hernandez et al. (2021) studied soil microbial communities along environmental stress gradients (water and nutrient availability), and measured modularity and cohesion (the negative:positive interactions ratio) of the co-occurrence networks. They found networks with lower water and nutrient availability (under higher stress) were less modular and had lower cohesion (i.e. dominated by positive co-occurrences) than those from lower stress environments. Using low modularity (and low cohesion) as an indicator of low stability, they took this as indicating lower network stability in high stress environments, thus it was concluded that environmental stress destabilizes microbial community networks (Hernandez et al. 2021). However, recent studies of microbial communities have shown both less modularity under stress (Wang et al. 2018 – precipitation in semi-arid grasslands; Hernandez et al. 2021 – water and nutrient availability; Price et al. 2021 – addition of an alkaline stabilized biosolid stressor; Ye et al. 2021 – water availability) and more modularity under stress (Zheng et al. 2017 – salinity; Yuan et al. 2021 – warming), suggesting network response in terms of its modular structure is as yet unclear.

Hernandez et al. (2021) also focuses on the role of the ratio of negative to positive interactions in response to stress. It uses a measure of cohesion originally developed by Herren & McMahon (2017), not as a network metric, but using the same network correlation matrix, essentially weighting positive and negative interactions of taxa (correlations) by their relative abundances in samples. Cohesion was shown to be a good predictor of community turnover, particularly negative cohesion (i.e. weighted negative interactions/correlations), better in fact, than using all of the available environmental variables. The importance of negative interactions on network stability is well established, especially in food webs (e.g. McCann et al. 1998; Neutal et al. 2002; Teng & McCann 2004), and it has also been shown mathematically in networks which reflect the types of interactions that are present in microbial networks (Coyte et al. 2015). There is also empirical evidence for this in microbial communities. Feng et al. (2017) found that the negative interactions between key functional taxa in a biofilm community might

fundamentally impact resilience. However, they found increased strength of this negative interaction was associated with a longer recovery time (Feng et al. 2017). In a complementary approach, it has also been shown empirically that negative interactions are lost in high stress environments (Hernandez et al. 2021).

A common way to consider the effects of interaction strengths across an ecological network is to directly consider linear stability of the system, particularly of a system described by a set of generalised Lotka-Volterra (gLV) equations. Linear stability considers the ability of the system to return to its equilibrium state after a perturbation, i.e. the resilience of a system (Pimm 1984). This method is very common in networks of macro-organisms such as food webs beginning with May (1972), but less so for microbial networks. Partly this must be due to the difficulties in parameterising true interactions between microbes, i.e. causal relationships where changing abundances of one taxon directly affects the abundance of another, something that is more easily done and better understood in macro-organisms. Guesseva et al. (2022) are particularly cautious of the gLV approach as microbial interactions are usually inferred from co-occurrences rather than from causal relationships. However, the use of time series data offers a method to achieve this. Stein et al. (2013) use time series data of simple (few species) intestinal bacterial communities training and testing on subsets of the data. Coyte et al. (2015) tested their theoretical model using the time series data of Stein et al. (2013). This approach, however is obviously quite data intensive and longer time series data for large communities do not exist in general for new NGB programmes. One other possible method is to directly establish interactions through culturing. A study of microbial UTI communities by de Vos et al. (2017) took this approach and also showed over representation of competitive interactions particularly between closely related isolates (within genus), suggesting such interactions promote stability in these networks. They parameterised a gLV model for interacting species based on their empirical data and identified the communities which had (linearly) stable communities, and those that, due to the presence of an additional unstable fixed point/community, could only be assembled in certain ways (otherwise at least one species would be absent from the final community – i.e. it would reach the alternative fixed point community) (de Vos et al. 2017). The obvious limitation to this approach is the microbes must be culturable, and secondly there will be a limit on the number of interactions it is feasible to investigate in this way.

A final word of caution in the use of networks to consider microbial community stability relates to the complexity-stability debate. The debate around whether complexity leads to stability in food webs was largely triggered by May (1972) and generated many discoveries, such as the type of links and strength of links being important, rather than just the presence of links (e.g. McCann et al. 1998; Neutal et al. 2002; Teng & McCann 2004). There are some studies of microbial networks that either show or assume that high connectivity (a measure of high complexity) means less stable networks (de Vries et al. 2018). Alternatively, there are some that show or assume high connectivity (more complexity) means more stable microbial networks (Wang et al. 2018; Yuan et al. 2021). Closer investigation, however, suggests that the apparent disagreement is a result of the types of interactions being included in the network analysis. The study of de Vries et al.(2018) includes only positive associations, and large numbers of these have been

shown to be destabilising. Wang et al. (2018) include all interactions, and we know from the diversity-stability debate that more negative links can be stabilising. It is therefore worth being very clear about the types of interactions included in any network before drawing conclusions about stability.

4.2 Keystone species in Microbial Ecological Networks

A lot of attention has been given to the identification of microbial keystone species/taxa using network measures, understandable given the clear links to stability and ecosystem health. In general ecology, keystone species are those that have a large effect relative to their abundance (Power et al. 1996) – i.e. they are found in low numbers but have a large effect. This is not necessarily true for microbes, however, as some taxa can have very important keystone roles and are often found at high abundances. An example of this given by Banerjee et al. (2018) is a species of bacteria found in the human intestine, which is an anaerobic symbiont. This prompted Banerjee at al. (2018) to provide the following definition of a microbial keystone: “keystone taxa are the taxa which have major influence on microbiome composition and function at a particular space or time. These taxa often, but not always, have an over-proportional influence in the community, relative to their abundance.”

The majority of the studies that identify keystone (or potential keystone) species/taxa do so using network metrics based on how candidate taxa are connected to others in the network. There is however some disagreement about what these should be. Berry & Widder (2014) use those with high mean degree, low betweenness centrality (few of the shortest paths between any two nodes in the graph pass through a keystone node), high closeness centrality (a keystone node has low average distance to any other node), and high transitivity. Many studies have used this classification to identify keystone species in microbial networks (e.g. Ma et al. 2016; Liu et al.2021c). While high degree is often accepted as a required property of a keystone species, even this is not necessarily generally true. For example, bridging nodes/connectors connect modules within the network, thus occupying a potentially important ecological role and whose loss could cause networks to become disconnected, they themselves tend to be connected to only a few other nodes (low degree). These too should be (and sometimes are) considered as potential keystones (e.g. Fang et al. 2023).

There is also difficulty in identifying an appropriate level of “betweenness” to classify keystone taxa. In contrast to the Berry & Widder (2014) approach of using low betweenness centrality, others use high betweenness centrality (alongside high degree) as an indicator (e.g. Tipton et al. 2018; Ishimoto et al. 2021). The argument for using high betweenness is a logical one, i.e. that a node appearing in more paths between other nodes shows a greater influence over more nodes. Berry and Widder (2014) used simulations to identify “keystoneness” of taxa, removing nodes from networks and measuring the impact, hence their requirement of a low level of betweenness stems from simulated data. Although this adds some weight to the argument to use low betweenness,

Guseva et al. (2022) caution against the use of network topology to identify keystone species specifically because removing a node in a simulated network based on co-occurrences will not necessarily have the same effects in a real-life networks, as co-occurrences do not imply dependencies. This disagreement in whether to use high or low values of betweenness, somewhat weakens the argument for using network properties to identify keystone species in microbial networks. It could be linked to the previously mentioned issues surrounding stability and complexity, that is whether all associations or only positive associations in the network are being considered. Whatever the underlying reason, is an issue that needs to be resolved before using "betweenness" in any meaningful way going forward.

Finally, keystone identification from networks of correlations has obvious limitations that we have previously mentioned as general limitations, such as habitat filtering and spurious associations (Banerjee et al. 2018) and lack of causal relationships between taxa (Banerjee et al. 2018; Guseva et al. 2022). To avoid these issues there is a sensible requirement to support theoretical suggestions of candidate keystone taxa with empirical evidence (Banerjee et al. 2018; Banerjee et al. 2019). There are, however, clear limitations to testing "keystoneness" empirically, in that not all microbes are culturable therefore identifying potential keystone taxa has value (Banerjee et al. 2019).

4.3 Incorporating Environmental Variables into Microbial Ecological Networks

The most obvious way to consider links between the environment and network structure is simply to carry out tests of association between the two (e.g. Wang et al. 2018). However, environmental variables have been included in the analysis as network nodes to identify positive and negative associations between taxa/OTUs and the environment (e.g. Ju & Zhang 2015; Zhao et al. 2016; Wang et al. 2018; Lin et al. 2019). As well as providing useful results in its own right, the method of including environmental variables as nodes is one way to deal with the issues of co-occurrences that are due to taxa having very similar niche requirements (Faust, 2021). Interactions between taxa can then be clearly identified as indirect (through the environmental variables) rather than direct. It is also possible to specifically consider environmental variables as nodes between modules in a network to understand how these affect the network and which variables are most important (Zhao et al. 2016). For this, eigengene network analysis has been used where an abundance profile is produced for each of the modules within the network (called the module eigengene) (Zhao et al. 2016). This method was developed for gene expression networks (Langfelder & Horvath 2007), but has been applied to microbial networks with some success (Zhou et al. 2011; Deng et al. 2012; Deng et al. 2016; Zhao et al. 2016; Zheng et al. 2017; Yuan et al. 2021). However, it is unclear if there is any particular advantage to approaching this question with a network methodology, or via statistical methods that model (e.g. via multivariate general linear models) the response of all microbial taxa within a community to changes in environmental conditions (e.g. Alzarhani et al. 2019; Ferguson et al. 2021).

4.4 Linking Microbial Ecological Networks to Ecosystem Functioning

Given the importance of microbes in ecosystem functioning, many studies have incorporated some aspect of functional role into the study of networks of microbes. Some of these present findings that are only possible because of taking the network approach to studying the microbes present. For example, Chen et al. (2022), sampled fungal and bacterial communities at increasing elevation, built co-occurrence networks at each elevation, and also measured species diversity (using standard diversity indices). At each elevation they also measured ecosystem multifunctionality, which combines 18 ecosystem functions known to be regulated by microbes (including soil carbon, nitrogen, and phosphorous, plant growth, mitigation of greenhouse gases, and control of potential fungal plant pathogens in soils). They found steeper relationships, with more variance explained, between network complexity and multifunctionality than between diversity and multifunctionality. Interestingly, once confounding variables (climate, soil and biotic factors) were accounted for in their analyses, there was little or no association between diversity and multifunctionality, but positive relationships between network measures and multifunctionality persisted.

Functional molecular ecological networks (fMENs) have been useful in studying ecosystem function in this way. This approach enabled detection of such things as the most highly connected functional gene nodes in an elevated CO_2 soil being different to those in corresponding ambient CO_2 networks, and the same functional gene nodes having very different network characteristics under the different treatments (Zhou et al. 2010). For example, N fixers interacted with nodes associated with N fixation, denitrification, C fixation, C degradation, sulphate reduction, sulphur oxidation and P utilization under elevated CO_2, but interacted with very few other functional gene nodes under ambient CO_2 (Zhou et al. 2010). This was correlated with changes in soil biogeochemical properties hence indicating the importance of network structure on ecosystem function (Zhou et al. 2010). Yuan et al. (2021) further support the link between network structure (particularly complexity) and microbial community functional structure and therefore ecosystem function. Deng et al. (2016) studied uranium degradation in groundwater microbial communities subjected to the addition of an emulsified vegetable oil (EVO). This would be expected to increase the “phylogenetic richness and diversity” of the network. However, only a bacterial taxon associated with sulphate reduction increased dramatically (Deng et al. 2016). Through network analysis from a functional trait point of view, Deng et al. (2016) showed that this taxon occupied a hub node and had predominantly negative interactions with its neighbours, in particular its eight neighbours which carried carbon cycling genes. The much higher abundance of this taxon after the EVO addition limited the growth of its neighbours, preventing the expected increase in richness and diversity. The network approach allowed insight into why the system did not behave in the expected way, and importantly, identified implications for carbon cycling, which would otherwise have been overlooked (Deng et al. 2016).

5. Methodological limitations to microbial ecological networks analysis

5.1 Underlying Problems with Microbial Ecological Network Analysis

As with all emergent methodologies, there are multiple issues, limitations and considerations with microbial ecological network analyses that have yet to be fully resolved (Röttjers & Faust 2019; Faust 2021). Within the context of NGB, this is further confounded by the inherent limitations and methodological biases associated with DNA/RNA based metabarcoding and environmental metagenomics/metatranscriptomics. We have covered specific limitations associated with various methods, approaches and interpretations of microbial ecological networks earlier in this think piece. Moreover, issues with the underlying molecular ecology approaches are addressed in detail elsewhere (e.g. Albertsen et al. 2015; Schirmer et al. 2015; Gołębiewski & Tretyn 2019; Porath-Krause et al. 2022) and numerous best practice guidelines have been published (e.g. Zinger et al. 2019; Bohmann et al. 2022; Tedersoo et al. 2022). Here we cover major background limitations general to all microbial ecological network analyses.

While addressing every molecular ecology issue would be superfluous, a very specific set of attributes of these approaches have a disproportionate impact on microbial ecological network analyses (Faust 2021). These include, the compositional nature of microbiome NGS datasets (Quinn et al. 2018; Gloor et al. 2017), the influence of data processing choices and bioinformatics pipelines on OTU/ASV count data (e.g. Schirmer et al. 2015; Siegwald et al.2019; Prodan et al. 2020), underlying issues with amplification biases and multi-copy genes (Dahllöf 2002; Gonzalez et al. 2012; Pierella Karlusich et al. 2022). In addition, to issues driven by the properties of NGS datasets, there are also fundamental issues with samples sizes and with spatial and temporal scales of analyses that influence both network inference and properties. Moreover, there are additional issues regarding the interpretation and meaning of microbial ecological networks. These range from linking inferred interactions to ecological/biological interactions and how interaction strength may or may not influence community composition (Faust 2021), how to deal with higher order interactions, “hairballs” and identifying core networks (Faust 2021) and importantly from a NGB perspective, the links between microbial networks and ecosystem-level properties (e.g. stability, functionality etc).

5.2 Limitations to Microbial Ecological Network Analysis from Molecular Data

Microbial OTU/ASV count data generated via NGS platforms is both sparse and compositional (Quinn et al. 2018; Gloor et al. 2017). The former refers to the high proportion of null values within a typically zero-inflated dataset; this can range between 70%-90% of values for a 16S rRNA gene metabarcoding dataset (Cappellato et al. 2021). The later refers to microbial OTU/ASV count data representing relative proportions of a

total rather than counts on a continuous scale. This is because all current NGS platforms have an upper maximum to the number of sequences they return, and these are then split across the OTUs/ASVs present in each sample sequenced, such that an increase in the relative abundance of one OTU/ASV will cause a decrease in the relative abundance of others (Quinn et al. 2018; Gloor et al. 2017). Both the sparse and compositional natures of these data can produce issues with microbial ecological network inference, but these are well known and a range of solutions has been provided (see, for example, SparCC, Friedman & Alm 2012; SPIEC-EASI, Kurtz et al. 2015). These data properties have a strong influence on correlation-based (and those based on other statistical methods for evaluating associations) network inferences, which if ignored (i.e. left unaccounted for in the network inference algorithms used) may lead to erroneous correlations with no underlying biological significance. This is because all pairwise associations within the OTU/ASV count data are non-independent (i.e. because they are proportions of a total, any change in one variable will automatically lead to a change in others regardless of whether this is biologically meaningful or not), violating the basic assumptions of many statistical approaches for evaluating associations. As Friedman & Alm (2012) note, this has been well known for >100 years, and consequently there are existing statistical approaches that can overcome these issues.

Whilst the majority of current network inference algorithms account for the compositionality and sparsity of microbial NGS data, it appears that different algorithms are better suited (in terms of sensitivity and precision) to data that are more vs less sparse, have higher vs lower compositionality (Weiss et al. 2016), alongside other properties associated with species richness and rarity. Moreover, while this issue is known for a subset of microbial network inference algorithms (for example, those explored in Weiss et al. 2016), it remains largely unknown as to the extent of this problem across all microbial network inference algorithms when applied to datasets with a range of properties. This is because a comprehensive validation of all microbial network algorithms against both real world, and simulated data, with known variation across a range of data properties does not currently exist. This wouldn’t be an issue if all microbial NGS data had similar properties, but these vary significant across datasets and this is due both to differences in the ecosystems and environments sampled, and difference in the bioinformatics pre-processing before networks are built. Differences between ecosystems and environments in their microbial communities, and the properties of microbial NGS data when surveying these is not a major issue for NGB programs, as these are likely to focus on specific ecosystems/environments in isolation (e.g. River Surveillance Network). However, it does raise the issue of how the NGS data are pre-processed before any network analysis is undertaken (Faust 2021).

The bioinformatics behind the processing of NGS data, which produce the species (OTUs/ASVs) count table on which network analyses are built, is covered elsewhere (e.g. Dumbrell et al. 2017). It is clear that choices associated with sequence trimming, quality checking, error correction, and pair-end alignments all influence the properties of the species (OTUs/ASVs) count table (Schirmer et al. 2015; Siegwald et al. 2019; Prodan et al. 2020; Barroso-Bergada et al. 2021). Currently the most popular pipeline (based on citations) for these analysis is DADA2 (Callahan et al. 2016), but whether this popularity is

due to superior performance, or ease of use (as a package for the R language, it avoids the need for additional computational knowledge), or a combination of both is unclear. From an NGB perspective, the use of a popular approach brings additional logistical benefits in terms of a wider user-base to support problem solving.

In terms of influencing results of ecological network analysis, two of the final stages in NGS metabarcoding bioinformatics are likely to have disproportional influences, these are data normalisation and the choice of generating either OTUs (and the associated clustering algorithm) or ASVs. Due to differences in sample processing and sample concentrations loaded on NGS platforms, the number of sequence reads returned will differ across samples. As this is related to the number of unique sequences each sample will have (i.e. richness of OTUs/ASVs), with more new sequence types likely to be observed in samples with greater sequencing depth, some form of data normalisation is recommended. As this influences the compositionality of the data, it is clear to see how this may influence network inference. Multiple approaches to this normalisation have been proposed and there is extensive literature discussing the relative merits of each method (McMurdie & Holmes 2014; Weiss et al. 2017; McKnight et al. 2019; Cameron et al. 2021). However, these have generally been compared from a community ecology perspective, with suggestions, for example, that the rarefaction method works best when comparing communities (McKnight et al. 2019), but it is still unclear which is the most appropriate method to support robust microbial ecological network analysis (Faust 2021).

In recent years, there has been a shift from OTU based methods to ASV based methods. OTUs work by clustering together sequences that are similar to each other (e.g. within 97% sequence similarity) and using this as the fundamental taxonomic measure (i.e. representing bacterial species). ASVs take exact sequence variants and use this to represent species, based on the assumption that current methods for removing errors from Illumina NGS data is sufficient to avoid artefactual ASVs (Callahan et al. 2017). The choice of approach affects estimates of species richness, and may have a greater influence on data properties than choice of normalisation approach (Chiarello et al. 2022). It is not clear which method (and how this should be combined with normalisation approaches) is most suited to microbial ecological network analysis, although it is clear due to the affects these have on data properties that it will be important. One thing that is worth noting, is that the choice to use OTUs or ASVs may be taxon-specific. For example, fungal metabarcoding based on the ITS region, is more suited to analysis via OTUs than ASVs (Tedersoo et al. 2022). Thus for any NGB programme that works across taxonomic groups (e.g. fungi, bacteria, algae), the decision to use OTUs or ASVs may well be determined by the taxon with diversity estimates most susceptible to this methodological decision.

5.3 Spatial and Temporal Scale of Analysis and Appropriate Sample Sizes

In a NGB context, there are a few considerations (as opposed to inherent limitations and issues) around samples sizes and the spatial and temporal scales of analyses. There is variation across algorithms for co-occurrence network inference in their required samples size. For example, correlation/regression (association-based) approaches and graphical model inference approaches may be able to operate on smaller sample sizes than methods based on Machine Learning. However, while many microbial ecological network inference methods have to some extent examined their own performance relative to different sample sizes (e.g. Kurtz et al. 2015; Lo & Marculescu 2017a), best practice recommendations across different inference algorithms and from microbial communities sampled from different ecosystems or operating under different ecological contexts have yet to emerge. In general, larger samples sizes produce more robust network inferences (Berry & Widder 2014). For example, Kurtz et al. (2015) showed optimal performance with n > 1300, but also adequate performance with fewer samples. For new NGB programmes, some preliminary analysis examining how their perceived microbial ecological networks are influenced by differences in sample size is probably needed. Ideally, this would be validated using microbial communities with known properties and interactions, but this is hard to achieve and doesn't currently exist (Faust 2021). There have been some recent developments in conducting power analyses for microbiome studies (Kelly et al. 2015; Ferdous et al. 2022). While these don't specifically focus on microbial ecological network analysis, they do examine sample size (power) issues across a range of commonly applied metrics and statistical tests, and this could be developed further to cover network inference.

Related to issues of sample size, is the spatial and temporal scale of sampling, and in some cases these issues are intertwined. Microbial ecological networks describe the microbial community at a far larger spatial (or temporal) scale than the individual samples were collected at. For example, a river might have n = 100 samples collected where each comes from a few grams of sediment at the millimetre scale, but a single co-occurrence network is inferred across these and thus describing something closer to the river scale. As species richness scales with space (i.e. Species Area Relationships) and the strength of this relationship is influenced by environmental heterogeneity, any changes in heterogeneity between area/habitats being compared will lead to scale dependent results (Dumbrell et al. 2008). How these issues of scale-dependencies translate to network inference is not well explored, but it is conceivable they may influence comparisons between inferred networks as species richness and evenness have minor but noticeable affects on network inference (Berry & Widder 2014). Moreover, underlying spatial structure in the sampled microbial communities, may influence the ability of microbial ecological network analysis to accurately identify interactions (Armitage & Jones 2019). These issues are likely caused by the disconnect between the spatial scale species interactions operate over, and the spatial scale of sampling. However, this could potentially be accounted for by measuring small-scale spatial heterogeneity and including this in part of the analyses (Armitage & Jones 2019). Potentially a bigger logistical challenge for NGB programmes is in the initial phases, where ecosystem process or ecosystem function measurements may

be collected and related to properties of the microbial ecological networks. Here it is likely that the disconnect between the spatial scale of sampling, and the spatial scale the network is constructed over is even more exaggerated. For example, if a single process measurement is taken at the location of each microbial sample, then issues of integrating the process data to match the spatial scale the network is constructed over will exist.

6. The major knowledge gap in microbial ecological networks analysis

Faust (2021) outlined 10 major challenges in microbial ecological network analysis, and these represent some of the major knowledge gaps that exist. Many of these are outlined in some manner within this think piece; for example, issues of NGS data pre-processing, the inclusion of environmental factors in networks, and what we can learn from various network properties. It is tempting to simply reiterate the various limitations, lacks of consensus, and underlying problems covered earlier and state that addressing these (alongside those in Faust 2021) will address the major knowledge gaps. To some extent this is true, however, most of these are specific methodological limitations (which we have already discussed). This would overlook the major issue that microbial ecological network analysis does not have an underlying theoretical basis and mechanistic framework that has been developed to the same extent as that of other ecological networks (e.g. food webs; Ings et al. 2009). This is a major issue for NGB programmes, as their goal should be to move away from phenomenological pattern fitting towards mechanistic, predictive frameworks.

Microbiology as a research area is counterintuitively both very old and very young. Microbes have been studied for hundreds of years, but due to the limitations of culture-dependent methods, this research was confined to a limited number of taxa (and their interactions) of which we now have considerable knowledge. In contrast, the ability to fully enumerate the diversity of natural microbial communities via NGS approaches is exclusively 21st century research. Thus for some particularly well studied species that are amenable to culture we have a theoretical foundation from which to interpret their role in microbial ecological networks, but for the other ~99% this is entirely absent. This is further confounded by trying to apply existing ecological theory to microbial networks. Within a food web, we have a clear understanding of interactions, such as predation, competition and mutualism, all defined clearly from observational studies of higher taxa. Whereas microbial interactions also encompass the transfer of genetic material, chemical signalling between cells, and various synergistic interactions not observed in higher taxa (Tshikantwa et al. 2018). There may even be potential for greater diversity of interaction types within microbial ecological networks, many of which are poorly understood. In addition to this, our understanding of microbial ecology in general requires a greater development of theory (Prosser et al. 2007) and while there is no doubt that the advent of NGS technologies has significantly advanced our understanding of microbial communities, it has typically generated far more methodological, descriptive, or exploratory studies than

those grounded in hypothetico-deductive rigour (Prosser 2020). While it is entirely possible to apply understanding gained from the analysis of other ecological networks to the interpretation of microbial ecological networks, without developing the theoretical and mechanistic framework specific to microbes to interpret these, we run the risk of falling short of a central goal in NGB.

Examples of where the issue of underdeveloped understanding of the mechanisms structuring and influencing microbial ecological networks is particularly acute are around higher-order interactions (Faust 2021), issues around stochastic vs deterministic assembly process in microbial communities, and links to ecosystem functions. Higher order interactions (i.e. the alteration of interactions between species by an additional species) are problematic to infer from association networks,yet may be important in promoting stability (Faust 2021). This adds an additional layer of complexity to understanding what aspects of microbial ecological networks contribute to ecological stability. In other ecological networks, we know interaction type and strength are potentially more important than the existence of links (McCann et al. 1998; Neutal et al. 2002; Teng & McCann 2004), but for microbial ecological networks there exist interactions (like higher order interactions) of which we know relatively little or can't capture. However, when microbial interactions are experimentally examined, we do see similar properties in microbial communities. For example, Ratzke et al. (2020) showed experimentally that the strength of interactions does determine stability in microbial communities. Further experimental validation of this nature is urgently needed to move microbial ecological network science inline with our broader understanding of ecological networks.

Other issues resolve around the relative influences of different assembly processes structuring microbial communities. A large volume of research has attempted to understand the relative influences of stochastic and deterministic processes in structing microbial communities (e.g. Dumbrell et al. 2010; Dini-Andreote et al. 2015; Zhou & Ning 2017). Most papers find a contribution of both processes, yet the importance of one over the other is still hotly debated (Zhou & Ning 2017). However, from a microbial ecological network context, networks inferred from more deterministically assembled (i.e. niche-based) communities may face different issues to those inferred from highly stochastically assembled (i.e. neutral-based) communities. For example, separating direct and indirect (via strong responses to environmental factors) interactions may be harder in more niche-structure communities. However, what remains unclear is how a shift from being deterministically to stochastically structured (e.g. in response to stressors) may influence network properties, or indeed ecologically what this means.

Finally, until a greater understanding of the mechanisms underpinning microbial ecological networks is developed, links to ecosystem-level properties in NGB programmes will remain tenuous. It is clear that aspects of ecological stability can be correlated to properties of microbial ecological networks (see **Ecological Stability Inferred from Microbial Ecological Network Analysis**), and our understanding of this is developing. However, the relationships between microbial ecological networks and ecosystem functions, for example, carbon and nitrogen transformations and biogeochemical cycling, are less well established. This is a clear example of where the need to develop the ecological

understanding of microbial networks is required, as many of the key taxa involved in these functions are already identified. For example, the microbial taxa (and their functional genes) for known transformations within the nitrogen cycle have been identified (e.g. Smith et al. 2016), as have microbes involved in methanogenesis and methanotrophy (Zhu et al. 2020). Thus, the knowledge gap is around how interactions across the microbial ecological network modulate these functions, whether network properties are mechanistically linked to these functions, and if these provide more robust functional predictions than alternative methods (e.g. based on qPCR copy numbers of functional marker genes).

7. Complementary approaches to microbial ecological network analysis

7.1 Molecular Ecology

There is a range of molecular approaches that can provide additional complementary data, or data that can be included in the inference of microbial ecological networks. Given ecosystem processes are driven largely by biomass or abundance and not just the identity of taxa present, it makes sense to apply any molecular approaches that can support estimating abundance to the same DNA extracts being used for NGS. For example, when incorporating microbial responses within a food-web based approach to examine pesticide spills, the abundance of key microbial groups (as opposed to richness or identity) initially changed, reflecting their functional role in the environment (Thompson et al. 2016). Methods such as qPCR or ddPCR can be used to estimate the gene-copy abundance of key microbial groups, especially those with clear links to ecosystems process. For example, the abundance (qPCR gene copy number) of the fungal cellobiohydrolase gene (cbh1) is positively related to freshwater fungal decomposition (Fell et al. 2021). Data from these assays can then be viewed either as an additional component of the network (similar to the inclusion of environmental data) or as response of the network (given their links to other ecosystem processes).

Other approaches that provide a link to functions and traits of the microbial taxa are also useful to help develop the mechanistic framework required to underpin microbial ecological network analysis. Metagenomic analysis provides functional information about the entire community, and as NGS costs decrease may well replace metabarcoding as the new data platform for microbial ecological network analyses. Potentially, the functional attributes of individual species as opposed to the entire community is more important for understanding microbial ecological networks and responses of these to environmental change. Metagenomics has the potential to provide metagenome-assembled genomes (MAGs) and thus provide species-specific functional information (Singleton et al. 2021), but this is likely to currently be cost prohibitive for most NGB programmes. Alternative methods to achieve this include Emulsion, Paired Isolation and Concatenation PCR (epicPCR), which provides a method to co-amplify both a phylogenetic and functional

marker gene (linking functions to individual species) in NGS metabarcoding (Spencer et al. 2016; Roman et al. 2021). This approach is highly suited to aquatic samples, and could potentially provide a greater functional framework for NGB. Another approach would be to infer functions of individual species bioinformatically. This could be achieved via tools such as PICRUSt2 (Douglas et al. 2020), Taxa4Fun2 (Wemheuer et al. 2020), and FUNGuild (Nguyen et al. 2016), which assign functions to data generated from phylogenetic marker genes, or via querying assembled trait databases such as BactoTraits (Cébron et al. 2021). The use of PICRUSt2, Taxa4Fun2 or FUNGuild could easily be incorporated into any NGS bioinformatics pipelines being used by NGB programmes, whereas open-source code to query BactoTraits is currently not available.

7.2 Statistical Ecology

One of the main limitations of studying the co-occurrence networks of microbes is that it is observational/correlation-based, and does not in itself provide information about causal links between taxa and network properties. Therefore, to understand underlying relationships, additional analyses need to be carried out. Structural equation modelling (SEM) has recently become a popular tool used alongside network analysis of microbial ecological networks to explore the causal relationships of interest (de Vries et al. 2018; Mamet et al. 2019; Kruk & Paturej 2020; Kaplan-Shabtai et al. 2021; Xue et al. 2022). For example, SEM allows relationships between environmental variables, microbial taxa and stability to be established (Xue et al. 2022). Whereas Banerjee et al. (2018) suggested that SEM should be used to explore keystone taxa. A possible drawback to SEM is that it requires a large sample size (Banerjee et al. 2018, Kaplan-Shabtai et al. 2021), which should not be a major limitation for microbial NGS data. However, it may be an issues when relating network properties to other variables, such as ecosystem process or function measurement, that are typically collected with lower samples sizes. Developments in multivariate general linear modelling (Warton 2011; Wang et al. 2012) have been presented as a solution to overreliance on reducing the dimensionality of community composition data via ordination approaches. It can be used to model the response of community composition and changes in the relative abundance of individual taxa (species/OTU/ASV) to environmental factors. This has already been used to show how every OTU within a microbial dataset has responded to changes in both the abiotic and biotic environment (although noting that biotic interactions were not directly measured) and the relative importance of each in structuring the microbial community (Alzarhani et al. 2019; Ferguson et al. 2021). Importantly, deploying an approach such as this may help reveal which OTUs have strong niche associations to different environmental variables. This provides both a more mechanistic link to changes in environmental conditions, which is important in a NGB context, and an additional method to help reveal indirect interactions within the microbial ecological network. In a similar manner, Species Distribution Models (SDMs) and Joint Species Distribution Models (JSDMs) are explicit environmental (or ecological) niche models and could potentially be used for all taxa within the microbial NGS data. This is particularly true for JSDMs which include co-occurrence patterns within their analyses, directly linking them to ecological networks. However, whether this is

computationally tractable within a large microbial NGS dataset spanning taxonomic domains, remains to be seen.

8. Recommendations for the River Surveillance Network (RSN) programme

Until RSN has had time to develop and generate longitudinal data, a network inference approach that utilise is cross-sectional data is most appropriate. Currently, the leading contender is likely to be SPIEC-EASI (Kurtz et al. 2015) or more efficient methods originating from it (e.g. gCoda; Fang et al. 2017), this is due to its ease of use, applicability to initial RSN data, and the volume of previous research using this method against which to contextualise findings. As Abductive/Inductive Logic Programming approaches develop, a switch to these may be beneficial due to the increase capacity to identify the form (negative, positive, neutral), magnitude and directionality of interactions (Barroso Bergadà 2022). Complementary to this, as the RSN data develop, a switch towards longitudinal analysis, which may help reveal the causal direction of responses, and avoid spurious interaction inferences, may be desirable (Cappellato et al. 2021). Approaches based around Local Similarity Analysis (LSA; e.g. eLSA; Xia et al. 2013 and FASTLSA; Durno et al. 2013) are widely used providing background research against which to contextualise findings, as well as being particularly suited to accurately identifying associations within data where interactions are sparse (Weiss et al. 2016).

The approach to the bioinformatics data pre-processing needs to be considered and some choices made. Once RSN data becomes available it would be entirely appropriate to explore a range of bioinformatics pipelines and how these influence properties of species (OTUs/ASVs) count tables that are of particular relevance to ecological network analysis and explore how these influence the inferred networks. This analysis should then guide subsequent choices. With regards to the two big choices around data normalisation methods and generating either OTUs or ASVs, some practical considerations may be involved until it is fully understood how these choices impact network inferences. Within the RSN, fungal communities will be examined via ITS metabarcoding, where OTUs outperform ASVs in recovering fungal diversity (Tedersoo et al. 2022). In contrast, using rarefaction and clustering OTUs at 99% similarities, avoids major differences between OTUs and ASVs for bacterial 16S rRNA data (Chiarello et al. 2022). Thus, choosing an OTU based approach may provide a consistent method across taxonomic groups. For NGS data normalisation, rarefaction may be the most practical method. Not only can it help reduce issues of using OTUs for 16S rRNA data, but it has been demonstrated to be the most appropriate method for analyses exploring ecological comparisons (McKnight et al. 2019), that may provide a complementary approach to microbial ecological network analysis.

As the RSN will measure a range of environmental data, including physical chemistry, hydrology, traditional invertebrate and macrophyte data, specific pressures (e.g., sewage treatment works or nutrient gradient) and organic chemical analytes such as pesticides, it will be beneficial to explore how multiple network properties respond to changes in these.

This initial exploratory approach can be used to generate specific hypotheses (Röttjers & Faust 2019), which may be followed up with analyses focused around identifying mechanisms (see below). Moreover, it provides the opportunity to explore the inclusion of environmental variables within the networks themselves, an approach which can help separate indirect from direct edges and reveal functional mechanisms underpinning patterns of co-occurrence (Röttjers & Faust 2019). While it is worth exploring a range of network properties (see Liu et al. 2021a and **Network Properties and Metrics** for a detailed list) from the RSN data, a few key metrics are likely to be disproportionality relevant to the RSN goals. Measurements of network modularity and overall network connectivity, connectance or density of links are likely to be important in revealing network responses to environmental perturbations (Wang et al. 2018, Hernandez et al. 2021, Yuan et al. 2021). There is also evidence that these properties may be linked with the stability (and potentially resilience) of the network, but also a lack of consensus around the direction (i.e. higher vs lower values promoting or reducing stability) of this link (de Vries et al. 2018, Wang et al. 2018, Hernandez et al. 2021, Yuan et al. 2021). Measures of highly connected nodes, or connector/bridging nodes, can identify nodes important for structuring the network (Banerjee et al. 2018, Fang et al. 2023). While there is considerable debate as to what these represent ecologically (e.g. are they keystone species or not; Banerjee et al. 2018; Röttjers & Faust 2019; Banerjee et al. 2019), it is clear these nodes are important in the network context. Once identified, complementary analyses can be used to explore these taxa further with the aim of providing a greater mechanistic understanding.

It may be beneficial to the RSN to expand its analysis to directly capture aspects of carbon and nitrogen cycling, and other biogeochemical processes driven by microbial communities. Microbial networks underpin key biogeochemical fluxes, processes, and nutrient transformations, although the nature of their relationship with microbial ecological networks is poorly understood. Taking direct measurements of these processes from the environment is often considered costly and time consuming but may be a worthwhile undertaking for a few key sites to establish links to microbial network properties. This may help to provide a mechanistic link between microbial ecological networks and key ecosystem processes that can be explored further as the RSN develops. Complementary to this, would be a targeted qPCR analysis of key functional genes that are known to correlate with ecosystem processes and functions. For example, the gene abundances (qPCR copy number) of cbh1 correlate to rates of fungal decomposition (Fell et al. 2021), mcrA correlates with CH_4 production (Morris et al. 2016), ureC and amoA correlate with NH_4^+ concentrations and nirK with NO_3^- concentrations (Yu et al. 2020). Gene abundance data can be obtained from the same DNA extracts as used for NGS data, and subsequently either linked to ecological network properties or included within the networks themselves (analogous to environmental data). This has the potential to provide additional mechanistic insights and directly causal relations between network properties and key ecosystem process and functions inferred from the abundances of genes driving them.

For the RSN to be effective as a NGB programme it must move away from phenomenological pattern fitting approaches of the past and provide an approach that both captures the diversity and functions of microbial communities, and provides a mechanistic, predictive framework linking this to ecosystem status. There is potential for

this with the proposed data being collected, but there remains as risk with microbial ecological network analysis that it simply becomes an approach for reducing the dimensionality of the NGS data describing microbial communities into a few key metrics (network properties) without a mechanistic link to ecosystem properties. While developing the underlying theory behind microbial ecological networks to the same extent that exist for food webs is likely outside the scope of RSN, deploying complementary analysis of the NGS data to reveal mechanistic insight is achievable. For example, developments in multivariate general linear modelling (Warton 2011; Wang et al. 2012) can be used to understand both community-level and population-level (species/OTU/ASV) responses to environmental conditions, helping to establish a causal relationship as to why certain nodes within the network behave the way they do. Whereas structural equation modelling can be used to establish causal relationships between changes in the environment and changes within the microbial ecological network and any links this has to ecosystem properties (Mamet et al. 2019).

Recommendations synthesis

Recommendations are made in the think pieces regarding the processing and analysis of data generated by the RSN. These include, how data could be analysed such as the construction and interrogation of microbial ecological networks, as well as practical considerations around their implementation. These recommendations are synthesized, compared, and contrasted herein.

1. Data processing

Prior to data exploration and analysis, data generated by metabarcoding and metagenomic analysis must be processed into a format that is suitable for analysis.

Dumbrell et al. outlined that an early decision should be made whether to work with OTU or ASV data because differences in how the data is treated to derive OTUs compared to ASVs may impact network inferences, although the extent to which this is a concern is unknown (p. 98). However, sequencing data can be normalised prior to analysis to ensure that comparisons between samples are not biased by technical artifacts or sequencing depth and OTUs can be clustered at a high percentage similarity to avoid major differences between OTUs and ASVs for 16S rRNA data (Chiarello et al., 2022). In addition, OTUs have been shown to outperform ASVs in recovering fungal diversity (Tedersoo et al., 2022). Dumbrell et al. suggested that an OTU-based approach for analysing the RSN microbial data may provide a more consistent approach across different taxonomic groups (p. 98).

2. Preliminary data analysis

Evans and Dumbrell et al. both recommended that preliminary analysis of the NGS data generated is undertaken and used to form hypotheses to be subsequently tested and explored (p. 68, 99). Think pieces by both authors suggested forming and testing hypotheses around ecosystem function and mechanisms (p. 68, 99).

Windsor suggested exploring the NGS data using more traditional data analysis techniques in the first instance, specifically suggesting the use of multivariate analyses (p. 37); these methods can be helpful in exploring and describing large, multi-dimensional datasets (Paily and Shankar, 2016). There are several challenges in applying some multivariate techniques to microbial datasets; however, specific techniques have been developed to negate these. For example, molecular analysis methods tend to generate the relative abundance of microbial taxa (i.e. data are highly compositional because they are based on relative as opposed to absolute abundances), which undermines statistical assumptions about the interdependence of variables. Methods, such as SparCC and newer iterations of this method (e.g., REBACCA), have been developed to construct networks by performing a centre log-ratio transformation on the data, which removes the effect of the constant-sum constraint (i.e., the assumption that the values for a certain

sample unit will always have the same sum) on co-variance and correlation matrices (p. 76). Similarly, microbial datasets tend to be sparse (i.e., have many variables with a value or abundance of 0); this is often the case for microbial datasets, particularly when using a taxonomy free approach (Martino et al., 2019). However, models have been developed to aid the analysis of zero inflated datasets (e.g., Zeng et al., 2022).

3. Inference networks

Windsor, Evans, and Dumbrell et al. all advise that networks can only be constructed on the basis of inferred organism interactions because microbial interactions cannot be directly observed (p. 37, 56, 74). Organism interactions are generally inferred based on patterns of species co-occurrence. Microbial interactions can be inferred from eDNA metabarcoding or metagenomic data (p. 74), with organism interactions inferred from presence/absence data or by correlation from species abundance data (p. 29). The think pieces highlighted some, albeit not many, studies that have developed inferred microbial networks in aquatic ecosystems (e.g., Widder et al., 2014).

The inference of interactions based on species co-occurrence or correlation has been the subject of significant criticism (Blanchet et al., 2020) and was described by Windsor as a significant limitation of taking a network approach (p. 28). For example, it has been shown that abundance correlations for microbes convey very limited information on network interactions for modelled microbial communities (p. 28). Current understanding of ecological interactions as represented by microbial co-occurrence or correlations is limited, meaning there is a missing link between species interactions and our understanding of mechanistic ecosystem function (p. 71).

There are a number of difficulties in inferring biotic interactions from microbial abundance or presence/absence data. Firstly, eDNA abundance data does not provide information on ecological interaction (p. 55). Secondly, studies have shown that microbial abundance correlations do not always match the direction of interactions (Pinto et al., 2022, p. 29). Another challenge is that species co-occurrence or correlation may not be a result of biotic interaction; instead, variation in species abundance or presence/absence may arise from habitat preference (p. 29). Furthermore, abiotic factors (e.g., flow, substrate, temperature, pH etc.) influences the concentration and degradation of eDNA in the environment, creating difficulty in interpreting the abundance or presence/absence of species from eDNA data (p. 55). In addition to this, the sparse and compositional nature of microbial datasets can present further issues, although there are solutions to help overcome this (e.g., Zang et al., 2020, p. 90-91). Studies have also highlighted issues with inferred network reproducibility (Barroso-Bergadà et al., 2021, p. 57).

Evans described microbial network science as being ‘in its infancy’ (p. 55) and advised caution when applying network analyses to NGS data generated by the RSN (p. 57). However, advances in network science that have addressed some of the problems with inference networks have been identified (synthesised in ‘4. Network inference methods’, p. 103). Additionally, Windsor highlighted that the spatial and temporal coverage of microbial

data collected by the RSN will help to ameliorate many of the concerns through sheer observational power (p. 37).

4. Network inference methods

Multiple methods for constructing ecological networks were described in the think pieces. These include correlation and regression approaches, graphical model inference, Bayesian and other statistical inference approaches, and machine learning. Dumbrell highlighted that these methods are mostly applicable to cross-sectional (i.e., spatial) data (p. 79) and Evans noted that many of these methods are rarely applied to microbial networks (p. 56).

While think piece authors recommended different approaches for inferring/constructing networks, both Evans and Dumbrell et al. advised testing and evaluating different network inference methods and data analysis pipelines (p. 56, 98). Evans noted that the RSN dataset is well suited to this due to its collection over spatial and temporal replicates (p. 56).

4.1 Graphical methods

Dumbrell et al. recommended the use of a graphical inference method, Sparse InversE Covariance estimation for Ecological Association and Statistical Inference (SPIEC-EASI), or newer methods that have built on this framework for inferring ecological networks from the RSN data, such as gCoda (p. 98).

Simpler, correlation-based methods often do not differentiate between direct and indirect microbial correlations (or associations). As such, graphical methods such as SPIEC-EASI have been developed to model conditional dependency and distinguish between direct and indirect interactions between organisms. These typically are more computational complex and have longer run times compared to correlation-based methods. The models create a unidirectional, weighted graph where edges imply the conditional dependency between two taxa (Kurtz et al., 2015; Matchado et al., 2021).

Firstly, SPIEC-EASI performs a transformation on the data to address compositionality issues with the data; this is required because OTU data is normalised to the count number meaning that microbial abundances are not independent of each other, and thus lead to spurious statistical results. An interaction graph is then estimated by one of two methods (either ‘Glasso’ or ‘Neighbourhood selection’) which effectively excludes OTUs that are correlated but not indirectly connected (Kurtz et al., 2015; Matchado et al., 2021).

SPIEC-EASI has been described as ‘a relatively robust method to infer microbiome networks’ (Birt and Dennis, 2021), and has been utilised by multiple published studies that have inferred microbial networks from SPIEC-EASI (e.g., Lam and Ye, 2022; Tipton et al., 2018). Dumbrell suggested that this is advantageous because it allows for the contextualisation of findings from the RSN (p. 98).

Other methods which take a similar, graphical approach to network inference have also been developed. gCoda was developed on the back of SPIEC-EASI and is reported to be more stable, accurate, and have a faster runtime compared to SPIEC-EASI. Other graphical methods include MDiNE, MixMPLN, NetComi, Environmental Driven Edge Detecton, Mint, mLDM, HARONIES, Hubs weighted graphical lasso, Flash Weave, and COZINE (Matchado et al., 2021). Each method offers slightly different advantages and have different limitations, which have been discussed further by Matchado et al., (2021).

4.2 Maximum Entropy

The Maximum Entropy (MaxEnt) method is an approach to network inference based on a statistical principle rooted in the principle of maximum entropy, whereby the probability distribution that 'best' fits a system or dataset is the one with the largest entropy. Effectively, this method determines the probability distribution from a dataset that is consistent with observations but is non-biased (De Martino and De Martino, 2018); MaxEnt has been shown to produce the least biased predictions of probability distributions consistent with prior knowledge of the constraints of those distributions (Harte and Newman, 2014). By selecting the most unbiased probability distribution that fits the dataset and known constraint, MaxEnt tends to avoid overfitting (Radosavljevic and Anderson, 2014).

This principle has been incorporated into multiple aspects of statistical modelling including niche modelling (Harte and Newman, 2014) and network inference (Caruso et al., 2022). MaxEnt is particularly useful in ecology and ecological network analysis because it allows for sparse or incomplete datasets to be analysed without making strong assumptions about data that is missing (De Martino and De Martino, 2019). Furthermore, it has been suggested that methods grounded in MaxEnt could help to differentiate between direct and indirect associations in networks (Menon et al., 2018 in Hirano and Takemoto., 2019).

Evans recommended 'Maximum Entropy' (MaxEnt) as a method of network inference to test on the NGS data (p.56). However, while MaxEnt approaches to network inferences have been developed for and applied to macro-ecology (e.g., networks of plants and pollinators; Caruso et al., 2022) and for the human microbiome (Li and Convertino, 2019), as have other approaches to network inference, there has been little application of MaxEnt to environmental NGS data.

4.3 Matrix Autoregression

Another network inference method that Evans suggested testing on the RSN data is Matrix Autoregression (p. 56). This method is typically applied to timeseries datasets and incorporates a species interaction matrix and a co-variate matrix capturing the impact of environmental drivers. A threshold is applied to identify significant interactions (Hampton et al., 2013). MAR models have been identified as particularly effective for studying interactions in dynamic communities (Hampton et al., 2013). Matrix autoregression is similar to multivariate or vector autoregression but looks at an entire matrix of variables instead of single time series variables.

As highlighted by Evans, MAR can be run with either single (e.g., Hampton et al., 2013) or multiple delays (i.e., time lags; e.g., Barraquand et al., 2021). Running the model with a single time point will analyse relationships between taxa at the current and immediately preceding timepoint, capturing more immediate dependencies and interactions in the data. Incorporating multiple delays into the MAR model allows exploration of longer term and perhaps more complex dependencies and interactions between taxa and environmental varies. MAR has been applied to studies looking at aquatic plankton community dynamics but has less readily been adopted in other aspects of ecology (e.g., Hampton et al., 2013). MAR modelling could be particularly useful in the analysis of the RSN data because it can be used to compute metrics of ecosystem stability (Hampton et al., 2013).

4.4 Joint Species Distribution Models

Windsor proposed the use of joint-species distribution models (J-SDM), specifically a type of model called ‘Hierarchical Modelling of Species Communities’ (HMSC), to explore the RSN data and infer ecological networks.

Joint-species distribution models are a type of species distribution model that can model the distribution of multiple species simultaneously (Pollock et al., 2014). They may be of value to this work because they take the presence of interactions between organisms into account (Ovaskainen et al., 2017). They combine habitat modelling and community ecology (Pollock et al., 2014) by incorporating spatially and temporally structured variables such as site-specific environmental conditions (temperature, water quality and quantity, pH), regional species pools and evolutionary histories (Tikhonov et al., 2020). Any remaining, unexplained co-occurrence is attributed to biotic interaction (Tikhonov et al., 2020). A study assessing the ability of J-SDMs to identify interactions found that they can identify mutualism and competition interactions of species well (Zurell et al., 2018).

HMSC is a Bayesian framework (available as an R package) that can integrate community ecology data with environmental covariates, species traits, phylogenetic relationships and the spatio-temporal context of the study to provide insight and predict multiple species distribution (Ovaskainen et al., 2017; Tikhonov et al., 2020) and estimate the percentage of community variability arising from different factors (Leite and Kuramae, 2020). Information on organism phylogeny can be incorporated from existing databases (e.g., rotl; Michonneau et al., 2016) by matching species taxonomy using metabarcoding data or by constructing phylogenetic trees based on sequenced gene regions (e.g., using BEAST; Drummond and Rambaut, 2007; Elias et al., 2013). Windsor advises that taking this approach would allow for the ‘most accurate estimations of ecological interactions’ because while the use of abundance correlations to construct microbial community networks has been heavily criticised, HMSC accounts for other abiotic and biotic influences making it more robust (p. 37). While HMSC has not been directly applied to microbial studies, hierarchical Bayesian joint distribution modelling has been developed and applied to microbial communities (e.g., Farrer et al., 2017; Yang et al., 2017; Bjork et al., 2018). Similar to other studies that have developed association networks using JSDM (e.g., Tikhonov et al., 2020), Windsor further suggests that merged networks of positive associations between OTUs could be created using HMSC (p. 38).

However, it would be challenging to apply J-SDMS to microbial data generated by the RSN. JSDMs are reliant on high quality data collected on appropriate temporal and spatial scales and a large amount of information on the target organisms, which is often not available for microbes (p. 30). Microbial communities usually have a higher proportion of rare taxa compared to macro-organisms communities. Furthermore, most taxa likely occur too sparsely to be modelled with a JSDM, although there are methods that can be used to help model rare taxa (Ruuskanen et al., 2021). J-SDMs also require a very high computer processing power. The diversity of microbe species in freshwater ecosystems would require computational power that is beyond the processing power that is readily available.

4.5 Machine learning and logic-based programming methods

Machine learning and logic-based programming were identified as methods for inferring and constructing ecological networks by Evans and Dumbrell (p. 59, 79) and may be applicable to the RSN data. Application of machine learning methods to network inference have been shown to reconstruct ecological networks as accurately as methods based on direct observations (Tamaddoni-Nezhad et al. 2013, p. 74). Similarly, machine learning methods have been able to learn ecological interactions from macro-organism species co-occurrence data when combined with background knowledge of the ecosystem (Bohan et al. 2011; Tamaddoni-Nezhad et al. 2013).

The major obstacle to the application of the machine learning methods to the RSN dataset is that they have not yet been applied extensively to microbial communities (Vacher et al., 2016), and even less so to freshwater microbial communities (p. 67). However, there has been significant development in the application of machine learning methods to microbial NGS data in the form of an R package called ‘Interaction Inference using Explainable Machine Learning’ (InfIntE; p. 79). InfIntE is described as a ‘generic, logic-based inference tool for learning networks in R’ and has been shown to perform well at identifying ecological interactions from microbial NGS data, as it is able to capture interactions between species and their forms (i.e., negative, positive, neutral; Barroso Bergadà 2022, p. 79).

Many artificial intelligence and machine learning methods and tools are underpinned by abductive and inductive logic programming respectively. Abductive logic programming derives the best hypothesis or explanation from a set of observations or dataset, whereas inductive logic programming derives a set of rules or principles from a dataset or specific examples. Dumbrell et al. suggested that as these methods develop, it could be advantageous to test out the use of AI and machine learning tools in inferring networks because of the methods’ increased capability to identify the form (positive, negative, neutral), magnitude, and directionality of interactions (Barros-Bergada et al., 2022, p.79).

4.6 Evaluating network inference methods

Both Evans and Dumbrell et al. suggested that different methods to network analysis should be applied to the RSN data and evaluated to determine the best method for inferring microbial networks (p. 56, p. 98). However, there is not extensive detail within the

think pieces as to how network inference methods should be evaluated. That being said, Dumbrell et al. highlighted a study by Faust and Reas (2012) that suggested that cultivation of unknown microorganisms, combinatorial labelling and parallel cultivation could facilitate systematic co-culturing and perturbation experiments, which can be linked into robustness analyses (see recommendations synthesis section 6.3). Evans suggested that mesocosm experiments could also be used to evaluate microbial networks, citing studies that have used mescosm experiments to look at microbial community response to copper contamination (Sutcliffe et al., 2019) and bacterial community response to warming (Yang et al., 2023; Evans, pers. comms, 02/05/2024).

5. Spatial multi-layer networks

Windsor recommended that multiple layers of ecological information and spatial information could be combined to better explore resilience on a river catchment scale, in what was termed a 'meta-community approach' (p. 39). The ability for network analysis to integrate multiple 'layers' of ecological information is considered by many, including the think piece authors, to be one of the strengths and benefits of a network approach. Windsor proposed that incorporating spatial 'rules' into the network (such as rules governing transport and dispersal distance of organisms) would allow for the resilience of individual sites in the context of the immigration and emigration of organisms and post-disturbance colonisation to be understood (p.39). Furthermore, Windsor suggested that different stressors could be simulated across the catchment, and that this approach could be tied into stability calculations (p. 40).

Evans highlighted the computational challenges that can be encountered when bringing multi-taxa microbial datasets into multi-layer networks due to the large number of interactions that can occur (p. 67).

6. Network interrogation and metrics

The think piece authors recommended or suggested several metrics or analyses that could be performed on networks inferred from the RSN data. This would allow for comparison of microbial networks and their properties between sites and through time.

6.1 Topological metrics

Topological metrics are related to the geometry and connectedness of a network and include metrics such as network modularity, connectivity, density and/or connectivity of links, network size, interaction evenness, and modularity. Both Windsor and Dumbrell acknowledge that the relationship between these network properties and ecosystem stability is complex with no consensus around the direction of relationship between metrics and ecosystem stability (p. 31, p. 81-83). This means that the interpretation of these metrics can be challenging. However, while Windsor recommends that this means these metrics are not particularly useful in analysis of the RSN data (p. 31), Dumbrell suggests

that they may be useful in exploring network responses to environmental perturbations (p. 99).

6.2 Identification of key nodes/hubs

Evans and Dumbrell et al. both suggested that key or highly connected taxa (nodes) or groups of taxa (hubs), which are important in the network structure (Banerjee et al. 2018, Fang et al. 2023) and may be involved in key ecological processes, could be identified in the networks inferred from the RSN data (p. 68, p. 99). However, while these nodes/hubs are important in the context of environmental networks, it is uncertain what they represent ecologically, for example, it is uncertain whether key nodes represent cornerstone species within an ecosystem (Banerjee et al. 2018; Röttjers & Faust 2019; Banerjee et al. 2019).

Dumbrell et al. suggested that following the identification of key taxa, complementary analyses could be used to explore these taxa further and develop a greater mechanistic understanding of their role in the ecosystem (p. 99). On a similar note, Evans suggested that it may be possible for future research to focus on these key hubs rather than generating entire networks (p. 68). The metagenomic data generated may also be helpful in determining likely functioning roles specific taxa perform.

6.3 Network robustness

Robustness is a measure of the tolerance of an ecological network to species extensions (i.e., removal of nodes; Dunne et al., 2002), and was identified as a possible avenue to explore in networks generated by the RSN data by Windsor (p. 38). Robustness analysis can also help to establish how network properties such as nestedness and connectance enhance or decrease resilience to loss of taxa/nodes and may be able to identify more less robust/sensitive groups within an ecosystem (Evans et al., 2013). Robustness analysis and metrics have previously been applied to a study looking at the robustness of a microbial community's functional profile to changes in community composition (Eng and Borenstein, 2018). Advances in methods for studying the robustness of ecological networks has facilitated the investigation of patterns across ecosystems and secondary extinctions (p. 61-62).

Windsor and Evans both illustrated how to assess the robustness of a network to extinctions (p. 31-33, 61-62). This involves removing a node from the network (primary extinction), identifying and subsequently removing any nodes that become separated from the network as a result (secondary extinction), and then repeating this until all nodes are lost from the network. This framework can be tailored to the context; from the data included to the inclusion of ecological knowledge (e.g., the likely order of species extinction), the introduction of thresholds for extinction based on the number of connections lost, and the introduction of interaction rewiring to simulate taxa interacting different taxa following an extinction. Evans noted that running robustness simulations on networks of any significant size requires large computer processing power (p. 68).

Evans suggested that adaptation network models could be used to investigate the how microbial networks respond to extinction events (p. 62-66). While robustness analysis tends to be concerned with the analysis stochastic networks (i.e., a snapshot of a network at a specific moment in time), adaptation network models include temporal dynamics Raimundo et al.. Specifically, adaptation network models include network rewiring, accounting for how species interactions adapt to species extinctions within a network (for example, predators may being to prey on other animals). Furthermore, by including network rewirings that we are likely to see in the real world, this model type may be able to help us identify or predict how a network may absorb or respond to environmental perturbations (Kaiser-Bunbury et al., 2010; Ramos-Jiliberto et al., 2012). However, this approach has not been previously applied to microbial interactions. Evans also highlighted that complexity-stability modelling could also be useful in simulating network response to species loss, and illustrated how this had been achieved through application of random matrix theory (p. 65). Evans suggests that there is scope in the future for developing freshwater 'vulnerability indices' by using adaptive network models and random matrix theory (p. 66).

6.4 Response diversity

Response diversity, a theory rooted in the insurance hypothesis of biodiversity (that greater biodiversity guarantees some organisms will maintain function even when others are lost; Yachi and Loreau, 1999), is described in Windsor's think piece as the range of potential responses or reactions of organisms to a given stressor or set of stressors (Elmqvist et al., 2023, p. 33). Windsor proposed that response diversity could be calculated for different environmental gradients represented in the RSN, specifically suggesting the use of the framework proposed by Ross et al., 2022 (p. 38-39). This is not a form of network analysis but would instead generate performance-environment curves for different taxon across different environmental conditions/gradients based on measured trait information from different species in the community.

At present this framework has only been applied to macro-ecological studies (e.g., White et al., 2023). However, Windsor noted that it is a flexible framework and could be coupled with biomonitoring data and secondary data, such as cell size for microbes (p. 34). It is possible for datasets to have multiple environmental variables within this framework and would therefore account for interacting stressors and interactive effects on the communities.

Windsor suggested that summaries of performance curves (provided by first derivatives of generalised additive models) can be used to calculate diversity using Hill numbers (p. 38). Diversity values could subsequently be compared across catchments/regions to understand the resilience of these areas to further change. It was also suggested that this approach could be applied to longitudinal data to calculate response diversity and combined with abiotic (water chemistry) data to calculate performance-environment curves for different (abiotic) drivers across the sites. Windsor suggested that further analysis of these performance-environment curves could provide site- and catchment-level insight into ecological resilience to future change (p. 38).

Windsor proposed that the outcomes of these analyses could be combined with aspects of ecological network analysis. For example, looking at the similarity of response between interlinked species could help to establish whether the co-occurrence of organisms is a result of their interaction (or not; p. 39). Similarly, outputs could be used to inform robustness simulations, whereby organisms with similar responses are lost in close succession.

6.5 Metrics linked to specific drivers and ecosystem function

Both Evans and Dumbrell et al. suggested that it could be beneficial to develop metrics linked to specific environmental perturbations and ecosystem services (p. 68, 99-100). Evans recommended developing metrics of ecosystem resilience linked to specific pressures, giving the example of analysing samples taken from RSN sites up and down stream of sewage outlets to measure the resilience of microbial networks to this perturbation (p. 68).

Microbial networks are known to underpin biogeochemical fluxes and nutrient transformations but relationships between the two are poorly understood. Dumbrell et al., suggested that it may be beneficial to link carbon and nitrogen cycling and other biogeochemical processes to microbial networks by taking direct measurements of these processes for a few key sites (p. 99). Additional to this, it was also suggested that targeted testing for key functional genes which are associated with ecosystem processes and functions is undertaken, as this may help provide insight into mechanisms and causal relationships between microbial networks and their properties and ecosystem processes and function. Dumbrell noted that this would require gene abundance data which can be obtained from the same DNA extracts as NGS and either linked to ecological network properties or included within networks.

Dumbrell stated that the real value of the RSN data would be if it was able to unravel mechanisms and causality and mechanistically link microbial communities and networks to ecosystem properties (p. 99-100). It was suggested that complementary analyses could be undertaken to do this; approaches invoking multivariate generalised linear modelling (Warton 2011; Wang et al. 2012) or structural equations (Mamet et al. 2019) may be able to do this (p. 97, p. 99).

6.6 Functional diversity

Windsor suggested that it may be beneficial to include information about the traits of organisms in (network) analyses because this can provide insight and used to develop metrics around functional diversity (Escalas et al., 2019, p. 36). This may also help add ecological realism to analyses such as robustness analysis.

Although trait data about micro-organisms is not as comprehensive as for macro-organisms, some trait data is available for diatoms and algae (e.g., Lange et al., 2016), bacteria (e.g., BactoTraits; Cébron et al., 2021, and PICRUSt2; Douglas et al., 2021), and fungi (e.g., FUNGuild; Nguyen et al., 2016, and FunFun; Krivonos et al., 2023).

7. Timeseries analysis

The recommendations outlined so far have focussed on the exploration and analysis of cross-sectional (i.e., spatial) data, however Dumbrell advised that it could be advantageous to undertake longitudinal (i.e., timeseries) data analysis as data is collected over time through the RSN (p. 80). Time series analysis of microbial ecology data will likely aid understanding of ecosystem dynamics, particularly because microbial communities can respond abruptly to perturbations in environmental conditions (Faust et al., 2015). This approach may be particularly helpful because it can help elucidate the causal direction of microbial responses (p. 98). Furthermore, it has also been suggested that interactions inferred through this approach may be more rigorous (p. 79).

Local Similarity Analysis (LSA) is valuable tool in understanding the varying dynamics of biological systems (Xia et al., 2013) and was identified by Dumbrell et al. as a possible LSA method to use in analysing the RSN data (p. 80, p. 98). LSA is able to identify the existence of local and lagged relationships and dependencies between taxa over short time periods from longitudinal data (Durno et al., 2013) and therefore the causal direction of responses (Cappellato et al., 2021). LSA was identified as a particularly applicable approach for analysing the RSN datasets because it is suitable for analysing sparse datasets (Weiss et al., 2016, p.80).

Dumbrell identified Extended Local Similarity Analysis (eLSA) as another possible method to apply to the RSN longitudinal data (p. 79, p. 98). eLSA was developed specifically for the analysis of microbial datasets, and has has been shown to identify statistically significant local and potentially time-delayed association patterns in replicated time series data (Xia et al., 2013). The incorporation of replicates facilitates understanding of the variability in local similarity metrics and obtain confidence intervals. As highlighted by Dumbrell et al., eLSA has been applied to numerous environmental microbial datasets to infer and analyse networks (e.g., Xia et al., 2011; Chow et al., 2013; Needham et al., 2017; Giner et al., 2018, Wang et al., 2021), thus providing a rich background of research against which to contextualise findings (p. 80).

An optimised version of eLSA, termed fastLSA has been developed and was also highlighted by Dumbrell et al. as a possible approach for the RSN data (p. 98). fastLSA was developed to perform LSA on big datasets. fastLSA takes a different approach in how statistical significance is assessed and calculated, which is much faster than preceding LSA methods (including eLSA). This means the computational time of fastLSA is much less because computationally intense permutation tests are not needed. Another difference is that fastLSA does not assume normality of the data, which is unlikely to be the case for ‘real-world’ datasets (Durno et al., 2013).

8. Complementary analyses

Dumbrell highlighted that a major challenge in the application of ENS to microbial ecology is the lack of mechanistic framework for the interpretation of microbial ecological networks (p. 95-96). This limits the interpretation of ecological network properties in terms of

ecological function. Dumbrell suggested that other, complementary analyses could aid microbial ecological network analysis. Specifically, they recommend the application of multivariate general linear modelling to understand community- and population-level responses to environmental pressures and establish causal relationships (e.g., Warton et al., 2011; Wang et al., 2012, p. 100), and structural equation modelling to establish causal relationships between environmental change, changes in the microbial network, and ecosystem properties (e.g., Mamet et al., 2019, p. 100).

9. Computing power and data storage

In his think piece, Evans recommends that requirements for storing data generated through this project are considered (p. 68). Metagenomic analysis generates huge volumes of data, with the volume of data generated by many projects employing NGS likely to exceed the storage capacity of individual hard drives (Hahn et al., 2016). Evans also highlighted that considerations should be made around about how data generated through the project is shared (p. 68). In an opinion piece addressing the infrastructure required a future in which large volumes of NGS data is produced, Hahn et al., 2016 proposed the use of cloud computing and distributed file storing systems as a solution to both storing large volumes of data and enabling access by multiple users; data is stored on the cloud and can then be accessed from and downloaded to personal devices. A number of publicly available online storage resources including the Earth Microbiome Project, the Joint Genome Institute's Genome Portal, and the National Centre for Biotechnology Information have been developed to enable large volumes of microbial data to be stored and publicly accessed (Hahn et al., 2016).

Evans also highlighted that some of the data analyses recommended (e.g., robustness analyses and simultaneous extinction modelling) will require large computer processing power and therefore access to High Power Computers is required (p. 68). Hahn et al., 2016 outlined that cloud-based computing could facilitate access to HPCs to analyse big microbial datasets as it allows individuals to remotely access to shared high-power computers from relatively basic personal computers. Cloud-based computing also others other benefits for data analysis, specifically for improving data reproducibility as it facilitates better documentation of how data has been manipulated and sharing of code (Hahn et al., 2016).

Next Steps

The think piece papers highlight the advantages and opportunities that ENS offers. Of particular value to the Environment Agency and the analysis of the RSN data is the potential it offers in furthering understanding of ecosystem function and dynamics and the development of metrics relating to these. Furthermore, it can be helpful in identifying important taxa or ‘hubs’ in ecosystem function, which could guide both future research and future monitoring. We therefore intend to experiment with ENS approaches and analysis pipelines in the analysis of the biofilm NGS data generated through the Environment Agency’s RSN.

The think pieces outline that microbial ecological networks must be inferred because interactions cannot be observed. There are multiple different methods for inferring networks; as advised by Evans and Dumbrell, we intend to test a number of different methods for inferring networks on the RSN data. Similarly, there are other aspects of the data analysis pipeline that we will experiment with, such as selecting to work with OTUs or ASVs, to find the optimum and most appropriate way to analyse the data.

The think piece authors highlight that microbial ENS is still in its infancy as a discipline and data analysis tool. We therefore intend to use other, more established data analysis methods alongside ENS and will compare the outputs from these methods.

References

ALBERT, R., JEONG, H., BARABÁSI, A.L., 2000. Error and attack tolerance of complex networks. Nature, 406, 378-382.

ALBERTSEN M., KARST, S.M., ZIEGLER, A.S., KIRKEGAARD, R.H., NIELSEN, P.H., 2015. Back to basics -The Influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. PLoS One 10, e0132783.

ALLESINA, S., TANG, S., 2015. The stability–complexity relationship at age 40: a random matrix perspective. Population Ecology, 57 (1) 63-75.

ALZARHANI, A.K., CLARK, D., UNDERWOOD, G.J.C., FORD, H., COTTON, T.E.A., DUMBRELL, A.J., 2019. Are drivers of root-associated fungal community structure context specific? The ISME Journal, 13, 1330-1344.

ARMITAGE, D.W., JONES, S.E., 2019. How sample heterogeneity can obscure the signal of microbial interactions. The ISME Journal, 13, 2639-2646.

ARROYO-CORREA, B., BARTOMEUS, I., JORDANO, P., 2021. Individual-based plant–pollinator networks are structured by phenotypic and microsite plant traits. Journal of Ecology 109, 2832-2844.

BAKSI, K.D., KUNTAL, B.K., MANDE, S.S., 2018. 'TIME': A web application for obtaining insights into microbial ecology using longitudinal microbiome data. Frontiers in Microbiology, 9, 36.

BALDASSANO, S., BASSETT, D., 2016. Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. Scientific Reports, 6, 26087.

BAN, Y., LINGLING, A., JIANG, H., 2015. Investigating microbial co-occurrence patterns based on metagenomic compositional data. Bioinformatics, 31, 3322-3329.

BANAVAR, J.R., HUBBELL, S.P., MARITAN, A., 2009. Inferring species interactions in tropical forests. Proceedings of the National Academy of Sciences of the United States of America, 106 (33), 13854-13859.

BANE, M.S., POCOCK, M.J.O., JAMES, R., 2018. Effects of model choice, network structure, and interaction strengths on knockout extinction models of ecological robustness. Ecology and evolution, 8 (22), 10794-10804.

BANERJEE, S., SCHLAEPPI, K., VAN DER HEIJDEN, M.G.A. 2019. Reply to 'Can we predict microbial keystones?'. Nature Reviews Microbiology 17, 194.

BANERJEE, S., SCHLAEPPI, K., VAN DER HEIJDEN, M.G.A., 2018. Keystone taxa as drivers of microbiome structure and functioning. Nature Reviews Microbiology, 16, 567–576.

BARABÁSI, A.-L., ALBERT, R., 1999. Emergence of Scaling in Random Networks. Science 286, 509-512.

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.

BARDGETT R.D., CARUSO T., 2020. Soil microbial community responses to climate extremes: resistance, resilience and transitions to alternative states. Philosophical Transactions of the Royal Society B 375, 20190112.

BARRAQUAND, F., PICOCHE, C., DETTO., M., HARTIG, F., 2021. Inferring species interactions using Granger causality and convergent cross mapping. Theoretical Ecology, 14 (1), 87-105.

BARROSO BERGADÀ, D., 2022. Automatic learning of interaction networks from next generation sequence data. PhD Thesis, University of Dijon.

BARROSO-BERGADA, D., PAUVERT, C., VALLANCE, J., DELIÈRE, L., BOHAN, D.A. BUÉE, M., VACHER, C., 2021. Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. Molecular Ecology Resources 21, 762-780.

BASCOMPTE, J., JORDANO, P., 2013. Monographs in Population Biology: Mutualistic Networks. New Jersey, USA: Princeton University Press.

BATTIN, T., BESEMER, K., BENGTSSON, M.., ROMANI, A.M., PACKMANN, A.I., 2016. The ecology and biogeochemistry of stream biofilms. Nature Reviews Microbiology, 14, 251-263.

BELL, T., NEWMAN, J., SILVERMAN, B., TURNER, S.L., LILLEY, A.K., 2005. The contribution of species richness and composition to bacterial services. Nature, 436, 1157-1160.

BENNETT, A.E., EVANS, D.M., POWELL, J.R., 2019. Potentials and pitfalls in the analysis of bipartite networks to understand plant–microbe interactions in changing environments. Functional ecology, 33 (1), 107-117.

BERLOW, E.L. DUNNE, J.A., MARTINEZ, N.D., STARK, P.B., WILLIAMS, R.J., BROSE, U., 2009. Simple prediction of interaction strengths in complex food webs. Proceedings of the National Academy of Sciences of the United States of America, 106, 187-191.

BERLOW, E.L., NEUTEL, A.-M., COHEN, J.E., DE RUITER, P.C., EBENMAN, B., EMMERSON, M., FOX, J.W., JANSEN, V.A.A., JONES, J.I., KOKKORIS, G.D., LOGOFET, D.O., MCKANE, A.J., MONTOYA, J.M., PETCHEY, O., 2004. Interaction Strengths in Food Webs: Issues and Opportunities. Journal of Animal Ecology 73, 585-598.

BERRY, D., WIDDER, S., 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in Microbiology 5, 219.

BERSIER, L.F., BANASEK-RICHTER, C., CATTIN, M.F., 2002. Quantitative descriptors of food-web matrices. Ecology, 83, 2394-2407.

BESEMER, K., 2015. Biodiversity, community structure and function of biofilms in stream ecosystems. Research in microbiology, 166 (10), 774-781.

BESSON, M., ALISON, J., BJERGE, K., GOROCHOWSKI, T.E., HØYE, T.T., JUCKER, T., MANN, H.M.R., CLEMENTS, C.F., 2022. Towards the fully automated monitoring of ecological communities. Ecology Letters 25, 2753-2775.

BJÖRK, J.R.K., HUI, F.K.C., O'HARA, R.B., MONTOYA, J.M. 2018. Uncovering the drivers of host-associated microbiota with joint species distribution modelling. Molecular Ecology,

BLANCHET, F.G., CAZELLES, K., GRAVEL, D., 2020. Co-occurrence is not evidence of ecological interactions. Ecology Letters, 23, 1050-1063.

BLONDER, B., WEY, T.W., DORNHAUS, A., JAMES, R., SIH., A., .2012. Temporal dynamics and network analysis. Methods in Ecology and Evolution, 3 (6), 958-972.

BODIN, Ö., ALEXANDER, S.M., BAGGIO, J., BARNES, M.L., BERARDO, R., CUMMING, G.S., DEE, L.E., FISCHER, A.P., FISCHER, M., MANCILLA GARCIA, M., GUERRERO, A.M., HILEMAN, J., INGOLD, K., MATOUS, P., MORRISON, T.H., NOHRSTEDT, D., PITTMAN, J., ROBINS, G., SAYLES, J.S., 2019. Improving network approaches to the study of complex social–ecological interdependencies. Nat Sustain, 2, 551-559.

Bohan, D.A., VACHER, C., TAMADDONI-NEZHAD, A., RAYBOULD, A., DUMBRELL., A.J., WOODWARD, G., 2017. Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. Trends in Ecology & Evolution, 32 (7), 477-487.

BOHAN, D.A., CARON-LORMIER, G., MUGGLETON, S., RAYBOULD, A., TAMADDONI-NEZHAD, A., 2011. Automated discovery of food webs from ecological data using logic-based machine learning. PLoS ONE 6, e29028.

BOHAN, D.A., VACHER, C., TAMADDONI-NEZHAD, A., RAYBOULD, A., DUMBRELL, A.J., WOODWARD, G., 2017. Next-Generation Global Biomonitoring: Large-scale, Automated Reconstruction of Ecological Networks. Trends in Ecology & Evolution 32, 477-487.

BOHMANN, K., ELBRECHT, V., CARØE, C., BISTA, I., LEESE, F., BUNCE, M., YU, D.W., SEYMOUR, M., DUMBRELL, A.J., CREER, S., 2022. Strategies for sample labelling and library preparation in DNA metabarcoding studies. Molecular Ecology Resources 22, 1231-1246.

BORRETT, S.R., MOODY, J., EDELMANN, A., 2014. The rise of Network Ecology: Maps of the topic diversity and scientific collaboration. Ecological Modelling, 293, 111-127.

BORRVALL, C., EBENMAN, B., 2006. Early onset of secondary extinctions in ecological communities following the loss of top predators. Ecology letters, 9 (4), 435-442.

BORRVALL, C., EBENMAN, B., TOMAS JONSSON, T.J., 2000. Biodiversity lessens the risk of cascading extinction in model food webs. Ecology letters, 3 (2), 131-136.

BROSE, U., DUNNE, J.A, HALL, R.O., HLADYZ, S.,KITCHING, R.L., MARTINEZ, N.D.,RANTALA, H., ROMANUK, T.N., STOUFFER, D.B., TYLIANAKIS, J.M., 2012. Food webs: reconciling the structure and function of biodiversity. Trends in Ecology & Evolution, 27 (12), 689-697.

BROWN, J.A., ROBERTSON, B.L., MCDONALD, T., 2015.Spatially Balanced Sampling: application to environmental surveys. Procedia Environmental Sciences, 27, (6-9).

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.

BULDYREV, S.V., PARSHANI, R., PAUL., G., STANLEY, H.E., HAVLIN., S., 2010. Catastrophic cascade of failures in interdependent networks. Nature, 464 (7291), 1025-1028.

CALLAHAN, B.J., MCMURDIE, P.J., ROSEN, M.J., HAN, A.W., JOHNSON, A.J., HOLMES, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13, 581-583.

CAMERON, E.S., SCHMIDT, P.J., TREMBLAY, B.J.-M., EMELKO, M.B., MÜLLER, K.M., 2021. Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. Scientific Reports, 11, 22302.

CAPPELLATO, M., BARUZZO, G., PATUZZI, I., DI CAMILLO, B., 2021. ModeLling Microbial Community Networks: Methods and Tools. Current Genomics, 22, 267-290.

CARDINALE, B.J., DUFFY, J.E., GONZALEZ, A., HOOPER, D.U., PERRINGS, C., VENAIL, P., NARWANI, A., MACE, G.M., TILMAN, D., WARDLE, D.A., KINZIG, A.P., DAILY, G.C., LOREAU, M., GRACE, J.B., LARIGAUDERIE, A., SRIVASTAVA, D.S., NAEEM, S. 2012. Biodiversity loss and its impact on humanity. Nature, 486, 59-67.

CARDINALE, B.J., WRIGHT, J.P., CADOTTE, M.W., CARROLL, I.T., HECTOR, A., SRIVASTAVA, D.S., LOREAU, M., WEIS, J.J., 2007. Impacts of plant diversity on biomass production increase through time because of species complementarity. Proceedings of the National Academy of Sciences, 104, 18123-18128.

CARDONA, C., WEISENHORN, P., HENRY, C., GILBERT, J.A., 2016. Network-based metabolic analysis and microbial community modelling. Current Opinion in Microbiology, 31, 124-131.

CARUSO, T., CLEMENTE, G.V., RILLIG, M.C., GARLASCHELLI, D., 2022. Fluctuating ecological networks: A synthesis of maximum-entropy approaches for pattern detection and process inference. Methods in Ecology and Evolution, 13 (11), 2306-2317.

CAZELLES, K., ARAÚJO, M.B., MOUQUET, N., GRAVEL., D., 2016. A theory for species co-occurrence in interaction networks. Theoretical Ecology, 9 (1), 39-48.

CÉBRON, A., ZEGHAL, E., USSEGLIO-POLATERA, P., MEYER, A., BAUDA, P., LEMMEL, F., LEYVAL, C., MAUNOURY-DANGER, F., 2021. BactoTraits – A functional trait database to evaluate how natural and man-induced changes influence the assembly of bacterial communities. Ecological Indicators, 130, 108047.

CHEN, W., WANG, J., CHEN, X., MENG, Z., XU, R., DUOJI, D., ZHANG, J., HE, J., WANG, Z., CHEN, J., LIU, K., HU, T., ZHANG, Y., 2022. Soil microbial network complexity predicts ecosystem function along elevation gradients on the Tibetan Plateau. Soil Biology and Biochemistry, 172, 108766.

CHIARELLO, M., MCCAULEY, M., VILLÉGER, S., JACKSON, C.R., 2022. Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. PLoS ONE, 17, e0264443.

CHOW, C.-E.T., SACHDEVA, R., CRAM, J.A., STEELE, J.A., NEEDHAM, D.M., PATEL, A., PARADA, A.E., FUHRMAN, J.E., 2013. Temporal variability and coherence of euphotic zone bacterial communities over a decade in the Southern California Bight. The ISME Journal, 7, 2259-2273.

CLARE, E.L., ECONOMOU, C.K., BENNETT, F.J., DYER, C.E., ADAMS, K., MCROBIE, B., DRINKWATER, R., LITTLEFAIR, J.E., 2022. Measuring biodiversity from DNA in the air. Current Biology, 32 (3), 693-700.

CLARE, E.L., FAZEKAS, A.J., IVANOVA, N.V., FLOYD, R.M., HEBERT, P.D.N., ADAMS, A.M., NAGEL, J., GIRTON, R., NEWMASTER, S.G., FENTON, M.B., 2019. Approaches to integrating genetic data into ecological networks. Molecular Ecology 28, 503-519.

CLARK, DR., FERGUSON, RMW., HARRIS, DN., MATTHEWS NICHOLASS, KJ., PRENTICE, HJ., RANDALL, KC., RANDELL, L., WARREN, SL., DUMBRELL, AJ., 2018. Streams of data from drops of water: 21st century molecular microbial ecology. WIREs Water, 5, e1280.

CODELLO, A., HOSE, G.C., CHARITON, A., 2023. Microbial co-occurrence networks as a biomonitoring tool for aquatic environments: a review. Marine and Freshwater Research, 74, 409-422.

Cohen, J.E., SCHITTLER, D.,N., RAFFAELLI., D.G., REUMAN, D.C., 2009. Food webs are more than the sum of their tritrophic parts. Proceedings of the National Academy of Sciences of the United States of America, 106 (52), 22335-22340.

CORDIER, T., ALONSO-SÁEZ, L., APOTHÉLOZ-PERRET-GENTIL, L., AYLAGAS, E., BOHAN, D.A., BOUCHEZ, A., CHARITON, A., CREER, S., FRÜHE, L., KECK, F., KEELEY, N., LAROCHE, O., LEESE, F., POCHON, X., STOECK, T., PAWLOWSKI, J., LANZÉN, A., 2020. Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Molecular Ecology, 30, 2937-2958.

CORDIER, T., LANZÉN, A., APOTHÉLOZ-PERRET-GENTIL, L., STOECK, T., and PAWLOWSKI, J. (2019). Embracing environmental genomics and machine learning for routine biomonitoring. Trends in Microbiology, 27, 387–397. doi: 10.1016/j.tim.2018.10.012.

CORDIER, T., ANGELES, I.B., HENRY, N., LEJZEROWICZ, F., BERNEY, C., MORARD, R., BRANDT, A., CAMBON-BONAVITA, M.A., GUIDI, L., LOMBARD, F., ARBIZU, P.M., 2022. Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. Science Advances, 85, p.eabj9309.

COSTA, A., MARTÍN GONZÁLEZ, A.M., GUIZIEN, K., DOGLIOLI, A.M., GÓMEZ, J.M., PETRENKO, A.A., ALLESINA, S., 2019. Ecological networks: Pursuing the shortest path, however narrow and crooked. Scientific Reports, 9, 17826.

COYTE, K.Z., SCHLUTER, J., FOSTER, K.R., 2015. The ecology of the microbiome: Networks, competition, and stability. Science, 350, 663-666.

CUFF, J.P., KITSON, J.J.N., HEMPRICH-BENNETT, D., TERCEL, M.P.T.G., BROWETT, S.S., EVANS, D.M., 2023. The predator problem and PCR primers in molecular dietary analysis: Swamped or silenced; depth or breadth? Molecular Ecology Resources, 23, 41–51.

CUFF, J.P., WINDSOR, F.M., TERCEL, M.P.T.G., KITSON, J.J.N., EVANS, D.M., 2022. Overcoming the pitfalls of merging dietary metabarcoding into ecological networks. Methods in Ecology and Evolution, 13, 545–559.

CURTIS, A.N. TIEMANN, J.S., DOUGLASS, S.A., DAVIS, M.A., LARSON, E.R., 2021. High stream flows dilute environmental DNA (eDNA) concentrations and reduce detectability. Diversity & distributions, 27 (10), 1918-1931.

CZECH, L., STAMATAKIS, A., DUNTHORN, M., BARBERA, P., 2022. Metagenomic Analysis Using Phylogenetic Placement-A Review of the First Decade. Frontiers in Bioinformatics, 2, 871393.

DAHLLÖF, I., 2002. Molecular community analysis of microbial diversity. Current Opinion in Biotechnology, 13, 213-217.

DE ANGELIS, D.L., 1975. Stability and Connectance in Food Web Models. Ecology, 56, 238-243.

DE MARTINO, A.,DE MARTINO, D., 2018. An introduction to the maximum entropy approach and its application to inference problems in biology. Heliyon, 4, e00596.

DE VOS, M.G.J., ZAGORSKY, M., MCNALLY, A., BOLLENBACH, T., 2017. Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections. Proceedings of the National Academy of Sciences, 114, 10666-10671.

DE VRIES, F.T., GRIFFITHS, R.I., BAILEY, M., CRAIG H., GIRLANDA M., GWEON H.S., HALLIN S., KAISERMANN A., KEITH A.M., KRETZSCHMAR M., BARDGETT R.D., 2018. Soil bacterial networks are less stable under drought than fungal networks. Nature Communications, 9, 3033.

DEAGLE, B.E., THOMAS, A.C., SHAFFER, A.K., TRITES, A.W., JARMAN, S.N., 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? Molecular Ecology Resources, 13, 620-633.

DELMAS, E., BESSON, M., BRICE, M.-H., BURKLE, L.A., DALLA RIVA, G.V., FORTIN, M.-J., GRAVEL, D., GUIMARÃES JR., P.R., HEMBRY, D.H., NEWMAN, E.A., OLESEN, J.M., PIRES, M.M., YEAKEL, J.D., POISOT, T., 2019. Analysing ecological networks of species interactions. Biological Reviews, 94, 16-36.

DENG, Y., JIANG, Y.H., YANG, Y., HE, Z., LUO, F., ZHOU, J., 2012. Molecular ecological network analyses. BMC Bioinformatics, 13, 113.

DENG, Y., ZHANG, P., QIN, Y., TU, Q., YANG, Y., HE, Z., SCHADT, C.W., ZHOU, J., 2016. Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. Environmental Microbiology, 18, 205-218.

DEROCLES, S.A., BOHAN, D.A., DUMBRELL, A.J., KITSON, J.J., MASSOL, F., PAUVERT, C., PLANTEGENEST, M., VACHER, C., EVANS, D.M., (2018) Biomonitoring for the 21st century: integrating next-generation sequencing into ecological network analysis. Advances in Ecological Research, 58, 1-62.

DEROCLES, S.A.P., EVANS, D.M., NICHOLS, P.C., EVANS, S.A., LUNT, D.H., 2015. Determining plant–leaf miner–parasitoid interactions: a DNA barcoding approach. PloS one, 10 (2), e0117872.

DEVOTO, M., BAILEY, S., CRAZE, P., MEMMOTT, J., 2012. Understanding and planning ecological restoration of plant–pollinator networks. Ecology Letters, 15, 319-328.

DINI-ANDREOTE, F., STEGEN, J.C., VAN ELSAS, J.D., SALLES, J.F. 2015. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. Proceedings of the National Academy of Sciences, 112, 1326-1332.

DIRZO, R., YOUNG, H.S., GALETTI, M., CEBALLOS, G., ISAAC, N.J.B., COLLEN, B., 2014. Defaunation in the Anthropocene. Science, 345, 401-406.

DONOHUE, I., PETCHEY, O.L., MONTOYA, J.M., JACKSON, A.L., MCNALLY, L., VIANA, M., HEALY, K., LURGI, M., O'CONNOR, N.E., EMMERSON, M.C., 2013. On the dimensionality of ecological stability. Ecology Letters, 16, 421–429.

DONOHUE, I.. HILLEBRAND, H., MONTOYA, J.M., PETCHEY, O.L., PIMM, S.L., FOWLER, M.S., HEALY, K., JACKSON, A.L., LURGI, M., MCCLEAN, M., O'CONNOR, N.E., O'GORMAN, E.J., YANG, Q., 2016. Navigating the complexity of ecological stability. Ecology letters, 19 (9), 1172-1185.

DORMANN, C.F., FRÜND, J., BLÜTHGEN, N., GRUBER, B., 2009. Indices, graphs and null models: analysing bipartite ecological networks. The Open Ecology Journal, 2, 7-24.

DOUGLAS, G.M., MAFFEI, V.J., ZANEVELD, J.R., YURGEL, S.N., BROWN, J.R., TAYLOR, C.M., HUTTENHOWER, C., LANGILLE, M.G., 2020. PICRUSt2 for prediction of metagenome functions. Nature Biotechnology, 38, 685-688.

DRAKE, L.E., CUFF, J.P., YOUNG, R.E., MARCHBANK, A., CHADWICK, E.A., SYMONDSON, W.O.C., 2022. An assessment of minimum sequence copy thresholds for identifying and reducing the prevalence of artefacts in dietary metabarcoding data. Methods in Ecology and Evolution, 13, 694-710.

DRUMMOND, A.J., RAMBAUT, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7, 214.

DUBART, M., ALONSO, P., BARROSO-BERGADA, D., BECKER, N., BETHUNE, K., BOHAN, D.A., BOURY, C., CAMBON, M., CANARD, E., CHANCEREL, E. AND CHIQUET, J., 2021. Coupling ecological network analysis with high-throughput sequencing-based surveys: Lessons from the next-generation biomonitoring project. Advances in Ecological Research, 65, 367-430.

DUMBRELL A.J., NELSON, M., HELGASON, T., DYTHAM, C., FITTER, A.H., 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. The ISME Journal, 4, 337-345.

DUMBRELL, A.J., CLARK, E.J., FROST, G.A., RANDELL, T.E., PITCHFORD, J.W., HILL, J.K. 2008. Changes in species diversity following habitat disturbance are dependent on spatial scale: theoretical and empirical evidence. Journal of Applied Ecology, 45, 1531-1539.

DUMBRELL, A.J., FERGUSON, R.M.W., CLARK, D.R., 2017. Hydrocarbon and Lipid Microbiology Protocols. Berlin, Heidelberg: Springer.

DUNNE, J.A., WILLIAMS, R.J., 2009. Cascading extinctions and community collapse in model food webs. Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 1711-1723.

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.

DURNO, W.E., HANSON, N.W., KONWAR, K.M., HALLAM, S.J., 2013. Expanding the boundaries of local similarity analysis. BMC Genomics 14, S3.

EBENMAN, B., LAW, R., BORRVALL, C., 2004. Community viability analysis: The response of ecological communities to species loss. Ecology, 85 (9), 2591-2600.

EKLÖF, A., TANG, S., ALLESINA, S. 2013. Secondary extinctions in food webs: a Bayesian network approach. Methods in Ecology and Evolution, 4 (8), 760-770.

ELIAS, M., FONTAINE, C., FRANK VAN VEEN, F.J., 2013. Evolutionary History and Ecological Processes Shape a local multilevel antagonistic network. Current Biology, 23, 1355-1359.

ELMQVIST, T., FOLKE, C., NYSTRÖM, M., PETERSON, G., BENGTSSON, J., WALKER, B., NORBERG, J., 2003. Response diversity, ecosystem change, and resilience. Frontiers in Ecology and the Environment, 1, 488-494.

ELTON, C.S., 1929. The Ecological Relationships of Certain Freshwater Copepods. Journal of Ecology, 17, 383-391.

EMARY, C., EVANS, D.M., 2021. Can a complex ecosystem survive the loss of a large fraction of its species? A random matrix theory of secondary extinction. Oikos, 130 (9), 1512-1522.

ENG, A., BORENSTEIN, E. 2018. Taxa-function robustness in microbial communities. Microbiome, 6 (1), 45.

ERŐS, T., LOWE, W.H., 2019. The Landscape Ecology of Rivers: from Patch-Based to Spatial Network Analyses. Curr Landscape Ecol Rep, 4, 103–112.

ESCALAS, A., HALE, L., VOORDECKERS, J.W., YANG, Y., FIRESTONE, M.K., ALVAREZ-COHEN, L., ZHOU, J., 2019. Microbial functional diversity: From concepts to applications. Ecology and Evolution, 9, 12000-12016.

EVANS, D.M., KITSON, J.J., 2020. Molecular ecology as a tool for understanding pollination and other plant–insect interactions. Current Opinion in Insect Science, 38, 26-33.

EVANS, D.M., KITSON, J.J.N., LUNT, D.H., STRAW, N.A., POCOCK, M.J.O., 2016. Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. Functional Ecology, 30, 1904-1916.

EVANS, D.M., POCOCK, M.J.O., MEMMOTT, J., 2013. The robustness of a network of ecological networks to habitat loss. Ecology Letters, 16, 844-852.

FALKOWSKI, P.G., FENCHEL, T., DELONG, E.F., 2008. The microbial engines that drive earth's biogeochemical cycles. Science, 320, 1034-1039.

FANG, H., HUANG, C., ZHAO, H., DENG, M., 2015. CCLasso: correlation inference for compositional data through Lasso. Bioinformatics, 31, 3172-3180.

FANG, H., HUANG, C., ZHAO, H., DENG, M., 2017. gCoda: Conditional Dependence Network Inference for Compositional Data. Journal of Computational Biology, 24, 699-708.

FANG, W., FAN, T., XU, L., WANG, S., WANG, X., LU, A., CHEN, Y., 2023. Seasonal succession of microbial community co-occurrence patterns and community assembly mechanism in coal mining subsidence lakes. Frontiers in Microbiology, 14, 1098236.

FARRER, E.C., PORAZINSKA, D.L., SPASOJEVIC, KING, A.J., BUENO ED MESQUITA, C.P., SARTWELL, S.A., M.J., SMITH, J.G., WHITE, C.T., SCHMIDT, S.K., SUDING, K.N., 2017. Soil Microbial Networks Shift Across a High-Elevation Successional Gradient. Frontiers in Microbiology, 10, 2887.

FAUST, K., 2021. Open challenges for microbial network construction and analysis. The ISME Journal, 15, 3111-3118.

FAUST, K., LAHTI, L., GONZE, D., DE VOS, W.M., RAES, J., 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current Opinion in Microbiology, 25, 56-66.

Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nature Reviews Microbiology, 10 (8), 538-550.

FAUST, K., SATHIRAPONGSASUTI, J.F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J., HUTTENHOWER, C., 2012. Microbial co-occurrence relationships in the human microbiome. PLoS Computational Biology, 8, e1002606.

FAVILA, N., MADRIGAL-TREJO, D., LEGORRETA, D., SÁNCHEZ-PÉREZ, J., ESPINOSA-ASUAR, L., EGUIARTE, L.E., SOUZA, V., 2022. MicNet toolbox: Visualizing and unraveling a microbial network. PLoS ONE, 17, e0259756.

FELIPE-LUCIA, M.R., GUERRERO, A.M., ALEXANDER, S.M., ASHANDER, J., BAGGIO, J.A., BARNES, M.L., BODIN, Ö., BONN, A., FORTIN, M.-J., FRIEDMAN, R.S., GEPHART, J.A., HELMSTEDT, K.J., KEYES, A.A., KROETZ, K., MASSOL, F., POCOCK, M.J.O., SAYLES, J., THOMPSON, R.M., WOOD, S.A., DEE, L.E., 2022. Conceptualizing ecosystem services using social–ecological networks. Trends in Ecology & Evolution, 37, 211-222.

FELL, S.C., CARRIVICK, J.L., CAUVY-FRAUNIÉ, S., CRESPO-PÉREZ, V., HOOD, E., RANDALL, K.C., MATTHEWS NICHOLASS, K.J., TIEGS, S.D., DUMBRELL, A.J., BROWN, L.E., 2021. Fungal decomposition of river organic matter accelerated by decreasing glacier cover. Nature Climate Change, 11, 349-353.

FENG, K., PENG, X., ZHANG, Z., GU, S., HE, Q., SHEN, W., WANG, Z., WANG, D., HU, Q., LI, Y., WANG, S., DENG, Y., 2022. iNAP: An integrated network analysis pipeline for microbiome studies. iMeta, 1, e13.

FENG, K., ZHANG, Y., HE, Z., NING, D., DENG, Y., 2019. Interdomain ecological networks between plants and microbes. Molecular Ecology Resources, 19, 1565-1577.

FENG, K., ZHANG, Z., CAI, W., LIU, W., XU, M., YIN, H., WANG, A., HE, Z., DENG, Y. , 2017. Biodiversity and species competition regulate the resilience of microbial biofilm community. Molecular Ecology, 26, 6170-6182.

FERDOUS, T., JIANG, L., DINU, I., GROIZELEAU, J., KOZYRSKYJ, A.L., GREENWOOD, C.M.T., ARRIETA, M.-C., 2022. The rise to power of the microbiome: power and sample size calculation for microbiome studies. Mucosal Immunology, 15, 1060-1070.

FERGUSON, R.M.W., O'GORMAN, E.J., MCELROY, D.J., MCKEW, B.A., COLEMAN, R.A., EMMERSON, M.C., DUMBRELL, A.J., 2021. The ecological impacts of multiple environmental stressors on coastal biofilm bacteria. Global Change Biology, 27, 3166-3178.

FISHER, C.K., MEHTA, P. 2014. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLOS ONE, 9, e102451.

FLEMMING, H.-C., WINGENDER, J., SZEWZYK, U., STEINBERG, P., RICE, S.A., KJELLEBERG, S., 2016. Biofilms: an emergent form of bacterial life. Nature reviews Microbiology 14, 563-575.

FORTUNA, M.A., STOUFFER, D.B., OLESEN, J.M., JORDANO, P., MOUILLOT, D., KRASNOV, B.R., POULIN, R., BASCOMPTE, J., 2010. Nestedness versus modularity in ecological networks: two sides of the same coin? Journal of Animal Ecology, 79, 811-817.

FOWLER, M.S., 2010. Extinction cascades and the distribution of species interactions. Oikos, 119 (5), 864-873.

FRIEDMAN, J., ALM, E.J. 2012. Inferring correlation networks from genomic survey data. PLoS Computational Biology, 8, e1002687.

GALIANA, N., BARROS, C., BRAGA, J., FICETOLA, G.F., MAIORANO, L., THUILLER, W., MONTOYA, J.M., LURGI, M., 2021. The spatial scaling of food web structure across European biogeographical regions. Ecography, 44, 653-664.

GALIANA, N., LURGI, M., CLARAMUNT-LÓPEZ, B., FORTIN, M.-J., LEROUX, S., CAZELLES, K., GRAVEL, D., MONTOYA, J.M., 2018. The spatial scaling of species interaction networks. Nature Ecology & Evolution, 2, 782-790.

GHANNAM, R.B., TECHTMANN, S.M., 2021. Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Computational and Structural Biotechnology Journal, 19, 1092-1107.

GINER, C.R., BLAGÚE, KRABBERØD, A.K., FERRERA, I., REÑE, A., GARCÉS, E., GASOL, J.M., LOGARES, R., MASSANA, R., 2018. Quantifying long-term recurrence in planktonic microbial eukaryotes. Molecular Ecology, 28 (5), 923-935.

GLOOR, G.B., MACKLAIM, J.M., PAWLOWSKY-GLAHN, V., EGOZCUE, J.J., 2017. Microbiome datasets are compositional: And this is not optional. Frontiers in Microbiology, 8, 2224.

GOLBECK, J., 2013. Analyzing the social web. Burlington, Massachusetts: Morgan Kaufmann.

GOŁĘBIEWSKI, M., TRETYN, A., 2020. Generating amplicon reads for microbial community assessment with next-generation sequencing. Journal of Applied Microbiology, 128, 330-354.

GÓMEZ, J.M., PERFECTTI, F., 2011. Fitness consequences of centrality in mutualistic individual-based networks. Proceedings of the Royal Society B: Biological Sciences 279, 1754-1760.

GONZALEZ, J.M., PORTILLO, M.C., BELDA-FERRE, P., MIRA, A., 2012 Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. PLOS ONE 7, e29973.

GRAHAM, N.R., KREHENWINKEL, H., LIM, J.Y., STANICZENKO, P., CALLAGHAN, J., ANDERSEN, J.C., GRUNER, D.S., GILLESPIE, R.G., 2023. Ecological network structure in response to community assembly processes over evolutionary time. Molecular Ecology, 32 (23), 6489-6506.

GRAY, C., BAIRD, D.J., BAUMGARTNER, S., JACOB, U., JENKINS, G.B., O'GORMAN, E.J., LU, X., MA, A., POCOCK, M.J.O., SCHUWIRTH, N., THOMPSON, M., WOODWARD, G., 2014. Ecological networks: the missing links in biomonitoring science. Journal of Applied Ecology, 51, 1444-1449.

GRAY, C., FIGUEROA, D.H., HUDSON, L.N., MA, A., PERKINS, D., WOODWARD, G., 2015. Joining the dots: An automated method for constructing food webs from compendia of published interactions. Food Webs, 5, 11-20.

GRENNI, P., 2022. Antimicrobial Resistance in Rivers: A Review of the Genes Detected and New Challenges. Environmental Toxicology and Chemistry, 41, 687-714.

GRILLI, J., ROGERS, T., ALLESINA, S., 2016. Modularity and stability in ecological communities. Nat Commun, 7, 12031.

GUIMARÃES, P.R., 2020. The Structure of Ecological Networks Across Levels of Organization. Annual Review of Ecology, Evolution, and Systematics, 51, 433-460.

GUIMARÃES, P.R., PIRES, M.P., JORDANO, P., BASCOMPTE J., THOMPSON, J.N., 2017. Indirect effects drive coevolution in mutualistic networks. Nature, 550 (7677), 511–514.

GUSEVA, K., DARCY, S., SIMON, E., ALTEIO, L., MONTESINOS-NAVARRO, A., KAISER, C., 2022. From diversity to complexity: Microbial networks in soils. Soil Biology and Biochemistry, 169, 108604.

HAHN, A.S., KONWAR, K.M., LOUCA, S., HANSON, N.W., HALLAM, S.J., 2016. The information science of microbial ecology. Current Opinion in Microbiology, 31, 209-216.

HAMILTON, M., FISCHER, A.P., AGER, A., 2019. A social-ecological network approach for understanding wildfire risk governance. Global Environmental Change, 54, 113–123.

HAMPTON, S.E., HOLMES, E.E., SCHEEF, L.P., SCHEUERELL, M.D., KATZ, S.L., PENDLETON, D.E., WARD, E.J. 2013. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. Ecology, 94 (12), 2663-2669.

HARRISON, J.B., SUNDAY, J.M. ROGERS, S.M. 2019. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings of the Royal Society B, 286 (1915), 20191409.

HARTE, J., NEWMAN, E.A., 2014. Maximum information entropy: a foundation for ecological theory. Trends in Ecology and Evolution, 29 (7), 384-389.

HARVEY, E., GOUNAND, I., WARD, C.L., ALTERMATT, F., 2017. Bridging ecology and conservation: from ecological networks to ecosystem function. Journal of Applied Ecology, 54, 371-379.

HECTOR, A., BAGCHI, R., 2007. Biodiversity and ecosystem multifunctionality. Nature, 448, 188-190.

HECTOR, A., SCHMID, B., BEIERKUHNLEIN, C., CALDEIRA, M. C., DIEMER, M., DIMITRAKOPOULOS, P. G., FINN, J. A., FREITAS, H., GILLER, P. S., GOOD, J., HARRIS, R., HÖGBERG, P., HUSS-DANELL, K., JOSHI, J., JUMPPONEN, A., KORNER, C., LEADLEY, P.W., LOREAU, M., MINNS, A., MULDER, C.P., O'DONOVAN, G., OTWAY, S.J., PEREIRA, J.S., PRINZ, A., READ, D.J., SCHERER-LORENZEN, M., SCHULZE, E.D., SIAMANTZIOURAS, A.S.D., SPEHN, E.M., TERRY, A.C., TROUMBIS, A.Y., WOODWARD, F.I., YACHI, S., LAWTON, J.H., 1999. Plant diversity and productivity experiments in European grasslands. Science, 286, 1123-1127.

HELMUS, M.R., BLAND, T.J., WILLIAMS, C.K., IVES, A.R., 2007. Phylogenetic Measures of Biodiversity. The American Naturalist, 169, 68-83.

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.

HERREN, C.M., MCMAHON, K.D., 2017. Cohesion: a method for quantifying the connectivity of microbial communities. The ISME Journal, 11, 2426-2438.

HIGINO, G.T., BANVILLE, F., DANSEREAU, G., MUÑOZ, N.R.F., WINDSOR, F., POISOT, T., 2023. Mismatch between IUCN range maps and species interactions data illustrated using the Serengeti food web. PeerJ, 11, e14620.

HIRANO, H., TAKEMOTO, K., 2019. Difficulty in inferring microbial community structure based on co-occurrence network approaches. BMC Bioinformatics, 20, 329.

HOLLING, C.S., 1973. Resilience and Stability of Ecological Systems. Annual Review of Ecology and Systematics, 4, 1-23.

HOLLING, C.S., 1996. Engineering Within Ecological Constraints. Washington D.C., USA : National Academies Press.

HUI, C., RICHARDSON, D., 2022. Invading Ecological Networks, Ecology, Biodiversity and Conservation. Cambridge, UK: Cambridge University Press.

HUTCHINSON, M.C., BRAMON MORA, B., PILOSOF, S., BARNER, A.K., KÉFI, S., THÉBAULT, E., JORDANO, P., STOUFFER, D.B., 2019. Seeing the forest for the trees: Putting multilayer networks to work for community ecology. Functional Ecology, 33, 206-217.

INGS, T.C., MONTOYA, J.M., BASCOMPTE, J., BLÜTHGEN, N., BROWN, L., DORMANN, C.F., EDWARDS, F., FIGUEROA, D., JACOB, U., JONES, J.I., LAURIDSEN, R.B., LEDGER, M.E., LEWIS, H.M., OLESEN, J.M., VAN VEEN, F.J.F., WARREN, P.H., WOODWARD, G., 2009. Ecological networks – beyond food webs. Journal of Animal Ecology, 78, 253-269.

ISAAC, N.J.B., BROTHERTON, P.N.M., BULLOCK, J.M., GREGORY, R.D., BOEHNING-GAESE, K., CONNOR, B., CRICK, H.Q.P., FRECKLETON, R.P., GILL, J.A., HAILS, R.S., HARTIKAINEN, M., HESTER, A.J., MILNER-GULLAND, E.J., OLIVER, T.H., PEARSON, R.G., SUTHERLAND, W.J., THOMAS, C.D., TRAVIS, J.M.J., TURNBULL, L.A., WILLIS, K., WOODWARD, G., MACE, G.M., 2018. Defining and delivering resilient ecological networks: Nature conservation in England. Journal of Applied Ecology, 55, 2537-2543.

ISAAK, D.J., LUCE, C.H., RIEMAN, B.E., NAGEL, D.E., PETERSON, E.E., HORAN, D.L., PARKES, S., CHANDLER, G.L., 2010. Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network. Ecological Applications, 20, 1350-1371.

ISHIMOTO, C.K., AONO, A.H., NAGAI, J.S., SOUSA, H., MIRANDA, A.R.L., MELO, V.M.M., MENDES, L.W., ARAUJO, F.F., DE MELO, W.J., KUROSHU, R.M., ESPOSITO, E., ARAUJO, A.S.F., 2021. Microbial co-occurrence network and its key microorganisms in

soil with permanent application of composted tannery sludge. Science of the Total Environment, 789, 147945.

JAKUSCHKIN, B., FIEVET, V., SCHWALLER, L., FORT, T., ROBIN, C., VACHER, C., 2016. Deciphering the Pathobiome: Intra- and Interkingdom Interactions Involving the Pathogen Erysiphe alphitoides. Microbial ecology, 72 (4), 870-880.

JAX, K., 2005. Function and 'functioning' in ecology: what does it mean? Oikos, 111 (3), 641-648.

JENSEN, H.J., 2022. Complexity Science: The Study of Emergence. Cambridge, UK: Cambridge University Press.

JO, T., MURAKAMI, H., YAMAMOTO, S., MASIDA, R., MINAMOTO, T., 2019. Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. Ecology and E volution, 9 (3), 1135-1146.

JOHNSON, M.D., FREELAND, J.R., PARDUCCI L., EVANS, D.M, MEYER, R.S., MOLANO-FLORES, B., DAVIS, M.A., 2023. Environmental DNA as an emerging tool in botanical research. American Journal of Botany, 110 (2), e16120.

JORDANO, P., 2016a. Chasing Ecological Interactions. PLOS Biology, 14, e1002559.

JORDANO, P., 2016b. Sampling networks of ecological interactions. Functional Ecology, 30, 1883-1893.

JU, F., ZHANG, T. 2015. Bacterial assembly and temporal dynamics in activated sludge of a full-scale municipal wastewater treatment plant. The ISME Journal, 9, 683-695.

KADOYA, T., MCCANN, K., 2015. Weak Interactions and Instability Cascades. Scientific Reports, 5, 12652.

KAISER-BUNBURY, C.N., MUFF, S., MEMMOTT, J., MÜLLER, C.B., CAFLISCH, A., 2010. The robustness of pollination networks to the loss of species and interactions: a quantitative approach incorporating pollinator behaviour. Ecology Letters, 13, 442-452.

KALLMEYER, J., POCKALNY, R., ADHIKARI, R.R., SMITH, D.C., D'HONDT, S., 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. Proceedings of the National Academy of Sciences, 109 (40), 16213-16216.

KANAGARAJ, R., WIEGAND, T., KRAMER-SCHADT, S., GOYAL, S.P., 2013. Using individual-based movement models to assess inter-patch connectivity for large carnivores in fragmented landscapes. Biological Conservation, 167, 298-309.

KANERYD,L., BORRVALL, C., BERG, S., CURTSDOTTER, A., EKLÖF, A., HAUZY, C., JONSSON, T., MÜNGER, P., SETZER, M., SÄTERBERG, T., EBENMAN, B., 2012. Species-rich ecosystems are vulnerable to cascading extinctions in an increasingly variable world. Ecology and Evolution, 2 (4), 858-874.

KAPLAN-SHABTAI, V., INDUGU, N., HENNESSY, M.L., VECCHIARELLI, B., BENDER, J.S., STEFANOVSKI, D., DE ASSIS LAGE, C.F., RÄISÄNEN, S.E., MELGAR, A., NEDELKOV, K., FETTER, M.E., FERNANDEZ, A., SPITZER, A., HRISTOV, A.N., PITTA, D.W. 2021. Using Structural Equation Modeling to Understand Interactions Between Bacterial and Archaeal Populations and Volatile Fatty Acid Proportions in the Rumen. Frontiers in Microbiology, 12, 611951.

KELLY, M.G., CAZAUBON, A., CORING, E., DELL'UOMO, A., ECTOR, L., GOLDSMITH, B., GUASCH, H., HÜRLIMANN, J., JARLMAN, A., KAWECKA, B., KWADRANS, J., LAUGASTE, R., LINDSTRØM, E.-A., LEITAO, M., MARVAN, P., PADISÁK, PIPP, E., PRYGEIL, J., ROTT, E., SABATER, S., VAN DAM, H., VIZINET, J., 1998. Recommendations for the routine sampling of diatoms for water quality assessments in Europe. Journal of Applied Phycology, 10, 215-224.

KELLY, B.J., GROSS, R., BITTINGER, K., SHERRILL-MIX, S., LEWIS, J.D., COLLMAN, R.G., BUSHMAN, F.D., LI, H., 2015. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. Bioinformatics, 31, 2461-2468.

KERMORVANT, C., D'AMICO, F., BRU, N.M., CAILL-MILLY, N., ROBERTSON, B., 2016. Spatially balance sampling designs for environmental surveys. Environmental Monitoring and Assessment, 191, 524.

KESSLER, E.J., ASH, K.T., BARRATT, S.N., LARSON, E.R., DAVIS, M.A.,2020. Radiotelemetry reveals effects of upstream biomass and UV exposure on environmental DNA occupancy and detection for a large freshwater turtle. Environmental DNA , 2 (1), 13-23.

KEYES, A.A. MCLAUGHLIN, J.P., BARNER, A.K., DEE, L.E., 2021. Author Correction: An ecological network approach to predict ecosystem service vulnerability to species losses. Nature communications, 12 (1), 5843.

KRIVONOS, D.V., KONANOV, D.N., ILINA, E.N., 2023. FunFun: ITS-based functional annotator of fungal communities. Ecology and Evolution, 13 (3), e9874.

KRUK, M., PATUREJ, E., 2020. Indices of trophic and competitive relations in a planktonic network of a shallow, temperate lagoon. A graph and structural equation modeling approach. Ecological Indicators, 112 (12), 106007.

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.

LALIBERTÉ, E., NIELSEN, A., BASCOMPTE, J., 2010. Conservation of species interaction networks. Biological Conservation, 143 (10), 2270-2279.

LAM T.J., YE, Y., 2022. Meta-analysis of microbiome association networks reveal patterns of dysbiosis in diseased microbiomes. Scientific Reports, 12, 17482.

LANDI, P., MINOARIVELO, H.O., BRÄNNSTRÖM, Å., HUI, C., DIECKMANN, U., 2018. Complexity and stability of ecological networks: a review of the theory. Popul Ecol, 60, 319-345.

LANGE, K., TOWNSEND, C.R., MATTHAEI, C.D., 2016. A trait-based framework for stream algal communities. Ecology and Evolution, 6, 23-36.

LANGFELDER, P., HORVATH, S., 2007. Eigengene networks for studying the relationships between co-expression modules. BMC Systems Biology, 1, 54.

LANGFELDER, P., HORVATH, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 29, 559.

LAU, M.K., BORRETT, S.R., BAISER, B., GOTELLI, N.J., ELLISON, A.M., 2017. Ecological network metrics: opportunities for synthesis. Ecosphere, 8 (8), e01900.

LAWTON, J.H., BROTHERTON, P.N.M., BROWN, V.K., ELPHICK, C., FITTER, A.H., FORSHAW, J., HADDOW, R.W., HILBORNE, S., LEAFE, R.N., MACE, G.M., SOUTHGATE, M.P., SUTHERLAND, W.J., TEW, T.E., VARLEY, J., WYNNE, G.R., 2010. Making space for nature. Report to Defra.

LEARY, D.J., PETCHEY, O.L., 2009. Testing a biological mechanism of the insurance hypothesis in experimental aquatic communities. Journal of Animal Ecology, 78, 1143-1151.

LEHTOVIRTA-MORLEY, L.E., 2018. Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. FEMS Microbiology Letters, 365 (9), fny058.

LEITE, M.F.A., KURAMAE, E.E., 2020. You must choose, but choose wisely: Model-based approaches for microbial community analysis. Soil Biology and Biochemistry, 151, 108042.

LEWIS, H.M., LAW, R., 2007. Effects of dynamics on ecological networks. Journal of Theoretical Biology 247, 64-76. https://doi.org/10.1016/j.jtbi.2007.02.006

LI, CONVERTINO, 2019

LIN, Y., ZHAO, D., ZENG, J., CAO, X., JIAO, C. 2019. Network analysis reveals seasonal patterns of bacterial community networks in Lake Taihu under aquaculture conditions. Water, 11, 1868.

LIND, B.M., CANDIDO-RIBEIRO, R., SINGH, P., LU, M., VIDAKOVIC, D.O., BOOKER, T.R., WHITLOCK, M.C., YEAMAN, S., ISABEL, N., AITKEN, S.N., (In press). How useful is genomic data for predicting maladaptation to future climate?

LIU Z., MA, A., MATHÉ, E., MERLING, M., MA, Q., LIU, B., 2021a. Network analyses in microbiome based on high-throughput multi-omics data., Briefings in Bioinformatics, 22, 1639-1655.

LIU, C., CUI, Y., LI, X., YAO, M., 2021b. microeco: an R package for data mining in microbial community ecology. FEMS Microbiology Ecology, 97, fiaa255.

LIU, Y., ZENG, M., XIE, Z., NING, D., ZHOU, J., YU, X., LIU, R., ZHANG, L., FANG, J., 2022. Microbial community structure and ecological networks during simulation of diatom sinking. Microorganisms, 10, 639.

LIU, Z., WEI, H., ZHANG, J., SALEEM, M., HE, Y., ZHONG, J., MA, R. 2021c. Higher sensitivity of soil microbial network than community structure under acid rain. Microorganisms, 9, 118.

LO, C., MARCULESCU, R., 2017a. MPLasso: Inferring microbial association networks using prior microbial knowledge. PLoS Computational Biology, 13, e1005915.

LO, C., MARCULESCU, R., 2017b. Inferring microbial interactions from metagenomic time-series using prior biological knowledge. Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics 2017., 168-177.

LOCEY, K.J., LENNON, J.T., 2016. Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences, 113, 5970-5975.

LU, X., GRAY, C., BROWN, L.E., LEDGER, M.E., MILNER, A.M., MONDRAGÓN, R.J., WOODWARD, G., MA, A., 2016. Drought rewires the cores of food webs. Nature Clim Change, 6, 875-878.

LUGO-MARTINEZ, J., RUIZ PEREZ, D., NARASIMHAN, G., BAR-JOSEPH, Z, 2019. Dynamic interaction network inference from longitudinal microbiome data. Microbiome, 7, 54.

LUNDBERG, P., RANTA, E., KAITALA, V. 2008. Species loss leads to community closure. Ecology letters, 3 (6), 465-468.

LUO, F., YANG, Y., ZHONG, J., GAO, H., KHAN, L., THOMPSON, D.K., ZHOU, J. 2007. Constructing gene co-expression networks and predicting functions of unknown genes by random matrix theory. BMC Bioinformatics, 8, 299.

LUO, F., ZHONG, J., YANG,Y., SCHEUERMANN, R.H., ZHOU, J. 2006. Application of random matrix theory to biological networks. Physics Letters A, 357, 420-423.

MA, B., WANG, H., DSOUZA, M., LOU, J., HE, Y., DAI, Z., BROOKES, P.C., XU, J., GILBERT, J.A., 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. The ISME Journal, 10, 1891-1901.

MACARTHUR, R., 1955. Fluctuations of Animal Populations and a Measure of Community Stability. Ecology, 36, 533-536.

MACGREGOR, C.J., EVANS, D.M., POCOCK, M.J.O., In press. Estimating sampling completeness of interactions in quantitative bipartite ecological networks: incorporating variation in species' specialisation. bioRxiv.

MAIA, K.P., MARQUITTI, F.M.D., VAUGHAN, I.P., MEMMOTT, J., RAIMUNDO, R.L.G., 2021. Interaction generalisation and demographic feedbacks drive the resilience of plant–insect networks to extinctions. Journal of Animal Ecology, 90, 2109-2121.

MAMET, S.D., REDLICK, E., BRABANT, M., LAMB, E.G., HELGASON, B.L., STANLEY, K., SICILIANO, S.D., 2019. Structural equation modeling of a winnowed soil microbiome identifies how invasive plants re-structure microbial networks. The ISME Journal, 13, 1988-1996.

MARTINO, C., MORTON, F.T., MAROTZ, C.A., THOMPSON, L.R., TRIPATHI, A., KNIGHT, R., ZENGLER, K., 2019. A Novel Sparse Compositional Technique Reveals Microbial Perturbations. Ecological and Evolutionary Science, 4 (1).

MATCHADO M.S., LAUBER M., REITMEIER S., KACPROWSKI T., BAUMBACH J., HALLER D., LIST, M., 2021. Network analysis methods for studying microbial communities: A mini review. Computational and Structural Biotechnology Journal, 19, 2687-2698.

MAY, R.M., 1972. Will a Large Complex System Be Stable. Nature, 238, 413–141.

MAY, R.M., 1974. Stability and Complexity in Model Ecosystems. 2nd ed. New Jersey, USA: Princeton University Press.

MAY, R.M., 1982. Mutualistic interactions among species. Nature, 296, 803-804.

MCCANN, K., HASTINGS, A., HUXEL, G. 1998. Weak trophic interactions and the balance of nature. Nature, 395, 794-798.

MCCANN, K.S., 2000. The diversity–stability debate. Nature 405, 228-233.

MCGREGOR, K., LABBE, A., GREENWOOD, C.M.T., 2020. MDiNE: a model to estimate differential co-occurrence networks in microbiome studies. Bioinformatics, 36, 1840-1847.

MCKNIGHT, D.T., HUERLIMANN, R., BOWER, D.S., SCHWARZKOPF, L., ALFORD, R.A., ZENGER, K.R. 2019. Methods for normalizing microbiome data: An ecological perspective. Methods in Ecology Evolution, 10, 389-400.

MCMURDIE P.J., HOLMES, S., 2014. Waste not, want not: Why rarefying microbiome data is inadmissible. PLOS Computational Biology, 10, e1003531.

MEMMOTT, J., 1999. The structure of a plant-pollinator food web. Ecology letters, 2 (5), 276-280.

MEMMOTT, J., CRAZE, P.G., WASER, N.M., PRICE, M.V., 2007. Global warming and the disruption of plant-pollinator interactions. Ecology letters, 10 (8), 710-717.

MEMMOTT, J., WASER, N.M., PRICE, M.V. 2004. Tolerance of pollination networks to species extinctions. Proceedings Royal Society B, 271 (1557), 2605–2611.

MERCADO, J.V., KOYAMA, M., NAKASAKI, K., 2022. Co-occurrence network analysis reveals loss of microbial interactions in anaerobic digester subjected to repeated organic load shocks. Water Research, 221, 118754.

MICHONNEAU, F., BROWN, J.W., WINTER, D.J., 2016. rotl: an R package to interact with the Open Tree of Life data. Methods in Ecology and Evolution 7, 1476-1481.

MILLER, K.E., POLASZEK, A., EVANS, D.M., 2021. A dearth of data: fitting parasitoids into ecological networks. Trends in Parasitology, 37, 863-874.S

MILNER, A.M., CONN, S.C., BROWN, L.E., 2006. Persistence and stability of macroinvertebrate communities in streams of Denali National Park, Alaska: implications for biological monitoring. Freshwater Biology, 51, 373-387.

MONTOYA, J.M., PIMM, S.L., SOLÉ, R.V., 2006. Ecological networks and their fragility. Nature, 442 (7100), 259-264.

MORALES-CASTILLA, I., MATIAS, M.G., GRAVEL, D., ARAÚJO, M.B., 2015. Inferring biotic interactions from proxies. Trends in Ecology & Evolution, 30, 347-356.

MORRIS, R.L., TALE, V.P., MATHAI, P.P., ZITOMER, D.H., MAKI, J.S., 2016. mcrA Gene abundance correlates with hydrogenotrophic methane production rates in full-scale anaerobic waste treatment systems. Letters in Applied Microbiology, 62, 111-118.

MORRISSEY, M.B., FERGUSON, M.M., 2011. Individual variation in movement throughout the life cycle of a stream-dwelling salmonid fish. Molecular Ecology 20, 235-248.

MOUGI, A., KONDOH, M., 2012. Diversity of Interaction Types and Ecological Community Stability. Science, 337, 349-351.

MOUTON, T.L., TONKIN, J.D., STEPHENSON, F., VERBURG, P., FLOURY, M., 2020. Increasing climate-driven taxonomic homogenization but functional differentiation among river macroinvertebrate assemblages. Global Change Biology, 26, 6904-6915.

MUGGLETON, S., DE RAEDT, L., 1994. Inductive Logic Programming: Theory and methods. The Journal of Logic and Algebraic Programming, 19-20, 629-679.

NEEDHAM, D.M., SACHDEVA, R., FUHRMAN, J.A., 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. The ISME Journal 11, 1614-1629.

NEIDEL, V., SINT, D., WALLINGER, C., TRAUGOTT, M., 2022. RNA allows identifying the consumption of carrion prey. Molecular Ecology Resources, 22, 2662-2671.

NEUTEL, A.-M., HEESTERBEEK, J.A.P., DE RUITER, P.C., 2002. Stability in Real Food Webs: Weak Links in Long Loops. Science, 296, 1120-1123.

NEUTEL, A.-M., HEESTERBEEK, J.A.P., VAN DE KOPPEL, J., HOENDERBOOM, G., VOS, A., KALDEWAY, C., BERENDSE, F., DE RUITER, P.C., 2007. Reconciling complexity with stability in naturally assembling food webs. Nature, 449, 599-602.

NEWMAN, M. E.J., 2018. Networks. Oxford, UK: Oxford University Press.

NEWMAN, M.E.J., 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103, 8577-8582.

NGUYEN, N.H., SONG, Z., BATES, S.T., BRANCO, S., TEDERSOO, L., MENKE, J., SCHILLING, J.S., KENNEDY, P.G. 2016. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology, 20, 241-248.

NIELSEN, A., BASCOMPTE, J., 2007. Ecological Networks, Nestedness and Sampling Effort. Journal of Ecology, 95, 1134-1141.

OLESEN, J.M., BASCOMPTE, J., DUPONT, Y.L., ELBERLING, H., RASMUSSEN, C., JORDANO, P., 2010. Missing and forbidden links in mutualistic networks. Proceedings of the Royal Society B: Biological Sciences 278, 725-732.

OVASKAINEN, O., ABREGO, N., HALME, P., DUNSON, D., 2016.Using latent variable models to identify large networks of species-to-species associations at different spatial scales. Method in Ecology and Evolution, 7 (5), 549-555.

OVASKAINEN, O., TIKHONOV, G., NORBERG, A., BLANCHET, F.G., DUAN, L.D., DUNSON, D., ROSLIN, T., ABREGO, N., 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology letters, 20 (5), 561-576.

PAILY, O., SHANKAR, V., 2016. Application of multivariate statistical techniques in microbial ecology. Molecular Ecology 25, 1032-1057.

PARRA, S.A., THÉBAULT, E., FONTAINE, C., DAKOS, V., 2022. Interaction fidelity is less common than expected in plant–pollinator communities. Journal of Animal Ecology, 91, 1842-1854.

PAULO R. GUIMARÃES, P.R., OLESEN, J.M., THOMPSON, J.N., 2014. Assembly of complex plant–fungus networks. Nature Communications, 5, 5273.

PENESYAN, A., PAULSEN, I.T., KJELLEBERG, S., GILLINGS, M.R., 2021. Three faces of biofilms: a microbial lifestyle, a nascent multicellular organism, and an incubator for diversity. Npj Biofilms and Microbiomes, 7, 80.

PERALTA, G., VÁZQUEZ, D.P., CHACOFF, N.P., LOMÁSCOLO, S.B., PERRY, G.L.W., TYLIANAKIS, J.M., 2020. Trait matching and phenological overlap increase the spatio-

temporal stability and functionality of plant–pollinator interactions. Ecology Letters, 23, 1107-1116.

PERES-NETO, P.R., 2004. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. Oecologia, 140, 352-360.

PESCHEL, S., MÜLLER, C.L., VON MUTIUS, E., BOULESTEIX, A.-L., DEPNER, M., 2021. NetCoMi: network construction and comparison for microbiome data in R. Briefings in Bioinformatics, 22, 1-18.

PETANIDOU, T., KALLIMANIS, A.S., TZANOPOULOS, J., SGARDELIS, S.P., PANTIS, J.D., 2008. Long-term observation of a pollination network: fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization. Ecology Letters, 11, 564-575.

PETERSON, E.E., VER HOEF, J.M., ISAAK, D.J., FALKE, J.A., FORTIN, M.-J., JORDAN, C.E., MCNYSET, K., MONESTIEZ, P., RUESCH, A.S., SENGUPTA, A., SOM, N., STEEL, E.A., THEOBALD, D.M., TORGERSEN, C.E., WENGER, S.J., 2013. Modelling dendritic ecological networks in space: an integrated network perspective. Ecology Letters, 16, 707-719.

PICHLER, M., BOREUX, V., KLEIN, A.-M., SCHLEUNING, M., HARTIG, F., 2020. Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. Methods in Ecology and Evolution, 11, 281-293.

PIERELLA KARLUSICH, J. J., PELLETIER, E., ZINGER, L., LOMBARD, F., ZINGONE, A., COLIN, S., GASOL, J. M., DORRELL, R. G., HENRY, N., SCALCO, E., ACINAS, S. G., WINCKER, P., DE VARGAS, C., BOWLER, C., 2022. A robust approach to estimate relative phytoplankton cell abundances from metagenomes. Molecular Ecology Resources, 23, 16-40.

PILLAR, V.D., BLANCO, C.C., MUELLER, S.C., SOSINSKI, E.E., JONER, F., DUARTE, L.D.S., 2013. Functional redundancy and stability in plant communities. Journal of Vegetation Science, 24 (5), 963-974.

PILOSOF, S., PORTER, M.A., PASCUAL, M., KÉFI, S., 2017. The multilayer nature of ecological networks. Nat Ecol Evol 1, 1-9.

PIMM, S.L., 1984. The complexity and stability of ecosystems. Nature, 307, 321-326.

PIMM, S.L., DONOHUE, I., MONTOYA, J.M., LOREAU, M., 2019. Measuring resilience is essential to understand it. Nat Sustain, 2, 895-897.

PINTO, S., BENINCÀ, E., NES, E.H. VAN, SCHEFFER, M., BOGAARDS, J.A., 2022. Species abundance correlations carry limited information about microbial network interactions. PLOS Computational Biology, 18, e1010491.

PIRES, M.M., O'DONNELL, J.L., BURKLE, L.A., DÍAZ-CASTELAZO, C., HEMBRY, D.H., YEAKEL, J.D., NEWMAN, E.A., MEDEIROS, L.P., DE AGUIAR, M.A.M., GUIMARÃES, P.R., 2020. The indirect paths to cascading effects of extinctions in mutualistic networks. Ecology, 101 (7), e03080.

POCOCK, M.J.O., EVANS, D.M., FONTAINE, C., HARVEY, M., JULLIARD, R., MCLAUGHLIN, Ó., SILVERTOWN, J., TAMADDONI-NEZHAD, A., WHITE, P.C.L., BOHAN, D.A., 2016. Advances in Ecological Research, Ecosystem Services: From Biodiversity to Society. Massachusetts, USA: Academic Press, 41-85.

POCOCK, M.J.O., EVANS, D.M., MEMMOTT, J., 2012. The Robustness and Restoration of a Network of Ecological Networks. Science, 335, 973-977.

POCOCK, M.J.O., SCHMUCKI, R., BOHAN, D.A., 2021. Inferring species interactions from ecological survey data: A mechanistic approach to predict quantitative food webs of seed feeding by carabid beetles. Ecology and Evolution, 11, 12858-12871.

POISOT, T., MOUQUET, N., GRAVEL, D., 2013. Trophic complementarity drives the biodiversity–ecosystem functioning relationship in food webs. Ecology Letters, 16, 853-861.

POISOT, T., STOUFFER, D.B., GRAVEL, D., 2014. Beyond species: Why ecological interaction networks vary through space and time. Oikos, 124, 243-251.

POLLOCK, 2014

POMERANZ, J.P.F., THOMPSON, R.M., POISOT, T., HARDING, J.S., 2019. Inferring predator–prey interactions in food webs. Methods in Ecology and Evolution, 10, 356-367

PORATH-KRAUSE, A., STRAUSS, A.T., HENNING, J.A., SEABLOOM, E.W., BORER, E.T., 2022. Pitfalls and pointers: An accessible guide to marker gene amplicon sequencing in ecological applications. Methods in Ecology and Evolution, 13, 266-277.

POTÉ, J., ACKERMANN, R., WILDI, W., 2009. Plant leaf mass loss and DNA release in freshwater sediments. Ecotoxicology and Environmental Safety, 72 (5), 1378-1383.

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 Keystones: Identifying keystone species is difficult—but essential to understanding how loss of species will affect ecosystems. BioScience, 46, 609-620.

PRICE, G.W., LANGILLE, M.G., YURGEL, S.N., 2021. Microbial co-occurrence network analysis of soils receiving short-and long-term applications of alkaline treated biosolids. Science of the Total Environment, 751, 141687.

PRODAN, A., TREMAROLI, V., BROLIN, H., ZWINDERMAN, A.H., NIEUWDORP, M., LEVIN, E., 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS ONE, 15, e0227434.

PROSSER, J.I., 2020. Putting science back into microbial ecology: a question of approach. Philosophical Transactions of the Royal Society B, 375, 20190240.

PROSSER, J.I., BOHANNAN, B.J., CURTIS T.P., ELLIS, R.J., FIRESTONE, M.K., FRECKLETON, R.P., GREEN, J.L., GREEN, L.E., KILLHAM, K., LENNON, J.J., OSBORN, A.M., SOLAN, M., VAN DER GAST, C.J., YOUNG, J.P. 2007. The role of ecological theory in microbial ecology. Nature Reviews Microbiology, 5, 384-392.

QUINCE, C., HIGGS, P.G., MCKANE, A.J., 2005. Deleting species from model food webs'. Oikos, 110 (2), 283-296.

QUINN, T.P. ERB, I., RICHARDSON, M.F., CROWLEY, T.M., 2018. Understanding sequencing data as compositions: an outlook and review. Bioinformatics, 34, 2870-2878.

RAFFERTY, N.E., IVES, A.R., 2013. Phylogenetic trait-based analyses of ecological networks. Ecology, 94 (10), 2321-2333.

RAIMUNDO, R.L.G., GUIMARÃES, P.R., EVANS, D.M., 2018. Adaptive Networks for Restoration Ecology. Trends in Ecology & Evolution, 33, 664-675.

RAMIREZ K.S., GEISEN S., MORRIËN E., SNOEK B.L., VAN DER PUTTEN W.H. 2018. Network Analyses Can Advance Above-Belowground Ecology. Trends in Plant Science 23, 759-768.

RAMMETTE, A., 2007. Multivariate analyses in microbial ecology. FEMS Microbiology Ecology, 62 (2), 142-160.

RAMOS-JILIBERTO, R., VALDOVINOS, F.S., DE ESPANÉS, P.M., FLORES, J.D. 2012. Topological plasticity increases robustness of mutualistic networks. The Journal of Animal Ecology, 81 (4), 896-904.

RATZKE C., BARRERE, J., GORE, J. 2020. Strength of species interactions determines biodiversity and stability in microbial communities. Nature Ecology and Evolution, 4, 376-383.

RAYFIELD, B., FORTIN, M.-J., FALL, A., 2011. Connectivity for conservation: a framework to classify network measures. Ecology, 92, 847-858.

REZENDE, E.L., LAVABRE, J.E., GUIMARÃES, P.R., JORDANO, P., BASCOMPTE, J., 2007. Non-random coextinctions in phylogenetically structured mutualistic networks. Nature, 448, 925-928.

RIDENHOUR, B., BROOKER, S., WILLIAMS, J., VAN LEUVEN, J.T., MILLER, A.W., DEARING, M.D., REMIEN, C.H., 2017. Modelling time-series data from microbial communities. The ISME Journal, 11, 2526-2537.

ROHR, R.P., SAAVEDRA, S., BASCOMPTE, J., 2014. On the structural stability of mutualistic systems. Science, 345 (6195), 1253497.

ROMAN, V.L., MERLIN, C., VIRTA, M.P., BELLANGER, X., 2021. EpicPCR 2.0: Technical and Methodological Improvement of a Cutting-Edge Single-Cell Genomic Approach. Microorganisms, 9, 1649.

ROMANUK, T.N., ZHOU, Y., BROSE, U., BERLOW, E.L., WILLIAMS, R.J., MARTINEZ, N.D., 2009. Predicting invasion success in complex ecological networks. Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 1743-1754.

ROSS, S.R.P.-J., PETCHEY, O.L., SASAKI, T., ARMITAGE, D.W., 2022. How to measure response diversity. Methods in Ecology and Evolution, 14 (5), 1150-1167.

ROSS, S.R.P.-J., SUZUKI, Y., KONDOH, M., SUZUKI, K., VILLA MARTÍN, P., DORNELAS, M., 2021. Illuminating the intrinsic and extrinsic drivers of ecological stability across scales. Ecological Research, 36, 364-378.

RÖTTJERS, L., FAUST, K., 2019. Can we predict keystones? Nature Reviews Microbiology, 17, 193.

RUUSKANEN, M.O., SOMMERIA-KLEIN, G., HAVULINNA, A.S., NIIRANEN, T.J., LAHTI, L., 2021. Modelling spatial patterns in host-associated microbial communities. Environmental Microbiology, 23 (5), 2374-2388.

SABATER, S., GUASCH, H., RICART, M., ROMANÍ, A., VIDAL, G., KLÜNDER, C., SCHMITT-JANSEN, M., 2007. Monitoring the effect of chemicals on biological communities. The biofilm as an interface. Analytical and Bioanalytical Chemistry, 387, 1425-1434.

SAGOVA-MARECKOVA, M., BOENIGK, J., BOUCHEZ, A., CERMAKOVA, K., CHONOVA,T., CORDIER, T., EISENDLE, U., ELERSEK, T., FAZI, S., FLEITUCH, T., FRÜHE, L., GAJDOSOVA, M., GRAUPNER, N., HAEGERBAEUMER, A., KELLY, A.-M., KOPECKY, J., LEESE, F., NÕGES P, ORLIC, S.,, PANKSEP K., PAWLOWSKI, J., PETRUSEK, A., PIGGOTT, J.J., RUSCH, J.C., SALIS, R., SCHENK, J., SIMEK, K., STOVICEK, A., STRAND, D.A., VASQUEZ, M.I., VRÅLSTAD, T., ZLATKOVIC, S., ZUPANCI, M., STOECK, T., 2021. Expanding ecological assessment by integrating microorganisms into routine freshwater biomonitoring. Water Research, 191, 116767.

SANDER, E.L., WOOTTON, J.T., ALLESINA, S., 2017. Ecological Network Inference From Long-Term Presence-Absence Data. Sci. Rep., 7, 7154.

SANDERS, D., THÉBAULT, E., KEHOE, R., VAN VEEN, F.J.K., 2018. Trophic redundancy reduces vulnerability to extinction cascades. Proceedings of the National Academy of Sciences of the United States of America, 115 (10), 2419-2424.

SAUVE, A.M.C., THÉBAULT, E., POCOCK, M.J.O., FONTAINE, C., 2016. How plants connect pollination and herbivory networks and their contribution to community stability. Ecology, 97, 908-917.

SAYLES, J.S., GARCIA, M.M., HAMILTON, M., ALEXANDER, S.M., BAGGIO, J.A., FISCHER, A.P., INGOLD, K., MEREDITH, G.R., PITTMAN, J., 2019. Social-ecological network analysis for sustainability sciences: a systematic review and innovative research agenda for the future. Environ. Res. Lett., 14, 093003.

SCHENEKAR, T., 2023. The current state of eDNA research in freshwater ecosystems: are we shifting from the developmental phase to standard application in biomonitoring? Hydrobiologia, 850 (6), 1263-1282.

SCHIRMER, M., IJAZ, U.Z., D'AMORE, R., HALL, N., SLOAN, W.T., QUINCE, C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Research, 43, e37.

SCHWAGER, E., MALLICK, H., VENTZ, S., HUTTENHOWER, C., 2017. A Bayesian method for detecting pairwise associations in compositional data. PLoS Computational Biology, 13, e1005852.

SCOTTI, M., CIOCCHETTA, F., JORDÁN, F., 2013. Social and landscape effects on food webs: a multi-level network simulation model. Journal of Complex Networks, 1, 160-182.

SEGAR, S.T., FAYLE, T.M., SRIVASTAVA, D.S., LEWINSOHN, T.M., LEWIS, O.T., NOVOTNY, V., KITCHING, R.L., MAUNSELL, S.C., 2020. The Role of Evolution in Shaping Ecological Networks. Trends in Ecology & Evolution, 35, 454-466.

SHADE, A., PETER, H., ALLISON, S.D., BAHO, D.L., BERGA, M., BÜRGMANN, H., HUBER, D.H., LANGENHEDER, S., LENNON, J.T., MARTINY, J.B.H., MATULICH, K.L., SCHMIDT, T.M. AND HANDELSMAN, J., 2012. Fundamentals of microbial community resistance and resilience. Frontiers in Microbiology, 3, 417.

SHAFIEI, M., DUNN, K.A., BOON, E., MACDONALD, S.M., WALSH, D.A., GU, H., BIELAWSKI, J.P., 2015. BioMiCo: a supervised Bayesian model for inference of microbial community structure. Microbiome, 3, 8.

SHAFIEI, M., DUNN, K.A., CHIPMAN, H., GU, H., BIELAWSKI, J.P., 2014. BiomeNet: A Bayesian model for inference of metabolic divergence among microbial communities. PLoS Computational Biology 10, e1003918.

SHAW, G.T.-W., PAO, Y.Y., WANG, D., 2016. MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. BMC Bioinformatics, 17, 488.

SHOKRALLA, S., SPALL, J.L., GIBSON, J.F., HAJIBABAEI, M., 2012. Next-generation sequencing technologies for environmental DNA research. Molecular Ecology 21, 1794-1805. https://doi.org/10.1111/j.1365-294X.2012.05538.x

SIEGWALD, L., CABOCHE, S., EVEN, G., VISCOGLIOSI, E., AUDEBERT, C., CHABÉ, M., 2019. The impact of bioinformatics pipelines on microbiota studies: Does the analytical “microscope” affect the biological interpretation? Microorganisms, 7, 393.

SILK, M.J., FINN, K.R., PORTER, M.A., PINTER-WOLLMAN, N., 2018. Can Multilayer Networks Advance Animal Behavior Research? Trends in Ecology & Evolution, 33, 376-378.

SILKNETTER, S., CREED, R.P., BROWN, B.L., FRIMPONG, E.A., SKELTON, J., PEOPLES, B.K., 2020. Positive biotic interactions in freshwaters: A review and research directive. Freshwater Biology, 65, 811-832.

SLADE, E.M., KIRWAN, L., BELL, T., PHILIPSON, C.D., LEWIS, O.T., ROSLIN, T., 2017. The importance of species identity and interactions for multifunctionality depends on how ecosystem functions are valued. Ecology, 98, 2626-2639.

SMITH, C.J., MCKEW, B.A., COGGAN, A., WHITBY, C., 2016. Hydrocarbon and Lipid Microbiology Protocols. Berlin, Heidelberg: Springer.

SONG, C., SIMMONS, B.I., FORTIN, M.-J., GONZALEZ, A., KAISER-BUNBURY, C.N., SAAVEDRA, S., 2022. Rapid monitoring for ecological persistence. Ecology, 120 (20), e2211288120.

SPENCER, S.J., TAMMINEN, M.V., PREHEIM, S.P., GUO, M.T., BRIGGS, A.W., BRITO, I.L., A WEITZ, D., PITKÄNEN, L.K., VIGNEAULT, F., VIRTA, M.P., ALM, E.J., 2016. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. The ISME Journal, 10, 427-436.

SRINIVASAN, U.T., DUNNE, J.A., HARTE, J., MARTINEZ, N.D., 2007. Response of complex food webs to realistic extinction sequences. Ecology, 88 (3), 671-682.

SRIVATHSAN, A., LEE, L., KATOH, K., HARTOP, E., KUTTY, S.N., WONG, J., YEO, D., MEIER, R., 2021. ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. BMC Biology, 19, 217.

STANICZENKO, P.P.A., KOPP, J.C., ALLESINA, S., 2013. The ghost of nestedness in ecological networks. Nat. Commun., 4, 1391.

STANICZENKO, P.P.A., LEWIS, O.T., JONES, N.S., REED-TSOCHAS, F., 2010. Structural dynamics and robustness of food webs. Ecology letters, 13 (7), 891-899.

STEELE, J., COUNTWAY, P., XIA, L., VIGIL P.D., BEMAN, J.M., KIM, D.Y., CHOW C.-E.T., SACHDEVA, R., JONES, A.C., SCHWALBACH, M.S., ROSE, J.M., HEWSON, I., PATEL, A., SUN, F., CARON, D.A., FUHRMAN, J.A., 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME Journal, 5, 1414-1425.

STEIN, R.R., BUCCI, V., TOUSSAINT, N.C., BUFFIE, C.G., RÄTSCH G., PAMER, E.G., SANDER, C., XAVIER, J.B., 2013. Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. PLoS Computational Biology, 9, e1003388.

STEVENS, D.L., OLSEN, A.R., 1999. Spatially Restricted Surveys Over Time for Aquatic Resources. Journal of Agricultural, Biological, and Environmental Statistics, 4, 415-428.

STEVENS, D.L., OLSEN, A.R., 2004. Spatially Balanced Sampling of Natural Resources. Journal of the American Statistical Association, 99 (456), 262-278.

STOUFFER, D.B., BASCOMPTE, J., 2011. Compartmentalization increases food-web persistence. Proceedings of the National Academy of Sciences, 108, 3648-3652.

STRICKLER, K.M., FREMIER, A.K., GOLDBERG, C.S. 2015. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biological Conservation, 183, 85-92.

STROGATZ, S., 2001. Exploring complex networks. Nature, 410, 268-276.

SUTCLIFFE, B., HOSE, G.C., HARFORD, A.J., MIDGLEY, D.J., GREENFIELD, P., PAULSEN, I.T., CHARITON, A.A. 2019. Microbial communities are sensitive indicators for freshwater sediment copper contamination. Environmental Pollution, 246, 1028-1038.

SYMONDSON, W.O.C., 2002. Molecular identification of prey in predator diets. Molecular Ecology, 11, 627-641.

TAMADDONI-NEZHAD, A., BOHAN, D.A., RAYBOULD, A., MUGGLETON, S., 2015. Inductive Logic Programming. New York, New York, USA: Springer International Publishing.

TAMADDONI-NEZHAD, A., CHALEIL, R., KAKAS, A., MUGGLETON, S., 2006. Application of abductive ILP to learning metabolic network inhibition from temporal data. Machine Learning, 64 (1), 209-230.

TAMADDONI-NEZHAD, A., MILANI, G.A., RAYBOULD, A., MUGGLETON, S., BOHAN, D., 2013. Construction and Validation of Food-webs using Logic-based Machine Learning and Text-mining. Advances in Ecological Research, 49, 225-289.

TAMADDONI-NEZHAD, A., MILANI, G.A., RAYBOULD,A., MUGGLETON, A.S., BOHAN, D.A., 2013. Advances in Ecological Research. Massachusetts, USA: Academic Press, pp. 225–289.

TAVAKOLI, S. YOOSEPH, S., 2019. Learning a mixture of microbial networks using minorization–maximization. Bioinformatics, 25, i23-i30.

TEDERSOO, L., BAHRAM, M., POLME, S., KOLJALG, U., YOROU, N.S., WIJESUNDERA, R., VILLARREAL RUIZ, L., VASCO-PALACIOS, A.M., THU, P.Q., SUIJA, A., SMITH, M.E., SHARP, C., SALUVEER, E., SAITTA, A., ROSAS, M., RIIT, T., RATKOWSKY, D., PRITSCH, K., POLDMAA, K., PIEPENBRING, M., PHOSRI, C., PETERSON, M., PARTS, K., PAERTEL, K., OTSING, E., NOUHRA, E., NJOUONKOU, A.L., NILSSON, R.H., MORGADO, L.N., MAYOR, J., MAY, T.W., MAJUAKIM, L., LODGE, D.J., LEE, S.S., LARSSON, K.-H., KOHOUT, P., HOSAKA, K., HIIESALU, I., HENKEL,

T.W., HAREND, H., GUO, L., GRESLEBIN, A., GRELET, G., GEML, J., GATES, G., DUNSTAN, W., DUNK, C., DRENKHAN, R., DEARNALEY, J., DE KESEL, A., DANG, T., CHEN, X., BUEGGER, F., BREARLEY, F.Q., BONITO, G., ANSLAN, S., ABELL, S., ABARENKOV, K., 2014. Global diversity and geography of soil fungi. Science, 346, 1256688.

TEDERSOO, L., BAHRAM, M., ZINGER, L., NILSSON, R. H., KENNEDY, P. G., YANG, T., ANSLAN, S., MIKRYUKOV, V., 2022. Best practices in metabarcoding of fungi: From experimental design to results. Molecular Ecology, 31, 2769-2795.

TENG, J., MCCANN, K. S., 2004. Dynamics of compartmented and reticulate food webs in relation to energetic flows. The American Naturalist, 164, 85-100.

TERCEL, M.P.T.G., SYMONDSON, W.O.C., CUFF, J.P., 2021. The problem of omnivory: A synthesis on omnivory and DNA metabarcoding. Molecular Ecology, 30, 2199-2206.

THEBAULT, E., HUBER, V., LOREAU, M., 2007. Cascading extinctions and ecosystem functioning: contrasting effects of diversity depending on food web structure. Oikos, 116, 163-173.

THIERRY, A., BECKERMAN, A.P., WARREN, P.H., WILLIAMS, R.J., COLE, A.J., PETCHEY, O.L., 2011. Adaptive foraging and the rewiring of size-structured food webs following extinctions. Basic and Applied Ecology, 12, 562-570.

THOMPSON, J.N., 2014. Interaction and Coevolution. Michigan, USA: University of Chicago Press.

THOMPSON, L., SANDERS, J., MCDONALD, D. AMIR A., LADAU J., LOCEY K., 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., XU Z.Z., JIANG L., HAROON M.F., KANBAR J., ZHU Q., JIN SONG S., KOSCIOLEK T., BOKULICH N.A., LEFLER J., BRISLAWN C.J., HUMPHREY G., OWENS S.M., HAMPTON-MARCELL J., MCKENZIE V., FIERER N., FUHRMAN J.A., CLAUSET A., STEVENS R.L., SHADE A., POLLARD K.S., GOODWIN K.D., JANSSON J.K., KNIGHT R., 2017. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature, 551, 457-463.

THOMPSON, M.S.A., BANKIER, C., BELL, T., DUMBRELL, A.J., GRAY, C., LEDGER, M.E., LEHMANN, K., MCKEW, B.A., SAYER, C.D., SHELLEY, F., TRIMMER, M., WARREN, S.L. AND WOODWARD, G., 2016. Gene-to-ecosystem impacts of a catastrophic pesticide spill: testing a multilevel bioassessment approach in a river ecosystem. Freshwater Biology, 61, 2037-2050.

TIKHONOV, G., OPEDAL, Ø.H., ABREGO, N., LEHIKOINEN, A., DE JONGE, M.M.J., OKSANEN, J., OVASKAINEN, O., 2020. Joint species distribution modelling with the r-package Hmsc. Methods in Ecology and Evolution 11, 442-447.

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

TRAUGOTT, M., KAMENOVA, S., RUESS, L., SEEBER, J., PLANTEGENEST, M., 2013. Advances in Ecological Research: Ecological Networks in an Agricultural World. Academic Press, 177-224.

TRAVESET, A., RICHARDSON, D.M., 2014. Mutualistic Interactions and Biological Invasions. Annual Review of Ecology, Evolution, and Systematics, 45, 89-113.

TSAI, S., ZHENG, Z., NGUYEN, N., LIEBERS, M., TOPKAR, V.V., THAPAR, V., WYVEKENS, N., KHAYTER, C., IAFRATE, A.J., LE, L.P., ARYEE, M.J., JOUNG, J.K. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology, 33, 187–197.

TSHIKANTWA, T.S., ULLAH, M.W., HE, F., YANG, G., 2018. Current trends and potential applications of microbial interactions for human welfare. Frontiers in Microbiology, 9, 1156.

TYLIANAKIS, J.M., DIDHAM, R.K., BASCOMPTE, J., WARDLE. D.A., 2008. Global change and species interactions in terrestrial ecosystems. Ecology Letters, 11 (12), 1351-1363.

VACHER, C., TAMADDONI-NEZHAD, A., KAMENOVA, S., PEYRARD, N., MOALIC, Y., SABBADIN, R., SCHWALLER, L., CHIQUET, J., SMITH, M.A., VALLANCE, J., FIEVET, V., 2016. Learning ecological networks from next-generation sequencing data. Advances in Ecological Research, 54, 1-39.

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.

VASSELON, V., RIMET, F., TAPOLCZAI, K., BOUCHEZ, A., 2017. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). Ecological Indicators, 82, 1–12.

VAUGHAN, I.P., GOTELLI, N.J., MEMMOTT, J.. PEARSON, C.E., WOODWARD, G., SYMONDSON, W.O.C., 2018. econullnetr: An r package using null models to analyse the structure of ecological networks and identify resource selection. Methods in Ecology and Evolution, 9 (3), 728-733.

VÁZQUEZ, D.P. BLÜTHGEN, N., CAGNOLO, L., CHACOFF, N.P., 2009. Uniting pattern and process in plant-animal mutualistic networks: a review. Annals of botany, 103 (9), pp. 1445-1457.

VIEIRA, M.C., ALMEIDA-NETO, M., 2015. A simple stochastic model for complex coextinctions in mutualistic networks: robustness decreases with connectance. Ecology Letters, 18, 144-152.

VIZENTIN-BUGONI, J., SPERRY, J.H., KELLEY, J.P., GLEDITSCH, J.M., FOSTER, J.T., DRAKE, D.R., HRUSKA, A.M., WILCOX, R.C., CASE, S.B., TARWATER, C.E., 2021. Ecological correlates of species' roles in highly invaded seed dispersal networks. Proceedings of the National Academy of Sciences, 118, e2009532118.

WANG, S., WANG, X., HAN, X., DENG, Y., 2018. Higher precipitation strengthens the microbial interactions in semi-arid grassland soils. Global Ecology and Biogeography, 27, 570-580.

WANG, Y., JUN, Y., JU, F., LIU, L., BOYD, J.A., DENG, Y.., PARKS, D.H., JIANG, X., YIN, X., WOODCROFT, B.J., TYSON, G.W., HUGENHOLTZ, P., POLZ, M.F., ZHANG, T., 2021. Successional dynamics and alternative stable states in a saline activated sludge microbial community over 9 years. Microbiome, 9, 199.

WANG, Y.I., NAUMANN, U., WRIGHT, S.T., WARTON, D.I., 2012. mvabund–an R package for model-based analysis of multivariate abundance data. Methods in Ecology and Evolution, 3, 471-474.

WARTON, D.I. 2011. Regularized sandwich estimators for analysis of high-dimensional data using generalized estimating equations. Biometrics, 67, 116-123.

WARTON, D.I., WRIGHT, S.T., WANG, Y., 2012. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution, 3, 89-101.

WASSERMAN, S., FAUST, K. 1994. Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press.

WATNICK, P., KOLTER, R., 2000. Biofilm, City of Microbes. Journal of Bacteriology, 182, 10.

WATTS, D., STROGATZ, S., 1998. Collective dynamics of 'small-world' networks. Nature, 393, 440-442.

WEISS, S., VAN TREUREN, W., LOZUPONE, C., FAUST, K., FRIEDMAN, J., DENG, Y. XIA, L.C., XU, Z.Z., URSELL, L., ALM, E.J., BIRMINGHAM, A., CRAM, J.A., FUHRMAN, J.A., RAES, J., SUN, F., ZHOU, J., KNIGHT, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. The ISME Journal, 10, 1669-1681.

WEISS, S., XU, Z.Z., PEDDADA, S., AMIR, A., BITTINGER, K., GONZALEZ, A., LOZUPONE, C., ZANEVELD, J.R., VÁZQUEZ-BAEZA, Y., BIRMINGHAM, A., HYDE, E.R., KNIGHT, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome, 5, 27.

WEMHEUER, F., TAYLOR, J. A., DANIEL, R., JOHNSTON, E., MEINICKE, P., THOMAS, T., WEMHEUER, B., 2020. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. Environmental Microbiome, 15, 1-12.

WHITE, H.J., BAILEY, J.J., BOGDAN, C., ROSS, S.R.P.-J., 2023. Response trait diversity and species asynchrony underlie the diversity–stability relationship in Romanian bird communities. Journal of Animal Ecology, 92, 2309-2322.

WHITMAN, W.B., COLEMAN, D.C., WIEBE, W.J., 1998. Prokaryotes: The unseen majority. Proceedings of the National Academy of Sciences, 95, 6578-6583.

WIDDER, S., BESEMER, K., SINGER, G.A., BATTIN., T.J., 2014. Fluvial network organization imprints on microbial co-occurrence networks. Proceedings of the National Academy of Sciences of the United States of America, 111 (35), 12799-12804.

WINDSOR, F.M., 2023. Expanding network ecology in freshwater ecosystems. Journal of Animal Ecology, 92 (8), 1575-1588.

WINDSOR, F.M., ARMENTERAS, D., ASSIS, A.P.A., ASTEGIANO, J., SANTANA, P.C., CAGNOLO, L., CARVALHEIRO, L.G., EMARY, C., FORT, H., GONZALEZ, X.I., KITSON, J.J.N., LACERDA, A.C.F., LOIS, M., MÁRQUEZ-VELÁSQUEZ, V., MILLER, K.E., MONASTEROLO, M., OMACINI, M., MAIA, K.P., PALACIOS, T.P., POCOCK, M.J.O., POGGIO, S.L., VARASSIN, I.G., VÁZQUEZ, D.P., TAVELLA, J., ROTHER, D.C., DEVOTO, M., GUIMARÃES, P.R., EVANS, D.M., 2022. Network science: Applications for sustainable agroecosystems and food security. Perspectives in Ecology and Conservation, 20, 79-90.

WINDSOR, F.M., PEREIRA, M.G., TYLER, C.R., ORMEROD, S.J., 2019. Biological Traits and the Transfer of Persistent Organic Pollutants through River Food Webs. Environ. Sci. Technol., 53, 13246-13256.

WINDSOR, F.M., VAN DEN HOOGEN, J., CROWTHER, T.W., EVANS, D.M., 2023. Using ecological networks to answer questions in global biogeography and ecology. Journal of Biogeography, 50, 57–69.

WISZ, M.S., POTTIER, J., KISSLING, W.D., PELLISSIER, L., LENOIR, J., DAMGAARD, C.F., DORMANN, C.F., FORCHHAMMER, M.C., GRYTNES, J.-A., GUISAN, A., HEIKKINEN, R.K., HØYE, T.T., KÜHN, I., LUOTO, M., MAIORANO, L., NILSSON, M.-C., NORMAND, S., ÖCKINGER, E., SCHMIDT, N.M., TERMANSEN, M., TIMMERMANN, A., WARDLE, D.A., AASTRUP, P., SVENNING, J.-C., 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. Biological Reviews, 88, 15-30.

WOODWARD G., SPEIRS, D.C., HILDREW, A., 2005. Quantification and resolution of a complex, size-structured food web. Advances in Ecological Research, 36, 85-135.

WOODWARD, G., BLANCHARD, J., LAURIDSEN, R.B., EDWARDS, F.K., JONES, J.I., FIGUEROA, D., WARREN, P.H., PETCHEY, O.L., 2010. Advances in Ecological Research, Integrative Ecology: From Molecules to Ecosystems. Massachusetts, USA: Academic Press

XIA, L.C., AI, D., CRAM, J., FUHRMAN, J.A., SU, F., 2013. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. Bioinformatics, 29, 230–237.

XIONG, W., JOUSSET, A., GUO, S., KARLSSON, I., ZHAO, Q., WU, H., KOWALCHUK, G.A., SHEN, Q., LI, R., GEISEN, S., 2018. Soil protist communities form a dynamic hub in the soil microbiome. The ISME Journal, 12, 634-638.

XUE, R., WANG, C., ZHAO, L., SUN, B., WANG, B., 2022. Agricultural intensification weakens the soil health index and stability of microbial networks. Agriculture, Ecosystems and Environment, 339, 108118.

XUE, Y., CHEN, H., YANG, J.R., LIU, M., HUANG, B., YANG, J., 2018. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. The ISME Journal, 12, 2263-2277.

YACHI, S., LOREAU, M., 1999. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. Proceedings of the National Academy of Sciences, 96, 1463-1468.

YANG, P., TAN, C., HAN, M., CHENG, L., CUI, X., NING, K., 2020. Correlation-Centric Network (CCN) representation for microbial co-occurrence patterns: new insights for microbial ecology. NAR Genomics and Bioinformatics, 2.

YANG, P., YU, S., CHENG, L., NING, K., 2019. Meta-network: Optimized species-species network analysis for microbial communities. BMC Genomics, 20, 187.

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 (1), 129-137.

YANG, Q, YANG, Y., HUANG, J., WANG., Z., FENG, M., CHENG, H., ZHANG, P., ZHANG, H., XU, J., ZHANG, M. 2023. The impact of warming on assembly processes and diversity patterns of bacterial communities in mesocosms. Microorganisms, 11 (11).

YE, Z., LI, J., WANG, J., ZHANG, C., LIU, G., DONG, Q., 2021. Diversity and co-occurrence network modularization of bacterial communities determine soil fertility and crop yields in arid fertigation agroecosystems. Biology and Fertility of Soils, 57, 809-824.

YOCCOZ, N.G., BRÅTHEN, K.A., GIELLY, L., HAILE, J., EDWARDS, M. E., GOSLAR, T., Von STEDINGK, H., BRYSTING, A.K., COISSAC, E., POMPANON, F., SØNSTEBØ, J.H., MIQUEL, C., VALENTINI, A., De BELLO, F., CHAVE, J., THUILLER, W., WINCKER, P., CRUAUD, C., GAVORY, F., RASMUSSEN, M., GILBERT, M.T.P., ORLANDO, L.,

BROCHMANN, C., WILLERSLEV, E., TABERLET, P. DNA from soil mirrors plant taxonomic and growth form diversity. Molecular ecology, 21 (15), 3647-3655.

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.

YU, H., XIE, B., KHAN, R., YAN, H., SHEN, G., 2020. The changes in functional marker genes associated with nitrogen biological transformation during organic-inorganic co-composting. Bioresource technology 295, 122197.

YUAN, M.M., GUO, X., WU, L. 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.

ZAPPELINI, C., KARIMI, B., FOULON, J., LACERCAT-DIDIER, L., MAILLARD, F., VALOT, B., BLAUDEZ, D., CAZAUX, D., GILBERT, D., YERGEAU, E., GREER, C., 2015. Diversity and complexity of microbial communities from a chlor-alkali tailings dump. Soil Biology and Biochemistry, 90, 101-110.

ZENG, Y., DAOLIN, P., ZHAO, H., WANG, T., 2022. A Zero-Inflated Logistic Normal Multinomial Model for Extracting Microbial Compositions. Journal of the American Statistical Association 118 (544), 2356-2369.

ZHAO, D., SHEN, F., ZENG, J., HUANG, R., YU, Z., WU, Q.L., 2016. Network analysis reveals seasonal variation of co-occurrence correlations between Cyanobacteria and other bacterioplankton. Science of The Total Environment, 573, 817-825.

ZHAO, K., XUE, R., LIU, Y., XU, J., MA, B. 2019. Strengthening Insights in Microbial Ecological Networks from Theory to Applications. Novel Systems Biology Techniques 4 (3).

ZHENG, W., XUE, D., LI, X., DENG, Y., RUI, J., FENG, K., WANG, Z., 2017. The responses and adaptations of microbial communities to salinity in farmland soils: a molecular ecological network analysis. Applied Soil Ecology, 120, 239-246.

ZHOU, J., DENG, Y., LUO, F., HE, Z., TU, Q., ZHI, X., 2010. Functional molecular ecological networks. mBio 1, e00169-10.

ZHOU, J., DENG, Y., LUO, F., HE, Z., YANG, Y. 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2. mBio 2, e00122-11.

ZHOU, J., NING, D., 2017. Stochastic community assembly: does it matter in microbial ecology? Microbiology and Molecular Biology Reviews 81, e00002-17.

ZHU, B. 2006. Degradation of plasmid and plant DNA in water microcosms monitored by natural transformation and real-time polymerase chain reaction (PCR). Water research, 40 (17), 3231-3238.

ZHU, Y., PURDY, K.J., EYICE, Ö. 2020. Disproportionate increase in freshwater methane emissions induced by experimental warming. Nature Climate Change, 10, 685-690.

ZINGER, L., BONIN, A., ALSOS, I.G., BÁLINT, M., BIK, H., BOYER, F., CHARITON, A.A., CREER, S., COISSAC, E., DEAGLE, B.E., DE BARBA, M., DICKIE, I.A., DUMBRELL, A.J., FICETOLA, G.F., FIERER, N., FUMAGALLI, L., GILBERT, M.T.P., JARMAN, S., JUMPPONEN, A., KAUSERUD, H., ORLANDO, L., PANSU, J., PAWLOWSKI, J., TEDERSOO, L., THOMSEN, P.F., WILLERSLEV, E., TABERLET, P, 2019. DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. Molecular Ecology, 28, 1857-1862.

ZULKEFLI, N.S., KIM, K.-H. AND HWANG, S.-J. 2019. Effects of Microbial Activity and Environmental Parameters on the Degradation of Extracellular Environmental DNA from a Eutrophic Lake. International Journal of Environmental Research and Public Health, 16 (18).

ZURELL, D., POLLOCK, L.J., THUILLER, W., 2018. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? Ecography, 41, 1812-1819.

List of abbreviations

AI - Artificial intelligence

ASV - Amplicon Sequence Variants

eDNA - Environmental DNA (Deoxyribonucleic acid)

eLSA - Extended Llocal Similarity Analysis

ENS - Ecological network science

EPS - Extracellular polymeric substance

GRTS - Generalised, randomised, tessellation, stratified

HMSC - Hierarchical Modelling of Species Vommunities

HPC - High power computers

J-SDM - Joint-species distribution model

LSA – Local Similarity Analysis

MAR – Matrix autoregression

MaxEnt – Maximum entropy

NGS – Next Generation Sequencing

OTU - Operational Taxonomic Units

RSN – River Surveillance Network

SPIEC-EASI - SParse InversE Covariance Estimation for Ecological Association Inference

TP – Think Piece

Glossary

Amensalisms - Interactions where one node is negatively affected whilst the other receives no benefit (e.g., animal trampling plants).

Antagonisms - Interactions where one node benefits at the expense of another (e.g., predator-prey or host-parasite interactions).

Binary network - A network comprised of interactions which are either present or absent (i.e., 1 or 0).

Biofilm - A diverse aggregate of microbial communities comprising bacteri, fungi, alage, and other protozoa within an extracellular polymeric substance matrix.

Commensalisms - Interactions where one node is positively affected whilst the other receives no benefit (e.g., cleaner-client interactions).

Connectance - The proportion of possible links between nodes that are realised.

Degree - The number of interactions associated with a node. For a directional graph this can be in degree (number of incoming interactions) and out degree (number of outgoing interactions).

Ecological network - A network which is comprised of living organisms and their interaction with other organisms and abiotic features in an ecosystem.

Ecosystem - A community of living organisms, which, together with their physical environment, interact as a functional unit.

Ecosystem function - The processes and activities carried out within an ecosystem. These functions contribute to the overall health of ecosystem.

Ecosystem services - The benefits to humans that are provided by ecosystems and ecosystem function.

Engineering resilience - Stability near a steady-state equilibrium, where the measure is the return time or speed of return to the equilibrium state.

Facilitation - A mutualistic or commensalistic interaction where the activity of one organism enables the activity of another (e.g., an organism feeding on leaves breaks the leaves into smaller pieces that another organism can feed on).

Forbidden links - Interactions between nodes that cannot possibly exist (e.g., two organisms do not occupy the same habitat, or coexist during the same seasons).

Functional diversity - The variety and range of functional traits and environmental processes exhibited by organisms within an ecosystem or community.

Interaction - The relationship and exchange that occurs between organisms or an organism and its abiotic environment.

Link - Interactions between the elements in a network (e.g., species interactions, movement, migration, foraging).

Metabarcoding - An eDNA data analysis method which determines the range and diversity of many taxa from a single sample. Species ID are assigned through comparison to a reference library.

Microbe - Micro-organisms including bacteria, archaea, fungi, and protists.

Modularity - A measure of the degree to which a network is divided into more strongly connected subnetworks. A modular network is one where subsets of nodes are more strongly or densely connected to one another, with weak or sparse links between these strongly connected subnetworks.

Multilayer network - An interaction network which is organised into different “layers” where each layer represents different interactions, or spatial and temporal segregation (e.g., networks across habitats or time points). There are intra-layer edges, those interactions that occur within layers, but also inter-layer edges, those interactions that link nodes across layers. Intra-layer edges are usually standard types of interaction, e.g., mutualistic and antagonistic, and inter-layer edges can represent links between the same individuals over time, or the same species across different habitats, or the same individual across different habitats (i.e., foraging or dispersal).

Multiplex network - An interaction network where individuals are connected across multiple networks of different types of interaction (i.e., merged mutualistic and antagonistic networks).

Multitrophic network - An interaction network where individuals are connected across different trophic levels. Interactions can be either mutualistic or antagonistic, however, the interactions between each group of organisms are consistent (i.e., plants-herbivores-parasitoids or microbes-plants-pollinators).

Multivariate - Datasets or statistical analyses which involve two or more variables.

Mutualisms - Interactions where both nodes benefit (e.g., plant-pollinator or host-symbiont).

Nestedness - A measure of the propensity for nodes to interact with subsets of the interaction partners of well-connected nodes. For example, a nested network is one where nodes with few connections all interact with nodes that well-connected nodes also interact with.

Network - A collection of interconnected elements (i.e., nodes) linked together through relationships or interactions.

Next generation sequencing - Synonym high-throughput sequencing. A broad term used to describe modern sequencing technologies that are able to sequence DNA and RNA more rapidly that Sanger sequencing.

Node - Distinct elements in a network (e.g., cells, species, habitats, ecosystems).

Response diversity - Range of potential responses or reactions of organisms to a given stressor (e.g., environmental change) or set of stressors.

Robustness - The tolerance of a network structure to node removal (I.e., in inter-specific ecological networks, the number of secondary extinctions resulting from a sequence of primary extinctions).

Sampling completeness - The degree to which the observed networks (constructed through sampling) represent the potential interactions that could occur in a given network.

Social-ecological networks - A merged network which includes social interactions, ecological interactions and the reciprocal interactions between these two interaction types (e.g., management and ecosystem service provision, respectively).

Spatial networks - A network where the nodes and links are organised spatially (i.e., the nodes are habitat patches and the links are the movements of organisms or populations).

Stability - The ability of an ecosystem to resist, or be resilient to, perturbations.

Topology - The arrangement of nodes and edges in a network. This is also commonly referred to as network structure.

Weighted network – A network where interactions are quantified (e.g., number of prey consumed, number of flowers visited, parasite abundance).

Would you like to find out more about us or your environment?

Then call us on

03708 506 506 (Monday to Friday, 8am to 6pm)

Email: enquiries@environment-agency.gov.uk

Or visit our website

www.gov.uk/environment-agency

incident hotline

0800 807060 **(24 hours)**

floodline

0345 988 1188 **(24 hours)**

Find out about call charges (https://www.gov.uk/call-charges)

Environment first

Are you viewing this onscreen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don’t forget to reuse and recycle.