PLOS ONE

RESEARCH ARTICLE

MicNet toolbox: Visualizing and unraveling a microbial network

Natalia Favila[1☯], David Madrigal-Trejo[2☯], Daniel Legorreta[1], Jazmín Sánchez-Pérez[2], Laura Espinosa-Asuar[2], Luis E. Eguiarte[2], Valeria Souza[2,3]*

1 Laboratorio de Inteligencia Artificial, Ixulabs, Mexico City, Mexico, **2** Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico, **3** Centro de Estudios del Cuaternario de Fuego-Patagonia y Antártica (CEQUA), Punta Arenas, Chile

☯ These authors contributed equally to this work.
* souza@unam.mx

Citation: Favila N, Madrigal-Trejo D, Legorreta D, Sánchez-Pérez J, Espinosa-Asuar L, Eguiarte LE, et al. (2022) MicNet toolbox: Visualizing and unraveling a microbial network. PLoS ONE 17(6): e0259756. https://doi.org/10.1371/journal.pone.0259756

Editor: Kazuhiro Takemoto, Kyushu Institute of Technology, JAPAN

Received: October 23, 2021

Accepted: April 5, 2022

Published: June 24, 2022

Data Availability Statement: MicNet toolbox is available to the scientific community as an open source project at https://github.com/Labevo/MicNetToolbox. All raw files are available at the GitHub repository, and the kombucha data is originally available from the ENA database (accession numbers ERP104502, ERP024546). In addition, we present an accompanying dashboard which can also be freely visited at http://micnetapplb-1212130533.us-east-1.elb.amazonaws.com, in which the files to build the kombucha network or any other compositional files

Abstract

Applications of network theory to microbial ecology are an emerging and promising approach to understanding both global and local patterns in the structure and interplay of these microbial communities. In this paper, we present an open-source python toolbox which consists of two modules: on one hand, we introduce a visualization module that incorporates the use of UMAP, a dimensionality reduction technique that focuses on local patterns, and HDBSCAN, a clustering technique based on density; on the other hand, we have included a module that runs an enhanced version of the SparCC code, sustaining larger datasets than before, and we couple the resulting networks with network theory analyses to describe the resulting co-occurrence networks, including several novel analyses, such as structural balance metrics and a proposal to discover the underlying topology of a co-occurrence network. We validated the proposed toolbox on 1) a simple and well described biological network of kombucha, consisting of 48 ASVs, and 2) we validate the improvements of our new version of SparCC. Finally, we showcase the use of the MicNet toolbox on a large dataset from Archean Domes, consisting of more than 2,000 ASVs. Our toolbox is freely available as a github repository (https://github.com/Labevo/MicNetToolbox), and it is accompanied by a web dashboard (http://micnetapplb-1212130533.us-east-1.elb.amazonaws.com) that can be used in a simple and straightforward manner with relative abundance data. This easy-to-use implementation is aimed to microbial ecologists with little to no experience in programming, while the most experienced bioinformatics will also be able to manipulate the source code's functions with ease.

Introduction

Microbiomes are not a mere collection of independent individuals, but rather, ensembles of intricate constituents, biotic and abiotic, that create highly complex systems where emergent interactions, structures, and functions are crucial for the survival and performance of the whole. The inference of microbial co-occurrence networks may help in understanding

the user inputs, can be explored and visualized in an interactive way.

Funding: This research was supported by DGAPA/UNAM-PAPIIT Project IG200319, CEQUA project ANID R20F0009, and PhD scholarship 970341 granted to J.S.P. by Consejo Nacional de Ciencia y Tecnologia (CONACyT).

Competing interests: The authors declare no competing interests.

emergent properties of these systems [1]: unravelling microbial interactomes [2], evaluating the effects of stress and perturbations in community stability [3], and providing a wide array of novel applications including diagnostics of environmental quality [4] and pathogen identification in disease management [5].

One promising contribution to unravel microbial ecological associations has been the application of network science as an increasingly used alternative to study complex systems [6], whose methods can handle the scale and diversity of high-throughput biological data [1]. Several approaches have been developed to infer microbial ecological associations, such that patterns can be visualized and analyzed within the schematic of a network. One of the most used family of methods is the inference by co-occurrence and correlations, such as: Pearson [7] or Spearman [8] correlation coefficient, Jaccard distance [9] or Bray-Curtis dissimilarity [10], Local Similarity Analysis [11, 12], Maximal Information Coefficient [13], MENA that adapts Random Matrix Theory [14, 15], SparCC based on Aitchison's log-ratio analysis [16, 17], and CoNet which combines information of several metrics [18, 19]. Other types of techniques, such as ordinary differential equations (ODE) models [20] have also been used as an alternative to capture microbial interactions, amongst which the generalized Lotka Volterra equations (gLV) are one of the most used [21] for two-species systems, and potentially useful for three-species systems or larger [22]. Finally, MetaMIS [23], LIMITS [24], and some variations which integrate forward stepwise regressions and bootstrap aggregation [2], offer a different implementation of the gLV equations.

Although the potential value of microbial co-occurrence networks is known, there are several caveats and limitations. To start with, high-throughput genomic data is often associated with low annotation resolution at the species level, which makes it difficult to differentiate between strains and species. This has led to the usage of ASVs (Amplicon Sequence Variants) or OTUs (Operational Taxonomic Units) to obtain a more reliable account of discrete ecological players, leading to hundreds and sometimes thousands of potential organisms [5, 25]. However, there are just a few techniques which enable the use of thousands of OTUs/ASVs in the construction of networks for the most diverse ecosystems, such as soil [19]. Furthermore, microbial abundances are normally presented as relative abundance matrices, which creates compositional data sets that are often sparse [1]. Some existing methods are commonly known to provide an efficient approach for compositional effects, spurious correlations, and sparse data handling; nonetheless, biological interaction inferences from compositional data alone should be taken with caution (see **Weiss et al. (2016)** [22]**, Dohlman & Shen (2019)** [2], **Hirano & Takemoto (2019)** [26] for a review on performance). Finally, noise or contamination are expected not be part of the real system portrayed in the network, and the identification of taxa that are not an integral part of a community could be essential for accurate biological interpretations [22].

The aforementioned issues have led to highly divergent results while trying to infer direct correlations between OTUs/ASVs [1]. Therefore, algorithms with a reliable statistical approach are needed. In addition to the intrinsic limitations of inferring interactions from microbial community data, there is a gap in the analysis of networks to obtain most of the biologically relevant information: many of the existing methodologies are not easily reachable to the research community, nor do they implement posterior analysis to retrieve information of the co-occurrence network in a clearly and accessible format. Furthermore, at the interpretation level, biologically-meaningful inferences derived co-occurrence networks is still a challenge to be untangled, as signals from co-occurrence may suggest a wide array of phenomena beyond interspecific ecological interactions [1], and most network metrics have debatable or unknown links to relevant concepts in microbial ecology [1, 4, 27]. In **Table 1**, we present a summary of several network analysis metrics and their current biological interpretation.

Table 1. Description of several network metrics and properties currently used in biological networks, including some of their prospective interpretations.

Metric/Network property	Definition	Prospective biological Interpretation in microbial co-occurrence networks	References
Total Nodes/Vertices	Total Entities within a network.	Total number of taxa (species, OTUs/ASVs) in a network (species richness); number of connected taxa; common measure of ecosystem state in response to perturbations	[4, 28–30]
Edges, Links, Relationship, connection	Relationship or associations between nodes. For co-occurrence networks, relationships exist between pairs of nodes.	Ecological associations, including interspecific interactions, niche overlap, cross-feeding, abiotic co-occurrence drivers, among others	[1, 4, 31, 32]
Density, Connectance, Complexity (network scale), Interactions diversity, Probability of connection	Fraction of edges that are actually present in the network with respect to all possible edges.	Reflection of the incidence of ecosystemic processes; Possible measure of ecological resilience; organization level of the community; measure of complexity in the microbial network	[4, 28, 33–36]
Connectivity	Total number of relationships in a network	Total number of ecological associations within a biological network	[28, 37]
Connected component	Sets of nodes, where every pair of nodes have a path between them.	Microbial network where every OTU/ASV have an indirect ecological association with every other OUT/ASV	[27, 38–40]
Average degree, Complexity (taxon scale), Connectedness (normalized degree)	Average number of edges connected to a node; average number of neighbors for a given node.	Measure of complexity in the microbial network	[4, 27, 35, 36, 41]
Degree centrality	Centrality of a node based on degree. i.e., nodes with higher degree are more central to the network. It is a measurement of popularity.	Keystone taxa; taxa that interacts the most within the community	[27, 36, 42, 43]
Closeness centrality	Centrality of a node based on its proximity to all other nodes in the network. It is a measure of broadcaster nodes, that is, nodes that can influence the network fastest	Keystone taxa; taxa that, if perturbed, influence the network the fastest.	[27, 43, 44]
Betweenness centrality	Centrality of a node based on how often a node is situated on paths between other nodes. It is a measurement of bridge nodes	Keystone taxa; taxa more important in communication in the network.	[4, 27, 41, 44]
PageRank	Centrality measure that computes a ranking of the nodes based on the structure of the incoming links. It identifies hub nodes.	Keystone taxa	[45–47]
Negative:Positive relationship ratio, Behavior	Ratio of positive and negative relationships. If > 1 there are more negative interactions, if < 1 there are more positive interactions present in the network.	Potential measure of cooperation level within the community; measure of community stability (ecological resilience and resistance)	[3, 4, 36, 41, 48–51]
Average shortest path length (AL), Average geodesic path	Average number of steps in the shortest paths from one node to another. It is calculated for all pairs and then averaged.	Microbial networks usually present small AL: measure of network's response speed to perturbations (ecological resilience); community cohesion; measure of information and substance flow	[4, 35, 37, 42, 52–54]
Diameter, Longest geodesic path	Length of the longest finite geodesic path anywhere in the network.	Measure of information and substance flow	[27, 36, 42, 54]
Small world index (SW)	Index based on a tradeoff between high clustering coefficient and short path length, the defining characteristics of small-world networks. Networks with SW > 1 are said to have more "small-worldness".	Microbial network topological property. Small-world microbial networks suggests that any two members in the community could interact with each other through a few intermediaries.	[6, 55]
Clustering coefficient, Transitivity	Average probability that two nodes neighbors of a third node are also connected between each other.	Presence of tripartite relationships (e.g., higher-order biological interactions) within the community; possible measure for redundance.	[4, 27, 36, 37, 42, 53]
Modularity, Assortativity (when normalized)	Quantification of compartmentalization into subgroups. Loosely speaking, high modularity means that there are more edges within groups and fewer between groups.	Modules/Clusters have been interpreted as niches; shared ecological functions among taxa; spatial compartmentalization; similar habitat preferences; measure of community stability (ecological resilience and resistance)	[1, 3, 18, 27, 36, 37, 42, 56–60]
Triad motifs and Balanced triads fraction	Motifs are overrepresented subnetworks (patterns). Triad motifs are classified by balanced or imbalanced based on the relationship types (positive or negative).	Motifs can be relevant in information flow (e.g., quorum sensing); potential biomarkers for microbiome perturbed state.	[1, 36, 61–63]

https://doi.org/10.1371/journal.pone.0259756.t001

In an attempt to capture the most relevant information from a microbial community in their co-occurrence network and to try to overcome some common issues, we have developed the MicNet toolbox, an open source code to create, analyze and visualize microbial co-occurrence networks. We implemented UMAP [64], a dimension reduction algorithm which has been previously used to identify unique clusters of data in several genomic projects [65, 66], given that it is both scalable to massive data and able to cope with high diversity [64]. Moreover, we coupled UMAP with different types of projections and HDBSCAN [67], an unsupervised clustering algorithm, able to identify both local and global relationships, as well as filtering out noise. Finally, we used and enhanced version of SparCC, a compositionally aware algorithm, to infer correlations for network construction [17, 22]. Additionally, the toolbox includes several analyses of network theory to inspect the topological properties, robustness, structural balance, communities, and hub nodes that arise in microbial co-occurrence networks. The development of the MicNet toolbox, as an integration of several analyses, attempts to provide an easy-to-use and straightforward implementation towards a comprehensive description of potential local and global patterns for a better understanding of microbial community systems.

Design and implementation

Python implementation

The code of the MicNet toolbox was built using python 3.9 [68]. The MicNet toolbox uses several standard packages in the Python ecosystem for matrices (pandas v1.3.2 [69, 70], numpy v1.20.3 [71], and dask v2021.8.0 [72]), to improve performance (numba v0.53.1 [73]), for temporary storage (h5py v3.2.1 [74]) and to create visualizations (bokeh v2.3.3 [75]). UMAP and HDBSCAN were implemented using packages umap-learn v0.5.1 [64] and hdbscan v0.8.27 [67], whereas network analyses were performed using functions from the networkx v2.6.2 package [76]. In the following sections we explain the different components implemented in the MicNet toolbox which are summarized in **Fig 1**.

Input data

MicNet toobox input data consists of relative abundance/compositional datasets from high-throughput sequencing methods, such as metagenomics or metabarcoding. Prior to building an abundance data table, raw assembled sequences should be OTU/ASV clustered. MicNet toolbox currently supports abundance data as .tsv files (separated by tabs) or .csv files (separated by commas). In the web dashboard, input abundance data table is filtered by default, removing singletons (< 5 total counts among all samples) and unique (only appearing in one sample) entries. If the user desires, singleton filtering could be deactivated. For SparCC and UMAP/HDBSCAN the first column of the table should contain the OTU/ASV ID and the following columns the abundance data. Taxonomic information is optional in the input abundance table since poor taxonomic assignment might hinder the interpretation of results. Hence, the user might prefer to work only with ASV/OUT IDs. If the user wishes to include taxonomic information in the resulting output, taxonomic annotation for the given OUT/ASV can be included in the second column, with ";" as a delimiter between each taxonomic hierarchy (e.g., Bacteria;Cyanobacteria;Cyanophyceae;Nostocales;Nostocaceae;Nostoc). In the case of network analyses, the correlation matrix output by SparCC should be input alongside the UMAP/HDBSCAN output datafile.

Data visualization

UMAP and HDBSCAN implementation. A common first step when visualizing high-dimensional data is applying a dimensionality reduction technique. In this toolbox, we

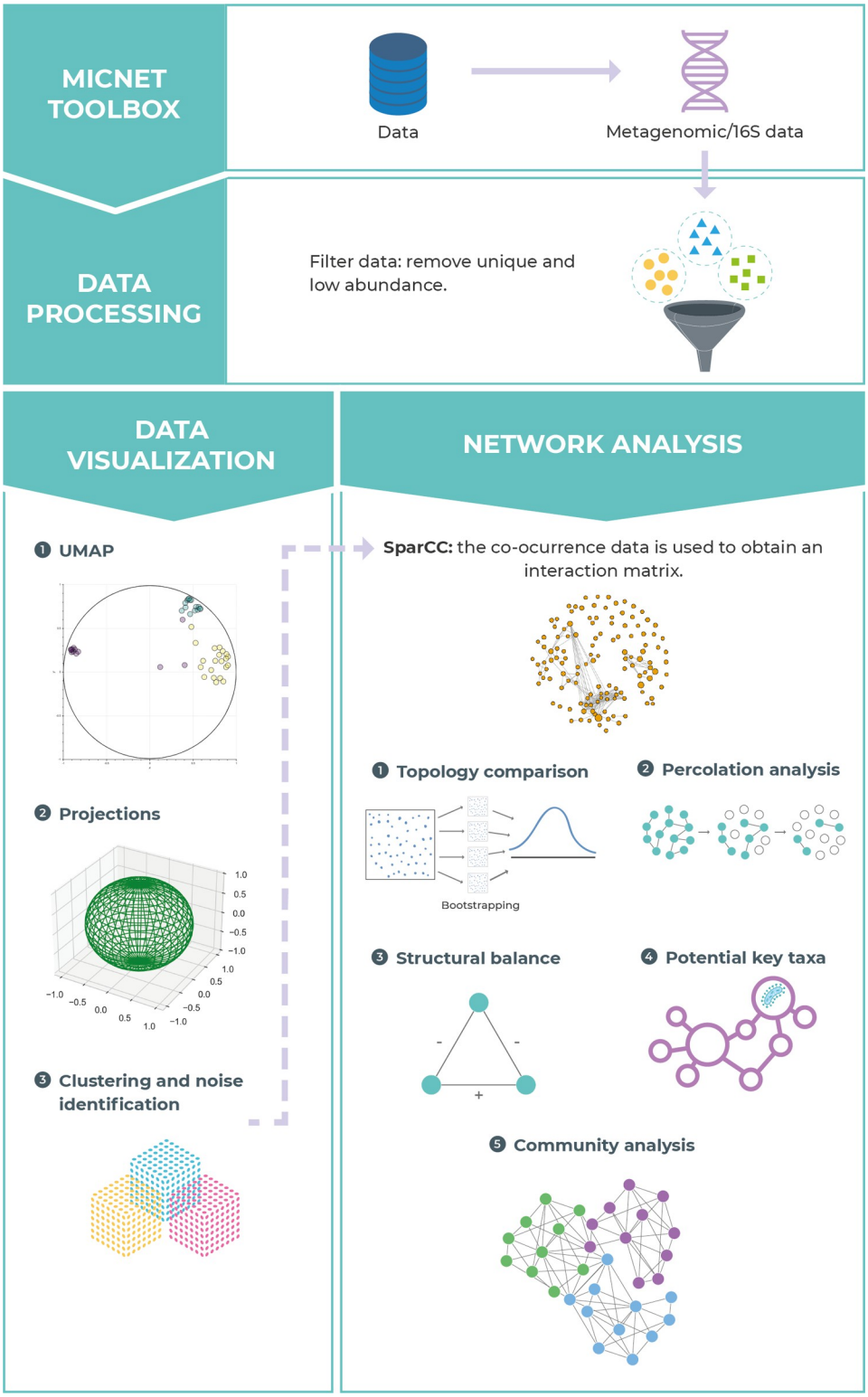

Fig 1. Overview of MicNet toolbox. MicNet Toolbox was designed for visualizing, creating and analyzing microbial networks obtained from compositional data (obtained through high-throughput sequencing methods such as metagenomic/16S surveys). Data filtering for singletons and low abundance taxa/ASV/OTU is supported if required. MicNet Toolbox includes two independent main modules: a data visualization module which uses UMAP and HDBSCAN to find local patterns in the data, and a network analysis module which implements an enhanced version

of SparCC to create a co-occurrence network; network analyses such as topology comparison, and community analysis are included in the aforementioned network analysis module. If desired by the user, HDBSCAN clustering output can be integrated into the network analysis module.

https://doi.org/10.1371/journal.pone.0259756.g001

implemented UMAP (Uniform Manifold Approximation and Projection), a non-linear dimension reduction technique that favors local data preservation, rather than global data, allowing a better identification of finer scale patterns [64]. We coupled UMAP with HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), a hierarchical clustering algorithm that partitions the data based on their density [67, 77]. This clustering technique has been shown to perform well when performed in combination with UMAP dimension reduction [78] and it has been tested with several dataset types, including genetic data, grouping genes into the correct known classes [79]. Thus, we implemented HDBSCAN on the data obtained from UMAP analysis, which represents the abundance data in a reduced space of two dimensions. HDBSCAN analysis not only classifies OTUs/ASVs as belonging to a cluster but also as noise (defined as any point that was not selected in any of the clusters) and outliers (detected with the GLOSH outlier detection algorithm, which works with local outliers).

When running UMAP we set as default a minimum distance of 0.1, number of components of 2, a Helligner output metric and number of neighbors of 15. In the case of HDBSCAN, the default parameters when running MicNet are Bray-Curtis metric, minimum cluster size of 15, minimum sample size of 5 and quantile limit of 0.9 for outlier detection. For our different datasets, the number of neighbors (UMAP), and minimum cluster/sample size (HDBSCAN) parameters were set depending on the input microbiome dataset. We used Bray-Curtis dissimilarity as the distance metric for UMAP, as it is a standard metric for biological datasets. In the web dashboard, UMAP and HDBSCAN parameters can be modified by the user to visualize the results according to each set of microbiome data.

New implementation of SparCC

To obtain the correlation matrix from relative abundance data, we implemented a modified version of the SparCC algorithm, a robust approach to discard spurious correlations when dealing with compositional data [17, 22]. Although the original SparCC algorithm was not altered, several modifications were made to improve and scale the SparCC estimation matrix. We made three main changes to the original code; first, the code changed from Python version 2.7 to 3.9. Second, we use Numba and Dask in some parts of the matrix processes, namely functions or operations, with two main improvements: parallelization of operations and scalability in the size of the estimated matrices. Finally, the original SparCC version stores each estimation step in RAM, as arrays in NumPy. Although storing in RAM is efficient for small data sets, with large data the required memory increases rapidly depending on the interaction numbers and the size of the dataset. Thus, we store each estimation step on disk as hdf5 binary format. These changes made it possible to calculate the SparCC estimates with good time performance in easily accessible computing resources. SparCC p-value test on the inferred correlation was not modified, it was calculated with a Monte-Carlo simulation (with default $n = 50$) as done by **Friedman & Alm (2012)** [17] and the default value is to calculate one-sided p-values, although this can be modified by the user.

To set SparCC parameter values, we perform a parameter search in our more complex study case from the communities reported by **Espinosa-Asuar et al. (2021)** [80]. We performed an independent parameter sweep on each parameter, varying the exclusion threshold from 0 to 1 in steps of 0.1, the number of iterations from 10 to 100 in steps of 10 and the

exclusion number from 10 to 100 in steps of 10. For each parameter value, we calculated the number of correlations found and selected the parameter value when this number stabilized (S1 Fig). The final parameter values used in the databases presented here were: 50 iterations, an exclusion number of 10, exclusion threshold of 0.10 and 100 simulations for p-value calculation. However, this can be modified by the user both in the dashboard and when running the code from the github repository.

Network analyses

Network analyses were performed to characterize both the overall structure and the local interactions of the microbial co-occurrence network, in which each OTU/ASV is represented as a node and the correlations found by SparCC as undirected weighted edges, such that an edge between two nodes implies a relationship between the two corresponding OTUs/ASVs. Given that most network analyses can only handle positive interactions, we normalized the SparCC correlation matrix from -1 to 1 to a range from 0 to 1, except for the structural balance analysis which directly uses the positive and negative correlation values. It is important to note that the normalization of values does not change the distribution of the correlation values, they are just mapped to another scale to allow running network analyses that do not handle negative values. All available network analyses in the MicNet toolbox are described as follows.

Network topology comparison. Networks have several large-scale structural measurements to characterize their topology. For this purpose, MicNet calculates the following structural metrics: 1) network density, using networkx function nx.density, 2) average degree, calculated as the mean of all nodes degree using numpy mean function, 3) degree standard deviation, using numpy std function, 4) ratio of positive-negative relationships calculated simply as the number of positively weighted edges divided by the number of negatively weighted edges, 5) average shortest path length using nx.average_shortest_path_length, 6) clustering coefficient nx.average_clustering function, 7) modularity, calculated with function nx.modularity, using as network modules those obtained with the nx.greedy_modularity_comminuties algorithm, and 8) the diameter, which was calculated using nx.diameter function. Finally, we have added a custom function that calculates a small-world (SW) index as suggested by **Humphries & Gurney (2008)** [55]. The SW index is calculated as:

$$SW = \frac{cc/cc_{rand}}{l/l_{rand}}$$

Where l and cc are the average shortest path length and clustering coefficient of the experimental co-occurrence matrix, respectively. Analogously, l_{rand} and cc_{rand} are the average shortest path length and clustering coefficient, accordingly, of a comparable random network with the same number of nodes and density. The random network was built using the function nx.erdos_renyi_graph. This is done several times, with default $n = 50$, and the mean value of SW is returned.

MicNet includes the computation of the distribution of several of this large-scale metrics under the assumption that the underlying topology is: 1) a random Erdos-Renyi network [81] built using function nx.erdos_renyi_graph, 2) a small world Watts-Strogatz network [82] built using nx.watts_strogatz_graph function, or 3) a scale-free Barabási-Albert network [83] built using nx.barabasi_albert_graph function. A short description of these canonical topologies can be found in Fig 2. This allows the comparison of the query data against the different topologies. These simulated networks are built with the same number of nodes, density and average degree as the experimental data, and correlations are drawn from a uniform distribution from -1 to 1. Finally, the simulated networks are made symmetrical to be comparable with the SparCC output correlation matrix.

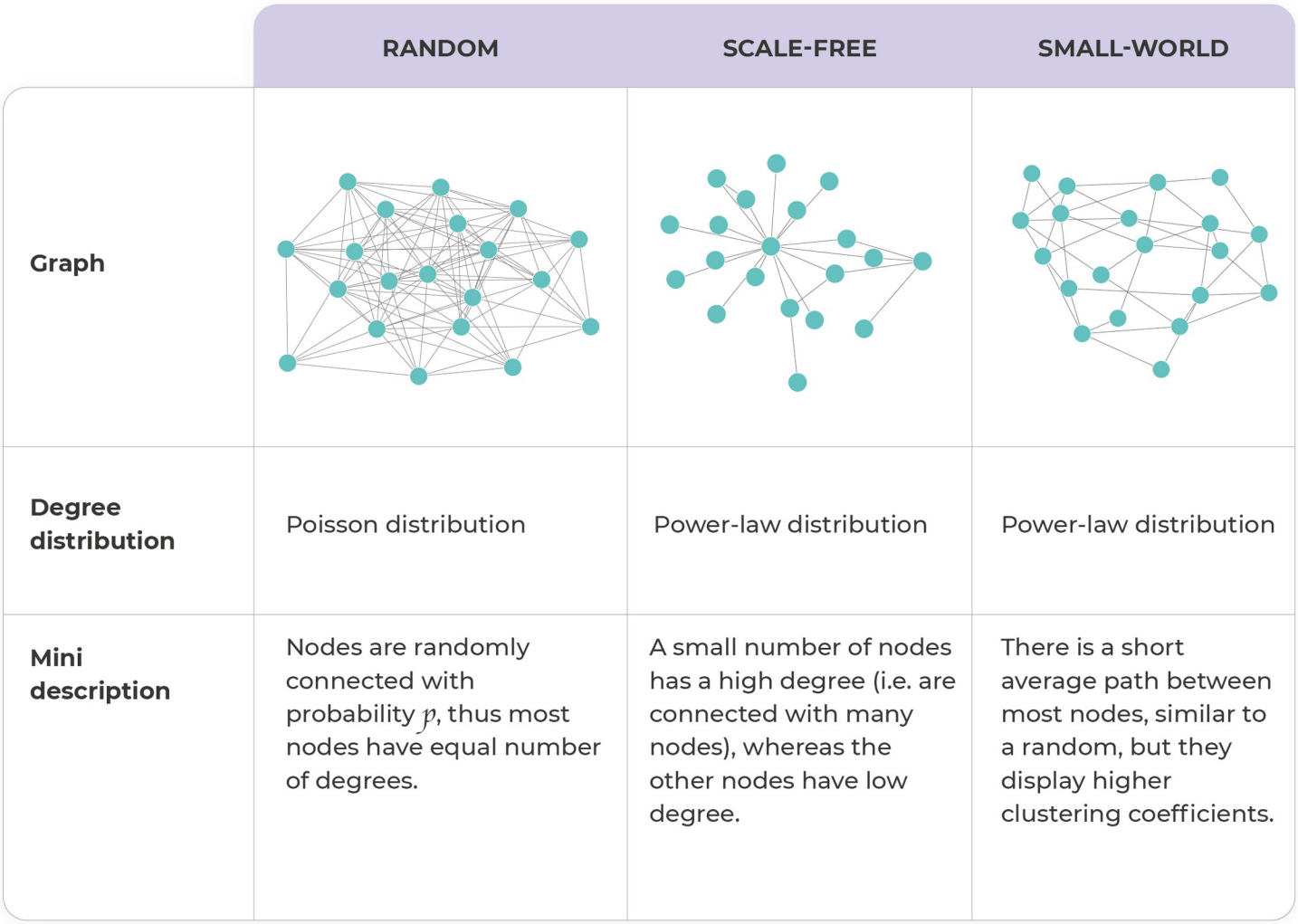

Fig 2. Description of three canonical network topologies. Random, scale-free, and small-world topologies have different network properties. A brief description along with their type of degree distribution in shown.

https://doi.org/10.1371/journal.pone.0259756.g002

Degree distributions can also be used to discriminate between network topologies. Thus, we have included in the MicNet toolbox a function that plots the Complementary Cumulative Distribution Function (CCDF) of the degrees of the given network and compares it with the CCDF of a simulated comparable random, scale-free and small-word network on a log-log scale. To calculate the CCDF we first divide the range of the degrees into bins; for each bin we obtain its probability as frequency/total. We used this discrete definition of the Probability Density Function (PDF) to calculate the Cumulative Density function (CDF) as the cumulative sum of the PDF:

$$CDF_i = \sum_{k=1}^{i} x_k$$

where x_k is the PDF of each bin k previously defined, such that the CCDF is calculated as:

$$CCDF = 1 - CDF$$

We used CCDF since it has been suggested as an easier way to visualize the difference between degree distributions [84, 85].

Community analysis. To analyze subnetworks, we used two ways of dividing the network into subunits: 1) We used Louvain method to detect communities in networks (using python-louvain library [86]), and 2) we used the clusters found with clustering algorithm HDBSCAN. Each subnetwork's nodes and edges were isolated as a subnetwork using function nx.subgraph, and then for each we obtained the following metrics (also used for network topology analysis): total number of nodes, total number of relationships, density, average degree, and clustering coefficient. Finally, we characterize the diversity of each subnetwork by looking at the total number of different taxa present in each subnetwork at the phylum level.

Percolation analysis. Depending on the network structure, networks can be more or less robust to disruptions. Networks are usually formed by a giant component, which includes between 50% and 90% of the nodes. The formation and dissolution of this giant component is called percolation transition in network theory [87]. The percolation approach consists of removing nodes and their corresponding edges and analyzing how much the network's properties are disrupted [88]. The percolation simulation implemented in MicNet consists of n iterations; in each iteration a percentage of the nodes (with default value of 0.1, but this can be specified by the user) is removed along with all of their edges. After removing the nodes and corresponding edges, the following metrics are calculated for the remaining network: density, average degree, number of connected components (this last one calculated using the nx.connected_components function), size of giant component, fraction of nodes belonging to the giant component, the communities found by the python-louvain algorithm and the network modularity. We implemented several percolation approaches: 1) random percolation, in which nodes are removed randomly; 2) centrality percolation, in which nodes are removed by centrality (whether degree, closeness or betweenness centrality), higher values first; and 3) group percolation, where groups of nodes are removed according to a grouping variable provided, such as taxonomic groups or HDBSCAN groups. Consequently, network robustness to different types of disruptions could be assessed by looking at changes in different network metrics.

Structural balance analysis. Structural balance analysis finds all triangle motifs in the network, that is, nodes that are interacting in triads, and then classifies them as balanced based on the simple analogy that "my friend's friend is my friend" and "my friend's enemy is my enemy" [89, 90]. This leads to classifying triads of interactions as balanced if they meet this criterion, or as imbalanced otherwise **Fig 3**. A network is considered to be balanced if most triads found in it are balanced. To calculate structural balance, we found all triads in the network using function nx.cycle_basis, and keeping only the cycles of length three. Then, we classified the found triangle motifs into balanced or imbalanced, depending on their mutual correlations. The output of the analysis is a percentage of balanced and imbalanced triangles with respect to all triangles found, and the exact percentage for each of the four types of triangles displayed in **Fig 3**.

Potential key taxa analysis. Four different centrality measures were implemented to characterize each node (OTU/ASV): degree centrality using function nx.degree_centrality, betweenness centrality using function nx.betweenness_centrality, closeness centrality using function nx.closeness_centrality and PageRank using function nx.pagerank. When running the code, a dataset is returned with each centrality metric for each node.

Dashboard interface

In addition to the freely accessible source code at the github repository, we have also developed a web dashboard at http://micnetapplb-1212130533.us-east-1.elb.amazonaws.com that can be

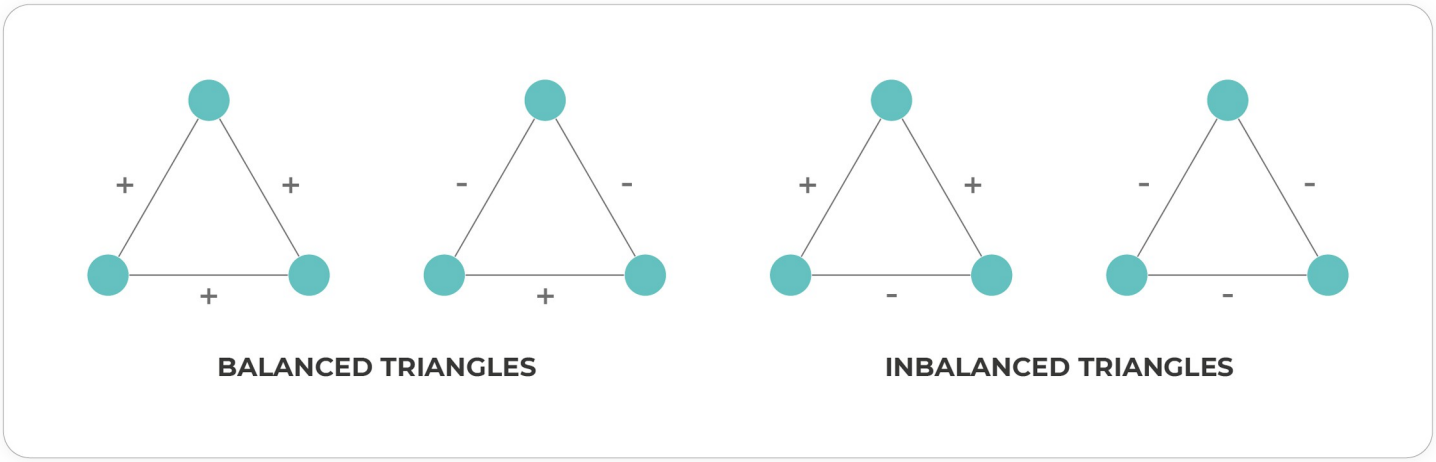

Fig 3. Classification of triad motifs according to structural balance theory. Balanced or imbalanced criterion is assigned according to their mutual correlations.

https://doi.org/10.1371/journal.pone.0259756.g003

used to run most of the analyses presented here. The dashboard consists of three main parts: 1) UMAP and HDBSCAN, 2) SparCC and 3) Network analyses. In the first component, the raw abundance data should be input (see Input Data section), and several parameters, as well as normalizations, for UMAP and HDBSCAN can be modified as desired. This first component returns an interactive visualization of the UMAP plot along with HDBSCAN clusters identified by color. Both the resulting plot and the file detailing cluster belonging can be downloaded from the dashboard.

The second component is for the estimation of the co-occurrence network with SparCC. In this section, abundance data should be input, and parameters can also be adjusted by the user. The resulting correlation matrix can be downloaded from the dashboard. Finally, the third component includes the post-processing analyses of the co-occurrence network. Thus, the input of this section should be the matrix obtained from the previously defined SparCC component or any square correlation matrix, and the UMAP/HDBSCAN output file. When run, this section allows the user to obtain the large-scale metrics of the network, the structural balance percentages, descriptions of the communities found, and two network graphs where the size of the node indicates degree centrality, green edges indicate positive relationships between two nodes, whereas red edges indicate negative ones. Finally, in the graph named HDBSCAN the color of the nodes refers to the HDBSCAN cluster they belong to; whereas in the graph named Community the colors indicate the color of the community they belong to, based on Louvain clustering algorithm. When the network of interest consists of less than 500 nodes an interactive visualization plot is deployed, but for larger networks a static plot is returned, given limited computational resources.

For the other network analyses presented, such as percolation analysis and topology comparison, or if the user wishes to run the complete pipeline directly with code, we have also provided a package called micnet that can be install via pip. All the functions necessary to run the complete example of Kombucha are available in the micnet package and their usage is explained in a notebook present in the github repository of the MicNet toolbox. It should be noted that, if more computational resources are needed, the dashboard itself can also be run after downloading the MicNet toolbox from github, creating the conda environment and deploying it with streamlit as suggested in the readme file.

Validations

Modified SparCC. We performed two validations with simulated communities. We validated the new version of SparCC on the dataset provided by **Friedman & Alm (2012)** [17], which consists of 50 OTUs in 200 samples drawn from a multinomial log-normal distribution. For this, we compared the real correlations to the estimations performed by SparCC, and then we calculated the RMSE (Root Mean Square Error). Secondly, to corroborate that our implementation of SparCC does indeed scales better to larger datasets than the previous version, we compared the execution times and the RAM consumption between our new version and the previous version of SparCC for datasets containing from 50 to 2600 OTUs.

Biological validation: Kombucha consortium. To further authenticate MicNet Toolbox's approach to analyze microbial co-occurrence networks, we needed to see if biological interactions previously described by experimental work could be replicated in the network. For this, we make use of the kombucha dataset described in **Arıkan et al. (2020)** [91]. Test data was downloaded from the European Nucleotide Archive (ENA) at EMBL-EBI under the accession numbers ERP104502 (https://www.ebi.ac.uk/ena/browser/view/ERP104502) and ERP024546 (https://www.ebi.ac.uk/ena/browser/view/ERP024546). The raw 16S amplicon reads were filtered, processed and annotated with QIIME 2 [92] and DADA2 [93]. Abundance and taxonomy for each ASV cluster was acquired. The obtained abundance table with all samples was filtered, as singleton and unique counts were removed from the data, as suggested by **Berry & Widder (2014)** [25]. Filtering unique and singletons resulted in 48 ASVs. For the visualization module, UMAP parameters were set as follows: number of neighbors of 5, minimum distance of 0.10, number of components of 2 and an Eucledian metric. In the case of HDBSCAN the parameters were: minimum cluster size of 5, minimum sample size of 3 and Bray-Curtis metric. Network construction and network analyses were performed as described in previous sections. The raw data from the kombucha database is in the github repository so that the main results can be replicated and the user could interact with them in the dashboard.

Case study: *Archean Domes*

A 16S amplicon dataset was provided from a highly diverse microbial community named *Archean Domes*. This dataset comes from a microbial mat located in the Cuatro Ciénegas Basin (CCB), Coahuila, Mexico (coordinates 26˚49'41.7" N, 102˚01'28.7" W). The sampling used in this case study, which consists of ten samples along a 1.5 m transect, represents a natural community with more than 6,000 ASVs [80]. Compositional data was acquired as raw reads form 16S amplicon sequencing. Reads were filtered and processed for clustering and taxonomic annotation in QIIME 2 platform, as shown by the authors [80]. Singletons and unique counts were subsequently filtered as suggested, and consequently, 2,600 ASVs remained [25]. The ASV abundance matrices along with a taxonomic annotation for these sequences were used as input for the MicNet toolbox. For the visualization module, UMAP and HDBSCAN parameters were set as follows: number of neighbors of 15, minimum cluster size of 15, and minimum sample size of 5. Network construction and analyses were implemented as described in previous sections.

Results

Validations

The enhanced SparCC. To validate that the modifications performed to SparCC did not affect its performance, we ran our version of SparCC on the dataset provided by **Friedman & Alm (2012)** [17]. We compared our estimated correlation with their true basis correlation

(Fig 4A and 4B). We found an overall RMSE of 0.08, and a consistent value of RMSE when estimating small and large correlation values from the simulated samples (Fig 4C). Thus, although the original pipeline of SparCC was not modified, by implementing several techniques that parallelized different parts of the code, our implementation of SparCC can now be used for large databases in a reasonable amount of time with relatively small RMSE.

We then moved on to characterize the execution times and RAM consumption of the new SparCC version in comparison to the previous one. In Fig 5 we show that for relatively small datasets, that is, those containing less than 1,000 OTUs, both versions do similarly in terms of execution times, with the previous version of SparCC being slightly faster. However, for larger datasets, the execution time of the algorithm increases in an exponential fashion, taking approximately 9 hours to run the largest dataset we had of 2,600 OTUs. In comparison, the execution times of the new version of SparCC scales better to larger datasets, with an execution time of around 2.5 hours for the 2,600 OTUs dataset. The consumption of RAM memory is more less constant in both versions, but there is a higher RAM consumption in the new version of SparCC as a result of the parallelization of some processes of the algorithm. However,

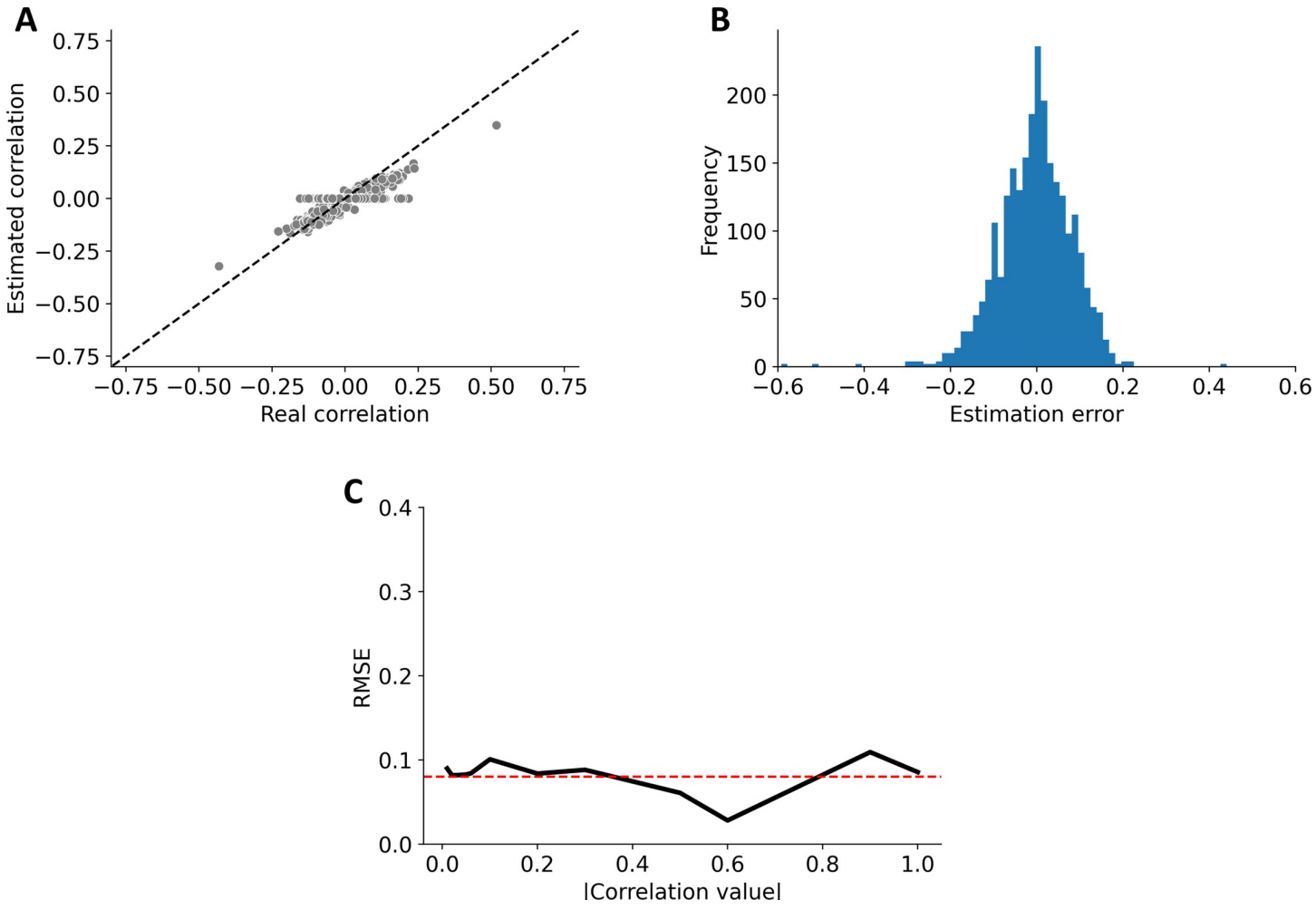

Fig 4. SparCC validation. The modified version of SparCC was validated using the database provided by **Friedman & Alm (2012)** [17]. **A**. Comparison of the estimated correlation with the real correlation. **B**. Histogram of the estimation error produced with the new SparCC version. **C**. RMSE across different absolute values of correlations, the overall RMSE error was 0.08 shown in the dashed red line.

https://doi.org/10.1371/journal.pone.0259756.g004

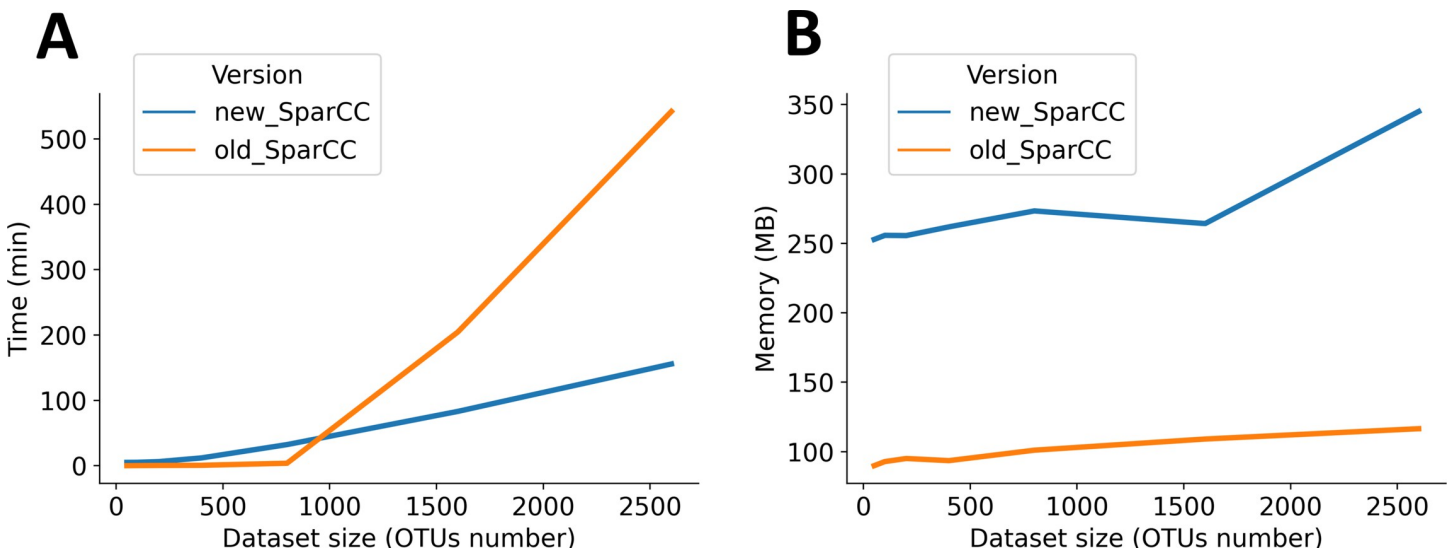

Fig 5. Comparison of execution times and RAM consumption between SparCC version. A. Execution time in minutes for datasets ranging from 50 to 2600 OTUs for the previous version of SparCC (orange) and the modified version (blue), **B.** RAM consumption in MB for the old and new version of SparCC for datasets ranging from 50 to 2600 OTUs.

https://doi.org/10.1371/journal.pone.0259756.g005

we can see that the RAM consumption does not grows exponentially with the size of the dataset and it has the advantage that by parallelizing some processes the execution time is significantly reduced.

Biological validation: The kombucha consortium. To demonstrate how MicNet tools could infer ecological associations, we used a kombucha data set to replicate main global and local behavior. Kombucha is a simple and well-studied microbial consortium of bacteria and yeast, which grow as biofilm due to cellulose production from acetic acid bacteria (AAB) [91, 94], but also develops as a liquid consortium. This consortium has been suggested as a convenient tractable system, whose general cooperative and antagonistic multi-species interactions have been previously described [94–96]. After filtering singletons and unique taxa from the raw data, only 48 ASVs remained in the analysis, corresponding to five annotated bacterial taxa and three fungal taxa.

As MicNet pipeline suggests, first, the community was visualized and analyzed with UMAP and HDBSCAN to uncover global patterns and noise taxa. Clusters from HDBSCAN showed one main group containing almost all ASV (34 of 48), a small group with 12 ASV and 2 ASV classified as noise shown in purple (**Fig 6C**). This could refer to a close-interacting community where highly stratified interactions are not common. Since the kombucha community has similar compositions in both homogeneous liquid and biofilms [91], physical closeness between all organisms is expected; this was reflected in the formation of one main group with the HDBSCAN algorithm. We then obtained the co-occurrence matrix of the kombucha samples using our modified version of SparCC. **S2 Table** shows the main metrics of the kombucha network and **Fig 6A and 6B** show the resulting co-occurrence network.

The resulting co-occurrence network was indeed a relatively connected one, with a connectance of 0.23. This is reflected in the high average degree, which indicates that each ASV is related on average with around 10 ASVs out of the 48 that are present in the community. Highly connected networks could point to a more homogenous environments, including liquid consortiums and slightly stratified biofilms in which kombucha develops, as opposed to more stratified environments, such as soil and microbial mat communities. The kombucha

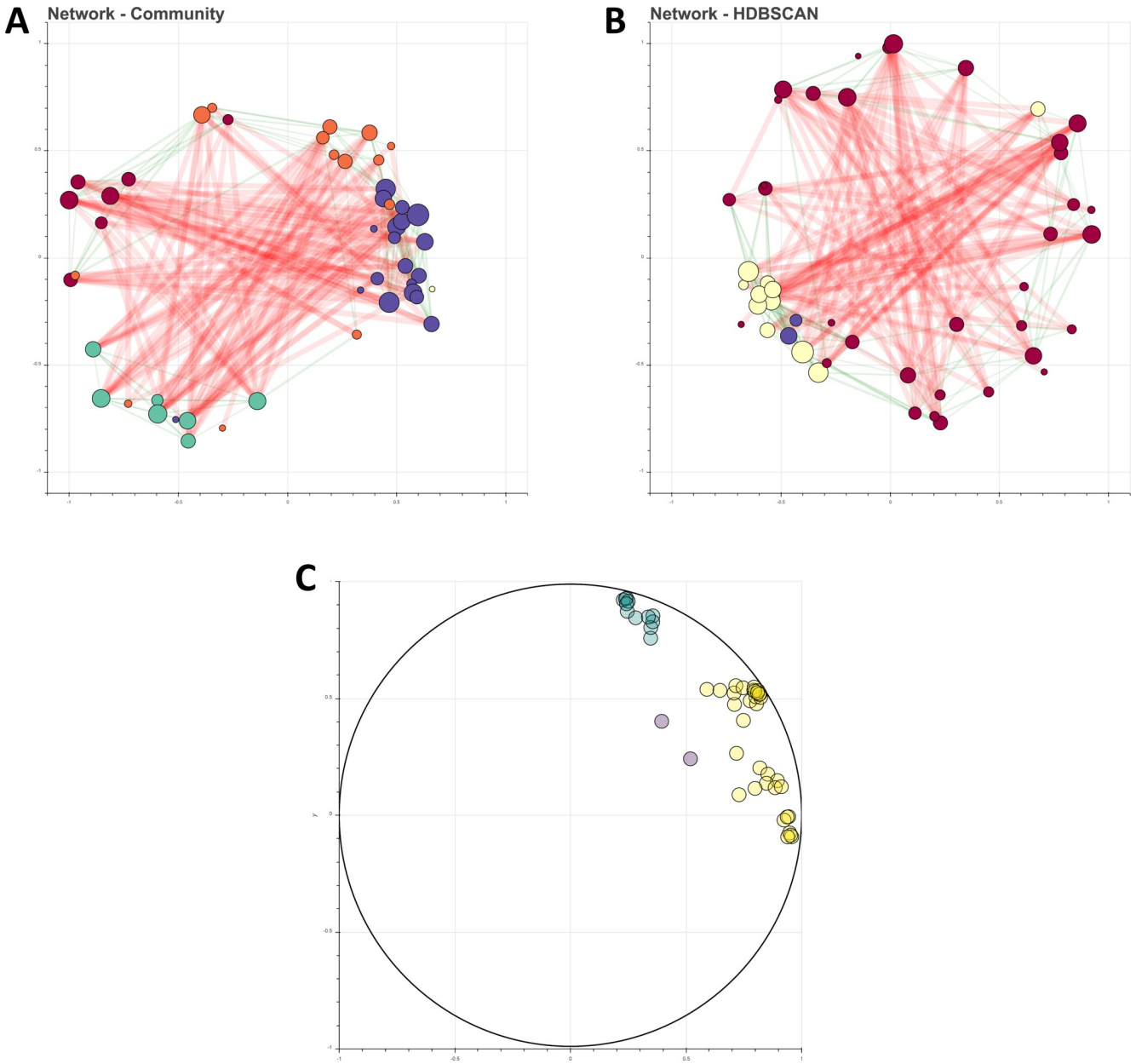

Fig 6. Kombucha microbial network. A. SparCC co-occurrence network, where colors indicate Louvain groups. **B**. SparCC co-occurrence network, where color indicates the resulting HDBSCAN clusters. In purple is shown the group depicted as noise. **C**. UMAP and HDBSCAN results show one main group and a smaller one of 12 ASVs.

https://doi.org/10.1371/journal.pone.0259756.g006

consortium is the result of a metabolic interplay between its microbial consortium, and though cheaters and antagonistic relationships are known [97], more cooperative relationships have been reported [94, 95]. In fact, the network did show slightly more positive (55%) than negative relationships (45%), and a major contribution due to mutualist or commensalist interactions is expected.

We perform a topology comparison analysis to explore if the kombucha consortium fits within three canonical networks. Based on metrics bootstrapping from comparable networks, the kombucha network's degree standard deviation, average path length, SW index suggest

some degree of small-worldness [55] (Fig 7A). Additionally, comparison of kombucha CCDF with these simulated networks' CCDF suggest that the degree distribution of kombucha is not following particularly any specific canonical topology, although it appears to fit best with a random network (Fig 7B). These results are consistent, as kombucha medium is a non-stratified environment when it develops as liquid consortium, explaining the random properties, but still a stratified microbial consortium when developed as biofilm, explaining the scale-free properties. Thus, kombucha community plausibly shows properties in which microorganisms are adapted to biologically interact with each other in both homogeneous or heterogeneous (to some degree) structures. Thus, real data rarely conforms to a single mathematical topology, but comparing the data's metrics and degree distribution can hint towards one or another structure and give us an idea of which metrics are better to discern between topologies.

Although some biological interactions in the kombucha consortium still need to be confirmed, the global interplay between AAB and yeasts is well-known. In kombucha fermentation, yeast produce invertase which cleave sucrose into glucose and fructose, and further use fructose to produce ethanol. Ethanol is a noxious compound for the consortium. Hence, as a mechanism to regulate ethanol concentrations in the media, AAB transforms glucose and ethanol into gluconic and acetic acid, respectively, exhibiting a straightforward case of syntrophy [91, 94]. This biological interplay was depicted in the ASV classified as *Zygosaccharomyces baillii*, the most abundant yeast in the sample, and the ASV with the greatest number of interactions (S3 Table). Second to *Z. bailii*, an ASV corresponding to *Komagataeibacter europaeus*

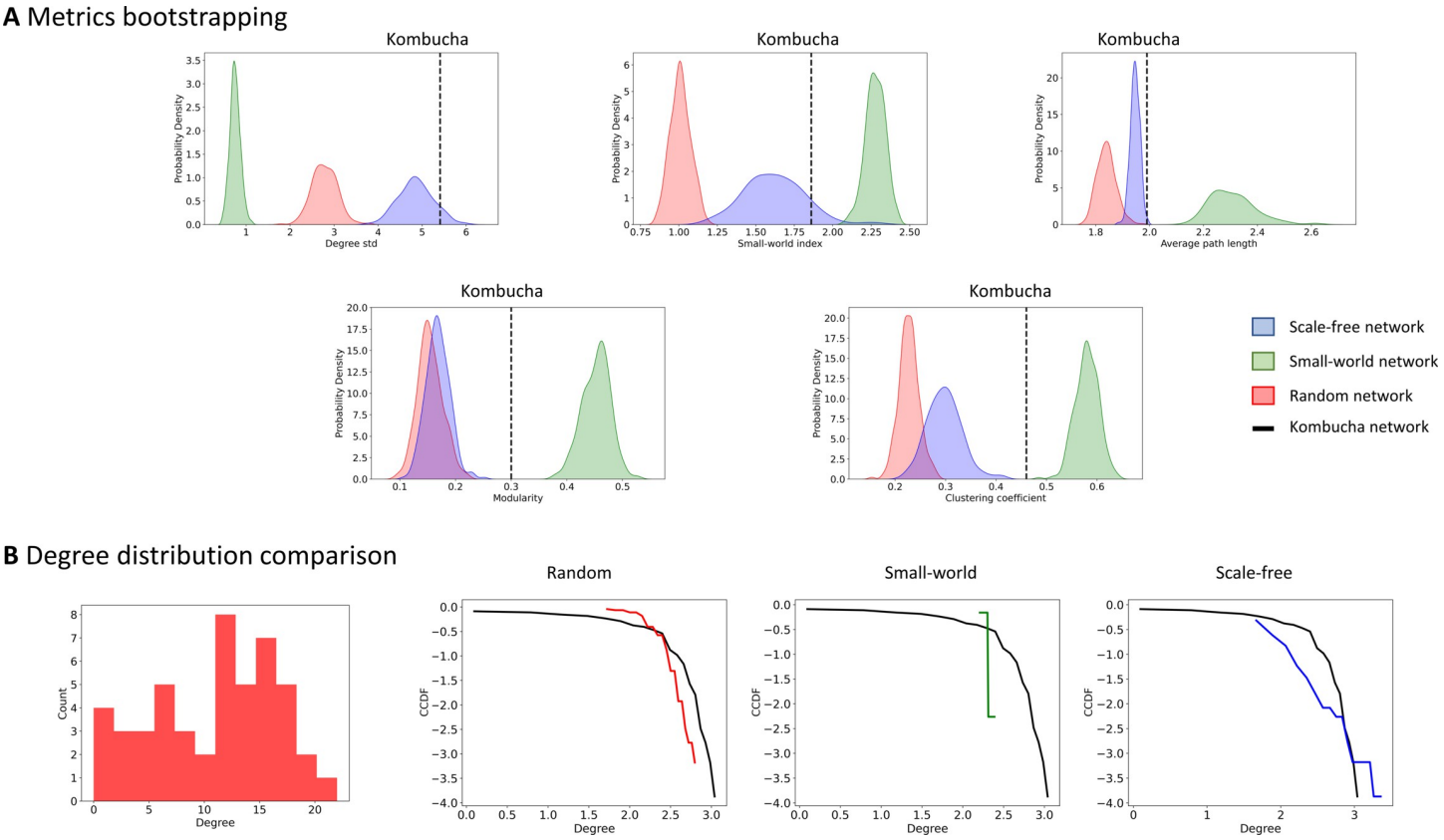

Fig 7. Kombucha topology comparison. A. Distributions obtained from simulated random (red), scale-free (blue) and small-world (green) networks and its comparison to the metrics found in the kombucha network for: degree variance, modularity, average path length, small-world index, and clustering coefficient. **B**. Degree distribution of the kombucha network. We also show the comparison of the kombucha CCDF with a random network CCDF, a small-world network, and a scale-free network.

https://doi.org/10.1371/journal.pone.0259756.g007

(an AAB) appears to be a key central taxa based on each centrality metric. According to the inferred correlations by the enhanced SparCC (correlation matrices at the genus and species level are provided in the in S1 File), the *Z. bailii* correlated to every other ASVs, and mainly positively co-occurring with *Komagataeibacter* (0.1812), and negatively correlating with *Cortinarius* (-0.2081). Actually, for the species within the *Komagataeibacter* show the highest positively mean correlation with *Z. bailii*, particularly *Komagataeibacter europeus* (0.5397), reflecting the cooperative interplay between yeast and AAB. Additionally, *Komagataeibacter*, an AAB genus, produces acetic acid which inhibits growth of other species, except for *Z. bailii* [91, 98]. More specifically, it has been reported that *K. rhaeticus* is one of the main producers of acetic acid compared to other microbial species [99]. As expected, *Komagataeibacter* genus shows negative interactions against some of other species different from their own genus, such as *Variovorax* that is negatively correlated, which might be explained by with its growth-inhibiting capability.

From mean relationships within the taxa present in kombucha, ASVs from the same species tend to have more positive relations between themselves, and this was reflected in the community clustering analysis, where we found that communities were appreciably grouped per species, according to the Louvain method (**S2 Fig**). Clustering resulting from phylogenetic relatedness is common in microbial data, and it may reflect niche overlapping to some degree [4, 31]. From the 5 communities predicted with the Louvain method, one of them is considered as noise as it consists of just 1 ASV. HDBSCAN group composition is variable, as most ASV belong to just one group (**S3 Fig**). Nonetheless, the smaller group with 12 ASV's from HDBSCAN is particularly interesting, as it includes most of *K. europaeus* and the *Z. bailii* ASV, probably depicting the core syntrophic interactions. This potential core group is similarly shown as a Louvain method group. Main metrics for each group of the community analysis (via Louvain or HDBSCAN groups) are reported in **S4 and S5 Tables**.

Another aspect of kombucha interactions is that even though yeast could be an important player in the metabolic interplay with AAB, AAB are not fully dependent on them for substrates, characterizing their interaction as some class of non-strict parasitism [96]. In the co-occurrence network, we found evidence that *Z. bailii* was indeed considered a key organism given its high centrality metrics, but in the percolation analysis where nodes were removed by degree centrality (beign Z. bailii), there was not a breakdown of the network (nor in the network density, the average degree, the number of components or the number of communities, as shown in **S6 Table**, along with other percolation analyses performed). In contrast, percolation analysis where the nodes are removed depending on their genus exhibit a network breakup of several components when most of *Komagataeibacter* nodes were removed. Even by removing six *Komagataeibacter* nodes, the network is disrupted into three components, further supporting the relevance of *Komagataeibacter* in the kombucha network. Percolation analysis and centrality metrics are consistent in positioning the *Komagataeibacter* as a crucial genus to the community, and this can be biologically understood due to 1) their independence (to some degree) from yeast to thrive, 2) their cellulose production capability (as a mechanism for protection and resource storage [94]), and 3) as regulators on ethanol concentration.

Case study: *Archean Domes* microbial mats

To further evaluate the performance of MicNet as a high-throughput toolbox capable of analyzing a highly diverse and complex environment, we tested it on a compositional dataset of ten samples from a microbial mat in Cuatro Ciénegas, Mexico, in the Chihuahuan Desert [80]. This microbial mat, The *Archean Domes*, thrives in a fluctuating hypersaline pond which has

been previously described as hyperdiverse [80, 100]. Like every microbial mat, it is a stratified community with an intricate metabolic interplay between their organisms [101]. **Espinosa-Asuar et al. (2021)** provided us their microbial data set of 6,063 ASVs, as they have reported. To begin the pipeline, the abundance matrix for all ASV was filtered to exclude low abundance and unique ASVs, remaining 2,600 amplicon sequence variants for the analysis.

First, we search for global patterns and local clustering within the ASV abundance matrix with UMAP and HDBSCAN. The community was grouped into 49 groups, with a mean of 50 nodes in each group (Fig 8C). With this approach, the HDBSCAN algorithm allowed us to

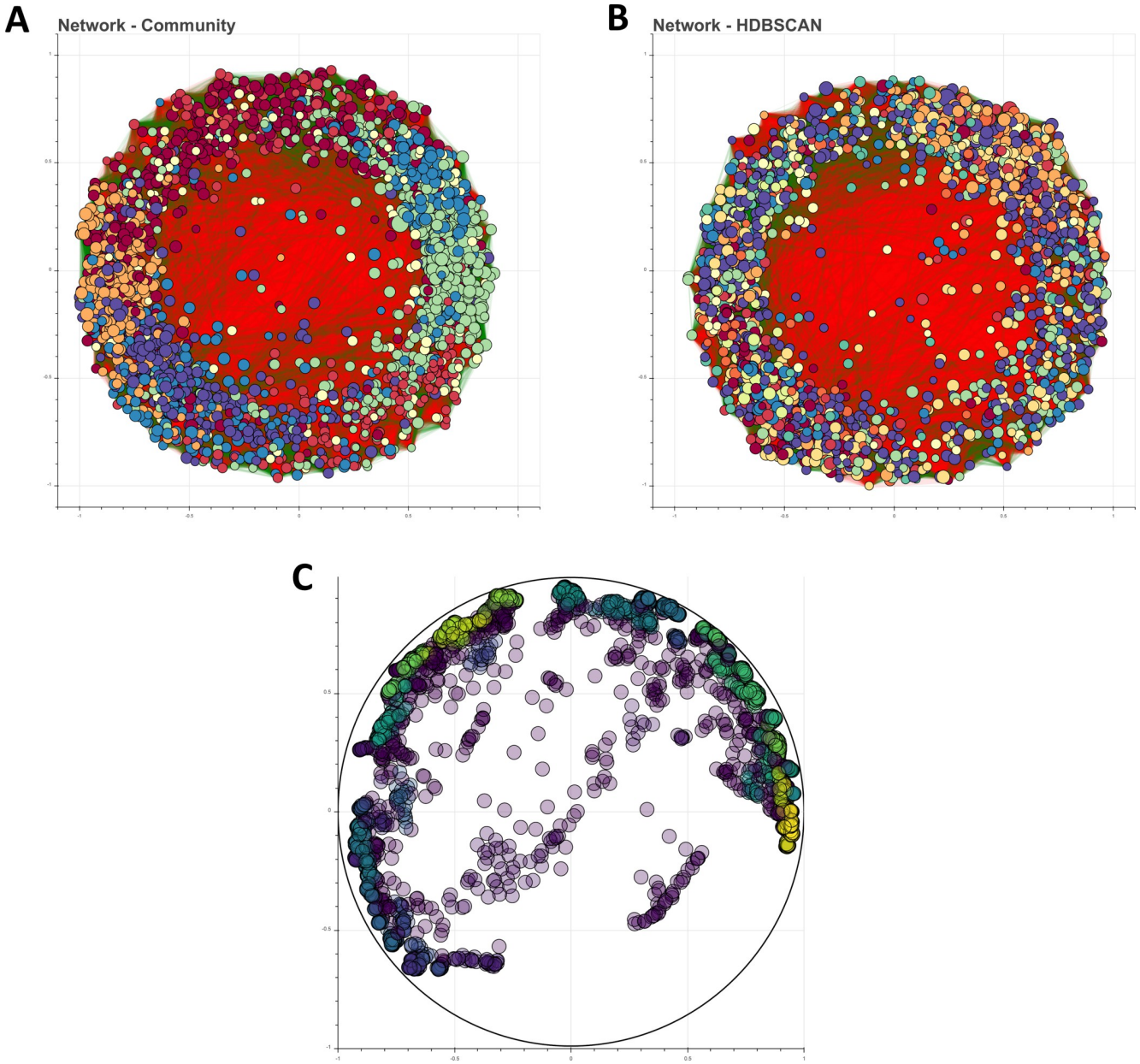

Fig 8. ***Archean Domes*** **biological network. A**. SparCC co-occurrence network, where different colors indicate Louvain groups. **B**. SparCC co-occurrence network, where color indicates the resulting HDBSCAN clusters. **C**. UMAP and HDBSCAN results show 50 clusters and several ASV classified as noise and outliers.

https://doi.org/10.1371/journal.pone.0259756.g008

categorize organisms as noise or outliers, as 553 ASV did not group with any cluster and were categorized as noise, and 260 behaved as outliers.

Afterwards, we computed the correlation matrix with the modified version of SparCC. Main large-scale characteristics of the network are shown in **S7 Table**. With the 2,600 ASV we recreated a network with 463,609 interactions with all of them aggregated in a big but sparsely connected network (density = 0.137) (**Fig 8A and 8B**). This value is consistent with the interaction density of an antagonist network between 37 Gammaproteobacteria strains isolated from water samples of *Churince* (reported density = 0.14), a lake situated also in the CCB [102], further suggesting the fact that microbial mats' networks are often sparse [1, 31].

Network topology comparison for the *Archean Domes* network display how high-complexity systems do not fit greatly simplified theoretical models. This was shown in the metrics bootstrapping from simulated comparable networks, where most *Archean Domes* metrics fall in between a scale-free and small-world network distributions (**S4 Fig**). Particularly, degree standard deviation, modularity, SW index, and clustering coefficient suggest that *Archean Domes* possess intermediate properties between scale-free and small-world networks. On the other hand, the average path length of 1.86 is typical of a scale-free network. These results show that this community is not randomly assembled, and that as expected from microbial data, the network is likely an intermediate between a scale-free and a small-world network [31, 55].

Potential key taxa analysis based on node centrality was performed for this high-complexity network. An ASV corresponding to a bacterium from the order MSBL9, class Phycisphaerae, phylum Planctomycetes, exhibit the highest centrality values, regardless of the centrality measure employed (**S8 Table**). This bacterial class has been previously described in hypersaline microbial mats as anoxic, fermenting, halotolerant and halophilic microorganisms [103, 104]. Looking at other top centrality nodes, most of them were associated to unassigned sequences, except for a node member of the class Parcubacteria and another one from the genus *Imperialibacter* (**S8 Table**). Moreover, unknown taxa as central nodes further suggest the relevance of "microbial dark matter" in ecosystems, and particularly, in hypersaline environments like the one studied here [105].

Local correlations between nodes at different taxonomic levels was inspected (correlation matrices at the phylum and species level are provided in the S1 File). From the 463,610 total relationships in the network, we only highlight some of them which might be insightful. At the phylum level, ASVs from Proteobacteria, the most abundant phylum in the sample, have positive mean correlations with most of the phyla, except Nanoarchaeota, Dependentiae and other unassigned prokaryotes. Cyanobacteria, a key phylum in microbial mats, on the other hand face more negative mean correlations with other phyla, including Synergistetes, Acetothermia and Chloroflexi. These negative correlations prospectively originate from overall antagonistic interactions or different niche requirements (such as wavelength, carbon metabolism or temperature adaptation for potential chloroflexi-cyanobacteria associations [106]). Lower taxonomic associations could be inspected. For example, the most central hub taxa (unassigned MSBL9, class Phycisphaerae) positively co-occurs the most (0.5903) with a bacterium from the genus *Dehalobium*, while negatively co-occurs the most with a deltaproteobacteria from the Syntrophobacteraceae (-0.6177).

Although most ecological associations (particularly biological interactions) are analyzed by pairs, species also associate and interact involving more than two organisms, which are vital for ecosystem diversity [107]. Triad motif identification and structural balance theory attempts to address this issue in microbial networks. For the *Archean Domes* network, we identify that most of the triads are balanced, with a highly balanced triad fraction of 0.9995. One property of structural balanced networks is group division, wherein all intergroup ties are negative, and all intragroup ties are positive [108]. Potentially applied to microbial ecology, we suggest that

structural balance could reflect niche differentiation, in conjunction with other network metrics and properties that have been previously described as useful [1, 4, 109].

Furthermore, we carried out a community analysis to inspect subnetwork properties. Our two clustering methods, Louvain and HDBSCAN groups, display contrasting results, which could reflect different ecological structures. Louvain grouping algorithm resulted in 7 network communities with mean total nodes of 371.43 and mean density of 0.4269, while HDBSCAN showed 50 clusters with mean total nodes of 40.94 and mean density of 0.73. Main metrics for each group, for each grouping algorithm, are shown in **S9 and S10 Tables**. Phyla composition within most groups (either Louvain or HDBSCAN) is highly diverse, which could mirror spatial structures or compartments at different scales [3, 59], which is consistent with the stratified structure of microbial mats [101]. **S5 and S6 Figs** display the phyla composition for Louvain and HDBSCAN clustering methods, accordingly.

High-diversity communities are commonly associated with an overall stable state. While the high number of ASV in *Archean Domes* probably reflects a highly diverse system, novel methods to assess ecosystem stability have been suggested. In microbiomes, positive relationships alone are prone to destabilize microbial networks, as they can create highly dependent and vulnerable feedback loops [3, 110]. *Archean Domes* co-occurrence network show a negative:positive ratio of 0.94, thus, positive and negative relationships are evenly present, suggesting an ecologically resilient, resistant and stablished community. Within this scheme, negative relationships in the community might be originating from antagonistic competitive taxa, lower abundance of facilitative taxa (mutualists), divergent niche requirements, or a combination of all of them [1, 3]. Modularity further bolster the stability hypothesis within the microbial mat. Modular groups (or clusters), whether product of biological interactions or habitat preference, plausibly aid in the stability of the system, as fewer links between groups likely ameliorate the spread of local perturbations to other groups. Modularity in *Archean Domes* shows a high value of 0.34, higher than the kombucha network and other published biological networks [3]. Low average path length (1.86 in *Archean Domes* network) likely function as a measure of response capacity to disturbances, hinting about the ecological resilience capabilities of this microbial mat [4, 31]. Finally, Percolation analysis delves deeper into community stability. We performed random, centrality and phylum percolation simulations, from which the resulting metrics are shown in **S11 Table**. While none of the node removals induce the breakup of the giant component, the total number of communities and overall modularity do perceive the effects of network perturbations. With this in mind, some degree of ecological robustness could be reflected by the impact of percolation simulations to the co-occurrence network.

Discussion

MicNet toolbox has shown to be a promising pipeline. Our new implementation of the SparCC algorithm allows larger datasets to be processed without overflowing the RAM in a reasonable time. Furthermore, UMAP and HDBSCAN, relatively new dimension reduction and clustering techniques, are promising in microbial ecology studies since, as suggested in this work, they are useful methods to identify metabolic groups, niche overlapping, or subcommunities. Finally, given the potential of processing co-occurrence networks with a graph theory approach, we have included several network analyses, both new and commonly used, to further describe and understand the resulting networks from SparCC.

One important aspect to have in mind when using the MicNet toolbox is sample size. UMAP and HDBSCAN are known to be quite sensitive to the size of the database used. With a very small dataset (less than 50 OTUs/ASVs) we recommend being cautious at the interpretation level. **Dalmaijer et al. (2020)** [78] have suggested that for optimal use of UMAP and

HDBSCAN, it is recommended to have around 20–30 data points per expected cluster or subgroup. Furthermore, the choice of parameters for these two techniques should be done with careful consideration, in particular the number of nearest neighbors and minimum distance [65]. We hope that the interactive dashboard will help in this aspect, since parameters can be modified in a simple way.

In terms of network topology, not all large-scale metrics of a network should be used to discern between topologies. As we show previously, some metrics, such as diameter, have no use to discern topologies, whereas others, like the small-word index provide useful information [55]. Moreover, it is unlikely that any biological system follows exactly a single topology, as shown by our case study: *Archean Domes*, the hyperdiverse microbial mats in CCB. However, we believe that knowing whether a network tends more towards a random, scale-free or small-world could give insightful pointers about its general behavior. For example, a small-world model (created with the Watts-Strogatz algorithm) suggests that the network will have short path lengths, because it is formed by highly connected clusters, which are weakly connected among each other; whereas a scale-free model (created with the Barabasi-Albert algorithm) suggests that networks will have short average path lengths as well, as a consequence that certain nodes that have very high degree and can act as hubs; in both cases the average distance between nodes would be expected to be small but for different reasons, which could give us insight of key biological network properties [85].

Microbial communities in the light of complexity have shown, once again, the potential in drawing biological conclusions from networks. Nonetheless, the biological interpretation of UMAP, HDBSCAN and network analysis should also be taken cautiously. As we mentioned before, the interpretation of the different metrics obtained is debatable, and there is a high diversity of interpretations and terms (see synonyms in Table 1) used for the same ecological concepts. As for today, researchers are encouraged to correlate network theory metrics to biological significance in an attempt to find fundamental metrics that could be useful to describe a particular biological phenomenon in whichever microbial system. With MicNet we suggest an analysis pipeline including visualization, co-occurrence network creation and postprocessing of the resulting network with graph theory analyses that could be used as a standard method for network analysis, offering an overview of a microbial community and enabling the comparison between different microbial systems. Potentially, this approach promises to aid in the search for biologically fundamental metrics.

Biological validation with a kombucha consortium was accomplished, as known local and global behavior, including key taxa and interspecific biological interactions empirically confirmed elsewhere [91, 94], were reproduced by the proposed toolbox. Moreover, our high-complexity case study, *Archean Domes*, displays the scope and usefulness of MicNet toolbox by deconstructing microbial co-occurrence networks to manageable biological knowledge. Aware of current caveats of the limitations microbial co-occurrence networks have [1, 26], we restate that this approach should be taken as a roadmap for further research on the microbiome system, rather than a conclusive analysis. For example, directed studies to Phycisphaerae bacteria (and other central taxonomic groups) can be performed, and consequently, assess the relevance of this taxon to the whole community structure and functioning. Similarly, "microbial dark matter" characterization and relevance could be further explored with the increasing technologies and databases [111]. Directed co-culture experiments and other novel strategies such as microdroplets are fundamental for biological interactions' validation [31], which could be applied to inferred correlations between taxa of interest. Moreover, module aimed experiments, including synthetic microbial communities, are tractable strategies that, although reducing complexity of the system, could be informative about mid-scale structures crucial to the system's stability [112], especially if the experiments include perturbations [113].

Given its potential usefulness, understanding both global and local patterns in microbial communities may be a wise strategy to delve deeper into their currently unknown properties. With the introduction of the MicNet toolbox, we hope that the research community will be able to implement several existing and new analysis techniques in a straightforward manner to further keep unravelling the intricate conundrums that microbiomes hold.

Supporting information

S1 Fig. SparCC parameter selection. SparCC parameters were set based on our most complex network: *Archean Domes*. We ran SparCC varying **A**. the number of iterations from 10 to 100, **B**. the exclusion number from 10 to 100 and **C**. the exclusion threshold from 0.1 to 0.9. We chose the final values based on the stabilization of the number of edges found, such that the final values used for our databases were: 50 iterations, 10 exclusion number and 0.1 for exclusion threshold.
(PNG)

S2 Fig. Bar plot for community analysis composition (Louvain groups) from the kombucha Network. Community ID is shown in the x-axis.
(PNG)

S3 Fig. Bar plot for community analysis composition (HDBSCAN groups) from the kombucha Network. Cluster ID is shown in the x-axis.
(PNG)

S4 Fig. *Archean Domes* topology comparison. A. Distributions obtained from simulated random (red), scale-free (blue) and small-world (green) networks and its comparison to the metrics found in the *Archean Domes* network for: degree variance, modularity, average path length, small-world index, and clustering coefficient. **B**. Degree distribution of the *Archean Domes* network. We also show the comparison of the kombucha CCDF with a random network CCDF, a small-world network, and a scale-free network.
(PNG)

S5 Fig. Bar plot for community analysis composition (Louvain groups) from the *Archean Domes* network. Community ID is shown in the x-axis. Unassigned bacteria and not annotated sequences are grouped in the NA category.
(PNG)

S6 Fig. Bar plot for community analysis composition (HDBSCAN groups) from the *Archean Domes* network. Cluster ID is shown in the x-axis. Unassigned bacteria and not annotated sequences are grouped in the NA category.
(PNG)

S1 Table. Network metrics for three canonical topologies. Large scale metrics of three simulated networks with a random, scale-free and small world topology.
(PNG)

S2 Table. Main metrics and network properties of the kombucha co-occurrence network.
(PNG)

S3 Table. Centrality measures from potential key players in kombucha network. Degree, closeness, betweenness, and PageRank centrality was calculated for the top 5 ASV respectively. If a given ASV was not among the top 5 in a centrality measure, the value is reported as NA.
(PNG)

S4 Table. Community analysis (Louvain groups) metrics on kombucha network. For each community, total nodes, diameter, clustering coefficient, and average shortest path were calculated. Clusters are ordered by increasing density.
(PNG)

S5 Table. Community analysis (HDBSCAN groups) metrics on kombucha network. For each community, total nodes, diameter, clustering coefficient, and average shortest path were calculated. Clusters are ordered by increasing density.
(PNG)

S6 Table. Network robustness analysis for the kombucha network. Random, by groups (Genus), and by degree centrality percolation simulations were performed on the Louvain groups. For the percolation by groups, only *Komagataeibacter* and is shown.
(PNG)

S7 Table. Main metrics and network properties of the *Archean Domes* co-occurrence network.
(PNG)

S8 Table. Centrality measures from potential key players in *Archean Domes* network. Degree, closeness, betweenness, and PageRank centrality was calculated for the top 10 ASV respectively. If a given ASV was not among the top 10 in a centrality measure, the value is reported as NA.
(PNG)

S9 Table. Community analysis (Louvain groups) metrics on *Archean Domes* network. For each community, total edges, total nodes, average degree, clustering coefficient, and density were calculated. Clusters are ordered by increasing density.
(PNG)

S10 Table. Community analysis (HDBSCAN groups) metrics on *Archean Domes* network. For each community, total edges, total nodes, average degree, clustering coefficient, and density were calculated. Clusters are ordered by increasing density.
(PNG)

S11 Table. Network robustness analysis from the *Archean Domes* network. Random, by groups (Phylum), and by degree centrality percolation simulations were performed on the Louvain groups. For the percolation by groups, only Cyanobacteria percolation is shown.
(PNG)

S1 File.
(ZIP)

Acknowledgments

We thank Laboratorio de Inteligencia Artificial, Ixulabs, for funding acquisition, Diego Nava for his contributions in the conceptualization, and Julian Trejo and Diana Fernandez Rosales for their contribution in the elaboration of the artwork. We also thank Rosalinda Tapia and Erika Aguirre for their technical support.

Author Contributions

Conceptualization: Natalia Favila, David Madrigal-Trejo, Daniel Legorreta, Jazmín Sánchez-Pérez, Luis E. Eguiarte.

Funding acquisition: Jazmín Sánchez-Pérez, Valeria Souza.

Investigation: Natalia Favila, David Madrigal-Trejo, Jazmín Sánchez-Pérez, Laura Espinosa-Asuar.

Methodology: Natalia Favila, David Madrigal-Trejo, Daniel Legorreta.

Project administration: Jazmín Sánchez-Pérez, Laura Espinosa-Asuar, Valeria Souza.

Resources: Luis E. Eguiarte.

Software: Natalia Favila, Daniel Legorreta.

Supervision: David Madrigal-Trejo, Laura Espinosa-Asuar, Valeria Souza.

Validation: Natalia Favila, David Madrigal-Trejo.

Visualization: Natalia Favila, David Madrigal-Trejo, Daniel Legorreta.

Writing – original draft: Natalia Favila, David Madrigal-Trejo.

Writing – review & editing: Natalia Favila, David Madrigal-Trejo, Jazmín Sánchez-Pérez, Laura Espinosa-Asuar, Luis E. Eguiarte, Valeria Souza.

References

1. Röttjers L, Faust K. From hairballs to hypotheses–biological insights from microbial networks. FEMS Microbiol. Rev. 2018; 42:761–80. https://doi.org/10.1093/femsre/fuy030 PMID: 30085090
2. Dohlman AB, Shen X. Mapping the microbial interactome: Statistical and experimental approaches for microbiome network inference. Exp. Biol. Med. 2019; 244:445. https://doi.org/10.1177/1535370219836771 PMID: 30880449
3. Hernandez DJ, David AS, Menges ES, Searcy CA, Afkhami ME. Environmental stress destabilizes microbial networks. ISME J. 2021 156 2021; 15:1722–34. https://doi.org/10.1038/s41396-020-00882-x PMID: 33452480
4. Karimi B, Maron PA, Chemidlin-Prevost Boure N, Bernard N, Gilbert D, Ranjard L. Microbial diversity and ecological networks as indicators of environmental quality. Environ. Chem. Lett. 2017 152 2017; 15:265–81.
5. Poudel R, Jumpponen A, Schlatter DC, Paulitz TC, Gardener B., Kinkel LL, et al. Microbiome Networks: A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management. Phytopathology 2016; 106:1083–96. https://doi.org/10.1094/PHYTO-02-16-0058-FI PMID: 27482625
6. Xiaofei L, Kankan Z, Ran X, Yuanhui L, Jianming X, Bin M. Strengthening Insights in Microbial Ecological Networks from Theory to Applications. mSystems 2010; 4:e00124–19.
7. Pearson K. Determination of the coefficient of correlation. Science (80-.). 1909; 30:23–5. https://doi.org/10.1126/science.30.757.23 PMID: 17838275
8. Spearman C. The Proof and Measurement of Association between Two Things. Am. J. Psychol. 1904; 15:72.
9. Jaccard P. The distribution of the flora in the alpine zone. New Phytol. 1912; 11:37–50.
10. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol. Monogr. 1957; 27:325–49.
11. Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 2006; 22:2532–8. https://doi.org/10.1093/bioinformatics/btl417 PMID: 16882654
12. Xia LC, Ai D, Cram J, Fuhrman JA, Sun F. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. Bioinformatics 2013; 29:230–7. https://doi.org/10.1093/bioinformatics/bts668 PMID: 23178636
13. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al. Detecting Novel Associations in Large Data Sets. Science (80-.). 2011; 334:1518–24. https://doi.org/10.1126/science.1205438 PMID: 22174245
14. Jizhong Z, Ye D, Feng L, Zhili H, Yunfeng Y, David R. Phylogenetic Molecular Ecological Network of Soil Microbial Communities in Response to Elevated CO2. MBio 2021; 2:e00122–11.

15. Deng Y, Jiang Y-H, Yang Y, He Z, Luo F, Zhou J. Molecular ecological network analyses. BMC Bioinforma. 2012 131 2012; 13:1–20. https://doi.org/10.1186/1471-2105-13-113 PMID: 22646978

16. Aitchison J. The statistical analysis of compositional data. London; New York: Chapman and Hall; 1986.

17. Friedman J, Alm EJ. Inferring Correlation Networks from Genomic Survey Data. PLOS Comput. Biol. 2012; 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687 PMID: 23028285

18. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial Co-occurrence Relationships in the Human Microbiome. PLOS Comput. Biol. 2012; 8:e1002606. https://doi.org/10.1371/journal.pcbi.1002606 PMID: 22807668

19. Faust K, Raes J. CoNet app: inference of biological association networks using Cytoscape. F1000Research 2016; 5:1519. https://doi.org/10.12688/f1000research.9050.2 PMID: 27853510

20. Succurro A, Ebenhöh O. Review and perspective on mathematical modeling of microbial ecosystems. Biochem. Soc. Trans. 2018; 46:403. https://doi.org/10.1042/BST20170265 PMID: 29540507

21. Volterra V. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. 1926.

22. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J. 2016 107 2016; 10:1669–81. https://doi.org/10.1038/ismej.2015.235 PMID: 26905627

23. Shaw GT-W, Pao Y-Y, Wang D. MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles. BMC Bioinforma. 2016 171 2016; 17:1–12. https://doi.org/10.1186/s12859-016-1359-0 PMID: 27887570

24. Fisher CK, Mehta P. Identifying Keystone Species in the Human Gut Microbiome from Metagenomic Timeseries Using Sparse Linear Regression. PLoS One 2014; 9:e102451. https://doi.org/10.1371/journal.pone.0102451 PMID: 25054627

25. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front. Microbiol. 2014; 0:219. https://doi.org/10.3389/fmicb.2014.00219 PMID: 24904535

26. Hirano H, Takemoto K. Difficulty in inferring microbial community structure based on co-occurrence network approaches. BMC Bioinforma. 2019 201 2019; 20:1–14. https://doi.org/10.1186/s12859-019-2915-1 PMID: 31195956

27. Raman K. Structure of Networks. In: An Introduction to Computational Systems Biology: Systems-Level Modelling of Cellular Networks. Chapman and Hall/CRC.; 2021. page 57–90.

28. Landi P, Minoarivelo HO, Brännström Å, Hui C, Dieckmann U. Complexity and stability of ecological networks: a review of the theory. Popul. Ecol. 2018 604 2018; 60:319–45.

29. Bouchez T, Blieux AL, Dequiedt S, Domaizon I, Dufresne A, Ferreira S, et al. Molecular microbiology methods for environmental diagnosis. Environ. Chem. Lett. 2016 144 2016; 14:423–41.

30. Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, et al. Biodiversity and ecosystem functioning: Current knowledge and future challenges. Science (80-.). 2001; 294:804–8. https://doi.org/10.1126/science.1064088 PMID: 11679658

31. Faust K, Raes J. Microbial interactions: from networks to models. Nat. Rev. Microbiol. 2012 108 2012; 10:538–50. https://doi.org/10.1038/nrmicro2832 PMID: 22796884

32. Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, et al. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. Sci. Reports 2018 81 2018; 8:1–12. https://doi.org/10.1038/s41598-018-23931-0 PMID: 29651160

33. Elmqvist T, Folke C, Nyström M, Peterson G, Bengtsson J, Walker B, et al. Response diversity, ecosystem change, and resilience. Front. Ecol. Environ. 2003; 1:488–94.

34. Tylianakis JM, Laliberté E, Nielsen A, Bascompte J. Conservation of species interaction networks. Biol. Conserv. 2010; 143:2270–9.

35. Zappelini C, Karimi B, Foulon J, Lacercat-Didier L, Maillard F, Valot B, et al. Diversity and complexity of microbial communities from a chlor-alkali tailings dump. Soil Biol. Biochem. 2015; 90:101–10.

36. De Anda V, Zapata-Peñasco I, Blaz J, Poot-Hernández AC, Contreras-Moreira B, González-Laffitte M, et al. Understanding the Mechanisms Behind the Response to Environmental Perturbation in Microbial Mats: A Metagenomic-Network Based Approach. Front. Microbiol. 2018; 0:2606. https://doi.org/10.3389/fmicb.2018.02606 PMID: 30555424

37. Zhou H, Gao Y, Jia X, Wang M, Ding J, Cheng L, et al. Network analysis reveals the strengthening of microbial interaction in biological soil crust development in the Mu Us Sandy Land, northwestern China. Soil Biol. Biochem. 2020; 144:107782.

38. Corel E, Lopez P, Méheust R, Bapteste E. Network-Thinking: Graphs to Analyze Microbial Complexity and Evolution. Trends Microbiol. 2016; 24:224–37. https://doi.org/10.1016/j.tim.2015.12.003 PMID: 26774999

39. Connor N, Barberán A, Clauset A. Using null models to infer microbial co-occurrence networks. PLoS One 2017; 12:e0176751. https://doi.org/10.1371/journal.pone.0176751 PMID: 28493918

40. Barroso-Bergadà D, Pauvert C, Vallance J, Delière L, Bohan DA, Buée M, et al. Microbial networks inferred from environmental DNA data for biomonitoring ecosystem change: Strengths and pitfalls. Mol. Ecol. Resour. 2021; 21:762–80. https://doi.org/10.1111/1755-0998.13302 PMID: 33245839

41. de Vries FT, Griffiths RI, Bailey M, Craig H, Girlanda M, Gweon HS, et al. Soil bacterial networks are less stable under drought than fungal networks. Nat. Commun. 2018 91 2018; 9:1–12.

42. Dong Y, Gao J, Wu Q, Ai Y, Huang Y, Wei W, et al. Co-occurrence pattern and function prediction of bacterial community in Karst cave. BMC Microbiol. 2020 201 2020; 20:1–13.

43. Hannigan GD, Duhaime MB, Koutra D, Schloss PD. Biogeography and environmental conditions shape bacteriophage-bacteria networks across the human microbiome. PLoS Comput. Biol. 2018;14. https://doi.org/10.1371/journal.pcbi.1006099 PMID: 29668682

44. Liu Z, Wei H, Zhang J, Saleem M, He Y, Zhong J, et al. Higher Sensitivity of Soil Microbial Network Than Community Structure under Acid Rain. Microorg. 2021, Vol. 9, Page 118 2021; 9:118. https://doi.org/10.3390/microorganisms9010118 PMID: 33419116

45. Schmoldt A, Benthe HF, Haberland G. Digitoxin metabolism by rat liver microsomes. Biochem. Pharmacol. 1975; 24:1639–41. PMID: 10

46. Layeghifard M, Hwang DM, Guttman DS. Disentangling Interactions in the Microbiome: A Network Perspective. Trends Microbiol. 2017; 25:217. https://doi.org/10.1016/j.tim.2016.11.008 PMID: 27916383

47. Estrada-Peña A, Cabezas-Cruz A, Pollet T, Vayssier-Taussat M, Cosson J-F. High Throughput Sequencing and Network Analysis Disentangle the Microbial Communities of Ticks and Hosts Within and Between Ecosystems. Front. Cell. Infect. Microbiol. 2018; 8:236. https://doi.org/10.3389/fcimb.2018.00236 PMID: 30038903

48. Herren CM, McMahon KD. Cohesion: a method for quantifying the connectivity of microbial communities. ISME J. 2017 1111 2017; 11:2426–38. https://doi.org/10.1038/ismej.2017.91 PMID: 28731477

49. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks, competition, and stability. Science (80-.). 2015; 350:663–6.

50. Suweis S, Grilli J, Maritan A. Disentangling the effect of hybrid interactions and of the constant effort hypothesis on ecological community stability. Oikos 2014; 123:525–32.

51. Mougi A, Kondoh M. Diversity of interaction types and ecological community stability. Science (80-.). 2012; 337:349–51. https://doi.org/10.1126/science.1220529 PMID: 22822151

52. Zhou J, Deng Y, Luo F, He Z, Tu Q, Zhi X. Functional molecular ecological networks. MBio 2010;1. https://doi.org/10.1128/mBio.00169-10 PMID: 20941329

53. Lupatini M, Suleiman AKA, Jacques RJS, Antoniolli ZI, de Siqueira Ferreira A, Kuramae EE, et al. Network topology reveals high connectance levels and few key microbial genera within soils. Front. Environ. Sci. 2014; 0:10.

54. Barberán A, Bates ST, Casamayor EO, Fierer N. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 2012 62 2011; 6:343–51. https://doi.org/10.1038/ismej.2011.119 PMID: 21900968

55. Humphries MD, Gurney K. Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. PLoS One 2008; 3:e0002051. https://doi.org/10.1371/journal.pone.0002051 PMID: 18446219

56. Dubin K, Callahan MK, Ren B, Khanin R, Viale A, Ling L, et al. Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. Nat. Commun. 2016 71 2016; 7:1–8. https://doi.org/10.1038/ncomms10391 PMID: 26837003

57. McHardy IH, Goudarzi M, Tong M, Ruegger PM, Schwager E, Weger JR, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. Microbiome 2013 11 2013; 1:1–19.

58. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. Nat. 2016 5327600 2016; 532:465–70. https://doi.org/10.1038/nature16942 PMID: 26863193

59. Cram JA, Xia LC, Needham DM, Sachdeva R, Sun F, Fuhrman JA. Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes. ISME J. 2015 912 2015; 9:2573–86. https://doi.org/10.1038/ismej.2015.76 PMID: 25989373

60. Zhang B, Zhang J, Liu Y, Shi P, Wei G. Co-occurrence patterns of soybean rhizosphere microbiome at a continental scale. Soil Biol. Biochem. 2018; 118:178–86.

61. Saberi M, Khosrowabadi R, Khatibi A, Misic B, Jafari G. Topological impact of negative links on the stability of resting-state brain network. Sci. Reports 2021 111 2021; 11:1–14. https://doi.org/10.1038/s41598-021-81767-7 PMID: 33500525

62. Ma Z (Sam), Ye D. Trios—promising in silico biomarkers for differentiating the effect of disease on the human microbiome network. Sci. Reports 2017 71 2017; 7:1–9. https://doi.org/10.1038/s41598-017-12959-3 PMID: 29038470

63. Srinivasan A. Local balancing influences global structure in social networks. Proc. Natl. Acad. Sci. 2011; 108:1751–2. https://doi.org/10.1073/pnas.1018901108 PMID: 21252302

64. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. 2018; 3:861.

65. Diaz-Papkovich A, Anderson-Trocmé L, Ben-Eghan C, Gravel S. UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. PLOS Genet. 2019; 15:e1008432. https://doi.org/10.1371/journal.pgen.1008432 PMID: 31675358

66. Diaz-Papkovich A, Anderson-Trocmé L, Gravel S. A review of UMAP in population genetics. J. Hum. Genet. 2020 661 2020; 66:85–91. https://doi.org/10.1038/s10038-020-00851-4 PMID: 33057159

67. McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. J. Open Source Softw. 2017; 2:205.

68. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.

69. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. SciPy; 2010. page 56–61.

70. The pandas development team. pandas-dev/pandas: Pandas 1.3.2. 2021.

71. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nat. 2020 5857825 2020; 585:357–62. https://doi.org/10.1038/s41586-020-2649-2 PMID: 32939066

72. Dask Development Team. Dask: Library for dynamic task scheduling. 2016.

73. Lam SK, Pitrou A, Seibert S. Numba: A llvm-based python jit compiler. In: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. 2015. page 1–6.

74. Collette A. Python and HDF5. O'Reilly; 2013.

75. Bokeh Development Team. Bokeh: Python library for interactive visualization. 2018.

76. Hagberg AA, Schult DA, Swart PJ. Exploring Network Structure, Dynamics, and Function using NetworkX. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of the 7th Python in Science Conference. Pasadena, CA USA: 2008. page 11–5.

77. Malzer C, Baum M. A Hybrid Approach To Hierarchical Density-based Cluster Selection. IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst. 2019;2020-September:223–8.

78. Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. 2020.

79. Campello RJGB, Moulavi D, Sander J. Density-Based Clustering Based on Hierarchical Density Estimates. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 2013;7819 LNAI:160–72.

80. Espinosa-Asuar L, Monroy C, Madrigal-Trejo D, Navarro M, Sánchez J, Muñoz J, et al. Ecological relevance of abundant and rare taxa in a high-diverse elastic hypersaline microbial mat, using a small-scale sampling. bioRxiv 2021;2021.03.04.433984.

81. Erdös P, Rényi A. On Random Graphs I. Publ. Math. Debrecen 1959; 6:290.

82. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nat. 1998 3936684 1998; 393:440–2. https://doi.org/10.1038/30918 PMID: 9623998

83. Barabási A-L, Albert R. Emergence of Scaling in Random Networks. Science (80-.). 1999; 286:509–12. https://doi.org/10.1126/science.286.5439.509 PMID: 10521342

84. Alizadeh M, Cioffi-Revilla C, Crooks A. Generating and analyzing spatial social networks. Comput. Math. Organ. Theory 2016 233 2016; 23:362–90.

85. Downey A. Think complexity: complexity science and computational modeling. Sebastopol, Calif.: O'Reilly; 2018.

86. Aynaud T. python-louvain x.y: Louvain algorithm for community detection. 2020.

87. Newman MEJ. Networks: an introduction. Oxford [u.a.]: Oxford Univ. Press; 2010.

88. Barabási A-L, Pósfai M. Network science. 2016. https://doi.org/10.1017/nws.2016.2 PMID: 27867518

89. Aref S, Wilson MC. Balance and frustration in signed networks. J. Complex Networks 2019; 7:163–89.

90. Estrada E. Rethinking structural balance in signed social networks. Discret. Appl. Math. 2019; 268:70–90.

91. Arıkan M, Mitchell AL, Finn RD, Gürel F. Microbial composition of Kombucha determined using amplicon sequencing and shotgun metagenomics. J. Food Sci. 2020; 85:455–64. https://doi.org/10.1111/1750-3841.14992 PMID: 31957879

92. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 2019 378 2019; 37:852–7. https://doi.org/10.1038/s41587-019-0209-9 PMID: 31341288

93. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample inference from Illumina amplicon data. Nat. Methods 2016; 13:581. https://doi.org/10.1038/nmeth.3869 PMID: 27214047

94. May A, Narayanan S, Alcock J, Varsani A, Maley C, Aktipis A. Kombucha: a novel model system for cooperation and conflict in a complex multi-species microbial ecosystem. PeerJ 2019;7.

95. BE W, RJ D. Fermented foods as experimentally tractable microbial ecosystems. Cell 2015; 161:49–55. https://doi.org/10.1016/j.cell.2015.02.034 PMID: 25815984

96. Tran T, Grandvalet C, Verdier F, Martin A, Alexandre H, Tourdot-Maréchal R. Microbial Dynamics between Yeasts and Acetic Acid Bacteria in Kombucha: Impacts on the Chemical Composition of the Beverage. Foods 2020;9. https://doi.org/10.3390/foods9070963 PMID: 32708248

97. Jayabalan R, Malbaša R V., Lončar ES, Vitas JS, Sathishkumar M. A Review on Kombucha Tea—Microbiology, Composition, Fermentation, Beneficial Effects, Toxicity, and Tea Fungus. Compr. Rev. Food Sci. Food Saf. 2014; 13:538–50. https://doi.org/10.1111/1541-4337.12073 PMID: 33412713

98. Dang TDT, Vermeulen A, Ragaert P, Devlieghere F. A peculiar stimulatory effect of acetic and lactic acid on growth and fermentative metabolism of Zygosaccharomyces bailii. Food Microbiol. 2009; 26:320–7. https://doi.org/10.1016/j.fm.2008.12.002 PMID: 19269576

99. Yamada Y, Yukphan P, Vu HTL, Muramatsu Y, Ochaikul D, Tanasupawat S, et al. Description of Komagataeibacter gen. nov., with proposals of new combinations (Acetobacteraceae). J. Gen. Appl. Microbiol. 2012; 58:397–404. https://doi.org/10.2323/jgam.58.397 PMID: 23149685

100. Medina-Chávez NO, De la Torre-Zavala S, Arreola-Triana AE, Souza V. Cuatro Ciénegas as an Archaean Astrobiology Park. Springer, Cham; 2020. page 219–28.

101. Prieto-Barajas CM, Valencia-Cantero E, Santoyo G. Microbial mat ecosystems: Structure types, functional diversity, and biotechnological application. Electron. J. Biotechnol. 2018; 31:48–56.

102. Aguirre-von-Wobeser E, Soberón-Chávez G, Eguiarte LE, Ponce-Soto GY, Vázquez-Rosas-Landa M, Souza V. Two-role model of an interaction network of free-living γ-proteobacteria from an oligotrophic environment. Environ. Microbiol. 2014; 16:1366–77. https://doi.org/10.1111/1462-2920.12305 PMID: 24128119

103. Spring S, Bunk B, Spröer C, Rohde M, Klenk H-P. Genome biology of a novel lineage of planctomycetes widespread in anoxic aquatic environments. Environ. Microbiol. 2018; 20:2438–55. https://doi.org/10.1111/1462-2920.14253 PMID: 29697183

104. Thomas F, Morris JT, Wigand C, Sievert SM. Short-term effect of simulated salt marsh restoration by sand-amendment on sediment bacterial communities. PLoS One 2019; 14:e0215767. https://doi.org/10.1371/journal.pone.0215767 PMID: 31034478

105. Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. A network approach to elucidate and prioritize microbial dark matter in microbial communities. ISME J. 2020 151 2020; 15:228–44. https://doi.org/10.1038/s41396-020-00777-x PMID: 32963345

106. Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, Heidelberg JF, et al. Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J. 2011 58 2011; 5:1262–78. https://doi.org/10.1038/ismej.2011.73 PMID: 21697961

107. Bairey E, Kelsic ED, Kishony R. High-order species interactions shape ecosystem diversity. Nat. Commun. 2016 71 2016; 7:1–7. https://doi.org/10.1038/ncomms12285 PMID: 27481625

108. Chiang YS, Chen YW, Chuang WC, Wu CI, Wu C Te. Triadic balance in the brain: Seeking brain evidence for Heider's structural balance theory. Soc. Networks 2020; 63:80–90.

109. Zhang Q, Acuña JJ, Inostroza NG, Duran P, Mora ML, Sadowsky MJ, et al. Niche Differentiation in the Composition, Predicted Function, and Co-occurrence Networks in Bacterial Communities Associated With Antarctic Vascular Plants. Front. Microbiol. 2020; 0:1036.

110. Venturelli OS, Carr A V, Fisher G, Hsu RH, Lau R, Bowen BP, et al. Deciphering microbial interactions in synthetic human gut microbiome communities. Mol. Syst. Biol. 2018; 14:e8157. https://doi.org/10.15252/msb.20178157 PMID: 29930200

111. Jiao J-Y, Liu L, Hua Z-S, Fang B-Z, Zhou E-M, Salam N, et al. Microbial dark matter coming to light: challenges and opportunities. Natl. Sci. Rev. 2021; 8:2021. https://doi.org/10.1093/nsr/nwaa280 PMID: 34691599

112. Bolhuis H, Cretoiu MS, Stal LJ. Molecular ecology of microbial mats. FEMS Microbiol. Ecol. 2014; 90:335–50. https://doi.org/10.1111/1574-6941.12408 PMID: 25109247

113. Tropini C, Moss EL, Merrill BD, Ng KM, Higginbottom SK, Casavant EP, et al. Transient Osmotic Perturbation Causes Long-Term Alteration to the Gut Microbiota. Cell 2018; 173:1742-1754.e17. https://doi.org/10.1016/j.cell.2018.05.008 PMID: 29906449