

# Network analysis methods for studying microbial communities: A mini review [1]

## 1 Background

Microbial communities consist of micro-organisms such as bacteria, protozoa, archaea, fungi and virus. Micro-organisms have built robust ecosystems in various environments such as soil, sea water and various human organs. For example, the human body serves as a home for most of these diverse micro-organisms.

However, these organisms have major influence on the health of their host. Therefore, understanding the microbial interactions within and between kingdoms will provide a great insight on how to combat complex human diseases such as rectal inflammatory. Microbial communities are dynamic over time in the sense that abundances of members could fluctuate from one point in time to the next. [2]

Extensive research has been done on interactions between host and its microbiome. These studies have revealed several cases of complex dynamics. Trans-kingdom interactions between these micro-organisms and their effects on the host have only recently been studied. Network-based analytical approaches have helped in the study of systems with complex microbial interactions. They have also been vital to infer complex networks in systems biology.

The literature gives an overview of simple correlation and complex conditional dependence-based methods used to infer intra-kingdom interactions as well as current limitations that encourage further method developments.

## 2 Previous Work

Several methods have been developed to research microbial interactions. Microbial interactions go beyond within-kingdom interactions. They also entail trans-kingdom analysis. However, only a few available methods tackle trans-kingdom interactions. The tools available, range from correlation-based methods and conditional dependence or graphical models to network-based methods for trans-kingdom analysis.

Pearson and Spearman correlations are some of the important correlation-based tools used in studying microbial interactions in the human organs and soil micro-biomes. The downside is that, they are unable to account for compositionality hence resulting in misleading associations between low-abundance organisms. Therefore, some methods have been adopted to curb this compositionality issue including the classical Centered Log Ratio (CLR) transformation. With the CLR transformation, the reference is taken as the geometric mean of the sample vector. Correlation methods also fail to distinguish between direct and indirect associations. Examples of some existing tools which implement the correlation-based method are SparCC, CCLasso, REBACCA,

Correlation-Centric Network Command line tool and MENAP online tool.

Conditional dependence on the other hand, has higher computational complexity and runtime than correlation-based methods. Due to this advantage, partial correlation methods have been made to model conditional dependence so that they can differentiate between direct and indirect associations. These are some R packages which have been written to implement conditional dependence : gCoda, MDiNE, MixMPLN, NetComi, HARMONIES, SPIEC-EASI and COZINE.

With regards to network-based methods for trans-kingdom analysis, methods using several data modalities have been adopted to study relationships between taxa from different kingdoms. Attaining multi-modal data is very challenging in the microbiome community. Thus, only a few methods have been developed over time for this purpose. The SPIEC-EASI method has been adopted for trans-kingdom analysis. This method concatenates two or more data sets for cross-kingdom networks analysis. The Multi-Omics Factor Analysis (MOFA) also uses group analysis to provide more insight about the variation across different samples. Conversely, the Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) method is a tool which only has support for continuous data and assumes a linear relationship between selected variables. While this function may not be beneficial in complex scenarios, it is able to reduce the number of predictors in simple systems by employing Partial Least Squares (PLS) regression technique.

### 3 Novelty

The ideas presented in the literature are both theoretical and empirical. There are many proposed methods by the literature for studying network-based models for microbial interactions. For simplicity of usage, there are also many packages in various programming languages like R and Python that replicate the functionality of these methods. However, each of them shows some form of drawbacks. The literature presents the fact that network analysis should not be limited to bacterial co-occurrence patterns but also viral or fungal interactions. As connections exist between different environments, different microbial interactions occur across various microbial niches. It also emphasis on the three major challenges with regards to microbiome network analysis and methods which have been adopted to solve them.

### 4 Results

Comprehensive benchmark datasets or accepted simulated datasets will serve as the needed datasets to systematically test the performance of existing methods. Currently, there exist a challenge of inadequate standard dataset. Hence, an optimal method of network analysis is chosen based on trade-offs between complex models and simple methods. The figure below shows the various network methods and their corresponding challenges and advantages. It can be observed that, a lot of factors might influence the method to be chosen to perform network analysis. It can be inferred that the complex methods are at the bottom of the tree while the simple methods are at the top. For Correlation based methods, REBACCA and CCLASSO handle compositionality

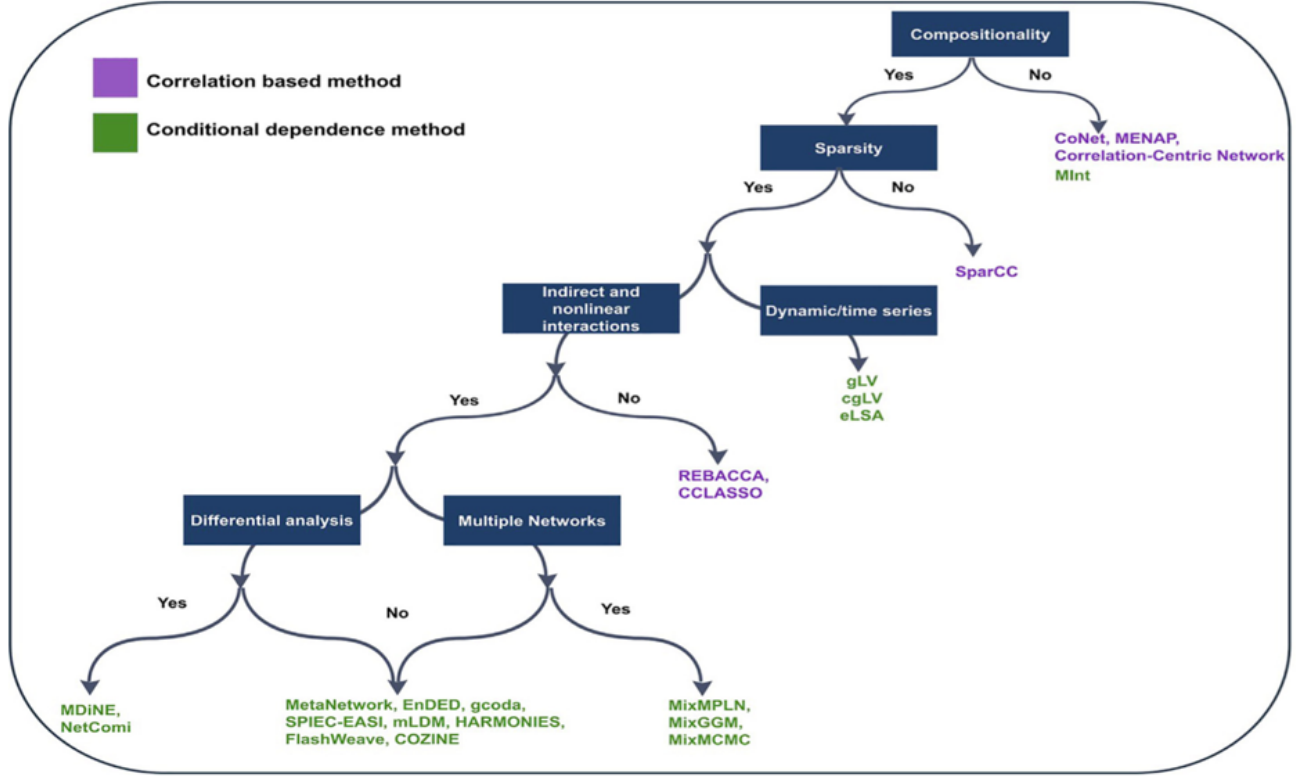


Figure 1: Workflow indicating the suitable network approaches [1]

bias and obtain relatively accurate results even though they cannot detect nonlinear interactions. Conversely, most of the conditional dependence methods are able to detect indirect and nonlinear interactions but have relatively high complexity.

## 5 Future work

It is a fact that existing methods are unable to overcome all of the challenges with respect to microbial interaction networks. Some of the issues include compositionality bias, over dispersion and poor sample to feature ratio. Another major issue that is yet to be tackled involves trans-kingdom interactions. More time-series analysis to study dynamic changes in microbial compositions should be made to show relevant patterns which leads to dynamism in community behavior. Further research should be conducted to investigate temporal changes in microbiome studies, cross correlation and time-varying network inference. Relevant analysis such as bistability and time-decay analysis on microbial relationships should also be done. Moreover, further studies should be done to validate existing methods by using universal benchmark or accepted simulated datasets. The literature would use both quantitative and qualitative methods with further testing and experimentation to address most of the challenges.

## References

- [1] Monica Steffi Machado, Michael Lauber, Sandra Reitmeier, Tim Kacprowski, Jan Baumbach, Dirk Haller, and Markus List. Network analysis methods for studying microbial communities: A mini review. *Computational and structural biotechnology journal*, 19:2687–2698, 2021.
- [2] Didier Gonze, Katharine Z Coyte, Leo Lahti, and Karoline Faust. Microbial communities as dynamical systems. *Current Opinion in Microbiology*, 44:41–49, 2018. Microbiota.