

BITS :: Call for Abstracts 2024 - Oral communication

| | |
|--------------------|--|
| <i>Type</i> | Oral communication |
| <i>Session</i> | Young BITS-RSG Symposium |
| <i>Title</i> | Limitations of the co-occurrence-based approach in the inference of microbiota interaction networks. |
| <i>All Authors</i> | Piero Mariotto(1), Matteo Baldan(1), Barbara Di Camillo(1),(2),(3), Giacomo Baruzzo(1) |

Affiliation

- (1) Department of Information Engineering, University of Padova, Padova (PD), Italy.
(2) Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro (PD), Italy.
(3) Padua Center of Network Medicine, University of Padova, Padova (PD), Italy.

Motivation

Microbial communities, called microbiota, inhabit all ecosystems, from the soil to the human body. These communities exert profound influence on the environment they inhabit. For example, in the human body community composition and structure have been correlated with multiple diseases like obesity, diabetes, IBD, cancer and HIV [1].

Communities' establishment and evolution are believed to be determined primarily by two phenomena: environment filtering (environment characteristics determine which species are able to inhabit it) and members cross-feeding (exchange of metabolites between members). These interactions can be represented as a network. The knowledge of these interaction networks is essential to understand community functions and to enable targeted intervention on the microbiota. For these reasons, much effort has been devoted, in recent years, to inferring these networks from experimental data.

To this end several methods have been proposed in the literature to infer the community interaction networks from the abundance matrices obtained from high-throughput sequencing techniques, such as 16S rDNA-seq experiments.

Unfortunately, the interaction networks of real communities are unknown, due to the complexity and large size of the communities, and the challenges of culture-based experiments. The lack of ground truth data makes it impossible to assess the accuracy of methods' prediction, thus leading to a scarce application of these methods in metagenomics' pipelines.

This work aims to benchmark in a uniform and consistent setting the performance of the major inference methods found in the literature using novel in-silico ground truth obtained from a recently developed simulator [2], thus providing a first comprehensive overview of current methods' accuracy.

Methods

We benchmarked 15 methods: 6 based on correlation (Pearson and Spearman correlation, proportionality, SparCC, CCLasso, SECOM-linear and SECOM-non-linear) and 7 based on partial-correlation (partial Pearson and Spearman correlation, SPIEC-EASI-MB, SPIEC-EASI-glasso, gCoda, SPRING, HARMONIES and COZINE).

To simulate the ground truth network and the corresponding 16S count table we used N2SIMBA simulator [2]. N2SIMBA allows the user to set a network topology, then it uses the network to model bacteria interactions and simulates the evolution of the bacteria community through a Microbial Consumer Resource Model [3], to finally simulate the sequencing count data through a Multivariate-Hypergeometric model.

We tested multiple simulated scenarios varying 2 network topologies (random or scale-free), 4 sample sizes ($n=50, 75, 100, 200$), 5 network densities ($=1, 2, 4, 6, 12$), 8 competition interaction percentages (from 0 to 75%) and 3 different types of metabolites intake rate. We then assessed the accuracy of the inferred network (measuring precision, recall and MCC) and the resemblance to the true network (degree distribution, diameter, radius, average distance, and average clustering coefficient). For each scenario we simulate a community of 50 species, and we run 50 iterations, for a total of 148500 tests.

Results

Fig. 1 shows part of the obtained results. In general, methods show low estimation power. Even in the best scenario, the inferred network has different topological structure compared to the true one (Fig. 1A), tending to be overly sparse and disconnected. Increasing the number of samples improves most method's performance up to 0.6 in terms of MCC (Fig. 1B). On the other hand, increasing network density or percentage of competition negatively affects method's accuracy, causing up to a 0.5 drop in MCC (Fig. 1B). These results highlight a narrow field of applications for the methods: only in a "optimal scenario" (high number of samples, low network density and competition percentage) few best-performing methods achieve ~ 0.8 MCC (CCLasso, Fig. 1C). Outside this scenario method's performance drops sharply, mainly due to the network sparsity assumption made by most methods.

Info

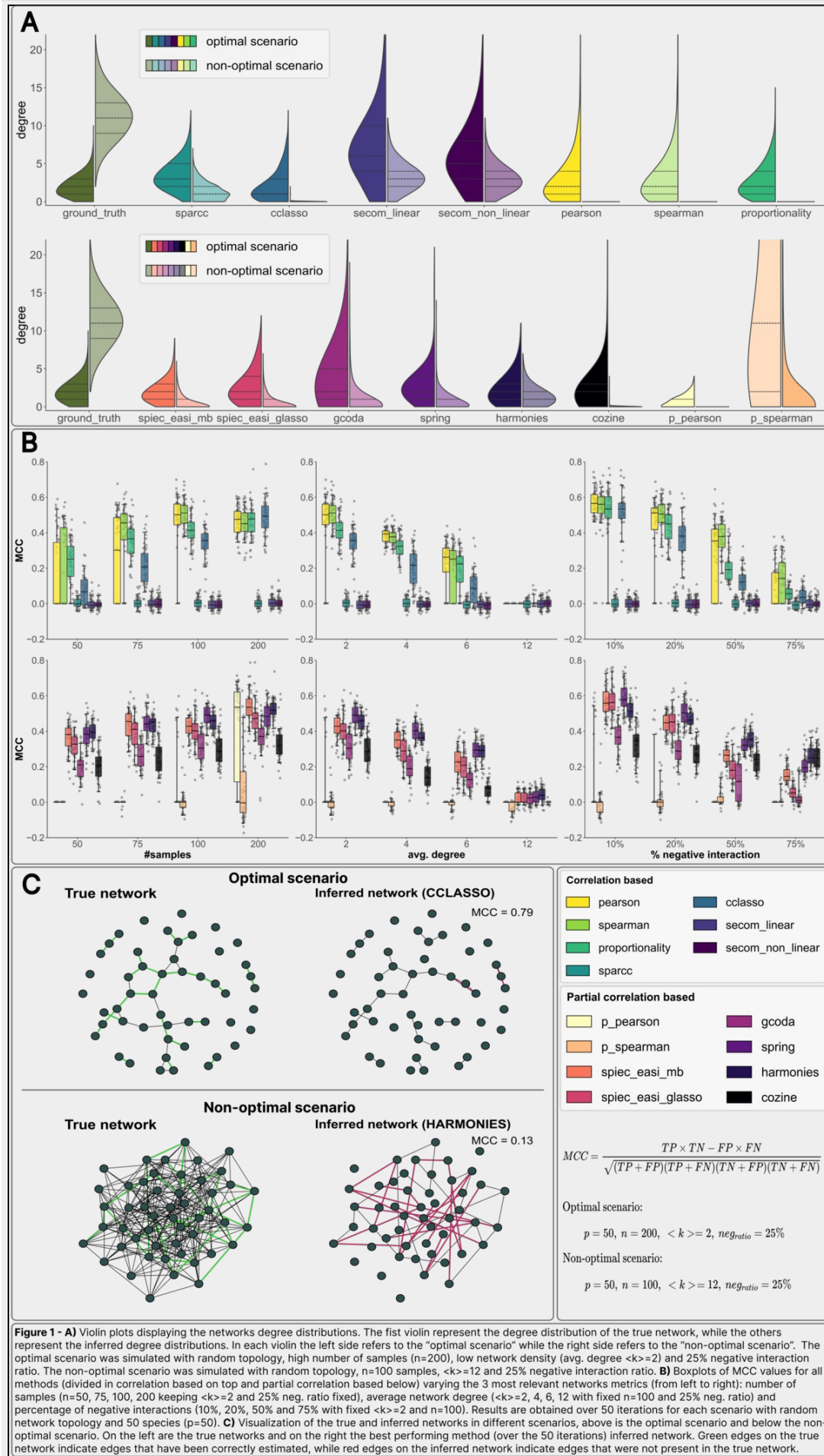
[1] Hou, K., Wu, ZX., Chen, XY. et al. Microbiota in health and diseases. *Sig Transduct Target Ther* 7, 135 (2022). <https://doi.org/10.1038/s41392-022-00974-4>

[2] Baldan et. al., N2BPC: an algorithmic approach from Networks to Bacteria's metabolite Production and Consumption. *Proceedings of the 18th Conference on Computational Intelligence Methods for Bioinformatics and Biostatistics (CIBB 2023)*.

[3] Marsland R III, Cui W, Goldford J, Sanchez A, Korolev K, et al. (2019) Available energy fluxes

filename Figure 1.png

Figure



Availability

Dissemination Material

Social

linkedin.com/in/sysbiobig
@sysbiobig

Summary

Microbial communities, known as microbiota, influence various ecosystems, including the human body, where they are linked to diseases like obesity and diabetes. Understanding their interaction network is crucial for targeted interventions. However, inferring these networks is challenging due to the complexity of communities. In this study we benchmarked 15 inference methods using simulated data, revealing low estimation power and sensitivity to network characteristics. Only under optimal conditions methods achieve good accuracy. #Microbiota #Benchmark #NetworkInference

Corresponding Author

| | |
|----------------------|--------------------------|
| <i>Name, Surname</i> | Giacomo, Baruzzo |
| <i>Email</i> | giacomo.baruzzo@unipd.it |
| <i>Submitted on</i> | 02.05.2024 |

Società Italiana di Bioinformatica
C.F. / P.IVA 97319460586
E-mail bits@bioinformatics.it

Sede legale Viale G. Mazzini, 114/B - 00195 Roma
Website bioinformatics.it

message generated by sciencedev.com for <https://bioinformatics.it> 19:00:21 02.05.2024
