

MICROBIAL NETWORK INFERENCES AND CLUSTERING: A NEW AGGREGATION METHOD ENABLES GUILDS EXPLORATION

CAMILLE CHAMPION¹ & RAPHAELLE MOMAL¹ & EMMANUELLE LE
CHATELIER¹ & MAHENDRA MARIADASSOU² & MAGALI BERLAND¹

¹ *Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas, France*
camille.champion@inrae.fr

² *Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France*
mahendra.mariadassou@inrae.fr

In recent years, research has been significantly directed towards the role of intestinal microbiota in health and disease. Composed of various bacterial species, there is some evidence that gut microbiota is bidirectionally linked with host immune system, well balanced in healthy individuals and disrupted in case of dysbiosis. Microbial communities are typically structured with local interactive bacterial organizations, called microbial guilds that may influence the host system.

We designed an approach based on specific inference and clustering methods dedicated to the modeling and analysis of microbial networks. The whole genome sequencing techniques yield species abundances, which by their specific properties (heterogeneity, extreme variability and high proportion of rare species), can be tricky to model, the network becoming sparse and poorly reproducible. The consequence of microbial data complexity is that we need to define adapted interaction measures. Conditional dependence relationships provide a clear separation between indirect and direct effects, and therefore yield sparse and easy to interpret networks when used as an interaction measure. Recently, many methods have been designed for the inference of co-occurrence networks and these methods rely on different mathematical concepts, making hard the selection of the most suitable inference technique to the microbial data modelling. Here, we selected seven inference techniques based on conditional dependency networks estimated using Gaussian Graphical Models (GGM): Magma [CBW19], SpiecEasi [KMM⁺15], gCoda [FHZD17], PLNetwork [CRM18], EMtree [MRA20], SPRING [YCG20] and ZiLN [PGB21]. In order to produce a stable consensus network, we aggregated the edges found to counter-balance each method's estimation defaults. The graph built was then clustered with the following three methods: CORE-clustering [CBB⁺21], ℓ_1 spectral clustering [CCB⁺] and Stochastic Block Model [AWF].

Running these methodologies on six similar microbial datasets and connecting guilds with associated clinical variables lead to the detection of reliable microbial guilds. A careful analysis of each guild allowed the identification of microbial species strongly involved in healthy individuals.

References

- [AWF] C.J. Anderson, S. Wasserman, and K. Faust. Building stochastic block models. *Social Networks*, 14.
- [CBB⁺21] C. Champion, A.C. Brunet, R. Burcelin, J.M. Loubes, and L. Risser. Detection of representative variables in complex systems with interpretable rules using core-clusters. *Algorithms*, 14(2):66 – 66, 2021.
- [CBW19] A. Cougoul, X. Bailly, and E.C. Wit. Magma: Inference of sparse microbial association networks. (538579), 2019.
- [CCB⁺] C. Champion, M. Champion, M. Blazere, R. Burcelin, and J.M. Loubes. L1-spectral clustering algorithm: a spectral clustering method using l1-regularization. (03095805).
- [CRM18] J. Chiquet, S. Robin, and M. Mariadassou. Variational inference for sparse network reconstruction from count data. In *International Conference on Machine Learning*. PMLR, 2018.
- [FHZD17] H. Fang, C. Huang, H. Zhao, and M. Deng. gcode: Conditional dependence network inference for compositional data. *Journal of Computational Biology*, 24(7):699–708, 2017.
- [KMM⁺15] Z.D. Kurtz, C.L. Müller, E.R. Miraldi, D.R. Littman, M.J. Blaser, and R.A. Bonneau. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computational Biology*, 11(5), 2015.
- [MRA20] R. Momal, S. Robin, and C. Ambroise. Tree-based inference of species interaction networks from abundance data. *Methods in Ecology and Evolution*, 11(5):621–632, 2020.
- [PGB21] V. Prost, S. Gazut, and T. Bröls. A zero inflated log-normal model for inference of sparse microbial association networks. *PLOS Computational Biology*, 17(6), 2021.
- [YCG20] G. Yoon, R.J. Carroll, and I. Gaynanova. Sparse semiparametric canonical correlation analysis for data of mixed types. *Biometrika*, 107(3):609–625, 2020.