AGGREGATING NETWORK INFERENCES: TOWARDS USEFUL NETWORKS

CAMILLE CHAMPION
Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas, France
RAPHAELLE MOMAL
Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas, France
EMMANUELLE LE CHATELIER
Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas, France.
MAHENDRA MARIADASSOU
Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France.
mahendra.mariadassou@inrae.fr
MAGALI BERLAND
Université Paris-Saclay, INRAE, MGP, 78350, Jouy-en-Josas, France.
magali.berland@inrae.fr

ABSTRACT

Modeling microbial systems as sparse and reproducible networks is a major challenge in network inference. Direct interactions between the microbial species of a given biome can help to understand how the microbial communities influence the system and through which mechanisms. Most state of the art methods reconstruct network from abundance data using Gaussian Graphical Models (GGM) for which statistically grounded and efficient inference approaches are available: Meinshausen-Bühlmann (MB) neighborhood selection, graphical Lasso (glasso), tree averaging approach, etc.

In this article, we examine eight inference techniques and propose a two-step consensus method to combine them. All methods methods rely on stability selection, a resampling-based procedure to choose a regularization parameter by evaluating the selection frequency of selecting edges in the constructed networks. However, stability selection can result in very different precision - recall trade-offs for different methods, even when they have identical precision - recall curves. Rather than selecting an optimal regularization parameter and disregarding the selection frequencies, as is performed in stability selection, we suggest to combine edge frequencies directly to reconstruct the network. This approach ensures that only robust and reproducible edges are included in the consensus network.

The effectiveness of the consensus method compared to individual methods is demonstrated on synthetic data derived from a set of healthy microbiota and then applied to a real study-case of microbiota from patients with liver cirrhosis. While the performances of all methods in terms of precision/recall are quite similar, the MB-based ones tend to outperform glasso-based ones and are in turn outperformed by the consensus method.

Keywords Network inference · glasso · neighborhood selection · microbial networks · ensemble methods

1 Introduction

The human gut microbiota is a complex ecosystem consisting of trillions of microorganisms, mainly bacteria, archeae and microbial eucaryotes that play critical roles in host physiology, including digestion, immune function and metabolism. virusRecent advances in sequencing technologies have enabled the characterization of gut microbiota composition and

function, providing unprecedented opportunities to understand the microbial communities that reside within the human gastrointestinal tract and decipher their role in health and disease. However, despite these technological advancements, understanding the functional interactions within the gut microbiota remains a major challenge, as is identifying the different microbial communities associated with a healthy microbiota. Microbial interactions within the gut are complex and dynamic, with individual microorganisms interacting with each other in a multitude of ways, including mutualism, parasitism, commensalism, and competition [1, 2]. Identifying these interactions and the microbial networks they give rise to is critical for understanding the ecology and dynamics of the gut microbiota.

To address this challenge, network-based approaches have been developed to infer microbial interactions and construct microbial interaction networks. These approaches use various computational techniques to model the relationships between microbial taxa, such as correlation analysis, co-occurrence analysis, and phylogenetic profiling. The resulting microbial networks can reveal potential interactions between microbial taxa, including positive and negative relationships and the identification of microbial guilds.

Microbial guilds are groups of microorganisms that co-occur and interact with each other, forming functional units within the gut microbiota, that may influence the host system. Identifying microbial guilds and the interactions between them is crucial for understanding the ecological dynamics of the gut microbiota and can provide insights into the role of the microbiota in health and disease. Indeed, perturbation of the microbiota with loss of health relevant function makes human body is more vulnerable to diseases and and intrusions of pathogens [3]. Those observations have opened up new prospects in medicine and alternate routes for therapeutic strategies to recover balance.

In this work, we present a novel network-based approach for inferring microbial guilds and their interactions within the human gut microbiota. Formally, microbial interaction networks consist of nodes, which correspond to bacterial species, and edges, which correspond to interactions between those species. Positive and negative interactions (resulting from parasitism, mutualism, or acommensalism) are rarely observed directly. They are instead often reconstructed from abundance data, using either longitudinal data (see the generalized Lotka-Volterra model in [4]) or co-occurrence data. We focus here on the latter suite of methods.

Network inference from co-occurrence data uses a variety of inference methods including (i) correlation-based approaches [5, 6, 2], (ii) approaches based on the inference of the latent correlation structure [7, 6] and (iii) approaches based on probabilistic graphical models [8, 9, 10]. Correlation-based methods model total dependencies and are therefore prone to confusion by environmental factors (*e.g.* shared habitat preferences or susceptibility to the same abiotic factors) and do not lend themselves to a clear separation between indirect and direct effects. By contrast, conditional dependency-based method prune out indirect correlations from direct interactions and lead to sparser and easier to interpret networks, at the cost of increased computational burden and more sophisticated models. The problem of network inference is compounded by the adverse characteristics of microbial abundance data (heterogeneity, heteroscedasticity, extreme variability and sparsity) which makes them tricky to model and leads to poorly reproducible or sparse networks, with many missing edges [11].

The complexity of reconstructing networks from co-occurrence data has spawned a rich literature with many additional methods being proposed in the last few years. The most recent methods include: `mixPLN` and `ms-mixPLN` [12, 13] which consider the problem of inferring multiple microbial networks (one per host condition) from a given sample-taxa count matrix when microbial associations are impacted by host factors. [14] proposed Hybrid Approach foR MicrobiOme Network Inferences via Exploiting Sparsity (`HARMONIES`) algorithm which address some critical aspects of abundance data (compositionality due to fixed sampling depth, over-dispersion and zero-inflation of the counts) while maintaining computational scalability and sparsity of the interaction network, in contrast to `mixPLN` and `ms-mixPLN`. Finally, Network Construction and comparison for Microbiome data (`NetCoMi`) [11], provides a one-stop platform for inference and comparison of microbial networks, where many existing methods for data preprocessing, network inference and edge selection procedures are implemented. All of the aforementioned inference methods rely on different mathematical models and implicit biological assumptions, making the selection of the most suitable inference technique for a given dataset perilous. We consider a two-step aggregation method for network reconstruction.

The first consensus step occurs at the method level. Indeed, a common point of all these methods is the use of a regularization parameter, usually denoted λ, which is selected using resampling approaches and produces many bootstrap networks for each value of λ. Most procedures relies on selecting the best value $\hat{\lambda}$ from those networks before doing a "refit" operation, where the network computed on the whole dataset for this $\hat{\lambda}$ is returned. We avoid this refit step and instead leverage the bootstrap networks to select only the edges recovered with high-frequency, typically more than 0.9, in the bootstrap networks. This is akin to bootstrap estimates [15] and should in principle lead to more robust and reproducible edges.

The second aggregation step occurs across methods and combines edges (and their inclusion frequency) over methods to build the final network. This is akin to an ensemble procedure where many methods are combined together. In order

to produce a stable consensus network, we explore different aggregation strategies to mitigate the drawback of each method and hopefully reconstruct a better network.

In this study, we included eight inference techniques, all of which rely on Gaussian Graphical Models (GGM) to estimate conditional dependencies networks: Magma [16], SpiecEasi [8], gCoda [17], PLNetwork [10], EMtree [18], SPRING [19], ZiLN [20] and cozine [21]. We excluded `mixPLN` and `ms-mixPLN` as they do not reconstruct a single network but a collection of networks and `NetCoMi` as it collects already existing methods rather than introducing a new one. Finally, we left out `Harmonies` from the comparison, as it does allow the user to set the regularization grid, a crucial step in the ensemble method.

The paper is organized as follows: Section 2 presents different common modeling strategies used to address the specificity of microbial abundance data and introduces the general framework of gaussian graphical models (GGM). Section 3 covers both the mathematical details of GGM-based network inference for abundance data and of the aggregation criteria used in the ensemble method. It also explains how to improve the precision and the reproducibility of each individual method with an original use of the stability-selection principle. Section 4 describes the inference performance criteria used to assess the efficiency and accuracy of each inference method and of the aggregation scheme and the synthetic data used for this purpose. Section 5 illustrates the methods on a gut microbiota data set and highlights the differences between the individuals and the consensus network. Finally, Section 6 discusses the limits of the methods and potential developments.

2 Network microbial interactions

Network-based analytical approaches have proven to be useful to study systems with complex interactions and represent powerful tools to infer complex networks. The most traditional and commonly used approach to identify microbial interactions is the construction of microbial co-occurrence networks but leads to spurious results as interactions occurring in are tricky to model. In this article, co-occurrence networks are modeled with conditional dependence measures as opposed to unconditional correlation measures [6]. The most common framework for the estimation of the conditional dependencies is Gaussian Graphical Models (GGM) [22] which describe the conditional dependency structure of multivariate Gaussian distributions. GGM are so popular because of the Markov faithful property which states that for a random vector X of size p such that $X \sim \mathcal{N}_p(\mu, \Sigma)$, the conditional dependencies are exactly given by the non-null entries of its precision matrix $\Omega = \Sigma^{-1}$.

Obviously, abundance data does not directly fit with the gaussian framework but many network inference methods dedicated to abundance data actually rely on latent gaussian structures. The following subsections describe the classical approaches to shift to gaussian universe.

2.1 Transformations

The first idea when it comes to transposing counts into gaussian framework is to use mathematical transformations. Let us consider a microbial system of p quantitative variables $X = (X^1, \ldots, X^p)$ and n observations $(X_1^j, \ldots, X_n^j), j \in \{1, \ldots, p\}$ of these p variables. Note that X_i^j denotes the count of species j in sample i.

2.1.1 Log transformation

A common transformation is the log function, with a small constant, usually 1, added to each count to avoid zeros. However, it does not stabilize data variability because of depth sequence variation from one sample to another, making dependencies modeling tricky.

2.1.2 CLR transformation

On the contrary, the Centered Log Ratio (CLR) transformation [23] guarantees the study of dependencies, and is defined as:

$$\forall i \in [\![1, n]\!],\ \widetilde{X}_i = \big(\log \frac{X_1}{m(X)}, ..., \log \frac{X_n}{m(X)}\big), \ \text{ where } \ m(X) = (\prod_{k=1}^{n} X_k)^{1/n}$$

It is however criticized when data contain a high proportion of zeros, like whole-metagenome or amplicon sequencing data. To circumvent this problem, [19] introduce a modified version of the CLR transformation that respects the original

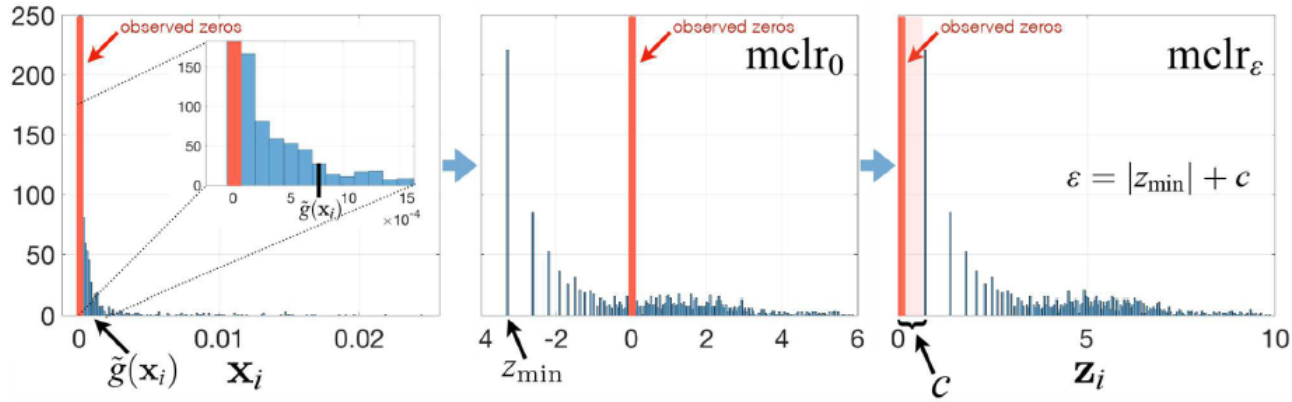

Figure 1: Modified CLR_ϵ transformation, extracted from figure 1 from [24]

ordering of the data but does not account anymore for the compositional nature of the data. The key steps of the $mCLR_\epsilon$ transform are described below and visualized in Figure 1:

$$\forall i \in [\![1,n]\!],\ \tilde{X}_i = \big(0,\ldots,0,\log\frac{X_{i(q+1)}}{\tilde{m}(X)}+\epsilon,\ldots,\log\frac{X_{in}}{\tilde{m}(X)}+\epsilon\big),$$

$$\text{where}\ \ \tilde{m}(X) = (\prod_{k=q+1}^{n} X_k)^{1/(n-q)},$$

$$\text{and}\ \ \epsilon = |min(\tilde{X}_{ij})| + c,\ \ \text{with}\ \ c\ \ \text{usually taken as}\ \ c = 1.$$

2.1.3 GMPR transformation

Another way to account for sampling efforts is to compute the Geometric Mean of Pairwise Ratios [25]. The GMPR normalization is designed for abundances with a high proportion of zero values. It compares pairs of samples based only on the species they share, and considers the geometric mean of the median ratio. This makes this technique robust to both differentially abundant species and extreme values. The normalization factor is computed as follows:

$$\forall i,k \in [\![1,n]\!],\ \tilde{X}_i = \left(\prod_{i=1}^{n} \underset{j\in\{1,\ldots,p\}|X_i^j.X_k^j\neq 0}{\text{Median}} \left(\frac{X_i^j}{X_k^j}\right)\right)^{1/n}$$

When two samples do not share any species, the computation of GMPR fails (this happens when samples come from very contrasted conditions with no or limited species overlap).

2.1.4 RLE transformation

In this case Relative Log Expression (RLE) normalization method is used [26]. This method is based on the assumption that most of the species are not differentially abundant. The RLE scaling factor of a given sample is calculated as the median of the ratio, for each species, of its abundance over its geometric mean across all samples:

$$\forall i \in [\![1,n]\!],\ \tilde{X}_i = \underset{j\in\{1,\ldots,p\}}{\text{Median}} \left\{ \frac{X_i^j}{\left(\prod_{i=1}^{n} X_i^j\right)^{1/n}} \right\}$$

Unfortunately, this normalization factor fails when no single species is shared across all samples, which is frequently the case in microbiome data. A modified version of RLE only considers positive counts to avoid this drawback.

2.2 Distributions and models

The second idea is to transpose counts into gaussian variables using distributions and models adapted to count data characteristics : overdispersion (excess of variability in the data) and zero-inflation(excess of zeros).

2.2.1 Poisson-log normal models

Poisson-log normal models count for both structuring factors (environmental characteristics or gradients, such as habitat type or nutrient availability) and potential interactions between the species (competition, mutualism, parasitism, etc.), which is instrumental in disentangling meaningful ecological interactions from mere statistical associations. Modeling the dependency between the species is challenging because of the count-valued nature of abundance data and most distributions for multivariate counts rely on a Gaussian latent layer to encode the dependencies between species in a covariance matrix.

The Poisson-log normal (PLN) model [27] is designed for the analysis of abundance tables, that is typically the $n \times p$ count matrix X introduced above. It accounts for both structuring factors and potential interactions between the species. It verifies:

$$\begin{cases} Z_i \sim \mathcal{N}(0, \Sigma), \\ X_{ij}|Z_i \sim \mathcal{P}(\exp(o_{ij} + Z_{ij})) \end{cases}$$

Where Z_i are iid, and the X_{ij} are independent conditionally on Z_i. Dependencies are all stored in the Σ covariance matrix which allows to transpose the problem to the latent Gaussian space.

2.2.2 Negative binomial

In the presence of overdispersion, the Poisson regression model is not adequate and can lead to biased parameter estimates and unreliable standard errors estimates. The Negative Binomial (NB) model is then often used [28]. It models the number of failures (denoted k) in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of successes (denoted r) occurs. If a random variable X follows a negative binomial distribution, then the probability of experiencing k failures before experiencing a total of r successes can be found by the following formula:

$$P(X = k) = \binom{k + r - 1}{r - 1}(1 - p)^k p^r,$$

where p is the probability of success. Note that this model can be written as a compound Poisson-Gamma hierarchical model which mirrors the Poisson-log normal.

2.2.3 Zero-Inflated model

Contrary to the NB model, the zero-inflated model [29] is often motivated by an excess of zeros in the data, but less flexible to that the zero outcome. The associated probability function is then:

$$\begin{aligned} P(X = 0) &= \pi + (1 - \pi) \times p(X = 0) \\ P(X = h) &= (1 - \pi) \times p(X = h) \end{aligned}$$

where π is the mixture weight that determine the amount of zero-inflation, $p(X)$ is a truncated probability distribution function and h is any value in the domain of $p(X)$.

2.2.4 Hurdle model

Hurdle models [30] are a class of models for count data that help handle excess zeros and overdispersion. An intuitive approach to analyzing zero-inflated count data is to view the data as arising from a mixture distribution of a point mass distribution at zero and a count distribution. In such a model, a random variable X is modelled as:

$$\begin{cases} P(X = 0) = \theta, \\ P(x \neq 0) = p_{X \neq 0}(X) \end{cases}$$

where $p(X)$ is a truncated probability distribution function, truncated at 0. In contrast to Zero inflated-models, hurdle models capture both an excess or a lack of zeros in the dataset.

2.2.5 zero-inflated negative binomial model

The zero-inflated negative binomial (ZINB) model [31], obtained by applying ZI to NB model, both overdispersion and excess of zeros. The probability distribution of the ZINB random variable X_i can be written either as:

$$P(X_i = x_i) = \begin{cases} \pi_i + (1-\pi_i)(\frac{\theta}{\mu_i+\theta})^\theta, \text{ if } x_i = 0 \\ (1-\pi_i)\frac{\Gamma(\theta+x_i)}{\Gamma(x_i+1)\Gamma(\theta)}(\frac{\mu_i}{\mu_i+\theta})^{x_i}(\frac{\theta}{\mu_i+\theta})^\theta, \text{ if } x_i > 0, \end{cases}$$

where $0 \leq \pi_i \leq 1$, $\mu_i \geq 0$ and $1/\theta$ is the positive overdispersion parameter.

2.2.6 Copulas

Copulas are a multivariate cumulative distribution functions for which the marginal probability distribution of each variable is uniform on the interval $[0, 1]$. As they fully describe the dependence structure, models with copulas allow to separate the modeling of marginal distributions from the modeling of dependencies. Recent developments use gaussian copula coupled with discrete marginal distributions to study multivariate count data [32].

Formally, let ϕ be the Cumulative Distribution Function (CDF) of the standard normal distribution $\mathcal{N}(0, 1)$. Then, corresponding Gaussian copula C_G verify the following property:

$$P(X_1 \leq x_1, \ldots, X_p \leq x_p) = C_G(\phi^{-1}(X_1) \leq \phi^{-1}(x_1), \ldots, \phi^{-1}(X_p) \leq \phi^{-1}(x_p))$$

The first step is then to estimate each of the p marginal univariate discrete distribution and then construct a gaussian copula which satisfies different types of dependence structures. [33] showed that Gaussian copulas are a relevant and promising approach to the problem of network inference from abundance data, even if the computational cost remains substantial. One way of taking advantage of the copula theory without having to actually estimate the joint distribution is to use copulas as a data transformation.

2.3 Latent variables

The third idea is then to model multivariate discrete data using latent variables. Recently, latent variables models have received attention as they provide a convenient way to model the dependence structure between species. Two specifications of latent variables stand out in community ecology [34]: the Multivariate Generalized Linear Mixed Model (GLMM) [35, 36], and the Latent Variable Model (LVM) [37, 38]. The difference between these models lies in the dimension of their respective random effects: there are as many latent variables as there are species in the GLMM, whereas in the LVM their number is a parameter of the model.

In more details, a GLMM models the latent variables and abundances as follows:

$$\begin{gathered} X_{ij}|Z_i \sim \mathcal{F}(m_{ij}, \phi_j), \\ Z_i \sim \mathcal{N}(0, \Sigma). \end{gathered}$$

where F is a distribution with mean m_{ij} and dispersion parameter ϕ_j and with given correlation matrix Σ.

3 Microbial network inference

As stated in the previous section, most methodologies to infer networks from count data first model count data in a way to transpose the problem to the Gaussian setting. There they take advantage of the GGM framework to perform network inference using penalized or tree approaches to estimate the precision matrix Ω.

3.1 Classical inference approaches

3.1.1 Penalized approaches

There exist two main penalized approaches for the estimation of GGM: the graphical LASSO (glasso) [39], and the neighborhood selection, also called the Meinshausen-Bühlmann approach (MB) [40]. Both are penalized approaches which perform a sparse estimation of Ω, either all at once for the glasso or row by row for MB.

In this section we denote by $X = (X^1, ..., X^p) \sim \mathcal{N}_p(0, \Omega^{-1})$ the $n \times p$ abundance matrix. The glasso algorithm introduces sparsity in the precision matrix by imposing ℓ_1 penalties on its non-diagonal entries. The Ω matrix is then estimated by solving the following optimization problem using the ℓ_1 penalty:

$$\underset{\Omega \geq 0}{\arg\max} \{ \log \mid \Omega \mid -tr(X^\intercal X \Omega) - \lambda \mid\mid \Omega \mid\mid_1 \}, \quad \mid\mid \Omega \mid\mid_1 = \sum_{j \neq k} \mid \omega_{jk} \mid .$$

where $\Omega = (w_{ij}^2)$ and λ is the penalty term to get a sparse network. The edges of the network are selected as non-null entries of Ω, by forcing some of them to zero in the precision matrix.

The neighborhood selection on the other hand, takes advantage of the regression interpretation of the precision coefficients : $-\frac{w_{jk}}{w_{jj}}$ is the regression coefficient of X^j against X^k. It estimates a sparse graphical model by fitting a collection of lasso regression models, using each variable as a response and the others as predictors. Namely, the regression of column X^j on all other columns of X writes:

$$\forall j \in [\![1, p]\!], \; X^j = \sum_{k \neq j} \beta_{jk} X^k + \varepsilon_j, \quad \text{where} \;\; \varepsilon_j \sim \mathcal{N}(0, \omega_{jj}^{-1}), \; \text{and} \;\; \beta_{jk} = -\omega_{jk}/\omega_{jj}.$$

Therefore a penalized regression of a node on all other nodes will reveal its neighbors by forcing some coefficients to zero. The neighborhood selection thus simply considers p penalized regression problems $X^j = \sum_{k \neq j} \beta_{jk} X^k - \lambda \sum_{k \neq j} X^k + \varepsilon_j$ with $j \in \{1, ..., p\}$.

With a complexity in the order of $\mathcal{O}(np^2)$, the MB algorithm slightly outperforms the glasso one ($\mathcal{O}(p^3)$).

3.1.2 Tree averaging approach

Another method to estimate GGM we consider in this article is the tree averaging approach [41], which leverages specific algebraic properties to perform a complete and efficient exploration of the space of spanning tree structures. It assumes a hierarchical model, where the data is Gaussian conditionally to a random tree T with β weights on its edges, such that :

$$\begin{cases} T \sim \prod_{jk} \beta_{jk}/B, \quad B = \sum_T \prod_{jk \in T} \beta_{jk}, \\ X|T \sim \mathcal{N}(0, \Sigma) \end{cases}$$

The approach highly relies on the Matrix Tree Theorem, which allows to efficiently sum on all spanning trees. An Expectation-Maximization algorithm maximizes the likelihood and yields the probability for edges to be part of T:

$$P(jk \in T \mid X) = \sum_{T \in \mathcal{T}, jk \in T} p(T \mid X).$$

Note that this approach does not need the GGM Markov faithful property. Thresholding probabilities then give a network.

3.2 Adapted network inference methods to abundance data

We considered in this article only inference methods representative of the different characteristics previously mentioned. The eight methods selected are: SpiecEasi [8], gcoda [17], SPRING [19], Magma [16], PLNetwork [10], EMtree [18], ZiLN [20] and cozine [21]. Table 1 summarizes the inference strategies adopted by each method selected, the computation time when applied on a 500×104 microbial count table, and the potential integration of covariables in the model.

3.3 Consensus bootstrap

Each inference method mentioned above assigns a score to the edges: either a probability (for the tree averaging method) or the maximal penalty level at which the edge is selected in the network. An optimal penalty λ^* on these scores is then needed for an edge to be selected in the final network.

Several approaches exist but the concept of stability selection [42] is the most interesting for this work as it yields a compromise between precision and recall, while fostering reproducibility. The associated method, called Stability Approach to Regularization Selection (StARS) and originally developed for regularization, uses a resampling strategy

Method	Counts and abundances	Distribution	Inference approach	Covariates	Reference
SpiecEasi	CLR	Multivariate gaussian	MB	No	[8]
gCoda	CLR	Multivariate gaussian	glasso	No	[17]
SPRING	CLR	Copulas	MB	No	[19]
Magma	GMPR + RLE	Copulas + ZINB	MB	Yes	[16]
PLNetwork	GMPR + RLE	PLN + Latent variables	glasso	Yes	[10]
EMtree	GMPR + RLE	Latent variables	Tree averaging	Yes	[18]
ZiLN	CLR	Latent variables	MB	No	[20]
cozine	CLR	Hurdle gaussian	MB	Yes	[21]
NetCoMi	All	All	All	Yes	[11]
HARMONIES		ZINB	glasso	No	[14]
mixPLN		multivariate PLN	glasso	No	[12]

Table 1: Network inference characteristics : general methods, counts and abundances normalization (Centred Log Ration (CLR), Geometric Mean of Pairwise Ratios (GMPR), Relative Log Expression (RLE)), distribution transformations, inference approaches (Meinshausen-Bühlmann (MB), glasso, tree averaging), covariates integration and references.

to select the value of λ leading to the most stable graph. This is achieved by creating subsamples, reconstructing a graph for each subsample and computing for each edge in the network its frequency of selection in the graphs reconstructed from the subsamples.

We describe the StARS algorithm formally in the following and the modification we propose in this work.

3.3.1 Edge selection frequencies

The original data$X$ is subsampled B times and the inference is conducted on each sub-sample for each value of λ in a grid $(\lambda_1, \ldots, \lambda_K$ to obtain a graph $G^{b,k}$, with $k \in \{1, ...K\}$.

The selection frequency of edge e for parameter λ_k, is computed as its frequency of selection in the subgraphs:

$$f_e^k = \sum_{b=1}^{B} \mathbf{1}_{\{e \in G^{b,k}\}}.$$

The selection frequency gives an idea of an edge reproducibility (selection percentage of edges over all resamples): frequency and robustness of the edges are clearly related.

3.3.2 Stability coefficients

The stability measures the overall variability of edges selection among the set of resampled inferences. For a given parameter λ_k, it is defined as follows:

$$S^k = 1 - 4 \underbrace{\frac{1}{q} \sum_e f_e^k (1 - f_e^k)}_{\text{Mean of Bernoulli variances}} ,$$

where $q = p(p-1)/2$ is the total number of possible edges.

Each value λ_k is linked to a single selection frequency vector, and a resulting stability value. Finding the right edge frequency is therefore equivalent to finding the right stability level. Classical choices for stability are $stab = 80\%$ or $stab = 90\%$ to have a good compromise between recall and precision and the optimal level λ^* is chosen as $\lambda^* = \min_{\lambda} S(\lambda) \geq stab$.

3.3.3 Refit

Once the optimal level λ^* is fixed, one solution is to return the refit coefficient (1 if the edge is selected and 0 otherwise) computed by running inference methods on the complete dataset, to get the final network.

3.3.4 High frequencies

Instead of using the refit, we propose to select directly the edges, presenting a high selection frequency for the optimal penalty λ^*. In this way, we guarantee both precision and reproducibility as high frequencies are associated with reproducible edges as selected many times in the resampling. The set $E^{\lambda^*}(c)$ of selected edges is defined as:

$$E^{\lambda^*}(c) = \{e, \quad f_e^{\lambda^*} > c\},$$

where c is a constant close to 1. Two advantages of using frequencies rather than refitting the network are (i) that we filter out edges with low support that could be included in the refit graph and (ii) it makes easier to combine edges reconstructed using different methods.

Remark: Note that λ^* depends on the value chosen for the target stability level.

3.4 Inference aggregation

All of the methods selected depend on tuning parameters that are estimated to find the best version of the network to rebuilt. As each inference technique rely on distinct mathematical mechanisms, we get as a final result a family of networks. We then propose to develop an ensemble learning method, called aggregation inference, to infer a consensus network considering each method advantages and counter-balancing each method's estimation defaults.

The aggregation is the process of taking multiple input values and then using them to produce a single output value via standard metrics. In this article, we compute three different aggregation measures. Denoting by f_m the selection frequency for the edge under study computed with method $m \in \{1, ..., M\}$, aggregations are computed as follows:

- *mean*: average of an edge selection frequencies $\sum_m f_m \Big/ M$,
- *norm2*: the euclidean norm (2-norm) $(\sum_m f_m^2)^{1/2} \Big/ M^{1/2}$,
- *IVW*: the inverse-variance weighted average $\left(\sum_m f_m \times \frac{1}{f_m(1-f_m)}\right) \Big/ \left(\sum_m \frac{1}{f_m(1-f_m)}\right)$, where f_m follows a Bernoulli with $\widehat{Var}(f_m) = f_m(1 - f_m)$.

The next sections are dedicated to experimental studies to assess the numerical performances of the aggregation inference algorithm. First, we introduce the criteria we need to evaluate inference performances.

4 Numerical experiments

4.1 Evaluation criteria

A simple way to assess each method performance is to use criteria based on the comparison between the infered network structure and the real network structure:

- Precision (positive predictive value): PPV= TP/(TP+FP),
- Recall (true positive rate): TPR= TP/(TP+FN),

where TP stands for True Positive (an edge that is rightfully detected), FP for False positive and FN for False negative. The precision measures the proportion of positives among what is detected, whereas the recall measures the proportion of positives detected.

4.2 Datasets

In this work we simulate data using the methodology described in [19] which is based on reshufflering an original dataset to preserve and mimic the peculiarity of count data. This method yields synthetic data with marginal distributions that are closer to the original empirical dataset, while enforcing a given correlation structure between the species.

4.2.1 Empirical dataset

The empirical dataset selected, from [43], corresponds to stool samples from 216 Chinese individuals sequenced using whole-metagenome sequencing techniques, a dataset deposited in the European Nucleotide Archive (ENA) under

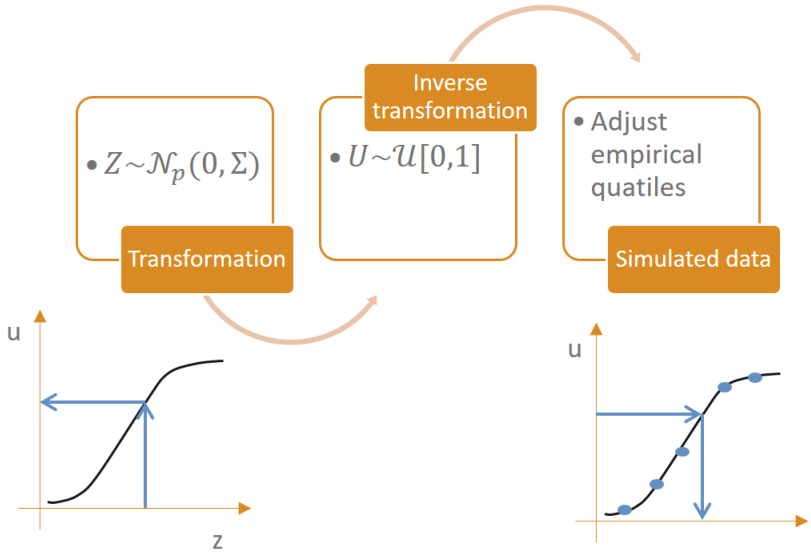

Figure 2: Simulation steps inspired from [24]

accession number PRJEB6337. Among this population, 102 individuals are healthy and 114 are suffering from liver cirrhosis. Abundances of all microbial species (metagenomic species or MSP) detected using 10.4 million IGC2 gut gene catalogue [44] are extracted using the Meteor software suite that create a gene count table by mapping high quality reads onto the gene catalogue, using Bowtie2. Abundance of each MSP is computed as the mean abundance of 100 marker genes selected for each MSP, where the gene abundance is the read count normalized by gene length. The final table of size 1990 MSP by 216 individuals which record all the normalized abundances [45].

4.2.2 Simulated dataset

The simulated dataset considers the framework of an unknown undirected graph $G(V, E)$, with no retroactive loop, consisting of p vertices $V = \{1, \ldots, p\}$ and a set of edges $E \subseteq V \times V$ connecting each pair of vertices. As usual, the graph G is represented by its associated adjacency matrix $A = (A_{ij})_{(i,j)\in E}$ of size $p \times p$, defined as:

$$\forall (i,j) \in [\![1,p]\!]^2, \quad A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

The package 'EMtree' v.1.1.0 [46] is used to generate a precision matrix Ω defined as the graphical Laplacian A, from a cluster graph. Using Ω, we generate the correlation matrix Σ by taking the inverse of the precision matrix. The idea is then to simulate variables with arbitrary marginal distributions from multivariate normal variables with correlation structure given by Σ. Specifically, we generate $n \times p$ matrix Z with independent normal rows $Z_i \sim \mathcal{N}(0, \Sigma)$. We then get uniform random vectors by applying standard normal cdf transformation to each column of Z, $u^j = \phi(Z^j)$ element-wise, and finally apply the quantile functions of the empirical data marginal distributions to each u^j. The function *synthData_from_ecdf* from the 'SPRING' package v.1.0.4 [47] is used for these simulation steps and summarized in the Figure 2.

4.3 Effect of the dimension on stability/density

In this sub-section, we study the relationship between each method stability and the dimension of the abundance table. Data is simulated from the liver cirrhosis dataset restricted to diseased individuals. Only metagenomic species with a prevalence greater than 50% are studied, yielding 159 species. In this experiment, the sample size varies in $\{50, 100, 500, 1000\}$ and we consider $B = 40$ resamples each time. Instead of fixing a target stab at 0.8 or 0.9 as usual, we let it vary and study the impact of stab on the reconstructed graph.

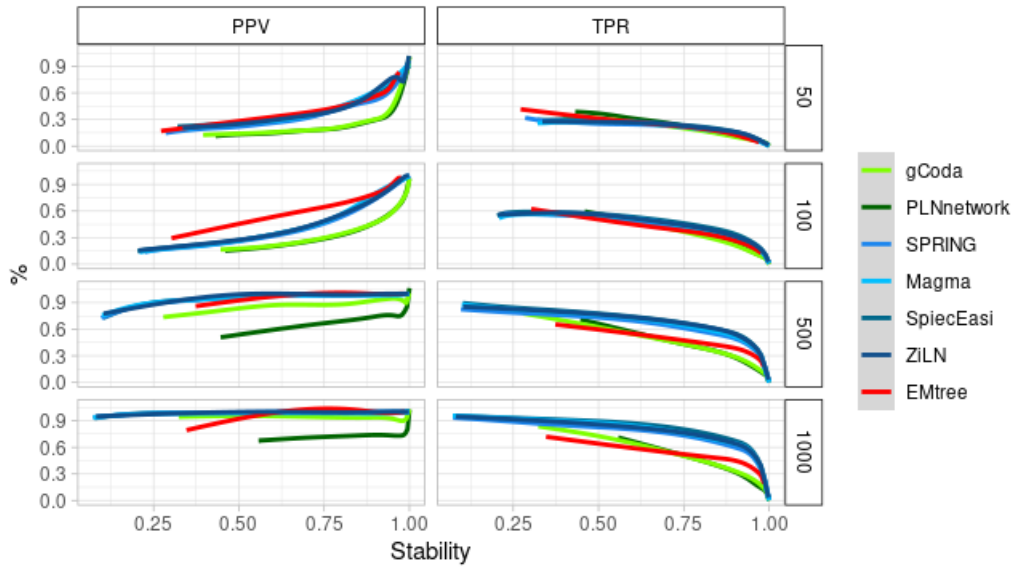

Figure 3: Stability with the edge set $E^{\lambda^*}(0.90)$ according to the sample size and the inference method

4.3.1 Precision/Recall vs. Stability

Figure 3 shows the relationship between the performance obtained with the edge set $E^{\lambda^*}(0.90)$ (precision PPV90 and recall TPR90), and the corresponding stability. The difference in patterns grows with the sample size n, revealing peculiarities inherent to each method. Clearly, methods have a very distinct performance for the same stability level which shows that methods do not share the same precision/recall compromise.

As a result, they produced edge sets that greatly differ in both size and quality. This suggests that we should not use the same stab value for all methods. Ideally, we would like to calibrate all methods to achieve the same precision level.

4.3.2 Precision/Recall vs. density

Precision is obviously unavailable when dealing with empirical datasets, but the number of detected edges, or density, is. Figure 4 shows the relationship between precision and density for all methods and increasing sample size. Curves are almost superimposed for values of n up to 100, and then different behaviors appear. However, the gap in performance between methods stays smaller when imposing the same density than when imposing the same stability. This encourages to bypass stability in the favor of density, meaning that density can act as a mediator to obtain close precision levels for all methods.This also means that the top n-edges of each method achieve the same graph reconstruction quality, even though they are characterized by different stabilities.

4.3.3 Mean stability

The previous observation prompted us to explore the link between stability and density for different values of λ.

Figure 5 shows how stability decreases with increasing density. Note that all methods have similar density precision curves. We search for a common density that leads to a target mean stability of 0.9.

Then, setting a target mean stability value (e.g. 90%) yields a corresponding target density level for each value of n (here between 100 and 250). As n grows, the target density level yields different stability values: for $n = 1000$, stabilities range from 0.8 to 1. Here we can see how targeting the mean stability rather than the same stability for all methods allows to adapt the precision level of each method through density to make them more similar. As an example, let's focus on Magma and SPRING. From previous results, they show different precision/stability curves yet the same precision/density curves. As expected, their final stability values are very different.

4.4 Point of stability per method, per dimension

We now aim at exploring the sample size on individual method stability. The following experiments result in simulated count tables computed from liver cirrhosis data for a number of samples n in $\{50, 100, 500, 1000\}$.

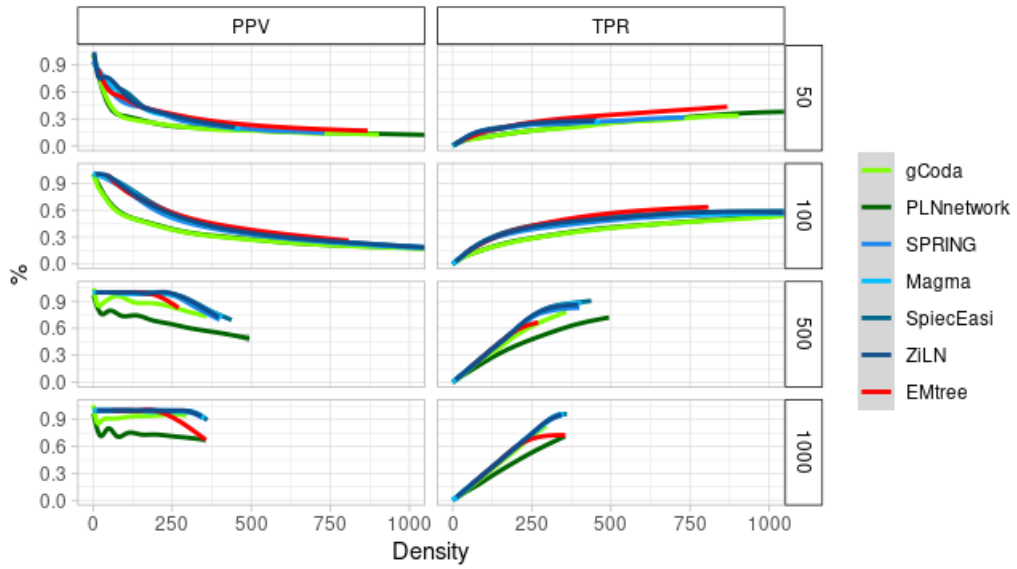

Figure 4: Density with the edge set $E^{\lambda^*}(0.90)$ according to the sample size

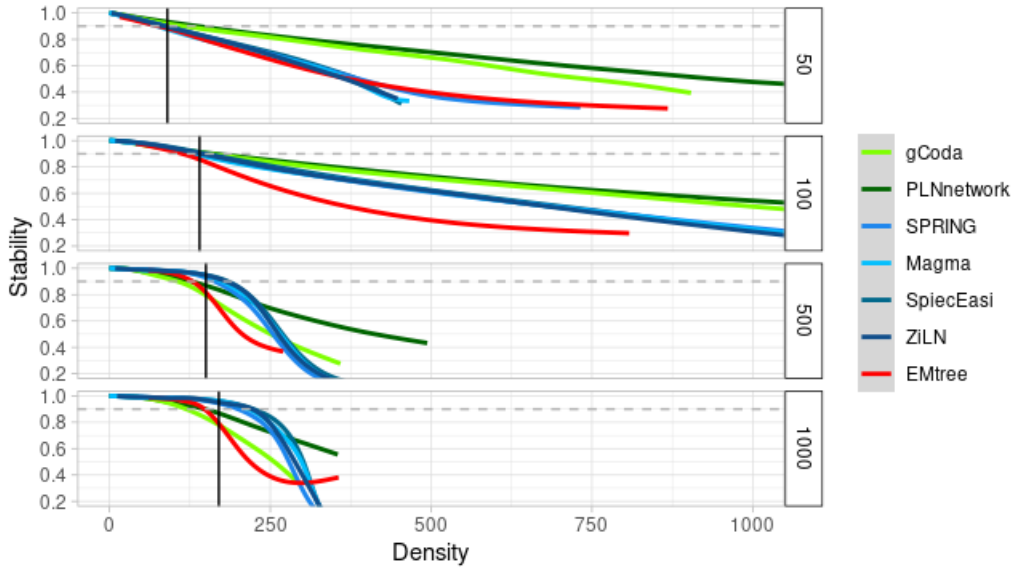

Figure 5: Stability/density curves for all methods

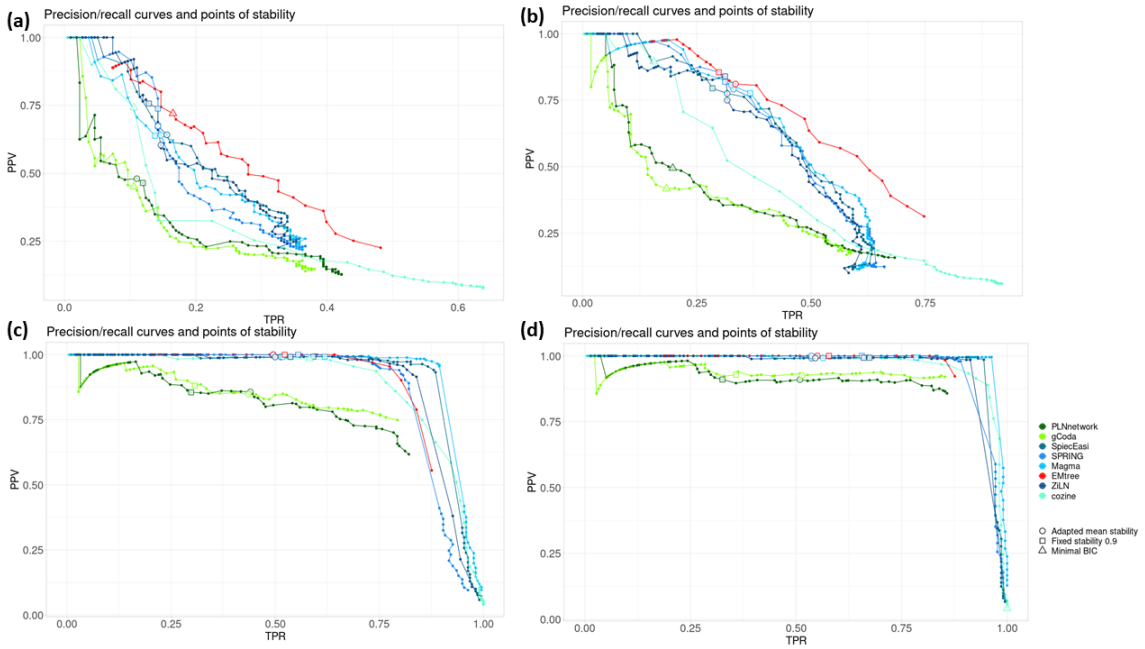

Figure 6: Performances of each methods and adapting stabilities according to the sample size: (a)50 (b)100 (c)500 (d)1000

This experiment consists in applying the different inference methods on simulated data with a varying sample size but a fixed prevalence threshold (0.7). Graphical results are presented in Figure 6 and one can obviously note that the sample size has a major impact on the precision and stability. From $n = 500$, curves are almost stabilized (Figure 6 c).

It is also noticeable that the adapted stabilities reduces the range of the method's precision. Furthermore, for glasso-based methods, new stability target leads to a 20 points improvement in TPR at almost no cost in PPV, for large sample sizes.

Note that we integrated differently cozine method in this experiment as it rely on a BIC criteria as a regularization parameter.

4.5 Quality of aggregated methods

In the previous subsections, we proved that targeting the same stability for each method resulted in different precision levels. We instead aim for a mean stability and select the density/λ value closest to that mean. We now assess the aggregation methods. Note that, because of the common density, each method should provide roughly the same number of edges to the aggregation procedures. This experiment is conducted with $n = 50$ to $n = 1000$. For the frequency threshold of 90%, the quality of both individual and aggregated frequencies are stored in terms of precision and recall, as well as the reproducibility. Results can be visualized in Figures 7 for the precision and 8 for the recall.

The sample size has a dramatic effect on all criteria and methods. With $n \geq 500$, the majority of methods shows median precision values above 99%, and recall between 30% and 60% (Figures 7, 8 (c) and (d)). When decreasing n under 500, the majority of median precision drops between 40% and 85% and the recall between 15% and 40% (Figures 7, 8 (a) and (b)).

Comparing aggregated frequencies to individual ones, we see that aggregations with mean and 2-norm present among the best precision levels, and among the worse recall values. This illustrates how they lead to sparser networks with higher precision.

Note the discrepancy in terms of precision and recall between the two previous of studies of cozine method. Due to the BIC criteria, we changed the procedure between the two experiments.

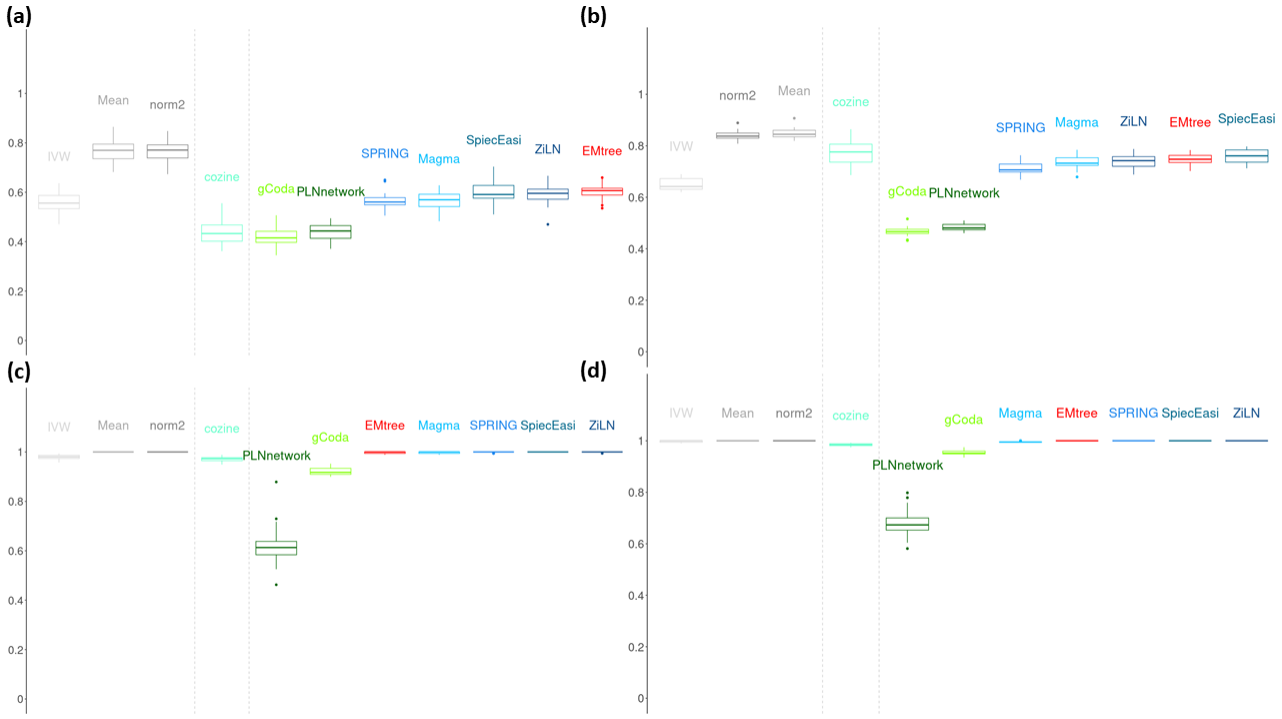

Figure 7: Quality of both individual and aggregated frequencies in terms of PPV for: (a)$n = 50$ (b) $n = 100$ (c) $n = 500$ (d) $n = 1000$

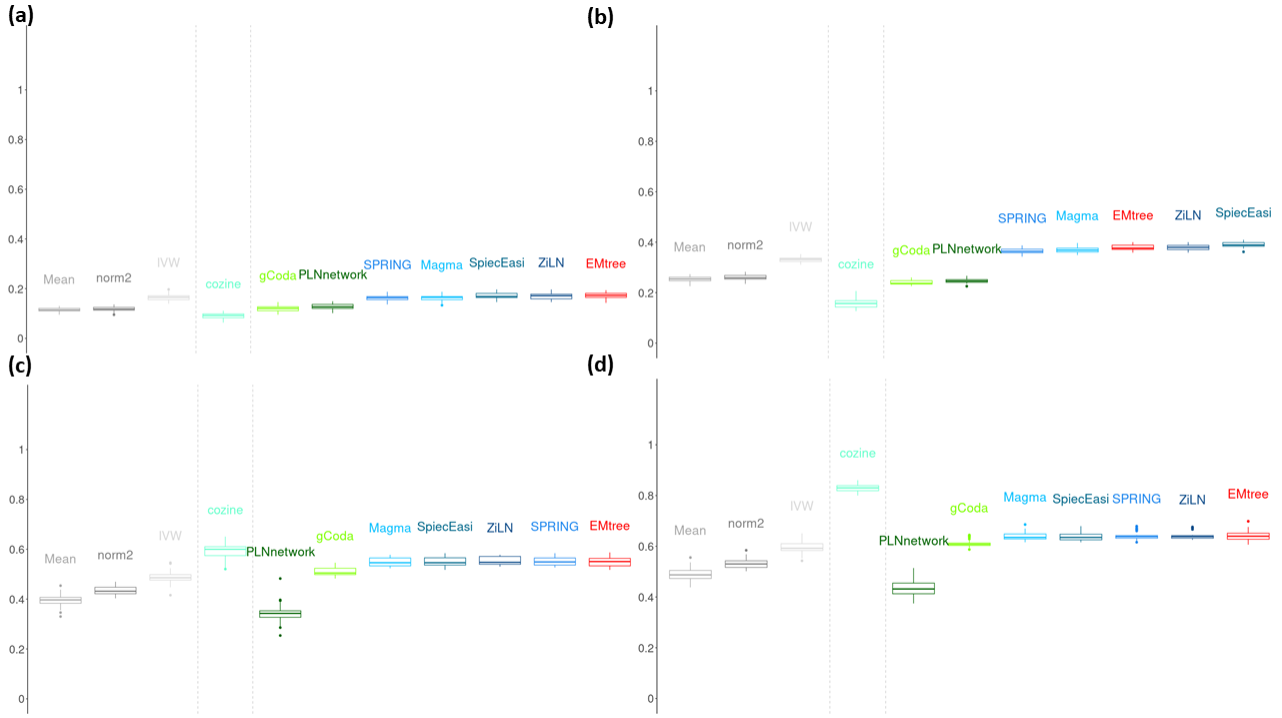

Figure 8: Quality of both individual and aggregated frequencies in terms of TPR for: (a)$n = 50$ (b) $n = 100$ (c) $n = 500$ (d) $n = 1000$

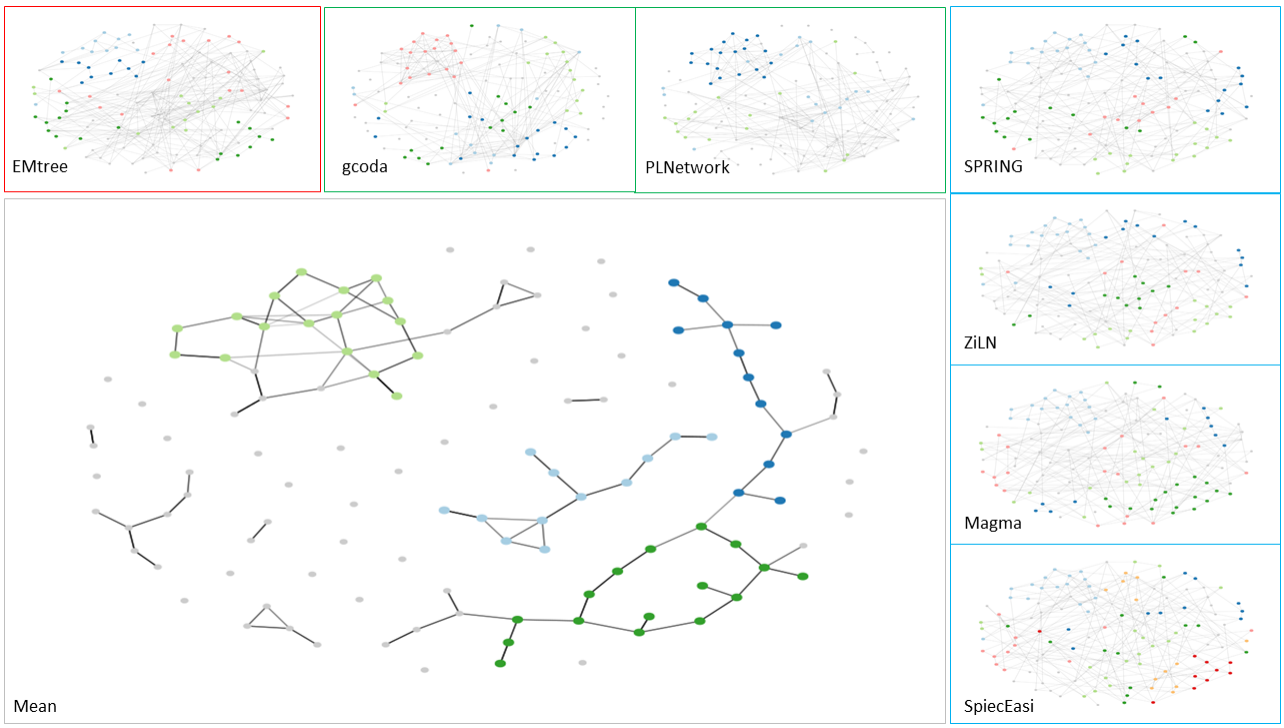

Figure 9: Networks from EMtree, gCoda, PLNnetwork, SPRING, ZiLN, Magma, SpiecEasi and Mean aggregated inference followed by Core-clustering algorithm to identify the microbial guilds. All graphs are shown using the same layout and the nodes are coloured according to their cluster. Methods are grouped based on their inference technique.

5 Application to cirrhosis data

This section is dedicated to the application of all method on the real liver cirrhosis dataset from [43], presented in section 4 followed by a clustering as in [48] to reconstruct microbial guilds associated with liver cirrhosis. Hence, only metagenomic species with a prevalence greater than 50% and ill samples are studied, yielding 155 species and 114 samples to create the networks across methods.

The main objective of this application is first to compare the inferred and clustered networks. Then, we fixed the number of clusters between 10 and 19 following the guidelines of [48]. Figure 9 illustrates the graphs inferred by the methods cited in this article and then clustered using CORE-clustering algorithm. We note that the mean graph inferred is the sparsest one and is closely related to the SpiecEasi and the SPRING's one. This result reflects the idea raised by Figure 7.

The second objective of this application is to characterize microbial species responsible for liver cirrhosis using the mean aggregated network constructed. One guild, represented in Figure 10, drew our attention as it mostly reflects the "cirrhotic guild" presented in [43].

6 Discussion

Microbial network inference is a challenging task, which prompted many works in the last decade. The importance of choosing an appropriate approach to deal with the peculiarities of microbial abundance properties and infer interaction from them must not be disregarded. The microbiota is typically structured with local sub-communities called microbial guilds that may influence the host system. The detection of such guilds using network theory may help to find alternate route for therapeutic strategies to recover microbial composition balance in non healthy people. The most widely used approaches to infer a microbial network are based on GGM, a framework which offers several important advantages (integration of environmental effects as covariates, statistically-grounded rationale, software availability,...). Eight graph inference methods to infer a sparse and reproducible microbial network using GGM were selected. All of them rely on one of three inferences approaches: glasso, neighborhood selection and tree averaging approach. All depend on a so-called regularization parameter that must be optimized to find the best version of the network. They however rely on

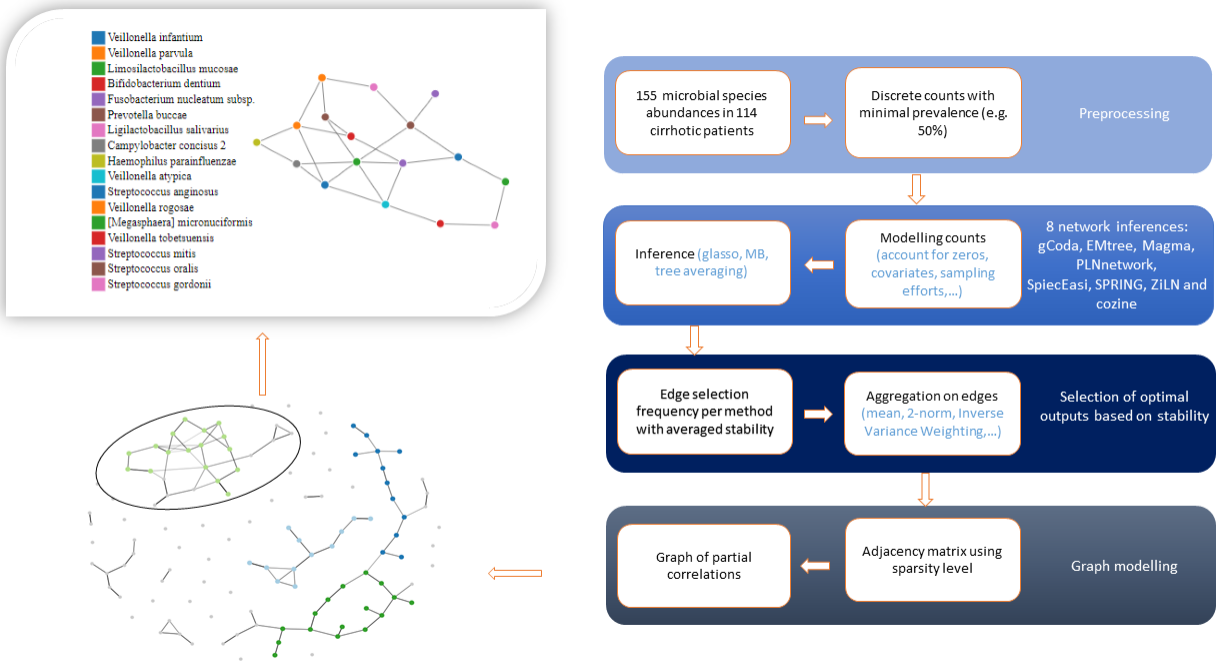

Figure 10: Statistical study steps to find "cirrhotic guild"

different mathematical models, making the selection of the most suitable inference technique for a particular dataset challenging.

Hence, we proposed a two-step aggregation method. The first consensus step relies on selecting edges with high inclusion frequency in the networks reconstructed from resampled data. This first step improves both precision and reproducibility of individual methods and allows us to bypass the fact that stability selection selects very different precision - recall compromises for different methods, even when they have similar precision - recall curves, and to achieve similar precision levels across methods before constructing the consensus. The second aggregation step is the inference of a consensus network considering each method advantages and counter balancing each estimation's default.

The experimental studies designed to compare the efficiency of individual and consensus methods, show that two of the three aggregated methods have the best precision values and among the best reproducibility levels. When considered individually, the neighborhood selection and the tree averaging approaches are really close in terms of performance and robust. In all numerical illustrations, we also prove the impact of the sample size on the precision. An acceptable threshold to maintain high robustness proved to be $n = 100$ samples. Obviously, this threshold can be adjusted according to the prevalence threshold selected, as interactions between highly prevalent species are easier to reconstruct.

The application part of this article is dedicated to the discovery of bacterial species acting on liver cirrhosis. The best aggregation method in terms of precision led to the sparsest graph, in which guilds are easiest to reconstruct, thus validating the usefulness of this method for applications. Therefore, we integrated a clustering method in our aggregated model to discover robust guilds. As a preliminary result, we recovered one guild known for being associated liver cirrhosis.

7 Supplementary Material

7.1 Illustrating the common number of edges selected

In this appendix, we explored the intersections between the edge sets identified by all methods, aggregations, and the true set of edges. The following graphics (Figure 11 and Figure 12) allow the exploration of major intersections of methods and aggregation against truth.

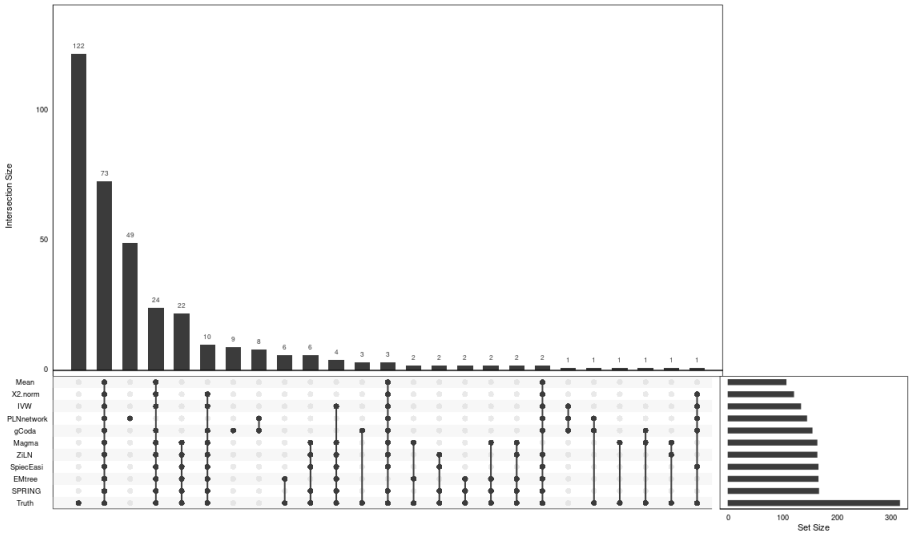

Figure 11: Major intersections of methods and aggregation against truth for $n = 300$

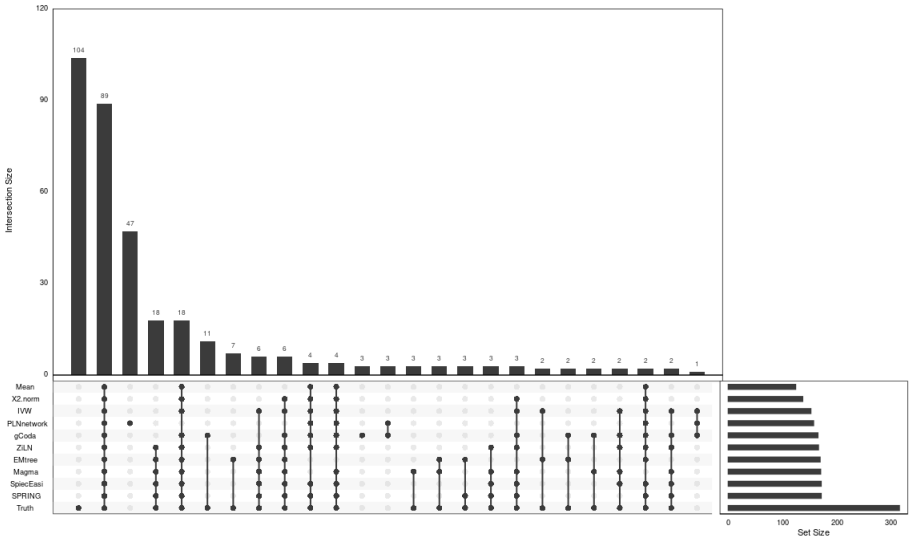

Figure 12: Major intersections of methods and aggregation against truth for $n = 500$

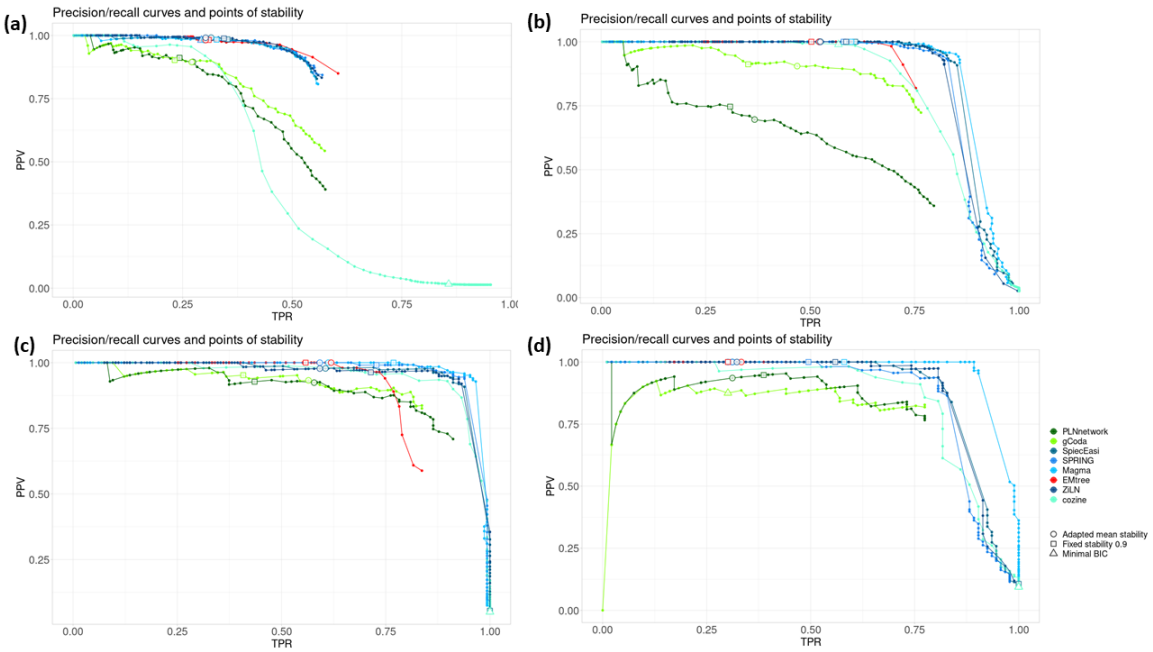

Figure 13: Performances of each methods and adapting stabilities according to the prevalence threshold: (a)0.20 (b)0.50 (c)0.8 (d)0.9

These outputs illustrate the first major and reassuring fact, that methods overall agree with one another and with the truth. With the mean, this represents above 33% for $n = 300$ and 39% for $n = 500$. The aggregation with Inverse Variance Weighting gives a significant additional quantity of edges compared to other aggregations, but at the cost of some additional errors too.

7.2 Illustrating the performances according to the prevalence threshold

This experiment rests on the application of the inference methods on simulated data with a varying feature size but a fixed sample size (500). The feature size is controlled by the prevalence threshold imposed at the very beginning of the analysis. Figure 13 presents the performances of each method depending on the remaining variables. One can notice that the precision is less impacted by the number of species in the system. However, with a prevalence threshold between 0.7 and 0.8, the distance between the precision curves of all methods is reduced (Figure 13 c) and d)).

7.3 Illustrating the quality without glasso methods

We tested the effect of PLNnetwork and gcoda methods on the aggregated ones by removing them from the experiment. Despite the gap in performance of the glasso depending methods with the rest of individual methods, aggregations still show the best precision levels equivalent to that of the best methods (Figures 14 and 15). This reflects the inherent robustness of aggregations to methods with outlying performance. Note that this is not true for the IVW measure, which makes more discoveries and therefore more mistakes.

References

[1] S. Weiss, Z.Z. Xu, S. Peddada, A. Amir, K. Bittinger, A. Gonzalez, and C. Lozupone. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1), 2017.

[2] K. Faust and J. Raes. Conet app: inference of biological association networks using cytoscape. *F1000Research*, (5), 2016.

[3] G. Wu, N. Zhao, C. Zhang, Y.Y. Lam, and L. Zhao. Guild-based analysis for understanding gut microbiome in human health and diseases. *Genome Medicine*, 13(22), 2021.

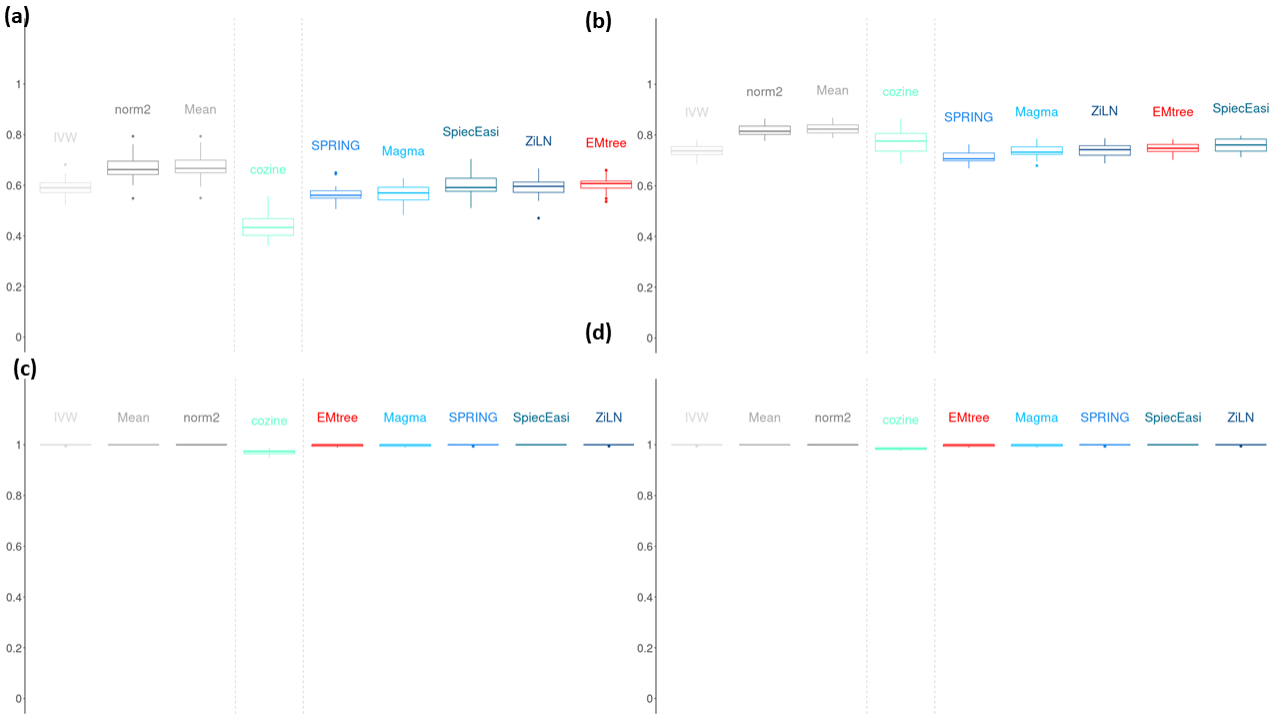

Figure 14: Quality of both individual and aggregated frequencies without glasso methods in terms of PPV for: (a)$n = 50$ (b) $n = 100$ (c) $n = 500$ (d) $n = 1000$

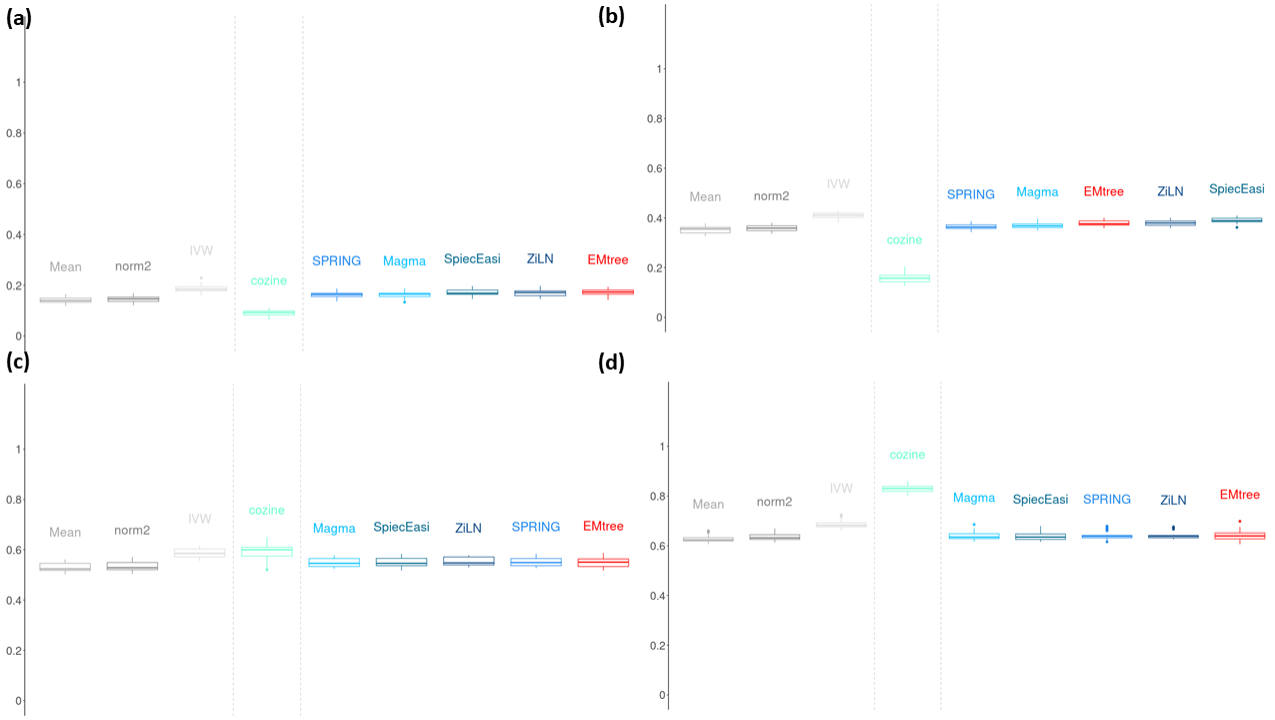

Figure 15: Quality of both individual and aggregated frequencies without glasso methods in terms of TPR for: (a)$n = 50$ (b) $n = 100$ (c) $n = 500$ (d) $n = 1000$

[4] V. Bucci, B.Tzen, N. Li, M. Simmons, T. Tanoue, E. Bogart, L. Deng, V. Yeliseyev, M.L. Delaney, Q. Liu, B. Olle, R.R. Stein, K. Honda, L. Bry, and G.K. Gerber. Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses. *Genome Biol*, 17(121), 2016.

[5] G.B. Gloor, J.M. Macklaim, V. Pawlowsky-Glahn, and J.J. Egozcue. Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology*, 8, 2017.

[6] J. Friedman and E. Alm. Inferring correlation networks from genomic survey data. *PLoS computational biology*, 8, 2012.

[7] H. Fang, C. Huang, H. Zhao, and M. Deng. Cclasso: correlation inference for compositional data through lasso. *Bioinformatics*, 31:3172–3180, 2015.

[8] 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.

[9] S. Biswas. Learning microbial interaction networks from metagenomic count data. *International Conference on Research in Computational Molecular Biology*, pages 32–43, 2015.

[10] J. Chiquet, S. Robin, and M. Mariadassou. Variational inference for sparse network reconstruction from count data. In *International Conference on Machine Learning*. PMLR, 2018.

[11] S. Peschel, C.L. Müller, E. Von Mutius, A.L. Boulesteix, and M. Depner. Netcomi: network construction and comparison for microbiome data in r. *Brief Bioinformatics*, 22(4), 2021.

[12] S. Tavakoli and S. Yooseph. Learning a mixture of microbial networks using minorization-maximization. *Bioinformatics*, 35(14):23–30, 2019.

[13] S. Yooseph and S. Tavakoli. Variational approximation-based model selection for microbial network inference. *Journal of computational biology*, 29(0), 2022.

[14] S. Jiang, G. Xiao, A.Y. Koh, Y. Chen, B. Yao, Q.Li, and X.Zhan. Harmonies: A hybrid approach for microbiome networks inference via exploiting sparsity. *Front. Genet.*, 11(445), 2020.

[15] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.

[16] A. Cougoul, X. Bailly, and E.C. Wit. Magma: Inference of sparse microbial association networks. (538579), 2019.

[17] H. Fang, C. Huang, H. Zhao, and M. Deng. gcoda: Conditional dependence network inference for compositional data. *Journal of Computational Biology*, 24(7):699–708, 2017.

[18] 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.

[19] G. Yoon, R.J Carroll, and I. Gaynanova. Sparse semiparametric canonical correlation analysis for data of mixed types. *Biometrika*, 107(3):609–625, 2020.

[20] V. Prost, S. Gazut, and T. Brüls. A zero inflated log-normal model for inference of sparse microbial association networks. *PLOS Computational Biology*, 17(6), 2021.

[21] M.J Ha, J. Kim, J. Galloway-Pena, K.A. Do, and C.B. Peterson. Compositional zero-inflated network estimation for microbiome data. *BMC Bioinformatics*, 21(581), 2020.

[22] S.L. Lauritzen. *Graphical Models*. Oxford Statistical Science Series, 1996.

[23] J. Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

[24] G. Yoon, I. Gaynanova, and C. Müller. Microbial networks in spring - semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data. *Frontiers in Genetics*, 10:516, 2019.

[25] L. Chen, J. Reeve, L. Zhang, S. Huang, X. Wang, and J. Chen. Gmpr: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ*, 6, 2018.

[26] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome biology*, 11(10), 2010.

[27] J. Chiquet, M. Mariadassou, and S. Robin. The poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *Frontiers in Ecology and Evolution*, 9, 2021.

[28] C. Forbes, M. Evans, N. Hastings, and B. Peacock. *Statistical distributions*. John Wiley and Sons, 2010.

[29] W.H. Greene. Accounting for excess zeros and sample selection in poisson and negative binomial regression models. *Research Papers in Economics*, 1994.

[30] J.G. Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39:829–844, 1971.

[31] Y.B. Cheung. Zero-inflated models for regression analysis of count data: a study of growth and development. *Statistics in Medicine*, 21, 2002.

[32] M.J. Anderson, P. de Valpine, A. Punnett, and A. E. Miller. A pathway for multivariate analysis of ecological communities using copulas. *Ecology and evolution*, 9(6):3276–3294, 2019.

[33] G. C. Popovic, D. I. Warton, F. J. Thomson, F. K. C. Hui, and A. T. Moles. Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution*, 10(9):1571–1583, 2019.

[34] D. I. Warton, F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C Walker, and F. KC. Hui. So many variables: joint modeling in community ecology. *Trends in Ecology and Evolution*, 30(12):766–779, 2015.

[35] O. Ovaskainen, J. Hottola, and J. Siitonen. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91(9):2514–2521, 2010.

[36] L. J. Pollockand R. Tingley, W. K. Morris, N. Golding, R. B. O'Hara, K. M. Parris, P. A. Vesk, and M. A. McCarthy. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (jsdm). *Methods in Ecology and Evolution*, 5(5):397–406, 2014.

[37] O. Ovaskainen, N. Abrego, P. Halme, and D. Dunson. Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, 7(5):549–555, 2016.

[38] O. Ovaskainen, G. Tikhonov, A. Norberg, F Guillaume B., L. Duan, D. Dunson, T. Roslin, and N. Abrego. How to make more out of community data? a conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5):561–576, 2017.

[39] J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[40] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, pages 1436–1462, 2006.

[41] M. Meila and M. I. Jordan. Learning with mixtures of trees. *Journal of Machine Learning Research*, pages 1–48, 2000.

[42] H. Liu, K. Roeder, and L. Wasserman. Stability approach to regularization selection (stars) for high dimensional graphical models. *Advances in Neural Information Processing Systems*, 24(2), 2010.

[43] N. Qin, F. Yang, A. Li, E. Prifti, Y. Chen, L. Shao, and J.Guo. Alterations of the human gut microbiome in liver cirrhosis. *Nature*, 513(7516):59–64, 2021.

[44] C. Wen, Z. Zheng, T. Shao, L. Liu, Z. Xie, E. Le Chatelier, Z. He, W. Zhong, Y. Fan, L. Zhang, H. Li, C. Wu, Q. Xu, J. Zhou, S. Cai, D. Wang, Y. Huang, M. Breban N. Qin, and S.D. Ehrlich. Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. *Genome biology*, 1(18):255–261, 2017.

[45] C. Champion, R. Momal, E. Le Chatelier, M. Mariadassou, and M. Berland. Microbial species abundances from public project PRJEB6337 on liver cirrhosis. `https://doi.org/10.57745/5YXXN1`, 2023.

[46] R. Momal. Emtree: Infers direct species association networks using tree averaging. `https://rdrr.io/github/Rmomal/EMtree/`, 2021.

[47] G. Yoon. Semi-parametric rank-based approach for inference in graphical model (spring). `https://rdrr.io/github/GraceYoon/SPRING/`, 2022.

[48] 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*, 2(14), 2021.