ARZIMM: A Novel Analytic Platform for the Inference of Microbial Interactions and Community Stability from Longitudinal Microbiome Study

Linchen He[1], Chan Wang[2], Jiyuan Hu[2], Zhan Gao[3], Emilia Falcone[4], Steven Holland[4], Martin J. Blaser[3], Huilin Li[2*]

[1]Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA
[2]Division of Biost2atistics, Department of Population Health, New York University School of Medicine, NY 10016, USA
[3]Department of Medicine and Pathology & Laboratory Medicine, Rutgers University, NJ, USA
[4]Division of Intramural Research, Immunopathogenesis Section, NIAID, NIH, MD, USA

Linchen He, PHD,	lh1790@nyu.edu
Chan Wang, PHD	chan.wang@nyulangone.org
Jiyuan Hu, PHD	jiyuan.hu@nyulangone.org
Zhan Gao, PHD	zg138@cabm.rutgers.edu
Emilia Falcone, MD, PhD	emilia.falcone@nih.gov
Steven Holland, MD	sholland@niaid.nih.gov
Martin J. Blaser, MD	martin.blaser@cabm.rutgers.edu
Huilin Li, PHD	huilin.li@nyulangone.org

*Corresponding Author: Huilin Li (huilin.li@nyulangone.org)

Word count: 4593

6 figures, 2 tables

Abstract:

Dynamic changes of microbiome communities may play important roles in human health and diseases. The recent rise in longitudinal microbiome studies calls for statistical methods that can model the temporal dynamic patterns and simultaneously quantify the microbial interactions and community stability. Here, we propose a novel autoregressive zero-inflated mixed-effects model (ARZIMM) to capture the sparse microbial interactions and estimate the community stability. ARZIMM employs a zero-inflated Poisson autoregressive model to model the excessive zero abundances and the non-zero abundances separately, a random effect to investigate the underlining dynamic pattern shared within the group, and a Lasso-type penalty to capture and estimate the sparse microbial interactions. Based on the estimated microbial interaction matrix, we further derive the estimate of community stability, and identify the core dynamic patterns through network inference. Through extensive simulation studies and real data analyses we evaluated ARZIMM in comparison with the other methods.

Key words: absolute abundance, autoregressive, longitudinal microbiome data, microbial community stability, microbial interactions, network analysis, mixed-effects model, and zero-inflated model.

1 Introduction

The human microbiota, a diverse array of microbial organisms living in and on human bodies, form a dynamic ecosystem that plays a critical role in human health. While temporally stable microbial communities are observed among healthy adults [1], the fluctuation of microbiome has been linked to increasing frailty [2] and declining immune function of hosts [3], and diseases such as inflammatory bowel disease [4, 5], colorectal cancer [6, 7], and irritable bowel syndrome [8, 9]. When a microbial community changes in response to an external perturbation, it undergoes a dynamic process and tends to evolve toward another stable state (Figure 1). This dynamic process is stochastic and varies according to the type and strength of perturbation, the community stability prior to the perturbation, and other subject-level relevant features. The recent rise in longitudinal studies, in which microbial samples are collected repeatedly over time, offers unique insights into the responses of such communities to perturbations and the associated dynamic patterns. For example, in our ongoing microbiome study evaluating the effects of antibiotic exposure as a short-term perturbation on microbial, immune, and metabolic physiology (MIME study), we are interested in determining how differently the microbial community responds to the antibiotic treatment.

Human microbiota studies have been accelerated by the advent of next-generation sequencing technologies which enabled the quantification of the composition of microbiomes, often by two common sequencing approaches—16S rRNA marker gene sequencing and shotgun metagenomics sequencing [10]. There are pros and cons to each of those techniques, which are discussed in recent reviews [11, 12]. But for both methods, because of the varying sequencing read counts obtained across samples, it is

necessary to employ various normalization tools to convert raw counts data into relative abundances [13]. However, the dependency of the compositional components greatly hampers the interpretation of microbiota changes in longitudinal studies. There is reason to believe that the absolute abundances of bacteria are biologically meaningful measures, especially in the study of microbial interactions. Thus, in our MIME study, we use an independent quantitative polymerase chain reaction (qPCR) technology [14-16] to quantify total bacterial load per unit sample, and then use these data to estimate absolute bacterial abundance by combining them with the relative abundance values obtained from 16S rRNA or shotgun sequencing methods. This MIME study motivated us to develop analytical methods to investigate microbial interaction and community stability after a strong external perturbation, and identify core active microbial taxa by modeling the absolute abundances of bacteria.

Although many well-developed statistical tools are widely used for assessing the diversity of microbial communities and its composition, there are only a few methods available for inferring the ecological networks of microbial communities. Here we briefly review the well-developed statistical methods for studying the dynamic microbial systems and their limitations.

A Bayesian network contains a set of multivariate joint distributions that exhibit certain conditional independences and a directed and acyclic graph (DAG) that encodes conditional independences among random variables. If the dependence relationships repeat and the signals at a certain time point only depend on the signals from previous time points, then the whole network can be formulated as a dynamic Bayesian network (DBN)[17] representation. McGeachie et al.[18] constructed a simplified two-stage DBN

(TS-DBN) which uses a Markov assumption that the observed values at time $t+1$ are independent of those at earlier time points ($t-1$ and earlier) given the variable values at time t. Lugo-Martinez et al. presented a computational pipeline which first aligns the data collected from all individuals, and then learns a dynamic Bayesian network from the aligned profiles[19]. However, DBN has several limitations in analyzing the longitudinal microbial data. (1) It can only model the microbial community subject-by-subject. (2) DBN cannot handle the exceeding zero structure of microbial counts. Most methods remove the taxa whose relative abundances exhibit zero entry (i.e., not present in a measurable amount at one or more of the measured time points) before the downstream analysis. (3) The assumed distributions are unrealistic. E.g. all continue variables are assumed to be normally distributed. (4) The computational cost is relatively high, since parent nodes are added sequentially for each bacterial node. Additionally, the maximum number of possible parents is imposed, which is not realistic. (5) Due to sampling and sequencing limitations, the compositionality bias in microbiome data may also cause inaccurate estimation of parameters. The existing methods ignore this compositionality bias, making parameter estimates difficult to interpret. (6) Irregular sampling time may also result in inaccurate parameter estimation. Therefore, it is advised to cautiously interpret the findings from DBN[20, 21].

The classical Lotka-Volterra equations has been used to model simple system such as two species in a predator-prey relationship, where the interactions are strictly assumed to be competitive. The generalized Lotka-Volterra (gLV) equations extend the classical predator-prey (Lotka-Volterra) equations, where the interacting species might have a wide range of relationships including competition, cooperation, or neutralism. Assuming that

the interaction (or the effect) of one species with another can be modeled by the corresponding coefficient in the equation, gLV equations provide a framework to analyze and simulate microbial populations. Mounier et al. used the gLV equations to model the interaction between bacteria and yeast in a cheese microbiome[22]. Other microbiome studies further extended and implemented the gLV equations[23-26].

Many software are available for applying gLV modeling on microbial time series data, such as LIMITS[27], MetaMis[28], and MDSINE[29]. LIMITS and MetaMis can be implemented to construct microbial interactions using the longitudinal microbiome data from one subject. MDSINE can jointly analyze multiple time series, but requires Matlab programming. Web-gLV (http://web.rniapps.net/webglv) can be used for modeling, visualization, and analysis of microbial populations, but can only handle limited number of samples. In summary, there are several limitations of gLV in analyzing the longitudinal microbial data. (1) gLV based models capture the interactions using a single averaged effect, thus they are not well-suited for noisy data. (2) Some methods estimate almost all possible edges without incorporating variable selection techniques. (3) gLV estimates the growth rate of each taxon marginally, therefore, ignores the intrinsic dynamic correlations of the repeated measurements. (4) gLV does not account for random processes which forms essential part of any biological system. (5) With the increased number of species and time span of prediction, the simulation output is prone to numerical errors. For example, Web-gLV can only simulate a maximum of 10 species at a time for at most 100 time points. (6) As DBN, gLV is not suitable for sparse, compositional, and irregular sampled microbiome data.

In Ives et al. [30], the stability of a microbial community is determined by three key

interrelated components of microbial community structure: diversity, species composition, and interaction pattern among species. They viewed the dynamics of a microbial community as a stochastic process and proposed to use a first-order multivariate autoregressive process (MAR (1)) time-series model to disentangle the effects of these three components on community stability and to estimate the stability properties of a community by estimating the strengths of interactions between species. This method is widely used to estimate the stability of ecosystems (e.g., lake, ocean) based on culture-dependent microbial data[31, 32]. Usually a few (four or five) key microbes are detected with high frequency in each ecosystem in time-series measurements over a long period, and their abundances are rarely zero. In contrast, our MIME study will yield microbiome data from approximately nine time points over half a year from 80 subjects in three groups in the complete study—a relatively smaller number of repeated microbiome samples but from a relatively larger number of microbial communities (subjects) than what would be the case for an ecosystem study. Moreover, the 16S rRNA sequencing and qPCR methods used in this study provide absolute abundances for a staggering number of taxa, which include a large number of zero values. Because the MAR modeling methods require the normality assumption, they are not appropriate for analyzing data from sequence-based longitudinal microbiome studies. Therefore, we propose an autoregressive zero-inflated mixed effects model (ARZIMM) to address the special features of data instead. Its novelties are threefold. First, we propose to use a zero-inflated Poisson autoregressive model to model the excessive zero abundances and the non-zero abundances separately. Second, the random effects in the proposed model can investigate the underlining dynamic pattern shared within the group. Third, the

164 employment of regularization techniques and network inference in our model enables the
165 identification of the core dynamic patterns. The proposed ARZIMM estimates the strength
166 of interactions between taxa, which is required to estimate the stability properties of a
167 community, and identify key active taxa efficiently by using all of the longitudinal
168 sequencing data. ARZIMM has been implemented in an open-source software package
169 (https://github.com/Hlch1992/ARZIMM), and provides a useful tool for formulating,
170 understanding, and implementing longitudinal microbiome data analysis.

171 In the following Material and Method section, we introduce the ARZIMM framework,
172 discuss the quantification of microbial stability based on the estimated microbial
173 interaction matrix, and investigate the conditions under which there exist a strict-sense
174 stationary distribution. Then in the Result section, we evaluate ARZIMM using extensive
175 simulation studies to show that it outperforms the conventional methods, and apply
176 ARZIMM to the MIME study to illustrate network visualization and inference. In the end,
177 we conclude with Discussion section.

178 ## 2. Material and Method

179 ### 2.1. ARIZMM Model

180 As illustrated in Figure 2, ARZIMM can be considered as a two-part model which
181 comprises a logistic component and an autoregressive component. To address zero
182 inflation, we consider the zero-inflated mixture model because it assumes both sampling
183 zeros (due to the low sequencing depth) and structural zeros (being truly absent) exist in
184 the data. Specifically, the logistic component models the structure zeros of taxa in the
185 samples, and the autoregressive component models the non-structure-zero abundances
186 of the taxa under the assumption that the changes in abundances from time $t-1$ to time

t depend only on the observed abundances at time $t-1$ and other time-independent covariates and the observed abundances before time $t-1$ have no direct effect. Since the goal of ARZIMM is to characterize microbial interactions and community stability during a short period after a strong external perturbation like the antibiotic usage in our MIME study, we assume there are no other time-dependent factors exist to affect the microbial stability.

Notation and Model Specifications

Let Y_{imt} denote the observed absolute abundance of bacterial taxon m $(m = 1,\ldots,M)$ for subject i at time t $(i = 1,2,\ldots,n, t = 1,\ldots,T_i)$, and we model Y_{imt} with a conditional mixture distribution as follow:

$$Y_{imt}|\mathrm{V}_{i(t-1)} \sim \begin{cases} 0 & p_{im} \\ F(y_{imt}|\mathrm{V}_{i(t-1)}; \theta_{itm}, \phi_m) & 1-p_{im} \end{cases} \quad (1)$$

where $\mathrm{V}_{i(t-1)}$ represents all information that is known at time $(t-1)$ for individual i, including the observed absolute abundance $Y_{im(t-1)}$ and later defined coviariates $\boldsymbol{W}_i$ and $\boldsymbol{Z}_i$. The parameter p_{im} represents the probability of the observation Y_{imt} being structural zero and is assumed time independent. Furthermore, F is assumed to be an exponential dispersion family distribution with the canonical parameter θ_{imt} and the dispersion parameter ϕ_m. Both Poisson and negative binomial (NB) distributions can be used as to model absolute abundance. Below we illustrate the detailed modelling using Poisson model.

The mixture probability parameters $\boldsymbol{p}_i = (p_{i1},\ldots, p_{iM})'$ are modeled by the logistic regression:

$$logit(\boldsymbol{p}_i) = \boldsymbol{A}\boldsymbol{W}_i + \boldsymbol{a}_i \quad (2)$$

where $\boldsymbol{W}_i = (1, w_{i1}, \ldots, w_{il})'$ consists of intercept and l time independent covariates for individual i, the parameter $\boldsymbol{A} = (\boldsymbol{A_1}, \ldots, \boldsymbol{A_M})'$ is an $M \times (l+1)$ matrix whose elements A_{mj} is the effect of covariate j on the zero proportion of taxon m. $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{iM})'$ is an $M \times 1$ vector of random intercepts to model the within-subject heterogeneity of being zero for individual i and has the joint multivariate normal distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_a)$.

The canonical parameters for Poisson distribution is $\theta_{imt} = \log E(Y_{imt})$. We introduce the auto-regressive model by relating $\boldsymbol{\theta}_{it} = (\theta_{i1t}, \ldots, \theta_{iMt})'$ to the i^{th} individual's observed log-transformed absolute abundance vector at time $t-1$: $\widetilde{\boldsymbol{Y}}_{i(t-1)} = (\log(Y_{i1(t-1)} + 1), \ldots, \log(Y_{iM(t-1)} + 1))'$ (where the pseudo count 1 is added to avoid the undefined logarithm when the absolute abudance is zero), and $\boldsymbol{Z}_i = (1, Z_{i1}, \ldots, Z_{iq})'$, the intercept and q time-independent covariates of individual i by

$$\boldsymbol{\theta}_{it} | \widetilde{\boldsymbol{Y}}_{i(t-1)} = \boldsymbol{B}\widetilde{\boldsymbol{Y}}_{i(t-1)} + \boldsymbol{C}\boldsymbol{Z}_i + \boldsymbol{\eta}_i \quad (3)$$

where $\boldsymbol{B}$ is an $M \times M$ matrix whose element B_{mj} gives the effect of the abundance of taxon j on the growth rate of taxon m, $\boldsymbol{C}$ is an $M \times (q+1)$ matrix whose element C_{mj} gives the effect of covariate j on taxon m, and $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iM})'$ is time-independent random intercepts. Note that, as an autoregressive model, $\boldsymbol{\eta}_i$ is correlated with the fixed effect $\widetilde{\boldsymbol{Y}}_{i(t-1)}$ and this dependency can be tracked all the way back to the initial observation $\widetilde{\boldsymbol{Y}}_{i0}$. Because the standard random effects model has assumption that the random effects are

independent to the other covariates in the model, in order to derive the random effect type maximum likelihood (ML) estimators, we use the Chamberlain type projections[33] to get around this correlation. Specifically, we project $\boldsymbol{\eta}_i$ onto the time 0 observations $\widetilde{\boldsymbol{Y}}_{i0}$ by:

$$\boldsymbol{\eta}_i = \boldsymbol{\Pi}\widetilde{\boldsymbol{Y}}_{i0} + \boldsymbol{b}_i \quad (4)$$

where $\boldsymbol{\Pi}$ is an $M \times M$ matrix with $\text{diag}(\boldsymbol{\Pi}) = (\pi_1, \dots, \pi_M)'$ and off-diagonal components being zero. The components of $\boldsymbol{\Pi}$ represent how much variation in $\boldsymbol{\eta}_i$ is due to the dependence on subject i's initial value $\widetilde{\boldsymbol{Y}}_{i0}$. $\boldsymbol{b}_i = (b_{i1}, \dots, b_{iM})'$ is an $M \times 1$ vector, representing the independent subject-specific random effect and follows a joint multivariate normal distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_b)$.

In the model, our primary interest is to estimate matrix $\boldsymbol{B}$, which measures the strengths of interactions between taxa. For a microbial community with a given number of species, its stability or dynamics status depends on the changes in the species' population growth rates due to perturbation, which immediately cause the changes in the population growth rates of other species via species-species interactions[34]. Interaction between species can be viewed as a filter that amplifies the variability in species' population growth rates caused by perturbation.

Note that we choose Poisson distribution because of its nice stationary distribution property in the autoregressive model which is crucial for our following stability investigation. To deal with the over-dispersion of microbiome data, we implemented the quasi-Poisson model [35] in the simulation and real data analysis.

Penalized ML Estimation and Variable Selection

To define the joint likelihood of the longitudinal microbial absolute abundance data $\boldsymbol{Y}_{it}$,

we assume that the vector of time independent random effects $\mathbf{c}_i = (\boldsymbol{a}_i', \boldsymbol{b}_i')'$ underlies both the zero and autoregressive generative processes and these random effects account for the within-subject group heterogeneity in the multivariate logistic component and the multivariate autoregressive component. Denote $\boldsymbol{D} = (\boldsymbol{B}, \boldsymbol{C}) = (\boldsymbol{D_1}, \ldots, \boldsymbol{D_M})'$, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_M)'$, and $1_{[\cdot]}$ as the indication function that when $[\cdot]$ meets, $1_{[\cdot]} = 1$, otherwise, $1_{[\cdot]} = 0$. Formally, we have the joint likelihood function as:

$$\mathcal{L}(\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{\Pi}, \boldsymbol{\phi}, \boldsymbol{\sigma})$$

$$= \prod_{i=1}^{n} \int \left\{ \left[\prod_{t=1}^{t_i} \prod_{m=1}^{M} f_y \left(y_{itm} | \theta_{itm}(b_{im}), \phi_m, p_{im}(a_{im}) \right) \right] g\left(\mathbf{c}_i | \boldsymbol{\Sigma}(\boldsymbol{\sigma}) \right) \right\} d\mathbf{c}_i \qquad (5)$$

where f_y is the conditional probability density function and given as

$$f_y(y_{itm}|\theta_{itm}(b_{im}), \phi_m, p_{im}(a_{im})) = [p_{im} + (1 - p_{im}) f(y_{itm} = 0|\theta_{itm}, \phi_m)]^{1_{[y_{itm}=0]}}$$

$$\times [(1 - p_{im}) f(y_{itm}|\theta_{itm}, \phi_m)]^{1_{[y_{itm}\neq 0]}}. \qquad (6)$$

The function $g(\mathbf{c}_i|\boldsymbol{\Sigma}(\boldsymbol{\sigma}))$ is the joint distribution of $\mathbf{c}_i$, and $\boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \begin{bmatrix} \boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ab} & \boldsymbol{\Sigma}_b \end{bmatrix}$ represents the corresponding $2M \times 2M$ covariance matrix, where $\boldsymbol{\sigma}$ accounts for all unique non-zero elements of $\boldsymbol{\Sigma}$. For the model and computational simplicity, we assume $Cov(\boldsymbol{a}_i, \boldsymbol{b}_i) = \Sigma_{ab} = 0$, i.e. $\boldsymbol{a_i}$ and $\boldsymbol{b}_i$ are independent.

Assuming that the true underlying fixed effects $\boldsymbol{A}$ and $\boldsymbol{D}$ are sparse, we advocate a Lasso-type approach, which adds an ℓ_1-penalty for the fixed-effects to the likelihood function. Thus, we consider the following objective function:

$$Q = -2log\mathcal{L} + \sum_{m=1}^{M} [\mu_{1m} ||\boldsymbol{D_m}||_1 + \mu_{2m} ||\boldsymbol{A_m}||_1]. \qquad (7)$$

Maximization of the penalized log-likelihood function corresponding to equation (7) with

respect to $(\boldsymbol{D}, \boldsymbol{A}, \boldsymbol{\Pi}, \boldsymbol{\phi}, \boldsymbol{\sigma})$ is a computationally challenging task. This is mainly because both integrals with respect to the random effects and the zero-inflated structure do not have analytical solutions. Following the conventional methods, we propose to implement a Laplace approximation on the integral of random effects in equation (7) and use the Expectation-Maximization (EM) algorithm to calculate the expectation and compute parameters iteratively, in which the label of zero is treated as "missing data". The tuning parameters are selected using Bayesian information criterion (BIC).

2.2. **Stability Properties**

The existence of a stationary distribution has been investigated for the log-linear Poisson auto-regression model based on the perturbation technique [36]. Here, we prove the existence of a stationary distribution of a zero-inflated Poisson mixed-effect auto-regression model in Theorem 1 utilizing the theory of Markov chains which has been proposed to prove the existence of a stationary distribution of a general class of time series count models [37]. The detailed proof is provided in the Supplementary Material, Section 3.

Theorem 1. Assuming that time-independent parameters $\boldsymbol{\eta}_i$ and $\boldsymbol{p}_i$ are known, if all eigenvalues of matrix $\boldsymbol{B}$ lie inside the unit circle, a strict-sense stationary ergodic process $\{\boldsymbol{Y}_{it}\}_{t\in\mathbf{N}}$ will exist, where $\mathbf{N}$ denotes the set of natural numbers.

With this Theorem, we can first show that for a microbial community, its dynamic process $\{\boldsymbol{Y}_{it}\}_{t\in\mathbf{N}}$ has a stationary distribution by proving that all eigenvalues of matrix $\boldsymbol{B}$ lie inside the unit circle. Then, following Ives et al. [30], we consider the return rate and reactivity as two stability measures based on the variability of the stationary distribution for MAR (1) model. Specifically, return rate depends on the rate at which the perturbed microbial

community approaches the stationary distribution and reactivity, assesses how strongly population-level microbiome abundances are pulled towards the mean of the stationary distribution. Both are bounded by the largest eigenvalue of $\boldsymbol{B}$, denoted by $\max(\boldsymbol{\lambda}_B)$. In general, a smaller $\max(\boldsymbol{\lambda}_B)$ indicates the perturbed microbial community approaches its stationary distribution faster, or a system is less reactive, then the microbial community is more stable. The detailed proof is deferred in the Supplementary Material, Section 3.2.

Based on the theory in Ives et al. [30], for a community with multiple species, the covariance matrix of the stationary distribution depends on the covariance matrix of the process error and the interactions between species captured in the matrix $\boldsymbol{B}$. As illustrated in our Figure 1, when the external perturbation(blue arrow) acts on the community, the ball(microbial community) sitting in a deep bowl in state 2 which represents a relatively stable system, will return to its stationary state faster than the ball sitting in a shallow bowl in state 1 which represents a less stable system. In a stable system, the variance of stationary distribution is only slightly greater than the variance of process error and the variance of species interaction is small. In contrast, in a less stable system, the species interaction will amplify the environmental variance and create large variance in the stationary distribution, therefore the variance of species interaction is large, assuming the process errors are similar in the compared two states. Thus, the difference between the variances of stationary distribution of different communities can be attributed to species interactions. The smaller of the variance of matrix $\boldsymbol{B}$, the more stable of the study microbial community.

3. Results

3.1. Simulation Study

We have conducted extensive simulation studies to evaluate the performance of ARZIMM in both model fitting and variable selection by comparing it with the competing methods: penalized Poisson auto-regression (Poisson), penalized log-normal multivariate auto-regression (MAR), and extended generalized Lotka-Volterra (gLV) equations using Bayesian algorithm (MDSINE) [38]. The brief descriptions of these methods are provided in the Supplementary Material: Section 2.

3.1.1. **Simulation Design**

We generated the longitudinal absolute abundances from zero-inflated Poisson distribution with parameters p_{im} and θ_{imt} for each taxon. Since our focus is on the estimation of the interaction matrix $\boldsymbol{B}$, which depends on the non-zero part, we adopted a simple simulation design for the zero inflation proportions $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{iM})'$. We ignored the individual variations in $\boldsymbol{p}_i$ by dropping the random effect term $\boldsymbol{a}_i$ in equation (2). With model $logit(\boldsymbol{p}_i) = \boldsymbol{A}\boldsymbol{W}_i$ and by controlling the values of $\boldsymbol{W}$ and $\boldsymbol{A}$ respectively, we set the zero inflation proportions $\boldsymbol{p}_i$ for 20 taxa to mimic the observed sparsity in real data as

$$\boldsymbol{p}_i = (0.72, 1.00, 0.96, 0.34, 0.50, 0.56, 0.94, 0.84, 0.98, 1.00, 0.78, 0.68, 0.96, 1.00, 0.38, 0.56, 0.82, 1.00, 0.28, 1.00)' \quad (8)$$

The detailed values of $\boldsymbol{W}$ and $\boldsymbol{A}$ are provided in the Supplementary Material, Section 4. We generated the non-zero absolute abundances from Poisson distribution with their $\boldsymbol{\theta}_{it} = (\theta_{i1t}, \ldots, \theta_{iMt})^T$ defined as $\boldsymbol{\theta}_{it} = \boldsymbol{B}\widetilde{\boldsymbol{Y}}_{i(t-1)} + \boldsymbol{b}_0 + \boldsymbol{b}_i$, where the intercept $\boldsymbol{b}_0$ was set to be the mean log-transformed non-zero absolute abundances of taxa in MIME real data, and the random effects $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, diag(\boldsymbol{\Sigma}_b))$ with $diag(\boldsymbol{\Sigma}_b) \sim 10^{\mathcal{N}(-1.5, 0.5)}$. We assumed that the interaction matrix $\boldsymbol{B}$ was sparse by randomly selecting 5% of its elements to be non-

zero. Three interaction matrices were considered with varied informative absolute effect strengths: high ($B_{jm}^{H}\sim 10^{\mathcal{N}(-0.5.0.5)}$), medium ($B_{jm}^{M}=\sqrt{0.1}\beta_{jm}^{H}$), and low ($B_{jm}^{L}=0.1B_{jm}^{H}$), for the non-zero elements B_{jm}. In addition, we designed four simulation scenarios: Scenario 1 with $diag(\boldsymbol{\Sigma}_b)=\mathbf{0}$ and $\boldsymbol{p}_i=\mathbf{0}$, considered as the benchmark situation where subjects are homogeneous and taxa are all presented; Scenario 2 with $diag(\boldsymbol{\Sigma}_b)\sim 10^{\mathcal{N}(-1.5.0.5)}$ and $\boldsymbol{p}_i=\mathbf{0}$, where subjects are heterogeneous and taxa are all presented; Scenario 3 with $diag(\boldsymbol{\Sigma}_b)=\mathbf{0}$ and $\boldsymbol{p}_i$ as in (8), where subjects are homogeneous and taxa have zero inflated structure; and Scenario 4 with $diag(\boldsymbol{\Sigma}_b)\sim 10^{\mathcal{N}(-1.5.0.5)}$ and $\boldsymbol{p}_i$ as in (8), where subjects are heterogeneous and taxa have zero inflated structure.

In each scenario, we generated 500 independent repetitions for $n=20\text{ or }50$ subjects, $T=10\text{ or }20$ time points, and $M=20$ taxa for each sample to evaluate the performance of ARZIMM.

3.1.2. **Simulation Results**

We first compared the model fittings of ARZIMM, Poisson, and MAR methods using mean normalized squared error score (MNSES), as suggested in the prior studies [39-42]. MNSES is defined as $\frac{1}{n\times T\times M}(\frac{y_{imt}-\hat{y}_{imt}}{\hat{\sigma}_{y_{imt}}})^2$ with $\hat{y}_{imt}$ being the estimated y_{imt} and $\hat{\sigma}_{y_{imt}}$ being the estimated standard error of y_{imt}. The closer the MNSES is to 1, the better model fitting the method has. Since MDSINE only provides the estimates of interactions among species without their variance estimates, it was excluded from this comparison. Table 1 and Supplementary Table S1 summarize the median and interquartile range (IQR) of MNSES over 500 replications for these three methods. Overall, the medians of MNSES for ARZIMM are all around the expected value of 1 in various settings across four scenarios, which indicates the good fitness and robustness of ARZIMM in dealing with

359 excess zeros and the correlation among repeated measures at the same time, as well as
360 its satisfying estimation accuracy on the microbial interaction parameters. However, the
361 other two methods: Poisson and MAR, both exhibit inferior performance. The Poisson
362 model is only competent in Scenario 1, when subjects are homogeneous and no excess
363 zeros are present. In Scenarios 2-4, when any factor, excess zero or subject
364 heterogeneity, presents, the predicted values based on the Poisson model deviate greatly
365 from the observed values. Comparing the considered two factors, Poisson model is more
366 sensitive to the subject heterogeneity and presents larger deviations with it. Due to the
367 invalid normality assumption and lack of consideration of the correlation among the
368 longitudinal measurements, the MAR model exhibits the worst performance among three
369 methods with enormous deviation especially in Scenarios 3 and 4, which confirms the
370 inappropriateness of using conventional statistical methods which require the normality
371 assumption to analyze the microbiome data.
372 Next, we evaluated the variable selection performance for ARZIMM, Poisson, MAR, and
373 MDSINE in terms of true positive rate (TPR; mathematically equals to the power) and
374 false positive rate (FPR; mathematically equals to the type I error). Specifically, TPR
375 quantifies the probability of a significant interaction identified by one method given that
376 the interaction effect is truly nonzero; and FPR quantifies the probability of a significant
377 interaction identified by one method given that the interaction effect is truly zero. The
378 simulation results for 50 subjects with 20 time points are summarized in Figure 3 and all
379 the other simulation results with different subject numbers and time points are deferred
380 to Supplementary Figure S1, because they have a similar pattern as seen in Figure 3.
381 Figure 3 shows that the FPRs of ARZIMM are all at or below the nominal level (5%) across

different simulation regimes and effect sizes, and its TPR estimates exhibit a sensible and consistent pattern as they increase as the interaction effect gets stronger across four scenarios. As expected, the FPR and TRP estimates of Poisson and ARZIMM models are coincident under Scenario 1, because when subjects are homogeneous and taxa don't have excess zeros, ARZIMM model is reduced to Poisson model. However, in Scenarios 2-4, because simple Poisson model fails to take care of the excess zeros or subject heterogeneity, it suffers from the inflated false positives, while ARZIMM does not. For the other two methods, both MAR and MDSINE perform poorly on controlling false positive rates for all simulation scenarios, because MAR fails to fit the skewed and highly sparse microbiome data, while MDSINE captures the interactions based on the averaged effect over subjects in a group but completely ignores the randomness at the subject level process which is the essential characteristic of any biological system. In summary, ARZIMM outperforms the other competitors in handling the excess zeros and subject heterogeneity well with controlled FPR and satisfactory TPR.

To further investigate the performance of informative interaction selection, we calculate Matthew correlation coefficient (MCC), defined as $\frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$, and F-score, defined as $\frac{TP}{TP+(FP+FN)/2}$, where TP gives the number of selected interactions being true positive, FP gives the number of selected interactions being false positive, TN gives the number of unselected interactions being true negative, and FN gives the number of selected interactions being false negative. MCC ranges from -1 to 1, where value 1 indicates perfect agreement between truth and selection, value -1 indicates perfect disagreement, and value 0 indicates that the selection is random with respect to the truth. F-score ranges from 0 to 1, where value 1 indicates that there are neither false negatives

nor false positives and value 0 only indicates no true positives are reported. As expected, MCC and F score are comparable to each other and increase as effect size increases (Supplementary Figure S2). This consistent pattern is observed across four scenarios for ARZIMM but not for Poisson nor MAR models. Similar to TPR and FPR estimates, the MCC and F score values of Poisson and ARZIMM models are coincident under Scenario 1. However, in other situations, both Poisson and MAR perform poorly with low MCC and F score values.

As for the computational cost, ARIZMM took about 2.4 hour to complete the estimation and bootstrap inference for a simulated dataset with 50 subjects, 20 timepoints, and 20 taxa.

3.2 Real Data Application

We applied ARZIMM methods to the MIME study. The MIME study is an ongoing randomized trial on 80 healthy volunteers with one control group (ctrl) and two antibiotic groups (amoxicillin, amx, and azithromycin, azm); antibiotics are provided for a 1-week period at the start of the trial. The main microbiome research goal of the MIME study is to evaluate the effects of antibiotics on microbial profiles at both the community and taxonomical levels. With ARIZMM, we propose a different perspective to evaluate the effect of antibiotics through the investigation of microbial interaction and community stability across groups. Because the clinical trial is still ongoing and only partial data are available, the following data analysis is done on a subset of MIME data including only 11 subjects who were randomized to two groups: 4 ctrls and 7 azms. The main purpose of this analysis is to illustrate how to use ARIZMM, not for the scientific conclusion. For each subject, we collected two baseline microbiome samples, three samples during the course

of antibiotics, and five post-antibiotic samples. The gut microbiota of these individuals were profiled using 16S rRNA gene targeted sequencing on the Illumina MiSeq platform. To obtain the microbial absolute abundances, we multiplied the relative abundances of OTUs by the sample density 1.1g/cm^3 and the number of universal 16S rRNA per gram measured using qPCR [43]. In our analysis, samples that collected before treatment in both antibiotic groups were excluded. The abundances of taxa were agglomerated at the genus level and taxa were further filtered if (1) the average relative abundances over all samples are less than 0.1%, and (2) the taxa are presented in less than 5 samples within each group.

First, Figure 4A shows a comparison of the relative abundance (top panel) and the absolute abundances determined by quantitative sequencing (bottom panel) of the dominant bacterial genera in 99 fecal samples from 11 subjects (blocks) across seven to nine time points (shown from left to right within each block) of this preliminary dataset. It is evident that the relative abundance and absolute abundance data present different information about the microbial profiles, and that the total bacterial load changes over time for each subject (i.e., within each block). Thus it is essential to study the microbial interactions using the absolute abundance data.

Then, we evaluate the model fitting of the log-normal distribution (used in MAR(1)) and zero-inflated over-dispersed Poisson distribution (used in ARZIMM) on the available subset of MIME data using chi-square goodness of fit test at 5% significance level taxon by taxon. Out of 45 taxa in the control group, 1 and 44 of their absolute abundances were fitted well (p>0.05) by log-normal distribution and zero-inflated over-dispersed Poisson distribution respectively. The log-normal distribution fails to fit the data well when

451 microbial taxa's absolute abundance data are left-skewed and sparse (two examples are
452 illustrated in Figure 4B).

453 Next we demonstrate how to conduct inference for microbial interactions and community
454 stability with ARZIMM on MIME data. First, we fit ARZIMM to ctrl and azm groups
455 separately, adjusting for age, gender, and BMI, to get their estimated interaction matrix
456 $\widehat{\boldsymbol{B}}$s. Table 2 reports the characteristics of microbial interaction matrix estimates $\widehat{\boldsymbol{B}}$s.
457 Defining the interaction effect as informative if its $\hat{B}_{mj}$'s 95% bootstrap confidence interval
458 (based on 100 bootstrap samples) does not contain zero, we identified 125 and 105
459 informative interactions, respectively, in azm and ctrl groups. Their interaction effects are
460 illustrated using networks in Figure 5. With more informative interactions, the azm groups
461 have bigger and more complex networks than the ctrl group (first row of Figure 5), while
462 the control group has more large estimated interaction effects than those in azm group
463 as showed in Table 2 and the last three rows of Figure 5. This observation indicates that
464 the antibiotic treatment reduce the strength of the interactions among the taxa and create
465 more variations with more weak interactions among taxa, thus reduce its stability. In the
466 last row of Table 2, based on our stability theory we report the stability properties of the
467 studied microbial communities. The ctrl group has the lower estimates of maximum
468 eigenvalue squared 0.11 comparing to the azm group's maximum eigenvalue squared
469 0.32, which indicates that the control microbial community is more stable than the
470 antibiotic communities.

471 Figure 6 provides additional information on the network feature comparison between ctrl
472 and azm groups. Figure6A displays the distribution of the positive and negative
473 informative interaction estimates separately. The ratios between the numbers of positive

and negative interactions are both around 1:1 in two groups. Figure 6B presents the frequency distribution of vertex degree of all the taxa in each group and they are all skew to the right. In the figure, a vertex represents a taxon in a community and its vertex degree is the number of informative interaction effect it has with the other taxa. By defining average neighbor degree as the average number of a given taxon's neighbor vertices' degrees, Figure 6C shows that the average neighbor degree is negatively correlated with the vertex degree in azm antibiotic treated group, but not in the control group. This indicates that there may be a group of taxa interacting with each other actively in the antibiotic group. It would be interesting to identify such sub-community with additional effort.

Discussion

In this paper, we propose ARZIMM, an analytic platform which estimates the microbial interactions and community stability using longitudinal microbiome data. ARZIMM tackles the zero-inflated absolute abundance with a mixture distribution of zero and exponential dispersion distribution family, and enhances statistical efficiency by utilizing a random-effects term to account for the correlations among repeated measurements.

It is well-known that microbial correlations calculated from relative abundances are distorted by the compositional nature of microbiome data, and are insufficient in tracking microbial dynamics[44]. We advocate to investigate the microbial correlations using longitudinal absolute abundances which can be determined by combining gene amplicon sequencing with auxiliary total DNA quantitation data. qPCR is one of the most commonly used strategies to quantify total DNA[45] and has been implemented in various statistical analyses[46, 47]. Other alternative methods to quantify the absolute abundances include

the combination of the sequencing approach (16S rRNA gene) with robust single-cell enumeration technologies (flow cytometry)[48] and the usage of synthetic chimeric DNA spikes[49].

Plenty of zero-inflated mixed effects models have been recently proposed to handle the excess zeros in microbiome abundance data such as zero-inflated Poisson, negative binomial and quasi-Poisson models[50, 51]. However, none of the existing methods estimates the microbial interactions and community stability. To fill this gap, we extended a zero-inflated Poisson model with auto-regression and random effects modeling, which plays crucial role in efficiently handling the individual heterogeneity and enable the investigation of microbial interactions.

We investigated two community stability measurements derived from ARZIMM: the return rate and reactivity, to further understand ecological dynamics. The estimated interaction matrix $\boldsymbol{B}$ from the ARZIMM model serves the basis to calculate the largest eigenvalue of $\boldsymbol{B}$: $\max(\lambda_{\boldsymbol{B}})$, which determines the return rate of the mean of the transition distribution from the departure to the mean of the stationary distribution. We proposed to measure the reactivity of a microbial community by the expected change of the stationary distribution's mean in distance from one time point to the next time point. In ARZIMM, higher reactivity coincides with larger eigenvalues of $\boldsymbol{B}$, thus governed again by $\max(\lambda_{\boldsymbol{B}})$. Other measures of community stability, such as variance of the stationary distribution [30], warrant further investigations.

It is worth noting that by utilizing the ARZIMM model framework, the time-dependent perturbation (for instance, diet) can also be assessed flexibly in both the autoregressive part and the logistic part in the model. However, the stability based on the microbial

interactions has to be interpreted with caution, since the mean of stationary distribution changes along with the time-dependent covariates.

We have demonstrated that ARZIMM outperforms the competing methods and exhibits its feasibility for examining microbial interactions and stability based on longitudinal microbial data. We applied our method to a real human microbiome study of antibiotic treatment and elucidated the microbial interaction network of bacteria from antibiotic and non-antibiotic groups separately. The application of ARZIMM to temporal microbiome data shows great promise. Still, the development of accurate predictive models will require further developments. For example, the method used here to infer microbial interactions may be expanded by adding functional information as well as phylogenetic information. Although this method is primarily developed for the gut microbiota, it may be potentially applied to longitudinal data from any ecological systems. Since interactions between members of microbial communities are primary driving forces for the long-term stability[52], the corresponding stability properties will provide useful principles for community dynamics.

Note that the proposed ARIZMM assumes the probability of observing a zero count for a taxon is constant over time. The reason is two-fold. 1) One major goal of ARIZMM is to derive the inference on the stability of the microbial community over a certain period. With the constant probability of observing a zero count assumption, the stability inference will solely depend on the estimation of the taxon-by-taxon interaction matrix *B*. Otherwise, a stationary distribution won't exit. 2) Using the MIME data, we estimated the proportions of zeros (denoted as q_{mt}) for all taxa by group at all time points, then calculated the mean($\bar{\hat{q}}_m$) and standard deviation ($SD_{\hat{q}_m}$) over all the time points and the coefficient of

variation ($CV_m = SD_{\hat{q}_m}/\bar{\hat{q}}_m, m = 1, \dots, M$) to evaluate their temporal variations. The median of CV_m over all taxa in the control, Amoxicillian and Azithromycin groups are 0.16, 0.12, and 0.34 respectively. This results reveal two observations: 1) the temporal variations of q_{mt} in most taxa are relative weak; and 2) the temporal variation of the proportions of zero is heterogeneous and there may be no one perfect model fitting all the taxa well. Thus, we believe our assumption that p_m is constant over time is valid and pragmatic. To further check the robustness of our proposed model, we conducted additional simulation by introducing extra randomness when we generate the probability of observing a zero count across the time points, while analyze the data using our proposed model. Our results show that the moderate temporal variation in probability of zero count does not affect ARIZMM's performance much in capturing the informative interactions by estimating *B* when the absolute effect strengths of interaction matrices is high(FDR<0.05) or medium (FDR <0.15). The detailed simulation design and results are reported in the Supplementary Material Section 4.2 and Figure S3.

The proposed method, ARZIMM has a few limitations and future works are needed to improve it. ARZIMM adopts a simple correlation structure that the random effects in the multivariate logistic component and the multivariate autoregressive component $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are assumed independent. We took this parsimonious model based on our experience[53, 54] in modeling the longitudinal microbiome data to ease the computational burden. The more general random effects structure with cross-part correlations can provide more robust modeling, however, can suffer from model convergence as well. Further investigation is warranted.

CONFLICT OF INTEREST

Author Linchen He is employed by Novartis Pharmaceuticals Corporation. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHORS CONTRIBUTIONS

LH and HL developed the methodological ideas. LH implemented the methods, performed the simulations and real data analysis, and developed the software package. CW and JH contributed to the simulation design and real data analysis. ZG, EF, SH, MJB contributed to the acquisition of utilized real microbiome data. MJB provided biological insights and interpretation of the real data analysis. LH and HL wrote the manuscript. All authors read, edited, and approved the final manuscript.

FUNDING

This work was supported in part by National Institutes of Health grants R01DK110014, P20CA252728, and U01AI22285.

DATA AVAILABILITY STATEMENT

Software for implementing the method described in this paper is publicly available on GitHub at https: https://github.com/Hlch1992/ARZIMM.

REFERENCES

1. Faith, J.J., et al., *The long-term stability of the human gut microbiota.* Science, 2013. **341**(6141): p. 1237439.
2. Jackson, M.A., et al., *Signatures of early frailty in the gut microbiota.* Genome Med, 2016. **8**(1): p. 8.
3. Claesson, M.J., et al., *Gut microbiota composition correlates with diet and health in the elderly.* Nature, 2012. **488**(7410): p. 178-84.

4. Martinez, C., et al., *Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission.* The American journal of gastroenterology, 2008. **103**(3): p. 643.
5. Zuo, T. and S.C. Ng, *The gut microbiota in the pathogenesis and therapeutics of inflammatory bowel disease.* Frontiers in microbiology, 2018. **9**.
6. Scanlan, P.D., et al., *Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis.* Environmental microbiology, 2008. **10**(3): p. 789-798.
7. Uronis, J.M., et al., *Modulation of the intestinal microbiota alters colitis-associated colorectal cancer susceptibility.* PloS one, 2009. **4**(6): p. 6026.
8. Carroll, I.M., et al., *Alterations in composition and diversity of the intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome.* Neurogastroenterology & Motility, 2012. **24**(6): p. 521-248.
9. Maukonen, J., et al., *Prevalence and temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria.* Journal of Medical Microbiology, 2006. **55**(5): p. 625-633.
10. Woo, P.C., et al., *Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories.* Clin Microbiol Infect, 2008. **14**(10): p. 908-34.
11. Gilbert, J.A., et al., *Current understanding of the human microbiome.* Nat Med, 2018. **24**(4): p. 392-400.
12. Shankar, J., *Insights into study design and statistical analyses in translational microbiome studies.* Ann Transl Med, 2017. **5**(12): p. 249.
13. Knight, R., et al., *Best practices for analysing microbiomes.* Nat Rev Microbiol, 2018. **16**(7): p. 410-422.
14. Nadkarni, M.A., et al., *Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set.* Microbiology, 2002. **148**(Pt 1): p. 257-266.
15. Kim, J., J. Lim, and C. Lee, *Quantitative real-time PCR approaches for microbial community studies in wastewater treatment systems: applications and considerations.* Biotechnol Adv, 2013. **31**(8): p. 1358-73.
16. Ott, S.J., et al., *Quantification of intestinal bacterial populations by real-time PCR with a universal primer set and minor groove binder probes: a global approach to the enteric flora.* J Clin Microbiol, 2004. **42**(6): p. 2566-72.
17. Russell, S. and P. Norvig, *Artificial intelligence: a modern approach.* 2002.
18. McGeachie, M.J., et al., *Longitudinal prediction of the infant gut microbiome with dynamic bayesian networks.* Scientific reports, 2016. **6**(1): p. 1-11.
19. Lugo-Martinez, J., et al., *Dynamic interaction network inference from longitudinal microbiome data.* Microbiome, 2019. **7**(1): p. 1-14.
20. Faust, K. and J. Raes, *Microbial interactions: from networks to models.* Nature Reviews Microbiology, 2012. **10**(8): p. 538-550.
21. Gerber, G.K., *The dynamic microbiome.* FEBS letters, 2014. **588**(22): p. 4131-4139.
22. Mounier, J., et al., *Assessment of the microbial diversity at the surface of Livarot cheese using culture-dependent and independent approaches.* International journal of food microbiology, 2009. **133**(1-2): p. 31-37.
23. Marino, S., et al., *Mathematical modeling of primary succession of murine intestinal microbiota.* Proceedings of the National Academy of Sciences, 2014. **111**(1): p. 439-444.
24. Dam, P., et al., *Dynamic models of the complex microbial metapopulation of lake mendota.* NPJ systems biology and applications, 2016. **2**(1): p. 1-7.

25. de Vos, M.G.J., et al., *Interaction networks, ecological stability, and collective antibiotic tolerance in polymicrobial infections.* Proceedings of the National Academy of Sciences, 2017. **114**(40): p. 10666-10671.
26. Venturelli, O.S., et al., *Deciphering microbial interactions in synthetic human gut microbiome communities.* Molecular systems biology, 2018. **14**(6): p. e8157.
27. Fisher, C.K. and P. Mehta, *Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression.* PloS one, 2014. **9**(7): p. e102451.
28. Shaw, G.T.-W., Y.-Y. Pao, and D. Wang, *MetaMIS: a metagenomic microbial interaction simulator based on microbial community profiles.* BMC bioinformatics, 2016. **17**(1): p. 1-12.
29. Bucci, V., et al., *MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses.* Genome biology, 2016. **17**(1): p. 1-17.
30. Ives, A.R., et al., *Estimating community stability and ecological interactions from time-series data.* Ecological monographs, 2003. **73**(2): p. 301-30.
31. Carpenter, S.R., et al., *Early warnings of regime shifts: a whole-ecosystem experiment.* 2011. **332**(6033): p. 1079-1082.
32. Shade, A., et al., *A meta-analysis of changes in bacterial and archaeal communities with time.* 2013. **7**(8): p. 1493-1506.
33. Chamberlain, G., *Multivariate regression models for panel data.* Journal of econometrics, 1982. **18**(1): p. 5-46.
34. Ives, A.R., J.L. Klug, and K. Gross, *Stability and species richness in complex communities.* 2000. **3**(5): p. 399-411.
35. Ver Hoef, J.M. and P.L. Boveng, *Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data?* Ecology, 2007. **88**(11): p. 2766-2772.
36. Fokianos, K. and D. Tjøstheim, *Log-linear Poisson autoregression.* Journal of Multivariate Analysis, 2011. **102**(3): p. 563-578.
37. Douc, R., P. Doukhan, and E. Moulines, *Ergodicity of observation-driven time series models and consistency of the maximum likelihood estimator.* Stochastic Processes and their Applications, 2013. **123**(7): p. 2620-2647.
38. Bucci, V., et al., *MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses.* Genome biology, 2016. **17**(1): p. 121.
39. Carroll, S.S. and N. Cressie, *Spatial modeling of snow water equivalent using covariances estimated from spatial and geomorphic attributes.* Journal of Hydrology, 1997. **190**(1-2): p. 42-59.
40. Liesenfeld, R., I. Nolte, and W. Pohlmeier, *Modelling financial transaction price movements: a dynamic integer count data model.* Empirical Economics, 2006. **30**(4): p. 795-825.
41. Tkacz, A., M. Hortala, and P.S. Poole, *Absolute quantitation of microbiota abundance in environmental samples.* Microbiome, 2018. **6**(1): p. 110.
42. Czado, C., T. Gneiting, and L. Held, *Predictive model assessment for count data.* Biometrics, 2009. **65**(4): p. 1254-1261.
43. Stein, R.R., et al., *Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota.* PLoS computational biology, 2013. **9**(12): p. 1003388.
44. Gloor, G.B., et al., *Microbiome Datasets Are Compositional: And This Is Not Optional.* Frontiers in Microbiology, 2017. **8**: p. 2224.

45. Dannemiller, K.C., et al., *Combining real-time PCR and next-generation DNA sequencing to provide quantitative comparisons of fungal aerosol populations.* Atmospheric environment, 2014. **84**: p. 113-121.
46. Bucci, V., et al., *MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses.* 2016. **17**(1): p. 1-17.
47. Stein, R.R., et al., *Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota.* 2013. **9**(12): p. e1003388.
48. Props, R., et al., *Absolute quantification of microbial taxon abundances.* The ISME journal, 2017. **11**(2): p. 584-587.
49. Tkacz, A., M. Hortala, and P.S. Poole, *Absolute quantitation of microbiota abundance in environmental samples.* Microbiome, 2018. **6**(1): p. 1-13.
50. Xia, Y., J. Sun, and D.-G. Chen, *Statistical analysis of microbiome data with R*. Vol. 847. 2018: Springer.
51. Zhang, X., et al., *Negative binomial mixed models for analyzing longitudinal microbiome data.* 2018. **9**: p. 1683.
52. Ratzke, C., J. Barrere, and J. Gore, *Strength of species interactions determines biodiversity and stability in microbial communities.* bioRxiv, 2019: p. 671008.
53. Hu, J., et al., *Joint modeling of zero-inflated longitudinal proportions and time-to-event data with application to a gut microbiome study.* bioRxiv, 2020: p. 2020.12.10.419945.
54. Wang, C., et al., *Microbial trend analysis for common dynamic trend, group comparison and classification in longitudinal microbiome study.* bioRxiv, 2020: p. 2020.01.30.926824.

Table 1: Simulation results for all settings under scenario 1 and 4 Poisson refers to the penalized Poisson autoregression model and MAR refers to penalized log-normal multivariate autoregression model. The reported value is median (IQR) of mean normalized squared error score (MNSES) calculated over 500 simulations for each setting. N refers to the number of subjects, and T refers to the number of time points. Scenarios 2 and 3 are deferred to Supplementary Material.

Methods		ARZIMM			Poisson			MAR		
Effect size		High	Median	Low	High	Median	Low	High	Median	Low
Scenario 1										
N	T	Median (IQR)								
20	10	0.98	0.98	0.98	1.00	0.99	1.00	47	50	52
		(0.97-0.99)	(0.97-0.99)	(0.97-0.99)	(0.99-1.01)	(0.984-1.00)	(0.99-1.00)	(33-77)	(35-80)	(37-80)
50	20	0.99	0.99	0.99	1.00	1.00	1.00	123	115	114
		(0.99-1.00)	(0.99-1.00)	(0.99-1.00)	(1.00-1.01)	(1.00-1.00)	(1.00-1.00)	(86-192)	(78-187)	(80-177)
Scenario 4										
N	T	Median (IQR)								
20	10	0.95	0.92	0.91	30.59	18.87	18.46	30071.82	29390	22435
		(0.87-2.30)	(0.86-1.10)	(0.86-1.02)	(21.90-41.51)	(15.26-21.95)	(14.77-21.80)	(8984-133153)	(13929-77371)	(10251-50171)
50	20	1.09	0.92	0.85	40.31	31.08	30.30	211141	110656	93579
		(1.05-1.20)	(0.90-0.93)	(0.85-0.86)	(36.80-43.63)	(30.25-31.76)	(29.65-30.95)	(118860-473551)	(80809-202068)	(66942-167227)

Table 2: The characteristics of networks.

Group description	Azithromycin	Control
Sample size	7	4
Number of time points	9	9
Number of taxa	49	45
Number of informative interactions	125	105
Number of $\lvert\hat{B}_{mj}\rvert < 0.1$	73	45
Number of $0.1 \leq \lvert\hat{B}_{mj}\rvert < 0.25$	30	29
Number of $0.25 \leq \lvert\hat{B}_{mj}\rvert < 0.5$	17	14
Number of $\lvert\hat{B}_{mj}\rvert \geq 0.5$	5	17
Informative interaction percentage (%)	5.21	5.19
Maximum eigenvalue squared	0.32	0.11

FIGURE LEGENDS

Figure 1: Schematic of the evolution of microbial community states in response to external perturbation. External perturbation (blue arrows) can affect microbial community composition (shown in a pie chart), defined as a community state. For each state, the ball-in-basin diagram portrays stability measured by the variance in the stationary distribution of the location of the ball. White arrows indicate the reaction of microbial community to the perturbation.

Figure 2: Graphical representation of ARZIMM model and analytic techniques.

Figure 3: Simulation results of variable selection performance. Poisson refers to the penalized Poisson auto-regression model and MAR refers to penalized log-normal multivariate auto-regression model. MDSINE refers to the method with extended generalized Lotka-Volterra (gLV) equations using a Bayesian algorithm. Mean (and 95% confidence interval) of false positive and true positive rates are reported for 500 simulations with 50 subjects and 20 time points in four scenario: (A) no zero-inflated structure and no heterogeneity, (B) heterogeneity but no zero-inflated structure, (C) zero-inflated structure but no heterogeneity, and (D) both zero-inflated structure and heterogeneity.

Figure 4: MIME study microbiome data. (A) Difference between relative abundances (top panel) and absolute abundances based on qPCR (bottom panel) of dominant genera in XX fecal

samples obtained from 21 subjects (block) at 7–9 time points (x-axis) each. (B) Distribution of absolute abundances of two representative genera from the MIME study, shown in the left and right panels, respectively. For each genus, the absolute abundance is fitted with a log-normal distribution (red line) or a two-part distribution: a zero part (dark green line shown in right panel) and a non-zero part fitted with an over-dispersion Poisson distribution (blue line).

Figure 5: Interaction network. Estimated interaction network for: (A) azithromycin (azm), and (B) control groups, displaying (1) all selected interactions, (2) interactions with $|\hat{B}_{mj}| \geq 0.1$, (3) interactions with $|\hat{B}_{mj}| \geq 0.25$, and (4) interactions with $|\hat{B}_{mj}| \geq 0.5$. Each node represents a taxon at the genus level, the size of which shows the degree of that taxa and the color of which shows the phylogenetic Order level for each taxon. Each edge with arrow represents an interaction effect, the width of which represents the absolute effect size on a $\log_{10}$ scale, with the color showing a positive (orange) or negative (blue) effect.

Figure 6: Characteristics of estimated interactions. (A) The effect size of estimated informative interactions, wherein the x-axis represents the $\log_{10}$ scaled absolute effect size, the y-axis represents the count of informative interactions, and the colors represent the positive or negative effects. (B) Histogram of vertex degree, wherein given a vertex, vertex degree is defined as the counts of edges upon the vertex. (C) The average neighbor degree (y-axis) versus vertex degree on a log-log scale (x-axis). The average neighbor degree is the average number of a given taxon's neighbor vertices' degrees. Dotted lines represent 95% confidence limits.

ARZIMM: An autoregressive zero-inflated mixed model

The zero part:
logit model

$$logit(p_i) \quad AW_i + a_i$$

The non-zero part:
a first order autoregressive mixed model

$$\theta_{i(t-1)} \quad B\ddot{Y}_{i(t-1)} + CZ_i + \eta_i$$

$$\eta_i = \Pi\ddot{Y}_{i0} + b_i$$

The random effects (a_i, b_i) investigate the underlining dynamic pattern shared within the group.

Model the excessive zero abundances

Joint likelihood function:

$$\mathcal{L}(D, A, \Pi, \phi, \sigma)$$

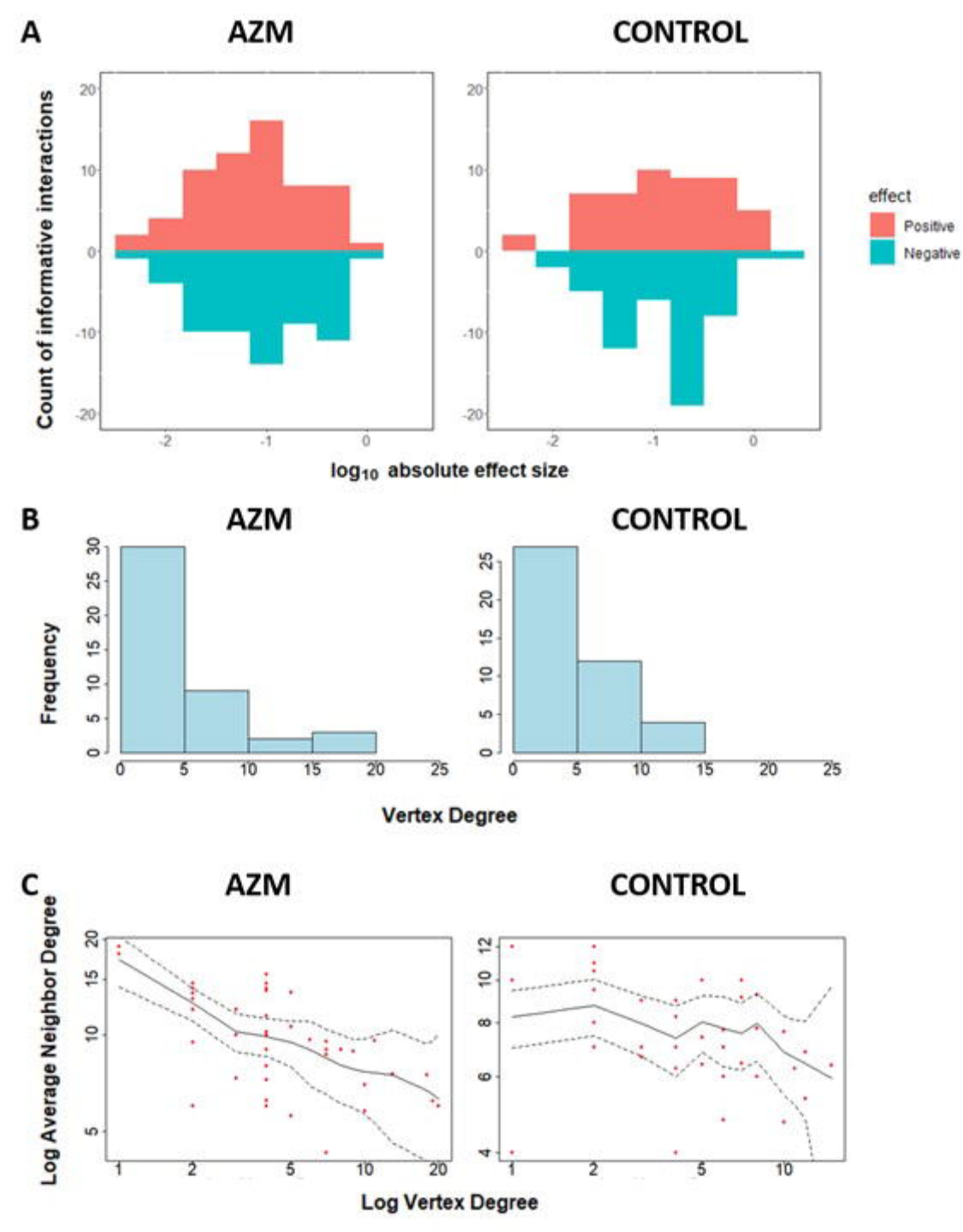
Chamberlain type projections to get around the correlation between random effects η_i

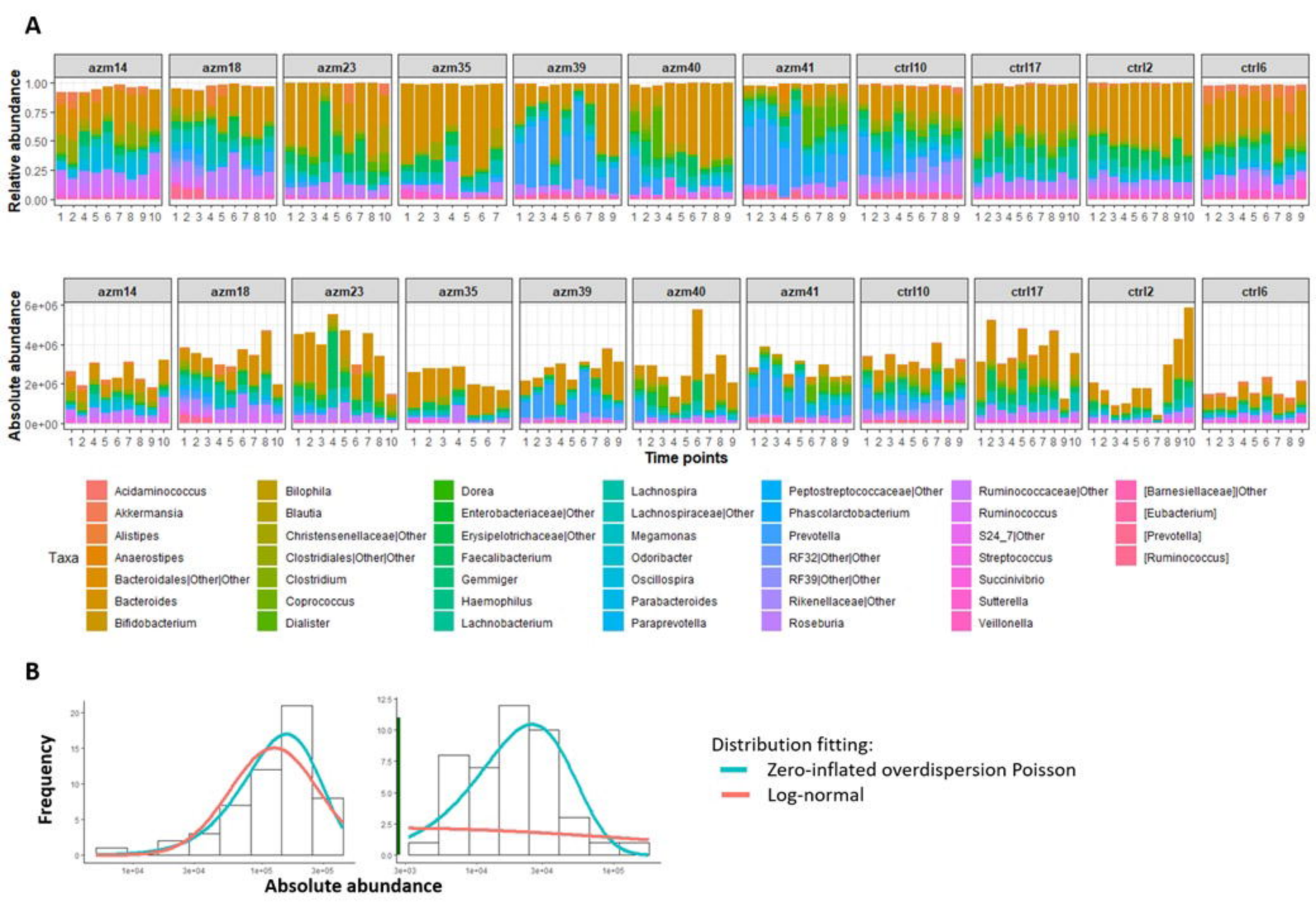
Penalized ML Estimation and Variable Selection

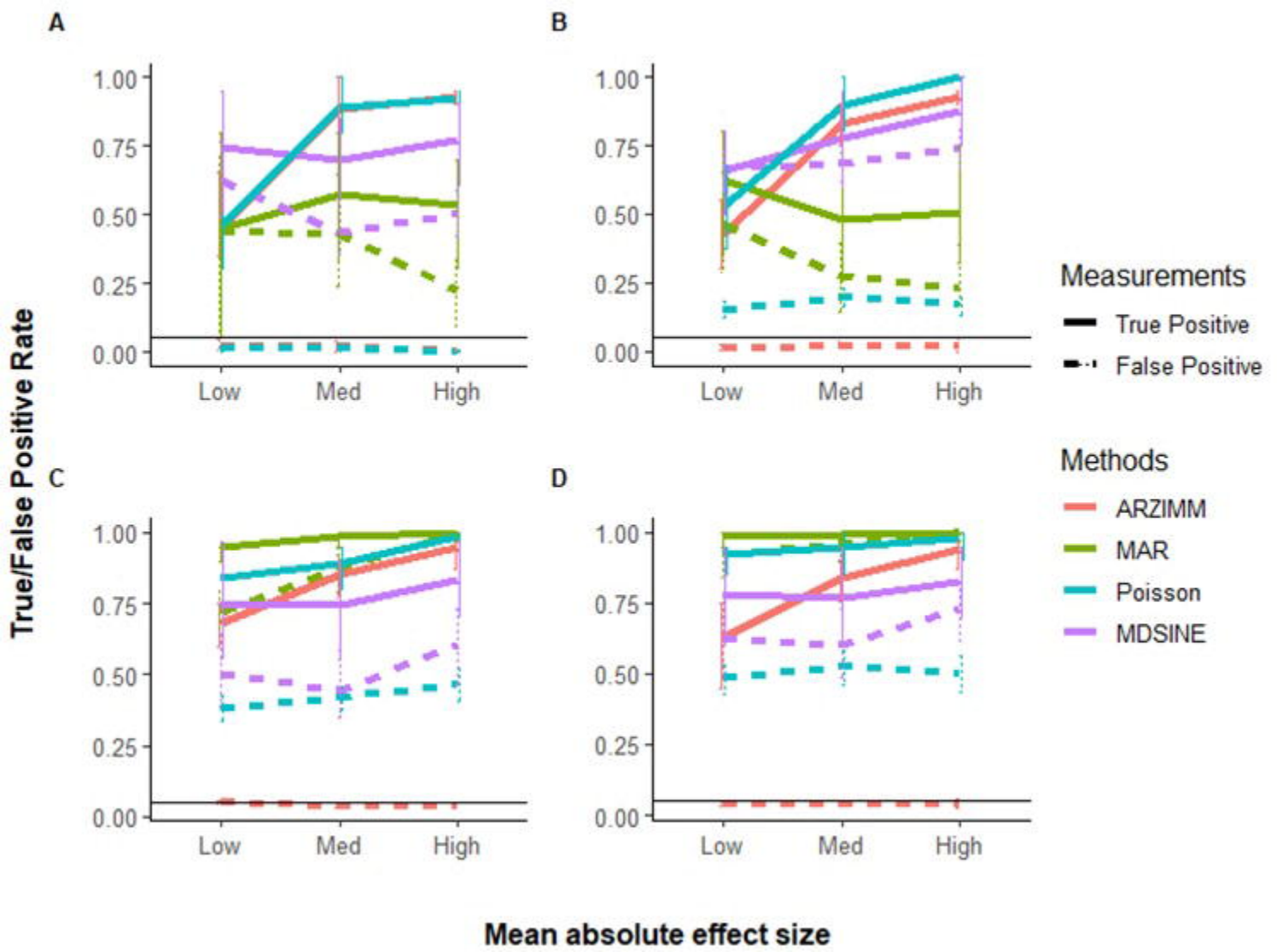
$$Q = -2log\mathcal{L} + \sum_{m=1}^{M}{}' [\mu_{1m}||D_m||_1 + \mu_{2m}||A_m||_1].$$

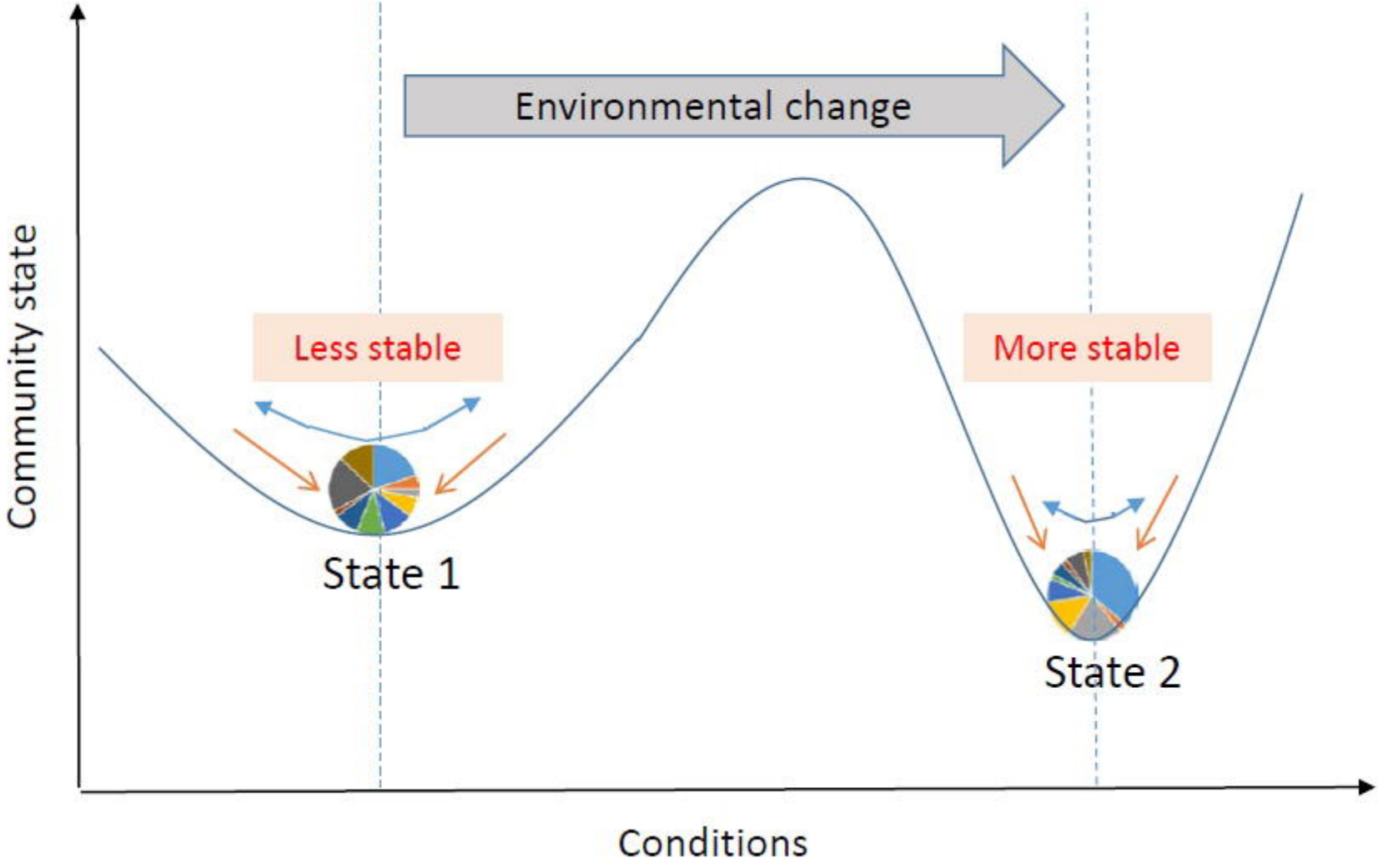
Computation Procedures:

- Lasso-type regularization approach
- Laplace approximation on random effect
- EM algorithm to calculate the expectation and compute parameters iteratively
- BIC for varialbe selection

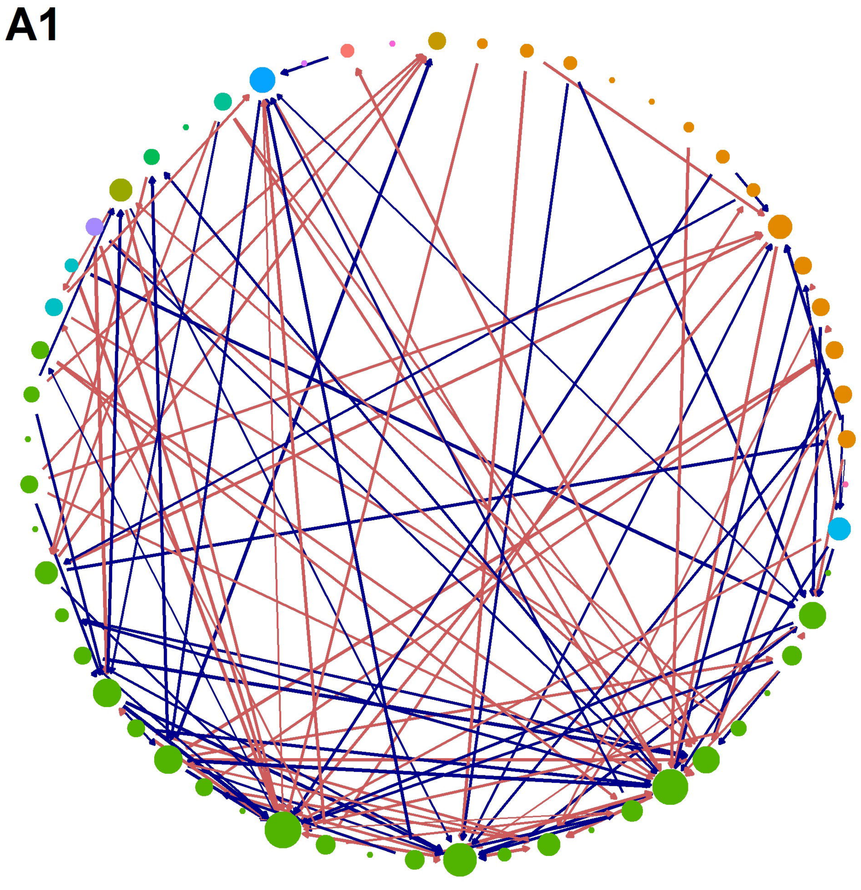

A
AZM
CONTROL
Count of informative interactions
log10 absolute effect size
effect
Positive
Negative
B
AZM
CONTROL
Frequency
Vertex Degree
C
AZM
CONTROL
Log Average Neighbor Degree
Log Vertex Degree


A
azm14
azm18
azm23
azm35
azm39
azm40
azm41
ctrl10
ctrl17
ctrl2
ctrl6
Relative abundance
Absolute abundance
Time points
Taxa
Acidaminococcus
Akkermansia
Alistipes
Anaerostipes
Bacteroidales|Other|Other
Bacteroides
Bifidobacterium
Bilophila
Blautia
Christensenellaceae|Other
Clostridiales|Other|Other
Clostridium
Coprococcus
Dialister
Dorea
Enterobacteriaceae|Other
Erysipelotrichaceae|Other
Faecalibacterium
Gemmiger
Haemophilus
Lachnobacterium
Lachnospira
Lachnospiraceae|Other
Megamonas
Odoribacter
Oscillospira
Parabacteroides
Paraprevotella
Peptostreptococcaceae|Other
Phascolarctobacterium
Prevotella
RF32|Other|Other
RF39|Other|Other
Rikenellaceae|Other
Roseburia
Ruminococcaceae|Other
Ruminococcus
S24_7|Other
Streptococcus
Succinivibrio
Sutterella
Veillonella
[Barnesiellaceae]|Other
[Eubacterium]
[Prevotella]
[Ruminococcus]
B
Frequency
Absolute abundance
Distribution fitting:
Zero-inflated overdispersion Poisson
Log-normal

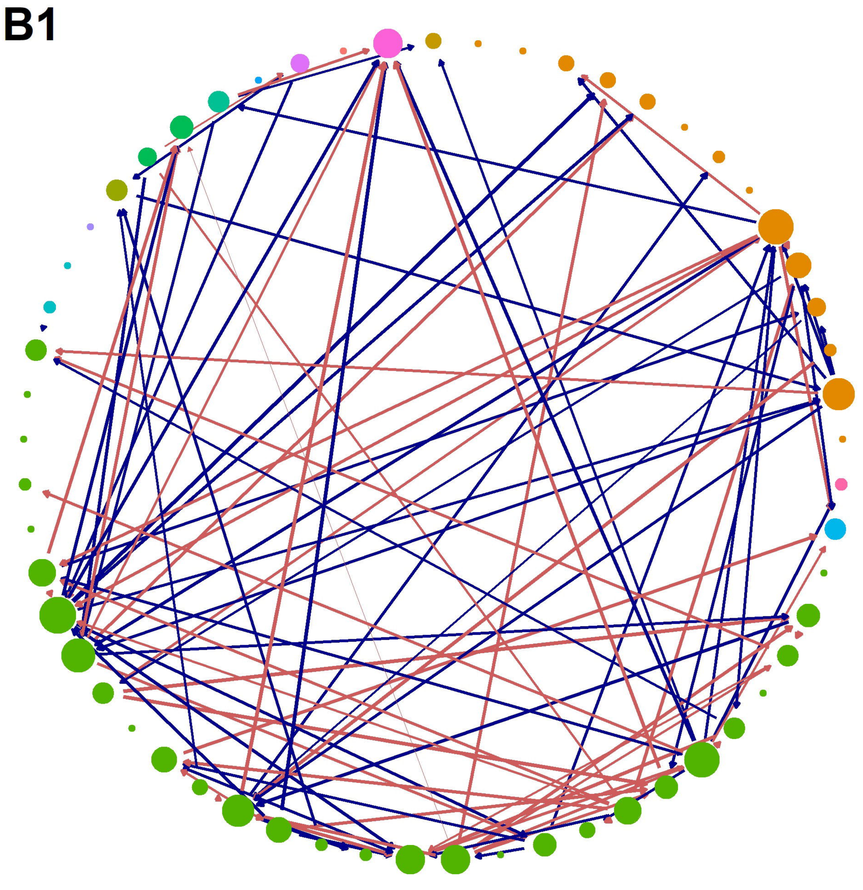

A
B
C
D
1.00
0.75
0.50
0.25
0.00
Low
Med
High
True/False Positive Rate
Mean absolute effect size
Measurements
True Positive
False Positive
Methods
ARZIMM
MAR
Poisson
MDSINE


Environmental change
Less stable
More stable
State 1
State 2
Community state
Conditions

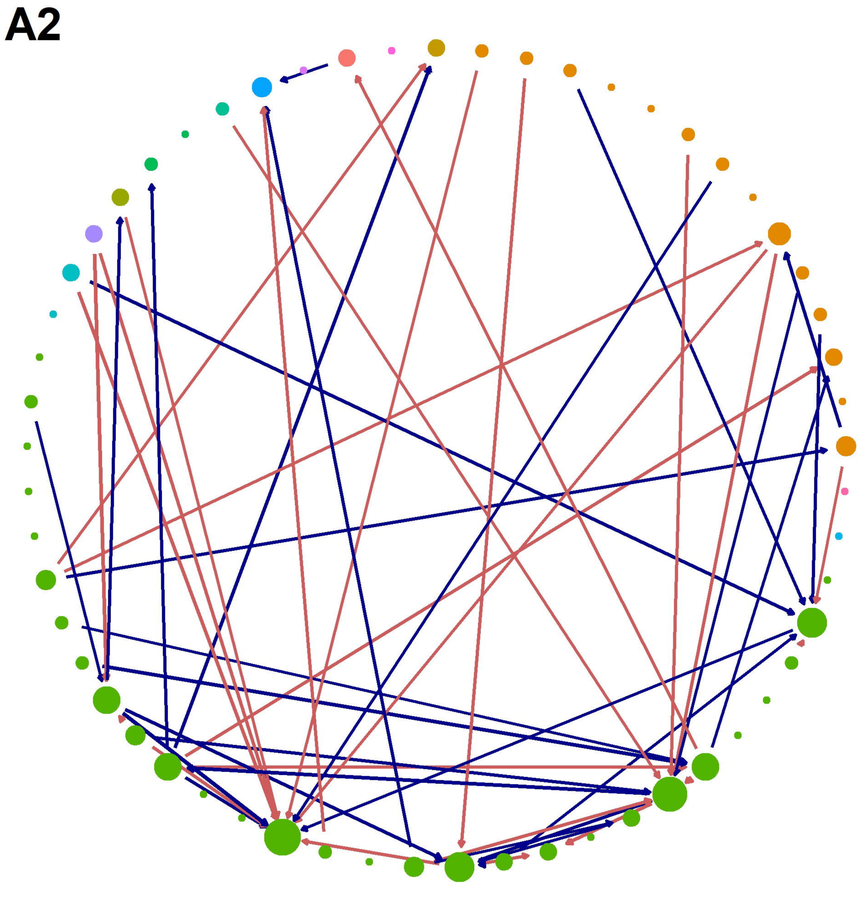
A1



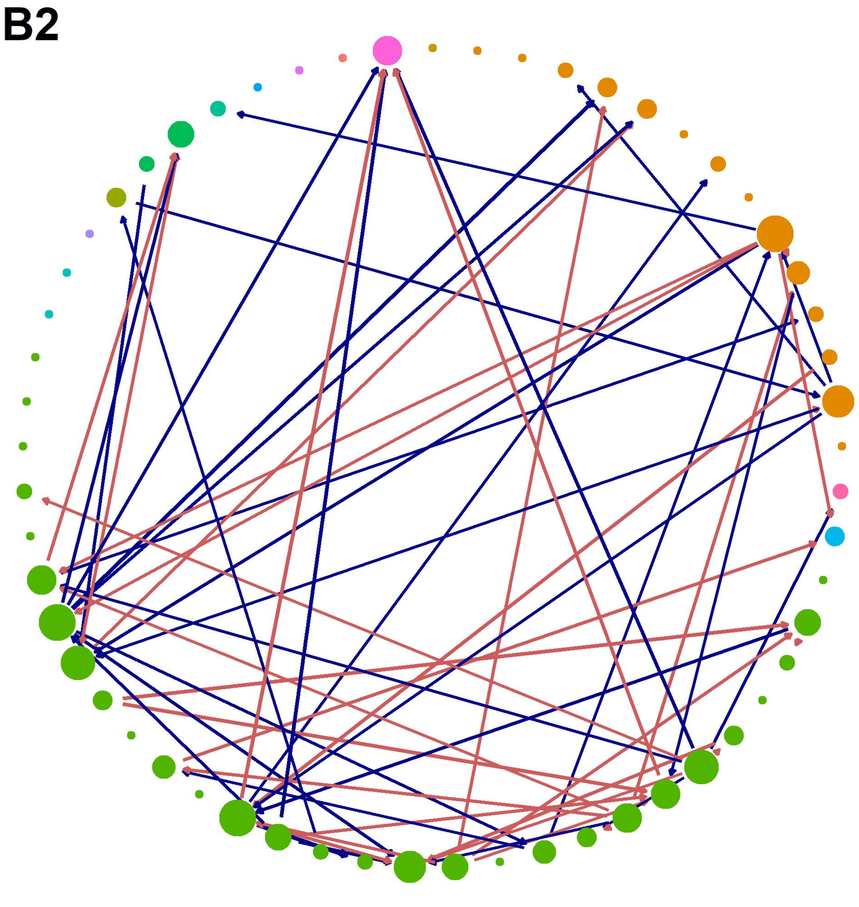
B1



A2



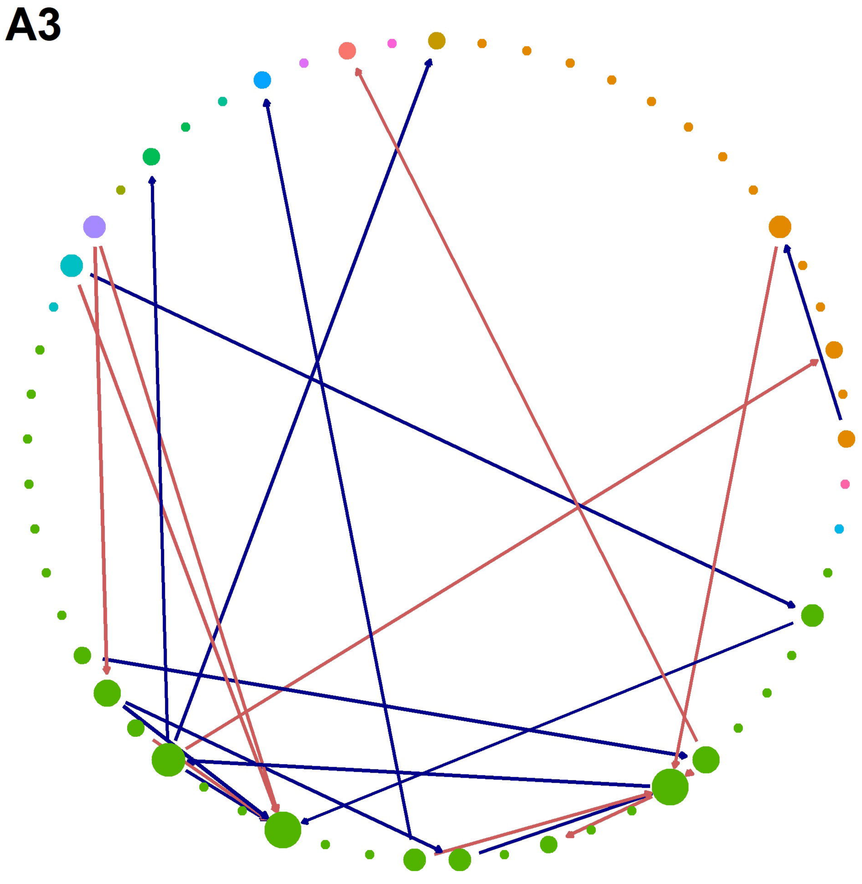
B2



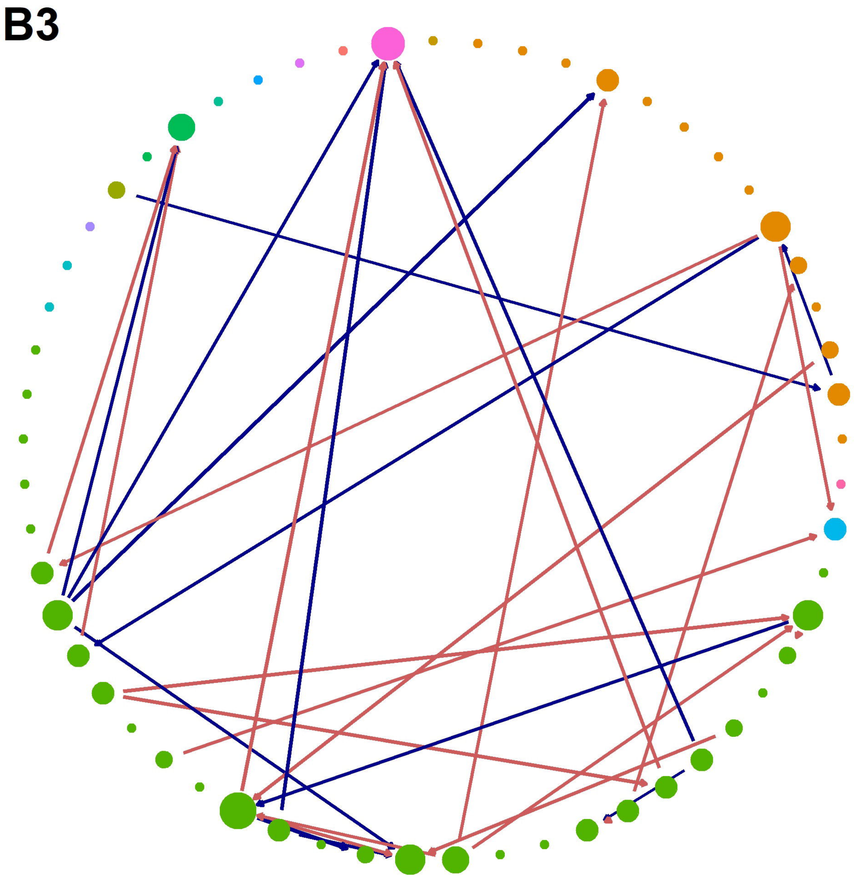
Order

- [Cerasicoccales]
- Bacteroidales
- Bifidobacteriales
- Burkholderiales
- Clostridiales
- Desulfovibrionales
- Enterobacteriales
- Erysipelotrichales
- Lactobacillales
- Pasteurellales
- RF32
- RF39
- Verrucomicrobiales
- YS2

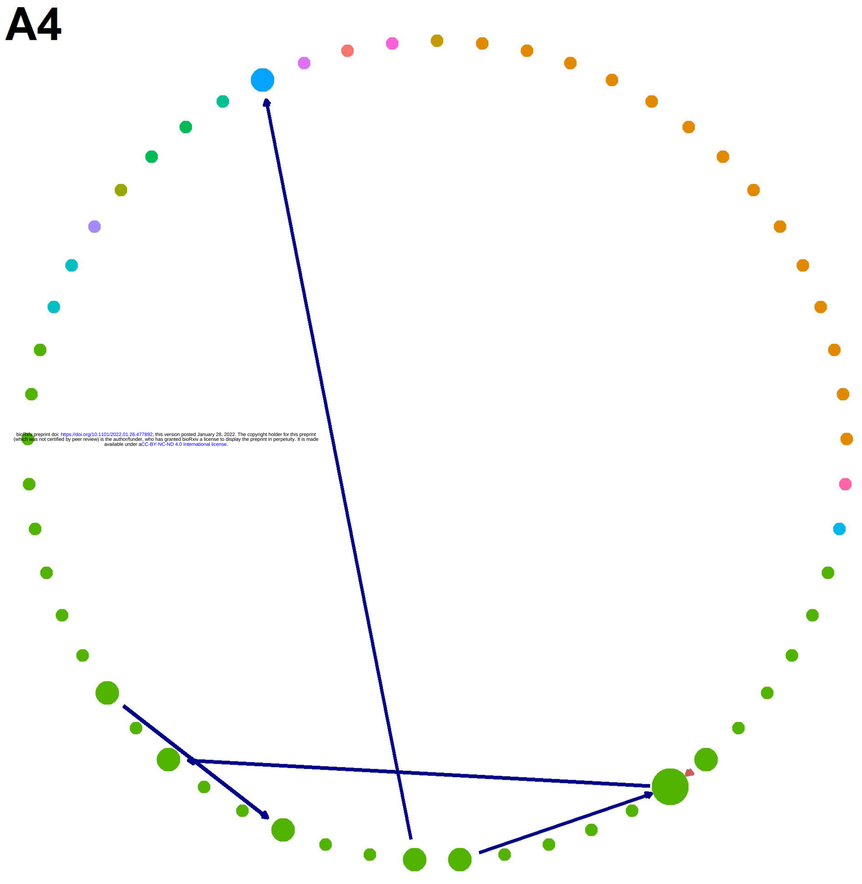
A3



B3



A4



B4