Capturing the dynamics of microbial interactions through individual-specific networks

Behnam Yousefi[1,2,3†], Federico Melograna[3†], Gianluca Galazzo[4], Niels van Best[4,5], Monique Mommers[6], John Penders[4,7], Benno Schwikowski[1] and Kristel Van Steen[3,8]*

[1]Computational Systems Biomedicine Lab, Institut Pasteur, University Paris City, Paris, France, [2]École Doctorale Complexite du vivant, Sorbonne University, Paris, France, [3]BIO3—Laboratory for Systems Medicine, Department of Human Genetics, Katholieke Universiteit Leuven, Leuven, Belgium, [4]Department of Medical Microbiology, Infectious Diseases and Infection Prevention, School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Center+, Maastricht, Netherlands, [5]Institute of Medical Microbiology, Rhine-Westphalia Technical University of Aachen, RWTH University, Aachen, Germany, [6]Department of Epidemiology, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, Netherlands, [7]Department of Medical Microbiology, Infectious Diseases and Infection Prevention, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center+, Maastricht, Netherlands, [8]BIO3—Laboratory for Systems Genetics, GIGA-R Medical Genomics, University of Lièvzge, Liège, Belgium

Longitudinal analysis of multivariate individual-specific microbiome profiles over time or across conditions remains dauntin. Most statistical tools and methods that are available to study microbiomes are based on cross-sectional data. Over the past few years, several attempts have been made to model the dynamics of bacterial species over time or across conditions. However, the field needs novel views on handling microbial interactions in temporal analyses. This study proposes a novel data analysis framework, MNDA, that combines representation learning and individual-specific microbial co-occurrence networks to uncover taxon neighborhood dynamics. As a use case, we consider a cohort of newborns with microbiomes available at 6 and 9 months after birth, and extraneous data available on the mode of delivery and diet changes between the considered time points. Our results show that prediction models for these extraneous outcomes based on an MNDA measure of local neighborhood dynamics for each taxon outperform traditional prediction models solely based on individual-specific microbial abundances. Furthermore, our results show that unsupervised similarity analysis of newborns in the study, again using the notion of a taxon's dynamic neighborhood derived from time-matched individual-specific microbial networks, can reveal different subpopulations of individuals, compared to standard microbiome-based clustering, with potential relevance to clinical practice. This study highlights the complementarity of microbial interactions and abundances in downstream analyses and opens new avenues to personalized prediction or stratified medicine with temporal microbiome data.

1. Introduction

The human gut is a complex ecosystem where microbes interact amongst themselves and with the host (Faust et al., 2012). Variations in the human gut microbial ecosystem can be caused by antibiotics (Isolauri et al., 2017) or by drugs such as metformin (Rodriguez et al., 2018) to prevent or treat Type 2 diabetes, or even by a sudden change in diet

(Singh et al., 2017). Dysbiosis, an imbalance of a microbial ecosystem, has been linked to several complex diseases (Walker, 2017). It may be reflected by alterations in microbial co-abundance (Chen et al., 2020a) or changes in how a community's microbes interact. These microbial perturbations may have short- or long-term health effects. Microbial interactions have been shown to exhibit rich complementary information about various health conditions (Faust et al., 2012). Their disease specificity has been demonstrated for conditions such as inflammatory bowel disease and obesity (Chen et al., 2020a). Furthermore, the gut microbiome involves a dynamic ecosystem, with microbial co-occurrences or interactions changing over time (Ji et al., 2020) and potentially indicating health-to-disease transitions (Baldassano and Bassett, 2016; Einarsson et al., 2019). For instance, perturbations to the microbiome during infancy have been associated with the development of chronic illnesses in later life, including infectious diseases and asthma or allergies (Gaufin et al., 2018; Alsharairi, 2020; Cukrowska et al., 2020).

Two important and highly studied determinants of early-life microbiome establishment are birth mode and infant diet. C-section delivery provides a barrier in the dispersal of maternal faecal and vaginal microbes during delivery and has been linked to various non-communicable diseases and least in part as a consequence of its perturbation in early-life microbial colonization (Bursuker and North, 1984; Galazzo et al., 2020; Stokholm et al., 2020). In addition, the shift from infant feeding (breastfeeding and/or formula) to a more diverse diet consisting of a wide variety of substrates in complementary foods, which largely takes place between 6 and 9 months post-partum, has been associated with a rapid diversification and maturation of the intestinal microbiome (Galazzo et al., 2020).

Capturing time-related patterns in data can be achieved via time series analysis (TSA) or longitudinal analysis (LDA). Such analyses involve time course data and extract additional information from the data compared to cross-sectional studies. The latter involves analysing data limited to a single time point only. The terms time series and longitudinal analyses are sometimes used interchangeably. However, there is a subtle difference, and some methods developed for TSA may not transfer well to LDA contexts. Whereas time series data refer to a sequence of data points, collected at multiple time points or intervals, longitudinal data refer to a subject's or object's measurement(s) taken over time. Standard statistical LDA approaches cannot simply be transferred as such to the microbiome field. This is because of the characteristics of metagenomics data. The noisiness, compositionality, and sparseness of the data pose a big challenge in modelling microbial structure; the complexity of multivariate repetitive data for each individual adds to this challenge. In general, methods for microbiome time course data typically aim to address either one or a combination of the following questions (see also Coenen et al., 2020). Is there a temporal trend? What is the similarity between multiple time-course profiles? Which microbial community members co-evolve? For examples of microbiome LDA analyses that address such questions, we refer to references in Lugo-Martinez et al. (2019).

The construction and interpretation of personalized networks have obtained renewed attention in the context of precision medicine. For instance, Menche et al. (2017) used a template network structure derived from knowledge about protein interactions and, for each individual, superimposed the individual's gene expression scores on nodes. For personalized networks derived in this way, node values are specific to individuals, but edge values are constant across individuals based on reference data. In contrast, we define an individual-specific network (ISN) as a network, for which both nodes and edges can be allocated to a single individual, and that can be seen as a realization of a new measure to describe within-individual activity Liu et al. (2016a); Jahagirdar and Saccenti (2020). One procedure to construct such ISNs, LIONESS, was proposed by Kuijjer et al. (2019a,b). LIONESS constructs an individual-specific network for each individual of interest from perturbations that subtracting the individual from a pool of individuals causes on the population-based interaction network. The development of individual-specific networks is recent, and given the challenges involved with compositional data, only a few applications exist with microbial data. For instance, Mac Aogáin et al. (2021) used LIONESS on microbiome abundances as nodes and Pearson correlations as edge strengths. Individual-specific interactions (edges) are taken as new predictors in models for the time to the next exacerbation in a chronic airway disease. An interesting individual-specific approach to the temporal analysis of microbiome data was introduced by Yu et al. (2019). Their method builds on earlier work, particularly *individual-specific edge-network analysis* (iENA) Yu et al. (2017), and has disease prediction as the objective. The iENA framework overcomes a critical practical difficulty when there are not enough longitudinal samples available for the same individual. An additional application on faecal microbiomes can be found in Chen et al. (2020b).

For the first time, we use Kuijjer's LIONESS method to study microbial dynamic patterns via individual-specific microbial networks. As explained in the Methods Section, their method relies on reference data, i.e., a collection of samples that can be pooled to compute a microbial co-occurrence network. Several strategies exist for constructing a pooled-data microbial co-occurrence network. These strategies either rely on Pearson-like correlations or graphical models and the inference of sparse variance-covariance matrices. For an overview and discussion of common co-occurrence network strategies, we refer to Hirano and Takemoto (2019) and Matchado et al. (2021). Generally, a microbial co-occurrence network takes microbial taxa as nodes and evidence for microbial association as edge strength. In the literature, the microbial association is often called "co-occurrence" or sometimes "interaction." Pooled-sample or population-based models have shown their utility to increase our understanding of underlying characteristics of individuals or to derive personalized predictions (Kosorok and Laber, 2019; Bzdok et al., 2021). However, from the perspective of personalized medicine, if association networks were available for each individual separately, then descriptions of such networks would readily be individual-specific. Moreover, taking those individual-specific networks as new units of analysis, one would use more information from the data than is typically done. Such analyses may involve association models to understand mechanisms, prediction models to estimate the risk of disease or treatment non-response, identification

of endotypes, or, more generally, homogeneous subgroups of individuals that may be targeted together during drug development processes.

In this work, we develop a novel approach to study individual-specific microbial neighborhood dynamics over time or across conditions. We illustrate the approach on microbial data from newborns with measurements at 6 and 9 months over time. We first describe the study design, and the components needed to compute individual-specific microbial networks. Second, we introduce a new microbiome analysis framework, based on representation learning, that we call *multiplex network differential analysis* (MNDA). MNDA generates new representations of local microbial interaction neighborhoods that can be used in supervised or unsupervised models or to identify stable or unstable microbial taxa over time or across interventions. Finally, we present and discuss the results of various microbiome dynamic analyses via MNDA. These analyses broadly fall into three broad classes. The first class covers dynamic analyses of global networks, where a global network refers to a microbial co-occurrence network constructed on a pooled set of individuals. The second class captures dynamic changes of microbial ISNs. These analyses include comparing fluctuations in microbial ISNs over time and tracking each individual in an embedding space. We use microbial neighborhood dynamics to identify subgroups of individuals that are similar in terms of their microbial interaction dynamics, which provides a different viewpoint than the classical *Dirichlet multinomial mixture* (DMM) clustering (Holmes et al., 2012). The third class of analyses aim to enhance the interpretation of findings: topologies of microbial association networks are studied within and between individuals and are linked to infant delivery mode and infant's diet changes between months 6 and 9.

2. Materials and methods

2.1. Study design and microbiome profiling

The LucKi Gut cohort is an ongoing study monitoring gut microbiota development throughout infancy, and early childhood (de Korte-de Boer et al., 2015). Pregnant women from the South Limburg area in the Netherlands were recruited via mother and childcare professionals, through the study website and Facebook. Women were eligible to participate if they gave birth at > 37 weeks of completed gestation. Study questionnaires and faecal samples of the infant were collected at different time points, e.g., 1–2, 4, and 8 weeks, 4, 5, 6, 9, 11, and 14 months. Parents were instructed to collect infant faeces from diapers and freeze them immediately at $20°C$ in their home freezer inside a cool transport container (Sarstedt, Hilden, Germany). Samples were transported to the laboratory, preserving the cold chain. Metagenomic DNA was extracted with a custom extraction protocol involving mechanical and enzymatic lysis (Stearns et al., 2015). The LucKi Birth Cohort Study was approved by the Medical Ethical Committee of Maastricht University Medical Centre (MEC 09-4-058). LucKi is designed according to the privacy rules that are stipulated in the Dutch "Code of Conduct for Health Research."

Microbiome profiling was performed by next-generation sequencing of the 16S rRNA V3-V4 hypervariable gene region.

Thereafter, a DADA2-based pipeline was used to identify *amplicon sequence variants* (ASVs; Stearns et al., 2015). Lastly, a *centered log ratio* (CLR) transformation of data using the ALDEx2 R package was performed to account for the compositional nature of microbiome data, whenever appropriate (Gloor et al., 2017). Additional details are provided in earlier publications (Bartram et al., 2011).

The current study focuses on months 6 and 9 after delivery. These moments in an infant's life are recognized milestones in the maturation of microbial communities, possibly influenced by the change of diet during 6–9 months after birth. For this reason, we include dietary information on infants in the applications of our new framework. Diet type was encoded as $\{0, 1, 2\}$, with "0" representing breast milk exclusively, "2" representing exclusively solid food, and "1" indicating a mix of both. We refer to a diet as *persistent* if it does not change during 6–9 months. Furthermore, we also had information about the mode of delivery (either C-section or vaginal delivery) available at months 6 and 9. We used this information in prediction models and stratified analyses.

2.2. Data pre-processing and exploratory analysis

Selecting informative individuals and taxa and filtering out random noise was achieved with the following prevalence filter: only ASVs with a prevalence exceeding 15% survived the filtering. Prevalence indicates the percentage of samples in which a microbe was detected. The average sequencing depth was 57,392 read counts, while the range was $[11, 123, 105, 921]$, with an interquartile range of 25,346. Pre-processing was done on the merged data of 155 newborns across months 6 and 9. No infants were removed. Out of 1,144 taxa, only 95 (8%) remained after data pre-processing. These 95 taxa were considered for subsequent analyses. In addition, we defined two new classes of microbes: *appearing* or *disappearing* microbes—taxa that meet the 15% threshold at month 9 (appearing), or at month 6 (disappearing), but were filtered out by joint-time point pre-processing.

The basic exploratory analysis involved computation of *α-diversity* (within-sample diversity), at each time point. In the spirit of a CoDA analysis, we used *Aitchison's distance* (Euclidean distance on CLR-transformed abundances), as implemented in the *vegan* R package. We also investigated *β-diversity* or between-sample diversity. In particular, CoDA ordination was implemented via PCA on CLR-transformed compositions. We used the implementation of Calle (2019).

2.3. Global microbial network construction

For the remainder of this manuscript, we refer to a *global network* as a network computed over a set of independent individuals. For instance, such a set can refer to all infants at month 6 after birth. We computed a global network for each time point, by selecting microbial taxa as nodes and calculating association strength via *microbial association graphical model analysis* (MAGMA; Cougoul et al., 2019) on all newborns at

each time point (81 at time point 6 m and 74 at 9 m). MAGMA uses *copula Gaussian graphical models* for which marginals are modeled with zero-inflated negative binomial generalized linear models, and sparseness is induced via a graphical Lasso strategy. From a practical point of view, we used the *rmagma* R package to derive MAGMA-networks (https://gitlab.com/arcgl/rmagma). To optimize its internal penalization parameter, we used the rotation information criterium (RIC; Zhao et al., 2012). For more details about the adopted MAGMA analysis, see Supplementary material—Global Network Construction.

We preferred MAGMA (Cougoul et al., 2019) over commonly used correlation-based measures due to its theoretical advantages (Calle, 2019) and flexibility to adjust for confounders. MAGMA has the advantage of yielding a sparse microbial co-occurrence network, acting as a natural sparsifier. It targets direct associations, removing indirect ones. Notably, in the case of microbiome data, characterized by zero inflation, high heterogeneity, and overdispersion, the popular Pearson correlation as a measure of association may give rise to false positives, as discussed in Friedman and Alm (2012). Several other approaches exist to infer or construct microbial co-occurrence networks, with remarkable wiring differences. See for instance Kishore et al. (2020) and Matchado et al. (2021). Hence, we verified the robustness of MAGMA-obtained global networks (Supplementary material—Stability Analysis). Alternatively, multiple inference methods may be employed in parallel, to assess the impact on conclusions or to construct a consensus network at the expense of computational complexity in downstream analyses (see Section 2.4). In this work, as an alternative to MAGMA, we applied *SparCC* (Friedman and Alm, 2012) via FastSpar's C++ implementation. Results are presented and discussed in Supplementary material—SparCC Analyses. SparCC can handle spurious associations better than its predecessor correlation-based microbial co-occurrence network inference methods, but it may fail to generate a positive definite covariance matrix. We emphasize that pre-processed data entered MAGMA and SparCC routines with default options. No data transformation was performed before MAGMA/SparCC as those analysis frameworks internally accommodate compositional data.

2.4. Individual-specific network construction

We used Kuijjer et al.'s LIONESS method (Kuijjer et al., 2019b) to infer individual-specific networks from global microbial co-occurrence networks, as implemented in Kuijjer et al. (2019a). Each individual-specific edge weight measures the impact of the individual observation on a global network edge. In particular, the edge weights of the n_{th} ISN (for individual n) were computed as

$$w_n^{ij} = Nw^{ij} - (N-1)w_{-n}^{ij} \qquad (1)$$

where w^{ij} and w_{-n}^{ij} are the edge weights of a global network (Section 2.3) and the n_{th} leave-one-out (LOO) network, respectively, for any pair of microbes ($i \neq j$). N is the total number of individuals in a reference population (Figure 1). For large populations ($N \rightarrow \infty$) and under a weight homogeneity condition, the average individual-specific edge weights w_n^{ij} ($n = \{1, ..., m\}$) converges to

the corresponding global edge weights. Namely, the global network can be seen as a weighted average of the ISNs. Weight homogeneity means that the proportion of the weights between individuals is constant between global (w^{ij}) and LOO (w_{-n}^{ij}) networks (Kuijjer et al., 2019b; Suppl. 5.2). We considered our 69 paired infants at months 6 and 9 after birth as two distinct populations. We do that by only considering the paired infants in both time points. Since, in this work, we are only interested in the unidirectional strengths of microbial associations, we replaced all individual-specific edge weights with their absolute values. Notably, even when population-based edge weights, w^{ij} and w_{-n}^{ij}, are positive, the derived individual-specific weight w_n^{ij} may be negative. This occurs when $w^{ij} < \frac{(N-1)w_{-n}^{ij}}{N}$.

2.5. Multiplex network differential analysis

The ISNs constructed in the previous section can be paired into 69 multiplex networks [i.e., networks with multiple layers of matching nodes (Hamilton, 2020; Hammoud and Kramer, 2020)]. Each multiplex refers to a single individual; within each multiplex, a layer refers to an individual-specific network at a particular time. More generally, we assume, as input to a representation learning algorithm, several multiplexes that represent an object, for which matched data are available. Matching may be performed based on repeated measures over time for the same individual (as is the case for our study example). Still, it may also refer to matched samples where two individuals are matched according to shared characteristics. In both scenarios, the availability of individual-specific edges offers the opportunity to investigate local neighborhood stability.

To capture how the local neighborhoods of microbial taxa change over time, we proposed the following algorithm, which we term *multiplex network differential analysis* (MNDA). First, multiplexes were formed. For our LucKi subcohort, ISNs derived in Section 2.4 at months 6 and 9 are paired. Each ISN has 95 nodes, representing the 95 microbial taxa retained in the study, after data pre-processing (Section 2.2). One individual and its multiplex ISN structure is shown in Figure 2A. Second, we developed a network representation algorithm, based on a shallow *encoding-decoding neural network* (EDNN) that forms the core of the MNDA framework (Figure 2B). For the implementation, we used the *Keras* R package. A graphical flow is given in Figure 2B. The inputs and outputs to the EDNN are vectors, one for each node in a multiplex layer. The input vector at the encoder side uses the edge information of each node's immediate (graph distance 1) neighbors. A binary such vector for a particular microbial taxon for individual 1 at month 9 would be a vector of ones and zeros indicating the direct neighboring taxa based on the individual's ISN at month 9. This assumes a binary ISN (i.e., an unweighted network with edges that are either present or not). It is worth mentioning that, in this method, we do not consider self-loops (i.e., the node itself is not seen as a direct neighbor of itself), which makes our model generalizable to accepting new nodes. A non-binary (weighted) input vector would be a vector for which the ones in the binary vector are replaced by their actual edge strengths. It is here that an additional argument can be made in favour of
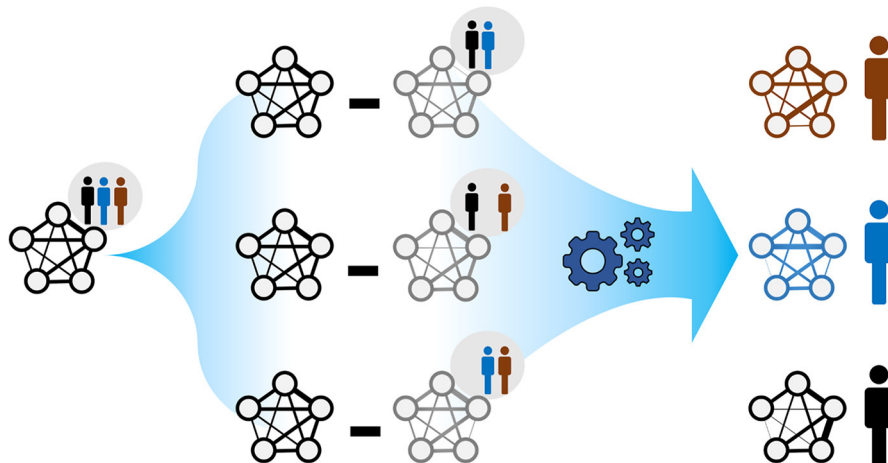
FIGURE 1
The rationale behind ISNs. A global network summarises node and edge information across the individuals belonging to the same population. The goal is to compute individual-specific networks with individual-specific edge weights (as in Kuijjer et al., 2019a); each individual is allowed to exhibit a particular network topology. Comparing such network topologies may not only reveal individual heterogeneity but may also indicate instances where health-related interventions based on the global network may be inappropriate.

microbial co-occurrence network inference methods that lead to sparse networks. MAGMA will generate a sparse weighted network, avoiding the need to work with fully connected networks. Fully connected networks may complicate the interpretability and are less computationally tractable than sparse networks.

The output vector of EDNN, which needs to be predicted at the decoder side, is a representation of more distant neighbors likely to be reached by a *random walk*. In particular, for a specific (seed) taxon, both binary or weighted versions of such an output vector would refer to probabilities that a microbial taxon is reached by a *fixed-length random walk* starting from the seed taxon. Binary and weighted versions differ in how these probabilities are computed (see Supplementary material). A customized implementation of repetitive weighted random walks). As activation functions for the hidden neurons and output neurons, we used *ReLU* and the *Logistic* functions, respectively. The dimension of the hidden space is equal to the number of hidden units and was chosen to be equal to 10 since this resulted in the least *mean squared error* of EDNN compared to the other choices (i.e., 2, 5, 10, 15, 20). Third, after learning the local structure of multiplex network layers and having created representations in a 10-dimensional embedding space, we tracked the positions of the same microbial taxon at months 6 and 9. We formed these pairs for all 95 taxa (Figure 2A), and do this for every individual. We then computed a distance between paired taxa (see next—*A new measure of microbial dynamics*).

2.5.1. A new measure of microbial dynamics

The low-dimensional learnt the local structure of the multiplex network layers could be analysed further by computing the angle θ between embedded vectors belonging to the same microbial taxon at different time points, or the corresponding *cosine similarity* $cos(\theta)$. In a positive space, the smaller $\theta \in (0, \frac{\pi}{2})$, the larger $cos(\theta)$ and thus the larger the similarity between vectors. Even though it is
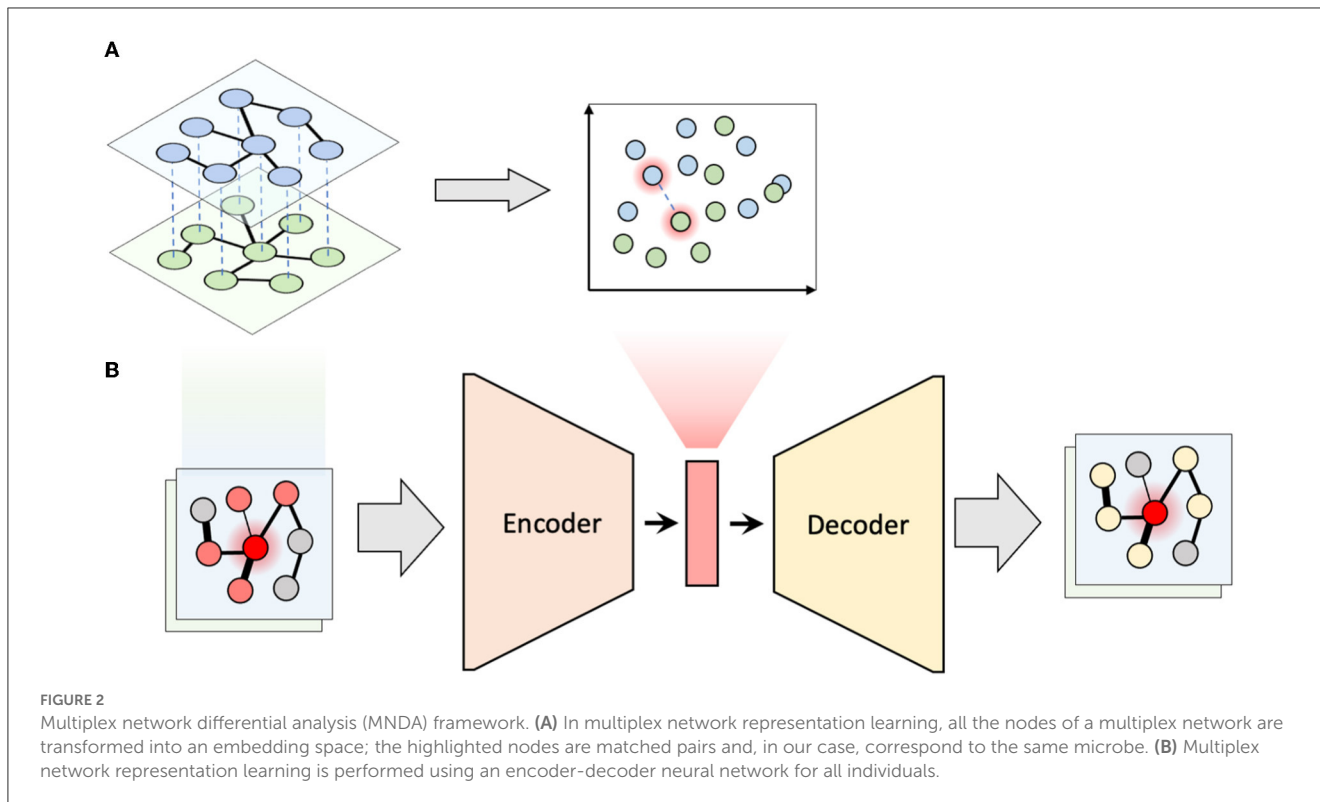
not a genuine distance metric, the cosine distance $d_{cos}(.,.)$ is often used as a complement of cosine similarity in a positive space, and defined as $1 - cos(\theta)$ (without any restrictions on θ):

$$d_{cos}(A, B) = 1 - \frac{A.B}{|A|.|B|} = 1 - \frac{\Sigma_{i=1}^{n} A_i B_i}{\sqrt{\Sigma_{i=1}^{n} A_i^2} \sqrt{\Sigma_{i=1}^{n} B_i^2}} \quad (2)$$

with A_i and B_i ($i \in \{1, ..., n\}$) components of A and B, respectively, and A, B time-specific representations of the same microbial taxon as vectors in the derived embedding space. In contrast, the normalized angle between A and B (called angular distance) is a metric but is defined as $\frac{2\theta}{\pi}$ in positive space and $\frac{\theta}{\pi}$ else. As cosine similarity varies in the range $[-1, 1]$, cosine distances vary in the range $[0, 2]$.

2.5.2. Simulation study

We conducted a simulation study to evaluate the ability of MNDA to capture the neighborhood variation in a multiplex network. To this end, we constructed a graph with node degree and weight distributions similar to our microbial graphs. Then we created a perturbed copy of this graph with a few (m) distinct randomly selected ("control") nodes having different neighboring nodes. This resulted in a multiplex network with control nodes whose neighborhood varies. We expect that in the embedding space, all the nodes will be very close to each other except for the control nodes. Supplementary Figure 1A shows the embedding space for a simulated 2-layer multiplex network; observations align with the aforementioned expectation. To quantify this observation, we first sorted the cosine distance between the node pairs and select the m most distant pairs. Next, we calculated the *Jaccard index* between the pre-set varying nodes and MNDA-based detected nodes and observed that it almost always equals one. Hence, our proposed measure can, with a high degree of confidence, detect nodes whose neighborhoods change from one graph to the other.

FIGURE 2
Multiplex network differential analysis (MNDA) framework. **(A)** In multiplex network representation learning, all the nodes of a multiplex network are transformed into an embedding space; the highlighted nodes are matched pairs and, in our case, correspond to the same microbe. **(B)** Multiplex network representation learning is performed using an encoder-decoder neural network for all individuals.

To further assess the robustness of our method under the noise, we added uniformly distributed noise, ranging between $min = 0$ and $max \in \{e^{-10}, ..., e^{-2}\}$, to the adjacency matrix of each graph. The Jaccard index was calculated and plotted against different noise levels in Supplementary Figure 1B (red diagram).

As for comparison, we used a method based on the eigenvectors of the *Laplacian matrix* of the graphs, which is a typical approach to measuring the distance between graphs. We represented the nodes of each graph layer to the eigenvector space of their Laplacian matrix and used the cosine distance between the node pairs. Compared to MNDA, only a few varying nodes were identified by this method (Jaccard index = 0.3). To investigate the robustness of MNDA against random perturbation or noise in the edge weights, we added uniform noise with different ranges to the adjacency matrices of each layer and assessed its performance in prioritising the dynamics. As shown in Supplementary Figure 1, MNDA is reasonably robust to such uniform noise and consistently performs better than the eigendecomposition-based method.

FIGURE 3
Phyla in the subset of the LucKi cohort. The share of the phyla in the 95 selected microbes is shown. Sample figure title.

3. Results

3.1. Exploratory data analysis

In general, it is well-known that the dominant gut microbiota phyla are *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Fusobacteria*, and *Verrucomicrobia*, with the two phyla *Firmicutes* and *Bacteroidetes* representing 90% of the adult gut microbiota (Rinninella et al., 2019). Moreover, the most prevalent phylum for our retained 95 microbial taxa across 6 and 9 months is *Firmicutes* (53 out of 95). *Bacteroidetes* (14), *Actinobacteria* (15),
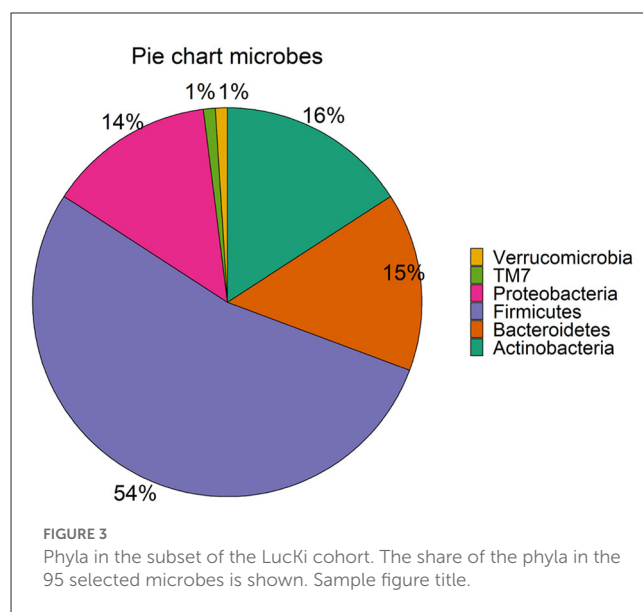
and *Proteobacteria* (13) are (almost) equally represented. *TM7* and *Verrucomicrobia* phyla are represented by a single microbe each (Figure 3). The fractional relative abundances differ between time points. *Actinobacteria* is the most abundant phylum at 6 m, while *Firmicutes* has the greatest share at 9 m. Fractional relative abundances per phylum, order, class, and family are provided in Supplementary Figures 2A-D, respectively.

Microbiota diversity was assessed via α- and β-diversity on the 69 paired newborns. Violin plots of α-diversity grouped per

time point, delivery type, and diet are depicted in Figure 4. Violin plots combine classical box plotting with kernel density graphs. We found evidence for a significant difference in α-diversity between months 6 and 9. Specifically, the paired Mann-Whitney U-test shows a low p-value ($3.843 * 10^{-8}$), rejecting the hypothesis of no difference between the time points. This agrees with Figures 6A, B, where we can see the increased connection strength at time points 9 m compared to 6 m. Figure 5 shows CoDa ordination plots grouped per time point, delivery type, and diet. These plots do not exhibit marked differences between different modes of delivery or diet, while the first principal component provides separation between the time points.

3.2. Cross-sectional global network analyses

Figures 4A, B show two global microbial networks obtained via MAGMA, one per time point, with edge strengths replaced by their absolute values. Microbial taxa are organized according to their phyla. Supplementary Figure 5 show similar plots but with microbes organized (colored) according to order, class, and family, respectively. The size of the node is proportional to relative microbial abundance. The stronger the edge strength, the thicker the edge line that connects corresponding nodes in these networks. Notably, these networks are not fully connected. This sparsification is due to the graphical Lasso in the MAGMA microbial network computation.

Stronger absolute correlations can be observed in the global MAGMA network at month 9 compared to month 6 (Figure 6A). In Figures 6C–F, we grouped microbes of the same phylum into a meta-node and depicted the connections based on the global network co-occurrences. Hence, a node represents a phylum, with a size related to the number of taxa in the phylum, with the exact number indicated in the node itself. The microbes for each phylum can have connections with microbes of the same phylum (indicated by self-loops) or with microbes in different phyla. Edge thickness for edges between phyla is related to the strength of phylum-phylum associations, normalized to account for potential differences in phylum sizes. In particular, for unweighted MAGMA edges (Figures 6C, D), the number of edges η between two phyla P_1 and P_2, with respective sizes n_1 and n_2, was benchmarked against the maximum number of edges $n_1 n_2$:

$$\hat{\eta}_{norm} = \frac{\eta}{n_1 n_2} \qquad (3)$$

For phylum self-loops, $e.g.$, in a phylum P of size n, the denominator in (4) was adapted to $\frac{n(n-1)}{2}$. This normalized count has a natural interpretation of a percentage. We obtained the unweighted MAGMA network by binarizing the MAGMA output, with a 1 for every non-zero entry. For weighted MAGMA edges (Figures 6E, F), the connection strength η between two phyla P_1 and P_2 was defined as a normalized sum of edge weights, with the same normalizing factor as for binary networks. In this scenario, the interpretation of $\hat{\eta}_{norm}$ as a percentage is no longer possible, since η is not normalized for its maximum value.
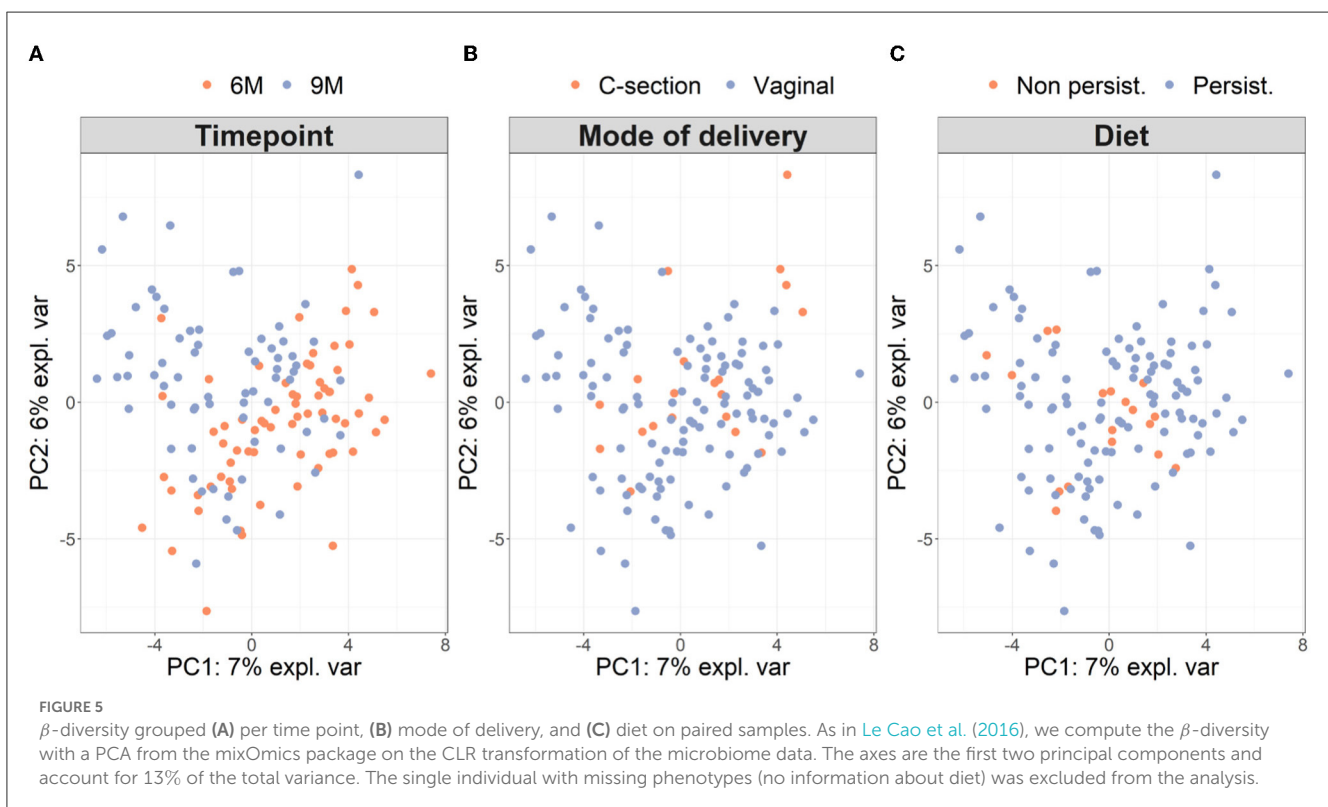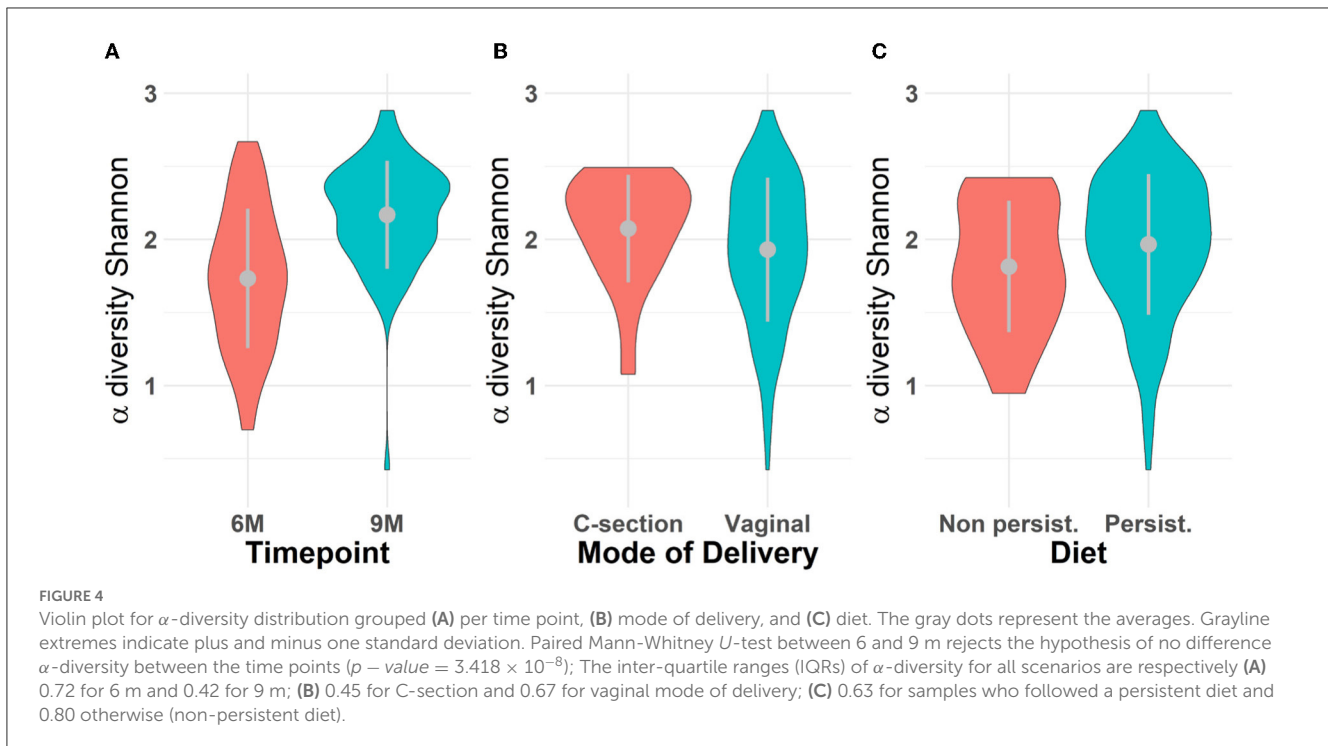
$Firmicutes$ is the largest phylum, with 51 microbes, and shows limited intra-phylum and inter-phylum associations.

$Proteobacteria$ and $Bacteroides$ are medium-sized phyla, with 13 and 14 taxa, respectively (Section 3.1), and have strong intra-phylum connectivity. In particular, intra-phylum association strength for $Bacteroides$ is close to 20% in the binary global networks (Figures 6C, D) in both time points, meaning that almost one-fifth of all the possible $Bacteroides$ interactions are present. The strongest binary associations are inter-phylum, appearing at month 9 (between $Firmicutes$, $TM7$, and $Proteobacteria$). Only one microbe (lowest annotation: Class $TM7$-3) belongs to the $TM7$ phylum, with a moderate—and decreasing in time—association with the $Actinobacteria$ phylum. Similarly, one microbe ($Akkermansia$ $muciniphila$) belongs to the $Verrucomicrobia$ phylum and shows increased connections with $Firmicutes$ and $Bacteroides$ phyla at 9 m. While these phyla associations may be strong in binary global networks, their corresponding association strengths in weighted global networks remains rather limited.

The corresponding global microbial networks obtained with SparCC are included in Supplementary Figures 6, 7. Note that no binarization is available for SparCC, i.e., is a weighted network. The same pre-processing steps of MAGMA network, i.e., prevalence and abundance filters, are applied to SparCC microbial network. We notice that intra-phylum connections are stronger than inter-phyla. SparCC connections are, on average, stronger than in the weighted MAGMA's counterparts. As for MAGMA, $Proteobacteria$ and $Bacteroides$ yield stronger intra-phylum connectivity than $Firmicutes$. Only one microbe (the lowest annotation: Class $TM7$-3) belongs to the $TM7$ phylum, and shows increased connections with $Firmicutes$, $Proteobacteria$, and $Verrucomicrobia$ phyla at 9 m. Moreover, it is interesting to note that for both MAGMA and SparCC networks, the connections involving $Proteobacteria$'s taxa are stronger at time point 9 m.

3.3. Longitudinal analysis: neighborhood dynamics in global networks

We applied MNDA to a multiplex network with two layers, each consisting of the global microbial co-occurrence networks at months 6 and 9. We also computed cosine distances between all possible pairs of microbial taxa in the embedding space (Section 3.2). The measures of dissimilarity can be used to cluster taxa pairs. Pairs may involve components from the same layer or different layers of the network. However, neural networks usually have a non-convex cost function; thus, reruns of the EDNN can lead to different outcomes. Therefore, multiple repeats of MNDA were used as inputs to a novel implementation of robust clustering (see Supplementary material—a novel implementation of ensemble consensus clustering). This led to two robust clusters, shown in Figure 7A as $cluster$ 1 and $cluster$ 2. The confusion table capturing how many matched months 6–9 pairs of microbial taxa belong to the same cluster or are spread over two clusters, is given in Figure 7B. Corresponding transition probabilities $[P_{i \rightarrow j} = P(\text{cluster } j \text{ at } 9 \text{ m} \mid \text{cluster } i \text{ at } 6m)]$ are visualized in Figure 7C. These show that most microbes have a similar global network neighborhood dynamic. For 24 out of 95 microbial taxa this is not the case. These are listed in Supplementary Table 1, with their corresponding genus-species names. The two microbial taxa with
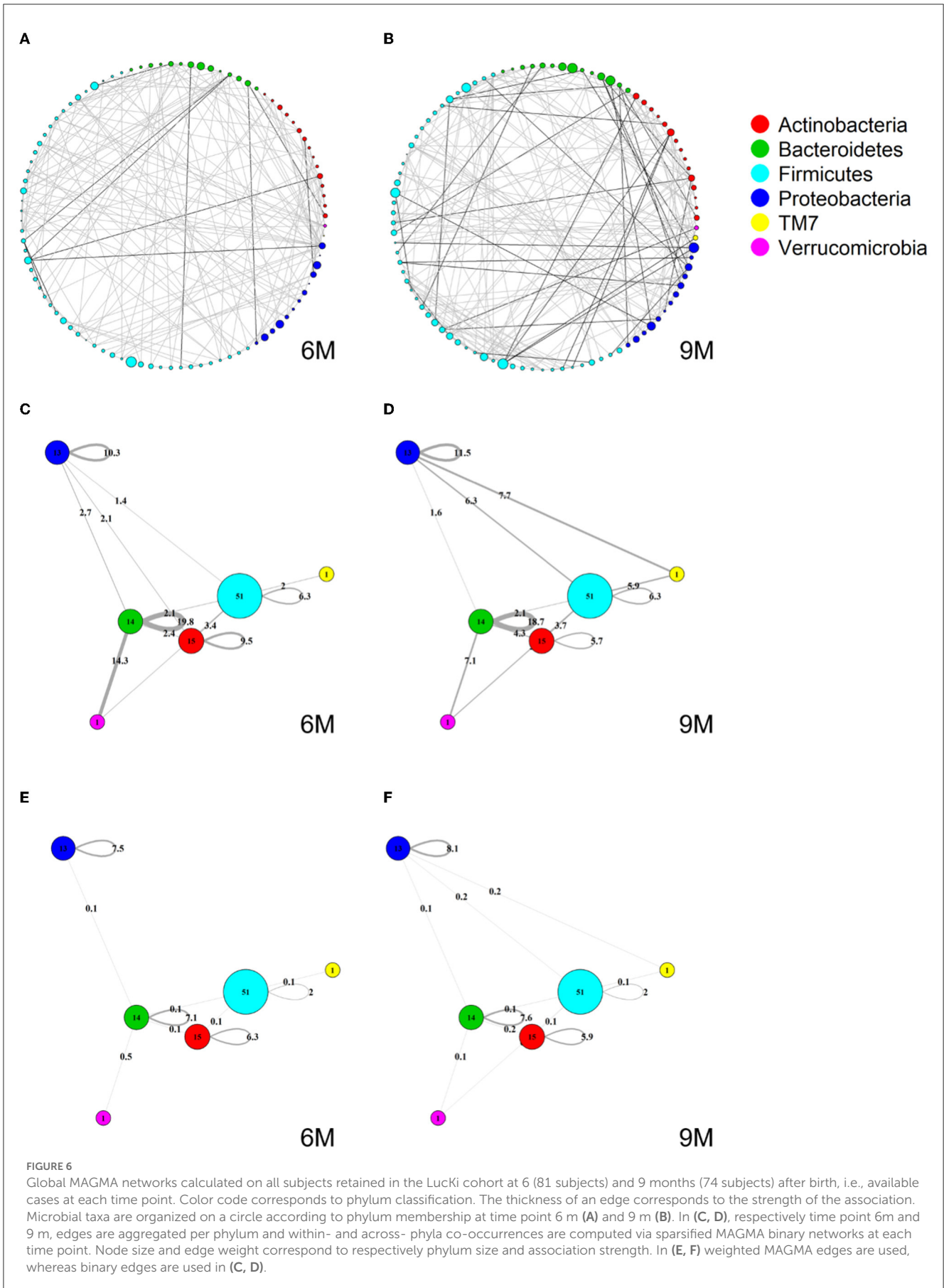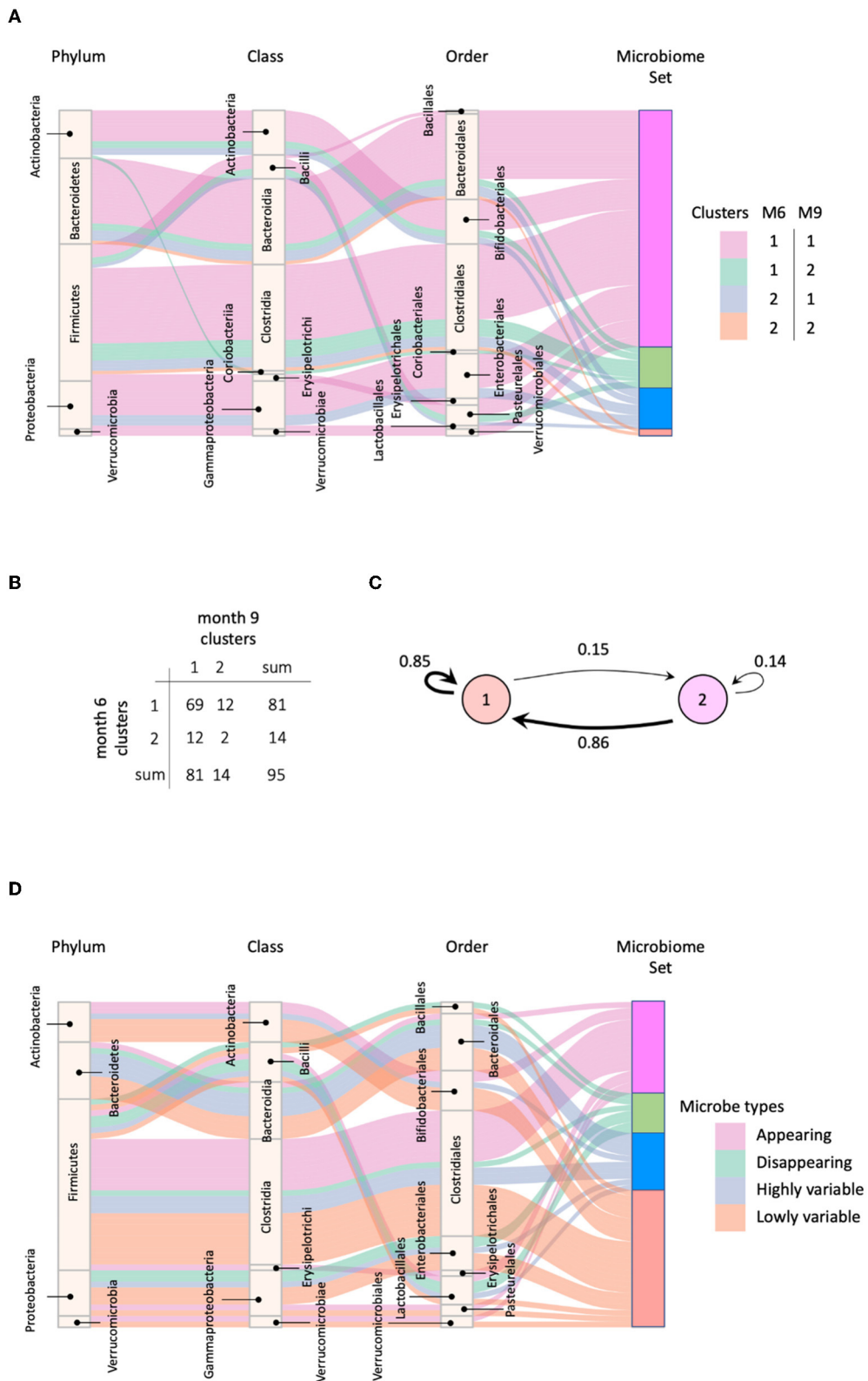
FIGURE 4
Violin plot for α-diversity distribution grouped **(A)** per time point, **(B)** mode of delivery, and **(C)** diet. The gray dots represent the averages. Grayline extremes indicate plus and minus one standard deviation. Paired Mann-Whitney U-test between 6 and 9 m rejects the hypothesis of no difference α-diversity between the time points ($p-value = 3.418 \times 10^{-8}$); The inter-quartile ranges (IQRs) of α-diversity for all scenarios are respectively **(A)** 0.72 for 6 m and 0.42 for 9 m; **(B)** 0.45 for C-section and 0.67 for vaginal mode of delivery; **(C)** 0.63 for samples who followed a persistent diet and 0.80 otherwise (non-persistent diet).

FIGURE 5
β-diversity grouped **(A)** per time point, **(B)** mode of delivery, and **(C)** diet on paired samples. As in Le Cao et al. (2016), we compute the β-diversity with a PCA from the mixOmics package on the CLR transformation of the microbiome data. The axes are the first two principal components and account for 13% of the total variance. The single individual with missing phenotypes (no information about diet) was excluded from the analysis.

similar global neighborhoods, yet different from those microbes in cluster 1, are *Bacteroides uniformis* and *Blautia sp*.

We identified microbes with high or low neighborhood dynamics by sorting the robust co-clustering similarities for each time-matched pair of taxa. In particular, we selected the first larger jumps at both extreme ends of the similarities, respectively, as shown in Supplementary Figure 8A. A complete list of microbes with extreme neighborhood dynamics is given in Figure 7D and Table 1, which also lists *appearing* and *disappearing* microbes (defined in Section 2.2—*pre-processing*). Notably, rather than using co-clustering similarities to rank taxa in terms of their dynamics, we could also have ranked taxa directly via their corresponding cosine

FIGURE 6
Global MAGMA networks calculated on all subjects retained in the LucKi cohort at 6 (81 subjects) and 9 months (74 subjects) after birth, i.e., available cases at each time point. Color code corresponds to phylum classification. The thickness of an edge corresponds to the strength of the association. Microbial taxa are organized on a circle according to phylum membership at time point 6 m **(A)** and 9 m **(B)**. In **(C, D)**, respectively time point 6m and 9 m, edges are aggregated per phylum and within- and across- phyla co-occurrences are computed via sparsified MAGMA binary networks at each time point. Node size and edge weight correspond to respectively phylum size and association strength. In **(E, F)** weighted MAGMA edges are used, whereas binary edges are used in **(C, D)**.

FIGURE 7
(A) Four groups of microbes based on their cluster shifts between 6 and 9 m; (B) The frequency of cluster shifts between 6 and 9 m; (C) Graphical illustration of microbial cluster shift rates; (D) Highly and lowly variable microbes.

TABLE 1 The list of highly dynamic microbes with their corresponding genus-species names.

	Microbial taxa*
Appearing at 9M	*Akkermansia muciniphila, Bifidobacterium sp., Coprobacillus cateniformis, Coprococcus sp., Dorea sp., Clostridium hathewayi, Clostridium perfringens, Haemophilus parainfluenzae, Oscillospira sp., Parabacteroides distasonis, Streptococcus sp., Veillonella dispar*
Disappearing at 9M	*Enterococcus sp., Klebsiella sp., Parabacteroides distasonis, Staphylococcus aureus, Streptococcus sp., Trabulsiella sp., Veillonella dispar*
High neighborhood dynamics	*Bacteroides ovatus, Bacteroides ovatus, Bacteroides sp., Bifidobacterium longum, Escherichia coli, Faecalibacterium prausnitzii, Faecalibacterium prausnitzii, Parabacteroides sp., SMB53 sp., Streptococcus sp.*
Low neighborhood dynamics	*Akkermansia muciniphila, Bacteroides sp., Bacteroides sp., Bacteroides sp., Bacteroides sp., Bifidobacterium longum, Bifidobacterium sp., Bifidobacterium sp., Bifidobacterium sp., Enterobacteriaceae unclassified, Coprococcus sp., Escherichia coli, Faecalibacterium prausnitzii, Haemophilus parainfluenzae, Klebsiella sp., Lachnospira sp., Roseburia sp., Rikenellaceae unclassified,Ruminococcus gnavus, Ruminococcaceae unclassified, SMB53 sp., Staphylococcus epidermidis, Streptococcus luteciae, Veillonella dispar, Veillonella sp.*

*Note that the same species name can occur multiple times as multiple amplicon sequence variants (ASVs) can belong to the same species.

distances in the MNDA embedding space, averaged across multiple MNDA runs and re-ranked. This procedure led to similar results (not shown).

Overall, *Firmicutes'* microbes constitute the vast majority of appearing (69%) and disappearing (57%) microbes, i.e., microbes that would pass the 15% prevalence threshold only at month 9 (appearing), or at month 6 (disappearing). Microbes with high (low) neighborhood dynamics constitute 11% (25%) of 95 original microbes. Phyla *Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria*, and *Verrucomicrobia* deliver respectively 1.6% (6.6%), 1.8% (1.8%), 0.6% (1.6%), 0.7% (2.8%), and 0.0% (25%) of high (low) dynamic microbes in terms of their global network neighborhoods.
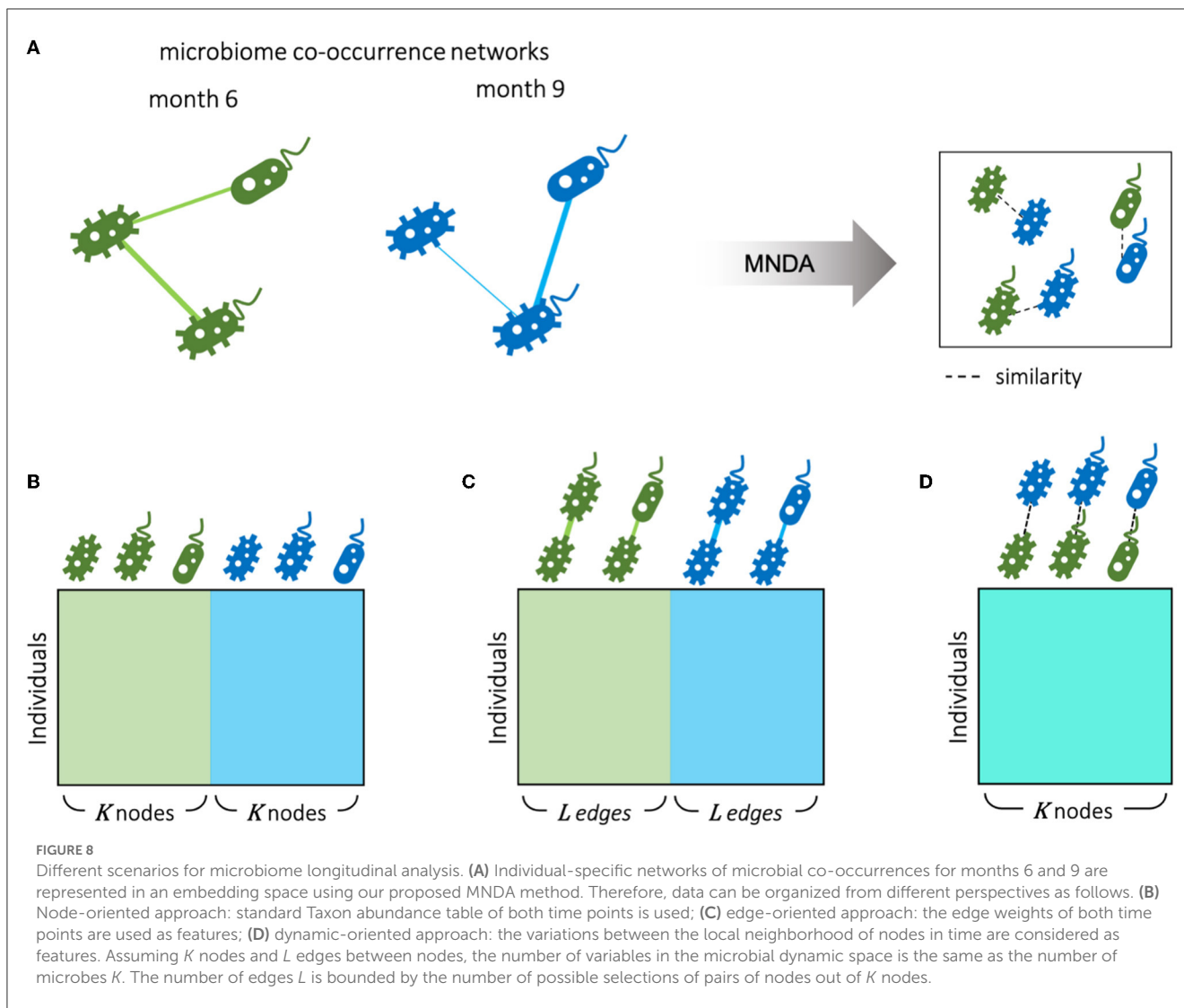
3.4. Longitudinal analysis: neighborhood dynamics in individual-specific microbial networks

An alternative view on microbiome data is given by ISNs, which provide edge information at an individual-specific level. Hence, where in Section 3.4, time-course analyses involved comparing global microbial co-occurrence networks at each time point, here, we do so at an individual-specific level (Figure 8A). We used the same MNDA framework, but instead of submitting a single multiplex network (Section 3.3), we submitted 69 multiplex ISNs simultaneously. We show how our new notion of microbiome neighborhood dynamics (explained in Section 2.5), when assessed on a per-individual level, may offer complementary views compared to standard data views. The most standard data view analyses (transformed) microbiome abundances for study subjects; data is organized according to a matrix depicted in

Figure 8B, where the microbial abundances of both time points are considered as features (node-oriented approach). Since the rise of individual-specific network construction techniques (outlined in the Background Section), data records may additionally (or only) involve information about individual-specific edges (presence or absence, or edge strength on an interval scale). Such an edge-oriented data format is depicted in Figure 8C, where the features are the microbial co-occurrences of both time points (edge-oriented approach). With the newly proposed MNDA framework, dynamics across time points are investigated in an embedding space, and each individual can be assigned a vector of cosine distances. As features, each cosine distance captures ISN local neighborhood dynamics across time points (Figure 8D—dynamic-oriented approach). The similar data formats between Figures 8B–D allow for adopting similar association modelling or prediction modelling strategies, yet interpretations will differ. It is noteworthy that our proposed method results in the least feature size as we have one feature for each taxon independent of the number of time points.

MNDA-induced prediction outperforms other methods and complements standard approaches. To support this statement, we developed prediction models for the mode of delivery (C-section vs. vaginal) and diet type (persistent versus non-persistent—as defined in Section 2.1). In particular, we applied *support vector machines* (SVM) with *radial basis function* (RBF) kernels to the data organized in each of the aforesaid structures (Figure 8). In the training phase, we balanced the classes' size via under-sampling the majority class. To reduce the dimensionality of the data, we used a *forward feature selection* framework. We repeated the modelling process, each time leaving out a single individual, as part of a leave-one-out cross-validation. The left-out individual was used to test the trained model. The entire process was repeated 100 times; AUCs were averaged across repeats and standard errors were computed. The results are reported in Figure 9. MNDA-informed prediction models consistently outperform models that only either use microbial abundance or MAGMA individual-specific edge weights as input features. The advantage of using individual-specific edges is context-dependent: depending on the time point and the phenotype, the classification performances vary. The edges at 9 m have the best performance among MAGMA individual-specific edge weights with AUC of 0.57 and 0.64 for the mode of delivery and diet. However, it is under-performing compared to directly using CLR-transformed microbial abundances at 9 m.

ISN dynamic analysis highlights microbes not identified via global network analyses. Important discriminative microbes for diet type or mode of delivery were identified by counting the number of times a microbe was selected by the adopted forward feature selection algorithm, mentioned above, out of 69 runs (every run had one individual being considered as a validation sample) and 100 repeats. It generates a ranking of microbial importance: the higher the selection count, the higher the microbe's importance. Results over 50 generated embedding spaces were summarized via summing 345,000 repeats per microbe, giving rise to a robust final ranking of important discriminative microbes. We emphasize that the selection of microbes in discriminative models was based on a measure of local neighborhood dynamics across time points. A list of top discriminators in this sense is provided in Table 2, using annotations of genus-species names. Among these taxa, *Lachnospira sp.* and *Bifidobacterium sp.* have
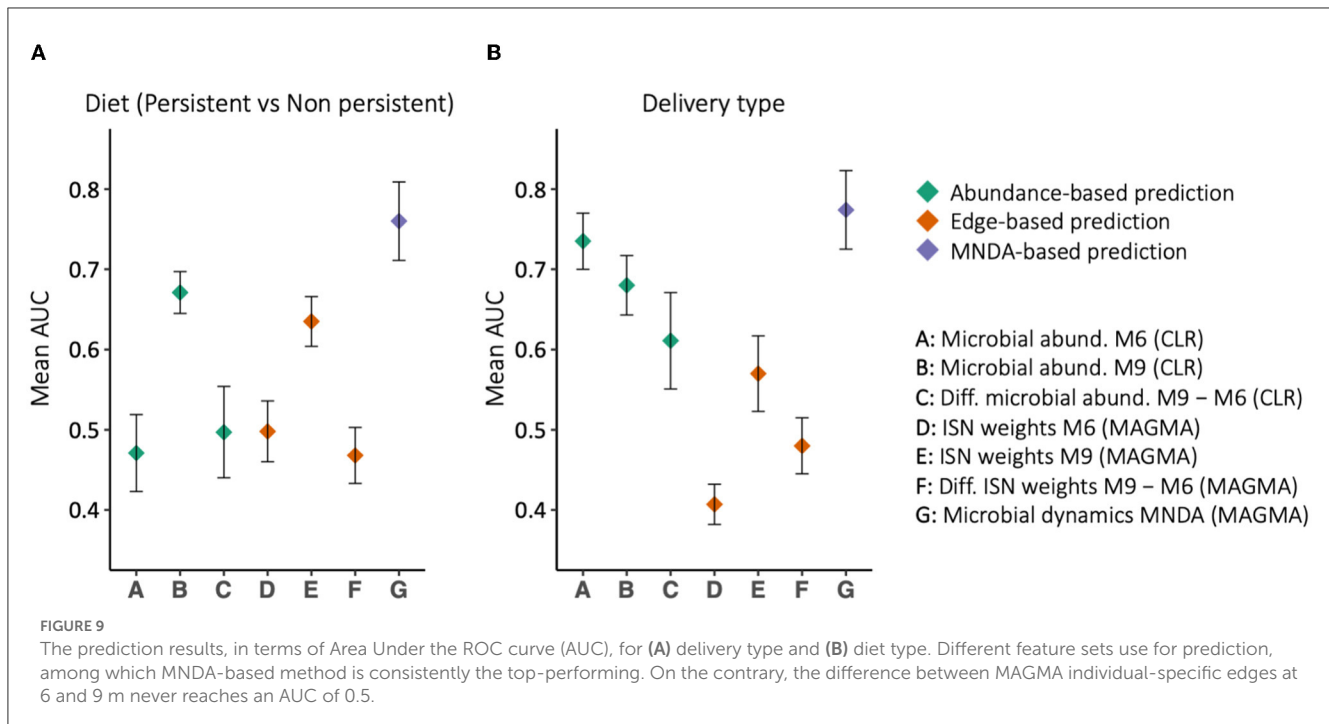
FIGURE 8
Different scenarios for microbiome longitudinal analysis. **(A)** Individual-specific networks of microbial co-occurrences for months 6 and 9 are represented in an embedding space using our proposed MNDA method. Therefore, data can be organized from different perspectives as follows. **(B)** Node-oriented approach: standard Taxon abundance table of both time points is used; **(C)** edge-oriented approach: the edge weights of both time points are used as features; **(D)** dynamic-oriented approach: the variations between the local neighborhood of nodes in time are considered as features. Assuming K nodes and L edges between nodes, the number of variables in the microbial dynamic space is the same as the number of microbes K. The number of edges L is bounded by the number of possible selections of pairs of nodes out of K nodes.

TABLE 2 The list of top-ranked discriminator microbes (genus-species names) for delivery type and diet mode.

	Microbe ASVs
Delivery type	*Streptococcus luteciae, Trabulsiella sp., Ruminococcus sp., Ruminococcus sp., Parabacteroides distasonis, Lachnospira sp., Bifidobacterium sp.*
Diet Mode	*Bacteroides ovatus, Clostridium citroniae, Bacteroides ovatus, Bifidobacterium bifidum, Ruminococcus sp., Streptococcus luteciae, unclassified Lachnospiraceae.*

low neighborhood dynamics; besides, *Streptococcus luteciae* and *Ruminococcus sp.* are important microbes for both delivery and diet types. The latter neighborhood dynamics analysis did not account for differences between infants by diet during months 6 and 9, nor delivery mode. Specifically, *Streptococcus luteciae* had previously been reported to be associated with infant feeding (Brown and Jaspan, 2020); moreover, its association with the delivery type can be explained by its relation to the skin bacterium.

Stratified analyses confirm the differential dynamic behavior of identified discriminators in Table 2. When a microbe is highly discriminative for diet type, and discrimination is based on ISN local neighborhood dynamics over time, then the change over time of its immediate "interaction" partners should also be markedly different between dietary strata; and the same for the mode of delivery. In Figure 10, we focus on the microbial taxa of Table 2 (dark nodes) and their distance-1 neighbors to investigate how the edge weights change over time. For each class of delivery mode (C-section or vaginal delivery) and diet mode (persistent or non-persistent) we obtained two subnetworks for each time point, months 6 and 9, by averaging ISNs from infants of the same class per time point. Then, we subtract the edge weights of networks in month 6 from the edge weights of networks in month 9, resulting in four networks for each class. In Figure 10, the strengths of the presented edge indicate the differences between averaged edge strengths between time points for each class (red for average edge weight at 9 m larger than 6 m and green for the reverse). Each node is annotated with its genus-species name. For further comparison of these difference networks within the delivery mode and diet, we refer to Supplementary Figure 9.

FIGURE 9
The prediction results, in terms of Area Under the ROC curve (AUC), for **(A)** delivery type and **(B)** diet type. Different feature sets use for prediction, among which MNDA-based method is consistently the top-performing. On the contrary, the difference between MAGMA individual-specific edges at 6 and 9 m never reaches an AUC of 0.5.

We observe that the differences in co-expression networks over time show that the connections at 6 M are much stronger than at 9 M for C-section delivery. Furthermore, we observe that a change in diet between the two time points results in apparent differences in co-expression networks (Figure 10C). In contrast, the networks are more stable when infants have a more stable dietary pattern.

Specifically for the diet mode, intra-class variation of ISNs restricted to the same considered taxa in Figure 10 is illustrated via *so-called graph filtration curves* (Figure 11; O'Bray et al., 2021). These curves provide a more refined or complementary view to the averaged ISN representations in Figure 10: Each individual's contribution to the average can be shown. These filtration curves show that the largest variation between the two time points is observed for non-persistent diet (i.e., the variation in the diet mode is correlated with the dynamics of the microbial co-occurrence). Hence, our analysis confirms that dietary shifts have a stronger impact on the dynamics of local neighborhood microbial co-occurrences than observed between groups characterized by different delivery modes.

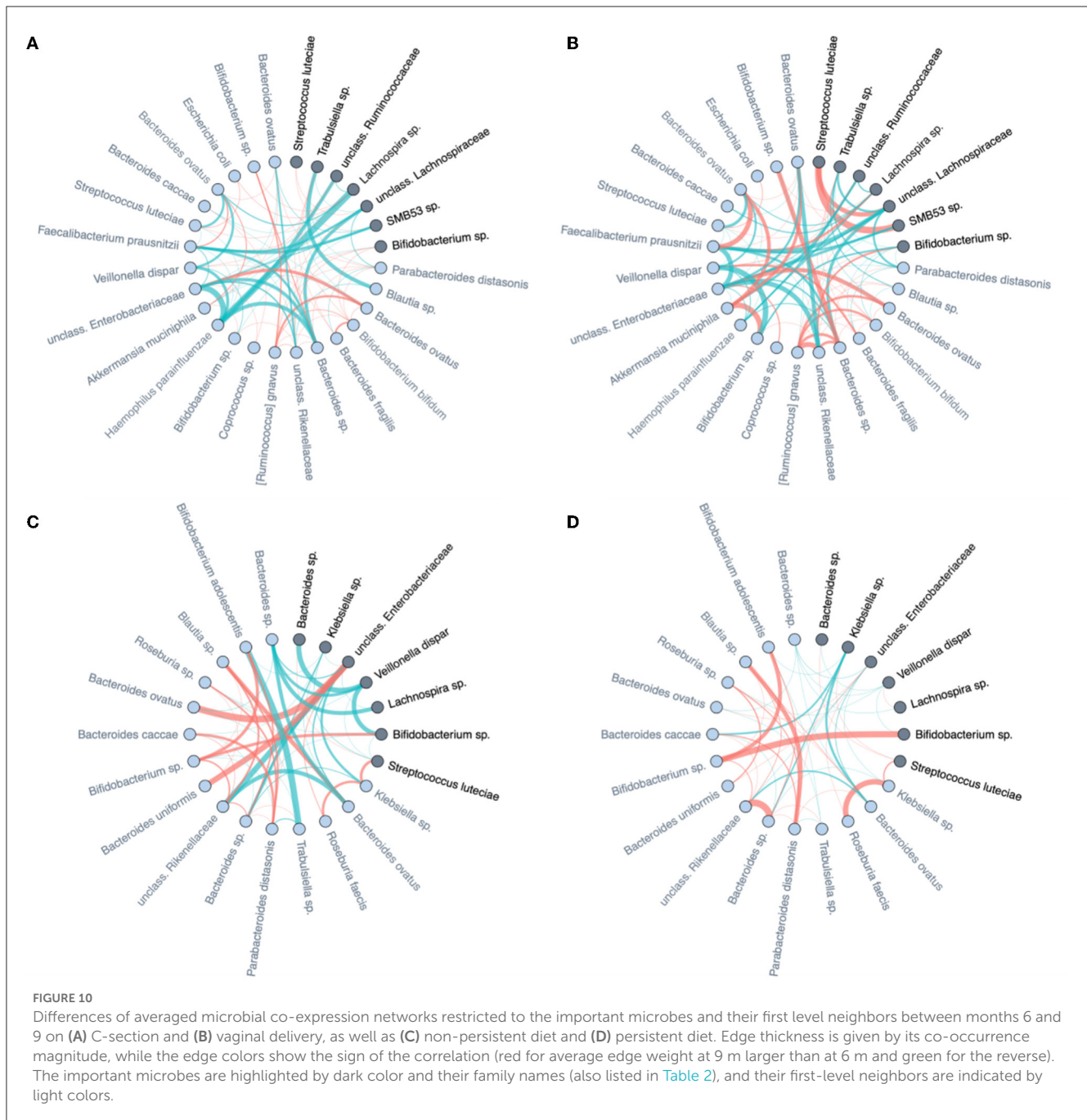## 3.5. Local neighborhood dynamics to identify between-individual heterogeneity

MNDA-based similarity measures can be used to cluster individuals into homogeneous groups according to similar microbial neighborhood dynamics. Using ISNs to define neighborhoods, we observed that clusters of individuals significantly differed from those obtained via *Dirichlet multinomial mixtures methodology* (DMM; Holmes et al., 2012). More specifically, DMM clustering on pooled data across time points

revealed two clusters. These are depicted in Figure 12A, together with corresponding transition information as infants grew older. In contrast to DMM, MNDA-induced clustering groups individuals according to their dynamics-similarity in microbial interaction patterns. Robust clustering was performed, as described in Supplementary material—a novel implementation of ensemble consensus clustering. This analysis also highlights two clusters, roughly of the same size (33 and 36 individuals). We used a Chi-square statistic to evaluate the degree of non-random correspondence between MNDA- and DMM-clusterings. This gave rise to a permutation-based *p-value* of 0.2814 (Figure 12B). It is worth noting that, unlike DMM, MNDA results in communities that do not change over time; they are based on dynamic information across time points. Microbial abundance changes over time, and the dynamics of microbial interactions are two distinct reflections of the same process.

4. Discussion

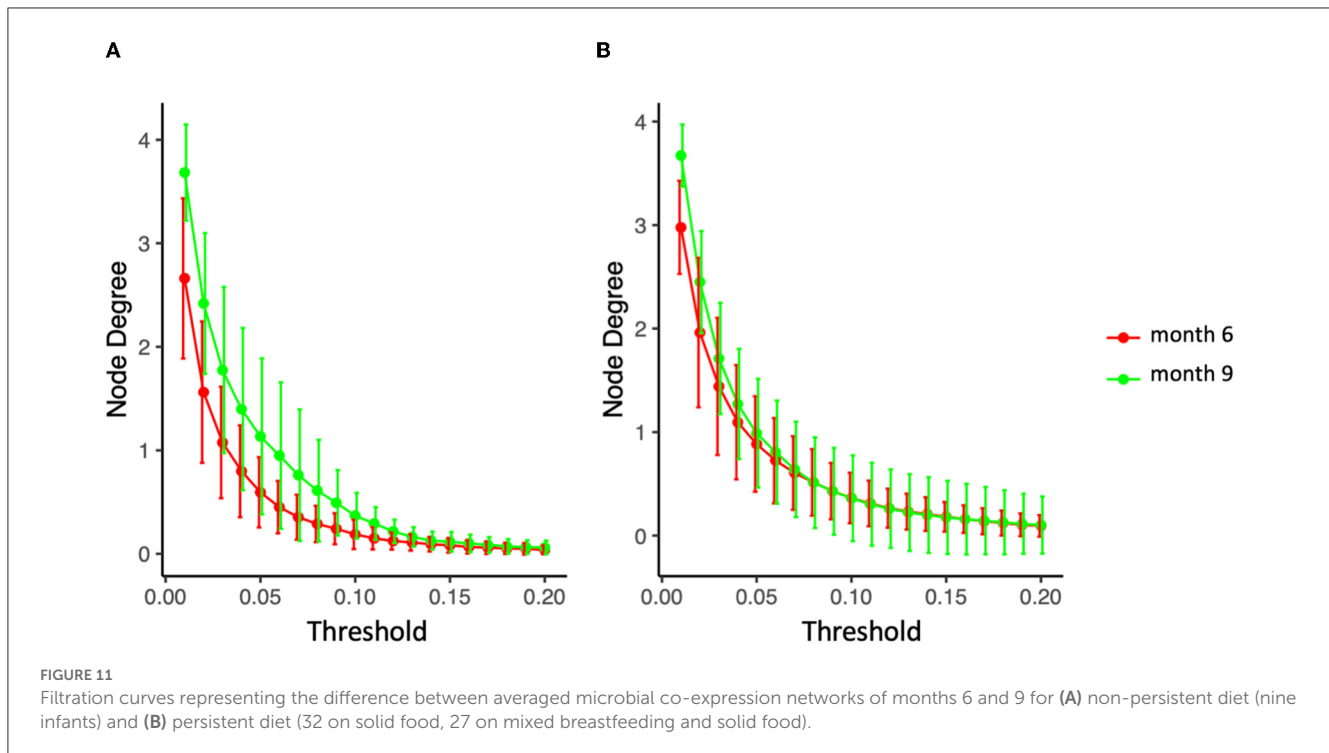### 4.1. The value of individual-specific dynamic microbial networks

Environmental factors may play a critical role in human complex diseases (Qin et al., 2012; Hoyles et al., 2018). One of these critical factors, the gut microbiota, has received special attention in recent years, for instance, in the context of disease development and progression (Tetz et al., 2019; Luna et al., 2020). Even though many studies have shown the resilience of the gut microbiota and its stability over time, the gut microbiota is subject to dramatic shifts due to person-oriented interventions such as changes in diet or medication use.

FIGURE 10
Differences of averaged microbial co-expression networks restricted to the important microbes and their first level neighbors between months 6 and 9 on **(A)** C-section and **(B)** vaginal delivery, as well as **(C)** non-persistent diet and **(D)** persistent diet. Edge thickness is given by its co-occurrence magnitude, while the edge colors show the sign of the correlation (red for average edge weight at 9 m larger than at 6 m and green for the reverse). The important microbes are highlighted by dark color and their family names (also listed in Table 2), and their first-level neighbors are indicated by light colors.

In this work, we proposed a novel work analysis framework, MNDA, to capture the dynamics of gut microbial co-occurrence, within and across individuals. The approach is built on individual-specific microbial networks, with microbial taxa as nodes and individual-specific connecting edges defined via microbial interactions. microbial ISNs were embedded in a shared space. The cosine distance of the same node (microbe) between two time points in the embedding space was taken to quantify local node-neighborhood dynamics. This information was subsequently exploited to stratify individuals into different homogeneous subpopulations, revealing new aspects of population heterogeneity from the microbial interactome perspective. The proposed strategy

was illustrated on data from the LucKi cohort (de Korte-de Boer et al., 2015), containing microbiome profiles of 69 newborns collected at two different time points (6 and 9 months after birth). Via comparison with baseline techniques in the field, we motivated the potential of microbial ISNs in microbiome time-course analyses.

Numerous studies have been performed on microbial longitudinal taxon abundance data, among others, associated with clinical outcomes. However, these studies typically ignore microbial interactions, the dynamics of which could also be highly informative. Taking such microbial co-occurrences into account when modelling temporal dynamics in bacterial

FIGURE 11
Filtration curves representing the difference between averaged microbial co-expression networks of months 6 and 9 for **(A)** non-persistent diet (nine infants) and **(B)** persistent diet (32 on solid food, 27 on mixed breastfeeding and solid food).

communities, *generalized Lotka-Volterra* models (Chung et al., 2017; Lo and Marculescu, 2017) and *dynamic Bayesian network* models (McGeachie et al., 2016; Lugo-Martinez et al., 2019) have been developed. Even though these methods may describe the development of dynamic microbial networks, data are aligned by assuming that patterns are similar across individuals, yet exhibit different rates of change by demographic and clinical variables. Individual-specific microbe neighborhoods in ISNs are ignored. Also, these methods typically require many time points, which may not always be available in microbial cohorts dealing with humans.

Our MNDA framework was exemplified on time course data with two time points. We selected cosine distance as a measure of microbial neighborhood dynamics across two time points, as we aimed to capture the amount of similarity between two data points in a two-dimensional subspace of the MNDA embedding space. Cosine distance is linked to, but not the same as, angular distance, which varies in the range [0, 1]. Angular distance is a true metric but requires the computation of $arccos(.)$. When the cosine distance between data points in the MNDA embedding space is small (or equivalently, cosine similarity is high), the two points will be located in the same general direction from the origin. In other words, one can be seen as a scaled-up version of the other.

4.2. The value of advanced representation learning

At the heart of MNDA lies an encoder-decoder algorithm to embed local microbial neighborhoods at multiple time points. Neighborhoods may be based on global microbial co-occurrence networks or on derived ISNs. Traditional embedding approaches (Hamilton et al., 2017), such as PCA of Laplacian matrix, *DeepWalk*

(Perozzi et al., 2014) and *node2vec* (Grover and Leskovec, 2016), have been previously proposed for single-layer networks. For these methods, the parameters in the encoder are not shared between nodes and cannot be generalized to new nodes. Using an EDNN structure, MNDA can introduce shared parameters while encoding the nodes. In addition, we solved the node generalisation problem by feeding the vector of node neighbors to the encoder. Using this trick, the encoder can generate embedding for a new node whose neighbors need to be determined. In addition, the method can easily be extended to multiple time points by increasing the number of layers of the multiplex network. Our MNDA implementation works with weighted networks by exploiting a fixed-length weighted random walk algorithm. Even though MNDA can be used on binarized networks, with binary input and binary output vectors for the encoder-decoder system (Ietswaart et al., 2021), we do not describe it here since our empirical evaluation indicates suboptimal results.

Computing ISNs can be computationally intensive when the number of edges grows. On a computational infrastructure with *Windows 10* and *R version 4.0.3 (2020-10-10)*, computing 5 ISNs for 95 microbes took 1.17 s. The MAGMA network calculation can rapidly become intractable: for 5 MAGMA networks on 95 microbes, it took 55.15 s. MNDA analysis, including random walk probability calculation and EDNN training, is the most time-consuming step in our analysis. In particular, on a *MacOS (version 12.4)* and *R version 4.0.5 (2021-03-31)*, random walk probability calculation took 28.1 s for each multilayer network of 95 microbes; each round of EDNN training took 5.7 min for 69 individuals. The measurements were done through the *Sys.time()* function in *R*. Embedding individuals into a joint space is less computationally intensive than creating an embedding space for each individual separately. It also ensures that cosine distances are comparable
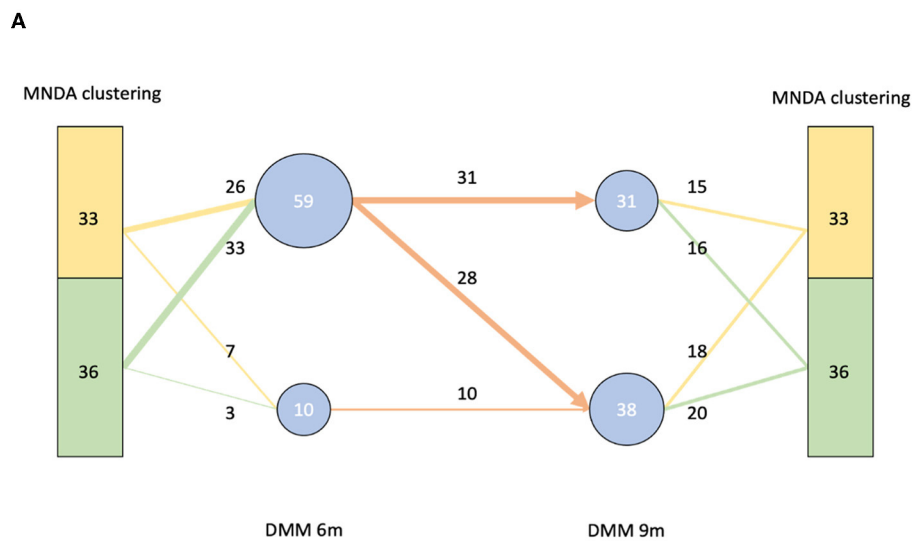
FIGURE 12
A comparison between our MNDA-induced clustering and DMM clustering. **(A)** The DMM clustering at each time point with transition information along with the corresponding MNDA-induced clustering. **(B)** The contingency table of DMM clustering transitions agains MNDA-induced clustering. Our analysis reveals no significant association, which means that MNDA provides an independent view of the data. It is also noteworthy that, unlike DMM which finds clusters for each time point, MNDA provides clustering of individuals based on their variation in both time points.

across individuals (and thus can be used to cluster individuals). However, on the downside, when the initial dataset is enriched with additional individuals, the joint embedding space will differ, and individual predictions may change.

4.3. Unique use of individual-specific network methodology

The ISNs were constructed using Kuijjer et al. (2019b). It builds on a global reference network from which individuals are iteratively extracted and perturbation effects are used to derive an ISN with individual-specific edges. Motivation for this particular way of ISN construction included that it is easy to implement via the LIONESS software (Kuijjer et al., 2019a) and has a straightforward interpretation: On average, and for an asymptotically large number of individuals, Kuijjer's ISNs average to the global reference network. Moreover, unlike for sample-specific networks, approaches such as those described in Liu

et al. (2016b), edges in Kuijjer's ISNs go beyond differential co-occurrences: for each individual a tailored "co-occurrence" network is reconstructed.

Given the compositional and zero-inflated nature of the microbial data, it is necessary to use a global network association algorithm that accommodates these data characteristics. One of the most common tailored microbial association network inference approaches in the field is SparCC (Friedman and Alm, 2012; Watts et al., 2019). Alternatively, Cougoul et al. (2019) proposed MAGMA (*rMAGMA R* package), which not only takes into account microbiomes' noisy structure, excess of zero counts, overdispersion (high skewness) and compositional nature (simplex constraint), but also ensures inferred association strengths within $[-1, 1]$. Moreover, it results in a sparse low dimensional matrix of co-occurrences, with edge strengths that can be adjusted for known confounders. On the contrary, SparCC does not have a native sparsification and needs to be added as an external step, with no established rule on how to perform it, hence, undermining the reproducibility. Since ISNs are computed outside the MNDA

framework, our pipeline can be combined with any type of ISN, as long as the edges are individual-specific.

Compared to the work done by Mac Aogáin et al. (2021), our work has some important differences. First, the authors construct microbial co-occurrence networks on CLR-transformed abundances and Pearson correlation. Pearson correlation is the base measure of association implemented in the LIONESS software, which was initially showcased on gene expression data. Friedman and Alm (2012) showed that even though correlations on the CLR transformation are more accurate than Pearson correlation, they are not as accurate as the SparCC algorithm. Second, the authors continued their analysis with edges as units of analysis. Our work exploits individual-specific networks to define individual-specific ASV neighborhoods (hence sets of edges). Third, MNDA is mainly developed to compare microbiomes across conditions or over time, by stacking ISNs into multiplex networks.

Compared to the iENA protocol (Yu et al., 2017) applied to multi-time point microbial analysis (Chen et al., 2020b), our newly proposed workflow is different in the following ways: (i) firstly, the MAGMA transformation is considered instead of a Pearson correlation on abundances; (ii) LIONESS aims to reconstruct a network with the same interpretation as the global network, i.e., the population-based network, while iENA computes an individual deviation from the average; (iii) edge-network in iENA is aggregated in a single sCI value quantifying the disease's risk, while in our work, the edges constitute the input of the MNDA pipeline. They state that only interactions, not abundances are significant and suggest remarkable disruption of the microbiome community when diseases occur. This understanding is reinforced by a follow-up work from the same group, characterising the personalized microbiome dynamics for disease classification by individual-specific edge-network analysis (Yu et al., 2019). Beneficial co-evolved interactions between host and microbiome can be disrupted by different environmental stresses such as changes in dietary habits, natural physiology, virus infections, and medical treatments (Dethlefsen et al., 2008; Wu et al., 2011; Pop et al., 2014). Generally, analysis techniques developed to process time-varying networks often require numerous temporal observations. The analysis of time-dependent multiplex networks with low temporal dimensions remains largely under-investigated. In this paper, we encapsulate our proposed time-course analysis of ISNs into a multiplex network differential analysis framework, where each network layer refers to a point in time. The framework quantifies the changes in the local neighborhood of each node for an ISN (i.e., a microbial taxon) between the time points. This is achieved by embedding the nodes of network layers into a shared embedding space.

4.4. Complementary novel findings

We uncovered previously unreported microbial taxa as biomarkers for temporal changes between months 6 and 9 after birth. In particular, the two microbial taxa that consistently cluster together in months 6 and 9 and are different from other microbes are *Bacteroides uniformis* and *Blautia sp.* We can find insights from the composition of the appearing/disappearing microbes.

The disappearing *Firmicutes* microbes' are mostly (*facultative*) aerobic microorganisms, including *Enterococcus*, *Streptococcus*, and *Staphylococcus* species, along with other *facultative aerobes* such as *Klebsiella* (phylum *Proteobacteria*). In contrast, the appearing taxa within the *Firmicutes* phylum include several strictly anaerobic species, including *clostridial* members, *Dorea sp.* and *Coprococcus*. This shift indicates a more reduced intestinal environment and more mature *microbiome* adapting to the fermentation of complex dietary carbohydrates. Moreover, *Verrucomicrobia* is recognized as the phylum with the least dynamic microbes in terms of their global network neighborhoods.

Regarding the individual-specific neighborhood dynamics, *Lachnospira sp.* and *Bifidobacterium sp.* have low neighborhood dynamics; *Streptococcus luteciae* and *Ruminococcus sp.* are important microbes for both delivery and diet types. *Streptococcus luteciae* had previously been reported to be associated with infant feeding (Brown and Jaspan, 2020); and, it is a skin bacterium, which can be the reason why in our study it was associated with birth mode (like other skin bacteria) and exhibited low neighborhood dynamics. Other markers, such as *Ruminococcus sp*, *Lachnospira* and *Bifidobacterium sp.*, are too general, and little biological interpretation can be extracted in the context of this study. However, the observed link of *S. luteciae* and *B. bifidum* with diet can potentially trace back to the depletion of these microbes once breastfeeding is ceased.

The differences in co-expression networks over time (as highlighted in Figure 10) show that the connections at 6 M are much stronger than at 9 M for C-section delivery. This is indicative of the waning effect of C-section delivery and temporal colonization of environmental and skin bacteria in C-section delivered infants that are displaced by other bacteria. The interaction networks of such typical C-section delivered microbes also appear to wane over time. This is exemplified by the edge between *Lachnospira sp.* and *Haemophilus parainfluenza*, a bacterium typically found to be temporally enriched in C-section infants, which is much stronger at 6 as compared to 9 months of age. Furthermore, we observed that a change in diet between the two time points results in clear differences in co-expression networks (Figure 10C), while the networks are apparently more stable when infants have a more stable dietary pattern. Note that even infants with persistent dietary patterns still have a diet that is gradually becoming more complex and diverse as more complementary foods are being introduced over time.

Moreover, filtration curves indicate that newborns shifting diets between 6 M and 9 M have the largest variation between the two time points. This would indicate that dietary shifts have a stronger impact on the dynamics of microbial co-occurrence in this age window when compared to C-section, which is in line with previous studies indicating that the impact of birth mode is mainly restricted to the first months of life.

4.5. Limitations and future work

We did not apply our approach to shotgun metagenomics (WMGS), which may be seen as a limitation. In favour of applying MNDA to 16S rRNA gene data, we first observe that we used an

Amplicon Sequence Variant approach on our data, rather than the Operational Taxonomic Unit-based clustering approaches that have been the default approach until a few years ago. As ASVs are exact sequences, this implies that comparisons can be made to a reference database at a much higher resolution allowing for more precise identification. Depending on the length of the 16S rRNA gene region being sequenced, identification down to the species level can even be achieved for many, but not all, ASVs. Indeed, a recent comparative analysis of 16S rRNA gene and metagenomic sequencing of the same pediatric faecal samples as in our study showed that the microbial diversity was similar between WMGS and 16S rRNA gene data and even slightly lower when using WMGS. This latter is likely because WMGS requires much deeper sequencing to cover the total diversity in a sample (Peterson et al., 2021). Altogether, these and other studies indicate that WMGS data do not necessarily recover more microbial taxa when compared to 16S rRNA gene amplicon data. The most significant difference between both methods is that the microbial taxa that are being identified can be more precisely taxonomically classified (e.g., most taxa can be classified down to the species level in WMGS data while ASVs can in some cases be classified down to the species level whereas in other cases taxonomic classification can only be done down to the genus or even family level). Note that this only relates to the naming of the identified taxa, and hence the biological interpretation, and not the identification of these taxa. Second, although WMGS becomes more popular, because of the possibility to look beyond the taxonomy and uncover the functional potential, 16S rRNA gene amplicon sequencing is still a widely applied and valid approach. Especially for extensive population-based studies, 16S rRNA gene amplicon sequencing is still more common, given the lower costs. For the same reason, microbiome studies from low- and middle-income countries are mainly based on 16S rRNA gene amplicon sequencing.

Microbiome data analysis results are particularly susceptible to choices made during virtually all steps of the analysis: for instance, during pre-processing (e.g., normalization Lin and Peddada, 2020), when adopting differential abundance strategies (Nearing et al., 2022), or when carrying out network analyses (Matchado et al., 2021). ISNs can be computed in several ways, with individual-specific edges being binary or weighted, sparse or rich, and positively weighted or not. In this work, we transformed ISN edge weights to their absolute values. Hence, we did not differentiate between positive and negative correlations. Loftus et al. (2021) noticed that taxonomically and functionally similar species tend to have positive associations. In the current version of MNDA, taxa with neighborhoods at months 6 or 9, only differing in sign, would be considered highly similar in terms of their neighborhood dynamics. In future work, we aim to adapt MNDA to accommodate positive and negative edge weights. Particularly linked to our MNDA framework, tuning the hyperparameters of EDNN may further enhance performance. These hyperparameters are the number of hidden units (dimension of the embedding space), the number of layers, L_1 and L_2 regularization parameters, and batch size. Although performance can be clearly defined in view of expected prediction accuracy in a supervised context (see Section 3.4—*MNDA inspired prediction*), it is less evident in unsupervised modelling contexts, in the absence of the ground truth. For instance, the relevance of homogeneous subgroups identified in Section 3.5 may become more apparent when associated with variation in extraneous data. For this study, we only had additional information about diet and mode of delivery. No significant association was observed between diet, mode of delivery and cluster membership (using chi-squared test $\alpha = 0.05$).

Several steps in our MNDA framework can be varied. More specifically, we could have chosen different strategies to construct the ISNs that serve as input to our MNDA framework. For instance, even when adopting Kuijjer's formula to construct individual-specific networks from a global network, the global network edges themselves can be inferred in several ways. Our stability analysis (more details in the Supplementary material) showed that different inference methods for microbial interactions might lead to different global networks and thus may impact the downstream construction of Kuijjer's ISNs. We argue that ISNs derived from valid inference methods, albeit different, can be seen to represent different views of the same data. Furthermore, we could have chosen Euclidean distance instead of cosine distance. However, Euclidean distance calculations are computationally intensive and are often replaced by Manhattan distance for high-dimensional data. In scenarios of high dimensionality, the approximation error introduced by the Manhattan distance may be unacceptably large and thus undesirable. For the user who does wish to adopt Euclidean distances instead, Cardarilli et al. (2019) proposed an approximation method to Euclidean distance in high-dimensional spaces. Building upon our default implementation of cosine similarity, in principle, we can generalize the adopted similarity measure to more than two time points by moving from angular similarity to similarity based on *dihedral angles* (i.e., angle between two intersection half-planes) or generalized solid angles of pointed convex cones [i.e., the intersection of a finite number of half-spaces whose corresponding hyperplanes meet in exactly one point (DeSario and Robins, 2007)]. Assessing the performance of multiple distance measures to capture microbiome neighborhood dynamics as a biomarker for prediction or stratified medicine is a logical next step.

Our methodology generally applies to any scenario for which a network is available for each individual in a targeted set of individuals measured at different conditions (e.g., time, location, and treatment regimen). The MNDA algorithm matches the networks for the same individual according to their nodes and uses the links between nodes (possibly weighted, but not necessarily) to assess whether local neighborhoods are consistent across conditions. In other words, the core of our methodology assumes that individual-specific links between nodes are available. The actual values of the nodes are not directly used in the core procedures. Thus in practice, ISNs can represent molecular interactions for a single (omics) data type or for multiple related (omics) data sources (Koh et al., 2019). For all these scenarios, we believe that our longitudinal analysis framework will be useful to identify novel biomarkers and advance precision medicine.

5. Conclusion

In this paper, we propose a novel framework to uncover microbial neighborhood dynamics. Our approach, MNDA,

combines representation learning and individual-specific microbial networks, which makes it unique in the current landscape of statistical methods for microbiome temporal data analysis. MNDA is not restricted to microbiome data but can handle any data type and measurements, as long as these can sensibly be organized into cross-sectional association networks. Our results show that MNDA can induce predictions that outperform standard approaches and that ISN dynamic analysis can identify microbes that are not identified by global network comparisons. Stratified analysis over clinical variables confirms the differential dynamic behavior of identified discriminators to diet type stability or mode of delivery. Standard microbial abundance changes over time and MNDA dynamics of microbial interactions can be seen as alternate representations of the same underlying process.

Data availability statement

The LucKi Gut Study dataset is available upon request from the Euregional Microbiome Center. Requests to access these datasets should be directed to www.microbiomecenter.eu.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethical Committee of Maastricht University Medical Centre (MEC 09-4-058). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

BY and FM performed the main analysis and contributed to writing the text. GG performed the DMM analysis. NB, MM, and JP contributed in the LucKi cohort and helped interpret the data and the results. JP, BS, and KV supervised the project and contributed to writing the text. All authors contributed to the article and approved the submitted version.

Conflict of interest

The 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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1170391/full#supplementary-material

References

Alsharairi, N. A. (2020). The infant gut microbiota and risk of asthma: The effect of maternal nutrition during pregnancy and lactation. *Microorganisms* 8, 1119. doi: 10.3390/microorganisms8081119v

Baldassano, S. N., and Bassett, D. S. (2016). Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Sci. Rep.* 6, 1–14. doi: 10.1038/srep26087

Bartram, A. K., Lynch, M. D., Stearns, J. C., Moreno-Hagelsieb, G., and Neufeld, J. D. (2011). Generation of multimillion-sequence 16s rrna gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl. Environ. Microbiol.* 77, 3846–3852. doi: 10.1128/AEM.02772-10

Brown, B. P., and Jaspan, H. B. (2020). Compositional analyses reveal correlations between taxon-level gut bacterial abundance and peripheral T cell marker expression in african infants. *Gut Microbes* 11, 237–244. doi: 10.1080/19490976.2019.1643673

Bursuker, I., and North, R. (1984). Generation and decay of the immune response to a progressive fibrosarcoma. II. failure to demonstrate postexcision immunity after the onset of T cell-mediated suppression of immunity. *J. Exp. Med.* 159, 1312–1321. doi: 10.1084/jem.159.5.1312

Bzdok, D., Varoquaux, G., Prediction, S. E., and Association, N. (2021). Paves the road to precision medicine. *J. Am. Med. Assoc. Psychiatr.* 78, 127–128. doi: 10.1001/jamapsychiatry.2020.2549

Calle, M. L. (2019). Statistical analysis of metagenomics data. *Genom. Informat.* 17, e6. doi: 10.5808/GI.2019.17.1.e6

Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Nannarelli, A., Re, M., and Spanò, S. (2019). *n*-dimensional approximation of euclidean distance. *IEEE Trans. Circuit. Syst. II.* 67, 565–569. doi: 10.1109/TCSII.2019.2919545

Chen, L., Collij, V., Jaeger, M., van den Munckhof, I. C., Vich Vila, A., Kurilshikov, A., et al. (2020a). Gut microbial co-abundance networks show specificity in inflammatory bowel disease and obesity. *Nat. Commun.* 11, 1–12. doi: 10.1038/s41467-020-17840-y

Chen, X.-Y., Fan, H.-N., Zhang, H.-K., Qin, H.-W., Shen, L., Yu, X.-T., et al. (2020b). Rewiring of microbiota networks in erosive inflammation of the stomach and small bowel. *Front. Bioeng. Biotechnol.* 8, 299. doi: 10.3389/fbioe.2020.00299

Chung, M., Krueger, J., and Pop, M. (2017). Identification of microbiota dynamics using robust parameter estimation methods. *Math. Biosci.* 29, 71–84. doi: 10.1016/j.mbs.2017.09.009

Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., and Weitz, J. S. (2020). A primer for microbiome time-series analysis. *Front. Genet.* 11, 310. doi: 10.3389/fgene.2020.00310

Cougoul, A., Bailly, X., and Wit, E. C. (2019). Magma: Inference of sparse microbial association networks. *BioRxiv* 2019, 538579. doi: 10.1101/538579v1

Cukrowska, B., Bierła, J. B., Zakrzewska, M., Klukowski, M., and Maciorkowska, E. (2020). The relationship between the infant gut microbiota and allergy. The role of bifidobacterium breve and prebiotic oligosaccharides in the activation of anti-allergic mechanisms in early life. *Nutrients* 12, 946. doi: 10.3390/nu12040946

de Korte-de Boer, D., Mommers, M., Creemers, H. M., Dompeling, E., Feron, F. J., Gielkens-Sijstermans, C. M., et al. (2015). Lucki birth cohort study: Rationale and design. *BMC Publ. Health* 15, 1–7. doi: 10.1186/s12889-015-2255-7

DeSario, D., and Robins, S. (2007). A solid angle theory for real polytopes. *arXiv preprint arXiv:0708.0042*. doi: 10.48550/arXiv.0708.0042

Dethlefsen, L., Huse, S., Sogin, M. L., and Relman, D. A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S RRNA sequencing. *PLoS Biol.* 6, e280. doi: 10.1371/journal.pbio.0060280

Einarsson, G. G., Zhao, J., LiPuma, J. J., Downey, D. G., Tunney, M. M., and Elborn, J. S. (2019). Community analysis and co-occurrence patterns in airway microbial communities during health and disease. *ERJ Open Res.* 5, 128. doi: 10.1183/23120541.00128-2017

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* 8, e1002606. doi: 10.1371/journal.pcbi.1002606

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8, 1002687. doi: 10.1371/journal.pcbi.1002687

Galazzo, G., van Best, N., Bervoets, L., Dapaah, I. O., Savelkoul, P. H., Hornef, M. W., et al. (2020). Development of the microbiota and associations with birth mode, diet, and atopic disorders in a longitudinal analysis of stool samples, collected from infancy through early childhood. *Gastroenterology* 158, 1584–1596. doi: 10.1053/j.gastro.2020.01.024

Gaufin, T., Tobin, N. H., and Aldrovandi, G. M. (2018). The importance of the microbiome in pediatrics and pediatric infectious diseases. *Curr. Opin. Pediatr.* 30, 117–124. doi: 10.1097/MOP.0000000000000576

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* 8, 2224. doi: 10.3389/fmicb.2017.02224

Grover, A., and Leskovec, J. (2016). "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* (San Francisco, CA), 855–864.

Hamilton, W. L. (2020). Graph representation learning. *Synth. Lectur. Artif. Intell. Machine Learn.* 14, 1–159. doi: 10.1007/978-3-031-01588-5

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*. doi: 10.48550/arXiv.1709.05584

Hammoud, Z., and Kramer, F. (2020). Multilayer networks: Aspects, implementations, and application in biomedicine. *Big Data Analyt.* 5, 1–18. doi: 10.1186/s41044-020-00046-0

Hirano, H., and Takemoto, K. (2019). Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformat.* 20, 1–14. doi: 10.1186/s12859-019-2915-1

Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* 7, e30126. doi: 10.1371/journal.pone.0030126

Hoyles, L., Snelling, T., and Umlai, U. (2018). Microbiome-host systems interactions: Protective effects of propionate upon the blood-brain barrier. *Microbiome* 6, 55. doi: 10.1186/s40168-018-0439-y

Ietswaart, R., Gyori, B. M., Bachman, J. A., Sorger, P. K., and Churchman, L. S. (2021). Genewalk identifies relevant gene functions for a biological context using network representation learning. *Genome Biol.* 22, 1–35. doi: 10.1186/s13059-021-02264-8

Isolauri, E., Sherman, P. M., and Walker, W. A. (2017). *Intestinal Microbiome: Functional Aspects in Health and Disease: 88th Nestlé Nutrition Institute Workshop, Playa del Carmen, September 2016*. Basel: Karger Medical and Scientific Publishers.

Jahagirdar, S., and Saccenti, E. (2020). Evaluation of single sample network inference methods for metabolomics-based systems medicine. *J. Proteome Res.* 20, 932–949. doi: 10.1021/acs.jproteome.0c00696

Ji, B. W., Sheth, R. U., Dixit, P. D., Tchourine, K., and Vitkup, D. (2020). Macroecological dynamics of gut microbiota. *Nat. Microbiol.* 5, 768–75. doi: 10.1038/s41564-020-0685-1

Kishore, D., Birzu, G., Hu, Z., DeLisi, C., Korolev, K. S., and Segrè, D. (2020). Inferring microbial co-occurrence networks from amplicon data: A systematic evaluation. *bioRxiv*. doi: 10.1101/2020.09.23.309781

Koh, H. W., Fermin, D., Vogel, C., Choi, K. P., Ewing, R. M., and Choi, H. (2019). iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. *NPJ Syst. Biol. Appl.* 5, 1–10. doi: 10.1038/s41540-019-0099-y

Kosorok, M., and Laber, E. (2019). Precision medicine. *Annu. Rev. Stat. Appl.* 6, 263–286. doi: 10.1146/annurev-statistics-030718-105251

Kuijjer, M., Hsieh, P., and Quackenbush, J. (2019a). lionessR: Single sample network inference in R. *BMC Cancer* 19, 1003. doi: 10.1186/s12885-019-6235-7

Kuijjer, M., Tung, M., Yuan, G., Quackenbush, J., and Glass, K. (2019b). Estimating sample-specific regulatory networks. *IScience* 14, 226–240. doi: 10.1016/j.isci.2019.03.021

Le Cao, K.-A., Costello, M.-E., Lakis, V. A., Bartolo, F., Chua, X.-Y., Brazeilles, R., et al. (2016). MixMC: A multivariate statistical framework to gain insight into microbial communities. *PLoS ONE* 11, e0160169. doi: 10.1371/journal.pone.0160169

Lin, H., and Peddada, S. D. (2020). Analysis of microbial compositions: A review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes* 6, 1–13. doi: 10.1038/s41522-020-00160-w

Liu, X., Wang, Y., Ji, H., Aihara, K., and Chen, L. (2016a). Personalized characterization of diseases using sample-specific networks. *Nucl. Acids Res.* 44, e164. doi: 10.1093/nar/gkw772

Liu, X., Wang, Y., Ji, H., Aihara, K., and Chen, L. (2016b). Personalized characterization of diseases using sample-specific networks. *Nucl. Acids Res.* 44, e164.

Lo, C., and Marculescu, R. (2017). "Inferring microbial interactions from metagenomic time-series using prior biological knowledge," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (Boston, MA), 168–177.

Loftus, M., Hassouneh, S. A.-D., and Yooseph, S. (2021). Bacterial associations in the healthy human gut microbiome across populations. *Sci. Rep.* 11, 1–14. doi: 10.1038/s41598-021-82449-0

Lugo-Martinez, J., Ruiz-Perez, D., Narasimhan, G., and Bar-Joseph, Z. (2019). Dynamic interaction network inference from longitudinal microbiome data. *Microbiome* 7, 1–14. doi: 10.1186/s40168-019-0660-3

Luna, P. N., Mansbach, J. M., and Shaw, C. A. (2020). A joint modeling approach for longitudinal microbiome data improves ability to detect microbiome associations with disease. *PLoS Comput. Biol.* 16, e1008473. doi: 10.1371/journal.pcbi.1008473

Mac Aogáin, M., Narayana, J. K., Tiew, P. Y., Ali, N., Yong, V. F. L., Jaggi, T. K., et al. (2021). Integrative microbiomics in bronchiectasis exacerbations. *Nat. Med.* 27, 688–699. doi: 10.1183/13993003.congress-2020.4102

Matchado, M. S., Lauber, M., Reitmeier, S., Kacprowski, T., Baumbach, J., Haller, D., et al. (2021). Network analysis methods for studying microbial communities: A mini review. *Comput. Struct. Biotechnol. J.* 19, 2687–2698. doi: 10.1016/j.csbj.2021.05.001

McGeachie, M. J., Sordillo, J. E., Gibson, T., Weinstock, G. M., Liu, Y.-Y., Gold, D. R., et al. (2016). Longitudinal prediction of the infant gut microbiome with dynamic bayesian networks. *Sci. Rep.* 6, 1–11. doi: 10.1038/srep20359

Menche, J., Guney, E., Sharma, A., Branigan, P. J., Loza, M. J., Baribaud, F., et al. (2017). Integrating personalized gene expression profiles into predictive disease-associated gene pools. *NPJ Syst. Biol. Appl.* 3, 1–10. doi: 10.1038/s41540-017-0009-0

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., et al. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nat. Commun.* 13, 342. doi: 10.1038/s41467-022-28034-z

O'Bray, L., Rieck, B., and Borgwardt, K. (2021). "Filtration curves for graph representation," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Virtual Event Singapore), 1267–1275.

Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY), 701–710.

Peterson, D., Bonham, K. S., Rowland, S., Pattanayak, C. W., Consortium, R., and Klepac-Ceraj, V. (2021). Comparative analysis of 16S RRNA gene and

metagenome sequencing in pediatric gut microbiomes. *Front. Microbiol.* 12, 670336. doi: 10.3389/fmicb.2021.670336

Pop, M., Walker, A. W., Paulson, J., Lindsay, B., Antonio, M., Hossain, M. A., et al. (2014). Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. *Genome Biol.* 15, R76. doi: 10.1186/gb-2014-15-6-r76

Qin, J., Li, Y., and Cai, Z. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi: 10.1038/nature11450

Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., et al. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* 7, 14. doi: 10.3390/microorganisms70 10014

Rodriguez, J., Hiel, S., and Delzenne, N. M. (2018). Metformin: Old friend, new ways of action–implication of the gut microbiome? *Curr. Opin. Clin. Nutr. Metabol. Care* 21, 294–301. doi: 10.1097/MCO.00000000000 00468

Singh, R. K., Chang, H.-W., Yan, D., Lee, K. M., Ucmak, D., Wong, K., et al. (2017). Influence of diet on the gut microbiome and implications for human health. *J. Trans. Med.* 15, 73. doi: 10.1186/s12967-017-1175-y

Stearns, J. C., Davidson, C. J., McKeon, S., Whelan, F. J., Fontes, M. E., Schryvers, A. B., et al. (2015). Culture and molecular-based profiles show shifts in bacterial communities of the upper respiratory tract that occur with age. *ISME J.* 9, 1246–1259. doi: 10.1038/ismej.2014.250

Stokholm, J., Thorsen, J., Blaser, M. J., Rasmussen, M. A., Hjelmsø, M., Shah, S., et al. (2020). Delivery mode and gut microbial changes correlate with an increased risk of childhood asthma. *Sci. Transl. Med.* 12, eaax9929. doi: 10.1126/scitranslmed.aax9929

Tetz, G., Brown, S. M., Hao, Y., and Tetz, V. (2019). Type 1 diabetes: an association between autoimmunity, the dynamics of gut amyloid-producing *E. coli* and their phages. *Sci. Rep.* 9, 9685. doi: 10.1038/s41598-019-46087-x

Walker, W. (2017). "Dysbiosis," in *The Microbiota in Gastrointestinal Pathophysiology*, eds M. Floch, Y. Ringel, and W. A. Walker (Boston, MA: Elsevier), 227–232.

Watts, S. C., Ritchie, S. C., Inouye, M., and Holt, K. E. (2019). FastSpar: Rapid and scalable correlation estimation for compositional data. *Bioinformatics* 35, 1064–1066. doi: 10.1093/bioinformatics/bty734

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344

Yu, X., Chen, X., and Wang, Z. (2019). Characterizing the personalized microbiota dynamics for disease classification by individual-specific edge-network analysis. *Front. Genet.* 10, 283. doi: 10.3389/fgene.2019.00283

Yu, X., Zhang, J., Sun, S., Zhou, X., Zeng, T., and Chen, L. (2017). Individual-specific edge-network analysis for disease prediction. *Nucl. Acids Res.* 45, e170. doi: 10.1093/nar/gkx787

Zhao, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L. (2012). The huge package for high-dimensional undirected graph estimation in R. *J. Machine Learn. Res.* 13, 1059–1062. doi: 10.48550/arXiv.2006.14781