

POLITECNICO DI TORINO

Degree Course
Mathematical engineering

Master's Degree Thesis

**Investigating microbial interactions in the chicken gut
microbiome: a comparative co-occurrence network
analysis of Antibiotic Growth Promoters (AGPs) and
Phytogenic Feed Additives (PFAs) over time**



**Politecnico
di Torino**



Supervisors

Dr. Stefano Berrone
Dr. Thomas Abeel
Dr. Chengyao Peng

Candidate

Giorgia delle Grazie

2024

Abstract

The use of antibiotics as growth promoters (AGPs) in the poultry industry raises global concerns about antimicrobial resistance (AMR). As an alternative, phytogenic feed additives (PFAs) are gaining interest for their potential to enhance intestinal health and animal performance without worsening AMR.

The effects of both the AGP and PFA treatment remain unclear.

For this reason, a thorough analysis is needed to understand how these feed additives impact the poultry microbial community.

In this study, we hypothesized that the poultry microbial microbiome would respond to the feed additives, leading to a healthier community structure with more efficient dynamics. To confirm this, we aim to use co-occurrence networks and network analysis to investigate the effects of antibiotic growth promoters (AGPs) and phytogenic feed additives (PFAs) on the cecum microbial communities in the chicken gut. This approach allows us to visualize and quantify complex interactions among various microbial taxa, providing a more detailed understanding of the dynamics within the microbial community. To achieve this, first we construct networks with microbiome samples collected from poultry at different ages, treated with either a common antibiotic growth promoter, a phytogenic feed additive, or just basal feed as the control group. Based on that, we conducted detailed topological analyses to identify differences among the networks to gain deeper insights into the impact of the feed additives on the cecum microbiome.

Our findings show that AGP treatment significantly reduces network connectivity compared to control samples, confirming the expected impact of antibiotics on microbial interactions. Besides, AGP-treated networks show enhanced resilience and a more distinct modular structure, indicating a robust response to disturbances. On the other hand, PFA treatment results in a moderate reduction in connectivity and robustness, suggesting a less disruptive effect on the microbial network.

Additionally, we identify a stable core of species within the networks that remain consistent across different treatments, indicating that the fundamental structure of the chicken gut microbiome is preserved.

Moreover, using Random Forest classification integrated with network analysis, we pinpoint key microbial species that effectively distinguish between the different treatments, highlighting the specific impacts of AGPs and PFAs on the gut microbiota.

This study provides a comprehensive analysis of the impacts of PFAs and AGPs on the chicken gut microbiome through co-occurrence network analysis. The results provide us with new insights that are crucial for finding and developing effective alternatives for sustainable poultry production.

Contents

1	Introduction	4
2	Releted Works	8
	Challengers for construct co-occurrence network	8
	Varying sequencing depth	8
	Rare Taxa	9
	Indirect Edge	10
	Scoring	11
	Null Model	11
	Analyze the presence of a Core Association Network (CAN)	12
	Anuran	13
	Biomarker	19
3	Material and methods	20
3.1	Study subjects and sample collection	20
3.2	Construct co-occurrence network	21
	Input	21
	Scoring	22
	Assessment of significance	22
	Network analysis	23
	Topological analysis and core identification	23
3.3	Keystone taxa	28
4	Results	30
4.1	Lower connectivity, higher modularity and higher clustering coefficient with AGP and PFA administration	30
4.2	Higher robustness of the microbial community with AGP and PFA ad- ministration	31
4.3	Modular structure of co-occurrence networks and its change	32
4.4	Statistically significant Core identification	38
4.5	Species that distinguish between different treatments	42
5	Discussion	44
5.1	Future work	48
6	Conclusion	49
7	Supplementary	50
	List of Figures	57

List of Tables	59
Bibliography	61

1 Introduction

In the last seventy years, the poultry industry has made significant strides in production efficiency, largely thanks to intensive breeding programs [1]. A major addition to these breeding efforts has been the use of sub-therapeutic doses of antibiotics, known as **antibiotic growth promoters** (AGPs), incorporated into chicken feed to enhance productivity [2].

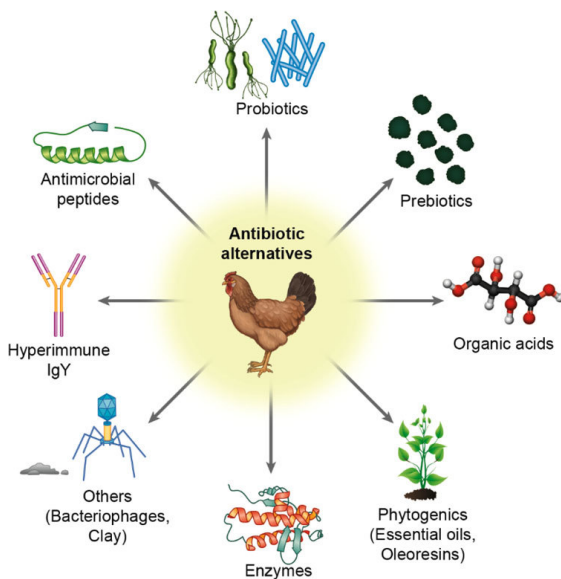


Figure 1. Antibiotics alternatives [3]

Growing global concerns about the correlation between AGPs and antimicrobial resistance (AMR) [4–6] have prompted increasing calls to eliminate their use. For instance, Europe banned AGPs in 2003, while recent efforts in the United States and Canada aim for their elimination [1]. This ban has resulted in increased infections in Europe [7] and reduced production efficiency in Canada [8]. For this reason, the livestock industry [9] is trying to find alternative strategies (as shown in Figure[1]) to promote animal health without aggravating antimicrobial resistance. In this scenario, Phytochemical feed additives (PFA) have been gaining considerable interest lately due to their ability to improve performance by sustaining a healthy gut environment [10].

According to European Union legislation (EC 1831/2003), Phytochemical feed additives (PFA), derived from plants, herbs, and spices, are utilized to enhance animal performance. They have proven highly effective due to their beneficial impact on growth, improved immune system, and influence on caecal microflora composition [9, 11]. Recent findings indicate that PFAs serve as promising alternatives to AGPs [9, 12, 13]. Overall, PFAs possess the capability to mitigate microbial threats and support intestinal health, essential for maximizing bird performance and profitability [9]. The findings indicate that both PFAs and AGPs contribute to increased body weight and improved feed conversion ratio [9].

In recent years, studies on complex microbial communities have made significant progress due to methodological improvements, such as high-throughput DNA sequencing technologies [14]. These technologies provide detailed information on the composition of microbial communities, thanks to sequence data derived from the segmentation of the rRNA gene into small subunits [15]. A wide range of techniques can be employed to analyze this sequence data to describe the composition and diversity of microbial communities, as well as to understand variations spatially, temporally, or in response to experimental treatments [15]. These techniques are employed to examine the impact

of growth-promoting antibiotics (AGPs) and phytogenic additives (PFAs) on the gut microbiota of chickens. The use of both these feed additives has shown to enhance the apparent total tract nutrient digestibility along the small intestine. This enhancement is facilitated by the increase in villus height, leading to a notable reduction in cecal coliform populations while fostering the growth of beneficial bacteria such as *Lactobacillus* spp and inhibiting the growth of pathogenic bacterial species such as *Clostridium* spp. [9].

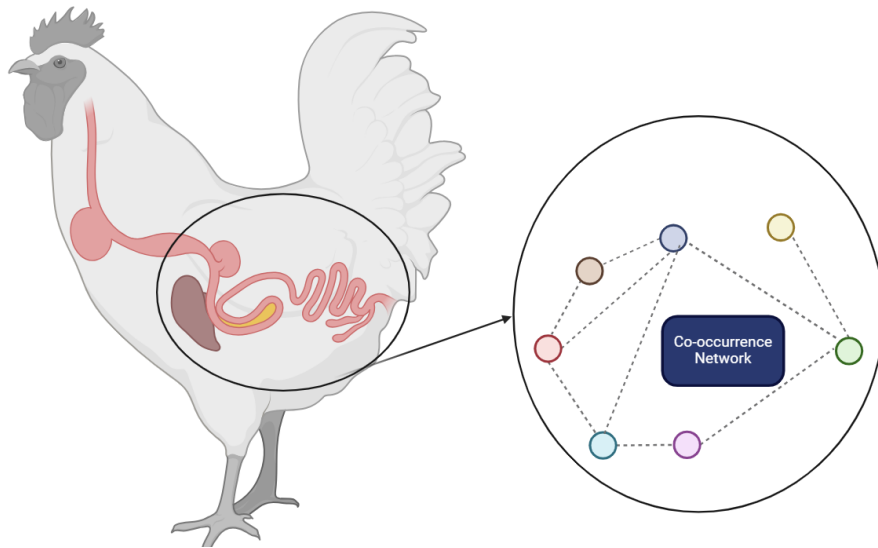


Figure 2. Gut of chickens modulate with a co-occurrence network In this picture the gut of chickens is modulated using a co-occurrence network in which each node represents a different species, and each arc indicates that these species co-occur.

These findings provide limited insights into the influence of growth-promoting antibiotics (AGP) and phytogenic feed additives (PFA) on the gut microbiota of chickens. Thousands of different microbes coexist in the gut and have complex and dynamic relationship between each other. Understanding these relationships is crucial to learn the impact and mechanisms of action of feed additives. However, previous studies have primarily focused on comparing the compositional levels of specific species, overlooking potential changes in microbial interactions [9, 12, 13]. Using network to model these interactions enhances our understanding, in particular co-occurrence networks are notably suitable for this analysis because they allow visualization of the relationships among different microorganisms present in the gut based on their co-occurrence. This approach enables the identification of patterns and interactions among microorganisms that may not emerge through other analytical techniques. Unfortunately, only a limited number of studies [16, 17] have utilized co-occurrence networks to assess the effects of AGPs and PFAs on the chicken gut. Furthermore, these studies have primarily conducted

basic measurements, such as calculating the number of connections or assessing modularity, to evaluate how the network responds to different treatments [18]. Modeling the chicken gut using networks can provide valuable information on the crucial nodes for the stability of the network itself, especially in relation to the effects of antibiotic growth promoters (AGP) and phytogenic feed additives (PFA). These nodes can also be used to distinguish between different treatments, identifying the species that are most affected by these treatments, defined as biomarkers, are measurable indicators of biological, pathological, or physiological processes that can be used to assess an individual's health status, diagnose diseases, monitor disease progression, or evaluate treatment responses [19]. However, so far, this type of analysis has been primarily conducted on humans [20, 21] and are not widely used to describe how different growth promoter treatments affect the gut of chickens.

As previously mentioned, the use of co-occurrence networks will visually depict the relationships among microbial species in various experimental contexts. Specifically, the connections between species in the network will be determined by correlation analysis, which assesses the tendency of species to co-occur across different treatments. Co-occurrence networks will enable the visualization and quantification of interactions between diverse microbial species, shedding light on how treatments may influence the structure and dynamics of the intestinal microbiota. So far, co-occurrence networks were mainly used to model changes in species interactions in soil, influenced by environmental conditions [22–27]. Previous research also applied co-occurrence network analysis to understand full-scale anaerobic digestion systems, highlighting the impact of operational factors on the microbial ecosystem [28]. Additionally, these networks were utilized to examine how they alter their structure in the human intestine by comparing networks constructed from samples of healthy individuals with those from individuals with intestinal diseases [3].

In this study, we will investigate how growth-promoting antibiotics (AGPs) and phytogenic feed additives (PFAs) affect microbial interactions in the intestines of chickens. Our hypothesis posits that AGP and PFA treatments will alter microbial interactions, leading to changes in the stability and functionality of the intestinal microbiota in distinct and measurable ways. To test this hypothesis, we will analyze microbial co-occurrence networks constructed from species present in different treatments. A more comprehensive network analysis will be conducted, encompassing metrics such as modularity, clustering coefficient, and average degree. These metrics will help us understand the initial differences between the networks of interactions among microbial species present in the intestines of chickens treated with AGP and those treated with PFA. Additionally, we will evaluate if our networks exhibit a modular structure and aim to identify consistent modules across various treatments. Furthermore, we will examine the types of species present within these modules to attribute biological significance to them and endeavor to understand the reasons why they may resist these treatments. Furthermore, we are interested in identifying "keystone taxa," species fundamental in supporting the stability of the microbiota, whose presence or absence can have disproportionate effects on the microbial ecosystem. We will aim to verify if these keystone taxa are useful in distinguishing between different treatments, this method can help to

gain a deeper understanding of the mechanisms by which AGP and PFA affect intestinal health, providing practical guidance for the application of these additives in poultry feeding programs. We will use a similar approach employed in two different studies, [29] and [30]. In the first study, Random Forest analysis highlighted the significant impact of key microbial taxa on soil microbiota. In the second study, the aim was to identify species that distinguish samples of monkeys fed high-fat or low-fiber diets, also through the same classification algorithm.

2 Releted Works

In this section, we will describe in detail the tools that we will use in our project.

Challengers for construct co-occurrence network

Data preparation for constructing a co-occurrence network poses several challenges that require careful consideration.

Effectively addressing these obstacles will ensure an accurate and meaningful representation of species relationships within the studied microbiota:

Varying sequencing depth

When DNA from biological samples undergoes sequencing, we obtain representations of the DNA fragments present in those samples [31]. However, the number of these sequences can vary between samples due to factors such as the quantity and quality of the genetic material extracted, the efficiency of preparation processes, and sequencing conditions. This variation in sequence numbers across samples does not always reflect the true biological diversity of the samples. Some samples might have more sequences simply because they were sequenced more efficiently, not because they contain a greater diversity of organisms [31].

This phenomenon can create apparent or false correlations between different entities (such as bacterial species) that are not biologically associated. Consequently, it could lead to erroneous interpretations in subsequent analyses. Differences in microbial composition between samples with varying sequencing depths might be mistakenly interpreted as genuine biological variations [31].

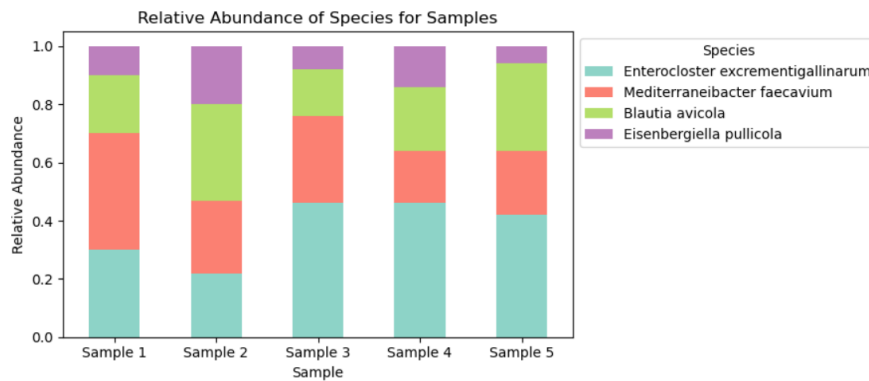


Figure 3. Relative Abundance. Graph of species distribution after calculation of relative abundance

To address the challenges posed by varying sequencing depths, techniques such as rarefaction or normalization can be employed. It is possible to use relative abundance to tackle the issue of varying sequencing depths. Relative abundance refers to the proportion of a particular species or group of organisms within a community relative to the total of the community itself.

When calculating relative abundances, data is normalized so that each species is represented proportionally to the total number of sequences in the sample. This process ensures that species are evaluated based on their relative abundance compared to others, regardless of the total number of sequences in each sample. By using relative abundances, we can more accurately assess the composition of microbial communities, while also reducing the impact of sequencing depth variations and allowing for a more precise comparison between different samples.

Rare Taxa

Microbial communities, with their hundreds or thousands of operational taxonomic units (OTUs), often include a large proportion of rare taxa. This composition can pose methodological challenges, as demonstrated by simulations highlighting issues in methods to detect associations between OTUs, which are believed to reflect interactions among them [31].

A significant portion of sequencing data comprises zeros. In ecological count data, a zero might indicate either a genuine absence or a presence below the detection threshold, meaning the taxon was present, but its DNA was not included in the count table [32].





	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
	0	5	2	0	3	0
	0	0	0	2	0	0
	58	56	45	129	81	18
	0	2	0	175	0	0

Figure 4. Rare Taxa [32]. For the construction of our network we only considered species that exceed a certain threshold, in our case we consider rare taxa the taxa that have an abundance less of 0.001%.

To address the problem of composition data, it is possible to apply a filtering and transformation strategy. Firstly, we filtered out taxa whose abundance level was greater than a certain value, thus eliminating taxa that are only present in minute and rare quantities. This filtering allows us to focus on the most significant taxa relevant to the analysis, reducing the impact of taxa with sporadic or irrelevant presence. Subsequently,




	S1	S2	S3	S4	S5	S6
	0.0	0.007	0.6	0.008	0.03	0.1
	0.002	0.04	0.0	0.1	0.005	0.4
	0.00001	0.0	0.00002	0.0	0.0	0.0

Figure 5. Remove Rare Taxa [32]

Once these rare taxa were selected, it is possible to not to consider them and therefore to remove them from our analysis.

it is possible to apply the *Centered Log-Ratio (CLR)* to further improve the distribution of compositional data. The choice of CLR was motivated by the need to handle the unique structure of compositional data, stabilizing variance and making the data more linear.

This transformation not only enhanced data quality but also made the interpretation of relationships between variables easier, resulting in clearer and more understandable analysis results [33].

The use of logarithmic ratio transformation is valuable in these situations as it takes into account the presence of zeros in our measurements. Completely ignoring data with zeros can lead to the loss of important information, while dealing with zeros without proper transformation can result in spurious or unrealistic associations between variables. The *CLR* transformation allowed us to retain significant information from these zero values, effectively managing their impact on the analysis. This ensured that relationships between variables were preserved meaningfully, avoiding misleading interpretations resulting from improper handling of zeros in compositional data [33].

The CLR transformation involves dividing each abundance value by the geometric mean of its sample and then taking the logarithm of this ratio. The geometric mean is the $n - th$ root of the product of all values in a sample, where n is the number of species.

Indirect Edge

Indirect edges in a microbial network represent associations that are not direct and do not indicate a direct interaction between two microbial taxa.

Instead, they are the result of an association that could be mediated by a third factor, such as an environmental factor or another taxon present in the environment.

These indirect links can complicate the interpretation of microbial networks because

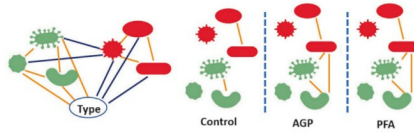


Figure 6. Indirect Edge [32]. To prevent the formation of indirect edges, we decided to create a co-occurrence network for each treatment.

they can make it appear that there are direct interactions when in fact they are mediated by other factors. For this reason, it is important to consider the environmental context and other variables when analyzing microbial networks to distinguish between direct and indirect interactions between taxa [32].

In this study, to address this challenge, we considered the environmental factor, and we calculated a separate network for each treatment as shown in Figure[6].

Scoring

In this study, we evaluated how each species relates to all others by calculating correlations. To determine these relationships, we used **Spearman’s correlation**, a method that tells us whether two variables are related and how strongly. A correlation coefficient close to 1 indicates that the two variables tend to vary together, a coefficient close to -1 indicates that they move in opposite directions, while a coefficient close to 0 suggests no relationship.

La formula utilizzata per il calcolo del coefficient di Sperman è la seguente

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

Where:

d_i are the differences in the ranks for each pair of species

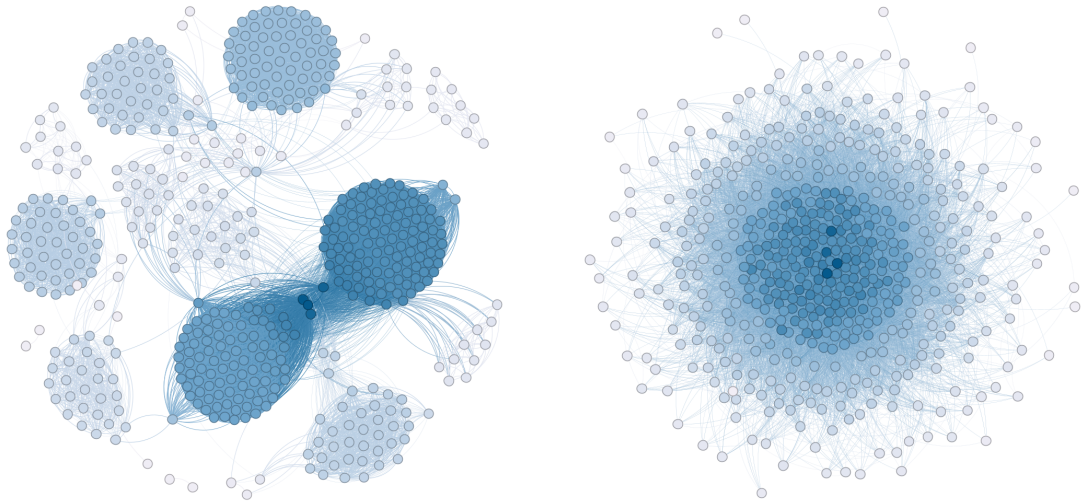
n is the total number of samples

Null Model

Microbial networks, which describe interactions within microbiomes, are typically constructed from statistically inferred connections. However, they may not always accurately represent genuine ecological interactions. As a result, microbial association networks are prone to errors and may not faithfully portray the true community structure [34].

This means that to give significance to the features obtained from our analyses, we compare the results derived from the observed networks with those obtained using a null model. We compare real data with data generated under the null hypothesis that all associations are random, allowing us to identify non-random patterns in groups of association networks.

We will compare the results obtained for modularity and cluster coefficient with those from 100 randomly generated networks, in which edges are exchanged instead of being



AGP at Day 21 observed network

AGP at Day 21 Random Network

Figure 7. Observed Network and Random Network On the left is the observed co-occurrence network of species found on day 21 in samples treated with AGP. On the right is the null model created using the Brain Connectivity tool keeping the number of nodes, edges, and the node degree distribution the same as an observed network. The color of the two networks indicates the distribution of node degrees.

removed and added back. For instance, if there are edges (a, b) and (c, d), they become the new edges (a, c) and (b, d). This ensures that the model maintains the degree distribution observed in the original network, with each node retaining the same degree as in the initial network [34].

However, certain centralities like betweenness centrality may vary, and we will do the same with the identification of the core adding a comparison also between the observed network and a totally random network obtained with edges randomly added until the total number of edges matches that of the input network [34].

Additionally, to highlight the significant differences, we will conduct a z-test to provide statistical significance to the obtained results.

Analyze the presence of a Core Association Network (CAN)

Modularity analysis not only unveils how this property adapts to environmental changes but also provides an opportunity to identify resilient modules (conserved subsets known as core association networks (CAN) [34]) capable of maintaining their integrity despite surrounding variations. This implies that we can better understand not only how species interactions change in response to alterations but also which groups of species maintain relative stability over time.

To identify the core association networks (CAN), we will use the *Anuran* [34] tool, which identifies overlapping edges across all networks considering all treatments and

ages. This tool compares the results obtained with 40 random networks, where the distribution of node degrees is preserved, and with another 40 random networks created with the same nodes as the original network [34]. This way, we can determine if the associations identified in the CAN are significantly different from those one would expect to obtain randomly.

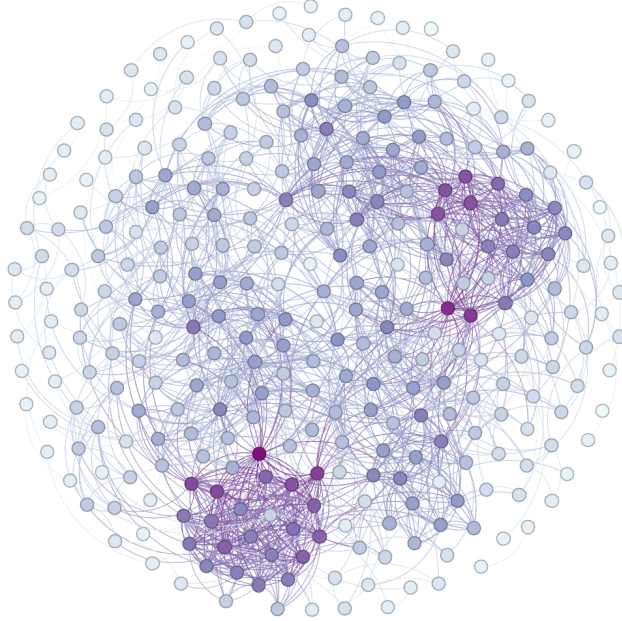


Figure 8. Core Example. Example of a core obtained via Anuran, present in co-occurrence networks constructed considering the species present in the PFA-treated samples. The color of the nodes corresponds to the degree within the core.

Anuran

It is possible to find the toolbox at the following link <https://github.com/ramellose/anuran>.

This software uses a set-based method to find core association networks (CAN), which are essential networks of connections. A set is a specific group of links, like the intersection set, containing links found in multiple networks. CANs are found by comparing differences between specific intersection sets. This software detects sets of links present only in certain network subsets, distinguishing less conserved from more conserved links. If we are going to consider four networks the tools pipeline is reported in the following image [34]:

The model uses a Venn diagram to identify intersections and differences between different networks [35]. This approach helps understand how the networks overlap and differ from each other. For example, it can find intersections that represent networks present

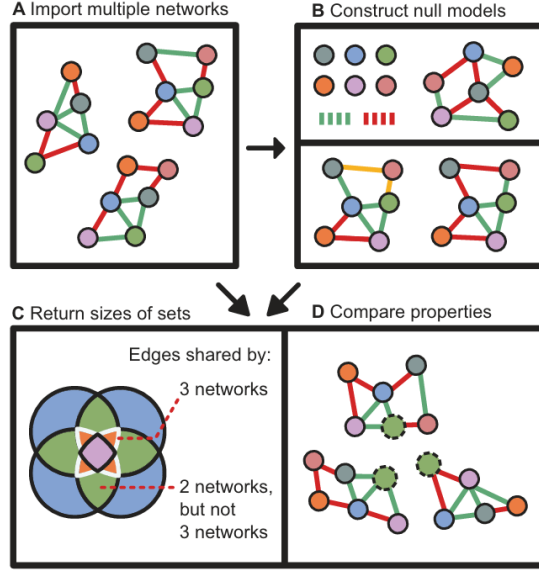


Figure 9. Anuran. The pipeline of Anuran tools consist in: **A** Import multiple networks, each representing microbial taxa depicted by node colors and interactions indicated by edge colors (red for negative and green for positive weights). **B** Random networks are generated for each imported network. **C** The toolbox returns various types of sets represented by a Venn diagram, illustrating the overlap between specific numbers of networks. Each color in the diagram corresponds to a set returned by the toolbox. **D** We compare the results to the null model to give significance to the results that we obtain. [34]

under certain conditions, such as 50% of the networks or less.

To calculate the differences between the networks, the model finds the edges (connections) unique to each network. The difference between the networks is the union of all sets D_i . The formula is:

$$\text{Difference} = D_i = \bigcup_{i=1}^n D_i$$

where D_i is the set of edges present in E_i but not in the other edge sets. The formula for D_i is:

$$D_i = \left\{ x : x \in E_i, x \notin \bigcup_{j=1, j \neq i}^n E_j \right\}$$

To calculate the k-intersections, the model analyzes which edges are shared among groups of networks [35]. The k-intersections represent the edges common to groups of k networks. To find the k-intersections, we calculate the intersections S_I of the groups I of k networks. The size of all possible groups is given by the binomial coefficient:

$$P_{n,k} = \binom{n}{k}$$

For example, for 40 networks and $k = 4$, the formula is:

$$\binom{40}{4}$$

For example, if you have three networks you can find the connections unique to each network and the connections common to all three networks or combinations of two of them. This model helps to better understand the relationships between networks, distinguishing between common and less common associations and provides a detailed analysis of the unique and shared connections among them [35]. The concept of keystone taxa (species), initially introduced by Paine in 1969, refers to native species that exert a significant influence on the stability of their ecosystem [36]. Over time, identifying keystone species has become a crucial component of ecosystem analysis, essential for comprehending ecosystem vulnerabilities and ensuring sustainability [36]. These species often play critical roles in maintaining the structure, diversity, and function of their ecosystems, making their identification and preservation paramount.

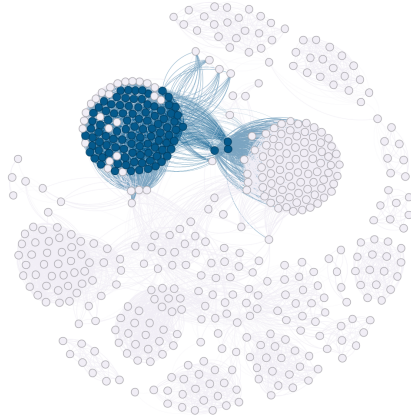


Figure 10. Keystone Taxa Example of keystone taxa distribution in a network. In particular, these species were identified as keystone taxa through a network analysis of the Co-occurrence network constructed by considering the species present in the PFA-treated samples at Day 21

In addition to examining co-occurrence patterns, microbial networks offer a statistical method for identifying keystone taxa. Microbial networks provide insights into the complex interactions between different species within a community, allowing researchers to pinpoint those taxa that have the most significant impact on network stability and functionality.

Utilizing network analysis provides a robust approach to deducing keystone taxa within microbial communities [37]. This method involves analyzing various centrality measures to understand the relative importance of each node (species) in the network. Previous studies have often relied on high betweenness centrality for the statistical identification of keystone taxa, recent research suggests that a combination of high mean degree, high closeness centrality, and low betweenness centrality can accurately identify keystone taxa with an 85% success rate [38]. This multi-metric approach reflects a more comprehensive understanding of network dynamics, considering not only the intermediary role of species but also their connectivity and influence across the entire network.

By integrating these analytical techniques, researchers can more effectively identify and target keystone species for conservation and management efforts. This is particularly important in microbial ecosystems, where keystone taxa can drive critical processes such as nutrient cycling, pathogen suppression, and ecosystem resilience. Ensuring the stability and functionality of these ecosystems is vital for broader environmental health and sustainability.

Random Forest

In our research, we aim to evaluate the significance of the keystone taxa identified via ecological network analysis. These keystone taxa, vital for the structure and function of the studied ecosystem, have emerged as critical elements in maintaining balance and stability in the environment. We aimed to use these keystone taxa to differentiate between various environmental treatments, referred to as AGP, CTR, and PFA.

To achieve this, we will employ the **Random Forest** algorithm.

The Random Forest is a powerful classification algorithm in the field of machine learning. It relies on a group of decision trees known as "random trees". These trees are trained on random subsets of the data and use a voting principle to make decisions, reducing the risk of overfitting and improving model generalization through ensemble learning. This approach makes the Random Forest particularly suitable for analyzing count data. Since each decision tree within the algorithm is trained on a random subset of the data and uses a random subset of features at each split.

By applying the Random Forest, we selected the species that contribute most to the classification of treatments, enabling us to identify species that play a significant role in ecological differentiation among treatments.

The pipeline of our work using the random forest will be as follows:

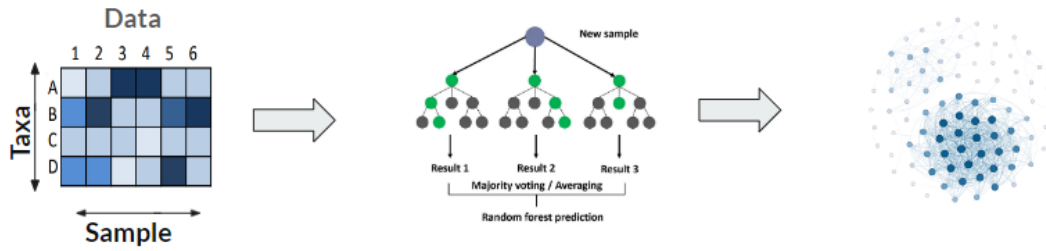


Figure 11. Pipeline Random Forest, [39]. We will begin by training our Random Forest model using the keystone abundance species found using the combination of high degree centrality high closeness centrality and low betweenness centrality, where the treatment associated with each sample will be our target to predict. Next, we will use the permutation method to identify the species that are most important in distinguishing between the different treatments. This process involves evaluating the importance of features by measuring their variation in the model’s results when the features are randomly permuted. The species that have a significant impact on the model’s ability to distinguish treatments will be identified through this process. These species that emerge as important for the classification model also have a significant role in the structure and function of the ecosystem, as identified through network analysis.

What we are going to do is an in-depth analysis of the key species for each treatment through network analysis.

First, we will calculate the keystone taxa using network analysis, identifying the species that are fundamental for the stability and robustness of the ecological network for each treatment.

Next, we will build a Random Forest model using the species abundance matrix, after a feature selection using mutual information considering the feature that explains the 90% of variance, and apply nested cross-validation to make the model more robust and reduce the risk of overfitting.

Permutation importance [40, 41] measures the importance of a feature by assessing how much the model’s prediction error increases after permuting that feature. If permuting the values of a feature causes an increase in the model’s error, then that feature is considered *important* because the model relied on it to make its predictions. On the other hand, if permuting the values of a feature does not change the model error, then that feature is considered *not important* because the model did not use it to make the predictions.

Mutual information (MI) serves as an indicator of statistical independence and has two fundamental characteristics. Firstly, it is capable of assessing different relationships between random variables, including non-linear associations. Furthermore, MI is unaffected by alterations in the feature space that maintain invertibility and differentiability, such as translations, rotations, and transformations that maintain the original order of the elements of the feature vectors [42].

We chose to use mutual information to identify the most relevant features in our dataset.

This allowed us to select features that accounted for 90% of the variance in the dataset, thus reducing the dimensionality of our dataset. The aim was to retain only the most informative features, improving the effectiveness and accuracy of the Random Forest classification model applied subsequently.

The Random Forest model will be a binary classification model, and our goal is to identify the species that help us distinguish between different treatments, determining whether a sample has been treated with AGP (antibiotic growth promoters), PFA (phytogenic feed additives), or the Control (CTR). To do this, we will rename the treatments in three distinct ways: we will set the AGP treatment to 1 and the PFA treatments to 0, then we will set the PFA treatment to 1 and the CTR treatments to 0, and finally, we will set the CTR treatment to 1 and the AGP treatments to 0.

The metric selected to assess the effectiveness of classification in the present study, where the dataset is balanced, is accuracy. This metric represents the fraction of correct predictions out of the total number of predictions made by the model on a test dataset. The accuracy is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + F} \quad (2)$$

where:

- TP (True Positives) represents the number of instances correctly classified as positive by the model.
- TN (True Negatives) denotes the number of instances correctly classified as negative by the model.
- FP (False Positives) denotes the number of instances incorrectly classified as positive by the model.
- FN (False Negatives) denotes the number of instances incorrectly classified as negative by the model.

Once we have obtained the influential species from each of the three classifications, we will select the common species among the Random forest classification and the first 100 keystone taxa identified through network analysis for each treatment. These common species will be those that help us to distinguish between the various treatments.

Biomarker

The utilization of biomarkers in both basic and clinical research, as well as in clinical settings, has proliferated to the extent that their inclusion as primary endpoints in clinical trials is now largely unquestioned [43]. For specific biomarkers that have been well-characterized and repeatedly shown to accurately predict relevant clinical outcomes across various treatments and populations, this use is entirely justified and appropriate [43]. The term "biomarker," derived from the combination of the words "biological marker," refers to a broad category of medical signs. These are objective indications of a patient's health state, observable from the outside, and can be measured accurately and reproducibly. Medical signs are distinct from symptoms, which are subjective perceptions of health or illness reported by the patients themselves. In the scientific literature, there are various specific definitions of biomarkers that, fortunately, largely overlap [43]. In 1998, the National Institutes of Health Biomarkers Definitions Working Group defined a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." A joint initiative on chemical safety, the International Programme on Chemical Safety, led by the World Health Organization (WHO) in collaboration with the United Nations and the International Labour Organization, defines a biomarker as "any substance, structure, or process that can be measured in the body or its products and that can influence or predict the incidence of outcomes or diseases" [43].

An even broader definition considers not only the incidence and outcomes of diseases but also the effects of treatments, interventions, and unintended environmental exposures, such as those to chemicals or nutrients. In their report on the validity of biomarkers in environmental risk assessment, the WHO stated that a true definition of a biomarker includes "almost any measurement reflecting an interaction between a biological system and a potential hazard, which may be chemical, physical, or biological [44]. The measured response can be functional and physiological, biochemical at the cellular level, or a molecular interaction" [43].

Biomarkers can be identified using features from classification algorithms. This approach is effective because classification algorithms are designed to identify the most relevant features that contribute to distinguishing between different classes. These features represent crucial indicators and potential biomarkers as they improve the accuracy of the predictive model.

3 Material and methods

3.1 Study subjects and sample collection

At the start of the experiment, 96 male broiler chickens [45] were randomly assigned to twelve pens with eight birds each. These twelve pens were then randomly assigned to one of three following feeding groups, each with four replicating pens.

- Control (basal diet)
- AGP (basal diet + 150mg/kg feed of ALBAC)
- PFA (basal diet + 150mg/kg feed of Digestarom®)

All birds received the same basal diet until day 3. At day 3 (right before the start of feed additive administration), day 14, day 21 and day 35, two birds per pen were euthanized as shown in the Figure[12]. Their cecum digesta and mucosa microbiota samples were taken and pooled as a replicate for metagenomic sequencing.

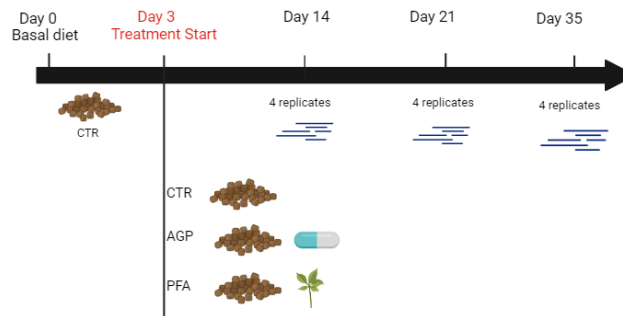


Figure 12. Chicken’s treatment. The figure describes when the treatment starts. Before day 3 all the chickens received the same treatment after that they were divided into three treatment groups and on day 14 microbiota samples were taken for each treatment group, this process was repeated also for days 21 and 35.

3.2 Construct co-occurrence network

In this work, the following pipeline was followed for the construction of the co-occurrence network: preparing the input matrix from taxonomic classification, scoring the inter-species correlation, assessment of the significance and network visualization. The detailed description of each step can be found below (in the Figure[13]).

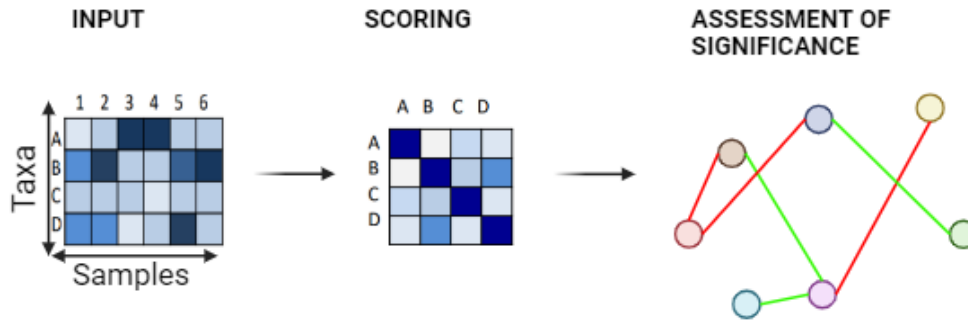


Figure 13. Co-occurrence networks pipeline. The pipeline consists of three main stages: **Input:** data preprocessing to address challenges (such as Varying sequencing depth, Rare taxa, and Indirect edges) present after obtaining the abundance matrix; **Scoring:** calculation of the correlation matrix; **Assessment of significance:** filtering of the data to include only significant relationships (means only significant edges), i.e., if the p-value is less than 0.05

Input

Data preparation for constructing a co-occurrence network poses several challenges that require careful consideration. Effectively addressing these obstacles will ensure an accurate and meaningful representation of species relationships within the studied microbiota:

- **Varying sequencing depth:** Differences in sequencing depth among samples can lead to variations in results. It is essential to address this challenge to ensure that analyses are consistent and not influenced by external factors. To overcome the challenges arising from varying sequencing depths, methods like rarefaction or normalization can be utilized. In this particular study, we will employ relative abundance to address the issue of varying sequencing depths. Relative abundance indicates the proportion of a specific species or group of organisms within a community in relation to the total community size.
- **Rare taxa:** The presence of species with very low relative abundance can be problematic as it may contribute to noise in the data. Removing these species allows focusing on the most relevant and representative species within the dataset.

To address the problem of composition data, we decided to apply a filtering and transformation strategy. Firstly, we filtered out taxa whose abundance level was greater than 0.01%, thus eliminating taxa that are only present less quantities. This filtering allows us to focus on the most significant taxa relevant to the analysis, reducing the impact of taxa with sporadic or irrelevant presence. Subsequently, we applied the *Centered Log-Ratio (CLR)* to further improve the distribution of compositional data. The choice of CLR was motivated by the need to handle the unique structure of compositional data, stabilizing variance and making the data more linear. This transformation not only enhanced data quality but also made the interpretation of relationships between variables easier, resulting in clearer and more understandable analysis results [33].

- **Indirect edges:**

In addition to direct connections between species, there are also indirect connections that can complicate network interpretation. Identifying and managing these indirect connections is crucial for an accurate understanding of species relationships. In this study, to address this challenge, we considered the environmental factor, we calculated a separate network for each treatment and we added two extra nodes that represent the different types digesta and mucosa.

Scoring

In this study, we evaluated how each species relates to all others by calculating correlations. To determine these relationships, we used Spearman’s correlation, a method that tells us whether two variables are related and how strongly. A correlation coefficient close to 1 indicates that the two variables tend to vary together, a coefficient close to -1 indicates that they move in opposite directions, while a coefficient close to 0 suggests no relationship.

The formula used for calculating Spearman’s coefficient is as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

Where d_i are the differences in the ranks for each pair of species and; n is the total number of samples

Assessment of significance

It is crucial to note that it is necessary to carefully select statistically significant correlations. This selection process is essential to ensure that the resulting network is informative and truly reflects the ecological dynamics at play.

If we did not only select statistically significant correlations, we would end up with a fully connected network. This type of network, although rich in connections, could be uninformative and would not provide the necessary insight into the specific relationships between species.

To select the statistically significant connections, or edges, as described in [27] using Spearman correlation, we performed permutation and bootstrap procedures with 1000 iterations each. Taxon abundances were shuffled for permutations while resampling from samples with replacements was conducted for bootstrapping. The resulting p-value was then derived as the probability of the null value under a Gaussian curve fitted to the mean and standard deviation of the bootstrap distribution. Renormalization was applied to Spearman correlation permutations to address compositionality bias, followed by multiple-testing correction using the Benjamini-Hochberg method [46]. Lastly, edges with an adjusted p-value above 0.05 were excluded.

Network analysis

Topological analysis, including metrics such as modularity, cluster coefficient, number of nodes, and number of edges, provides us with initial tools to compare our networks and understand the initial differences or similarities observed among the networks constructed from samples treated with AGP and PFA. Our approach will involve comparing the results obtained from networks constructed using only samples treated as Controls (CTR) with those constructed using samples treated with AGP and PFA. All topological measurements and module calculations were performed using a dedicated tool **Brain Connectivity Tool** [47].

For graph visualization, we used **Gephi** software, widely employed for analyzing and visually representing networks.

Topological analysis and core identification

We will begin by examining the main differences between the various networks in terms of **clustering coefficient**, **modularity**, and **average degree (average degree)**. These topological parameters will help us understand how the connectivity and modular structure of the networks vary following different treatments.

Mean degree This metric employed in network analysis assesses the typical level of connectivity among nodes within the network. In a co-occurrence network or any other type of network, a node's "degree" denotes the number of connections it has with other nodes. The "mean degree" represents the average degree across all nodes in the network.

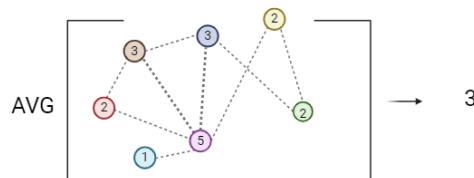


Figure 14. **Average degree.** This measure is calculated computing the average of the degree of all nodes in the networks

Mean clustering coefficient The clustering coefficient $CC(i)$ for a specific vertex i within a network is calculated as follows [48]:

$$CC(i) = \frac{1}{N} \sum_i \frac{\text{number of triangles connected to } i}{\text{number of possible triangles connected to } i} = \frac{1}{N} \sum_i \frac{2E_i}{k_i(k_i - 1)} \quad (4)$$

Where E_i is the number of triangles centered in vertex i and k_i is the degree of that vertex, N is the total number of nodes.

The clustering coefficient provides insight into the network's intricacy, showcasing the strength of interactions among microorganisms [49]. To prove that these results are statistically significant, we will perform a z-test between the value obtained from this computation and the clustering coefficient derived from 100 random networks (computed using *Brain Connectivity Tool* [47]) that maintain the same degree distribution as the observed network.

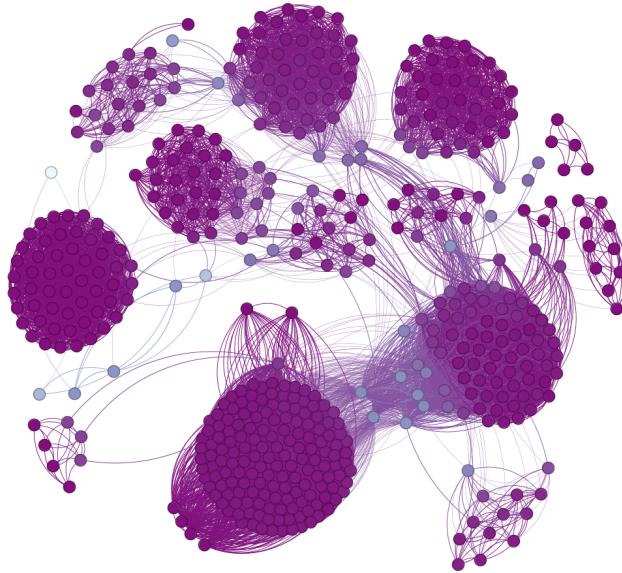


Figure 15. Clustering coefficient. Calculated for each node, considering the species present in the samples at day 21 and treated with PFA.

Network modularity The modularity involves grouping network nodes into distinct categories or subsets based on their connectivity patterns within the graph, rather than inherent node characteristics. We examined if the microbial co-occurrence networks exhibited modular properties by assessing their community structure through modularity maximization using a locally greedy algorithm, Louvain’s method [25].

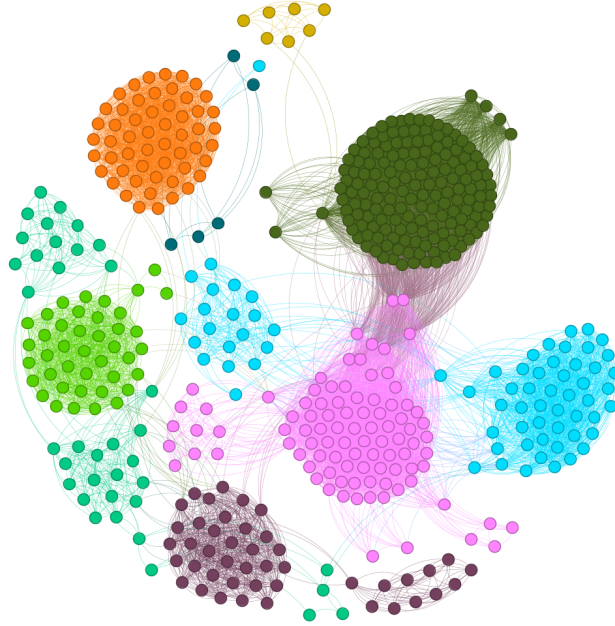


Figure 16. Modularity. Modularity of node clustering at day 21 for species present in the sample treated with PFA, after applying the Louvain algorithm 100 times and calculating the consensus partition.

This approach divides nodes into distinct groups without overlap and is based on a two-step heuristic:

1. To allocate nodes to modules (or communities), a greedy algorithm is employed to maximize the modularity index Q , which is defined as [25]:

$$Q = \frac{1}{2A} \sum_i \sum_j \left(A_{ij} - \gamma \frac{A_i A_j}{2A} * \delta(C_i, C_j) \right) \quad (5)$$

Where A_{ij} is the adjacency matrix, A_i is the degree of node i , A is the average of the degree of the network, C_i is the community to which node i is assigned, and γ is a scaling parameter used to adjust the size of the modules.

2. Some networks exhibit numerous local maxima, meaning there are many potential ways to divide the network with similar but not identical modularity values. This variability can complicate determining the optimal or most significant division.

To overcome this obstacle, modularity optimization is iterated multiple times (in this case, 1000 times). With each iteration, a new division of the network into communities is generated. The consensus partition is then derived by aggregating these diverse partitions, highlighting nodes consistently assigned to the same community across all iterations. This process stabilizes the identified community structure and mitigates the impact of randomness-induced variability.

We also will compare the modularity of the observed network with that of 100 random networks, each preserving the same degree distribution as the co-occurrence network under study by computing a z-test.

To highlight how the distribution of modules changes with different treatments, we will explore how the specific module assignment of the networks varies in response to different treatments, for each age. To achieve this, we will utilize the communities identified within the Control network as our reference. In other words, we will keep the assignment of nodes to modules of the Control network constant in both AGP and PFA for each age.

We will exclude the two nodes *digesta* and *mucosa* to make the graphs clearer and visualize the various modules obtained after applying the Louvain algorithm. This method allows us to verify if there are modules that survive changes in the environment, we define these modules *core*. To identify a core association network (CAN) that is statistically significant, we will use the *Anuran* [34] tool, which identifies overlapping edges across all networks considering all treatments and ages. Then, this tool compares the results obtained with 40 totally random networks, and with another 40 random networks created with the same degree distribution as the study networks [34].

To identify a potential common core shared by all networks, we focus on edge intersection that appearing in at least 80% of the networks and in at least 50% of the samples.

We will also measure their survival rate by calculating the Jaccard similarity. The Jaccard similarity is defined as the ratio between the number of species common to both sets (that is, those that survive both before and after treatment) and the total number of species present in at least one of the two sets (that is, those present before the treatment plus those added or remaining after the treatment). This measure allows us to quantify how much the composition of the Core modules remains stable or changes as a result of the applied treatments.

To understand why these species survive under different treatments, we will examine the scientific literature to explain the role and importance of these species. We will review previous studies and publications to determine the biological and ecological factors that contribute to their resilience and stability in various treatment contexts. This approach will help us better understand the characteristics that enable these species to maintain their presence and function despite changes in environmental conditions.

To enhance comprehension of the stability and robustness of networks treated with AGP and PFA compared to the Control group, and to ascertain how these treatments influence the stability of the gut microbial ecosystem and its response to changes, we conducted two types of experiments. First, we randomly removed 300 nodes: this helps us evaluate how well the network performs when random nodes are lost. Secondly, we

conducted an experiment to selectively eliminate 300 nodes from the co-occurrence network. These nodes were identified according to specific criteria: high degree centrality, low betweenness centrality, and high proximity centrality. This criterion aims to identify nodes considered influential in the ecological framework of the network. To evaluate the effect of these two experiments, we calculate the network:

Residual global efficiency: the global efficiency represents the average of the inverse of the shortest path lengths between all pairs of nodes in the graph and is inversely proportional to the characteristic path length. The residual global efficiency is the difference between the global efficiency before the removal of the nodes and after the removal of the nodes.

Residual edges will allow us to determine how the network structure withstands perturbations. The residual edge is the difference between the number of edges before the removal of the nodes and after the removal of the nodes.

3.3 Keystone taxa

In this section, our goal is to distinguish the key species that characterize each network and that may be crucial for differentiating the various treatments. To do this, we will identify the species that have the greatest influence within the network by calculating the so-called *keystone taxa* [38] as nodes that have a high degree of centrality and closeness centrality but low betweenness [38]:

Betweenness centrality measures the importance of a node in facilitating communication within the network. It is defined as the number of times a node acts as a bridge along the shortest path between two other nodes.

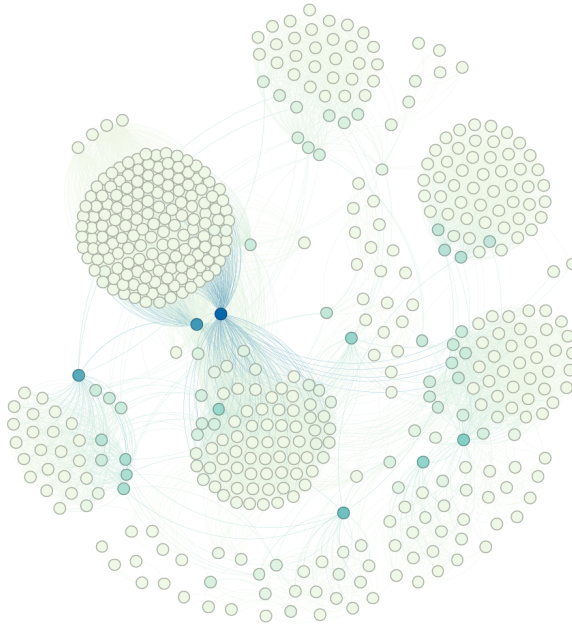


Figure 17. Between Centrality Calculated for each node, considering the species present in the samples at day 21 and treated with PFA. Lighter colors indicate a lower value, while darker values indicate a higher value.

Degree is the number of connections or edges it has with other nodes in the network. In other words, it represents the number of direct connections a node possesses. It is a fundamental measure to understand the local connectivity of a node within the network.



Figure 18. Degree Calculated for each node, considering the species present in the samples at day 21 and treated with PFA. Lighter colors indicate a lower value, while darker values indicate a higher value.

Closeness Centrality measures how close a node is to all other nodes in the network. It is defined as the inverse of the sum of the shortest path distances from the given node to all other nodes. A node with high closeness centrality can efficiently spread information throughout the network, being "close" to all other nodes.

The calculation of the keystone taxa will be performed using these criteria and we will obtain 9 distinct groups of keystone taxa, one for each combination of treatment and age group. Subsequently, the keystone taxa from the same treatment will be merged into a single set, resulting in 3 different groups. This method will allow the identification of species fundamental to the stability of the ecological network for each treatment. Next, a Random Forest model will be constructed using the species identify as keystone taxa, after feature selection using mutual information [42]. Features explaining 90% of the variance in the dataset will be considered, reducing the dimensionality of the dataset. The aim is to retain only the most informative features, thus improving the effectiveness and accuracy of the Random Forest classification model. Nested cross-validation will be applied to enhance the model's robustness and reduce the risk of overfitting. This will be followed by the use of permutation importance to select the species, which involves measuring the impact of randomly shuffling each species on the model's performance. This method allows us to identify which species have the most significant influence on the model's accuracy. The Random Forest model will be a binary classification model. Treatments will be renamed in three distinct ways: first, the AGP

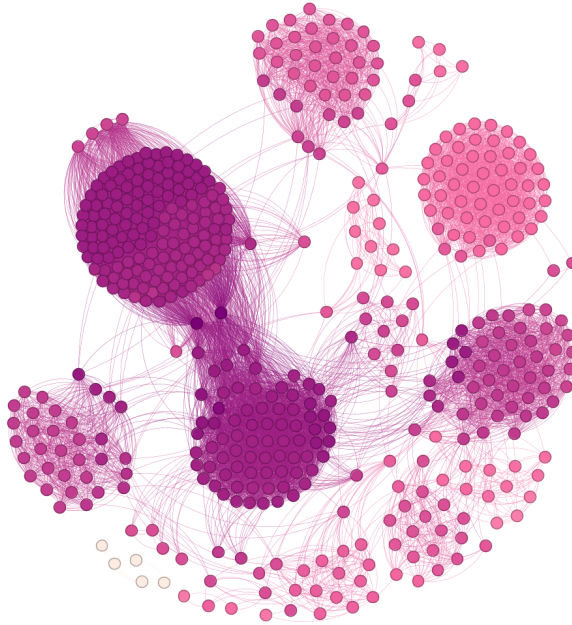


Figure 19. Closeness Centrality Calculated for each node, considering the species present in the samples at day 21 and treated with PFA. Lighter colors indicate a lower value, while darker values indicate a higher value.

treatment will be set to 1 and the PFA treatments to 0, then the PFA treatment will be set to 1 and the CTR treatments to 0, and finally, the CTR treatment will be set to 1 and the AGP treatments to 0. In this way, we will enable the model to identify the most relevant species for distinguishing between AGP, PFA, and CTR. Accuracy will be the selected metric to assess the effectiveness of classification in the present study, where the dataset will be balanced.

4 Results

4.1 Lower connectivity, higher modularity and higher clustering coefficient with AGP and PFA administration

After obtaining the co-occurrence networks computed for each treatment and age group, we proceeded to analyze them using topological measures. We discovered significant differences in connectivity, modularity and clustering coefficient among the three networks at different time points.

At all time points, we noted that when comparing with the networks from the Control group, PFA-treated and AGP-treated co-occurrence networks had significantly fewer connections and a lower average degree than other groups. In these results, we can also notice that the AGP-treated co-occurrence network showed more extreme results than

the one obtained with the PFA-treated network. Although interactions between species are reduced, the higher modularity observed in AGP- and PFA-treated co-occurrence shown in the Table[1] networks suggests a clear division into distinct groups of microorganisms, with robust interactions within each module. This is further supported by the high value of the clustering coefficient, indicating intensified microbial interactions within these groups.

Age	Treatment	Nodes	Edges	Modularity		AVG Clustering coefficient		AVG Degree
				Observed Networks	Random Networks	Observed Networks	Random Networks	
14	CTR	503	16798	0.4514 (>0.01)	0.0707 (>0.01)	0.889	0.4535 (0.003)	66.7913
	PFA	510	14697	0.6101 (>0.01)	0.0843 (>0.01)	0.907	0.3203 (0.002)	57.6353
	AGP	496	13168	0.655 (>0.01)	0.0856 (>0.01)	0.899	0.2601 (0.018)	53.0968
21	CTR	511	32201	0.216 (>0.01)	0.0398 (>0.01)	0.863	0.7729 (>0.01)	126.0313
	PFA	513	20337	0.540 (>0.01)	0.0698 (>0.01)	0.9242	0.2470 (>0.01)	79.2865
	AGP	520	16657	0.6849 (>0.01)	0.0853 (>0.01)	0.919	0.2470 (>0.01)	64.0654
35	CTR	531	36300	0.2155 (>0.01)	0.0397 (0.0014)	0.871	0.6893 (0.07)	136.7231
	PFA	517	22287	0.556 (>0.01)	0.0673 (>0.01)	0.903	0.3662 (>0.01)	86.2166
	AGP	520	20523	0.6179 (>0.01)	0.0739 (>0.01)	0.914	0.3097 (0.03)	78.9346

Table 1. Network analysis results from treatment groups across the time. For each constructed network, the average clustering coefficient (avg clustering coefficient), modularity, and the number of edges and nodes were listed. For modularity and clustering coefficient, standard deviations are shown in parentheses. Besides, we compared these observed values to those of random networks with preserved degree distribution to assess significance.

4.2 Higher robustness of the microbial community with AGP and PFA administration

Based on the results of modularity and cluster coefficient we wanted to test if these two values are related to robustness. Figure[20] panel **A** represents the change in connectivity (residual edges) and global efficiency of the graph before and after a perturbation, in this case, the random removal of 300 nodes. From the results, it is clear that the greatest loss of edges occurs in the graph constructed using Control data. This suggests that the co-occurrence network based on the Control data is more sensitive to node perturbation than the networks constructed using other data. Furthermore, it is observed that the lowest loss of edges occurs in the AGP data while an intermediate value is shown in the PFA data. Similarly, we observe the same trend for the removal of 300 target nodes in Figure[20] panel **B**, which includes nodes with a high degree of neighborhood centrality but low intermediary centrality. The greatest loss of edges is recorded in the network built using the Control data, while the lowest loss is observed in the AGP data. Regarding the residual global efficiency, this is consistently higher for Control data, while it is very similar for AGP and PFA.

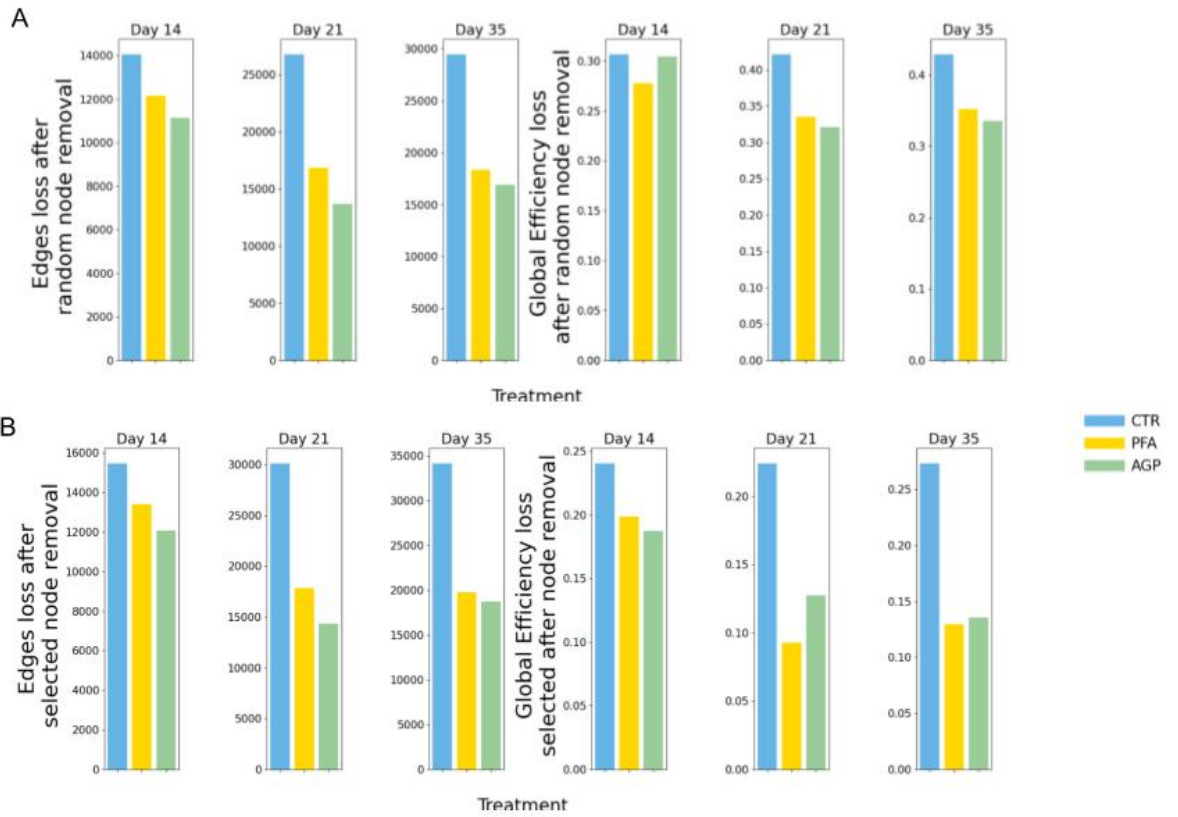


Figure 20. Residual edges and global efficiency after random and selected node removal. The bar plots illustrate the loss in terms of residual edges and global efficiency of networks subjected to the removal of 300 random nodes that are important for the stability of the networks, in panel A while in panel B the bar plots illustrate the same loss but with selected removal of 300 nodes. The x-axis represents the different treatments, while the y-axis shows the residual values. The first plot, located at the left, shows the residual edges after node removal, and the second plot, located at the right, depicts the residual global efficiency after node removal. The three plots are arranged side by side showing the results for the three different days (Day 14, Day 21, Day 35). Each bar represents one of three treatments.

4.3 Modular structure of co-occurrence networks and its change

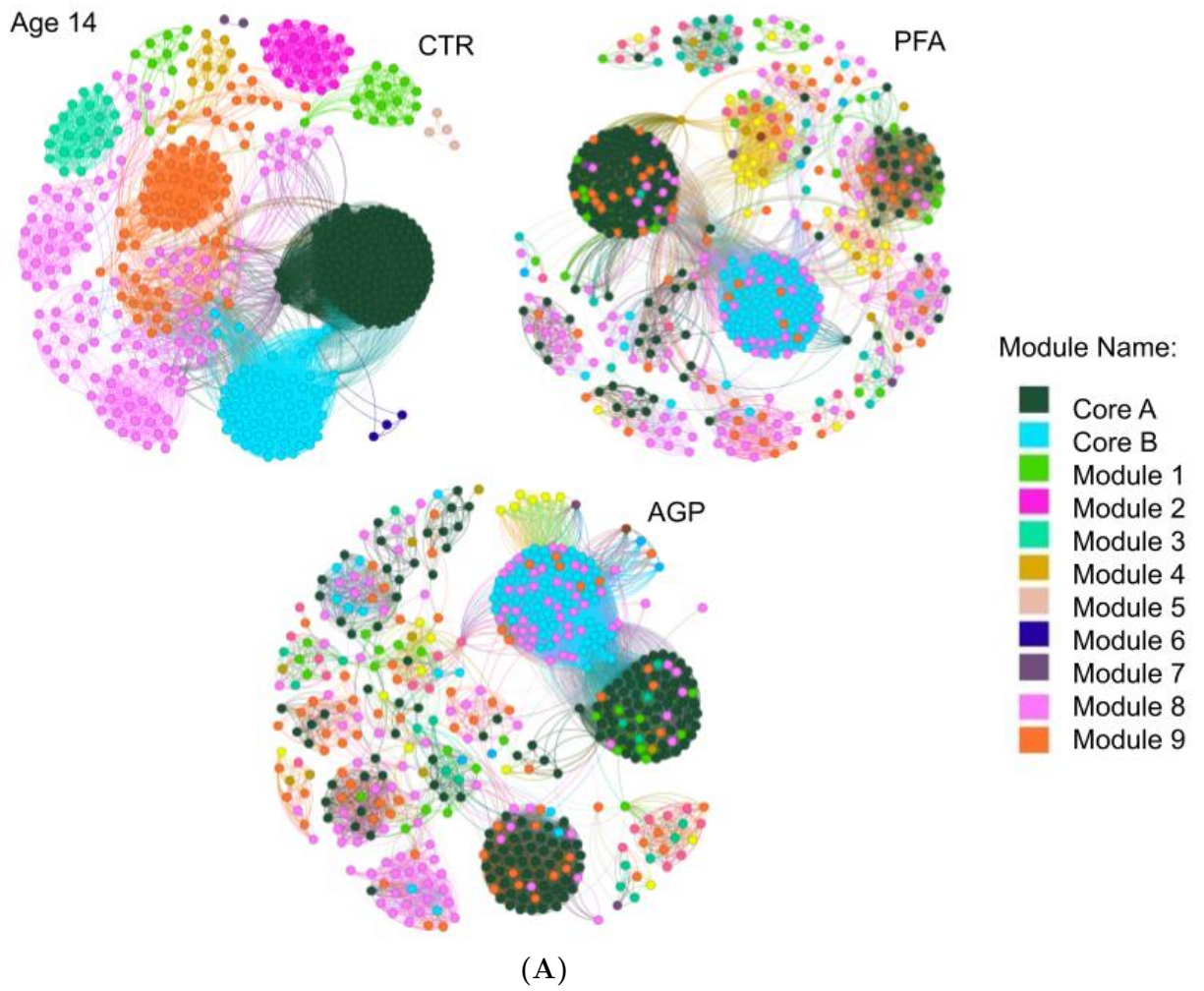
The predominant trend of module dispersion is evident from Figure [21] in the various age groups, the colors of the modules vary according to their modules name, the colors and names of the modules were assigned randomly. Except for the modules that we named *Core A* and *Core B*, which not only persist across all three figures, meaning they remain even when we change the treatment and their age, but also show more or less the same species. We kept their colors and names constant. While other modules, such as Module 1 at age 14 in the control group, disperse, suggesting that the species of Module 1 no longer co-occur together in the networks constructed considering the treated AGP and PFA samples.

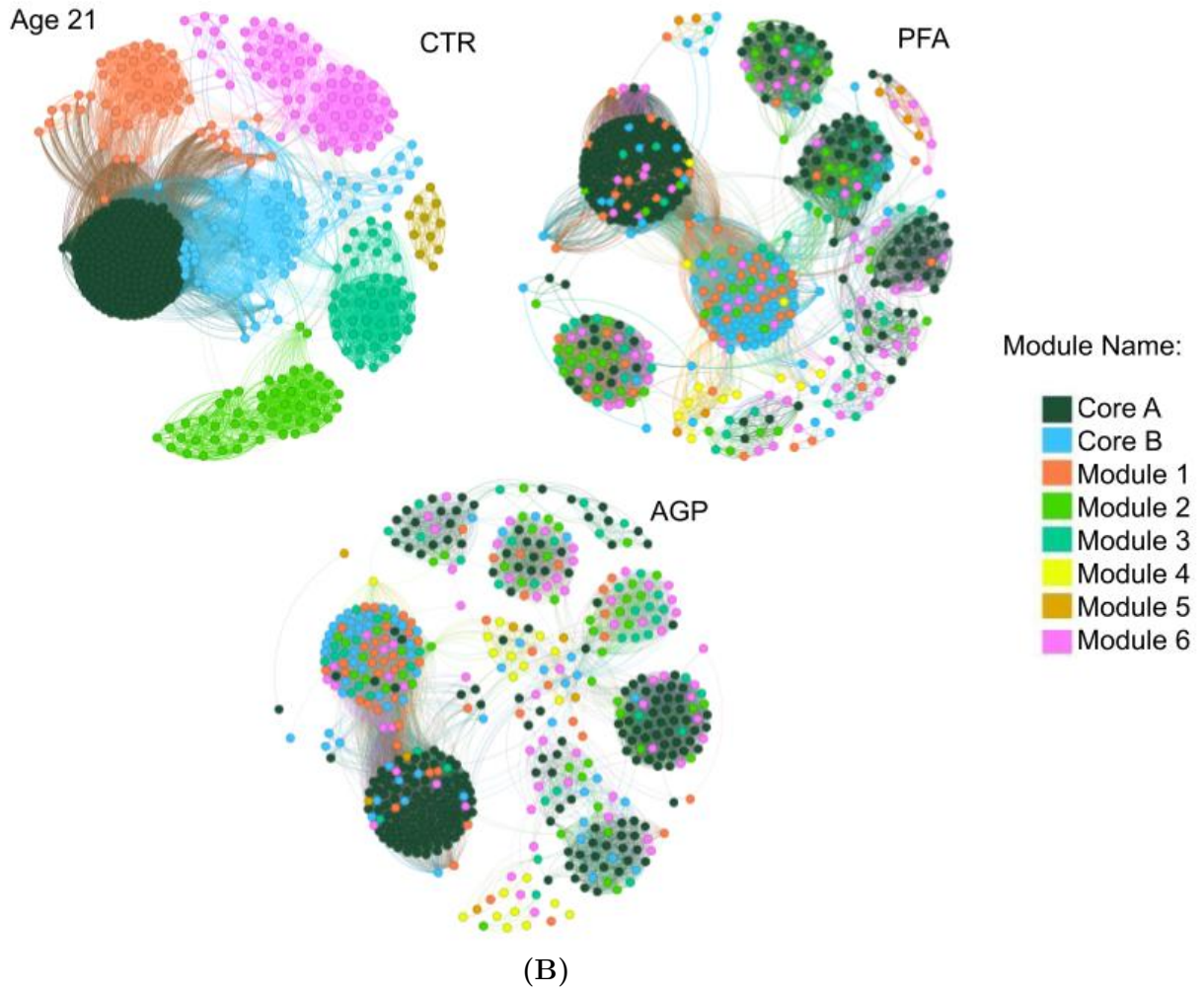
We measured the percentage of species surviving in the Core modules at varying treatments using Jaccard similarity.

AGE	Core A_PFA	Core B_PFA	Core_A_AGP	Core_B_AGP
14	49%	57%	53%	40%
21	49%	29%	51%	22%
35	60%	46%	56%	38%

Table 2. Jaccard similarity of Core A and Core B. This table displays the percentage of shared elements between Core A and Core B within the Control, AGP, and PFA networks. The first column presents the age data of the samples considered. The following two columns indicate the percentage of shared elements between Core A and Core B in the Control network, while the last two columns show the percentage of shared elements considering Core B in the PFA network and Core A in the AGP network.

During our examination of persistent modules, we observed that in certain networks, such as AGP at ages 14 and 21, and PFA at age 35 the Core A, is no longer confined within a single entity but instead divided into two separate components. Upon incorporating nodes digesta and mucosa into our network analysis, we noted that in Control samples, this module was exclusively present in digestawhile, in AGP (ages 14 and 21) and PFA (age 35) samples, it also appeared in mucosa, as illustrated in Figure [22]. The treatments with AGP and PFA seemed to impact this distribution, potentially facilitating bacterial colonization of the mucosa. This phenomenon might be attributed to changes in intestinal morphology or mucosal composition, creating a more conducive environment for these bacteria.





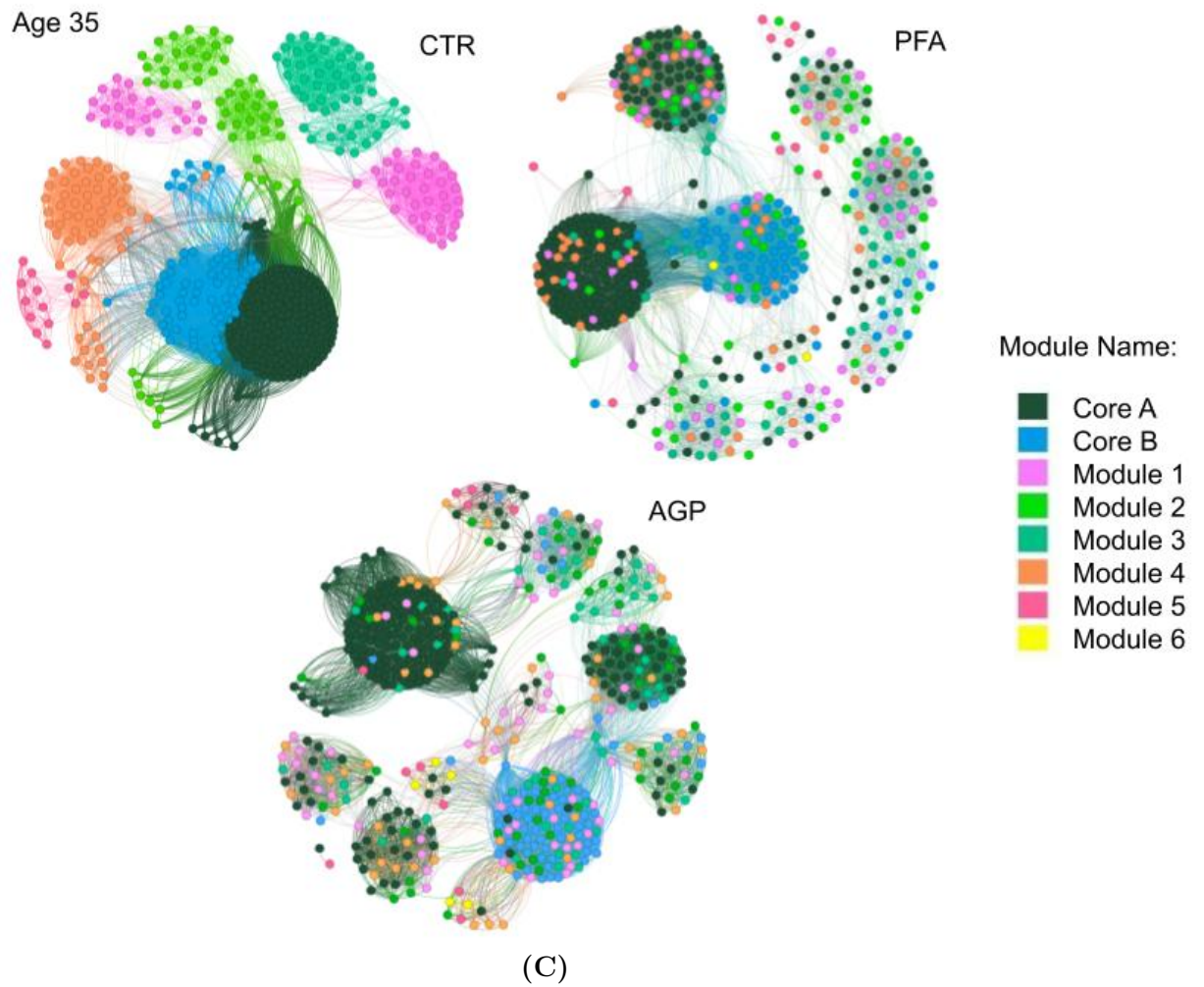


Figure 21. Comparison of graph modularity across different treatments on day 14, 21, 35. Nodes are colored according to the modules of the Control treatment (CTR) at day 14 in the panel A at day 21 in the panel B and at day 35 in the panel C, allowing for the assessment of how module structure varies under the influence of different treatments. This approach highlights the dynamic organization of relationships among entities within the studied system.

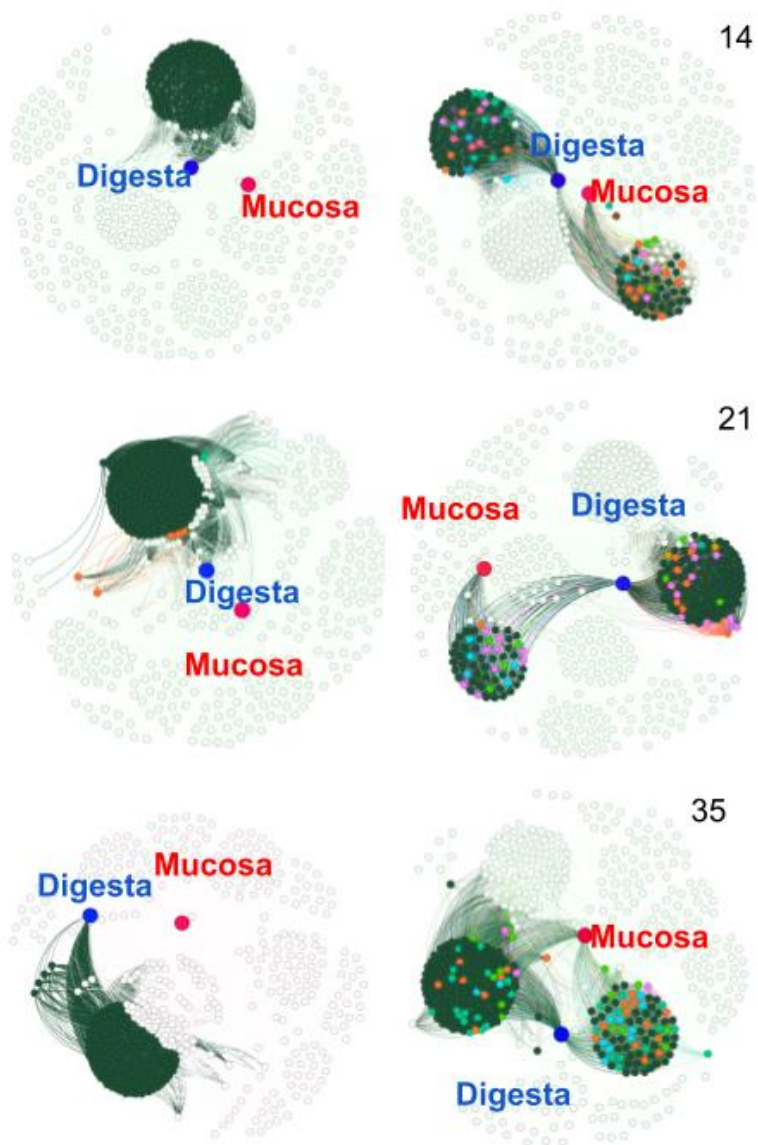


Figure 22. Distribution Green Module in digesta and mucosa under different treatments. The green module shows distinct distribution patterns in different treatments. In control samples, this module is exclusively present in Digesta. However, in AGP-treated samples at ages 14 and 21 days, and in PFA-treated samples at age 35 days, the green module splits and is present in both digesta and mucosa. This indicates that treatments with AGP and PFA alter the distribution, facilitating the colonization of the mucosa by these bacteria.

4.4 Statistically significant Core identification

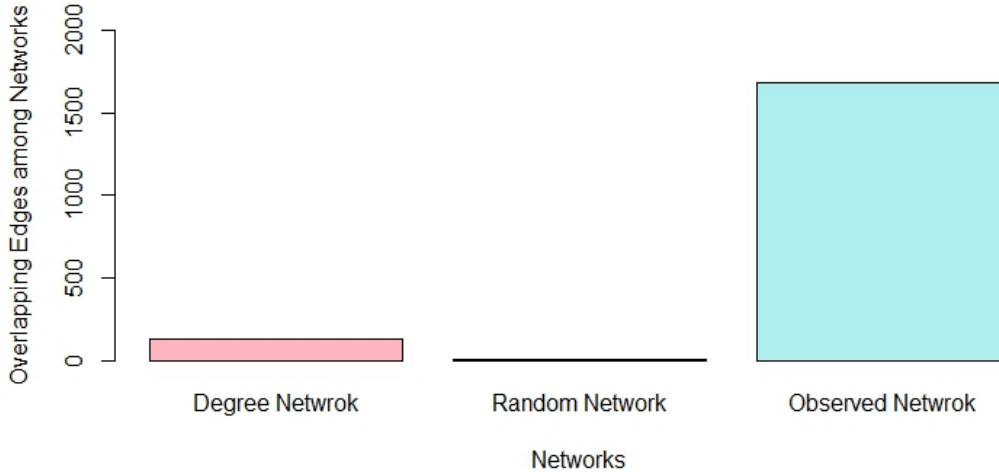


Figure 23. Edge intersection values appearing in at least 80% of the networks.
Degree Network: Represents the intersection of at least 40 networks whose degree, nodes, and arcs have the same distribution concerning the observed networks.
Random Network: Represents the intersection of at least 40 completely random networks maintaining the same nodes and arcs as the observed networks.
 The overlap of the observed networks shows a significant difference (P-value < 0.001) to the intersections obtained in the randomized networks.

The results obtained after applying "Anuran" to our network reveal the presence of a central core that cannot be attributed to chance. This conclusion is based on the observed edge intersection values between the real networks and the random networks. We found that the intersection of edges in the real networks is 1684. In comparison, the intersection in random networks maintaining the degree distribution is 130, while in completely random networks it is only 4. It is clearly evident that these edge intersection values in the observed networks are significantly higher compared to those in the two types of random networks. Subsequently, we analyzed the species present in the core, shown in Table[3]. We identified species from various genera, and our analysis will focus on some of the most well-known genera with familiar functions. These include *Flavonifractor* [50], *Blautia* [50], *Eisenbergiella* [51], *Lachnoclostridium* [51], *Fournierella* [52], and *Mediterraneibacter* [53].

Genus	Core species
Flavonifractor	Flavonifractor avistercoris, Flavonifractor intestinigallinarum, Flavonifractor sp017811815
Blautia	Blautia merdaviium, Blautia ornithocaccae, Blautia pullistercoris, Blautia stercoravium, Blautia stercorigallinarum, Blautia_A avistercoris, Blautia_A excrementigallinarum, Blautia_A gallistercoris, Blautia_A intestinipullorum
Eisenbergiella	Eisenbergiella intestinigallinarum, Eisenbergiella intestinipullorum, Eisenbergiella merdaviium, Eisenbergiella merdigallinarum, Eisenbergiella pullistercoris, Eisenbergiella sp900555195, Eisenbergiella sp904392525, Eisenbergiella stercorigallinarum
Lachnoclostridium	Lachnoclostridium_A pullistercoris, Lachnoclostridium_A stercoripullorum, Lachnoclostridium_B faecipullorum, Lachnoclostridium_B phocaeense, Lachnoclostridium_B stercoravium
Fournierella	Fournierella merdipullorum
Mediterraneibacter	Mediterraneibacter cottocaccae, Mediterraneibacter excrementigallinarum_A, Mediterraneibacter faecipullorum, Mediterraneibacter glycyrrhizinolyticus_A, Mediterraneibacter intestinigallinarum, Mediterraneibacter sp900541505, Mediterraneibacter sp900761655, Mediterraneibacter vanvlietii

Table 3. Core species Species identified in the core identification using the *Anuran* tool, categorized by gene. Species are listed in the first column, while various species are listed in the second column.

After identifying the core, which includes associations present in 80% of the networks and at least 50% of the samples, we delve into the significance of these findings. These core species are pivotal for the gut microbiota's adaptation throughout the chicken's growth stages. Each species within this core plays a distinct role crucial for maintaining the overall balance and functionality of the gut microbiota.

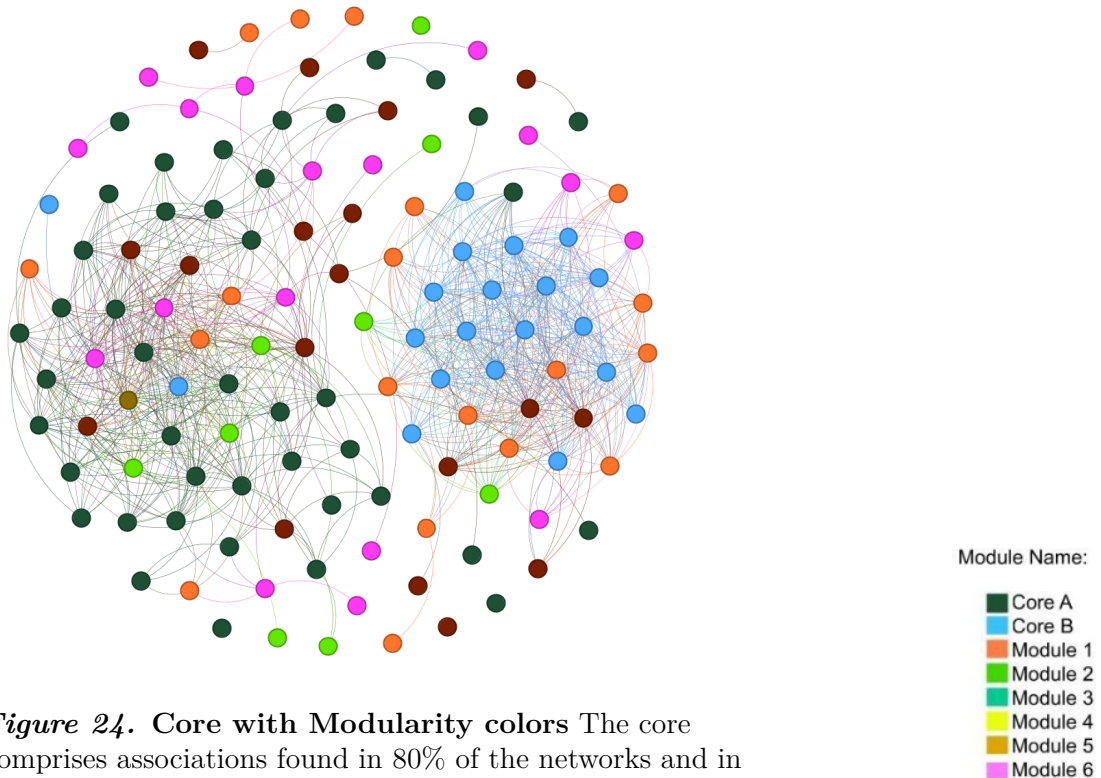


Figure 24. Core with Modularity colors The core comprises associations found in 80% of the networks and in at least 50% of the samples, with the color of the module names on Day 21 to further emphasize that the two modules detected during the modularity analysis are indeed cores.

The colors of the nodes in the graph represent the modules of the partition obtained from the control group at the age of 21. We have maintained this partition to highlight how the modularity analysis showed that Core A and Core B remain unchanged even when the treatment is altered. This means that, despite the change in treatment, the identified modular structures (modules 1 and 3) are stable.



Figure 25. Names of Core Species. Names of the species identified in the core: the colors represent their corresponding genes. We have chosen to focus exclusively on those species whose role in the chicken gut is known.

4.5 Species that distinguish between different treatments

After the identification for the keystone taxa for each treatment using the combination of the different nodes property, we finally can train the Random forest.



Figure 26. Keystone taxa in CTR, PFA, AGP at Day 14

This image shows all the keystone taxa present in the three different treatments, using network analysis on day 14. Highlighted in blue are the top 100 keystone taxa, identified based on the topological characteristics of the network.

Firstly, we will select the most meaningful and informative features of our model through Mutual Information, a method of evaluating relationships between variables. We repeated the experiment for the CTR, PFA, and AGP metrics, selecting the features that are able to describe 90 percent of the variance in the data.

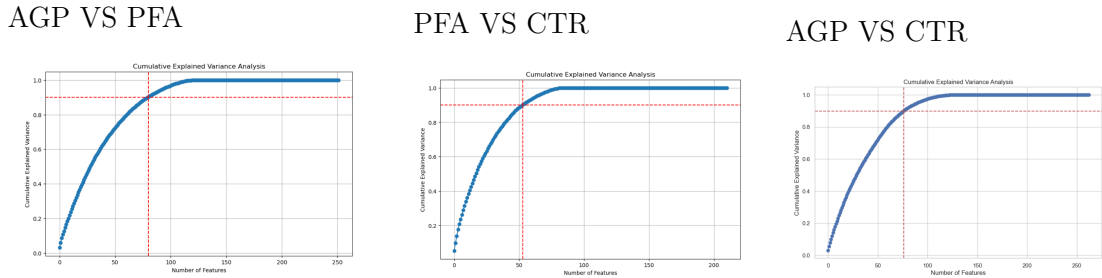


Figure 27. Cumulative Explained Variance Analysis Using Mutual Information This figure presents the cumulative explained variance analysis utilizing Mutual Information for feature selection. The Mutual Information scores are arranged in descending order, and the cumulative explained variance is computed based on the number of features under consideration. Additionally, red lines are utilized to indicate the number of selected features necessary to achieve 90% cumulative explained variance.

The results from this training, shown in the Table[4] indicate that the model can accurately distinguish samples in AGP group from the ones in PFA and Control group. However, there are difficulties in distinguishing between PFA and CTR. Subsequently, using permutation importance analysis, we identified crucial features for the model in classification, shown in the Figure[28]. These features, selected from the various species in the dataset, contribute significantly to its ability to distinguish between different experimental treatments.

	AGP vs PFA	AGP vs CTR	PFA vs CTR
Mean Accuracy	0.785 (0.130)	0.847 (0.170)	0.542 (0.140)

Table 4. Random Forest Classification Accuracy for Different Treatments

This table presents the accuracy of Random Forest classification for various treatments. The classification was conducted three times: initially focusing solely on samples treated with AGP and PFA (identified as AGP = 1 and PFA = 0), followed by an expanded analysis including samples with other treatments. Nested cross-validation was performed, repeating the test 32 times. Standard deviations of the test are reported in brackets.

The reported species are those that are most affected by the change due to the treatment and they could provide important information to better understand the biological effects and ecological dynamics involved. The classification algorithm used is able to identify the most significant species, also known as biomarkers. Biomarkers, or biological markers, are measurable indicators of some biological state or condition. They can be molecules, genes, gene products, cells, enzymes, or other substances in the body that can be measured to indicate a normal or pathological biological process, or a response to a therapeutic intervention.

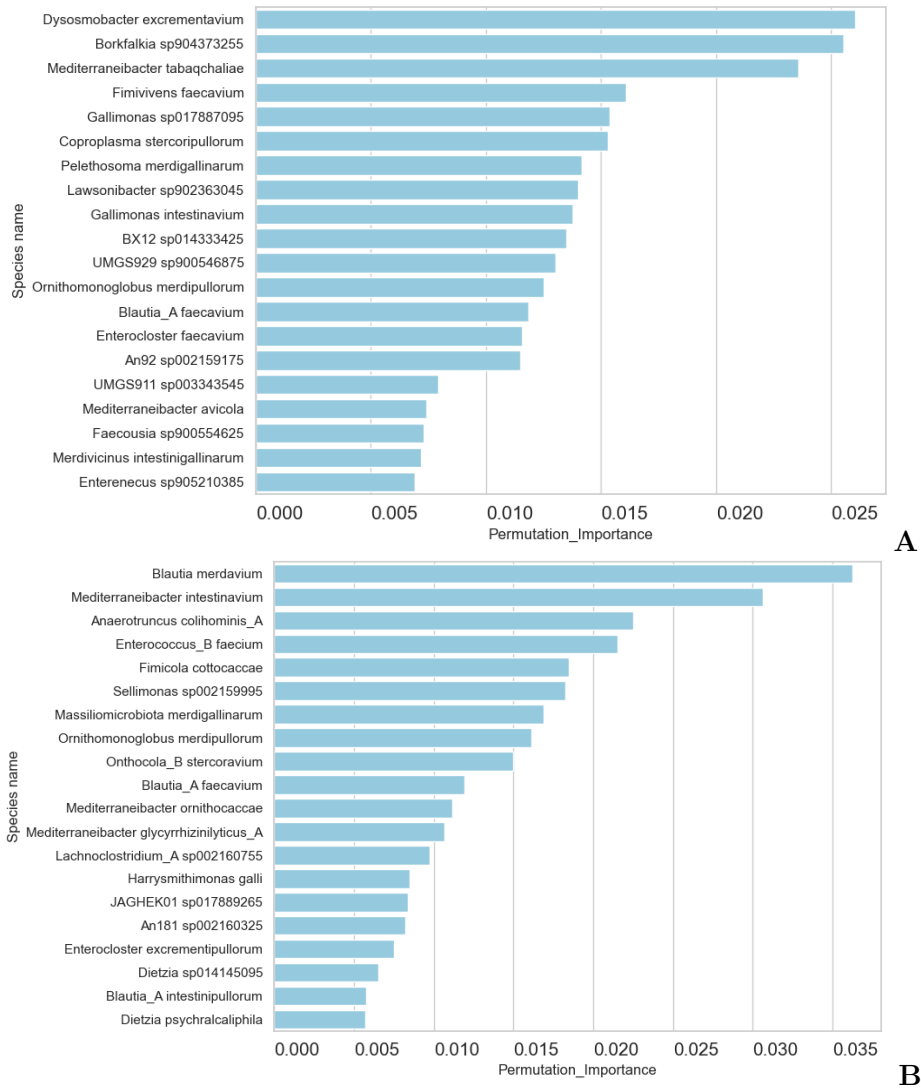


Figure 28. Important species distinguishing between treatments. This figure highlights the top 20 species obtained from permutation tests, focusing on AGP and CTR treatments in the panel A and on PFA and CTR treatments in the panel B. Species names are represented on the x-axis, while their respective calculated importance values are shown on the y-axis.

5 Discussion

We conducted a network analysis on samples treated with different treatments (AGP, PFA, CTR) and of different ages. For each network, we calculated several measures, including modularity, clustering coefficient, average degree, and number of edges. We compared the results obtained for the networks constructed with samples from AGP and

PFA treatments to the control samples (CTR), repeating this comparison for each age group. From the analysis, we found that networks constructed with AGP samples have a lower number of edges and a lower average degree, but show higher values of clustering coefficient and modularity compared to control networks. Networks constructed with PFA samples show similar results to those with AGP, but less pronounced. We also examined how the networks divide into modules using the Louvain algorithm. By keeping the association between nodes and modules fixed for the control, AGP, and PFA treatments, we observed how AGP and PFA treatments can influence modularity. During this analysis, we noticed the presence of a constant module, which remained intact even when considering AGP and PFA. This suggested the presence of a stable core in our networks, further confirmed by the Anuran tool. Additionally, we used network analysis to identify species that maintain the stability and robustness of the networks, known as keystone taxa. We considered keystone taxa to be those species with high degree centrality, high closeness centrality, and low betweenness centrality. Among these species, we selected only those that contribute to distinguishing the different treatments, using a classification algorithm like Random Forest to identify the features that contributed to this classification.

From the topological analysis of the networks, we noticed a reduction in connectivity in the networks constructed from samples treated with AGP and PFA with respect to the Control group. This highlights how AGP and PFA influence the structure of intestinal microbial networks. This result is consistent with previous studies on rumen microbiomes in cows and pigs [16], [17], where it was highlighted that antibiotics reduce network connectivity. This suggests that AGPs play a key role in changing microbial community structure, which has been rarely studied and reported so far because of the lack of thorough network analysis in previous studies. This may indicate increased specialization of interactions or a reduction in the complexity of microbial networks [18]. Interestingly, from the topological analysis, PFAs seem to impact the microbiome in a similar way as AGPs but to a lesser degree: they reduce connectivity, but not as drastically as antibiotics. This phenomenon is understudied, as there are few studies that have investigated the effects of PFAs on chicken intestinal microbiota using co-occurrence network analysis.

Furthermore, significantly high modularity and clustering coefficients were found with both the AGP and PFA groups, suggesting that the cecum microbiome might be more stable and robust. The clustering coefficient measures how close a microorganism's neighbors are to each other, indicating that with feed additives, microorganisms tend to form tightly connected groups. A high clustering coefficient is a sign of a robust and resilient network [28], as such structures facilitate cooperation and communication among microorganisms, contributing to the stability of the network itself. Modularity, on the other hand, indicates the probability that a network will divide into modules, i.e., groups of nodes with strong internal interactions and few interactions between different modules. High modularity suggests that the network is composed of well-defined communities with strong internal connections and weak external connections. This type of structure can enhance network robustness, as shown in other studies [18, 54–57], since modules can function relatively independently, reducing the impact of local perturbations on the entire network [55, 58].

To confirm that AGP and PFA improved the stability and robustness of the microbiome community, we conducted two types of experiments. First, we used random node removal to assess its overall robustness [16]. Secondly, we selectively removed 300 nodes from the co-occurrence network based on specific criteria: high degree centrality, low betweenness centrality, and high proximity centrality. As a result, we found that the network with the greatest losses, and thus the least robust, was the one constructed using the Control data. On the other hand, the network associated with AGP use was the most robust, while the network related to PFA use showed intermediate behavior between the two treatments. This observation suggests that the co-occurrence network from the AGP-treated sample is more resilient to external factors and disturbances. The PFA treatment has a less noticeable effect on network strength, but there is still an improvement compared to the network constructed using Control data.

Furthermore, the consistency of the similar results across different ages (e.g., at 21 and 35 days) indicates that the effects of AGP and PFA are strong and can influence the gut microbiome at different stages of development.

During our analysis of intestinal microbial co-occurrence networks, we observed an interesting phenomenon related to network modularity. Specifically, we noticed that while maintaining the Control group (CTR) module distribution, treatments with PFA and AGP exhibited stable modules, a core that despite variations in treatments and sample ages. This consistent modular behavior may indicate a certain resistance to external factors or significant functional importance within the intestinal microbial ecosystem. The core constitutes a fundamental component of the network that deserves further investigation to fully understand its role and influence on the dynamics of the intestinal microbiota. Subsequently, using the Anuran tool [34], we confirmed our hypothesis by finding a statistically significant core in the data, indicating that the co-occurrence networks do not overlap more than expected by chance.

We then investigated the species that characterize this core and found the presence of species, including *Flavonifractor*, *Blautia*, *Eisenbergiella*, *Lachnoclostridium*, and *Mediterraneibacter*. *Flavonifractor* is known to be beneficial in improving feed conversion efficiency and, in the presence of *Blautia*, may have contributed to the improved growth performance of broilers [50]. *Eisenbergiella* and *Lachnoclostridium* play an important role in the production of butyrate, which is the preferred energy source for gut epithelial cells [51]. *Fournierella* may be involved in the maintenance of intestinal homeostasis [52]. *Mediterraneibacter* could be involved in the fermentation of complex food substrates, such as polysaccharides and plant fibers present in the diet of chickens. It might also play a role in sugar metabolism and the balance of the intestinal ecosystem [53].

These species are considered essential for animals, as indicated by other studies. Besides aiding in food digestion and nutrient production, they play a crucial role in maintaining intestinal balance, regulating microbial populations, and safeguarding the gut from harmful pathogens. Their presence serves as a sign of stability, indicating that even with various treatments, the normal gut development of chickens remains unaffected. [50, 51, 53]

Continuing our analysis of how AGPs and PFAs affect the chicken gut microbiota, we searched for bacterial species that could distinguish between different treatments. To

achieve this, we followed a similar approach to a study by [30], where they distinguished between high-fat/low-fiber and low-fat/high-fiber monkey samples using a Random Forest algorithm on bacterial data. We applied this method to assess if our model could differentiate between AGP and PFA, AGP and CTR, and PFA and CTR treatments. From the results obtained comparing the *accuracy* of the Random Forest model during the three binary classifications, it is clear the model can effectively distinguish AGP-treated samples from those with PFA and CTR treatments. However, it struggles more with PFA-treated samples compared to CTR, though no serious errors are seen. When a model can distinguish between AGP and PFA or CTR treatments, it suggests significant differences in the gut microbiota response to each treatment. This implies bacteria in the gut react differently to AGPs compared to PFAs and Controls. If the model struggles with distinguishing between PFA and CTR, it could imply that PFA-induced changes in the gut microbiota are more subtle compared to AGP effects. This suggests that AGPs and PFAs might exert different impacts on the gut microbiota. For example, AGPs could lead to significant alterations in bacterial composition, while PFAs might affect diversity or stability without causing obvious changes. The model's difficulty in discerning between PFA and CTR treatments may indicate that PFAs have more nuanced effects on the gut microbiota. This was confirmed by network analysis, which consistently showed differential behavior between PFAs and AGPs, suggesting that AGPs might have more pronounced effects. After using permutation importance and selecting the top 20 species with the highest importance, as well as choosing only those that were also identified as keystone taxa in our networks, these species not only differentiate the effects of the different treatments but also provide insights into gut stability. Future studies could delve into how these species specifically regulate the effects of AGPs and PFAs on the chicken gut and their interactions with other bacteria, thereby influencing gut dynamics.

5.1 Future work

To deepen the understanding of the impact of AGPs (antimicrobial growth promoters) and PFAs (plant-based feed additives) on the intestinal microbiota, it would be essential to examine the functional consequences of changes in the structures of microbial networks with animal phenotype data. For instance, it is important to evaluate how alterations in connectivity and modularity influence intestinal health, nutrient absorption, and immune responses in chickens.

Exploring the robustness and resilience of microbial networks under various stress conditions, such as pathogenic challenges and environmental stresses, is another crucial aspect. Understanding how different treatments prepare the intestinal microbiota to withstand external disturbances will help develop more effective strategies to maintain the stability and health of the microbiota.

Additionally, including negative correlations in the co-occurrence network analysis can provide a more comprehensive picture of how relationships between species change in response to different treatments. In this study, we focused exclusively on positive correlations between species, excluding negative ones. This approach might have reduced the completeness and accuracy of our analysis by ignoring some important relationships.

We chose to exclude negative correlations mainly because they were rare compared to positive ones. While negative correlations can indicate competition or antagonism between species, they are often more difficult to detect and interpret. In our dataset, these negative correlations were so sparse that including them would have added little value and complicated the interpretation of the results without providing significant information. cite people that do not consider negative association.

Moreover, future work with an increasing number of samples could improve the understanding of data patterns and enhance the robustness of the analysis. More samples would better represent the variability within the intestinal microbiota and could reveal subtler but significant correlations between species.

Combining these research directions can provide a more complete and accurate view of how AGPs and PFAs influence the intestinal microbiota, improving the health and performance of animals.

6 Conclusion

The experiments described in this study show the significant impact of treatments with AGP (antimicrobial growth promoters) and PFA (plant-based feed additives) on the structure and resilience of intestinal microbial networks. Our results show that AGP treatments greatly reduce connectivity and interactions within the microbial network while maintaining the number of species. This suggests that interactions between bacterial species become simpler and more specialized under the influence of AGPs, in contrast to some studies that suggest minimal impact of AGPs on the cecal microbiota of animals.

PFAs show intermediate behavior, reducing connectivity but not as drastically as AGPs. This new insight, supported by the analysis of modularity and clustering coefficient, highlights the complex and resilient structure of microbial networks under PFA and AGP treatments. High values of clustering coefficient and modularity indicate robust and well-defined microbial communities, improving network stability and resilience. Further resilience tests through random and targeted node removal confirmed that AGP-treated networks are the most robust, while PFA-treated networks show intermediate resilience and Control networks are the least robust. This indicates that AGP treatments promote a more resilient microbial ecosystem against disturbances, while PFA treatments offer moderate improvements compared to controls.

A key observation is the constant modularity among treatments, suggesting a stable core of microbial species. This core verified using the Anuran tool, includes beneficial species like *Flavonifractor*, *Eisenbergiella*, *Lachnoclostridium*, *Fournierella*, and *Mediterraneibacter*, essential for intestinal stability and homeostasis. The presence of these species demonstrates that, despite variations in treatment, the normal development of the chicken's intestine is not affected.

Using a Random Forest algorithm for bacterial abundance data, we were able to effectively distinguish between AGP, PFA, and Control treatments. AGP treatments produced distinct changes in the microbiota, while PFA effects were more subtle. This aligns with our network analysis, which showed intermediate effects of PFAs compared to AGPs.

Our results highlight the differential impacts of AGP and PFA treatments on the chicken gut microbiota, with AGPs inducing more pronounced changes in networks and resilience, while PFAs offer moderate and beneficial effects.

Understanding these dynamics is crucial for optimizing gut health and growth performance in poultry.

7 Supplementary

The following tables show the values of connectivity loss and efficiency, derived from a random node removal process. The experiment was repeated 100 times, showing the average of both losses along with their respective standard deviations.

Age: 14	AGP	PFA	CTR
Residual Edges	11141 (158)	12161 (200)	14058 (258)
Residual Global Efficiency	0.3043 (0.0068)	0.2774 (0.0048)	0.3067 (0.0041)

Age:21	AGP	PFA	CTR
Residual Edge	13695 (196)	16825 (288)	26757 (482)
Residual Global Efficiency	0.3211 (0.0082)	0.3354 (0.0062)	0.4214 (0.0049)

Age:35	AGP	PFA	CTR
Residual Edge	16935 (228)	18371 (295)	29484 (513)
Residual Global Efficiency	0.3352 (0.0075)	0.3522 (0.0049)	0.4288 (0.0052)

Table 5. Edge and Global Efficiency loss values for age 14, 21, 35 samples, random nodes This table presents the values of edge and global efficiency loss, as illustrated in Figure [20]. The data is specific to samples at age 14. For this analysis, 100 repetitions of removing 300 randomly random nodes were performed. The reported values represent the averages of these 100 repetitions, with the respective standard deviations shown in parentheses.

The following tables illustrate the connectivity loss and efficiency values generated by the removal of nodes with a high degree of centrality in closeness and betweenness measures. These data provide an overview of the consequences of removing the most influential nodes within the network:

Age:14	AGP	PFA	CTR
Residual Edge	12065	13401	15464
Residual Global Efficiency	0.1876	0.1989	0.2406

Age:21	AGP	PFA	CTR
Residual Edge	14335	17837	30127
Residual Global Efficiency	0.1275	0.0927	0.2242

Age:35	AGP	PFA	CTR
Residual Edge	18705	19748	34133
Residual Global Efficiency	0.1355	0.1295	0.2732

Table 6. Edge and global efficiency loss values for age 14,21,35 samples, targeted node This table presents the values of edge and global efficiency loss, as illustrated in Figure [20], panel B. The data is specific to samples at age 14,21,35. For this analysis 300 targeted nodes with high degree centrality, high closeness centrality e low betweenness centrality, were removed.

In the *Results* section, we previously presented the diagram of the top 20 most important features selected by Permutation Importance after conducting binary classification by Random Forest. However, we the display to only those features deemed important for the Random Forest in distinguishing between CTR and PFA, as well as between CTR and AGP. In this section, the features used to distinguish between AGP and PFA will also be reported.

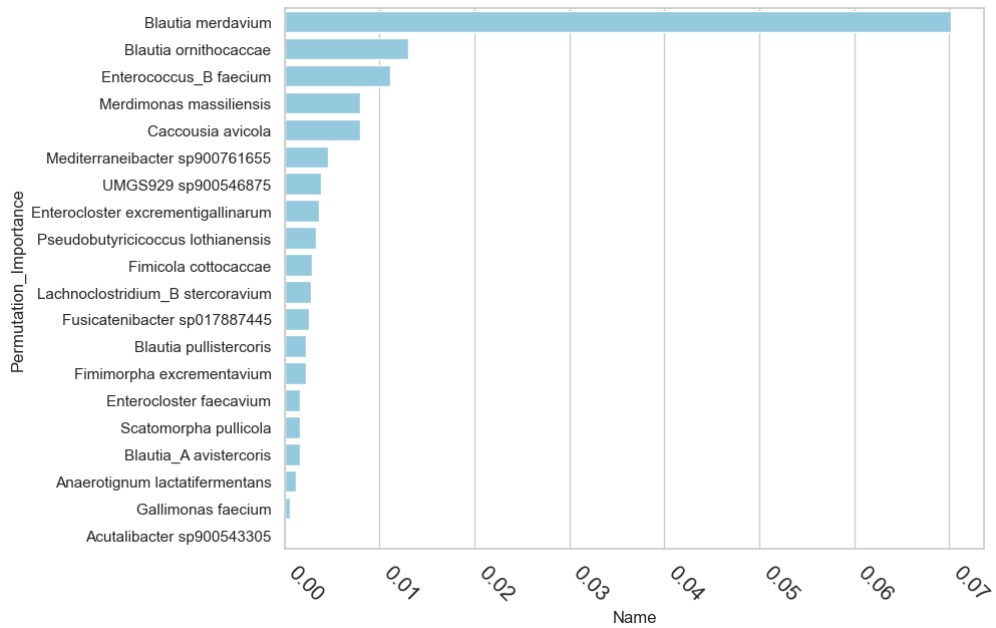


Figure 29. Top 20 Species from Permutation Test in AGP and PFA Treatments. This figure displays the top 20 species obtained from the permutation test, focusing solely on AGP and PFA treatments. Species names are represented on the x-axis, while their calculated importance values are shown on the y-axis.

In the following two pictures, we explore how modularity changes in networks derived from samples treated with PFA and AGP at age 14. As depicted in the image, two core modules remain consistent (Core a and Core B), even when maintaining fixed modules identified in the CTR across different ages. While all the other modules disappear in the other networks. This finding underscores the presence of stable core structures within our networks, suggesting that the species involved play crucial roles in the baseline functioning of chicken intestines.

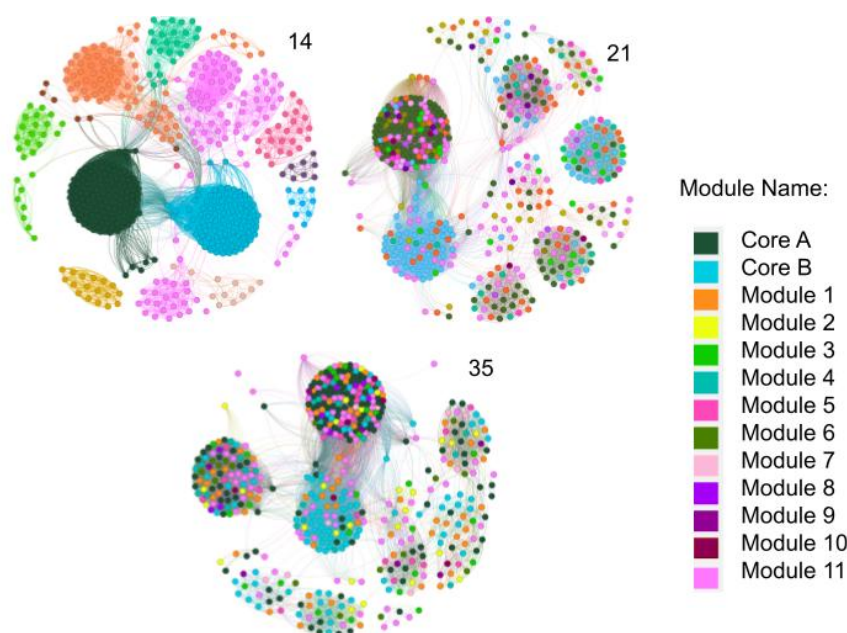


Figure 30. PFA modularity. In this image, it is evident that even while maintaining fixed modules found in PFA at age 14, the previously identified core modules A and B remain stable even in ages 21 and 35. The colors represent the modules found in PFA at age 14.

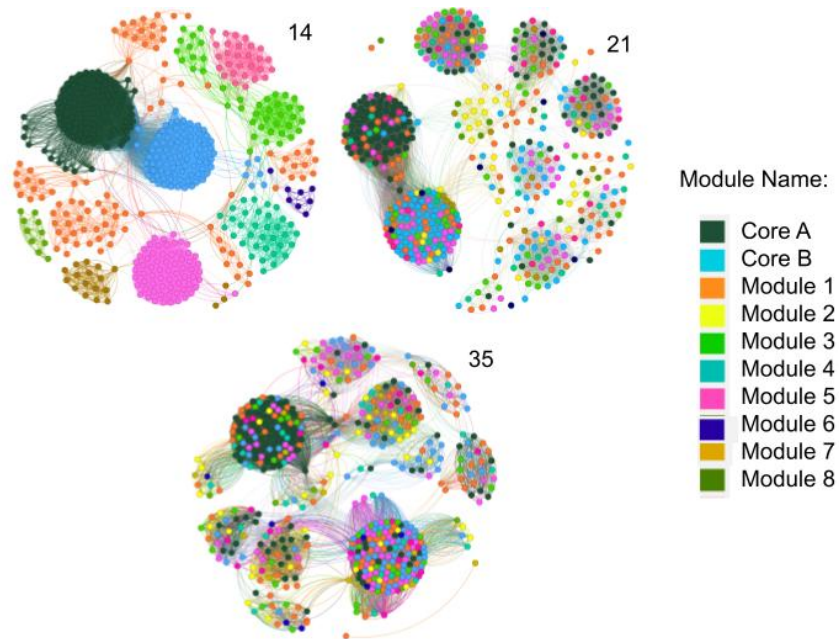


Figure 31. AGP modularity. In this image, it is evident that even while maintaining fixed modules found in AGP at age 14, the previously identified core modules A and B remain stable even in ages 21 and 35. The colors represent the modules found in AGP at age 14."

For the choice of the gamma parameter in the modularity formula (Eq. 5), various tests were conducted and the following Figure[32]. To ensure an appropriate choice of the gamma parameter, we decided to examine how the consensus partition of our networks changed as this parameter, γ , varied. This parameter, in Louvain's formula, represents the structural resolution parameter. This parameter governs the size of the modules detected in the graph modularity analysis. Increasing the value of gamma encourages the creation of larger modules while decreasing the value of gamma encourages the division of the graph into smaller modules. Therefore, the appropriate selection of this parameter is crucial to obtain a meaningful and representative consensus partition of the graph structure.

In our specific dataset, we observed that the maximum modularity index of the network is inversely correlated with the resolution of the communities. This means that with larger and less resolved communities (lower values of γ), the modularity index tends to be higher, as there is a higher proportion of connections within the communities.

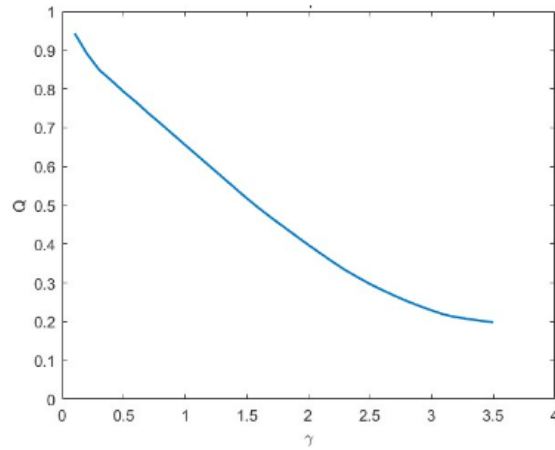
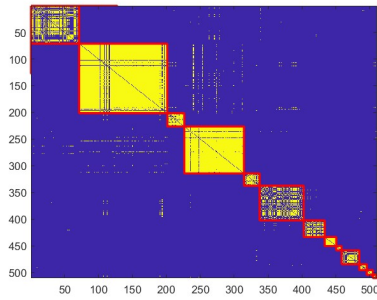


Figure 32. Modularity Index (Q) Determination Using the Louvain Algorithm The Q index of network modularity, focusing on species present in samples treated by CTR, was determined utilizing the local greedy algorithm Louvain. This index varies according to the structural resolution parameter, γ . On the y-axis, the values of Q are depicted, while on the x-axis, the values of γ are represented.

A traditional value for the parameter γ is 1. This value is commonly used as a starting point in many studies. To ensure that the chosen value of γ is appropriate for also for our dataset, γ after compute the communities for γ equal to 1 and equal to 2, this values were visualized on an adjacency heat map. From the adjacency matrix heatmap (with γ set to 2), the red lines delineating the community boundaries lie outside the yellow squares representing the communities, it suggests a mismatch between the community delineation and the density of connections within the matrix. Thus, we opted for γ equal to 1 because it strikes a balance in community resolution, steering clear of fragmentation or undue generalization.

$\gamma = 1$



$\gamma = 2$

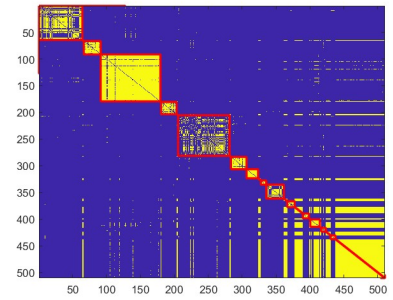


Figure 33. Heatmaps of Adjacency Matrix for Different γ Values These heatmaps display the adjacency matrix (for γ equal to 1 in the first figure and equal to 2 in the second), with red lines outlining community boundaries outside the yellow squares representing the communities themselves.

List of Figures

1	Antibiotics alternatives [3]	4
2	Gut of chickens modulate with a co-occurrence network	5
3	Relative Abundance. Graph of species distribution after calculation of relative abundance	8
4	Rare Taxa	9
5	Remove Rare Taxa	10
6	Indirect Edge	11
7	Observed Network and Random Network	12
8	Core Example	13
9	Anuran	14
10	Keystone Taxa	15
11	Pipeline Random Forest	17
12	Chicken's treatment	20
13	Co-occurrence networks pipeline	21
14	Average degree	23
15	Clustering coefficient	24
16	Modularity	25
17	Between Centrality	28
18	Degree	29
19	Closeness Centrality	30
20	Residual edges and global efficiency after random and selected node removal.	32
21	Comparison of graph modularity across different treatments on day 14, 21, 35	36
22	Distribution Core A in digesta and mucosa under different treatments	37
23	Edge intersection values appearing in at least 80% of the networks.	38
24	Core with Modularity colors	40
25	Names of Core Species	41
26	Keystone taxa in CTR, PFA, AGP at Day 14	42
27	Cumulative Explained Variance Analysis Using Mutual Information	42
28	Important species distinguishing between treatments	44
29	Top 20 Species from Permutation Test in AGP and PFA Treatments	52
30	PFA modularity	53
31	AGP modularity	54
32	Modularity Index (Q) Determination Using the Louvain Algorithm	55

33 Heatmaps of Adjacency Matrix for Different γ Values 56

List of Tables

- 1 Network analysis results from treatment groups across the time 31
- 2 Jaccard similarity of Core A and Core B 33
- 3 Core species 39
- 4 Random Forest Classification Accuracy for Different Treatments 43
- 5 Edge and Global Efficiency loss values for age 14, 21, 35 samples, random nodes 50
- 6 Edge and global efficiency loss values for age 14,21,35 samples, targeted node 51

Bibliography

- [1] A. Zou, K. Nadeau, X. Xiong, P.W. Wang, J.K. Copeland, J.Y. Lee, et al. Systematic profiling of the chicken gut microbiome reveals dietary supplementation with antibiotics alters expression of multiple microbial pathways with minimal impact on community structure. *Microbiome*, 10(1):127, 2022.
- [2] J.H. Taylor and W.S. Gordon. Growth-promoting activity for pigs of inactivated penicillin. *Nature*, 176:312–313, 1955.
- [3] U. Gadde, Woohyun Kim, Sungtaek Oh, and Hyun Lillehoj. Alternatives to antibiotics for maximizing growth performance and feed efficiency in poultry: a review. *Animal Health Research Reviews*, 18:1–20, 2017.
- [4] M.S. Diarra and F. Malouin. Antibiotics in canadian poultry productions and anticipated alternatives. *Front Microbiol*, 5:282, 2014.
- [5] A. Qu, J.M. Brulc, M.K. Wilson, B.F. Law, J.R. Theoret, L.A. Joens, et al. Comparative metagenomics reveals host-specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PLoS One*, 3:e2945, 2008.
- [6] J.L. Danzeisen, H.B. Kim, R.E. Isaacson, Z.J. Tu, and T.J. Johnson. Modulations of the chicken cecal microbiome and metagenome in response to anticoccidial and growth promoter treatment. *PLoS One*, 6:e27949, 2011.
- [7] F. Van Immerseel, J.I. Rood, R.J. Moore, and R.W. Titball. Rethinking our understanding of the pathogenesis of necrotic enteritis in chickens. *Trends Microbiol*, 17:32–36, 2009.
- [8] M.L. Gaucher, S. Quessy, A. Letellier, J. Arsenault, and M. Boulianne. Impact of a drug-free program on broiler chicken growth performances, gut health, clostridium perfringens and campylobacter jejuni occurrences at the farm level. *Poult Sci*, 94:1791–1801, 2015.
- [9] G.R. Murugesan, B. Syed, S. Haldar, and C. Pender. Phytogetic feed additives as an alternative to antibiotic growth promoters in broiler chickens. *Front Vet Sci*, 2:21, 2015.

- [10] G. R. Murugesan, Basharat Syed, Subir Haldar, and Clare Pender. Phytogetic feed additives as an alternative to antibiotic growth promoters in broiler chickens. *Frontiers in Veterinary Science*, 2:21, 2015.
- [11] J. Wang, S. Su, C. Pender, R. Murugesan, B. Syed, and W.K. Kim. Effect of a phytogetic feed additive on growth performance, nutrient digestion, and immune response in broiler-fed diets with two different levels of crude protein. *Animals (Basel)*, 11(3):775, 2021.
- [12] Youcef Mehdi, Marie-Pierre Létourneau-Montminy, Marie-Lou Gaucher, Younes Chorfi, Gayatri Suresh, Tarek Rouissi, Satinder Kaur Brar, Caroline Côté, Antonio Avalos Ramirez, and Stéphane Godbout. Use of antibiotics in broiler production: Global impacts and alternatives. *Animal Nutrition*, 4(2):170–178, 2018.
- [13] K.C. Mountzouris, V. Paraskevas, P. Tsirtsikos, I. Palamidi, T. Stenier, G. Schatzmayr, et al. Assessment of a phytogetic feed additive effect on broiler growth performance, nutrient digestibility, and caecal microflora composition. *Anim Feed Sci Technol*, 168:223–231, 2011.
- [14] Y. Shang, S. Kumar, B. Oakley, and W.K. Kim. Chicken gut microbiota: Importance and detection technology. *Frontiers in Veterinary Science*, 5:420757, 2018.
- [15] Albert Barberán, Steven T. Bates, Emilio O. Casamayor, and Noah Fierer. Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal*, 6:343–351, 2012.
- [16] Shiyu Ji, Tao Jiang, Huilong Yan, et al. Ecological restoration of antibiotic-disturbed gastrointestinal microbiota in foregut and hindgut of cows. *Frontiers in Cellular and Infection Microbiology*, 8, 2018.
- [17] Susanne A. Kraemer, Arjun Ramachandran, and Gabriel G. Perron. Antibiotic pollution in the environment: from microbial ecology to public policy. *Microorganisms*, 7(6):180, 2019.
- [18] Karoline Faust and Jeroen Raes. Microbial interactions: from networks to models. *Nature Reviews Microbiology*, 10:538–550, 2012.
- [19] Kyle Strimbu and Jorge A. Tavel. What are biomarkers? *Current Opinion in HIV and AIDS*, 5(6):463–466, 2010.
- [20] Livia J. Marcos-Zambrano, Kristina Karaduzovic-Hadziabdic, Tatjana Loncar Turukalo, et al. Applications of machine learning in human microbiome studies: A review on feature selection, biomarker identification, disease prediction and treatment. *Frontiers in Microbiology*, 12:634511, 2021.
- [21] Qusay Al-Tashi, Mohammed B. Saad, Arif Muneer, et al. Machine learning models for the identification of prognostic and predictive cancer biomarkers: A systematic review. *International Journal of Molecular Sciences*, 24(9):7781, 2023.

- [22] D. Tilman. *Resource Competition and Community Structure*. Princeton University Press, New Jersey, NJ, 1982.
- [23] J. HilleRisLambers, P.B. Adler, W.S. Harpole, J.M. Levine, and M.M. Mayfield. Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics*, 43:227–248, 2012.
- [24] R.J. Williams, A. Howe, and K.S. Hofmockel. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology*, 5:358, 2014.
- [25] S. Baldassano and D. Bassett. Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Scientific Reports*, 6:26087, 2016.
- [26] M.S. Matchado, M. Lauber, S. Reitmeier, T. Kacprowski, J. Baumbach, D. Haller, and M. List. Network analysis methods for studying microbial communities: A mini review. *Computational and Structural Biotechnology Journal*, 19:2687–2698, 2021.
- [27] B. Ma, Y. Wang, S. Ye, S. Liu, E. Stirling, J.A. Gilbert, K. Faust, R. Knight, J.K. Jansson, C. Cardona, L. Röttjers, and J. Xu. Earth microbial co-occurrence network reveals interconnection pattern across microbiomes. *Microbiome*, 8(1):82, 2020.
- [28] B. Guo, L. Zhang, H. Sun, et al. Microbial co-occurrence network topological properties link with reactor parameters and reveal the importance of low-abundance genera. *npj Biofilms Microbiomes*, 8:3, 2022.
- [29] Yanshuo Pan, Surina Borjigin, Ye Liu, et al. Role of key-stone microbial taxa in algae amended soil for mediating nitrogen transformation. *Science of The Total Environment*, 823:153547, 2022.
- [30] Y. Yan, H. Li, A. Fayyaz, and Y. Gai. Metagenomic and network analysis revealed wide distribution of antibiotic resistance genes in monkey gut microbiota. *Microbiological Research*, 254:126895, 2022.
- [31] S. Weiss, Z.Z. Xu, S. Peddada, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5:27, 2017.
- [32] Karoline Faust. Open challenges for microbial network construction and analysis. *The ISME Journal*, 15, 2021.
- [33] John Aitchison. *The Statistical Analysis of Compositional Data: Monographs in Statistics and Applied Probability*. Chapman & Hall, London, 1986.
- [34] Lisa Röttjers, Doris Vandeputte, Jeroen Raes, and Karoline Faust. Null-model-based network comparison reveals core associations. *ISME Communications*, 1(1):36, 2021.
- [35] Zahra Rashid, Zainab A. Mirani, Sehrish Zehra, et al. Enhanced modulation of gut microbial dynamics affecting body weight in birds triggered by natural growth promoters administered in conventional feed. *Saudi Journal of Biological Sciences*, 27(10):2747–2755, 2020.

- [36] Guy Amit and Amir Bashan. Top-down identification of keystone taxa in the microbiome. *Nature Communications*, 14:3951, 2023.
- [37] Samiran Banerjee, Klaus Schlaeppi, and Marcel G. A. van der Heijden. Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology*, 16:567–576, 2018.
- [38] David Berry and Stefanie Widder. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology*, 5:1–14, 2014.
- [39] Weida Gang. Demystifying machine learning models: Random forest, 2020.
- [40] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [41] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable’s importance by studying an entire class of prediction models simultaneously, 2018.
- [42] J. R. Vergara and P. A. Estévez. A review of feature selection methods based on mutual information. *Neural Computing & Applications*, 24:175–186, 2014.
- [43] Kyle Strimbu and Jorge A. Tavel. What are biomarkers? *Current Opinion in HIV and AIDS*, 5(6):463–466, Nov 2010.
- [44] Tarek H Mouhieddine, Leila El Houjeiri, Mahmoud Sabra, Ronald L Hayes, and Stefania Mondello. CNS trauma biomarkers and surrogate endpoints pipeline from bench to bedside: A translational perspective. In Firas H Kobeissy, editor, *Brain Neurotrauma: Molecular, Neuropsychological, and Rehabilitation Aspects*, chapter 20. CRC Press/Taylor & Francis, Boca Raton, FL, 2015.
- [45] M. Segura-Wang, N. Grabner, A. Koestelbauer, V. Klose, and M. Ghanbari. Genome-resolved metagenomics of the chicken gut microbiome. *Front Microbiol*, 12:726923, Aug 2021.
- [46] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29:1165–1188, 2001.
- [47] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52:1059–1069, 2010.
- [48] Salvador A. Alcalá-Corona, Santiago Sandoval-Motta, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Modularity in biological networks. *Frontiers in Genetics*, 12:701331, 2021.
- [49] Bingqiang Guo, Liang Zhang, Haijie Sun, et al. Microbial co-occurrence network topological properties link with reactor parameters and reveal importance of low-abundance genera. *npj Biofilms and Microbiomes*, 8(3), 2022.

- [50] Shengbo Zhang, Gaoxiang Zhong, Dengke Shao, et al. Dietary supplementation with bacillus subtilis promotes growth performance of broilers by altering the dominant microbial community. *Poultry Science*, 100(3):100935, 2021.
- [51] Victor Clavijo, Tania Morales, Maria J. Vives-Flores, et al. The gut microbiota of chickens in a commercial farm treated with a salmonella phage cocktail. *Scientific Reports*, 12:991, 2022.
- [52] Ying Liu, Yifan Feng, Xiaodan Yang, et al. Mining chicken ileal microbiota for immunomodulatory microorganisms. *The ISME Journal*, 17(5):758–774, 2023.
- [53] Jong Soo Kim, Kyung Chul Lee, Mi Kyung Suh, et al. *Mediterraneibacter butyricigenes* sp. nov., a butyrate-producing bacterium isolated from human faeces. *Journal of Microbiology*, 57(1):38–44, 2019.
- [54] Jens M. Olesen, Jordi Bascompte, Yoko L. Dupont, and Pedro Jordano. The modularity of pollination networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104(50):19891–19896, 2007.
- [55] Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469:351–355, 2011.
- [56] Mark E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103:8577–8582, 2006.
- [57] Qianqian Gao, Shuang Gao, Chad Bates, et al. The microbial network property as a bio-indicator of antibiotic transmission in the environment. *Science of The Total Environment*, 758:143712, 2021.
- [58] Shengwei Liu, Huang Yu, Yuhe Yu, et al. Ecological stability of microbial communities in lake donghu regulated by keystone taxa. *Ecological Indicators*, 136:108695, 2022.