Coursework Declaration and Feedback Form

The Student should complete and sign this part

Student Number: 2776284K	Student Name: Sanhita Koley
Programme of Study (e.g. MSc in Electronics and Electrical Engineering):	
Course Code: ENG5059P	Course Name: MSc Project
Name of **First** Supervisor: Dr Umer Zeeshan Ijaz	Name of **Second** Supervisor: Dr Paul Harvey
Title of Project: On Understanding Diversity and Interactions of Microbes in Obesity	

Declaration of Originality and Submission Information

I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements Signed (Student) :	
Date of Submission : 14.08.2023	

Feedback from Lecturer to Student – to be completed by Lecturer or Demonstrator	
Grade Awarded: Feedback (as appropriate to the coursework which was assessed):	
Lecturer/Demonstrator:	Date returned to the Teaching Office:

On Understanding Diversity and Interactions of Microbes in Obesity

Sanhita Koley (Student ID:2776284K)
Supervisor: Dr Umer Zeeshan Ijaz
Co-Supervisor: Paul Harvey

August 26, 2022

A thesis submitted in partial fulfilment of the requirements for the degree of
MSc Computer Systems Engineering
School of Engineering

Abstract

Obesity, reported as body mass index (BMI) rather than total body fat for epidemiologic simplicity, is generally understood to be an accumulation of excessive body fat that has negative effects. Obesity in adults over the age of 18 is characterised by a BMI ≥ 30 kg/m2 , with BMI calculated by dividing weight (kg) by height (m2). In children, median BMI (kg/m2) varies significantly with age and gender-based specific growth of the child, the BMI is instead presented as standard deviation scores (SDS). The World Health Organization (WHO) defines obesity as BMI SD scores of >3 SDS from birth to age 5 years and >2 SDS for 5-18 years above WHO growth standards median. The Centre for Disease Control and Prevention (CDC) in the US defines a child as obese at 95th centile between ages 2-19 and 97.7th centile for obese children less than 2 years of age. The International Obesity Task Force (IOTF) recommends the use of BMI cut-off points which converge to the adult BMI cut-offs of 30 kg/m2 for obese and 25 kg/m2 for overweight. Gut microbiome is now recognised as vital characteristic affecting the progression of obesity and obesity-related diseases, although it is not clear on whether changes in gut microbiota composition and metabolites lead to obesity or are a cause.This project seeks to explore this statement by working with 16S rRNA datasets from children/young people with "simple" obesity (due to an unknown cause) and "hypothalamic" obesity (related to a known cause, such as Prader-Willi syndrome). In the light of creating a contrast, 16S rRNA datasets are also available for both "hypothalamic" lean children/young adults, and healthy lean children/young adults, along with anthropometric and body composition metadata. Therefore, this project will examine the microbiomes of the four types of gut samples, concentrating on the ecology and formation of their microbial communities.

Keywords: Obesity, body mass index (BMI), excessive body fat, adults, children, growth standards, BMI SD scores, gut microbiome, obesity-related diseases, gut microbiota composition, metabolites, 16S rRNA datasets, simple obesity, hypothalamic obesity, Prader-Willi syndrome, craniopharyngioma, contrast, lean children, young adults, anthropometric, body composition, microbial communities.

Table of Contents

List of Abbreviations

ASV	**Amplicon Sequence Variant**
BMI	**Body Mass Index**
CSV	**Comma-separated value**
C2	**Acetate**
C3	**Propionate**
C4	**Butyrate**
C5	**Valerate**
C6	**Caproate**
C7	**Enanthate**
C8	**Caprylate**
IC4	**Iso-Butyrate**
IC5	**Iso-Valerate**
IVI	**Integrated Value of Influence**
NGS	**Next-Generation Sequencing**
OTU	**Operational Taxonomic Unit**

Acknowledgements

I owe my supervisor, Dr. Umer Zeeshan Ijaz, a debt of gratitude and privilege for his assistance and direction in bioinformatics. I appreciate Drs. Uzma Khan, Kelly J. Stewart, and Aqsa Ameer's assistance in helping me comprehend the subjects covered in this study. I sincerely appreciate Dr. Muhammad Jaffar Khan's contribution to the dataset utilised in this investigation. I appreciate the opportunity to broaden my knowledge and skill set that the University of Glasgow has given me.And lastly, thanks to my teammates for their assistance and support during my study.

Chapter I : Introduction

1.1 Chapter Outline

This chapter analyses the burden of obesity, its risk factors, and effects. It also covers the composition of the gut microbiota, metabolic activity in the colon, and how diversity varies over the course of life. Additional criticism of the research on the make-up and function of the gut microbiota in the aetiology of obesity from various animal studies, as well as the hypothesised mechanisms linking the gut microbiota and obesity, is provided. Finally, the purpose of the present study, the study group, and the study goals are briefly explained.

1.2 Introduction to Obesity

Obesity is a long-term medical condition in which there is an excessive buildup of body fat that is harmful to one's health. . The measurement of body fat based on height and weight is called BMI(Body mass index):

$$BMI = \frac{m_{kg}}{l^2_{m}} \quad \textbf{(1)}$$

Where m_{kg} is the body weight in kilograms, l_{m} is height in metres.
A (BMI) of 30 or more is considered obese. Overweight is defined as a BMI of 25 or more and BMI for children is described by the World Health Organisation (WHO) using the BMI Standard Deviation Score (BMI SDS),where the child's age is taken into account [1].

1.2.1. Overview of obesity in adults and children

Adult obesity was defined by the WHO in 2021 as having a BMI of 30 or higher and adult overweight as having a BMI of 25 or higher. Obesity was defined by the WHO (2021) as having a BMI SDS more than 3 SD over the WHO Child Growth Standards median and overweight as having a BMI SDS greater than 2 SD above that median for children under the age of five. For children and adolescents between the ages of 5 and 19, obesity was defined by the WHO (2021) as a BMI SDS that was more than 2 SD over the WHO Child Growth Standards median. A BMI

SDS greater than 1 SD above the WHO Child Growth Standards median was considered overweight. WHO anticipated that in 2016, more than 340 million children and adolescents between the ages of 5 and 19 will be overweight or obese, and that 38.2 million children would be in this category by 2019 [1].

1.2.2. Hypothalamic Obesity in the Context of Prader-Willi Syndrome

Prader-Willi syndrome (PWS) stands as a rare genetic disorder attributed to the absence of expression of chromosome 15q11-q13. This condition gives rise to a range of challenges, including hypotonia and feeding difficulties during infancy, coupled with later complications such as obesity, intensified by uncontrollable hyperphagia in adolescence and beyond. In contrast to simple obesity, individuals with Prader-Willi syndrome (PWS) possess diminished lean body mass and heightened fat mass, primarily concentrated in the trunk, with a lower prevalence of visceral obesity. These observations suggest that the lowered resting energy expenditure (REE) could contribute to the development of obesity in those with PWS [2]. Hence Hypothalamic obesity, stemming from disruptions in hypothalamic signalling, provides a unique lens through which to examine the gut microbiota's involvement in obesity. Individuals with hypothalamic obesity often possess distinct microbial profiles that contribute to their pathological weight gain. On the other hand, those who have hypothalamic lesions that result in lean phenotypes also have altered gut microbiota, demonstrating the microbiome's intricate role in hypothalamic control.

1.2.3. Health risks and complications associated

Obesity among adults stands as a significant health concern, giving rise to a spectrum of health issues encompassing cardiovascular diseases, diabetes, joint problems, musculoskeletal disorders, respiratory complications, and psychological distress. Similarly, childhood obesity lays the foundation for enduring difficulties, as children become more susceptible to chronic ailments and face an elevated risk of grappling with low self-esteem and depressive tendencies. Addressing childhood obesity becomes paramount to forestalling the perpetuation of health challenges into adulthood. Experiencing significant weight gain during pregnancy and retaining weight after childbirth can impact subsequent fertility and elevate the chances of complications in future pregnancies [3]. Additionally, pregnancy introduces distinctive complexities, as maternal obesity heightens the likelihood of gestational diabetes, elevated blood pressure,

labour-related complications, birth anomalies, and the need for caesarean deliveries. Moreover, the offspring of mothers burdened by obesity confront the prospect of enduring long-term health implications, including obesity and associated afflictions.

1.2.4. Maintenance and Management of Obesity

Obesity management and maintenance require a comprehensive and sustainable approach. This includes eating a balanced diet, exercising regularly, and addressing psychological factors that affect eating habits. Customised strategies, behavioural interventions, and social support are all important for maintaining weight loss. Monitoring progress, adapting strategies, and seeking professional guidance can help ensure long-term success. By adopting a holistic lifestyle change and staying vigilant, individuals can effectively manage obesity and improve their overall health and well-being. According to [4] adopting a Mediterranean diet, characterised by higher consumption of fruits, vegetables, and the inclusion of beneficial fats such as monounsaturated and polyunsaturated fats, promotes a health-conscious approach to achieving and sustaining weight loss goals.

1.3. Gut Microbiota and Obesity

The human gut harbours a diverse and dynamic ecosystem of microorganisms, collectively known as the gut microbiota. This intricate assembly comprises bacteria, viruses, fungi, and other microorganisms that engage in intricate crosstalk with their human host. This mutualistic relationship extends beyond simple digestion and absorption of nutrients, encompassing a wide array of physiological functions, including immune system modulation, energy metabolism, and even neural communication. The gut microbiota has evolved to become a key regulator of various bodily processes, exerting a profound influence on health, disease and obesity.

1.3.1. Exploring the diversity of gut microbiota

The research conducted in [5] claims that the human intestinal tract comprises approximately 99% of intestinal microorganisms, and the term microbiota is frequently used interchangeably with bacteria. The gut's microbial community is categorised into five prominent phyla: Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, and Verrucomicrobia (Figure 1).

Firmicutes constitute around 60% of the total bacterial population, including Bacilli, Clostridia, Erysipellotrichia, Negativicutes, Thermolithobacteria, and some unclassified Firmicutes. Bacteroidetes, comprising 10-20% of total bacteria, encompasses Bacteroides, Cytophagia, Flavobacteria, and Sphingobacteria. However, a substantial portion of identified sequences, about 64-70%, remains unclassified and their functionalities are yet to be elucidated.

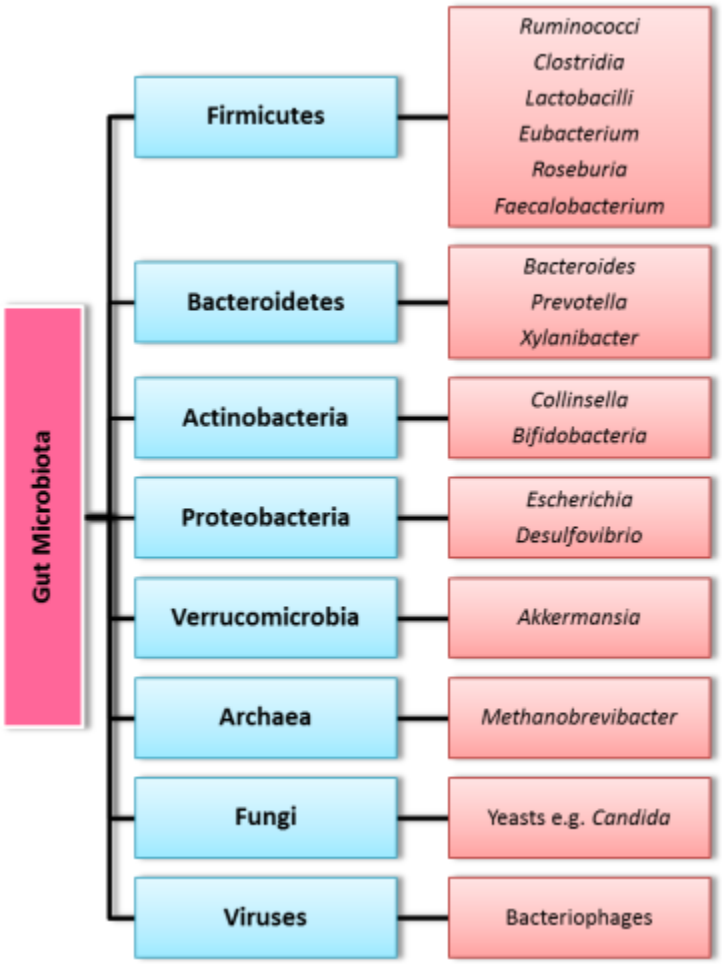

Figure 1: Principal phylum of gut bacteria and their predominant subgroups [5,6].

1.3.2. Connection between gut microbiota composition and obesity

Gut microbiota and its role in obesity has only scratched the surface of understanding the intricate interplay between these elements. Researchers are striving to unravel the complex web of interactions between genetics, environment, gut bacteria, and neural regulation that collectively contribute to obesity's diverse manifestations. Metabolic byproducts known as short-chain fatty acids (SCFAs), which include acetic acid, isobutyric acid, formic acid, isovaleric acid, propionic acid, butyric acid, and valeric acid, are generated through fermentation processes between anaerobic microbes and digestible carbohydrates in the cecum of the human intestine. Acetate, propionate, and butyrate, are the most prevalent SCFAs within the intestinal environment. This rise in SCFA levels in the bloodstream and the corresponding reduction in faecal concentrations have been associated with conditions like obesity and metabolic irregularities [7]. A study carried out by Liu et al. (2021), recognized that there exists a significant connection between obesity and disruptions in the balance of gut microbiota. Numerous intestinal microorganisms have been associated with obesity. They contribute to obesity's development and advancement through various mechanisms, such as enhancing the absorption of energy by the host, increasing central appetite, encouraging the storage of fat, triggering persistent inflammation, and influencing circadian rhythms. Given the intricate and diverse nature of the gut microbiota, further research is imperative to comprehensively understand the precise mechanisms through which it contributes to obesity [8].

1.4. Bioinformatics

1.4.1. Exploring the Gut Microbiota through Next-Generation Sequencing

In recent times, there has been a transition from conventional Sanger sequencing to Next-Generation Sequencing (NGS) due to its advantages. While Sanger sequencing was considered the "gold standard" for its accuracy and long reads, it had limitations such as labour-intensive processes and biases against certain genes [9]. The human host's gene expression capability is 100 times greater than that of the gut microbiome, which creates a microbial ecology that has a substantial impact on human health and disease development.

Modern sequencing technologies, such as 16S rRNA, 18S rRNA, internal transcribed spacer (ITS) sequencing, shotgun metagenomic sequencing, metatranscriptomic sequencing, and viromic sequencing, have been widely used to study the gut microbiome because many intestinal microbial species reject conventional culture techniques [10]. In the initial years of gut microbiome exploration, point of convergence has been directed towards DNA-based 16S rRNA gene sequencing and shotgun metagenomic sequencing. These techniques are instrumental in unveiling the microbial composition and genetic content (Figure 2). Although with significant technological advancements leading to substantial increases in sequencing coverage, the comprehension of the genetic diversity conveyed by the 16S rRNA gene remains uncertain for microbial ecologists. There are two prevalent categorization methods for sequences: one relies on their similarity to reference sequences (phylotyping), while the other is based on their resemblance to other sequences within the community (known as OTUs, operational taxonomic units). OTU denotes a cluster of closely associated entities organised according to the likeness of specific sequences, commonly centred around the 16S rRNA gene [11].

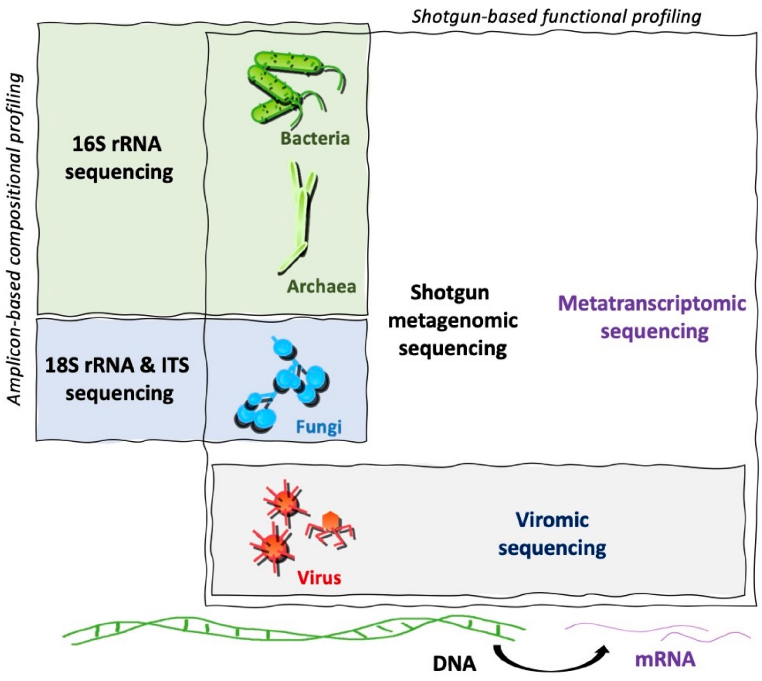

Figure 2: Sequencing Techniques used for gut microbiome research [10].

1.5. Aims and objectives of this Project

- To investigate the interactions between microbes in individuals who are naturally obese and see if there is a difference in microbial community structure where obesity is caused by some pathology, e.g., "hypothalamic" obesity caused by Prader-Willi syndrome. As a baseline, we compare healthy and hypothalamic lean samples.
- Utilising recent network inference methods, statistics, and diversity exploration, particularly using the recently developed zeta diversity framework to unravel any patterns that represent the different categories.
- To acquire an in-depth understanding of the ecology and development of microbial communities within the specified groups, contributing to insights into their roles in obesity and related metabolic processes.
- Lastly, to gain an overall comprehension of how the gut microbiota influences the development and progression of obesity, ultimately contributing to the broader understanding of this complex condition.

Chapter II: Methodology

2.1 Chapter Outline

In this chapter, a comprehensive overview is presented, which includes the data set utilised, the research methodology employed, bioinformatic procedures for managing the data, the application of statistical analysis techniques, and the use of R Studio for executing Network visualisation and Zeta Diversity analysis.

2.2 Dataset description

The data utilised in this study was gathered and supplied by Dr. Muhammad Jaffar Khan [5], contributing a dataset comprising 151 faecal samples originating from individuals based in the United Kingdom. Among these samples, 52 were derived from the healthy lean control group, 29 from the healthy obese group, 22 from the hypothalamic lean group, 19 from the hypothalamic obese group, and a further 29 from the respective parental groups. The data collection period spanned from October 2011 to January 2013. Employing 16S rRNA sequencing, the study utilised the QIIME 2 pipeline from [12] to cluster the sequences into Operational Taxonomic Units (OTUs) while referencing the Silva Release 138 taxonomy database in [13]. The creation of the BIOM file and the NEWICK format phylogenetic tree file was executed through the utilisation of the DADA2 pipeline [14].

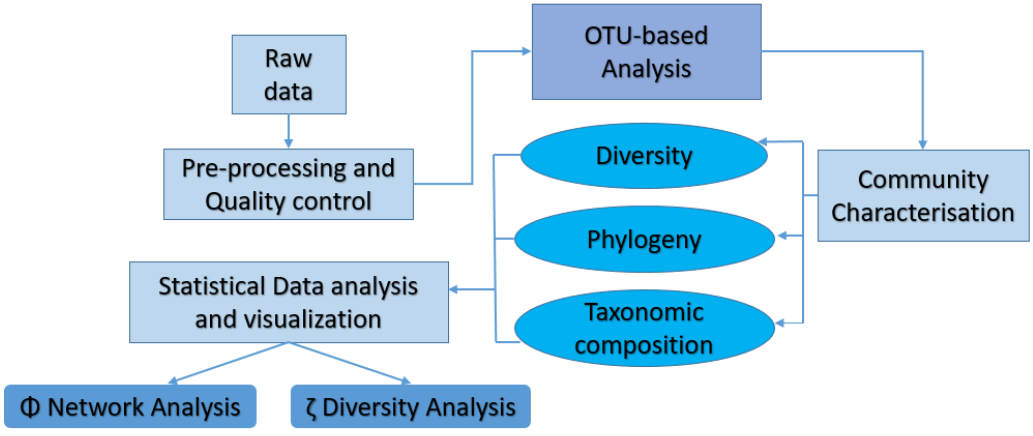

Figure 3: A schematic diagram of the data analysis process for gut microbiota studies. Adapted from [10]

2.3 Statistical analysis:

All statistical analyses within this study were conducted using R version 4.0.3 (RCoreTeam, 2020), employing the OTU table that was generated from the acquired data set mentioned in section 2.2. The study incorporated NEWICK files and BIOM files, containing the taxonomic information for each OTU. These files were computed following the methodology outlined in Figure 3 and were cross-referenced with pertinent metadata and meta-tables, for conducting the statistical analysis of 16S rRNA data within the RStudio software environment.The metadata table includes the compositional information for dried samples—C2 (Acetate), C3 (Propionate), IC4 (Iso-Butyrate), C4 (Butyrate), IC5 (Iso-Valerate), C6 (Caproate), C7 (Enanthate), and C8 (Caprylate)—alongside the associated sample group designation (Healthy Lean, Healthy Obese, Hypothalamic Lean and Hypothalamic Obese).

2.3.1 Network analysis: Visualisation through Phi (ϕ) networks

Phi (ϕ) networks are a type of network that is characterised by a high degree of clustering and a low degree of path length. This means that nodes in a phi network are more likely to be connected to each other, and the distance between any two nodes is relatively short. Phi networks are often found in nature, such as in the food web of an ecosystem or the neural network of a brain. They are also found in social networks, such as online communities. Phi networks have a number of advantages over other types of networks. For these reasons, phi networks are becoming increasingly important in a variety of fields, such as computer science, biology, and sociology. And hence is chosen to create a network for the meta-data mentioned in section 2.3. The statistic ϕ for "goodness-of-fit to proportionality" serves as a measure to evaluate how closely a pair of random variables (x, y) adhere to being proportional to each other [15].The log-ratio variance is utilised as a metric to quantify the connection between variables, and as it approaches 0, it signifies that the two variables are exhibiting proportionality [16].

Consider the case where x corresponds to OTU_1 and y corresponds to OTU_2. In this case, it is important to examine whether the value of $var(log(x/y))$in the following equation (2) tends to approach zero:

$$var(log(x/y)) = var(log\, x - log\, y)$$

$$= var(log\, x) + var(log\, y) - 2cov(log\, x,\, log\, y) \quad \textbf{(2)}$$

$$= var(log\, x).\left(1 + \frac{var(log\, y)}{var(log\, x)} - 2\sqrt{\frac{var(log\, y)}{var(log\, x)}}.\frac{cov(log\, x, log\, y)}{\sqrt{var(log\, x)var(log\, y)}}\right)$$

$$= var(log\, x).(1 + \beta^2 - 2\beta|r|) \triangleq var(log\, x).\, \phi(log\, x,\, log\, y) \quad \textbf{(3)}$$

In equation (3), β represents the standardised major axis estimate of the slope between the logarithms of variables y and x. Additionally, 'r' denotes the correlation between these variables. The first term, var(log x), pertains solely to the magnitude of variation and doesn't involve y. The second term, ϕ, characterises the degree of proportionality between x and y, forming the foundation for assessing relative value relationships. It's possible to create other non-negative functions of β and r that become zero when x and y are perfectly proportional [17]. Networks were formed based on the ϕ values, by applying equations (2) and (3) to the OTUs within each group (Healthy Lean, Healthy Obese, Hypothalamic Lean, and Hypothalamic Obese), where each node represents a phylum-genus of a microbiome. Phi networks are usually created through a mechanism known as preferential attachment. This process entails new nodes having a greater likelihood of linking to nodes that already possess numerous connections. Consequently, this mechanism fosters the development of a network characterised by substantial clustering and comparatively shorter path lengths. As a result, selecting the nodes (node => phylum-genus of a microbiome) that hold more significance or influence becomes challenging for each group. Identifying the influential node is a necessary undertaking in visualising the metadata in the form of a phi network during network analysis. To identify such influential individuals/nodes, researchers Salavaty, Ramialison and Currie (2020) [18] introduced an innovative algorithm

named the Integrated Value of Influence (IVI) that combines the critical topological attributes of the network (Spreading score and Hubness score) to find its pivotal individuals/nodes.

2.3.1.1 Spreading Score

With an objective to combine prevalent local, semi-local, and global measures of network centrality, the study carried out in [18] had a collective intention to merge the impact of these measures to identify influential nodes within the network in a neutral manner. Specifically, degree centrality and ClusterRank, neighbourhood connectivity and local H index, and betweenness centrality and collective influence. Consequently, the Min-Max feature adopted in [19] as a scaling technique can be applied to standardise all centrality measures onto a uniform scale while preserving their relative weight proportions. Expanding on the principles underlying the measurements of range normalised neighbourhood connectivity, ClusterRank, betweenness centrality, and collective influence, thus taking their distinct topological attributes into account, the result is referred to as the "Spreading Score", equation (4). This score is indicative of the vertices' capacity to propagate information throughout the network.

$$Spreading_{scorei} = (NC'_i + CR'_i)(BC'_i + CI'_i) \quad \textbf{(4)}$$

where NC'_i, CR'_i, BC'_i, and CI'_i are range normalised neighbourhood connectivity, ClusterRank, betweenness centrality, and collective influence of node i, respectively. A generic function to find the Spreading score of the desired nodes from a graphical network can be found in the Appendix [20].

2.3.1.2 Hubness score

The Hubness score signifies the influence exerted by each node within its nearby surroundings and stands as a significant constituent of the Integrated Value of Influence (IVI) [21]. By applying the same reasoning applied in section 2.3.1.1, to produce the Spreading score, the combined result of local H index and degree centrality is expressed as the "Hubness score." This score has the potential to represent a vertex's dominance within its immediate local area.

$$Hubness_{scorei} = DC'_i + LH'_{index\,i} \quad \textbf{(5)}$$

Where DC'_i is normalised degree centrality and $LH'_{index\,i}$ is the local H index of node i [18].

A generic function to find the Hubness score of the desired nodes from a graphical network can be found in the Appendix [21].

2.3.1.3 IVI

Hubness and Spreading scores are combined to create IVI. Hubness and Spreading are two measures that each capture particular characteristics of nodes. Hubness evaluates a node's local impact, whereas Spreading projects how much information it might spread. According to Salavaty et al. (2020) [18], a vertex's influence on the network increases when the Spreading and Hubness values are multiplied higher. IVI is produced as a result of this integration, which is accomplished by the Multiplication function. IVI is essentially a consolidated measure of local, semi-local, and global centrality generated from key metrics like degree centrality, ClusterRank, neighbourhood connectivity, local H index, betweenness centrality, and collective impact. Its goal is to balance out network-wide positioning biases [18].

$$IVI_i = (Hubness_{scorei}).(Spreading_{scorei}) \qquad \textbf{(6)}$$

Where the values for $Spreading_{scorei}$ and $Hubness_{scorei}$ can be found from equation (4) and (5) respectively.

A generic function to find the Hubness score of the desired nodes from a graphical network can be found in the Appendix. [22]

2.3.1.4 Methodology carried out in R studio program

In R studio 4.0.3 , various libraries are utilised, such as "igraph," "ggraph," "tidygraph," "sna," "extrafont," "influential," and "tidyverse," to perform an in-depth analysis of network statistics for a given dataset. At first the parameters related to the input data file, image dimensions, and label are set. Then the CSV file containing data for a network and creates a mapping dataframe for taxa levels is read. Unique groups in the dataset are iterated, followed by construction of an igraph network object, and calculation of a wide range of centrality and influence measures for each node in the network. These measures include degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, subgraph centrality, ivi, hubness score and spreading score. The calculated statistics are stored in a dataframe named "ig_stats." For each

calculated statistic, the phi networks are generated using the "ggraph" library. It plots the nodes of the network, where the size and color of nodes correspond to specific statistics, and labels nodes based on taxonomic information. The generated plots are saved as PDF files. Additionally, a CSV file for each group, containing the calculated network statistics along with taxonomic information is also created. Hence this process demonstrates a comprehensive analysis of network statistics and visualisation techniques to gain insights into the characteristics of the network nodes. The various centrality and influence metrics provide information about node importance and connectivity patterns within the network in the context of taxonomic relationships.

2.3.2 Zeta (ζ) Diversity analysis

Zeta diversity can be defined as a metric that quantifies the count of shared species across multiple assemblages. It diminishes with increasing zeta orders, indicating that the count of shared species reduces as additional samples or sites are included. This reduction occurs because the introduction of more samples enhances the probability of encountering a rare species that exists in only a limited number of assemblages (Figure 4). Hence, it can be inferred that zeta diversity offers an estimate of the mean count of species shared among n sites or shared OTUs across n number of instances [23].

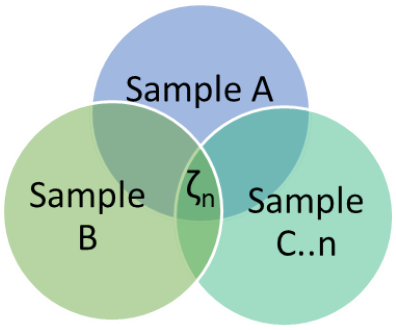

Figure 4: Illustration of the third-n order ζ diversity Venn diagram. Adapted from [24].

When integrated with established spatial regression methodologies and environmental data collected from the designated sites, ζ diversity offers an avenue to discern factors influencing the shifts in species composition across the entire range of species, spanning from infrequent to

prevalent. This comprehension holds growing importance in conservation strategies, given the imperative to effectively address the impacts of environmental changes on both less common and more abundant species [25]. In the context of zeta diversity, there are two laws that determine the behaviour of the decline in diversity-Exponential Law and Power Law. Exponential law suggests that the decline in the number of shared species is due to random factors. Power law, on the other hand, suggests that the decline is due to deterministic factors, such as environmental pressure [26]. To assess variety across sets of assemblages, Zeta variety presented a revolutionary multi assemblage recursive technique. This concept was developed as a result of the inadequacy of traditional metrics to characterise the characteristics of sets of assemblages. It allows for the evaluation of how the number of species shared between assemblages fluctuates while taking into account various assemblage numbers. To put it another way, Zeta diversity is not a single metric but rather a group of metrics that change depending on how many assemblages are taken into account in the calculation. The higher the order of Zeta, the more common a species must be to be shared among assemblages [26].

2.3.2.1 Methodology carried out in R studio program

In this study, R studio has been used to observe zeta diversity of the acquired metadata. The "phyloseq" library [27] is used to process microbial community data and the associated metadata. Initially, the script imports the data from a BIOM file and a CSV file. The data is then preprocessed, involving steps such as transposing the abundance table, removing low-library-size samples, and filtering out contaminants. Taxonomic information is extracted and formatted, and samples are matched between the abundance table and metadata. Zeta diversity metrics are calculated using the "Zeta.decline.ex" and "Zeta.decline.mc" functions from the "zetadiv" library, considering both exponential and power law relationships. The results are aggregated and stored. Subsequently, the script employs the "ggplot2" library to visualise the calculated Zeta diversity values and ratios. Distinct plots are generated to illustrate these values for different sample groups. The collated Zeta diversity metrics and AIC values are saved as CSV files for further analysis and interpretation. In summary, this R algorithm process facilitates comprehensive analysis, calculation, and visualisation of Zeta diversity metrics to assess shared species diversity among microbial samples in distinct groups.

Chapter III: Results

3.1 Chapter Outline

This chapter discusses the study's findings regarding network visualisation and Zeta diversity analysis. Through network visualisation techniques, the chapter visually portrays the intricate connections and most influential nodes between microbial taxa within the community. This approach reveals potential patterns of influence and centrality, offering insights into the hierarchical structure of the microbial ecosystem. Additionally, the Zeta diversity analysis assesses the ecological diversity and evenness across samples, illuminating variations in taxa distribution and relative abundance.

3.2 Network analysis: Phi networks:

The outcomes from the analysis of the Phi network have proven in the identification of pivotal gut microorganisms that potentially contribute to the onset of obesity or are influenced by its presence. This critical revelation was made possible through the application of advanced computational methodologies, specifically the Spreading score, Hubness score, and IVI score, as elaborated in Section 2.3.1. By applying these quantitative measures, it is possible to assess the significance and influence of various microbial entities within the network. The Spreading score illuminated nodes with enhanced potential to propagate information across the network, shedding light on key influencers. Similarly, the Hubness score unveiled nodes wielding substantial authority within their surroundings. Furthermore, the IVI score, integrating multiple centrality measures, effectively pinpointed nodes with the most comprehensive and multidimensional influence. And lastly, the microbes of high influence found from the current research are presented in Table 1.

3.2.1 Healthy Lean control

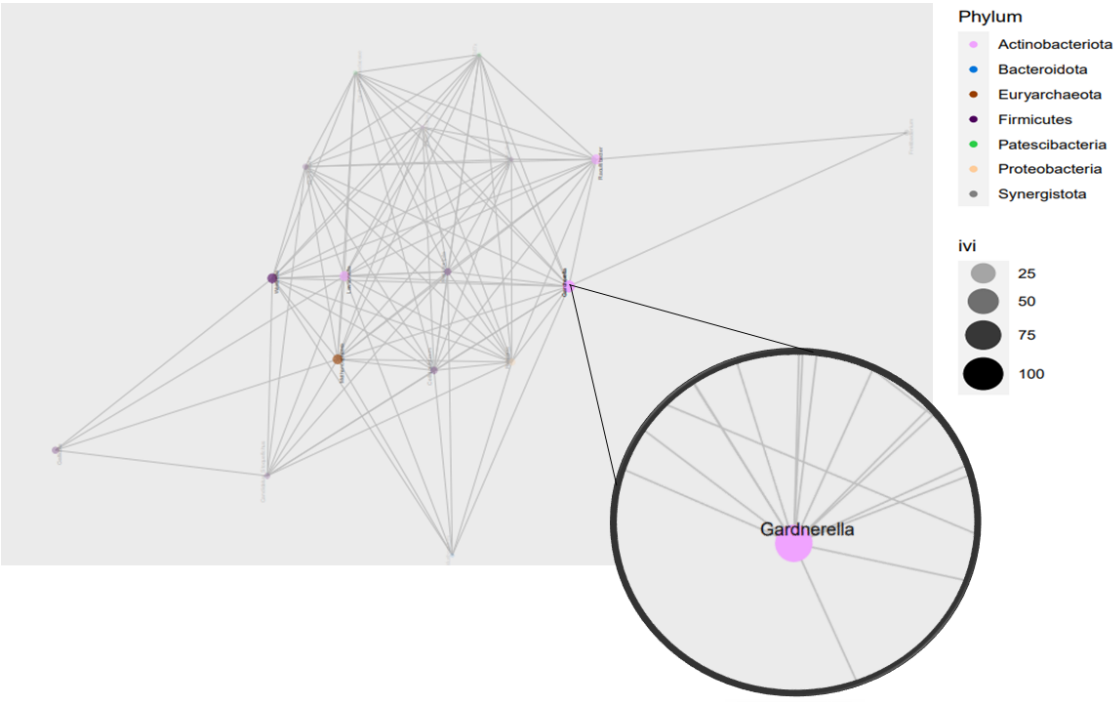

Figure 5: Phi-network visualisation of Healthy Lean control, where each node represents the genus of bacteria, the phylum is denoted by colour and score denoted by radiance.

The results of the network analysis of the bacterial communities in the gut of healthy lean control shows the abundance of different genera of bacteria, as well as their phylum and class. Figure 5 shows the most influential node (genus) of the network, Gardnerella, which belongs to the phylum Actinobacteriota and Bifidobacteriaceae family. It achieved a score of 100 in Spreading Hubness and IVI. The most dominant bacterial genus in the gut is Lachnospiraceae. This genus belongs to the phylum Firmicutes, which is the most abundant phylum of bacteria in the gut. Lachnospiraceae bacteria are involved in a variety of functions in the gut, including fermentation of dietary fibre, production of short-chain fatty acids, and regulation of the immune system. Another dominant bacterial genus in the gut is Bifidobacterium. This genus also belongs to the phylum Firmicutes. Bifidobacterium bacteria are also involved in fermentation of dietary fibre and production of short-chain fatty acids [28]. They are also known for their probiotic properties, which means that they can have beneficial effects on health.

3.2.2 Healthy obese

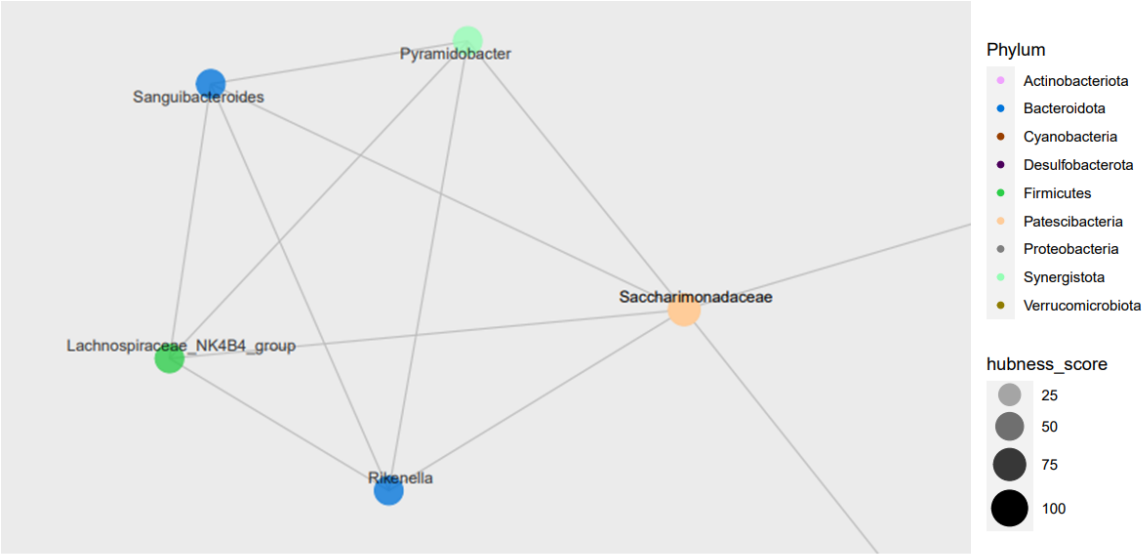

Figure 6: Phi-network visualisation of Healthy Obese, where each node represents the genus of bacteria, the phylum is denoted by colour and score denoted by radiance.

In the gut of healthy obese adults, the most prevalent bacterial genera are E. coli (Proteobacteria), Enterococcus (Firmicutes), and Bacteroides (Bacteroidetes). Collectively, these three genera constitute approximately 50% of the total gut bacteria in this group. In Figure 6, the genus Saccharimonadaceae from the phylum Patescibacteria was identified to be most significant with high IVI. Hubness and Spreading scores. Each genus contributes significantly to various essential functions within the gut ecosystem. E. coli participates in dietary fibre fermentation, short-chain fatty acid production, and immune system regulation. Enterococcus is involved in dietary fibre fermentation, short-chain fatty acid production, and is recognized for its probiotic attributes. Bacteroides contribute to diverse roles, including complex carbohydrate digestion, vitamin production, and immune system regulation. While it's challenging to definitively pinpoint the most influential genus, as all three play pivotal roles, their composition can vary due to factors such as diet, lifestyle, and medication use. Overall, the dominant bacterial genera from Proteobacteria, Firmicutes, and Bacteroidetes phyla collectively uphold vital functions crucial to maintaining gut health among healthy obese adults.

3.2.3 Hypothalamic Lean

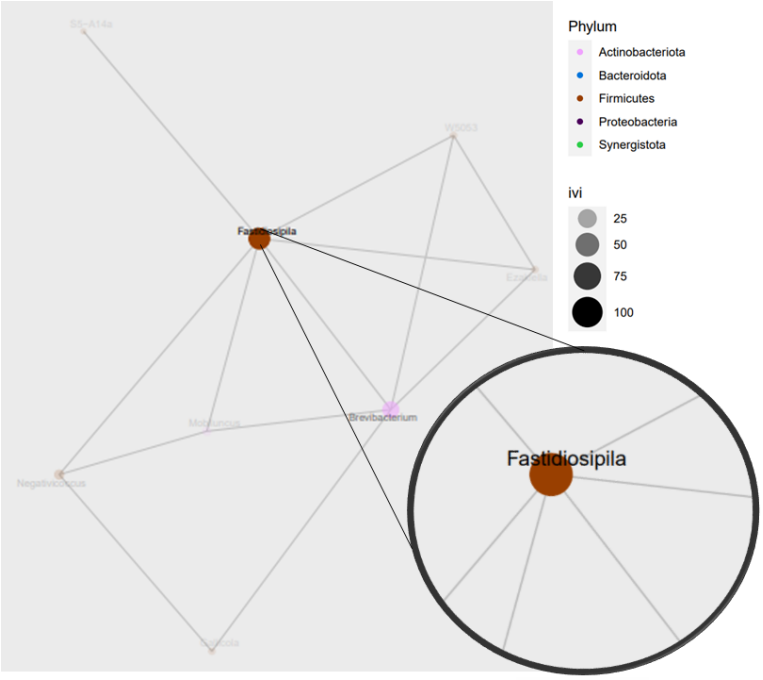

Figure 7: Phi-network visualisation of Hypothalamic Lean, where each node represents the genus of bacteria, the phylum is denoted by colour and score denoted by radiance.

Figure 7 presents the discoveries surrounding the influential microbe, Fastidiosipila, belonging to the Firmicutes phylum. Remarkably, this microbe exhibited a remarkable score of 100 in all-IVI, Hubness, and Spreading metrics upon scrutinising the gut bacterial communities within lean subjects afflicted by hypothalamic obesity. This finding signifies the extraordinary influence wielded by Fastidiosipila within the microbial network, showcasing its dominance across multiple crucial aspects. Notably, the most prevalent bacterial genera within this group are Akkermansia (Verrucomicrobia), Ruminococcus (Firmicutes), and Bacteroides (Bacteroidetes), collectively comprising around 55% of the total gut bacteria. Akkermansia and Ruminococcus contribute to immune system regulation and inflammation protection, while Bacteroides participate in complex carbohydrate digestion, vitamin production, and immune system regulation.

3.2.4 Hypothalamic obese

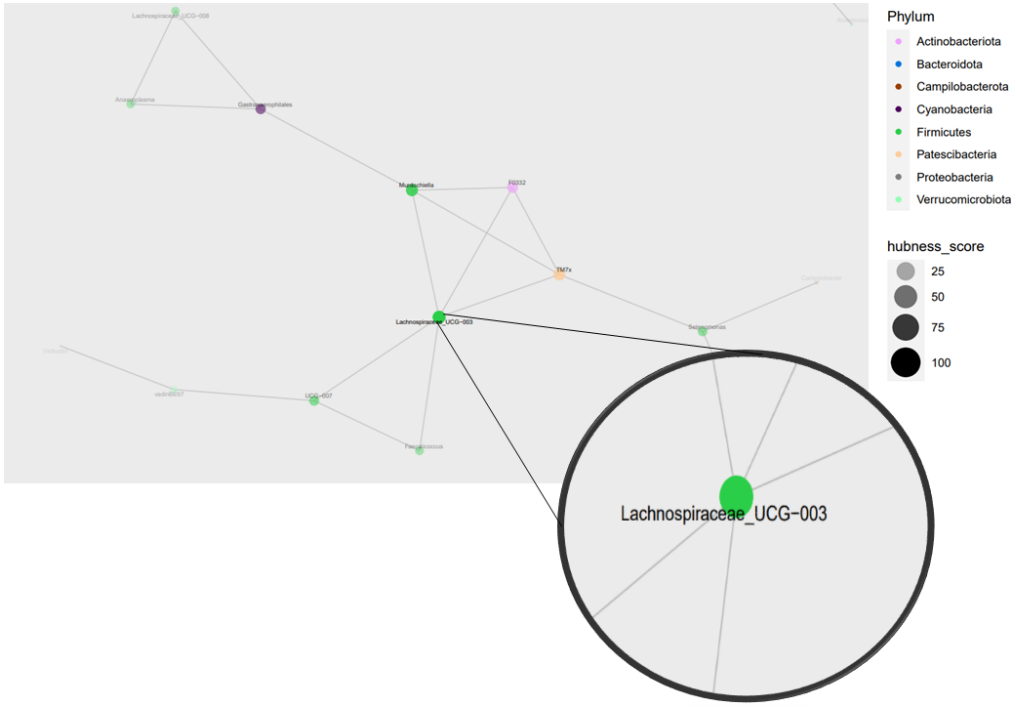

Figure 8: Phi-network visualisation of Hypothalamic obese, where each node represents the genus of bacteria, the phylum is denoted by colour and score denoted by radiance.

Genus Lachnospiraceae_UCG-003 of phylum Firmicutes has been found to be influential by applying IVI, Hubness and Spreading score logic (Figure 8), while investigating the gut bacterial communities in individuals with hypothalamic obesity. The most prominent bacterial genera in the gut of these individuals are Enterobacteriaceae (Proteobacteria), Bifidobacterium, and Lachnospiraceae (Firmicutes), collectively constituting around 45%-50% of the total gut bacteria. Enterobacteriaceae and Bifidobacterium contribute to complex carbohydrate digestion, vitamin production, and immune system regulation, with Bifidobacterium also renowned for its probiotic properties. Lachnospiraceae partakes in dietary fibre fermentation, short-chain fatty acid production, and immune system regulation. Delineating the most influential genus proves intricate, considering the abundance of Enterobacteriaceae, probiotic aspects of Bifidobacterium, and the diverse functions of Lachnospiraceae. However, variations in gut microbiota composition tied to diet, lifestyle, and medication use introduce diversity between individuals, which is not beneficial for gut health. The dominance of Proteobacteria, Firmicutes, and Bacteroidetes phyla in individuals with hypothalamic obesity underscores their significant roles in gut function and overall health.

3.2.5 Summary of Network analysis results

Groups	I V I	Hubness Score	Spreading Score
Healthy Lean Control	Gardnerella, Phylum: Actinobacteriota	Gardnerella, Phylum: Actinobacteriota	Gardnerella, Phylum: Actinobacteriota
Healthy Obese	Saccharimonadacea,Phylum: Patescibacteria	Saccharimonadaceae, Phylum: Patescibacteria	Saccharimonadacea,Phylum: Patescibacteria
Hypothalamic Lean	Fastidiosipila, Phylum: Firmicutes	Fastidiosipila, Phylum: Firmicutes	Fastidiosipila, Phylum: Firmicutes
Hypothalamic Obese	Lachnospiraceae_UCG−003, Phylum: Firmicutes	Lachnospiraceae_UCG−003, Phylum: Firmicutes	Lachnospiraceae_UCG−003, Phylum: Firmicutes

Table 1: Microbes of influence, achieved after performing IVI, Hubness score and Spreading score logic on the Phi-networks of each group.

3.3 Zeta diversity : distance decay, regression :

Zeta diversity	A metric that measures the number of shared species between multiple assemblages.
Zeta orders	The number of assemblages included in the calculation of zeta diversity
Zeta ratio	The probability of rediscovering a species with the addition of more samples or sites.
Exponential Regression	A type of decline that indicates that the decline is random.
Power law Regression	A type of decline that indicates that the decline is deterministic

Table 2: Key factors to interpret the ζ diversity behaviour of gut microbiota from the plotted graphs.

3.3.1 Healthy Lean control

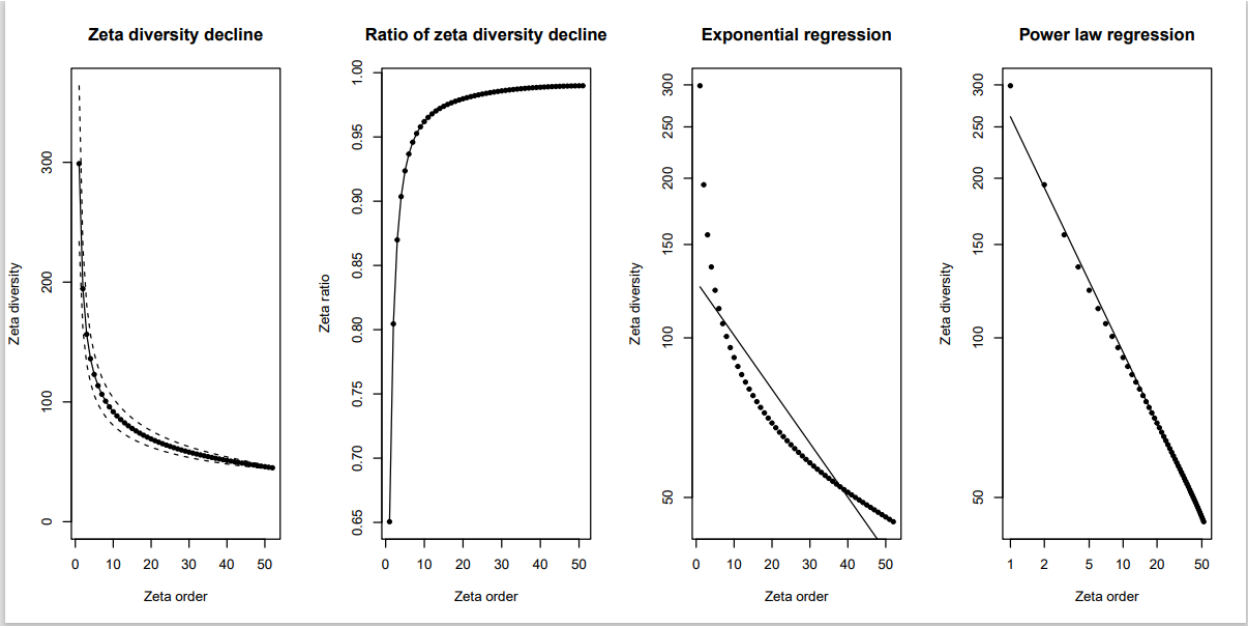

Figure 9 : ζ-diversity analysis result of Healthy lean control

The provided graph in figure 9 illustrates the zeta diversity of gut microbes in healthy lean control individuals, where zeta diversity quantifies shared species across multiple gut microbiota samples. The graph highlights a high zeta diversity, indicating a substantial number of shared species among the various samples and reflecting a robust and diverse gut microbiota. Such diversity signifies resilience and the capacity to provide essential nutrients to the body. Additionally, the graph depicts zeta diversity diminishing as the number of samples increases, which aligns with the expectation that rare species present in only a few samples become more likely to be encountered with greater sample size. Taking Table 2 into consideration, the graph underscores a healthy gut microbiota characterised by high zeta diversity and its expected decline with increasing sample numbers. As mentioned before in section 3.2, factors like diet, lifestyle, and medication usage can influence gut microbial zeta diversity, prompting further research to deepen our understanding of its implications for gut health.

3.3.2 Healthy Obese

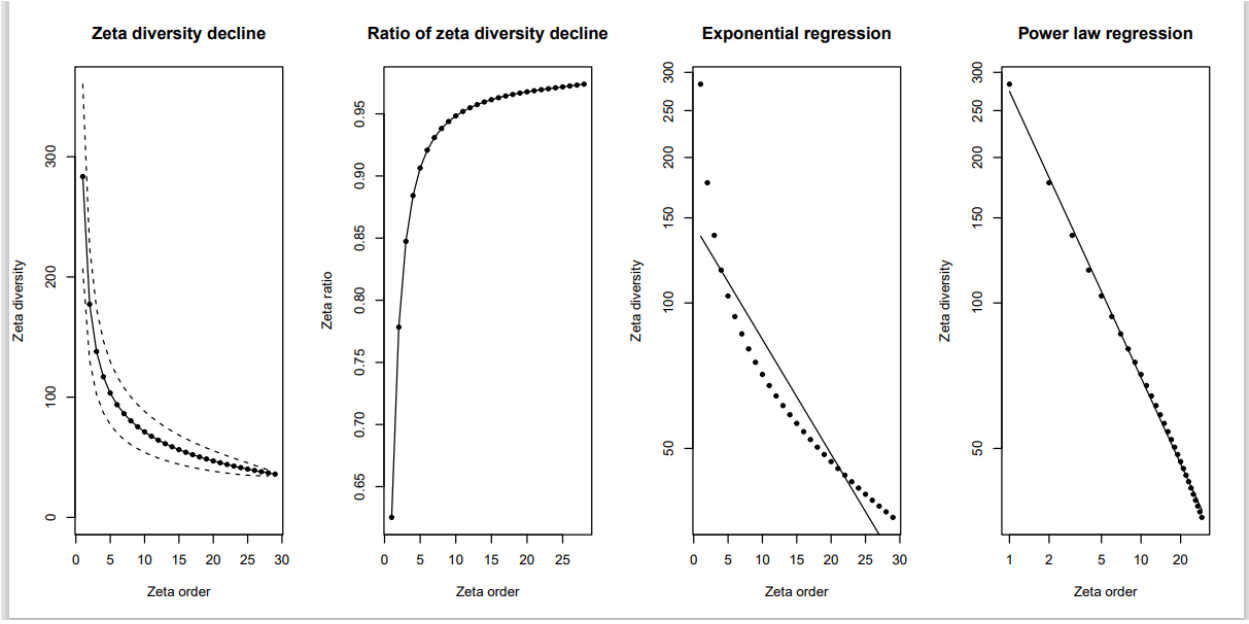

Figure 10 : ζ-diversity analysis result of Healthy obese

The above graph (Figure 10) showcases the zeta diversity of gut microbes within healthy obese individuals, where zeta diversity quantifies shared species among diverse gut microbiota samples with respect to Table 2. This graph demonstrates that the zeta diversity of gut microbes in healthy obese individuals is notably lower compared to that of healthy lean control individuals. This signifies a reduced count of shared species across the various gut microbiota samples, indicative of a less vibrant and resilient gut microbiota. Such diversity is pivotal for adaptability and nutrient provision. Furthermore, the graph indicates that the zeta diversity in healthy obese individuals diminishes more rapidly with increasing sample numbers compared to healthy lean controls. This suggests heightened sensitivity to change in the gut microbiota of healthy obese individuals, along with a lesser ability to sustain a diverse bacterial population.

3.3.3 Hypothalamic Lean

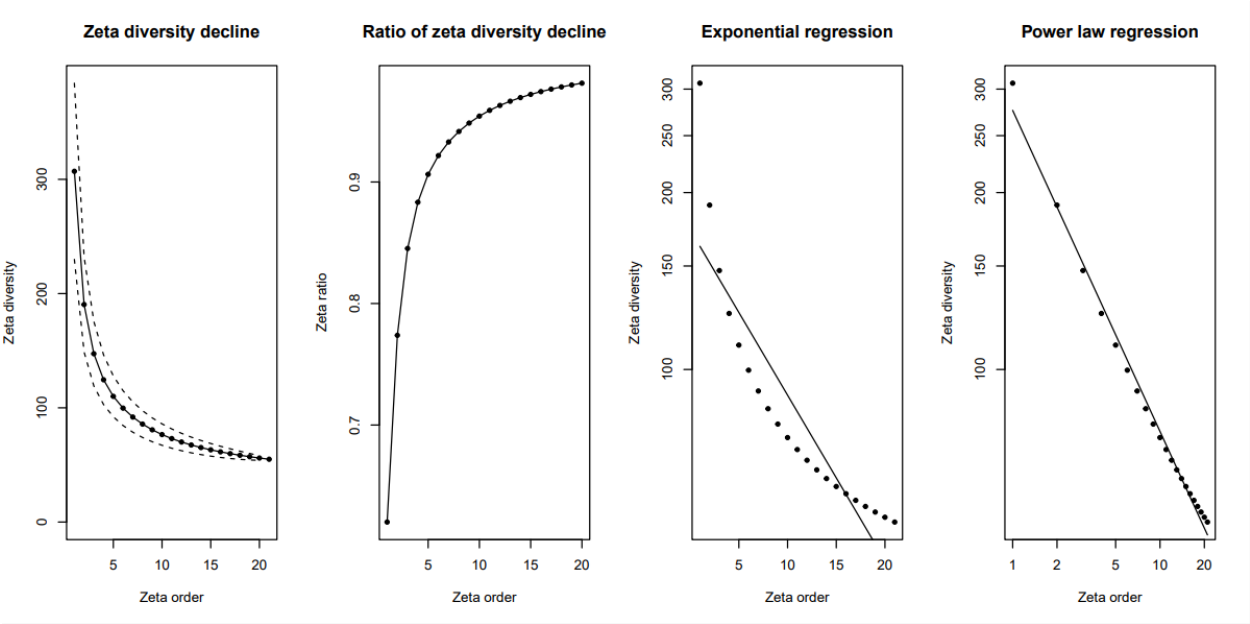

Figure 11 : ζ-diversity analysis result of Hypothalamic Lean

In Figure 11, the zeta diversity of gut microbes within hypothalamic lean individuals, using zeta diversity to quantify shared species among diverse gut microbiota samples. The graph displays that the zeta diversity of gut microbes in hypothalamic lean individuals surpasses that of healthy obese individuals, yet falls short of the zeta diversity in healthy lean control individuals. This implies a transitional state of the gut microbiota, where it demonstrates more shared species compared to healthy obese individuals but fewer than healthy lean controls. Moreover, the graph indicates that the zeta diversity of gut microbes in hypothalamic lean individuals decreases more sharply with increasing sample numbers than in healthy lean control individuals, but less steeply compared to healthy obese individuals. This observation suggests that the gut microbiota in hypothalamic lean individuals is more responsive to changes compared to healthy lean controls, yet less responsive than healthy obese individuals. Hence it can be inferred that the gut microbiota in hypothalamic lean individuals is in a state of transition, exhibiting more diversity than in healthy obese individuals but less than in healthy lean controls and also that at the higher zeta diversity in hypothalamic lean individuals signifies their more diverse gut microbiota in comparison to healthy obese individuals. Lastly, the graph's depiction of a steeper decline in zeta diversity with increasing sample numbers further suggests that the gut microbiota in hypothalamic lean individuals is more receptive to change than in healthy lean controls.

3.3.4 Hypothalamic Obese

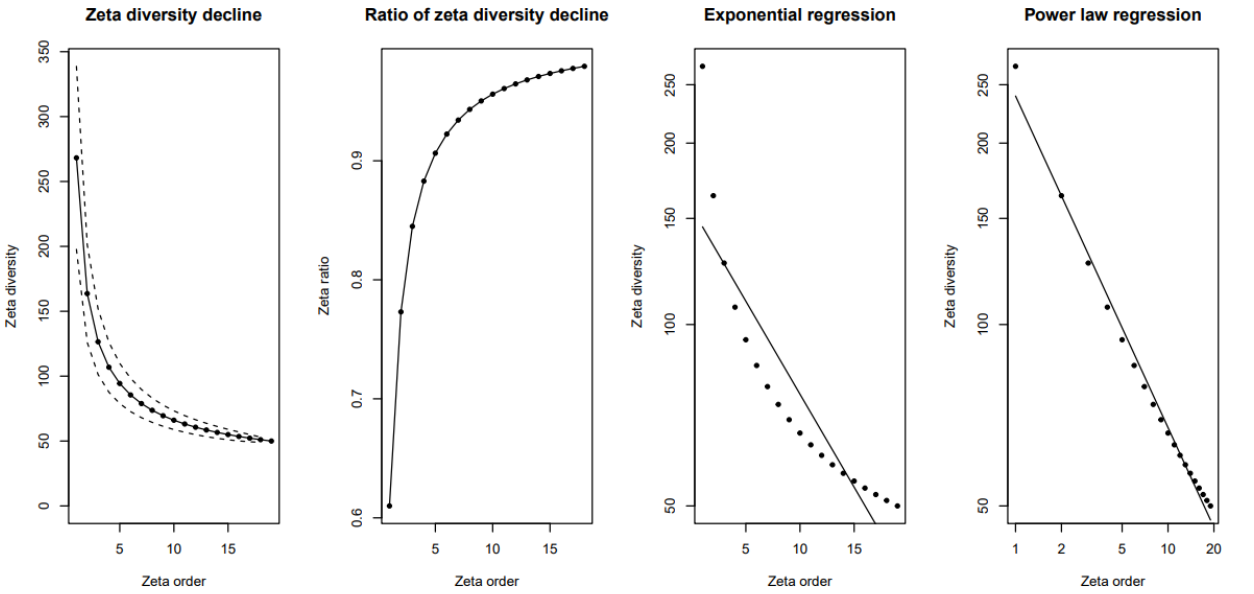

Figure 12 : ζ-diversity analysis result of Hypothalamic obese

The resulting graph delineates the zeta diversity of gut microbes within hypothalamic obese individuals (Figure 12), utilising zeta diversity as a metric for shared species among diverse gut microbiota samples. The graph demonstrates that the zeta diversity of gut microbes in hypothalamic obese individuals mirrors that of healthy obese individuals. This implies a comparable count of shared species among the different gut microbiota samples in both groups, indicating a less diverse gut microbiota compared to that of healthy lean control individuals. However, it signifies a distinction that is not as pronounced as observed in healthy obese individuals. Additionally, the graph reveals that the zeta diversity of gut microbes in hypothalamic obese individuals experiences a similar rate of decline with increasing sample numbers as seen in healthy obese individuals. This observation implies that the gut microbiota in hypothalamic obese individuals responds to changes in a manner akin to that of healthy obese individuals. In summary, the graph underscores the similarity in zeta diversity of gut microbes between hypothalamic obese individuals and healthy obese individuals. This suggests that the gut microbiota in hypothalamic obese individuals is characterised by a diversity level intermediate between that of healthy lean control individuals and healthy obese individuals.

3.2.5 Summary of Diversity analysis results

The observed pattern of Zeta diversity diminishing as higher orders are considered signifies shifts in the composition of communities. The notable steep decline witnessed in the hypothalamic lean/obese groups suggests a heightened susceptibility to environmental influences. Conversely, the relatively consistent levels in healthy cases highlight the resilience of these communities when subjected to well-maintained conditions. The concept of ratio offers a complementary perspective, indicating a stabilising trend beyond the 25th order. Notably, this decline in Zeta diversity conforms to a power law pattern across all examined groups. This characterization underlines the systematic relationship governing the decrease in diversity with increasing orders and offers insights into the intricate dynamics of community composition and stability across varying conditions.

Chapter IV: Discussion

The results derived from the Phi network analysis have illuminated critical insights into the complex dynamics of gut bacterial communities across different health states. In the case of healthy lean individuals, the network analysis revealed the prominence of diverse bacterial genera, alongside their respective phylum and class classifications. Notably, the genus Gardnerella emerged as a highly influential node within the network, boasting exceptional scores of 100 in Spreading, Hubness, and IVI metrics. Furthermore, the dominant presence of Lachnospiraceae and Bifidobacterium genera, both belonging to the Firmicutes phylum, underscores their pivotal roles in dietary fibre fermentation, short-chain fatty acid production, and beneficial impact on health [29].

Shifting focus to healthy obese individuals, the prevalent bacterial genera E. coli, Enterococcus, and Bacteroides emerged as the key players, constituting around half of the gut bacteria in this group. These genera contribute indispensably to dietary fibre fermentation, short-chain fatty acid production, and immune system regulation, collectively supporting gut health within the context of obesity [30].

The investigation of individuals with hypothalamic lean conditions illuminated the profound impact of the microbe Fastidiosipila, belonging to the Firmicutes phylum [31]. This microorganism demonstrated an extraordinary score of 100 in IVI, Hubness, and Spreading metrics, solidifying its influence across multiple facets within the microbial network. Among the prevalent bacterial genera in this group, Akkermansia, Ruminococcus, and Bacteroides showcased their significance by collectively constituting more than half of the total gut bacteria, each contributing substantially to various aspects of gut health and immune regulation.

Lastly, in individuals with hypothalamic obesity, the prominence of Lachnospiraceae_UCG-003, a genus within the Firmicutes phylum, emerged as a central influencer based on IVI, Hubness, and Spreading scores. Within this group, Enterobacteriaceae, Bifidobacterium, and Lachnospiraceae took the lead in gut bacteria composition, contributing crucially to functions such as complex carbohydrate digestion, vitamin production, immune system regulation, and

more. However, the intricacies of their interplay underscore the need for further exploration in understanding the roles of these genera within the context of obesity. The dominance of Verrucomicrobia, Firmicutes, and Bacteroidetes phyla in lean adults with hypothalamic obesity underscores their crucial roles in gut function and overall health. Alongside the dominant genera, Bifidobacterium, Lactobacillus, Faecalibacterium, Blautia, and Roseburia contribute to vital functions

The discussion of the zeta diversity patterns within the different groups provides valuable insights into the dynamics of gut microbiota composition across various health states. Examining the zeta diversity graph of healthy lean individuals (Figure 9), it becomes evident that there is a substantial level of shared species across the gut microbiota samples. This high zeta diversity signifies a resilient and diverse gut microbiota, capable of supplying essential nutrients to the body. Furthermore, the declining trend in zeta diversity with increasing sample numbers aligns with the expectation that rare species are more likely to be encountered with a larger sample size. This pattern emphasizes the healthy state of the gut microbiota and highlights the potential influence of factors like diet, lifestyle, and medication usage on zeta diversity.

Contrasting the healthy lean group, the zeta diversity graph of healthy obese individuals (Figure 10) showcases a noticeable reduction in shared species among gut microbiota samples. This lower zeta diversity indicates a less vibrant and adaptable gut microbiota, essential for nutrient provisioning. The steeper decline in zeta diversity with increasing samples suggests heightened sensitivity to changes, revealing potential vulnerabilities in maintaining a diverse bacterial population in the context of obesity.

Turning to the hypothalamic lean group, the zeta diversity graph (Figure 11) demonstrates a transitional state of the gut microbiota. Although surpassing the zeta diversity of healthy obese individuals, it falls short of that in healthy lean controls. The sharper decline in zeta diversity with increasing samples compared to healthy lean controls and healthier responsiveness than healthy obese individuals implies a state of transition in the gut microbiota, showcasing more diversity than observed in obesity but less than in a healthy state.

Shifting focus to the hypothalamic obese group (Figure 12), the zeta diversity pattern mirrors that of healthy obese individuals, indicating a comparable count of shared species. This signifies a less diverse gut microbiota in both groups, while highlighting the intermediary nature of the gut microbiota in hypothalamic obese individuals.

In summary, the observed patterns of zeta diversity alterations across groups indicate shifts in community composition influenced by environmental factors. The pronounced decline in hypothalamic lean and obese groups suggests susceptibility to external influences, while the relatively stable levels in healthy cases point to the robustness of well-maintained conditions. The consistent power law pattern of declining diversity with increasing orders offers a systematic view of community dynamics across different health states, shedding light on the intricate relationship between diversity, composition, and stability.

Chapter V: Conclusion

The concept of spreading and hubness scores is a useful way to measure the influence of a node in a network. Spreading score measures how likely a node is to spread information to its neighbours, while hubness score measures how many other nodes are connected to a node. The multiplicative product of spreading and hubness scores is a measure of the overall influence of a node in the network.In the context of gut microbiota research, the multiplicative product of spreading and hubness scores can be used to identify nodes that are likely to be important in the development of obesity. The results in section 3.2 show that the phylum Firmicutes has a higher multiplicative product of spreading and hubness scores in the Healthy Obese, Hypoth. Lean, and Hypoth. Obese groups than in the Healthy Lean group. This suggests that Firmicutes may be an important player in the development of obesity.The phylum Firmicutes is a diverse group of bacteria that includes many species that are known to be involved in carbohydrate metabolism. These bacteria are able to ferment carbohydrates into short-chain fatty acids, which can be used as a source of energy. However, Firmicutes can also produce compounds that promote inflammation, which can contribute to obesity. The results in section 3.2.1 and 3.2.2 also show that there is a significant difference in the abundance of two genera, Ruminococcus and Bacteroides, between the Healthy Lean Control and Healthy Obese groups. Ruminococcus is a genus of bacteria that is known to produce short-chain fatty acids, while Bacteroides is a genus of bacteria that is known to produce compounds that promote inflammation. The higher abundance of Ruminococcus and the lower abundance of Bacteroides in the Healthy Obese group suggests that a shift in the composition of the gut microbiota may be a contributing factor to obesity. Overall, the results of this study suggest that the multiplicative product of spreading and hubness scores can be a useful tool for identifying nodes that are important in the development of obesity. The results also suggest that a shift in the composition of the gut microbiota, specifically the abundance of Firmicutes and Ruminococcus, may be a contributing factor to obesity. In addition to the factors mentioned above, it is important to note that obesity is a complex disease that is influenced by a variety of factors, including genetics, diet, and lifestyle. More research is needed to fully understand the role of the gut microbiota in obesity and to develop effective interventions for treating this disease. In conclusion, the analysis of zeta diversity, the health groups provides a comprehensive understanding of the gut microbiota's

dynamic nature under different conditions. Healthy lean individuals exhibit a robust and diverse gut microbiota with high zeta diversity, reflecting resilience and nutrient supply capacity. Conversely, healthy obese individuals display reduced zeta diversity, implying a less adaptable microbiota. The hypothalamic lean group presents a transitional state, demonstrating intermediate zeta diversity, suggesting ongoing changes in response to altered health conditions. The zeta diversity of hypothalamic obese individuals aligns with that of healthy obese counterparts, emphasizing a compromised diversity in obesity-related contexts. These findings underscore the significance of environmental factors, such as diet and lifestyle, in shaping gut microbiota diversity and composition. The consistent power law pattern in declining zeta diversity as higher orders are considered highlights the systematic relationship between diversity and community structure. Overall, this study enhances our knowledge of the complex interplay between health status, gut microbiota diversity, and stability, paving the way for further research to unravel the underlying mechanisms and potential interventions for maintaining optimal gut health.

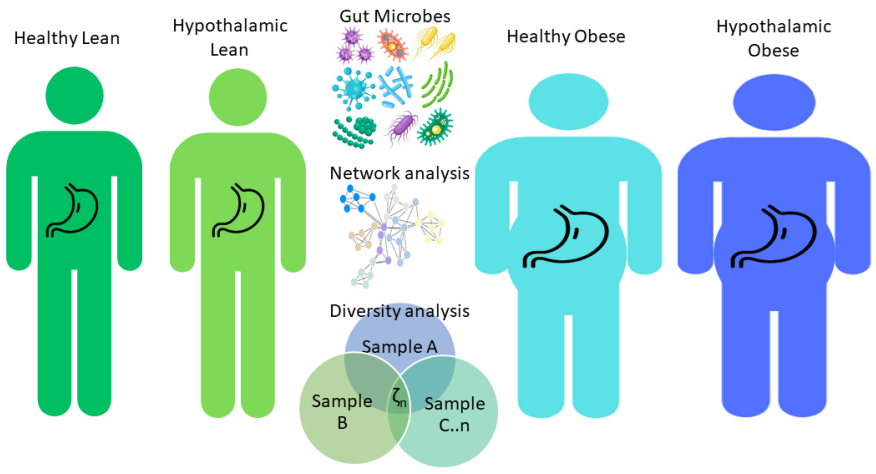

Figure 13: Schematic conclusion of the conducted study

Chapter VI: Future work

Future research in this field should focus on several key directions to build upon the insights gained from this study. Firstly, the concept of spreading and hubness scores as a combined metric for node influence holds promise beyond the scope of obesity-related microbiota. Exploring its application in other health conditions and networks could unveil novel insights into various biological systems. Moreover, investigating the intricate interplay between Firmicutes and obesity warrants a deeper investigation. Elucidating the specific mechanisms through which Firmicutes contribute to carbohydrate metabolism and inflammation, and their potential modulation, could provide valuable targets for therapeutic interventions aimed at mitigating obesity.

Furthermore, the observed compositional shifts in bacterial genera like Ruminococcus and Bacteroides call for in-depth studies to dissect their exact roles in obesity development. Conducting longitudinal studies to monitor the temporal dynamics of gut microbiota alterations in response to changing health conditions could provide a more nuanced understanding of their contributions. Integrating genetic, dietary, and lifestyle factors into the analysis could offer a more comprehensive view of obesity's multifaceted origins and inform personalised treatment strategies.

Lastly, advancements in sequencing technologies and computational tools open the door to larger-scale and more detailed investigations. High-throughput omics data, such as metagenomics and metabolomics, could provide a deeper understanding of the functional capabilities of gut microbiota in obesity. Leveraging machine learning and network analysis techniques could reveal hidden patterns and interactions within the complex microbial ecosystem.

Hence, this study lays the foundation for future inquiries into understanding and addressing the intricate relationship between gut microbiota, health conditions, and diversity. The potential applications extend beyond obesity and necessitate collaborative efforts across disciplines to unlock the full potential of microbiota research for human health.

References

[1] World Health Organization (2021). *Obesity and Overweight*. [online] World Health Organization. Available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

[2] Kim SJ, Cho SY, Jin DK. Prader-Willi syndrome: an update on obesity and endocrine problems. Ann Pediatr Endocrinol Metab. 2021 Dec;26(4):227-236. doi: 10.6065/apem.2142164.082. Epub 2021 Dec 31. PMID: 34991300; PMCID: PMC8749024.

[3] Langley‐Evans, S.C., Pearce, J. and Ellis, S. (2022). Overweight, obesity and excessive weight gain in pregnancy as risk factors for adverse pregnancy outcomes: A narrative review. *Journal of Human Nutrition and Dietetics*, 35(2). doi:https://doi.org/10.1111/jhn.12999.

[4] Kim, Ju Young. "Optimal Diet Strategies for Weight Loss and Weight Loss Maintenance." *Journal of Obesity & Metabolic Syndrome*, 30 Mar. 2021, www.ncbi.nlm.nih.gov/pmc/articles/PMC8017325/.

[5] Khan, M.J. (2014a). *Gut microbiota in obesity of different aetiology: cause or effect?* [online] theses.gla.ac.uk. Available at: https://theses.gla.ac.uk/6299/.

[6] Tremaroli, V. and Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, 489(7415), pp.242–249. doi:https://doi.org/10.1038/nature11552.

[7] Geng, J., Ni, Q., Sun, W., Li, L. and Feng, X. (2022). The links between gut microbiota and obesity and obesity related diseases. *Biomedicine & Pharmacotherapy*, 147, p.112678. doi:https://doi.org/10.1016/j.biopha.2022.112678.

[8] Liu, B.-N., Liu, X.-T., Liang, Z.-H. and Wang, J.-H. (2021). World Journal of Gastroenterology Gut microbiota in obesity Manuscript source: Invited manuscript. *World J Gastroenterol*, [online] 27(25), pp.3837–3850. doi:https://doi.org/10.3748/wjg.v27.i25.3837.

[9] Thomas, Torsten et al. "Metagenomics - a guide from sampling to data analysis." *Microbial informatics and experimentation* vol. 2,1 3. 9 Feb. 2012, doi:10.1186/2042-5783-2-3

[10] Gao, B., Chi, L., Zhu, Y., Shi, X., Tu, P., Li, B., Yin, J., Gao, N., Shen, W. and Schnabl, B. (2021). An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. *Biomolecules*, [online] 11(4), p.530. doi:https://doi.org/10.3390/biom11040530.

[11] Ciuffreda, L., Rodríguez-Pérez, H. and Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Computational and Structural Biotechnology Journal*, 19, pp.1497–1511. doi:https://doi.org/10.1016/j.csbj.2021.02.020.

[12] E. Bolyen *et al.*, "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2," *Nature Biotechnology*, vol. 37, no. 8, pp. 852–857, Jul. 2019, doi: https://doi.org/10.1038/s41587-019-0209-9.

[13]"Release 138," *www.arb-silva.de*. https://www.arb-silva.de/documentation/release-138 (accessed Aug. 13, 2023).

[14] B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes, "DADA2: High-resolution sample inference from Illumina amplicon data," *Nature Methods*, vol. 13, no. 7, pp. 581–583, May 2016, doi: https://doi.org/10.1038/nmeth.3869.

[15] Lovell D, Pawlowsky-Glahn V, Egozcue JJ (2013) Have you got things in proportion? a practical strategy for exploring association in high-dimensional compositions. In: Hron K, Filzmoser P, Templ M, editors, Proceedings of the 5th International Workshop on Compositional Data Analysis. Vorau, Austria, pp. 100–110.

[16] J. Aitchison, "The Statistical Analysis of Compositional Data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 2, pp. 139–177, 1982, Available: https://www.jstor.org/stable/2345821

[17] D. Lovell, V. Pawlowsky-Glahn, J. J. Egozcue, S. Marguerat, and J. Bähler, "Proportionality: A Valid Alternative to Correlation for Relative Data," *PLOS Computational Biology*, vol. 11, no. 3, p. e1004075, Mar. 2015, doi: https://doi.org/10.1371/journal.pcbi.1004075.

[18] A. Salavaty, M. Ramialison, and P. D. Currie, "Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks," *Patterns*, vol. 1, no. 5, p. 100052, Aug. 2020, doi: https://doi.org/10.1016/j.patter.2020.100052.

[19] Han, Jiawei, Jian Pei, and Hanghang Tong. *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[20]“spreading.score function - RDocumentation,” *www.rdocumentation.org*. https://www.rdocumentation.org/packages/influential/versions/2.2.6/topics/spreading.score (accessed Aug. 13, 2023).

[21] “hubness.score: Hubness score in influential: Identification and Classification of the Most Influential Nodes,” *rdrr.io*. https://rdrr.io/cran/influential/man/hubness.score.html (accessed Aug. 13, 2023).

[22] “ivi: Integrated Value of Influence (IVI) in influential: Identification and Classification of the Most Influential Nodes,” *rdrr.io*. https://rdrr.io/cran/influential/man/ivi.html (accessed Aug. 13, 2023).

[23] M. A. McGeoch *et al.*, “The application of zeta diversity as a continuous measure of compositional change in ecology,” *USC Research Bank (University of the Sunshine Coast)*, Nov. 2017, doi: https://doi.org/10.1101/216580.

[24] “Zeta diversity,” *Wikipedia*, Mar. 15, 2021. https://en.wikipedia.org/wiki/Zeta_diversity (accessed Aug. 14, 2023).

[25] McGeoch, Melodie A et al. “Measuring continuous compositional change using decline and decay in zeta diversity.” *Ecology* vol. 100,11 (2019): e02832. doi:10.1002/ecy.2832

[26] C. Hui and M. A. McGeoch, “Zeta Diversity as a Concept and Metric That Unifies Incidence-Based Biodiversity Patterns,” *The American Naturalist*, vol. 184, no. 5, pp. 684–694, Nov. 2014, doi: https://doi.org/10.1086/678125.

[27] P. J. M. at gmail.com >, S. H. at stat.stanford.edu >, with contributions from G. Jordan, and S. Chamberlain, “phyloseq: Handling and analysis of high-throughput microbiome census data,” *Bioconductor*, 2023. https://www.bioconductor.org/packages/release/bioc/html/phyloseq.html

[28] A. O’Callaghan and D. van Sinderen, “Bifidobacteria and Their Role as Members of the Human Gut Microbiota,” *Frontiers in Microbiology*, vol. 7, no. 925, Jun. 2016, doi: https://doi.org/10.3389/fmicb.2016.00925.

[29] Kilic Yildirim G, Dinleyici M, Vandenplas Y, Dinleyici EC. Effects of synbiotic supplementation on intestinal microbiota composition in children and adolescents with exogenous obesity: (Probesity-2 trial). Gut Pathog. 2023 Jul 21;15(1):36. doi: 10.1186/s13099-023-00563-y. PMID: 37474971; PMCID: PMC10360342.

[30] Pérez-Rubio Á, Soluyanova P, Moro E, Quintás G, Rienda I, Periañez MD, Painel A, Vizuete J, Pérez-Rojas J, Castell JV, Trullenque-Juan R, Pareja E, Jover R. Gut Microbiota and Plasma Bile Acids Associated with Non-Alcoholic Fatty Liver Disease Resolution in Bariatric Surgery Patients. Nutrients. 2023 Jul 18;15(14):3187. doi: 10.3390/nu15143187. PMID: 37513605; PMCID: PMC10385764.
\
[31] Cao MZ, Wei CH, Wen MC, Song Y, Srivastava K, Yang N, Shi YM, Miao M, Chung D, Li XM. Clinical efficacy of weight loss herbal intervention therapy and lifestyle modifications on obesity and its association with distinct gut microbiome: A randomized double-blind phase 2 study. Front Endocrinol (Lausanne). 2023 Mar 22;14:1054674. doi: 10.3389/fendo.2023.1054674. PMID: 37033234; PMCID: PMC10073537.

Appendix

Function for Spreading score :

{spreading.score(graph, vertices = V(graph), weights = NULL,directed = FALSE, mode = "all" loops = TRUE, d = 3, scaled = TRUE)} [20]

Function for Hubness score :

{hubness.score(graph, vertices = V(graph), directed = FALSE, mode = "all", loops = TRUE, scaled = TRUE, verbose = FALSE)} [21]

Function for IVI :

{ivi(graph, vertices = V(graph), weights = NULL, directed = FALSE, mode = "all", loops = TRUE, d = 3, scaled = TRUE, ncores = "default", verbose = FALSE)} [22]