Identifying hub nodes and sub-networks from cattle rumen microbiome multilayer networks

Mengyuan Wang[1], Haiyang Wang[1], Huiru Zheng[1]*, Richard J. Dewhurst[2], Rainer Roehe[2]

[1] School of Computing, Ulster University, United Kingdom
[2] Scotland's Rural College, Edinburgh, United Kingdom
h.zheng@ulster.ac.uk

Abstract.

Purpose: The purpose of this research is to represent the complex system of the rumen microbiome through a multi-layer network, to explore the main biological functions related to methane metabolism.

Methods: A three-layer heterogeneous network has been constructed based on rumen metabolites, rumen microbial genes and rumen microbial communities. Node association is calculated based on linear and non-linear association p-value voting combined with the permutation test. This research proposed a method to generate key hubs based on the topological properties of nodes in a multilayer network.

Results: 1) A total of 59 individuals (including metabolites, microbial genera, and microbial genes) were found in the topological hubs. Among them, 23 individuals appeared in more than one topologically ranked hub. 2) The metabolite with the highest topological centrality is methane, which is the top one in the 9 topological property rankings. Among the top 10 topologically central 35 microbial genes, based on the reconstruction of the metabolic network, the most involved metabolic pathway is methane metabolism which included 14 microbial genes. 3) There are 3 genera of microbes (including *Aeropyrum*, *Desulfurococcus* and *Thermosphaera*) from the same family and are ranked top 10 in at least 2 topological characteristics. In the topological hub of the Bottleneck, we found that from *Aeropyrum* and *Desulfurococcus* of the same family, their indirect paths are associated with methane. And SufD (K09015), a protein related to the family in the same hub, was also found to be positively associated with methane.

Conclusion: The nodes with important topological properties at the metabolite level and the microbial level are associated with methane functions, indicating the consistency of the biological functions of the topological properties between the layers. This method provided a solution for extracting related hubs of highly central nodes and shows potential in the understanding of biological functions. The topological property-Bottleneck showed the most potential by identified hub nodes were found as exhibiting the consistent sulfur reduction functions throughout multilayer networks.

1   Introduction

Dairy and beef cattle are responsible for 9.5 percent of all anthropogenic greenhouse gas emissions, with ruminal microbial fermentation accounting for 40-50 percent of this total [1]. The rumen microbiome is a complex system closely related to the host and regulated by multiple factors [2]. The exploration of the mechanism of microbial methane metabolism in the rumen is key to reducing the greenhouse effect and improving animal nutrition. There have been some network studies of the rumen microbiome, including the identification of biomarkers of methane emission and diet digestibility based on the correlation network of microbial gene abundance [3-5]. Other studies have used microbial taxonomy abundance correlated network community niche exploration and analysis [6], metabolite-related diffusion analysis based on a multi-layer network [7] or network analysis related to hosting phenotype and genetic characteristics of rumen methane emission [8]. However, most of the current research is based on single-layer networks.

Research in this field has urgent to find a way which can integrate comprehensive information to represent the complex host-microbial metabolic system. The multilayer network is through the embedded representation of the interaction structure of the complex system, to provide a research approach to understand the nature and function of the complex system. A complex system is composed of individuals which interact non-linearly in static or dynamic ways. These interactions make it produce various forms of association structure and represent the system's multi-scale characteristics [9]. The identification and prediction of the influence of nodes in complex networks is a hot research area in complex networks [9], which helps us understand the internal structural characteristics of biological complex systems [10]. Studies have shown that the topological properties and functions of proteins in protein networks the degree of importance are significantly related [10]. However, the identification and prediction of node influence are limited to specific network topology. Once the method is not significant in the corresponding topology, then the accuracy of the recognition and prediction of the method is questionable [11]. Moreover, the topological properties and the biological significance of multi-layer network embedding for the rumen complex system are also unknown.

This research mainly develops a method for identifying and analyzing key hubs based on multi-layer network topology ranking. The main innovations and contributions of this research are: 1) The results of the study found important topological nodes with consistent biological functions between the metabolite layer/microbial community layer and the microbial gene layer (e.g., methane and its microbial functional genes). Further verified the multi-layer network method is practical to represent the complex system of rumen microbiome, which is of great significance for the exploration and understanding of the correlation pattern of the rumen microbiome. 2) In this research, the unknown and uncertainty about the characteristics of the multi-layer network structure and the driving biological mechanism, as well as the huge dense connections of the

target node, make the mining of key modules more challenging. The methodology applied in the study showed huge potential to conquer the challenge by extracting the key hubs consisting of the heterogeneous nodes involved in the same biological metabolism.

2   Method

In this research, a framework for extracting key topological hubs based on the ruminal metabolome and microbiome multilayer network was developed.

Dataset: The 30 beef cattle rumen fluid samples used in this study were carried out at Scotland's Rural College's Beef and Sheep Research Centre (SRUC) [6]. The SRUC Animal Experiment Committee permitted the experiment, which was conducted in compliance with the Animals (Scientific Procedures) Act of 1986 in the United Kingdom. The rumen fluids come from samples of cattle breeds, diets, additives, and feed efficiency that is balanced, and have been analyzed by metabolomics and metagenomics. Refer to the following papers [7-8] for the experimental process and steps.

- Metabolite dataset: A total of 119 rumen metabolites were identified and quantified in the NMR-based targeted metabolomics experiment.
- Metagenomics dataset: The relative concentration of 119 rumen metabolites, the relative abundance of 1461 microbial genes (>0.01%), and the relative abundance of 1178 microbial genera (excluding zero samples) were selected.

Construction of multilayer network:

The study constructed a multi-layer network consisting of microbial genera, rumen metabolites, and microbial genes. First, nodes of the same type are connected by a linear or non-linear combination of correlations with each other to form a single-layer network. Secondly, the metabolite-microbial gene, metabolite-microbial genera linear or non-linear combination of correlation is used to connect each layer separately. The measurements and procedure of edges can refer to the following article [10-12]. The calculation and visualization are processed by Cytoscape and the plug-in CoNet [12].

The metabolite layer was used to reflect the results of rumen microbes involved in host digestion and metabolism. The interaction layer between metabolites and microbial genes is expected to represent the metabolic results of microbial function mapping. The interaction between metabolites and the microbial community layer is expected to reflect the composition and interaction of the microbial community.

Hub collection based on topological properties:

We selected the top 10 nodes of each topological property and generated the shortest connection path between them to get the sub-networks. The topological properties used to filter the key nodes are listed in Table 1. The calculation process is completed by Cytoscape plug-in cytoHubba [13]. The subsequent metabolite and microbial gene biological function analysis was completed using the tools of the KEGG mapper.

Table 1. List of topological properties.

Topological properties
Degree

Edge Percolated Component (EPC)
Maximum Neighbourhood Component (MNC)
Density of Maximum Neighbourhood Component
(DMNC)
Maximal Clique Centrality (MCC)
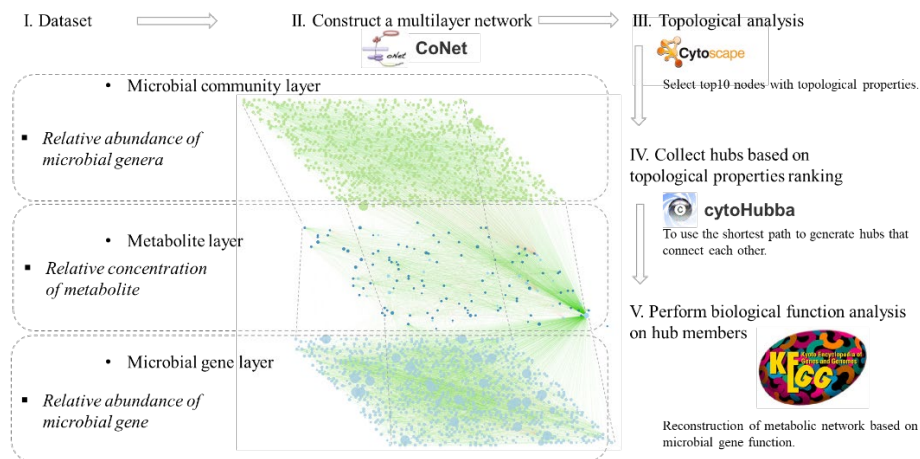Bottleneck
Radiality
Betweenness
Stress
Closeness

Fig. 1. Research pipeline.

3 Results

In this session, the hub composed of the top 10 nodes of each topological property was extracted from the constructed multilayer network, and the members in the hub were analyzed with biological functions. Among them, Degree, MNC, DMNC and MCC are local-based measures, that is, only the direct neighborhood of a node is considered. Closeness, Radiality, Bottleneck, Stress, Betweenness are based on the shortest path, while EPC is based on the theory of percolation.

3.1 Multilayer network

The final multi-layer network (see Fig. 2) contains 2092 nodes and 9770 edges. Topological information is summarized in Table 2. The network finally contains 981 microbial genera, 994 microbial genes and 117 metabolites.
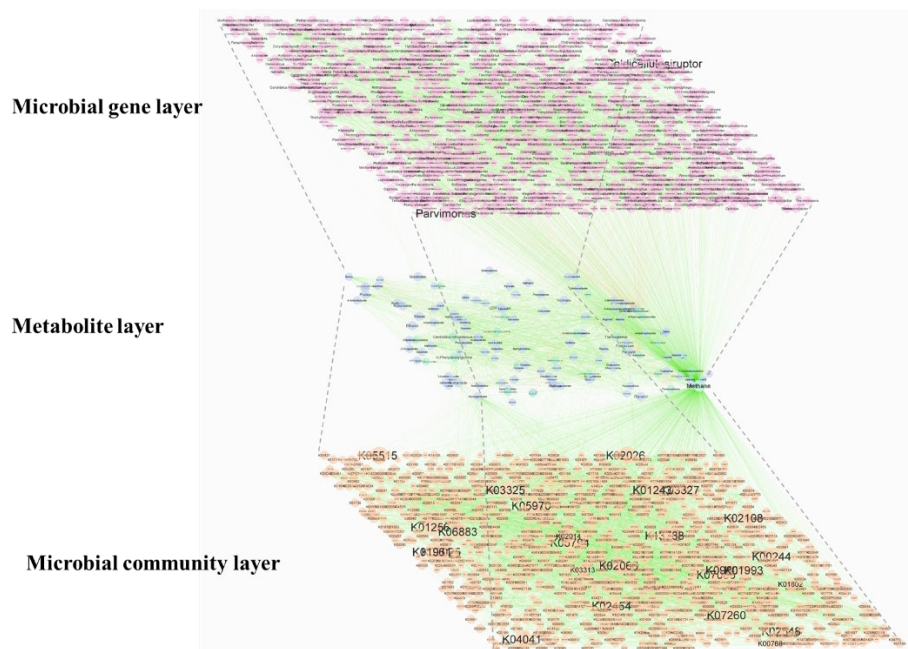
Fig. 2. Multilayer network (The orange nodes represent microbial genes. The purple nodes represent metabolites. The blue node represents the microbial community. The green edges indicate positive associations, and the red edges indicate negative associations).

Table 2. Topological properties of multilayer networks.

| Item	Value
Number of nodes	2092
Number of edges	9770
Average number of neighbours	9.453
Network diameter	15
Network radius	8
Characteristic path length	3.279
clustering coefficient	0.306
Network density	0.005
Network heterogeneity	3.426
Network centralization	0.639
Connected components	14

3.2   Node hub ranked based on topological properties

This research selects the topological properties of 10 nodes (Table 2) to analyze the network. A total of 59 individuals were obtained from the top 10 topological rankings of each feature, of which 23 individuals appeared more than once in the top 10 topological rankings. All the hubs include 35 microbial genes, 6 metabolites and 18 microbial genera.

The members of the Betweenness, Degree, MNC and Stress modules are not completely directly connected. The relationship between them is shown in Fig. 3. through indirect edges. Bottleneck, Closeness, EPC and Radiality are all methane-centric hubs. In these hubs, methane is only positively associated with other nodes. As the highest node in the network, methane connects 1327 nodes.

It is worth noting that except for DMNC, methane ranks top in nine topological rankings (Fig. 3). Methane becomes the most important node in the multilayer network. The microbial gene K01810 is in the top 5 of the rankings of five topological properties, including Closeness, Degree, EPC, Radiality and MNC. There are two microbial genera *Aeropyrum* and *Desulfurococcus*, which are ranked top 5 in the rankings of Betweenness, Bottleneck, Degree and Stress (Fig. 3). They are two genera belonging to the *Desulfurococcaceae* family. *Thermosphaera* is also a member of the *Desulfurococcaceae* family, which is included of the top 5 in Betweenness and Stress, respectively.

There are only microbial genes in the hub extracted by DMNC, and these microbial genes only appear in this topological hub. The hubs extracted by EPC and MCC are only methane and microbial genes, and they have three common members. There are four common members in the hubs of Betweenness, Stress and Bottleneck. The hub of Betweenness and Bottleneck includes three types of nodes. The stress hub does not contain microbial genes. There are eight common members in the hub of Radiality and Closeness. Degree and MNC have eight members in common.
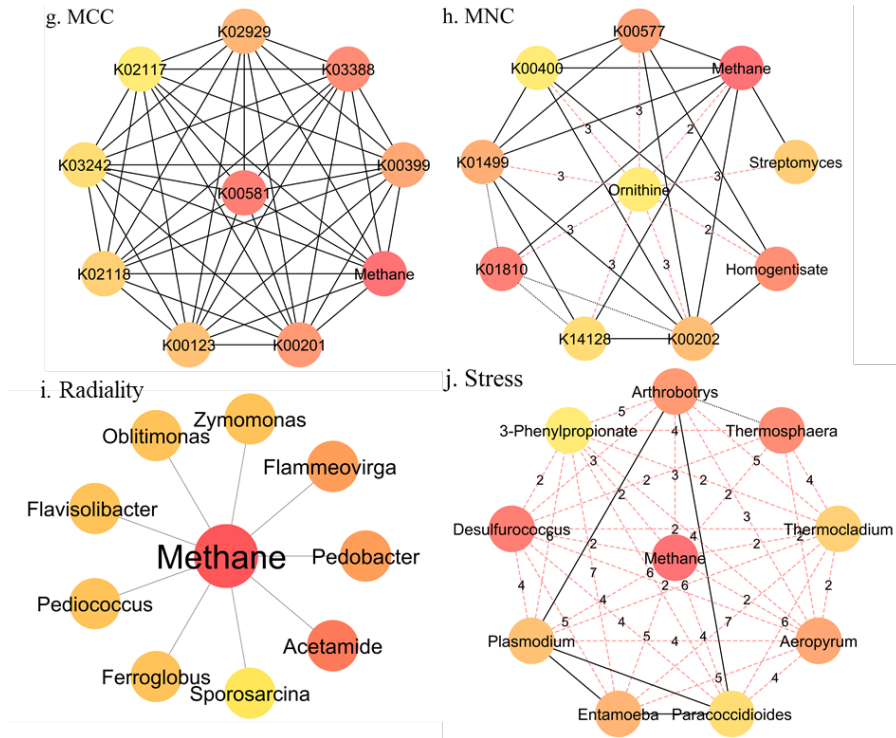
a. Betweeness

b. BottleNeck

c. Closeness

d. Degree

e. DMNC

f. EPC
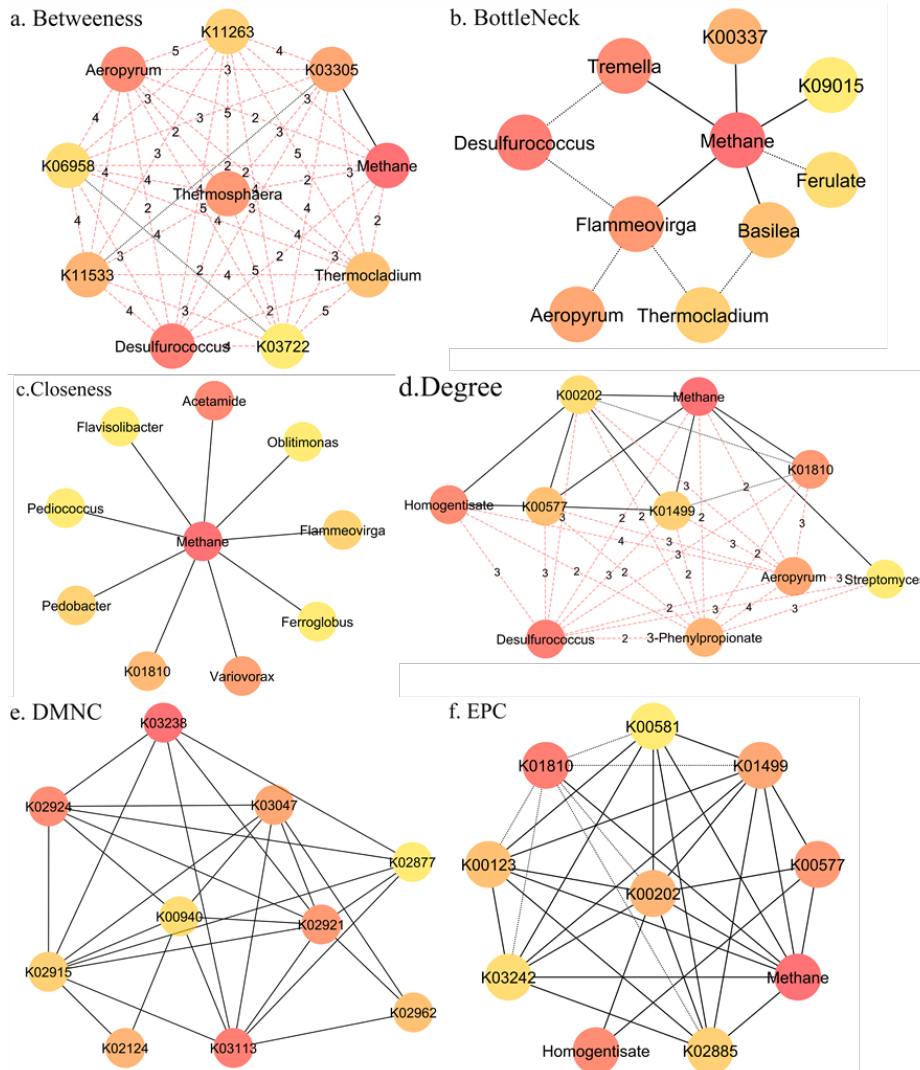
Fig. 3. Hub based on topological properties (The color of the node closer to red indicates the higher the topological property ranking. The color of the node closer to yellow indicates the lower the topological property ranking. The black line represents the direct connection, the solid line represents the positive connection, and the dotted line represents the negative connection. The red line represents the indirect connection, and the number on the line represents the number of steps between two nodes.).

3.3   Biological function analysis

We reconstructed the metabolic network in the KEGG database using 35 microbial genes from the topological hub. The metabolite pathways designed include Carbohydrate metabolism, Energy metabolism, Lipid metabolism, Amino acid metabolism and Genetic Information Processing. Among them, 14 microbial genes are enriched in the methane metabolism network (Fig. 4), followed by the Genetic Translation Ribosome, which contains 7 microbial genes (Fig. 5). Oxidative phosphorylation contains 4 microbial genes (Fig. 5). The hub members of the five metabolites are involved in amino acid metabolism, such as ornithine, which is the central part of the urea cycle to discharge excess nitrogen.
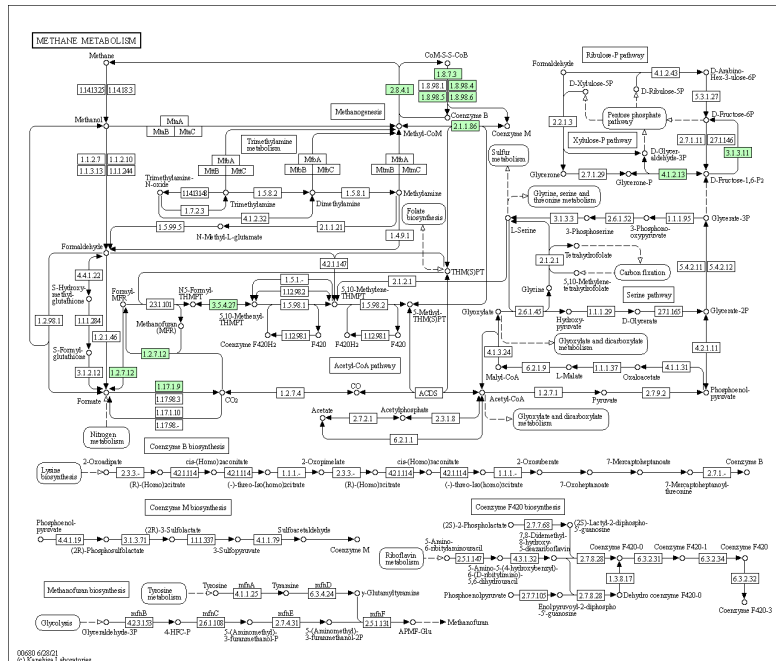
Fig. 4. Reconstructed methane metabolism (The green box represents the location of microbial functional genes).
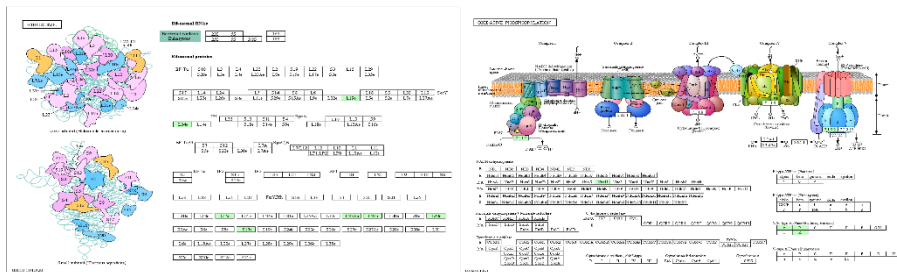

Fig. 5. Reconstructed genetic translation processing and Oxidative phosphorylation (The green box represents the location of microbial functional genes).

4 Discussion

Research on multi-layer networks has begun to develop from simply extending the concepts and methods of single-layer networks to defining corresponding topological properties and dynamic behaviors for multi-layer network structures and practical problems. A lot of work can be done in the research of multi-layer network topology. The potential geometric properties of complex networks have attracted much attention in recent years [14]. Exploring the internal relationship between the geometric properties and

topological structure of multi-layer networks will help promote the progress of related research, such as community structure, link prediction, etc. [14].

In this study, the main analysis is based on a multi-layer heterogeneous network containing rumen metabolites, rumen microbial genes and rumen microbes. Both methane and methane-related microbial genes are in the topological center of the network, indicating that the multi-layer network representation has potential biological consistency. The finding indicated the reliability of the multi-layer network representing the complex system of the rumen microbiome. And other key topological nodes are highly possible to be importantly related to the same biological function.

In the multilayer network, methane connected 1,327 nodes. The denseness of the associated attributes in the first stage makes it very difficult to analyze the network in methods, such as module detection or diffusion analysis. Only considering the analysis of the nodes directly related to methane is already very complicated. There is an urgent need to find a way to collect the most important individuals among all the close contacts. In this study, the topological attribute rankings of nodes are calculated, and subnetworks are generated according to the rankings. Most of these topological indicators have overlapping results. DMNC, MCC and EPC are topological measurements based on direct field and percolation theory, respectively. The results show that there are dense components in the microbial gene layer. But the core members of these three components are different. Betweenness, Stress and Bottleneck are all based on the shortest path to consider the extent to which a node is passed in the network. According to previous research on protein networks. In the protein signal network, high-stress proteins indicate that their functions may be relevant to the protein in linking regulatory molecules. But it may also mean that the protein is involved in a large number of cellular processes. The higher the betweenness centrality indicates that in higher the relevance of the protein as a tissue regulator. High betweenness may be required to keep the signal mechanism functioning. Proteins with higher radiality have a higher chance to become the core of other proteins in the network. Signal networks with higher average radiality are more likely to include functional units and modules. But this needs to be accompanied by evidence of eccentricity and Closeness. The overlap of the hub members of Radiality and Closeness can show that these nodes all have a central regulatory role. The high degree of coincidence between Degree and MNC points out central nodes that are highly connected to other nodes, and dense subgraphs are likely to be generated near them.

Among these topological attributes, we have still extracted valuable hubs. The hub members of the five metabolites are involved in amino acid metabolism, such as ornithine, which is the central part of the urea cycle to discharge excess nitrogen [15]. Acetamide is a substrate that produces Acetate [16]. The associations between microbial communities seem to be more complicated, but the common associations of the same family of microbial genera seem to point to the close connection between methane and this function. *Aeropyrum*, *Desulfurococcus* and *Thermosphaera* are three genera in the *Desulfurococcaceae* family. In the bottleneck hub, *Aeropyrum* and *Desulfurococcus* are negatively related to two nodes that are positively related to methane, and the K09015 gene (SufD) appears in the same hub. Because of the crucial importance of SufD in Fe-S cluster biogenesis [17]. SufBCD stimulates SufS's cysteine desulfurase

activity by working in tandem with SufE [18]. The anaerobic oxidation of methane appears to be linked to the sulfur reduction process [19].

5 Conclusion

The highly central nodes are interconnected, making the entire network super dense. It is very challenging to extract the most critical information from these dense associations. The research developed a multi-layer network hub identification based on topological property rankings to explore individuals associated with methane functions in the complex rumen microbial system. The results found the consistency of the biological functions and topological properties of the multi-layer network. The Bottleneck was identified as one of the most promising topological properties in the study by finding the sulfur-reducing family *Desulfurococcaceae* and related protein across the multi-layer of the network. The results also provided clues to understanding the relationship between the topological structure of the multilayer network and the biological meaning of the complex system. The future direction of research will be dedicated to verification on larger data sets, combined with knowledge-driven networks.

References

1. Huws, S. A. et al. Addressing Global Ruminant Agricultural Challenges Through Understanding the Rumen Microbiome: Past, Present, and Future. Frontiers in Microbiology 9, 2161 (2018).
2. Chaucheyras-Durand, F. & Ossa, F. REVIEW: The rumen microbiome: Composition, abundance, diversity, and new investigative tools. The Professional Animal Scientist 30, 1–12 (2014).
3. Lima, J. et al. Identification of Rumen Microbial Genes Involved in Pathways Linked to Appetite, Growth, and Feed Conversion Efficiency in Cattle. Frontiers in Genetics 10, 701 (2019).
4. Auffret, M. D. et al. Identification of Microbial Genetic Capacities and Potential Mechanisms Within the Rumen Microbiome Explaining Differences in Beef Cattle Feed Efficiency. Frontiers in Microbiology 11, 1229 (2020).
5. Martínez-Álvaro, M. et al. Identification of Complex Rumen Microbiome Interaction Within Diverse Functional Niches as Mechanisms Affecting the Variation of Methane Emissions in Bovine. Frontiers in Microbiology 11, 659 (2020).
6. Wang, M., Wang, H., Zheng, H., Dewhurst, R. J. & Roehe, R. A heat diffusion multilayer network approach for the identification of functional biomarkers in rumen methane emissions. Methods 192, 57–66 (2021).
7. Mengyuan, W., Wang, H., Zheng, H., Dewhurst, R. & Roehe, R. A multilayer co-occurrence network reveals the systemic difference of diet-based rumen microbiome associated with methane yield phenotype. 2860 (2020).
8. 2016 CoNet app inference of biological association networks using Cytoscape.pdf.
9. Chin, C.-H. et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 8, S11 (2014).
10. Chin et al. - 2014 - cytoHubba identifying hub objects and sub-network.pdf.

11. Lee, B., Zhang, S., Poleksic, A. & Xie, L. Heterogeneous Multi-Layered Network Model for Omics Data Integration and Analysis. Frontiers in Genetics 10, 1381 (2020).
12. Kivelä, M. et al. Multilayer networks. Journal of Complex Networks 2, 203–271 (2014).
13. Pilosof, S., Porter, M. A., Pascual, M. & Kéfi, S. The multilayer nature of ecological networks. Nat Ecol Evol 1, 1–9 (2017).
14. Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data | Briefings in Bioinformatics | Oxford Academic.
15. Barmore, W., Azad, F. & Stone, W. L. Physiology, Urea Cycle. in StatPearls (StatPearls Publishing, 2021).
16. KEGG REACTION: R00321. https://www.genome.jp/entry/R00321.
17. Yuda, E. et al. Mapping the key residues of SufB and SufD essential for biosynthesis of iron-sulfur clusters. Sci Rep 7, 9387 (2017).
18. sufD - FeS cluster assembly protein SufD - Bacillus subtilis (strain 168) - sufD gene & protein. https://www.uniprot.org/uniprot/O32165.
19. Timmers, P. H. et al. Anaerobic oxidation of methane associated with sulfate reduction in a natural freshwater gas source. ISME J 10, 1400–1412 (2016).