

A Network Analysis of Methane and Feed Conversion Genes in the Rumen Microbial Community

Fiona Browne^{1*}, Haiying Wang¹, Huiru Zheng¹, Rainer Roehle², Richard J. Dewhurst², Paul Walsh³

¹School of Computing and Mathematics, Computer Science Research Institute, Ulster University, UK

²Future Farming Systems, Scotland's Rural College, Edinburgh, United Kingdom

³NSilico Life Science Ltd., Cork, Ireland

*f.browne@ulster.ac.uk

Abstract— Metagenomics involves the genetic analysis of microbial DNA extracted from communities in an environment sample. Advent and falling costs of next-generation sequencing technologies has accelerated metagenomics research providing an improved understanding of microbial communities. In this study we investigate if the traits methane production and feed conversion rates in the rumen microbial community overlap with top genes ranked by topological metrics in a co-abundance network. A co-abundance network was constructed from abundance values of 1570 microbial genes in rumen samples of 8 cattle identified in a metagenomics study at the Beef and Sheep Research Centre of Scotland's Rural College. We used 4 different topological measures: Degree Centrality, Betweenness Centrality, Bonacich Power Centrality and PageRank to the network. Using permutation testing, we discovered, methane production trait genes significantly overlapped with top ranked genes obtained using the metrics PageRank and Bonacich Power Centrality. Feed conversion trait genes overlapped with top ranked genes using Bonacich Power Centrality and Betweenness. Furthermore, we observed the top ranked genes from PageRank and Bonacich Power Centrality significantly overlapped with genes involved in the KEGG methane metabolism pathway and ranked highly key methanogenesis genes such as *mcrA* and *fmdB*. Identified functional clusters containing most methane and feed conversion genes were also analyzed in terms of overlap with top ranked genes from topological metrics.

Keywords—metagenomics; co-abundance network; topological analysis; rumen microbial analysis

I. INTRODUCTION

It is estimated that the world's population of ruminates is approximately 75.3 million with over 200 species [1]. Ruminates are an important source of high quality food, but also a significant contributor to the emissions of methane gas from anthropogenic activities [2]. Microbes in the rumen are a complex ecosystem consisting of bacteria, protozoa, fungi and archaea (methanogens). These organisms breakdown complex plant polysaccharides, indigestible by humans, into absorbable nutrients. However, this process also produces microbial biomass, and greenhouse gases such as methane [3]. Methane is a major contributor to global warming and 28 times more potent than carbon dioxide [4]. Therefore understanding and

developing approaches to mitigate methane emissions is an important worldwide issue. With advances in rumen microbiology our knowledge of rumen microbial communities is improving [3]. Unraveling variations among rumen microbial communities in terms of methane production and feed conversion rates is important in aiding our understanding of the transformation of plant polysaccharides into desirable nutrients with less environmental impact [3].

Metagenomics is a powerful tool to analyze microbial communities [5]. Metagenomics, as defined by Thomas et al. [6] is based on the genetic analysis of microbial DNA extracted from communities in an environment sample. Initial approaches in the field included cloning and functional expression screening of environmental DNA [7] through to random shotgun sequencing [8][9]. Sequence based metagenomics has been accelerated with the advent of next-generation sequencing (NGS) technologies [10] coupled with the falling costs of these technologies [5]. NGS technologies are advantageous in providing a high volume of data, improved resolution of the community under study in regards to abundance and metabolic activity [11][12].

A number of metagenomic studies have investigated rumen microbial populations [13][14][15]. These include research by Callaway et al. [16] who used bacterial tag-encoded FLX amplicon pyrosequencing to analyse the rumen community of 6 cattle. The cattle were randomly assigned 3 different diets with different portion concentrations of dried distillers grain. It was observed that for samples with diets containing dried distillers grain, ruminal and fecal Firmicute:Bacteroidetes ratios were smaller compared to controls. Similarly, by varying a host's diet, Faith et al. [17] predicted a human gut microbiota's response to diet in gnotobiotic mice. A total of 10 sequenced human gut bacteria were introduced into gnotobiotic mice. The abundance and microbial gene expression were measured across samples consisting of different host diets. It was predicted that 60% of the variation in species abundance was due to the differences in diet.

The recent studies by [18] and [19] applied metagenomics to investigate the rumen microbial community of cattle. The studies aimed to determine if there were links between

microbial genes and the traits methane emissions and feed conversion efficiency. Metagenomic analysis was performed on the rumens of the 8 cattle samples and relative abundance values were measured for 3970 microbial genes of which 1570 had a relative abundance of more than 0.001%. This set of genes was further investigated using partial least squares analysis, which identified 20 and 49 genes associated with the traits methane production and feed conversion ratio respectively. A co-abundance network was constructed using the 1570 identified microbial genes network analysis was performed identifying distinct functional clusters of gene networks. Two clusters of which contained most genes known to be associated with methane metabolism [19].

In this paper, we will extend the network analysis of the co-abundance network performed in [19]. Specifically, we will focus on topological analysis to determine if the traits methane production and feed conversion rates in the rumen microbial community overlap with top genes ranked by topological metrics in a co-abundance network. Analysis of network topologies has been applied in many domains from social networks [20] to drug-target networks [21]. A number of studies [22][23] have shown how the application of network topological metrics can detect key biological processes. For example, the study by Yu et al. [24] showed using protein interaction networks identified that nodes with the highest betweenness values control most of the information flow in the network. Furthermore, topology-based analysis of metabolic networks has been important in studying their impact on modularity [25] and function and regulation [26].

In this paper section II describes the methods, including the dataset, co-abundance network construction and topological metrics. This is followed by results in Section III and conclusion and future works in Section IV.

II. DATA AND METHODS

A. Metagenomic Data

The metagenomic data applied in this research has been obtained from [19] and [27] through experiments undertaken at the Beef and Sheep Research Centre of Scotland's Rural College (SRUC). In these previous studies, Roehe et al. [19] estimated the genetic control of a host on methane emissions caused by changes in microbial community. An overview of the experimental design, sequencing and KEGG analysis is presented below.

1) Experiment Design

A 2×2 factorial design experiment was performed using two breed types (Aberdeen Angus (AA) and Limousin (LIM) crosses) and two diets (defined as concentrate (CON) and forage (FOR)). The steers ($n=72$) were from a two-breed rotational cross between AA and LIM. The progeny groups were from 5 AA and 4 LIM sires. The average number of progenies per sire were 7 for AA and 9 for LIM.

2) Methane Measurements

Methane emissions of individual animals were measured in respiration chambers. Six indirect open-circuit respiration chambers were used (No Pollution Industrial Systems Ltd., Edinburgh, UK)

3) Deep sequencing of DNA

Deep sequencing (Table 1) was carried out in Edinburgh Genomics for 4 pairs of beef cattle selected from 72 animals with extreme high and low methane emissions matched for breed (Aberdeen-Angus or Limousin cross) and diet (CON or FOR). All raw data associated with the study is publically available and can be found in the European Nucleotide Archive under accession PRJEB10338 (<http://www.ebi.ac.uk/ena/data/search?query=PRJEB10338>).

TABLE I. CHARACTERISTICS OF 8 SAMPLES USED IN THE SRUC STUDIES. AA: ABERDEEN ANGUS; LIM: LIMOUSIN CROSS; CON: CONCENTRATE BASED DIET; FOR: FORAGEBASED DIET; DMI: DRY MATTER INTAKE; AND FCR: FEED CONVERSION RATIO

Sample	Breed	Diet	Methane emissions group	Methane (kg/DMI)	FCR (kg intake/kg gain)
1	AA	CON	LOW	7.635	6.102
2	AA	CON	HIGH	18.137	6.096
3	LIM	CON	LOW	9.290	9.327
4	LIM	CON	HIGH	20.130	8.039
5	AA	FOR	LOW	17.412	10.381
6	AA	FOR	HIGH	32.415	6.719
7	LIM	FOR	LOW	19.373	8.065
8	LIM	FOR	HIGH	30.372	8.118

Sequence data were assembled *de novo*. Across all eight samples, there were 1,500,390 predicted proteins, 729,736 of which can be annotated using a Pfam domain. All protein predictions and annotations are available as a Meta4 database (<http://www.ark-genomics.org/tools/meta4>).

4) KEGG analysis

To perform functional analysis, the genomic reads were aligned to the KEGG genes database allowing for up to a 10 % mismatch. The read and best hits belonging to a single KEGG orthologue group (KO) were retained, otherwise the read were disregarded. Read counts for KEGG orthologues were summed and normalised to the total number of hits. A total of 3970 KEGG genes were identified, 1570 of which has the relative abundance of more than 0.001%.

5) Methane emission and Feed Conversion Genes

Using the set of KEGG genes described above, the study in [19] applied partial least squares analysis to identify genes associated with the traits methane production and feed conversion ratio. This resulted in a set of 20 and a set of 49 microbial genes found to be associated with methane emission and feed conversion efficiency in cattle respectively.

B. Network Construction

Using the abundance values of the microbial genes across the 8 samples, a co-abundance network is constructed using Algorithm 1. This network consists of nodes modeling the microbial genes and their interactions modeled by similarity values. The co-abundance similarity network is then analyzed using topological features described below.

$$c_i(\alpha, \beta) = \sum_j (\alpha + \beta c_j) A_{ij} \quad (3)$$

where A_{ij} is the element of row i and column j of the adjacency matrix. Parameter α is used to normalize the measure. Parameter β sets the dependence of each nodes centrality to the adjacent nodes.

1) PageRank

PageRank ranks genes according to importance, i.e. connection to other genes. This metric was recently applied to associate gene connectivity with different subtypes of ovarian cancer tumors [32]. PageRank can be described as:

$$PR_{v_n} = \frac{1-d}{N} + d \sum_{v_j \in M(v_n)} \frac{PR(v_j)}{L(v_j)} \quad (4)$$

where d represents the damping factor, n the index of node of interest, v_n the node, $M(v_i)$ the set of nodes linking to v_n and $L(v_j)$ the out-link counts from node v_j .

Each topological metric is applied to the co-abundance network producing a ranked list, for example, using Degree, we rank the microbial genes from high to low based on Degree value. The top 20% ranked genes from each list are then selected for further analysis. This includes KEGG pathway analysis using top ranked genes and comparison to the methane emission and feed conversion gene sets.

III. RESULTS

A. Network Construction

The microbial gene abundance network was constructed from the set of 1570 KEGG genes identified in rumen contents samples in the metagenomic study by [19] and [27]. Using the gene abundance data across the 8 samples (described in Table I), a co-abundance similarity network was constructed in Cytoscape version 3.2.1 using the plugin ExpressionCorrelation. The nodes in the network represent the KEGG genes. The links represent the similarity between vectors of the abundance levels of genes across all samples. The similarity matrix is computed using Pearson's Correlation Co-efficient with a threshold of 0.90 as these have been applied in the previous study in [19]. This resulted in a network consisting of 1521 nodes and 52,207 edges. The global topology of the co-abundance network is summarized in Table II.

TABLE II. SUMMARY OF CO-ABUNDANCE NETWORK PROPERTIES

Network Features	Values
Nodes	1521
Edges	52,207
Clustering Co-efficient	0.567
Centralization	0.143
Heterogeneity	1.127
Density	0.045
Characteristic Path Length	4.373
Diameter	12
Radius	1

The methane and feed conversion gene sets described in the Methods section were mapped onto the network and colored yellow and red respectively. A detailed network view is presented in Fig. 1 with the methane set in yellow and feed conversion set in red. It can be seen from this figure that the two sets occupy different areas of the global network.

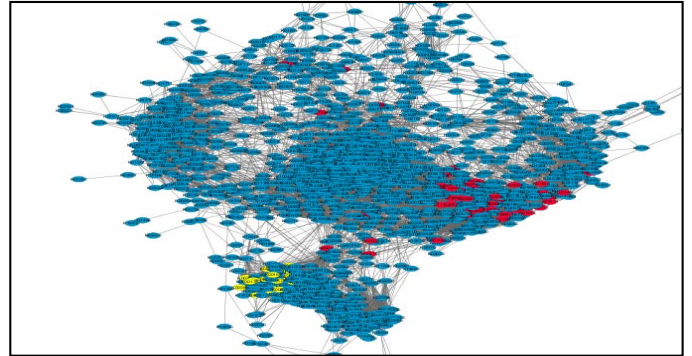


Fig. 1 Overview of co-abundance network with methane represented by yellow nodes and feed conversion genes red

B. Topological Analysis

In this study, we are interested in discerning if top ranked genes using different topological analysis metrics on a co-abundance similarity network overlap with trait methane and feed conversion genes in the rumen microbial community. We hypothesize that application of these metrics are important as previous studies have uncovered key players from biological networks using metrics such as degree (hubs), whereby network hubs are often essential [33]. Four topological metrics: Degree, Betweenness centrality, PageRank and Bonacich Power Centrality as described in the Methods were applied to the co-abundance network. Application of each metric produced a ranked list, the top 20% ranked genes from each metric list was then analyzed. This percentage was selected as it has been previously applied as a cut-off threshold in the study [24].

1) KEGG Pathway Analysis

The top ranked KEGG genes for each topological metric were firstly compared to KEGG pathways detailed in Table III. A total of 8 KEGG pathways are presented representing the top 5 pathways enriched with genes obtained from the various topological analyses. All ranked genes from the topological metrics overlapped with KEGG pathways including: metabolic pathways carbon and methane metabolism. We can see that the metrics Bonacich Power Centrality, PageRank and Betweenness Centrality have high overlap with the 8 pathways. Interestingly, we can see that Bonacich Power Centrality and PageRank have the highest overlap with the methane metabolism pathway (which is of interest as it is related to methane production). Furthermore, these overlaps are statistically significant (Fisher Exact Test, $p < 0.05$).

TABLE III. OVERLAP OF TOP TOPOLOGICAL FEATURES WITH KEGG PATHWAYS

Pathway	Degree	Betweenness	Page Rank	Bonacich
KO00680 Methane metabolism	2	7	15	21
KO01100 Metabolic pathways	82	113	99	106
KO01110 Biosynthesis of secondary metabolites	35	52	43	42
KO01130 Biosynthesis of antibiotics	24	36	26	39
KO02010 ABC transporters	22	13	14	19
KO01120 Microbial metabolism in diverse environments	18	32	24	39
KO01230 Biosynthesis of amino acids	16	21	25	25
KO01200 Carbon metabolism	9	24	18	29

Additional analysis was performed using the 15 and 21 genes from the KEGG Methane Metabolism pathway highly ranked by the PageRank and the Bonacich Power Centrality metrics. Both sets of genes were combined resulting in 27 unique genes. We then calculated the average abundance across the 8 samples. As presented in Fig. 2, higher abundance values are observed for the samples obtained from the higher methane emissions group represented by the blue columns.

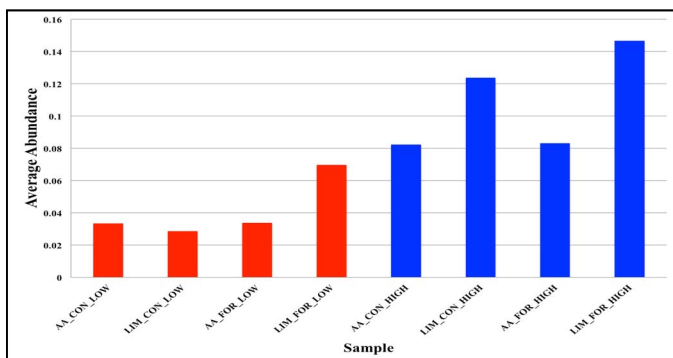


Fig. 2 Average co-abundance values across 8 samples using 27 KEGG methane pathway genes overlapping with top genes identified using topological analysis.

2) Comparison to Methane and Feed Conversion Set

Using the set of methane emission genes and feed conversion efficiency genes identified in study [19] and summarized in the Methods a comparison between the top ranked genes from topological analysis and these sets were performed.

TABLE IV. OVERLAP OF TOP TOPOLOGICAL FEATURES WITH METHANE AND FEED CONVERSION SETS

	Methane	Feed Conversion
Degree	0	3
Betweenness	1	15
Bonacich	8	14
PageRank	6	4

From Table IV, it can be seen that top ranked genes obtained using the topological metrics Bonacich Power Centrality and PageRank overlapped with the methane set (similar to the overlap observed with the KEGG methane metabolism pathway). A heatmap in Fig. 3 illustrates the abundance values of the 8 unique genes identified using these two approaches across the samples. Red represents low abundance through to green representing high abundance. The labels on the X axis present the different samples and they are labeled in terms of breed, diet and methane emissions respectively. The last 4 samples labeled high, represent the samples from the high methane emissions group. From Fig. 3 we can see that these also correspond with high abundance values for the 10 identified genes.

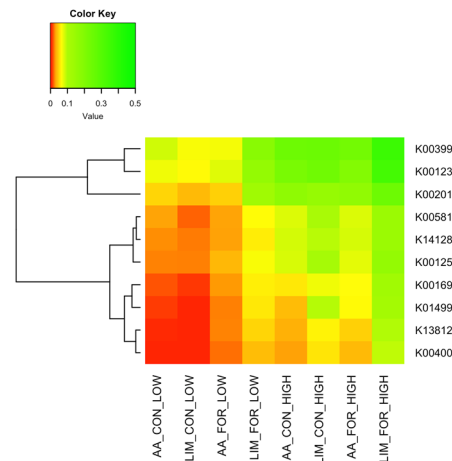


Fig. 3 Heat map illustrating abundance values of 10 microbial genes across 8 samples.

Overlap was also observed between the feed conversion set and top ranked genes from Betweenness Centrality and Bonacich Power Centrality topological metrics. The heatmap in Fig. 4 illustrates the abundance values of the 24 unique genes identified using these two approaches across the samples. Red represents low abundance through to green representing high abundance. It can be seen from the heatmap that the samples LIM_CON_LOW and AA_FOR_LOW obtain high abundance values across the top ranked genes. From Table II, we can see that these 2 samples also have the highest feed conversion ratios.

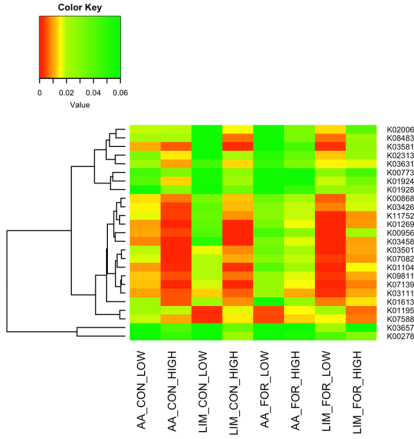


Fig. 4 Heat map illustrating abundance values of 24 microbial genes across 8 samples.

To determine the statistical significance of these results we performed permutation testing where we randomly selected 304 genes from the set of 1570 genes in the rumen dataset and compared these to the methane and feed conversion set respectively. Using equation below we calculated the p value

$$\text{as: } \frac{\sum \text{Rand} > \text{TOP}}{1000} \text{ where } \text{TOP} \text{ represents the number of}$$

methane or feed conversion genes identified by a topological metric (such as Bonacich Power Centrality and PageRank) and $RAND$ the number of methane or feed conversion genes identified by the randomly selected set. Performing permutation for the methane set resulted in $p=0.013$ and $p=0.016$ which indicates the statistical significance of the overlap between the top ranked genes using topological metrics and the methane and feed conversion sets.

C. Cluster Analysis

Further analysis was performed to identify functional clusters within the co-abundance similarity network. In this analysis we compare key clusters identified with the top genes identified using topological metrics of the network. Cluster analysis was performed in Cytoscape using the ClusterOne plugin. A total of 135 clusters were identified with 26 significant clusters ($p<0.05$). We observed for 2 of these clusters, referred to as Cluster A and Cluster B, an overlap with the methane and feed conversion gene set respectively. Cluster A consists of 130 genes and overlaps with all 20 genes in the methane set. Cluster B consists of 120 genes and overlaps with 25 genes of the 49 genes from the feed conversion set. Not all genes were overlapping for feed conversion may be due to the lower power of identification of genes of feed conversion ratio due to selection of animals based on methane emissions and not feed conversion efficiency. Table V presents an overview of the overlap of the clusters with the topological features.

TABLE V. OVERLAP OF CLUSTERS WITH TOP RANKED GENES FROM TOPOLOGICAL ANALYSIS

	Cluster A	Cluster B
Degree	0	19
Betweenness	18	29
Bonacich	38	36
PageRank	45	18

We analyzed the top ranked topological genes that overlapped with genes in Cluster A and Cluster B respectively. In Cluster A (enriched with the methane set), a total of 75 unique topologically ranked genes were observed. We measured the average abundance of these microbial genes across the 8 samples. It can be seen that higher abundance is noted for samples with high methane emissions illustrated in Fig. 5 where the blue columns represent the samples with high methane emissions.

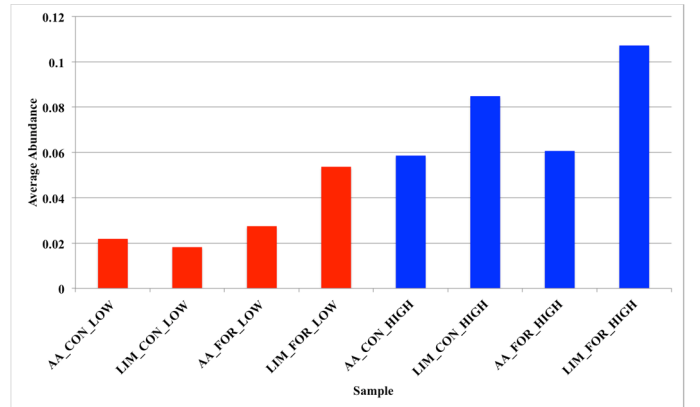


Fig. 5 Average co-abundance values across 8 samples using overlapping genes in Cluster A with top ranked genes from topological metrics.

The same analysis was performed using topologically top ranked genes which overlapped with Cluster B (enriched with the feed conversion set), a total of 64 unique topologically ranked genes was observed. We measured the average abundance of these microbial genes across the 8 samples for the 64 identified genes illustrated in Fig. 6. Here we can see the top 2 highest samples LIM_CON_LOW and AA_FOR_LOW represented by red columns achieved the highest average abundance scores.

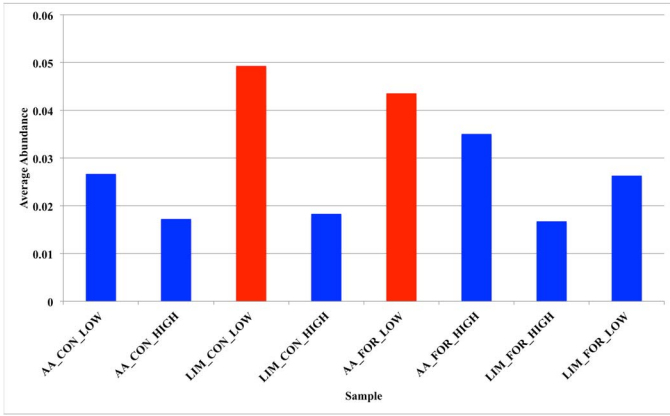


Fig. 6 Average co-abundance values across 8 samples using overlapping genes in Cluster B with top ranked genes from topological metrics.

We extended this analysis by using the identified clusters in the study [19] which contained most genes known to be associated with methane metabolism (Cluster C) and feed conversion efficiency (Cluster D) respectively. Cluster C contains in total 177 genes obtained from Cluster 4 and 6 in [19]. Cluster D contains in total 243 genes obtained from Cluster 2 and 5 in [19]. An overlap with the top ranked genes from the topological metrics is presented in Table VI below.

TABLE VI. OVERLAP OF CLUSTERS FROM [19] WITH TOP RANKED GENES FROM TOPOLOGICAL ANALYSIS

	Cluster C	Cluster D
Degree	0	20
Betweenness	20	54
Bonacich	40	40
PageRank	45	41

Interestingly, we can see agreement in overlap of top ranking genes obtained from topological metrics and the two clustering analysis techniques (Cluster One analysis and MCL clustering in [19]). The average abundance of genes which overlapped with identified top ranked topological metric was obtained for Cluster C. A total of 76 unique microbial genes was observed. Fig. 7 illustrates the average abundance across the different samples for these genes. It can be seen that higher abundance values for the high methane samples in blue was observed, a result which corresponds to Fig. 5.

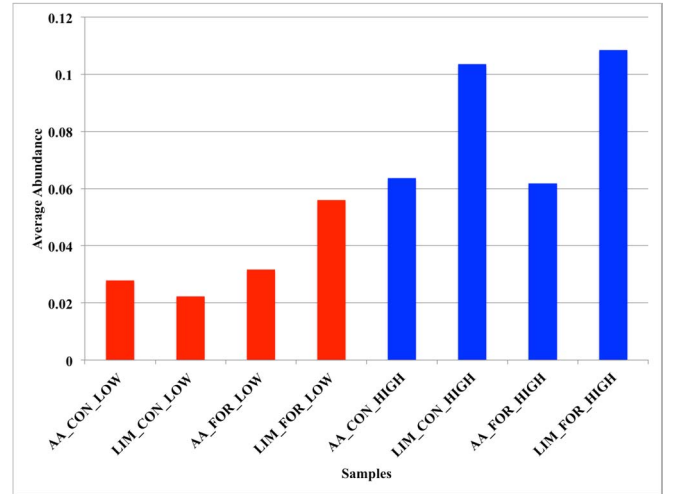


Fig. 7 Average co-abundance values across 8 samples using overlapping genes in Cluster C with top ranked genes from topological metrics.

A total of 110 unique topologically ranked genes were observed in Cluster D. Using these genes, the average abundance of microbial genes across the 8 samples was calculated. Fig. 8. It can be seen that higher abundance is noted for samples with high methane emissions illustrated in Fig. 8 where the red columns represent the samples with highest feed conversion rates. This is similar to the results presented in Fig. 6.

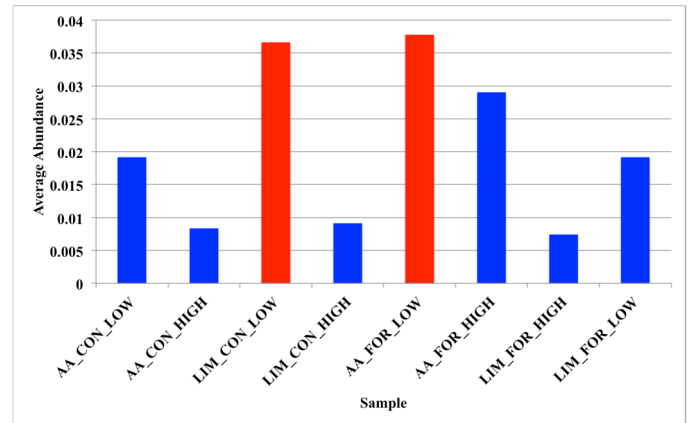


Fig. 8 Average co-abundance values across 8 samples using overlapping genes in Cluster D with top ranked genes from topological metrics.

These results highlight how top genes identified from co-abundance similarity networks using topological measures are consistent with the traits methane emissions and feed conversion.

IV. CONCLUSIONS

Metagenomic approaches are providing the capabilities to investigate the phylogenetic and genomic content of microbial communities in the rumen and beyond [10]. This is important as researchers can identify the functional gene sets responsible for certain traits such as methane emissions and feed conversion rates. In this study we have analyzed the

topological metrics of a co-abundance network constructed using abundance values of 1570 genes from 8 cattle samples. Our objective in this research was to determine if trait genes, namely methane emission and feed conversion rate are observed in top ranked genes using topological metrics. Using 4 different topological measures: Degree, Betweenness Centrality, Bonacich Power Centrality and PageRank, we discovered, depending on trait, some topological metrics were overlapped with more trait genes than others. Specifically, we observed the measures PageRank and Bonacich Power Centrality highly ranked genes involved in the KEGG methane metabolism pathway and also genes from the methane set. The Bonacich Power Centrality extends Degree Centrality through measuring both centrality and power. This combination in relation to the analyzed co-abundance network overlapped with both methane emissions and feed conversion genes. PageRank ranks genes according to importance in a network, i.e. connection to other genes. Therefore using measures of importance and power, key methanogenesis genes were ranked highly. The topological metrics Bonacich Power Centrality and Betweenness Centrality overlapped with more feed conversion genes from the set of 49 genes. Using cluster analysis, we identified 2 distinct clusters containing methane and feed conversion genes from the 2 gene sets respectively. Using the overlapping top ranked genes from the topological metrics we calculated the average co-abundance values across the 8 samples based on the identified genes. Using this approach higher abundance values were observed for the high methane emission samples and higher feed conversion rate samples. This trend was also observed using the identified clusters containing most methane and feed conversion genes from the study in [19].

These results illustrate the potential for the application of topological measures to identify key trait genes in co-abundance networks. This work focused on the high abundance values on genes, in future work we will further analyze down regulated genes with low average abundance which may increase a trait of interest. Furthermore, we will extend this initial analysis to look at other topological features and the role of importance and power in these networks.

ACKNOWLEDGMENT

This research is supported by the MetaPlat project, Horizon 2020 Marie Skłodowska-Curie Research And Innovation Staff Exchange (RISE) <http://www.metaplat.eu/>.

REFERENCES

- [1] T. J. Hackmann, J. N. Spain, J. Z. Adamczewski, et al., "Invited review: Ruminant ecology and evolution: Perspectives useful to ruminant livestock research and production," *J. Dairy Sci.*, vol. 93, no. 4, pp. 1320–1334, Apr. 2010.
- [2] G. Henderson, F. Cox, S. Ganesh, A. Jonker, et al., "Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range," *Sci. Rep.*, vol. 5, p. 14567, Oct. 2015.
- [3] M. B. Lengowski, K. H. R. Zuber, M. Witzig, et al., "Changes in Rumen Microbial Community Composition during Adaption to an In Vitro System and the Impact of Different Forages," *PLoS One*, vol. 11, no. 2, p. e0150115, Feb. 2016.
- [4] E. R. Morgan, E. J. Milner-Gulland, P. R. Torgerson, and G. F. Medley, "Ruminating on complexity: macroparasites of wildlife and livestock," *Trends Ecol. Evol.*, vol. 19, no. 4, pp. 181–188, 2004.
- [5] S. E. Denman, G. Martinez Fernandez, T. Shinkai, M. Mitsumori, and C. S. McSweeney, "Metagenomic analysis of the rumen microbial community following inhibition of methane formation by a halogenated methane analog," *Front. Microbiol.*, vol. 6, p. 1087, Oct. 2015.
- [6] T. Thomas, J. Gilbert, and F. Meyer, "Metagenomics - a guide from sampling to data analysis," *Microb. Inform. Exp.*, vol. 2, no. 1, p. 3, 2012.
- [7] J. Handelsman, M. R. Rondon, S. F. Brady, et al., "Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products," *Chem. Biol.*, vol. 5, no. 10, pp. R245–R249, Oct. 1998.
- [8] G. W. Tyson, J. Chapman, P. Hugenholtz, E. E. Allen, et al., "Community structure and metabolism through reconstruction of microbial genomes from the environment," *Nature*, vol. 428, no. 6978, pp. 37–43, Mar. 2004.
- [9] J. G. Caporaso, C. L. Lauber, W. A. Walters, et al., "Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample," *Proc. Natl. Acad. Sci. U. S. A.*, no. Suppl 1, pp. 4516–22, Mar. 2011.
- [10] S. Shokralla, J. L. Spall, J. F. Gibson, and M. Hajibabaei, "Next-generation sequencing technologies for environmental DNA research," *Mol. Ecol.*, vol. 21, no. 8, pp. 1794–1805, Apr. 2012.
- [11] M. Hess, A. Sczyrba, R. Egan, et al., "Metagenomic discovery of biomass-degrading genes and genomes from cow rumen," *Science*, vol. 331, no. 6016, pp. 463–7, Jan. 2011.
- [12] H. J. Lee, J. Y. Jung, Y. K. Oh, S.-S. Lee, E. L. Madsen, and C. O. Jeon, "Comparative survey of rumen microbial communities and metabolites across one caprine and three bovine groups, using bar-coded pyrosequencing and ¹H nuclear magnetic resonance spectroscopy," *Appl. Environ. Microbiol.*, vol. 78, no. 17, pp. 5983–93, Sep. 2012.
- [13] R. R. Stein, V. Bucci, N. C. Toussaint, et al., "Ecological Modeling from Time-Series Inference: Insight into Dynamics and Stability of Intestinal Microbiota," *PLoS Comput. Biol.*, vol. 9, no. 12, p. e1003388, Dec. 2013.
- [14] C. T. Brown, A. G. Davis-Richardson, A. Giongo, et al., "Gut Microbiome Metagenomics Analysis Suggests a Functional Model for the Development of Autoimmunity for Type 1 Diabetes," *PLoS One*, vol. 6, no. 10, p. e25792, Oct. 2011.
- [15] D. W. Pitta, N. Indugu, S. Kumar, B. et al., "Metagenomic assessment of the functional potential of the rumen microbiome in Holstein dairy cows," *Anaerobe*, vol. 38, pp. 50–60, 2016.
- [16] T. R. Callaway, S. E. Dowd, T. S. Edrington, et al., "Evaluation of bacterial diversity in the rumen and feces of cattle fed different levels of dried distillers grains plus solubles using bacterial tag-encoded FLX amplicon pyrosequencing," *J. Anim. Sci.*, vol. 88, no. 12, pp. 3977–3983, Dec. 2010.
- [17] J. J. Faith, N. P. McNulty, F. E. Rey, et al., "Predicting a human gut microbiota's response to diet in gnotobiotic mice," *Science*, vol. 333, no. 6038, pp. 101–4, Jul. 2011.
- [18] R. J. Wallace, J. A. Rooke, C.-A. Duthie, et al., "Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle," *Sci. Rep.*, vol. 4, p. 5892, Jan. 2014.
- [19] R. Roehle, R. J. Dewhurst, C.-A. Duthie, J. et al., "Bovine Host Genetic Variation Influences Rumen Microbial Methane Production with Best Selection Criterion for Low Methane Emitting and Efficiently Feed Converting Hosts Based on Metagenomic Gene Abundance," *PLoS Genet.*, vol. 12, no. 2, p. e1005846, Feb. 2016.
- [20] M. C. Waumans, T. Nicodème, H. Bersini, et al., "Topology Analysis of Social Networks Extracted from Literature," *PLoS One*, vol. 10, no. 6, p. e0126470, Jun. 2015.
- [21] M. A. Yıldırım, K.-I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, "Drug–target network," *Nat. Biotechnol.*, vol. 25, no. 10, pp. 1119–1126, Oct. 2007.
- [22] D. Berry and S. Widder, "Deciphering microbial interactions and detecting keystone species with co-occurrence networks," *Front. Microbiol.*, vol. 5, p. 219, May 2014.
- [23] S. Greenblum, P. J. Turnbaugh, and E. Borenstein, "Metagenomic

- systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 2, pp. 594–599, Jan. 2012.
- [24] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and M. Gerstein, “The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics,” *PLoS Comput. Biol.*, vol. 3, no. 4, pp. 713–720, 2007.
- [25] K. R. Patil and J. Nielsen, “Uncovering transcriptional regulation of metabolism by using metabolic network topology,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 8, pp. 2685–9, Feb. 2005.
- [26] R. Guimerà and L. A. Nunes Amaral, “Functional cartography of complex metabolic networks,” *Nature*, vol. 433, no. 7028, pp. 895–900, Feb. 2005.
- [27] R. J. Wallace, J. A. Rooke, N. McKain, et al, “The rumen microbial metagenome associated with high methane production in cattle,” *BMC Genomics*, vol. 16, no. 1, p. 839, Dec. 2015.
- [28] D. Koschützki and F. Schreiber, “Centrality analysis methods for biological networks and their application to gene regulatory networks,” *Gene Regul. Syst. Bio.*, vol. 2, pp. 193–201, 2008.
- [29] S. Bergmann, J. Ihmels, and N. Barkai, “Similarities and Differences in Genome-Wide Expression Data of Six Organisms,” *PLoS Biol.*, vol. 2, no. 1, p. e9, Dec. 2003.
- [30] B. H. Junker, D. Koschützki, F. Schreiber, R. Albert, et al., “Exploration of biological network centralities with CentiBiN,” *BMC Bioinformatics*, vol. 7, no. 1, p. 219, 2006.
- [31] D. Ghersi, M. Singh, K. Goh, M. Cusick, D. et al, “Disentangling function from topology to infer the network properties of disease genes,” *BMC Syst. Biol.*, vol. 7, no. 1, p. 5, 2013.
- [32] C. Ma, Y. Chen, D. Wilkins, X. Chen, and J. Zhang, “An unsupervised learning approach to find ovarian cancer genes through integration of biological data,” *BMC Genomics*, vol. 16 Suppl 9, no. Suppl 9, p. S3, 2015.
- [33] H. Yu, D. Greenbaum, H. Xin Lu, X. Zhu, M. et al., “Genomic analysis of essentiality within protein networks,” *Trends Genet.*, vol. 20, no. 6, pp. 227–231, Jun. 2004.