PhD Project 1: Network Inference in microbial communities (metagenomics) to understand microbial activity in drinking water distribution systems

Supervisors:

University of Glasgow:

Dr Umer Zeeshan Ijaz

http://userweb.eng.gla.ac.uk/umer.ijaz

Professor Cindy Smith

https://www.gla.ac.uk/schools/engineering/staff/cindysmith/

Professor William Sloan

https://www.gla.ac.uk/schools/engineering/staff/williamsloan/

University of Sheffield:

Dr Isabel Douterelo Soler

https://www.sheffield.ac.uk/civil/staff/academic/isabel-soler

Description

The aim of the proposed research is to investigate novel network-inference strategies that will provide an integrated view of microbial communities comprising of actors and their mutual interdependences in drinking water distribution systems. The PhD project develops a sequence of increasingly sophisticated methods for fast inference of evolving networks both at spatial and temporal scales. A network is a set of nodes (taxa) with links. These links describe, possible associations and/or causal relationships, and can even be weighted. The work will fuse some of the key concepts from the latest developments in the *Inverse Problems* with those in network science.
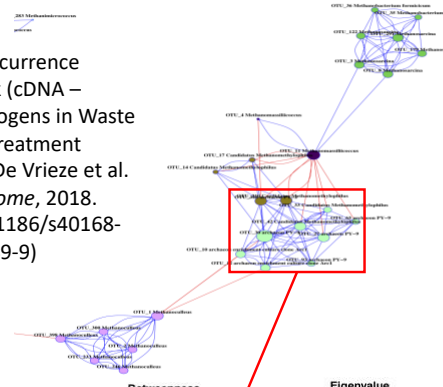

a) GlobalView (http://projectglobalview.blogspot.co.uk/)
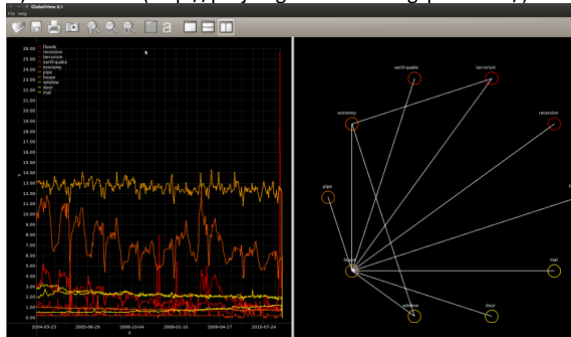
b) microbiomeSeq (http://www.github.com/umerijaz/microbiomeSeq)

c) Co-occurrence network (cDNA – methanogens in Waste Water Treatment Plants; De Vrieze et al. *Microbiome*, 2018. DOI:10.1186/s40168-018-0449-9)

d) Network-wide statistics to identify keystone subcommunities
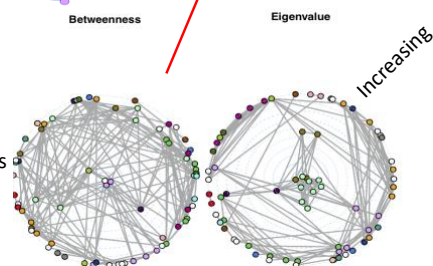
Betweenness   Eigenvalue   Increasing

Figure 1: a) GlobalView software (developed by Dr Ijaz) for static and temporal network inference; b) microbiomeSeq implementing co-occurrence network generation; c) utilising a correlation-based co-occurrence network to identify subcommunities (same color labels) for an existing study; and d) utilising *Betweenness* and *Eigenvalue* centralities to identify keystone subcommunities.

Dr Ijaz has a vast experience in developing network inference strategies, for instance, GlobalView software developed by him at University of Oxford (Figure 1a) implements *Probabilistic Graphical Models* (those that are based on *Markov Random Fields*, and some that are based on *Dynamic Bayesian Networks*). These were initially implemented to work with time series data, however, *Graphical Lasso Method* [1] was successfully applied to 16S rRNA datasets. More recently, microbiomeSeq, an R package (http://userweb.eng.gla.ac.uk/umer.ijaz/projects/microbiomeSeq_Tutorial.html) from Dr Ijaz's lab implements generation of co-occurrence networks through simple correlations as well as *Biweight Midcorrelations* to infer network topologies, on which subcommunity detection algorithms are applied. Furthermore, network wide statistics such as centrality measures as well as recent work by Guimera et al. [2] enable tagging subcommunities whether they serve as peripheral, provincial, or connectors between other subcommunities, as well as identification of subcommunities with disruptive powers to the network (Figure 1d). A recent review paper [3] considers information-theoretic approaches to microbiome networks, marries neutrality with networks, positing that a shift from neutral to niche states transition occurs when species network exhibit scale-free behaviour. The project team have published on neutral community assemblage: fitting Neutral Community model for prokaryotes to the distribution of microbial taxa (Professor Sloan) [4]; and recent work involving Dr Ijaz on fitting the *Unified Neutral Theory of Biodiversity* with Hierarchical Dirichlet Process (NMGS package [5]). These works and thereby lessons learnt, will be expanded to consider neutrality in ecological species networks within the scope of this PhD.

A second aim of the PhD project is to overlay the microbiome networks obtained through species abundances with those obtained from their genomic characteristics (e.g., Antimicrobial Peptides (AMP) signatures). A recent unpublished work by Dr Ijaz ([6] in submission) with colleagues at Cyprus Institute of Neurology & Genetics, utilises AMP similarities to obtain a "Die" score on network topology to predict bacterial compatibility. This work will also be expanded to consider other genomic characteristics. The research will be part of current research taking place at the Civil & Structural Engineering Department at The University of Sheffield by Dr Douterelo. This research will aid Dr Douterelo's to understand microbiome networks ,particularly in biofilms formed in distribution networks, to avoid its adverse consequences on water quality and promote beneficial microbial associations in these ecosystems [7].

The prospective student, ideally someone with a computational background: will become part of Environmental'Omics lab within the Water & Environment group (School of Engineering); will take the existing work forward by consolidating the existing network inference strategies; will be given access to high-performance computing facility maintained at Dr Ijaz's lab; and will be provided numerous datasets from existing and past microbial community studies to test their methods. Further, programming experience in R is required as the secondary aim of the project is to port the developed methods to microbiomeSeq package that both Dr Ijaz and Professor Sloan are contributing to. Dr Douterelo, will provide data from real drinking water distribution systems to optimise network-inference strategies and will assist in understanding the implications and the applicability of the research for water utilities.

References:
[1] U. Z. Ijaz. Graphical Lasso Method for 16SrRNA datasets http://userweb.eng.gla.ac.uk/umer.ijaz/bioinformatics/GraphicalLasso.tar.gz

[2] Guimera, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *nature*, *433*(7028), 895.

[3] Li, J., & Convertino, M. (2019). Optimal Microbiome Networks: Macroecology and Criticality. *Entropy*, *21*(5), 506.

[4] W.T. Sloan et al. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ Microbiol 8: 732–740, 2006.

[5] K. Harris et al. Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process. Proceedings of the IEEE, 105(3):516-529, 2017.

[6] Bacterial Wars – a method for bacterial compatibility based on sequence similarity. *In preparation* (2019).

[7] Douterelo, I, Sharpe, RL, Husband, S, Fish, KE, Boxall, JB. Understanding microbial ecology to improve management of drinking water distribution systems. *WIREs Water*. 2019; 6:e01325. https://doi.org/10.1002/wat2.1325