Project Proposal
BMI826/CS838 Fall 2016
10/6/2016
Erika Da-Inn Lee

Review of Network Modelling Methods for the Human Gut Microbiome & Exploration of Dependency Network Modelling for Probiotic Agent Identification

Introduction

The project goal is twofold: first to review and evaluate existing methodology for modelling the network of microbial interactions within the human gut microbiome; then to implement and experiment with a dependency network learning algorithm, GENIE3, as a framework to generate hypotheses on potential probiotic agents against pathogenic microbes (e.g. *Clostridium difficile*).

Significance

The bacterial, fungal, and viral microbes harbored within the human gastrointestinal tract are referred to as the gut microbiome. Extensive research has established its involvement in metabolism, nutrition, physiology, and immune function, with perturbations or imbalance in the 'normal' microbiota linked to inflammatory bowel disease (IBD), irritable bowel syndrome (IBS), obesity, type 2 diabetes, and atopy (Bull & Plummer, 2014). One of the key and endemic perturbations in modern healthcare landscape has been overuse of antibiotics, effectively selecting for multidrug resistant strains of microbes (Ventola, 2015), and accelerating the need and search for probiotic alternatives to infection prevention and treatment (D'Souza, Rajkumar, Cooke, & Bulpitt, 2002). In the context of plant and soil microbiome, network models have been identified as tools to generate testable hypotheses about candidate microbes affecting plant health, with "pathogen-focused analysis" of networks aiming to find taxa with direct or indirect associations with taxa known a priori as pathogens (Poudel et al. 2016). This project aims to review work done so far on network modelling of the human gut microbiome, evaluating their efficacy in hypothesis generation and predictive power. An dependency network learning algorithm, GENIE3, which has been particularly effective in learning gene regulatory networks (Huynh-Thu, Irrthum, Wehenkel, & Geurts, 2010), will be appropriated for the purpose of finding potential probiotic agents against pathogenic microbes.

Approach

For the literature review portion, recent publications employing correlation and Bayesian network modelling of microbiome will be reviewed (see preliminary list of publications for review below). The preliminary list of publications to be reviewed are not limited to learning a network from 16S rRNA sequenced data of the human gut microbiome, but also that of soil, human lung, milk, and plant, so that any viable method of network construction in a microbial community can be evaluated. There not having been identified a common metric to measure the accuracy of learned networks in the preliminary review of the selected publications, the review will also look at different validation methods for the learned network, along with how each paper measures the predictive power of the learned model; if any a priori knowledge was integrated, and if so, how; and finally, how time-series or context-specific interactions are modelled from the sample data.

The experimentation/implementation portion is contingent upon identifying a training and test data set (most likely 16S rRNA sequenced data) that can be used for implementing a dependency network learning algorithm, GENIE3, against microbiome data. The goal is to identify microbial agents with high interaction with one another, which could be used to identify potential probiotic agents against pathogenic microbial agents (e.g. *C. difficile*). GENIE3 is a particularly effective methodology for constructing gene regulatory networks from gene expression data. The plan is to adjust the algorithm to use Operational Taxonomic Units (OTUs) or some equivalent representation of microbial agents (most likely limited to bacterial agents) as its 'features' in the fixed-length feature vector, instead of individual genes/regulatory factors. If reality allows for ambition, exploration of ensemble methods beyond random forests or extra trees (Ruyssinck et al., 2014) will be included in the implementation.

References

Bull, M. J., & Plummer, N. T. (2014). Part 1: The Human Gut Microbiome in Health and Disease. *Integrative Medicine: A Clinician's Journal*, *13*(6), 17–22.

D'Souza, A. L., Rajkumar, C., Cooke, J., & Bulpitt, C. J. (2002). Probiotics in prevention of antibiotic associated diarrhoea: meta-analysis. *BMJ : British Medical Journal*, *324*(7350), 1361.

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS One*, *5*(9), e12776. doi:10.1371/journal.pone.0012776

Poudel, R., Jumpponen, A., Schlatter, D. C., Paulitz, T. C., Gardener, B. B., Kinkel, L. L., & Garrett, K. A. (2016). Microbiome Networks: A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management. *Phytopathology*, *106*(10), 1083-1096.

Ruyssinck, J., Huynh-Thu, V. A., Geurts, P., Dhaene, T., Demeester, P., & Saeys, Y. NIMEFI: Gene Regulatory Network Inference using Multiple Ensemble Feature Importance algorithms *PLoS ONE*, *9*(3), e92709. doi:10.1371/journal.pone.0092709

Ventola, C. L. (2015). The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics*, *40*(4), 277–283.

Preliminary List of Publications to Review

Barberán, A., Bates, S. T., Casamayor, E. O., & Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *The ISME Journal, 6*(2), 343–351. http://doi.org/10.1038/ismej.2011.119

Chen, E. Z., & Li, H. (2016) A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics, 32*(17), 2611-2617.

Kelder, T. Stroeve, J. H. M., Bijlsma, S., Radonjic, M., & Roeselers, G. (2014). Correlation network analysis reveals relationships between diet-induced changes in human gut microbiota and metabolic health. *Nutrition & Diabetes, 4,* e122. doi:10.1038/nutd.2014.18

Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., Brookes, P. C., Xu, J., & Gilbert, J. A. (2016). Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *The ISME Journal*, *10*, 1891–1901; doi:10.1038/ismej.2015.261

Ma, Z., Guan, Q., Ye, C., Zhang, C., Foster, J. A., & Forney, L. J. (2014). Network analysis suggests a potentially 'evil' alliance of opportunistic pathogens inhibited by a cooperative network in human milk bacterial communities. *Scientific Reports, 5,* 8275. doi:10.1038/srep08275

McGeachie, M. J., Sordillo, J. E., Gibson, T., Weinstock, G. M., Liu, Y.-Y., Gold, D. R., … Litonjua, A. (2016). Longitudinal Prediction of the Infant Gut Microbiome with Dynamic Bayesian Networks. *Scientific Reports*, *6*, 20359. http://doi.org/10.1038/srep20359

McHardy, I. H., Goudarzi, M., Tong, M., Ruegger, P. M., Schwager, E., Weger, J. R., … Braun, J. (2013). Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*, *1*, 17. http://doi.org/10.1186/2049-2618-1-17

Morris, A., Paulson, J. N., Talukder, H., Tipton, L., Kling, H., Cui, L., … Ghedin, E. (2016). Longitudinal analysis of the lung microbiota of cynomolgous macaques during long-term SHIV infection. *Microbiome*, *4*, 38. http://doi.org/10.1186/s40168-016-0183-0

Naqvi, A., Rangwala, H., Keshavarzian, A., & Gillevet, P. (2010). Network-based Modeling of the Human Gut Microbiome. *Chemistry & Biodiversity*, *7*(5), 1040–1050. http://doi.org/10.1002/cbdv.200900324

Steinway, S. N., Biggs, M. B., Loughran, T. P., Papin, J. A., & Albert, R. (2015). Inference of Network Dynamics and Metabolic Interactions in the Gut Microbiome. *PLoS Computational Biology*, *11*(6), e1004338. http://doi.org/10.1371/journal.pcbi.1004338

Tsai, K.-N., Lin, S.-H., Liu, W.-C., & Wang, D. (2015). Inferring microbial interaction network from microbiome data using RMN algorithm. *BMC Systems Biology*, *9*, 54. http://doi.org/10.1186/s12918-015-0199-2

Van der Heijden, M. G. A., & Hartmann, M. (2016). Networking in the Plant Microbiome. *PLoS Biology*, *14*(2), e1002378. http://doi.org/10.1371/journal.pbio.1002378