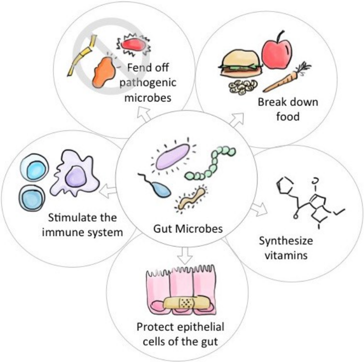
A Novel Graph-Theoretical Approach for Identifying Inter-correlations and Functional Pathways in Microbiome Data

Suyeon Kim

College of Information Science and Technology
Biomedical Informatics

March 1st, 2019

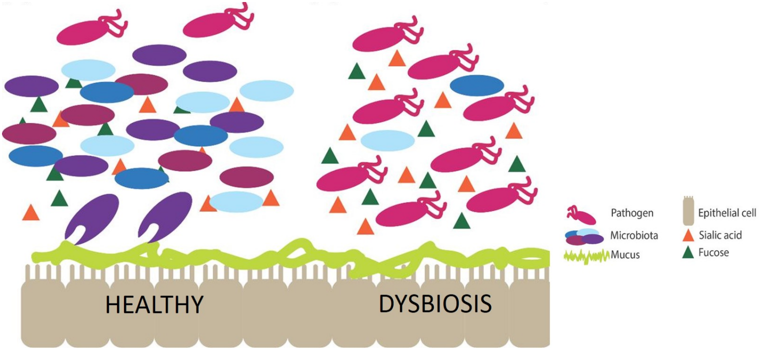
What do our gut microbes do for us?



- The human microbiome represents a vastly complex ecosystem that is tightly linked to our development, physiology, and health.

Figure by Shannon MacArdel

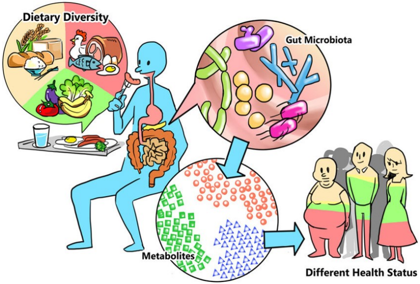
Microbiome, key controller of health



- Several studies highlighted the role of microbiome in human health and disease.

McKenney et al, Oxford, 2016

Gut microbiome and Disease



Examples of disease associated with altered gut microbiota diversity:

- Inflammatory bowel disease (IBD)
- Parkinson's disease
- Depression and Anxiety
- Diabetes and obesity
- Colorectal cancer
- Autism
- Irritable bowel syndrome
- *C.difficile* infection

Lyu, Q., Hsu, C. C. (2018). Can Diet Influence Our Health by Altering Intestinal Microbiota-Derived Fecal Metabolites?. MSystems, 3(2), e00187-17.

Motivation

- Interactions between species shape its host microbial community.
- These interactions are important to understand how microbiome related diseases.
- Metagenomics-based studies of microbiome
  - Characterize the composition of species in a microbiome
  - Complex co-occurrence patterns
- Host and microbiota do not operate alone
  - The interactions between the host and bacterial systems need to be considered through a more holistic, systematic approach

Literature Review

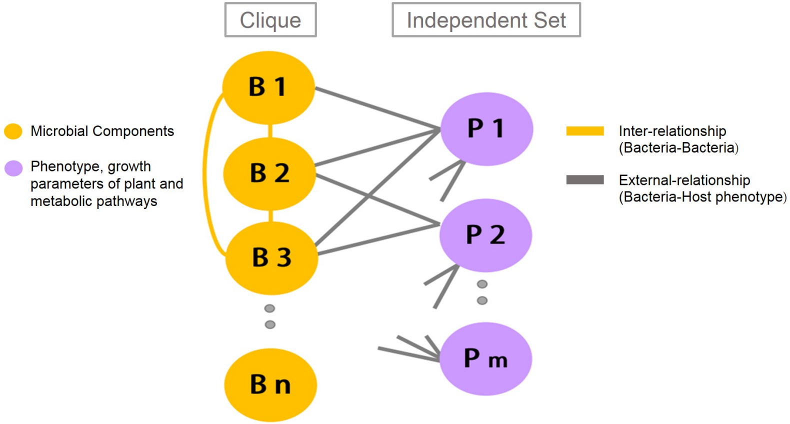
- Multi high-throughput methods for studying the human microbiome are available. [Buschart, A et al., 2017, Borenstein, E et al., 2013]
- With the advanced high throughput technqiues, it calls for the development of systems-level methods in microbiome study. [Borenstein, E et al., 2013]
- Difficult to identify the actual contribution of microbial traits to human physiology with metagenomics study alone. [Greenblu, S et.al 2013, Buschart, A et al., 2017]
- Preliminary systems-level modeling show promise but many challenges remain. [Greenblum, S et al., 2017]

Goal of the project

Developing a comprehensive framework that integrates microbe-microbe and microbe-host interactions:

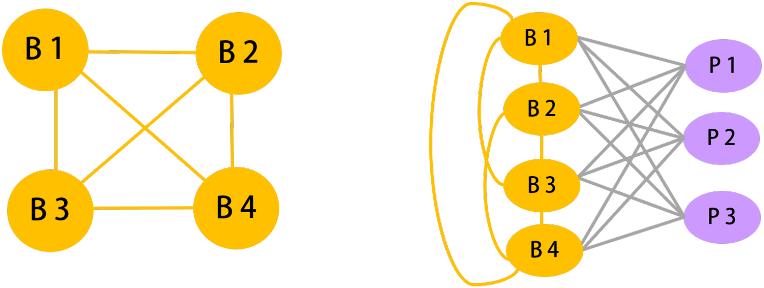
- To create a graph theoretic pipeline for identifying bacterial inter-relations and functional pathways in microbial ecosystem.

Proposed Solution: Split Graph Model



Advantages of Split Graph

- Two different relationships in split graph



● Bacteria

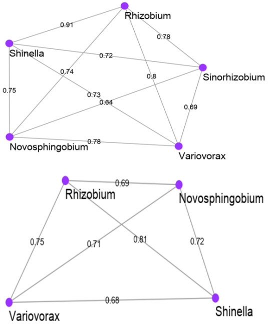
▬ Inter-relationship (Bacteria-Bacteria)

● Phenotype

▬ External-relationship (Bacteria-Phenotype)

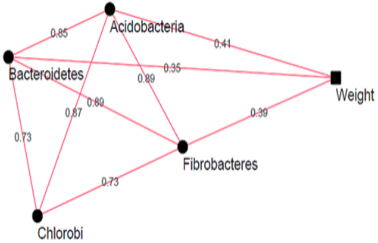
Split graphs in multiple ecosystems

- Soil



Healthy (upper) vs. Unhealthy(bottom)

- Fish

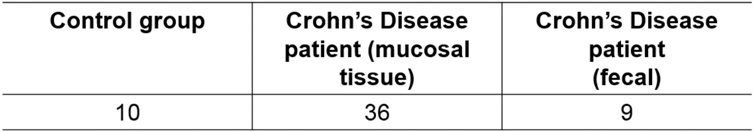


FCB group with *Acidobacteria* in fish gut

Kim, S et al., IEEE BIBM 2017

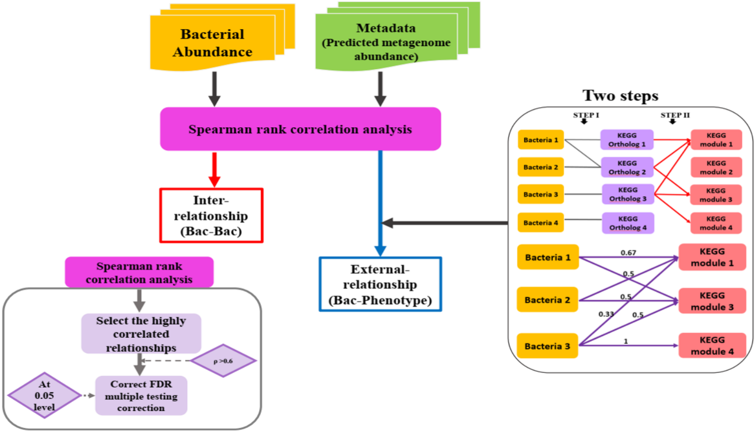
Microbiome in Human gut

- Human gut microbiome data
- Bacterial relative abundance (family level)
    - All samples from a Korean population



Control group	Crohn's Disease patient (mucosal tissue)	Crohn's Disease patient (fecal)
10	36	9

Split Graph Analysis



Bacterial Abundance

Metadata (Predicted metagenome abundance)

Spearman rank correlation analysis

Inter-relationship (Bac-Bac)

External-relationship (Bac-Phenotype)

Spearman rank correlation analysis

Select the highly correlated relationships

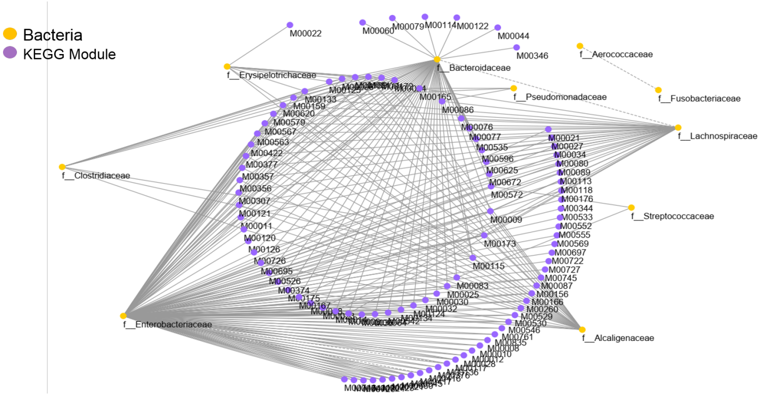
$\rho > 0.6$

Correct FDR multiple testing correction

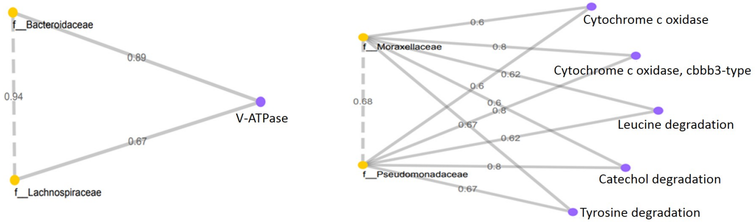
At 0.05 level

Two steps

Integrated Correlation Network (CDS)



Split graph in Crohn's disease (stool and tissue)



- All Bacteria are reported to increased risk of IBD [Hasler et al., 2016]
- V-ATPase is involved in a number of human diseases [Hinton et al., 2009]
- Modules/Pathways in CDT are closely associated with energy metabolism, amino acid degradation, and energy deficiency (low ATP levels).

Results

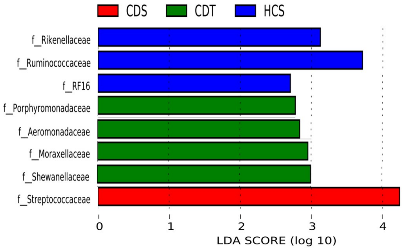
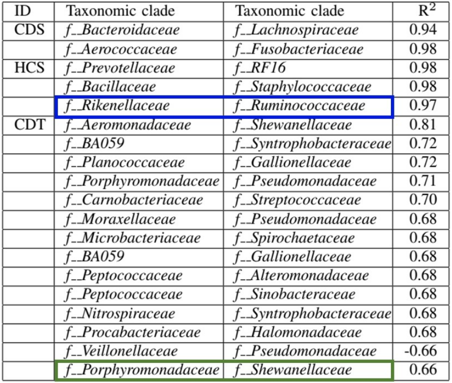


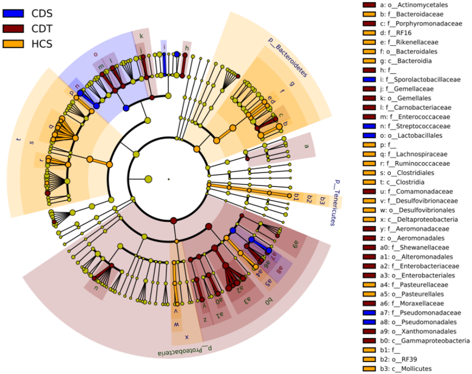
Fig: Differentially abundant bacterial taxa

TABLE I: Inter-bacteria correlations in all sample groups



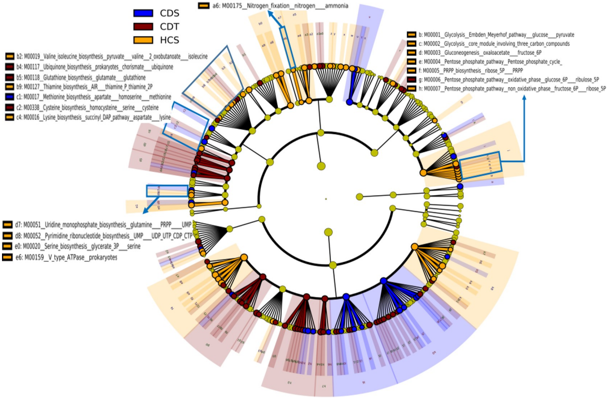
ID	Taxonomic clade	Taxonomic clade	R^2
CDS	*f_Bacteroidaceae*	*f_Lachnospiraceae*	0.94
	f_Aerococcaceae	*f_Fusobacteriaceae*	0.98
HCS	*f_Prevotellaceae*	*f_RF16*	0.98
	f_Bacillaceae	*f_Staphylococcaceae*	0.98
	f_Rikenellaceae	*f_Ruminococcaceae*	0.97
CDT	*f_Aeromonadaceae*	*f_Shewanellaceae*	0.81
	f_BA059	*f_Syntrophobacteraceae*	0.72
	f_Planococcaceae	*f_Gallionellaceae*	0.72
	f_Porphyromonadaceae	*f_Pseudomonadaceae*	0.71
	f_Carnobacteriaceae	*f_Streptococcaceae*	0.70
	f_Moraxellaceae	*f_Pseudomonadaceae*	0.68
	f_Microbacteriaceae	*f_Spirochaetaceae*	0.68
	f_BA059	*f_Gallionellaceae*	0.68
	f_Peptococcaceae	*f_Alteromonadaceae*	0.68
	f_Peptococcaceae	*f_Sinobacteraceae*	0.68
	f_Nitrospiraceae	*f_Syntrophobacteraceae*	0.68
	f_Procabacteriaceae	*f_Halomonadaceae*	0.68
	f_Veillonellaceae	*f_Pseudomonadaceae*	-0.66
	f_Porphyromonadaceae	*f_Shewanellaceae*	0.66

Identify candidate taxonomic biomarkers between Control and CD groups



- Some of the bacterial taxa identified by the split graph are candidate biomarkers.

Identify candidate functional biomarkers between Control and CD group



- Some of the functional features(phenotype) identified by the split graph are also candidate biomarkers.

Key findings

- The split graph model easily captured the structure and function of microbial communities in Crohn's disease patients and healthy individuals.
- The taxonomic and metabolic biomarkers in different conditions were also highlighted in split graph analysis.

Conclusion

- Provided new insights into identifying two distinct relationships with split graph analysis.
- Microbial elements within the relationship in CDT are closely associated with the metabolic mechanisms that lead to the Crohn's disease.
- Identified the distinct structure or function of microbial community in different health conditions.
- Pipeline was extended to allow additional microbial study by overlaying of multi-omics data.