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.
Several studies highlighted the role of microbiome in human health 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

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 techniques, 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. [Greenblum, 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

- Microbial Components
- Phenotype, growth parameters of plant and metabolic pathways

Clique

Independent Set

B1 -> P1
B2 -> P1, P2
B3, Bn -> P1, P2, Pm

Inter-relationship (Bacteria-Bacteria)
External-relationship (Bacteria-Host phenotype)

Suyeon Kim
Student Research and Creativity Activity Fair 2019
Advantages of Split Graph

- Two different relationships in split graph

Bacteria

Inter-relationship (Bacteria-Bacteria)

B 1

B 2

B 3

B 4

Phenotype

External-relationship (Bacteria-Phenotype)

B 1

B 2

B 3

B 4

P 1

P 2

P 3

Suyeon Kim

Student Research and Creativity Activity Fair 2019
Split graphs in multiple ecosystems

- **Soil**
  - Healthy (upper) vs. Unhealthy (bottom)
  - Split graph with various bacterial species like Rhizobium, Shinella, Sinorhizobium, Novosphingobium, Variovorax, and Shinella.

- **Fish**
  - Split graph with various bacterial species like Acidobacteria, Bacteroidetes, Chlorobi, Fibrobacteres, and Weight.

  **FCB group with Acidobacteria in fish gut**

---

Kim, S et al., IEEE BIBM 2017

Suyeon Kim

Student Research and Creativity Activity Fair 2019
Microbiome in Human gut

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

<table>
<thead>
<tr>
<th>Control group</th>
<th>Crohn’s Disease patient (mucosal tissue)</th>
<th>Crohn’s Disease patient (fecal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>36</td>
<td>9</td>
</tr>
</tbody>
</table>
Split Graph Analysis

**Bacterial Abundance**

**Metadata**
(Predicted metagenome abundance)

**Spearman rank correlation analysis**

**Inter-relationship**
(Bac-Bac)

**External-relationship**
(Bac-Phenotype)

**Two steps**

**STEP I**
- Bacteria 1
- Bacteria 2
- Bacteria 3
- Bacteria 4

**KEGG module 1**

**STEP II**
- KEGG Ortholog 1
- KEGG Ortholog 2
- KEGG Ortholog 3
- KEGG Ortholog 4

**KEGG module 1**

- Bacteria 1, 0.67
- Bacteria 2, 0.5
- Bacteria 3, 0.33
- Bacteria 4, 1

**KEGG module 2**

- Bacteria 1, 0.5
- Bacteria 2, 0.5
- Bacteria 3, 0.5

**KEGG module 3**

- Bacteria 1, 0.5
- Bacteria 2, 0.5
- Bacteria 3, 0.5

**KEGG module 4**

- Bacteria 1, 0.5
- Bacteria 2, 0.5
- Bacteria 3, 0.5

**Suyeon Kim**

Student Research and Creativity Activity Fair 2019
Integrated Correlation Network (CDS)
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>
<thead>
<tr>
<th>ID</th>
<th>Taxonomic clade</th>
<th>Taxonomic clade</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDS</td>
<td>f._Bacteroidaceae</td>
<td>f._Lachnospiraceae</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>f._Aerococcaceae</td>
<td>f._Fusobacteriaceae</td>
<td>0.98</td>
</tr>
<tr>
<td>HCS</td>
<td>f._Prevotellaceae</td>
<td>f._RF16</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>f._Bacillaceae</td>
<td>f._Staphylococcaceae</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td><strong>f._Rikenellaceae</strong></td>
<td><strong>f._Ruminococcaceae</strong></td>
<td><strong>0.97</strong></td>
</tr>
<tr>
<td>CDT</td>
<td>f._Aeromonadaceae</td>
<td>f._Shewanellaceae</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>f._BA059</td>
<td>f._Syntrophobacteriaceae</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>f._Planococcaceae</td>
<td>f._Gallionellaceae</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>f._Porphyromonadaceae</td>
<td>f._Pseudomonadaceae</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>f._Carnobacteriaceae</td>
<td>f._Streptococcaceae</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>f._Moraxellaceae</td>
<td>f._Pseudomonadaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._Microbacteriaceae</td>
<td>f._Spirochaetaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._BA059</td>
<td>f._Gallionellaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._Peptococcaceae</td>
<td>f._Aleronadaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._Peptococcaceae</td>
<td>f._Sinobacteriaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._Nitrospiraceae</td>
<td>f._Syntrophobacteriaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._Procbacteriaceae</td>
<td>f._Halomonadaceae</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>f._Veillonellaceae</td>
<td>f._Pseudomonadaceae</td>
<td>-0.66</td>
</tr>
<tr>
<td></td>
<td><strong>f._Porphyromonadaceae</strong></td>
<td><strong>f._Shewanellaceae</strong></td>
<td><strong>0.66</strong></td>
</tr>
</tbody>
</table>
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.