Using Pairwise Difference Features to Measure Temporal Changes in the Microbial Ecology

M. Yazdani[1], L. Smarr[2]

[1]*California Institute for Telecommunications and Information Technology, University of California San Diego USA*
[2]*Harry E. Gruber Professor, Department of Computer Science and Engineering, University of California San Diego USA*

The exponential affordability of DNA sequencing technologies has enabled not only the accessible and rapid sequencing of the human genome, but the opportunity to sequence vast numbers of multi-cellular organism and the tiniest microbes. Studying the microbes in the human body has led us to the view that the human body is not a single organism but rather a symbiotic ecology of microbes and human cells. The microbiome is referred to the collection of microorganisms in our body (consisting mostly of bacteria, archaea, and some eukarya) and outnumber our human cells ten to one. Numerous studies suggest that the microbiome may play a critical role in a number of autoimmune diseases, with most research on inflammatory bowel disease. The microbiome ecology can be thought of as a composition of the relative abundance of a large number of organisms that reside in our bodies. Numerous studies suggest that disruptions in the relative abundance of a healthy microbial ecology leads to a disease state. The challenge remains, however, in understanding how and what changes in the composition of the ecology leads to a disease state.

The relative abundance of a microbial ecology can be thought of as residing in a high dimensional vector space (specifically, a simplex). Typically, a dimensionality reduction technique (such as Principle Component Analysis or similar method) is used to project the high dimensional abundance distributions of organisms to a two or three dimensions to compare the microbial ecology of healthy patients against non-healthy. While such techniques can successfully illustrate that ecologies of healthy and sick patients are "different," a drawback of this approach is losing interpretability and assessing what (and how) microbes are responsible for discrimination between disease and healthy states. Interpreting dimensionality reduction techniques in the context of temporal changes also becomes especially challenging.

In this work we propose interpretable features, what we refer to as "pairwise differences", for understanding temporal changes of the microbial ecology. This method is inspired by the "distance" correlation originally proposed by Szelky [1]. The data that we use in this work are 40 samples over the course of 3 years from the gut of one patient (patient "LS"). The samples were processed by uBiome who derived the relative microbial taxonomic abundances for each sample. We processed these relative abundances using our proposed method of pairwise differences to identify changes in the microbiome ecology. Our method of pairwise differences enables us to also easily identify outliers and other interesting changes. Using our pairwise difference method we can also derive features for other data analysis techniques. Using such features, we then applied Topological Data Analysis (TDA) [2] to identify the specific microbes that are responsible for the temporal changes in the microbial ecology.

978-1-4799-9926-2/15/$31.00 ©2015 IEEE

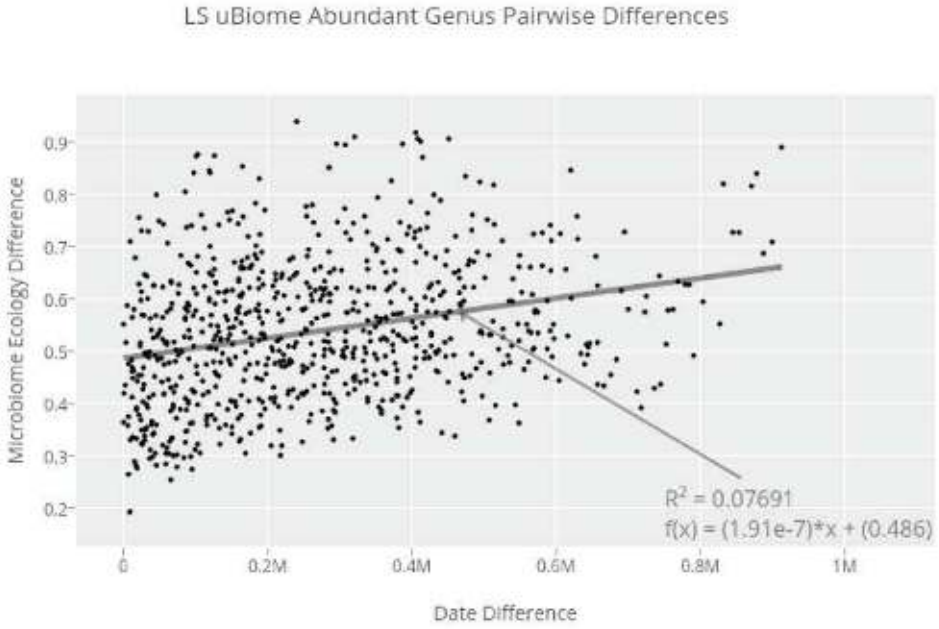


Fig.1 – Pairwise differences of microbial ecological changes versus changes in time. Small time changes that yield large ecological change are of interest.

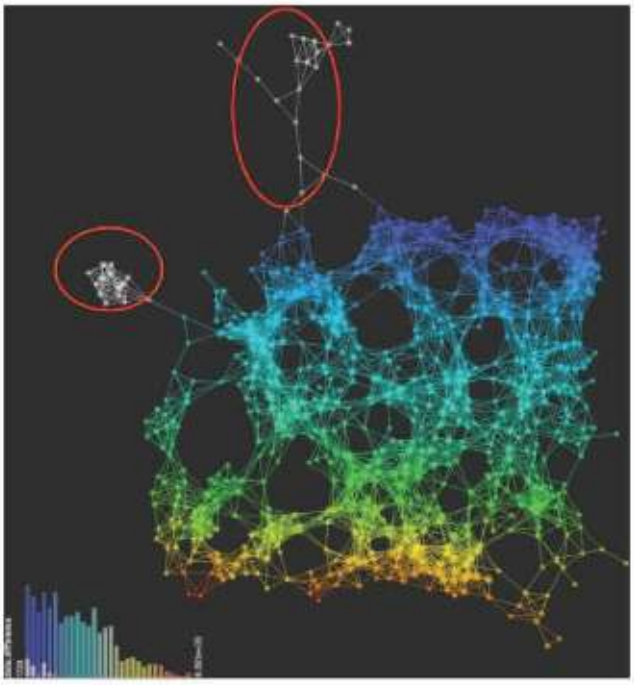


Fig.2 – Topological analysis of microbial ecological change colored by the changes of time. Red ovals indicate changes of time where the differences in time is small, but the changes in the ecology is large.

References

[1] Székely, Gábor J., Maria L. Rizzo, and Nail K. Bakirov *The Annals of Statistics* **35.6** (2007): 2769-2794.

[2] Epstein, Charles, Gunnar Carlsson, and Herbert Edelsbrunner. *Inverse Problems* **27.12** (2011): 120201.