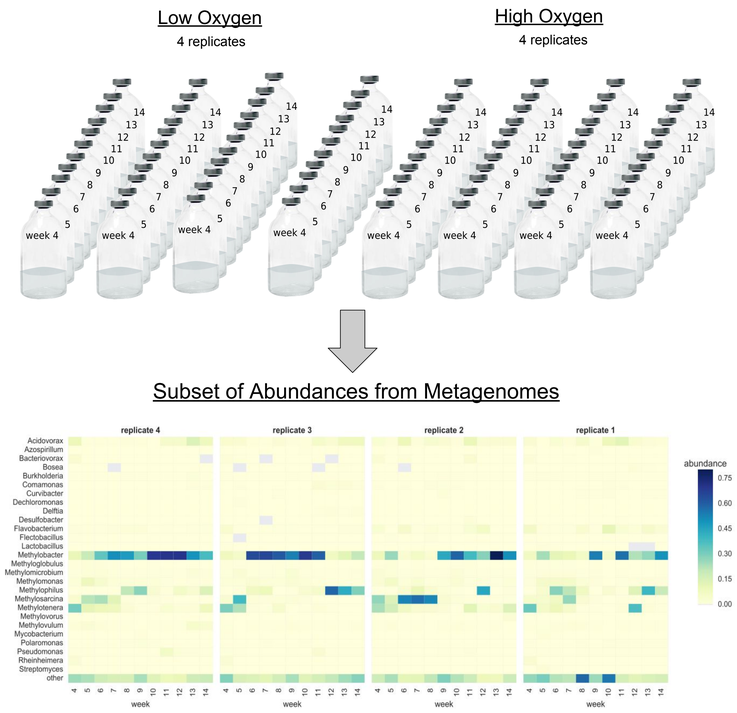
Time Series Analysis of Bacterial Community Composition

Software Engineering for Data Scientists (CSE 599 B1) Project
Saghar Hosseini , Yuan Gao, Janet Matsen , Pearl Philip

Background

How do bacteria cooperate to remediate methane in Nature? We can discover their interactions by observing population dynamics.

Lake sediment was collected from a 180-foot deep study site in Lake Washington using this sediment corer. The samples were propagated with methane, and either high or low oxygen for 14 weeks. Metagenomes were sampled for weeks 4-14. Relative abundances of bacteria were inferred from this 5 terabyte sequencing data set.



Methods

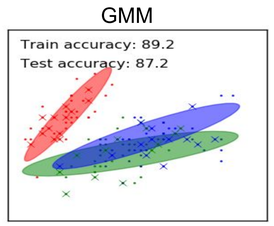
Dynamic Mode Decomposition (DMD)

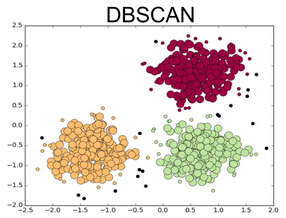
- Powerful tool for analyzing the dynamics of nonlinear systems
- Given a sequential time series of data, DMD computes a linear mapping A, relating the snapshots:

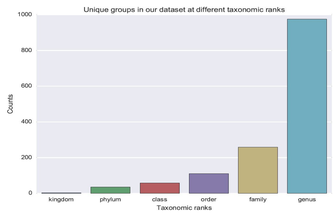
$$\begin{bmatrix} \vdots & \vdots & & \vdots \\ x_1 & x_2 & \dots & x_{t+1} \\ \vdots & \vdots & & \vdots \end{bmatrix} = A \begin{bmatrix} \vdots & \vdots & & \vdots \\ x_0 & x_1 & \dots & x_t \\ \vdots & \vdots & & \vdots \end{bmatrix}$$

Clustering algorithms

1. Gaussian Mixture Models: Generates data points from weighted sum of Gaussian component densities.
2. DBSCAN: Clusters points that are closely packed, marking as outliers points that in low-density regions.
3. Clustering through logical taxonomic reduction: Summing abundances across the sample space below a cutoff taxonomic rank to reduce data size.

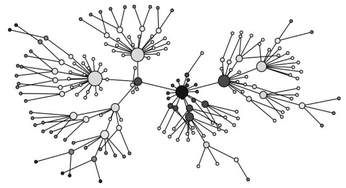






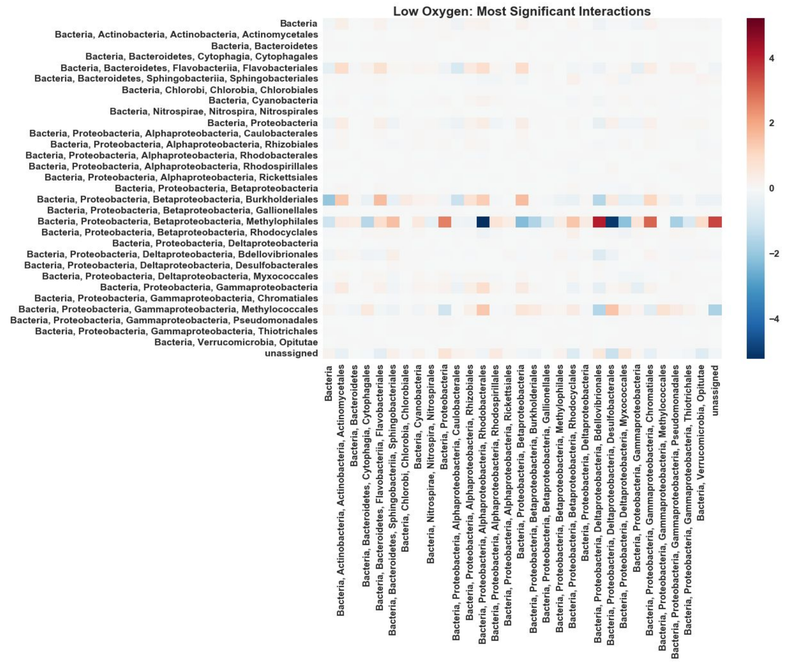
Visualization tools

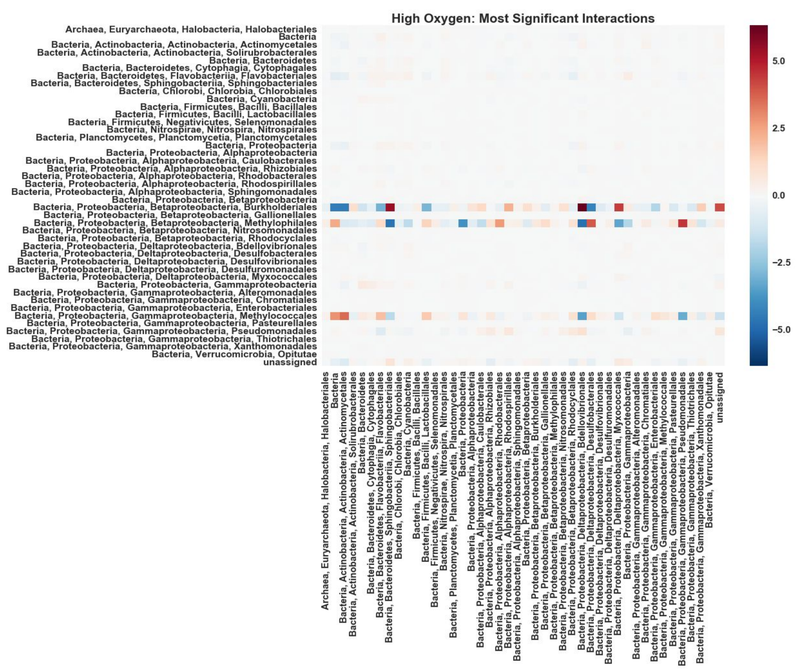
1. NetworkX:Creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



Results

Bacterial Correlations from Interaction matrices





Learn more at github.com/JanetMatsen/bacteriopop

Seattle, WA, March 2016