

THE MICROBIAL INTERACTION DATABASE

Dileep Kishore

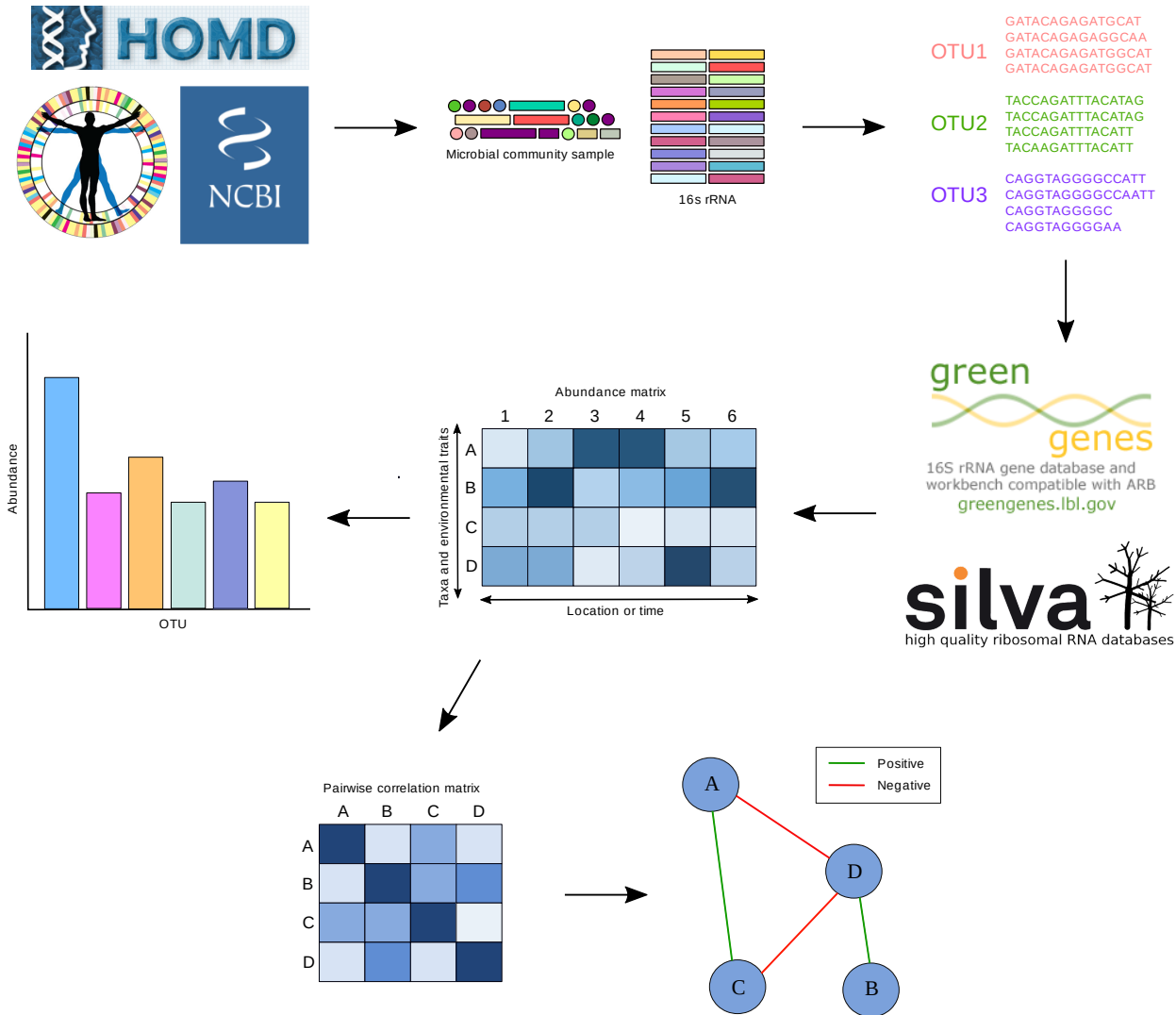
INTRODUCTION

- A comprehensive database of microbial interactions from different data sources
- A platform for users to compare, analyze and visualize their data
- Data population has started with associations derived from 16S data and literature

TERMINOLOGY

- Operational Taxonomic Units: Clusters of micro-organisms grouped by DNA sequence similarity in the 16S region
- Co-occurrence Networks: Networks obtained from the associations inferred in 16S rDNA samples from various environments
- Exact Sequence Variants: Amplicon sequence variants (ASVs) that are resolved exactly, down to the level of single-nucleotide differences over the sequenced gene region.

GOALS



GOALS

- Data collection:
 - Collect 16S rDNA sequence data
 - Human Microbiome Project, American Gut, etc...
- User data processing
 - Provide access to a pipeline for 16S data analysis
 - Data stored in the database is processed using the same standardized pipeline and parameters
- Data exploration
 - Disease vs. Healthy microbiome
 - Identify core or differential microbiota
 - Query microbial interactions by taxa level or based on environment or across
 - Allow the user to compare his network with networks with matching metadata

Demonstration

Toolbox × ↗

☐ Labels ☐ Auto Fit

Clear

Reset



Vibrio × Enter microbial name

Taxonomy filters:

☒ All

☒ ↑ phylum

☒ ↑ order

☒ ↑ genus

☒ ↑ class

☒ ↑ family

☒ ↑ species

Edge filters:

Weight cutoff

0

Number of Links

1

Contexts:

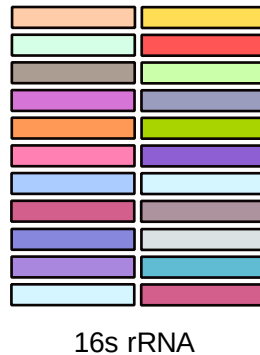
☒ All

Enter keywords of contexts

	<input checked="" type="checkbox"/>	0	Gut; T2D and non-diabetic control; Homo Sapiens; male, mid-age and normal-weight cohort; metabolic influence[C0006]
	<input checked="" type="checkbox"/>	0	Tooth surface; Healthy; Homo Sapiens; Early colonizers; binding; Experiment info:Manual curation based on Review: Communication among Oral Bacteria[C0007]
	<input checked="" type="checkbox"/>	0	Tooth surface; Healthy; Homo Sapiens; Late colonizers; binding; Experiment info:Manual curation based on Review: Communication among Oral Bacteria[C0008]
	<input checked="" type="checkbox"/>	0	lateral gene transfer by transduction from donors ??? recipients mediated by the phages; transduction; Experiment info:17158 recent transduction events in 3982 genomes[C0009]
	<input checked="" type="checkbox"/>	0	Colon mucosae; Healthy/Normal; Homo Sapiens; Gut mucosal microbiome of 61 healthy controls; correlation; 0.05; fdr; Experiment info:16S rRNA, SparCC[C0300]
	<input checked="" type="checkbox"/>	0	Colon adenoma mucosae; Colon adenoma; Homo Sapiens; Gut mucosal microbiome of 47 paired samples of adenoma and adenoma-adjacent mucosae; correlation; 0.05; fdr; Experiment info:16S rRNA, SparCC[C0300]
	<input checked="" type="checkbox"/>	0	Colon adenoma-adjacent mucosae; Colon adenoma; Homo Sapiens; Gut mucosal microbiome of 47 paired samples of adenoma and adenoma-adjacent mucosae; correlation; 0.05; fdr; Experiment info:16S rRNA, SparCC[C0300]
	<input checked="" type="checkbox"/>	0	Colon carcinoma mucosa; Colon carcinoma; Homo Sapiens; Gut mucosal microbiome of 52 paired samples of carcinoma and carcinoma-adjacent mucosae; correlation; 0.05; fdr; Experiment info:16S rRNA, SparCC[C0300]
	<input checked="" type="checkbox"/>	0	Colon carcinoma-adjacent mucosa; Colon carcinoma; Homo Sapiens; Gut mucosal microbiome of 52 paired samples of carcinoma and carcinoma-adjacent mucosae; correlation; 0.05; fdr; Experiment info:16S rRNA, SparCC[C0300]
	<input checked="" type="checkbox"/>	0	Skin, axilla; axillary odour; Homo Sapiens; 24 Caucasian male and female non-antiperspirant volunteers; correlation; Experiment info:16S rRNA, Spearman test[C0305]
	<input checked="" type="checkbox"/>	0	Skin, forehead; Healthy; Homo Sapiens; 200 skin samples from Chinese individuals living in Hong Kong; correlation; 0.05; Experiment info:16S rDNA sequence, SparCC[C0500]
	<input checked="" type="checkbox"/>	0	Skin, left forearm; Healthy; Homo Sapiens; 200 skin samples from Chinese individuals living in Hong Kong; correlation; 0.05; Experiment info:16S rDNA sequence, SparCC[C0501]
	<input checked="" type="checkbox"/>	0	Skin, right forearm; Healthy; Homo Sapiens; 200 skin samples from Chinese individuals living in Hong Kong; correlation; 0.05; Experiment info:16S rDNA sequence, SparCC[C0502]
	<input checked="" type="checkbox"/>	0	Skin, left palm; Healthy; Homo Sapiens; 200 skin samples from Chinese individuals living in Hong Kong; correlation; 0.05; Experiment info:16S rDNA sequence, SparCC[C0503]
	<input checked="" type="checkbox"/>	0	Skin, right palm; Healthy; Homo Sapiens; 200 skin samples from Chinese individuals living in Hong Kong; correlation; 0.05; Experiment info:16S rDNA sequence, SparCC[C0504]
	<input checked="" type="checkbox"/>	0	Adenoid; Uncertain; Homo Sapiens; 67 individuals who underwent adenoidectomy, weight=MIC-??? > 0.2; correlation; 0.05; Experiment info:16S rDNA sequence, MINE(maximal info)
	<input checked="" type="checkbox"/>	0	Gut, stool sample; Healthy; Homo Sapiens; CS1-3 following birth, total 25 at-term caesarean (CS) delivered neonates; correlation; 0.05; Experiment info:HITChip microarray data, Pearson co
	<input checked="" type="checkbox"/>	0	Gut, stool sample; Healthy; Homo Sapiens; CS7-30 following birth, total 25 at-term caesarean (CS) delivered neonates; correlation; 0.05; Experiment info:HITChip microarray data, Pearson co
	<input checked="" type="checkbox"/>	0	Gut, stool sample; Healthy; Homo Sapiens; V1-3 days following birth, total 6 vaginally (V) delivered neonates; correlation; 0.05; Experiment info:HITChip microarray data, Pearson co
	<input checked="" type="checkbox"/>	0	vagina: Bacterial vaginosis: Homo Sapiens: 220 women. 34% were Black and 44% were White women. 98 (43%) had BV by Amsel???s criteria. and 117 (53%) by Gram stain: correl

METHODS

Sequencing method



- 454 vs. illumina
- Different sequencing depths
- Different pre-processing
- Different quality filtering

METHODS

Clustering method

OTU1

GATACAGAGATGCAT
GATACAGAGAGGCAA
GATACAGAGATGGCAT
GATACAGAGATGGCAT

OTU2

TACCAGATTTACATAG
TACCAGATTTACATAG
TACCAGATTTACATT
TACAAGATTTACATT

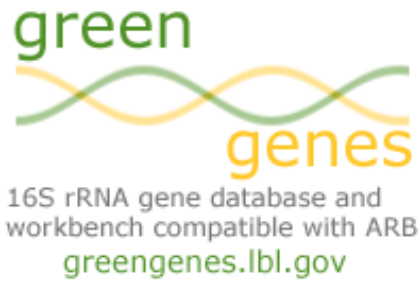
OTU3

CAGGTAGGGGCCATT
CAGGTAGGGGCCAATT
CAGGTAGGGGC
CAGGTAGGGGAA

- Closed reference
- Open reference
- De novo reference
- Error modeling

METHODS

Alignment method



- Greengenes
- Silva
- RDP

METHODS

Processing method

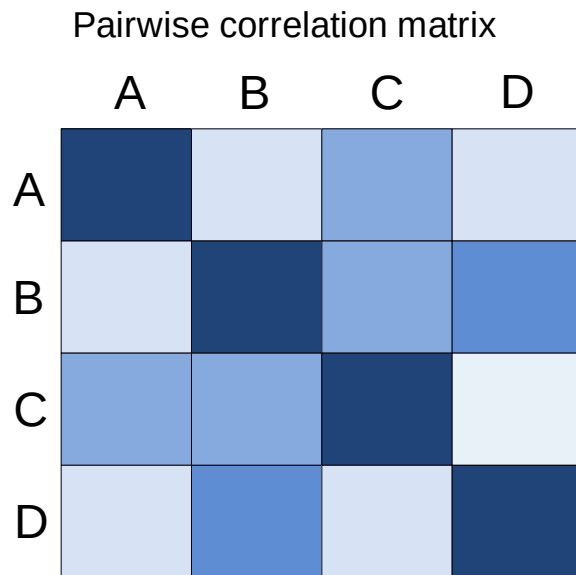
Original Abundance			Rarefied Abundance		
	A	B		A	B
OTU1	62	500	OTU1	62	50
OTU2	38	500	OTU2	38	50
Total	100	1000		100	100

Standard Tests for Difference			
P-value	chi-2	Prop	Fisher
Original	0.0290	0.0290	0.0272
Rarefied	0.1171	0.1171	0.1169

- Normalization
- Rarefying
- Mixture models
- Removing sparse data

METHODS

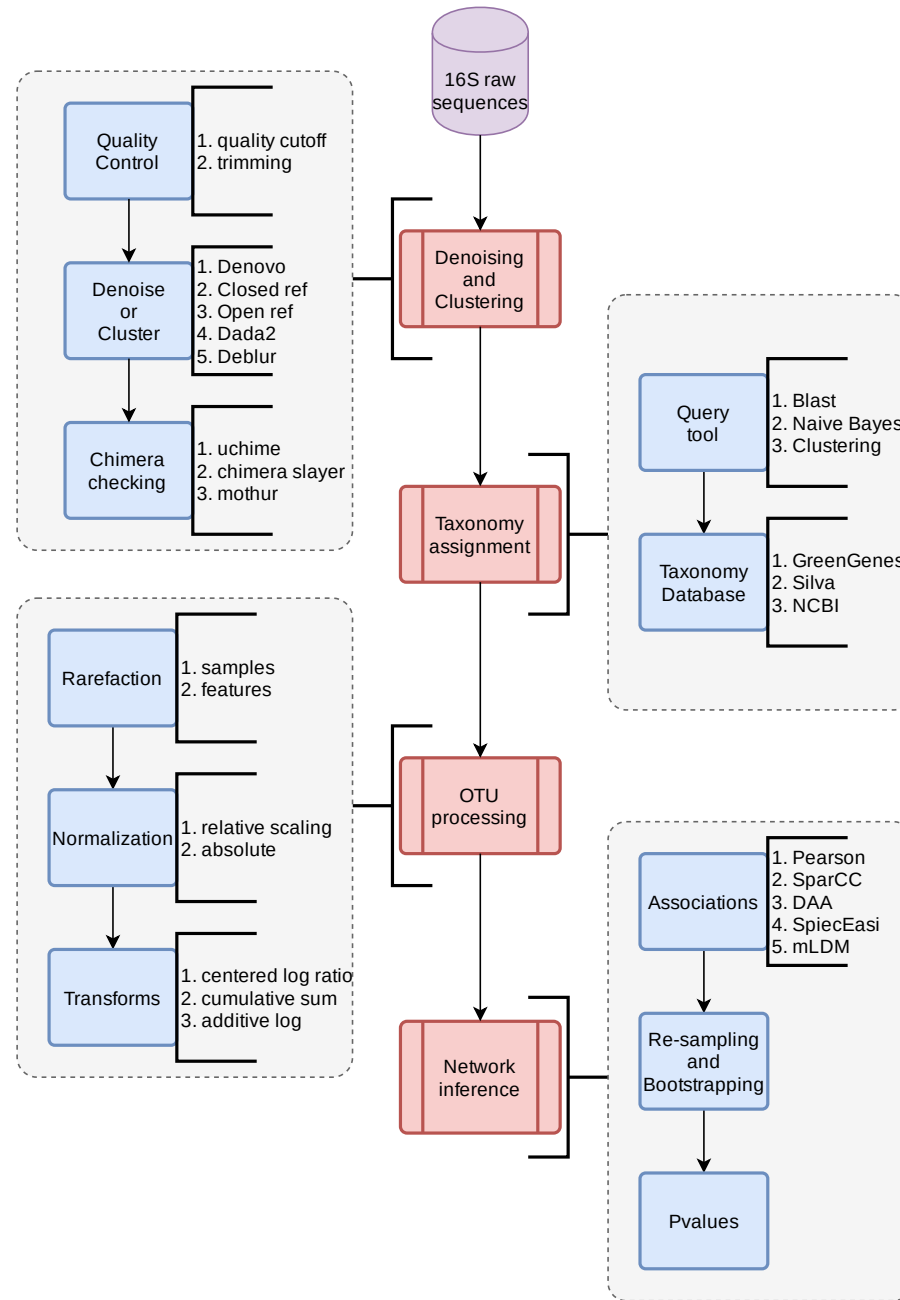
Calculating associations



- Compositionality
- Sparsity
- Pairwise associations
 - Pearson
 - Spearman
 - SparCC
- Direct associations
 - SpiecEasi
 - CCLasso
 - mLDM

PIPELINE

- Develop a standard pipeline - identify tools and good parameters
- This pipeline would be used to process all the data to be stored in the database
- Create a set of guidelines using a decision tree
- A consensus network created using a weighted voting scheme



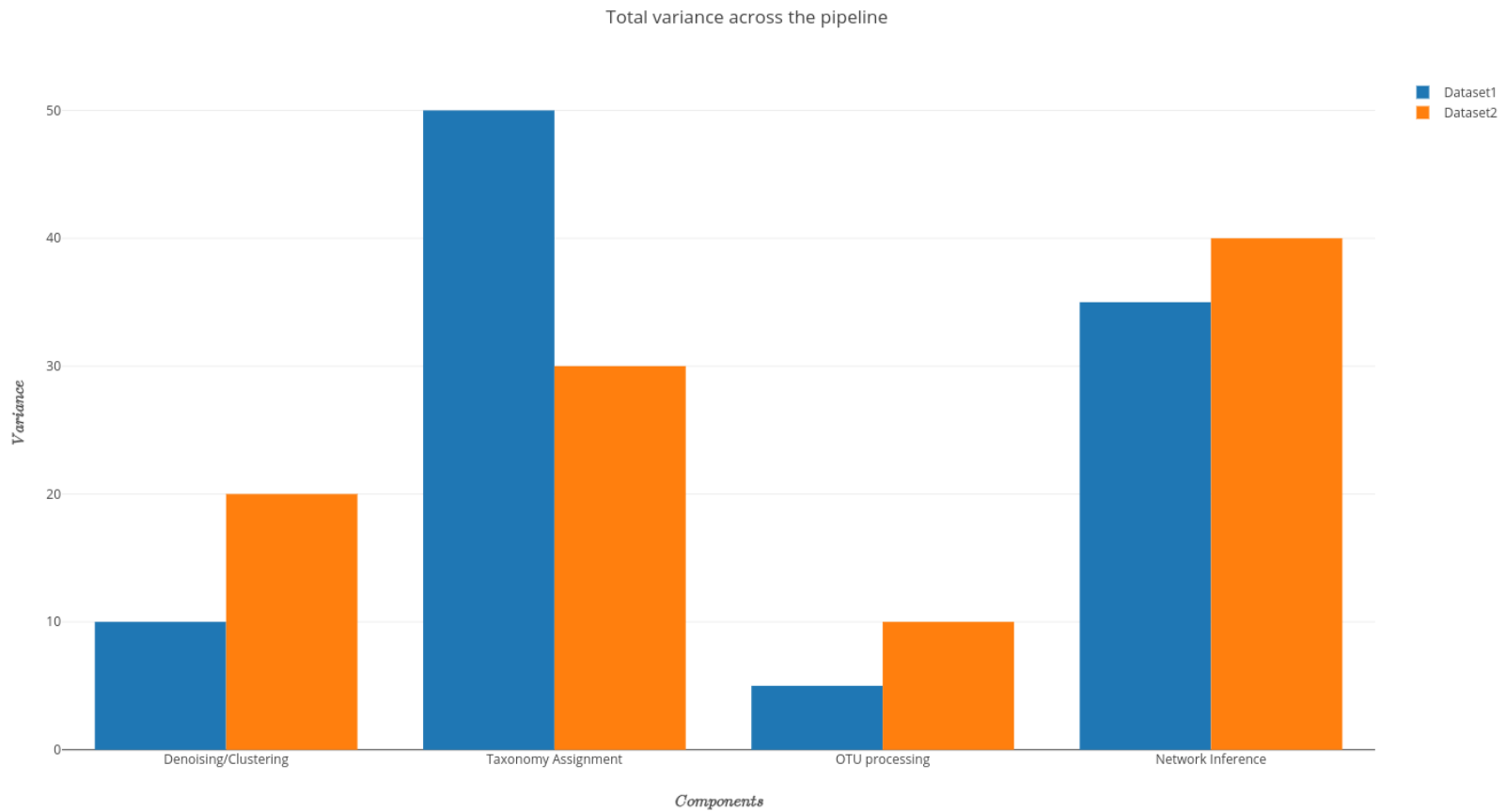
HOW DO WE FIND THE BEST METHODS?

- Is there a standard to compare against?
- Can we make synthetic data?
- What metrics can we use for comparisons?
 - Diversity
 - Abundances
 - Degree distributions
 - Motifs or connected components in the network
 - Metabolic models

WHAT CAN WE DO THEN?

- Identify parts of the pipeline that are most crucial
- Show how data/errors propagate through the pipeline
- Identify downstream steps that make choices in upstream steps inconsequential

MAIN FIGURE/IDEA



OTHER FIGURES

- Similarity, diversity, rank abundance and pairwise correlations after the denoising/clustering
- Diversity and correlations after assigning taxonomy at various taxonomy levels. Also, correlation between sequence similarity and assigned taxa
- Change in similarities after OTU filtering and normalization
- Edge and node overlaps, similarity and degree distribution after network inference
- A embedding of all the networks in a 2D plane

DISTANCE METRICS

- Between abundances in OTU tables
- Between assigned taxonomies in OTU tables
- Between associations in networks

Demonstration

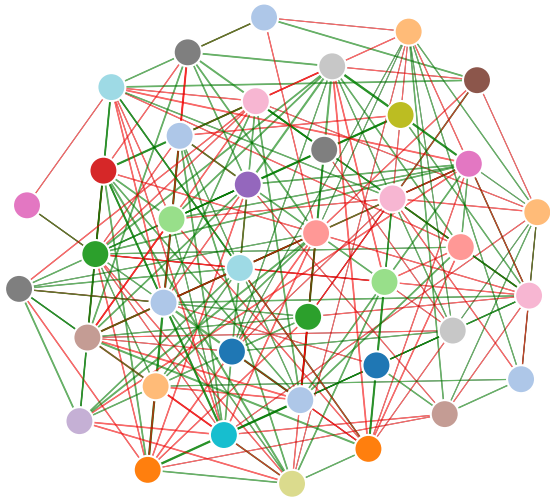
Comparing various methods in 16S data analysis

Enter the location of the source directory

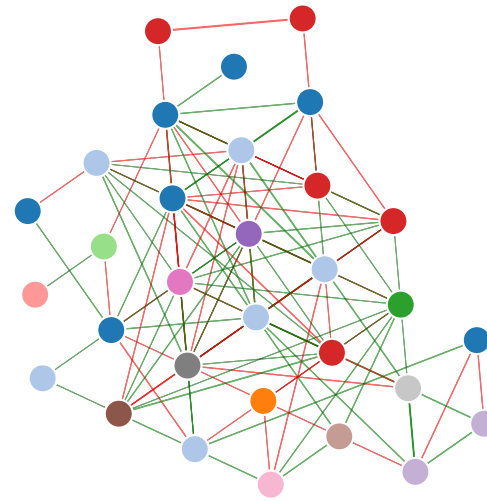
/home/dileep/Documents/boston_university/mind-analysis/mind_analysis/data/example_source

Enter source

NETWORKS

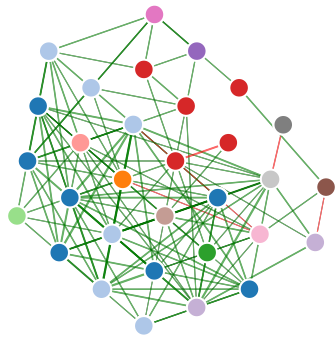


Saliva Network

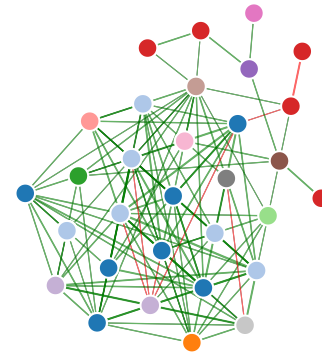


Stool Network

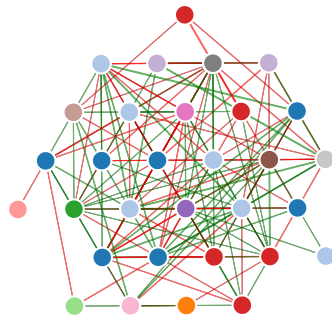
NETWORK INFERENCE METHODS



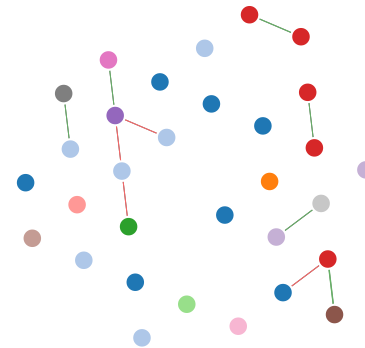
Pearson



Spearman



SparCC



SpiecEasi

DENOSING/CLUSTERING METHODS



Closed Reference



Open Reference

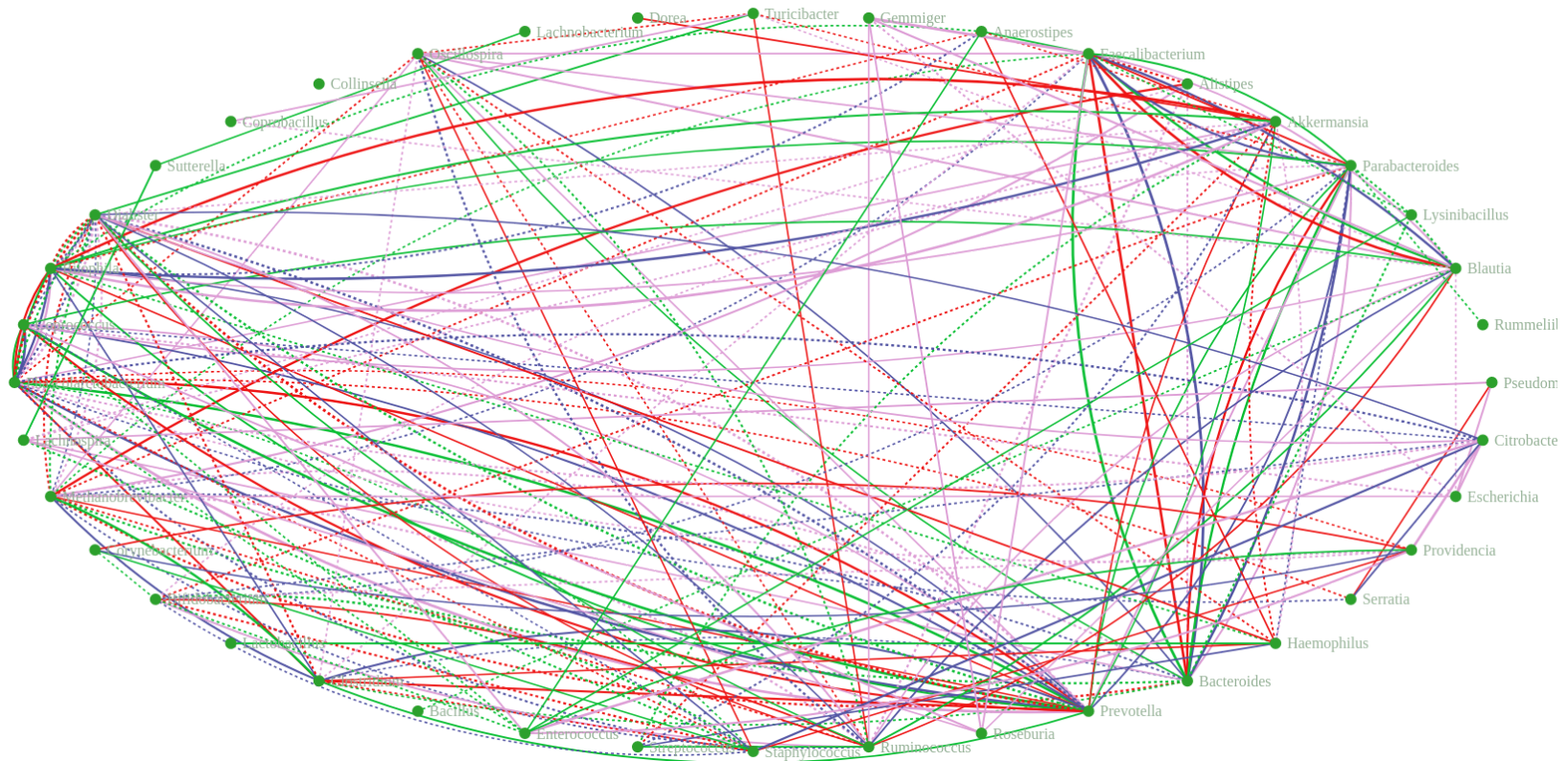


Denovo

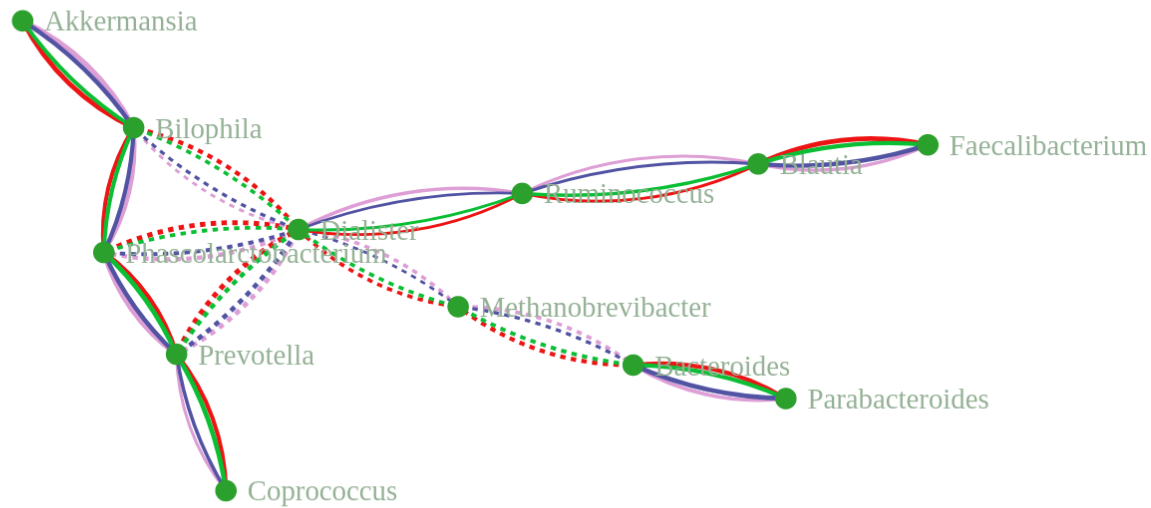


Dada2

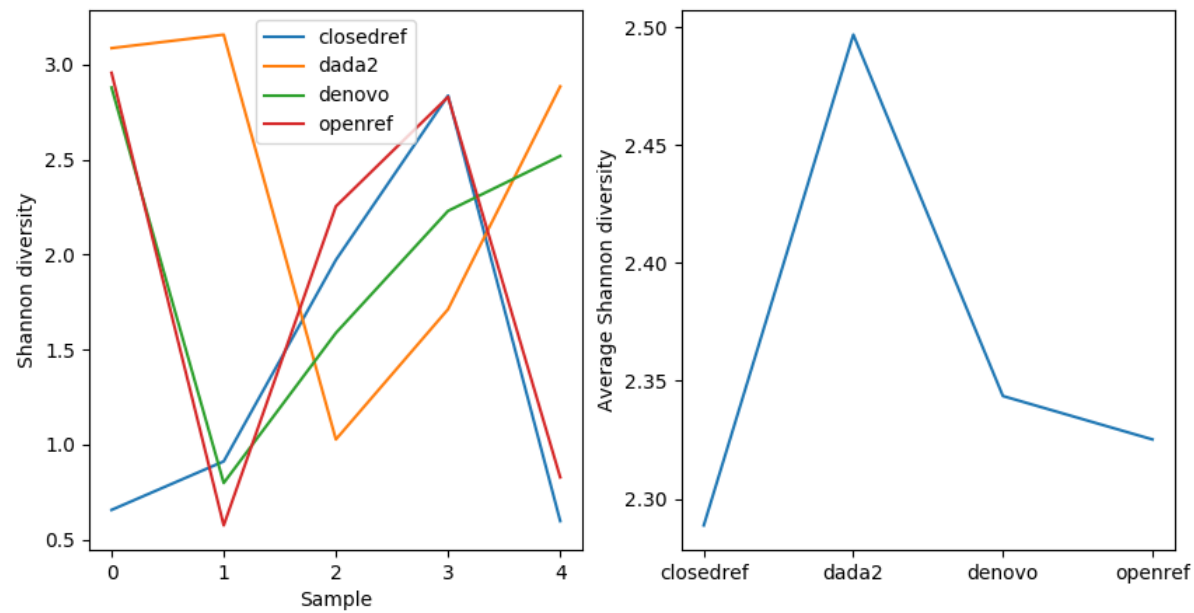
DENOSING/CLUSTERING METHODS



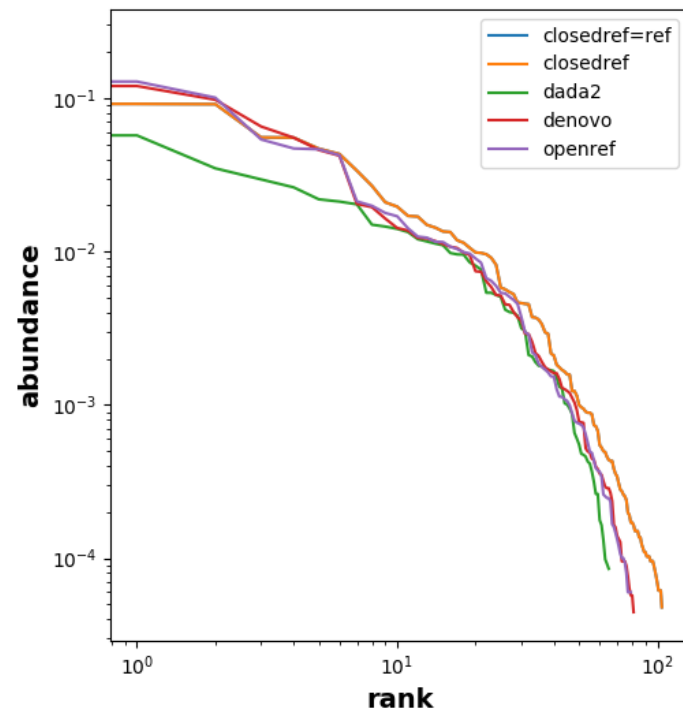
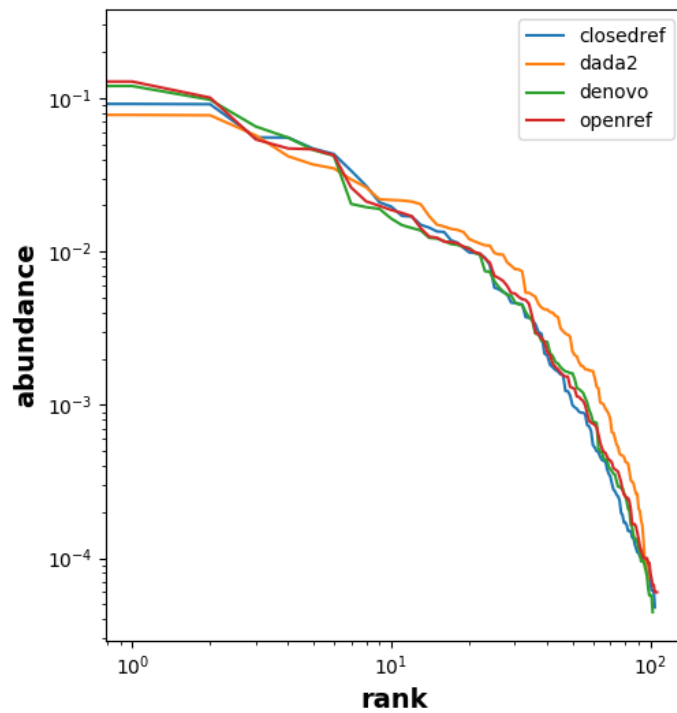
DENOSING/CLUSTERING METHODS



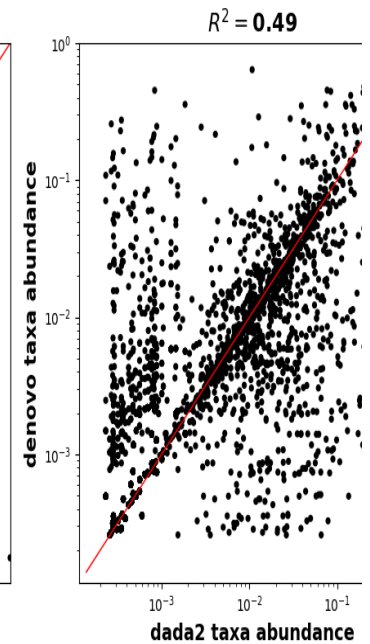
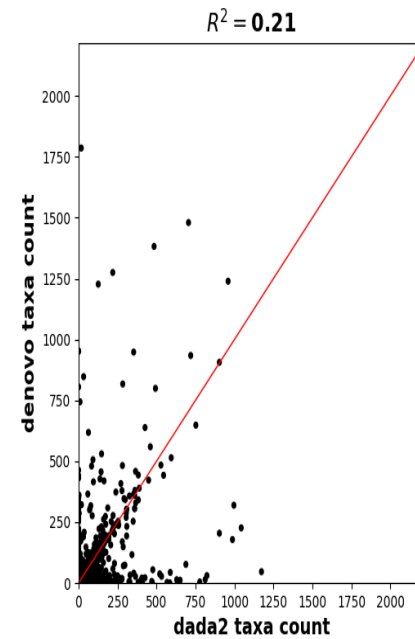
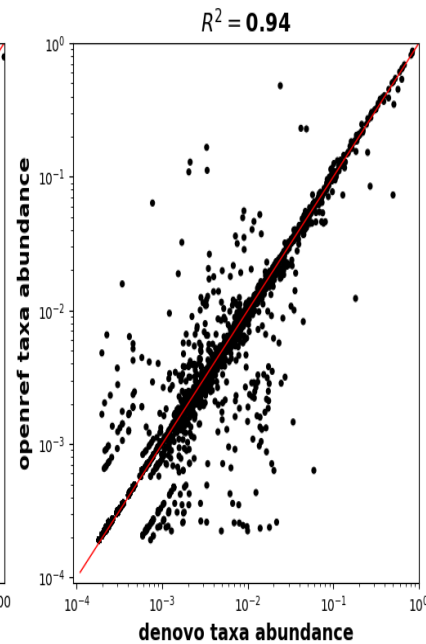
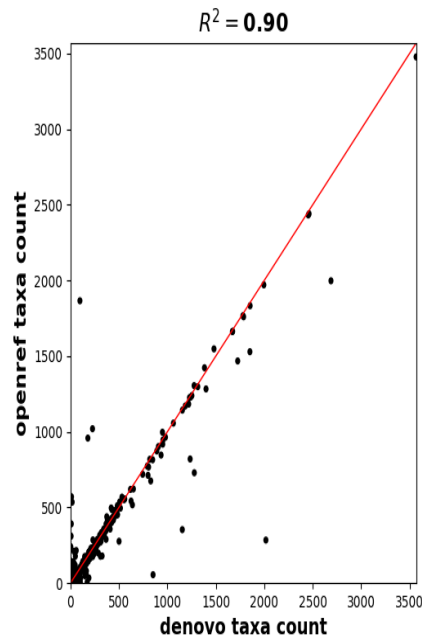
DIVERSITY



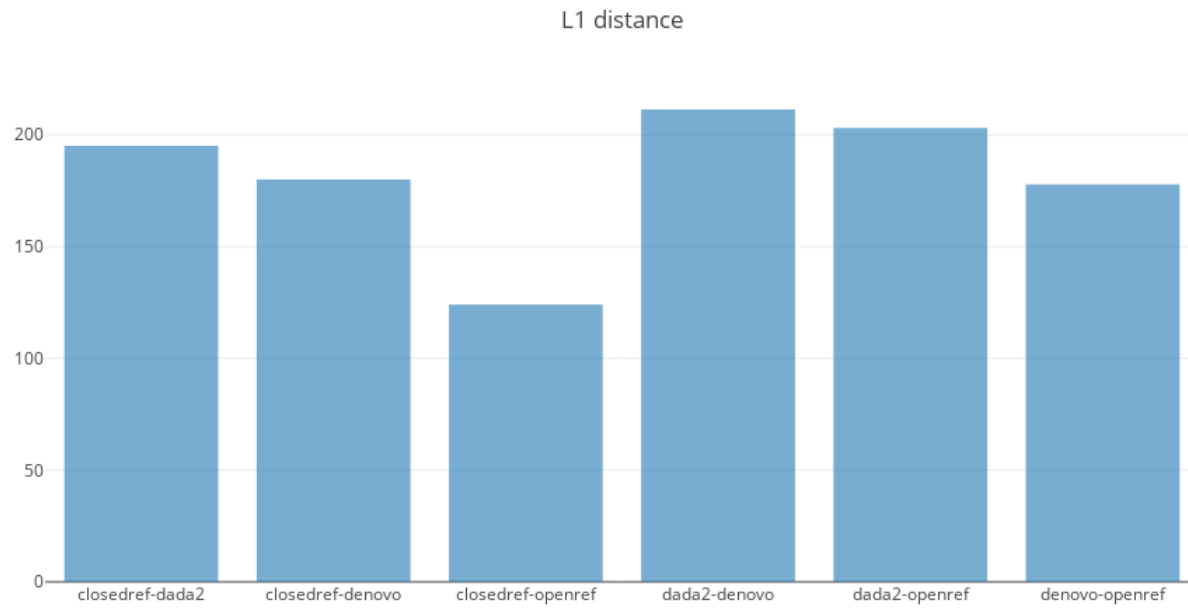
RANK ABUNDANCE



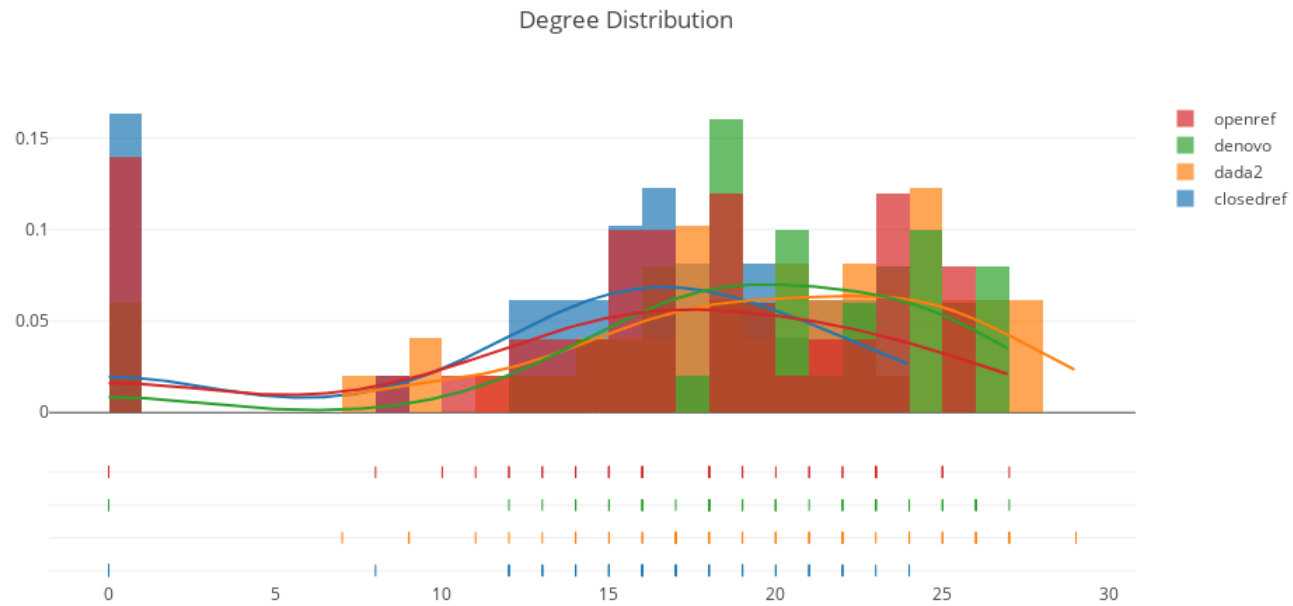
ABUNDANCE CORRELATION ON COMMON TAXA



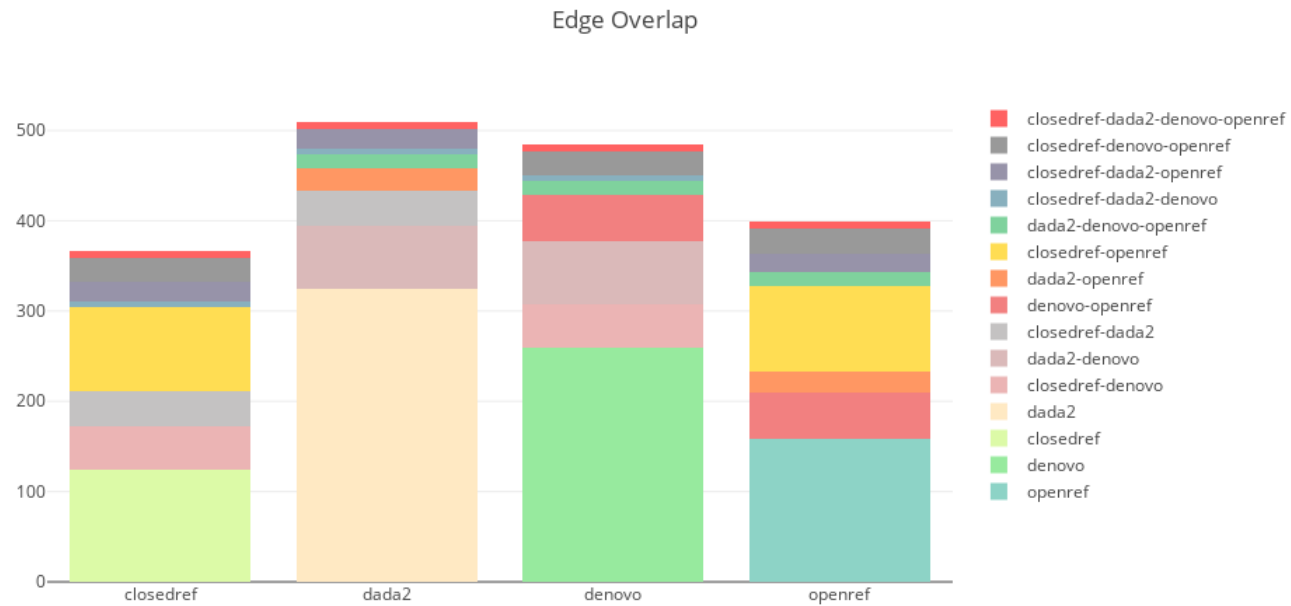
NETWORK PROPERTIES



NETWORK PROPERTIES



NETWORK PROPERTIES



| Thank you!