**Pre 8/19/14 Meeting work**

Network analysis continued:

Now that I have an association network for soil genera from the

Wilson dataset using ccrepe and have attempted to visualize this network

with gephi, next objective is to apply basic network analysis methods to this network.

To do’s in bold.

Presentation:

1. slide going backwards: What is network analysis good for and what general questions can be answerered – shown below.
2. Specific points bolded below made into graphs/tables.
3. References to T Zhang, Wilson, and huttenhower

Python network analysis tool/data format?

Some general questions that I hope to answer:

What is the overall network structure? Does the connectivity follow a power-law?

Is this network a "small-world" network?

**data requirements: Degree distribution, mean path length, diameter, clustering coefficient, modularity index**

Are there "hubs" within this network? Do hubs behave differently than non-hubs?

**Data requirements: nodes sorted by degree, some cutoff criteria. Then identification of within and between module hubs. Degree vs abundance chart.**

Do phylogenetically more related groups interact differently than random groups?

**Data requirements: inter vs intra clade positive and negative association graph**

"Specialists" vs generalists: are organisms that are highly abundant in one ecosystem but not present in others

different (statistically, over all node/edges) than organisms that exist in most environments?"""

Association network properties:

Node connectivity in association network from Wilson paper follows a power law distribution: Proability of OTU with k connections ~ kY, Y =-.6. “Small-world” networks often have Y between 2-3.5, these types of networks have “hubs”, some nodes with much higher connectivity that act as centers for preferential network attachment.

Because the power law coefficient is low, this association network is more even or random than typical small world networks.

Next up: Modules/clusters within this network.

**Tuesday, August 19, 2014**

Hub identity: Given that some nodes are more connected (hubs) than others in this network, are there any commonalities between hubs? Are they generalists/specialists/etc?

Goals:

dataframe appended with specialist /abundance data

plots of degree vs abundance, degree vs specialist/generalist

results added to update docx.

Code commented and doc’ed somewhere

run home

work in 30 minute chunks

done:

responded to Mary Brand, signed up for data analysis class.

Thursday:

**Added degree vs abundance, degree vs # of occurrence.**

Inter va intra class associations? How to handle this:

I have an edge list with “k;p;c…” naming convention.

Filter this based on rules (positive/negative R>.9, q<.05)

For inter vs. intra class:

Function that takes an edge dataframe (or row?) and returns all unique classes + counts?

Function that takes an edge dataframe

**Monday, August 25, 2014**

Sampling stuff for tomorrow:

Bring sampling kit + completed forms

9:30 AM Kirie: Ask for James Kaminski + John Smoody

address: 3, contact #:

Thursday, September 4, 2014

Story up to now:

Used ccrepe on soil microbiome study

* In original paper, they used unifrac (define?) to demonstrate consistent differences between endo/rhizo/bulk soil microbiomes. They used RDA/ANOVA to show statistically significant differences between communities from different soil sites/strains.
* My approach was to use co-occurrence networks to identify potential interactions in soil dataset.

Procedure:  
Generate co-occurrence network in R-CCREPE implementation – point is to avoid compositional effects. Compositional effects strongest in environments with low alpha diversity. (I think one thing I’m missing in my work to date is a control – what does the network look like if I use Pearson and Pearson with shuffling?)

Analyze network generated by this procedure? Small-world etc.

Found power-scaling of connectivity

Find hubs – determine if hubs are overrepresented by some biological group/explanation

**Friday, September 5, 2014**

Inter vs intra class associations:

Which are overrepresented and why?

C\_Actinobacteria is highly intra-class connected. Why is this?