Using network models for analyzing microbiome data

Outline:

- **Brief introduction to metagenomics**

- **Case study: Phyllosphere fungal networks, and
                        Bare Patch Disease of Wheat**

- **Hands-on example**

By
Ravin Poudel
Ricardo I. Alcala-Briseño
Garrett Lab
ICPP, Boston, MA

UF|IFAS
UNIVERSITY of FLORIDA

Institute for
Sustainable Food
Systems

University of Florida
Plant Pathology

Brief introduction to metagenomics

Genomics: Study of structure, function, evolution, and mapping of genomes.

Metagenomics (Environmental Genomics or **Community Genomics**) is the study of genomes recovered from environmental samples without the need for culturing them.

Culture-independent analysis
- ❖ 16S ribosomal RNA (rRNA) sequencing
- ❖ Whole genome sequencing
- ❖ **Metagenomics: No need for culturing/ use samples directly from environment**

Example: 16S ribosomal RNA (rRNA) sequencing

Sample

↓

PCR amplification

↓

Sequencing

↓

Identification

OTUs / Taxa

Sample

Count Data
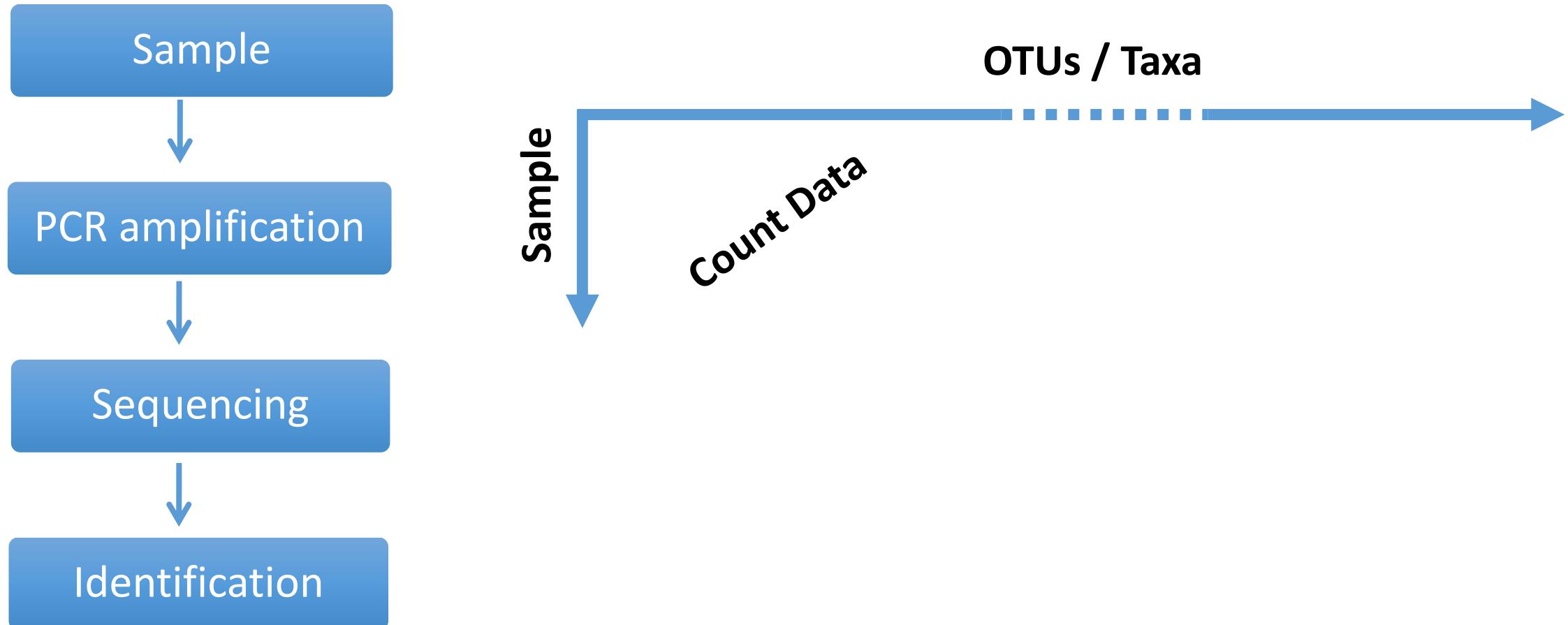
Case study- metagenomics and network models

- OTU/ taxon is represented as a node

- Links defines the relationship between two OTUs

- Various methods can be used to define the links

  - Presence or absence of association (co-occurrence)

  - Statistical approach to define the association
    - Such as correlation, proportionality

    - Methods robust for metagenomics data ~ deal with compositional bias and spurious correlation.
      - SparCC, SpiecEasi….. and so on…

   - Mostly, to reduce compositional bias associated with data type.
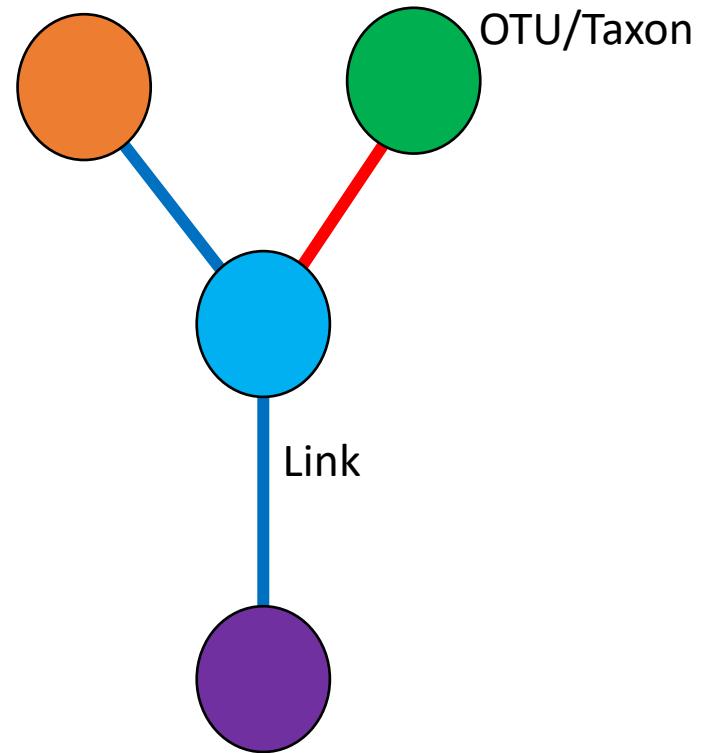
OTU/Taxon

Link

General framework

Microbiome Networks: A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management
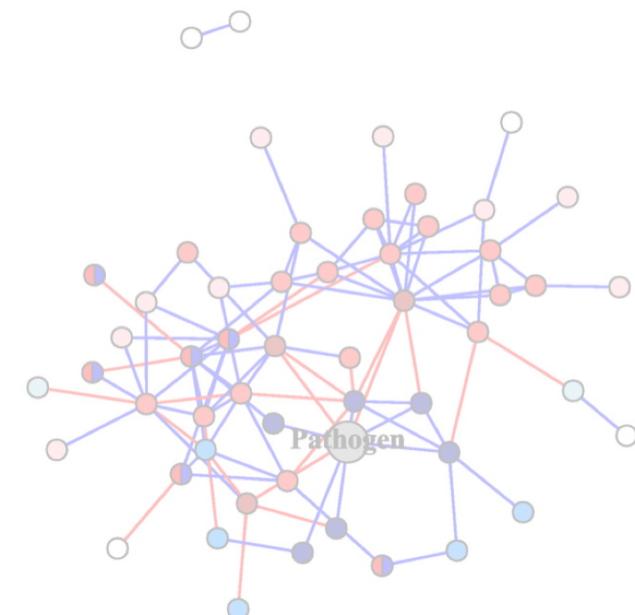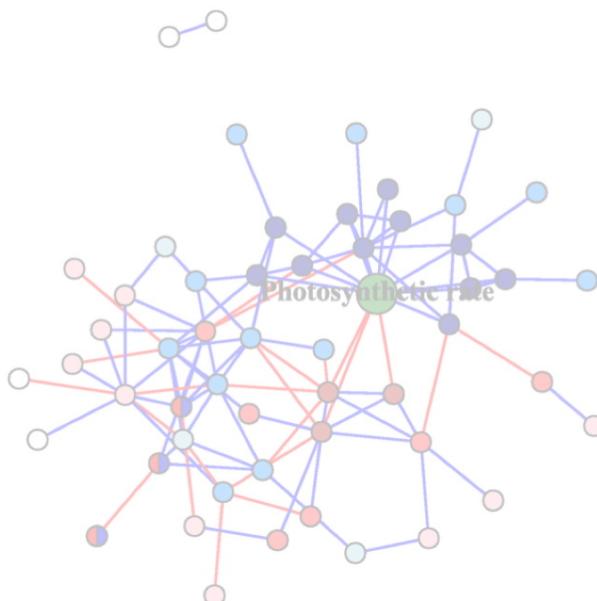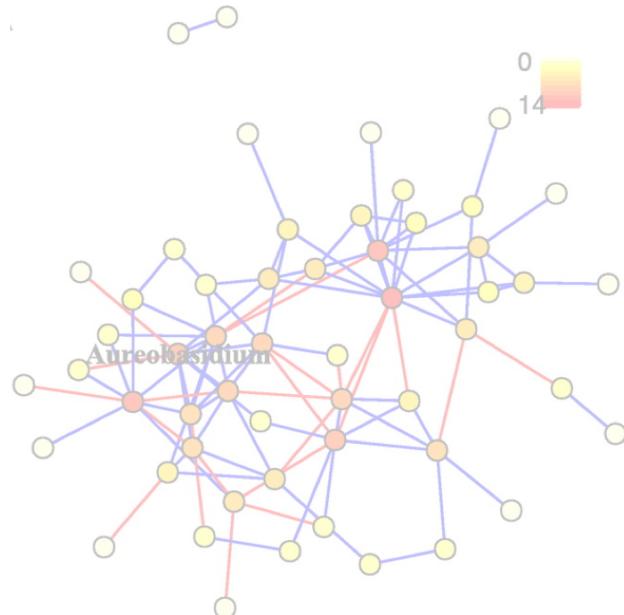
Goal: To select potential candidate taxa based on – interactions and network attributes
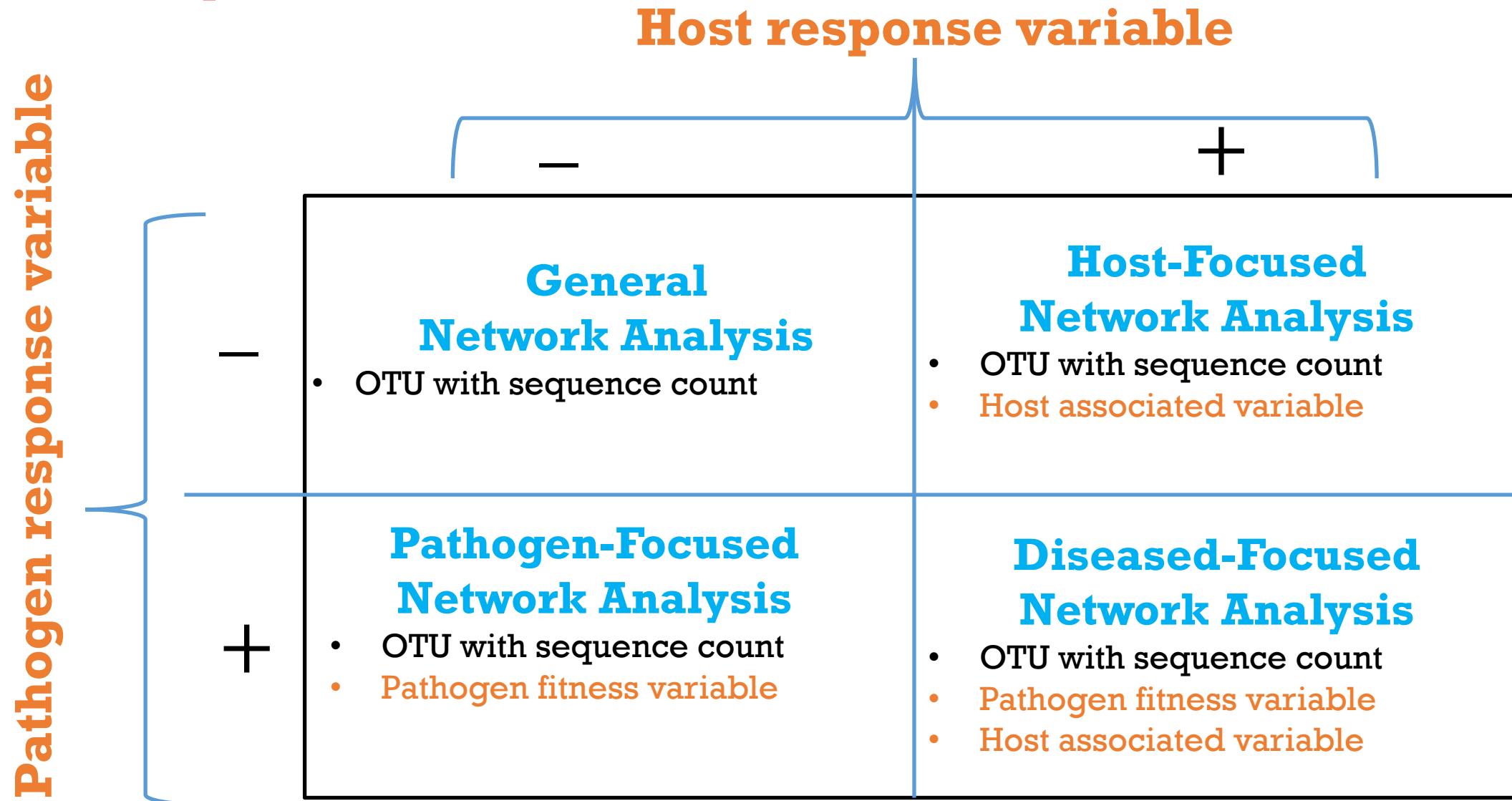
Host response variable

Pathogen response variable

	−	**+**
−	**General Network Analysis** • OTU with sequence count	**Host-Focused Network Analysis** • OTU with sequence count • Host associated variable
+	**Pathogen-Focused Network Analysis** • OTU with sequence count • Pathogen fitness variable	**Diseased-Focused Network Analysis** • OTU with sequence count • Pathogen fitness variable • Host associated variable

Microbiome networks: A systems framework for identifying candidate microbial assemblages for disease management

- Fungal phyllosphere data (*Jumpponen et. al., 2010*)

- Node = Fungal taxon

- Links = association calculated using SparCC

Data Source:

Research

New
Phytologist

Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments

A. Jumpponen[1,2] and K. L. Jones[2,3]

Microbiome networks: A systems framework for identifying candidate microbial assemblages for disease management


General analysis
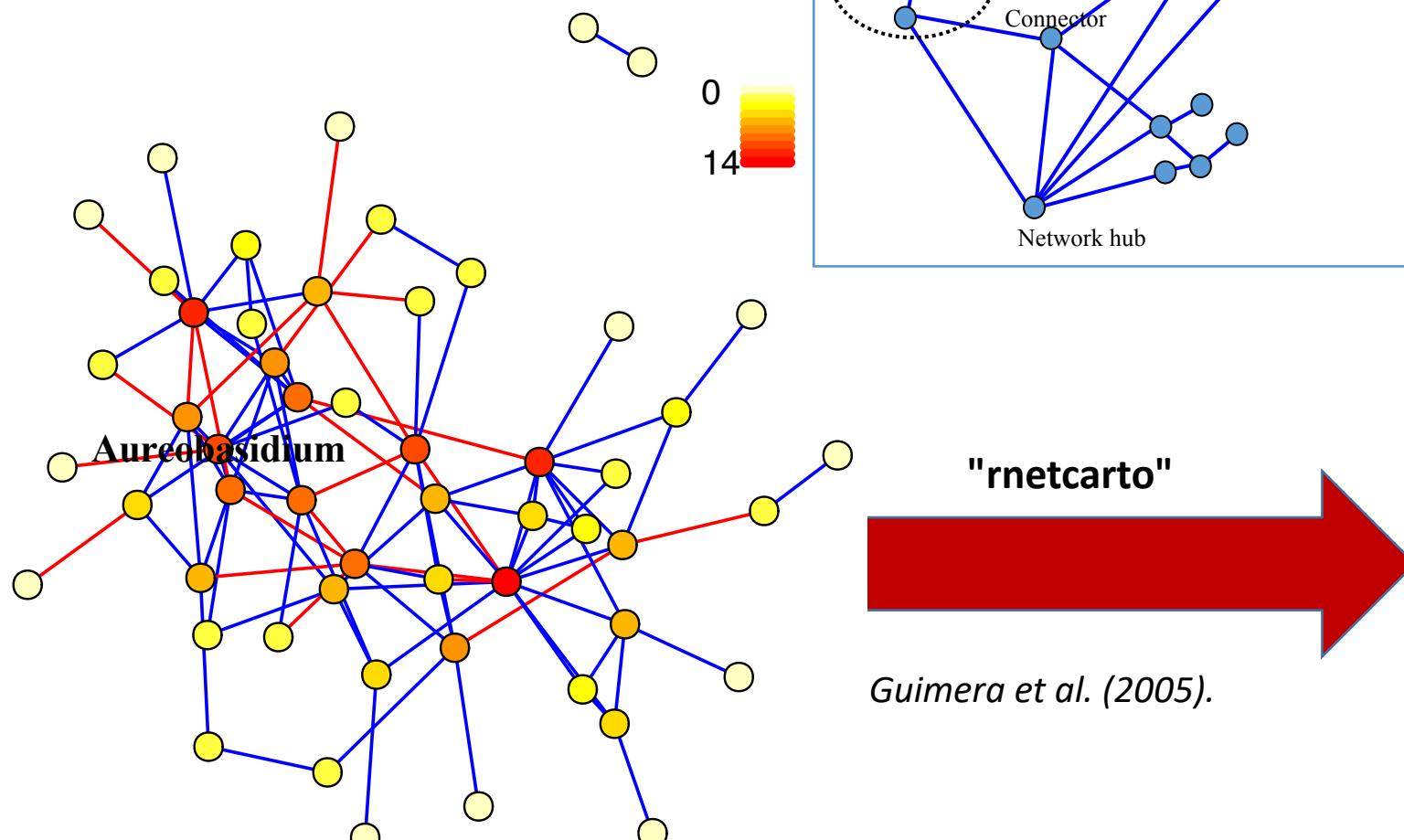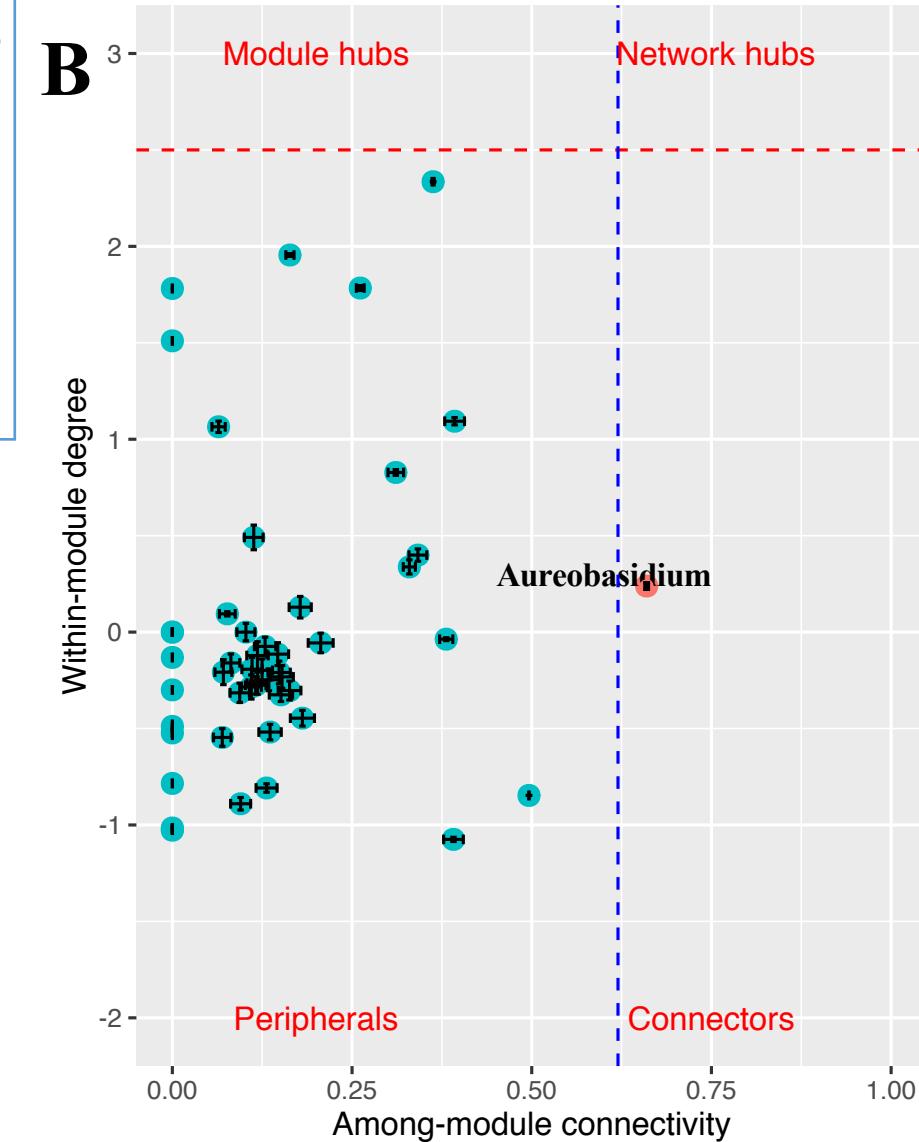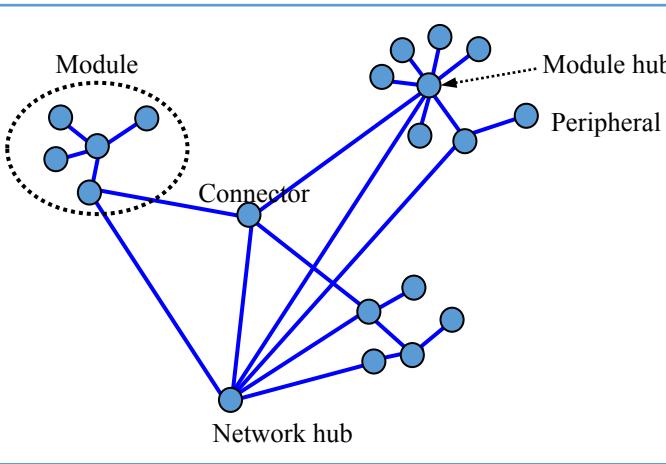
Module
Module hub
Peripheral
Connector
Network hub

0
14

"rnetcarto"

Guimera et al. (2005).

B

Module hubs
Network hubs

Aureobasidium

Peripherals
Connectors

Within-module degree

Among-module connectivity

Aureobasidium

Microbiome networks: A systems framework for identifying candidate microbial assemblages for disease management

Host-focused analysis
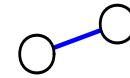
Blocking→ Negatively Linked Nodes

First degree neighbors
Second degree neighbors
Third degree neighbors
Mixed Association

Photosynthetic rate
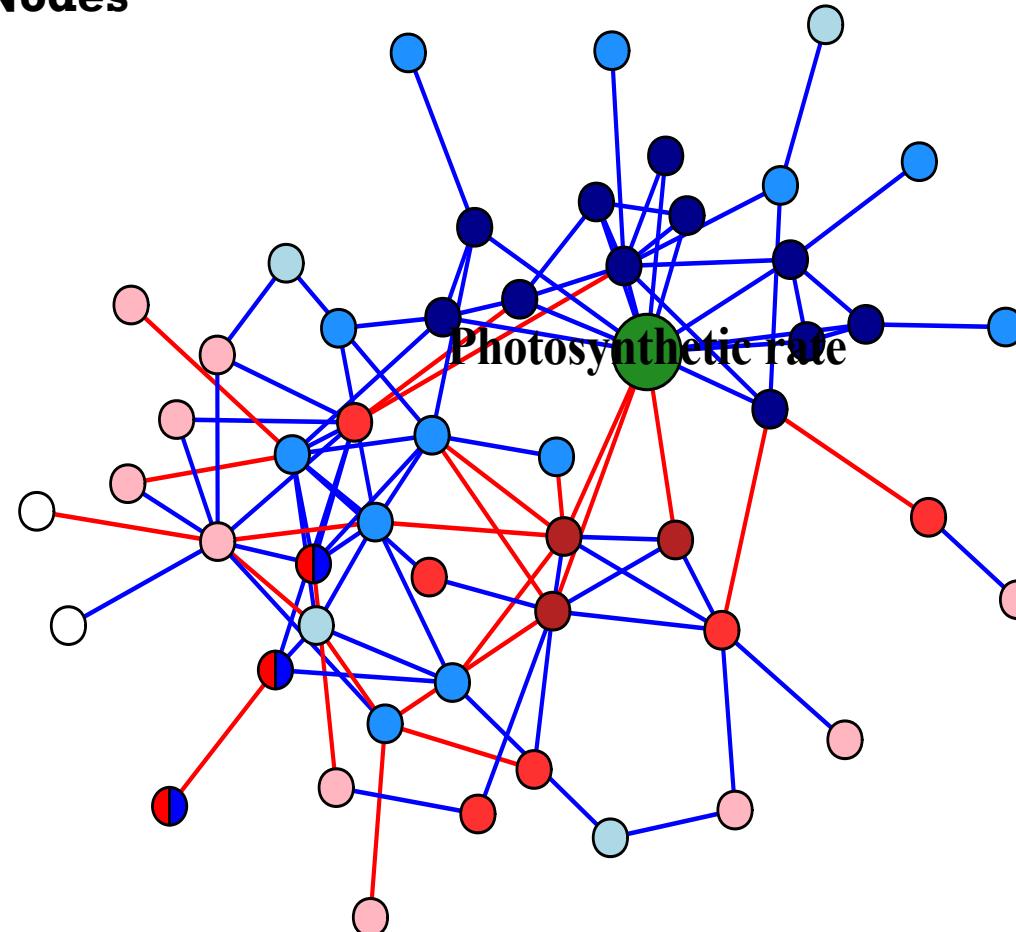
Goal → Biofertilizers

First degree neighbors
Second degree neighbors
Third degree neighbors
Mixed Association

Microbiome networks: A systems framework for identifying candidate microbial assemblages for disease management

Pathogen-focused analysis

Goal → Biocontrols

First degree neighbors
Second degree neighbors
Third degree neighbors
Mixed Association

Blocking→ Positively Linked Nodes

First degree neighbors
Second degree neighbors
Third degree neighbors
Mixed Association
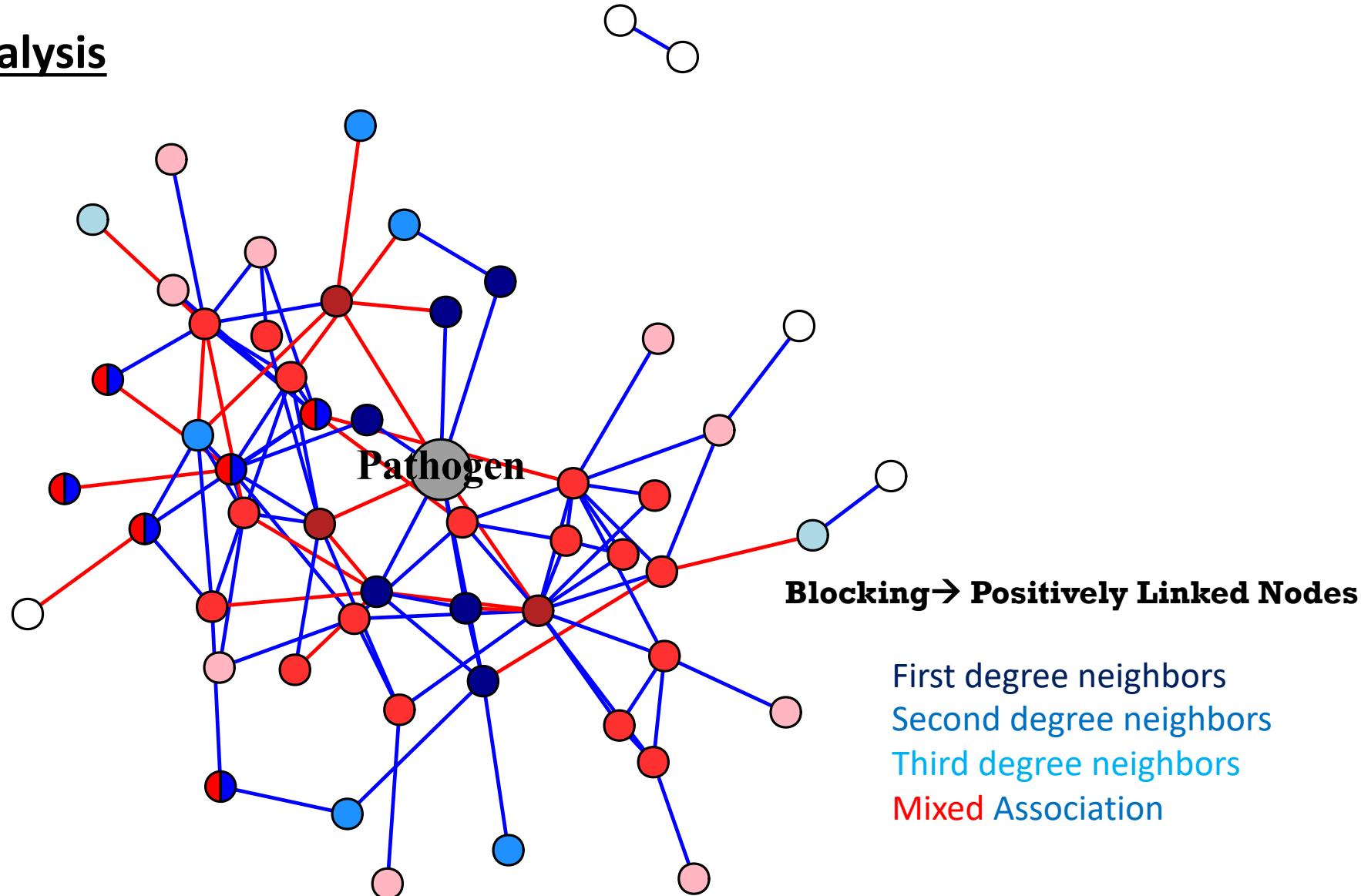
Disease-focused analysis

Data source:

Role of Bacterial Communities in the Natural Suppression of *Rhizoctonia solani* Bare Patch Disease of Wheat (*Triticum aestivum* L.)

Chuntao Yin,[a] Scot H. Hulbert,[a] Kurtis L. Schroeder,[a] Olga Mavrodi,[a] Dmitri Mavrodi,[a] Amit Dhingra,[b] William F. Schillinger,[c] Timothy C. Paulitz[d]

Applied and Environmental Microbiology, 2013

Berendsen, 2012

TRENDS in Plant Science
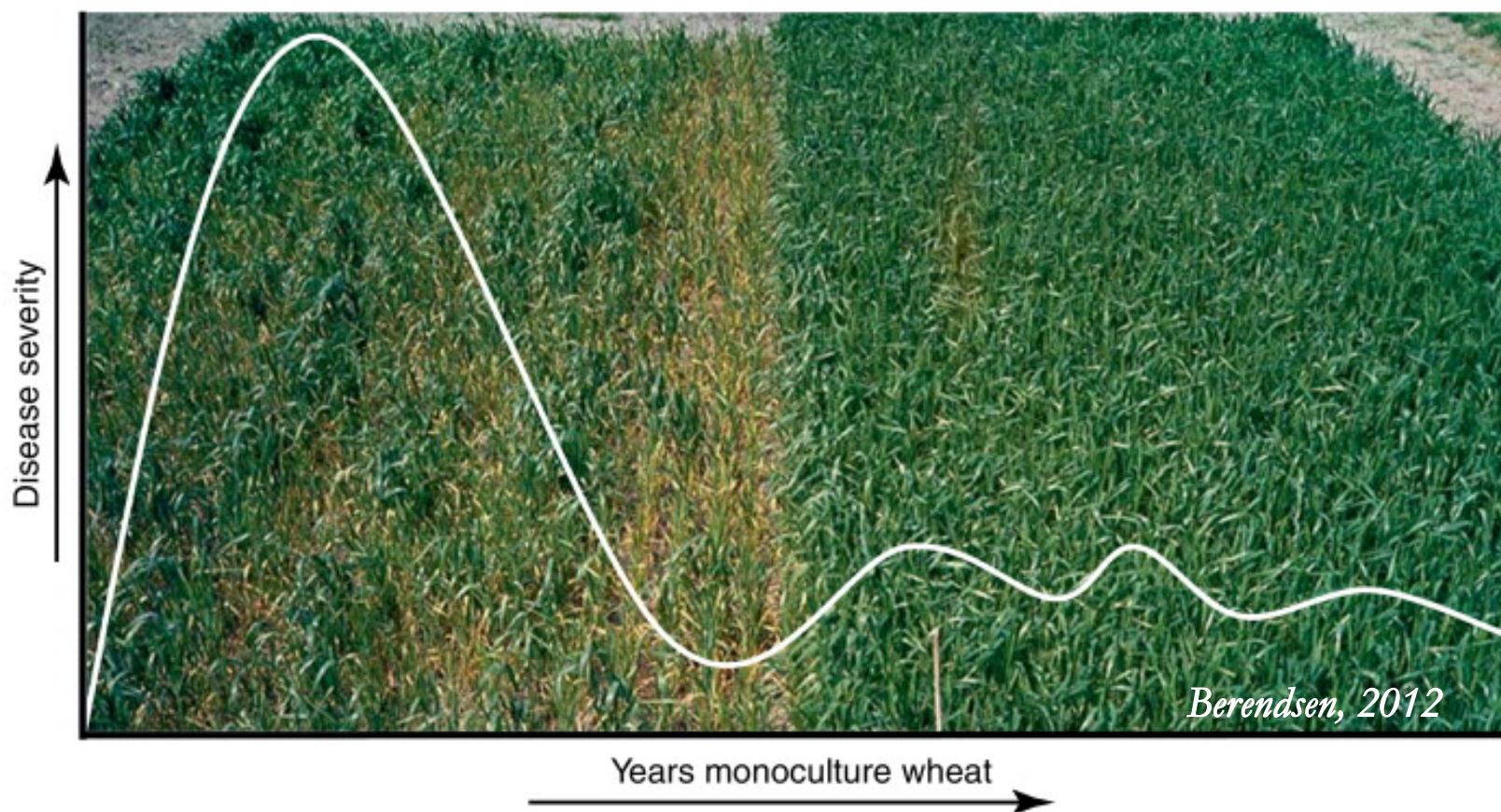
Disease-focused analysis

- Blue link = positive association with *Rhizoctonia solani*

- Red link = negative association with *Rhizoctonia solani*

- Node size= Abundance

- Node associated with disease

- Node associated with healthy state
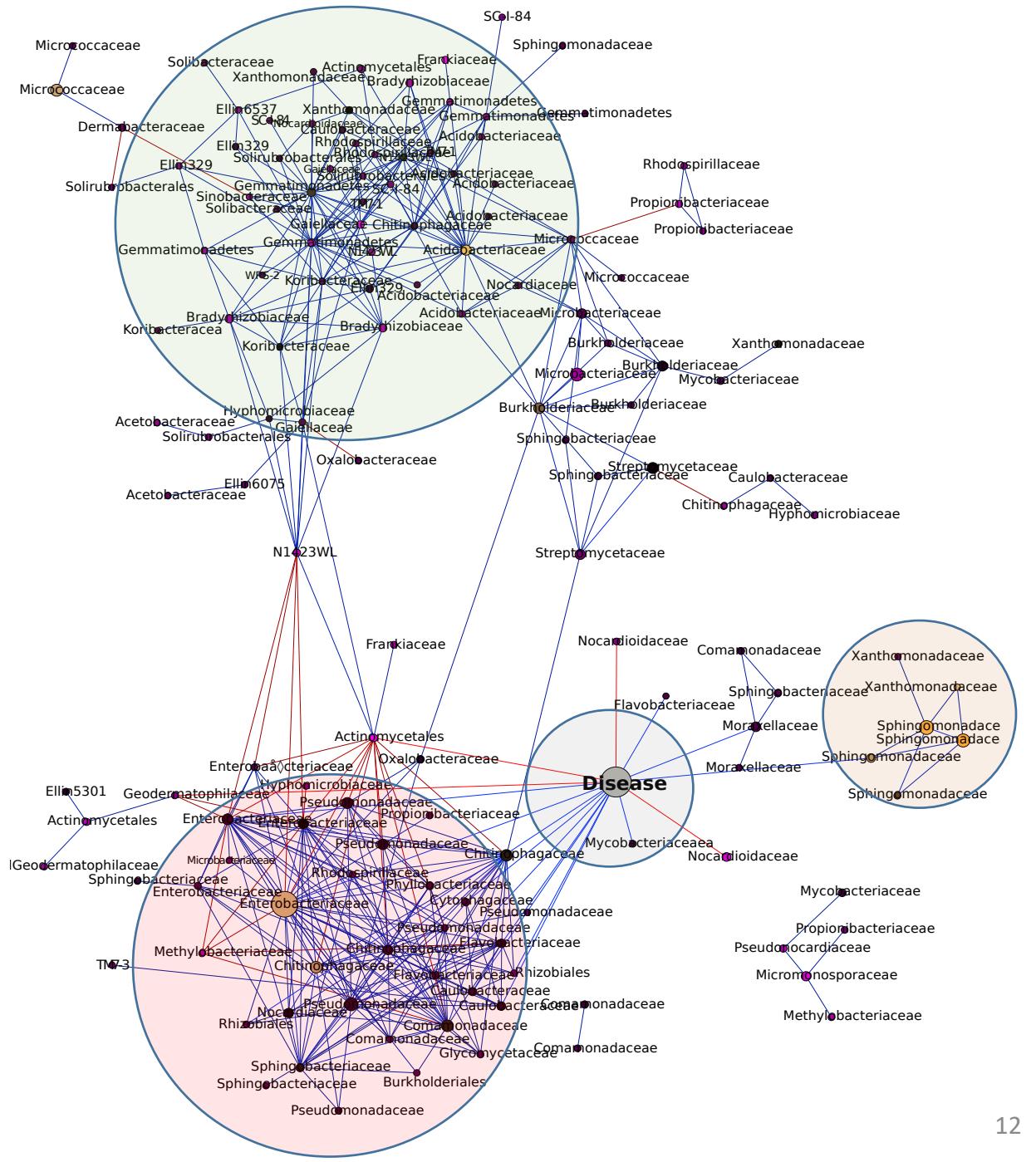
Do grafting and rootstock genotype affect the rhizobiome? A study of tomato systems
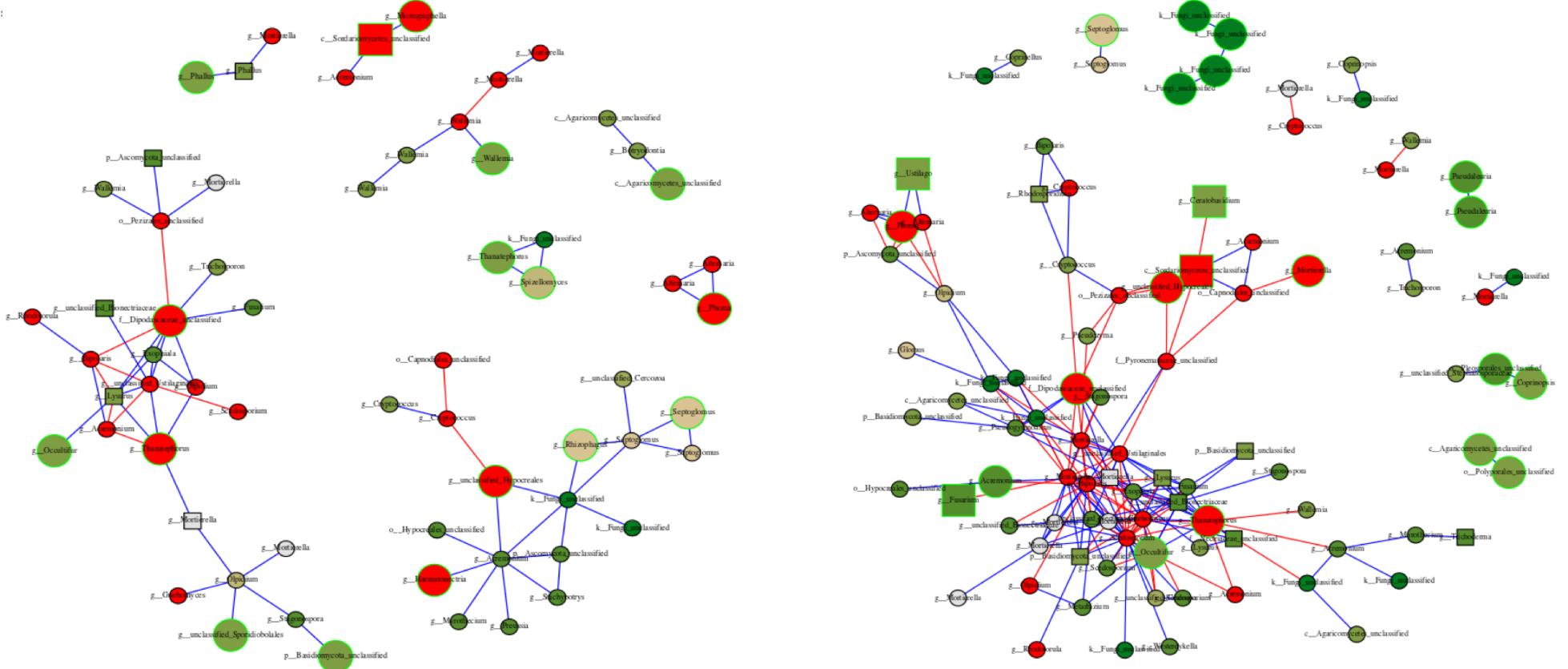
Hands-on experience on building network models with microbiome date

Let's switch to RStudio

# Loading required packages
library(igraph)
library(Hmisc)
library(Matrix)

```
# Load the data with the OTU table: otudata.csv
otu.table<-read.csv(file.choose(), header=T, row.names = 1)

# Read taxonomy file associated with OTU table into new object: otu_taxonomy.csv
tax<-read.csv(file.choose(),header=T, row.names = 1)

# Check how many OTUs we have
dim(otu.table)

# Check for the decrease in the number of OTUs
dim(otu.table.filter)

# Keep the OTUs with more than 10 counts
otu.table.filter<-otu.table[,colSums(otu.table)>10]
```

```
# Calculate pairwise correlation between OTUs
otu.cor<-rcorr(as.matrix(otu.table.filter), type="spearman")

# Get p-value matrix
otu.pval <- forceSymmetric(otu.cor$P) # Self-correlation as NA

# Select only the taxa for the filtered OTUs by using rownames of otu.pval
sel.tax <-tax[rownames(otu.pval),,drop=FALSE]

# Sanity check
all.equal(rownames(sel.tax), rownames(otu.pval))
```

```
# Filter the association based on p-values and level of correlations
p.yes<-otu.pval<0.001

# Select the r values for p.yes
r.val=otu.cor$r # select all the correlation values
p.yes.r<-r.val*p.yes # only select correlation values based on p-value criterion

# Select OTUs by level of correlation
p.yes.r<-abs(p.yes.r)>0.75 # output is logical vector
p.yes.rr<-p.yes.r*r.val # use logical vector for subscripting.
```

```
# Create an adjacency matrix
adjm<-as.matrix(p.yes.rr)

# Add taxonomic information associated with adjacency matrix
colnames(adjm)<-as.vector(sel.tax$Family)
rownames(adjm)<-as.vector(sel.tax$Family)

# Create an adjacency matrix in igraph format
net.grph=graph.adjacency(adjm,mode="undirected",weighted=TRUE,
diag=FALSE)
```

```
# Calculate edge weight == level of correlation
edgew<-E(net.grph)$weight

# Identify isolated nodes
bad.vs<-V(net.grph)[degree(net.grph) == 0]

# Remove isolated nodes
net.grph <-delete.vertices(net.grph, bad.vs)
```

```
# Plot the graph object
plot(net.grph,
        vertex.size=4,
        vertex.frame.color="black",
        edge.curved=F,
        edge.width=1.5,
        layout=layout.fruchterman.reingold,
        edge.color=ifelse(edgew<0,"red","blue"),
        vertex.label=NA,
        vertex.label.color="black",
        vertex.label.family="Times New Roman",
        vertex.label.font=2)
```
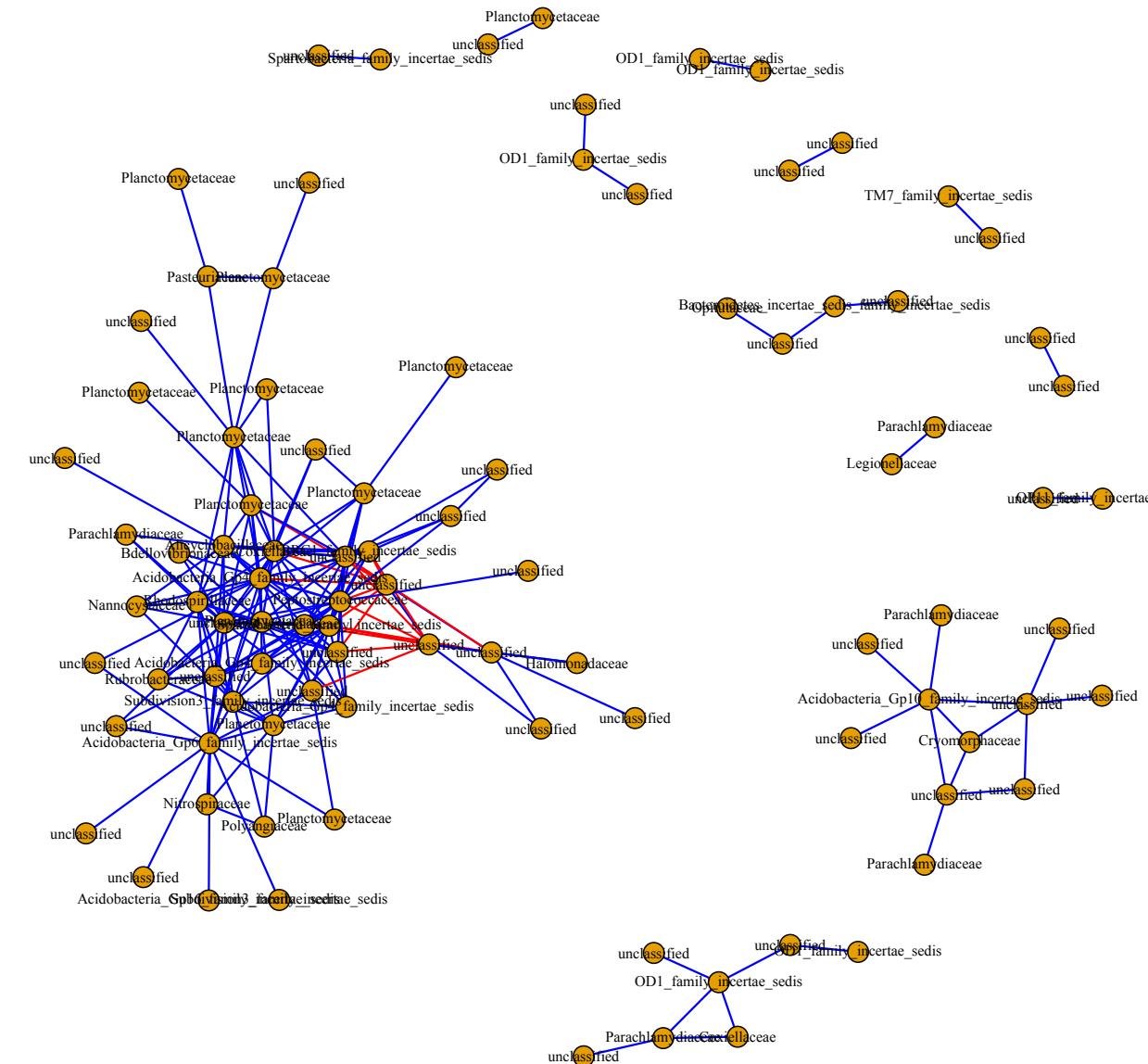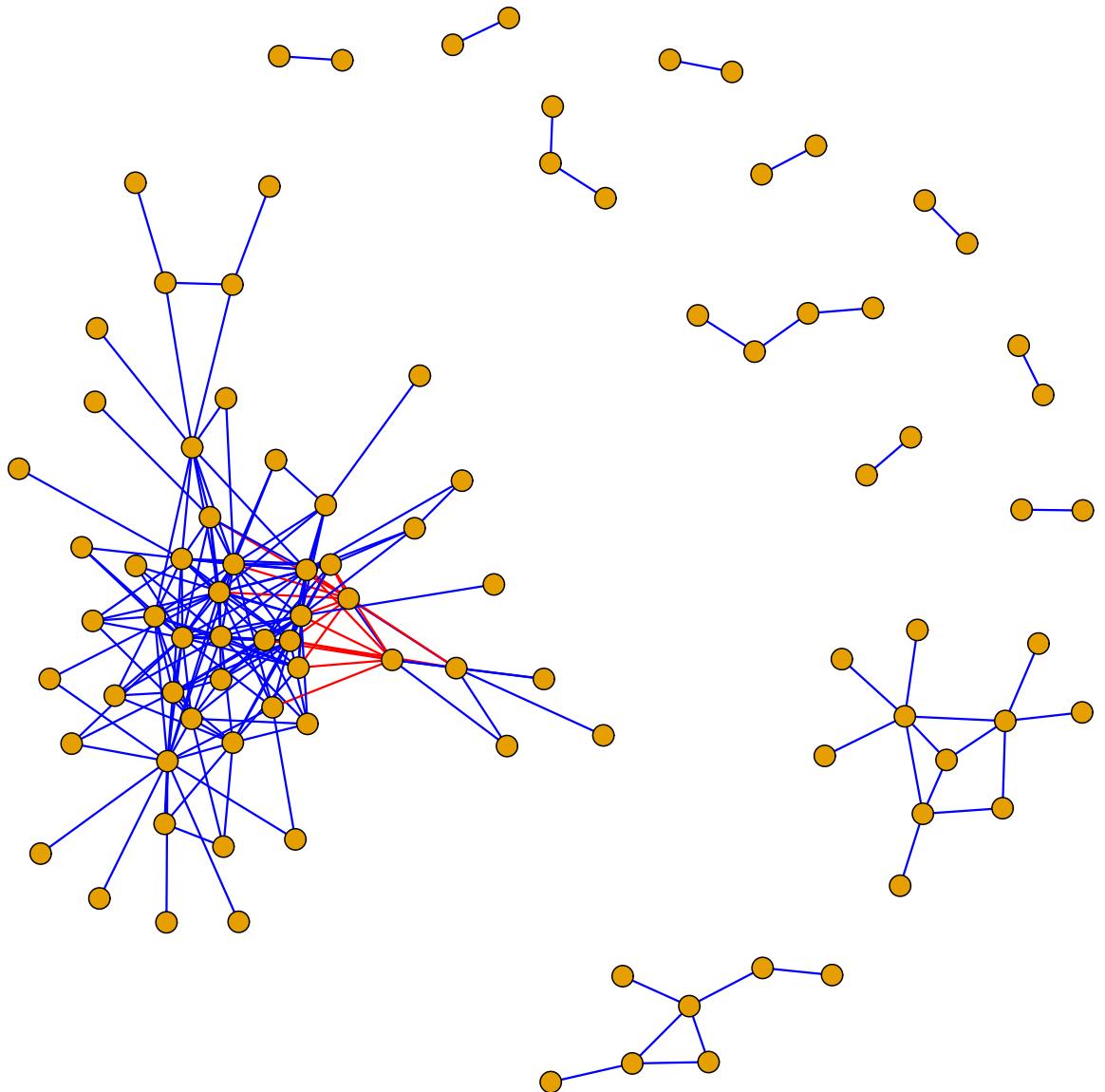
# Summaries of network traits

V(net.grph)

E(net.grph)

vcount(net.grph)

ecount(net.grph)

plot(degree(net.grph))

max(degree(net.grph))

min(degree(net.grph))



igraph | igraph R package

## Get started with R igraph

Install and start using the igraph R package

http://igraph.org/r/