HW05

Emalee Schuler

Objective

Use a Gaussian graphical model to construct a network for the first 10 microbes (OTUs) in the data set "tobacco_clr$data" (the first 10 columns of the data matrix). Visualize the network and interpret your results.

Isolating first 10 microbes in data set

```
load("/Users/eschuler/Downloads/tobacco_clr.Rdata")

tobacco_data <- tobacco_clr$data

tobacco_data10  <- tobacco_data[,1:10]

tobacco_names <- tobacco_clr$otu.names

tobacco_names_10 <- tobacco_names[1:10]
```

Constructing a network for the first 10 microbes (OTUs) in the data set

```
library(ppcor)

Loading required package: MASS

library(igraph)


Attaching package: 'igraph'

The following objects are masked from 'package:stats':

    decompose, spectrum

The following object is masked from 'package:base':

    union

round(pcor(tobacco_data10)$p.value,3)

        643311 4444760 687185 192795 164413 176468 255657 181589 219151
4425571
643311   0.000   0.294  0.012  0.621  0.395  0.002  0.837  0.414  0.661
0.006
4444760  0.294   0.000  0.170  0.205  0.096  0.001  0.772  0.706  0.010
0.324
```

687185	0.012	0.170	0.000	0.202	0.118	0.000	0.646	0.165	0.115
0.036									
192795	0.621	0.205	0.202	0.000	0.762	0.904	0.865	0.668	0.032
0.793									
164413	0.395	0.096	0.118	0.762	0.000	0.043	0.119	0.069	0.240
0.234									
176468	0.002	0.001	0.000	0.904	0.043	0.000	0.783	0.085	0.610
0.002									
255657	0.837	0.772	0.646	0.865	0.119	0.783	0.000	0.005	0.092
0.951									
181589	0.414	0.706	0.165	0.668	0.069	0.085	0.005	0.000	0.261
0.524									
219151	0.661	0.010	0.115	0.032	0.240	0.610	0.092	0.261	0.000
0.099									
4425571	0.006	0.324	0.036	0.793	0.234	0.002	0.951	0.524	0.099
0.000									

```r
network <- (pcor(tobacco_data10)$p.value <= 0.05) *1

diag(network) <- 0

network.plot <- graph_from_adjacency_matrix(network)

plot(network.plot,vertex.label.dist = 0.5, vertex.size= 20, edge.arrow.size =
0.5, edge.arrow.width = 0.5, vertex.label=tobacco_names_10)
```
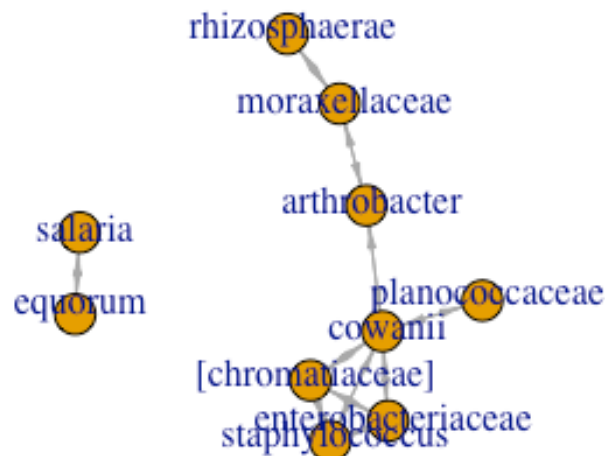
Interpreting Results

```
degree(network.plot)

 643311 4444760  687185  192795  164413  176468  255657  181589  219151
4425571
      6       4       6       2       2      10       2       2       4
6

betweenness(network.plot)

 643311 4444760  687185  192795  164413  176468  255657  181589  219151
4425571
      0      20       0       0       0      30       0       0      12
0

closeness(network.plot)

    643311    4444760     687185     192795     164413     176468     255657
0.07142857 0.08333333 0.07142857 0.04545455 0.06250000 0.10000000 1.00000000
    181589     219151    4425571
1.00000000 0.06250000 0.07142857

clusters(network.plot)
```

```
$membership
 643311 4444760  687185  192795  164413  176468  255657  181589  219151
4425571
      1        1       1       1       1       1       2       2       1
1

$csize
[1] 8 2

$no
[1] 2
```

Degree Centrality:

The degree centrality values indicate the number of direct connections each node has in the network. Node 176468 (cowanii) stands out with a high degree centrality of 10, suggesting it acts as a significant hub with many direct connections. Other nodes have varying degrees, with values ranging from 2 to 6, indicating diverse levels of connectivity. Overall, the network appears to have nodes with distinct roles in terms of direct connections, with node 176468 (cowanii) playing a central role as a key connector.

Betweenness Centrality

The betweenness centrality values represent the extent to which each node serves as a bridge in the network. Node 176468 (cowanii) has a substantial betweenness centrality of 30, indicating its crucial role as a connector along many shortest paths. Nodes 4444760 (arthrobacter) and 219151 (moraxellaceae) also contribute significantly to bridging different parts of the network with betweenness values of 20 and 12, respectively. Other nodes have lower or zero betweenness centrality, suggesting less involvement in facilitating communication between nodes. Overall, these values highlight key nodes that act as important intermediaries in the network.

Closeness Centrality

Closeness centrality measures how close a node is to others in the network. Nodes 255657 (salaria) and 181589 (equorum) have perfect closeness centrality (1.0), indicating they are the closest to all other nodes. Node 176468 (cowanii) also has high closeness centrality (0.1), suggesting it is generally close to others. Nodes 192795 (rhizosphaerae) and 219151 (moraxellaceae) have moderate values, indicating moderate accessibility. Nodes 643311 (chromatiaceae), 4444760 (arthrobacter), 687185(staphylococcus), 164413(planococcaceae), and 4425571 (enterobacteriaceae) have lower closeness centrality, suggesting they are relatively farther away from other nodes on average. These values highlight the varying levels of accessibility and centrality in the network.

Clusters:
The network has two distinct clusters. Nodes 643311(chromatiaceae), 4444760(arthrobacter), 687185 (staphylococcus), 192795 (rhizosphaerae), 164413(planococcaceae), 176468(cowanii), and 4425571(enterobacteriaceae) form

Cluster 1, while nodes 255657(salaria), 181589(equorum), and 219151(moraxellaceae) constitute Cluster 2. Cluster sizes are 8 and 2, respectively. This clustering suggests that nodes within the same cluster are more densely connected to each other than to nodes in other clusters. The network's structure is characterized by the presence of two cohesive groups, providing insights into the underlying organization or functional units within the network.