Microbiome_Data_Analysis

Set Working Directiory :

```r
setwd("/Users/Vartika_Bisht/Individual_Project")
```

Load all source codes :

```r
source("Penalty_Function.R")
source("Incorporate_Groups.R")
source("Required_Libraries.R")
```

```
##
## ---------------------
## Welcome to dendextend version 1.13.4
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## Or contact: <tal.galili@gmail.com>
##
##  To suppress this message use:  suppressPackageStartupMessages(library(dendextend))
## ---------------------


##
## Attaching package: 'dendextend'

## The following object is masked from 'package:stats':
##
##     cutree


## Loading required package: Matrix


## Loaded glmnet 4.0


## network: Classes for Relational Data
## Version 1.16.0 created on 2019-11-30.
## copyright (c) 2005, Carter T. Butts, University of California-Irvine
##                     Mark S. Handcock, University of California -- Los Angeles
##                     David R. Hunter, Penn State University
##                     Martina Morris, University of Washington
##                     Skye Bender-deMoll, University of Washington
##  For citation information, type citation("network").
##  Type help("network-package") to get started.
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa


##
## Attaching package: 'gplots'


## The following object is masked from 'package:stats':
##
##     lowess


## corrplot 0.84 loaded


##
## Attaching package: 'plotly'


## The following object is masked from 'package:ggplot2':
##
##     last_plot


## The following object is masked from 'package:stats':
##
##     filter


## The following object is masked from 'package:graphics':
##
##     layout


## Loading required package: viridisLite


## NOTE: Either Arial Narrow or Roboto Condensed fonts are required to use these themes.


##       Please use hrbrthemes::import_roboto_condensed() to install Roboto Condensed and


##       if Arial Narrow is not on your system, please see https://bit.ly/arialnarrow


##
## Attaching package: 'igraph'


## The following object is masked from 'package:plotly':
##
##     groups


## The following objects are masked from 'package:network':
##
##     %c%, %s%, add.edges, add.vertices, delete.edges, delete.vertices,
##     get.edge.attribute, get.edges, get.vertex.attribute, is.bipartite,
##     is.directed, list.edge.attributes, list.vertex.attributes,
##     set.edge.attribute, set.vertex.attribute


## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union


## Loading required package: usethis


## Loading required package: grid


## =======================================
## ComplexHeatmap version 2.2.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite:
## Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
##   genomic data. Bioinformatics 2016.
## =======================================


##
## Attaching package: 'ComplexHeatmap'


## The following object is masked from 'package:plotly':
##
##     add_heatmap


## The following object is masked from 'package:network':
##
##     %v%


## Loading required package: lattice
```

```r
source("Borrowed_Functions.R")
source("MicrobiomeAnalyst.R")
source("best_epsilon_DBSCAN.R")
```

Load Data Set :

```r
# Load Dataset 2
Data_Set_1 <- read.xlsx("Data_all_micobiome.xlsx",1)
# Choose Microbiome Data
df_data1 <- Data_Set_1[3:49]
data1 <- data.matrix(Data_Set_1[3:49])
# Create labels for prediction ( 2nd column )
label_dat <- label_creator(Data_Set_1)
```

Input Dataset :

```r
head(Data_Set_1)
```

```
##    Sample Class c__Gammaproteobacteria f__Enterobacteriaceae
## 1   HC10    HC              0.026749019           0.024576879
## 2   HC11    HC              0.020946539           0.019442081
## 3    HC1    HC              0.009181709           0.007950845
## 4    HC2    HC              0.722407840           0.713623395
## 5    HC3    HC              0.021028725           0.019669925
## 6    HC4    HC              0.022512010           0.014574446
##   o__Enterobacteriales f__Prevotellaceae f__Paraprevotellaceae
## 1          0.024576879        0.005810750            0.000077200
## 2          0.019442081        0.194510412            0.000188030
## 3          0.007950845        0.002392202            0.006313053
## 4          0.713623395        0.003941737            0.000000000
## 5          0.019669925        0.001692905            0.000069100
## 6          0.014574446        0.243289740            0.006035345
##   f__Myxococcales.0319.6G20 s__Myxococcales.sp c__Lentisphaeria
## 1               0.000000000        0.000000000      0.000000000
## 2               0.000000000        0.000000000      0.000053700
## 3               0.000099300        0.000099300      0.000000000
## 4               0.000488025        0.000488025      0.000000000
## 5               0.000000000        0.000000000      0.000000000
## 6               0.000000000        0.000000000      0.000524512
##   o__Victivallales p__Lentisphaerae g__Streptococcus c__Coriobacteriia
## 1      0.000000000       0.000000000      0.000419031        0.000319753
## 2      0.000053700       0.000053700      0.000716475        0.000071700
## 3      0.000000000       0.000000000      0.000406972        0.000000000
## 4      0.000000000       0.000000000      0.002008409        0.000000000
## 5      0.000000000       0.000000000      0.000218811        0.000149713
## 6      0.000524512       0.000524512      0.002734435        0.000776317
##   o__Coriobacteriales f__Coriobacteriaceae g__Paraprevotella
## 1         0.000319753          0.000319753       0.000077200
## 2         0.000071700          0.000071700       0.000134330
## 3         0.000000000          0.000000000       0.006283253
## 4         0.000000000          0.000000000       0.000000000
## 5         0.000149713          0.000149713       0.000000000
## 6         0.000776317          0.000776317       0.001076998
##   g__Paraprevotella.1  g__Blautia g__Erysipelotrichaceae.cc_115
## 1         0.000077200 0.008721539                             0
## 2         0.000134330 0.009367283                             0
## 3         0.006283253 0.003275630                             0
## 4         0.000000000 0.001952098                             0
## 5         0.000000000 0.001117087                             0
## 6         0.001076998 0.007070425                             0
##   s__Erysipelotrichaceae.cc_115.sp s__symbiosum. s__Desulfovibrionaceae.sp
## 1                                0   0.000000000               0.000165389
## 2                                0   0.000000000               0.000000000
## 3                                0   0.000000000               0.000000000
## 4                                0   0.000000000               0.000000000
## 5                                0   0.000000000               0.000000000
## 6                                0   0.000524512               0.000014000
##   g__Desulfovibrionaceae.unidentified f__Brevibacteriaceae s__Brevibacterium.sp
## 1                         0.000165389              6.62e-05             6.62e-05
## 2                         0.000000000              0.00e+00             0.00e+00
## 3                         0.000000000              2.98e-05             2.98e-05
## 4                         0.000000000              0.00e+00             0.00e+00
```

```
## 5                          0.000000000          0.00e+00           0.00e+00
## 6                          0.000014000          1.40e-05           1.40e-05
##   s__ramosum. s__Sphingomonas.sp s__Tepidimonas.sp f__Microbacteriaceae
## 1 0.005336568      0.000308727      0.000055100          0.000066200
## 2 0.003591098      0.000000000      0.000000000          0.000205973
## 3 0.000774240      0.000000000      0.000069500          0.000138966
## 4 0.000000000      0.000000000      0.001426534          0.000337863
## 5 0.000092100      0.000184262      0.000518236          0.001658356
## 6 0.000000000      0.000000000      0.000000000          0.000000000
##   g__Tepidimonas s__hathewayi. g__Staphylococcus s__adolescentis.   g__Rothia
## 1    0.000055100   0.000121286       0.001157726        0.000165389 0.000044100
## 2    0.000000000   0.000125375       0.000385080        0.000277616 0.000107464
## 3    0.000069500   0.000000000       0.002193679        0.000000000 0.000000000
## 4    0.001426534   0.000000000       0.012256926        0.000000000 0.000000000
## 5    0.000518236   0.000000000       0.000978891        0.000000000 0.000000000
## 6    0.000000000   0.002769424       0.000475558        0.002230925 0.000049000
##   s__citroniae. s__mucilaginosa. s__distasonis. g__Actinomyces s__Actinomyces
## 1  0.001686973      0.000044100    0.000000000    0.000000000      0.000000000
## 2  0.001011955      0.000107464    0.000779116    0.000000000      0.000000000
## 3  0.006164139      0.000000000    0.000000000    0.000119114      0.000119114
## 4  0.000112621      0.000000000    0.000000000    0.000187702      0.000187702
## 5  0.010353207      0.000000000    0.006667972    0.000046100      0.000046100
## 6  0.000594447      0.000049000    0.000000000    0.000111896      0.000111896
##   s__stutzeri. o__Alteromonadales g__Shewanella f__Shewanellaceae  c__Bacilli
## 1 0.000000000                  0             0                 0 0.001885557
## 2 0.000000000                  0             0                 0 0.001289555
## 3 0.000000000                  0             0                 0 0.003504017
## 4 0.000375404                  0             0                 0 0.023687962
## 5 0.000138196                  0             0                 0 0.002268725
## 6 0.000069900                  0             0                 0 0.006783623
##   s__Roseburia.sp s__Parvimonas.sp g__Parvimonas s__fragilis.
## 1    0.003837036      0.000000000   0.000000000  0.000264623
## 2    0.003313482      0.000000000   0.000000000  0.000000000
## 3    0.000000000      0.000049600   0.000049600  0.392966331
## 4    0.000000000      0.000000000   0.000000000  0.000563105
## 5    0.000000000      0.000000000   0.000000000  0.000391556
## 6    0.002944262      0.000804252   0.000804252  0.017022169
```

Input Data for Module 1 (Features) :

```r
head(as.data.frame(data1))
```

```
##   c__Gammaproteobacteria f__Enterobacteriaceae o__Enterobacteriales
## 1           0.026749019          0.024576879          0.024576879
## 2           0.020946539          0.019442081          0.019442081
## 3           0.009181709          0.007950845          0.007950845
## 4           0.722407840          0.713623395          0.713623395
## 5           0.021028725          0.019669925          0.019669925
## 6           0.022512010          0.014574446          0.014574446
##   f__Prevotellaceae f__Paraprevotellaceae f__Myxococcales.0319.6G20
## 1      0.005810750          0.000077200          0.000000000
## 2      0.194510412          0.000188030          0.000000000
## 3      0.002392202          0.006313053          0.000099300
```

```
## 4       0.003941737         0.000000000              0.000488025
## 5       0.001692905         0.000069100              0.000000000
## 6       0.243289740         0.006035345              0.000000000
##   s__Myxococcales.sp c__Lentisphaeria o__Victivallales p__Lentisphaerae
## 1        0.000000000       0.000000000       0.000000000       0.000000000
## 2        0.000000000       0.000053700       0.000053700       0.000053700
## 3        0.000099300       0.000000000       0.000000000       0.000000000
## 4        0.000488025       0.000000000       0.000000000       0.000000000
## 5        0.000000000       0.000000000       0.000000000       0.000000000
## 6        0.000000000       0.000524512       0.000524512       0.000524512
##   g__Streptococcus c__Coriobacteriia o__Coriobacteriales f__Coriobacteriaceae
## 1      0.000419031       0.000319753         0.000319753          0.000319753
## 2      0.000716475       0.000071700         0.000071700          0.000071700
## 3      0.000406972       0.000000000         0.000000000          0.000000000
## 4      0.002008409       0.000000000         0.000000000          0.000000000
## 5      0.000218811       0.000149713         0.000149713          0.000149713
## 6      0.002734435       0.000776317         0.000776317          0.000776317
##   g__Paraprevotella g__Paraprevotella.1  g__Blautia
## 1       0.000077200         0.000077200 0.008721539
## 2       0.000134330         0.000134330 0.009367283
## 3       0.006283253         0.006283253 0.003275630
## 4       0.000000000         0.000000000 0.001952098
## 5       0.000000000         0.000000000 0.001117087
## 6       0.001076998         0.001076998 0.007070425
##   g__Erysipelotrichaceae.cc_115 s__Erysipelotrichaceae.cc_115.sp s__symbiosum.
## 1                             0                               0   0.000000000
## 2                             0                               0   0.000000000
## 3                             0                               0   0.000000000
## 4                             0                               0   0.000000000
## 5                             0                               0   0.000000000
## 6                             0                               0   0.000524512
##   s__Desulfovibrionaceae.sp g__Desulfovibrionaceae.unidentified
## 1               0.000165389                         0.000165389
## 2               0.000000000                         0.000000000
## 3               0.000000000                         0.000000000
## 4               0.000000000                         0.000000000
## 5               0.000000000                         0.000000000
## 6               0.000014000                         0.000014000
##   f__Brevibacteriaceae s__Brevibacterium.sp s__ramosum. s__Sphingomonas.sp
## 1             6.62e-05             6.62e-05 0.005336568         0.000308727
## 2             0.00e+00             0.00e+00 0.003591098         0.000000000
## 3             2.98e-05             2.98e-05 0.000774240         0.000000000
## 4             0.00e+00             0.00e+00 0.000000000         0.000000000
## 5             0.00e+00             0.00e+00 0.000092100         0.000184262
## 6             1.40e-05             1.40e-05 0.000000000         0.000000000
##   s__Tepidimonas.sp f__Microbacteriaceae g__Tepidimonas s__hathewayi.
## 1       0.000055100          0.000066200    0.000055100   0.000121286
## 2       0.000000000          0.000205973    0.000000000   0.000125375
## 3       0.000069500          0.000138966    0.000069500   0.000000000
## 4       0.001426534          0.000337863    0.001426534   0.000000000
## 5       0.000518236          0.001658356    0.000518236   0.000000000
## 6       0.000000000          0.000000000    0.000000000   0.002769424
##   g__Staphylococcus s__adolescentis.   g__Rothia s__citroniae. s__mucilaginosa.
## 1       0.001157726      0.000165389 0.000044100   0.001686973      0.000044100
```

```
## 2        0.000385080        0.000277616 0.000107464     0.001011955        0.000107464
## 3        0.002193679        0.000000000 0.000000000     0.006164139        0.000000000
## 4        0.012256926        0.000000000 0.000000000     0.000112621        0.000000000
## 5        0.000978891        0.000000000 0.000000000     0.010353207        0.000000000
## 6        0.000475558        0.002230925 0.000049000     0.000594447        0.000049000
##   s__distasonis. g__Actinomyces s__Actinomyces s__stutzeri. o__Alteromonadales
## 1    0.000000000    0.000000000     0.000000000 0.000000000                    0
## 2    0.000779116    0.000000000     0.000000000 0.000000000                    0
## 3    0.000000000    0.000119114     0.000119114 0.000000000                    0
## 4    0.000000000    0.000187702     0.000187702 0.000375404                    0
## 5    0.006667972    0.000046100     0.000046100 0.000138196                    0
## 6    0.000000000    0.000111896     0.000111896 0.000069900                    0
##   g__Shewanella f__Shewanellaceae  c__Bacilli s__Roseburia.sp s__Parvimonas.sp
## 1             0                 0 0.001885557     0.003837036     0.000000000
## 2             0                 0 0.001289555     0.003313482     0.000000000
## 3             0                 0 0.003504017     0.000000000     0.000049600
## 4             0                 0 0.023687962     0.000000000     0.000000000
## 5             0                 0 0.002268725     0.000000000     0.000000000
## 6             0                 0 0.006783623     0.002944262     0.000804252
##   g__Parvimonas s__fragilis.
## 1   0.000000000  0.000264623
## 2   0.000000000  0.000000000
## 3   0.000049600  0.392966331
## 4   0.000000000  0.000563105
## 5   0.000000000  0.000391556
## 6   0.000804252  0.017022169
```

Input Data for Module 1 (Output Variable: Given Variable) :

```
#Labels
Data_Set_1$Class
```

```
##  [1] HC     HC     HC     HC     HC     HC     HC     HC     HC     HC
## [11] PSC-UC PSC-UC PSC-UC PSC-UC PSC-UC PSC-UC PSC-UC PSC-UC PSC-UC PSC-UC
## [21] UC     UC     UC     UC     UC     UC     UC     UC     UC     UC
## Levels: HC PSC-UC UC
```

Input Data for Module 1 (Output Variable: Desired Labels) :

```
#Desired Labels
label_dat
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3
```

Module 1:

```
## Input : Numeric Labels(label_dat) and Microbiome Abundance Data(data1)
source("Module_1.R")
```

```
##   |                                                                       |
## [1] "ANFIS DONE!"
## [1] "New labels have been assigned!"
## [1] "Rule based matrix is saved!"
## [1] "Scaled Ruled Based Matrix saved"
```

Module 1 Output (Rule Based Matrix):

```
head(as.data.frame(rules_int))
```

```
##   c__Gammaproteobacteria f__Enterobacteriaceae o__Enterobacteriales
## 1                      7                    14                   21
## 2                      7                    14                   21
## 3                      1                     8                   15
## 4                      1                     8                   15
## 5                      1                     8                   15
## 6                      4                    11                   18
##   f__Prevotellaceae f__Paraprevotellaceae f__Myxococcales.0319.6G20
## 1                22                    29                        42
## 2                24                    29                        38
## 3                22                    29                        37
## 4                22                    29                        36
## 5                22                    29                        36
## 6                28                    30                        36
##   s__Myxococcales.sp c__Lentisphaeria o__Victivallales p__Lentisphaerae
## 1                 49               50               57               64
## 2                 45               50               57               64
## 3                 44               50               57               64
## 4                 43               50               57               64
## 5                 43               50               57               64
## 6                 43               56               63               70
##   g__Streptococcus c__Coriobacteriia o__Coriobacteriales f__Coriobacteriaceae
## 1               72                78                   85                   92
## 2               71                78                   85                   92
## 3               71                78                   85                   92
## 4               72                78                   85                   92
## 5               71                78                   85                   92
## 6               72                78                   85                   92
##   g__Paraprevotella g__Paraprevotella.1 g__Blautia
## 1                99                 106        113
## 2                99                 106        113
## 3               105                 112        113
## 4                99                 106        113
## 5                99                 106        116
## 6                99                 106        113
##   g__Erysipelotrichaceae.cc_115 s__Erysipelotrichaceae.cc_115.sp s__symbiosum.
## 1                           120                              127           134
## 2                           120                              127           134
## 3                           120                              127           134
## 4                           120                              127           134
## 5                           120                              127           134
## 6                           120                              127           134
##   s__Desulfovibrionaceae.sp g__Desulfovibrionaceae.unidentified
## 1                       141                                 148
## 2                       141                                 148
## 3                       141                                 148
## 4                       141                                 148
```

```
## 5                                  141                                   148
## 6                                  141                                   148
##    f__Brevibacteriaceae s__Brevibacterium.sp s__ramosum. s__Sphingomonas.sp
## 1                   155                    162         169              176
## 2                   155                    162         169              176
## 3                   155                    162         170              176
## 4                   155                    162         169              176
## 5                   155                    162         171              176
## 6                   155                    162         169              176
##    s__Tepidimonas.sp f__Microbacteriaceae g__Tepidimonas s__hathewayi.
## 1               189                  190            203           204
## 2               183                  193            197           204
## 3               183                  190            197           204
## 4               185                  190            199           207
## 5               183                  192            197           207
## 6               183                  191            197           204
##    g__Staphylococcus s__adolescentis. g__Rothia s__citroniae. s__mucilaginosa.
## 1               217              218       225           232              239
## 2               211              218       225           232              239
## 3               212              218       225           235              239
## 4               214              220       225           238              239
## 5               212              219       225           235              239
## 6               213              218       230           232              244
##    s__distasonis. g__Actinomyces s__Actinomyces s__stutzeri. o__Alteromonadales
## 1            246            255            262          267               274
## 2            246            253            260          267               275
## 3            246            254            261          267               274
## 4            246            256            263          267               274
## 5            246            253            260          267               274
## 6            246            253            260          268               276
##    g__Shewanella f__Shewanellaceae c__Bacilli s__Roseburia.sp s__Parvimonas.sp
## 1           281              288        300            302             309
## 2           281              288        296            302             309
## 3           281              288        296            302             309
## 4           281              288        297            302             315
## 5           281              288        295            302             309
## 6           281              288        297            302             309
##    g__Parvimonas s__fragilis.
## 1           316         323
## 2           316         323
## 3           316         329
## 4           322         323
## 5           316         328
## 6           316         323
```

Module 1 Output (Labels):

```r
as.data.frame(label_dat)[,1]
```

```
##  [1] -1.204159  1.204159 -1.204159  0.000000  1.204159  1.204159 -1.204159
##  [8]  1.204159 -1.204159 -1.204159 -1.204159 -1.204159 -1.204159 -1.204159
## [15] -1.204159  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
## [22]  0.000000  0.000000  0.000000  1.204159  1.204159  1.204159  1.204159
## [29]  1.204159  1.204159
```

Module 2:

```
## Input : Scaled Rule Based Matrix (scaled_rules_int)
source("Module_2.R")
```

```
## [1] "Epsilon value used : 3.5"
## [1] "13 cluster(s) found!"
## [1] "Clustering Done!"
## [1] "Feature's cluster number saved"
## [1] "Grouping Highly Colinear Features Together :-"
## [1] "Clubbing features in a group together"
## [1] "Features Clubbed and incorporated in a new Data Frame!"
## [1] "Rule Based matrix with Colinearity Handled saved"
## [1] "PCA Loadings used to combine groups saved"
```

```
## Output : Rule Based matrix with Colinearity Handled (new_data1) and PCA Loadings used to combine gro
```

Module 2 Output (Clusters):

```
groups_we_need
```

```
## [[1]]
## [1] "c__Gammaproteobacteria" "f__Enterobacteriaceae"  "o__Enterobacteriales"
##
## [[2]]
## [1] "f__Myxococcales.0319.6G20" "s__Myxococcales.sp"
##
## [[3]]
## [1] "c__Lentisphaeria" "o__Victivallales" "p__Lentisphaerae"
##
## [[4]]
## [1] "g__Streptococcus"     "f__Brevibacteriaceae" "s__Brevibacterium.sp"
## [4] "c__Bacilli"
##
## [[5]]
## [1] "c__Coriobacteriia"    "o__Coriobacteriales"  "f__Coriobacteriaceae"
##
## [[6]]
## [1] "g__Paraprevotella"    "g__Paraprevotella.1"
##
## [[7]]
## [1] "g__Erysipelotrichaceae.cc_115"    "s__Erysipelotrichaceae.cc_115.sp"
##
## [[8]]
## [1] "s__Desulfovibrionaceae.sp"           "g__Desulfovibrionaceae.unidentified"
##
## [[9]]
## [1] "s__Sphingomonas.sp" "g__Actinomyces"     "s__Actinomyces"
## [4] "s__Roseburia.sp"
##
## [[10]]
## [1] "s__Tepidimonas.sp" "g__Tepidimonas"
```

```
## 
## [[11]]
## [1] "g__Rothia"         "s__mucilaginosa."
## 
## [[12]]
## [1] "g__Shewanella"     "f__Shewanellaceae"
## 
## [[13]]
## [1] "s__Parvimonas.sp" "g__Parvimonas"
```

Module 2 Output (PCA Loadings):

```
head(as.data.frame(t(PCA_loadings)))
```

```
##                            PCA Loadings
## c__Gammaproteobacteria    -0.573412643272275
## f__Enterobacteriaceae     -0.579309045560097
## o__Enterobacteriales      -0.579309045560097
## f__Myxococcales.0319.6G20 -0.661430751941418
## s__Myxococcales.sp        -0.750006240231513
## c__Lentisphaeria           0.577350269189626
```

Module 2 Output (New Data Frame):

```
head(as.data.frame(new_data1))
```

```
##   f__Prevotellaceae f__Paraprevotellaceae g__Blautia s__symbiosum. s__ramosum.
## 1                22                    29        113           134         169
## 2                24                    29        113           134         169
## 3                22                    29        113           134         170
## 4                22                    29        113           134         169
## 5                22                    29        116           134         171
## 6                28                    30        113           134         169
##   f__Microbacteriaceae s__hathewayi. g__Staphylococcus s__adolescentis.
## 1                  190           204               217              218
## 2                  193           204               211              218
## 3                  190           204               212              218
## 4                  190           207               214              220
## 5                  192           207               212              219
## 6                  191           204               213              218
##   s__citroniae. s__distasonis. s__stutzeri. o__Alteromonadales s__fragilis.
## 1           232            246          267                274          323
## 2           232            246          267                275          323
## 3           235            246          267                274          329
## 4           238            246          267                274          323
## 5           235            246          267                274          328
## 6           232            246          268                276          323
##   c__Gammaproteobacteria~f__Enterobacteriaceae~o__Enterobacteriales
## 1                                                         -24.28971
## 2                                                         -24.28971
## 3                                                         -13.89752
## 4                                                         -13.89752
```

```
## 5                                                      -13.89752
## 6                                                      -19.09361
##   f__Myxococcales.0319.6G20~s__Myxococcales.sp
## 1                              -64.53040
## 2                              -58.88465
## 3                              -57.47321
## 4                              -56.06178
## 5                              -56.06178
## 6                              -56.06178
##   c__Lentisphaeria~o__Victivallales~p_Lentisphaerae
## 1                               98.7269
## 2                               98.7269
## 3                               98.7269
## 4                               98.7269
## 5                               98.7269
## 6                              109.1192
##   g__Streptococcus~f__Brevibacteriaceae~s__Brevibacterium.sp~c__Bacilli
## 1                                            -361.5970
## 2                                            -358.5616
## 3                                            -358.5616
## 4                                            -359.6767
## 5                                            -357.9215
## 6                                            -359.6767
##   c__Coriobacteriia~o__Coriobacteriales~f__Coriobacteriaceae
## 1                                    147.2243
## 2                                    147.2243
## 3                                    147.2243
## 4                                    147.2243
## 5                                    147.2243
## 6                                    147.2243
##   g__Paraprevotella~g__Paraprevotella.1
## 1                         144.9569
## 2                         144.9569
## 3                         153.4422
## 4                         144.9569
## 5                         144.9569
## 6                         144.9569
##   g__Erysipelotrichaceae.cc_115~s__Erysipelotrichaceae.cc_115.sp
## 1                                            174.6554
## 2                                            174.6554
## 3                                            174.6554
## 4                                            174.6554
## 5                                            174.6554
## 6                                            174.6554
##   s__Desulfovibrionaceae.sp~g__Desulfovibrionaceae.unidentified
## 1                                      204.3539
## 2                                      204.3539
## 3                                      204.3539
## 4                                      204.3539
## 5                                      204.3539
## 6                                      204.3539
##   s__Sphingomonas.sp~g__Actinomyces~s__Actinomyces~s__Roseburia.sp
## 1                                       497.1475
## 2                                       494.9695
```

```
## 3                                                496.0585
## 4                                                498.2365
## 5                                                494.9695
## 6                                                494.9695
##   s__Tepidimonas.sp~g__Tepidimonas g__Rothia~s__mucilaginosa.
## 1                       -277.1859                 328.0975
## 2                       -268.7006                 328.0975
## 3                       -268.7006                 328.0975
## 4                       -271.5290                 328.0975
## 5                       -268.7006                 328.0975
## 6                       -268.7006                 335.1686
##   g__Shewanella~f__Shewanellaceae s__Parvimonas.sp~g__Parvimonas
## 1                       402.3438                 441.9417
## 2                       402.3438                 441.9417
## 3                       402.3438                 441.9417
## 4                       402.3438                 450.4270
## 5                       402.3438                 441.9417
## 6                       402.3438                 441.9417
```

Module 3:

```
## Input : Rule Based matrix with Colinearity Handled (new_data1) and PCA Loadings used to combine grou
source("Module_3.R")
```

```
## [1] "Feature Parameters computed and saved"
```

```
## Output : Feature Parameters (feature_parameters)
```

Module 3 Output (Adaptive LASSO Results):

```
head(as.data.frame(feature_parameters))
```

```
##                          ADres
## f__Prevotellaceae      5.0036851
## f__Paraprevotellaceae -0.1920410
## g__Blautia            -0.1920407
## s__symbiosum.         -0.1922304
## s__ramosum.           -0.1924006
## f__Microbacteriaceae  -0.1875031
```

Module 4 (TSEA - Specify Disease):

```
#Diseases to look for in TSEA
disease <- c("Colorectal","Crohn")
```

Module 4 (TSEA Type of feature):

```
TSEA_feature <- "Microbes"
```

Module 4 (TSEA - OTU): If Features are OTU and need to be changed into appropriate Microbes for TSEA

```
if(TSEA_feature == "OTU"){
  #List of Microboes from selected features (OTU)
#OTU to Microbes
OTU_file <- read.table("final.csv", header = 1)
OTU_index <- which(OTU_file$OTU %in% rownames(feature_parameters))
selected_OTU <- OTU_file[OTU_index,]
feature_inorder <- selected_OTU$OTU
write.csv(selected_OTU,"OTU Microbes Selected Table.csv")

#Valid Microbe Names
OTU_network <- c()
taxa <- strsplit(as.character(selected_OTU$Taxonomy),";")
for(i in 1:length(taxa)){
  taxa_name <- taxa[[i]][length(taxa[[i]])]
  OTU_network <- c(OTU_network, substr( taxa_name , 1 , nchar(taxa_name)-5) )
}

Name_Change <- as.data.frame(OTU_network)
rownames(Name_Change) <- feature_inorder
write.csv(Name_Change,"Features to Microbes for TSEA.csv")


}
```

Module 4 (TSEA - Microbes of Different Taxa Level): If Features are Microbes of Different Taxa Level and need to be changed into appropriate Microbes for TSEA

```
if(TSEA_feature == "Microbes"){
  #List of Microboes from selected features (Microbes)
  Microbes_name <- substring(colnames(rules_int),4)
  OTU_network <- c()
  for(i in Microbes_name){
    n <- strsplit(i,split='.', fixed=TRUE)[[1]]
    if((length(n)>1)&&(n[2] == "unidentified")){
      OTU_network <- c(OTU_network,sprintf("%s.%s",n[1],n[2]))
    }else{
      OTU_network <- c(OTU_network,n[1])
    }
  }
  feature_inorder <- colnames(rules_int)

  Name_Change <- as.data.frame(OTU_network)
  rownames(Name_Change) <- feature_inorder
  write.csv(Name_Change,"Features to Microbes for TSEA.csv")
}
```

Module 4 (TSEA - The names used for TSEA with the feature associated):

```
head(as.data.frame(Name_Change))
```

```
##                             OTU_network
## c__Gammaproteobacteria    Gammaproteobacteria
## f__Enterobacteriaceae      Enterobacteriaceae
```

```
## o__Enterobacteriales        Enterobacteriales
## f__Prevotellaceae              Prevotellaceae
## f__Paraprevotellaceae       Paraprevotellaceae
## f__Myxococcales.0319.6G20      Myxococcales
```

Module 4 (TSEA - The names used for TSEA with the feature associated):

```
Microbes <- unique(OTU_network)
Microbes
```

```
##  [1] "Gammaproteobacteria"              "Enterobacteriaceae"
##  [3] "Enterobacteriales"                "Prevotellaceae"
##  [5] "Paraprevotellaceae"               "Myxococcales"
##  [7] "Lentisphaeria"                    "Victivallales"
##  [9] "Lentisphaerae"                    "Streptococcus"
## [11] "Coriobacteriia"                   "Coriobacteriales"
## [13] "Coriobacteriaceae"                "Paraprevotella"
## [15] "Blautia"                          "Erysipelotrichaceae"
## [17] "symbiosum"                        "Desulfovibrionaceae"
## [19] "Desulfovibrionaceae.unidentified" "Brevibacteriaceae"
## [21] "Brevibacterium"                   "ramosum"
## [23] "Sphingomonas"                     "Tepidimonas"
## [25] "Microbacteriaceae"                "hathewayi"
## [27] "Staphylococcus"                   "adolescentis"
## [29] "Rothia"                           "citroniae"
## [31] "mucilaginosa"                     "distasonis"
## [33] "Actinomyces"                      "stutzeri"
## [35] "Alteromonadales"                  "Shewanella"
## [37] "Shewanellaceae"                   "Bacilli"
## [39] "Roseburia"                        "Parvimonas"
## [41] "fragilis"
```

Module 4 (TSEA - The names used for TSEA with the feature associated):

```
## Input : List of Microbes
source("Module_4(TSEA Network).R")
```

```
## [1] "----Microbiome Analyst----"
## [1] "Init MicrobiomeAnalyst!"
## [1] "Loaded files from MetaboAnalyst web-server."
## [1] "Loaded files from MetaboAnalyst web-server."
## [1] "No matches were found in the selected taxon set library!"
## [1] "Mixed-level taxa set was selected!"
## [1] "Mix Taxa TSEA Results Calculated"
## [1] "Mix Taxa TSEA Disease Specific Results Calculated"
## [1] "Calculating Adjacency Matrix for Network"
```

```
## Output : Network and Network Legends with Node size (Legends)
```

Module 4 (TSEA Network):

```r
plot(g, layout=layout_in_circle, vertex.size=vertex_wt,edge.width = E(g)$weight)
```
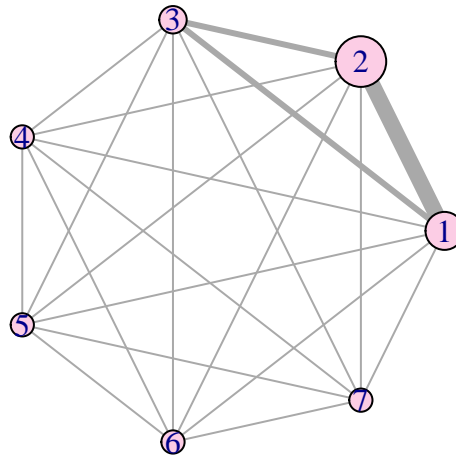

Module 4 (TSEA Network Legends):

```r
head(as.data.frame(Network_Info))
```

```
##   Node        Microbe Names Node Size
## 1    1     Enterobacteriaceae         8
## 2    2 Gammaproteobacteria        14
## 3    3     Enterobacteriales         3
## 4    4        Prevotellaceae         1
## 5    5     Paraprevotellaceae         1
## 6    6         Myxococcales         1
```

Module 4 (Infusing Data Driven Information): TSEA Network and Adaptive LASSO Results

```r
## Input : TSEA Network and Adaptive LASSO Results
Cluster_Parameters <- c()
Cluster_OTU_name <- c()
for (i in Network_Info[,"Microbe Names"]) {
  index <- which(OTU_network %in% i)
  OTU <- as.character(feature_inorder[index])
  Cluster_OTU_name <- append(Cluster_OTU_name,list(OTU))
  if(length(OTU) > 1){
    OTUs_val <- c()
```

```r
  for (j in OTU) {
    OTUs_val <- c( OTUs_val , abs(feature_parameters[j,]) )
  }
  CP <- (sum(OTUs_val)/length(OTUs_val))[1]
} else {
  CP <- abs(feature_parameters[OTU,])
}
Cluster_Parameters <- c( Cluster_Parameters , CP )
}
Data_Bio_Driven <- cbind(Network_Info,Cluster_Parameters)
rownames(Data_Bio_Driven) <- NULL
write.csv(Data_Bio_Driven,"Biological Network with Data Driven Results fused.csv")
## Output : Data Driven Cluster Parameters added
```

Module 4 (Infusing Data Driven Information): TSEA Network and Adaptive LASSO Results

```r
head(as.data.frame(Data_Bio_Driven))
```

```
##   Node       Microbe Names Node Size Cluster_Parameters
## 1    1    Enterobacteriaceae         8  0.111358080108876
## 2    2 Gammaproteobacteria        14  0.110224640119713
## 3    3    Enterobacteriales         3  0.111358080108876
## 4    4       Prevotellaceae         1   5.00368509982012
## 5    5   Paraprevotellaceae         1  0.192040963509009
## 6    6          Myxococcales         1  0.135627340111481
```

Module 4 (Infusing Data Driven Information): Module 2 Clusters and TSEA Network

```r
## Input : Module 2 Clusters and TSEA Network
source("Module_4(Data Driven Network).R")
```

```
## [1] "Calculating Edges and Nodes to be added for the Data Driven Network"
```

```
## [1] "Final Fused Network Saved!"
## [1] "Final Fused Network Cluster Information Saved!"
```
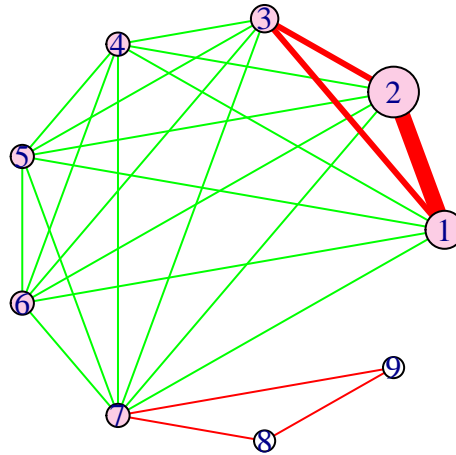
```r
## Output : Network with Data Driven Clusters
```

Module 4 (Infusing Data Driven Information): Module 2 Clusters and TSEA Network

```r
#Green Edges <- TSEA
#Red Edges <- Data Driven Cluslters
#Pink nodes <- TSEA
#White nodes <- Data Driven Cluslters
plot(gh, layout=layout_in_circle, vertex.size=vertex_wt_gh,edge.width = E(gh)$weight,edge.color=col_edge
```

Module 4 (Infusing Data Driven Information): Module 2 Clusters and TSEA Network

```r
head(as.data.frame(Data_Bio_Driven_with_clusters))
```

```
##   Node       Microbe Names Node Size Cluster_Parameters
## 1    1   Enterobacteriaceae         8  0.111358080108876
## 2    2 Gammaproteobacteria        14  0.110224640119713
## 3    3    Enterobacteriales         3  0.111358080108876
## 4    4        Prevotellaceae         1   5.00368509982012
## 5    5   Paraprevotellaceae         1  0.192040963509009
## 6    6          Myxococcales         1  0.135627340111481
##                         Features in the Cluster
## 1                        f__Enterobacteriaceae
## 2                       c__Gammaproteobacteria
## 3                         o__Enterobacteriales
## 4                           f__Prevotellaceae
## 5                        f__Paraprevotellaceae
## 6 f__Myxococcales.0319.6G20, s__Myxococcales.sp
```