Package ‘netmes’

July 26, 2013

Version 1.0

Title Exploratory Analysis of Gene Network Infoerence Algorithms R package

Author Gokmen Altay, Frank Emmert-Streib

Maintainer Gokmen Altay <altayscience@gmail.com>

Date 2010-09-22

Description This package implements various algorithms for an explanatory analysis of network based evaluation measures to assess network inference algorithms systematically.

License GPL (>= 3)

Depends minet, infotheo, igraph, Hmisc, gregmisc

URL http://www.bio-complexity.com/

Repository R-Forge

Date/Publication 2010-09-06 15:04:40

R topics documented:

`evalnetworks` *Computing all the statistics from data of sythetic network*

Description

This function takes the adjacency matrix, the data sets folder path and number of mutual information steps for the therhold mutual information. It returns averaged mutual information matrix, averaged mutual information matrix for the inferred edges, the averaged mutual information values for the inferred edges, the averaged mutual information values for the non-inferred edges, maximum F-scores, threshold mutual information values that gives maximum F-score, and also the same input adjacency matrix of genes.

Usage

```
evalnetworks(E, G, net, infilepath)
```

Arguments

`E` number of experiments.

`G` Number of genes.

`net` Adjacency matrix of genes (not symmetric).

`infilepath` The path where all the data files kept.

Details

It reads all the mutual information matrices from the folders and take the average and then derive the other parameters from it.

Value

`evalnetworks` returns an object of class ’"res"’ which is a list with components:

`aveMImat` averaged mutual information matrix of the overall experiments.

`rmat` averaged mutual information matrix for the inferred edges of the overall experiments.

`tpre` All the averaged mutual information values for the inferred edges.

`fnre` All the averaged mutual information values for the non-inferred edges.

`Fscores` Maximum F-score values for all the experiments.

`Ithr` The threshold mutual information values, which gives maximum F-score, for all the experiments.

`net` Adjacency matrix of genes (not symmetric).

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`readNet`

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

# infilepath <- "./netmes/data/syn/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/syn/"

#This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworks(E, G, net, infilepath)

#boxplot(res$Fscores) #optional

#save(res, file = "resDAG")  # is used for saving boxplots
```

`evalnetworksNames` *Computing all the statistics from data of biological network*

Description

This function takes the adjacency matrix, the data sets folder path and number of mutual information steps for the therhold mutual information. It returns averaged mutual information matrix, averaged mutual information matrix for the inferred edges, the averaged mutual information values for the inferred edges, the averaged mutual information values for the non-inferred edges, maximum F-scores, threshold mutual information values that gives maximum F-score, and also the same input adjacency matrix of genes.

Usage

```
evalnetworksNames(E, G, net, infilepath, namesSif)
```

Arguments

`E` number of experiments.

`G` Number of genes.

`net` Adjacency matrix of genes (not symmetric).

`infilepath` The path where all the data files kept.

`namesSif` names of the genes corresponding to the adjacency matrix.

Details

It reads all the mutual information matrices from the folders and take the average and then derive the other parameters from it. This function is designed for biological networks where the gene names need to be derived along with the gene interactions from the files.

Value

`evalnetworksNames` returns an object of class ’"res"’ which is a list with components:

`aveMImat` averaged mutual information matrix of the overall experiments.

`rmat` averaged mutual information matrix for the inferred edges of the overall experiments.

`tpre` All the averaged mutual information values for the inferred edges.

`fnre` All the averaged mutual information values for the non-inferred edges.

`Fscores` Maximum F-score values for all the experiments.

`Ithr` The threshold mutual information values, which gives maximum F-score, for all the experiments.

`net` Adjacency matrix of genes (not symmetric).

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`readSifNet`, `readNet`, `evalnetworksNames`

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments

# infilenet <- "./netmes/data/networkEcoli.sif"
infilenet <- "/home/zeyneb/Desktop/netmes/data/networkEcoli.sif"
```

```
resEC <- readSifNet(infilenet, G)   # true network (+1 edge/ activator - -1 edges/repr.
net <- resEC$net

# infilepath <- "./netmes/data/bio/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/bio/"

# This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworksNames(E, G, resEC$net, infilepath, resEC$names)

boxplot(res$Fscores)

# save(res, file = "res_EColi")
# if desired for later analysis.
```

motifcount3 *Motif statistics for regulatory networks*

Description

This function takes the adjacency matrix, averaged mutual information matrix for the inferred edges, averaged mutual information matrix and number of genes as input. It returns

the data sets folder path and number of mutual information steps for the therhold mutual information. Considering 3 gene case, it computes number of motifs, probability to observe a motif, average mutual information of motifs and node indices for motifs.

Usage

```
motifcount3(net, rmat, MImat, G)
```

Arguments

net	adjacency matrix of genes (not symmetric).
rmat	averaged mutual information matrix for the inferred edges of the overall experiments.
MImat	averaged mutual information matrix of the overall experiments.
G	number of genes in the network.

Details

It only consider 3 genes case and search for all the possible motifs with 3 genes.

Value

motifcount3 returns an object of class ’"res"’ which is a list with components:

moty	number of motifs.
motyprob	probability to observe a motif.
motyMI	average mutual information of motifs.
emotnodes	node indices for motifs.

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

plotMotifs, evalnetworks, evalnetworksNames

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

# infilepath <- "./netmes/data/syn/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/syn/"

#This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworks(E, G, net, infilepath)

res2 <- motifcount3(net, res$rmat, res$aveMImat, G)

plotMotifs(res2)

# if True Reconstruction Rate wished to be plotted
envm <- res2$motyprob
y <- c()
xg <- c()
i <- 2
aux <- get(as.character(i), envir = envm)
y <- append(y, aux, after = length(y))
xg <- append(xg, rep(1, length(aux)), after = length(xg))
boxplot(y ~ xg, names = c("DAG"), ylab = expression(p), cex.lab = 1.7, cex.axis = 1.6, col =
```

`plotDs`	*TPR for all of the re or ac types of edges*

Description

This function takes the object that is the result of TPRall function and Sm and computes Ds values and plot it.

Usage

```
plotDs(TPReALL, Sm, fig=TRUE)
```

Arguments

`TPReALL`	Object that is the result of TPRall function.
`Sm`	Sum of degrees of nodes.
`fig`	By default it saves the figures of the results. If not wanted then set it FALSE.

Details

The influence of the degree of nodes on the inferability of edges is computed and plotted as Ds.eps.

Value

`plotDs` returns an object of class '"ds"' which is a list with components:

`gem`	Mean values of the degree of nodes.
`ges`	Standard deviation values of the degree of nodes.
`Ds.eps`	(If fig is set as TRUE or by default) Saves the figure Ds.eps to the current directory where x-axis sum of degree values, Ds, and y-axis is for corresponding TPR values.

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`plotMotifs`, `evalnetworks`, `evalnetworksNames`

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

Sm <- sumdeg(net)

# infilepath <- "./netmes/data/syn/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/syn/"

# This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworks(E, G, net, infilepath)

TPRs <- TPRall(res)  #also save the fig

Ds <-plotDs(TPRs$TPReALL, Sm)
```

---

plotMotifs *Print to screen in latex format of motif statistics*

Description

This function takes the object res (that includes all the explored statistics) and outputs the table in latex format that includes standard deviation, mean and the sum of edges.# ***

Usage

```
plotMotifs(res)
```

Arguments

`res` an object of class ’"res"’ which is a list with components: # *** .

Value

`plotMotifs` returns the table in latex format.

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

motifcount3, evalnetworksNames

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

# infilepath <- "./netmes/data/syn/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/syn/"

# This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworks(E, G, net, infilepath)

res2 <- motifcount3(net, res$rmat, res$aveMImat, G)

plotMotifs(res2)
```

readNet *Fetching true network from .sif file*

Description

This function takes the sythetic .sif file, which has interactions of genes, as input and converts it into an adjanceny matrix.

Usage

```
readNet(infile, G)
```

Arguments

`infile`	The path for the .sif file that has true interactions of the genes.
`G`	Number of genes.

Details

It assigns 1 for 'ac' and and -1 for 're' and then fills by zeros for the rest of the matrix. This function is for synthetic networks as the gene names are expilicit as numbered. Note that the matrix is not symmetric. An example of .sif file can be obtained from SynTRen where the underlying network has gene labels as 'numbers'.

Value

`readNet` returns the genes' relation matrix where activator edges, repressor edges and absence of edges are represented by 1, -1 and 0 respectively.

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

Examples

```
# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

G <- 100 # number of genes in the network
net <- readNet(infile, G)
tnet <- abs(net) + abs(t(net)) # true network (make symmetric)
tnet <- 1*(tnet > 0)           # in case there were bi-directional connections
```

readSifNet — *Fetching true biological network from .sif file*

Description

This function is for biological networks where the names of the genes are arbitrary and needs to be derived along with the interaction of the genes. It provides the adjanceny matrix and and the corresponding gene names of the matrix.

Usage

```
readSifNet(infilenet, G)
```

Arguments

`infilenet` The path for the .sif file that has true interactions of the genes.

`G` Number of genes.

Details

It assigns 1 for 'ac' and and -1 for 're' and then fills by zeros the rest of the matrix. Note that the matrix is not symmetric.

Value

`readSifNet` returns an object of class '"res"' which is a list with components:

`net` the genes' relation matrix where activator edges, repressor edges and absence of edges are represented by 1, -1 and 0 respectively.

`names` the gene names corresponding to the row and column names of the adjanceny matrix.

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`readSifNet`, `evalnetworks`, `evalnetworksNames`

Examples

```
# infilenet <- "./netmes/data/networkEcoli.sif"
infilenet <- "/home/zeyneb/Desktop/netmes/data/networkEcoli.sif"


G <- 100 # number of genes in the network
resEC <- readSifNet(infilenet, G)   # true network (+1 edge/ activator - -1 edges/repr.
net <- resEC$net
tnet <- abs(resEC$net) + abs(t(resEC$net)) # true network (make symmetric)
tnet <- 1*(tnet > 0)            # in case there were bi-directional connections
```

sumdeg *Sum of degree calculation for each node*

Description

This function takes the adjacency matrix, and calculate sum of degrees.

Usage

```
sumdeg(net)
```

Arguments

net	adjacency matrix of genes (not symmetric).

Details

It only considers all edges.

Value

`sumdeg` returns sum of degrees of noes as Sm

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`plotMotifs`, `evalnetworks`, `evalnetworksNames`

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

Sm <- sumdeg(net)
```

TPRall *TPR for all of the re or ac types of edges*

Description

This function takes the object that is the result of evalnetworks or evalnetworknames. It results the True positive rate (TPR) of edges for respressor and activator types.

Usage

```
TPRall(res)
```

Arguments

`res`	Object that is the result of evalnetworks or evalnetworknames.

Details

It only considers ac and re edges.

Value

`TPRall` returns an object of class '"TPRs"' which is a list with components:

`TPReALL`	TPR of all edges.
`TPReALLac`	TPR of ac types of edges.
`TPReALLre`	TPR of re types of edges.

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`plotMotifs`, `evalnetworks`, `evalnetworksNames`

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"
```

```
net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

# infilepath <- "./netmes/data/syn/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/syn/"

# This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworks(E, G, net, infilepath)

TPRs <- TPRall(res)  #also save the fig
```

visnet *Plotting the inferability of a network with colored edges*

Description

This function takes the interaction file, names of genes, and averaged mutual information matrix for the inferred edges as input and outputs a coloured edge graph for the true gene network.

Usage

```
visnet(net, names, mat, Sm)
```

Arguments

net	Adjacency matrix of genes (not symmetric).
names	Names of the genes.
mat	Averaged mutual information matrix for the inferred edges of the overall experiments.
Sm	Edge specific measure, which is the sum of outgoing edges of the regulating gene and the incoming edges of the regulated gene for an edge between them. Do not enter this parameter if you do not want it to be labelled on each edge in the network.

Details

It uses igraph software to plot the directed and coloured true gene network.

Value

`visnet` returns graphical representation of genes with the edge specific measure labeled on each edge and the edges are coloured with four colour (red, green, blue and black from worst to best respectively) regardin the average mutual information of the edge..

References

Gokmen Altay and Frank Emmert-Streib. Revealing differences in gene network inference algorithms on the network level by ensemble methods. Bioinformatics, 2010, 26(14),1738-1744.

Frank Emmert-Streib, Gokmen Altay. Local network-based measures to assess the inferability of different regulatory networks. IET Syst. Biol. 2010, Volume 4, Issue 4, p.277-288.

See Also

`readSifNet`, `evalnetworks`, `evalnetworksNames`

Examples

```
G <- 100 # number of genes in the network
E <- 5 # number of experiments (data sets)

# infile <- "./netmes/data/networkDAG.sif"
infile <- "/home/zeyneb/Desktop/netmes/data/networkDAG.sif"

net <- readNet(infile, G)   # true network (+1 edge/ activator - -1 edges/repr.)

# infilepath <- "./netmes/data/syn/"
infilepath <- "/home/zeyneb/Desktop/netmes/data/syn/"

#This is an example path. Change this path wrt the path of data in your computer.

res <- evalnetworks(E, G, net, infilepath)

Sm <- sumdeg(net)

visnet(net, c("no"), res$rmat, Sm)

# Sm is optional if wished to be labelled on edges.
# Adjust the plot by selecting "Layout" -> "Fruchterman-Reingold"

# visnet(net, resEC$names, res$rmat, Sm)  # for biological networks-resEC$names containes ge
# This can be used with evalnetworksNames funtion.
```

Index