Package 'DiffGraph'

November 19, 2017

Type Package

Title DiffGraph: An R Package for Identifying Gene Network Rewiring using Differential Graphical Models

Version 1.1.1

Author Xiao-Fei Zhang

Maintainer Xiao-Fei Zhang <zhangxf@mail.ccnu.edu.cn>

Description We develop DiffGraph, an R package that integrates four influential differential graphical models for identifying gene network rewiring between two different conditions from gene expression data. The input and output of different models are packaged in the same format, making it convenient for users to compare different models using a wide range of datasets and carry out follow-up analysis. Furthermore, the inferred differential networks are visualized in both non-interactive and interactive manners. The package is useful for identifying gene network rewiring from an input dataset, comparing the predictions of different well-established methods, and visualizing the results.

License GPL (>= 2)

Encoding UTF-8

NeedsCompilation no

Depends R (>= 2.10), igraph, MASS, Matrix

R topics documented:

DiffGraph-package *DiffGraph: An R Package for Identifying Gene Network Rewiring using Differential Graphical Models*

Description

DiffGraph is an R package for identifying gene network rewiring between two different conditions from gene expression data.

Details

DiffGraph integrates four influential differential graphical models which identify gene network rewiring through estimating the difference of condition specific precision matrices: fused graphical lasso (FGL), lasso penalized D-Trace loss (Dtrace), perturbed-node joint graphical lasso (PNJGL) and prior information-induced differential network analysis (pDNA). DiffGraph also provides three approaches to compute the condition specific covariance matrices (Pearson, Spearman and Kendall). To fit normal data using Gaussian graphical models, we suggest using Pearson correlation. To fit non-normal data using nonparanormal graphical models, Spearman's rho or Kendall's tau correlation is suggested. DiffGraph outputs the estimated differential networks as igraph graphs, which can be visualized in both non-interactive and interactive manners using the igraph R package.

Author(s)

Xiao-Fei Zhang

Maintainer: Xiao-Fei Zhang <zhangxf@mail.ccnu.edu.cn>

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), DiffGraph: An R package for identifying gene network rewiring using differential graphical models

See Also

FGL, Dtrace, PNJGL, pDNA, TCGA.BRCA, TCGA.GBM

Dtrace *Lasso penalized D-Trace loss*

Description

This function is implemented to identify gene network rewiring using the Dtrace model.

Usage

```
Dtrace(X, lambda, covType = "pearson", tol = 1e-05, maxiter = 500)
```

Arguments

X	A list (length=2) of input matrices. They can be data matrices or covariance matrices. If every matrix in the X is a symmetric matrix, the matrices are assumed to be covariance matrices among genes. If the input are data matrices, the rows represent the observations, and the columns represents the genes.
lambda	A nonnegative number. This tuning paramter controls the sparsity level of the estimated differential network.
covType	A parameter to decide which approach is selected to compute the sample covariance matrices. If covType = "pearson", it means that we compute (when input X represents data) the sample covariance matrices using Pearson correlation and that Dtrace is implemented based on Gaussian graphical models. If data is normal, we suggest covType = "pearson". If covType = "kendall" or "spearman", it means that we compute (when input X represents data) the sample covariance matrices using kendall's tau correlation or Spearman's rho correlation and that Dtrace is implemented based on nonparanormal graphical models. If data is non-normal, we suggest covType = "kendall" or "spearman".
tol	The tolerance parameter for convergence criteria.
maxiter	The maximum number of iterations.

Details

Lasso penalized D-Trace loss (Dtrace) estimates the precision matrix difference directly without estimating the condition specific precision matrices. One major difference of Dtrace as compared to FGL is that the Dtrace method only requires sparsity assumption on the differential network while FGL applies sparsity penalties to both the condition specific networks and the differential network. Dtrace also has an advantage of using much smaller sample size to achieve competitive performance.

We solve the optimization problem of Dtrace using the accelerated proximal gradient method.

Value

Delta	The estimated difference of condition specific precision matrices.
Delta.graph.full	The estimated differential network over all nodes.
Delta.graph.connected	The estimated differential network over only the connected nodes.

Author(s)

Xiao-Fei Zhang

Maintainer: Xiao-Fei Zhang <zhangxf@mail.ccnu.edu.cn>

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), DiffGraph: An R package for identifying gene network rewiring using differential graphical models.

Huili Yuan, Ruibin Xi, Chong Chen and Minghua Deng (2017). Differential network analysis via lasso penalized D-trace loss. Biometrika, 104(4), 755–770.

Dechao Tian, Quanquan Gu and Jian Ma (2016). Identifying gene regulatory network rewiring using latent differential graphical models. Nucleic Acids Research, 44(17), e140.

See Also

FGL, PNJGL, pDNA, TCGA.BRCA, TCGA.GBM

Examples

```
# Identify differential network between breast cancer subtypes
data(TCGA.BRCA)
X = TCGA.BRCA$X[1,]
dtrace.results= Dtrace(X, 0.45, covType = "spearman")
net.dtrace = dtrace.results$Delta.graph.connected
# Visualize the estimated differential network in an interactive manner.
tkid <- tkplot(net.dtrace, vertex.size= degree(net.dtrace)*1.5, layout =layout_with_fr,
    vertex.color="red", vertex.label.cex=0.8, edge.width =1.5, edge.color="orange")
# Visualize the estimated differential network in a non-interactive manner.
# grab the coordinates from tkplot
l.dtrace <- tkplot.getcoords(tkid)
plot(net.dtrace, layout=l.dtrace, vertex.size= degree(net.dtrace)*1.5, vertex.color="red",
    vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")



# Identify differential network between glioblastoma subtypes
data(TCGA.GBM)
X = TCGA.GBM$X[1,]
dtrace.results= Dtrace(X, 0.35, covType = "spearman")
net.dtrace = dtrace.results$Delta.graph.connected
# Visualize the estimated differential network in an interactive manner.
tkid <- tkplot(net.dtrace, vertex.size= degree(net.dtrace)*1.5, layout =layout_with_fr,
    vertex.color="red", vertex.label.cex=0.8, edge.width =1.5, edge.color="orange")
# Visualize the estimated differential network in a non-interactive manner.
# grab the coordinates from tkplot
l.dtrace <- tkplot.getcoords(tkid)
plot(net.dtrace, layout=l.dtrace,  vertex.size= degree(net.dtrace)*1.5,  vertex.color="red",
    vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")
```

FGL *Fused graphical lasso*

Description

This function is implemented to identify gene network rewiring using the FGL model.

Usage

```
FGL(X, lambda1, lambda2, covType = "pearson", weights = "equal",
    penalize.diagonal = FALSE, tol = 1e-05, maxiter = 500, rho = 0.1,
    rho.incr = 1.05, rho.max = 1e+05)
```

Arguments

X	A list (length=2) of input matrices. They can be data matrices or covariance matrices. If every matrix in the X is a symmetric matrix, the matrices are assumed to be covariance matrices among genes. If the input are data matrices, the rows represent the observations, and the columns represents the genes.
lambda1	A nonnegative number. The hyperparameter controls the sparsity level of the estimated condition specific networks.
lambda2	A nonnegative number. The hyperparameter controls the sparsity level of the estimated differential network.
covType	A parameter to decide which approach we choose to compute the sample covariance matrices. If covType = "pearson", it means that we compute (when input X represents data directly) the sample covariance matrices using Pearson correlation and that FGL is implemented based on Gaussian graphical models. If data is normal, we suggest covType = "pearson". If covType = "kendall" or "spearman", it means that we compute (when input X represents data directly) the sample covariance matrices using kendall's tau correlation or Spearman's rho correlation and that FGL is implemented based on nonparanormal graphical models. If data is non-normal, we suggest covType = "kendall" or "spearman".
weights	Determines the putative sample size of each condition's data. Allowed values: a vector with length two; "equal", giving each condition weight 1; "sample.size", giving each condition weight corresponding to its sample size.
penalize.diagonal	Determines whether the sparsity penalty is applied to the diagonal of condition specific precision matrices.
tol	The tolerance parameter for convergence criteria.
maxiter	The maximum number of iterations.
rho	The penalty parameter in the ADMM algorithm.
rho.incr	The increase step parameter for varying penalty parameter rho.
rho.max	The maximum value of rho.

Details

Fused graphical lasso (FGL) is based on a penalized log likelihood approach. The ℓ_1 penalties are applied to both condition specific precision matrices and their difference, encouraging sparsity not only in the condition specific networks but also in the differential network. Similarities between different conditions are exploited by encouraging similar edge values across conditions.

We solve the optimization problem of FGL using the ADMM algorithm. We accelerate the ADMM iterations by adaptively changing ρ in iterations.

Value

Delta	The estimateed difference of condition specific precision matrices.
Delta.graph.full	The estimated differential network over all nodes.
Delta.graph.connected	The estimated differential network over only the connected nodes.

`Theta` A list of the estimated condition specific precision matrices.

`Theta.graph.full`
A list of the estimated condition specific networks over all nodes.

`Theta.graph.connected`
A list of the estimated condition specific networks over only the connected nodes.

Author(s)

Xiao-Fei Zhang

Maintainer: Xiao-Fei Zhang <zhangxf@mail.ccnu.edu.cn>

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), DiffGraph: An R package for identifying gene network rewiring using differential graphical models.

Patrick Danaher, Pei Wang and Daniela M. Witten (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 76(2), 373-397

See Also

`Dtrace`, `PNJGL`, `pDNA`, `TCGA.BRCA`, `TCGA.GBM`

Examples

```
##Identify differential network between breast cancer subtypes
data(TCGA.BRCA)
X = TCGA.BRCA$X[1,]
fgl.results= FGL(X, 0.05, 0.2, covType = "spearman")
net.fgl = fgl.results$Delta.graph.connected
# Visualize the estimated differential network in an interactive manner
tkid <- tkplot(net.fgl, vertex.size= degree(net.fgl)*1.5, layout =layout_with_fr,
    vertex.color="red", vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")
# grab the coordinates from tkplot
l.fgl <- tkplot.getcoords(tkid)
# Visualize the estimated differential network in a non-interactive manner.
plot(net.fgl, layout=l.fgl, vertex.size= degree(net.fgl)*1.5,  vertex.color="red",
    vertex.label.cex =0.9, edge.width =1.5, edge.color="orange")


## Identify differential network between glioblastoma subtypes
data(TCGA.GBM)
X = TCGA.GBM$X[1,]
fgl.results= FGL(X, 0.1, 0.16, covType = "spearman")
net.fgl = fgl.results$Delta.graph.connected
# Visualize the estimated differential network in an interactive manner
tkid <- tkplot(net.fgl, vertex.size= degree(net.fgl)*1.5, layout =layout_with_fr,
    vertex.color="red", vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")
# grab the coordinates from tkplot
l.fgl <- tkplot.getcoords(tkid)
# Visualize the estimated differential network in a non-interactive manner.
plot(net.fgl, layout=l.fgl, vertex.size= degree(net.fgl)*1.5, vertex.color="red",
```

```
vertex.label.cex =0.9, edge.width =1.5, edge.color="orange")
```

---

pDNA *Prior information-induced differential network analysis*

Description

This function is implemented to identify gene network rewiring using the pDNA model.

Usage

```
pDNA(X, lambda, covType = "pearson", tol = 1e-05, maxiter = 500)
```

Arguments

X	A matrix (K x 2) of list of input matrices, where K is the number of data types. The (k, g)-th element is the data matrix corresponding to the k-th data type and the g-th condition. They can be data matrices or covariance matrices. If every matrix in X is a symmetric matrix, the matrices are assumed to be covariance matrices among genes. If the input are data matrices, the rows represent the samples, and the columns represents the genes.
lambda	A nonnegative number. The tuning controls the sparsity level of the estimated differential networks.
covType	A parameter to decide which approach we choose to compute the sample covariance matrices. If covType = "pearson", it means that we compute (when input X represents data) the sample covariance matrices using Pearson correlation and that pDNA is implemented based on Gaussian graphical models. If data is normal, we suggest covType = "pearson". If covType = "kendall" or "spearman", it means that we compute (when input X represents data) the sample covariance matrices using kendall's tau correlation or Spearman's rho correlation and that pDNA is implemented based on nonparanormal graphical models. If data is non-normal, we suggest covType = "kendall" or "spearman".
tol	The tolerance parameter for convergence criteria.
maxiter	The maximum number of iterations.

Details

Prior information-induced differential network analysis (pDNA) is an extension of Dtrace to deal with the settings where gene expression measurements are collected using multiple data types. Information across different data types are integrated. Furthermore, pDNA can capture important genes that drive the changes of network in a similar manner to PNJGL.

We solve the optimization problem of pDNA using the local linear approximation method and the accelerated proximal gradient method.

Value

`Delta` List of the estimateed differences of condition specific precision matrices for different data types.

`Delta.graph.full`
List of the estimated differential network over all nodes for different data types.

`Delta.graph.connected`
List of the estimated differential network over only the connected nodes for different data types.

`Delta.weight` The difference of condition specific precision matrices averaged over different data types.

`Delta.graph.weight.full`
The weighted differential network over all nodes, where the weight of a pair of genes is the frequency of data types in which they are identified as a differential edge.

`Delta.graph.connected`
The weighted differential network over only the connected nodes, where the weight of a pair of genes is the frequency of data types in which they are identified as a differential edge.

Author(s)

Xiao-Fei Zhang

Maintainer: Xiao-Fei Zhang <zhangxf@mail.ccnu.edu.cn>

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), DiffGraph: An R package for identifying gene network rewiring using differential graphical models.

Xiao-Fei Zhang, Le Ou-Yang and Hong Yan (2017). Incorporating prior information into differential network analysis using non-paranormal graphical models. Bioinformatics, 33(16), 2436-2445

See Also

`Dtrace`, `FGL`, `PNJGL`, `TCGA.BRCA`,`TCGA.GBM`

Examples

```
# Identify differential network between breast cancer subtypes
data(TCGA.BRCA)
pdna.results= pDNA(TCGA.BRCA$X, 0.9, covType = "spearman")
net.pdna = pdna.results$Delta.graph.weight.connected
# Visualize the estimated differential network in an interactive manner.
tkid <- tkplot(net.pdna, vertex.size= degree(net.pdna)*1.5, layout =layout_with_fr,
               vertex.color="red", vertex.label.cex =0.9, edge.width =1.5, edge.color="orange")
# Visualize the estimated differential network in a non-interactive manner.
# grab the coordinates from tkplot
l.pdna <- tkplot.getcoords(tkid)
plot(net.pdna, layout=l.pdna,  vertex.size= degree(net.pdna)*1.5,  vertex.color="red",
     vertex.label.cex=0.9,  edge.width =1.5, edge.color="orange")
```

```
# Identify differential network between glioblastoma subtypes
data(TCGA.GBM)
pdna.results= pDNA(TCGA.GBM$X, 0.64, covType = "spearman")
net.pdna = pdna.results$Delta.graph.weight.connected
# Visualize the estimated differential network in an interactive manner.
tkid <- tkplot(net.pdna, vertex.size= degree(net.pdna)*1.5, layout =layout_with_fr,
               vertex.color="red", vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")
# Visualize the estimated differential network in a non-interactive manner.
# grab the coordinates from tkplot
l.pdna <- tkplot.getcoords(tkid)
plot(net.pdna, layout=l.pdna,  vertex.size= degree(net.pdna)*1.5,  vertex.color="red",
     vertex.label.cex =0.9, edge.width =1.5, edge.color="orange")
```

PNJGL *Perturbed-node joint graphical lasso*

Description

This function is implemented to identify gene network rewiring using the PNJGL model.

Usage

```
PNJGL(X, lambda1, lambda2, covType = "pearson", weights = "equal",
    penalize.diagonal = FALSE, tol = 1e-05, maxiter = 500, rho = 0.1,
    rho.incr = 1.05, rho.max = 1e+05)
```

Arguments

X	A list (length=2) of input matrices. They can be data matrices or covariance matrices. If every matrix in the X is a symmetric matrix, the matrices are assumed to be covariance matrices among genes. If the input are data matrices, the rows represent the observations, and the columns represents the genes.
lambda1	A nonnegative number. The hyperparameter controls the sparsity level of the estimated condition specific networks.
lambda2	A nonnegative number. The hyperparameter controls the similarities between the estimated condition specific networks.
covType	A parameter to decide which approach we choose to compute the sample covariance matrices. If covType = "pearson", it means that we compute (when input X represents data directly) the sample covariance matrices using Pearson correlation and that PNJGL is implemented based on Gaussian graphical models. If data is normal, we suggest covType = "pearson". If covType = "kendall" or "spearman", it means that we compute (when input X represents data directly) the sample covariance matrices using kendall's tau correlation or Spearman's rho correlation and that PNJGL is implemented based on nonparanormal graphical models. If data is non-normal, we suggest covType = "kendall" or "spearman".
weights	Determines the putative sample size of each condition's data. Allowed values: a vector with length two; "equal", giving each condition weight 1; "sample.size", giving each condition weight corresponding to its sample size.

`penalize.diagonal`
: Determines whether the sparsity penalty is applied to the diagonal of condition specific precision matrices.

`tol`
: The tolerance parameter for convergence criteria.

`maxiter`
: The maximum number of iterations for the ADMM algorithm.

`rho`
: The penalty parameter in the ADMM algorithm.

`rho.incr`
: The increase step parameter for varying penalty parameter rho.

`rho.max`
: The maximum value of rho.

Details

Perturbed-node joint graphical lasso (PNJGL) is similar to FGL. The major difference between the two models is that FGL assumes that the differences between two condition specific networks arise from individual edges, while PNJGL assumes that network differences arise from certain nodes perturbed across conditions. A row-column overlap norm penalty is applied to the precision matrix difference to identify the crucial nodes driving network rewiring.

We solve the optimization problem of PNJGL using the ADMM algorithm. We accelerate the ADMM iterations by adaptively changing ρ in iterations.

Value

`Delta`
: The estimateed difference of condition specific precision matrices.

`Delta.graph.full`
: The estimated differential network over all nodes.

`Delta.graph.connected`
: The estimated differential network over only the connected nodes.

`Theta`
: A list of the estimated condition specific precision matrices.

`Theta.graph.full`
: A list of the estimated condition specific networks over all nodes.

`Theta.graph.connected`
: A list of the estimated condition specific networks over only the connected nodes.

Author(s)

Xiao-Fei Zhang

Maintainer: Xiao-Fei Zhang <zhangxf@mail.ccnu.edu.cn>

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), DiffGraph: An R package for identifying gene network rewiring using differential graphical models.

Karthik Mohan, Palma London, Daniela Witten and Su-In Lee (2014). Node-based learning of multiple Gaussian graphicalmodels. Journal of Machine Learning Research, 15(1), 445-488.

See Also

`Dtrace`, `FGL`, `pDNA`, `TCGA.BRCA`, `TCGA.GBM`

Examples

```
# Identify differential network between breast cancer subtypes
data(TCGA.BRCA)
X = TCGA.BRCA$X[1,]
pnjgl.results= PNJGL(X, 0.3, 1.2, covType = "spearman")
net.pnjgl = pnjgl.results$Delta.graph.connected
# Visualize the estimated differential network in an interactive manner.
tkid <- tkplot(net.pnjgl, vertex.size= degree(net.pnjgl)*.3, layout =layout_with_fr,
    vertex.color="red", vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")
# grab the coordinates from tkplot
l.pnjgl <- tkplot.getcoords(tkid)
plot(net.pnjgl, layout=l.pnjgl,  vertex.size= degree(net.pnjgl)*.3,  vertex.color="red",
    vertex.label.cex =0.9, edge.width =1.5, edge.color="orange")



# Identify differential network between glioblastoma subtypes
data(TCGA.GBM)
X = TCGA.GBM$X[1,]
pnjgl.results= PNJGL(X, 0.2, 1.5, covType = "spearman")
net.pnjgl = pnjgl.results$Delta.graph.connected
# Visualize the estimated differential network in an interactive manner.
tkid <- tkplot(net.pnjgl, vertex.size= degree(net.pnjgl)*.3, layout =layout_with_fr,
    vertex.color="red", vertex.label.cex=0.9, edge.width =1.5, edge.color="orange")
# grab the coordinates from tkplot
l.pnjgl <- tkplot.getcoords(tkid)
plot(net.pnjgl, layout=l.pnjgl,  vertex.size= degree(net.pnjgl)*.3,  vertex.color="red",
    vertex.label.cex =0.9,edge.width =1.5, edge.color="orange")
```

TCGA.BRCA *TCGA breast cancer data*

Description

The TCGA breast cancer gene expression datasets used in our case study. The datasets are obtained from the TCGA database. They are collected from three platforms: mRNA expression (Agilent G450 microarray), mRNA expression (RNA sequencing) and copy number variants (Affymetrix genome-wide human SNP Array 6.0). They include expression profiles for 207 luminal A cancers and 86 basal-like cancers. The data only include expression measurement of genes that overlap with the breast cancer pathway collected from the Kyoto Encyclopedia of Genes and Genomes database.

Usage

```
data("TCGA.BRCA")
```

Author(s)

Xiao-Fei Zhang

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), DiffGraph: An R package for identifying gene network rewiring using differential graphical models.

The Cancer Genome Atlas Research Network (2012), Comprehensive molecular portraits of human breast tumors. Nature. 490(7418), 61-70. (http://cancergenome.nih.gov/)

See Also

Dtrace, FGL, PNJGL, pDNA, TCGA.GBM

Examples

```
data(TCGA.BRCA)
## maybe str(TCGA.BRCA) ...
```

TCGA.GBM *TCGA glioblastoma data*

Description

The TCGA glioblastoma gene expression data used in our case study. The data are obtained from the TCGA database. They are collected from three platforms: Agilent 244K Custom Gene Expression G450, Affymetrix HT Human Genome U133 Array Plate Set and Affymetrix Human Exon 1.0 ST Array. They include expression measurments for 81 proneural tumors and 124 mesenchymal tumors. The data only include expression measurements of genes that overlap with the RTK/PI3K, p53, Rb signaling pathways which consists of the most frequently altered genes in glioblastoma.

Usage

```
data("TCGA.GBM")
```

Author(s)

Xiao-Fei Zhang

Source

The Cancer Genome Atlas Research Network (2012), Comprehensive molecular portraits of human breast tumors. Nature. 490(7418), 61-70. (http://cancergenome.nih.gov/)

References

Xiao-Fei Zhang, Le Ou-Yang, Shuo Yang, Xiaohua Hu and Hong Yan (2017), JGNI: A graphical model for joint gene network inference across multiple subpopulations and data types.

Roel GW Verhaak, Katherine A Hoadley, Elizabeth Purdom, Victoria Wang, Yuan Qi, Matthew D Wilkerson, C Ryan Miller, Li Ding, Todd Golub, Jill P Mesirov, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in pdgfra, idh1, egfr, and nf1. Cancer cell, 17(1):98-110, 2010.

See Also

Dtrace, FGL, PNJGL, pDNA, TCGA.BRCA

Examples

```
data(TCGA.GBM)
## maybe str(TCGA.GBM) ...
```

Index