

# Package ‘HeteroGGM’

October 12, 2022

**Type** Package

**Title** Gaussian Graphical Model-Based Heterogeneity Analysis

**Version** 0.1.0

**Date** 2021-2-6

**Description** The goal of this package is to user-friendly realizing Gaussian graphical model-based heterogeneity analysis. Recently, several Gaussian graphical model-based heterogeneity analysis techniques have been developed. A common methodological limitation is that the number of subgroups is assumed to be known a priori, which is not realistic. In a very recent study (Ren et al., 2021), a novel approach based on the penalized fusion technique is developed to fully data-dependently determine the number and structure of subgroups in Gaussian graphical model-based heterogeneity analysis. It opens the door for utilizing the Gaussian graphical model technique in more practical settings. Beyond Ren et al. (2021), more estimations and functions are added, so that the package is self-contained and more comprehensive and can provide “more direct” insights to practitioners (with the visualization function). Reference:  
Ren, M., Zhang S., Zhang Q. and Ma S. (2021). Gaussian Graphical Model-based Heterogeneity Analysis via Penalized Fusion. *Biometrics*, <doi:10.1111/biom.13426>.

**License** GPL-3

**Encoding** UTF-8

**Imports** igraph, Matrix, MASS

**LazyData** true

**LazyLoad** yes

**RoxygenNote** 7.1.1

**Depends** R (>= 3.4)

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr, rmarkdown

**NeedsCompilation** no

**Author** Mingyang Ren [aut, cre],  
 Sanguo Zhang [aut],  
 Qingzhao Zhang [aut],  
 Shuangge Ma [aut]

**Maintainer** Mingyang Ren <renmingyang17@mails.ucas.ac.cn>

**Repository** CRAN

**Date/Publication** 2021-02-11 09:30:05 UTC

## R topics documented:

example.data . . . . .	2
FGGM . . . . .	3
FGGM.refit . . . . .	4
genelambda.obo . . . . .	5
GGMPF . . . . .	6
linked_node_names . . . . .	8
PGGMBC . . . . .	8
plot_network . . . . .	9
summary_network . . . . .	10
<b>Index</b>	<b>11</b>

---

example.data	<i>Some example data</i>
--------------	--------------------------

---

### Description

Some example data

### Format

A list including: data: The 600 x 20 matrix, the design matrix. L0: The subgroup to which each sample truly belongs. Mu0: The true mean parameters of 3 subgroups. Theta0: The true precision matrices of 3 subgroups. n\_all: The total sample size. K0: The true number of subgroups 3.

### Source

Simulated data (see examples in the function tuning.lambda.FGGM)

### Examples

data(example.data)

**Description**

The base function of Gaussian graphical model-based heterogeneity analysis via penalized fusion: identifying the order of subgroups and reconstructing the network structure.

**Usage**

```
FGGM(data, K, lambda1 = 0.5, lambda2 = 0.2, lambda3 = 2, a = 3, rho = 1,
      eps = 5e-2, niter = 20, maxiter=10, maxiter.AMA=5, initialization=TRUE,
      initialize, average=FALSE, asymmetric=TRUE, local_appro=TRUE,
      penalty = "MCP", theta.fusion=TRUE)
```

**Arguments**

data	n * p matrix, the design matrix.
K	Int, a selected upper bound of $K_0$ .
lambda1	A float value, the tuning parameter controlling the sparse of the mean parameter.
lambda2	A float value, the tuning parameter controlling the sparse of the precision matrix.
lambda3	A float value, the tuning parameter controlling the number of subgroup.
a	A float value, regularization parameter in MCP, the default setting is 3.
rho	A float value, the penalty parameter in ADMM algorithm of updating precision matrix Theta, the default setting is 1.
eps	A float value, algorithm termination threshold.
niter	Int, maximum number of cycles of the EM algorithm, the default setting is 20.
maxiter	Int, maximum number of cycles of the ADMM algorithm.
maxiter.AMA	Int, maximum number of cycles of the AMA algorithm.
initialization	The logical variable, whether to calculate the initial value, the default setting is T, if initialization = F, the initial value uses initialize.
initialize	A given initial value used if initialization = F.
average	The logical variable, whether to use averaging when integrating parameters that are identified as identical subgroups, the default setting is F, which means the estimated parameters for the subgroup with the largest sample size among the subgroups identified as identical subgroups is used as the final parameter for this subgroup.
asymmetric	The logical variable, symmetry of the precision matrices or not, the default setting is T.
local_appro	The logical variable, whether to use local approximations when updating mean parameters, the default setting is T.
penalty	The type of the penalty, which can be selected from c("MCP", "SCAD", "lasso").
theta.fusion	Whether or not the fusion penalty term contains elements of the precision matrices. The default setting is T.

**Value**

A list including all estimated parameters and the BIC value.

**Author(s)**

Mingyang Ren, Sanguo Zhang, Qingzhao Zhang, Shuangge Ma. Maintainer: Mingyang Ren <ren-mingyang17@mails.ucas.ac.cn>.

**References**

Ren, M., Zhang S., Zhang Q. and Ma S. (2020). Gaussian Graphical Model-based Heterogeneity Analysis via Penalized Fusion. *Biometrics*, Published Online.

---

 FGGM.refit

*Refitting of FGGM*


---

**Description**

Refitting when  $K_0$  is identified using FGGM().

**Usage**

```
FGGM.refit(data, K, lambda1 = 0.5, lambda2 = 0.2, lambda3 = 2, a = 3, rho = 1,
            eps = 5e-2, niter = 20, maxiter=10, maxiter.AMA=5,
            initialization=TRUE, initialize, average=FALSE,
            asymmetric=TRUE, local_appro=TRUE, penalty = "MCP", theta.fusion=TRUE)
```

**Arguments**

data	n * p matrix, the design matrix.
K	Int, a selected upper bound of $K_0$ .
lambda1	A float value, the tuning parameter controlling the sparse of the mean parameter.
lambda2	A float value, the tuning parameter controlling the sparse of the precision matrix.
lambda3	A float value, the tuning parameter controlling the number of subgroup.
a	A float value, regularization parameter in MCP, the default setting is 3.
rho	A float value, the penalty parameter in ADMM algorithm of updating precision matrix Theta, the default setting is 1.
eps	A float value, algorithm termination threshold.
niter	Int, maximum number of cycles of the algorithm, the default setting is 20.
maxiter	Int, maximum number of cycles of the ADMM algorithm.
maxiter.AMA	Int, maximum number of cycles of the AMA algorithm.
initialization	The logical variable, whether to calculate the initial value, the default setting is T, if initialization = F, the initial value uses initialize.

initialize	A given initial value used if initialization = F.
average	The logical variable, whether to use averaging when integrating parameters that are identified as identical subgroups, the default setting is F, which means the estimated parameters for the subgroup with the largest sample size among the subgroups identified as identical subgroups is used as the final parameter for this subgroup.
asymmetric	The logical variable, symmetry of the precision matrices or not, the default setting is T.
local_appro	The logical variable, whether to use local approximations when updating mean parameters, the default setting is T.
penalty	The type of the penalty, which can be selected from c("MCP", "SCAD", "lasso").
theta.fusion	Whether or not the fusion penalty term contains elements of the precision matrices. The default setting is T.

**Value**

A list including all estimated parameters and the BIC value after refitting.

**Author(s)**

Mingyang Ren, Sanguo Zhang, Qingzhao Zhang, Shuangge Ma. Maintainer: Mingyang Ren <ren-mingyang17@mails.ucas.ac.cn>.

**References**

Ren, M., Zhang S., Zhang Q. and Ma S. (2020). Gaussian Graphical Model-based Heterogeneity Analysis via Penalized Fusion. *Biometrics*, Published Online.

---

genelambda.obo                      *Generate tuning parameters*

---

**Description**

Generating a sequence of the tuning parameters (lambda1, lambda2, and lambda3).

**Usage**

```
genelambda.obo(nlambda1=10, lambda1_max=1, lambda1_min=0.05,
               nlambda2=10, lambda2_max=1, lambda2_min=0.01,
               nlambda3=10, lambda3_max=5, lambda3_min=0.5)
```

**Arguments**

<code>nlambda1</code>	The numbers of lambda 1.
<code>lambda1_max</code>	The maximum values of lambda 1.
<code>lambda1_min</code>	The minimum values of lambda 1.
<code>nlambda2</code>	The numbers of lambda 2.
<code>lambda2_max</code>	The maximum values of lambda 2.
<code>lambda2_min</code>	The minimum values of lambda 2.
<code>nlambda3</code>	The numbers of lambda 3.
<code>lambda3_max</code>	The maximum values of lambda 3.
<code>lambda3_min</code>	The minimum values of lambda 3.

**Value**

A sequence of the tuning parameters (`lambda1`, `lambda2`, and `lambda3`).

**Author(s)**

Mingyang Ren

**Examples**

```
lambda <- genelambda.obo(nlambda1=5,lambda1_max=0.5,lambda1_min=0.1, nlambda2=15,lambda2_max=1.5,
                        lambda2_min=0.1, nlambda3=10,lambda3_max=3.5,lambda3_min=0.5)
lambda
```

---

GGMPF

*GGM-based heterogeneity analysis.*


---

**Description**

The main function of Gaussian graphical model-based heterogeneity analysis via penalized fusion.

**Usage**

```
GGMPF(lambda, data, K, initial.selection="K-means",
       initialize, average=FALSE, asymmetric=TRUE,
       eps = 5e-2, maxiter=10, maxiter.AMA=5, local_appro=TRUE,
       trace = FALSE, penalty = "MCP", theta.fusion=TRUE)
```

**Arguments**

<code>lambda</code>	A list, the sequences of the tuning parameters ( <code>lambda1</code> , <code>lambda2</code> , and <code>lambda3</code> ).
<code>data</code>	$n * p$ matrix, the design matrix.
<code>K</code>	Int, a selected upper bound of <code>K_0</code> .
<code>initial.selection</code>	The different initial values from two clustering methods, which can be selected from <code>c("K-means", "dbscan")</code> .
<code>initialize</code>	A given initial values, which should be given when <code>initial.selection</code> is not in <code>c("K-means", "dbscan")</code> .
<code>average</code>	The logical variable, whether to use averaging when integrating parameters that are identified as identical subgroups, the default setting is F, which means the estimated parameters for the subgroup with the largest sample size among the subgroups identified as identical subgroups is used as the final parameter for this subgroup.
<code>asymmetric</code>	The logical variable, symmetry of the precision matrices or not, the default setting is T.
<code>eps</code>	A float value, algorithm termination threshold.
<code>maxiter</code>	Int, maximum number of cycles of the ADMM algorithm.
<code>maxiter.AMA</code>	Int, maximum number of cycles of the AMA algorithm.
<code>local_appro</code>	The logical variable, whether to use local approximations when updating mean parameters, the default setting is T.
<code>trace</code>	The logical variable, whether or not to output the number of identified subgroups during the search for parameters.
<code>penalty</code>	The type of the penalty, which can be selected from <code>c("MCP", "SCAD", "lasso")</code> .
<code>theta.fusion</code>	Whether or not the fusion penalty term contains elements of the precision matrices. The default setting is T.

**Value**

A list including all estimated parameters and the BIC values with all choices of given tuning parameters, and the selected optional parameters.

**Author(s)**

Mingyang Ren, Sanguo Zhang, Qingzhao Zhang, Shuangge Ma. Maintainer: Mingyang Ren <ren-mingyang17@mails.ucas.ac.cn>.

**References**

Ren, M., Zhang S., Zhang Q. and Ma S. (2020). Gaussian Graphical Model-based Heterogeneity Analysis via Penalized Fusion. *Biometrics*, Published Online, <https://doi.org/10.1111/biom.13426>.

---

linked_node_names	<i>Indexes the names of all nodes connected to some particular nodes in a subgroup.</i>
-------------------	---

---

**Description**

Indexes the names of all nodes connected to some particular nodes in a subgroup.

**Usage**

```
linked_node_names(summ, va_names, num_subgroup=1)
```

**Arguments**

summ	A list, the summary of the resulting network structures.
va_names	A vector, the names of nodes of interest.
num_subgroup	Int, the subgroup numbering.

**Value**

A list including the names of connected nodes to the nodes of interest in a subgroup.

---

PGGMBC	<i>Penalized GGM-based clustering.</i>
--------	--

---

**Description**

The main function of penalized Gaussian graphical model-based clustering with unconstrained covariance matrices.

**Usage**

```
PGGMBC(lambda, data, K, initial.selection="K-means", initialize, average=FALSE,
        asymmetric=TRUE, eps = 5e-2, maxiter=10,
        maxiter.AMA=5, local_appro=TRUE, trace = FALSE, penalty = "MCP")
```

**Arguments**

lambda	A list, the sequences of the tuning parameters (lambda1 and lambda2).
data	n * p matrix, the design matrix.
K	Int, a given number of subgroups.
initial.selection	The different initial values from two clustering methods, which can be selected from c("K-means","dbscan").



initialize	A given initial values, which should be given when initial.selection is not in c("K-means","dbscan").
average	The logical variable, whether to use averaging when integrating parameters that are identified as identical subgroups, the default setting is F, which means the estimated parameters for the subgroup with the largest sample size among the subgroups identified as identical subgroups is used as the final parameter for this subgroup.
asymmetric	The logical variable, symmetry of the precision matrices or not, the default setting is T.
eps	A float value, algorithm termination threshold.
maxiter	Int, maximum number of cycles of the ADMM algorithm.
maxiter.AMA	Int, maximum number of cycles of the AMA algorithm.
local_appro	The logical variable, whether to use local approximations when updating mean parameters, the default setting is T.
trace	The logical variable, whether or not to output the number of identified subgroups during the search for parameters.
penalty	The type of the penalty, which can be selected from c("MCP", "SCAD", "lasso").

**Value**

A list including all estimated parameters and the BIC values with all choices of given tuning parameters, and the selected optional parameters.

**Author(s)**

Mingyang Ren, Sanguo Zhang, Qingzhao Zhang, Shuangge Ma. Maintainer: Mingyang Ren <ren-mingyang17@mails.ucas.ac.cn>.

**References**

Ren, M., Zhang S., Zhang Q. and Ma S. (2020). Gaussian Graphical Model-based Heterogeneity Analysis via Penalized Fusion. *Biometrics*, Published Online, <https://doi.org/10.1111/biom.13426>.  
 Zhou, H., Pan, W., & Shen, X. (2009). Penalized model-based clustering with unconstrained covariance matrices. *Electronic journal of statistics*, 3, 1473.

---

plot_network	<i>Visualization of network structures.</i>
--------------	---

---

**Description**

Visualization of network structures.

**Usage**

```
plot_network(summ, num_subgroup = 1,
             vertex.size=2, vertex.label.cex=0.7,
             vertex.label.dist=0.75, edge.width = 0.1, l=0)
```

**Arguments**

summ	A list, the summary of the resulting network structures.
num_subgroup	Int/vector, the subgroup numbering.
vertex.size	The vertex size.
vertex.label.cex	The vertex label size.
vertex.label.dist	The distance of vertex labels.
edge.width	The edge width.
l	Node Coordinates.

**Value**

Visualization of network structure

---

summary_network	<i>The summary of the resulting network structures.</i>
-----------------	---

---

**Description**

Summarize the characteristics of the resulting network structures.

**Usage**

```
summary_network(opt_Mu_hat, opt_Theta_hat, data)
```

**Arguments**

opt_Mu_hat	A $p \times K0\_hat$ matrix, the optional mean vectors of $K0\_hat$ subgroups.
opt_Theta_hat	$n \times p \times K0\_hat$ matrix, the optional precision matrices of $K0\_hat$ subgroups.
data	A $n \times p$ matrix, the design matrix.

**Value**

A list including the overlap of edges of different subgroups, the number of edges, and the names of connected nodes to each nodes in each subgroup.

# Index

`example.data`, [2](#)

`FGGM`, [3](#)

`FGGM.refit`, [4](#)

`genelambda.obo`, [5](#)

`GGMPF`, [6](#)

`linked_node_names`, [8](#)

`PGGMBC`, [8](#)

`plot_network`, [9](#)

`summary_network`, [10](#)