

# Package ‘RNAseqNet’

October 12, 2022

**Type** Package

**Title** Log-Linear Poisson Graphical Model with Hot-Deck Multiple Imputation

**Version** 0.1.4

**Date** 2020-06-08

**Maintainer** Nathalie Vialaneix <nathalie.vialaneix@inrae.fr>

**Description** Infer log-linear Poisson Graphical Model with an auxiliary data set. Hot-deck multiple imputation method is used to improve the reliability of the inference with an auxiliary dataset. Standard log-linear Poisson graphical model can also be used for the inference and the Stability Approach for Regularization Selection (StARS) is implemented to drive the selection of the regularization parameter. The method is fully described in <[doi:10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819)>.

**License** GPL (>= 3)

**Repository** CRAN

**Depends** R (>= 3.1.0), ggplot2

**Imports** igraph (>= 1.0), hot.deck, PoiClaClu, glmnet, methods, utils

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.0

**NeedsCompilation** no

**Author** Alyssa Imbert [aut],  
Nathalie Vialaneix [aut, cre]

**Date/Publication** 2020-06-08 15:50:03 UTC

## R topics documented:

chooseSigma	2
GLMnetToGraph	3
GLMnetwork	4
GLMpath	6
HDImputed	6
HDpath	7
imputedGLMnetwork	8
imputeHD	9
lung	11
RNAseqNetUsersGuide	11
stabilitySelection	12
stars	13
thyroid	14
varIntra	14
<b>Index</b>	<b>16</b>

---

chooseSigma	<i>Select the threshold sigma for hd-MI.</i>
-------------	--

---

### Description

chooseSigma computes the average intra-donor pool variance for different values of sigma. It helps choosing a sigma that makes a good trade-off between homogeneity within the pool of donors and variety (large enough number of donors in every pool).

### Usage

```
chooseSigma(X, Y, sigma_list, seed = NULL)
```

### Arguments

X	n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
Y	auxiliary dataset (n' x q numeric matrix or data frame)
sigma_list	a sequence of increasing positive values for sigma (numeric vector)
seed	single value, interpreted as an integer, used to initialize the random number generation state

### Details

The average intra-donor pool variance is described in (Imbert *et al.*, 2018).

### Value

a data frame with the values of sigma and the corresponding intra-donor pool variances

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>  
Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**References**

Imbert, A., Valsesia, A., Le Gall, C., Armenise, C., Lefebvre, G. Gourraud, P.A., Viguerie, N. and Villa-Vialaneix, N. (2018) Multiple hot-deck imputation for network inference from RNA sequencing data. *Bioinformatics*. doi: [10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819).

**See Also**

[varIntra](#)

**Examples**

```
data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
sigma_stats <- chooseSigma(lung, thyroid, 1:5)
## Not run: plot(sigma_stats, type = "b")
```

---

GLMnetToGraph

*Convert the result of imputedGLMnetwork or a matrix into a network.*

---

**Description**

GLMnetToGraph combines the  $m$  inferred networks, obtained from  $m$  imputed datasets, into a single stable network or convert a matrix of coefficients of a GLM model into a network (non zero coefficients are converted to edges)

**Usage**

```
GLMnetToGraph(object, threshold = 0.9)
```

**Arguments**

object	an object of class HDpath as obtained from the function <a href="#">imputedGLMnetwork</a> or a squared matrix with zero and non zero values
threshold	the percentage of times, among the $m$ imputed networks, that an edge has to be predicted to be in the final network. Used only for objects of class HDpath. Default to 0.9

**Value**

an 'igraph' object. See [igraph](#)

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>

Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**References**

Imbert, A., Valsesia, A., Le Gall, C., Armenise, C., Lefebvre, G. Gourraud, P.A., Viguerie, N. and Villa-Vialaneix, N. (2018) Multiple hot-deck imputation for network inference from RNA sequencing data. *Bioinformatics*. doi: [10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819).

**See Also**

[imputedGLMnetwork](#), [igraph](#)

**Examples**

```
data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
lambdas <- 4 * 10^(seq(0, -2, length = 10))
## Not run:
lung_hdmi <- imputedGLMnetwork(lung, thyroid, sigma = 2, lambdas = lambdas,
                              m = 10, B = 5)
lung_net <- GLMnetToGraph(lung_hdmi, 0.75)
lung_net
plot(lung_net)

## End(Not run)
```

---

GLMnetwork

*Infer a network from RNA-seq expression.*

---

**Description**

GLMnetwork infers a network from RNA-seq expression with the log-linear Poisson graphical model of (Allen and Liu, 2012).

**Usage**

```
GLMnetwork(counts, lambdas = NULL, normalize = TRUE)
```

**Arguments**

counts	a n x p matrix of RNA-seq expression (numeric matrix or data frame)
lambdas	a sequence of decreasing positive numbers to control the regularization (numeric vector). Default to NULL
normalize	logical value to normalize predictors in the log-linear Poisson graphical model. If TRUE, log normalization and scaling are performed prior the model is fit. Default to TRUE

**Details**

When input lambdas are null the default sequence of [glmnet](#) for the first model (the one with the first column of count as the target) is used.

**Value**

S3 object of class GLMnetwork: a list consisting of

- lambda regularization parameters used for LLGM path(vector)
- path a list having the same length than lambda. It contains the estimated coefficients (in a matrix) along the path

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>

Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**References**

Allen, G. and Liu, Z. (2012) A log-linear model for inferring genetic networks from high-throughput sequencing data. In *Proceedings of IEEE International Conference on Bioinformatics and Biomedecine (BIBM)*.

**See Also**

[stabilitySelection](#)

**Examples**

```
data(lung)
lambdas <- 4 * 10^(seq(0, -2, length = 10))
ref_lung <- GLMnetwork(lung, lambdas = lambdas)
```

---

GLMpath

*Methods for 'GLMpath' objects.*

---

### Description

Methods for the result of [GLMnetwork](#) (GLMpath object)

### Usage

```
## S3 method for class 'GLMpath'  
summary(object, ...)
```

```
## S3 method for class 'GLMpath'  
print(x, ...)
```

### Arguments

object	GLMpath object
...	not used
x	GLMpath object

### Author(s)

Alyssa Imbert, <alyssa.imbert@gmail.com>  
Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

### See Also

[GLMnetwork](#)

---

HDImputed

*Methods for 'HDImputed' objects.*

---

### Description

Methods for the result of [imputeHD](#) (HDImputed object)

### Usage

```
## S3 method for class 'HDImputed'  
summary(object, ...)
```

```
## S3 method for class 'HDImputed'  
print(x, ...)
```

**Arguments**

object	HDImputed object
...	not used
x	HDImputed object

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>  
Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**See Also**

[imputeHD](#)

---

HDpath	<i>Methods for 'HDpath' objects.</i>
--------	--------------------------------------

---

**Description**

Methods for the result of [imputedGLMnetwork](#) (HDpath object)

**Usage**

```
## S3 method for class 'HDpath'  
summary(object, ...)  
  
## S3 method for class 'HDpath'  
print(x, ...)  
  
## S3 method for class 'HDpath'  
plot(x, ...)
```

**Arguments**

object	HDpath object
...	not used
x	HDpath object

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>  
Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**See Also**

[imputedGLMnetwork](#)

**Examples**

```

data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
lambdas <- 4 * 10^(seq(0, -2, length = 10))
## Not run:
lung_hdmi <- imputedGLMnetwork(lung, thyroid, sigma = 2, lambdas = lambdas,
                               m = 10, B = 5)

plot(lung_hdmi)

## End(Not run)

```

---

imputedGLMnetwork	<i>Multiple hot-deck imputation and network inference from RNA-seq data.</i>
-------------------	--

---

**Description**

imputedGLMnetwork performs a multiple hot-deck imputation and infers a network for each imputed dataset with a log-linear Poisson graphical model (LLGM).

**Usage**

```
imputedGLMnetwork(X, Y, sigma, m = 50, lambdas = NULL, B = 20)
```

**Arguments**

X	n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
Y	auxiliary dataset (n' x q numeric matrix or data frame)
sigma	affinity threshold for donor pool
m	number of replicates in multiple imputation (integer). Default to 50
lambdas	a sequence of decreasing positive numbers to control the regularization (numeric vector). Default to NULL
B	number of iterations for stability selection. Default to 20

**Details**

When input lambdas are null the default sequence of [glmnet](#) for the first model (the one with the first column of count as the target) is used. A common default sequence is generated for all imputed datasets using this method.



**Value**

S3 object of class HDpath: a list consisting of

- path a list of  $m$  data frames, each containing the adjacency matrix of the inferred network obtained from the corresponding imputed dataset. The regularization parameter is selected by StARS
- efreq a numeric matrix of size  $p \times p$ , which indicates the number of times an edge has been predicted among the  $m$  inferred networks

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>

Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**References**

Imbert, A., Valsesia, A., Le Gall, C., Armenise, C., Lefebvre, G. Gourraud, P.A., Viguerie, N. and Villa-Vialaneix, N. (2018) Multiple hot-deck imputation for network inference from RNA sequencing data. *Bioinformatics*. doi: [10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819).

**Examples**

```
data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
lambdas <- 4 * 10^(seq(0, -2, length = 10))
## Not run:
lung_hdmi <- imputedGLMnetwork(lung, thyroid, sigma = 2, lambdas = lambdas,
                              m = 10, B = 5)

## End(Not run)
```

---

imputeHD

*Impute missing row datasets with multiple hot deck.*

---

**Description**

imputeHD performs multiple hot-deck imputation on an input data frame with missing rows. Each missing row is imputed with a unique donor. This method requires an auxiliary dataset to compute similarities between individuals and create the pool of donors.

**Usage**

```
imputeHD(X, Y, sigma, m = 50, seed = NULL)
```

## Arguments

X	n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
Y	auxiliary dataset (n' x q numeric matrix or data frame)
sigma	threshold for hot-deck imputation (numeric, positive)
m	number of replicates in multiple imputation (integer). Default to 50
seed	single value, interpreted as an integer, used to initialize the random number generation state. Default to NULL (not used in this case)

## Details

Missing values are identified by matching rownames in X and Y. If rownames are not provided the missing rows in X are supposed to correspond to the last rows of Y.

## Value

S3 object of class HDImputed: a list consisting of

- donors a list. Each element of this list contains the donor pool for every missing observations
- draws a data frame which indicates which donor was chosen for each missing samples
- data a list of m imputed datasets

## Author(s)

Alyssa Imbert, <alyssa.imbert@gmail.com>  
Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

## References

Imbert, A., Valsesia, A., Le Gall, C., Armenise, C., Lefebvre, G. Gourraud, P.A., Viguerie, N. and Villa-Vialaneix, N. (2018) Multiple hot-deck imputation for network inference from RNA sequencing data. *Bioinformatics*. doi: [10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819).

## See Also

[chooseSigma](#), [imputedGLMnetwork](#)

## Examples

```
data(lung)
data(thyroid)
nobs <- nrow(lung)
miss_ind <- sample(1:nobs, round(0.2 * nobs), replace = FALSE)
lung[miss_ind, ] <- NA
lung <- na.omit(lung)
imputed_lung <- imputeHD(lung, thyroid, sigma = 2)
```



**Value**

Character string giving the file location. If `html = TRUE` and `view = TRUE`, the HTML document reader is started and the User's Guide is opened in it.

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>  
 Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**Examples**

```
RNAseqNetUsersGuide(view = FALSE)
RNAseqNetUsersGuide(html = FALSE)
## Not run: RNAseqNetUsersGuide()
```

---

stabilitySelection      *Selection of the regularization parameter by StARS (Liu et al., 2010).*

---

**Description**

`stabilitySelection` implements the regularization parameter selection of (Liu et al., 2010) called 'Stability Approach to Regularization Selection' (StARS).

**Usage**

```
stabilitySelection(counts, lambdas = NULL, B = 20)
```

**Arguments**

<code>counts</code>	a $n \times p$ matrix of RNA-seq expression (numeric matrix or data frame)
<code>lambdas</code>	a sequence of decreasing positive numbers to control the regularization (numeric vector). Default to NULL
<code>B</code>	number of iterations for stability selection. Default to 20

**Details**

When input `lambdas` are null the default sequence of `glmnet` (see [GLMnetwork](#) for details).

**Value**

S3 object of class `stabilitySelection`: a list consisting of

- `lambdas` numeric regularization parameters used for regularization path
- `B` number of iterations for stability selection
- `best index` of the regularization parameter selected by StARS in `lambdas`
- `variabilities` numeric vector having same length than `lambdas` and providing the variability value as defined by StARS along the path

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com> Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**References**

Liu, H., Roeder, K. and Wasserman, L. (2010) Stability approach to regularization selection (StARS) for high dimensional graphical models. In *Proceedings of Neural Information Processing Systems (NIPS 2010)*, **23**, 1432-1440, Vancouver, Canada.

**See Also**

[GLMnetwork](#)

**Examples**

```
data(lung)
lambdas <- 4 * 10^(seq(0, -2, length = 5))
stability_lung <- stabilitySelection(lung, lambdas = lambdas, B = 4)
## Not run: plot(stability_lung)
```

---

stars

*Methods for 'stars' objects.*

---

**Description**

Methods for the result of [stabilitySelection](#) (stars object)

**Usage**

```
## S3 method for class 'stars'
summary(object, ...)

## S3 method for class 'stars'
print(x, ...)

## S3 method for class 'stars'
plot(x, ...)
```

**Arguments**

object	stars object
...	not used
x	stars object

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>  
 Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**See Also**

[stabilitySelection](#)

---

thyroid	<i>RNA-seq expression from thyroid tissue (GTEx).</i>
---------	---

---

**Description**

This data set is a small subset of the full data set from GTEx. It contains RNA-seq expressions measured from thyroid tissue. The RNA-seq expressions have been normalized with the TMM method.

**Format**

a data frame with 221 rows and 50 variables (genes). Rownames are identifiers for individuals.

**Author(s)**

Alyssa Imbert <alyssa.imbert@gmail.com>

**Source**

The raw data were downloaded from <https://gtexportal.org/>. The TMM normalisation of RNA-seq expression was performed with the R package edgeR.

---

varIntra	<i>Average intra-donor pool variance.</i>
----------	---

---

**Description**

varIntra computes the average intra-donor pool variance.

**Usage**

```
varIntra(X, Y, donors)
```

**Arguments**

X	n x p numeric matrix containing RNA-seq expression with missing rows (numeric matrix or data frame)
Y	auxiliary dataset (n' x q numeric matrix or data frame)
donors	donor pool (a list, as given \$donors obtained from the function <a href="#">imputeHD</a> )

**Value**

varIntra returns a numeric value which is the average intra-donor pool variance, as described in (Imbert *et al.*, 2018).

**Author(s)**

Alyssa Imbert, <alyssa.imbert@gmail.com>

Nathalie Vialaneix, <nathalie.vialaneix@inrae.fr>

**References**

Imbert, A., Valsesia, A., Le Gall, C., Armenise, C., Lefebvre, G. Gourraud, P.A., Viguerie, N. and Villa-Vialaneix, N. (2018) Multiple hot-deck imputation for network inference from RNA sequencing data. *Bioinformatics*. doi: [10.1093/bioinformatics/btx819](https://doi.org/10.1093/bioinformatics/btx819).

**See Also**

[imputeHD](#), [chooseSigma](#)

# Index

`chooseSigma`, [2](#), [10](#), [15](#)

`glmnet`, [5](#), [8](#), [12](#)

`GLMnetToGraph`, [3](#)

`GLMnetwork`, [4](#), [6](#), [12](#), [13](#)

`GLMpath`, [6](#)

`GLMpath-class (GLMpath)`, [6](#)

`HDImputed`, [6](#)

`HDImputed-class (HDImputed)`, [6](#)

`HDpath`, [7](#)

`HDpath-class (HDpath)`, [7](#)

`igraph`, [4](#)

`imputedGLMnetwork`, [3](#), [4](#), [7](#), [8](#), [10](#)

`imputeHD`, [6](#), [7](#), [9](#), [14](#), [15](#)

`lung`, [11](#)

`plot.HDpath (HDpath)`, [7](#)

`plot.stars (stars)`, [13](#)

`print.GLMpath (GLMpath)`, [6](#)

`print.HDImputed (HDImputed)`, [6](#)

`print.HDpath (HDpath)`, [7](#)

`print.stars (stars)`, [13](#)

`RNAseqNetUsersGuide`, [11](#)

`stabilitySelection`, [5](#), [12](#), [13](#), [14](#)

`stars`, [13](#)

`stars-class (stars)`, [13](#)

`summary.GLMpath (GLMpath)`, [6](#)

`summary.HDImputed (HDImputed)`, [6](#)

`summary.HDImputed (HDImputed)`, [6](#)

`summary.HDpath (HDpath)`, [7](#)

`summary.stars (stars)`, [13](#)

`thyroid`, [14](#)

`varIntra`, [3](#), [14](#)