Package 'neatmaps'

July 22, 2025

Title Heatmaps for Multiple Network Data

Version 2.1.0

Author Philippe Boileau [aut, cre]

Maintainer Philippe Boileau <philippe_boileau@berkeley.edu>

Description Simplify the exploratory data analysis process for multiple network
data sets with the help of hierarchical clustering, consensus
clustering and heatmaps. Multiple network data consists of multiple
disjoint networks that have common variables (e.g. ego networks).
This package contains the necessary tools for exploring such data,
from the data pre-processing stage to the creation of dynamic
visualizations.

Depends R (>= 3.5.0)

Imports igraph, heatmaply, ConsensusClusterPlus, ggplot2, grDevices,
dplyr

License MIT + file LICENSE

URL <https://github.com/PhilBoileau/neatmaps>

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-05-12 19:10:03 UTC

Contents

---

aggNodeAttr   *Node Attribute Aggregater*

Description

aggNodeAttr creates a data frame that summarizes node attributes.

Usage

```
aggNodeAttr(node_df, measure_of_cent = "mean")
```

Arguments

node_df   A data frame containing all the characteristics of the nodes in the network. If there are n networks, a maximum of x nodes per network and y variables for each node, the data frame should have n rows and x*y columns. The column names of each variable should be written as follows: var1, var2, ... , varX.

measure_of_cent

    A vector that contains the measures of centrality with which to summarize the node attributes. The supported measures are "mean" and "median". Note that missing values are excluded from the calculations.

Author(s)

Philippe Boileau , <philippe_boileau@berkeley.edu>

---

calcICLNoPlots	*Consensus Cluster Plus without Plots*

Description

consensusClusterNoPlots is a wrapper function for `ConsensusClusterPlus` that suppresses the creation of the plots that are created automatically.

Usage

```
calcICLNoPlots(consensus_results)
```

Arguments

consensus_results

> Results of consensus clustering. The second item in the list returned by `neatmap`.

Author(s)

Philippe Boileau , <philippe_boileau@berkeley.edu>

---

consClustResTable	*Consensus Cluster Results in a Table*

Description

consClustResTable create a dataframe of the consensus cluster results. The dataframe presents the results of each iteration of the `ConsensusClusterPlus` algorithm, the cluster consensus of each cluster and the list of the cluster elements with their corresponding item consensus. The item consensus is taken with respect to the variable's cluster allocation.

Usage

```
consClustResTable(neatmap_res)
```

Arguments

neatmap_res   Output from the `neatmap` function.

Value

A dataframe of the results of the consensus clustering.

Author(s)

Philippe Boileau , <philippe_boileau@berkeley.edu>

References

For more information on the consensus cluster and item consensus statistics, see Monti et al..

Examples

```
# create the data frame using the network, node and edge attributes
df <- netsDataFrame(network_attr_df,
                     node_attr_df,
                     edge_df)

# run the neatmap code on df
neat_res <- neatmap(df, scale_df = "ecdf", max_k = 3, reps = 100,
                     xlab = "vars", ylab = "nets", xlab_cex = 1, ylab_cex = 1)

# get the consensus cluster results for each iteration
consensus_res_df <- consClustResTable(neat_res)
```

---

consensusChangeECDF   *Change in Area Under the ECDF*

Description

consensusChangeECDF plots the relative change in area under empirical cumulative distribution function for consecutive consensus cluster matrices produced using the neatmap function.

Usage

```
consensusChangeECDF(neatmap_res)
```

Arguments

neatmap_res   Output from the neatmap function.

Value

A ggplot of the change in consecutive area under the ECDFs of the consensus cluster matrices.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

References

For more information on the consensus matrices, see Monti et al..

Examples

```
#' # create the data frame using the network, node and edge attributes
df <- netsDataFrame(network_attr_df,
                    node_attr_df,
                    edge_df)

# run the neatmap code on df
neat_res <- neatmap(df, scale_df = "ecdf", max_k = 3, reps = 100,
                    xlab = "vars", ylab = "nets", xlab_cex = 1, ylab_cex = 1)

# visualize the relative change in AU ECDF of consecutive consensus cluster
# iterations
consensusChangeECDF(neat_res)
```

---

consensusClusterNoPlots

Consensus Cluster Plus without Plots

Description

consensusClusterNoPlots is a wrapper function for [ConsensusClusterPlus](#)that suppresses the creation of the plots that are created automatically.

Usage

```
consensusClusterNoPlots(df, link_method, dist_method, max_k, reps, p_var,
  p_net, cc_seed)
```

Arguments

df	A dataframe of network attributes containing only numeric values. The columns of the dataframe should likely be normalized.
link_method	The agglomeration method to be used for hierarchical clustering. Defaults to the average linkage method. See other methods in [hclust](#).
dist_method	The distance measure to be used between columns and between rows of the dataframe. Distance is used as a measure of similarity. Defaults to euclidean distance. See other options in [dist](#).
max_k	The maximum number of clusters to consider in the consensus clustering step. Consensus clustering will be performed for max_k-1 iterations, i.e. for 2, 3, ..., max_k clusters. Defaults to 10.
reps	The number of subsamples taken at each iteration of the consensus cluster algorithm. Defaults to 1000.
p_var	The proportion of network variables to be subsampled during consensus clustering. Defaults to 1.
p_net	The proportion of networks to be subsampled during consensus clustering. Defaults to 0.8.

cc_seed   The seed used to ensure the reproducibility of the consensus clustering. Defaults
                 to 1.

                 @author Philippe Boileau , <philippe_boileau@berkeley.edu>

                 @importFrom ConsensusClusterPlus ConsensusClusterPlus @importFrom grDe-
                 vices png dev.off

---

consensusECDF   *Consensus Matrix ECDFs*

Description

consensusECDF plots the empirical cumulative distribution functions (ECDF) of the consensus ma-
trices produced during the consensus clustering step of the [neatmap](#) function.

Usage

```
consensusECDF(neatmap_res)
```

Arguments

neatmap_res   Output from the [neatmap](#) function.

Details

This function visualizes the ECDFs of the consensus matrices for each each iteration of consensus
clustering that is carried out as part of the [neatmap](#) function.

Value

Returns a ggplot depicting the ECDFs of each iteration of the consensus clustering, i.e. one ECDF
per number of clusters used in each iteration.

Author(s)

Philippe Boileau , <philippe_boileau@berkeley.edu>

References

For more information on the consensus matrices, see [Monti et al.](#).

Examples

```
#' # create the data frame using the network, node and edge attributes
df <- netsDataFrame(network_attr_df,
                    node_attr_df,
                    edge_df)

# run the neatmap code on df
neat_res <- neatmap(df, scale_df = "ecdf", max_k = 3, reps = 100,
```

```
                    xlab = "vars", ylab = "nets", xlab_cex = 1, ylab_cex = 1)

# create the ECDF plot
consensusECDF(neat_res)
```

---

consensusMap   *Create Heatmaps of Consensus Matrices*

Description

consensusMap produces a list of heatmaps from the consensus matrices produced during the consensus clustering step of the [neatmap](#) function.

Usage

```
consensusMap(neatmap_res, link_method = "average")
```

Arguments

neatmap_res	Output from the [neatmap](#) function.
link_method	The agglomeration method to be used for hierarchical clustering. Defaults to the average linkage method. See other methods in [hclust](#).

Details

This function will create a list of heatmaps of the consensus matrices produced during the consensus clustering step of the [neatmap](#) function. The default clustering method used in the heatmaps is hierarchical clustering using the average linkage method, though other linkage methods can be used. The consensus cluster matrix is used as a measure of similarity. The heatmaps are produced using [heatmaply](#).

Value

Returns of a list of heatmaps depicting the consensus matrices of each

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

References

For more information on the consensus matrices, see [Monti et al.](#).

Examples

```
# create the data frame using the network, node and edge attributes
df <- netsDataFrame(network_attr_df,
                    node_attr_df,
                    edge_df)

# run the neatmap code on df
neat_res <- neatmap(df, scale_df = "ecdf", max_k = 3, reps = 100,
                    xlab = "vars", ylab = "nets", xlab_cex = 1, ylab_cex = 1)

# create the list of heatmaps for each iteration
hm_list <- consensusMap(neat_res)
```

---

createNetworks   *Create Networks Using Edge Data Frame*

Description

createNetworks creates Igraph network objects using an edge data frame. This is important for computing structural properties of the networks to be explored by neatmap.

Usage

```
createNetworks(edge_df)
```

Arguments

edge_df   A data frame where each row represents a different network and where each
                column represents a potential edge between node A and node B. The column
                names should be of the form "XA_B", where A and B are the node numbers in
                the network. If Node A or B do not exist in the specific network, the cell should
                have a value of NA. If there is no edge between A and B, place a value of 0.
                Avoid redundant column names since all edges are assumed to be undirected,
                e.g. avoid "XA_B" and "XB_A".

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

edge_df   *Edge List Data Frame*

Description

A dataset containing a list of undirected edges for ten different networks. Each network has a maximum size of 5 nodes. The network and node attribute data are saved in their respective files.

Usage

```
edge_df
```

Format

An object of class data.frame with 10 rows and 10 columns.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

getStructureAttr   *Structural Attributes of Networks Data Frame*

Description

getStructureAttr produces a data frame of the structural attributes of a list of networks.

Usage

```
getStructureAttr(net_list)
```

Arguments

net_list   A list of Igraph network objects that represent the collection of networks.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

hierarchy   *Hierarchy*

Description

hierarchy calculates the hierarchy of a network

Usage

```
hierarchy(net)
```

Arguments

net   An igraph object representing a network

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

neatmap   *Explore Multi-Network Data*

Description

neatmap produces a heatmap of multi-network data and identifies stable clusters in its variables.

Usage

```
neatmap(df, scale_df, link_method = "average",
  dist_method = "euclidean", max_k = 10, reps = 1000, p_var = 1,
  p_net = 0.8, cc_seed = 100, main_title = "", xlab, ylab,
  xlab_cex = 1, ylab_cex = 1, heatmap_margins = c(50, 50, 50, 100))
```

Arguments

df   a dataframe of network attributes containing only numeric values.

scale_df   A string indicating whether the columns of the data frame should be scaled, and,
                    if so, which method should be used. The options are "none", "ecdf", "normalize"
                    and "percentize". If "none" is selected, then the columns are not scaled. If "ecdf"
                    is selected, then the columns are transformed into their empirical cumulative
                    distribution. If "normalize" is selected, each column is centered to have a mean
                    of 0 and scaled to have a standard deviation of 1. If "percentize" is selected,
                    column values are transformed into percentiles.

link_method   The agglomeration method to be used for hierarchical clustering. Defaults to the
                    average linkage method. See other methods in `hclust`.

dist_method	The distance measure to be used between columns and between rows of the dataframe. Distance is used as a measure of similarity. Defaults to euclidean distance. See other options in [dist]().
max_k	The maximum number of clusters to consider in the consensus clustering step. Consensus clustering will be performed for max_k-1 iterations, i.e. for 2, 3, ..., max_k clusters. Defaults to 10.
reps	The number of subsamples taken at each iteration of the consensus cluster algorithm. Defaults to 1000.
p_var	The proportion of network variables to be subsampled during consensus clustering. Defaults to 1.
p_net	The proportion of networks to be subsampled during consensus clustering. Defaults to 0.8.
cc_seed	The seed used to ensure the reproducibility of the consensus clustering. Defaults to 1.
main_title	The title of the heatmap.
xlab	The x axis label of the heatmap.
ylab	The y axis label of the heatmap.
xlab_cex	The font size of the elements on the x axis.
ylab_cex	The font size of the elements on the y axis.
heatmap_margins	
	The size of the margins for the heatmap. See [heatmaply]().

Details

This function allows users to efficiently explore their multi-network data by visualizing their data with a heatmap and assessing the stability of the associations presented within it. neatmap requires that the data frame be processed into an appropriate format prior to use. Data is then scaled (if necessary) using of the built in methods. See (list functions) for further details on how to prepare multi-network data for use with neatmap. The heatmap is created using [heatmaply]() and the consensus clustering is performed using [ConsensusClusterPlus]()

Value

A named list containing the heatmap of the multi-network data and a list of length max_k-1 where each element is a list containing the consensus matrix, the consensus hierarchical clustering results and the consensus class assignments. The list of results produced by the consensus clustering can be parsed using following functions in the [neatmaps]() package: [consClustResTable](), [consensusECDF]() and [consensusChangeECDF]().

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

References

For more information on the consensus clustering, see [Monti et al.]().

Examples

```
# create the data frame using the network, node and edge attributes
df <- netsDataFrame(network_attr_df,
                    node_attr_df,
                    edge_df)

# run the neatmap code on df
neat_res <- neatmap(df, scale_df = "ecdf", max_k = 3, reps = 100,
                    xlab = "vars", ylab = "nets", xlab_cex = 1, ylab_cex = 1)

# extract the heatmap
heatmap <- neat_res$heatmap

# extract the consensus clustering results
consensus_res <- neat_res$consensus_clust
```

---

neatmaps   neatmaps *package*

Description

A package for exploring multi-network data.

Details

See the README on [CRAN](#) or [GitHub](#)

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

netsDataFrame   *Networks Data Frame*

Description

netsDataFrame produces data frames of collections of networks.

Usage

```
netsDataFrame(net_attr_df, node_attr_df, edge_df,
  cent_measure = c("mean"))
```

Arguments

`net_attr_df`	A data frame consisting of all of the networks' graph attributes. The first column should contain the name of the network, and all other columns should be numeric. All empty entries should be filled as "NA".
`node_attr_df`	A data frame consisting of all of the networks' nodes' attributes. All columns should be numeric. All empty entries should be filled in as "NA".
`edge_df`	A data frame consisting of the edge matrix for each ego network. Edges are assumed to be undirected and unweighted. 1 indicates the existence of an edge between nodes, 0 indicates the lack of an edge.
`cent_measure`	A vector of the measures of centrality to be used for the summary of the node attributes data. The supported measures of centrality are: "mean" and "median".

Details

The function produces data frames of collections of networks. The function requires the input of three data frames: a data frame containing the graph attributes, a data frame containing the node characteristics and a data frame containing the edge list of each network. The rows in each of these data frames must represent individual networks, and must therefore have identical row length. Measures of centrality used in the summarization of the node attributes must also be furnished.

Value

The function returns a data frame that offers an overview of all of the ego networks.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

Examples

```
df <- netsDataFrame(network_attr_df,
                    node_attr_df,
                    edge_df)
```

`network_attr_df`	*Network Attributes Data*

Description

A data set containing four randomly generated variables used to mimic the network attributes of ten different networks. Attribute 1 and 2 have a correlation of 0.41, 1 and 3 have a correlation of 0.91, 1 and 4 have a correlation of 0.34, 2 and 3 have a correlation of 0.07, 2 and 4 have a correlation of 0.17 and 3 and 4 have a correlation of 0.32.

Usage

```
network_attr_df
```

Format

An object of class `data.frame` with 10 rows and 4 columns.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

node_attr_df   *Node Attribute Data*

Description

A dataset containing randomly generated node attributes for each node in each of the ten networks. Attributes A and B have a correlation of roughly 0.8, A and C have a correlation of roughly -0.2, B and C have a correlation of roughly 0.5. Attributes D and E were generated completely randomly, and should not be strongly correlated with any of the other attributes.

Usage

node_attr_df

Format

An object of class `data.frame` with 10 rows and 25 columns.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

---

scaleColumns   *Scale Between 0 and 1*

Description

`scaleColumns` scales the columns of a data frame object between the values of 0 and 1 without changing the underlying distribution of the columns.

Usage

scaleColumns(df)

Arguments

df   The data frame of numerical values to be scaled.

Author(s)

Philippe Boileau, <philippe_boileau@berkeley.edu>

Index