Package ‘nda’

October 14, 2022

Type Package
Title Network-Based Dimensionality Reduction and Analysis
Version 0.1.6
Maintainer Zsolt T. Kosztyan <kosztyan.zsolt@gtk.uni-pannon.hu>
License GPL (>= 2)
Encoding UTF-8
LazyData true
URL https://github.com/kzst/nda
Depends R (>= 3.10)
Imports leidenAlg, energy, psych, stats, igraph, Matrix, visNetwork
RoxygenNote 7.2.0
NeedsCompilation no
Author Zsolt T. Kosztyan [aut, cre],
Marcell T. Kurbucz [aut],
Attila I. Katona [aut]
Repository CRAN
Date/Publication 2022-10-14 13:10:02 UTC

R topics documented:

biplot.nda ......................................................... 2
COVID19_2020 .................................................. 3
CrimesUSA1990.X .............................................. 3
CrimesUSA1990.Y .............................................. 4
CWTS_2020 ......................................................... 5
data_gen .......................................................... 5
dCor ............................................................... 6
GOVDB2020 ....................................................... 7
Biplot function for Network-based Dimensionality Reduction and Analysis (NDA)

## S3 method for class 'nda'

```r
biplot(x, main=NULL, ...)
```

### Arguments

- `x` an object of class `NDA`.
- `main` main title of biplot.
- `...` other graphical parameters.

### Value

No return value

### Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona  
e-mail*: kzst@gtk.uni-pannon.hu

### See Also

`plot`, `summary.ndr`.

### Examples

```r
# Biplot function without feature selection

data("CrimesUSA1990.X")
df<-CrimesUSA1990.X
p<-ndr(df)
biplot(p)
```
# Biplot function with feature selection
# minimal eigen values (min_evalue) is 0.0065
# minimal communality value (min_communality) is 0.1
# minimal common communality value (com_communalities) is 0.1

p<-ndr(df,min_evalue = 0.0065,min_communality = 0.1,com_communalities = 0.1)
biplot(p)

---

COVID19_2020

Covid'19 of countries (2020), where the data frame has 138 observations of 18 variables.

Description

Sample datasets for Network-based Dimensionality Reduction and Analysis (NDA)
Covid’19 of countries (2020), where the data frame has 138 observations of 18 variables.

Usage

data("COVID19_2020")

Format

A data frame with 138 observations 18 variables.

Source


Examples

data(COVID19_2020)

---

CrimesUSA1990.X

Crimes in USA cities in 1990. Independent variables (X)

Description

Sample datasets for Network-based, non-parametric dimensionality reduction and analysis (NDA)
Crimes in USA cities in 1990. Independent variables (X)

Usage

data("CrimesUSA1990.X")
Format

A data frame with 1994 observations 123 variables.

Source


Examples

data(CrimesUSA1990.X)

data("CrimesUSA1990.Y")

Description

Sample datasets for Network-based Dimensionality Reduction and Analysis (NDA)

Crimes in USA cities in 1990. Dependent variable (Y)

Usage

data("CrimesUSA1990.Y")

Format

A data frame with 1994 observations 1 variables.

Source


Examples

data(CrimesUSA1990.Y)
CWTS Leiden’s 2020 dataset, where the data frame has 1176 observations of 42 variables.

Description

Sample datasets for Network-based Dimensionality Reduction and Analysis (NDA)

CWTS Leiden’s 2020 dataset, where the data frame has 1176 observations of 42 variables.

Usage

data("CWTS_2020")

Format

A data frame with 1176 observations of 42 variables.

Source

CWTS Leiden Ranking 2020: https://www.leidenranking.com/ranking/2020/list

Examples

data(CWTS_2020)

Generate random block matrix for NDA

data_gen

Description

Generate random block matrix for Network-based Dimensionality Reduction and Analysis (NDA)

Usage

data_gen(n,m,nfactors=2,lambda=1)

Arguments

- **n**: number of rows
- **m**: number of columns
- **nfactors**: number of blocks (factors, where the default value is 2)
- **lambda**: exponential smoothing, where the default value is 1
Details

n, m, nfactors must be integers, and they are not less than 1; lambda should be a positive real number.

Value

M

a dataframe of a block matrix

Author(s)

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary
e-mail: kzst@gtk.uni-pannon.hu

Examples

# Specification 30 by 10 random block matrices with 2 blocks/factors
df<-data_gen(30,10)
library(psych)
screen(df)
biplot(ndr(df))
# Specification 40 by 20 random block matrices with 3 blocks/factors
df<-data_gen(40,20,3)
library(psych)
screen(df)
biplot(ndr(df))
plot(ndr(df))

# Specification 50 by 20 random block matrices with 4 blocks/factors
# lambda=0.1
df<-data_gen(50,15,4,0.1)
screen(df)
biplot(ndr(df))
plot(ndr(df))

dCor Calculating distance correlation of two vectors or columns of a matrix

Description

Calculating distance correlation of two vectors or columns of a matrix for Network-based Dimensionality Reduction and Analysis (NDA).

The calculation is very slow for large matrices!

Usage

dCor(x,y=NULL)
Arguments

- **x**: a numeric vector, a numeric matrix (in this case `y=NULL`), or a numeric data frame (in this case `y=NULL`)
- **y**: a numeric vector (optional)

Details

If `x` is a numeric vector, `y` must be specified. If `x` is a numeric matrix or numeric data frame, `y` must be ignored from the parameters.

Value

Either a distance correlation value of vectors `x` and `y`, or a distance correlation matrix of `x`.

Author(s)

Prof. Zsolt T. Kosztyan, Department of Quantitative Methods, Institute of Management, Faculty of Business and Economics, University of Pannonia, Hungary

e-mail: kzst@gtk.uni-pannon.hu

References


Examples

```r
# Specification of distance correlation value of vectors x and y.
x<-rnorm(36)
y<-rnorm(36)
dCor(x,y)
# Specification of distance correlation matrix.
x<-matrix(rnorm(36),nrow=6)
dCor(x)
```

GOVDB2020

Governmental and economic data of countries (2020), where the data frame has 138 observations of 2161 variables.

Description

Sample datasets for Network-based Dimensionality Reduction and Analysis (NDA)

Governmental and economic data of countries (2020), where the data frame has 138 observations of 2161 variables.

Usage

```r
data("GOVDB2020")
```
**Format**

A data frame with 138 observations of 2161 variables.

**Source**


**Examples**

```r
data(GOVDB2020)
```

---

<table>
<thead>
<tr>
<th>I40_2020</th>
<th>NUTS2 regional development data (2020), where the data frame has 414 observations of 101 variables.</th>
</tr>
</thead>
</table>

**Description**

Sample datasets for Network-based, non-parametric dimensionality reduction and analysis (NDA) NUTS2 regional development data (2020), where the data frame has 414 observations of 101 variables.

**Usage**

```r
data("COVID19_2020")
```

**Format**

A data frame with 414 observations of 101 variables.

**Source**


**Examples**

```r
data(I40_2020)
```
Package of Network-based Dimensionality Reduction and Analyses

Description

Network-based Dimensionality Reduction and Analysis. Biplot function.

Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona

e-mail*: kzst@gtk.uni-pannon.hu

References


See Also

ndr, plot, biplot, summary, dCor.

Network-based Dimensionality Reduction and Analysis (NDA)

Description

Network-based Dimensionality Reduction and Analysis (NDA).

Usage

```r
ndr(data, cor_method=1, min_R=0, min_comm=2, Gamma=1, null_modell_type=4, mod_mode=6, min_evalue=0, min_communality=0, com_communalities=0, use_rotation)
```

Arguments

data: A numeric data frame

cor_method: Correlation method (optional). ‘1’ Pearson’s correlation (default), ‘2’ Spearman’s correlation, ‘3’ Kendall’s correlation, ‘4’ Distance correlation

min_R: Minimal square correlation between indicators (default: 0)

min_comm: Minimal number of indicators in a community (default: 2)

Gamma: Gamma parameter in multiresolution null_modell (default: 1)
null_modell_type
'1' Differencial Newmann-Grivan’s null model, '2' The null model is the mean of square correlations between indicators, '3' The null model is the specified minimal square correlation, '4' Newmann-Grivan’s model (default)

mod_mode
'1' Louvain modularity, '2' Fast-greedy modularity, '3' Leading Eigen modularity, '4' Infomap modularity, '5' Walktrap modularity, '6' Leiden modularity (default)

min_evalue
Minimal eigenvector centrality value (default: 0)

min_communality
Minimal communality value of indicators (default: 0)

com_communalities
Minimal common communalities (default: 0)

use_rotation
FALSE no rotation (default), TRUE varimax rotation

Details
NDA both works on low and high simple size datasets. If min_evalue=min_communality=com_communalities=0 than there is no feature selection.

Value

communality
Communality estimates for each item. These are merely the sum of squared factor loadings for that item.

loadings
A standard loading matrix of class “loadings”

uniqueness
Uniqueness value of indicators.

factors
Number of found factors

scores
Estimates of the factor scores are reported

n.obs
Number of observations specified or found

fn
NDA

Author(s)
Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona

e-mail*: kzst@gtk.uni-pannon.hu

See Also
plot, biplot, summary.

Examples

data(swiss)
df<-swiss
p<-ndr(df)
summary(p)
plot(p)
plot.nda

```r
biplot(p)
```

---

**Description**

Plot variable network graph

**Usage**

```r
## S3 method for class 'nda'
plot(x, cuts=0.3,...)
```

**Arguments**

- `x`: an object of class 'NDA'.
- `cuts`: minimal correlation square value for an edge in the variable network graph.
- `...`: other graphical parameters.

**Value**

No return value

**Author(s)**

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona
e-mail*: kzst@gtk.uni-pannon.hu

**See Also**

`biplot`, `summary.ndr`.

**Examples**

```r
# Plot function with feature selection

data("CrimesUSA1990.X")
df<-CrimesUSA1990.X
p<-ndr(df)
biplot(p,main="Biplot of CrimesUSA1990 without feature selection")

# Plot function with feature selection
# minimal eigen values (min_evalue) is 0.0065
# minimal communality value (min_communality) is 0.1
# minimal common communality value (com_communalities) is 0.1
```
p<-ndr(df,min_evalue = 0.0065,min_communality = 0.1,com_communalities = 0.1)

# Plot with default (cuts=0.3)
plot(p)

# Plot with higher cuts
plot(p,cuts=0.6)

summary.nda

Summary function of Network-based Dimensionality Reduction and Analysis (NDA)

Description

Print summary of Network-based, non-parametric dimensionality reduction and analysis (NDA)

Usage

## S3 method for class 'nda'
summary(object, digits = getOption("digits"), ...)

Arguments

object an object of class 'nda'.
digits the number of significant digits to use when add.stats = TRUE.
... additional arguments affecting the summary produced.

Value

No return value

Author(s)

Zsolt T. Kosztyan*, Marcell T. Kurbucz, Attila I. Katona
e-mail*: kzst@gtk.uni-pannon.hu

See Also

biplot, plot, ndr.
Examples

# Example of summary function of NDA without feature selection

data("CrimesUSA1990.X")
df<-CrimesUSA1990.X
p<-ndr(df)
summary(p)

# Example of summary function of NDA with feature selection
# minimal eigen values (min_evalue) is 0.0065
# minimal communality value (min_communality) is 0.1
# minimal common communality value (com_communalities) is 0.1

df<-CrimesUSA1990.X
p<-ndr(df,min_evalue = 0.0065,min_communality = 0.1,com_communalities = 0.1)
summary(p)
Index

* array
  data_gen, 5
dCor, 6
* control chart
  plot.nda, 11
* correlation matrix
dCor, 6
* datasets
  COVID19_2020, 3
  CrimesUSA1990.X, 3
  CrimesUSA1990.Y, 4
  CWTS_2020, 5
  GOVDB2020, 7
  I40_2020, 8
* dimensionality
  ndr, 9
* distance correlation
dCor, 6
* multivariate
  data_gen, 5
dCor, 6
  ndr, 9
  plot.nda, 11
  summary.nda, 12
* nonparametric
  ndr, 9
* package
  nda, 9
* plot
  biplot.nda, 2
* random block matrix
  data_gen, 5
* reduction
  ndr, 9
  biplot, 9–12
  biplot.nda, 2
  COVID19_2020, 3
  CrimesUSA1990.X, 3
  CrimesUSA1990.Y, 4
  CWTS_2020, 5
  data_gen, 5
dCor, 6, 9
  GOVDB2020, 7
  I40_2020, 8
  nda, 9
  ndr, 2, 9, 9, 11, 12
  plot, 2, 9, 10, 12
  plot.nda, 11
  summary, 2, 9–11
  summary.nda, 12