# 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>

**Description** Non-parametric dimensionality reduction function. Reduction with and without feature selection. Plot functions. Kosztyan et. al. (2022) <doi:10.1016/j.knosys.2022.109180>.

**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:

**Index**                                                                                                            **14**

---

biplot.nda                          *Biplot function for Network-based Dimensionality Reduction and*
                                    *Analysis (NDA)*

---

### Description

Biplot function for Network-based Dimensionality Reduction and Analysis (NDA)

### Usage

```
## S3 method for class 'nda'
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

```
# 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

Kurbucz, M. T. (2020). A joint dataset of official COVID-19 reports and the governance, trade and competitiveness indicators of World Bank group platforms. Data in brief, 31, 105881.

## 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**

UCI - Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/communities+and+crime

**Examples**

```
data(CrimesUSA1990.X)
```

---

CrimesUSA1990.Y        *Crimes in USA cities in 1990. Dependent variable (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**

UCI - Machine Learning Repository: https://archive.ics.uci.edu/ml/datasets/communities+and+crime

**Examples**

```
data(CrimesUSA1990.Y)
```

| CWTS_2020 | *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)
```

| data_gen | *Generate random block matrix for NDA* |
| --- | --- |

### 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 | exponental smoothing, where the default value is 1 |

## Details

n, m, nfactors must beintegers, 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)
scree(df)
biplot(ndr(df))
# Specification 40 by 20 random block matrices with 3 blocks/factors
df<-data_gen(40,20,3)
library(psych)
scree(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)
scree(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

Rizzo M, Szekely G (2021). _energy: E-Statistics: Multivariate Inference via the Energy of Data_.
R package version 1.7-8, <URL: https://CRAN.R-project.org/package=energy>.

## Examples

```
# Specification of distance correlation value of vectors x and y.
x<-rnorm(36)
y<-rnorm(36)
dCor(x,y)
# Specification of distance correlaction 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

```
data("GOVDB2020")
```

## Format

A data frame with 138 observations of 2161 variables.

## Source

Kurbucz, M. T. (2020). A joint dataset of official COVID-19 reports and the governance, trade and competitiveness indicators of World Bank group platforms. Data in brief, 31, 105881.

## Examples

```
data(GOVDB2020)
```

---

| I40_2020 | _NUTS2 regional development data (2020), where the data frame has 414 observations of 101 variables._ |
|---|---|

---

## 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

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

## Format

A data frame with 414 observations of 101 variables.

## Source

Honti, G., Czvetkó, T., & Abonyi, J. (2020). Data describing the regional Industry 4.0 readiness index. Data in Brief, 33, 106464.

## Examples

```
data(I40_2020)
```

---

nda | *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

Kurbucz, M. T., Katona, A. I., Lantos, Z., & Kosztyan, Z. T. (2021). The Role of Societal Aspects in the Formation of Official COVID-19 Reports: A Data-Driven Analysis. International journal of environmental research and public health, 18(4), 1505.

### See Also

ndr, plot, biplot, summary, dCor.

---

ndr | *Network-based Dimensionality Reduction and Analysis (NDA)*

---

### Description

Network-based Dimensionality Reduction and Analysis (NDA).

### Usage

```
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 modell, '2' The null model is the mean of square correlations between indicators, '3' The null modell is the specified minimal square correlation, '4' Newmann-Grivan's modell (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)
```

```
biplot(p)
```

---

| plot.nda | *Plot function for Network-based Dimensionality Reduction and Analysis (NDA)* |
| --- | --- |

---

### Description

Plot variable network graph

### Usage

```
## 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

```
# 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

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

# Index