

# Package ‘DCA’

April 11, 2017

**Type** Package

**Title** Dynamic Correlation Analysis for High Dimensional Data

**Version** 1.0

**Date** 2017-04-07

**Author** Tianwei Yu <tianwei.yu@emory.edu>

**Maintainer** Tianwei Yu <tianwei.yu@emory.edu>

**Description**

Finding dominant latent signals that regulate dynamic correlation between many pairs of variables.

**Imports** methods, GPArotation, PMA, coin, elasticnet, locfdr, modeest,  
mvtnorm,

**License** GPL-2

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-04-11 21:04:59 UTC

## R topics documented:

DCA-package . . . . .	2
dca . . . . .	2
find.xy . . . . .	4
la.simu.gen . . . . .	5
plot_la . . . . .	6
<b>Index</b>	<b>8</b>

DCA-package

*Dynamic Correlation Analysis for high dimensional data***Description**

Given a data matrix with variables in rows and samples in the columns, the method DCA finds dominant latent signals that regulate the dynamic correlation between many pairs of variables.

**Details**

The subroutine `dca()` computes dynamic correlation signals from the data matrix. It can use PCA, SPCA, and kmeans clustering to find dominant signals. The subroutine `find.xy()` subsequently finds variable pairs that are associated with each latent signal. The subroutine `plot.la()` plots the dynamic correlation of two variables X and Y given the Z vector.

**Author(s)**

Tianwei Yu <tianwei.yu@emory.edu>

**References**

Li, K.C. (2002) Genome-wide coexpression dynamics: theory and application, Proceedings of the National Academy of Sciences of the United States of America, 99, 16875-16880.

dca

*Dynamic Correlation Analysis***Description**

The method finds a series of latent vectors, which serve as the LA scouting vectors for large numbers of variable pairs.

**Usage**

```
dca(array, top.pairs.prop = 0.95, max.pairs = 1e+06, n.fac = 10,
sumabsv = sqrt(max.pairs)/10, normalization = "standardize", method = "PCA")
```

**Arguments**

<code>array</code>	The data matrix, with variables in the rows and samples in the columns.
<code>top.pairs.prop</code>	The method ranks all variable pairs from the most likely to have dynamic correlation relationship to the least likely. The top pairs are used for detection of latent signals. This parameter controls the percentage of pairs used in the computation.

max.pairs	The maximum number of pairs to use. When the data contains too many variables, such as tens of thousands of variables in a gene expression matrix, this parameter limits the maximum number of variable pairs to enter the calculation.
n.fac	The number of top latent factors to report. If the method "kmeans" is used, this parameter is used as the number of clusters.
sumabsv	The sumabsv parameter to be passed on to the SPC() method.
normalization	The way the data matrix is to be row-normalized. The method requires each row to have mean 0 and SD 1. There are two options, "standardize", or "normal score".
method	The method for finding the latent factors. Current choices are "PCA", "SPCA", and "kmeans".

### Details

After finding the factors, the method attempts to rotate the factor using oblique rotation to achieve more interpretable results.

### Value

The method returns a list.

fac	The original factors found. This is the PC, SPC, or cluster mean vector depending on the method chosen.
rotated	The factors after rotation.
ss.proj	The sum of squared attributed to each rotated factor.

### Author(s)

Tianwei Yu <tianwei.yu@emory.edu>

### See Also

find.xy()

### Examples

```
x<-la.simu.gen(n=100,p=200,n.grp=2, n.noise.gene=100, rho=0.5, pwr=0.5)
z<-dca(x$dat, n.fac=2)
cor(z[[2]], x$z, method="spearman")
```

---

find.xy

*Find variable pairs for a given set of LA scouting vectors.*


---

## Description

After finding the latent vectors, this function can be used to find the pairs of variables (rows of the original data matrix) that are associated with each latent vector in terms of dynamic correlation.

## Usage

```
find.xy(array, z, fdr.cut = 0.05, normalization = "standardize",
center.z = FALSE, low.cor.percentile = c(0.2, 0.8))
```

## Arguments

array	The data matrix with variables in the rows and samples in the columns.
z	The matrix of latent variables. Each column is a latent vector.
fdr.cut	The threshold of local fdr for the selection of variable pairs.
normalization	The way the data matrix is to be row-normalized. The method requires each row to have mean 0 and SD 1. There are two options, "standardize", or "normal score".
center.z	Whether to remove mean from each z vector.
low.cor.percentile	The variable pairs that are considered low correlation will enter the computation. This is determined by the percentile of the correlation among all variable pairs. If the percentile is within the provided range, then they are considered lowly correlated.

## Value

A list is returned. Each numbered item is a matrix with three columns: row number 1, row number 2, and local fdr value. Only those selected by fdr threshold are returned.

## Author(s)

Tianwei Yu <tianwei.yu@emory.edu>

## See Also

dca()

**Examples**

```
x<-la.simu.gen(n=100,p=80,n.grp=2, n.noise.gene=100, rho=0.5, pwr=0.25)
z<-dca(x$dat, n.fac=2)
xy<-find.xy(x$dat, z[[2]], fdr.cut=0.01)
summary(xy)
xy[[1]][1:5,]
xy[[2]][1:5,]
```

---

la.simu.gen	<i>Simulate a data matrix with underlying dynamic correlation signal in the Liquid Association (LA) framework</i>
-------------	---

---

**Description**

The simulation follows the LA framework, namely dynamic correlation is in the form of  $X \sim N(0,1)$ ,  $Y \sim N(0,1)$ ,  $Z \sim N(0,1)$ ,  $E(XY|Z)$  is a function of  $Z$ .

**Usage**

```
la.simu.gen(n, p, n.grp, n.noise.gene, rho, pwr)
```

**Arguments**

n	Sample size (number of columns of the data matrix).
p	The number of genes in each LA module, i.e. a group of genes regulated by the same latent dynamic correlation factor.
n.grp	The number of LA modules to simulate.
n.noise.gene	The number of pure noise genes to add to the matrix.
rho	The standard deviation of the Gaussian noise to be added to the simulated data in the modules.
pwr	The power for the transformation (see details)

**Details**

Between modules, the latent LA factor  $z$ 's are independent.

Within each module, 10 sub-modules are simulated. For each sub-module, we first generate a pair of  $X$  and  $Y$  vectors, which follows:

$X \sim N(0,1)$ ,  $Y \sim N(0,1)$   $u = (\text{pnorm}(z) - 0.5) * 2 \ E(XY|z) = \text{sign}(u) * \text{abs}(u)^{\text{pwr}}$

Then white noise with SD of  $\rho$  is added to the hidden  $X$ ,  $Y$  pair to generate pairs of observed  $X$ ,  $Y$  vectors.

**Value**

A list is returned.

dat	The data matrix.
z	The true $z$ vectors.

**Author(s)**

Tianwei Yu <tianwei.yu@emory.edu>

**See Also**

dca()

**Examples**

```
x<-la.simu.gen(n=100,p=200,n.grp=3, n.noise.gene=100, rho=0.5, pwr=1)
x$dat[1:5,1:5]
x$z[1:5,]
```

---

plot_la	<i>Visualizing the dynamic correlation between a pair of genes given the AL scouting vector.</i>
---------	--

---

**Description**

Given three vectors, x, y and z, the function produces a color scatter plot of x and y, colored by the grouping of z.

**Usage**

```
plot_la(x, y, z, use.locfdr = FALSE, cols = c("red", "green", "blue"), cex = 0.5)
```

**Arguments**

x	The x vector.
y	The y vector.
z	The z vector, conditioned on which X and Y have dynamic correlation.
use.locfdr	Whether to use local fdr to group the z values. If TRUE, locfdr() is used and the fdr (posterior probability of belonging to the more extreme groups) threshold is 0.5. If FALSE, the z values are cut at the 0.33 and 0.67 quantiles into three groups.
cols	The colors of the groups.
cex	The point size to be passed to plot().

**Details**

The locfdr approach is only to be used when it is clear z has heavy tails, and it is believed such points are driving the dynamic correlation.

**Value**

A vector, each item is a string, recording the LA score and the correlation in each of the point groups as determined by z values.

*plot\_la*

7

**Author(s)**

Tianwei Yu <tianwei.yu@emory.edu>

**Examples**

```
x<-la.simu.gen(n=100,p=20,n.grp=2, n.noise.gene=10, rho=0.25, pwr=1)
plot_la(x$dat[1,], x$dat[2,], x$z[,1],use.locfdr=FALSE)
```

# Index

\*Topic **package**

DCA-package, [2](#)

DCA (DCA-package), [2](#)

dca, [2](#)

DCA-package, [2](#)

find.xy, [4](#)

la.simu.gen, [5](#)

plot\_la, [6](#)