Package ‘dSimer’

March 22, 2017

**Type** Package

**Title** Integration of Disease Similarity Methods

**Version** 1.0.0

**Date** 2015-12-10

**Author** Min Li <limin@mail.csu.edu.cn>, Peng Ni <nipeng@csu.edu.cn>

with contributions from Zhihui Fei and Ping Huang.

**Maintainer** Peng Ni <nipeng@csu.edu.cn>

**Description** dSimer is an R package which provides computation of nine methods for measuring disease-disease similarity, including a standard cosine similarity measure and eight function-based methods. The disease similarity matrix obtained from these nine methods can be visualized through heatmap and network. Biological data widely used in disease-disease associations study are also provided by dSimer.

**Depends** R (>= 3.3.0), igraph (>= 1.0.1)

**Imports** stats, Rcpp (>= 0.11.3), ggplot2, reshape2, GO.db, org.Hs.eg.db, AnnotationDbi, graphics

**Suggests** knitr, rmarkdown, BiocStyle

**LinkingTo** Rcpp

**License** GPL (>= 2)

**biocViews** Software, Visualization, Network

**VignetteBuilder** knitr

**RoxygenNote** 5.0.1

**NeedsCompilation** yes

**R topics documented:**

- dSimer-package ........................................... 2
- BOG ..................................................... 3
- CosineDFV .............................................. 4
- d2go_sample .......................................... 5
- d2g_fundo_entrezid .................................. 5
- d2g_fundo_symbol ................................... 6
- d2g_separation ...................................... 6
- d2s_hsdn ............................................. 7
dSimer-package

Integration of Disease Similarity Methods

Description

dSimer is an R package which provides computation of nine methods for measuring disease-disease similarity, including a standard cosine similarity measure and eight function-based methods. The disease similarity matrix obtained from these nine methods can be visualized through heatmap and network. Biological data widely used in disease-disease associations study are also provided by dSimer.

Details

Package: dSimer
Type: Package
Version: 0.99.6
Date: 12-10-2015
biocViews: Software, Visualization, Network
Depends: R (>= 3.3.0), igraph (>= 1.0.1)
BOG

Imports: stats, Rcpp (>= 0.11.3), ggplot2, reshape2, GO.db, AnnotationDbi, org.Hs.eg.db, graphics
Suggests: knitr, rmarkdown, BiocStyle
LinkingTo: Rcpp
License: GPL (>= 2)

Author(s)

Min Li, Peng Ni

BOG  calculate disease similarity by BOG

Description

given two vectors of diseases and a list of disease-gene associations, this function will calculate disease similarity by method BOG.

Usage

BOG(D1, D2, d2g)

Arguments

D1 a vector consists disease ids
D2 another vector consists disease ids
d2g a list of disease-gene associations

Value

a matrix of disease disease similairity which rownames is D1 and colnames is D2

Author(s)

Peng Ni, Min Li

References


See Also

Normalize

Examples

data(d2g_separation) #get disease-gene associations
ds<-sample(names(d2g_separation),5)
sim<-BOG(ds,ds,d2g_separation)
Normalize(sim) #normalize BOG sim scores
CosineDFV  
*calculate disease similarity by using feature vectors*

**Description**

given two (lists of) disease names, this function will calculate cosine similarity between these diseases’ feature vectors.

**Usage**

```r
CosineDFV(D1, D2, d2f, dcol = 2, fcol = 1, ccol = 3)
```

**Arguments**

- **D1**: a vector consists of disease ids/names
- **D2**: another vector consists of disease ids/names
- **d2f**: data.frame, contains term co-occurrences between features and diseases
- **dcol**: integer, disease column number in d2f
- **fcol**: integer, feature column number in d2f
- **ccol**: integer, co-occurrences column number in d2f

**Value**

a matrix of disease disease similarity which rownames and colnames are the disease names

**Author(s)**

Zhihui Fei, Peng Ni, Min Li

**References**


**Examples**

```r
### this is a disease-symptom-cooccurrence sample, if you want to use
### the complete data, please use "data(d2s_hsdn)" command
data(d2s_hsdn_sample)

ds <- sample(unique(d2s_hsdn_sample[,2]), 10)
simmat <- CosineDFV(ds, ds, d2s_hsdn_sample)
```
**d2go_sample**

**Description**

A sample list of disease-GO term associations.

**Value**

d2go_sample is a named list of length 3. The names are the DOIDs (DOIDs are ids of terms in Disease Ontology, e.g. "DOID:4") and list elements are vectors of GO term ids. The entire data of disease-GO term associations can be obtained by function HypergeometricTest.

**See Also**

HypergeometricTest

**Examples**

```r
data(d2go_sample)
```

---

**d2g_fundo_entrezid**

**Description**

A list of disease-gene associations from FunDO.

**Value**

d2g_fundo_entrezid is a named list of length 1855 which stored disease-gene associations from FunDO. The names are the DOIDs (DOIDs are ids of terms in Disease Ontology, e.g. "DOID:4") and list elements are vectors of Entrez gene IDs.

**References**


**Examples**

```r
data(d2g_fundo_entrezid)
```
Description

A list of disease-gene associations from FunDO.

Value

d2g_fundo_symbol is a named list of length 1855 which stored disease-gene associations from FunDO. The names are the DOIDs (DOIDs are ids of terms in Disease Ontology, e.g. "DOID:4") and list elements are vectors of gene symbols.

References


Examples

data(d2g_fundo_symbol)

description

A list of disease-gene associations from the reference paper (see below).

Value

d2g_separation is a named list of length 299 which stored disease-gene associations from the reference paper (see below). The names are diseases and list elements are vectors of gene entrez ids.

References


Examples

data(d2g_separation)
**d2s_hsdn**

<table>
<thead>
<tr>
<th><strong>d2s_hsdn</strong></th>
<th><strong>d2s_hsdn</strong></th>
</tr>
</thead>
</table>

**Description**

diseases, symptoms and their co-occurrences in PubMed

**Value**

d2s_hsdn is a data.frame of 73726 rows and 3 columns, contains PubMed co-occurrences of diseases and symptoms, will be used in method CosineDFV.

**References**


**See Also**

CosineDFV

**Examples**

data(d2s_hsdn)

---

<table>
<thead>
<tr>
<th><strong>d2s_hsdn_sample</strong></th>
<th><strong>d2s_hsdn_sample</strong></th>
</tr>
</thead>
</table>

**Description**

a sample of d2s_hsdn

**Value**

d2s_hsdn_sample is a data.frame of 1480 rows and 3 columns, contains PubMed co-occurrences of diseases and symptoms. It is a sample of d2s_hsdn.

**References**


**See Also**

d2s_hsdn, CosineDFV

**Examples**

data(d2s_hsdn_sample)
FunSim

*calculate disease similarity by FunSim*

**Description**

given two vectors of diseases, a list of disease-gene associations, and a list of gene-gene log-likelihood score from HumanNet, this function will calculate disease similarity by method FunSim.

**Usage**

```r
FunSim(D1, D2, d2g, LLSnList)
```

**Arguments**

- `D1` a vector consists disease ids
- `D2` another vector consists disease ids
- `d2g` a list of disease-gene associations, while gene ids should be entrez id.
- `LLSnList` a list of gene-gene log-likelihood score from HumanNet

**Value**

a matrix of disease disease simialrity which rownames is D1 and colnames is D2

**Author(s)**

Peng Ni, Min Li

**References**


**See Also**

LLSn2List

**Examples**

```r
## in this method, we must use disease-gene associations
## which genes are represented by entrez ids because of HumanNet
data(d2g_fundo_entrezid)
data(HumanNet_sample)
## we specified 5 DOIDs to match Human_sample
ds<-c("DOID:8176","DOID:2394","DOID:3744","DOID:8466","DOID:5679")
llsnlist<-LLSn2List(HumanNet_sample)
FunSim(ds,ds,d2g_fundo_entrezid,llsnlist)
```
**get_GOterm2GeneAssos**

**Description**
get GO-gene associations from GO.db and org.Hs.eg.db

**Usage**

```r
get_GOterm2GeneAssos(GOONTOLOGY = c("BP", "MF", "CC"),
                     geneid = c("ENTREZID", "SYMBOL"),
                     rm.IEAs = TRUE,
                     rm.termlessthan3genes = TRUE)
```

**Arguments**

- `GOONTOLOGY` "BP" or "MF" or "CC"
- `geneid` gene id type, "ENTREZID" or "SYMBOL"
- `rm.IEAs` logical value, remove GO terms with evidence "IEA" or not
- `rm.termlessthan3genes` logical value, remove terms whose number of annotated genes are less than 3 or not

**Value**
a list which names are GO term IDs and elements are gene ids or symbols annotated with GO terms

**Author(s)**
Peng Ni, Min Li

**References**

**See Also**

`PSB`, `Sun_function`

**Examples**

```r
go2g<-get_GOterm2GeneAssos(GOONTOLOGY="BP", geneid="SYMBOL")
go2g
```
Description

a sample list of GO term-gene associations.

Value

go2g_sample is a named list of length 465. The names are GO term ids (GOIDs) and list elements are vectors of gene symbols. The entire data of GO term-gene assos can be obtained by function get_GOterm2GeneAssos.

See Also

get_GOterm2GeneAssos

Examples

data(go2g_sample)

Description

graphlet signature of nodes in HPRD PPI network.

Value

#’ graphlet_sig_hprd is a matrix of 9270 rows and 73 rows. The rownames of graphlet_sig_hprd are gene symbols of nodes from HPRD. Each row indicates a graphlet signature of one node. Graphlet signatures of nodes in HPRD PPI network were calculated by ORCA tool, will be used in method Sun_topology.

References


See Also

Sun_topology

Examples

data(graphlet_sig_hprd)
**HumanNet_sample**

**Description**

A sample of HumanNet likelihood score data which will be used in method FunSim.

**Value**

HumanNet_sample is a data.frame has 22708 rows and 3 columns. Each row indicates a pair of genes and their normalized likelihood score in HumanNet. HumanNet_sample will be used in method FunSim after being converted to list by method LLSn2List. The entire data of HumanNet can be downloaded from the website http://www.functionalnet.org/humannet/.

**References**


**See Also**

FunSim, LLSn2List

**Examples**

```r
data(HumanNet_sample)
```

**HypergeometricTest**

**Hypergeometric test and multiple testing**

**Description**

Given disease-gene associations and go-gene associations, return disease-go associations by using hypergeometric test and fdr multiple testing.

**Usage**

```r
HypergeometricTest(d2g, go2g, method = "BH", cutoff = 0.05)
```

**Arguments**

- `d2g`: a list of disease-gene associations
- `go2g`: a list of GOterm-gene associations
- `method`: multiple testing method, the same as parameter in method p.adjust
- `cutoff`: multiple testing cut off value

**Value**

A list of disease-GO term associations.
ICod

Author(s)

Peng Ni, Min Li

See Also

PSB, Sun_function, get_GOterm2GeneAssos

Examples

```r
## see more examples in function PSB or Sun_function
data(d2go_sample)
data(go2g_sample)
data(d2g_fundo_symbol)
HypergeometricTest(d2g_fundo_symbol[names(d2go_sample)], go2g_sample)
```

ICod

*calculate disease similarity by ICod*

Description

Given two vectors of diseases, a list of disease-gene associations and a PPI network, this function will calculate disease similarity by method ICod.

Usage

```r
ICod(D1, D2, d2g, graph, A = 0.9, b = 1, C = 0)
```

Arguments

- **D1**: a vector consists disease ids
- **D2**: another vector consists disease ids
- **d2g**: a list of disease-gene associations
- **graph**: an igraph graph object of PPI network
- **A**: a parameter used in ICod to calculate transformed distance of node pair, default 0.9
- **b**: a parameter used in ICod to calculate transformed distance of node pair, default 1
- **C**: a parameter used in ICod to calculate disease similarity, default 0

Value

A matrix of disease disease similarity which rownames is D1 and colnames is D2

Author(s)

Peng Ni, Min Li
**InformationContent**

**References**


**Examples**

```r
data(d2g_fundo_symbol)
data(PPI_HPRD)

graph_hprd<-graph.data.frame(PPI_HPRD,directed=FALSE) # get a igraph object based on HPRD data
ds<-sample(names(d2g_fundo_symbol),5)
ICod(ds,ds,d2g_fundo_symbol,graph_hprd)
```

**Descripion**

calculate information content of all term ids in a term list

**Usage**

```r
InformationContent(T2G)
```

**Arguments**

- **T2G**
  a list of Term-Gene associations which names are term ids

**Value**

a list of IC values of inputted term ids

**Author(s)**

Peng Ni, Min Li

**Examples**

```r
data(d2g_fundo_symbol)
InformationContent(d2g_fundo_symbol[1:5])
```
Description

interactome data

Value

interactome is a data.frame of 141296 rows and 2 columns. Each row indicates an interaction of two gene entrez ids. It was obtained from the reference below.

References


Examples

data(interactome)

jaccardindex calculating Jaccard Index

Description

calculate Jaccard Index of two terms by using their annotated genes

Usage

jaccardindex(x1, x2, x2y)

Arguments

x1 a disease id
x2 another disease id
x2y a list of disease-gene associations which consists x1 and x2

Value

numeric value of a jaccard index of x1 and x2

Author(s)

Peng Ni, Min Li

Examples

## this function is not just for disease-gene associations
data(d2go_sample)
d1<-names(d2go_sample)[1]
d2<-names(d2go_sample)[2]
jaccardindex(d1,d2,d2go_sample)
**LLSn2List**

- **convert data.frame of HumanNet log-likelihood Score to list**

**Description**

convert HumanNet normalized log-likelihood score from data.frame to list, which will be used in FunSim method

**Usage**

```r
LLSn2List(LLSn)
```

**Arguments**

- `LLSn` data.frame of gene-gene normalized log-likelihood score in HumanNet

**Value**

a list of normalized log-likelihood score

**Author(s)**

Peng Ni, Min Li

**References**


**See Also**

- `FunSim`

**Examples**

```r
## see examples in function FunSim
data(HumanNet_sample)
llsnlist<-LLSn2List(HumanNet_sample[1:100,])
llsnlist
```
Normalize

**Description**

normalize a vector or a matrix based on the formula from SemFunSim

**Usage**

```
Normalize(data)
```

**Arguments**

- `data`: a numeric/integer vector or matrix

**Value**

normalized vector or matrix

**Author(s)**

Peng Ni, Min Li

**References**


**Examples**

```
sim<-matrix(1:9,3,3)
Normalize(sim)
```

---

orbit_dependency_count

**Description**

orbit dependency count

**Value**

`orbit_dependency_count` is a 73-dim vector, indicating 73 orbits’ dependency count in graphlet theory, used to calculate weight factor in method `setWeight`.

**References**

plot_bipartite

See Also

setWeight

Examples

data(orbit_dependency_count)

plot_bipartite(xylist, vertex.size = 12, vertex.shape1 = "circle",
vertex.shape2 = "square", vertex.color1 = "darkseagreen",
vertex.color2 = "turquoise1", vertex.label.font = 2,
vertex.label.dist = 0, vertex.label.color = "black",
vertex.label.cex = 0.8, edge.color = "black",
layout = layout.kamada.kawai)

Description

plot a bipartite graph which visualizes associations between diseases and genes (or GO terms etc.)

Usage

plot_bipartite(xylist, vertex.size = 12, vertex.shape1 = "circle",
vertex.shape2 = "square", vertex.color1 = "darkseagreen",
vertex.color2 = "turquoise1", vertex.label.font = 2,
vertex.label.dist = 0, vertex.label.color = "black",
vertex.label.cex = 0.8, edge.color = "black",
layout = layout.kamada.kawai)

Arguments

xylist a named list object which names are diseases and each element of the list is a
gene set with respect to each disease.
vertex.size vertex size
vertex.shape1 shape for one kind of vertex
vertex.shape2 shape for another kind of vertex
vertex.color1 color for one kind of vertex
vertex.color2 color for another kind of vertex
vertex.label.font label text font
vertex.label.dist label text dist
vertex.label.color label text color
vertex.label.cex label text cex
data(orbit_dependency_count)
edge.color edge color
layout layout

Value

an igraph plot object

Author(s)
Peng Ni, Min Li
Examples

```r
data(d2g_fundo_symbol)
d2g_sample<-sample(d2g_fundo_symbol, 3)
plot_bipartite(d2g_sample)
```

---

**plot_heatmap**

Similarity matrix heatmap plotting

Description

plot heatmap of a disease similarity matrix

Usage

```r
plot_heatmap(simmat, xlab = "", ylab = "", color.low = "white",
            color.high = "red", labs = TRUE, digits = 2, labs.size = 3,
            font.size = 14)
```

Arguments

- `simmat`: a similarity matrix
- `xlab`: xlab
- `ylab`: ylab
- `color.low`: color of low value
- `color.high`: color of high value
- `labs`: logical, add text label or not
- `digits`: round digit numbers
- `labs.size`: label size
- `font.size`: font size

Value

a ggplot object

Author(s)

Peng Ni, Min Li

References

plot_net

Examples

data(d2g_separation)
data(interactome)

graph_interactome <- graph.data.frame(interactome, directed = FALSE)

ds <- c("myocardial ischemia", "myocardial infarction", "coronary artery disease", "cerebrovascular disorders", "arthritis, rheumatoid", "diabetes mellitus, type 1", "autoimmune diseases of the nervous system", "demyelinating autoimmune diseases, cns", "respiratory hypersensitivity", "asthma", "retinitis pigmentosa", "retinal degeneration", "macular degeneration")

sep <- Separation(ds, ds, d2g_separation, graph_interactome)
sim <- Separation2Similarity(sep)
plot_heatmap(sim)

---

plot_net

plot a network based on a symmetric disease similarity matrix

Description

plot a network/graph of a symmetric disease similarity matrix, note that a unsymmetric matrix can’t be visualized into a network by this method.

Usage

plot_net(simmat, cutoff = 1, vertex.label.font = 2,
vertex.label.dist = 0.5, vertex.label.color = "black",
vertex.label.cex = 0.8, vertex.shape = "circle",
vertex.color = "paleturquoise", vertex.size = 20, edge.color = "red",
layout = layout.fruchterman.reingold)

Arguments

simmat a symmetric similarity matrix
cutoff a cutoff value, only disease pairs have similarity scores no less than cutoff will be visualized in the network
vertex.label.font label text font
vertex.label.dist label text dist
vertex.label.color label text color
vertex.label.cex label text cex
vertex.shape vertex shape
vertex.color vertex color
vertex.size vertex size
edge.color edge color
layout layout
plot_topo

Value
an igraph plot object

Author(s)
Peng Ni, Min Li

Examples

data(d2g_separation)
data(interactome)

graph_interactome<-graph.data.frame(interactome,directed=FALSE)
ds<-c("myocardial ischemia","myocardial infarction","coronary artery disease","cerebrovascular disorders","arthritis, rheumatoid","diabetes mellitus, type 1","autoimmune diseases of the nervous system","demyelinating autoimmune diseases, cns","respiratory hypersensitivity","asthma","retinitis pigmentosa","retinal degeneration","macular degeneration")

sep<-Separation(ds,ds,d2g_separation,graph_interactome)
sim<-Separation2Similarity(sep)
plot_net(sim,cutoff=0.2)

plot_topo

plot topological relationship of two gene sets

Description
plot topological relationship of two gene sets (which are associated with two diseases respectively).

Usage
plot_topo(geneset1, geneset2, graph, vertexcolor = c("tomato", "orange", "lightsteelblue"), vertex.shape = "circle", vertex.size = 14, vertex.label.font = 1, vertex.label.dist = 0, vertex.label.color = "black", vertex.label.cex = 0.5, edge.color = "black", layout = layout.auto)

Arguments
geneset1 a character vector contains gene ids
geneset2 another character vector contains gene ids
graph an igraph graph object which represents a gene network
vertexcolor a character vector contains 3 colors for vertexs
vertex.shape vertex shape
vertex.size vertex size
vertex.label.font label text font
vertex.label.dist label text dist
### PPI_HPRD

vertex.label.color

- label text color

vertex.label.cex

- label text cex

dataframe: 36867 rows and 2 columns

- Each rows indicates an interaction of two gene symbols.
- It was fetched from HPRD.

```r
# Examples

data("PPI_HPRD")
g <- graph.data.frame(PPI_HPRD, directed = FALSE) # get an igraph graph

data(d2g_fundo_symbol)
a <- d2g_fundo_symbol[["DOID:8242"]]
# get gene set a
b <- d2g_fundo_symbol[["DOID:4914"]]
# get gene set b

plot_topo(a, b, g)
```

---

### Description

PPI data from HPRD

### Value

PPI_HPRD is a data.frame of 36867 rows and 2 columns. Each rows indicates an interaction of two gene symbols. It was fetched from HPRD.

### References


### Examples

```r
data(PPI_HPRD)
```
**PSB**

*calculate disease similarity by PSB*

**Description**

Given two vectors of diseases, a list of disease-GO term associations and a list of GO term-gene associations, this function will calculate disease similarity by method PSB.

**Usage**

```r
PSB(D1, D2, d2go, go2g)
```

**Arguments**

- **D1**: A vector consists disease ids.
- **D2**: Another vector consists disease ids.
- **d2go**: A list of disease-go associations.
- **go2g**: A list of go-gene associations.

**Value**

A matrix of disease disease similarity which rownames is D1 and colnames is D2.

**Author(s)**

Peng Ni, Min Li

**References**


**See Also**

- `get_GOterm2GeneAssos`
- `HypergeometricTest`
- `Normalize`

**Examples**

```r
## these are samples of GO-gene associations and disease-GO associations
data(go2g_sample)
data(d2go_sample)

##### the entire associations can be obtained by follows:
## go2g<-get_GOterm2GeneAssos(GOONTOLOGY = "BP", geneid="SYMBOL") #get go-gene associations
## data(d2g_fundo_symbol)
## d2go<-HypergeometricTest(d2g = d2g_fundo_symbol,go2g = go2g)

##### #################################################################

d<-names(d2go_sample)
sim<-PSB(ds,ds,d2go_sample,go2g_sample)
Normalize(sim)
```
Separation

**Description**

given two vectors of diseases, a list of disease-gene associations and a PPI network, this function will calculate network-based separation by method Separation.

**Usage**

Separation(D1, D2, d2g, graph)

**Arguments**

- **D1** a vector consists disease ids
- **D2** another vector consists disease ids
- **d2g** a list of disease-gene associations
- **graph** an igraph graph object of PPI network

**Value**

a matrix of disease disease network-based separation which rownames is D1 and colnames is D2

**Author(s)**

Peng Ni, Min Li

**References**


**See Also**

Separation2Similarity

**Examples**

data(d2g_separation)
data(interactome)

graph_interactome<-graph.data.frame(interactome,directed=FALSE)
ds<-sample(names(d2g_separation),5)
sep<-Separation(ds,ds,d2g_separation,graph_interactome)
sim<-Separation2Similarity(sep)
sim
**Separation2Similarity**  
*a method which convert separation to similarity*

**Description**  
convert a separation matrix to a similarity matrix

**Usage**  
Separation2Similarity(data)

**Arguments**  
- **data**  
a numeric/integer matrix calculated by method Separation

**Value**  
a similarity matrix

**Author(s)**
Peng Ni

**See Also**
Separation

**Examples**

```r
a<-matrix(c(-4:4),3,3)
Separation2Similarity(a)
```

---

**setWeight**  
*set weight factor*

**Description**  
set weight factor of 73-orbits in graphlet theory

**Usage**  
setWeight(orbit_dependency_count)

**Arguments**  
- **orbit_dependency_count**  
a vector which each element are the dependency count of each orbit

**Value**  
a vector which contains weight factors to each orbit
Sun_annotation

Author(s)

Peng Ni

References


Examples

data(orbit_dependency_count)
setWeight(orbit_dependency_count)

Sun_annotation (Sun’s annotation measure of disease similarity calculating)

Description

given two vectors of diseases and a list of disease-gene associations, this function will calculate disease similarity by method Sun_annotation

Usage

Sun_annotation(D1, D2, d2g)

Arguments

D1 a vector consists disease ids
D2 another vector consists disease ids
d2g a list of disease-gene associations

Value

a matrix of disease disease similarity which rownames is D1 and colnames is D2

Author(s)

Peng Ni, Min Li

References


Examples

data(d2g_separation)
ds<-sample(names(d2g_separation),5)
Sun_annotation(ds,ds,d2g_separation)
Sun_function

Sun’s function measure of disease similarity calculating

Description

given two vectors of diseases and a list of disease-go term associations, this function will calculate
disease similarity by method Sun_function

Usage

Sun_function(D1, D2, d2go)

Arguments

D1  a vector consists disease ids
D2  another vector consists disease ids
d2go  a list of disease-go term associations

Value

a matrix of disease disease simialrity which rownames is D1 and colnames is D2

Author(s)

Peng Ni, Min Li

References

Sun K, Goncalves JP, Larminie C. Predicting disease associations via biological network analy-

See Also

get_GOterm2GeneAssos, HypergeometricTest

Examples

## get a sample of disease-GO associations
data(d2go_sample)

######## the entire disease-GO associations can be obtained by follows:
## go2g<-get_GOterm2GeneAssos(GOONTOLOGY = "BP", geneid="SYMBOL") #get go-gene associations
## data(d2g_fundo_symbol)
## d2go<-HypergeometricTest(d2g = d2g_fundo_symbol,go2g = go2g)
######## ###################################

ds<-names(d2go_sample)
Sun_function(ds,ds,d2go_sample)
Sun_topology

Sun's topology measure of disease similarity calculating

Description

given two vectors of diseases, a list of disease-gene associations, a matrix of genes’ graphlet signature in a PPI network and a weight vector of 73 orbits in graphlet theory, this function will calculate disease similarity by method Sun_function

Usage

Sun_topology(D1, D2, d2g, graphlet_sig_mat, weight)

Arguments

D1
a vector consists disease ids
D2
another vector consists disease ids
d2g
a list of disease-gene associations
graphlet_sig_mat
matrix of graphlet signature of nodes in a ppi network calculated by orca, see examples below.
weight
a vector which elements are weight factors to each orbit in graphlet theory

Value

a disease disease similarity matrix

Author(s)

Peng Ni, Min Li

References


Examples

data(d2g_fundo_symbol)
data(graphlet_sig_hprd) #get graphlet signatures of genes in HPRD PPI network
data(weight)
dsc<-sample(names(d2g_fundo_symbol),5)
Sun_topology(ds,ds,d2g_fundo_symbol,graphlet_sig_hprd,weight)
weight  

**Description**

weight factor

**Value**

weight is a 73-dim vector, indicating 73 orbits’ weight factor, will be used in method Sun_topology.

**References**


**See Also**

setWeight, Sun_topology

**Examples**

data(weight)

---

x2y_conv2_y2x  

**convert x2ylist to y2xlist**

**Description**

convert list of x-y associations to list of y-x associations

**Usage**

x2y_conv2_y2x(x2ylist)

**Arguments**

x2ylist  

a list which the names are xs and the elements are ys of each x

**Value**

a list of y2x

**Author(s)**

Peng Ni, Min Li

**Examples**

data(go2g_sample)

g2go_sample<-x2y_conv2_y2x(go2g_sample[1:100])
Description
concert x-y associations (e.g., disease-gene associations) from data.frame to list

Usage
x2y_df2list(x2ydf, xcol = 1, ycol = 2)

Arguments
x2ydf data.frame of x-y associations
xcol col of x in x2ydf
ycol col of y in x2ydf

Value
a list of x-y associations

Author(s)
Peng Ni, Min Li

Examples
options(stringsAsFactors = FALSE)
d2g_fundo_sample<-read.table(text = "DOID:5218   IL6
DOID:8649   EGFR
DOID:8649   PTGS2
DOID:8649   VHL
DOID:8649   ERBB2
DOID:8649   PDCD1
DOID:8649   KLRC1
DOID:5214   MPZ
DOID:5214   EGR2
DOID:5210   AMH")
d2g_fundo_list<-x2y_df2list(d2g_fundo_sample)
Index

*Topic dataset
  d2g_fundo_entrezid, 5
  d2g_fundo_symbol, 6
  d2g_separation, 6
  d2go_sample, 5
  d2s_hsdn, 7
  d2s_hsdn_sample, 7
  go2g_sample, 10
  graphlet_sig_hprd, 10
  HumanNet_sample, 11
  interactome, 14
  orbit_dependency_count, 16
  PPI_HPRD, 21
  weight, 28
*Topic package
  dSimer-package, 2

BOG, 3

CosineDFV, 4, 7

  d2g_fundo_entrezid, 5
  d2g_fundo_symbol, 6
  d2g_separation, 6
  d2go_sample, 5
  d2s_hsdn, 7, 7
  d2s_hsdn_sample, 7
  dSimer (dSimer-package), 2
  dSimer-package, 2

FunSim, 8, 11, 15

  get_GOterm2GeneAssos, 9, 10, 12, 22, 26
  go2g_sample, 10
  graphlet_sig_hprd, 10

HumanNet_sample, 11

HypergeometricTest, 5, 11, 22, 26

ICod, 12

InformationContent, 13

interactome, 14

jaccardindex, 14

LLSn2List, 8, 11, 15

Normalize, 3, 16, 22

orbit_dependency_count, 16

plot_bipartite, 17
plot_heatmap, 18
plot_net, 19
plot_topo, 20
PPI_HPRD, 21
PSB, 9, 12, 22

Separation, 23, 24
Separation2Similarity, 23, 24
setWeight, 17, 24, 28
Sun_annotation, 25
Sun_function, 9, 12, 26
Sun_topology, 10, 27, 28

weight, 28

x2y_conv2_y2x, 28
x2y_df2list, 29

weight, 28