

Package ‘graphite’

April 12, 2018

Version 1.24.1

Date 2017-12-14

Title GRAPH Interaction from pathway Topological Environment

Author Gabriele Sales <gabriele.sales@unipd.it>, Enrica Calura
<enrica.calura@gmail.com>, Chiara Romualdi
<chiara.romualdi@unipd.it>

Maintainer Gabriele Sales <gabriele.sales@unipd.it>

Description Graph objects from pathway topology derived from Biocarta,
HumanCyc, KEGG, NCI, Panther, Reactome and SPIKE databases.

License AGPL-3

Depends R (>= 2.10), methods

Imports AnnotationDbi, checkmate, graph, httr, rappdirs, stats, utils

Suggests a4Preproc, ALL, BiocStyle, clipper, codetools,
hgu133plus2.db, hgu95av2.db, impute, knitr, org.Hs.eg.db,
parallel, R.rsp, RCy3, rmarkdown, SPIA (>= 2.2), testthat,
topologyGSA (>= 1.4.0)

Collate pathway.R fetch.R conversion.R plot.R utils.R clipper.R
graph.R spia.R tables.R topologygsa.R build.R

LazyData yes

VignetteBuilder R.rsp

biocViews Pathways, ThirdPartyClient, GraphAndNetwork, Network,
Reactome, KEGG, BioCarta, Metabolomics

NeedsCompilation no

R topics documented:

as.list.DeprecatedPathwayList	2
as.list.PathwayList	3
biocarta	3
buildPathway	4
convertIdentifiers	5
cytoscapePlot	6
DeprecatedPathwayList-class	7
humancyc	7

kegg	8
nci	9
panther	10
Pathway-class	11
pathwayDatabases	12
pathwayGraph	12
PathwayList-class	13
pathways	14
Pathways-class	14
pharmgkb	15
prepareSPIA	16
reactome	16
runClipper	17
runSPIA	18
runTopologyGSA	20
smpdb	21

Index 22

as.list.DeprecatedPathwayList

Conversion of DeprecatedPathwayLists into lists.

Description

Converts a [DeprecatedPathwayList](#) into a list of [Pathways](#).

Usage

```
## S3 method for class 'DeprecatedPathwayList'
as.list(x, ...)
```

Arguments

x a [DeprecatedPathwayList](#) object
 ... extra arguments to as.list

Value

A list of pathways.

Author(s)

Gabriele Sales

See Also

[DeprecatedPathwayList](#)

Examples

```
as.list(reactome)
```

as.list.PathwayList *Conversion of PathwayLists into lists.*

Description

Converts a [PathwayList](#) into a list of [Pathways](#).

Usage

```
## S3 method for class 'PathwayList'  
as.list(x, ...)
```

Arguments

x a [PathwayList](#) object
... extra arguments to as.list

Value

A list of pathways.

Author(s)

Gabriele Sales

See Also

[PathwayList](#)

Examples

```
as.list(kegg)
```

biocarta *BioCarta pathways*

Description

BioCarta pathways.

Direct access to this object is deprecated. Call `pathways(species, "biocarta")` instead.

Pathway topology conversion

BioCarta pathways were retrieved in BioPax format from the PDI database web page.

We define a pathway for each BioPax tag “pathway”. Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag ‘complex’. However, it often happens that the ‘protein’ tag contains multiple ‘xref’ pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element ‘A’ is linked to compound ‘c’ and compound ‘c’ is linked to element ‘B’, element ‘A’ should be linked to element ‘B’. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

buildPathway	<i>Build a Pathway object.</i>
--------------	--------------------------------

Description

This function creates a new object of type Pathway given a data frame describing its edges.

Usage

```
buildPathway(id, title, species, database, proteinEdges,
             metaboliteEdges = NULL, mixedEdges = NULL,
             timestamp = NULL)
```

Arguments

id	the pathway identifier.
title	the title of the pathway.
species	the species the pathway belongs to.
database	the name of the database the pathway derives from.

proteinEdges	a data.frame of edges between proteins (or genes). Must have the following columns: src_type, src, dest_type, dest, direction and type. Direction must be one of the two strings: "directed" or "undirected".
metaboliteEdges	interactions between metabolites. Can be NULL. Otherwise, it must have the same structure as proteinEdges.
mixedEdges	interactions between metabolites and proteins. Can be NULL. Otherwise, it must have the same structure as proteinEdges.
timestamp	when the pathway was annotated, by default the time buildPathway is called.

See Also

[Pathway-class](#)

Examples

```
edges <- data.frame(src_type = "ENTREZID", src="672",
                   dest_type = "ENTREZID", dest="7157",
                   direction="undirected", type="binding")
pathway <- buildPathway("#1", "example", "hsapiens", "database", edges)

# Example with metabolites:
edges <- data.frame(src_type = "ENTREZID", src="672",
                   dest_type = "ENTREZID", dest="7157",
                   direction="undirected", type="binding")
mixed <- data.frame(src_type = "CHEBI", src="77750",
                   dest_type = "ENTREZID", dest="7157",
                   direction="undirected", type="binding")
pathway <- buildPathway("#1", "example", "hsapiens", "database",
                       edges, mixedEdges = mixed)
```

convertIdentifiers *Convert the node identifiers of a pathway.*

Description

Converts the node identifiers of pathways.

If the option Ncpus is set to a value larger than 1 and the package parallel is installed, the conversion procedure will automatically use multiple cores.

Usage

```
convertIdentifiers(x, to)
```

Arguments

x can be a list of pathways or a single pathway

to a string describing the type of the identifier. Can assume the values "entrez", "symbol" or the name of one of the columns provided by an Annotation package (for example, "UNIPROT").

Value

A Pathway object.

See Also

[Pathway](#)

Examples

```
r <- pathways("hsapiens", "reactome")
convertIdentifiers(r$`mTOR signalling`, "symbol")
```

cytoscapePlot	<i>Plot a pathway graph in Cytoscape</i>
---------------	--

Description

Renders the topology of a pathway as a Cytoscape graph.

Usage

```
cytoscapePlot(pathway, ..., cy.ver = 3)
```

Arguments

pathway	a Pathway object.
...	optional arguments forwarded to pathwayGraph .
cy.ver	select a Cytoscape version. Only version 3 is supported in this release.

Details

Requires the RCy3 package.

See Also

[Pathway](#)

[pathwayGraph](#)

Examples

```
## Not run:
r <- pathways()
cytoscapePlot(convertIdentifiers(reactome$`Unwinding of DNA`, "symbol"))

## End(Not run)
```

DeprecatedPathwayList-class
Class "DeprecatedPathwayList"

Description

Represents deprecated objects for accessing pathway databases.

Extends

Class "[Pathways](#)", directly.

Methods

`l[i]`: Returns a selection of the pathways contained in the pathway list.

`l[[i]]` Access one of the pathways contained in the pathway list.

`l$title` Access one of the pathways by its title.

`convertIdentifiers(l, to)` Returns a new list of pathways using a different type of node identifiers.

`length(l)` Returns the number of pathways contained in the list.

`names(l)` Returns the titles of the pathways contained in the list.

`prepareSPIA(l, pathwaySetName, print.names=FALSE)` Prepares the pathways for a SPIA analysis.

`runClipper(l, expr, classes, method, maxNodes=150, ...)` Runs a clipper analysis over all the pathways in the list.

`runTopologyGSA(l, test, exp1, exp2, alpha, maxNodes=150, ...)` Runs a topologyGSA analysis over all the pathways in the list.

Author(s)

Gabriele Sales

See Also

[PathwayList](#)

humancyc *Humancyc pathways*

Description

HumanCyc pathways.

Direct access to this object is deprecated. Call `pathways(species, "humancyc")` instead.

Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* 38:D473-9 2010.

Pathway topology conversion

HumanCyc pathways were retrieved in the BioPax format downloaded by the Pathway Commons web page <http://www.pathwaycommons.org>.

We define a pathway for each BioPax tag “pathway”. Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag ‘complex’. However, it often happens that the ‘protein’ tag contains multiple ‘xref’ pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element ‘A’ is linked to compound ‘c’ and compound ‘c’ is linked to element ‘B’, element ‘A’ should be linked to element ‘B’. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

kegg

KEGG pathways

Description

KEGG pathways.

Direct access to this object is deprecated. Call `pathways(species, "kegg")` instead.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 1999 Jan 1;27(1):29-34.

Pathway topology conversion

KEGG pathway were retrieved in KGML format from the KEGG ftp site.

KEGG database provides separate xml files, one for each pathway. A pathway is therefore define by all the reactions described within each file.

Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these

groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the KGML format there are two ways of defining nodes with multiple elements: protein complexes (group AND defined by entry type="group") and groups with alternative members (group OR defined by entry type="gene").

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element 'A' is linked to compound 'c' and compound 'c' is linked to element 'B', element 'A' should be linked to element 'B'. Within the KGML format there are two different ways of describing a compound mediated interaction: i) direct interaction type="PPrel" ('A' interacts with 'B' through compound 'c') and ii) indirect one type="PCrel" ('A' interacts to compound 'c' and 'c' interacts with 'B').

Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

nci

NCI pathways

Description

NCI pathways.

Direct access to this object is deprecated. Call `pathways(species, "nci")` instead.

Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. PID: the Pathway Interaction Database. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D674-9. Epub 2008 Oct 2.

Pathway topology conversion

NCI pathways were retrieved in BioPax format from the PDI database web page.

We define a pathway for each BioPax tag "pathway". Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag 'complex'. However, it often happens that the 'protein' tag contains multiple 'xref' pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element 'A' is linked to compound 'c' and compound 'c' is linked to element 'B', element 'A' should be linked to element 'B'. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

panther

Panther pathways

Description

PANTHER pathways.

Direct access to this object is deprecated. Call `pathways(species, "panther")` instead.

PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Huaiyu Mi, Anushya Muruganujan and Paul D. Thomas Nucl. Acids Res. (2012) doi: 10.1093/nar/gks1118

Pathway topology conversion

Panther pathways were retrieved in the BioPax format downloaded by the Pathway Commons web page <http://www.pathwaycommons.org>.

We define a pathway for each BioPax tag "pathway". Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag 'complex'. However, it often happens that the 'protein' tag contains multiple 'xref' pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element 'A' is linked to compound 'c' and compound 'c' is linked to element 'B', element 'A' should be linked to element 'B'. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly

frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

Pathway-class	Class "Pathway"
---------------	-----------------

Description

A biological pathway.

Variants

A Pathway instance actually stores multiple variants of the same biological data.

This is the list of included variants:

- `proteins`: includes only interactions among proteins;
- `metabolites`: includes only interactions among metabolites;
- `mixed`: includes all available interactions.

Methods

`pathwayId(p)`: Returns the native ID of the pathway.

`pathwayTitle(p)`: Returns the title of the pathway.

`pathwayDatabase(p)`: Returns the name of the database the pathway was derived from.

`pathwaySpecies(p)`: Returns the name of the species in which the pathway was annotated.

`pathwayTimestamp(p)`: Returns the date of pathway data retrieval.

`pathwayURL(p)`: Returns the URL of the pathway in its original database, if available.

`convertIdentifiers(p, to)`: Returns a new pathway using a different type of node identifiers.

`edges(p, which = c("proteins", "metabolites", "mixed"), stringsAsFactors = TRUE)`: Returns a data.frame describing the edges of this pathway.

The option which selects the desired pathway variant (see section "Variants" above).

If `stringsAsFactors` is TRUE, strings are converted to factors.

`nodes(p, which = c("proteins", "metabolites", "mixed"))`: Returns the names of the nodes belonging to this pathway.

The option which selects the desired pathway variant (see section "Variants" above).

`plot(p)`: Shows the pathway topology in Cytoscape.

`runClipper(p, expr, classes, method, ...)`: Runs a clipper analysis over the pathway.

`runTopologyGSA(p, test, exp1, exp2, alpha, ...)`: Runs a topologyGSA analysis over the pathway.

Author(s)

Gabriele Sales

See Also[pathways](#)

pathwayDatabases	<i>List the available pathway databases.</i>
------------------	--

Description

Obtains the list of pathway databases available through graphi te.

Usage

```
pathwayDatabases()
```

Value

Returns a data.frame with two columns: species and database.

Author(s)

Gabriele Sales

See Also[pathways](#)**Examples**

```
pathwayDatabases()
```

pathwayGraph	<i>Graph representing the topology of a pathway</i>
--------------	---

Description

Builds a graphNEL object representing the topology of a pathway.

Usage

```
pathwayGraph(pathway, which = "proteins", edge.types = NULL)
```

Arguments

pathway	a Pathway object.
which	the pathway variant you want. See Pathway documentation for a list of the supported variants.
edge.types	keep only the edges matching the type names in this vector.

Value

A graphNEL object.

See Also

[Pathway](#)
[graphNEL](#)

Examples

```
r <- pathways("hsapiens", "reactome")
pathwayGraph(r$mTOR signalling`, edge.types="Binding")
```

PathwayList-class	Class "PathwayList"
-------------------	---------------------

Description

A collection of pathways from a single database.

Extends

Class "[Pathways](#)", directly.

Methods

`l[i]`: Returns a selection of the pathways contained in the pathway list.

`l[[i]]` Access one of the pathways contained in the pathway list.

`l$title`` Access one of the pathways by its title.

`convertIdentifiers(l, to)` Returns a new list of pathways using a different type of node identifiers.

`length(l)` Returns the number of pathways contained in the list.

`names(l)` Returns the titles of the pathways contained in the list.

`prepareSPIA(l, pathwaySetName, print.names=FALSE)` Prepares the pathways for a SPIA analysis.

`runClipper(l, expr, classes, method, maxNodes=150, ...)` Runs a clipper analysis over all the pathways in the list.

`runTopologyGSA(l, test, exp1, exp2, alpha, maxNodes=150, ...)` Runs a topologyGSA analysis over all the pathways in the list.

Author(s)

Gabriele Sales

See Also

[pathways](#)

pathways	<i>Retrieve a list of pathways.</i>
----------	-------------------------------------

Description

This function retrieves a list of pathways from a database for a given species.

Usage

```
pathways(species, database)
```

Arguments

species	one of the supported species
database	the name of the pathway database

Value

A PathwayList object.

See Also

[PathwayList](#)

Examples

```
pathways("hsapiens", "reactome")
```

Pathways-class	<i>Class "Pathways"</i>
----------------	-------------------------

Description

A virtual class acting as a common parent to all other classes representing pathway databases.

Objects from the Class

A virtual Class: No objects may be created from it.

Methods

No methods defined with class "Pathways" in the signature.

Author(s)

Gabriele Sales

See Also

[PathwayList](#)

pharmgkb

PharmGKB pathways

Description

PharmGKB pathways.

Direct access to this object is deprecated. Call `pathways(species, "pharmgkb")` instead.

Pharmacogenomics Knowledge for Personalized Medicine. M. Whirl-Carrillo, E.M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C.F. Thorn, R.B. Altman and T.E. Klein. *Clinical Pharmacology & Therapeutics* (2012) doi: 10.1038/clpt.2012.96

Pathway topology conversion

PharmGKB pathways were retrieved in the BioPax format from the PharmGKB download page <https://www.pharmgkb.org/downloads/>.

We define a pathway for each BioPax tag "pathway". Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag 'complex'. However, it often happens that the 'protein' tag contains multiple 'xref' pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element 'A' is linked to compound 'c' and compound 'c' is linked to element 'B', element 'A' should be linked to element 'B'. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

prepareSPIA *Prepare pathway dataset needed by runSPIA.*

Description

Prepare pathway dataset needed by runSPIA. See [runSPIA](#) and [spia](#) for more details.

Usage

```
prepareSPIA(db, pathwaySetName, print.names = FALSE)
```

Arguments

db a [PathwayList](#) object or a list of [Pathways](#).
pathwaySetName name of the output pathway set.
print.names print pathway names as the conversion advances.

References

Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, Kim CJ, Kusanovic JP, Romero R. A novel signaling pathway impact analysis. *Bioinformatics*. 2009 Jan 1;25(1):75-82.

Adi L. Tarca, Sorin Draghici, Purvesh Khatri, et. al, A Signaling Pathway Impact Analysis for Microarray Experiments, 2008, *Bioinformatics*, 2009, 25(1):75-82.

Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R.: A systems biology approach for pathway level analysis. *Genome Research*, 17, 2007.

See Also

[runSPIA](#)
[spia](#)
[PathwayList](#)

reactome *Reactome pathways*

Description

Reactome pathways.

Direct access to this object is deprecated. Call `pathways(species, "reactome")` instead.

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res*. 2009 Jan;37(Database issue):D619-22. Epub 2008 Nov 3.

Pathway topology conversion

Reactome pathways were retrieved in the BioPax format from the Reactome database web page.

We define a pathway for each BioPax tag “pathway”. Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag ‘complex’. However, it often happens that the ‘protein’ tag contains multiple ‘xref’ pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element ‘A’ is linked to compound ‘c’ and compound ‘c’ is linked to element ‘B’, element ‘A’ should be linked to element ‘B’. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

runClipper

Run a topological analysis on an expression dataset using clipper.

Description

clipper is a package for topological gene set analysis. It implements a two-step empirical approach based on the exploitation of graph decomposition into a junction tree to reconstruct the most relevant signal path. In the first step clipper selects significant pathways according to statistical tests on the means and the concentration matrices of the graphs derived from pathway topologies. Then, it “clips” the whole pathway identifying the signal paths having the greatest association with a specific phenotype.

Usage

```
runClipper(x, expr, classes, method, which = "proteins", seed = NULL, ...)
```

Arguments

x a [PathwayList](#), a list of [Pathways](#) or a single [Pathway](#) object.
 expr a matrix (size: number p of genes x number n of samples) of gene expression.

classes	a vector (length: n) of class assignments.
method	the kind of test to perform on the cliques. It could be either "mean" or "variance".
which	the pathway variant you want. See Pathway documentation for a list of the supported variants.
seed	if not NULL, set the seed for the random number generator used by clipper.
...	additional options: see for details easyClip . When invoked on a PathwayList , you can use the named option <code>maxNodes</code> to limit the analysis to those pathways with at most a given number of nodes.

Details

The expression data and the pathway have to be annotated in the same set of identifiers.

References

Martini P, Sales G, Massa MS, Chiogna M, Romualdi C. Along signal paths: an empirical gene set approach exploiting pathway topology. *Nucleic Acids Res.* 2013 Jan 7;41(1):e19. doi: 10.1093/nar/gks866. Epub 2012 Sep 21. PubMed PMID: 23002139; PubMed Central PMCID: PMC3592432.

See Also

[clipper](#)

Examples

```
if (require(clipper) & require(ALL) & require(a4Preproc)) {
  data(ALL)
  pheno <- as(phenoData(ALL), "data.frame")
  samples <- unlist(lapply(c("NEG", "BCR/ABL"), function(t) {
    which(grepl("^B\\d*", pheno$BT) & (pheno$mol.biol == t))[1:10]
  }))
  classes <- c(rep(1,10), rep(2,10))

  expr <- exprs(ALL)[,samples]
  rownames(expr) <- paste("ENTREZID", featureData(addGeneInfo(ALL))$ENTREZID,
    sep = ":")

  k <- as.list(pathways("hsapiens", "kegg"))
  selected <- k[c("Bladder cancer", "Cytosolic DNA-sensing pathway")]

  runClipper(selected, expr, classes, "mean", pathThr = 0.1)
}
```

runSPIA

Run SPIA analysis

Description

Run a topological analysis on an expression dataset using SPIA.

Usage

```
runSPIA(de, all, pathwaySetName, ...)
```

Arguments

de A named vector containing log₂ fold-changes of the differentially expressed genes. The names of this numeric vector are Entrez gene IDs.

all A vector with the Entrez IDs in the reference set. If the data was obtained from a microarray experiment, this set will contain all genes present on the specific array used for the experiment. This vector should contain all names of the 'de' argument.

pathwaySetName A list of pathways like [kegg](#), [nci](#) or [reactome](#).

... Additional options to pass to [spia](#).

Details

The spia option "organism" is internally used. It is an error use it in the additional options.

Value

The same of spia, without KEGG links. A data frame containing the ranked pathways and various statistics: pSize is the number of genes on the pathway; NDE is the number of DE genes per pathway; tA is the observed total perturbation accumulation in the pathway; pNDE is the probability to observe at least NDE genes on the pathway using a hypergeometric model; pPERT is the probability to observe a total accumulation more extreme than tA only by chance; pG is the p-value obtained by combining pNDE and pPERT; pGFdr and pGFWER are the False Discovery Rate and respectively Bonferroni adjusted global p-values; and the Status gives the direction in which the pathway is perturbed (activated or inhibited).

References

Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, Kim CJ, Kusanovic JP, Romero R. A novel signaling pathway impact analysis. *Bioinformatics*. 2009 Jan 1;25(1):75-82.

Adi L. Tarca, Sorin Draghici, Purvesh Khatri, et. al, A Signaling Pathway Impact Analysis for Microarray Experiments, 2008, *Bioinformatics*, 2009, 25(1):75-82.

Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R.: A systems biology approach for pathway level analysis. *Genome Research*, 17, 2007.

See Also

[spia](#)

Examples

```
if (require(SPIA) && require(hgu133plus2.db)) {
  data(colorectalancer)

  top$ENTREZ <- mapIds(hgu133plus2.db, top$ID, "ENTREZID", "PROBEID", multiVals = "first")
  top <- top[!is.na(top$ENTREZ) & !duplicated(top$ENTREZ), ]
  top$ENTREZ <- paste("ENTREZID", top$ENTREZ, sep = ":")
  tg1 <- top[top$adj.P.Val < 0.05, ]
```

```

DE_Colorectal = tg1$logFC
names(DE_Colorectal) <- tg1$ENTREZ
ALL_Colorectal <- top$ENTREZ

biocarta <- pathways("hsapiens", "biocarta")[1:20]
biocarta <- convertIdentifiers(biocarta, "ENTREZID")
prepareSPIA(biocarta, "biocartaEx")
runSPIA(de = DE_Colorectal, all = ALL_Colorectal, "biocartaEx")
}

```

runTopologyGSA	<i>Run a topological analysis on an expression dataset using topologyGSA.</i>
----------------	---

Description

Use graphical models to test the pathway components highlighting those involved in its deregulation.

Usage

```
runTopologyGSA(x, test, exp1, exp2, alpha, ...)
```

Arguments

x	a PathwayList , a list of Pathways or a single Pathway object.
test	Either "var" and "mean". Determine the type of test used by topologyGSA.
exp1	Experiment matrix of the first class, genes in columns.
exp2	Experiment matrix of the second class, genes in columns.
alpha	Significance level of the test.
...	Additional parameters forwarded to topologyGSA.

When invoked on a [PathwayList](#), can use the named option "maxNodes" to limit the analysis to those pathways having up to this given number of nodes.

Details

This function produces a warning and returns NULL when the number of genes in common between the expression matrices and the pathway is less than 3.

References

Massa MS, Chiogna M, Romualdi C. Gene set analysis exploiting the topology of a pathway. BMC System Biol. 2010 Sep 1;4:121.

See Also

[pathway.var.test](#) [pathway.mean.test](#)

Examples

```
if (require(topologyGSA)) {
  data(examples)
  colnames(y1) <- paste("SYMBOL", colnames(y1), sep = ":")
  colnames(y2) <- paste("SYMBOL", colnames(y2), sep = ":")

  k <- pathways("hsapiens", "kegg")
  p <- convertIdentifiers(k[["Fc epsilon RI signaling pathway"]], "SYMBOL")
  runTopologyGSA(p, "var", y1, y2, 0.05)
}
```

smpdb

SMPDB pathways

Description

SMPDB pathways.

Direct access to this object is deprecated. Call `pathways(species, "smpdb")` instead.

SMPDB 2.0: Big Improvements to the Small Molecule Pathway Database. Jewison T, Su Y, Disfany FM, et al. Nucl. Acids Res. (2014) doi: 10.1093/nar/gkt1067

Pathway topology conversion

SMPDB pathways were retrieved in the BioPax format from the SMPDB download page <http://smpdb.ca/downloads>.

We define a pathway for each BioPax tag “pathway”. Pathway nodes often correspond to multiple gene products. These can be divided into protein complexes (proteins linked by protein-protein interactions) and groups made of alternative members (genes with similar biochemical functions). Thus, when considering signal propagation these groups are considered differently. The first kind (hereafter group AND) should be expanded into a clique (all proteins connected to the others), while the second (hereafter group OR) should be expanded without connection among them. In the BioPax format only one type of group is allowed: protein complexes (group AND) with the tag ‘complex’. However, it often happens that the ‘protein’ tag contains multiple ‘xref’ pointing to alternative elements of the process (group OR).

Compound mediated interactions are interactions for which a compound acts as a bridge between two elements. Since chemical compounds are not usually measured with high-throughput technology, they should be removed from the network to analyse gene signals. However, the trivial elimination of the compounds, without signal propagation, will strongly bias the topology interrupting the signals that pass through them. If element ‘A’ is linked to compound ‘c’ and compound ‘c’ is linked to element ‘B’, element ‘A’ should be linked to element ‘B’. Not all compounds are considered for the propagation because some of them (for example: H₂O, ATP, ADP) are highly frequent in map descriptions and the signal propagation through them would lead to chains too long. Compounds not considered for propagation are not characteristic of a specific reaction, but act as secondary substrates/products widely shared among different processes.

graphite allows the user to see the single/multiple relation types that characterized an edge. The type of edges have been kept as much as possible similar to those annotated in the original data format. Some new types have been introduced due to topological conversion needs.

See Also

[PathwayList](#)

Index

- *Topic **analysis**
 - runClipper, [17](#)
 - runSPIA, [18](#)
 - runTopologyGSA, [20](#)
- *Topic **classes**
 - DeprecatedPathwayList-class, [7](#)
 - Pathway-class, [11](#)
 - PathwayList-class, [13](#)
 - Pathways-class, [14](#)
- *Topic **clipper**
 - runClipper, [17](#)
- *Topic **spia**
 - runSPIA, [18](#)
- *Topic **topologyGSEA**
 - runTopologyGSA, [20](#)
- *Topic **topology**
 - runClipper, [17](#)
 - runSPIA, [18](#)
 - runTopologyGSA, [20](#)
- [, DeprecatedPathwayList-method
(DeprecatedPathwayList-class),
[7](#)
- [, PathwayList-method
(PathwayList-class), [13](#)
- [[, DeprecatedPathwayList-method
(DeprecatedPathwayList-class),
[7](#)
- [[, PathwayList-method
(PathwayList-class), [13](#)
- \$, DeprecatedPathwayList-method
(DeprecatedPathwayList-class),
[7](#)
- \$, PathwayList-method
(PathwayList-class), [13](#)

- as.list.DeprecatedPathwayList, [2](#)
- as.list.PathwayList, [3](#)

- biocarta, [3](#)
- buildPathway, [4](#)

- clipper, [18](#)
- convertIdentifiers, [5](#)

- convertIdentifiers, DeprecatedPathwayList-method
(DeprecatedPathwayList-class),
[7](#)
- convertIdentifiers, Pathway-method
(Pathway-class), [11](#)
- convertIdentifiers, PathwayList-method
(PathwayList-class), [13](#)
- cytoscapePlot, [6](#)

- DeprecatedPathwayList, [2](#)
- DeprecatedPathwayList-class, [7](#)

- easyClip, [18](#)
- edges, Pathway, character-method
(Pathway-class), [11](#)
- edges, Pathway, missing-method
(Pathway-class), [11](#)

- graphNEL, [13](#)

- humancyc, [7](#)

- kegg, [8](#), [19](#)

- length, DeprecatedPathwayList-method
(DeprecatedPathwayList-class),
[7](#)
- length, PathwayList-method
(PathwayList-class), [13](#)

- names, DeprecatedPathwayList-method
(DeprecatedPathwayList-class),
[7](#)
- names, PathwayList-method
(PathwayList-class), [13](#)
- nci, [9](#), [19](#)
- nodes, Pathway-method (Pathway-class), [11](#)

- panther, [10](#)
- Pathway, [2](#), [3](#), [6](#), [12](#), [13](#), [16–18](#), [20](#)
- Pathway-class, [11](#)
- pathway.mean.test, [20](#)
- pathway.var.test, [20](#)
- pathwayDatabase (Pathway-class), [11](#)
- pathwayDatabases, [12](#)

pathwayGraph, [6](#), [12](#)
pathwayId (Pathway-class), [11](#)
PathwayList, [3](#), [4](#), [7–11](#), [14–18](#), [20](#), [21](#)
PathwayList-class, [13](#)
Pathways, [7](#), [13](#)
pathways, [12](#), [13](#), [14](#)
Pathways-class, [14](#)
pathwaySpecies (Pathway-class), [11](#)
pathwayTimestamp (Pathway-class), [11](#)
pathwayTitle (Pathway-class), [11](#)
pathwayURL (Pathway-class), [11](#)
pharmgkb, [15](#)
plot, Pathway, ANY-method
 (Pathway-class), [11](#)
prepareSPIA, [16](#)
prepareSPIA, DeprecatedPathwayList-method
 (DeprecatedPathwayList-class),
 [7](#)
prepareSPIA, list-method (prepareSPIA),
 [16](#)
prepareSPIA, PathwayList-method
 (PathwayList-class), [13](#)

reactome, [16](#), [19](#)
runClipper, [17](#)
runClipper, DeprecatedPathwayList-method
 (DeprecatedPathwayList-class),
 [7](#)
runClipper, list-method (runClipper), [17](#)
runClipper, Pathway-method
 (Pathway-class), [11](#)
runClipper, PathwayList-method
 (PathwayList-class), [13](#)
runClipperMulti (runClipper), [17](#)
runSPIA, [16](#), [18](#)
runTopologyGSA, [20](#)
runTopologyGSA, DeprecatedPathwayList-method
 (DeprecatedPathwayList-class),
 [7](#)
runTopologyGSA, list-method
 (runTopologyGSA), [20](#)
runTopologyGSA, Pathway-method
 (Pathway-class), [11](#)
runTopologyGSA, PathwayList-method
 (PathwayList-class), [13](#)
runTopologyGSAMulti (runTopologyGSA), [20](#)

smpdb, [21](#)
spia, [16](#), [19](#)