

Package ‘ANUBIX’

August 25, 2024

Title Accurate Network Enrichment Analysis Test

Version 2.0.0

Author Miguel Castresana-Aguirre and Erik L.L. Sonnhammer

Maintainer Miguel Castresana-Aguirre <miguel.castresana@scilifelab.se>

Description Includes examples with data for accurate network enrichment analysis of gene sets against pathways.

Depends R (>= 3.4.0)

Imports dplyr,
optimr,
TailRank,
parallel,
igraph,
purrr,
stats

License GPL-3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

R topics documented:

anubix	2
anubix_clustering	3
anubix_clustering_website	4
anubix_constrained_website	6
anubix_links	7
anubix_transitivity	8
example_anubix	9

Index	11
--------------	-----------

anubix

*Perform ANUBIX for the provided gene sets***Description**

Computes ANUBIX, an accurate test for network enrichment analysis between query sets and pathway sets. Instead of normal random sampling it does constrained random sampling, taking the degree of the nodes into account.

Usage

```
anubix(network, links_matrix, genesets, pathways, cores = 2, cutoff = 0.8,
        sampling = 2000, network_type = "weighted")
```

Arguments

network	A data.frame. Two columns if the network has no weights. Where column 1 and column 2 are genes. Each row means a link between genes. If the network is weighted, then the third column are the weights of the links.
links_matrix	A numeric matrix. Links_matrix stores the links each gene has to each pathway. The number of rows equals to the number of genes in the network and the number of columns equals to the number of pathways under study.
genesets	A data.frame of two columns. Column 1, are the genes and column 2 the experiment where they belong to.
pathways	A data.frame of two columns. Column 1 are the genes and second column the pathway where they belong to.
cores	A numeric value. Cores is defined as the number of cores used by the algorithm. The default value is 2
sampling	A numeric value. Sampling is defined as the number of random samplings required to construct the null distribution. The default value is 2000
network_type	Either "weighted" or "unweighted". The default value is weighted .
A	numeric value. Cutoff is defined as the link confidence threshold of the weights between genes. The default value is 0.8 .

Value

A data frame with the following columns:

- geneset - Gene set under study.
- pathway - Pathway under study.
- obv_links - Observed number of links between the gene set and the pathway.
- exp_mean - Expected number of links between the gene set and the pathway.
- overlap - Number of genes shared by the gene set and the pathway
- p-value - p-value of the test.
- q-value - Corrected p-value using Benjamini-Hochberg.
- FWER - Corrected p-value using Bonferroni correction.

Note

anubix_links() needs to be run beforehand

See Also

[anubix_links](#), [anubix_transitivity](#), [anubix_clustering](#), [example_anubix](#)

Examples

```
# We provide with example data to be able to run ANUBIX.
## Not run:
anubix(network = example_anubix$network, links_matrix = example_anubix$links_genes, genesets = example_anubix$genesets, pathways = example_anubix$pathway_set, cores = 2, cutoff = 0.8, sampling = 2000, network_type = "weighted")

## End(Not run)
```

anubix_clustering

Network clustering of the gene set and then ANUBIX is applied

Description

Clusters the gene set using Infomap and then applies ANUBIX.

Usage

```
anubix_clustering(network, links_matrix, genesets, pathways, cores = 2,
  sampling = 2000)
```

Arguments

network	A data.frame. Two columns if the network has no weights. Where column 1 and column 2 are genes. Each row means a link between genes. If the network is weighted, then the third column are the weights of the links.
links_matrix	A numeric matrix. Links_matrix stores the links each gene has to each pathway. The number of rows equals to the number of genes in the network and the number of columns equals to the number of pathways under study.
genesets	A data.frame of two columns. Column 1, are the genes and column 2 the experiment where they belong to.
pathways	A data.frame of two columns. Column 1 are the genes and second column the pathway where they belong to.
cores	A numeric value. Cores is defined as the number of cores used by the algorithm. The default value is 2
sampling	A numeric value. Sampling is defined as the number of random samplings required to construct the null distribution. The default value is 2000

Value

A list with two dataframes:

- gene - Gene belonging to the module.
- module - Module from the clustering approach.
- geneset - Gene set under study.
- pathway - Pathway under study.
- obv_links - Observed number of links between the gene set and the pathway.
- exp_mean - Expected number of links between the gene set and the pathway.
- overlap - Number of genes shared by the gene set and the pathway
- p-value - p-value of the test.
- q-value - Corrected p-value using Benjamini-Hochberg.
- FWER - Corrected p-value using Bonferroni correction.
- module - Module from the clustering approach.

Note

anubix_links() needs to be run beforehand

See Also

[anubix](#), [anubix_transitivity](#), [anubix_old](#), [anubix_links](#), [example_anubix](#)

Examples

```
# We provide with example data to be able to run ANUBIX.
## Not run:
anubix_clustering(network = example_anubix$network, links_matrix = example_anubix$links_genes, genesets = example_anubix$genesets, pathways = example_anubix$pathway_set, cores = 2, sampling = 2000, cutoff = 0.75, network_type = "weighted")

## End(Not run)
```

anubix_clustering_website

Network clustering of the gene set and then ANUBIX is applied

Description

Clusters the gene set using Infomap and then applies ANUBIX.

Usage

```
anubix_clustering_website(
  network,
  links_matrix,
  genesets,
  pathways,
  cores = 2,
  sampling = 2000,
  cutoff = 0.75,
  network_type = "weighted",
  callback = NULL
)
```

Arguments

network	A data.frame. Two columns if the network has no weights. Where column 1 and column 2 are genes. Each row means a link between genes. If the network is weighted, then the third column are the weights of the links.
links_matrix	A numeric matrix. Links_matrix stores the links each gene has to each pathway. The number of rows equals to the number of genes in the network and the number of columns equals to the number of pathways under study.
genesets	A data.frame of two columns. Column 1, are the genes and column 2 the experiment where they belong to.
pathways	A data.frame of two columns. Column 1 are the genes and second column the pathway where they belong to.
cores	A numeric value. Cores is defined as the number of cores used by the algorithm. The default value is 2
sampling	A numeric value. Sampling is defined as the number of random samplings required to construct the null distribution. The default value is 2000

Value

A list with two dataframes:

- gene - Gene belonging to the module.
- module - Module from the clustering approach.
- geneset - Gene set under study.
- pathway - Pathway under study.
- obv_links - Observed number of links between the gene set and the pathway.
- exp_mean - Expected number of links between the gene set and the pathway.
- overlap - Number of genes shared by the gene set and the pathway
- p-value - p-value of the test.
- q-value - Corrected p-value using Benjamini-Hochberg.
- FWER - Corrected p-value using Bonferroni correction.
- module - Module from the clustering approach.

See Also

[anubix](#), [anubix_constrained](#), [anubix_links](#), [example_anubix](#)

anubix_constrained_website

Perform ANUBIX for the provided gene sets

Description

Computes ANUBIX, an accurate test for network enrichment analysis between query sets and pathway sets. Instead of normal random sampling it does constrained random sampling, taking the degree of the nodes into account.

Usage

```
anubix_constrained_website(
  links_matrix,
  genesets,
  pathways,
  cores = 2,
  sampling = 2000,
  callback = NULL
)
```

Arguments

links_matrix	A numeric matrix. Links_matrix stores the links each gene has to each pathway. The number of rows equals to the number of genes in the network and the number of columns equals to the number of pathways under study.
genesets	A data.frame of two columns. Column 1, are the genes and column 2 the experiment where they belong to.
pathways	A data.frame of two columns. Column 1 are the genes and second column the pathway where they belong to.
cores	A numeric value. Cores is defined as the number of cores used by the algorithm. The default value is 2
sampling	A numeric value. Sampling is defined as the number of random samplings required to construct the null distribution. The default value is 2000
new_deg_list	A list with the genes that fall into certain degree
map_pos_deg_list	A data frame which maps the position of a degree value in the new_deg_list list

Value

A data frame with the following columns:

- geneset - Gene set under study.
- pathway - Pathway under study.
- obv_links - Observed number of links between the gene set and the pathway.
- exp_mean - Expected number of links between the gene set and the pathway.
- overlap - Number of genes shared by the gene set and the pathway
- p-value - p-value of the test.
- q-value - Corrected p-value using Benjamini-Hochberg.
- FWER - Corrected p-value using Bonferroni correction.

See Also

[anubix_links](#), [example_anubix](#), [anubix_clustering](#), [anubix](#)

anubix_links	<i>Creates the link_matrix data</i>
--------------	-------------------------------------

Description

Computation of all the links that each gene in the network has to each of the pathways.

Usage

```
anubix_links(network, cores = 2, pathways, cutoff = 0.75, network_type = "weighted")
```

Arguments

network	A data.frame. Two columns if the network has no weights. Where column 1 and column 2 are genes. Each row means a link between genes. If the network is weighted, then the third column are the weights of the links.
pathways	A data.frame of two columns. Column 1 are the genes and second column the pathway where they belong to.
cutoff	A numeric value. Cutoff is defined as the link confidence threshold of the weights between genes. The default value is 0.75 .
network_type	Either "weighted" or "unweighted". The default value is weighted .

Value

A matrix with the following dimensions:

- nrows - Total number of genes in the studied network.
- ncols - Total number of pathways under study.

See Also

[anubix](#), [example_anubix](#), [anubix_clustering](#)

Examples

```
# Example with a tiny network:
## Not run:
anubix_links(example_anubix$network, example_anubix$pathway_set, cutoff = 0.75, "weighted")

## End(Not run)
```

anubix_transitivity *Perform ANUBIX for the provided gene sets*

Description

Computes ANUBIX, an accurate test for network enrichment analysis between query sets and pathway sets. Instead of normal random sampling it does constrained random sampling, taking the degree of the nodes into account. Additionally, it takes into account the gene set's transitivity to compute enrichment.

Usage

```
anubix_transitivity(network, links_matrix, genesets, pathways, cores = 2, cutoff = 0.8,
sampling = 2000, network_type = "weighted")
```

Arguments

network	A data.frame. Two columns if the network has no weights. Where column 1 and column 2 are genes. Each row means a link between genes. If the network is weighted, then the third column are the weights of the links.
links_matrix	A numeric matrix. Links_matrix stores the links each gene has to each pathway. The number of rows equals to the number of genes in the network and the number of columns equals to the number of pathways under study.
genesets	A data.frame of two columns. Column 1, are the genes and column 2 the experiment where they belong to.
pathways	A data.frame of two columns. Column 1 are the genes and second column the pathway where they belong to.
cores	A numeric value. Cores is defined as the number of cores used by the algorithm. The default value is 2
sampling	A numeric value. Sampling is defined as the number of random samplings required to construct the null distribution. The default value is 2000
network_type	Either "weighted" or "unweighted". The default value is weighted .
A	numeric value. Cutoff is defined as the link confidence threshold of the weights between genes. The default value is 0.8 .

Value

A data frame with the following columns:

- geneset - Gene set under study.
- pathway - Pathway under study.
- obv_links - Observed number of links between the gene set and the pathway.
- exp_mean - Expected number of links between the gene set and the pathway.
- overlap - Number of genes shared by the gene set and the pathway
- p-value - p-value of the test.
- q-value - Corrected p-value using Benjamini-Hochberg.
- FWER - Corrected p-value using Bonferroni correction.

Note

anubix_links() needs to be run beforehand

See Also

[anubix_links](#), [example_anubix](#), [anubix_clustering](#), [anubix](#)

Examples

```
# We provide with example data to be able to run ANUBIX.
## Not run:
anubix_transitivity(network = example_anubix$network, links_matrix = example_anubix$links_genes, genesets = ex
pathways = example_anubix$pathway_set, cores = 2, cutoff = 0.8, sampling = 2000, network_type = "weighted")

## End(Not run)
```

example_anubix

Example data provided with the package

Description

List collecting example data to be able to run ANUBIX

- **links_genes** - The number of links each human gene has to KEGG pathways using a link confidence cutoff in the functional association network Funcoup 4.1 of 0.75. Dimensions 12444 rows (genes) and 288 columns (pathways)
- **pathway_set** - 288 KEGG pathways for *Homo sapiens*.
- **gene_set** - Example gene set with 42 genes.
- **network** - A subset from the Funcoup network for *Homo sapiens*. 2000 links and 1788 genes.

Usage

```
example_anubix
```

Format

An object of class list of length 4.

Source

- **pathway_set:** Kanehisa, M., and Goto, S. (2002). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res., 28(1), 27-30.
- **network:** Schmitt, T., Ogris, C. and Sonnhammer, E.L.L. (2014) FunCoup 3.0: database of genome-wide functional coupling networks. Nucleic Acids Res., 42, D380–8.

References

- Kanehisa, M., and Goto, S. (2002). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res., 28(1), 27-30.
- Schmitt, T., Ogris, C. and Sonnhammer, E.L.L. (2014) FunCoup 3.0: database of genome-wide functional coupling networks. Nucleic Acids Res., 42, D380–8.

See Also

[anubix](#), [anubix_links](#), [anubix_clustering](#)

Examples

```
## Not run:
data(example_anubix)

# We provide with the precomputed matrix for Funcoup with a cutoff of 0.75 to run ANUBIX directly.
# However, to show an example for the link computation we use an a subset of funcoup network.

anubix_links(example_anubix$network,2,example_anubix$pathway_set,cutoff = 0.75, "weighted")

# Performs ANUBIX. We provide with the data necessary to run it directly.

anubix(links_matrix = example_anubix$links_genes,genesets = example_anubix$gene_set,
pathways = example_anubix$pathway_set,cores = 2,total_genes = 20000, sampling = 2000)

## End(Not run)
```

Index

* datasets

example_anubix, 9

anubix, [2](#), [4](#), [5](#), [7](#), [9](#), [10](#)

anubix_clustering, [3](#), [3](#), [7](#), [9](#), [10](#)

anubix_clustering_website, [4](#)

anubix_constrained, [5](#)

anubix_constrained_website, [6](#)

anubix_links, [3–5](#), [7](#), [7](#), [9](#), [10](#)

anubix_old, [4](#)

anubix_transitivity, [3](#), [4](#), [8](#)

example_anubix, [3–5](#), [7](#), [9](#), [9](#)