Package ‘CoCiteStats’

March 17, 2025

Title Different test statistics based on co-citation.

Version 1.78.0

Author B. Ding and R. Gentleman

Description A collection of software tools for dealing with co-citation data.

Maintainer Bioconductor Package Maintainer

`<maintainer@bioconductor.org>`

License CPL

Depends R (>= 2.0), org.Hs.eg.db

Imports AnnotationDbi

biocViews Software

git_url https://git.bioconductor.org/packages/CoCiteStats

git_branch RELEASE_3_20

git_last_commit 14fc718

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2025-03-17

Contents

actorAdjTable 2
gene.gene.statistic 3
gene.geneslist.sig 4
gene.geneslist.statistic 5
paperLen 6
twowayTable 7
twTStats 8

Index **9**

`actorAdjTable` *Compute actor size adjustment on a two way table*

Description

When two objects are related through a bipartite graph it is sometimes appropriate to carry out special adjustments. One of the adjustments is called actor size adjustment. In this case the counts are adjusted according to how often the objects are referenced.

Usage

```
actorAdjTable(twT, eps = 1e-08)
```

Arguments

`twT`	A two way table as produced by `twowayTable`.
`eps`	A small quantity used to assess approximate equality.

Details

When testing for associations between entities, the social networks literature has developed a number of tools to help measure such associations. We can think of genes (actors) as being joined by citation in papers (events) and having two genes cited in the same paper (equivalent to two actors attending the same event) suggests that they are related to each other. However, some genes are cited in many papers and so we might want to discount the level of importance, as compared to genes that are cited less often. And additionally, some papers cite very many genes, and hence typically say less about them than a paper that cites rather fewer genes.

Value

An adjusted two way table, with elements named `u11`, `u12`, `u21` and `u22`.

Author(s)

R. Gentleman

References

Testing Gene Associations Using Co-citation, by B. Ding and R. Gentleman. Bioconductor Technical Report, 2004

See Also

`paperLen`, `twowayTable`

Examples

```
tw1 = twowayTable("10", "100", FALSE)
actorAdjTable(tw1)
```

`gene.gene.statistic` *Compute gene-gene statistics*

Description

Computes gene gene statistics.

Usage

```
gene.gene.statistic(g1, g2, paperLens = paperLen())
```

Arguments

`g1`	The Entrez Gene identifier for one of the genes.
`g2`	The Entrez Gene identifier for the other gene.
`paperLens`	A vector with the number of citations for each paper.

Details

For the two genes identified by their Entrez IDs a number of two-way table statistics, i.e. those computed via `twTStats`, are returned, as are their gene and paper size adjusted variants.

Value

A list with entries

`original`	The output of `twTStats` on the observed data.
`gs`	The output of `twTStats` on the data scaled for gene size.
`ps`	The output of `twTStats` on the data scaled for paper size.
`both`	The output of `twTStats` on the data scaled for both paper and gene size.

Author(s)

B. Ding and R. Gentleman

References

Testing Gene Associations Using Co-citation, by B. Ding and R. Gentleman. Bioconductor Technical Report, 2004

See Also

`twowayTable`

Examples

```
g1 = "10"  #Entrez ID for gene 1
g2 = "101"  #Entrez ID for gene 2
pLens = paperLen()
gene.gene.statistic(g1, g2, pLens)
```

`gene.geneslist.sig` *Compute 3 two-way table statistics and p-values, with 4 different adjustments, for PubMed co-citation*

Description

This function calculates Concordance, Jaccard's index and Hubert's Γ with no adjustment, adjusting for paper size (PS), adjusting for gene size (GS) and both, to evaluate the significance of co-citation of a gene of interest and a gene list

Usage

```
gene.geneslist.sig(gene, geneslist, paperLens = paperLen(), n.resamp=100)
```

Arguments

`gene`	The Entrez Gene ID for the gene of interest.
`geneslist`	The list of Entrez Gene IDs for genes with which the co-citation of the gene of interest is to be evaluated.
`paperLens`	The sizes of the PubMed papers for consideration.
`n.resamp`	Number of resampling for generating empirical p-values.

Value

Statistics and resampling p-values for all 3 two-way tables along with the 4 adjustments for gene and `geneslist` based on `n.resamp` resamplings.

Author(s)

Beiying Ding

References

Testing Gene Associations Using Co-citation, by B. Ding and R. Gentleman. Bioconductor Technical Report, 2004

See Also

`actorAdjTable`, `paperLen`, `twTStats`, `twowayTable`

Examples

```
gene <- "705"
geneslist <- "7216"

gene.geneslist.sig(gene, geneslist, n.resamp=50)
```

`gene.geneslist.statistic` *A function to compute association measures between a gene of interest and a list of genes.*

Description

Whether or not a gene has an association with another gene, or a set of genes is measured using co-citation in PubMed as a basis for measuring that association.

Usage

```
gene.geneslist.statistic(gene, geneslist, paperLens = paperLen())
```

Arguments

`gene`	The Entrez Gene ID for the gene of interest.
`geneslist`	A vector of Entrez Gene ID for the set of genes.
`paperLens`	A vector containing the number of genes cited by each paper.

Value

To be filled in later.

Author(s)

R. Gentleman

References

Testing Gene Associations Using Co-citation, by B. Ding and R. Gentleman. Bioconductor Technical Report, 2004

See Also

`twowayTable`, `link{gene.gene.statistic}`

Examples

```
g1 = "101"
gl = c("10014", "10015", "10016", "10017", "10018")
pL = paperLen()
s1 = gene.geneslist.statistic(g1, gl, pL)
s1
```

`paperLen` *Find the number of papers cited*

Description

The set of papers that cite the input Entrez Gene identifiers are found, and for each of these the number of genes cited in that paper is computed and returned.

Usage

```
paperLen(x)
```

Arguments

`x`	A vector of Entrez Gene identifiers.

Details

This function first finds the set of unique PMIDs associated with the input set of Entrez Gene IDS. Then for each PMID it finds the number of Entrez Gene identifiers associated with that paper. The function uses different sets of variable mappings from the **org.Hs.eg.db** package.

If `x` is missing then all Entrez gene identifiers in the org.Hs.egPMID are used.

Value

`counts`	For each paper the number of Entrez Gene identifiers referred to.
`papers`	A list of the same length as `x`, each element contains the papers that refer to the corresponding Entrez Gene identifier.

Author(s)

R. Gentleman

See Also

`twTStats`

Examples

```
ans = paperLen(c("10", "1001"))
ans$counts
ans$papers
```

`twowayTable` *Compute a two way co-citation table for 2 genes.*

Description

This function computes a two way table for comparing co-citation, in PubMed for the two input genes. The values in the table can be adjusted according to either the paper size or the gene size.

Usage

```
twowayTable(g1, g2, weights = TRUE, paperLens=paperLen())
```

Arguments

`g1`	The EntrezGene identifier for gene 1.
`g2`	The EntrezGene identifier for gene 2.
`weights`	`TRUE` or `FALSE` indicating whether paper size weights should be used.
`paperLens`	A vector containing the number of genes each paper refers to, or cites.

Details

To determine the association between two genes one can use co-citation in the medical literature. When `weights` is `FALSE` this function computes the number of papers that cite only gene 1, only gene 2, both and neither.

By default, we use the **org.Hs.eg.db** package to define the set of papers that are used in the computations. For other organisms, or for more restricted sets of papers the user will need to supply the vector `paperLens` explicitly.

One can consider papers which cite many genes to be less informative than those that cite only a few genes. If `weights` is `TRUE` (the default) then papers are weighted by the inverse of the number of citations.

Value

A vector of length four, with entries `n11`, `n12`, `n21` and `n22`. These correspond to the number of papers that cite both genes, the number that cite only gene 1, the number that cite only gene 2, and the total number of papers minus those counted in `n11`, `n21`, `n12`, or in the default case the weighted versions of these quantities.

Author(s)

R. Gentleman

See Also

`paperLen`, `twTStats`

Examples

```
pL = paperLen()
twowayTable("10", "100", paperLens=pL)
twowayTable("10", "100", FALSE, paperLens=pL)
```

`twTStats` *Computes 3 two way table statistics.*

Description

For two way tables based on co-citations four different test statistics are reported, the odds ratio, the Concordance, the Jaccard index and Hubert's Gamma.

Usage

```
twTStats(twT)
```

Arguments

`twT`	A two way table, as produced by `twowayTable`.

Details

The entries in the presumed 2 by 2 table are labeled n11, n12, n21, n22, corresponding to the entries in the first row first column, first row second column etc. The odds ratio is the product of n11 and n22 divided by the product of n12 and n21. The Conordance is simply the n11 entry. The Jaccard index is the n11 entry divided by the sum of n11, n12, and n21. Hubert's Gamma is slightly more complicated.

Value

`Concordance`	The concordance statistic.
`Jaccard`	The Jaccard index.
`Hubert`	Hubert's Gamma
`OddsRatio`	The odds ratio.

Author(s)

R. Gentleman

References

Testing Gene Associations Using Co-citation, by B. Ding and R. Gentleman. Bioconductor Technical Report, 2004

See Also

`paperLen`, `twowayTable`

Examples

```
tw1 = twowayTable("10", "101", FALSE)
twTStats(tw1)
```

Index

∗ **manip**
    actorAdjTable, 2
    gene.gene.statistic, 3
    gene.geneslist.sig, 4
    gene.geneslist.statistic, 5
    paperLen, 6
    twowayTable, 7
    twTStats, 8

actorAdjTable, 2, *4*

gene.gene.statistic, 3
gene.geneslist.sig, 4
gene.geneslist.statistic, 5

paperLen, *2*, *4*, 6, *7*, *8*

twowayTable, *2–5*, 7, *8*
twTStats, *3*, *4*, *6*, *7*, 8