

Package ‘cliqueMS’

January 11, 2019

Type Package

Title Annotation of Isotopes, Adducts and Fragmentation Adducts for in-Source LC/MS Metabolomics Data

Version 0.2.4

Date 2019-01-10

Description Annotates data from liquid chromatography coupled to mass spectrometry (LC/MS) metabolomics experiments. Based on a network algorithm (O.Senan, A. Aguilar-Mogas, M. Navarro, O. Yanes, R.Guimerà and M. Sales-Pardo, Metabolomics Conference (2016), Dublin), 'CliqueMS' builds a weighted similarity network where nodes are features and edges are weighted according to the similarity of this features. Then it searches for the most plausible division of the similarity network into cliques (fully connected components). Finally it annotates metabolites within each clique, obtaining for each annotated metabolite the neutral mass and their features, corresponding to isotopes, ionization adducts and fragmentation adducts of that metabolite.

License GPL (>= 2)

Depends R (>= 3.4.0)

Imports Rcpp (>= 0.12.15), xcms, igraph, qlcMatrix, Matrix, CAMERA, methods

LinkingTo Rcpp, BH, RcppArmadillo

URL <https://github.com/osenan/cliqueMS>

BugReports <https://github.com/osenan/cliqueMS/issues>

SystemRequirements C++11

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

NeedsCompilation yes

Author Oriol Senan Campos [aut, cre],
 Antoni Aguilar-Mogas [aut],
 Jordi Capellades [aut],
 Miriam Navarro [aut],
 Oscar Yanes [aut],
 Roger Guimerà [aut],
 Marta Sales-Pardo [aut]

Maintainer Oriol Senan Campos <oriol.senan@urv.cat>

Repository CRAN

Date/Publication 2019-01-11 00:00:04 UTC

R topics documented:

anClique	2
cliqueMS	3
computeCliques	4
createNetwork	5
ex.cliqueGroups	6
getAnnotation	6
getCliques	8
getIsotopes	9
negative.adinfo	10
positive.adinfo	10

Index **12**

anClique	<i>'anClique-class' for annotating isotopes and adducts</i>
----------	-------------------------------------------------------------

Description

S3 Class anClique-class for annotating isotopes and adducts in processed m/z data. Features are first grouped based on a similarity network algorithm and then annotation of isotopes and adducts is performed in each group.

anClique creates an 'anClique' object from processed m/z data.

Usage

```
anClique(msSet)
```

```
anClique(msSet)
```

Arguments

msSet A 'xcmsSet' object with processed m/z data.

Details

anClique-class contains the following elements:

'peaklist' Is a data.frame with m/z, retention time and intensity information for each feature. It also contains adduct and isotope information if annotation has been performed.

'network' Is an igraph undirected network of similarity used to compute groups of features before annotation.

'cliques' Is a list that contains the groups of features. Each id corresponds to a row in the peaklist.

'isotopes' Is a data.frame with the column 'feature' for feature id, column 'charge' for the charge, column 'grade' that starts with 0 and it is 1 for the first isotope, 2 for the second and so on and column 'cluster' which labels each group of features that are isotopes.

'cliquesFound' is TRUE if clique groups have been computed, 'isoFound' is TRUE if isotopes have been annotated, 'anFound' is TRUE if annotation of adducts have been computed.

See Also

[anClique](#)

[anClique-class](#)

Examples

```
library(cliqueMS)
mzfile <- system.file("standards.mzXML", package = "cliqueMS")
msSet <- xcms::xcmsSet(files = mzfile, method = "centWave",
ppm = 15, peakwidth = c(5,20), snthresh = 10)
ex.anClique <- anClique(msSet = msSet)
summary(ex.anClique)
```

cliqueMS

'cliqueMS' annotates isotopes and adducts in m/z data

Description

'cliqueMS' first separates features in the data into different groups. To do this it computes a similarity weighted network from the data, and searches clique groups. This cliques are fully connected components that have higher similarity in inner edges than edges outside cliques.

Once clique groups are computed, annotation of isotopes is first performed. After isotope annotation, adducts are annotated within each group

computeCliques	<i>Computes clique groups from a similarity network</i>
----------------	---------------------------------------------------------

Description

This function splits the features in the network in clique groups. The cliques are fully connected components that have high similarity for inner edges and low similarity for edges outside the clique. This function finds the clique groups that better fit this criteria, moving nodes to different groups until we find the groups that have the best log-likelihood.

Usage

```
computeCliques(anclique, tol = 1e-05, silent = TRUE)
```

Arguments

anclique	This function uses S3 'anClique' object. Gives warning if clique groups have already been computed.
tol	Minimum relative increase in log-likelihood to do a new round of log-likelihood maximisation.
silent	If 'FALSE' print on the console the log-likelihood maximization progress. Default is 'TRUE'.

Value

It returns an 'anClique' object with the computed clique groups. It adds the column 'cliqueGroup' to the 'peaklist' in the 'anClique' object.

See Also

[getCliques](#)

Examples

```
library(cliqueMS)
mzfile <- system.file("standards.mzXML", package = "cliqueMS")
msSet <- xcms::xcmsSet(files = mzfile, method = "centWave",
ppm = 15, peakwidth = c(5,20), snthresh = 10)
ex.anClique <- anClique(msSet = msSet)
summary(ex.anClique)
netlist <- createNetwork(msSet, msSet@peaks, filter = TRUE)
ex.anClique$network <- netlist$network
ex.anClique$peaklist <- netlist$peaklist
ex.cliqueGroups <- computeCliques(ex.anClique)
summary(ex.cliqueGroups)
```

createNetwork	<i>Function to create a similarity network from processed m/z data</i>
---------------	------------------------------------------------------------------------

Description

This function creates a similarity network with nodes as features and weighted edges as the cosine similarity between those nodes. Edges with weights = 0 are not included in the network. Nodes without edges are not included in the network. This network will be used to define clique groups and find annotation within this groups.

Usage

```
createNetwork(msSet, peaklist, filter = TRUE, mzerror = 5e-06,  
             intdiff = 1e-04, rtdiff = 1e-04)
```

Arguments

msSet	A 'xcmsSet' object with processed m/z data
peaklist	Is a data.frame feature info for m/z data. put each feature in a row and a column 'mz' for mass data, retention time column 'rt' and intensity in column 'maxo'.
filter	If TRUE, filter out very similar features that have a correlation similarity > 0.99 and equal values of m/z, retention time and intensity.
mzerror	Relative error for m/z, if relative error between two features is below that value that features are considered with similar m/z value.
intdiff	Relative error for intensity, if relative error between two features is below that value that features are considered with similar intensity
rtdiff	Relative error for retention time, if relative error between two features is below that value that features are considered with similar retention time

Details

Signal processing algorithms may output artefact features. Sometimes they produce two artefact features which are almost identical This artefacts may lead to errors in the computation of the clique groups, so it is recommended to set 'filter' = TRUE to drop repeated features.

Value

This function returns a list with the similarity network and the filtered peaklist if 'filter' = TRUE. If filter = FALSE the peaklist is returned unmodified.

See Also

[getCliques](#)

Examples

```
library(cliqueMS)
mzfile <- system.file("standards.mzXML", package = "cliqueMS")
msSet <- xcms::xcmsSet(files = mzfile, method = "centWave",
ppm = 15, peakwidth = c(5,20), snthresh = 10)
netlist = createNetwork(msSet, msSet@peaks, filter = TRUE)
```

ex.cliqueGroups

Example m/z processed data

Description

This dataset contains a mass spectrometry data of metabolite standards MS1 analyses were performed using an UHPLC system (1290 series, Agilent Technologies) coupled to a 6550 ESI-QTOF MS (Agilent Technologies) operated in positive (ESI+) electrospray ionization mode.

The metabolites in this set are the following: (-)riboflavine, 1,2-distearoyl-sn-glycero-3-phosphocholine, biotin, cholic acid, deoxycholic acid, L-methionine sulfoxide, thymine and uracil

Usage

```
ex.cliqueGroups
```

Format

It is an 'xcmsSet' object of one sample with 276 features. Has obtained with parameters ppm = 15, method = "centWave", peakwidth = c(5,20), snthresh = 10. then features have been splitted into cliques with getCliques, with default parameters and filter = T. Before getCliques it was used set.seed(1).

getAnnotation

Annotate adducts and fragments

Description

This function annotates adducts after isotope annotation. For each clique group, it searches for combinations of two or more features compatible with the same neutral mass and two or more adducts in 'adinfo' list. For clique groups than have more than one annotation solution, it scores all possibilities and returns the top five solutions.

Usage

```
getAnnotation(anclique, adinfo, polarity, topmasstotal = 10,
topmassf = 1, sizeanG = 20, ppm = 10, filter = 1e-04,
emptyS = 1e-06)
```

Arguments

anclique	Object of class 'anClique' with isotope annotation
adinfo	data.frame with columns 'adduct' with adduct name, column 'log10freq' with the log10 frequency of each adduct in the list, column 'massdiff' with the adduct mass diff, column 'nummol' with the number of metabolite's molecule necessary for that adduct and column 'charge' with the charge of that adduct.
polarity	Polarity of the adducts, choose between 'positive' or 'negative'
topmasstotal	All neutral masses in the group are ordered based on their adduct log-frequencies and their number of adducts. From that list, a number of "topmasstotal" masses are selected for the final annotation.
topmassf	In addition to 'topmasstotal', for each feature the list of ordered neutral masses is subsetted to the masses with an adduct in that particular feature. For each sublist, a number 'topmassf' neutral masses are also selected for the final annotation.
sizeanG	After neutral mass selection, if a clique group has a number of monoisotopic features bigger than 'sizeanG', the annotation group is divided into non-overlapping annotation groups. Each subdivision is annotated independently.
ppm	Relative error in ppm in which we consider two or more features compatible with a neutral mass and two or more adducts in 'adinfo'.
filter	This parameter removes redundant annotations. If two neutral masses in the same annotation group have a relative mass difference smaller than 'filter' and the same features and adducts, drop the neutral mass with less adducts
emptyS	Score given to non annotated features. If you use your own 'adinfo', do not set 'emptyS' bigger than any adduct frequency in your list.

Details

Reported scores do not always refer to the entire clique group. There might be features whose annotation is independent from other features of the clique group. This occurs when there are no neutral masses with adducts in both groups of features. Therefore, the clique group is divided in non overlapping regions, called annotation groups. Scores report for these annotation groups. The score is intended to compare annotations within the same group, but do not compare scores of different groups because the score is smaller if the number of features in the group is bigger.

If clique groups have a lot of features, there are many combinations of neutral masses and adducts. This could lead to long running times to score the top annotations. Parameters 'topmassf' and 'topmasstotal' are relevant in those cases to drop the less likely neutral masses to speed up the time of computation and still obtain the most plausible annotation. If the clique group is small usually no neutral masses are discarded for the scoring.

Value

An 'anClique' object with annotation columns added to the peaklist

See Also

[getCliques](#) [getIsotopes](#)

Examples

```
library(cliqueMS)
summary(ex.cliqueGroups)
ex.isoAn <- getIsotopes(ex.cliqueGroups)
summary(ex.isoAn)
ex.adductAn <- getAnnotation(ex.isoAn, positive.adinfo, 'positive')
```

getCliques

Compute clique groups from processed m/z data

Description

This function splits features in groups to find isotope and adduct annotation within each group. To find them it uses a similarity network. This similarity network has nodes as features and weighted edges as the cosine similarity between features. Once the network is obtained we find clique groups in this network. The clique groups are fully connected components with high similarity in inner edges and lower similarity in edges outside the clique. We move nodes to different groups until we find the groups with the maximum log-likelihood.

Usage

```
getCliques(msSet, filter = TRUE, mzerror = 5e-06, intdiff = 1e-04,
           rtdiff = 1e-04, tol = 1e-05, silent = TRUE)
```

Arguments

msSet	A 'xcmsSet' object with processed m/z data
filter	If TRUE, filter out very similar features that have a correlation similarity > 0.99 and equal values of m/z, retention time and intensity.
mzerror	Relative error for m/z, if relative error between two features is below that value that features are considered with similar m/z value.
intdiff	Relative error for intensity, if relative error between two features is below that value that features are considered with similar intensity.
rtdiff	Relative error for retention time, if relative error between two features is below that value that features are considered with similar retention time.
tol	Minimum relative increase in log-likelihood to do a new round of log-likelihood maximisation.
silent	If 'FALSE' print on the console the log-likelihood maximization progress. Default is 'TRUE'.

Details

Signal processing algorithms may output artefact features. Sometimes they produce two artefact features which are almost identical. These artefacts may lead to errors in the computation of the clique groups, so it is recommended to set 'filter' = TRUE to drop repeated features.

Value

It returns an 'anClique' object with the computed clique groups. It adds the column 'cliqueGroup' to the 'peaklist' in the 'anClique' object.

See Also

[computeCliques](#) [createNetwork](#) [anClique](#)

Examples

```
library(cliqueMS)
mzfile <- system.file("standards.mzXML", package = "cliqueMS")
msSet <- xcms::xcmsSet(files = mzfile, method = "centWave",
  ppm = 15, peakwidth = c(5,20), snthresh = 10)
ex.cliqueGroups <- getCliques(msSet)
```

getIsotopes

Annotate isotopes

Description

This function annotates features that are carbon isotopes based on m/z and intensity data. The monoisotopic mass has to be more intense than the first isotope, the first isotope more intense than the second isotope and so one so forth. Isotopes are annotated within each clique group.

Usage

```
getIsotopes(anclique, maxCharge = 3, maxGrade = 2, ppm = 10,
  isom = 1.003355)
```

Arguments

anclique	An 'anClique' object with clique groups computed
maxCharge	Maximum charge considered when we test two features to see whether they are isotopes
maxGrade	The maximum number of isotopes apart from the monoisotopic mass. A 'maxGrade' = 2 means that we have the monoisotopic mass, first isotope and second isotope
ppm	Relative error in ppm to consider that two features have the mass difference of an isotope
isom	The mass difference of the isotope

Value

It returns an 'anClique' object with isotope annotation. it adds the column 'isotope' to the peaklist in the anClique object

See Also

[getCliques](#)

Examples

```
library(cliqueMS)
summary(ex.cliqueGroups)
ex.isoAn <- getIsotopes(ex.cliqueGroups)
summary(ex.isoAn)
```

negative.adinfo *Default list of negative charged adducts*

Description

This is a sorted list of adducts ordered by adducts with charge < -1, adducts with charge = -1 and number of molecules = 1, and adducts with number of molecules > 1. Each of the tree groups of adducts is sorted from smaller to bigger mass difference

Usage

```
negative.adinfo
```

Format

This is a dataset of 11 rows and 5 columns, corresponding to 11 different adducts. Column 'adduct' contains adduct names, column 'log10freq' contains the log10 frequency of that adduct in the list, column 'massdiff' contains the mass difference of that adduct, column 'nummol' has the number of molecules of that adduct and column 'charge' has the charge of that adduct.

positive.adinfo *Default list of positive charged adducts*

Description

This is a sorted list of adducts ordered by adducts with charge > 1, adducts with charge = 1 and number of molecules = 1, and adducts with number of molecules > 1. Each of the tree groups of adducts is sorted from smaller to bigger mass difference

Usage

`positive.adinfo`

Format

This is a dataset of 39 rows and 5 columns, corresponding to 39 different adducts. Column 'adduct' contains adduct names, column 'log10freq' contains the log10 frequency of that adduct in the list, column 'massdiff' contains the mass difference of that adduct, column 'nummol' has the number of molecules of that adduct and column 'charge' has the charge of that adduct.

Index

*Topic **datasets**

- ex.cliqueGroups, 6
- negative.adinfo, 10
- positive.adinfo, 10

anClique, 2, 3, 9

anClique-class (anClique), 2

cliqueMS, 3

cliqueMS-package (cliqueMS), 3

computeCliques, 4, 9

createNetwork, 5, 9

ex.cliqueGroups, 6

getAnnotation, 6

getCliques, 4, 5, 7, 8, 10

getIsotopes, 7, 9

negative.adinfo, 10

positive.adinfo, 10