Package 'xMWAS'

September 19, 2017

Type Package

Title xMWAS: R package for data-integration and differential network
analysis

Version 0.41

Date 2017-09-18

Author Karan Uppal

Maintainer Karan Uppal <kuppal2@emory.edu>

Description

xMWAS allows integration of omics data using multivariate methods such as PLS. The package performs community detection and network analysis to allow visualization of positive or negative associations between different datasets generated using samples from the same individuals.

License GPL2.0

LazyLoad no

Depends mixOmics, WGCNA, snow, igraph, graph, RBGL, plyr, plsgenomics

R topics documented:

---

xMWAS_0.41-package *xMWAS*

Description

xMWAS allows integration of omics data using multivariate methods such as PLS. The package performs community detection and network analysis to allow visualization of positive or negative associations between different datasets generated using samples from same individuals. The run_xmwas() function performs the integrative, community detection, and network analysis. The output files include .gml files that can be used with software like Cytoscape.

Details

Package:	xMWAS
Type:	Package
Version:	0.41
Date:	2017-09-18
License:	gpl2.0
LazyLoad:	yes

Author(s)

Karan Uppal Maintainer: <kuppal2@emory.edu>

---

run_xmwas   *run_xmwas*

Description

The function uses sPLS or PLS and network functions in mixOmics package to perform pairwise integrative and correlation analysis. The pairwise graphs are merged using igraph and community detection is performed using the Multilevel clustering algorithm. Association networks can be visualized in R or using Cytoscape.

Usage

```
run_xmwas(xome_fname = NA, yome_fname = NA, zome_fname = NA,
wome_fname = NA, outloc = NA, class_fname = NA, Xome_data = NA,
Yome_data = NA, Zome_data = NA, Wome_data = NA, classlabels = NA,
xmwasmethod = "spls", plsmode = "canonical", max_xvar = 10000,
max_yvar = 10000, max_zvar = 10000, max_wvar = 10000,
rsd.filt.thresh = 1, corthresh = 0.4, keepX = 100, keepY = 100,
keepZ = 100, keepW = 100, pairedanalysis = FALSE, optselect = TRUE,
rawPthresh = 0.05, numcomps = 10, net_edge_colors = c("blue", "red"),
net_node_colors = c("orange", "green", "blue", "purple"), Xname = "X",
Yname = "Y", Zname = "Z", Wname = "W",
net_node_shape = c("circle", "rectangle", "triangle", "star"),
maxnodesperclass = 1000, seednum = 100, label.cex = 0.3,
vertex.size = 6, graphclustering = FALSE, interactive = TRUE,
max_connections = 1e+05, centrality_method = "betweenness",
use.X.reference = FALSE, removeRda = TRUE, compare.classes=TRUE, ...)
```

Arguments

xome_fname	Full path with filename for dataset A. Default: NA; The software uses the value provided for Xome_data when this is set to NA.
yome_fname	Full path with filename for dataset B Default: NA; The software uses the value provided for Yome_data when this is set to NA.

`zome_fname`	Full path with filename for dataset C Default: NA; The software uses the value provided for Zome_data when this is set to NA.
`wome_fname`	Full path with filename for dataset D Default: NA; The software uses the value provided for Wome_data when this is set to NA.
`Xome_data`	Data matrix for dataset A. Run: data(exnci60); head(exnci60$mrna) to see how to format data matrices.
`Yome_data`	Data matrix for dataset B
`Zome_data`	Data matrix for dataset C
`Wome_data`	Data matrix for dataset D
`outloc`	Output directory
`classlabels`	Data matrix with phenotype information. Set to NA if this information is not available. see: data(classlabels_casecontrol) for case vs control design or data(classlabels_repeatmeas for repeat measures
`class_fname`	File with phenotype information. Set to NA if this information is not available. see: data(classlabels_casecontrol) for case vs control design or data(classlabels_repeatmeasures) for repeat measures
`xmwasmethod`	Method for data integration. eg: "pls": partial least squares regression "spls": sparse partial least squares regression "o1pls": orthogonal partial least squares regression
`plsmode`	"canonical" for bi-directional relationships; "regression" for regression/predictive relationships
`max_xvar`	Maximum number of X variables to select based on relative standard deviation (RSD). e.g. 10000
`max_yvar`	Maximum number of Y variables to select based on relative standard deviation (RSD). e.g. 10000
`max_zvar`	Maximum number of Z variables to select based on relative standard deviation (RSD). e.g. 10000
`max_wvar`	Maximum number of W variables to select based on relative standard deviation (RSD). e.g. 10000
`rsd.filt.thresh`	Relative standard deviation (sd/mean) threshold
`corthresh`	Correlation threshold. eg: 0.7
`keepX`	Maximum number of X variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spls"
`keepY`	Maximum number of Y variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spls"
`keepZ`	Maximum number of Z variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spls"
`keepW`	Maximum number of W variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spls"
`pairedanalysis`	Are their repeated measurements? TRUE or FALSE
`optselect`	Find optimal number of PLS components. TRUE or FALSE
`rawPthresh`	p-value threshold calculated using Student's t-test. eg: 0.05
`numcomps`	Number of components to use in PLS model. eg: 3

net_edge_colors

        Colors for edges.

net_node_colors

        Colors for nodes.

Xname   Name for X dataset. eg: "Genes"

Yname   Name for Y dataset. eg: "Proteins"

Zname   Name for Z dataset. eg: "Metabolites"

Wname   Name for W dataset. eg: "EnvironmentalExposures"

net_node_shape

        Shapes for nodes.

maxnodesperclass

        Maximum number of nodes per dataset. eg: 100

seednum   Seed for random number generator used for plotting the network.

label.cex   Size of the labels. eg: 0.8

vertex.size Size of the nodes.

graphclustering

        Should community clustering be performed? eg: TRUE or FALSE. Multilevel clustering method is used for community detection.

interactive Generate an interactive network using the TK plot.

max_connections

        Maximum number of associations to include in the network.

centrality_method

        Method for centrality measures. eg: "eigenvector" for eigenvector centrality or "betweenness" for betweenness centrality

use.X.reference

        TRUE or FALSE if you want to use Xome_data as reference. If TRUE, only X<->Y, X<->Z, and X<->W pairwise analysis will be performed.

removeRda   TRUE or FALSE; set to TRUE if you want to remove the intermediate files.

compare.classes

        TRUE or FALSE; set to TRUE if you want to compare individual classes as provided in class labels file.

Author(s)

Karan Uppal kuppal2@emory.edu