

Package ‘xMWAS’

July 23, 2019

Type Package

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

Version 0.553

Date 2019-07-23

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 WGCNA, snow, igraph, graph, plyr, RBGL, plsgenomics, mixOmics, shiny, shinyBS, visNetwork

R topics documented:

xMWAS_0.553-package	1
run_xmwas	2
Index	6

xMWAS_0.553-package	<i>xMWAS</i>
---------------------	--------------

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: 0.553
 Date: 2019-07-23
 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 = "pls", plsmode = "regression", max_xvar = 10000,
  max_yvar = 10000, max_zvar = 10000, max_wvar = 10000, rsd.filt.thresh = -1,
  all.missing.thresh = 0, missing.val = 0, corthresh = 0.4, keepX = 1000,
  keepY = 1000, keepZ = 1000, keepW = 1000, pairedanalysis = FALSE,
  optselect = TRUE, rawPthresh = 0.05, numcomps = 10,
  net_edge_colors = c("blue", "red"), net_node_colors = c("orange", "green", "blue", "gold"),
  Xname = "X", Yname = "Y", Zname = "Z", Wname = "W",
  net_node_shape = c("square", "circle", "triangle", "star", "rectangle", "csquare",
    "crectangle", "vrectangle"), seednum = 100,
  label.cex = 0.3, vertex.size = 6, max_connections = NA,
  centrality_method = "eigenvector", use.X.reference = FALSE,
  removeRda = TRUE, compare.classes = TRUE, class.comparison.allvar = TRUE,
  modularity.weighted = FALSE, html.selfcontained = TRUE, globalcomparison = TRUE,
  plot.pairwise = TRUE, apply.sparse.class.comparison = TRUE, layout.type = "fr1",...)
```

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 "spl": sparse partial least squares regression "olpls": 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
all.missing.thresh	Remove variables (rows) that do not meet the minimum threshold for presence of non-missing values. e.g. 0.8
missing.val	How are the missing values represented in the input data files? Default: 0
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 "spl"
keepY	Maximum number of Y variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spl"
keepZ	Maximum number of Z variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spl"
keepW	Maximum number of W variables to select in sPLS. Note: keepX, keepY, keepZ, and keepW are only used when xmwasmethod is set to "spl"
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.
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.
max_connections	Maximum number of associations to include in the network. The connections between nodes are ranked based on the strength of association (+ve and -ve). Only the top "max_connections" connections are shown and used for centrality and community detection analyses. Set max_connections=NA if you want to use all connections. e.g. 1e5, 1e6, or NA
centrality_method	Method for centrality analysis. Options: 1) "eigenvector" for eigenvector centrality, which is based on the number and quality of connections - centrality scores range from 0 to 1, where 1 means high centrality and 0 means low or no centrality. Nodes/vertices with high centrality scores are connected to many other nodes, which are in turn connected to many other nodes. Please see <code>igraph::eigen_centrality</code> function for more details. 2) "betweenness" for betweenness centrality, which is based on the number of shortest paths going through a node/vertex - centrality scores are normalized and scaled to 0 to 1 range, where 1 means high betweenness centrality and 0 means low or no betweenness centrality. Please see <code>igraph::betweenness</code> function for more details. 3) "degree.count" based on the number of connections of a node. -centrality scores are normalized and scaled to 0 to 1 range. High centrality means more connections. Please see <code>igraph::degree</code> function for more details. 4) "degree.weight" is based on the sum of absolute weights of connections of a node. -centrality scores are normalized and scaled to 0 to 1 range. High centrality means stronger connections. 5) "closeness" based on the reciprocal of the sum of the distances of a node/vertex to all other nodes. -centrality scores are normalized. High centrality means the node is closer to all other nodes.
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.

class.comparison.allvar
TRUE or FALSE; set to TRUE if all nodes shown

modularity.weighted
Use edge weights during modularity analysis. TRUE or FALSE. Default: FALSE

layout.type
Different layout options: "fr1": Fruchterman-Reingold layout in igraph with weights=absolute(edge weights) "fr2": Fruchterman-Reingold layout in igraph with weights=1-absolute(edge weights) "fr": Fruchterman-Reingold layout in igraph with weights=NULL "lgl": Large Graph Layout in igraph Other options as implemented in igraph: "drl", "tree", "sphere", "nicely", "graph opt", "randomly", "kk"

Karan Uppal kuppal2@emory.edu

Index

*Topic **xMWAS**

xMWAS_0.553-package, [1](#)

run_xmwas, [2](#)

xMWAS (xMWAS_0.553-package), [1](#)

xMWAS_0.553-package, [1](#)