Package: pairkat (via r-universe)

November 6, 2026

Type Package

Title PaIRKAT

Version 1.18.0

Description PaIRKAT is model framework for assessing statistical relationships between networks of metabolites (pathways) and an outcome of interest (phenotype). PaIRKAT queries the KEGG database to determine interactions between metabolites from which network connectivity is constructed. This model framework improves testing power on high dimensional data by including graph topography in the kernel machine regression setting. Studies on high dimensional data can struggle to include the complex relationships between variables. The semi-parametric kernel machine regression model is a powerful tool for capturing these types of relationships. They provide a framework for testing for relationships between outcomes of interest and high dimensional data such as metabolomic, genomic, or proteomic pathways. PaIRKAT uses known biological connections between high dimensional variables by representing them as edges of 'graphs' or 'networks.' It is common for nodes (e.g. metabolites) to be disconnected from all others within the graph, which leads to meaningful decreases in testing power whether or not the graph information is included. We include a graph regularization or 'smoothing' approach for managing this issue.

License GPL-3

Encoding UTF-8

BugReports https://github.com/Ghoshlab/pairkat/issues

Depends R (>= 4.1)

Imports SummarizedExperiment, KEGGREST, igraph, data.table, methods, stats, magrittr, CompQuadForm, tibble

biocViews Software, Metabolomics, KEGG, Pathways, Network, GraphAndNetwork, Regression

RoxygenNote 7.1.2

Suggests rmarkdown, knitr, BiocStyle, dplyr

VignetteBuilder knitr

Config/pak/sysreqs libglpk-dev libpng-dev libxml2-dev libssl-dev zlib1g-dev

Repository Bioconductor 3.23

Date/Publication 2026-04-28 12:57:08 UTC

RemoteUrl https://github.com/bioc/pairkat

RemoteRef RELEASE_3_23

RemoteSha a28d1352e49b5ac8c8a663bf4203dfd3bcb858d9

Contents

GatherNetworks 2
PaIRKAT 3
plotNetworks 4
smokers 5

Index **6**

`GatherNetworks` *Gather pathway information from KEGG through the KEGGREST API.*

Description

Takes a `SummarizedExperiment` and constructs a list with KEGG pathways

Usage

```
GatherNetworks(SE, keggID = "KEGG", species = "hsa", minPathwaySize = 5)
```

Arguments

`SE`	A `SummarizedExperiment` with the features of interest in the first `assay`.
`keggID`	column name in pathway data containing KEGG IDs
`species`	The three letter KEGG organism ID
`minPathwaySize`	Filter pathways that are below a minimum size

Details

Queries KEGG database for known molecular interactions between included metabolites via the KEGGREST API.

Value

a list object containing the original SummarizedExperiment and igraph network objects

Examples

```
data(smokers)
# Query KEGGREST API

networks <- GatherNetworks(SE = smokers, keggID = "kegg_id",
species = "hsa", minPathwaySize = 5)
```

PaIRKAT *Perform PaIRKAT on the output from the GatherNetworks function*

Description

Pathway Integrated Regression-based Kernel Association Test (PaIRKAT) is a model framework for assessing statistical relationships between networks and some outcome of interest while adjusting for potential confounders and covariates.

Use of PaIRKAT is motivated by the analysis of networks of metabolites from a metabolomics assay and the relationship of those networks with a phenotype or clinical outcome of interest, though the method can be generalized to other domains.

Usage

```
PaIRKAT(formula.H0, networks, tau = 1)
```

Arguments

formula.H0	The null model in the "formula" format used in lm and glm functions
networks	networks object obtained with GatherNetworks
tau	A parameter to control the amount of smoothing, analagous to a bandwidth parameter in kernel smoothing. We found 1 often gave reasonable results, as over-smoothing can lead to inflated Type I errors.

Details

The PaIRKAT method is to update the feature matrix, Z, with the regularized normalized Laplacian, L_R, before performing the kernel association test. L_R is calculated using a "linear" regularization,

$$L_R = (I + \tau L)^{-}1,$$

where I is the identity matrix, τ is a regularization parameter that controls the amount of smoothing, and L is the graph's normalized Laplacian. The updated feature matrix, $Z * L_R$ is matrix used for the kernel association test.
The linear regularization and Gaussian kernel is used for all tests. See Carpenter 2021 for complete details on PaIRKAT and Smola 2003 for information about graph regularization

Value

a list object containing the formula call and results by pathway

References

Carpenter CM, Zhang W, Gillenwater L, Severn C, Ghosh T, Bowler R, et al. PaIRKAT: A pathway integrated regression-based kernel association test with applications to metabolomics and COPD phenotypes. bioRxiv. 2021 Apr 26;2021.04.23.440821.

Smola AJ, Kondor R. Kernels and Regularization on Graphs. In: Schölkopf B, Warmuth MK, editors. Learning Theory and Kernel Machines. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 144–58. (Goos G, Hartmanis J, van Leeuwen J, editors. Lecture Notes in Computer Science; vol. 2777). http://link.springer.com/10.1007/978-3-540-45167-9_12

Examples

```
data(smokers)

# Query KEGGREST API
networks <- GatherNetworks(SE = smokers, keggID = "kegg_id",
species = "hsa", minPathwaySize = 5)

# Run PaIRKAT Analysis
output <- PaIRKAT(log_FEV1_FVC_ratio ~ age, networks = networks)

# View Results
output$results
```

`plotNetworks` *Plot networks created by GatherNetworks function*

Description

Helper function for plotting networks of metabolites gathered from the KEGG pathways database using the `GatherNetworks` function.

Usage

```
plotNetworks(networks, pathway = "all", ...)
```

Arguments

`networks`	networks object obtained with `GatherNetworks`
`pathway`	Pathway to be plotted. Leaving this as 'all' will plot all pathways in 'networks'
`...`	Parameters to be passed to the `plot` (i.e. `plot.igraph`) function

Details

Plots the specified network(s) as an `igraph`

Value

a plot or list of plots generated by igraph

Examples

```
data(smokers)

# Query KEGGREST API
networks <- GatherNetworks(SE = smokers, keggID = "kegg_id",
species = "hsa", minPathwaySize = 5)

# Plot all networks
plotNetworks(networks)

# Plot specific network
plotNetworks(networks,
pathway = "Glycerophospholipid metabolism",
layout = igraph::layout_with_kk,
main = "Glycerophospholipid Metabolism")
```

---

smokers *Smokers - PaIRKAT Example Data*

Description

A synthetic data set of human subjects with phenotype variables related to lung health among smokers. Subjects have associated metabolomics assay data that are linked to KEGG pathway database IDs.

Usage

```
data(smokers)
```

Format

A SummarizedExperiment S4 object containing the following components

`phenotype` A dataframe containing phenotype variables and outcomes of interest. Row names are subject IDs.

`pathways` A dataframe containing pathway database identifiers (i.e. KEGG IDs). Row names are metabolite names

`metabalome` A dataframe containing a metabolomics assay. Row names are metabolite names and column names are subject IDs.

Index

∗ **datasets**
smokers, 5

assay, *2*

GatherNetworks, 2, *3*, *4*
glm, *3*

igraph, *4*

lm, *3*

PaIRKAT, 3
plot, *4*
plot.igraph, *4*
plotNetworks, 4

smokers, 5
SummarizedExperiment, *2*