Package 'GraphRankTest'

May 7, 2026

Type Package

Title Rank in Similarity Graph Edge-Count Two-Sample Test (RISE)

Version 0.1

Maintainer Doudou Zhou <ddzhou@nus.edu.sg>

Description Implements the Rank In Similarity Graph Edge-count two-sample test (RISE) for high-dimensional and non-Euclidean data.
The method constructs similarity-based graphs, such as k-nearest neighbor graph (k-NNG), k-minimum spanning tree (k-MST), and k-minimum distance non-bipartite pairing (k-MDP), and evaluates rank-based within-sample edge counts with asymptotic and permutation p-values.
For methodological details, see Zhou and Chen (2023) <https://proceedings.mlr.press/v195/zhou23a.html>.

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 3.5.0)

Imports stats, ade4, nbpMatching

Suggests testthat (>= 3.0.0), rmarkdown

RoxygenNote 7.3.2

URL https://proceedings.mlr.press/v195/zhou23a.html

NeedsCompilation no

Author Doudou Zhou [aut, cre] (ORCID: <https://orcid.org/0000-0002-0830-2287>), Hao Chen [aut] (ORCID: <https://orcid.org/0000-0002-4597-2773>)

Repository CRAN

Date/Publication 2025-11-12 21:10:07 UTC

Contents

RISE *Rank-based Two-Sample Tests on Similarity / Graph-Induced Rank Matrices*

Description

RISE constructs a nonnegative, symmetric rank/graph matrix R from two samples X and Y (or from a pre-computed similarity matrix S), then computes a Hotelling-type quadratic statistic with an asymptotic chi-square p-value. Optionally, a permutation p-value is returned.

Usage

```
RISE(
  X = NULL,
  Y = NULL,
  S = NULL,
  sample1ID = NULL,
  sample2ID = NULL,
  k = 10,
  rank.type = "RgNN",
  perm = 0
)
```

Arguments

X	Numeric matrix of size $m \times p$ (first sample). Optional if S is supplied.
Y	Numeric matrix of size $n \times p$ (second sample). Optional if S is supplied.
S	Numeric similarity matrix of size $N \times N$ with $N = m + n$ (larger values indicate greater similarity). If X and Y are provided, S is constructed internally as -dist(rbind(X, Y)).
sample1ID	Integer indices (length m) for sample X in S. Ignored if X and Y are given.
sample2ID	Integer indices (length n) for sample Y in S. Ignored if X and Y are given.
k	Positive integer tuning parameter. For "RgNN"/"RoNN", it is the neighborhood size in the k-nearest-neighbor graph (k-NNG); for "RgMST"/"RoMST", it controls the number of minimum-spanning-tree layers (k-MST); for "RoMDP", it specifies the number of rounds of minimum-distance non-bipartite matching (k-MDP).
rank.type	Character, one of c("RgNN","RoNN","RgMST","RoMST","RoMDP"). Prefix "Rg" denotes graph-induced ranks; prefix "Ro" denotes overall ranks obtained by ordering all selected edges. See the references for precise definitions.
perm	Integer, number of permutations for a permutation p-value (default 0).

Details

From S (or from X, Y), the procedure constructs a symmetric matrix R with zero diagonal using one of the supported graph/ranking schemes. It then forms the within-group edge sums $U_x = \sum_{i,j \in X} R_{ij}$ and $U_y = \sum_{i,j \in Y} R_{ij}$. The expectation vector and covariance matrix of (U_x, U_y) are derived under the permutation null distribution. The test statistic is

$$T = (U_x - \mu_x, U_y - \mu_y)\Sigma^{-1} \begin{pmatrix} U_x - \mu_x \\ U_y - \mu_y \end{pmatrix},$$

where μ_x, μ_y are the expected values and Σ is the covariance matrix. Under the null hypothesis, T is asymptotically chi-square distributed with 2 degrees of freedom.

Value

A list with components:

- `test.statistic`: quadratic form T_R.
- `pval.approx`: asymptotic p-value (chi-square, df = 2).
- `pval.perm`: permutation p-value (present only if `perm > 0`).

References

Zhou, D. and Chen, H. (2023). *A new ranking scheme for modern data and its application to two-sample hypothesis testing*. In *Proceedings of the 36th Annual Conference on Learning Theory (COLT 2023)*, PMLR, pp. 3615–3668.

See Also

rTests.base, Cov.asy

Examples

```
set.seed(1)
X <- matrix(rnorm(50*100, mean = 0), nrow=50)
Y <- matrix(rnorm(50*100, mean = 0.3), nrow=50)
# RgNN: graph-induced ranks from the k-nearest-neighbor graph
out.RgNN <- RISE(X = X, Y = Y, k = 10, rank.type = "RgNN", perm = 1000)
out.RgNN

# RoNN: overall ranks obtained by ordering edges from the k-NN graph
out.RoNN <- RISE(X = X, Y = Y, k = 10, rank.type = "RoNN", perm = 1000)
out.RoNN

# RgMST: graph-induced ranks from layered minimum spanning trees

out.RgMST <- RISE(X = X, Y = Y, k = 10, rank.type = "RgMST", perm = 1000)
out.RgMST


# RoMST: overall ranks obtained by ordering edges in the MST
```

```
out.RoMST <- RISE(X = X, Y = Y, k = 10, rank.type = "RoMST", perm = 1000)
out.RoMST


# RoMDP: overall ranks obtained by ordering edges from minimum-distance pairings

out.RoMDP <- RISE(X = X, Y = Y, k = 10, rank.type = "RoMDP", perm = 1000)
out.RoMDP
```

Index