

Package ‘mstknnclust’

May 13, 2026

Type Package

Title MST-kNN Clustering Algorithm

Version 1.0.0

Description Implements the MST-kNN clustering algorithm proposed by Inostroza-Ponta (2008) <<https://trove.nla.gov.au/work/28729389>>. The algorithm determines the number of clusters automatically by recursively intersecting the Minimum Spanning Tree (MST) and the k-Nearest Neighbor (kNN) proximity graphs constructed from a pairwise distance matrix. The value of k is selected via a connectivity criterion (the smallest k such that the kNN graph is connected, bounded by floor(log(n))). The package requires only a distance matrix as input and returns cluster assignments, an 'igraph' network, and partition metadata.

License GPL-2

URL <https://github.com/jorgeklz/package-mstknnclust>,
<https://jorgeklz.github.io/package-mstknnclust/>

BugReports <https://github.com/jorgeklz/package-mstknnclust/issues>

Depends R (>= 3.5.0)

Imports igraph

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Encoding UTF-8

LazyData true

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

NeedsCompilation no

Author Jorge Parraga-Alava [aut, cre] (ORCID:
<<https://orcid.org/0000-0001-8558-9122>>),
Pablo Moscato [aut],
Mario Inostroza-Ponta [aut]

Maintainer Jorge Parraga-Alava <jorge.parraga@utm.edu.ec>

Repository CRAN

Date/Publication 2026-05-13 07:10:02 UTC

Contents

dslanguages	2
dsyeastexpression	3
mst.knn	3
Index	5

dslanguages	<i>Indo-European languages dataset</i>
-------------	----------------------------------------

Description

It contains the distances between 84 Indo-European languages based on the mean percent difference in cognacy, using the 200 Swadesh words.

Usage

```
data(dslanguages)
```

Format

An data frame with 84 rows and 84 columns containing a distance matrix.

Details

Once the data set is loaded, it can be accessed as an object of class dataframe called dslanguages.

References

Dyen, I., Kruskal, J., and Black, P. (1992). An indoeuropean classification: A lexicostatistical experiment. Transactions of the American Philosophical Society. 82, (5).

dsyeastexpression	<i>Budding Yeast dataset</i>
-------------------	------------------------------

Description

It contains the expression levels of 2467 genes on 79 samples corresponding to 8 different experiments of the budding yeast: alpha factor (18 samples), cdc15 (15 samples), cold shock (4 samples), diauxic shift (7 samples), DTT shock (4 samples), elutriation (14 samples), heat shock (6 samples) and sporulation (11 samples).

Usage

```
data(dsyeastexpression)
```

Format

An data frame with 2467 rows and 79 columns.

Details

Once the data set is loaded, it can be accessed as an object of class dataframe called dsyeastexpression.

Source

<https://www.pnas.org/content/suppl/1998/12/08/95.25.14863.DC1/3917data.xls>

References

M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, 95(25):14863–14868

mst.knn	<i>Performs the MST-kNN clustering algorithm</i>
---------	--------------------------------------------------

Description

Performs the MST-kNN clustering algorithm which generates a clustering solution with automatic number-of-clusters determination by recursively intersecting the Minimum Spanning Tree (MST) and the k -Nearest Neighbor (k NN) graphs.

Usage

```
mst.knn(distance.matrix, suggested.k)
```

Arguments

- `distance.matrix` A numeric matrix or data.frame with equal numbers of rows and columns representing pairwise distances between objects.
- `suggested.k` Optional. A numeric value representing the suggested number of nearest neighbours.

Value

A list with elements `cnumber`, `cluster`, `partition`, `csize`, `network`.

Author(s)

Mario Inostroza-Ponta, Jorge Parraga-Alava, Pablo Moscato

Examples

```
set.seed(1987)
n <- 100; m <- 15
x <- matrix(runif(n * m, min = -5, max = 10), nrow = n, ncol = m)
d <- base::as.matrix(stats::dist(x, method = "euclidean"))
library("mstknnclust")
results <- mst.knn(d)
library("igraph")
plot(results$network,
      vertex.size = 8,
      vertex.color = igraph::components(results$network)$membership,
      layout = igraph::layout_with_fr(results$network, niter = 10000),
      main = paste("MST-kNN | clusters =", results$cnumber))
```

Index

* **datasets**

 dslanguages, [2](#)

 dsyeastexpression, [3](#)

dslanguages, [2](#)

dsyeastexpression, [3](#)

mst.knn, [3](#)