Package 'scStability'

June 17, 2025

Title Measuring the Stability of Dimension Reduction and Cluster
Assignment in scRNA-Seq Experiments

Version 1.0.0

Description Provides functions for evaluating the stability of low-dimensional embeddings and clus-
ter assignments in single-cell RNA sequencing (scRNA-seq) datasets. Starting from a princi-
pal component analysis (PCA) object, users can generate multiple repli-
cates of t-Distributed Stochastic Neighbor Embedding (t-SNE) or Uniform Manifold Approxi-
mation and Projection (UMAP) embeddings. Embedding stability is quantified by comput-
ing pairwise Kendall's Tau correlations across replicates and summarizing the distribution of cor-
relation coefficients. In addition to dimensionality reduction, 'scStability' assesses clustering con-
sistency using either Louvain or Leiden algorithms and calculating the Normalized Mutual Infor-
mation (NMI) between all pairs of cluster assignments. For background on UMAP and t-SNE al-
gorithms, see McInnes et al. (2020, <doi:10.21105/joss.00861>) and van der Maaten & Hin-
ton (2008, <https://lvdmaaten.github.io/tsne/>), respectively.

License MIT + file LICENSE

Language en-US

Encoding UTF-8

RoxygenNote 7.3.2

Imports aricode, future, future.apply, ggplot2, magrittr, pcaPP,
rlang, Rtsne, Seurat, stats, uwot, vegan

Suggests spelling, knitr, rmarkdown, scRNAseq, SummarizedExperiment,
testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

NeedsCompilation no

Author Ben Abrahams [aut, cre]

Maintainer Ben Abrahams <ben.abrahams.de@gmail.com>

Repository CRAN

Date/Publication 2025-06-17 06:50:06 UTC

Contents

---

clustStable   *Create and compare multiple clustering runs on scRNA-seq data*

Description

Generate multiple clustering iterations on a Seurat object containing scRNA-seq data using the provided dimensionality reduction. The function creates a shared nearest neighbor (SNN) graph and assigns clusters using the specified algorithm, then calculates stability metrics across iterations.

Usage

```
clustStable(
  n_runs,
  seurat_obj,
  method = c("louvain", "leiden"),
  resolution = 0.8,
  dims = 1:10,
  n_cores = 1,
  verbose = TRUE,
  print_plot = TRUE,
  seeds = NULL
)
```

Arguments

n_runs	Integer specifying the number of cluster assignments to generate (default: 100)
seurat_obj	A Seurat object containing scRNA-seq data with a PCA reduction
method	Character string specifying the clustering algorithm to use: either "louvain" or "leiden"
resolution	Numeric value specifying the clustering resolution parameter (default: 0.8)
dims	Integer vector specifying which PCA dimensions to use (default: 1:10)
n_cores	Integer specifying the number of CPU cores to use for parallelization (default: 1)
verbose	Whether the function should print summary statistics as it calculates them
print_plot	Whether the final violin plot should be automatically printed
seeds	A set of seeds of length n_runs for creating clusters

Value

A list containing the following components:

per_index_means

          Numeric vector of NMI values for each clustering iteration

ci   Numeric vector containing the lower and upper bounds of the 95% confidence interval

cluster_labels List of cluster assignments for each iteration

---

compareEmb	*Compare dimensional reduction embeddings and calculate stability statistics*

Description

Evaluates the stability of a set of dimension reduction embeddings by performing pairwise Procrustes alignment and calculating Kendall's Tau correlation between each pair. This function quantifies the consistency of embeddings generated with the same algorithm but different random initializations.

Usage

```
compareEmb(emb_list, n_cores = 1, verbose = TRUE, print_plot = TRUE)
```

Arguments

emb_list   A list of 2D embeddings (each typically containing coordinates for UMAP or t-SNE) created by the createEmb function

n_cores   Integer specifying the number of CPU cores to use for parallelization (default: 1)

verbose   Whether the function should print summary statistics as it calculates them

print_plot   Whether the final violin plot should be automatically printed

Value

A list containing the following components:

mean   Numeric value representing the overall mean correlation across all pairwise comparisons

mean_per_embedding

          Numeric vector of mean correlation values for each embedding

all_pairwise_correlations

          Numeric vector containing all pairwise correlation values

range   Numeric vector with minimum and maximum of mean correlation per embedding

ci   Numeric vector containing the lower and upper bounds of the 95% confidence interval

---

createEmb   *Create multiple dimension reduction embeddings*

Description

Generates multiple dimension reduction embeddings using either UMAP or t-SNE algorithms. Each embedding is created with different random initializations to assess stability. The function returns a list of embeddings, each represented as a data frame or matrix.

Usage

```
createEmb(
  dr_input,
  n_runs = 100,
  method = c("umap", "tsne"),
  n_neighbors = 15,
  min_dist = 0.1,
  perplexity = 30,
  theta = 0.5,
  n_cores = 1,
  seeds = NULL
)
```

Arguments

dr_input	A numeric matrix or data frame containing the input data for dimension reduction, with rows representing observations (cells) and columns representing PCA components
n_runs	Integer specifying the number of embeddings to generate (default: 100)
method	Character string specifying the dimension reduction method to use: either "umap" or "tsne"
n_neighbors	Integer specifying the number of neighbors to consider when constructing the initial graph (used for UMAP only, default: 30)
min_dist	Numeric value specifying the minimum distance between points in the embedding (used for UMAP only, default: 0.1)
perplexity	Numeric value controlling the effective number of neighbors (used for t-SNE only, default: 30)
theta	Numeric value between 0 and 1 controlling the speed/accuracy trade-off (used for t-SNE only, default: 0.5)
n_cores	Integer specifying the number of CPU cores to use for parallelization (default: 1)
seeds	A set of seeds of length n_runs to be used for each embedding

Value

A list of dimension reduction embeddings, each represented as a data frame with rows corresponding to observations (cells) and two columns representing the x and y coordinates in the reduced space.

---

scStability	*A user friendly wrapper function that runs the entire scRNA-seq stability workflow and shows statistics for each step*

Description

A wrapper function that runs all other stability analysis functions in order. Statistics for each step are printed accordingly and a final DR and cluster plot is shown which represents the mean embeddings and cluster assignments that were generated.

Usage

```
scStability(
  seurat_obj,
  n_runs = 100,
  dr_method = "umap",
  clust_method = "louvain",
  n_cores = 1,
  verbose = TRUE,
  print_plot = TRUE,
  seeds = NULL
)
```

Arguments

seurat_obj	A Seurat object containing scRNA-seq data and a PCA
n_runs	Number of DR embeddings and number of cluster assignments to be generated (< 250 recommended)
dr_method	Method to use for dimension reduction, either "umap" or "tsne"
clust_method	Algorithm used for clustering, either "louvain" or "leiden"
n_cores	Number of CPU cores to use for parallelising functions
verbose	Whether the function should print summary statistics as it calculates them
print_plot	Whether the final violin plot should be automatically printed
seeds	A set of seeds of length n_runs used for generating embeddings and clusters

Value

A list containing:

`mean_emb`	Data frame containing the mean embedding coordinates
`mean_clust`	Vector of the mean cluster assignments
`plot`	ggplot2 object with the mean embedding plot and cluster assignments
`embedding_stats`	
	List of embedding statistics
`cluster_stats`	List of clustering statistics
`seurat_object`	Seurat object now containing mean embeddings and mean clusters

Index