Package 'NESS'

April 3, 2025

Title What the Package Does (One Line, Title Case)

Version 0.0.0.9000

Description What the package does (one paragraph).

License `use_mit_license()`, `use_gpl3_license()` or friends to pick a license

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

R topics documented:

NESS *NESS: Neighbor Embedding Stability Scoring*

Description

Performs dimensionality reduction (t-SNE, UMAP, or PHATE) multiple times to evaluate local neighbor stability across repeated embeddings. This function is useful for evaluating the robustness of low-dimensional embeddings for high-dimensional data.

Usage

```
NESS(
  data,
  ...,
  data.name = "",
  GCP = NULL,
  cell_type = NULL,
  rareness = FALSE,
  method = "tsne",
  initialization = 1,
  stability_threshold = 0.75,
  early_stop = TRUE,
  seed_base = 1e+10,
  N = 30,
```

```
  k = 50,
  svd_cutoff_ratio = 1.1,
  svd_max_k = 30,
  stop_global_stability_threshold = 0.9,
  stop_relative_change_threshold = 0.05
)
```

Arguments

data	A numeric matrix or data frame with rows as observations and columns as features.
...	Additional arguments passed to the dimensionality reduction methods (Rtsne, uwot::umap, or phateR::phate), such as theta for t-SNE, min_dist for UMAP, or decay for PHATE.
data.name	Character string used in plot titles to label the dataset.
GCP	Optional numeric vector of neighborhood sizes (e.g., perplexity for t-SNE or number of neighbors for UMAP/PHATE). If NULL, a default sequence is generated.
cell_type	Optional vector of cell type labels for coloring the embedding plots.
rareness	Logical; if TRUE, computes rareness metrics based on neighbor consistency across embeddings.
method	Dimensionality reduction method to use. One of "tsne", "umap", or "phateR".
initialization	Initialization method: 1 for random, 2 for PCA-based initialization.
stability_threshold	Quantile threshold (default = 0.75) for determining local neighbor stability.
early_stop	Logical; if TRUE, stops early if global stability saturates.
seed_base	Base random seed used for reproducibility.
N	Number of repeated embeddings runs.
k	Number of nearest neighbors to use when computing stability metrics (default=50).
svd_cutoff_ratio	Threshold ratio used to estimate intrinsic dimensionality via SVD (default = 1.1).
svd_max_k	Maximum number of SVD components to check when estimating dimensionality (default = 30).
stop_global_stability_threshold	Early stopping threshold for global stability (default = 0.9).
stop_relative_change_threshold	Early stopping threshold for relative improvement in global stability (default = 0.05).

Value

A list containing:

local_stability A vector of local kNN stability scores for the selected GCP value.

plot_list_stability A ggplot2 object showing the embedding colored by stability.

global_stability_plot A ggplot2 line plot showing global stability across GCP values.

plot_list_cell_type (optional) Embedding plot colored by cell type if cell_type is provided.

rareness_mean (optional) A ggplot2 plot of rareness score (mean) if rareness = TRUE.