Package 'NESS'

April 4, 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 helps assess the robustness of low-dimensional embeddings for high-dimensional data.

Usage

```
NESS(
  data,
  ...,
  data.name = "",
  GCP = NULL,
  cluster = NULL,
  rareness = FALSE,
  method = "tsne",
  initialization = 1,
  stability_threshold = 0.75,
  early_stop = TRUE,
  seed_base = 1,
  N = 20,
```

```
  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.
cluster	Optional vector of cluster 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 embedding 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:

GCP Vector of neighborhood sizes used for evaluation.

GCP.optim The selected GCP value corresponding to the median rareness score.

rare.mean (optional) Vector of rareness mean scores across GCP values (if rareness = TRUE).

rare.var (optional) Vector of rareness variance scores across GCP values (if `rareness = TRUE`).

embedding Embedding coordinates for the optimal GCP value.

local_stability Vector of local kNN stability scores (no names).

global_stability Vector of global stability scores across GCP values.

embedding_stability_colored A ggplot2 plot of the embedding colored by local stability score.

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

embedding_cluster_colored (optional) Embedding plot colored by cluster labels, if `cluster` is provided.

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

par A list of input parameters used to run the function for reproducibility.