Package ‘SIMLR’

February 1, 2017

Version 1.0.1
Date 2016-10-20
Title SIMLR: Single-cell Interpretation via Multi-kernel LeaRning
Maintainer Daniele Ramazzotti <daniele.ramazzotti@yahoo.com>
Depends R (>= 3.3),
Imports parallel, Matrix, stats, methods,
Suggests BiocGenerics, BiocStyle, testthat, knitr, igraph, scran,
Description Single-cell RNA-seq technologies enable high throughput gene expression measurement of individual cells, and allow the discovery of heterogeneity within cell populations. Measurement of cell-to-cell gene expression similarity is critical to identification, visualization and analysis of cell populations. However, single-cell data introduce challenges to conventional measures of gene expression similarity because of the high level of noise, outliers and dropouts. We develop a novel similarity-learning framework, SIMLR (Single-cell Interpretation via Multi-kernel LeaRning), which learns an appropriate distance metric from the data for dimension reduction, clustering and visualization. SIMLR is capable of separating known subpopulations more accurately in single-cell data sets than do existing dimension reduction methods. Additionally, SIMLR demonstrates high sensitivity and accuracy on high-throughput peripheral blood mononuclear cells (PBMC) data sets generated by the GemCode single-cell technology from 10x Genomics.

Encoding UTF-8
LazyData TRUE
License file LICENSE
URL https://github.com/BatzoglouLabSU/SIMLR
BugReports https://github.com/BatzoglouLabSU/SIMLR
biocViews Clustering, GeneExpression, Sequencing, SingleCell
RoxygenNote 5.0.1
VignetteBuilder knitr
NeedsCompilation yes
Author Bo Wang [aut],
Daniele Ramazzotti [aut, cre],
Luca De Sano [aut],
Junjie Zhu [ctb],
Emma Pierson [ctb],
Serafim Batzoglou [ctb]
**R topics documented:**

- BuettnerFlorian .................................................. 2
- SIMLR ................................................................. 2
- SIMLR_Feature_Ranking ........................................... 3

---

### Description

example dataset to test SIMLR from the work by Buettner, Florian, et al.

### Usage

```
data(BuettnerFlorian)
```

### Format

gene expression measurements of individual cells

### Value

- list of 6: `in_X` = input dataset as an (m x n) gene expression measurements of individual cells, `n_clust` = number of clusters (number of distinct true labels), `true_labs` = ground true of cluster assignments for each of the `n_clust` clusters, `seed` = seed used to compute the results for the example, `results` = result by SIMLR for the inputs defined as described, `nmi` = normalized mutual information as a measure of the inferred clusters compared to the true labels

### Source


---

### Description

perform the SIMLR clustering algorithm

### Usage

```
SIMLR(X, c, no.dim = NA, k = 10, if.impute = FALSE, normalize = FALSE, cores.ratio = 1)
```
**SIMLR_Feature_Ranking**

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>an (m x n) data matrix of gene expression measurements of individual cells or and object of class SCESet</td>
</tr>
<tr>
<td>c</td>
<td>number of clusters to be estimated over X</td>
</tr>
<tr>
<td>no.dim</td>
<td>number of dimensions</td>
</tr>
<tr>
<td>k</td>
<td>tuning parameter</td>
</tr>
<tr>
<td>if.impute</td>
<td>should I traspose the input data?</td>
</tr>
<tr>
<td>normalize</td>
<td>should I normalize the input data?</td>
</tr>
<tr>
<td>cores.ratio</td>
<td>ratio of the number of cores to be used when computing the multi-kernel</td>
</tr>
</tbody>
</table>

**Value**

clusters the cells based on SIMLR and their similarities

list of 8 elements describing the clusters obtained by SIMLR, of which y are the resulting clusters:

- y = results of k-means clusterings,
- S = similarities computed by SIMLR,
- F = results from network diffusion,
- ydata = data referring the the results by k-means,
- alphaK = clustering coefficients,
- execution.time = execution time of the present run,
- converge = iterative convergence values by T-SNE,
- LF = parameters of the clustering |

**Examples**

```r
SIMLR(X = BuettnerFlorian$in_X, c = BuettnerFlorian$n_clust, cores.ratio = 0)
```

```r
library(scran)
ncells = 50
ngenes = 25
mu <- 2*runif(ngenes, 3, 10)
gene.counts <- matrix(rnbinom(ngenes*ncells, mu=mu, size=2), nrow=ngenes)ownames(gene.counts) = paste0("X", seq_len(ngenes))
sce = newSCESet(countData=data.frame(gene.counts))
output = SIMLR(X = sce, c = 8, cores.ratio = 0)
```

**Description**

perform the SIMLR feature ranking algorithm. This takes as input the original input data and the corresponding similarity matrix computed by SIMLR

**Usage**

```r
SIMLR_Feature_Ranking(A, X)
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>an (n x n) similarity matrix by SIMLR</td>
</tr>
<tr>
<td>X</td>
<td>an (m x n) data matrix of gene expression measurements of individual cells</td>
</tr>
</tbody>
</table>

**SIMLR Feature Ranking**

**Usage**

```r
SIMLR_Feature_Ranking(A, X)
```
Value

- a list of 2 elements: pvalues and ranking ordering over the n covariates as estimated by the method

Examples

SIMLR_Feature_Ranking(A = BuettnerFlorian$results$, X = BuettnerFlorian$in_X)
Index

BuettnerFlorian, 2
SIMLR, 2
SIMLR_Feature_Ranking, 3