Package: doubletrs (via r-universe)

October 16, 2026

Title Memory-Efficient Doublet Detection for Single-Cell RNA Sequencing

Version 0.1.0

Description Detects doublets in single-cell RNA sequencing data with the 'DoubletFinder' method of McGinnis, Murrow and Gartner (2019) <doi:10.1016/j.cels.2019.03.003>, implemented in 'Rust' to use less memory and time. Provides artificial doublet simulation, exact nearest-neighbour search and pANN (proportion of artificial nearest neighbours) scoring, and reproduces 'DoubletFinder' scores exactly.

License MIT + file LICENSE

URL https://github.com/mayankgandhi13/doublet-rs

BugReports https://github.com/mayankgandhi13/doublet-rs/issues

SystemRequirements Cargo (Rust's package manager), rustc >= 1.71

Encoding UTF-8

Imports graphics, KernSmooth, Matrix, methods, parallel, stats

Suggests knitr, rmarkdown, Seurat (>= 5.0.0), SeuratObject (>= 5.0.0), testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/version 8.1.0

NeedsCompilation yes

Author Mayank Gandhi [aut, cre, cph], Authors of the dependency Rust crates [cph] (see inst/AUTHORS for the bundled crates, their authors and licenses)

Maintainer Mayank Gandhi <135.mayankgandhi@gmail.com>

Config/pak/sysreqs libclang-dev

Repository https://cran.r-universe.dev

Date/Publication 2026-10-16 14:40:08 UTC

RemoteUrl https://github.com/cran/doubletrs

RemoteRef HEAD

RemoteSha 8223523bee5d0c06704cf6837b42bc0ebcc3ec66

Contents

compute_pann 2
doubletFinder 3
doubletrs_threads 4
find.pK 4
find_neighbors 5
modelHomotypic 6
paramSweep 6
simulate_doublets 7
summarizeSweep 8

Index **9**

`compute_pann` *Compute pANN scores*

Description

The proportion of artificial nearest neighbours: for each real cell, the fraction of its `k` nearest neighbours that are artificial doublets (rows after `n_real` in `pcs`).

Usage

```
compute_pann(pcs, n_real, k, threads = doubletrs_threads())
```

Arguments

`pcs`	Numeric matrix, points x PCs, with the real cells in the first `n_real` rows and artificial doublets after them.
`n_real`	Number of real cells.
`k`	Number of neighbours per cell (excluding the cell itself).
`threads`	Number of threads to use. See `doubletrs_threads()`.

Value

Numeric vector of length `n_real`, named by `rownames(pcs)`.

Examples

```
pcs <- matrix(rnorm(200), ncol = 2)
pann <- compute_pann(pcs, n_real = 80, k = 5)
summary(pann)
```

doubletFinder *Detect doublets in a Seurat object*

Description

Drop-in replacement for DoubletFinder::doubletFinder(). Simulates artificial doublets, preprocesses the merged real and artificial cells with the parameters already used on seu, computes pANN (the proportion of artificial nearest neighbours) for every cell, and calls the nExp highest-scoring cells doublets. With the same seed, results are identical to DoubletFinder's, but the dense cells x cells distance matrix is never built.

Usage

```
doubletFinder(
  seu,
  PCs,
  pN = 0.25,
  pK,
  nExp,
  reuse.pANN = NULL,
  sct = FALSE,
  annotations = NULL,
  threads = doubletrs_threads()
)
```

Arguments

seu	A Seurat object on which NormalizeData(), FindVariableFeatures(), ScaleData() and RunPCA() have been run (or SCTransform() and RunPCA() when sct = TRUE).
PCs	Principal components to use, e.g. 1:10.
pN	Artificial doublets as a proportion of the merged real and artificial data.
pK	Neighbourhood size as a proportion of the merged data. Choose it with paramSweep(), summarizeSweep() and find.pK().
nExp	Number of cells to call doublets, e.g. the expected doublet rate times the number of cells, adjusted with modelHomotypic().
reuse.pANN	Name of an existing pANN metadata column. If given, no pANN is computed and only new calls for nExp are made.
sct	Whether seu was preprocessed with SCTransform().
annotations	Not supported yet; must be NULL.
threads	Number of threads to use. See doubletrs_threads().

Value

seu with metadata columns pANN_<pN>_<pK>_<nExp> and DF.classifications_<pN>_<pK>_<nExp> ("Singlet" or "Doublet"), named as DoubletFinder names them.

Examples

```
seu <- SeuratObject::pbmc_small
set.seed(1)
seu <- doubletFinder(seu, PCs = 1:5, pN = 0.25, pK = 0.1, nExp = 5)
table(seu$DF.classifications_0.25_0.1_5)
```

`doubletrs_threads` *Number of threads used by doubletrs*

Description

The default for every `threads` argument. It is `getOption("doubletrs.threads")` if set, otherwise the number of CPU cores, capped by the `OMP_THREAD_LIMIT` environment variable if set. During `R CMD check --as-cran` it is at most 2, as CRAN policy requires.

Usage

```
doubletrs_threads()
```

Value

A single positive integer.

Examples

```
doubletrs_threads()
old <- options(doubletrs.threads = 1)
doubletrs_threads()
options(old)
```

`find.pK` *Choose pK from a parameter sweep*

Description

Drop-in replacement for `DoubletFinder::find.pK()`: the mean-variance normalised bimodality coefficient (BCmvn) for each pK across all pN. The pK with the highest `BCmetric` is DoubletFinder's recommended choice.

Usage

```
find.pK(sweep.stats, plot = TRUE)
```

Arguments

sweep.stats	Output of summarizeSweep().
plot	Whether to plot BCmvn against pK, as DoubletFinder does.

Value

A data frame with columns ParamID, pK, MeanBC, VarBC and BCmetric (plus MeanAUC when the sweep was summarised with ground truth).

---

find_neighbors *Find nearest neighbours of real cells*

Description

Exact k-nearest-neighbour search in PC space, giving the same neighbours as DoubletFinder.

Usage

```
find_neighbors(pcs, n_real, k, threads = doubletrs_threads())
```

Arguments

pcs	Numeric matrix, points x PCs, with the real cells in the first n_real rows and artificial doublets after them.
n_real	Number of real cells.
k	Number of neighbours per cell (excluding the cell itself).
threads	Number of threads to use. See doubletrs_threads().

Value

Integer matrix, n_real x k: 1-based row indices of each real cell's neighbours in pcs, nearest first.

Examples

```
pcs <- matrix(rnorm(200), ncol = 2)
nn <- find_neighbors(pcs, n_real = 80, k = 5)
dim(nn)
```

modelHomotypic *Estimate the proportion of homotypic doublets*

Description

Drop-in replacement for `DoubletFinder::modelHomotypic()`: the expected fraction of doublets formed by two cells of the same type, given cell type annotations. Homotypic doublets are hard to detect, so the expected doublet count is often scaled by `1 - modelHomotypic(annotations)`.

Usage

```
modelHomotypic(annotations)
```

Arguments

`annotations`	Cell type or cluster label for each cell.

Value

A number between 0 and 1: the sum of squared type frequencies.

Examples

```
modelHomotypic(c("T", "T", "B", "B"))
```

paramSweep *pN-pK parameter sweep*

Description

Drop-in replacement for `DoubletFinder::paramSweep()`. For each pN from 0.05 to 0.3, simulates doublets once and computes pANN at every pK from 0.0005 to 0.3. Each cell's neighbours are found once, up to the largest pK, instead of sorting all distances. Datasets with more than 10,000 cells are down-sampled to 10,000, as in DoubletFinder.

Usage

```
paramSweep(
  seu,
  PCs = 1:10,
  sct = FALSE,
  num.cores = 1,
  threads = doubletrs_threads()
)
```

Arguments

seu	A Seurat object on which NormalizeData(), FindVariableFeatures(), ScaleData() and RunPCA() have been run (or SCTransform() and RunPCA() when sct = TRUE).
PCs	Principal components to use, e.g. 1:10.
sct	Whether seu was preprocessed with SCTransform().
num.cores	Accepted for compatibility with DoubletFinder and ignored; use threads.
threads	Number of threads to use. See doubletrs_threads().

Value

A named list with one data frame (column pANN, one row per cell) per pN-pK pair, named pN_<pN>_pK_<pK>, for summarizeSweep().

Examples

```
seu <- SeuratObject::pbmc_small
set.seed(1)
sweep <- paramSweep(seu, PCs = 1:5)
stats <- summarizeSweep(sweep)
bcmvn <- find.pK(stats)
```

---

simulate_doublets *Simulate artificial doublets*

Description

Builds n artificial doublets by averaging pairs of cells drawn independently and with replacement, as DoubletFinder does. Parent cells are drawn with R's random number generator, in the same order as DoubletFinder, so set.seed() makes results reproducible and identical to DoubletFinder's. The count matrix is read directly from R memory; it is not copied.

Usage

```
simulate_doublets(counts, n, threads = doubletrs_threads())
```

Arguments

counts	Genes x cells count matrix, ideally a dgCMatrix. Other matrix types are converted to one first.
n	Number of doublets to simulate.
threads	Number of threads to use. See doubletrs_threads().

Value

A genes x n dgCMatrix of doublets, with an attribute "parents": a data frame of the two parent cell indices (1-based columns of counts) for each doublet.

Examples

```
counts <- abs(Matrix::rsparsematrix(100, 50, density = 0.1))
set.seed(1)
doublets <- simulate_doublets(counts, n = 20)
dim(doublets)
head(attr(doublets, "parents"))
```

---

summarizeSweep *Summarise a parameter sweep*

Description

Drop-in replacement for DoubletFinder::summarizeSweep(): the bimodality coefficient of each pANN distribution (from a Gaussian kernel density estimate), and optionally the AUC against known doublets.

Usage

```
summarizeSweep(sweep.list, GT = FALSE, GT.calls = NULL)
```

Arguments

sweep.list	Output of paramSweep().
GT	Whether ground-truth doublet calls are available.
GT.calls	If GT = TRUE, a vector of "Singlet" / "Doublet" for each cell, in the order of the sweep.

Value

A data frame with columns pN, pK (factors), BCreal, and AUC when GT = TRUE, for find.pK().

Index

compute_pann, 2

doubletFinder, 3
doubletrs_threads, 4
doubletrs_threads(), *2*, *3*, *5*, *7*

find.pK, 4
find.pK(), *3*, *8*
find_neighbors, 5

modelHomotypic, 6
modelHomotypic(), *3*

paramSweep, 6
paramSweep(), *3*, *8*

simulate_doublets, 7
summarizeSweep, 8
summarizeSweep(), *3*, *5*, *7*