

Package ‘PubMatrixR’

August 21, 2026

Title PubMed Pairwise Co-Occurrence Matrix Construction and Visualization

Version 1.0.1

Description Queries the 'NCBI' (National Center for Biotechnology Information) Entrez 'E-utilities' API to count pairwise co-occurrences between two sets of terms in 'PubMed' or 'PubMed Central'. It returns a matrix-like data frame of publication counts and can export hyperlink-enabled results in CSV or ODS format. The package also provides heatmap helpers for exploratory visualization of overlap patterns. Based on the method described in Becker et al. (2003) ``PubMatrix: a tool for multiplex literature mining" <doi:10.1186/1471-2105-4-61>.

Depends R (>= 4.1.0)

License MIT + file LICENSE

Encoding UTF-8

URL <https://github.com/ToledoEM/PubMatrixR-v2>,
<https://toledoem.github.io/pubmatrixr/>

BugReports <https://github.com/ToledoEM/PubMatrixR-v2/issues>

Imports pbapply, pheatmap, readODS, xml2

Suggests covr, dplyr, ggplot2, kableExtra, knitr, rmarkdown, testthat
(>= 3.0.0), withr

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Tyler Laird [aut],
Enrique Toledo [aut, cre] (ORCID:
<https://orcid.org/0000-0002-1460-4708>)

Maintainer Enrique Toledo <enriquetoledo@gmail.com>

Repository CRAN

Date/Publication 2026-08-21 05:46:09 UTC

Contents

plot_pubmatrix_heatmap	2
PubMatrix	4
pubmatrix_heatmap	5

Index	7
-------	---

plot_pubmatrix_heatmap	<i>Create a formatted heatmap from PubMatrix results</i>
------------------------	--

Description

This function creates a heatmap displaying overlap percentages derived from a PubMatrix result matrix, with Euclidean distance clustering for rows and columns.

Usage

```
plot_pubmatrix_heatmap(  
  matrix,  
  values = c("raw", "row_pct", "relative"),  
  title = "PubMatrix Co-occurrence Heatmap",  
  cluster_rows = TRUE,  
  cluster_cols = TRUE,  
  show_numbers = TRUE,  
  color_palette = NULL,  
  filename = NULL,  
  width = 10,  
  height = 8,  
  cellwidth = NA,  
  cellheight = NA,  
  scale_font = TRUE  
)
```

Arguments

matrix	A data frame or matrix from PubMatrix results containing publication co-occurrence counts
values	Character scalar selecting what is plotted in each cell. One of "raw" (default, the co-occurrence counts as returned by [PubMatrix()]), "row_pct" (each cell as a percentage of its row total), or "relative" (see Details).
title	Character string for the heatmap title. Default is "PubMatrix Co-occurrence Heatmap"
cluster_rows	Logical value determining if rows should be clustered using Euclidean distance. Default is TRUE

cluster_cols	Logical value determining if columns should be clustered using Euclidean distance. Default is TRUE
show_numbers	Logical value determining if overlap percentage values should be displayed in cells. Default is TRUE
color_palette	Color palette for the heatmap. Default uses a red gradient color scale
filename	Optional filename to save the heatmap. If NULL, displays the plot
width	Width of saved plot in inches. Default is 10
height	Height of saved plot in inches. Default is 8
cellwidth	Optional numeric cell width for pheatmap (in pixels). Default 'NA' lets pheatmap auto-size.
cellheight	Optional numeric cell height for pheatmap (in pixels). Default 'NA' lets pheatmap auto-size.
scale_font	Logical value determining if font size should scale with cell size. Default is TRUE

Details

Rows and columns are clustered with Euclidean distance. NA values in the input matrix are converted to 0 before any calculation.

The 'values' argument controls what each cell shows:

- "raw" (default) - the co-occurrence counts themselves.
- "row_pct" - each count as a percentage of its row total, useful for comparing how one row's attention is distributed across columns.
- "relative" - $\text{count} / (\text{row_total} + \text{col_total} - \text{count}) * 100$, where the totals are sums over the supplied matrix.

Note that "relative" is **not** a Jaccard index and is **not** comparable across runs. Its totals are sums over whichever partner terms happen to be in the matrix, not the marginal publication counts for each term, so adding an unrelated term changes the value reported for every existing cell. A true Jaccard index would require the single-term counts, which [PubMatrix()] does not fetch. Use it only to compare cells within one fixed matrix.

Value

A pheatmap object (invisible). The matrix actually plotted is attached as the "plotted_values" attribute.

Examples

```
# Create a small test matrix
test_matrix <- matrix(c(1, 2, 3, 4), nrow = 2, ncol = 2)
rownames(test_matrix) <- c("Gene1", "Gene2")
colnames(test_matrix) <- c("GeneA", "GeneB")

# Create heatmap using the helper (plots the raw counts by default)
plot_pubmatrix_heatmap(test_matrix, title = "Test Heatmap")
```

```
# Percentage views are available too:
plot_pubmatrix_heatmap(test_matrix, values = "row_pct", title = "Row %")

# Equivalent using pheatmap directly:
pheatmap::pheatmap(
  test_matrix,
  main = "Test Heatmap (pheatmap)",
  color = colorRampPalette(c("#fee5d9", "#cb181d"))(100),
  display_numbers = TRUE,
  fontsize = 16,
  fontsize_number = 14,
  border_color = "lightgray",
  show_rownames = TRUE,
  show_colnames = TRUE
)
```

PubMatrix	<i>Query 'PubMed' or 'PMC' and Build a Pairwise Co-occurrence Matrix</i>
-----------	--

Description

'PubMatrix()' counts publications for all pairwise combinations of two term sets using the 'NCBI' Entrez 'E-utilities' API. It returns a matrix-like data frame with rows corresponding to terms in 'B' and columns corresponding to terms in 'A'.

Usage

```
PubMatrix(
  file = NULL,
  A = NULL,
  B = NULL,
  API.key = NULL,
  Database = "pubmed",
  daterange = NULL,
  outfile = NULL,
  export_format = NULL
)
```

Arguments

file	Optional path to a text file containing search terms. The file must contain a '#' separator line between the 'A' and 'B' term lists. Used only when 'A' and 'B' are both 'NULL'.
A	Character vector of search terms for matrix columns.
B	Character vector of search terms for matrix rows.
API.key	Optional 'NCBI' API key.

Database	Character scalar. One of "pubmed" or "pmc".
daterange	Optional numeric vector of length 2 giving 'c(start_year, end_year)'.
outfile	Optional output file stem used when 'export_format' is set.
export_format	Optional export format: "csv" or "ods".

Details

Examples and vignettes should avoid live web queries during package checks. This function performs live requests to 'NCBI' and may fail when there is no internet connectivity or when the service is unavailable.

Terms are trimmed of surrounding whitespace, and duplicate terms within 'A' or within 'B' are rejected. When reading from 'file', blank lines on either side of the '#' separator are ignored.

Requests are paced to respect the 'NCBI' rate limits (3 requests per second without an API key, 10 with one). Two options tune the request behaviour:

- 'PubMatrixR.n_tries' - number of attempts per query before giving up (default '3'). Failed attempts back off exponentially.
- 'PubMatrixR.min_interval' - seconds to wait between requests (default '0.34' without an API key, '0.11' with one). Set to '0' to disable pacing.

Value

A data frame of publication counts with rows named by 'B' and columns named by 'A'.

Examples

```
## Not run:
A <- c("WNT1", "WNT2")
B <- c("FZD1", "FZD2")
result <- PubMatrix(A = A, B = B, Database = "pubmed", daterange = c(2020, 2023))
print(result)

## End(Not run)

try(PubMatrix(A = NULL, B = NULL, file = NULL))
try(PubMatrix(A = "a", B = "b", Database = "invalid_db"))
```

pubmatrix_heatmap

Create a simple heatmap from PubMatrix results

Description

A simplified version of plot_pubmatrix_heatmap for quick visualization

Usage

```
pubmatrix_heatmap(  
  matrix,  
  title = "PubMatrix Results",  
  values = c("raw", "row_pct", "relative")  
)
```

Arguments

matrix	A numeric matrix from PubMatrix results
title	Character string for the heatmap title
values	Character scalar passed to [plot_pubmatrix_heatmap()]. Defaults to "raw" (co-occurrence counts).

Value

A pheatmap object (invisible)

Examples

```
# Create a small test matrix  
test_matrix <- matrix(c(1, 2, 3, 4), nrow = 2, ncol = 2)  
rownames(test_matrix) <- c("Gene1", "Gene2")  
colnames(test_matrix) <- c("GeneA", "GeneB")  
  
# Create simple heatmap (wrapper)  
pubmatrix_heatmap(test_matrix, title = "Simple Test Heatmap")  
  
# Equivalent pheatmap call  
pheatmap::pheatmap(  
  test_matrix,  
  main = "Simple Test Heatmap (pheatmap)",  
  color = colorRampPalette(c("#fee5d9", "#cb181d"))(100),  
  display_numbers = TRUE,  
  fontsize = 16,  
  fontsize_number = 14  
)
```

Index

`plot_pubmatrix_heatmap`, [2](#)

`PubMatrix`, [4](#)

`pubmatrix_heatmap`, [5](#)