Package 'xxdi'

June 22, 2024

Type Package

Title Calculate Expertise Indices

Version 1.1.0

Maintainer Nilabhra Rohan Das <nr.das@yahoo.com>

Description Institutional performance assessment remains a key challenge to a multitude of stakeholders. Existing indicators such as h-type indicators, g-type indicators, and many others do not reflect expertise of institutions that defines their research portfolio. The package offers functionality to compute and visualise two novel indices: the x-index and the xd-index. The x-index evaluates an institution's scholarly expertise within a specific discipline or field, while the xd-index provides a broader assessment of overall scholarly expertise considering an institution's publication pattern and strengths across coarse thematic areas. These indices offer a nuanced understanding of institutional research capabilities, aiding stakeholders in research management and resource allocation decisions. Lathabai, H.H., Nandy, A., and Singh, V.K. (2021) <doi:10.1007/s11192-021-04188-3>. Nandy, A., Lathabai, H.H., and Singh, V.K. (2023) <doi:10.5281/zenodo.8305585>.

License GPL-3

Depends R (>= 4.3.3.0)

Imports agop (>= 0.2.4), dplyr (>= 1.1.4), ggplot2 (>= 3.5.0), Matrix (>= 1.6.1.1), stats (>= 4.3.3), tidyr (>= 1.3.1)

Encoding UTF-8

RoxygenNote 7.3.1

NeedsCompilation no

Author Nilabhra Rohan Das [cre, aut] (<https://orcid.org/0000-0001-8187-0080>),
Abhirup Nandy [aut] (<https://orcid.org/0000-0001-8618-0847>)

Repository CRAN

Date/Publication 2024-06-21 23:30:02 UTC

Contents

---

g_index   *g_index*

Description

This function calculates the g-index for an institution using bibliometric data from an edge list, with an optional plot visualisation.

Usage

```
g_index(df, id, cit, plot = FALSE)
```

Arguments

df	Data frame object containing bibliometric data. This data frame must have at least three columns: one for keywords, one for unique IDs, and one for citation counts. Each row in the data frame should represent a document or publication.
id	Character string specifying the name of the column in "df" that contains unique identifiers for each document. Each cell in this column must contain a single ID (unless missing) and not multiple IDs.
cit	Character string specifying the name of the column in "df" that contains the number of citations each document has received. Citations must be represented as integers. Each cell in this column should contain a single integer value (unless missing) representing the citation count for the corresponding document.
plot	Logical value indicating whether to generate and display a plot of the g-index calculation. Set to "TRUE" or "T" to generate the plot, and "FALSE" or "F" to skip plot generation. The default is "FALSE".

Value

g-index value and plot for institution.

Examples

```
# Create an example data frame
dat1 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a; b; c", "b; d", "c", "d", "e; g", "f", "g"),
             id = c("abc123", "bcd234", "def345", "efg456", "fgh567", "ghi678", "hij789"),
                   categories = c("a; d; e", "b", "c", "d; g", "e", "f", "g"))
# Calculate g-index
g_index(df = dat1, id = "id", cit = "citations")
```

```
# Create another example data frame
dat2 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a/ b/ c", "b/ d", "c", "d", "e/ g", "f", "g"),
                   id = c("123", "234", "345", "456", "567", "678", "789"),
                   categories = c("a/ d/ e", "b", "c", "d/ g", "e", "f", "g"))
# Calculate g-index
g_index(df = dat2, id = "id", cit = "citations", plot = FALSE)

# Create another example data frame
dat3 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a, b, c", "b, d", "c", "d", "e, g", "f", "g"),
                   id = c(123, 234, 345, 456, 567, 678, 789),
                   categories = c("a: d: e", "b", "c", "d: g", "e", "f", "g"))
# Calculate g-index and produce plot
g_index(df = dat3, id = "id", cit = "citations", plot = TRUE)
```

---

h_index   *h_index*

Description

This function calculates the h-index for an institution using bibliometric data from an edge list, with an optional plot visualisation.

Usage

```
h_index(df, id, cit, plot = FALSE)
```

Arguments

df	Data frame object containing bibliometric data. This data frame must have at least three columns: one for keywords, one for unique IDs, and one for citation counts. Each row in the data frame should represent a document or publication.
id	Character string specifying the name of the column in "df" that contains unique identifiers for each document. Each cell in this column must contain a single ID (unless missing) and not multiple IDs.
cit	Character string specifying the name of the column in "df" that contains the number of citations each document has received. Citations must be represented as integers. Each cell in this column should contain a single integer value (unless missing) representing the citation count for the corresponding document.
plot	Logical value indicating whether to generate and display a plot of the h-index calculation. Set to "TRUE" or "T" to generate the plot, and "FALSE" or "F" to skip plot generation. The default is "FALSE".

Value

h-index value and plot for institution.

Examples

```
# Create an example data frame
dat1 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a; b; c", "b; d", "c", "d", "e; g", "f", "g"),
              id = c("abc123", "bcd234", "def345", "efg456", "fgh567", "ghi678", "hij789"),
                   categories = c("a; d; e", "b", "c", "d; g", "e", "f", "g"))
# Calculate h-index
h_index(df = dat1, id = "id", cit = "citations")

# Create another example data frame
dat2 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a/ b/ c", "b/ d", "c", "d", "e/ g", "f", "g"),
                   id = c("123", "234", "345", "456", "567", "678", "789"),
                   categories = c("a/ d/ e", "b", "c", "d/ g", "e", "f", "g"))
# Calculate h-index
h_index(df = dat2, id = "id", cit = "citations", plot = FALSE)

# Create another example data frame
dat3 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a, b, c", "b, d", "c", "d", "e, g", "f", "g"),
                   id = c(123, 234, 345, 456, 567, 678, 789),
                   categories = c("a: d: e", "b", "c", "d: g", "e", "f", "g"))
# Calculate h-index and produce plot
h_index(df = dat3, id = "id", cit = "citations", plot = TRUE)
```

---

xc_index	*xc_index*

Description

This is a standalone function that specifically calculates the xc-index for an institution using bibliometric data from an edge list, with an optional plot visualisation. The function is suitable for including inside loops when plotting parameter is set to "FALSE" or "F".

Usage

```
xc_index(df, kw, cat, id, cit, kdlm = ";", cdlm = ";", plot = FALSE)
```

Arguments

df	Data frame object containing bibliometric data. This data frame must have at least three columns: one for keywords, one for unique IDs, and one for citation counts. Each row in the data frame should represent a document or publication.
kw	Character string specifying the name of the column in "df" that contains keywords. Each cell in this column may contain no keywords (missing), a single keyword or multiple keywords separated by a specified delimiter.
cat	Character string specifying the name of the column in "df" that contains categories. Each cell in this column may contain no categories (missing), a single category or multiple categories separated by a specified delimiter.

id	Character string specifying the name of the column in "df" that contains unique identifiers for each document. Each cell in this column must contain a single ID (unless missing) and not multiple IDs.
cit	Character string specifying the name of the column in "df" that contains the number of citations each document has received. Citations must be represented as integers. Each cell in this column should contain a single integer value (unless missing) representing the citation count for the corresponding document.
kdlm	Character string specifying the delimiter used in the "kw" column to separate multiple keywords within a single cell. The delimiter should be consistent across the entire "kw" column. Common delimiters include ";", "/", ":", and ",". The default delimiter is set to ";".
cdlm	Character string specifying the delimiter used in the "cat" column to separate multiple categories within a single cell. The delimiter should be consistent across the entire "cat" column. Common delimiters include ";", "/", ":", and ",". The default delimiter is set to ";".
plot	Logical value indicating whether to generate and display a plot of the xc-index calculation. Set to "TRUE" or "T" to generate the plot, and "FALSE" or "F" to skip plot generation. The default is "FALSE".

Value

xc-index value and plot for institution.

Examples

```
# Create an example data frame
dat1 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a; b; c", "b; d", "c", "d", "e; g", "f", "g"),
            id = c("abc123", "bcd234", "def345", "efg456", "fgh567", "ghi678", "hij789"),
                   categories = c("a; d; e", "b", "c", "d; g", "e", "f", "g"))
# Calculate xc-index and produce plot
xc_index(df = dat1, kw = "keywords", cat = "categories", id = "id", cit = "citations", plot = TRUE)
```

---

xdc_index   *xdc_index*

Description

This is a general function that calculates the x-index, xd-index, and xc-index for an institution using bibliometric data from an edge list. Returns the x-index or the xd-index depending on the input vector in "p1". Returns x-index when "p1" is the dataframe column containing keywords and xd-index when "p1" the dataframe column containing categories. Returns a summary table listing x-index, xd-index, and xc-index when both "p1" and "p2" are supplied. In this case, "p2" must be the higher level, i.e., categories and "p1" must be the lower level, i.e., keywords. The function is suitable for including inside loops when only one input vector is provided. However, for looping xc-index, a separate 'xc_index' function is provided which also includes an option to produce plots. Similarly, 'x_index' and 'xd_index' are standalone functions for calculating x-index and xd-index respectively as well as producing plots.

Usage

```
xdc_index(df, p1, p2 = NULL, id, cit, dlm1 = ";", dlm2 = ";")
```

Arguments

df	Data frame object containing bibliometric data. This data frame must have at least three columns: one for keywords/categories, one for unique IDs, and one for citation counts. Each row in the data frame should represent a document or publication.
p1	Character string specifying the name of the column in "df" that contains the first parameter of interest. Generally, this will be a column for keywords/categories. Each cell in this column may contain no keywords/categories (missing), a single keyword/category or multiple keywords/categories separated by a specified delimiter. Input the column for keywords to calculate the x-index, and the column for categories to calculate the xd-index.
p2	Character string specifying the name of the column in "df" that contains the second parameter of interest. This is an optional parameter only required when calculating xc-index. When an input is provided, it is expected that "p1" is the lower level, generally keywords, and "p2" is the higher level, generally categories. Each cell in this column may contain no categories (missing), a single category or multiple categories separated by a specified delimiter. The default is set to "NULL".
id	Character string specifying the name of the column in "df" that contains unique identifiers for each document. Each cell in this column must contain a single ID (unless missing) and not multiple IDs.
cit	Character string specifying the name of the column in "df" that contains the number of citations each document has received. Citations must be represented as integers. Each cell in this column should contain a single integer value (unless missing) representing the citation count for the corresponding document.
dlm1	Character string specifying the delimiter used in the "p1" column to separate multiple keywords/categories within a single cell. The delimiter should be consistent across the entire "p1" column. Common delimiters include ";", "/", ":", and ",". The default delimiter is set to ";".
dlm2	Character string specifying the delimiter used in the "p2" column to separate multiple categories within a single cell. The delimiter should be consistent across the entire "p2" column. Common delimiters include ";", "/", ":", and ",". The default delimiter is set to ";".

Value

x-index value, and/or xd-index value, and/or xc-index value for institution.

Examples

```
# Create an example data frame
dat1 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                    keywords = c("a; b; c", "b; d", "c", "d", "e; g", "f", "g"),
```

```
                 id = c("abc123", "bcd234", "def345", "efg456", "fgh567", "ghi678", "hij789"),
                      categories = c("a; d; e", "b", "c", "d; g", "e", "f", "g"))
# Calculate x-index
xdc_index(df = dat1, p1 = "keywords", id = "id", cit = "citations")
# Calculate xd-index
xdc_index(df = dat1, p1 = "categories", id = "id", cit = "citations")
# Calculate x-index, xd-index, and xc-index together
xdc_index(df = dat1, p1 = "keywords", p2 = "categories", id = "id", cit = "citations")
```

---

xd_index   *xd_index*

Description

This is a standalone function that specifically calculates the xd-index for an institution using bibliometric data from an edge list, with an optional plot visualisation. The function is suitable for including inside loops when plotting parameter is set to "FALSE" or "F".

Usage

```
xd_index(df, cat, id, cit, dlm = ";", plot = FALSE)
```

Arguments

df	Data frame object containing bibliometric data. This data frame must have at least three columns: one for categories, one for unique IDs, and one for citation counts. Each row in the data frame should represent a document or publication.
cat	Character string specifying the name of the column in "df" that contains categories. Each cell in this column may contain no categories (missing), a single category or multiple categories separated by a specified delimiter.
id	Character string specifying the name of the column in "df" that contains unique identifiers for each document. Each cell in this column must contain a single ID (unless missing) and not multiple IDs.
cit	Character string specifying the name of the column in "df" that contains the number of citations each document has received. Citations must be represented as integers. Each cell in this column should contain a single integer value (unless missing) representing the citation count for the corresponding document.
dlm	Character string specifying the delimiter used in the "cat" column to separate multiple categories within a single cell. The delimiter should be consistent across the entire "cat" column. Common delimiters include ";", "/", ":", and ",". The default delimiter is set to ";".
plot	Logical value indicating whether to generate and display a plot of the xd-index calculation. Set to "TRUE" or "T" to generate the plot, and "FALSE" or "F" to skip plot generation. The default is "FALSE".

Value

xd-index value and plot for institution.

Examples

```
# Create an example data frame
dat1 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                    keywords = c("a; b; c", "b; d", "c", "d", "e; g", "f", "g"),
                id = c("abc123", "bcd234", "def345", "efg456", "fgh567", "ghi678", "hij789"),
                    categories = c("a; d; e", "b", "c", "d; g", "e", "f", "g"))
# Calculate xd-index
xd_index(df = dat1, cat = "categories", id = "id", cit = "citations")

# Create another example data frame
dat2 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a/ b/ c", "b/ d", "c", "d", "e/ g", "f", "g"),
                   id = c("123", "234", "345", "456", "567", "678", "789"),
                   categories = c("a/ d/ e", "b", "c", "d/ g", "e", "f", "g"))
# Calculate xd-index
xd_index(df = dat2, cat = "categories", id = "id", cit = "citations", dlm = "/", plot = FALSE)

# Create another example data frame
dat3 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a, b, c", "b, d", "c", "d", "e, g", "f", "g"),
                   id = c(123, 234, 345, 456, 567, 678, 789),
                   categories = c("a: d: e", "b", "c", "d: g", "e", "f", "g"))
# Calculate xd-index and produce plot
xd_index(df = dat3, cat = "categories", id = "id", cit = "citations", dlm = ":", plot = TRUE)
```

---

x_index   *x_index*

Description

This is a standalone function that specifically calculates the x-index for an institution using bibliometric data from an edge list, with an optional plot visualisation. The function is suitable for including inside loops when plotting parameter is set to "FALSE" or "F".

Usage

```
x_index(df, kw, id, cit, dlm = ";", plot = FALSE)
```

Arguments

df   Data frame object containing bibliometric data. This data frame must have at
                  least three columns: one for keywords, one for unique IDs, and one for citation
                  counts. Each row in the data frame should represent a document or publication.

kw
: Character string specifying the name of the column in "df" that contains keywords. Each cell in this column may contain no keywords (missing), a single keyword or multiple keywords separated by a specified delimiter.

id
: Character string specifying the name of the column in "df" that contains unique identifiers for each document. Each cell in this column must contain a single ID (unless missing) and not multiple IDs.

cit
: Character string specifying the name of the column in "df" that contains the number of citations each document has received. Citations must be represented as integers. Each cell in this column should contain a single integer value (unless missing) representing the citation count for the corresponding document.

dlm
: Character string specifying the delimiter used in the "kw" column to separate multiple keywords within a single cell. The delimiter should be consistent across the entire "kw" column. Common delimiters include ";", "/", ":", and ",". The default delimiter is set to ";".

plot
: Logical value indicating whether to generate and display a plot of the x-index calculation. Set to "TRUE" or "T" to generate the plot, and "FALSE" or "F" to skip plot generation. The default is "FALSE".

Value

x-index value and plot for institution.

Examples

```
# Create an example data frame
dat1 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a; b; c", "b; d", "c", "d", "e; g", "f", "g"),
              id = c("abc123", "bcd234", "def345", "efg456", "fgh567", "ghi678", "hij789"),
                   categories = c("a; d; e", "b", "c", "d; g", "e", "f", "g"))
# Calculate x-index
x_index(df = dat1, kw = "keywords", id = "id", cit = "citations")

# Create another example data frame
dat2 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a/ b/ c", "b/ d", "c", "d", "e/ g", "f", "g"),
                   id = c("123", "234", "345", "456", "567", "678", "789"),
                   categories = c("a/ d/ e", "b", "c", "d/ g", "e", "f", "g"))
# Calculate x-index
x_index(df = dat2, kw = "keywords", id = "id", cit = "citations", dlm = "/", plot = FALSE)

# Create another example data frame
dat3 <- data.frame(citations = c(0, 1, 1, 2, 3, 5, 8),
                   keywords = c("a, b, c", "b, d", "c", "d", "e, g", "f", "g"),
                   id = c(123, 234, 345, 456, 567, 678, 789),
                   categories = c("a: d: e", "b", "c", "d: g", "e", "f", "g"))
# Calculate x-index and produce plot
x_index(df = dat3, kw = "keywords", id = "id", cit = "citations", dlm = ",", plot = TRUE)
```

Index