Package 'hindex'

February 22, 2020

Title Simulating the Development of h-Index Values

Version 0.2.0

Description H-index and h-alpha are a bibliometric indicators. This package provides functions to simulate how these indicators may develop over time for a given set of researchers and to visualize the simulation data.
The implementation is based on the 'STATA' ado h-index and is described in more detail in Bornmann et al. (2019) <arXiv:1905.11052>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Suggests testthat

Imports foreach, stats, ggplot2, purrr

RoxygenNote 7.0.2

NeedsCompilation no

Author Alexander Tekles [aut, cre],
Lutz Bornmann [ctb],
Christian Ganser [ctb]

Maintainer Alexander Tekles <alexander.tekles@soziologie.uni-muenchen.de>

Repository CRAN

Date/Publication 2020-02-22 22:20:02 UTC

R topics documented:

plot_hsim *Plot the result of simulate_hindex*

Description

Plot the result of a simulation computed by simulate_hindex.

Usage

```
plot_hsim(
  simdata,
  plot_hindex = FALSE,
  plot_halpha = FALSE,
  plot_toppapers = FALSE,
  plot_mindex = FALSE,
  subgroups = FALSE,
  group_boundaries = NULL,
  exclude_group_boundaries = FALSE,
  plot_group_diffs = FALSE
)
```

Arguments

simdata	The result of a simulation returned by simulate_hindex.
plot_hindex	If this parameter is set to TRUE, the h-index values are plotted.
plot_halpha	If this parameter is set to TRUE, the h-alpha values are plotted.
plot_toppapers	If this parameter is set to TRUE, the numbers of top-10% papers are plotted.
plot_mindex	If this parameter is set to TRUE, the mindex values are plotted.
subgroups	If this parameter is set to TRUE, the subgroups in simdata are considered for grouping plotting the index values separately for each of these groups.
group_boundaries	Alternative to subgroups for specifying groups of scientists for plotting the index values separately for these groups. Here, the groups are specified based on the initial h-index of the agents. group_boundaries must be a list of vectors or a vector of integers specifying the groups. If a list is specified, each element must be a vector of length 2 representing the lower and the upper bound for the initial h-index (if the boundaries are included in the corresponding intervals is specified by the exclude_group_boundaries parameter). If a vector of integers is specified, each element in group_boundaries separates two groups such that all agents with an initial h-index below this boundary (and equal to or above any lower boundary; if exclude_group_boundaries is set to TRUE, the initial h-index has to be above any lower boundary) are in the first group, and all agents with an initial h-index equal to or above this boundary (and below any higher boundary) are in the second group.

exclude_group_boundaries
: If this parameter is set to TRUE, the scientists are grouped such that those scientists whose initial h-index is equal to a boundary are not included.

plot_group_diffs
: If this parameter is specified, the difference between the groups that are specified by group_boundaries is plotted.

Value

A ggplot object (ggplot).

Examples

```
set.seed(123)
simdata <- simulate_hindex(runs = 2, n = 20, periods = 3)
plot_hsim(simdata, plot_hindex = TRUE, plot_halpha = TRUE)
```

---

simulate_hindex *Simulate h-index and h-alpha values*

Description

Simulate the effect of publishing, being cited, and (strategic) collaborating on the development of h-index and h-alpha values for a specified set of agents.

Usage

```
simulate_hindex(
  runs = 1,
  n = 100,
  periods = 20,
  subgroups_distr = 1,
  subgroup_advantage = 1,
  subgroup_exchange = 0,
  init_type = "fixage",
  distr_initial_papers = "poisson",
  max_age_scientists = 5,
  dpapers_pois_lambda = 2,
  dpapers_nbinom_dispersion = 1.1,
  dpapers_nbinom_mean = 2,
  productivity = 80,
  distr_citations = "poisson",
  dcitations_speed = 2,
  dcitations_peak = 3,
  dcitations_mean = 2,
  dcitations_dispersion = 1.1,
  coauthors = 5,
  strategic_teams = FALSE,
```

```
  diligence_share = 1,
  diligence_corr = 0,
  selfcitations = FALSE,
  update_alpha_authors = FALSE,
  boost = FALSE,
  boost_size = 0.1,
  alpha_share = 0.33
)
```

Arguments

runs	Number of times the simulation is repeated.
n	Number of agents acting in each simulation.
periods	Number of periods the agents collaborate across in each period.
subgroups_distr	Share of scientists in the first subgroup among all scientists
subgroup_advantage	Factor by which citations of papers published by agents of subgroup 2 exceed those of papers published by subgroup 1. This option is intended to reflect subdisciplines with different citation levels.
subgroup_exchange	Share of agents publishing (alone or in collaboration) with the other subgroup in each period. For example, when specifying subgroup_exchange = .1, 10% of each subgroup join the other subgroup each period.
init_type	Type of the initial setup. May be 'fixage' or 'varage'. For init_type = 'fixage', all initial papers have the same age (specified by max_age_scientists). For init_type = 'varage', papers get a random age which is less than or equal to max_age_scientists.
distr_initial_papers	Distribution of the papers the scientists have already published at the start of the simulation. Currently, the poisson distribution ("poisson") and the negative binomial distribution ("nbinomial") are supported.
max_age_scientists	Maximum age of scientists at the start of the simulation. For init_type = varage, a random age less than or equal to max_age_scientists is assigned to the initial papers. For init_type = fixage, all papers are max_age_scientists old.
dpapers_pois_lambda	The distribution parameter for a poisson distribution of initial papers.
dpapers_nbinom_dispersion	Dispersion parameter of a negative binomial distribution of initial papers.
dpapers_nbinom_mean	Expected value of a negative binomial distribution of initial papers.
productivity	The share of papers published by the 20% most productive agents in percentage. This parameter is only used for init_type = 'varage'. For init_type = 'fixage', diligence_share and diligence_corr can be used to control the productivity of scientists.

`distr_citations`
: Distribution of citations the papers get. The expected value of this distribution follows a log-logistic function of time. Currently, the poisson distribution ("poisson") and the negative binomial distribution ("nbinomial") are supported.

`dcitations_speed`
: The steepness (shape parameter) of the log-logistic time function of the expected citation values.

`dcitations_peak`
: The period after publishing when the expected value of the citation distribution reaches its maximum.

`dcitations_mean`
: The maximum expected value of the citation distribution (at period dcitations_peak after publishing, the citation distribution has dcitations_mean).

`dcitations_dispersion`
: For a negative binomial citation distribution, dcitations_dispersion is a factor by which the variance exceeds the expected value.

`coauthors`
: Average number of coauthors publishing papers.

`strategic_teams`
: If this parameter is set to TRUE, agents with high h-index avoid co-authorships with agents who have equal or higher h-index values (they strategically select co-authors to improve their h-alpha index). This is implemented by assigning the agents with the highest h-index values to separate teams and randomly assigning the other agents to the teams. Otherwise, the collaborating agents are assigned to co-authorships at random.

`diligence_share`
: The share of agents publishing in each period. Only used for init_type = 'fixage'.

`diligence_corr`
: The correlation between the initial h-index value and the probability to publish in a given period. This parameter only has an effect if diligence_share < 1. Only used for init_type = 'fixage'.

`selfcitations`
: If this parameter is set to TRUE, a paper gets one additional citation if at least one of its authors has a h-index value that exceeds the number of previous citations of the paper by one or two. This reflects agents strategically citing their own papers with citations just below their h-index to accelerate the growth of their h-index.

`update_alpha_authors`
: If this parameter is set to TRUE, the alpha author of newly written papers is determined every period based on the current h-index values of its authors. Without this option, the alpha author is determined when the paper is written and held constant from then on.

`boost`
: If this parameter is set to TRUE, papers of agents with a higher h-index are cited more frequently than papers of agents with lower h-index. For each team, this effect is based on the team's co-author with the highest h-index within this team.

`boost_size`
: Magnitude of the boost effect. For every additional h point of a paper's co-author who has the highest h-index among all of the paper's co-authors, citations of the paper are increased by boost_size, rounded to the next integer.

`alpha_share`
: The share of previously published papers where the corresponding agent is alpha author.

Value

For each run, the h-index values and the h-alpha values for each period are stored in a list of lists.

Examples

```
set.seed(123)
simdata <- simulate_hindex(runs = 2, n = 20, periods = 3)
plot_hsim(simdata, plot_hindex = TRUE)
```

Index