

# Package ‘hindex’

July 12, 2019

**Title** Simulating the Development of h-Index Values

**Version** 0.1.0

**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

**RoxygenNote** 6.1.1

**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** 2019-07-12 16:10:05 UTC

## R topics documented:

plot_hsim . . . . .	2
simulate_hindex . . . . .	3

<b>Index</b>	<b>5</b>
--------------	----------

---

`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,  
          group_boundaries = NULL, exclude_group_boundaries = FALSE,  
          plot_group_diffs = FALSE)
```

**Arguments**

<code>simdata</code>	The result of a simulation returned by <code>simulate_hindex</code> .
<code>plot_hindex</code>	If this parameter is set to <code>TRUE</code> , the h-index values are plotted.
<code>plot_halpha</code>	If this parameter is set to <code>TRUE</code> , the h-alpha values are plotted.
<code>group_boundaries</code>	A list of vectors or a vector of integers specifying the groups for plotting the h-index/h-alpha values separately for each of these groups. The groups are defined based on the initial h-index of the agents. 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 <code>exclude_group_boundaries</code> parameter). If a vector of integers is specified, each element in <code>group_boundaries</code> separates two groups such that all agents with an initial h-index below this boundary (and equal to or above any lower boundary; if <code>exclude_group_boundaries</code> is set to <code>TRUE</code> , 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.
<code>exclude_group_boundaries</code>	If this parameter is set to <code>TRUE</code> , the scientists are grouped such that those scientists whose initial h-index is equal to a boundary are not included.
<code>plot_group_diffs</code>	If this parameter is specified, the difference between the groups that are specified by <code>group_boundaries</code> 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	<i>Simulate h-index and h-alpha values</i>
-----------------	--

---

### 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,
  distr_initial_papers = "poisson", dpapers_pois_lambda = 2,
  dpapers_nbinom_dispersion = 1.1, dpapers_nbinom_mean = 2,
  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.
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.
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.
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.
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.
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

`ggplot`, [2](#)

`plot_hsim`, [2](#)

`simulate_hindex`, [2, 3](#)