Eur. Phys. J. B (2013) 86: 426
DOI: 10.1140/epjb/e2013-40207-0

THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

An agent-based model for the bibliometric h-index

Georgia Ionescu and Bastien Chopard[a]

Computer Science Department, University of Geneva, 1211 Genève, Switzerland

Received 13 March 2013 / Received in final form 2 July 2013
Published online 9 October 2013 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2013

Abstract. We model a virtual scientific community in which authors publish and cite articles. Citations are attributed according to a preferential attachment mechanism. From the numerical simulations, the h-index can be computed. This bottom-up approach reproduces well real bibliometric data. We consider two versions of our model. (1) The *single-scientist* is controlled by two parameters which can be tuned to reproduce the value of the h-index of many real scientists. Moreover, this model shows how the h-index grows with the number of citations, for a fixed number of articles. We also define an average h-index that can be used to compare the scientific productivity of institutions of different sizes. (2) The *multi-scientist* model considers a population of scientists and allows us to study the impact of removing citations from the low h-index researchers on the community. Simulations on real bibilometric data, as well as the predictions of the model, show that the h-index eco-system can be strongly affected by such a filtering.

1 Introduction

Bibliometric indexes are more and more used to evaluate the performance and quality of scientists through a statistical analysis of their publications. The h-index is a well-known example. It has been introduced by Hirsch in 2005 [1] as a combined measure of productivity and impact of a researcher. It is defined as the larger number h of papers published by a researcher (or an institution) that have each received at least h citations.

Reducing the capability of a scientist to a set of numbers has been criticized by many authors [2,3]. Apart from the fact that it gives little recognition to the complexity of a human being and its interaction with a social structure, it has been argued (Goodhart's law) that a metrics stops being a metrics as soon as it becomes a target [4]. In other words, a high bibliometric index may only reflect the capability of a person to make this index large. A nice example of how to artificially increase the impact factor of a journal is described in reference [5].

In the present paper we do not want to further argue on the relevance of the h-index. Rather we investigate the possibility to describe the publications and citations of the scientific community with a multi-agent model. In such an approach, a component of the model is typically a scientist producing new papers, giving citations to his own and other already published papers, and receiving citations from the community.

As opposed to the previous approaches found in the literature (e.g. [6]), we do not extract properties by analyzing bibliometric databases but we propose a bottom-up, constructive approach, based on a stochastic process, to see whether real data can be explained as the emergence of a simple behavior of the scientists involved in the community.

The key idea of our approach, described in Section 2, is that a new citation is given to an existing paper with a probability proportional to the number of citations this paper has already received. This is the cumulative advantage described in reference [7] or the mechanism of *preferential attachment*, very common in the theory of complex network [8]. We show in Section 3 that this assumption is sufficient to reproduce the distribution of citations observed in real data. Moreover, by adjusting two parameters in the model, we can accurately reproduce the h-index of a given scientist, knowing his total number of papers N and total number of citations M.

From the proposed model we show how the h-index of a group of scientists can be predicted from the number of papers and citations of each individual. This suggests a way to define an "average" h-index, which is independent of the size of the group. This offers a possibility to compare the scientific performance of institutions of different sizes.

Asymptotically our model agrees with the empirical law

$$h = \left(\frac{M^2}{4N}\right)^{1/3} \tag{1}$$

found in reference [9] and valid for large values of N and M, as obtained when aggregating all the papers from a community. However, our model also gives correct results for N and M corresponding to those of a single scientist.

[a] e-mail: bastien.chopard@unige.ch

In addition, equation (1) can possibly predict an h-index larger than N if $M \geq 2N^2$, which is impossible since, by definition, $h \leq N$.

Also, equation (1) wrongly predicts the evolution of h which is observed when M increases, with N fixed. Real data indicate (see Sect. 3) that h grows as $M^{1/2}$ whereas equation (1) predicts $M^{2/3}$. Our model, on the other hand, correctly captures the exponent 1/2.

Finally, in Section 4, we consider a numerical model, coined the *multi-scientist* model, to describe a complete scientific community. In Section 5 we study the impact of filtering out low h-index researchers from a community. By analyzing both this model and the Stanford High Energy data base, we investigate the change in h-indexes when all the papers and citations from scientists with an h-index below a given threshold are removed. Our results suggest a stratified structure of the scientific community, in which the lower h levels mostly cite papers from the upper h levels.

2 The single-scientist model

The model discussed in this section simulates the work of a single scientist (author) by generating iteratively its publications and the citations received either from himself (through self-citations) or from the external scientific community. Several parameters are needed to initialize the model: (i) the total amount N of publications of the author; (ii) the total number M of citations this author has received for the N articles; (iii) the number p of self-citations a new published article is giving to the old ones; and (iv) the number q of citations given by the external scientific world at each iteration of the process, as explained below.

In our model, the articles are created one by one. At any stage, the $n \leq N$ articles already published are described by two sets of n integers, $X = \{X_1, \ldots, X_n\}$ and $Y = \{Y_1, \ldots, Y_n\}$, where the subscripts are indexes to the articles. The value X_k represents the amount of citations received by publication k from the scientific community, and Y_k the amount of self-citation this same paper has received.

Each time a new paper is added, two new entries $X_{n+1} = 0$ and $Y_{n+1} = 0$ are created, and n is incremented by one. Then, p self-citations and q external citations are distributed to the existing papers. This increases by 1 the value X_k and/or Y_j of the corresponding recipient papers k or j.

Citations are given based on preferential attachment. This means that articles with a high number of citations have a larger chance to be cited again, compared to the articles which have less citations. Preferential attachment depends only on the distribution of external citations. When a new citation is given (whether self-citation or external one) it goes to paper k with a probability P_k computed as

$$P_k = \frac{X_k + \Delta}{n\Delta + \sum_{j=1}^{n} X_j} \tag{2}$$

where $\Delta = 1$ is a score added to all articles to prevent papers without citations ($X_k = 0$) from having a zero probability of being cited.

The process starts with an initial number of articles N_0, without citation. N_0 should be small enough not to influence the rest of the process, but large enough in order to have enough articles to cite at the first iteration. Here we choose $N_0 = p + q$.

Once the number of publications n is equal to N, only external citations are distributed, q at each step, until a total of M citations has been given. The scientist's bibliometric profile is then obtained by summing X_k and Y_k for each article, $k = 1, \ldots, N$. From this profile, the h-index is easily computed.

3 Simulations with the single-scientist model

In order to validate our model we have computed the h-index for a set of 120 scientists and compared the values obtained with the real data from the Publish or Perish (PoP) software [10]. For each selected scientist (from computer science), we know the value of N (the number of articles), M (the total amount of citations a scientist received for all his/her papers) and the actual h-index, according to PoP.

The model is stochastic and its results vary from one execution to the next, even for the same values of p, q, M and N. For each scientist, we run the model several times and average the produced values instead of giving the result of a single execution. Note that the standard deviation of a simulation is found to be around 8% for an h-index below 15, and around 3% for an h-index higher than 60.

The values of p and q are however to be determined by exploring exhaustively a range of possible values. We observed that the number of self-citations has little impact on h, as opposed to the number of external citations. We found that $p = 1$ (self-citations) and $q = 2$ (external citations) give the best global agreement between the model and the actual h-index of the 120 scientists [11]. This is illustrated in the scatter plot in Figure 1 where the estimated h-index (produced with our model) is plotted against the actual one. In this figure, each point represents a scientist. Ideally, the estimated h-index should be equal to the h-index, and all points should fit on the main diagonal. On the same plot, the result of the empirical formula, equation (1) is also shown.

By computing the standard deviation with respect to the identity line, we observe that the prediction of our model is 20% better than that of equation (1), as reported in Table 1. Note that we also consider the profile of 70 biologists, and the same values $p = 1$ and $q = 2$ are again the best choice to explain their h-indexes.

Note that, by construction, Np is the number of self-citations generated by our model. With $p = 1$ we have that, on average, an author cites one of his previous paper each time he writes a new article. On the other hand, the total number of external citations, $M - Np$, is not easily related to q because after the first N rounds where $N(p + q)$

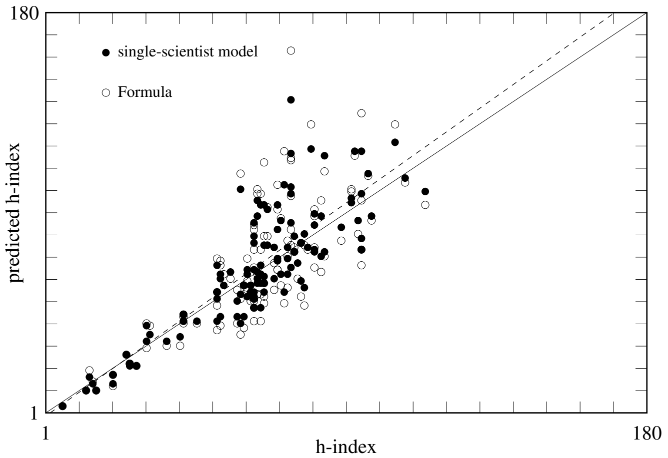

Fig. 1. Computed versus actual h-index for 120 computer scientists (points). Black circles show the mean values of the h-index, averaged over 50 executions for each scientist, obtained using our model. The white circles show the h-index computed using equation (1). The solid line is the identity function. The dashed line is a mean-squares fit of the white and black points (the two fits are indistinguishable).

Table 1. Dispersion from the actual values of the h-index computed with our model and equation (1).

	Slope	Intercept	Dispersion
Model with $p = 1$, $q = 2$	1.06	–0.88	15.06
$\left(\frac{M^2}{4N}\right)^{\frac{1}{3}}$	1.054	–1.202	19.19
Ideally	1	0	0

citations have been distributed, additional rounds with q citations are usually necessary to complete the simulation. Therefore our model should not be considered as a dynamical system in which the building of the h-index follows the actual time evolution of the scientist. It is better viewed as an iterative process that produces the h-index at a given moment in time, with known values of N and M, as does equation (1). In a forthcoming study, based on ideas discussed in [11], we plan to propose another model in which the time evolution of h can be linked to the time evolution of N and M.

It turns out that the two parameters p and q of our model can be tuned to fit almost any scientific profile (actually up to 93% of the 120 selected scientists, the other 7% being atypical researchers with very high h-index). Table 2 illustrates this result. The h-index of eight randomly selected researchers can be re-obtained from our model, with the indicated values of p and q, after an average over 20 runs. We observe that the p and q of an individual can depart from the average behavior $p = 1$, $q = 2$. However, for the first six scientists in the table, who have rather standard values of N and M, the model with $p = 1$, $q = 2$ gives an h-index which is within 1 or 2 units of the actual one. But, for the last two scientists of the table, the model with $p = 1$, $q = 2$ predicts $h = 41$ and $h = 42$, which is very far from reality. Note

Table 2. Examples of values of p and q which reproduce the h-index of eight actual scientists, whose values of N and M are given in the first two columns. M', the number of self-citations, is shown in the third column. The fourth column represents the h-index of the scientist. The fifth and the sixth columns give, respectively, the values of p and q which produce the desired h-index. The last two rows are atypical in the sense that $N(p + q)$ is larger than M. So the *single-scientist* algorithm stops once M is reached, possible without generating all N papers.

N	M	M'	h-index	p	q
18	180	3	6	0.15	1.5
394	2272	106	17	0.27	2.5
302	1069	44	11	0.14	3.7
93	507	21	10	0.22	1.3
210	838	68	12	0.32	1.4
594	2 784	144	19	0.25	1
675	10 730	281	19	0.41	20
769	11 694	259	22	0.33	19.5

that equation (1) gives $h = 35$ for these two scientists, also an erroneous result.

In Table 2 we accept non-integer values for p and q, so that we can better resolve the number of self-citations $M' \approx Np$ for each scientist. With a non-integer p one distributes, at each round, $\lceil p \rceil$ citations with probability α, or $\lceil p \rceil - 1$ citations with probability $1 - \alpha$. By choosing $\alpha = 1 + (p - \lceil p \rceil)$, the average number of citations is p. One does the same for q. For integer values of p and q, this procedure is identical to the previous one as $\alpha = 1$.

This result shows that N and M are not enough to predict accurately the h-index of a given scientist. The other parameters, p and q are needed to better specify its profile. Thus, equation (1), or the values $p = 1$, $q = 2$ are good approximations only when describing an average behavior.

Our model also allows us to investigate the evolution of the value of the h-index for a fixed number of papers, but with a number M of citations that keeps increasing. Simulations produce the results shown in Figure 2. It predicts that h grows with $M^{1/2}$. In addition, one can see the transition between the power law behavior and the saturation value as h gets closer to N. As shown by the gray line in Figure 2, the $M^{1/2}$ behavior is also observed with real data. Real data have been obtained by considering papers published by a given author before 2000 and cited between 2000 and 2010 (data gathered from SPIRES). Note that we considered several authors in the dataset and they all have a similar behavior.

As a last application of the *single-scientist* model, we now discuss the h-index of an institution or a group of researchers. It is a common practice to compute the cumulative h-index of a group by simply collecting all the papers authored by one member of that group and by applying the standard procedure to determine h. It is then tempting to compare the bibliometric quality of several groups by comparing their respective cumulative h-index.

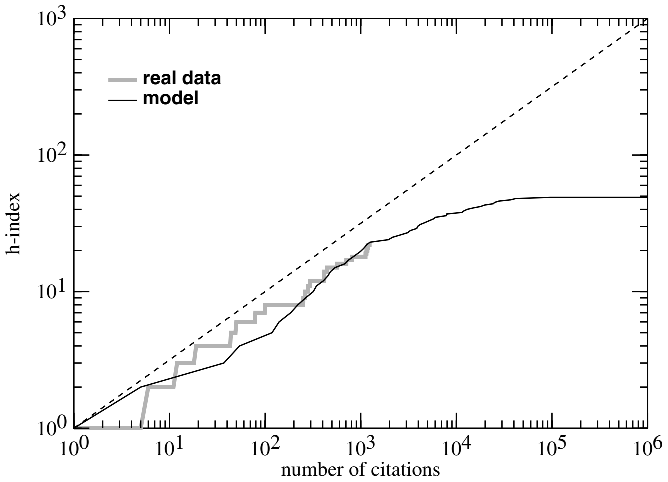

Fig. 2. Evolution of the h-index as a function of the number M of citations, with $N = 59$ fixed. The dashed line shows a slope of 1/2. The gray line is the evolution of h for a randomly picked real scientist from the SPIRES dataset and the black line is the output of our model.

However, a fair comparison should take into account the size of the group, as large groups are more likely to have a large h.

An interesting question is whether the cumulative h-index can be related to the individual h-indexes. Clearly, there is no exact solution to this problem as, depending on the distribution of citations, the cumulative h-index of two scientists can range between the sum of the individual h-indexes and the larger of the two h-indexes.

However, in general, a satisfactory prediction can be obtained, not from the individual h-indexes, but from the numbers N_i of papers and M_i citations each scientist i brings to the group. By using our model with $N_{tot} = \sum N_i$ and $M_{tot} = \sum M_i$ we obtain a good approximation of the cumulative h value.

Table 3 shows the cumulative h-index of groups built by randomly choosing scientists from the dataset used previously. If, among the authors, there are common papers, they are counted only once. The value of the h-index and that predicted by *single-scientist* model, with $p = 1$ and $q = 2$ are indicated. The table also shows the prediction of formula (1), with $N = N_{tot}$ and $M = M_{tot}$. Both equation (1) and the *single-scientist* model estimate well the h-index of a group. However, the model produces significantly better results as shown by the value of Δ, the discrepancy with respect to the actual values of h.

In order to provide a fair comparison between two institutions of different sizes, we propose to define $\bar{h}$, an "average" h-index as follows: let k denote the number of scientists in the group. We define $\bar{N} = N_{tot}/k$ the average number of papers per scientist, and $\bar{M} = M_{tot}/k$ the average number of citations. Then, $\bar{h}$ is defined as the h-index of a representative individual having $\bar{N}$ papers and $\bar{M}$ citations.

This average value $\bar{h}$ can be computed with our *single-scientist* model and will provide a metrics which takes into account the size of the group. However, equation (1) offers an analytical expression which is a good approximation, with an explicit dependency upon k. We have

$$\bar{h} = \left(\frac{(M_{tot}/k)^2}{4(N_{tot}/k)}\right)^{\frac{1}{3}} = k^{-\frac{1}{3}}\left(\frac{M_{tot}^2}{4N_{tot}}\right)^{\frac{1}{3}} = k^{-\frac{1}{3}}h \qquad (3)$$

where h is the h-index of the group. In short we propose to define the representative h-index of a group by dividing its h-index by the cubic root of its size. Using this metrics, we have, for instance, $\bar{h} = 85$ for the third group (with $h = 154$ and $k = 6$), whereas the sixth group gets $\bar{h} = 83$ (with $h = 188$ and $k = 12$). Thus, in terms of our representative h, these two groups perform quite identically.

Table 3. h-indexes of groups of scientists. Each row corresponds to a group, randomly chosen from the database. The first column shows the h-indexes of each individual in the group. The second column is the h-index of the group, computed from all the papers authored by the group. The third and fourth columns are the total number of papers and the total number of citations in the group. Finally, the fifth and the sixth columns are the predicted h-indexes. The quantities $\langle\Delta\rangle$ are the averages of the deviations of the predicted h to the actual one.

Individual h-indexes	h-index of the group	N_{tot}	M_{tot}	Eq. (1)	Model
113 35 75 58 7 14	**172**	1 215	150 239	**167**	**173**
113 42 75 95 106 54 58	**252**	2 558	279 685	**197**	**290**
7 14 25 58 54 8	**154**	1 122	109 940	**139**	**151**
23 7 34 25 14 13 8 35	**75**	931	26 980	**58**	**68**
5 5 62 28 18 13 8 13 16 13	**77**	621	23 656	**61**	**69**
79 106 95 5 5 62 28 18 13 8 13 16	**188**	1 910	178 936	**161**	**178**
5 5 28 62 18 61 22 25 47	**117**	1 099	66 518	**100**	**109**
3 4 5 7 9 12 13 16 19 20 21 30 32 33 37 46 53 55 58 62 64	**174**	4 785	198 622	**127**	**160**
				$\langle\Delta\rangle =$ 24.9	$\langle\Delta\rangle =$ 11.1

4 The multi-scientist model

In this section we consider a virtual community of K scientists having produced a total of N articles, together with a total of M citations.

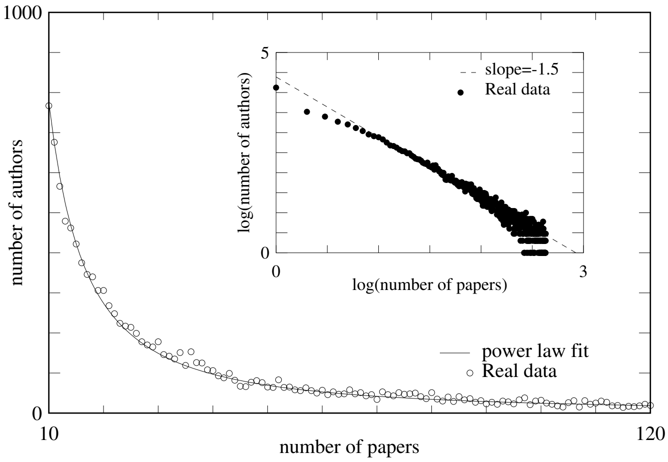

Fig. 3. Fit of the of the number of authors who published a given number of papers (data from SPIRES) within 10 years. The inset shows the entire dataset in a base 10 log-log plot. The main figure gives the fit on a smaller range of publications where we can better see that the expression $y = ax^{-3/2}$ represents well the data, with $a = 24\,500$.

In our model, each author is defined by its publication profile, i.e. the probability that he publishes a new article at each time period. In order to use a significant value for this quantity, we have extracted the frequency distribution of publications from the SPIRES [12] High Energy Physics dataset (containing 120 465 articles, 453 223 citations and 47 115 authors). From our study, we have obtained that the number y of authors having published x papers over a given period of time goes as

$$y = ax^{-3/2} \tag{4}$$

with a constant a that can be fitted from the data shown in Figure 3. This result is compatible with Lotka's law for author productivity.

The model then proceeds as follows. For each time step t, from $t = 1$ to some given T_{max}, all the K authors are considered one by one. Each of them adds a new paper to the existing ones with a fixed probability, which is determined from distribution (4). If a paper is added, it also comes with a reference list that gives r new citations to be distributed among the papers already produced since the beginning of the process. In this model, the number of self-citation is included in r.

On average, the value of r is simply M/N, so that once N papers have been published, there is a total of M citations. However, we may expect that the value of r varies around M/N for each published article. Therefore we draw r from a probability distribution which is again determined from real data. Using the Stanford dataset [13] (High Energy Physics papers published between 1993 and 2003) we can extract a histogram of the number s of papers with r references. We obtain the exponential distribution

$$s \sim \exp(-0.077r). \tag{5}$$

The last stage of one iteration of the *multi-scientist* model is to distribute the r references to the existing papers,

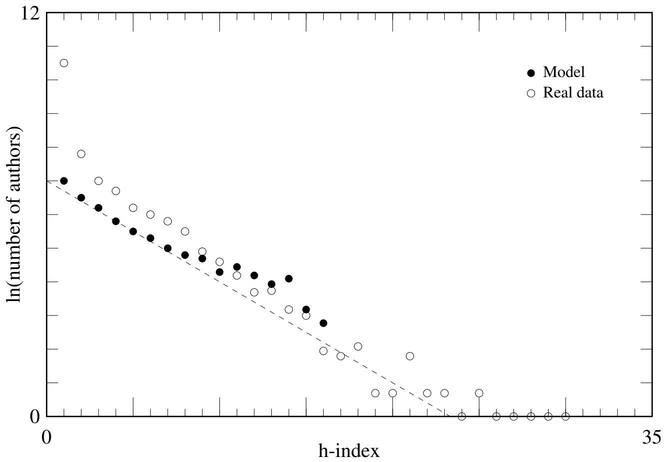

Fig. 4. Distribution of the h-index for both the *multi-scientist* model (black dots) and SPIRES data (white dots). To guide the eyes a line of slope –0.3 is shown.

according to the preferential attachment mechanism already discussed. The process terminates when a total of N papers have been produced. Due to the way r is drawn, we expect to obtain a number of citations close to M.

To validate the *multi-scientist* model, we show in Figure 4 the number K_h of authors with a given h-index, for both the model and the real data obtained from SPIRES. This result suggests that the model captures reality in a satisfactory way. It also suggests that the distribution of h follows an exponential law $K_h = a\exp(-bh)$, with $a = 1097$ and $b = 0.3$.

5 Impact on the h-index due to the removal of scientists

In this section we investigate how the h-index of scientists is affected when authors with low h-index are removed from the community. Removing a scientist means to also eliminate his papers and the citations he gave. As a result, the h-index of the remaining researchers is expected to decrease. The motivation of such an experiment is to better understand the social structure of the scientific community and the impact of filtering scientists on the value of their h-index.

A first hypothesis is that the structure of the scientific world is pyramidal. This means that the h-index of scientists at some h level is mostly built out of the citations coming from the community of authors with a lower h value. Thus, removing these lower layers should decrease significantly the h of the upper layers. The results we obtain with both our *multi-scientist* model and from the SPIRES dataset tend to confirm this assumption.

In Figure 5 the gray line shows the function $F(h)$ defined as number of authors having an h-index larger than or equal to h. These curves are obtained from the SPIRES dataset, but for the sake of clarity, only the 1120 authors with $h \geq 6$ are shown. Then, we removed from the database all the scientists with an h value lower than or equal to 5, as well as all the citations they gave to the

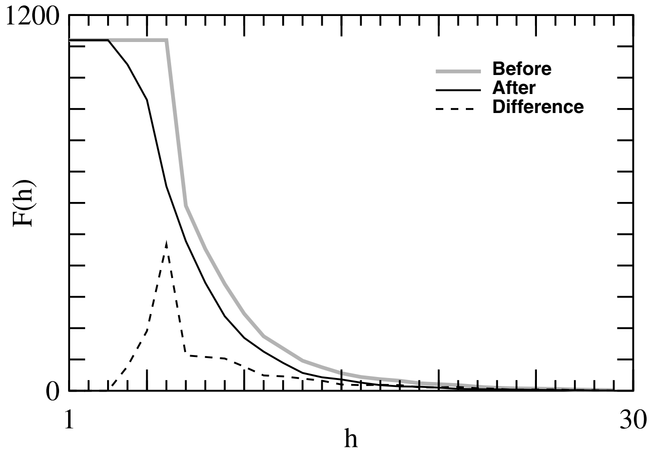

Fig. 5. Number $F(h)$ of authors with an h-index larger than or equal to h. The gray line corresponds to the SPIRES dataset, in which only authors with $h \geq 6$ are shown. The black line gives the same quantity after eliminating, from the dataset, the citations of scientists with an h-index lower than or equal to 5. The dashed line is the difference of the gray and black lines.

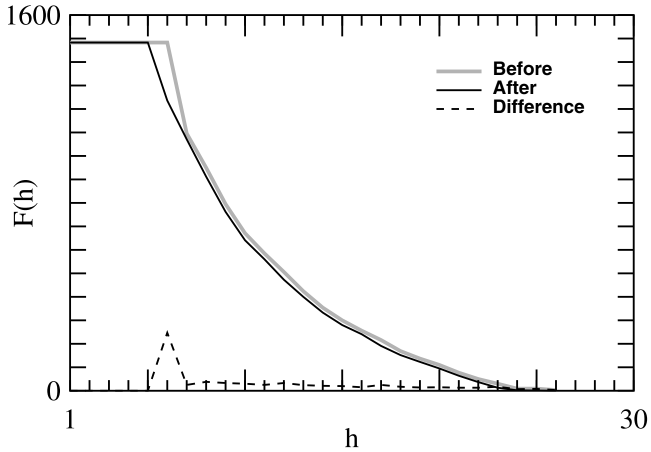

Fig. 6. Number $F(h)$ of authors with an h-index larger than or equal to h, as obtained with the *multi-scientist* model. The gray line is the original situation, were only authors with $h \geq 6$ are shown. The black line gives the same quantity after eliminating, from the simulation results, the citations of scientists with an h-index lower than or equal to 5.

community. As a result, 97.5% of the scientists were removed from the dataset along with 61% of the articles.

The number $F(h)$ of authors with an index larger than h is then affected as shown by the black line in Figure 5. We observe that, out of the 1120 scientists who had previously $h \geq 6$, 467 of them are now below the threshold, with h-indexes $h = 3$, $h = 4$ or $h = 5$.

Globally, all authors are affected by the process, as the removed low-h scientists are numerous and are feeding the entire pyramid of citations. This effect is very visible for scientists with an h-index close to the threshold $h = 5$, as shown by the peak of the dashed line in the figure, which gives the difference of the curves before and after the suppression. For authors with an original h larger than 7, $F(h)$ is roughly shifted to the left by an amount $\Delta h \approx 2$. A closer look at the values of the curves reveals a variation of F between 30% to 50% of its original value, thus showing the impact of the process on the entire community. Note that a variation of $F(h)$ of 50% means that the number of authors with an h-index larger than h has decreased by half.

When the threshold value for h is set to 10 instead of 5, a similar behavior is observed. For instance, a randomly chosen scientist with an initial $h = 15$, sees his h-index drop to 13 after setting the threshold to 5, and to 11 for $h_{thres} = 10$. As another example we can consider an author with $h = 36$ in the dataset. When setting the removal threshold at $h_{thres} = 5$, his h-index decreases to 33, thus losing 3 units of h. With $h_{thres} = 10$, his h-index becomes 27, a drop of 9 units. Finally, with $h_{thres} = 20$, his new h value is 19, corresponding to a reduction of 17 units.

A similar behavior, however less pronounced, is also visible in the simulations we performed with the *multi-scientist* model (see Fig. 6). As with the real data, only scientists with initially $h \geq 6$ are displayed. A peak in the variation of $F(h)$ before and after filtration shows up around the threshold value (corresponding to a drop of 16% for $h = 6$). For higher values of h, i.e. $7 \leq h \leq 24$, the relative variation of $F(h)$ due to applying the threshold ranges between 2% and 78%, with an average of 15%. The high relative variations of $F(h)$ are not visible in Figure 6 because they concern small quantities. For instance, for $h = 24$, $F(h)$ varies from 9 to 2, leading to a 78% drop.

We have not yet studied which parameters of the *multi-scientist* model should be tuned to reproduce the same amplitude of values as observed with real data.

From a general point of view, the strong impact of filtering scientists with respect to a minimum h-index can be explained by assuming a stratified scientific community, where the value of h in each level depends mostly on the citations from the lower h levels. Accordingly, scientists with a low h value are essential to maintain the h-index of more productive researchers. For instance, the following damaging scenario could be envisaged: if, for some political reasons, scientists with low h-index are suppressed from the community, it will prevent scientists who were above the threshold from increasing their h value over time. In turn, these scientists will be seen as low h-index researchers and will not receive funding anymore. Little by little, the entire pyramid might collapse, destroying all the scientific community.

6 Conclusions

In this paper we have presented a new approach to study the properties of the bibliometric h-index. We showed that we can capture very well the real properties of this metrics with simple agent-based models, composed of virtual scientists publishing papers and giving citations according to a preferential attachment mechanism.

Our approach goes beyond the state of the art because (i) it gives sensible predictions for any number of papers and publications; (ii) it explains the evolution of the h-index as a function of time; (iii) it proposes an “average”

h-index that can be used to compare the scientific quality of institutions of different sizes.

Our approach also allows us to study the impact of given bibliometric policies on the scientific community. In particular our study reveals interesting properties about the publication eco-system when structured by the h-index. It shows that inter-dependencies exist across the h levels, suggesting that a scientific community can be seen as a pyramid where scientists with low h-indexes feed the higher levels with their citations.

We thank the reviewer for his relevant remarks and questions that allowed us to improve our paper.

References

1. J.E. Hirsch, Proc. Natl. Acad. Sci. **102**, 16569 (2005)
2. R. Costas, M. Bordons, J. Informetrics **401**, 193 (2007)
3. F. Laloe, R. Mosseri, Europhys. News **40**, 26 (2009)
4. M. Strathern, Eur. Rev. **5**, 305321 (1997)
5. D.N. Arnold, K.K. Fowler, Notices of the AMS **58**, 434 (2011)
6. S. Redner, Eur. Phys. J. B **4**, 131 (1998)
7. D. de Solla Price, J. Amer. Soc. Inf. Sci. **27**, 292 (1976)
8. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
9. J.E. Iglesias, C. Pecharromán, Scientometrics **73**, 303 (2007)
10. A.W. Harzing, Publish or Perish http://www.harzing.com/pop.htm
11. G. Ionescu, A numerical model for the bibliometric-index, Technical report, Computer Science department, University of Geneva, Switzerland, 2012
12. Stanford physics information retrieval system, http://inspirehep.net/ (2009)
13. J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: Densification laws, shrinking diameters and possible explanations, in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2005*