Open Access

Open Journal of Web Technologies (OJWT)
Volume 4, Issue 1, 2017

http://www.ronpub.com/ojwt
ISSN 2199-188X

The mf-index: A Citation-Based Multiple Factor Index to Evaluate and Compare the Output of Scientists

Eric Oberesch, Sven Groppe

Institute of Information Systems (IFIS), University of Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, Germany, oberesch@ieee.org, groppe@ifis.uni-luebeck.de

Abstract

Comparing the output of scientists as objective as possible is an important factor for, e.g., the approval of research funds or the filling of open positions at universities. Numeric indices, which express the scientific output in the form of a concrete value, may not completely supersede an overall view of a researcher, but provide helpful indications for the assessment. This work introduces the most important citation-based indices, analyzes their advantages and disadvantages and provides an overview of the aspects considered by them. On this basis, we identify the criteria that an advanced index should fulfill, and develop a new index, the mf-index. The objective of the mf-index is to combine the benefits of the existing indices, while avoiding as far as possible their drawbacks and to consider additional aspects. Finally, an evaluation based on data of real publications and citations compares the mf-index with existing indices and verifies that its advantages in theory can also be determined in practice.

Type of Paper and Keywords

Regular research paper: *bibliometric indices, bibliometric indicators, citation-based indices, multiple factor index, mf-index, evaluation of science, research evaluation, scientific output, research performance, scholarly impact*

1 Introduction

The relevance of bibliometric indicators for the evaluation of the output of researchers has increased continuously over time, due to their promise to provide an objective assessment in adequate time. Consequently, there exist many scientific publications dealing with that subject. The topic of the publications is often the development of new indicators or the improvement of existing ones.

The most important category of indicators is the one containing numeric, citation-based indices. The essential calculation basis of typical indices consists of the citations received by a publication. Additionally, these indices return a number as result value, which is usually either an integer or a real number. However, the currently existing indices often have some drawbacks, which may result in an unjustified preference or disadvantage of single scientists if these indices are used. Furthermore, the existing indices often only consider one or two different factors and disregard other, potentially important aspects like the influence of the field of study of a publication on the amount of received citations. Section 2 introduces the most important bibliometric indices with their advantages and disadvantages and provides an overview of the aspects considered by them. The analysis of the existing indices is a good basis to conceive a new index, which avoids, as far as possible, their drawbacks and combines several known and new aspects to assess the output of a scientist.

To achieve this, the general criteria for the new index are defined in Section 4.1, and Section 4.2 describes the factors that should be taken into account by the new index. We define the new mf-index in Section 4.3 and discuss the properties and benefits of the new index in the subsequent Section 4.4. Afterwards, we compare in Section 5 the mf-index with existing indices based on data of real publications and citations. At the end of this paper, Section 6 summarizes the achieved results and addresses possible future improvements.

2 Existing Indices

Nearly every decision in the academic context (like approval of research funds, filling of open positions at universities, and conferment of scientific awards) now depends on the output of the researchers. Therefore, the importance grows to measure this output [2]. The best way to assess the output of a scientist is to read and understand her/his publications. However, the number of papers has increased significantly over the years as shown by the number of publications recorded in the ISI Science Citation Index, which has tripled from 1971 to 2003. Hence, there is an increasing need to find less time-consuming, but still good and meaningful evaluation mechanisms [5].

The section surveys a series of indices, which evaluate the research performance of scientists. Since the number of existing indices is extremely high and the main focus of our paper is on the definition and analysis of the new mf-index, we have to limit the number of indices discussed in this section. In Section 3, we refer interested readers to six other contributions offering an extensive overview of the currently used bibliometric indicators and to three papers providing a detailed comparison of existing indices based on real-world data and partially also on well-constructed theoretical cases.

To select the indices presented in this paper, we first conducted a detailed literature survey on bibliometric indices. We then selected a subset of these indices based on the visibility and categories. We include those indices that are most visible and discussed in many other publications. We chose the indices from different categories, where each category focuses on adjusting a certain weakness of the h-index, such as the ignored excess citations, the influence of career length, publication age, citation age, co-authors or self-citations. This subset of indices is later used in Section 4 as the basis to design the mf-index that incorporates the aspects from the different categories and adds additional aspects.

We first provide an overview of classical indices in Section 2.1 and then present modern and more complex indices. Section 2.2 describes the well-known h-index in detail. Thereafter, different categories of h-index variants are presented, where each category addresses a certain weakness of the h-index such as the ignored excess citations. Since some of the variants consider multiple aspects, we list them in the category that refers to their most important aspect. The complete list of the aspects of each index is later shown in Section 2.11. Each index is presented with a formal definition, followed by a description of its advantages and disadvantages. After presenting the single indices, Section 2.11 gives an overview of the different factors considered in the indices and summarizes their advantages and disadvantages.

2.1 Classical Indices

A series of classical indices has been developed to evaluate both productivity and impact of scientists. To measure productivity, indicators like the total number of publications or the number of publications within a specific time frame are used [2]. The impact is determined by using indicators like the total number of citations, the average number of citations per paper, the number and percentage of papers with a minimum amount of citations or the average impact factor of the scientific journals in which the articles were published [2]. With the rise of new databases to record scientific papers (e.g., ISI Web of Science, Scopus and Google Scholar), it became possible to develop more complex indices considering multiple aspects of the output of scientists [2].

2.2 h-index

The h-index [28] was developed in 2005 by a physicist, Jorge Eduardo Hirsch. Hirsch's h-index attracts a large amount of research work on indices and is itself extensively cited.

2.2.1 Definition

The h-index is defined in [28] as follows:

> *"A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq h$ citations each."*

To graphically display the h-index, one has to plot on the x-axis the publications ordered descending by their number of citations and on the y-axis the respective number of citations. The h-index is equivalent to the point where the number of citations intersects with the number of publications [28]. The first h papers are also referred to as *Hirsch core* or *h-core* [34].

2.2.2 Advantages of the h-index

The h-index is easy to calculate and considers both productivity and impact of a scientist [28]. Furthermore, it is an objective indicator and is therefore a reasonable basis for important decisions in the academic context like the appointment of professors [17]. It also has some advantages over other numeric comparison criteria such as the journal impact factor (JIF), the total number of publications, the total number of citations, the average number of citations per paper, the amount of highly cited publications and the number of citations to each of the q most cited publications [28]. The h-index is not vulnerable to a set of rarely cited publications [66]. To increase the h-index, all most cited papers have to receive new citations, which corresponds to a logarithmic growth of the h-index such that a higher h-index is difficult to achieve [2]. In addition, the errors in the bibliographic databases mainly concern the less cited publications, which has hardly any effect on the calculation of the h-index. Furthermore, the h-index does not give the impression of extreme precision, as the result is an integer and not a real number [66].

2.2.3 Disadvantages of the h-index

The h-index suffers from problems with the comparability of scientists who work in different fields of study, due to the fact that productivity and citation practice differ in part considerably among the individual disciplines [28]. The differences concern the amount of publications within a specific time frame as well as the average number of received citations. For instance, papers in the field of mathematics usually receive significantly less citations than publications in the field of physics [6].

It is often not possible to obtain the complete set of publications of a researcher with a very common name [28]. The h-index also depends on the age of the scientist, due to the fact that the amount of publications and received citations increases over time [37], which prevents a meaningful comparison of scientists of different ages [2]. In addition, the h-index ignores how much more than h citations the most cited papers have received [21, 22]. A scientist with some highly cited publications may therefore obtain a nearly identical h-index as a significantly less cited researcher [2].

The h-index does not consider the impact of the papers outside the h-core [9]. Two scientists may receive the same h-index, although one researcher published exactly h papers, while the other published significantly more than h papers with all of them having an impact on the scientific community [1]. Additionally, scientists are disadvantaged by the h-index when they publish only a small number of papers that nonetheless have gained a large international influence. A reason for this is that the h-index has an upper bound being the total number of publications. This means that the highest possible h-index value of a scientist corresponds with her/his number of papers [17].

The h-index considers self-citations, which could change the publication behavior of researchers [2]. This could lead scientists to intentionally increase the amount of citations of the papers, which are responsible for the next increment of the h-index [51]. Another aspect is that the h-index ignores the context of the citations [2]. Scientists cite a certain paper for different reasons: A publication may be cited because it was relevant for the researcher [69], but there are also papers which are only superficially cited [16, 44]. Some publications are only mentioned because a historical overview of the topic is often expected [46, 49].

Citations are also used to embellish the introduction of a publication, without having any influence on the scientist's research work [70]. Some papers are even just cited to increase the size of the references section, pretending that the author is familiar with the matter and has read a lot of literature on the topic [36]. Additionally, some citations are negative [16, 44, 69, 70]. They may occur within a critical analysis of a research paper [16, 44, 69], but may also be used to hint at scientific misconduct [35, 70].

Another aspect is that well-known scientists get significantly more citations than less known scientists [70]. Citations are often distributed on the scientists according to the Pareto principle. This means that very few researchers receive a lot of citations, while the majority of the scientists gets significantly lesser cited [19]. Furthermore, review articles usually obtain more citations than original research papers [1, 42, 70]. This puts innovative scientists at a disadvantage and rewards researchers who simply summarize already existing findings [42]. Additionally, there remains the risk of uncritical use of the h-index. The same applies to nearly every single indicator which is easy to calculate [2]. The evaluation of research performance is a complex task which cannot adequately be expressed by a single indicator [41].

2.3 Variants Considering h-core Citations

In this section, we list h-index variants that only introduce small changes to the definition of the original h-index such that their main aspect still considers the h-core citations. Section 2.3.1 presents the h_α-index allowing a better comparison of researchers with high h-indices. Section 2.3.2 describes the h^n-index which normalizes the h-index to compare scientists with different publication counts.

2.3.1 h_α-index

The h_α-index is defined similarly to the h-index, with the difference that a coefficient $\alpha \in (0, \infty)$ is placed before the citation count of a publication. It is quite difficult for researchers to further increase their already high h-index, which makes it a challenging task to compare those researchers due to their similar h-index values. The usage of the h_α-index by varying α has the advantage that it becomes possible to compare scientists with such a high h-index [65].

2.3.2 h^n-index

The h^n-index [60] is defined as $h^n = h/n$. Dividing by the amount of publications n of the considered researcher compensates the problem that the h-index does not consider the total number of publications of a scientist. However, this approach has the drawback that it rewards less productive researchers [2].

2.4 Variants Considering Excess Citations

As pointed out in Section 2.2.3, the h-index ignores the citations above h. In this section, we present different approaches, which consider excess citations. Sections 2.4.1, 2.4.2 and 2.4.3 introduce the $h^{(2)}$-index, g-index and o-index focusing on the most cited papers. In contrast to this, the m-index (see Section 2.4.4) uses the median citation count in the h-core. The hg-index (see Section 2.4.5) combines the previously introduced h- and g-indices, while the q^2-index (see Section 2.4.6) connects the h- and m-indices. Other indicators like the A-index and R-index (see sections 2.4.7 and 2.4.8) use the average citation count or the square root of the citation counts, respectively. Furthermore, we describe the j-index in Section 2.4.9, which is an index that determines the exact distribution of the citations in the h-core. Section 2.4.10 and Section 2.4.11 introduce the h_w- and h_d-indices detecting dynamic changes in the research performance of scientists. Finally, Section 2.4.12 introduces the h_{rat}-index considering a part of the excess citations and non h-core citations to express the difference to the next h-index level.

2.4.1 $h^{(2)}$-index

The $h^{(2)}$-index is defined in [38] as follows:

> *"A scientist's $h^{(2)}$-index is defined as the highest natural number such that his $h^{(2)}$ most cited papers received each at least $[h^{(2)}]^2$ citations."*

The $h^{(2)}$-index has the advantage that less work is needed to verify the correctness of the publication data, which is especially relevant when the data set contains different scientists with identical names. Additionally, the $h^{(2)}$-index emphasizes the highly cited papers [38].

However, it is difficult to compare researchers with high $h^{(2)}$-indices if their total number of publications differs and the citation frequencies differ. This disadvantage is due to the fact that the $h^{(2)}$-index only considers a limited subset of the papers [2].

2.4.2 g-index

The g-index is equivalent to the highest number g of publications which together received at least g^2 citations. It is always at least as high as the h-index. The g-index has the advantage that it considers the citations above h, while they are ignored by the h-index [21, 22].

A disadvantage of the g-index is that it increases significantly if a scientist achieves a big success with a single publication, even if most of her/his papers obtain hardly any citations. This puts other researchers, who have a higher average number of citations per publication, at a disadvantage [3].

2.4.3 o-index

The motivation behind the o-index is to avoid that some extraordinary scientists, who published fewer, but ground-breaking contributions, receive a lower rank, which is a disadvantage of the h-index [20]. For this purpose, the o-index considers the most cited publication cit_1 of a researcher besides the h-index:

$$o = \sqrt{cit_1 \cdot h} \tag{1}$$

The disadvantages of the o-index favoring researchers with a big success of even only one highly cited publication are reinforced in comparison to the g-index.

2.4.4 m-index

The m-index corresponds to the median of the number of citations to the publications in the Hirsch core. The median provides a more accurate picture of the amount of received citations, which is not necessarily the case when using the arithmetic mean, since citations are often unevenly distributed among the different papers [10]. The usage of the m-index reduces the influence of the few papers that are highly cited [2].

2.4.5 hg-index

The hg-index is defined as $hg = \sqrt{h \cdot g}$. Its value lies between the h-index and the g-index, i.e., $h \leq hg \leq g$.

The hg-index is not as vulnerable to a small subset of highly cited papers as the g-index. As noted in Section 2.4.2, the g-index increases significantly if a scientist achieves a big success with a single publication, even if most of her/his papers obtain hardly any citations, which is shown by a very low h-index. In such a case, the hg-index does not increase significantly, which no longer results in a disproportionate benefit for researchers that have a few highly cited papers and a lot of hardly cited papers. Additionally, the hg-index is more fine-grained and allows a comparison of scientists who have a similar h-index or g-index [3].

2.4.6 q^2-index

The q^2-index is defined as $q^2 = \sqrt{h \cdot m}$. One of its advantages is that the q^2-index combines two categories: the number of publications in the Hirsch Core (by using the h-index) and the impact of these publications (by using the m-index) [14].

2.4.7 A-index

The A-index is defined in [33] as follows:

$$A = \frac{1}{h} \sum_{j=1}^{h} cit_j \tag{2}$$

cit_j corresponds to the citation count of the j-th paper (in a list sorted by most citations). The A-index expresses the average number of citations within the h-core [33]. An advantage of the A-index is that it considers all citations in the h-core, and this allows a comparison of scientists with similar h-index [2]. A major drawback of the A-index is that it disadvantages researchers with a higher h-index, since it performs a division by h [34].

2.4.8 R-index

Using the same notation as in the A-index, the R-index is defined in [34] in the following way:

$$R = \sqrt{\sum_{j=1}^{h} cit_j} \tag{3}$$

The R-index compensates the major disadvantage of the A-index, since it takes the square root instead of dividing by h. This approach no longer disadvantages researchers with higher h-indices. However, the R-index is vulnerable to a small subset of highly cited papers [34].

2.4.9 j-index

The j-index is defined in [62] as follows:

$$j = h + \frac{\sum_{k=1}^{12} w_k \cdot N_k(h \cdot \Delta h_k)}{\sum_{k=1}^{12} w_k}, w_k = \frac{1}{k} \tag{4}$$

Δh_k consists of 12 different intermediate stages (500, 250, 100, 50, 25, 10, 5, 4, 3, 2, 1.5, 1.25). At the beginning, Δh_k is a large number that is getting smaller with increasing k. $N_k(h \cdot \Delta h_k)$ corresponds to the number of publications which received at least $h \cdot \Delta h_k$ citations. $h \cdot \Delta h_k$ can therefore be regarded as a threshold for the number of citations, which increases when the h-index rises. Since with increasing k, the value of Δh_k decreases continuously, more and more publications fulfill the threshold for the minimum number of citations. However, since the value of w_k decreases at the same time, the papers do not count as much as before [62].

The above described construction of the j-index makes it possible to detect the exact distribution of the citations in the Hirsch core. Thus, highly cited papers can be considered without easily affecting the result, because the threshold for the needed citations (i.e., $h \cdot \Delta h_k$) is quite high for small k values. When k increases, the j-index also takes into account if some less cited publications have a higher citation count than other rarely cited papers within the Hirsch core. The less cited publications are considered without distorting the result, since their amount is multiplied with a significant lower value w_k [62].

2.4.10 h_w-index

The h_w-index is defined in [25] in the following way:

$$h_w = \sqrt{\sum_{j=1}^{r_0} cit_j} \tag{5}$$

cit_j corresponds to the citation count of the j-th publication (in a list sorted by most citations), while r_0 represents the highest possible index i which satisfies the inequality $r_w(i) \leq cit_i$. $r_w(i)$ is calculated according to the following equation:

$$r_w(i) = \frac{1}{h} \sum_{j=1}^{i} cit_j \tag{6}$$

The h_w-index has the advantage that it is able to detect dynamic changes in research performance [25].

2.4.11 h_d-index

The h_d-index is defined in [54] as $h_d = R(T) \cdot v_h(T)$ and consists of the R-index and the h-velocity v_h at the time T. $v_h(T)$ is defined as follows:

$$v_h(T) = \frac{dh}{dt}(T) = \lim_{t \to 0} \frac{h(T+t) - h(T)}{t} \quad (7)$$

$h(t)$ can be a linearly increasing function of t, but may also be another type of function. Exemplary definitions of $h(t)$ can be found in [54].

The main advantage of the h_d-index is that it takes into account if the h-index of a scientist has recently increased significantly or if it is unchanged for a long time [54].

2.4.12 h_{rat}-index

The h_{rat}-index is defined in [55] as follows:

$$h_{rat} = (h+1) - \frac{n_c}{2 \cdot h + 1} \quad (8)$$

n_c corresponds to the number of missing citations to obtain an h-index of $h+1$, while $2 \cdot h + 1$ represents the maximum number of needed citations to increase the h-index by one step. The design of the index leads to the inequality $h \leq h_{rat} < h + 1$. h_{rat} lies closer to $h+1$ if only few citations are missing, whereas it lies nearer to h if still a lot of citations are missing [2]. The h_{rat}-index has the benefit to increase in smaller steps than the h-index, which is due to the fact that the value of the h_{rat}-index is a rational number. This allows a more accurate comparison of scientists with a similar h-index [55].

2.5 Variants Considering all Citations

The previously introduced h-index and its variants consider only a subset of the received citations of a scientist: the h-core citations and the excess citations in the h-core. The h_{rat}-index (see Section 2.4.12) differs slightly from the other indices, since it also takes into account non h-core citations by considering publications that have less than or equal to h citations. However, the h_{rat}-index considers only a part of the excess citations and the non h-core citations. In the following section, we describe the tapered h-index which is different than the earlier presented indices, since it considers the h-core citations as well as the complete excess citations and non h-core citations.

2.5.1 Tapered h-index

Definition

The tapered h-index h_T is defined in [4] as follows:

$$h_T = \sum_{j=1}^{n} h_{T(j)} \quad (9)$$

In Equation 9, n represents the scientist's paper count. The terms of the sum $h_{T(j)}$ are defined as follows:

$$h_{T(j)} = \begin{cases} \frac{cit_j}{2j-1} & cit_j \leq j \quad (10a) \\ \frac{j}{2j-1} + \sum_{i=j+1}^{cit_j} \frac{1}{2i-1} & cit_j > j \quad (10b) \end{cases}$$

We assume that the papers are sorted descending by their citation count, i.e., a paper j is the j-th most cited paper of the considered author. Hence, the tapered h-index considers all n papers of a scientist, all received citations cit_j of a paper j and the number of needed citations to increase the h-index by one step, respectively. If the currently considered paper j has not more than j citations, the inverse of the required citations is multiplied with the already existing citations cit_j of this paper (see Equation 10a). If there are more than j citations, the inverse of the needed citations is multiplied with j. In addition, a sum is added which considers all citations above j and assigns a value to them (see Equation 10b).

Advantages

According to Equation 10b, a citation counts less when it is more away from j. This prevents that highly cited publications disproportionately affect the final result [4]. Additionally, h_T has a higher accuracy than the h-index. This is due to the fact that h_T has decimal places, while the h-index is just an integer. The decimal places of h_T directly express small changes, while the h-index does not change for a long time, and when it changes, it is a sudden jump to the next h-index level [4]. Another advantage is that all publication and citation data is taken into account and has a direct impact on the value of the index. This means that two categories of previously unconsidered aspects are now taken into account: The first category concerns the citations above h and within the h-core, and the second category deals with the citations below h and outside the h-core [4].

Disadvantages

However, the calculation of the tapered h-index is a bit difficult because exact data of the citation frequency is needed also for the less cited papers [2].

2.6 Variants Considering Career Length

One of the disadvantages of the h-index mentioned in Section 2.2.3 is the discrimination of young scientists, since the number of publications and received citations increases over time, and with them the h-index. Therefore, it is not possible to compare older scientists with young researchers who did not have the same time to publish papers and obtain citations. The following section describes the m-quotient allowing a comparison of scientists of different seniority.

2.6.1 m-quotient

The m-quotient is defined as $m = h/y$, whereby y represents the number of years since the first publication of the author. Dividing by the length of the scientific career allows to compare scientists of different ages [28].

2.7 Variants Considering Publication Age

The original h-index does not take into account that older publications benefit from the fact that they had a long time to receive citations. Without considering the age of the papers, the index value also does not reflect if a scientist is still actively publishing or if she/he has stopped publishing new contributions. In this section, we present two indices incorporating the publication age in the calculation of the index value. Section 2.7.1 introduces the AR-index, and Section 2.7.2 describes the h^c-index.

2.7.1 AR-index

The AR-index is defined in [34] as follows:

$$AR = \sqrt{\sum_{j=1}^{h} \frac{cit_j}{a_j}} \qquad (11)$$

cit_j corresponds to the number of citations of the j-th publication and a_j represents the years since the publication of the paper. It is an advantage that the age of the publications is taken into account. The AR-index can therefore dynamically rise and fall [34].

2.7.2 h^c-index

To receive an h^c-index of h^c, h^c of the publications of a scientist need to have a score of $S^c(j) \geq h^c$ each [60]. The score can be determined as follows:

$$S^c(j) = \gamma \cdot (Y(now) - Y(j) + 1)^{-\delta} \cdot cit_j \qquad (12)$$

$Y(j)$ represents the publication year of the j-th paper and $Y(now)$ corresponds to the current year. The equation also contains the parameter δ, which may, for example, be set to 1. Furthermore, the citation count of the j-th paper cit_j and the coefficient γ are used in the equation. γ may serve as scaling factor to prevent a result being too small to obtain a meaningful index value. Thus, $S^c(j)$ represents the citation count of a publication divided by its age. In this way, older papers gradually lose their impact on the h^c-index, even if they still receive a few new citations [60].

2.8 Variants Considering Citation Age

The h-index has the drawback to ignore the age of the citations. This neglects that the citation age can show if a publication is still relevant in terms of receiving many recent citations or if the paper obtains hardly any new citations and is probably not relevant any more. The following section presents the h^t-index which considers the citation age.

2.8.1 h^t-index

To obtain an h^t-index of h^t, h^t of the papers of a researcher need to have a score of $S^t(j) \geq h^t$ each [60]. The score can be calculated with the following equation:

$$S^t(j) = \gamma \cdot \sum_{\forall x \in cit_j} (Y(now) - Y(x) + 1)^{-\delta} \qquad (13)$$

The used parameters are defined to be identical to those of the h^c-index (Section 2.7.2). $Y(x)$ represents the publication year of the x-th citation. The score is higher if the papers receive many recent citations and is lower if most of the citations are old. A benefit of the h^t-index is that it can show if a scientist's research work deals with current trends [60].

2.9 Variants Considering Publication Authors

Another disadvantage of the h-index is that it treats papers with many authors the same as papers with few authors ignoring the individual contribution of the author for whom the index value is calculated. In this section, we list different approaches to consider the authors of the papers. The h_m- and h_I-indices (see sections 2.9.1 and 2.9.2) are based on fractional counting of the publication authors, whilst the $\bar{h}$-index (see Section 2.9.3) uses an approach which compares the h-core of the co-authors with the h-core of the scientist for whom the index is calculated.

2.9.1 h_m-index

The h_m-index corresponds to the following equation defined in [59]:

$$h_m = max(j) \mid j_{eff}(j) \leq cit_j \qquad (14)$$

$j_{eff}(j)$ can be determined in this way:

$$j_{eff}(j) = \sum_{j'=1}^{j} \frac{1}{auth_{j'}} \quad (15)$$

The papers are counted by considering the number of authors $auth_{j'}$ of each paper j', where the papers are sorted descending by the number cit_j of citations. The h_m-index has the advantage that the influence of a paper on the index value decreases when the paper has more authors. This approach rewards researchers who publish alone or together with only a few colleagues [59].

2.9.2 h_I-index

The h_I-index is defined in [7] as follows:

$$h_I = \frac{h}{\overline{auth}(h)} \quad (16)$$

The divisor $\overline{auth}(h)$ contains the average number of authors. The h_I-index has the same advantages as the h_m-index, since the h_I-index decreases if a publication has more authors.

2.9.3 $\bar{h}$-index

The $\bar{h}$-index is defined in [29] as follows:

> *"A scientist has index $\bar{h}$ if $\bar{h}$ of his/her papers belong to his/her $\bar{h}$ core. A paper belongs to the $\bar{h}$ core of a scientist if it has $\geq \bar{h}$ citations and in addition belongs to the h-core of each of the coauthors of the paper."*

In order to determine the $\bar{h}$-index, the h-index is first calculated, which leads in the beginning to $\bar{h} = h$. Those papers, which do not belong to the h-core of the co-authors, are then removed from the $\bar{h}$-core. If there are any publications whose citation count is greater than or equal to the new $\bar{h}$ value and which also belong to the h-core of the co-authors, then these publications are added to the $\bar{h}$-core [29]. The $\bar{h}$-index has the advantage that it can detect if a scientist disproportionately benefits from his more widely known co-authors by receiving more citations than if she/he had co-authored the paper with researchers who are at the same scientific level as her-/himself [29].

2.10 Variants Considering Citation Authors

As pointed out in Section 2.2.3, the h-index has the drawback that it does not distinguish between self-citations and citations from independent researchers. In this section, we introduce two indices that take into account the citation authors. Section 2.10.1 describes the aH-index which reflects the responses to the work of a scientist in terms of how many other researchers cited the papers of this scientist. Section 2.10.2 presents the c-index which determines the collaboration distance between the cited scientist and the citation authors to weight self-citations differently than citations from independent researchers. Finally, Section 2.10.3 describes indices based on the PageRank algorithm [13, 47], which favors citations from highly influential papers and scientists.

2.10.1 aH-index

The aH-index is defined in [39] as follows:

> *"A scientist has aH-index a if a of the N_c researchers, that cite his or her work, cite at least a his or her publications each and the other $(N_c - a)$ researchers cite $\leq a$ publications."*

A high aH-index means that a scientist publishes a large number of papers and at the same time achieves a significant impact on the scientific community, due to the fact that many other researchers cite a great amount of her/his publications [39]. However, the construction of the aH-index may in some cases lead to a distorted perception of a researcher. As an example, if 20 scientists publish together an article in which they cite 20 different papers of a researcher, then this researcher would obtain an aH-index of 20, although this merely results from only one article with a high number of co-authors [39].

2.10.2 c-index

Definition

The c-index evaluates the collaboration distance between a cited researcher and citing researchers. The collaboration is close if the cited scientist has already published a paper together with the citing scientist, whereas the collaboration is loose if there are a couple of researchers between the cited author and the citing author. In other words, the principle of the index is based on the length of the shortest path, which is defined as follows: "Scientist a_0 has collaborated with scientist a_1, scientist a_1 has collaborated with scientist a_2, ..., scientist a_{l-1} has collaborated with scientist a_l." Note that a_0 corresponds to the cited author and a_l to the citing author [12].

After determining the length of the shortest path, the c-index considers the individual scientific collaborations along the path. According to [12], this can be described in a formal way as follows:

$$\sum_{i=0}^{l-1} \frac{1}{\mid p(a_i) \cap p(a_{i+1}) \mid} \quad (17)$$

The divisor corresponds to the number of papers published together by scientist a_i and scientist a_{i+1} [12]. If there is no collaboration path between cited and citing authors, the collaboration distance is defined as ∞ representing a citation with the highest degree of independence, whereas a self-citation corresponds to a distance of 0, which is a citation with the lowest degree of independence. To calculate the final value of the c-index, the citations to all publications are first sorted descending according to their collaboration distance. The further steps are similar to the calculation of the h-index, with the difference that the c-index additionally uses a coefficient α [12].

α is used to increase the expressiveness of the c-index, since most of the cited researchers have a short distance of not more than 10 to the citing researchers. For $\alpha = 1$ the c-index has analogous results compared to the h-index. Higher values of α put authors at an advantage whose citations have a high collaboration distance, whereas smaller α values result in a benefit for scientists with a large number of received citations. The more papers a researcher has published together with other scientists, the more likely it is that the received citations have a short collaboration distance, which narrows the interval in which the final c-index value lies. Hence, the determination of a useful value for α is not easy. α should therefore not be too high so that the c-index keeps its expressiveness. For a c-index of $c_\alpha(a)$ the scientist a needs to obtain $c_\alpha(a)/\alpha$ citations, which have at least a collaboration distance of $c_\alpha(a)$, while the rest of the citations has a distance of no more than $c_\alpha(a)$ [12].

Advantages

One of the benefits of the c-index is the compensation of the disadvantage of young scientists against older scientists, because scientists early in their career have a higher collaboration distance to other researchers. Hence, citations received by young scientists count more than those obtained by older scientists who already have close relationships with other researchers [12].

Calculation

The calculation of the c-index is based on Floyd's algorithm and uses a graph with n nodes and edge weights that indicate the distance between neighboring nodes. To obtain the collaboration distances at the time t when a citation occurs, the algorithm considers all nodes i and calculates for each of these nodes the shortest distance to all other nodes j. The result is a matrix $P = \{p_{i,j}\}$ with the size of $n \times n$, where $p_{i,j}$ represents the shortest distance between the nodes i and j [12].

The database that stores the information for the c-index corresponds to the above-described graph. Note that each node of the graph represents a researcher and that an edge between two researchers exists if they have at least one joint paper. The edge weight is defined as 1 divided by the number of joint papers of these two researchers. The calculation of the distances at a specific time t has to be performed at regularly recurring time intervals. The choice of the time interval depends on how accurately the date of the citations is recorded. Since at least the year of the citation is known, the calculation should be performed at least annually. If the citation record also contains the month and the day, the calculation could be performed monthly or weekly, which would lead to a more precise result [12].

When the calculation is finished, the matrix P_t is stored in the database. The matrix contains for each researcher i the collaboration distance to any other researcher j at the time t. If t_0 is the date when the distances are calculated for the first time, then another calculation has to be performed after the chosen time interval (weekly, monthly or annually) has elapsed. However, the first calculation of the distances takes more time, since not only the distances at the starting time t_0 have to be determined, but also the distances at any other time $t < t_0$. The smallest t, for which the distances have to be calculated, corresponds to the date of the oldest citation which is stored in the database [12].

To calculate a matrix P_t with $t < t_0$, a subgraph of the original graph has to be constructed. The subgraph is created by recalculating the edge weights, which were defined as 1 divided by the number of joint publications of two scientists. The number of joint publications is decreased by one for each paper that the two researchers published at a time $t^{'} > t$. If the calculation result shows that the two researchers do not have a single joint publication at the time t, then the edge between these two researchers is completely removed. Here, it is assumed that in the future only new publications and citations are inserted into the database and not old papers and citations. This aspect is quite important, since regularly adding old data requires the costly recalculation of the distances of the past every time. To sum it up, the accuracy of the c-index crucially depends on the size of the underlying database, while the efficiency mainly depends on whether or not old data is inserted into the database on a regular basis [12].

2.10.3 CiteRank and SARA-index

In [53], the authors propose the *Science Author Rank Algorithm (SARA)* and in [15] *CiteRank*, which are both inspired by the standard ranking procedure PageRank [13, 47] for webpages. CiteRank and SARA define the input graph for variants of the PageRank algorithm in a different way: CiteRank defines the publications themselves as nodes and the edges reflect citations. The nodes in the input graph of SARA are the authors and edges are created from the citing to the cited author. According to the PageRank algorithm each node is initialized with a unit of credit, which is then redistributed to their neighbors in several iterations. Hence not only publications with many citations, but also publications cited by highly influential publications receive higher ranks in CiteRank. SARA takes into account that citations from highly ranked authors are often considered to be more important than citations from low-ranked authors. The disadvantage of these approaches based on PageRank are that more or less closed communities of highly ranked scientists are favored, whereas it is for scientists outside these communities even harder to get a high rank.

2.11 Overview of the Considered Factors

This section provides an overview of the factors taken into account by the existing indices. For reasons of clarity, only the modern and complex indices are analyzed and not the classical indices. Table 1 lists the considered factors in the first column and contains the name of the respective index in the second column. Furthermore, the advantages and disadvantages of the respective index are shown in the third and fourth column. Since most variants only correct one or two drawbacks of the h-index, the majority of the drawbacks of the h-index still applies to its variants. For reasons of clarity, these disadvantages are only listed for the h-index. If one of the variants corrects a specific disadvantage, then this aspect is listed in the "Advantages" column of the respective variant. If a variant introduces a new disadvantage, then this aspect is shown in the "Disadvantages" column of the variant.

3 Further Related Work

Besides the previously discussed proposals of citation-based indices, some contributions like [2, 24, 48, 63, 68, 71] offer an extensive overview of the currently used bibliometric indicators. Additionally, they discuss some general problems like inter-field differences concerning the citation practice, the role of self-citations and the effect of multiple co-authorship. [10, 56, 62] provide a detailed comparison of existing indices based on real-world data and partially also on well-constructed theoretical cases.

Surveys like [6, 35] provide overviews over the scholarly bibliographic area and shed light on important aspects like collaborations and co-authorship, scientific journals and databases, trend and citation analysis. [18] examines the various aspects of the citation process. Other contributions like [1, 16, 44, 49, 61, 69, 70] deal with the importance of the citation context and the meaning of citations and offer classification schemes for citations. [46] investigates the reasons why some old papers are still highly cited many years after their publication. Furthermore, [42] addresses concrete difficulties for a bibliographic analysis of computer science publications.

The authors of [43] propose to use bibliometric data, which has been artificially generated through a model of citation dynamics calibrated on empirical data, in order to compare indicators for the impact of scientists. They envision their framework to become a standard tool for the assessment of impact metrics. Other contributions are dedicated to scholarly bibliographic databases. [40] discusses the various data formats, protocols and technical requirements of getting indexed by widely used bibliographic databases in the area of computer science and provides hints for maximal database inclusion. [19] analyzes the data in Google Scholar.

4 Conception of the mf-index

We describe the conception of the new index step by step in this section. We first work on general criteria to be fulfilled by the new index in Section 4.1. We describe in Section 4.2 the concrete factors that should be considered by the new index. Section 4.3 contains the design of the new index. Finally, we discuss the properties and benefits of the new index in Section 4.4.

4.1 General Criteria

We already enumerated the most important existing indices with their specific advantages and disadvantages in Section 2. We also examine the factors considered by the existing indices and these factors are summarized in Table 1. This table shows that researchers develop the existing indices in order to compensate certain disadvantages of the h-index. Indeed, most indices achieve this goal, but bring new disadvantages at the same time.

For example, the g-index (see Section 2.4.2) is one of the indicators considering the additional citations above h, which do not increase the h-index, but — due to the construction of the g-index — those researchers benefit disproportional, who have a great success with a single publication, although the rest of their publications has only few citations. This is even worser for the o-index (see Section 2.4.3), where the number of citations of the most cited paper is a direct factor.

Table 1: Considered factors, advantages and disadvantages of the existing indices

Factors	Index	Advantages	Disadvantages
h-core citations	h	- easy to calculate - balances productivity and impact - not vulnerable to rarely cited papers	- not comparable across different scientific fields - not comparable when career length is different - ignores excess citations above h - ignores non h-core citations - undervalues selective scientists - treats self-citations equal to independent citations - ignores the citation context - ignores the bias that favours well-known scientists - ignores the bias that favours review articles
	h_α	- similar h-values better comparable	- only comparable when the same α-value is used
	h^n	- normalizes the publication count	- rewards less productive scientists
h-core citations excess citations	$h^{(2)}$	- fast to calculate - considers citations above h	- ambiguous if citation distributions differ too much
	g	- considers citations above h	- vulnerable to a small set of highly cited papers
	m	- reduces impact of excess citations	- the used citation median is not always meaningful
	hg	- balances the h- and g-index - higher accuracy than h and g	- requires to calculate both h and g
	q^2	- balances the h- and m-index - higher accuracy than h and m	- requires to calculate both h and m
	A	- similar h-values better comparable	- discriminates scientists with a high h-index
	R	- similar h-values better comparable - eliminates the disadvantage of A	- vulnerable to a small set of highly cited papers
	j	- detects the exact citation distribution - not vulnerable to highly cited papers	- complicated to calculate by hand
	h_w	- shows research performance changes - the index can rise and fall	- complicated to calculate by hand
	h_d	- shows research performance changes - can detect aspiring scientists - the index can rise and fall - flexible configuration	- only comparable if the same $h(t)$ function is used - complicated to calculate by hand
h-core citations most cited paper	o	- considers most cited publication	- vulnerable to the most cited paper
h-core citations excess citations non h-core citations	h_{rat}	- higher accuracy than the h-index	- considers only a part of the excess citations - considers only a part of the non h-core citations
	h_T	- higher accuracy than the h-index - considers all excess citations - considers all non h-core citations - not vulnerable to extreme values	- complicated to calculate by hand
h-core citations career length	m-quo.	- no discrimination of young scientists	- discrimination of retired or dead scientists
h-core citations excess citations publication age	AR	- rewards actively publishing authors - the index can rise and fall	- discrimination of retired or dead scientists - missing threshold devalues old but seminal papers
	h^c	- rewards actively publishing authors - the index can rise and fall - too small values can be scaled	- discrimination of retired or dead scientists - missing threshold devalues old but seminal papers
h-core citations excess citations citation age	h^t	- shows authors dealing with trends - the index can rise and fall - too small values can be scaled	- discrimination of retired or dead scientists - missing threshold devalues old citations too much
h-core citations publication authors	h_m	- rewards publishing in small teams - discourages honorary authorship	- may discourage collaboration - treats main and minor contributors the same
	h_I	- rewards publishing in small teams - discourages honorary authorship	- may discourage collaboration - treats main and minor contributors the same
	$\bar{h}$	- rewards publishing in small teams - discourages honorary authorship - fairer than fractional counting	- may discourage collaboration - discriminates in some cases young scientists
h-core citations citation authors	aH	- shows the number of citing authors	- unjustifiably favors certain authors in some cases
	c	- emphasizes independent citations - too small values can be scaled	- completely excludes self-citations - calculation needs far more data than the h-index - complicated to calculate by hand
citation graph	SARA/ CiteRank	- weights citations of highly-influential authors/publications	- favors communities of highly ranked scientists - high calculation costs

However, also those indices, which do not have new serious drawbacks, have typically at least the weakness to focus on adjusting only one or at most two disadvantages of the h-index. Although some indices shrewdly adjust a certain weakness of the h-index, they cannot serve as single indicator of the performances of scientists, because they leave out many other important aspects. Hence, there is a need for a new index, which not only takes more than one or two factors into account, but combines several of them in order to offer balanced and objective performance evaluations and comparisons of researchers.

The existing bibliometric indicators can be divided into two basic categories: The first category contains indices like the h-index, m-quotient, g-index, o-index and $h^{(2)}$-index, which describe the most productive core of the researcher's performance and determine the number of publications belonging to that core. The indices of the second category like the A-index, m-index, R-index, AR-index and h_w-index measure the impact of the publications in that core [10]. The newly to be developed index should combine the two dimensions of performance evaluation.

The new index should also fulfill some other criteria. Its concept should be easily understandable and its calculation should not be too expensive. Moreover, the new index should be robust against incomplete information, such that its result is not strongly affected by incomplete information. It should also be considered that the concept of the new index can influence the (future) publication behavior of scientists. Hence, it needs to be decided carefully if the new index should focus on quantity or quality, or if it should balance these aspects. We should also investigate how the new index is influenced by extreme values like single, often cited publications and if these extreme values distort the results. In addition, the new index should have an informative value [56].

4.2 Factors to Consider

This section examines the concrete factors to be considered by the new index. Each factor is presented in detail in its own section including a discussion which of the existing indices consider this factor and which of these indices deal with this factor in the best way. In addition, we describe the difficulties mentioned in the literature regarding certain factors. Table 1 lists the basis of the choice of the factors, as it contains all aspects considered by the existing indices. Because of overlapping factors in the table, we summarize similar items in such a way that the new factor covers all relevant areas. Additionally, we refine some factors in order to increase their meaningfulness in comparison to the existing indices and to compensate their disadvantages. Furthermore, we introduce new factors, which are currently not considered by the existing indices, e.g., the comparability of scientists who work in different fields of study.

4.2.1 Total Number of Papers and Citations

The total number of publications and the complete set of citations are an essential factor to be considered in order to get a complete picture of the researcher's performance. The h-index only provides information about the most productive core of the scientist, the Hirsch core. For this Hirsch core, the quantity in the form of the number h is declared, which is specified to be the number of publications in the Hirsch core. At the same time, the minimum number of citations in the Hirsch core is expressed, which is at least h^2, because — based on the definition of the h-index — each publication in the Hirsch core must have at least h citations.

The information about all additional citations above h in the Hirsch core as well as those from papers outside the Hirsch core, which are cited less than h times, is lost. On the one hand, this is the purpose of the h-index and one of its advantages considering only the most productive core of a researcher. On the other hand, many not considered citations lead to discrimination of certain scientists and make it difficult to compare researchers with similar h-indices, if no other data is investigated.

The design of the new index should aim to incorporate all publications and citations of a scientist without weighting all in the same way, because a pure quantity like the total number of publications and citations is not reasonable. A large number of papers is a sign for a high productivity, but not necessarily for a high quality. A large number of citations can be caused by some few publications, although most of the publications may have only few citations.

Some of the indices like g, h_{rat}, h_T, j, $h^{(2)}$, A, R, m, hg and q^2 try to incorporate the citations above h in the Hirsch core and partially also the citations outside the Hirsch core. From these indices, the tapered h-index h_T considers the total number of publications and of citations in the best way. As explained in Section 2.5.1, all papers and all citations influence the calculation of the tapered h-index. Nevertheless, the tapered h-index offers a protection against extreme values like single, very often cited publications, as the citations decreasingly influence the index if the distance from the current rank is higher. By expressing the proportion between the received and the needed citations, each citation directly affects the index, such that the comparability of scientists with the same h-index is increased.

The calculation of h_T is certainly more expensive than those of other indices. Many indices are originally designed for manual calculations and hence should not be too costly in calculation. However, if the indices can be automatically calculated (as in publication repositories or bibliographic databases like Google Scholar), slightly increased calculation costs are no crucial disadvantage.

4.2.2 Age of the Scientist

The age of the scientist is a distortive element in the comparison of different researchers, as already explained in Section 2.2.3 about the disadvantages of the h-index. The m-quotient (see Section 2.6.1) is targeting at this problem by dividing the index value with the number of years since the first publication of the author. Young scientists are in this way not discriminated any more by the calculation of the index value. By the design of the m-quotient, researchers must have a constant productivity and the same or increasing impact in form of citations if they want to have the same or an increasing index value.

However, after a certain time older scientists are discriminated: Even if a researcher, who is, e.g., retired or even passed away, receives new citations, the number of new citations will be decreasing as the researcher does not publish any new papers any more. Generally, a flexible increasing and decreasing of the index value, depending on the current productivity and the impact of the researcher, is reasonable. However, the index value should not sink in an arbitrary way because of the age of the scientist in order to honor the lifetime performance of dead researchers like Alan Turing, even if the papers of these researchers only receive few new citations.

We propose hence to choose an upper bound for the distance to the first publication in order to avoid discrimination of those scientists, who indeed had an important impact in their area, but are already for some time inactive or dead. For the purpose of determining the upper bound we look at the length of the most productive phase of scientific careers. We assume the start of the publication time of a researcher in the age of approximately 25 years. In many cases the end of the publication time is in the age of 65 years. Hence the phase of scientists, in which they publish most of their papers, has a length of 40 years.

We could conclude that — analogous to the definition of the m-quotient — we should divide the index value by the career length until a career length of 40 years, and afterwards just only by 40. Older researchers have indeed a higher probability to be cited due to their higher name recognition, but this probability is seldom higher than the factor of 40 compared to young researchers even in extreme cases. For this reason, a lower upper bound than 40 is reasonable. We propose to use half of the originally conceived upper bound, i.e., 20 seems to be an adequate value[1]: In the first 20 years the index value should be divided by the exact career length and afterwards by 20. In this way, we can avoid to discriminate scientists at the end of their career by dividing their index value only by 20 instead of the high number of 40.

4.2.3 Age of the Publications

The age of the publications should also influence the calculation of the index. Typically, older publications hold more citations than younger ones, which did not have yet so much time to be cited. This aspect is different from the age of scientists. In fact, considering the age of the publications serves also for avoiding the discrimination of young researchers, but there are more aspects to look at.

Usually, two scientists have different productivities even if both publish already for the same number of years. One of the researchers could have finished or decreased her/his publication activities, because of a new job in industry, other time-consuming duties like teaching, work in standardization organizations, offering scientific services, or other reasons. Another researcher may also have a smaller need for publishing, because her/his successful papers of the past still receive many new citations. The other researcher may still publish many new results, because, e.g., the researcher could get funding for a big team. The different publication productivities should be considered in the calculation of the new index.

Especially the AR-index and the h^c-index take this aspect into consideration. Basically, both indices divide the number of citations by the age of the publication. The decreasing impact of the work can be compensated by new citations for some time, but most publications receive much less citations after some day leading to a drop of the index. It seems to be reasonable to also introduce an upper bound to the AR-index and the h^c-index, such that seminal papers are not depreciated in an arbitrary way only because they were published a long time ago. A reasonable upper bound could be 10 years. A higher value is unreasonable, especially because our new index considers the age of a publication as well as the career length of a scientist.

Although both factors address different aspects, they

[1] Other references assume an average career length of about 20 (with a high standard deviation of over 10 years), which corresponds to our assumption, see e.g. [45], [8] (for long-term scientists) and `http://www.sciencemag.org/careers/2013/07/want-be-professor-choose-math`

influence the index value of older scientists, as older scientists typically have a higher career length and some older publications. A higher upper bound for the publication age would hence discriminate older scientists, since we divide by the career length as well as by the publication age. The choice of 10 years as upper bound for the publication age should be an appropriate value independent of additionally considering the career length. An upper bound of 10 years has still enough precision, as it weights the publications concerning if they are from the last decade or at least 10 years old. Moreover, the exact age is used for younger publications. Hence, in total, an upper bound of 10 years has a good informative value.

4.2.4 Age of the Citations

Another important aspect is the age of the citations. The already mentioned issue of considering the age of the scientists compensates the disadvantage of younger researchers, while the influence of the age of the publications in index rank calculations balances the rank of newer publications in comparison to older ones. On the contrary, the age of the citations is an indicator for the publication being relevant independent of the publication age. Some old publications may have relevance for the current research and may still be cited, while other papers are not cited any more after some decades.

Among the mentioned indices, the h^t-index (see Section 2.8.1) considers the age of citations and constructs an index for trend detection. For combining the citation age with other factors, an approach, which divides each of the citations by its citation age, is also reasonable. According to [64], over 70% of total citations are not more than 10 years old. The analysis in [11] also supports the claim that most citations are in the time period up to 10 years after the publication of the cited work. Hence, an upper bound of 10 years is feasible, as in this way a citation is not devalued in an arbitrary way and the common consideration of the three factors (career length, publication age and citation age) does not lead to discrimination of older researchers.

4.2.5 Co-Authors of the Publications

The number of co-authors of a publication should also influence the calculation of the index value: There should be a difference if a contribution of a scientist is authored by the scientist on its own or by a big group of researchers. There are some few approaches to consider the number of co-authors during the calculation of the index value: One approach counts the number of publications subject to their number of authors, i.e., as $1/auth_j$ (the h_m-index does this in a similar way) [23, 59]. Another approach divides the number of citations of a paper by the number of its authors, i.e., $c_j/auth_j$ [23].

Furthermore, it is also possible to divide the final index value by the average number of authors calculated on all papers of the considered scientist (the h_I-index follows this approach) [7]. In contrast to the other approaches, the $\bar{h}$-index counts a publication only in the case if the publication belongs to the Hirsch core of the co-authors [29]. The most practicable approach should be the division of the number of citations of a publication by its number of authors.

4.2.6 Self-Citations and Colleague Citations

Self-citations and citations from colleagues should also be taken into consideration. First of all, we should investigate the existing literature about the influence of self-citations. Some researchers are of the opinion that self-citations strongly affect the h-index [57, 67], especially in the case of young scientists with low h-index [57]. According to [30], self-citations influence the h-index significantly in the case of an h-index lower than 10. The g-index is even more prone to self-citations compared to the h-index [58].

However, other researchers are of the different opinion that self-citations do not significantly affect the calculated index value. In order to increase the h-index by self-citations, a researcher must cite many of her/his own publications. However, it is difficult to predict which publications must be concretely cited in order to increase the h-index [26]. As the debate about the impact of self-citations on an index is controversial, also the drawn conclusions are different. Some researchers propose to exclude self-citations from the calculations of bibliometric indices [57, 58, 67], but other researchers do not see a need for it because of the limited influence of the self-citations [26].

As the opinions concerning the influence of self-citations are ambiguous, it seems to be reasonable to not fully exclude self-citations from the calculation of the index value. Analogous remarks are valid for citations of colleagues, as colleague citations have a higher significance than self-citations, although they are less significant than citations of external researchers. Before looking at existing indices, which consider self- and colleague citations, we should define the term colleague in this context.

An index considering the proximity between cited and citing author should take the date of the citation into account. In this way, only those scientists are defined to be colleagues, who worked together resulting in a joint paper in the past, i.e., in the time before the citation. It is important to consider this aspect, as otherwise cited and citing authors could rethink future cooperations in the form of joint papers, because of a retroactive negative effect on their indices [12]. As a consequence, colleagues as defined by the index do not necessarily have the same affiliation. The only determining fact is the presence or absence of joint publications in the past before the citation. As shown in Table 1, the c-index is the only index so far which considers self- and colleague citations. As described in Section 2.10.2, the c-index enlarges the investigation and determines the collaboration distance between cited and citing authors. The c-index marks self-citations with a distance of 0, and citations of completely foreign researchers with a distance of ∞.

In addition to direct colleagues (who published together with the cited researcher), the c-index considers also "colleagues of colleagues". Note again that colleagues as defined are also scientists having not necessarily the same affiliation. In this way, we can search for a shortest path from the cited scientist over the researchers with joint publications and their colleagues and so on to the citing scientist. If such a path exists, the distance is determined by considering the individual scientific collaborations along the path and summing up the fractions defined as 1 divided by the number of joint papers of two researchers. If no path exists, the distance from the cited to the citing scientist is defined to be ∞ as mentioned above.

As described in Section 2.10.2, the calculation of the c-index has in general a high complexity. The calculation of the c-index must be repeated periodically. Moreover, the costs are increased in case of adding especially old publications and citations to the database. The c-index defines the distance of self-citations to be 0, which is cutting the influence of self-citations too much compared to the controversial debate about self-citations as discussed above. Before incorporating the ideas of the c-index into the newly proposed index, we hence have to modify its concept, as the c-index cannot be combined with other indices like the tapered h-index. It seems to be reasonable to limit the properties of the c-index on a subset because of practicable reasons. We propose to weight citations according to three categories. The first category includes citations of completely foreign researchers, with whom the cited scientist did not publish together. The second category contains citations of direct colleagues, with whom the cited scientist has at least one joint paper before the citation date. The third category covers self-citations.

As discussed at the beginning of this section, self-citations should not be completely ignored in the calculation, such that we propose to weight them with 5%. Citations of colleagues could get a weight of 25% and the weight of remaining citations of non-colleagues, i.e., citations with the highest value, could be 100% in order to avoid a reduction of the index value in this case. The differences of the weights between citations of non-colleagues and those of colleagues and self-citations are in fact high, but they are feasible because of the investigated coherences as described in the literature. As pointed out at the beginning of this section, self-citations falsify the results significantly if the authors are young researchers or, in more general terms, scientists with a low number of publications and citations. Hence, the differences between the weights for the different categories of citations should be reasonably high.

To determine how large the differences between the weights should be, we performed an analysis after defining the new mf-index in Section 4.3: We investigated data sets with publications and citations of real scientists to identify under which circumstances the mf-index is vulnerable to self-citations and colleague citations. We refined our initial weights as long as the mf-index values calculated on our data sets could significantly be affected by these two types of citations. The resulting weights (25% for colleague citations and 5% for self-citations) that we propose in this section are low enough to eliminate the bias that these two citation types caused in our data sets, but are also high enough to prevent that too many citations are not considered during the index calculation. In Section 5.3.2, we show the results of an evaluation of the three weights for self-citations, colleague citations and non-colleague citations.

4.2.7 Comparability of Fields of Study

Over the years, researchers have developed different approaches to standardize the h-index in order to enable a comparison of scientists working in different domains [7, 31, 32, 50, 52]. The original h-index offers only a limited comparability, because the publication and citation frequencies as well as citation practices differ in different scientific communities of different domains [28]. For instance, papers in the field of mathematics usually receive significantly less citations than publications in the field of physics [6].

One method is to use normalization factors and especially the ISI field of study, which the considered researcher contributes to. By looking at the average number of citations per publication in the particular field of study and by declaring physics as reference field, a normalization coefficient can be constructed. The coefficient can be multiplied with the index to enable a comparison between scientists of different fields [31]. The method is a good basis for researchers of different fields. However, it only considers to multiply the final calculated index value of the scientist with the

Table 2: Correction factors for the fields of study [31]

ISI field of study	Factor
Agricultural Sciences	1.27
Biology & Biochemistry	0.60
Chemistry	0.92
Clinical Medicine	0.76
Computer Science	1.75
Economics & Business	1.32
Engineering	1.70
Environment/Ecology	0.88
Geosciences	0.88
Immunology	0.52
Materials Science	1.36
Mathematics	1.83
Microbiology	0.63
Molecular Biology & Genetics	0.44
Neuroscience & Behavior	0.56
Pharmacology & Toxicology	0.84
Physics	1.00
Plant & Animal Science	1.08
Psychiatry/Psychology	0.88
Social Sciences, general	1.60
Space Science	0.74

normalization coefficient. This method neglects that researchers can contribute to different fields.

An improvement of the method is correcting the factor for each publication according to the field of study. For example, if a scientist publishes in the field of mathematics as well as in the field of physics, then during index calculation the higher normalization coefficient for mathematics is considered for mathematical publications (receiving typically less citations), and the lower coefficient for physics for publications in the field of physics, as their citation frequency is usually higher.

A reasonable way is to normalize the number of received citations of the considered publication. In this way, the normalization coefficient is directly multiplied with that measurement, which its definition is referred to, i.e., in this case the number of citations. Table 2 enumerates the normalization factors of [31] with the first column containing the ISI field of study and the second column the correction coefficient.

4.3 Definition of the mf-index

We already developed general criteria for our proposed index and described factors to be considered in the previous sections. We present the design of our proposed index in this section. Section 4.1 contains not only the enumeration of the general criteria, but also the requirements for the new index like considering more than one or two factors in contrast to the existing indices by combining several aspects in order to achieve a balanced and objective performance evaluation of scientists.

For this purpose, we suggest the name multi-factor-index (abbreviated as mf-index) for our proposed index. We point out the most significant difference to the existing indices with this name: the consideration of multiple factors during the calculation of the index. We first define the mf-index in a general form in Section 4.3.1, which considers the factors described in Section 4.2, but let their weighting functions be abstract functions. We define a concrete mf-index in Section 4.3.2 by proposing concrete weighting functions for the abstract ones of Section 4.3.1. In this way, it will be easy to define variants of our mf-index in order to analyze different aspects by applying different weighting functions dependent on the main focus of the analysis.

4.3.1 Generalized mf-index

Equation 18 defines the generalized mf-index:

$$mf = \omega_{career}(y_{career}) \cdot h_T \tag{18}$$

h_T represents the tapered h-index, which we introduced in Section 2.5.1, but we redefine cit_j used in equations 10a and 10b later in Equation 20, such that we will receive another value for h_T in comparison to the original tapered h-index. ω_{career} represents a weighting function (defined in Equation 22) of the number y_{career}, which represents the number of years since the first publication of the considered researcher. The following equation contains the formal definition of y_{career}, where $Y(now)$ represents the current year and $Y(1)$ the year of the first publication:

$$y_{career} = Y(now) - Y(1) + 1 \tag{19}$$

In Equation 20, we redefine and calculate the sum cit_j of citations for each publication used in equations 10a and 10b:

$$cit_j = \left\lfloor \omega_{pub}(f_j, auth_j, y_j) \cdot \sum_{k=1}^{l_j} cit_{j,k} \right\rceil \tag{20}$$

We round the value of cit_j to the nearest integer, as the tapered h-index expects integer values for cit_j. We do not use the floor function in order to avoid that too many citations are not considered in the calculation of the tapered h-index, which would be a big disadvantage especially for young scientists who usually have lower citation counts compared to older researchers. f_j

contains the normalization coefficient for the field of study in which the j-th paper was published. $auth_j$ represents the number of authors of this paper. y_j represents the age of the j-th paper and is calculated by using $Y(pub_j)$ (the publication year of the paper) in an analogous way as y_{career}. These factors f_j, $auth_j$ and y_j are the input of the weighting function ω_{pub} (defined in Equation 23).

Equation 20 also contains a sum which is calculated by considering the original number l_j of citations to the j-th publication and summing the citation values $cit_{j,k}$ by recalculating them in the following way:

$$cit_{j,k} = \omega_{cit}(y_{j,k}, auth_{j,k}) \cdot \sum_{m=1}^{auth_{j,k}} \omega_{auth}(a_{j,k,m}) \quad (21)$$

$y_{j,k}$ represents the age of the citation, which is calculated in an analogous way as y_{career}, but considering the year $Y(cit_{j,k})$ in which the citation occurs instead of the publication year of the first paper $Y(1)$. $auth_{j,k}$ represents the number of authors of the k-th paper which cites the j-th publication. The weighting function ω_{cit} (defined in Equation 24) calculates a factor based on the age $y_{j,k}$ of the citation and the number $auth_{j,k}$ of citation authors. $a_{j,k,m}$ contains the m-th citation author of the k-th citing publication of the considered j-th cited publication. The weighting function ω_{auth} (defined in Equation 25) returns a calculated weight for this citation author.

4.3.2 Weighting functions of the concrete mf-index

We propose and discuss concrete weighting functions for our mf-index in this section.

Weighting function for career length: Our proposed weighting function correlates to the requirement described in Section 4.2.2 of avoiding an arbitrary decreasing of the index of inactive or dead scientists in order to value their lifetime performances in a reasonable way. We already justified in Section 4.2.2 an upper bound of 20 for the career length, which should be taken over for the value of t_{career} in the definition of the weighting function ω_{career}:

$$\omega_{career}(y_{career}) = \frac{s_{career}}{min(y_{career}, t_{career})} \quad (22)$$

As long as y_{career} contains at most the value of the threshold t_{career}, the current career length y_{career} is considered. If the value is greater than t_{career}, then only t_{career} is considered. s_{career} is a scaling factor for y_{career}, as the result of the division could result in too small values for a reasonable comparison between scientists. A reasonable value for s_{career} is the maximum value of y_{career} because of its upper bound for the career length. A suitable value for s_{career} is hence the number 20.

Weighting function for publications: The used upper bound t_{pub} for the publication age is the same as for the citation age and corresponds to 10 years as previously discussed in Section 4.2.3. The usage of a scaling factor is also here reasonable, as the divisions would otherwise lead to a number cit_j of citations which would be too small for determining the tapered h-index as defined in Equation 18. Like s_{career}, the value of the scaling factor s_{cit} should not be too small. Because of the high number of divisions in Equation 20 in comparison to Equation 18, s_{cit} should at least double s_{career}. Hence, we propose s_{cit} to be 50.

$$\omega_{pub}(f_j, auth_j, y_j) = s_{cit} \cdot f_j \cdot \frac{1}{auth_j} \cdot \frac{1}{min(y_j, t_{pub})} \quad (23)$$

Weighting function for citations: As citations are also publications and can also have more than one author, the average of the weighted citation authors is determined by using this weighting function ω_{cit}. The larger the fragment of non-colleagues is under the citation authors, the higher is also the overall summand in Equation 21. The reversal conclusion is that the summand is smaller when the fragment of citation authors contributing to the cited publication is larger. The upper bound t_{cit} for the citation age should be 10, as described in Section 4.2.4:

$$\omega_{cit}(y_{j,k}, auth_{j,k}) = \frac{1}{min(y_{j,k}, t_{cit})} \cdot \frac{1}{auth_{j,k}} \quad (24)$$

Weighting function for citing authors: There are two possibilities for using the weight for self-citations: The first possibility is that the citation's origin is directly the researcher for whom the index is calculated. In the second possibility, the author of the citation is a co-author of the cited publication, such that this author not only cites the considered researcher, but also her-/himself.

The weight for colleague citations is used whenever a citation author worked together with the considered scientist in the past (in form of a joint paper), but is not a direct co-author of the j-th publication. All citation authors neither citing themselves nor being colleagues

Table 3: Overview of thresholds and scaling factors

Name	Type	Used for	Value
t_{career}	upper bound	career length	20
t_{pub}	upper bound	publication age	10
t_{cit}	upper bound	citation age	10
s_{career}	scaling factor	career length	20
s_{cit}	scaling factor	citation count	50

of the considered researcher are considered as non-colleagues, i.e., as foreign scientists.

$$\omega_{auth}(a_{j,k,m}) = \begin{cases} 1 & a_{j,k,m} \text{ is a non-colleague} \quad (25a) \\ \frac{1}{4} & a_{j,k,m} \text{ is a colleague} \quad (25b) \\ \frac{1}{20} & a_{j,k,m} \text{ is author of } j \quad (25c) \end{cases}$$

With this definition, the value $cit_{j,k}$ in Equation 21 of a citation can be at most 1, which is the case when the citation is from the current year and when all citation authors are non-colleagues.

Used upper bounds and scaling factors: During the previous paragraphs, we defined some upper bounds and scaling factors and proposed values for them. Table 3 provides an overview of them. The first column contains the name as used in the previous equations. The second column shows the type (e.g., upper bound) and the third column the aspect (e.g., career length) for which it is used. Finally, the fourth column contains the proposed value.

Discussion of variants of the mf-index: Variants of the mf-index can be easily developed by using other weighting functions. Indeed with the right choice of weighting functions, the mf-index variant becomes equivalent to the original tapered h-index, i.e., the general mf-index is a generalization of the tapered h-index. Other examples include to apply logarithmic functions instead of using divisions with an upper bound in order to support soft transitions of the values.

By redefining the weighting functions, one can activate or deactivate single factors of the proposed mf-index. Other weighting functions could also emphasize different aspects like, e.g., valuing collaborations. Valuing collaborations may result in more co-authors, whereas it may result in fewer co-authors claiming that the contribution of each single co-author to a paper with many co-authors is less compared to papers with fewer co-authors (according to which our current weighting functions are designed). The investigation of these variants of the mf-index will be part of our future work.

In the following, we analyze the proposed mf-index and weighting functions in more detail. We use the name mf-index in the following sections for the concrete mf-index using the weighting functions and the thresholds of Section 4.3.2, and generalized mf-index for our generalization proposed in Section 4.3.1.

4.4 Properties and Benefits of the mf-index

We discuss the properties and benefits of the proposed mf-index in this section. The mf-index considers the general criteria as described in Section 4.1, and fulfills the requirement to combine several aspects instead of only focusing on one or two aspects. The mf-index connects both dimensions of performance evaluation, since it measures productivity (the first dimension) and impact (the second dimension). The mf-index achieves this goal by using the tapered h-index, as the tapered h-index considers all publications and weights at the same time the received citations.

The mf-index also fulfills the other general criteria. Although it considers many different factors, its conception is still easily understandable. Its calculation is indeed more costly in comparison to the h-index, but the costs are at an appropriate rate because of the many additionally considered aspects. The mf-index is also robust against missing information. Especially in border areas data holes may occur. On the one hand, if there are some citations missing for less cited papers, this does not falsify the result much because of the construction of the used tapered h-index, which indeed considers seldom cited publications. However, the impact of less cited papers on the final result is limited in comparison to the most productive core of the researcher. On the other hand, if citations of frequently cited publications are missing, then this also has no significant influence on the final result: The tapered h-index weights citations less if they are more away from the current rank, thus weakening the impact of extreme values and missing citations.

The design of the mf-index also addresses the additional criterion of the effect of the index on the publication behavior of scientists. As the mf-index has a focus on a balanced rate of quantity and quality, this should lead to no negative change in the publication behavior of researchers. Nevertheless, from the construction of the mf-index one can infer which publication and citation behavior has a positive or negative respectively influence on the index value. Due to the weights of the citations, the index value increases if a scientist gets attention of foreign researchers (non-colleagues), while the index value is lowered if the scientist's papers are only noticed by her/his colleagues and are only cited by them.

Another general criterion is the impact of extreme values. Extreme values do not falsify the result much, because, due to its construction, the used tapered h-index less weights highly cited publications in a reasonable manner. The last criterion is that an index should have a good informative value. This should also be the case for the mf-index, as it considers all factors listed in Section 4.2 and hence takes a lot of aspects like productivity, impact, age of the scientist, age of the publications, age of the citations, number of co-authors, distance between the cited author and the citing authors, field of study of the publication into consideration. If a scientist receives a high index value under the impact of all these aspects, then this should be a sign for the quality of her/his work.

5 Evaluation

This section deals with the evaluation of the proposed mf-index. Section 5.1 describes the chosen methodical approach and Section 5.2 addresses the data set and its properties chosen for the evaluation. We describe and analyze the results in Section 5.3.

5.1 Methodical Approach

We will first investigate existing indices and their values for some real-world data of scientists before we analyze the mf-index separately in more detail. Thereafter, we compare the existing indices with the mf-index.

In order to receive meaningful results, we will consider the real-world data of researchers of different types. We propose to investigate three different classes of scientists (Nobel laureate, professor, research assistant) and to choose scientists of different fields of study. Based on the chosen scientists, we first calculate their values of the existing indices and analyze these values. Afterwards, we investigate the values of the mf-index based on the same chosen researchers. As the mf-index consists of a high number of factors, we will investigate each factor separately in more detail.

For this purpose, we will check the influence of each factor by activating the considered factor (e.g., the correction according to the field of study) and deactivating all the other factors. The determined different index values can then be interpreted in the context of the scientific careers of the considered researchers like number of publications, number of citations and career length in order to discuss why the activation of specific factors influences the index value of some researchers strongly and of others not significantly.

We propose to additionally analyze the chronological sequence of the index values of different scientists by calculating the values for different years. In this way, we investigate if the index value of a certain researcher has strongly risen, stagnated or fallen in comparison to previous years, which leads to conclusions of changes of the productivity or of the impact of a researcher.

Finally, we analyze the calculated index values of the mf-index after activating all factors based on the complete career lengths of the considered researchers. The succeeding comparison of the mf-index values with the values of existing indices will shed light on questions like why the index values of the mf-index differ much for some scientists and why for some other researchers the differences are not significant. By using real-world data of existing scientists, we will show that the practical results of the mf-index verify the theoretical advantages discussed in Section 4.4.

5.2 Selection of the Data Sets

This section deals with the choice of the concrete data sets for the evaluation. We will deeply analyze the index values for these chosen data sets. Because of the extensive analysis, its deep interpretation and the extensive description of the results, the data set covers the data of only some researchers. We hence choose for each category two researchers and in total six researchers: two Nobel laureates, two professors and two research assistants.

The fields of study of the scientists should come from Table 2, such that the correction of the mf-index for the field of study can be evaluated. Since some fields of study may not have suitable candidates for the different categories, it is possible that the chosen researchers are not all from different fields of study. The choice of a scientist from a too special field of study takes the risk of having too few data of publications, citations and so on for a reasonable comparison with other researchers. Hence, we choose scientists of fields of study with a bigger community like physics, chemistry and mathematics. Although a data set of six researchers seems to be limited on a first view, we consider the analysis results nevertheless meaningful, because we carefully choose the researchers based on the criteria described above.

We use the publication and citation data of the Microsoft Academic Graph (MAG)[2] for our evaluation. MAG contains about 130 million papers authored by about 115 million authors [27]. MAG supports an own taxonomy of field of study. However, according to [27], only about 33% of the publications in MAG are assigned to one or more field of study entities. This rate of publications linked to field of study entities is too low for a serious analysis. Hence, we neglect the correction concerning the field of study in our analysis. Special

[2] `https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/`

Table 4: Data of the selected scientists

	N1	N2	P1	P2	R1	R2
Career length	52	49	50	23	14	13
Number of publications	8	54	251	23	45	24
Number of co-authors A	1	3	1	1	3	5
Number of co-authors M	0	2	0	1	2	4
Publication age A	31	21	20	12	8	7
Publication age M	25	19	22	12	8	7
Number of citations	1215	3981	3787	586	117	182
Number of self-citations	1%	1%	7%	5%	53%	3%
Number of colleague citations	0%	1%	0%	4%	6%	2%
Number of non-colleague citations	99%	98%	93%	91%	41%	95%
Citation age A	9	11	9	12	7	6
Citation age M	5	10	6	12	8	6

care must be taken for retrieving all the publication and citation data of a concrete author without having wrongly associated publications and citations: The authors are not necessarily disambiguated (i.e., MAG mixtures the data of two or more authors with the same name) and one author might be written in different ways (i.e., abbreviating the first name or using the full first name, with or without the middle name and so on).

Table 4 contains the properties of the chosen scientists, which we have made anonymous in order to protect the chosen scientists, because our detailed analysis may show sensible information. We provide the average (abbreviated as A) as well as the median value (abbreviated as M) of the number of co-authors, publication age and citation age, as there could be wrong conclusions drawn for some researchers when considering only the average value. The value of self-citations in Table 4 is determined according to the same rules as for weighting the citation authors in the mf-index, because this simplifies the interpretation of the index values. The value of self-citations represents the number of citation authors citing a paper of the considered researcher, which they co-authored.

The value of self-citations is hence not (!) the same as the number of citations in which the considered scientist cites an own publication, but how many self-citations are among her/his received citations by also looking at her/his co-authors. Or in other words: Self-citations of co-authors to a publication of the considered scientist are also counted! Furthermore, we cannot assume completely correct data of a researcher in case of automatically determined data sets like MAG, which is at least partly based on crawled data. Random samples support this claim even in the data of the chosen scientists: For example, there is one paper assigned to one of the chosen scientists, which does not have a reasonable publication date in comparison to the career of the considered scientist.

Looking at the researcher's webpage, the mentioned paper is not listed. However, also other bibliographic databases associate the mentioned paper with our chosen scientist, such that wrong data is not only a problem of MAG, but looking at the large bibliographic data sets probably of all bibliographic databases. Wrong data may influence the index value of a researcher and hence also the preciseness of our analysis in Section 5.3. For some cases, wrong data may lead to a higher index value, because additional publications are considered for a scientist. For other cases, wrong data may lead to a lower index value because of, e.g., an increased career length, which is one of the factors of the mf-index (see Equation 18). However, in our considered case, the career length without the mentioned wrongly assigned paper is above the upper bound of 20 years, such that the mentioned wrongly assigned paper does not have a big influence on the calculated index value.

For showing the practicability of the proposed mf-index, we do not manually correct the data, as the index rank calculations are anyway done based on these not completely correct bibliographic databases. However, the quality of the bibliographic databases is still high, because most of the publications and citations are correctly assigned and the few wrongly assigned papers hence do not significantly influence the calculated index values. Indeed, wrong data and incomplete information typically only vary the index values slightly compared to their absolute value based on our observations. For the comparison between the different indices, we argue that the data is for each considered index the same and hence the comparison is valid. If the indices are calculated based on the data of institutional repositories, then the data quality may be better (because of a better administration of the data) leading to preciser calculated index values. However, today's institutional repositories

Table 5: Values of the existing indices

	N1	N2	P1	P2	R1	R2
h	5	18	18	10	6	4
h_T	12.19	32.6	35.04	18.52	9.73	8.6

do not collect citations which are essential for the calculation of the index values.

5.3 Results

This section deals with the evaluation results and their analysis. First, we discuss in Section 5.3.1 the results of the existing indices. Afterwards, we analyze the mf-index in Section 5.3.2 in more detail. Finally, we compare the mf-index with the existing indices in Section 5.3.3.

5.3.1 Analysis of the Existing Indices

As argued above for the limited data set of six scientists, we will perform an extensive analysis, give deep interpretation and present the results in detail. Therefore, since some highly recommendable papers like [10, 56, 62] already deal with an extensive comparison of existing indices based on real-world data and partially also on well-constructed theoretical cases, we only consider two of the existing indices for the separate analysis and their comparison with our mf-index.

We select the h-index and the tapered h-index for the following reasons: The h-index is the basis for most of the indices developed after the initial h-index proposal by Hirsch in 2005 and should hence be used as reference in an analysis. The tapered h-index is, to our knowledge, the one of the existing indices which is best designed and has the most advantages and the fewest disadvantages. Furthermore, the tapered h-index is used by the mf-index (as shown in Equation 18) and should therefore also be used as reference.

In sum, we argue that the selection of two existing indices is sufficient for the purpose of this paper and that we carefully chose those two indices which are best suited for a meaningful analysis and comparison with the mf-index. Table 5 contains the index values of the h-index and the tapered h-index for the chosen scientists.

The index values of the two Nobel laureates differ much from each other. Both have a similar career length (N1: 52 years, N2: 49 years), but the Nobel laureate N2 published 54 contributions and N1 only 8. Indeed, the Nobel laureate N1 once claimed in an interview that she/he is not productive enough for today's research systems, where the number of published contributions is one of the most important factors.

It is accordingly not astonishing that the value of indices with the number of publications as one of their key factors is relatively low for researchers like N1. For example, the highest possible h-index value corresponds to the number of papers, which is in the case of N1 not an appropriate upper bound. The tapered h-index is a little bit higher, because the 1215 received citations of N1 influence more the index value than the one of the h-index.

An interesting observation is that the Nobel laureate N2 and the professor P1 (who did not receive the Nobel prize) have identical h-index values and similar tapered h-index values. Obviously, P1 published more contributions (251) than N2 (54). However, both researchers received a similar number of citations (N2: 3981, P1: 3787). P1 published an important contribution in one of her/his research areas, which may explain why she/he obtained about the same number of citations as the Nobel laureate N2.

The index values of professor P2 are about half of those of P1 and N2 respectively. One reason could be that P2 has published only for 23 years, which is about half of the time that P1 and N2 are active in research. Furthermore, P2 received less citations (586). P2 is mathematician and her/his publications discussing mathematical problems receive according to [31] on average less citations than those of physicians (like P1) and chemists (like N2). As the h-index and the tapered h-index (in contrast to the mf-index) do not consider the field of study, P2 is discriminated in the calculation of the index values (of the h-index and the tapered h-index) receiving too low index values because of her/his field of study. As a matter of fact, the h-index and the tapered h-index should only be used to compare scientists within the same field of study and not across different fields of study.

According to Table 5, the index values of the research assistants R1 and R2 are significantly lower than those of the professors P1 and P2 and of the Nobel laureate N2. The Nobel laureate N1 is an exception as we have discussed before. Lower index values of research assistants in comparison to professors and Nobel laureates are reasonable.

5.3.2 Analysis of the mf-index

We will evaluate and analyze the proposed mf-index in this section. First, we investigate each factor of the mf-index before we analyze how the index values changed over time for the considered scientists. Finally, we look at the total values, i.e., those index ranks, which are determined when all factors are activated for all available publications of the considered scientists.

Table 6: Effects of the age factor

	N1	N2	P1	P2	R1	R2
Before	12.19	32.6	35.04	18.52	9.73	8.6
After	12.19	32.6	35.04	18.52	12.97	12.29

Table 7: Effects of the citation age factor

	N1	N2	P1	P2	R1	R2
Before	12.19	32.6	35.04	18.52	9.73	8.6
After	11.8	28.11	28.86	17.17	8.58	8.75

Analysis of the Different Factors

In order to examine each factor of the mf-index on its own and its impact on the index values of the considered scientists, we activate only the currently examined factor and deactivate all the other factors. In this way, we determine those researchers whose index values are (not) significantly influenced by the considered factor of the mf-index. Based on the data concerning the scientists' careers given in Table 4, we interpret the influence of each factor on the mf-index for different types of researchers.

Table 6 presents the effects of the age factor on the mf-index. The row before shows values for deactivated career length correction (also with deactivated scaling factor s_{career}, i.e., $s_{career} = 1$). The row after shows values for activated career length correction/age factor (also with activated scaling factor s_{career}). Not surprisingly, the mf-index values without correction are identical with the tapered h-index values, as the definition of the mf-index without correction is equivalent to the definition of the tapered h-index. In the following analysis, we always compare the index values after deactivating all factors (which are identical to the tapered h-index values) with the index values after activating one of the examined factors.

The age correction does not have any effect on the index values of the researchers N1, N2, P1 and P2. These scientists already passed a career length of 20 years (see Table 4). Hence, the age correction results in a division by 20 (the upper bound of the career length). Since the scaling factor s_{career} is also 20, the index value remains the same as before the division. However, the index values of the research assistants are significantly increased by the age correction. The reason is the shorter career length of R1 (14 years) and R2 (13 years). Obviously, the goal of the age correction is fulfilled: Younger researchers get a compensation for shorter research time than the older scientists. The effect is even bigger for researchers at the beginning of their career, who are younger than R1 and R2. Scientists with a longer career length are not discriminated either, as after a career length of 20 years the divisor remains 20 and is not the actual career length any more.

We now examine the influence of corrections based on the properties of citations and the citation count of publications. To scale the values after the corrections, the scaling factor for citations s_{cit} is used according to Equation 23. As described in Section 4.3.2, the default value of s_{cit} is 50, since usually all factors of the mf-index are activated, which leads to a high number of divisions in Equation 20. Without an adequately high scaling factor, the citation count would get too small.

However, in this analysis only one division is performed, since we evaluate each factor on its own by only activating the currently analyzed factor and deactivating all the other factors. The highest possible divisor in the following citation age correction is the upper bound of 10 years for the citation age. For some of the six scientists, the median of the citation age is even significantly lower than 10 (e.g., Nobel laureate N1 with a median of 5 years). Hence, it seems reasonable to choose the scaling factor s_{cit} to be half of the upper bound, i.e., $s_{cit} = 5$.

Table 7 lists the index values after activating the citation age correction. Comparing the index values with and without activated citation age correction, the values of N1, P2 and R1 differ only slightly. The index value of R2 is even slightly increased. In contrast, the index values of N2 and P1 are significantly dropped. A first assumption is that the scientists with about the same index value with and without correction have newer citations than those researchers with a significantly decreased index value.

According to Table 4, N1 and P1 have about the same average citation age, but the index value of N1 remains nearly stable while the index value of P1 is significantly decreased. Analogous remarks apply for researchers N2 with a significantly decreasing index value and P2 with a nearly stable index value. These phenomena are explainable: After the citation age correction we still apply the tapered h-index for determining the final value of the mf-index. The tapered h-index weights citations differently (see Equation 10a and Equation 10b).

In fact, all citations are considered for the calculation of the tapered h-index, but the most important papers of the scientists (i.e., the Hirsch core) influence the index value most (as for the h-index). Citations of less cited papers have a decreased weight in order to avoid similar index values for researchers with a large set of rarely cited papers in comparison to scientists with fewer publications but having many citations. Furthermore, according to Equation 10b, citations of highly cited papers gain less weights if the citation count of the considered publication is higher. This avoids similar index values of researchers with a single successful publication in comparison to scientists who continuously publish contributions that receive a high attention in the scientific community.

Taking the previous thoughts into account, we can draw some conclusions on why the division by the citation age influences the index values of some researchers more than ones of other scientists. If a citation of a less cited publication is already quite old, then this citation decreases only slightly the index value, as citations of seldom cited publications anyway do not influence the index value much because of the tapered h-index. However, this is a desired result: If a rarely noticed early contribution of a researcher receives only few citations over some decades, then the index value is later only little lowered, as these early publications anyway do not belong to the set of high impact papers of the scientist. Similar remarks apply in the case of older citations of an often cited paper if the number of the older citations is low. In this case, the age of these citations does not influence much the index value, since the citations are less considered by the tapered h-index when the citation count is higher.

This phenomenon should also be considered: An old publication being cited over hundred times is still seen as up to date, even if some citations are already older, but most of the citations are newer ones. For example, the Nobel laureate N1 published one contribution in the 1960s, which received only few citations in the 1970s and 1980s, but half of the citations (of over 350 in total) are from the last 5 years. This may be a hint why the index value of N1 decreased only a little bit.

However, if the index value of a scientist is much lowered by the citation age correction, then the reason could be that one or more of her/his most important papers has many older citations. The decrease of the index value can be justified considering that an important contribution of the scientist is obviously not up to date any more, such that this contribution is not cited in the recent time any more. Taking all into account, the activation of the citation age leads to the desired effect that the index value is influenced by how up to date the publications are.

We now examine the publication age and its influence on the value of the mf-index. Table 8 lists the index values of the six chosen scientists after activating the publication age in the calculation of the mf-index. As before, the value of the scaling factor s_{cit} is chosen to be half of the upper bound (in this case of the publication age), i.e., $s_{cit} = 5$.

It can be seen that the index values after the publication age correction have fallen for all six researchers. However, for some scientists the decrease is only slight, while it is significant for other researchers. The decrease of the values correlates relatively well with the median of the publication age: The publication age's median of R2 is 7 years, which is the smallest value in comparison to all chosen researchers. Indeed, the index value of R2 has only slightly fallen (0.97). The median of R1 is already 8 years leading to a difference between the index values of 2.56. The increased median of 12 years of professor P2 results in an increased difference of 3.41 between the two index values.

Table 8: Effects of the publication age factor

	N1	N2	P1	P2	R1	R2
Before	12.19	32.6	35.04	18.52	9.73	8.6
After	10.4	25.28	25.76	15.11	7.17	7.63

N2's median of 19 years decreases the index value by about 7.32. A difference of 9.28 of the index values in the case of P1 is caused by her/his median of 22 years. The index value of N1 is, surprisingly, only lowered by about 1.79, although N1's median of the publication age is the highest (25 years). The decrease for N1 is comparable to the ones of R2 and R1, which have the smallest medians of 7 and 8 years. This could be caused by the fact that N1 published only 8 papers. Conclusions based on the median are often wrongly drawn in case of such small numbers in combination with irregular publication behavior. Two of N1's publications are quite young (8 and 9 years old) in comparison to the other publications being over 23 years old. Her/his three oldest publications are even more than 50 years old.

Because of the upper bound of the publication age, also for N1's papers, the publication ages will be at most divided by the value 10. As discussed in Section 4.2.3, the upper bound of 10 for the publication age seems to be reasonable. A larger upper bound would devalue, e.g., one of the contributions of N1 dated in the 1960s, for which N1 got the Nobel prize later on. With an upper bound of 10 years, six publications of N1 are divided by this upper bound 10, but two contributions only by 8 and 9 respectively. Considering all these facts, we conclude that the correction of the publication age serves its purpose to balance the index values between older papers having more time to receive citations (without devaluing by having an upper bound of the publication age) and younger ones.

In the following paragraphs we examine the influence of the author number of a publication on the index value. Table 9 lists the values of the mf-index before and after activating the correction concerning the author number of a publication. Since the median of the co-author number of five out of the six scientists is at most 2, we choose the scaling factor s_{cit} to be

Table 9: Effects of the publication authors factor

	N1	N2	P1	P2	R1	R2
Before	12.19	32.6	35.04	18.52	9.73	8.6
After	13.16	30.3	42.77	18.18	7.83	6.39

Table 10: Effects of the citation authors factor

	N1	N2	P1	P2	R1	R2
Before	12.19	32.6	35.04	18.52	9.73	8.6
After	12.09	32.05	32.34	17.8	6.63	8.11

2. The number of publication authors is typically less than the citation age or the publication age. Hence, we divide by smaller numbers in comparison to the previous considered corrections concerning the citation and publication ages.

After performing the correction of author number, some researchers (N1 and P1) have higher index values than before. The index values of other scientists (N2, R1 and R2) are decreased or are about the same (P2). N1 and P1 have mostly published alone and only in rare cases with co-authors: The average number of co-authors of both scientists is 1 and the median even 0. P1's index value increased more than the one of N1, because her/his large number of publications have mostly no or at most one co-author (besides P1). In comparison to P1, N1 published only 8 papers, one of which has 2 co-authors and another one even 7 co-authors. This explains the smaller increase of N1's index value compared to P1's index value.

The median and the average of the number of co-authors in the case of P2 is 1, which explains that her/his index value is not much influenced by this correction. The number of co-authors of publications of the researchers N2, R1 and R2 is higher than those of N1, P1 and P2, such that their index values decreased. However, the decrease is only small, even in the case of R2 who has an average of 5 co-authors. This seems to be justified, since the number of co-authors should affect the index value, but not in a too strong way to avoid discrimination of scientists who publish in large teams. Hence, also the correction concerning the number of publication authors serves its purpose: The index values of scientists publishing alone or only with fewer co-authors increase, and a high number of co-authors decrease the index value, since we assume that each author has a smaller contribution in case of many co-authors.

Finally, we investigate the effects of weighting the citation authors. Table 10 presents the mf-index values before and after activating the citation authors weights. We do not additionally scale the retrieved values, as we usually do not divide by large numbers for the correction according to the citation authors.

According to Table 10, the index values of N1 and N2 are only slightly decreased. This fits to the observation that N1 received 99% citations of non-colleague authors and N2 98%. We notice also for R2 only a slight difference after the decimal point, which is reasonable looking at the high rate of non-colleague citations of 95%. On a first view, it seems unusual that R2's value has even slightly increased, since, according to the equations 25a, 25b and 25c, the value for a citation only ranges from 0.05 (for self-citations) to at most 1 (for non-colleague citations) and no scaling factor is used in this particular analysis. However, this can be explained by the fact that the citation count of a paper is rounded before calculating the tapered h-index in Equation 18, since the original tapered h-index expects the citation count to be a natural number and not a real number.

The citation count is rounded depending on the decimal places and not always rounded down to the next natural number, since otherwise too many citations would be excluded from the calculation, which would deteriorate the accuracy of the mf-index. When the citation count is rounded up, this may lead to a slightly higher value of the final mf-index. Since the rounding usually only affects the decimal places of the final index value (as for R2), this should be a negligible effect.

P2's difference of the index values is 0.72 and this is caused by a higher rate of self-citations (5%) and colleague citations (4%). P1's index value is even reduced by 2.7. This fits to the fact that P1's rate of self-citations is already 7%. R1's index value is decreased most with a difference of 3.1. This is mainly caused by the high rate of self-citations of 53%. Moreover, R1 received 6% colleague citations, such that only 41% of the citations are from non-colleagues. However, we notice that the absolute decrease of the index values is about the same for R1 (3.1) and P1 (2.7), although R1 has much more self-citations (53%) than P1 (7%). Looking at the relative differences, the picture needs to be redrawn: P1's decrease is only 8% of her/his index value, while R1's decrease is about 32%.

The decrease of 32% of the index value is justified by the high number of self-citations, such that the comparison to the other of our chosen researchers with a rate of at least 91% non-colleague citations becomes fair. Furthermore, R1's index value of 6.63 is still in a reasonable range, which is not out of scale despite the high rate of self-citations. Looking at the analysis, we conclude that the citation author weights achieve the aim of calculating fair index values considering the rate of non-colleague, colleague and self-citations.

Table 11: Changes of the mf-index over time

	N1	N2	P1	P2	R1	R2
1970	11.8	9.4	16.32	0	0	0
1980	7.58	11.1	7.05	0	0	0
1990	6.73	7.19	25.03	0	0	0
2000	7.06	20.97	17.54	25.7	0	0
2010	10.69	22.48	29.83	17.64	11.71	17.55
2011	12.01	21.8	29.8	18	8.88	15.69
2012	12.16	21.55	28.22	15.98	9.33	13.28
2013	12.82	21.56	29.89	15.26	9.97	12.15
2014	12.39	21.83	28.66	14.89	9.31	11.6
2015	11.95	20.8	27.26	14.46	6.1	9
2016	11.07	19.45	24.05	12.63	4.39	7.23

Table 12: Total values of the mf-index

	N1	N2	P1	P2	R1	R2
Values	11.07	19.45	24.05	12.63	4.39	7.23

Analysis of the Temporal Development

The productivity of researchers and their impact to the scientific community change over the years. Published contributions and received citations of scientists may be at a peak for some years, and in other years a researcher's productivity may stagnate holding at least a constant level of research. In some years, the scientist may also publish not so many contributions any more or may deal with a cold topic receiving only few citations for this reason. We can also calculate the mf-index value by limiting the considered publications and citations to a certain range of years. In this way, we can determine the impact of a researcher in a given decade (e.g., in the 1990s) or even in a given single year, and can examine the changes of the researcher's productivity over time.

Table 11 enumerates the index values of the chosen scientists for several years. Note that all factors of the mf-index are activated for these values except of the correction of the field of study. Only few of the chosen researchers have been active in research for many decades, such that the index values are given in bigger intervals for the first years.

Looking at Table 11 and ignoring small decreases in some years, the index values of the Nobel laureates N1 and N2 remain on a constant level for the first decades. Only since the turn of the millennium, the index values of N1 and N2 increase significantly, where the increase of N2 is much more than the one of N1. We could draw the conclusion that it took a while until the scientific community recognized the contributions of N1 and N2 respectively or until their contributions have been verified and accepted to be right. After recognizing their contributions, their index values remain relatively constant. As they currently only publish few contributions because of their age, a stable index value indicates that they still receive many citations. Otherwise, the index value would have already been fallen, as the mf-index includes many temporal dependent factors in its calculation.

Looking at the temporal development of the index value of P1, we can sometimes detect a slight decrease in her/his value, but overall the index value is increased and afterwards it is stable. This can be explained by P1's high productivity and hence stable impact on the scientific community. In contrast to these observations, the value of P2 has been increased for the first years, but afterwards it has fallen. However, there are only slight differences of the values between some years, which also indicates a stable productivity and impact on the scientific community. The index values of R1 and R2 decrease over time and are only in few time periods stable (e.g., compare the index value of R1 between 2011 and 2014).

It seems to be promising to determine the changes of the mf-index values over time in order to draw conclusions about different career developments of scientists like varying productivity and number of citations.

Analysis of the Total Values

We now analyze the total mf-index values of the chosen scientists by activating all factors (except of the correction according to the field of study due to the low rate of papers assigned to field of study entities [27]) over all publications and citations. Table 12 lists the total mf-index values of the considered scientists.

P1 obtains the highest mf-index value (24.05) of all the six considered researchers. N2's mf-index value is the next highest one with 19.45. In order to analyze this difference in the mf-index values, we look at the data about the researchers' career given in Table 4. According to the MAG data, P1 and N2 have both a similar career length (N2: 49 years, P1: 50 years). The number of citations of N2 (3981 citations) is only slightly higher than the one of P1 (3787 citations). The rate of non-colleague citations is high for both researchers (N2: 98%, P1: 93%). However, P1's citations are newer than those of N2: The average of the citation age is 9 years and the median even 6 years for P1; the average is 11 years and the median 10 years for N2.

The publication ages of both scientists are similar. However, P1 (251 papers) published many more contributions than N2 (54 papers). Furthermore, P1's papers are authored by less scientists: The average of the number of co-authors is 3 (and its median 2) for N2, but only 1 for P1 (0 respectively). Following this argumentation, a higher index value of P1 (having newer citations, less co-authors and higher productivity) in comparison to N2 is explainable and justifiable. N2 obtained the second highest mf-index value, which values N2's lifetime performance in a reasonable way.

P2's mf-index value (12.63) is the third highest being only slightly above the one of N1 (11.07). N1's career length is already 52 years and hence doubles the one of P2 having a career length of 23 years. However, P2 published already 23 papers, which is about three times more contributions than N1 with 8 publications. Moreover, the publication age of P2 (in average 12 years, which is also the median) is lower than the one of N1 (in average 31 years with a median of 25 years). As already discussed in the analysis for the publication age, average and median values in combination with small sample sizes (like in N1's case of having only 8 publications) must be handled with care.

The differences of the number of co-authors is marginal: The average number of co-authors is 1 for both scientists, but N1's median is even 0 in comparison to the median 1 of P2. The number of citations differ much for both scientists: N1 received 1215 citations (99% of non-colleagues) and P2 586 (91% of non-colleagues). The citations of N1 (with an average of 9 years and a median of 5 years) are younger than those of P2 (average and median: 12 years). However, N1 received 99% of her/his citations with only 3 papers, but 99% of P2's citations are distributed among 13 publications.

The mf-index basically uses the tapered h-index (see Equation 10b), which weights the citations of highly cited publications less when single publications have more citations. This avoids extremely high index values of scientists with only few successful papers. Moreover, if we divide the number of citations by the career length, we notice that P2 received 25 citations per year in comparison to 23 citations per year of N1. Considering additionally the well distribution of citations to publications of P2, it seems to be reasonable that P2 has a higher index value than N1. On the other hand, N1's data covers some positive aspects like higher total number of citations, higher rate of non-colleague citations and lower citation age leading to only a slightly lower mf-index value of N1 in comparison to P2.

However, the index value still seems to be too low for a Nobel laureate. The main reason for the low mf-index value is missing data (and especially citations) in MAG. For example, according to Google Scholar, one of N1's papers already received 5172 citations, but MAG includes only 548 citations, such that about 90% of the citations of Google Scholar are missing in MAG. MAG is the most comprehensive publicly available data set about publications and especially their citations [27], but this example shows that there are still many drawbacks when using MAG. In other words: The determined index value can be preciser only when the underlying data set is high-quality in order to obtain reasonable results for comparisons of researchers.

According to Table 12, there is a certain gap between N1's index value and the ones of the research assistants R1 (of 6.68) and R2 (of 3.84). However, these gaps should be justifiable due to the fact that N1 is a Nobel laureate. When comparing the mf-index values of the research assistants, we can also detect a gap of 2.84 between both scientists. The career lengths of R1 (of 14 years) and R2 (of 13 years) are nearly the same, although the number of R1's publications (45 contributions) nearly doubles the one of R2 (24 contributions). Furthermore, R1 published with less co-authors (average: 3, median: 2) than R2 (average: 5, median: 4). However, R2's publications are slightly younger (average and median: 7 years) than the ones of R1 (average and median: 8 years). Moreover, R2 received more citations (182) than R1 (117), and R2's citation age is lower (average and median: 6 years) in comparison to R1 (average: 7 years, median: 8 years).

We notice the biggest difference between both scientists when looking at the kind of citations: R1 cited her-/himself in 53% of R1's citations, received 6% colleague citations, and 41% non-colleague citations. On the contrary, R2's citations are mainly non-colleague citations (95%). In total, the difference in the mf-index values of R1 and R2 is reasonable and justifiable because of differences in their data like more and younger citations of R2 mainly from non-colleagues.

Overall, we draw the conclusion that the mf-index evaluates the performances of scientists in a fair way leading to reasonable, justifiable and comparable index values of researchers.

5.3.3 Comparison of all Indices

The previous sections deal with the mf-index and especially its single factors as well as the existing indices. In this section, we compare our mf-index with the existing indices. We examine for which scientists the index values of the different indices differ more or have similar values and indicate possible reasons. Finally, we discuss whether or not the theoretical advantages described in Section 4.4 can be verified by our experimental evaluation based on the data of the six chosen researchers. Table 13 contains the index values of the existing indices as well as of our mf-index (activating all factors except for the correction for the field of study).

According to Table 13, the index value of the h-index

Table 13: Values of all indices

	N1	N2	P1	P2	R1	R2
h	5	18	18	10	6	4
h_T	12.19	32.6	35.04	18.52	9.73	8.6
mf	11.07	19.45	24.05	12.63	4.39	7.23

is usually lower in comparison to the tapered h-index and the mf-index for all chosen scientists. The reason for the lower index values is that the h-index is upper bounded by the number of publications of a researcher. For example, the h-index of N1 can be at most 8, which is N1's number of publications. Furthermore, the tapered h-index values are higher than those of the mf-index. This can be explained by the fact that the mf-index considers much data of the scientist and on this basis performs some divisions during the index calculation, which may lead to a slightly or significantly lower index value.

However, this property of the mf-index is welcome, especially because the consideration of temporal factors helps to reflect dynamical changes in the researcher's performance in contrast to the other existing indices. Basically, the calculation of the h-index and of the tapered h-index is based on the publication and citation counts. Thereby the h-index and the tapered h-index do not consider important aspects like the career length, the publication and citation age, the citation source (the researcher on her/his own, colleagues or non-colleagues) and so on.

We now compare differences between the index values of the mf-index and the existing indices of the different scientists. We focus on the comparison between the tapered h-index and our mf-index, since the mf-index is based on the tapered h-index and we modified the tapered h-index by considering additional aspects.

According to Table 13, the mf-index values are lower for N1, N2, P1 and P2 in comparison to the ones of the tapered h-index. The lower values are mainly caused by the greater career length and the higher publication and citation ages. We already discussed in the previous section the influence of each factor on the mf-index rank calculations. Table 13 shows that the difference of the index values of N2 and P1 is 2.44 for the tapered h-index, but for the mf-index already 4.6. Reasons are N2's higher average citation age and number of co-authors in comparison to P1.

The index values of the research assistants are also lowered: In contrast to R1, the decrease for R2 is only small, because R2's ages of publications and citations are still moderate. Although some data of R1 is comparable to the data of R2, R1's mf-index value is much decreased in comparison to the tapered h-index. As a matter of fact, R1's high self-citation rate of 53% is crucial. The other scientists (inclusive R2) have at least 91% non-colleague citations, whereas only 41% of R1's citations are from non-colleagues. Thereby R1's mf-index value is even below the h-index, which does not occur for the other scientists. However, we welcome this phenomenon, because such a high self-citation rate should have an immense effect on the index value. Existing indices do not consider this aspect, such that the mf-index offers an appropriate ranking of scientists.

Summarizing the previous analyses of the existing indices in Section 5.3.1, of the mf-index in Section 5.3.2 and of the direct comparison, we recognize several advantages of the mf-index in this experimental evaluation based on data of the six chosen scientists.

The mf-index values are more appropriate than those of the in many cases too low h-index and are only in justified exceptions (like the 53% self-citation rate of R1) lower than the h-index. Furthermore, the mf-index offers a fair comparison between older and younger researchers by considering temporal dependent factors and moderately lowering the index values of older researchers. The mf-index considers other factors like the co-author number, which enriches the analysis of researcher comparisons with additional aspects (like in the comparison between N2 and P1). The mf-index has the potential advantage of the correction according to the field of study, which could not be evaluated here in detail, because the MAG data has a too low rate of papers assigned to field of study entities [27].

We suppose the following effects of activating the correction according to the field of study: The scientists N1 and P1 are physicists, such that their index values remain the same after a correction according to the field of study, since physics is the reference field for the correction (see Section 4.2.7). N2 is active in the field of chemistry and biology. Hence, N2's index value will be lowered according to Table 2 balancing the higher citation rates in these fields. P2's index value will be increased, because P2's field is mathematics receiving typically less citations. Similar remarks apply to R1 and R2, both of which are computer scientists taking an advantage of the correction based on the field of study.

In total, the experimental evaluation based on data of real scientists verifies numerous advantages of the mf-index in comparison to existing indices.

6 Summary and Outlook

In this work, we propose a new bibliometric indicator called mf-index, which combines multiple factors for evaluating the overall performance of researchers as objective and precise as possible.

The new contribution of this index is especially the consideration of multiple factors like career length, publication and citation age, citation weights for different kinds of citations, field of study and number of co-authors. We show the practicability of the new mf-index by a detailed analysis of each of its factors and their impact on the resulting index value by comparing the data of real-world scientists. Furthermore, we compare the mf-index values with those of other existing bibliometric indicators and verify that the mf-index much better balances different aspects of researchers and results in a fairer comparison of their performances.

Our evaluation and analysis show that our mf-index has many advantages, but there is still room for future improvements, which may however result in higher computational costs. As a matter of fact, there are only three weights of the citation authors for self-citations, colleague and non-colleague citations. However, we can also search for a shortest path from the cited scientist over the researchers with joint publications and their colleagues to the citing scientist. In this way, we can determine a weight between 0% and 100% reflecting the collaboration distance between the cited and the citing researcher by a preciser value.

We can also improve the consideration of the co-authors of a scientist. The mf-index divides the received citations of a publication by the number of its authors. In this way, the rate of contribution of a researcher to the considered publication is neglected. We can also weight authors with a higher contribution more than other authors, although we then need to find a method to detect the rate of contributions, as conventions like the main contributor being the first author or listing the authors in alphabetical order differ from working group to working group. Furthermore, the division by the number of authors disregards that many co-authors are also often a positive sign, as it indicates a high reputation, such that many colleagues want to cooperate for authoring joint papers, and/or a high sociability. A modification of considering the co-authors can be based on the $\bar{h}$-index, which considers a publication of a researcher only in the case that this publication also belongs to the h-core of the co-authors.

If a publication is not in the h-core of the co-authors (but belongs to the h-core of the considered researcher), then this is a sign for different scientific levels of the considered researcher and the co-authors. Possibly, the publication would have received less citations if the well-known co-authors did not author the considered publication. In such cases, we can still include the publication in the index calculation of the considered scientist, but assign a smaller weight to this paper with the intention to allow fairer comparisons of researchers at different scientific levels.

We already incorporated a strong abstraction of the factors and their calculations by introducing weighting functions. In this way, our generalized mf-index can serve for analyses of different purposes and for different domains by just changing the weighting functions, which also reduces the implementation costs for index calculation routines. For example, instead of a division by the number of publication authors a logarithmic function can be used in order to encourage researchers to work together with other scientists while not overweighting masses of publication authors. In this way, one set of weighting functions may be used to detect scientists with a high sociability while other weighting functions could be used to detect researchers with an impact to the scientific community, which is similar to the one of Nobel laureates. Also other factors not considered so far can be easily integrated into the mf-index.

Overall, our proposed mf-index provides a new metric to measure and compare the performances of researchers by well balancing important factors for a fair comparison. The explanatory power for different purposes might be improved in future work by extensively studying the effects of different variants and by modifying the weighting functions of our generalized mf-index.

References

[1] R. Adler, J. Ewing, and P. Taylor, "Citation Statistics: A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)," *Statistical Science*, vol. 24, no. 1, pp. 1–14, 2009.

[2] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "h-index: A Review Focused in its Variants, Computation and Standardization for Different Scientific Fields," *Journal of Informetrics*, vol. 3, no. 4, pp. 273–289, 2009.

[3] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices," *Scientometrics*, vol. 82, no. 2, pp. 391–400, 2010.

[4] T. Anderson, R. Hankin, and P. Killworth, "Beyond the Durfee square: Enhancing the h-index to score total publication output," *Scientometrics*, vol. 76, no. 3, pp. 577–588, 2008.

[5] R. Ball, “Der Hirschfaktor - ein neuer ‘Stern’ am bibliometrischen Indikatorenhimmel,” *BIT-Online*, vol. 9, no. 4, pp. 309–312, 2006.

[6] R. Ball and D. Tunger, *Bibliometrische Analysen - Daten, Fakten und Methoden: Grundwissen Bibliometrie für Wissenschaftler, Wissenschaftsmanager, Forschungseinrichtungen und Hochschulen*, ser. Schriften des Forschungszentrums Jülich, Reihe Bibliothek / Library. Jülich: Forschungszentrum Jülich, 2005, vol. 12.

[7] P. D. Batista, M. G. Campiteli, O. Kinouchi, and A. S. Martinez, “Is it possible to compare researchers with different scientific interests?” *Scientometrics*, vol. 68, no. 1, pp. 179–189, 2006.

[8] M. Biryukov and C. Dong, “Analysis of Computer Science Communities Based on DBLP,” in *Research and Advanced Technology for Digital Libraries, 14th European Conference, ECDL 2010, Glasgow, UK, September 6-10, 2010. Proceedings*, 2010, pp. 228–235.

[9] L. Bornmann and H.-D. Daniel, “The state of h index research,” *EMBO reports*, vol. 10, no. 1, pp. 2–6, 2009.

[10] L. Bornmann, R. Mutz, and H.-D. Daniel, “Are There Better Indices for Evaluation Purposes than the h Index? A Comparison of Nine Different Variants of the h Index Using Data from Biomedicine,” *Journal of the American Society for Information Science and Technology*, vol. 59, no. 5, pp. 830–837, 2008.

[11] H. Bouabid, “Revisiting citation aging: a model for citation distribution and life-cycle prediction,” *Scientometrics*, vol. 88, no. 1, pp. 199–211, 2011.

[12] M. Bras-Amorós, J. Domingo-Ferrer, and V. Torra, “A Bibliometric Index Based on the Collaboration Distance between Cited and Citing Authors,” *Journal of Informetrics*, vol. 5, no. 2, pp. 248–264, 2011.

[13] S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine,” *Computer Networks and ISDN Systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[14] F. J. Cabrerizo, S. Alonso, E. Herrera-Viedma, and F. Herrera, “q2-Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core,” *Journal of Informetrics*, vol. 4, no. 1, pp. 23–28, 2010.

[15] P. Chen, H. Xie, S. Maslov, and S. Redner, “Finding scientific gems with Googles PageRank algorithm,” *Journal of Informetrics*, vol. 1, no. 1, pp. 8–15, 2007.

[16] D. E. Chubin and S. D. Moitra, “Content Analysis of References: Adjunct or Alternative to Citation Counting?” *Social Studies of Science*, vol. 5, no. 4, pp. 423–441, 1975.

[17] R. Costas and M. Bordons, “The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level,” *Journal of Informetrics*, vol. 1, no. 3, pp. 193–203, 2007.

[18] B. Cronin, *The Citation Process: The Role and Significance of Citations in Scientific Communication*. London: Taylor Graham, 1984.

[19] A. Dilger and H. Müller, “Ein Forschungsleistungsranking auf der Grundlage von Google Scholar,” *Zeitschrift für Betriebswirtschaft*, vol. 82, no. 10, pp. 1089–1105, 2012.

[20] S. N. Dorogovtsev and J. F. Mendes, “Ranking scientists,” *Nature Physics*, vol. 11, no. 11, pp. 882–883, 2015.

[21] L. Egghe, “An improvement of the h-index: the g-index,” *ISSI Newsletter*, vol. 2, no. 1, pp. 8–9, 2006.

[22] L. Egghe, “Theory and practise of the g-index,” *Scientometrics*, vol. 69, no. 1, pp. 131–152, 2006.

[23] L. Egghe, “Mathematical Theory of the h- and g-Index in Case of Fractional Counting of Authorship,” *Journal of the American Society for Information Science and Technology*, vol. 59, no. 10, pp. 1608–1616, 2008.

[24] L. Egghe, “The Hirsch Index and Related Impact Measures,” *Annual Review of Information Science and Technology*, vol. 44, no. 1, pp. 65–114, 2010.

[25] L. Egghe and R. Rousseau, “An h-index weighted by citation impact,” *Information Processing & Management*, vol. 44, no. 2, pp. 770–780, 2008.

[26] L. Engqvist and J. G. Frommen, “The h-index and self-citations,” *Trends in Ecology & Evolution*, vol. 23, no. 5, pp. 250–252, 2008.

[27] D. Herrmannova and P. Knoth, “An Analysis of the Microsoft Academic Graph,” *D-Lib Magazine*, vol. 22, no. 9/10, 2016.

[28] J. E. Hirsch, “An index to quantify an individual’s scientific research output,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 46, pp. 16 569–16 572, 2005.

[29] J. E. Hirsch, “An index to quantify an individual’s scientific research output that takes into account the

effect of multiple coauthorship," *Scientometrics*, vol. 85, no. 3, pp. 741–754, 2010.

[30] J. E. Hirsch and G. Buela-Casal, "The meaning of the h-index," *International Journal of Clinical and Health Psychology*, vol. 14, no. 2, pp. 161–164, 2014.

[31] J. E. Iglesias and C. Pecharromán, "Scaling the h-index for different scientific ISI fields," *Scientometrics*, vol. 73, no. 3, pp. 303–320, 2007.

[32] J. Imperial and A. Rodríguez-Navarro, "Usefulness of Hirsch's h-index to evaluate scientific research in Spain," *Scientometrics*, vol. 71, no. 2, pp. 271–282, 2007.

[33] B. Jin, "H-index: an evaluation indicator proposed by scientist," *Science Focus*, vol. 1, no. 1, pp. 8–9, 2006.

[34] B. Jin, L. Liang, R. Rousseau, and L. Egghe, "The R- and AR-indices: Complementing the h-index," *Chinese Science Bulletin*, vol. 52, no. 6, pp. 855–863, 2007.

[35] M. Jokić and R. Ball, *Qualität und Quantität wissenschaftlicher Veröffentlichungen: Bibliometrische Aspekte der Wissenschaftskommunikation*, ser. Schriften des Forschungszentrums Jülich, Reihe Bibliothek / Library. Jülich: Forschungszentrum Jülich, 2006, vol. 15.

[36] J. Kaube, "Wissenschaftsevaluation: Die bibliometrische Verblendung," July 24, 2008. [Online]. Available: http://www.faz.net/-gqz-zt2d/

[37] C. D. Kelly and M. D. Jennions, "The h index and career assessment by numbers," *Trends in Ecology & Evolution*, vol. 21, no. 4, pp. 167–170, 2006.

[38] M. Kosmulski, "A new hirsch-type index saves time and works equally well as the original h-index," *ISSI Newsletter*, vol. 2, no. 3, pp. 4–6, 2006.

[39] M. Kudělka, J. Platoš, and P. Krömer, "Author Evaluation Based on H-index and Citation Response," *CoRR*, vol. abs/1511.05709, 2015.

[40] A. Kusserow and S. Groppe, "Getting Indexed by Bibliographic Databases in the Area of Computer Science," *Open Journal of Web Technologies (OJWT)*, vol. 1, no. 2, pp. 10–27, 2014. [Online]. Available: https://www.ronpub.com/ojwt/OJWT_2014v1i2n02_Kusserow.html

[41] B. R. Martin, "The use of multiple indicators in the assessment of basic research," *Scientometrics*, vol. 36, no. 3, pp. 343–362, 1996.

[42] F. Mattern, "Zur Evaluation der Informatik mittels bibliometrischer Analyse," *Informatik-Spektrum*, vol. 25, no. 1, pp. 22–32, 2002.

[43] M. Medo and G. Cimini, "Model-based evaluation of scientific impact indicators," *Physical Review E*, vol. 94, no. 3, pp. 1–12, 2016.

[44] M. J. Moravcsik and P. Murugesan, "Some Results on the Function and Quality of Citations," *Social Studies of Science*, vol. 5, no. 1, pp. 86–92, 1975.

[45] B. A. Nosek, J. Graham, N. M. Lindner, S. Kesebir, C. B. Hawkins, C. Hahn, K. Schmidt, M. Motyl, J. Joy-Gaba, R. Frazier, and E. R. Tenney, "Cumulative and Career-Stage Citation Impact of Social-Personality Psychology Programs and Their Members," *Personality and Social Psychology Bulletin*, vol. 36, no. 10, pp. 1283–1300, 2010.

[46] C. Oppenheim and S. P. Renn, "Highly Cited Old Papers and the Reasons Why They Continue to be Cited," *Journal of the American Society for Information Science*, vol. 29, no. 5, pp. 225–231, 1978.

[47] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," Stanford InfoLab, Tech. Rep., 1999.

[48] J. Panaretos and C. Malesios, "Assessing scientific research performance and impact with single indices," *Scientometrics*, vol. 81, no. 3, pp. 635–670, 2009.

[49] B. C. Peritz, "A classification of citation roles for the social sciences and related fields," *Scientometrics*, vol. 5, no. 5, pp. 303–312, 1983.

[50] I. Podlubny and K. Kassayova, "Towards a better list of citation superstars: compiling a multidisciplinary list of highly cited researchers," *Research Evaluation*, vol. 15, no. 3, pp. 154–162, 2006.

[51] A. Purvis, "The h index: playing the numbers game," *Trends in Ecology & Evolution*, vol. 21, no. 8, p. 422, 2006.

[52] F. Radicchi, S. Fortunato, and C. Castellano, "Universality of citation distributions: toward an objective measure of scientific impact," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 45, pp. 17 268–17 272, 2008.

[53] F. Radicchi, S. Fortunato, B. Markines, and A. Vespignani, "Diffusion of scientific credits and the ranking of scientists," *Physical Review E*, vol. 80, no. 5, pp. 1–10, 2009.

[54] R. Rousseau and F. Y. Ye, "A Proposal for a Dynamic h-Type Index," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 11, pp. 1853–1855, 2008.

[55] F. Ruane and R. S. J. Tol, "Rational (successive) h-indices: An application to economics in the Republic of Ireland," *Scientometrics*, vol. 75, no. 2, pp. 395–405, 2008.

[56] J. Ruscio, F. Seaman, C. D'Oriano, E. Stremlo, and K. Mahalchik, "Measuring Scholarly Impact Using Modern Citation-Based Indices," *Measurement: Interdisciplinary Research and Perspectives*, vol. 10, no. 3, pp. 123–146, 2012.

[57] M. Schreiber, "Self-citation corrections for the Hirsch index," *EPL (Europhysics Letters)*, vol. 78, no. 3, pp. 1–6, 2007.

[58] M. Schreiber, "The influence of self-citation corrections on Egghe's g index," *Scientometrics*, vol. 76, no. 1, pp. 187–200, 2008.

[59] M. Schreiber, "To share the fame in a fair way, hm modifies h for multi-authored manuscripts," *New Journal of Physics*, vol. 10, no. 4, pp. 1–9, 2008.

[60] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos, "Generalized Hirsch h-index for disclosing latent facts in citation networks," *Scientometrics*, vol. 72, no. 2, pp. 253–280, 2007.

[61] S. Teufel, A. Siddharthan, and D. Tidhar, "An annotation scheme for citation function," in *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, ser. SIGdial '06. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 80–87.

[62] R. Todeschini, "The j-index: a new bibliometric index and multivariate comparisons between other common indices," *Scientometrics*, vol. 87, no. 3, pp. 621–639, 2011.

[63] R. Todeschini and A. Baccini, *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research*. Weinheim: Wiley-VCH, 2016.

[64] M.-Y. Tsay, "Library journal use and citation age in medical science," *Journal of Documentation*, vol. 55, no. 5, pp. 543–555, 1999.

[65] N. J. van Eck and L. Waltman, "Generalizing the h- and g-indices," *Journal of Informetrics*, vol. 2, no. 4, pp. 263–271, 2008.

[66] J. K. Vanclay, "On the robustness of the h-index," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 10, pp. 1547–1550, 2007.

[67] P. Vinkler, "Eminence of scientists in the light of the h-index and other scientometric indicators," *Journal of Information Science*, vol. 33, no. 4, pp. 481–491, 2007.

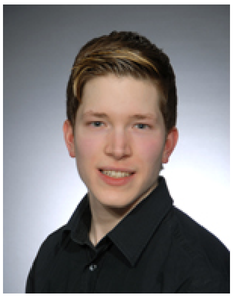
[68] L. Waltman, "A review of the literature on citation impact indicators," *Journal of Informetrics*, vol. 10, no. 2, pp. 365–391, 2016.

[69] M. Weinstock, "Citation Indexes," in *Encyclopedia of Library and Information Science*, A. Kent and H. Lancour, Eds. New York: Marcel Dekker, Inc., 1971, vol. 5, pp. 16–40.

[70] M. C. Wendl, "H-index: however ranked, citations need context," *Nature*, vol. 449, no. 7161, p. 403, 2007.

[71] L. Wildgaard, J. W. Schneider, and B. Larsen, "A review of the characteristics of 108 author-level bibliometric indicators," *Scientometrics*, vol. 101, no. 1, pp. 125–158, 2014.

AUTHOR BIOGRAPHIES



Eric Oberesch was born in Munich, Germany, in 1991. He earned his bachelor's degree in Informatik (Computer Science) in 2016 from the University of Lübeck. He wrote his bachelor thesis about citation-based indices which allow to compare the output of different scientists. An essential part of the thesis was the development of a new, better index and its implementation in an existing system for managing scientific publications.

Sven Groppe earned his diploma degree in Informatik (Computer Science) in 2002 and his Doctor degree in 2005 from the University of Paderborn. He earned his habilitation degree in 2011 from the University of Lübeck. He worked in the European projects B2B-ECOM, MEMPHIS, ASG and TripCom. He was a member of the DAWG W3C Working Group, which developed SPARQL. He was the project leader of the DFG project LUPOSDATE, an open-source Semantic Web database, and one of the project leaders of two research projects, which research on FPGA acceleration of relational and Semantic Web databases. He is also the chair of the Semantic Big Data workshop series, which is affiliated with the ACM SIGMOD conference (so far in 2016 and 2017), and of the Very Large Internet of Things workshop in conjunction with the VLDB conference in 2017. His research interests include databases, Semantic Web, query and rule processing and optimization, Cloud Computing, peer-to-peer (P2P) networks, Internet of Things, data visualization and visual query languages.