The Scientific Influence Passivity Score (SIP): Bibliometrics learning from Twitter

Bernd Markscheffel
Dept. of Business Informatics
Technische Universität
Ilmenau
+493677694053
bernd.markscheffel@tu-ilmenau.de

Johannes Schmidt
Dept. of Business Informatics
Technische Universität
Ilmenau
+493677694053
johannes.schmidt@tu-ilmenau.de

ABSTRACT

Bibliometrics as the application of mathematical and statistical methods to books and other media of communication research [1] has already been established since a couple of years as a basis for funding decisions, and as an accepted instrument for evaluation of persons and institutions or recognizing trends and tendencies in science and technology. A wide range of methods and indicators are used to achieve this goal. In this paper, we present a new bibliometric indicator – the SIP Score – which is an adoption of the Influence Passivity Score (invented for the microblogging platform Twitter). We explain the adoption process, show the first results on a special created bibliographical dataset and discuss comparison approaches with 'classical' indicators (h-index and its derivate).

Categories and Subject Descriptors

A.m MISCELLANEOUS **General Terms**
Performance, Economics, Verification.

Keywords

Bibliometrics, Twitter Ecosystem, Impact Indicator, Time Related Aspects, Scientific Influence-Passivity-Score (SIP-Score)

1. INTRODUCTION

Developments in research administration and in technology have pushed Bibliometrics on the center stage in research assessment [2] Therefor a lot of methods, metrics and indicators were developed. The range reaches thereby from simple, but expressive indicators like the h-index [3] to complex models and algorithms like the Yule-process [4, 5]. Especially in the context of New- or Alt-metrics [6, 7, 8, 9] some concerns about citation based metrics have been expressed e.g.:

- the lack of a semantically interpretation of the connection between two documents because citation based metrics ignore context and reasons for citations [9];
- the lack of processing and recognition speed, citation metrics are slower, compared to Web2.0 driven reflections, because it takes a long time to accumulate. Mostly, scholarly literature first lead to a measurable mention within the Web 2.0 ecosystem before, later, it is also reflected in increased citations and/or policy changes and social impact [10];
- the lack of transparency of citation based metrics, because they are not independent from the underlying dataset e.g. the h-index of a researcher varies widely depending on the database used to calculate it [10, 11]. E.g the journal impact factor (JIF), has often been criticized for not being transparent and significant gaming is relatively easy [9, 12, 13];
- the lack of a detailed or inconsequent consideration of time as a factor for influence and impact of authors [14].

Especially the last point marks a very difficult detail in recognition of scientific performance. Many indicators (e.g. h-index) did not distinguish time related aspects, they do not differentiate between creative period or citation period or if and how the acknowledgement phase is delayed et cetera.

In this paper we introduce a new indicator – the SIP-Score (Scientific Influence Passivity-Score) which will include these problems. The SIP-Score is an adoption of the existing IP-Score (Influence Passivity-Score) [15] that was invented for the microblogging service Twitter.

After discussing related work, we will introduce the IP-Score. In the following section we will describe the bibliometric database that was created for this special purpose on basis of Google Scholar. Then we will describe the adoption process, its results and we will summarize and discuss the further developments and future research challenges.

2. RELATED WORK

We performed a literature study to address the time related aspects of citation indicators and its solutions.

MEDES'14, September 15-17, 2014, Buraidah Al Qassim, Saudi Arabia.

http://dx.doi.org/10.1145/2668260.2668310

After reviewing the essential papers we went backward by screening the citations for these articles to determine a core set which was analyzed to answer the following questions: *Which papers addressing time related aspects?* and *Are there solutions for the time related problems?*

2.1 h-index

The h-index, introduced by Hirsch [3] measures the productivity and the impact of the published work of a researcher. It is based on the set of the researcher's most cited papers and the number of citations that they have received in other publications. Hirsch described the index as follows:

“A scientist has index h if h of his/her Np papers have at least h citations each, and the other (Np − h) papers have no more than h citations each” [3].

Beside from the well-known criticism of the h-index [16, 17] like:

- it does not take into account the number of authors of a paper,
- it does not take into account the citation context,
- it does not take into account the source of the citation (book or article),
- it is easy to manipulate through self citation [18]

for our work is another main point of criticism essentially: the time independent interpretation of h-index values.

As one can see in Figure 1 both authors (red and blue) have the same h-index (5) over a period of 30 years but the productivity is totally different. So, it is necessary to make a differentiation between different publishing and citing periods to make statements about the impact of an author in a distinct time range.

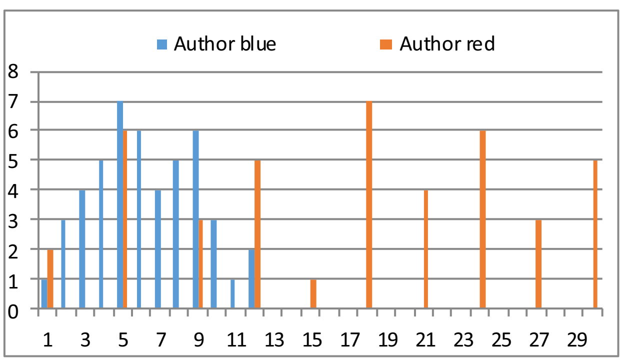

Figure 1. h-Index sample

2.2 m-index

Main aim of the m-index was to compare research output of scientists of different seniority [3]. The m-index is the relation of the h-index divided by the number of years the academic has been active (measured as the number of years since the first published paper). The impact statements for the sample in Figure 1 about author blue and red changes to:
Author *blue* has a **m-index of 0.416** (h-index=5 / 12) and Author *red* has a **m-index of 0,166** (h-index=5 / 30). The impact of author blue is higher than the impact of author red. Therefore, the m-index provides a more realistic assessment of the academic achievement of scientists [19].

2.3 hc-index

The contemporary h-index (hc-Index) adds an age-related weighting to each cited article. It evaluates the articles with the time difference between publishing year and actual year.
The definition of the hc-index is:
“..a researcher has contemporary h-Index hc, if hc of its N articles get a score of $S_i^c \geq hc$ each, and the rest ($N - hc$) articles get a score of $S_i^c \leq hc$” [20],

where S_i^c, the so called Novel Score, is calculated for each article i according to $S_i^c = \frac{\gamma}{(Y_{now} - Y_i + 1)^{\delta}} * |C_i|$ with Y_i as the publication year of an article i and C_i are the articles citing the article i.

2.4 Influence Passivity Score - IP-Score [15]

The IP-Score in original is a Social Media Score, developed for the digital ecosystem Twitter. It was developed to analyze the influence and passivity of users based on their information forwarding activity. The outstanding feature of the IP-Score is to take into account not only the influence of an author. Influence in context of Twitter is the impact to other user to force a reaction in terms of retweeting a message. Moreover, the IP-Score takes into account the passivity, as a measure of how difficult it is for other users to influence him. In general, the model makes four assumptions:

1. "A user's influence score depends on the number of people she influences as well as their passivity” [15] p.3. In Figure 2 Fritz is the user for which the impact should be calculated and N1 and N2 and Paul are the user who where influenced by Fritz, and respectively their passivity has to overcome by Fritz**.**
2. “A user's influence score depends on how dedicated the people she influences are. Dedication is measured by the amount of attention a user pays to a given one as compared to everyone else” [15] p.3.. In Figure 2 is exemplary illustrated for user Paul, who is influenced by Fritz, which user are affecting Paul (N3, N4, N5) and therefore also influencing the ‘quality’ of the influence from Paul by Fritz. The ‘quality’ of the impact is a measure how strong the influence (of Fritz) is compared to all other influences (Paul, N3, N4, N5 …).

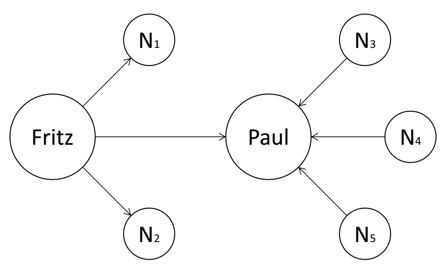

Figure 2: Illustration of Assumption 1 & 2

3. “A user's passivity score depends on the influence of those who she's exposed to but not influenced by” [15] p.3 As we see in Figure 3, Fritz is the user for which the passivity score should be calculated, and his passivity score depends on the

user (N1, N2, Paul) who are trying to affect Fritz unsuccessfully.

4. A user's passivity score depends on how much she rejects other user's influence compared to everyone else" [15] p.3. As we can see, not the individual value of the passivity of Fritz is calculated, moreover it is the ratio of the rejection of the impact of Paul by Fritz compared to the overall (Fritz, N3, N4, N5) rejection value. If a user rejects the impact of a highly influential user then his passivity is higher as the passivity of a user who rejects the impact of less influential user.

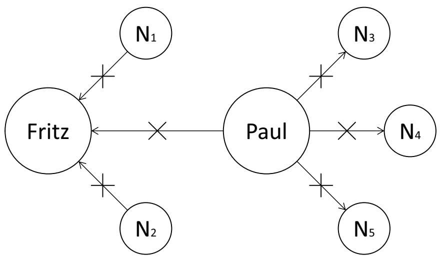

Figure 3: Illustration of Assumption 3 & 4

As we can see from assumption 1-4 influence and passivity are strongly related to each other. The user and the relationships to other user building a graph structure with nodes (n) and arcs (e) and arc weights (w). The weights w_{ij} on arc e_{ij} represent the ration of influence that i exerts on j to the total influence that i exerts on j. The acceptance rate e_{ij} is a value that represents the user j accepted from user i normalized by the total influence accepted by j from all users in the network: $u_{ij} = \frac{w_{ij}}{\sum_{k:(k,j)\in E} w_{kj}}$ as a value for the dedication or loyalty user j has to user i.

The rejection rate is defined by $v_{ji} = \frac{1-w_{ji}}{\sum_{k:(j,k)\in E}(1-w_{jk})}$. Since the value $1 - w_{ji}$ is the amount of influence that user i rejected from j, then the value v_{ji} represents the influence that user i rejected from user j normalized by the total influence rejected from j by all users in the network [15].

The final values for influence I_i and passivity P_i for a specific user I can calculated by the following formulas:

$$I_i = \sum_{j:(i,j)\in E} u_{ij} P_j \qquad P_i = \sum_{j:(j,i)\in E} v_{ji} I_j$$

The final algorithm consist of an iterative calculation of I_i and P_i. Romero shows that the variance of the I_i and P_i values converge to zero after approximately 50 iterations [15].

3. DATA SET

In order to compare the new score with already existing indices (h-index, m-index, hc-index) there was a need of a bibliometric database. The main aim was to create a playgroundset of bibliometric information with the following requirements:

- real data to ensure objectivity,
- enough data to allow universal results,
- not too many data to keep interpretability,
- a publication time span big enough to allow the analysis of different time frames.

This task was solved by generating a dataset based on Google Scholar. The set was started with the author Henk F. Moed and all of his listed publications. After that, the first 98 authors, who cited his paper "Citation analysis in research evaluation", as well as their publications, were subsequently added to the database. Finally, all citations between these 99 authors and their articles in the database were saved.

In the final version, the database consisted of 99 authors, 15,374 articles and 38.977 citations.

4. SCIENTIFIC IP-SCORE (SIP-SCORE)

The SIP-Score is a direct adaption of the IP-Score. Main aim of this score was to adopt the advantages of the original score of the Twitter-ecosystem to bibliometrics and try to solve the explained problems. Especially the problem of not considered time effects in bibliometric indices and models.

To achieve this, a couple of assumptions had to be made:

- Analogy between the different authors: users on Twitter are as authors as scientific writers.
- Analogy between the different publications: Tweets on Twitter and scientific books and papers can be treated the same way. The enormous difference in size between the publications (tweets and scientific publications) is relative to their creation time and time of presence in the observed area. While tweets are being written within some seconds and only last for a couple of hours, scientific publications are being written in several days/months/years but also present for a longer time.
- The follower problem: This was a crucial part of the adaption. The IP-Score is based on the number of followers of a user and the ability to decide who read (and maybe answered/retweeted) the post, and who does not. In bibliometrics, this is not possible since there are not follower lists or something alike. In this case, a special assumption was made: As soon as an author cited another author he is being seen as a follower of the cited author. So, whenever the second author releases a book, we assume the first author (who cited the author once) will know about the book.

But, there are obvious downsides, which need to get explained.

One is the fact that the citing author might have stopped his scientific career while the authors he cited are still active. This would end in a rising passivity score of the author, because he's not citing them anymore, which would affect his score unreasonably negative.

The solution of this was only respecting the cited authors' publications as long as they are publicized within the citer's own publication time.

With these assumptions made it was possible to adopt the IP-Score to the Bibliometrics area according to the - in the last section explained - algorithms but with adopted interpretations e.g. that a connection e_{ij} between author i and author j exists only if author j has cited author i at least on time.

5. RESULTS

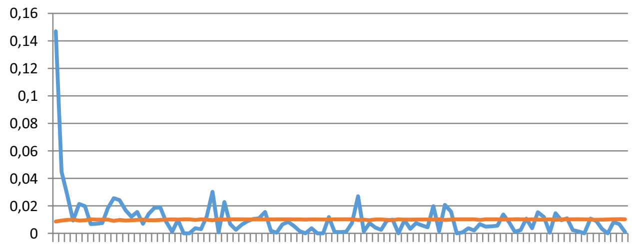

Figure 4. Influence (blue) and Passivity (red) score of the sample dataset for 1968 - 2014 1

Figure 4 shows the influence- and passivity values of the sample dataset for all 99 authors after 50 iteration. We can see that author 1 has an extraordinary high influence score compared to all other authors. The author is Moed and its paper was the starting point for the generation of the dataset sample because it was cited from every other author. So it is reasonable that author1=Moed has the highest influence score of the sample data set. The second clear finding is that the absolute values of the passivity score are relatively constant compared to the absolute values of the influence score. This is due to the regularly evolved structure of the citations in the sample compared to the randomness retweed nature in the Twitter ecosystem.

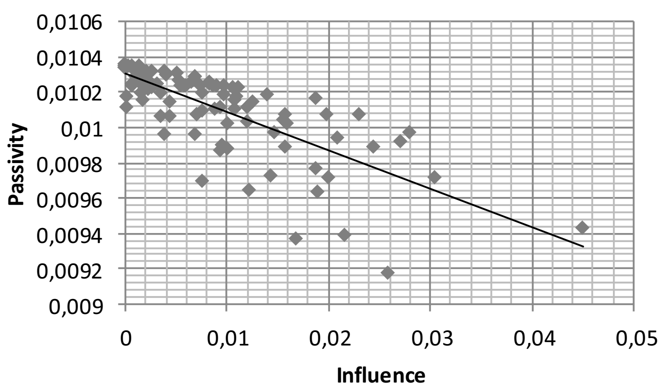

Figure 5: Influence and Passivity relation of the sample dataset after 50 iterations

Figure 5 indicates the trend of an increasing influence by decreasing passivity. Successful scientific work is based on the involvement and perception of other authors work. The more one recognizes the more it increases his influence.

If we compare the results of the SIP-Score with the h-; hc- and m-index we indicate no significant correlation between these measures. The correlation coefficient has a value of 0.4 (h-index); 0.38 (m-index) and 0.29 (hc-index). But this finding is not a negative one, moreover the no-correlation describes and emphasizes the distinction to the other indicators.

One can see an example for the significance and meaningfulness of the SIP-Score regarding authors self-citations. Table 1 shows a cut-out (the first five and the last five) of the table where the authors sorted according to its SIP-influence rank and compared with the other indicators.

Table 1: First five and last five authors according to its SIP-Influence Rank

Author	SIP-Influence Score	SIP-Influence Rank	h-Index	m-Index	hc-Index
1	0,1470	1	18	0,60	10
2	0,0449	2	21	0,50	12
28	0,0304	3	12	0,44	7
3	0,0279	4	8	0,50	5
53	0,0270	5	13	0,34	6
70	0,0001	95	8	0,22	4
23	0,0000	96	3	0,14	2
24	0,0000	97	8	0,31	4
92	0,0000	98	2	0,25	2
96	0,0000	99	1	0,08	1

Particular attention should be paid to authors 23, 24, 92, 96. These authors have no citation by other authors in our sample bibliography. That is why the SIP-influence score is zero, but the values of the other indices are unequal zero because they are considering self-citations. Therefore, we can see the SIP-Score as an indicator, which is elaborating and recognizing the self-citation behaviour of authors more than the 'classical' indicators. The influence of an author is solely determined by his impact to other authors.

Time related Aspects

We have separated four timeframes as proposals, which are representing different productive periods of one's work. For each timeframe we have calculated the SIP-influence and -passivity score, authors rank and the values of the reference indicators. E.g. Table 2 shows the first ten leading authors according to its h-index in timeframe IV.

Table 2: First ten leading authors according to its h-index in timeframe IV

Author	SIP-Influence Rank	h-Index Rank	m-Index Rank	hc-Index Rank	Self citation rate
11	4	1	1	1	0,57
5	5	2	2	2	0,36
6	3	3	3	3	0,20
13	10	4	5	6	0,65
19	7	5	6	4	0,29
2	2	6	7	5	0,14
18	6	7	4	6	0,20
4	13	7	7	12	0,63
51	20	7	7	19	0,97

The main results not only dealing with the interpretation of the results regarding various timeframes are:

a. We have found different results for the SIP-score for various timeframes. The interpretation must be done with actual and complete data set. Our sample set can only help to express basic features of the SIP-score.

b. But it is clear visible that both the SIP-influence- and the SIP-passivity score do not correlate with the other reference indicators. As we can see from Table 2 the SIP-influence score shows an influence value which is completely different to the reputation values of the reference indicators (for example author11: h-index=26; m-index=2,89). That means, there must be other influence signals, which are affecting the influence or reputation of an author measured by the reference indicators (h-index, m-index…). In other words, the effective influence is not as high as faked by the other indicators. One reason for that is the self-citation rate (which is e.g. for author11 57%) and which is not affecting the value of the SIP-influence score but the reference indicators.

c. We can found a clear correlation between decreasing self-citation rate and the SIP-influence score. As considered above the influence of an author is solely calculated by the influence to others authors.

d. This can be seen as a decoupling between productivity features = quantitative features (like number of publications) and influence signals = quality signals. With the help of the SIP-score an author has little role to affect his influence with a large number of publication.

6. SUMMARY AND FUTURE WORK

The adaption of a social-media-ecosystem-indicator for "Big Science" evaluation was a successful one. We introduced with the SIP-score a new indicator for bibliometric purposes, which might have verifiable advantages compared to conventional indicators. It provides advantages, not only in general terms, but especially in treating time related aspects. Moreover, it is a more objective way to calculate users influence without distortion caused by self-citation and productivity features. In addition to that, we have detected some capabilities to improve the SIP-score. We have discussed briefly two specific ways of perfecting the SIP score. The time corrected SIP-score as an approach, which takes into consideration the actual possible perception of the works of other authors. The contemporary SIP-score is a more time-related improvement, which adopts the weighting of the authors within the algorithm towards a higher value for actual high productive authors versus authors, which have their productivity timeframe in the past. Nevertheless, there are some unsolved problems left. First, we must have a closer look to the problem of multi-authorship. In our actual approach, we calculate influence values for each author independently how many authors exist. One possible solution can be a sophisticated weighting according to the number of authors. Due to the iterative nature of the algorithm, which is comparable with the PageRank-algorithm, we should discuss the introduction of a damping factor according to the PageRank [21], to accelerate the calculation of the SIP-score. Finally, we must awake the interest of the major database provider to make a large dataset available for verification and improvement of the SIP-Score.

7. REFERENCES

[1] Pritchard, A. (1969). Statistical bibliography or bibliometrics? In: *Journal of Documentation*, Vol. 24, pp. 348–349.

[2] Glänzel, W. (2010). Bibliomtrics come of age. In: *Research Trends*, Vol. 15, Januar 2010. Retrieved 02.02.2014 from http://www.researchtrends.com/issue15-january-2010/research-trends-7/.

[3] Hirsch, J. G. (2005). *An index to quantify an individual's scientific research output*. San Diego 2005. Retrieved 02.02.2014 from http://arxiv.org/abs/physics/0508025.

[4] Simon, H. A. (1955). On a class of skew distribution functions. In: *Biometrika*, Vol. 42, No. 3, 1955, pp. 425-440.

[5] Chen, Y.-S., Chong, P.P., Tong, M.Y. (1994). The Simon-Yule approach to bibliometric modelling. In: *Information Processing & Management*, Vol. 30, No. 4, July–August 1994, pp. 535–556.

[6] Markscheffel, B. (2013). New Metrics, a Chance for Changing Scientometrics! A Preliminary Discussion of Recent Approaches. In: *Scientometrics – Status and Prospect for Development*. Moscow 10-12. October 2013.

[7] Priem, J., Groth, P., & Taraborelli, D. (2012). The Altmetrics Collection. In: *PLoS ONE*, Vol. 7, No. 11, e48753. doi:10.1371/journal.pone.0048753.

[8] Priem, J., & Hemminger, B. M. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. In: *First Monday*, Vol. 15, No. 7. Retrieved 02.02.2014 from http://www.firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874/2570

[9] Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (n.d.). *altmetrics: a manifesto – altmetrics.org*. Retrieved 12.03.2014, from http://altmetrics.org/manifesto/.

[10] Eysenbach, G. (2011). Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. In: *Journal of Medical Internet Research*, Vol. 13, No. 4, e123. doi:10.2196/jmir.2012.

[11] Bornmann, L., Daniel, H.-D. (2009). The state of h index research.In: *EMBO Reports*, 10(1), pp. 2-6.

[12] Notkins, A. (2008) Neutralizing the impact factor culture. In: *Science* 322: 191.

[13] Petsko GA (2008) Having an impact (factor). In: *Genome Biol* 9: pp 107.

[14] Costas, R., van Leeuwen, T. N., van Raan, A. F. N. (2011). The "Mendel syndrome" in science: durability of scientific literature and its effects on bibliometric analysis of individual scientists. In: *Scientometrics*, Vol. 98, No. 1, 2011, pp. 177-205.

[15] Romero, D. M., Galuba, W., Asur, S., Huberman, B. A. (2010). *Influence and Passivity in Social Media*. Retrieved 02.02.2014 from http://arxiv.org/abs/1008.1253, Palo Alto 2010.

[16] Wendl, M. (2007). H-index: however ranked, citations need context. In: *Nature,* 449 (7161).

[17] Waltman, L., van Eck, N. J. (2012). The inconsistency of the h-index. In: *Journal of the American Society for Information Science and Technology*, Vol. 63, No. 2, 2012, pp. 406-415.

[18] Bartneck, C., Kokkelmans, S. (2011). Detecting h-index manipulation through self-citation analysis. In: *Scientometrics,* Vol. 87, No. 1, pp. 85–98.

[19] Harzing, A.-W. (2008). *Reflections on the h-index*. Retrieved 02.04.2014 from http://www.harzing.com/pop_hindex.htm.

[20] Sidiropoulos A., Katsaros D., Manolopoulos Y. (2007) Generalized Hirsch h-index for disclosing latent facts in citation networks. In: *Scientometrics*, Vol. 72, No. 2, pp. 253-280.

[21] Page, L., Brin, S., Motwani, R. and Winograd, T. (1999). The PageRank citation ranking: Bringing order to the Web. Retrieved 04.05.2014 from http://ilpubs.stanford.edu:8090/422/1/1999-66.p