UNIVERSITY OF COPENHAGEN

Alt-Index: A proposed Index for measuring the social activity of scientific research

Ul Hassan, Saeed; Imran, Mubashir; Zuccala, Alesia Ann

Published in:
Proceedings of the 23rd International Conference on Science and Technology Indicators

Publication date:
2018

Citation for published version (APA):
Ul Hassan, S., Imran, M., & Zuccala, A. A. (2018). Alt-Index: A proposed Index for measuring the social activity of scientific research. In *Proceedings of the 23rd International Conference on Science and Technology Indicators*

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

Alt-Index: A proposed Index for measuring the Social Activity of Scientific Research[1]

Saeed-Ul Hassan*, Mubashir Imran*, and Alesia Zuccala**

*saeed-ul-hassan@itu.edu.pk; mubashir.imran@itu.edu.pk
Scientometrics Lab, Information Technology University, Lahore, 54770, (Pakistan)

***a.zuccala@hum.ku.dk*
Department of Information Studies, University of Copenhagen, Njalsgade 76, 2300 København S, (Denmark)

Introduction

Altmetrics is a broad term used to refer to article level metrics, which focus on a more timely measurement of interest in scholarly documents made visible through social media (Priem et al., 2010). As with any new metric, concerns have been raised regarding its use (Prime, 2013; Kwok, 2013), due to the fact that an Altmetric score might potentially be manipulated or gamed, just as it is possible to game citations (Bartneck & Kokkelmans, 2011; Wilhite et al., 2012). However, given that there are so many diverse measures now compiled within Altmetrics, the tampering process is actually not that easy (Piwowar, 2013). Altmetric scores have therefore attracted the attention of the scientific community as part of their daily practices (Piwowar, 2013). Thus far, a few studies have shown weak to medium correlations between bibliometric measures and Altmetric scores (Costas et al., 2015). More recently, a comprehensive study using data from Altmetic.com and Scopus has shown that when compared to journal citation scores, Altmetric scores demonstrate a higher-level of accuracy for identifying highly cited publications (Hassan et al., 2017).

With this paper, we would like to propose a new measure, termed the alt-index. It is analogous to the h-index (Hirsch 2005), in that it may be defined as follows: "*a scholar has an alt-index of **a**, if **a** of her/his N_p papers have at least **a** social mentions, and the other (N_p-**a**) papers have no more than **a** mentions each*".

Social mentions, in this context, means that the publication is mentioned in at least one social media channel, regardless of the type of online platform. Similar to the h-index, the alt-index can be computed at different levels, such as the level of the researcher, the source journal/conference, the institution, or at the level of a country.

In order to analyze the effectiveness of our proposed index, we focus on a dataset of more than 1.1 million publications received from Altmetric.com, and on measuring the relationship between the h-index and our newly proposed alt-index at the source level of a publication (see for example Braun et al (2006) who have done this for the h-index), utilizing a Spearman's

[1] This work was partially supported by the faculty research & development funds at Information Technology University. The authors are also thankful to Digital Science & Research Solutions Inc. who provided Altmetrics data free of cost for this research.

coefficient of correlation. Our results show a strong positive correlation between the h-index and the alt-index both in general, and across 16 individual research disciplines. We conclude that the alt-index can be deployed to measure the social activity of publications in situations where agreement with social media channels is an important consideration.

Dataset

We used jun-4-2016.tar.gz version of Altmetric.com data containing 4.5 million publications in JSON format. Each publication in this data is mapped according to the social mention(s) it had received from different online platforms, including *Blogs, Downloads, F1000, Facebook, Googleplus, Linkedin, News, Peer Review (Publons and PubPeer), Pinterest, Policy, Q&A, Reddit, Twitter, Weibo and Wikipedia*. Note that Altmetric.com had only began to amass data from online platforms during the second half of 2011, hence to ensure a complete dataset, the set of publications that we focused on were published no earlier than July 2011.

Our working dataset for this study consisted of 1,104,275 publications, having at least on citation and at least one social mention, each of which had been indexed in 10,188 unique source (journal) titles, from the period of July 2011 to December 2015. The citation window for each publication was from the time of publication to February 2017. This gives more than a year time window to all publications that had been published in 2015.

Methodology

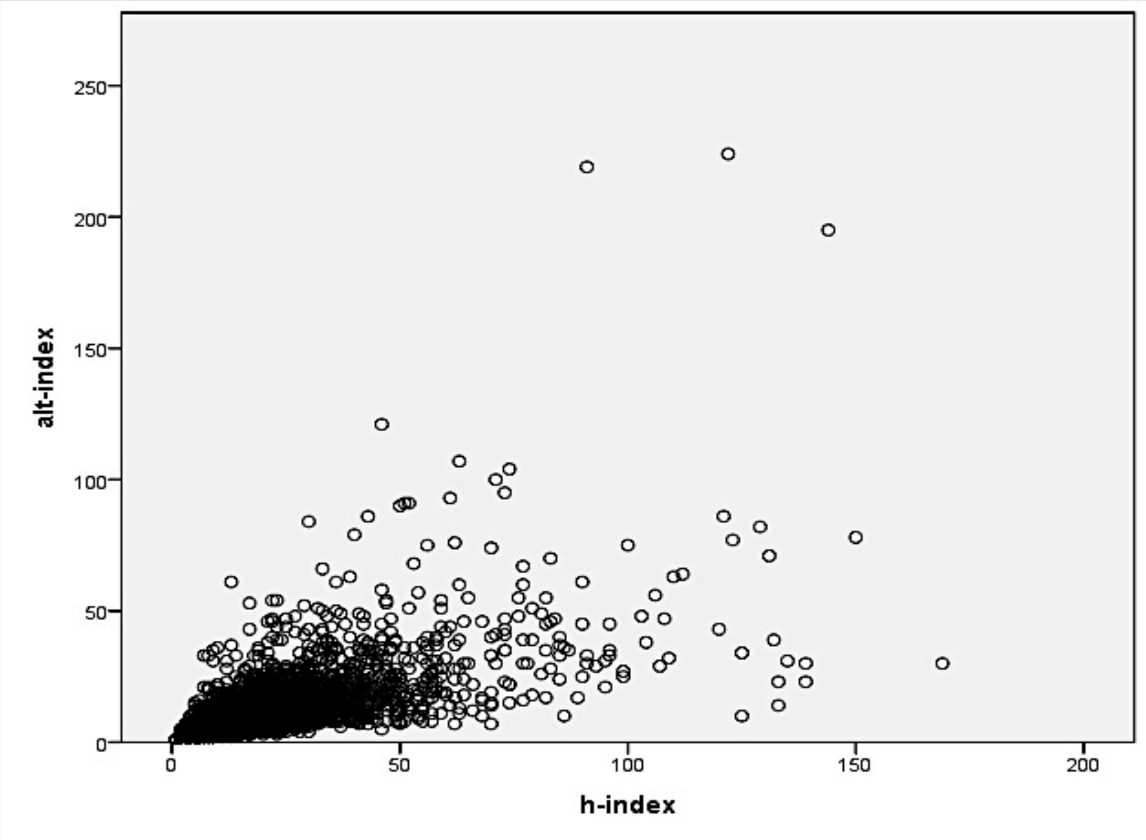
Using the original definition of h-index (Hirsch, 2005) and our newly proposed alt-index, we computed the indices for all 10,188 unique source titles. In Figure 1, below, we present a visual descriptive of both indices using a Scattered Plot. Next, we applied the Spearman's coefficient of correlation with bootstrap to study the relationship between the indices. And finally, we calculated the correlation between the indices across a selection of research disciplines, defined or categorized according to the Scopus All Science Journal Category (ASJC).

Using All Science Journal Category (ASJC), employed Scopus, the top level of the ASJC disciplines were merged (see Haddawy et al., 2017) to establish 16 top level subject disciplines, ranging from "Agricultural and Biological Sciences" through to "Social Science". This reduced mapping normalizes the double counting effect of journals being mapped in more than one subject categories. Note that all statistical analyses were prepared using IBM SPSS v.22.

Table 1. Top six source titles based on their h-index scores indexed in our dataset.

Source Title	**alt-index [rank]**	**h-index [rank]**
Nature	415 [1]	291 [1]
New England Journal of Medicine	240 [2]	263 [2]
Science	227 [3]	259 [3]
Cell	113 [10]	192 [4]
The Lancet	187 [7]	189 [5]
Proceedings of the National Academy of Sciences	182 [8]	170 [6]

Figure 1. Scatter plot of h-index and alt-Index source titles from Altmetric.com and Scopus.



Results

Figure 1 shows the alt-index and h-index scatterplot of all selected source titles, excluding only six listed separately in Table 1 for better visualization. Note from the ranked list, shown in Table 1, that *Nature*, *New England Journal of Medicine* and *Science* occupy the top positions based on both their h-index and alt-index scores. Similarly, we see that *Cell, The Lancet* and *Proceedings of the National Academy of Sciences* are listed amongst the top ten both in terms of their alt-index and h-index ranks (see Table 1). This shows that scientific achievements published in interdisciplinary venues like *Nature, Science* and *PNAS* often receive as much recognition from selected social media channels as they do from traditional scholarly communication channels. Moreover, *PLoS One* is ranked 4th in terms of its alt-index (i.e. 224), and this is next to Science. We also find *PLoS Biology* and *PLoS Medicine* at ranked at 11th and 13th place with alt-index scores of 107 and 100 respectively. This very high impact of *PLoS* source titles could be related to their open access business model.

Table 2, below, shows the Spearman's correlation values for the alt-index and h-index using Bootstrap. Here, we see a high positive correlation i.e. $\rho = 0.891$ between the indices with the following stability intervals between lower bound as $\rho = 0.886$ and upper bound as $\rho = 0.896$. In contrast to other studies, which show that absolute citation counts do not correlate well with Altmetric scores (Costas et al., 2015; Hassan et al., 2017) we find, from Figure 1, that our alt-index and the h-index do similarly identify high impact sources. These are results are encouraging since the alt-index and h-index examine quite different types of impact.

Table 2. Spearman's correlation using bootstrap for the h-index and alt-index of source titles from Altmetric.com and Scopus.

					alt_index	h_index
Spearman's rho	alt_index	Correlation Coefficient			1.000	.891*
		Sig. (2-tailed)				0.000
		N			10188	10188
		Bootstrap**	Bias		0.000	.000
			Std. Error		0.000	.002
			95% Confidence Interval	Lower	1.000	.886
				Upper	1.000	.896
	h_index	Correlation Coefficient			.891*	1.000
		Sig. (2-tailed)			0.000	
		N			10188	10188
		Bootstrap**	Bias		.000	0.000
			Std. Error		.002	0.000
			95% Confidence Interval	Lower	.886	1.000
				Upper	.896	1.000

*. Correlation is significant at the 0.01 level (2-tailed).
**. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

In Table 3, all disciplinary Spearman's rho values are presented (see Table 3). Note that every discipline shows a strong positive correlation ($\rho > 0.8$), although the absolute highest value ($\rho = 0.972$) is noted for the *General (Multidisciplinary)* category, which includes source titles such as *Science, Nature* and *PNAS.* It is clear here that multidisciplinary activities and publications tend to have the most influence, thus often lead to both high citations and high social media mentions. The following disciplines also show a strong positive correlation: *Agricultural, Biological Sciences* and Veterinary ($\rho = 0.906$), *Chemistry* ($\rho = 0.924$), *Earth and Planetary Sciences* ($\rho = 0.901$), *Materials Sciences* ($\rho = 0.927$), *Physics and Astronomy* ($\rho = 0.908$) and *Social Science* (ρ = 0. 904).

With Figure 2, below, we present scatterplots for the alt-index and h-index of source titles related to the sixteen broad disciplines of ASJC (see Fig. 2). Note that each of the scatterplots illustrate further the positive correlation between the indices.

Table 3: Cross-disciplinary analysis of Spearman's correlation using bootstrap between alt-index and h-index.

Discipline	N	Spearman's rho	95% Confidence Interval**	
			Upper Bound	**Lower Bound**
Agricultural, Biological Sciences and Veterinary	979	*0.906**	.919	.891
Biochemistry, Genetics and Molecular Biology	1212	0.892*	.906	.876
Chemistry	456	*0.924**	.938	.907
Computer Science	655	0.864*	.888	.837
Earth and Planetary Sciences	429	*0.901**	.923	.874
Economics, Business and Decision Sciences	861	0.871*	.889	.848
Engineering	1084	0.877*	.894	.894
Environmental Sciences	535	0.892*	.911	.869
General (Multidisciplinary)	26	**0.972***	.993	.909
Health Professions and Nursing	543	0.870*	.895	.842
Materials Sciences	446	*0.927**	.942	.909
Mathematics	578	0.883*	.905	.855
Medicine and Medical Sciences	3247	0.877*	.888	.867
Other Life and Health Sciences	956	0.877*	.887	.867
Physics and Astronomy	523	*0.908**	.925	.887
Social Science	2754	*0.904**	.912	.895

*. Correlation is significant at the 0.01 level (2-tailed).
**. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Concluding remarks

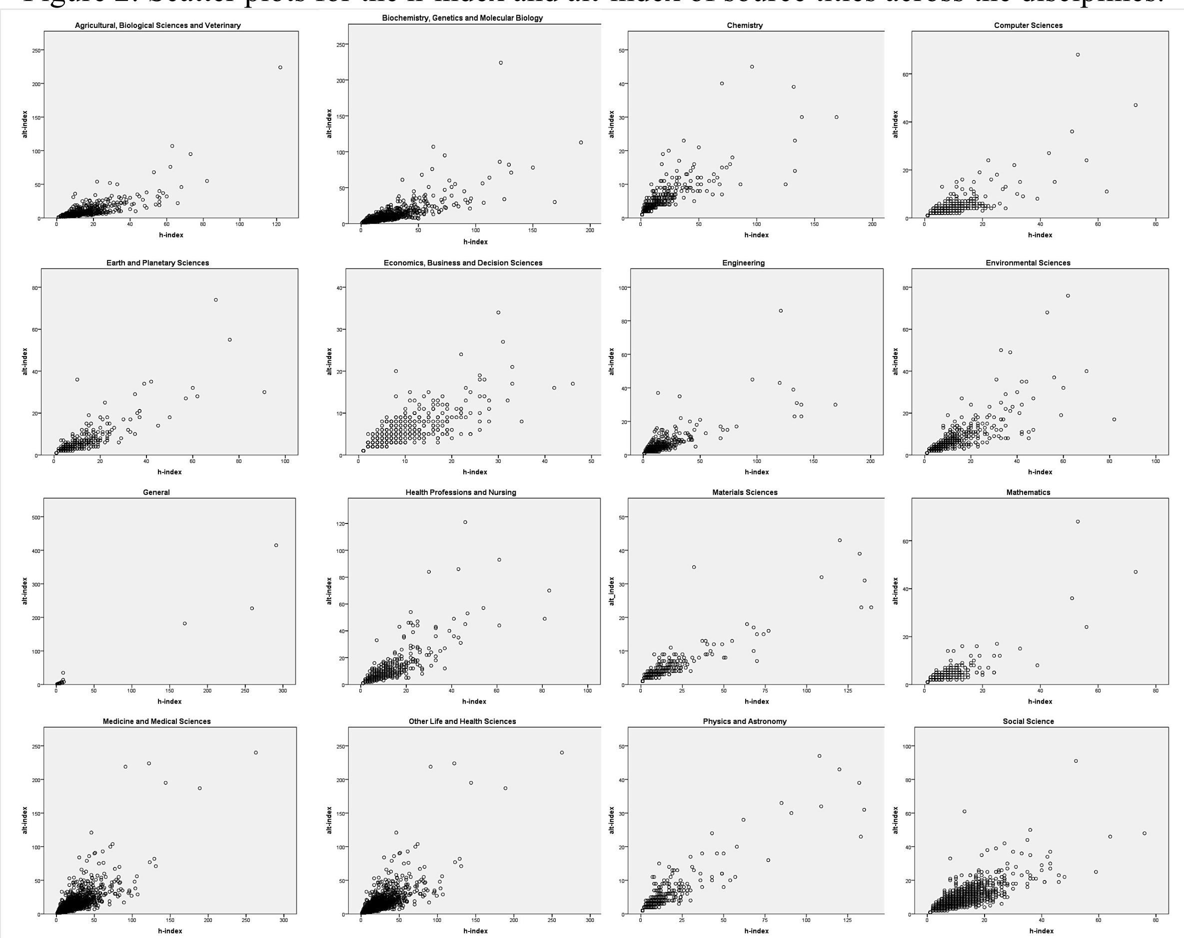
Few indicators have been developed so far for use with social media; however, with the alt-index, we see at least four specific advantages. The first advantage rest with our methodology, including a brief comparison to another research paper, where a similar index was proposed, called the t-factor (Bornmann & Haunschild, 2015). The focus of the t-factor was on Twitter mentions, but because the authors utilized a restricted dataset (i.e., only one publications with several Twitter mentions over time); it serves primarily as a proof-of-concept study, with no statistical claims like those shown in this paper. Unlike our approach, the authors also chose not to make any comparisons to any other traditional or conventional bibliometric measure. This alt-index study, by contrast, is statistically relevant because it is the result of a much more exhaustive analytic process, involving 1.1 million papers indexed in 10,188 source titles. By using such a large dataset, we have been able to show more comprehensively its comparative strength in relation to the journal h-index.

Another benefit of the alt-index is that if we also compare it to the simple Altmetrics.com donut, we see how the absolute donut count can be misleading, given that social media platforms are prone to bot based activities, in addition to organic ones. The proposed alt-index is well suited to tackle this issue, because it is designed to handle skewed data; thus, mentions attributed to bots will not influence the counting process. For example, if a journal and its research papers are more highly Tweeted due to bots, such highly skewed mentions are less likely to be included, given the reverse-order of ranking, which is utilized in a similar way with the h-index.

The third advantage pertains the idea that the alt-index is equally well suited to the assessment of mature as well as newly established source title (i.e., conference or journal). If we compare our proposed index to a traditional citation-based measure, which requires several years (5 to 7) before any scholarly impact can be established, the alt-index can quickly provide early feedback, allowing editors to determine how well a newly established source is performing, at least from the perspective of visibility and attention.

And last but not the least, the fourth advantage of the alt-index is that it may be computed using different variations analogous to the h-5 index, i10-index, g-index or applying different normalizations originally used with the h-index.

Figure 2: Scatter plots for the h-index and alt-index of source titles across the disciplines.



References

Bartneck, C. & Kokklemans, S. (2011). Detecting h-index manipulation through self-citation analysis. *Scientometrics, 87,* 85-98.

Bornmann, L., & Haunschild, R. (2015). t factor: A metric for measuring impact on Twitter. *arXiv preprint arXiv:1508.02179*.

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, 69(1), 169-173. Chicago.

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003-2019.

Haddawy, P., Hassan, S. U., Abbey, C. W., & Lee, I. B. (2017). Uncovering fine-grained research excellence: The global research benchmarking system. *Journal of Informetrics*, *11*(2), 389-406.

Hassan, S. U., Imran, M., Gillani, U., Aljohani, N. R., Bowman, T. D., & Didegah, F. (2017). Measuring social media activity of scientific literature: an exhaustive comparison of scopus and novel altmetrics big data. *Scientometrics, 113(2),* 1037-1057.

Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America, 102(46),* 16569–16572.

Kwok, R. (2013). Research impact: Altmetrics make their mark. *Nature,* 500(7463) 491-493.

Mingers, J., Macri, F., & Petrovici, D. (2012). Using the h-index to measure the quality of journals in the field of business and management. *Information Processing & Management*, *48*(2), 234-241.

Piwowar, H. (2013). Altmetrics: value all research products. *Nature*, *493*(7431), 159-159.

Priem, J. (2013). Scholarship: Beyond the paper. *Nature,* 495 (7442), 437-440.

Priem, J., Taraborelli, D., Groth, P., and Neylon, C. (2010). Altmetrics: a manifesto. Retrieved 26 April 2017 from: http://altmetrics.org/manifesto/.

Wilhite, Allen W., and Eric A. Fong. (2012). Coercive citation in academic publishing. Science, 335 (6068), 542-543.