

23rd International Conference on Science and Technology Indicators "Science, Technology and Innovation Indicators in Transition"

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 <u>https://hdl.handle.net/1887/64521</u>

ISBN: 978-90-9031204-0

© of the text: the authors © 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Atribution-NonCommercial-NonDetivates 4.0 International Licensed

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

Studying the Velocity Index for various Altmetric.com data sources¹

Zhichao Fang*, Rodrigo Costas**

**z.fang@cwts.leidenuniv.nl; rcostas@cwts.leidenuniv.nl* Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62A, Leiden, 2333 AL (The Netherlands)

**DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University (South Africa)

Abstract

In this study the velocity of 12 Altmetric.com data sources in disseminating newly published research outputs is investigated. The *Velocity Index* is proposed to make a comparison of velocity among Altmetric.com data sources across document types and subject fields. Some altmetric posts accumulated very fast within the first few days after publication, such as Reddit, Twitter, News, and Facebook, while posts of Policy documents, Wikipedia, Q&A, and Peer review with low Velocity Index values accrued relatively slowly. Most data sources' velocity degree also change by document types and subject fields. The velocity of most data sources confronted with the type of Review is lower than the overall and Article, while Editorial Material and Letter are higher. In general, most altmetric data sources show higher velocity values in the fields of *Multidisciplinary Journals* and *Natural Sciences*.

Introduction

"Speed", has been highlighted as one of the most important properties of altmetrics (Wouters & Costas, 2012; Bornmann, 2014). Speed in the context of altmetrics is related to the idea that the impact of a given scientific output can be measured and analysed soon after its publication with altmetric data sources. As a result, compared with citations, which has been often criticized for its time delay, altmetrics are assumed to be more *immediate*, so that online activities on newly published papers can be tracked much earlier (Priem, et al., 2010). The immediacy of some specific altmetric data sources have been discussed elsewhere. Maflahi & Thelwall (2016) found that Mendeley readership counts may be useful in measuring impact for both newer and older articles in the field of Library and Information Sciences. The results based on *PeerJ* social referrals data of Wang, Fang, & Guo (2016) suggested that the number of "visits" to papers from social media (Twitter and Facebook) accumulates very quickly after publication. Yu, et al. (2017) found that Twitter and Weibo are more immediate than citations, however they also suggested that not all altmetric data sources have the same degree of immediacy.

In order to measure the velocity degree of different altmetric data sources in disseminating newly published research outputs, we proposed the Velocity Index. Thus, this paper aims to answer the following research questions:

¹ This research is partially funded by the South African DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), and Zhichao Fang is partially supported by the China Scholarship Council (CSC).

- 1. Which data sources show the highest velocity in disseminating newly published research outputs and which are relatively low?
- 2. How do the Velocity Indexes of different Altmetric.com data sources vary across document types and subject fields?

Data

In order to exhibit the velocity of different Altmetric.com data sources in disseminating newly published research outputs at the day level, it is necessary to find a precise proxy for *the first publication date*² of research outputs and *post date*³ of altmetric records. In this study "created date" and "issued date" of DOIs collected from Crossref are combined to be used as the proxy for the first publication date, while the "posted on date" recorded by Altmetric.com for each altmetric event are collected to represent the post date of altmetric records.

There are four steps to select and clean Altmetric IDs matched with Crossref publication date and Web of Science (WoS) bibliographic information. Table 1 presents 13 data sources with posted on date information tracked by Altmetric.com⁴ together with the statistics of data selection and cleaning process in 4 steps.

Data sources	Step 1	Step 2		Step 3		Step 4	
	Altmetric IDs with posted on date	Altmetric IDs with DOI and without arXiv ID		Altmetric IDs with DOI indexed by Crossref and WoS		Altmetric IDs cleaned by the first seen date	
	N	N	%	Ν	%	Ν	%
Blog	716,934	578,186	80.65%	403,721	56.31%	359,730	50.18%
F1000	140,266	139,970	99.79%	125,384	89.39%	116,455	83.02%
Facebook	1,387,610	1,196,040	86.19%	829,797	59.80%	748,348	53.93%
Google+	239,112	186,942	78.18%	129,055	53.97%	109,365	45.74%
News	587,079	496,039	84.49%	377,215	64.25%	327,934	55.86%
Peer review	71,296	71,213	99.88%	46,911	65.80%	41,931	58.81%
Policy documents	722,126	697,604	96.60%	237,496	32.89%	224,615	31.10%
Q&A	28,973	18,149	62.64%	8,903	30.73%	7,785	26.87%
Reddit	90,428	69,933	77.34%	51,189	56.61%	43,407	48.00%
Syllabi	675,457	9,636	1.43%	1	0.00%	1	0.00%
Twitter	5,392,073	4,377,364	81.18%	3,136,167	58.16%	2,910,690	53.98%
Video	50,421	40,789	80.90%	27,521	54.58%	23,746	47.10%
Wikipedia	819,691	630,962	76.98%	272,050	33.19%	261,227	31.87%

Table 1. Statistics of data selection and cleaning process in 4 steps.

Step 1: Altmetric IDs with posted on date were selected. Until October 2017, there are 8,157,486 Altmetric IDs (account for 99.90%) have at least one record from these 13 data sources.

 $^{^2\,\}text{Date}$ on which a publication was first formally accessible and available to the scientific community or the public.

³ Date on which an altmetric event (e.g. tweets, blog mentions, news mentions) was posted online or published (for policy documents).

⁴ <u>https://help.altmetric.com/support/solutions/articles/6000136884-when-did-altmetric-start-tracking-attention-to-each-attention-source-</u>

Step 2: In order to match with Crossref publication date and WoS bibliographic information through DOIs, 6,221,669 Altmetric IDs which have DOI were selected. However, among these Altmetric IDs, there exists 79,761 Altmetric IDs (account for 1.29%) that have preprint version (i.e. with arXiv IDs). The existence of preprint version makes research outputs available to social media before they are formally published (Darling et al., 2013), which may lead to the altmetric post date to be earlier than the publication date. Therefore, Altmetric IDs with arXiv IDs were excluded.

Step 3: Altmetric IDs were matched with Crossref publication date and WoS bibliographic information. Finally 3,892,610 Altmetric IDs have DOIs recorded by Crossref until August 2017 and indexed by Web of Science until December 2017 at the same time.

Step 4: Altmetric.com "first seen date" was used as the benchmark. As mentioned above, except for the influence of preprint version, the first publication date of a publication should be expected to be earlier than its altmetric first seen date⁵, as in principle an altmetric post cannot mention a publication before it exists online. Consequently, first seen date of each Altmetric ID among 13 Altmetric.com data sources were aggregated to serve as the benchmark to examine whether the first publication date is reliable or not. After comparison, there are 245,567 Altmetric IDs (6.31%) with altmetric first seen date earlier than the first publication date. The possible reasons for the existence of these unreliable cases are the following:

- 1. Crossref "created date" and "issued date" are not always precise in reflecting the first publication date.
- 2. Publication date may be updated by publishers for many reasons (e.g. publisher mergers).

These Altmetric IDs with a first seen date before their first publication date were excluded. As a result, all of the altmetric posts about these 3,647,043 Altmetric IDs were analysed in our study. As Syllabi only has 1 Altmetric ID matches the conditions, Syllabi were excluded in this study and other 12 data sources were compared.

Results and Discussion

Velocity Index for measuring the velocity of altmetric data sources

From the view of altmetric data, in consideration of the diverse nature, scale, and user types of data sources, they also show different velocity degrees in face of newly published research outputs. To reflect the velocity differences among different altmetric data sources, we propose the *Velocity Index* for different Altmetric.com data sources: The proportion of altmetric posts accrued in a specific time interval (e.g. 1 day, 1 month, 1 year, etc.) after the publication of papers. The calculation method is shown in the formula below.

$$Velocity \ Index = \frac{P_i}{TP_i}$$

In a specific observed time window, *Pi* number of posts accrued in a time interval after publication (e.g. 1 day, 1 month, 1 year, etc.), *TPi* total number of posts during the observed

⁵ Date on which Altmetric.com captures the first event for a paper. Recorded for 99.9% of all the records in Altmetric.com

time window. The Velocity Index of each altmetric source symbolizes the preference of each source (e.g. Twitter, Facebook, etc.) to disseminate posts of publications in a given time. In general, the closer the Velocity Index to 1, the more immediately (faster) the data source is disseminating new publications in a given observation period. Conversely, the closer to 0, the slower (i.e. more posts beyond that period happened in the altmetric sources).

The Velocity Index of each Altmetric.com data source at the day, month, and year level are calculated respectively in an open time window, and the ranks are shown in Figure 1.The ranking varies at different time scale. Reddit, Twitter, and News are the most immediate data sources in disseminating newly published research outputs at the day, month, and year level. Followed by Facebook, Google+, and Blog. While Policy documents, Q&A, Peer review, Wikipedia, and Video perform weakly in Velocity Index. F1000, as one of the slowest data sources at the day level, ranks the first place at the year level, which means although F1000 is weak in disseminating newly published research outputs in a short time, it prefers to recommend those published without a long history.



Figure 1: Velocity Index ranking at the day, month, and year level.

Velocity Index variations across document types

The change trend of the Velocity Index of data sources at the month level across the top 4 document types with most number of publications: Article (N=3,022,507, Coverage=82.88%), Review (N=322,767, Coverage=8.85%), Editorial Material (N=182,094, Coverage=4.99%), and Letter (N=61,074, Coverage=1.67%), are illustrated in Figure 2. The type of Article dominates in the quantity of publications, so its Velocity Index is very close to the overall Velocity Index of each data source. Review, Editorial Material, and Letter, in comparison, show obvious discrepancy with overall Velocity Index, especially for data sources with relatively high Velocity Index values. In principle, for the type of Review, the Velocity Index is lower than the overall. Newly published Review is not as attractive as other main document

types for most altmetric data sources to disseminate immediately. Conversely, Editorial Material and Letter are more likely to be disseminated within a short time after publication. Their Velocity Indexes are higher than general level among those immediate data sources, such as Reddit, Twitter, News, and Facebook. In particular, Editorial Material and Letter hold relatively high Velocity Index on Peer review platforms (Publons and PubPeer⁶), which is classified into quite slow data source based on the overall Velocity Index. Review also has a higher Velocity Index than overall and Article on Peer review. On the on hand, compared with Article, Peer review platforms may notice and comment on Editorial Material, Letter and Review more quickly. On the other hand, the limited number of publications and the small coverage of Peer review posts (0.50% - 0.67%) of these three document types may intensify the performance of Velocity Index.





Altmetric.com data sources

Velocity Index variations across subject fields

The coverage of publications in Altmetric.com from different data sources differs by subject fields (Zahedi, Costas, & Wouters, 2014). In this study (Figure 3) we analysed the changes in the Velocity Index at the month level of different Altmetric.com data sources across 7 major fields of science (using the NOWT classification (Tijssen, Hollanders, & van Steen, 2010) developed by CWTS). Each row presents the Velocity Index of different altmetric data sources ranked from high to low in each NOWT field. Each altmetric data source in Figure 3 is indicated with the same colour, together with their specific Velocity Index. On the top of Figure 3, altmetric data sources are ranked by their overall Velocity Indexes at the month level. And colourful lines between two Velocity Indexes in the same colour display the rank changes for the same data source across fields. According to these results, Twitter, Facebook,

⁶ <u>https://www.altmetric.com/blog/a-tour-of-the-peer-reviews-tab/</u>

Reddit, and News are the most immediate data sources to newly published research outputs in all subject fields. By contrast, the overall Velocity Indexes of all altmetric sources in Multidisciplinary Journals and Natural Sciences are the highest. In these two fields, Reddit reaches the first place as the most immediate source, although it only covers very small shares of publications in most fields (0.41% - 7.40%). In the fields of Engineering Sciences, Language, Information and Communication, Medical and Life Sciences, and Social and Behavioural Sciences, Twitter ranks first. Facebook shows the highest velocity degree in the fields of *Law*, *Arts and Humanities*, although overall, the Velocity Index values of this field is comparatively low. News has relatively high Velocity Index in the fields of Engineering Sciences, Medical and Life Sciences, Multidisciplinary Journals, and Natural Sciences, while performs differently in humanities and social sciences, with much lower Velocity Index. The Velocity Index of Google+ also fluctuate to some extent across fields, it gets a relatively high Velocity Index in Language, Information and Communication, ranks only second to Twitter. As to other data sources, they keep a quite steady medium or low Velocity Index in all subject fields. For example, Policy documents and Q&A have the lowest Velocity Index across most fields, suggesting that these data sources are comparatively less focused on more recent publications as compared to the other sources whatever in which fields.



Figure 3: Velocity Index variations across seven NOWT subject fields.

Preliminary conclusions and outlook

In this study the velocity of 12 Altmetric.com data sources in disseminating newly published research outputs were investigated based on the proposed Velocity Index. The property of speed is not found to be owned by all of Altmetric.com data sources, existing a relevant differentiation between the fast sources (e.g. Reddit, Twitter, News) and the slow sources (e.g. Policy documents, Q&A, Wikipedia), which may also have implications for their analytical uses and applications.

The performance of velocity of Altmetric.com data sources varies across document types and subject fields. The velocity of most data sources confronted with the type of Review is lower than the overall and Article, while Editorial Material and Letter are higher. From the perspective of fields, the velocity ranking of different data sources changes across fields, and

most altmetric data sources show higher velocity values in the fields of *Multidisciplinary Journals* and *Natural Sciences*.

The main limitation of this study lies in the precision of Crossref "created date" and "issued date" as proxy for the first publication of research outputs. Although altmetric first seen date was used as the benchmark to exclude some unreliable data, Crossref cannot be seen as a perfect proxy for publication dates. There might still be a small distance between the date on which DOI was created and research output was actually made publicly available, which could result in some negative influence on our results. Future research will focus on these issues as well as on the study of advanced time-based analytics of altmetric data sources.

References

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, 8(4), 895-903.

Darling, E. S., Shiffman, D., Côté, I. M., & Drew, J. A. (2013). The role of Twitter in the life cycle of a scientific publication. *PeerJ PrePrints*, 1, e16v11.

Haustein, S., Bowman, T. D., & Costas, R. (2015). When is an article actually published? An analysis of online availability, publication, and indexation dates. *Proceeding of the 15th International Conference on Scientometrics and Informetrics (ISSI)*, (pp. 1170-1179), 29 Jun-4 July 2015, Istanbul, Turkey. <u>http://doi.org/arXiv:1505.00786v1</u>

Maflahi, N., & Thelwall, M. (2016). When are readership counts as useful as citation counts? Scopus versus Mendeley for LIS journals. *Journal of the Association for Information Science and Technology*, 67(1), 191-199.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. Retrieved from <u>http://altmetrics.org/manifesto/</u>

Tijssen, R., Hollanders, H., van Steen, J. (2010). Wetenschaps en Technologie Indicatoren 2010. Neederlands Observatorium Wetenschap en Technologie (NOWT).

Wang, X., Fang, Z., & Guo, X. (2016). Tracking the digital footprints to scholarly articles from social media. *Scientometrics*, 109(2), 1365-1376.

Wouters, P., & Costas, R. (2012). Users, narcissism and control-tracking the impact of scholarly publications in the 21st century. Utrecht: SURFfoundation. Retrieved from http://research-acumen.eu/wp-content/uploads/Users-narcissism-and-control.pdf

Yu, H., Xu, S., Xiao, T., Hemminger, B. M., & Yang, S. (2017). Global science discussed in local altmetrics: Weibo and its comparison with Twitter. *Journal of Informetrics*, 11(2), 466-482.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A crossdisciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics*, 101(2), 1491-1513.