



Universiteit  
Leiden

The Netherlands

## **Towards advanced social media metrics: understanding the diversity and characteristics of Twitter interactions around science**

Fang, Z.

### **Citation**

Fang, Z. (2021, December 21). *Towards advanced social media metrics: understanding the diversity and characteristics of Twitter interactions around science*. Retrieved from <https://hdl.handle.net/1887/3247587>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3247587>

**Note:** To cite this publication please use the final published version (if applicable).

# CHAPTER 3

## Studying the accumulation velocity of altmetric data tracked by Altmetric.com<sup>1</sup>

---

<sup>1</sup> **This chapter is based on:**

Fang, Z., & Costas, R. (2020). Studying the accumulation velocity of altmetric data tracked by Altmetric.com. *Scientometrics*, 123(2), 1077–1101. <https://doi.org/10.1007/s11192-020-03405-9>

**Author contributions:**

Fang, Z. (Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Data Curation, Writing - Original Draft, Writing - Review & Editing)

Costas, R. (Conceptualization, Methodology, Investigation, Supervision, Writing - Review & Editing)

## Abstract

This paper investigates the data accumulation velocity of 12 Altmetric.com data sources. DOI *created date* recorded by Crossref and altmetric event *posted date* tracked by Altmetric.com are combined to reflect the altmetric data accumulation patterns over time and to compare the data accumulation velocity of various data sources through three proposed indicators, i.e., *Velocity Index*, *altmetric half-life*, and *altmetric time delay*. Results show that altmetric data sources exhibit different data accumulation velocity. Some altmetric data sources have data accumulated very fast within the first few days after publication, such as Reddit, Twitter, news, Facebook, Google+, and blogs. On the opposite spectrum, research outputs are at relatively slow pace in accruing data on some data sources, like policy documents, peer review, Q&A, Wikipedia, video, and F1000Prime. Most altmetric data sources' velocity degree also changes by document types, subject fields, and research topics. The type Review is slower in receiving altmetric mentions than Article, while Editorial Material and Letter are typically faster. In general, most altmetric data sources show higher velocity values in the fields of *Physical Sciences and Engineering* and *Life and Earth Sciences*. Within each field, there also exist some research topics that attract social attention faster than others.

## Keywords

Altmetrics, Crossref, data accumulation speed, Velocity Index, altmetric half-life, time delay

### 3.1 Introduction

“Speed” has been highlighted as one of the most important characteristics of altmetrics (Bornmann, 2014a; Wouters & Costas, 2012). Compared to citations, which has been often criticized for its time delay in providing reliable measurement for research impact (J. Wang, 2013), speed in the context of altmetrics is related to the idea that the impact of a given scholarly output can be measured and analyzed much earlier (Mohammadi & Thelwall, 2014; Priem et al., 2010). Publication delays are considered to substantially slow down the formal communication and dissemination of scientific knowledge (Amat, 2008; Björk & Solomon, 2013). In contrast, interactions around science on social media platforms are likely to happen within a very short time-frame. For instance, Twitter mentions of scientific papers may occur immediately within hours or even minutes after they were available online (Haustein, Bowman, et al., 2015; Shuai et al., 2012).

However, because of the strong heterogeneity of altmetrics (Haustein, 2016), which incorporate a wide range of metrics based on different types of data sources, it is difficult to establish a clear-cut and unified conceptual framework for the temporal analysis of all altmetrics. Each altmetric indicator, typically with unique functions and aimed at different audiences, may tell different stories about the reception of scientific information, and show distinguishing patterns in varying contexts. Lin and Fenner (2013a) concluded that altmetrics are very likely representing very different things. From this point of view, we argue that the interpretation of the characteristic properties of different altmetrics should be made for each metric separately, including among these properties also their “speed”.

#### 3.1.1 Accumulation patterns and immediacy measurement of citations and usage metrics

In contrast to altmetric data, the accumulation patterns of citations have already been widely discussed in previous studies from several perspectives, such as their “obsolescence” (Line, 1993), “ageing” (Aversa, 1985; Glänzel & Schoepflin, 1995), “durability” (Costas et al., 2010), or “delayed recognition” (Garfield, 1980; Min et al., 2016). Citation histories, which relate to the analysis of the distribution of citations over time, were mainly studied from the synchronous or diachronous perspectives (Stinson & Lancaster, 1987). The former considers the distribution of the publication years of cited references, while the latter focuses on the distribution of received citations over time (Colavizza & Franceschet, 2016; Sun et al., 2016), which are also referred to as “retrospective citations” and “prospective citations”, respectively (Glänzel, 2004). These two approaches have been applied to study the accumulation patterns of usage metric data as well. With the development of digital publishing, usage metrics have been proposed and adopted by publishers during the last decades to supplement citations in reflecting how frequently scientific outputs are used and measuring their early impact to some extent (Schloegl & Gorraiz, 2011). From the

3

synchronous perspective, Kurtz et al. (2005) concluded that most studies of obsolescence found that the use of literature declines exponentially with age. The diachronous accumulation patterns of usage metrics, like views, downloads, or reads, were investigated and often compared with citations. On the basis of page views data of *Nature* publications, X. Wang et al. (2014) explored the dynamic usage history over time and found that papers are used most frequently within a short period after publication, finding that in median it only takes 7 days for papers to reach half of their total page views. Schlögl et al. (2014) reported that citations take several years until they reach their peak, however most downloads of papers are quickly accrued in the same publication year. In a similar fashion, Moed (2005) already found that citations and downloads show different patterns of obsolescence, and about 40% of downloads accumulated within the first 6 months after publication. More recently X. Wang, Fang, & Sun (2016) used the article-level “usage counts” provided by Web of Science to investigate the usage patterns of indexed papers, and identified that newly published papers accumulated more Web of Science usage counts than older papers.

As to the measurement of the “speed” of citations and usage metrics, several indicators have been created and applied in practice. For example, based on the time elapsed between the publication date and the date of the first citation of a paper, Schubert and Glänzel (1986) developed the indicator *mean response time* (MRT) in order to measure the citation speed of journals, understood as the properly formed average number of years between the publication of articles in a journal and the time of their first citation. In order to measure how quickly articles in a journal are cited, the *Journal Citation Reports* (JCR) calculates the indicator named *Immediacy Index* for each journal in each year. This indicator is defined as the average number of times an article is cited in the same year it is published.<sup>1</sup> Besides, at the journal level, *Cited Half-Life* and *Citing Half-Life* are also calculated by JCR to measure how fast journals are accumulating half of their citations and how far back that citing relationship extends.<sup>2</sup> Analogous to the citation-based Immediacy Index and half-life, the “usage immediacy index” and “usage half-life” (Rowlands & Nicholas, 2007), “download immediacy index” (Wan et al., 2010) were proposed to describe the life cycle of usage metrics. By analyzing usage data in the field of Oncology collected from Science Direct, Schloegl and Gorraiz (2010) calculated the mean usage half-life and found that it is much shorter than the average cited half-life, observing also different obsolescence patterns between downloads and citations.

---

<sup>1</sup> See more information about *Immediacy Index* at: <https://clarivate.com/webofsciencegroup/blog/know-your-metrics-immediacy-index/> (Accessed January 29, 2020).

<sup>2</sup> See more information about *Cited Half-Life* and *Citing Half-Life* at: <https://clarivate.com/webofsciencegroup/blog/a-closer-look-at-cited-and-citing-half-lives/> (Accessed January 29, 2020).

### 3.1.2 Accumulation patterns and immediacy measurement of altmetric data

Since the emergence of altmetrics, most related studies have focused on the coverage of scientific papers across altmetric sources and their correlation with citation counts (Costas et al., 2015a; Haustein, Peters, Bar-Ilan, et al., 2014; Thelwall, Haustein, et al., 2013). Less attention was paid to the study of the accumulation velocity of altmetric data over time. Only a few altmetric data sources were investigated from the perspective of their immediacy. Maflahi and Thelwall (2018) conducted a longitudinal weekly study of the Mendeley readers of articles in six library and information science journals and found that they start to accrue early from when articles are first available online and continue to steadily build over time, being this the case even for journals with large publication delays. Thelwall (2017) also found that articles attracted between 0.1 and 0.8 Mendeley readers on average in the month they first appeared in Scopus, with some variability across subject fields. The results based on *PeerJ* social referrals data of X. Wang, Fang, & Guo (2016) suggested that the number of “visits” to papers from social media (Twitter and Facebook) accumulates very quickly after publication. By comparing the temporal patterns of Twitter mentions and downloads of arXiv papers, Shuai et al. (2012) found that Twitter mentions have shorter delays and narrower time spans than arXiv downloads. Ortega (2018b) made a comparison of temporal distribution at the month time interval among citations, views, downloads, Mendeley readership, tweets, and blog mentions recorded by PlumX, and concluded that tweets and blog mentions are the quickest available metrics. 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 contrast to citation histories, which are mainly analyzed at year or month levels, for altmetrics it is insufficient to use such large time aggregations, since the real-time update of social media metric data makes altmetric events around research outputs visible within smaller time scales (e.g. hours or days). Nevertheless, a large-scale quantitative analysis comparing the data accumulation patterns of different altmetric data sources at the micro-level time interval (i.e., day) is still missing in the literature in altmetrics, probably caused by the absence of a reliable and precise proxy for publication dates, a piece of information that is critical in order to study the accumulation patterns of altmetric data (Haustein, Bowman, et al., 2015). Crossref provides several publication dates for its recorded DOIs, such as DOI *created date* (date on which the DOI was first registered), *published-online date* (date on which the work was published online), and *published-print date* (date on which the work was published in print). The distribution and potential of these date information for altmetrics have been compared and analyzed in a previous study (Fang & Costas, 2018), as suggested by Haustein, Bowman, et al. (2015), the value of DOI created date as a fine-grained benchmark of publication date in the context of altmetrics was highlighted.

In this paper, on the basis of DOI *created date* recorded by Crossref, as well as the altmetric event *posted date*<sup>3</sup> recorded by Altmetric.com, we compare the accumulation velocity amongst different types of altmetric data from a diachronous perspective.

### 3.1.3 Objectives

The main objectives of this study are: (1) to measure the accumulation velocity of altmetric data of scientific papers on 12 Altmetric.com data sources, here velocity referring to the pace at which altmetric events accumulate over time, and (2) to compare altmetric data accumulation velocity of different altmetric data sources across document types, subject fields, and research topics. The specific research questions are as follows:

**RQ1.** What are the altmetric data accumulation patterns of various Altmetric.com data sources?

**RQ2.** On which data sources do newly published research outputs show higher velocity in accruing altmetric data (and which ones are relatively lower)?

**RQ3.** How do the data accumulation velocity of different Altmetric.com data sources vary across document types, subject fields, and research topics?

## 3.2 Data and methods

### 3.2.1 Altmetric.com data sources with altmetric event posted date

In this study altmetric event records of 12 Altmetric.com data sources with *posted date* are selected as research objects. The altmetric data for this study were provided by Altmetric.com in a dump file with their data until October 2017. Table 1 presents these 12 data sources with event posted date information tracked by Altmetric.com together with the date when they started their coverage.

---

<sup>3</sup> This is the date on which a given altmetric event (e.g., a tweet, a news mention, or a blog citation) was posted online or published (for policy documents).

**Table 1.** Altmetric.com data sources with altmetric event posted date<sup>4</sup>

Data source	Concept measured with regard to research outputs	Coverage began date
Twitter	Twitter mentions, including original tweets, reply tweets, quote tweets, and retweets.	Oct 2011
Facebook	Facebook mentions, including posts on a curated list of public pages only.	Oct 2011
News	News media mentions on a list of news sources tracked by Altmetric.com, which contains over 2,900 English and non-English global news outlets.	Oct 2011 & Dec 2015
Blogs	Blog citations on a list of blogs tracked by Altmetric.com, which contains over 14,000 academic and non-academic blogs.	Oct 2011
Google+	Google+ mentions.	Oct 2011
Wikipedia	Wikipedia citations on English Wikipedia pages only.	Jan 2015
Policy documents	Policy documents citations on a wide range of public policy documents tracked by Altmetric.com, including policy, guidance, or guidelines documents from a governmental or non-governmental organization.	Jan 2013
F1000Prime	F1000Prime recommendations.	May 2013
Reddit	Reddit mentions on all sub-reddits, including original posts only.	Oct 2011
Peer review	Post-publication peer review comments collected from two forums: PubPeer and Publons.	Mar 2013
Video	Video mentions on YouTube.	Apr 2013
Q&A	Q&A mentions on Stack Overflow.	Oct 2011

**Note:** As of 2017, Altmetric.com has stopped collecting data from CiteULike, Sina Weibo, LinkedIn, and Pinterest. Syllabus data only posted in 2015 were provided by Altmetric.com and almost all publications mentioned by Syllabus are not indexed by Web of Science. Mendeley and CiteULike, two online reference managers, lack proper posted date information. Therefore, these data sources have not been included in this study. Although Google+ has also been discontinued and thereby Altmetric.com has stopped tracking it since January 2019, it is still considered as one of the data sources to be studied in this paper due to the availability of data during our observation time window.

### 3.2.2 Dataset

Considering the posted dates of the different altmetric events, we could know the exact date on which an altmetric event was posted. In addition, in order to study the accumulation patterns of altmetric data at the day time interval, DOI created dates of research outputs recorded by Crossref are collected to serve as the proxy of publication dates. To obtain both

<sup>4</sup> See more information about different Altmetric.com data sources at: <https://help.altmetric.com/support/solutions/articles/6000060968-what-outputs-and-sources-does-altmetric-track-> (Accessed November 26, 2019); and their coverage dates at: <https://help.altmetric.com/support/solutions/articles/6000136884-when-did-altmetric-start-tracking-attention-to-each-attention-source-> (Accessed November 26, 2019).



altmetric event posted date and DOI created date for measuring accumulation velocity, Web of Science (WoS) papers with the following criteria were selected as research objects:

- Papers with DOI recorded by Crossref. In order to get the DOI created dates, selected papers must have DOIs recorded by Crossref.
- Papers with publication date ranging from 2012 to 2016 according to both WoS publication year and Crossref DOI created date. To filter out old papers with newly registered DOIs (Fang & Costas, 2018), WoS publication year is also used as a benchmark to restrict the publication year of samples.
- Papers with at least one altmetric event recorded from any altmetric data source listed in Table 1.
- Papers without arXiv preprint version tracked by Altmetric.com. 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 record posted dates to be earlier than the publication date. Therefore, papers with arXiv IDs tracked by Altmetric.com are not included in this study.

According to the above criteria, there are 2,597,339 papers extracted from the CWTS in-house WoS database. However, 204,387 of them (accounting for 7.9%) have at least one altmetric event posted date earlier than their DOI created dates. Except for the influence of preprint versions, in theory an altmetric event cannot mention a DOI before it exists. The possible reasons for the existence of these unreliable cases are the following:

- Crossref DOI created dates may contain errors and not always accurately reflecting the publication date.
- Papers' DOI created dates may be updated by publishers due to different reasons (e.g., publisher mergers).<sup>5</sup>

In order to ensure the highest precision in our analysis, papers with any altmetric event posted date before their DOI created date are excluded from our analysis, resulting in a total set of 2,392,952 papers that are finally analyzed in this study. Table 2 lists the number of papers mentioned by each data source and the total number of altmetric events they have accumulated in the dataset. Twitter contributes the most majority of altmetric data to the selected papers, followed by Facebook.

---

<sup>5</sup> Extracted from personal communication with Euan Adie from Altmetric.com.

**Table 2.** General presence of altmetric data for the dataset

Data source	Number of papers	Number of altmetric events	Coverage	Intensity
Twitter	2,157,556	14,853,823	90.2%	6.9
Facebook	545,370	1,375,880	22.8%	2.5
News	224,036	1,037,719	9.4%	4.6
Blogs	200,784	360,736	8.4%	1.8
Google+	84,754	216,787	3.5%	2.6
Wikipedia	75,693	106,917	3.2%	1.4
Policy documents	56,296	73,523	2.4%	1.3
F1000Prime	39,981	48,517	1.7%	1.2
Reddit	31,726	43,805	1.3%	1.4
Peer review	20,783	33,599	0.9%	1.6
Video	12,918	18,643	0.5%	1.4
Q&A	2,369	2,474	0.1%	1.0

**Note:** Coverage refers to the proportion of papers with at least one corresponding altmetric event of all papers in our dataset. Intensity refers to the mean number of altmetric events of papers with at least one corresponding altmetric event (Haustein, Costas, et al., 2015).

### 3.2.3 Indicators and analytical approaches

Considering the diverse nature, scale, and user types of different altmetric data sources, it is very likely that they exhibit also very different velocity degrees of accumulation in face of newly published research outputs. To reflect the velocity differences among altmetric data sources, we use three indicators to measure velocity from both flexible and fixed perspectives, including *Velocity Index*, *altmetric half-life*, and *altmetric time delay*.

For altmetric data accumulated on a specific data source, the *Velocity Index* (VI) refers to the proportion of altmetric events that happened in a specific time interval (e.g., 1 day, 1 month, or 1 year) after the publication of the papers. The calculation method is shown in the formula below.

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

$P_i$  is the number of events accrued in a specific time interval after publication (e.g., 1 day, 1 month, or 1 year) for a set of papers,  $TP_i$  indicates the total number of events during the observed time window. In general, the closer to 1 of the Velocity Index, the more immediate (faster) the altmetric data of new papers accumulated in the given observation period.

Conversely, the closer to 0, the lower the accumulation velocity (i.e., more events happened beyond the specified period of time).

Besides, in line with the *Twitter half-life* and *Twitter time delay* proposed by Haustein (2019), which refer to the number of days until 50% of all tweets have appeared and the number of days between the publication of a document and its first tweet, respectively, we generalize these indicators for all altmetric data sources. Consequently, the *altmetric half-life* of an altmetric data source is defined as the number of days until half of its events have appeared, and *altmetric time delay* of a research output on an altmetric data source is defined as the number of days between its publication and its first altmetric event on that data source.

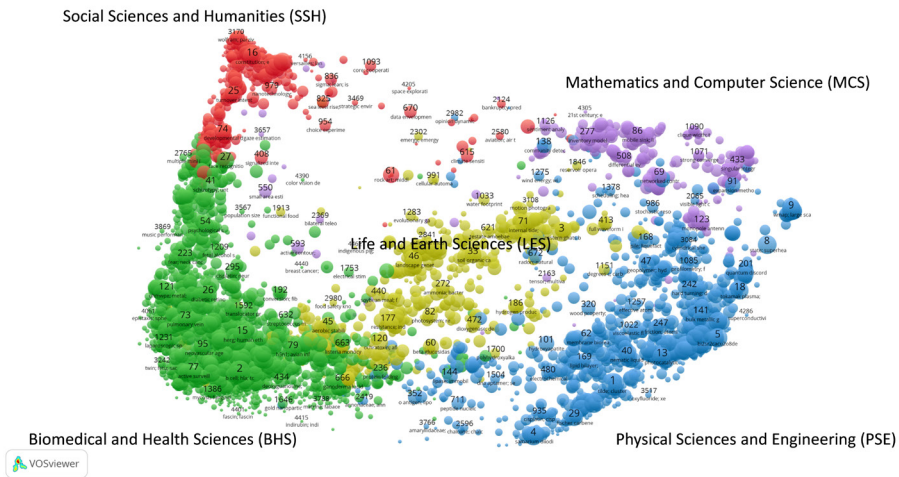
Both Velocity Index and altmetric half-life are based on overall data distribution of all events received by a paper, while altmetric time delay focuses on a special altmetric event (the first one). Velocity Index provides a flexible perspective for the measurement of data accumulation velocity, since it allows for more nuanced time accumulation discussions considering different time intervals (i.e., days, months, or years). By comparison, altmetric half-life and altmetric time delay provide a fixed perspective at the day level. Therefore, these indicators work as relevant complements to each other in order to better characterize the tempo of altmetric data accumulation.

In addition, the Spearman correlation analysis is performed with IBM SPSS Statistics 25 to explore the relationships among Velocity Index, altmetric half-life, and altmetric time delay. Also, at the research topic level, in order to testify whether or not research topics with fewer papers and altmetric events are more likely to reach higher values of Velocity Index, the Spearman correlation analysis is applied to exhibit the relationships among number of papers, number of altmetric events, and the Velocity Index.

### 3.2.4 CWTS publication-level classification system

The CWTS classification is a publication-level subject field classification system developed by Waltman and Van Eck (2012). It has not only been applied in Leiden Ranking (<https://www.leidenranking.com/>), but has also been employed by many previous studies for subject field related analysis (Costas et al., 2015a; Didegah & Thelwall, 2018). In the 2019 version of the publication-level classification, only citable items (Article, Review, and Letter) indexed by Web of Science are clustered into 4535 micro-level fields. These micro-fields correspond to small research topics (*micro-topics*), and they are assigned to five main subject fields of science algorithmically obtained, including *Social Sciences and Humanities* (SSH), *Biomedical and Health Sciences* (BHS), *Physical Sciences and Engineering* (PES), *Life and*

*Earth Sciences (LES)*, and *Mathematics and Computer Science (MCS)*,<sup>6</sup> which are illustrated in Figure 1 with VOSviewer. The layout of Figure 1 is also used to exhibit the Velocity Index of each micro-topic in the Result section. For the selected papers in our dataset, 2,189,708 of them (accounting for 91.5%) have CWTS classification information. This set of papers is drawn as our final sample of papers for the comparison of altmetric data accumulation velocity across subject fields and research topics. Statistics on the general presence of different altmetric data across five main subject fields can be found in Table 4 in Appendix.



**Figure 1.** Five main subject fields of science of the CWTS publication-level classification system. Each circle represents a micro-level field clustered by papers with similar research topics (micro-topics)

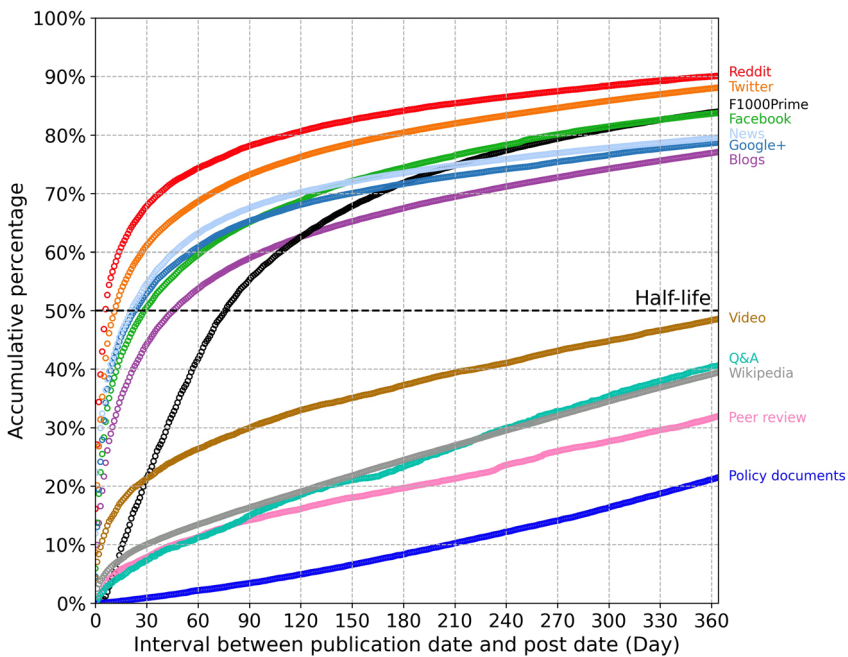
### 3.3 Results

#### 3.3.1 Altmetric data accumulation patterns

The intervals between publication dates and altmetric events posted dates are calculated for all altmetric events on each data source. Thus we can investigate the altmetric data accumulation patterns at the day time interval. Figure 2 shows the different data accumulation patterns of the 12 data sources within 1-year time interval (365 days) after publication. Data

<sup>6</sup> See more details about the CWTS classification system at: <https://www.leidenranking.com/information/fields> (Accessed January 29, 2020).

sources show different data accumulation patterns. Altmetric events to newly published research outputs on some data sources accumulated very fast, such as Reddit and Twitter, since half of their data accrued in the first 2 weeks (14 days) after the research outputs were published, and over 85% of their data happened within a year (365 days). Following Twitter and Reddit we have other pretty fast altmetric data sources including news, Google+, Facebook, and blogs. In contrast, policy documents, Wikipedia, Q&A, and peer review show much slower data accumulation patterns similar to that of traditional citations. Only 21.5% of policy document citations, 31.9% of peer review comments, 39.4% of Wikipedia citations, and 40.6% of Q&A mentions are accumulated within 1 year, which means that most of the events from these data sources happened more than a year after publication. Among these data sources, F1000Prime presents some uniqueness. In the first month after research outputs are published, the accumulation of F1000Prime recommendations is not very fast, but it speeds up over time, with more than 84% of data accrued within the first year.



**Figure 2.** Altmetric data accumulation patterns of 12 Altmetric.com data sources within the first year (365 days) after publication

The dashed line at accumulative percentage of 50% in Figure 2 indicates the altmetric half-life, and Table 3 lists the altmetric half-lives of the 12 data sources analyzed. Reddit ranks

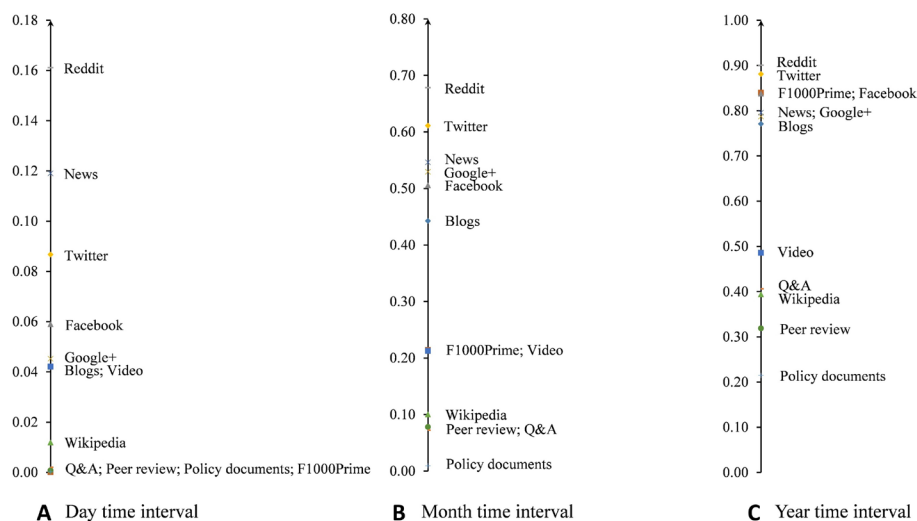
first, with a half-life of 7 days, followed by Twitter (13 days), news (22 days), Google+ (25 days), and Facebook (30 days). Over half of altmetric events on these data sources happened within 1 month after the publication of research outputs. Other sources such as Wikipedia, peer review, and policy documents, need over 500 days to accumulate half of their event data. On the one hand, these data sources show lower reaction speed towards newly published papers. On the other hand, it suggests that they also pay more attention to papers with older publication time.

**Table 3.** Altmetric half-lives of 12 Altmetric.com data sources

Rank	Data source	Altmetric half-life (day)
1	Reddit	7
2	Twitter	13
3	News	22
4	Google+	25
5	Facebook	30
6	Blogs	47
7	F1000Prime	77
8	Video	394
9	Q&A	498
10	Wikipedia	515
11	Peer review	633
12	Policy documents	716

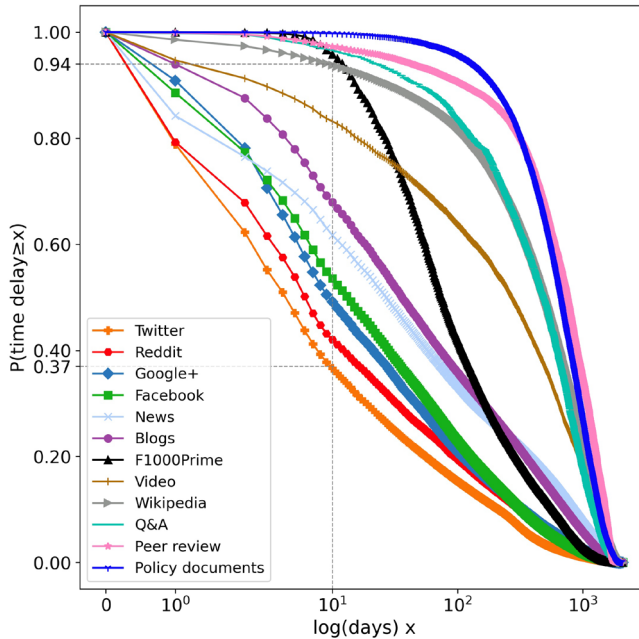
### 3.3.2 Generalizing the Velocity Index and altmetric time delay

The Velocity Indexes of each Altmetric.com data source at the day, month, and year time intervals are calculated respectively, and the rankings of sources by their Velocity Index are shown in Figure 3. The rankings vary at different time intervals. Reddit, Twitter, and news are the data sources showing the most immediate data accumulation patterns at the day, month, and year time intervals. Followed by Facebook, Google+, and blogs. While policy documents, peer review, Wikipedia, Q&A, and video perform more slowly in their Velocity Index values. F1000Prime, as mentioned above, although one of the slowest data sources at the day time interval, ranks the third at the year time interval. This means that the accumulation of F1000Prime recommendations of newly published papers is relatively slow in the short term, but it is faster at the year time interval (see also Figure 2). The case of F1000Prime highlights the importance of considering together the altmetric half-life of data sources and their Velocity Index, since both bring two different perspectives about the tempo of altmetric data.



**Figure 3.** Velocity Index rankings at the **A** day, **B** month, and **C** year time intervals

Besides the Velocity Index and altmetric half-life which are based on overall altmetric data of each data source, we also consider the time delay of papers until they accrued their first altmetric event from different data sources, in which case only one specific altmetric event of papers is considered. The number of days between being published and being mentioned for the first time on a certain data source is calculated for each paper, and the distribution of altmetric time delays of the 12 Altmetric.com data sources is plotted in Figure 4. Each curve shows, for each specific data source, the proportion of papers that accrued the first altmetric event beyond certain number of days since being published. For instance, only about 37% of papers received their first Twitter mentions after the 10th day after being published (the vertical dashed line in Figure 4), while 94% of papers received their first Wikipedia citations after the 10th day after publication. In other words, around 63% of papers obtained their first Twitter mentions within 10 days after publication, and only 6% of papers got the first Wikipedia citations within the same time period. The more skewed the curve, the higher the proportion of papers accrued their first altmetric event after a long time. As a result, papers are faster to be visible on Twitter compared to other data sources, followed by Reddit, Google+, and Facebook. For various altmetric data sources, the patterns of accumulating the first altmetric event are quite similar with their Velocity Indexes at the month time interval and altmetric half-lives (in Appendix Table 5 provides the spearman correlations for the rankings based on these three indicators).



**Figure 4.** Distribution of altmetric time delay of 12 Altmetric.com data sources in log(days)

Overall, Twitter, Reddit, Google+, Facebook, news, and blogs can be categorized as *fast sources*, while in general, F1000Prime, video, Wikipedia, Q&A, peer review, and policy documents show lower velocity in mentioning scientific papers. These six data sources can be classified as *slow sources*.

### 3.3.3 Velocity Index variations across document types

For different document types, their altmetric data accumulation velocity might show some differences. So we utilize the Velocity Index at the month time interval to measure the altmetric data accumulation velocity for different document types across data sources. The differences in the Velocity Index across the four main document types with most number of papers: Article (N = 1,951,197, Coverage = 81.5%), Review (N = 196,722, Coverage = 8.2%), Editorial Material (N = 139,950, Coverage = 5.8%), and Letter (N = 52,038, Coverage = 2.2%), are illustrated in Figure 5. The presence of altmetric data across these four document types is listed in Table 6 in Appendix. The type of Article is the largest in number of papers, so its Velocity Index is very close to the overall Velocity Index of each data source. Review, Editorial Material, and Letter, in comparison, show differences with the overall Velocity Index, especially for data sources with relatively high Velocity Index values. Reviews are not as fast in accumulating altmetric data as compared to the other document



types. Conversely, Editorial Material and Letter are document types more likely to be mentioned faster after publication. The Velocity Indexes of these two document types are higher than the overall Velocity Index for most data sources. In particular, Editorial Material and Letter hold relatively high Velocity Indexes on peer review platforms (Publons and PubPeer), which is among the group of “slower” data sources based on the overall Velocity Index (Figure 3) and its altmetric half-life (Table 3). The Review type also has a slightly higher Velocity Index than the overall and Article type on peer review events. Results show that peer review platforms seem to notice and comment on Editorial Materials, Letters and Reviews more quickly than regular Articles. Although the coverage of these three document types with peer review data is limited (0.20-0.27%), there are larger shares of peer review comments that happened soon after their publication compared to other altmetric events of slow sources.

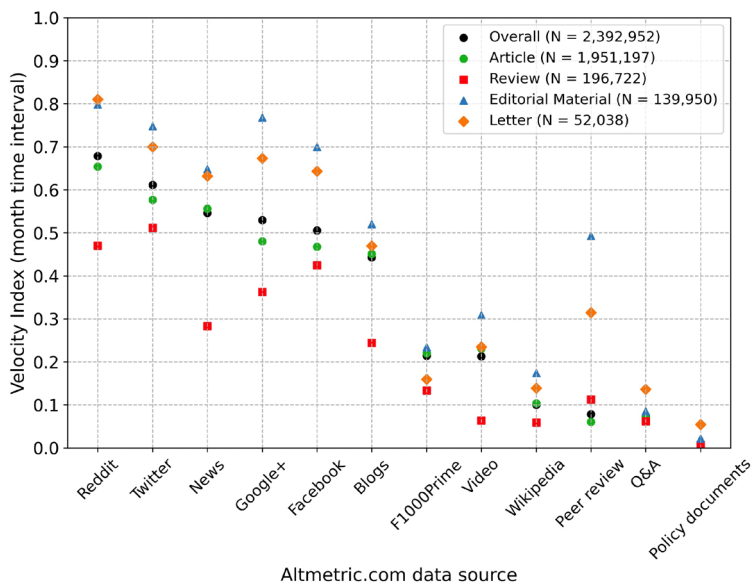


Figure 5. Velocity Index variations across four document types

### 3.3.4 Velocity Index variations across subject fields

The coverage of papers in Altmetric.com from different data sources differs by subject field (Costas et al., 2015b; Zahedi et al., 2014). In this study (Figure 6) we analyze the changes in the Velocity Index at the month time interval of different Altmetric.com data sources across five major subject fields of science (using the CWTS classification). Each row presents the

Velocity Indexes of different altmetric data sources ranked from high to low in each subject field. Each altmetric data source in Figure 6 is indicated with the same color, together with their specific Velocity Index. On the top of Figure 6, altmetric data sources are ranked by their overall Velocity Indexes at the month time interval. Colorful lines between two Velocity Indexes in the same color display the rank changes for the same data source across subject fields. According to these results, Twitter and Reddit are the most immediate data sources to newly published papers in all subject fields. By subject fields, the overall Velocity Indexes of all altmetric sources in *Physical Sciences and Engineering* (PSE) and *Life and Earth Sciences* (LES) are the highest. Facebook shows the higher immediacy degree in the fields of *Social Sciences and Humanities* (SSH) and *Mathematics and Computer Science* (MCS), although overall, the Velocity Index values of these subject fields are comparatively low. Conversely, news has relatively high Velocity Index in the fields of *Physical Sciences and Engineering*, *Life and Earth Sciences*, and *Biomedical and Health Sciences* (BHS), while it is slower in *Social Sciences and Humanities*. As to other data sources, they keep quite steady medium or low Velocity Indexes in all subject fields. For example, policy documents, peer review, and Q&A have the lowest Velocity Indexes across most subject fields, suggesting that these data sources are comparatively less focused on more recent papers as compared to the other sources regardless the subject fields of the papers.

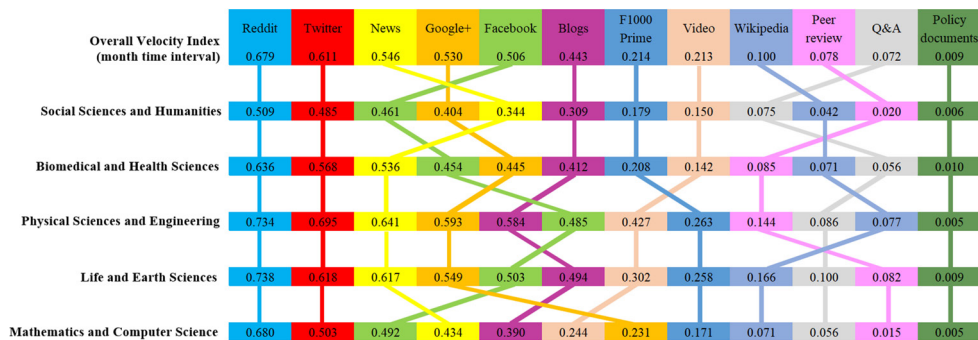
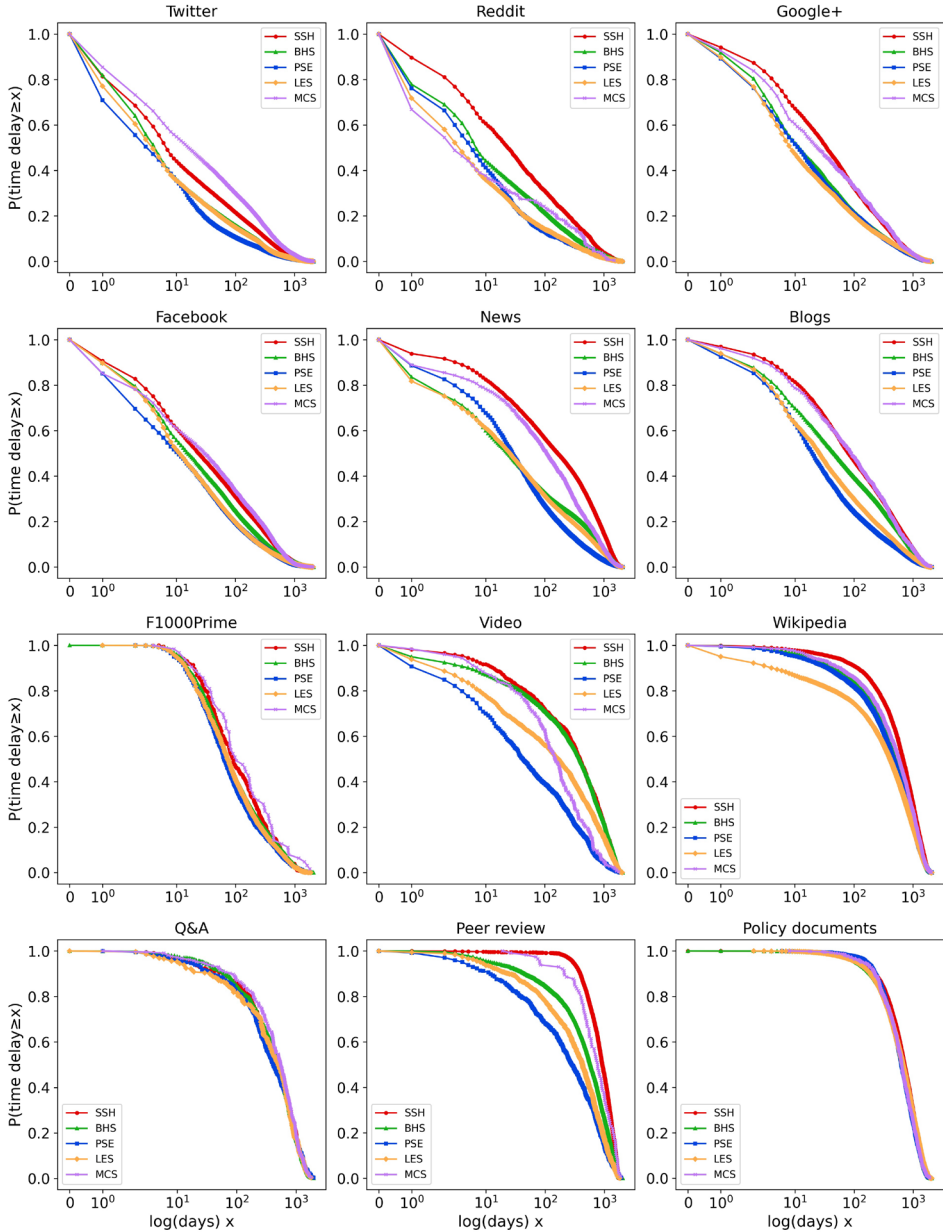


Figure 6. Velocity Index variations across the five subject fields

From the perspective of altmetric time delay, Figure 7 shows the distribution of altmetric time delay across the five subject fields for 12 Altmetric.com data sources respectively. For most data sources, although to different degrees, scientific papers in the fields of PSE and LES are faster to receive their first altmetric mention. In contrast, it took more days for papers in the fields of SSH and MCS to accumulate the first altmetric event record. Altmetric time delays of papers in BHS are in the middle on most data sources. Still, the accumulation

velocity across subject fields in terms of altmetric time delay is similar with the results observed through the lens of Velocity Index.



**Figure 7.** Distribution of altmetric time delays of 12 Altmetric.com data sources in log(days) across the five subject fields

### 3.3.5 Velocity Index variations across research topics

Considering the Velocity Index at the month time interval, we further investigate the variations across research topics to study which topics accumulated altmetric data faster than others. Twitter and Wikipedia are selected as two representatives for fast sources and slow sources, respectively, because they hold the largest data volume among their same types of data sources. Velocity Indexes are calculated for papers within each micro-level field sharing the similar research micro-topics based on Twitter mention data (Figure 8) and Wikipedia citation data (Figure 9). In both Figures 8 and 9, size of each circle is determined by the number of papers with Twitter mention/Wikipedia citation data in this micro-level field, while color is determined by the Velocity Index at the month time interval. Within micro-level fields, number of papers and number of altmetric events are very weakly correlated with the Velocity Index values based on Twitter data, and are moderately and positively correlated with those based on Wikipedia data (see Table 7 in Appendix), indicating that not all of micro-level fields with fewer papers are more likely to reach high Velocity Index, and vice versa. Some prominent research micro-topics with relatively high Velocity Index values in every main subject field are highlighted with annotation texts.

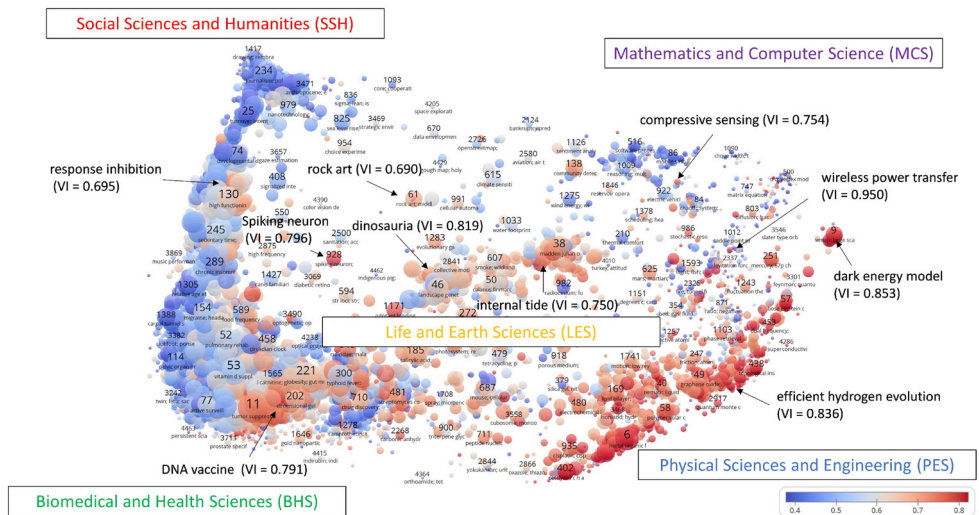
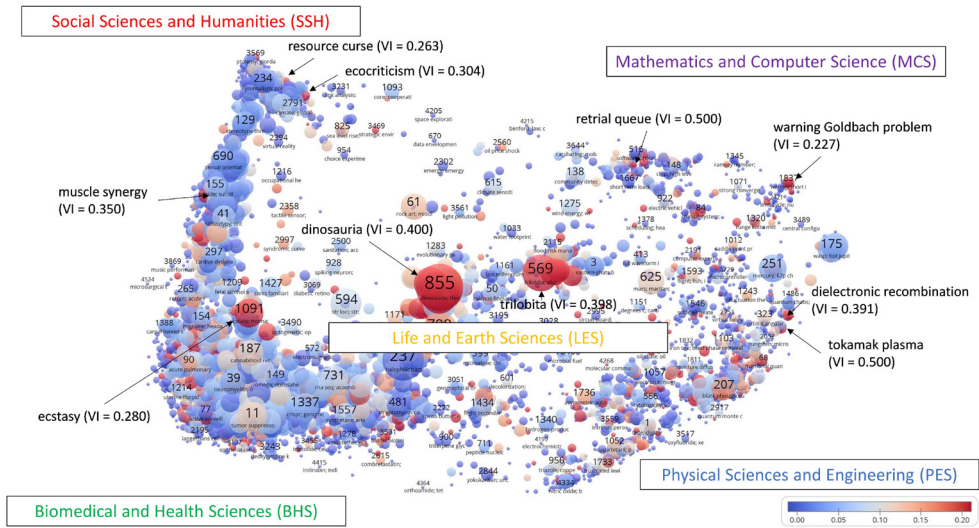


Figure 8. Velocity Index variations across research micro-topics (Twitter)



**Figure 9.** Velocity Index variations across research micro-topics (Wikipedia)

From the point of view of Twitter data, research micro-topics in the fields of PSE exhibit the highest Velocity Index values in contrast to the other fields, which is in correspondence with the above observations. Within the other subject fields, there are some research micro-topics that show quite high Twitter mention accumulation velocity as well. For example, “wireless power transfer” and “compressive sensing” in MCS accumulated the majority of their Twitter mentions in a short time, as well as “dinosauria” and “internal tide” in LES. In the fields of BHS and SSH, “DNA vaccine”, “spiking neuron”, “response inhibition”, and “rock art” drew attention on Twitter relatively fast too.

Compared to Twitter mentions, the overall accumulation velocity of Wikipedia citations is much lower, and the difference among main subject fields is not as obvious as Twitter. However, there also exist some research micro-topics showing higher data accumulation velocity. For instance, “dinosauria” and “trilobita” in LES are two micro-topics faster in Wikipedia. Papers about these two topics received more Wikipedia citations in a short time period compared to the others. Similarly, “ecstasy” (caused by drugs), “muscle synergy”, “warning Goldbach problem” and some other research micro-topics accumulate Wikipedia citations also relatively fast. In the field of SSH, although most research micro-topics were quite slow to be cited by Wikipedia, some environmental protection related micro-topics, such as “ecocriticism” and “resource curse”, show higher Velocity Index values.

### 3.4 Discussion

Speed has always been assumed as a characteristic property of altmetrics, however not much research has been done in characterizing the accumulation velocity of different altmetric data on a large scale. This study fills this gap by describing the immediacy of altmetric data accrued after the publication of scientific papers. Using the DOI created date and altmetric event posted date enables the possibility of studying the altmetric data accumulation patterns at the day level. The date when a DOI was assigned to a paper provided by Crossref has already been used to show the life cycle of some altmetric events at the month level by Ortega (2018b). This study investigates further on the accumulation velocity of various altmetric data at a more micro-level time interval and considering a larger scale of data samples.

As observed by Sun et al. (2016), citation histories typically show a pattern of just a few citations accrued within the first few years after publication, reaching a citation peak after 3-4 years, and then a decrease afterwards. Yet most kinds of altmetric data exhibit a different accumulation pattern compared with citations. We found that the accumulation velocity of different altmetric data vary substantially across data sources, document types, and subject fields.

#### 3.4.1 Variations across altmetric data sources

It is demonstrated that various altmetric data sources vary in their data accumulation patterns, and the property of speed is not found to be owned by all of altmetric data sources. Some of the altmetric data sources accrue a considerable proportion of events very soon after the publication date of scientific papers. Among these outputs we have Reddit, Twitter, news, Facebook, Google+, and blogs. All these altmetric data sources exhibit short altmetric half-lives, short altmetric time delays, and relatively high Velocity Indexes. Therefore, it can be argued that their velocity aligns with the property of speed that altmetrics are expected to have, being possible to label these as fast sources. However, for policy documents, Q&A, peer review, Wikipedia, video, and F1000Prime events, only a very limited share of these altmetric events happened within a short time after publication, being these slow sources. The data accumulation velocity of some slow sources are similar to that of citations, with important delayed patterns after publication. For example, based on our dataset, half of policy document citations happened after 716 days since publication. Older papers, however, seem also to still be attractive for these slow data sources, so that their attention is not concentrated on just newly published scientific papers. As a whole, most social media platforms and mainstream media are more immediate in sharing, discussing, and reporting new research outputs.

Interestingly, different time windows may also show different sources as being fast or slow. For example, although F1000Prime is seen as a slow source in the short term (e.g., day or

month level), it is one of the sources that accumulated the largest share of its events within 1 year. This reinforces the importance of combining different perspectives (e.g., different indicators, different time windows) to study the tempo of altmetrics to provide the most complete picture.

As a result, assumptions about the “speed” of types of events classified under the umbrella term “altmetrics” should be taken with particular caution. Not all of them are fast sources, and not all of them have the same accumulation pace. Thus, it is important to take the social media environment in which these events are produced into consideration (Alperin, 2015). Once again, caution about the merging of altmetric sources in compound metrics or global indicators must be observed, particularly considering that time affects differently to different sources. Keeping altmetric events separate seems to be an important recommendation, this given not only their fundamental differences (Haustein, Bowman, & Costas, 2016; Wouters et al., 2019) but also their time accumulation patterns as demonstrated in this study. Moreover, the pace and tempo of different altmetrics cannot be seen as equivalent and, similar to what happens with citations, these time differences need to be taken into account when considering different time windows in altmetric research.

### 3.4.2 Variations across document types

Zahedi et al. (2014) concluded that the coverage of several altmetric data sources varies across document types and subject fields. In this study, it is shown that the same type of variations applies also to the data accumulation velocity of different altmetric data sources. In terms of document types, Reviews (this document type mainly focuses on retrospectively reviewing existing findings) are overall the slowest in accumulating altmetric events. A possible reason for this slowest reception lies in the less innovative nature of Reviews. In other words, Review papers are less prone to provide new research discoveries and more to condense the state-of-the-art in a subject field or research topic, therefore lacking the novelty component of other document types. For example, the research topics presented in Editorial Materials and Letters may be more likely to evoke social buzz immediately, since they cover more novel topics, debates, scientific news, etc., without using a too complicated and technical language (Haustein, Costas, et al., 2015). The thematic property of these two document types might facilitate the users’ attention received more immediately, particularly on peer review platforms, a type of altmetric data source which is mainly used by researchers, who are faster to take notice of controversial topics emerging in the scientific community. This finding is quite similar with the ageing patterns of citations to different document types: Editorial Materials and Letters were found more likely to be the “early rise-rapid decline” papers with most citations accumulated in a relatively short time period, while Review was observed to be the delayed document type with a slower growth (Costas et al., 2010; J. Wang, 2013).

### 3.4.3 Variations across scientific fields and topics

In terms of scientific fields, research outputs from the fields of PSE and LES are more attractive to social media audiences shortly after publication, accruing altmetric events faster compared to other fields. Research outputs from the fields of both SSH and MCS are relatively slower to be disseminated on altmetric data sources, although papers in these two fields hold different altmetric data coverage, with the former much higher than the latter (Costas et al., 2015a; Fang, Costas, et al., 2020). Such field-related data accumulation dynamics was also observed in the context of citations, for instance, citation ageing in the social sciences and mathematics journals is similarly slower than in the medical and chemistry journals (Glänzel & Schoepflin, 1995), the physical, chemical, and earth sciences, fields in which the research fronts are fast-moving, have more papers showing rapidly declining citation pattern (Aksnes, 2003). From the perspective of first-citation speed, papers in the field of physics are faster in receiving the first citation, followed by biological, biomedical, and chemical research, while mathematics papers show lower first-citation speed (Abramo et al., 2011). Even though the overall accumulation patterns between citation data and most altmetric data are obviously different, they share very similar tempos across scientific fields.

Furthermore, the variations do not only exist at the main subject field level, but also the research topic level. Within each subject field, different research topics also show various velocity patterns in receiving altmetric attention, both on fast sources or slow sources. This signifies the thematic dependency of users in following up-to-date research outputs around some topics, just like some certain research topics drive more social attention over others (Robinson-Garcia et al., 2019). Thus, further research should focus on identifying the main distinctive patterns of papers and research topics to determine their faster/slower reception across altmetric sources, and how different observation time windows, and the selection of different data sources, may affect real-time assessment in altmetric practice.

### 3.4.4 Limitations

The main limitation of this study lies in the precision of Crossref's DOI created date as the proxy of actual publication date of scientific papers. There might still be a small distance between the date on which a DOI was created and the paper was actually made publicly available, which could result in some inaccuracies in our results. Besides, as we mentioned in the data part, DOI created dates might be updated due to the change of DOI status, thereby causing the unreliable time intervals. One of the effects of these inaccuracies is that some papers may have altmetric event posted date even earlier than DOI created dates. Therefore, papers with such unexpected time intervals have been excluded from this study to lower the negative influence made by questionable DOI created dates. Future research should focus on refining accurate methods of identifying the effective publication date of research outputs.



As shown in this study, they have important repercussion to determine accurate time windows for altmetric research.

### 3.5 Conclusions

Several conclusions can be derived from this study. First, we conclude that not all altmetrics are fast and that they do not accumulate at the same speed, existing a fundamental differentiation between fast sources (e.g., Reddit, Twitter, news, Facebook, Google+, and blogs) and slow sources (e.g., policy documents, Q&A, peer review, Wikipedia, video, and F1000Prime). Another important conclusion of this study is that the accumulation velocity of different kinds of altmetric data varies across document types, subject fields, and research topics. The velocity of most altmetric data of Review papers is lower than that of Articles, while Editorial Material and Letter are generally the fastest document types in terms of altmetric reception. From the perspective of scientific fields, the velocity ranking of different data sources changes across subject fields, and most altmetric data sources show higher velocity values in the fields of PSE and LES, and lower in SSH and MCS. Finally, with regards to individual research topics, substantial differences in the velocity of reception of altmetric events across topics have been identified, even among topics within the same broader field. Such topical difference in velocity suggests that it is worth studying the underlying reasons (e.g., hotness, controversies, scientific debates, media coverage) of why some topics within the same research area do receive social (media) attention much faster than others.

### 3.6 Appendix

**Table 4.** Descriptive statistics of altmetric events and papers mentioned by different altmetric data sources across the five main subject fields

Data source	SSH		BHS		PSE		LES		MCS	
	NP	NE	NP	NE	NP	NE	NP	NE	NP	NE
Reddit	3,349	5,258	18,137	22,554	1,814	2,405	4,462	5,906	244	316
Twitter	220,454	1,681,028	1,166,583	7,593,795	253,035	715,855	300,754	1,951,100	31,053	103,535
News	22,043	107,446	120,697	565,035	28,232	119,520	31,965	162,326	1,433	5,777
Google+	9,556	20,722	43,036	101,916	7,245	20,207	11,614	28,980	1,045	4,415
Facebook	54,884	107,947	305,928	774,662	44,676	79,992	75,564	198,963	5,143	7,105
Blogs	27,055	49,873	91,945	161,242	21,337	35,628	40,614	76,231	1,207	2,037
F1000Prime	361	424	34,422	41,976	1,018	1,214	3,674	4,335	63	70
Video	1,041	1,357	7,271	10,860	1,748	2,332	1,865	2,688	250	352
Wikipedia	7,379	9,588	35,517	47,549	9,003	13,783	17,806	28,290	1,491	1,882
Peer review	10,123	11,019	8,505	16,609	614	1,815	822	1,913	114	131
Q&A	225	239	1,109	1,149	380	397	293	299	214	232
Policy documents	10,001	13,368	32,519	42,244	1,802	2,136	8,862	11,816	549	650

**Note:** NP refers to number of papers with corresponding altmetric data; NE refers to total number of corresponding altmetric events.

**Table 5.** Spearman correlations for the rankings of altmetric data sources from the perspectives of Velocity Index at the month time interval, altmetric half-life, and altmetric time delay (proportion of altmetric events with altmetric time delays no more than 10 days as the benchmark)

Indicator	Velocity Index	Altmetric half-life	Altmetric time delay
Velocity Index	1.000	0.979	0.944
Altmetric half-life		1.000	0.937
Altmetric time delay			1.000

**Table 6.** Descriptive statistics of altmetric events and papers mentioned by different altmetric data sources across the four main document types

Data source	Article		Review		Editorial Material		Letter	
	NP	NE	NP	NE	NP	NE	NP	NE
Reddit	25,330	32,919	2,271	3,056	2,083	3,415	510	601
Twitter	1,752,296	10,359,094	179,941	1,463,407	130,161	1,866,897	48,355	272,291
News	183,647	884,239	17,353	59,932	14,283	58,626	3,740	17,269
Google+	62,463	153,770	8,826	20,556	8,991	21,786	1,435	2,433
Facebook	424,922	1,009,847	52,502	140,350	44,091	144,962	11,116	22,996
Blogs	162,979	293,559	17,684	28,813	12,724	23,720	2,349	4,066
F1000Prime	36,381	44,471	2,970	3,359	373	406	223	238
Video	10,619	14,873	1,448	2,553	550	733	133	200
Wikipedia	58,005	82,082	12,844	18,609	2,802	3,717	687	836
Peer review	19,706	31,171	535	847	276	1,030	137	340
Q&A	1,849	1,926	355	374	100	106	22	22
Policy documents	46,796	60,821	6,386	8,708	2,134	2,818	613	752

**Note:** NP refers to number of papers with corresponding altmetric data; NE refers to total number of corresponding altmetric events.

**Table 7.** Spearman correlations for micro-level fields (based on Twitter and Wikipedia data)

	Twitter			Wikipedia		
	NP	NE	VI	NP	NE	VI
NP	1.000	0.949	0.031	NP	0.983	0.424
NE		1.000	-0.058	NE	1.000	0.432
VI			1.000	VI		1.000

**Note:** NP refers to number of papers with corresponding altmetric data in the micro-level field; NE refers to total number of corresponding altmetric events in the micro-level field; VI refers to the Velocity Index at the month time interval.