

# Ageing Factor: a Potential Altmetric for Observing Events and Attention Spans in Microblogs

Victoria Uren<sup>1</sup> and Aba-Sah Dadzie<sup>2</sup>

<sup>1</sup>Aston Business School, Aston University, The Aston Triangle, Birmingham, UK

<sup>2</sup>Department of Computer Science, 211 Portobello, The University of Sheffield, Sheffield, UK

We present an initial examination of the (alt)metric ageing factor to study posts in Twitter. Ageing factor was used to characterize a sample of tweets, which contained a variety of astronomical terms. It was found that ageing factor can detect topics that both cause people to retweet faster than baseline values, and topics that hold people's attention for longer than baseline values.

## 1 Introduction

Our long-term goal is to study public communication about science: understanding what the public thinks about scientific research is important for many reasons, from developing science policy, through making the case for technological developments and determining the impact of research, to simply being able to characterize the vibrant public discourse which marks a healthy society, in which science plays as important a part as politics, popular culture or the arts. The new media of the social web, which are open to all, offer a fresh insight into public opinion to supplement the surveys, and so forth, used in Public Understanding of Science (PUS) research [20].

Because of the scale of data available (Twitter claims there are 340M tweets per day<sup>1</sup>), quantitative metrics are needed to aggregate the contributions of many individuals. Informetrics research has developed and used quantitative measures to study scholarly communication in traditional media for decades. Informetric methods have been shown to transfer to communication on the web [1] and latterly social media / Web 2.0 [2], in which field they are coming to be known as altmetrics. Altmetrics adapt tried and tested informetric methods to the analysis of scientific communication in social media. Used along with natural language methods, such as clustering and summarization, we believe they have much to offer analysts. Our aim, in the experiments reported in this paper, was to study a metric called ageing factor as a means to characterize whether people's interest in discussion of scientific topics on Twitter is sustained or transient.

---

<sup>1</sup> <http://blog.twitter.com/2012/03/twitter-turns-six.html>

## 1.1 Science in Social Media

While reservations about the use of social media in formal work still exist [3], many scientific communities have embraced social media as a mode of communication [4]. In one survey [5], Twitter was one of the highest ranked services for semantic web researchers, but other communities favour different services, e.g., of informetrics researchers only 16% had Twitter accounts while 70% used LinkedIn [2]. Ponte & Simon [4] similarly found, in a survey of researchers from different fields, that nearly 18% used micro-blogging services such as Twitter, while 40% used science-based blogs and social networks. The ways in which scientific communities use social media vary. In computer science, delegates tweet or blog running commentary on conferences, opening up the proceedings to delegates in other rooms as well as colleagues who cannot attend [5] [6]. Whereas in biology, Mandavilli reports examples of intensive public critique of contentious articles [7].

However, scientists play only one part in the bigger picture of science communication - science organizations, journalists (both science and public-interest media), lobbyists and the general public also have important contributions to make. The current ethos of science communication, as discussed by Nisbet & Scheufele [8] and many others, advocates engaging the public in dialogue. Social media support open access to discussion, and for science organizations, Web 2.0 has become an essential part of their public relations operation. As one white paper boldly states “*the people formerly known as the audience are now at the center of media*” [9]. This is echoed by Eysenbach [10], who describes the “traces” left by both scientists and the lay public as they interact with information on the web. Individual scientists are increasingly aware of the public engagement agenda: Ponte and Simon [4] report scholars’ desire to make greater use of Web 2.0 methods for peer review and communication of research. Letierce [5] has shown that although researchers’ main motivation for using Twitter is to communicate with members of their own community (89%), some are also trying to reach general audiences (45.9%). Consequently, we argue that it is necessary to expand the scope of analysis of science communication in social media beyond the boundaries of scientists’ communication with each other to encompass wider public communication about science.

Compared to the numerous works on politics and marketing, relatively few studies exist about the public’s (as opposed to researchers’) scientific communication in microblogs. Hubmann-Haidvogel et al. [11] present a visualization tool using climate change posts as a use case, and Chew and Eysenbach [12], report the timeliness of social media for highlighting trends in the development of pandemics. In our own experience [13], the lack of research may be because scientific topics typically have few posts compared to current affairs, popular culture etc., so that simple trend spotting methods can be ineffective. Furthermore, there is a high level of noise, with many “scientific” search terms being used in non-scientific contexts [13]. Chew and Eysenbach [12], and Weller and Puschmann [14] also recognise the negative influence of noise in Twitter data. Chew and Eysenbach remark that “*spam and popular news articles that contained key phrases can influence search results and create peaks in activity that may not be reflective of the concept*” [12]. Both suggest the use of ad-

vanced natural language processing to support the identification of tweets containing scientific information and to filter out non-relevant or spurious use of domain-specific terms.

Weller et al. [15] have identified three classes of scientific microposts for their study of communication in scientific conferences: posts (tweets on Twitter) with links to scientific content, posts published by a scientist, and posts with a science related hashtag. To these we add posts that contain scientific terms used within a scientific context, as a more general definition of the sorts of posts we need to identify in order to analyse general public communication about science.

## 1.2 Ageing Factor

Our earlier work [13] which looked at changes in the usage of a sample of scientific terms over time, determined that the basic trend spotting approach, which looks for peaks of tweets occurring for a term on a timeline, is not effective for scientific data because the level of tweeting about science is very low compared to the level of tweets on popular culture. Furthermore, scientific terms are frequently used in non-scientific contexts. The combination of these two factors means that it is difficult to reliably identify peaks of discussion about science topics, because they are small and obscured by noise (irrelevant tweets using the same words). In order to study public communication about science in public media, we need an altmetric which is sensitive even on moderate sized samples of data, because big trend peaks, such as those seen in pandemics, will be relatively rare. We are interested in studying microblogging, for which time is a critical parameter, therefore, a metric which looks at temporal effects is needed.

The metric we test in the experiments presented here, called aging factor, is based on a well-established informetric measure. We follow the convention in which, by analogy, retweets are treated as citations (tweetations) by researchers working on altmetrics. For example, Eysenbach calculated a range of metrics for retweets of announcements by the Journal of Medical Internet Research about the publication of new papers [10]. He uses these to determine whether initial interest indicated by retweets can be correlated with later citation rates. The metrics Eysenbach used include tweet half-life defined as “*the point in time after publication by which half of all tweetations of that article within the first  $n$  days occur*”.

Half-life is similar to the ageing factor metric used in this paper with a number of important differences, which we consider make ageing factor more suitable for the kind of data we are studying. Half-life takes what is called the *diachronous view*, i.e. the metric observes a fixed set of documents, such as one year’s articles in a given journal or one year’s tweets from a particular organization). It is therefore useful for organizations which want to judge the impact of their own tweets and are monitoring the occurrence of their Twitter name in retweets on an on-going basis. This is necessary in order to harvest every tweetation of a particular tweet in the first  $n$  days. By contrast, we want to take a snapshot of general discussion on scientific topics in a given time period. This is what is termed the *synchronous view* and requires a metric which does not rely on the originating tweet being present in the sample. For this we

adapted Avremescu’s ageing factor measure as presented in [16], changing the counts of citations to retweets and changing time windows measured in years to windows measured in hours to suit the fast pace of communication on Twitter. Aging factor,  $AF$ , is defined below, where  $i$  is the cut-off time in hours,  $k$  is the number of retweets originating at least  $i$  hours ago and  $l$  is the number of retweets originating less than  $i$  hours ago.

$$AF = \sqrt[i]{\frac{k}{k+l}}$$

We examined two values of  $i$ ,  $i=1$  giving the one hour ageing factor ( $1hAF$ ), and  $i=24$  giving the 24 hour ageing factor ( $24hAF$ ). A convenient feature of  $1hAF$  is that it is simply the ratio of retweets in a sample that originated more than one hour after the original *createdAt* time over the total number of retweets in the sample; this makes it easy to understand. The  $24hAF$  is the 24th root of the similar ratio for a 24 hour cut-off. In either case,  $AF$  values are produced in the range 0-1 with higher values indicating more retweets originating after the cut-off.

## 2 Experiments

In the context of communication on Twitter, low values of  $AF$  would suggest a flurry of activity typical of a trending topic, such as might happen following the posting of tweets about an exciting topic. This might be a special event - in our experiments we looked at retweeting about meteor showers. This fits with the findings in [10] that even for the most interesting or highly cited articles, twitter citations quickly fall off soon after announcement of publication. On the other hand, we interpret high values of  $AF$  as an indication that people have shown sustained interest in a topic and continue to read and retweet posts for a long (in terms of Twitter) time after they appear. We argue that, for science, being able to show the public has a long attention span for ongoing developments in a field is as important as showing you can get a reaction to a hot news item. Based on these interpretations of what ageing factor could tell us, we made two assumptions about how to interpret ageing factor.

*Assumption 1:* ageing factors for topics which concern special events will be lower than suitable baselines.

*Assumption 2:* ageing factors which are higher than suitable baselines are associated with topics in which interest is sustained over time.

The question of what constitutes a “suitable” baseline therefore arises. Unfortunately no benchmark corpus of Twitter presently exists (plans for a corpus to be held at the United States Library of Congress are believed to be underway at the time of writing). In this experiment, we have taken a pragmatic approach. We know there is a high level of noise in the samples (see table 1) - the majority of posts for selected terms are *not* about science. Therefore, we take the superset of tweets as a sample of general usage of that term on Twitter at that time, and use ageing factors for these sets as our baselines.

We chose the topic of astronomy for the experiment, because it has an enthusiastic following of amateur stargazers who own their own telescopes and are interested in observing events in the night sky for themselves. This led us to believe that it would be possible to harvest posts from Twitter, which discussed astronomical events and might provide evidence about the validity of assumption 1. We collected data on two nights when meteor showers were expected. Our data collection harvested tweets containing a broad range of astronomical terms in order to compare ageing factors for meteor showers with those for other astronomical topics.

## 2.1 Experiment 1 – Geminid Meteor Shower

As in our previous experiment [13], the UNESCO Thesaurus<sup>2</sup> was used as a source of scientific terminology. The starting point was the terms under the subheading “*Astronomical Systems*”. We used 32 of 33 terms for the initial filtering, a mix of single words and two word phrases (see Table 1). The 33rd term, *Time*, produced an unacceptably high level of noise and was therefore removed from the set after an initial test run of the harvesting program. Using the public Twitter stream, two data sets were collected: a training set, comprising 8980 tweets collected between Dec 14th 2011 at 22:36 GMT and Dec 14th 2011 at 23:18 GMT, and a test set, comprising 81891 tweets collected between Dec 14th 2011 at 23:18 GMT and Dec 15th 2011 at 03:30. Dec 13-14th 2011 were the nights on which the annual Geminid meteor shower was expected to take place.

**Table 1.** Occurrence of retweets (RT) containing UNESCO terms in the training data, and the number of retweets judged to have scientific content (Sci) (where RT > 98).

UNESCO Term	RT	Sci	UNESCO Term	RT	Sci
Celestial bodies	0		Solar activity	0	
Cosmic matter	0		Solar disturbances	0	
Interstellar matter	0		Sunspots	0	
Galaxies	1		<b>Stars</b>	<b>174</b>	<b>7</b>
Stellar systems	0		Quasars	0	
Interstellar space	0		<b>Universe</b>	<b>99</b>	<b>5</b>
Black holes	2		Cosmos	5	
Meteorites	0		Astronomy	14	
Comets	9		Astrophysics	0	
Meteors	13		Gravitation	1	
Solar system	1		Celestial mechanics	0	
Planets	28		Cosmology	0	
<b>Earth</b>	<b>213</b>	<b>8</b>	Cosmogeny	0	
Satellites	1		<b>Space</b>	<b>166</b>	<b>27</b>
<b>Moon</b>	<b>241</b>	<b>9</b>	Outer space	3	
<b>Sun</b>	<b>565</b>	<b>6</b>	Space sciences	0	

<sup>2</sup> <http://www2.ulcc.ac.uk/unesco/thesaurus.htm>

For the AF calculations we needed to pick terms with reasonable levels of retweets. Our previous experiment [13] with a range of scientific terms, lead us to predict that many of the tweets that used UNESCO terms would not have scientific content. This proved to be true for the astronomical terms (see Table 1). Of the UNESCO terms identified in retweets in significant numbers, most are words used in daily life, which do not necessarily have an astronomical meaning: Sun, Moon, Stars etc. are used in a multitude of colloquial ways. The proportion of retweets judged to be scientific, from the six terms categorised, was 0.043. This is substantially lower than levels reported elsewhere: e.g., Mejova and Srinivasan [17] report 0.389 tweets judged to be topical, for a collection of tweets with the categories movies, music albums, smart phones, computer games, and restaurants and note that this is low compared to 0.60 for their sample of blogs. They identify “*the need for more precise retrieval strategies for Twitter*”. We suggest this is even more important for scientific communication.

We considered terms with 99 or more retweets in the test data to be worth considering in the experiment. These were classified by reading the tweets and making a judgement about whether or not they had scientific content. For example, “*when I was little I thought the sun and moon followed me around everywhere!*” was judged not scientific, whereas “*If the Sun exploded we wouldn't know for 8m 20s. Light & gravity take that long to reach us. Then we'd vaporize*” was judged scientific. Some retweets needed more research, for example, “*RT @VirtualAstro: Make sure You watch a Night with the stars with your illustrious leader on Sunday night :)*” was judged scientific after establishing the @VirtualAstro describes himself as “*The Basil Fawlty of Astronomy, Science, Nature and more.*”; he fits the profile of an amateur stargazer. As can be seen from Table 1, this classification exercise made clear the low level of scientific retweets.

Having identified UNESCO terms for which we could harvest reasonable levels of scientific retweets (albeit along with significant amounts of noise), we selected two baselines: the UNESCO thesaurus term *Space*, and a compound term we labelled *Astro*, which bundled together the UNESCO terms Earth, Moon, Sun, Stars, Universe and Space. In their raw condition, these both contain high levels of non-scientific usage of terms. Therefore, each can be considered as a sample of general use of those terms on Twitter at the sampled point in time. Subsets of the baseline, selected to filter out noise and represent scientific usage of the terms then had to be extracted.

### **Identifying Scientific Retweets.**

Ideally, we would use a natural language processing method to identify scientific use of the terms. However, these experiments have the objective of testing whether AF is an appropriate metric for studying scientific communication. Therefore, we took a simple approach to identifying scientific retweets using SQL queries to reduce the noise in samples by adding narrower terms. We accept that this approach, based on human interpretation of the language of the domain, has limitations for practical implementation on the large scale and will need to be replaced in future work with advanced NLP methods as advocated by [12].

The training data was a small enough sample to be analysed by hand. The retweets that had previously been identified as scientific were reviewed and topically related terms, which co-occurred with more than one UNESCO term, were identified. The terms were then sorted into topical queries. For example, terms related to space exploration equipment (e.g. Hubble) were in one set and terms which would be ambiguous (e.g. program) were all grouped together in another. One search (e.g. Space and Bodies+) added the names of the planets (plus Pluto) based on background knowledge.

Table 2 presents short form versions of the queries used, in which | represents OR, and standard parts of the query have been omitted for clarity. An actual search statement for the search *Space AND sci* for the training data set would read:

```
SELECT statusid, createdat, retweetid, retweetcreatedat
FROM 'twitter' WHERE (retweetid != "" AND batch =
"1323902158593" AND (text like '%space%') AND (text like
'%nasa%' OR text like '%science%' OR text like
'%station%' OR text like '%soyuz%' OR text like
'%satellite%' OR text like '%hubble%' OR text like
'%interstellar%' OR text like '%program%' OR text like
'%physics%' OR text like '%plane%' OR text like
'%voyager%')) ORDER BY retweetid ASC;
```

**Table 2.** Terms used in searches

Search label	Terms
Batch	Batch number only
Space	space
Space AND sci	Space AND (nasa science station soyuz satellite hubble interstellar program physics plane voyager)
Space AND gear	Space AND (nasa soyuz satellite spaceflight orbit hubble telescope spacecraft voyager)
Space AND amb	Space AND (agency program plane rock beam aircraft station aero astro launch deep outer travel)
Space AND bodies	Space AND (interstellar black hole comet moon geminid)
Space AND bodies+	Space AND Bodies AND (planet mercury venus mars jupiter saturn neptune uranus pluto)
Astro	Earth moon sun stars universe space
Astro AND events	Astro AND (meteor shooting star launch phaethon geminid)
Astro AND @	Astro AND (@universetoday @sciencemagazine @brainpicker @NASA_GoddardPix @doctorjeff @earthskyscience @anditurner @Sky_Safari, @VirtualAstro @NASA Ames @NASA_Lunar,)
Astro AND tech	Astro AND (light year astronomy galactic gravity astronaut)
Astro NOT meteor	Astro AND (nasa science astro hubble) NOT (meteor geminid shooting AND star)
Meteor	meteor geminid (shooting AND star)

Table 3 shows both 1hAF and 24hAF for the searches. The 24hAF values for this dataset were all in the range 0.8-0.95 (zero values were assigned when all retweets collected were within the 24 hour window), whereas 1hAF ranged from 0.25-0.65. In general, 24hAF tracks 1hAF. The culture of Twitter places high value on currency, and 24h is a long time for many Twitter users. 24hAF appears to be an insensitive metric and we used 1hAF only for the remainder of the experiments.

1hAF for the training data searches are often based on small samples of retweets, such that just two or three retweets can make a big difference to the 1hAF. For example, the 1hAF value for Space and gear of 0.73 in the training data is based on 11 retweets, and consequently the 0.4 difference compared to 1hAF in the test data (0.33) is unlikely to be significant. Therefore, the following observations look only at test data, and use Batch, Space and Astro test values as baselines. In this first experiment, we made a naïve interpretation of the results, simply looking for values of 1hAF that appeared high or low, then trying to explain them in terms of the content of retweets.

**Table 3.** Number of retweets, 1hAF and 24hAF values for experiment 1. Baselines for *Space Astro* and the whole batch are **0.34**, **0.37** and **0.37** respectively

Search Label	Training Data			Test Data		
	RT	1hAF	24hAF	RT	1hAF	24hAF
Batch	1526	0.31	0.88	15583	0.37	0.91
Space	166	0.43	0.90	2594	0.34	0.90
Space AND sci	18	0.44	0.89	396	0.33	0.85
Space AND gear	11	0.73	0.90	375	0.33	0.83
Space AND amb	25	0.48	0.87	487	0.37	0.87
Space AND bodies	7	0.57	0	42	<b>0.45</b>	0.88
Space AND bodies+	7	0.57	0	315	<b>0.26</b>	0.82
Astro	1425	0.31	0.88	14634	0.37	0.91
Astro AND events	25	0.44	0.93	597	<b>0.26</b>	0.86
Astro AND @	14	0.43	0.90	23	<b>0.65</b>	0
Astro AND tech	16	0.25	0	122	<b>0.58</b>	0.90
Astro NOT meteor	29	0.55	0.91	511	0.36	0.88
Meteor	27	0.15	0	364	<b>0.22</b>	0.84

For the Space set, two searches have 1hAF values that look different to the baselines: 1hAF for Space AND bodies is increased (0.45 compared to 0.34 in the Space baseline and 0.37 for the batch), 1hAF for Space AND bodies+ is decreased (0.26 compared to the same baselines). For the Astro set of queries, all queries except Astro NOT Meteor show differences when compared to the baseline Astro. Searches with increased 1hAF are: Astro AND @ (0.65 compared to 0.37 for the Astro baseline, but with only 23 retweets in the sample we should be cautious about its significance), and Astro AND tech (0.58). Astro AND events shows decreased 1hAF (0.26). The search Meteor was run to isolate tweets concerning the Geminid meteor shower. As can be clearly seen, it has a low value of 1hAF (0.22 compared to 0.37 for the batch).



Assumption 1 would associate low 1hAF with an event of some kind. The text of the retweets was examined and we found a high level of retweets of “NASA launch new rover to Mars” tweet in both Space AND bodies+ (which contains the term Mars) and Astro AND events (which contains the term launch). It seems the high level of retweeting of this post brings the 1hAF down for these two subsets. These initial results were sufficiently encouraging to make us want to study 1hAF in more detail with a larger dataset.

## 2.2 Experiment 2: Quadrantid Meteor Shower

A larger sample of the public Twitter stream was then collected. This was filtered using the same 32 UNESCO astronomy terms and covers the full 24 hours of the 3rd of January 2012. This was the night on which the annual Quadrantid meteor shower was expected and our aim was to see if 1hAF values were low for this event, as per assumption 1, and whether the time of day matters (it must be dark to see meteors).

Initially we filtered out subsets using the searches we had developed using the training data for the first experiment (see Table 2). The day was divided into four periods 0:00-5:59 GMT (labelled 6), 6:00-11:59 GMT (12), 12:00-17:59 GMT (18) and 18:00-23:59 (24). Figure 1 shows the 1hAF values for these searches.

The first observation is that although the batch baseline 1hAF is steady in the range 0.32-0.37 through the day, the other two baselines each have one quarter of the day when they are high or low (24 for Space and 12 for Astro).

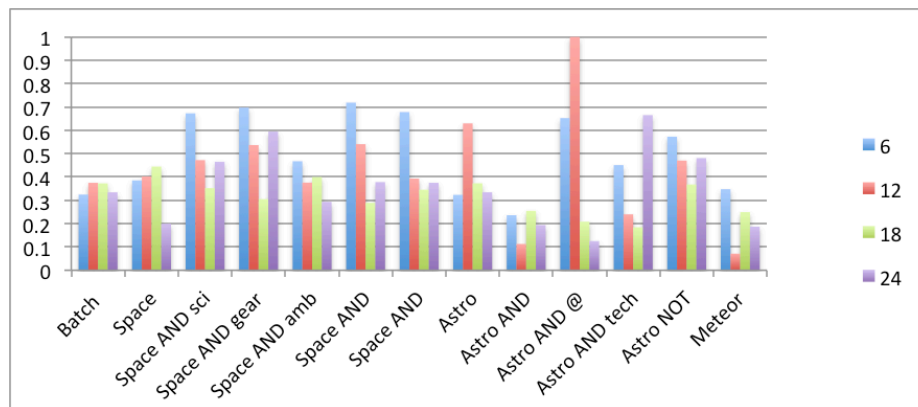


Fig. 1. 1hAF for 4\*6 hour periods over the course of 3rd Jan. 2012

As in the previous experiment, the 1hAF values for Astro are more variable than those for Space. For example, both the Astro AND Events 1hAF and the Meteor 1hAF (both of which contain the term *shooting star*) are relatively low compared to the Astro baselines, especially for period 12 (2nd quarter of the day). It would be easy, but incorrect, to infer that 1hAF had identified a flurry of retweets about the meteor shower. Examination of the text of posts in the second quarter show that of 275 total retweets 18 contain the term *quadrantid* while 213 contain the term *wish*.

There are various original tweets, but “@iQuoteFresh: #IfYouWereMine I’d stop wishing on 11:11’s, birthday candles, dandelions and shooting stars...Because I’d have my wish ...” is a typical example. Noise from non-scientific posts clearly remains an important issue, a fact underlined by examination of the high 1hAF values observed for four of the Space searches. These turned out to be due largely to retweets of variants of humorous posts on the lines of “Oh really? You need space? You might as well join NASA.”, for which the original tweets were more than one hour old.

The searches were based on the sample of training data collected about two weeks earlier around the Geminid meteor shower. It seems that even in this short time, the ways terms were being used had changed. We therefore took further steps to remove noise from our samples. Three astronomical events that took place around the 3<sup>rd</sup> of Jan. 2012 were used as background knowledge to add narrower terms to three of the original searches. The events were the Quadrantid meteor shower on the night of 3-4 Jan., the second of the twin Grail spacecraft moving into orbit around the Moon on the 2<sup>nd</sup> of Jan., and the proximity of the Moon and the planet Jupiter in the night sky on the 2<sup>nd</sup> of Jan. Searches excluding the event related terms were also conducted (see Table 4) as non-overlapping sets in order to assess the significance of results.

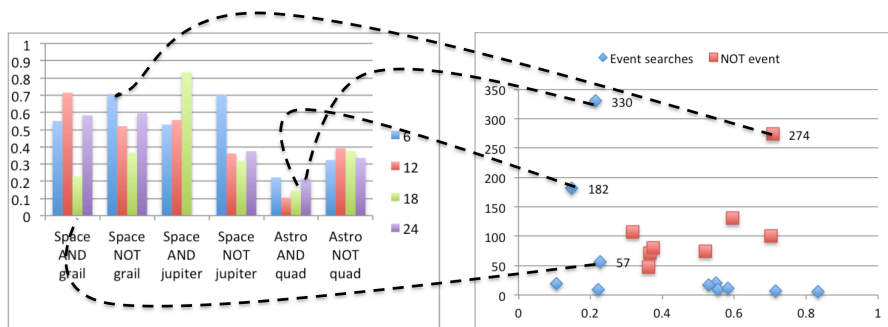
**Table 4.** Terms used in modified searches, refer to table 2 for details of original searches

Search label	Terms
Space AND grail	Space AND gear AND (grail lunar moon)
Space NOT grail	Space AND gear AND NOT (grail lunar moon)
Space AND jupiter	Space AND bodies+ AND (jupiter AND moon)
Space NOT jupiter	Space AND bodies+ AND NOT (jupiter AND moon)
Astro AND quad	Astro AND (quadrantid meteor shower)
Astro NOT quad	Astro AND NOT (quadrantid meteor shower)

Naïve interpretation of the results in Figure 2 (left) now seems promising. There is a low 1hAF for the Space and Grail search in the third quarter of the day (@18, 0.23), which contains retweets of posts about Grail tweeted by NASA and SETI in the first half of the (USA) working day. 1hAF for Space AND jupiter is generally high, and particularly in the third quarter (@18, 0.83), with retweets typically of links to pictures taken the previous night. Finally, the 1hAF values for Astro and quad are in the range 0.11-0.22, some of the lowest we saw, compared to between 0.32 and 0.39 for Astro AND NOT quad. These retweets are of messages from several sources reminding people to get up before dawn in order to see the meteors.

However, the differences in sample size between the searches about events and the exclusion searches we are using for comparison could be extreme: the largest (Astro NOT quad @24) contains 26327 retweets, the smallest (Space and jupiter @18) which generates 1hAF of 0.83 contains just 6 retweets. We therefore require a method of determining whether the 1hAF values we are seeing are significant or are merely the effect of small samples. To gain insight into the significance of results we used a funnel plot (see Figure 2 right). Funnel plots are employed in meta-analyses to detect publication bias and other biases [18]. The rationale is that small samples are ex-

pected to show higher variance. Therefore, if the measured values (in this case 1hAF) are plotted on the x axis and sample size on the y axis, then if all the data points come from the same population a triangular spread of points around the mean would be expected. Asymmetry in the plot suggests the data points may not all come from the same population. Figure 2 (right) presents a funnel plot for the events searches, excluding the points for Astro NOT quad, which have retweets in the thousands (plotting these would force us to use a log scale for the y axis making the funnel plot much harder to interpret).



**Fig. 2.** Left: Ageing Factor at one hour, 1hAF, for searches modified to detect specific astronomical events, measured for 4\*6 hour periods over the course of 3rd Jan. 2012. Right: Funnel plot of 1hAF vs number of retweets for the same data (excluding Astro NOT quad).

Based on the funnel plot, two of the Astro AND quad points still look interesting (@time 1hAF RT: @18 0.15 182, @24 0.22 330). Therefore, we can infer that retweeting activity around the Quadrantid meteor shower was significant in the hours of darkness for the UK and USA, where the largest concentrations of English speaking Twitter users in the northern hemisphere would be expected to be. Other data points which previously looked interesting, such as (@18 0.23 57) for Space and grail, on the funnel plot look like normal fluctuation within the expected variance of the population. This suggests that, at least on the 3<sup>rd</sup> of Jan 2012, the progress of Grail was not exciting the public to a significant degree.

A third data point also deserves investigation, that for Space NOT grail (@6 0.71 274). Examination showed that 216 of the retweets in this set contained the phrase “*join NASA*”, from retweets of the humorous posts we identified earlier. We infer that people had a long attention span for that particular joke.

### 3 Discussion and Future Directions

Our objective in carrying out the two exploratory experiments described here was not to rate jokes, but to test ageing factor as an altmetric for analysing scientific communication in social media, and specifically to test whether it can give any insights for the smaller datasets typical of scientific content. For ageing factor calculat-

ed with a window of one hour (1hAF), several of our naïve observations fitted both assumption 1 (ageing factors for topics which concern special or exciting events will be lower than suitable baselines) and assumption 2 (ageing factors which are higher than suitable baselines are associated with topics in which interest is sustained over time). However, when a funnel plot was used to identify data points which lay outside the area of expected variance, only three data points appear significant: two with low 1hAF for retweets about the Quadrantid meteor shower, and one with high 1hAF for a (non-scientific) humorous post. We conclude that further investigation of 1hAF would be worthwhile, but that interpretation of the metric without reference to sample size must be avoided.

In these exploratory experiments, topics were identified using SQL searches. It would be intriguing to compare topics with low and high AF to the statistical linguistics approach used by Hu et al. [19] to classify tweets into *episodic* and *steady* categories. Hu defines episodic tweets as “*tweets that respond specifically to the content in the segments of the events*” and steady tweets as those “*that respond generally about the events*”. Our intuitions about how ageing factor works suggest that episodic tweets would be more likely to show low 1hAF and steady tweets more likely to show high 1hAF.

24hAF gave similar values for all the searches in experiment 1. Therefore, we did not use 24hAF in experiment 2. As in [10], when analysing twitter data, especially about a specific event or topic, there is an upper limit beyond which relevant tweets tail off. In our experiments this was 24 hours, in [10], which examined scientific publications, with a significantly longer lifetime, this started at 7 days, and up to 10, for publications released within a three month window. Future studies will look at a wider range of time windows to see if they give more sensitive results than the 24h window and will apply the funnel plot technique to check significance. For example, a six hour window (6hAF) might be interesting to observe for studies like experiment 2 which divide the day into quarters.

The overall aim of this work is to contribute to the nascent development of methods and metrics that will support analysis of public online scientific communications. It is clear that the big issue in achieving this is the level of noise in samples coupled with low actual levels of scientific communication in social media. These combine to make it difficult to get big enough samples to get statistically significant results. As an additional problem, the usage of terms on Twitter clearly varies considerably even over a few weeks: our experiments used data collected only a few weeks apart, but the searches developed in experiment 1 proved useless in experiment 2. This may make it difficult to devise standard filters for on-going monitoring of scientific communication. Noise was addressed in this study by writing SQL queries to produce disambiguated subsets. However, in the future we will need to identify, and possibly develop, more subtle, NLP-based techniques for classifying tweets on science related topics. These techniques will need to adjust dynamically to pick up new topics as they arise.

As for future work, although our interest in ageing factor was stimulated by the small sample sizes we found for typical scientific topics, we are investigating the application of the technique to larger datasets and longer sequences of events. Furthermore, we have not explored the differences between types of participants. For

example, is there a difference between ageing factors observed for private individuals' tweets vs professional scientists' vs organizations'? Techniques for distinguishing these groups will be particularly important in achieving our overall goal of analysing public opinion about science.

## 4 Acknowledgements

A-S Dadzie is currently funded by the UK MRC project Time to Change (129941). Thanks to Mike Thelwall and the Statistical Cybermetrics Research Group for encouragement with pursuing the informetrics approach, and to Gary Simpson at Aston University for advice on the statistical issues.

## 5 References

1. M. Thelwall, "Webometrics," in *Annual Review of Information Science and Technology*, 2005, pp. 81-135.
2. J. Bar-Ilan, S. Haustein, I. Peters, J. Priem, H. Shema, and J. Terliesner, "Beyond citations: Scholars' visibility on the social Web," 2012.
3. J. Priem, H. A. Piwowar, and B. M. Hemminger, "Altmetrics in the wild: Using social media to explore scholarly impact," *arXiv12034745v1 csDL 20 Mar 2012*, vol. 1203.4745, pp. 1-23, 2012.
4. D. Ponte and J. Simon, "Scholarly Communication 2.0: Exploring Researchers' Opinions on Web 2.0 for Scientific Knowledge Creation, Evaluation and Dissemination," *Serials Review*, vol. 37, no. 3, pp. 149-156, 2011.
5. J. Letierce, A. Passant, J. Breslin, and S. Decker, "Understanding how Twitter is used to widely spread Scientific Messages," in *In Proceedings of the WebSci10 Extending the Frontiers of Society OnLine*, 2010.
6. K. Weller, E. Dröge, and C. Puschmann, "Citation Analysis in Twitter. Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences," in *MSM2011 - 1st Workshop on Making Sense of Microposts*, 2011, no. May, pp. 1-12.
7. A. Mandavilli, "Trial by Twitter," *Nature*, vol. 469, pp. 286-287, 2011.
8. D. A. Nisbet, Matthew C., Scheufele, "What's new for science communication? Promising directions and lingering distractions," *American Journal of Botany*, vol. 10, pp. 1767-1778, 2009.
9. J. Clark and P. Aufderheide, "Public Media 2.0: Dynamic, Engaged Publics," Center for Social Media, 2009.
10. G. Eysenbach, "Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact," *Journal of Medical Internet Research*, vol. 13, no. 4, p. e123, 2011.
11. A. Hubmann-Haidvogel, A. M. P. Brasoveanu, A. Scharl, M. Sabou, and S. Gindl, "Visualizing Contextual and Dynamic Features of Micropost Streams," in *Proceedings of the WWW'12 Workshop on "Making Sense of Microposts"*, , April 16, 2012. *CEUR Workshop Proceedings V838*, 2012.
12. C. Chew and G. Eysenbach, "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak," *PLoS ONE*, vol. 5, no. 11, p. 13, 2010.

13. V. Uren and A.-S. Dadzie, "Relative Trends in Scientific Terms on Twitter," in *altmetrics11: Tracking scholarly interest on the Web, Workshop at ACM WebSci'11*, 2011.
14. K. Weller and C. Puschmann, "Twitter for Scientific Communication: How Can Citations/References be Identified and Measured?" in *Proceedings of the ACM WebSci11*, 2011, pp. 1-4.
15. K. Weller, E. Dröge, and C. Puschmann, "Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences," in *Proceedings of the WWW'12 Workshop on "Making Sense of Microposts"*, , April 16, 2012. *CEUR Workshop Proceedings V838*, 2012.
16. L. Egghe and R. Rousseau, *Introduction to Informetrics*. Elsevier, 1990.
17. Y. Mejova and P. Srinivasan, "Crossing Media Streams with Sentiment: Domain Adaptation in Blogs, Reviews and Twitter," in *Sixth International AAI Conference on Weblogs and Social Media*, 2012, pp. 234-241.
18. M. E. Egger, G. Davey Smith, M. Schneider, and C. Minder, "Bias in meta-analysis detected by a simple, graphical test," *British Medical Journal*, 1997.
19. Y. Hu, A. John, D. e D. Seligmann, and F. Wang, "What Were the Tweets About? Topical Associations between Public Events and Twitter Feeds," in *Sixth International AAI Conference on Weblogs and Social Media*, 2012, pp. 154-161.
20. J. D. Miller, "Public Understanding of, and Attitudes toward, Scientific Research: What We Know and What We Need to Know", in *Public Understanding of Science*, v. 13, 2004, pp. 273-294.