

# Scientific Bloggers under the Altmetric Microscope

Isabella Peters, Lisa Beutelspacher, Parinaz Maghferat, Jens Terliesner

Heinrich Heine University Düsseldorf

Department of Information Science

Universitätsstr. 1, D-40225 Düsseldorf

{isabella.peters | lisa.beutelspacher | parinaz.maghferat | jens.terliesner}@hhu.de

## ABSTRACT

In this paper we present a preliminary altmetric study of scientific bloggers and how they use different social media (i.e. blogs, social bookmarking systems, and Twitter) for scholarly communication, information dissemination, and creation of visibility. We analyzed linking behavior in blog posts and tweets, number of comments assigned to blog posts and share of publications found in social bookmarking systems. Results show that heavy tweeting and blogging do not result in large numbers of followers and comments, tweets and blog posts contain lots of URLs and self-citations, and share of publications found in social bookmarking systems varies between different platforms.

## Keywords

Altmetrics, social bookmarking, blogs, Twitter, CiteULike, Bibsonomy, self-citation, scholarly communication.

## INTRODUCTION

Simply spoken, working of scholarly communication and impact of published articles are mainly dependent on visibility of authors and their publications as only awareness of papers induces readership (and along with it citations). With social media new tools for information dissemination and communication enter the scene which help researchers to address new audiences, publish study results, or guide readers to relevant publications. Impact of such social media-based scholarly practices are analyzed under the umbrella of “altmetrics” (Priem et al., 2010) which uses Web data (i.e., tweets, bookmarks, blog posts) and Web tools (i.e., social networks, or social bookmarking systems) to fully understand the characteristics of scholarly communication on the Web. Moreover, altmetrics credit scholarly activities (i.e., discussing or linking to journal articles) carried out on the Web which is not yet being acknowledged by traditional metrics of scientific impact (such as citation indicators). Altmetrics yield at complementing existing impact metrics instead of replacing

them. Although altmetrics is becoming popular, the critical discussion on adequate application scenarios, data sources, and indicators for measuring impact of authors, papers, or journals is still ongoing (Priem & Hemminger, 2010).

In order to establish and evaluate data sources or indicators for altmetrics we first have to understand how researchers use blogs, Twitter, or traditional publication lists in scholarly practice and how they raise readers’ awareness of their research. Among the first in this area is the work of Shema, Bar-Ilan, and Thelwall (2012) who studied the demography, topics, and disciplines of blogs and bloggers from [researchblogging.org](http://researchblogging.org), a blog platform for discussion of peer-reviewed research publications. In contrast to that, this preliminary study examines what characteristics of scholarly communication and practice are found in blogs, Twitter, and personal publication lists, which may be exploited for altmetrics. Three research questions are guiding our study:

- How productive are and how much discussion arises around scientific blogs (in terms of comments) and which linking behavior do analyzed blogs show (in terms of outgoing links and self-citations)?
- How many publications from self-maintained publication lists can be found in social bookmarking systems?
- How communicative are scientific bloggers on Twitter (in terms of tweets, retweets, and @-messages) and which linking behavior do analyzed tweets show (in terms of outgoing links and self-citations)?

## METHODS

It follows the description of methods used for data acquisition.

## Blogs

Scientific blogs are our key information source in this study as they determine the selection of analyzed authors. We used two blog portals, [www.scienceblogs.com](http://www.scienceblogs.com) and [www.scienceblogs.de](http://www.scienceblogs.de), which host blogs of scientific writers. The bloggers of these portals are not necessarily researchers employed by universities or other research institutions but are interested in science in general (e.g. science journalists). For our study we only considered such scientific writers who are affiliated with universities or other research institutions. This limitation resulted in 33 English-writing authors and eleven German-writing bloggers. Because some blogs are maintained from more than one author, we combined the authors of each blog and

This is the space reserved for copyright notices.

ASIST 2012, October 28-31, 2012, Baltimore, MD, USA.  
Copyright notice continues right here.

analyzed data of 30 English and ten German blogs indicated with respective author names. For all of the chosen blogs we manually collected blog name, name(s) of author(s), blog starting date, and number of blog posts, comments, and unique commentators<sup>1</sup>. The first entry of a blog marks its starting date. The number of comments is the sum of comments attached to each blog post. The number of unique commentators is the sum of individuals commenting each blog post. Moreover, we automatically extracted URLs of blogs posts to analyze linking behavior of bloggers. The analysis is based on 19,721 blog posts.

### Publication Lists and Social Bookmarking Services

We used [www.mendeley.com](http://www.mendeley.com), [www.bibsonomy.org](http://www.bibsonomy.org), and [www.citeulike.com](http://www.citeulike.com) for extracting social bookmarking data. Because chosen social bookmarking systems are mostly used as web-based reference managers typically research papers are saved there. Social bookmarking reflects via reader numbers how interested a community is in particular publications (Haustein, 2012). To gain article-based metrics as well as reading or bookmarking statistics we first had to search for official publication lists of chosen bloggers on institutional or private websites. Here we worked with individuals and not blogs. We considered publication lists found on institutional or private websites as gold standard as we assumed that scientific authors are strongly interested in maintaining their publication lists on a regular basis to be visible in the scientific community. However, some authors did not have any publication lists so that we had to create such lists from publications found in analyzed social bookmarking systems. We also cross-checked social bookmarking systems to find missing articles on publication lists and to determine the share of “official” papers (recorded in self-maintained publication lists) in social bookmarking systems. Authors without publication lists or articles saved in social bookmarking systems were excluded from analyses, which resulted in 936 publications from 41 authors.

### Twitter

Based on the personal information found in blogs we also searched for Twitter-accounts of analyzed bloggers. We found five non-ambiguous accounts for authors from scienceblogs.de and 29 accounts for bloggers from scienceblogs.com<sup>2</sup>. Via Twitter-API we collected starting date of Twitter account, number of tweets since starting date, and number of followers at download date. Numbers of @-messages and retweets were counted. For sake of simplicity we only considered the first 3 characters of tweets, meaning that when string “RT” was found in first 3 characters we counted this tweet as retweet (the same applies to @-messages). We also extracted and analyzed URLs published in tweets. Because Twitter’s API only allows download of about a user’s last 3,200 tweets our study is limited to this maximum number of tweets. This means that in some cases we could not analyze tweeting behavior since starting date

<sup>1</sup> Blog and social bookmarking data was collected between December 15, 2011 and January 15, 2012.

<sup>2</sup> Data from scienceblogs.de was collected on January 17, 2012 and from scienceblogs.com on December 14, 2011.

of Twitter account because some bloggers publish more than 3,200 tweets the month. The analysis is based on a set of 50,019 tweets.

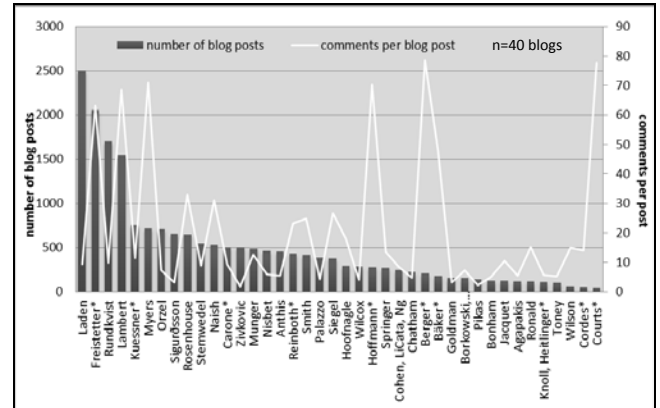


Figure 1. Number of blog posts and comments per blog post. Scienceblogs.de-authors are marked with \*.

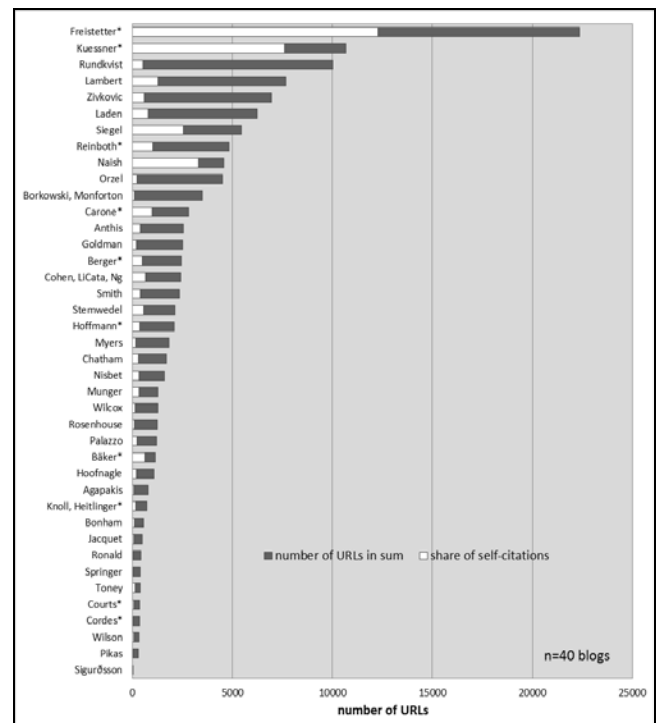


Figure 2. Number of URLs in blog posts and self-citations. Scienceblogs.de-authors are marked with \*.

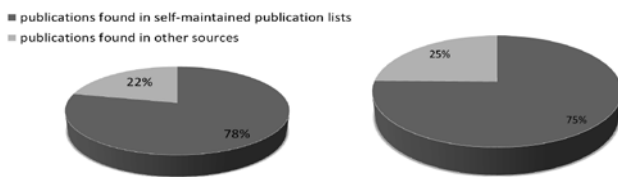
## RESULTS

In this section we present the results of our analyses.

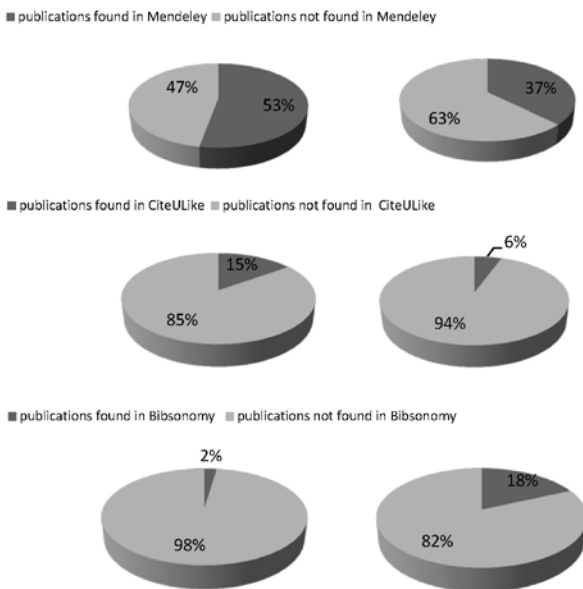
### Blogs

Figure 1 shows that analyzed blogs vary in the number of published blog posts and generated comments. As some blogs are maintained by more than one author, we grouped author names with their own blogs. Few blogs are very productive (e.g., 2,502 posts from *Laden* since October 7, 2007 and 2,056 posts from *Freistetter* since April 16, 2008) although blogs publish 11 posts per month on average. Interestingly, heavy publishing does not necessarily lead to lots of comments of

readers. The blog with most comments per blog post (*Berger* with 78.59 comments on average) has only published 212 blog posts since July 1, 2008. The use of URLs is common practice in blogs as shown in Figure 2. Especially heavy bloggers are distributing URLs via blog posts (e.g., *Lambert*). However, linking behavior of bloggers differs fundamentally between blogs and shows power law-like characteristics. Some bloggers mainly link to other sources (e.g., *Rundkvist*) outside scienceblogs.com whereas others often give reference to their own blog posts or to blog posts also published at scienceblogs.de or scienceblogs.com (e.g., *Freistetter*). We consider latter behavior as “self-citation” in Figure 2. Tables 1 and 2 show the ten most linked to top-level domains from scienceblogs.com and scienceblogs.de. Independent of the language used on the blog portal other social media platforms, such as *Wikipedia*, *YouTube*, or *Twitter*, and news platforms (e.g. *New York Times* or *Spiegel*) are mostly referenced in blog posts – besides self-reference to scienceblogs.de or scienceblogs.com which both are the top-link destinations.



**Figure 3. Articles in self-maintained publication lists and other sources. Left: scienceblogs.com; right: scienceblogs.de. n=41 authors.**

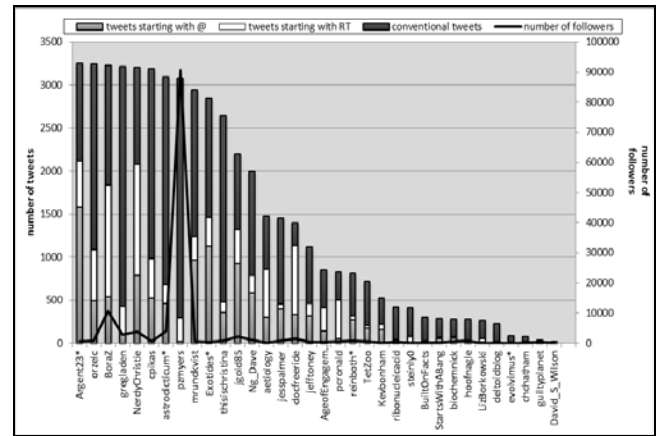


**Figure 4. Articles in social bookmarking systems. Left: scienceblogs.com; right: scienceblogs.de. n=41 authors.**

### Social Bookmarking Systems

Surprisingly, it turned out that self-maintained publication lists are not complete or updated frequently from authors. Figure 3 shows that 22% of publications from authors of scienceblogs.com and 25% of publications from authors of scienceblogs.de are only findable via author name searches in

other sources (i.e., Scopus, CiteULike, Mendeley, and BibSonomy). The detailed analyses of the three social bookmarking systems showed that for both author groups Mendeley is the service where most of the publications can be found (see Figure 4). The reason for such good coverage may lie in Mendeley’s broad topical scope and the variety of users. On the other hand, BibSonomy is more popular in Germany which might be the reason for its increased usage.



**Figure 5. Tweeting behavior of authors. scienceblogs.de-authors are marked with \*. n=34 twitterers.**

### Twitter

Figure 5 displays tweeting behavior of bloggers (indicated with their Twitter-names). The most active twitterer is *Argent23* (author name: *Knoll*), who is among the five least active bloggers (see Figure 1). Although this may lead to the conclusion that active twitterers cannot simultaneously be active bloggers, *gregladen* proves the opposite (author name: *Laden*). The distribution of tweets is not as skewed as the distribution of blog posts maybe because the costs of publishing tweets are less than this of blog posts and therefore induce increased tweeting. In the analyzed data set the maximum number of tweets per day is 54.97, the minimum is 0.07. Heavy tweeting does not result in large follower numbers (see *NerdyChristie*). However, tweeting behavior between authors differs greatly. Mostly, twitterers distribute conventional tweets not intended to reflect conversations with followees or followers. On the other hand, some twitterers use their tweets to directly communicate with readers via @-messages or cite people they follow via retweets (RTs) and re-distribute their publications. In terms of followers it is the same like tweet frequency: many @-messages or RTs do not necessarily result in many followers (see *NerdyChristie* or *pzmyers*). Almost every third tweet (2.89) contains a URL. The ten most referenced top-level domains are displayed in Tables 1 and 2. Authors promote their own blog posts (or this of colleagues) on scienceblogs.com and scienceblogs.de via Twitter. Commonly used are also Twitter-centred services which allow for sharing photos and videos (e.g., *Twitpic*). As such Twitter seems to be used as unidirectional channel pushing information to the tweetosphere.

### DISCUSSION

The findings presented in this paper indicate that the different tools (publication lists, social bookmarking systems, blogs, and

scienceblogs.de bloggers					
outgoing links from blogposts			outgoing links from tweets		
link destination	absolute	%	link destination	absolute	%
scienceblogs.de	25019	52.45	scienceblogs.de	1095	30.85
de.wikipedia.org	3709	7.78	twitpic.com	88	2.48
en.wikipedia.org	882	1.85	youtube.com	73	2.06
amazon.de	517	1.08	derstandard.at	56	1.58
flattr.com	393	0.82	spiegel.de	46	1.30
esowatch.com	342	0.72	forschungs-blog.de	40	1.13
arxiv.org	304	0.64	de.wikipedia.org	39	1.10
spiegel.de	238	0.50	yfrog.com	38	1.07
youtube.com	225	0.47	boingboing.net	32	0.90
twitter.com	202	0.42	flickr.com	32	0.90

**Table 1. Top 10 link destinations from blogs and tweets from scienceblogs.de.**

scienceblogs.com bloggers					
outgoing links from blogposts			outgoing links from tweets		
link destination	absolute	%	link destination	absolute	%
scienceblogs.com	18041	23.40	scienceblogs.com	3522	25.52
technorati.com	3008	3.90	freethoughtblogs.com	1608	11.65
bloggar.se	2873	3.73	blogs.scientificamerican.com	930	6.74
en.wikipedia.org	2430	3.15	friendfeed.com	340	2.46
delicious.com	2044	2.65	bigthink.com	297	2.15
amazon.com	1088	1.41	yfrog.com	289	2.09
nytimes.com	746	0.97	popperfont.net	265	1.92
researchblogging.org	632	0.82	boingboing.net	176	1.28
del.icio.us	625	0.81	twitpic.com	175	1.27
ncbi.nlm.nih.gov	519	0.67	scientificamerican.com	145	1.05

**Table 2. Top 10 link destinations from blogs and tweets from scienceblogs.com.**

Twitter) serve different purposes. We assumed that scientific bloggers use blogs, Twitter, and personal publication lists to raise awareness for their work and to improve scholarly communication. Our results proved correct for blogs and Twitter as in both strong promotion of own writings is observable in linking behavior of authors. But there is still potential for updating publication lists and building presence in social bookmarking systems to guide readers to publications. As in traditional bibliometrics here we also undergo the problems of author disambiguation which may result in incorrect publication and reader numbers. Moreover, this initial study suffers from small numbers of analyzed data, but demonstrates in which direction studies on alternative ways of scholarly communication and altmetrics may go.

## CONCLUSION

We showed how scientific authors from the blog portals scienceblogs.com and scienceblogs.de use publication lists, blogs, and Twitter for information sharing and how often they were engaged in discussions via comments or @-messages or followed by Twitter-users. It was also presented how visible are scientific publications of analyzed authors in social bookmarking systems. When searching for complete publication lists of authors several sources should be combined as self-maintained publication lists are sometimes patchy. Mendeley is the most popular social bookmarking service by now and should therefore be fed with own publications to make them more visible to the community. On the other hand, usage of social bookmarking systems must further increase (especially on the reader side) to gain reliable reader or bookmarking statistics in order to meaningfully complement traditional bibliometric indicators. Blogs and tweets are frequently used for promotion of own blog posts (i.e. self-citation) or other websites providing further information on the topic (e.g. *Wikipedia*) indicated by the large numbers of URLs published in them. Heavy blogging and tweeting do not necessarily result in large numbers of comments or followers. The same applies to extensive conversational use of tweets (i.e. @-messages and retweets).

Further research should comprise detailed analyses of blog posts', tweets' and scientific articles' content revealing whether bloggers blog and tweet about the same topics they study professionally. It is also open which indicators are appropriate for measuring impact of authors on the blogosphere or tweetsphere and how they should be transferred into the field of scientometrics.

## ACKNOWLEDGMENTS

Presented work is associated with the Junior Researchers Group "Science and the Internet" funded by the Strategic Research Fund of the Heinrich Heine University.

## REFERENCES

- Haustein, S. (2012). *Multidimensional Journal Evaluation. Analyzing Scientific Periodicals beyond the Impact Factor*. Berlin: De Gruyter Saur.
- Priem, J., & Hemminger, B.M. (2010). *Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web, First Monday, 15(7), 7-5*.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A Manifesto*. Retrieved June 10, 2012 from <http://altmetrics.org/manifesto>.
- Shema, H.; Bar-Ilan, J., & Thelwall, M. (2012) Research blogs and the discussion of scholarly information, *PLoS One, 7(5)*, doi: 10.1371/journal.pone.0035869.