Blog Citations as Indicators of the Societal Impact of Research: Content Analysis of Social Sciences Blogs

Hamid R. Jamali*, Dariush Alimohammadi**

ARTICLE INFO

ABSTRACT

This article analyzes motivations behind social sciences blog posts citing journal articles in order to find out whether blog citations are good indicators for the societal impact or benefits of research. A random sample of 300 social sciences blog posts (out of 1,233 blog posts) from ResearchBlogging.org published between 01/01/2012 to 18/06/2014 were subjected to content analysis. The 300 blog posts had 472 references including 424 journal articles from 269 different journals. Sixty-one (22.68%) of all cited journals were from the social sciences and most of the journals with high frequency were highly cited general science journals such as PNAS and Science. Seventy-five percent of all journals were referenced only once. The average age of articles cited at the time of citation was 5.8 years. Discussion and criticism were the two main categories of motivations. Overall, the study shows the potential of blog citations as an altmetric measure and as a proxy for assessing the research impact. A considerable number of citation motivations in blogs such as disputing a belief, suggesting policies, providing a solution to a problem, reacting to media, criticism and the like seemed to support gaining societal benefits. Societal benefits are considered as helping stimulate new approaches to social issues, or informing public debate and policy-making. Lower self-citation (compared to some other altmetric measures such as tweets) and the fact that blogging involves generating content (i.e. an intellectual process) give them an advantage for altmetrics. However, limitations and contextual issues such as disciplinary differences and low uptake of altmetrics, in general, in scholarly communication should not be ignored when using blogs as a data source for altmetrics.

1. Introduction

According to Kostoff (1995, p. 869), in research sponsoring organizations, the selection of research programs is based on the quality of science and potential contribution to the organization's mission. In other words, funders expect outcomes and impact from research projects. In the last two decades,

 * PhD, Associate professor, Department of Library and Information Studies, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran (h.jamali@gmail.com) (Corresponding author)
** Lecturer, Department of Library and Information Studies, Faculty of Psychology and Education, Kharazmi University, Tehran, Iran (webliographer@gmail.com)

the shift in research evaluation has moved from output to outcome and impact. However, measuring outcome and impact of research is not as easy as measuring output, and it is a challenging task. Several methods have been used for the research impact assessment and one can find a good review of them in Boaz, Fitzpatrick, and Shaw (2008). Also several frameworks have been developed for research impact assessment, especially in health sciences (for a review see Brutscher, Wooding, and Grant 2008, Marjanovic, Hanney, and Wooding 2009). However, most of these methods and frameworks are mainly based on qualitative methods and data such as interviews and so on. This makes implementing these methods and frameworks expensive and difficult. As a result, scientometricians focus on quantitative metrics such as citation-based metrics as proxies for impact. For example if an article is cited in another article we can assume that the cited article has had some scientific impact, or if it is cited in a legislative or government document, it has had impact on policy-making or legislation. However, citation is limited to specific types of use (research purposes, mainly) and specific types of users (authors only).

With the growth of Web 2.0 and social media, altmetrics has emerged as an alternative set of metrics. Some of these alternative metrics include numbers of social bookmarks, usage data or numbers of downloads, numbers of links received, and numbers of microblogging instances (e.g. Twitter).

One of these metrics is the number of times a journal article is mentioned or cited in weblog posts. The assumption is that the more people write about an article in their blogs, the more likely that article is to be popular, interesting, and probably useful at least for the community to which bloggers belong. This could be taken as an indication or proxy for the societal impact of a study. Societal impact or benefits refers "to the contribution of research to the social capital of a nation, in stimulating new approaches to social issues, or in informing public debate and policy-making" (Bornman, 2012, p. 673). Therefore, one can argue that if blog posts use research studies to help promote and enhance the discussion of social issues and public debates, they are increasing the societal impact of those studies.

During the last few years an increasing number of studies (such as Priem et al., 2012; and Priem, Piwowar, and Hemminger, 2011) have dealt with altmetrics. However, few studies have dealt with the context in which articles are mentioned or cited in blogs and the actual meaning of a citation to an article in a blog post. The only content analysis study in this area is the study by Shema, Bar-Ilan, and Thelwall (2014b) on health research blogs.

This article aims to analyze motivations behind social sciences blog posts citing journal articles in order to find out whether blog citations of scholarly journal articles are good indicators for the societal impact of research. We chose social sciences because, according to Costas, Zahedi, and Wouters (2014), publications from the social sciences, humanities, and the medical and life sciences show the highest presence of altmetrics, indicating their potential value and interest for these fields. Additionally, research in social science is supposed to have an impact on and be beneficial for society as society is the main subject of study in social sciences. With the aid of content analysis of blog posts citing journal articles, we hope to shed some light on the context in which journal articles are used in weblogs. Specifically, we seek to answer the following questions:

• What are the motivations for social sciences blog posts citing journal articles?

• What journal titles are frequently cited in social sciences blog posts?
• What are the subject categories of journals cited in social science blogs?
• What is the average age of the articles cited in social sciences blogs?

2. Literature review

Blogs, as a kind of social media, are used by researchers to support the research lifecycle (Nicholas & Rowlands, 2011). Blogs were studied as a source of information before they were considered a source of data for altmetrics. For instance Bar-Ilan (Bar-Ilan, 2005) studied library and information science blogs as sources of information, and Kim (2009) did a content analysis of 485 cancer blog posts to see if they are good sources of information for patients.

After blogs and microblogs were introduced as a source of data for altmetrics, there were a few quantitative studies to show if they correlate with the well-established citation metrics. Eysenbach (2011) studied tweets citing medical articles and showed that tweets can predict highly cited articles within the first 3 days of article publication. He maintained that social media activity either increases citations or reflects the underlying qualities of the article that also predict citations, but the true use of these metrics is to measure the distinct concept of social impact. Shuai, Pepe, and Bollen (2012) studied the correlation between citations, downloads, and Twitter mentions for a set of arXiv preprints. They found that the volume of Twitter mentions was statistically correlated with arXiv downloads and early citations just months after the publication of a preprint, with a possible bias that favors highly mentioned articles. Thelwall et al (2013a) compared 11 altmetrics (including blog mentions) with Web of Science citations for a set of PubMed articles. Their results provided strong evidence that six of the eleven altmetrics (tweets, Facebook wall posts, research highlights, blog mentions, mainstream media mentions, and forum posts) associate with citation counts, at least in medical and biological sciences and for articles with at least one altmetric mention. However, they admitted that their methods did not shed light on the magnitude of any correlation between the altmetrics and citations (i.e., the correlation effect size was unknown). They also concluded that the coverage of all the altmetrics except for Twitter seems to be low and so it is not clear if they are prevalent enough to be useful in practice.

Shema, Bar-Ilan, and Thelwall (2014a) showed a correlation between blog citations and future citations. An analysis of blog posts aggregated at ResearchBlogging.org, which cited peer-reviewed articles published in 2009 and 2010 revealed statistically significant evidence that articles receiving blog citations close to their publication time receive more journal citations later on than the articles in the same journal published in the same year that did not receive such blog citations. Another study on the correlation of altmetrics and citations (Costas, Zahedi, and Wouters, 2014) confirmed previous claims of positive correlations but relatively weak, thus supporting the idea that altmetrics do not reflect the same concept of impact as citations. They also maintained that altmetric counts do not always present a better filtering of highly cited publications than journal citation scores. Altmetrics scores (particularly mentions in blogs) are able to identify highly cited publications with higher levels of precision than journal citation scores (JCS), but they have a lower level of recall.

Finally, a recent meta-analysis of altmetric studies by Bornman (2015) revealed that the correlation with traditional citations for micro-blogging counts is negligible (pooled r = 0.003) and for blog counts it is small (pooled r = 0.12).

Once the correlation of blog citations and traditional metrics were partially confirmed, some studies focused on the context in which a journal article is mentioned or cited in blogs and microblogs. Analysis of a sample of 270 tweets by Thelwall et al. (2013b) showed that 42% of the tweets echoed an article title and another 41% presented a brief summary. One reason for summarising an article seemed to be to translate it for a general audience. Few tweets explicitly praised an article and none were critical. Most tweets did not directly refer to the article author, but some did, and others were clearly self-citations. They maintained that "tweets containing links to scholarly articles generally provide little more than publicity, and so whilst tweet counts may provide evidence of the popularity of an article, the contents of the tweets themselves are unlikely to give deep insights into scientists' reactions to publications, except perhaps in special cases". Shema, Bar-Ilan, and Thelwal (2014b) used content analysis to discover health research bloggers' motivations for citing journal articles. A random sample of 391 blog posts from ResearchBlogging.org was analyzed using a coding scheme they developed which included ten categories (discussion, criticism, advice, trigger, extensions, self, controversy, data, ethics, and other). Multidisciplinary and general medical journals were the main publications referred. Few self-citations were identified, and 90% of posts provided a general report of the research cited. About 30% of bloggers also gave health advices. The same number tended to criticize findings of studies cited. Results could be interpreted as the motivation for widening health debates and popularizing medical issues on the one hand, and on the other hand, gaining more readers.

Past studies also tell us about the characteristics of research blogs and bloggers. Shema, Bar-Ilan, and Thelwall (2012) studied 135 bloggers of ResearchBlogging.org and their blogs. The study revealed that Life Science blogs were the most popular in the sample, followed by the Psychology, Psychiatry, and Neurosciences & Behavioral Science blogs. Blogs about Social Sciences, Humanities, Computer Science, and Engineering were the least represented in the sample. About two-thirds of the blogs had one male author, 18% had one female author, 5% had two male authors, and 4% had one female and one male author. Out of the 126 blogs in the sample, ninety (72%) had at least one active, public Twitter account. References in blog posts included 913 references to articles of 429 journals, 9 references to articles uploaded to arxiv.org, 3 references to conference proceedings, and 2 references to books. Science, Nature, and PNAS were the most cited journals. The science bloggers were highly educated. In general, most of the bloggers (59%) were either students or researchers in an academic institute. Less than a third (30%) were not affiliated with an academic institute, and 10% remained unknown. Holmberg & Thelwall (2014) studied tweeting behavior of researchers in ten disciplines and found subject differences. For example researchers in digital human-ities and cognitive science used Twitter more for conversations, while researchers in economics shared the most links. Another study on tweeting behavior by Haustein et al. (2014a) focused on 37 astrophysicists and showed that different user groups had different tweeting behaviors.

Overall, the number of content analysis and qualitative studies on research blogs are limited

and only one study (Shema, Bar-Ilan, and Thelwall, 2014b) exists in this area. Most of the altmetric studies also focus on health or basic sciences and rarely deal with social sciences.

3. Methods

ResearchBlogging.org (launched in 2007) is an aggregator of blog posts referencing peer-reviewed research in a structured manner, similar to the references in scholarly discourse. Research bloggers can register their blogs with ResearchBlogging, however, editors of ResearchBlogging make sure that the posts of participating blogs are about research and are not commercial in nature. ResearchBlogging shows a snippet of blog posts with the cited articles and readers can click and read the full post on the blog sites. Blogs registered in ResearchBlogging are categorized by topic (e.g. social sciences, health etc.)

We used the advanced search of ResearchBlogging, restricting the topic to Social Sciences, and searched for all blog posts between 01/01/2012 to 18/06/2014. This time frame was chosen because we wanted to study recent blog posts, and therefore, we chose a 30-month time frame up to the time of data collection, which was mid-2014. In, total 1,233 blog posts were retrieved. The results were presented in 62 pages (20 posts in each page). Using the Cochrane (1977) formula, the sample size with 95% confidence level and confidence interval of 5 was calculated to be 297 blog posts. However we made the number round and added 3 to the needed sample size to make it 300. A systematic random sampling method was used. One of the authors went through all pages of the results and chose every fourth blog post and collected the information of 300 blog posts including URL of the post, cited references, blog title, and post date. The 300 posts belonged to 103 unique blogs. Using the URL of the post, every single blog post was visited for further data collection and content analysis. The dates of all cited articles were recorded along with the cited journal titles. Scopus second level subject categories were used to check the subject of the cited journals.

Citation motivations can be done using a few different methods including surveys of authors (e.g. Kapseon, 2004), critical incident interviews (e.g. Thornley et al., 2015), and content analysis of citing documents (e.g. Case and Higgins, 2000). A review of citation motivation studies (Tohidinasab & Jamali, 2013) has shown that content analysis is a common method for such studies. This study uses content analysis, especially as it is difficult to access bloggers located around the world for survey or interview purposes.

For content analysis we used the categories in the coding scheme developed by Shema, Bar-Ilan, and Thelwall (2014b) as the base. However, since their scheme was developed for analyzing health blogs and was partly inapplicable to social sciences blogs, we inductively created a few other categories. For example, we did not find any post related to the 'ethics' category or any post related to 'advocacy against a certain treatment' motivation. Therefore, we did not use these. Instead we came across motivations such as introducing or criticizing a book, and so we added these to the original schema. The lead author did the content analysis and the second author also coded 30 blog posts (10% of the sample) to check the inter-coder reliability which turned out to be 83%. The measure for inter-coder reliability was the similarity and overlap between the two coders' analyses.

There are nine categories which include 35 motivations. To find out about the motivation for citing an article, we read all blog posts and looked at the context in which articles were cited. Some of the blog posts had a list of structured references but did not have any in-text citation so it was not clear what the role of the specific paper was in the content of the blog post. In-text citation did not have to be an 'Author-Date' sign of the reference. Some bloggers used phrases such as 'a new study⋯' or 'a recent article⋯' in the post as a reference to the paper in their reference list. There were 14 of those references. It should be noted that some of the blog posts could be categorized under more than one category, but only the main category was used in the analysis. The nine categories and sub-categories (motivations) are presented below:

- *Discussion*: consideration and examination of an issue.
  - Discussing a social phenomenon
  - Praising and giving detailed description of an article
  - Review of current knowledge about a subject
- *Criticism*: finding fault with a research-related issue.
  - Disputing a common belief
  - Raising methodological issues
  - Overall criticism
  - Criticism of a practice
  - Criticism of a book
  - Criticism of a third work
  - Criticism of a public policy
  - Criticizing an article's conclusion/findings
  - Criticizing an article for missing references
  - Criticizing an article for not sharing data
- *Advice*: recommending actions for the readers.
  - Providing advice
  - Providing a solution to a social problem
  - Suggesting policy
  - Recommending a behavior
- *Trigger*: a direct stimulus for post writing that was mentioned in the post.
  - Neutral detailed description of an article
  - Reaction to another blog post
  - Reaction to a topic in news/media
  - Announcing the publication of a new article
  - Invitation for further discussion in the comments
  - Answering a common question
  - Introducing a book
- *Extension*: suggesting possibilities beyond the post's scope.
  - Suggesting further reading or other material
  - Suggestions for possible future research

- *Self* : the bloggers added post content which was specifically related to them.
  - ◦ Self-citation
  - ◦ Sharing personal experience
  - ◦ Used as the basis of own work
- *Controversy*: discussing controversy; explaining and/or discussing disagreements.
- *Data*: providing data and facts with practical implications; background factual information about the blog post subject.
- *Unclear*: the motivation for citation was not clear.

In this study we only analyzed structured references, i.e. references that are presented in a structured manner similar to the references in journal articles. We did not include unstructured references (also called blog mentions) in the study.

4. Findings

From 300 blog posts, 145 (48%) belonged to 2012, and 39 (13%) posts were from 2014. The lower number of posts from 2014 was because we only included the first six months of 2014 in our analysis (See Figure 1.) The 300 blog posts had 472 references (on average, 1.57 references per blog post). Links to other blog posts or news websites or other general websites were not included in this count. From 472 references, 424 were journal articles (on average, 1.41 references to journal articles per blog post). The rest of references were books (35), book chapters (5), theses (3), conference papers (4), and e-print (1).

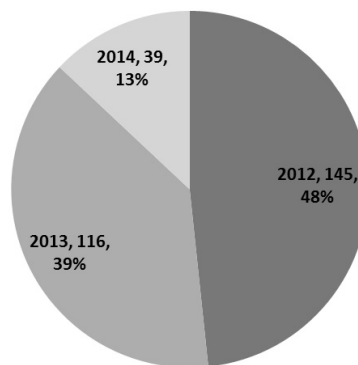**Fig. 1.** Number of blog posts analyzed by publication year

4.1. Motivations for citing journal articles

Table 1 presents the apparent motivations for citing journal articles in blog posts. The most frequent (38, 12.67%) motivation was to 'neutrally present details of a study'. Blog posts in this sub-category did not criticize the article; they simply presented a great deal of details about method,

findings, and other parts of an article. Some of them were similar to press releases.

The second most frequent (35, 11.57%) motivation was to 'discuss a social phenomenon'. Discussion sometimes (23) aimed at explaining a phenomenon i.e. giving reason why a phenomenon happens. For example a blog post cited research to explain why restaurants in California require Spanish language skills for those working in the kitchen.

> "Thinking sociologically about this, though, perhaps speaking Spanish *is* a requirement for back of the house work in Los Angeles. Waldinger (1998) suggests that as immigrants of a particular group get concentrated in particular sectors of particular industries, employers prefer not to hire workers from other groups because they have trouble fitting in" (Calvin N. Ho, 2013)

In some other cases (8) of discussions, bloggers discussed pros and cons of phenomena (not the reason for their occurrence). For instance a blog post discussed the phenomenon or tradition of fox hunting in the UK by delineating on its advantages and disadvantages from the eyes of its supporters and opponents. In some other cases (4), the discussions took the form of simple description and bloggers simply introduced and gave examples of occurrences of a phenomenon.

'Providing advice' was the motivation in twenty-seven (9%) cases. For example, a blog post cited research to advise people to spend their bonus on other people if they wanted to be happier. (References after each quotation include blogger's name and year of the post).

> "In fact, the research indicates that spending the money on someone other than yourself actually leads to *greater* happiness. More than that, it can lead to your *improved* performance at work." (Stanghini, 2013)

Blogs in the sub-category of 'disputing a common belief' (24, 8%) used references to dispute a common belief. In the example below, the blogger cited a reference to argue against the common belief that members of the Mafia are psychopaths.

> "The view that the Mafia is an organization of especially ruthless psychopaths is wrong – in fact, members of 'Cosa Nostra' have lower psychopathic traits than other criminals. That's according to a new study from Italian researchers Schimmenti and colleagues, who, appropriately enough, are based in Sicily, the Mafia's birthplace." (Neuroskeptic, 2014)

'Praising and giving detailed description of an article' accounted for 19 (6.33%) blog posts. Posts in this sub-category, similar to the first sub-category, gave a lot of details about the study (method, findings etc.), but they were not neutral. They positively praised and admired the study that was being discussed.

'Raising methodological issues' (12, 4%) was used when blogger criticized a study for its methodology. For instance in the following post, the blogger criticized the sampling method of the study.

"I'm actually surprised they found the findings they did; not because I disagree with them, but because of the sample they drew from. I would expect those who participate in a home-based, voluntary program first thing in the morning to be more confident and healthier than those who do not participate." (Mr Eepidemiology, 2012)

Blogs in the sub-category of 'Review of current knowledge about a subject' (12, 4%) simply presented an overview of the current knowledge about a topic. 'Discussing controversy' (11, 3.67%) means discussing disagreement.

"Earlier this year a major study of almost one million Canadian children found that rates of diagnosed ADHD- as well as use of ADHD medications like Ritalin- were higher in kids born later in the year···. Now a new paper published in the Journal of Attention Disorders rejects the immaturity hypothesis: Is the Diagnosis of ADHD Influenced by Time of Entry to School?" (Neuroskeptic, 2012)

'Overall criticism' was given to 11 (3.67%) posts where bloggers generally criticized a work with no particular emphasis on its findings or methods. For instance, the following blogger criticized different aspects of an article.

"It begins with a statement so out of place, or out of date, that one wonders whether the article is even worth reading ···" (Campbell, 2012)

Nine posts were written as a 'Reaction to another blog post' and eight were the reactions of the blogger to a topic in news or media. Eight posts cited articles to suggest further reading or recommend other material.

In 'Criticism of a practice' (7, 2.33%) bloggers criticized a current practice among researchers or the public. For instance a blogger criticized the reliance on self-reported data in psychological studies. Seven blog posts used data and facts presented in an article in order to discuss a topic. Six bloggers cited articles to support their arguments in reviewing a book, and five posts used articles for supporting their arguments in criticizing the idea or opinion of a third work or person. Five of the posts were dedicated to criticism of a public policy and articles were cited in this context. The other categories all had a frequency of four or fewer; they included motivations such as announcing the publication of a new article, providing a solution to a social problem, and self-citation.

Table 1. Frequency of motivations of bloggers for citing journal articles

Category	Motivation (sub-category)	N	%
Discussion	Discussing a social phenomenon	35	11.67
Discussion	Praising and giving detailed description of an article	19	6.33
Discussion	Review of current knowledge about a subject	12	4
	Overall discussion frequency	66	22
Criticism	Disputing a common belief	24	8
Criticism	Raising methodological issues	12	4
Criticism	Overall criticism	11	3.67
Criticism	Criticism of a practice	7	2.33
Criticism	Criticism of a book	6	2
Criticism	Criticism of a third work	5	1.67
Criticism	Criticism of a public policy	5	1.67
Criticism	Criticizing an article's conclusion/findings	3	1
Criticism	Criticizing an article for missing references	2	0.67
Criticism	Criticizing an article for not sharing data	1	0.33
	Overall criticism frequency	76	25.34
Advice	Providing advice	27	9
Advice	Providing a solution to a social problem	3	1
Advice	Suggesting policy	3	1
Advice	Recommending a behavior	2	0.67
	Overall advice frequency	35	11.67
Trigger	Neutral detailed description of an article	38	12.67
Trigger	Reaction to another blog post	9	3
Trigger	Reaction to a topic in news/media	8	2.67
Trigger	Announcing the publication of a new article	5	1.67
Trigger	Invitation for further discussion in the comments	4	1.33
Trigger	Answering a common question	3	1
Trigger	Introducing a book	2	0.67
	Overall trigger frequency	69	23
Extension	Suggesting further reading or other material	8	2.67
Extension	Suggestions for possible future research	7	2.33
	Overall extension frequency	15	5
Self	Self-citation	3	1
Self	Sharing personal experience	3	1
Self	Used as the basis of own work	1	0.33
	Overall self frequency	7	2.33
Controversy	Discussing controversy	11	3.67
Data	Providing data and facts	7	2.33
Other	Unclear	14	4.67
	Total	300	100

4.2. Cited journals

The 424 journal articles cited in blogs appeared in 269 unique journal titles. The list of journal titles with the frequency of 4 or more is presented in Table 2. It should be noted that higher frequency does not always mean diverse usage of that journal in blog posts, since one post might cite several articles of a single journal. For instance, all four references to the '*Australian Veterinary Journal*' were in a single blog post. The subject category of journals is presented in the last column and it is clear that none of the top 13 journals are from the category of Social Sciences.

The content analysis showed that although social science blogs cite journal articles from domains such as medicine or neuroscience, they are cited because the research presented in those articles has social aspect. The list of Social Sciences journals with the frequency of two or more is given in Table 3.

Figure 2 shows the cumulative percentage of unique journals (green line) and the cumulative percentage of citations (blue line). The horizontal axis represents the number of references to journals. From 269 journals, 203 journals (75.5% of all journals) were referenced only once, 35 journals (13%) were referenced twice, and 18 journals (6.7%) appeared among the references three times. In other words, journals with 2 or more references accounted for 24.5% of all cited journals and 87.7% of references. Most of the Social Sciences journals were among those journals that have been cited only once.

Table 2. List of journals cited 4 or more times in the blog posts

	Journal title	Frequency	%	Subject category
1	PLoS ONE	19	4.48	Agricultural and Biological Sciences
2	PNAS	18	4.24	General
3	Science	9	2.12	General
4	Animal Behavior	7	1.65	Agricultural and Biological Sciences
5	Nature	7	1.65	General
6	Psychological Science	7	1.65	Psychology
7	Journal of Human Evolution	5	1.18	Agricultural and Biological Sciences
8	Journal of The Interactions of People & Animals	5	1.18	Agricultural and Biological Sciences
9	Applied Animal Behavior Science	4	0.94	Veterinary
10	Australian Veterinary Journal	4	0.94	Veterinary
11	New England Journal of Medicine	4	0.94	Medicine
12	Philosophical Transactions of the Royal Society B: Biological Sciences	4	0.94	Biochemistry, Genetics and Molecular Biology
13	Trends in Cognitive Sciences	4	0.94	Neuroscience

Table 3. Social Sciences journals cited in blog posts

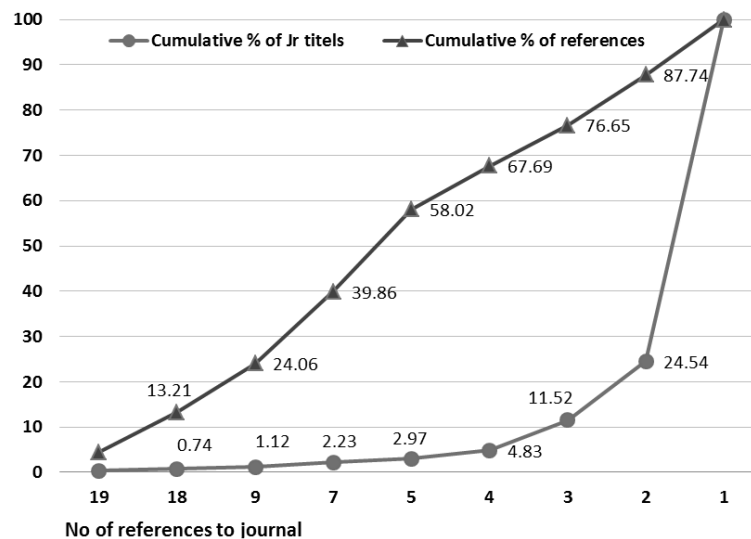|	Journal title	Frequency	%
1	Current Anthropology	3	0.71
2	Public Understanding of Science	3	0.71
3	American Anthropologist	2	0.47
4	Death Studies	2	0.47
5	Evolutionary Anthropology: Issues, News, and Reviews	2	0.47
6	Journal of Politics	2	0.47
7	Tobacco Control	2	0.47

Fig. 2. Percentage of journal titles and references by number of references to journals

4.3. Subject categories of cited journals

Table 4 shows the subject categories of the referenced journals. Sixty-one (22.68%) of all journals cited were from the category of social sciences, 38 from psychology, and 28 from humanities. Although journals from social sciences did not receive a high frequency of citation each, they were diverse and a large number of journals (61) received low numbers of citation each.

Table 4. Subject areas of journal articles cited in the blog posts

Subject area	No of journal	% of journals
Social sciences	61	22.68
Psychology	38	14.13
Medicine	29	10.78

Subject area	No of journal	% of journals
Neuroscience	29	10.78
Humanities	28	10.41
Business, Management, and Accounting	12	4.46
Economics, Econometrics, and Finance	12	4.46
Veterinary	11	4.09
Biochemistry, Genetics, and Molecular Biology	10	3.72
Environmental Science	10	3.72
Agricultural and biological sciences	8	2.97
General	6	2.23
Decision Sciences	3	1.12
Computer sciences	2	0.74
Earth and Planetary Sciences	2	0.74
Energy	2	0.74
Nursing	2	0.74
Chemistry	1	0.37
Dentistry	1	0.37
Engineering	1	0.37
Immunology and Microbiology	1	0.37
Total	269	100

4.4. Age of cited articles

In terms of the age of the cited articles, the oldest cited paper was published in 1951. Table 5 shows the distribution of the cited articles by their publication year for blog posts 2012-2014. Figure 3 shows all of the cited papers since 1951 by their publication year.

Taking into account the time period between the date of blog posts and the publication year of cited articles, the age of each reference was calculated and on average, references were 5.8 years old. Removing outliers (papers published during the 1950s and 1960s), the average age was 5.27. In this calculation, if a blog post in 2012 cited a paper published in 2012, the age was zero and if the cited paper was published in 2011, the age was 1 year. If the cited paper was 2013 (as some journals are published ahead of time) the age was -1. There were 8 references with the age of minus one (-1) (2012 blog posts that cited papers published in 2013). There were also 150 cases where blog posts and cited articles were both from the same year. 13.7% of articles cited in blogs were at least 14 years old (this figure was 7% for health blogs (Shema, Bar-Ilan, and Thelwall, 2014b)).

Table 5. Blog post references by cited article publication year (rows) and blog post publication year (columns)

| 2012 Year	n	%	2013 Year	n	%	2014 Year	n	%
2013	8	3.69	2014	0	0	2015	0	0
2012	66	30.41	2013	52	33.99	2014	19	35.19
2011	34	15.67	2012	25	16.34	2013	4	7.41
2010	17	7.83	2011	10	6.54	2012	5	9.26
2009	13	5.99	2010	6	3.92	2011	2	3.70
2008	11	5.07	2009	10	6.54	2010	3	5.56
2006-2007	12	5.53	2007-2008	12	7.84	2009	4	7.41
2004-2005	6	2.76	2005-2006	8	5.23	2004-2008	6	11.11
2000-2003	15	6.91	2000-2004	12	7.84	2000-2004	4	7.41
1996-1999	10	4.61	1996-1999	5	3.27	1996-1999	1	1.85
1990-1995	7	3.23	1990-1995	5	3.27	1990-1995	3	5.56
1980-1989	6	2.76	1980-1989	2	1.31	1980-1989	2	3.70
1971-1979	8	3.69	1971-1979	5	3.27	1971-1979	1	1.85
1955-1970	4	1.84	1951-1970	1	0.65	1951-1970	0	0
Total	217	100		153	100		54	100
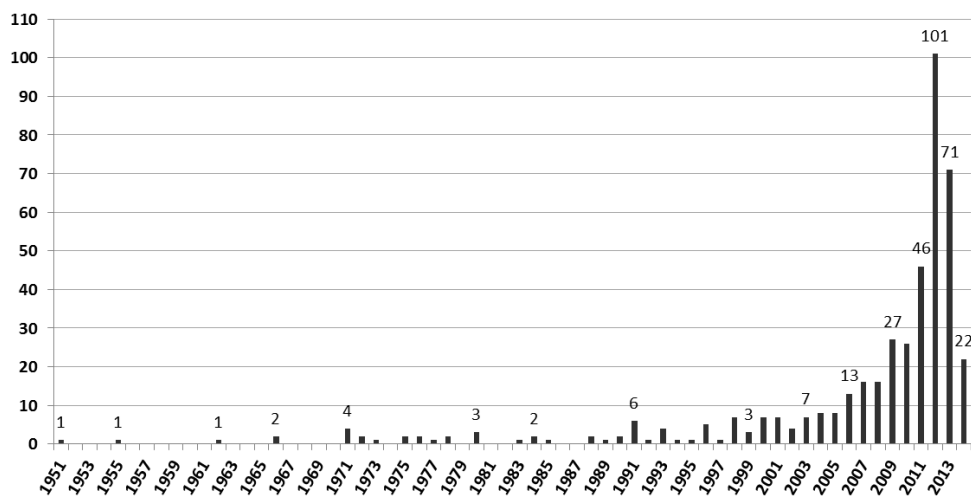
Fig. 3. Number of articles cited by publication year

5. Discussion and conclusion

This article was a content analysis of a sample of 300 posts (2012-2014) from 103 blogs registered as social sciences blogs in ResearchBlogging.org. Going back to the main aim of the study, we can conclude that blogs can indeed serve as proxies for the societal impact of research. However,

their limitations and context should be taken into account. Looking at the bloggers' motivations for citing journal articles (Table 1, above), one can see that a considerable number of them serve purposes that could be linked to societal benefits. As Bornman (2012) defined it, societal impact is partly about stimulating new approaches to social issues, or informing public debate and policy-making. Discussion and criticism were the two main categories of motivations. Motivations such as disputing a belief, suggesting policies, providing a solution to a problem, reacting to media, criticism and the like, all mean that blogs serve that definition of societal benefits. Moreover, the discourse that takes place on blogs as a form of social media also helps improve the societal benefits of research.

The other reason blogs are potentially useful as an indicator of research impact is that unlike some other altmetric measures where users do not engage in any form of intellectual process and simply do activities such as clicking, liking, linking, or bookmarking, bloggers engage in an intellectual process, i.e. generating content. Even tweets, although including contents, have not proved to be very useful as a data source for altmetrics (Thelwall et al., 2013b). However, since blog posts are usually longer and providing a context, they are more helpful as a source of data for altmetrics. Moreover, the low rate of self-citation makes blogs even more attractive as a proxy measure for research impact. We also know from past studies (see Groth & Gurney, 2010) that scientific discourse on blogs is more immediate, contextually relevant, and has a larger non-technical focus than the academic literature.

Comparing the motivations for citations in blogs with the citation motivations of authors of journal articles (of which a review and classification was presented in Tohidinasab and Jamali (2013)) one can see the differences are probably more than the similarities. The reason might be that the objective and function of a blog post is different from that of a journal article. Blogs are mainly discussion about others' works while articles are meant to present authors' own research. These differences imply that citations in blogs and in journal articles capture and manifest two different kinds of research impact and are complementary to each other. Citation in journal articles are considered as signs of scientific impact of research as they show the benefit one paper has had for the scientific community. We argued that blog citations, on the other hand, could show societal impact.

However, as mentioned above, limitations and contextual issues should not be ignored when using blogs as a data source for altmetrics. One issue is disciplinary differences. Comparing the findings about social sciences blogs with those of health blogs (Shema, Bar-Ilan, and Thelwall, 2014), some subject differences are revealed. For instance, self-citation is lower in social science blogs (1%) than in health blogs (3.1%); the number of posts including a general discussion of the issue covered in the article is much lower in social sciences compared to health sciences blogs; books are more frequently cited in social sciences blogs than in health blogs; and on average, articles cited in social sciences are older compared to health blogs. Other studies (e.g. Holmberg & Thelwall, 2014) have also confirmed the existence of subject differences.

Another issue is that altmetrics themselves are still in their infancy. In spite of the increasing number of studies on altmetrics, the scientific evidence about their reliability and validity is not yet robust. Previous studies such as those by Thelwall et al. (2013b), Costas, Zahedi, and Wouters (2014), and Haustein et al. (2014b) generally confirmed low uptake of altmetrics in scholarly

publications. Therefore, any use of blogs (and all other altmetric data) for research assessment purposes should be done cautiously, until they become prevalent enough to be useful in practice.

Overall, social science blogs are rather subject-oriented than article oriented. This means a considerable number of blog posts are not driven or triggered simply by writing about an article. In other words, similar to references in journal articles, bloggers write about their subject of interest and in that process use and cite articles they find relevant or helpful for their discourse. Although, some of the blog posts are triggered by the publication of new articles, this is not the case in many posts. The fact that the average age of articles cited in blogs are 5.8 years confirms this.

Finally, a word of caution about the limitations of this study, and future directions. One limitation was that the sample included only English blogs, and only blogs that are registered in ResearchBlogging. This might not be representative of all research-related blogs in social sciences. ResearchBlogging is very limited in nature, and studies should also be done on other blogs. For example, it would be helpful to see how research is reflected in blogs by non-researchers and the public.

We did not do any systematic analysis of the characteristics of journal articles cited in blog posts. However, while reading the list of the articles it was clear that at least most of the article titles sounded attractive. Titles such as 'The Neurobiology of Love', 'Has the tobacco industry evaded the FDA's ban on Light cigarette descriptors?', 'You see, the ends don't justify the means: visual imagery and moral judgment', and 'The illusion of saving face: how people symbolically cope with embarrassment'. Characteristics of article titles cited in blogs should be studied to see if they are any different from those not cited in blogs. Article title typing presented in Jamali and Nikzad (2011) might be helpful for this matter. Surveys also can be done to find out on what basis bloggers decide to blog about an article or to cite one.

Acknowledgement

References

Bar-Ilan, J. (2005). Information hub blogs. *Journal of Information Science*, 31(4), 297-307.

Boaz, A., Fitzpatrick, S., & Shaw, B. (2008). Assessing the impact of research on policy: A review of the literature for a project on bridging research and policy through outcome evaluation. *Policy Studies Institute & King's College London, February*. Available at <http://www.psi.org.uk/pdf/2008/bridgingproject_report.pdf> Retrieved 2014.08.20.

Bornman, L. (2012). Measuring the societal impact of research. *EMBO Reports*, 13(8), 673-676.

Bornmann, L. (2015). Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics. *Scientometrics*. In press, DOI: 10.1007/s11192-015-1565-y

Brutscher, P-B., Wooding, S., & Grant, J. (2008). *Health Research Evaluation Frameworks: an*

international comparison. California, Santa Monica: RAND Corporation.

Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51(7), 635-645.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York: John Wiley & Sons.

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *arXiv preprint arXiv:1401.4321*. Available at <http://arxiv.org/ftp/arxiv/papers/1401/1401.4321.pdf> Retrieved 2014.08.20.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.

Groth, P., & Gurney, T. (2010). Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study. In: *Proceedings of the WebSci10*, Raleigh, NC, US. Available: Available at <http://journal.webscience.org/308> Retrieved 2015.03.30.

Haustein, S., Bowman, T. D., Holmberg, K., Peters, I., & Larivière, V. (2014a). Astrophysicists on Twitter: An in-depth analysis of tweeting and scientific publication behavior. *Aslib Journal of Information Management*, 66(3), 279-296.

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014b). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4), 656-669.

Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, 101(2), 1027-1042.

Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88(2), 653-661.

Kapseon, K. (2004). The motivation for citing specific references by social scientists in Korea: The phenomenon of co-existing references. *Scientometrics*, 59(1), 79-93.

Kim, S. (2009). Content analysis of cancer blog posts. *Journal of the Medical Library Association: JMLA*, 97(4), 260-266.

Kostoff, R. N. (1995). Research requirements for research impact assessment. *Research Policy*, 24(6), 869-882.

Marjanovic, S., Hanney, S., & Wooding, S. (2009). *A historical reflection on research evaluation studies, their recurrent themes and challenges*. California, Santa Monica: RAND Europe Technical Report. Available at <http://www.fgcasal.org/aeets/Documentos/evaluacion%20de%20investigacion-RAND2009.pdf> Retrieved 2014.08.20.

Nicholas, D., & Rowlands, I. (2011). Social media use in the research workflow. *Information Services and Use*, 31(1), 61-83.

Priem, J., Parra, C., Piwowar, H., Groth, P., & Waagmeester, A. (2012). *Uncovering impacts: a case study in using altmetrics tools*. Available at <http://ceur-ws.org/Vol-903/paper-05.pdf> Retrieved 2014.08.20.

Priem, J., Piwowar, H. A., & Hemminger, B. H. (2011). *Altmetrics in the wild: An exploratory study of impact metrics based on social media*. Available at <http://jasonpriem.org/self-archived/PLoS-altmetrics-sigmetrics11-abstract.pdf> Retrieved 2014.08.20.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research blogs and the discussion of scholarly information. *PLoS ONE*, 7(5), e35869.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014a). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, 65(5), 1018-1027.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014b). How is Research Blogged? A Content Analysis Approach. *Journal of the Association for Information Science and Technology*, (in press). doi: 10.1002/asi.23239.

Shuai, X., Pepe, A., & Bollen, J. (2012). How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations. *PLoS ONE*, 7: e47523. doi: 10.1371/journal.pone.0047523

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013a). Do altmetrics work? Twitter and ten other social web services. *PloS one*, 8(5), e64841. 10.1371/journal.pone.0064841

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013b). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 17, 1-8.

Thornley, C., Watkinson, A., Nicholas, D., Volentine, R., Jamali, H. R., Herman, E., Allard, S., Levine, K. J., & Tenopir, C. (2015). The role of tust and authority in the citation behaviour of researchers. *Information Research*, 20(1), accepted in press.

Tohidinasab, F., & Jamali, H. R. (2013). Why and Where Wikipedia is cited in journal articles? *Journal of Scientometric Research*, 2(3), 231-238.