Do Tweets Indicate Scholarly Communication?

Ashraf Maleki

Master graduate at Scientometrics, University of Tehran, Iran

Although some characteristics of tweets linking to academic articles have been recognized in previous studies, less research yet has been conducted to characterize features of the Twitter users involved (Maleki, 2014). Twitter is widely used by non-academic audience, however, no large scale statistical evidence seems to be found as to whether these tweets are significantly used for informal scholarly communications or information-sharing between users outside academia. Twitter users may tweet articles for publicity, discussion, value of research findings, or even simply for amusing title of an article which are forms of communication independent from direct scholarly usage of research (Thelwall, 2014). However, the purpose of current study is to investigate various academic and non-academic Twitter users’ contribution to tweets and the significance of the relationship that might exist between each one’s contribution and conventional citations to articles.

Method

To address research issue, Altmetric.com was used as the aggregator of tweet mentions since it identifies tweets to articles based on disambiguated links of academic articles in various web resources and provides tweet counts based on unique contribution of each tweeter per article. To harness a reasonably large extent of tweeted articles, the popular interface of Altmetric Explorer was first used and a filter was created for articles at any time receiving at least one tweet. Although not comprehensive, the resultant export returned 99,965 papers out of which 81,247 papers were true matches. Based on the index of Scopus journals, 61,802 tweeted papers was determined using source title match which enabled to identify articles across four major subjects. Then, a full database of Scopus journal articles published between 2011 and 2014 was acquired by August 2016 which helped to identify 6,445 tweeted articles with their Scopus citations using DOIs (Digital Object Identifier).

Twitter users or tweeters are categorised by Altmetric.com in four cohorts of ‘Members of the Public’, ‘Researchers’, ‘Practitioners (doctors, other healthcare professionals)’, and ‘Science communicators (journalists, bloggers, editors)’. Information for these cohorts are already gathered and categorised by Altmetric.com from details provided by users in their profiles. Many profiles, meanwhile, are ambiguous or unknown such as automatized robots (Adie & Roe, 2013) and may automatically be categorised in the cohort of ‘Members of the Public’, as well as any other profile that exclude words that can explicitly identify them in one of the other three cohorts (Altmetric Support, 2015). So a major limitation of the study is that the cohort of ‘Members of the Public’ includes all accounts that could not be classified and thus does not practically seem to represent only the public audience.

To collect Twitter user’s information a software was developed to harness data through Altmetric.com API (Application Programming Interface) using articles’ Altmetric identifier. A sample of tweeted articles consisting of 6,445 articles published between 2011 and 2014 and 2,533 random tweeted articles in other years was selected to gather tweeter details. Given that the tweet counts provided show unique number of user accounts involved in tweets, there was discrepancy between sum of tweeters at cohorts and total unique tweeters posting each article. These included 474 (5%) papers whose cohorts were undercounted, perhaps indicating newer tweeters’ contribution that were in the process of classification; and only 27 (0.5%) articles whose tweeters at cohorts slightly exceeded total
unique tweeters, probably as false results. Thus, for conducting the correlation analysis an extra category of ‘unknown’ tweeters was created to represent uncategorised tweeters.

**Findings**
The primary results demonstrated that with 96% (8,755 articles), members of the public tweeted the highest number of articles. Researchers, meanwhile, were the second largest cohort to tweet articles (with approximately 69%, 6,251 articles), followed by Practitioners (49%, 4,426 articles) and Science Communicators (48%, 4,407 articles). Consistent with previous findings, more recent articles have found to be tweeted more frequently (Haustein et al., 2014). Based on articles published in 2011 to 2014, both median and mean tweeters across four broader fields have almost doubled over four years (except in Physical Science where median tweets almost tripled from 4 in 2011 to 12.5 in 2014), whereas Scopus citations turned one forth. Health Sciences, meanwhile, showed the largest median tweeter counts increasing from 11 in 2011 to 20 in 2014, while its median Scopus citation dropped from 36 to 9. The major cause of rise in tweets was the additions to the cohort of the public audience across all fields, particularly in Health Sciences where median members of the public had improved from 8 to 13, as well as researchers and practitioners (both from 1 to 2). As given in Table 1, the Spearman’s correlation between cohorts showed that tweets from public audience was moderately associated with researchers (r=0.726), practitioners (r=0.678), and science communicators (r=0.656). As the most common category, it possibly shows that cohort of members of the public involves unclassified tweeters which actually relate to the other three cohorts.

**Table 2.** Spearman’s correlation coefficients between frequencies of tweeters involved from various cohorts in the whole sample including 8,978 articles

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>Researchers</th>
<th>Members of the Public</th>
<th>Practitioners</th>
<th>Science Communicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Members of the Public</td>
<td>.726**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Practitioners</td>
<td>.514**</td>
<td>.678**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Science Communicators</td>
<td>.603**</td>
<td>.656**</td>
<td>.516**</td>
<td>-</td>
</tr>
<tr>
<td>Unknown Tweeters</td>
<td>.218**</td>
<td>.272**</td>
<td>.223**</td>
<td>.229**</td>
</tr>
</tbody>
</table>

** Significant at p < 0.01.

Low negative (Thelwall et al., 2013a) and positive (Haustein et al., 2014) correlations between tweet and citation counts observed in previous research are found with marginal enforcement to higher levels of correlations for articles with non-zero tweets (Haustein, Costas & Lariivi, 2015). Current research has seen low positive, but significant Spearman’s correlations ranging between 0.214 and 0.352 across four broader fields, except in 2011 physical and life sciences. It was found that the more recent years showed higher levels of correlations (see Table 2), presumably reflecting the impact of rise in tweet uptakes and more prevalence of citations accrued in comparison to previous studies which had aggregated data by shorter time spanned after publication of articles. As shown in Table 2, correlation between researcher tweets and Scopus citations was low positive, though the highest among the cohorts, ranging between 0.306 and 0.353 across years. Likewise, correlations between tweets from cohort of members of the public and formal citations is only slightly at lower range (between 0.242 and 0.332) probably because the counts overestimate the presence of public audience (see also Thelwall et al., 2013b; Tsou, et al., 2015). Because practitioners predominantly tweet articles in Health sciences (72%), the correlations in this field are distinctly offered (varying between .238 and .349), which demonstrates its weak positive, but significant association with scholarly impact. Overall, there is some evidence to suggest that tweets are in association with scholarly communication, though articles tweeted by researchers, and health experts seem to weakly reflect scholarly impact. Therefore, tweets might be cautiously used to interpret scholarly impact of research. In addition, a
more accurate stratification might be required to identify the extent to which the cohort of members of the public can translate into engagement of general public, although it seems their engagement is less significantly relevant to scholarly impact than other cohorts.

Table 3. Spearman’s correlation coefficients between unique tweeter counts and Scopus citations across cohorts for articles with non-zero tweets

<table>
<thead>
<tr>
<th>Publication Year</th>
<th>Researchers</th>
<th>Members of the Public</th>
<th>Practitioners All fields</th>
<th>Health Sciences</th>
<th>Science Communicators</th>
<th>Unknown</th>
<th>Total</th>
<th>No. of Sampled Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>.306**</td>
<td>.242**</td>
<td>.245**</td>
<td>.238**</td>
<td>.194**</td>
<td>.197**</td>
<td>.259**</td>
<td>834</td>
</tr>
<tr>
<td>2012</td>
<td>.315**</td>
<td>.281**</td>
<td>.228**</td>
<td>.254**</td>
<td>.238**</td>
<td>.104**</td>
<td>.294**</td>
<td>1,251</td>
</tr>
<tr>
<td>2013</td>
<td>.330**</td>
<td>.310**</td>
<td>.283**</td>
<td>.347**</td>
<td>.264**</td>
<td>.119**</td>
<td>.330**</td>
<td>1,820</td>
</tr>
<tr>
<td>2014</td>
<td>.353**</td>
<td>.332**</td>
<td>.314**</td>
<td>.349**</td>
<td>.275**</td>
<td>.104**</td>
<td>.354**</td>
<td>2,540</td>
</tr>
</tbody>
</table>

** Significant at p < 0.01.

Acknowledgements

The author would like to thank Dr. Kayvan Kousha, Statistical Cybermetrics Research Group of the University of Wolverhampton, for his useful comments. Also, the author appreciates Altmetric.com team for supporting access to data.

References


