

# Research Trends

---

Volume 1  
Issue 35 *Developing Research in Developing  
Countries*

Article 7

---

12-1-2013

## Towards a common model of citation: some thoughts on merging altmetrics and bibliometrics

Mike Taylor  
Elsevier

Follow this and additional works at: <https://www.researchtrends.com/researchtrends>

---

### Recommended Citation

Taylor, Mike (2013) "Towards a common model of citation: some thoughts on merging altmetrics and bibliometrics," *Research Trends*: Vol. 1 : Iss. 35 , Article 7.

Available at: <https://www.researchtrends.com/researchtrends/vol1/iss35/7>

This Article is brought to you for free and open access by Research Trends. It has been accepted for inclusion in Research Trends by an authorized editor of Research Trends. For more information, please contact [r.herbert@elsevier.com](mailto:r.herbert@elsevier.com).

Section 6:

Expert Opinion

# Towards a common model of citation: some thoughts on merging altmetrics and bibliometrics

Mike Taylor

## Reporting back

This article is based on presentations that Mike Taylor gave at the PLoS article level metrics workshop in San Francisco and at the World Social Science Forum (WSSF) in Montreal, both in October 2013.

The increasing visibility of scholarly communication and discussion has led to a dramatic increase in the complexity of understanding its academic impact and social reach.

Although the nature of the communication has many different forms, with radically different attributes, it is generally treated as a singular entity: that of altmetrics.

In fact, it is arguable that the creation of altmetrics as a singular entity was technocratic (driven by what is technically possible) and thus pragmatic (built from what is available), rather than rooted in a theoretical discipline, and, had the different sources emerged at different times, or been accessed via different technical solutions, they would have been kept discrete.

The fundamental differences are readily apparent. For example, when one tweets a reference to a paper, it can be observed that the communication is necessarily brief, and is unlikely to have taken much time or thought. Frequently it is in the form of a 'retweet' and can be classified as the mere repetition of a message through personal networks.

The effort taken to tweet a link or reference may be contrasted to a blog post, where the intended recipient may well be the original research team, as well as others interested in this academic area. Other forms of scholarly blogs link to papers when attempting to précis the content for a non-academic audience (<http://realclimate.org>), or engage misleading and mendacious uses of research to promote commercial and political aims - a less scholarly endeavor that nonetheless still contains links and discussion.

Nevertheless, both blogs and tweets can be said to have the explicit intention of being public: this can be contrasted with anonymous data that can be harvested and interpreted from many other sites. Of course, formal citation in a peer-reviewed article is also a public act, and this serves to introduce two other important criteria: that of context and immediacy. A tweet may have virtually no context (being only a reference to a paper), whereas a blog post may be several thousand words long. Similarly, a tweet may be an immediate act of impetuosity, whereas a citation in a peer-reviewed paper will necessarily take a longer period.

However, focusing on the issue of privacy: reading or downloading of articles may be considered as a private act in a study room, but user activity counts (and other demographic information) aggregating such acts and provided by tools such as Mendeley, Citeulike, GitHub and DataDryad are often included in publicly available altmetric data, as can be article-level-usage figures from publisher sites.

With the exception of people who are trying deliberately to distort data (for example, by repeatedly downloading an article – a practice which publishers work hard to counter), little is known of how mindful people are of the public nature or use of their activity and how this affects their behavior.

Therefore altmetrics consists of a wide variety of data with different characteristics, linked by a common set of tools. Data is typically accessed via an API (application programming interface), papers referenced by DOIs (digital object identifiers), and the platforms from which the data is gathered are social: this defines the set of data, rather than provides a theoretical foundation. It is not surprising, therefore, that little is known about the intentional, motivational or experiential motives of the users.

When a user posts a paper on Mendeley, we can hypothesize various motives including (but not limited to) the following:

- Other people might be interested in this paper.
- I might read this paper in the future.
- I have read this paper and want it to be easily findable.
- I want other people to think I have read this paper.
- It is my paper, and I maintain my own library.
- It is my paper, and I want people to read it.
- It is my paper, and I want people to see that I wrote it.
- I might skim read this paper in the future because I suspect it might back up an argument I'm thinking about making and it looks like it would make a useful citation.

With Twitter, the poster may choose to call attention to their tweet, to direct people to their response, may address the tweet to the authors, or may add inflections by the arbitrary (or organized) use of hashtags.

Each example of altmetric data has its own set of potential underlying motives, and each example requires different research: tweets may be subject to qualitative research, but are less easily studied by user surveys, for example. It would, of course, be possible (although time-consuming) to monitor tweets and ask the tweeter to complete a survey on their motivations for the individual tweet, but the time taken to survey would probably be disproportionately longer than the time taken to compose and post the original tweet.

To date, altmetric research has focused more on correlation (Priem et al, 1) than on motivation, and has relied upon assumptions rather than empirical evidence to postulate the relative level of engagement with an article (Fenner and Lin, 2).

#### Fifty years of relevant research

The related field of bibliometrics has – since 1962 – conducted a significant quantity of research in the field of motivation of citation. Amongst the many intellectual assets available for potential re-purpose are theoretical models, methodologies, data sets and references. Bornmann and Daniel's 2008 article, "What do citation counts measure? A review of studies on citing behavior" (3) reviews the extensive literature and reports the conclusions of this research. However, with the exception of Priem et al's passing

reference to this review, a search on [Scopus](#) reveals that of the 162 citations made to this paper, not one of them appears to be related to altmetrics.

The scholarly research into reference and citation attempted to test two potential theories of citation motivation: normative and social constructivist. Broadly speaking, the two camps may be positioned as:

1. "Scientists give credit to colleagues whose work they use by citing that work" versus
2. "Scientific knowledge is socially constructed through the manipulation of political and financial resources and the use of rhetorical devices" (reported in 3).

After fifty years of research, Cronin was able to summarize the weight of evidence in favor of the normative view:

"The weight of empirical evidence seems to suggest that scientists typically cite the works of their peers in a normatively guided manner and that these signs (citations) perform a mutually intelligible communicative function" (4).

Shortly after the inception of bibliometrics, Eugene Garfield (1962, as reported in 3) listed fifteen possible motivations to cite:

1. Paying homage to pioneers;
2. Giving credit for related work (homage to peers);
3. Identifying methodology, equipment, etc.;
4. Providing background reading;
5. Correcting one's own work;
6. Correcting the work of others;
7. Criticizing previous work;
8. Substantiating claims;
9. Alerting to forthcoming work;
10. Providing leads to poorly disseminated, poorly indexed, or uncited work;
11. Authenticating data and classes of fact (physical constants, etc.);
12. Identifying original publications in which an idea or concept was discussed;
13. Identifying original publication or other work describing an eponymic concept or term (...);
14. Disclaiming work or ideas of others (negative claims); and
15. Disputing priority claims of others (negative homage).

All of these are as relevant to social citation in 2013 as they were to formal citation in 1962; and the added visibility and speed of activity in social networks only adds to the list, for example:

16. Building a network of related researchers;
17. Building a reputation as a good networker;
18. Paying visible homage to a senior researcher;
19. Seeking the attention of a senior researcher;
20. Demonstrating that one's reading is up to date; and
21. Intimidating critics with the breadth of one's reading.

There are many more motivations that can be added to this list.

That there should be general agreement on the nature of formal citation should come as little surprise: learning how to reference, or "show your reading" is a skill that is taught from an early age. Many websites exist to support and develop best citation practice, even going to the length of invoking the law to encourage completion:

"If you do not include your references both in your essay and on a reference sheet at the end of your essay, you could face legal action for being in violation of plagiarism laws." How to Add Citations in an Essay, Allison Boyer, [http://www.ehow.com/how\\_7472938\\_add-citations-essay.html](http://www.ehow.com/how_7472938_add-citations-essay.html)

Various Google searches on October 22, 2013 for equivalent guidelines for tweeting scholarly references produced no relevant results, beyond guidance on structuring the actual form of the citation (<http://ucanr.edu/blogs/blogcore/postdetail.cfm?postnum=11505>). However, there are many resources to support the use of Twitter in the K-12 teaching environment (e.g. <http://www.teachhub.com/50-ways-use-twitter-classroom>). It seems like a reasonable assumption that people's first contact with social media will be away from the support of the academic community, and that individual practice will develop in a varied social environment.

Although statistics relating to negative citation are well-known (Bornmann and Daniel report a 5% incidence) there is a distinct contrast when it comes to abusive expression of power relations in social media. [Scopus](#) has indexed 30 papers with "cyberbully" in the title or abstract, and Schenk and Fremouw (5) report that 8.6% of college students have been subjected to cyberbullying.

The two observations: that people learn to use social networks away from an academic environment, and that the expression of power relations (at least in Twitter) is common may lead us to conclude that social citation – at least in the sense of public reference - may be less characteristic of normative citation practice.

### Developing a methodology

Altmetric data is complex and varied: in order to study it, it is necessary to simplify and normalize the data. For example, the usage figures of social networks vary across time, with networks drifting in and out of fashion, being subject to phases of organic growth and early adopter use, and with operators controlling access to data via their APIs.

#### Increasing engagement with the article, Fenner and Lin (1 is lowest level of engagement, 5 is maximum):

1. Viewing: the activity of accessing the article online.
2. Saving: storing and referencing of articles (or references) in online tools such as Mendeley or Citeulike.
3. Discussing: Ranging from tweeting to blogging.
4. Recommending: formal endorsement of a paper, e.g. F1000Prime.
5. Citing: formal citation of an article in another article.

Developing the idea of Lin and Fenner's taxonomy of social citation / usage behavior – albeit with some critical changes and without the idea of developing engagement with the article – it is possible to make sense of types of altmetric behavior. Rather than attributing motivation - or assuming that tweets are a deeper level of engagement than reading the article - I propose classifying activity according to the level of engagement with the behavior, as defined by the user's choice of platform:

- **Social activity** – characterized by rapid, brief engagement by users on platforms used by the general population – Twitter, Facebook, Delicious, etc.
- **Component re-use** – the re-use of the constituent elements of the research product – data, figures and code.
- **Scholarly commentary** – in-depth engagement by people using scholarly platforms, such as Science Blogs, F1000Prime reviews, etc.
- **Scholarly activity** – indirect measurement of activity by people using scholarly platforms, e.g., Mendeley, Zotero, Citeulike.
- **Mass media coverage** – coverage of research output in the mass media.

Any well-defined and meaningful collection of data should present two characteristics:

1) the sources that comprise an instance of data (e.g., social activity) should correlate well – for example, if the data is measuring the same class of activity, we should see tight correlation of activities between Twitter, Facebook, Delicious, etc.

and

2) each class should show discrete phenomena of activity.

Both of these are readily testable, and as altmetrics grows to encompass more datasets, it should be able to accommodate further classes of data. For example:

- Social activity surrounding mass media – comments, tweets, etc., linking to mass media coverage of scholarly output.
- References in books and monographs.
- Use of scholarly research in commercial activity, e.g., patents.
- Use of scholarly research in legislation and governmental context.
- Self-promotion, e.g., additional content to support use of research, press releases.

In each case, the legitimacy of the distinctness of the classes and the difference between the classes can be readily tested. In order to validate the uses of the classes to describe motivational behavior and to discover causal patterns between the different types of activity, it is necessary to engage in qualitative research – methods that have been exhaustively researched by the bibliometric researchers reported in Bornmann and Daniel. It is possible that some of this work may be aided by text-mining and entity-recognition techniques, as used in natural language processing research, but any attempt to ascribe motivation to social users will require surveys and interviews.

If the classes of altmetric activity are validated as distinct and internally consistent, then several research steps might follow:

- Identifying statistical trends between the classes.
- Qualitative analysis to understand causation.
- Surveys to acquire evidence of motivation.
- Understanding the likely consequences of 'gaming' behavior, e.g. buying tweets, encouraging colleagues to load papers into Mendeley, etc.
- Understanding how behavior changes as a consequence of legitimate promotion.
- Qualifying social citation / social network activity between disciplines, professionals and as the platforms develop.
- Discovering how combinations of classes can contribute to the understanding of potential use cases for altmetric data.

Considering this last point, there are many different issues that might be understood via a properly formulated study of altmetrics and bibliometrics. Given the pragmatic nature of altmetrics, the potential methodologies are varied, and this list is advanced as a discussion point.

1. **Prediction of ultimate citation** – although it has been speculated that some altmetric data might enable a prediction of future citation rates, research has not yet demonstrated a correlation between Twitter counts and citation (Haustein et al 2013) (6). However, disciplines are likely to vary in their adoption of different types of activity, so this work – which may be added to other research that attempts to predict citation rates – will continue to look for correlations in data (7).

2. **Measuring / recognizing component re-use / preparatory work / reproducibility** – a distinctive strand of altmetrics research is focused on measuring re-use of scholarly materials. This is of interest to funders and institutions in its own right; however, making data, code, etc. freely available may lead to increases in reproducibility and reliability. Nevertheless, work would need to be undertaken to understand the extent to which data (etc.) is reused simply because it is available, or well curated, rather than driven by scholarly need.

3. **Hidden impact (impact without citation)** – there has been speculation that some articles may have an impact that is not detected using bibliographic citation analysis. For example, “How to choose a good scientific problem” (8) has only been cited 4 times, according to Scopus, but has been shared on Mendeley nearly 42,000 times as of October 31, 2013.

4. **Real-time filtering / real-time evaluation** of important / impactful articles relies on both a qualitative and quantitative analysis of real-time data. However, it is unknown if there is sufficient data to make this work at a sufficiently fine granularity, whether this is of use to scholars and whether they would trust such a system.

5. **Platform / publisher / institution comparison** – although altmetrics can be used to gauge how effective organizations and authors are at providing social sharing tools, there has been no research on what this data might mean in terms of quality of research, rather than the more obvious values of being a ‘good read’, titivation or scandal.

6. **Measuring social reach / estimating social impact** – evidently a crucial part of communicating research outcomes to society is the ability to communicate, and altmetrics could be used as a starting point to understand the flow of research impact in society – if it expands its remit, issues of privacy remain of low concern and if citation practices outside academia improve (<http://www.researchtrends.com/issue-33-june-2013/the-challenges-of-measuring-social-impact-using-altmetrics>).

## Conclusion

The outcome of research in this area should be to align the studies of altmetrics and bibliometrics by developing a common theoretical model that allows for analysis of all forms of accessible reference to scholarly objects: in short, a model of the scholarly network.

Such an ambition would allow for the commonalities between formal citation and altmetric activity, and for understanding the differences. By accepting that different forms of citation or reference take place in environments with different attributes and motivations, we will achieve a richer view of both bibliometric activity and social citation.

## Acknowledgements

The author is indebted to the many people who are passionate about understanding scholarly communication and always ready to spend time discussing the issues, most notably Dr Stefanie Haustein, Dr Henk Moed, Euan Adie of Altmetric.com, Gregg Gordon of SSRN and many others.

## References:

1. Priem et al, “Altmetrics in the Wild”, Available at: <http://jasonpriem.org/self-archived/PLoS-altmetrics-sigmetrics11-abstract.pdf>
- (2) Lin, J., Fenner, M. (2013) “Altmetrics in Evolution: Defining and Redefining the Ontology of Article-Level Metrics”, *Information Standards Quarterly*, Vol. 25, No. 2, pp. 20, Available at: [10.3789/isqv25no2.2013.04](http://dx.doi.org/10.3789/isqv25no2.2013.04)
- (3) Bornmann, L., Daniel, H. (2008) “What do citation counts measure? A review of studies on citing behavior”, *Journal of Documentation*, Vol. 64, No. 1, pp. 45-80, Available at: [0.1108/00220410810844150](http://dx.doi.org/10.1108/00220410810844150)
- (4) Cronin, B. (2005) “A hundred million acts of whimsy?”, Available at: <http://www.iisc.ernet.in/currsci/nov102005/1505.pdf>
- (5) Schenk, A.M., Fremouw, W.J. (2012) “Prevalence, Psychological Impact, and Coping of Cyberbully Victims Among College Students”, *Journal of School Violence*, Vol. 11, No. 1, pp. 21-37, Available at: [10.1080/15388220.2011.630310](http://dx.doi.org/10.1080/15388220.2011.630310)
- (6) Haustein, S., Peters, I., Sugimoto, C.R., Thelwall, M., Larivière, V., “Tweeting biomedicine: an analysis of tweets and citations in the biomedical literature”, Available at: <http://arxiv.org/ftp/arxiv/papers/1308/1308.1838.pdf>
- (7) Yan, R., Huang, C., Tang, J., Zhang, Y., & Xiaoming, L. (2012) “To Better Stand on the Shoulder of Giants: Learning to Identify Potentially Influential Literature” Available at: <http://keg.cs.tsinghua.edu.cn/jietang/publications/JCDL12-Yan-et-al-To-Better-Stand-on-the-Shoulder-of-Giants.pdf>
- (8) Alon, U. (2009) “How To Choose a Good Scientific Problem”, *Molecular Cell*, Vol. 35, No. 6, pp. 726-728.

## Related presentations:

- Taylor, M. (2013) “140 Characters in Search of a Meaning: Incorporating Motivation into Altmetrics”, Available at: <http://dx.doi.org/10.6084/m9.figshare.821283> (Retrieved 12:10, Oct 23, 2013 (GMT))
- Taylor, M. (2013) “The Many Faces of Altmetrics Mapping the Social Reach of Research”, Available at: <http://dx.doi.org/10.6084/m9.figshare.820136> (Retrieved 12:13, Oct 23, 2013 (GMT))