The 2017 Altmetrics Workshop: The dependencies of altmetrics (Toronto, 26 September 2017)

Altmetric-Citation Disequilibria in Alzheimer-Disease Papers: Might Such Discrepancies Influence Journal Choice in Manuscript-Submission Decisions?

Colin Magdamo* and Aaron A. Sorensen**

* University of Minnesota (magda007@umn.edu)
** Digital Science Consultancy (a.sorensen@digital-science.com)

Introduction

Upon completion of a manuscript, authors must go through the process of choosing a journal to which they will submit their work. The choice of journal can be difficult as there is a tradeoff between total-time-to-publication and the prestige of the journal to which they first submit. Specifically, if a group of authors "aims too high" they can find themselves caught in a frustrating cycle of "major-changes-requested" or of outright rejections which can delay the publication of the article for months. The prestige of a journal has come to be quantified by its Journal-Impact-Factor (JIF), and the manuscript-submission options weighed by a group of authors are likely bounded by a minimum JIF they would be willing to accept and by a maximum JIF above which, it would be foolish to aim. Within a given manuscript's JIF-submission band, there will likely be multiple journals between which the authors in question find themselves unable to discriminate in an informed manner. As the use of altmetrics has become more widespread within the word of scientific publishing, a natural question that arises is whether altmetric-based journal characteristics might help submitting authors decide between competing journals with neighboring JIF values.

Objectives

The objective of this study is the development of a metric which might help submitting authors discriminate between journals within their current manuscript's perceived JIF-feasibility-band by quantifying the degree to which citation and altmetric measures for the journal in question are in (dis)equilibrium.

Because one of us (AAS) is the Scientometrics Editor of the Journal of Alzheimer's Disease (AD), the decision was made to focus this preliminary analysis on the 169 journals which have contributed at least thirty papers to AD research between 1/1/2015 and 3/21/2017.

Methods

In last year's Altmetrics16 workshop, (see Zahedi et al, 2016), a Principal Components Analysis (PCA) was presented in which the algorithms were run on WoS publications spanning numerous domains. A reviewer of our original submission suggested that conducting a similar analysis for the AD literature might provide us with new insight as to the source and structure of the underlying signal in our data set. We heartily agreed, and have conducted a PCA on the data in which our proposed altmetric-citation disequilibrium metric (see below) was the variable exhibiting the highest correlation with the second principal component.

Citation and altmetric data were extracted from Digital Science's Dimensions and Altmetric Explorer applications respectively.

Because this study focused AD papers published in recent years, the practice of taking the maximum of citation count or Mendeley-reader count as an estimator of future scientific impact was employed (see Thelwall, 2015).

The analysis of the data was performed in *R* (https://figshare.com/s/8be3225ce0dcd3daea42), version 3.3.0, running on OSX 10.10. We used the *data.table* package to help with both importing raw publication data (https://figshare.com/s/3af72febbc570bb2c666) and for the journal-level aggregation needed in the output file (https://figshare.com/s/72b08c6cce85bd4776fb). Additionally, the *Altmetric_Final.rmd* script performs citation-v-altmetric density estimations (https://figshare.com/s/c15c1fa79678002e8940) as well as plotting the social/non-social-media split of Altmetric scores over time (https://figshare.com/s/067e5a41e28c9494dc58).

Results

As a metric to help AD investigators understand the degree to which a given journal's Alzheimer's publications are in altmetric-citation equilibrium, we propose δ, or delta, which captures the disparity between a given publication's Altmetric-score percentile ranking and its citation-count percentile ranking within the AD literature of a given year. Paper-level δs are then aggregated at the journal level, resulting in a journal-specific delta that captures the typical disparity between a the Altmetric percentile ranking and the corresponding citation-count percentile ranking for the AD-related articles published in that journal (for the mathematical formalism see https://figshare.com/s/aba7389e5e61c65627c5).

The PCA results (see https://figshare.com/s/09dacb33b402934bc53d)

suggest that within the AD literature, in particular, it may be too difficult to separate the scholarly impact and altmetric effects for clustering purposes. However, it appears that journals

which frequently publish AD papers tend to have higher Altmetric percentile rankings than citation rankings. There is evidence that extending PCA (which assumes linear structure) to non-linear techniques may be better suited for the task of reducing publication dimensionality in a prediction context.

The dependencies we faced were the sparseness of citation counts for recent publications and the overall sparseness of altmetric counts leading to multimodal artifacts in the distributions. Our solution for citation sparseness was to use Mendeley reader counts as a citation proxy when higher than the corresponding citation count. Our altmetric-dependency work-around was to substitute the averaging of yearly medians with a median-of-medians approach.

Discussion

The AD journals with deltas in the top and bottom deciles (https://figshare.com/s/ee0ac4a03b6a916e3b8f) provide some anecdotal evidence worth highlighting.

In the top (green) decile, we observe that two of the journals with Altmetric-heavy disequilibria, are either non-traditional in media type (JoVE) or in topical focus (eCAM).

Additionally, by applying the Australian Research Council's Field of Research (FOR) categories to the two deciles (https://figshare.com/s/346a35f2289fe2c1c6c5), we observe an overrepresentation of chemistry-focused journals (i.e. "0601-Biochemistry-and-Cell-Biology," "03-Chemical-Sciences," "0304-Medicinal-and-Biomolecular-Chemistry," and "1101-Medicinal-Biochemistry-and-Metabolomic") in the citation-heavy decile (bottom, orange).

An interesting follow-up analysis would be to see whether chemistry, as a category, and JoVE and eCAM, as disruptive journals, would be similarly placed in altmetric-citation disequilibria analyses of other disease literature sets.

We would like to discuss with other workshop participants the above findings within the context of our decision to use a median-of-medians method to compensate for the altmetric data's sparseness. Secondarily, it would be of interest to discuss the use of Mendeley-reader counts, an altmetric measure, to compensate for a citation sparseness when trying to analyze newer articles.

References

1) Thelwall, M. (2015). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa?. *Journal of Librarianship and Information Science*, 0961000615594867.
2) Zahedi, Z., Haustein, S., Larivière, V., & Costas Comesana, R. (2016). On the relationships between bibliometric and altmetric indicators: The effect of discipline and density level. The 2016 Altmetrics Workshop, 27 September 2016, Bucharest, Romania.