Disciplinary differences in altmetrics for humanities research

Or, are gender studies researchers trolled more online?

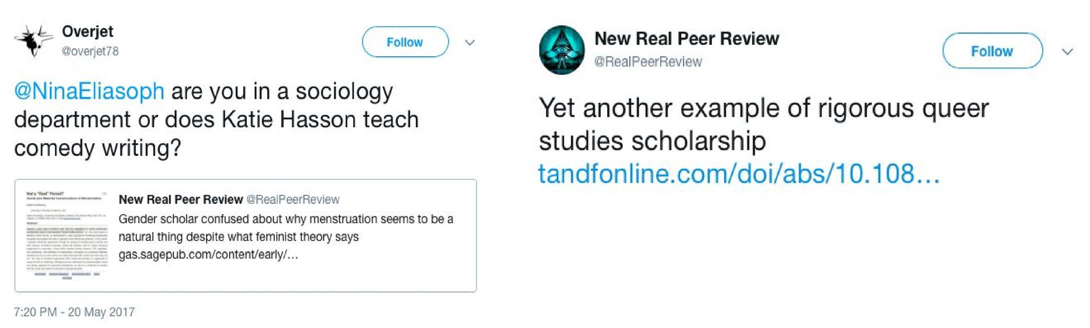
Alex Enachioaie, Site Reliability Engineer, Altmetric

Stacy Konkiel, Director of Research & Education, Altmetric

Introduction

Among the uninitiated, more metrics for research is often interpreted to be positive. After all, if more people are talking about a study, isn't that a good thing?

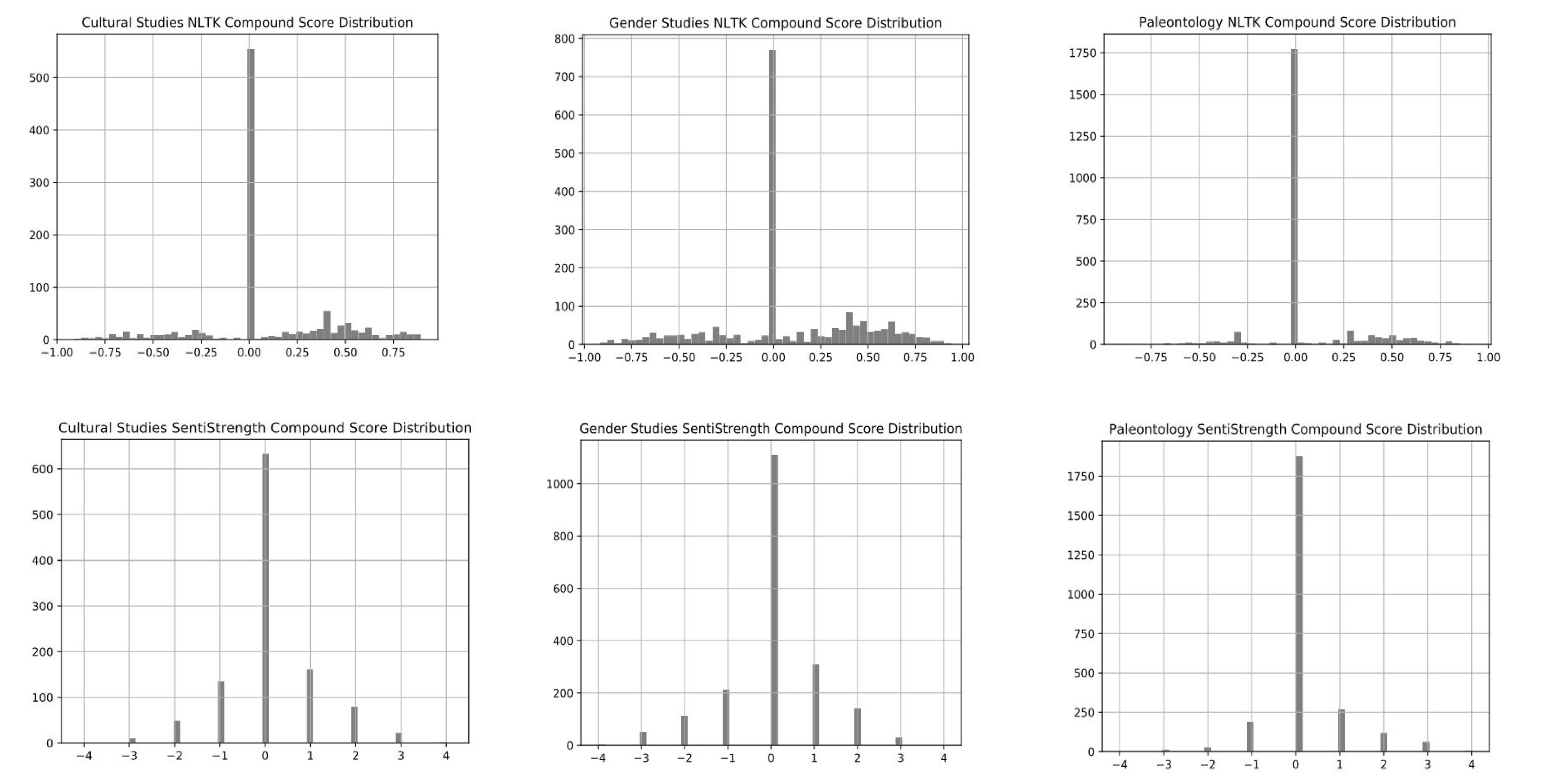
We suggest that more metrics are not always good, especially for certain humanities disciplines. In this study, we examine the effects of disciplinarity upon altmetrics for humanities research, comparing data for gender studies research published in the field's top ten (by SciMago Journal Rank) journals with articles published in the top ten cultural studies (humanities) and paleontology (STEM) journals. Specifically, we test our assumption that gender studies research is subject to more negative discussions than other research areas.



Methods

1. Extracted top ten journals and their ISSNs from SciMago Journal Rank
2. Added ISSNs to Altmetric Explorer to find all content with Altmetric attention published in those journals in 2016
3. Analyzed journal-level attention by discipline
4. Extracted full-text Twitter mentions directly from Altmetric database for sentiment analysis
5. Cleaned Twitter dataset for each discipline by removing non-english tweets, duplicates, article titles, usernames and other smaller contaminating artifacts
6. Analyzed each Twitter dataset using the Natural Language Tool Kit's VADER sentiment analysis tools
7. Analyzed each Twitter dataset again using SentiStrength

Sentiment Scores (NLTK & SentiStrength)



The difficulties of sentiment analysis for tweets

Manual coding - As opposed to other sentiment analysis and altmetrics studies which manually coded tweets to understand the accuracy of automated sentiment analysis tools like SentiStrength, we decided instead to take a comparative approach and test two different automated tools against each-other and assess the resulting level of agreement. It is likely that some types of sentiment expression which may be prevalent in our data sets (subtle irony, sarcasm, obscure terminology) may be poorly interpreted by such automated tools and manual coding could offer further insights.

Removing duplicate tweets - We took the potentially controversial step of removing duplicate tweets from our data set after inspecting the data and concluding that there was a high degree of spam-like repetition of neutral promotional style tweets (e.g. "Check out my new paper..."). This made the datasets more manageable for analysis and eliminates a source of neutral sentiment skew but can also affect the other sentiment scores if it eliminates positive or negative retweets (e.g. "Feminists are idiots. What rubbish!"). The extent of the general skew effect is debatable on the basis of whether a retweet can be considered to have the same emotional intensity as the original. To check the magnitude of this potential mean skew we compared the scores on the gender studies data set with and without duplicates removed and did not find any meaningful differences.

Removing titles - Another problematic aspect of the data set is the presence of titles in tweet text which overall tend to be neutral sentiment, thus possibly skewing the mean in its favor. One approach is to search for and remove titles as a whole which we did, however this had a fairly low hit rate because of human-introduced artifacts in tweets. The other would be to tokenize (split into words) titles and tweets and remove all the tokens present in the title from the tweets. This can itself skew the results however, as it has the potential to heavily alter text semantic/sentiment.

Differences in platforms for discussion

In all three disciplines, Twitter was the platform most used to discuss research (a trend seen in previous studies, as well).

There was no difference in the other most popular platforms used to discuss disciplinary research: all disciplines received the most attention in the news, on blogs, and on public Facebook pages.

Gender Studies research was mentioned in unique attention sources: Faculty of 1000 Prime and YouTube. It was not mentioned in policy sources, though Cultural Studies and Paleontology papers were.

Key findings

Cultural studies and gender studies research differed from paleontology in that those fields received more negative *and* positive attention, relatively speaking.

Surprisingly, all fields received more positive than negative attention.

Neutral attention to research strongly prevailed, confirming previous studies. Both positive and negative attention tends to be clustered in low to moderate intensity ranges (∓0.25 - 0.75; ∓ 1-3) with little to none in the extremes.

Overall the relatively high amount of manual review necessary to ensure clean mention data would likely preclude altmetrics services from confidently offering sentiment analysis in the near future.

References

Bird, Steven, Edward Loper and Ewan Klein (2009), Natural Language Processing with Python. O'Reilly Media Inc.

Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology, 61*(12), 2544–2558.

Acknowledgements

Many thanks to Dr. Mike Thelwall for granting us access to SentiStrength software, and Maciej Gajewski for his assistance in data retrieval. Thanks also to Jean Liu for this poster's visual precedent.

Contact: Stacy Konkiel - stacy@altmetric.com - @skonkiel
Alex Enachioaie – alex@altmetric.com - @quasi_quasar

Check out our code: github.com/thisisjaid/4am