



Universiteit
Leiden
The Netherlands

On the relationships between bibliometric and altmetric indicators: The effect of discipline and density level

Zahedi, Z.; Haustein, S.; Larivière, V.; Costas Comesana, R.

Citation

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. Retrieved from <https://hdl.handle.net/1887/47066>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/47066>

Note: To cite this publication please use the final published version (if applicable).

The 2016 Altmetrics Workshop (Bucharest, 27 September, 2016)
Moving beyond counts: integrating context

On the relationships between bibliometric and altmetric indicators: the effect of discipline and density level

Zohreh Zahedi (a), Stefanie Haustein (b), Vincent Larivière (b) and Rodrigo Costas (a)

(a) Centre for Science and Technology Studies (CWTS), Leiden University, Leiden, the Netherlands

(b) École de Bibliothéconomie et des Sciences de l'Information, Université de Montréal, Québec, Canada

Introduction

The analysis of the characteristics of social media metrics and their correlation with other bibliometric indicators has been well discussed in the literature (see Haustein et al, 2015). These studies have shown that various altmetric indicators exhibit different degrees of correlation with citations, and that this varies according to document types and discipline. This study builds on these previous analyses, and aims to answer the following research questions:

- What are the statistical dimensions of indicators that can be identified when combining bibliometric and altmetric indicators? ? Are these dimensions stable across disciplines?
- Does the relationship between indicators change when different density levels (i.e. higher or lower levels of activity) are considered?

To explore these questions, the following biblio/altmetric characteristics of papers are considered for all 2012 papers with a DOI indexed in the Web of Science: number of authors in the paper [n_authors], number of institutions [n_institutes], number of countries [n_countries], number of citations [n_cits], number of Mendeley readers¹ [mr], number of tweets [tw], number of Facebook mentions [fb], number of blog mentions [b]², number of citations in Wikipedia [w], number of pages [pg] and number of cited references [n_refs]. Publications and citation data are extracted from the Web of Science. A total of 1,791,169 paper-discipline combinations are considered, and 35 fields are considered (cf. Ruiz-Castillo & Costas, 2015).

Methodology

Two main methodological approaches are applied.

1) Factor Analysis³ (implemented in SPSS®) is used to study the dimensions of indicators globally and by field, for all the variables considered.

¹ From the Mendeley REST API (July 2016).

² Twitter, Facebook, blog and Wikipedia citations were obtained from Altmetric.com (June 2016).

³ Factor analysis is a statistical method to reduce the dimensionality of the data, in order to discover its underlying structure and interpret dependencies among sets of variables.

2) Characteristic Scores and Scales⁴ (hereafter CSS—Schubert et al, 1987; see also Crespo et al, 2012), to study potential differences in the correlations among indicators, according to different indicator density levels..

Results

What dimensions can be identified when combining bibliometric and altmetric indicators?

Table 1 presents the rotated Varimax solution of the Factor Analysis for all the records and variables included in the study. Four main dimensions (latent variables) are identified: Component 1 is the ‘collaboration’ dimension. Component 2 corresponds to the ‘scholarly impact’ dimension (such as the total number of citations and Mendeley readership). Component 3 is composed by the ‘social media’ indicators (Twitter, Facebook and blogs), and component 4 comprises bibliographic characteristics of documents (number of pages and references).

Table 1. Rotated Component Matrix for all observations^a

	Component			
	1	2	3	4
n_institutes	,968	,043	,020	,021
n_authors	,911	-,013	,011	-,017
n_countries	,860	,078	,015	,069
mr	,005	,836	,064	,113
n_cits	,080	,830	,011	,111
w	,007	,212	,127	,018
tw	,019	,168	,806	-,023
fb	-,002	-,099	,712	,081
b	,030	,353	,670	-,077
pg	,042	-,005	,028	,879
n_refs	,018	,264	-,016	,814

Extraction Method: Principal Component Analysis (66% of total variance explained).

Rotation Method: Varimax with Kaiser Normalization.^a

Are these dimensions stable across disciplines?

In order to answer this question, we have performed a Factor Analysis by discipline. It shows that the same dimensional composition is observed for most fields. More specifically, 19 disciplines exhibit the same composition of dimensions as observed for the all disciplines combined. Seven disciplines exhibit

⁴ In this method, publications are classified in three categories: ‘type 1’ regroups publications with the lowest density of an indicator; ‘type 2’ gathers publications with an intermediate density, and ‘type 3’ are the publications with the highest density. The groups are defined by two means: *m1* (overall mean of the distribution of an indicator in a given discipline) and *m2* (mean of the metrics of the publications above *m1*).

slight variations of the general pattern, with blogs (Creative arts, Culture and Music'), Facebook ('General and Industrial Engineering') and Wikipedia citations ('Electrical Engineering and Telecommunications'; 'Health Sciences', 'Multidisciplinary journals', 'Computer Sciences' and 'Physics and Materials Science') having a stronger relationship with the 'scholarly impact' dimension.

Does the relationship between indicators change according to different density levels??

Figures 1-3 present the distribution of the Spearman correlation coefficients for citations, Mendeley readership and Twitter, by levels of density. The box plots depict the distribution of the correlation coefficients of the 35 disciplines considered.

Figure 1. Boxplot of the correlation coefficient between citations and Mendeley/Twitter for the 35 disciplines, by density class

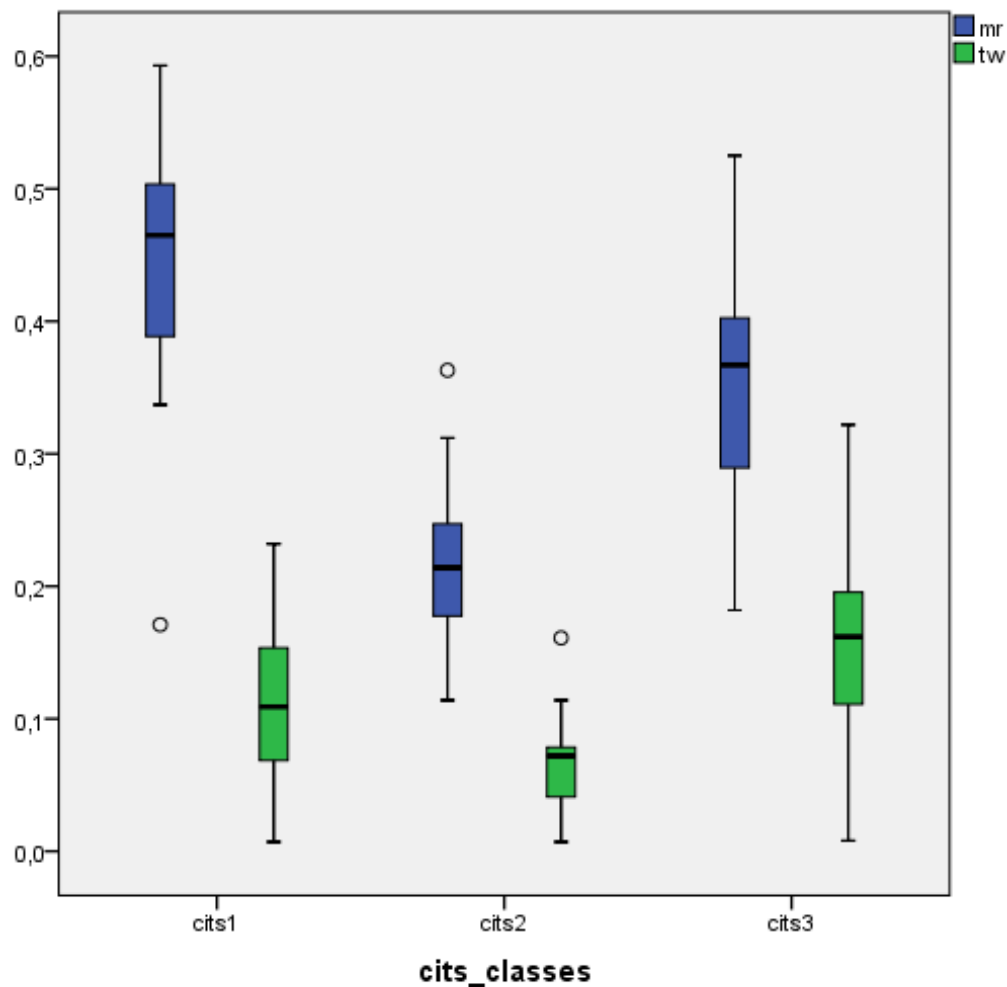


Figure 2. Boxplot of the correlation coefficient between Mendeley and citations/Twitter for the 35 disciplines, by density class

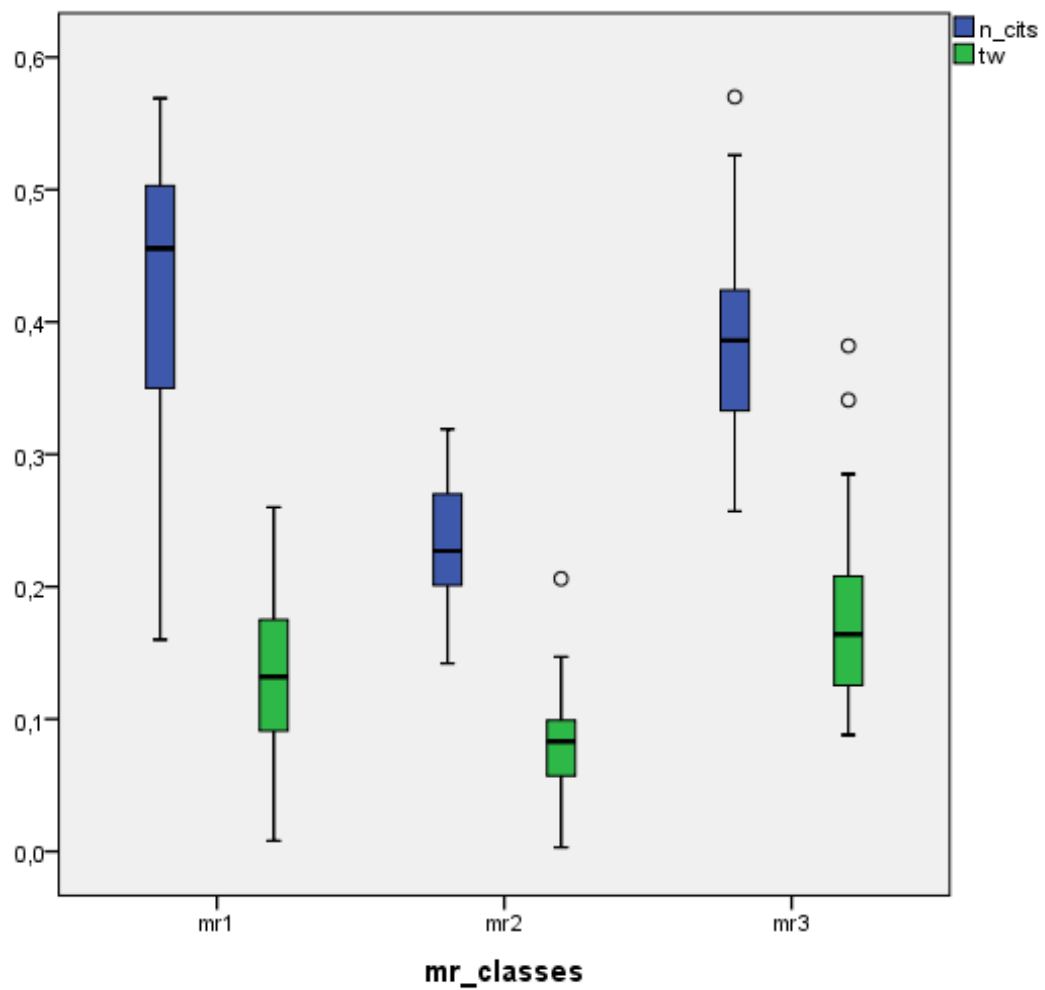
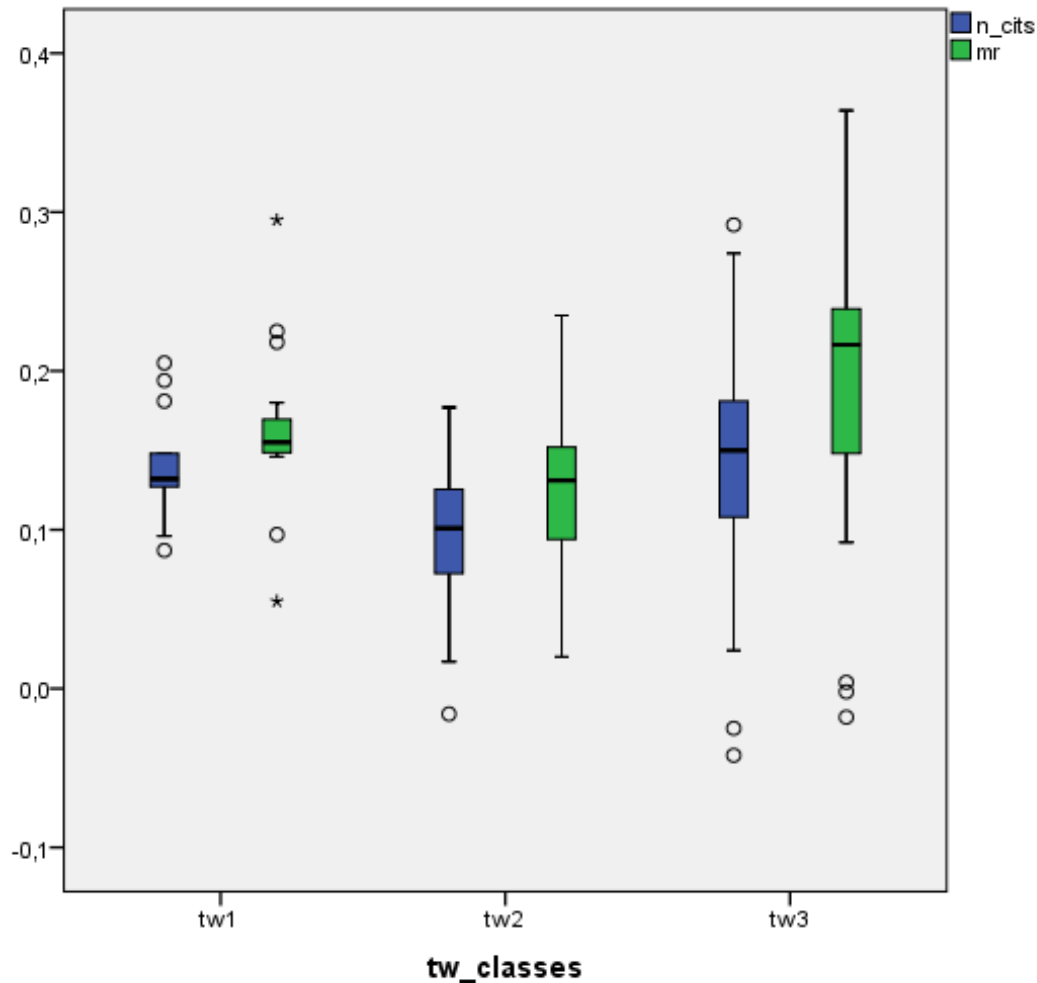


Figure 3. Boxplot of the correlation coefficient between Twitter and citations/Mendeley for the 35 disciplines, by density class



Figures 1 and 2 show a clear pattern. Publications with the lowest (type 1) and highest (type 3) density exhibit the highest correlations between citations and the other two indicators. Mendeley and citations show a higher correlation in the group of lowest density, followed by the highest density class and finally by the intermediated density class. In the case of Twitter the higher correlations (with citations and Mendeley) are found in the highest density groups of both citations and Mendeley. This is corroborated by Figure 3, in which we can see that the correlation between Twitter and the two other indicators increases with the density of Twitter activity.

Discussion, conclusions and further research

Our results show that the various dimensions of altmetric and bibliometric indicators are mostly the same across the various disciplines. They also provide evidence that ‘scholarly impact’ indicators (i.e. Mendeley, citations and occasionally Wikipedia citations) tend to appear as a separated dimension of ‘social media’ indicators (i.e. Twitter, blogs and Facebook). These results support results previously

obtained by Costas et al (2015 and Zahedi et al (2014). On the whole, this suggests that social media metrics are significantly different from scholarly metrics, and that this difference is consistently observed in the various disciplines.

Our results also suggest that the distribution of 'social media' indicators across publications is critical to better understand their meaning and value. The subset of publications with higher Twitter scores tend to have a higher correlation with scholarly indicators. However, the fact that these correlations are still quite low (usually below .200) reinforces the idea that Twitter mentions, even for the highest levels of density, capture a different type of impact as that of citations.

Acknowledgements

This research has been partly supported by funding from the South African DST-NRF Centre of Excellence in Scientometrics and STI Policy (SciSTIP).

References

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. doi:10.1002/asi

Crespo, J. A., Li, Y., & Ruiz-castillo, J. (2012). *Differences in citation impact across scientific fields*, (December), 1–32.

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. *PLoS ONE*. doi:10.1371/journal.pone.0120495

Ruiz-Castillo, J., & Costas, R. (2014). The skewness of scientific productivity. *Journal of Informetrics*, 8(4), 917–934. doi:10.1016/j.joi.2014.09.006

Schubert, A., Glänzel, W., & Braun, T. (1987). Subject field characteristic citation scores and scales for assessing research performance. *Scientometrics*, 12(5), 267–292.

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? A cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications. *Scientometrics*, 101, 1491–1513. doi:10.1007/s11192-014-1264-0