Study on Most Prolific Authors' and their Association with Citation

S Ravikumar

Department of Library and Information Science, North Eastern Hill University, Shillong, INDIA.

ABSTRACT

This paper tries to find the association between altmetric variable and citation for the articles written by most prolific authors in the field of Clinical medicine, Neuroscience, Micro and Molecular biology, 2014. A total of 713 authors had written 4886 articles in the above four subject areas. From the study it was found that five variables (Citation, Altmertric, Twitter Facebook and Mendeley) contribute more than 54% of variance among the seventeen variables. The subject wise correlation between the majority altmetric variables and citation is between 0.158 - 0.737. It was found that medical science attracts more attention in the social media compared to other science domain.

Keywords: Altmetric, Citation, Attention score, Prolific authors, Social media, New media.

Correspondence
S Ravikumar, Ph.D
Assistant Professor Department of Library and Information Science, North Eastern Hill University, Shillong, INDIA.
Phone: +91 9862305939
E-mail: ravikumar@nehu.ac.in

INTRODUCTION

Research articles are published in all stream of subject but all subjects doesn't contribute same quantum of literature to its subject domain. The literature output varies subject to subject and within a particular domain; all its sub-domain does not attract same level of research input which leads to scientific literature. Scientific publication is strongly related with R and D spending [1] which differs subject wise and nation wise. All subject doesn't get the same type attention as hot subjects like Genomic, Energy or Information Technology.[2] In every subject there are few prolific authors whose contribution to the domain knowledge is substantial. In past decade an academic writing was recognized only after a year or so, in the form of citation; but in the recent past the trend has changed dramatically due to the penetration of mobile technology with reference to Social Networking Site (SNS).These new tools have enriched the visibility of research outputs and academic writing. The penetration of ICT had give birth to various article level metrics under the umbrella of Altmertics.

This paper contributes to this discussion by comparing seventeen different altmetrics variables sources with citation data of 4989 articles of four different subjects namely Clinical Medicine, Microbiology, Molecular Biology and Neuroscience Published in 2014which were penned by 713 authors. Specifically,

this study seeks to answer the following research question: To what extent do the altmetric indicators associate with citation counts of most prolific authors?[3]

LITERATURE REVIEW

Scientometric analysis involves the construction and application of a series of indicators of the 'impact', 'influence', or 'value' of a scholarly work from citation data.[4] Bibliometrics is a standard method for measuring the impact of research on other research.[5] Evaluation of scholarly communication in present environment is not confined to citation alone as new media influence is substantial in digitalcommunication mode. The new media had given birth to new metric which is defined as alternate metrics to evaluate the reach of new research output on various platforms like "twitter", "facebook", "mendeley", "citeulike", "google +" , "Wikipedia" and many more which contribute to a new form of score called Altmertic score. Altmetrics is proposed as an alternative and the extension of the traditional bibliometric indicators (such as Journal Impact Factor or h-index).[6] These techniques include both "intrinsic" measures linked to the author's scientific community, and "extrinsic" measures from the broader context and outside the research community. Altmetric enables to study the impact of a paper just a few days or weeks after its has appeared.[5] Various studies had been done on individual altmetric variables and its association with citation like Twittes, Facebook,[7,8,9] Mendeley.[10,11] All the seventeen variables have association with citation count but it's more prominent among the science stream compared to physical and social science.

METHODS

Most prolific authors in the field of Clinical medicine, Neuroscience, Micro and Molecular biology were retrieved from WoS during 2014. Articles published in year 2014 were taken for the present study because a minimum windowperiod of two year was considered for citation time window.[9] The final list contained 713 authors, who together had authored 4886 papers; 2169 from Clinical Medicine, 1269 from Molecular Biology , 822 from Neuro Science and 616 from Microbiology- which is covered in Google scholar.[13] Author's affiliation was used to counter check the author's credentials. Currently three big altmetrics data providers: ImpactStory, Altmetric.com, and PlumX. Individual author's publication metrics was retrieved from Altmertric.com. Altmetric scores are largely based on New Media like Twitter, facebook ,Google+ and reference management tool like Mendeley and CiteUlike.[14] Data was collected between Januarys – February, 2017. To avoid redundancy manual method was used instead of API. Downloaded altmetric data was tabulated using spreadsheet according to the variable needed for the study.[15]

For data analysis "Factor analysis" was used; which is an explorative analysis much as that of cluster analysis. Factor analysis groups variable into dimensions based on their similarity. This means taking a mass data and shirinking it to a smaller data set, which will help to find the hidden patterns.

Data Analysis

At the outset, inter correlation Table 1 for all variables was computed to carry out factorial analysis. In respect of the study variable;the following five variables with more than one eigenvalue was considered. Citation, Altmertric, Twitter, Facebook and Mendeley, with 5.382,1.173,1.055, 1.003. Also the "scree test" Figure 1 suggested the retention of five factors. The total variance contributed by the eigenvalues more than one were 29.898,6.517,6.302,5.860,5.572. The cumulative percentage of variance explained by the five eigenvalues was 54.149%,[16] which means half of the variance is accounted for by the first five factors.

From the 17 study variables five was extracted using principal component method. In the present analysis the of "Direct oblimin" rotation was used.[17] Oblimin rotation ensures maximization of variance of a variable under a particular factor.

For interpretation of the factors variables with high factor loading and high communality were taken into consideration. The limits were arbitrarily set; and the rotated factor (Oblimin) matrix of altmetric variables is presented in the Table 2.

a. Kaiser-Meyer-Olkin Measure of Sampling Adequacy – This measure varies between 0 and 1, and the values computed from all the variable is .868 which is closer to 1 is a good sign.

Table 1: Correlation coefficient r measures between altmetric variable vs citation.

Citation Vs Altmetric variables	Correlation (One tailed test)
Twitter	0.449
Facebook	0.461
Mendeley	0.127
Bolg	0.432
ResearchHig	0.278
Newsoutlet	0.464
Redditor	0.199
Wiki	0.209
Citeulike	0.091
Wiebo	0.167
Peerreview	0.153
Videoupload	0.204
Google	0.33
Policy	0.359
QAthread	0.047
Connotea	0.009
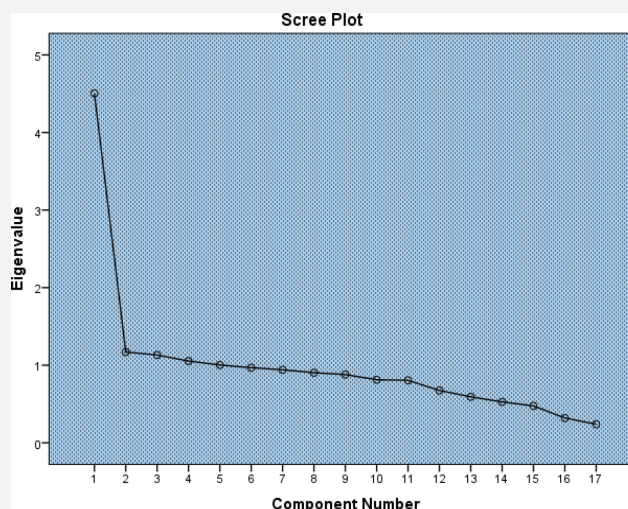
Figure 1: Scree plot shows the five variables having eigenvalue more than 1.

b. Bartlett's Test of Sphericity – This tests the correlation matrix is an identity matrix. An identity matrix is matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0. This is significant (less than 0.05), indicating that the correlation matrix is significantly different from an identity matrix, in which correlations between variables are all zero.

Table 2: Sampling adequacy and sphericity of the data.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.868
Bartlett's Test of Sphericity	Approx. Chi-Square	14905.671
	df	136
	Sig.	0.000

Taken together, above two result ascertains that sample data of the present study has a minimum standard which should be passed before a principal components analysis or a factor analysis.[18]

These initial communalities represent the relation between the variable and all other variables before rotation. Out of 17 variable 7 variables communalities are above (< 0.50)

The Rotated Component Matrix, which contains all the loadings (even those < 0.3) for each component, Table 3 is similar to the rotated factor matrix in Output Figure 2. The Component Plot in rotated Space gives one a visual representation of the loadings plotted in a 2–dimensional space. The plot shows how closely related.

Rotated Component Matrixa Rotation Method clearly explains the cluster formation pattern in three component and how each variables are loading.

Subject wise analysis

Subject wise correlation analysis Table 4 was performed between Altmetric variables and Citation . From the table value for the subject "Clinical Medicine" it was found that apart from QAthread and Cannotea all other variables was significantly correleated with citation at 0.0l level, the association between mendely readership count and citation was found to be very strong. In case of "Microbiology" subject; same pattern of "Clinical Medicine" was found. The trend in Molecular Biology was also same; exculding the variable Policy which showed a weak assocation compared to other variables. Apart from Connotea all other variables was strongly correlated with citation in case "Neuro Science" subject.

Table 3: Pattern matrix displays how individual variables are loading differently in each component.

Pattern Matrix					
	Component				
	1	2	3	4	5
Facebook	0.842				
Twitter	0.680				
Citation	0.639				
Google	0.632				
Policy	0.467			0-.361	
Redditor	0.360	0-.326			
Videoupload		0-.662			
Wiebo		0-.636			
Bolg	0.422	0-.575			
Newsoutlet	0.468	0-.503			
Wiki		0-.426			
Citeulike			0.611		
ResearchHig	0.484		0.485		
Mendeley			0.484		0.405
QAthread				0-.744	
Peerreview				0-.542	
Connotea					0.914

Figure 2: Component plot rotated space explains the loading pattern of each variables.

Table 4: Subject wise correlation coefficient r measures between altmetric variable vs citation.

	Citation			
	Clinical Medicine	**Micro Biology**	**Molecular Biology**	**Neuro Science**
Altmetric	0.574**	0.430**	0.563**	0.527**
Twitter	0.547**	0.386**	0.559**	0.475**
Facebook	0.386**	0.218**	0.393**	.314**
Mendeley	0.724**	0.605**	0.737**	0.724**
Bolg	0.441**	0.304**	0.462**	0.418**
ResearchHig	0.312**	0.306**	0.341**	0.198**
Newsoutlet	0.463**	0.394**	0.417**	0.406**
Redditor	0.129**	0.141**	0.156**	0.155**
Wiki	0.232**	0.276**	0.296**	0.226**
Citeulike	0.305**	0.350**	0.518**	0.430**
Wiebo	0.268**	0.269**	0.326**	0.266**
Peerreview	0.143**	0.060	0.182**	0.081*
Videoupload	0.122**	0.099*	0.158**	0.131**
Google	0.306**	0.184**	0.367**	0.230**
Policy	0.174**	0-.002	0.059	0.078*
QAthread	0.053*	0.052	0.127**	0.078*
Connotea	0.030			

RESULTS AND DISCUSSION

From the present study it is inferred that the new media tools are penitrating into the traditional area of citation which was considered the tool for evaluating scientific literature. Now this new media tools are complimenting to the citation count. The study indicates that almost all the sample authors article have at least one almertic measure.[19] The results provide strong evidence Twitter, Facebook, Blog, Mendeley , Google

+ are strongly associate with citation counts.[3] Discussion about a particular article on social media is short lived but its contrary to citation.[20] The trend ofnew media variable with four subjects are unbiased; more or less all the social media variable are correlated with the four subject in same manner. It was found that among the new media tools Mendeley readership count is storngly correalted with Citation count. [11,12,21] Though new media tools are highly used in various environment to disseminate information; still these channals are in nascent stage. From the data its evident that all the altmetric variable do not have same kind of proximity with the sample subject area; baring few variables. Hence at this junture visibility of citation is more compared to altmetric variables and its apparent that citation cout is more authentic as it grows with the time on other hand the new media variable presence is only in the initial stage and the discourse on this platform decreased over the years. So the altmetric variables can be campared with litmus paper which helps to find a solutions acidic nature; same way these new media tools can help the authors and publisher to prediect how much citation an article can attract in future [22] according to attention economics theory.[23]

CONCULSION

The new media brings the content to its reader; subsequently this channel opens a bridge to publish the views of a reader. While citation is one way path but its considered more authentic in assessing the impact of an article; in same manner altmetric be regarded as authentic and unobtrusive.[24] Altmetric data are hetrogenious in nature and its not clear in what way they represents engagements with scholarly works.[25] Till now these social media tools are notembraced by academicans; only citation and h–index are used for their career growth, still the altmetric variables has not been eastablsied as sub domain of bibliometrics.[5] Early research had proved that altmetrics variable has a significant association with citation but it differ subject wise. Early studies had conculed that science subjects are more popular in new media platform. The study further assertains the postive correlation for all four study subject domain; but medical science attract more attention in the social media.[25,26] The study can be expanded by including more prolific author and articles and to explore disciplinary differences.[27]

REFERENCES

1. Kogleck L, Priyadarshini S, Pincock S, Bocquet A, Gilloch C. A Nature Index analysis ABOUT NATURE PUBLISHING GROUP (NPG). In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016 [cited 2017 Mar 13]. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

2. Research outputs 2016 An overview of outcomes and impacts of EPSRC-supported re. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016 [cited 2017 Mar 13]. Available from: http://www.informationr.net/

ir/22-1/colis/colis1626.html

3. Thelwall M, Haustein S, Larivière V, Sugimoto CR. Do Altmetrics Work? Twitter and Ten Other Social Web Services. Bornmann L, editor. PLoS One [Internet]. 2013 May 28 [cited 2017 Mar 6];8(5):e64841. Available from: http://dx.plos.org/10.1371/journal.pone.0064841

4. Moed HF. Citation analysis in research evaluation [Internet]. Springer; 2005 [cited 2017 Mar 14]. 346 p. Available from: https://books.google.co.in/books?hl=en&lr=&id=D9SaJ6awy4gC&oi=fnd&pg=PR9&dq=%22research+evaluation%22&ots=FGp1Nz3QkW&sig=d9uFhTBxEjlPbR7J5l8tLXeE0-M#v=onepage&q=%22research evaluation%22&f=false

5. Bornmann L. Do altmetrics point to the broader impact of research ? An overview of benefits and disadvantages of altmetrics. J Informetr [Internet]. 2014;8(4):895–903. Available from: http://dx.doi.org/10.1016/j.joi.2014.09.005

6. Poplašen LM, Grgic IH. Altmetric and Bibliometric Scores : Does Open Access. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016. p. 451–60. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

7. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res [Internet]. 2011 [cited 2016 Aug 23]; Available from: http://www.jmir.org/2011/4/e123/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+JMedInternetRes+(Journal+of+Medical+Internet+Research+(atom))

8. Ortega L. with the tweeting and citation of research papers. 2016;(Ortega, L. (2016). with the tweeting and citation of research papers, 1353–1364. https://doi.org/10.1007/s11192-016-2113-0):1353–64.

9. Ringelhan S, Wollersheim J, Welpe IM, Iamarino A, Nahas T, Munger D. I Like, I Cite? Do Facebook Likes Predict the Impact of Scientific Work? In: Dorta-González P, editor. Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016 [cited 2017 Mar 14]. p. e0134389. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

10. Mohammadi E, Thelwall M. Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. J Assoc Inf Sci Technol [Internet]. 2014 Aug [cited 2017 Mar 14];65(8):1627–38. Available from: http://doi.wiley.com/10.1002/asi.23071

11. Shrivastava R, Mahajan P. New Library World Relationship between citation counts and Mendeley readership metrics A case of top 100 cited papers in Physics. Phys New Libr World New Libr World New Libr World. 11710(3):564–77.

12. Dorta-González P, Dorta-González M. Impact maturity times and citation time windows: The 2-year maximum journal impact factor. 2013 [cited 2017 Mar 30]; Available from: https://arxiv.org/ftp/arxiv/papers/1304/1304.5101.pdf

13. Haustein S, Peters I, Jason JB. Coverage and adoption of altmetrics sources in the bibliometric community. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016. p. 1145–63. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

14. Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz J. Research Data Explored : Citations versus Altmetrics.

15. Costas R, Zahedi Z, Wouters P. Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. J Assoc [Internet]. 2015 [cited 2016 Aug 22]; Available from: http://onlinelibrary.wiley.com/doi/10.1002/asi.23309/full

16. Ray GLS mondal. Research Mthods in Social Sciences and Extension Education. 2nd ed. New Delhi: Kalyani Publishiers; 1999. 1-292 p.

17. Ruch W, Weber M, Park N, Peterson C. Character strengths in children and adolescents: Reliability and initial validity of the german values in action inventory of strengths for youth (German VIA-Youth). Eur J Psychol Assess. 2014;30(1):57–64.

18. Principal Components Analysis | SPSS Annotated Output - IDRE Stats [Internet]. [cited 2017 May 24]. Available from: https://stats.idre.ucla.edu/spss/output/principal-components-analysis/

19. Costas R. Do " Altmetrics " Correlate With Citations ? Extensive Comparison of Altmetric Indicators With Citations From a Multidisciplinary Perspective. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016. p. 2003–19. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

20. Haustein S, Costas R, Larivière V, Cole F, Eales N, Gross P, et al. Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. Glanzel W, editor. PLoS One [Internet]. 2015 Mar 17 [cited 2016 Aug 23];10(3):e0120495. Available from: http://dx.plos.org/10.1371/journal.pone.0120495

21. Haustein S, Bowman TD, Costas R. Interpreting " altmetrics ": viewing acts on social media through the lens of citation and social theories. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016. p. 1–24. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

22. Brody T, Harnad S, Carr L. Earlier Web usage statistics as predictors of later citation impact. J Am Soc Inf Sci Technol [Internet]. 2006 Jun 1 [cited 2017 Jun 21];57(8):1060–72. Available from: http://doi.wiley.com/10.1002/asi.20373

23. Kortelainen TMK and A-SL. Attention and altmetrics. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

24. Haustein S. Grand challenges in altmetrics: heterogeneity, data quality and dependencies. In: Proceedings of the Ninth International Conference on Conceptions of Library and Information Science, Uppsala, Sweden, June 27-29, 2016 [Internet]. Information Research; published quarterly by the university of borås, sweden; 2016 [cited 2017 Mar 28]. p. 413–23. Available from: http://www.informationr.net/ir/22-1/colis/colis1626.html

25. Torres-Salinas D, Robinson-Garcia N, Jiménez-Contreras E. Can we use altmetrics at the institutional level? A case study analysing the coverage by research areas of four Spanish universities. In: International Conference on Science and Technology Indicators [Internet]. Valencia; 2016 [cited 2017 Mar 28]. Available from: http://dx.doi.org/10.4995/STI2016.2016.xxxx

26. Thelwall M, Wilson P. Mendeley Readership Altmetrics for Medical Articles : An Analysis of 45 Fields. J Assoc Inf Sci Technol [Internet]. 2016 Aug 1 [cited 2017 Jun 22];67(8):1962–72. Available from: http://doi.wiley.com/10.1002/asi.23501

27. Alhoori H, Giles CL. On the Relationship between Open Access and Altmetrics. 2015 [cited 2017 Jun 22]; Available from: https://www.ideals.illinois.edu/bitstream/handle/2142/73451/212_ready.pdf?sequence=2