A Bibliometric Analysis of Global Research Trends on Social Tagging over the Last Fifteen Years

Arindam Sarkar and Udayan Bhattacharya

Bibliometric analysis of publications helps to quantify researchers' interests in a specific research area. This method enables the identification of year wise growth of literature, geographical distribution of research, authorship pattern of publications and degree of collaborations among the authors etc. It is important to define research trends, which have developed in this specific field. It enables to determinate the main research topics, discoveries and correlations in the global knowledge network. The primary aim of this study was to identify the trends of research regarding the topic social tagging based on Scopus database from 2004-2018.

Keywords: *Bibliometric, Social tagging, Scopus, Authorship pattern, Knowledge network*

Introduction

Tagging, also known as social tagging or user tagging or collaborative tagging, has gained in mileage since the first social bookmarking system named del.icio.us was started in 2003. As an application of web 2.0 social tagging increases access points; more entry points, that are helpful for easy retrieve of resources i.e. documents. The annotated resources can be of any type or in any format, such as videos (e.g. YouTube), photos (e.g. Flickr), academic papers (e.g. CiteULike), books (e.g. LibraryThing) and so on. Social tagging can be viewed as a technique, by which many users add metadata in the form of keywords, genres, subjects to shared content or resources (Golder and Huberman, 2006). Social tagging is different from controlled process, social tagging is done in a totally uncontrolled environment. Social taggers do not

Arindam Sarkar, Research Scholar, Department of Library and Information Science, Jadavpur University, Kolkata- 700032.
E-Mail: infoarindam83@gmail.com

Prof. Udayan Bhattacharya, Professor, Department of Library and Information Science, Jadavpur University, Kolkata- 700032.
E-Mail: udayanbhattacharya1967 @hotmail.com

need to any expertise rules for tagging. They apply their own verbal descriptors to resources that interest them. Though social tags or controlled vocabularies have the common purpose of helping users share, store, organize and retrieve the resources they are interested in. The bibliometric analysis was first used by Pritchard in 1969 as a method of mathematical and statistical analysis applied to books, organizations, persons and other communication media. Citation and content analyses are now widely used bibliometric techniques. Bibliometrics is an effective and important tool to determine the trends of research in various fields of science as well as social science. Bibliometric studies include a number of quantitative and visual procedures to generalize patterns and dynamics of publications.

The goal of the work was to identify the trends of research regarding the topic social tagging based on Scopus database from 2004-2018. The method of bibliometric analysis was applied.

Research Questions

Research questions for this study are as follows:

RQ1. How many papers on social tagging were published between 2004 and 2018? Is there any trend in research activity?

RQ2. What publication types and what languages dominate research output?

RQ3. Which countries all over the world are notably active in this area of research?

RQ4. What is the authorship pattern of related to social tagging research? What are the degrees of collaboration among authors?

RQ5. What are the chronological distributions of citations of published documents within the selected period of study?

Scope and Limitation

The study is restricted within a particular database, i.e. Scopus.com (Scopus, 2019). In this study the documents published within 2004 to 2018 were collected.

Methodology

To find out the objectives of this study a general bibliometric process has been used. As a registered user after log in within Scopus database an advanced search was done at 11.30 a.m. on 18th March 2019 by using a search string TITLE (social tag) AND PUBYEAR > 2003 AND PUBYEAR < 2019 (Noruzi,2017). Also by using another search string TITLE (social tag) AND PUBYEAR > 2003 AND PUBYEAR < 2019 AND SUBJAREA (SOCI) for subject area Social Science, data have been collected on that specific domain. After retrieval, data have been collected, then analysed and tabulated keeping in view the objectives of the study.

Data Analysis and Findings

In the following few paragraphs retrieved data are presented and analysed through some tables and figures.

RQ 1: *Year-wise distribution of Publication*

Year wise distribution of publications helps us to identify the research trends regarding a particular topic.

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Only Social science	1	0	3	0	2	3	7	6	7	7	6	6	6	1	8
All domain	1	2	3	7	32	49	57	56	56	53	42	30	29	25	35

Figure 1: Year-wise distribution of Publication

Figure-1 reveals the year wise distribution of publications which were based on social tagging. As per this figure, total 477 number of publications were identified in Scopus. Among the total 477, the highest number of publications (57) were in 2010 and lowest number of publications (1) in 2004. In case of social science domain, total 63 number of publications were identified and out of 63 highest in 2018 and lowest in 2004. In case of year 2007 and 2005 there were no publications.

RQ 2: ***Language wise and Document-type wise distribution of Publication***

Following Table-1 reveals the language wise distribution of published documents.

Table-1: Language Wise Distribution of Published Documents

Language	For All Domain	Only Social Science
English	464	62
Chinese	6	-
German	4	1
French	1	-
Spanish	1	-
Turkish	1	-
TOTAL	477	63

From the above figure we can conclude that, among the published documents mostly were in English (464) and then followed by Chinese (6), German (4), French (1), Spanish (1) and Turkish (1). Also in case of Social Science (Figure-4B) the published documents were in English (62) and only one (1) in German.

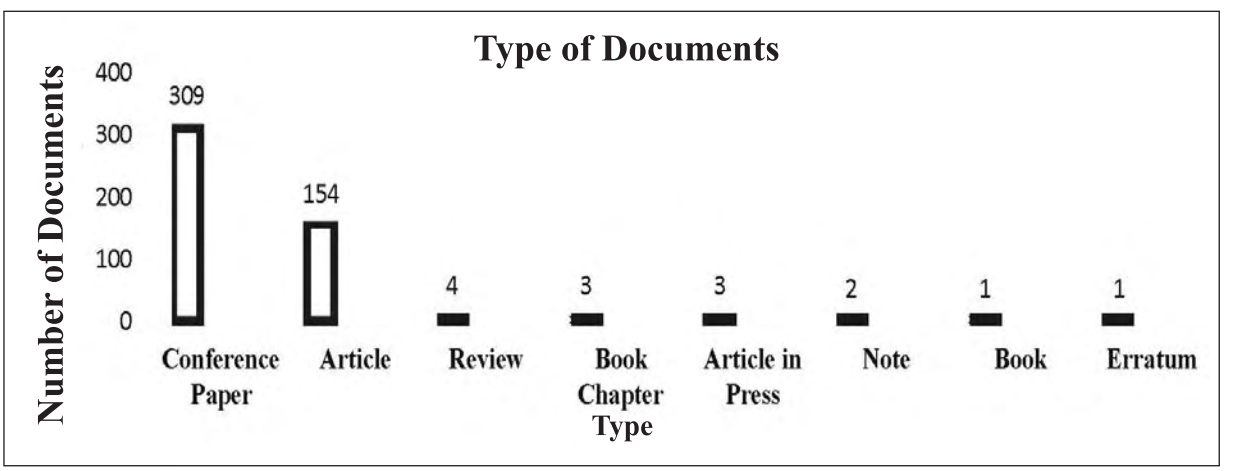

Figure 2: Document-type wise distribution of Publication

"One of the main issues in compiling bibliometric data is the choice of the types of documents to include" (Archambault, Campbell, Gingras, & Larivière, 2009, p. 1321), and Figure 2 shows the differences in document type distributions for RQ 2.

RQ 3: ***Country wise distribution of publication***

Following figure 3 measures the published documents according to affiliations of the authors. Researchers from China top the list with 155 publications on social tagging in their works while USA hold the 2nd position with 84 and Germany and Singapore jointly hold the 3rd ranks with 30 publications.

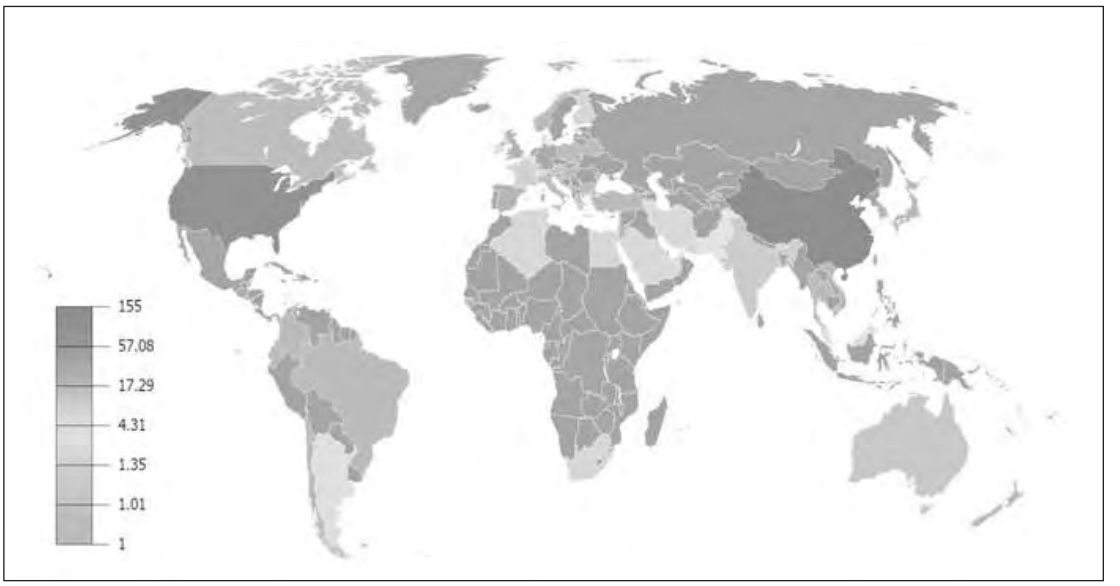

Figure-3: Heat map of countries with active research (Red: Top most, Green: Lowest and Grey: countries with no publications on social tagging). Tool: GunnMap

RQ 4: ***Distribution of Authorship Pattern and Degree of Collaboration among the Authors***

Table-2: Distribution of Authorship Pattern

Year	Single Author	More than one Author	Total Author	Year	Single Author	More than one Author	Total Author
2004	1	0	1	2012	5	51	56
2005	0	2	2	2013	1	52	53
2006	1	2	3	2014	3	39	42
2007	0	7	7	2015	0	30	30
2008	2	30	32	2016	1	28	29
2009	3	46	49	2017	2	23	25
2010	7	50	57	2018	7	28	35
2011	3	53	56				

From the above table (Table-2) it becomes clear that only 7.55% documents were published by single author whereas 92.45% articles contributed by more than one authors. So, this table concluded that multiple authorship has become a prominent issue in the 21st century research regarding this topic.

Degree of collaboration among the authors

The following table and paragraph describes the degree of collaboration among the authors. In this study the Degree of Collaboration (C) of the contributors has been calculated using the Subramanyam formula. The formula is as follows:

Degree of Collaboration **(C) = Nm / Nm+Ns**

Where,

C = Degree of Collaboration

Nm = Number of multiple authors

Ns = Number of single authors

Table-3: Degree of collaboration among the authors

Year	Single Authors (Ns)	Multiple Authors (Nm)	Total (Ns+Nm)	Degree of Collaboration (C)
2004	1	0	1	0
2005	0	2	2	1
2006	1	2	3	0.66
2007	0	7	7	1
2008	2	30	32	0.93
2009	3	46	49	0.93
2010	7	50	57	0.87
2011	3	53	56	0.94
2012	5	51	56	0.91
2013	1	52	53	0.98
2014	3	39	42	0.92
2015	0	30	30	1
2016	1	28	29	0.96
2017	2	23	25	0.92
2018	7	28	35	0.8
Total	36	441	477	0.92

So, Degree of Collaboration (C) = Nm / Nm + Ns,

Here, Degree of Collaboration (C) = 441 / 36+441,

= 441/477

= 0.92

Above table shows the individual year wise Degree of Collaboration and as a whole (from the year 2004 to 2018) Degree of Collaboration. In this case Nm = 441 and Nm + Ns= 477 and Degree of Collaboration (C) = 0.92 that indicates large number of collaborative works among the authors.

RQ 5: ***Year wise / Chronological distribution of citations***

Table-4: Chronological distribution of citation

Year	Total Number. of Citation	Year	Total Number. of Citation
2004	1	2012	288
2005	3	2013	682
2006	61	2014	328
2007	142	2015	349
2008	1758	2016	171
2009	1138	2017	113
2010	840	2018	7
2011	760		

The above table demonstrates that year 2008 has received maximum number of citations i.e. 1758. Using Pearson correlation formula the relationship among year and number of citation can be tested. Pearson correlation is a test used to know the correlation (degree of association) between two variables. In this study correlation has been observed between time (year) and growth of citation (number of citation).

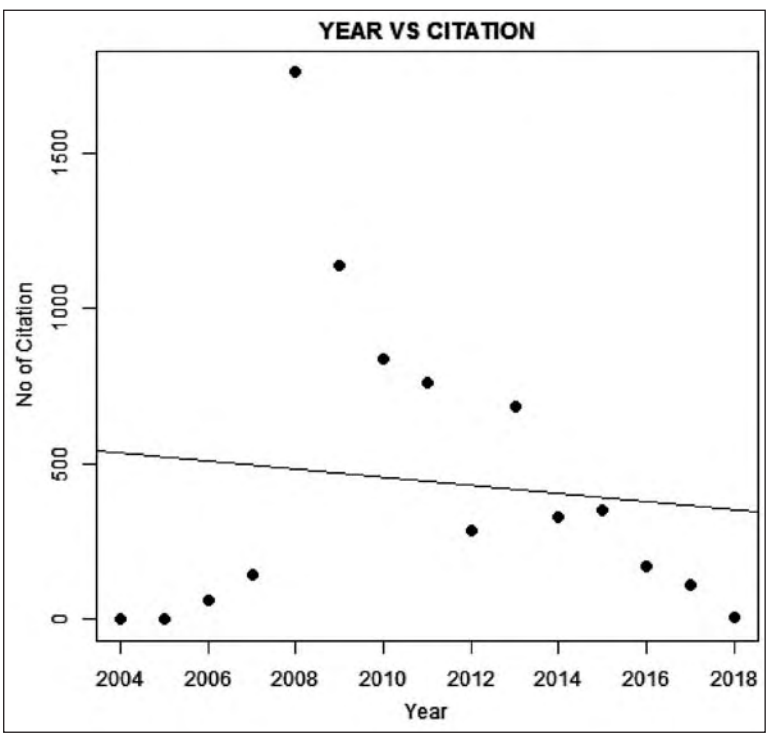

Figure- 4: Relationship between Year Vs Citation (Using R Statistical Software)

Pearson correlation as per R-Statistical Software:

> Year<- c (2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018)
> Citation<- c (1, 3, 61, 142, 1758, 1138, 840, 760, 288, 682, 328, 349, 171, 113, 7)
>cor (Year, Citation, method="pearson")
[1] -0.1161724

There is a significant negative relationship between time and growth of citation. In this case Pearson's **r =-0.1161724**. This figure also shows the negative correlation between two variables.

Conclusion

The study examines the global research trends on social tagging as indexed by Scopus database during the year 2004-2018. Most of the research contributions were multiple authors' articles. Degree of collaboration among the authors was 0.92. Most of the contributing authors are from China. Regarding citation time has an inverse relationship with the growth of citation. This study will helps to understand the trends of research activities on social tagging in the global perspective.

References

Archambault, E., Campbell, D., Gingras, Y., & Lariviere, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology, 60*(7), 1320-1326.

Golder S. & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems, *Journal of Information Science,* 32(2), 198–208.

Kumar, M. (2014). Library Herald journal: A bibliometric study. *Journal of Education & Social Policy, 1* (2), 123-134. Retrieved March 21, 2019 from http://jespnet.com/journals/Vol_1_No_2_December_2014/17.pdf

Noruzi, A. (2017). YouTube in scientific research: A bibliometric analysis. *Webology*, *14*(1), Retrieved from http://www.webology.org/2017/v14n1/editorial23.pdf

Scopus (2019). *Scopus*. Retrieved March 18, 2019, from https://www.scopus.com

Subramanyam, K. (1983). Bibliometric studies of research in collaboration: A review. *Journal of Information Science, 6* (1), 33-38.