Bibliographic Analysis of *Nature* Based on Altmetrics

Xiaoyan Su

DUT MSCLab

Contents

1/Introduction

2/Data and Methods

3/Results and discussion

4/Conclusion

Impact of academic publications

- Citation
- H-index or derivative index
- Data from social media platforms

Altmetrics

As a generalization of article level metrics, altmetrics can assess the **popularity or social impact** of publications based on data collected by social media platforms.
Compared with the traditional citation based metrics, altmetrics can **reduce the delay** for accumulation and cover **new forms** of scholarly content (e.g., datasets, software, and research blogs) to achieve more **broad**, **diversiform**, and **rapid** impact analysis.

Related work for Altmetrics

1. Representativeness（代表性） and validity（有效性） of data for Altmetrics

Altmetrics are a very broad group of metrics.

Classification:

- Viewed - HTML views and PDF downloads
- Discussed - journal comments, science blogs, Wikipedia, Twitter, Facebook and other social media
- Saved - Mendeley, CiteULike and other social bookmarks
- Cited - citations in the scholarly literature, tracked by Web of Science, Scopus, CrossRef and others
- Recommended - for example used by F1000Prime

M. Thelwall, S. Haustein, V. Larivire, and C. R. Sugimoto, “Do altmetrics work? twitter and ten other social web services,” PLoS ONE,vol. 8, no. 5, p. e64841, 05 2013.
Z. Zahedi, R. Costas, and P. Wouters, “How well developed are altmetrics? a cross-disciplinary analysis of the presence of alternative metrics in scientific publications,” Scientometrics, vol. 101, no. 2, pp. 1491–1513, 2014.
P. Wouters and R. Costas, Users, narcissism and control: tracking the impact of scholarly publications in the 21st century. Utrecht: SURFfoundation, 2012.

Related work for Altmetrics

2. Correlation between citations and various social media event counts（citation和社会数据之间的关系）

Citation VS Social media data

Social media data VS Social media data

Can social media data predict citation?

Whether both types of metrics measure similar concepts?

X. Shuai, A. Pepe, and J. Bollen, "How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations," PLoS ONE, vol. 7, no. 11, p. e47523, 2012.
S. Haustein, I. Peters, C. R. Sugimoto, M. Thelwall, and V. Larivi` ere, "Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature," Journal of the Association for Information Science and Technology, vol. 65, no. 4, pp. 656–669, 2014.
G. Eysenbach, "Can tweets predict citations ? metrics of social impact based on twitter and correlation with traditional metrics of scientificimpact," Journal of medical Internet research, vol. 13, no. 4, 2011.

Limitations

1. ignore the influence of **journal**, **discipline** and **publication date** on the validity of altmetrics.

2. do not analyze the correlation across **disciplines** for a **comprehensive scientific magazine**.

3. do not explore the correlation by **publication year** and **role of social user**.

My work

- **Representativeness（代表性）** and **validity（有效性）** of Twitter and Facebook as data sources of Altmetrics
- **Correlation** between citations and tweets（citation 和社会数据之间的关系）

Innovation points

- Based on a comprehensive scientific magazine *Nature*
- Consider the impact of publication year and discipline for the analysis
- Relatively long time (2010-2015)
- Consider the Twitter user type at the first time

Contents

1/Introduction

2/Data and Methods

3/Results and discussion

4/Conclusion

Part 2 Data and Methods

Data

We downloaded the metadata for all **Nature research papers** from the online literature database over the period between **January 2010 and June 2015**, including **title**, **publication date**, **discipline**, **keywords**, accumulated number of **tweets, Twitter user types** and **Facebook posts** from nature.altmetric.com and **citations** from the Web of Science.

TABLE I: Statistics of Nature publications in the data set

	2010	2011	2012	2013	2014	2015	Total
Biology Sciences	366	576	572	557	560	254	2876
Chemical Sciences	101	161	132	25	32	22	473
Earth & Environment Sciences	78	113	112	96	95	32	526
Physical Sciences	113	146	171	160	165	76	831
Total	552	789	846	846	842	401	4276

Part 2 Data and Methods

Data

Disciplines:

Biology sciences
Chemical sciences
Earth & environment sciences
Physical sciences

Twitter user types:

Member of the public: somebody who doesn't link to scholarly literature and doesn't otherwise fit any of the categories below.
Scientist: somebody who is familiar with the literature.
Practitioner: a clinician, or researcher who is working in clinical science.
Science communicator: somebody who links requently to scientific articles from a variety of different journals or publishers.

Part 2 Data and Methods

Methods

In order to evaluate the representativeness and validity of **Twitter** and **Facebook** as data sources for altmetrics, we analyze the **distribution** of academic information about Nature articles on Twitter and Facebook.

Definition II.1. ***Twitter Coverage*** $\mathrm{Cov}^{\mathrm{T}}$ *is defined as the proportion of articles tweeted at least once, i.e.,*

$$\mathrm{Cov}^{\mathrm{T}} = \frac{N^{\mathrm{T}}}{N} \tag{1}$$

where N is the total amount of articles for the analysis, and N^{T} is the amount of articles tweeted at least once.

Definition II.2. ***Facebook Coverage*** $\mathrm{Cov}^{\mathrm{F}}$ *is defined as the proportion of articles mentioned on Facebook at least once i.e.,*

$$\mathrm{Cov}^{\mathrm{F}} = \frac{N^{\mathrm{F}}}{N} \tag{2}$$

where N^{F} is the amount of articles mentioned on Facebook at least once.

Part 2 Data and Methods

Methods

In order to evaluate the representativeness and validity of **Twitter** and **Facebook** as data sources for altmetrics, we analyze the **distribution** of academic information about Nature articles on Twitter and Facebook.

Definition II.3. ***Twitter Mention Rate*** MR^{T} *is defined as the mean number of tweets per tweeted paper, i.e.,*

$$\mathrm{MR}^{\mathrm{T}} = \frac{\sum_{i=1}^{N} C_i^{\mathrm{T}}}{N^{\mathrm{T}}} \tag{3}$$

where C_i^{T} *is the tweeted count of the paper i.*

Definition II.4. ***Facebook Mention Rate*** MR^{F} *is defined as the mean number of Facebook posts per posted paper i.e.,*

$$\mathrm{MR}^{\mathrm{F}} = \frac{\sum_{i=1}^{N} C_i^{\mathrm{F}}}{N^{\mathrm{F}}} \tag{4}$$

where C_i^{F} *is the posted count of the paper i on Facebook.*

Part 2 Data and Methods

Methods

The **coverage** is used to evaluate the **concern degree** of social users on a Nature article and the development of the social media platform on the academic field.

The **mention rate** is used to examine the **impact** of a Nature article on a social media platform.

We also analyze the **relationship** between tweets and citations for Nature publications to determine whether both types of metrics measure similar concepts.

We evaluate the **Spearman correlation (measure of statistical dependence between two variables *S*)** between tweets and citations.

$0<S<1$ positive correlation
$-1<S<0$ negative correlation
$S=0$ uncorrelated
$|S|=1$ perfect monotone function

Contents

1/Introduction

2/Data and Methods

3/Results and discussion

4/Conclusion

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook

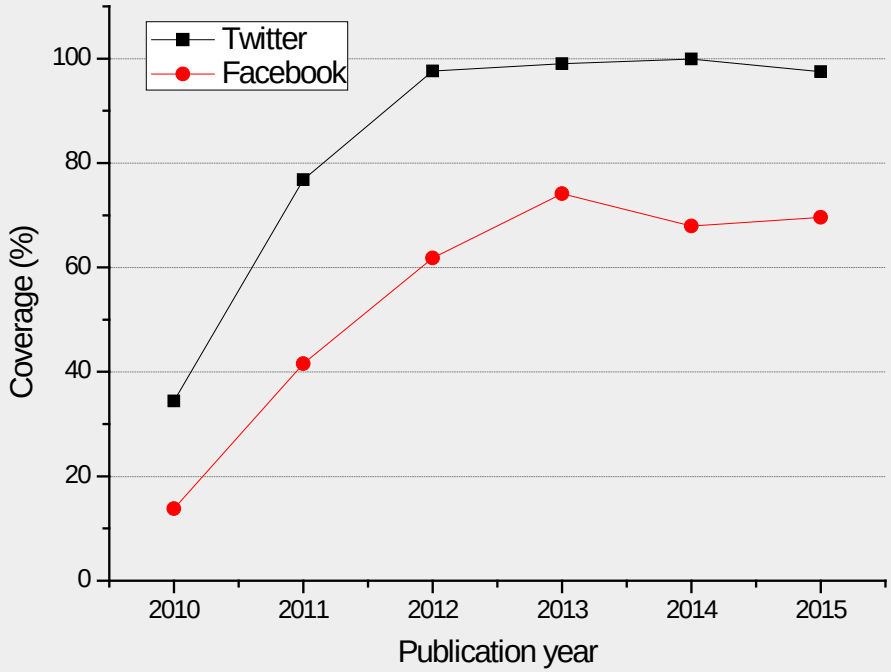


Fig. 1 Twitter and Facebook coverages by publication year

We can find that both Twitter users and Facebook users are interested in a **few Nature articles** published in 2010. As Twitter and Facebook evolve, social users increasingly focus on the scholarly documents, and thus Twitter and Facebook coverages show an **increasing trend** over the publication time. Twitter develops more **rapidly** than Facebook for the academic field.

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook

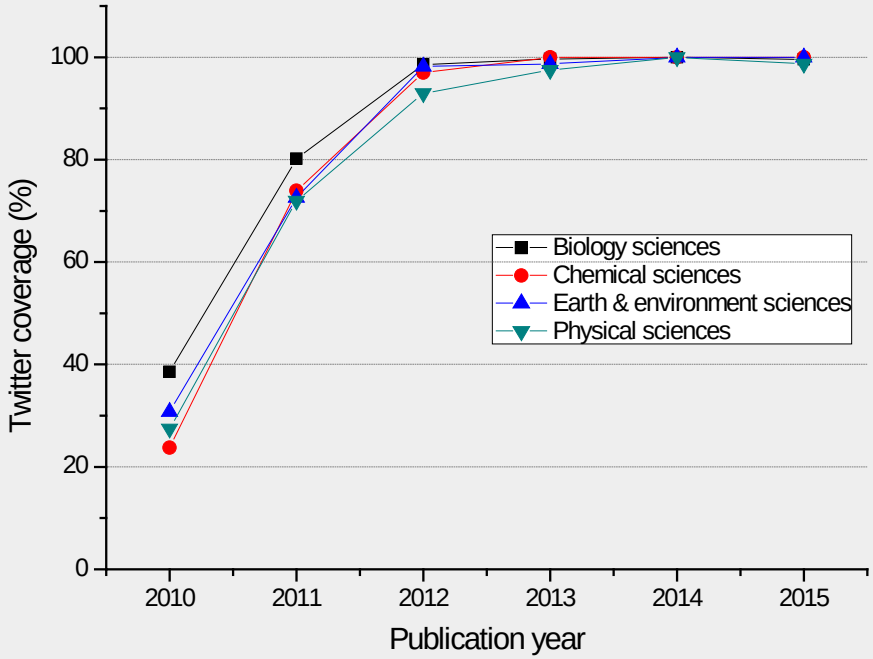


Fig. 2 Twitter coverage by publication year and discipline

Twitter coverage for **biology sciences** is significantly **higher** than other disciplines and Twitter coverage for other three disciplines show a similar lower growth trends. For Nature articles **published after 2012**, Twitter coverage for all disciplines approaches **100 percent**.

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook

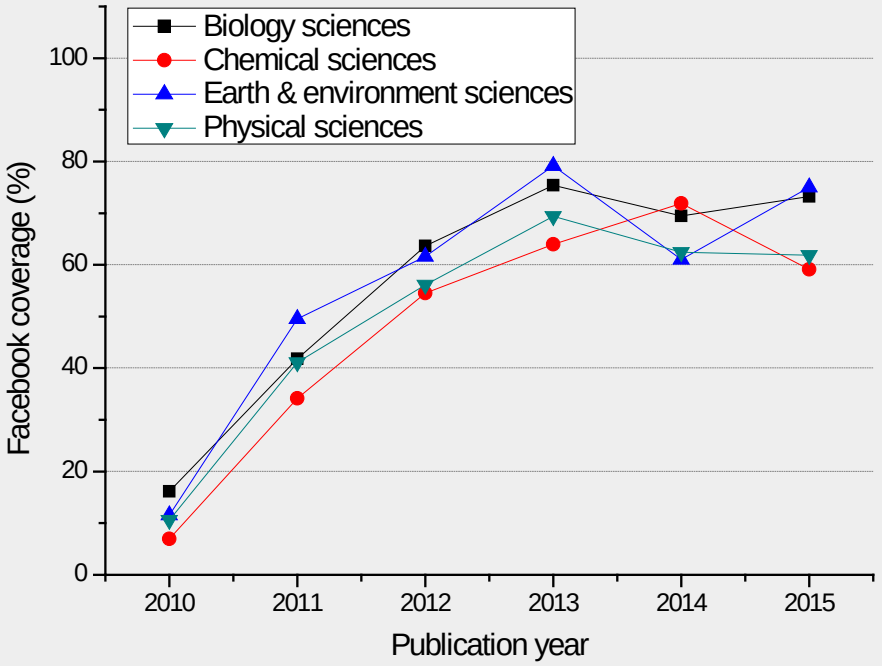


Fig. 3 Facebook coverage by publication year and discipline

Compared with Twitter coverage, the Facebook coverage **differences** among distinct disciplines are relatively **larger**.
For the articles which are **not published in 2014**, the Facebook has **a lower** coverage for **chemical sciences** than other disciplines and a relatively **high** coverage **for biology sciences and earth & environment sciences**.

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook

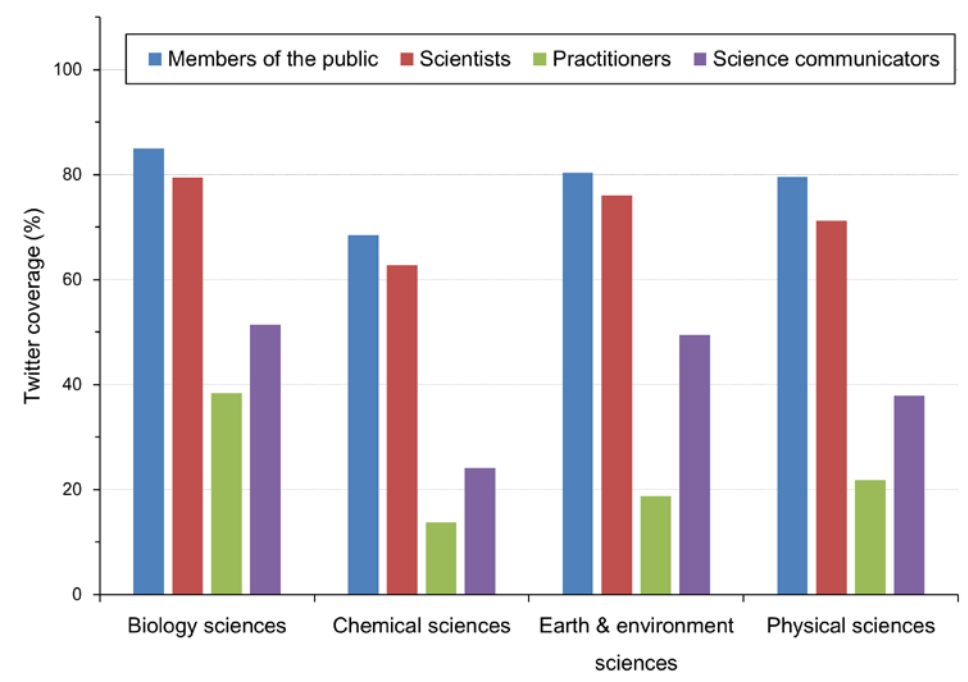


Fig. 4 Twitter coverage by user type and discipline

For all disciplines, **members of the public** have the **highest** concern degree.
Practitioners have the lowest concern degree.
Biology sciences draw more concern degree of four user types.

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook

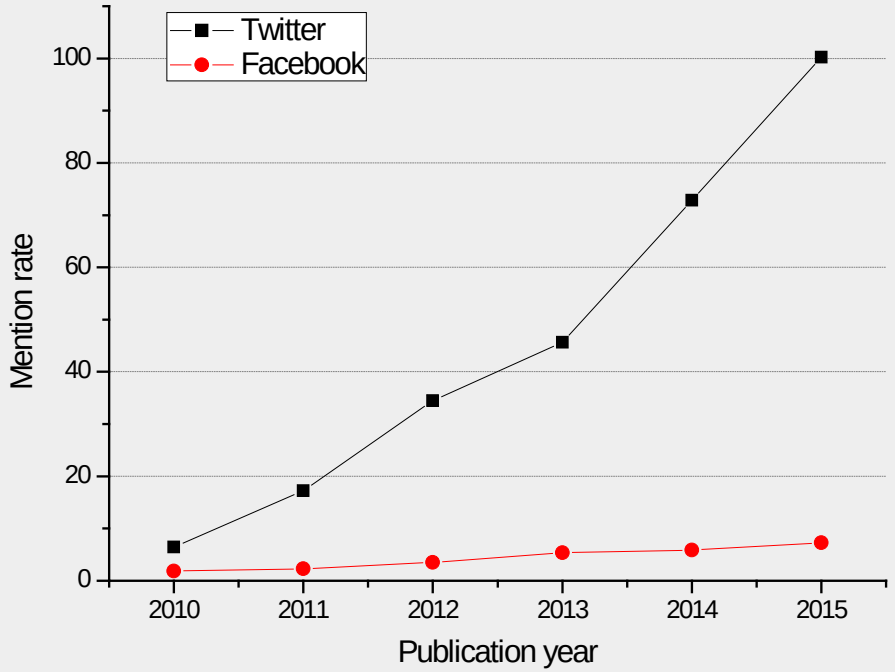


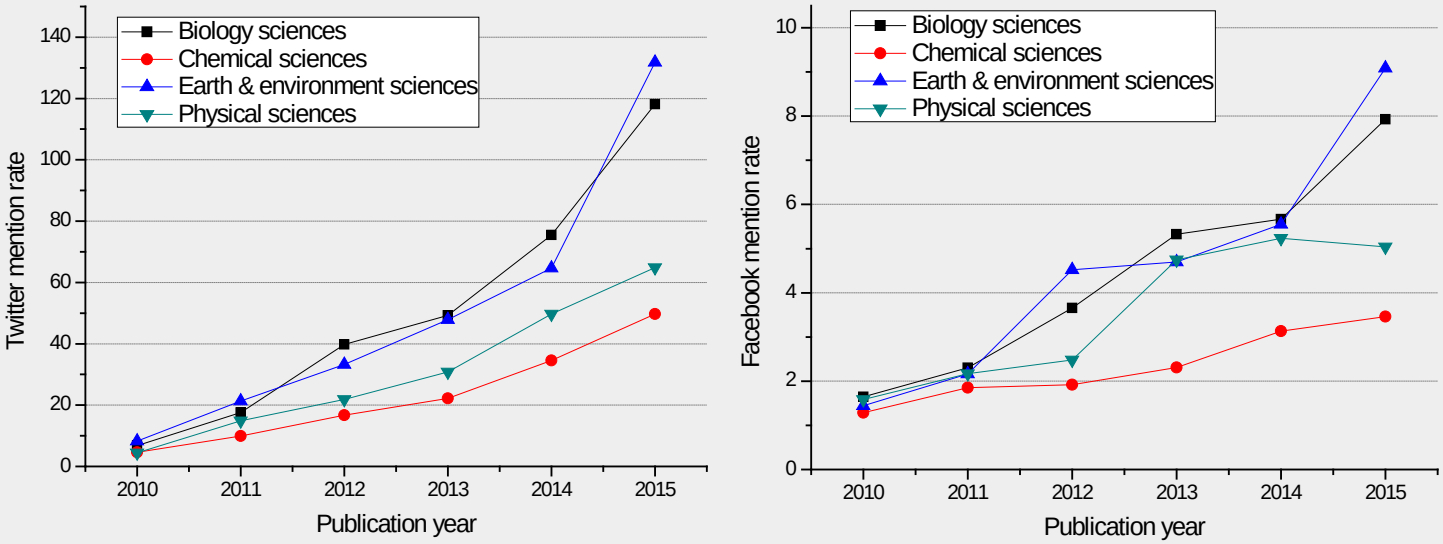
Fig. 5 Twitter and Facebook mention rate by publication year

There is a **continuous growth** for both Twitter and Facebook mention rates because of the developmet of social media platforms.
Compared with Twitter, the growth of Facebook mention rate relatively **slow.**

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook



Twitter and Facebook mention rate by publication year and discipline

There is an **ascending** trend of both Twitter and Facebook mention rates for articles about all disciplines.
For all articles published from 2010 to 2015, we also can see that the articles about **biology sciences and earth & environment sciences** have **higher** Twitter and Facebook mention rate than the other two disciplines.

Part 3 Results and discussion

1. Distribution of academic information

A. Bibliographic Analysis Based on Twitter and Facebook

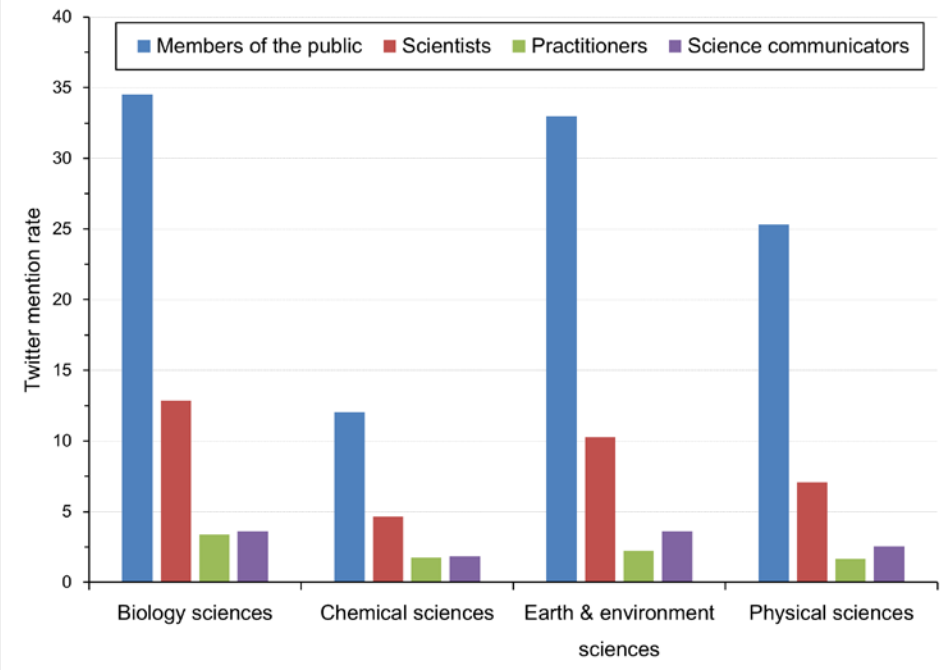


Fig. 8 Twitter mention rate by user type and discipline

For all disciplines, there is a highest impact on **members of the public**.
For members of the public, scientists and science communicators, the impact of the articles about **chemical sciences** is much **lower** than the articles of other three disciplines. Moreover, for all disciplines, there is a relatively small impact on **practitioners and science communicators**.

Part 3
Results
and
discussion

2. Relationship Analysis

B. Relationship Analysis between Tweets and Citations

	2010	2011	2012	2013	2014
Biology Sciences	0.258**	0.153**	**0.366****	0.365**	0.181**
Chemical Sciences	0.467**	0.065	**0.313****	0.35	**-0.22**
Earth & Environment Sciences	0.23*	0.298**	**0.42****	0.419**	0.301**
Physical Sciences	0.06	0.022	**0.233****	0.122	0.017
Total	0.259**	0.161**	**0.354****	0.321**	0.16**

* Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.01 level.

TABLE 2 Spearman correlation between tweets and citations

For the articles published from 2011 to 2014, the correlation coefficient shows **first increasing then decreasing** as the publication time passed. This finding suggests that the relationship analysis between tweets and citations can be influenced by changes in **Twitter use** and **citation delays**. Moreover, the correlation for the articles **about biology sciences and earth & environment sciences** is **positive** and there is a relatively **higher** positive correlation for papers of all disciplines **published in 2012**.

Part 3 Results and discussion

2. Relationship Analysis

B. Relationship Analysis between Tweets and Citations

		Biology Sciences		Chemical Sciences		Earth & Environment Sciences		Physical Sciences		Total	
		Correlation	Coverage	Correlation	Coverage	Correlation	Coverage	Correlation	Coverage	Correlation	Coverage
Member of the public	2010	0.184**	0.295	0.292**	0.149	0.151	0.231	0.106	0.195	0.205**	0.260
	2011	0.153**	0.739	0.073	0.634	0.293**	0.717	0.026	0.616	0.158**	0.692
	2012	0.324**	0.976	0.289**	0.970	0.404**	0.920	0.213**	0.912	0.318**	0.960
	2013	0.326**	0.995	0.416*	1.000	0.396**	0.979	0.144	0.956	0.299**	0.983
	2014	0.164**	0.995	-0.242	1.000	0.269**	1.000	0.018	0.994	0.132**	0.872
Scientist	2010	0.258**	0.210	0.366**	0.129	0.294**	0.179	0.068	0.150	0.239**	0.194
	2011	0.138**	0.681	0.06	0.627	0.238*	0.593	0.053	0.603	0.145**	0.635
	2012	**0.389****	0.935	0.235**	0.879	**0.424****	0.920	**0.219****	0.749	**0.367****	0.897
	2013	0.373**	0.939	0.2	0.840	0.379**	0.948	0.046	0.850	0.306**	0.916
	2014	0.211**	0.930	0.023	0.813	0.301**	0.979	0.006	0.867	0.185**	0.830
Practitioner	2010	0.177**	0.055	0.156	0.030	0.095	0.167	0.001	0.027	0.136**	0.045
	2011	0.231**	0.199	0.084	0.093	0.363**	0.097	0.029	0.075	0.248**	0.158
	2012	0.340**	0.421	0.207*	0.205	0.123	0.232	0.074	0.175	0.314**	0.348
	2013	0.348**	0.488	-0.111	0.160	0.14	0.229	0.006	0.231	0.315**	0.405
	2014	0.171**	0.513	0.174	0.250	0.246*	0.284	0.153*	0.376	0.185**	0.463
Science communicator	2010	0.023	0.055	0.115	0.020	0.195	0.064	0.208*	0.027	0.075	0.044
	2011	0.061	0.220	0.146	0.149	0.157	0.265	0.033	0.212	0.062	0.214
	2012	0.254**	0.615	0.255**	0.424	0.306**	0.625	0.144	0.392	0.245**	0.567
	2013	0.321**	0.628	**0.519****	0.360	0.39**	0.604	0.2*	0.481	[illegible]	0.585
	2014	0.169**	0.721	0.014	0.438	0.166	0.758	0.07	0.564	[illegible]	0.597

* Correlation is significant at the 0.05 level. ** Correlation is significant at the 0.01 level.

TABLE 3 Spearman Correlation between Tweets and Citations by Twitter User Type

The **Twitter user type** and the discipline have a **great influence** on correlation between tweets and citations.

Contents

1/Introduction

2/Data and Methods

3/Results and discussion

4/Conclusion

Part 4 Conclusion

Five important conclusions

This study presents a great many findings, but five are perhaps especially salient:

- The development of social media platforms makes people **more interested** in academic information.
- Twitter users have a higher and faster-growing **concern degree** on the Nature articles compared with Facebook users.
- Nature articles have higher and faster-growing **impact** on Twitter than on Facebook
- The **correlation** between tweets and citations for Nature articles is **positive** and appears quite sensitive to the publication date, discipline and Twitter user type.
- Although tweets and citations are somewhat related, they mostly measurea different type of impact.

Thank you

谢谢大家