What does Twitter Measure? Influence of Diverse User Groups in Altmetrics

Simon Barthel, Sascha Tönnies, Benjamin Köhncke, Patrick Siehndel, Wolf-Tilo Balke
L3S Research Center
Hannover, Germany
{barthel, toennies, koehncke, siehndel, balke}@l3s.de

ABSTRACT

The most important goal for digital libraries is to ensure high quality search experience for all kinds of users. To attain this goal, it is necessary to have as much relevant metadata as possible at hand to assess the quality of publications. Recently, a new group of metrics appeared, that has the potential to raise the quality of publication metadata to the next level – the altmetrics. These metrics try to reflect the impact of publications within the social web. However, currently it is still unclear if and how altmetrics should be used to assess the quality of a publication and how altmetrics are related to classical bibliographical metrics (like e.g. citations). To gain more insights about what kind of concepts are reflected by altmetrics, we conducted an in-depth analysis on a real world dataset crawled from the Public Library of Science (PLOS). Especially, we analyzed if the common approach to regard the users in the social web as one homogeneous group is sensible or if users need to be divided into diverse groups in order to receive meaningful results.

Categories and Subject Descriptors

H.3.7 [**Information Systems**]: Digital Libraries—*Standards*; H.3.3 [**Information Systems**]: Information Search and Retrieval

Keywords

Altmetrics, Twitter, Correlation Analysis, Social Media, Expert Mining

1 INTRODUCTION

The most important goal for digital libraries is to ensure a high quality search experience for the user. One central aspect to reach this goal is the assessment of the impact and quality of scientific publications, which is of course far from being trivial. In the scientific field this judgment is mostly performed with respect to the reputation of the publication venue and the number of citations the publication has. The reputation of a researcher is analogously assessed by scanning publication venues and number of citations in the researcher's publication list.

Since the 1960s with the release of the science citation index [1] several metrics were introduced to measure the success of publications, researchers or even whole journals in a deterministic way. Famous examples that are based on this index are e.g. the impact factor [2] and the h-index [3]. However, nowadays, the practicability of such metrics seems more and more questionable due to increasing agility in scientific progresses and communication. When relying on citation based metrics, this agility is impossible to provide since citation counts take several years to become stable [4].

When on the other hand observing the impact of an article within the social Web like bookmarking services, micro blogging platforms or social networks, reactions can be detected immediately after the date of publication [5], [6], [7]. Also this reaction originates from a much more diverse set of users in contrast to the science citation index, where only citations by peers can be regarded. The general meaningfulness and the basic ideas of altmetrics [8], [9], [10], [11] has been confirmed by empirical studies [12], [13] and therefore the usage of these measures is continuously increasing. This is demonstrated by the progression of first Web 2.0 tools, e.g. PlumX[1] and Altmetric[2] and the implementation by several information providers, e.g. PLOS[3] and Nature[4].

The main problem regarding altmetrics is however that it is still not clear what general conclusions can be drawn when an article is frequently mentioned within the social web. It is also not clear how altmetrics are related to classical bibliographical metrics like e.g. citations. It is certain however, that this relationship is not trivial like e.g. more tweets mean more citations. This becomes clear when comparing the tweeting behavior within different communities. For example, the average social science article in our corpus is mentioned 16 times on twitter while the average chemistry article is only mentioned 2.5 times. To judge, whether a certain number of tweets is "high" therefore always depends on the context.

The problem to associate altmetrics to citation counts has been studied a lot recently. This is an interesting problem since citation counts are currently the best naive estimate of scientific quality. If a connection between citation counts and altmetrics existed, this connection could be used to judge the scientific quality of an article much faster as it were possible with citation counts. Of course, before such concepts were introduced in practice several problems have

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
JCDL '15 June 21 - 25, 2015, Knoxville, TN, USA
© 2015 ACM. ISBN 978-1-4503-3594-2/15/06...$15.00
DOI: http://dx.doi.org/10.1145/2756406.2756913

[1] http://www.plumanalytics.com
[2] http://www.altmetric.com
[3] http://article-level-metrics.plos.org/
[4] http://www.nature.com/

to be solved. Possible negative indications are e.g. the appearance of spam in the social web to boost the reputation of bad articles or on the other hand good articles suffering from false predictions. Previous experiments investing the relation between altmetrics and citation metrics are however very diverse in their conclusions, which can be explained by the diverse communities and journals the experiments were performed on.
In this paper we therefore attempt to examine the effect of diverse user groups within the social web. For this purpose we proposed several methods to separate users by their field of interest, by their expertise or by other statistical properties. Our experiments are based on a real world dataset (PLOS) with more than 70,000 articles. In the first part of the experiments, we analyze and introduce the used corpus. In this part we illustrate typical progressions of different article metrics and the correlations between different metrics. In the second part we will take a deeper look into the available twitter data to identify different groups of users and we will analyze if these groups behave differently.

The contributions of our paper are:

- An in-depth correlation analysis between Web 2.0 metrics, view metrics and bibliographic metrics with respect to different scientific domains
- The identification of diverse user groups in the social media and an analysis of their impact on correlations to article metrics
- The introduction of a method for finding and characterizing a user group that maximizes the correlation between tweets and citations counts

The rest of the paper is organized as follows: In chapter 2 we introduce important related work introducing basic fundamentals of altmetrics and recent works searching for correlations between bibliographic metrics and altmetrics. In chapter 3 we describe our corpus and the conducted experiments. Afterwards, in chapter 4 the results obtained in chapter 3 are discussed and conclusions are drawn.

2 RELATED WORK

Recently, data from social media services has been frequently used to provide a new way of measuring the early impact of scholarly publications. This approach gets more and more popular under the term of altmetrics [5], [6], [7]. These metrics try to reflect activity in social media services with the purpose of gathering scholarly impact besides the traditional citation based metrics. By tracking these activities, it is possible to monitor the manner in which scholarly documents are disseminated and discussed in almost real-time [6], [17], [18]. The role of social media in scholarly communication has been investigated in several studies including their use in dissemination [19], conference chatter [20], science popularization [21], and promotion of scholarly products [22]. In addition, several tools have been introduced to facilitate the use of Altmetrics, e.g. PlumX, ImpactStory, Altmetrics and Scholarometer [23]. Finally, the authors of [24] conclude that the potential of altmetrics is nothing short of a complete map of scholarly activity and influence.
Twitter, as one of the most popular social media services which claims more than 200 million active users in March 2013 creating 400 million Tweets each day [32], is a very interesting source for altmetrics, because in addition to the links, contextual information is also available, which can be further used for content analysis. There are already many studies analyzing Twitter, for example in the area of expert finding. The authors of [27] analyzed different indicators for expertise and interest in an enterprise social software suite. Their results show that micro blogs produced good results in terms of recall and precision for expert mining. In [28] it was shown that the use of the Wisdom of the Crowds provided by the list information in twitter is a very effective way for mining expertise in Twitter. The authors showed that their system which relies on list information can compete with the proprietary who-to-follow system provided by Twitter.
Also, studies investigating the prediction of citations based on published tweets are available. However, the results of these papers are sometimes contradictory. In [29] the authors discovered that tweets correlate with citations. The evaluations shown in [29] are however based on very low numbers of articles (ranging from 8 to 55 articles). Also in [30] the authors concluded that tweets and citations are correlated. The authors also analyzed progression patterns of several metrics and found out that download patterns and response patterns in the social web differ significantly. The number of regarded articles in that study were however also very low (70 articles) and the selection of articles to calculate the correlations with was biased, since only the 70 top tweeted articles in their corpus were regarded. While we can support the findings considering progression patterns of different metrics, in our dataset we could not find a direct correlation between tweets and citations.
There are also other papers supporting our findings. In [31] the authors worked with 134,000 articles from the PubMed corpus ranging across different scientific fields. The articles were annotated with altmetrics provided by Altmetric.com. Their results show that the presence and density of altmetric counts are still very low for scientific publications. Only 15%-24% of the publications presenting some altmetrics activity. However, this depends on the domain of the publication: publications from social science, humanities and the social and life sciences showed the highest percentage of alt-metrics. The correlations between tweets and citations (ranging from 0.104 to 0.183) can be considered as very low. Also in [32] only weak or no correlations between tweets and citations were found. The goal of their research was to provide an overview about how often Twitter is used to disseminate information about journal articles in the biomedical sciences. Only 10% of the articles were mentioned in Twitter. The experiments regarding the correlation between these tweets and citations show almost no correlations.
Nevertheless, large-scale studies of altmetrics are still too rare to provide systematic evidence about the reliability, validity, and expressiveness of these measures [15], [16]. One study investigated the spread of scientific information by their Web usage statistics [25]. This study, as well as our study, is based on the statistics obtained from the PLOS Article Level Metrics dataset compiled by PLOS. The authors have shown that the cumulative number of HTML views follows a long tail distribution. In addition, they have shown that the spread of information displays two distinct decay regimes: a rapid downfall in the first month after publication, and a gradual power law decay afterwards.

3 EXPERIMENTS

For our experiments, we used a corpus that we have crawled from the PLOS journal consisting of 74,130 articles that have been published between October 2008 and August 2013. The

distribution of articles over time is shown in Figure 1.
For each document the corpus contains a collection of metadata fields, like the DOI, the title and the publication date. Additionally, the corpus contains log summaries for different sources (like Twitter or Facebook). The log summaries consist of a list of "events", where each event is an entity crawled by the respective crawler of a source (e.g. a tweet, a citation or a comment). Furthermore, the log contains several aggregated values to summarize the list of events (like the *current number* of tweets, or citations). Finally, the progression of this aggregations are logged in a history. The history is updated irregularly every time the respective source has been crawled. The crawler then appends the current timestamp and the current aggregated value for the respective source to the history of that source. Unlike the current aggregated values, the values for the history only consist of a single value at a certain time per source. For example, the Facebook event log contains the current number of shares, comments and likes for a document separately, but the Facebook event history only contains the *sum* of shares, comments and likes at different points in time.

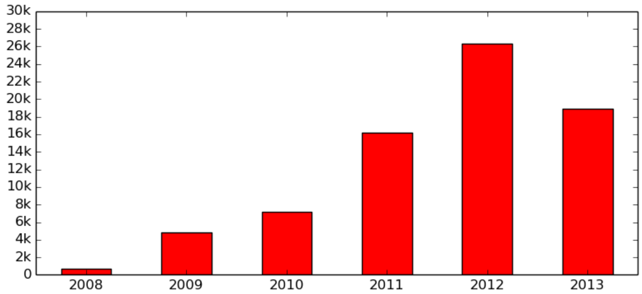


Figure 1: Distribution of articles over time

The sources contained in our corpus cover several Web 2.0 metrics like Twitter, different Facebook measures and CiteULike. From Twitter tweets referencing articles in our corpus with a link are available. For Facebook the corpus distinguishes between likes, comments and shares (a mechanism to share a link on Facebook). CiteULike is a social bookmarking service, allowing users to save and share citations to academic papers. Here, the number of shares for an articles is of interest. The Mendeley data covers the amount of readers for an article and groups working with that article. In addition, the corpus contains various citation metrics. These metrics count the number of references to an article within the respective data source. The sources taken into account for citation metrics are CrossRef, PubMed, Scopus, Nature and Postgenomic. Finally, the corpus contains multiple view metrics from PLOS and PubMed Central (PMC). PLOS is a nonprofit open access project for publishing peer reviewed journals with a strong focus on biology and medicine. Nevertheless, PLOS also contains articles in other domains published in the PLOS ONE journal. Table 3 shows the proportion of disciplines represented in the PLOS corpus for the most frequent disciplines. The disciplines were derived from the Mendeley metadata contained in the PLOS corpus. Since multiple disciplines can be assigned to one article the sum of the percentages is larger than 100%. PMC is an open access branch of PubMed with a strong focus on medicine. From PLOS and PMC both HTML views and PDF views are available.
To provide an initial overview of the data, Table 1 summarizes the most important statistical values for each metric. The since-field corresponds to the first occurrence of a nonzero entry logged in any history for that metric in the whole corpus. The min-value for each metric is not listed, since the min value is zero or close to zero for each metric and therefore not meaningful. The maximum, average and standard deviation fields were determined by using all documents published after the metric was crawled (after the respective since-date).

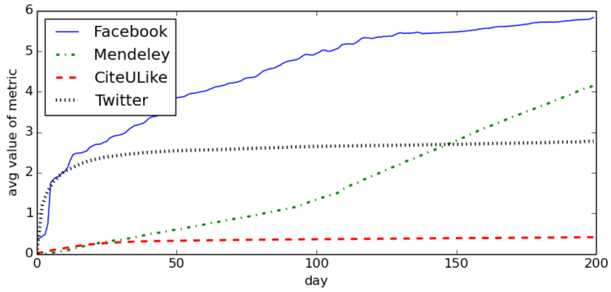
Table 1: Summary of available metrics

Metric	Max	Avg	Std	Since
Tweets	1,179	2.82	16.4	05/12
Facebook shares	2,325	2.56	24.0	12/11
Facebook comments	4,979	1.89	24.2	12/11
Facebook likes	4,068	3.86	30.4	12/11
Mendeley readers	627	6.5	9.90	01/12
Mendeley groups	1	0.19	0.39	01/12
CiteULike shares	138	0.45	1.97	03/09
PLOS HTML views	274k	1,637	3,529	09/09
PLOS PDF views	24k	330	382	09/09
PMC HTML views	59k	239	346	06/11
PMC PDF views	3,689	118	130	06/11
CrossRef citations	842	3.71	8.32	03/09
PubMed citations	392	2.34	5.80	03/09
Scopus citations	1,069	4.85	11.4	03/09
Nature citations	8	0.005	0.09	03/09
Postgenomic citations	8	0.004	0.09	03/09

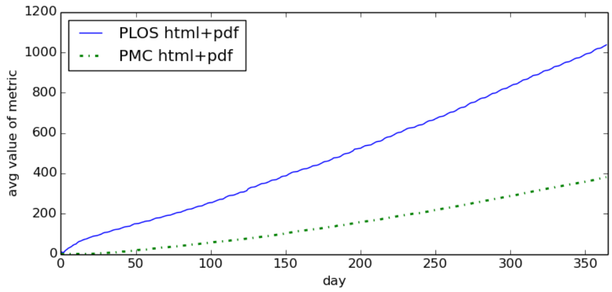
For several metrics the statistical overview indicates that further investigations are not very promising. For example, for the Nature and Postgenomic citations the number of citations is zero for almost all articles. Also the Mendeley groups are not suitable for further investigations, since for this field only a Boolean value (is contained in a group or not) is available. Therefore, these metrics were left out of further experiments.
To provide an impression of the progression of the regarded metrics, the aggregation of histories of each source was plotted in Figure 2. The plots were created by selecting every history from the corpus that belongs to the corresponding metric, starting after the metric was crawled (after the since-date) and lasting at least for the period of time that the plot shows. The value for the n'th day was then determined by fetching from each history the record that was closest to the n'th days after the corresponding article was published. The average of these values is the y-value for day n.
The graphs show typical behavior for the respective metrics. The citation metrics from Scopus, PubMed and CrossRef show a typical long term progression pattern indicating that the number of references to an article continues to grow after 500 days and more. The Web 2.0 metrics from Facebook, CiteULike and Twitter grow very quickly immediately after an article was published and very slowly afterwards. The view metrics from PLOS and PMC grow almost linearly and the progression of Mendeley readers are in between citation progressions and view count progressions.

3.1 Correlations of different metrics

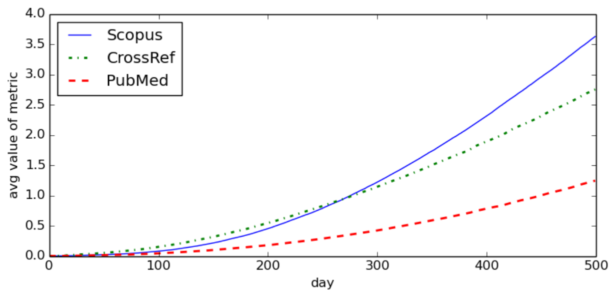
To detect possible interplay between different metrics, we analyzed the correlation between each pair of metrics. Known methods for measuring correlations are the Pearson correlation, Kendall's τ rank correlation coefficient and Spearman's rank correlation coefficient. The Pearson correlation is however not suitable for this task, since the data must be normal distributed. In our corpus this prerequisite does not comply. Kendall's τ or Spearman's rank correlation coeffi-



(a) Facebook, Mendeley, CiteULike and Tweets



(b) PLOS views+pdf and PMC views+pdf



(c) Scopus, CrossRef, PubMed

Figure 2: Progression of different metrics over time

cient are both applicable for this task but since former work mainly used Spearman's method, we also used Spearman's rank correlation to deliver comparable results. We also calculated the p-values for statistical significance testing which was below 0.01 for each correlation. This suggests that the resulting correlations are statistically significant.

For our correlation analysis, we only used articles from a certain time window, since articles from different time windows are hardly comparable. Take for instance the progression of Tweets in Figure 2a and the CrossRef citations in Figure 2c. The figures show that for most articles the number of tweets is stable after the first few weeks, while the number of citations continues to grow even after 500 days. Hence, the proportion of tweets and citations is highly dependent on the age of the article. For our correlation analysis, we chose all documents between June 2012 and August 2012 (6,731 articles). We performed the last update of the article metrics for these articles in July 2014. The article metrics for this analysis were therefore sampled roughly two years after the articles' publication.

Table 2 shows the pairwise correlation between all examined metrics using Spearman's rank correlation coefficient. Most of the strong correlations are rather obvious like Facebook Comments being correlated to Facebook Likes or HTML views being correlated PDF views. It is interesting however that the correlation between PDF and HTML views on PubMed Central is much higher than on PLOS. It is also interesting that correlations between Facebook and the remaining metrics only seem to rely on Facebook shares since correlations between Facebook shares and other metrics are consistently higher than to any other Facebook metric. It is also noticeable that strong correlations between citation metrics and other metrics are very rare. There are some connections to the view metrics and also to Mendeley but these correlations are rather weak.

Additionally to the analysis on the whole corpus we also analyzed different subsets of the corpus by regarding only document from certain scientific disciplines. Since another set of matrixes would however be far too verbose, we visualized the correlations of these metrics as graphs shown in Figure 3. To build these graphs, we had to choose a threshold defining whether two nodes should be connected by an edge or not. We set this threshold to a correlation of 0.5 which can be considered as a decent correlation. The weight and the color of the edges indicate the strength of the correlation. The size of the nodes indicate the number of non-zero entries for that metric.

The graphs for computer science and medicine show in general much more correlations passing the threshold of 0.5. Especially the view metrics and citation metrics are much more connected to each other. The computer science graph is also the only graph containing correlations to CiteULike, while the social sciences graph is the only graph that has a connection to Twitter. It is also noticeable that the social science graph has no connection to any PubMed Central metric. One commonality of all graphs is however that no correlation between Web 2.0 metrics and citation metrics exist but only between citation metrics and view metrics and between view metrics and Web 2.0 metrics.

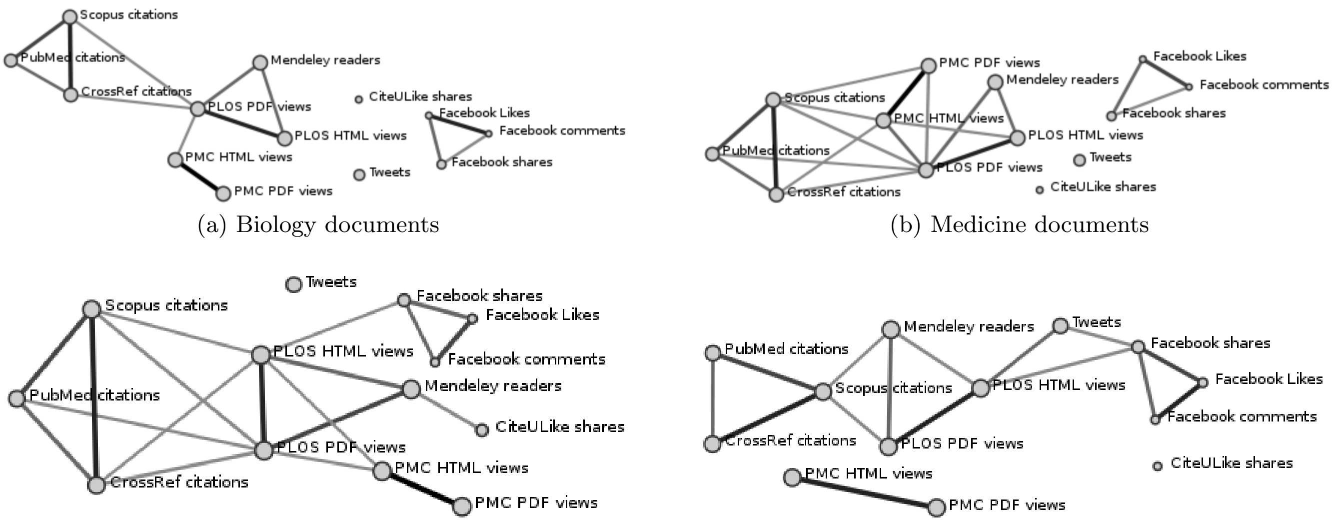
This high variance in the correlation of different metrics when observing different communities can be an explanation to the high discordance in the experiments mentioned in the related work. Table 3 further shows a high discordance in the average number of tweets for articles in different domains. Having in mind that the main focus of PLOS is biology and medicine, it is also likely that this effect is even stronger in corpora of domain specific journals in other domains than biology. Another factor that characterises PLOS the open access to its content. Therefore not only scientists have access to the documents in our corpus, but everyone. This influence will also be analyzed.

In the following chapter we therefore plan to analyze the influence of diverse user groups in the social web. For our experiments we mainly used Twitter as one of the most popular social media services, claiming more than 200 million active users in March 2013 creating 400 million Tweets each day [33]. Additionally, Twitter provides very useful metadata for each tweet like the author of the tweet, the actual tweet text, the exact date etc. Besides that, the Twitter API provides detailed profile information for every user.

In the experiments we will first verify if scientist contribute more to the correlation between Web 2.0 metrics and citation metrics. Then we partition the twitter users into diverse groups to verify if different groups of users show different behavior with respect to correlations between the article metrics. In the end we will present an approach to identify

Table 2: Pairwise correlation between each regarded metric

	Tweets	Facebook shares	Facebook comments	Facebook likes	Mendeley readers	CiteULike shares	PLOS HTML views	PLOS PDF views	PMC HTML views	PMC PDF views	CrossRef citations	PubMed citations	Scopus citations
Tweets	1.000	**0.406**	0.254	0.280	0.313	0.221	**0.425**	0.286	0.060	0.021	0.140	0.110	0.118
Facebook shares	**0.406**	1.000	**0.592**	**0.671**	0.263	0.159	**0.412**	0.294	0.053	0.001	0.125	0.095	0.110
Facebook comments	0.254	**0.592**	1.000	**0.800**	0.193	0.128	0.279	0.216	0.030	-0.005	0.092	0.058	0.097
Facebook likes	0.280	**0.671**	**0.800**	1.000	0.208	0.143	0.301	0.228	0.029	-0.017	0.087	0.072	0.096
Mendeley readers	0.313	0.263	0.193	0.208	1.000	0.309	**0.527**	**0.564**	0.106	0.082	0.276	0.229	0.270
CiteULike shares	0.221	0.159	0.128	0.143	0.309	1.000	0.276	0.249	0.015	-0.017	0.112	0.124	0.108
PLOS HTML views	**0.425**	**0.412**	0.279	0.301	**0.527**	0.276	1.000	**0.797**	**0.406**	0.338	0.383	0.312	0.392
PLOS PDF views	0.286	0.294	0.216	0.228	**0.564**	0.249	**0.797**	1.000	**0.520**	**0.492**	**0.436**	0.341	**0.452**
PMC HTML views	0.060	0.053	0.030	0.029	0.106	0.015	**0.406**	**0.520**	1.000	**0.926**	0.348	0.272	0.347
PMC PDF views	0.021	0.001	-0.005	-0.017	0.082	-0.017	0.338	**0.492**	**0.926**	1.000	0.338	0.263	0.338
CrossRef citations	0.140	0.125	0.092	0.087	0.276	0.112	0.383	**0.436**	0.348	0.338	1.000	0.388	**0.710**
PubMed citations	0.110	0.095	0.058	0.072	0.229	0.124	0.312	0.341	0.272	0.263	0.388	1.000	**0.621**
Scopus citations	0.118	0.110	0.097	0.096	0.270	0.108	0.392	**0.452**	0.347	0.338	**0.710**	**0.621**	1.000



(a) Biology documents

(b) Medicine documents

(c) Computer Science documents

(d) Social Science documents

Figure 3: Correlations between metrics as graph

Table 3: Proportion of documents in scientific domains with the average number of tweets referencing papers the domain

Discipline	**% Docs**	**∅ Tweets**
Biological Science	73.48%	3.19
Medicine	32.53%	4.07
Environm. Science	7.82%	7.32
Psychology	7.41%	10.82
Computer Science	5.87%	7.51
Chemistry	5.08%	2.64
Engineering	4.91%	2.92
Physics	3.68%	3.35
Social Science	3.44%	15.98
Earth Science	2.71%	8.28

and characterize users that correlate most to the regarded citation metrics.

3.2 Identifying diverse user groups in Twitter

In our corpus twitter related data is available since the middle of March 2012. There is a total amount of 115,892 tweets referencing articles in the PLOS corpus where 10,811 of these tweets reference articles published before the twitter data was available. These tweets were disregarded for the following experiments since otherwise incomplete tweet progressions would interfere with complete tweet progressions. Overall, 53% of all articles was referenced by at least one tweet that were posted by 10,858 distinct users.

In the following experiments we will introduce several methods to identify special user groups. For this user groups we analyzed if the tweets posted by that user group correlate more or less to the remaining article metrics. This is done by ignoring all tweets in the corpus that were not posted by a user in the regarded user group. Afterwards the correlations to all other metrics are recalculated and compared to a baseline.

Used baseline for the following experiments. It is worth mentioning that in order to receive a perfect rank correlation of 1.0 e.g. between tweets and citations there is a minimum amount of tweets needed, since all best cited papers also must have the most tweets, all second best cited papers must have at least one tweet less, and so on. For our corpus, this minimum amount of tweets would be 29,716. Thus, when regarding all tweets in our corpus, it would in theory be possible to reach a rank correlation of 1.0. But if we for instance only regard tweets posted by a very specialized group of domain experts who only composed 5% of all tweets, a rank correlation of 1.0 would even in theory not be achievable anymore. In fact, removing random tweets reduces the correlation between tweets and the remaining metrics significantly. Therefore, when we want to assess the correlations resulting from a certain user group we need a baseline that aware of this condition. In order to build this baseline we picked a random sample of tweets with the same size as the number of tweets that the analyzed user group has tweeted. Additionally, to ensure that this random sample has no undesired properties, we repeated this random sampling 100 times and averaged the correlations gained from the individual runs. This also ensures reproducible results to a satisfying degree.

3.2.1 Detect scientists by exploring user descriptions

In our first experiment we plan to find out if tweets posted by scientists correlate more to citation metrics than tweets posted by other users. To identify scientists we analyzed the twitter descriptions given by the users in their twitter profile. This description is a short text that twitter users can provide to describe themselves and their interests in a short text. In this experiment we assume that scientists mention some characteristic words in their profile descriptions. These words include e.g. "university", "Ph.D." (in different ways of spelling), "professor", "doctor" or "institute". By searching for user descriptions containing at least one of the words in our word list, we found 1232 users.

To ensure that the returned user descriptions really belong to scientists, we manually examined a random sample of 100 user descriptions. In our random sample we had a precision rate of 88%. False positives included for instance twitter accounts directly associated to certain institutes or universities as well as non-researchers hired in a university (e.g. technical staff). A recall could not be determined as for this we must know the true identity of each twitter user (even of those who didn't provide a description). In average the returned users posted 2 tweets referencing articles in the PLOS corpus. Table 4 shows the changes in correlation when only tweets by the identified scientists were used.

Table 4: Differences in correlations when only regarding scientists compared to the baseline described in section 3.2.1

Metric	Δ_{baseline}	Metric	Δ_{baseline}
Crossref	+0.020	PubMed	+0.025
Scopus	+0.020	PLOS pdf	+0.049
PLOS HTML	+0.043	PMC pdf	-0.045
PMC HTML	-0.042	Facebook Shares	+0.027
Facebook Comments	+0.028	Facebook Likes	+0.033
Mendeley Readers	+0.066	CiteULike	+0.073

The results show that the tweets posted by the identified scientists have a slightly higher correlation to citation metrics than average users. Bigger impacts can be found when observing the metrics for scientific document management i.e. Mendeley and CiteULike. Here the increase lies at 7 percent points. It is also noticeable that the identified scientists seam to prefer PLOS over PMC since the increase in correlation to the PLOS view metrics is almost identical to the loss in correlation to the PMC metrics.

3.2.2 Detect experts by annotation of user tweets

One problem with the previous experiment was that not only scientists in the field of biology (which is the main focus of the corpus) were regarded, but scientists in arbitrary domains. The expertise of a user in a certain domain would however be of great interest, since a domain experts recommendation to an article is certainly more valuable. Since another word filtering approach for different domains would however be very prone to errors and would further reduce the number of analyzable users, we instead analyzed the users tweets as an aditional data source for each user. The basic idea is, that if users tweet very often in a certain domain, they certainly have some expertise in that domain. In the following we will use this assumption to identify experts in diverse domains and analyze how the correlations change if only one group of experts is regarded.

Annotating user tweets with Wikipedia categories. The domains that we assign to the twitter users should be as diverse as possible and universally applicable. A good candidate for this purpose are the 24 top categories provided by Wikipedia's category graph. These are Medicine, Sports, Culture, Technology, Education, Health, Business, Belief, Humanities, Society, Life, Arts, Language, Law, History, Geography, Agriculture, Politics, Mathematics, Science, Nature, Environment, People and Chronology. To assign one of this top categories to a tweet, we used the Wikipedia Miner framework [34]. The Wikipedia Miner provides for each tweet a set of concepts (articles in the Wikipedia) that are related to the tweet. By using the Wikipedia category graph, the category of this concept could then be generalized to identify the respective top category. An expert of a top category was then defined as a user who posts more tweets than in that category than an average user. A more detailed description of the method is given in [35]. In this experiment we only used the 200 most recent tweets posted by a user. This both increases the feasibility of the approach due to restrictions from the Twitter API and this also functioned as a normalization since users with a low tweet frequency had an equal chance of being identified as expert as users with high tweet frequency.

Since biology is the main focus in our corpus, we had to ensure that this category is well represented by the used categories. Unfortunately, in the Wikipedia category hierarchy biology firstly occurs in the third level (Nature $\rightarrow$ Natural Science $\rightarrow$ Biology) and is therefore not very well represented. Thus, we introduced an exception and added the biology category to the list of regarded categories. If the category of a tweet can be generalized to biology, the tweet is then annotated both with Biology and Nature. Table 5 shows the number of users considered as experts in the respective discipline.

Table 5: Number of experts w.r.t. Wikipedia's top categories

Category	#users	Category	#users	Category	#users
Biology	1466	Medicine	2986	Sports	2506
Culture	5094	Technology	3861	Education	4349
Health	3158	Business	4367	Belief	4929
Humanities	4391	Society	4775	Life	4236
Arts	4292	Language	4340	Law	2967
History	4131	Geography	3691	Agriculture	3020
Politics	3838	Mathematics	4354	Science	3918
Nature	4463	Environment	2785	People	3598
Chronology	3920				

It is striking that the number of users is significantly higher than the number of identified scientists in our previous experiment. Fortunately, the distribution of experts is also very uniform with the exception of Biology experts. This can however be explained by the fact that Biology is a much more specific category in the Wikipedia category graph. In the following we will analyze the differences in correlations when only tweets posted by those experts are regarded.

Analyze correlations for diverse user groups. In our experiments we first analyzed every category independently and considered only the tweets posted by the experts of a certain category and analyzed the differences in correlations compared to the baseline (as explained in section 3.2). As a whole list of all differences in correlation would be far to verbose, we instead listed for each metric the 3 categories that yielded the most positive impact to that metric and the 3 categories that yielded the most negative impact (see Table 6).

The experiment shows that different user groups can have a significant influence on the metrics in a positive as well as in negative way. For instance when selecting Health and/or Medicine experts, the view correlations to PMC increased significantly while the view correlation to PLOS decreased significantly. Also, most negative correlations considering Facebook occurred when Medicine and/or Health experts were regarded. Some improvements to the Facebook correlations could be observed when People, Arts or Culture experts were regarded.

Our next experiment is inspired by two findings that are observable in the previous experiment. The first observation is that the negative influences in Table 6 is significantly higher than the positive influences. For instance the correlation to Facebook metrics decreases by 10 percent points when only Health experts were regarded. This raises the question if *excluding* certain domain experts from the set of all users would also increase the correlation to certain metrics.

The other finding refers to Table 5, showing the number of experts for each category. Having in mind that the total amount of users in the considered time window is 10,858 and each of the 25 categories contain about 3,000-4,000 users, it is easy to see that the overlap of these expert groups must be very high. To decrease this overlap, it is necessary to increase the threshold that defines a user as expert. E.g. users could be identified as experts if they tweeted more than the average user plus a multiple of the standard deviation in that category. A method to normalize values with respect to their average value and standard deviation is the z-score, where z is the multiple of a standard deviation in the equation $x = \mu + z\sigma$. To give an impression of the selectivity of different z-scores, Table 8 shows the number of users for some categories for different z-scores.

Table 7 shows the results of this improved method. For this results we determined expert groups for each category for different z-scores between -1 and 3 and either included or excluded the tweets of these users from the set of regarded tweets. The results are shown in a similar manner as in the previous experiment: For each metric we show the three sets of users that resulted in the biggest increase in correlation and the three sets of users that resulted in the biggest decrease in correlation. The number in index shows the z-score used to define the experts and the leading sign shows, whether the respective expert group was excluded or included from the set of regarded users ("–": excluded, "+": included). E.g. "+Life$_{3.0}$: -0.22" in the row "Facebook Likes" means that when only regarding users with a z-score of 3.0 or higher in the Life category, the correlation to Facebook Likes decreases by 22 percent points. And "–Sports$_{-1.0}$: 0.03" in the CrossRef row means that when *disregarding* users with a z-score of -1.0 or higher in Sports, the correlation to CrossRef citations increases by 3 percent points.

The results show that the effect of the previous experiment were significantly increased by changing the selectivity of the regarded expert groups and by allowing expert groups to be excluded from the set of regarded users. E.g. if only the expert group Life$_{3.0}$ is regarded, the correlation to all Facebook metrics fall by more than 20 percent points. Also the increase in citation metrics is slightly higher. The best increase for PubMed citations could for instance be achieved by selecting a very focused group in Biology (Bio$_{2.0}$). And the biggest increase for CrossRef and Scopus citations results when disregarding every user that occasionally talk about topics that are rather unrelated to science like Sports, Law, Arts or Nature. Also the category Mathematics seem to have a negative impact on Scopus citations, which can be seen in both directions: When excluding the user group Math$_{0.5}$, the correlation to Scopus citations increases and when *only* regarding the user group Math$_{2.0}$, the correlation to Scopus citations decreases. This tendency accords with the distribution of indexed articles in Scopus which has a clear focus on life science, social science and medicine.

Table 8: Number of experts w.r.t. some exemplarily categories for different z-scores

Category	**z=-1**	**z=0**	**z=1**	**z=2**	**z=3**
Biology	2967	1466	728	363	189
Medicine	9054	2986	1623	1030	653
Arts	9267	4292	599	120	44
History	9362	4131	608	136	55
Mathematics	9354	4354	987	288	111
Sports	8801	2506	839	421	235
Life	9387	4236	1487	523	175

3.2.3 Identify users that maximize the correlation to citations

Even though the previous experiments provided some interesting insights on different kinds of user groups, it is noticeable that no experiment had a significant positive impact on the correlation to citation metrics. Therefore, in this experiment we will perform a bottom up approach and first find a user group that positively influences the correlation to citation metrics and will then analyze this set of users to find communalities.

Sort users by impact to citation metrics. In this experiment we will order the users with an insertion sort like algorithm with respect to their impact to citation metrics. An insertion sort algorithm is one of the simplest ordering algorithms where in each step the "smallest" element is appended to a new list of elements and is removed from the original list. This procedure is repeated until the original list of elements is empty.

In our case, the definition of "small" for users is their negative impact on citation metrics. To determine the negative impact of a user in each iteration step, the correlations between tweets and citations are recalculated, ignoring tweets posted by that user. The "smallest" user is then defined as the user where the removal of tweets resulted to the biggest positive impact on the correlation between tweets and citations.

Aggregate citation metrics. One last problem that needs to be addressed is that we do not have a metric, to determine a correlation to *the citation metrics*, but only to every citation metric *independently*. We therefore aggregated the

Table 6: Differences in correlations when only regarding experts of certain categories

Metric	**Best cat.**		**2^{nd} best cat.**		**3^{rd} best cat.**			**3^{rd} worst cat.**		**2^{nd} worst cat.**		**Worst cat.**	
Crossref	Biology	0.01,	Agriculture	0.00,	Sports	-0.00,	...,	Language	-0.03,	Law	-0.03,	Humanities	-0.04
PubMed	Health	0.03,	Biology	0.02,	Medicine	0.02,	...,	Humanities	-0.03,	Language	-0.03,	Arts	-0.03
Scopus	Biology	0.01,	Agriculture	0.00,	Environm.	0.00,	...,	Culture	-0.03,	Language	-0.03,	Humanities	-0.04
PLOS pdf	People	0.02,	Agriculture	0.00,	Chronology	-0.00,	...,	Society	-0.05,	Health	-0.05,	Medicine	-0.05
PLOS HTML	People	0.02,	Chronology	0.02,	Environm.	0.01,	...,	Science	-0.03,	Health	-0.05,	Medicine	-0.05
PMC pdf	Health	0.11,	Medicine	0.10,	Biology	0.05,	...,	Humanities	-0.09,	Geography	-0.09,	Language	-0.09
PMC HTML	Health	0.09,	Medicine	0.08,	Biology	0.03,	...,	Humanities	-0.08,	Language	-0.09,	Geography	-0.09
Facebook Shares	People	0.05,	Culture	0.03,	Sports	0.02,	...,	Biology	-0.08,	Medicine	-0.10,	Health	-0.11
Facebook Comments	People	0.04,	Arts	0.03,	Culture	0.02,	...,	Biology	-0.06,	Medicine	-0.08,	Health	-0.09
Facebook Likes	People	0.05,	Arts	0.03,	History	0.02,	...,	Biology	-0.07,	Medicine	-0.09,	Health	-0.10
Mendeley Readers	Humanities	0.04,	Language	0.03,	Arts	0.03,	...,	Business	-0.04,	Medicine	-0.08,	Health	-0.08
CiteULike	Language	0.05,	Humanities	0.04,	Arts	0.03,	...,	Agriculture	-0.04,	Health	-0.05,	Medicine	-0.06

Table 7: Maximal differences in correlations when including or excluding experts of certain categories. The leading sign indicates whether the respective expert group was excluded or included ("–": excluded, "+": included). The number in the index shows the z-score used to define the selectivity of the expert group (See section 3.2.2 for explanation)

Metric	**Best cat.**		**2^{nd} best cat.**		**3^{rd} best cat.**			**3^{rd} worst cat.**		**2^{nd} worst cat.**		**Worst cat.**	
Crossref	$-$Sports$_{-1.0}$:	0.03,	$-$Law$_{-1.0}$:	0.03,	$-$Human$_{-1.0}$:	0.03,	...,	$+$Math$_{2.0}$:	-0.05,	$+$Society$_{1.5}$:	-0.05,	$+$Politics$_{1.0}$:	-0.06
PubMed	$+$Bio$_{2.0}$:	0.04,	$+$Bio$_{1.5}$:	0.03,	$+$Life$_{1.5}$:	0.03,	...,	$+$Human$_{1.0}$:	-0.04,	$+$Nature$_{2.0}$:	-0.04,	$+$Nature$_{1.5}$:	-0.05
Scopus	$-$Nature$_{-0.5}$:	0.03,	$-$Arts$_{-0.5}$:	0.03,	$-$Math$_{0.5}$:	0.03,	...,	$+$Math$_{2.0}$:	-0.04,	$+$Politics$_{1.0}$:	-0.05,	$+$Belief$_{0.5}$:	-0.05
PLOS pdf	$-$Society$_{0.0}$:	0.03,	$-$Society$_{0.5}$:	0.03,	$-$Health$_{3.0}$:	0.03,	...,	$+$Health$_{2.5}$:	-0.12,	$+$Health$_{3.0}$:	-0.14,	$+$Life$_{3.0}$:	-0.14
PLOS HTML	$-$Med$_{2.5}$:	0.03,	$+$Arts$_{1.0}$:	0.03,	$-$Health$_{2.5}$:	0.03,	...,	$+$Health$_{2.5}$:	-0.15,	$+$Nature$_{2.0}$:	-0.16,	$+$Life$_{3.0}$:	-0.19
PMC pdf	$+$Health$_{2.0}$:	0.15,	$+$Health$_{2.5}$:	0.15,	$+$Med$_{3.0}$:	0.15,	...,	$+$Geo$_{1.0}$:	-0.11,	$+$Geo$_{1.5}$:	-0.12,	$+$Env$_{1.0}$:	-0.12
PMC HTML	$+$Health$_{1.5}$:	0.12,	$+$Health$_{2.0}$:	0.12,	$+$Health$_{2.5}$:	0.11,	...,	$+$Geo$_{1.0}$:	-0.11,	$+$Geo$_{1.5}$:	-0.12,	$+$Env$_{1.0}$:	-0.13
Facebook Shares	$-$Med$_{0.5}$:	0.06,	$-$Med$_{0.0}$:	0.06,	$-$Health$_{0.5}$:	0.06,	...,	$+$Bio$_{2.5}$:	-0.20,	$+$Bio$_{3.0}$:	-0.22,	$+$Life$_{3.0}$:	-0.23
Facebook Comments	$-$Health$_{-0.5}$:	0.06,	$-$Med$_{0.0}$:	0.05,	$-$Health$_{0.0}$:	0.05,	...,	$+$Life$_{2.5}$:	-0.16,	$+$Bio$_{3.0}$:	-0.17,	$+$Life$_{3.0}$:	-0.20
Facebook Likes	$-$Health$_{0.0}$:	0.06,	$-$Med$_{0.0}$:	0.06,	$-$Health$_{0.5}$:	0.06,	...,	$+$Bio$_{3.0}$:	-0.18,	$+$Life$_{2.5}$:	-0.19,	$+$Life$_{3.0}$:	-0.22
Mendeley Readers	$+$Env$_{2.0}$:	0.05,	$-$Health$_{0.5}$:	0.05,	$-$Health$_{0.0}$:	0.05,	...,	$+$Med$_{3.0}$:	-0.15,	$+$Health$_{2.5}$:	-0.17,	$+$Life$_{3.0}$:	-0.18
CiteULike	$+$Belief$_{1.0}$:	0.06,	$+$Math$_{1.5}$:	0.05,	$+$Lang$_{0.0}$:	0.05,	...,	$+$Med$_{3.0}$:	-0.13,	$+$Health$_{2.5}$:	-0.14,	$+$Life$_{3.0}$:	-0.16

citation metrics known for an article as described in the following section.

The citation metrics given by the available sources Scopus, PubMed and CrossRef correspond to the amount of times an article was cited *within the respective corpus*. Thus, the provided number of citations is not necessary the "real" number of citations to an article, since the article can also be cited by articles not contained in the respective corpus. Therefore, the number of citations for an article given by a source can only be an estimate of the real number of citations to an article which is always smaller or equal to the real number of citations. To get the best estimate of citations to an article, the best approach would be to compose the union of the citations found by each source, but unfortunately the exact articles citing the article in focus are only provided by CrossRef, while Scopus and PubMed only provide the *number* of citations. Therefore, we can either assume that the citations found in each corpus are mostly distinct or mostly the same. In the first case, a good measure to the aggregate citation based metrics would be the *sum* of the respective metrics and in the latter case the *maximum value* would be more suitable. As we can assume that the overlap of the regarded sources is quite high, we will in the following use the maximum value of Scopus citations, CrossRef citations and PubMed citations as an estimate for the real number of citations for an article.

Characterize users with high impact to citation metrics. For the suggested sorting of users by their impact to citation metrics we only regarded users who posted at least 5 tweets, as for users who posted less than 5 tweets no significant changes in correlations is expected when ignoring their tweets. The pure removal of all users with less than 5 tweets did not influence the correlation between tweets and citations significantly (-0.009).

After the removal of all users who posted less than 5 tweets, 534 users remained to perform the proposed sorting method. Figure 4 shows the differences in the correlation to citations when only the first n users with respect to our sorting are regarded. When only the tweets posted by the first 180 users are regarded, the correlation increases by 18.1 percent points to 27.4%, which can be considered as a weak correlation. The users contained in that set of 180 users will in the following be referenced as the *positive users* (the remaining users as negative users respectively). In our last experiment we will analyze the positive users to find characteristics featuring these users.

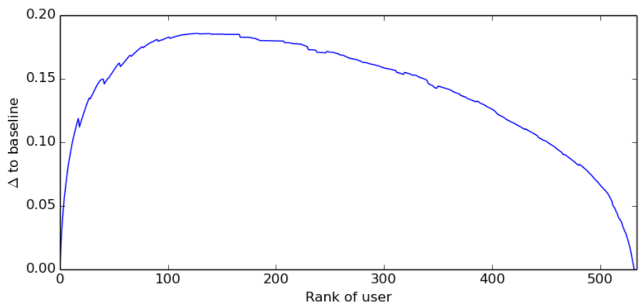


Figure 4: Correlation of first n users to citations metrics w.r.t. the introduced sorting method

In our experiments we identified features that occur more often for positive users than for negative users. The first

feature examined were the descriptions given by the users. Each description was translated into a set of words whereupon frequent words for positive and negative users were determined. To determine frequent words we identified the odds ratio between every word and the positive users, so that the unequal amount of positive and negative user descriptions is of no importance.
The second feature regarded in this experiment are the Twitter lists the users are in. Again, we find for each user the lists they are listed in and determine for each list the odds ratio to positive users.
Finally, as the last feature we analyzed the tweets of a user to find concepts annotated by the Wikipedia miner. The process is analogous to the previous features: The tweets of each user were annotated with Wikipedia concepts (Article titles) to get a set of concepts for each user. Afterwards, the odds ratio between every concept and positive users was calculated.
To ensure that only representative features were selected, we additionally introduced an occurrence threshold that has to be passed before a feature is considered as representative. This step is necessary since the number of distinct features is very high and many inexpressive features with low occurrence can have a high odds ratio just by chance. Table 9 shows for each type of feature a list of the most positive and most negative features.
The list of the most positive features contained in the table are very significant. The positive concepts include the biomedical research foundation Wellcome Trust and several biology related articles. The negative concepts on the other hand seem to be more or less random and unrelated. The most positive lists are also very relevant to biology and science. The positive words are a little bit harder to interpret. There are some biology/medicine relevant words like NCDs (Non-communicable deseases), microbiology or globalhealth. Also some words occur that indicate that our first description filtering approach heads into the right direction. This words include e.g. study, school, education and MD (Doctor of Medicine).
However, this features can currently only to be used to estimate if the returned group of users is reasonable (which seems to be the case). If it is sensible to use this features to actively search for users with high correlation with citations has to be evaluated in future work.

4 CONCLUSION AND FUTURE WORK

In our experiments we identified several diverse user groups and showed the difference in correlation that occur when only one group of users is regarded. We identified these diverse groups by their field of interest, by their expertise and by their influence on the resulting correlations. Our experiments showed that the identified user groups show significant differences in the character of article level metrics. Even though the main focus of the PLOS corpus is biology and medicine, the differences in correlations and average metric values in different domains were significant. The proposed methods are general enough to be applied to arbitrary corpora and provide the chance to view article-level-metrics on a corpus on various different ways.
The high variance in correlations when regarding several diverse user groups also indicates that universal statements regarding altmetrics (e.g. under what conditions tweets might correlate to citations) cannot be made by analyzing just one corpus. Before such universal statements can be proposed, it is necessary to have a large variety of studies conducted on a diverse set corpora. Based on this studies it would be possible to conduct a large meta-study where the experience and knowledge of previous work is aggregated. Unfortunately, the diversity and amount of such studies is currently still insufficient. The correlation analysis conducted on the PLOS corpus in the first part of our experiment section can therefore be seen as a contribution to attain this goal in the future.
The correlation analysis on the PLOS corpus showed that there are three separated classes of metrics that have strong correlations within the same class, but low correlations between different classes. On the one hand there are classic bibliographic measures, like citations, and on the other hand there are Web 2.0 metrics, like tweets or Facebook likes. These two groups have no correlations with each other. However, they are connected by the view metrics (HTML and PDF page views). Although we were able to show that these correlations show significant discordance among different scientific domains, this property still holds for all regarded domains.
In a second series of experiments we identified several diverse groups of users on Twitter and compared their impact on the correlations to the article metrics. First we identified a set of scientists with a simple keyword search approach on profile descriptions. Even though the rate of scientists found with this approach was quite high, we could not confirm that scientists contribute significantly more to citations than other users. This either means that the assumption that scientists contribute more to citation metrics is wrong or that the profession of the scientists was not taken into account appropriately, since also scientists with other research professions than biology or medicine were returned in this experiment.
In subsequent experiments we partitioned the users by their fields of interest that were designed to be as diverse as possible. We did this by identifying users as experts with respect to the Wikipedia's top-level categories. By including and excluding user groups defined by that approach, we could show that significant changes to several metrics could be achieved. For content provider this approach can be interesting, since it can be used as tool to identify particularities and preferences of their community. If a large change in correlation can be observed when a certain group of users is regarded, this indicates that this group is a predominant group within the regarded corpus. E.g. using this approach it was very clear that medicine and health experts are an important user group to PubMed Central, which is certainly true.
In our last experiment we used a bottom up approach to first find a set of users that enhances the correlation between tweets and citations significantly and afterward analyzed and characterized that set of users. In our experiment we used this method to identify a group of users that raises the correlation between tweets and citation metrics from about 10% to about 30%. Afterwards we characterized that set of users by analyzing their profile descriptions on twitter, the Twitter lists the users are listed on and the concepts the Wikipedia miner annotated to their tweets. The features found for the group of users were related to biology, medicine or science in general.
Currently, these identified features are only to be used to estimate if the returned group of users is reasonable. In future work we plan to find out if these features can also be used to actively search users that are more appropriate for an early estimate of scientific quality for articles. Of course this approach can also be used to find and characterize users

Table 9: List of positive and negative features characterizing users with high correlation to citation metrics

Words	pos: ncds, people, microbiology, general, conservation, study, school, practice, evidence, globalhealth, md, education, tropical, ... neg: global, working, neuroscientist, technology, leading, cell, work, founder, into, media, daily, psychologist, dad, math, ...
Lists	pos: pharma, computational biology, genetics, genomics, bioinfo, genome, medicina, top scientists, twitter science, ... neg: blinded me with science, cognitive, my feed, neuro/psych, social media, colleagues, journals, mental health, technology, ...
Concepts	pos: Commentary (magazine), Wellcome Trust, Whole genome sequencing, Long non-coding RNA, Biome, Omics, Humanos, ... neg: Hypertext Transfer Protocol, Spacetime, Tears, Point and click, Wall Street, Synchronization, Matrix (mathematics), ...

that have a high correlation to other metrics like e.g. view counts. This could be used to measure an effective hubness of a user i.e. find users whose tweets really result in more views. For future work it will be interesting to find out if this effective hubness differs conceptually from the number of followers.

The proposed methods can also be considered as fundamentals to distinguish more important and less important tweets. This will be a very relevant problem in the moment when altmetrics are used as a means to judge scientific quality in a large scale. In this case, it is a matter of time until services that publish tweets for money (REF) also specialize on boosting the reputation of scientists for money. In this scenario the proposed methods can be good fundamentals to detect and ignore spam tweets, as they will probably not correlate to any metric that is interesting for an information provider.

5 References

[1] A. Pritchard, Statistical Bibliography or Bibliometrics, Journal of Documentation, vol. 25, pp. 348–349, 1969

[2] E. Garfield, The Thomson Reuters Impact Institute, Technical Report, Factor for Scientific Information, 1994

[3] J. E. Hirsch, An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. U. S. A., vol. 102, pp. 16569–16572, 2005

[4] T. Brody and S. Harnad, Earlier Web Usage Statistics as Predictors of Later Citation Impact, Journal of the American Society for Information Science and Technology, vol. 57, no. 8, pp. 1060–1072, 2006

[5] J. Priem, K. Costello and T. Dzuba, Prevalence and use of Twitter among scholars, figshare, DOI: http://dx.doi.org/10.6084/m9.figshare.104629, 2012

[6] J. Priem and K. L. Costello, How and why scholars cite on Twitter, Proc. Am. Soc. Inf. Sci. Technol., vol. 47, no. 1, pp. 1–4, 2010

[7] J. Bar-Ilan, S. Haustein, I. Peters, J. Priem, H. Shema, and J. Terliesner, Beyond citations: Scholars' visibility on the social web, in Proceedings of 17th International Conference on Science and Technology Indicators, vol. 52900, pp. 98–109, 2012

[8] J. Priem and B. M. Hemminger, Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web, First Monday, vol. 15, no. 7, 2010

[9] C. Neylon and S. Wu, Article-level metrics and the evolution of scientific impact, PLoS Biol'09, vol. 7, no. 11, 2009

[10] D. Taraborelli, Soft peer review: Social Software and Distributed Scientific Evaluation, in International Conference on the Design of Cooperative Systems, 2008

[11] M. Fenner, What can article-level metrics do for you?, PLoS Biol., vol. 11, no. 10, p. e1001687, 2013

[12] P. Groth and T. Gurney, Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study, WebSci'10, 2010

[13] M. Thelwall, S. Haustein, V. Larivíere, and C. R. Sugimoto, Do altmetrics work? Twitter and ten other social web services, PLoS One, vol. 8, no. 5, 2013

[14] R. C. Roemer and R. Borchardt, Institutional Altmetrics and Academic Libraries, Inf. Stand. Q., vol. 25, no. 2, pp. 14–19, 2013

[15] J. Liu and E. Adie, Five challenges in altmetrics: A toolmaker's perspective, Bull. Am. Soc. Inf. Sci. Technol., vol. 39, pp. 31–34, 2013

[16] P. Wouters and R. Costas, Users' narcissism and control – tracking the impact of scholarly publications in the 21 st century, Image Rochester NY, p. 50 pages, 2012

[17] H. A. Piwowar, Altmetrics: Value all research products, Nature, vol. 493, p. 159, 2013

[18] X. Li, M. Thelwall, and D. Giustini, Validating online reference managers for scholarly impact measurement, Scientometrics, vol. 91. pp. 461–471, 2012

[19] E. S. Darling, D. Shiffman, I. M. Côté, and J. a. Drew, The role of Twitter in the life cycle of a scientific publication, PeerJ Prepr., vol. 1, p. 16, 2013

[20] K. Weller, E. Dröge, and C. Puschmann, Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences, Proc. ESWC'11, pp. 1–12, 2011

[21] C. R. Sugimoto and M. Thelwall, Scholars on Soap Boxes: Science Communication and Dissemination via TED Videos, J. Am. Soc. Inf. Sci. Technol., 2012

[22] B. Cronin, Metrics á la mode, J. Am. Soc. Inf. Sci. Technol., vol. 64, no. 6, pp. 1091–1091, 2013

[23] J. Kaur, D. T. Hoang, X. Sun, L. Possamai, M. JafariAsbagh, S. Patil, and F. Menczer, Scholarometer: A Social Framework for Analyzing Impact across Disciplines, PLoS One, vol. 7, 2012

[24] M. Taylor, Exploring the Boundaries: How Altmetrics Can Expand Our Vision of Scholarly Communication and Social Impact, Inf. Stand. Q., vol. 25, no. 2, p. 27, 2013

[25] K.-K. Yan and M. Gerstein, The spread of scientific information: insights from the web usage statistics in PLoS article-level metrics, PLoS One, vol. 6, no. 5, 2011

[26] W. Gunn, Social signals reflect academic impact: What it means when a scholar adds a paper to Mendeley, Inf. Stand. Q., vol. 25, no. 2, pp. 33–39, 2013

[27] I. Guy, U. Avraham, D. Carmel, S. Ur, M. Jacovi, and I. Ronen, Mining expertise and interests from social media, WWW'13, pp. 515–526, 2013

[28] S. Ghosh, N. Sharma, and F. Benevenuto, Cognos: crowdsourcing search for topic experts in microblogs, SIGIR'12, no. 1, pp. 575–584, 2012

[29] G. Eysenbach, Correction: Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact, Journal of Medical Internet Research, 2012

[30] 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, 2012

[31] R. Costas, Z. Zahedi, and P. Wouters, Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective, CoRR'14, 2014

[32] S. Haustein, I. Peters, C. R. Sugimoto, M. Thelwall, and V. Larivíere, Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature, J. Assoc. Inf. Sci. Technol., 2013

[33] K. Wickre, Celebrating #Twitter7, [Accessed: 10-Mar-2014]. https://blog.twitter.com/2013/celebrating-twitter7, 2013

[34] D. Milne and I. H. Witten, An open-source toolkit for mining Wikipedia, Artif. Intell., vol. 194, pp. 222–239, 2013

[35] P. Siehndel and R. Kawase, Twikime! User profiles that make sense, In Proceedings of the 11th International Semantic Web Conference (Posters and Demos), 2012