Journal of Informetrics 9 (2015) 39–49

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members

José Luis Ortega *

Cybermetrics Lab, CCHS-CSIC, Albasanz, 26–28, 28037 Madrid, Spain

ARTICLE INFO

Article history:
Received 28 July 2014
Received in revised form 24 October 2014
Accepted 3 November 2014
Available online 24 November 2014

Keywords:
CSIC
Altmetrics
Bibliometrics
Academic social network sites
Academic search engines

ABSTRACT

This study explores the connections between social and usage metrics (altmetrics) and bibliometric indicators at the author level. It studies to what extent these indicators, gained from academic sites, can provide a proxy for research impact. Close to 10,000 author profiles belonging to the Spanish National Research Council were extracted from the principal scholarly social sites: ResearchGate, Academia.edu and Mendeley and academic search engines: Microsoft Academic Search and Google Scholar Citations. Results describe little overlapping between sites because most of the researchers only manage one profile (72%). Correlations point out that there is scant relationship between altmetric and bibliometric indicators at author level. This is due to the almetric ones are site-dependent, while the bibliometric ones are more stable across web sites. It is concluded that altmetrics could reflect an alternative dimension of the research performance, close, perhaps, to science popularization and networking abilities, but far from citation impact.

1. Introduction

Many papers have been written about the success of social network sites and how the quickly a platform emerges or disappears (Banbersta, 2010; Elder, 2014). These sites can experience an unexpected boom and then lose popularity (Garcia, Mavrodiev, & Schweitzer, 2013). It is used to say that these sites go with the fashions and the craving of the public. However, these services mostly depend on the use that people make of them, which not only means to contact with other members, but also to produce the content to share (Kordestani, Limayem, Salehi-Sangari, Blomgren, & Afsharipour, 2011). This would explain that some specific social network sites did not succeed (e.g. MySpace, Friendster), while other platforms, not oriented to social networking themselves, emerge indirectly as social network platforms to spread messages (Twitter), pictures (Flickr, Instagram) or documents (Scribd) (Wallace, 2013). In this sense, scholars seem to demand similar media for sharing and spreading their own research results to the scientific community. This could explain that the first initiatives of academic social networks did not become established, because they only had contacting tools such as forums and groups (e.g. Nature Network, BiomedExpert). The success of a social site for scholars will thus depend on its ability to promote the uploading and sharing of research documents between their members.

The current generation of academic social sites has focused its attention on the most important results that any researcher could create: publications. In this way, while ResearchGate and Academia.edu act as document posting services for full-text papers, Mendeley is meanly a bibliographic tool to share scientific references. This makes that scholars can explore the

* Tel.: +34 916022603.
E-mail address: jortega@orgc.csic.es

http://dx.doi.org/10.1016/j.joi.2014.11.004
1751-1577/

publications of their partners, provoking the interaction (views, downloads, followers, etc.) between them. The inclusion of publications in social sites turns the spotlight on research performance, and therefore, on research evaluation. In this way, academic social sites link bibliometric indicators with social and usage metrics in a same context, suggesting that these alternative indicators could be used as proxies or predictors of research impact (Priem & Hemminger, 2010).

The appearance of these alternative indicators (altmetrics) is causing a great transformation in the research evaluation. Because of they introduce a new perspective on the research activity, relating research impact and social skill. This makes possible the comparison of personal and organization profiles with the resulting development of public-access rankings (Thelwall & Kousha, 2014b). Following this idea, this work explores the relationship among all these indicators at author level to make clear the meaning of these metrics and their interactions.

2. Related research

The literature on social network sites for scholars was initially focused on the motivations and preferences of researchers in using these platforms (Almousa, 2011; Jeng, He, & Jiang, 2014; Kelly & Delasalle, 2012), and how their utilities could improve the research activity (Jeng, He, Jiang, & Zhang, 2012; Jiang, Ni, He, & Jeng, 2013; Thelwall & Maflahi, 2014). Scientometrics also paid its attention on these services, studying their significance for research evaluation. In this way, counting hits (Perneger, 2004), downloads (Brody, Harnad, & Carr, 2006; Watson, 2009), tweets (Eysenbach, 2011), recommendations (Li & Thelwall, 2012) and bookmarking of articles (Li, Thelwall, & Giustini, 2012) were correlated with citations to explore whether these new measures had some relationship with research impact. However, although many of these indices have shown statistically significant correlations, these are poor and low. Bar-Ilan et al. (2012) found poor correlation between citations and bookmarked papers in Mendeley. Thelwall and Kousha (2014b) detected moderated rank correlations between ResearchGate indicators for organizations and university rankings. Thelwall, Haustein, Larivière, and Sugimoto (2013) neither observed important correlations among a set of indicators from different social network sites. These results do not make clear whether these measures could be an expression of research impact or they truly arise from a different research activity. In this line, Bollen, Van de Sompel, Hagberg, and Chute (2009) used PCA and discovered that usage metrics form an opposed component regarding bibliometric counts. Priem, Piwowar, and Hemminger (2012) claimed that "altmetric indicators seem to measure impact mostly orthogonal to citation" to suggest that altmetric measures make up a component not correlated with the bibliometric ones. Mohammadi and Thelwall (2014) also settled the numbers of Mendeley readers could "reflect different aspects of the research impact". Similar conclusions were adopted by Costas, Zahedi, and Wouters (2014) and Zahedi, Costas, and Wouters (2013).

These studies manage at article level, where the relationship between research impact and usage is directly observable. However, these sites also offer a wide range of personal metrics that describe the social position of researchers within their closest immediate surroundings. In this sense, far fewer works have explored the presence of researchers in these online media. Haustein et al. (2014) tracking the presence of 57 scientometricians on the Web, found that 23% were in Google Scholar Citations and 16% had a Twitter account; whereas Mas-Bleda, Thelwall, Kousha, and Aguillo (2014) followed 1517 researchers in several social sites, uncovering a low adoption rate and a limited overlapping of academic social sites. Unfortunately, only Thelwall and Kousha (2014a) explored the linkage between social and usage metrics with bibliometric measurements at author level, revealing scant correlations between citations and Academia.edu's metrics.

According to academic search engines, and concretely, to their profiling tools, some studies have described its useful (Jacsó, 2011, 2012), coverage (Ortega, 2014) and limitations (Delgado Lopez-Cozar, Robinson-García, & Torres-Salinas, 2012). Ortega and Aguillo (2014) compared profiles between Microsoft Academic Search and Google Scholar Citations detecting a larger proportion of citations (327%) and papers (158%) in Google Scholar by comparison with Microsoft Academic Search. Haley (2014) also observed similar results collating 50 economists in the aforesaid academic search engines. Khabsa and Giles (2014) also compared both search engines to estimate the number of web-accessible research papers, finding that Google Scholar is the most comprehensive engine with close to 87% of the scientific literature.

Spite of this, fewer articles have studied the relationships among different metrics into and between social sites at author level.

3. Objectives

The main intent of this work is to explore the relationship between social and usage metrics, also known as altmetrics, and traditional bibliometric indicators at author level. The aim is to discuss whether these measures are similar between them, and therefore, they can be considered a proxy for research evaluation; or in contrast, they are because of a different view of research activity independent of research impact and closer to research popularization and social skills. Several research questions were formulated at this point:

- Do these indicators depend more or less on the web sites that produce them? And does this fact imply any limitation for research assessments?
- Is there any connection between altmetric (social and usage) and bibliometric indicators? And could the first ones be used as proxies for research impact?

Besides, it is intended to see how profiles from a same institution populate different academic social sites and how these sites are overlapped between them. This would respond if researchers tend to use several social sites or contrarily they only use one or two platforms.

A third objective is, because of different sites contain similar indicators, to identify the most relevant ones. The aim is to obtain indicators that statistically contain more information and, at the same time, avoid the harvesting of repeated data.

4. Methods

4.1. Object of study: CSIC

Consejo Superior de Investigaciones Científicas (CSIC) is the largest research organization in Spain, which comprises around 13,000 staff members and more than 120 research institutes and centres (CSIC, 2013). The reasons to choose CSIC for this study are:

- Size: Being one of the largest research institutions in the world ensures the obtaining of a wide and statistically representative sample.
- Controlled population: A sample from a specific institution makes easy the retrieval of profiles, the unambiguous identification of users and disambiguating researchers with similar names.
- Multidisciplinary centre: CSIC is divided in 8 research areas, going from Humanities and Social Sciences (Area 1) to Chemical Science and Technologies (8). This allows to ensure the sample is well-adjusted by subject matter and it represents every type of research activity.

4.2. Sources and data extraction

Two academic search engines (Microsoft Academic Search and Google Scholar Citations) were used to extract bibliometric information, while three academic social sites (ResearchGate, Academia.edu and Mendeley) were employed to extract the altmetrics data. The reason to take these web sites is that they are the academic sites that included most researchers' profiles and they are the most popular in the scholarly community at the moment (Nentwich & König, 2014).

Microsoft Academic Search (MAS) is an academic search engine which automatically creates a personal profile from metadata extracted from CrossRef and other publishing sources. These profiles display bibliographic information (publications list, co-authors, keywords, etc.) and bibliometric indicators (citations and papers). At the same time, they are thematically classified and assigned to an organization. In this way, all profiles linked to "Consejo Superior de Investigaciones Cientificas" were harvested. However, MAS presents some problems. The most important is the high presence of repeated profiles, which was estimated at 11% (Ortega & Aguillo, 2014). This causes a hard cleaning process, removing repeated profiles and merging paper and citation counts. Evidence of the seriousness of this problem is the project ALIAS which attempts to improve the disambiguation of authors (Pitts, Savvana, Basu Roy, & Mandava, 2014). Another important disadvantage is the slow updating (once by year) of new records. For example, only 8.8% of the papers were published after 2010 which it brings a few recent publications (Ortega, 2014). Finally, another problem of the automatic creation of profiles is that many of these profiles belong to distant periods and therefore they are inactive profiles. Due to this, only active profiles since 2000 were selected (75.2%) because it is not possible to compare the research performance before 2000 with current social activity.

Google Scholar Citations (GSC) presents a brief curriculum vitae where researchers list their publications indexed in Google Scholar with some bibliometric indicators. Unlike to MAS, these profiles are created and edited by the users themselves so the information on each researcher is optional and written in natural language. This provokes the principal disadvantage of GSC, that is, the hard and tedious task of normalization and identification of affiliations. Three queries were therefore launched to retrieve the largest number of CSIC's researchers: CSIC, "Consejo Superior de Investigaciones Cientificas" and "Spanish National Research Council". In consequence, profiles without this institutional information were not retrieved because they do not belong to CSIC.

ResearchGate (RG) is a social network site that allows uploading papers, taking part in discussions and following other researchers. RG is the site that most indicators show at author level, going from social measurements (followers, following) and usage metrics (page view, document downloads) to bibliometric indicators (impact points, papers and citations). RG score is a compound index calculated from all the other ones. Impact Points is the addition of the Journal Impact Factor (JIF) of the sources where each paper was published. However, this indicator takes their values from the last JIF updating independently from the publication date, so articles from different years have the same JIF. Authors are optionally able to link with their academic institutions; therefore the Institutions section does not ensure that all researchers from CSIC are actually attached to their organization. For example, 4% of CSIC's researchers are not linked to "Spanish National Research Council". Other problem is that researchers from mixed centres are linked to university. To solve these problems, the three above queries were then launched to retrieve all the CSIC's researchers and to take the widest picture.

Academia.edu is a web platform centred in hosting academic papers that can be shared among their users. Academica.edu allows to build an own profile along with the list of documents uploaded to Academia.edu. This profile is completed with statistics on usage (views) and social interactions (followers/following). As in RG, each author profile is assigned to an

Table 1
Number of profiles according to academic profiling service.

	MAS	MAS from 2000	GSC	RG	Academia.edu	Mendeley
Total	12,115	8616	1342	3480	778	805
Unique	9448	7106	1325	3429	751	787
Duplicate	2667	1510	17	51	27	18
%	22	17.5	1.3	1.5	3.5	2.2

institution, but in this case the action is compulsory. All the profiles linked to "CSIC (Consejo Superior de Investigaciones Científicas-Spanish National Research Council)" were then extracted.

Mendeley is focused on sharing bibliographic references and discussing in thematic groups. Unlike the rest of social sites, Mendeley is more oriented to papers than authors. In fact, the number of readers is being used by Scopus as impact measure of articles, while author profiles are being less explored. Due to this, Mendeley is the academic social site that displays less information on profiles, including only number of followers/following as social indicators. This is the hardest service to retrieve authors belonging to a certain institution. Because of its search interface is designed to retrieve solely author names. A list of names from the previous sites (MAS, GSC, RG and Academia.edu) was then launched to Mendeley to return profiles from CSIC's researchers.

A specific SQL script was written for each service to automatically extract profiles and their indicators. This process was done from April to May 2014. Only for RG this task was done offline because of some limits in its HTML code.

4.3. Disambiguation and data cleaning

The first step to tie each author with their profiles was to split their names in three parts: name, surname 1 and surname 2 (in Spanish it is usual to use two surnames). After a manual revision, abbreviations and misspellings were corrected to make uniform these names and detect duplicate profiles. These profiles were treated in a different way in each service. In MAS' case, for example, they were summed because the profile's names are extracted from papers. It is not then possible that two profiles contain the same paper. However, in the rest of services, profiles with less content were removed.

Table 1 summarises the total count of profiles, duplicate profiles and percentage. The site with the largest number of profiles is MAS, followed by far by RG and GSC. This large amount of profiles in MAS could be due to several reasons. The first one is that MAS creates profiles from each article since any date. This causes that there are old profiles from inactive researchers. From 12,115 profiles only 8616 (7106 unique) were considered because they had published at least one paper since 2000. The second reason is the slow updating rate of their databases which causes that many profiles are outdated or with former affiliations. Authors that have thus left CSIC are still listed in the organization, while new researchers keep their former addresses. The third problem is that a profile can be linked only to one organization, preventing mixed research centres. Therefore, many of those researchers are assigned to the university instead to the CSIC. Mendeley's list could have some limits because it was generated from the profiles of the other services. Therefore, only profiles from authors previously registered in MAS, GSC, RG and Academia.edu were retrieved.

The next step was to compare each list of profiles through crossed queries that match the presence of a same author in different web services. In cases of authors which fit in with several profiles, different heuristics were employed to individuate each profile:

- A picture of each profile was downloaded to compare if that image also looks like to the same person in other services.
- Affiliations, interests, labels and thematic classifications were compared to confirm different profiles from a same author.
- Co-authors and followings/followers lists were explored to identify similar partners. This is because profiles with the same name and similar contributors are assumed that belong to the same person.
- Finally, publications list was also examined to identify affiliations and research interests.

4.4. Indicators

Table 2 gathers indicators from academic social sites and groups them in three categories: Bibliometric, Social and Usage indicators. The first one alludes to measures related to the research impact (#citations) and productivity (#papers). Social indicators describe the way in which users of a social site interact among them. And finally, Usage indicators express the use that is made of each profile.

5. Results

5.1. Overlapping

Table 3 and Fig. 1 describe the number of profiles in each service and the percentage of overlapping between them. Venn diagrams were generated thanks to Tim Hulsen's application (http://www.cmbi.ru.nl/~timhulse/venn/). Diagonal values

Table 2
Indicators extracted from academic profiling services and grouped by classes.

	Bibliometric	Social	Usage
Academia.edu	Academia_Papers	Academia_Followers Academia_Following	Academia_Document_views Academia_Profiles_views
GSC	GSC_Citations GSC_Hindex GSC_Papers		
MAS	MAS_Citations MAS_Papers		
Mendeley		Mendeley_Followers Mendeley_Following	
RG	RG_Impact points RG_Publications RG_Citations	RG_Following RG_Followers	RG_Views RG_Downloads

correspond to unique profiles in each service. In this way, MAS is the service with less overlapping, with 71% of their profiles not included in other services; while GSC, with 26%, is the site that most profiles have in common. In Mendeley's case, there are not unique profiles because the Mendeley's list was taken from the already existing profiles in other services. Due to this, Mendeley is the service with most overlapping profiles. From a two-way view, 80% of Mendeley members are also included in MAS, whereas only 8% of MAS profiles are GSC members as well. In general, Fig. 1 displays that there are not many overlapped profiles and that only a few researchers manage various profiles in different sites. In fact, 7093 (72%) researchers are only present at one, 2055 (21%) in two and 508 (5%) in three services, which it confirms that the maintenance of user profiles is hard and time-consuming.

5.2. Correlations

Fig. 2 plots the correlation matrix between the metrics from the academic social sites. All the correlations were calculated with the Spearman's rank correlation coefficient and results above $\rho = \pm 0.2$ are significant at 0.05 level. Pair-wise deletion procedure was used to avoid spurious correlations caused by high number of zeros and null values. The first impression is that each indicator depends on its own service because the best correlations are obtained between measures from the same site. For example, similar indicators such as followers/followings from different sites do not show high correlations (see Academica.edu and RG and, in less extend, Academia.edu and Mendeley). Contrarily, different indicators from a same site strongly correlate between them (see RG). In general, Fig. 2 shows that Academia.edu and Mendeley indicators scarcely correlate with similar indices from the other services, while GSC, MAS and RG indeed give significant correlations between them. This could owe to bibliometric indicators, because these three services are the only ones that include those measurements. In fact, bibliometric indicators are those that best correlate across platforms.

According to the influence between groups of indicators, the results show that social metrics present poor correlations with bibliometric ones, excepting *RG_Followers* with *RG_Citations* ($\rho = 0.64$) and *GSC_Citations* ($\rho = 0.62$). This could suggest that profiles with an elevated impact attract other scholars interested in the value of their works. Usage metrics have better correlations with bibliometric indices such as *RG_Views* with *GSC_Citations* ($\rho = 0.83$) and *RG_Downloads* with *RG_Citations* ($\rho = 0.69$). These indicators correlate better with the impact than with production, showing that the visits and documents downloading depend more on the quality of the papers than on the amount of these.

It is worthwhile stopping in the indicator *RG_Score*, which is a composite index drawn from RG's indicators, but whose algorithm is unknown (ResearchGate, 2014). Correlations show that *RG_Impact points* ($\rho = 0.98$) and *RG_Views* ($\rho = 0.9$) are the most related measures, beside *RG_Publications* and *RG_Citations* ($\rho = 0.87$). This points out that RG score relies more on bibliometric and usage factors, than on social ones (*RG_Followers* $\rho = 0.67$; *RG_Following* $\rho = 0.32$).

5.3. Principal component analysis (PCA)

One of the objectives of this work is to summarise this large set of indicators with different meaning and origins. The aim is to detect the most significant and informative indices. PCA is used to perform this task because it allows to reduce correlated variables to a limited group of factors or components with the highest amount of information. After a log

Table 3
Overlapping of academic profiling services.

	MAS	GSC	RG	Academia.edu	Mendeley	Total
MAS	5042 (71%)	567 (8%)	1448 (20%)	207 (3%)	627 (9%)	7109
GSC	567 (43%)	338 (26%)	627 (47%)	165 (12%)	354 (27%)	1325
RG	1448 (42%)	627 (18%)	1416 (41%)	275 (8%)	488 (14%)	3429
Academia.edu	207 (28%)	165 (22%)	275 (37%)	357 (48%)	120 (15%)	751
Mendeley	627 (80%)	359 (45%)	488 (61%)	120 (15%)	0	787

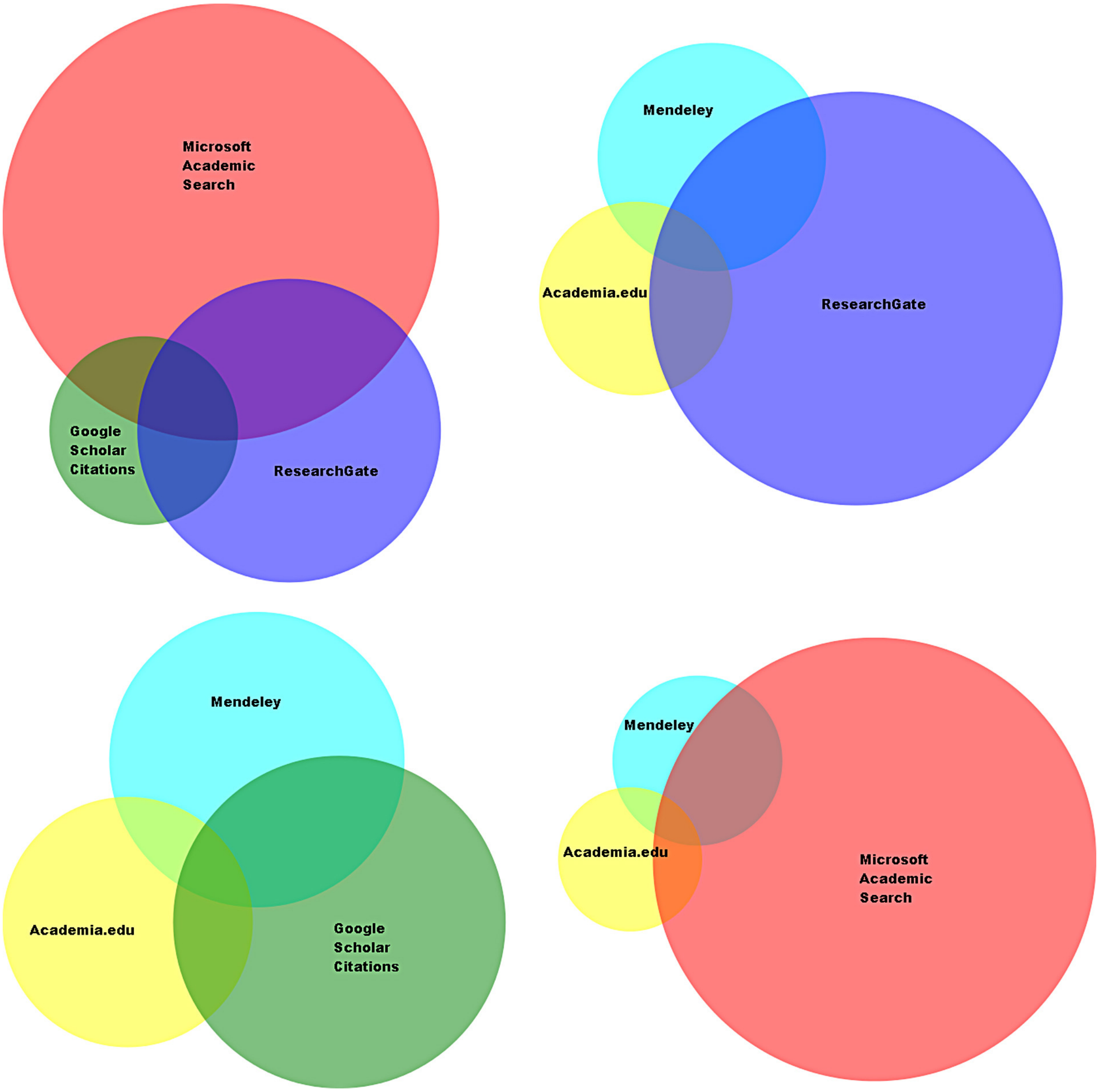

Fig. 1. Venn diagrams describing overlapping between academic profiling services.

transformation and the use of Pearson's correlation coefficient, PCA held four components with a cumulated variance of 79%. This makes clear that not all the indicators are strongly correlated between them and therefore they contain different information.

Table 4 and Fig. 3 describes correlations between profile's indicators and PCA's components. Remember that each component contains different information to the other ones, so they are uncorrelated between them. The results confirm previous findings about the dependence of these indicators on their corresponding sites. The first component groups indicators from GSC, MAS and RG; the second one from Academia.edu; the third one from Mendeley; and finally, the fourth one gathers the residual information of *RG_Following*. According to the first component, the most important correlations are with *RG_Views* ($r = 0.95$), *RG_Publications* ($r = 0.94$) and *GSC_hindex* ($r = 0.93$), which it suggests that this component is mainly built on bibliometric indicators. Then, the first component gathers indicators bound to the scientific production and impact; while the second one is closer to social indicators, although the site-dependence distorts this view. For example, an analysis with only RG's indicators reproduces this distinction between social and bibliometric measures.

Table 4
The first four components and their correlations with the profile's indicators.

	F1	F2	F3	F4
Academia_Papers	0.447	**0.547**	−0.056	0.296
Academia_Document_views	−0.054	**0.810**	−0.357	0.228
Academia_Profiles_views	−0.100	**0.840**	−0.212	−0.098
Academia_Followers	−0.125	**0.864**	−0.265	−0.132
Academia_Following	−0.159	**0.831**	−0.072	−0.057
GSC_Citations	**0.921**	−0.039	−0.097	−0.260
GSC_Hindex	**0.930**	−0.033	−0.110	−0.238
GSC_Papers	**0.837**	0.114	−0.184	−0.215
MAS_Citations	**0.705**	−0.025	−0.244	−0.264
MAS_Papers	**0.706**	0.055	−0.363	−0.195
Mendeley_Followers	−0.091	0.469	**0.782**	−0.316
Mendeley_Following	−0.135	0.497	**0.738**	−0.387
RG_Score	**0.827**	−0.156	0.231	0.049
RG_Impact points	**0.853**	−0.174	0.205	0.054
RG_Publications	**0.939**	0.088	0.021	0.026
RG_Citations	**0.902**	−0.053	0.067	−0.135
RG_Views	**0.956**	0.073	0.078	0.070
RG_Downloads	**0.772**	0.173	0.110	0.230
RG_Following	0.362	0.243	0.392	**0.589**
RG_Followers	**0.748**	0.180	0.184	0.375

Values in bold correlations higher than 0.5.

PCA was also used to answer the third question of this study, detecting the most informative variables. Contribution analysis allows to reject variables that provide little information, reducing the effort in extracting these data. In this way, the results produce that MAS and Mendeley indicators are the variables that less information gives to their components (Table 5). Thus, *MAS_Citations* only provides 5.5% to the first component, while *GSC_Citations* (9.4%) and *RG_Citations* (8.9%) almost double the MAS contribution. The same is found with Mendeley measures, *Mendeley_Followers* (5.9%) and *Mendeley_Following* (6.7%) contribute much less than *Academia_Followers* (19.9%) and *Academia_Following* (18.5%). These results suggest removing MAS as bibliometric source. In the same way, Mendeley can be avoided as altmetric resource due to its low contribution and the aforesaid problems to extract data on their profiles.

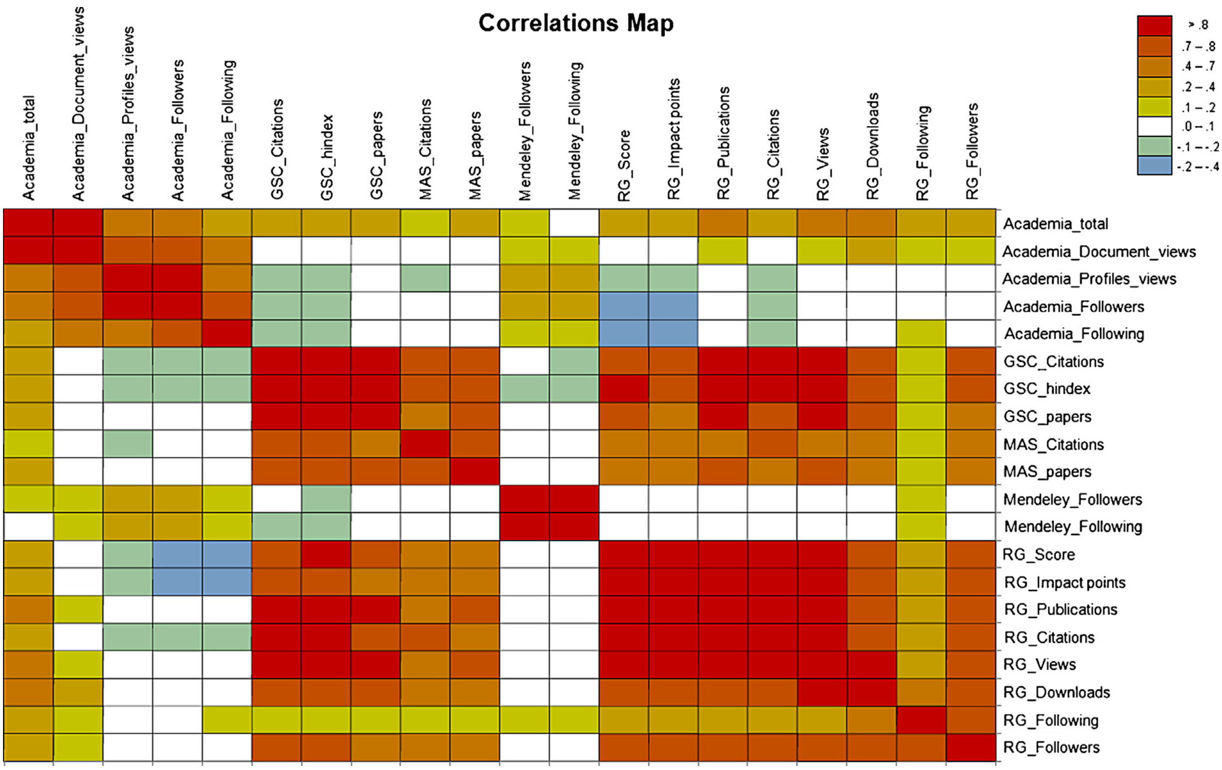

Fig. 2. Correlation matrix among indicators from academic profiling services (Spearman's rank).

Table 5
Contribution of the variables (%) after Varimax rotation.

	D1	D2
Academia_Papers	2.523	**7.168**
Academia_Document_views	0.001	**17.361**
Academia_Profiles_views	0.032	**18.794**
Academia_Followers	0.066	**19.963**
Academia_Following	0.143	**18.551**
GSC_Citations	**9.379**	0.211
GSC_Hindex	**9.561**	0.187
GSC_Papers	**7.888**	0.121
MAS_Citations	**5.499**	0.107
MAS_Papers	**5.584**	0.006
Mendeley_Followers	0.047	**5.914**
Mendeley_Following	0.129	**6.688**
RG_Score	**7.440**	1.065
RG_Impact points	**7.901**	1.283
RG_Publications	**9.890**	0.035
RG_Citations	**8.965**	0.278
RG_Views	**10.233**	0.011
RG_Downloads	**6.782**	0.446
RG_Following	**1.565**	1.306
RG_Followers	**6.372**	0.505

Values in bold contributions higher than 5%.

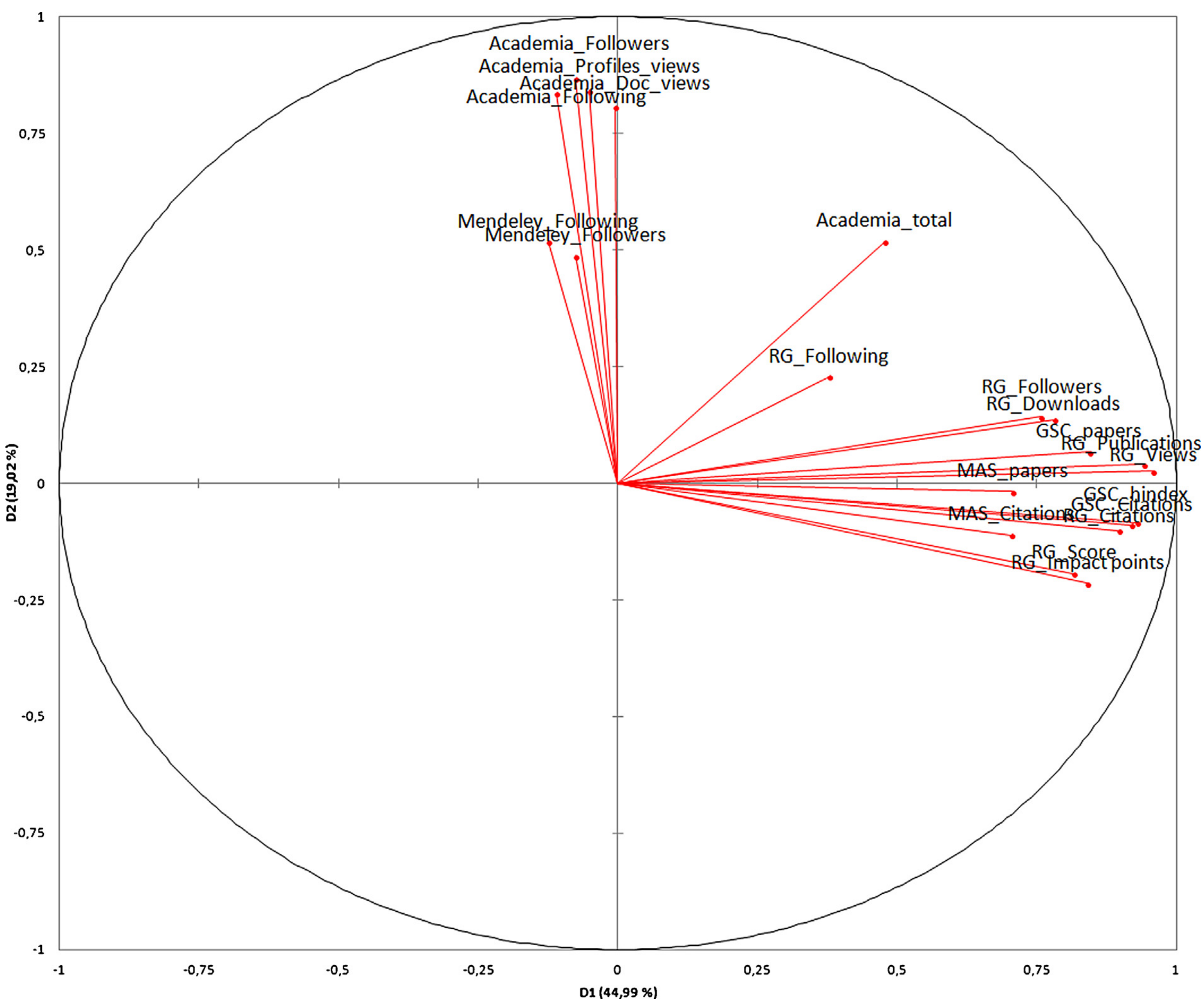

Fig. 3. PCA plot. Variance 64%.

6. Discussion

Results present little overlapping between profiling services because to manage these profiles entails time-consuming and extra efforts to update information and documents, as well as interacting with other users. Previous studies evidenced this fact to find little presence of researchers in social sites (Ponte & Simon, 2011; Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011) and very poor overlapping between these sites (Mas-Bleda et al., 2014). This also suggests the behaviour of a researcher in one social site could be very different across platforms. Further, authors with a very active profile in a service, at the same time, possibly maintain inactive profiles in other sites.

This disparity would influence the correlations between indicators and explain why the correlations are influenced by the profiling site where those indicators are implemented. Results show that different indicators from a same site correlate better than similar indicators from different sites. However, these differences are only evident in social and usage metrics because these measures are determined by inherent factors in each site. For example, the amount of users in the network, disciplinary biases in the distribution of the community (Ortega & Aguillo, 2012, 2014) or how each measure is counted and generated, are factors that determine the performance of a profile in a web site. Besides, the personal attitude of each author in distinct services can produce very different numbers. For instead, one author with no uploaded papers in Mendeley would hardly attract followers, while a very active member in RG, following and viewing profiles, could produce a reflection effect.

Nevertheless, bibliometrics indicators have demonstrated that they are more stable across services. Thus, papers and citations from MAS, GSC and RG describe high correlations. This could be due to that impact and production indices are independent of the site (one author can be cited independently if he/she has a profile in a web service). This causes that researchers show similar figures in different environments.

From a research evaluation view, social and usage indicators do not consistently correlate with the bibliometric ones, excepting between RG's metrics. Even in this case, these high correlations could be due to multiple interactions between variables. For example, it is quite logical that a prestigious author, with many papers and citations, attracts views, downloads and followers as well. Further, these correlations could be interpreted more as a result than as a cause of the research impact. Previous studies either found no significant correlations between social and bibliometric indicators. Thelwall and Kousha (2014a) did not see large correlations between Academica.edu indicators and bibliometric ones. Similarly, Thelwall et al. (2013) did not observe relevant relationships among these measurements either.

However, the poor correlations between social and usage indicators and bibliometric ones, positively tell us that they contain different information that has to be appreciated in the research evaluation of a scientist. PCA has shown that the second component, with social and usage metrics, describes a different perspective about the scientific habits, different from the usual environments of citing and publishing papers. This statement is confirmed by Bollen et al. (2009), who found that usage metrics frame a differentiate component regarding traditional citation-based indices. I consider that this societal dimension cannot be avoided and has to be valued, not as instruments for bibliometric research evaluation exercises, but as indicators of information spreading, science popularization or social influence (Cronin, 2013).

Nevertheless, this study presents some limitations that could affect to the results. The first one is that this sample is only representative of CSIC's researchers. Recent studies have shown that many academic social sites are skewed to certain disciplines (Ortega & Aguillo, 2012; Thelwall & Kousha, 2014a, 2014b), countries or organizations (Menendez, de Angeli, & Menestrina, 2012; Ortega, 2014). This unbalanced distribution could affect to the presence and overlapping of profiles in each service. However, the overlap between MAS and GSC (8%), similar to found by Ortega and Aguillo (2014), could be a sign of well-adjusted and representative sample. Even so, more studies on population distribution in academic social sites would be advisable to take a more precise picture on the coverage of these sites. But perhaps the most important limitation comes from the creation date of profiles. Social and usage indicators are very sensitive to this fact because as older a profile is the more followers/following could have and more views and downloads would receive. Due to this, new analyses would be welcome to precise the relationship between these indicators and solve these limitations.

7. Conclusions

Exploring social, usage and bibliometric indicators at author level allows to draw several relevant conclusions:

- Usage (views and downloads) and social (followers/followings) indicators are influenced by their own social sites, while bibliometric (papers, citations) indices are independent and therefore more stable across services.
- Correlations between social and usage metrics regarding bibliometric measurements are poor in every site analysed and they do not permit to claim that altmetrics could be a proxy for research evaluation, at least at author level. On the contrary, they could describe an alternative dimension of the academic uses, close to science popularization, networking abilities and social skills.

According to the peopling of these scholarly sites, the results show that most of the researchers are included in only one academic site, confirming a limited number of overlapping profiles. This behaviour could be motivated by time-consuming and maintenance efforts.

Finally, PCA suggests that indicators from MAS and Mendeley are low contributory sources. This means that the information that these sites provide are barely relevant to the model, so they might be removed from altmetric and bibliometric studies at author level.

References

Almousa, O. (2011). Users' classification and usage-pattern identification in academic social networks. In *Proceedings of 2011 IEEE Jordan conference on applied electrical engineering and computing technologies*. Amman, Jordania: IEEE.

Banbersta, M. (2010). *The success factors of the social network sites "Twitter"*. Utrecht: Utrecht University of Applied Sciences. http://crossmedialab.nl/files/SNS_Research.pdf

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations: Scholars' visibility on the social Web. In É. Archambault, Y. Gingras, & V. Larivière (Eds.), *Proceedings of 17th international conference on science and technology indicators* (p. 14). Montréal: Science-Metrix and OST.

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, *4*(6), e6022.

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology*, *57*(8), 1060–1072.

Costas, R., Zahedi, Z., & Wouters, P. (2014). *Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective*. arXiv:1401.4321

Cronin, B. (2013). Metrics à la mode. *Journal of the American Society for Information Science and Technology*, *64*(6), 1091.

CSIC. (2013). *Data 2012: Research centers and institutes*. Madrid, Spain: CSIC. http://documenta.wi.csic.es/alfresco/downloadpublic/direct/workspace/SpacesStore/d7069f1e-b26e-46ae-935f-313c44d875d0/CSIC_DATA_2012.pdf

Delgado Lopez-Cozar, E., Robinson-García, N., & Torres-Salinas, D. (2012). *Manipulating Google Scholar citations and Google Scholar metrics: Simple, easy and tempting*. arXiv:1212.0638

Elder, J. (2014). Social media fail to live up to early marketing hype. *The Wall Street Journal*, http://online.wsj.com/articles/companies-alter-social-media-strategies-1403499658

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, *13*(4), e123.

Garcia, D., Mavrodiev, P., & Schweitzer, F. (2013). Social resilience in online communities: The autopsy of Friendster. In S. Muthu Muthukrishnan, A. El Abbadi, & B. Krishnamurthy (Eds.), *Conference on online social networks* (pp. 39–50). Boston, USA: ACM.

Haley, M. R. (2014). Ranking top economics and finance journals using Microsoft Academic Search versus Google scholar: How does the new publish or perish option compare? *Journal of the Association for Information Science and Technology*, *65*(5), 1079–1084.

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 1–19. http://dx.doi.org/10.1007/s11192-013-1221-3

Jacsó, P. (2011). The pros and cons of Microsoft Academic Search from a bibliometric perspective. *Online Information Review*, *35*(6), 983–997.

Jacsó, P. (2012). Google Scholar author citation tracker: Is it too little, too late? *Online Information Review*, *36*(1), 126–141.

Jeng, W., He, D., & Jiang, J. (2014). *User participation in an academic social networking service: A survey of open group users on Mendeley*. arXiv:1401.6495

Jeng, W., He, D., Jiang, J., & Zhang, Y. (2012). Groups in Mendeley: Owners' descriptions and group outcomes. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–4.

Jiang, J., Ni, C., He, D., & Jeng, W. (2013). Mendeley group as a new source of interdisciplinarity study: How do disciplines interact on Mendeley? In J. S. Downie, R. H. McDonald, T. W. Cole, R. Sanderson, & F. Shipman (Eds.), *Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries* (pp. 135–138). Indianapolis, USA: ACM.

Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *PLOS ONE*, *9*(5), e93949.

Kelly, B., & Delasalle, J. (2012). Can LinkedIn and Academia.edu enhance access to open repositories? In *OR2012: The 7th international conference on open repositories* Edinburgh, Scotland.

Kordestani, A. A., Limayem, M., Salehi-Sangari, E., Blomgren, H., & Afsharipour, A. (2011). Why a few social networking sites succeed while many fail. In M. Conway (Ed.), *Proceedings of the annual conference of the academy of marketing science, Coral Gables, USA*.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In É. Archambault, Y. Gingras, & V. Larivière (Eds.), *Proceedings of 17th international conference on science and technology indicators* (pp. 451–551). Montréal: Science-Metrix and OST.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, *91*(2), 461–471.

Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? *Scientometrics*, http://dx.doi.org/10.1007/s11192-014-1345-0

Menendez, M., de Angeli, A., & Menestrina, Z. (2012). Exploring the virtual space of academia. In J. Dugdale, J. Dugdale, et al. (Eds.), *From research to practice in the design of cooperative systems: Results and open challenges*. London: Springer-Verlag.

Mohammadi, E., & Thelwall, M. (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology*, http://dx.doi.org/10.1002/asi.23071

Nentwich, M., & König, R. (2014). Academia goes Facebook? The potential of social network sites in the scholarly realm. In S. Bartling, & S. Friesike (Eds.), *Opening Science* (pp. 107–124). London: Springer International Publishing.

Ortega, J. L. (2014). *Academic search engines: A quantitative outlook*. Cambridge, UK: Chandos Publishing. ISBN: 1843347911.

Ortega, J. L., & Aguillo, I. F. (2012). Science is all in the eye of the beholder: Keyword maps in Google Scholar Citations;. *Journal of the American Society for Information Science and Technology*, *63*(12), 2370–2377.

Ortega, J. L., & Aguillo, I. F. (2014). Microsoft Academic Search and Google Scholar Citations: A comparative analysis of author profiles. *Journal of the American Society for Information Science and Technology*, *65*(6), 1149–1156.

Perneger, T. V. (2004). Relation between online 'hit counts' and subsequent citations: Prospective study of research papers in the BMJ. *British Medical Journal*, *329*(7465), 546–547.

Pitts, M., Savvana, S., Basu Roy, S., & Mandava, V. (2014). ALIAS: Author disambiguation in Microsoft academic search engine dataset. In S. Amer-Yahia, S. Amer-Yahia, et al. (Eds.), *Proceedings of the 17th international conference on extending database technology* (pp. 648–651). Athens, Greece: OpenProceedings.

Ponte, D., & Simon, J. (2011). Scholarly communication 2.0: Exploring researchers' opinions on Web 2.0 for scientific knowledge creation, evaluation and dissemination. *Serials Review*, *37*(3), 149–156.

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday*, *15*(7.) http://dx.doi.org/10.5210/fm.v15i7.2874

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). *Altmetrics in the wild: Using social media to explore scholarly impact*. arXiv:1203.4745

ResearchGate. (2014). *RG score FAQ – ResearchGate*. https://www.researchgate.net/publicprofile.RGScoreFAQ.html

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, *24*(3), 183–195.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLOS ONE*, *8*(5), e64841.

Thelwall, M., & Kousha, K. (2014a). Academia.edu: Social network or academic network? *Journal of the Association for Information Science and Technology*, *65*(4), 721–731.

Thelwall, M., & Kousha, K. (2014b). ResearchGate: Disseminating, communicating, and measuring scholarship? *Journal of the Association for Information Science and Technology*, http://dx.doi.org/10.1002/asi.23236

Thelwall, M., & Maflahi, N. (2014). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology*, http://dx.doi.org/10.1002/asi.23252

Wallace, D. (2013). R.I.P. – Top 10 failed social media sites. *Search Engine Journal*, http://www.searchenginejournal.com/r-i-p-top-10-failed-social-media-sites/57554/

Watson, A. B. (2009). Comparing citations and downloads for individual articles. *Journal of Vision*, *9*(4), http://www.journalofvision.org/content/9/4/i.

Zahedi, Z., Costas, R., & Wouters, P. F. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? In *PLoS ALM Workshop* San Francisco, USA.