ResearchGate Articles: Age, Discipline, Audience Size and Impact[1]

Mike Thelwall and Kayvan Kousha
Statistical Cybermetrics Research Group, University of Wolverhampton.

The large multidisciplinary academic social web site ResearchGate aims to help academics to connect with each other and to publicise their work. Despite its popularity, little is known about the age and discipline of the articles uploaded and viewed in the site and whether publication statistics from the site could be useful impact indicators. In response, this article assesses samples of ResearchGate articles uploaded at specific dates, comparing their views in the site to their Mendeley readers and Scopus-indexed citations. This analysis shows that ResearchGate is dominated by recent articles, which attract about three times as many views as older articles. ResearchGate has uneven coverage of scholarship, with the arts and humanities, health professions, and decision sciences poorly represented and some fields receiving twice as many views per article as others. View counts for uploaded articles have low to moderate positive correlations with both Scopus citations and Mendeley readers, which is consistent with them tending to reflect a wider audience than Scopus-publishing scholars. Hence, for articles uploaded to the site, view counts may give a genuinely new audience indicator.

Introduction

Researchers can join many different websites in order to publicise their research. A new paper might be published in the journal publisher's website and may be free to anyone if the journal is gold Open Access (OA) (i.e., all of its articles are open access), or the author may pay the journal an OA fee. Alternatively, or in addition, the author may self-archive a preprint in a subject or institutional repository (Swan & Brown, 2005), on their own website (Kousha & Thelwall, 2014) or in an academic social web site like ResearchGate or Academia.edu. Nevertheless, whichever strategies are chosen, there is a trade-off between the time needed to upload information and the benefits of the extra publicity (Ward, Bejarano, & Dudás, 2015). In this context it is important to assess the benefits of each site in order to allow rational choices about which, if any, to use.

There is some research about the benefits of online publicity for academic articles. OA articles tend to be more cited (e.g., Wang, Liu, Mao, & Fang, 2015) but this could be because they tend to be better rather than because they are easier to access (Davis, 2011), and there is no evidence that is specific to articles uploaded to academic social web sites. There is also evidence that citations and usage metrics from digital libraries or subject repositories, such as views or downloads of articles, correlate with each other, suggesting that usage metrics partially reflect scholarly impact (e.g., Kurtz et al., 2005; Brody, et al., 2006; Duy & Vaughan, 2006). This seems likely to be also true for articles in academic social web sites but has not been tested.

In addition to the above knowledge gaps, no investigation has assessed the disciplinary or age coverage of ResearchGate to characterise typical articles uploaded and to assess whether there are different levels of interest in them. This is an important omission

[1] This is a preprint of an article to be published in the Journal of the Association for Information Science and Technology © copyright 2014 John Wiley & Sons, Inc.

because 1) this information can be used to guide recommendations for the types of authors and articles that would most benefit from using the site (see also: Ward, Bejarano, & Dudás, 2015), 2) levels of uptake can give contexts to institutional analyses of associated ethical and intellectual property issues (Arènes, 2015), for proposals to use ResearchGate as a data source for alternative metrics (Campos Freire & Valencia, 2015; Scarlat, Mavrogenis, Pećina, & Niculescu, 2015) and 3) some research uses ResearchGate itself as an important source of articles for research (e.g., Kamath, Setlur, & Yerlagudda, 2015; Łaczmański, Jakubik, Bednarek-Tupikowska, Rymaszewska, Słoka, & Lwow, 2015; McKellar & Currie, 2015; Velik, 2015). In response, this study analyses a large sample of publications recently uploaded to ResearchGate in order to assess their age and discipline and the relationship between views in the site and other academic indicators.

Background

ResearchGate launched in 2008 with the stated aim of helping researchers to communicate, cooperate and share information. It is free to join and (as of August 2015) each member has a profile page giving brief biographical information and a publication list. Every publication registered also has its own page giving metadata and, in some cases, a preview and a link to a full text version, if the author has uploaded one to the site and the publisher has not requested that it be removed for copyright reasons (Clarke, 2013). The publication page also reports article-level metrics, such as the number of times that a work has been downloaded and viewed. ResearchGate members can also connect to others in the site, affiliate with a specific institution, and register an interest in academic topics. Social network support is apparently an important feature of the site (Kadriu, 2013; Ovadia, 2014) and ResearchGate has tried different interfaces to help members to communicate effectively (Goodwin, Jeng, & He, 2014). The site also sends automatic email alerts to people about activities related to their profile and publications. Within the social part of the site, offline regulation seems to be important because, for example, answers from more authoritative figures tend to be more highly regarded (Li, He, Jeng, Goodwin, & Zhang, 2015). Nevertheless, few users seem to use the social side of the site (e.g., Alheyasat, 2015), with members seeing it mainly as a means to disseminate their research (Corvello, Genovese, & Verteramo, 2014).

ResearchGate calculates a range of statistics about members and institutions. These statistics seem to broadly reflect other academic-related rankings although they are skewed by factors such as greatly differing international levels of uptake of the site (Thelwall & Kousha, 2015). According to the site, the RG Score "measures scientific reputation based on how all of your research is received by your peers" (http://www.researchgate.net/publicprofile.RGScoreFAQ.html). The score calculations are not transparent, however, and depend on Journal Impact Factors and so are inappropriate for the assessment of individual academics (Jordan, 2015; Kraker & Lex, 2015). An investigation of the ResearchGate presence of 23 South African universities found significant positive correlations between average Web of Science citations per paper and average ResearchGate downloads, views, impact points and score for each institution (i.e., n=23), suggesting that ResearchGate use reflects academic interest or impact well, at least at the level of entire institutions (Onyancha, 2015).

An increasing number of academics seem to be using ResearchGate. Currently (August 2015), ResearchGate has more than 7 million users (http://www.researchgate.net/about), which is an increase of 3 million compared with July 2014 (see, Hoffmann, Lutz, & Meckel, 2015). It seems also to be the most popular academic

social web site. A survey of 160 academics in an Indian university found that 54% had ResearchGate profiles compared with 51% for Academia and 39% for LinkedIn and 35% for CiteULike (Madhusudhan, 2012), although only a third of researchers surveyed used ResearchGate at three other Indian universities (Mahajan, Singh, & Kumar, 2013). An international online survey of academics in January 2014 with a bias towards English-speaking younger online social researchers found that more used Academica.edu (48%) than used ResearchGate (32%) for academic purposes, with LinkedIn (60%), Facebook (41%) and Twitter (90%) also used more than ResearchGate, although only 13% of academics found ResearchGate to be useful for their work (Lupon, 2014). Three quarters of 315 Arab survey respondents used at least one academic social web site, with ResearchGate being the most popular (Elsayed, in press). An analysis of statistics reported by ResearchGate for universities in Bolivia, Ecuador, Columbia and Peru found them all to have at least one member, with one having over 4,000 (Campos Freire, Rivera Rogel, Rodríguez, 2014). A study of members of a large Spanish research agency found that many more had ResearchGate profiles than had Academia.edu profiles, with about a third of the latter also having a ResearchGate profile (Ortega, 2015). An international survey of science and engineering scholars found that out of more than 3,579 responses from 95 different countries (a response rate of 3.2%), ResearchGate was the most popular academic social web site, although Google Scholar had more users (Van Noorden, 2014). Just under half of the science and engineering researchers declared that they were aware of ResearchGate and used it regularly whereas only 35% of the arts and humanities and social sciences respondents stated that they frequently used ResearchGate for scholarly communication. Nevertheless, little is known about why scholars sign up to ResearchGate. A survey of 100 academics in one Indian institution found that the common motivations for using ResearchGate are finding out about others' research, current awareness, and study groups (Chakraborty, 2012). A study of ResearchGate statistics from 55 academics of a Swiss university found that platform engagement, seniority, and publication impact contribute to members' centrality within the site and the membership duration has a significant impact on the number of followers in the site (Hoffmann, Lutz, & Meckel, 2015).

Academia.edu has similar goals to ResearchGate but with less emphasis on collaborative uses of the site. Like ResearchGate, popularity in the site probably reflects traditional academic hierarchies to some extent, although there is some evidence that women may be more popular in the site than offline in some fields (Thelwall & Kousha, 2014). Moreover, senior researchers rarely joined the site in the early years (Mas-Bleda, Thelwall, Kousha, & Aguillo, 2014), suggesting that it may not have adequate representation from the most important researchers.

Mendeley is a social reference manager and it has a different approach because its focus is on helping authors manage their references but it also has a social component and so it has a substantial overlap in functionality and targets a similar audience (Zaugg, West, Tateishi, & Randall, 2011). Users are typically junior scholars, however, such as PhD students, postgraduates and postdocs (Haustein & Larivière, 2014; Mohammadi, Thelwall, Haustein, & Larivière, 2015). Mendeley is also useful for scientometric purposes (Gunn, 2013) because its Applications Programming Interface (API) has allowed many studies to analyse the reader counts on large sets of articles, showing that they have properties similar to those of citation counts (Haustein, Larivière, Thelwall, Amyot, & Peters, 2014;. Li, & Thelwall, 2012; Mohammadi & Thelwall, 2014). Mendeley reports the number of "readers" for each article, but this refers to the number of Mendeley users that have saved the article

within their profile, whether or not they have read it. Since most Mendeley users probably have read their saved articles or intend to read them in the future (Mohammadi, Thelwall, & Kousha, 2015) it seems reasonable to refer to these users as the "Mendeley readers" of an article.

Research Questions

The first objective of this paper is to investigate the coverage of ResearchGate in terms of the ages and disciplines of the articles uploaded to the site. This is useful basic information for any analysis of publications in the site. The second objective is to assess how the age of an article affects its usefulness in the site, in terms of the number of views that it attracts. If, for example, older articles were rarely viewed then it would not be worth the effort of researchers to upload them. The third objective is to assess the viewership data in ResearchGate. This information is not quality controlled and so it is not clear that it is meaningful. Nevertheless, if useful, then it might be valuable as a readership indicator to supplement the current raft of alternative indicators.

- What is the disciplinary range and age range of articles posted to ResearchGate?
- Does the number of views that an article attracts in ResearchGate depend upon the age and subject area of the article?
- Do Scopus citations and Mendeley readers for articles correlate positively with their ResearchGate views?

Methods

ResearchGate does not provide an exhaustive list of publications uploaded to the site and so an ad-hoc method was used to generate samples for analysis. This used a large number of automatic Bing searches to identify publications that were uploaded to ResearchGate on specific dates. Each publication page contains an upload date and since these pages are indexed by commercial search engines they can be found by appropriate targeted Bing queries. The syntax below was used to retrieve lists of publications from specific dates from Bing.

```
site:researchgate.net "uploaded on" [Date]
```

Here, [Date] is the date searched for. Since some queries had more hits than the maximum returned by Bing for a single query, 1000, query splitting (Thelwall, 2008) was used to retrieve additional results. This method constructs and submits derivative queries that refine the original query and then merges the results of all derivative queries with the original query in order to give a larger overall set of results. The technique is not exhaustive and so the results are likely to be underestimates for days having more than 1000 uploads. The results are also likely to be underestimated for all queries because search engines do not comprehensively index the web (Lewandowski, Wahlig, & Meyer-Bautor, 2006; Lewandowski, Wahlig, & Meyer-Bautor, 2006; Vaughan & Thelwall, 2004). As of July 2015, upload dates are no longer presented in ResearchGate pages in this way but are reported as "Available from" dates and so the method described above no longer works.

ResearchGate publications were selected from three different recent months. Earlier periods were not analysed because ResearchGate reported the current date on these publication pages rather than the upload date of the publication and so it was not possible to identify accurate upload dates for them. This is presumably due to a technology change within ResearchGate before 2014. Three evenly spaced recent months were chosen for

comparison: January 2014, July 2014 and January 2015. Queries were submitted to Bing with its API using Webometric Analyst on June 8, 2015 for the last 28 days in both Januarys and July to retrieve papers uploaded on those dates. This produced a total of 122,424 ResearchGate publication URLs. A 28 day period was used to ensure that weekdays were included equally with weekends in all years in order to ensure broad comparability of the data, although Julys may be used differently from Januarys due to the academic calendar.

The pages identified by the Bing searches were downloaded by SocSciBot (http://socscibot.wlv.ac.uk) between June 10 and June 24, 2015 and a program was written to extract the key information from them (added to the Webometric Analyst Services Menu). The initial research plan included analyses of follower counts but since most of these were zero (74%), the numbers were low (the most followers for an article was 16), and during the analysis (July 2015) ResearchGate stopped reporting them. Hence, they would not be useful or give interesting results and were excluded.

Out of the downloaded pages, 68,731 (56%) contained a full-text publication and were uploaded during the three selected months. Most of these, 39,406 (57%), contained a DOI for the publication. For articles with DOIs in ResearchGate, the reader counts of the articles in Mendeley and their Scopus-indexed citation counts were extracted using DOI searches. Articles without DOIs were ignored because there seemed to be frequent cases of incorrect and abbreviated journal names in ResearchGate and this could lead to false matches if the articles were searched for in Scopus or Mendeley. When zero results were returned from Mendeley, the article may still have Mendeley readers (Zahedi, Fenner, & Costas, 2014) for a version without DOI information but it is more likely that it does not have readers and so these articles were treated as unread. In contrast, within Scopus an article without a DOI match was assumed to be missing from the database and was ignored as missing data. Documents that were not registered in Scopus were ignored, as were articles categorised by Scopus as anything other than a journal article (e.g., books, conference papers, reviews, editorials).

The three data sources were correlated against each other separately for each subject and year because merging either can substantially reduce the strength of correlations (Thelwall & Fairclough, 2015b). Spearman correlations were used because citation (de Solla Price, 1976) and alternative web data (Thelwall & Wilson, 2015) is typically skewed. Similarly, for measures of average, either median or geometric means are recommended and were used (Thelwall & Fairclough, 2015a; Thelwall & Wilson, 2014). Geometric means were calculated by adding one to the data, taking the natural log, then taking the arithmetic mean of the result. This mean was taken to the power *e* and 1 subtracted (Thelwall & Fairclough, 2015a). The offset of 1 is required to avoid problems caused by 0s in the data. Geometric mean confidence intervals were calculated using the standard normal distribution confidence intervals on the data after the log transformation with the limits being transformed using the same exponential (inverse) transformation.

For the time trends, 1981 was selected as the first year because one of the three data sets for the time analysis had less than 3 articles in 1980. Journal names were cross-referenced with Scopus field classification codes in order to give each article one of the 36 broad Scopus codes. Although these are imperfect and other methods using references, citations or keywords might give more accurate results for individual articles (Glänzel, Schubert, & Czerwon, 1999), the Scopus categories have the advantage of being transparent and reproducible. For simplicity, only the first subject classification was given to journals with multiple Scopus classifications.

Results and Discussion

A minority of publications in ResearchGate could be matched with journal articles having a DOI in Scopus (Table 1), with a smaller proportion for more recently uploaded articles. Since the coverage of Scopus does not seem to have decreased substantially, the most likely cause of the decrease is that authors are uploading a wider variety of documents to ResearchGate, perhaps including preprints and reports, so that the refereed journal articles form an increasingly small proportion of the documents uploaded. An alternative explanation is that publishers might be taking action against uploading to ResearchGate that breaches copyright, leading authors to be more cautious with journal article full text uploading.

Table 1. Sample statistics from the valid downloaded ResearchGate pages.

Upload date	Publications found in ResearchGate	Journal articles with DOI and in Scopus	Journal articles with DOI and in Scopus with category	ResearchGate views: Minimum; Median; Maximum	Scopus citations: Minimum; Median; Maximum	Mendeley readers: Minimum; Median; Maximum
Jan 2014	9,791	3,669 (37%)	3,439 (35%)	2; 81; 5,644	0; 8; 824	0; 7; 533
July 2014	34,878	8,239 (24%)	7,415 (21%)	1; 56; 2,226	0; 8; 5,217	0; 4; 1,298
Jan 2015	24,062	4,344 (18%)	3,935 (16%)	1; 46; 6,863	0; 7; 3,038	0; 3; 376
Total	68,731	16,252 (24%)	14,789 (22%)	-	-	-

Age and subject area of uploaded publications

Although the most common publication year for uploaded documents is the year of uploading (except for January 2014; Figure 1), in total, more documents from previous years are uploaded than from the current year. Hence, ResearchGate is commonly being used to archive older research as well as current research. No explanation could be found for the bump at 2001 in July 2014. It is possible that there was an attempt to systematically archive older articles in ResearchGate that was preceding in chronological order and had reached 2001 in July 2014.

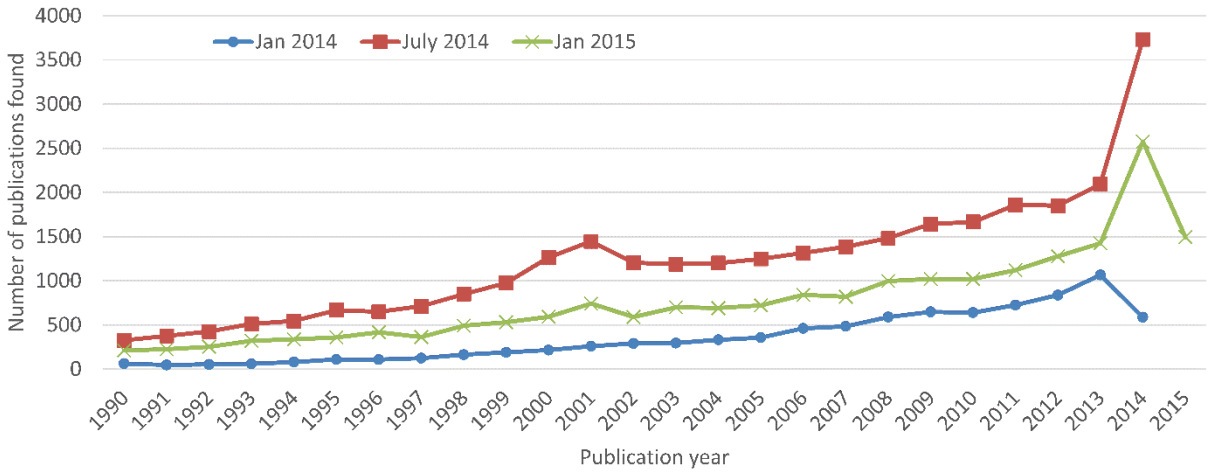

Figure 1. Publication years of articles uploaded to ResearchGate during three selected months (n=68,731).

Based upon the Scopus category of the journals of ResearchGate articles with a DOI and present in Scopus, there are substantial numbers of articles from medicine, physical sciences, life sciences, and engineering articles but the social sciences (other than psychology, business and economics to some extent), arts and humanities are not well represented (Figure 2). These results are affected by the coverage of Scopus, however, and the restriction to journal articles. The Scopus categories are also not equally large and so the results are somewhat misleading. Comparing the number of Scopus articles in ResearchGate with the number of Scopus articles overall for the largest year (2014) and for all years (Figure 3), ResearchGate's coverage of Scopus-indexed journal articles is uneven, with Arts and Humanities being particularly poorly represented. Arts and humanities research may be more represented in ResearchGate than the figures suggest, but with many reports, preprints, book information and book chapters as well as non-English journal articles that might not be Scopus-indexed.

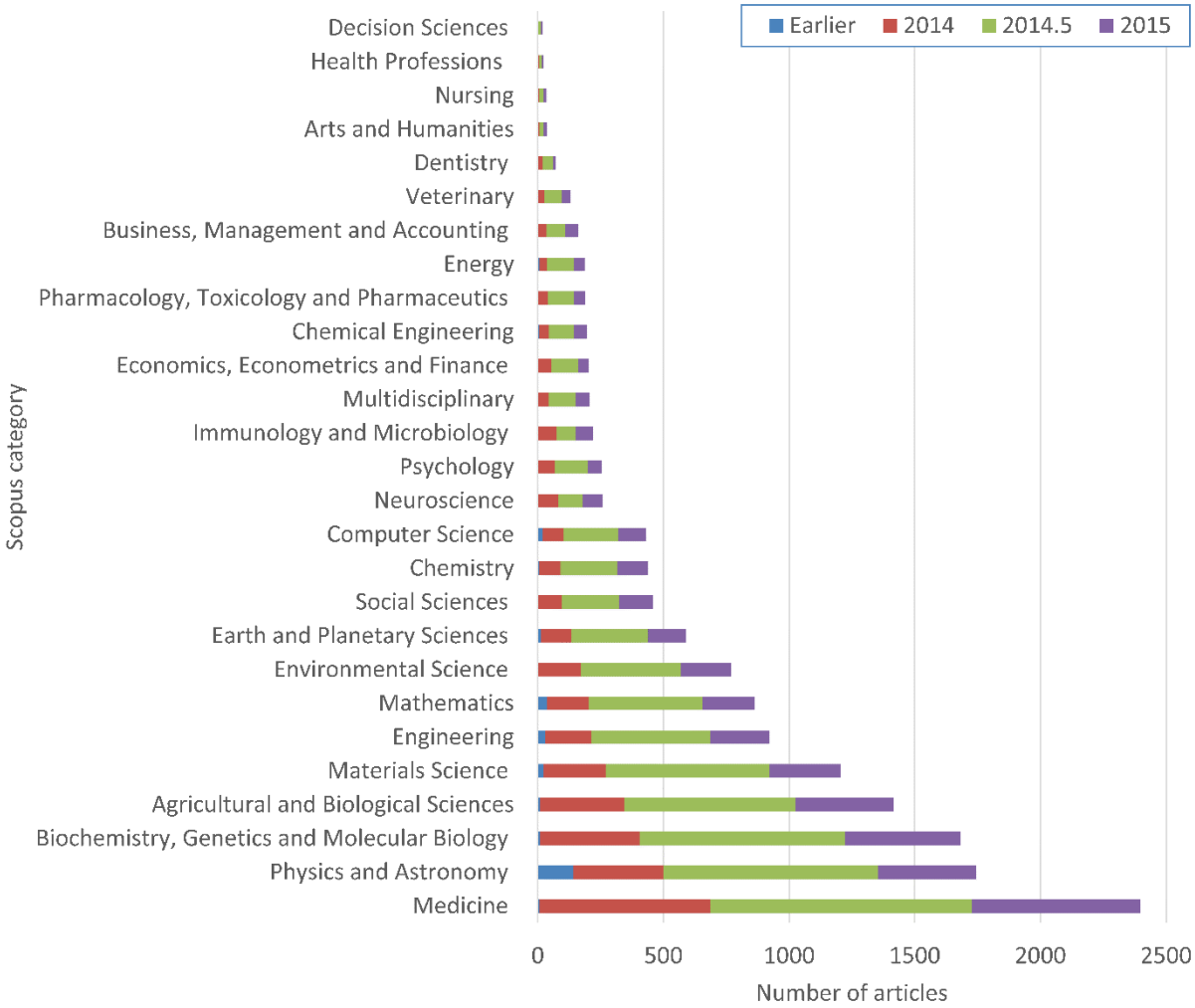

Figure 2. Scopus categories of journal articles with DOIs uploaded to ResearchGate during one of the selected months (n=14,789).

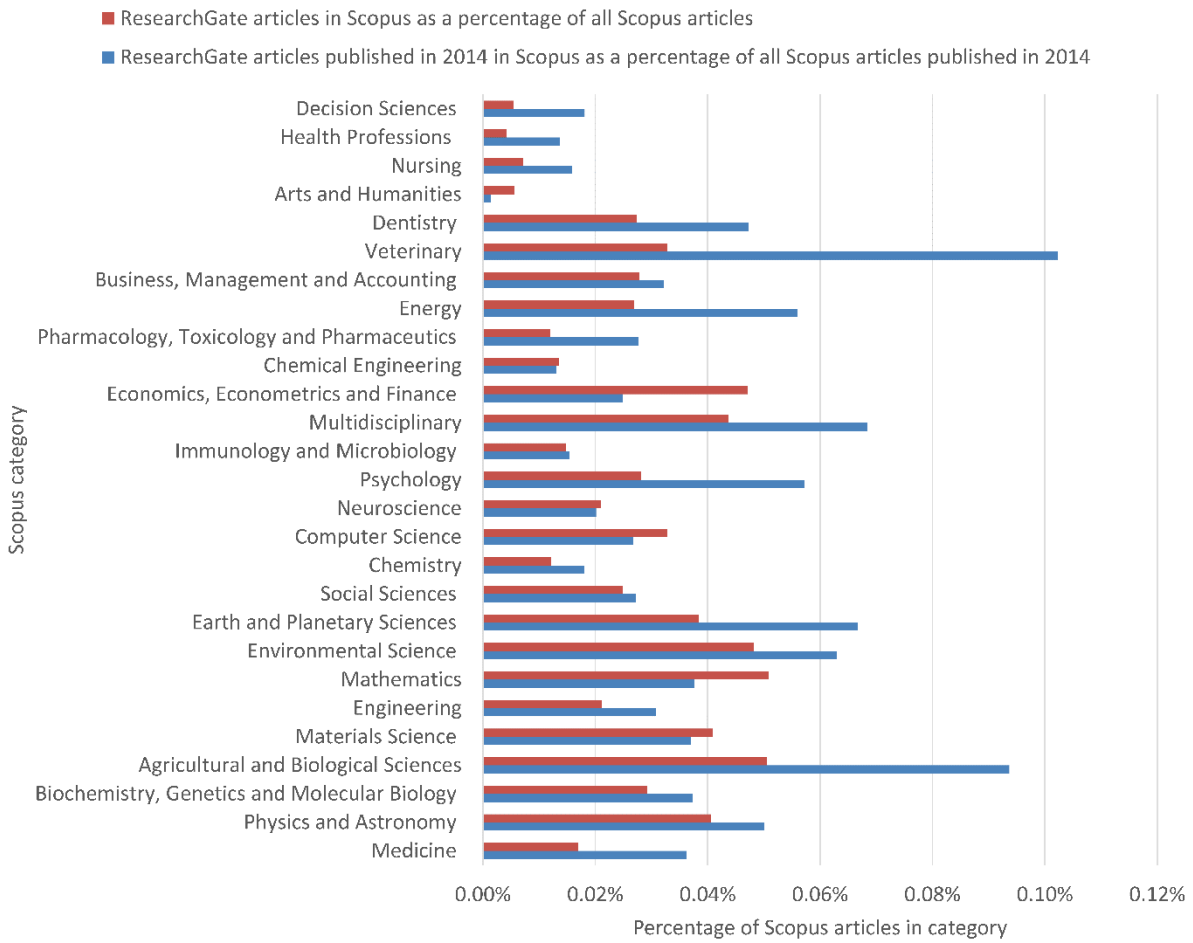

Figure 3. The number of journal articles with DOIs uploaded to ResearchGate during one of the selected months as a proportion of all Scopus articles in the category.

View counts and article ages

More recent articles tended to be more viewed on ResearchGate than were older articles uploaded during the same month (Figures 4-6). Older articles still attracted a substantial amount of views, however. For example, articles from the 1980s attracted views at about a third to a half of the rate for articles from the most recent decade. Whilst the higher level of interest in more recent articles is unsurprising, the trend depends partly on the selection of the older articles. If researchers upload all of their recent articles but only their best older articles, for example, then the view counts would not be comparable.

The trend for citations is for older articles to be more cited than are newer articles, which is to be expected since citations accrue over time. The increase for older articles is only slight and suggests that they have not been uploaded more selectively than newer articles. The readership trend confirms this because the older articles attract, on average, fewer readers than the newer articles, which reflects a natural preference for reading more current studies.

This trend reverses in about 2001 because articles published since 2001 tend to be viewed less often if they are more recent. The reason for the decreasing interest in the newest of the most recent articles may be that the average quality of the most recent articles has decreased. This would occur if a substantial fraction of authors attempted to be exhaustive in their uploading of articles, starting with their most recent ones. This might be

motivated, for example, by the prominent display of the RG Score alongside all members, with this score being affected by the number of publications uploaded. Another possibility is that the membership of ResearchGate is widening from an initial base of researchers that tended to produce research with higher academic impact. This could also include spreading from Europe and the USA to a wider international base. Another possibility is that the membership of ResearchGate has spread across more disciplines to include those that produce work that attracts a more niche audience.

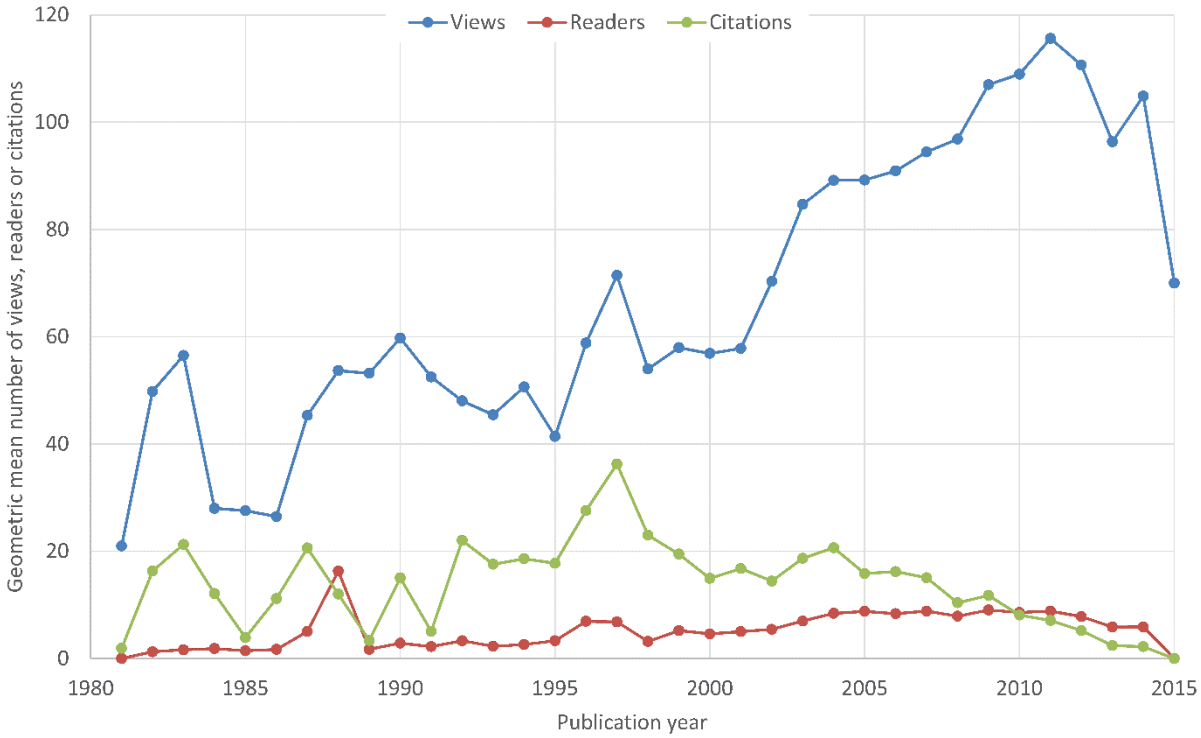

Figure 4. Average (geometric mean) number of views, readers and citations per article by publication year for journal articles with DOIs and registered in Scopus, within the set of articles uploaded to ResearchGate in January 2014 (n=3,669).

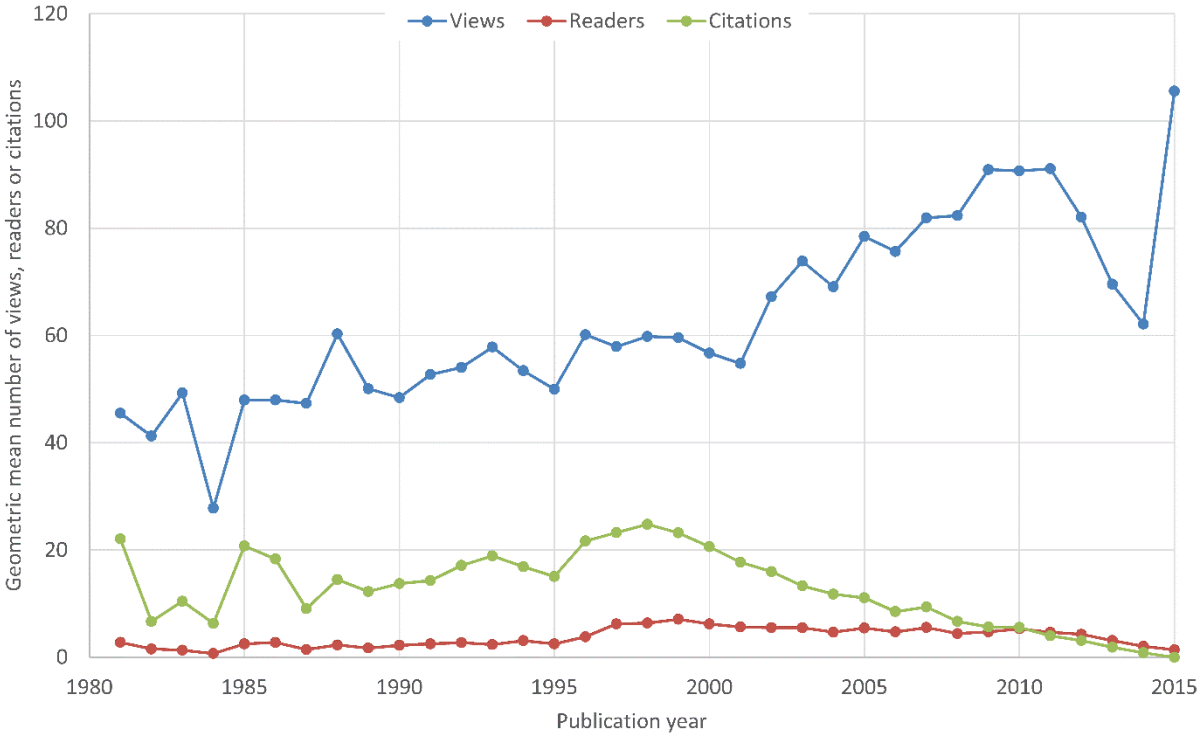

Figure 5. Average (geometric mean) number of views, readers and citations per article by publication year for journal articles with DOIs and registered in Scopus, within the set of articles uploaded to ResearchGate in July 2014 (the 2015 data is an outlier based on only 3 publications) (n=8,239).

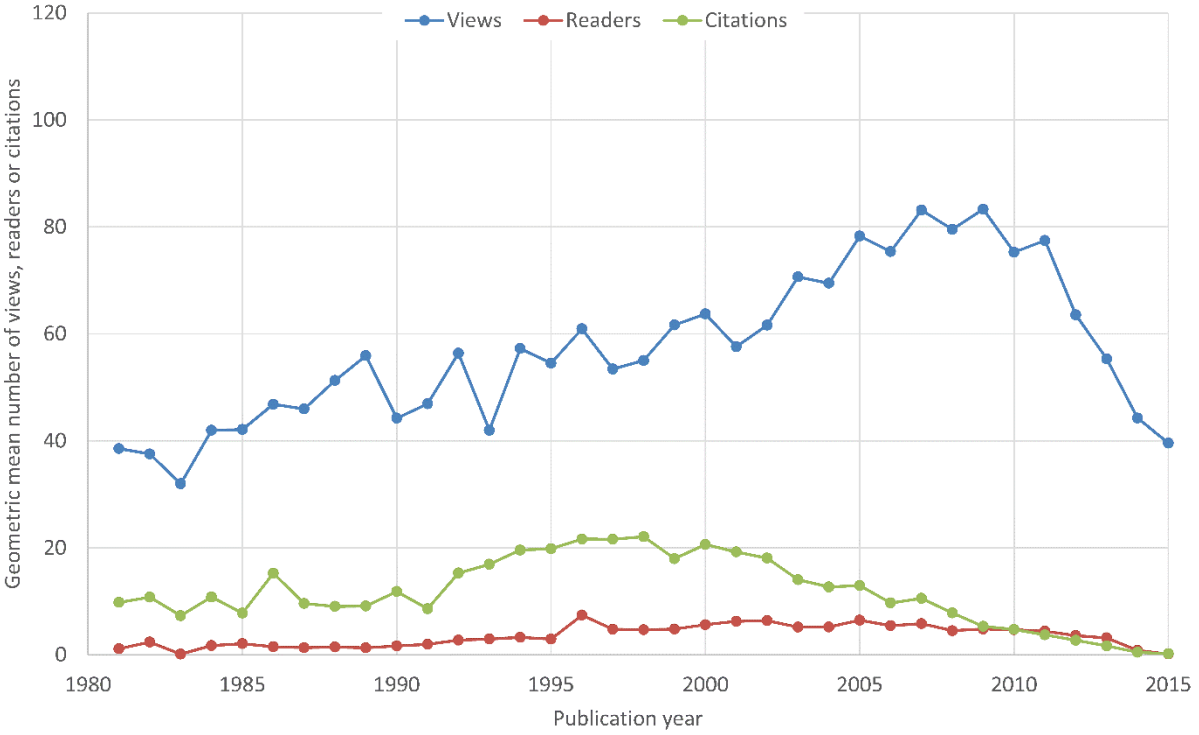

Figure 6. Average (geometric mean) number of views, readers and citations per article by publication year for journal articles with DOIs and registered in Scopus, within the set of articles uploaded to ResearchGate in January 2015 (n=4,344).

View counts and article subject areas

For the remainder of the article time factors are of minor importance and so results are given only for the middle period (articles uploaded in July 2014). The average number of views per article varies substantially and significantly between subject areas (Figure 7). Ignoring the areas with wide confidence intervals, the variation is from 64 views per article (Social Sciences) to almost double at 126 views per article (Dentistry). In general, the health and life sciences articles were the most viewed.

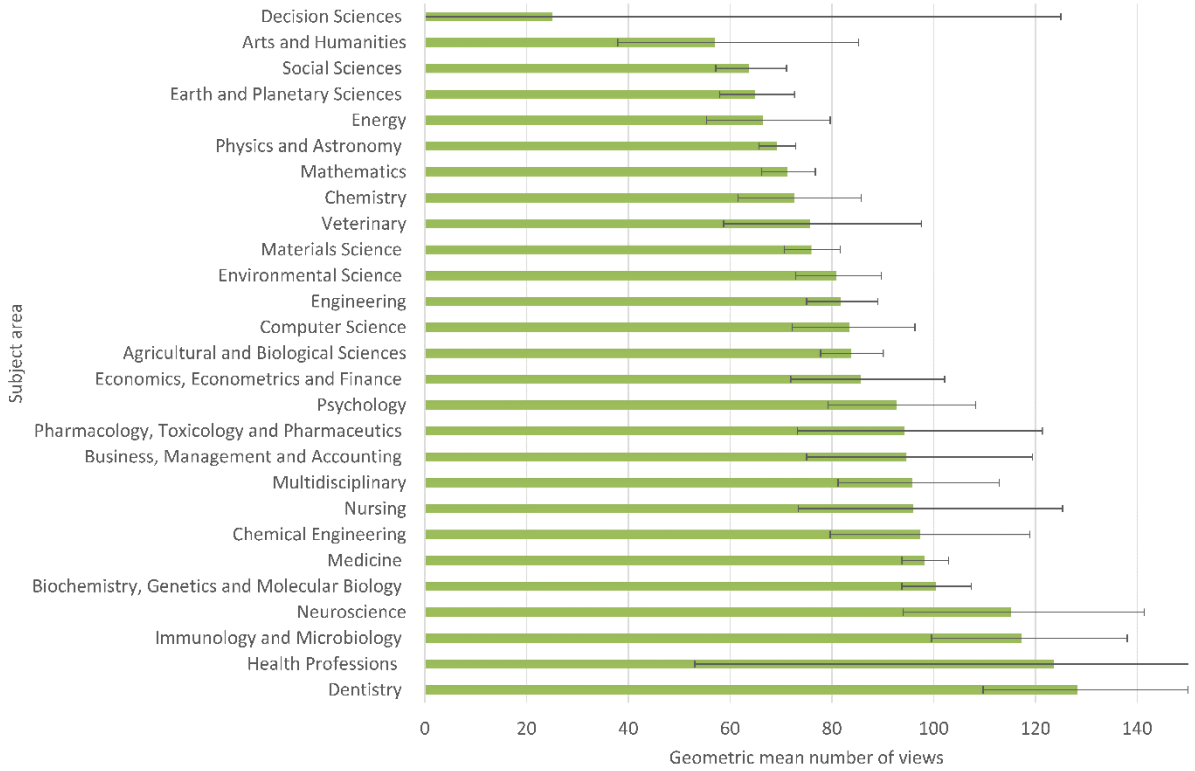

Figure 7. Average (geometric mean) number of views per article by subject area for journal articles with DOIs and registered in Scopus, within the set of articles uploaded to ResearchGate in July 2014 (n=3,439). The error bars show 95% confidence intervals.

Comparing the disciplinary viewing results (Figure 7) with an equivalent graph for Mendeley readers (Figure 8), there are disciplinary differences in the extent to which these two reflect article readership. For example, Mathematics is similar to Physics and Astronomy in the average number of views per article, but has many fewer Mendeley readers per article. The most likely cause of this is disciplinary differences in the uptake of the two sites.

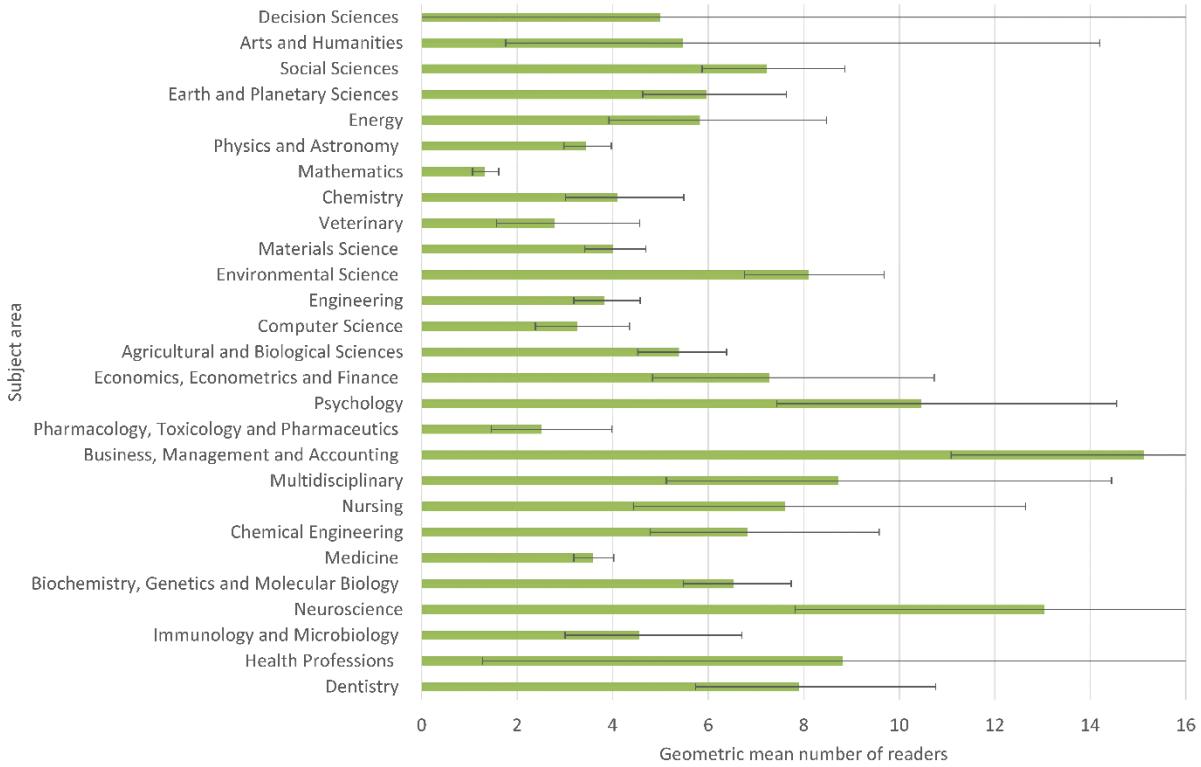

Figure 8. Average (geometric mean) number of Mendeley readers per article by subject area for journal articles with DOIs and registered in Scopus, within the set of articles uploaded to ResearchGate in July 2014 (n=7,415). The error bars show 95% confidence intervals. Subject areas are in the same order as for Figure 7.

There are also substantial disciplinary differences between ResearchGate views (Figure 7) and Scopus citations (Figure 9). For example, Medicine attracts more views per article than does Engineering but fewer Scopus citations per article. It is perhaps surprising that the difference between Figure 9 and Figure 7 is not more marked than the difference between Figure 8 and Figure 7. A possible explanation is that both ResearchGate and Scopus have reasonably broad coverage of all areas of scholarship and so the results are not affected by disciplinary differences in uptake. The differences between disciplines are unexpected in Figure 9, however. For example, medicine is a higher citation subject area than engineering (Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011) but the reverse is true in the graph. This might be due to higher profile medical journals taking steps to prevent authors posting copies of their published articles to ResearchGate (as did Elsevier, which publishes The Lancet: Clarke, 2013), and so the medical sample might not be representative of the field as a whole.

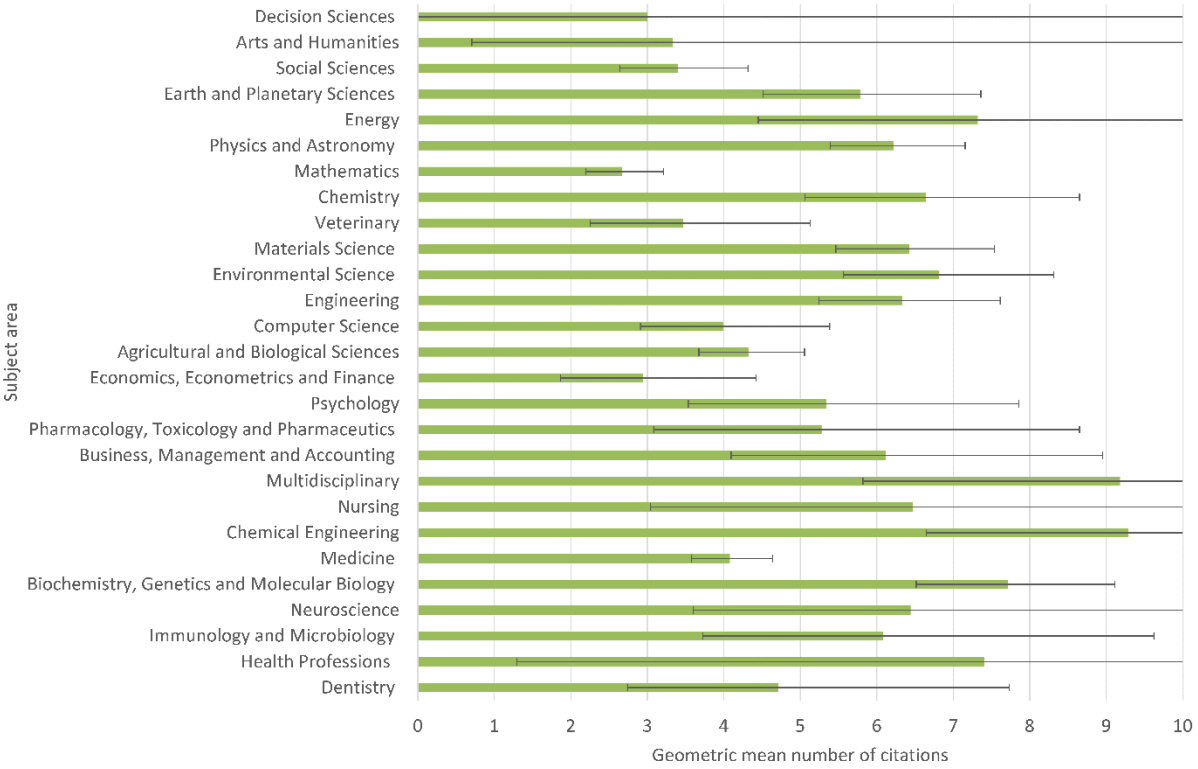

Figure 9. Average (geometric mean) number of Scopus citations per article by subject area for journal articles with DOIs and registered in Scopus, within the set of articles uploaded to ResearchGate in July 2014 (n=3,935). The error bars show 95% confidence intervals. Subject areas are in the same order as for Figure 7.

View counts, Scopus citations and Mendeley readers

For an effective check of the correlation between different sets of articles, the articles should have been published in the same subject area and year. This is because citations tend to increase over time and average numbers of citations per paper vary substantially between fields (Moed, 2006). Because of these differences, mixing articles from different fields will tend to reduce correlation strengths (Thelwall & Fairclough, 2015b) and mixing articles from different years may increase or decrease correlation strengths depending on whether the readership or viewer counts increase or decrease in line with citations. To avoid these problems, correlations were calculated for articles with a common subject area and year. Most such sets of homogenous articles were too small to analyse and so only the largest sets were analysed. These largest sets were from articles uploaded in July 2014, with a publication date of 2014 and in one of four fields (Table 2). As a backup test, correlation tests were also performed for sets of old articles for the same subject areas (Table 3). These old articles could reasonably assumed to have accumulated a large majority of their eventual citations since they had been published before 2005 (Larivière, Archambault, & Gingras, 2008), although this is not true for all articles (van Raan, 2004).

The correlations between numbers of ResearchGate views and Scopus citations tend to be positive, low and statistically significant (Tables 2, 3). The same is true for the correlations between ResearchGate views and Mendeley readers. This is perhaps surprising since viewing is more closely related to reading than to citing. In general, the correlations between Mendeley readers and Scopus citations are much higher than the other two correlations, giving figures in line with previous comparisons of Scopus and Mendeley (e.g.,

Haustein, Larivière, Thelwall, Amyot, & Peters, 2014; Mohammadi & Thelwall, 2014). A possible explanation for these differences is that Mendeley readership is more close to Scopus citation than to ResearchGate viewing because Mendeley is a reference manager and therefore might be used by people that are likely to cite an article, albeit not necessarily in an academic journal article. An alternative explanation is that ResearchGate view counts are subject to more noise in the form of spam and casual browsing than are Mendeley reader counts.

Table 2. Spearman correlations for articles uploaded to ResearchGate in July 2014 and with a 2014 publication date.

Field	Views-Cites	Views-Readers	Cites-Readers	Sample size
Medicine	0.16	0.24**	0.20*	116
Physics and Astronomy	0.35**	0.12	0.18	56
Biochemistry	0.38**	0.40**	0.62**	59
Agricultural	0.30**	0.33**	0.29**	103
Materials Science	0.06	0.23	0.41**	45

* Significant with $p<0.05$; ** significant with $p<0.01$.

Table 3. Spearman correlations for articles uploaded to ResearchGate in July 2014 and with a publication date of 2004 or before.

Field	Views-Cites	Views-Readers	Cites-Readers	Sample size
Medicine	0.28**	0.29**	0.61**	367
Physics and Astronomy	0.32**	0.31**	0.58**	381
Biochemistry	0.45**	0.40**	0.58**	515
Agricultural	0.28**	0.25**	0.55**	275
Materials Science	0.45**	0.33**	0.49**	205

* Significant with $p<0.05$; ** significant with $p<0.01$.

The relatively low correlations between Mendeley reader counts and ResearchGate views for homogenous sets of articles are surprising. To follow up on this, the Table 2 data for medicine was investigated again. The main cause of the low correlation was a number of articles with high view counts but few or zero readers or citations. For example, “An Acute care surgery dilemma: Emergent laparoscopic cholecystectomy in patients on aspirin therapy” in the American Journal of Surgery had no Mendeley readers or Scopus citations but 438 views. It is possible that this article was read by surgical practitioners because of its focus on an issue that would face any surgeon intending to remove the gall bladder of a patient on aspirin therapy. Hence, it is plausible that this article would attract many readers but no citations and also no Mendeley readers because, as a reference manager, practitioners may not use Mendeley. Related to this, the article may also have been added to the course reading lists for doctors being trained in surgery, although a Google search for the article found no evidence of this. The second most viewed article, “Domain-specific physical activity and health-related quality of life in university students” in the European Journal of Sport Science had 258 views, 3 Mendeley readers and 1 Scopus citation. This article may have been relatively frequently viewed because its subjects (students) are amongst the users of ResearchGate, or its message (keeping fit promotes happiness for students) is simple but useful. Thus, this article’s readership could also reflect its relatively high applied value. The same is true for the third and fourth most viewed articles,

"International Evidence-Based Recommendations for Focused Cardiac Ultrasound", and "Novel Characterization of Gait Impairments in People with Multiple Sclerosis by Means of the Gait Profile Score". Overall, then, it is plausible that ResearchGate view counts tend to reflect the readership of an article, albeit only those readers who visited the site to access the article. These might include ResearchGate members as well as casual visitors after a Google search. Presumably, also, a proportion of viewers do not read an article or read just the title and abstract and then discard it as irrelevant.

Limitations

An important limitation of the DOI analysis is that publications with DOIs in ResearchGate may have a different character to other publications. For example, a much greater proportion of these may have been found by ResearchGate crawlers than initially uploaded by members and a smaller fraction of older articles may have DOIs. This affects the DOI analyses but not the analyses that use all publications found. Another limitation is that the ResearchGate publications were identified through Bing queries and since search engines do not crawl the whole web, it is possible that Bing has found, or chosen to index, a biased subset of ResearchGate. If this is true, then the subset is likely to contain more popular articles than average for the site. Another selection bias is that some of the analyses used articles with DOIs that are indexed in Scopus, and this probably gives a bias towards articles in English language higher impact journals.

A limitation of the view count statistic is that its absolute size may be misleading if some of the numbers were derived from robots crawling the site or from authors or administrators checking pages for viewing statistics. The statistically significant positive correlations found between the viewing, citing and reading data are safeguards that at least some of the viewing data is genuine but this does not protect against artificial inflation of the data (e.g., Delgado López-Cózar, Robinson-García, Torres-Salinas, 2014). This is most problematic when comparing the magnitude of the view counts between years and between subjects. For example if the view count for one set is double the view count of another set but the same number of viewers in both sets are artificial then the underlying difference will be more than double.

Conclusions

ResearchGate seems to have a wide coverage of articles from different disciplines and years, although its coverage of recent years is more substantial than its coverage of older years and some disciplines, such as the arts and humanities and some areas of the social sciences, are poorly covered. Despite the older articles in ResearchGate apparently not having been selected for higher research impact, they were viewed at least a third as often, on average, as recent articles and so it makes sense for scholars to upload older articles to the site in order to attract and audience, although newer articles should be the priority. Similarly, articles in all disciplines in ResearchGate attract at least half as many views as articles from the most viewed discipline and so there is no reason for authors within any particular discipline to avoid ResearchGate in the belief that their articles would attract a negligible numbers of views. These conclusions rely on the unproven assumption that the ResearchGate viewing figures are not artificially inflated.

ResearchGate article views have low to moderate correlations with both Scopus citations and Mendeley readers, indicating that they are related to scholarly impact but that the relationship is not close. A small scale post-hoc investigation of one subject found that it

is plausible that ResearchGate views of articles could reflect their wider readership better than can Mendeley reader counts, which would make them the best current publically available readership indicator. The obvious source of readership information, download counts, can be private or calculated differently by publishers that are not COUNTER compliant (e.g., Botero, Carrico, & Tennant, 2011). Nevertheless, ResearchGate view counts would only be available for articles that had been uploaded to ResearchGate and, even for these articles, would be affected by the availability of alternative full text article sources (e.g., in the journal website or on the author's home page). These limitations, and the potential for gaming the figures, would rule out ResearchGate data from formal evaluations but it can still be useful for self-evaluations (Wouters & Costas, 2012) and investigations of patterns of readership within academia (e.g., Mohammadi & Thelwall, 2014). Hence, ResearchGate views are a promising potential source of readership evidence for academic articles, but a full-scale study is needed to assess this claim robustly, including an assessment of outliers (Thelwall, 2015) and other issues (Sud & Thelwall, 2014).

References

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. Scientometrics, 88(2), 385-397.

Alheyasat, O. (2015). Examination expertise sharing in academic social networks using graphs: The case of ResearchGate. Contemporary Engineering Sciences, 8(1-4), 137-151.

Arènes, C. (2015). Les réseaux sociaux numériques pour chercheurs: Quelles pratiques, quels enjeux? Ethics, Medicine and Public Health, 1(2), 239-247.

Bar-Ilan, J. (2004). The use of Web search engines in information science research. Annual Review of Information Science and Technology, 38(1), 231-288.

Botero, C., Carrico, S., & Tennant, M. R. (2011). Using comparative online journal usage studies to assess the Big Deal. Library Resources & Technical Services, 52(2), 61-68.

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.

Campos Freire, F., Rivera Rogel, D., Rodríguez, C. (2014). La presencia e impacto de las universidades de los países andinos en las redes sociales digitales. Revista Latina de Comunicacion Social, 69, 571-592. DOI: 10.4185/RLCS-2014-1025

Campos Freire, F., & Valencia, A. (2015). Managing academic profiles on scientific social networks. In New Contributions in Information Systems and Technologies (pp. 265-273). Berlin: Springer.

Chakraborty, N. (2012). Activities and reasons for using social networking sites by research scholars in NEHU: A study on Facebook and ResearchGate. http://ir.inflibnet.ac.in/handle/1944/1666

Clarke, M. (2013). The end of an era for Academia.edu and other academic networks? The Scholarly Kitchen. http://scholarlykitchen.sspnet.org/2013/12/11/has-elsevier-signaled-a-new-era-for-academia-edu-and-other-professional-networks/

Corvello, V., Genovese, A., & Verteramo, S. (2014). Knowledge sharing among users of scientific social networking platforms. Frontiers in Artificial Intelligence and Applications, 261, 369-380.

Davis, P.M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. The FASEB Journal, 25(7), 2129-2134.

Delgado López-Cózar, E., Robinson-García, N., Torres-Salinas, D. (2014). The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators. Journal of the Association for Information Science and Technology, 65(3), 446-454. doi:10.1002/asi.23056

de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science, 293.

Duy, J., & Vaughan, L. (2006). Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination. Journal of Academic Librarianship, 32(5), 512-517.

Elsayed, A.M. (in press). The use of academic social networks among Arab researchers: A survey. Social Science Computer Review, DOI: 10.1177/0894439315589146

Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. Scientometrics, 44(3), 427-439.

Goodwin, S., Jeng, W., & He, D. (2014). Changing communication on ResearchGate through interface updates. Proceedings of the American Society for Information Science and Technology, 51(1), 1-4.

Gunn, W. (2013). Social signals reflect academic impact: What it means when a scholar adds a paper to Mendeley. Information Standards Quarterly, 25(2), 33-39.

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? IT-Information Technology, 56(5), 207-215.

Haustein, S. & Larivière, V. (2014). Mendeley as a source of readership by students and postdocs? Evaluating article usage by academic status. In: Proceeding of the IATUL Conference. Paper 2. http://docs.lib.purdue.edu/iatul/2014/altmetrics/2/

Hoffmann, C. P., Lutz, C. & Meckel, M. (2015). A relational altmetric? Network centrality on ResearchGate as an indicator of scientific impact. Journal of the Association for Information Science and Technology. doi: 10.1002/asi.23423

Jordan, K. (2015). Exploring the ResearchGate score as an academic metric: Reflections and implications for practice. In: Quantifying and Analysing Scholarly Communication on the Web (ASCW15), 30 June 2015, Oxford. http://oro.open.ac.uk/43538/1/ASCW15_jordan_response_kraker-lex.pdf

Kadriu, A. (2013). Discovering value in academic social networks: A case study in ResearchGate. In Proceedings of the 35th International Conference on Information Technology Interfaces (ITI2013) (pp. 57-62). Los Alamitos: IEEE Press.

Kamath, V. V., Setlur, K., & Yerlagudda, K. (2015). Oral lichenoid lesions-A review and update. Indian Journal of Dermatology, 60(1), 102. DOI: 10.4103/0019-5154.147830

Kousha, K. & Thelwall, M. (2014). Disseminating research with web CV hyperlinks. Journal of the Association for Information Science and Technology, 65(8), 1615–1626.

Kraker, P. & Lex, E. (2015). A critical look at the ResearchGate score as a measure of scientific reputation. In Proceedings of the Quantifying and Analysing Scholarly Communication on the Web workshop (ASCW'15), Web Science conference 2015 (Oxford, UK, June 28 – July 1, 2015).

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S.S., Martimbeau, N., & Elwell, B. (2005). The bibliometric properties of article readership information. Journal of the American Society for Information Science and Technology, 56(2), 111-128.

Łaczmański, Ł., Jakubik, M., Bednarek-Tupikowska, G., Rymaszewska, J., Słoka, N., & Lwow, F. (2015). Vitamin D receptor gene polymorphisms in Alzheimer's disease patients. Experimental Gerontology, 69, 142-147.

Larivière, V., Archambault, É., & Gingras, Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). Journal of the American Society for Information Science and Technology, 59(2), 288-296.

Lewandowski, D., Wahlig, H., & Meyer-Bautor, G. (2006). The freshness of web search engine databases. Journal of Information Science, 32, 131-148.

Li, L., He, D., Jeng, W., Goodwin, S., & Zhang, C. (2015). Answer quality characteristics and prediction on an academic Q&A Site: A case study on ResearchGate. In Proceedings of the 24th International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland: IW3C2 (pp. 1453-1458).

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. In Proceedings of the 17th International Conference on Science and Technology Indicators (Vol. 2, pp. 451-551).

Lupton, D. (2014). 'Feeling better connected': Academics' use of social media. Canberra: News & Media Research Centre, University of Canberra.
http://www.canberra.edu.au/about-uc/faculties/arts-design/attachments2/pdf/n-and-mrc/Feeling-Better-Connected-report-final.pdf

Madhusudhan, M. (2012). Use of social networking sites by research scholars of the University of Delhi: A study. International Information and Library Review, 44(2), 100-113.

Mahajan, P., Singh, H., & Kumar, A. (2013). Use of SNSs by the researchers in India: A comparative study of Panjab University and Kurukshetra University. Library Review, 62(8/9), 525-546.

Mas-Bleda, A., Thelwall, M., Kousha, K., & Aguillo, I. F. (2014). Do highly cited researchers successfully use the social web? Scientometrics, 101(1), 337-356.

McKellar, C., & Currie, G. M. (2015). Publication productivity in nuclear medicine. Journal of Nuclear Medicine Technology, 43(2), 122-128.

Moed, H. F. (2006). Citation analysis in research evaluation. Berlin: Springer.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. Journal of the Association for Information Science and Technology, 66(9), 1832-1846. DOI: 10.1002/asi.23286

Mohammadi, E., Thelwall, M. & Kousha, K. (2015). Can Mendeley bookmarks reflect readership? A survey of user motivations. Journal of the Association for Information Science and Technology. DOI: 10.1002/asi.23477

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, 65(8), 1627-1638.

Onyancha, O. B. (2015). Social media and research: An assessment of the coverage of South African universities in ResearchGate, Web of Science and the Webometrics Ranking of World Universities. South African Journal of Libraries and Information Science, 81(1), 8-20.

Ortega, J. L. (2015). Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. Journal of Informetrics, 9(1), 39-49.

Ovadia, S. (2014). ResearchGate and Academia.edu: Academic social networks. Behavioral & Social Sciences Librarian, 33(3), 165-169.

Scarlat, M.M., Mavrogenis, A.F., Pećina, M., & Niculescu, M. (2015). Impact and alternative metrics for medical publishing: Our experience with International Orthopaedics. International Orthopaedics, 39(8), 1459-1464.

Sud, P. & Thelwall, M. (2014). Evaluating altmetrics. Scientometrics, 98(2), 1131-1143. DOI: 10.1007/s11192-013-1117-2.

Swan, A., & Brown, S. (2005). Open access self-archiving: An author study. http://users.ecs.soton.ac.uk/harnad/Temp/jisc2.pdf

Thelwall, M. & Fairclough, R. (2015a). Geometric journal impact factors correcting for individual highly cited articles. Journal of Informetrics, 9(2), 263–272.

Thelwall, M. & Fairclough, R. (2015b). The influence of time and discipline on the magnitude of correlations between citation counts and quality scores. Journal of Informetrics, 9(3), 529–541. DOI: 10.1016/j.joi.2015.05.006

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

Thelwall, M., & Kousha, K. (2015). ResearchGate: Disseminating, communicating, and measuring Scholarship? Journal of the Association for Information Science and Technology, 66(5), 876-889.

Thelwall, M. & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. Journal of Informetrics, 8(4), 963–971.

Thelwall, M. & Wilson, P. (2015). Mendeley readership altmetrics for medical articles: An analysis of 45 fields, Journal of the Association for Information Science and Technology. DOI: 10.1002/asi.23501

Thelwall, M. (2008). Extracting accurate and complete results from search engines: Case study Windows Live. Journal of the American Society for Information Science and Technology, 59(1), 38-50.

Thelwall, M. (2015). Why do papers have many Mendeley readers but few Scopus-indexed citations and vice versa? Journal of Librarianship & Information Science. DOI:10.1177/0961000615594867

Van Noorden, R. (2014). Scientists and the social network. Nature, 512(7513), 126-129. http://www.nature.com/news/online-collaboration-scientists-and-the-social-network-1.15711

van Raan, A. F. (2004). Sleeping beauties in science. Scientometrics, 59(3), 467-472.

Vaughan, L., & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. Information Processing & Management, 40(4), 693-707.

Velik, R. (2015). An objective review of the technological developments for radial pulse diagnosis in Traditional Chinese Medicine. European Journal of Integrative Medicine. doi:10.1016/j.eujim.2015.06.006

Wang, X., Liu, C., Mao, W., & Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. Scientometrics, 103(2), 555-564.

Ward, J., Bejarano, W., & Dudás, A. (2015). Scholarly social media profiles and libraries: A review. Liber Quarterly, 24(4), 174-204.

Wouters, P., & Costas, R. (2012). Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century. Utrecht: SURF foundation. http://www.surffoundation.nl/nl/publicaties/Documents/Users%20narcissism%20and%20control.pdf

Zahedi, Z., Fenner, M., & Costas, R. (2014). How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIs. In altmetrics 14. Workshop at the Web Science Conference, Bloomington, USA.

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. TechTrends, 55(1), 32-36.