

For which disciplines are papers covered in F1000Prime interesting?

An analysis of discipline-specific reader data from Mendely

Robin Haunschild\* & Lutz Bornmann\*\*

\* Corresponding author:

Max Planck Institute for Solid State Research

Heisenbergstr. 1,

70569 Stuttgart, Germany.

E-mail: R.Haunschild@fkf.mpg.de

\*\* Division for Science and Innovation Studies

Administrative Headquarters of the Max Planck Society

Hofgartenstr. 8,

80539 Munich, Germany.

E-mail: bornmann@gv.mpg.de

(submitted to F1000Research on 01/19/2015)

## **Abstract**

We have used the F1000Prime paper set to inquire the number of Mendeley readers per (sub-) discipline for each of the 114,582 bio-medical papers via the Mendeley Application Programming Interface (API). Although the (sub-) discipline of Mendeley readers is self-assigned and not mandatory, we find that a large share (99.9%) of F1000Prime paper readers at Mendeley does share their (sub-) discipline. As expected we find most readers in the disciplines biology and medicine. A network analysis reveals strong connections of the disciplines engineering, chemistry, and physics to biology and medicine.

## Introduction

Interest in the broad impact of research (Bornmann, 2012, 2013) has resulted in new forms of impact measurements. Traditional forms of impact measurements using bibliometrics only allow the measurement of impact on research itself. These new forms which have been named as altmetrics (abbreviation of alternative metrics) pretend to measure the impact of research on other areas of society (than research) by counting mentions of papers in social media: “Alternative metrics, sometimes shortened to just altmetrics, is an umbrella term covering new ways of approaching, measuring and providing evidence for impact” (Adie, 2014, p. 349). As altmetrics, the number of readers (on Mendeley), micro-bloggers (on Twitter), and other consumers of research using social media are counted. Although scientometrics research on altmetrics is still in a very early phase (comparable to research on bibliometrics in the 1970s), the use of this data in research evaluation is already an issue. For example, altmetrics are considered in the Snowball Metrics project (Colledge, 2014). This project compiled a set of clearly defined indicators which will be used by participating universities (mostly Anglo-American universities) for research evaluation purposes. It seems that altmetrics will be used in practice before scientometrics research has produced standards on their reliable, fair and valid application (Weller, 2015).

This study uses one of the most important sources for altmetrics data, namely Mendeley. Mendeley “claims 3.1 million members. It was originally launched as software for managing and storing documents, but it encourages private and public social networking” (Van Noorden, 2014, p. 126). Since data from Mendeley can be received by an API without any problems and the coverage of the scientific literature has been pointed out as high (Priem, 2014), Mendeley is a very attractive data source for the reception of research. “Mendeley records the number of users that have listed it [i.e. an article], describing them as readers,

whether or not they actually read it. Presumably, listing an article in Mendeley tends to reflect that an article has been read or will be read in the future, although there is no evidence that this assumption is true” (Thelwall & Maflahi, in press).

In this study, we match Mendeley data with data from F1000Prime. F1000Prime is a database with biomedical papers and their reviews by peers. It is intended as a support tool for researchers to receive hints for the most important literature. Since it is not clear who actually reads the F1000 papers, we investigated by which disciplines these papers have been read. We are mainly interested in two questions: Are F1000 papers only read by people from biomedicine or are people from other disciplines also interested? Which disciplines read F1000 papers frequently or seldom together? The latter question will be answered by using social network techniques.

## Methods

### **Peer ratings provided by F1000Prime**

F1000 is a post-publication peer review system of the biomedical literature (papers from medical and biological journals). This service is part of the Science Navigation Group, a group of independent companies that publishes and develops information services for the professional biomedical community and the consumer market. F1000 Biology was launched in 2002 and F1000 Medicine in 2006. The two services were merged in 2009 and today form the F1000 database. Papers for F1000 are selected by a peer-nominated global "Faculty" of leading scientists and clinicians who then rate them and explain their importance (F1000, 2012). This means that only a restricted set of papers from the medical and biological journals covered is reviewed, and most of the papers are actually not (Kreiman & Maunsell, 2011; Wouters & Costas, 2012).

The Faculty nowadays numbers more than 5,000 experts worldwide, assisted by 5,000 associates, which are organised into more than 40 subjects (which are further subdivided into

over 300 sections). On average, 1,500 new recommendations are contributed by the Faculty each month (F1000, 2012). Faculty members can choose and evaluate any paper that interests them; however, "the great majority pick papers published within the past month, including advance online papers, meaning that users can be made aware of important papers rapidly" (Wets, Weedon, & Velterop, 2003, p. 254). Although many papers published in popular and high-profile journals (e.g. *Nature*, *New England Journal of Medicine*, *Science*) are evaluated, 85% of the papers selected come from specialised or less well-known journals (Wouters & Costas, 2012). "Less than 18 months since Faculty of 1000 was launched, the reaction from scientists has been such that two-thirds of top institutions worldwide already subscribe, and it was the recipient of the Association of Learned and Professional Society Publishers (ALPSP) award for Publishing Innovation in 2002 (<http://www.alpsp.org/about.htm>)" (Wets et al., 2003, p. 249). The F1000 data base is regarded as a useful aid for researchers to obtain indications of the most relevant papers in their subject area: "The aim of Faculty of 1000 is not to provide an evaluation for all papers, as this would simply exacerbate the 'noise', but to take advantage of electronic developments to create the optimal human filter for effectively reducing the noise" (Wets et al., 2003, p. 253).

### **Use of the Mendeley API**

Within the first half of 2014 the reference manager Mendeley provided a new version of its Application Programming Interface (API). Some restrictions of the previous API were lifted: For example, the usage statistics were previously provided in relative terms and only for the top three entries (Haustein & Larivière, 2014). The new API provides results in absolute numbers and not only for the top three but for all entries. Mendeley provides access to the readership status (e. g. professor, postdoc, or student) and the distribution of the Mendeley readership across scientific disciplines as well as countries via the API. Those sets

of data can be correlated with other information available about papers (e. g. citations or Twitter counts).

Before one can start to use the Mendeley API, one has to register as a Mendeley user. Afterwards, registration of the desired application is necessary (<http://dev.mendeley.com>). Authentication with the API is done via OAuth 2.0. The credentials are set during registration of the application.

We used R (<http://www.r-project.org/>) to interface to the Mendeley API. It seems to us that using other interfaces does not change the functionality or responsiveness. Mendeley provides sample codes for Javascript, Python, R, and Ruby ([http://dev.mendeley.com/code/sample\\_code.html](http://dev.mendeley.com/code/sample_code.html)), whereby all requests to the API use HTTP GET and POST requests. Therefore, we suppose that any other scripting or programming language may be used. The reply is sent in Javascript Object Notation (JSON).

We requested user statistics for the F1000 publication set ( $n = 114,582$  papers) using the PubMedID and DOI between the 4<sup>th</sup> and 6<sup>th</sup> of December 2014. We observed seemingly random connection problems. Sometimes those problems occurred after a few hundred or a few thousand requests. The largest chunk of requests we were able to get through the API without connection problems consisted of 47,629 papers. This large number of records is contrasted with smaller chunks of requests (between 1,049 and 9,307 records).

Mendeley provides a breakdown of the user count into sub-disciplines. The possible values for disciplines and sub-disciplines can be obtained directly from the API via the GET /disciplines endpoint. Each discipline has a certain number of sub-disciplines. The sub-discipline “miscellaneous” occurs in every discipline. Each Mendeley user can select a discipline and a sub-discipline from a drop-down menu. This piece of information is not mandatory, like the user’s location.

## Network analysis

Pajek is used to create the F1000 readership network (<http://pajek.imfm.si/doku.php>; de Nooy, Mrvar, & Batagelj, 2011) applying the spring embedder of Kamada and Kawai (1989). For detecting communities in the common readership of F1000 papers, we used the VOS Clustering algorithm (Waltman, van Eck, & Noyons, 2010), which is available in Pajek. The aim of this algorithm is to provide further insight into the structure of the network (Milojević, 2014).

## Results

We find 6,263,913 Mendeley readers for the F1000Prime publication set. 99.9% (n=6,257,603) of them share their discipline and sub-discipline. This is a much higher percentage than those who share their geographical location (Haunschild, Stefaner, & Bornmann, in preparation). For the F1000Prime publication set, the vast majority (74.94%) of Mendeley users is found in the “miscellaneous” sub-discipline of all disciplines. Therefore, we added up all the readers of all sub-disciplines for each discipline. The results of our study are presented in Table 1. Nine disciplines have at least 1% of the readers of the F1000Prime publication set. The remaining 16 disciplines have less than 1%. As expected, most readers (81.78%) of the F1000Prime literature assign themselves to the bio-medical (sub-) disciplines. All other disciplines comprise the remaining 15.19% of the F1000Prime readership at Mendeley. The third largest readership is found in the discipline psychology which is related to medicine. After chemistry, which is also related to biology, five other disciplines show readership values above 1% within the F1000Prime literature. Those disciplines seem rather unrelated to the field of bio-medical research, especially environmental sciences (according to Figure 1). 3.05% of the F1000Prime readers at Mendeley come from other disciplines (not shown in Table 1). The disciplines with most readers below the threshold of 1% are: social

sciences (0.67%), mathematics (0.42%), electrical and electronic engineering (0.27%), education (0.22%), and materials sciences (0.21%).

Table 1

Mendeley users from different disciplines reading F1000Prime papers. Only disciplines are shown with more than 1% readers (sorted in decreasing order)

Discipline	% of all Mendeley readers of F1000Prime papers
Biology	66.32
Medicine	15.46
Psychology	3.98
Chemistry	2.92
Environmental sciences	2.71
Engineering	1.61
Computer and information science	1.54
Physics	1.42
Earth sciences	1.01
Others	3.05

We also analyzed connections between the disciplines. They are shown in Figure 1. A paper which is read by Mendeley users of different disciplines (e.g. biology and physics) constitutes a connection between these disciplines. Therefore a paper which is read by Mendeley users of the same discipline does not contribute to the network system, but a paper which is read by Mendeley users of many different disciplines contributes many connections to the network. The size of the vertices in Figure 1 reflects the numbers of readers for each discipline. The thicker and darker the edges between two disciplines, the more frequently they have read a F1000 papers jointly. The location of the discipline vertex also informs about the connectivity: The closer the vertex is located towards the center, the more connections are found to different disciplines. There are 25 disciplines and 300 links among those disciplines in the dataset. With a density of 1, the network is rather dense. The average node degree is 24. According to Figure 1, the strongest connection shows up between biology and medicine. The disciplines computers and information science (CIS), engineering (Eng), and chemistry (Chem) have rather strong connections to biology and medicine. The discipline arts and

literature (AnL) shows a low amount of readers (0.15%, close to sports and recreation with 0.16%) as well as a good connection to other disciplines in the network.

The community detection algorithm detected two communities in the network with Biology, Medicine, Engineering, Chemistry, and Physics as one community (yellow vertices) and all other disciplines as the other (green vertices).

## Discussion

The (sub-) discipline of Mendeley readers is self-assigned and not mandatory. Still, we find that a large share (99.9%) of F1000Prime paper readers at Mendeley do share their (sub-) discipline. Most readers (74.94%) assign the “miscellaneous” sub-discipline of their discipline to themselves. As the F1000Prime publication set is a collection of high-quality bio-medical papers, it is expected that we find most readers in the disciplines biology and medicine. We find strong connections of the disciplines engineering, chemistry, and physics to biology and medicine as well as their rather high reader percentages. The connections of arts and literature (AnL) to biology and medicine are much weaker. The disciplines arts and literature (AnL), engineering (Eng), materials sciences (Mate), and computer and information science (CIS) have connections to many other disciplines, as the central location in the network indicates.

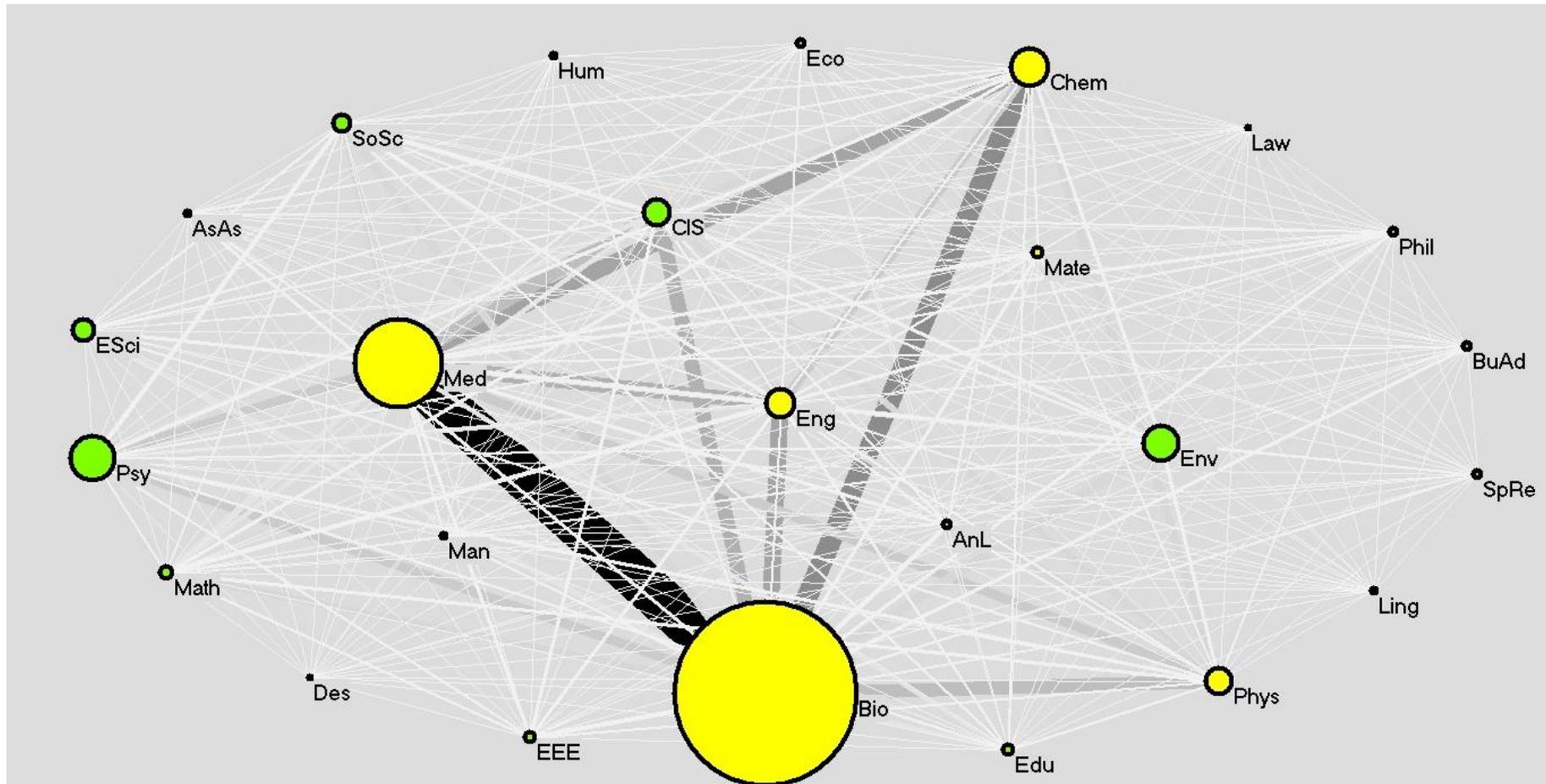


Figure 1

Network of F1000 papers' readers from arts and literature (AnL), astronomy and astrophysics (AsAs), biology (Bio), business administration (BuAd), chemistry (Chem), computer and information science (CIS), design (Des), earth sciences (ESci), economics (Eco), education (Edu), electrical and electronic engineering (EEE), engineering (Eng), environmental sciences (Env), humanities (Hum), law (Law), linguistics (Ling), management (Man), materials sciences (Mate), mathematics (Math), medicine (Med), philosophy (Phil), physics (Phys), psychology (Psy), social sciences (SoSc), sports and recreation (SpRe).

## References

- Adie, E. (2014). Taking the Alternative Mainstream. *Profesional De La Informacion*, 23(4), 349-351. doi: DOI 10.3145/epi.2014.jul.01
- Bornmann, L. (2012). Measuring the societal impact of research. *EMBO Reports*, 13(8), 673-676.
- Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society of Information Science and Technology*, 64(2), 217-233.
- Colledge, Lisa. (2014). Snowball Metrics Recipe Book. Amsterdam, the Netherlands: Snowball Metrics program partners.
- de Nooy, Wouter, Mrvar, Andrej, & Batagelj, Vladimir. (2011). *Exploratory social network analysis with Pajek*. New York, NY, USA: Cambridge University Press.
- F1000. (2012). What is F1000? Retrieved October 25, from <http://f1000.com/about/whatis>
- Haunschild, Robin, Stefaner, Moriz, & Bornmann, Lutz. (in preparation). *Who publishes, reads, and cites papers? An analysis of country information*.
- Haustein, Stefanie, & Larivière, Vincent. (2014). A multidimensional analysis of Aslib proceedings – using everything but the impact factor. *Aslib Journal of Information Management*, 66(4), 358-380. doi: doi:10.1108/AJIM-11-2013-0127
- Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7-15.
- Kreiman, G., & Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, 5. doi: 10.3389/fncom.2011.00048
- Milojević, Staša. (2014). Network Analysis and Indicators. In Y. Ding, R. Rousseau & D. Wolfram (Eds.), *Measuring Scholarly Impact* (pp. 57-82): Springer International Publishing.
- Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance*. Cambridge, MA, USA: MIT Press.
- Thelwall, M., & Maflahi, N. (in press). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology*.
- Van Noorden, R. (2014). Scientists and the Social Network. *Nature*, 512(7513), 126-129.
- Waltman, Ludo, van Eck, Nees Jan, & Noyons, Ed C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629-635. doi: <http://dx.doi.org/10.1016/j.joi.2010.07.002>
- Weller, Katrin. (2015). Social Media and Altmetrics: An Overview of Current Alternative Approaches to Measuring Scholarly Impact. In I. M. Welpé, J. Wollersheim, S. Ringelhan & M. Osterloh (Eds.), *Incentives and Performance* (pp. 261-276): Springer International Publishing.
- Wets, K., Weedon, D., & Velterop, J. (2003). Post-publication filtering and evaluation: Faculty of 1000. *Learned Publishing*, 16(4), 249-258.
- Wouters, Paul, & Costas, Rodrigo. (2012). Users, narcissism and control – tracking the impact of scholarly publications in the 21st century. Utrecht, The Netherlands: SURFfoundation.