The Impact of Recommenders on Scientific Article Discovery: The Case of Mendeley Suggest

Minh Le
Subhradeep Kayal
Andrew Douglas
{m.le,a.douglas}@elsevier.com
deep.kayal@pm.me
Elsevier
Amsterdam, Netherlands

ABSTRACT

Mendeley Suggest is a popular academic paper recommender, serving over 1.5M researchers in 2018. We attempt to assess the extent Mendeley Suggest helps its users in their research in two areas: helping researchers keep up with the most prominent development in the field and help researchers find relevant literature. Our findings indicate that the recommender significantly increases the chance that a user finds important research and decreases the amount of time she needs to spend on searching. We observe that the effect is much greater than the number of accepted recommendations and propose that it is due to an increase in reading activity that Mendeley Suggest recommendations spur. Time-series analyses are presented to back up this hypothesis. Our results highlight the potential of academic paper recommenders in furthering science.

CCS CONCEPTS

• **Information systems** → *Collaborative filtering*; *Digital libraries and archives*; • **Applied computing** → *Digital libraries and archives*.

KEYWORDS

Scientometrics, Recommender Systems, Mendeley Suggest

ACM Reference Format:
Minh Le, Subhradeep Kayal, and Andrew Douglas. 2019. The Impact of Recommenders on Scientific Article Discovery: The Case of Mendeley Suggest. In *Proceedings of 1st Workshop on the Impact of Recommender Systems (ImpactRS '19).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The International Association of Scientific, Technical, and Medical Publishers reported in mid-2018[1] that there were about 33,100 active scholarly peer-reviewed English-language journals, collectively publishing over 3 million articles a year, with a steady 3-5 % yearly growth, for about 7 million researchers in the globe. With such staggeringly large numbers of scientific articles, the need for efficient mechanisms of discovery is real and pressing. One promising approach to alleviate this need is the use of academic paper recommendation systems to help researchers save time while staying on top of the latest development in their field of research. However, how far existing recommenders meet this need and fulfill their promise is still an open question.

In this work, we attempt to answer this question for the case of Mendeley Suggest *(MS)*,[2] an article recommender that is used within the popular social reference manager, Mendeley.[3] Mendeley was inaugurated in 2008 and has grown to 6.5 million users in 2017,[4] and MS has accompanied it since early 2016 and attracted over 1.5 million users last year.

2 BACKGROUND

2.1 Related Work

Several research papers have investigated proxies for citations garnered by published articles, such as the work of Haustein et al. [5] and Sotudeh et al. [12], who found weak correlations between published articles and their mentions in Tweets or their CiteULike[5] bookmarks, respectively. In terms of studying the predictive power of Mendeley readership, Haustein et al. [4] and Schlögl et al. [11] both found a moderate correlations between the Mendeley readership and Scopus citations in bibliometric literature and information systems journals. Improving upon previous studies in terms of scale, Zahedi et al. [13] studied 9 million documents on Web of Science and found that Mendeley readership is a better proxy for identifying highly cited articles, in comparison with journal-based citation scores, although they cannot be considered as equivalent indicators [14], while Costas et al. [2] show that such altmetrics have higher precision but lower recall, when it comes to being able to identify high-impact articles, as compared to journal based citation scores.

Additionally, there is also substantial existing literature studying the effects of recommender systems, both analytically and for in-use cases. For example, Fleder et al. [3] make an analytical model for recommenders and show that recommenders might decrease the overall sales diversity, as they push popular products in an online store, while the overall sales was shown to increase due to the effect of cross-selling, as shown by the empirical study of Pathak et al. [9]. Hostler et al. [6] showed both theoretically and empirically

[1] https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf

[2] https://www.mendeley.com/suggest/
[3] https://www.mendeley.com
[4] https://www.elsevier.com/__data/assets/pdf_file/0011/117992/Mendeley-Manual-for-Librarians_2017.pdf
[5] https://en.wikipedia.org/wiki/CiteULike

ImpactRS '19, September 19, 2019, Copenhagen, Denmark

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

that the use of a recommender system enhances the consumers' satisfaction with the website and provides a more effective product search process. Zhou et al. [15] used crawled data from YouTube to reveal that there is a strong correlation between the view count of a video and the average view count of its top referrer videos. Apart from these specific works, Pu et al. [10] provide a survey of evaluation procedures for recommender systems from a user's perspective.

2.2 A Brief Overview of Mendeley Suggest

Mendeley is a free reference manager and an academic social network where users can manage their interests by creating a personal repository, called *library*, of articles which they find useful. Mendeley also provides a reader equipped with highlight and annotation functionalities on desktop, web, and mobile.

All Mendeley users automatically have access to Mendeley Suggest *(MS)*, an article recommender that uses collaborative and content-based approaches [7]. The tool exists as a separate tab on Mendeley website and mobile app. On the desktop application, a user can click on the "Related" button to retrieve suggestions based on the currently selected article. To encourage a focused reading experience, however, the button is *not* available while reading and there is also no tab for MS in the Mendeley reader. In addition, MS recommendations are integrated into Mendeley newsfeed and people can opt for receiving recommendations via email.

The MS recommender comprises different types of recommenders, which tackle the various disciplines and levels of seniority of researchers who use Mendeley. The primary recommender is based on a collaborative filter which makes use of similarities between users' libraries, i.e. predicting whether a user is interested in a paper based on whether similar users have the document in their libraries. One of the drawbacks of a collaborative filtering approach is its susceptibility to the cold-start problem, wherein newly added articles cannot be immediately recommended and new users cannot be served recommendations. To circumvent this problem, Suggest also has a content-based recommender, based on ElasticSearch *more-like-this* queries, and weighted by the popularity of articles.[6]

In addition to the recommenders, MS also applies dithering and impression discounting [8] to the set of produced recommendations to promote a feeling of freshness, so that users, on successive logins within very short periods of time, do not see the same static list.

3 METHOD

We attempt to quantify the value MS brings to its users along two dimensions: coverage and time. If we know the set $\mathcal{D} = \{(p, u)\}$ of all the papers $\{p\}$ that each researcher $\{u\}$ ought to read, we could measure how much of them she covers at a certain point in time, both through recommendations and other means, and we would hope that MS users reach higher coverage in a shorter amount of time compared to non-users. Although this ideal cannot be attained, we will later propose relaxations that capture some perspectives of the set.

[6]https://www.elastic.co/

3.1 Terminologies

Before presenting our experimental design, we will introduce a few terms used in this paper that requires specification beyond what is given by common sense.

MS works by giving users recommendations on papers to read, presented, for example, as a list on a web page or on a tab integrated in the Mendeley mobile application. We track anonymized interactions with MS via, among others, two types of events: **recommendation viewing** and **recommended addition**. Viewing a recommendation entails a user *clicking* on a link in the recommendation list, upon which a document page will be *opened*. At this point, the document is not added to the user's library yet. The user can actively do so by clicking on a button that says "Add to library". She can also add the same paper through other means (e.g. importing a PDF or pasting a bibtex entry) which are not captured as a recommended addition.

The routine of a user includes collecting documents to build up her library. We will refer to this activity as **additions**, which can include articles recommended by MS. For all papers, we have the timestamp of the last time they are added to a user's library. We also track events related to **annotations** performed on documents in users' library. When a line is highlighted or a note edited, we record the timestamp and action type for analytic purposes.

Throughout the paper, we assume the same notion of **articles** in MS, Mendeley libraries, and the literature as represented by Scopus. Similarly, in the scope of this paper, **citations** are treated as a given. Behind the scene, they are extracted via the machinery internal to Scopus [1].

3.2 User Groups

A common technique to measure the performance of recommender systems is A/B test. A control group A and an experimental group B are typically served two versions of a system that differ in a single feature. Although highly effective in measuring short-term direct effect, sustaining a long A/B test is often difficult in a commercial setting because of its negative effect on customer experience. More importantly, the approach is only suited to study versions of a recommender but not the very effect of using it because we cannot, in normal circumstances, bar users from using the product to create a control group.

As an alternative, we study groups of users differing in Mendeley Suggest usage. By measuring at a user-group level during an extended period of time, we can capture both direct and indirect effects of our recommender system.

Measured by the number of recommendation views between January 2018 and July 2019, the distribution of Mendeley users resembles a Zipfian curve, with most users opening less than one article per week. To study the effect of different degrees of usage, we divide this population into four chunks:

- **S-heavy** Users who clicked on the most recommendations, belonging to the 95% quantile,
- **S-frequent** Users who are less active than the first group but belong to the 75% quantile. This group of users viewed more than 2.5 recommendations per week during the period we observed.

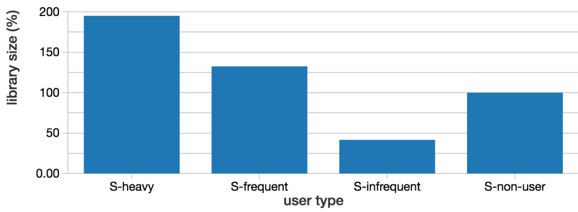

Figure 1: The median number of articles added to Mendeley library per user type, normalized to that of non-users

S-infrequent The remaining users who clicked on at least one recommendation, and

S-non-user Mendeley users who did not open any recommendation. To reduce computational complexity, we extract a random sample of 400,000 users.

We only include in **S-non-users** people who added at least one article to their Mendeley library since 2018. There can be various reasons an active user of Mendeley does not use Mendeley Suggest. Since the platform is most known for its reading and reference managing functionalities, a user might simply never encounter Mendeley Suggest. She might also have decided not to use it in the past. We leave an in-depth examination of the non-user group for future work.

Figure 1 shows the relative library size of user types, normalized to that of non-users. It can be observed that higher MS usage coincides with higher Mendeley usage overall, except between infrequent Suggest users and non-users. This is a factor affecting coverage that we will comment on later.

3.3 Coverage of Most-cited Recent Papers

As the first relaxation of the ideal paper assignment set $\mathcal{D}$, we propose to study the set $\mathcal{D}_1$ of recent and most-cited articles in the literature. Arguably, it is important for a researcher to be aware of the latest major development in her field, regardless of whether she is going to use it directly in her research.

To construct $\mathcal{D}_1$, we sort articles published in 2018 onward according to the number of times they are cited. For each field as codified by Scopus's All Science Journal Classification Codes (ASJC)[7], an excerpt of which can be found in Table 1, we extract the 100 most cited articles that is unambiguously in the field (i.e., being assigned to only one ASJC code). A sample of the papers we extracted can be seen in Table 2. We do not possess an up-to-date mapping from researchers to their field of research, therefore, we treat articles from every field equally. Given the contrast between the broad scope of ASJC codes and the narrow specialization of researchers, we do not expect a researcher to have read many of the extracted articles. We choose not to calculate Peason correlation because counting the number of articles might mistake broad-mindedness (or a lack of focus) for the coverage of useful literature.

The extent that a group of users capture the latest literature is therefore defined as the proportion of its members who added to their library at least one of the extracted articles:

$$\text{coverage}_1 = \frac{|\{\text{user who added at least one extracted article}\}|}{|\{\text{all users}\}|}$$

[7] http://www.researchbenchmarking.org/files/subject_hierarchy.pdf

Name	Code
Arts and Humanities (miscellaneous)	1201
Colloid and Surface Chemistry	1505
Geotechnical Engineering and Engineering Geology	1909
Ocean Engineering	2212
Oncology	2730

Table 1: Some ASJC codes picked at random

We checked that papers from $\mathcal{D}_1$ are reachable by MS, with the number of articles recommended to at least one user spreads relatively evenly across fields between 1 and 100 (mean=55, stddev=29).

3.4 Coverage of Personalized Citable Papers

In the second perspective, we attempt to measure the effectiveness of MS in helping users find papers that they might want to cite later on. To evaluate this, we construct the set $\mathcal{D}_2 = \{(u, p)\}$ of papers $\{p\}$ that Mendeley users $\{u\}$ cited between January 2018 and July 2019. This information is available to us via a feature in Mendeley that allows users to claim their Scopus profile. Publications of an author and the out-going citations were automatically extracted and can be readily queried via Scopus.

The coverage of citable articles for a group of users is proportional to the number of papers p they added to their Mendeley libraries *before* the publication of any of their articles citing p:

$$\text{coverage}_2 = \frac{|\{\text{pairs of } \langle\text{user, added paper that is later cited}\rangle\}|}{|\{\text{pairs of } \langle\text{user, cited paper}\rangle\}|}$$

Because users can combine Mendeley with other means of reference management and distribute references across co-authors, we do not expect the coverage to reach 100%. Ideally, we would like to measure literature added to Mendeley library before the submission of a paper but this data is not available to us. The delay between submission and publication might artificially increase coverage. However, we expect it to be the same across groups of users.

4 RESULTS AND DISCUSSIONS

In this section, we will present the results of experiments outlined in the previous section and their implications.

4.1 Staying Up to Date

Figure 2 shows the adoption curves of different groups of users w.r.t. our set of most cited recent papers. It is clear that the more a researcher uses MS, the more likely she finds the latest important paper.

The difference cannot be explained by the level of activity alone. Although **S-heavy** users added only twice as many articles into their Mendeley libraries compared to **S-non-user** (see Figure 1), they reached a coverage of 0.3843 compared to 0.0057 of **S-non-user** in July 2019 (68 times higher). Moreover, **S-infrequent** users who added to their library 50% less articles than **S-non-user** (see Figure 1) still got 23 times higher chance of staying up-to-date with

Field	Title	#cit.
Fluid Flow and Transfer Processes	Analytical and numerical solution of non-Newtonian second-grade fluid flow on a stretching sheet	26
Biochemistry	Directed Evolution of Protein Catalysts	33
Emergency Medicine	Low Accuracy of Positive qSOFA Criteria for Predicting 28-Day Mortality in Critically Ill Septic Patients During the Early Period After Emergency Department Presentation	24
Pharmacology, Toxicology and Pharmaceutics (all)	An updated overview on the development of new photosensitizers for anticancer photodynamic therapy	45
Hepatology	The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases	342

Table 2: Examples of most cited recent papers in a field

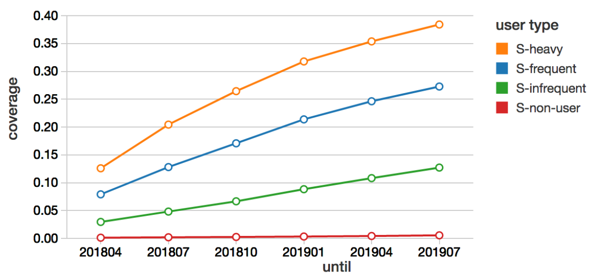

Figure 2: The proportion of users who discovered at least one article in our set of most-cited recent articles

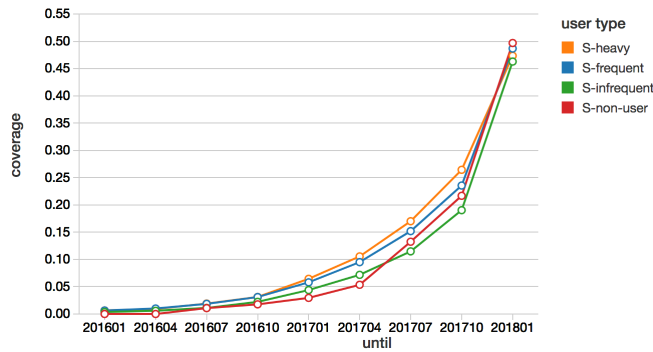

Figure 3: The amount of cited papers that were discovered by the author w.r.t. time.

the most important research (comparing a coverage of 0.1274 with 0.0028 in July 2019).

This result demonstrates the benefit of using MS, even in a non-frequent basis.

4.2 Finding Articles To Cite

Figure 3 shows that more frequent MS users discover more citable papers in early stages (the largest difference is 0.06 between **S-heavy** and **S-non-user** in April 2017). The effect dissipates with time and reverses with **S-non-user** performing the best in January 2018, right before their first citations of papers in $\mathcal{D}_2$. In this last time point, the largest different is 0.04 between **S-non-user** and **S-infrequent**.

Although the effect of MS is smaller and varies more with time in this use case, it is encouraging that **S-infrequent** users exert less effort on curating their library while still discovering comparably relevant papers.

We hypothesize that the observed dynamics reflect stages of a research project which we shall call *discovery*, *development*, and *finalization*. During the discovery phase, a researcher maintains a small number of "seed" articles related to the research topic. This collection is enlarged in every direction during the development phase. Finally, close to submission time, the researcher focuses on adding a lot of related literature and supporting articles. Whereas the last miles are characterized by deliberate and directed searches, the early stages are when recommender systems can make the largest impact via undirected discovery and serendipity.

4.3 Direct and Indirect Effects

The results in the previous sections are surprising when we consider the relatively small amount of viewed recommendations per week (see Section 3.2). Upon conducting a quick analysis, we found that, for **S-infrequent** Suggest users, the number of all additions to Mendeley library is 172 times that of additions recommended by Suggest. In the case of **S-frequent** and **S-heavy**, there are, respectively, 46 and 24 library additions for each recommendation by Suggest.

We hypothesize that the indirect effect of MS is much bigger than the direct one. In one scenario, upon reading a relevant paper, a researcher might follow forward and backward citations to gain a more exhaustive understanding of her field. Alternatively, a researcher might discover a new topic by serendipity, broadening her coverage. If this is the case, we expect an increase in additions to library when people use MS.

To validate this hypothesis, we study the usage pattern of **S-infrequent** users in the first quarter of 2019. As mentioned in Section 3.1, we have records of anonymized addition events in Mendeley. For the analysis, they are divided into two categories: those that occur on the same day as a recommendation viewing event and those do not. The results can be seen in Figure 4. In line with our prediction, days that people use MS see 1.55 times more articles added to their library.

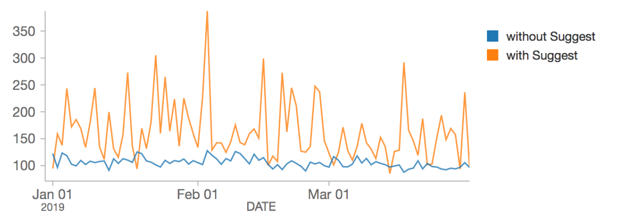

Figure 4: Additions into the Mendeley library of S-infrequent users in Q1 2019 in two scenarios: when they use and do not use MS. Numbers of additions are normalized such that the average activity without using MS is 100%.

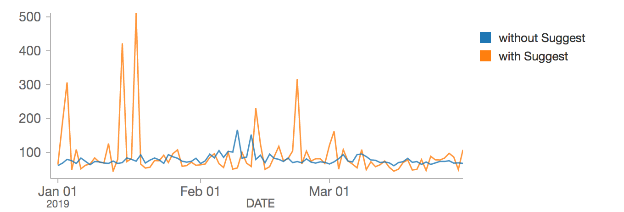

Figure 5: Annotation-related events of S-infrequent users in Q1 2019 in two scenarios: using and not using MS. Numbers of events are normalized such that the mean activity level without using MS is 100%.

We also check if MS usage coincides with deeper reading by looking at annotation events. Following the same procedure, our analysis shows that, on average, people annotate much more around the time they use the recommender. Although Suggest is observed together with increased annotating in only 42 days as opposed to the 47 days that it sees less activity, the peaks are much higher than the depth of the troughs (Figure 5). We repeated the experiments with **S-frequent** and **S-heavy** and obtained similar results although the effect is less pronounced: articles added together with MS usage are 1.14 and 1.13 times as many as without.

5 CONCLUSIONS

In the current research, we study the impact of Mendeley Suggest on scientific researchers. Through various analyses, we showed that MS increases the chance that a researcher finds important and relevant literature, in a more timely manner. We propose a mechanism to explain this effect in which a researcher does not stop at adding a recommended article to her library but read the content in depth and explore further to deepen and broaden her grasp of the literature. Evidences from Mendeley usage log are presented to support our hypothesis.

The results of our research highlight the positive effect a scientific article recommender can have on researchers' professional lives. Considering that MS is composed of standard techniques such as nearest-neighbor collaborative filtering and ElasticSearch-based content recommendations, without a reranking step, there is much room for improvement.

A limit of the current research is its observational nature. There are alternative explanations that, given the limited resources the authors possess, we could not eliminate. For example, the correlation between MS usage and reading activities might be because users tend to open recommendations when they have more time to read. Further research is needed to disentangle factors and reach a clearer picture of the recommender's impact.

REFERENCES

[1] Judy F. Burnham. 2006. Scopus database: a review. *Biomedical Digital Libraries* 3, 1 (2006), 1. https://doi.org/10.1186/1742-5581-3-1
[2] Rodrigo Costas, Zohreh Zahedi, and Paul Wouters. 2015. Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology* 66, 10 (2015), 2003–2019. https://doi.org/10.1002/asi.23309
[3] Daniel M. Fleder and Kartik Hosanagar. 2007. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *SSRN eLibrary* (2007). https://doi.org/10.1287/mnsc.1080.0974
[4] Stefanie Haustein, Isabella Peters, Judit Bar-Ilan, Jason Priem, Hadas Shema, and Jens Terliesner. 2014. Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics* 101, 2 (2014), 1145–1163. https://doi.org/10.1007/s11192-013-1221-3
[5] Stefanie Haustein, Isabella Peters, Cassidy R Sugimoto, Mike Thelwall, and Vincent Larivière. 2014. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology* 65, 4 (2014), 656–669. https://doi.org/10.1002/asi.23101
[6] R. Eric Hostler, Victoria Y. Yoon, Zhiling Guo, Tor Guimaraes, and Guisseppi Forgionne. 2011. Assessing the impact of recommender agents on on-line consumer unplanned purchase behavior. *Information Management* 48, 8 (2011), 336 – 343. https://doi.org/10.1016/j.im.2011.08.002
[7] Maya Hristakeva, Daniel Kershaw, Marco Rossetti, Petr Knoth, Benjamin Pettit, Saúl Vargas, and Kris Jack. 2017. Building recommender systems for scholarly information. In *Proceedings of the 1st workshop on scholarly web mining*. ACM, 25–32. https://doi.org/10.1145/3057148.3057152
[8] Pei Lee, Laks V.S. Lakshmanan, Mitul Tiwari, and Sam Shah. 2014. Modeling Impression Discounting in Large-scale Recommender Systems. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. 1837–1846. https://doi.org/10.1145/2623330.2623356
[9] Bhavik Pathak, Robert Garfinkel, Ram Gopal, Rajkumar Venkatesan, and Fang Yin. 2010. Empirical Analysis of the Impact of Recommender Systems on Sales. *J. Manage. Inf. Syst.* 27, 2 (2010), 159–188. https://doi.org/10.2753/MIS0742-1222270205
[10] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the user's perspective: Survey of the state of the art. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 317–355. https://doi.org/10.1007/s11257-011-9115-7
[11] Christian Schlögl, Juan Gorraiz, Christian Gumpenberger, Kris Jack, and Peter Kraker. 2014. Comparison of downloads, citations and readership data for two information systems journals. *Scientometrics* 101, 2 (2014), 1113–1128. https://doi.org/10.1007/s11192-014-1365-9
[12] Hajar Sotudeh, Zahra Mazarei, and Mahdieh Mirzabeigi. 2015. CiteULike bookmarks are correlated to citations at journal and author levels in library and information science. *Scientometrics* 105, 3 (2015), 2237–2248. https://doi.org/10.1007/s11192-015-1745-9
[13] Zohreh Zahedi, Rodrigo Costas, and Paul Wouters. 2017. Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology* 68, 10 (2017), 2511–2521. https://doi.org/10.1002/asi.23883
[14] Zohreh Zahedi and Stefanie Haustein. 2018. On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications. *Journal of Informetrics* 12, 1 (2018), 191–202. https://doi.org/10.1016/j.joi.2017.12.005
[15] Renjie Zhou, Samamon Khemmarat, and Lixin Gao. 2010. The Impact of YouTube Recommendation System on Video Views. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*. 404–410. https://doi.org/10.1145/1879141.1879193