SRHE Society for Research into Higher Education

‘Sort by relevance’: Exploring assumptions about algorithm-mediated academic literature searches

Research report - January 2023

Dr Katy Jordan & Dr Sally Po Tsai
Faculty of Education, University of Cambridge

UNIVERSITY OF CAMBRIDGE
Faculty of Education

Contents

Acknowledgements

We would like to thank the Society for Research in Higher Education for funding this project, without which the project would not have been possible. Many thanks to Dr Ibrar Bhatt as the critical friend for the project, and to Rob Gresham for his support.

We would also like to thank the members of the Faculty of Education who have supported the project.

We are extremely grateful to the participants who took part in the survey and the interviews, and to the individuals and organisations which helped to publicise the call for participation.

About the Authors

Katy Jordan is a Senior Research Associate at the Faculty of Education, University of Cambridge. She holds a PhD from the Institute of Educational Technology at the Open University, UK. Her research interests broadly sit at the intersection of educational research and internet studies, and she has published research on a range of topics, including social media in higher education, massive open online courses, and gender equity and educational technology. klj33@cam.ac.uk

Sally Po Tsai is a Research Assistant at the Faculty of Education, University of Cambridge. She has participated in a number of research projects. Her research interests are related to educational effectiveness research, educational technology and digital scholarship. pt374@cam.ac.uk

To cite this report:
Jordan, K. & Tsai, S.P. (2022). 'Sort by relevance': Exploring assumptions about algorithm-mediated academic literature searches. Society for Research into Higher Education (SRHE) Research Report.

Aim

Access to scholarly literature has arguably never been easier, given the ease of searching for academic publications through the internet. An increasing number of online platforms are now available for academics - and other users - to search the literature. However, platforms often utilise algorithms to manage how search query results are prioritised and presented to users, presenting results ranked by ill-defined criteria such as 'relevance'. This introduces an opaque layer to how academics and the public engage with the literature, with potentially important implications for rigour and equity. To date, no studies have explored academics' perceptions and assumptions of how processes such as the Google Scholar algorithm operate and influence their access to information; this study has sought to help address this gap.

Background to the study

The Internet has revolutionised access to information, and a range of online scholarly databases and academic social networking sites provide platforms to search for academic literature. However, an ever-increasing volume of information online brings challenges in searching effectively. Such challenges have been observed in other types of platforms, but have not been explored previously in the context of academic literature platforms. As the early World Wide Web grew, so did the challenge of being able to search it effectively (König & Rasch, 2014). Information retrieval has long been a core concern within Information Sciences, and a range of applications developed in the form of early Web search engines (Büttcher et al., 2016). Notably, Google emerged and dominated the market, through its innovative development and application of the PageRank algorithm. The algorithm extended beyond text matching and exploited the fact that the Web is a network, accounting for the features of how connected a page or website is as well as the content (Page et al., 1999).

Algorithms are now embedded into many of the ways in which people access information and content online, including social media, which has led to folk theories among users about the role that algorithms play (DeVito et al., 2017). In this context, 'the algorithm' determines whether or not content is promoted within news feeds, with a potentially dramatic impact on how wide an audience the content reaches (Kim, 2017). Such algorithms are usually intended to aid the user, by providing a calculated way to present the most 'relevant' material from an unmanageably large number of search results – but can also obscure exactly why particular results have been prioritised. There is increasing recognition of the role that algorithms are playing in mediating users' experiences online, and the hidden dangers that this brings. Depending on the sources used to construct or train the algorithm, there is a risk that it will amplify biases present in the data, and as such there is a growing call for transparency in the design and use of algorithms (Kossow et al., 2021).

While the role of algorithms in mediating access and visibility of content online is a much broader societal issue, it is also a potential risk for academic work. A view of these issues, from an academic perspective, were recently the focus of an article in the Times Higher Education (Matthews, 2021). Depending on the types of information used in the ranking algorithms, there is a risk that the way in which results are prioritised may exacerbate

existing biases within academic publishing (ibid.). If a ranking algorithm utilises data such as the number of citations or favours journals with a high impact factor, for example, it will serve to amplify the inequalities present in scholarly publishing. A lack of transparency about how ranking algorithms work carries methodological risks for literature reviews, as it is not clear exactly why a particular article has been deemed to be of high relevance in search results. This absence of transparency could have negative impacts on the rigour of literature reviews and potentially risk creating 'filter bubbles' - that is, that different users receive a different view on a field, if personal data is used to 'personalise' search results (Pariser, 2011). However, the article also highlighted the lack of research on how academics use these platforms and navigate these risks in practice. The present study was designed to help address this gap.

Google Scholar is the most prominent example of an online academic literature search platform which utilises ranking 'by relevance'. It is extremely popular, and free to use. One of the creators of the platform credits ranking by 'relevance' as having been a key reason why the platform gained an edge in the market (Van Noorden, 2014). In conducting the literature review for this project[1], we also looked for literature relating to relevance ranking in other platforms, but Google Scholar is the only example which has been addressed so far[2]. The publicly-available definition of how the ranking works is as follows:

> *"Google Scholar aims to rank documents the way researchers do, weighing the full text of each document, where it was published, who it was written by, as well as how often and how recently it has been cited in other scholarly literature." (Google Scholar, 2022).*

The definition supplements the content of articles with social information, which may introduce biases. For example, numbers of citations and reputation metrics for journals reflect biases in academic publishing (e.g. Czerniewicz, 2016; Larivière et al, 2013) - and combining such factors in determining search ranking may exacerbate this.

Findings from previous studies provide further insights into the Google Scholar algorithm, and confirm that the definition of relevance combines both the content and social information about articles. Beel and Gipp reported the earliest studies to examine the factors affecting the Google Scholar ranking algorithm, through a series of articles published in 2009 and 2010 (Beel et al., 2010). Citation counts were found to be the most influential factor affecting the ranking (Beel & Gipp 2009a; Beel & Gipp 2009c). Weighting differs according to where the search terms are found within the text, with greater weighting if a term is present in the title, compared to whether it is located within the full text (Beel & Gipp 2009a). The age of articles was not found to be significantly related to position in the ranking (Beel & Gipp 2009b). More recently, Rovira et al. (2018) confirm that citation counts remain the most influential factor, and report a bias favouring articles written in English (Rovira et al., 2021). Rovira et al. (2019) expand their earlier work to apply a similar approach to Microsoft

[1] The full literature review bibliography has been published separately (Jordan & Tsai, 2023a).

[2] A notable exception was Fiorini et al. (2018), a full paper reporting the novel 'Best Match' algorithm implemented by Pubmed.

Academic (now discontinued), Web of Science and Scopus, alongside Google Scholar. This comparative analysis revealed that citations were also the key factor in Microsoft Academic, and did not influence ranking in Scopus, while citations appeared to be intermittently involved in Web of Science ranking (Rovira et al., 2019). Only one study has addressed whether there is any evidence of a risk of 'filter bubbles' - that is, different results being presented to different users through personalisation - being formed when using Google Scholar (Yu et al., 2017).

Research questions

This study was guided by the following research questions[3]:

1. How prevalent is the use of ranking by 'relevance' in academic literature platforms, and how is relevance defined by the sites?
2. What are academics' assumptions about how algorithm-mediated literature searches (such as Google Scholar) work?
3. Does the perceived quality of results vary by platform, and what are the practical implications of this?
4. Is there any evidence to suggest that there is a risk of 'filter bubbles' being formed?

[3] Note that initially the project was based upon addressing three research questions (2, 3 and 4). Research question 1 was added at the start of the project; while Google Scholar was the starting point for the research enquiry, there was an initial question of whether other platforms should be included in the study too. No literature was found on this issue in relation to other platforms, so we needed to collect data first.

Methodology

An explanatory mixed-methods research design was used, collecting both quantitative and qualitative data, and using qualitative data to provide insight into quantitative data (Creswell & Plano Clark, 2011). Data collection included:

1. Information from platforms, about whether sorting 'by relevance' is used, and if so, how it is defined (quantitative and qualitative);
2. Online survey of academics who use the platforms (quantitative and qualitative);
3. Semi-structured co-interpretive interviews (qualitative).

Analysis of definitions of 'ranking by relevance'

To establish how prevalent the use of ranking 'by relevance' is, beyond Google Scholar, a sample of 14 of the largest academic bibliographic databases were examined (Table 1). For each platform, two types of information were recorded: first, whether or not ranking 'by relevance' is used as a way of presenting search results; and second, whether a definition of 'relevance' was provided. If so, the text was recorded, and categorised according to the types of information used.

Table 1: *Online academic literature search platforms included in the sample.*

Academia.edu	Lens.org
Aminer	Pubmed
Arxiv	Science Direct
Core.ac.uk	Scopus
Dimensions.ai	Semantic Scholar
Google Scholar	Web of Science
JSTOR	Zenodo

Survey

The online survey was undertaken during June and July 2022, using Qualtrics[4]. 101 participants fully completed the survey. While the survey did have an international reach, responses from the UK were over-represented, accounting for approximately half of the sample (53.3%). Google Scholar was the most popular platform, used to some extent by all the participants.

The survey questions have been published online as a research instrument (Jordan & Tsai, 2023c). The online survey served three purposes. First, to generate background data on academics' levels of use and search priorities when using online academic databases; second, to ask participants about their assumptions about ranking used by the platforms they use the most; and third, to recruit potential interview participants. Short-form text responses about participants' assumptions about how ranking is defined on the platforms that they use most frequently were analysed through arranging into categories through an approach using elements of grounded theory (Charmaz, 2014). This provided a short definition of relevance from a wider pool of participants, which complemented by in-depth insights from the interviews.

[4] https://www.qualtrics.com/

Interviews

Interviews were held with a sub-sample of survey participants who had indicated that they would be willing to take part. Twelve interviews were carried out, between August and October 2022. It had been hoped that a larger sample would respond but nonetheless, this is a reasonable sample size for in-depth qualitative interviews (Baker & Edwards, 2012).

The interview protocol takes a lead from a recent study undertaken in order to explore user assumptions about the algorithm behind content presentation on the TikTok social media platform, through qualitative interviews with users (Klug et al., 2021). The interviews were semi-structured (Wengraf, 2001), for a combination of consistency and also allowing for the discussions to be flexibly adapted according to the participants' responses.

All interviews were carried out online; video and audio of the interviews were recorded, and automatically transcribed. In the first part, screen-sharing was used. Participants were asked to carry out a search on a familiar specialist topic and talk through their perceived value or limitations of the particular results generated. In the second part, participants were also asked to carry out a pre-defined search, to explore whether results change for different individuals (to look for evidence of 'filter bubbles'). The interview protocol has been published online (Jordan & Tsai, 2023b).

The discussions around the first part were examined through thematic analysis, to identify themes in relation to the research questions. This was first undertaken by one of the researchers, and independently verified by the other. This process, in combination with triangulation with the open coded text responses from the online survey, served to ensure rigour in the qualitative analytical process (Miles & Huberman, 1994). The second part – regarding filter bubbles - was analysed using a novel approach, exporting the list of the first ten results from Google Scholar in order, and using colour-coding to visualise differences in the ranking across the sample.

Ethical considerations

At the start of the project, an application was made in accordance with the Faculty of Education ethical approval process. This involved completion of the Faculty ethics checklist (which is aligned with the British Educational Research Association's ethical guidelines; BERA, 2018). The checklist and supporting documents (draft survey, interview protocol and consent form) were reviewed by a senior member of the Faculty. The submission was approved, and a copy lodged with the Faculty research office.

Core ethical considerations for the project related to informed consent from participants, and secure handling and protection of identities in research data. The first page of the online survey provided information about the project, and purposes of the data collection and analysis. Participants were asked to indicate that they had read this information and consented to participate before proceeding to the questions. For interviews, consent forms were shared with potential participants and completed ahead of the sessions. At the start of each interview, verbal confirmation was sought to check that the participants were still happy to proceed on this basis. Once collected, research data were stored on secure online storage only accessible to the two researchers involved. The identities of participants were protected in the process of analysis and writing up of results.

Findings

RQ1: Prevalence and definition of ranking ‘by relevance’

Of the 14 platforms examined, all were found to offer the option to sort by ‘relevance’ or ‘best match’, and this was the default setting in all but two cases. However, the availability of definitions varied. Just over half of the platforms surveyed (8 of 14) provided information about how relevance was defined, and the amount of detail varied considerably.

For the platforms which did provide a definition of relevance, the main types of information which featured in those definitions are mapped in Figure 1. Broadly, the sample comprises two groups: one of platforms which use models based on aspects of the text, and a second group which combine text features with other data, such as the publication, author, citations, or date.

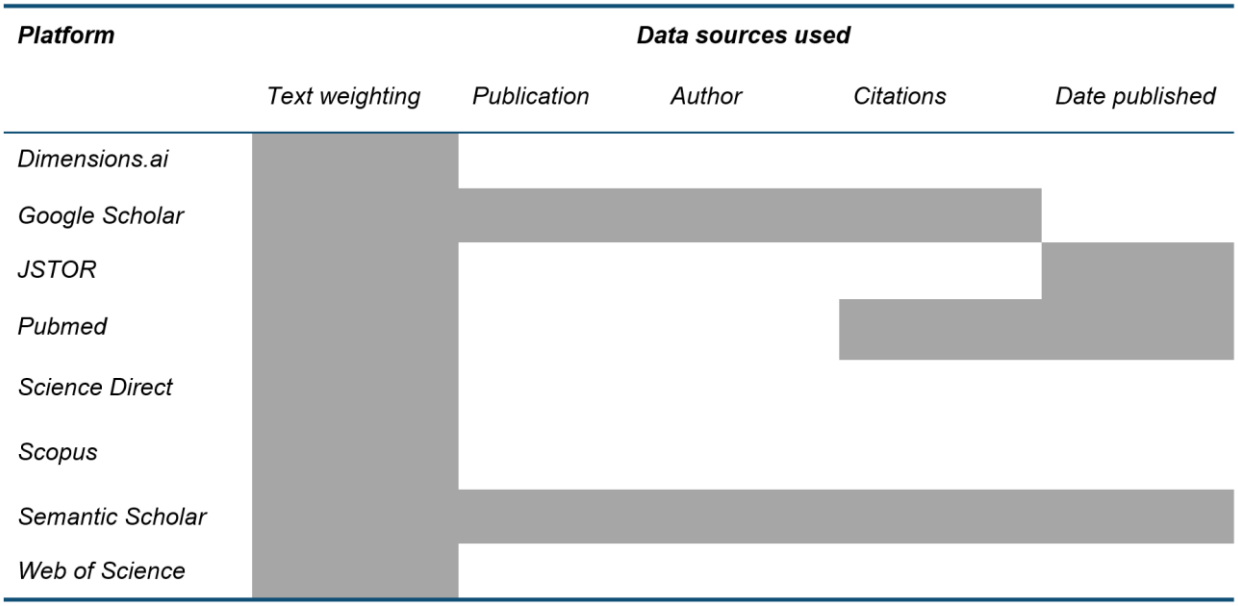

Figure 1: *Mapping of types of data used in definitions of ‘relevance’.*

RQ2: Assumptions about relevance

RQ2 is the core issue we sought to address through the project, and was answered primarily through the survey and supported by the interview data.

In the survey, 85 participants completed the section about benefits, constraints and how results are ranked on up to three platforms of their choice. In total, 63 unique platforms or other online sources were reported by at least one participant. However, the distribution of responses was steeply unequal and included a large proportion of specialist or institution-specific platforms, and only nine had a frequency of five or more, with the most frequent platform by far being Google Scholar. Focusing on Google Scholar, from the first round of categorisation of responses to the question ‘*Thinking about the last time you used [platform] for a literature search, how do you think the platform defines the order in which search results are ranked?*’, it was clear that a wider range of assumptions were held by the participants. A first round of coding yielded 15 unique codes (excluding responses which were off topic, or responses which simply noted that they choose a different ranking

- e.g. by date), codes were combined into nine similar categories, and grouped according to four more general themes (Figure 2).

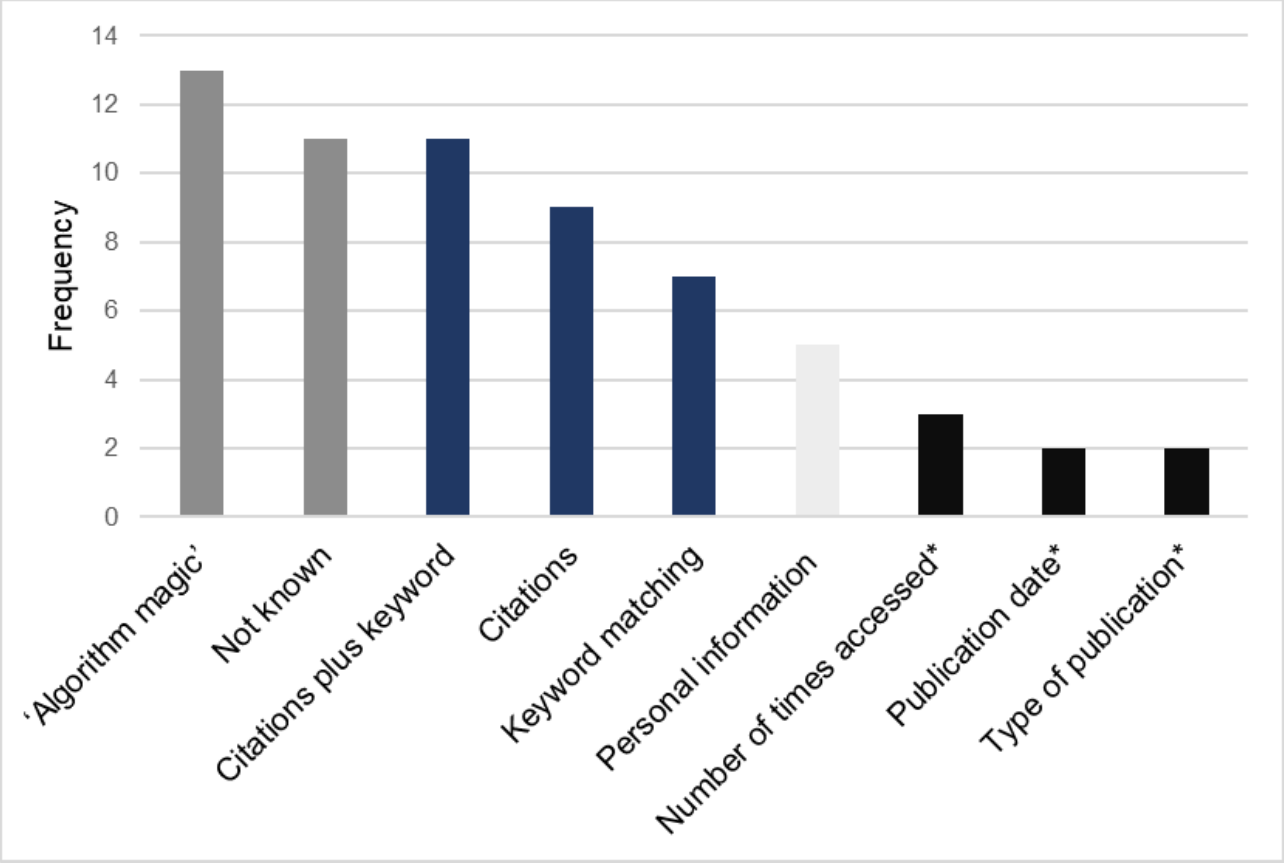

Figure 2: *Frequency of coding in the responses to how Google Scholar ranks search results. Note that in some instances, one response may have been included in more than one category (most often the categories marked with an asterisk). Bars are colour-coded to denote the four themes: dark grey, not defined; blue, citations and/or keywords; light grey, personal information; and black, publication information.*

Each of the four main themes will now be discussed in turn, including discussion of the categories which contributed to each theme. Three illustrative examples from the text responses will be shown for each category as appropriate.

Not defined

The first theme - 'not defined', shown in grey in Figure 2 - was the single most frequent theme, with 20 unique responses contributing to it (36%). Three of the twelve interview participants also related to this theme in their discussion about Google Scholar. This theme can be simply defined as instances where the participants reported not knowing how the ranking functions, although it comprised two slightly different reasons why. It included responses where either participants stated that they do not know ('not known'), or that they acknowledged that it is ranked by relevance or even explicitly that it is an algorithm (which itself is not known; 'algorithm magic' as one participant put it).

> *"I think this is a question we would [all] like answered. It's a total black box."*
>
> *"Algorithm magic."*

"Searches in Google Scholar [are] totally influenced by algorithms."

Citations and/or keywords

The second theme reflects that there were two factors which were identified with particular frequency in the text responses: keyword matching and the number of citations, or a combination of the two. 29 participants' responses contributed to this theme, and accounted for the largest proportion of responses. This was also the theme which was discussed most frequently in relation to Google Scholar during the interviews.

Even in terms of keyword matching, there is variation in terms of the specific text this applies to. The importance of citations in this theme aligns with previous studies which have identified citations as the most important single factor for ranking in Google Scholar (Rovira et al., 2018).

"By how the words that you are looking for appear in the title [of] the article, book or chapter. Secondly, by how those words appear in the content of those articles, books or chapters."

"I think by default the most cited/influential papers come out first."

"appearance of search terms in title/abstract; number of citations"

Personal information

The third and fourth themes occurred less frequently, and were often mentioned in conjunction with keywords and/or citations. The theme of personal information was mentioned in five of the survey responses, so about 10% of cases, and one of the interviewees. The use of personal information is something which isn't explicitly included in the definition provided by Google, but there is a distinct perception that it may be playing a role. Examples of personal information included previous search history, location of the user, and whether or not they are signed in to Google Scholar at the time.

"I think results are determined by my other google searches."

"I've often wondered that. I think its searches give different results depending on whether I'm logged in. I've got a curated Google Scholar profile and I think it uses that to help filter when I'm signed in."

"Based on popularity I guess. May also be affected by my use of VPN and which location I have given."

Publication information

The final theme - publication information - was also often used in conjunction with search terms and citations. This theme was the least frequently used, with only nine survey responses and three interviewees discussing it, but it is distinct from the more general theme of citations and search terms. The theme includes publication metadata such as date published, the type of publication the article is, or how frequently the article has been accessed. Interestingly, particular journals or their metrics such as impact factors did not come up. There were some contradictory beliefs surfaced; for examples, there are contrasting examples here about Google Books:

"Possibly based on how often the articles are accessed."

"Seems to favour citations, would seem to rely on metadata more than abstracts, favours Google Books"

"It almost never puts a Google Book at the top of the search ranking, although that's often what I'm looking for."

RQ3: Variation by platform and implications

While the first two research questions focused on how relevance is defined, RQ3 focuses on what this means in practice. This question was addressed through the interview data and discussion of 'live' search results. The findings here focus on Google Scholar as it was the only platform consistently used by the interview participants. The discussions yielded rich and valuable insights from the participants' perspectives. While it is not possible to represent this level of detail fully here, we focus on the themes which were common across the body of interviews, and draw upon examples in the participants' own words to illustrate the issues.

A core theme common throughout the interviews was the tension between highly ranked results being too broad or reflecting established viewpoints, while they would be more interested in seeking new or cutting-edge perspectives on their specialist research topics. The ranking was generally perceived to provide a good starting point for someone new to a topic. However, it is often perceived to favour broader, foundational information and tends not to reflect recent or contested advances and current debates. Given the importance of citation counts in the Google Scholar ranking, newly-published articles will initially be at a disadvantage.

> "They're the ones you might [find] classically. Like on some reading list, [...] It's a pretty secure list of articles there from recognisable journals with one exception, slightly dated. There's only I think one within the last 10 years. [...] It's giving a good background, but if a student was writing an essay for me, based on these alone, I would write in the comments. You need something a bit more current." - Participant I

However, relevance can still be perceived to be quite variable despite high ranking. The weighting toward citation counts may also skew interdisciplinary queries towards different fields' views on the topic, which may be related to documented differences in the

frequency of citations in different fields (Crespo et al., 2012). High relevance ranking may not be perceived to align with academics' views of quality or rigour.

> "It doesn't tell you what the quality of the work is [...] sometimes it tells you how many people cite it. [...] My background is in Social Sciences and Humanities, but somehow I feel like, if somebody is, say, we're doing interdisciplinary work and if somebody from the Sciences or Engineering backgrounds and published on a similar topic, and I think the articles tend to be ranked higher" - Participant B

> "I can see that the ranking is OK, but you have to look at the results carefully. [...] One of the disadvantages is getting too many results which are not related to the topic. I can't see a clear relevance in the listing of results." - Participant K

The inclusion of a wider range of sources - such as grey literature - is both an advantage and a limitation. The lack of flexibility in terms of filtering or organising search results makes it more difficult to customise results and control for factors in relation to perceived quality. While the reasons why an individual may be cautious about using Google Scholar may vary, the use of multiple platforms was commonplace in order to compensate for the limitations or opacity of any one platform.

> "One of the things that always makes me a little bit wary, when I'm using Google Scholar [...] I don't know how Google decides which results to surface here in which not so that's one of the things that as I say, makes me a bit cautious, when I'm using Google Scholar and why it's never the only source that I would use. [...] Because it's finding things from kind of open access sources, it's finding maybe things from university print sites or you know people's PhD theses or other kind of grey literature sources as well. If I'm just looking for the kind of peer-reviewed public published literature on the topic, then Google Scholar probably isn't best for that. […] I make up for the little limitations by using other platforms as well. So I don't use Google Scholar in isolation, you know, it would never be the only source I would use." - Participant E

RQ4: Filter bubbles

We addressed RQ4 through the interviews, by asking all of the participants to run the same search query and observing whether there were any noticeable differences in the order in which search results were presented. Given that Google Scholar was the only platform which was consistently used as an example by interview participants and one of the most opaque in terms of how results are ranked by 'relevance' (RQ1), it is a useful example to focus upon.

The ranking of the first ten articles (the first page of results) in response to a search for the term "climate change", are shown in Figure 3. Each column represents one interviewee, and arranged in chronological order according to interview date. Overall, the visualisation of data in Figure 3 is inconclusive in terms of providing evidence of filter bubbles. However, the data do demonstrate variation in results, and clearly show changes over time.

Ranking	22-Aug	05-Sept	08-Sept	09-Sept	13-Sept	15-Sept	15-Sept	23-Sept	05-Oct	06-Oct	24-Oct	27-Oct
1st	1	1	1	1	1	1	1	1	1	1	1	1
2nd	2	2	2	2	2	2	2	2	2	2	2	2
3rd	3	3	3	3	3	3	3	4	4	8	4	4
4th	4	4	5	4	4	4	4	5	5	5	5	5
5th	5	5	4	5	5	5	5	6	6	11	6	6
6th	6	6	6	6	6	6	6	7	7	4	7	7
7th	7	7	7	7	7	7	7	8	8	13	8	8
8th	8	8	8	8	8	8	8	11	11	7	11	11
9th	9	9	11	9	9	11	11	10	10	12	10	10
10th	10	10	10	10	10	10	10	12	12	14	12	15

Key					
1	Change et al. (2006) Avoiding dangerous climate change	6	Berrang-Ford et al. (2011) Are we adapting to climate change?	11	Taylor et al. (2013) Ground water and climate change
2	Change (2007) Climate change 2007: The physical science basis	7	Change (2014) Mitigation of climate change	12	Change I.P.O.C. (2001) Climate change 2007: Impacts, adaptation and vulnerability
3	McCarthy et al. (2001) Climate change 2001	8	Thomas (2010) Climate, climate change and range boundaries	13	Yusoff & Gabrys (2011) Climate change and the imagination
4	Houghton (1992) Climate change 1992	9	Chapman (2007) Transport and climate change: a review	14	Metz et al. (2007) Climate change 2007-mitigation of climate change
5	Bernstein (2013) Climate change 2007	10	Thuiller (2007) Climate change and the ecologist	15	Garnaut (2008) The Garnaut Linate change review

Figure 3: *Ranking of first page results in response to a search for the term "climate change" in Google Scholar.*

Conclusions and recommendations

Ranking by relevance is now widespread. Although Google Scholar is arguably the best known example and was the starting point for the inquiry, it is not unique to Google Scholar and all platforms surveyed used some kind of ranking 'by relevance'. For the majority of platforms (all but two), sorting 'by relevance' was the default setting for the presentation of search results.

However, definitions of relevance are not always provided, and depending on the types of information included, may carry a risk of bias. Although use of Google Scholar is now ubiquitous, a range of assumptions are held about how it works and many academics view the ranking as a black box. It is notable too that while the opaque nature of the algorithm was acknowledged by many users in relation to Google Scholar, this did reflect an awareness that an algorithm is being used in this case. 'Algorithm magic' was not mentioned in relation to other platforms, although ranking 'by relevance' is no longer limited to Google Scholar.

Although awareness of the technical risks of relevance ranking varied, in practice, many participants exercise caution in relation to relying on online platforms. Multi-platform use is a frequently adopted strategy in order to compensate for the questionable reliability of any one platform. This may include the use of other online sources such as social networking sites, as a way to look for cutting-edge work which is yet to be registered in databases.

Recommendations for Higher Education institutions and academics:

There is a need for raising awareness among academics and research students about the extent and risks of using 'ranking by relevance' when carrying out literature searches online. There is a particular methodological risk to systematic reviews or other forms of literature-focused research, of introducing bias or gaining an inaccurate view of a field if relying on resources deemed to be the 'most relevant' and not screening all eligible results. This issue is also more prominent at the moment, due to renewed recent interest in rapid reviews as a research approach (Wollscheid & Tripney, 2021) – which could be particularly vulnerable, due to the shortened timelines involved.

Staff development activities would be appropriate for research students and staff at all levels, as the use of relevance ranking is relatively new. The risks of bias should be highlighted, and practical tips shared such as equitable citation practices (e.g. Dworkin et al., 2020). Support may also benefit from being designed to account for different disciplinary perspectives, as the sources and online networked communities playing a role in multi-platform use may vary.

Recommendations for platforms and technical development:

Greater transparency on the part of platforms would be highly recommended. Brief definitions should be provided as a minimum, and this information should be readily available (e.g. as a link for further information, next to the option to sort by relevance). Platforms should critically consider whether the factors used in their relevance ranking risk compounding biases in academic publishing. There would also be scope for a radical rethinking of ranking algorithms, to potentially actively compensate for the biases in citation practices.

References

Baker, S.E., & Edwards, R. (2012). *How many qualitative interviews is enough*. NCRM. https://eprints.ncrm.ac.uk/id/eprint/2273/

Beel, J., & Gipp, B. (2009a). Google Scholar's ranking algorithm: An introductory overview. In: B. Larsen and J. Leta (Eds.), Proceedings of *ISSI 2009: 12th International Conference on Scientometrics and Informetrics*, International Society for Scientometrics and Informetrics, Rio de Janeiro (Brazil), July 2009. https://docear.org/papers/Google%20Scholar's%20Ranking%20Algorithm%20--%20An%20Introductory%20Overview%20--%20preprint.pdf

Beel, J., & Gipp, B. (2009b). Google Scholar's ranking algorithm: The impact of articles' age (an empirical study). *2009 Sixth International Conference on Information Technology: New Generations*, Las Vegas, NV, USA, April 2009, 160–164. https://doi.org/10.1109/ITNG.2009.317

Beel, J. & Gipp, B. (2009c) Google Scholar's ranking algorithm: The impact of citation counts (an empirical study). In: A. Flory and M. Collard (Eds.), *Proceedings of the 3rd IEEE International Conference on Research Challenges in Information Science* (RCIS'09) (Morocco), April 2009. IEEE. http://doi.org/10.1109/RCIS.2009.5089308

Beel, J., Gipp, B., & Wilde, E. (2010). Academic search engine optimization (ASEO). *Journal of Scholarly Publishing*, *41*(2), Article 2. https://doi.org/10.3138/jsp.41.2.176

BERA. (2018). *BERA ethical guidelines for educational research* (4th ed.). British Educational Research Association. https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-2018-online

Büttcher, S., Clarke, C.L.A. & Cormack, G.V. (2016) *Information retrieval: Implementing and evaluating search engines*. MIT Press.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE.

Crespo, J.A., Yunrong, L. & Ruiz-Castillo, J. (2012) Differences in citation impact across scientific fields, UC3M Working papers. Economics we1206, Universidad Carlos III de Madrid. Departamento de Economía.

Creswell, J.W., & Plano Clark, V.L. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE.

Czerniewicz, L. (2016) Knowledge inequalities: A marginal view of the digital landscape. Keynote presentation at Open Repositories Conference 2016, Dublin, Ireland, 14 June 2016. https://www.slideshare.net/laura_Cz/laura-czerniewicz-open-repositories-conference-2016-dublin

DeVito, M.A., Gergle, D., & Birnholtz, J. (2017). "Algorithms ruin everything": #RIPTwitter, folk theories, and resistance to algorithmic change in social media.

Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 3163–3174. https://doi.org/10.1145/3025453.3025659

Dworkin, J., Zurn, P., & Bassett, D.S. (2020). (In)citing action to realize an equitable future. *Neuron*, *106*(6), 890–894. https://doi.org/10.1016/j.neuron.2020.05.011

Fiorini, N., Canese, K., Starchenko, G., Kireev, E., Kim, W., Miller, V., Osipov, M., Kholodov, M., Ismagilov, R., Mohan, S., Ostell, J., & Lu, Z. (2018). Best Match: New relevance search for PubMed. *PLoS Biology*, *16*(8), e2005343. https://doi.org/10.1371/journal.pbio.2005343

Jordan, K. (2022). 'Sort by relevance' – whose relevance? A critical examination of algorithm-mediated academic literature searches. Paper presented at *AoIR 2022: The 23rd Annual Conference of the Association of Internet Researchers*. Dublin, Ireland.

Jordan, K. & Tsai, S.P. (2023a) 'Sort by relevance' project: Bibliography. Zenodo. http://doi.org/10.5281/zenodo.7328625

Jordan, K. & Tsai, S.P. (2023b) 'Sort by relevance' project: Interview protocol. Zenodo. http://doi.org/10.5281/zenodo.7317192

Jordan, K. & Tsai, S.P. (2023c) 'Sort by relevance' project: Survey structure. Zenodo. http://doi.org/10.5281/zenodo.7317201

Kim, S.A. (2017). Social media algorithms: Why you see what you see. *Georgetown Law Technology Review*. https://georgetownlawtechreview.org/social-media-algorithms-why-you-see-what-you-see/GLTR-12-2017/

Klug, D., Qin, Y., Evans, M., & Kaufman, G. (2021). Trick and please: A mixed-method study on user assumptions about the TikTok algorithm. *13th ACM Web Science Conference 2021*, 84–92. https://doi.org/10.1145/3447535.3462512

König, R. & Rasch, M. (2014) *Society of the Query reader: Reflections on web search*. Institute of Network Cultures.

Kossow, N., Windwehr, S., & Jenkins, M. (2021). *Algorithmic transparency and accountability*. Transparency International. https://knowledgehub.transparency.org/assets/uploads/kproducts/Algorithmic-Transparency_2021.pdf

Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C.R. (2013) Bibliometrics: Global gender disparities in science. *Nature* 504, 211–213.

Matthews, D. (2021). Will a Facebook-style news feed aid discovery or destroy serendipity? *Times Higher Education (THE)*. https://www.timeshighereducation.com/features/will-facebook-style-news-feed-aid-discovery-or-destroy-serendipity

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis* (2nd ed.). SAGE.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the Web.* Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/

Pariser, E. (2011). *The Filter Bubble: How the new personalized web is changing what we read and how we think*. Penguin.

Rovira, C., Codina, L., & Lopezosa, C. (2021). Language bias in the Google Scholar ranking algorithm. *Future Internet*, *13*(2), Article 2. https://doi.org/10.3390/fi13020031

Rovira, C., Guerrero-Solé, F., & Codina, L. (2018). Received citations as a main SEO factor of Google Scholar results ranking. *El Profesional de La Información*, *27*(3), Article 3. https://doi.org/10.3145/epi.2018.may.09

Rovira, C., Codina, L., Guerrero-Solé, F. & Lopezosa, C. (2019). Ranking by relevance and citation counts, a comparative study: Google Scholar, Microsoft Academic, WoS and Scopus. *Future Internet*, *11*(9), Article 9. https://doi.org/10.3390/fi11090202

Van Noorden, R. (2014). Google Scholar pioneer on search engine's future. *Nature*. https://doi.org/10.1038/nature.2014.16269

Wengraf, T. (2001). *Qualitative research interviewing*. SAGE.

Wollscheid, S. & Tripney, J. (2021) Rapid reviews as an emerging approach to evidence synthesis in education. *London Review of Education*, 19(1). http://doi.org/10.14324/LRE.19.1.32

Yu, K., Mustapha, N., & Oozeer, N. (2017). Google Scholar's filter bubble: An inflated actuality? In *Esposito, A. (Ed.) Research 2.0 and the impact of digital technologies on scholarly inquiry*. IGI Global. pp. 211-229.