Received: 10 April 2024 | Revised: 12 August 2024 | Accepted: 30 September 2024

DOI: 10.1002/asi.24959

RESEARCH ARTICLE

JASIST WILEY

The role of online search platforms in scientific diffusion

Kyriakos Drivas

Department of Economics, University of Piraeus, Piraeus, Greece

Correspondence
Kyriakos Drivas, Department of Economics, University of Piraeus, Piraeus, Greece.
Email: drivas@unipi.gr

Funding information
Hellenic Foundation for Research and Innovation, Grant/Award Number: 2988

Abstract

After the launch of Google Scholar older papers experienced an increase in their citations, a finding consistent with a reduction in search costs and introduction of ranking algorithms. I employ this observation to examine how recombination of science takes place in the era of online search platforms. The findings show that as papers become more discoverable, their knowledge is diffused beyond their own broad field. Results are mixed when examining knowledge diffusion within the same field. The results contribute to the ongoing debate of narrowing of science. While there might a general reduction in recombination of knowledge across distant fields over the last decades, online search platforms are not the culprits.

1 | INTRODUCTION

The birth and rise of digital technologies have had a profound impact on information. Goldfarb and Tucker (2019) highlighted that such technologies reduce the search costs of finding information and the acquisition costs of accessing it. This technological wave has not left the process of scientific research unaffected. Since the 1990s, the scientific community saw the launch of many influential online search platforms where researchers could find works without incurring the time and effort associated with physical search (Detmer & Shortliffe, 1997). As a result, online visibility of articles resulted in their unprecedent diffusion. In an early study, Lawrence (2001) found that online articles from proceedings would receive 336% more citations than articles from proceedings that were only available in print.

This facilitation of the search process however could be a double-edged sword. In addition to reducing search costs, online search platforms provide rankings of search results which in turn inevitably favor certain works at the expense of others.[1] More importantly, via a physical search across library aisles researchers would be forced to go over older and broader knowledge. In contrast, the exceedingly cheaper, in terms of time and effort, online search will direct researchers to contemporary and mainstream literature making more out-of-the-box search relatively costly (Evans, 2008).

The concern of narrowing science is supported by a general trend of papers becoming less disruptive (Park et al., 2023). While this trend is under scrutiny (Petersen et al., 2023), the overall concern with respect to online search platforms still remains unanswered. In other words, are works that become discoverable more likely to be used in their own narrow field in contrast to distant fields?

To answer the above question, the core objective of the paper is outlined. That is, I examine the impact of increased visibility on the depth and scope of recombination of knowledge. In doing so, I first argue that older papers were favorited by the launch of Google Scholar (GS) as they became more discoverable. The mechanisms involve (i) a reduction in search costs and (ii) the implementation of ranking algorithms placing such older works in visible positions.

First, I establish that older, published between 1950 and 1969, papers became more discoverable and therefore more cited in comparison to more recent papers after carefully matching pre-citation patterns between the two groups. Second, I turn to the core objective of the paper and examine whether this increase in citations came from the papers' own or distant fields.

It is argued that for any given older paper, researchers in distant fields are more likely to discover it than researchers in the same field as the latter may already be aware of this knowledge. Further, the reduction in search costs plays an additional positive role in using distant knowledge. Therefore, the net benefits and consequently incentives to cite distant knowledge increase. Indeed, results show that papers, that became more discoverable, to disproportionately receive more citations from other fields. Such a result is robust when providing a narrower definition of a science field.

With respect to own fields, a potential effect is more nuanced. If one were to hold the overall volume of scholarly work constant, then one would expect that increasing citedness by distant fields would displace science from the same field. However, as research output grows such a displacement may be not applicable. Indeed, when focusing on citations from the same field, results are mixed. While I find a strong positive effect on citations in the same broad field, that effect turns insignificant and negative when narrowing the definition of the science field.

The paper is related to the literature of how digital availability, as opposed to physical availability, influences scientific diffusion. The early study by Evans (2008) showed that online availability can have detrimental effects in making older knowledge less relevant, and its diffusion will be impeded. Further, research works would tend to be cited within their own narrow fields thereby reducing recombination of unrelated knowledge bits. McCabe and Snyder (2015) examined a platform that functions as an article repository rather than a search engine; that is, JSTOR. They challenged Evans' premise that digital availability reduces impact of older knowledge. They showed that the inclusion of older works in JSTOR boosted the citations of those works. We complement this literature by examining a reduction in search cost rather access cost where the previous studies focus on. In doing so, we take into account the current elephant in the room: that is, GS. As McCabe and Snyder (2015) conclude that technologies such as "Google Scholar… promise to continue making measurable contributions to scholarly productivity" (p. 163). GS came online in 2004 and very quickly became a staple in a researcher's literature search. The coverage of information, while not as clear when compared with other search platforms, is considered to be extensive covering an immense scope of scholarly and non-scholarly work and an unprecedent, in terms of years, coverage of articles (Gusenbauer, 2019). For instance, a simple search by years reveals that GS has indexed more than 200,000 works prior to 1800.

Given that the paper studies how citation patterns change within and beyond fields, the findings are also relevant to the vast literature that examines interdisciplinary research. Interdisciplinarity can provide new approaches to solutions and research avenues while it can address complex societal challenges (Barry et al., 2008; Hollingsworth & Hollingsworth, 2000; Larivière et al., 2015; Page, 2008; Rafols et al., 2012). To this end, it is not surprising that research has focused on the factors that promote or hinder interdisciplinarity. Jones (2009) poses that the overwhelming accumulation of knowledge will drive researchers to become more specialized. This will potentially result to less broad and interdisciplinary science produced given the knowledge burden that needs to be lifted by researchers across many fields. In addition to the evolution of science, researchers have also examined factors that are related to evaluation criteria of research (Rafols et al., 2012), the team composition (Wu et al., 2019), field heterogeneity (Yan, 2014), organizational and social dimensions (Lélé & Norgaard, 2005), and the researcher's tradeoff between productivity and curiosity (Palmer, 1999). This paper contributes to this literature by examining whether the introduction of online search platforms changed citation patterns across fields. While there may be other revolutionary technologies in several disciplines of science, online search platforms cut through the universe of science reducing drastically the costs of search. The question of whether such a technology hinders or promotes recombination of distant bits of knowledge is the objective of this paper.

2 | LITERATURE REVIEW

2.1 | A brief history

During the 1990s, with the rise of the internet, several academic search engines launched to meet the demand by researchers for easier and more comprehensive literature search. Perhaps, the two most notable were the Web Of Science (WOS), launched in 1997, and PubMed launched the year before. WOS was the search platform created by the Institute for Scientific Information which encompassed the well-known Journal Citation reports. These included the Science Citation Index (SCI), the Social Science Citation Index (SSCI) and finally the Arts & Humanities Citation Index (AHCI). PubMed on the other hand stemmed from a US government initiative originating in the early 1970s with the goal of listing medicine literature (Falagas et al., 2008); the coverage however has expanded since then. In addition, to the scope of the two search engines, a major difference was that PubMed was free while WOS was subscription based. Further, there was no single subscription as different models would alter the coverage years of publications.[2]

The year 2004 was a turning point in literature search. Elsevier launched Scopus while Google launched GS. Scopus stemmed from Elsevier but was covering works from other publishers as well. Similar to WOS, Scopus was subscription-based while GS was free. In addition to GS being a search engine, it also made an effort to add to the search result links that would direct the reader to a free version of the paper (Butler, 2004). Scopus initially had limited year coverage (Giles, 2004). GS on the other hand, to this day, does not disclose its coverage prompting criticism and significant amount of work comparing the coverage of the aforementioned search platforms (Delgado López-Cózar et al., 2014). It is indicative that via a simple search in Scopus, we located more than a 100 articles where in the title, the keyword of GS appeared at least with one of the following words; "PubMed," "WOS," or "Scopus."

Several scholars initially were critical of GS coverage (Jacsó, 2005; Falagas et al., 2008) while others were finding that GS had better coverage in subject fields (Levine-Clark & Gil, 2008; Walters, 2007). While more recent works have found that the coverage has grown substantially and perhaps overshadows its competitors (Halevi et al., 2017; Orduña-Malea et al., 2015) researchers still raise concerns over the appropriateness of GS as a primary search tool for systematic reviews (Gusenbauer & Haddaway, 2020).

Regardless of the above debate, the penetration of GS was rapid. For a journal such as Nature, by December of 2005 GS was already directing "more online traffic to Nature websites than any other multidisciplinary science search engine" (Giles, 2005). The widespread adoption of GS has been documented in additional studies (Van Noorden, 2014).

2.2 | Discoverability and scientific diffusion after GS

2.2.1 | Hypothesis 1

This rapid rise of GS could have benefited papers that to date were less discoverable. There are two reasons to expect that older papers on average benefited more than recent papers. First, recent literature would be readily available to the literature published at the time. For instance, if a researcher was reading a paper from the 1990s or early 2000s, s/he will be very likely to review references from the 1980s. This argument finds further support once we consider the citation lags of academic papers. The citations they receive follow an inverse U-shaped relationship with respect to the years after publication. This curve is right skewed indicating that the bulk of citations a paper receives takes place within the first years from publication (Seglen, 1992). Therefore, discoverability for recent papers would not apply thereby rendering any benefits minimal. On the opposite side, older papers are more difficult to already be known to researchers conducting literature review. Therefore, discoverability applies to such papers that will likely be translated to their increased citedness.

The second reason relates to rankings. GS to this day, does not disclose how it ranks results. While research has tried to examine the ranking algorithm (Beel & Gipp, 2009), a clear-cut answer remains elusive. Generally, GS's ranking algorithm is likely using a combination of relevance and existing ranking which may include clicks, views, and citations.[3] Therefore, with such rankings in place it could be the case that for certain keyword searches, older papers would be placed in the top 10 search results list. Indeed, Giustini and Barsky (2005) in an early evaluation of GS were concerned with how it ranks results concluding that: "GS's PageRank algorithm makes a calculated guess at what it believes is scholarly and lists articles by how relevant and popular they are—not how current."

To further corroborate this argument, I identified 2000 generic keywords that are likely to disclose fields of knowledge. These keywords are basically research concepts as defined by OpenAlex (Priem et al., 2022) and are discussed later in the paper. The full list of keywords is available upon request. Each of these keywords was searched in GS during February through March of 2023. Then I counted the publication years of the 10 results of the first page. Figure A1 shows that approximately 15.7% are published before 1990—i.e., papers 33 years and older. While I cannot perform a similar analysis as if we were in 2004, a safe prediction can be made that GS back then also ranked several results that were older thereby enhancing their discoverability and citedness.

The above discussion points to a favorable outcome of older papers vis-à-vis their recent counterparts. Nevertheless, it is useful to consider a potential counterargument. With the breadth of information available at low access costs, researchers could then more easily focus on mainstream and more recent science (Evans, 2008). In other words, while older science may become more visible, the breadth of recent and mainstream science has also become easier to navigate.

While the above may be true, it does not negate the increased visibility of older science which can be relevant even decades after its inception. Therefore, the first hypothesis of the paper can be stated as follows:

> **Hypothesis 1.** Older papers will receive more citations than expected after GS is launched.

2.2.2 | Hypothesis 2

When knowledge becomes more discoverable the natural next question is how it will be used. For science perhaps the most important question relates to whether the knowledge will be employed in the same or different fields to the one is incepted.

To state the hypothesis, it is important to provide a framework over searching for knowledge. Building on theories of searching for knowledge both within fields and across distant ones (Cyert & March, 1963; Nelson & Winter, 1982), Schilling and Green (2011) made an interesting distinction by referring to search scope and search depth. Search scope refers to a process where the researcher will look across fields to combine distant bits of knowledge to generate new knowledge. In contrast, in search depth, the researcher will dwell into a specific field and seek to combine bits of knowledge that are related. When operationalizing these two concepts, high search scope refers to papers that cite literature that is not in their own field while search depth refers to papers that cite literature within their own field. To this end, a paper can have both high search scope and depth. Indeed, in their study, they found that both types of searches can deliver research output of high impact.

A paper that is most closely related to this study, but as Schilling and Green (2011) employ backward citations, is by Evans (2008). He argued that the rise of the internet and online availability will divert researchers to prevailing opinion. Contrary to a physical search, where researchers would be forced to browse through older and perhaps less relevant knowledge, the more efficient online access makes it easier for researchers to locate the knowledge specific to their field.

While the setup by Evans (2008) contributes to the role of access costs, this paper focuses on search costs as it examines the launch of GS. While access costs could reduce the use of knowledge beyond their own fields, search costs that increase discoverability may have the opposite effect via two mechanisms. First, increased discoverability, for any given paper, will more likely make researchers in distant fields aware of this work than researchers in the same field as the latter may already be accustomed with this knowledge. Second, let us consider two environments that are different in only one dimension. In the first, search costs are high (offline environment) and in the second they are low (online environment). However, the benefits of employing distant knowledge are the same between both environments. Therefore, in the online environment the net benefit of searching in scope is higher. Thus, one would expect that when moving from an offline to an online environment for search scope to increase ceteris paribus.

Therefore, the second hypothesis of the paper can be stated as follows:

> **Hypothesis 2.** As older papers become more discoverable they will be used by works in different fields than their own more than expected.

This change however begs the question of what the implications may be for search depth. Having assumed ceteris paribus then one would expect for search scope to displace search depth. However, such an assumption is difficult to make given that the number of published works has been increasing steadily over the last decades (Bornmann & Mutz, 2015). To this end, displacement may not take place as long as the pool of published works that may cite prior literature increases. Therefore, it could be the case the search depth may not be impeded conditioned that the number of works published increases from shifting from the offline to the online environment.

3 | DATA AND METHODS

3.1 | Data compilation

I employ data from the recently released OpenAlex dataset (Priem et al., 2022). OpenAlex was created after Microsoft Academic Graph (MAG) was discontinued and is a free dataset that aims to document all scholarly work while recently researchers have already started making use of this resource for bibliometric analysis (Williams et al., 2023). I downloaded a snapshot of this dataset during the summer of 2023.

I collect all the works published between 1950 and 1989 along with all the works that are citing them. I also collect the Concepts that each work is tagged with. There is a large literature that aims to categorize papers based on fields. However, up until the recent past, studies were using journal classifications (Park et al., 2023; Schilling & Green, 2011). Nevertheless, a classification problem arises. Papers within a journal can be in distinct fields of

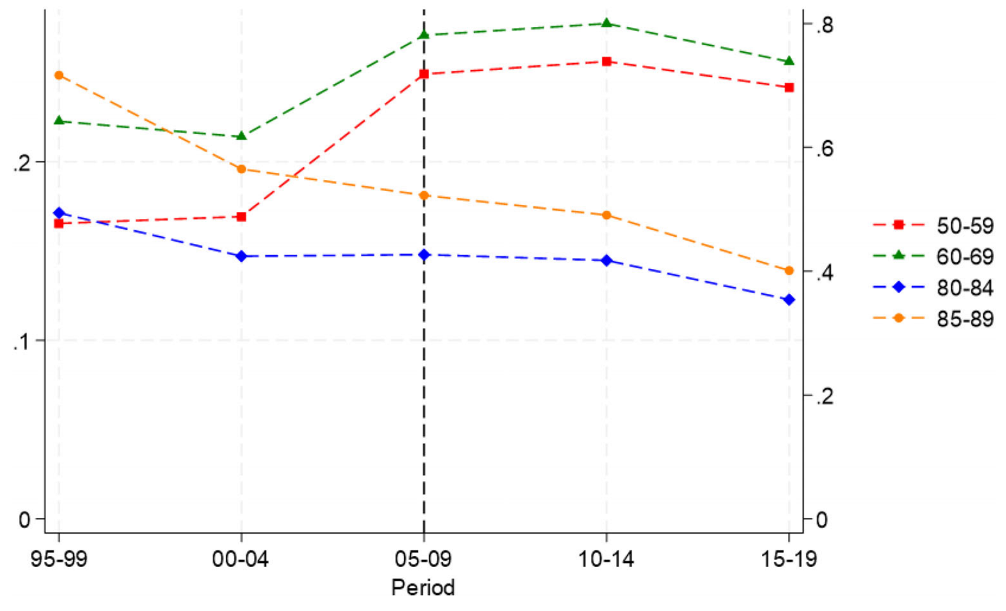

FIGURE 1 Citations by group. Each dashed line shows the average citation rate of papers that received cumulative over 5-year intervals starting from 1995 to 1999 ending to 2015 to 2019.

science as many journals host a variety of fields and sub-fields. Additionally, journals as they evolve encompass additional fields or drop others from their focus (Leydesdorff & Bornmann, 2016). To this end, OpenAlex by building on the work made by MAG and by employing each paper's abstract was able to assign it in one or multiple fields, referred to as Concepts (Shen et al., 2018). What is more, these Concepts are layered in five levels where Level-0 is the broadest encompassing 19 concepts, Level-1 is further disaggregated into 284 concepts while Level-2 is further disaggregated into a little over than 21,000 concepts (Version of the Concept Level Tree: August 17, 2022). Researchers have already been using such taxonomy to either identify papers in specific fields (Murphy et al., 2020), classify groups of papers in scientific fields (Lin et al., 2023) or even classifying papers within the same journal in various fields (Prati et al., 2024).

The core sample is journal articles published between 1950–1969 and 1980–1989. I treat the 50s and 60s papers as early papers that are less discoverable and 80s papers as recent and on average more discoverable. As a first exploration of Hypothesis 1, Figure 1 shows, for the core sample, the average citations received. For early groups (papers published between 1950 and 1969), GS had a positive effect on the citations they received. What is more, there is no significant pre-trend the decade leading up to GS launch. Further, this increase in citations has persisted at least until 2019. On the contrary, the cohorts of papers published between 1980 and 1989 (1980–1984 and 1985–1989) continued the downward trend in their citations received even after GS was launched.

Note that I exclude from the analysis papers published during 1970–1979 to provide a clear-cut comparison between groups. As will be shown below, the empirical design is based on matching between early and recent papers. Therefore, by including the 1970–1979 group it would be possible to end up with incoherent matches (e.g., a paper from 1979 matched with a paper from 1980). Such a comparison would therefore be spurious resulting to misleading interpretation. Figure A2 displays the citation rates for the group of papers published between 1970 and 1979. As can be seen, there is no clear-cut trend after GS.

Four groups are considered: the 50s and 60s papers and the papers published during 1980–1984 (first 80s) and those published during 1985–1989 (second 80s) as two separate groups. The reason for considering two separate groups during the 80s is due to the nature of citation lags which is discussed later in paper.

Further, the focus is on journal articles that disclose at least one Level-1 Concept. To this end I assign each paper to one of the four period groups and in the field based on the Level-1 Concept with the highest score. Approximately 88% of the papers disclose at least one Level-1 Concept. Table A1 shows the data cuts and samples of interest. The sets of papers from each period group along with the associated Level-1 Concept are referred to as a *cohort*. I then examine the citations during the period 1995–2019 to ensure a common period interval for all cohorts. Also, to reduce the dimensionality of the problem and provide better matching between early and recent papers, citations are aggregated across five intervals: 95–99, 00–04, 05–09, 10–14 and 15–19. The first two relate to pre-GS while the next three, post-GS. Finally, I exclude from the baseline analysis, papers which receive zero citations throughout the 1995–2019 period or are located at the top one percentile of their cohort with respect to citations. Robustness checks, presented below, show that such sample drop does not change the results. Overall, the baseline sample is comprised of approximately 4.3 million papers.

3.2 | Testing Hypothesis 1

The matching procedure is based on the Coarsened Exact Matching process proposed by Blackwell et al. (2009) and Iacus et al. (2012). I match each 50–59 cohort paper to a paper from 80–84 period with the same Level-1 Concept based on the citations it receives during 95–99 and 00–04; that is, Cites_{95-99} and Cites_{00-04}. The same process is repeated for each of the 60–69 cohort papers. I then repeat both of the aforementioned matches with the cohort papers from the 85–89 period. Tables A2 and A3 show the citations across periods by cohort pairs both for the unmatched and matched samples. Three observations stand out. First, the early published papers experience a drastic increase in the 05–09 period. This post-GS citation change remains constant up until the 15–19 period signifying a long-term effect. Second, after the matching process, the recently published papers (80–84 and 85–89) experience an increase post-GS. However, this increase is smaller when compared to early published papers. Finally, the matching rate is above 90% indicating an overall satisfying matching of the early published papers.

The diff-in-diffs specification is as follows:

$$\begin{aligned}\text{Cites}_{i,t} = {} & b_0 + b_1\text{Treat}_i + b_2\text{Treat}_i\_x\_00-04_t \\ & + b_3\text{Treat}_i\_x\_05-09_t + b_4\text{Treat}_i\_x\_10 \\ & - 14_t + b_5\text{Treat}_i\_x\_15-19_t + b_6 00-04_t \\ & + b_7 05-09_t + b_8 10-14_t + b_9 10-19_t \\ & + \text{Pair}_{i,j},\end{aligned} \tag{1}$$

where $\text{Cites}_{i,t}$ is the citations paper i receives in period t. Treat_i takes the value of 1 if the paper belongs in the 50–59 or 60–69 cohorts and 0 otherwise. $00-04_t$, $05-09_t$, $10-14_t$ $10-19_t$, are time period dummies. Finally, $\text{Pair}_{i,j}$ is a set of dummies that identifies each early paper with its matched recent paper. I estimate this regression separately for each of the following matched samples; that is, 50–59 with 80–84, 50–59 with 85–89, 60–69 with 80–84 and 60–69 with 85–89. At this point it is useful to discuss the reason separate matching took place for the 80–84 and 85–89 papers with the early cohorts. As research has shown citation patterns usually follow an inverse U-shaped curve with the peak being within the first years from publication (Hall et al., 2007; Mehta et al., 2010). To this end, for recent papers the average pattern would be a natural decline in citations as the unmatched samples of 80–84 and 85–89 of Tables A2 and A3 show. While matching has already isolated papers that resemble the early published papers, examining different age groups of the recent papers would provide robustness that the citation lag is not the driver behind any change found between early and recently published papers post-GS.

The focal interest is on the interaction terms between Treat_i and the period dummies. All interaction terms are compared with the period that is absent: that is, 95–99. One would expect that for the matching to be accurate to observe no visible pre-trend before GS. In other words, we should observe for the coefficient $\text{Treat}_i\_x\_00-04_t$ to be close to zero and statistically insignificant. Next, to validate Hypothesis 1 we expect for at least the sum of the next three interactions to be positive and statistically significant. Such a result would indicate that post-GS early papers experienced a relative increase in citedness compared to recently published papers. In other words, older papers should experience a more than expected increase in citations.

A note with respect to estimation both for Hypothesis 1 and 2 is warranted. The dependent variables across all estimations are count variables. To this end, an ideal estimator would be Poisson. However, given the large number of fixed effects (hundreds of thousand), the computational challenges are immense. Thankfully, recent work has shown that as long as the dependent variable is not transformed, a linear estimator will deliver similar results to a Poisson even if the mass is at zero (Mullahy & Norton, 2023). To this end, all estimations are performed using ordinary least squares (OLS). Finally, note that the standard errors are clustered at the pair level to avoid any serial correlation (Bertrand et al., 2004).

3.3 | Testing Hypothesis 2

The pairs of cohorts where matching takes place remain the same. However, for the purpose of Hypothesis 2 the matching takes place at a finer level. To this end, I match early and recently published papers based on Cites_{95-99} and Cites_{00-04} and citations they receive from papers that do not share any common Level-0 Concept; denote these two variables as $\text{NonSame}L0_{95-99}$ and $\text{NonSame}L0_{00-04}$.

The diff-in-diffs specification is as follows:

$$\begin{aligned}\mathrm{NonSame}L0_{i,t} = {} & b_0 + b_1\mathrm{Treat}_i + b_2\mathrm{Treat}_{i}\_x\_00 \\ & -04_t + b_3\mathrm{Treat}_i\_x\_05-09_t \\ & + b_4\mathrm{Treat}_i\_x\_10-14_t \\ & + b_5\mathrm{Treat}_i\_x\_15-19_t + b_6 00-04_t \\ & + b_7 05-09_t + b_8 10-14_t + b_9 10-19_t \\ & + \mathrm{Pair}_{i,j}.\end{aligned} \tag{2}$$

To validate Hypothesis 2, we expect for the interaction terms post-GS to be positive and statistically significant while the interaction term pre-GS to be close to zero and statistically insignificant. To provide further support for Hypothesis 2, we examine citations at Level-1 Concept and perform a similar matching and regression analysis.

Finally, while no hypothesis was stated for the citation pattern within the same field, I nevertheless perform the same process of matching for all combinations of groups based on Cites_{95-99} and Cites_{00-04}; in this case however, I further match based on the citations they receive from papers that share at least one same Level-0 cohort; denote these two variables as $\mathrm{Same}L0_{95-99}$ and $\mathrm{Same}L0_{00-04}$.

The diff-in-diffs specification is as follows:

$$\begin{aligned}\mathrm{Same}L0_{i,t} = {} & b_0 + b_1\mathrm{Treat}_i + b_2\mathrm{Treat}_i\_x\_00-04_t \\ & + b_3\mathrm{Treat}_i\_x\_05-09_t \\ & + b_4\mathrm{Treat}_i\_x\_10-14_t \\ & + b_5\mathrm{Treat}_i\_x\_15-19_t + b_6 00-04_t \\ & + b_7 05-09_t + b_8 10-14_t + b_9 10-19_t \\ & + \mathrm{Pair}_{i,j}.\end{aligned} \tag{3}$$

The raw and matched samples across all levels are displayed in Tables A4–A7. Overall, for Level-0 Concept (Tables A4 and A5), one can observe a higher increase in citations both from the same and distant field for early papers in comparison to recent papers. While a clear-cut comparison is not straightforward as the levels of citations are not the same for within and outside the field, it is worth highlighting that the change in the citation increase, percentage-wise, is bigger in the case of *Non-Same* citations vis-à-vis *Same* citations. The same also holds when examining citations for Level-1 Concept (Tables A6 and A7). These simple comparisons show that the increase is more pronounced from citations stemming from distant fields regardless the level one focuses on.

4 | RESULTS

4.1 | Hypothesis 1

Table 1 tests Hypothesis 1. Column 1 compares 50–59 with 80–84 papers. The coefficient of the interaction term pre-GS (i.e., *Treat_x_00-04*) is close to zero and statistically insignificant. However, post-GS the coefficients of the interaction terms are positive and statistically significant. We observe the same pattern when comparing 60–69 with 80–84 papers (Column 2), 50–59 with 85–89 papers (Column 3), and 60–69 with 85–89 papers (Column 4). I also display the coefficients of these Columns, along with their 99% confidence intervals in Figure 2. Overall, there is significant support for Hypothesis 1; post-GS older papers experienced a disproportionate increase compared to their matched recent counterparts. What is more, this increase is rather permanent. Even up until recently the citation uptick remains constant highlighting the permanent discoverability of older papers in the digital era.

To provide robustness, I include all citations available in OpenAlex and re-perform the matching procedure. Regression results are displayed in Table A8. Results are qualitatively similar to the results of Table 1. A potential bias with respect to the OpenAlex database could be that older papers are not properly logged and therefore the papers or their citations may be missing. To provide robustness, I include both the group of papers that have received zero citations and the outliers. After performing the same matching procedure for each pair of groups, results are displayed in Table A9. First note that the sample increase is substantial since many older papers receive no citations during the observed period. For instance, for the 50–59 papers the size increased more than 400%. Second, the coefficients reduce in size due to the inclusion of zeros. However, the significance remains the same indicating that even by including the entirety of older and recent papers the change post-GS persists.

As an additional robustness test, I examine the cited half-life of papers by group. The cited half-life in year of analysis t of papers published in year y is calculated as follows. We take the all the citations that papers published in year y up until year t and assign them in an ascending order. The cited half-life is the years it takes for half of the citations to be accumulated to the focal papers. For instance, for the group of papers that were published in 1950 the cited half-life in 1999 is 17.1. This implies that for papers published in 1950 it took them 17.1 years to accrue half of the citations they have garnered by 1999. The cited-half life for the same group in 2009 is 22.13 implying that it took them 22.13 years to accrue half of the citations they have garnered by 2009.

TABLE 1 Testing Hypothesis 1. Baseline results.

	(1)	(2)	(3)	(4)
Treat	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Treat_x_00 – 04	−0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)
Treat_x_05 – 09	0.106*** (0.004)	0.066*** (0.003)	0.131*** (0.004)	0.103*** (0.003)
Treat_x_10 – 14	0.073*** (0.005)	0.052*** (0.004)	0.109*** (0.006)	0.097*** (0.004)
Treat_x_15 – 19	0.143*** (0.006)	0.107*** (0.004)	0.189*** (0.006)	0.163*** (0.004)
00 – 04	0.009*** (0.003)	−0.059*** (0.002)	0.010*** (0.003)	−0.059*** (0.002)
05 – 09	0.292*** (0.004)	0.147*** (0.003)	0.270*** (0.004)	0.104*** (0.003)
10 – 14	0.352*** (0.004)	0.191*** (0.003)	0.322*** (0.004)	0.141*** (0.003)
15 – 19	0.211*** (0.004)	0.035*** (0.003)	0.170*** (0.005)	−0.030*** (0.004)
Constant	0.810*** (0.002)	1.129*** (0.002)	0.827*** (0.002)	1.160*** (0.002)
Observations	3,137,305	7,197,345	3,133,800	7,188,500
R-squared	0.566	0.568	0.580	0.583
adj R-squared	0.518	0.520	0.533	0.537
# of pairs	313,731	719,735	313,381	718,851

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Columns 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The dependent variable is citations made by journal articles. All Columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
***$p < 0.01$.

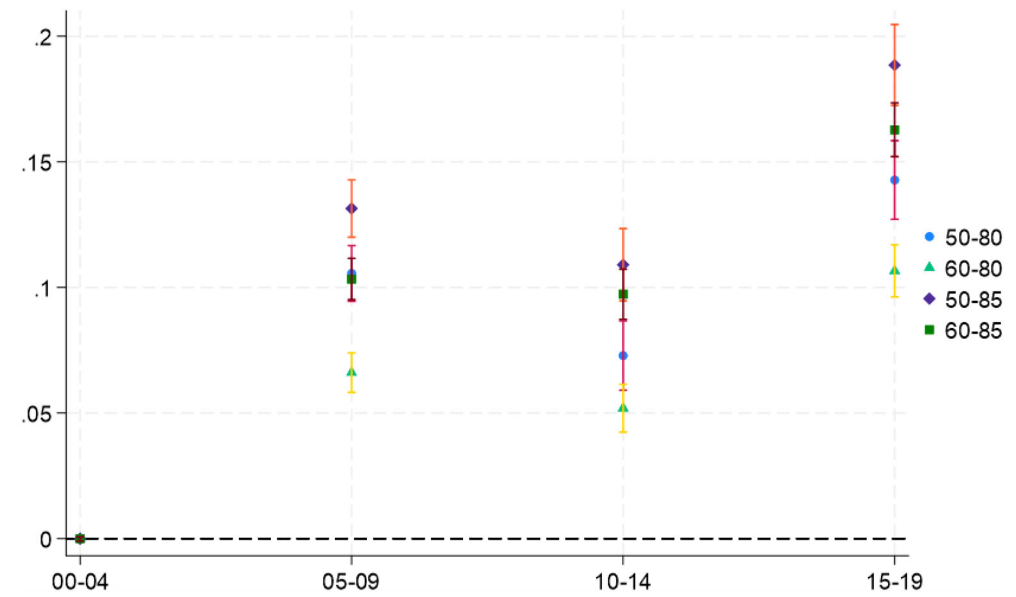

FIGURE 2 Testing Hypothesis 1. Plotting interactions. The dots display the coefficients of Table 1. Specifically, the dots in 00–04 represent the Treat_x_00 – 04 coefficients. The dots in 05–09 represent the Treat_x_05 – 09 coefficients. The dots in 10–14 represent the Treat_x_10 – 14 coefficients. The dots in 15–19 represent the Treat_x_15 – 19 coefficients. The lines display the 99% confidence interval for each of the coefficients. 50–80, 60–80, 50–85 and 60–85 represent the coefficients of the interaction terms and their associated confidence intervals from Column 1, Column 2, Column 3 and Column 4 of Table 1, respectively.

Initially, we observe citations for the 50–59 and 60–69 groups since 1970 in Figures A3 and A4. In Figure A3, we calculate the weighted average of the cited half-life by group. Figure A4 shows the growth of this cited half-life. From these two figures we can observe that the cited half-life increases at a descending rate. However, that

TABLE 2 Testing Hypothesis 2. Examine impact on citations within and beyond the same Level-0 Concept.

	(1) Citations in the same Level-0 Concept	(2)	(3)	(4)	(5) Citations in the different Level-0 Concept	(6)	(7)	(8)
Treat	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Treat_x_00 − 04	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Treat_x_05 − 09	0.062*** (0.004)	0.044*** (0.003)	0.084*** (0.004)	0.074*** (0.003)	0.038*** (0.001)	0.017*** (0.001)	0.042*** (0.001)	0.021*** (0.001)
Treat_x_10 − 14	0.016*** (0.005)	0.024*** (0.003)	0.046*** (0.005)	0.063*** (0.003)	0.037*** (0.001)	0.022*** (0.001)	0.041*** (0.001)	0.026*** (0.001)
Treat_x_15 − 19	0.090*** (0.005)	0.081*** (0.004)	0.128*** (0.006)	0.132*** (0.004)	0.030*** (0.001)	0.016*** (0.001)	0.033*** (0.001)	0.020*** (0.001)
00 − 04	−0.002 (0.002)	−0.062*** (0.002)	−0.002 (0.002)	−0.066*** (0.002)	0.002*** (0.001)	0.001** (0.000)	0.002** (0.001)	−0.000 (0.000)
05 − 09	0.261*** (0.003)	0.134*** (0.002)	0.240*** (0.003)	0.093*** (0.003)	0.029*** (0.001)	0.026*** (0.001)	0.023*** (0.001)	0.019*** (0.001)
10 − 14	0.315*** (0.004)	0.172*** (0.003)	0.289*** (0.004)	0.122*** (0.003)	0.038*** (0.001)	0.033*** (0.001)	0.032*** (0.001)	0.025*** (0.001)
15 − 19	0.200*** (0.004)	0.045*** (0.003)	0.165*** (0.004)	−0.022*** (0.003)	0.020*** (0.001)	0.017*** (0.001)	0.015*** (0.001)	0.010*** (0.001)
Constant	0.671*** (0.002)	0.960*** (0.001)	0.690*** (0.002)	1.009*** (0.002)	0.068*** (0.001)	0.059*** (0.000)	0.071*** (0.001)	0.065*** (0.000)
Observations	3,085,110	7,030,150	3,089,050	7,054,320	3,084,490	7,032,820	3,088,310	7,052,610
R-squared	0.536	0.524	0.547	0.546	0.417	0.419	0.416	0.428
adj R-squared	0.484	0.471	0.496	0.496	0.352	0.354	0.351	0.364
# of pairs	308,511	703,015	308,905	705,432	308,449	703,282	308,831	705,261

Note: For Columns 1–4 the dependent variable is citations by journal articles that share at least one Level-0 Concept as the focal paper. For Columns 5–8 the dependent variable is citations by journal articles that share no Level-0 Concept as the focal paper. Columns 1 and 5 compare matched papers from the 50–59 group with the 80–84 group. Columns 2 and 6 compare matched papers from the 60–69 group with the 80–84 group. Columns 3 and 7 compare matched papers from the 50–59 group with the 85–89 group. Columns 4 and 8 compares matched papers from the 60–69 group with the 85–89 group. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
$^{**}p < 0.05$. $^{***}p < 0.01$.

rate shifts during the 05–09 period, after GS is launched, a finding consistent with the baseline results.

Figures A5 and A6 compare the cited half-lives for all groups for the periods 95–99 through 15–19. As can be seen the cited half-lives of the 80–84 and 85–89 groups are increasing at a declining rate. Contrary to the early published papers however, the rate does not change after the introduction of GS. The above findings show that older papers experienced an abrupt change in their citations after GS; a finding that is not present in recent papers.

4.2 | Hypothesis 2

Table 2 displays the results of the matched samples for specification 2 (Columns 5–8) and specification 3 (Columns 1–4). Throughout all columns, one can observe an increase in citations post-GS. Note that such a result is robust to including all citations (Table A10). This implies that post-GS, both papers from within the broad field of Level-0 Concept and beyond cited the focal papers more. The finding with respect to distant fields provides support for Hypothesis 2 indicating that discoverability helped these papers transcend beyond their own field.

With these findings in mind, I examine the citations that papers receive from the same Level-0 Concept but not the same Level-1 Concept. To this end, I focus on a narrower field of science and examine how citation patterns change therein. Table A11 paints a similar picture as with citations beyond the broad field of science. Overall, I find significant evidence in support of Hypothesis 2. Upon discoverability of older papers, these works

TABLE 3 Examine impact on citations within the same Level-1 Concept.

	(1)	(2)	(3)	(4)
Treat	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Treat_x_00 – 04	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Treat_x_05 – 09	−0.025*** (0.003)	−0.001 (0.002)	−0.018*** (0.003)	0.020*** (0.002)
Treat$_{x_{10}}$ – 14	−0.038*** (0.004)	−0.016*** (0.003)	−0.030*** (0.004)	0.011*** (0.003)
Treat_x_15 – 19	0.020*** (0.004)	0.037*** (0.003)	0.042*** (0.004)	0.073*** (0.003)
00 – 04	−0.008*** (0.002)	−0.059*** (0.002)	−0.008*** (0.002)	−0.064*** (0.002)
05 – 09	0.231*** (0.003)	0.118*** (0.002)	0.222*** (0.003)	0.088*** (0.002)
10 – 14	0.266*** (0.003)	0.145*** (0.002)	0.255*** (0.003)	0.109*** (0.002)
15 – 19	0.186*** (0.003)	0.049*** (0.003)	0.161*** (0.003)	−0.000 (0.003)
Constant	0.474*** (0.002)	0.726*** (0.001)	0.487*** (0.002)	0.763*** (0.001)
Observations	3,054,570	6,890,620	3,060,770	6,940,630
R-squared	0.491	0.493	0.492	0.504
adj R-squared	0.435	0.437	0.435	0.449
# of pairs	305,457	689,062	306,077	694,063

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Column 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The dependent variable is citations by journal articles that share at least one Level-1 Concept as the focal paper. All Columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
$p < 0.05$.*$p < 0.01$.

TABLE 4 Examine impact on citations within the same Level-2 Concept.

	(1)	(2)	(3)	(4)
Treat	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)
Treat_x_00 – 04	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)
Treat_x_05 – 09	−0.045*** (0.003)	−0.014*** (0.002)	−0.040*** (0.003)	−0.004** (0.002)
Treat_x_10 – 14	−0.057*** (0.003)	−0.031*** (0.002)	−0.055*** (0.003)	−0.019*** (0.002)
Treat_x_15 – 19	−0.009*** (0.003)	0.013*** (0.002)	0.001 (0.003)	0.032*** (0.002)
00 – 04	−0.012*** (0.002)	−0.049*** (0.001)	−0.013*** (0.002)	−0.053*** (0.001)
05 – 09	0.176*** (0.002)	0.096*** (0.002)	0.170*** (0.002)	0.077*** (0.002)
10 – 14	0.198*** (0.003)	0.111*** (0.002)	0.194*** (0.003)	0.090*** (0.002)
15 – 19	0.137*** (0.003)	0.039*** (0.002)	0.125*** (0.003)	0.009*** (0.002)
Constant	0.306*** (0.001)	0.485*** (0.001)	0.320*** (0.001)	0.515*** (0.001)
Observations	2,654,170	6,217,720	2,660,960	6,263,000
R-squared	0.418	0.440	0.440	0.452
adj R-squared	0.353	0.377	0.377	0.391
# of searches	265,417	621,772	266,096	626,300

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Column 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The dependent variable is citations by journal articles that share at least one Level-2 Concept as the focal paper. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
$p < 0.05$.*$p < 0.01$.

appear to be re-deployed beyond the broad or narrow field of inception. Borrowing the definitions by Schilling and Green (2011) we observe for search scope to increase due to the online search platforms such as GS.

Framing a hypothesis for search depth was more nuanced given the potentially opposing forces (displacement and increase in published works). Nevertheless, it is useful to examine the overall citation behavior. First, by examining Columns 1–4 of Table 2, we see that citations from within the same broad field to also increase. To provide further insight, I perform analogous matching and regressions at the Level-1 Concept for citations within the field. Results are displayed in Table 3. Here results are mixed. Several of the interaction coefficients post-GS are negative and significant indicating that at this narrower definition of a field search depth for older compared to recent papers does not appear to have a clear-cut trend post-GS. Results are also mixed when including all citations as robustness in Table A12.

Finally, I examine citation changes within the same Level-2 Concept. However, approximately 15% of the sample is dropped as such information is not available for several papers. Results are displayed in Table 4; results with citations by all types of works are displayed in Table A13. Apart from the comparison of the 60–69 with the 80–84 group, the pattern is similar showing that post-GS, citations by works in the same narrow field as the focal older paper reduce compared to citations to the focal recent counterpart. The overall results of the effect of GS on search depth are mixed providing some support for the discussion in section 2.2. Resources may be limited and using distant knowledge is at the expense of using more familiar knowledge. However, over the decades there has been an influx of research output. The overall net effect is therefore harder to pinpoint.

5 | DISCUSSION

Overall, we observe empirical evidence that validate both Hypotheses. In other words, discoverability, initiated by lower search costs, boosted overall citedness of papers and in particular beyond their own field. While an online search platform such as GS has contributed to this change in citation behavior during the last two decades, it may not be the sole driver of the changing patterns of citedness. There can be other factors, intertwined or independent to the rise of online search platforms, that can contribute to the follow-on use of older papers.

A first related factor is how information consumption has changed with the digitization of information and launch of the world wide web (Marchionini, 1995; O'Brien & Toms, 2008). In an early study, Teevan et al. (2004) found that users are likely to forego time-consuming specific searches in favor of less precise and broader searches. This accords to early theories that view users as information foragers (Pirolli & Card, 1999). Through this lens, users are likely to hunt for information beyond their own area of expertise actively seeking new knowledge. While the above discussion points to a similar prediction to Hypothesis 2, the mechanism is different. Nevertheless, this change in information consumption can be attributed, at least to an extent, to the influx of unprecedented information caused by digitization and the world wide web. To this end, both factors, GS and information consumption, can be contributing factors in higher citation rates from distant fields.

A second factor relates to the growing tendency of stakeholders promoting interdisciplinary research as they view it as an effective way of tackling complex problems (National Academy of Sciences, 2005; National Research Council, 2014). There are several cases where universities across the globe have included interdisciplinary research as a criterion for promotion and tenure decisions (Cornell et al., 2013; Klein & Falk-Krzesinski, 2017). For instance, in an early survey of US medical schools Bunton and Mallon (2007) showed that institutions had already started recognizing interdisciplinary research. In terms of funding, evidence points to lower success rates of interdisciplinary research proposals (Bromham et al., 2016). However, in the grant scheme, more funds are directed towards interdisciplinary research in the twenty-first than the twentieth century (Porter et al., 2006). Coupled with the fact that interdisciplinary researchers can attract more funding in the long-term (Sun et al., 2021) we should expect for interdisciplinary research to increase. To this end, we would expect for knowledge to be more likely to be cited beyond its own field.

A third factor relates to the shift from the Mode 1 to Mode 2 knowledge production. Gibbons et al. (1994) in their study urged for research to move away from discipline-based scientific conduct towards problem solving that can encompass society's challenges adopting interdisciplinary and transdisciplinary approaches. In a follow-up study Nowotny et al. (2003) urged the involvement of all stakeholders in an environment where problems can be identified and subsequently solved. Such an approach of knowledge production is likely to encourage interdisciplinary research (Holmwood, 2010) and therefore re-deployment of knowledge in distant fields. This is also evident by the tendency of many researchers shifting topics throughout their careers (Zeng et al., 2019).

6 | CONCLUSION

The launch and rise of online search platforms have made literature review search more efficient thereby reducing the overall costs of engaging in research. There have been concerns however that the more efficient search may promote a narrowing of science as researchers can now more easily find works relevant to them. In other words, the process of randomly finding knowledge that is distant to one's own research is now impeded given the reduction of search costs.

I employ the launch of Google Scholar to first examine how researchers reacted to the knowledge embedded in older knowledge. Results are robust and show that the citations for those papers increased on average. Further, this newfound discoverability of these papers is associated with an increase in citations different than their own fields. With respect to citations within the paper's own field results are mixed.

The results of this study contribute to the role of online availability in the diffusion of science (Evans, 2008; McCabe & Snyder, 2015). While they focused primarily on access costs, we contribute by examining the reduction in search costs pertaining to the elephant in the room—i.e., GS. While both studies performed an unprecedent at the time data collection, GS after 2004 brought all these references under one roof. In other words, while papers may have been available at the time, researchers still needed to find them in commercial datasets or the publisher's website. To this end, GS reduced the overall costs of searching.

I should stress that the empirical findings do not answer the question whether a narrowing of science has been taking place over the last decades. Recently, a debate has been ignited over whether science is indeed narrowing and becoming less disruptive (Park et al., 2023; Petersen et al., 2023). Such a debate has deep roots in the theory of scientific and technological progress. While scientific advances create opportunities to stand on the shoulders of giants (Scotchmer, 1991), the growing number of scientific advancements across fields creates a burden of knowledge that in turn forces researchers to specialize and therefore less likely to seek knowledge outside their field (Jones, 2009). Adding to the latter organic evolution of science, several works have highlighted reasons for which interdisciplinary research will be impeded (Tarafdar & Davison, 2018; Van Rijnsoever & Hessels, 2011). This paper contributes to this ongoing debate by highlighting that the reduction in search costs is not likely to impede the use of knowledge from distant fields. If anything, it is likely to promote it. Further, there could be other contributing factors including a change in scientific information consumption, policies fostering interdisciplinary research and changes in scientific production from Mode 1 to Mode 2. Distinguishing the role of each of these forces is beyond the scope of this paper and could be future avenues of research.

ORCID

Kyriakos Drivas https://orcid.org/0000-0002-9567-6083

ENDNOTES

[1] The implicit power of rankings is well documented in the marketing literature in addition to case studies such the airline ticket market (Derakhshan et al., 2022; Ursu, 2018).

[2] See for instance a news blog from Georgia State University Campus which informs its community over the years it has acquired subscription for: https://blog.library.gsu.edu/1999/07/13/the-web-of-science/.

[3] Note that GS is not the only search platform that uses an algorithm to rank results. For instance, PubMed also employs its own relevance algorithm (Kiester & Turp, 2022).

REFERENCES

Barry, A., Born, G., & Weszkalnys, G. (2008). Logics of interdisciplinarity. *Economy and Society*, *37*(1), 20–49.

Beel, J., & Gipp, B. (2009). Google Scholar's ranking algorithm: An introductory overview. In *Proceedings of the 12th international conference on scientometrics and informetrics (ISSI'09)* (Vol. 1, pp. 230–241).

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, *119*(1), 249–275.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524–546.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222.

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, *534*(7609), 684–687.

Bunton, S. A., & Mallon, W. T. (2007). The continued evolution of faculty appointment and tenure policies at US medical schools. *Academic Medicine*, *82*(3), 281–289.

Butler, D. (2004). Science searches shift up a gear as Google starts Scholar engine. *Nature*, *432*(7016), 423.

Cornell, S., Berkhout, F., Tuinstra, W., Tàbara, J. D., Jäger, J., Chabay, I., de Wit, B., Langlais, R., Mills, D., Moll, P., Otto, I. M., Petersen, A., Pohl, C., & van Kerkhoff, L. (2013). Opening up knowledge systems for better responses to global environmental change. *Environmental Science & Policy*, *28*, 60–70.

Cyert, R., & March, J. (1963). *A behavioral theory of the firm*. Prentice-Hall.

Delgado López-Cózar, E., Robinson-García, N., & Torres-Salinas, D. (2014). The G oogle scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*, *65*(3), 446–454.

Derakhshan, M., Golrezaei, N., Manshadi, V., & Mirrokni, V. (2022). Product ranking on online platforms. *Management Science*, *68*(6), 4024–4041.

Detmer, W. M., & Shortliffe, E. H. (1997). Using the Internet to improve knowledge diffusion in medicine. *Communications of the ACM*, *40*(8), 101–108.

Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science*, *321*(5887), 395–399.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal*, *22*(2), 338–342.

Gibbons, M., Limoges, C., Scott, P., Schwartzman, S., & Nowotny, H. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. Sage.

Giles, J. (2004). Publishers go head-to-head over search tool. *Nature*, *428*(6984), 683–684.

Giles, J. (2005). Start your engines. *Nature News*, *438*, 554–555.

Giustini, D., & Barsky, E. (2005). A look at Google Scholar, PubMed, and Scirus: Comparisons and recommendations. *Journal of the Canadian Health Libraries Association*, *26*(3), 85–89.

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43.

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181–217.

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, *118*(1), 177–214.

Halevi, G., Moed, H., & Bar-Ilan, J. (2017). Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of Informetrics*, *11*(3), 823–834.

Hall, B., Mairesse, J., & Turner, L. (2007). Identifying age, cohort, and period effects in scientific research productivity: Discussion and illustration using simulated and actual data on French physicists. *Economics of Innovation and New Technology*, *16*(2), 159–177.

Hollingsworth, J. R., & Hollingsworth, E. J. (2000). Major discoveries and biomedical research organizations: Perspectives on interdisciplinarity, nurturing leadership, and integrated structure and cultures P. In N. S. Weingart (Ed.), *Practising interdisciplinarity* (pp. 215–244). University of Toronto Press.

Holmwood, J. (2010). Sociology's misfortune: Disciplines, interdisciplinarity and the impact of audit culture. *British Journal of Sociology*, *61*(4), 639–658.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, *20*(1), 1–24.

Jacsó, P. (2005). Google Scholar: The pros and the cons. *Online Information Review*, *29*(2), 208–214.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *Review of Economic Studies*, *76*(1), 283–317.

Kiester, L., & Turp, C. (2022). Artificial intelligence behind the scenes: PubMed's Best Match algorithm. *Journal of the Medical Library Association*, *110*(1), 15–22.

Klein, J. T., & Falk-Krzesinski, H. J. (2017). Interdisciplinary and collaborative work: Framing promotion and tenure practices and policies. *Research Policy*, *46*(6), 1055–1061.

Larivière, V., Haustein, S., & Börner, K. (2015). Long-distance interdisciplinarity leads to higher scientific impact. *PLoS One*, *10*(3), e0122565.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, *411*(6837), 521.

Lélé, S., & Norgaard, R. B. (2005). Practicing interdisciplinarity. *Bioscience*, *55*(11), 967–975.

Levine-Clark, M., & Gil, E. L. (2008). A comparative citation analysis of Web of Science, Scopus, and Google Scholar. *Journal of Business & Finance Librarianship*, *14*(1), 32–46.

Leydesdorff, L., & Bornmann, L. (2016). The operationalization of "fields" as WoS subject categories (WCs) in evaluative bibliometrics: The cases of "library and information science" and "science & technology studies". *Journal of the Association for Information Science and Technology*, *67*(3), 707–714.

Lin, Y., Frey, C. B., & Wu, L. (2023). Remote collaboration fuses fewer breakthrough ideas. *Nature*, *623*(7989), 987–991.

Marchionini, G. (1995). *Information seeking in electronic environments* (Vol. 9). Cambridge University Press.

McCabe, M. J., & Snyder, C. M. (2015). Does online availability increase citations? Theory and evidence from a panel of economics and business journals. *Review of Economics and Statistics*, *97*(1), 144–165.

Mehta, A., Rysman, M., & Simcoe, T. (2010). Identifying the age profile of patent citations: New estimates of knowledge diffusion. *Journal of Applied Econometrics*, *25*(7), 1179–1204.

Mullahy, J., & Norton, E. C. (2023). *Why transform Y? The pitfalls of transformed regressions with a mass at zero*. Oxford Bulletin of Economics and Statistics.

Murphy, M. C., Mejia, A. F., Mejia, J., Yan, X., Cheryan, S., Dasgupta, N., Destin, M., Fryberg, S. A., Garcia, J. A., Haines, E. L., Harackiewicz, J. M., Ledgerwood, A., Moss-Racusin, C. A., Park, L. E., Perry, S. P., Ratliff, K. A., Rattan, A., Sanchez, D. T., Savani, K., … Pestilli, F. (2020). Open science, communal culture, and women's participation in the movement to improve science. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(39), 24154–24164.

National Academy of Sciences, Committee on Science, Public Policy, & Committee on Facilitating Interdisciplinary Research. (2005). *Facilitating interdisciplinary research*. National Academies Press.

National Research Council, Division on Earth, Life Studies, Board on Life Sciences, & Committee on Key Challenge Areas for Convergence. (2014). *Convergence: Facilitating transdisciplinary integration of life sciences, physical sciences, engineering, and beyond*. National Academies Press.

Nelson, R. R., & Winter, S. (1982). *An evolutionary theory of economic change*. Harvard University Press.

Van Noorden, R. (2014). Online collaboration: Scientists and the social network. *Nature News*, *512*(7513), 126–129.

Nowotny, H., Scott, P., & Gibbons, M. (2003). Introduction: "Mode 2" revisited: The new production of knowledge. *Minerva*, *41*(3), 179–194.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955.

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A., & Delgado López-Cózar, E. (2015). Methods for estimating the size of Google Scholar. *Scientometrics*, *104*, 931–949.

Page, S. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies-new edition*. Princeton University Press.

Palmer, C. L. (1999). Structures and strategies of interdisciplinary science. *Journal of the American Society for Information Science*, *50*(3), 242–253.

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature*, *613*(7942), 138–144.

Petersen, A. M., Arroyave, F., & Pammolli, F. (2023). *The disruption index is biased by citation inflation*. arXiv preprint: arXiv: 2306.01949.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675.

Porter, A. L., Roessner, J. D., Cohen, A. S., & Perreault, M. (2006). Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation*, *15*(3), 187–195.

Prati, R. C., Rodrigues, B. S., Aragão, I., Soares, T. A., Quiles, M. G., & Da Silva, J. L. (2024). The impact of interdisciplinary, gender and geographic distributions on the citation patterns of the Journal of Chemical Information and Modeling. *Journal of Chemical Information and Modeling*, *64*(4), 1107–1111.

Priem, J., Piwowar, H., & Orr, R. (2022). *OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts*. arXiv preprint: arXiv:2205.01833.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinarity. The case of innovation studies and business and management. *Research Policy*, *41*, 1262–1282.

Schilling, M. A., & Green, E. (2011). Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, *40*(10), 1321–1331.

Scotchmer, S. (1991). Standing on the shoulders of giants: Cumulative research and the patent law. *Journal of Economic Perspectives*, *5*(1), 29–41.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628–638.

Shen, Z., Ma, H., & Wang, K. (2018). *A web-scale system for scientific knowledge exploration*. arXiv preprint: arXiv:1805.12216.

Sun, Y., Livan, G., Ma, A., & Latora, V. (2021). Interdisciplinary researchers attain better long-term funding performance. *Communications Physics*, *4*(1), 263.

Tarafdar, M., & Davison, R. M. (2018). Research in information systems: Intra-disciplinary and inter-disciplinary approaches. *Journal of the Association for Information Systems*, *19*(6), 551.

Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 415–422). ACM.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, *37*(4), 530–552.

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472.

Walters, W. H. (2007). Google scholar coverage of a multidisciplinary field. *Information Processing & Management*, *43*(4), 1121–1132.

Williams, K., Michalska, S., Cohen, E., Szomszor, M., & Grant, J. (2023). Exploring the application of machine learning to expert evaluation of research impact. *PLoS One*, *18*(8), e0288469.

Wu, L., Wang, D., & Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, *566*(7744), 378–382.

Yan, E. (2014). Finding knowledge paths among scientific disciplines. *Journal of the Association for Information Science and Technology*, *65*(11), 2331–2347.

Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E., & Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature Communications*, *10*(1), 3439.

How to cite this article: Drivas, K. (2024). The role of online search platforms in scientific diffusion. *Journal of the Association for Information Science and Technology*, 1–24. https://doi.org/10.1002/asi.24959

APPENDIX A A

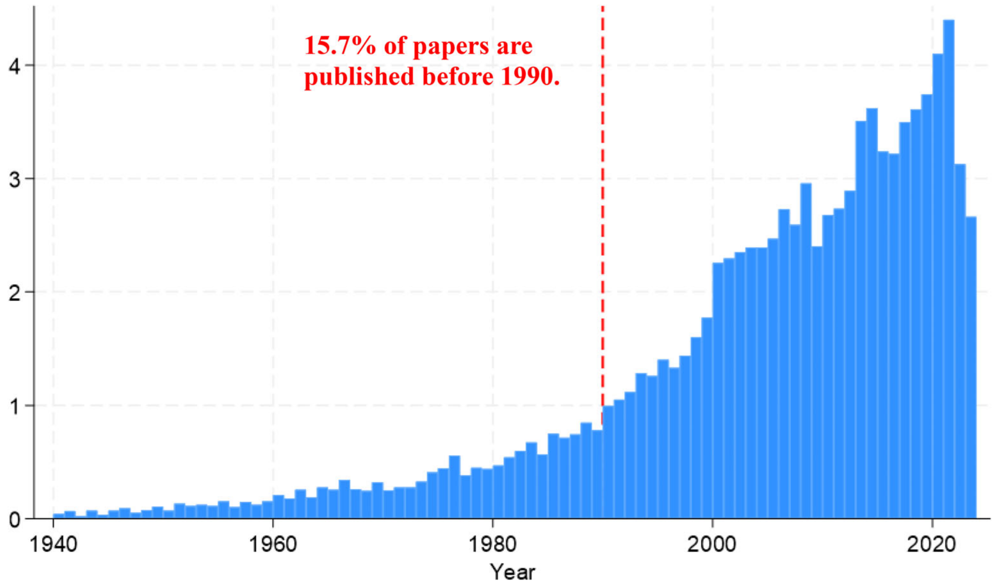

FIGURE A1 Histogram of GS search results by publication year. Two thousand keywords were searched in GS and noted the publication years of the first 10 search results. The figure displays the histogram of the years for these results. For exposition purposes, results prior to 1940 are not displayed. These constitute 9.7% of the sample.

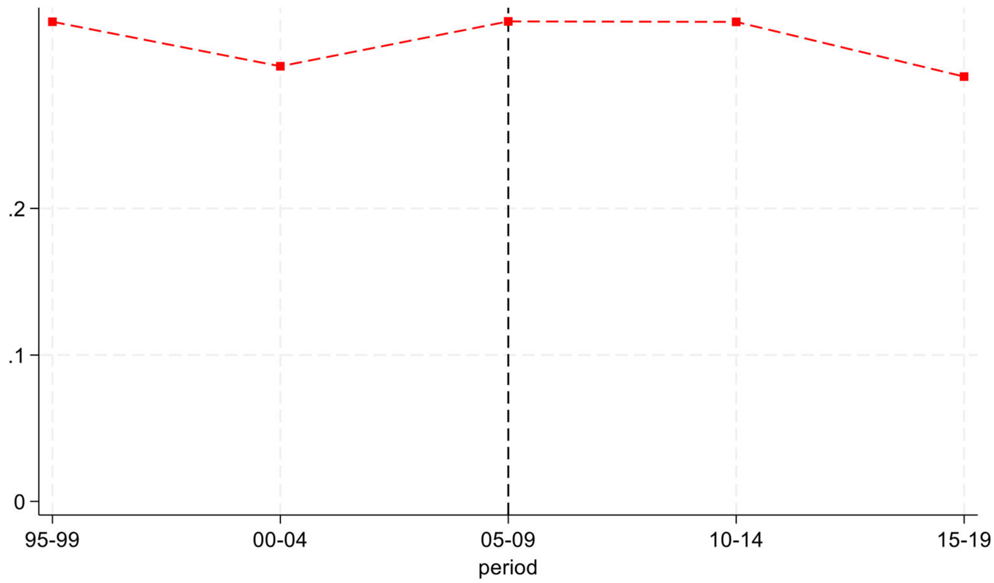

FIGURE A2 Citations for the 70–79 group. For papers published between 1970 and 1979 the dashed line displays the average cumulative citations that were received over 5-year intervals starting from 1995 to 1999 ending to 2015–2019.

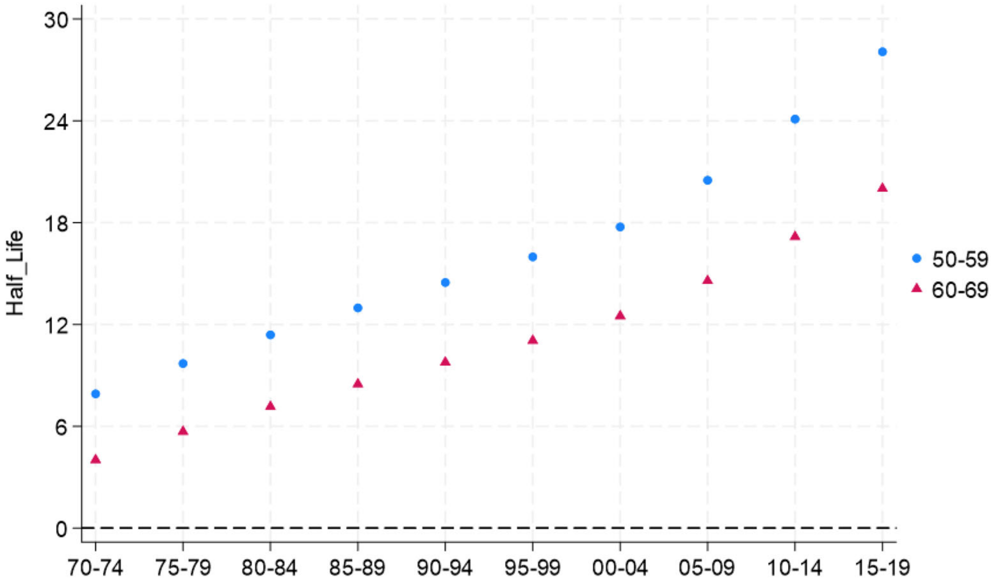

FIGURE A3 Cited half-life by group of older papers.

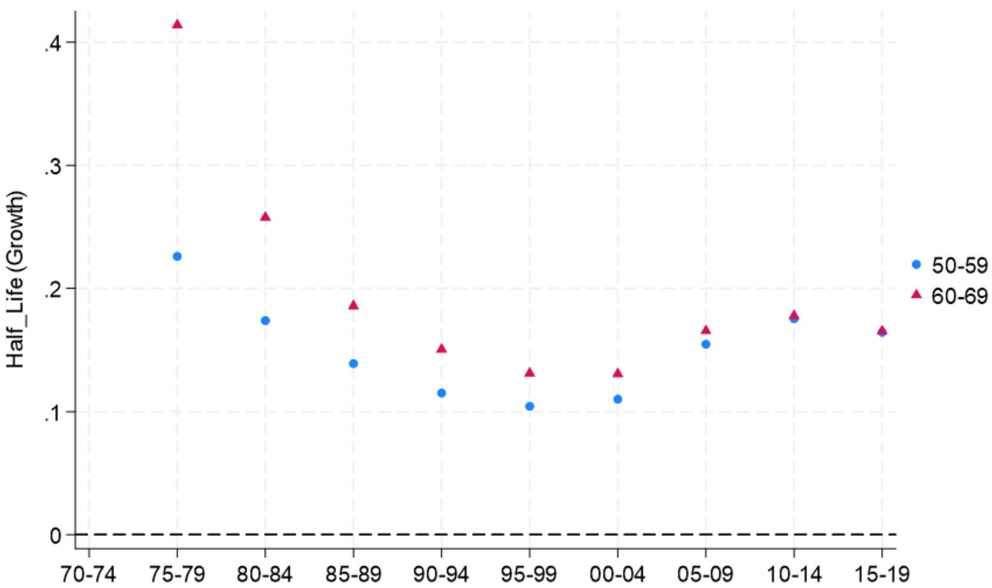

FIGURE A4 Growth of cited half-life by group of older papers.

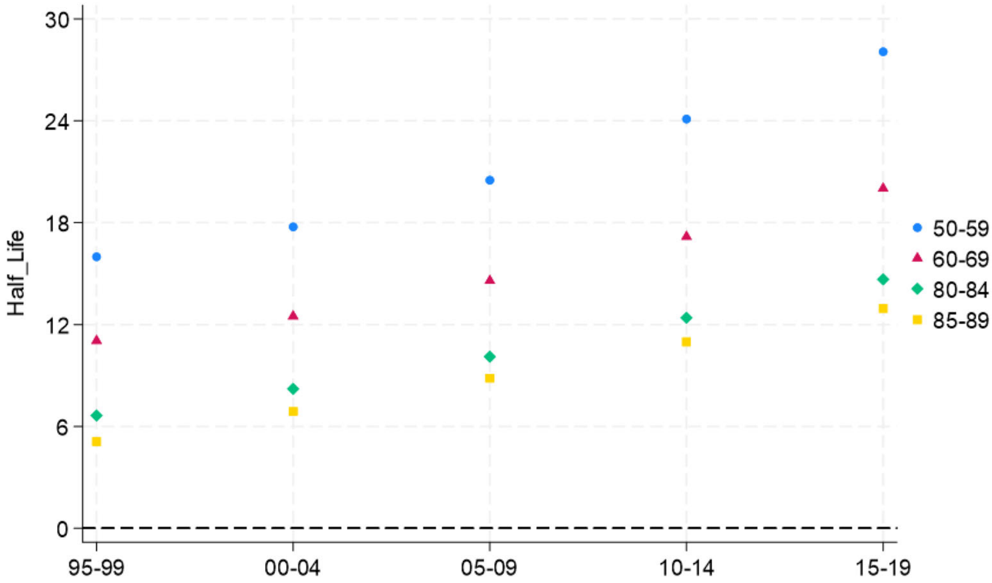

FIGURE A5 Cited half-life by group of papers. All core papers.

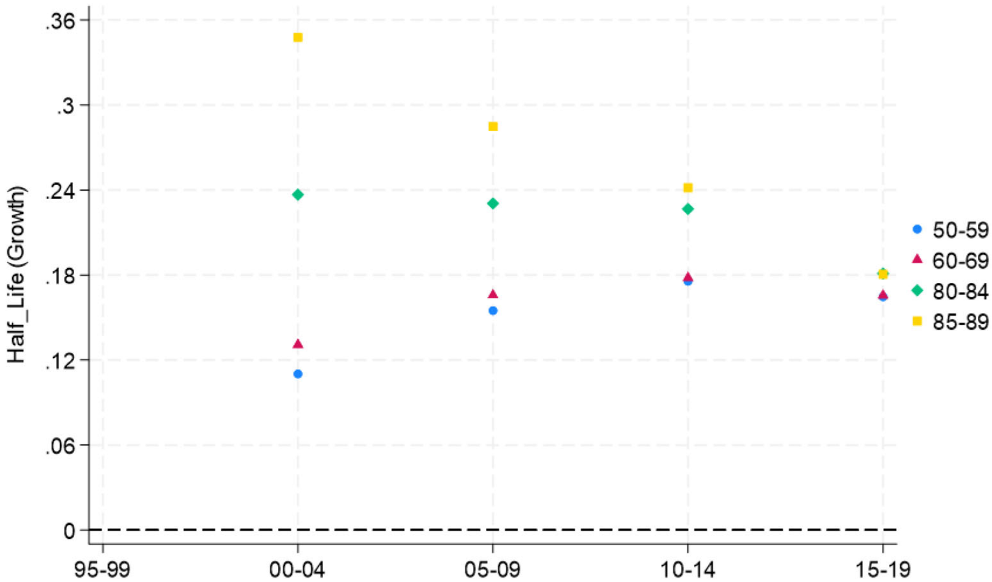

FIGURE A6 Growth of cited half-life by group of papers. All core papers.

TABLE A1 Sample size and cuts by group of papers.

Year published	Frequency of all papers	Frequency of papers that disclose at least one Level-1 Concept	Frequency of papers that disclose at least one Level-1 Concept and have garnered at least one citation during 1995–2019. Also take out outliers
1950–1959	1,792,428	1,434,758	322,720
1960–1969	3,108,737	2,623,911	770,870
1980–1984	6,738,590	6,031,696	1,355,112
1985–1989	8,337,446	7,526,872	1,887,924
Entire period	19,977,201	17,617,237	4,336,626

TABLE A2 Comparison of citation rates for papers published between 1950–1959 and 1980–1989. Unmatched and matched groups.

	Unmatched comparison		Matched comparison (97%)		Unmatched comparison		Matched comparison (97%)	
Citing period	**80–84**	**50–59**	**80–84**	**50–59**	**85–89**	**50–59**	**85–89**	**50–59**
95–99	2.47	0.83	0.81	0.81	3.58	0.83	0.83	0.83
00–04	2.12	0.85	0.82	0.82	2.83	0.85	0.84	0.84
05–09	2.13	1.25	1.10	1.21	2.61	1.25	1.10	1.23
10–14	2.09	1.28	1.16	1.23	2.45	1.28	1.15	1.26
15–19	1.77	1.21	1.02	1.16	2.01	1.21	1.00	1.19

Note: The percentage in the parenthesis denotes the percent of papers matched from the 50–59 group of papers.

TABLE A3 Comparison of citation rates for papers published between 1960–1969 and 1980–1989. Unmatched and matched groups.

	Unmatched comparison		Matched comparison (93%)		Unmatched comparison		Matched comparison (93%)	
Citing period	**80–84**	**60–69**	**80–84**	**60–69**	**85–89**	**60–69**	**85–89**	**60–69**
95–99	2.47	1.11	1.13	1.13	3.58	1.11	1.16	1.16
00–04	2.12	1.07	1.07	1.07	2.83	1.07	1.10	1.10
05–09	2.13	1.35	1.28	1.34	2.61	1.35	1.26	1.37
10–14	2.09	1.39	1.32	1.37	2.45	1.39	1.30	1.40
15–19	1.77	1.28	1.16	1.27	2.01	1.28	1.13	1.29

Note: The percentage in the parenthesis denotes the percent of papers matched from the 60–69 group of papers.

TABLE A4 Citations within and beyond the same Level-0 Concept as the focal paper. Compare papers published between 1950 and 1959 to papers published between 1980 and 1984 and to papers published between 1985 and 1989. Unmatched and matched groups.

	Citations from a different Level-0 Concept (NonSame*L*0)							
	Unmatched comparison		**Matched comparison (96%)**		**Unmatched comparison**		**Matched comparison (96%)**	
	80–84	**50–59**	**80–84**	**50–59**	**85–89**	**50–59**	**85–89**	**50–59**
95–99	0.13	0.09	0.07	0.07	0.18	0.09	0.07	0.07
00–04	0.13	0.10	0.07	0.07	0.16	0.10	0.07	0.07
05–09	0.14	0.16	0.10	0.14	0.17	0.16	0.10	0.14
10–14	0.15	0.18	0.11	0.14	0.17	0.18	0.10	0.14
15–19	0.12	0.15	0.09	0.12	0.13	0.15	0.09	0.12
	Citations from a same Level-0 Concept (Same*L*0)							
	Unmatched comparison		**Matched comparison (96%)**		**Unmatched comparison**		**Matched comparison (96%)**	
	80–84	**50–59**	**80–84**	**50–59**	**85–89**	**50–59**	**85–89**	**50–59**
95–99	2.34	0.74	0.68	0.68	3.40	0.74	0.70	0.70
00–04	1.99	0.75	0.68	0.68	2.67	0.75	0.70	0.70
05–09	1.99	1.08	0.95	1.01	2.44	1.08	0.95	1.03
10–14	1.94	1.11	1.00	1.02	2.29	1.11	0.99	1.04
15–19	1.65	1.06	0.88	0.98	1.88	1.06	0.87	1.00

Note: The percentage in the parenthesis denotes the percent of papers matched from the 50–59 group of papers.

TABLE A5 Citations within and beyond the same Level-0 Concept as the focal paper. Compare papers published between 1960 and 1969 to papers published between 1980 and 1984 and to papers published between 1985 and 1989. Unmatched and matched groups.

	Citations from a different Level-0 Concept (NonSame*L*0)							
	Unmatched comparison		**Matched comparison (91%)**		**Unmatched comparison**		**Matched comparison (91%)**	
	80–84	**60–69**	**80–84**	**60–69**	**85–89**	**60–69**	**85–89**	**60–69**
95–99	0.13	0.09	0.36	0.36	0.18	0.09	0.39	0.39
00–04	0.13	0.10	0.34	0.34	0.16	0.10	0.36	0.36
05–09	0.14	0.13	0.44	0.46	0.17	0.13	0.44	0.48
10–14	0.15	0.14	0.48	0.52	0.17	0.14	0.50	0.54
15–19	0.12	0.12	0.46	0.50	0.13	0.12	0.48	0.51
	Citations from a same Level-0 Concept (Same*L*0)							
	Unmatched comparison		**Matched comparison (91%)**		**Unmatched comparison**		**Matched comparison (91%)**	
	80–84	**60–69**	**80–84**	**60–69**	**85–89**	**60–69**	**85–89**	**60–69**
95–99	2.34	1.02	1.93	1.93	3.40	1.02	2.01	2.00
00–04	1.99	0.97	1.81	1.81	2.67	0.97	1.89	1.88
05–09	1.99	1.22	2.11	2.16	2.44	1.22	2.15	2.23
10–14	1.94	1.24	2.33	2.38	2.29	1.24	2.36	2.47
15–19	1.65	1.16	2.39	2.51	1.88	1.16	2.38	2.58

Note: The percentage in the parenthesis denotes the percent of papers matched from the 60–69 group of papers.

TABLE A6 Citations within and beyond the same Level-1 Concept as the focal paper. Compare papers published between 1950 and 1959 to papers published between 1980 and 1984 and to papers published between 1985 and 1989. Unmatched and matched groups.

	Citations from a different Level-1 Concept (*NonSameL1*)							
	Unmatched comparison		**Matched comparison (95%)**		**Unmatched comparison**		**Matched comparison (95%)**	
	80–84	**50–59**	**80–84**	**50–59**	**85–89**	**50–59**	**85–89**	**50–59**
95–99	0.41	0.25	0.59	0.60	0.55	0.25	0.62	0.62
00–04	0.38	0.26	0.58	0.58	0.47	0.26	0.61	0.61
05–09	0.42	0.43	0.75	0.87	0.49	0.43	0.77	0.90
10–14	0.44	0.43	0.85	0.95	0.49	0.43	0.89	0.97
15–19	0.38	0.40	0.86	1.03	0.41	0.40	0.89	1.07
	Citations from a same Level-1 Concept (Same*L*1)							
	Unmatched comparison		**Matched comparison (95%)**		**Unmatched comparison**		**Matched comparison (95%)**	
	80–84	**50–59**	**80–84**	**50–59**	**85–89**	**50–59**	**85–89**	**50–59**
95–99	2.04	0.57	1.24	1.24	3.00	0.57	1.35	1.32
00–04	1.71	0.57	1.21	1.21	2.31	0.57	1.28	1.28
05–09	1.68	0.79	1.46	1.56	2.08	0.79	1.51	1.64
10–14	1.62	0.82	1.65	1.79	1.93	0.82	1.65	1.93
15–19	1.38	0.79	1.69	1.96	1.58	0.79	1.66	2.10

Note: The percentage in the parenthesis denotes the percent of papers matched from the 50–59 group of papers.

TABLE A7 Citations within and beyond the same Level-1 Concept as the focal paper. Compare papers published between 1960 and 1969 to papers published between 1980 and 1984 and to papers published between 1985 and 1989. Unmatched and matched groups.

	Citations from a different Level-1 Concept (NonSame$L$1)							
	Unmatched comparison		Matched comparison (89%)		Unmatched comparison		Matched comparison (90%)	
	80–84	60–69	80–84	60–69	85–89	60–69	85–89	60–69
95–99	0.41	0.27	0.61	0.61	0.55	0.27	0.65	0.65
00–04	0.38	0.27	0.58	0.58	0.47	0.27	0.61	0.61
05–09	0.42	0.37	0.74	0.80	0.49	0.37	0.76	0.82
10–14	0.44	0.39	0.85	0.88	0.49	0.39	0.87	0.91
15–19	0.38	0.35	0.84	0.91	0.41	0.35	0.88	0.93
	Citations from a same Level-1 Concept (Same$L$1)							
	Unmatched comparison		Matched comparison (89%)		Unmatched comparison		Matched comparison (90%)	
	80–84	60–69	80–84	60–69	85–89	60–69	85–89	60–69
95–99	2.04	0.83	1.54	1.54	3.00	0.83	1.64	1.63
00–04	1.71	0.78	1.42	1.42	2.31	0.78	1.49	1.49
05–09	1.68	0.95	1.69	1.71	2.08	0.95	1.73	1.78
10–14	1.62	0.97	1.88	1.87	1.93	0.97	1.91	1.96
15–19	1.38	0.91	1.93	1.97	1.58	0.91	1.94	2.05

Note: The percentage in the parenthesis denotes the percent of papers matched from the 60–69 group of papers.

TABLE A8 Testing Hypothesis 1. All citations.

	(1)	(2)	(3)	(4)
Treat	0.000 (0.000)	0.000 (0.000)	0.000 (0.002)	0.000 (0.001)
Treat_x_00 – 04	−0.000 (0.000)	−0.000 (0.000)	−0.002 (0.002)	−0.001 (0.001)
Treat_x_05 – 09	0.319*** (0.006)	0.238*** (0.004)	0.390*** (0.007)	0.331*** (0.004)
Treat_x_10 – 14	0.168*** (0.007)	0.120*** (0.005)	0.223*** (0.008)	0.203*** (0.005)
Treat_x_15 – 19	0.294*** (0.010)	0.211*** (0.006)	0.364*** (0.011)	0.318*** (0.006)
00 – 04	0.078*** (0.004)	0.007*** (0.003)	0.086*** (0.004)	0.014*** (0.003)
05 – 09	0.505*** (0.005)	0.376*** (0.003)	0.440*** (0.006)	0.281*** (0.004)
10 – 14	0.522*** (0.005)	0.369*** (0.004)	0.478*** (0.006)	0.289*** (0.004)
15 – 19	0.363*** (0.007)	0.182*** (0.005)	0.300*** (0.007)	0.075*** (0.005)
Constant	0.963*** (0.003)	1.321*** (0.002)	0.986*** (0.003)	1.366*** (0.002)
Observations	3,125,300	7,139,020	3,120,140	7,124,800
R-squared	0.508	0.535	0.558	0.560
adj R-squared	0.453	0.484	0.508	0.511
# of pairs	312,532	713,904	312,019	712,485

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Columns 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The dependent variable is all citations available at OpenAlex. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
***$p < 0.01$.

TABLE A9 Testing Hypothesis 1. Include all papers.

	(1)	(2)	(3)	(4)
Treat	0.018 (0.016)	−0.000 (0.000)	0.016 (0.017)	−0.000 (0.005)
Treat_x_00 – 04	0.003 (0.004)	−0.000 (0.000)	0.004 (0.004)	0.003 (0.003)
Treat_x_05 – 09	0.058*** (0.007)	0.047*** (0.001)	0.080*** (0.008)	0.086*** (0.008)
Treat_x_10 – 14	0.087*** (0.012)	0.059*** (0.002)	0.124*** (0.013)	0.128*** (0.012)
Treat_x_15 – 19	0.125*** (0.016)	0.086*** (0.004)	0.176*** (0.016)	0.169*** (0.013)
00 – 04	0.019*** (0.003)	−0.011*** (0.001)	0.020*** (0.004)	−0.000 (0.003)
05 – 09	0.120*** (0.006)	0.065*** (0.001)	0.102*** (0.007)	0.070*** (0.007)
10 – 14	0.168*** (0.009)	0.094*** (0.002)	0.136*** (0.010)	0.104*** (0.011)
15 – 19	0.134*** (0.011)	0.040*** (0.003)	0.090*** (0.010)	0.040*** (0.012)
Constant	0.301*** (0.009)	0.442*** (0.001)	0.314*** (0.010)	0.518*** (0.005)
Observations	13,806,890	22,488,180	13,954,330	23,411,690
R-squared	0.540	0.712	0.529	0.528
adj R-squared	0.489	0.680	0.477	0.475
# of pairs	1,380,689	2,248,818	1,395,433	2,341,169

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Column 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The matching process and the subsequent regressions include papers with no citations and the outliers with respect to citation counts. The dependent variable is citations by journal articles. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.

***$p < 0.01$.

TABLE A10 Counterpart of Table 2. All cites.

	(1) Citations in the same Level-0 Concept	(2)	(3)	(4)	(5) Citations in the different Level-0 Concept	(6)	(7)	(8)
Treat	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Treat_x_00 − 04	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Treat_x_05 − 09	0.062*** (0.004)	0.044*** (0.003)	0.084*** (0.004)	0.074*** (0.003)	0.038*** (0.001)	0.017*** (0.001)	0.042*** (0.001)	0.021*** (0.001)
Treat_x_10 − 14	0.016*** (0.005)	0.024*** (0.003)	0.046*** (0.005)	0.063*** (0.003)	0.037*** (0.001)	0.022*** (0.001)	0.041*** (0.001)	0.026*** (0.001)
Treat_x_15 − 19	0.090*** (0.005)	0.081*** (0.004)	0.128*** (0.006)	0.132*** (0.004)	0.030*** (0.001)	0.016*** (0.001)	0.033*** (0.001)	0.020*** (0.001)
00 − 04	−0.002 (0.002)	−0.062*** (0.002)	−0.002 (0.002)	−0.066*** (0.002)	0.002*** (0.001)	0.001** (0.000)	0.002** (0.001)	−0.000 (0.000)
05 − 09	0.261*** (0.003)	0.134*** (0.002)	0.240*** (0.003)	0.093*** (0.003)	0.029*** (0.001)	0.026*** (0.001)	0.023*** (0.001)	0.019*** (0.001)
10 − 14	0.315*** (0.004)	0.172*** (0.003)	0.289*** (0.004)	0.122*** (0.003)	0.038*** (0.001)	0.033*** (0.001)	0.032*** (0.001)	0.025*** (0.001)
15 − 19	0.200*** (0.004)	0.045*** (0.003)	0.165*** (0.004)	−0.022*** (0.003)	0.020*** (0.001)	0.017*** (0.001)	0.015*** (0.001)	0.010*** (0.001)
Constant	0.671*** (0.002)	0.960*** (0.001)	0.690*** (0.002)	1.009*** (0.002)	0.068*** (0.001)	0.059*** (0.000)	0.071*** (0.001)	0.065*** (0.000)
Observations	3,085,110	7,030,150	3,089,050	7,054,320	3,084,490	7,032,820	3,088,310	7,052,610
R-squared	0.536	0.524	0.547	0.546	0.417	0.419	0.416	0.428
adj R-squared	0.484	0.471	0.496	0.496	0.352	0.354	0.351	0.364
# of pairs	308,511	703,015	308,905	705,432	308,449	703,282	308,831	705,261

Note: For Columns 1–4 the dependent variable is citations by works that share at least one Level-0 Concept as the focal paper. For Columns 5–8 the dependent variable is citations by works that share no Level-0 Concept as the focal paper. Columns 1 and 5 compare matched papers from the 50–59 group with the 80–84 group. Columns 2 and 6 compare matched papers from the 60–69 group with the 80–84 group. Columns 3 and 7 compare matched papers from the 50–59 group with the 85–89 group. Columns 4 and 8 compares matched papers from the 60–69 group with the 85–89 group. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
$p < 0.05$.*$p < 0.01$.

TABLE A11 Further testing Hypothesis 2. Examine citations within the same Level-0 Concept but not the same Level-1 Concept.

	Journal citations				Citations by all works in OpenAlex			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Treat_*x*_00 − 04	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Treat_*x*_05 − 09	0.076*** (0.002)	0.042*** (0.001)	0.083*** (0.002)	0.050*** (0.001)	0.127*** (0.002)	0.088*** (0.001)	0.146*** (0.002)	0.109*** (0.002)
Treat_*x*_10 − 14	0.044*** (0.002)	0.032*** (0.001)	0.050*** (0.002)	0.040*** (0.001)	0.067*** (0.002)	0.048*** (0.002)	0.075*** (0.002)	0.064*** (0.002)
Treat_*x*_15 − 19	0.048*** (0.002)	0.034*** (0.001)	0.057*** (0.002)	0.045*** (0.001)	0.076*** (0.003)	0.055*** (0.002)	0.088*** (0.003)	0.073*** (0.002)
00 − 04	0.002** (0.001)	−0.004*** (0.001)	0.002 (0.001)	−0.006*** (0.001)	0.010*** (0.001)	0.002*** (0.001)	0.009*** (0.001)	0.000 (0.001)
05 − 09	0.052*** (0.001)	0.040*** (0.001)	0.044*** (0.001)	0.028*** (0.001)	0.108*** (0.002)	0.097*** (0.001)	0.088*** (0.002)	0.072*** (0.001)
10 − 14	0.068*** (0.001)	0.055*** (0.001)	0.060*** (0.001)	0.043*** (0.001)	0.109*** (0.002)	0.098*** (0.001)	0.100*** (0.002)	0.079*** (0.001)
15 − 19	0.044*** (0.001)	0.033*** (0.001)	0.034*** (0.001)	0.018*** (0.001)	0.086*** (0.002)	0.071*** (0.001)	0.072*** (0.002)	0.050*** (0.001)
Constant	0.127*** (0.001)	0.139*** (0.000)	0.134*** (0.001)	0.151*** (0.000)	0.153*** (0.001)	0.167*** (0.001)	0.161*** (0.001)	0.182*** (0.001)
Observations	2,948,250	6,625,120	2,954,750	6,670,250	2,890,470	6,455,740	2,900,670	6,500,350
R-squared	0.309	0.323	0.321	0.330	0.276	0.297	0.291	0.305
adj R-squared	0.232	0.248	0.246	0.256	0.196	0.219	0.212	0.228
# of searches	294,825	662,512	295,475	667,025	289,047	645,574	290,067	650,035

Note: For Columns 1–4 the dependent variable is journal article citations by works that share at least one Level-0 Concept but do not share any Level-1 Concept as the focal paper. For Columns 5–8 the dependent variable is citations by works that share at least one Level-0 Concept but do not share any Level-1 Concept as the focal paper. Note that Columns 5–8 consider citations by any type of work in OpenAlex. Columns 1 and 5 compare matched papers from the 50–59 group with the 80–84 group. Columns 2 and 6 compare matched papers from the 60–69 group with the 80–84 group. Columns 3 and 7 compare matched papers from the 50–59 group with the 85–89 group. Columns 4 and 8 compares matched papers from the 60–69 group with the 85–89 group. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.
$p < 0.05$.*$p < 0.01$.

TABLE A12 Counterpart of Table 3. All cites.

	(1)	(2)	(3)	(4)
Treat	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Treat_x_00 – 04	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Treat_x_05 – 09	0.036*** (0.004)	0.053*** (0.003)	0.069*** (0.004)	0.103*** (0.003)
Treat_x_10 – 14	−0.026*** (0.005)	0.010*** (0.003)	0.002 (0.005)	0.051*** (0.003)
Treat_x_15 – 19	0.059*** (0.005)	0.080*** (0.004)	0.105*** (0.005)	0.137*** (0.004)
00 – 04	0.012*** (0.002)	−0.038*** (0.002)	0.012*** (0.002)	−0.038*** (0.002)
05 – 09	0.331*** (0.003)	0.230*** (0.002)	0.299*** (0.003)	0.175*** (0.003)
10 – 14	0.355*** (0.004)	0.233*** (0.003)	0.331*** (0.004)	0.188*** (0.003)
15 – 19	0.255*** (0.004)	0.113*** (0.003)	0.212*** (0.004)	0.049*** (0.003)
Constant	0.522*** (0.002)	0.796*** (0.001)	0.546*** (0.002)	0.841*** (0.001)
Observations	3,018,130	6,767,270	3,026,750	6,814,240
R-squared	0.489	0.499	0.505	0.517
adj R-squared	0.432	0.443	0.450	0.463
# of searches	301,813	676,727	302,675	681,424

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Column 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The dependent variable is citations by journal articles that share at least one Level-1 Concept as the focal paper. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.

***$p < 0.01$.

TABLE A13 Counterpart of Table 4. All cites.

	(1)	(2)	(3)	(4)
Treat	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Treat_x_00 – 04	0.000 (0.000)	−0.000 (0.000)	0.000 (0.000)	−0.000 (0.000)
Treat_x_05 – 09	−0.017*** (0.003)	0.018*** (0.002)	−0.004 (0.003)	0.044*** (0.002)
Treat_x_10 – 14	−0.049*** (0.004)	−0.018*** (0.003)	−0.037*** (0.004)	0.005** (0.003)
Treat_x_15 – 19	0.015*** (0.004)	0.036*** (0.003)	0.036*** (0.004)	0.069*** (0.003)
00 – 04	−0.004** (0.002)	−0.038*** (0.001)	−0.006*** (0.002)	−0.042*** (0.002)
05 – 09	0.233*** (0.003)	0.159*** (0.002)	0.217*** (0.003)	0.124*** (0.002)
10 – 14	0.248*** (0.003)	0.161*** (0.002)	0.234*** (0.003)	0.129*** (0.002)
15 – 19	0.175*** (0.003)	0.077*** (0.002)	0.153*** (0.003)	0.032*** (0.002)
Constant	0.332*** (0.001)	0.524*** (0.001)	0.352*** (0.001)	0.560*** (0.001)
Observations	2,627,750	6,122,100	2,637,880	6,160,180
R-squared	0.424	0.442	0.455	0.463
adj R-squared	0.360	0.380	0.395	0.403
# of searches	262,775	612,210	263,788	616,018

Note: Column 1 compares matched papers from the 50–59 group with the 80–84 group. Column 2 compares matched papers from the 60–69 group with the 80–84 group. Column 3 compares matched papers from the 50–59 group with the 85–89 group. Columns 4 compares matched papers from the 60–69 group with the 85–89 group. The dependent variable is citations by works that share at least one Level-2 Concept as the focal paper. All columns are estimated via OLS. All regressions include paper pair fixed effects. Standard errors are clustered at the paper pair level.

$p < 0.05$.*$p < 0.01$.