

Let Me Begin Again

Devin Smith, devin.smith.work@gmail.com

Keywords: Uncitedness, uncited, citation analysis, informetrics, qualitative, quantitative, discovery.

Abstract

This is a literature review of work on uncitedness, with an emphasis on publications from the last decade. Uncitedness varies widely between subject areas and document types, and is trending upward along with the total number of publications per year. There have been significant changes in the geographical distribution of uncitedness since the 1990s, reflecting the increasing presence of European and Asian authors in highly-cited journals. Although a few field-specific uncitedness surveys are now published each year, it remains infrequently studied as a standalone phenomenon: Uncitedness analyses typically occur within work on citedness, which frequently responds to performance-based administrative contexts. This paper identifies a significant knowledge gap in the lack of qualitative research on the topic, advocates for increased engagement with uncited works, and calls for librarian interventions supporting non-quantitative content delivery frameworks.

Terminology Notes

- An uncited item exists in a state of “uncitedness.”
- Uncitedness is temporal: An item may be uncited in one study’s time window, and cited in another’s.
- When considered within a corpus (e.g., a journal or field of research), uncitedness refers to the number of uncited items in the corpus; The uncitedness “rate,” “ratio,” or “factor” refer to the percentage of the corpus which is uncited.
- Qualifier terms such as “seldom-cited” or “highly-cited” are not standardized; They are typically defined on a per-study basis.
- In this paper, I will use the broader term “informetrics” to encompass narrower fields like bibliometrics and scientometrics.

- The descriptor “performance-based” refers to an administrative framework utilizing metrics to assess the quality of a subject’s (e.g., an author, a research program, or funding) output.
- The term “impact” as it typically appears in informetrics and performance-based contexts refers to measurable changes in the metrics being considered.

Into the Void

Nicolaisen and Frandsen’s (2019) *Zero Impact: A large-scale study of uncitedness* is the most expansive work on the subject to date, encompassing Scopus data for 29.5 million documents published between 1996 and 2016, 7.5 million of which are uncited. This study groups the population by subject areas and document type, revealing a “somewhat tacitly agreed upon genre hierarchy existing in all subject areas, yet with important local traits and differences.” For example, in physics and astronomy, conference papers typically precede a journal article, whereas in arts and humanities, they’re often the final research output. Accordingly, uncitedness rates in the former are much higher than the latter. Similarly, letters and notes serve different functions across subject areas’ publishing norms, resulting in wide variations in uncitedness.

The study’s 20-year window also clarified Hamilton’s (1991) earlier findings of arts and humanities’ high uncitedness rates: These articles take longer to accrue citations, but within 15 years attain similar citation rates to other subject areas. Nicolaisen and Frandsen conclude by introducing the concept of a document’s “citation potential,” taking into account its type, subject area, and time since publication.

This study found that uncitedness rates are trending upward across all subject areas and document types. Given the year-over-year increase in total publications (Pan, et al. 2018) and typical delays in citation accrual, this makes intuitive sense — but Kozłowski et al. (2024) reveal striking changes in the geography of uncitedness. Analyzing Web of Science’s core collection from 1980 to 2010, they found that when North American publications were excluded, the uncitedness rate almost halved, from 44% to 23%. This reflects the increased scholarly output of Europe starting in the 1990s and Asia in the 2000s, and variations in international authorship and referencing patterns. The rising dominance of Asian papers is evidenced by excluding their publications, which produces a near-exponentially rising curve. Given these

more recent findings, an updated replication of Xia et al.'s (2014) earlier study on the effects of cross-country authorship (which showed that uncitedness tended to decrease when Indian authors partnered with authors from “developed countries”) would likely be a worthwhile endeavor.

Informetrics is a Winner’s Game

But dedicated studies of uncitedness are more the exception than the rule. Uncitedness most often appears as the Jungian Shadow within studies of the more frequent, prestigious, and lucrative phenomenon of citedness.

Prior to the introduction of citation indexes in the 1960s, large-scale citation analysis was impractical — and thereafter, possible but laborious. With the introduction of online informetrics databases like *Web of Science* in 1997 and *Scopus* in 2004, network speed and computer power became the limiting factors; Informetrics research has blossomed in tandem with big data processing techniques. Field-specific citation analyses (which usually include some consideration of uncitedness) are now regularly published: Here’s three recent examples from linguistics (CheshmehSohrabi & Mashhad, 2023), high-energy physics (Katchanov, et al., 2023), and ophthalmology (Whitescarver, et al., 2023).

Polymath Eugene Garfield (1925–2017), founder of the Institute for Scientific Information (ISI), is a key figure in the field of informetrics. Drawing inspiration from the case law index *Shepard’s Citations*, in 1964 he launched the *Scientific Citation Index*, which alongside ISI’s *Social Sciences Citation Index* became the authoritative sources for academic publishing. To assist with selecting journals for inclusion, he developed the Impact Factor metric (now more commonly rendered as “Journal Impact Factor,” or JIF), which considers (among other things) a journal’s citation count in its calculation.¹ Contrary to Garfield’s intention for the metric, JIF began to take an outsized role in performance-based administrative functions like hiring decisions, research funding, science policy, and university rankings; It has since been joined by a bouquet of citation- and impact-related metrics like the h-index and the Eigenfactor, and an elaborate parallel universe of altmetrics. (Baykoucheva, 2019; Pakkan, et al., 2021)

¹ The ISI-published *Journal Citation Reports* publishes these metrics. As of 2019, it covers over 11,500 journals in 230 disciplines.

It's worth recounting this background, because uncitedness analysis frequently play a role in evaluating or developing impact metrics — and even when not explicitly summoned, the specter of performance-based funding lurks between the lines and stanzas of this oeuvre. Researchers have experimented with removing uncited, lowly-cited, and highly-cited papers from various Scopus categories, finding that the results sometimes align better with certain statistical models. (Thelwall & Wilson, 2014; Thelwall, 2016) Miranda and Garcia-Carpintero's (2019) study on citedness and uncitedness across different research fields is framed specifically around the increasing use of first-quartile journals in “performance-based funding of public research;” Its fascinating and wide-ranging findings emphasize the limitations of using informetrics for this purpose. In Amez's (2012) thorough quantitative analysis of several different variants of the h-index as applied to Leibniz Prize winners, the introduction and conclusion clarify that the paper's purpose is advancing critical discourse around the equitability of relying on citation-based metrics for the allocation of grant funding, and the selection of panels to administer this process.

Hu, Wu, and Sun (2017), on the other hand, describe how certain characteristics of a journal — specifically, its “number of publications” and “amount of interdisciplinarity” — can be “improved” to decrease its uncitedness factor. Abramo et al.'s (2016) study of uncitedness rates among 3,386 “Top” Italian scientists found wide variation based on the authors' subject areas, and that authors with higher citation distribution are likelier to have lower uncitedness than the those with more concentrated distributions. (Read: Be wary of publishing too much on a single topic.) This emphasis on decreasing uncitedness is unsurprising, given the financial stakes at play. In a longitudinal analysis of the efficacy of public subsidies on published research in Turkey, Tonta and Akbulut (2020) considered both citation and non-citation as key indicators.

After all, as Abramo et al. (2016) state in their very first sentence, “One of the aims of universities and research institutions is to contribute to socio-economic development through the production of new knowledge.” Viewed through this lens, the identification or development of advantageous metrics (or goosing existing ones) is easily justified as an economic form of *realpolitik*. As long as informetrics are intertwined with economic incentives, we can expect scrupulous institutions to play whack-a-mole with evermore novel techniques for their distortion; For example, see the recent

busting of “citation cartels” in mathematics. (Catanzaro, 2024) If anything, this literature review highlights the ongoing struggle of academics working within a performance-based milieu: Without premeditated selection, most of the papers herein include at least some discussion of the problems inherent in using informetrics for this purpose, and academics are increasingly comfortable voicing their criticism. (Waitere, et al., 2011)

Sidebar: Sleeping Beauties

Slumbering in the lilac twilight between citedness and uncitedness are “sleeping beauties:” Papers which accrue little or no recognition for years, then are suddenly “awakened” into a bevy of citations by a “prince” (an influential paper citing the sleeper). Along with painting this delightful linguistic tableau — a much snappier replacement for “delayed recognition” — van Raan (2004) proposed a set of quantifiers for the “depth” of a paper’s sleep and its “awakening intensity.” Li et al. (2014) analyzed the “heartbeats” (annual citation rates) of sleeping papers prior to their awakening, finding that certain patterns correlate with a greater probability of a prince’s arrival. Sadly, they also found that if a paper is in a “vegetative state” — accruing no citations in a five year span — its “awakening probability is rather low.” Ho and Hartley (2017) studied the citation trajectories of three sleeping beauties in psychology, and Li and Shi (2016) examined the awakenings of sleeping “genius works” (papers with exponential citation accrual). Sleeping beauty research is perhaps the most optimistic (and dreamy) brach of informetrics: Your paper is *not uncited*, your paper has *not yet received true love’s kiss*.

Absence Upon Absence

But rarest of all are qualitative studies of uncitedness — they are vanishingly scarce,² and even to some extent orbit around citation-based frameworks, nimbly pointing out their blind spots.

Building on their earlier work considering uncited influences on published literature, MacRoberts and MacRoberts’ (2009) *Problems of citation analysis: A study of uncited and seldom-cited influences* includes several fascinating and worthwhile tangents. They begin by considering Szava-Kovat’s (2008)

² Searching Google Scholar for required words “uncited” and “qualitative” returned only two results, neither of which were relevant. SJSU’s OneSearch returned 13 results, none of which were relevant.

work on “nonindexed reference stock” (academic work influencing a paper but uncited within it) and his view that “The oppressive weight” of accrued literature has become “so vast that it is ... impossible to cite;” and briefly discuss Merton’s (1973) perspectives on “obliterated” work — that is, “work that has become so accepted that it is no longer cited.”

Their qualitative analysis begins with a study of 22 biogeography articles, whose bibliographies have been compared to their uncited influences; Two of the 22 papers are discussed in detail. The authors found, essentially, a thread of citation heavily favoring theoretical/analytical work, which itself cites mostly other theoretical/analytical work (and so on), noting that “if we did not know that biogeographic papers cited very few of their influences, we might come to the same conclusion as [Cole and Cole (1972)]: that only a small number of individuals contribute to scientific progress.”

Moving into an examination of the influence of an uncited paper and a “seldom-cited” paper,³ the authors discuss how the information therein has nonetheless been “quickly incorporated into the biological literature,” and elaborate on modes of knowledge transfer operating outside the capture of informetrics. (Put another way, there is a healthy, thriving universe of scholarly communication with a citation potential of zero.) The authors conclude that in biogeography, “most of the work used is not cited” and “biogeographers are not unique in their citing practices,” listing several fields of study and scientific methodologies which exist almost entirely without citation.

It is worth noting that a study like this would likely be impossible without the authors’ significant domain expertise: The lure of expediency is, perhaps, the simplest explanation for the dearth of qualitative studies on uncitedness. On the other hand, perhaps this knowledge gap is a subconscious form of empathic self-care, as we — authors — know all too well the silent dread of casting our work into the churning maw of The Discourse and hearing only silence in return. Or perhaps we harbor a creeping, subconscious fear that uncitedness is somehow virulent; That an intentional, serious engagement with uncited work can naught but tarnish our sterling academic reputations.

³ The former is a note published in *Phytologia* concerning a moss found outside its previously identified range; The later is a study of vegetation in an area of Texas with only a handful of citations.

Sidebar: The Demarcation Problem

And here, at the gooey caramel center of uncitedness is the demarcation problem, “one of the most high-profile, perennial, and intractable issues in the philosophy of science:” How to distinguish scholarly from un-scholarly work. (Resnik, 2000) Frandsen and Nicolaisen (2022) gamely stepped into this discussion, beginning with a high-level overview of the philosophy, before discussing attempts by Solla Price (1970), Windsor and Windsor (1973), and most recently Visser et al. (2021) to use a document’s reference count “as a rough proxy of [its] scientific contribution.” Frandsen and Nicolaisen sought to employ quantitative techniques to identify a natural demarcation breakpoint by comparing documents’ citation rates to their reference counts. The verdict? No dice. The demarcation problem remains quantitatively elusive.

Conclusion

Uncited authors inhabit a split-screen reality, with one side looping a reminder that their material survival requires putting numbers on the board, while the other soothingly coos “to be an uncited scientist is no cause for shame.” (Garfield, 1991) It seems counterproductive to put the individuals tasked with advancing human knowledge in this environment, even if it does make performance-based assessments more convenient for administrators.

No — let me begin again. In the years I lost to the haze of depression, there were days, in the descent of orange skies, of purple air, of a million acres burning, that I felt hopeless, immobilized in the suffocating grip of apocalyptic visions, collapsing time, the ambiguous loss of our future weighing like a stone on my chest, a nightmare where you see yourself, in slow motion, driving the car toward the ledge, unable to turn the wheel, the time for meaningfully addressing the anthropocenic collapse well behind us, the Rubicon irreversibly crossed,

No — let me begin again. Quantitative analytic frameworks are self-reinforcing: As the corporate bromide goes, “what gets measured gets managed.” As the total number of publications increases and time-strapped academics reach for convenient ways to filter the mass, we can expect to see readership and citations further consolidate to Q1 journals; authors reading authors everyone reads. And the designs of our web interfaces and pedagogy re-enforce this tendency.

The mind is a strange and unpredictable place; non-linear in its digestion and processes. Focusing on “improving” a journal’s uncitedness rates by reducing its interdisciplinarity is winning the battle but losing the war. The survival not just of our species, but of the culture, the art we’ve assembled, requires a different way of thinking about the value of what we ingest. Digital librarians are ideally-positioned to provide interventions here: The gilded levers of recommendation and UI sit plainly before us, but we have become complacent, pulling them in the same order purely out of habit.

No one knows what mix of input will coalesce into the ideas that save the world, and we should embrace that. Uncited papers, dusty theses, pulp novels, garbage TV — something about the struggle of drafting this longhand in a coffee shop blasting Wu Tang at six a.m. — the momentum of caffeine carrying me outside — the sunrise reflecting in vertical sheets of glass — breath in the lungs, the body moving itself uphill — a moment of transcendent joy — *real, transcendent joy* — almost psychotropic in its character — the mind experiencing itself unspooling, beautifully. For years, I lived without this — now finally balanced on medication and once again participating in the world, I find myself thinking of 18th c. symphonies overflowing their four movements into five, six, then exploding, like all art in the 20th, into a complete freedom of form, the fragmenting → fast → of Jorie Graham. Focus on this feeling. Remember how to feel it. Cherish its return.

References

- Abramo, G., D'Angelo, C. A., & Soldatenkova, A. (2016a). The dispersion of the citation distribution of top scientists' publications. *Scientometrics*, *109*(3), 1711–1724. <https://doi.org/10.1007/s11192-016-2143-7>
- Amez, L. (2012). Citation measures at the micro level: Influence of publication age, field, and uncitedness. *Journal of the American Society for Information Science and Technology*, *63*(7), 1459–1465. <https://doi.org/10.1002/asi.22687>
- Baykoucheva, S. (2019). Eugene Garfield's Ideas and Legacy and Their Impact on the Culture of Research. *Publications*, *7*(2), 43. <https://doi.org/10.3390/publications7020043>
- Cantanzaro, M. (2024). Citation cartels help some mathematicians—and their universities—climb the rankings. *Science*, *383*(6682). <https://www.science.org/content/article/citation-cartels-help-some-mathematicians-and-their-universities-climb-rankings>
- CheshmehSohrabi, M., & Mashhadi, A. (2023). Using Data Mining, Text Mining, and Bibliometric Techniques to the Research Trends and Gaps in the Field of Language and Linguistics. *Journal of Psycholinguistic Research*, *52*(2), 607–630. <https://doi.org/10.1007/s10936-022-09911-6>
- Cole, J.R., & Cole, S. (1972). The Ortega hypothesis. *Science*, *178*, 368–375. <https://doi.org/10.1126/science.178.4059.368>
- Frandsen, T. F., & Nicolaisen, J. (2023). Defining the unscholarly publication: A bibliometric study of uncited and barely cited publications. *Scientometrics*, *128*(2), 1337–1350. <https://doi.org/10.1007/s11192-022-04610-4>
- Garfield, E. (1991). To be an uncited scientist is no cause for shame. *The Scientist*, *5*(6), 12. <http://www.garfield.library.upenn.edu/essays/v14p390y1991.pdf>
- Hamilton, D. P. (1991). Research Papers: Who's Uncited Now? *Science*, *251*(4989), 25–25. <https://doi.org/10.1126/science.1986409>
- Ho, Y.-S., & Hartley, J. (2017). Sleeping beauties in psychology. *Scientometrics*, *110*(1), 301–305. <https://doi.org/10.1007/s11192-016-2174-0>
- Hu, Z., Wu, Y., & Sun, J. (2018a). A quantitative analysis of determinants of non-citation using a panel data model. *Scientometrics*, *116*(2), 843–861. <https://doi.org/10.1007/s11192-018-2791-x>

- Katchanov, Y. L., Markova, Y. V., & Shmatko, N. A. (2023). Uncited papers in the structure of scientific communication. *Journal of Informetrics*, 17(2), 101391. <https://doi.org/10.1016/j.joi.2023.101391>
- Kozłowski, D., Andersen, J. P., & Larivière, V. (2024). Uncited articles and their effect on the concentration of citations. *Journal of the Association for Information Science and Technology*.
- Li, J., & Shi, D. (2016). Sleeping beauties in genius work: When were they awakened? *Journal of the Association for Information Science and Technology*, 67(2), 432–440. <https://doi.org/10.1002/asi.23380>
- Li, J., Shi, D., Zhao, S. X., & Ye, F. Y. (2014). A study of the “heartbeat spectra” for “sleeping beauties.” *Journal of Informetrics*, 8(3), 493–502. <https://doi.org/10.1016/j.joi.2014.04.002>
- MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1–12. <https://doi.org/10.1002/asi.21228>
- Merton, R.K. (1973). *The sociology of science*. Chicago: University of Chicago Press.
- Miranda, R., & Garcia-Carpintero, E. (2019). Comparison of the share of documents and citations from different quartile journals in 25 research areas. *Scientometrics*, 121(1), 479–501. <https://doi.org/10.1007/s11192-019-03210-z>
- Nicolaisen, J., & Frandsen, T. F. (2019). Zero impact: A large-scale study of uncitedness. *Scientometrics*, 119(2), 1227–1254. <https://doi.org/10.1007/s11192-019-03064-5>
- Pakkan, S., Sudhakar, C., Tripathi, S., & Rao, M. (2021). Quest for Ranking Excellence Impact Study of Research Metrics. *DESIDOC Journal of Library & Information Technology*, 41(1), 61–69. <https://doi.org/10.14429/djlit.41.1.16328>
- Pan, R. K., Petersen, A. M., Pammolli, F., & Fortunato, S. (2018). The memory of science: Inflation, myopia, and the knowledge network. *Journal of Informetrics*, 12(3), 656–678. <https://doi.org/10.1016/j.joi.2018.06.005>
- Solla Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3–22). Heath.

- Szava-Kovats, E. (2008). Phenomenon and manifestation of the “Author’s effect of showcasing” (AES): A literature science study, I. Emergence, causes and traces of the phenomenon in the literature, perception and notion of the effect. *Journal of Information Science*, *34*, 30–44.
- Resnik, D. B. (2000). A pragmatic approach to the demarcation problem. *Studies in History and Philosophy of Science Part A*, *31*(2), 249–267. [https://doi.org/10.1016/S0039-3681\(00\)00004-2](https://doi.org/10.1016/S0039-3681(00)00004-2)
- Thelwall, M. (2016). Are there too many uncited articles? Zero inflated variants of the discretised lognormal and hooked power law distributions. *Journal of Informetrics*, *10*(2), 622–633. <https://doi.org/10.1016/j.joi.2016.04.014>
- Thelwall, M., & Wilson, P. (2014). Distributions for cited articles from individual subjects and years. *Journal of Informetrics*, *8*(4), 824–839. <https://doi.org/10.1016/j.joi.2014.08.001>
- Tonta, Y., & Akbulut, M. (2020). Does monetary support increase citation impact of scholarly papers? *Scientometrics*, *125*(2), 1617–1641. <https://doi.org/10.1007/s11192-020-03688-y>
- van Raan, A. F. J. (2008). Self-citation as an impact-reinforcing mechanism in the science system. *Journal of the American Society for Information Science and Technology*, *59*(10), 1631–1643. <https://doi.org/10.1002/asi.20868>
- Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41.
- Waitere, H. J., Wright, J., Tremaine, M., Brown, S., & Pausé, C. J. (2011). Choosing whether to resist or reinforce the new managerialism: The impact of performance-based research funding on academic identity. *Higher Education Research & Development*, *30*(2), 205–217. <https://doi.org/10.1080/07294360.2010.509760>
- Windsor, D. A., & Windsor, D. M. (1973). Citation of the literature by information scientists in their own publications. *Journal of the American Society for Information Science*, *24*, 377–381.
- Xia, X., Wang, Z., Wu, Y., Ruan, L., & Wang, L. (2014). Country of Authorship and Collaboration Affect Citations of Articles by South and East Asian Authors in Agronomy Journals: A Case Study of China, Japan, and India. *Serials Review*, *40*(2), 118–122. <https://doi.org/10.1080/00987913.2014.929610>