

# Measuring Transfer of In-Practice CS Literature through Patent-Research Citation Analysis

*Keywords: Patent; Citations Analysis; Knowledge Transfer; Innovation; Big Data*

## Extended Abstract

Only a few scientific research ideas yield practical applications, even though billions of dollars are invested in basic research to bridge the gap between research and practical uses. This investment has led to technological and social advances in the public and private sectors. Consequently, there has been an emerging line of research that investigates how academic research influences science and technological invention. The primary aims of such research are to gain insights into the type of research and development being conducted, to learn how inventors and scientists look for basic science that can be commercialized, and to learn the process of exploiting university inventions by firms.

The field of biology and chemistry has a noticeable amount of research on these aims. However, there is a lack of literature that measures the impact of computer science (CS) research beyond academia. It is unclear how much CS research is recognized in practice or the extent and rate of diffusion from research to inventions. Therefore, this study aims to fill this gap by conducting empirical analysis of CS literature through academic publications and studying how those papers are adopted in patent documents, which is a proxy for industrial practice. We aim to answer two questions within this project: (1) what percentage of CS research gets transferred into patents? (2) How long does it take for CS literature to be cited within inventions?

We used five datasets, four of which came from the United States Patent and Trademark Office (USPTO) PatentsView Database (-2022): the `g_application` dataset that provides information on granted patent applications, the `g_ipc_at_issue` dataset that includes action dates and classification information, the `g_inventor_disambiguated` dataset that provides inventor names, and the `g_us_application_citation` dataset that includes all essential information about citing (such as the order of reference being cited and whether the examiner or applicant cited it). The last dataset from Marx et al. (2020) connects each patent reference to USPTO (1927-2018) to academic papers (1800-2018) from Microsoft Academic Graph (MAG) through a disambiguation matching method. As MAG stopped operating during 2021 and the MAG-ID are not usable, we only utilized the DOI, month duration, and patent-ID. The benefit of combining the USPTO PatentsView Database with Marx et al.'s dataset is that most patents in the latter have either been granted or rejected, given that a patent's award time length is typically 3-4 years. To match self-citations by author names, we scrape the Crossref website and parse the author names.

The preliminary results reveal several key findings. Firstly, a number of academic papers are cited by patents, but the transfer rate is below 20% and varies across different fields. Secondly, the average time lag between patent award and paper publication is 7-10 years, also with variation across fields. For instance, the H04 class of Electric Communication Techniques has

the lowest time lag (<8.3 years), while H05 Electric Techniques has the highest (<13 years) (see Figure 1). Thirdly, CS patents tend not to rely on very old papers. For reference, the average time lag for Section C Chemistry and Metallurgy patents is within 12-16 years, and some extreme classes (e.g., C14 Skins and Hides) having nearly 20 years of lag time, with old papers being "recycled" frequently. Lastly, patents that have at least one author in common with the original paper tend to have lower time lags than those that do not, although this difference does not exceed a year in most classes.

The findings of our study suggest that patent inventions do not necessarily follow the latest research trends, possibly due to trade secrets that prevent knowledge from being publicly patented. It is interesting to note that, while the review time for patents has decreased over the years (see Figure 2), the time lag between patent award and paper publication has increased by 50%. Given the relatively low transfer rate, our study provides a valuable method for measuring the impact of computer science research and understanding its contributions to innovation in industry. To further promote the inventive step of a patent, we recommend that patent applicants consider using relatively new research. We also suggest that the USPTO encourage the use of newer literature during the reviewing process besides merely reducing review time.

## References

- Hsiao, T. K., & Torvik, V. I. (2019). Knowledge transfer from technology to science: The longevity of paper-to-patent citations. *Proceedings of the Association for Information Science and Technology*, 56(1), 417-421. <https://doi.org/10.1002/pra2.41>
- Li D, Azoulay P, Sampat BN. The applied value of public investments in biomedical research. *Science*. 2017 Apr 7;356(6333):78-81. doi: 10.1126/science.aal0010. Epub 2017 Mar 30. PMID: 28360137.
- U.S. Patent and Trademark Office. "Data Download Tables." PatentsView. Accessed [Jan 2023]. <https://patentsview.org/download/data-download-tables>.
- M. Marx & A. Fuegi, "Reliance on Science by Inventors: Hybrid Extraction of In-text Patent-to-Article Citations." forthcoming in *Journal of Economics and Management Strategy*. (<http://doi.org/10.1111/jems.12455>)
- M. Marx, & A. Fuegi, "Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles" (2020), *Strategic Management Journal* 41(9):1572-1594. (<https://onlinelibrary.wiley.com/doi/full/10.1002/smj.3145>)

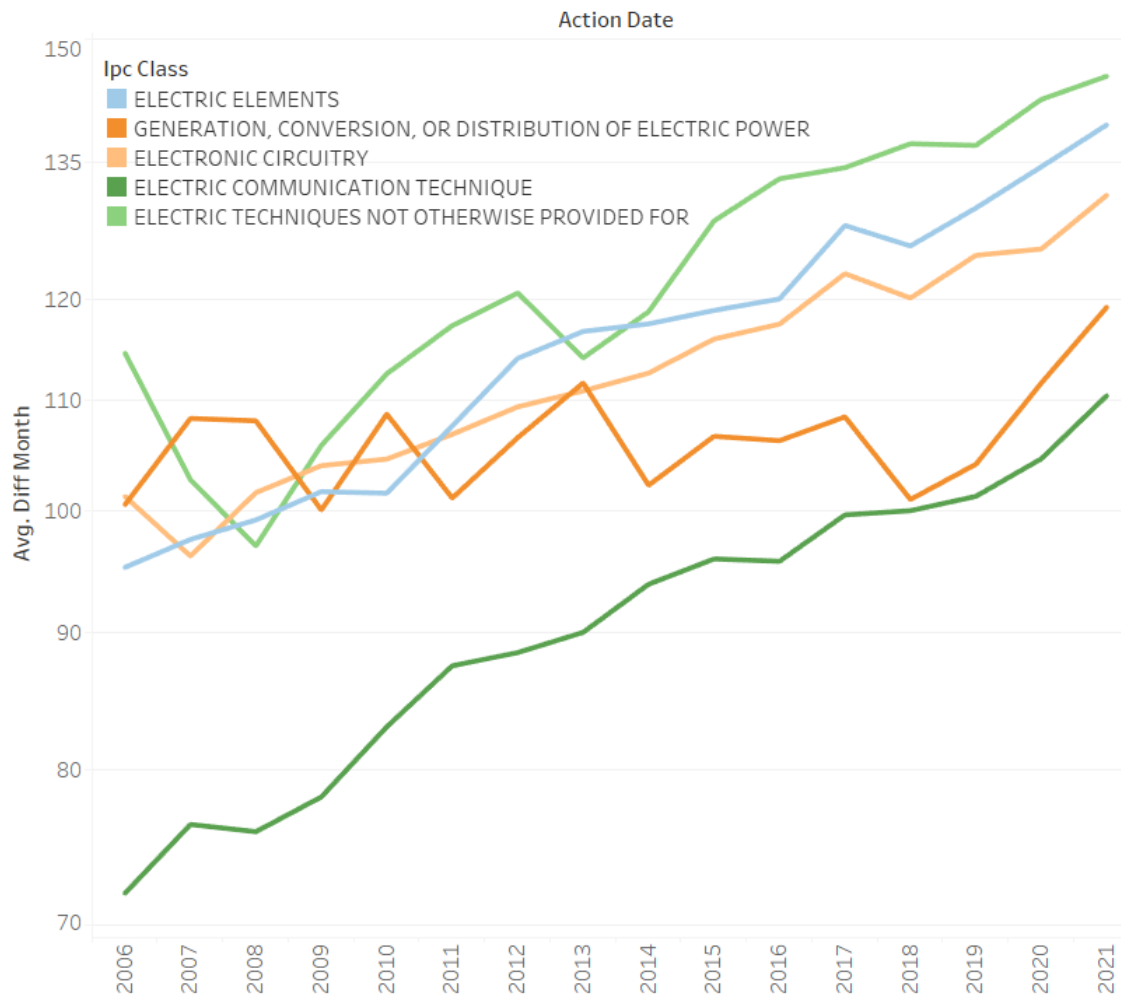


Figure 1. Average Time Lag of Literature-to-CS-Patents in order of Patent Action Date

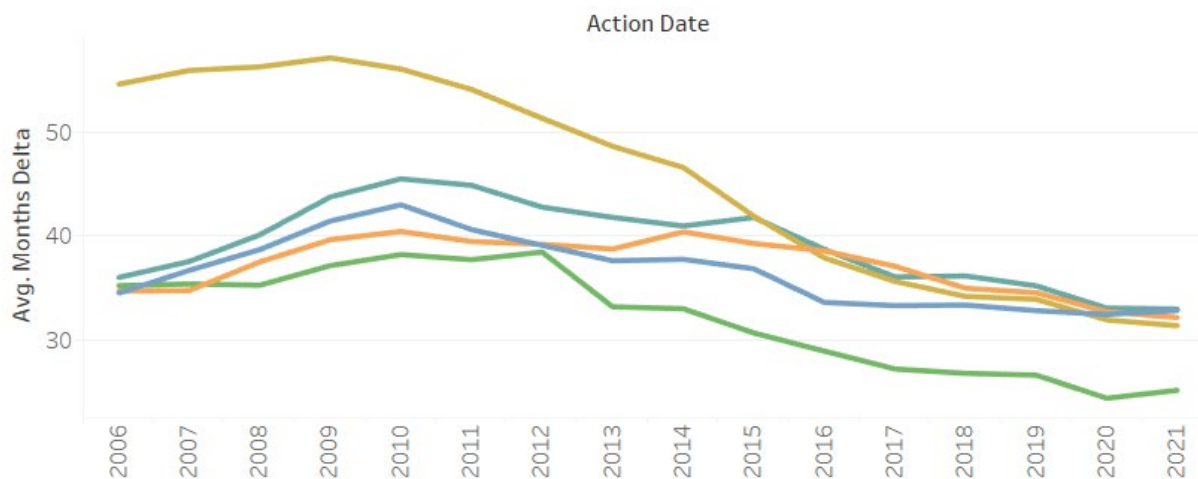


Figure 2. Average Review Time of CS Patents that is in the in order of Patent Action Date