

Uncovering the universal nature of citation networks: From science of science to law of law and patterns of patents

Keywords: Legal citations, bibliometry, knowledge-systems, collective citation model, science of science

Extended Abstract

Human culture is built on our ability to record and accumulate knowledge. Perhaps one of the most sophisticated examples is the scientific system. Science accumulates knowledge over time by building on existing work through citations, which allow scientific communities to compress and use existing knowledge. Examining how scientists cite existing work has revealed many insights into the ways scientists combine existing knowledge to produce new knowledge [1-4]. Although it is tempting to consider these insights as universal “laws” of citation, these patterns may be a result of the unique procedures and incentives of the scientific enterprise and may not generalize to other systems.

We explore the universality of citation dynamics by focusing on two additional, sophisticated knowledge systems – the common-law legal system and the U.S. patent system. All three systems – science, law, and patents – rely on citations to build on the past. In all three systems, citations are employed as the primary mechanism to draw upon the existing base of knowledge. While these systems are collaborative knowledge systems built on citations, they are distinct in terms of procedures and incentives. Anyone can attempt to publish science or file a patent and merit it determined at the time of publication, while judges are carefully preselected but can publish opinions without review; scientists and inventors choose their own research problems while judges are assigned to cases; the number of scientists and inventors has been growing rapidly while the entry to judicial system is limited; science aims to be egalitarian whereas the legal system has a codified hierarchy. These contrasting differences in how the systems are organized and operate provide an ideal opportunity to test whether the “laws” of citation can be generalized beyond science.

We show that, despite the stark differences between the three systems, the fundamental citation dynamics are remarkably universal, suggesting that the citation dynamics are largely shaped by intrinsic human constraints and robust against the numerous factors that distinguish the three systems. We demonstrate that the systems share similar characteristics across preferential attachment (Fig 1. F, J, N), citation recency (Fig 1. G, K, O), and diamonds in a rough (Fig 1. H, L, P). While some of the observed patterns can be explained by preferential attachment models, some emerge from collective behavior. We propose a new Collective Citation Model (Fig 1. A-D) that bases citation dynamics on the entire knowledge system, rather than just individual publications, and show that it gives rise to the empirical dynamics we observe. This model is able to better predict the trajectory of the most successful papers, the emergence of diamonds in the rough, and the growing number of references per publication. Our results build a strong bridge across three disparate systems, suggesting that theories and tools that describe human-based reference mechanisms (e.g., science, common law) can be translated into one another.

References

1. Wang, Dashun, Chaoming Song, and Albert-László Barabási. “Quantifying long-term scientific impact.” *Science* 342.6154 (2013): 127-132.
2. Uzzi, Brian, et al. “Atypical combinations and scientific impact.” *Science* 342.6157 (2013): 468-472.
3. Fortunato, Santo, et al. “Science of science.” *Science* 359.6379 (2018): eaao0185.
4. Price, Derek J. De Solla. “Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front.” *Science* 149.3683 (1965): 510-515.

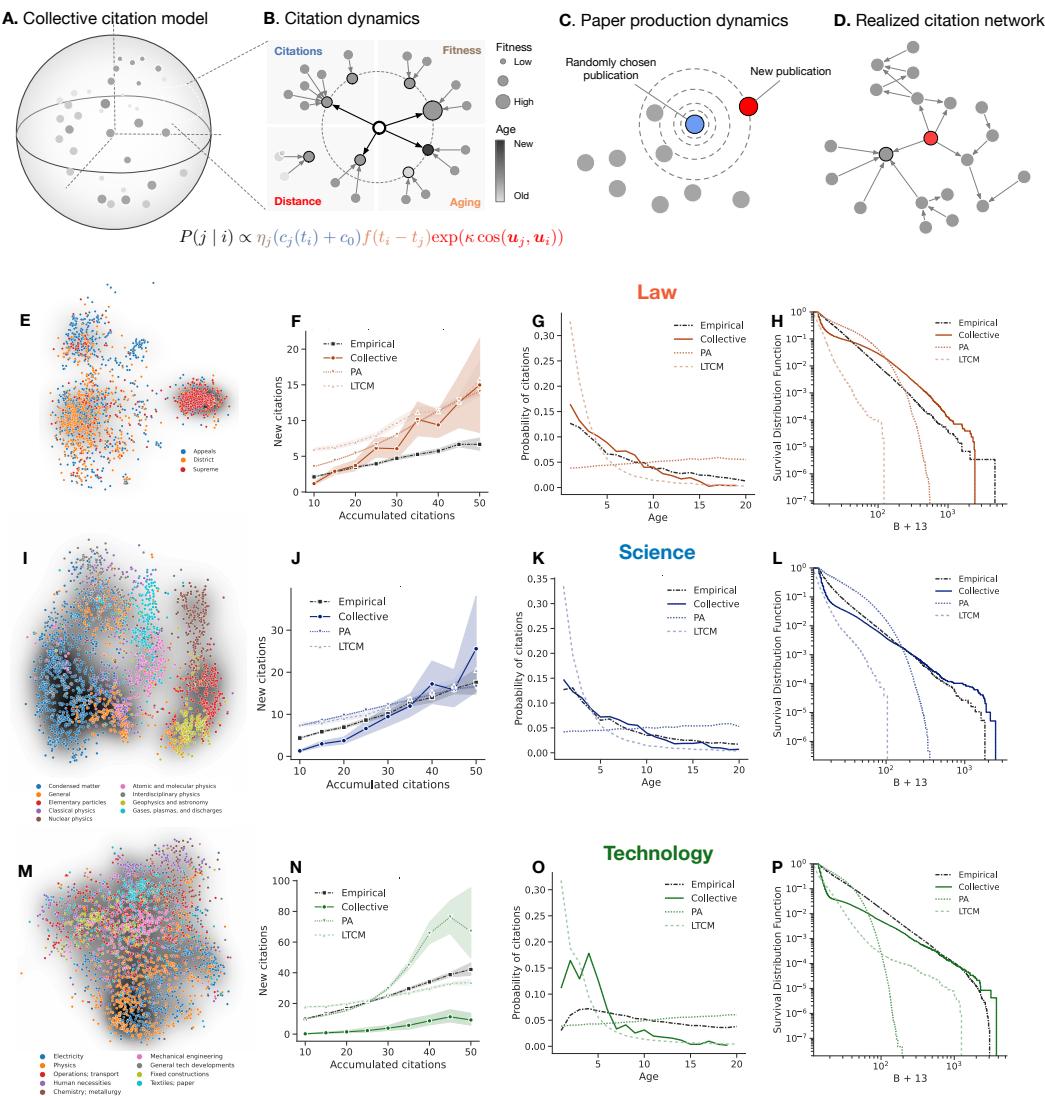


Figure 1. The Collective Citation Model: Authors of new publications aim to publish close to existing knowledge communities and papers cite each other based on closeness and preferential attachment (A-D). Embedding space for each knowledge system where points represent publications and the distance between points reflects the cosine distance in the original embedding space. Community structures form around topics (science, patents) and hierarchy (law) (E, I, M). Preferential attachment is observed in all systems and captured by both publication-centric and collective models (F, J, N). In all systems, older publications are less likely to be cited – this effect is captured by both publication-centric and collective models (G, K, O). In all systems, some impactful publications go initially unnoticed and then experience a surge of interest (diamonds in the rough), only the collective model captures this (H, L, P).