

Review of: "Strategic Citations in Patents: Analysis Using Machine Learning"

Pengfei Li¹

¹ University of Calgary

Potential competing interests: No potential competing interests to declare.

It is a great idea to apply machine learning in patent analysis and the paper has offered a lot of analyses to digest! With an original method to measure patent text similarity, you have discussed various topics, such as the relationship between patent text similarity and citation, localization of citation, self-citation, and citation after job change or city relocation. As these topics touch various dimensions of knowledge flows, I feel your ambitious analyses would fit in with several papers, rather than a single one. When focusing on a specific topic, you could engage more deeply with the specific stream of literature and discuss more implications of your findings with the new method.

In several places of your interesting results, I am wondering whether it is possible to interpret in a way other than strategic consideration. For example, first, you found an increase of patent citation while a decrease of text similarity of cited patents, and concluded, "These trends would indicate that the relevance of citations have been diluted by the addition of less related citations." Note your analysis covers a long period from 1985 to 2015. Imagine knowledge creation in a field. In the early days, due to limited knowledge, more general words may be used. Later, new words are created to describe our nuanced understanding of the field. In this perspective, we could expect patent texts would be more complex and specific over a long period. If so, would this imply a decline of text similarity of patents, unrelated to citation?

Second, you found patent citation increases with patent text similarity within cities but not across cities and interpreted the results from a strategic perspective that firms cite prior art in geographical proximity to reduce an infringement risk and omit relevant patents in other cities strategically. The explanation makes sense but requires an assumption that the risk of patent infringement is higher when firms are in the same city. The results may also be interpreted in another perspective. A firm or an inventor has cognitive restraints to know the prior art. In geographical proximity, such as cities, knowledge spillovers (Jaffe et al., 1993; Almeida and Kogut, 1999) so the inventor may become more aware of previous patents in the same city and cite them more likely.

Third, note there is a time window of patent citation. A patent received majority citations in its early years (Hall et al., 2001). In your analysis of self-citation before and after firm change, as the cited patents (A,B,C) are created at firm 1, due to time-lapse, we could also expect they would be cited less when the inventor moves to the second firm.

References:

Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917, 1999.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3), 577-598.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools.