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

Gianluca Zanella[1]

1 University of Texas at San Antonio

Potential competing interests: No potential competing interests to declare.

Thank you for the opportunity to review the paper "Strategic Citations in Patents: Analysis Using Machine Learning". The topic is relevant to the increasing body of literature on patent analysis and the ML approach in intriguing. I think the paper has a lot of potential, but I have some concerns that should be addressed to increase the relevance of the paper.

1. "Evidence of external influences on rate of local citations" - "Sample construction": I understand the process of finding potentially citable patents, but I think it needs more explanation. For example, did you analyze the date of publication to check if the patent was truly "citable"? In my knowledge, the legal offices check only published patents for similarities, and does not check the corpora of applications.
2. "Evidence of strategic omissions": The finding that "provide(s) evidence that external influences on the selection of citations tends to favour local citations overall" is intriguing. However, a potential explanation is a technical collaboration between small/medium companies that may be part of a technolgy hub. It would be interesting to check the finding against the size of the company because I believe big companies are not geographically constrained.
3. One additional concern is abouit the use of the p-value to support the hypotheses. As the seminal paper "ASA Statement on Statistical Significance and P-values" explains (http://dx.doi.org/10.1080/00031305.2016.1154108), the p-value should be coupled with estimates of confidence, credibility, or prediction intervals. In other words, the analysis should also estimate in some ways the effect size to support the hypotheses. BI attach some references that can provide ideas on how to improve the statistical validity.
4. The manuscript needs a troughout proof-reading to correct typos and syntax issues.

Good luck with your research.

References:

Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p< 0.05”. The American Statistician, 73(sup1), 1-19.

Sander Greenland (2019) Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values, The American Statistician, 73:sup1, 106-114, DOI: 10.1080/00031305.2018.1529625

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. Annals of internal medicine, 167(4), 268-274.