Review of: "A method to reduce false positives in a patent query"

Hung Tseng

Potential competing interests: No potential competing interests to declare.

I like the idea of using network analysis to clean up a collection of samples, in this case patents, using their intrinsic relations, in this case the technology domain classifications. Sample collection and sample purity is of general importance to virtually all research projects and experiments because they will influence, if not determine, the validity of the outcome (conclusion) of the research conducted, therefore this article is of general interest, in my opinion. I have not done patent analysis, but I do bibliographic analysis where research papers (not patents) are the subject of study. In bibliographic analysis a similar situation exists, i.e., a keyword search often results in a collection of research papers that is "contaminated" by irrelevant or false- positive papers, regarding the topic area of the analysis. The Co-citation network analysis (Small 1973) has been used to "cluster" papers, analogue to the "community" in this article, which in turn can identify some of the irrelevant papers. The co-citation network is built from papers that shared citations in their bibliography, which is considered an indication of the topic (conceptual) relationship among them, like the technology domain classification used in this article. Based on this comparison, I believe the idea presented here is reasonable and sound.

The article stops at presenting the idea, comes short testing it, except mentioning that the authors have been using it for sometime. Like the other reviewer (Tatiana Martins), I wish the authors could present some case studies where this idea is applied to actual patent analysis, and its effectiveness analyzed. My experience has been that network analysis (co-citation) cannot clean up all the irrelevant papers in a bibliographic sample and manual curation is often needed as a final step. The authors may have plans to do such analysis in the future, if so, this point can be disregarded.

The writing clarity of this article is probably sufficient for readers in the field of patent analysis. If the authors wish to make this article more friendly to readers outside this field (such as myself), thus enhancing its impact, they may consider spell out IPC (there are many IPCs from a google search), and explain its origin and meaning in a patent, because how a patent's technical domain is determined, can be critical for understanding this article and for the success of this method.

Henry Small (1973)
Co-citation in the scientific literature: A new measure of the relationship between two documents.
Journal of the American Society for Information Science
https://doi.org/10.1002/asi.4630240406