

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

Thomas Scheidsteger<sup>1</sup>

<sup>1</sup> Max Planck Society

**Potential competing interests:** No potential competing interests to declare.

The identification and removal of false positives is an important issue in the retrieval of patent data sets on, e.g., emerging technologies, but in removing them one should be reasonably sure not to remove relevant patents, too, especially if the researchers are not experts in the field investigated. The authors explain the main reasons for the occurrence of false positives using illuminating examples.

As practical solution proven helpful in their experience, they propose to build networks of connected classifications (e.g. IPC) in the patent data set and to analyse them with established methods. Thus, those classes that are not all connected to a core network component, can easily be identified, then checked for possible errors in the query and in that case removed.

The remaining classes are algorithmically assigned to communities which need to be examined intellectually concerning their relation to the core. In a last step, single nodes (e.g. IPC classes) are examined. Especially the role of being a "gatekeeper" as connection node between the core and other communities seems to be interesting.

In the conclusions, the authors point to some limitations that one has to keep in mind while applying the proposed method. To that end it would also be helpful for the interested reader to know the exact query leading to Figure 4.

Overall, I think this piece is a valuable contribution for practical scientometric work with patents.

Additionally, I have some suggestions for corrections:

- section 3, paragraph 1: "we use a classification network" appears to having been doubled
- Figure 2: the node in the bottom right corner should probably be "B82Y-030/00"
- section 4.1 : I count six, not seven components in the top right
- section 4.2, paragraph 2: the community "urinals" is displayed in the top right (not left)
- the reference to Bastian et al (2009) now has a DOI assigned:

<https://doi.org/10.1609/icwsm.v3i1.13937> ; the URL is no longer valid

