

Dear Editor,

Thank you for considering our manuscript "Classifying Patents Based on their Semantic Content" for possible publication in PLOS ONE. We are also very grateful to your suggestions and comments. This will undoubtedly be of great value to the paper.

We have read carefully your suggestions and comments, and have updated the paper accordingly. We provide you now the point-by-point response to the Editor and referees' reports.

We deal first with the Editor's comments.

1. We made adjustments to fully meet PLOS ONE requirements. All figures were converted and assessed using the PACE tool.
2. "Authors used co-occurrence of keywords to construct a patent network. Is this a new way? Or at least a discussion of the advantages should be provided."  
→ The use of co-occurrences to construct a semantic network has already been used, and is the best way to extract the endogenous semantic structure. We added a discussion on this point.
3. "Authors introduced a measure to correct the network topology. But how to properly determine the threshold? Why choose the value 0.06 ?"  
→ Thank you for pointing out this issue. We have been more specific about it now, adding three sentences. Indeed, the explanation at the end of Section 3.3 was sloppy.
4. "The research background include complex network analysis, community detection and data analysis, some recent progress in these areas should be reviewed"  
→ We included some of the references suggested in Section 3.5, as time series complex network analysis is indeed an interesting potential development.
5. "All the parameters should be clearly explained."  
→ Thank you for also pointing this out. We have been clearer on the definition of  $K_w$ ,  $k$  and  $t_i$  (Section 3.2),  $\theta_c$ ,  $\theta_w$  and  $\theta_w^{(0)}$  (Section 3.3).

We also provide you responses to Reviewer #3. Reviewers #1 and #2 did not ask anything specific but their comments were taken into account in the adjustments we made.

1. "Maybe a minor typo in Page 24 - "availability of these data" -> "this data" or "these datasets".  
→ We corrected accordingly.
2. "I also did not see a caption for the figures and it is quite hard to read the text in the figures due to their current small size. New readers are attracted to tables and figures, and thus it is useful to have descriptive captions - the current caption for Table 1 does not describe the variables being used (I know the text does it) - a brief description of theta being the likelihood estimates (or similar) would be useful."  
→ We have made the according changes. All captions inside figures were magnified as large as possible to ensure readability.
3. "I like the authors' approach overall, but would also recommend adding some discussion on how a semantic approach enables information integration and reuse - possibly with how their dataset /

ontology can be linked to others already existing in Linked Open Data. If such linking already exists, it should be shown - otherwise, this is a strong direction for future work."

→ This is indeed a very good suggestion and we add ideas of interesting potential developments by joining our database with existing open databases. We added accordingly a part to the discussion.

Yours faithfully,  
Juste Raimbault  
Université Paris 7 - UMR CNRS 8504 Géographie-cités