(TODO)

- *A rebuttal letter that responds to each point raised by the academic editor and reviewer(s). This letter should be uploaded as separate file and labeled 'Response to Reviewers'.*

- *A marked-up copy of your manuscript that highlights changes made to the original version. This file should be uploaded as separate file and labeled 'Revised Manuscript with Track Changes'.*

- *An unmarked version of your revised paper without tracked changes. This file should be uploaded as separate file and labeled 'Manuscript'.*

Reviewer comments :

(Editor) Authors extended some usual techniques of classification resulting from a large-scale data-mining and network approach. The analysis is interesting. There are many works about patent network analysis. 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. Authors introduced a measure to correct the network topology. But how to properly determine the threshold? Why choose the value 0.06. The research background include complex network analysis, community detection and data analysis, some recent progress in these areas should be reviewed, e.g., Europhysics Letters, 2016, 116: 50001; Scientific Reports, 2016, 6: 20052; Europhysics Letters, 2015, 109: 30005; Scientific Reports, 2015, 5: 8222. All the parameters should be clearly explained.

(TODO)

- *Add a discussion on the advantages of our approaches $\rightarrow$ already here no ?*

- *For parameter description and optimization, clarify and highlight more ?*

- *Add literature : we already cite what is necessary on the techniques used (modularity), I'm not sure adding what he says is relevant ; we should find a compromise*

(Reviewer 1) First authors have defined how they built a network of patents based on a classification method that uses semantic information from patent abstracts. Second, authors provide researchers with materials resulting from their analysis. An evaluation using the correspondence between citation links and classes to provide a measure of accuracy for classifications is performed.

Nothing to do for this one, no ?

(Reviewer 2)In this paper, the authors present a semantic approach to patent classification. This is an interesting approach and one I have not seen previously

applied to patents. The authors are correct that semantic analysis has been used in other limited studies over scholarly publications. Their decision to test their approach over the entire US Patent database is thus significant in both its scope and subject and lends credibility to the studys rigor and final appropriability. The study is both well-crafted and explained in appropriate detail. Although the primary function of a patent is legal protection, as a document they have been used to serve other functions both scholarly and social. Classification of these documents is thus complex and ever-changing. The approach detailed here thus adds to both the theoretical discussion of use of semantic analysis as well as being potentially applicable in the further study of science through patent analysis. Additionally, the paper is well written and the conclusions drawn by the authors are appropriate given the analysis.

Also nothing to do ?

(Reviewer 3) This is one of the rare occasions that I am quite satisfied reviewing a paper - this paper tackles an interesting challenge in semantic classification of patents based on their text content and meaning of text, it is written in good easy-to-understand English, the dataset/code used is available for reproducibility, and the evaluation methods are appropriately carried out. The appendices also form a useful addition to the manuscript. Maybe a minor typo in Page 24 - "availability of these data" –¿ "this data" or "these datasets". 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. 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.

(TODO)

- *typos*

- *table caption*

- *figures caption : are in text. text sizes. for figure 2, explain that the png is an illustration and that the same zoomable svg figure is available as supplementary.*

- *discussion on coupling with existing databases.*