Reviewer's report

Title: A Bibliometric Analysis on Tobacco Regulation Investigators

Version: 2 Date: 11 October 2014

Reviewer: Yaoyun Zhang

Reviewer's report:

This paper employed author topic modeling (ATM) on MEDLINE citations published by currently funded TRS principle investigators (PIs). Topics generated with ATM on dataset collected with a TRS keyword list and topics generated with ATM on dataset collected with TRS PIs are analyzed from various perspectives, such as topic distribution, author-topic relation, temporal trends, etc.

Overall this paper did a comprehensive analysis of authors and topics related to TRS. The findings based on this analysis may lead to a better understanding of the nature and scope of current relevant research.

Major revision:

1. In the “Author Topic Modeling Experiment” section, the sentence “Topic number T is selected as 400 for KWset and as 20 for TRSAwardeeSet respectively.”, why the topic number T is selected as 400 for and as 20 for TRSAwardeeSet? Given that KWset is selected by 300 tobacco-related keywords, and TRSAwardeeSet is selected according to 2,740 authors, what is the rationale behind the selection of topic numbers of these two sets?

2. In the “Summary” section, the authors stated that “Author topic modeling can associate topics to authors in a high accuracy.” Since all the analysis in this article is based on the output of the author topic model, an evaluation of the model on the two datasets should be conducted, and the performance is suggested to be reported forehead, so that to make the analysis part more convincing.

Minor revision:

3. In the “Top Topic Clusters” section, the sentence “The top topic clusters for all TRS investigators are metabolism, pharmacology, and legal & statistics in Figure 8...” A description of Figure 8 should be given first before this sentence to make it clear. Moreover, unfortunately, the reviewer cannot find these three topics clusters in Figure 8. Should it be Figure 9?

4. In the 9th page, the first sentence “At the first glance, Figure 6 is similar to Figure 5, aiming at showing how many authors appear in one topic.” seems to be confusing. A brief description of Figure 6 first would be clearer.

5. There are quite a few typos across the paper. A careful refinement of the article is needed before publication. Some typos are listed as follows:
“are shown are quite” should be “are quite”
“Figure 8” appears twice in the “Temporal Trend of Key Words” section. Are they referring to Figure 12 and Figure 13, respectively?
“the edges connecting the to three topics…” should be “the edges connecting them to the three topics…”
“The results can thus we used to assess …” should be “The results can thus be used to assess …”

6. The quality of figures is poor. An improvement of those figures in the final version is recommended.

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests