Reviewer's report

Title: Analyzing hidden populations online: topic, emotion, and social network of HIV-related users in the largest Chinese online community

Version: 0 Date: 24 Aug 2017

Reviewer: Jiang Bian

Reviewer's report:

The study is well motivated. Using data collected from online community to study hidden HIV population is an important topic.

1) The paper is easy to follow, but it will benefit from some professional editing. Some of the language and word choices feel weird. There are quite a few grammar errors as well.

2) Missing citations: e.g., "Currently, Baidu Tieba has more than 20 million bars and the number of active users has reached 300 million." Please cite the source of this information. Note this is just an example. There are other places.

3) "all records until August 2016", what is the start time? Since 2003?

4) Have the results of the sentiment analysis methods been validated? Sentiment analysis based on word counts is known for its inaccuracy.

5) How are the sentiment words extracted? Based on a known sentiment dictionary? Or the authors have reviewed all the words in the entire corpus to assign a sentiment score to every word? It is not very clear in the paper.

6) How is the number of clusters determined? The same case for how is the number of topics determined in LDA?
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal