Review of Automatic Controversy Detection in Social Media: A content-independent motif-based approach

Mauro Coletto, Kiran Garimella, Aristides Gionis, Claudio Lucchese,
Online Social Networks and Media, Issue 3, Volume 4, pages 22-31, 2017.

Summary

This paper deals about the problem of classifying a graph as controversial or non-controversial [1]. The author state that the majority of the research has been focused on utilizing the contents of the tweets, and less attentions has been on the network structure. They further argue that the their approach is independent of content and language which makes it versatile. The collect data from twitter, construct graphs, manually label them as controversial or not, and extract features as they hypothesize as relevant. The most novel aspect of the this work is using the motifs as a predictor of controversy. They conclude the paper with examining different classifiers, and exploring feature importance.

Reasons to accept the paper: Pros and Strength

- The manuscript is well-written, and easy to read.
- The idea is novel, and advances the field as most solutions solely rely on content and language features.

Reasons to reject the paper: Cons and Weakness

In my perspective, the research is well-performed, and I suggest the acceptance under major revision.

- There are some vagueness about the data collection process. The process makes the reader confused without providing details, and the correct Twitter data field names. Please read the comments 2 for more detail.
- There are some issue with interpretation of the observations since some of them don't fully narrate the observations. Please see the comments 3 and 5 for further discussion.

Comments (Recommendations)

The research is of great importance in the field of social networks, and controversy quantification. This specific field still has more room for improvement, and the acceptance of this paper helps nourishing the field. The idea is novel and creative. The manuscript is well-written, and comprehensible. If the authors can address the comments, the paper should be accepted.

Major Comments

1. The paper depends wholly on the "motifs" definition in the context of networks; yet authors only introduce this term without providing the reader with a complete precise definition. In addition, it would be great to provide a simple example of a motif in the context of social networks.
2. I have some comments on the data collection process on page 24 to enhance readability and reproduciblilty. I suggest the authors to rephrase this whole section, and provide more details using exact Twitter Data Field Dictionary to make it reproducible, and easier to understand. Another extra improvement is to accompany this section with a psudo-code algorithm for ultimate transparency.

(a) Labeling tweets' process is a bit ambiguous; as it is not clear whether the page labeling occurred first, and then tweet labeling. The other pathway is that the tweet labeling is performed before the page labeling.

(b) Another ambiguous part of the data collection is the twitter pages itself. Are they pointing towards a specific user who frequently posts controversial content? or are they simply a post not a user who generated controversial replies and reactions from other users. Since in one paragraph, it is stated they collected the last 200 tweets for each page. Did they mean replies or retweets? or did they mean the 200 last tweets of a user? In the following paragraph, they state that for each tweet in each page (root post) which implies it is a tweet not a user's tweeter homepage. So I conclude that they gathered controversial original tweets, and then collected the last 200 replies who sounded as controversial, then for each reply; the extended the conversational thread. This is a narrative problem which can be improved by a clearer description and the correct use of the data field and dictionary of Twitter's platform.

(c) The definition of the *user graph* on page 25, is based on $e = (u_i, u_j) \in E$ which means u_i and u_j being friends or if user u_i is following user u_j. However; it is vague if this information is also collected in the Data Collection process to construct the graphs. I suggest that this specific information is also included in the manuscript for clarity.

3. It is written that *"Fig. (2d) reports the distribution of the degree for the root, as well as the node with the larger degree excluding the root in T. We see that in this case the controversial and non-controversial discussions have similar distributions"*. However, it is not easy to confidently state that two distributions are similar as the boxes overlap with a small portion. I suggest authors provide a Q-Q plot for a more confident statement in the response letter. It's up to the authors to provide this information in the manuscript.

4. The legend in Figure 2 is incomplete. Although the authors refer the reader to the electronic version, the print version needs to stand on its own. So for a quick fix, I suggest adding the whisker and median colors, and as well as outlier shapes to the legend.

5. The box plots in Figure (2e) depicting "avg reply time" and "max reply time" shows that the reply time has less dispersion of controversial topics than non-controverial at least around approx. 2.5 times. However, it is stated that there is no significant difference between the distributions. In addition, it is nice to add the time metric (seconds or minutes or hours) for clarity. Furthermore, the number '1e6' showing the scale of the figure is tiny and difficult to read. It is better to increase the font size for visibility.

6. The feature which is defined as the number of closed triangles with the number of all possible triangles, on page 27, has a close relation with the term called *clustering coefficient* which is a well-known metric for graph density. It would be interesting that the authors explore this connection.

Minor comments

1. The sentence *"To classify them the content of the tweet and the received user replies were considered."* on page 25 sounds vague. I suggest to add a comma after *"them"* and before *"the content"* to make it easier to grasp.

2. The *user reply graph* is a directed graph according to the definition on page 25; however, the figures (1a) and (1b) do not seem to be directed. Probably it would have been too messy if the authors would have depicted the arrows. For clarity, I suggest that the authors state this in the caption that they have removed the arrows in the drawings for better representation; however, the graphs are directed.

3. The wiskers of the box plot are a bit invisible because they align with the maximum and minimum sides of the plot. I suggest users increase the line width to make them more visible.

References

[1] Mauro Coletto, Kiran Garimella, Aristides Gionis, and Claudio Lucchese. Automatic controversy detection in social media: A content-independent motif-based approach. *Online Social Networks and Media*, 3-4:22–31, 2017.