OUR RESPONSES TO THE COMMENTS OF EDITORS AND REVIEWERS

-Reviewer #1:

1. The authors propose a method for detecting trending topics on Indonesian tweets. In their experiments, they examine the effects of varying the number of topics, n-grams, stemming, and aggregation on the quality of the resulting trending topics.
2. Modeling of the problem with some proves are good presented in this paper. The author shows clarity of expression and the technical level of the paper is well presented. The paper also has an authoritative list of references.
3. There is still lacking on survey and analysis of related work especially on the benchmark approach and discuss their strength and weaknesses. There is no verification done between the proposed method and benchmark method.

Response:

In Section 2 (Related Work), we have revised the related work about document pivot and explained the verification done between the proposed method and benchmark method. In our paper, we only applied BN-Grams and Doc-P for Indonesian tweets. In the work of Aiello et al. (2013), BN-Grams and Doc-p have been compared but for English tweets. Detection of the trending topics in Indonesian tweets requires different stemming and stop words during the preprocessing stage. While Aiello et al. (2013) uses Porter stemming and English stop words, in this study, we use Adriani et al.’s (2007) stemming and Tala’s Indonesian stop words (Tala, 2003). The impact of the use of different stemming and stop words will be investigated in this paper. (We have revised Introduction, accordingly).

4. Twitter has its own trending detection mechanism, which is one of the several natural and essential features of Twitter. However, the paper doesn't mention Twitter's own trending feature at all, which means it also doesn't mention what differentiate this paper with Twitter's trending detection, or how is the performance of this paper's approach/algorithm, compared to Twitter's.

Response:

There are two differences between the trending topics of our paper and Twitter’s. First, the trending topics of our method is generated for general topics, but in this paper, we want to produce trending topics especially in political events. Second, trending topics of our method are shown with a set of keywords, while Twitter’s trending topics are formatted with hashtags and a set of keywords.

We have compared the trending topics of our method with Twitter’s trending topics in Section 4.4.5.

-**Reviewer#2:**

1. I think it is interesting approach for analyzing Twitter messages for obtaining the emotional reaction of persons and their sentiments. The graphics have a good explanation There are an excellent comparison between Doc-p. BN-grams and Doc-p in all three datasets.

Response:
Thank you very much for the suggestion. The suggestion for analyzing Twitter messages for obtaining the emotional reaction of persons and their sentiments will be studied in our next paper. In this paper, our scope is for comparing BN-Grams with Doc-p in Indonesian tweets for trending topics detection.

-Reviewer#3:

Overall this is a well-written paper on an interesting and important topic. However, the paper would benefit greatly from improvement in above mentioned aspects.

The manuscript would benefit from the following considerations:
1) The benchmark and comparisons
a. The benchmark with local news needs justification: The main thrust of the evaluation rests on comparing the results of the trending algorithms investigated by the authors with trending topics found in local news, as a bench mark. This is a reasonable approach but it is only one side of the story. There may well be trending topics that originate in social media and then get reported in mainstream news media. e.g. particularly in political news and elections. [See: ……] There is a two relationship: items in mainstream news can be reported in social media and vice versa, in other words.

Response:
We have compared trending topics of our method with Twitter's trending topics in Section 4.4.5.

The suggestion has been added in Section 5 (The experiments also indicate that trending topics generated by our method and trending topics in local news complement each other. The trending topics of our method form a material for the trending topics in local news, beside the tweets collected based on the trending topics in local news. There are two relationships: the trending topics in local news can be reported in trending topics of our method and vice versa.).

We have revised Section 4.1 (The ground truth consists of ten topics, built based on the trending topics in local news. The trending topics in local news is the most read news by news readers as called the most popular news. Ground truth contains a set of keywords based on the most popular news taken the next day after the trending topic is detected.).

There does not seem to be an acknowledgement of the above, in the paper. There should be some discussion and justification as to why the comparison with local news has been chosen.

Although it may be less likely, if there is any comparison of topics focused on breaking news originating from social media then that would be very good. Otherwise, a strong rationale of the approach, along with some acknowledgement of improvements possible / weaknesses would be good.

Response:
Breaking news is news that has either just happened or is currently happening (Phuvipadawat & Murata, 2010). While, trending topic, which is also called emerging trend or emerging topic (Becker, 2011), is a topic area that is growing in interest and utility over time (Kontostathis, Galitsky, Pottenger, & Phelps, 2004). Based on the definitions, trending topics and breaking news are two different meanings. Therefore, our paper does not contain a comparison of trending topics with breaking news.

b. The use of one expert needs justification and background: The paper states that one expert was used. Why only one and not 2 or 3 as is sometimes the case, and use the commonly agreed ones? This may be an issue of resource but awareness of this is important. What is the background of the expert. Is it an Indonesian news agency worker? An editor? Or is it someone expert in tweet headlines? Or is it a computer science student (in which case further discussion is needed)? Or is it a journalism student? Etc.

Response:
We have clarified in Section 4.3 (In this study, we employ two experts, namely a lecturer with a Ph.D. in political sciences, and an Indonesian news agency worker who has contributed to the three most popular news websites in Indonesia (Kompas.com, Tempo, and Detik.com).).

2) The language and cultural context: There is some good relevant referencing to other work on Indonesian tweets. However, more could be said on what is interesting to investigate and why with Indonesian tweets. Is there something about the language and grammar that we could learn and extrapolate to other languages? Is there something about the topics investigated that we should pay attention to when making comparisons in other languages, including English, for example?

- However, more could be said on what is interesting to investigate and why with Indonesian tweets

Response:
We have added in Section 1, Introduction (Detection of the trending topics in Indonesian tweets requires different stemming and stop words during the preprocessing stage. While Aiello et al. (2013) uses Porter stemming and English

stop words, in this study, we use Adriani et al.'s (2007) stemming and Tala's Indonesian stop words (Tala, 2003). The impact of the use of different stemming and stop words will be investigated in this paper.).

- Is there something about the language and grammar that we could learn and extrapolate to other languages?

Response:
We have added in Conclusion (The pattern in Indonesian writing is similar to the language pattern in the Indonesian subgroups: Melayu (Malaysia), Malagasy (Madagascar), Formosa and Philippines (Darmini, 2012). Therefore, Indonesian trending topics research has an excellent opportunity to be applied to trending topics in Indonesian language subgroup. Also, experimental results show that topics generated by BN-grams and Doc-P from Indonesian tweets do not have subject, predicate, object and adverb (SPOK) pattern, as Indonesian sentences should be; this will become a challenge for future research.).

Is there something about the topics investigated that we should pay attention to when making comparisons in other languages, including English, for example?

Response:
The focus of this study is limited only to comparing trending topics of our method with trending topics of mainstream news media and social media based on Indonesian tweets. Comparisons with other languages other than Indonesian will serve as input for future research.

3) Tweet examples: The paper would benefit greatly from more sample tweets, topics, and headlines – both in Indonesian and where possible English translation and phonetics. Table 3 seems to be the only example provided and that is only in Indonesian so it is difficult to conclude topically relevant assessments for the readers with limited or no knowledge of Indonesian.

Response:
We have revised in Table 3.

The language and style of the paper is clear.

Structurally the paper would benefit from a more focused discussion and analysis section prior to conclusions (which should be a highlight/summary of what has already been said). In other words, there could be a clear account of the results presented with the figures and tables as well as comparison with others' works and the implications. For example, this is done with the comparison and of results relating to stemming in section 4. More of this type of cross-referencing would benefit the paper. This can be done either within relevant parts in section 4 and/or by adding another section before conclusion as appropriate.

Response:
We have added discussion and analysis in Section 5.

Literature: There may be a few more papers the authors find relevant in last few years' ECIR, SIGIR, and JASIST papers on social media and news where the topic has featured more lately too.

Response:
During literature study, we did not find papers that are related to document pivot and feature pivot in the journals.

More information on the datasets is needed. Over what period were the tweets gathered for each collection? Were there any omitted and if so how/why? The authors state that keywords from 'political figures, executive agencies, …. ' were used. Were the same keywords applied for all collections or were some keywords used for one or two of the collections and not the other(s)? If so, why?

Response:
We have clarified in Section 4.1 (The datasets are constructed based on keywords from political figures, executive agencies, legislative assembly, judicial bodies, political event hashtags, names of the governor or vice governor candidates, and names of political parties. Moreover, in the absence of keywords relevant to political events emerging during the period of dataset collection, new keywords were added.).

The possible effect of the nature of the collection on any results discussed in section 4 should then be discussed accordingly too.

Response:
1. We have revised in Section 4.4.1 (Topic recall values increase as the number of topics increases. P1 produces more certain trending topics than P2 and P3. This is because P1 has fewer tweets and also has relatively shorter drawing period than P2 and P3. Thus, P1 has less difficulty and higher accuracy than P2 and P3.).
2. We have explained in Section 4.4.2 (Fig. 4 depicts the accuracy of trending topics detection by varying the n-grams used. The n-grams used are unigram to sixgrams. Trigrams produces higher accuracy than other n-grams. In P1 and P2 trigrams has the highest topic recall compared to unigram and bigrams; in P3 however trigrams produces lower topic recall than bigrams. This is because the number of tweets is smaller and the period of tweet collection is also shorter for P1 and P2 compared with those of P3.).

The authors refer to campaign accounts 'always generating same tweet for a specific purpose'. Work on the Scottish Referendum showed that different campaigns used their accounts and topics differently (see Pedersen et al further below).

Response:
We have clarified in Section 1, Introduction.

Actually, campaign accounts do not always generate the same tweet. However, campaign accounts in Indonesia generate different tweets, but still have the same meaning for a particular campaign objective. While in Pedersen et al.'s research (Pedersen et al., 2015a), Twitter was used to see the public response of a political debate aired on television.

Also, as in point 1) mentioned early above, what are the limitations of a collection that is based on news title form mainstream media. (i.e. this disadvantages consideration of news items discussed in society because of the good exposure in social media – even if it is not reported in mainstream media. So what is the implication of this?)

Response:
We have added in Section 5 (The critical finding in our experiments is the contrary between trending topics of our method and the trending topics in local news, which is evident in the local election of governor and vice governor of Jakarta in 2017. The trending topics generated by our method is a direct opinion of the society without any manipulation. Therefore, the trending topics produces by our method can be an early warning system for political events in Indonesia.).

Further clarity on the topics. The paper refers to 'The objective of the first category is to generate topics from an event, political movements, and the rate of urbanization..' (p2). Is this a reference to a category of topics that the authors have created/used in their work described or is this a category of topics mentioned in the literature that authors simply want to refer to or both? The authors refer to 'disruptive events, popular topics, and daily routines' (earlier on in p1). How do these categories relate to those? What exactly are the categories for each collection?

Response:
We have clarified in Section 1, Introduction (In this research, we want to apply BN-grams and document pivot for trending topics detection for general use, even though our case studies are limited only to detection of trending topics in the political field.).

Hence, as mentioned above, overall this is a well-written paper on an interesting and important topic. However, the paper would benefit greatly from improvement in above mentioned aspects.

References from other domains (media and journalism) that the authors may find relevant:

Pedersen, S., Baxter, G., Burnett, S., Göker, A., Corney, D., and Martin, C. (2015). Backchannel Chat: Peaks and Troughs in a Twitter Response to Three Televised Debates during the 2014 Scottish Independence Referendum Campaign.

Response:
We have added (Pedersen et al., 2015) in Section 1, Introduction.

Schifferes, S., Newman, N., Thurman, N., Corney, D.P.A., Göker, A. and Martin C. (2013) "Identifying and verifying news through social media: Developing a user-centred tool for professional journalists," The Future of Journalism Conference, 12-13 September 2013, Cardiff, UK.

Response:
Based on the literature study, this paper is too far from our topic.