False News Classification and Dissemination: The Case of the 2019 Indonesian Presidential Election

Rayan Suryadikara
r.suryadikara@umail.leidenuinv.nl

Suzan Verberne
s.verberne@liacs.leidenuniv.nl

Frank W. Takes
takes@liacs.nl

Abstract

In this paper we investigate automated methods for understanding false news dissemination on Twitter in relation to one particular event: the 2019 Indonesian presidential election. We collected a sample of 2,360 tweets related to topics addressed by fact-checking websites. The tweets were hand-labeled according to their trustworthiness. We trained several classification models on the human-labelled data, using three groups of text features. The word n-gram features appeared to be the most effective, reaching a recall of 85% for true news and 62% for false news. With this classifier we labeled a larger sample of tweets related to fact-checking topics in the context of the 2019 Indonesian presidential elections. We then analysed the dissemination of true news and false news in the underlying Twitter network using community detection and centrality measures. The top influential users in the network disseminate more false news, including a government institution account and a verified politician's account. Our results show that the combination of text features and social network analysis can provide valuable insights in detecting and preventing the dissemination of false news. Moreover, we make the dataset used in this research available for reuse by the community.

Title of the Proceedings: "Proceedings of the CIKM 2020 Workshops October 19-20, Galway, Ireland". Editors of the Proceedings: Stefan Conrad, Ilaria Tiddi

1 Introduction

A recent study strictly defined fake news as news articles that are intentionally and verifiably false and could therefore mislead readers [2]. In a political context the definition can be considered a bit wider. One study argues that politicians tend to label any news sources which do not support their positions as fake news [23]. This is especially common in the context of a large political event, e.g., an election. For example, there was an allegation that Joko Widodo was both a communist and Chinese in the Indonesia 2014 presidential election [10]. In this paper, we focus on the 2019 presedential election in Indonesia.

Social media flourishes as an alternative information source, in particular during elections, where many politicians utilize social media as means to reach out to the public more directly. Politicians prefer Twitter because of its efficiency in spreading messages, sparking conversations, building public opinion, or gaining support [19]. Especially in volatile political times, there are so-called buzzer teams that attempt to amplify messages and creates a "buzz" on social networks to spread positive content about one side of the political spectrum, while disseminating negative content about the other [11]. Hashtags are often used to increase their visibility to Indonesian Twitter users, which often become trending topics that then gain even more attention [11].

Because of these problems and their political impact, there is an urgent need to automatically identify and analyze false news in social media. This process could then result in the identification of the actors involved, as well as their networks that disseminated the false news. This research studies how false news can be detected based on the content of the messages posted, and then analyses its dissemination using social network analysis. The particular case that is considered is the 2019 Indonesian presidential election on Twitter, for which data was manually gathered and labeled in light of this study.

The contributions of this paper are:

- A new hand-labeled dataset of 2,360 tweets for the detection of false news in the Indonesian language;
- A method based on word features that can reasonably distinguish true news and false news in this data.
- An analysis of how true news and false news disseminate in the Twitter network related to the 2019 Indonesian elections, and what role particular communities, accounts, and hashtags play in the dissemination of false news.

The remainder of the paper is organized as follows. In Section 2 we discuss related work. In Section 3 we introduce the data and the annotation process. In Section 4 we present the methods we use, followed by experimental results in Section 5. Finally, the conclusions of the research are outlined in Section 6.

2 Related Work

In this section, we discuss work on false news on social media as well as methods for identifying this false news.

A recent study examined fake news from a political perspective, inspired by the 2016 US presidential elections [2]. They differentiated fake news and its close cousins in the political subject: unintentional reporting mistakes, rumors, conspiracy theories, satires, false statements by politicians, and slanted or misleading reports. The nature of the political world itself where a great number of critical reports have been discredited as fake news leads to redefining fake news which spread on social media [2]. A relevant study by Vosoughi et al. [23] focused on the veracity of Twitter posts which have been true or false.

In addition, they also defined news (either true or false) as any story or claim with an assertion in it, especially in social media. This extends the definition scope of false news from 'intentional' characteristics, allowing to incorporate aforementioned fake news' close cousins [2] into a single term. Therefore, the 'false news' term will be used throughout the paper which incorporates fake news and its close cousins.

In the text classification field for Indonesian language, most research focuses on hate speech identification. One of the first researches on Indonesian hate speech was conducted with multiple text features (character n-grams and negative sentiment) and classifiers (Naive Bayes, SVM, and Random Forest) [1]. This research and data set were expanded with adding abusive language and hate speeches' target and levels [8]. However, there has not been conducted research to detect false news in the Indonesian language, despite they are usually associated with hate speech.

A study analyzed Australia's Department of Immigration and Citizenship (DIAC) Twitter data to identify topics over the DIAC Twitter account and the spread of tweets, particularly the most retweeted tweets [26]. Another study further explored the analysis by taking the mention feature into account and term co-occurrence analysis with Korean Presidential Election on Twitter [18]. It marked the possibility to analyse the real political situation from the social network. On the other hand, one research utilized and built hashtag co-occurrence graph [24] to discover semantic relations between words in a tweet.

Another study [7] investigated filter bubble effects which tend to be generated by recommender systems that personalize and filter tweets via community detection. Regarding influential actors in a network, a recent study with the main topic is the 2014 Malaysian floods [14] utilized betweenness centrality to identify the potentially key Twitter users during information dissemination. Another study analyses false news based on the impact of emotion [5] or the profiling of Twitter users [4].

While these works present the analysis of filter bubbles or the influential users, our study will utilize actual true news and false news labels of news messages to assess which type of news is circulated inside certain communities and/or spread of particular influential actors.

3 Data

3.1 Data collection

For crawling tweets we use the GetOldTweets Library[1] to bypass the limitations of the official Twitter API. This allows us to to download historical Twitter data within a specific date range for a particular query. The queries we used for crawling Twitter data are built on topics that were published by two Indonesian fact-checking websites[2]. The tweets are in the Indonesian language. We gathered data from the first day of the 2019 Indonesian presidential campaign (September 23, 2018) to a week after the election result was publicized (May 28, 2019).

We selected 281 topics related to the presidential elections from the above referenced fact-checking websites with their corresponding supporting URLs. For each topic we created a query. For example, for the supporting URL that examines whether the 23 European Union ambassadors support Prabowo-Sandi or

[1] `https://github.com/Jefferson-Henrique/GetOldTweets-python`

[2] `https://cekfakta.tempo.co/` (Cek Fakta Tempo from Tempo) and `https://turnbackhoax.id/` (Turn Back Hoax from Mafindo)

not[3], we used the topic "European Ambassadors Support Prabowo" as the query to extract the relevant tweets.

To ensure alignment between the extracted tweets and the supporting URL, tweets from the first time the news aired in social media until its seventh day are selected. After removal of duplicate tweets, this resulted in a set of 8,784 tweets for the 281 topics. For annotation, tweets that one retweet, one like, and one reply, or less are removed resulting in a set of 2,360 that we use for annotation.

3.2 Annotation

We recruited 10 native Indonesian speakers to annotate the data. They do not have political job, political affiliation, or belong to a political party to facilitate the impartiality. Having 2,360 tweets as original data set, and two annotators per tweet, each annotator had to label 472 tweets.

The information provided to the annotators was the topic, the supporting URL, and the tweet text. One topic is linked to one supporting URL and to multiple tweets. We wrote an extensive annotation guideline for Indonesian false news and validated it in several short iterations before starting the actual annotation process. [4] Annotators are asked to assign one of four classes to each tweet:

- True: Tweets that relate to the topic and are true or accurate according to the supporting URLs;
- False: Tweets that relate to the topic and are false or inaccurate according the to supporting URLs;
- Misleading: Tweets that relate to the topic and have accurate information according to supporting URLs but lead to wrong conclusions;
- Other: Tweets that do not relate to the topic or are not discussed within supporting URLs.

While misleading news is sometimes considered a subset of false news, we decided to distinguish it separately for text classification. According to [21], misleading news tends to use correct facts and data, but how the news is delivered or how conclusions are drawn is false and therefore leads to the wrong interpretation. This is consistent with other definitions that misleading news conceives false facts by topic changes, irrelevant information, and equivocations to mislead the audience [22].

[3] `https://cekfakta.tempo.co/fakta/111/fakta-atau-hoax-benarkah-23-dubes-uni-eropa-dukung- prabowo-sandi`, determined to be false news

[4] The annotation guideline can be found here: `https://github.com/rayansuryadikara/false_news_detection_and_dissemination_analysis`

Class	Statistics
True News	896
False News	648
Misleading News	189
Other	627
Total	**2,360**

Table 1: The 2019 Indonesian Presidential Election News Data Set Size for Annotation

The annotation process was conducted in two stages. In the first stage, two annotators annotated the data. In the second stage, a third annotator (the first author of this paper) acted as a final judge for any tweet where two previous annotators disagreed. We analyzed the inter-rater reliability of the annotated data using Cohen's κ. Out of 10 annotator pairs, there are five pairs with moderate agreement ($\kappa = 0.41 - 0.60$), four pairs with fair agreement ($\kappa = 0.21 - 0.40$), and one pair with slight agreement ($\kappa = 0.01 - 0.20$). The highest κ score is 0.52 and the lowest is 0.07. As a whole, we obtain fair agreement with a mean κ of 0.33. The statistics of the annotated data are outlined in Table 1.

3.3 Network Data

We extract two different networks from our Twitter collection of 8,748 tweets. The first is the **mention network**. In literature, it is suggested that mentioning other usernames in a tweet represents a more direct form of communication than what is obtained from a network based on follower connections [18]. The second network that we create is the **hashtag co-occurrence network** The frequency of use for a hashtag indicates its popularity. In the 2019 Indonesian presidential election, there are certain hashtags created to support or oppose certain figures, such as `#jokowiamin` to support Joko Widodo, the incumbent, and `#2019gantipresiden` ("2019 change the president") to support Prabowo, the challenger.

The mention network is a weighted directed network where posting usernames are defined as the source and mentioned usernames are the target of a directed link. Link weight is determined by how many times the source username mentions the target username. The hashtag co-occurrence network is a weighted undirected network in which two hashtags are connected if they occur together in a tweet. Link weight is determined by counting how many times the tags co-occur.

In our experiments, we visualize the two networks to analyse how true news and false news spread in presidential election settings. For the network data, misleading news will be merged under false news to

keep it straightforward and to simplify the contrasting visualization between true news and false news. In doing so, we actually model both networks as a multigraph in which two nodes can be connected based on how often they communicate or co-occur in both true and fake news.

4 Methods

In this section, we first present our text classification methods using three different content-based feature sets (Section 4.1) and voting ensembles to combine the feature representations. Next, we present the network analysis features that we use to analyze the dissemination of true and false news in the Twitter network (Section 4.2).

4.1 Text classification

Features. For the content-based classification, we compare three types of features: orthography features, sentiment lexicon features, and word n-grams.

Social media such as Twitter is a common example wherein there the conventions of orthographies are sometimes lacking [6]. Therefore, orthography patterns are commonly used for social media analysis [8, 17]. We define five **orthography features**: counts of exclamation marks (E), question marks (Q), uppercase letters (U), lowercase letters (L), and emojis (M).

For **sentiment features**, we use the Indonesian Sentiment Lexicon (InSet) [9] which comprises 3,609 positive words and 6,609 negative words[5]. The sentiment scores range from -5 to 5, where negative scores indicate negative words and positive scores indicate positive words. Words with score 0 are disregarded since the lexicon excludes neutral category. Along with InSet, we use an Indonesian abusive lexicon [8],which comprises 126 words that are considered abusive.[6] Thus, we have three sentiment lexicon features: the positive word count (P), the negative word count (N), and the abusive word count (A). Before applying the sentiment lexicons, we apply stop words removal and text normalization[7]. The stop words dictionary is adopted from [20].[8] The text normalization dictionary comprises of 11,034 terms which are mapped to a normalized form. The dictionary is a continuous, collective work from researches [1, 8, 16] on the Indonesian language. In addition to lemmatization, the dictionary also facilitates Indonesian abbreviations, slangs, misspelled words, and even political figures' names. Therefore, the normalized form often consists of more than one word.

For the **word n-gram features** the text was lowercased, and URLs and punctuation were removed. For mentioned usernames and hashtags, we removed the @ and # symbols while the usernames and the hashtag words themselves were kept because both are instrumental parts of tweets to be identified and distinguished [13, 15]. Some of the usernames and hashtags are also included in the text normalization dictionary and therefore are normalized as well. We used six subsets of word n-grams to create vocabularies: Unigram, bigram, trigram, uni-bigram, bi-trigram, and uni-bi-trigram. In all n-gram feature sets we use tf-idf as term weight.

Classification models. We used the same classifiers as prior work on Indonesian text classification [1, 8]: Multinomial Naive Bayes (MNB), Support Vector Machines (SVM) with SGD optimization [25], and Random Forest (RF), all implemented in Scikit-learn. We used the default hyperparameter settings for each classifier. For SVM, this means that $C = 1$. For RF, the number of estimators is 100 with no maximum depth for the trees. The final precision and recall scores of each set of text feature are the average scores of these three classifiers. Meanwhile, F1 scores are calculated according to average precision and recall scores.

Voting ensembles. We assembled the results of from each experiment with different text features. The final precision, recall, and F1 scores of each ensemble follow the same approach with the text feature sets after the voting ensemble is performed. We use majority voting: the numbers for each label are compared and the most voted label is selected. If there is not one label with the most votes, the class will be determined according to a text feature that has the best performance. We construct two different ensembles: Ensemble I is arranged from all combinations of each feature, Ensemble II is arranged from the best combination of each feature.

4.2 Social network analysis

We aim to analyse how true news and false news spread between actors in the two networks described in Section 3.3. For visualization, we use Gephi [3], an open-source tool for social network analysis. While we do not directly model the precise diffusion of the news as the network evolves, we do believe that these two methods provide crucial insights in the reach of different types of news and the network effects involved in the process.

Community detection is a method capable of partitioning the network into communities (more

[5] https://github.com/fajri91/InSet

[6] https://github.com/okkyibrohim/id-multi-label-hate-speech-and-abusive-language-detection

[7] https://github.com/okkyibrohim/id-multi-label-hate-speech-and-abusive-language-detection/blob/master/new_kamusalay.csv

[8] https://github.com/stopwords-iso/

tightly connected groups with fewer connections to other communities). Here, we use the well-known Louvain modularity maximization algorithm to perceive the potential of filter bubble effects in a community [7]. Filter bubbles are a phenomenon in which a person is exposed to ideas, people, facts, or news that adhere to or are consistent with a particular political or social ideology, leaving alternative ideas unconsidered and in some cases outrightly rejected [12]. We propose to systematically identify every community to see the type of news circulating in that community.

Centrality measures assign a ranking to nodes in a network based on their topological position in the network. Here, we choose to use betweenness centrality to identify the most influential nodes. Betweenness centrality measures for a particular node how many other nodes are connected via a shortest path that runs through that node. Therefore for the mention network, the node or username acts as an important hub in receiving and spreading information to other nodes [14]. On an individual node level, betweenness centrality captures information from neighboring users who both consume and generate false news. For the hashtag co-occurrence network, the hashtag is also an important hub where it frequently co-occurs lot with other hashtags.

5 Results and analysis

We first present results on the comparison of the effectiveness of the three different text feature types (Section 5.1). After finding the most effective text features, we investigate the dissemination of true and false news, using the network analysis metrics (Section 5.2).

5.1 Results — text classification

Experimental settings. We evaluate our classifiers in two different types of experimental settings. The first setting is the data set with three classes, namely True News, False News, and Misleading News. The second setting is the data set with four classes: True, False, Misleading, Other, and Unclear (where the three annotators all assigned a different label) While the 3-class setting is easier for the classifier to learn, the 5-class setting is more realistic because it includes the tweets that are irrelevant but will occur in a real Twitter stream as well. We used a fixed random train–test split of the data for evaluation of the models, with 20% of the data for testing.

Comparison of feature sets. We find that in the 3-class classification, the best n-gram feature set is the combination of unigrams and bigrams; in the 5-class classification the best n-gram feature set is the use of bigrams alone. The best orthography feature set for the 3-class classification is the feature set with counts of exclamation marks, question marks, lowercase letters, and emojis; for the 5-class classification having the uppercase letter count instead of the question mark count is the most effective set. Of the sentiment lexicons, using a combination of positive and negative sentiment words gives the best results for both settings. The assemble of the best feature combinations performed the best in the 3-class, while the assemble from all feature combinations performed the best in the 5-class. We compare the best feature combination for each feature type in Table 2.

The table shows that the n-gram features outperform orthographies and sentiment lexicons in each setting and each class. The ensemble methods are also not able to improve over the n-gram features alone. Nevertheless, the ensembling method allows orthography and sentiment lexicons to be included as features in text classification with better performance than independently, especially from social media sphere.

Final quality of text classification With the best text features in the 5-class setting (which is more difficult, but also more realistic than the 3-class setting), we obtain precision scores of 55% for true news, 71% of false news, and 68% for misleading news. Recall is 85% for true news, 62% for false news, and 26% for misleading news. The low recall for misleading news is caused by the small number of items in this category.

We analyzed the full collection of 8,784 tweets where the unannotated data set (6,424 tweets) is labelled by the SVM classifier with SGD optimization in the 5-class setting with the best-performing feature set (word bigrams). We then do the social network analysis on the automatically labelled dataset, which we discuss in the next section.

5.2 Results — social network analysis

Table 3 shows the counts of nodes and edges (full network, and for true and false news) in the labelled Twitter networks. The last line of the table shows the number of communities. For the 10 largest communities, the distribution of true and false news by community as well as the top 10 influential actors are shown in Figure 1 and 2 for the mention network and in Figure 5 and 6 for the hashtag co-occurrence network.

The distributions are stacked column of true news and false news, listing the number of nodes and edges in each discovered community or actor (usernames for the mention network work and hashtags for the hashtag co-occurrence network). True news is defined by blue color while false news is defined by orange color.

In the visualization, communities are represented by colours and betweenness centrality determined node size, as shown in Figure 3 and 4 for the mentioned net-

	Features	True News			False News			Misleading News		
		P	R	F1	P	R	F1	P	R	F1
3 Classes	Uni-bigram	**0.730**	**0.903**	**0.807**	**0.811**	**0.692**	**0.747**	**0.830**	**0.246**	**0.380**
	EQLM	*0.374*	*0.512*	*0.432*	0.437	0.523	0.476	0.133	0.079	0.099
	PN	0.552	0.836	0.665	*0.299*	*0.221*	*0.254*	*0.064*	*0.044*	*0.052*
	Ensemble II	0.671	0.899	0.768	0.796	0.569	0.664	0.643	0.237	0.346
5 Classes	Bigram	0.562	0.790	0.657	**0.707**	**0.621**	**0.661**	**0.683**	**0.263**	**0.380**
	EULM	*0.354*	*0.285*	*0.316*	0.308	0.528	0.389	*0.051*	*0.035*	*0.042*
	PN	0.414	0.786	0.542	*0.455*	*0.179*	*0.257*	0.077	0.070	0.074
	Ensemble I	**0.551**	**0.849**	**0.668**	0.638	0.569	0.602	0.471	0.211	0.291

Table 2: Comparison of all text feature sets plus the ensemble methods. For each text feature type in each classification setting, only the most effective feature combination is shown. The evaluation scores are average scores over the three classifiers (NB, SVM, RF).

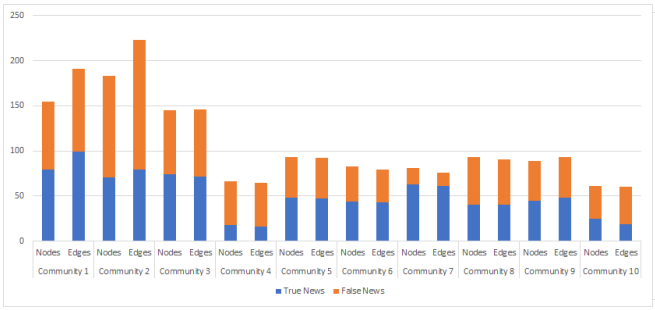

Figure 1: Distribution of true news and false news - top 10 communities of mention network

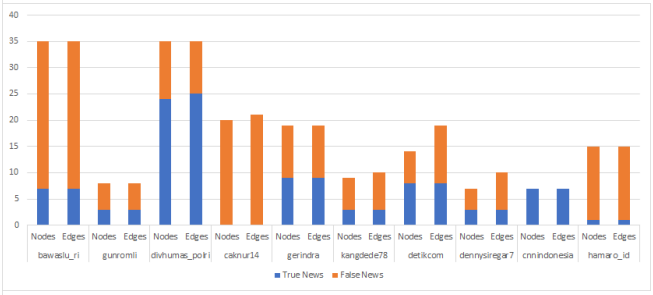

Figure 2: Distribution of true news and false news - top 10 influential usernames of mention network

Statistics	Mention Network	Hashtag Co-occurrence
# Nodes	1,891	1,302
# Edges	2,582	4,315
# True news edges	841	2,213
# False news edges	1,043	1,655
# Communities	165	133

Table 3: Network Data Properties

work and Figure 7 and 8 for the hashtag co-occurrence network. The visualization is formed by applying ego network to the ego (determined username or hashtag) within level 1 or its direct connection.

Mention network Based on the analysis of the mention network for the 2019 Indonesian presidential elections on Twitter, we find that:

- False news is more prevalent in the largest communities and also being disseminated and received more by top influential usernames. However, there are still more communities with a balanced proportion between true news and false news. Many news source accounts are found in these balanced communities.

- While the proportions of true news and false news are quite balanced in general, some usernames show a very strong tendency towards false news over true news, in particular a verified government institution account `bawaslu_ri` (shown in Figure 3 and 4) and two unverified accounts, `caknur14` and `hamaro_id`. One predominantly "true news" username is `cnnindonesia`, which is a verified news source account.

- Verified accounts tend to spread more false news than true news, where three of the top four influential usernames disseminate more false news than true news. The two largest, `bawaslu_ri`[9] (shown in Figure 3 and 4) and `gunromli`[10], are verified and politically-related account.

- One of the top "true news" influential usernames is `divhumas_polri`[11]. This is to be expected since

[9]The official account of an Indonesian government institution.
[10]The official account of an Indonesian politician.
[11]The official account of Indonesian republic police force.

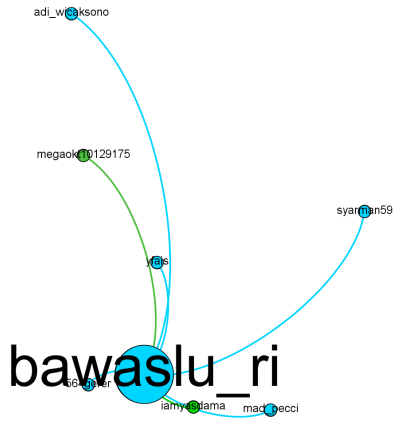

Figure 3: Network of `bawaslu_ri`'s - true news dissemination

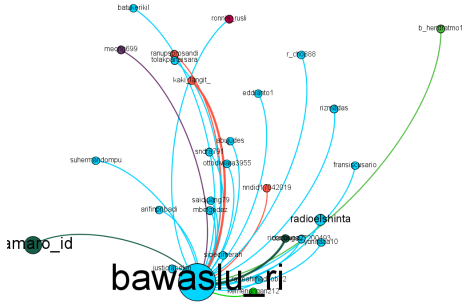

Figure 4: Network of `bawaslu_ri`'s - false news dissemination

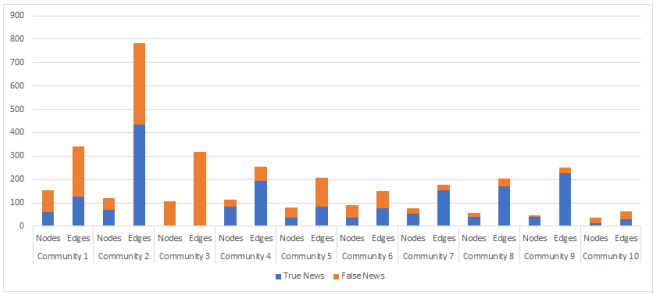

Figure 5: Distribution of true news and false news - top 10 communities (by size) of hashtag co-occurrence network

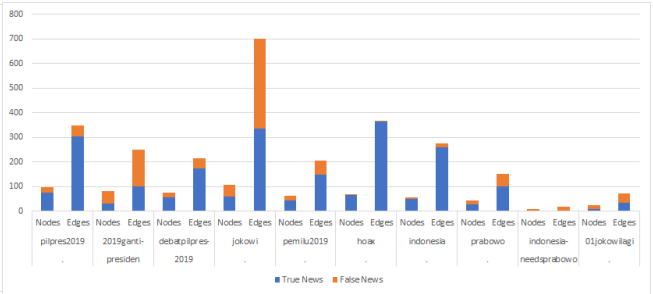

Figure 6: Distribution of true news and false news - top 10 influential hashtags of hashtag co-occurrence network

they have a cyber division dedicated to fight back hoax.

Hashtag co-occurrence network Based on the analysis of the hashtag co-occurrence network for the 2019 Indonesian presidential elections on Twitter, the interesting findings are:

- True news is more strongly associated with top influential hashtags.

- False news is more strongly associated with sentiment-induced hashtags than with hashtags about events or occurrences. Examples are `2019gantipresiden` (2019 change the president, shown in Figure 7 and 8), `indonesianeedsprabowo` and `01jokowilagi` (01 Jokowi again), which show support for both candidates. These results confirm the finding of previous work [5] that emotions are important in detecting false information.

- There is a community formed (Community 3) where only false news circulate in it. This community is filled with many slandering hashtags towards the incumbent Jokowi, such as `jaekingoflies` (Jae is one of derogatory title to Jokowi), `jaengibuldimanalagi` (Where does Jae lie again) and `uninstalljaenow`. However, none of them is a hashtag with enough influence.

- The inclined "true news" influential hashtags are very general terms and not directly about the presidential election, such as `hoax` and `Indonesia`. Hashtag `hoax` is especially noteworthy because any tweet which includes this hashtag mostly warns that the topic is a hoax, therefore fighting back hoax and is categorized as true news. The particular case of this hashtag was also outlined in the annotation guideline.

The mentioned-based network shows that the influential users are not only receive more false news, but also spread them as well. These usernames consists of unverified and verified ones, with the top two influential usernames are verified and "false news" inclined. This indicates that accounts with verification mark are not always clean from hoaxes.

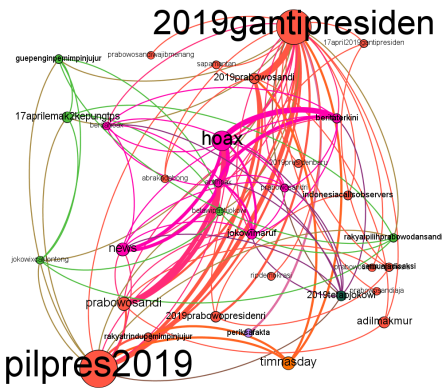

Figure 7: Network of `2019gantipresiden` - true news dissemination

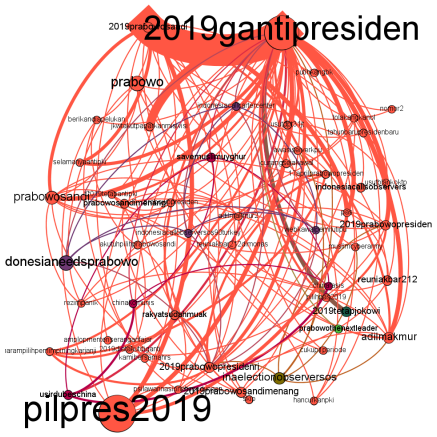

Figure 8: Network of `2019gantipresiden` - false news dissemination

Meanwhile, the hashtag-based network shows that supportive or sentiment-induced hashtags tend to relate more with false news, rather than more general events or terms. This indicates that these hashtags are more prone to information bias. Especially the supportive hashtags for each candidate, where users show fanatic support and attack the opposite candidate as well, often with false information.

As a reminder, these results illustrate the circumstances of the 2019 Indonesian presidential election event on Twitter. Furthermore, the news are selected based on fact-checking websites, which confirming circulating, trending topics on social media whether it is true or false.

6 Conclusions

In this paper we trained classifiers for detecting false news on Twitter and we analysed its dissemination related to the 2019 Indonesian presidential elections. We created a labelled dataset for true, false, and misleading news that we publish for use by other researchers.[12]

We found that the most prominent text feature to detect and distinguish true news, false news, and misleading news is word n-grams, in particular unigrams and bigrams. We also experimented with orthography features and sentiment features, but those did not improve the n-gram baseline. Nevertheless, the ensemble method allows the possibility to include and further refine these two text features in the future research.

From the social network analysis perspective, we found that the largest communities with top influential usernames tend to have more false news circulating rather than true news. Some of these influential users are also verified accounts. Regarding the hashtags, the hashtags that relate to explicit support of an election candidate occur more in false news messages than hashtags related to general events. These supportive or favouring hashtags tend to contain names or have strong sentiments.

In the 2019 Indonesian presidential election case, our results show that the combination of text features with social network analysis can provide valuable insights for the study of false news on social media. Hopefully these findings pave the way for not only detecting but also preventing the dissemination of false news in elections.

[12]The URL of the data repository will be added after anonymous peer review.

References

[1] I. Alfina et al. Hate Speech Detection in The Indonesian Language: A Dataset and Preliminary Study. *2017 International Conference on Advanced Computer Science and Information Systems*, 233–238, October 2017.

[2] H. Allcott and M. Gentzkow. Social Media and Fake News in The 2016 Election. *Journal of Economic Perspectives*, 31(2):211–236, May 2017.

[3] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third International AAAI Conference on Weblogs and Social Media*, March 2019.

Duan, Xinhuan, Elham Naghizade, Damiano Spina, and Xiuzhen Zhang. "RMIT at PAN-CLEF 2020: Profiling Fake News Spreaders on Twitter." CLEF, 2020.

[4] X. Duan et al. RMIT at PAN-CLEF 2020: Profiling Fake News Spreaders on Twitter. *CLEF*, 2020.

[5] B. Ghanem, P. Rosso and F. Rangel. An Emotional Analysis of False Information in Social Media and News Articles. *ACM Transactions on Internet Technology (TOIT)*, 20(2):1–18, April 2020.

[6] K. Gimpel et al. Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 42–47, June 2011.

[7] Q. Grossetti, C. Du Mouza and N. Travers. Community-Based Recommendations on Twitter: Avoiding the Filter Bubble. *International Conference on Web Information Systems Engineering*, 212–227, November 2019.

[8] M. O. Ibrohim and I. Budi. Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter. *Proceedings of The Third Workshop on Abusive Language Online*, 46–57, August 2019.

[9] F. Koto and G. Y. Rahmaningtyas. Inset Lexicon: Evaluation of A Word List for Indonesian Sentiment Analysis in Microblogs. *2017 International Conference on Asian Language Processing*, 391–394, December 2017.

[10] K. Lamb. Fake News Spikes in Indonesia ahead of Elections. `www.theguardian.com/world/2019/mar/20/fake-news-spikes-in-indonesia-ahead-of-elections`.

[11] K. Lamb. 'I felt disgusted': Inside Indonesia's Fake Twitter Account Factories. `www.theguardian.com/world/2018/jul/23/indonesias-fake-twitter-account-factories-jakarta-politic`.

[12] N. Lum. The Surprising Difference between Filter Bubble and Echo Chamber. `www.medium.com/@nicklum/the-surprising-difference-between-filter-bubble-and-echo-chamber-b909ef2542cc`.

[13] N. Naveed et al. Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter. *Proceedings of the 3rd International Web Science Conference*, 1–7, June 2011.

[14] A. T. Olanrewaju and A. Rahayu. Examining The Information Dissemination Process on Social Media during The Malaysia 2014 Floods Using Social Network Analysis. *Journal of Information and Communication Technology*, 17(1):141–166, January 2020.

[15] Y. Ruan et al. Prediction of Topic Volume on Twitter. *Proceedings of the 4th International ACM Conference on Web Science*, 397–402, 2012.

[16] N. A. Salsabila et al. Colloquial Indonesian Lexicon. *2018 International Conference on Asian Language Processing*, 226–229, November 2018.

[17] M. S. Saputri, R. Mahendra, and M. Adriani. Emotion classification on indonesian Twitter Dataset. *2018 International Conference on Asian Language Processing*, 90–95, November 2018.

[18] M. Song, M. C. Kim, and Y. K. Jeong. Analyzing The Political Landscape of 2012 Korean Presidential Election in Twitter. *IEEE Intelligent Systems*, 29(2):18–26, June 2014.

[19] S. Suraya and F. E. D. Kadju. Jokowi Versus Prabowo Presidential Race for 2019 General Election on Twitter. *Saudi Journal of Humanities and Social Sciences*, 4(3):198–212, April 2019.

[20] F. Z. Tala. A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia. *Institute for Logic, Language and Computation Universiteit van Amsterdam*, December 2003.

[21] Tempo. Metodologi. `www.cekfakta.tempo.co/metodologi`.

[22] S. Volkova and J. Y. Jang. Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media. *Companion Proceedings of The Web Conference 2018*, 29(2):575–583, April 2018.

[23] S. Vosoughi, D. Roy, and S. Aral. The Spread of True and False News Online. *Science*, 359(6380):1146–1151, March 2018.

[24] Y. Wang et al. Using Hashtag Graph-based Topic Model to Connect Semantically-related Words without Co-occurrence in Microblogs. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1919–1933, February 2016.

[25] R. G. J. Wijnhoven and P. H. N. de With. Fast Training of Object Detection using Stochastic Gradient Descent. *20th International Conference on Pattern Recognition*, 424–427, August 2010.

[26] Y. Zhao. Analysing Twitter Data with Text Mining and Social Network Analysis. *Proceedings of The 11th Australasian Data Mining and Analytics Conference*, 23–29, November 2013.