Universiteit Leiden

Master Computer Science

[False News Classification and Dissemination Analysis: The 2019 Indonesian Presidential Election]

Name: [Rayan Suryadikara]
Student ID: [S2432234]

Date: [28/08/2020]

Specialisation: [Computer Science and Science Communication and Society]

1st supervisor: [Dr. Suzan Verberne]
2nd supervisor: [Dr. Frank W. Takes]

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

LEIDEN UNIVERSITY

Abstract

Faculty of Science
Leiden Institute of Advanced Computer Science

Master of Computer Science

False News Classification and Dissemination Analysis: The 2019 Indonesian Presidential Election

by Rayan SURYADIKARA

False news has been intensely circulating and fooling its readers. It is especially prevalent in a political event, such as an election. In Indonesia, the political environment polarized people and fake news were frequently disseminated to flame the increasing tension, especially by social media such as Twitter. This research aims to identify and analyze the dissemination of false news in social media regarding one particular event: the 2019 Indonesian presidential election. The contributions of this research are classifying true, false, and misleading news by text features, examining the correlation between false news with hate speech and abusive language, and analyzing the dissemination of the 2019 Indonesian presidential news, particularly filter bubbles and influential actors. A sample of 2,360 tweets related to topics are collected and annotated by ten annotators who are Indonesian native speakers. This research found that the n-grams feature performed the best overall, while the ensemble method opened a greater possibility to incorporate the other inferior text features. Another finding is that abusive languages are more closely correlated with false news than hate speech. Finally, with the 2019 Indonesian presidential election as the main subject, the top influential usernames are prone to disseminate more false news rather than true news and the hashtags which have strong sentiments, either supporting or opposing, are more related with false news. The 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.

Acknowledgements

I would like to start by thanking the God. For my parents, my family, and my friends who have always been there with their support. For the annotators who participated in this research. For my supervisors, Suzan and Frank, who are encouraging and tireless in their insights and guidance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Fake news has been intensely circulating and fooling its readers, especially with the rising of disruptive social technologies such as social media (Tandoc Jr., Lim, and Ling, 2018). A recent study strictly defined fake news as news articles that are intentionally and verifiably false which could mislead readers (Allcott and Gentzkow, 2017). The main motivations are mostly financial or ideological. Flashy or even outrageous fake stories attract more attentions and clicks to inflate the revenue, or it can be launched to staunchly promote particular ideas, people, or affiliations while often discredit others (Allcott and Gentzkow, 2017).

However, the original purport of fake news has constantly deviated in the political arena. Fake news hides under a veneer of legitimacy and credibility by trying to appear like real news and also is often invoked to discredit some critical reporting from news organizations, further complicating in distinguish it from actual news (Tandoc Jr., Lim, and Ling, 2018). One study argues (Vosoughi, Roy, and Aral, 2018) that politicians tend to label any news sources which do not support their positions as fake news, whereas the ones that do are labelled as real news. This urges several researches to use false news term instead to conduct the term as objective as possible (Vosoughi, Roy, and Aral, 2018). This case is especially common in a huge political event, e.g. election. Few examples are Pope Francis' endorsement for then U.S.A. president candidate, Donald Trump, in 2016 (Silverman and Singer-Vine, 2016), and an allegation that Joko Widodo was both a communist and Chinese in Indonesia 2014 presidential election (Lamb, 2019).

In Indonesia, the political environment tension is severely exacerbated following Ahok's blasphemy case (Setijadi, 2017) . Basuki Tjahja Purnama (Ahok's real name), was an incumbent governor of DKI Jakarta, the capital city of Indonesia, who was lost decisively in 2017 DKI Jakarta gubernatorial election. This is a ramification over his public comments made in September 2016 that allegedly insulted the Al-Maidah 51 verse of the Quran and subsequently he was found guilty of blasphemy against Islam two months after his election lost (Setijadi, 2017). This incident aggravates sectarianism persecution to minorities in ethnicity and religion, flamed by fake news which were frequently hateful and abusive in nature (Azali, 2017). They were extensively spread on social media and messaging applications, particularly Twitter, Facebook and WhatsApp (Azali, 2017). Inevitably, this stream of false news kept maintained to 2019 Indonesia presidential election, where the incumbent president Joko Widodo (normally called Jokowi) is a close political ally with Ahok, have acted as DKI Jakarta governor with Ahok as his vice before ran for presidency in 2014 and won (Lamb, 2019).

Social media flourishes as an alternative information source regarding the presidential election due to print and television media could no longer able to maintain their independence and neutrality (Suraya and Kadju, 2019). Politicians also recognize this trend, and many of them utilized social media as means to reach the public more closely. They prefer Twitter because of its efficiency in spreading messages, sparking conversations, building public opinion, or gaining supports (Suraya and Kadju, 2019). Conversations that are widely discussed on Twitter can make its way to become a trending topic, where it will be easier for more people to access, read, and join the conversation thread.

Indonesia is one of the countries with the most social media users, emerged as the fifth-largest country in terms of Twitter users in 2017 (Herman and Mononimbar, 2017). However, it is estimated that roughly 15 percent of those users are bots account (Herman and Mononimbar, 2017). In especially volatile political times, there are buzzer teams that amplify messages and creates a "buzz" on social networks to spread positive content about the candidate they support while also disseminate negative ones about their opposition (Lamb, 2018a). The worst manifestation is Muslim Cyber Army, a network that spreads fake news and hate speech to inflame religious and ethnic and spread defamatory content to undermine the president (Lamb, 2018b). These buzzers also utilized hashtags to increased their visibility to Indonesian Twitter users, which often refers to trending hashtags (Lamb, 2018a).

Based off on these problems, there is an urgent need to identify and analyze false news in Indonesian social media, which some are induced by hateful or abusive intended words. This identification could lead up to recognizing the networks of usernames and hashtags which disseminate false news. Therefore, it will be possible to detect these accounts or hashtags right away to mitigate it in some way, either by authorities, Twitter regulators, or even individuals.

One previous research with Indonesian political topic was undertaken to detect Indonesian hate speech on Twitter (Alfina et al., 2017). This research covered hate speech for religion, ethnicity, race, and gender, where the data set from annotation is made available online (Alfina et al., 2017). This study on hate speech further developed by incorporate abusive languages and split hate speech based on target, category, and level, essentially assembled as multi-labelled problem (Ibrohim and Budi, 2019). However, in addition to having not explored true or false news realm, these approaches only focused on classify tweets by text features without considering the accounts and its networks, which is crucial to perceive how social media post spread.

A political-themed research analyzed the Korean Presidential Election by SNA approach which highlighted the importance of mention-based network and term co-occurrence aspects (Song, Kim, and Jeong, 2014). Meanwhile, another research studied the information dissemination on Twitter about Malaysia 2014 floods from SNA perspective (Olanrewaju and Rahayu, 2020). It focused on three aspects, namely the type of network structure formed and its density, the influential people involved, and the kind of information shared. These aspects could be adapted on the Indonesian presidential election problem where false news generating and reverberating accounts will be identified and analysed.

This research will study how to detect false news based on text mining approach and analyse its dissemination based on social network analysis approach regarding the 2019 Indonesian presidential election on Twitter. In general, the main research questions of this research:

- To what extent are the text features able to distinguish false news from true news and subsequently what is the most informative text feature for Indonesian social media posts?
- To what extent does false news correspond with hate and abusive speech in Indonesian social media posts?

- To what extent are network features able to provide information and visualization of false news dissemination about Indonesian social media posts?

As the preliminary information, this research associates "buzzers" with "influencers". Buzzer term has more broad and derogatory meaning, while this research particularly define the importance of username and therefore "influential usernames" term is more suitable. In addition, the influential usernames which spread true news are also observed, different with buzzers that mostly defined as false news spreaders. Therefore, this research is intended to be capable in drawing a similarity between the "false news" inclined influential usernames and the buzzers without grouping them together in the same definition.

This research is organized as follows. In Chapter 2 we will delve into false news background, its evolution within social media and political arena, and previous, related work as means to conduct the experiment. In Chapter 3 we will introduce the data and how it was retrieved, and the annotation process and how it was conducted. In Chapter 4 we will elaborate the methodologies, respectively to each experiment. In Chapter 5 we will present the experiment results, along with exhaustive analysis to answer the research questions. Finally, the conclusions and possible future directions of the research are outlined in Chapter 6.

Chapter 2

Related Work

2.1 False News Definition in Research Field

"Fake news" term has now become a murky area to pin an accurate definition of it. There is an exhaustive investigation to trace the origin of fake news definition according to other scholars and the types of fake news that have been identified in any published literature (Tandoc Jr., Lim, and Ling, 2018). They identified six types: satire, parody, fabrication, manipulation, propaganda, and advertising. Except for advertising, the rest of the types are closely associated with politics, for instance the fabrication that Pope Francis endorse Donald Trump as U.S. President (Silverman and Singer-Vine, 2016) and official Russian news channel called Channel One which was analyzed as a deliberate propaganda tool (Khaldarova and Pantti, 2016). They conclude the common characteristic across these definitions is to present false credibility in order to imitate real news.

Another study examined fake news from a political perspective, inspired by the 2016 US presidential elections (Allcott and Gentzkow, 2017). 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 (Allcott and Gentzkow, 2017). A study (Vosoughi, Roy, and Aral, 2018) focused on the veracity of Twitter posts which have been true or false, objectively referred the term as ’true news’ and ’false news’. In addition, they also defined news (whether they are true or false) as any story or claim with an assertion in it, especially in social media. It extends its own definition scope from ’intentional’ characteristics, allowing to incorporate fake news’ close cousins (Allcott and Gentzkow, 2017) into a single term. Therefore, the ’false news’ term will be used throughout the paper which incorporates fake news and its close cousins.

2.2 False News Evolution in Social Media and Political Domain

While a precise definition of social media is hard to define, in a general sense it is internet-based software applications that facilitate user-generated and user-based content and connections (Kelly and Autry, 2011). The development of social media with the increasing number of users and real-time connections inevitably leads to a large amount of data generation, due to information and update can be posted and shared from anywhere and at any time (Zubiaga, Liakata, and Procter, 2016). This is helpful for newsgathering in the journalistic field (Zubiaga, Ji, and Knight, 2013), criminal surveillance, and handling (Procter et al., 2013), as well as disaster management (Carley et al., 2016). In the political field, Barack Obama is perhaps the first and therefore the most well-known example of an electoral support mobilization via social media (Crawford, 2009). It is an effective vessel to collect and analyze public opinion on policies or candidates where there are much-disseminated information and political opinions shared and read (Stieglitz and Dang-Xuan, 2013; Zeng et al., 2010). These available opinions are mostly examined by sentiment analysis approach, whether to link a connection of public opinions from traditional polls and sentiment from posts (O'Connor et al., 2010) or directly predict election results by relevant features (Budiharto and Meiliana, 2018; Sang and Bos, 2012).

However, regardless of all these benefits, there is no denying that it enables questionable and outright inaccurate news to reach audiences without a credible filter. Particularly in Indonesia, these false news aggravate belief and identity sectarian tensions after or towards political events (Azali, 2017). An analysis suggested that the online spread of false news in Indonesia mostly consisted of political motives, reverberated by anonymous accounts or buzzers as a campaign tool to shift people's political preferences (Suraya and Kadju, 2019). These false news have almost always been equipped with hate speech and negative content, inadvertently inflaming persecution to minorities which impacted in the real world (Lamb, 2018a; Lamb, 2018b).

2.3 Previous Related Work on False News Identification

One of the early attempts to identify false news, specifically rumors, was based on its prominent aspects of diffusion (Kwon et al., 2013). They introduced temporal fluctuations to analyze rumor propagation aside from structural and linguistic aspects, which corresponds with rumour's characteristic of having a short cycle. Another attempt focused on a reporting series of breaking news which is mostly fast-paced and previously unknown (Zubiaga, Liakata, and Procter, 2016). They used a sequential classifier Conditional Random Fields (CRF) to learn the context during propagation time. A deep learning neural network approach was also performed for detecting fake news from text and images (Ajao, Bhowmik, and Zargari, 2018) and for coordinating a classical classification algorithm on lexical level (da Silva, Vieira, and Garcia, 2019).

2.3.1 Previous Related Work: Text Mining Approach

However, there has not been conducted research to detect fake news in the Indonesian language, at most hate speech and abusive language. 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) (Alfina et al., 2017). This research and data set were expanded with adding abusive language and hate speeches' targets, categories and levels (Ibrohim and Budi, 2019). A manual annotation process through human was conducted to initially assign the labels on each of their data sets. Both of the annotation processes emphasize the diversity of annotators to reduce bias, thorough annotation guidelines to ensure the annotators have a good understanding of the labels, and layered, multiple judgements to produce robust agreement results.

Along with word n-grams, orthographies and sentiment lexicon features were increasingly applied as alternatives to train the Indonesian language text classification models (Saputri, Mahendra, and Adriani, 2018; The, Wicaksono, and Adriani, 2015). Particularly for sentiment lexicons, there are some Indonesian lexicon dictionaries for sentiment built and improved, from Indonesian sarcasm sentiment lexicon (Lunando and Purwarianti, 2013) to Indonesian Sentiment (InSet) Lexicon (Koto and Rahmaningtyas, 2017). These three features were applied, measured, and compared together for the Indonesian hate speech and abusive language research. (Ibrohim and Budi, 2019). Other features were explored further (Ibrohim, Setiadi, and Budi, 2019) for the same topic, such as vectors generated from Word2Vec, part-of-speech tagging, and emoji features.

2.3.2 Previous Related Work: Social Media Analysis Approach

Twitter's nature as social media encourages researches to incorporate social network analysis alongside text mining techniques. One of the related preliminary researches 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 (Zhao, 2013). Another study further explored the potential analysis by taking the formation and modelling of a topic, mention network, and term co-occurrence analysis (Song, Kim, and Jeong, 2014). The study topic was also South Korean Presidential Election on Twitter, which 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 (Wang et al., 2016) to discover semantic relations between words in a tweet.

The community detection method has been exerted to examine and investigate echo chambers or filter bubble effects in a network. A research (Grossetti, Du Mouza, and Travers, 2019) investigated filter bubble effects which tend to be generated by recommender systems that personalize and filter tweets according to a community's interests. In addition, the research also developed a ranking strategy based on the communities' profiles to reduce the filter bubble effects. Regarding influential actors in a network, a recent study with the main topic is the 2014 Malaysian floods (Olanrewaju and Rahayu, 2020) utilized betweenness centrality to identify the potentially key Twitter users during information dissemination. Other than identifying key users, the research also studied the categories of exchanged information and the structural dynamics of the social network during the flood. This research will utilize the labels of tweets to assess which type of news is circulated inside certain communities and/or spread by particular influential actors.

Chapter 3

Data Collection & Annotation

3.1 Data Overview

In this research, the tweet data was crawled using Twitter Search API[1]. To stream the tweets, GetOldTweets Library[2] was used because its capacity to stream old Twitter data with date range and to bypass time constraints limitation of Twitter Official API. The queries which were used for crawling Twitter data are built on topics that were fact-checked by Cek Fakta Tempo from Tempo and Turn Back Hoax from Mafindo, two Indonesian fact-checking websites. The tweets are Indonesian language. The date range started from the first day of the 2019 Indonesian presidential campaign (September 23rd, 2018) to a week after the result was publicized (May 28th, 2019).

In more detail, queries to extract data are based on the topic of a supporting URL. In order to ensure that the extracted tweets relate to the news of the supporting URL as close as possible, the date per topic is set from the first time the news aired in social media until its seventh day. Suppose a supporting URL that examines whether the 23 European Union ambassadors support Prabowo-Sandi or not, which was determined to be false. The topic is "European Ambassadors Support Prabowo" and afterwards these words were set as the query to extract the relevant tweets. The news is recorded to be aired for the first time in social media realm on 19th of January, therefore the date range was set from 19th to 25th of January.

The duplicates of a tweet are removed where the one that remains is the tweet with the most retweets, likes, and replies, respectively. Retweets are prioritized because they enable a post to have a greater reach of audience, followed by likes and replies (Mavrck, 2014). The size of final data set is 8,784 tweets over 281 topics. For annotation, a filtering is applied to dismiss any tweet which has the number of retweets and likes and replies equal or less than one, resulting in a set of 2,360 tweets. The complete list of topics with extraction dates is supplemented in Appendix A.

3.2 Annotation Guideline

Annotation files are created by spreading the data set into ten annotation spreadsheet files. The files are arranged in such a manner that a half data of a file intersects with a half data in another file, thus every file has two halves of data set which intersect with halves from two different files. The annotators are 10 native Indonesian speakers, consist of of 8 males and 2 females within age range 21 - 32 years. They mostly have a bachelor degree for the last education, with one person for each a senior high school degree and a master degree. They also does not have political

[1] https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets

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

affiliation or belong to a political party to facilitate the impartiality. Having 2,360 tweets as original data set to be spread into ten files, each file has 472 tweets. This will ensure that every tweet will be annotated by two different annotators.

The columns of each file are "Topic", "Supporting URL", "Tweet", "True News", "False News", "Misleading News", and "Other". One topic is linked to one supporting URL and one to many tweets. Annotators are tasked to decide a class for a tweet based on the following definitions:

- True: Tweets that relate to the topic and are true or accurate according to supporting URLs.
- False: Tweets that relate to the topic and are false or inaccurate according 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 a one subset of false news, this research attempts to represent the class and draws its distinction from false news. According to Cek Fakta Tempo in methodology section, misleading news tends to use correct facts and data, but how they deliver or conclude is wrong and therefore leads to the wrong interpretation (Tempo, 2018). This is consistent with other definition that misleading news conceives false facts by topic changes, irrelevant information, and equivocations to fool the audience (Volkova and Jang, 2018).

The annotation process will be conducted in two stages. In the first stage, two annotators will annotate the given pair of tweet data. Afterwards, in the second stage, a third annotator will act as a final judge for any tweet where two previous annotators dissented. There are two sequential conditions for a tweet to be exempted. The first condition is when the first two annotators agree to classify it into Other class, where Other is for tweets that are not associated with the topic and therefore not discussed in the given supporting URL. The second condition is when the judgement from the third annotator dissents with either of two annotators' as well.

Before examining the tweets, each annotator is advised to read the supporting URLs for a moment to understand the trace of the facts. Supporting URLs are the only place for annotators to verify the tweet, so there is no need to manually look for other supporting evidence. Each annotator checks whether the tweet is a statement or claim mentioned in the exposition of the supporting URL. What needs to be checked is only claims or statements that are related to the topic or discussed in the supporting URL. Hence, if there is a tweet contains several other claims that are not related to the topic, then these claims are unnecessary to be checked. These limitations are principal to regulate the focus of annotators and to ensure that they will not deviate too far from the topic.

For instance, let us review once more the topic "European Ambassadors Support Prabowo" with supporting URLs Is it true that 23 EU Ambassadors Support Prabowo - Sandi?. The article concludes the rumor is false. As a summary, the twenty three EU Ambassadors visited Prabowo-Sandi BPN (Name of Prabowo-Sandi campaign team) headquarters to discuss various issues such as tax reform, openness for business and investment, social welfare issues and the permanent voters list polemic issued by the KPU. Afterwards, the EU ambassadors also visited the

Jokowi-Ma'ruf TKN (Name of Jokowi-Ma'ruf campaign team) headquarters to discuss European Union's partnership with Indonesia. It was stated that there was no discussion about support for candidate pair 02 during the visit.

Some examples of tweets based on this topic:

- "Prabowo has never sought support from the European ambassadors"
  The main statement relates to the topic and states refutation that Prabowo is seeking support from the European ambassadors. According to the conclusion of the supporting URL, there is no dialogue of support for candidate pairs 02, so this tweet belongs in True class.

- "The ambassadors have supported Prabowo by asked him to declare his vision and missions!"
  The main statement relates to the topic and states that the ambassadors have supported Prabowo and asked him to declare his vision and missions. According to the conclusion of the supporting URL, there is no support declaration from the ambassadors, only dialogue on tax and business matters, so this tweet belongs in the False class.

- "The EU ambassador instead came to Prabowo's headquarters, what else if it were not to give moral support to the campaign!"
  The main statement relates to the topic and states that the EU ambassadors came to Prabowo's headquarters to provide support. According to the conclusion of the supporting URL, the visit of the EU ambassador is factual, but it misled with the wrong conclusion because it has been explained and refuted that there is no campaign support, so this tweet belongs in Misleading class.

- "Prabowo was appointed as Indonesian ambassador to the Netherlands in the meeting of the European Union Ambassador."
  The main statement does not relate to the topic or is not addressed in the supporting URL, so this tweet belongs in Other class.

- "Prabowo has received a visit from the Saudi Ambassador https://news.detik.com/berita/d-4781593/usai-bertemu-dubes-as-prabowo-terima-kunjungan-dubes-saudi"
  Although the tweet includes a link that could give grounds for the statement, it can not be used as a guideline because the only valid link to check the veracity is simply the supporting URL and already supplemented for the annotators. In this case, the statement does not even relate to the topic or is not addressed in the supporting URL, so this tweet belongs in Other class.

- "Yesterday Prabowo was able to gain support from the ambassadors, Europe has begun to target domestic assets, China and America also will not remain silent."
  The statement that relates to the topic is Prabowo gained support from the ambassador, which is proven wrong, so this tweet belongs in False class. Other statements, such as Europe began targeting domestic assets or Chinese and Americans do not need to be checked because they are independent statements that do not relate to the topic.

- "[FALSE] EU Ambassador supports Prabowo Campaign"
  The statement relates with the topic and confirms the rumour that EU ambassadors support Prabowo's campaign is false according to the supporting URL, so this tweet belongs in True class.

3.3 Annotation Result & Final Data

The annotation process proved to be quite a challenging task. Out of 10 pairs, there are five pairs with moderate agreement, four pairs with fair agreement, and one pair with slight agreement as shown in Table 3.1. The highest Kappa score is 0.5208 while the lowest is 0.0721. As a whole, the Kappa score falls into slight agreement with 0.3333. The total data for each class consist of 643 True News, 316 False News, 75 Misleading News, and 189 Other.

Pair of Annotators	Observed Agreement	Cohen's Kappa
A1 - A2	0.610	0.450
A2 - A3	0.458	0.255
A3 - A4	0.394	0.225
A4 - A5	0.674	0.521
A5 - A6	0.572	0.407
A6 - A7	0.500	0.270
A7 - A8	0.292	0.072
A8 - A9	0.644	0.506
A9 - A10	0.470	0.306
A10 - A1	0.568	0.424
As a Whole	**0.518**	**0.333**

TABLE 3.1: Cohen's Kappa of Agreements

After the first stage annotation, the second annotation was conducted for differing tweets. The processed data from the second stage consist of 253 True News, 332 False News, 114 Misleading News, and 438 Other tweets. Overall, the total data set to be trained are 1,733 data, consist of 896 True News, 648 False News, and 189 Misleading News, shown in Table 3.2. Other tweets contain tweets which both annotators in the first stage agreed to pin it in Other class and tweets that the third annotator judged differently from previous annotators.

Class	First Stage	Second Stage	Total
True News	643	253	896
False News	316	332	648
Misleading News	75	114	189
Other	189 (Others 1)	438 (Others 2)	627
Total Data Set (Without Other)			**1,733**
Total Data Set (With Other)			**2,360**

TABLE 3.2: The 2019 Indonesian Presidential Election News Data Set Size

There are few remarks which can shed light on the challenges, as shown in Figure 3.1. The figure presents the distribution of agreements percentage from each category. Each cluster, which are denoted by established categories, maps percentage values to each categories. Therefore, each cluster has one bar agreement and three bars of disagreement. Out of all categories, True News (TN) is the category with the least disagreements on the rest of categories, while also has the highest agreement. False News (FN) also has the highest agreement in its clusters. Misleading News

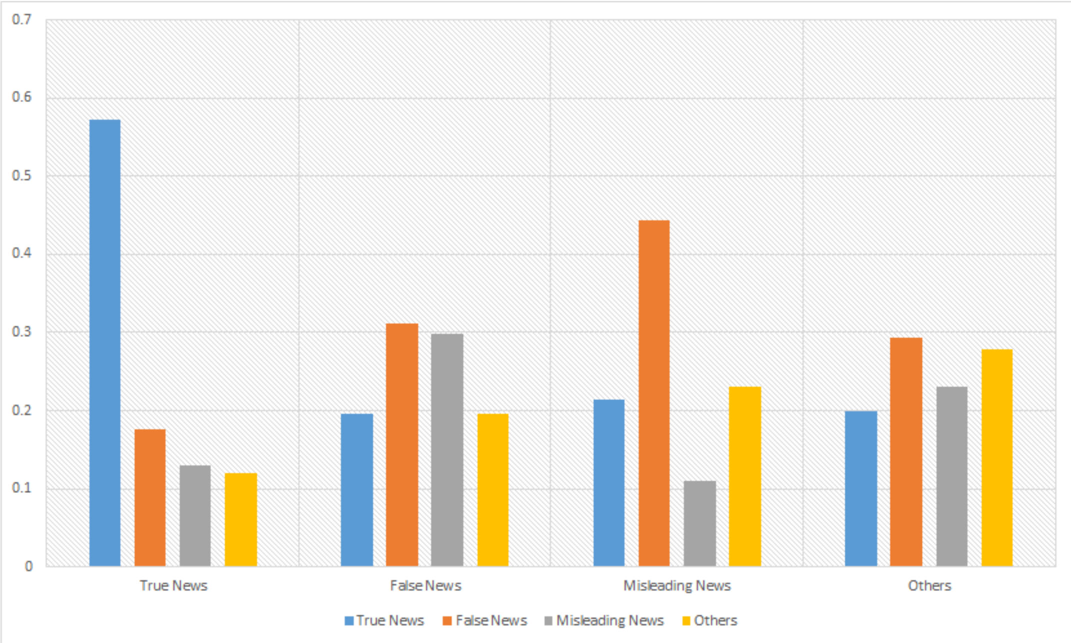

FIGURE 3.1: Distribution of Categories' Agreements Percentage

(MN) ironically has the lowest bar as agreement in its cluster and hence lower than any disagreements. Others' (OT) agreement sits in the second place after disagreement with FN bar.

Both MN and OT categories were difficult to be pinned down. MN is especially hard since it comprises elements of TN and FN. MN might have true statements or information according to supporting URLs but ultimately concludes to wrong statement. From the figure, the ratio of FN and MN disagreement in MN cluster is the highest, even the highest after TN agreement as a whole. This is, in all likelihood, one of the contributing factors why the class is small in size and instead often get confused with FN. In addition, TN comprises the most of data since quite a few of the tweets are clarification or refutation against news which are categorized as false. A number of tweets that contain FN, especially the originators, also have been deleted, removed, or took down from Twitter.

3.4 Hate Speech and Abusive Language Data Set

Hate Speech and Abusive Language data set is a multi-label data set constructed in Indonesian hate speech and abusive language detection research (Ibrohim and Budi, 2019). Apart from defining hate speech and abusive language, they assemble hate speech into two categories based on targets (individual, group), five categories based on characteristics (religion, race, physical, gender, other), and three categories based on levels (weak, moderate, strong). According to this setting, a tweet could have maximum eight labels (hate speech, abusive, a category from targets, four categories from characteristics apart from other, a category from levels).

The data set was built from several previous research (Alfina et al., 2017; Ibrohim and Budi, 2018) added by tweets crawling. The crawling set the date from March 20th, 2018 to September 10th, 2018. For the annotation process, they used crowdsourcing with a paid mechanism in two stages. The first stage was conducted on hate speech and abusive language, while the second stage was directed on hate speech targets, characteristics, and levels. They employed annotation guidelines with gold standard annotation to test whether the annotators fully understand the task before performing the annotation. The annotators were selected from various ages, jobs, ethnicities, and religions. The total number of data set is 13,169 tweets.

This research attempts to observe how similar false news with hate speech or abusive language by text classification, given false news tends to be decorated with hate speech, especially in political settings (Lamb, 2018b). The classes of hate speech targets, characteristics, and levels are excluded, leaving hate speech and abusive language. The detailed numbers of each class are shown in Table 3.3.

Categories	**Numbers**
Hate Speech	5,561
Abusive Language	5,043
Hate Speech + Abusive Language	3,295

TABLE 3.3: Hate Speech and Abusive Language Numbers

3.5 Network Data

There are two kinds of network data created for this research. The first network data is built by utilizing one of Twitter's most signature feature which are adapted by other social media, the mention feature. Mentioning other usernames represent a more real form of communication than creating a network based on following connections (Song, Kim, and Jeong, 2014). Furthermore, a user is able to mention other users without having to follow them in the first place. Buzzers are well-known to disseminating news, facts or information that glorify or slander figures by mention their usernames (Lamb, 2018b). On the other hand, more cautious and fact-checking users also tend to verify or confirm the truth by mentioning the figures whom circulating rumors refer to.

The second network data is built by observing the hashtag co-occurrences in a tweet. In a similar vein with the mention feature, the hashtag feature is another of Twitter's signature features which links every tweet that contains it, effectively useful in cataloguing topics. A hashtag which has been frequently posted will be elected in trending features and thus indicates its popularity. In campaigns for election, there are certain hashtags created to support or oppose certain figures. In the 2019 Indonesian presidential election, two examples of popular, clashing hashtags are *jokowiamin* to support Joko Widodo, the incumbent, along with his vice president Ma'aruf Amin, and *2019gantipresiden* (2019 change the president) to support Prabowo, the challenger.

The mention network and the hashtag co-occurrence network are compared and observed to analyse the visualization of 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. For the mention network, it is a directed network where posting usernames are defined as Source and mentioned usernames as Target. Weight is determined by counting how many times username A mentions username B. For the hashtag co-occurrence network, it is an undirected network, hence Source and Target terms do not matter. Weight is determined by counting how many times hashtag A occurs together with hashtag B. Both of the networks are associated with true, false, or misleading news attribute.

The statistics of both networks are outlined in Table 3.4 and the figures of degree distribution are shown in Figure 3.2, 3.3, and 3.4.

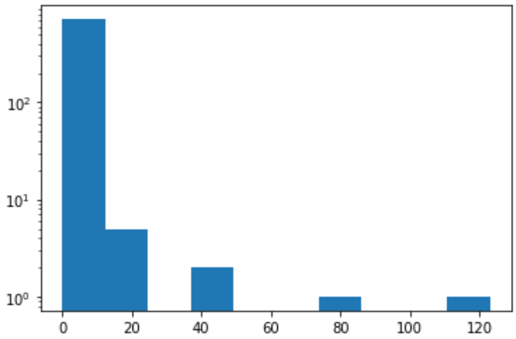

FIGURE 3.2: Logarithmic In-degree Distribution of the Mention Network

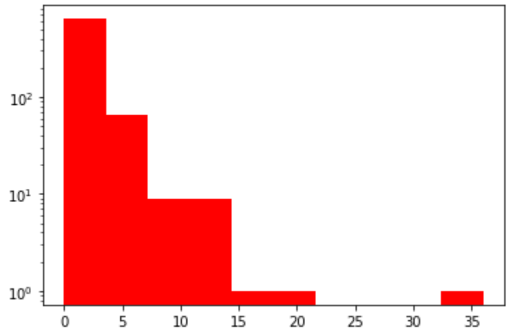

FIGURE 3.3: Logarithmic Out-degree Distribution of the Mention Network

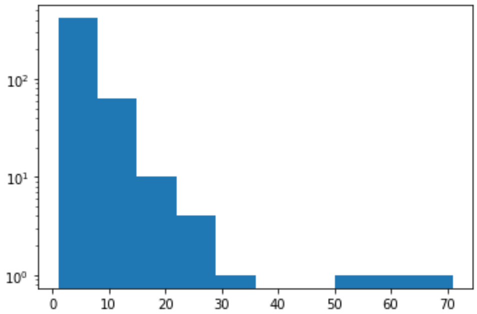

FIGURE 3.4: Logarithmic Degree Distribution of the Hashtag Co-occurrence Network

Statistics	**Mention (Directed)**	**Hashtag Co-occurrence (Undirected)**
# Nodes	724	498
# Edges	992	1,135
Diameter	3	10
Average Degree	1.37	4.56
# Strongly Connected Components	723	-
# Weakly Connected Components	45	75
Average Clustering Coefficient	0.01	0.86

TABLE 3.4: Network Data Properties

Chapter 4

Methodology

4.1 Methodology Overview

The experiments are conducted within two distinctive fields, which are text mining and social network analysis, to detect false news and analyse its dissemination. In the text mining field, this research aims to examine the relevance of text features to distinguish true news, false news, and misleading news. The text features are word n-grams, orthographies, and sentiment lexicons. As a complement, these text features will be assembled by an ensemble method as well. Furthermore, this research also attempts to inspect the relationship between Indonesian false news with Indonesian hateful and abusive language with data set provided by Ibrohim et al. (Ibrohim and Budi, 2019).

To perform text classification, this research utilizes three distinctive classifiers from Python Sklearn module: Multinomial Naive Bayes (MNB), Support Gradient Descent (SGD), and Random Forest (RF). Previous works for Indonesian text classification (Alfina et al., 2017; Ibrohim and Budi, 2019) also utilized these classifiers except SGD, where they utilized Support Vector Machine (SVM). However, instead of reapplying SVM, this research utilizes Support Gradient Descent optimization for SVM to enhance the optimization process (Wijnhoven and de With, 2010).

From the social network analysis perspective, this research aims to provide meaningful visualization to analyse how true news and false news spread between actors in the network. Gephi, which is an open-source software for network analysis, is applied for the task. Few key features of Gephi are navigating and manipulating network data to present interactive and informative visual results (Bastian, Heymann, and Jacomy, 2009). To perceive and analyze how the news disseminate in social media, the community detection and betweenness centrality features are used.

This research aims to amalgamate both fields in distinguishing false news. First, the tweets are analyzed in the textual realm to deduce which text features are significant in discerning false news. Second, the networks of hashtag co-occurrence and mentioned usernames are formed and visualized to present a better understanding on how false news spread. The topic scope of this research is the 2019 Indonesian presidential election where the political nature prompts false news to spread intensely in social media (Azali, 2017).

4.2 Data Preprocessing for Text Classification

The data preprocessing was done differently according to each requirement of the classification experiment. To be discussed in an orderly fashion, the procedure starts with low-dimensional experiments: orthographies and sentiment lexicons. For orthography features, there was no necessity to perform data cleaning since the orthography features (such as exclamation mark or uppercase letters) are simply extracted without modifying the data.

Next, for sentiment lexicon features, stop words removal and text normalization[1] were applied. The stop words dictionary is adopted from (Tala, 2003) and has been further increased in Github stop words project[2]. The text normalization dictionary comprises of 11,034 terms which are mapped to a normalized form. The dictionary is continuous, collective works from researches (Ibrohim and Budi, 2019; Alfina et al., 2017; Salsabila et al., 2018) 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. The comparison of the number of unique terms between original and after preprocessing is shown in Table 4.1.

Word n-grams	**# Unique Terms**	
	Original	After Normalization & Stop Word Removal
Unigram	9,767	8,246
Bigram	35,987	23,869
Trigram	41,909	27,407
Uni-bigram	45,754	32,115
Bi-trigram	77,896	51,276
Uni-bi-trigram	87,663	59,522

TABLE 4.1: The Number of Unique Terms

Finally, there are combinations of word n-grams which are high-dimensional experiments. In this experiment, lower case conversion, link removal and punctuation removal were applied. For mentioned usernames and hashtags, the removal process only targeted at and hash symbol while the usernames and the tags were kept. Keeping mentioned usernames and hashtags are increasingly included as terms given the arguments (Naveed et al., 2011; Ruan et al., 2012) that both are instrumental parts of tweets to be identified and distinguished. Some of usernames and hashtags are also included in the text normalization dictionary and therefore are normalized as well.

4.3 Text Features Extraction

The extracted features are word n-grams, orthographies, and sentiment lexicons. Word n-grams created high-dimensional matrix according to the number of terms in the data set. Meanwhile, both orthographies and sentiment lexicons features created low-dimensional matrices.

4.3.1 Word n-grams

An n-gram is a sequence of N words which form a unit or term (Zhai and Massung, 2016). Hence, a term could consists of one, two, or more words. This research

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

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

establishes and implements six subtypes of word n-grams to create vocabularies: Unigram, bigram, trigram, uni-bigram, bi-trigram, and uni-bi-trigram.

The n-grams example given a document,

"Malam ini, debat pertama akan diselenggarakan antara calon presiden dan wakil presiden."

"Tonight, the first debate will be held between the presidential and vice presidential candidates."

are shown in Table 4.2.

N-Grams	**Terms**
Unigrams	malam:1, debat:1, pertama:1, selenggara:1, calon:1, presiden:2, wakil:1
Bigrams	malam debat:1, debat pertama:1, pertama selenggara:1, selenggara calon:1, calon presiden:1, presiden wakil:1 wakil presiden:1
Trigrams	malam debat pertama:1, debat pertama selenggara:1, pertama selenggara calon:1, selenggara calon presiden:1, calon presiden wakil:1, presiden wakil presiden:1
Uni-bigrams	malam:1, debat:1, pertama:1, selenggara:1, calon:1, presiden:2, wakil:1, malam debat:1, debat pertama:1, pertama selenggara:1, selenggara calon:1, calon presiden:1, presiden wakil:1 wakil presiden:1
Bi-trigrams	malam debat:1, debat pertama:1, pertama selenggara:1, selenggara calon:1, calon presiden:1, presiden wakil:1 wakil presiden:1, malam debat pertama:1, debat pertama selenggara:1, pertama selenggara calon:1, selenggara calon presiden:1, calon presiden wakil:1, presiden wakil presiden:1
Uni-bi-trigrams	malam:1, debat:1, pertama:1, selenggara:1, calon:1, presiden:2, wakil:1, malam debat:1, debat pertama:1, pertama selenggara:1, selenggara calon:1, calon presiden:1, presiden wakil:1 wakil presiden:1, malam debat pertama:1, debat pertama selenggara:1, pertama selenggara calon:1, selenggara calon presiden:1, calon presiden wakil:1, presiden wakil presiden:1

TABLE 4.2: N-Grams Example

First of all, the document is converted into lower case, punctuation and stop words, which are *'ini'* (*this*), *'akan'* (*will*), *'antara'* (*between*), and *'dan'* (*and*) are removed, and normalization from *'diselenggarakan'* (*be held*) as passive form into infinitive form *'selenggara'* (*hold*). The preprocessed document:

"malam debat pertama selenggara calon presiden wakil presiden"

"night first debate hold president vice president candidate"

Furthermore, the table presents the Term Frequency (TF) of each term in the document. On the first glance, term that frequents often naturally has a significance in the document. However, the rarer terms were observed to be as significant as common terms, especially in the longer document or larger collection. Thus, alongside TF, the Inverse Document Frequency (IDF) will be taken into account in calculating the weight of terms (Zhai and Massung, 2016).

The ranking function of TF-IDF weighting scheme is derived as (Zhai and Massung, 2016):

$$f(q,d) = \sum_{q \cup d} c(q,d) \log \frac{M+1}{df(q)} \tag{4.1}$$

where q denotes term, d denotes document, and c denotes count function. M denotes the total number of documents, while $df(q)$ denotes the total number of documents which contain term q.

The ten most frequent terms based on n-grams are outlined in the Table 4.3.

Word n-grams	**Most Frequent Terms: Top 10**
Unigram	prabowo, indonesia, pemilihan *(election)*, jokowi, presiden *(president)*, hoaks *(hoax)*, suara *(vote)*, komisi *(commision)*, calon *(candidate)*, 2019
Bigram	komisi pemilihan, pemilihan presiden, surat suara *(ballot)*, calon presiden, nasional indonesia *(indonesian national)*, tentara nasional *(national army)*, pemungutan suara *(voting)*, republik indonesia *(indonesian republic)*, presiden 2019, komunis indonesia *(indonesian communist)*
Trigram	tentara nasional indonesia, partai komunis indonesia *(indonesia communist party)*, pemilihan presiden 2019, front pembela islam *(islam defender front)*, 2019 ganti presiden *(2019 change president)*, susilo bambang yudhoyono, calon wakil presiden *(vice president candidate)*, kepolisian republik indonesia *(indonesian republic police force)*, badan pengawas pemilihan *(election supervisory institute)*, situng komisi pemilihan *(election commision vote count)*
Uni-bigram	prabowo, indonesia, pemilihan, jokowi, presiden, hoaks, suara, komisi, komisi pemilihan, calon
Bigram	komisi pemilihan, pemilihan presiden, surat suara, calon presiden, nasional indonesia, tentara nasional, tentara nasional indonesia, pemungutan suara, republik indonesia, presiden 2019
Uni-bi-trigram	prabowo, indonesia, pemilihan, jokowi, presiden, hoaks, suara, komisi, komisi pemilihan, calon

TABLE 4.3: N-Grams Top 10 Terms

4.3.2 Orthographies

Orthographies are a set of conventions for writing, such as punctuation or capitalization. Social media such as Twitter is a common example wherein there the conventions of orthographies are sometimes lacking (Gimpel et al., 2011). This is due to people showing their emotions and sentiments more expressively by contorting the formal rules of writing. Therefore, orthographies pattern in social media is increasingly inspected (Saputri, Mahendra, and Adriani, 2018; Ibrohim and Budi, 2019). This research examines the frequency of five orthographies: Exclamation marks, question marks, uppercase letters, lowercase letters, and emojis.

The document example in applying orthographies feature extraction:

Gimana nih, Pak @prabowo ....katanya: Untuk apa jalan tol bagus, kalau mobil dan truk produk asing? Lha...Bapak malah punya Lexus, Alphard, CR-V, Land Cruiser, Land Rover dan Pajero, semuanya produk asing?? Beli mobil Esemka dong yg produk lokal...!☺

Exclamation Mark: Total count: 1.
Question Mark: Total count: 3.
Upper Case Letter: Total count: 17.
Lower Case Letter: Total count: 171.
Emoji: Total count: 1.

4.3.3 Sentiment Lexicons

Sentiment lexicons are vocabulary lists where sentiment scores are allocated to each word. Sentiment analysis or opinion mining has been increasingly conducted to gauge public opinions on certain topics, especially in the social media sphere (Agarwal et al., 2011; Thelwall, Buckley, and Paltoglou, 2012). Particularly for Indonesian language, one of the early sentiment lexicons was created by building seed words from translating English sentiment (Vania, Ibrahim, and Adriani, 2014). This is further refined by adding stemming and synonym set to adjust with social media, where abbreviations and trendy slang words are abundant (Koto and Rahmaningtyas, 2017).

This research exerts Indonesian Sentiment Lexicon (InSet) (Koto and Rahmaningtyas, 2017) which comprises 3,609 positive words and 6,609 negative words[3]. 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. The weighting was done by two annotators by finding the average of both scores and rounding it into the greater nearest integer. Afterwards, the variations of a word were searched in Indonesian stemming library and Indonesian synonym dictionary to enrich the vocabulary. Along with InSet, an Indonesian abusive lexicon (Ibrohim and Budi, 2019) is applied as well which comprises 126 words[4]. These abusive words were built from abusive words that they used as queries. This lexicon will be useful in comparing the 2019 Indonesian Presidential Election News data set with hate speech and abusive language data set experiment.

The positive and negative sentiment scores for each tweet are calculated by sum total. For negative sentiment, every score is converted into positive prior to sum calculation. For abusive sentiment, since it does not contain scaling scores, every occurrence of word equals with 1.

The document example in applying sentiment lexicons feature extraction (after preprocessing):

[3] https://github.com/fajri91/InSet
[4] https://github.com/okkyibrohim/id-multi-label-hate-speech-and-abusive-language-detection

"meriang nih cebonger pidato prabowo pujian tuan erdogan puji erdogan waras dungu kayak cebong"

Positive: *puji (praise)* = 4, *waras (sane)* = 2. Total score: 6.
Negative: *meriang (feverish, agitated)* = 4, *waras (sane)* = 3, *dungu (idiot)* = 5, *kayak (like)* = 3. Total score: 15.
Abusive: *dungu (idiot)* = 1, *cebong*[5] *(tadpole)* = 1. Total score: 2.

4.3.4 Voting Ensembles

Majority voting ensemble or voting ensemble is among the simplest and most effective ensemble learning methods by combining the predictions from multiple other models. The problem type of this research is a classification, therefore the numbers for each label are compared and the most voted label is selected. If each label has equal votes, the class will be determined according to a feature which has the best performance. Voting ensemble has been applied to detect Indonesian hate speech and could reduce the jeopardy of choosing a poor classifier from available options (Fauzi and Yuniarti, 2018). Another recent example is to utilize hard-voting for branches of a convolutional neural network to create a multi-focus image fusion (Amin-Naji, Aghagolzadeh, and Ezoji, 2019).

4.4 Social Network Analysis

Social network visualization will present clear image on how nodes interact and the news dissemination which edges represent. Gephi provides modularity and centrality features, among other things, to present the data accordingly in inspecting how true and false news disseminate. The community detection is capable in partitioning the network into communities to help perceive the potential of filter bubble effects in each community (Grossetti, Du Mouza, and Travers, 2019). Betweenness centrality is a useful centrality metric to identify the most influential nodes which are influencers or even buzzers (Olanrewaju and Rahayu, 2020).

4.4.1 The Community Detection

Community detection is conducted by partitioning a network into communities of densely connected nodes where a community is sparsely connected with others. Precise formulations are incomputable, therefore previous research has been developing heuristic-based methods, such as edge betweenness (Newman and Girvan, 2004) and label propagation (Raghavan, Albert, and Kumara, 2007). To measure and compare the performance of these methods, The modularity of a partition is a standard to calculate the density of edges inside communities as compared to edges between communities (Newman and Girvan, 2004).

For weighted networks, the modularity is defined as (Newman, 2004):

$$Q = \frac{1}{2m}\sum_{i,j}[A_{i,j} - \frac{k_i k_j}{2m}]\delta(c_i, c_j) \tag{4.2}$$

where m represents the sum of the weights of all the edges, $A_{i,j}$ represents the weight of the edge between node i and j, k_i represents the sum of the weights of the edges

[5] A derogatory term intended to Jokowi's supporters

attached to node i, c_i is the community to which node i belongs and the function $\delta(u, v)$ equals 1 if $u = v$ and otherwise 0.

Gephi itself employs fast community enfolding or the Louvain method (Blondel et al., 2008). The main concept of the method is a heuristic approach to greedily optimize modularity by relocating nodes between preliminary communities. In every iteration, the migrating nodes will be placed in the community which maximizes its modularity scores. Afterwards, the discovered communities are aggregated to build a new network of communities. This process repeat itself until a convergence of modularity is attained.

By utilizing Gephi's modularity feature, every community could be identified and partitioned to see the variance of news circulate in each community. Since every community is distinguished according to each density inside and between them, it is a useful feature to observe echo chambers or filter bubble effect (Grossetti, Du Mouza, and Travers, 2019). This is 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 (Lum, 2017).

4.4.2 Betweenness Centrality

Centrality measures are developed to determine the most prominent actor or node in a network according to its position. One of these measures is betweenness centrality, which calculates the number of shortest paths going through a node. Unlike degree centrality which values nodes that have the most direct contacts with other nodes or closeness centrality which values the shortest path to other nodes, betweenness centrality favors nodes that are located on many shortest paths (Brandes, 2001), essentially act as an information bridge from one node to another.

Betweenness centrality is defined as (Brandes, 2001):

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{4.3}$$

where i represents the node which betweenness centrality is measured, σ_{st} represents the total number for the shortest paths in the network, and $\sigma_{st}(i)$ represents all of the shortest paths which traverse the node i.

Due to its nature, betweenness centrality is increasingly applied in determining the influencers of a network (Abbasi and Altmann, 2011; Olanrewaju and Rahayu, 2020).

Chapter 5

Analysis & Results

5.1 Analysis Overview

The analysis is expanded on four experiments. The first experiment is the most meaningful text features, where word n-grams, orthographies, and sentiment lexicons features are compared and analysed. The second experiment is the correspondence with hate speech or abusive language, where the 2019 Indonesian Presidential Election News (2IPEN) data set is compared with the Hate Speech and Abusive Language (HSAL) data set to investigate if there is a correspondence between false news with hate speeches or abusive languages. The third experiment is the social network analysis where the networks are built from the annotated data based on the mention and hashtag co-occurrence to investigate how false news disseminate in the networks by highlighting filter bubble effects and influential actors. Finally, the fourth experiment is the combination of text classification and social network analysis where a text classification model is created to assign the labels into the new 2IPEN data set. This new labelled data set will be incorporated to create new networks and comparisons with the previous networks will be drawn. The fourth experiment will further validate or rebuke the insights derived from the previous networks with new data set processed by the text classification model.

5.2 Most Informative Text Features Experiment

This experiment is conducted to find out which one is the most informative text features for distinguishing true news, false news, and misleading news. In order to enrich the perspective, the 2019 Indonesian Presidential Election News (2IPEN) data set is generated into two different 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 five classes, with additional classes are Other 1 and Other 2. Other 1 represents the data where the two annotators judge them as not news or having out-of-the-topic tweets. Other 2 represents the data where three annotators differ in annotating them. Hence, the second setting emulates real-world case of text classification where there are some tweets that are not news, do not correlate, or highly ambiguous.

The text features consist of word n-grams, orthographies, and sentiment lexicons. Word n-grams are the most prevalent features to be applied in text classification, where the bag of words created by terms with $n-$length. This bag of words is transformed by TF-IDF weighting to regard the significance of a term based on frequent and rare occurrences. Orthography is considered since a tweet's contexts are not only gauged from the words but from other constituents such as punctuation or capitalization as well. Since Twitter is a social media where people communicate themselves more expressively, they tend to utilize these constituents to fully convey what they really mean. Lastly, sentiment lexicons is a set of vocabularies linked to scores which defined to a certain sentiment. Alike with orthographies, sentiment lexicons are considered since expressive or heartfelt tweets are inclined to use certain words that have strong sentiment. The final setting is a classification voting ensemble, where the prediction will be determined by the most voted predicted class from the three text features. If there is no majority class, the prediction from the best-performed text feature will be selected.

5.2.1 Word N-Grams

Word n-grams consist of unigram, bigram, trigram, the combination of unigram and bigram (uni-bigram), the combination of bigram and trigram (bi-trigram), and the combination of unigram, bigram, and trigram (uni-bi-trigram). The scores of precision, recall, and f1 score of word n-grams feature are outlined in Table 5.1 and 5.2.

Word N-Grams	True News			False News			Misleading News		
Features	P	R	F1	P	R	F1	P	R	F1
Unigram	0.725	0.884	0.797	0.777	0.656	0.718	**0.788**	**0.280**	**0.414**
Bigram	0.693	0.886	0.778	0.805	0.613	0.696	0.726	0.263	0.386
Trigram	*0.645*	*0.918*	*0.757*	*0.821*	*0.485*	*0.609*	*0.731*	*0.210*	*0.327*
Uni-bigram	**0.730**	**0.903**	**0.807**	**0.811**	**0.692**	**0.747**	0.830	0.246	0.380
Bi-trigram	0.673	0.879	0.762	0.779	0.592	0.673	0.770	0.210	0.330
Uni-bi-trigram	0.712	0.888	0.790	0.801	0.667	0.728	0.781	0.246	0.374

TABLE 5.1: Metrics Score of Three Classes 2IPEN Data Set - Word N-Grams

Word N-Grams	True News			False News			Misleading News		
Features	P	R	F1	P	R	F1	P	R	F1
Unigram	**0.601**	**0.816**	**0.692**	0.627	0.669	0.647	0.645	0.202	0.307
Bigram	0.562	0.790	0.657	**0.707**	**0.621**	**0.661**	0.683	0.263	0.380
Trigram	*0.532*	*0.739*	*0.619*	*0.758*	*0.495*	*0.599*	*0.705*	*0.172*	*0.277*
Uni-bigram	0.580	0.825	0.681	0.646	0.659	0.652	0.757	0.228	0.350
Bi-trigram	0.557	0.799	0.656	0.718	0.608	0.658	**0.745**	**0.289**	**0.417**
Uni-bi-trigram	0.563	0.816	0.666	0.650	0.651	0.651	0.731	0.211	0.327

TABLE 5.2: Metrics Score of Five Classes 2IPEN Data Set - Word N-grams

The scores of precision and recall are the average scores of the three classifiers (Multinomial Naive Bayes, Support Gradient Descent, and Random Forest) generated for each word n-gram feature. F1 scores are calculated according to average precision and recall scores. In each class, the bold-styled numbers defined the best feature based on the F1 score, while the italic-styled numbers defined the worst. This is the re-occurring format of the following text features experiments.

As shown in the tables, True News data tend to have a better recall while False News and Misleading News tend to have more precise results. While the variance between precision and recall scores of True News and False News is slight, Misleading News has significant low recall scores. Misleading News has the smallest size in number compared to the other two classes (21% of True News size and 29% of False News size) and this striking imbalance favored precision over recall according to the experiment. Meanwhile, True News and False News showed more favorable performance according to their F1 scores. As shown in Figure 5.1, the reduction of F1 scores between three classes and five classes settings is also small, with the largest gap is 0.138 in trigram for True News. In one case, bi-trigram for Misleading News performed better in the five classes setting.

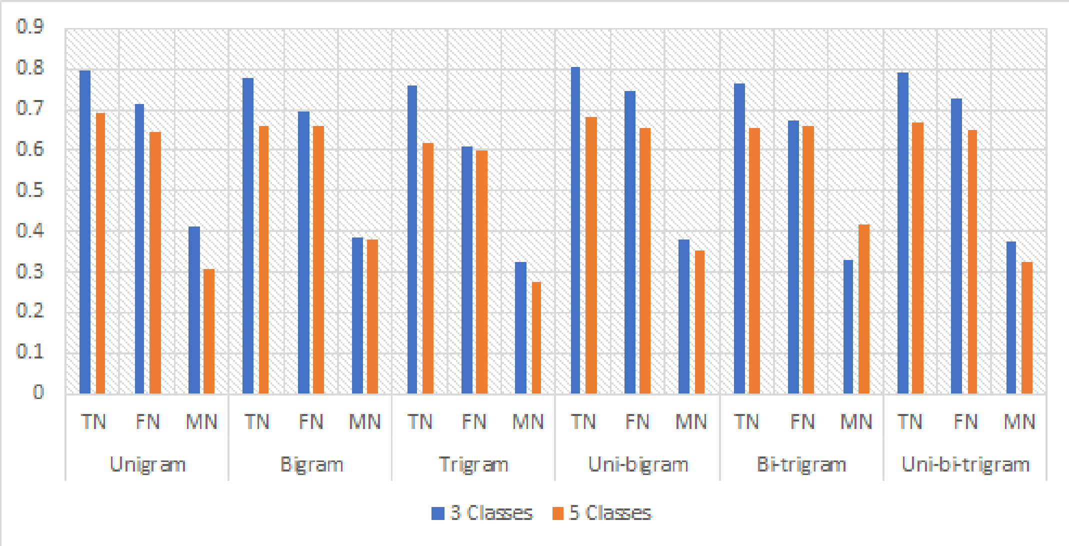

FIGURE 5.1: F1 Score Comparison of Two 2IPEN Data Sets - Word N-Grams

In the three classes setting, any gram composition which includes unigram majorly performed better rather than the compositions which do not. Interestingly, there are some intriguing occurrences when adding Other 1 and Other 2 classes. For the True News class, unigram composition still performed superior as the three class setting suggests. However, False News and Misleading News favored bigram and bi-trigram.This suggests that false news and misleading news tend to associate with two-word phrases, or abbreviated words and slangs which if normalized translated into terms consist of two words or three words, more than true news. Finally, trigram composition performed the worst in all cases, indicating that bigram is the best maximum composition of n-grams without adding other grams for the Indonesian language.

5.2.2 Orthographies

Orthographies consist of the combination of exclamation marks (E), question marks (Q), uppercase letters (U), lowercase letters (L), and emojis (M): EQULM, EQUL, EQUM, EQLM, EULM, and QULM. The scores of precision, recall, and f1 score of orthographies feature are outlined in Table 5.3 and 5.4.

Orthographies Features	True News P	R	F1	False News P	R	F1	Misleading News P	R	F1
EQULM	0.553	0.780	0.647	0.635	0.287	0.396	0.056	0.035	0.043
EQUL	*0.367*	*0.434*	*0.398*	**0.402**	**0.595**	**0.480**	0.056	0.035	0.043
EQUM	**0.533**	**0.844**	**0.653**	*0.473*	*0.200*	*0.281*	*0.039*	*0.018*	*0.024*
EQLM	0.374	0.512	0.432	0.437	0.523	0.476	**0.133**	**0.079**	**0.099**
EULM	0.534	0.765	0.629	0.460	0.267	0.338	0.157	0.061	0.088
QULM	0.359	0.452	0.400	0.407	0.567	0.474	0.048	0.026	0.034

TABLE 5.3: Metrics Score of Three Classes 2IPEN Data Set - Orthographies

Orthographies Features	True News P	R	F1	False News P	R	F1	Misleading News P	R	F1
EQULM	0.305	0.322	0.313	*0.340*	*0.285*	*0.310*	**0.079**	**0.044**	**0.057**
EQUL	0.327	0.280	0.302	0.294	0.526	0.377	*0.020*	*0.018*	*0.019*
EQUM	**0.449**	**0.507**	**0.476**	0.298	0.436	0.354	0.037	0.018	0.024
EQLM	0.274	0.445	0.339	0.300	0.503	0.375	0.039	0.035	0.037
EULM	0.354	0.285	0.316	0.308	0.528	0.389	0.051	0.035	0.042
QULM	*0.291*	*0.294*	*0.292*	**0.311**	**0.564**	**0.401**	0.057	0.053	0.055

TABLE 5.4: Metrics Score of Five Classes 2IPEN Data Set - Orthographies

In the three classes setting, True News data tend to have a better recall, Misleading News tend to have more precise results, and False News results are varied. In the five classes setting, True News results turn to be varied as well. The performances of orthographies are considerably lower than n-grams, especially for Misleading News where all F1 scores are below 0.1. Unlike word n-grams, Misleading News' precision and recall scores are roughly equal. As shown in Figure 5.2, True News for the combination of EQULM, EQUM, and EULM have the largest reduction of F1 scores between three classes and five classes settings, where the F1 scores for three classes setting are fairly high. Few results where the F1 scores of five classes setting are higher belong to False News and Misleading News classes.

From an orthographic perspective, removing lowercase letters gives the best performance for True News, yet frequently tends to give the inferior or worst performance for False News and Misleading News. On the other hand, in the three classes setting, removing emojis boosts the performance of False News and Misleading News yet undermines True News'. In the five classes setting, removing exclamation marks gives a similar outcome with removing emojis in three classes setting. All of this indicates that the orthographic combinations between True News with False and Misleading News tend to contrast to each other. Therefore, the complete combination tends to produce average results over the settings.

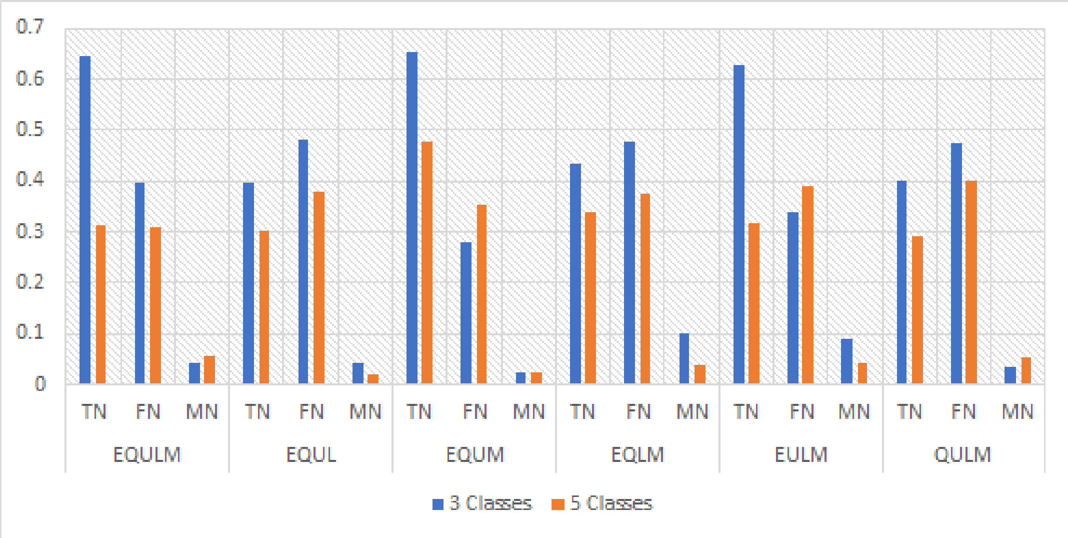

FIGURE 5.2: F1 Score Comparison of Two 2IPEN Data Sets - Orthographies

5.2.3 Sentiment Lexicons

Sentiment lexicons consist of the combination of positive sentiments (P), negative sentiments (N), and abusive sentiments (A): PNA, PN, PA, and NA. The scores of precision, recall, and f1-score of sentiment lexicons feature are outlined in Table 5.5 and 5.6.

Sentiment Lexicons	True News			False News			Misleading News		
Features	P	R	F1	P	R	F1	P	R	F1
PNA	0.554	0.829	0.664	0.525	0.251	0.340	0.063	0.035	0.045
PN	**0.552**	**0.836**	**0.665**	0.299	0.221	0.254	0.064	0.044	0.052
PA	0.564	0.726	0.635	**0.423**	**0.351**	**0.384**	*0.000*	*0.000*	*0.000*
NA	*0.531*	*0.622*	*0.573*	*0.309*	*0.136*	*0.189*	**0.074**	**0.307**	**0.120**

TABLE 5.5: Metrics Score of Three Classes 2IPEN Data Set - Sentiment Lexicons

As shown in the tables, True News data tend to have a better recall while False News and Misleading News tend to have more precise results. However, Misleading News has roughly equal precision and recall scores, although its scores are significantly low compared to the other two classes, even has few zero scores. The performances of sentiment lexicons are also considerably lower than n-grams and approximately alike with orthographies. As shown in Figure 5.3, PA for False News is the single occurrence that has a significant reduction from three classes to five classes, and PN for Misleading News has better performance in five classes setting.

Sentiment Lexicons	True News			False News			Misleading News		
Features	P	R	F1	P	R	F1	P	R	F1
PNA	0.405	0.790	0.535	**0.396**	**0.192**	**0.259**	0.044	0.178	0.025
PN	0.414	0.786	0.542	0.455	0.179	0.257	**0.077**	**0.070**	**0.074**
PA	*0.429*	*0.674*	*0.524*	*0.231*	*0.108*	*0.147*	*0.000*	*0.000*	*0.000*
NA	**0.394**	**0.908**	**0.549**	0.229	0.126	0.162	*0.000*	*0.000*	*0.000*

TABLE 5.6: Metrics Score of Five Classes 2IPEN Data Set - Sentiment Lexicons

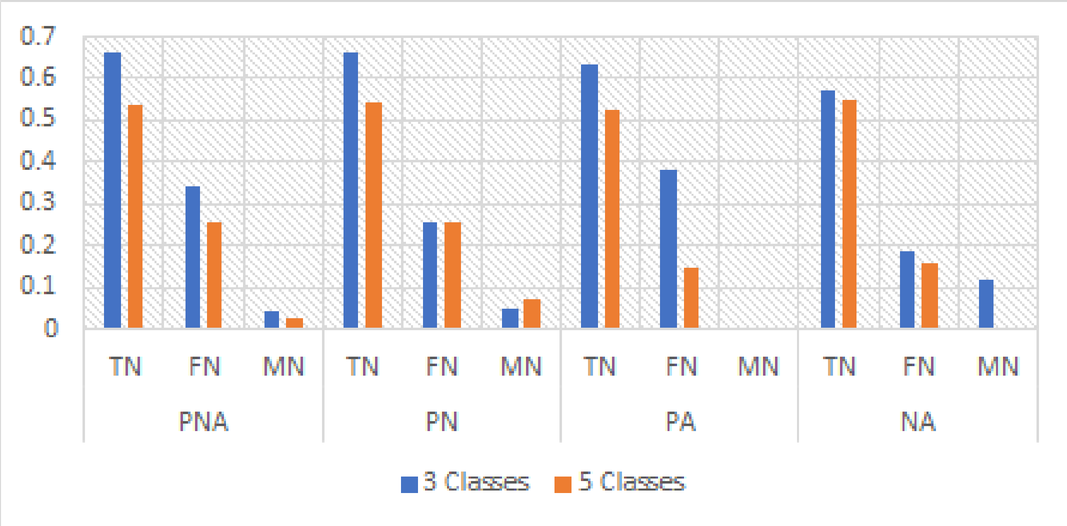

FIGURE 5.3: F1 Score Comparison of Two 2IPEN Data Sets - Sentiment Lexicons

False News in three classes setting shows to be an anomaly of the emerged pattern. In that setting, removing abusive lexicons performs the worst where in other settings performs very well or the best, and removing negative lexicons performs the best, but becomes the worst in almost all of the remaining settings. Removing either positive or negative lexicons gives zero scores for Misleading News. The complete combination tends to produce average results over the settings. Overall, the mixed result make the insights based on sentiment lexicon are very difficult to be pinned down to conclude them in a general conclusion.

5.2.4 Voting Ensembles

Voting ensembles consists of three parts, assembled from three models from each text feature. There are three voting ensembles configured for each data set. The first ensemble is arranged from all combinations of each feature, the second ensemble is arranged from the best combination of each feature and the third ensemble is arranged from the worst combination. The best and worst combinations adhere to each data set setting and are determined by the sum of their F1 scores in each class, where the minimum is the best.

The combinations of the ensembles are outlined in Table 5.7. The scores of precision, recall, and f1-score of sentiment lexicons feature are outlined in Table 5.8 and 5.9.

Data Set	**Ensembles**	**Combinations**
Three-classes	Ensemble I	Uni-bi-trigram, EQULM, PNA
	Ensemble II	Uni-bigram, EQLM, PN
	Ensemble III	Trigram, EQUM, NA
Five-classes	Ensemble I	Uni-bi-trigram, EQULM, PNA
	Ensemble II	Bigram, EULM, PN
	Ensemble III	Trigram, EQUL, PA

TABLE 5.7: The Ensembles and Features Combinations

While outperformed orthographies and sentiment lexicons, the voting ensembles performed still slightly below the text features. In Figure 5.4, all performances are reduced from three-classes to five-classes except False News and Misleading News in Ensemble III.

Voting Ensembles	True News P	R	F1	False News P	R	F1	Misleading News P	R	F1
Ensemble I	0.685	0.888	0.774	0.728	0.577	0.644	**0.833**	**0.263**	**0.400**
Ensemble II	**0.671**	**0.899**	**0.768**	**0.796**	**0.569**	**0.664**	0.643	0.237	0.346
Ensemble III	*0.576*	*0.927*	*0.711*	*0.709*	*0.300*	*0.422*	*0.500*	*0.053*	*0.095*

TABLE 5.8: Metrics Score of Three Classes 2IPEN Data Set - Cross-features Ensembles

Cross-features Ensembles	True News P	R	F1	False News P	R	F1	Misleading News P	R	F1
Ensemble I	**0.551**	**0.849**	**0.668**	**0.638**	**0.569**	**0.602**	0.471	0.211	0.291
Ensemble II	0.505	0.872	0.639	0.670	0.500	0.573	**0.476**	**0.263**	**0.339**
Ensemble III	*0.462*	*0.872*	*0.603*	*0.671*	*0.423*	*0.519*	*0.429*	*0.158*	*0.231*

TABLE 5.9: Metrics Score of Five Classes 2IPEN Data Set - Cross-features Ensembles

Ensemble II (the text features ensemble from the best compositions) mostly performed the best in the three-classes setting while Ensemble I (the text features ensemble from all compositions) in the five-classes setting. Ensemble III which is assembled from the worst combination of each feature performed the worst in all occasions. From this experiment, voting ensemble method paves possibilities to include under-performing text features such as orthographies and sentiment lexicons to give more desirable results. The best and worst compositions from the previous experiments also remain valid, while encompassing all combinations gives better performance to the larger data set or the data set with more classes.

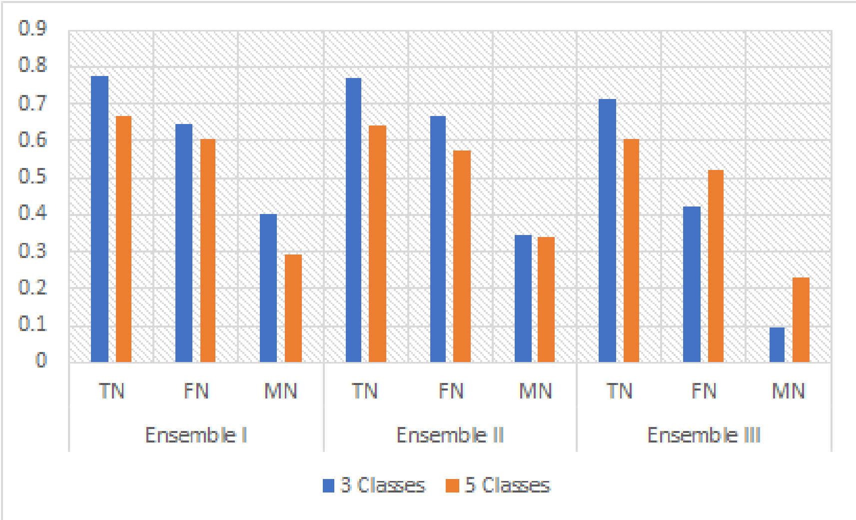

FIGURE 5.4: F1 Score Comparison of Two 2IPEN Data Sets - Cross-features Ensembles

5.2.5 Final Comparison of Text Features Experiment

The final comparison table of the best text features from previous experiments are outlined in Table 5.10.

Data Set	Features	F1 True News	F1 False News	F1 Misleading News
Three-classes	**Uni-bigram**	**0.807**	**0.747**	**0.380**
	EQLM	0.432	0.476	0.099
	PN	*0.665*	*0.254*	*0.052*
	Ensemble II	0.768	0.664	0.346
Five-classes	**Bigram**	**0.657**	**0.661**	**0.380**
	EULM	*0.316*	*0.389*	*0.042*
	PN	0.542	0.257	0.074
	Ensemble I	0.668	0.602	0.291

TABLE 5.10: Metrics Score of The Final Comparison of All Text Features plus Ensemble

The best text feature for detecting false news and misleading news is word n-grams, specifically uni-bigram for three-classes and bigram for five-classes data set. Orthographies and sentiment lexicons performed fairly undesirable, but these features are possible to be applied together by a voting ensemble with word n-grams to leverage the performance.

5.3 Correspondence with Hate Speech or Abusive Language Experiment

This experiment is conducted to investigate how similar false news tweets with hate speech and abusive language induced tweets. The general idea is to create two training models from 2IPEN data set and hate speech and abusive language (HSAL) data set. These models will be processed by the same feature processing and will evaluate the same testing data set by the same classification method. The testing data set is derived and concatenated from the aforementioned two data sets. Afterwards, the resulting prediction models will be arranged through a co-occurrence matrix where its p-value is measured by the chi-squared test to investigate whether the aforementioned data sets are not correlated or correlated to each other.

Regarding the composition of HSAL data set, there are three divisions of data set defined to be processed along with 2IPEN data set, which are hate speech composition, abusive language composition, and hate speech plus abusive language composition. In this experiment, Others 1 and Others 2 class will be disregarded. False News and Misleading News data in the 2IPEN data set will be merged together and adapted into the same category according to the composition. For example, HSAL data set with hate speech composition has every tweet either classified as 1 (a hate speech) or 0 (not a hate speech), and 2IPEN data set will convert every tweet that classified as False News and Misleading News as 1, while True News as 0.

For the text features, uni-bigram is selected as the best text features of the previous experiment for three classes data set (disregarding Others classes). The detail size of HSAL and FN data sets can be seen in Table 3.3 and 3.2. The scores of co-occurrences, chi-square, and p-values between HSAL and 2IPEN data sets in every compositon are outlined in Table 5.11, 5.12, and 5.13.

Hate Speech Composition		**2IPEN Data Set Yes**	**2IPEN Data Set No**
HSAL Data Set	**Yes**	930	262
	No	1,325	453
Chi-square			4.778
p-value			0.029

TABLE 5.11: Co-occurrences Scores, Chi-squares, and P-values - Hate Speech Composition

Abusive Language Composition		**2IPEN Data Set Yes**	**2IPEN Data Set No**
HSAL Data Set	**Yes**	746	132
	No	1,506	586
Chi-square			56.818
p-value			0.478×10^{-13}

TABLE 5.12: Co-occurrences Scores, Chi-squares, and P-values - Abusive Language Composition

Hate Speech + Abusive Language Composition		2IPEN Data Set Yes	2IPEN Data Set No
HSAL Data Set	**Yes**	337	50
	No	1,917	666
Chi-square			30.443
p-value			0.344×10^{-7}

TABLE 5.13: Co-occurrences Scores, Chi-squares, and P-values - Hate Speech + Abusive Language Composition

The co-occurrence scores are gained from the average scores of the three classifiers generated for each composition. The total amount of co-occurrences is 2,970. The dimension of matrix is 2×2 and therefore the degrees of freedom is 1. The null hypothesis is that both of these data sets are not correlated.

First of all, let us define 'Hate Speech', 'Abusive Language', and 'Hate Speech + Abusive Language' as the topic. All of the composition shares a similar outcome where HSAL predicts as not the topic while 2IPEN predicts as the topic co-occur the most. Conversely with a same pattern, all outcomes where HSAL predicts as the topic while 2IPEN predicts as not the topic co-occur the least. Regarding the correlation, all compositions have p-values under 0.05, thus the null hypothesis is rejected and 2IPEN data set correlates with every composition. Particularly, Abusive Language and Hate Speech + Abusive Language compositions produce very small p-values. Meanwhile, although still correlate, the p-value of Hate Speech composition is nearly the critical value, suggesting the correlation is fairly modest. These findings indicate that abusive languages are heavily correlated with false news and hate speeches are moderately correlated. By compounding hate speeches and abusive languages, it is still strongly correlated with false news.

5.4 Social Network Analysis Experiment

This experiment is conducted to illustrate how true news and false news disseminate through a network of Twitter data with the 2019 Indonesian Presidential Election as the main topic. The two networks are the mention usernames and the hashtag co-occurrence networks. The mention network is a directed network where the nodes are Twitter usernames. The hashtag co-occurrence network is an undirected network where the nodes are hashtags.

Other 1 and Other 2 classes will be included in the network formation but sidelined in the analysis. In the visualization, communities are represented by colours and the betweenness centralities are defined by nodes' size. The detailed statistics of both networks are outlined in Table 5.14.

Both networks have a quite balanced distribution of true news and false news, especially the hashtag co-occurrence network. Both networks also have fairly high modularity scores. This indicates that the networks are dense and closely connected, enabling a faster rate of information dissemination through the networks.

In the main report, only selected few of the network visualizations are included in order to adhere to the analysis narration. The complete visualization of each community is placed in Appendix B.

Statistics	Mention (Directed)	Hashtag Co-occurrence (Undirected)
# True News Nodes	268	220
# False News Nodes	344	216
# True News Edges	292	401
# False News Edges	384	424
Modularity	0.714	0.856
# Communities	62	81

TABLE 5.14: Detailed Network Data Properties

5.4.1 Mention Network

The mention usernames network is illustrated in Figure 5.5. This is a directed network and the colors of edges are defined according to the colors of source nodes. From the network, some influential usernames of the 2019 Indonesian Presidential Election could be identified such as *caknur14* (a personal account), *divhumas_polri* (an official account of *Kepolisian Republik Indonesia* or shortened *Polri*, Indonesian National Police), and *bawaslu_ri* (an official account of *Badan Pengawas Pemilihan Umum Republik Indonesia* or shortened *Bawaslu RI*, Indonesian General Election Supervisory Agency).

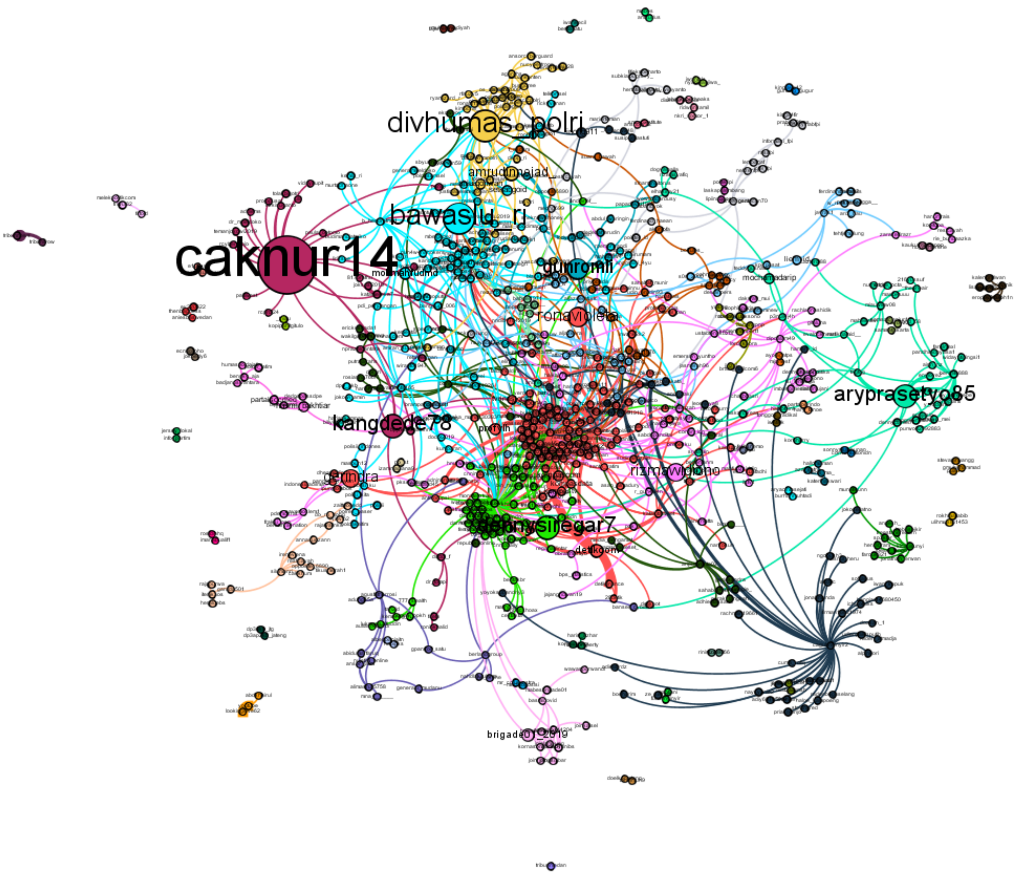

FIGURE 5.5: Mention Network

The first analysis is community-based to investigate the presence of a filter bubble effect in each community. Table 5.15 outlines the top ten largest communities according to the number of the nodes and the influential usernames that each community has. The selected usernames are only for those that have a betweenness centrality score larger than zero and limited to the top five usernames. Bold-styled usernames indicate that they belong to the top ten influential usernames according to the betweenness centrality score. This table format will re-occur in the hashtag co-occurrence network. In addition, there are six verified accounts: two online news sources (*detikcom* and *kompascom*), two government institutions (*bawaslu_ri* and *divhumas_polri*), one political party (*gerindra*), and one politician (*mohmahfudmd*).

For the centrality analysis, figure 5.6 outlines the number comparison between True News and False News of nodes and edges in each community for the mention network. This figure format will re-occur in the following comparisons between True News and False News. From the top ten communities of the mention network, there are five communities that have more false news circulating rather than true news, three communities with more true news, and two communities which have fairly balanced proportion of both news.

Communities	**Influential Usernames**
1st Community	**ronavioleta, gerindra (Ver.),** detikcom (Ver.), profylh, kompas-com (Ver.)
2nd Community	**bawaslu_ri (Ver.),** mohmahfudmd (Ver.)
3rd Community	**dennysiregar7**
4th Community	**rizmawidiono**
5th Community	inong911
6th Community	**aryprasetyo85,** mochamadarip
7th Community	**divhumas_polri (Ver.),** amrudinnejad_, ahlulqohwah
8th Community	liem_id
9th Community	**caknur14, kangdede78**
10th Community	-

TABLE 5.15: Top 10 Communities with Influential Usernames - Mention Network

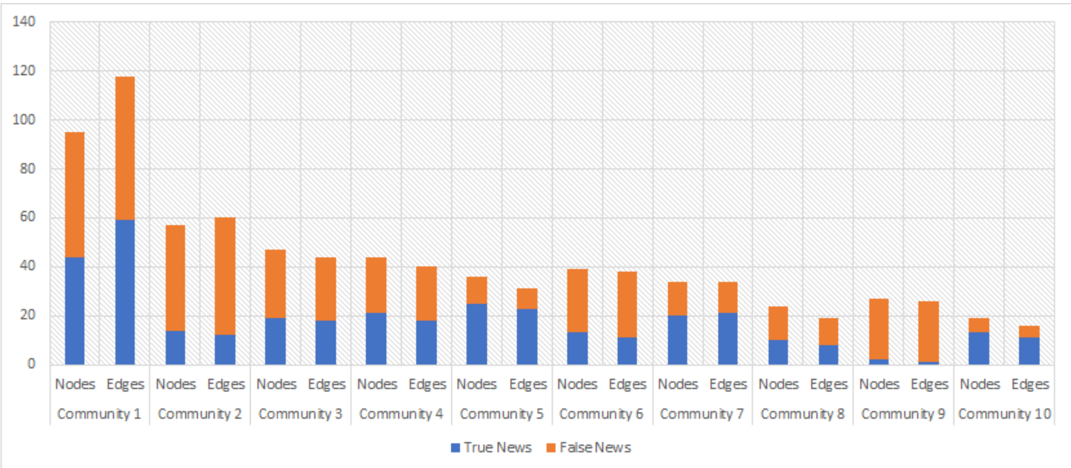

FIGURE 5.6: Distribution of True News and False News - Top 10 Communities of Mention Network

One of the communities which is strongly vulnerable to filter bubble of false news is the 9th Community as shown in Figure 5.7 and 5.8. This community has two influential users, *caknur14* and *kangdede78*, that are unverified accounts, and additionally only contain one true news tweet and the remaining ones are entirely false news. Another community with a similar disposition of false news circulation is the 2nd Community, shown in Figure B.3 and B.4. Different from the 9th Community, this community mostly comprises of verified, governmental figures and institutions. The influential usernames are a government institution *Bawaslu* official account, and a politician Moh. Mahfud MD. The accounts of *Bawaslu* and Mahmud MD are also the influential users of this community.

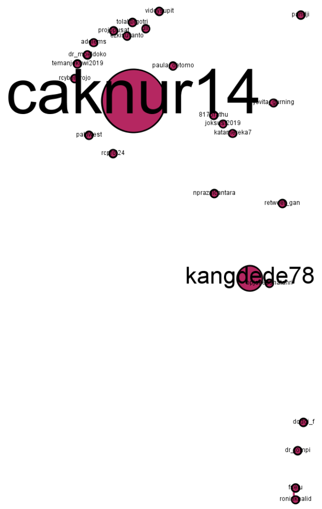

FIGURE 5.7: 9th Community of Mention Usernames Network - True News Dissemination

FIGURE 5.8: 9th Community of Mention Usernames Network - False News Dissemination

On the other hand, the 7th Community is one of top communities which has more true news circulating than false news, as shown in Figure 5.9 and 5.10. This community mostly comprises of police and military accounts. *Polri*'s official account is the most influential actor of this community, along with few unverified, personal accounts *amrudinnejad_* and *ahlulqohwah*. It would be reasonable for this community to have true news filter bubble, since the police force has a cyber-based department to fight hoaxes. Another community with a similar disposition of true news circulation is the 5th Community, as shown in Figure B.9 and B.10. However, this community only has one influential unverified account.

It is also worth mentioning that Joko Widodo, one of the president candidate, belongs to the 3rd Community that is vulnerable to false news filter bubble. This is different with Prabowo Subianto, the other candidate, who belongs to the 1st Community, the most populated community with balanced number of true and false news. By the raw numbers, however, Prabowo is attacked more by False News than Jokowi, yet is also supported more by True News to create a fairly balanced community.

FIGURE 5.9: 7^{th} Community of Mention Usernames Network - True News Dissemination

FIGURE 5.10: 7^{th} Community of Mention Usernames Network - False News Dissemination

The second analysis focuses on identifying the influential actors, which in this case are the users, by observing centrality-based visualization. Slightly different from community-based visualization, the centrality-based visualization is able to disregard the user nodes which do not have edges except the influential user itself. This is because Gephi features for topology only enable filtering the nodes out by nodes, not communities.

Table 5.15 outlines the top ten influential usernames according to the betweenness centrality score and other influential usernames each reaches, along with the classes. The selected reached usernames are only for those that have a betweenness centrality score larger than zero and limited to the top five usernames. This table format will re-occur in the hashtag co-occurrence network. A new seen influential username is *gunromli* which belongs to the 13^{th} Community yet ranks as the 7^{th} influential username. In addition, *gunromli* is a politician and his account is verified as well.

Usernames	Community	Reached Influential Usernames
caknur14	9^{th}	kangdede78 (FN)
divhumas_polri (Ver.)	7^{th}	-
bawaslu_ri (Ver.)	2^{nd}	-
dennysiregar7	3^{rd}	-
kangdede78	9^{th}	caknur14 (FN)
aryprasetyo85	6^{th}	rizmawidiono (TN), mochamadarip (TN)
gunromli (Ver.)	13^{th}	-
ronavioleta	1^{st}	-
gerindra (Ver.)	1^{st}	-
rizmawidiono	4^{th}	aryprasetyo85 (TN)

TABLE 5.16: Top 10 Influential Usernames with Other Reached Influential Usernames - Mention Network

Figure 5.11 outlines the number comparison between True News and False News of nodes and edges in each username for mention network. From the top ten influential users of the mention network, there are five users that disseminate or receive more false news rather than true news, two users with more true news, and three users which have fairly balanced proportion of both news.

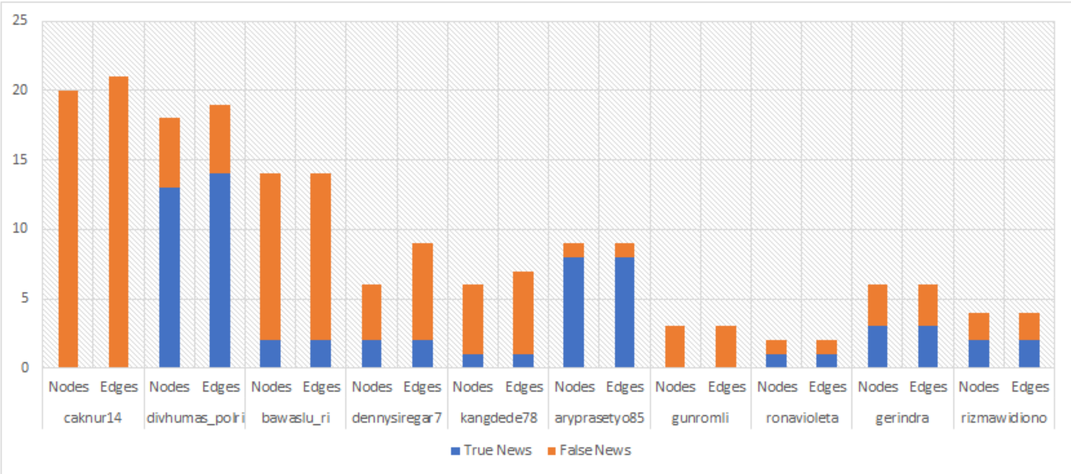

FIGURE 5.11: Distribution of True News and False News - Top 10 Influential Usernames of Mention Network

The top influential actor, *caknur14*, only reverberates and receives false News as shown in Figure 5.12 and 5.13. This user also reaches users in other communities and another influential user in the same community, *kangdede78*. Both of these users are personal, unverified accounts which are prominent information receivers and disseminators, particularly false news information. Due to their nature and connection, there is a high possibility that these users are buzzers that intentionally spread false news. Another influential user of false news is *bawaslu_ri*, which happens to be an official account of a government institution and its visualization is shown in Figure B.17 and B.18. While this user reaches users in other communities as well, it does not connect with more influential users. The lack of connection with other influential users suggests that the user spread the false news unknowingly or believing it is true before proven to be false. This is in accordance with their status as a government institution.

FIGURE 5.12: *caknur14*'s Network - True News Dissemination

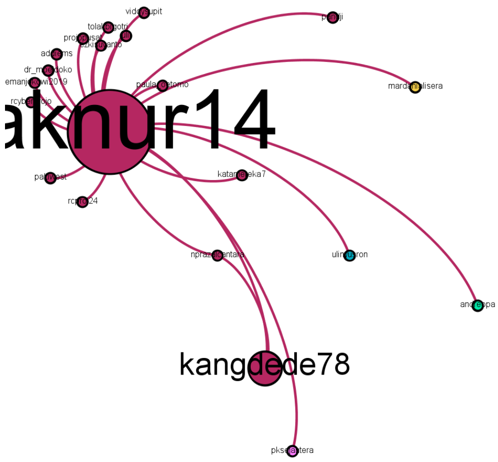

FIGURE 5.13: *caknur14*'s Network - False News Dissemination

The top influential actor who circulates more true news is *divhumas_polri*, a verified account of Indonesian national police, as shown in Figure 5.14 and 5.15. This user reaches users in other communities but not other influential users. Meanwhile, user *aryprasetyo85* is an opposite, as shown in Figure B.23 and B.24. It is a personal, unverified account, and belongs to a community which has more false news circulating. However, it is a prominent true news receiver and disseminator in its community, and reaches an other influential actor of a different community, *rizmawidiono*. However, it needs to be mentioned that *rizmawidiono* user does not side to either news. According to these findings, influential users which are personal and unverified tend to reach other influential users than verified, institutional users.

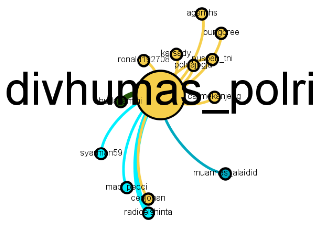

FIGURE 5.14: *divhumas_polri*'s Network - True News Dissemination

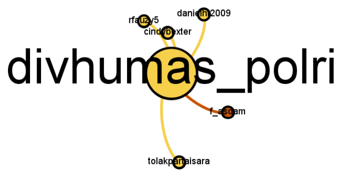

FIGURE 5.15: *divhumas_polri*'s Network - False News Dissemination

Based on the social network analysis experiment of the mention network with the 2019 Indonesian presidential election setting, the interesting findings are:

- False news spread more in the top communities of usernames than true news. False news are also more likely being disseminated or received by most of the top influential usernames, especially verified accounts.
- The communities that are infested with more false news tend to have more influential usernames than the inclined "true news" communities.
- The total number of false news in an inclined "false news" community or from an inclined "false news" username tends to have a large margin from true news, while the opposite case only has a moderate margin.
- Regardless of the news' label, unverified influential usernames tend to be more connected with other influential usernames than the verified usernames.

5.4.2 Hashtag Co-occurrence Network

The hashtag co-occurrence network is illustrated in Figure 5.16. Different from mention network, this is an undirected network, therefore the colors of edges do not hold a important information and are randomly mixed by Gephi. From the network, some influential hashtags of the 2019 Indonesian Presidential Election could be identified such as *debatpilpres2019* (2019 presidential election debate), *pemilu2019* (2019 general election), and *2019gantipresiden* (2019 change the president).

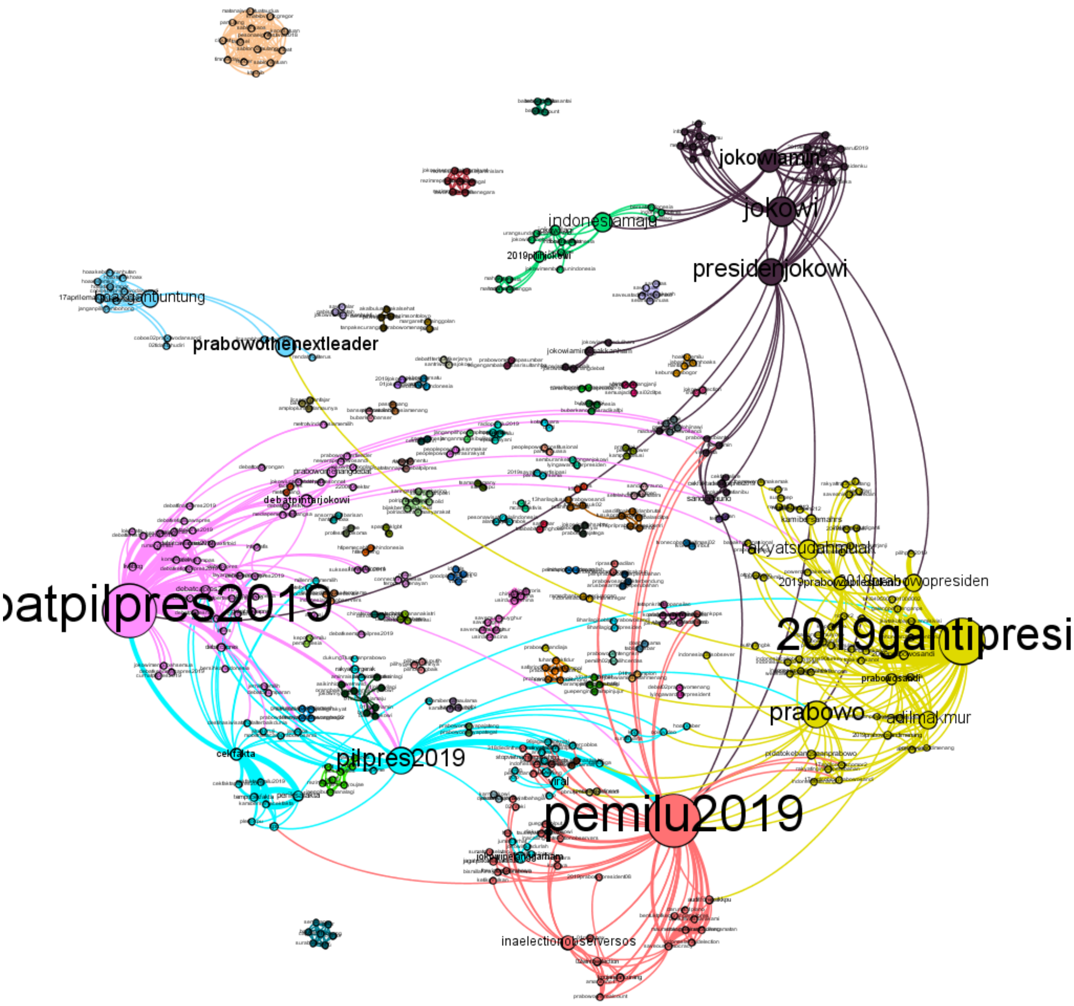

FIGURE 5.16: Hashtag Co-occurrence Network

Table 5.17 outlines the top ten largest communities according to the number of the nodes and the influential hashtags that each community has.

Communities	Influential Hashtags
1st Community	**2019gantipresiden, prabowo, rakyatsudahmuak,** adilmakmur, 2019prabowopresiden
2nd Community	**debatpilpres2019,** debatpintarjokowi, prabowomenangdebat, debatcapres2019, debatcawapres2019
3rd Community	**pilpres2019,** viral, cekfakta, jokowipelanggarham, periksa-fakta
4th Community	**pemilu2019,** inaelectionobserversos, auditforensikkpu, kpu-jangancurang, 02wintheelection
5th Community	**jokowi, presidenjokowi, jokowiamin,** sandiagauno, jokowi-amintegakkanham
6th Community	indonesiamaju, 2019pilihjokowi, jokowilagi
7th Community	**prabowothenextleader,** hoaxgantiuntung, 17aprilemak2kepung-tps
8th Community	-
9th Community	01indonesiamaju, jokowinyatakerjanya
10th Community	savemuslimuyghur, usirdubeschina, fasischina

TABLE 5.17: Top 10 Communities with Influential Hashtags - Hashtag Co-occurrence Network

Figure 5.17 outlines the number comparison between True News and False News of nodes and edges in each community for the hashtag co-occurrence network. From the top ten communities of the hashtag co-occurrence network, there are four communities that have more false news circulating rather than true news, five communities with more true news, and one community which have neither of the news, indicating all the news in there is classified as Other 1 and Other 2.

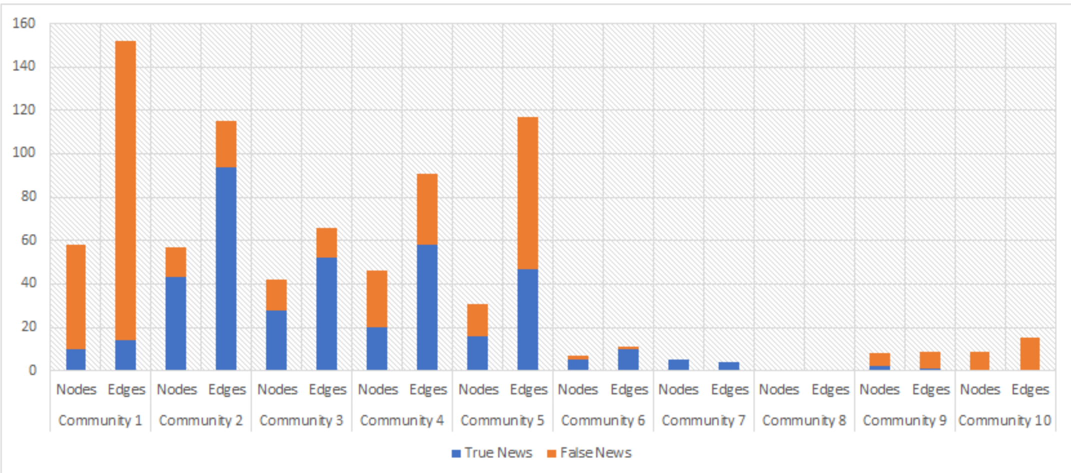

FIGURE 5.17: Distribution of True News and False News - Top 10 Communities of Hashtag Co-occurrence Network

The top community is the 1st Community which tends to be infested with false news as shown in Figure 5.18 and 5.19. Few prominent influential hashtags of this community are *2019gantipresiden* (2019 changes president), *prabowo* (The name of one of the president candidates), *rakyatsudahmuak* (people have had enough), and *2019prabowopresiden* (2019 prabowo is the president). It is quite clear that this hashtags community supports Prabowo Subianto as the opposing candidate to defeat Joko Widodo as the incumbent. Meanwhile, the 5th Community supports Joko Widodo with some influential hashtags *jokowi*, *presidenjokowi* (president Jokowi), and *jokowiamin* (Jokowi and Amin, his couple as vice president, but this also could be interpreted as "Jokowi, Amen."). This community also tends to favor false news as well, shown in Figure B.37 and B.38. Other two "false news" communities show support to Jokowi (the 9th Community) and strong negative sentiment to China (the 10th Community).

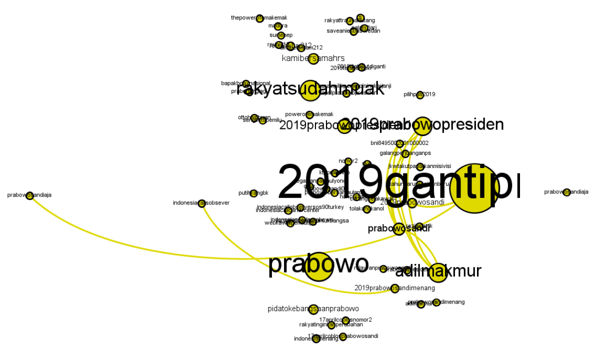

FIGURE 5.18: 1st Community of Hashtag Co-occurrence Network - True News Dissemination

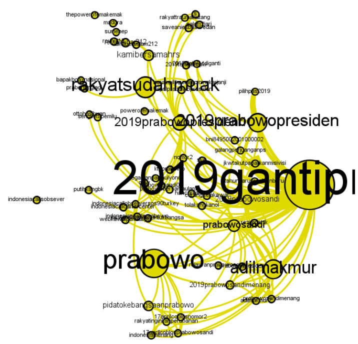

FIGURE 5.19: 1st Community of Hashtag Co-occurrence Network - False News Dissemination

The top ten communities of hashtag co-occurrences consist of more true news communities, contrary to the mention communities. One of the communities is the 2nd Community that has *debatpilpres2019* (2019 presidential debate) hashtag as the most influential hashtag. This community mostly consists of debate-related events. Other top true news communities such as the 3rd and 4th Community also tends to be related with events by influential hashtags *pilpres2019* (2019 presidential election) and *pemilu2019* (2019 general election), shown in Figure B.33, B.34, B.36, and B.36. This indicates that hashtag communities which discuss figures, in particular president candidates, tend to be related with false news filter bubble than communities which discuss events. Most likely, these hashtags are sentimental in their support or rejection and therefore prone to embellish news which have strong sentimental factors and not thoroughly undergo fact-checking.

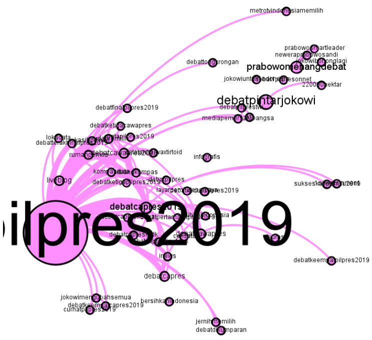

FIGURE 5.20: 2nd Community of Hashtag Co-occurrence Network - True News Dissemination

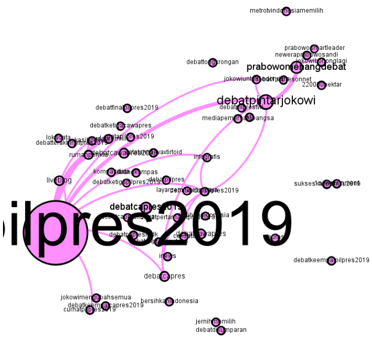

FIGURE 5.21: 2nd Community of Hashtag Co-occurrence Network - False News Dissemination

Table 5.18 outlines the top ten influential hashtags according to the betweenness centrality score and other influential hashtags each reaches, along with the classes.

Usernames	**Community**	**Reached Influential Hashtags**
debatpilpres2019	2^{nd}	pemilu2019 (TN), pilpres2019 (TN, FN), presidenjokowi (TN), rakyatsudahmuak (FN), debatpintarjokowi (FN)
pemilu2019	4^{th}	debatpilpres2019 (TN), prabowo (FN), pilpres2019 (TN, FN), presidenjokowi (TN, FN), adilmakmur (FN)
2019gantipresiden	1^{st}	prabowo (FN), prabowothenextleader (FN), rakyatsudahmuak (FN), adilmakmur (FN), 2019prabowopresiden (FN)
jokowi	5^{th}	prabowo (O), presidenjokowi (FN), jokowiamin (FN), 2019prabowo-presiden (O), jokowiamintegakkanham (FN)
prabowo	1^{st}	pemilu2019 (FN), 2019gantipresiden (FN), jokowi (O), pilpres2019 (TN), adilmakmur (FN)
pilpres2019	3^{rd}	debatpilpres2019 (TN, FN), pemilu2019 (TN, FN), prabowo (TN), viral (TN), cekfakta (FN)
presidenjokowi	5^{th}	-debatpilpres2019 (TN), pemilu2019 (TN, FN), jokowi (FN), jokowiamin (FN), sandiagauno (TN)
jokowiamin	5^{th}	jokowi (FN), presidenjokowi (FN), indonesiamaju (O)
prabowothenextleader	7^{th}	2019gantipresiden (FN), hoaxgantiuntung (O)
rakyatrsudahmuak	1^{st}	debatpilpres2019 (FN), 2019ganti-presiden (FN), 2019prabowopresidenri (FN)

TABLE 5.18: Top 10 Influential Hashtags with Other Reached Influential Hashtags - Hashtag Co-occurrence Network

For the centrality analysis, figure 5.22 outlines the top ten influential hashtags based on the value of each betweenness centrality. From the top ten influential hashtags, there are six hashtags relate to more false news rather than true news, three hashtags with more true news, and one hashtag which relate fairly balanced with both news. There are two communities which have more than one influential hashtag and each of these communities relates with each of president candidates.

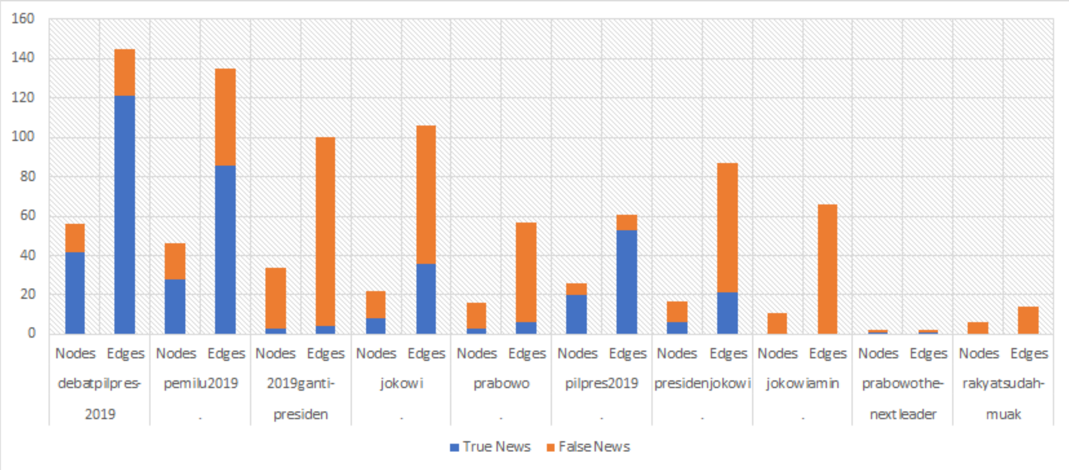

FIGURE 5.22: Distribution of True News and False News - Top 10 Influential Hashtags of Hashtag Co-occurrence Network

Note that the six influential hashtags of false news are belonged to two communities which relate to each of the president candidates. Hashtag *2019gantipresiden* is the most influential hashtag of false news, shown in Figure 5.23 and 5.24. This hashtag, along with *prabowo* and *rakyatsudahmuak* hashtags are connected by false news and all three of them also reach influential users in a various communities. One interesting example is *prabowothenextleader* hashtag that is reached by *2019gantipresiden*. Although it belongs to a different community, the hashtag strongly suggests a support for Prabowo and all of the influential hashtags are connected by false news posts. On the other hand, *jokowi*, *presidenjokowi*, and *jokowiamin* hashtags are also connected by false news posts yet have closed false news circulation exclusively in their community, as shown in Figure B.51, B.52, B.57, B.58, B.59, and B.60. It is important to mention that both of these influential users sets are not connected through true news posts. In addition, both communities have hashtags that do not reverberate or receive any true news, which are *jokowiamin* and *rakyatsudahmuak*.

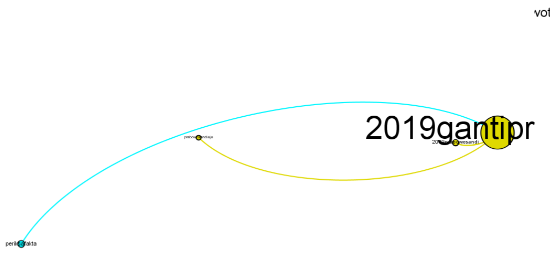

FIGURE 5.23: *2019gantipresiden*'s Network - True News Dissemination

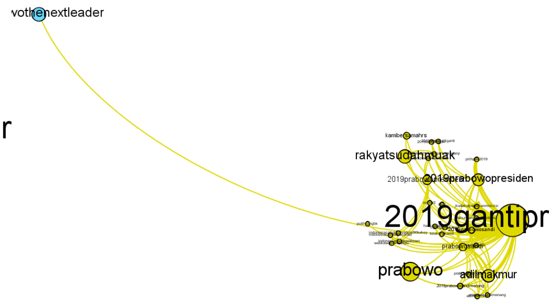

FIGURE 5.24: *2019gantipresiden*'s Network - False News Dissemination

The three influential hashtags of true news belong to different communities. Nevertheless, they are still directly connected to each other through true news posts. One example is *debatpilpres2019* as shown in Figure 5.25 and 5.26, and the others are shown in Figure B.49, B.50, B.55, and B.56. Either true news or false news, both influential hashtags tend to be connected to each of the respective classes. Be that as it may, false news influential hashtags have a greater chance to flock together in the same community rather than true news influential hashtags. False news influential hashtags also have more personal and sentimental nature since they relate with the candidates. This indicates that the set of false news hashtags tends to be written together in the same post, more than the set of true news hashtags.

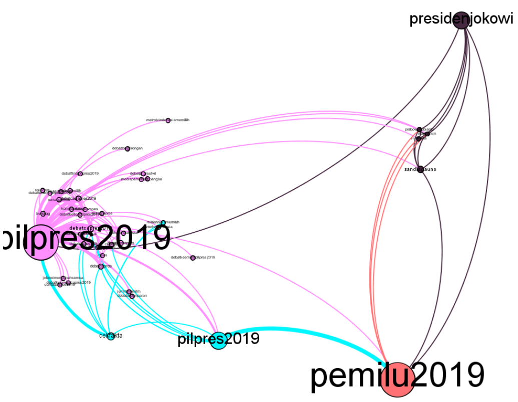

FIGURE 5.25: *debatpilpres2019*'s Network - True News Dissemination

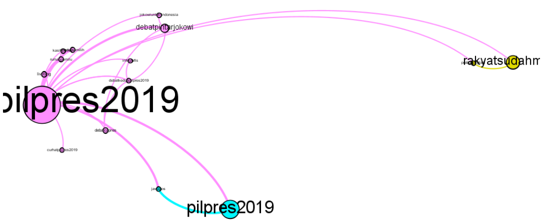

FIGURE 5.26: *debatpilpres2019*'s Network - False News Dissemination

Based on the social network analysis experiment of the hashtag co-occurrence network with the 2019 Indonesian presidential election setting, the interesting findings are:

- False news associate more with top influential hashtags rather than true news.
- False news are more likely related to supportive or sentiment-induced hashtags than hashtags about events or occurrences.
- The inclined "false news" influential hashtags tend to be connected to each other more than inclined "true news" influential hashtags, implying the inclined "false news" hashtags are more frequent and uniform to be posted together.

5.5 Updated Social Network Analysis Experiment

This experiment is conducted on a new 2IPEN data set to observe how the new networks can be created and analyse the alteration from the previously established networks. While the previous networks are generated from the 2IPEN data set that are labelled by annotators, this new 2IPEN data set is labelled by a text classification model trained from the original 2IPEN data set with five classes setting. The total data of the new 2IPEN data set is 6,424. The model is created with bigram text feature, the feature with best performance for five classes setting. The detailed statistics of both networks are outlined in Table 5.19.

Statistics	**Updated Mention (Directed)**	**Updated Hashtag Co-occurrence (Undirected)**
# Nodes	1,891	1,302
# Edges	2,582	4,315
Diameter	3	12
Average Degree	1.365	6.628
# Strongly Connected Components	1,890	-
# Weakly Connected Components	145	118
# True News Nodes	768	625
# False News Nodes	859	631
# True News Edges	841	2,213
# False News Edges	1,043	1,655
Modularity	0.737	0.817
Communities	165	133

TABLE 5.19: Updated Network Data Properties

5.5.1 Updated Mention Network

The updated mention network is illustrated in Figure 5.27. Some influential usernames, which also appeared in the original mention network, could be identified such as *bawaslu_ri*, *gunromli*, abd *divhumas_polri*.

Table 5.20 outlines the top ten largest communities according to the number of the nodes and the influential usernames that each community has. There are ten

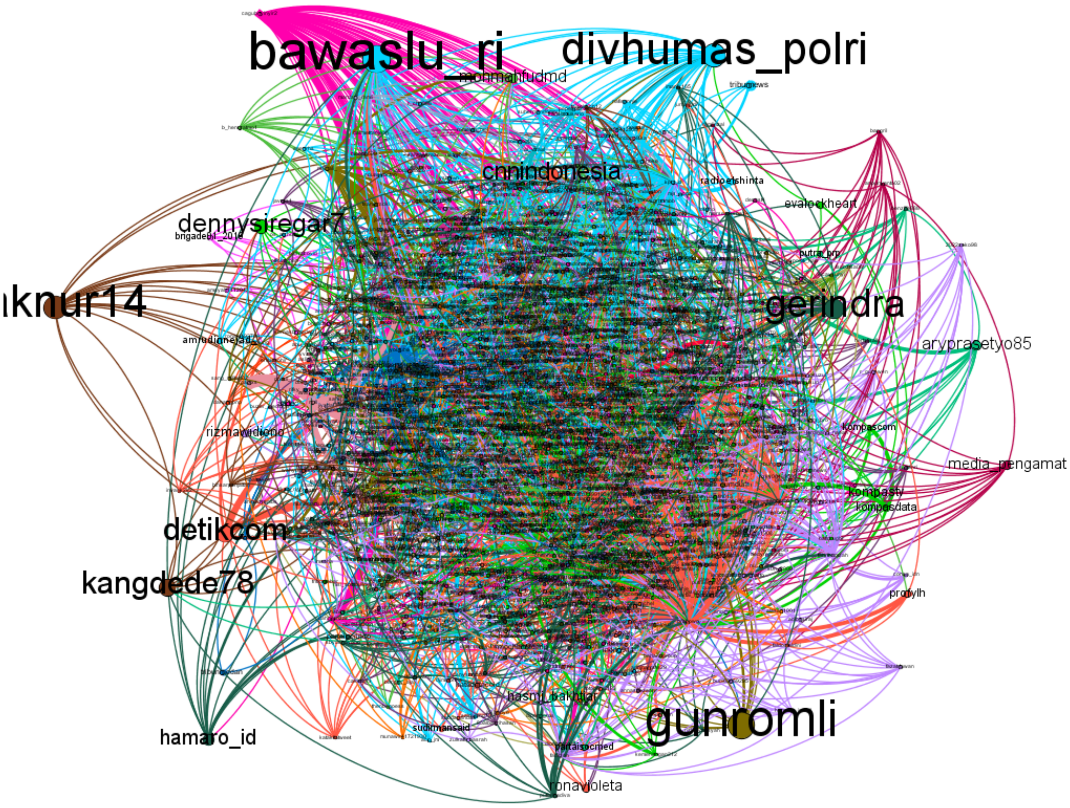

FIGURE 5.27: Updated Mention Network

verified accounts: six online news sources (*detikcom*, *cnnindonesia*, *radioelshinta*, *kompastv*, *kompasdata*, and *kompascom*), two government institutions (*bawaslu_ri* and *divhumas_polri*), one political party (*gerindra*), and one politician (*gunromli*).

Communities	Influential Usernames
1st Community	**detikcom (Ver.), cnnindonesia (Ver.),** ronavioleta, profylh, sudirmansaid
2nd Community	**bawaslu_ri (Ver.), divhumas_polri (Ver.),** radioelshinta (Ver.), amrudinnejad_, tribunnews
3rd Community	**dennysiregar7,** kompastv (Ver.), evalockheart, kompasdata (Ver.), kompascom (Ver.)
4th Community	rizmawidiono
5th Community	**gerindra (Ver.), hamaro_id,** partaiperindo (Ver.)
6th Community	**gunromli (Ver.),** muannas_alaidid, yusuf_dumdum
7th Community	aadec, semiaji_w, valencia_amara
8th Community	tribunmedan
9th Community	aryprasetyo85, hasmi_bakhtiar, partaisocmed, putra_prp, mochamadarip
10th Community	liem_id

TABLE 5.20: Top 10 Communities with Influential Usernames - Updated Mention Network

Figure 5.28 outlines the number comparison between True News and False News of nodes and edges in each community for the updated mention network. There are four communities that have more false news circulating rather than true news, one community with more true news, and five communities which have fairly balanced proportion of both news.

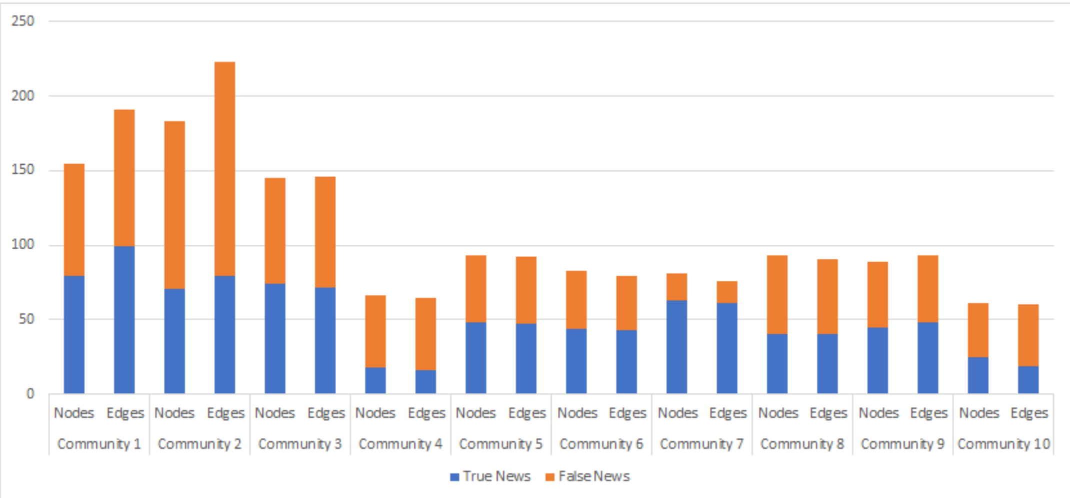

FIGURE 5.28: Distribution of True News and False News - Top 10 Communities of Updated Mention Network

Table 5.20 outlines the top ten influential usernames according to the betweenness centrality score and other influential usernames each reaches. The new seen influential usernames, but are included in the previous experiment's network, are *caknur14* and *kangdede78* which belong to the 11^{th} Community.

Usernames	Community	Reached Influential Usernames
bawaslu_ri (Ver.)	2	hamaro_id (O), radioelshinta (Ver., FN)
gunromli (Ver.)	6	-
divhumas_polri (Ver.)	2	evalockheart (TN), radioelshinta (Ver., TN), muannas_alaidid (TN)
caknur14	11	kangdede78 (FN)
gerindra (Ver.)	5	cnnindonesia (Ver., TN), putra_prp (O)
kangdede78	11	caknur14 (FN)
detikcom (Ver.)	1	-
dennysiregar7	3	-
cnnindonesia (Ver.)	1	gerindra (Ver., TN)
hamaro_id	5	bawaslu_ri (Ver., O), putra_prp (O)

TABLE 5.21: Top 10 Influential Usernames with Other Reached Influential Usernames - Updated Mention Network

Figure 5.29 outlines the number comparison between True News and False News of nodes and edges in each username for the updated mention network. From the top ten influential users of the mention network, there are seven users that disseminate or receive more false news rather than true news, two users with more true news, and one user which has fairly balanced proportion of both news.

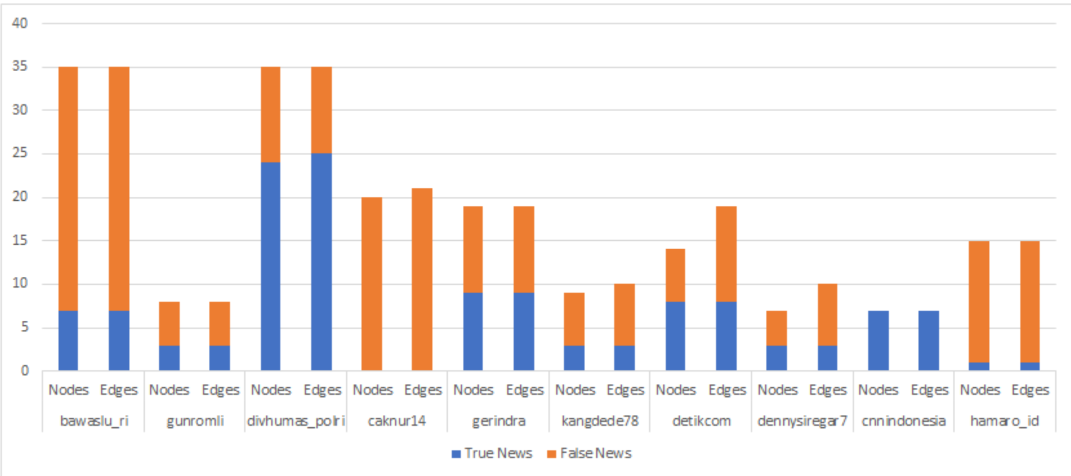

FIGURE 5.29: Distribution of True News and False News - Top 10 Influential Usernames of Mention Network

The comparison of insights between the original and updated mention network with the 2019 Indonesian presidential election setting are:

- False news are still spread more in the top communities of usernames and more likely being disseminated or received by top influential usernames rather than true news. However, there are more communities with a balanced proportion between true news and false news in the updated network. Two of these communities have many news source accounts. For verified accounts, they still tend to spread more false news than true news in the updated setting, where three of the top four influential usernames disseminate false news more. The two largest, *bawaslu_ri* and *gunromli*, are official and politically-related account.

- The communities that are prone with more false news still have more influential usernames than the inclined "true news" communities, especially considering only one inclined "true news" community is formed. However, there are more influential usernames found in a "balanced" community than an inclined "false news" community.

- While the margin of true news and false news based on the communities are mostly balanced, on influential usernames they still have large margin, especially *bawaslu_ri*, *caknur14*, and *hamaro_id*, an unverified account, for false news. One inclined "true news" username who has a large margin with false news is *cnnindonesia*, which is an official news source account. This suggests that influential actors are prone to spread or receive false news longer and false news filter bubbles do not stay very long.

- One of the top “true news” influential usernames is *divhumas_polri*[1]. This is to be expected since they have a cyber division dedicated to fight back hoax.

- In the updated network, verified usernames also have more connection with other influential usernames as well, alike unverified usernames from the previous network. This is most probably caused by the growing size of data.

5.5.2 Updated Hashtag Co-occurrence Network

The updated hashtag co-occurrence network is illustrated in Figure 5.30. From the network, some influential hashtags, which also appeared in the original hashtag co-occurrence network, could be identified such as *pilpres2019*, *2019gantipresiden*, and *debatpilpres2019*.

Table 5.22 outlines the top ten largest communities according to the number of the nodes and the influential hashtags that each community has.

Communities	Influential Hashtags
1st Community	**pilpres2019, 2019gantipresiden, prabowo,** 2019prabowopresiden, berita
2nd Community	**jokowi,** informasi, klarifikasi, hoaks, nasional
3rd Community	jaekingoflies, jaengibuldimanalagi, uninstalljaenow, shameonyoujae, jaerajahoax
4th Community	**debatpilpres2019,** cekfakta, debatpintarjokowi, debatcapres, debatcapres2019
5th Community	**01jokowilagi,** indonesiamaju, jokowilagi, 01indonesiamaju, 2019tetapjokowi
6th Community	**pemilu2019,** inaelectionobserversos, saveourdemocracy, kpujangancurang, jokowipelanggarham
7th Community	**hoax,** news, jokowimaruf, beritaterkini, pan
8th Community	**indonesia,** madura, jatim, surabaya, rabu
9th Community	viral, jakarta, reuni212, fakta, foto
10th Community	**indonesianeedsprabowo,** tutup01tusuk02, tusukprabowosandi, uasdifitnahkejibalasditps, tanpakecuranganprabowomenang

TABLE 5.22: Top 10 Communities with Influential Hashtags - Updated Hashtag Co-occurrence Network

[1]The official account of Indonesian republic police force.

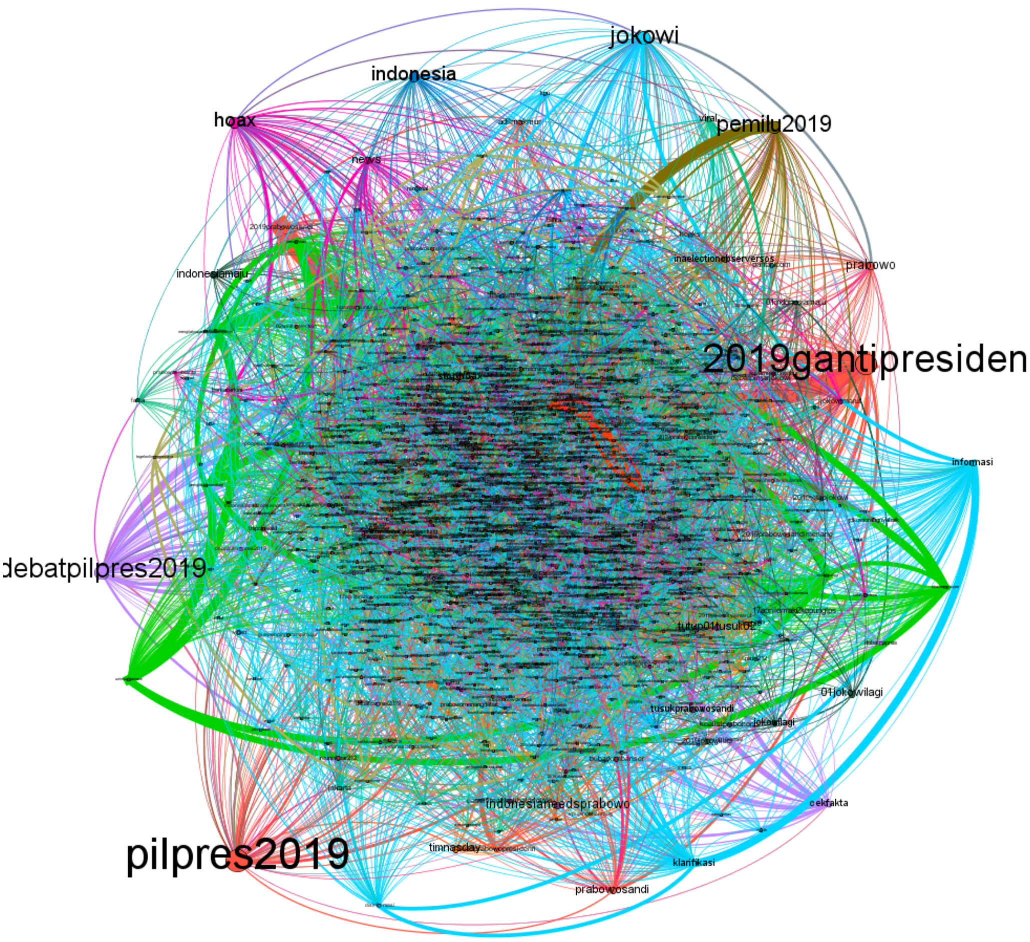

FIGURE 5.30: Updated Hashtag Co-occurrence Network

Figure 5.31 outlines the number comparison between True News and False News of nodes and edges in each community for the hashtag co-occurrence network. From the top ten communities of the hashtag co-occurrence network, there are three communities that have more false news circulating rather than true news, five communities with more true news, and two community which relate fairly balanced with both news.

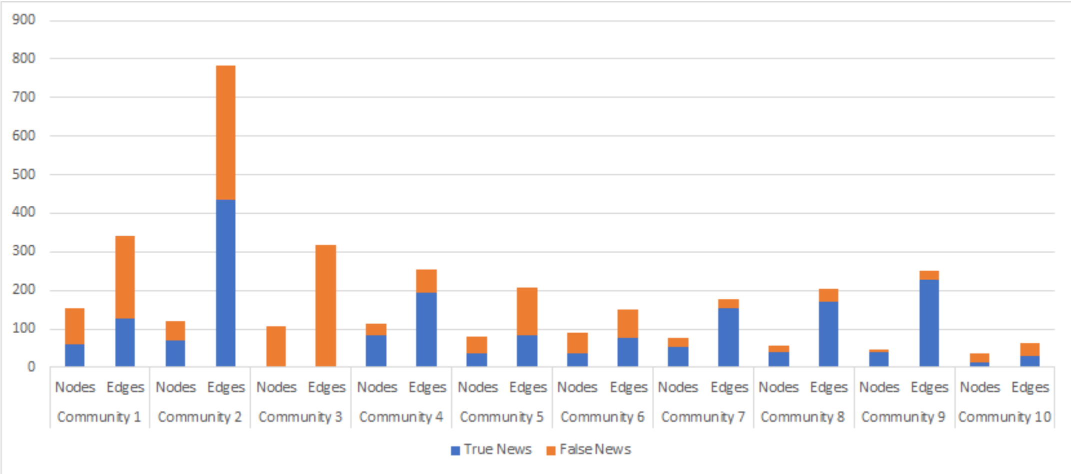

FIGURE 5.31: Distribution of True News and False News - Top 10 Communities of Updated Hashtag Co-occurrence Network

Table 5.23 outlines the top ten influential hashtags according to the betweenness centrality score and other influential hashtags each reaches, along with the classes.

Usernames	Community	Reached Influential Hashtags
pilpres2019	1	2019gantipresiden (TN, FN), debatpilpres-2019 (TN, FN), jokowi (TN, FN), pemilu-2019 (TN, FN), hoax (TN)
2019gantipresiden	1	pilpres2019 (TN, FN), hoax (TN), prabowo (FN), indonesianeedsprabowo (FN), news (TN)
debatpilpres2019	4	pilpres2019 (TN, FN), pemilu2019 (TN), indonesia (TN), prabowo (TN), 01jokowi-lagi (FN)
jokowi	2	pilpres2019 (TN, FN), pemilu2019 (TN), hoax (TN), indonesia (TN), prabowo (TN),
pemilu2019	6	pilpres2019 (TN, FN), debatpilpres2019 (TN), jokowi (TN), indonesia (TN), prabowo (FN)
hoax	7	pilpres2019 (TN), 2019gantipresiden (TN), jokowi (TN), prabowo (TN), news (TN)
indonesia	8	pilpres2019 (TN), debatpilpres2019 (TN), jokowi (TN), pemilu2019 (TN), prabowo (TN)
prabowo	1	pilpres2019 (FN), 2019gantipresiden (FN), debatpilpres2019 (TN), jokowi (TN), pemilu2019 (FN)
indonesianeeds-prabowo	10	2019gantipresiden (FN), tutup01tusuk02 (FN), 2019prabowopresidenri (FN), ranpa-kecuranganprabowomenang (FN), yusril-norakdankasarmainnya (FN)
01jokowilagi	5	debatpilpres2019 (FN), jokowilagi (FN), 01indonesiamaju (TN, FN), debatpintar-jokowi (FN), koalisiprabohong (TN)

TABLE 5.23: Top 10 Influential Hashtags with Other Reached Influential Hashtags - Updated Hashtag Co-occurrence Network

Figure 5.32 outlines the top ten influential hashtags based on the value of each betweenness centrality. From the top ten influential hashtags, there are three hashtags relate to more false news rather than true news, six hashtags with more true news, and one hashtag which relate fairly balanced with both news.

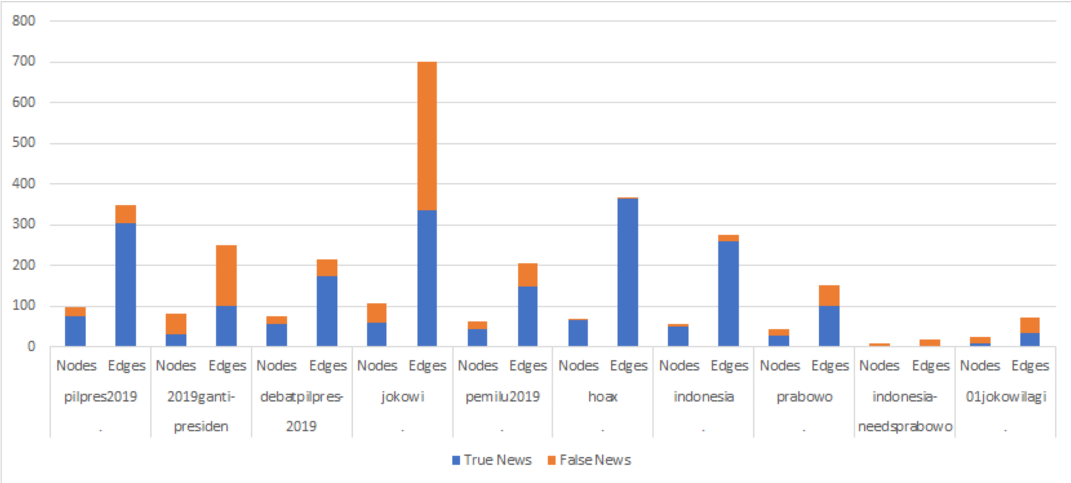

FIGURE 5.32: Distribution of True News and False News - Top 10 Influential Hashtags of Updated Hashtag Co-occurrence Network

The comparison of insights between the original and updated hashtag co-occurrence network with the 2019 Indonesian presidential election setting are:

- True news are more likely being associated with top influential hashtags, different from the result of the original hashtag co-occurrence network.
- False news are still more likely to infest supportive or sentiment-induced hashtags than hashtags about events or occurrences. The new sentiment-induced hashtags are *indonesianeedsprabowo* and *01jokowilagi* (01 Jokowi again), although they co-occur little with other hashtags.
- The inclined "false news" and "true news" influential hashtags tend to co-occur in a mixture in this updated network, implying that there are also occurrences where "true news" and "false news" hashtags are posted together.
- 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.

Chapter 6

Conclusion & Future Works

6.1 Conclusion

This research is conducted to detect false news and analyse its dissemination regarding the 2019 Indonesian presidential election on Twitter through text mining and social network analysis disciplines, with proposed research questions are outlined in Chapter 1. To answer those questions, we collected a Twitter data set regarding the 2019 Indonesian presidential election. To make the topics more focused and within a clear boundary, the topics were based on Indonesian fact-checking websites. Afterwards, a sample of roughly quarter of the whole data set is assigned to annotators where the final results produce five categories: True News, False News, Misleading News, Others 1, and Others 2.

From text mining perspective, the most prominent text feature to detect and distinguish true news, false news, and misleading news is word n-grams, particularly any composition that includes unigram. Bigram compositions performed fairly well for Indonesian texts, on some occasions outperform unigram compositions. Trigram compositions performed the worst in any setting of word-ngrams, yet still performed better than orthographies or sentiment lexicon features. Orthography features present that the best composition of True News is the worst composition of False News and Misleading News, and contrariwise. The voting ensemble of these three text features produce more favourable results, enabling the possibility to keep incorporating orthographic and sentiment lexicon feature. For Indonesian social media posts, abusive languages strongly correlate with false news and misleading news. Meanwhile, hate speeches have moderate correlation with false news and misleading news.

From social network analysis perspective, the communities with the most users tend to have more false news circulating rather than true news. The inclined "false news" communities also have more influential usernames. The top influential users based on their betweenness centrality also tend to disseminate more false news rather than true news, which many of them are verified. The inclined "false news" influential users spread false news in a large gap compared with true news, while the inclined "true news" influential users tend to have smaller gap. Regarding the hashtags, the strong sentiment hashtags to either support or oppose the candidates are more related with false news, while hashtags about general events or terms are not. Furthermore, the inclined "true news" hashtags are too general or do not directly related with the presidential election itself.

6.2 Future Works

Regarding the future works, this research suggests the following improvements:

- This research utilized news fact-checking websites to define a clear boundary of the topics and facilitate limiting factors of the annotation process, such as manpower and time. The result is the 2019 Indonesian Presidential Election (2IPEN) data set, which has five classes: True News, False News, Misleading News, Others 1, and Others 2. As the main focus only aimed to true, false, and misleading news, this Indonesian data set could be expanded further with false news from various topics. The original and additional 2IPEN data sets are stored in GitHub[1].

- This research utilized word n-grams, orthography, and sentiment lexicon features to detect and distinguish true news, false news, and misleading news. Other text features such as vector and part-of-speech tagging, with a deep learning approach, are suggested for further research. Orthography and sentiment lexicon were underperformed compared to word n-grams but nevertheless are still potential for social media data set and have many areas to be investigated and improved, especially their transformation.

- This research utilized social network visualizations by Gephi to observe how the news disseminate, according to usernames and hashtags. The objects of observation are filter bubble effects and influential usernames or hashtags, by applying community detection and betweenness centrality, respectively. Other network measures could be employed to learn and probe various phenomena in the network, for example PageRank for directed type network or Eigen Centrality for undirected to observe the most influential actors based on their reference from other actors.

[1]https://github.com/rayansuryadikara/false_news_detection_and_dissemination_analysis

Appendix A

List of Topics

Topic/Query (in Indonesian)	Translation and Supporting Link	First Posted
Jokowi dan Tol Trans Papua	Jokowi and Trans Papua Highway	23/09/2018
Lawan Berat Prabowo	Prabowo's Fierce Opponent	23/09/2018
Najwa Dukung Prabowo-Sandi	Najwa Supports Prabowo-Sandi	23/09/2018
Polri Blokir Situs Skandal-sandiaga	Police Blocks Skandalsandi-aga Site	23/09/2018
Prabowo Produk Asing	Prabowo is Foreign Product	23/09/2018
Angka Nol Nomor Urut	Zero Ballot Number	25/09/2018
Beras Toko Tani	Toko Tani Rice	25/09/2018
TNI AU Disiksa Cina	Air Force Tortured by China	25/09/2018
Muhammadiyah Dukung Jokowi	Muhammadiyah Supports Jokowi	26/09/2018
Nahdliyin Dukung Prabowo	Nahdliyin Supports Prabowo	28/09/2018
Presiden Petugas Partai	President is Party Officer	28/09/2018
Rampok Rumah Terbakar	Ransacking The Burning House	28/09/2018
Siap Pindah Warga Negara	Ready to Emigrate	28/09/2018
Tabligh Akbar Doa Bangsa	Great Tabligh Nation Prayer	28/09/2018
Infrastruktur Palu	Palu's Infrastructure	30/09/2018
FPI Bantu Gempa Palu	FPI Helps The Palu Earth-quake	01/10/2018
Pesindo Organisasi PKI	Pesindo is PKI Organization	01/10/2018
Anis Matta Dukung Jokowi	Anis Matta Supports Jokowi	02/10/2018
Tokoh PKI Ratna Sarumpaet	PKI Figure Ratna Sarumpaet	03/10/2018
Jokowi Bersulang Wine	Jokowi Toasts Wine	06/10/2018
Batas Pengecekan Pemilih Presiden	Time Limit of President Vot-ers Check	07/10/2018
Kaos Khabib Nurmagome-dov	Khabib Nurmagomedov's T-Shirt	10/10/2018
Bantuan Turki untuk Gempa Palu	Turkey's Aid for The Palu Earthquake	11/10/2018
Bonus Pelatih Asian Games	Asian Games Coaches' Bonus	16/10/2018
Mahasiswa Kendari Dipukul Aparat	Kendari Students Beaten by Officers	18/10/2018

Jokowi Hapus Kementerian Agama	Jokowi Removes Ministry of Religion	19/10/2018
Data Pemilih Kemendagri	Ministry of Affairs' Voters Data	20/10/2018
Soal CPNS Bocor	CPNS Tests Leaking	20/10/2018
Banser Bakar Bendera Tauhid	Banser Burns The Tawhid Flag	22/10/2018
Bendera Arab Saudi dan Markas PBB	Saudi Arabia Flag and UN Headquarters	22/10/2018
Bentrokan Ormas Cikarang	Clash of Cikarang Mass Organizations	22/10/2018
Anies Dipayungi Jokowi	Anies under The Umbrella of Jokowi	23/10/2018
CFD Membiru	CFD Turns Blue	25/10/2018
Tagar PrabowobersamaHTI	Hashtag PrabowobersamaHTI	25/10/2018
Muslim Rusia dan Bendera Tauhid	Russian Muslims and Tawheed Flag	27/10/2018
Guru Honorer Demonstran Meninggal	Honorary Teacher Demonstrator Dies	31/10/2018
Korupsi Dana Desa	Village Fund Corruption	04/11/2018
BDCS dan Internet	BDCS and Internet	05/11/2018
BPN dan Hotman Paris	BPN and Hotman Paris	06/11/2018
Aksi 911 Bela Tauhid	911 Act to Defend Tawhid	07/11/2018
Turki Melacak Pemfitnah Rizieq	Turkey Traces The Slanderer of Rizieq	07/11/2018
Tiket Gratis Film Ahok	Free Tickets for Ahok Movie	08/11/2018
Bendera Tauhid Tidak Dilarang	Tawhid Flag is Not Prohibited	09/11/2018
Al-Jazeera dan Kasus Rizieq	Al-Jazeera and Rizieq Case	10/11/2018
Mahfud Mendukung Prabowo-Sandi	Mahfud Supports Prabowo-Sandi	13/11/2018
Foto Syur Grace Natalie	Sexy Photo of Grace Natalie	14/11/2018
Revolusi Mental PSI	PSI's Mental Revolution	15/11/2018
Gigi dan Prabowo-Sandi	Gigi and Prabowo-Sandi	16/11/2018
Kartu Nikah Poligami	Polygamy Marriage Card	16/11/2018
BIN Suap Mahasiswa	BIN Bribes Students	17/11/2018
Erdogan dan Utang indonesia	Erdogan and Indonesia Debt	17/11/2018
Ridwan Kamil dan Film Ahok	Ridwan Kamil and Ahok Movie	17/11/2018
KAI dan Rombongan 212	KAI and 212 Group	22/11/2018
Cina Kirim Warganya	China Sends Its Citizens	23/11/2018
Pidato Prabowo Profesor Fisika	Prabowo's Speech on Professors of Physics	24/11/2018
KTP Prabowo	Prabowo ID card	26/11/2018
Foto Prabowo Presiden	Photo of President Prabowo	27/11/2018
Jalan Kaki dari Solo ke Jakarta	On Foot from Solo to Jakarta	27/11/2018

Reuni 212 Ojol dan Guru	212 Reunion with Online Ojeks and Teachers	29/11/2018
Indosat SMS Reuni 212	Indosat Short Messages 212 Reunion	30/11/2018
Nusron Peserta Aksi 212	Nusron 212 Act Participants	01/12/2018
Padang Kirim Rendang	Padang Sends Rendang	01/12/2018
Posko Puan 212	Puan's Post 212	01/12/2018
Karni Ilyas Dipanggil Istana	Karni Ilyas is Summoned by Istana	02/12/2018
PBNU Kedubes dan Arab Saudi	PBNU and Embassy of Saudi Arabia	04/12/2018
Ma'ruf Jatuh Sakit	Ma'ruf Fell Ill	05/12/2018
WNA Status WNI	Foreign Citizens with Indonesian Citizen Statuss	05/12/2018
Jusuf Kalla Dukung Prabowo	Jusuf Kalla Supports Prabowo	07/12/2018
Rekening Ketua KPK	KPK Chairperson's Account	07/12/2018
Bom Reuni 212	212 Reunion Bomb	08/12/2018
Media Jepang 212	Japanese Media on 212	08/12/2018
Prabowo Berpakaian Pendeta	Prabowo Dresses as Priest	11/12/2018
Puan Negara Maju	Puan and Advanced Country	11/12/2018
Kinerja SBY Pembangunan Sumut	SBY Performance of North Sumatra Development	16/12/2018
Munir Dibunuh Setelah Kesaksian	Munir was Killed after Testimony	16/12/2018
Ikut Campur Uighur	Interference on Uyghurs	18/12/2018
Jepang Memprotes China Uighur	Japanese Protests on Chinese Uyghurs	19/12/2018
Tentara Cina Menghanguskan Indonesia	Chinese Army Scorches Indonesia	19/12/2018
Penyiksaan Muslim Uighur	Torture of Uyghur Muslims	22/12/2018
Jokowi Pilihan UAS	Jokowi is UAS Choice	23/12/2018
Simulasi Orang Gila TPS	Polling Place Simulation with Mad Men	23/12/2018
Tempo Bencana Tertinggi Jokowi	Tempo and Most Disasters on Jokowi	24/12/2018
Jangan Pilih Pemimpin Suap	Do Not Vote A Bribing Leader	25/12/2018
Panwaslu Membuka Kecurangan KPU	Election Commitee Reveals KPU Fraud	26/12/2018
Syekh Ali Jaber Jokowi	Sheikh Ali Jaber and Jokowi	26/12/2018
Jokowi Resmikan Patung Yesus	Jokowi Inaugurates Jesus Statue	28/12/2018
Tes Membaca Al-Qur'an Prabowo	Reading Qoran Test and Prabowo	31/12/2018
MUI Potong Tangan Koruptor	MUI Cuts Corruptors' Hands Off	01/01/2019

Viking dan Capres Cawapres 01	Vikings and 01 Presidential Candidate	01/01/2019
Sandiaga Bangun Cikopo-Palimanan	Sandiaga Builds Cikopo-Palimanan	02/01/2019
Surat Suara 01 Tercoblos	Pierced 01 Ballots	02/01/2019
Pecat TNI Sweeping PKI	Fire TNI who Swept PKI	03/01/2019
Jokowi Cucu Sunan Kalijaga	Jokowi is Sunan Kalijaga's Grandson	04/01/2019
Surat Suara Tercetak	Ballots were Printed	04/01/2019
Kader PKS Penyewa Vanessa Angel	PKS Politician is Vanessa Angel's Renter	06/01/2019
Ketua KPU Saudara Soe Hok Gie	KPU Chairperson is Soe Hok Gie's Brother	06/01/2019
Suara Terkecil Ketua KPU	The Smallest Vote of KPU Chairperson	06/01/2019
Khofifah dan Ganjar Satu Jari	Khofifah and Ganjar One Finger	08/01/2019
Ustad Arifin Naik Pesawat Prabowo	Ustad Arifin Boards Prabowo's Airplane	09/01/2019
Panglima TNI Dua Jari	TNI Commander Two Fingers	11/01/2019
Garbi Dukung Jokowi	Garbi Supports Jokowi	12/01/2019
Alumni UNS Dukung Jokowi	UNS Graduates Support Jokowi	13/01/2019
Tabligh Solo Dihadang Polisi	Solo Tabligh Blocked by Police	13/01/2019
Prabowo dan Kasus Bunuh Diri	Prabowo and Suicide Cases	14/01/2019
Tempo dan Elektabilitas Prabowo	Tempo and Prabowo's Electability	14/01/2019
Erdogan Puji Pidato Prabowo	Erdogan Praises Prabowo's Speech	15/01/2019
Ijazah SMA Jokowi Palsu	Jokowi's High School Diploma is Fake	15/01/2019
Pelanggaran HAM Berat	Severe Human Rights Violations	16/01/2019
Gerindra dan Caleg Eks-Koruptor	Gerindra and Ex-Corruptor Legislative Candidates	17/01/2019
Kalah Maka Masuk Neraka	Lose Then Enter Hell	17/01/2019
Menteri Jokowi dan impor Beras	Jokowi's Ministers and Rice Import	17/01/2019
Pekerjaan Penyandang Disabilitas	Jobs for People with Disabilities	17/01/2019
Pria Asing di Debat Pilpres	Foreign Man in Presidential Election Debate	18/01/2019
Dubes Eropa Dukung Prabowo	European Ambassadors Support Prabowo	19/01/2019
PWNU Jatim Dukung Prabowo	PWNU of East Java Supports Prabowo	22/01/2019

SBY Dukung Pemerintahan Jokowi	SBY Supports Jokowi's Government	22/01/2019
Shalat Jumat Prabowo	Prabowo's Friday Prayer	23/01/2019
Jokowi dan Bank Menarik Dana Nasabah	Jokowi and Bank Withdraw Customers' Funds	26/01/2019
Kartu E-Money Prabowo-Sandi	Prabowo-Sandi E-Money Card	26/01/2019
Rizieq Menarik Dukungan Prabowo	Rizieq Withdraws Support to Prabowo	28/01/2019
PSI Pasang Spanduk LGBT	PSI Puts LGBT Banners	30/01/2019
Iwan Fals Mendukung Jokowi	Iwan Fals Supports Jokowi	31/01/2019
Menteri Kominfo Intimidasi ASN	Minister of Communication and Information Intimidates Civil Servant	31/01/2019
Prabowo Berfoto dengan Cucu PKI	Prabowo Takes Photo with PKI's Grandchild	02/02/2019
TNI Menembaki Pemuda PKI	TNI Shoots on PKI Youths	02/02/2019
Bocah Bercita-cita Prabowo Menang	Boy Aspires for Prabowo to Win	04/02/2019
Sindiran Pembangunan Tutut Soeharto	Tutut Soeharto's Innuendo on Development	04/02/2019
Survei Pilpres Kompas	Kompas Presidential Election Survey	06/02/2019
Tulisan Ridwan Kamil tentang Kubu Sebelah	Ridwan Kamil's Writing about Other Faction	10/02/2019
Ali Ngabalin Sakit Stroke	Ali Ngabalin Has Stroke	11/02/2019
Mobil Berstiker Prabowo Dilarang Masuk	Prabowo-stickered Car is Denied Entry	13/02/2019
Hotman Paris Membela Prabowo	Hotman Paris Defends Prabowo	17/02/2019
Indonesia Mengembangkan Biodiesel 20	Indonesia Develops Biodiesel 20	17/02/2019
Indonesia Mengimpor Air	Indonesia Imports Water	17/02/2019
Kebakaran Lahan Gambut	Peatland Fire	17/02/2019
Pemerintah Membagikan Sertifikat Tanah	The Government Distributes Land Certificates	17/02/2019
Perusahaan Besar Melanggar Lingkungan	Large Corporations Violate The Environment	17/02/2019
Perusahaan Reklamasi dan Lubang Tambang	Reclamation Corporations and Mining Holes	17/02/2019
Alat Bantu Dengar Jokowi	Jokowi's Hearing Aid	18/02/2019
Jokowi Tersangka Penyebar Hoaks	Jokowi is Hoax Spreader Suspect	18/02/2019
Kacamata Google Prabowo	Prabowo's Google Glasses	19/02/2019
PDIP Senam di Sajadah	PDIP Does Gymnastics on Prayer Rugs	19/02/2019
Impor Beras Meningkat	Rice Import Increases	20/02/2019

Ibu Jokowi PKI	Jokowi's Mother is PKI	22/02/2019
Kedatangan TKA Berambut Cepak	Arrival of Short-Haired Foreign Workers	23/02/2019
Relawan Jokowi Hampir Bunuh Diri	Jokowi's Volunteer Almost Commits Suicide	23/02/2019
Yusuf Mansur Kampanyekan Jokowi	Yusuf Mansur Campaigns on Jokowi	23/02/2019
Prabowo Menegur Ulama	Prabowo Reprimands Ulamas	27/02/2019
Hari Poligami PKS	PKS Polygamy Day	02/03/2019
TKA Berseragam Tentara	Foreign Workers in Army Uniform	02/03/2019
Gus Sholah dan Acara Dukungan Prabowo	Gus Sholah and Supporting Prabowo Event	03/03/2019
Buzzer Polisi Dukung Jokowi	Police Buzzers Support Jokowi	05/03/2019
Kondom Jokowi-Ma'ruf	Jokowi-Ma'ruf Condoms	07/03/2019
Uang Kampanye Ditilep	Campaign Money is Embezzled	08/03/2019
Ibu Pingsan Depan Presiden	Woman Fainted before President	09/03/2019
Pasukan Asing Masuk NKRI	Foreign Troops Enter Indonesia Republic	09/03/2019
Demo Hoaks Ganti Untung	Compensation Hoax Demo	10/03/2019
Betawi Pilih Nomor 2	Betawi Chooses Number 2	11/03/2019
Pesawat Prabowo Dilarang Mendarat	Prabowo's Airplane is Denied Landing	11/03/2019
Maklumat Ponpes Dukung Prabowo	Pondok Pesantren's Announcement to Support Prabowo	13/03/2019
Mobil Prabowo dari Pemimpin ISIS	Prabowo's Car from ISIS leader	13/03/2019
Mobil Jokowi Masuk Got	Jokowi's Car Falls in Sewer	14/03/2019
Angka Kematian Ibu Indonesia	Indonesian Maternal Mortality Rate	17/03/2019
BPJS Tidak Membiayai Kanker	BPJS Does Not Cover for Cancer	17/03/2019
Hasil Riset Kurang Dimanfaatkan	Research Results are Underutilized	17/03/2019
Ibu Lis dan Debat Sandiaga	Mrs. Lis and Sandiaga's Debate	17/03/2019
Metro TV tentang Eggboy Pengagum Jokowi	Metro TV about Eggboy An Admirer of Jokowi	17/03/2019
Pengangguran Didominasi Lulusan SMK	Unemployment Dominated by Vocational School Graduates	17/03/2019
Penurunan Angka Pengangguran	Unemployment Rate Decreases	17/03/2019
Penurunan Angka Stunting	Stunting Rate Decreases	17/03/2019

Pendukung Jokowi di Palembang	Jokowi Supporters in Palembang	18/03/2019
Gerakan 150 Juta	150 Million Movement	20/03/2019
Camat Aman Guru Honorer Dipecat	Sub-district Chiefs are Safe Honorary Teachers are Fired	21/03/2019
Mobil Kampanye Berplat TNI	Campaign Car with TNI License Plate	21/03/2019
Papan Masjid Unsri Diretas	Unsri's Mosque Board is Hacked	21/03/2019
Polisi Acungkan Satu Jari	Police Raise One Finger	21/03/2019
PDIP Tidak Butuh Umat Islam	PDIP Does Not Need Muslims	23/03/2019
Hasil Survei LIPI	LIPI Survey Results	24/03/2019
Prabowo Tidak Sholat dan Minum Bir	Prabowo Does Not Pray and Drinks Beer	24/03/2019
Baliho Pilpres Gowa	Gowa Presidential Election Billboard	26/03/2019
Presiden Chechnya Dukung Prabowo	Chechnya President Supports Prabowo	27/03/2019
Baliho Berbahasa Cina di Kendari	Chinese Language Billboard in Kendari	29/03/2019
Anggaran Militer Indonesia Kecil	Indonesian Military Budget is Small	30/03/2019
Indonesia Penengah Konflik Negeri	Indonesia as The Intermediary of States Conflict	30/03/2019
Korupsi Indonesia Stadium Empat	Indonesian Corruption is Stage Four	30/03/2019
Pemerintah Membubarkan Lembaga Negara	Government Dissolves State Institutions	30/03/2019
Penguatan Pertahanan Pulau Jawa	Strengthening of Java's Defense	30/03/2019
Menlu Mengimbau KBRI Arab Saudi	Minister of Foreign Affairs Urges Indonesian Embassy in Saudi Arabia	31/03/2019
Ratusan Ribu Orang Kampanye Prabowo	Hundreds of Thousands in Prabowo's Campaign	01/04/2019
Spanduk PKS-HTI Ganti Presiden	PKS-HTI Replace President Banner	01/04/2019
Orasi Prabowo di Padang	Prabowo's Oration in Padang	02/04/2019
Surat Suara tanpa Tanda Tangan	Ballots without Signature	02/04/2019
Pesawat Prabowo Dihalangi Jet Tempur	Prabowo's Airplane is Obstructed by Fighter Jet	03/04/2019
Server KPU Diatur 57%	KPU Server is Fixed by 57%	03/04/2019
Lembaga Survei Pengkondisian 57%	Conditioning Survey Institutes 57%	04/04/2019
Anggaran Kotak Suara Kardus	Cardboard Ballot Box Budget	06/04/2019

Dana Haji Dipakai infrastruktur	Hajj Funds are Spent for infrastructure	06/04/2019
Coblos Tanpa Identitas di Arab Saudi	Vote Without Identity in Saudi Arabia	07/04/2019
Satu Juta Orang di GBK	One Million People in GBK	07/04/2019
Hasil Perhitungan Suara Luar Negeri	Results of Overseas Vote Counting	08/04/2019
Sweeping Soloraya	Soloraya Sweeping	08/04/2019
Kapal Patroli Dikejar Malaysia	Patrol Boat is Chased by Malaysia	09/04/2019
Panwaslu Malaysia Berpose Dua Jari	Malaysian Election Overseer Committee Poses Two Fingers	10/04/2019
Konvoi Menuju GBK	Convoy Departs to GBK	11/04/2019
Jokowi Diruqiyah	Jokowi is Ruqyah-ed	12/04/2019
Deindustrialisasi Indonesia	Indonesian De-industrialization	13/04/2019
Indonesia Importir Produk Halal	Indonesia is The Importer of Halal Products	13/04/2019
Indonesia Negara Wisata Halal	Indonesia is Halal-tourism Country	13/04/2019
Kekayaan Indonesia Ke Luar Negeri	Indonesia's Wealth Goes Abroad	13/04/2019
Mekaar Menjangkau Nasabah	Mekaar Reaches Out to Customers	13/04/2019
Pemerintah Menguasai SDA Strategis	Government Controls Strategic Natural Resources	13/04/2019
Rumah UAS Pemberian Prabowo	UAS' House is Prabowo's Gift	13/04/2019
Exit Poll Hasil Pemilu	Exit Poll Election Results	14/04/2019
Kecurangan Pemilu Hong Kong	Hong Kong Election Cheating	14/04/2019
Ketua KPU Umroh dengan Jokowi	KPU Chairperson Does Umrah with Jokowi	14/04/2019
Pendukung 02 Pemilu di Australia	Supporters of 02 Election in Australia	14/04/2019
Polisi Malaysia Tangkap Pencoblos	Malaysian Police Arrest Voters	14/04/2019
Ahok Mencoblos Di Osaka	Ahok Votes in Osaka	15/04/2019
Jokowi Bertemu Habib Rizieq	Jokowi Meets Habib Rizieq	15/04/2019
Abu Bakar Ba'asyir Golput Pemilu	Abu Bakar Ba'asyir Abstains in Election	16/04/2019
Grace Natalie Ajak Makan Babi	Grace Natalie Invites to Eat Pork	16/04/2019
Suara Pemilih Prabowo Hangus	Prabowo Voters' Ballots are Forfeited	16/04/2019
Metro Mengubah Hasil Quick Count	Metro Changes Quick Count Results	17/04/2019

NCID Lembaga Survey Akurat	NCID is Accurate Survey Institute	17/04/2019
Selebaran Palu Arit di Banyumas	Hammer Sickle Flyers in Banyumas	17/04/2019
Surat Suara Tercoblos Bantul	Pierced Ballots in Bantul	17/04/2019
Surat Suara Tercoblos Banyuasin	Pierced Ballots in Banyuasin	17/04/2019
Surat Suara Tercoblos Cipondoh	Pierced Ballots in Cipondoh	17/04/2019
Surat Suara Tercoblos Surabaya	Pierced Ballots in Surabaya	17/04/2019
Megawati Mengakui Kemenangan Prabowo	Megawati Admits Prabowo's Victory	18/04/2019
Surat Suara Tercoblos Riau	Pierced Ballots in Riau	18/04/2019
TV Luar Negeri dan Prabowo Presiden	Foreign Television and Prabowo as President	18/04/2019
Acara PA 212 Dibatalkan TNI	212 Brotherhood Event is Canceled by TNI	19/04/2019
Kotak Suara Dibakar di Jambi	Ballot Boxes Burned in Jambi	19/04/2019
Pesan jusuf Kalla	Jusuf Kalla's Message	19/04/2019
Petugas KPPS Mencoblos Surat Suara	KPPS Officers Pierce Ballots	19/04/2019
Erdogan Tidak Mengakui Kemenangan Jokowi	Erdogan Does Not Acknowledge Jokowi's Victory	20/04/2019
SBY Telpon Moeldoko	SBY Calls Moeldoko	20/04/2019
Syarat Presiden Menang 17 Provinsi	President Requirement Winning in 17 Provinces	20/04/2019
Bentrok Warga Sampang KPU	Sampang Residents Clash with KPU	22/04/2019
Dukun KPU Wonogiri	Shamans of Wonogiri KPU	22/04/2019
Mobil KPU Datangi Ruko	KPU Car Visits Shophouse	22/04/2019
Data TNI Prabowo Menang	TNI Data Shows Prabowo Wins	23/04/2019
Hasil Situng KPU Excel	KPU Excel Count Results	23/04/2019
Pencurian Form C1	C1 Form Theft	23/04/2019
Sekjen PBB Menyelamati Prabowo	UN Secretary General Congratulates Prabowo	25/04/2019
Petugas KPPS Lampung Tewas	Lampung KPPS Officer Dies	28/04/2019
Menpora Imam Nahrawi Mundur	Minister of Youth and Sports Imam Nahrawi Resigns	30/04/2019
Beda Hasil C1 Pleno	Different Results of C1 Plenary	01/05/2019
Rancangan Istana Negara	Design of Istana Negara	01/05/2019
Tri Rismaharini Protes Kpu	Tri Rismaharini Protests Kpu	01/05/2019
Anak MTS Bobol Situs KPU	Madrasa Student Breaches KPU Site	02/05/2019

Jihad Bela Prabowo Masuk Neraka	Defending Prabowo Jihad Enters Hell	02/05/2019
Saksi Prabowo-Sandi Dikeroyok	Prabowo-Sandi Witnesses are Ganged Up	02/05/2019
Prabowo Menguasai Semua Provinsi	Prabowo Dominates All Provinces	03/05/2019
Situng KPU Dikendalikan Bareskrim	Situng KPU Count Site is Controlled by Criminal Investigation Agency	03/05/2019
Kesulitan Pemberangkatan Haji Kemenag	Ministry of Religion's Difficulties in Hajj Departure	04/05/2019
Jokowi Bilang Korban Meninggal Takdir	Jokowi Says Dead Victims is Fate	05/05/2019
Petugas KPPS Meninggal Diracun	KPPS Officer Died by Poison	06/05/2019
Ketua KPUD Bekasi Meninggal	Bekasi KPUD Chairperson Dies	07/05/2019
Pemilu Paling Berdarah	The Bloodiest Election	07/05/2019
Emak-emak Telanjang dan KPU	Naked Mothers and KPU	10/05/2019
Saksi Ahli Bawaslu	Bawaslu's Expert Witnesses	10/05/2019
Guru Ancam Penggal Jokowi	Teacher Threatens to Behead Jokowi	11/05/2019
Mahasiswa Trisakti Bergerak	Trisakti Students Take Action	11/05/2019
HUT PKI dan Pengumuman KPU	PKI Anniversary and KPU Announcement	12/05/2019
Pelaporan Kecurangan Pemilu ke Mahkamah Internasional	Election Fraud Reporting to International Court	13/05/2019
Tsamara Amany Mengubah Data Situng	Tsamara Amany Changes KPU Count Site Data	13/05/2019
Al-Jazeera Menyiarkan Kecurangan Pemilu	Al-Jazeera Broadcasts Election Fraud	14/05/2019
Aksi Baduy Ke Jakarta	Baduys March Action to Jakarta	19/05/2019
Kapal Perang Menjatuhkan Bom	Battleships Drop Bombs	19/05/2019
Mujahidin Maluku dan Baduy Berunjuk Rasa	Maluku Mujahideen and Baduys Demonstrate	19/05/2019
Brimob Disusupi Tentara Cina	Brimob is Infiltrated by Chinese Army	21/05/2019
Otto Hasibuan Kuasa Hukum Prabowo	Otto Hasibuan is Prabowo's Attorney	21/05/2019
SPDP Prabowo Tersangka Makar	SPDP Prabowo as Plot Suspect	21/05/2019
Pengumuman KPU Tidak Sah	Invalid KPU Announcement	22/05/2019
60 Korban Kerusuhan Meninggal	60 Riot Victims Died	23/05/2019

Harun Remaja Syahid	Harun is Syahid Boy	23/05/2019
Jokowi Dijuluki Tangan Berdarah	Jokowi is Nicknamed Bloody Hand	23/05/2019
Margaretha Nainggolan Meninggal	Margaretha Nainggolan Dies	23/05/2019
Prabowo Menang Di Mahkamah Konstitusi	Prabowo Wins at The Constitutional Court	25/05/2019

TABLE A.1: List of topics

Appendix B

Network Visualization

B.1 Mentioned Usernames Network - Communities

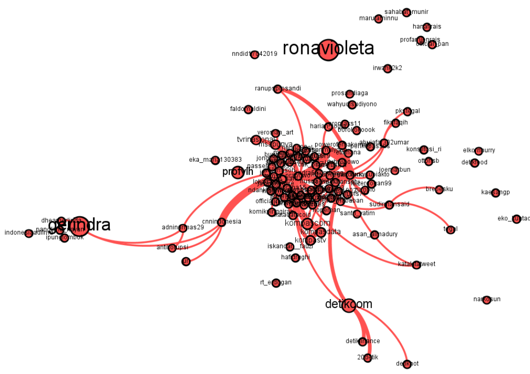

FIGURE B.1: 1st Community of Mention Usernames Network - True News Dissemination

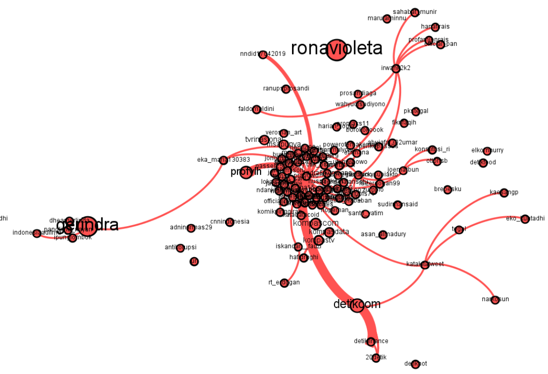

FIGURE B.2: 1st Community of Mention Usernames Network - False News Dissemination

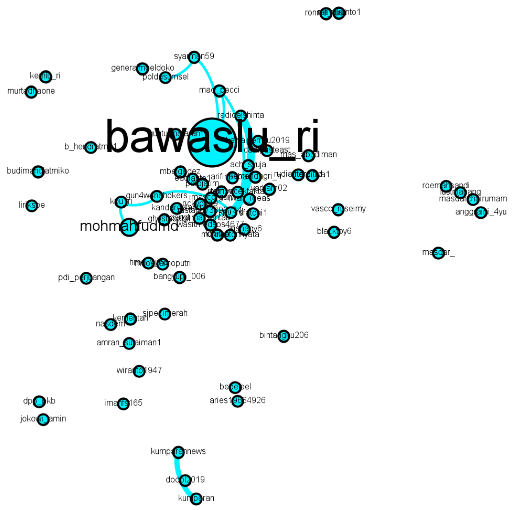

FIGURE B.3: 2[nd] Community of Mention Usernames Network - True News Dissemination

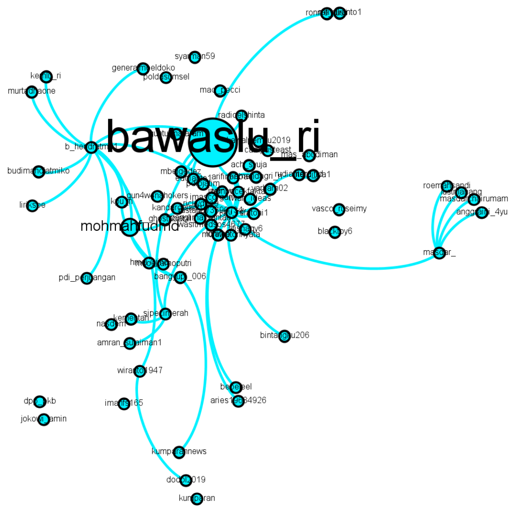

FIGURE B.4: 2[nd] Community of Mention Usernames Network - False News Dissemination

FIGURE B.5: 3[rd] Community of Mention Usernames Network - True News Dissemination

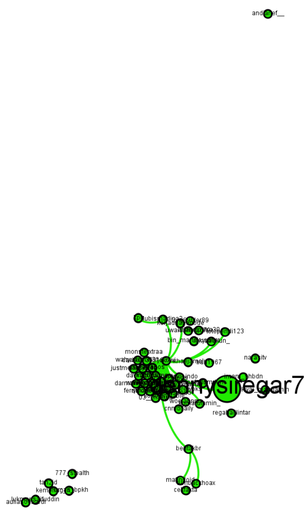

FIGURE B.6: 3[rd] Community of Mention Usernames Network - False News Dissemination

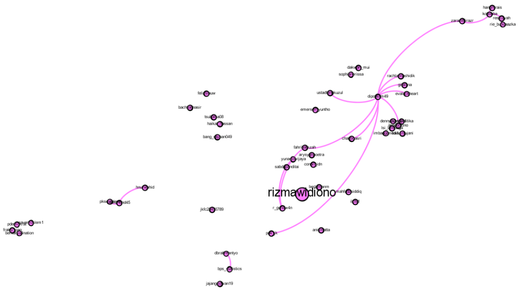

FIGURE B.7: 4^{th} Community of Mention Usernames Network - True News Dissemination

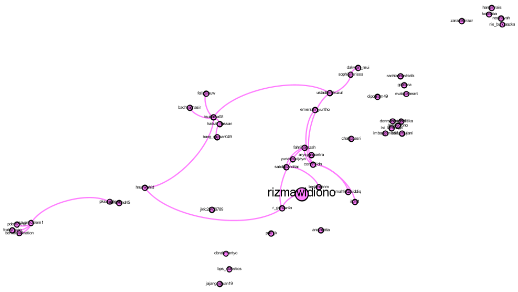

FIGURE B.8: 4^{th} Community of Mention Usernames Network - False News Dissemination

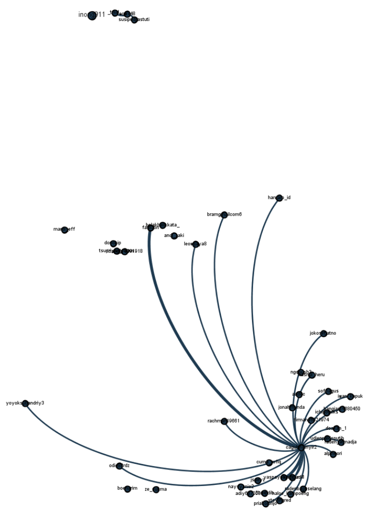

FIGURE B.9: 5^{th} Community of Mention Usernames Network - True News Dissemination

FIGURE B.10: 5^{th} Community of Mention Usernames Network - False News Dissemination

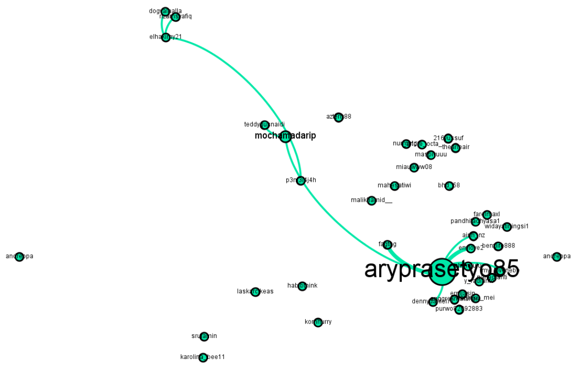

FIGURE B.11: 6th Community of Mention Usernames Network - True News Dissemination

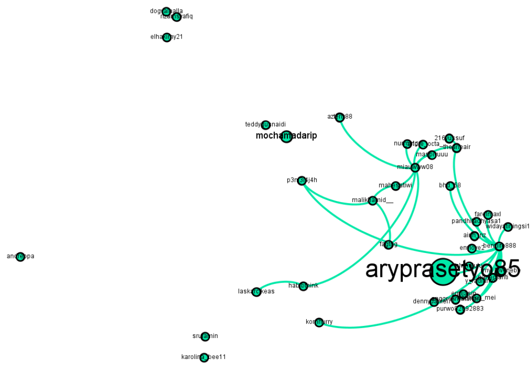

FIGURE B.12: 6th Community of Mention Usernames Network - False News Dissemination

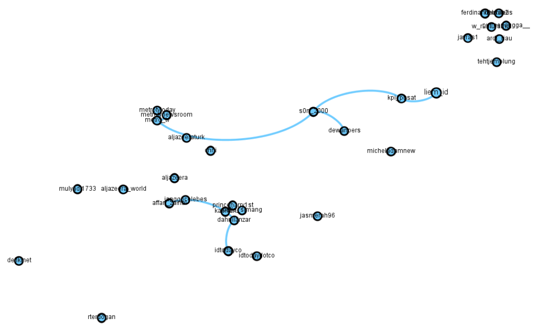

FIGURE B.13: 8th Community of Mention Usernames Network - True News Dissemination

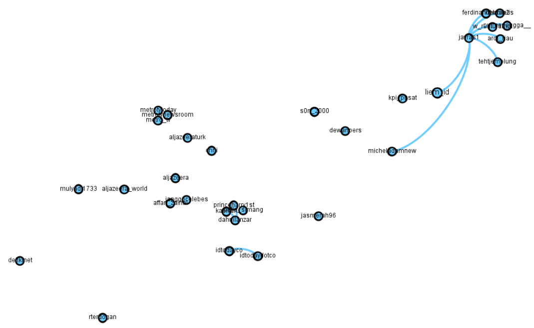

FIGURE B.14: 8th Community of Mention Usernames Network - False News Dissemination

FIGURE B.15: 10th Community of Mention Usernames Network - True News Dissemination

FIGURE B.16: 10th Community of Mention Usernames Network - False News Dissemination

B.2 Mentioned Usernames Network - Influential Users

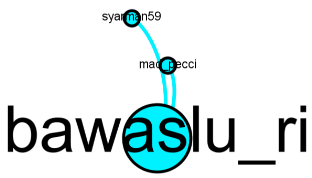

FIGURE B.17: *bawaslu_ri*'s Network - True News Dissemination

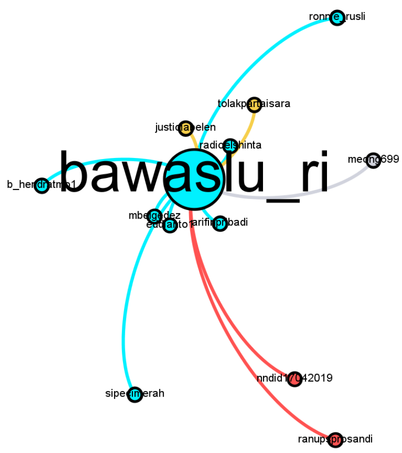

FIGURE B.18: *bawaslu_ri*'s Network - False News Dissemination

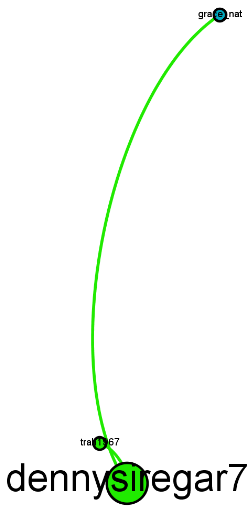

FIGURE B.19: *dennysiregar7*'s Network - True News Dissemination

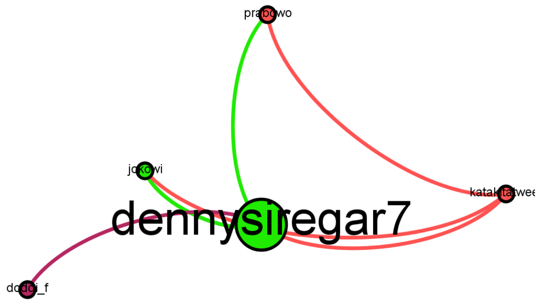

FIGURE B.20: *dennysiregar7*'s Network - False News Dissemination

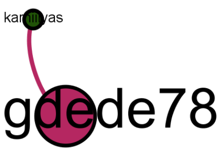

FIGURE B.21: *kangdede78*'s Network - True News Dissemination

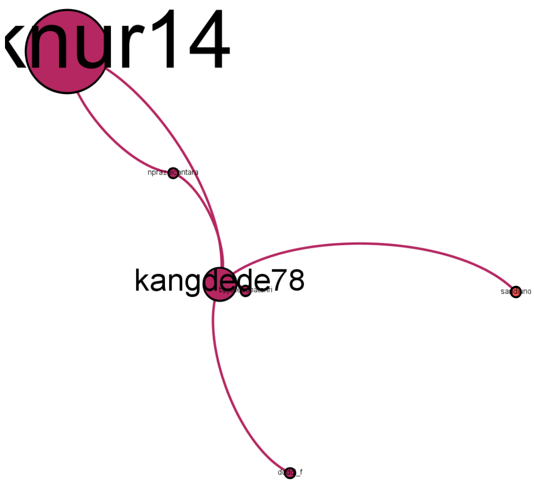

FIGURE B.22: *kangdede78*'s Network - False News Dissemination

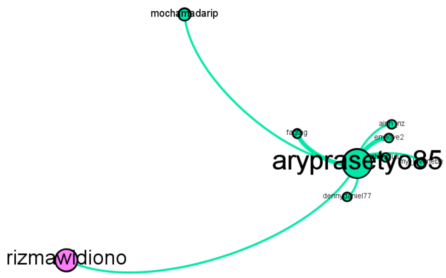

FIGURE B.23: *aryprasetyo85*'s Network - True News Dissemination

FIGURE B.24: *aryprasetyo85*'s Network - False News Dissemination

FIGURE B.25: *gunromli*'s Network - True News Dissemination

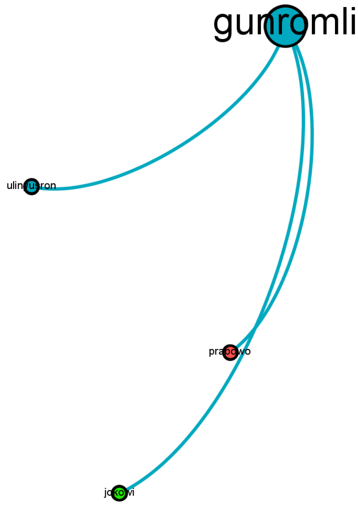

FIGURE B.26: *gunromli*'s Network - False News Dissemination

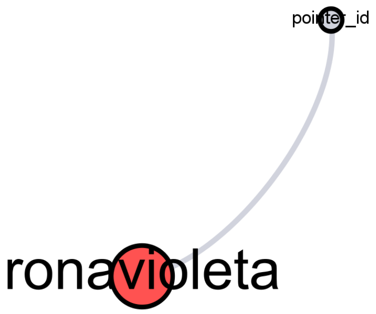

FIGURE B.27: *ronavioleta*'s Network - True News Dissemination

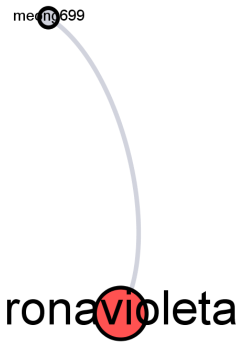

FIGURE B.28: *ronavioleta*'s Network - False News Dissemination

FIGURE B.29: *gerindra*'s Network - True News Dissemination

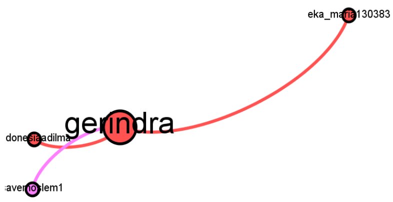

FIGURE B.30: *gerindra*'s Network - False News Dissemination

FIGURE B.31: *rizmawidiono*'s Network - True News Dissemination

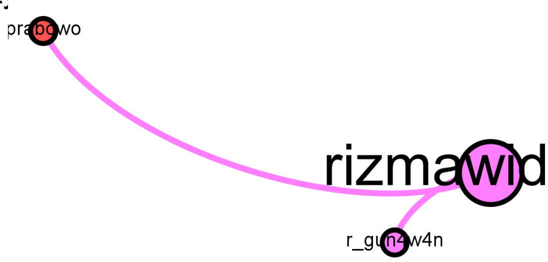

FIGURE B.32: *rizmawidiono*'s Network - False News Dissemination

B.3 Hashtags Co-occurrence Network - Communities

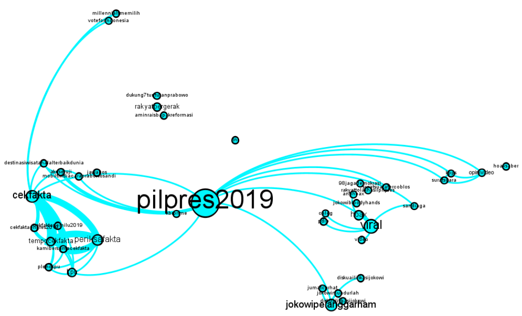

FIGURE B.33: 3rd Community of Hashtags Co-occurrence Network - True News Dissemination

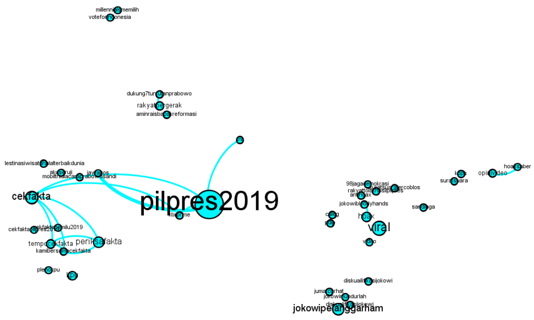

FIGURE B.34: 3rd Community of Hashtags Co-occurrence Network - False News Dissemination

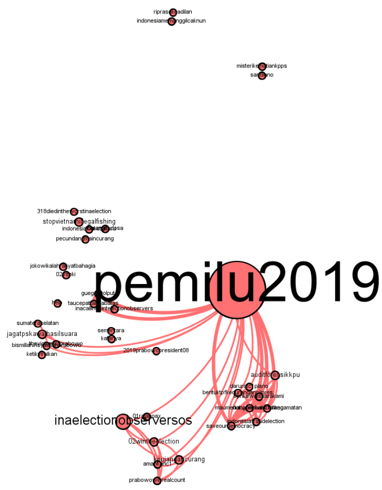

FIGURE B.35: 4th Community of Hashtags Co-occurrence Network - True News Dissemination

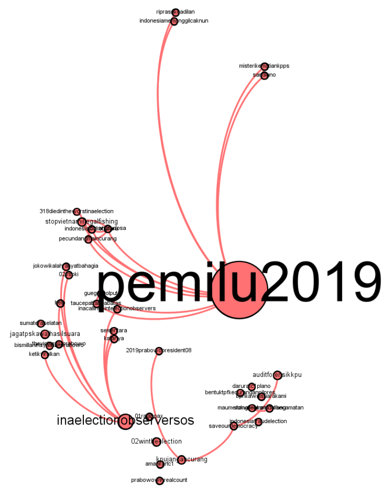

FIGURE B.36: 4th Community of Hashtags Co-occurrence Network - False News Dissemination

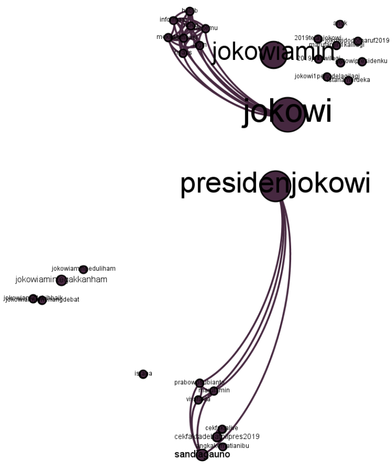

FIGURE B.37: 5th Community of Hashtags Co-occurrence Network - True News Dissemination

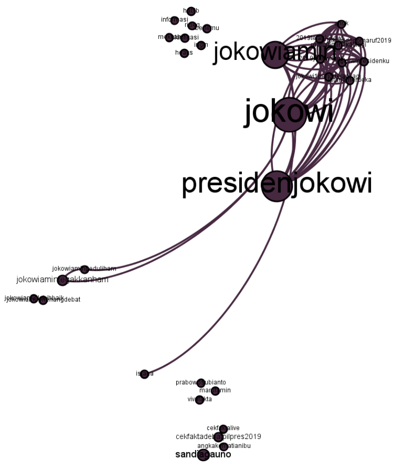

FIGURE B.38: 5th Community of Hashtags Co-occurrence Network - False News Dissemination

FIGURE B.39: 6th Community of Hashtags Co-occurrence Network - True News Dissemination

FIGURE B.40: 6th Community of Hashtags Co-occurrence Network - False News Dissemination

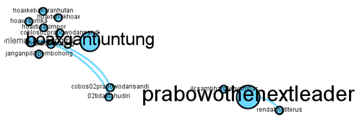

FIGURE B.41: 7th Community of Hashtags Co-occurrence Network - True News Dissemination

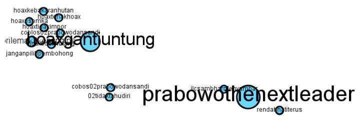

FIGURE B.42: 7th Community of Hashtags Co-occurrence Network - False News Dissemination

FIGURE B.43: 8th Community of Hashtags Co-occurrence Network - True News Dissemination

FIGURE B.44: 8th Community of Hashtags Co-occurrence Network - False News Dissemination

FIGURE B.45: 9th Community of Hashtags Co-occurrence Network - True News Dissemination

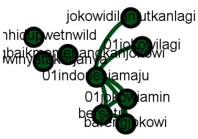

FIGURE B.46: 9th Community of Hashtags Co-occurrence Network - False News Dissemination

FIGURE B.47: 10th Community of Hashtags Co-occurrence Network - True News Dissemination

FIGURE B.48: 10th Community of Hashtags Co-occurrence Network - False News Dissemination

B.4 Hashtags Co-occurrence Network - Influential Hashtags

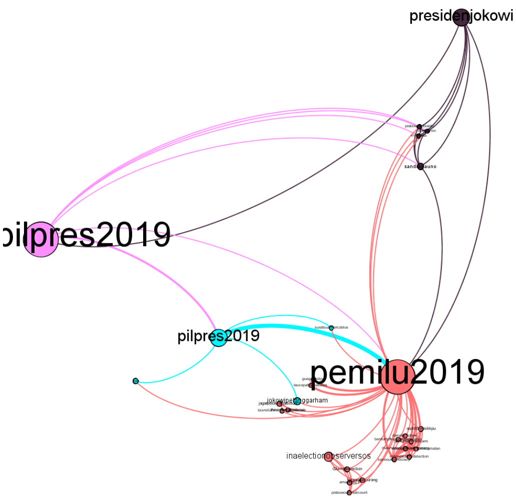

FIGURE B.49: *pemilu2019*'s Network - True News Dissemination

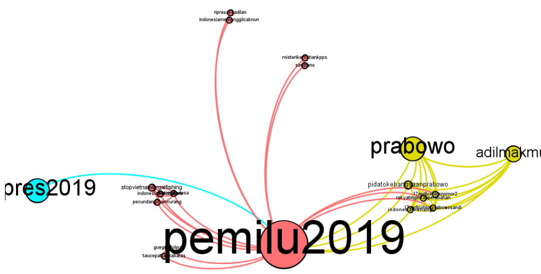

FIGURE B.50: *pemilu2019*'s Network - False News Dissemination

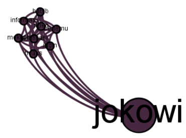

FIGURE B.51: *jokowi*'s Network - True News Dissemination

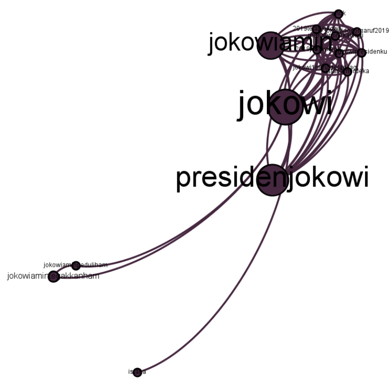

FIGURE B.52: *jokowi*'s Network - False News Dissemination

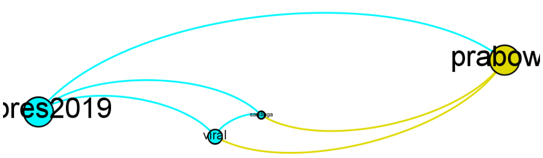

FIGURE B.53: *prabowo*'s Network - True News Dissemination

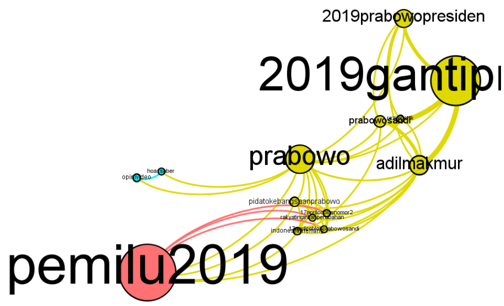

FIGURE B.54: *prabowo*'s Network - False News Dissemination

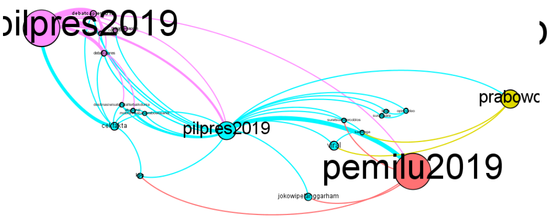

FIGURE B.55: *pilpres2019*'s Network - True News Dissemination

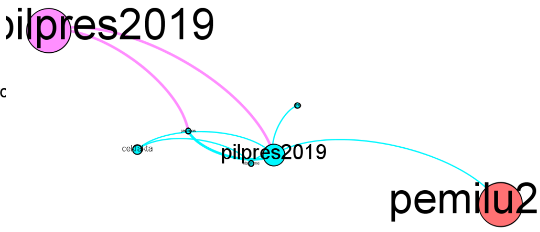

FIGURE B.56: *pilpres2019*'s Network - False News Dissemination

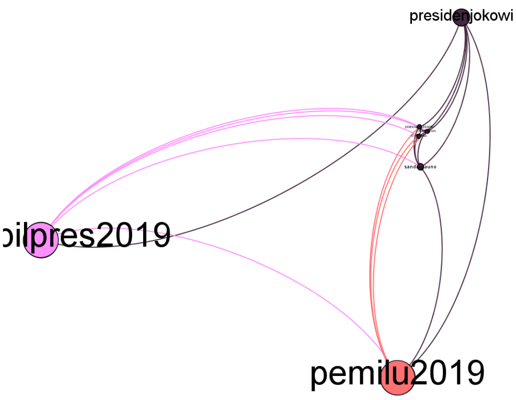

FIGURE B.57: *presidenjokowi*'s Network - True News Dissemination

FIGURE B.58: *presidenjokowi*'s Network - False News Dissemination

FIGURE B.59: *jokowiamin*'s Network - True News Dissemination

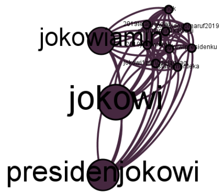

FIGURE B.60: *jokowiamin*'s Network - False News Dissemination

FIGURE B.61: *prabowothenextleader*'s Network - True News Dissemination

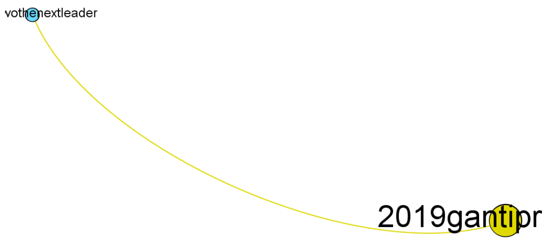

FIGURE B.62: *prabowothenextleader*'s Network - False News Dissemination

FIGURE B.63: *rakyatsudahmuak*'s Network - True News Dissemination

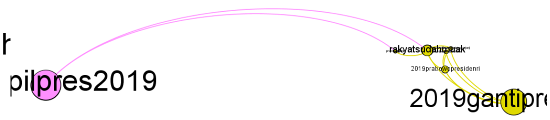

FIGURE B.64: *rakyatsudahmuak*'s Network - False News Dissemination

Bibliography

Abbasi, Alireza and Jorn Altmann (2011). "On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks". In: *44th Hawaii International Conference on System Sciences*, pp. 1–10.

Agarwal, Apoorv et al. (2011). "Sentiment Analysis of Twitter Data". In: *Proceedings of the Workshop on Language in Social Media 2011*, pp. 30–38.

Ajao, Oluwaseun, Deepayan Bhowmik, and Shahrzad Zargari (2018). "Fake News Identification on Twitter with Hybrid CNN and RNN Models". In: *Proceedings of The 9th International Conference on Social Media and Society*, pp. 226–230.

Alfina, Ika et al. (2017). "Hate Speech Detection in The Indonesian Language: A Dataset and Preliminary Study". In: *2017 International Conference on Advanced Computer Science and Information Systems*, pp. 233–238.

Allcott, Hunt and Matthew Gentzkow (2017). "Social Media and Fake News in The 2016 Election". In: *Journal of Economic Perspectives* 31.2, pp. 211–236.

Amin-Naji, Mostafa, Ali Aghagolzadeh, and Mehdi Ezoji (2019). "CNNs Hard Voting for Multi-focus Image Fusion". In: *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–21.

Azali, Kathleen (2017). "Fake News and Increased Persecution in Indonesia". In: *IYI Institute* 61.

Bastian, Mathieu, Sebastian Heymann, and Mathieu Jacomy (2009). "Gephi: An Open Source Software for Exploring and Manipulating Networks". In: *Third International AAAI Conference on Weblogs and Social Media*.

Blondel, Vincent D. et al. (2008). "Fast Unfolding of Communities in Large Networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 10.

Brandes, Ulrik (2001). "A Faster Algorithm for Betweenness Centrality". In: *Journal of Mathematical Sociology* 25.2, pp. 163–177.

Budiharto, Widodo and Meiliana (2018). "Prediction and Analysis of Indonesia Presidential Election from Twitter using Sentiment Analysis". In: *Journal of Big data* 5.1, pp. 51–60.

Carley, Kathleen M. et al. (2016). "Crowd Sourcing Disaster Management: The Complex Nature of Twitter Usage in Padang Indonesia". In: *Safety Science* 90, pp. 48–61.

Crawford, Kate (2009). "Following You: Disciplines of Listening in Social Media". In: *Continuum* 23.4, pp. 525–535.

da Silva, Fernando C.D., Rafael Vieira, and Ana Cristina Garcia (2019). "Can Machines Learn to Detect Fake News? A Survey Focused On Social Medias". In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Fauzi, Ali M. and Anny Yuniarti (2018). "Ensemble Method for Indonesian Twitter Hate Speech Detection". In: *Indonesian Journal of Electrical Engineering and Computer Science* 1.1, pp. 294–299.

Gimpel, Kevin et al. (2011). "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 42–47.

Grossetti, Quentin, Cédric Du Mouza, and Nicolas Travers (2019). "Community-Based Recommendations on Twitter: Avoiding the Filter Bubble". In: *International Conference on Web Information Systems Engineering*, pp. 212–227.

Herman and Donny A. Mononimbar (2017). *Indonesia Fifth-Largest Country in Terms of Twitter Users*. URL: https://jakartaglobe.id/news/indonesia-fifth-largest-country-in-terms-of-twitter-users/. (accessed: 04.02.2020).

Ibrohim, Muhammad O. and Indra Budi (2018). "A Dataset and Preliminaries Study for Abusive Language Detection in Indonesian Social Media". In: *Procedia Computer Science* 135, pp. 222–229.

— (2019). "Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter". In: *Proceedings of The Third Workshop on Abusive Language Online*, pp. 46–57.

Ibrohim, Muhammad O., Muhammad A. Setiadi, and Indra Budi (2019). "Identification of Hate Speech and Abusive Language on indonesian Twitter using The Word2vec, Part of Speech and Emoji Features". In: *Proceedings of the International Conference on Advanced Information Science and System*, pp. 1–5.

Kelly, Ashley R. and Meagan K. Autry (2011). "A Humanistic Approach to The Study of Social Media: Combining Social Network Analysis and Case Study Research". In: *Proceedings of The 29th ACM International Conference on Design of Communication*, pp. 257–260.

Khaldarova, Irina and Mervi Pantti (2016). "Fake News: The Narrative Battle over The Ukrainian Conflict". In: *Journalism Practice* 10.7, pp. 891–901.

Koto, Fajri and Gemala Y. Rahmaningtyas (2017). "Inset Lexicon: Evaluation of A Word List for Indonesian Sentiment Analysis in Microblogs". In: *2017 International Conference on Asian Language Processing*, pp. 391–394.

Kwon, Sejeong et al. (2013). "Prominent Features of Rumor Propagation in Online Social Media". In: *IEEE 13th International Conference on Data Mining*, pp. 1103–1108.

Lamb, Kate (2018a). *'I felt disgusted': Inside Indonesia's Fake Twitter Account Factories*. URL: https://www.theguardian.com/world/2018/jul/23/indonesias-fake-twitter-account-factories-jakarta-politic. (accessed: 04.02.2020).

— (2018b). *Muslim Cyber Army: A 'Fake News' Operation Designed to Derail Indonesia's Leader*. URL: https://www.theguardian.com/world/2018/mar/13/muslim-cyber-army-a-fake-news-operation-designed-to-bring-down-indonesias-leader. (accessed: 04.02.2020).

— (2019). *Fake news spikes in Indonesia ahead of elections*. URL: https://www.theguardian.com/world/2019/mar/20/fake-news-spikes-in-indonesia-ahead-of-elections. (accessed: 17.03.2020).

Lum, Nick (2017). *The Surprising Difference between Filter Bubble and Echo Chamber*. URL: https://medium.com/@nicklum/the-surprising-difference-between-filter-bubble-and-echo-chamber-b909ef2542cc. (accessed: 21.06.2020).

Lunando, Edwin and Ayu Purwarianti (2013). "Indonesian Social Media Sentiment Analysis with Sarcasm Detection". In: *2013 International Conference on Advanced Computer Science and Information Systems*, pp. 195–198.

Mavrck (2014). *'How A Reply, Favorite, Or Retweet Reflects Influence*. URL: https://www.mavrck.co/how-a-reply-favorite-or-retweet-reflects-influence/. (accessed: 25.04.2020).

Naveed, Nasir et al. (2011). "Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter". In: *Proceedings of the 3rd International Web Science Conference*, pp. 1–7.

Newman, Mark E. J. (2004). "Analysis of Weighted Networks". In: *Physical Review E* 70.5.

Newman, Mark E. J. and Michelle Girvan (2004). "Finding and Evaluating Community Structure in Networks". In: *Physical Review E* 69.2.

O'Connor, Brendan et al. (2010). "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series". In: *Fourth International AAAI Conference on Weblogs and Social Media*.

Olanrewaju, Abdus-Samad T. and Ahmad Rahayu (2020). "Examining The Information Dissemination Process on Social Media during The Malaysia 2014 Floods Using Social Network Analysis". In: *Journal of Information and Communication Technology* 17.1, pp. 141–166.

Procter, Rob et al. (2013). "Reading The Riots: What were The Police Doing on Twitter?" In: *Policing and Society* 23.4, pp. 413–436.

Raghavan, Usha N., Réka Albert, and Soundar Kumara (2007). "Near Linear Time Algorithm to Detect Community Structures in Large-scale Networks". In: *Physical Review E* 76.3.

Ruan, Yiye et al. (2012). "Prediction of Topic Volume on Twitter". In: *Proceedings of the 4th International ACM Conference on Web Science*, pp. 397–402.

Salsabila, Nikmatun A. et al. (2018). "Colloquial Indonesian Lexicon". In: *2018 International Conference on Asian Language Processing*, pp. 226–229.

Sang, Erik T.K. and Johan Bos (2012). "Predicting The 2011 Dutch Senate Election Results with Twitter". In: *Proceedings of The Workshop on Semantic Analysis in Social Media*, pp. 53–60.

Saputri, Mei S., Rahmad Mahendra, and Mirna Adriani (2018). "Emotion classification on indonesian Twitter Dataset". In: *2018 International Conference on Asian Language Processing*, pp. 90–95.

Setijadi, Charlotte (2017). "Ahok's Downfall and The Rise of Islamist Populism in Indonesia". In: *ISEAS Perspective* 38.8.

Silverman, Craig and Jeremy Singer-Vine (2016). *The True Story Behind The Biggest Fake News Hit Of The Election*. URL: https://www.buzzfeednews.com/article/craigsilverman/the-strangest-fake-news-empire. (accessed: 17.03.2020).

Song, Min, Meen C. Kim, and Yoo K. Jeong (2014). "Analyzing The Political Landscape of 2012 Korean Presidential Election in Twitter". In: *IEEE Intelligent Systems* 29.2, pp. 18–26.

Stieglitz, Stefan and Linh Dang-Xuan (2013). "Social Media and Political Communication: A Social Media Analytics Framework". In: *Social Network Analysis and Mining* 3.4, pp. 1277–1291.

Suraya, Suraya and Fransiskus E. D. Kadju (2019). "Jokowi Versus Prabowo Presidential Race for 2019 General Election on Twitter". In: *Saudi Journal of Humanities and Social Sciences* 4.3, pp. 198–212.

Tala, Fadillah Z. (2003). "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia". In: *Institute for Logic, Language and Computation Universiteit van Amsterdam*.

Tandoc Jr., Edson C., Zheng W. Lim, and Richard Ling (2018). "Defining "Fake News" A Typology of Scholarly Definitions". In: *Digital Journalism* 6.2, pp. 137–153.

Tempo (2018). *Metodologi*. URL: https://cekfakta.tempo.co/metodologi. (accessed: 27.05.2020).

The, Johanes E., Alfan F. Wicaksono, and Mirna Adriani (2015). "A Two-stage Emotion Detection on Indonesian Tweets". In: *2015 International Conference on Advanced Computer Science and Information Systems*, pp. 143–146.

Thelwall, Mike, Kevan Buckley, and Georgios Paltoglou (2012). "Sentiment Strength Detection for the Social Web". In: *Journal of the American Society for Information Science and Technology* 63.1, pp. 163–173.

Vania, Clara, Mohammad Ibrahim, and Mirna Adriani (2014). "Sentiment Lexicon Generation for an Under-Resourced Language". In: *Int. J. Comput. Linguistics Appl.* 5.1, pp. 59–72.

Volkova, Svitlana and Jin Y. Jang (2018). "Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media". In: *Companion Proceedings of The Web Conference 2018* 29.2, pp. 575–583.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018). "The Spread of True and False News Online". In: *Science* 359.6380, pp. 1146–1151.

Wang, Yuan et al. (2016). "Using Hashtag Graph-based Topic Model to Connect Semantically-related Words without Co-occurrence in Microblogs". In: *IEEE Transactions on Knowledge and Data Engineering* 28.7, pp. 1919–1933.

Wijnhoven, Rob G. J. and Peter H. N. de With (2010). "Fast Training of Object Detection using Stochastic Gradient Descent". In: *20th International Conference on Pattern Recognition*, pp. 424–427.

Zeng, Daniel et al. (2010). "Social Media Analytics and Intelligence". In: *IEEE Intelligent Systems* 25.6, pp. 13–16.

Zhai, Chengxiang and Sean Massung (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. Association for Computing Machinery and Morgan Claypool. ISBN: 9781970001198.

Zhao, Yanchang (2013). "Analysing Twitter Data with Text Mining and Social Network Analysis". In: *Proceedings of The 11th Australasian Data Mining and Analytics Conference*, pp. 23–29.

Zubiaga, Arkaitz, Heng Ji, and Kevin Knight (2013). "Curating and Contextualizing Twitter Stories to Assist with Social Newsgathering". In: *Proceedings of the 2013 international Conference on Intelligent User Interfaces*, pp. 213–224.

Zubiaga, Arkaitz, Maria Liakata, and Rob Procter (2016). "Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media". In: *arXiv: 1610.07363* 1.