International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 11, Issue 1, 2020

# Deciphering Social Opinion Polarization towards Political Events Based on Content and Structural Analysis

**Andry Alamsyah[a*], Wachda Yuniar Rochmah[b], Arina Nahya Nurnafia[c],** [a,b,c]School of Economics and Business, Telkom University, Bandung, Indonesia, Email: [a*]andrya@telkomuniversity.ac.id

There is evidence to suggest that social opinion polarization leads to the breakup of relationships, sometimes on a scale of small communities, but it can occasionally divide large organizations or even a nation. The methodology often used to investigate the root cause of opinion polarization in society is random sampling or a questionnaire-based approach, which are considered expensive and time-consuming. On the contrary, a large-scale approach using social media can provide a rich source of data for the investigation of several questions, such as: how does social opinion polarization form, what changes occur in social network mechanisms, and who are the dominant actors and communities? The power of a big data approach lies in the volume of data analysed; the more data involved in the process, the more accurately it describes the population. Today, computing power is no longer an obstacle to the large-scale processing of data, and thus, the observation of social opinion polarization processes becomes possible. This research investigates the root cause of social opinion polarization by using a topic modelling methodology. The dynamic social network mechanism is measured using social network properties. Identification of influential actors and communities are provided by social network analysis metrics and methodologies. By answering the three major questions above, this study examines opinion polarization and its growth over time, both for their topology structure and conversational content. This knowledge gives insight into how and when opinion separation takes place. As a case study, two massive adverse political campaign in Indonesia are used: the pro and contra opinions on whether the incumbent president should continue his presidentship after the 2019 presidential election. Indonesians' habit of expressing their voice on social media by producing user-generated content is an advantage that strengthens this study.

**Key words:** *Opinion Polarization, Social Network Analysis, Topic Modelling.*

## Introduction

Technology has permitted us to communicate efficiently, and it has thus provided medium for societal to growth. Human interaction brings knowledge, influences opinion, disseminates information, and groups people with similar interests. This online activity leaves digital traces behind, which allow us to capture human behaviour. Data analytics provide a set of tools and a methodology that allow us to extract knowledge regarding a specific social problem. Large-scale data generated from social interactions has led to the formation of a new social science, called network science.

Social media contents, whether in the form of images, videos, testimonials, tweets, blog posts, reviews, and otherwise, are referred to as User Generated Content (UGC). Many people participate in UGC creation as part of their effort to build social reputation, or just find information that is of interest to them. UGC is a substantial resource, as it reveals many precious insights that social scientists seek. Data analytics methods can be employed to identify patterns in UGC data, which is mainly in the form of unstructured data. The nature of unstructured data makes it complicated to process, and in many cases, the data cannot be processed at all using legacy procedures such as data mining (Alamsyah et al, 2017). One of the fastest strategies to process unstructured data from social media is by using a Social Network Analysis (SNA) methodology. SNA models actors and their surrounding relationship with their neighbours' actors. To capture the growth and shrinkage of relationships between actors, we use Dynamic Network Analysis (DNA) (Leskovic et al, 2007).

There are many network measurements, called network properties, that can quantify a social network. We are able to track the evolution of a network based on these metrics: nodes, which represent actors; edges, which represent relationships among actors; network average degree, which is the number of relationships each actor has divided by the total number of relationships that occur in the overall network; network diameter, which shows the maximum distance between actors at the ends of network; modularity, which measures the tendency of a network to cluster; and network density, which measures the ratio between current edges in the network and maximum number of possible edges (Newman, 2011).

Some of the most dynamic social conversations are found in the political domain. Close to an election event, social actors will tend to support a political figure according to their preferences. Some consider a candidate's vision and mission, track record, achievements, or

even political party background. Eventually, these preferences will lead to opinion polarization in our society. The development of technology has greatly facilitated the expression of public opinion through social media. This phenomenon gives us the advantage of being able to see the tendencies that occur in the community, either in favour of or against a political figure.

Jokowi is the 7th president of Indonesia who will join the presidency election in 2019. A large number of people are supporting his run for president for the second time. Nonetheless, there are people who do not. The people who oppose Jokowi express their disapproval through a campaign called *Ganti Presiden*. To know public opinion on this matter, the authors extracted some information from *Twitter*, a popular social media platform in Indonesia. The datasets contain "tweets", sorted according to their different hashtags. The authors separate the sentiments through a suitable hashtag, which may be pro or contra to Jokowi. In order to understand what is being discussed by certain communities on Twitter, the authors utilize a method called Topic Modelling, which enables us to examine the topics being discussed through the most probable terms within those topics. Finally, the authors use SNA and DNA methodology to measure network metrics created by pro and contra hashtags.

User interaction in social media forms a social network in a manner similar to real-world networks, and such a network can be represented by a graph with set of nodes to represent actors and edges to represent an actor's relationship (Newman, 2011). SNA is a way to quantify various interaction patterns in a social network. It is able to enhance a researcher's understanding of the actors' characteristics, the most influential actors, the network community, and several other elements (Hanneman and Riddle, 2005).

DNA is an emergent scientific field that combines traditional SNA with network science and network theory. There are two aspects to this field. The first one is the statistical analysis of DNA data. The second is the utilization of simulation to address issues of network dynamics. The differences between DNA networks and traditional social networks are that DNA's are larger scale, more dynamic, and more complex networks that may contain varying levels of uncertainty (White, 1992). DNA also takes the interactions, social features, conditioning structure, and behaviour of networks into account. The evolution of dynamics in actor interactions gives valuable insights into actors' online social behaviour. This DNA permits us to understand how relationships thrive from time to time, how relationships are created among actors, and how information diffuses (Alamsyah et al, 2018).

In this paper, the authors investigated the social opinion polarization mechanism using Topic Modelling (TM). TM helps identify dominant terms on the opposing sides of an opinion. It is followed by Text Network Analysis (TNA) methodology that summarizes the conversations on each side. And lastly, the DNA mechanism is measured using social network properties.

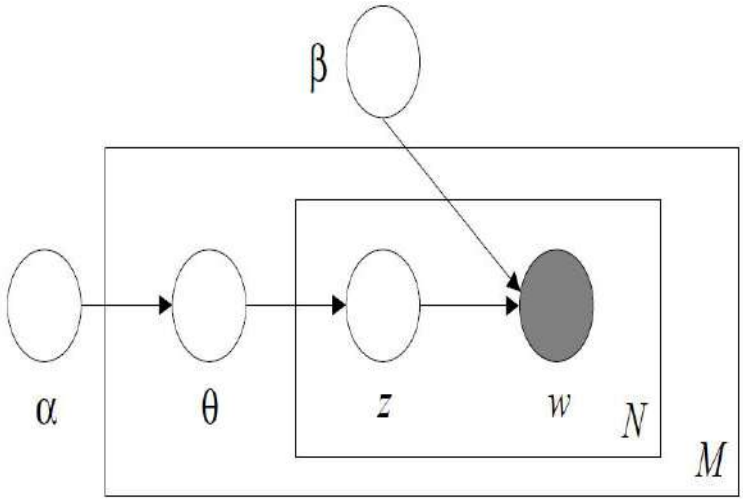
Identification of influential actors and communities are provided by SNA metric and methodology. By answering three major questions: how social opinion polarizations are formed, dynamic social network mechanism during time-windows observation, and who are the dominant actors and communities, we are able to explain opinion polarization and its growth over time, both for their topology structure and conversational content. The process knowledge also gives us an insight into how and when the separation of opinion takes place.

## Literature Reviews

### *A. Topic Modelling (TM)*

Topic Modelling (TM) is an approach to infer the topic in a document. The algorithm uses the Latent Dirichlet Allocation (LDA), which is a generative probabilistic model of a corpus. Documents are represented as random mixtures over latent topics, where each topic is indicated by a distribution over words (Blei at al, 2003).

**Figure 1.** The Latent Dirichlet Allocation Model Illustration




The LDA model is portrayed as a probabilistic graphical model in Figure 1. The boxes are plates that represents replicates. The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document. The parameters α and β are parameters in the corpus-level, assumed to be sampled once in the corpus generating process. The variables $\theta_d$ are variables in the document-level, sampled once per document. Lastly, the variables $Z_{dn}$ and $W_{dn}$ are word-level variables and are sampled once for each word in each document (Blei at al, 2003).

### *B. Social Network Analysis (SNA)*

Social Network Analysis (SNA) is an exploration and describing patterns approach for understanding how social relationships are formed by an individual or group. The interactions

that occur in social networks can be represented by two elements: those that represents actors or individuals, called nodes; and those that represent relationships or interactions among actors or groups within the network, called edges. SNA has become an important mode of research that focuses on many areas, such as management, sociology, health care, and many more. When people interact with each other online, they barely rely on paper-based questionnaires to build a network, mainly due to the fact that this limits the acquisition of data. SNA covers four main concepts: actors' interactions on social networks; strength of relationships between actors; identification of key or important actors in the networks; and the measurement of network structure and cohesion (Leskovic at al, 2007).

### *C. Dynamic Network Analysis (DNA)*

Dynamic Network Analysis (DNA) consists of analytic and algorithmic models that identify the overall process of social network evolution to forecast individual as well as group behaviour and their relationships to each other. DNA is the latest approach to understanding interactions in a network. DNA captures the dynamics of network structures through complex systems as sequences of time within interactions (Alamsyah et al, 2018) (del Val et al, 2015).

Several examples of dynamic network properties are as follows. The edge evolution shows changes in relationships that occur within a given period of time. Node evolution shows changes in the number of nodes or actors that exist from time to time. Some other dynamic measurements are network diameter evolution, average clustering coefficient evolution, modularity evolution, and network density evolution.

### *D. Text Network Analysis*

Text Network Analysis (TNA) is one of the possible ways to represent text occurrence complexity. The nodes are the texts and the relationship of the texts are represented by the edge (Hunter, 2014). One way to summarize a document is by representing it as a network of document text. This network contains the nodes, or the texts, and the edges, or the relationships, of the texts, which represent text occurrences on the same phrase. Higher frequencies of texts occurrence next to each other in a phrase results in a stronger connection between them. The meanings and agendas that can arise from that interconnectedness are diverse. However, each expression of the text has a certain purpose, related to a certain moment in time. Having the text as a network allows for a much more holistic views of the text and for many other expressions of the same agenda that could be more related to a specific context. (Paranyuskhin, 2011).

## Methodology

The methodology starts with the data collection process, followed by data pre-processing, the main process, and, lastly, summarization of overall process. The research framework is shown in Figure 2. The details of each process are explained in sub-chapters A, B, C, and D.

### *A. Data Collection Process*

The data collected from *Twitter* consisted of data stream over a period of ten days, from April 27th until May 2nd, 2018. The author filtered or classified the tweets by pro and contra opinion using the selection hashtags shown in Table. 1. The *Twitter* Application Programming Interface (API) was used as the gate to access public *Twitter* data. The number of tweets collected from each side is also shown in Table. 1.

**Figure 2.** The Research Workflow

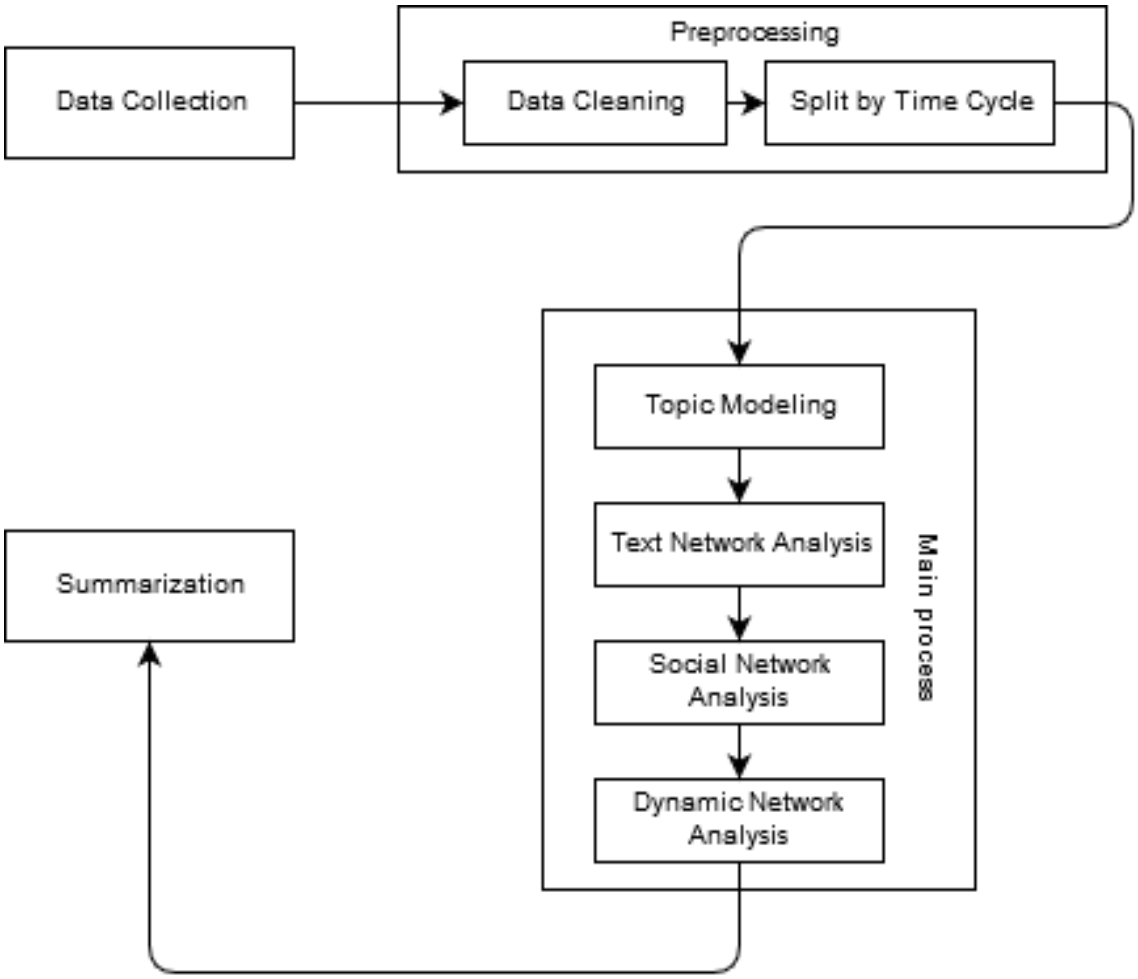


**Table 1:** The Hashtag List

| | **Pro** | **Contra** |
|---|---|---|
| **Hashtag** | *jokowi2periode, JKW2P, jokowipresiden2019, 2019tetapjokowi, jokowisekalilagi, diasibukkerja, rakyatmaujokowi2019, jokowiduaperiode, salam2jari, ogah2019gantipresiden,* | *2019gantipresiden, 2019presidenbaru, gantipresidenyuk, gantipresiden, 2019asalbukanjokowi, gantipresiden2019, asalbukanjkw, 2019gantirezim, 2019wajibgantipresiden,* |
| **Number of Tweets** | 24097 | 418256 |

***B. Data Pre-Processing***

**1) Data Cleaning**

Data cleaning is the first sub process of data pre-processing, which is a crucial stage in reducing the complexity of the quantification process, which leads to better input for the next process. Basically, in this stage, for TM and TNA, it entails the removal of unnecessary words, as well as processes such as stopwords filtering, tokenization to separate phrases into words, stemming to transform the word into its original form, and more. For SNA and DNA, datasets are cleaned of bots, users who randomly send spam to sell things online, and all tweets without any interaction. This filters the source data and automatically targets actor interaction. Table 2 gives examples of the cleaning process for the TM and TNA.

**Table 2:** The Example of Data Cleaning

| **Before** | **After** |
|---|---|
| *Mengintimidasi* | *intimidasi* |
| *Memilih* | *pilih* |
| *Kaus* | *kaos* |

**2) Split by Time Cycle**

To see the evolution of network properties within the overall network from both sides, datasets are split day by day. From the splitting process, we are able to see the growth or shrinkage of the topology within the network over the ten days of observation time. This pre-processing step is done only for the DNA process.

### *C. Main Process*

The main process consists of four methodologies, executed in the following order: TM, TNA, SNA, and DNA. Each process has specific functions to support the final summarization of research conclusions. TM is used to detect the topics on each pro and contra hashtag tweet. SNA produces network models to detect the properties of each pro and contra network. DNA is used to track network evolution during observation time. Finally, TNA fulfils basically the same function as SNA, but texts is used to replace the actors. The explanation of each part of the main process is as follows:

#### 1) Topic Modelling (TM)

LDA methodology to detect topics is implemented on each opponent's side. LDA generate topics based on word frequency from each side's tweet data. Here, that refers to the 24,097 pro tweets and the 418,256 contra tweets. LDA is a reasonably fast and accurate way to find mixed topics in the collection of documents. From each opponent's side, the authors map the topics generated and see the terms arranging each topic based on the frequency of appearance. Hence, it may be seen what kind of topic emerges from those collections of terms.

#### 2) Social Network Analysis (SNA)

The social network construction is based on actors' interaction tweets. Once the tweet data is collected, it is transformed into raw tweets data, then into node source, and then a node target list, represented as edge list. The procedure only collects usernames who tweet and usernames who are mentioned in particular tweets. These mentions could be in the form of conversations or retweets. If someone tweets without generating conversation, then that tweet can be omitted. Figure 3a shows the raw tweet format and 3b shows the list of node sources and targets. The edge list in Figure 3b is the main ingredient in the construction of social networks related to politics and the election event.

**Figure 3.** (a) Raw Tweet Format. (b) Node Source and Target List

(a)

| created_at | screen_name | text | source | display_text_width | reply_to_status_id | reply_to_user_id | reply_to_screen_name | is_quote | is_retweet |
|---|---|---|---|---|---|---|---|---|---|
| 16/03/2019 12:29 | AgnesAlexandri1 | @jokowi #TanganJoko | Krile2 | 99 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:29 | AgnesAlexandri1 | @jokowi Solidaritas U | Krile2 | 201 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 08:25 | AgnesAlexandri1 | @jokowi Lembaga sur | Krile2 | 227 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 09:27 | AgnesAlexandri1 | @jokowi HEHEHEHE. I | Krile2 | 216 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 12:24 | asiapasific15 | @jokowi teri medan d | Twitter for And | 154 | x1106857676607127552 | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 12:05 | ArtaKahisha | @jokowi Bersatu Duk | Krile2 | 217 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:41 | borshiregar1805 | @Elina_Vay Tau gak m | Twitter for And | 58 | x1106378625811509250 | x3253266487 | Elina_Vay | FALSE | FALSE |
| 16/03/2019 11:37 | BolaBolin | @deninovi @dipoala | Twitter for And | 104 | x1106876532679598082 | x203821863 | deninovi | FALSE | FALSE |
| 16/03/2019 11:33 | IndonesiaR4Y4 | @_kiranalara @edhi_ | Twitter for And | 94 | x1106777887951118337 | x1173169994 | _kiranalara | FALSE | FALSE |
| 16/03/2019 08:23 | IndonesiaR4Y4 | @MurtadhaOne #Pres | Twitter for And | 70 | x1106802475208593408 | x363128593 | MurtadhaOne | FALSE | FALSE |
| 16/03/2019 11:33 | Elina_Simamora | @Bi4nkaR4ra pak Joko | Twitter for And | 106 | x1106862520596422656 | x986184193 | Bi4nkaR4ra | FALSE | FALSE |
| 16/03/2019 10:41 | AghnaValerie15 | @jokowi Jokowi dijad | Mixero | 214 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 08:08 | AghnaValerie15 | @jokowi 17 Aprel Mai | Mixero | 217 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:30 | AghnaValerie15 | @jokowi PENA 98 Sur | Mixero | 87 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:29 | Jhoni_Santuy | @Bi4nkaR4ra Untuk k | Twitter for And | 228 | x1106862520596422656 | x986184193 | Bi4nkaR4ra | FALSE | FALSE |
| 16/03/2019 11:26 | Kaston_Sinaga | @jokowi @pekaklont | Twitter Web Cl | 79 | x1106857676607127552 | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:12 | FaniFau17663257 | @jokowi #KataEmakG | Krile2 | 255 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:11 | FaujiahSafira | @jokowi #KataEmakG | Krile2 | 255 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 11:11 | AndikaPuri5 | @jokowi #KataEmakG | Echofon | 255 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 10:49 | Songgek1 | @woelannnn @jokow | Twitter Web Ap | 124 | x1106818743575822337 | x1930932188 | woelannnn | FALSE | FALSE |
| 16/03/2019 10:38 | daradiputeri | @jokowi Tim Kampan | Manja Menang | 264 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 10:22 | nadi_bagus | @slankdotcom @hex | Tween | 277 | | x17166826 | slankdotcom | FALSE | FALSE |
| 16/03/2019 06:46 | mrs_herlambang | @addiems Selamat da | Twitter for iPh | 198 | x1106759717571878914 | x94805910 | addiems | FALSE | FALSE |
| 16/03/2019 10:02 | mrs_herlambang | @detikcom Biarpun n | Twitter for iPh | 79 | x1106855663358144512 | x69183155 | detikcom | FALSE | FALSE |
| 16/03/2019 10:02 | OKOK20357253 | @BiLLRaY_gak ngeri l | Twitter for iPh | 65 | x1106735503963914240 | x902798881192263681 | BiLLRaY_ | FALSE | FALSE |
| 16/03/2019 09:43 | sutovick | @Elina_Vay <U+0001F | Twitter Web Cl | 240 | x1106765023265710081 | x3253266487 | Elina_Vay | FALSE | FALSE |
| 16/03/2019 09:40 | UraniaSyua | @jokowi #KataEmakG | Silver Bird | 255 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 09:39 | IlhamRe21452541 | @jokowi #KataEmakG | Twitter for iPad | 255 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 09:39 | balqisfayruz03 | @jokowi Mantap sem | TweetCaster | 171 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 09:35 | ArmiJulia | @jokowi insfratruktu | Twitter for iPad | 229 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 09:20 | irmaang12670197 | @jokowi Anda main T | Silver Bird | 252 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 09:19 | carina_niken | @Namaku_Mei #Joko | Twitter for And | 15 | x1106823384455798784 | x3259106634 | Namaku_Mei | FALSE | FALSE |
| 16/03/2019 09:12 | SyamsulAlifah | @jokowi menang dua | Twitter for iPad | 191 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 08:43 | NelaAfra | @jokowi PENA 98 Sur | Krile2 | 87 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 08:38 | CahyaAdiSetiya1 | @jokowi Survei SMR | Twitter for iPad | 154 | | x366987179 | jokowi | FALSE | FALSE |
| 16/03/2019 08:27 | fatahismail82 | @jokowi @hexagrap | Twitter for iPad | 255 | | x366987179 | jokowi | FALSE | FALSE |

(b)

| Source | Target |
|---|---|
| AgnesAlexandri1 | jokowi |
| AgnesAlexandri1 | jokowi |
| AgnesAlexandri1 | jokowi |
| AgnesAlexandri1 | jokowi |
| asiapasific15 | jokowi |
| ArtaKahisha | jokowi |
| borshiregar1805 | Elina_Vay |
| BolaBolin | deninovi |
| IndonesiaR4Y4 | _kiranalara |
| IndonesiaR4Y4 | MurtadhaOne |
| Elina_Simamora | Bi4nkaR4ra |
| AghnaValerie15 | jokowi |
| AghnaValerie15 | jokowi |
| AghnaValerie15 | jokowi |
| Jhoni_Santuy | Bi4nkaR4ra |
| Kaston_Sinaga | jokowi |
| FaniFau17663257 | jokowi |
| FaujiahSafira | jokowi |
| AndikaPuri5 | jokowi |
| Songgek1 | woelannnn |
| daradiputeri | jokowi |
| nadi_bagus | slankdotcom |
| mrs_herlambang | addiems |

**3) Dynamic Network Analysis (DNA)**

DNA is used to identify the dynamic interactions between actors in the tweet universe. DNA captures each SNA measurement on different observation time. In this research, the measurement is based on daily observation. As a result, graph or network evolutions are produced over time. The chart's x-axis represents the period of time and the y-axis represents the network properties value. This visualization enables us to see network and actor's behaviour over time.

**4) Text Network Analysis (TNA)**

TNA basically work like an SNA model, where nodes represent texts or terms instead of actors. In TNA, edges represent co-occurrent texts in the same phrase or document. This method helps us to summarize a large-scale document or tweet into a condensed piece of network information. Compared to the common method of extracting information from documents, such as from a word cloud, TNA are easier to interpret, thanks to the presence of links between texts, which show how relatable those texts are to the document or tweet. TNA

is applied to both the pro and contra side by the following work order: 1. finding the co-occurrence between texts in the same tweet; 2. measuring each text's frequency of appearance; 3. measuring the intensity of relationships; 4. detecting network modularity to identify the grouping of texts between different topics; 5. constructing the network, where node size represents frequency of appearance, edge size represents intensity of relations, and node-edge colour represents the text group modularity.

## Results and Analysis

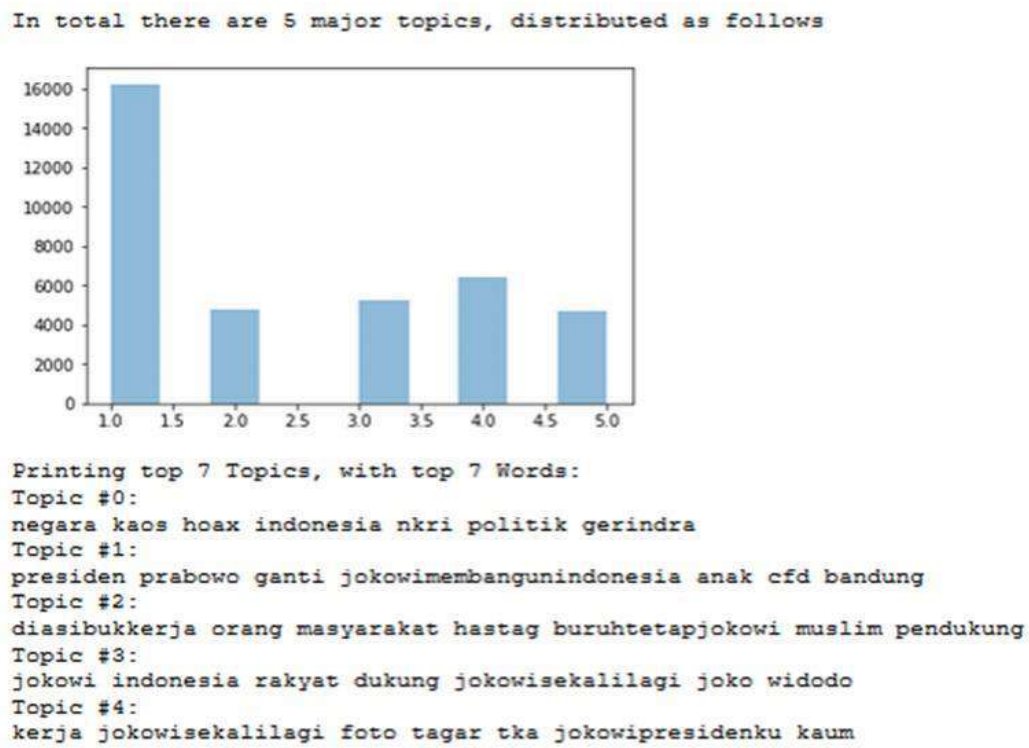
### *A. Topic Modelling (Tm)*

#### *1) Pro Hashtags*

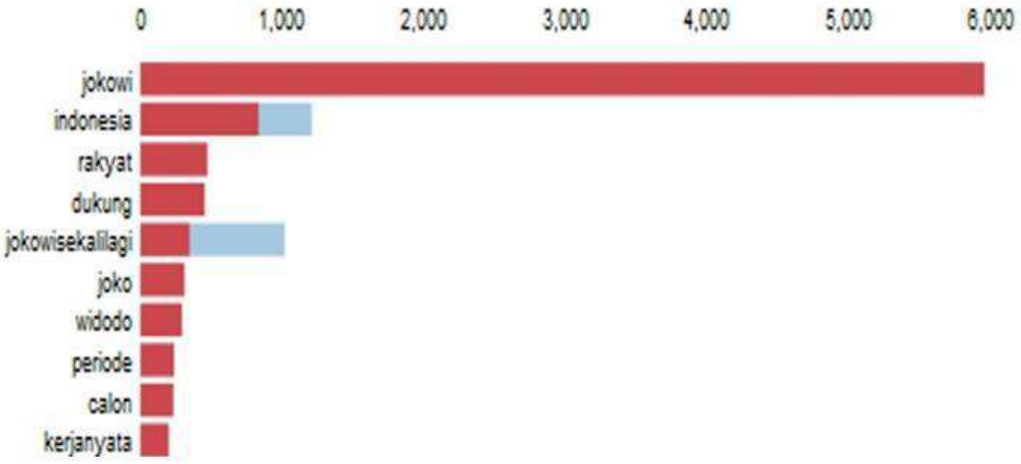
The authors obtained the top 5 biggest topics for each pro and contra hashtag network. Figure. 4a shows the distribution of the top 5 topics in the pro hashtag network. Each topic consists of several domain texts or terms. Figures 4b, 4c, 4d, and 4e show the distribution of terms in each topic, ordered from the highest to the lowest. The blue coloured bar represents the frequency of overall terms in the topics; meanwhile, the red coloured bar represents the frequency of terms in the current topic.

**Figure 4.** (a). The Distributions of Top 5 Biggest Topics in Pro Hashtag Network, (b). The First Topic, (c). The Second Topic, (d) The Third Topic, (e) The Fourth Topic.

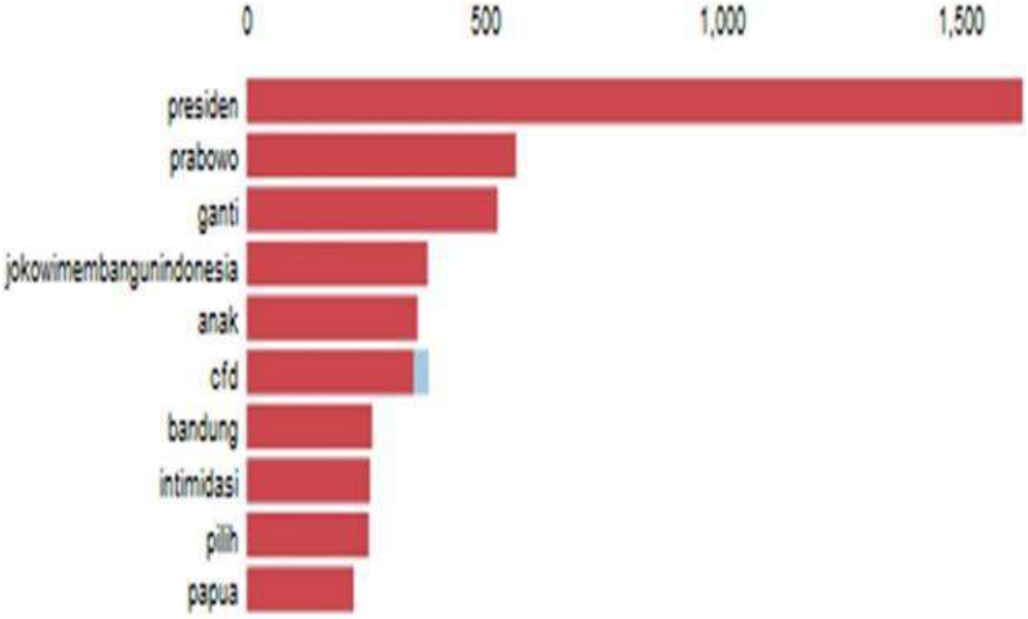
(a)



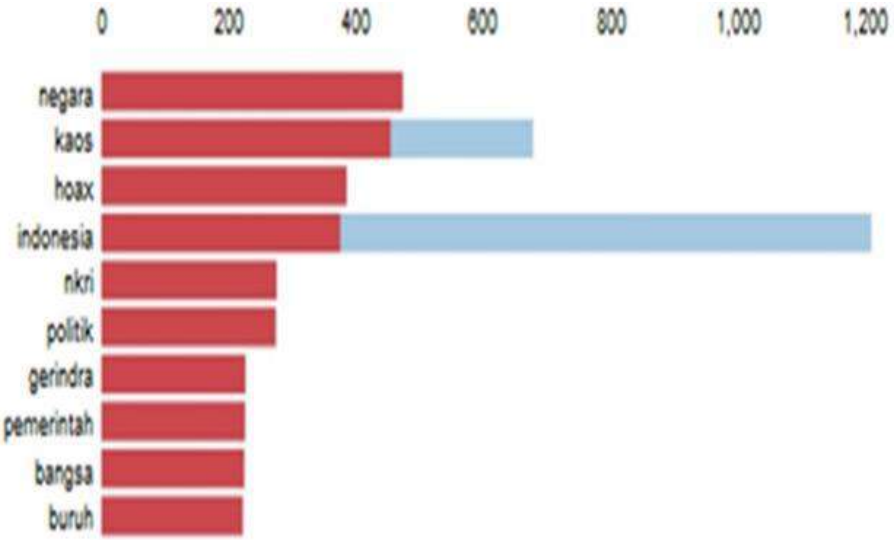
(b)


0
1,000
2,000
3,000
4,000
5,000
6,000
jokowi
indonesia
rakyat
dukung
jokowisekalilagi
joko
widodo
periode
calon
kerjanyata

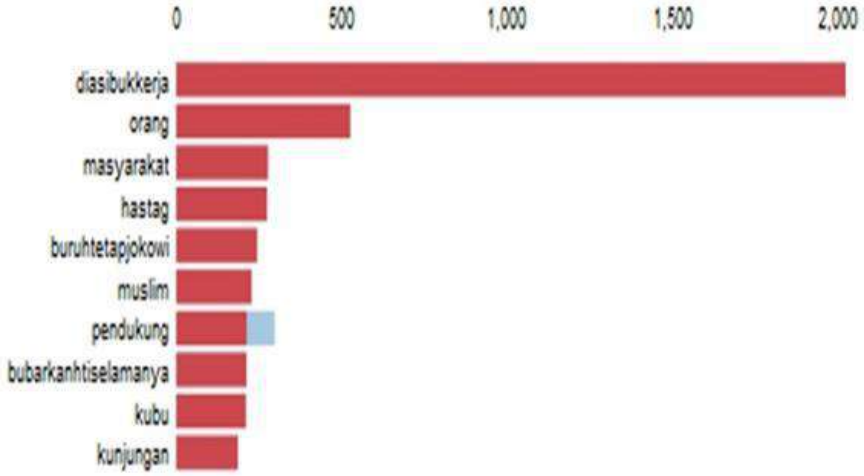
(c)


0
500
1,000
1,500
presiden
prabowo
ganti
jokowimembangunindonesia
anak
cfd
bandung
intimidasi
pilih
papua

(d)




(e)




The first topic in the pro dataset shows the public support for the incumbent president Joko Widodo or Jokowi to step forward for the next period. We draw this conclusion from the presence of these terms: *Jokowi*, *dukung*, *jokowisekalilagi*, *periode*, and *kerjanyata*. *Jokowi* is the most dominant term, showing up 6,000 times. The public mostly talk about the positive qualities that Jokowi has shown during his tenure as president, including his strength, his spirit, and his determination to maintain transparency and fight corruption. Figure 4b shows the distribution of words in the first topic.

From figure 4c, the authors summarize that the second topic concerns one important incident, where some people wearing shirts labelled "*ganti presiden*" were intimidating a mother and her son during a car free day event. Public using pro hashtags assume that the culprits behind this incident come from supporters of Prabowo, the opponent of the incumbent. Public using pro hashtags also claim that they will not retaliate because it goes against their moral or ethical values.

As shown in figure 4d, the third topic concerns the spreading of false rumours or hoaxes in the news, and is represented by the terms *hoax* and *pemerintah*. This mainly refers to efforts by the government to criminalize several religious leaders in Indonesia. Some also talk about Jokowi and his ex's picture on the internet, a rumour that is unfounded. These people use the pro hashtag to ask those who are involved in such hoaxes to cease their action.

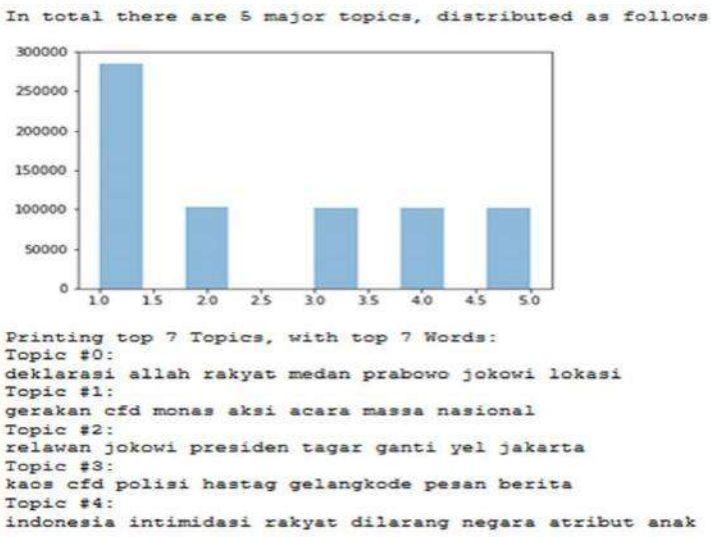
The fourth topic in figure 4e concerns the support given by some laborers to Jokowi. The term *buruhtetapjokowi* refers to the labour unions who stand with a political party named *Partai Demokrasi Indonesia Perjuangan* (PDIP). They claim that they will stand by PDIP because it has the same goal as the laborers in Indonesia. The aspiration of laborers is, admittedly, to be listened to, and this is the main reason for their support on this side.

### 2) Contra Hashtags

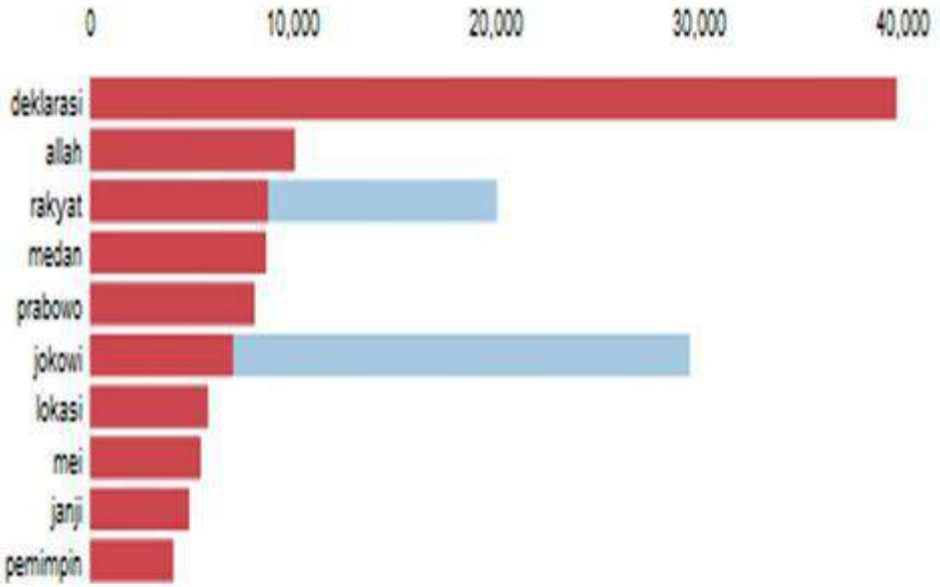
Figure 5a shows the distribution of the top 5 topics in the dataset generated by the contra hashtag. The results of figure 5b show that the most dominant topic is the declaration of *#gantipresiden*, which takes place in some cities in Indonesia. This is supported by the presence of the term *deklarasi*, which means declaration, as the most dominant term in the topic. We can also see the term *Medan*, which is the name of the city where the declaration event occurs. Medan becomes attentional because some local citizens who attend car free day events wear shirts saying "*ganti presiden*", a symbol against the incumbent. This is probably in response to the letter issued by the local government that promised to discipline those people who wear "*ganti presiden"* attire to the car free day event.

**Figure 5.** (a). The Distributions of Top 5 Biggest Topics in Contra Hashtag Network, (b). The First Topic, (c). The Second Topic, (d) The Third Topic, (e) The Fourth Topic.

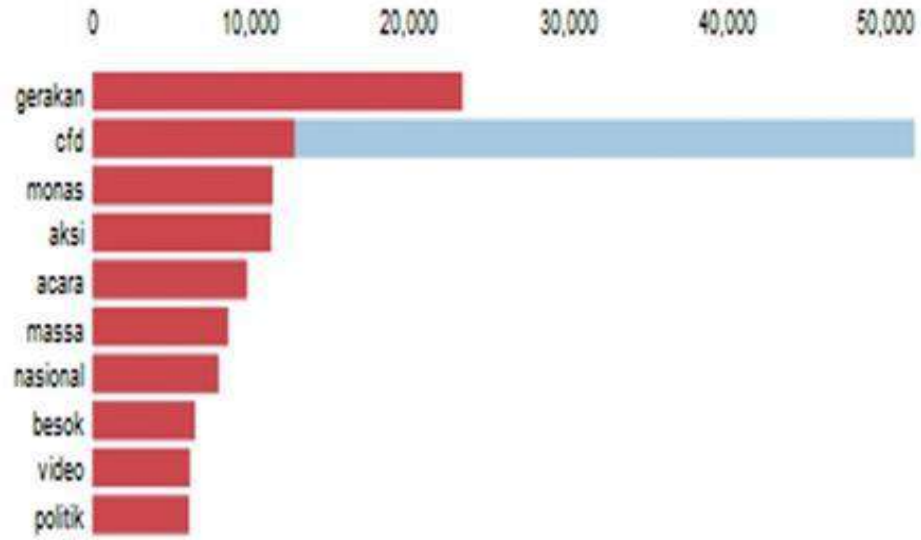
(a)



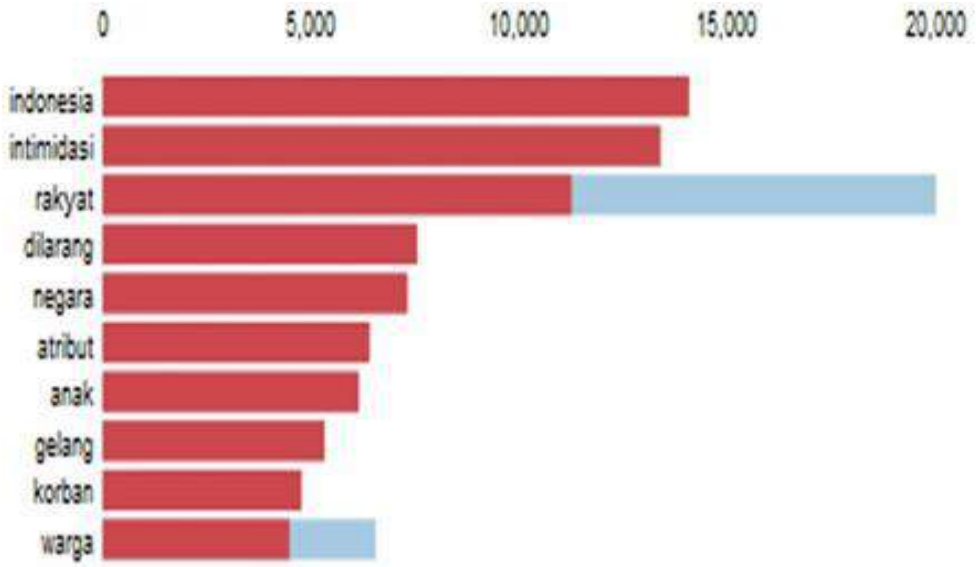
(b)


0
10,000
20,000
30,000
40,000
deklarasi
allah
rakyat
medan
prabowo
jokowi
lokasi
mei
janji
pemimpin

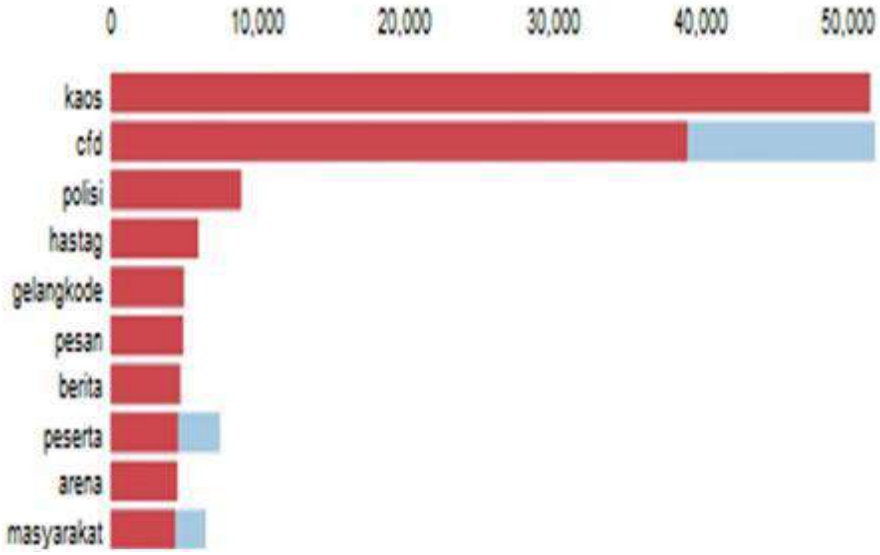
(c)


0
10,000
20,000
30,000
40,000
50,000
gerakan
cfd
monas
aksi
acara
massa
nasional
besok
video
politik

(d)




(e)




The second dominant topic in the contra hashtag is the demonstration by people who joined the march at National Monument. According to the media, this demonstration was held on May 6th, 2018, which is outside the range of the data collection scope. Hence, the authors concluded that there were many people talking about the demonstration before the event had occurred. The words distribution of second topic is shown in Figure 5c.

The next topic of the contra hashtags concerns the intimidation incident, the same as the second biggest topic of the pro hashtag. Both sides of the opinion discuss the same event using a completely different tone. Mostly, the contra hashtag groups reject the association

between the intimidators and Prabowo supporters, although they were wearing Prabowo attire at the time. The words distribution of the second topic is shown in Figure 5d.

The fourth topic tells us about the dismissal of people wearing the "*ganti presiden"* t-shirt by the police. Figure 5e shows the most dominant term is *kaos*, which means t-shirt, followed by *cfd*, which means car free day event. The term *polisi*, which mean police, is also present.

From the TM model, the authors have concluded that the conversational topics on each side are different. Each side picks the most beneficial topic to promote for the advantage of their own side. Even when they discussed the same event, each side presented it in a way that contradicted the other. This shows that social polarization happens at a topics level.

### *B. Social Network Analysis (SNA)*

The authors recorded a total of 118,144 actors and 544,075 relations based on pro and contra hashtags. Figure 6 shows the whole network visualization formed from the ten days of data collection. The network consists of actors and edges from the pro network in Figure 6a, which looks much denser compared to Figure 6b. This shows that the pro network has much more relations between actors and that this network generates much more conversation.

**Figure 6.** The Social Network of (a). Pro Network, (b). Contra Network

(a) (b)

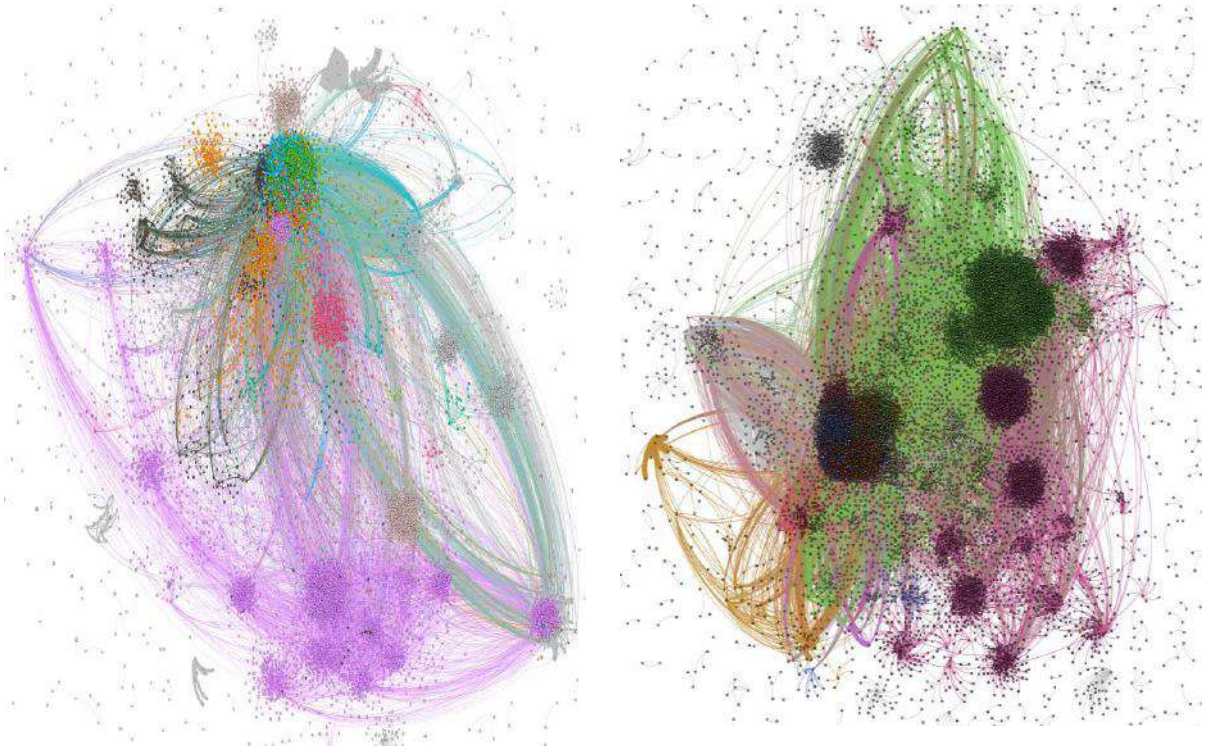


**Table 3:** Network Properties of Pro and Contra Networks

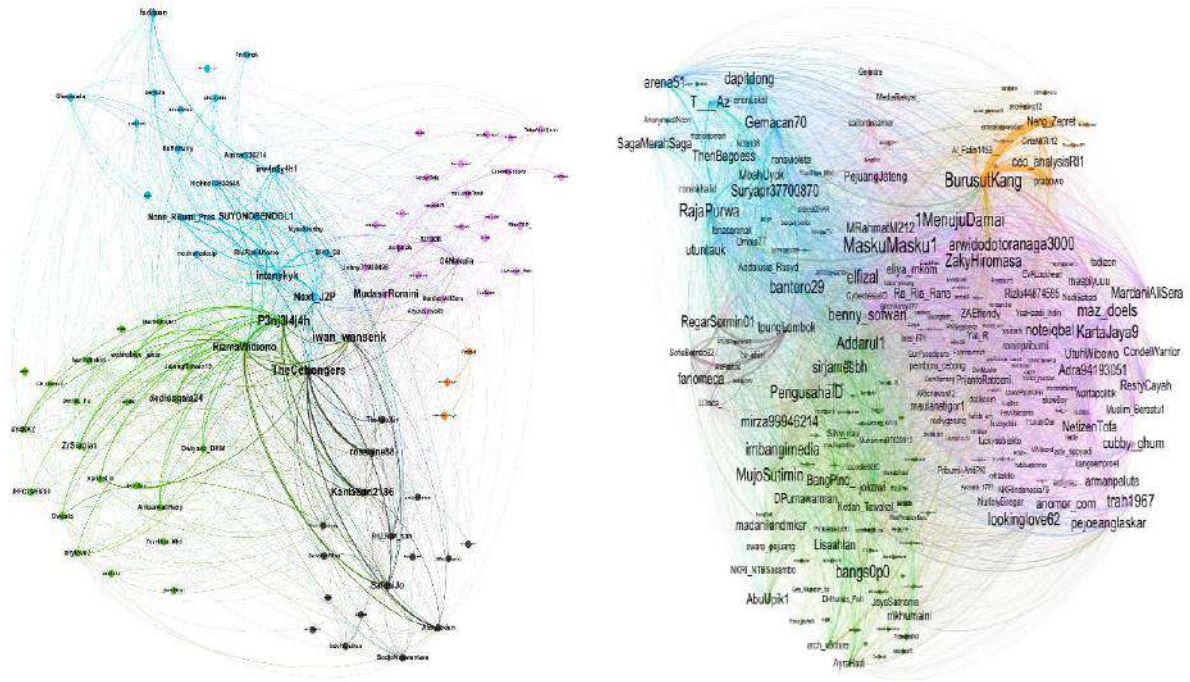
| Network Properties | Pro | Contra |
|---|---|---|
| Tweets | 24,097 | 418,256 |
| Nodes | 29,791 | 88,353 |
| Edges | 81,851 | 462,224 |
| Average Degree | 7.509 | 6.79 |
| Diameter | 12 | 19 |
| Density | 0.607 | 0.219 |
| Modularity | 0.423 | 0.713 |

From a technical aspect, the author employed the undirected network type, one that does not need to consider the direction of interactions. Repeated parallel interactions are merged in this network visualization. After visualizing the social network, the author elaborates on the insights extracted from the data from both sides. The network measurement metric of each opposing side is shown in Table 3.

The number of actors from contra network is almost 4 times higher than from pro network, which is represented by the nodes measurement. The number of actor interactions in the contra network is almost eighteen times higher than in the pro one, as shown from by edges measurement. The average number of actor interactions is illustrated by average degree, which is slightly higher in the pro network than in the contra network. The higher the average degree, the faster the information is diffused process from one actor to the rest of the network; hence, we can conclude that the pro network disseminates information slightly faster. Network diameter shows the distance between the furthest-separated actors in the network. This is a considerably shorter distance in the pro network, which equates to a faster transfer process. The modularity measurement shows how distinct the separation is between groups. A higher modularity value means that the members of a network are distinctly separated. Modularity values range between 0 and 1. The modularity value of the contra network shows it to be more distinct than the pro network. This means that actors in the contra network exclusively belong to a particular group; while in the pro network, actors tend to be a member of multiple different groups.

**Figure 7.** The Social Network with Actor Identification of (a). Pro Network, (b). Contra Network

(a) (b)




The identification of influential actors is shown in Figure 7. The key actor's measurement is based on number of connections or acquaintances and is called the degree centrality metric. The highest degree centrality in the pro network belongs to an actor named *@P3NJ3L4J4H*; the highest in the contra network is *@RajaPurwa*. These two actors have the highest capacity to oversee the dissemination of information, since they are the greatest influencers within their own network.

From the SNA model, the authors conclude that although the contra network is much bigger than pro network, most of its participant do not engage in meaningful or impactful conversation. This is shown by the lower density value, lower average degree, and higher diameter value. The authors suspect that a large number of the network participants in the contra network are not real people but rather engineered accounts or bots. In term of social polarization, the author is able to identify a set of prominent actors on each side of the network.

### *C. Dynamic Network Analysis (DNA)*

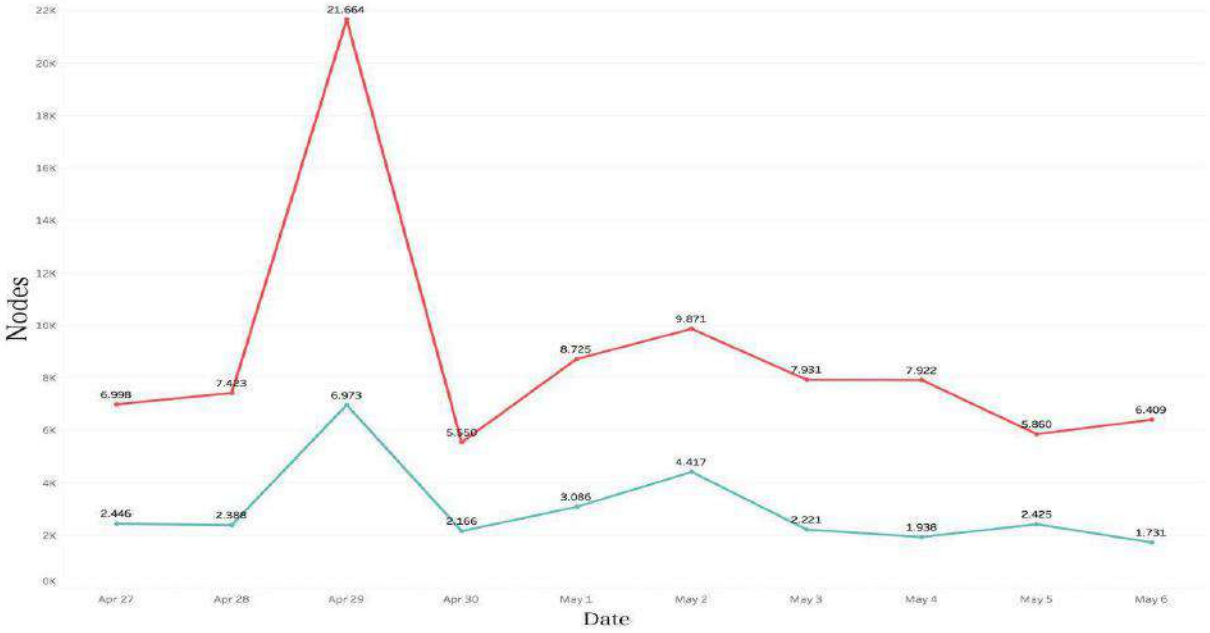
#### 1) Node Evolution

Node evolution shows the dynamic number of actors involved in a conversation network over an observed period. Figure 8a shows the overtime daily chart between the pro and contra

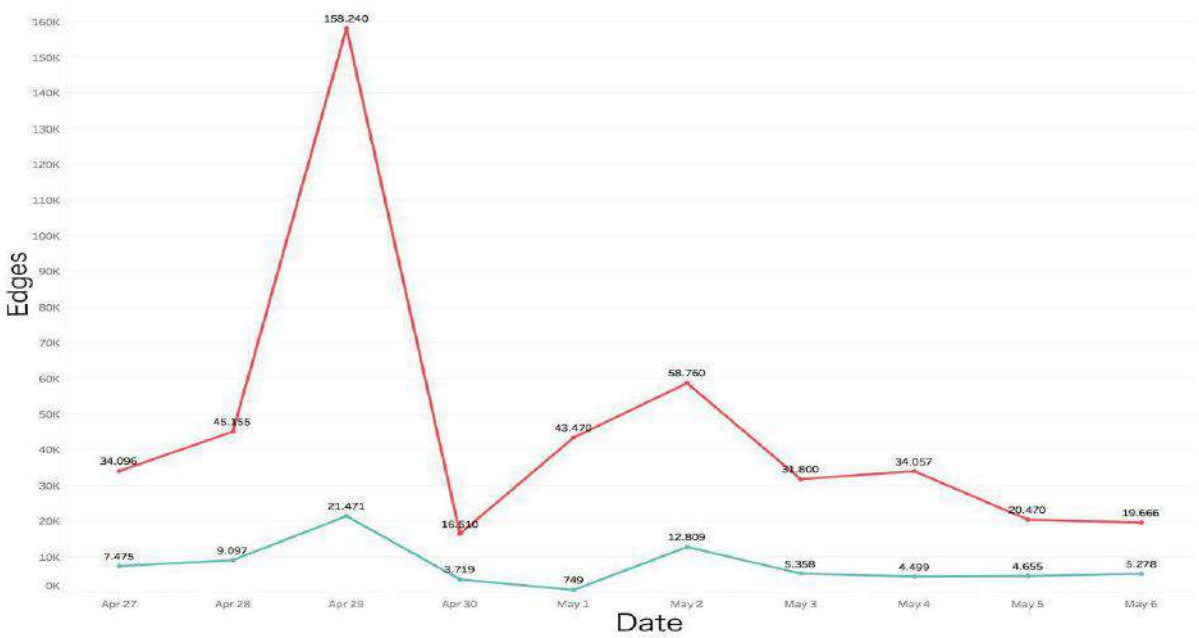
network. At some point on the 3rd day of observation, the number of actors in the contra network suddenly rose sharply. By looking at the dataset, we can see that on April 29, 2018 (Friday), the dominant topic is the Labour Day celebration, which was due in the coming few days. The emergence of the Labour Day topic comes from the efforts of the labour union movement to celebrate the economic and social achievements of the worker. For two days straight, this topic was talked about incessantly. Moreover, the spike in the popularity of this topic occurred on the weekend when most people have more free time, which might explain the high frequency of interactions among actors in both the pro and contra networks.

**Figure 8.** The DNA Chart from each SNA Measurement, Green is Pro Network and Red is Contra Network. The Metrics are (a). Nodes Evolution, (b). Edges Evolution, (c). The Average Degree, (d). The Network Diameter, (e). The Number of Communities, (f). The Network Diameter.

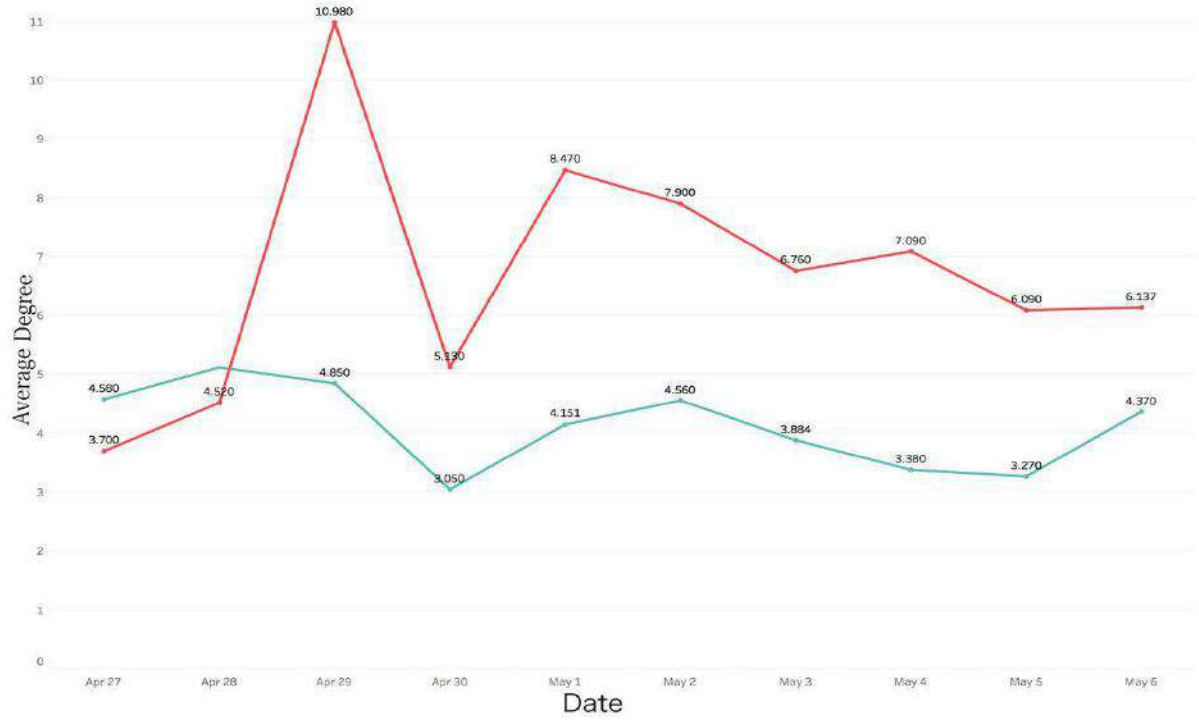
(a)



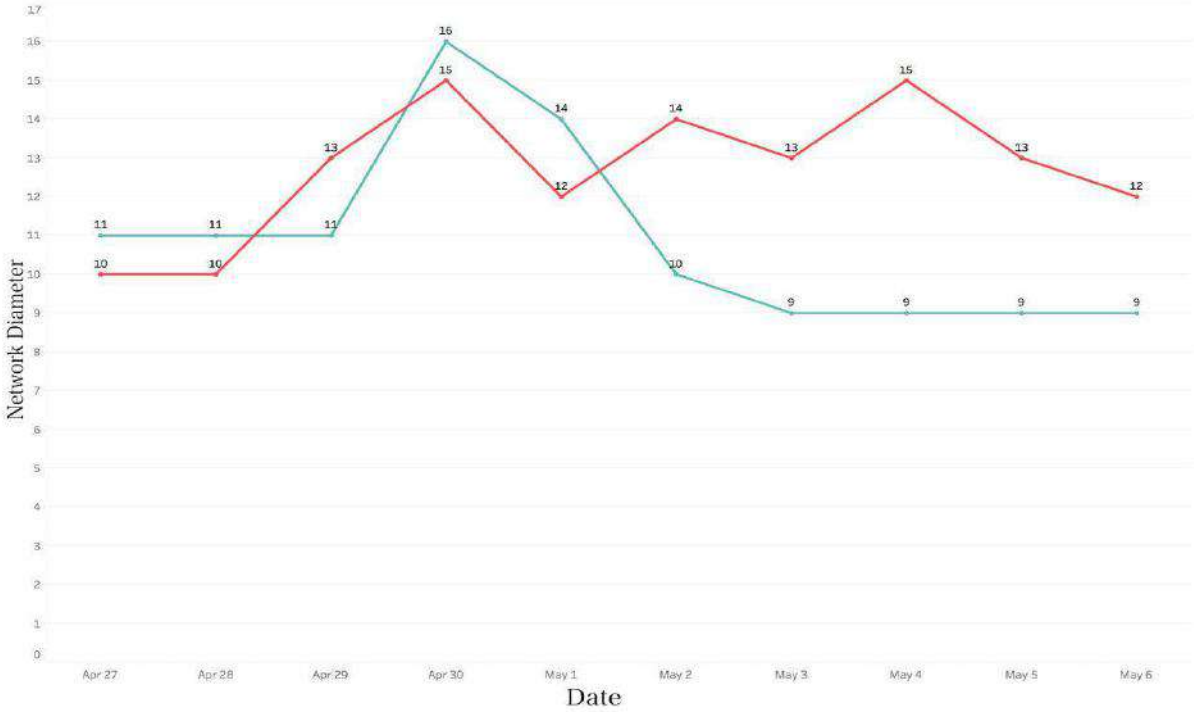
(b)


Edges
Date
Apr 27
Apr 28
Apr 29
Apr 30
May 1
May 2
May 3
May 4
May 5
May 6
34.096
45.155
158.240
16.610
43.470
58.750
31.800
34.057
20.470
19.666
7.475
9.097
21.471
3.719
749
12.809
5.358
4.499
4.655
5.278

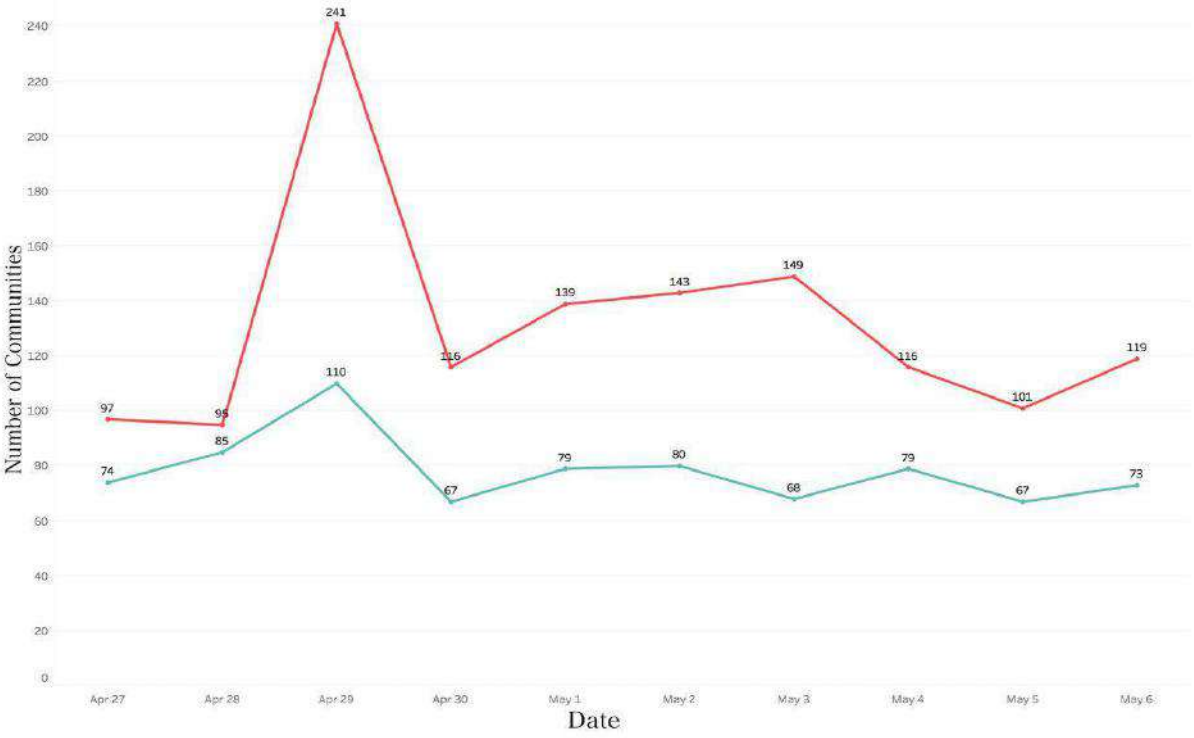
(c)


Average Degree
Date
Apr 27
Apr 28
Apr 29
Apr 30
May 1
May 2
May 3
May 4
May 5
May 6
3.700
4.520
10.980
5.130
8.470
7.900
6.750
7.090
6.090
6.137
4.580
4.850
3.050
4.151
4.560
3.884
3.380
3.270
4.370

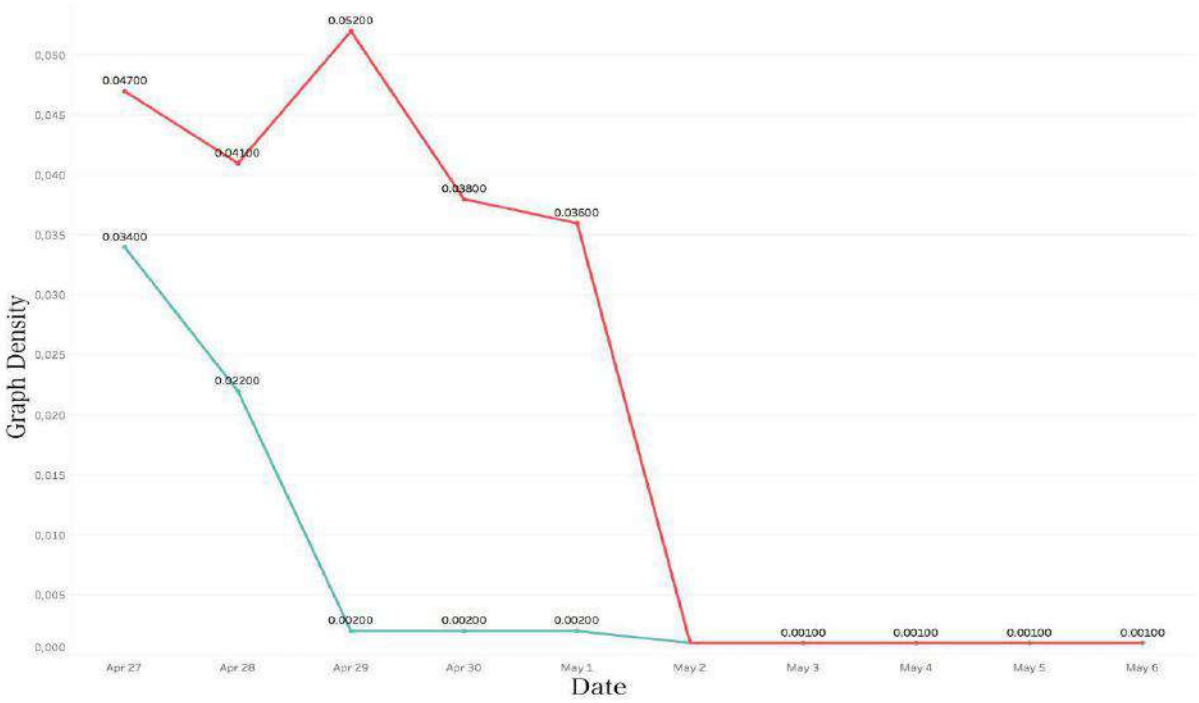
(d)


Network Diameter
Date
Apr 27
Apr 28
Apr 29
Apr 30
May 1
May 2
May 3
May 4
May 5
May 6

(e)


Number of Communities
Date
Apr 27
Apr 28
Apr 29
Apr 30
May 1
May 2
May 3
May 4
May 5
May 6

(f)




**2) Edge Evolution**

Similar to node evolution, the edge evolution peaked on the 3rd day of observations. This is reasonably normal and indicates the higher tendency for actors to generate conversations. Figure 8b shows a direct comparison across all days of observation. Edges represent interactions, which include tweets, replies, mentions, and retweets. The conversations on the 3rd day are about Labour Day and Jokowi's achievement over the past four years. Conversation is prompted by the asking of and responding to what impacts Jokowi's actions have had on Indonesia. It can be seen that a fluctuation of edges tends to occur more on the weekend.

**3) Average Degree Evolution**

Figure 8c shows the daily dynamic of average degree measurement. The purpose of this metric is to see the average interactions of all actors in the network. For the contra network, the 3rd day still holds the highest average degree, followed by the 5th day, and the 7th day. Whereas in the pro network, the 2nd day has the highest value, followed by the 6th day, and the 10th day. We detected that central nodes or key important actors were integral to the spread of information to others in their immediate neighbourhood.

**4) Network Diameter Evolution**

This analysis is used to determine the development of the social network, characterized by the network diameter during the observation period. Network diameter is determined by the

smallest number of steps that must be taken to connect the pair of furthest actors in the network. Smaller network diameter means that information can travel faster, making it easier to disseminate information throughout the network. From Figure 8d, the contra network's diameter tends to fluctuate more compared to the pro network. The lowest network diameter value for the pro network is 9 and the contra network is 10. The dynamic pattern of the network diameter does not form a trend in either network; it can be seen that the pro network line graph tends to be more stable compared to the contra network. This means that the pro network is easier to predict, as shown by its capability to retain the network diameter value.

**5) Modularity and Communities Evolution**

We found that the number of communities varies according to daily interactions. Communities are determined by modularity measurement, the main function of which is to detect the presence of communities. Modularity itself measures the tendency for nodes to group. A higher modularity means that the communities are distinct and that nodes belong exclusively to a single community. Figure 8e shows that the contra network has the higher number of communities on the 3$^{rd}$ day, while the highest modularity value in both networks occurred on the 4$^{th}$ day. We concluded that the distinct grouping on the 4$^{th}$ day reflects the stronger relationship inside the communities.

**6) Network Density Evolution**

Network density measurements show the potential for a network to become a strong, fully connected network. Network density calculates the ratio between the current number of edges and the maximum number of possible edges. The higher the network density value, the closer the relationship among actors in the network. Figure 8f shows that network density tends to decrease over time. From the 6$^{th}$ day, network density had reached almost zero, meaning that the network is either far less connected than it had previously been or that the number of nodes had increased significantly, while the number of edges had not.

DNA models enable the authors to see the time evolution of each opponent network from the measurement given by SNA metrics. Whether the pro and contra network reach their potential network configuration depends on the several opportunities driven by available events on a particular day. As a compliment methodology to the content analysis (TM) and structure analysis (SNA), the authors concluded that social polarization occurs based on which topics are beneficial to each network, hence the different peaks reached by the opposing sides of the network.

***D. Text Network Analysis (TNA)***

The pattern of relations between texts or terms in the pro and contra networks can be seen in Figure 9. The edges connecting the texts give us sense of the document's context. The edge

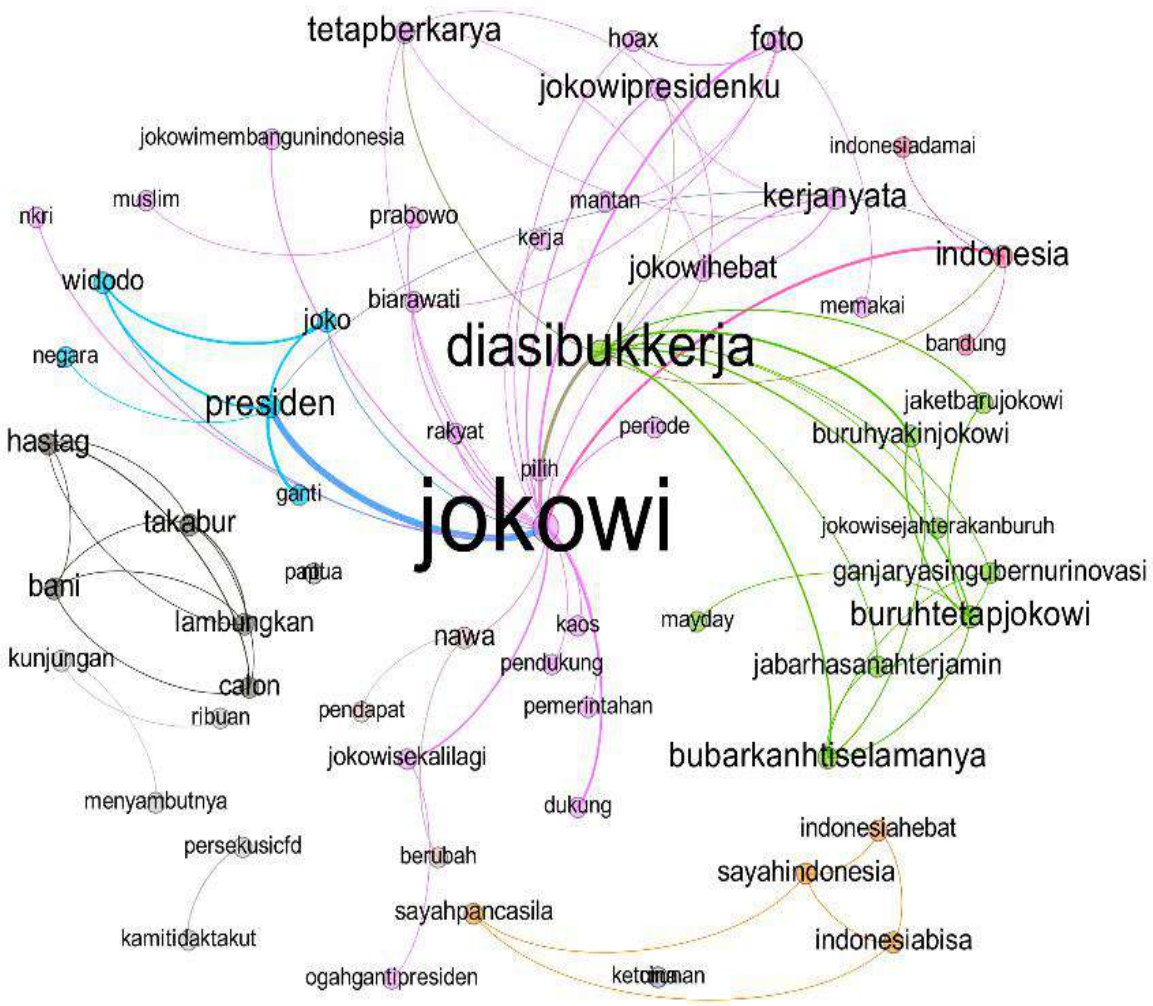
thickness shows the edge's strength, represented by weight. The thicker the edges, the more frequently a pair of texts share phrases. The nodes' thickness represents the frequency with which a particular text shows up in a phrase or document. The weight of both networks is shown in Table 4. To verify the results' consistency, Table 4a shows that *jokowi* and *president* text has a frequency of 818, and in Figure 9a, *jokowi* and *president* have the thickest edges.

Figure 9 also shows that we are able to detect network communities, which are distinguished by colour. In this sense, the communities formed in the text network often correlate with the topics found in a particular network. For example, in Figure 9a, the first topic, represented by the colour purple, tells us about public support for Jokowi to run for a second term of presidency. In the contra network, the text *ganti* and *presiden* have the thickest edge. Based on Table 4b, both terms have the highest weight, which is 12,828. The contra network produces several topics, the first of which, also denoted by the colour purple contain terms related to the incident at the car free day event.

The TNA model gives the authors the opportunity to drill deeper into the content analysis aspect. Even though the topics have been discovered by TM methodologies in the initial discussions, here TNA provides more easily interpretable topics through text network constructions. The TNA results confirm similar social opinion polarization to that which was found in the TM methodologies.

**Figure 9.** The Text Network of: (a) Pro Network, (b) Contra Network

(a)



*(b)*

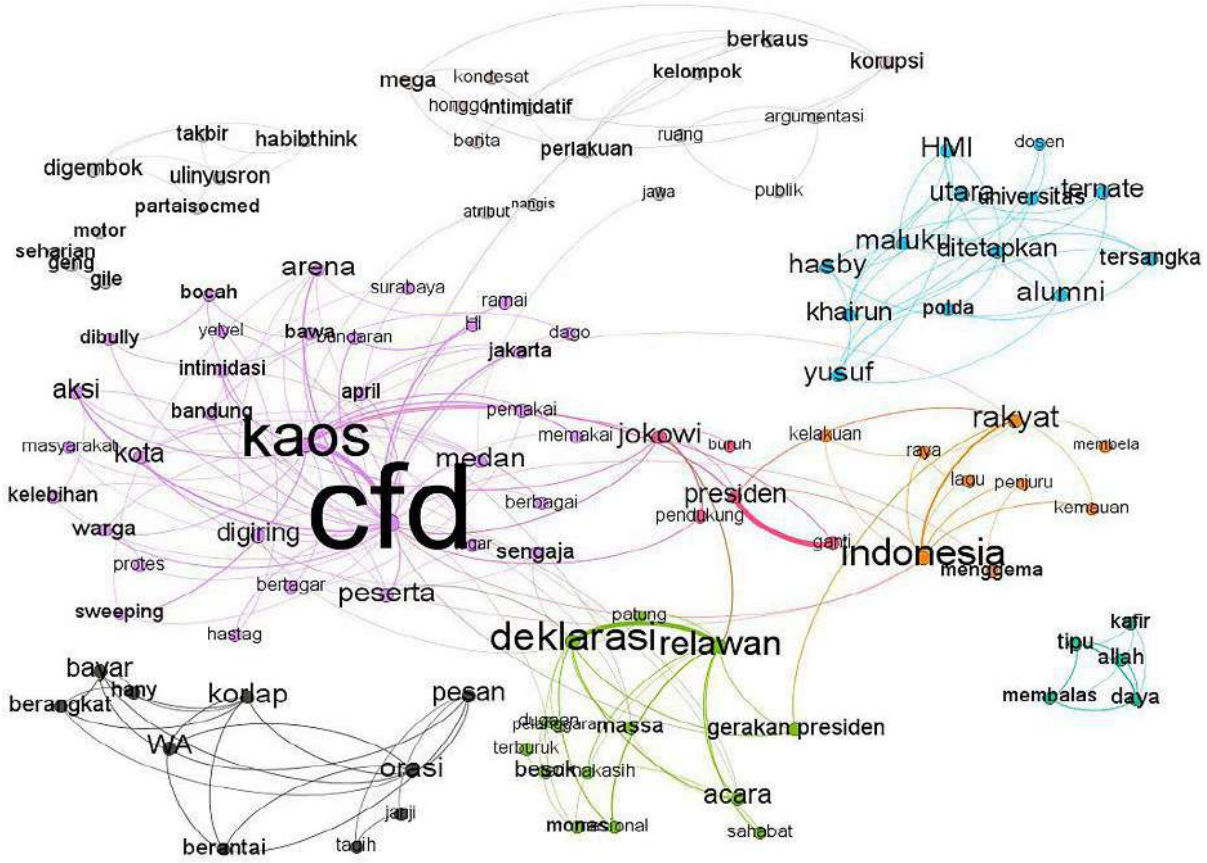



**Table 4:** The Top 14 Relations of Texts in: (a) Pro Network, (b) Contra Network

(a)

| No. | Source | Target | Weight |
|---|---|---|---|
| **1** | jokowi | presiden | 818 |
| **2** | jokowi | diasibukkerja | 484 |
| **3** | ganti | Presiden | 389 |
| **4** | jokowi | indonesia | 351 |
| **5** | jokowi | dukung | 314 |
| **6** | jokowi | foto | 306 |
| **7** | joko | widodo | 277 |
| **8** | presiden | Joko | 276 |
| **9** | presiden | widodo | 272 |

(b)

| No | Source | Target | Weight |
|---|---|---|---|
| **1** | ganti | presiden | 12828 |
| **2** | kaos | cfd | 11000 |
| **3** | deklarasi | relawan | 10536 |
| **4** | indonesia | rakyat | 4710 |
| **5** | cfd | aksi | 4546 |
| **6** | cfd | HI | 3958 |
| **7** | cfd | arena | 3936 |
| **8** | cfd | jakarta | 3891 |
| **9** | kaos | presiden | 3778 |

| | | | |
|---|---|---|---|
| **10** | diasibukkerj | buruhyakinjoko | 268 |
| **11** | jokowi | Jokowisekalilag | 257 |
| **12** | diasibukkerj | Bubarkanhtisela | 227 |
| **13** | jokowi | Jokowipresiden | 214 |
| **14** | diasibukkerj | buruhtetapjoko | 212 |

| | | | |
|---|---|---|---|
| **10** | tipu | daya | 3640 |
| **11** | kaos | memakai | 3601 |
| **12** | nasional | relawan | 3316 |
| **13** | jokowi | relawan | 3302 |
| **14** | jokowi | pendukung | 3261 |

## Conclusion

Opinion polarization is happening in Indonesia around political events in 2019. We proved this by showing how one particular event was discussed by opposing sides with a differing tones. The contradictory opinion examined in this study was whether or not (pro or contra) to support Jokowi. The four methodologies, TM, SNA, DNA, and TNA, were used to reveal a comprehensive understanding on the polarization process. By understanding conversation content and topology, we are able to decipher the issues extracted from the content, as well as the dynamics of conversation network. By connecting to the right people, we can see how topics and ideas disseminate and go viral and how the interactions between actors in the social network dynamically respond. The network dynamics behaviour gives us insights into how the network evolves when it is stimulated. It also helps us to understand the evolution of opinions in each network, how the contra side suddenly changes to the pro side and vice versa.

This research gives us an understanding of the whole process of virality and allows us to observe the timely interactions between actors. Social media forms a social network that is in some ways similar to real-world networks, as it represents social behaviour. This correlation is shown by the tendency for topics that go viral in the real world trending in social media. In addition, high traffic commonly occurs during weekends and national celebration days, as compared to weekdays. For example, the actions of people who are contra Jokowi in some cities of Indonesia spiked on April 29$^{th}$, and the national celebration day of labour on May 1$^{st}$ had a significant impact on many important actors during that time.

The authors recommend to separate polarization directly using Sentiment Analysis methodology based on machine learning, as it allows for the possibility of detecting contradictory opinions on the pro and contra sides simultaneously. The second recommendation is the removal of the pre-processing step to opponent classification; instead, this can be implemented on real time application or processing stream data. Furthermore, this will also help us identify false campaigns, which are campaigns on the right hashtag but carrying divergent messages. A final suggestion is to obtain more data over a longer period of observation.

## REFERENCES

Alamsyah, A., Priyana, Y., Rahardjo, B., and Kuspriyanto. (2017). Fast Summarization of Large-Scale Social Network using Graph Pruning based on K-Core Property. Journal of Applied and Theoretical Information Technology, Vol. 95, Issue 16.

Alamsyah, A., Bratawisnu, M.K., and Sanjani, P.H. (2018). Finding Pattern in Dynamic Network Analysis. The 6th International Conference on Information and Communication Technology.

Blei, B.M., Ng, A., and Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research 3, pp. 993-1022.

del Val, E., Rebollo, M., and Botti, V. (2015). Does the Type of Event Influence How User Interactions Evolve on Twitter? PLoS ONE 10(5): e0124049.

Hanneman, R.A., and Riddle, M. (2005). Introduction to Social Network Methods. Riverside, United States: University of California.

Hunter, S. (2014). A Novel Method of Network Text Analysis. Open Journal of Modern Linguistics, 4, 350-366, 2014.

Leskovic, J., Kleinberg, J., and Faloutsos, C. (2007). Graph Evolution: Densification and Shrinking Diameters. Journal ACM Transactions on Knowledge Discovery from Data, Vol 1, Issue 1, Article No. 2.

Newman, M.E.J. (2011). Network: An Introduction. University Michigan and Santa Fe Institute. Oxford University Press.

Paranyushkin, D. (2011). Identifying the Pathways for Meaning Circulation using Text Network Analysis. Berlin: Nodus Labs.

White, H. C. (1992). Identity and control: A Structural Theory of Social Action. Princeton University Press.