Hashtag activism and message frames: social network analysis of Instagram during the COVID-19 pandemic outbreak in Indonesia

A Priadana$^1$ and S P Tahalea$^{2*}$

$^1$Center of Data Analytic Research and Services, Universitas Jenderal Achmad Yani Yogyakarta, Indonesia
$^2$Department of Informatics Engineering, Universitas Pembangunan Nasional "Veteran" Yogyakarta, Indonesia

E-mail: $^1$dana@unjaya.ac.id, $^{2*}$sylvert@upnyk.ac.id

Abstract. During the Coronavirus pandemic, social media played an essential role in disseminating information online by utilizing various message frames and hashtags. Instagram is a popular social media worldwide that are used to spread various information. This study applied the social network analysis (SNA) approach as a theoretical framework and explored how the relationship between users on the Instagram platform related to Coronavirus outbreaks. Moreover, we also investigate how the relationship between users and their hashtags. SNA techniques were used for visualizing network models using an undirected graph, measuring network attributes, and centrality measures to find the most influential users in the network. A total of 10,403 posts based on #wabahcorona on Instagram from February 28, 2020, to May 18, 2020, were analyzed. Based on the analysis results, hashtags play an important role in this topic. Degree centrality measure as connectivity number shows that only two user accounts can make it into the list’s top ten. When using the Eigenvector centrality measure, there are no users in the top ten. The modularity measure detects 122 distinct communities that show the dense connection of nodes in the networks. Betweenness centrality measure shows that there are six related hashtags to COVID-19 pandemic out of ten hashtags. They are #wabahcorona, #covid19, #corona, #viruscorona, #dirumahaja, and #coronavirus and the four are Islamic related hashtags. The Islamic related hashtags are caused by the date of Ramadhan month.

1. Introduction
Corona virus diseases or COVID-19 is a rapidly spreading global pandemic, which, as of July 23, 2020, has affected more than 15 million people and claimed over 600,000 deaths globally [1]. Since President Joko Widodo reported the first confirmed two cases of COVID-19 infection on March 2, 2020, in Indonesia [2], COVID-19 has affected about 93 thousand people and claimed over 4,000 deaths on July 23, 2020 [3]. This outbreak had an impact on various fields such as economics, health, tourism, and education. This phenomenon motivated many people, especially researchers, to conduct researches related to COVID-19.

Research by utilizing data related to COVID-19 is one of the topics that has been done. Data used in [4] collected from the online platform of the Ministry of Health and Family Welfare in India to predict infection. The research claims that India is likely to witness an increased spreading rate of COVID-19 in June and July. The internet search data from the Baidu Index Platform and China CDC were used to obtain the search volume (SV) of keywords for symptoms associated with COVID-19[5].
The results of this study suggested that the patients who searched relevant symptoms on the internet may begin to see doctors in 2–3 days later and be confirmed in 3–4 days later. In 2020 there is research that used the data daily new confirmed cases of the COVID-19 outbreaks in Japan and South Korea. The daily new confirmed cases data of the COVID-19 outbreaks in Japan and South Korea that are available from the Wind Database[6]. The data were analyzed using the statistical software to predict the daily new confirmed cases in the next weeks.

Furthermore, enactment of a lockdown policy during the COVID-19 pandemic period, people got informed mostly on Online Social Media [7]. This fact has prompted several studies to use social media data such as Twitter, Instagram, etc. to conduct research related to COVID-19. Twitter data also is used to analyze the tone of officials’ tweet text as alarming and reassuring and capture the response of Twitter users to official communications[8]. Over 20 million COVID-19-related Twitter posts are used to examine worldwide trends of four sadness, anger, fear, and joy—and the narratives underlying those emotions during the COVID-19 pandemic[9]. This research found that public emotions shifted strongly from fear to anger throughout the pandemic, while sadness and joy also surfaced. Another Twitter data set is used to analyze emotional public response during the COVID-19 pandemic situation[10]. This study can be used by authorities to understand the mental health of the people.

In addition to research conducted on the Twitter platform, research related to COVID-19 on social media has also been carried out on the Instagram platform that as one of the popular social media [11]. A multilingual coronavirus (COVID-19) Instagram dataset consisting of the post, publisher profile, like, and comment data is provided. It could support diverse research activities such as analyzing rumors spreading and behavioral change during the pandemic and information sharing associated with COVID-19[12]. Google Trends and Instagram hashtags (#) data is used to investigate the internet search behavior related to COVID-19 and the extent of infodemic monikers circulating in Google and Instagram during the pandemic period in the world[13]. Instagram data is also used to analyze how effective #stayhome hashtag as a social campaign to prevent COVID-19[14]. Unlike Twitter, Instagram is a popular social media that is used to share images [15]. The paper aimed to use Instagram data for analyzing hashtag activism and message frames related to COVID-19 by applying the social network analysis (SNA) approach, considering the importance of image-based content in the dissemination of news (and misinformation) [16], [17]. The SNA approach is applied to determine a suitable indicator for characterizing places, along with the tourist's online activities, in terms of sharing pictures on Instagram[18], while another shows the most famous distribution of tourist places and the center popularity of tourist destinations based on an Instagram account[19].

The Social Web provided opportunities for people to share their ideas or opinions with people connected through the internet[20]. Instagram networks can be constructed by using a relationship using the following features between users or using hashtags relation. Some research studies used follow features to detect accounts' popularity and how their networks look on Instagram[21], [22]. Popularity on Instagram is not only about how to collect followers but also about influence. Users with many followers are known as social influencers who can influence their followers about advertising or, more specifically, like their opinion about something[23]. However, the hashtag has a unique way to connect users or other hashtags. Every hashtag can represent similar or different events like politics, comedy, or disaster[24], [25]. This research aims to find the most influential hashtags and communities that get involved in disseminating COVID-19 information on Instagram using Social Network Analysis (SNA) from networks built using users and hashtags network.

2. Method
The process of social network analysis in this study consists of four stages: data extraction and preprocessing, building a network model, measuring centrality value, and measuring modularity value.

2.1. Data Extraction and Preprocessing
We implemented a web scraping technique to extract post data from Instagram web-based on a particular hashtag. We got attributes such as caption, owner id, number of likes, number of comments, post URL, and post time from the post data. The steps in extracting web data using web scraping techniques are as follows [26]. In the analysis phase, the researcher studies the structure of HTML and
JSON from the Instagram website. This process aims to determine the data structure and elements that will be downloaded from an Instagram post. In the coding phase, the researcher makes a crawling code by using the Python programming language. In the implementation phase, the researcher extracts the data on the Instagram web by sending a request to an Instagram web page and then extracting JSON (JavaScript Object Notation) data containing data from an Instagram post.

After we collected the Instagram post data, we prepared and cleaned data called the pre-processing step. In this step, we extracted the hashtag and user accounts from the caption of the post data.

2.2. Network Model
This network is built by using user accounts, and hashtags as nodes and connections between them are edges. The connection between hashtag happens when the hashtags are published in the same post. The connection between user and hashtag happens when a user publishes some hashtags at a post. The network representation of this case is shown in Figure 1.

![Figure 1. Network representation.](image)

2.3. Centrality Measure
Centrality measures used in this research are degree centrality, betweenness centrality, and eigenvector centrality. These centrality measures can show the central or highly influence nodes of the network in some fields, such as criminal networks, social networks, marketing, e-sport, and even the researcher's network[27]–[31].

2.3.1. Degree centrality
This research used degree centrality to identify most connected users or hashtags. More connection means more related posts to hashtags used in this research. It can be measured using equation 1.

\[ C_D(i) = \sum \frac{d_i}{n - 1} \] (1)

Users or hashtags represent as \( i \) and total nodes (users and hashtags) in the network represent as \( n \).

2.3.2. Betweenness centrality
Betweenness centrality measures can show how many users or hashtags are acting as bridges in the network. It can be measured using equation 2.

\[ C_B(i) = \sum \frac{g_{j,k}(i)}{g_{j,k}} \] (2)

Users or hashtags represent as \( i \), the number of shortest paths from actor \( j \) to actor \( k \) shown as \( g_{j,k} \), and the number of shortest paths from actor \( j \) to actor \( k \) through actor \( i \) shown as \( g_{j,k}(i) \).

2.3.3. Eigenvector centrality
Eigenvector shows the importance of users or hashtags based on their connection. This centrality of a node relies on the centrality of its neighbor's centrality [32]. Eigenvector centrality can be measured using equation 3.
\[ C_e(i) = \frac{1}{\lambda} \sum_{j \in G} a_{i,j} \]  

(3)

Users or hashtags represent as \( i \), constant represents as \( \lambda \), and \( a_{i,j} \) is shown adjacency matrix of the network.

2.4. Modularity

Modularity measure is an objective function for network cluster analysis [33]. Modularity can be measured from weighted or unweighted networks [34] as it can be measured using equation 4.

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

(4)

Weight of the edge between nodes \((i,j)\) represent as \( A_{i,j} \), the sum of the weights of the edges attach to node \( i \) represents as \( k_i \), a community of node \( i \) represents as \( c_i \) and \( m = \frac{1}{2} \sum_{i,j} A_{i,j} \).

3. Result and Discussion

This section discusses the results of the calculations that have been carried out. For the first time, we did data profiling from the data extraction. Then a social network analysis is carried out to find most influence users, hashtags, and communities that play an important role in the dissemination of COVID-19 information on Instagram.

3.1. Data Profiling

We have collected 10404 public posts form Instagram based on #wabahcorona hashtag. That meant we extracted posts that mentioned #wabahcorona hashtag. The 10404 posts were gathered from March 2 to May 18, 2020, i.e., 77 days. We selected those dates since President Joko Widodo reported the first confirmed two cases of COVID-19 infection on March 2, 2020, in Indonesia [35]. Then, the data collection was closed on May 18, 2020, when this research analysis begins.

3.2. Degree centrality

Degree centrality shows the connections of hashtags or users. In this network, hashtags have the highest degree centrality. It can happen because the connection between hashtags can happen in one post and have meant that people or users tend to use more than one hashtag in an Instagram post. There are only two user accounts that can make it into the list’s top ten, as shown in Table 1. The highest degree centrality is owned by #wabahcorona, which is the spotlight in Indonesia with 2888 connections. In the top ten, there are eight related hashtags about pandemic COVID-19 and two user accounts that are not social influencer accounts but news portal accounts. They are @teras_sumut with 448 connections and @katadatacoid with 309 connections.

| No. | Id          | Username       | Degree |
|-----|-------------|----------------|--------|
| 1   | #wabahcorona| -              | 2888   |
| 2   | #covid19    | -              | 1124   |
| 3   | #corona     | -              | 1008   |
| 4   | #dirumahaja | -              | 891    |
| 5   | #viruscorona| -              | 748    |
| 6   | #coronavirus| -              | 520    |
| 7   | 31353651301 | @teras_sumut  | 448    |
| 8   | #covid_19   | -              | 349    |
| 9   | 2272159145  | @katadatacoid  | 309    |
| 10  | #wabah      | -              | 303    |
3.3. Betweenness Centrality

Betweenness centrality can show user accounts or hashtags possibilities to become the bridges or medium of information flow. In this network model, there are only three user accounts in the top ten list, as shown in Table 2, which means hashtags still play an important role in information flow. The connection between hashtags in the same post makes them show in some user’s homepage because the Instagram algorithm will think they are related. The more posts involving a hashtag means that hashtags will have a high possibility of becoming a bridge and a high betweenness centrality score, which will increase the possibility of being shown on the user’s homepage.

| No. | Id             | Username        | Betweenness |
|-----|----------------|-----------------|-------------|
| 1   | #wabahcorona   | -               | 0.647968    |
| 2   | #dirumahaja    | -               | 0.076363    |
| 3   | #covid19       | -               | 0.061249    |
| 4   | #corona        | -               | 0.041044    |
| 5   | 31353651301    | @teras_sumut   | 0.03072     |
| 6   | #viruscorona   | -               | 0.025566    |
| 7   | #coronavirus   | -               | 0.019149    |
| 8   | 2272159145     | @katadatacoid   | 0.017233    |
| 9   | #WabahCorona   | -               | 0.016964    |
| 10  | 1366854394     | @fameela9412   | 0.016852    |

The hashtags that have high betweenness centrality scores are hashtags that are related to the COVID-19 pandemic topic. There are several differences between where some hashtags that high degree centrality value is missed in the top ten and replaced by other hashtags and a user account. Table 2 also shows that #wabahcorona has the highest betweenness centrality score with a large gap to the second-highest score #dirumahsaja means that #wabahcorona plays a vital role in information flow. As for user accounts, there are only three user accounts that do not signify that an account can play a crucial role in a topic or challenging issue. The result is similar to degree centrality scores, where @teras_sumut have the highest betweenness centrality score among user accounts.

3.4. Eigenvector centrality

Eigenvector centrality is able to identify the most popular nodes in the network. The popularity is not only based on several connected nodes but also using the importance. The result of eigenvector centrality is shown in Table 3.

| No. | Id             | Username        | Eigenvector |
|-----|----------------|-----------------|-------------|
| 1   | #wabahcorona   | -               | 1           |
| 2   | #covid19       | -               | 0.493444    |
| 3   | #corona        | -               | 0.463667    |
| 4   | #viruscorona   | -               | 0.359933    |
| 5   | #dirumahaja    | -               | 0.350106    |
| 6   | #coronavirus   | -               | 0.209334    |
| 7   | #ramadhan2020  | -               | 0.145182    |
| 8   | #shahihfiqih   | -               | 0.136329    |
| 9   | #muhasabah     | -               | 0.136065    |
| 10  | #ramadhan1441  | -               | 0.134911    |
Table 3 shows the most popular hashtags as a node. This condition happens because of a lack of mention between users in the posts. Users tend to use more hashtags than mention other user accounts in the posts. There are six related hashtags to COVID-19 pandemic out of ten hashtags. They are #wabahcorona, #covid19, #corona, #viruscorona, #dirumahaja, and #coronavirus and the four are Islamic related hashtags. The hashtags related to the COVID-19 pandemic tend to have COVID-19 related words like corona, COVID, and virus, while #dirumahaja is the only appeal associated with the hashtag. That hashtag is an appeal to engage people to stay in their houses to limit the virus’s spread. The Islamic related hashtags caused by the date of Ramadhan month and the other are appeals to other people to stay safe.

3.5. Modularity
Modularity is simply a method to divide the network into some clusters. The modularity score is 0.719, with 122 communities detected in the networks, the smallest community has two nodes, and the biggest community has 1936 nodes. The size distribution is shown in Figure 2.

![Size Distribution](image)

**Figure 2.** Modularity distribution.

4. Conclusion and Further Research
Throughout this paper, the main goal was to identify the most influential hashtags and use Instagram accounts based on COVID-19 topics. The result shows that hashtags play an important role in this topic. The hashtags #wabahcorona is the most influential hashtags based on the result because it happened to be the 1st in every centrality measure. This fact is certainly appropriate because we are taking data based on the hashtag #wabahcorona. However, we can see the ranking of the hashtags below, which are the most influential. In the top ten, there are eight related hashtags about pandemic COVID-19 and two user accounts which are not social influencer account but news portal account. They are @teras_sumut with 448 connections and @katadaticoid with 309 connections. There are 122 communities detected in the networks of different size. The smallest community happens to have two nodes, and the biggest community has 1936 nodes. There are six related hashtags to COVID-19 pandemic out of ten hashtags. They are #wabahcorona, #covid19, #corona, #viruscorona, #dirumahaja, and #coronavirus and the four are Islamic related hashtags.
Although this research was designed to identify most influenced hashtags and user accounts, it also has limitations. Further research can be designed to explore the structure or model the information dissemination using other attributes based on an Instagram feature such as likes and comments.

Acknowledgment
The authors would like to thank the Center of Data Analytic Research and Services, Universitas Jenderal Achmad Yani Yogyakarta, for supporting the data and the Department of Informatics Engineering, Universitas Pembangunan Nasional "Veteran" Yogyakarta, for supporting data visualization and exploration of this research.

References
[1] World Health Organization (WHO) 2020 The Coronavirus Disease 2019 (COVID-19) Situation Report [Online] Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
[2] Djalante R et al 2020 Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020 Prog. Disaster Sci. 6 100091
[3] Peta Sebaran Kasus COVID-19 di Indonesia 2020 [Online] Available: https://covid19.go.id/peta-sebaran [Accessed: 23-Jul-2020]
[4] Kumar A, Rani P, Kumar R, Sharma V, and Purohit S R 2020 Data-driven modelling and prediction of COVID-19 infection in India and correlation analysis of the virus transmission with socio-economic factors Diabetes Metab. Syndr. Clin. Res. Rev. 14(5) 1231–1240
[5] Qiu H J et al. 2020 Using the internet search data to investigate symptom characteristics of COVID-19: A big data study World J. Otorhinolaryngol. - Head Neck Surg.
[6] Duan X and Zhang X 2020 ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data Data Br 31 105779
[7] De Santis E, Martino A, and Rizzi A 2020 An Infoveillance System for Detecting and Tracking Relevant Topics from Italian Tweets During the COVID-19 Event IEEE Access 1(1)
[8] Rao H R, Vemprala N, Akello P, and Valecha R 2020 Retweets of officials’ alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management Int. J. Inf. Manage. 102187
[9] Lwin M O et al. Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends. JMIR public Heal. Surveill. 6(2) e19447
[10] Mathur A, Kubde P, and Vaidya S 2020 Emotional Analysis using Twitter Data during Pandemic Situation: COVID-19 5th International Conference on Communication and Electronics Systems (ICCES) 845–848
[11] Akrianto M I, Hartanto A D, and Priadana A 2019 The Best Parameters to Select Instagram Account for Endorsement using Web Scraping 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITSEE) 40–45
[12] Zarei K, Farahbakhsh R, Crespi N, and Tyson G 2020 A First Instagram Dataset on COVID-19
[13] Rovetta A and Bhagavathula A S 2020 Global Infodemiology of COVID-19: Focus on Google web searches and Instagram hashtags medRxiv
[14] Umar N J 2020 (#)Stayhome hashtag As A Social Campaign To Prevent The Covid 19 through Instagram in Makassar City Communication and Islamic Broadcasting Study Program Faculty of Ushuluddin and Da’wah
[15] Priadana A and Habibi M 2019 Face Detection using Haar Cascades to Filter Selfie Face Image on Instagram International Conference of Artificial Intelligence and Information Technology (ICAIT) 6–9
[16] Zannettou S et al. 2017 The web centipede: Understanding how web communities influence each other through the lens of mainstream and alternative news sources Proc. ACM SIGCOMM Internet Meas. Conf. IMC, F1319 405–418
[17] Zannettou S et al. 2018 On the origins of memes by means of fringe web communities Proc. ACM SIGCOMM Internet Meas. Conf. IMC 188–202
[18] Giordano G, Primerano I, and Vitale P2020 A Network-Based Indicator of Travelers
Performativity on Instagram Soc. Indic. Res. 1–19

[19] Iswandhani N and Muhajir M 2018 K-means cluster analysis of tourist destination in special region of Yogyakarta using spatial approach and social network analysis (a case study: Post of @explorejogja instagram account in 2016) Journal of Physics: Conference Series 974(1) 12033

[20] Pozzi F A, Fersini E, Messina E, and Liu B 2016 Sentiment Analysis in Social Networks Elsevier

[21] Nurulain S, Rum M, Yaakob R, and Affendey L S 2018 Detecting Influencers in Social Media Using Social Network Analysis (SNA) Int. J. Eng. Technol. 7(4) 950

[22] Vrana V, Kydros D, Kehris E, Theocharidis A I, and Karavasilis G 2019 A Network Analysis of Museums on Instagram Springer, Cham 1–10

[23] Kim S, Han J, Yoo S, and Gerla M 2017 How are social influencers connected in Instagram? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10540 257–264

[24] Trisminingsih R and Kurniawan R A 2019 Social Media Listening pada Instagram untuk Kasus Kebakaran Hutan di Indonesia Menggunakan Graph Clustering J. Teknol. Inf. dan Ilmu Komput. 6(2) 149

[25] Highfield T and Leaver T 2015 A methodology for mapping Instagram hashtags First Monday 20(1)

[26] Fatmasari, Kunang Y N, and Purnamasari S D 2019 Web Scraping Techniques to Collect Weather Data in South Sumatera Proceedings of 2018 International Conference on Electrical Engineering and Computer Science, ICECOS 2018

[27] Tahalea S P and SN A 2019 Central Actor Identification of Crime Group using Semantic Social Network Analysis Indones. J. Inf. Syst. 2(1) 24

[28] Antoniadis I and Charmantzi A 2016 Social network analysis and social capital in marketing: theory and practical implementation Int. J. Technol. Mark. 11(3) 344

[29] Tsai F C, Hsu M C, Chen C T, and Kao D Y 2019 Exploring drug-related crimes with social network analysis Procedia Computer Science

[30] Tahalea S P 2020 Identifikasi Peran Hero DOTA2 Menggunakan Social Network Analysis TEKNOMATIKA 12(2) 81–86

[31] Fu C, Chuang L, Jiahai Y, and Yuemei X 2016 A study on online social networks theme semantic computing model Proc. - 2016 IEEE Int. Conf. Web Serv. ICWS 2016 244–251

[32] Gómez S 2019 Centrality in Networks: Finding the Most Important Nodes Business and Consumer Analytics: New Ideas, Springer International Publishing 401–433

[33] Li H, Li W, and Tan J 2014 Modularity-based community detection in large networks: An empirical evaluation IEEE International Conference on Information and Automation, ICIA 2014 1131–1136

[34] Blondel V D, Guillaume J L, Lambiotte R, and Lefebvre E 2008 Fast unfolding of communities in large networks J. Stat. Mech. Theory Exp. 2008(10) P10008

[35] Djalante R et al. 2020 Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020 Prog. Disaster Sci. 6 100091