

# A Text Mining and Network Analysis of Tweets Related to Variants of COVID-19 on Twitter in Japan: Focus on the Delta Variant

Kazunori Minetaki

*Faculty of Business Administration, Kindai University, Japan*

## Abstract

This study examines how people's concerns, emotions, and behavior related to variants of COVID-19 are expressed on Twitter and its network structure in Japan. 3,523,728 tweets related to several variants were collected from May 9 to August 31, 2021. The Delta variant of COVID-19 caused the fifth wave of infections. The first case of the Delta variant was identified in Tokyo on May 15, 2021, spreading widely in Japan. The concern about the Delta variant on Twitter surged in May and June 2021. After that, the emotion of fear, anger, and condemnation on Twitter peaked in early August 2021. This anger was toward the Japanese government because their measurement to prevent invading the Delta variant was not enough and too late. The attitude of holding the Tokyo Olympics during the pandemic was criticized on Twitter.

**Keywords:** SARS-COV-2, COVID-19, variants, pandemic, Twitter, morphological analysis, network analysis

## 1. INTRODUCTION

The first cases of COVID-19, which was formally named SARS-COV-2, emerged in Wuhan, China, in late December 2019. The Chinese government quarantined the city of Wuhan on January 23, 2020. The World Health Organization named this disease SARS-COV-2 on February 11, 2020. Subsequently, many countries have been taking quarantine policies to prevent the spread of COVID-19 worldwide. Japan experienced the first wave of COVID-19 in April 2020 and successfully contained the first wave. The number of infections and deaths in Japan has been small compared to cases in the United States and several European countries.

First, this study examines how people's concerns, emotions, and behavior related to the variants of COVID-19 from April 2021 to August 2021 have been expressed on Twitter in Japan. The

infection and tolls of COVID-19 were worse than ever before in Japan. This study introduces how the Japanese think and feel, especially the Delta variant of COVID-19 compared to the cases of other countries. Even if the conspiracy theory is included in tweets, it is helpful for policymakers in each country to analyze tweets appropriately.

The Japanese government tries to prevent the spread of COVID-19 infections and restore socioeconomic activities simultaneously. This policy mix may confuse based on a misunderstanding (Minetaki, 2020). Preventative measures are sometimes implemented by the Japanese government and local governments inconsistently. The typical case is the Go To Travel Campaign. The Japanese government promotes the Go To Travel Campaign, which provides up to 50 percent on the costs for travel within Japan except for Tokyo. This campaign aims to promote domestic travel and boost local business.

Several prefectures in rural areas disagree with this campaign because of concerns about increasing infected persons.

This study focuses on the fifth wave caused by the Delta variant of COVID-19. Tweets related to the threat of the Delta variant began to increase in early May 2021. Considering that the shortage of hospitals became an issue in July and August in 2021, people quickly responded to the Delta variant on Twitter. Tweets announced that people had already been aware of the threat of the Delta virus before the breakdown of the medical system in Japan.

The risks of holding the Tokyo Olympics & Paralympics amid the COVID-19 pandemic were discussed in June 2021. The Japanese government has consistently taken measures to decline the COVID-19 pandemic and prepare for Tokyo Olympics & Paralympics. This study investigates how the Japanese people have felt such complicated situations during the COVID-19 pandemic in tweets.

Second, this study investigates the structure of the social network in Twitter. Not only the contents of tweets but also the topology of the social network influences how the information is transmitted.

Centrality metrics are calculated in several days when emotional tweets related to fear, anger, and condemnation surge. The results indicate that such emotional tweets go viral instantly in several days because of the network properties.

Therefore, the policymaker can predict people's behavior by recognizing what people tweet and the structure of the network on Twitter.

## 2. LITERATURE REVIEW

The spreading of COVID-19 misinformation on Twitter could confuse people. As WHO Director-General Tedros Adhanom Ghebreyesus proclaimed, "We're not just fighting an epidemic; we're fighting an infodemic. Misinformation can be defined as a "claim of fact that is currently false due to lack of scientific evidence" (Kouzy, 2020; Chou et al., 2018). The damaging misinformation spreads faster than the positive one. Misinformation on COVID-19 appears to be spreading rapidly on social media.

Misinformation can be identified by using several fact-check programs. Misinformation of COVID-19 are categorized as follows: (1) Cures, prevention and treatments, (2) Virus characteristics and numbers, (3) Vaccine and test kits, (4) Political measures, (5) Conspiracies, and (6) Others (Charquero-Ballester Marina et al., 2020).

Chen Emily et al. (2020) built the Twitter dataset, which helped track the COVID-19 misinformation and enable understanding of fear and panic emotions. Kouzy et al. (2020) indicated that the ratio of tweets including misinformation to total tweets is 24.8 %, and 17.4% including unverifiable information regarding damaging the COVID-19 epidemic. Online platforms communications have included various conspiracies. Ahmed Wasim et al. (2020) developed an understanding of the drivers of the 5G COVID-19 conspiracy theory and strategies to deal with such misinformation. Ahmed Wasim et al. (2020) concluded 34.8 % of total tweets contained views that 5G and COVID-19 were linked. Twitter has already introduced warning messages when tweets contain contents that provide disputed or misleading information related to COVID-19. This study also adequately considers the possibility that tweets include misinformation.

This study refers to Alaa Abd-Alrazaq et al. (2020) to classify words collected by Twitter API. Alaa Abd-Alrazaq et al. (2020) identified the main topics, which are (1) origin of the virus, (2) its sources, (3) impact on people & countries & economy, and (4) the ways of mitigating the risk of infection, posted by Twitter users related to the COVID-19 pandemic.

Sentiment analysis can describe people's emotions on Twitter. Boon-Itt et al. (2020) investigated sentiments of (1) anger, (2) anticipation, (3) disgust, (4) fear, (5) joy, (6) sadness, (7) surprise, (8) trust. Manguri et al. (2020) conducted sentiment analysis on worldwide COVID-19 outbreak categorized by (1) happy and joy, (2) confident, (3) optimistic, (4) hopeful, (5) calm and content, (6) neutral and relaxed, (7) relieved, (8) pessimistic and impatient, (9) worry and boredom, (10) discouraged and difficulty, (11) depressed and fear. Lwin et al. (2020) examined worldwide trends of four emotions: fear, anger, sadness, and joy on tweets.

The view of mental health is also essential to

recognize the effects of COVID-19. Ahmad Araz Ramazan et al. (2020) determined how social media affected self-reported mental health and the spread of panic about COVID-19 in the Kurdistan Region of Iraq by conducting an online questionnaire.

As for the early stage of this pandemic, Su Yue et al. (2020) compared the cases of Wuhan, China, and Lombardy, Italy, where cities were locked down in the early stage of the COVID-19 pandemic and mentioned that Weibo users and Titter users are popular respectively in Wuhan, and Lombardy.

### 3. MORPHOLOGICAL ANALYSIS OF TWEETS RELATED TO COVID-19

3,523,728 tweets were collected from May 9 to August 25 in 2021 by leveraging Twitter's streaming API and Tweepy (python library). The query on Twitter was "Indian strain OR delta variant OR lambda variant."

Morphological analysis was conducted to find words highly frequently<sup>1)</sup> tweeted related to those keywords. Noun, verb, and adjective extracted from 3,523,728 tweets by morphological analysis were

**Table 1: Results of morphological analysis**

1. virus	2. infection	3. vaccine	4. medicine (excluding vaccine), hospital, doctor	5. symptoms, side effects
corona virus alfa beta gammer delta lambda kappa hybrid variant mutation spike	infection cluster outbreak pandemic rampant explotion community spread replace positive	vaccine Pfizer neutralization buster	steroid antibody cocktails antifebrile hospital hospital bed healthcare center doctor home doctor nurse pediatrics	asthma cough pneumonia thrombus infarction fever emesis fatigue complicating disease aftereffects side effects anaphylaxis symptom serious illness inflammation taste
6. prevention	7. death	8. fear, anger, condemnation	9. politics, government	10. Olympics, Paralympics
prevention isolation quarantine thermometry examination ventilation hand-washing disinfection mask shield lockdown time shorting restraint edge of the water many loopholes	death die passing away sudden death obituary funeral suicide	fear terrible terrifying scare dangerous anger angry demonstration condemnation riot	poilitics government cabinet official residence prime minister minister Liberal Democratic Party Komeito governor Governor of Tokyo Metropolis	Olympics Paralympics IOC Bach

classified into ten categories which are (1) virus, (2) infection, (3) vaccine, (4) medicine (excluding vaccine), hospital, doctor symptom, medical treatment, and medical resource, (5) prevention, (6) death, and serious condition, (7) symptoms, side effect, (8) fear, anger, condemnation (9) politics, government, (10) Olympics, Paralympics. Those categories are based on the previous studies, such as adding (10) Olympics, Paralympics which is the specific situation in Japan. This study included retweets data, and therefore reaction by people on Twitter was possibly more significant than the case without retweets.

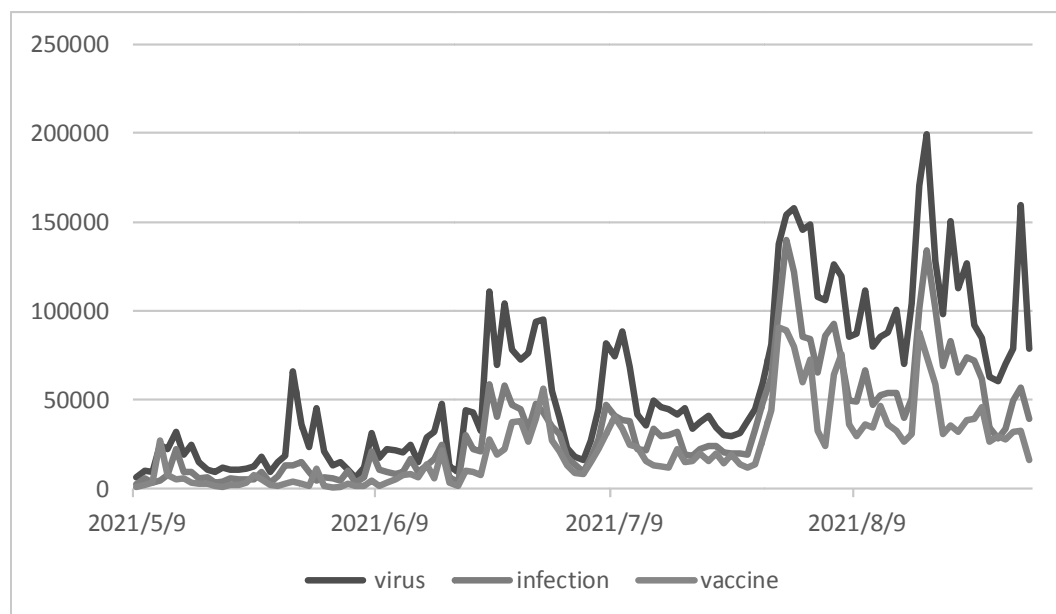
The results of morphological analysis and classification of 10 categories are summarized in each period (Table 1<sup>2)</sup>). The results of morphological analysis are shown in Figure 1–5.

The virus, infection, and vaccine trends have increased, and the increment of the virus was more significant than others (Figure 1). The number of words related to medicine (excluding vaccine), hospital, and a doctor has not changed sharply until early August and leaped at the about middle in August 2021 (Figure 2). Both symptoms & side drug adverse reactions and prevention have increased gradually, and their trends began to decline after early August (Figure 3). Prevention

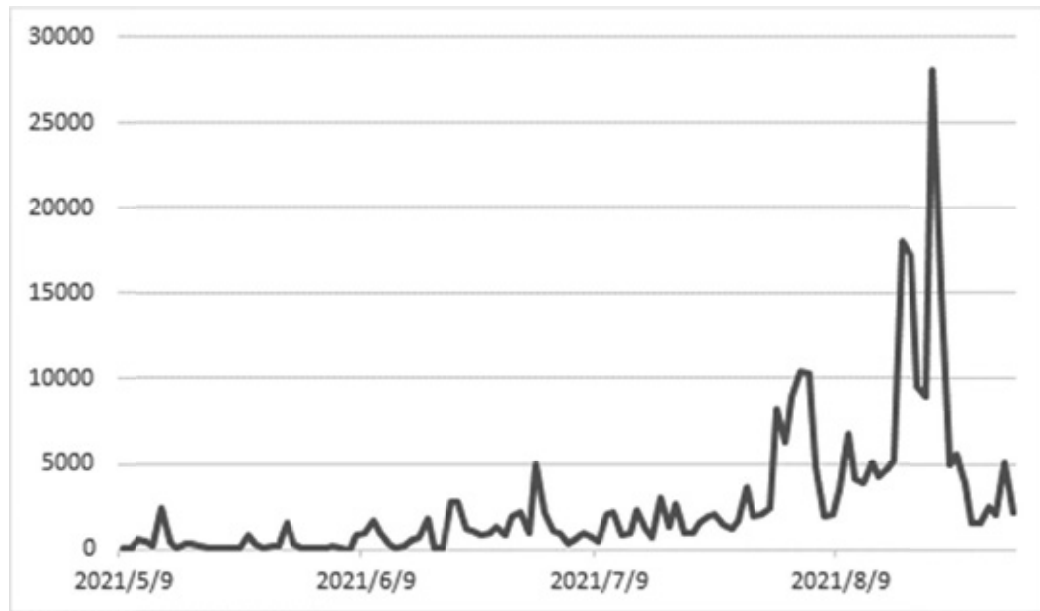
has soared in early July and August (Figure 3). Fear, anger, and condemnation have jumped in early July and early August (Figure 4). The first wave of death was observed in the latter half of June, and subsequent death soared from the latter half of July (Figure 4). There were three peaks in politics and government, and the peak of the Olympics & Paralympics was the latter half of June (Figure 5).

The delta and lambda variants are more transmissible than pre-existing variants of COVID-19 and are likely to cause severe symptoms in elderly persons and the younger generation. The background and tweets of symptoms have increased, and fear, anger, and condemnation surged rapidly. Many infected persons could not appropriately care in July and August 2021, and it was recognized as a social problem. People's emotions on Twitter reflected on this issue.

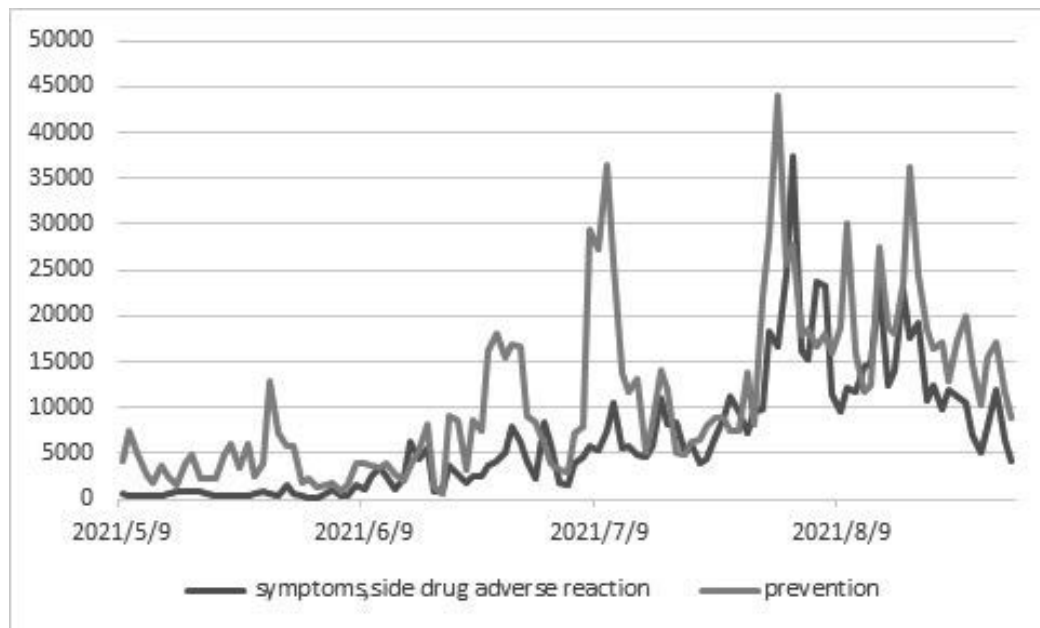
Many tweets that TV media inflamed people on Twitter to fear the delta virus and the vaccine would fully contain the pandemic were seen on May 11, 2021. Tweets on July 10, 2021, which recorded the highest of fear, anger, and condemnation, repeatedly quoted the CNN news that the Tokyo Olympics had no spectators due to the city's coronavirus state of emergency through the Games the Japanese Olympic Committee. Also, many tweets related



**Figure 1: Transition of virus, infection, and vaccine**



**Figure 2: Transition of medicine (excluding vaccine), hospital, and doctor**



**Figure 3: Transition of symptoms and side drug adverse reaction**

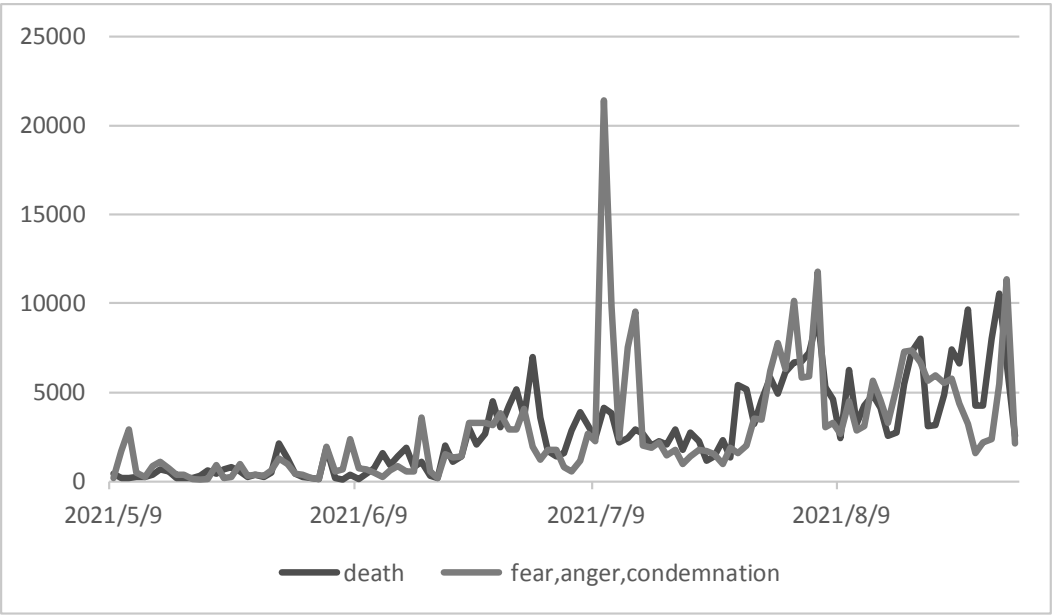


Figure 4: Transition of death, and fear, anger, and condemnation

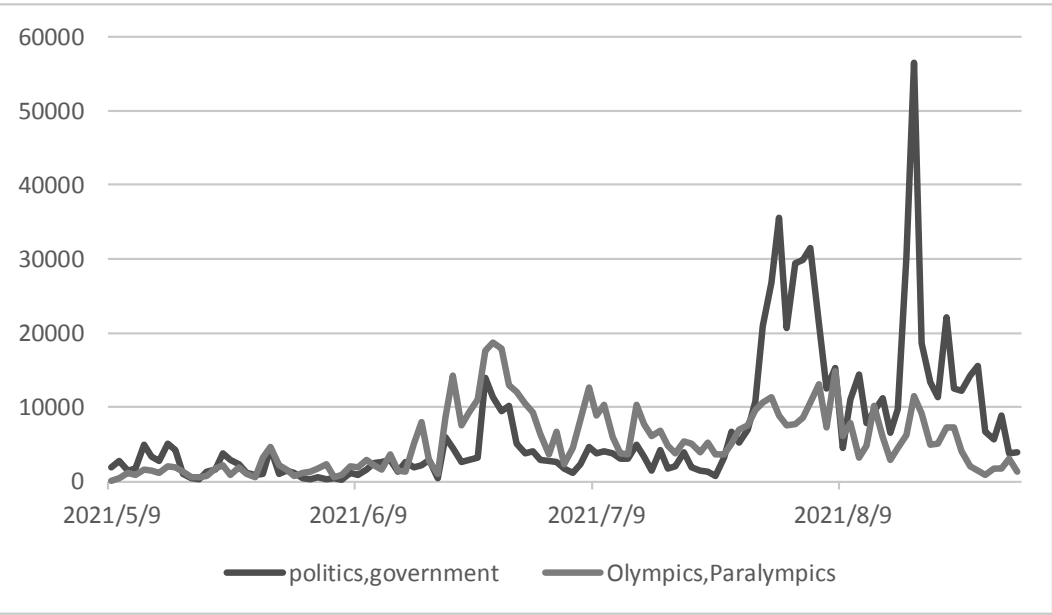
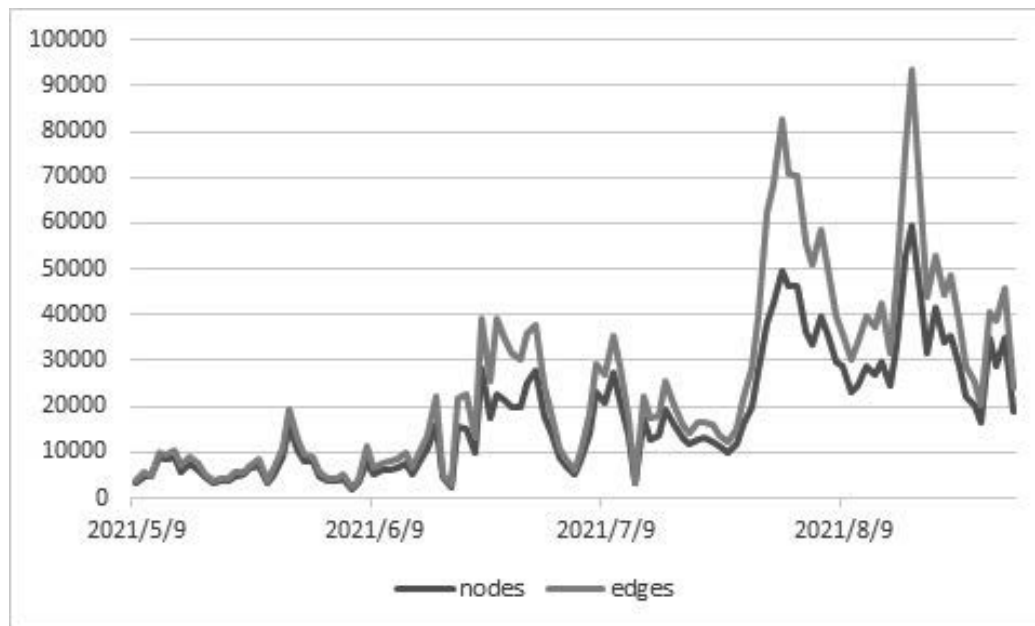
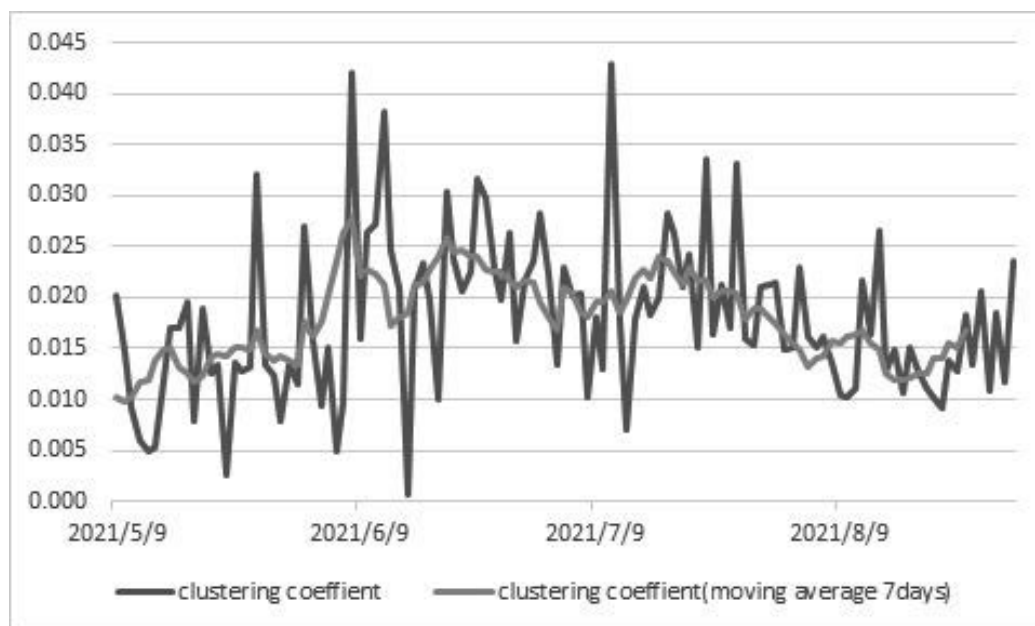


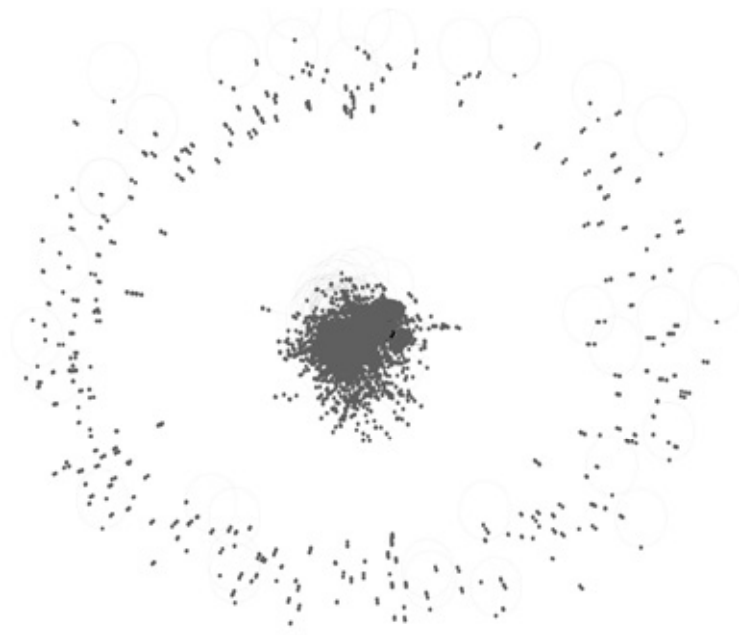
Figure 5: Transition of politics & government, and Olympics & Paralympics



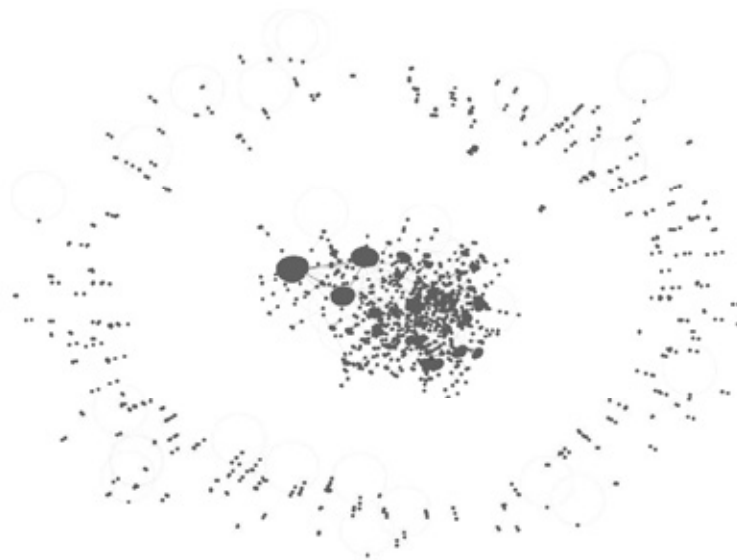
**Figure 6: Network analysis: nodes, edges**



**Figure 7: Network analysis: Clustering coefficient**

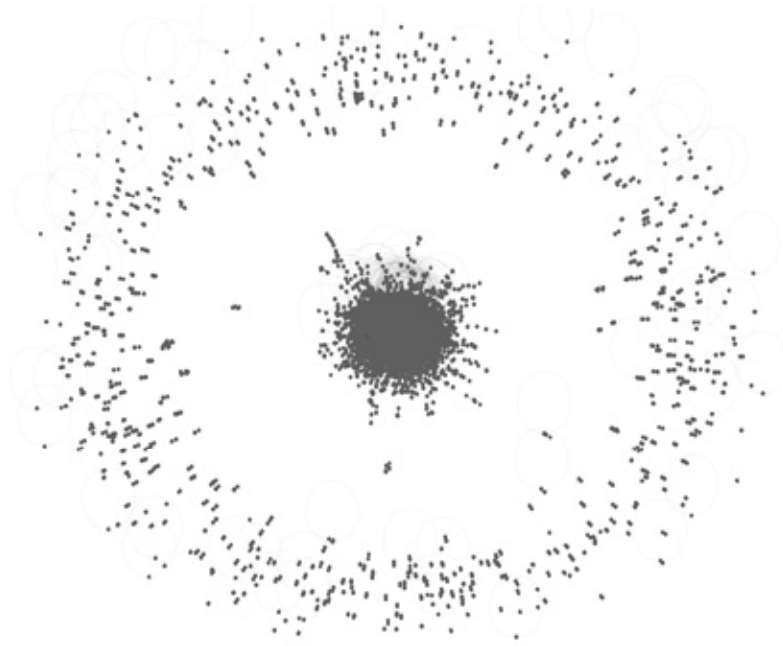


**Figure 8: Network of May 11, 2021**



**Figure 9: Network of July 10, 2021**





**Figure 10: Network of August 6, 2021**

to the anger and condemnation for the reports released by the committee of minister of Health, Labor, and Welfare were repeatedly observed on August 6, 2021.

#### 4. NETWORK ANALYSIS

Retweet<sup>2)</sup> data can provide the networks on Twitter, and therefore, the daily network of tweets & retweets from May 9 to August 31, 2021, were analyzed. Network analysis is one of the graphical theories. A network structure consists of nodes and edges. Here, nodes represent objects we will explore, while edges represent the relationships between those objects. The transitions of nodes and edges were parallel (Figure 6). The number of edges jumped in the first half of August 2021.

The clustering coefficient is the index that shows the density of the network.

$$C_i = \frac{\text{Number of triangles including node}_i}{k_i(k_i - 1)/2}$$

$k_i$ : number of edges of node<sub>*i*</sub>

$$C = \frac{1}{n} \sum_{i=1}^n C_i$$

The moving average of clustering coefficient in 7 days started to increase at the beginning of June and showed a downward trend from early July (Figure 7). It can be interpreted that the density of the network was relatively high from June to early July and became relatively low.

This study focuses on the transition of fear, anger, condemnation, and therefore its several peaks, specifically May 11, July 10, and August 10, 2021, were highlighted. Figure 8–10 describes the retweet networks of those days. Figure 8 seems to be different from Figure 9 and 10. Accumulations of nodes in the central area and the periphery in Figure 9 and 10 are more than those in Figure 8.

Centrality is an indicator of the connectivity which a node has. This study calculates degree, closeness, betweenness for centrality (Figure 11–19).

The degree is the simplest centrality which counts how many edges each node has. Degree centrality is calculated more sophisticatedly to divide the number of edges by the number of nodes to standardize.

Degree centrality is formulated as follows.

$$DC_i = \frac{k_i}{n-1}$$

$n$ : number of nodes

Closeness centrality indicates how close a node is to all other nodes in the network. Closeness centrality is formulated as follows.

$$CC_i = \frac{n-1}{\sum_{j=1}^{n-1} d(i,j)}$$

$d(i,j)$ : the minimum distance between  $node_i$  and  $node_j$

Betweenness centrality measures how often a node is in the shortest path between two other nodes in the network. This measure describes which nodes are ‘bridges’ between nodes in a network.

Betweenness centrality is formulated as follows.

$$BC_i = \sum_{s=1(s \neq i)}^n \sum_{t=1(t \neq s,i)}^n \frac{\sigma_{st}(i)}{\sigma_{st}}$$

$\sigma_{st}$ : number of shortest paths

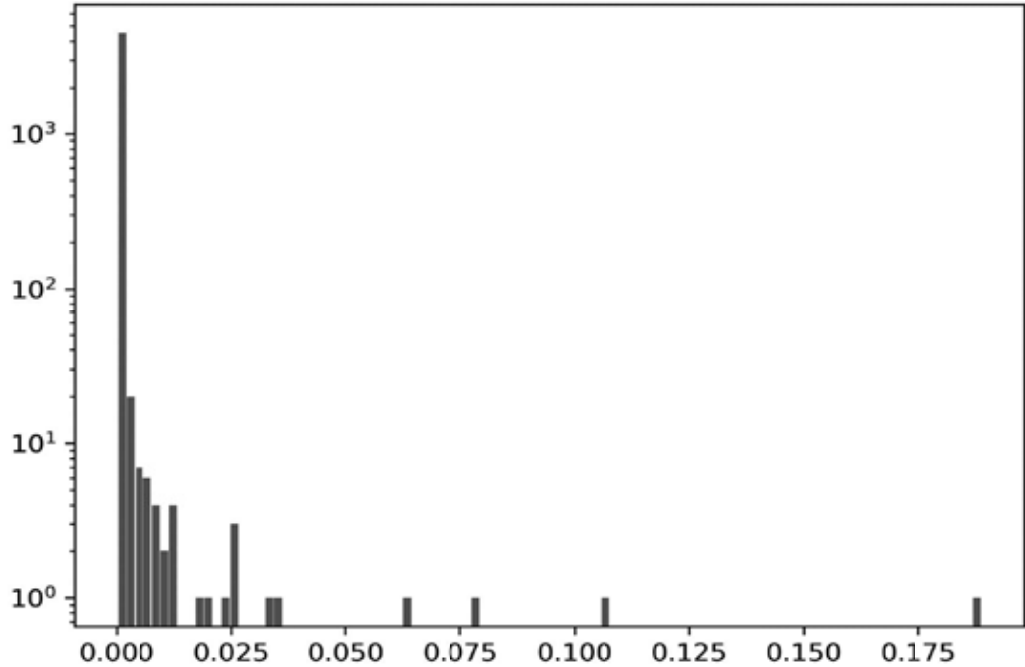
$\sigma_{st}(i)$ : number of paths passing through some  $node_i$

Degree, closeness, and betweenness of the centrality of May 11, 2021, are described in Figure 11–13. Figure 14–16 are for July 10 and Figure 17–19 for August 6. As for degree centrality, the case of July 10.

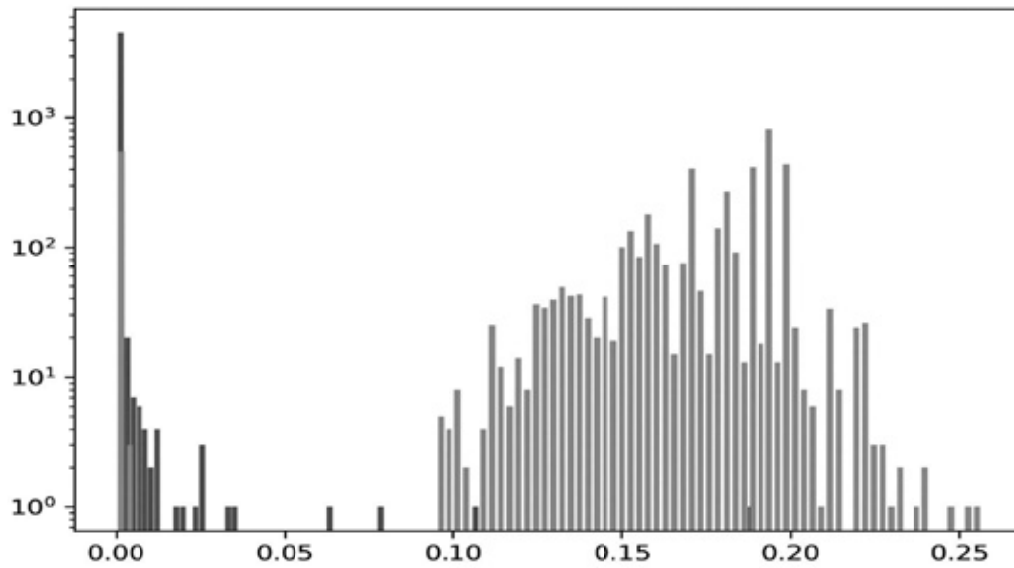
Degree centrality can identify the hub of nodes, and therefore the network where the average degree centrality is comparatively high suggests the existence of hubs connected with many nodes. It seems that many nodes have a few edges (Figure 11, 14, 17). The case of July 10 indicates that nodes of high degree centrality exist (Figure 14). The peak of the tweets related to fear, anger, and condemnation was July 10, 2021. The person who had the highest degree of centrality tweeted the fear of the anti-vaccine group without masks on July 10, 2021.

Closeness centrality describes the distance among nodes, and it can be considered a concept of the size of a network. The shape of a graph of closeness centrality is that nodes of closeness centrality close to zero exist; on the other hand, many nodes have higher values.

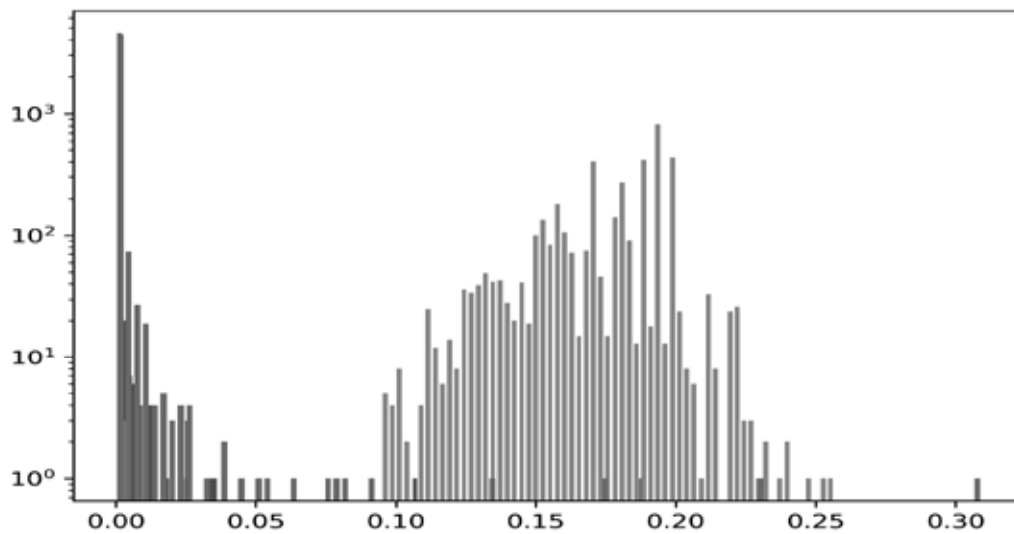
Comparing three cases of July 10, August 6, and May 11, 2021, two instances of July 10 August 6 show higher closeness centrality than the case of



**Figure 11: Network analysis: Degree centrality, May 11, 2021**



**Figure 12: Network analysis: Closeness centrality, May 11, 2021**



**Figure 13: Network analysis: Betweenness centrality, May 11, 2021**

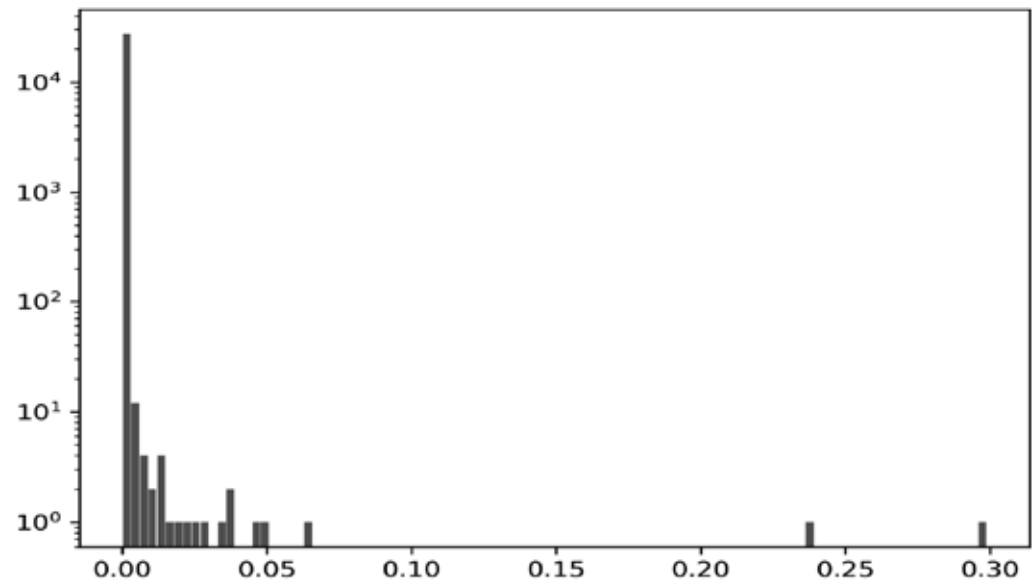


Figure 14: Network analysis: Degree centrality, July 10, 2021

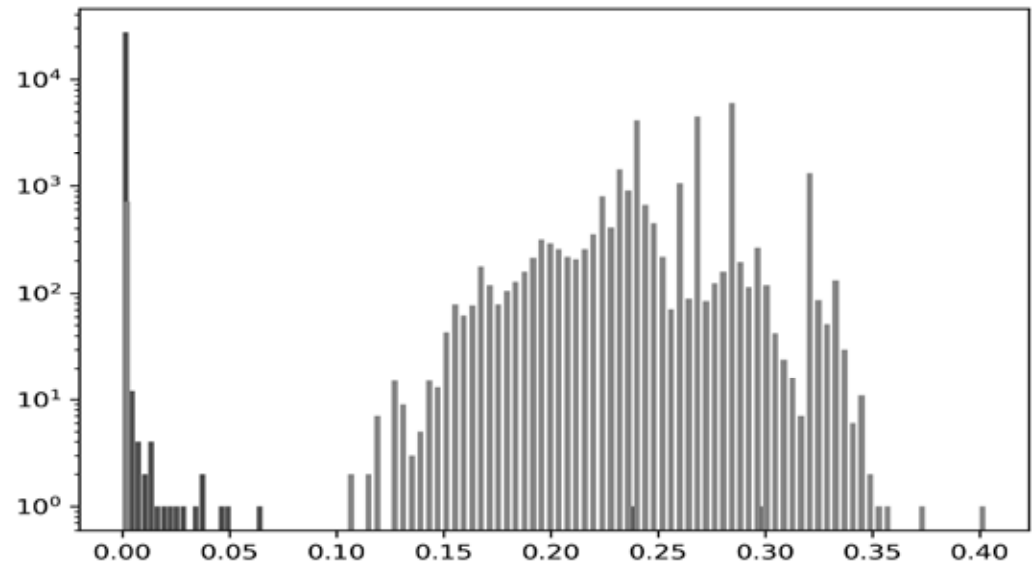
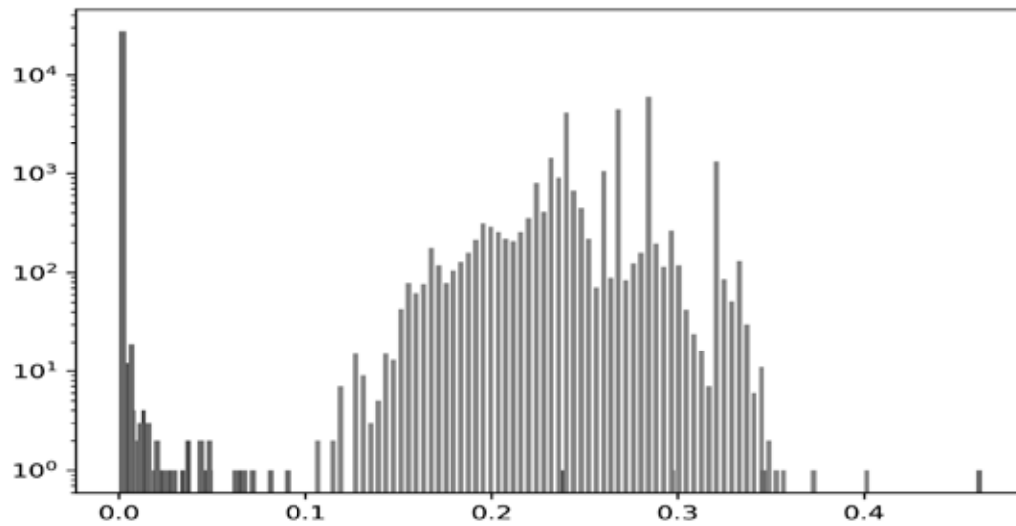
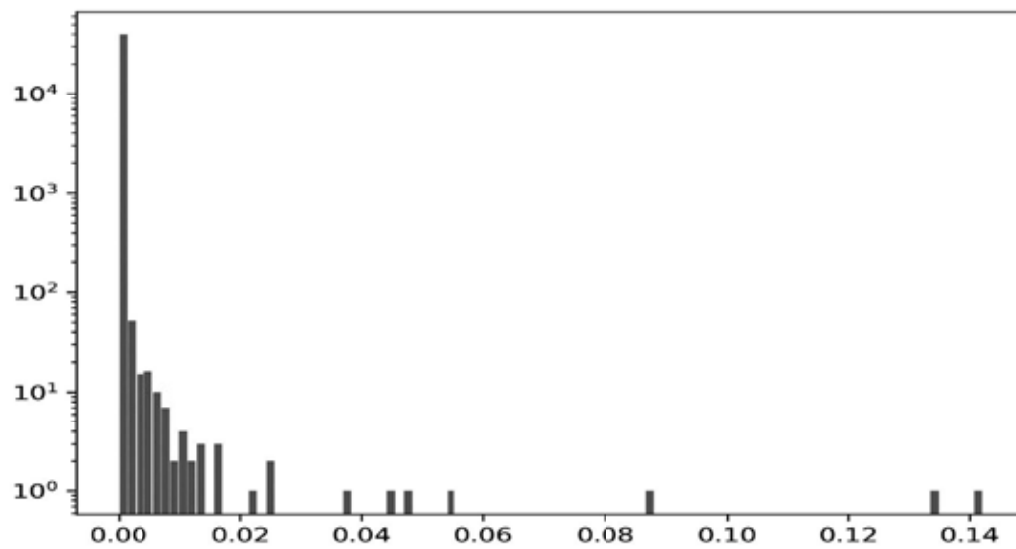


Figure 15: Network analysis: Closeness centrality, July 10, 2021



**Figure 16: Network analysis: Betweenness centrality, July 10, 2021**



**Figure 17. Network analysis: Degree centrality, August 6, 2021**

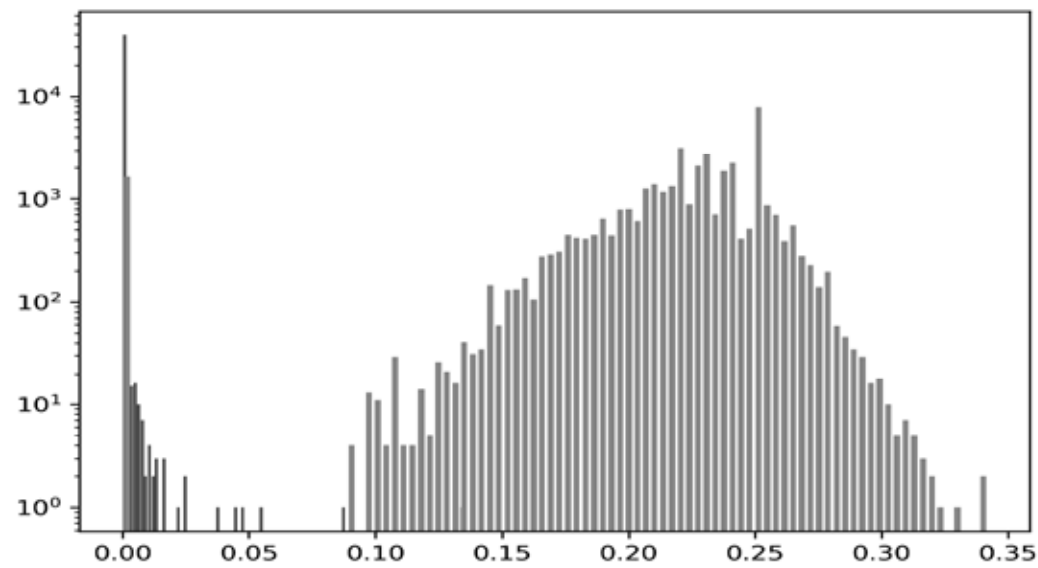


Figure 18: Network analysis: Closeness centrality, August 6, 2021

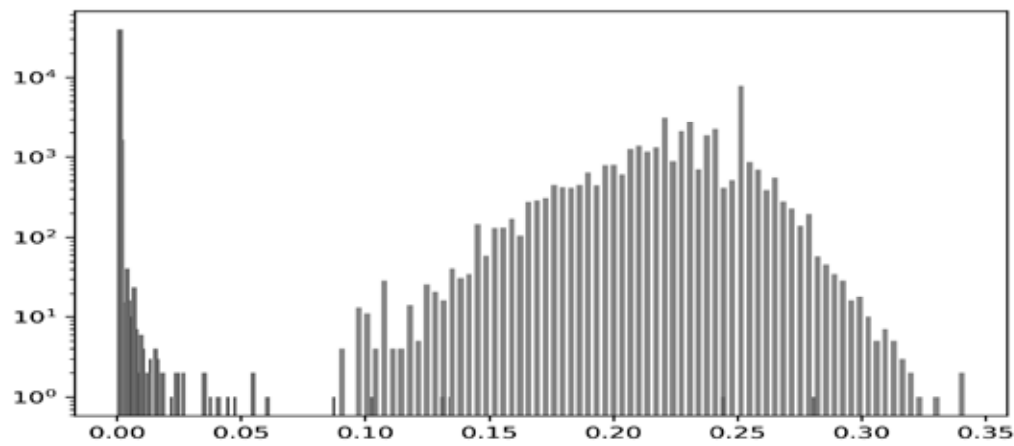


Figure 19: Network analysis: Betweenness centrality, August 6, 2021

May 11. It possibly implies sharing information would be easier.

The graph of betweenness centrality was shaped by two kinds of accumulations of nodes and closeness centrality. The July 10 and August 6, 2021 cases showed higher betweenness centrality than May 11, 2021. It indicates that many bridges to connect other networks could make it easy to share information.

The network analysis of centrality suggests that emotional tweets would go viral instantly in the networks of July 10 and August 6, 2021. Mainly the same nodes were top three centralities for the degree, closeness, and betweenness on July 10, 2021. It meant that specified nodes were the hubs in the July 10, 2021 network and connected to other networks.

## 5. CONCLUSION

The issues of delta variant have been discussed in early May 2021 on Twitter. The delta virus is more transmissible than pre-existing variants of COVID-19. Severe symptoms and many infected persons who could not be hospitalized because of the capacity of the Japanese healthcare system have been seen in July and August 2021. This study can examine the reaction of Twitter to such difficult situations in Japan. Tweets including words of virus, infection, vaccine, symptoms have gradually increased. The peak of fear, anger, and condemnation tweets was early July 2021. The fear for the delta variant and the antipathy to the attitude of the Japanese government were observed on Twitter.

The structure of the network influences how the information will be transmitted. A large number of degrees indicated the hub in the network. The short distance in the network and the many bridges among nodes could be the factors that promote the transmission of information.

This study investigated the words of fear, anger, and condemnation, focusing on the three peaks of May 11, July 10, and August 8. This study took up those three peaks because the case of May 11 was the first peak to be thought to reflect on the initial emotional change to the delta virus, and the instances of July 10 and August 6 showed a considerable increase where people faced severe medical

conditions. The handle names with high centrality of degree, closeness, and betweenness were identified. This study found the persons who were the influencers.

People repeatedly tweeted that TV media incited people on Twitter to fear the delta virus on May 11, 2021. Many tweets that the Tokyo Olympics had no spectators due to the city's coronavirus state of emergency, and the fear of no mask demonstration in Sapporo city was seen on July 10, 2021. The anger and condemnation for the shorthand released by the minister of Health, Labor, and welfare committee because the risk of delta variant could be predicted before several months was repeatedly observed on Twitter, August 6, 2021.

Twitter is a valuable communication tool to suggest people's emotions and behavior. This study concludes that tweets announced that people had already been aware of the threat of the delta virus before the breakdown of the medical system in Japan. With the misinformation, policymakers can utilize Twitter to know what citizens fear and criticize amid the COVID-19 pandemic. Network analysis shows how the information will spread and indicates a significant influence.

Though this study focused on three days from the transition of fear, anger, and condemnation, further investigation of other days by investigating the centrality of network analysis is needed to predict the following critical issues.

## NOTES

- 1) 2000 words by day were extracted. The morphological analysis focus on the words frequently used, not ensuring to grasp the whole words related to the keywords.
- 2) The categories and words to pick up are different from the tables in Minetaki (2020) and Minetaki (2021) because Minetaki (2021) covered COVID-19 as a whole. For example, "ventilation" was included in Minetaki (2020) and Minetaki (2021), not in this study.
- 3) This study used the dataset, including retweets. Retweets provide the relations among nodes, and therefore the network analysis can be conducted. The characteristics of tweets and

retweets are different in the time and effort required. Minetaki (2020) and Minetaki (2021) focused on tweets only.

## REFERENCES

- Abd-Alrazaq Alaa, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, Zubair Shah (2020), "Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study," *Journal of Medical Internet Research*, 22(4).
- Ahmad Araz Ramazan, Hersh Rasool Murad (2020), "The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study," *Journal of Medical Internet Research*, 22(5).
- Ahmed Wasim, Josep Vidal-Alaball, Joseph Downing, Francesc López Seguí (2020), "COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data," *Journal of Medical Internet Research*, 22(5).
- Bavel Jay J Van, Katherine Baicker, Paulo S Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J Crockett, Alia J Crum, Karen M Douglas, James N Druckman, John Drury, Oeindrila Dube, Naomi Ellemers, Eli J Finkel, James H Fowler, Michele Gelfand, Shihui Han, S Alexander Haslam, Jolanda Jetten, Shinobu Kitayama, Dean Mobbs, Lucy E Napper, Dominic J Packer, Gordon Pennycook, Ellen Peters, Richard E Petty, David G Rand, Stephen D Reicher, Simone Schnall, Azim Shariff, Linda J Skitka, Sandra Susan Smith, Cass R Sunstein, Nassim Tabri, Joshua A Tucker, Sander van der Linden, Paul van Lange, Kim A Weeden, Michael J A Wohl, Jamil Zaki, Sean R Zion, Robb Willer (2020), "Using social and behavioural science to support COVID-19 pandemic response," *Nature Human Behaviour*, 4.
- Boon-Itt Sakuma, and Yukolpat Skunkan (2020), "Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study," *JMIR Public Health and Surveillance*, 6(4).
- Chen Emily, Kristina Lerman, Emilio Ferrara (2020), "Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set," *JMIR Public Health and Surveillance*, 6 (2).
- Kouzy Ramez, Joseph Abi Jaoude, Afif Kraittem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, Khalil Baddour (2020), "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter," *Cureus*, 12(3).
- Lwin May Oo, Jiahui Lu, Anita Sheldenkar, Peter Johannes Schulz, Wonsun Shin, Raj Gupta, Yinping Yang (2020), "Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends," *JMIR Public Health Surveill*, 6(2).
- Minetaki Kazunori (2021), "Analysis of words related to the COVID-19 on Twitter—Focusing on the declaration of the second state of emergency," *Journal of Business Studies*, 67(3).
- National Institution of Infectious Diseases (2021), "Current Situation of Infection, August 18, 2021."
- Peter Johannes Schulz, Wonsun Shin Raj Gupta, Yinping Yang (2020), "Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends," *JMIR Public Health and Surveillance*, 6(2).
- Su Yue, Jia Xue, Xiaoqian Liu, Peijing Wu, Junxiang Chen, Chen Chen, Tianli Liu, Weigang Gong, Tingshao Zhu (2020), "Examining the Impact of COVID-19 Lockdown in Wuhan and Lombardy: A Psycholinguistic Analysis on Weibo and Twitter," *International Journal of Environmental Research and Public Health*, 17.

Dr. Kazunori Minetaki is professor of the Faculty of Business Administration and the Institute for Creative Management and Innovation, Kindai University, Japan. Email: kminetaki@bus.kindai.ac.jp