

How did Twitter Users React to the Pandemic? Social Network Analysis of Public Tweets on CoViD-19 Outbreak

Asia Pacific Journal of
Multidisciplinary Research
Vol. 8 No.3, 52-59
August 2020
P-ISSN 2350-7756
E-ISSN 2350-8442
www.apjmr.com
ASEAN Citation Index

Jasten Keneth D. Treceñe¹, Ralph Jerico P. Abides²

¹Eastern Visayas State University – Tanauan Campus, ²Visayas State University – Tolosa Campus, Philippines
jastentrecene5@gmail.com¹, abidesr@gmail.com²

Date Received: May 7, 2020; Date Revised: July 15, 2020

Abstract –Several researchers have presented several studies on the CoViD-19 outbreak like on the epidemiological aspects of the disease, diagnostics method of the novel coronavirus, clinical characteristics, transmission, and vaccines. However, the sentiments and behaviour of the people online particularly in twitter remain unexplored. In this paper we focused on exploring peoples' tweets to uncover their attitudes, sentiments, and find out the network effects of peoples' tweets and the heated topics. Text mining approach was utilized using sentiment and social network analysis. Term document matrix, word cloud, nrc_sentiment dictionary, histogram, community edge betweenness algorithm, and network graph were used in the study. An API account was created wherein 15000 tweets were extracted from March 22, 2020 to March 31, 2020 containing the keyword #COVID-19 to make a working data for analysis. Results from the social network analysis showed a close relationship between tweets where people are globally talking part by sharing information about the CoViD-19. The peoples' attitude showed the willingness to follow government precautionary measures to lessen the impact of the virus. Despite of the fear and sadness felt by the people over twitter, sentiment analysis revealed positive emotion towards the crisis. Such insights are significant when guiding people to respond appropriately and helping them to learn to cope with the sudden infectious disease as it promotes social stability. This will also help the authorities understand the sentiments and anxieties of the people, giving a strong direction to enact policies beneficial to the people. Moreover, social network analysis can be used as a method of understanding the behaviour of the people online and how these people are talking towards an issue.

Keywords –Social Network, Sentiment Analysis, Text Mining, covid-19, coronavirus

INTRODUCTION

As the new coronavirus outbreak hits the world and people need to stay at home to avoid of being infected with the virus and lessen the spreading of the contagion, thus where most of the conversations are taking place online. People take the opportunity of using internet to share information, raise their concerns, and consume most of their time in the internet while in quarantine. The time when those online discussions light up also tell us a lot about how their feelings around the pandemic are growing. With the advent and the rapid growth of technology, there has been a considerable change on the information landscape and information-consumption of the people [1]. Discussion of the CoViD-19 has been flooded across various social media platforms as reported by media analytics [2].

At the early stage, literatures emphasized that we still have limited data about the outbreak, this can be found in the study of Fong where they only have a

handful of datasets to develop a model. This is also because we still have few studies about the disease [24]. It is important to know the peoples' sentiments to CoViD-19 during the current situation. Such insights are significant when guiding people to respond appropriately and helping them to learn to cope with the sudden infectious disease as it will also promote social stability. Furthermore, this study will also help authorities to know peoples' worries and anxieties, having them a strong direction and ratify new policies helpful to the people. This study used various text mining techniques and algorithms mainly, sentiment analysis using the nrc_sentiment dictionary in R and network analysis using community edge betweenness.

The rise of the discussions of the corona virus online has been followed as the pandemic has been infecting more and more people around the world. Sprinklr, a media analytics noted that several emotions were also been expressed online based on the emojis that most

commonly associated with the corona virus. Peoples concern on the corona virus is evidently seen in the search history of Google. An exponential increase of search terms relating to corona virus has been recorded. With the different analytics presented by Sprinkl and google trends, twitter has been also widely used by many people and as basis also for understanding the most talked issue around the world. Twitter has developed rapidly in recent years, increasing number of public individuals are already using this social media platform to communicate, share information, and raise their concerns and opinions towards a specific issue. Twitter has become an important channel for promoting risk communication during crisis [3-4]. The use of social media particularly twitter to measure public attention has also been gradually applied in research on infectious diseases [3]-[7], [9].

Currently, the world is experiencing a Corona virus (CoViD-19) outbreak and has now spread to more than 50 countries [10]. It was already declared by WHO as a pandemic and a Public Health Emergency of International Concern (PHEIC). With the onset of CoViD-19, many people are turning to twitter to assess the severity of the situation, raise concerns to the current condition, and to the government policies and actions. Presently, various text mining techniques particularly sentiment and social network analysis has become an important tool for understanding people's behaviour online and come up with meaningful insights from them [3]-[4]-[5]. Various researchers made efforts in different aspects to fight against CoViD-19 and promote the prevention and mitigation of the pandemic like on the epidemiological aspects of the disease, diagnostics method of the novel corona virus, clinical characteristics of the disease, disease transmission, and virus vaccines [11]-[13]. A study on the CoViD-19 outbreak submitted to the bulletin of World Health Organization used the predictive modelling approach to forecast CoViD-19 outbreak within and outside China based on daily observation [14]. They also analysed the sentiments from news articles and classify these articles based on the polarity, this is also to understand the influence of the news to the behaviour of the people, politically and economically.

Pastor [15] also studied on the sentiments to the CoViD-19 pandemic. Both qualitative and quantitative method was used in the study with the application of sentiment analysis. However, the study was just limited to only a specific group of people, similar to the research who also studied in the CoViD-19 outbreak where they analysed the sentiments of Chinese from the

extracted data in a microblog hot search list [5]. But the study focused only on a microblog wherein more people are accessing other social media platform like twitter. In this paper we focused on peoples' tweets around the world from March 22, 2020 to March 31, 2020 wherein people express more of their opinions in the site.

The rest of the paper is structured as follows: Section II presents the objectives of the study, section III elucidates the methodology of the study, including the research approach used, the data collection, the research process and the data analysis, section IV illustrates the results of the study, it also includes the data exploration, sentiment analysis, and network analysis, section V provides the summary of findings and the discussions, section VI summarizes the conclusions, and finally section VII presents the limitations of the study and future works..

OBJECTIVES OF THE STUDY

The propagation of social media usage for discussion of opinions and feelings by the public has created possibilities of analyzing such sentiments about any prevalent discourse. This study analyzed the sentiments and attitudes about the CoViD-19 pandemic expressed globally over twitter. Specifically, this study explored the data towards peoples' attitude on the CoViD-19 pandemic, analyzed and presented the sentiments of people towards the pandemic, and identified how people's tweets are closely connected to each other using community detection algorithm, identified the most influential words inside the graph using the measures of betweenness centrality and degree, and find out the heated topics.

MATERIALS AND METHODS

Research Approach

To achieve the objectives of the study, the researchers used the text mining techniques such as sentiment analysis (SA) and network analysis. The text mining area has been widely used in computer science which adopts the concepts of natural language processing, knowledge management, data mining, and machine learning [16]. It explores interesting patterns from the useful unstructured data that has been extracted [17].

Data Collection

The tweets were extracted into a working data for analysis using R programming. Also, an API account on tweeter was created first to allow us to harvest tweets. We extracted 15000 tweets from different tweeter users

globally from March 22, 2020 to March 31, 2020 containing the keyword #COVID-19 for the website. Replies and retweets were not included. The data consists of 16 columns and 15000 rows where it contains the tweet, followed by information such as the like engagement, time, user id, and etc.

Research Process and Data Analysis

The research process includes of extracting first the tweets, after, the data will undergo on the pre-processing and data cleaning stage, then will go on the process of analysing the sentiments and network analysis together with data visualization.

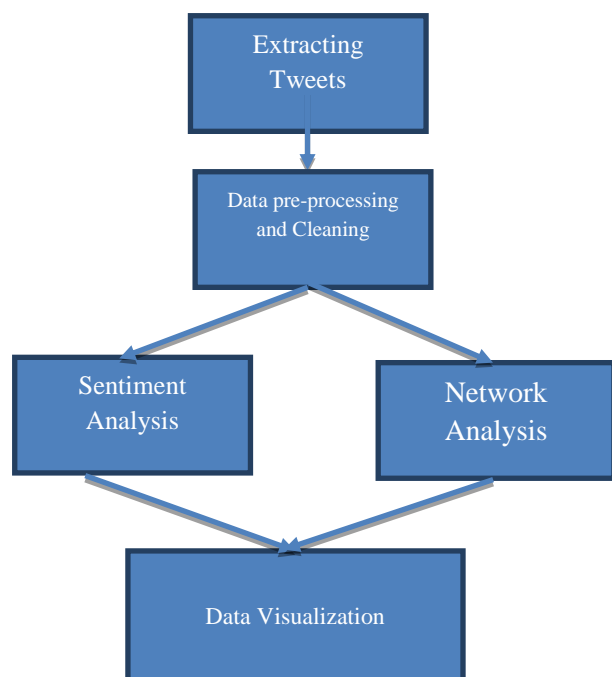


Fig. 2. Research process

Data Pre-processing and Cleaning

The data set is transformed into a corpus, a corpus is a group of text known in R. Then, the corpus, was pre-processed using tokenization and text normalization. This stage is very important when dealing with large amount of data.

Tokenization—In this stage of pre-processing, all the characters were transformed into lowercase, punctuations and numbers were removed, English stop words and white spaces were also removed. Moreover, uniform resource locator, emojis, and unnecessary words were also removed such as names mentioned in tweets.

Text normalization – before further processing of the text, it needs to be normalized. It is generally

referring to allowing the words on equal footing and allows the processing to continue uniformly. Two tasks were used to normalize the text such as text stemming and lemmatization. In the stemming process, words like need, needed, and needing were stemmed to the word “need”. In the lemmatization part, words like corona virus, ncov and virus were transformed into its citation form to “covid”. This idea is used to reduce the distinct number of words in the corpus that will improve the analysis.

Sentiment Analysis

After the pre-processing stage, sentiment analysis was done to reveal the emotions behind people’s tweets. Sentiment analysis (SA) is a natural language process that creates meaningful information out of the textual data [18]. The technique was used to identify the emotions expressed by the people from the tweets. These emotions focused on eight emotions such as trust, joy, sadness, fear, anger, surprise, disgust and anticipation[18]. To obtain the sentiment scores of the tweets, “nrc_sentiment” dictionary was used to calculate the presence of eight emotions and their corresponding valence. The sentiment analysis helped to learn individual’s emotion and attitudes to the CoViD-19 outbreak.

Network Analysis

In text network analysis, a text is represented as graph. It helps identifies relationships of text in social media platforms. The words are the nodes and co-occurrences of the words are the connections between them [19]. Then, the community detection algorithm was used on the constructed graph to identify the groups of nodes that are more densely connected to one another than to the rest of the network as well as the most influential words inside the graph using the measures of betweenness centrality and degree.

To avoid the messy display of the data, we only cover terms that appeared more than 30 times in the text. The connected terms are those that appear together on Twitter. Then edge betweenness was utilized to cluster all the words. Betweenness signifies how recurrently a node is between other nodes’ paths. Edge betweenness is the number of shortest paths that go through an edge in a network graph [20].

Data Visualization

Different data visualization techniques were used to obtain the objectives of the study. First, to know the attitude of the people to the CoViD-19 pandemic, the

Network Analysis of Tweets

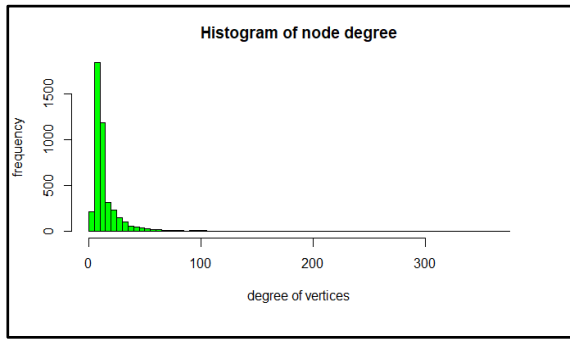


Fig. 6. Histogram of degree nodes

The histogram in figure 6 shows the degree values of the tweets. It indicates that the histogram is skewed on the right for most degree of values of the tweets. Few extreme values are seen in the other side of the histogram. This implies that terms in the tweets have close connection with others.

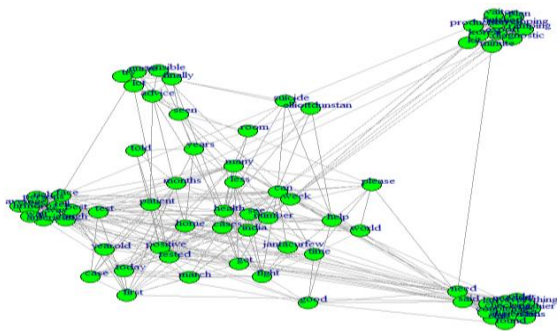


Fig. 7. Term visualization

The network graph in figure 7 provides the cleaner look of the terms using only the terms with the frequency of more than 30. The connected terms above are those terms appeared together on the tweets. The term health is probably at the centre of the network graph and closely connected to other terms such as home, patient, months, case, and number.

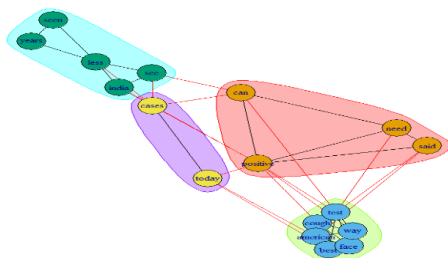


Fig. 8. Nodes clustering based on edge betweenness

Figure 8 indicates the clustered networked terms based on edge betweenness. The four clusters talk about the various measures to lessen the impact of the coronavirus, CoViD-19 cases across countries, the need to be tested and understand the disease, the need of protective equipment, and the common symptoms of the disease.

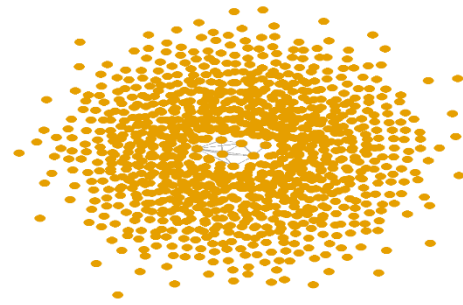


Fig. 9. Tweets vertices

The plot in figure 9 shows the network impact of tweets and its distribution. The points that are already far from the dense area of the plot shows no connection among others while points near and at the centre are tweets that are related. It shows that only few from the extracted tweets do not have connection to each other.

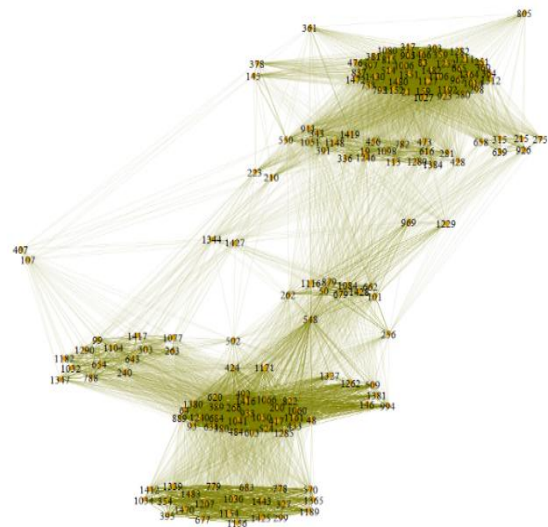


Fig. 10. Network of tweets

The network graph in figure 10 shows the detailed network of tweets. The numbers in each of the points show the twitter ID of tweets in raw data. Tweets in the dense area are most recurrently liked, retweeted, and commented. Below are the tweets randomly picked in every dense area to see what people are commonly

talking about the CoViD-19 outbreak. Based on the most talked tweets online, we now understand how people react to the current situation the world is facing.

People are talking what are the best way to tell if someone has been infected by the disease. People are taking part on sharing information about CoViD-19 confirmation tests. Giving awareness to others by sharing information of the current number of CoViD-19 cases. They believe that the measures and actions made by the government will help lessen the impact of the disease. Reminding people to be prepared as the cases of infected people are increasing rapidly. Lastly, they are talking about the effects of the community quarantine.

DISCUSSIONS

People's Attitudes to the CoViD-19 Pandemic

People are taking part on this crisis by sharing information on twitter, however, subject to this is the proliferation of false news. With the spread of the disease is also the spread of false news. Several social media platforms are making their moves to live up their responsibilities as they have the medium of what information should appear on their sites. Twitters are doing their ways to lessen the spread of false news [21]. Moreover, people should be mindful of sharing information in social media particularly twitter, as the study on the sentiment of tweets on covid-19 confirms that there are misleading stories tend to misinform readers [22]. They are also actively talking about how different private individuals, businesses and governments are doing to help lessen the impact of the disease. Lastly, they call everyone the need for prayers, help each other and follow the precautionary measures imposed by the authorities because of the rapid spread of the disease.

People's Sentiments towards CoViD-19 Pandemic

They trust on the measures imposed by the government that it will make us free from the disease. They also express fear of being infected especially those who are more vulnerable like the children and old ones. Moreover, fear was also expressed by the people for the front-liners of being infected by the disease especially the doctors and nurses. Fear was also an expressed emotion because of the rapid spread of the disease and their still no clear treatment and vaccine for the corona virus [26]. However, this is on contrary to an article, which she claims that people are becoming less fearful. Based on the analysis of tweets, people are not anymore expressing fear about the corona virus, they

become more knowledgeable about the disease [23]. This is also in support of the attitudes revealed by the people, this is because of constantly sharing of information about the contagion. Nevertheless, subject to this is a confirmatory study about the emotions of people expressed based on analysis of tweets. This study revealed an interesting result where despite of the crisis, an optimistic emotion was more expressed from the outcome of the analysis. This is similar to a study in India where the results of the sentiment analysis revealed a positive emotion toward the covid19 outbreak [25][27]. They also trust their government that the measures implemented will be successful and people will not struggle [25].

Heated Topics based on the Network of Tweets

Based on the results of the histogram and tweet vertices, it showed that most of the tweets have close connection with each other. The tweets were clustered into four; it revolves on the discussions about precautionary measures to lessen the impact of the corona virus, cases who are infected by the disease, the need to be tested and understand the disease, and the need to have enough protective equipment.

CONCLUSION AND RECOMMENDATION

Text mining has been widely used across fields, from business, education, and in health issues. This study used the approach particularly sentiment and social network analysis to uncover the attitudes and sentiments of the people towards the CoViD-19 pandemic. This study also looked into how the tweets are connected to each other and find out the popular topics. Results of the study showed how people are taking part on the crisis, by sharing reliable information for the awareness of everyone, calling to help each other and follow the precautionary measures imposed by the authorities. Interestingly, despite the fear felt by the people, the sentiment scores revealed positive emotion towards the crisis. As twitter has been widely used by researchers in various fields, knowing the sentiments and what people are talking online will help authorities understand what is happening and how is people reacting to the current situation, this will also help enact policies that would be valuable to the everyone.

LIMITATIONS AND FUTURE WORKS

Based on the results of the study, the different text mining approach used successfully revealed the attitudes and sentiments of the people toward the

CoViD-19 pandemic. It was also successfully identified the network of tweets and the popular topics. However, there are improving points of the study. First, the data are limited only to the number of tweets. Second, the cluster analysis used only one algorithm to group the terms. Future works could extract a larger number of tweets to gain more interesting results of the study. They can also consider comparing other algorithms for cluster analysis.

REFERENCES

- [1] Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600).<https://doi.org/10.1145/1772690.1772751>
- [2] Molla R. (2020), "How coronavirus took over social media", Voxmedia. Accessed March 20, 2020 [online]. Available at <https://www.shorturl.at/qDJT1>
- [3] Househ, M. (2016). Communicating Ebola through social media and electronic news media outlets: A cross-sectional study. *Health informatics journal*, 22(3), 470-478.[doi: 10.1177/1460458214568037](https://doi.org/10.1177/1460458214568037)
- [4] Gui, X., Wang, Y., Kou, Y., Reynolds, T. L., Chen, Y., Mei, Q., & Zheng, K. (2017). Understanding the Patterns of Health Information Dissemination on Social Media during the Zika Outbreak. In AMIA Annual Symposium Proceedings (Vol. 2017, p. 820). American Medical Informatics Association. Available at <https://www.ncbi.nlm.nih.gov/pubmed/29854148>
- [5] Zhao, Y., & Xu, H. (2020). Chinese public attention to COVID-19 epidemic: Based on social media. *medRxiv*.[doi: https://doi.org/10.1101/2020.03.18.20038026](https://doi.org/10.1101/2020.03.18.20038026)
- [6] Wong, R., & Harris, J. K. (2018). Geospatial distribution of local health department tweets and online searches about Ebola during the 2014 Ebola outbreak. *Disaster medicine and public health preparedness*, 12(3), 287-290.[doi: 10.1017/dmp.2017.69](https://doi.org/10.1017/dmp.2017.69)
- [7] Seltzer, E. K., Horst-Martz, E., Lu, M., & Merchant, R. M. (2017). Public sentiment and discourse about Zika virus on Instagram. *Public Health*, 150, 170-175.[doi: 10.1016/j.puhe.2017.07.015](https://doi.org/10.1016/j.puhe.2017.07.015)
- [8] Fung, I. C. H., Fu, K. W., Chan, C. H., Chan, B. S. B., Cheung, C. N., Abraham, T., & Tse, Z. T. H. (2016). Social media's initial reaction to information and misinformation on Ebola, August 2014: facts and rumors. *Public Health Reports*, 131(3), 461-473.[doi: 10.1177/003335491613100312](https://doi.org/10.1177/003335491613100312)
- [9] Fu, K. W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P., & Fung, I. C. H. (2016). How people react to Zika virus outbreaks on Twitter? A computational content analysis. *American journal of infection control*, 44(12), 1700-1702.[doi: 10.1016/j.ajic.2016.04.253](https://doi.org/10.1016/j.ajic.2016.04.253)
- [10] WHO, "Advice for Public," WHO Int., 2020. [Online]. Available:<https://www.who.int/news-room/feature-stories/2020/02/2020-02-20-advice-for-public> [Accessed 27 February 2020].
- [11] World Health Organization, "Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)," World Health Organization, 2020. Available at <https://www.who.int/docs/default-source/coronavirus/situation-reports/20200211-who-china-joint-mission-report-on-2019-ncov-epidemiology-and-risk-assessment.pdf?sfvrsn=3>
- [12] Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *International journal of antimicrobial agents*, 105924.[doi: https://doi.org/10.1016/j.ijantimicag.2020.105924](https://doi.org/10.1016/j.ijantimicag.2020.105924)
- [13] Kementerian Kesihatan Malaysia, "COVID-19," Kementerian Kesihatan Malaysia, 2020. [Online]. Available: <https://www.facebook.com/kementeriankesihatanmalaysia/>. [Accessed 27 February 2020].
- [14] Hamzah F. A. B., Lau C. H., Nazhri H., Ligot D. V., G. Lee, C. L. Tan, M. K. B. Mohd Shaib, U. H. B. Zaidon, A. B. Abdullah, M. H. Chung, C. H. Ong, P. Y. Chew, R. E. Salunga (2020). CoronaTracker: Worldwide COVID-19 Outbreak Data Analysis and Prediction. [Submitted]. *Bull World Health Organ*. [doi: http://dx.doi.org/10.2471/BLT.20.255695](https://doi.org/10.2471/BLT.20.255695)
- [15] Pastor, C. K. L. (2020). Sentiment Analysis on Synchronous Online Delivery of Instruction due to Extreme Community Quarantine in the Philippines caused by COVID-19 Pandemic. *Asian Journal of Multidisciplinary Studies*, 3(1), 1-6.<https://asianjournal.org/online/index.php/ajms/article/download/207/89>
- [16] Radovanović M., & Ivanović M." Text mining: Approaches and applications". *Novi Sad J. Math*, 2008, 38(3), 227-234. Available at [ibt.ly/O9Z1](https://www.ibt.ly/O9Z1)
- [17] Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press. Available at [ibt.ly/zKEI](https://www.ibt.ly/zKEI)
- [18] Treceñe, J. K. (2019). Delving The Sentiments To Track Emotions In Gender Issues: A Plutchik-Based Sentiment Analysis In Students' Learning Diaries. *International Journal of Scientific & Technology Research*, 8(12), 1134 – 1139. [doi: 10.13140/RG.2.2.35883.80164](https://doi.org/10.13140/RG.2.2.35883.80164).
- [19] Paranyushkin, D. (2019, May). Infranodus: generating insight using text network analysis. In *The World Wide Web Conference* (pp. 3584-3589). [Doi: https://doi.org/10.1145/3308558.3314123](https://doi.org/10.1145/3308558.3314123)
- [20] Girvan M., & Newman ME (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 99(12):7821-6.[doi: https://doi.org/10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799)

- [21] Coronavirus Fake News: How Facebook, Twitter, And Instagram Are Tackling The Problem. <http://ibit.ly/yi1a>, Date accessed March 31, 2020
- [22] How did Twitter react to the Coronavirus pandemic?. <http://ibit.ly/nEwa>, Date accessed March 31, 2020
- [23] David, T. (2020), "What social media tells us about sentiment during COVID-19", PR Week Region Asia. Accessed March 31, 2020 [online]. Available at <http://ibit.ly/9rbA>
- [24] Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Finding an accurate early forecasting model from small dataset: A case of 2019-ncov novel coronavirus outbreak. arXiv preprint arXiv:2003.10776.
- [25] Barkur, G., & Vibha, G. B. K. (2020). Sentiment Analysis of Nationwide Lockdown due to COVID 19 Outbreak: Evidence from India. *Asian Journal of Psychiatry*. <https://doi.org/10.1016/j.ajp.2020.102089>
- [26] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *Journal of Medical Internet Research*, 22(4), e19016. <http://www.jmir.org/2020/4/e19016/doi:10.2196/19016>
- [27] Dubey, A. D. (2020). Twitter Sentiment Analysis during COVID19 Outbreak. Available at SSRN 3572023. Available at SSRN: <https://ssrn.com/abstract=3572023> or <http://dx.doi.org/10.2139/ssrn.3572023>

COPYRIGHTS

Copyright of this article is retained by the author/s, with first publication rights granted to APJMR. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4>).