International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-3 September 2020

Research Method of Clustering of COVID-19 with Text-mining

Junhwan Moon

Abstract: ***A suspected patient with symptoms similar to coronavirus infection-19 was identified for the first time on January 8, 2020 in Korea. After that, the world was dominated by COVID-19. People all around the world must face a new society that will be changed by COVID-19. To prepare for such future, this study collected related words around the keyword COVID-19 and has predicted what risk factors and opportunity factors occur. As a result of SNA analysis by collecting news data from January to May, 2020, when COVID-19 was rapidly spreading, the key words "Prevention of epidemics", "Inspection", "Quarantine", "Infection", "Government", Keywords such as "Patient", "Addition", "Diffusion", "Judgment" and "Prohibition" have had important influences. Furthermore, COVID-19 has been affecting the daily lives of individual citizens, and their interest in the government response process increased. Therefore, the response to the new infectious disease must be quarantine based on science and technology and data, and it is imperative to establish a legal basis for using social facilities as treatment facilities.***

Keywords**: Coronavirus, COVID-19, Big data, Social Network Analysis, Cluster Analysis, Future Prediction, South Korea.***

I. INTRODUCTION

On December 31, 2019, it was officially reported that a pneumonia of unknown origin was prevalent in Wuhan, China. Three months later, on March 5, 2020, 80,555 confirmed patients in China and 17,821 confirmed patients in 90 other countries. The World Health Organization declared on March 11, 2020 that COVID-19 was a pandemic. In addition, that day, the symptoms spread to 114 countries, and more than 4,000 people died. This phenomenon differs from the fact that SARS-CoV was first identified on November 16, 2002 and then caused 8,096 confirmed patients and 774 deaths in 29 countries until July 2003, but it was judged that there was no possibility of a global pandemic. On February 23, 2020, the Korean government announced that they would significantly strengthen the response system by raising the COVID-19 crisis warning stage to the highest level, the critical stage. In the case of Daegu and Cheongdo in Gyeongbuk, Korea, they were classified as a "special management area for infectious diseases" [1]. Afterwards, on March 15, 2020, Daegu was declared a special disaster area, and medical support such as fatigue reduction and replacement plans for dispatched medical personnel, expansion of medical institution budgets, and loss compensation were expanded [2]. Due to the rapidly spreading COVID-19, the Korean government has implemented a strengthened social distancing from March 23, 2020, urging all the people to actively participate [3]. School openings across the country have been continuously postponed and businesses are encouraged to implement a telecommuting system if possible [4, 5]. Some businesses, such as religious facilities, indoor sports facilities, and entertainment facilities, were subject to restrictions on operation, and social life of the people was restricted. As this period of social isolation continues, people have complained of depression, anxiety, and stress [3, 6]. This study analyses news articles on Coronavirus and finds trend from the main keywords and their relationships. Social network analysis is used to find major keywords and their relationships. It will help us find current trend of Coronavirus in short-term period, and it also help us predict future trend of them in long-term view. This study suggests below research questions:

Research Question 1. What are major keywords as Coronavirus in South Korea?

Research Question 2. What are the relationships among these keywords?

II. LITERATURE REVIEW

A. Coronavirus Pandemics

Coronavirus-19 (covid-19) spread worldwide, resulting in 4,013,728 infections and 278,993 deaths in 215 countries (2020. 5. 11, WHO). Because of this, people all over the world are trembling with fear. As COVID-19, first confirmed in China in December 2019, occurred in other countries such as Taiwan, the UK, and Japan and continued to spread, WHO declared Pandemic a pandemic, a global epidemic of infectious diseases on March 11, 2020. Several countries, including Taiwan, have implemented measures such as closing borders and prohibiting the operation of certain airline routes to block the entry of infected patients. The technology of the future information society we will meet is a hyper-connected era where everything is connected and interacts in common. Intelligence through technologies such as artificial intelligence and big data will not only enable on-demand services, but also ultimately produce and develop new technologies and content with surreal forms, attributes, and structures.

Revised Manuscript received on August 01, 2020.
Revised Manuscript received on August 05, 2020.
Manuscript published on September 30, 2020.

Junhwan Moon, Senior Research Engineer of Department of Business Research, Sogang University, Seoul. Email: mjhpioneer@gmail.com

Table- I: COVID-19 outbreaks by country

Country	Number of Patients	Number of deaths
South Korea	10,909	256
England	219,183	31,855
France	139,063	26,380
China	82,918	4,633
Canada	67,996	4,728
Japan	15,798	621
Taiwan	440	6

※ Ministry of Health and Welfare website (2020.5.11.)

Although the power of COVID-19 has not been accurately known to date, it is true that its spread is very fast. When 12 studies were analysed among the papers published from January 1 to February 7, 2020, the basic reproduction number (R0, a mathematical term that indicates how contagious an infectious disease is) of SARS-CoV-2 was estimated to be 1.4–6.49, and the average of R0 was 3.23, with a median of 2.79 [7]. R0 may change as more epidemiological data accumulate in the future, but it is predicted to remain around 2–3. Considering the pattern of COVID-19 transmission in Korea, it seems that transmission occurs mainly in the early stages of infection. In addition, a case of transmission within a family from a family member without symptoms has recently been reported [8]. In other words, the transmission of COVID-19 occurs from the beginning of the infection and can occur even when symptoms are absent or mild. Therefore, until effective vaccines and antiviral agents are developed, the state of quarantine tension will continue and in the long term may become indigenous. Globally, COVID-19 is soaring, and it will take some time before countries can effectively control the spread.

B. Social Network Analysis

Social network theory is based on the point of view that the structural characteristics of social relations influence behavior. It is the method of modelling the relationship between individuals and groups as nodes and links, and quantitatively analysing the structure, diffusion, and evolution of their phase. It is used in various fields as it shows how independent entities are connected within a network and represents the flow between knowledge, information, people, and organizations. (Newman and Girvan, 2004; Wasserman and Faust, 1994). In the existing Supplier relationship management (SRM), it has been analysed according to the traditional data analysis procedure using statistical methodology according to the properties of the data of each supplier. However, social network analysis is the technique that quantitatively expresses the connection relationship between entities, not the properties of each entity's data. If a node is connected to a node, it is marked as 1, and if it is not connected, it is marked as 0, and based on this, a quantitative analysis is performed. This technique considers individuals as nodes and individual social relationships as links to obtain social networks, and is used to find out how it spreads through analysing the structure and strength of social networks, who is an influencer that acts as a hub or center to word of mouth on the network [9]. This social network analysis began in the 1950s for the purpose of small group analysis in the fields of sociology and anthropology, and then research has been steadily continued, and in the field of business administration, it was introduced in earnest from the 1970s [10, 11, 12, 13, 14]. This explains the impact of social relationships. The structural characteristics between members are analysed in detail across multiple dimensions, such as the characteristics of the entire network, the characteristics of individuals in the network, and the characteristics of one-to-one relationships between individuals. There are various Social network analysis indicators for each dimension such as Centralization, whether the network is concentrated among the few at the level of the group as a whole, Centrality, how many individuals are in relationship, Strength of Ties, how close the relationship between individuals and individuals is [14, 15]. The indicators for centrality analysis include Degree centrality and Closeness centrality, based on connectivity between entities, Betweenness centrality, based on inter-individual mediation, and Eigenvector centrality based on the weight of the relative connectivity of individuals [15, 16, 17]. Most of the studies that utilize network information for the analysis target grasp the network characteristics of each node through centrality analysis with various attributes presented above.

III. METHODOLOGY

This study used social network analysis to solve the research problem presented in Chapter 1. In order to predict the future that will be significantly different due to COVID-19, we collected related words around the keyword COVID-19 and tried to predict what risk factors and opportunity factors will occur. We collected news data from January to May 2020, when coronavirus was rapidly spreading, to understand what causes this virus to emerge. In big data analysis, we tried to secure public confidence by using public news data rather than private data, and to grasp the new future environment created by coronavirus. The analysed data was collected by writing code through the R program, and Web crawling was performed. The data collected in text form was cleaned with the R program, and Text mining technique was applied for frequency analysis and correlation. In addition, Centrality analysis and Cluster were performed using NodeXL program, and Network structure was examined by visualizing the research results.

A. DATA COLLECT

Big data refers to collecting, managing, and analysing a large amount of data accumulated in today's digitalized information exchange space to extract valuable information. Gartner defined big data as "a large-capacity, high-speed, and diverse information asset that requires new types of information processing for better decision-making, discovery of implications, and process optimization." Big data analysis has various advantages. Since computer analysis is performed based on algorithms, time, manpower, and costs can be saved after the system is built. Since human coding is not performed when coding data, there is no problem of trust between coders, and since large-scale data is used, various relationships can be identified. As a result, the likelihood of making a type II error is lowered.

In addition, since it is data-driven, it has the advantage of reducing the possibility of the researcher's bias to intervene in the interpretation. When collecting big data, an automated method using computers is used. Methods for collecting large-scale data include those such as using Open Application Programming Interface (Open API), Web-crawling and Log collector.

We used Web crawling technique that is widely used. Web crawler is a technology that automatically collects web documents from a specific site, and mainly used when collecting unstructured data. We used the R program as a Web crawler tool to collect data from January 2020 to May 2020 from NAVER News, a representative portal site in Korea.

B. DATA PREPROCESSING

Words and phrases that we do not need for analysis were eliminated from the collected text data, and the Normalization process of the same and similar but differently expressed words was performed many times. Various special characters and stop-words that do not have meanings were removed from the collected text data through Web crawling. Afterward, the words that have the same meanings but different form of abbreviation were unified. In the normalization work, we wrote the R program code and performed the data preprocessing step.

C. DATA Analysis and Visualization

In this study, the subject of data analysis was limited to unstructured text. In addition, in order to evaluate the status according to the network topology of each supplier and the potential important supplier, the centrality analysis was conducted to analyse the network topology of each node with the highest utilization. Visualization analysis to determine the network structure identifies the frequency of linkages between key words through the thickness of the connecting line between nodes, and at this time, it is determined that each word is strongly connected as the connection line is thicker [18]. Besides, the size of each word node means the number of connection lines between nodes, and it is interpreted that the larger the node size, the more words are connected. In this study, data were analysed by Text mining, Association rule, and Clustering techniques and visualized through the NodeXL program.

IV. RESULTS

A. Analysis of Network Structure

The figures by computing through the NodeXL program are shown in <Table 2>. A total of 109 nodes have been collected in the network, which have been connected by 212 links, of which 218 have been broadly associated. The density, which means the degree of connectivity between people in the network, has been found at 0.02429. The larger the value of the density index of the network, the more active the information exchange between individuals in the network, and that is interpreted as the active exchange of information between entities [19]. Network density figures have been calculated from 0 to 1, and the analysis indicates that the figures are significantly lower. It means that the exchanges are concentrated around few specific key words, rather than organically associated and exchanged between words about the future technology of information society. NodeXL Version is used when computing the network structure. The purpose of network analysis is to find out what role the node's role plays or influences on the network according to the relationship between nodes in the network. Centrality analysis method consists of a node that represents each actor and a link that represents mutual relations, and is identifying their correlations. In particular, it is possible to analyse the center, such as which node acts as a hub and which node acts as a relay, and it allows to understand the role, location, characteristics, and influence of each node in the entire network.

Table- II: Network statistic value

Graph Metric	Value
Vertices	109
Unique Edges	212
Edges With Duplicates	218
Total Edges	430
Self-Loops	0
Reciprocated Vertex Pair Ratio	0.036232
Reciprocated Edge Ratio	0.069930
Connected Components	1
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	109
Maximum Edges in a Connected Component	430
Maximum Geodesic Distance (Diameter)	6
Average Geodesic Distance	3.1933
Graph Density	0.02429
Modularity	-
NodeXL Version	1.0.1.439

We conducted Degree, Betweenness, Closeness and Eigenvector Centrality analysis to identify the entities that have the great effects in the network, and which means they act as mediators such as gatekeepers, opinion leaders. The main keywords in the Coronavirus news regarding Research Question 1 of this study are induced using text mining technologies. The primary purposes of this study are to identify patterns Coronavirus in South Korea. On reviewing the top 20 major keywords shown in Table 3. Looking at this, it can be seen that words such as Prevention of epidemics, Inspection, Quarantine, Infection, Government, Patient, Addition, Diffusion, Judgment, and Prohibition were derived as representing COVID-19. The high degree centrality means a word that is frequently used together with other representative words in a main technology with high word correlation, which can be regarded as a key word with high possibility of convergence with other technologies. As the result of analysing Betweenness Centrality, it can be seen that “Prevention of epidemics”, “Area”, “Prohibition”, “Inspection”, “Aid”, “Quarantine”, and “Medical treatment” show high values. On the other hand, “Consolidation”, “Symptom”, “Judgment”, and “Confirmed case” show relatively low values. When compared with the results of Degree Centrality analysis, the top 20 words of Bethweenness Centrality include “Prevention of epidemics (1463.306)”, “Area (1135.720)”, “Prohibition (1096.297)”,

"Inspection (1034.524)", and "Aid (1022.384)" that have relatively high values. This means that they serve as links for fusion with the existing key words. In other words, these words are interpreted as acting as intermediaries such as gatekeepers or opinion leaders that exchange various information within the network.

Eigenvector centrality is an extended concept from Degree centrality, and it considers the centrality of not only directly connected entities but also indirectly connected entities, so it is possible to identify words that have total influence in the network. Assuming that a word with a high Eigenvector centrality value affects other words, a user with a high prestige centrality in this relationship can be regarded as the most influential core user. "Infection (0.03632)", "Inspection (0.03514)", "Quarantine (0.03357)", "Hospital (0.03005)", "Patient (0.02987)", and "Prevention of epidemics (0.02686)" are interpreted as having high impact. Closeness centrality is a high-impact word for accessing other words. It is confirmed that words such as "Prevention of epidemics", "Inspection", "Quarantine", "Area", and "Addition" have high Closeness centrality values with similar levels of Closeness centrality.

Table- III: Results of centrality (Top 20)

No	Core Keyword	D.C.	B.C.	C.C.	E.C.
1	Prevention of epidemics	17	1463.306	0.00391	0.02686
2	Inspection	17	1034.524	0.00386	0.03514
3	Quarantine	17	989.418	0.00383	0.03357
4	Infection	16	481.147	0.00366	0.03632
5	Government	15	767.455	0.00369	0.02195
6	Patient	14	725.550	0.00377	0.02987
7	Addiction	14	750.564	0.00380	0.02537
8	Diffusion	14	672.307	0.00369	0.02199
9	Judgment	14	310.631	0.00362	0.02586
10	Prohibition	13	1096.297	0.00375	0.02129
11	Area	12	1135.720	0.00383	0.01679
12	Hospital	12	455.181	0.00358	0.03005
13	Aid	11	1022.384	0.00351	0.01141
14	Confirmed case	11	316.813	0.00357	0.02422
15	New type	11	395.005	0.00356	0.02009
16	Medical treatment	10	916.933	0.00332	0.01384
17	Safety	10	630.297	0.00341	0.01178
18	China	10	504.395	0.00353	0.01665
19	Consolidation	10	260.083	0.00315	0.00975
20	Symptom	10	276.047	0300339	0.01968

D.C.: Degree Centrality (Sum of In-Degree and Out-Degree)
B.C.: Betweenness Centrality
C.C.: Closeness Centrality
E.C.: Eigenvector Centrality

Based on words with high Betweenness centrality, which can serve as links for fusion with existing core keywords, the relationship between them is identified. The result of applying the Harel-Koren Fast Multiscale algorithm for network visualization is shown in <Figure 1>. As a result of the analysis, keywords such as "Prevention of epidemics", "Area", "Prohibition", "Inspection", and "Aid" have high Betweenness Centrality. Betweenness Centrality means that it can serve as a link for fusion with existing key keywords. From this point of view, words such as "Prevention of epidemics", "Area", "Prohibition", "Inspection", and "Aid" are serving as gatekeepers or opinion leaders exchanging various information within the network. As the result of examining their network relationships, "Prevention of epidemics" is affected by "Virus", "Response", "Self", "Area", "Coronavirus", "Definite Diagnosis", "Confirmation", "Group", and "Concern", and has the effect on "Mask", "Safety", "Distance", "Aid", "Government", "Consolidation." This is interpreted to a growing interest in self-isolation and infection concerns as COVID-19 spreads. Therefore, efforts are needed to prevent the spread of the virus by supplying masks and strengthening preventive measures. "Area" is found to be affected by "Cancellation", "Addition", "Progress", "Patient", "New type", "Countermeasure", "Concern" and affect "Economy", "Authorities", "Aid", "Daegu", "Prevention of epidemics." This result can be interpreted as increasing interest in "Daegu" as COVID-19 spreads rapidly in Korea, and it means that the keyword "Area" becomes important when those infected appear in one country. "Prohibition" is affected by "Overseas", "Occurrence", "Patient", "Inspection", "Self", "New type", and has the influence on "Nation", "Action", "China", "Entry", "Government." Because COVID-19 was initially originated in China, it means there is growing interest in quarantine and banning entry for people entering China or other countries. "Inspection" is affected by "Hospital", "Hospitalization", "Infectee", "Prearrangement", "Coronavirus", "Entry", "Safety", "Group", and has the impact on "Consequence", "Judgment", "Symptom", "Diagnosis", "Doubt", "Prohibition", "Negative." "Aid" is affected by "Area", "Government", "Diagnosis", "Prevention of epidemics", "Response" and affects "Urgency", "Expansion", "Support fund", "Disaster", "Case." That is, the government's prompt support and response is important, and the public wants support funds and policies to be implemented urgently.

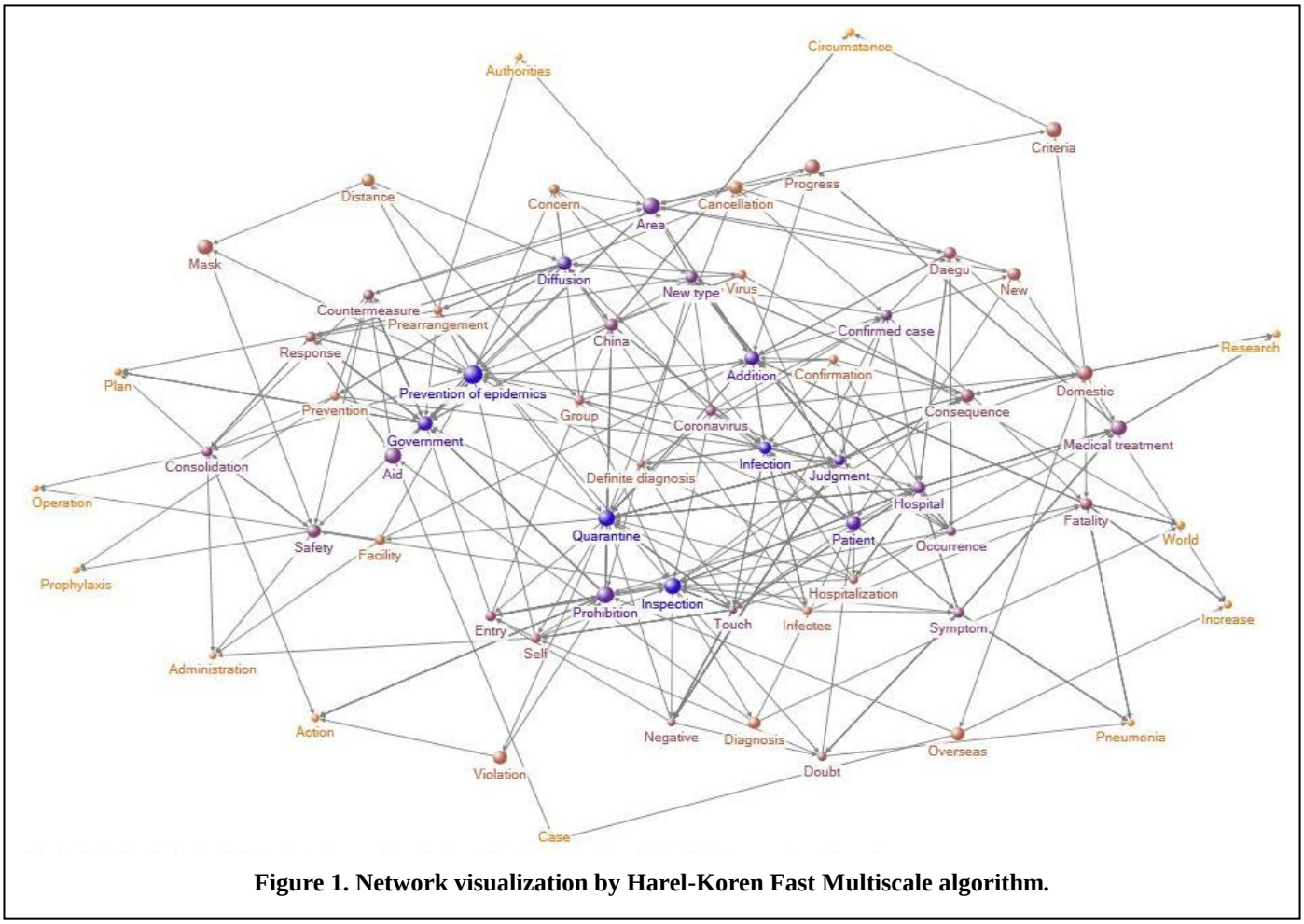

Figure 1. Network visualization by Harel-Koren Fast Multiscale algorithm.

Table- IV: Clustering node and list.

Cluster	Keyword category	Number of Nodes
G1	Entry, Restriction, Quarantine, Self, Group, Inspection, Patient, Infection, Medical treatment, Effect, Clinical, Medicine, Pneumonia, Doubt, Symptom, Fever, Negative, Hospital, Medical staff, Hospitalization, Definite diagnosis, Infectee, Touch, Judgment, Positive	25
G2	Facility, Administration, Safety, Operation, Prevention of epidemics, Countermeasure, Area, Consolidation, Site, Diffusion, Prophylaxis, Prevention, Government, Plan, Prearrangement, The public, Concern, Possibility, Authorities, Economy, Response, Conference	22
G3	Cancellation, Criteria, New, Overseas, Consolidation of quarantine, Case, Influx, Mitigation, Prevention level, New measures, Announcement, Increase, Service, Business	14
G4	Violation, Fine, Accusation, Police, Action, Prohibition, Circumstance, World, New type, China, Wuhan, Nation, Virus	13
G5	Addition, Confirmed case, Domestic, Daegu, Last month, Sinchonji(new world), Fatality, Daily, Confirmation, Occurrence, confirmed case, Coronavirus	12
G6	Distance, Life, Social, Mask, Public, Supply, Purchase, Sale	8
G7	Aid, Expansion, Disaster, Support fund, Urgency, Diagnosis, Kit	7
G8	Consequence, Analysis, Seoul, Research	4
G9	Progress, Continuation, Max, Various	4

G1 which is the largest number of nodes, is grouped into "Medical treatment", "Inspection", "Infection", "Patient", and "Hospital" with the center of “Quarantine”. It means that hospital facilities for quarantine, medical treatment, and examination are important, and efforts to minimize contact with and management of entrants and suspicious symptoms are needed. In order to secure safety, it is necessary to strengthen the government-led quarantine plan and immediately implement it, and to prevent the spread of the virus, occurrence confirmation, management, and support in new areas are necessary. In addition, it was confirmed that social distancing for quarantine and expansion of public mask sales and supply are important in Group 6.

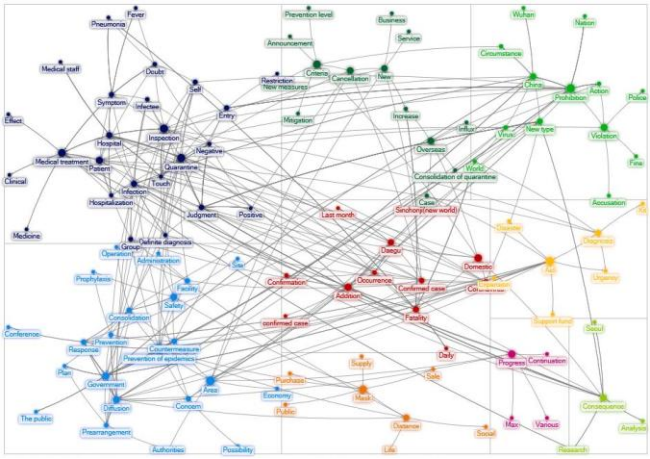

Figure 2. Network visualization by Clustering Analysis.

V. DISCUSSION AND CONCLUSION

This study has conducted big data analysis to identify various threats and opportunity factors arising from the rapid spread of COVID-19. With the results of Social network analysis using news data, key keywords related to COVID-19 have been discovered and the relationship between them has been identified. As a result of SNA analysis, the keywords "Prevention of epidemics", "Inspection", "Quarantine", "Infection", "Government", "Patient", "Addition", "Diffusion", "Judgment" and "Prohibition" are important. It is difficult for anyone to predict exactly how long the current COVID-19 outbreak will last. In the case of Korea, it is difficult to determine whether the spreading power can be lowered due to the case of an explosion in the number of infected people in a specific group. Korea Centers for Disease Control and Prevention has prepared for the outbreak of a new infectious disease before patients appeared in Korea based on its experience in responding to MERS, and actively taken quarantine measures from the beginning. Prior to the outbreak of patients, they made efforts to develop a method for diagnosing COVID-19, and as a result, RT-PCR method that diagnoses within 6 hours was developed. In addition, unlike the MERS outbreak, epidemiological investigation results have been disclosed in detail from the time of the first patient outbreak, and Korea's coping ability is highly evaluated by foreign media through tasks such as separation of confirmed cases and operation of screening clinics by dedicated health professionals. Foreign public health experts appreciate Korea's ability to cope, and Scott Gottllieb, former director of the US Food and Drug Administration (FDA), said, “The Korean health authorities' reports on COVID-19 are very detailed, and they have a great capacity to inspect more than 3,000 cases a day." Since the first COVID-19 confirmed case in Korea was reported, Korea has been blocking the spread of the virus through thorough quarantine, epidemiological investigations, and extensive contact management. However, it is difficult to rule out the possibility of spreading nationwide due to the outbreak of infection clusters in Daegu and Gyeongbuk regions and their effects, so response strategies appropriate to the current situation are needed.

Up to now, we have seen that COVID-19 has been affecting the daily lives of individual citizens, and that interest in government countermeasures is high. With the spread of COVID-19, the problem of shortage of supply and demand for "masks" intensified, and to solve this problem, the Korean government increased the supply of public masks. Currently, in order to prevent the spread of the infectious disease, objective information such as the current status of confirmed patients by region and movement information are delivered in real time, but a government-led data-based system that can view this overall is needed. Response to new infectious diseases should be quarantine based on scientific technology and data. For various social facilities, it is necessary to establish a legal basis for procedures, management, areas of services provided, and manpower management for use as treatment facilities. In addition, an operating system for providing information on Korean medical facilities to overseas patients should be prepared. Central to this study is to discover the main keywords according to the rapid spread of COVID-19 and to prove the relationship between them through Social Network Analysis. Additionally, it is significant to provide basic research data for future prediction and appropriate countermeasures in hard time due to the COVID-19 outbreak.

REFERENCES

1. C. Batini, C. Cappiello, C. Francalanci, A. Maurino, and G. Viscusi. (2011, August). “A capacity and value based model for data architectures adopting integration technologies,” In AMCIS.
2. Central Accident Remediation Headquarters. Central Defense Response Headquarters, 2020. [Online]. Available: http://ncov.mohw.go.kr/tcmBoardView.do?brdId=3&brdGubun=31&dataGubun=&ncvContSeq=583&contSeq=583&board_id=311&gubun=ALL
3. Central Accident Remediation Headquarters. Central Defense Response Headquarters, 2020. [Online]. Available: http://ncov.mohw.go.kr/tcmBoardView.do?brdId=3&brdGubun=31&dataGubun=&ncvContSeq=1241&contSeq=1241&board_id=311&gubun=ALL
4. Central Accident Remediation Headquarters. Central Defense Response Headquarters, 2020. [Online]. Available: http://ncov.mohw.go.kr/tcmBoardView.do?brdId=&brdGubun=&dataGubun=&ncvContSeq=353682&contSeq=353682&board_id=&gubun=ALL
5. Ministry of Education, 2020. [Online]. Available: https://www.moe.go.kr/sub/info.do?m=020102&s=moe
6. Ministry of Employment and Labor, 2020. [Online]. Available: http://www.moel.go.kr/news/enews/report/enewsView.do?news_seq=10754
7. National Trauma Center, 2020. [Online]. Available: https://nct.go.kr/serviceCenter/pressDetail.do?currentPageNo=2&refnceSeq=339&searchKeyword1=
8. Ying Liu, Albert A Gayle, Annelies Wilder-Smith and Joacim Rocklöv, “The reproductive number of COVID-19 is higher compared to SARS coronavirus”, Journal of travel medicine, 2020.
9. Yan Bai, Lingsheng Yao, Tao Wei, Fei Tian, Dong-Yan Jin, Lijuan Chen and Meiyun Wang, “Presumed asymptomatic carrier transmission of COVID-19”, Jama, 323(14), 1406-1407, 2020.
10. Freeman, Linton C. (Ed.)., Research methods in social network analysis. Routledge, 2017.
11. Milgram, Stanley, “The small world problem”, Psychology today, Vol. 2, no. 1, pp.60-67, 1967.
12. Freeman, Linton, “The development of social network analysis”, A Study in the Sociology of Science, 1, 687, 2004.
13. Marin, Alexandra and Barry Wellman, "Social network analysis: An introduction", The SAGE handbook of social network analysis, 11, 2011.

14. Otte, Evelien, and Ronald Rousseau, "Social network analysis: a powerful strategy, also for the information sciences", Journal of information Science, Vol. 28, no. 6, pp.441-453, 2002.
15. Kim, Jooho, and Makarand Hastak, "Social network analysis: Characteristics of online social networks after a disaster", International Journal of Information Management, Vol. 38, no. 1, pp.86-96, 2018.
16. Wasserman, Stanley, and Katherine Faust, Social network analysis: Methods and applications, Vol. 8, Cambridge university press, 1994.
17. Lee, Min-Joong, Sunghee Choi, and Chin-Wan Chung, "Efficient algorithms for updating betweenness centrality in fully dynamic graphs", Information Sciences, 326, pp.278-296, 2016.
18. Scott, Robert A, The Making of Blind Men, Deviance: A Symbolic Interactionist Approach, 236, 1995.
19. De Nooy, Wouter, Andrej Mrvar, and Vladimir Batagelj, "Exploratory social network analysis with Pajek: Revised and expanded edition for updated software", Vol. 46, Cambridge University Press, 2018.
20. Rowley, Timothy J, "Moving beyond dyadic ties: A network theory of stakeholder influences", Academy of management Review, Vol. 22, no. 4, pp.887-910, 1997.
21. Ng, Andrew Y., Michael I. Jordan, and Yair Weiss, "On spectral clustering: Analysis and an algorithm", In Advances in neural information processing systems, pp. 849-856, 2002.
22. Xie, Junyuan, Ross Girshick, and Ali Farhadi, "Unsupervised deep embedding for clustering analysis", In International conference on machine learning, pp. 478-487, 2016.
23. Lee, Jae Yun, Heejung Kim, and Pan Jun Kim, "Domain analysis with text mining: Analysis of digital library research trends using profiling methods", Journal of Information Science, Vol. 36, no. 2, pp.144-161, 2010.
24. Kim, Heejung, and Jae Lee, "Archiving research trends in LIS domain using profiling analysis", Scientometrics, Vol. 80, no. 1, pp75-90, 2009.
25. Huang, Wei, and Sung-Kwun Oh, "Optimized polynomial neural network classifier designed with the aid of space search simultaneous tuning strategy and data preprocessing techniques", Journal of Electrical Engineering & Technology, Vol. 12, no. 2, pp.911-917, 2017.

AUTHORS PROFILE

Dr. Jun Hwan Moon is working as Senior Research Engineer in Sogang University. He has received Ph.D. degree in Management Information System from Sogang University, Korea. He holds a M.S. in Global Service Management from Sogang University, Korea. His research interests include data science, data mining, knowledge management, business model and consumer behavior in online.