Assessing geographic spread of COVID-19 with Social-Contact Network Analysis: Case of South Korea

Mshomba Elias[1][0000-0002-4466-1375] , Kabalisa Rene[1][0000-0002-8450-9378] Enkhbaatar Batbayar[1][0000-0002-0000-0000] , Kassahun Gelana[1][0000-0003-2821-1443] , Mubarkoot Mohammed[1][0000-0001-9265-4375] and Altmann Jorn[1][0000-0002-8880-9546]

[1]Technology Management, Economics and Policy, College of Engineering, Seoul National University, Seoul 08826, South Korea

Abstract. The COVID-19 pandemic has challenged global health systems, resulted in a massive health crisis and cost the lives of many people. The way countries handle such pandemic depends on respective governments' strategies, policies, cultural aspects, and regional contexts. Determining the right action to take is critical and can be game-changing in the battle against COVID-19. Understanding common characteristics of the most infectious super-spreading patients and their clusters can be critical in determining the right move which can help contain the spread of the virus at an early stage. This paper uses social network analysis as a tool to investigate most infectious super-spreading patients and determine social behavior of those patients. Anonymized COVID-19 data has been collected from the Korea Center for Disease Control & Prevention. Analysis shows that populated areas with very close interaction are confirmed to contribute more infections compared to areas with less social contacts. We recommend monitoring infectious individuals in real-time and disseminate information to the public, alerting on the most infected locations. These results can help in directing policy actions and strategic resource allocation in a more urgent and efficient way and therefore mitigate the impact of COVID-19.
Keywords: Super-spreader, pandemic analysis, SNA, Social Network Analysis, COVID-19, Korea, Resource allocation.

1   Introduction

On March 11, 2020, The World Health Organization (WHO) officially acknowledged the novel coronavirus (COVID-19) outbreak as a global pandemic. The ways for the transmission of COVID-19 include direct infections from person to person through; cough, sneeze, speak, droplet inhalation, saliva, fetal-oral and contact transmission, such as the contact with

oral, nasal, and eye mucous membranes [3]. These droplets are relatively heavy, do not travel far and quickly sink to the ground [21]. There is still ongoing research on the ways the virus spreads, the treatment, and vaccination. The rapid spread of the virus has become a global challenge since it has infected around 31 million people in 215 countries by 19th September, 2020 [21].

Korea was no exception in this global pandemic and in early February, 2020, experienced an outbreak of COVID 19. The virus started spreading at the local and community level. The number of new coronavirus cases increased exponentially peaking at 909 new infections on February 29. However, Korea was able to successfully level out the curve on COVID 19 after a few days [6]. Information and Communications Technology (ICT) infrastructure played a role in fighting Covid-19 in Korea, where mobile devices and robots were used for early testing, tracing and spread of essential information. Different other strategies were also used such as enforcement for social distancing, wearing of protective masks, compulsory quarantine for suspected cases and people returning from abroad [7]. Other measures taken include, informational apps and public web-sites in addition to free testing centers and call centers. Countries tend to have different policies and strategies to tackle and contain this pandemic. Early response time together with real time information dissemination measures has proven to be key in containing the virus [8].

Infections of the disease are often transmitted via social contacts of individuals in the small-world network [9]. The patterns of such social contact can be represented as a network structure [10]. Network measures can be used to identify high-risk individuals and assess their attributes. Such measures can help in setting up the stepping stone towards establishment of containment strategies. Identification of most central individuals in the contagion can help in surveillance and develop better control strategies [9]. Within this context, many models are introduced to depict the diffusion of diseases taking into account the nature of the spread of the disease. Traditional pandemic modeling considers full mixing of people but considering other aspects can have different dimensions including cultural and social aspects [10]. There is a need to understand characteristics and patterns that may lead to the diffusion of the virus, taking into account the local context. A global response to establish strong worldwide health systems is needed in order to reduce the number of new infections [5].

The objective of this study is to explore the fundamental attributes for most super-spreading contracting patients and their characteristics to help overcoming the spread of the virus using social network analysis approach.

The subsequent sections of this paper introduce the geographic assessment of COVID-19 with the social network analysis of contracting patients in the South Korean dataset. The remaining part of this paper is structured as follows; Section 2 covers reviews of related literature. The methodology of the research is discussed in section 3 followed by description of the dataset in section 4.

Section 5 analyzes and discusses the results of the study. The last section draws conclusions from the study findings with outline on study implications, recommendations for further research.

2   Literature Review

2.1   The COVID-19 outbreak

WHO defines a pandemic as a worldwide spread of a new disease [21]. In the 18s and 19s, epidemics killed between 50 to 100 million people [21]. Saunders-Hastings (2006) discusses how advances in medicine and epidemiology transformed how we deal with pandemics. Studies by McCullers (2008) also reveal that viruses of this kind such as SARS and MERS are able to adapt and mutate by evading the host's defense systems which in turns makes it difficult to cure. By officially acknowledging the novel coronavirus (COVID-19) outbreak as a global pandemic [21], it was then crucial to understand the geo-environmental determinants which accelerate diffusion of COVID-19 which in turns lead into high deaths rate. There is no known vaccine or treatment for COVID-19 at the moment. One of the key successes to curb the transmission of the COVID-19 was the early screening. [8]. In their study on COVID-19, Wells et al. [29] emphasizes that identification of persons who may have been infected within the epicenter is very important in order to block the person-to-person transmission outside of the infected area or countries. They point out that applying isolation seems to be a better option. The interventions to stop the diffusion of the virus play a great role in response to the spread where early symptoms and infectiousness are concurrent [24][28].

2.2   Countries' initiatives on fighting against the spread of COVID-19

Mario Coccia [33] discusses different strategies for combating the virus. He emphasizes that the spread dynamics of COVID-19 is the result of air pollution-to-human transmission (airborne viral infectivity) rather than human-to-human trans-mission. This study stresses that cities with more than 100 days of air pollution have a very high average number of infected individuals as the transmission dynamics show the existence of high association of air pollution. He also points out general roots for the transmission for all regions including biological characteristics of a virus and incubation period while there are different specific factors for locations such as close ties between air pollution, meteorological conditions and biological characteristics of viral infectivity and health level of individuals (habits, immune system, age, sex). Governments responded differently in fighting against the COVID-19. Pisano (2020) highlights how effective Korea was in responding to the virus by the way they coherently orchestrated actions simultaneously. Minyoung [15] argues that

culture also played an important role in flattening the spread of the virus; Koreans willingness to cooperate and obey the guide-lines was a key to its success in contrast to other countries where people violate rules.

2.3   Policy measures, a tool for curbing COVID-19

The degree of countries affected by the virus in terms of human lives and economies varies depending on their policy preparedness and response actions. The containment and mitigation mechanism also vary from country to country with well-prepared ICT infrastructure and policy mitigation mechanisms. Loayza, et.al, 2020 argue that government policies play a big role on the impact of effectiveness of social distancing and handling pandemics. Various countries consider two categories of policy measures; where the first group of countries took policy measures such as; extended lock down, monetary policy related decisions, and international travel restrictions which were severely affected by the pandemic. Furthermore, the second group imposed restrictions on internal movement and higher fiscal policy spending had a positive impact on the level of economic activities, although the increasing number of confirmed coronavirus cases did not have a significant effect on the level of economic activities [11]. Governments implemented unique policy measures depending on their peculiar situations as per respective countries on measures like; the timing of implementation, the level of social restriction they impose, dealing with the contracted patients, and the impacts of the infectious cases [13].

2.4   Social Network Analysis, a tool for visualizing social interactions

Extant studies show that the outbreak of pandemics can be represented and visualized using Social Network Analysis (SNA) tools whereby nodes are infected patients and connections represent contact of infections [16]. Such a representation helps to look into connections of those nodes and measure how each node contributes to the spread of the pandemic. These measures include degree centrality in which most active nodes can be identified, betweenness centrality to measure the extent to which a node lies in between others. Other measures including closeness centrality, which highlights how contagious the infected patient is, and eigenvector centrality which shows how important the node is within its surrounding context [10]. These measures help identify key nodes which account for the highest portion of spreading the virus which can help looking into further characteristics that those nodes have in common to deliver more understanding of the phenomena. Moreover, they can also lead to revealing some of the unobserved characteristics of nodes [17].

Our study attempts to answer the following questions; which is the mostly infected location based on data analysis? Who are the most infectious

patients in super-spreading of COVID-19? Who is the most vulnerable with regard to attributes (age and gender)?

3   Methodology

Review of literature is conducted on pandemics and their nature of diffusion. Then the study explained how SNA tools can be utilized to understand the spread of the virus through analysis of social contact. . After that data of COVID-19 of South Korea were collected and prepared for analysis. The analysis of the dataset is conducted using SNA tool R programming with "igraph" and "statnet" packages. In this regard, network centrality measures selected including: degree centrality, betweenness centrality, eigenvector centrality and closeness centrality. The reason behind selecting those measures is to extract the most superspreading patients and calculate to what extent they contributed to the spreading of the virus, moreover, these measures also help visualize network structure of super-spreaders and makes it easy to analyze numbers generated by these measures. Based on the analysis, the study then investigates characteristics of most infectious patients using SNA tools and looking through other attributes of those super-spreaders to explore more what they have in common. Results are then presented in forms of network graphs, mathematical and network measures. Based on analysis of results, findings were then discussed and interpreted. Finally, conclusions and policy implications were drawn.

4   Data

Korea has a total population of 51,225,320 across ten cities by 2019 based on data from the United Nation [31]. The statistics from the Korea Center for Disease Control & Prevention (KCDC) on health and epidemiological data are being provided to researchers conducting predictive research and developing implications based on those data [21]. KCDC Categorized contact tracing information into 4 stages: investigation, exposure risk assessment, contact classification, and contact management. During the investigation phase, basic information including whereabouts of the patient for a certain period of time are collected through the process of interviewing patients and suspected contacts. Moreover, if supplementary information is needed, other objective information may be collected during the risk assessment stage [24]. According to [23], data of COVID-19 in South Korea were collected using contact tracing of different channels as following:

● Medical records of the contacting citizens used in a professional and as per the Act of KCDC for giving due attention during treating them.

- Credit card transactions referring to where and when the individual visited those are areas for tracing the contacted peoples.
- Cell phone records for tracing citizens who had possibility of close contacts with the contracted person
- CCTV footages for collecting in real time the information citizens contracting with the identified patients
- GPS tracking in order to get the possible citizens involved and the routes to get the possible contracting once.
- Data of public transportation for gathering possible exposures and provision of information on the public transport users

Furthermore, contact of people who were close to the infected ones are traced instantly in that an automated contact tracing and reporting takes place within 10 minutes [23]. Cases of COVID-19 in South Korea surpassed 22,000 by the time of writing this research [21]. For our study COVID-19 dataset of around 5,166 with the last update on June 18th, 2020. After some preprocessing of data and removal of incomplete and missing records, the focus is narrowed down to 1,347 records. The preprocessing involved eliminating records that have no information related to tracing of infection and those with incomplete data. Preprocessing also included minimizing the dataset and limiting only to records which contain details about most super-spreaders patients.

The data sample for our study consists of data about the patient demographics, cases of infection according to geographical areas and patient routes. Patient information contains information about patient ID, gender, birth year, country of origin, province, city, infected by, symptom onset date, confirmed date, released date, deceased date and state of the patient. Cases data include infection case, places where people infected, confirmed, latitude and longitude. Other data include patient route information including the type of facilities (places) the patient visited.

5   Analysis and Discussion

5.1   Descriptive Analysis

COVID-19 infections vary by provinces. Fig. 1 shows that Daegu is the highly infected province which addresses one of the research questions. Other highly infected provinces include: Gyeongsangbuk-do followed by Seoul and then Gyeonggi-do.
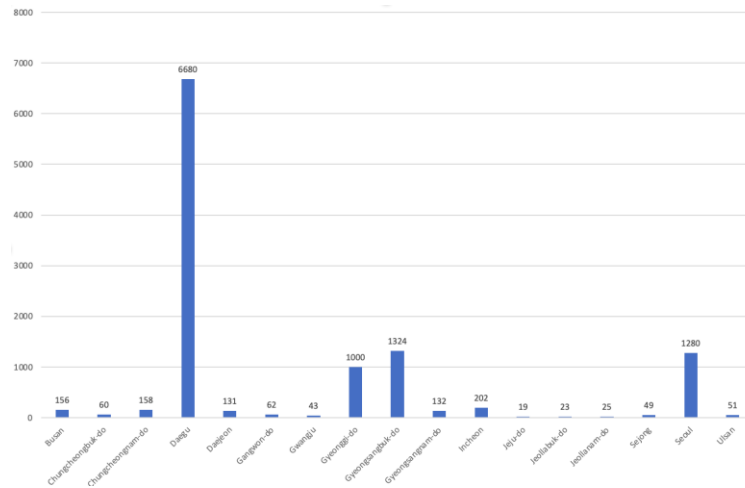
Fig. 1. Cases by province

Findings also revealed an inequality in the distribution with regard to gender. The number of infected females were much higher than those of males. The Fig. 2 illustrates the distribution of patients with regard to gender over time. Data also shows that females were always recording the highest infection over males across time. The figure also shows that the end of March recorded the highest new cases of COVID-19. Referring to the research question, which tends to find patients who are most vulnerable with regard to attributes of gender; the results in Fig. 2 shows that females are more infected than males. However, due to limitations of the dataset, it is hard to conclude that either males or females are more vulnerable than the other.
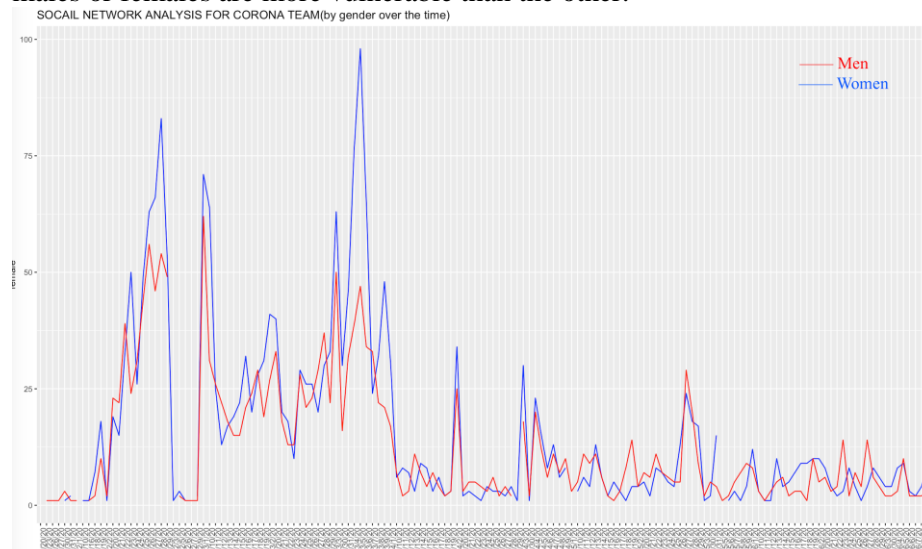
Fig. 2. Representation of male and female infections over time

Looking through the distribution of infection with regard to age; results show that most infections are evenly distributed. Fig. 3, displays that the majority of infected COVID-19 patients fall between the 20s and 60s. Patients of the age 20s occupy around 26% of all infections. Fig. 3 responds to the third research question which tends to find patients who are most vulnerable with regard to attributes of age and gender. Compared to the age distribution of South Korea depicted in Fi. 4, results indicate there is no significant vulnerability of COVID-19 to a certain age. Why is there no significant vulnerability with regard to age? Comparing Korean population of 51,255,320 [31] and the population sample of the dataset which is only 5,166 sourced from kaggle website [32]. Therefore, results cannot conclude higher vulnerability or infection to a certain age.
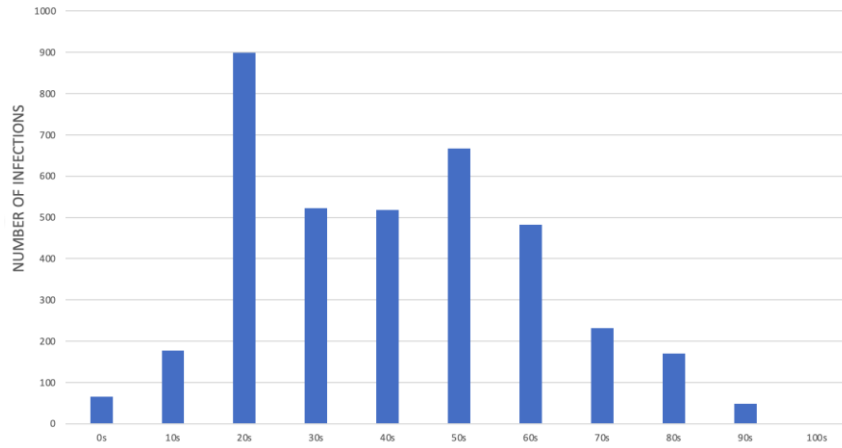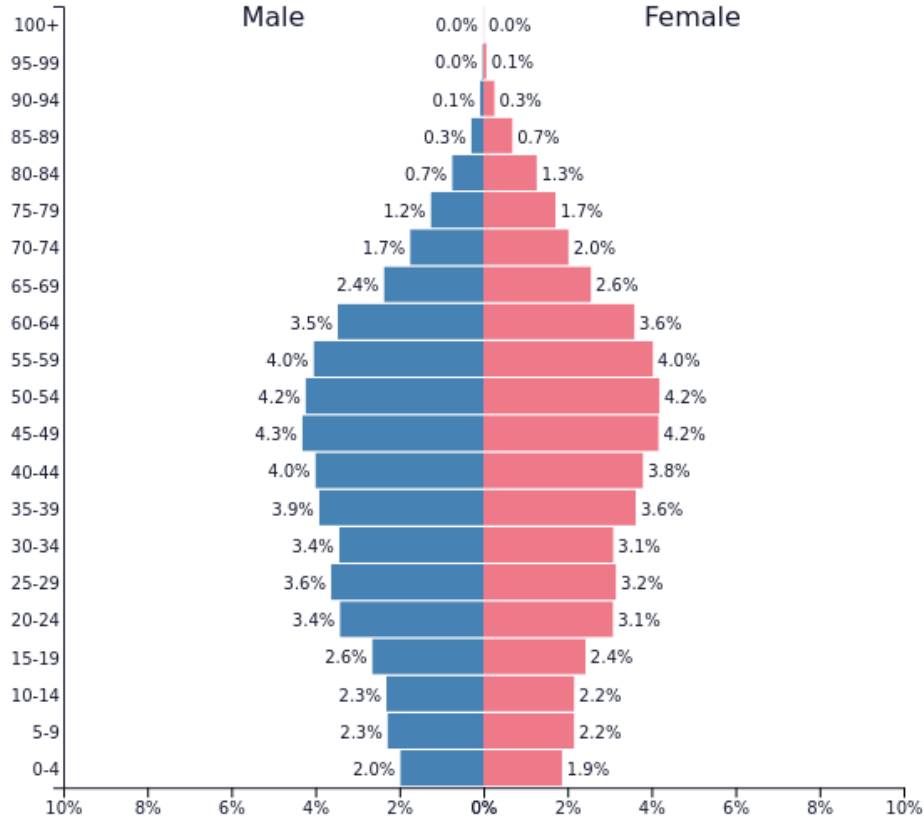

Fig. 3. Distribution of cases by age

Fig. 4 Age distribution of Korea [31]

5.2 Network structure analysis

To identify the most influential patients that have a big impact in the super-spreading of COVID-19 in Korea. The following Fig. 5 shows the representation of the full network structure of the COVID-19 in Korea as of May 25th 2020. The figure depicts the entire network of infections which represents the diffusion of COVID-19 infection in which nodes are infected people and the edges are the spread of infection. The network contains many clusters which show the spread of the virus.
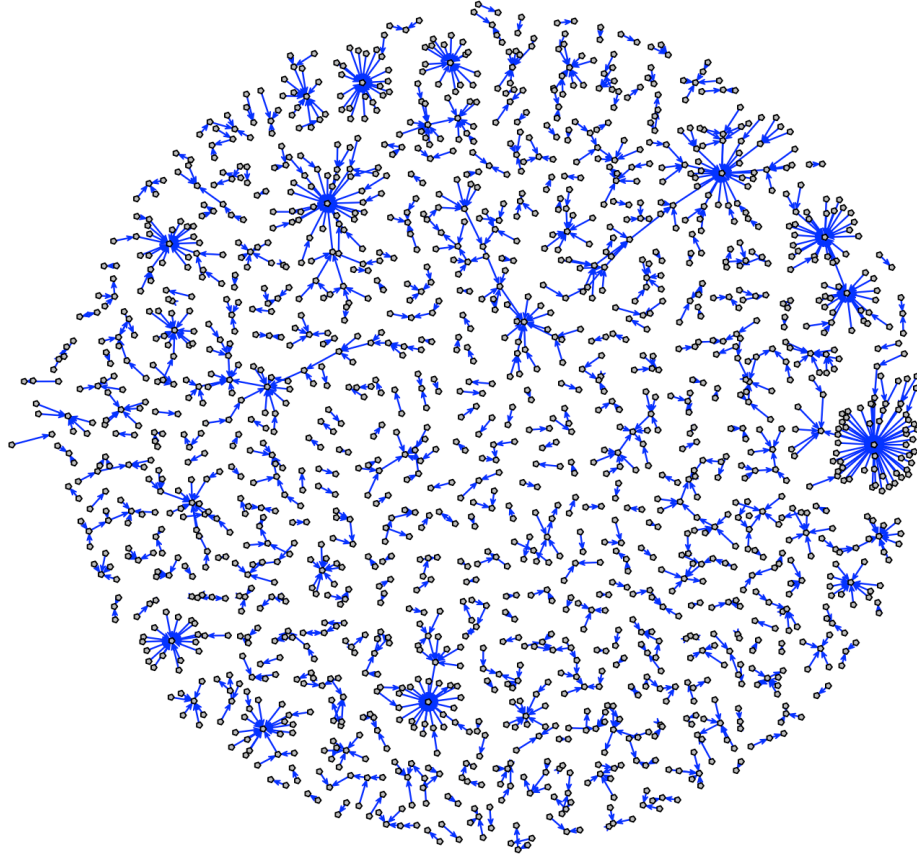
Fig. 5. Network structure of social-contact infections

Studying the network structure of the COVID-19 patients, the graph (Fig. 5) shows different network measures. The structure shows that nodes with the highest degree (with many links), are the most spreading nodes. The graph also reveals network structure with various components (clusters) of different sizes.

Clusters of infection can help reveal and develop a better understanding of the attributes of the infected patients. Analysis of some clusters as illustrated in Fig. 6, shows that contacts with relatively similar age range tend to have similar patterns. As shown in Fig. 6, not all clusters are connected, which brings limitation for extracting complete information about the clusters.
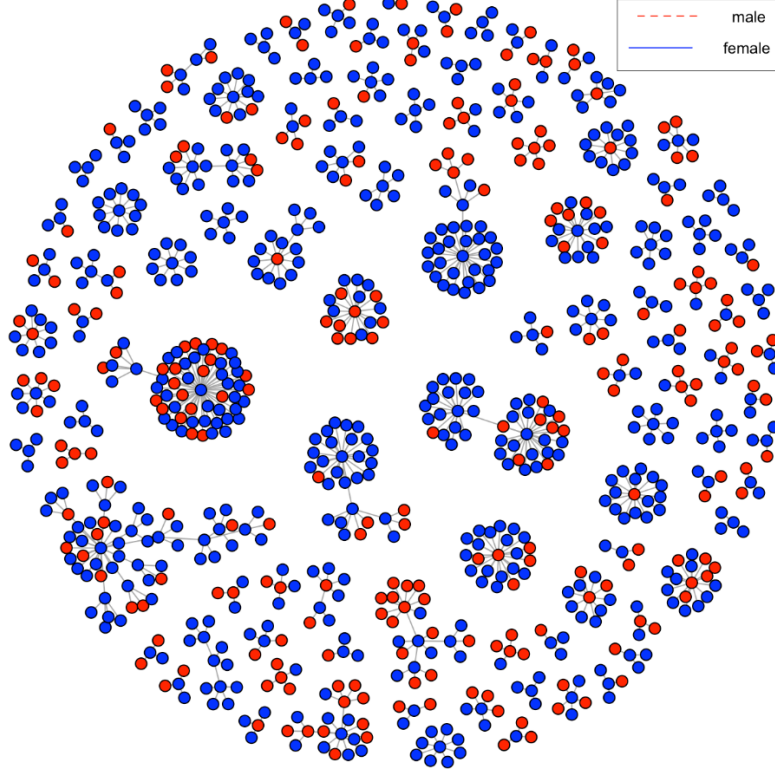
Fig. 6. Network contact structure depicting network measures: (a) degree (b) betweenness (c) closeness and (d) eigenvector-centrality metrics for COVID-19 infection

Table 1. Summary of Contacts of network metrics of COVID-19

Degree Centrality		Betweenness Centrality		Closeness Centrality		Eigenvector Centrality	
Node	Value	Node	Value	Node	Value	Node	Value
2000000205	52	2000000217	51	2000000217	1.37E+08	1000000138	0.153238
4100000008	27	2000000218	24	2000000218	1.37E+08	2000000217	0.138534
2000000167	25	2000000219	22	2000000219	1.37E+08	2000000218	0.138534
1400000209	24	2000000220	15	2000000220	1.37E+08	2000000219	0.138534
4100000006	21	2000000221	12	2000000221	1.37E+08	2000000220	0.138534
2000000309	21	2000000222	9	2000000222	1.37E+08	2000000221	0.138534
1000000125	18	2000000223	9	2000000223	1.37E+08	2000000222	0.138534
2000000508	17	2000000224	9	2000000224	1.36E+08	2000000223	0.138534
1200000031	17	2000000225	8	2000000225	1.36E+08	2000000224	0.138534
2000001153	15	2000000226	7	2000000226	1.36E+08	2000000225	0.138534

Table 1, represents the measures of network centrality of the most 10 infectious nodes (patients) for each centrality measure. Degree centrality metric was identified based on the node with the highest number of connected contacts (connected nodes). Table 1 shows that the patient 2000000205 directly infected 52 patients, he is the super-spreading patient with the highest degree centrality in our network graph. Respectively, patients with ID number 4100000008, 2000000167, 1400000209 directly infected 27, 25 and 24 patients.

Betweenness centrality which measures the extent to which a node lies between others indicates the number of other patients infected by a given patient. Patients 2000000217, 2000000218, 2000000219, and 2000000220 have the highest betweenness score with 51, 24, 22 and 15 respectively. These nodes with high betweenness centrality can be thought of as bridges which participate in transmitting the infection to other super-spreaders ending up creating a new cluster of infections. Therefore, identifying those nodes is critical in the fight of the pandemic. For example, the patient 2000000205, is a 70 years old female from Gyeonggi-do, Seoul, and was the most infectious patient. Moreover, the patient 2000000217 has the highest betweenness centrality, hence acted as a bridge connecting to other individuals which could lead to more infections and creating more clusters. Consequently, proper measures can be taken based on such findings to stop creating new bridges.
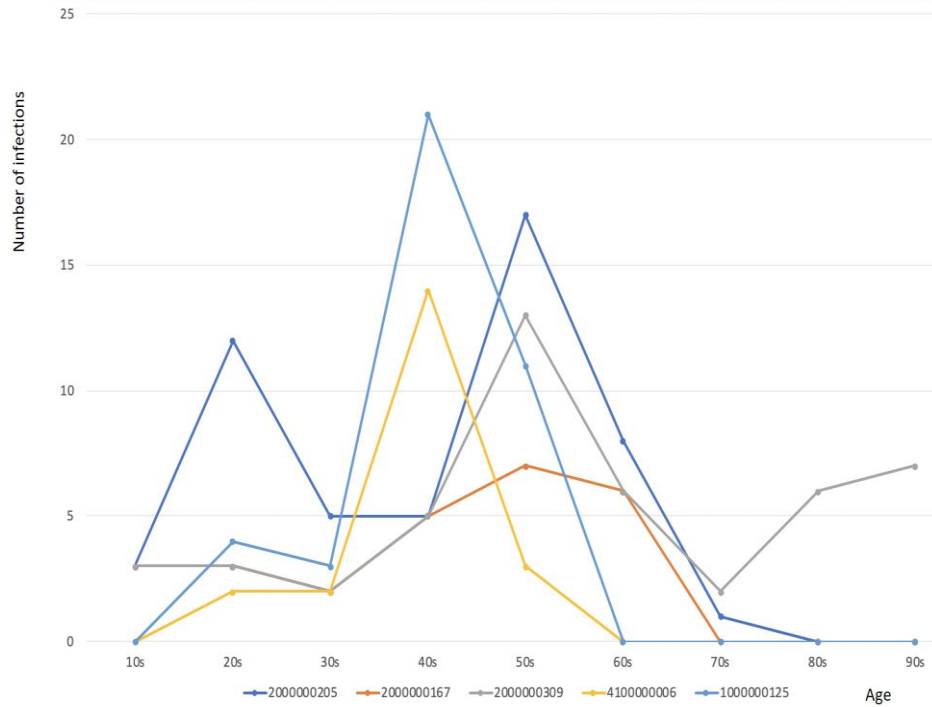

Fig. 7. Super-spreaders within clusters

The Fig. 7 represent the top five super-spreaders within their respective clusters. The figure shows that the majority of infected people range from 30s to 60s. With this analysis, policy should be established to mitigate the spread of the infection. Looking into age ranges of super-spreaders, Fig. 8 shows that they range from 40s to 80s.
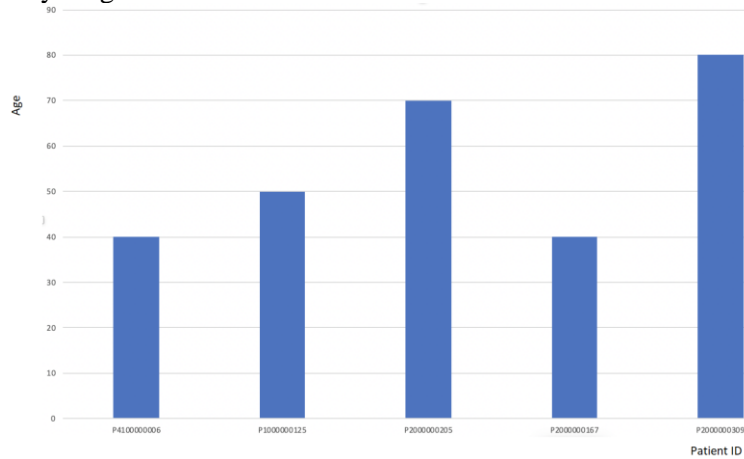

Fig. 8. Super-spreaders by age

With the aid of the social network analysis tool the study focused on answering the three research questions. The four network measures used in the analysis include; degree centrality, betweenness centrality, closeness centrality and Eigenvector centrality. Aligning to the research questions, network measures identify super-spreading nodes (patients) and the extent in which they spread the virus. Descriptive analysis of the dataset shows gender, age and location of COVID-19 infections. The analysis shows that Daegu is the mostly infected location followed by Gyeongsangbuk-do, Seoul and then Gyeonggi-do. Females are the most vulnerable compared to males as per the study results. Results also show the age with high infections ranges from 20-60 years. The network structure also shows the top five clusters of infections linked by five super-spreaders. As per the results, the age of the super-spreaders range from 40-80 years. The results and analysis drawn from this study can act as input for the policy design in an effort to prevent or manage the spread of COVID-19 infections in a more efficient and effective way.

6   Conclusions and Recommendations

The study analyzed different characteristics of COVID-19 data including; geographic and gender distribution, and network measures. KCDC is an organization under the Ministry of Welfare and Health of South Korea. The

social contact network shows how individuals are connected through network measures such as betweenness centrality, degree centrality, closeness centrality and Eigenvector centrality. These net-work measures depict how individuals are connected, the distance between them and the way they interact with each other. The graph therefore helps to identify the most infectious super-spreaders as well as patients who establish bridges in between and creates more infections through new clusters. With this analysis from the graph, it is possible to take quick measures in preventing infectious individuals from spreading the virus. In this regard, the study shows the importance of understanding the social-network tracing in the transmission of COVID-19. Due to limited sample data given compared to the total data of infections in South Korea, the study can not make a general conclusion on the degree of vulnerability. The total data of infected people in Korea is 21,000, but the data given is 5,166. Other sources emphasize the issue of not disclosing privacy of patients as KCDC issued a guideline to municipal and local governments to limit the scope and detail of the information disclosed to the public [30]. This makes it challenging to draw a clear conclusion and generalize findings. Furthermore, the findings recommend directing policy actions and strategic resource allocation in a more urgent and efficient way to mitigate the impact of COVID-19 outbreak. In addition, results provide a room for technology experts to come up with innovative technology to curb the spread of the virus. The data published for research purposes contains less cases compared to the total number of confirmed COVID-19 cases published daily from WHO website. In this regard, the study recommends further study with more data in order to draw extra findings.

References

1. Sohrabi, Catrin, et al. "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)." International Journal of Surgery (2020).
2. Fehr, Anthony R, and Stanley Perlman. "Coronaviruses: An Overview of Their Replication and Pathogenesis." Coronaviruses. Springer, 2015. 1-23. Print.
3. Peng, Xian, et al. "Transmission Routes of 2019-Ncov and Controls in Dental Practice." International Journal of Oral Science 12.1 (2020): 1-6. Print.
4. Perlman, Stanley, and Jason Netland. "Coronaviruses Post-Sars: Update on Replication and Pathogenesis." Nature reviews microbiology 7.6 (2009): 439-50. Print.

5. Remuzzi, Andrea, and Giuseppe Remuzzi. "Covid-19 and Italy: What Next?" The Lancet (2020). Print.
6. Adhikari, Sasmita Poudel, et al. "Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review." Infectious diseases of poverty 9.1 (2020): 1-12.
7. Kim, Pan Suk. "South Korea's fast response to coronavirus disease: implications on public policy and public management theory." Public Management Review (2020): 1-12.
8. Pisano, Gary P., Raffaella Sadun, and Michele Zanini. "Lessons from Italy's response to coronavirus." (2020).
9. Christley, Robert M., et al. "Infection in social networks: using network analysis to identify high-risk individuals." American journal of epidemiology 162.10 (2005): 1024-1031.
10. Newman, Mark. Networks. Oxford university press, 2018.
11. Ozili, Peterson K., and Thankom Arun. "Spillover of COVID-19: Impact on the Global Economy." Available at SSRN 3562570 (2020).
12. Loayza, Norman V., and Steven Pennings. "Macroeconomic policy in the time of COVID-19: A primer for developing countries." (2020).
13. Magal, Pierre and Webb, Glenn, Predicting the Number of Reported and Unreported Cases for the COVID-19 Epidemic in South Korea, Italy, France and Germany (March 19, 2020). Available at SSRN: http://dx.doi.org/10.2139/ssrn.3557360
14. Pisano, Gary P., Raffaella Sadun, and Michele Zanini. "Lessons from Italy's response to coronavirus." (2020).
15. Her, Minyoung. "How is COVID-19 affecting South Korea? What is our current strategy?." Disaster medicine and public health preparedness (2020): 1-3.
16. Borgatti, Stephen P., Martin G. Everett, and Linton C. Freeman. "Ucinet for Windows: Software for social network analysis." Harvard, MA: analytic technologies 2006 (2002).
17. Jackson, Matthew O., Brian W. Rogers, and Yves Zenou. "The economic consequences of social-network structure." Journal of Economic Literature 55.1 (2017): 49-95.
18. McCullers, Jonathan A. "Preparing for the next influenza pandemic." The Pediatric infectious disease journal 27.10 Suppl (2008): S57.
19. Saunders-Hastings, Patrick R., and Daniel Krewski. "Reviewing the history of pandemic influenza: understanding patterns of emergence and transmission." Pathogens 5.4 (2016): 66.
20. Osterholm, Michael T. "Preparing for the next pandemic." New England Journal of Medicine 352.18 (2005): 1839-1842.

21. WHO, 2020, "Coronavirus disease (COVID-2019) situation reports", https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed May 2020).
22. Korean Government, 2020, "Flattening the curve on COVID-19 1 - How Korea responded to a pandemic using ICT", http://viewer.moj.go.kr/skin/doc.html?rs=/result/bbs/51&fn=temp_1587012429386100 (accessed May 2020).
23. Hoon-Sang Lee, 2020, "How South Korea flattened its COVID-19 curve", Global Health Security, Yonsie university, CBC TV interview
24. KCDC, "Press Release", available at https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&act=view&list_no=366764&nPage=1 , April 2020 (accessed on June 2020)
25. https://www.populationpyramid.net/republic-of-korea/2019/ (accessed on June 2020)
26. https://worldpopulationreview.com/countries/south-korea-population/ (accessed on June 2020)
27. Coccia, Mario. "Factors Determining the Diffusion of Covid-19 and Suggested Strategy to Prevent Future Accelerated Viral Infectivity Similar to Covid." Science of the Total Environment (2020): 138474. Print.
28. Swanson, Krista C., et al. "Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015." PLoS neglected tropical diseases 12.9 (2018): e0006762.
29. Wells, Chad R., et al. "Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak." Proceedings of the National Academy of Sciences 117.13 (2020): 7504-7509.
30. Park, Sangchul, Gina Jeehyun Choi, and Haksoo Ko. "Information technology–based tracing strategy in response to COVID-19 in South Korea—privacy controversies." Jama (2020).
31. Population Pyramids of the World from 1950 to 2100. (n.d.). PopulationPyramid.Net. Retrieved September 19, 2020, from https://www.populationpyramid.net/ republic-of-korea/2019
32. Data Science for COVID-19 (DS4C). (n.d.). Retrieved September 19, 2020, from https://kaggle.com/kimjihoo/coronavirusdataset
33. Coccia, M. (2020). Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. Science of The Total Environment, 729, 138474. https://doi.org/10.1016/j.scitotenv.2020.138474