

South Korea as the Role Model for Covid-19 Policies? An Analysis of the Effect of Government Policies on Infection Chain Structures

Alexander Haberling¹, Jakob Laurisch², and Jörn Altmann⁴

¹ University of Mannheim, ahaberli@mail.uni-mannheim.de

² Dresden University of Technology, jakob.laurisch@tu-dresden.de

³ Seoul National University, jorn.altmann@acm.org

Abstract. The fast increase of Covid-19 cases led to high attention from local and international authorities to mitigate and reduce the propagation of the disease. Concerning the risks and the negative effects inflicted by the spread of the pandemic, many countries established a series of policies reinforcing public protection from the virus. With respect to these policies, this study characterizes the infection chain structure in Korea and identifies changes in the structure over time. Furthermore, using multiple linear regressions, the impact of government policy interventions on the infection chain structure is measured. The analysis shows a high fluctuation in infection chain structures at the beginning of the pandemic, which decreases with the implemented policies. The findings serve as a foundation for policymakers to evaluate the success of policies and strategies for reducing the diffusion of Covid-19 and to make optimized resource allocation decisions.

Keywords: Covid-19 · South Korea · Policy · Infection Chain Structure · Network Analysis · Graph Mining · Network Diameter · Network Diffusion · Time Series Analysis · Multiple Linear Regression.

1 Introduction

The first cases of the novel corona-virus Sars-Cov-2 emerged near the end of 2019. After spreading rapidly through China, Korea, and many other countries, the World Health Organization (WHO) declared it a pandemic on March 11th, 2020 [13]. With an estimation of 7 billion infected people and 40 million deceased by the end of 2020 worldwide, assuming no countermeasures were implemented [10], national governments have been urged to find and implement effective and comprehensive responses. As one of the countries praised for its successful handling of the pandemic, Korea serves as a role model for pandemic policy worldwide.

The Korean Center for Disease Control & Prevention (KCDC) declared the first Covid-19 case in Korea on January 20th, 2020[5]. During the following

months, a variety of policies concerning immigration procedures, self-quarantine, mask distribution, and the educational sector were implemented [9]. Even though a lot of data was gathered by the KCDC, a comprehensive assessment of how effective the policies were to contain the Covid-19 pandemic is missing. This paper introduces a framework for characterizing the diffusion mechanisms of the pandemic and for measuring the effect of protective policies. By applying network analysis methods, the framework enhances existing methods to monitor the spread of diseases. In particular, this framework proposes a way to characterize infection chains by their network parameters, making resource allocation and decisions on adequate policies in time possible.

This paper shows that linking key events of a pandemic to changes in network parameters is possible. Additionally, a model analyzing the effect of legislative decisions on the network diameter is introduced. These provided tools enable policymakers and researchers to better understand and prevent the spread of Covid-19. The identification of growing infection chains helps to allocate resources (e.g., medical, computing, experts, medical staff) to areas where they are most needed and most effective.

1.1 Literature Review

To assess available measurement methods and understand legislation aimed at containing the disease, a literature review was conducted and resulted in the following observations. Eunha et al. [8] used publicly available patient data of South Korean cases to simulate a generalized growth model. Their results showed a reproduction number of 1.5, which quantifies the time-dependent variations in the average number of secondary cases generated per case, considering different intrinsic and extrinsic factors. They used an intrinsic growth rate of 0.6, describing the natural increase in population, and a scaling of growth parameter at 0.8, modeling exponential growth dynamics. Their findings also indicate a sub-exponential growth of Covid-19, four major clusters, and a crude case-fatality rate that was estimated to be increasing with older age. Chimmula and Lei [3] focused on the spread of the virus in Canada and made use of data provided by the John Hopkins University and Canadian health authorities. The development of a deep learning long short-term memory (LSTM), a recurrent neural network model frequently used for time-series predictions, enabled them to forecast the Covid-19 outbreak. Under certain conditions, they indicated a possible ending point of the outbreak at around June 2020.

With respect to policy research, Chang et al. [2] were concerned with the Covid-19 pandemic in Australia and calibrated an agent-based model to reproduce key characteristics of Covid-19 transmissions. Their results showed that intervention strategies such as travel restrictions, self-quarantining, and school closures were individually not bringing decisive benefits. Containment of the diffusion was only achieved when coupling several measures with a high level of

social distancing compliance.

Wang et al.[11] conducted social network related research. They observed the network of relationships among patients and hospitals in the Chinese city Henan. Based on this, they estimated an average incubation time of 7.4 days, contrary to the 5-6 day estimate by the WHO[13], and identified 208 cases of cluster infections. As an recommendation for preventing the revival of the diffusion due to patients with long incubation times, the authors suggested a necessity for strong measures until the prevalence of the infection reduces dramatically.

1.2 Problem Description

Despite the impact of the inherent network structure of epidemics, only one research article could be found that analyzes the transmission structure of the underlying network of the Sars-CoV-2 virus, at the time of this research. The use of the network analysis methods to model Covid-19 diffusion mechanism is highly useful, as it allows identifying patterns of diffusion at an early stage. This knowledge allows governments to keep the Covid-19 diffusion under control and protect the society from further spreading of the virus.

Identifying patterns of diffusion of the virus and understanding how it can be contained with specific policies are the main objectives of this research.

In order to achieve that, infection chains are identified and their network structure is quantified with respect to Covid-19 policies. The resulting network measurements are analyzed over time, to identify the changes in the infection chain structure with respect to the policies in place and to identify the most effective policies.

The remaining paper is organized into three sections: the description of the methodology applied, the analysis of the data, and a brief discussion of the results.

2 Methodology

To quantify the infection chain structure, all network components of the network of infections are extracted. A component represents all infected people (i.e., cases) that can be linked with each other. Linking cases requires the analysis of past movements, activities, and interactions of infected people with other people. Figure 1 depicts an exemplary component (i.e., an infection chain). Components with only one infected person represent isolated cases, which could not be linked to other cases.

For the resulting directed trees, the component diameter, which is defined as the longest shortest path between any two nodes in a network, is calculated

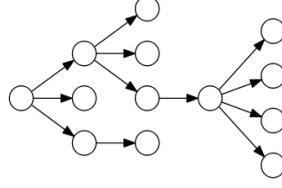


Fig. 1. Exemplary Infection Chain with Diameter 6

according to equation 1. With respect to the Covid-19 infections, the component diameter represents the upper limit for the length of the infection chain. Thereby, the length of the diameter is a representation of how far the virus spreads through the population without being detected. Components with a small diameter represent the successful early isolation of patients and the mitigation of further infections. The earlier the infection chain is cut, the fewer victims will be claimed by the pandemic. The feasibility of using the chain diameter for policy evaluation is grounded in the imperfect nature of tracking Covid-19 cases. What is most likely one huge infection chain in reality is represented as many fractured chains in the available data sets. With this approach, deriving legislative actions from incomplete data is made possible.

As the network diameter represents the depth of the infection chain, it is preferred to other available network parameters (e.g., the average out-degree, which represents the width of the infection chain). The depth of the infection chain is important due to the rapid spreading of the virus, which requires authorities to be able to act quickly after the discovery of cases.

$$Diameter = \max(d(u, v) | u, v \in C) \quad (1)$$

$d(u, v)$ = geodesic distance between the nodes u and v
 C = the nodes in a network component

Changes in the infection chain structure over time are analyzed by calculating the average network diameter (equation 2) for each day. By averaging over the number of active chains, the changing amount of network components is taken into account for the calculation of the network diameter. This way, the diffusion of the pandemic can be compared over different time periods.

$$AverageDiameter_t = \frac{1}{N_t} * \sum_{i=1}^{N_t} Diameter_{i,t} \quad (2)$$

i = network component

t = time

N_t = number of active network components at time t

An estimation of correlations between Covid-19 related government policies and structural changes over time in the extracted network components is performed with multiple linear regressions following equation 3. This linear regression is designed to compare the focal policy with all other policies that are at work at the same time period.

$$y_{t+t^*} = \beta_0 + \beta_s * x_{s,t} + \beta_{-s} * \sum_{p, p \neq s}^P x_{p,t} + \epsilon_t \quad \text{with } s, p \in P \quad (3)$$

y = average network diameter
 x = status of policy $\in \{0, 1\}$
 p = policy
 s = focal policy
 P = set of policies
 t^* = incubation time

Choosing the average component diameter at time $t + 14$ as the dependent variable accounts for the prescribed quarantine that most countries require for infected individuals [?]. To model the influence of virus-related legislation, the binary variable x_t indicates for each time step t whether a policy is active or not, with x_s representing the status of the focal policy and x_p representing all other policies. The model results in the estimation of β_s , representing the effect of the focal policy s on the average network diameter, β_{-s} , representing the effect of all other policies, and β_0 , the y-intercept. Analyzing the values of these coefficients allows for the inference on the relative effect of s on the infection chain structure.

3 Analysis

3.1 Data Set

The data set used is provided by the KCDC [5] and the Data Science for Covid-19 (DS4C) project team [4]. It consists of various sub data sets that contain information on patient-level, as well as macro-level information on population movements, weather, and legislation. In this paper, two sub data sets on patient information and government policy, namely 'PatientInfo.csv' [4] and 'Policy.csv' [4] are used. The former is mostly restricted to patients outside of Daegu, since the local government of Daegu does not provide respective information[4].

The policy data set includes 53 policy decisions categorized in the subsections of 'Alert', 'Immigration', 'Health', 'Social', 'Education', 'Technology' and 'Administrative'. The data set on patient information contains the patient ID, the demographics of the patient, and the patient ID of the infecting patient for the period from January 26th, 2020 to May 12th, 2020.

Overall the patient data set contains information about 3519 patients and 829 transmissions from provinces all around South Korea, namely: Gyeongsangbuk-do, Gyeonggi-do, Seoul, Chungcheongnam-do, Busan, Gyeongsangnam-do, Incheon, Daegu, Gangwon-do, Sejong, Chungcheongbuk-do, Ulsan, Daejeon, Gwangju, Jeollabuk-do, Jeollanam-do, Jeju-do. The provinces are sorted in descending order of the number of cases.

3.2 Descriptive Network Analysis

The complete network extracted using the whole time window displays an average out-degree of 0.78 and is considered very sparse with a density of 0.0014%. It consists of 244 components, representing the infection chains. The average chain size is 4.37 with the largest chain containing 67 cases and the smallest chain 2 cases. The average diameter of the chains is 1.33 ranging from a diameter of 1 to 6. The variance of the diameter is 0.52, with a standard deviation of 0.72. Its median and mode fall on 1.

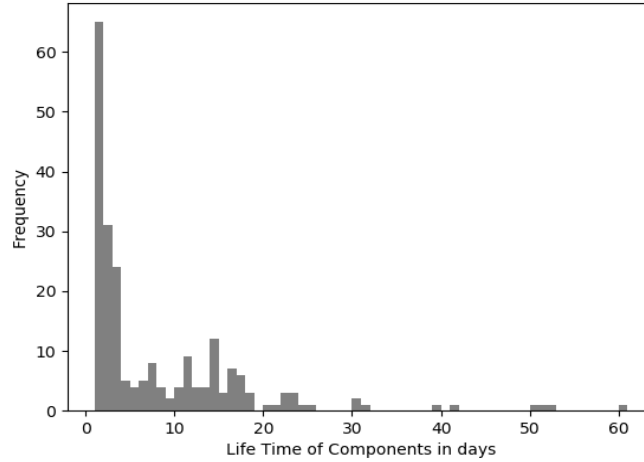


Fig. 2. Histogram of the Component Lifetime in Days

The time between the first and the last confirmed infection of a component is considered its lifetime and averages around 6.99 days. The longest lifetime in the data set is 61 days, while the shortest is within a single day. Figure 2 provides more details on the distribution of the component lifetimes. The bin of

the shown graph is 1.

Figure 3 displays the daily average diameter of all active chains. A infection chain is considered to be active during all days between the dates of its first and last confirmed case. Figure 4 displays the daily number of active chains and provides context to the dynamics of the daily average diameter.

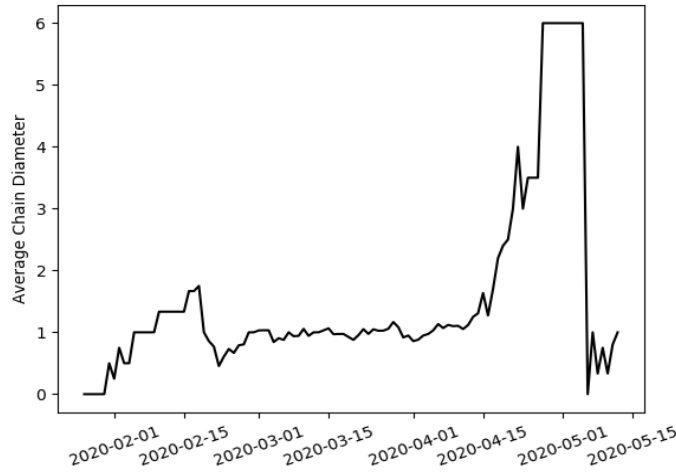


Fig. 3. Average Diameter over Time

The volatile growth in the early stages of the virus diffusion in Figure 3 is followed by a steep decline. This can be explained by the prominent case of ‘patient 31’, the first patient labeled as a super spreader, accused of starting the cluster infections in the Shincheonji Church and Daegu [1], as well as by the increase in testing in Korea. Government efforts, involving a major increase in testing supplies and facilities, screening clinics, and the introduction of newly developed, more efficient, testing kits [1], enabled finding more infected people that were not related to previous components. As a result, the number of active chains increased rapidly (Figure 4). In the following weeks, the daily average diameter levels off around a value of 1. The end of the time series is characterized by a sudden peak and an even steeper decline in the daily average Diameter. These outliers are also better understood by considering Figure 4. The steep increase of the daily average diameter is caused by a steep decline in the number of active chains. The fewer chains are active, the less the daily average of these chains is regularized. At the highest point of Figure 3 only one chain remains active. The chain with the highest diameter in the data set. The rock bottom

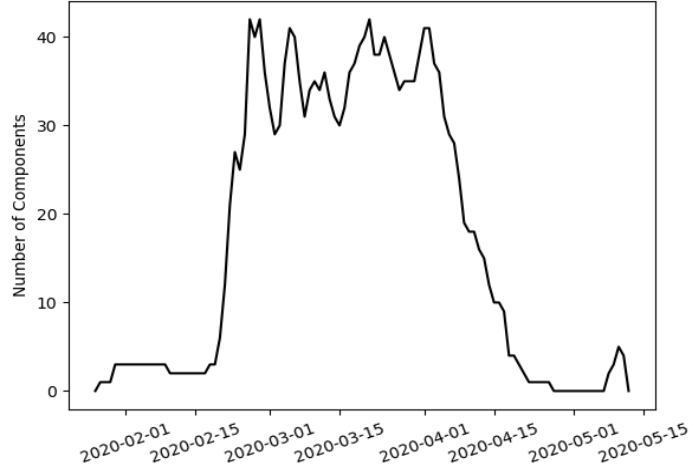


Fig. 4. Number of active Components over Time

is reached afterward, when this last active chain goes inactive, resulting in zero active components and, hence, a daily average diameter of zero.

The mean of all daily calculated average diameters is reported as 1.53 ranging from 6, when only the longest chain is active in a day, to 0 when no chain is active in a day. The variance is 2.3 with a standard deviation of 1.52. Both median and mode are 1.

3.3 Regression Analysis

To quantify the relationship between policy decisions and the daily average diameter, the regression model is employed. It assesses the effect of one policy (β_s) compared to the effect of all other policies (β_{-s}). This model has been estimated for all 53 policies listed in the data set.

In order to prevent the distortion of the regression estimation, the highly volatile end of the time frame was excluded. The time window used for the regression analysis spans from January 26th, 2020 to April 12th, 2020.

Table 1 displays the four legislative decisions that show a significant correlation with the daily average diameter, while the effects of the reference legislation do not. The estimated parameters of the four policies are identical, since all were passed on the same day and span the same time period. The negative β_s -coefficient of -0.1740 indicates a mitigating effect on the infection chain length.

Policy	β_0	β_s	$\beta_{\neg s}$	R^2
Immigration HongKong	1.1207***	-0.1740*	0.0023	0.081
Immigration Macau	1.1207***	-0.1740*	0.0023	0.081
EM ¹ Kit Authorization 1	1.1207***	-0.1740*	0.0023	0.081
Diagnosis App	1.1207***	-0.1740*	0.0023	0.081

Table 1. Regression Models of Significant Focal Policies

Table 2 displays the models, in which the reference policies $\beta_{\neg s}$ are correlated with the average diameter. A negative β_s -coefficient of -0.4702 indicates a mitigating effect on the length of infection chains.

Policy	β_0	β_s	$\beta_{\neg s}$	R^2
Immigration China	1.3609***	-0.4702***	0.0047*	0.438
EM ¹ Kit Authorization 1	1.3609***	-0.4702***	0.0047*	0.438

Table 2. Regression Models of Significant Focal Policies and Significant Reference Policies

With these regressions, it was shown that it is possible to characterize Covid-19 infection chains with the help of network parameters. The dynamics of the daily average diameter was successfully linked to key events in Korea’s Covid-19 course. In spite of major limitations (section 4.2), the implemented model was useful to identify policies that show a mitigating effect on the daily average diameter and, thereby, on the length of infection chains.

4 Conclusion

In this paper, Covid-19 infection chains are characterized by component diameters. Changes in the component structure are analyzed by taking the daily average diameter. The resulting time-series shows large fluctuation at the beginning of the observation period. For the majority of the observed time frames, the diameter fluctuates around the value of 1. The last weeks are characterized by two outliers. Using a multiple linear regression model, 4-6 policies are identified that show a mitigating effect on infection chain length, implicating a high influence on the prevention of spreading of the virus.

4.1 Contributions

The framework proposed in this paper provides the first step towards an in-depth analysis of the network properties of the infection chain structure of Covid-19 in Korea. It proposes a way to predict how the length of infection chains and,

¹ Emergency

thereby, the state of diffusion of the pandemic is impacted through legislation.

This is useful for policymakers and researchers that want to understand the transmission and containment of the Covid-19 disease. It enables them to identify effective policies that could stop the rapid spread of the virus. These approaches could be used to compare the pandemic policies between countries and suggest measures for countries that struggle to contain the disease.

Furthermore, the results help designing mechanisms for allocating resources (e.g., medical, computing, experts, medical staff) to areas where there are components with large diameters. These components represent a large spike in demand for resources. Knowing the number of large components and their location will help to make better use of the available resources.

4.2 Limitations and Future Work

The research presented in this article comes with a few limitations that are listed in the following paragraphs. At the same time, these limitations allow for sophisticated extensions of this research.

First, the analysis of the network diameter over time is performed on the average value for every point in time t , the structure of the used data inheres a highly volatile daily average diameter in the last fraction of the time window (section 3.1). When applying the regression model 3 to the whole time window, the volatility of these outliers will bias the estimated correlation of policies implemented in the later stage of the time series. Future efforts need to handle this bias by regularizing the daily average diameter in a different way.

Second, the lack of a control group, consisting of infection chains that are not affected by policies limits the regression model. The significance of focal policies can only be accessed relative to all other placed policies.

Third, the implemented model allows identifying legislation that is correlated with a shrinking infection chain length. However, policies that display a positive significant correlation might be of interest as well. When estimated with a control group, some of these policies might display a mitigating effect. For dealing with the lack of a control group based on empirical data, simulating the unhindered spread of the virus and, hence, creating a control group with no placed policies might also be a promising way.

Forth, for this paper, the only network measurement applied is the daily average component diameter. While it provides a good evaluation of the infection chain structure, further network measurements such as component size or degree centrality could provide additional information. Focusing on infection rates of chains by analyzing the average time needed for an infection to take place may grant interesting insights as well.

Fifth, the explainability of infection chain structure by other attributes, such as demographic information of the patients or the weather of the infection day, could provide a deeper understanding of the reproduction mechanisms of Covid-19 and allow controlling noise in the dependent variable.

Finally, it is important to take into account, that this study has been conducted in the midst of the Covid-19 pandemic, for which no long-term data on the infection development is available. With more data available in the future, analysis results will assumably improve.

Acknowledgements

We would also like to thank the reviewers of this article, who gave highly valuable feedback for improving our research. Furthermore, we are grateful for the initial discussions and contributions of Antoine Marie Zacharie Kundimana. This research has partly been supported by the Institute of Engineering Research at Seoul National University, which provided research facilities.

References

1. Center For Strategic & International Studies. "A Timeline of South Korea's Response to COVID-19" <https://www.csis.org/analysis/timeline-south-koreas-response-covid-19>. Last accessed 5 June 2020
2. Chang, Sheryl, L., et al.: Modelling transmission and control of the Covid-19 pandemic in Australia. arXiv preprint arXiv:2003.10218 (2020)
3. Chimmula, Vinay Kumar Reddy, and Lei, Zhang.: Time series forecasting of Covid-19 transmission in canada using LSTM networks. *Chaos, Solitons & Fractals* **135** (2020)
4. Kaggle: Data Science for Covid-19 (DS4C). "DS4C: Data Science for COVID-19 in South Korea" <https://www.kaggle.com/kimjihoo/coronavirusdataset>. Last accessed 5 June 2020
5. Korea Centers for Disease Control & Prevention. http://www.cdc.go.kr/cdc_eng. Last accessed 5 June 2020
6. Perc, Matjaž, et al.: Forecasting Covid-19. *Frontiers in Physics* **8**(127) (2020)
7. Sahasranaman, Anand, and Kumar, Nishanth: Network Structure of Covid-19 Spread and the Lacuna in India's Testing Strategy. Available at SSRN 3558548 (2020).
8. Shim, Eunha, et al.: Transmission potential and severity of Covid-19 in South Korea. *International Journal of Infectious Diseases* **93**, 339–344 (2020)
9. The Government of the Republic of Korea: How Korea responded to a pandemic using ICT: Flattening the curve on COVID-19. http://www.undp.org/content/seoul_policy_center/en/home/presscenter/articles/2019/flattening-the-curve-on-covid-19.html. Last accessed 7 June 2020
10. Walker, Patrick, et al.: Report 12: The global impact of Covid-19 and strategies for mitigation and suppression. Imperial College London (2020),
11. Wang, Pei, et al.: Statistical and network analysis of 1212 Covid-19 patients in Henan, China. *International Journal of Infectious Diseases* **95**, 391–398 (2020)

12. World Health Organization: "Coronavirus disease 2019 (Covid-19) Situation Report – 73". <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200402-sitrep-73-Covid--19.pdf>. Last accessed 6 June 2020
13. World Health Organization: "Rolling updates on coronavirus disease (COVID-19)". <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>. Last accessed 6 June 2020