



Research Article

Network-based community detection of early-stage COVID-19 pandemic based on the international geographical distance

Duygu BAĞCI DAŞ^{1,*}, Zerrin IŞIK²

¹Ege University, Ege Vocational School, Department of Computer Programming, İzmir, Türkiye

²Dokuz Eylül University, Department of Computer Engineering, İzmir, Türkiye

ARTICLE INFO

Article history

Received: 30 March 2021

Revised: 14 June 2021

Accepted: 21 July 2021

Keywords:

COVID-19; SARS-CoV-2;

Network Analysis; Pandemic

Analysis; Community Detection

ABSTRACT

COVID-19 pandemic erupted from Wuhan/China, has impacted the whole world in many ways drastically. Understanding such a pandemic, taking necessary precautions, and controlling a potential epidemic is essential. Modules and communities are typically and frequently represented in networks and analyzed with specific methods. Epidemic network analysis is a powerful method that provides us different information to interpret and make accurate decisions for pandemic events. The randomness of epidemic networks is considerably high; hence, it is essential to obtain consistent information from such networks. In this study, we proposed the network-based community detection of the early-stage COVID-19 associated with the basic reproduction number and the geographical distance between country locations. Worldwide confirmed cases between 22/01/2020 and 08/06/2020 had been analyzed for ten-day periods. For this purpose, the Community Detection method was used. Therefore, (i) the community regions, (ii) the change of these regions and weighted nodal degrees of confirmed cases during the COVID-19, (iii) the relations between locations, and (iv) locations or regions which played an important role during the spread of this disease were obtained. The results of this study may help to reduce the reproduction number by lowering the average rate of contact in the early stage of a new pandemic.

Cite this article as: Bağcı Daş D, Işık Z. Network-based community detection of early-stage COVID-19 pandemic based on the international geographical distance. Sigma J Eng Nat Sci 2023;41(4):665–676.

INTRODUCTION

Coronavirus disease (COVID-19) shows up in Wuhan/China on 19 November, which caused millions of people ill and thousands of deaths in a brief period [1]. This pandemic has revealed that humanity is entirely unprepared

for such incidents. Therefore, many studies, including the origin, diagnosis, treatment, and effects of COVID-19, have been conducted. The fundamental goal is to investigate the virus's behavior, cure the disease, develop medicine, slow the spreading of the disease, and examine the effects of COVID-19 on social life and the economy. It is essential to

*Corresponding author.

*E-mail address: duygu.bagci.das@ege.edu.tr

This paper was recommended for publication in revised form by Regional Editor Hijaz Ahmad



take precautions to prevent the spreading and minimize the transmission risk since in pandemics like COVID-19, the individuals' behavior and actions play a significant role in spreading such disease. Governments took several precautions during the COVID-19 epidemic, such as lockdowns, regional or national quarantines, suspending schools and malls, travel restrictions, or bans. These precautions changed the course of the spreading of the disease. Still, the disease spreads rapidly. Investigating the transmission or the consequences of the disease's spreading may provide important information about social, economic, and medical precautions against such a pandemic. At this point, epidemic network analysis takes place. It may be effective to create a strategy to slow down the spreading of such diseases by using outcomes of epidemic network analyses. Besides, the severity of a potential pandemic on a location that may be affected and affects other locations under a possible pandemic can be determined.

Since it is still a living pandemic, few studies about COVID-19 are reported in the literature. Milano and Cannataro [2] took Italian COVID-19 data into account to investigate the communities and their changes over time by performing statistical network analysis. Zhao [3] implemented weighted networks in which vertices represent individuals and the contact is reflected via edges. The study uses the model parameters from China, then applied to the United States. So et al. [4] utilized network analysis to visualize the COVID-19 pandemic risk. Wang et al. [5] investigated COVID-19 patients in Henan, China, by using statistical and network analysis. They concluded that due to the incubation period of the disease, strong measures have to be employed. Besides, they indicated that migrant workers and college students are at high risk. Ceylan [6] performed an estimation of the epidemiological trend of COVID-19 for Spain, Italy, and France using Auto-Regressive Integrated Moving Average models. Huang and Qiao [7] considered the COVID-19 pandemic as a case study to investigate the tracking of an epidemic by using transmission rate. Shi et al. [8] investigated the effects of the travel restrictions on the transmission of SARS-COV-2 by calculating the effective distance between every airport and Wuhan, China. They concluded that the travel restrictions were not sufficient to prevent the spread of the pandemic. Jia et al [9] examined the flow of the population in Wuhan, China on the transmission of COVID-19 by using mobile phone data. Yi et al. [10] conducted a geospatial network analysis of the COVID-19 pandemic considering the migrant population in Singapore. They concluded that the social and geospatial connections of migrant workers have to be taken into account for lockdown conditions while the fundamental daily services should be provided. Javeed et al. [11] proposed a novel mathematical model to express the transmission paths of COVID-19 considering the non-constant transmission rates, epidemiology, and conditions of the environment. They concluded that the developed method can successfully analyze the transmission of

the COVID-10 pandemic in Pakistan, Italy, Japan, Spain, and as well as other regions. Owusu-Mensah et al. [12] employed a fractional-order type susceptible-exposed-infected-recovered model to have a better understanding of the COVID-19 pandemic. They concluded from the conducted simulations that tracing and moving testing up is beneficial since it helps to reduce the spread of the pandemic. Iboi et al. [13] examined the effectiveness of the Public Health Education on the COVID-19 pandemic in the United States by proposing a mathematical model. They concluded that an effective Public Health Education program successfully reduces both daily and total mortality of the pandemic. Besides, they suggested that the public health measurements should be obeyed since the loss of willingness would reverse the positive aspects of the public health education program. Gibbs et al. [14] investigated the travel patterns in China during the early stages of the COVID-19 epidemic by employing network analysis. They specifically examined whether the Lunar New Year and holidays might have a larger role in the transmission of the epidemic. They found that there were no major changes in the transportation after the Lunar New Year. Epidemic network analysis is a powerful method to help to control pandemics.

Due to its significance, there are several studies about epidemic network analysis. Buddenhagen et al. [15] investigate the epidemiological role of seeds by performing epidemic network analysis. They applied an analysis for a regional potato farmer consortium. Wang et al. [16] presented epidemic spread analysis based on the group basis where vertices are grouped in directed networks based on their connectivity. Miller [17] investigated the percolation and epidemics in randomly clustered networks to prove that such networks may have a lower epidemic threshold than configuration model networks, including the same degree distribution. Brockmann and Helbing [18] simplified the complex spatiotemporal patterns by replacing the traditional geographic distance with a probabilistically motivated effective distance considering the air-traffic-mediated epidemics. Iyiola et al. [19] proposed a fractional-order type generalized Chagas vectors re-infestation model to develop insecticide-based strategies to control vectors of Chagas disease.

In this study, network analysis of the COVID-19 pandemic was performed by considering the geographical distance between locations in the world. Worldwide confirmed cases between 22/01/2020 and 08/06/2020 were analyzed for ten-day periods. The community regions for each period were founded by using the Infomap algorithm [20]. These regions were determined by node weights, which were the number of confirmed cases. The edge weights were calculated by taking geographical distances into account. Besides, the nodal degree distribution values and the confirmed case-weighted nodal degrees of key countries were examined to interpret the transmission speed of COVID-19. All analyses were conducted via Python 3.7 and igraph library [21].

One major cause affecting the transmission speed of an infectious disease is the contact rate among individuals living in the same or different locations. The motivation of this study is to examine the transmission behavior of COVID-19 to lower the reproduction number for prevention of a possible future pandemic in early-stage by understanding which countries show similar and relational characteristics, as a community. Therefore, the purpose of this study is to model an approach that evaluates the contact rate based on geographical location to diminish the transmission speed of such a disease by performing community detection analysis.

Apart from the aspects of the various studies that cover the COVID-19 pandemic, we proposed the community detection of the early-stage COVID-19 associated with the basic reproduction number. Besides, the international key locations that played a major role in the transmission speed of the pandemic were examined by employing a network-based community detection analysis. Other contributions of this proposed approach are as follows: (i) The relations between countries in the transmission of COVID-19 were examined by performing community detection analysis based on geographical locations. (ii) The behavior of each country on the tendency of being an isolated location or forming a community with other countries based on the geographical distance and number of confirmed cases was examined. (iii) The effect of geographical locations on the transmission speed of the COVID-19 pandemic was measured. (iv) Locations that had key roles in the transmission speed during the early-stage of the pandemic were investigated.

The rest of the study was organized as follows. In Section 2, the dataset and the considered methodology including the corresponding mathematical expressions were given. The analysis results were given in Section 3. Finally, the conclusions of the study were presented in Section 4.

MATERIAL AND METHOD

The Dataset

The dataset used within this study's scope belongs to COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at John Hopkins University [22]. The dataset contains the daily confirmed, recovered, and death cases of provinces/states, including their countries and geographical coordinates, between 22/01/2020 and 08/06/2020. In this dataset, the data of the U.S. were presented by a single node (single location). For this reason, the transmission of COVID-19 throughout the provinces of the U.S. and the relation between other countries with the U.S. in terms of spreading the disease was not able to be examined. Nevertheless, the U.S. appeared in each network analysis result as a single-country or isolated community.

Methodology

It is estimated that the outcomes of this study will help to reduce the reproductivity of the disease by lowering the average rate of contact between susceptible and infected individuals, which is one of the key elements of the basic reproduction number (R_0) given as [23]

$$R_0 = \left(\frac{\text{infection}}{\text{contact}}\right) \cdot \left(\frac{\text{contact}}{\text{time}}\right) \cdot \left(\frac{\text{time}}{\text{infection}}\right) \quad (1)$$

or

$$R_0 = \tau \cdot \bar{c} \cdot d \quad (2)$$

where τ is the transmissibility, \bar{c} is the average rate of contact between susceptible and infected individuals, and d is the duration of infectiousness, respectively.

Geographical Distance Based Graph Modelling

In this study, confirmed cases of COVID-19 were investigated. The dataset was divided into ten-day periods. Therefore, the analyses were performed in ten-day periods from 22/01/2020 and 08/06/2020. The data was converted into an undirected weighted network. Each location was represented as weighted nodes based on confirmed cases, while each distance between these locations was given as weighted edges as a function of geographical distance. The geographical distance was accepted as a similarity/closeness measure, which helped to evaluate whether two locations should be connected considering the distance between them. According to this method, the connectivity between those locations diminishes as the distance increases [4,8,9,24]. Since the community detection was performed based on international distances, the average country size was taken into account. To perform such an analysis considering national (interprovincial) distances, the distance (approx. 210 km) given in Ref. [25] was considered. Hence, an assumption was made to determine whether an edge should have been drawn between two locations. If the distance between those locations was bigger than 876 km, which is the square root of an average country size in km square [26] the locations were accepted as "not connected". On the other hand, for smaller distances between two places, the edges were drawn and weighted by considering Eq. (3).

$$W_e = \frac{1}{d_{loc}} \quad (3)$$

where W_e denotes the weight value of the edge; d_{loc} is the geographical distance considering the latitude and longitude of the first and the second location. The geographical distance of two locations was calculated by taking Haversine Formula [22]:

$$d_{loc} = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\theta_2 - \theta_1}{2} \right) + \cos(\theta_1) \cos(\theta_2) \sin^2 \left(\frac{\alpha_2 - \alpha_1}{2} \right)} \right) \quad (4)$$

where r denotes the radius of the Earth, θ_i and α_j are the latitude and longitude of location i and j , respectively.

Community Detection

The network analysis was performed via Python 3.7 using the igraph library. The Infomap Community Detection algorithm was implemented to obtain the community regions. Besides, central locations and important places of the pandemic were evaluated. Infomap is a community clustering algorithm based on the map equation, as given in Eq. (5) [20]

$$L(M) = q^*H(\sigma) + \sum_{i=1}^m p^i H(\rho^i) \tag{5}$$

where $L(M)$ is the map equation for a module partition M that includes α nodes ($\alpha = 1, 2, 3, \dots, n$) into i modules ($i = 1, 2, 3, \dots, m$), q^* is the sum of the probability of the exit module i . $H(\sigma)$ denotes the frequency-weighted average length of codewords within the index codebook. The sum of p^i represents the probability of the corresponding module. Finally, $H(\rho^i)$ is the entropy for the module codebook i . Therefore,

$$H(\sigma) = - \sum_{i=1}^m \frac{q_i^*}{\sum_{j=1}^m q_j^*} \log \left(\frac{q_i^*}{\sum_{j=1}^m q_j^*} \right) \tag{6}$$

and,

$$H(\rho^i) = - \frac{q_i^*}{q_i^* + \sum_{\beta \in i} p_\beta} \log \left(\frac{q_i^*}{q_i^* + \sum_{\beta \in i} p_\beta} \right) - \sum_{\alpha \in i} \frac{p_\alpha}{q_i^* + \sum_{\beta \in i} p_\beta} \log \left(\frac{p_\alpha}{q_i^* + \sum_{\beta \in i} p_\beta} \right) \tag{7}$$

where p_α and p_β are the probability rates of nodes α and β which are the elements of the module codebook i .

For undirected networks, the relative weight w_α of the links connected to the node α , is the node visit frequency. The relative weight is denoted as the division of the total weight of the links connected to the node to double the total weight of all links existing in the network. The relative weight of the module i can be written as

$$w_i = \sum_{\alpha \in i} w_\alpha \tag{8}$$

Considering the relative weight of exiting links of the module w_i^* and the total relative weight of the connections between modules, w^* , the map equation becomes

$$L(M) = w^* \log(w^*) - 2 \sum_{i=1}^m w_i^* \log(w_i^*) - \sum_{\alpha=1}^n w_\alpha \log(w_\alpha) + \sum_{i=1}^m (w_i^* + w_i) \log(w_i^* + w_i) \tag{9}$$

Figure 1 shows an example for detected communities by using Infomap based on the geographical locations of confirmed cases.

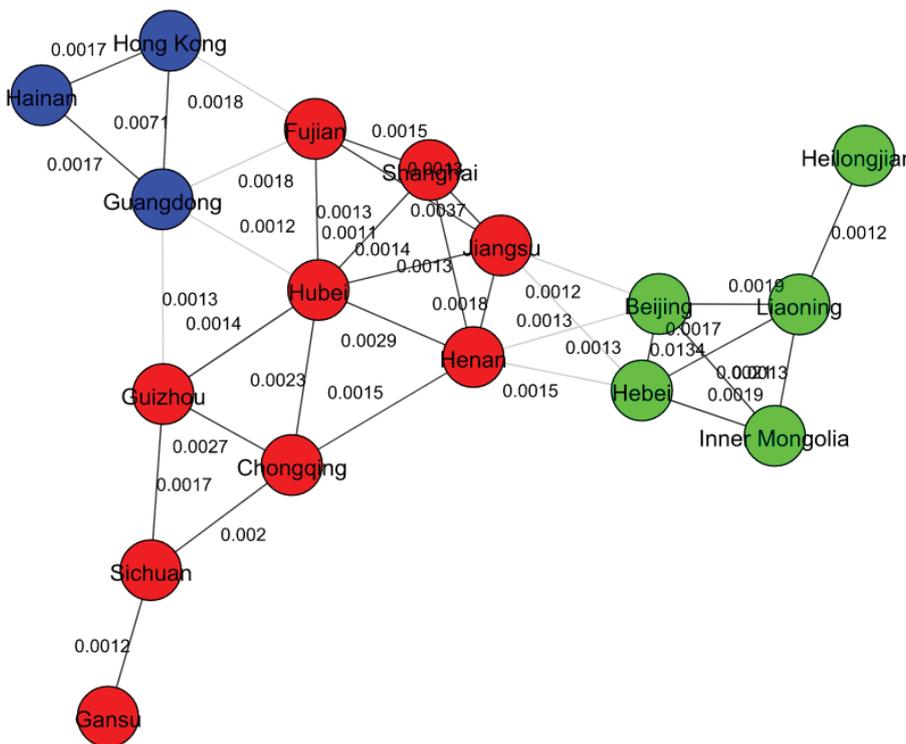


Figure 1. A sample for communities for Chinese.

EXPERIMENTAL RESULTS

We present the community detection results of the network analysis of the COVID-19 pandemic based on geographical distance. Although all periods are analyzed, only the networks of the first, second, fourth to seventh, and fourteenth periods are presented due to critical events (i.e., spreading to a new country/continent or increment of the impact of the disease) of the pandemic occurred within those periods. The third and eighth to thirteenth periods

were generally stable in terms of both spreading characteristics and the impact of the pandemic. Besides, communities having one to four locations were listed only for the first four periods to show the spreading regions of COVID-19. After the fourth period, these communities were also considered, but not listed within the figures since the presence of the disease in these communities was stable and not spreading or not having a further impact that resulted in the occurrence of bigger communities. Figures 2 and 3

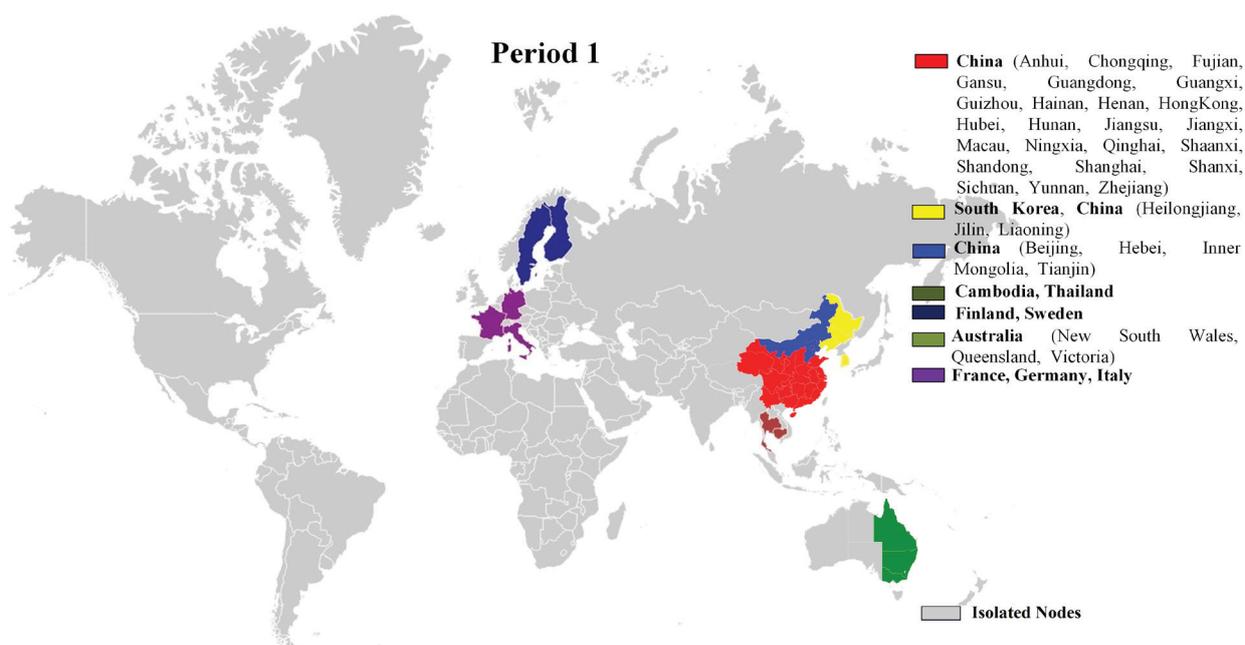


Figure 2. Community detection results of the first period of the COVID-19 pandemic.

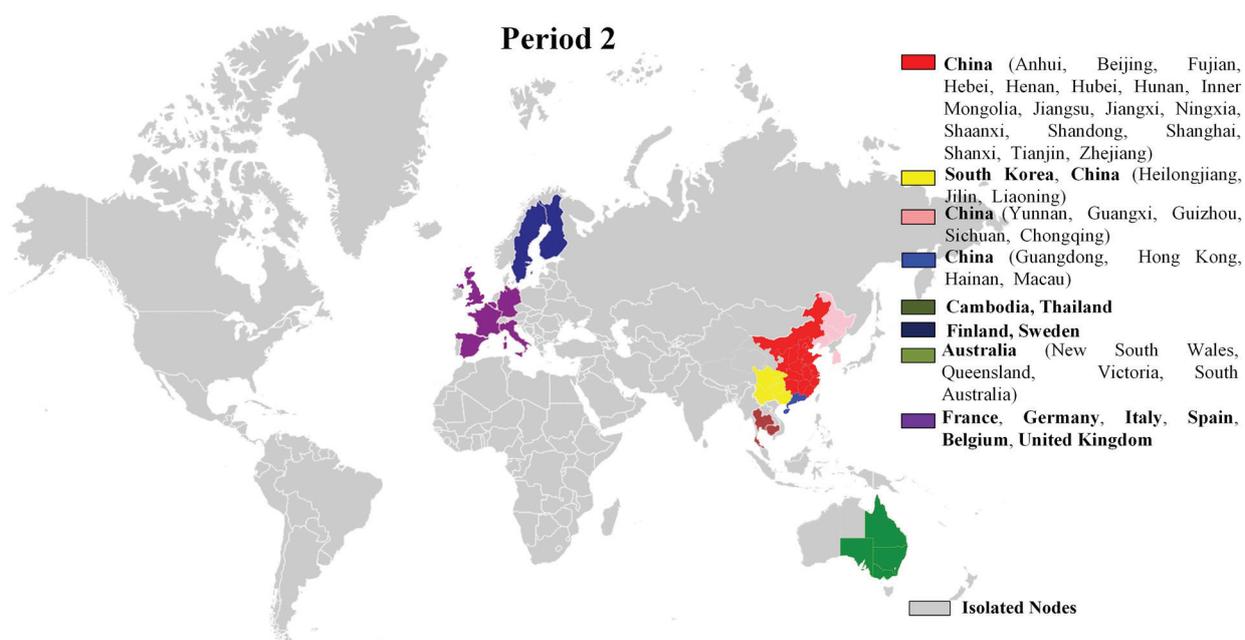


Figure 3. Community detection results of the second period of the COVID-19 pandemic.

show the Infomap community detection results for the first two periods of the COVID-19 pandemic.

According to the results shown in Figures 2 and 3, there are 22 communities for both periods. For the first period, there were 14 isolated locations (i.e., communities with a single location), while for the second period, it was 13. The isolated locations indicate that the disease was present but not effective enough to spread to other countries. The most crowded community was obtained in the Chinese provinces including Hubei, in which Wuhan, the city where the COVID-19 began, is located. It was expected to obtain such a network, dominated by communities, which comprise Chinese provinces since the disease emerged in that location. Some other Asian countries like South Korea, Cambodia, Thailand, and Nepal constituted communities in those two periods, indicating that the disease may also have been spread to other Asian countries from these locations besides China. In these early periods, it is seen that the disease was reached in the Pacific Region and Europe. In the first period, it is understood that Australia was affected by COVID-19. According to the increment of the members of the “Australia Community”, the disease kept spreading within the country during the second period. Besides, the disease was spread to two different communities in the European Region. Countries from Europe, which was the second central region of the disease, showing up in the first period as small communities. According to Figure 2, the disease emerged in France, Germany, Italy, and Scandinavian countries, including Sweden and Finland. During the second period, while the community comprising Sweden and Finland was stable, the UK, Spain, and Belgium joined

France, Germany, and Italy within only ten days. Hence, it can be concluded that France, Germany, and Italy played a major role in the transmission of COVID-19 within the European Region.

The tendency of the disease and the number of the third-period communities were similar to the first and second periods. On the other hand, there were significant changes when the disease was in its fourth period. Figure 4 shows the community detection results of the fourth period of the COVID-19 pandemic.

According to Figure 4, the number of communities increased to 31, of which 19 countries were isolated communities during the fourth period. The transmission of COVID-19 speeded up during the fourth period since there are new major communities where the disease increased its effectiveness. For instance, the community whose members are Bahrain, Iran, Iraq, Kuwait, Oman, Qatar, and the United Arab Emirates, shows that the transmission of the disease was started to spread in the Middle East. The disease was also transmitted to Western Asian countries such as Pakistan and Afghanistan, which constituted a new community during the fourth period. In addition to the Middle East and Western Asia, COVID-19 increased its effects in North America, especially in Canada during this period. It is concluded from the analysis of the fourth period that China was still the critical region since the country did not only constitute a large community itself, but also was a member of other communities, as seen from Figure 4.

The transmission of COVID-19 increased until the end of the seventh period. The disease remained stable in terms of spreading to another country between the eighth

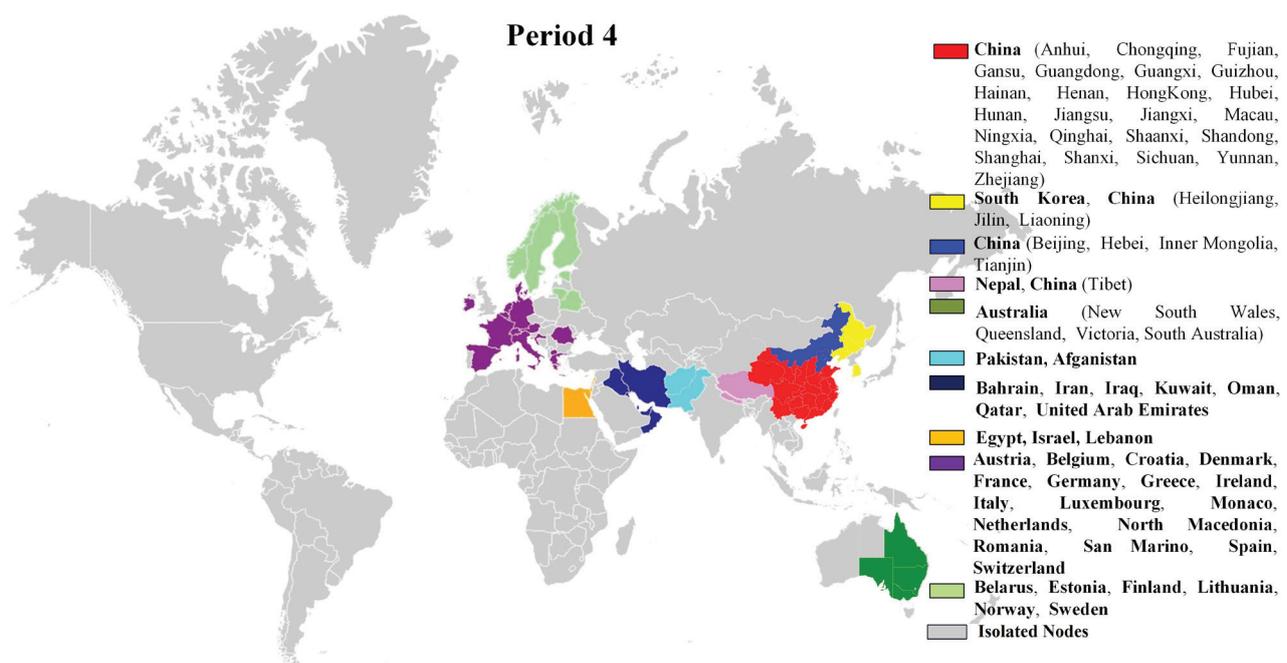


Figure 4. Community detection results of the fourth period of the COVID-19 pandemic.

– fourteenth periods. Figures 5 – 8 show the community detection results of the transmission of COVID-19 during the fifth, sixth, seventh, and fourteenth periods.

According to Figure 5, there are 41 communities, of which 21 are isolated countries. It is concluded that the transmission of COVID-19 still increased during the fifth period since the communities got crowded when compared to the previous periods. Besides, the disease reached to the countries that were not affected by COVID-19 earlier. The virus increased its effect in Africa during this period as another community is constituted by Burkina Faso,

Cameroon, Nigeria, and Togo. Besides, it is seen that the disease also had a significant impact in North Africa, since a community, including Gibraltar and Morocco, emerged within the fifth period. This period was critical because the center of the disease was switched to Europe. As seen from Figure 5, some communities were constituted by European countries. Two of those communities already existed in the fourth period. However, these communities got crowded within the fifth period significantly. This situation made Europe the new center of COVID-19.

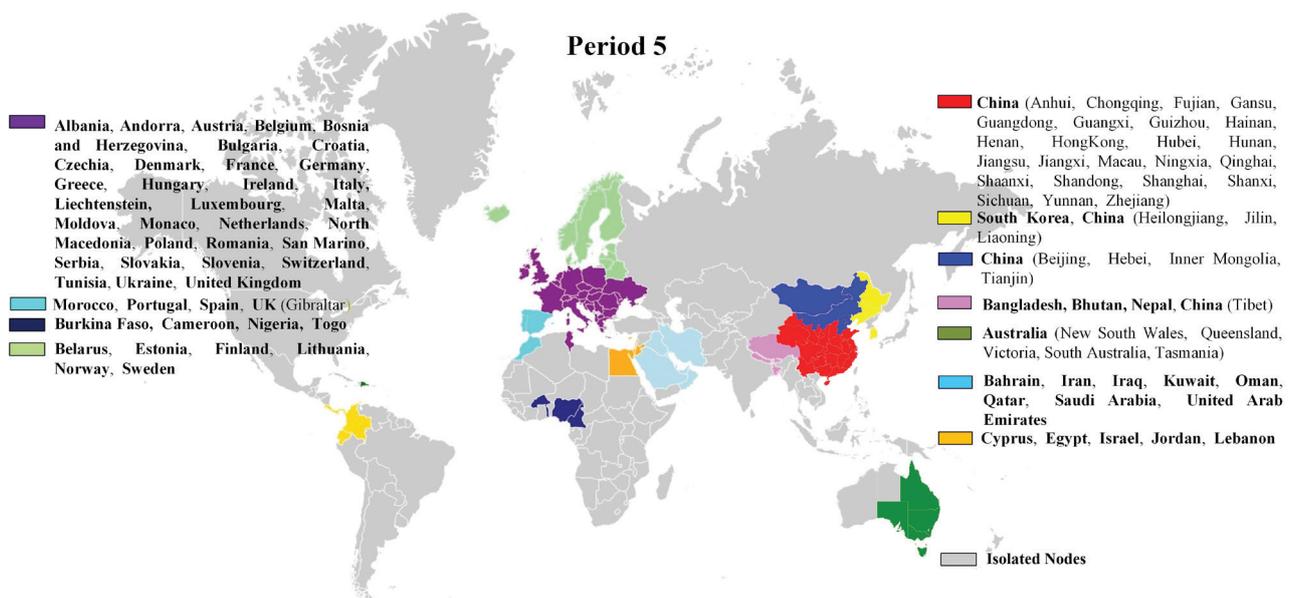


Figure 5. Community detection results of the fifth period of the COVID-19 pandemic.

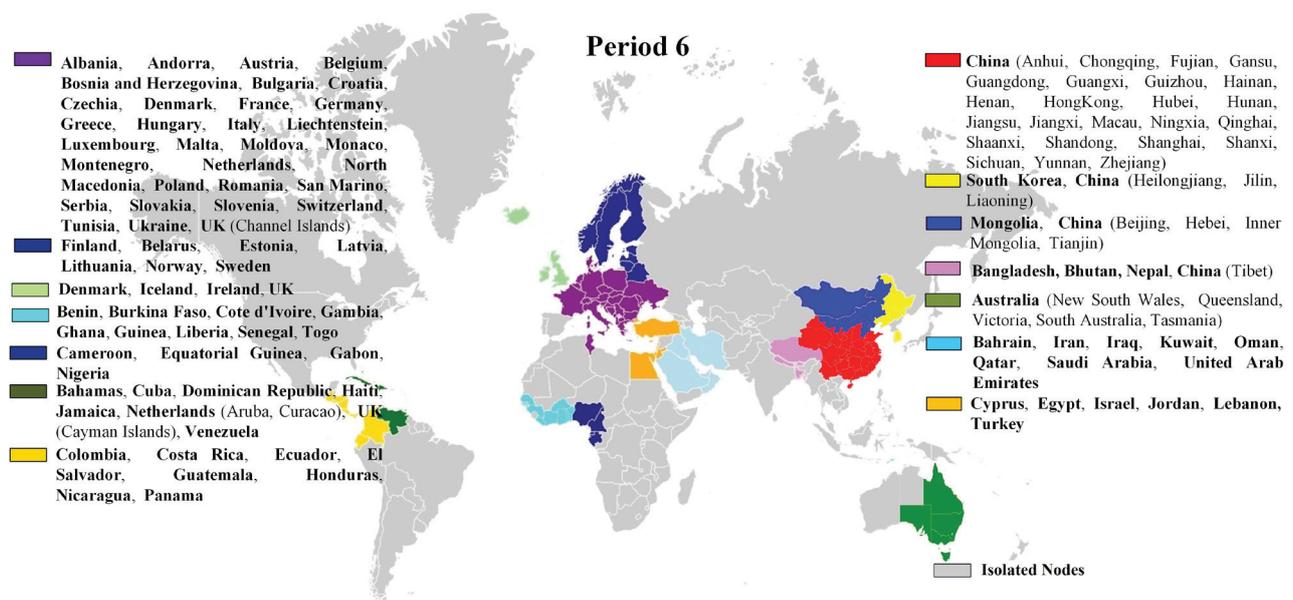


Figure 6. Community detection results of the sixth period of the COVID-19 pandemic

According to Figures 6 and 7, the community number was increased from 41 to 66 during the sixth and seventh periods, respectively. Among these communities, 34 were isolated or single-country communities for the sixth period, while the number of those groups was 31 during the seventh period. The transmission effects of the disease slowed down in China. On the other hand, the spreading of COVID-19 was speeded up in America, Middle East, Western Pacific, and Africa. The disease almost enclosed Canada and countries in the Middle East. The transmission trend was still high in Europe as European communities'

amount and density was high during these periods. South Korea joined the first community, which was covering only China before the seventh period. Following the seventh period, the transmission trend remained stable throughout the world. Hence, the communities of the eighth to fourteenth periods were almost identical to each other.

Figure 8 shows the community detection analysis results of the fourteenth period of the COVID-19 pandemic. There were 66 communities, of which 33 are single-country communities during this period. As seen from Figure 8, the communities were similar to those obtained

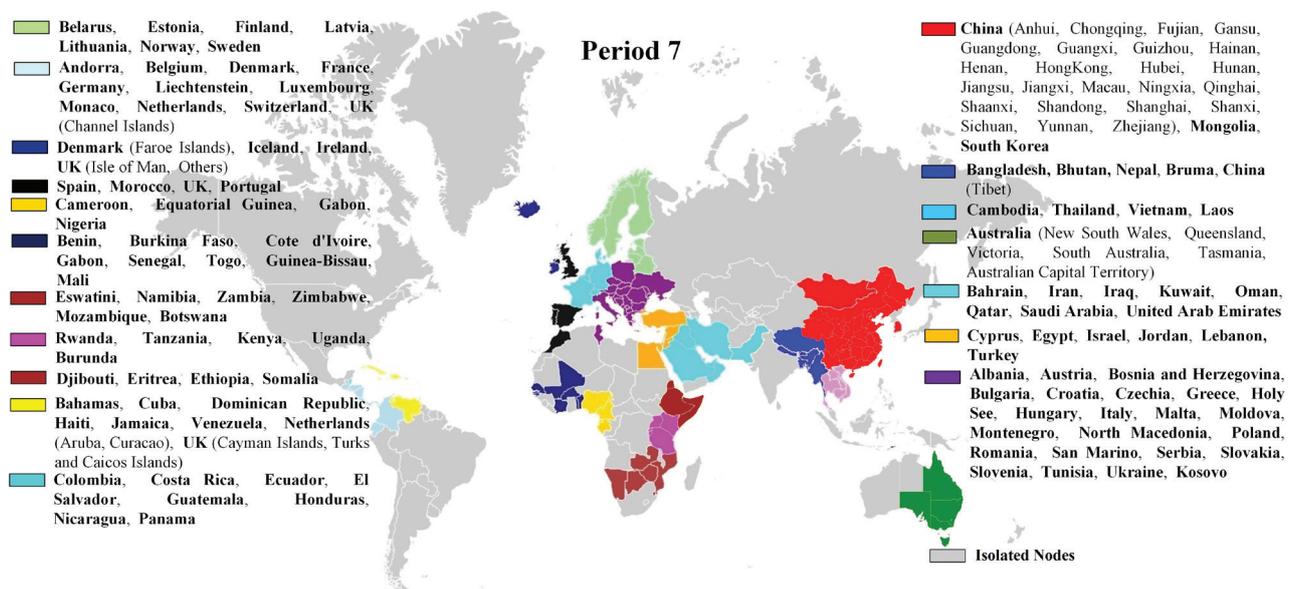


Figure 7. Community detection results of the seventh period of the COVID-19 pandemic.

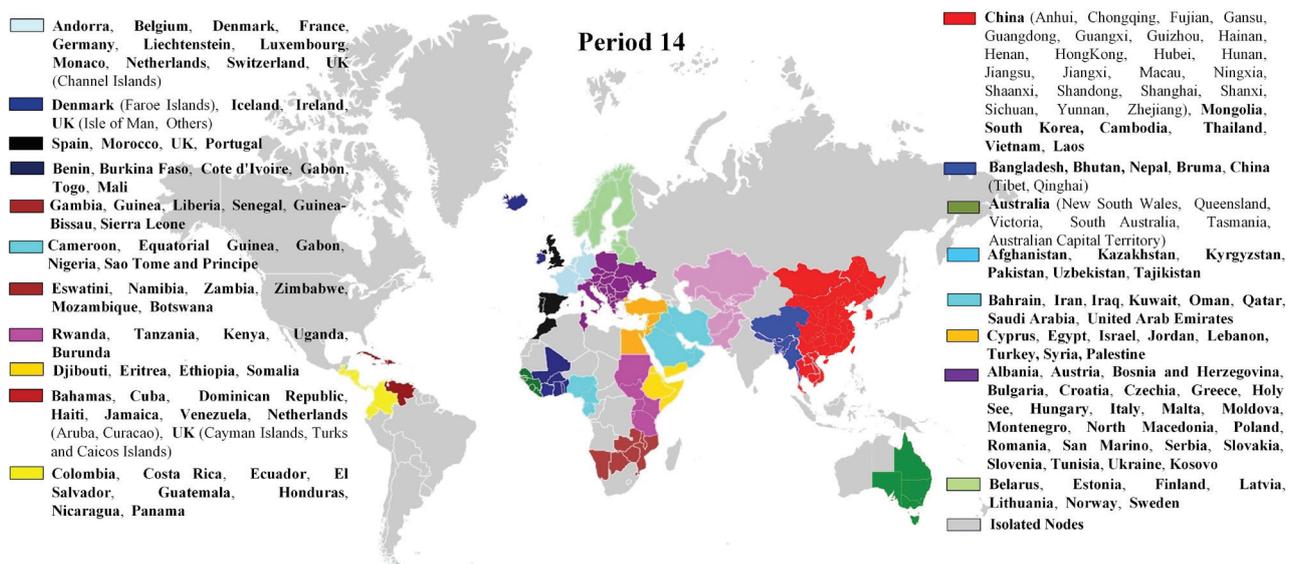


Figure 8. Community detection results of the fourteenth period of the COVID-19 pandemic.

for the seventh period. Beginning from the eighth period, Taiwan, Thailand, Vietnam, Laos, and Cambodia joined the first community. The COVID-19 transmission has begun to slow down everywhere except in some parts of Asia and Africa after the eighth to the fourteenth period since two communities emerged in these regions during the eighth period of the pandemic.

In order to understand the roles of locations in the transmission of COVID-19, we also considered the nodal degree of each location. The nodal degree reflects the number of connections of a location with other locations. Table 1 gives the periodic nodal degree distribution of the locations (nodes). Here, zero value means that the corresponding nodes were isolated, i.e., not connected to any location. For example, there were 9 locations, which have only one neighbor (nodal degree=1), in the first period; this increased to 27 locations in the fourteenth period.

According to the analysis results in Table 1, the average nodal degree increased significantly during the fourth period and kept growing until the seventh period. Afterward, the increment diminished and slowly stabilized.

The average nodal degree trend indicates that the pandemic was out of control and spread speed is too high, especially between the fourth and seventh periods. Table 2 gives the most central locations during the first fourteen periods of the COVID-19 pandemic according to nodal degrees weighted by confirmed cases.

It is seen from Table 2 that the disease increased its effect and spread to the different locations until the seventh period. During the first period, COVID-19 was already present in many locations in Europe. However, as it was spread throughout Europe, Italy became the center of the disease of Europe during the fifth period. Similarly, Iran, where the disease emerged during the third period, was another key location in the Middle East in the same period. According to Table 2 and community detection analysis results, the first five periods of the COVID-19 pandemic were the most critical in terms of transmission, since the disease increased its effect in not only Asia but also in Europe and the Middle East.

Table 1. Periodic nodal degree distribution of COVID-19 pandemic

Nodal Degree	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
0	14	13	15	19	21	34	32	33	33	33	33	33	33	33
1	9	10	10	11	18	28	30	30	30	29	29	27	27	27
2	4	5	5	11	17	27	22	21	21	24	24	25	25	25
3	2	3	3	5	19	23	31	28	28	27	27	29	29	29
4	4	5	5	10	9	23	22	28	28	29	29	29	29	29
5	1	1	1	4	5	10	13	14	14	14	14	14	14	14
6	1	1	1	2	10	15	18	21	21	21	21	21	21	21
7	2	2	2	4	4	5	7	7	7	7	7	7	7	7
8	3	3	3	4	3	4	9	9	9	9	9	9	9	9
9	3	3	3	6	7	5	4	4	4	4	4	4	4	4
10	6	6	6	11	6	12	7	7	7	7	7	7	7	7
11	3	3	3	4	5	9	4	4	4	4	4	4	4	4
12	2	2	2	2	3	5	6	6	6	6	6	6	6	6
13	2	2	2	2	3	3	3	3	3	3	3	3	3	3
14	1	1	1	1	3	4	4	4	4	4	4	4	4	4
15	2	2	2	2	7	3	3	3	3	3	3	3	3	3
16	0	0	0	0	2	5	8	8	8	8	8	8	8	8
17	0	0	0	0	0	2	12	12	12	12	12	12	12	12
18	0	0	0	0	5	1	4	4	4	4	4	4	4	4
19	0	0	0	0	3	5	2	2	2	2	2	2	2	2
20	0	0	0	0	1	2	6	6	6	6	6	6	6	6
21	0	0	0	0	2	1	1	1	1	1	1	1	1	1
22	0	0	0	0	0	2	1	1	1	1	1	1	1	1
23	0	0	0	0	0	0	2	2	2	2	2	2	2	2
Avg	2.46	2.59	2.67	4.09	6.38	9.50	10.46	10.75	10.75	10.84	10.84	10.88	10.88	10.88

Table 2. The key locations according to weighted nodal degrees

	P1	P2	P3	P4	P5	P6	P7
1	Hubei	Hubei	Hubei	Henan	Hubei	Hubei	Italy
2	Henan	Henan	Henan	Hubei	Italy	Italy	Germany
3	Guangdong	Guangdong	Guangdong	Jiangxi	Iran	Germany	Hubei
	P8	P9	P10	P11	P12	P13	P14
1	Italy	Italy	Italy	Italy	Italy	Italy	Italy
2	Germany	Germany	Germany	Germany	Germany	Germany	Germany
3	France	France	France	France	France	France	France

CONCLUSIONS

We applied the community detection analysis based on the geographical locations of the confirmed cases of the COVID-19 pandemic to determine the locations that constituted communities due to the transmission of COVID-19. Besides, we aimed to evaluate the contact rate in terms of the nodal degree to have a foresight for diminishing the transmission speed of such pandemics.

As seen from Figures 2-8, the number of communities increased during the first fourteen periods of the COVID-19 pandemic. The most considerable increments occurred during the fourth and sixth periods of the epidemic. The reason behind this situation may differ (e.g., insufficient precautions, lack of information about the disease, insufficient travel restrictions, etc.). Figures 2-8 also indicated that some communities became denser since they are geographically close to each other. For instance, European countries such as Italy, France, Germany, United Kingdom, Spain, and Belgium constituted a community during the second period of the COVID-19. On the other hand, although they are geographically close to each other, they became a part of separate communities from the beginning of the seventh period to the fourteenth period. Therefore, it can be concluded that the transmission occurred in a more local area, which may be due to the precautions that were taken by those countries.

According to the distribution of nodal degree values, the transmission speed of COVID-19 kept increasing at different ratios during each period. The highest transmission speed of COVID-19 occurred between the third and seventh periods (20.02.2020 – 31.03.2020). After the seventh period, the spreading speed and effectiveness of the diseases remained almost stable. However, it was not fully controlled since the number of communities and nodal degrees increased between the seventh and fourteenth periods. Considering community detection results, nodal degree distributions, and weighted nodal degrees, the first four periods were critical for slowing the transmission of COVID-19 down. The average nodal degree increased rapidly after the third period. Therefore, it can be interpreted that the precautions were not sufficient in the early stage

of the pandemic since the disease still transmitted rapidly after the fourth period, especially within the sixth and seventh periods.

The nodal degree distribution should have exponential or positively skewed distribution to reduce the transmission speed. To achieve such distribution, it is needed to divide big communities and increase the number of isolated locations. According to the weighted nodal degree results, China was the epicenter of the diseases in terms of transmissibility for the first six periods. However, the epicenter switched to Italy after the sixth period and did not change for the rest of the periods. Considering the communities between the first and fifth period in which Italy existed, the rapid population in that community was actually a signal that indicates the epicenter was going to be a European country after China. This European country was going to be Italy since the country was the second key location during the fifth period of the pandemic. The fifth period was critical since the key locations are located in Asia, Europe, and the Middle East. Hence, it can be concluded that the disease was significantly effective and kept spreading in those regions. The analysis results according to weighted nodal degrees indicate that Iran was the third key location during the fifth period in addition to China and Italy. As seen from Figure 4, the disease was already spread to Eastern Asia, Middle East, and Africa. However, as seen from Figures 5-8, the transmission speed significantly increased in the fifth period and after, when Iran was the third key country. Considering all the above conditions, isolating the key locations by taking strong and accurate precautions may have considerably diminished the transmission speed, especially after the fifth period. On the other hand, the isolated nodes kept increased between the first and fourteenth periods. However, the increment in the isolated nodes may indicate the uninfected countries, recently infected countries, and/or countries that take necessary precautions, which resulted an isolation from an existed community. Therefore, supporting the conclusions above, the most critical period was the first four period of the epidemic since the outbreak could have been rapidly prevented by taking the optimal precautions.

It is concluded that the adverse effects of the COVID-19 pandemic could have been reduced significantly within four periods (in a month) if robust and accurate precautions (e.g., complete quarantine, lockdown, travel ban, etc.) were taken in locations whose nodal degree values were high. This study proposed a network-based community detection approach using international geographical distance to understand a potential future epidemic and isolating the key locations or dividing the dense communities before the disease spreads throughout the world by taking necessary precautions. Based on network analysis results, the real-time application of this model can be used to predict possible future spreading locations. Besides, it is estimated that this model will be able to provide effective predictions for other possible pandemics.

Future works may focus on the community detection of provinces and cities of a country to develop a model that can evaluate the key locations, number, and size of the communities to provide a better understanding of the transmission of COVID-19 in a country.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

- [1] Wu Z, McGoogan JM. Characteristics of and important lessons from THE Coronavirus DISEASE 2019 (COVID-19) outbreak in China. *JAMA* 2020;323:1239. [\[CrossRef\]](#)
- [2] Milano M, Cannataro M. Statistical and network-based analysis of Italian COVID-19 data: communities detection and temporal evolution. *Int J Environ Res Public Health* 2020;17:4182. [\[CrossRef\]](#)
- [3] Zhao PJ. A Social Network Model of the COVID-19 Pandemic. *medRxiv* 2020; 03.23.20041798. Preprint. doi: 10.1101/2020.03. [\[CrossRef\]](#)
- [4] So MKP, Tiwari A, Chu AMY, Tsang JTY, Chan JNL. Visualizing COVID-19 pandemic risk through network connectedness. *Int J Infect Dis* 2020;96:558–561. [\[CrossRef\]](#)
- [5] Wang P, Lu JA, Jin Y, Zhu M, Wang L, Chen S. Statistical and network analysis of 1212 COVID-19 patients in Henan, China. *Int J Infect Dis* 2020;95:391–398. [\[CrossRef\]](#)
- [6] Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci Total Environ* 2020;729:138817. [\[CrossRef\]](#)
- [7] Huang NE, Qiao F. A data driven time-dependent transmission rate for tracking an epidemic: a case study of 2019-nCoV. *Sci Bull* 2020;65:425–427. [\[CrossRef\]](#)
- [8] Shi S, Tanaka S, Ueno R, Gilmour S, Tanoue Y, Kawashima T, et al. Travel restrictions and SARS-CoV-2 transmission: an effective distance approach to estimate impact. *Bull World Health Organ* 2020;98:518–529. [\[CrossRef\]](#)
- [9] Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* 2020;582:389–394. [\[CrossRef\]](#)
- [10] Yi H, Ng ST, Farwin A, Pei Ting Low A, Chang CM, Lim J. Health equity considerations in COVID-19: geospatial network analysis of the COVID-19 outbreak in the migrant population in Singapore. *J Travel Med* 2021;28:taaa159. [\[CrossRef\]](#)
- [11] Javeed S, Anjum S, Alimgeer KS, Atif M, Khan MS, Farooq WA, et al. A novel mathematical model for COVID-19 with remedial strategies. *Results Phys* 2021;27:104248. [\[CrossRef\]](#)
- [12] Owusu-Mensah I, Akinyemi L, Oduro B, Iyiola OS. A fractional order approach to modeling and simulations of the novel COVID-19. *Adv Differ Equ* 2020;2020:683. [\[CrossRef\]](#)
- [13] Iboi E, Richardson A, Ruffin R, Ingram D, Clark J, Hawkins J, et al. Impact of Public Health Education Program on the Novel Coronavirus Outbreak in the United States. *Front Public Health* 2021;9:630974. [\[CrossRef\]](#)
- [14] Gibbs H, Liu Y, Pearson CAB, Jarvis CI, Grundy C, Quilty BJ, et al; LSHTM CMMID COVID-19 working group; Eggo RM. Changing travel patterns in China during the early stages of the COVID-19 pandemic. *Nat Commun* 2020;11:5012. [\[CrossRef\]](#)
- [15] Buddenhagen CE, Hernandez Nopsa JF, Andersen KF, Andrade-Piedra J, Forbes GA, Kromann P, et al. Epidemic network analysis for mitigation of invasive pathogens in seed systems: Potato in Ecuador. *Phytopathology* 2017;107:1209–1218. [\[CrossRef\]](#)
- [16] Wang X, Song B, Ni W, Ping Liu R, Guo J, Niu X, et al. Group-based susceptible-infectious-susceptible model in large-scale directed networks. *security and communication networks* 2019;2019:1657164. [\[CrossRef\]](#)

- [17] Miller JC. Percolation and epidemics in random clustered networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2009;80(2 Pt 1):020901. [\[CrossRef\]](#)
- [18] Brockmann D, Helbing D. The hidden geometry of complex, network-driven contagion phenomena. *Science* 2013;342:1337–1342. [\[CrossRef\]](#)
- [19] Iyiola O, Oduro B, Akinyemi L. Analysis and solutions of generalized Chagas vectors re-infestation model of fractional order type. *Chaos Solit Fractals* 2021;145:110797. [\[CrossRef\]](#)
- [20] Rosvall M, Axelsson D, Bergstrom CT. The map equation. *Eur Phys J Spec Top* 2009;178:13–23. [\[CrossRef\]](#)
- [21] Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006; 1695.
- [22] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–534. [\[CrossRef\]](#)
- [23] Jones J. H. Notes on R0. California: Department of Anthropological Sciences, 2007; 323.
- [24] Han J, Kamber M, Pei J. Advanced Cluster Analysis. *Data Mining* 2012; 497–541. [\[CrossRef\]](#)
- [25] Brandão MA, Moro MM, Lopes GR, Oliveira JPM. Using link semantics to recommend collaborations in academic social networks. *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*. 2013; doi: 10.1145/2487788.2488058. [\[CrossRef\]](#)
- [26] Largest Countries in the World (by area). In: *Worldometer*. Accessed 2 Jun 2020; <https://www.worldometers.info/geography/largest-countries-in-the-world/>.