IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 6, 2020, accepted November 16, 2020, date of publication November 19, 2020,
date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039260

Trajectory Linkage and Spreader Centrality for Social Epidemic Networks

IMAM MUSTAFA KAMAL[1], HYERIM BAE[2], (Member, IEEE), AND KI HAING CHO[3]
[1]Department of Big Data, Pusan National University, Busan 46241, South Korea
[2]Department of Industrial Engineering, Pusan National University, Busan 46241, South Korea
[3]Statistics and Big Data Division, Busan Metropolitan Government, Busan 1001, South Korea

Corresponding author: Hyerim Bae (hrbae@pusan.ac.kr)

ABSTRACT Constructing social networks for real epidemic cases is very challenging. Many mathematical models have been proposed to model such networks using simulation models, such as susceptible-infected (SI), susceptible-infected-recovered (SIR), and susceptible-exposed-infectious-removed (SEIR). Nonetheless, social network analyses can fail to capture real conditions, as such models are constructed based on many assumptions. Furthermore, unlike standard online social networks (OSNs), a social epidemic network requires different treatment from both model construction and social network analysis perspectives, especially for the detection of superspreaders. To address these issues, we propose a trajectory linkage method to automatically discover social networks from historical patient-trajectory data, wherein relations among patients are determined by spatial proximity and time-windows. Moreover, we introduce a novel spreader centrality measure that is devised to identify superspreaders in a social epidemic network. Extensive experiments were performed using real epidemic data. The results revealed that trajectory linkage can obtain a denser social network model than is possible by only incorporating patient data (the ''who is infected by whom'' relationship). By performing a social network analysis, the trajectory linkage model can express the real conditions of the patient relationship. Furthermore, our spreader centrality can capture the real superspreaders more effectively than can the existing centrality measure in social epidemic networks.

INDEX TERMS Social networks, epidemic networks, network centrality, spreader centrality, social network analysis.

I. INTRODUCTION

The modeling of social networks that are representative of the propagation of any of the many novel infectious diseases that may arise is an important research task [1]. The modeling of the epidemics on social contact networks has been studied in recent years. Many scholars use typical infectious models as social epidemic networks, such as susceptible-infectious (SI) [2], susceptible-infected-susceptible (SIS) [3], susceptible-infected-recovered (SIR) [4], and susceptible-exposed-infectious-removed (SEIR) [5]. The SI model is the simplest form of all epidemic models, and it categorizes the entire population into two groups, i.e., susceptible (S) and infectious (I). It also assumes that individuals are born into the simulation with no immunity (susceptible). Once infected, if there is no treatment, individuals remain infected and infectious throughout their lives and remain in contact with the susceptible population. In contrast, in SIS, the disease is transmitted only when a susceptible individual is in contact with an infected individual, and after infection, infected individuals return to the susceptible state. The SIR model was first used by Kermack and McKendrick in 1927. The components S, I, and R, respectively, represent the number of susceptible, infected, and recovered individuals in the population. Many diseases have a latent phase during which an individual is infected but not yet infectious. This delay between the acquisition of infection and the infectious state can be incorporated within the SIR model by adding a latent/exposed population, E, and letting infected (but not yet infectious) individuals move from S to E, and from E to I. Hence, the SEIR model is more suitable and comprehensive for the delineation of the information propagation mechanism. Those models are computationally simple, theoretically

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

tractable, and relatively easy for the fitting of observational data. Nonetheless, they can be limited by their fundamental assumption systems [6].

Unlike the aforementioned studies, in this paper, we propose a novel trajectory linkage that utilizes both spatial proximity and a time-window for the construction of social networks from the historical trajectory of patient data. We utilized a trace from each patient's data obtained during the outbreak of COVID-19 in South Korea. The patient is linked if they meet each other within a given specific distance (especially the same place) and within a given time-window. Such a trajectory linkage results in undirected-graph social networks because the original infector is assumed to be unknown. We can also combine trajectory linkage with prior information, such as patient information (the ''who is infected by whom'' relationship), to construct directed and denser networks in the form of a hybrid model.

In addition to modeling, one of the important issues in the social epidemic networks is detecting the superspreader. The term superspreader refers to an individual who is particularly effective in transmitting infectious diseases or spreading information. In epidemiology, a superspreader is an infected organism that infects disproportionally more secondary contacts than others who are also infected with the same disease [7]. Incorporating this knowledge, we propose a novel centrality method to identify superspreaders in social epidemic networks. Our spreader centrality is inspired by the PageRank method [8]. Accordingly, a page is more important if it has more in-links, which are votes from important pages that have a higher weightage. In other words, our spreader centrality determines the importance of a node based on the number of its out-links.

The main contributions of this work are summarized as follows:

- We introduce a trajectory linkage algorithm to construct automatically social epidemic networks from the patient's historical trajectory data given radius and time-window constraints.
- We proposed spreader centrality, a new centrality measure dedicated to identifying superspreaders in social epidemic networks.
- From the extensive experiments conducted using the COVID-19 case in South Korea, we find that our social network representation obtained by trajectory linkage can result in a more complete and meaningful interpretation than existing patient relational data.

To the best of our knowledge, there has not been an approach incorporating radius and time-window in constructing epidemic networks. Moreover, unlike the existing spreader centrality measure, our spreader centrality considers the communication flow (the ''who is infected by whom'' relationship) among the patients, making it possible to identify the real superspreaders.

The rest of this paper is organized as follows. Section II describes some previous works related to modeling social epidemic networks and centrality measures. Section III provides a comprehensive explanation of the proposed trajectory linkage and spreader centrality method. Section IV presents the experimentation and a discussion of the method's performance. Finally, section V concludes and looks ahead to future research.

II. RELATED WORKS

In this section, several previous studies about modeling social epidemic networks and network centrality measure are described.

A. MODELING SOCIAL EPIDEMIC NETWORK

Modeling the social epidemic network is an important issue in this decade. To make a near realistic social epidemic model, some studies have proposed some complex methods, such as EpiRep, which adopt SI in dynamic networks [9]. ISIR [4], [10] considers the real heterogeneous contact networks among people in society rather than homogeneous ones, as in the standard SIR model. By adopting a nonlinear dynamical system (NLDS), which originated with the representation of the propagation of viruses on computer networks [11], the improved non-linear dynamical system (INLDS) [12] improved the SIS model by utilizing the probability of social contact. The authors in [3] improved SIS to enable it to consider the contact probability between nodes, as determined by their social distances and their degrees of activity. In [5], the authors adopted SEIR by introducing a recovery method for incomplete relationship data in a really complex social network according to statistical sampling. Moreover, some immunity strategies have also been added to epidemic modeling, such as in [13], where a hierarchical targeted immunization strategy was proposed to control epidemic spread in a crowd with modular and hierarchical social contact networks. In [14], the authors proposed a novel community-based immunization strategy for the selection of targeted immunization nodes based on optimization. Nonetheless, there remain many challenges to the creation of models that can capture real-world complex dynamics [6].

Related to our research, several studies have utilized the spatial context in social epidemic networks. In [15], social contact graphs are utilized to model different contact patterns within the population for the COVID-19 pandemic. The authors in [16] studied social ties between family members, friends, and neighbors, as well as the probability of obesity within those networks. The work in [17] focused on Los Angeles gang networks and the location of each gang, combining social and spatial methods to better understand violence and youth behavior. In [18], autism in California was examined, and the authors reported that children living in closer proximity to children with autism were more likely to be diagnosed with the disorder. The authors in [19] used social networks and spatial analytical methods simultaneously to model disease transmission. Social clustering, on its own, can be influenced by the shared environment [20], [21]. After the outbreak of equine influenza in Australia, social networks were constructed by combining contact-tracing data on horse movements with a distance matrix between all

premises holding infected horses. In addition, a proximity network was constructed based on a matrix of the distances between each pair of infected premises in the contact network. They obtained a distance cut-off dichotomized at a minimum of 5 km, while 15.3 km was the maximal network based on empirical research.

B. NETWORK CENTRALITY

The identification of superspreaders in a social epidemic network is a key problem affecting the design of an effective mitigation strategy against the spread of an epidemic disease. A corresponding strategy to identify spreaders can also be established to accelerate or hinder information dissemination, increase the exposure range of products, detect contagious outbreaks, and support the execution of early intervention strategies [22]. Hence, the identification of key spreaders in a network has become an important research issue in its own right. There are many classic centrality methods, such as the degree centrality [23], closeness centrality [24], and betweenness centrality [25]. While most of the current superspreader identification research is under the SIR model, the author in [26] proposed $R0$-adjusted centrality, which incorporates the $R0$ value of a node into the existing network centrality measure to quantify a node's importance from two perspectives, namely, the network topology and the amplification/attenuation of the intensity of disease spreading. In [27], the authors devised a method to detect nodes that are capable of exerting a strong influence over a huge multilayer network solely based on the local knowledge of a network's topology, for speed and scalability. The k-shell method proposed by Kitsak *et al.* [28] is the most widely used method [29] and utilizes a node's location as one of the essential factors in determining the most efficient spreaders. However, as the k-shell index does not provide sufficient information on the topological locations of nodes [30], the authors in [31] proposed the measurement of a normalized local centrality based on normalized local structure attributes. This model considers the topology of the local network around a node and the influence feedback of the node's nearest-neighbor nodes. The authors in [32] combined the local and global performances of nodes for the measurement of nodal spreading abilities.

Recently, several significant kinds of research addressing to identify the spreader nodes in various social networks have been conducted. Chen *et al.* [33] claim that influential nodes can be identified by extracting and synthesizing topology feature information of traditional centrality indices and spreading influence. Yang *et al.* [34] identify the influential nodes by incorporating the degree and clustering-coefficient of neighbor nodes. Berahmand *et al.* [35] incorporate the natural characteristics of complex networks to capture the spreader node. Wang *et al.* [36] identify the influential spreader by considering the weight neighborhood nodes in complex networks. Eeti *et al.* [37] enhance The so-called Temporal Threshold Page Rank Opinion Formation model (TTPROF) by incorporating temporal evolution to identify the influential node. Ahmad *et al.* [38] introduce a community-based hybrid approach for identifying influential nodes. Unlike the previous studies, we consider the communication flow (the ''who is infected by whom'' relationship) of each node by adopting the PageRank metric to measure the spreader rank. It is in line with the spreader term in epidemiology. Wherein a spreader is an infected organism that infects disproportionally more secondary contacts than others who are also infected with the same disease [7].

III. METHODOLOGY

Below, we describe our proposed data-driven method to automatically obtain a social network model based on the trajectories of patient data. After that, we explain our centrality measure devised for the identification of superspreaders in a social epidemic network.

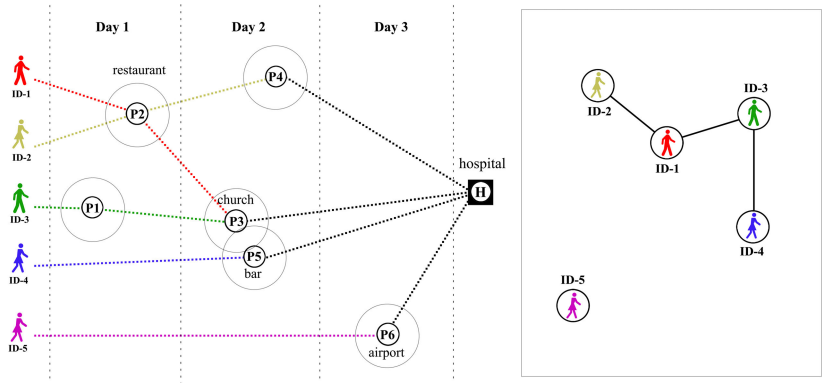
A. PROBLEM FORMULATION

Let us denote $G(V, E)$ as a graph G that consists of node V and edge E, where $V = \{v_1, v_2, \ldots, v_N\}$ and N is the total number of nodes in G. Each node has an edge with $E = \{(v_1, v_2), (v_1, v_3), \ldots, (v_{N-1}, v_N)\}$, which represents the set of links between corresponding nodes. For instance, if node v_1 is connected with node v_2, the value of (v_1, v_2) is 1; otherwise, it is 0. Note that there is no self-connection of v in G. In our case, a node represents a patient, where each patient has some visited places. We denote this as $v_i = \{P^i_1, P^i_2, \ldots, P^i_{M_i}\}$, where P^i_z corresponds to the z-th visited place of patient i and M_i represents the maximum number of visited places of patient i. Note that the number of visited places of each patient may differ. The visited place information includes the date of the visited place (dt), the place name (pn), as well as the longitude (lo) and latitude (la) attributes, and we can thus denote $P^i_1 = \{dt, pn, lo, la\}$.

B. TRAJECTORY LINKAGE

A trajectory linkage is a data-driven method for automatically discovering a social network model. Given the history of visited places by each patient (as shown in Fig.1a), it will construct a social network model (as depicted in Fig.1b). The link between two patients is determined based on their spatial proximity when visiting a place within a given time-window. We define our trajectory linkage output as an undirected and unweighted edge. The indirect relation among nodes is chosen because the first virus carrier among them is unknown. For instance, in the case of COVID-19, some symptoms are only detectable by a Lab. Therefore, the objective of the trajectory linkage method is to find relations among patients. An unweighted graph is desired because all patients are equally capable of spreading the virus.

The spatial proximity is determined by calculating the radius of the visited place between two patients. We utilize the Haversine formula [39] to determine the *radius* between two points on a sphere given their longitudes and latitudes. Meanwhile, in the time-window (day), we use the day as a minimum time unit.



(a) Patient trace data (b) Social network model

FIGURE 1. Discovering social network from trace data using Trajectory Linkage.

Algorithm 1 Trajectory Linkage

Input: $radius, day, M, A$

Output: M

Initialisation: M matrix $N \times N$ size with all value 0

1: **for** $i \in \{v_1, \ldots, v_{N-1}\}$ **do**
2: **for** $j \in \{v_2, \ldots, v_N\}$ **do**
3: **for** $k \in \{P^i_1, \ldots, P^i_{M_i}\}$ **do**
4: **for** $l \in \{P^j_1, \ldots, P^j_{M_j}\}$ **do**
5: **if** $i \leqslant j$ **then**
6: calculate R from $\{P^i_k.lo, P^i_k.la\}$ and $\{P^j_l.lo, P^j_l.la\}$
7: calculate D from $P^i_k.dt$ and $P^j_l.dt$
8: **if** $(R \leqslant radius)$ AND $(D \leqslant day)$ **then**
9: $M_{i,j+1} = 1$
10: **end if**
11: **end if**
12: **end for**
13: **end for**
14: **end for**
15: **end for**

The overall process of the trajectory linkage method is described in Algorithm 1. First, it requires two input parameters, namely *radius* as spatial proximity and day as a time-window. While the output is the adjacency matrix M, we initialize matrix M of size $N \times N$ with all values set to 0, which means that no node has any connection in the initial phase. The first and second loop represents the iteration to find the possible matchings between a pair of nodes. Note that in the first loop, the node is started from node 1 to $N-1$, whereas in the second loop, it begins from node 2 to N. This mechanism avoids overlapping matching. The third and fourth loops correspond to the iteration of each visited place for each corresponding node; because we want to generate an undirected graph, the link between nodes i and j is equal to the link between nodes j and i. Therefore, for simplicity, we conduct matching from node i to j only; this mechanism is conditioned on line 5 in the algorithm. The distance R is obtained by calculating the radius of the visited place between the longitude and latitude of nodes i and j. The time-window D in which two nodes meet is calculated by obtaining the absolute value of the difference between $P^i_k.dt$ and $P^j_l.dt$, as shown in lines 6 and 7, respectively. If the nodes i and j have met each other within a given less-or-equal-to radius and on an earlier-or-equal-to day, they will be connected by updating the value of matrix $M_{i,j+1}$ as 1. Note that i and j are equal in value. Therefore, in matrix M, it will be in index i and $j+1$ instead of index i and j.

Moreover, given $radius = 0$, which represents the exact same place where two patients meet, we can rank the visited place using our trajectory linkage algorithm. This can be realized by adding a list variable to save a frequent of the visited place of each patient after line 9 in Algorithm 1. Hence, by appending the $P^i_k.pn$ or $P^j_l.pn$ value to the variable, we can rank it based on the number of occurrences. The most frequently visited place means that many patients are infected because they visit this place. Therefore, this information can be used for prevention and mitigation purposes to minimize any increase in the rate of infection.

C. HYBRID

Besides using the trajectory linkage method, we can also construct a social network model using given ''patient information'' data. We refer to this as prior information. From prior information, we can approximately know (assume) the ''who is infected by whom'' relationship between patients, and we can, therefore, obtain a direct relation. Unfortunately, this will be only a very sparse connection (as will be discussed in Section IV). However, by combining the prior information with our trajectory linkage method, we derive the hybrid-undirected and hybrid-directed methods.

Obtaining a social network by a Hybrid-undirected method is very simple. Suppose matrices M and A represent the adjacency matrix by trajectory linkage and prior information, respectively. We can perform a special merging operation ($\oplus$) between M and A, as shown in Fig. 2. Note that M and A

$$\underset{H}{\begin{bmatrix}1&1&1\\1&1&0\\0&1&0\end{bmatrix}}=\underset{M}{\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}}\oplus\underset{A}{\begin{bmatrix}1&1&1\\0&1&0\\0&1&0\end{bmatrix}}$$

FIGURE 2. Matrix H is obtained from element-wise OR operation ($\oplus$) between matrix M and A.

are binary matrices; value 1 indicates that the two nodes are connected, while 0 means otherwise. The $\oplus$ indicates that we perform an OR operation in an element-wise manner of the matrices M and A. Therefore, if one of them indicates that the two nodes are connected, they will be connected.

For the hybrid-directed method, the mechanism is more complex than with the hybrid-undirected method. The basic idea of the hybrid-directed method is the use of given patient information as prior knowledge. We want to infer a direct relationship between patients. For instance, if patients i and j meet in the same place and at the same time, we want to know who is the first virus carrier among them. There are two mechanisms to tackle this issue. In the first mechanism, we can use matrix A as our prior knowledge. If patient i was ever infected by another patient before he/she met patient j, we can assume that patient j was infected by patient i. Unfortunately, this information is limited because matrix A is sparse (as will be discussed in Section IV. In the second mechanism, we can trace the history of places visited by patient i. If patient i ever met one of the infected patients before meeting patient j, we can assume that patient i infected patient j. We can obtain this information using the trajectory linkage algorithm 1. It can be realized by adding a variable to save the minimum day ($P^i_k.dt$) of the infected patient i. We use both mechanisms to obtain the day of infection for each patient. For simplicity, we denote this information as *dayInfected*.

The overall mechanism of the hybrid-directed method is described in Algorithm 2; the inputs are matrices M and A, while the output is matrix H. Note that M contains an indirect relationship, while A contains a direct relationship. From the input matrices M and A, we will obtain matrix H, which has a direct relationship. The condition in lines 5 to 6 represents obtaining a direct relationship from the ground truth matrix A. The condition in lines 8 to 9 corresponds to the inferring of a direct relationship based on matrix M and dayInfected. For instance, if patients i and j meet each other and the patient *dayInfected* of patient i is earlier than the day of their meeting (patient j meets patient i), we can assume that patient i infected patient j. Whereas the condition in lines 12 to 13 represents the case where they met each other and had never been infected before; therefore, we assume that they infected each other.

Algorithm 2 Hybrid-Directed

Input: M, A
Output: H
1: **for** $i \in \{v_1, \ldots, v_{N-1}\}$ **do**
2: **for** $j \in \{v_2, \ldots, v_N\}$ **do**
3: **for** $k \in \{P^i_1, \ldots, P^i_{M_i}\}$ **do**
4: **for** $l \in \{P^j_1, \ldots, P^j_{M_j}\}$ **do**
5: **if** $(A_{i,j+1} = 1)$ **then**
6: $H_{i,j+1} = 1$
7: **end if**
8: **if** $(M_{i,j+1} = 1)$ AND $(dayInfected(i) < P^j_M.dt)$ **then**
9: $H_{i,j+1} = 1$
10: **end if**
11: **if** $(M_{i,j+1} = 1)$ AND $(dayInfected(i) = P^j_M.dt)$ **then**
12: $H_{i,j+1} = 1$
13: $H_{j+1,i} = 1$
14: **end if**
15: **end for**
16: **end for**
17: **end for**
18: **end for**

D. PORTION OF CONNECTION

We introduce a portion of connection as P_c to analyze the connectivity among nodes (patients) in the social network model. P_c can be obtained by using a fraction between the number of connected nodes (N_c) and all possible connections (N_a), as denoted in Eq. 1. Note that the self-connection node is excluded in N_a, since there is no self-infected case in our problem. We can say that P_c will increase if the number of connections increases in our social network model. In Trajectory Linkage Algorithm 1, we can analyse the input variable's *radius* and *day* effects on relations between patients. For instance, is P_c increasing if we add the number of *radius*, and also, is P_c increasing if we add the number of *day*?.

$$P_c(radius, day) = \frac{N_c(radius, day)}{N_a} \tag{1}$$

E. SPREADER CENTRALITY

The term superspreaders refers to those who are particularly effective in transmitting infectious diseases [7]. In simpler terms, we could say that they infect more people than do most others and that they are therefore most responsible for the spread of disease. This becomes our basic foundation for the identification of superspreaders in a social epidemic network. The basic idea of our spreader centrality is that a patient will be highly ranked if he/she infects many patients and that the infected patients also infect many other patients, and so on, recursively, in our social epidemic network. In other words, the node will be important if it has a more important outlink. This mechanism is denoted in Eq. 2, and is illustrated in Fig. 3. Note that, even if it is inspired by PageRank, its traits and objectives are different. In PageRank, the node has a high score if it has many in-links. However, in our spreader

centrality, the node has a high score If it has many out-links.

$$r_j = \sum_{i \leftarrow j} r_i(1 + d_i) \tag{2}$$

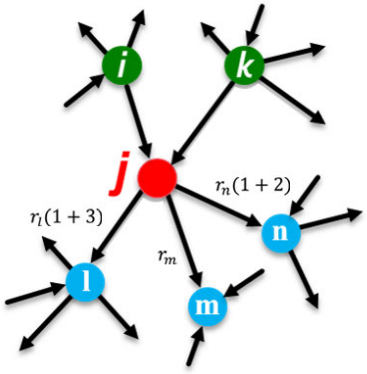


FIGURE 3. Spreader centrality.

where d_i represents the number of out-link and r_i spreader score of node i. As shown in Fig. 3, we can obtain the spreader score of node j by using equation $r_j = r_l(1+3)+r_m+r_n(1+2)$. We can initialize adjacency matrix C based on the spreader score. For instance, if node j has out-link i, then $r_j = r_i(1+d_i)$; otherwise, the index of nodes j to i will be valued as 0. Note that our idea was inspired by the PageRank method of determining the important node [8]. Therefore, the calculation of the rank that we utilize is the same mechanism as in PageRank. We utilize the power iteration method to obtain the spreader rank of each node. To guarantee convergence of the power iteration, matrix C should satisfy stochastic, aperiodic, and irreducibility conditions [40], [41]. Stochastic means that the matrix column sum must be 1. To realize this, we update Eq. 2 by normalizing it based on the column sum. We denote matrix C^* as a normalized form of the spreader score matrix C.

$$S = \beta \cdot C^* + (1 - \beta) \cdot \frac{1}{N} \tag{3}$$

The aperiodic and irreducibility properties can be satisfied by implementing a random surfer model, as denoted by Eq. 3. β represents the free contribution for all nodes to avoid spider traps (a node for which all out-links are within the group case) and dead ends (a node that has no out-links). This technique is known as the random surfer model, whereby in each step, the random surfer has two options: to follow a link randomly with probability β and to jump onto random nodes with probability $(1 - \beta)$. The optimal values of β are within the 0.8 to 0.9 range [8]. Finally, we have a matrix S, which already satisfies all of the aforementioned requirements. We can subject S to the power iteration method, as denoted in Eq. 4,

$$r^{(t+1)} = S \cdot r^t \tag{4}$$

where r is a vector of spreader rank, having length N. The index order of vector r represents the spreader rank of the node. For instance, index 0 of vector r represents the first rank of the superspreader. We can initialize $r^0 = [\frac{1}{N}, \ldots, \frac{1}{N}]^T$. After some iterations, we can obtain the spreader rank of each node. The process terminates when $|r^{t+1} - r^t| < \epsilon$.

IV. EXPERIMENT

The data used to develop our framework are described below. Then, we present various social epidemic network models and centrality measures, a discussion of the results, and their comparison.

A. COVID-19 DATASET

To validate the effectiveness of the proposed algorithm, it was evaluated using real datasets. We used COVID-19 infection cases in South Korea as our case study. The data were obtained from the KCDC (Korea Centers for Disease Control & Prevention), and is available at https://www.kaggle.com/kimjihoo/coronavirusdataset. The data are obtained using both GPS (cell phone inspection) and interviews with the corresponding patients. The spatial data were examined and recorded manually using the corresponding stack holder. Because the data were manually compared with the data recorded automatically by the system, we believe that these data are less vulnerable to noise compared with the data automatically acquired by the system. Furthermore, because the Korean data resulted in a very complex network, for simplicity in analyzing the network centrality, we utilized a Busan dataset. Actually, this data is a subset of the KCDC data but is from a different source, namely Busan Metropolitan City. There are two main datasets, *patient information* and *patient route* data. The *patient information* consists of attributes *patient id*, *infected by*, *symptoms on-site date* and *confirmed date*. The patient route corresponds to the history of places visited by each patient a few days after they have suspected symptoms. The *patient route* consists of *patient id*, *date*, *place name*, *longitude*, and *latitude*. The recorded data covers the period from mid-January until mid-May 2020.

We performed some data preprocessing because the use of raw data has some problems, such as matching issues between *patient information* and *patient route*. Sometimes, *patient id* exists in *patient information* but not in *patient route* data. Therefore we used only the *patient id*, which exists in both of the main datasets. Data incompleteness exists in *patient information* data. Some patients do not have *symptoms on-site date* data. This happens, because in some cases, COVID-19 patients display no symptoms, or the reason might be due to human error. We handle this condition by assuming that if the symptoms on-site date is empty, we should replace it by *confirmed date* data. Data incompleteness also exists in *patient route* data. For some patients, the only place visited is the hospital, with no *longitude* or *latitude* data. Therefore, those kinds of records were excluded. Finally, we omitted the hospital as a place visited by a patient because in reality, when people display symptoms, they tend to go to a hospital. Moreover, the last place visited by the patient must be a

hospital and represents the place where they are treated until cured. Therefore, the hospital may have many links. In reality, the hospital has high standards for how patients are treated during the COVID-19 outbreak. Hence, we assumed that the probability of spreading the virus in a hospital was very low.

B. SOCIAL NETWORK MODEL

The social epidemic network model can be constructed based on patient data, trajectory linkage, hybrid-undirected, and hybrid-directed. In the following, we present and compare all of them.

1) BASELINE

As a baseline, we utilized merely *patient information* data to construct the social epidemic network model. As shown in Fig. 4, this resulted in a very sparse network model, wherein the majority of the patients are not connected to anyone. Therefore, it was nearly impossible to conduct a social network analysis with this model. The network sparseness might have been caused by any of three factors. First, patients might not realize that he/she has met an infected patient in a given place. Therefore, the staff administrator cannot infer this information in recording *patient information* data. Second, an infected patient might be an overseas-inflow case. Such people have tested positive after visiting other countries. Third, with respect to the data incompleteness issue, data is sometimes not recorded, due to issues such as patient privacy or human error.

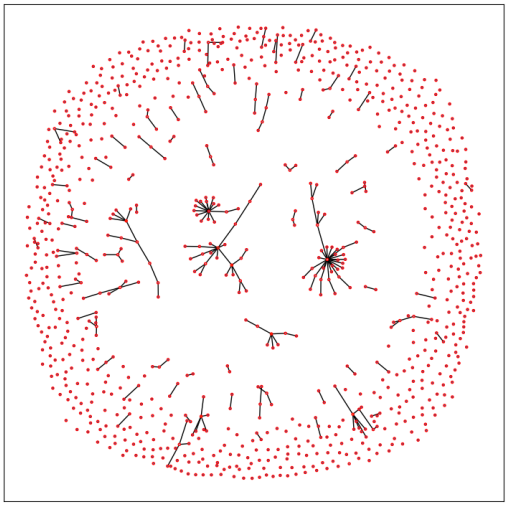


FIGURE 4. Social network model constructed by *patient information* data.

2) TRAJECTORY LINKAGE

The social network model obtained from the trajectory linkage method was denser than the baseline network, as depicted in Fig. 5. With $radius = 0$ (exactly the same place) and time-window $day = 14$, the P_c value of the baseline model was 4.32×10^{-5}, while that of trajectory linkage was $1,915 \times 10^{-5}$. Therefore, we can reveal that the portion of the connection of trajectory linkage was 443 times higher than that of the baseline model.

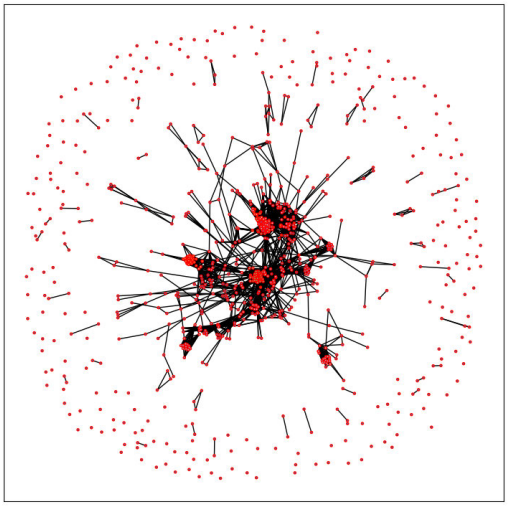


FIGURE 5. Social network model by Trajectory Linkage method, with $N_c(0, 14)$.

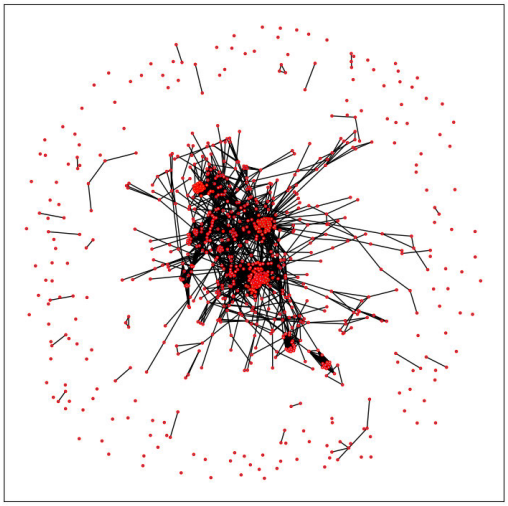


FIGURE 6. Social network model constructed by Hybrid-undirected method.

3) HYBRID

In the hybrid method, to obtain a still-denser social network model, we combined the patient information data and our trajectory linkage output. The hybrid method obtains two kinds of network models, namely, hybrid-undirected and hybrid-directed. The output of the hybrid-undirected method is shown in Fig. 6. Note that the P_c value of trajectory linkage is $1,915 \times 10^{-5}$, while that of hybrid-undirected is $1,917 \times 10^{-5}$. We can conclude that the hybrid-directed method adds only a few connections. Therefore, its P_c value was not significantly improved relative to that of the trajectory linkage output.

In analyzing how the disease spreads in the epidemic network, the directed connections have to be observed,

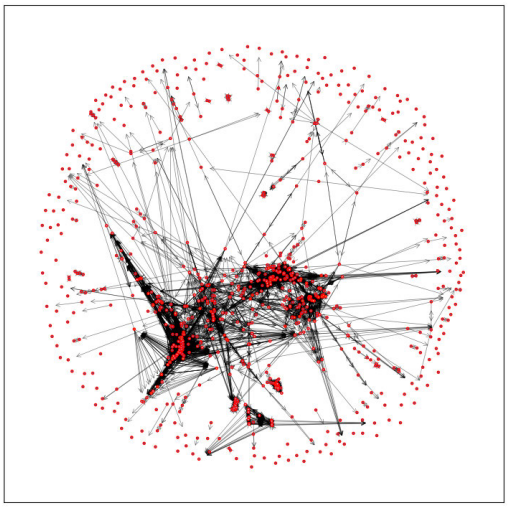


FIGURE 7. Social network model constructed by hybrid-directed method.

especially the out-links. Therefore, we also constructed a directed graph of the social epidemic network, as shown in Fig. 7. As is apparent, the connections are fewer than in trajectory linkage and hybrid-undirected with $P_c = 0.082 \times 10^{-5}$. This was caused by the ''infected before relationship'' condition, which narrowed down the connections among the patients. As described in section 2, a link between patients i and j can be established only if patient i has symptoms before he/she meets patient j or if neither of them has symptoms before they meet each other.

4) MODEL SUMMARY

As seen in Figs. 5 and 6, the social networks resulting from the use of the trajectory linkage and hybrid-undirected methods are not much different. The hybrid-directed method obtained fewer connections than the trajectory linkage and hybrid-undirected methods because there were more constraints to consider with regard to the meeting of two patients. Note that the baseline model was constant along the changing *radius* and *day* because it had been purely constructed from *patient information* data. By contrast, in trajectory linkage, there was a dynamic link among patients because it depended only on parameter *radius* and *day*.

We analyzed the social network (SN) topology generated by each method, as shown in Table 1. We can conclude that there is a decrease in the number of connected components, transitivity, and isolated nodes of the SN that are obtained from patient information, trajectory linkage, and the hybrid-undirected method. Therefore, the SN model, from patient information to the hybrid model, becomes denser. The network density of an SN obtained using the hybrid-directed method is lower than that of the hybrid-undirected method, while the number of isolated nodes exceeds that of the hybrid-undirected method. This is because there were more constraints to consider a connection with respect to the meeting of two patients, as shown in Algorithm 2. Note that with the exception of the SN generated using the hybrid-directed method, all of them are undirected graphs. Therefore, there are different criteria, e.g., the hybrid-directed method has strongly and weakly connected components.

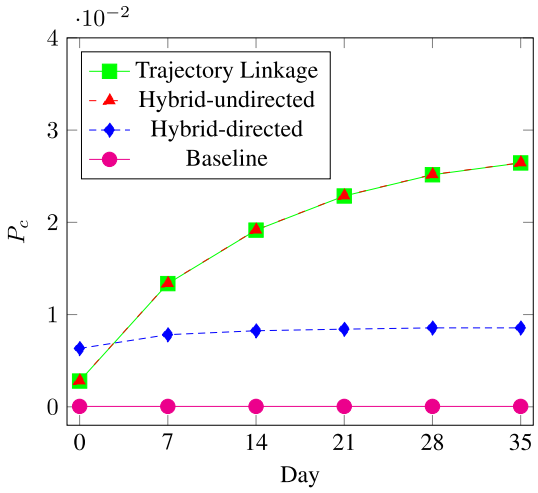


FIGURE 8. The effects changing time-window *day* while the radius is fixed radius $N_c(0, day)$ in Trajectory Linkage output.

We observe the values of parameter *radius* and *day* in our trajectory linkage method, and we analyzed the changes of the portion of connection P_c. With the *radius* fixed as 0, representing the exact same place, we changed the value of the time-window (*day*). As shown in Fig. 8, with an increasing number of *day*, the P_c value tended to increase. The interesting pattern there is that the number of patients sharply increases in the first week, whereas over the following 14, 21, 28, and 35 days, the difference is smaller. This represents the speed with which the corresponding government broadcasts the information. Actually, the government broadcasted warnings for citizens to avoid the corresponding place after some cases related to this place were officially detected. However, this information was delayed owing to several issues, such as the patient having symptoms but not going to a hospital (for test purposes), the test requiring at least one day to deliver the result, and the time required by the government to broadcast information to the citizen. Therefore, the total delay of the broadcast may be approximately one week. After obtaining this information, the government will lock down the corresponding place and conduct testing for each person who visited the corresponding place. Subsequently, citizens will avoid that place after being informed to do so. Ideally, the earlier the delivery of information to citizens, the less likely it will be their chances of being infected. As shown in Fig. 9, Busan was better in handling this issue than was the Korean nation overall. For example, whereas P_c was nearly flat after the first week in Busan, in the national dataset, it was still increasing. It can be specifically concluded that overall, the Busan government broadcasted information faster than the Korean government.

TABLE 1. Summary of network topological properties of each method.

Criteria	*Patient info* data	Trajectory linkage	Hybrid-undirected	Hybrid-directed
Connected components	732	259	182	-
Transitivity	0	0.8234	0.8266	-
Density	0.0005	0.0193	0.0196	0.0127
Isolates	651	215	156	176
Average degree	0.4435	18.045	18.4180	11.9060
Strongly connected components	-	-	-	383
Weakly connected components	-	-	-	219

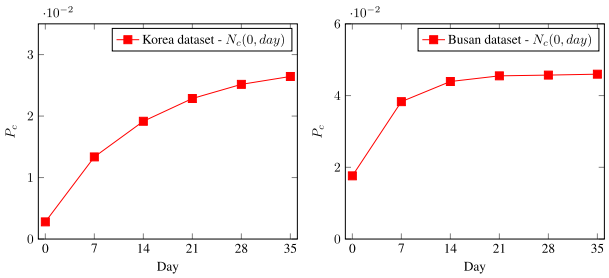


FIGURE 9. National (Korea) and Busan city comparison in handling COVID-19 with Trajectory Linkage model $P_c(0, day)$.

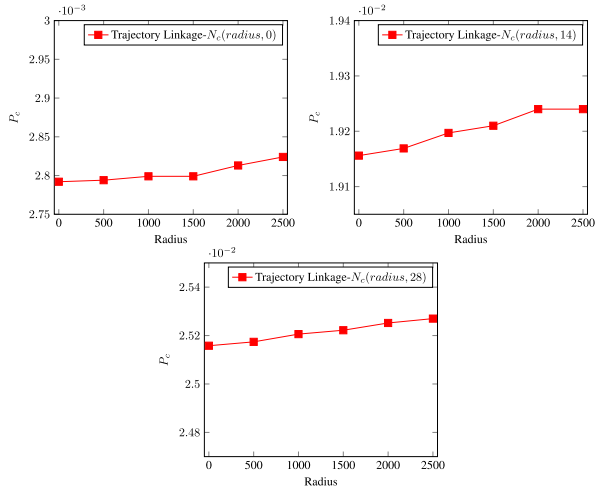


FIGURE 10. The effects changing *radius* while the *day* is fixed in Trajectory Linkage output.

We also observed the effects of varying *radius* under a fixed time-window *day* = 0,14,28. As shown in Fig. 10, the number of connected patients was not significantly increased or changed by increasing the *radius*. It, therefore, appears that the virus cannot live long and migrate to a nearby place. As revealed in [42], COVID-19 remains viable in aerosols for only 3 h. Furthermore, as shown in Fig. 10, in the model with an increasing number of days denoted as $N_c(radius, 0)$, $N_c(radius, 14)$, and $N_c(radius, 28)$, the value of P_c is increasing. Therefore, we can conclude that the parameter *radius* is dominated by the parameter *day*.

Starting from here, we will utilize the trajectory linkage model with $N_c(1.5, 14)$ for both the hybrid-undirected and hybrid-directed methods to conduct a social network analysis. Those parameter values were selected based on the Elbow method [43]. This represents the beneficial points that system designers have long selected to best balance inherent trade-offs [43]. For example, the higher the *radius* value, the more the patients tend to be linked to each other, as is likely to be the case in a crowded city. Meanwhile, the higher the *day* value, the more likely is it that false links will appear. In reality, the government will lock down a place after many related cases are reported. Consequently, citizens will avoid that place. Therefore, there have to be no new confirmed cases for that place unless, by coincidence, new patients infected at other places visit nearby corresponding places.

TABLE 2. Place (visited place) rank based on Trajectory Linkage on Busan data.

Rank	Place name
1	OC Church
2	SS Song Practice Station
3	SY Elementary School
4	Incheon Airport
5	Busan Station

By using our trajectory linkage method, we ranked places based on the frequency of visits by infected patients. As shown in Table 2, the place with the greatest superspreader effect was OC Church, followed by SS Song Practice Station, SY Elementary School, Incheon Airport, and Busan Station. This information can be used for mitigation measures such as the locking down of spreader places.

The running time of Algorithm 1 and 2 is 1205.1 sec. (20.085 min.) and 260.2 sec. (4.337 min.), respectively. Algorithm 2 requires a faster running time than Algorithm 1 because in Algorithm 2, there is no calculation of radius and time-window. Moreover, since it is dedicated to finding a direct relationship, it results in a fewer number of links compared with Algorithm 1. The implementation of this study is conducted using a computer with Windows Server 2012 R2 operating system, Intel Core i7 4790K CPU, 32 Gb

TABLE 3. Spreader rank based on Hybrid-undirected method on Busan dataset.

Rank	Patient id	Degree	Patient id	Betweenness	Patient id	Closeness	Patient id	Eigenvector
1	B-7	0.04670	B-15	0.09309	B-7	0.03187	B-7	0.06716
2	B-30	0.04670	B-47	0.08147	B-15	0.03187	B-30	0.06689
3	B-15	0.04396	B-7	0.07518	B-30	0.03081	B-15	0.06669
4	B-8	0.04121	B-30	0.06160	B-8	0.02934	B-8	0.06623
5	B-19	0.03846	B-1	0.04370	B-19	0.02844	B-19	0.06500

RAM, and Python (ver. 3.6.0). While the number of patients is 938, with the total number of visited places is 4,786.

C. CENTRALITY MEASURE

Different kinds of network models, such as undirected and directed graphs, employ different methods to calculate the centrality. The reasons for utilizing both of them are that the undirected graph can capture centrality based on the network topology, while in terms of the analysis of superspreader centrality in the epidemic network, a directed graph, especially its out-links, is essential. Therefore, using those two models, we can learn how a virus spreads in our society.

For our comparisons, we utilized some commonly used centrality measures, such as the degree, closeness, betweenness, and eigenvector centrality. The degree centrality of a node in a graph is simply a count of the number of edges that connect to it. The advantage of using degree centrality is that it can identify and rank superspreaders. In the closeness centrality, a node is considered important if it is relatively close to all other nodes. It can identify most rapidly the individuals who are best placed to influence the entire network. Based on the betweenness, a node is important if it lies in between many shortest-paths. This measure can indicate which nodes are ''bridges'' between nodes in a network. The eigenvector measures a node's influence based on the number of links it has to other nodes in the network. Therefore, it can be used to identify nodes with influence over the whole network. Nonetheless, in a social epidemic network, it may not be as effective.

1) UNDIRECTED GRAPH

In the present study, we were more interested in the superspreaders than in all of the nodes in the network. Thus, only the top nodes on the ranking list were considered. For this purpose, the top-5 nodes of each ranking algorithm were selected. As shown in Table 3, we could determine *B-7* as the top superspreader based on the degree, closeness, and eigenvector centrality. As shown in Fig. 11, Patient *B-7* was connected to all of the networks of the OC Church cluster as well as to other clusters.

Moreover, to identify the cluster of COVID-19 infection in Busan city, we utilize Clauset-Newman-Moore Greedy modularity maximization [44] to detect the community. As shown in Fig. 12, there are twelve communities detected, different colors represent the distinct community. The biggest cluster

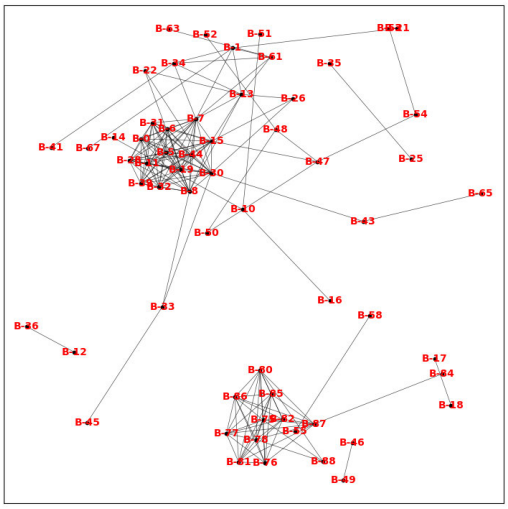


FIGURE 11. Undirected graph for centrality measure. It is constructed by Hybrid-undirected method.

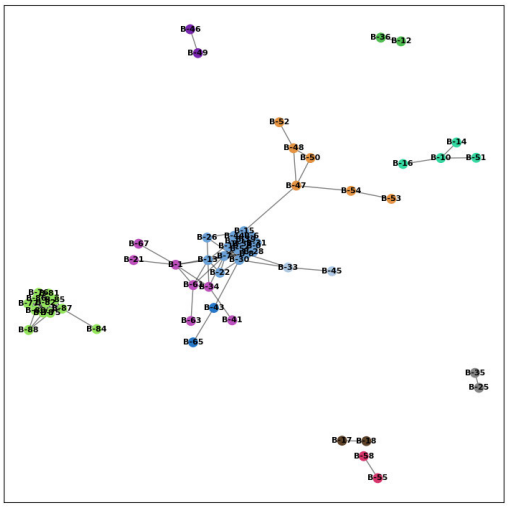


FIGURE 12. Twelve communities are detected using the Clauset-Newman-Moore greedy modularity maximization in the Busan dataset.

(blue node) related to the OC Chruch case, since the majority of the patients of this case, are lied on this community. In this community, from 16 members, 12 of them are related to the OC Chruch case. The second biggest community (green

TABLE 4. Spreader rank based on Hybrid-directed method on Busan dataset.

Rank	Patient id	Out-degree	Patient id	Betweenness	Patient id	Closeness	Patient id	Eigenvector	Patient id	Spreader centrality
1	B-7	0.07296	B-15	0.10070	B-15	0.27549	B-31	0.08820	B-30	0.05169
2	B-30	0.07296	B-47	0.08335	B-5	0.24794	B-5	0.08820	B-7	0.05101
3	B-0	0.05579	B-7	0.07767	B-31	0.24794	B-28	0.08810	B-76	0.04764
4	B-5	0.05150	B-30	0.06261	B-8	0.24445	B-15	0.07718	B-77	0.04764
5	B-6	0.05150	B-1	0.04352	B-28	0.24445	B-8	0.07473	B-75	0.04554

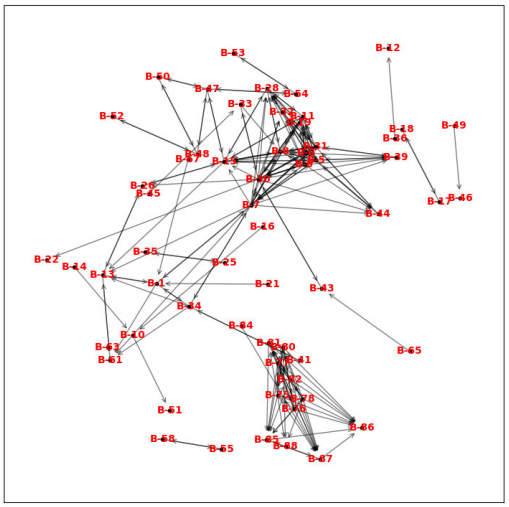


FIGURE 13. Directed graph for centrality measure. It is constructed by Hybrid-directed method.

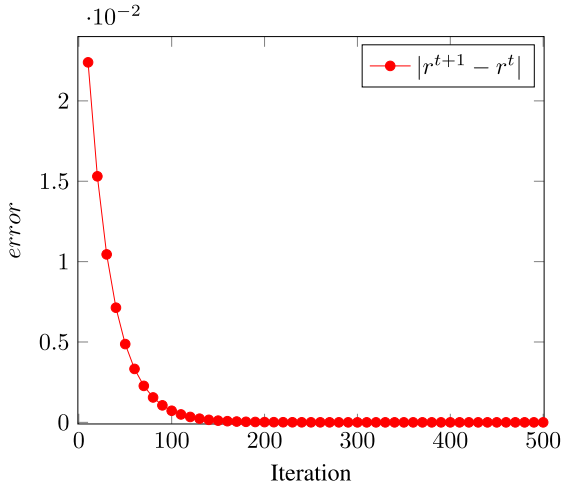


FIGURE 14. A convergence of calculation Spreader centrality in power method iteration.

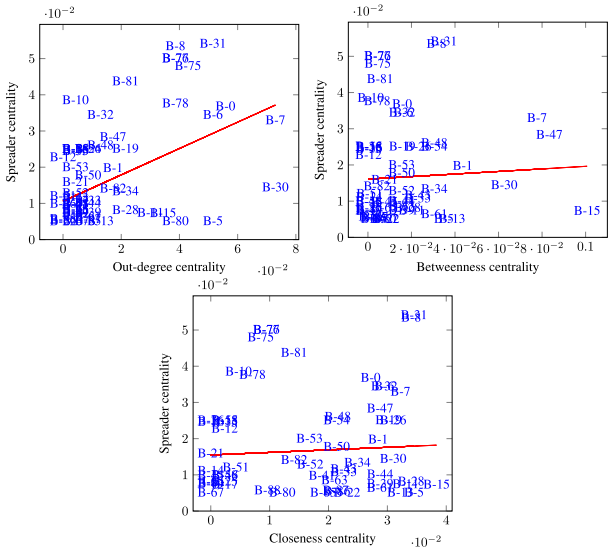


FIGURE 15. Correlation between Spreader centrality and classic centrality (Out-degree, Betweenness, and Closeness centrality).

color) consists of 12 members, while 7 of them are included as SY Elementary School case. While in other communities, they are mixed with each other in various cases in Korea.

2) DIRECTED GRAPH

In the directed graph, as shown in Table 4, our Spreader centrality identified *B-30* as the top superspreader. It was also related to the OC Church cluster. Rather than *B-7* based on the out-degree centrality, or *B-15* based on the betweenness and closeness, our proposed method chose *B-7* because it had fewer out-links than *B-30*. Note that *B-7* infected many patients both in the OC Church cluster and in other clusters. Therefore, compared with classic centrality measures, our proposed method can better capture the "who/whom" (infect/infected) relationship between nodes in a network.

Note that in the ranking, the order is more important than the values themselves. In this experiment, we set $\epsilon = 1 \times 10^{-5}$. In the power iteration, our spreader centrality calculation converged after iteration 130, as shown in Fig. 14.

We normalized each centrality measure to make the sum of the values equal to 1. Subsequently, we compared our spreader centrality with the classic centrality in a correlation graph, as shown in Fig. 15. As can be seen, our spreader centrality was highly correlated with the out-degree centrality. This was because the spreader centrality and out-degree centrality both considered out-links when calculating the rank of each node. This similarity was also shown in the top-2 spreader nodes, as shown in Table 3.

The spreader centrality is an eigenvector-based centrality that utilizes the power iteration method. Therefore, we also compared our proposed method with those of the same family, such as eigenvector, PageRank, and inversion of PageRank. Note that the inversion of PageRank is a PageRank in

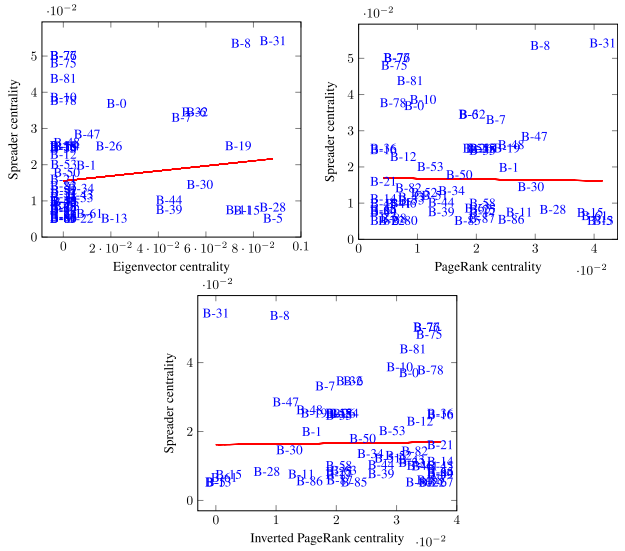


FIGURE 16. Correlation between Spreader centrality and power method based centrality (Eigenvector, PageRank, and inversion of PageRank).

which the rank value is inverted. For instance, the highest rank in PageRank will be the lowest rank in the inversion of PageRank. Note that our spreader centrality has the opposite objective to that of PageRank. Therefore, this mechanism ensures that our algorithm differs from the inversion version of the PageRank algorithm. As shown in Fig. 16, our proposed method did not correlate with any of them, not even with the inversion version of PageRank

V. CONCLUSION

The modeling of a social network model from epidemic data is important to gain an understanding of how a disease spreads in our society. Nonetheless, most of the previous studies construct social network models based on stochastic models with many assumptions. Unlike previous research, we derived and herein propose a data-driven method called the trajectory linkage method for the automatic discovery of social networks. The proposed method considers the spatial proximity and a time-window to isolate links between patients. In an experiment on COVID-19 datasets, the number of connected patients was increased as the length of the time-window was increased. In contrast, increasing the radius resulted in only small changes in the number of connected patients. The social network model constructed by the trajectory linkage method can capture relations between patients better than existing baseline data by obtaining a denser network. Furthermore, using the proposed method, we can rank the places most visited by patients, which represent dangerous places to visit. Such information is useful for disease-spread-mitigation purposes. Although the proposed method was originally devised for the construction of an undirected graph, using the baseline data, we can infer a directed graph as well.

One of the important tasks of social network analysis is the identification of centrality. However, over the past decade, there have been few reported studies on the identification of superspreaders in social epidemic networks. Therefore, we introduced our novel spreader centrality measure to identify the real superspreaders in social epidemic networks. Our spreader centrality is with a member of the PageRank centrality family. Hence, using the power iteration method, we formulated our spreader centrality equation to satisfy stochastic, aperiodic, and irreducibility properties to ensure applicability. Based on experiment results, our spreader centrality could capture the "who/whom" infect/infected relationship well because it considers communication among patients in a social network, specifically with respect to the transfer of disease among them. In the future, we can implement the trajectory linkage method for other epidemic cases. We believe that spreader centrality can be implemented in other domains, such as in the analysis of how rumors spread in social networks. Moreover, we realize that our current study has a limitation regarding the scalability of algorithm 1. The quadratic time complexity is inevitable since finding a link possibility between two patients requires a self-join operation, where we need to calculate the *radius* and time-window (*day*) among the patients. Nonetheless, we can incorporate distributed computing techniques such as MapReduce to deal with the large-scale dataset to tackle this limitation.

REFERENCES

[1] M. K. Parvez and S. Parveen, "Evolution and emergence of pathogenic viruses: Past, present, and future," *Intervirology*, vol. 60, nos. 1–2, pp. 1–7, 2017.

[2] T. Zhou, J.-G. Liu, W.-J. Bai, G. Chen, and B.-H. Wang, "Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 74, no. 5, 2006, Art. no. 056109.

[3] Y. Wang and W. Cai, "Epidemic spreading model based on social active degree in social networks," *China Commun.*, vol. 12, no. 12, pp. 101–108, 2015.

[4] Z. Zhang, H. Wang, C. Wang, and H. Fang, "Modeling epidemics spreading on social contact networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 3, pp. 410–419, Sep. 2015.

[5] B. Hu, J. Gong, and B. Hu, "Simulation of epidemic spread in social network," in *Proc. Int. Conf. Manage. Service Sci.*, Sep. 2009, pp. 1–4.

[6] J. Leitch, K. A. Alexander, and S. Sengupta, "Toward epidemic thresholds on temporal networks: A review and open questions," *Appl. Netw. Sci.*, vol. 4, no. 1, pp. 1–21, Dec. 2019.

[7] Y. Liu, B. Wang, B. Wu, S. Shang, Y. Zhang, and C. Shi, "Characterizing super-spreading in microblog: An epidemic-based information propagation model," *Phys. A, Stat. Mech. Appl.*, vol. 463, pp. 202–218, Dec. 2016.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the Web," Stanford InfoLab, Stanford, CA, USA, Tech. Rep. 1999-66, Nov. 1999.

[9] B. Shi, J. Zhong, Q. Bao, H. Qiu, and J. Liu, "EpiRep: Learning node representations through epidemic dynamics on networks," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, Oct. 2019, pp. 486–492.

[10] P. R. Thakare and S. S. Mathurkar, "Modeling of epidemic spread by social interactions," in *Proc. IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2016, pp. 1320–1324.

[11] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Trans. Inf. Syst. Secur.*, vol. 10, no. 4, pp. 1–26, Jan. 2008.

[12] Y. Wang, W. Cai, and R. Wang, "Propagation dynamics based on social contact probability on social network," in *Proc. Int. Conf. Estimation, Detection Inf. Fusion (ICEDIF)*, Jan. 2015, pp. 340–343.

[13] Z. Song, Y. Ge, L. Luo, H. Duan, and X. Qiu, "An effective immunization strategy for airborne epidemics in modular and hierarchical social contact network," *Phys. A, Stat. Mech. Appl.*, vol. 439, pp. 142–149, Dec. 2015.

[14] S. Wang, M. Gong, W. Liu, and Y. Wu, "Preventing epidemic spreading in networks by community detection and memetic algorithm," *Appl. Soft Comput.*, vol. 89, Apr. 2020, Art. no. 106118.
[15] M. Small and D. Cavanagh, "Modelling strong control measures for epidemic propagation with networks-a covid-19 case study," *IEEE Access*, vol. 8, pp. 109719–109731, 2020.
[16] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England J. Med.*, vol. 357, no. 4, pp. 370–379, Jul. 2007, doi: 10.1056/NEJMsa066082.
[17] S. M. Radil, C. Flint, and G. E. Tita, "Spatializing social networks: Using social network analysis to investigate geographies of gang rivalry, territoriality, and violence in los angeles," *Ann. Assoc. Amer. Geographers*, vol. 100, no. 2, pp. 307–326, Mar. 2010, doi: 10.1080/00045600903550428.
[18] K.-Y. Liu, M. King, and P. S. Bearman, "Social influence and the autism epidemic," *AJS; Amer. J. Sociol.*, vol. 115, 5, pp. 1387–1434, 2010.
[19] M. Emch, E. D. Root, S. Giebultowicz, M. Ali, C. Perez-Heydrich, and M. Yunus, "Integration of spatial and social network analysis in disease transmission studies," *Ann. Assoc. Amer. Geographers*, vol. 102, no. 5, pp. 1004–1015, Sep. 2012.
[20] S. M. Firestone, M. P. Ward, R. M. Christley, and N. K. Dhand, "The importance of location in contact networks: Describing early epidemic spread using spatial social network analysis," *Preventive Veterinary Med.*, vol. 102, no. 3, pp. 185–195, Dec. 2011.
[21] S. M. Firestone, R. M. Christley, M. P. Ward, and N. K. Dhand, "Adding the spatial dimension to the social network analysis of an epidemic: Investigation of the 2007 outbreak of equine influenza in australia," *Preventive Veterinary Med.*, vol. 106, no. 2, pp. 123–135, Sep. 2012.
[22] Y.-H. Fu, C.-Y. Huang, and C.-T. Sun, "Using global diversity and local features to identify influential social network spreaders," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2014, pp. 948–953.
[23] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Netw.*, vol. 1, no. 3, pp. 215–239, Jan. 1978.
[24] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, Dec. 1966.
[25] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977.
[26] T. Lee, H.-R. Lee, and K. Hwang, "Identifying superspreaders for epidemics using R0-adjusted network centrality," in *Proc. Winter Simulations Conf. (WSC)*, Dec. 2013, pp. 2239–2249.
[27] P. Basaras, G. Iosifidis, D. Katsaros, and L. Tassiulas, "Identifying influential spreaders in complex multilayer networks: A centrality perspective," *IEEE Trans. Netw. Sci. Eng.*, vol. 6, no. 1, pp. 31–45, Jan. 2019.
[28] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Phys.*, vol. 6, no. 11, pp. 888–893, Aug. 2010.
[29] L. Jiang, X. Zhao, B. Ge, W. Xiao, and Y. Ruan, "An efficient algorithm for mining a set of influential spreaders in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 516, pp. 58–65, Feb. 2019.
[30] C. Salavati, A. Abdollahpouri, and Z. Manbari, "Ranking nodes in complex networks based on local structure and improving closeness centrality," *Neurocomputing*, vol. 336, pp. 36–45, Apr. 2019.
[31] X. Zhao, F. Liu, S. Xing, and Q. Wang, "Identifying influential spreaders in social networks via normalized local structure attributes," *IEEE Access*, vol. 6, pp. 66095–66104, 2018.
[32] D. Zhang, Y. Wang, and Z. Zhang, "Identifying and quantifying potential super-spreaders in social networks," *Sci. Rep.*, vol. 9, no. 1, Oct. 2019, Art. no. 14811.
[33] X. Chen, M. Tan, J. Zhao, T. Yang, D. Wu, and R. Zhao, "Identifying influential nodes in complex networks based on a spreading influence related centrality," *Phys. A, Stat. Mech. Appl.*, vol. 536, Dec. 2019, Art. no. 122481.
[34] Y. Yang, X. Wang, Y. Chen, M. Hu, and C. Ruan, "A novel centrality of influential nodes identification in complex networks," *IEEE Access*, vol. 8, pp. 58742–58751, 2020.
[35] K. Berahmand, A. Bouyer, and N. Samadi, "A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks," *Chaos, Solitons Fractals*, vol. 110, pp. 41–54, May 2018.
[36] J. Wang, X. Hou, K. Li, and Y. Ding, "A novel weight neighborhood centrality algorithm for identifying influential spreaders in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 475, pp. 88–105, Jun. 2017.
[37] Eeti, A. Singh, and H. Cherifi, "Centrality-based opinion modeling on temporal networks," *IEEE Access*, vol. 8, pp. 1945–1961, 2020.
[38] A. Ahmad, T. Ahmad, and A. Bhatt, "HWSMCB: A community-based hybrid approach for identifying influential nodes in the social network," *Phys. A, Stat. Mech. Appl.*, vol. 545, May 2020, Art. no. 123590.
[39] G. Brummelen, *Heavenly Mathematics:The Forgotten Art of Spherical Trigonometry*. Princeton, NJ, USA: Princeton Univ. Press, 2013.
[40] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, Apr. 1998.
[41] I. M. Kamal, H. Bae, L. Liu, and Y. Choi, "Identifying key resources in a social network using f-PageRank," in *Proc. IEEE Int. Conf. Web Services (ICWS)*, Jun. 2017, pp. 397–403.
[42] N. van Doremalen, T. Bushmaker, D. Morris, M. Holbrook, A. Gamble, B. Williamson, A. Tamin, J. Harcourt, N. Thornburg, S. Gerber, J. Lloyd-Smith, E. de Wit, and V. Munster, "Aerosol and surface stability of hcov-19 (sars-cov-2) compared to sars-cov-1," *Medrxiv*, Dec. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7217062/
[43] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a, 'kneedle' in a haystack: Detecting knee points in system behavior," in *Proc. 31st Int. Conf. Distrib. Comput. Syst. Workshops*, 2011, pp. 166–171.
[44] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 6, Dec. 2004, Art. no. 066111.

IMAM MUSTAFA KAMAL received the B.S. degree in informatics and computer engineering from the Electronic Engineering Polytechnic Institute of Surabaya-Sepuluh Nopember Institute of Technology (ITS), Indonesia, in 2015. He is currently pursuing the Ph.D. degree with Pusan National University, Busan, South Korea. His research interests include machine learning, deep learning, and data science.

HYERIM BAE (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Industrial Engineering Department, Seoul National University (SNU), South Korea. He is currently a Professor major in industrial data science and engineering with the Industrial Engineering Department, Pusan National University (PNU), South Korea. He is currently leading BAB project (http://www.babcloud.org) which is an open source operational big data analysis tool. He is doing research projects with many large enterprises, including Samsung Heavy Industry, CyberLogitech, and Busan Port Authority, in the area of operational intelligence using Big data and process analytics techniques. His research interests include big data analytics, process analytics for industry 4.0, smart logistics, and smart factory. He is currently an Associate Editor of ICIC Express Letters, *International Journal of Innovation in Enterprise System*, and *International Journal of Innovative Computing, information and Control*.

KI HAING CHO is currently the Director of the Statistics and Big Data Division, Busan Metropolitan Government. As a Big Data Specialist, he oversees a various range of data-related functions of the Busan Metropolitan Government. He served as a Big Data Analytics Specialist at the City of Namyangu. Before joining the public sector, he worked with Korea Credit Bureau, CIGNA, AIA, and Hyundai Card Company.