

# Human-centric Computing and Information Sciences

December 2021 | Volume 11



[www.hcisjournal.com](http://www.hcisjournal.com)



# Modelling Prevention and Control Strategies for COVID-19 Propagation with Patient Contact Networks

Yixuan Yang<sup>1</sup>, Fei Hao<sup>2,3,\*</sup>, Doo-Soon Park<sup>1,\*</sup>, Sony Peng<sup>1</sup>, Hyejung Lee<sup>4</sup>, and Makara Mao<sup>1</sup>

## Abstract

Due to the coronavirus disease 2019 (COVID-19) outbreak, there is an urgent need to research the spread of disease and prevention strategies. As the spread of COVID-19 is closely related to the structure of human social networks, there are a lot of existing works that use a topological structure to analyze the characteristics of spread. Several studies have proposed certain strategies to prevent COVID-19 by analyzing the topological structure of the contact network, but most of the existing works have focused on detecting dense groups such as cliques; however, as the clique is the densest subgraph, it is easy for it to be influenced when the data has noise or lacks some edges. To reduce the influences of noise or lacks of data, there is a concept of  $\gamma$ -quasi-cliques is considered in this paper.  $\gamma$ -quasi-cliques is less restrictive and denser than cliques, and it is thus more suitable for analyzing and detecting communities in social networks to identify the close contacts of patients and achieve timely control under high levels of epidemic prevention strategies. Therefore, this paper proposed an algorithm based on the traditional formal concept analysis method for detecting  $\gamma$ -quasi-cliques, and also designed a model for detecting and mining close contacts and sub-close (secondary) contacts in the patient's contact network. Consequently, manual intervention occurs in response to the asymptomatic close or sub-close contacts detected by this model, and nucleic acid testing and home isolation are performed to prevent the widespread of COVID-19. In our experiments, a real-life contact network is used to determine the ideal value of  $\gamma$  for the detection of quasi-clique, which is 0.6, and the results show the validity and feasibility of the model.

## Keywords

Epidemic Prevention, COVID-19, Contact Network,  $\gamma$ -Quasi-Clique

## 1. Introduction

With the global outbreak of coronavirus disease 2019 (COVID-19), the World Health Organization has defined the pandemic as a public health emergency of international concern. In human societies, the spread of disease is influenced by three key factors: the biological structure of the virus, the social structure of the society, and the individual's knowledge of the virus. The spread of the virus reflects the movement of the population and the composition of its corresponding social networks. A social network

\* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

\*Corresponding Author: Doo-Soon Park (parkds@sch.ac.kr), Fei Hao (feihao@gmail.com)

<sup>1</sup>Department of Software Convergence, Soonchunhyang University, Asan, South Korea

<sup>2</sup>School of Computer Science, Shaanxi Normal University, Xi'an, China

<sup>3</sup>Department of Computer Science, University of Exeter, Exeter, UK

<sup>4</sup>Department of Innovation and Convergence, Hoseo University, Asan, South Korea

is a network system formed by social relationships between individual members of society. Many of the existing social network analysis is to collect user data through online social media for analysis [1–5], we can also obtain the social network constituted by people's actual geographic location through the user's GPS positioning, etc., to build a contact network. Existing studies [6] have established a network of contacts and interactions between the carriers of the virus, which is known as the contact network. In a contact network, nodes represent users and edges represent various types of contact relationships between users, which can represent strong family ties and connections between strangers [7, 8], as well as the geographical distance between people [9]. The virus can spread through human society through these connections. Analyzing a patient's contact network is key to elucidating the spread of the disease. Such analysis helps identify the transmission route of the virus, and it is an important basis for preventive measures. However, the current studies have mostly been based on retrospective epidemiological studies of COVID-19 patients, and there has been a lack of research into patient contact networks.

Before the COVID-19 outbreak, there had been several studies analyzing the spread of various viruses through contact networks, including sexually transmitted diseases (STDS) [7], acquired immune deficiency syndrome (AIDS) [10], and plague [8].

Many diseases contact networks, such as those in the cases described above, can be characterized as small-world networks [11]. Such contact networks have two characteristics depending on the location distance: one consists of dense local clusters and is formed by many local connections, and the other consists of many different local clusters in different regions and groups that are connected by the occasional long-distance linkage [12]. In such a network structure, even if each person has a limited number of contacts, the virus can still quickly spread throughout the network, thus causing a disease outbreak [7]. This characteristic leads to the spread of the virus beyond the original infection site, and it also leads to synchrotron oscillations that cause distant disease outbreaks [6]. The spread of COVID-19 in the contact network also has such characteristics.

Classical disease transmission models in social networks include the SI model, SIR model, SIS model, SEIR model, etc. The SEIR model considers the exposed state (E) on the basis of the SIR model. The susceptible state transfers to the latent state with the infection probability per unit time  $\beta$ , and the latent state transfers to the infection state with the infection probability per unit time  $\gamma$  [13]. Considering the incubation period of COVID-19 patients, this study classified COVID-19 as a SEIR model [14]. According to this model, individuals have five states in total: susceptible (S), exposed (E), infected-infectious (I), infected-asymptomatic (A), and recovered (R). The SEIR infectious disease model is illustrated in Fig. 1.



**Fig. 1.** The model of SEIR infectious disease.

With the COVID-19 virus as a backdrop, an increasing number of studies have shown that there are a certain number of asymptomatic carriers among COVID-19 patients [15], and preliminary evidence has proved their infectiousness [16]. Sun et al. [17] shows that asymptomatic persons infected with COVID-19 are also a possible source of infection, and that emphasis should be placed on isolating asymptomatic infected persons and protecting the surrounding people. The authors of [18] found that 54 of the 391 patients (about 13.8%) had no obvious symptoms. Qui [19] pointed out that the proportion of asymptomatic patients is very high, and that 30%–60% of infected persons are asymptomatic or have mild symptoms.

Due to a large number of asymptomatic infected persons, it is not safe to prevent outbreaks merely by controlling close contacts of confirmed patients, as some sub-close contacts may have been infected [20]. Therefore, we need nucleic acid testing and home quarantine not only for close contacts of confirmed

cases but also for certain sub-close contacts (i.e., people who have been in contact with close contacts, but who have not been in contact with a confirmed patient).

The commonly used method of close contact mining and detection is to discover cliques in the contact network and obtain the close contacts of confirmed users. However, this method has two disadvantages: (i) a clique is comprised of a completed connected graph with the most rigorous definition; the obtained group is the densest group, and it is susceptible to interference from data noise or missing information, and (ii) only detecting cliques will ignore those sub-close contacts who may have been infected. Therefore, in this paper, we use the concept of a quasi-clique, which has a relatively loose definition and a smaller constraint than cliques in mining the group.

In summary, to present a COVID-19 prevention and control model to improve the COVID-19 response, this paper considers the possible infection of some asymptomatic patients, close contacts and sub-close contacts through the contact network, and recommends they undergo nucleic acid testing and quarantine measures.

The contribution of this paper is as follows:

- A  $\gamma$ -quasi-clique enumeration method based on formal concept analysis (FCA) is designed. The FCA method and concept lattice are applied to the mining problem of  $\gamma$ -quasi-clique, the correlation between concept lattice and  $\gamma$ -quasi-clique is studied and discussed, some concepts in the concept lattice are filtered and screened according to the specific conditions, and the corresponding  $\gamma$ -quasi-clique is finally obtained.
- Since the quasi-clique is not a completed connected graph, some nodes which are not in the clique can also be classified into the group related to targeting users. This paper uses the unique advantages of the quasi-clique concept to design a model to track all possible infected close contacts and sub-close contacts from the entire contact network by using the algorithm mentioned above by providing a target confirmed patient.
- The real contact network is used in experiments to test the model proposed in this paper, and according to the results, we study the  $\gamma$ -quasi-clique quantities of the contact network under different threshold  $\gamma$  and determine the most appropriate threshold value. The experimental results show that the model performs best when the threshold value is 0.6.

The organizational structure of this paper is as follows: Section 1 introduces and analyzes the research background and motivation; Section 2 provides the definition of  $\gamma$ -quasi-clique and FCA as well as the existing research work. Section 3 presents the problem description and the proposed method, and it proposes and explains the algorithm used in this paper. Section 4 describes the experiments are conducted for evaluating the proposed algorithm. Section 5 presents the conclusion and suggestions for future work.

## 2. Related Work

### 2.1 $\gamma$ -Quasi-Clique

#### 2.1.1 Basic definition

Graphs are some of the most commonly used data structures, and they are often used to describe complex relationships [21]. Graph mining is an important research direction in the data mining field. Graph mining technology can be applied in the fields of biology, computers, social systems, and others, such as protein interactive network analysis, gene biological network analysis, community mining, social network analysis [22], academic map construction, etc.

Clique is an important concept in graph theory. In a graph, a clique is a set of vertices that can form a complete subgraph with the adjacency of any two points. It has the best connectivity and is the core of the dense substructure in the graph, and it contains centralized knowledge information. Clique mining is one of the most basic problems in the field of graph mining [23]. The concept of a clique can be used in many areas, such as k-clique community mining, movie recommendation systems [24], and so on.

Since the clique is the densest subgraph form, its definition is very strict, and all nodes must be connected in pairs. As a result, any data loss will seriously affect the results of clique mining, and the loss of edges in the graph caused by data noise or statistical errors will substantially reduce the value of any mined cliques; as a result, its use in practical problems is limited. Compared to clique, which has a strict definition, the definition of quasi-clique is relatively loose. Quasi-clique is a concept derived from the extension of the clique which was first proposed by Abello et al. [25]. Quasi-clique is typically defined in one of two ways. One is based on the definition of edge number, which requires that the proportion of the number of edges in the subgraph and the number of edges in the complete graph formed by its nodes be greater than the given threshold value [26]. The other is the definition based on degrees, which requires that the ratio of the degree of each node to the total number of nodes in the subgraph be greater than the given threshold [27].

The definition of quasi-clique adopted in this paper is based on degrees; that is, the degree of each node is not less than  $\gamma^*(n-1)$ , where  $n$  is the number of nodes in a quasi-clique, and  $\gamma$  is the parameter, which is typically in the range of  $(0.5, 1)$ , thus indicating the density of quasi-clique.

**Definition 1 ( $\gamma$ -quasi-clique).** Given a graph  $G$  and  $G = (V, E)$  if  $\forall v \in V, deg^G(v) \geq \gamma(|V| - 1)$ , then the graph  $G$  is a quasi-clique and the parameter is  $\gamma$ , by  $\gamma$ -quasi-clique. When  $V$  is the set of nodes and  $V = \{v_1, v_2, v_3, \dots, v_n\}$ ,  $E$  is the set of edges and  $E = \{(u, v) | u, v \in V\}$ .

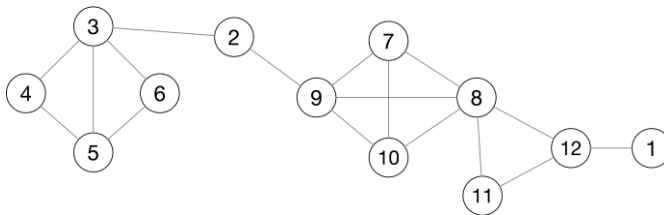
**Definition 2 (Maximal  $\gamma$ -quasi-clique).** For the graph  $G = (V, E)$  and point set  $X \subseteq V$ , if  $G(X)$  is a  $\gamma$ -quasi-clique and there is no point set  $Y \supset X$  that makes  $G(Y)$  a  $\gamma$ -quasi-clique, we call  $G(X)$  a maximal  $\gamma$ -quasi-clique.

The problem discussed in this paper is: Input graph  $G = (V, E)$ , and the degree threshold  $\gamma \in (0.5, 1)$ ; output all point sets  $X$  such that  $G(X)$ , and  $G(X)$  is the maximal  $\gamma$ -quasi-clique when  $\gamma$  is the minimum threshold.

A quasi-clique formed by two nodes can only consist of the two nodes and the edge connecting them, which can be directly obtained by enumerating all the edges in the graph. Since the quasi-clique of two nodes is simple and meaningless, it is equivalent to an edge in the graph, so the number of nodes in the quasi-clique discussed in this paper is greater than or equal to 3 by default.

Here is an example (Example 1) illustrating the  $\gamma$ -quasi-clique.

**Example 1.** Taking graph  $G$  in Fig. 2 as an example, if  $\gamma = 0.6$ , all  $\gamma$ -quasi-cliques in graph  $G$  are:  $\{\{3, 4, 5, 6\}, \{7, 8, 9, 10\}, \{8, 11, 12\}\}$ ; if  $\gamma \in [0.7, 1]$ , all  $\gamma$ -quasi-cliques in graph  $G$  are:  $\{\{3, 4, 5, 6\}, \{7, 8, 9, 10\}, \{8, 11, 12\}\}$ .



**Fig. 2.** An example of graph  $G$ .

### 2.1.2 Existing quasi-clique algorithms

The earliest publication on this topic is attributed to Abello et al. [25], who defined the concept of  $\gamma$ -quasi-clique and proposed greedy randomized adaptive search procedures (GRASP) for detecting large quasi-cliques in graphs illustrating telecommunications data. The Crochet algorithm, proposed by Pei et al. [26] and presented in a paper at the 2005 SIGKDD Conference, was the first algorithm for the maximal  $\gamma$ -quasi-clique enumeration problem. Based on the depth-first search framework, the algorithm constructs

a  $\gamma$ -quasi-clique depth-first search tree. Zeng et al. [27] proposed two new pruning strategies based on the Crochet algorithm proposed by Pei et al. [26], and also designed a Cocain algorithm based on Crochet. The runtime time of the Cocain algorithm is significantly less than that of the Crochet algorithm. The Quick algorithm proposed by Liu and Wong [28] is a high efficiency maximal  $\gamma$ -quasi-clique enumeration algorithm. In this algorithm, which is based on the other algorithms described above, five new pruning strategies are proposed, which greatly improves the efficiency of the algorithm. There have also been several other papers, some of which have used modified definitions of quasi-cliques, that have presented heuristic approaches to detecting large quasi-cliques in graphs for use in various applications.

## 2.2 Formal Concept Analysis

### 2.2.1 Basic definition

FCA is a typical data analysis method; it defines a set of formal concepts that represent the relationships between objects and attributes in an information system. These formal concepts are organized as concept lattices with a partial order. FCA has been widely used in data mining, machine learning, software engineering, data analysis, ontology, and other fields. In this paper, we apply this method for enumerating maximal  $\gamma$ -quasi-clique.

**Definition 3 (Formal context [21]).** Let  $F = (O, A, I)$  be a formal context, where  $O = \{x_1, x_2, \dots, x_n\}$  is the set of objects,  $A = \{a_1, a_2, \dots, a_m\}$  is the set of attributes, and  $I$  is a binary relation between  $O$  and  $A$ . Each  $x_i$  ( $i \leq n$ ) is called an object, and each  $a_j$  ( $j \leq m$ ) is called an attribute. If an object  $o$  has an attribute  $a$ , we note  $ola$  or  $(o, a) \in I$ . This is presented in Table 1.

**Table 1.** A formal context  $F$

$O \times A$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$	$A_7$	$A_8$	$A_9$	$A_{10}$	$A_{11}$	$A_{12}$
$O_1$	1	0	0	0	0	0	0	0	0	0	0	1
$O_2$	0	1	1	0	0	0	0	0	1	0	0	0
$O_3$	0	1	1	1	1	1	0	0	0	0	0	0
$O_4$	0	0	1	1	1	0	0	0	0	0	0	0
$O_5$	0	0	1	1	1	1	0	0	0	0	0	0
$O_6$	0	0	1	0	1	1	0	0	0	0	0	0
$O_7$	0	0	0	0	0	0	1	1	1	1	0	0
$O_8$	0	0	0	0	0	0	1	1	1	1	1	1
$O_9$	0	1	0	0	0	0	1	1	1	1	0	0
$O_{10}$	0	0	0	0	0	0	1	1	1	1	0	0
$O_{11}$	0	0	0	0	0	0	0	1	0	0	1	1
$O_{12}$	1	0	0	0	0	0	0	1	0	0	1	1

**Definition 4 (Operators  $\uparrow$  and  $\downarrow$  [21]).** Let  $F = (O, A, I)$  be a formal context. The operators  $\uparrow$  and  $\downarrow$  on  $X \subseteq O$  and  $Y \subseteq A$  are respectively defined as:

$$X^\uparrow = \{a \in A \mid \forall o \in O, (o, a) \in I\} \quad (1)$$

$$Y^\downarrow = \{o \in O \mid \forall a \in A, (o, a) \in I\} \quad (2)$$

$\forall o \in X$ , let  $\{o\}^\uparrow = o^\uparrow$ . For  $\forall a \in A$ , let  $\{a\}^\downarrow = a^\downarrow$ .

**Definition 5 (Formal concept [21]).** For a formal context  $F = (O, A, I)$ , if  $(O, A)$  satisfies  $O^\uparrow = A$  and  $A^\downarrow = O$ ,  $(O, A)$  is called a formal concept (FC),  $O$  is the extent of the concept, and  $A$  is the intent of the concept.  $FC(F)$  is an operation for obtaining a formal concept from formal context  $F$ .

**Definition 6 ([21]).** Let  $F(K)$  denote the set of all formal concepts of the formal context  $F = (O, A, I)$ . If  $(O_1, A_1), (O_2, A_2) \in FC(K)$ , then let

$$(O_1, A_1) \leq (O_2, A_2) \leftrightarrow O_1 \subseteq O_2 (\leftrightarrow A_1 \supseteq A_2) \tag{3}$$

Then, “ $\leq$ ” is a partial relation of  $F(K)$ .

**Definition 7 (Concept lattice [21]).** For a set of all concepts of the formal context  $F = (O, A, I)$ , a concept lattice,  $CL(F)$ , satisfies the following constraints:

$$FC(F) \subseteq CL(F), \text{ and } CL(F) = \sum FC(F) \tag{4}$$

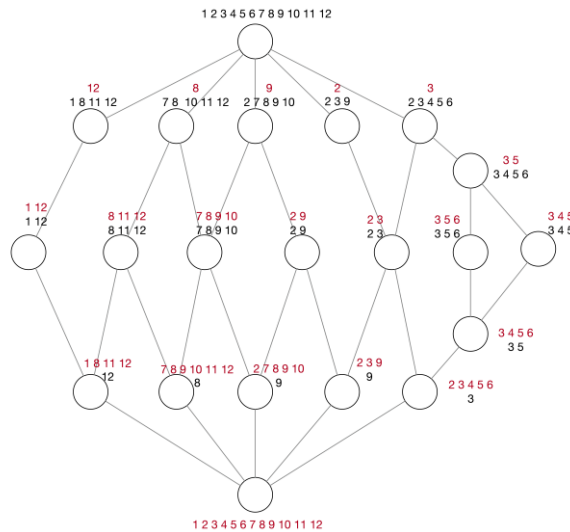
All concepts  $FC(F)$  can form a Concept Lattice  $CL(F) = (FC(F), \leq)$  by the partial ordered relation  $\leq$ .

Here, a graphical representation of the partial ordered relation  $\leq$  is presented in the form of a Hasse diagram. We provide an example to facilitate the understanding of a formal context and formal concept lattice.

Below, we provide an example to elucidate the generation process of the formal concept lattice.

**Example 2.** Table 1 presents a formal context  $F = (O, A, I)$ . The set of objects is  $O = (o_1, o_2, o_3, o_4, o_5, \dots, o_{11})$ , and the set of attributes is  $A = (a_1, a_2, a_3, a_4, \dots, a_{11})$ . “ $I$ ” represents the binary relationship between  $O$  and  $A$ . For example,  $I(o_1, a_{12}) = 1$  means that there is an attribute of  $a_{12}$  on the object  $o_1$ , and  $I(o_2, a_1) = 0$  means that object  $o_2$  has no attribute  $a_1$ . Further, the attributes “ $a_3$ ”, “ $a_4$ ”, “ $a_5$ ”, and “ $a_6$ ” have the common objects “ $o_3, o_5$ ”, so we can obtain a concept of  $((3, 5), (3, 4, 5, 6))$ .

The Hasse diagram shown in Fig. 3 illustrates the concept lattice for the context of Table 1. As shown in Fig. 3, each node indicates a concept, and the labelled numbers in black represent the intent of the concept, while the labelled numbers in red represent the extent of the concept.



**Fig. 3.** The concept lattice of formal context  $F$ .

### 2.2.2 Existing algorithms of FCA

The FCA theory was proposed by R. Wille [29] in 1982. As obtaining the concept lattice is an NP-completed problem, the algorithm for generating concept lattices is a challenge, and it represents an

important research issue in FCA. Ganter and Wille [29] proposed a Next Closure algorithm for obtaining the concept lattice corresponding to a given formal context. The authors of [30] proposed scalable algorithms for large datasets using Spark to save consumption. Yang et al. [31] proposed an FCA generation algorithm for the attribute-incremental formal context and a generation algorithm for the dynamic social network.

Beyond the proposed generation method of FCA, there have been many application cases to detect community and clique using FCA. For example, Hao et al. [32] constructed the formal context by taking the nodes in social networks as the objects and attributes of the formal context; importantly, the  $k$ -equiconcept in concept lattice is defined, and the equivalent theorem between the  $k$ -equiconcept and  $k$ -clique is presented for efficiently detecting  $k$ -cliques. They also applied FCA and the three-way concept for knowledge discovery in social networks [33] and COVID-19 knowledge navigation [34].

However, in this paper, we apply FCA to the detection and enumeration of  $\gamma$ -quasi-clique.

### 3. Quasi-Clique-based COVID-19 Virus Prevention Model

#### 3.1 Model Overview

To strengthen the prevention and control of COVID-19, this paper aims to establish a model based on  $\gamma$ -quasi-clique and FCA to determine whether certain users in the contact network are close contacts or secondary contacts of COVID-19 patients, and whether nucleic acid testing and home quarantine should be suggested.

Fig. 4 depicts a big picture diagram of the model. In this big picture, the first step is to preprocess the data set and convert it into a two-dimensional matrix (contact network) with a safe distance of  $Nm$  as the judgment condition. The second step is to call the  $\gamma$ -quasi-clique detection method provided in this paper to obtain all  $\gamma$ -quasi-cliques of the contact network. The next step is to enter the ID of some given confirmed cases (patients) and obtain their close and sub-close contacts users' ID. Finally, we will provide the nucleic acid testing and home quarantine suggestions based on the close and sub-close contacts users' ID.

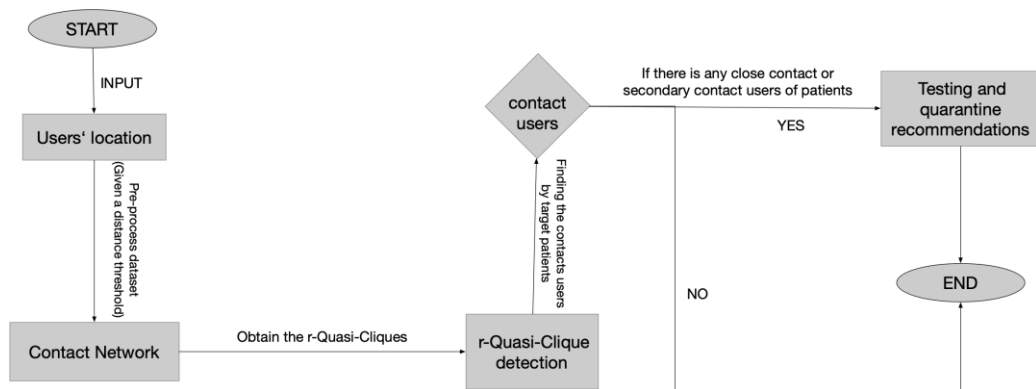


Fig. 4. A big picture diagram of the prevention and control of COVID-19.

#### 3.2 Model Implementation

##### 3.2.1 $\gamma$ -quasi-clique detection with FCA

Based on the basic knowledge in Section 2, we know that each node in a concept lattice represents a concept, and that the extent and intent of the concept contain certain relationships between object and attribute.



In Fig. 2, a social network  $G$  is formulated as a graph with the vertices indicating a set of individuals and the edges representing the binary relations between vertices. In this paper, we adopt the modified adjacency matrix to represent the formal context of graph  $G$ ; that is, Formal context( $G$ ) = ( $V, V, I$ ), where  $I$  is the binary relationship between two vertices.

**Definition 8 (Modified adjacency matrix [32]).** Let a social network be a graph with  $n$  vertices that are assumed to be ordered from  $v_1$  to  $v_n$ . The  $n \times n$  matrix  $M$  is called a modified adjacency matrix, in which

$$M(F) = \begin{cases} m_{ij} = 1, & \text{if } (v_i, v_j) \in I \text{ and } i \neq j \\ m_{ij} = 1, & \text{if } i = j \\ m_{ij} = 0, & \text{if } (v_i, v_j) \notin I \text{ and } i \neq j \end{cases} \quad (5)$$

According to formula (5), the social network  $G$  in Fig. 2 should be converted to a formal context  $F$ , as presented in Table 1. The formal concept lattice is shown in Fig. 3.

Here, it is obvious that when the extent and intent of the concept are both greater than 2, every concept is a  $\gamma$ -quasi-clique ( $\gamma \in (0, 1)$ , but we keep  $\gamma \in (0.5, 1)$ ).

We calculate and screen out the concepts conforming to our given threshold  $\gamma$ , after which we can obtain the corresponding  $\gamma$ -quasi-clique. In addition, regarding the symmetry of the social network, the modified adjacencies matrix is symmetric, and the concept lattice in this paper must be symmetric. Therefore, to maximize the efficiency of the algorithm, we only consider half of the concepts in the concept lattice.

Algorithm 1, which is used to calculate and screen out the  $\gamma$ -quasi-clique, is described as follows:

---

**Algorithm 1.**  $\gamma$ -quasi-clique detection by FCA

---

```

1: Input a threshold  $\gamma$ , formal context adjMat
2: Initialize the sets of QuasiClique, candidate
//QuasiClique is the set of  $\gamma$ -Quasi-Cliques //Candidate is the set of candidate
3: begin
4:  $FL \leftarrow$  FCA generation method( $G$ ) // Generating the formal concept lattice.
5: for each concept  $c \in FL$ :
6:   if  $c$ .extent.length > 1 and  $c$ .intent.length > 1 and  $c$ .extent.length >=
     c.inttent.length
7:     if  $c$ .extent.length == 2 and  $c$ .extent.equales( $c$ .intent) is True:
8:       pass
9:     else:
10:      candidate  $\leftarrow$  add( $c$ )
// Calculate the threshold value of whether the node is a quasi-clique in this concept
11: for each  $ce \in$  candidate:
12:    $NR \leftarrow \gamma * c$ .extent.length
13:   flag = TRUE
14:   for  $n \in ce$ :
15:     temp = -1
16:     for  $m \in ce$ :
17:       if adjMat[ $n$ ][ $m$ ]==1:
18:         tm +=1
19:       if  $NR > tm$ :
20:         flag= FALSE
21:       if flag:
22:         QuasiClique.add( $ce$ )
23: return QuasiClique
24: end
25: Output QuasiClique Set

```

---

In Algorithm 1, lines 1–2 give the input and initializes the set of QuasiClique and candidate; line 4 obtain the formal concept lattice by traditional FCA algorithm; lines 5–10 filter the candidate sets that meet the criteria; lines 11–23 travel the candidate set and compute the NR value of each concept, then determining whether the concept is a quasi-clique by comparing with the threshold value; line 25 output the result of the algorithm.

### 3.2.2 The model of prevention and control of COVID-19

According to the process in the big picture diagram given in Section 3.1, and the method based on FCA detection  $\gamma$ -quasi-clique that was designed in subsection 3.2.1, this section described the overall process of the COVID-19 prevention and control model in pseudocode. The procedure followed in this model can be described as in the following algorithm:

---

**Algorithm 2.** The model of prevention and control of COVID-19

---

```

1: begin
2: Pre-processing Data to generate a formal context  $F$ 
// Given different distance threshold, it will be obtained different formal context
3: Generating Formal Concept Lattice  $L$  of formal context  $F$ 
// Generating the formal concept lattice.
4: Detecting  $\gamma$ -Quasi-Clique from Concept Lattice  $L$  by Algorithm 1.
5: Input a patient ID
6:   if the ID in set of  $\gamma$ -Quasi-Clique:
7:     print all  $\gamma$ -Quasi-Cliques which has the ID
8:     //give a recommend result of Nucleic acid testing and Home quarantine for the
       users in above  $\gamma$ -Quasi-Cliques.
9:   else:
10:    print "there is not close contact users and secondary contact users"
11: end

```

---

In Algorithm 2, lines 1–2 do the pre-process for the dataset to suit for our method; lines 3–4 obtain the formal concept lattice and detect the  $\gamma$ -quasi-cliques by Algorithm 1; lines 5–11 input a target patient ID and detect the contact users or sub-contact users from the  $\gamma$ -quasi-cliques, then we will print the all users who are the contact ID or sub-contact ID.

## 4. Empirical Study

The experimental environment is a workstation equipped with the macOS Big Sur operating system, two Intel 2.3 GHz Quad-Core Intel Core i5 CPUs, and 16G RAM; the programming language is Python 3.8.

### 4.1 Dataset

In this paper, we use an open data set of human social interactions dedicated to modeling the dynamics of infectious diseases [9]. High-resolution GPS data from residents of the town of Haslemere, UK, was collected over three days, and in [9], analysis of this dataset shows that it is both structurally relevant to modeling disease transmission and has great potential for understanding and controlling the real-world disease.

The Haslemere dataset consists of pairwise distances of up to 1 m resolution between 469 volunteers from Haslemere, UK, at 5-minute intervals over three consecutive days (Thursday Oct 12 – Saturday Oct 14, 2017), excluding the hours between 11 PM and 7 AM.

In our experiment, we took the first-day data of only 469 volunteers and turned it into a contact network matrix. Fig. 5 shows a scatterplot of the data of the 469 volunteers on the first day.



Fig. 5. Scatterplot of 469 volunteers on the first day.

## 4.2 Experimental Results

### 4.2.1 Parameters configuration

According to relevant research, the most dangerous distance for COVID-19 transmission is 2 m, and the maximum distance across which droplets can splash in the air is 6 m, so we constructed two contact networks based on 2 m and 6 m.

Fig. 6(a) shows the contact network constructed with a safe distance of 2 m, and it has a total of 469 nodes and 1,135 edges; Fig. 6(b) shows the contact network constructed with a safe distance of 6 m, and it has a total of 469 nodes and 1,355 edges. It can be seen that the contact network with a safe distance of 6 m is denser than that in Fig. 6(a). To improve the level of epidemic prevention, we selected the contact network in Fig. 6(b) as the test dataset to apply for the model to experiment.

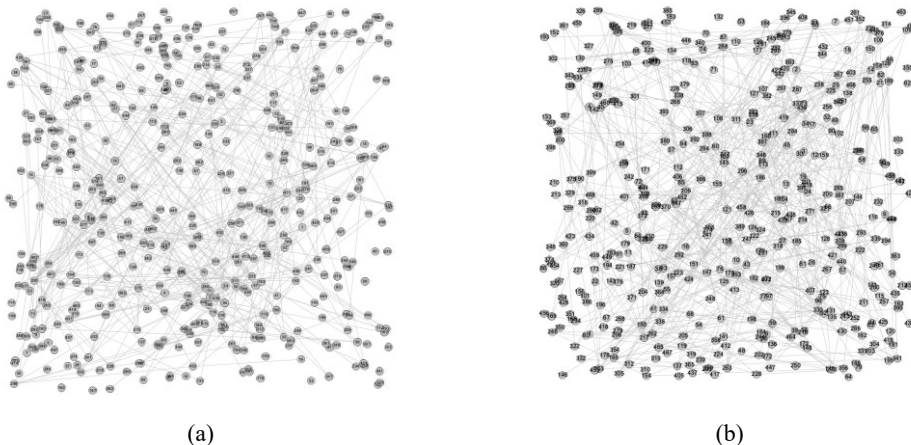


Fig. 6. The contact networks with different safe distances of the Haslemere dataset: (a) 2 m and (b) 6 m.

Obviously, Fig. 6(b), which uses a safe distance of 6 m is denser than Fig. 6(a), and more  $\gamma$ -quasi-cliques are detected, so that the prevention and control efforts in this case will be greater.

Table 2 lists the number of  $\gamma$ -quasi-cliques and the number of total nodes in all  $\gamma$ -quasi-cliques in the Haslemere contact network (with 6 m safe distance).

**Table 2.** Number of  $\gamma$ -quasi-cliques and nodes with varying parameter  $\gamma$ 

	$\gamma=0.5$	$\gamma=0.6$	$\gamma=0.7$	$\gamma=0.8$	$\gamma=0.9$	$\gamma=1.0$
Number of $\gamma$ -quasi-cliques	62	64	75	77	78	78
Total number of nodes in all $\gamma$ -quasi-cliques	133	133	133	133	133	133

According to the experiment results in Table 2, the smaller the value of  $\gamma$ , the higher the density of the quasi-clique. In other words, the smaller the  $\gamma$  value, the greater the number of close contacts and sub-close (secondary) contacts found by target users, and the greater the prevention and control efforts. Here, we provide an average quasi-clique density formula with which to evaluate the quality of the threshold.

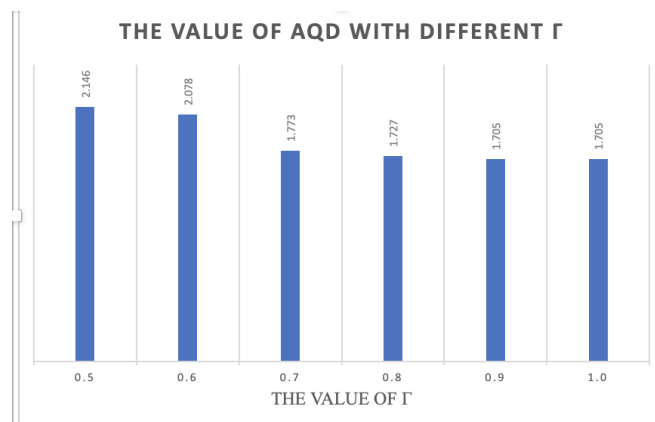
**Definition 9 (Average quasi-clique density [AQD]).**

$$AQD = \frac{\text{The total number of nodes in all } \gamma\text{-Quasi-Cliques}}{\text{The number of } \gamma\text{-Quasi-Cliques}} (\gamma = 0.5, 0.6, 0.7, \dots, 1) \quad (6)$$

According to formula (6), we calculated the AQD with different  $\gamma$ , and the results are presented in Table 3 and Fig. 7. Since 0.5 and 1 are the two boundary values of  $\gamma$ , we do not consider the AQD values of 0.5 and 1.0; the AQD value is the largest when  $\gamma = 0.6$ , so 0.6 is considered to be the ideal threshold for  $\gamma$ .

**Table 3.** Number of AQD with varying parameter  $\gamma$ 

	$\gamma=0.5$	$\gamma=0.6$	$\gamma=0.7$	$\gamma=0.8$	$\gamma=0.9$	$\gamma=1.0$
AQD	2.146	2.078	1.773	1.727	1.705	1.705

**Fig. 7.** AQD with different  $\gamma$ .

#### 4.2.2 Analysis results

Fig. 7 shows all 0.6-quasi-cliques in the Haslemere contact network (Fig. 6(b)), where the number is the ID of the node in the contact network. The visualized nodes are marked in Fig. 8, where the blue nodes are the nodes contained in the 0.6-quasi-cluster.

When the user with ID 73 is a COVID-19 patient, the 0.6-quasi-cliques which contain ID 73 can be detected by our model as follows: (73, 109,330), (73, 217, 276, 330, 375, 416), (73, 268, 330), (72, 73, 414), (20, 39, 73, 330), (73, 315, 330), (73, 166, 330) (72, 73, 217, 301, 330), and (73, 217, 228, 276, 330, 375) as all contain user of the target user 73, users (416, 228, 166, 39, 72, 73, 330, 268, 109, 301, 276, 20, 375, 217, 315, 414) are advised to undergo nucleic acid testing and follow a home quarantine. Fig. 9 shows these nodes in orange.

- |                               |                               |
|-------------------------------|-------------------------------|
| (73, 109, 330)                | (23, 36, 239, 299)            |
| (83, 163, 380)                | (35, 99, 102, 153)            |
| (295, 318, 332)               | (57, 76, 87, 370, 425)        |
| (112, 311, 342)               | (215, 265, 342, 361)          |
| (15, 238, 371)                | (72, 73, 217, 301, 330)       |
| (35, 99, 102, 298)            | (12, 217, 330)                |
| (14, 31, 426)                 | (73, 217, 228, 276, 330, 375) |
| (216, 221, 292, 469)          | (131, 239, 407)               |
| (411, 426, 442)               | (221, 292, 341, 469)          |
| (73, 217, 276, 330, 375, 416) | (86, 127, 190)                |
| (73, 268, 330)                | (57, 90, 439)                 |
| (23, 239, 407)                | (137, 184, 439)               |
| (239, 285, 305)               | (217, 228, 283)               |
| (72, 73, 414)                 | (126, 153, 341)               |
| (20, 39, 73, 330)             | (83, 217, 330, 393)           |
| (174, 217, 239)               | (144, 157, 388)               |
| (35, 253, 400)                | (64, 280, 368, 404)           |
| (31, 90, 184, 439)            | (14, 215, 311, 342)           |
| (50, 403, 411, 426)           | (215, 265, 361, 375)          |
| (178, 263, 345)               | (14, 25, 289, 311)            |
| (73, 315, 330)                | (216, 269, 316)               |
| (23, 239, 255, 285)           | (110, 130, 352)               |
| (26, 30, 60)                  | (14, 307, 337)                |
| (73, 166, 330)                | (181, 189, 276, 378)          |
| (211, 239, 407)               | (32, 180, 182, 195)           |
| (110, 242, 325)               | (118, 129, 302)               |
| (35, 102, 298, 316)           | (144, 153, 411)               |
| (8, 10, 392)                  | (86, 90, 439)                 |
| (154, 164, 246, 356)          | (236, 239, 285, 407)          |
| (198, 199, 200)               | (14, 34, 307, 426)            |
| (123, 223, 449)               | (139, 149, 339)               |
| (95, 106, 107)                | (38, 149, 184, 439)           |

Fig. 7. List of all 0.6-quasi-cliques in the Haslemere contact network.

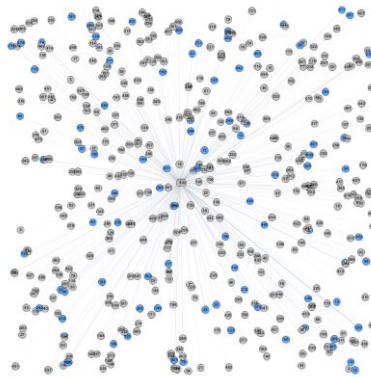


Fig. 8. All nodes in the 0.6-quasi-clique.

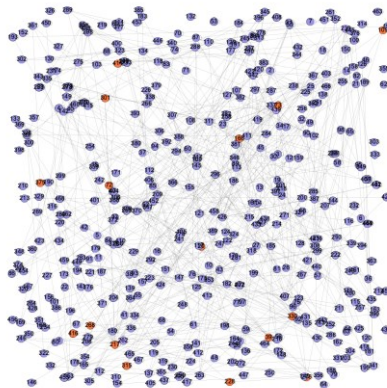


Fig. 9 The contact and sub-contact nodes of the node ID 73 in the 0.6-quasi-cliques.

## 5. Conclusion

In this paper, with the backdrop of the COVID-19 pandemic, an algorithm suitable for the detection of  $\gamma$ -quasi-cliques is designed using the traditional FCA method, and a model for detecting and mining close contacts and sub-close (secondary) contacts of patients in the patient's contact network is proposed. Asymptomatic close or sub-close contacts detected by the model should be responded to with manual intervention, and they should undergo nucleic acid tests and home quarantine to prevent the widespread of the COVID-19 virus. In future work, we will conduct further research using the approach of directed graphs as well as the simulation of virus propagation speed and evolution to identify the key nodes in the contact network.

### Author's Contributions

Conceptualization, YY, FH, DSP, SP, HL, MM. Funding acquisition, DSP. Investigation and methodology, YY, DSP, FH. Writing of the original draft, YY. Formal analysis, YY, DSP, FH. Data curation, YY, DSP, FH.

### Funding

This research was supported by the National Research Foundation of Korea (No. NRF-2020RIA2B5B01002134) and the BK21 FOUR (Fostering Outstanding Universities for Research) (No. 5199990914048), European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant (Agreement No. 840922), the National Natural Science Foundation of China (No. 61702317), and the Fundamental Research Funds for the Central Universities (No. GK202103080).

### Competing Interests

The authors declare that they have no competing interests.

## References

- [1] M. M. Liu, Q. C. Hu, J. F. Guo, and J. Chen, "Link prediction algorithm for signed social networks based on local and global tightness," *Journal of Information Processing Systems*, vol. 17, no. 2, pp. 213-226, 2021.
- [2] Z. Zhang, J. Jing, X. Wang, K. K. R. Choo, and B. B. Gupta, "A crowdsourcing method for online social networks security assessment based on human-centric computing," *Human-centric Computing and Information Sciences*, vol. 10, article no. 23, 2020. <https://doi.org/10.1186/s13673-020-00230-0>
- [3] J. Salminen, M. Hopf, S. A. Chowdhury, S. G. Jung, H. Almerexhi, and B. J. Jansen, "Developing an online hate classifier for multiple social media platforms," *Human-centric Computing and Information Sciences*, vol. 10, article no. 1, 2020. <https://doi.org/10.1186/s13673-019-0205-6>
- [4] C. Blundo, C. De Maio, M. Parente, and L. Siniscalchi, "Targeted advertising that protects the privacy of social networks users," *Human-centric Computing and Information Sciences*, vol. 11, article no. 18, 2021. <https://doi.org/10.22967/HGIS.2021.11.018>
- [5] H. Hong, B. Hu, and Z. Sun, "An efficient and secure attribute-based online/offline signature scheme for mobile crowdsensing," *Human-centric Computing and Information Sciences*, vol. 11, article no. 26, 2021. <https://doi.org/10.22967/HGIS.2021.11.026>
- [6] Z. Yang, "Analysis of dynamic contact network of patients with COVID-19 in Shaanxi Province of China," *Scientific Reports*, vol. 11, article no. 4889, 2021. <https://doi.org/10.1038/s41598-021-84428-x>
- [7] P. S. Bearman, J. Moody, and K. Stovel, "Chains of affection: the structure of adolescent romantic and sexual networks," *American Journal of Sociology*, vol. 110, no. 1, pp. 44-91, 2004.
- [8] S. A. Marvel, T. Martin, C. R. Doering, D. Lusseau, and M. E. Newman, "The small-world effect is a modern phenomenon," 2013 [Online]. Available: <https://arxiv.org/abs/1310.2636>.

- [9] S. M. Kissler, P. Klepac, M. Tang, A. J. Conlan, and J. R. Gog, "Sparkling "The BBC Four Pandemic": leveraging citizen science and mobile phones to model the spread of disease," 2020 [Online]. Available: <https://doi.org/10.1101/479154>.
- [10] H. W. Jaffe, "The early days of the HIV-AIDS epidemic in the USA," *Nature Immunology*, vol. 9, no. 11, pp. 1201-1203, 2008.
- [11] M. Kupennan and G. Abramson, "Small world effect in an epidemiological model," in *Physical Review Letters*, vol. 86, article no. 2909, 2001. <https://doi.org/10.1103/PhysRevLett.86.2909>
- [12] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440-442, 1998.
- [13] R. Fam, Y. Wang, M. Luo, Y. Zhang, and C. Zhu, "SEIR-based COVID-19 transmission model and inflection point prediction analysis," *Journal of University of Electronic Science and Technology of China*, vol. 49, no. 3, pp. 369-374, 2020.
- [14] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. Oxford, UK: Oxford University Press, 1992.
- [15] X. Pan, D. Chen, Y. Xia, X. Wu, T. Li, X. Ou, L. Zhou, and J. Liu, "Asymptomatic cases in a family cluster with SARS-CoV-2 infection," *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 410-411, 2020.
- [16] Y. Bai, L. Yao, T. Wei, F. Tian, D. Y. Jin, L. Chen, and M. Wang, "Presumed asymptomatic carrier transmission of COVID-19," *JAMA*, vol. 323, no. 14, pp. 1406-1407, 2020.
- [17] H. Sun, M. Xu, and X. Xu, "Infection and prevention of COVID-19 in schools based on real-life interpersonal contact data," *Journal of the University of Electronic Science and Technology of China*, vol. 49, no. 3, pp. 399-407, 2020.
- [18] W. W. Sun, F. Ling, J. R. Pan, J. Cai, Z. P. Miao, S. L. Liu, W. Cheng, and E. F. Chen, "Epidemiological characteristics of 2019 novel coronavirus family clustering in Zhejiang Province," *Chinese Journal of Preventive Medicine*, vol. 54, no. 6, pp. 625-629, 2020.
- [19] J. Qiu, "Covert coronavirus infections could be seeding new outbreaks," *Nature*, 2020. <https://doi.org/10.1038/d41586-020-00822-x>
- [20] J. A. Firth, J. Hellewell, P. Klepac, S. M. Kissler, A. J. Kucharski, and L. G. Spurgin, "Combining fine-scale social contact data with epidemic modelling reveals interactions between contact tracing, quarantine, testing and physical distancing for controlling COVID-19," 2020 [Online]. Available: <https://doi.org/10.1101/2020.05.26.20113720>.
- [21] Y. Yang, F. Hao, B. Pang, G. Min, and Y. Wu, "Dynamic maximal cliques detection and evolution management in social Internet of Things: a formal concept analysis approach," *IEEE Transactions on Network Science and Engineering*, 2021. <https://doi.org/10.1109/TNSE.2021.3067939>
- [22] J. M. Almedros-Jimenez, A. Becerra-Teron, and M. Torres, M. (2021). The Retrieval of Social Network Data for Points-of-Interest in OpenStreetMap," *Human-centric Computing and Information Sciences*, vol. 11, article no. 10, 2021. <https://doi.org/10.22967/HICIS.2021.11.010>
- [23] J. Cheng, L. Zhu, Y. Ke, and S. Chu, "Fast algorithms for maximal clique enumeration with limited memory," in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Beijing, China, 2012, pp. 1240-1248.
- [24] P. Vilakone, K. Xinchang, and D. S. Park, "Movie recommendation system based on users' personal information and movies rated using the method of k-clique and normalized discounted cumulative gain," *Journal of Information Processing Systems*, vol. 16, no. 2, pp. 494-507, 2020.
- [25] J. Abello, M. G. Resende, and S. Sudarsky, "Massive quasi-clique detection," in *LATIN 2002: Theoretical Informatics*. Heidelberg, Germany: Springer, 2002, pp. 598-612.
- [26] J. Pei, D. Jiang, and A. Zhang, "On mining cross-graph quasi-cliques," in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, Chicago, IL, 2005, pp. 228-238.
- [27] Z. Zeng, J. Wang, L. Zhou, and G. Karypis, "Coherent closed quasi-clique discovery from large dense graph databases," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, 2006, pp. 797-802.
- [28] G. Liu and L. Wong, "Effective pruning techniques for mining quasi-cliques," in *Machine Learning and Knowledge Discovery in Databases*. Heidelberg, Germany: Springer, 2008, pp. 33-49.

- [29] R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts," in *Formal Concept Analysis*. Heidelberg, Germany: Springer, 2009, pp. 314-339.
- [30] R. K. Chunduri and A. K. Cherukuri, "Scalable formal concept analysis algorithms for large datasets using Spark," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 11, pp. 4283-4303, 2019.
- [31] Y. Yang, F. Hao, B. Pang, K. Qin, Z. Pei, and B. Li, "A quick algorithm on generating concept lattice for attribute-incremental streaming data," in *Proceedings of 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Zhangjiajie, China, 2019, pp. 2811-2816.
- [32] F. Hao, G. Min, Z. Pei, D. S. Park, and L. T. Yang, "K-clique community detection in social networks based on formal concept analysis," *IEEE Systems Journal*, vol. 11, no. 1, pp. 250-259, 2017.
- [33] F. Hao, Y. Yang, G. Min, and V. Loia, "Incremental construction of three-way concept lattice for knowledge discovery in social networks," *Information Sciences*, vol. 578, pp. 257-280, 2021.
- [34] F. Hao and D. S. Park, "CoNavigator: a framework of FCA-based novel coronavirus COVID-19 domain knowledge navigation," vol. 11, article no. 6, 2021. <https://doi.org/10.22967/HGIS.2021.11.006>