Future Generation Computer Systems 107 (2020) 408–417

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Diversified top-k maximal clique detection in Social Internet of Things

Fei Hao [a,b,*], Zheng Pei [c], Laurence T. Yang [d]

[a] Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an, China
[b] School of Computer Science, Shaanxi Normal University, Xi'an, China
[c] Center for Radio Administration & Technology Development, Xihua University, Chengdu, China
[d] Department of Computer Science, St. Francis Xavier University, Antigonish, Canada

ARTICLE INFO

Article history:
Received 27 January 2019
Received in revised form 1 September 2019
Accepted 6 February 2020
Available online 8 February 2020

Keywords:
SIoT
Maximal clique
Diversified top-k maximal clique
Formal concept analysis
Coverage of clique

ABSTRACT

Social Internet of Things (SIoT), an IoT where things are autonomously capable of establishing relationships with other smart objects related to humans, allows them to interact within a social structure based on relationships. Importantly, exploiting the social structures of smart objects in SIoT is important for supervision and management of various services. Diversified top-k maximal clique, as a novel social structure, can be used for anomaly detection, and smart community detection from SIoT. However, the scalability of the existing approaches for detecting diversified top-k maximal cliques is becoming a significant challenge faced in the big graph. To this end, this paper proposes a novel diversified top-k maximal clique detection approach based on formal concept analysis. Specifically, we firstly prove the existence of equivalence relation between maximal cliques and equiconcepts which are a class of special concepts where the extent and intent are the same. Based on this equivalence relation, an efficient and innovative approach based on formal concept analysis for identifying diversified top-k maximal cliques is then further presented. Finally, three real-world social network datasets are adopted in experiments for the validation of effectiveness of our approach in SIoT.

1. Introduction

The Internet of Things (IoT) exhibits variety great benefits to our lives, such as Smart Medical, Smart Campus and so forth. The smart objects in IoT are simple connected via its communication infrastructure. However, the advancement and popularity of social networks enable these smart objects are socially connected. There, a novel communication paradigm, called Social Internet of Things (SIoT) [1], is emerging. It is a socially IoT where the things are capable of building social relationships with other objects, autonomously with respect to humans. In this way, a corresponding social network of objects can be created [2]. The objectives of SIoT are to separate the two layers of people and things; to allow objects to form their social networks; to allow people to use rules to protect their privacy and access the result of autonomous inter-object interactions occurring on the social network of objects. Importantly, exploiting the topological structures of smart objects in SIoT is important for supervision and management of various services [3,4]. In addition, The structure of SIoT can be shaped as required to guarantee the network navigability, so as that the discovery of smart objects and services is performed effectively and the scalability is guaranteed like in the human social networks [2].

Graph, a main information representation model for various ubiquitous networks, such as SIoT, has been widely used for networked information modeling and analysis. With this model, a number of algorithms on topological structure mining (k-clique mining [5], k-clique community detection [6,7], and maximal cliques enumeration [8,9]) in graph have recently been investigated. The purpose of maximal clique detection is to find all maximal cliques from a given network. Importantly, enumerating the maximal clique has huge potential application value in graph modeling, sociology, biological science, and chemistry [10]. The relationships between smart objects in SIoT have the similar feature with users interaction in a social network. Smart objects can collaboratively communicate with context-aware SIoT applications for diversified purposes [11]. Formally, this type of communication between smart objects is related to the underlying structure, termed diversified maximal cliques. Considering the computational complexity of maximal clique enumeration, a recent work [12] presented a new topic about diversified top-k clique mining which can significantly save the amount of useful overlapped information. The big difference between the maximal clique enumeration and diversified top-k maximal clique search lies in finding k maximal cliques that have the largest number

* Corresponding author at: School of Computer Science, Shaanxi Normal University, Xi'an, China.
E-mail address: fhao@snnu.edu.cn (F. Hao).

https://doi.org/10.1016/j.future.2020.02.023
0167-739X/

of nodes but also overlapping with each other as less as possible [12]. Thus, exploring the diversified top-k maximal cliques from SIoT is urgently to be addressed. To this end, this paper addresses this problem and presents an efficient solution based on formal concept analysis.

1.1. Motivating example

Actually, the diversified top-k maximal clique detection can be deployed in various applications, such as motif discovery [13], anomaly detection, and community detection [14,15] in large-scale complex networks. Recently, the research on biological networks such as co-expression, and protein interaction has become a hot issue in the field of bioinformatics. This motivating example is described with Arabidopsis gene coexpression network (CEGs) for discovering the hidden motifs. Mathematically, CEGs can be represented as maximal cliques [16]. The working principle of motif discovery is to identify the large CEGs with low overlaps.

Obviously, the above problem can be transformed as diversified top-k maximal clique detection problem. To address this problem, some existing algorithms have been proposed. Intuitively, the random algorithm and max k-cover algorithm [17] are two straightforward solutions. But, these algorithms suffer the huge consumption of memory for keeping all maximal cliques. To cope with this shortcoming, Yuan et al. [12] developed a novel algorithm to maintain k candidates when detecting the maximal clique from a given graph.

Different from the existing algorithms, this paper aims to address the diversified top-k maximal clique detection problem and hardness of it from the formal concept analysis point of view. We prove an equivalence relation between maximal cliques and equiconcepts included in the formal concept lattice of the given graph, then all maximal cliques are founded based on this equivalence relation (Section 3); We explore the overlapped/non-overlapped relationships among the equiconcepts in the formal concept lattice (Section 4); We push the overlapped equiconcepts into the stack and obtain the Top-k cliques (Section 4). The major contributions are summarized as follows.

1.2. Contributions

- To address the challenging issue of diversified top-k maximal cliques detection from a graph, a novel technical solution is proposed for finding the diversified top-k cliques from a graph. The novelty of our solution lies in detecting the diversified top-k maximal cliques based on formal concept analysis methodology which can generate the searching index of maximal cliques efficiently.
- To figure out the relation between maximal clique and formal concept and further to construct the searching index for maximal cliques from a graph G, a significant finding on equivalence relation between maximal clique and equiconcept is presented. In other words, when an undirected graph G is given, the maximal cliques existing in G exactly match with the equiconcepts in the formal concept lattice of G. The formal proof is also provided.
- Based on various operations of the equiconcepts over the stack, we propose an efficient FCA-based approach for diversified top-k maximal cliques detection. With this approach, a wise-greedy algorithm on detecting the diversified top-k maximal clique is devised. The extensive simulation experiments are conducted for demonstrating that the proposed wise-greedy algorithm can maximize the coverage of cliques compared to the other two algorithms, *i.e.*, random and greedy algorithms. Especially, as k increases, the improvement on coverage of cliques increases significantly.

1.3. Paper organization

The rest of this paper is structured as follows. Section 2 provides the preliminaries knowledge of this research as well as the problem formulation about diversified top-k clique detection. Section 3 firstly provides an equivalence relation theorem between maximal clique and equiconcept. By virtue of the equivalence relation, a FCA-based diversified Top-k maximal clique detection approach and its corresponding algorithm are detailed in Section 4. Section 5 shows the experimental results and performance evaluation. Eventually, Section 6 concludes this paper including our future work.

2. Preliminaries and problem formulation

In this section, we firstly recall the preliminary knowledge about FCA methodology and its relevant important properties; then the problem about diversified top-k clique detection (DKC) is mathematically formulated.

2.1. Preliminaries

FCA is an alternative effective computational intelligence methodology for characterizing the relation between nodes of the graph. This section briefly revisits the basics of formal concept analysis (FCA) [18,19]. In which, formal context is an important input of FCA, it is formalized as a triple including objects, attributes and their relations [6,20]. Besides, two key operations are defined for extracting the common attributes/common objects for a given set of objects/attributes. With these two operations, a formal concept is further formed by a pair containing extent A and intent B, denoted as $H = (A, B)$ [20].

Definition 1 (*Concept Lattice*)**.** Given a formal context $K = (G, M, I)$, a concept lattice $L(G, M, I)$ can be formed by organizing all concepts in $T(K)$ according to the hierarchical partial order $\leq$.

The formal concepts are then organized with partial order by Hasse diagram, called concept lattice. For any two formal concepts $H_1 = (A_1, B_1), H_2 = (A_2, B_2)$ in Hasse diagram, we define $H_1 \leq H_2 \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$, then H_2 is the father concept of H_1, and H_1 is the son concept of H_2.

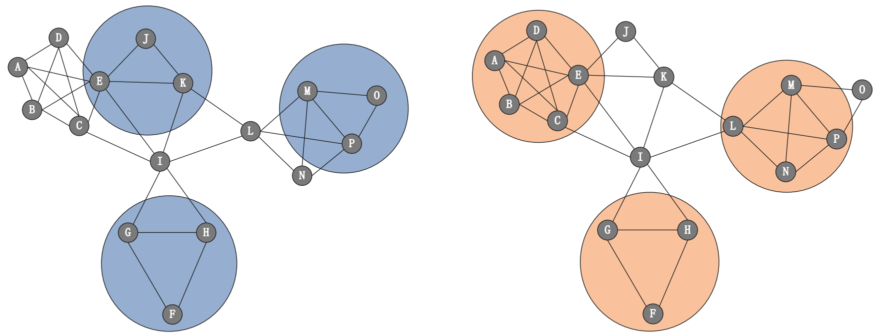
2.2. Problem formulation

In this section, we firstly revisit the related concepts of clique, maximal clique that are used throughout this paper. The formalism of the diversified top-k maximal clique detection problem (*DKC* detection problem) addressed in this paper is then mathematically described after introducing a new concept, named "*coverage of cliques*". To clarify the presentation, an illustrative case about the DKC detection problem is provided. At last, the idea proposed for solving this problem is briefly discussed.

2.2.1. Related concepts

Definition 2 (*Clique [6,20]*)**.** Given an undirected graph $G = (V, E)$ with V indicating the vertices and E indicating the edges between vertices. A clique c in G is a subset $S \subset V$ such that for any two vertices $v_i, v_j \in S$ there exists an edge $(v_i, v_j) \in E$. Obviously, a clique is a complete connected subgraph in the graph.

Definition 3 (*Maximal Clique [21]*)**.** A maximal clique is a clique that cannot be extended by including one more adjacent vertex. Formally, for a clique c if there exists no clique c' in G such that $c \subset c'$. Then, the clique c is called as maximal clique.



(a) Top-3 maximal cliques (b) Diversified Top-3 cliques

Fig. 1. An illustrative case for demonstrating the *DKC* detection problem.

Definition 4 (*Coverage of Cliques [12]*)**.** For a set of cliques $C = \{c_1, c_2, \ldots, c_m\}$ in an undirected graph $G = (V, E)$, the coverage of cliques of C, denoted as $Cov(C)$, is the set of nodes in G which can be covered by the cliques in C.

$$cov(C) = \bigcup_{c_i \in C} c_i \tag{1}$$

2.2.2. Problem descriptions

With the above related concepts, the formalism and descriptions of the *DKC* detection problem addressed in this paper are give as follows.

Input A graph G and an integer k;

Obj. ***Max*** $Cov(C)$;

St. $|C| \leq k$;

Output A set of maximal cliques C.

Clearly, this problem aims to find a set of maximal cliques C in which the maximal cliques c_i $(i = 1, 2, \ldots, k)$ can maximize the coverage of cliques $\bigcup_{c_i \in C} c_i$ in the graph G.

Proposition 1. *The DKC detection problem is NP-complete.*

Proof. To demonstrate the hardness of this problem, we prove the proposition by transforming and reducing the current problem to the existing well-known NP-hard complete problem, *i.e.*, connected set-covering problem [22,23]. The DKC detection problem can be solved using a baseline solution: (1) Enumerating the maximal cliques from the graph G; (2) Outputting the set of maximal cliques is transformed into a set-covering problem; note that the set in this problem refers to the different combinational set of maximal cliques. Since the set-covering problem as a subproblem of DKC detection problem, is a NP-hard complete problem. Therefore, the DKC detection problem is NP-complete as well. □

2.2.3. An illustrative case

Example 1. Let us see Fig. 1, a simple social network g including 16 nodes, *i.e.*, $A, B, C, \ldots, P$ and 31 edges. With the help of R language (*maximal.cliques*(g)), we can obtain 9 maximal cliques. However, Fig. 1(a) demonstrates a Top-3 maximal cliques: $c_1 = \{E, J, K\}$, $c_2 = \{G, H, F\}$, $c_2 = \{M, O, P\}$, then the set of Top-3 maximal cliques is denoted as $C_a = \{c_1, c_2, c_3\}$. Therefore, the $Cov(C_a) = |C_a| = 9$; Fig. 1(b) presents a diversified Top-3 cliques: $c_4 = \{A, B, C, D, E\}$, $c_2 = \{G, H, F\}$, $c_5 = \{L, M, N, P\}$, then the set of Top-3 diversified cliques is denoted as $C_b = \{c_4, c_2, c_5\}$. Consequently, the $Cov(C_b) = |C_b| = 12$.

3. On equivalence relation between maximal clique and equiconcept

This section presents one of the significant contributions about equivalence relation between maximal clique and equiconcept in this research. In order to pave a technical way for the efficient detection of diversified top-k maximal cliques, an equivalence theorem between maximal clique and equiconcept is provided and proved.

Before presenting the proposed equivalence theorem between maximal clique and equiconcept, the following relevant techniques are briefly provided.

- (**Technique 1:** Formal Context Construction of a Graph) The formal context of a given graph $g = (V, E)$, can be constructed by regarding the vertices as the objectives and attributes. For any $v_i, v_j \in V$, $I(v_i, v_j) = 1$ if $(v_i, v_j) \in E$, otherwise, $I(v_i, v_j) = 0$, formally described as $K = FC(g) = (V, V, I)$. Specially, $I(v_i, v_i) = 1$ for any $i \in \{1, 2, \ldots, n\}$. Hence, $FC(g) = (V, V, I)$ is a symmetry formal context. Formally, the construction of formal context of $g = (V, E)$ is equivalent to devising a modified adjacency matrix M from $g = (V, E)$ as follows:

$$M = \begin{cases} m_{ij} = 1 & if\ \exists e(v_i, v_j) \&\& i \neq j \\ m_{ij} = 1 & if\ i = j \\ m_{ij} = 0 & otherwise \end{cases}$$

  Clearly, $FC(g)$ is equivalent to the modified adjacency matrix of g, i.e., $FC(g) \equiv M$. Similar to the properties of M, $FC(g)$ also has the following properties:
- (**Technique 2:** Equiconcepts Extraction) Equiconcept is a special concept where the extent is exact the same as the intent, i.e, for a concept $H = (A, B)$ appearing in the corresponding generated concept lattice $L(V, V, I)$, if $A = B$, then this concept $H = (A, B)$ is an equiconcept. We denote a set of all equiconcepts with $EC(K)$.

In order to demonstrate the working process of the above critical techniques, the following example is then provided.

Example 2. By virtual of technique 1, we can easily construct the formal context of the following small social network g (as shown in Fig. 2), i.e, $FC(g)$ is formally represented as shown in Table 1. Then, we apply the Formal Concept Lattice generation algorithm to the above constructed formal context of g. Consequently, the concept lattice of $FC(g)$ can be generated as shown in Fig. 3. Obviously, there are 8 formal concepts generated via formal concept lattice generation algorithm. Interestingly, 3 concepts i.e., $(\{D, G\}, \{D, G\})$, $(\{B, C, E\}, \{B, C, E\})$, $(\{B, E, F\}, \{B, E, F\})$ are equiconcepts since they all have the same extent and intent.

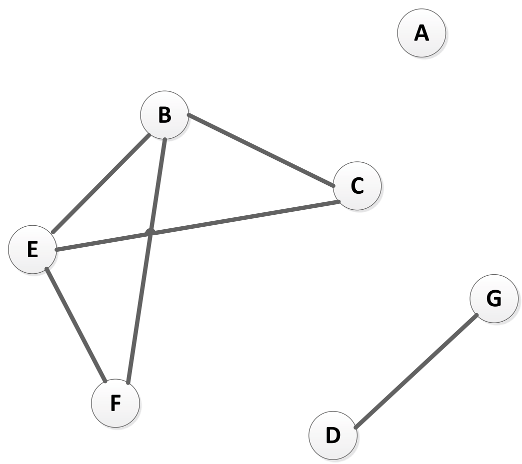


Fig. 2. A Social Network g.

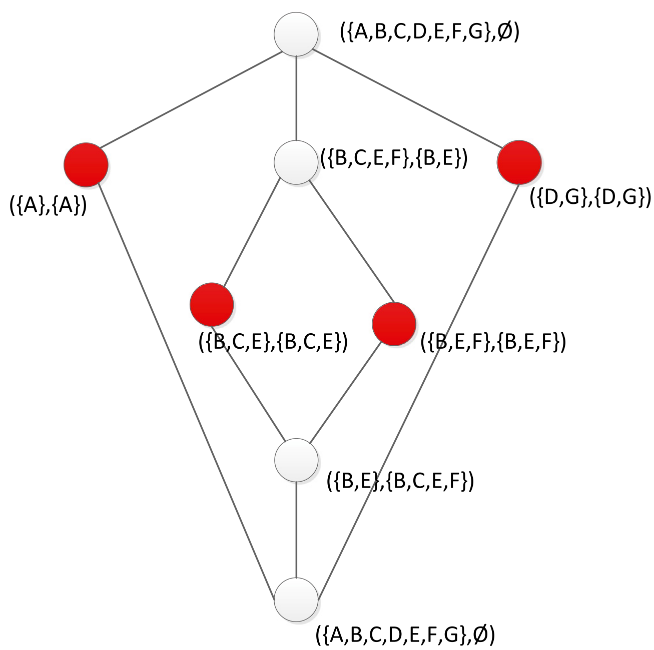


Fig. 3. Concept Lattice of $FC(g)$ (Remark: $(\{D,G\},\{D,G\})$, $(\{B,C,E\},\{B,C,E\})$, $(\{B,E,F\},\{B,E,F\})$) are the equiconcepts.

Table 1
Constructed formal context of g.

User	A	B	C	D	E	F	G
A	1	0	0	0	0	0	0
B	0	1	1	0	1	1	0
C	0	1	1	0	1	0	0
D	0	0	0	1	0	0	1
E	0	1	1	0	1	1	0
F	0	1	0	0	1	1	0
G	0	0	0	1	0	0	1

Theorem 1 (*Equivalence Theorem between Maximal Clique and Equiconcept*)**.** *Given a social graph $G = (V, E)$, the maximal cliques existing in graph G exactly match with the equiconcepts in the formal concept lattice of G. Formally, an equivalence relation holds.*

$$maximal.cliques(G) \equiv EC(K) \tag{2}$$

where K is the formal context of G; $maximal.cliques(G)$ returns a set of maximal cliques of G; $EC(K)$ indicates the set of equiconcepts with respective to the formal context K.

Proof. In order to validate the existence of this equivalence relation, we need to prove it towards two directions: (1) $maximal.cliques(G) \Rightarrow EC(K)$; (2) $EC(K) \Rightarrow maximal.cliques(G)$.

1. ($maximal.cliques(G) \Rightarrow EC(K)$): Given a complex network G, a maximal clique is a special clique in which all vertices are connected with each other. Suppose a maximal clique $\widehat{C}$ contains vertices $v_1, v_2, \ldots, v_k$, for any two vertices v_i, v_j, they connected with a link. Since a maximal clique is a subgraph, it is easily to construct its formal context using a modified adjacency matrix. Clearly, the formal context of a maximal clique is a matrix of 1's. With this all 1's formal context, we can extract a special concept $(\{v_1, v_2, \ldots, v_k\}, \{v_1, v_2, \ldots, v_k\})$, which satisfies $A = B$ (A is the extent and B refers to the intent of a concept, respectively). This special concept is an equiconcept. Hence, $maximal.cliques(G) \Rightarrow EC(K)$ holds.
2. ($EC(K) \Rightarrow maximal.cliques(G)$): According to the equiconcept definition shown in Technique 2, all equiconcepts $EC(K) = \{(A_i, B_i)|i = 1, 2, \ldots, r\}$ extracted from the generated concept lattice of given complex network g. Here, (A_i, B_i) is the i_{th} equiconcept where $A_i = B_i$. Let us assume one more vertex v_k be deleted or added into the extracted equiconcept (A_i, B_i), i.e., $(A_i \cup \{v_k\}, B_i \cup \{v_k\})$ and $(A_i - \{v_k\}, B_i - \{v_k\})$. With Definition 4, we can clearly see that the father concepts/son concepts of the concept (A_i, B_i) satisfying the properties: if $A_i \cup \{v_k\} \supseteq A_i$, then $B_i \cup \{v_k\} \subseteq B_i$, that is to say, we cannot find any father concepts so that the extent is the same with the intent. Similarly, if $(A_i - \{v_k\} \subseteq A_i$, then $B_i - \{v_k\}) \supseteq Bi$, we also cannot find any son concepts so that the extent is the same as the intent. By adding or deleting vertex from the equiconcepts, we cannot find any father/son concepts that equivalent to the cliques. Therefore, $EC(K) \Rightarrow maximal.cliques(G)$ holds.

Since we have already proved that $maximal.cliques(G) \Rightarrow EC(K)$ and $EC(K) \Rightarrow maximal.cliques(G)$ simultaneously, $maximal.cliques(G) \equiv EC(K)$ holds. □

Actually, a formal concept is a maximal pair of set of extent and its intent closed with Galois connection [24,25]. That is to say, the equiconcept as a special formal concept, is also a maximal pair of set of extent and its intent (Note that, a maximal pair is corresponding to a maximal clique). It is coinciding with the above proof.

Lemma 1. *Finding the number of maximal cliques in G is equivalent to detecting the number of equiconcepts $EC(K)$ from all formal concepts $T(K)$. Formally, it is described as*

$$N(maximal.cliques(G)) \equiv N(EC(K)) \tag{3}$$

where $N(.)$ is the function for outputting the total number.

Proof. According to Theorem 1, this proof is thus straightforward. □

Example 3. Let us continue using Example 1 for better understanding the proposed equivalence theorem. On one hand, we run the command "***maximal.cliques(g)***" in R language compiler. Eventually, we obtain 9 maximal cliques (e.g., red curly braces, such as $\{b, a, e, d, c\}$ and $\{j, e, k\}$) as shown in Fig. 4. On the other hand, a formal context of $g = (V, E)$ can be constructed, i.e., $K = FC(g) = (V, V, I)$ by using **Technique 1**. Then, the corresponding concept lattice $L(V, V, I)$ is built according to Definition 1. With **Technique 2**, the set of equiconcepts (e.g., pairs, such as $(\{g, h, i\}, \{g, h, i\})$ and $(\{l, m, n, p\}, \{l, m, n, p\})$) existing in L is denoted as $EC(K)$ as shown in Fig. 4.

As can be seen from Fig. 4, there exists an equivalence relation or one to one mapping from equiconcepts to maximal cliques. In other words, there exist 9 mappings that match the detected 9 equiconcepts with the mined 9 maximal cliques in the graph.

4. FCA-based diversified top-k maximal clique detection

Aiming to detect the diversified top-k clique from a graph, this section is dedicated to elaborating a novel approach for addressing the *DKC* problem based on formal concept analysis. Differs from the existing approaches, the proposed approach firstly explores the relationships among the equiconcepts (*e.g.*, maximal cliques) obtained with the above equivalence relation and devises a wise-greedy algorithm for detecting the diversified top-k clique from the given graph. Accordingly, the concrete detection method and its algorithm are presented as follows.

4.1. DKC problem based on FCA

Our proposed approach in this paper basically follows the idea of baseline solution mentioned in [12]. The originality of this approach lies in twofolds:

- Optimizing the complexity of searching the maximal cliques from a graph. Since we convert the maximal cliques detection problem into the set of equiconcepts detection in concept lattice, the procedure of enumeration is thus eliminated.
- A wise-greedy algorithm is devised for reducing the time complexity. Particularly, a stack data structure is adopted for obtaining and storing the diversified top-k cliques in this paper.

The working procedures of searching the diversified top-k clique based on FCA from a given graph g are described as follows:

4.1.1. Step 1-Extracting maximal cliques

We extract all equiconcepts from the concept lattice of a given graph g. According to Theorem 1, the set of equiconcepts here are exactly mapping to all maximal cliques in g.

4.1.2. Step 2-Pushing the largest maximal clique into a stack

Among the set of equiconcepts $EC(K)$, we firstly push the largest maximal cliques (*i.e.*, the cardinality of extent/intent of equiconcepts should be the largest). Suppose a stack S which is used to store the maximal cliques, this process is formally represented as

$$S.push(\arg\max_{H_i \in \mathcal{EC}(\mathcal{K})} |A_i|) \qquad where\, H_i = (A_i, B_i)$$

Suppose $k = 3$, Step 2 is clearly explained with the following Fig. 5. Among 9 obtained equiconcepts, Step 2 pushes the equiconcepts whose cardinality of extent or intent is the largest. Therefore, equiconcepts $\{(a, b, c, d, e), (a, b, c, d, e)\}$ are founded. Finally, the extent/intent of this equiconcept is pushed into the stack S.

4.1.3. Step 3-Pushing the i_{th} non-overlapped largest maximal clique into a stack

In order to guarantee the diversification of the top-k cliques, the proposed approach attempts to avoid a phenomenon on possible maximal cliques at the i_{th} step overlapping with the maximal cliques pushed into the stack S at the $(i-1)_{th}$ step. To this end, the following definition is provided.

Definition 5 (*Brother Equiconcept*)**.** Given a formal context $K = (G, M, I)$, regarding to three equiconcepts $H_1 = (A_1, B_1)$, $H_2 = (A_2, B_2)$, $H_3 = (A_3, B_3) \in EC(K)$, we define $H_1 \le H_3 \Leftrightarrow A_1 \subseteq A_3 \Leftrightarrow B_1 \supseteq B_3$, then H_3 is the father concept of H_1, denoted as $father(H_3, H_1)$ and $H_2 \le H_3 \Leftrightarrow A_2 \subseteq A_3 \Leftrightarrow B_2 \supseteq B_3$, then H_3 is also the father concept of H_2, termed $father(H_3, H_2)$. Then, H_3 is the common father concept of H_1 and H_2, and H_1 is the brother equiconcept of H_2, denoted as $brother(H_1, H_2)$ or $brother(H_2, H_1)$.

Let H_{i-1} be the $(i-1)_{th}$ non-overlapped largest maximal clique which has been pushed into stack S. Then, we need to eliminate the brother equiconcepts of H_{i-1} iteratively, such that these eliminated equiconcepts share the same father's father concept. Consequently, the remaining equiconcepts are regarded as the potential candidates for the i_{th} non-overlapped largest maximal clique. Eventually, we recall Step 2 for further pushing the i_{th} non-overlapped largest maximal clique into a stack S.

Continuing with Fig. 5 as an example, if $k = 3$, then Step 3 will help us to find the other two maximal cliques plus the largest maximal clique (so called Top-3 diversified cliques) that can maximize the coverage of cliques. First of all, Fig. 6 shows how to push the second maximal clique which can maximize the coverage of cliques. Technically, this step avoids the largest maximal clique $\{a, b, c, d, e\}$'s brother equiconcepts which share the same father's father concept (as shown in red lines) and then the possible candidates for the 2nd largest maximal clique appeared in rectangle box are preserved. Further, we recall Step 1 for pushing the 2nd largest maximal clique $\{l, m, n, p\}$ into stack S.

Similarly, Step 3 will continue to find the next maximal clique which can maximize the coverage of cliques. And, it is easy to get two maximal cliques $\{g, h, i\}$ and $\{f, g, h\}$ which can maximize the coverage. Therefore, $\{g, h, i\}$ and $\{f, g, h\}$ are pushed into the stack S eventually. Up to now, we have obtained diversified top-3 cliques: $C_1 = (\{a, b, c, d, e\}, \{l, n, m, p\}, \{g, h, i\})$ and $C_2 = (\{a, b, c, d, e\}, \{l, n, m, p\}, \{f, g, h\})$. Then, $Cov(C_1) = Cov(C_2) = 12$ (see Fig. 7).

4.2. Algorithm

Based on the above proposed approach for detecting the diversified top-k cliques from a graph, the corresponding algorithm is devised in Algorithm 1. Note that, Algorithm 1 invokes Algorithm 2 which is used to detect the equiconcepts from a given graph G.

The Algorithm 1 is working as follows. Firstly, a graph G and a parameter k are taken as the initial inputs in this algorithm. At the initialization stage, we initialize a set of maximal cliques C, a set of Equiconcepts EC, and a stack (Line 1). Then, the proposed algorithm starts to search all the equiconcepts by invoking **Equiconcepts(G)** as shown in Algorithm 2 (Line 3). According to Theorem 1, all equiconcepts will be assigned to C (Line 4); Lines 5–8 are in charge of pushing the equiconcepts into stack S. Finally, all pushed elements are pop out and inserted into the set of maximal cliques C (Line 9–11). Algorithm 2 as a sub-algorithm of Algorithm 1, the detailed working principle and procedure are similar with our previous work [6].

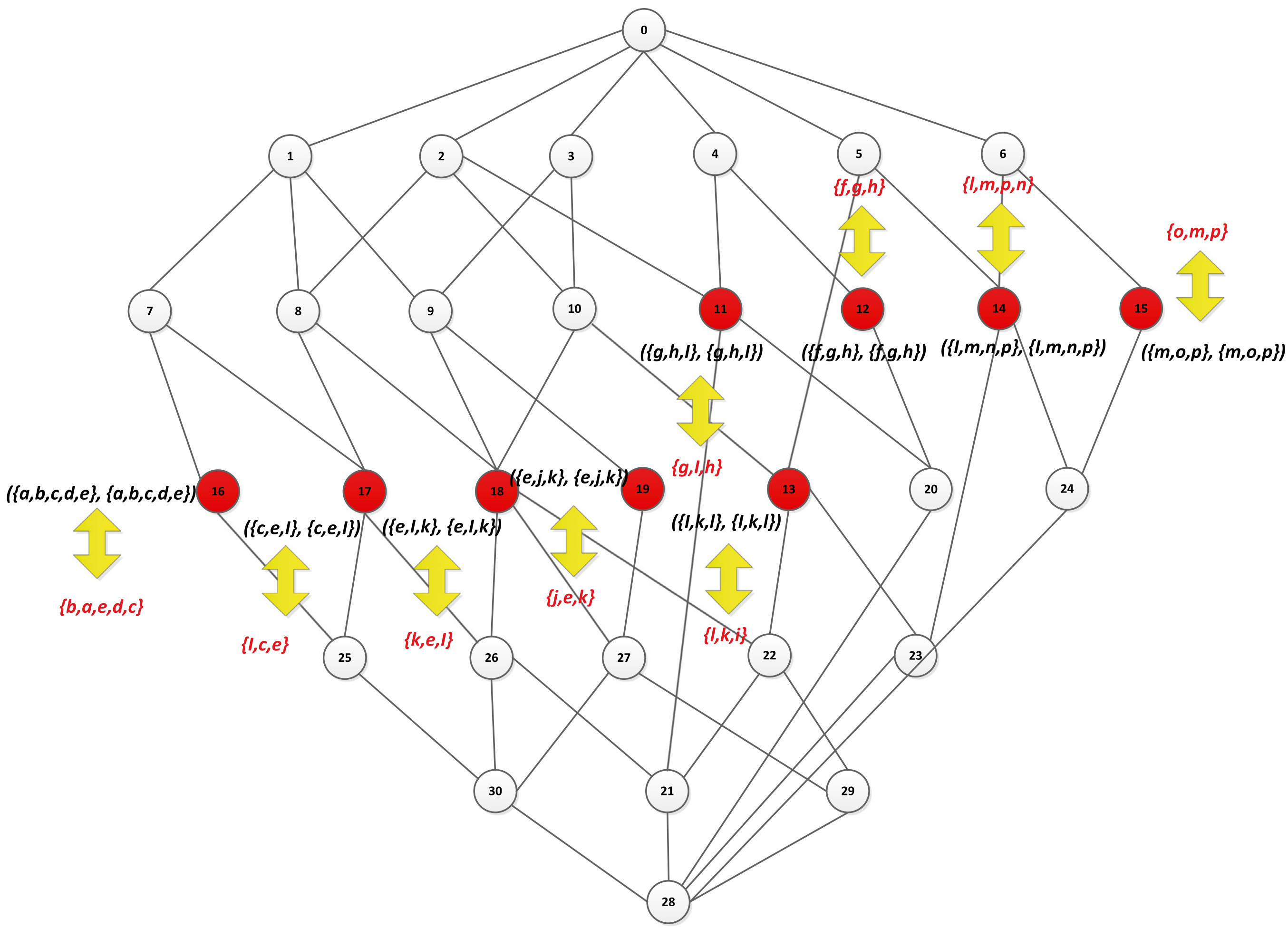


Fig. 4. Equivalence relation between equiconcepts and maximal cliques. Note that yellow two-way arrow denotes an equivalence relation.

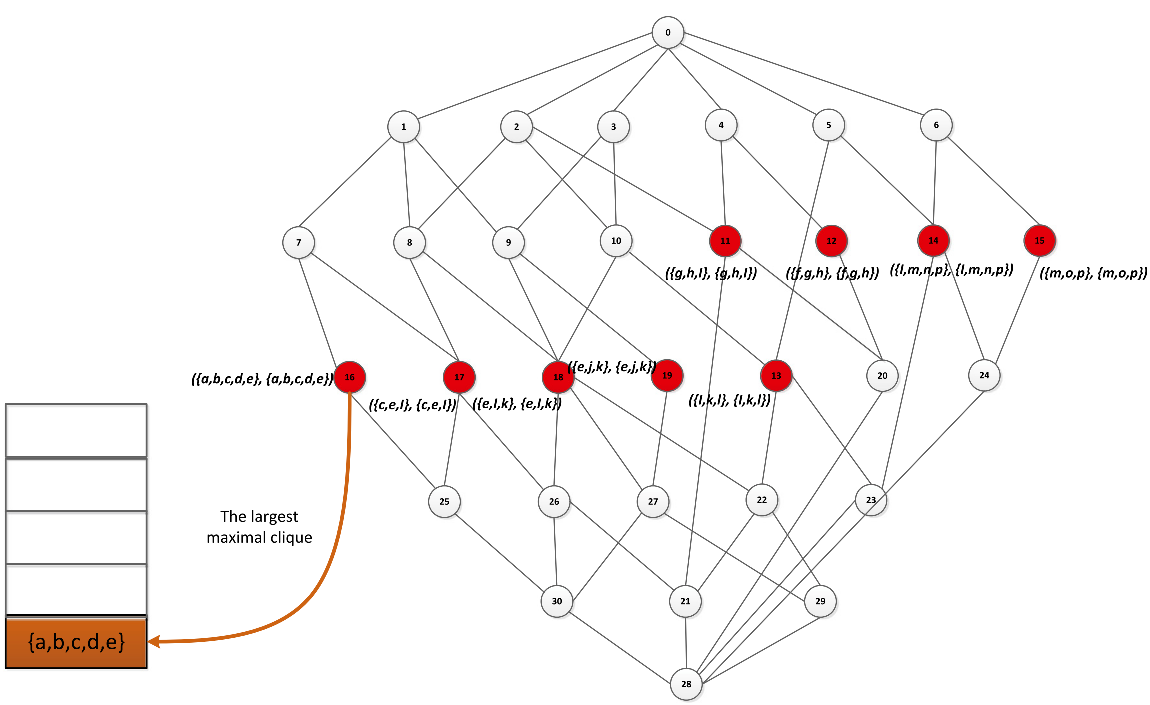


Fig. 5. Step 2: Pushing the largest maximal clique into a stack *S*.

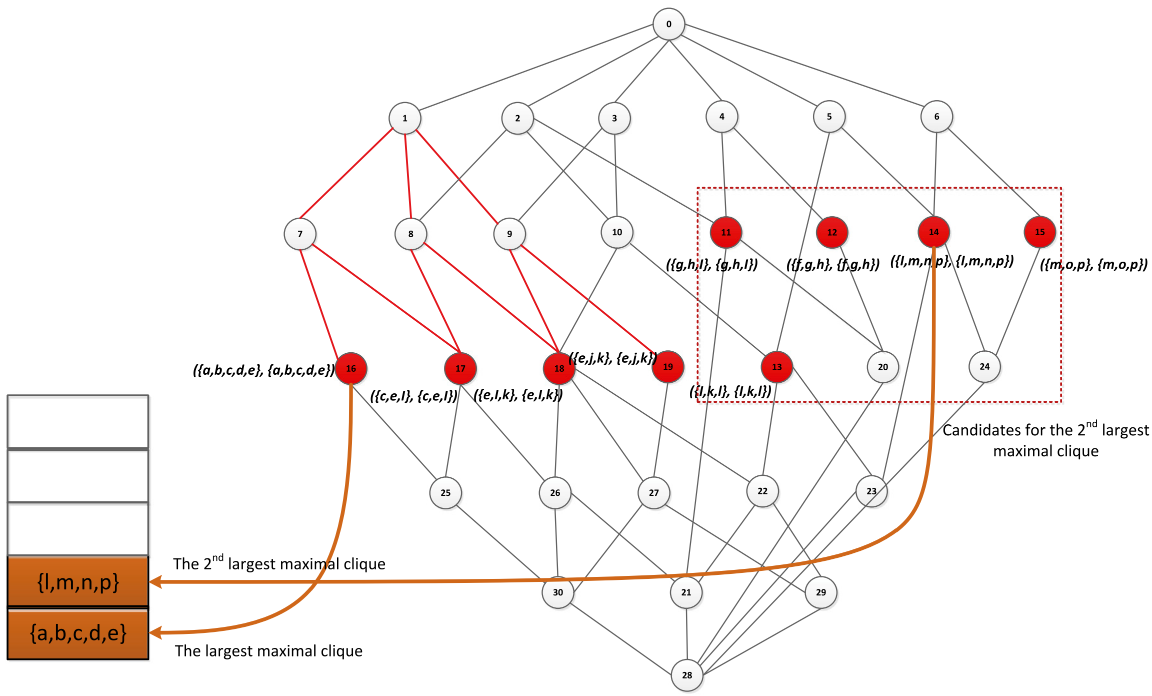


Fig. 6. Pushing the 2nd largest maximal clique into a stack *S*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

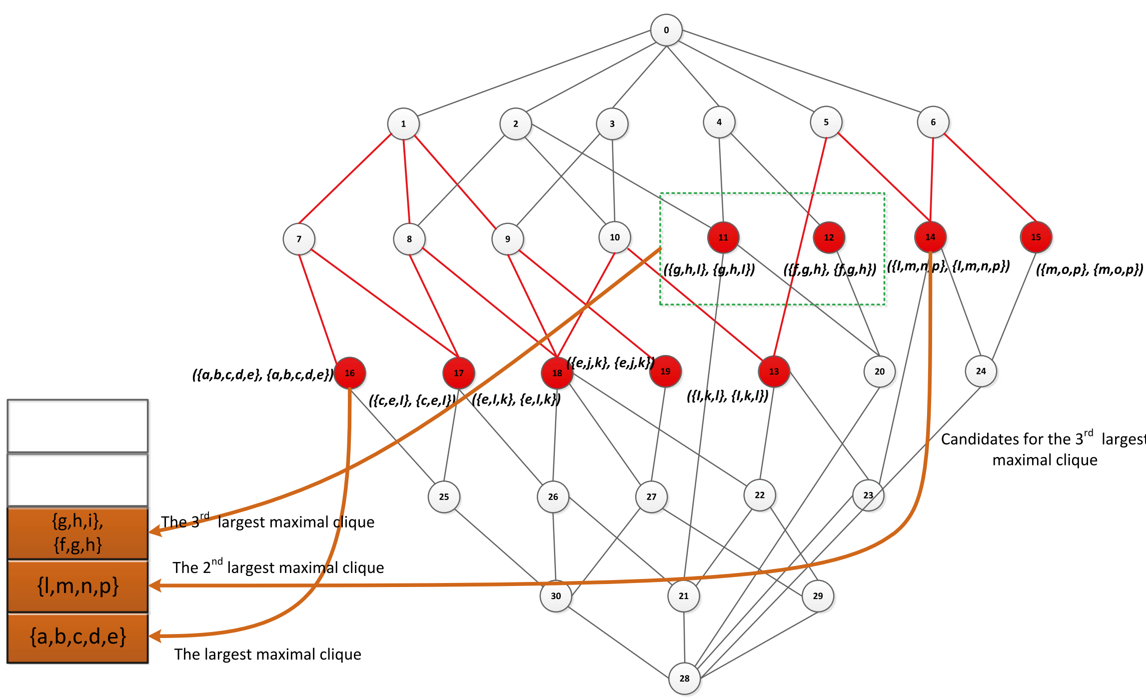


Fig. 7. Pushing the 3rd largest maximal clique into a stack *S*.

5. Experiments

This section focuses on evaluation of the proposed FCA-based diversified top-*k* maximal cliques detection approach through extensive experiments with the goal of validating its effectiveness for maximizing the coverage of cliques in graphs.

5.1. Setup

Note that all of experiments are conducted on a machine with an Intel core i5-3740 processor, 3.6 GHZ, 8 GB RAM. We utilize three datasets (as shown in Table 2) with different network sizes for evaluation. Dataset I describes a small-size social network of 34 karate club members at a US university. Dataset II is a

Algorithm 1 Diversified Top-k Maximal Clique Detection: A Wise-Greedy Algorithm

Require:
$G = (V, E)$;
Parameter k;
Ensure:
Set of Maximal Cliques C
Set of Equiconcepts EC
Stack S
Initialize $C = \emptyset$, $EC = \emptyset$ and S;
begin
$EC \leftarrow$ **Equiconcepts**(G)
$C \Leftrightarrow EC$ /*according to Theorem 1*/
S.push(arg $\max_{H_i \in \mathcal{EC}(\mathcal{K})} |A_i|$) where $H_i = (A_i, B_i)$
for $i = 2$ ***to*** k ***do***
  if $A_i \bigcap A_{i-1} = \emptyset$
    S.push(A_i)
for $i = 1$ ***to*** k ***do***
  $C \leftarrow C \cup S.pop(i)$
end

Algorithm 2 Equiconcepts(G): Equiconcepts Extraction Algorithm

Require:
$G = (V, E)$;
Ensure:
Set of Equiconcepts Γ
Initialize $\Gamma = \emptyset$
begin
Construct a formal context $FC(G)$ according to modified adjacency matrix
Build a concept lattice $C(FC(G))$
end
for each concept $(X, B) \in C(FC(G))$
begin
  if $X = B$ and $|X| = |B|$
    $\Gamma \leftarrow \Gamma \cup (X, B)$
end

Table 2
Datasets statics.

Dataset G	Type	$\|V(G)\|$	$\|E(G)\|$	Ave degree
Karate	Physical	34	78	2.29
Jazz	Musicians	198	5484	2.38
Yeast	Biology	1484	4406	2.4

social network composed of 198 Jazz musicians in a community.[1] The last dataset, a relative large network, depicts a Yeast protein interactive network [26].

Fig. 8 shows the degree distributions of three experimental data sets, respectively. In general, they all follow the power-law distribution.

5.2. Results

After the necessary configurations of experiments, we run and compare the following related algorithms with the proposed algorithm.

- **Random**: As a baseline algorithm, it randomly selects k maximal cliques from a given graph; In order to guarantee a stable coverage of clique, 10 simulation experiments are conducted and return the average value as the final coverage of clique with this algorithm.
- **Greedy**: This is a greedy algorithm for detecting the top-k diversified cliques which can cover more nodes;
- **Wise-Greedy**: We termed the proposed algorithm in this paper: Wise-greedy, a wise greedy algorithm for diversified top-k clique detection. The main difference between this algorithm and others is that we adopted a powerful soft computing methodology—formal concept analysis. We firstly adopt the equivalence relation between maximal cliques and equiconcepts of a given graph; then a stack is utilized for extracting the top-k diversified cliques from the graph.

Table 3
Running time on formal concept lattice generation for experimental datasets.

Dataset	Num. of concepts	Time
Karate	136	19 ms
Jazz	68 880	3059 ms
Yeast	2 205 857	3 454 358 ms

The performance comparison of the above algorithms are carried out with an important evaluation metric: coverage of clique as shown in Definition 4. Fig. 9(a) shows the performance evaluation for Karate dataset. Clearly, it reported that our proposed wise-greedy algorithm can achieve the maximal coverage of cliques compared to the other two algorithms. Especially, as k increases, the improvement on coverage of cliques increases significantly.

Similarly, Fig. 9(b) depicts the performance evaluation results based on Jazz dataset. As can be seen, it demonstrated that our proposed wise-greedy algorithm can maximize the coverage of cliques compared to the other two algorithms. Especially, as k increases, the improvement on coverage of cliques increases significantly.

From the efficiency of our proposed algorithm, we also compare the running time on formal concept lattice generation for experimental datasets.

From Table 3, we know that as the size of dataset increases, the running time also increases. In addition, the number of formal concepts are increasing as well.

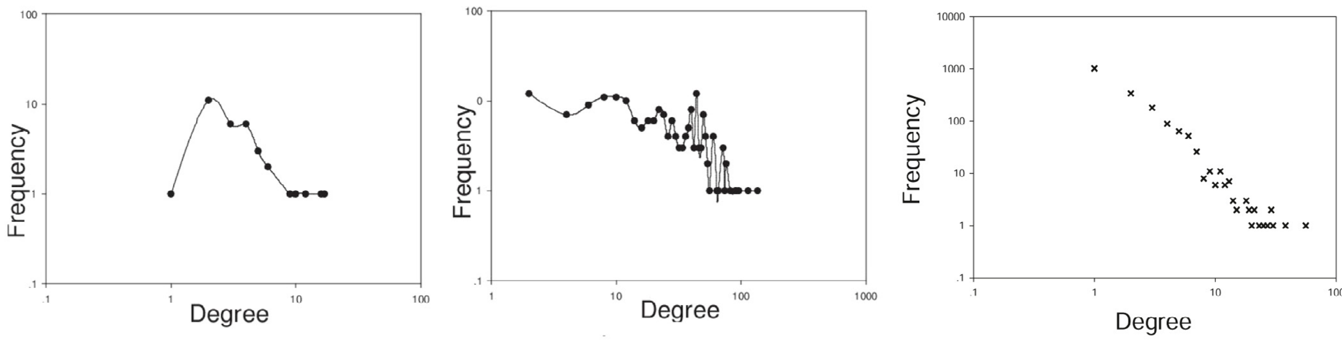
5.3. Discussions

To adapt this DKC problem with Formal Concept Analysis methodology, we firstly evaluate the running time on formal concepts generation for three datasets. Table 3 reveals an important finding: if we would like to detect diversified top-k maximal clique from a larger SIoT, it will cost much execution time for getting them. The reason is that most of formal concepts generation algorithms are costly since it is a typical NP-hard problem. However, an incremental algorithm could be an effective solution for saving the running time for formal concepts generation. Regarding to the results on the coverage of cliques from a given SIoT, our proposed algorithm Wise-Greedy outperforms other existing algorithms since an equivalence relation between maximal cliques and equiconcepts is adopted for simplifying the DKC detection procedure.

6. Conclusions

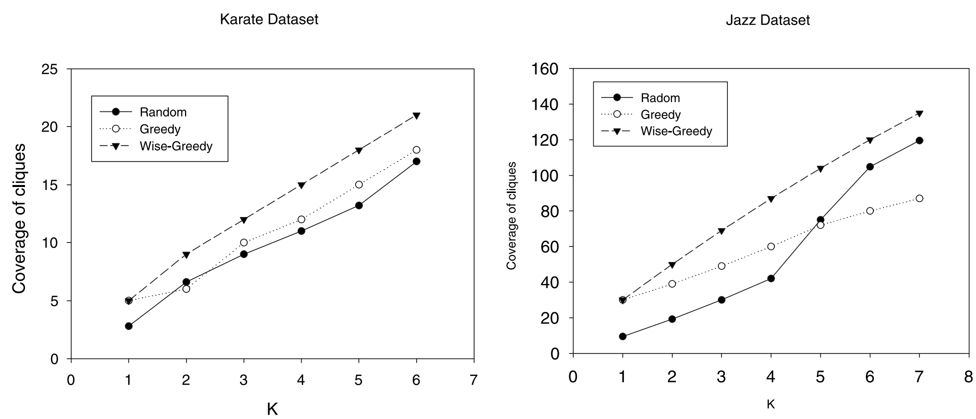
This paper has investigated the diversified top-k maximal clique detection problem in SIoT. We proved that the hardness of this problem is NP-hard. Therefore, it is necessary to devise an efficient approximate algorithm for addressing this problem. To this

[1] http://www-personal.umich.edu/~mejn/netdata/.



(a) Karate Dataset (b) Jazz Dataset (c) Yeast Dataset

Fig. 8. Degree distributions (log–log scale) of data sets.



(a) Dataset I (b) Dataset II

Fig. 9. Performance evaluation for dataset I and dataset II.

end, we have firstly provided and proved an equivalence theorem on maximal clique and equiconcept. Then, a wise-greedy algorithm is further devised for reducing the time complexity of the process on diversified top-k maximal clique detection. However, the topological structure of SIoT is usually dynamically changing as the time elapses. For example, a certain smart object might join or leave from SIoT. Hence, the dynamic characteristics of SIoT will lead to much difficulty for addressing and implementing our DKC problem. In the future, we plan to propose an incremental algorithm for extracting equiconcepts and dynamically update and adjust the maximal cliques from SIoT. Further, a corresponding diversified top-k maximal clique dynamic detection approach will be explored.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by National Natural Science Foundation of China (Grant No. 61702317) and the Fund Program for the Scientific Activities of Selected Returned Overseas Professionals in Shaanxi Province, China (Grant No. 2017024), and was also supported by the Natural Science Basic Research Plan in Shaanxi Province of China (2019JM-379).

References

[1] L. Atzori, A. Iera, G. Morabito, From “smart objects” to “social objects”: The next evolutionary step of the internet of things, IEEE Commun. Mag. 52 (1) (2014) 97–105.
[2] L. Atzori, A. Iera, G. Morabito, M. Nitti, The social internet of things (siot)-when social networks meet the internet of things: Concept, architecture and network characterization, Comput. Netw. 56 (16) (2012) 3594–3608.
[3] D. Zhang, D. Zhang, H. Xiong, L.T. Yang, V. Gauthier, NextCell: predicting location using social interplay from cell phone traces, IEEE Trans. Comput. 64 (2) (2015) 452–463.
[4] M. Dong, H. Li, K. Ota, L.T. Yang, H. Zhu, Multicloud-based evacuation services for emergency management, IEEE Cloud Comput. 1 (4) (2014) 50–59.
[5] I. Derenyi, G. Palla, T. Vicsek, Clique percolation in random networks, Phys. Rev. Lett. 94 (16) (2005) 160202.
[6] F. Hao, G. Min, Z. Pei, D.S. Park, L.T. Yang, k-clique communities detection in social networks based on formal concept analysis, IEEE Syst. J. 11 (1) (2017) 250–259.
[7] E. Gregori, L. Lenzini, S. Mainardi, Parallel (k)-clique community detection on large-scale networks, IEEE Trans. Parallel Distrib. Syst. 24 (8) (2013) 1651–1660.
[8] J. Cheng, L. Zhu, Y. Ke, et al., Fast algorithms for maximal clique enumeration with limited memory, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 1240–1248.
[9] J. Wang, J. cheng, A.W.C. Fu, Redundancy-aware maximal cliques, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 122–130.
[10] P.S. Segundo, J. Artieda, D. Strash, Efficiently enumerating all maximal cliques with bit-parallelism, Comput. Oper. Res. 92 (2018) 37–46.
[11] P. Doody, A. Shields, Mining network relationships in the internet of things, in: International Workshop on Self-aware Internet of Things, 2012.

[12] L. Yuan, L. Qin, X. Lin, L. Chang, W. Zhang, Diversified top-k clique search, in: Proceedings of 2015 IEEE 31st International Conference on Data Engineering, ICDE, 2015 pp. 387–389.
[13] P. Li, H. Dau, G. Puleo, O. Milenkovic, Motif clustering and overlapping clustering for social network analysis, in: Proceedings of IEEE International Conference on Computer Communications, Infocom, 2017, pp. 1–9.
[14] X. Wu, Z. Hu, X. Fu, L. Fu, X. Wang, S. Lu, Social network de-anonymization with overlapping communities: Analysis, algorithm and experiments, in: Proceedings of IEEE International Conference on Computer Communications, Infocom, 2018.
[15] K. Gu, L. Wang, B. Yin, Social community detection and message propagation scheme based on personal willingness in social network, Soft Comput. 23 (15) (2019) 6267–6285.
[16] X. Zheng, T. Liu, Z. Yang, J. Wang, Large cliques in Arabidopsis gene coexpression network and motif discovery, J. Plant Physiol. 168 (6) (2011).
[17] U. Feige, A threshold of in n for approximating set cover, J. ACM 45 (4) (1998).
[18] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundation, Springer-Verlag, New York, 1999.
[19] S. Andrews, In-close: a fast algorithm for computing formal concepts, in: Proceeding of International Conference on Conceptual Structures, ICCS, 2009.
[20] F. Hao, D.S. Park, G. Min, Y.S. Jeong, J.H. Park, k-cliques mining in dynamic social networks based on triadic formal concept analysis, Neurocomputing 209 (C) (2016) 57–66.
[21] S. Sun, Y. Wang, W. Liao, W. Wang, Mining maximal cliques on dynamic graphs efficiently by local strategies, in: Proceedings of IEEE 33rd International Conference on Data Engineering, ICDE, 2017, pp. 115–118.
[22] E. Balas, M.W. Padberg, On the set-covering problem, Oper. Res. 20 (6) (1972) 1152–1161.
[23] Z. Zhou, S. Das, H. Gupta, Connected k-coverage problem in sensor networks, in: Proceedings of 13th IEEE International Conference on Computer Communications and Networks, ICCCN, 2004, pp. 373–378.
[24] P.K. Singh, A.K. Cherukuri, J. Li, Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy, Int. J. Mach. Learn. Cybern. 6 (1) (2014) 1–11.
[25] Z. Zhang, Constructing L-fuzzy concept lattices without fuzzy Galois closure operation, Fuzzy Sets and Systems 333 (2018) 71–86.
[26] B. Schwikowski, P. Uetz, S. Fields, A network of protein–protein interactions in yeast, Nature Biotechnol. 18 (2000) 1257–1261.

Fei Hao received the B.Sc. degree in Information and Computing Science and the M.Sc. degree in Computer Software and Theory from Xihua University, China, in 2005 and 2008, respectively, and the Ph.D. degree in Computer Science and Engineering from Soonchunhyang University, South Korea, in 2016. He is currently an associate professor with the School of Computer Science, Shaanxi Normal University, China. He has authored over 80 papers in international journals and conferences. He received five best paper awards from KISM 2012, GreenCom 2013, MUE 2015, UCAWSN 2015, and CUTE 2016. He is a recipient of the IEEE Outstanding Leadership Award at CPSCom 2013, the 2015 Chinese Government Award for Outstanding Self-Financed Students Abroad and the IEEE Outstanding Service Awards at SmartData 2017 and DSS 2018. His research interests include social computing, ubiquitous computing, big data analysis and processing and mobile cloud computing. He is a member of ACM and KIPS.

Zheng Pei received the M.S. and Ph.D. degrees from Southwest Jiaotong University, Chengdu, China, in 1999 and 2002, respectively. He is currently a Professor with the School of Computer and Software Engineering, Xihua University, Chengdu. His research interests are rough set theory, fuzzy set theory, logical reasoning, and linguistic information processing. He has nearly 70 research papers published in academic journals or conference.

Laurence T. Yang is a fellow of the Canadian Academy of Engineering. He received the B.E. degree in Computer Science and Technology from Tsinghua University, Beijing, China, and the Ph.D. degree in Computer Science from the University of Victoria, Victoria, BC, Canada. He is a professor in the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, and in the Department of Computer Science, St. Francis Xavier University, Antigonish, NS, Canada. He is also a researcher in the Research Institute of Huazhong University of Science and Technology in Shenzhen, Shenzhen, China. His research interests include parallel and distributed computing, embedded and ubiquitous/pervasive computing, and big data. His research has been supported by the National Sciences and Engineering Research Council, Canada, and the Canada Foundation for Innovation.