This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3107529, IEEE Transactions on Network Science and Engineering

Learning Concept Interestingness for Identifying Key Structures from Social Networks

Jie Gao, Fei Hao, Zheng Pei, and Geyong Min, *Member, IEEE*

***Abstract*—Identifying key structures from social networks that aims to discover hidden patterns and extract valuable information is an essential task in the network analysis realm. These different structure detection tasks can be integrated naturally owing to the topological nature of key structures. However, identifying key network structures in most studies has been performed independently, leading to huge computational overheads. To address this challenge, this paper proposes a novel approach for handling key structures identification tasks simultaneously under the unified Formal Concept Analysis (FCA) framework. Specifically, we first implement the FCA-based representation of a social network and then generate the fine-grained knowledge representation, namely concept. Then, an efficient concept interestingness calculation algorithm suitable for social network scenarios is proposed. Next, we then leverage concept interestingness to quantify the hidden relations between concepts and network structures. Finally, an efficient algorithm for jointly key structures detection is developed based on constructed mapping relations. Extensive experiments conducted on real-world networks demonstrate that the efficiency and effectiveness of our proposed approach.**

***Index Terms*—Social Networks, Structure Identification, Formal Concept Analysis, Concept Interestingness**

I. Introduction

THE rapid development of the Internet and the popularity of smart terminals in recent years promote the proliferation of Online Social Networks (OSNs). In particular, the emerging social media represented by short video platform attracts an increasing number of users, which generates massive social media data and results in a more complicated network structure. OSNs are considered complex networks with non-trivial topological properties because the link patterns between nodes is not random or purely regular [1]. As a result, key structures identification is essential to discovering unobserved patterns and understanding structural characteristics. And it also plays an important role in various application domains, such as social recommendation [2], information diffusion analysis [3] and privacy protection [4].

The term of key structures in this paper refers to cohesive structures including maximal clique, isolated maximal clique and community, and bridging structures including bridge and structure hole spanner. Cohesive subgraph detection is one of the flourishing issue in understanding the characteristics of real-world networks. Maximal clique is one of the most fundamental cohesive subgraph models in network analysis, which requires each pair of vertices to have an edge and cannot be extended by including one more adjacent vertex [5] [6]. A maximal clique is considered isolated if it has no edges connecting it to the rest of the graph [7]. Although the definition of community structure is ambiguity, the community is generally a cohesive subgraph in which nodes are more likely to be connected due to sharing common attributes or characteristics [8]. Besides, identifying bridging structures is helpful to analyze information diffusion between communities and then better serve social networking applications. This is because bridges or bridging nodes commonly located in the borders of communities are able to enable information exchange among different communities [9]. Specifically, the bridge represents the essential structure of a social networking scenario where users join different OSNs even though they are not familiar with each other but still can interact. In sociology, bridging nodes, known as structural hole spanners in structural hole theory [10], perform similar functions in ensuring communication between different communities.

Due to the topology properties of cohesive structures and bridging structures, these key structures are generally interdependent, and their identification tasks are naturally tangled with each other. In the literature, considerable approaches have been proposed for independently discovering key structures in the complex graph, which will be elaborated in related work. However, some researchers [11]–[13] have noticed the possibility of jointly solving these detection tasks and try to tackle them simultaneously. For example, Shen *et al* [11] studied exploring intrinsic structural regularities in networks using a general stochastic block model. However, this model focuses on dividing nodes into cohesive groups while ignoring the bridging structures, and the number of cohesive group needs to be specified. To jointly detect communities and structural hole spanners, He *et al* [13] proposed a harmonic modularity method using the topological structure of the social network. From a practical point of view, this model fails to have high flexibility when the detection task needs to be changed.

In this paper, we adopt Formal Concept Analysis (FCA) [14] to analyze and tackle key structures detection simultaneously under the unified and flexible framework. FCA, a powerful theory for knowledge discovery, provides techniques to effectively discover the fine-grained knowledge (namely formal concept) from binary relations and then organize them in a lattice-based structure, namely concept lattice [14]. The

J. Gao and F. Hao are with the Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an 710062, China & School of Computer Science, Shaanxi Normal University, Xi'an 710119, China. F. Hao is also with the Department of Computer Science, University of Exeter, Exeter EX4 4QF, U.K.

Z. Pei is with the Center for Radio Administration & Technology Development, Xihua University, Chengdu 610039, China.

G. Min is with the Department of Computer Science, University of Exeter, Exeter EX4 4QF, U.K.

*Corresponding author: Fei Hao; Email:feehao@gmail.com.

general graph characterizes the relationship between nodes or entities as a binary relation, which provides the potential for FCA to identify hidden patterns in social networks. A great deal of recent research has focused on using FCA-related technologies for social network analysis [15]–[17]. In practice, when the concept generated from the formal context (binary relation) of a social network, how to interpret the implicit association between the concept and the network structure becomes crucial. As more FCA-related technologies such as concept interestingness are applied to learn this hidden relationships, this model's high flexibility makes it possible to discover key structures or other structures in social networks.

In practice, we adopt concept interestingness measures including stability and separation to select interesting concepts and quantify the correlation between concepts and key structure. More concretely, the stability and separation are introduced to construct the mapping relations between concept and key structure, and then the quantified mapping relation is used to identify multiple key structures simultaneously. The reasons for picking stability and separation are as follows: (1) the combining stability and separation is suitable for quantifying the characteristics of key structures, that is, internal entities are cohesive and separated from external entities; (2) the existing experiments have proved that combining these methods is promising to handle noisy data and improve the readability of concept lattice [18]. While the primary challenge we encountered in practice is stability calculation which has been proved NP-complete [19]. Most previous works calculate stability values using approximate calculations. Although recent work [20] can handle accurate stability calculation, it does not make full use of the relationship between concepts and is not suitable for social network scenarios. Therefore, considering the uniqueness of social networks, we advance a concept interestingness calculation algorithm CICal suitable for social network scenarios.

In this paper, we propose a novel approach to identify key structures simultaneously under the unified FCA framework. The main contributions of this paper are summarized as follows:

- **A Concept Interestingness Calculation Algorithm suitable for Social Network Scenario**: Considering the uniqueness of social network and the knowledge hidden in generated concept lattice, we propose a concept interestingness calculation algorithm (CICal) suitable for bridging the gap in how to leverage concept-cognitive learning technologies for social network analysis efficiently. The experimental results show that our algorithm can effectively handle concept interesting calculations of formal concepts generated from social networks.
- **Mapping Relations between Concepts and Key Structures of a Social Network**: In concept-cognitive learning, we introduce stability and separation to interpret the hidden association between the concept and the key structures. Specifically, we have proved the mapping relations between particular concepts and key structures and learned a mapping relation table, paving the way for the subsequent identification approach.
- **FCA-based Key Structures Detection Approach**: Based on the quantified mapping relations between concepts and key structures, we devise a key structures identification approach under a unified FCA framework to detect maximal cliques, bridges, structural holes, isolated maximal cliques and communities.
- **Evaluation**: Furthermore, we conduct extensive experiments to evaluate the proposed identification method on four real-network datasets. Experimental results show that our approach has a higher NMI value and F1-measure score than other approaches. It demonstrates that our proposed approach can accomplish multiple key structures identification tasks in a unified FCA framework.

The outline of this paper is organized as follows: Section II overviews the related work. The preliminary knowledge is provided in Section III. Section IV defines the problem and then presents a solution framework. Section V thoroughly describes the approach for identifying key structures from a social network. Section VI reports the experimental results and analysis. Finally, Section VII concludes this paper and presents future work.

II. Related Work

In this section, we conduct an in-depth review of key structures detection in social network analysis, concept cognition and concept interesting measures.

A. Key Structures Detection

There have been a lot of state-of-art studies on key structures detection in network analysis.

In the literature, a large number of algorithms have been proposed for maximal clique enumeration, although it is NP-hard theoretically. Lu *et al* [5] devised a randomized algorithm for discovering the maximum clique which employs a binary search and a novel iterative method to determine the maximum clique. Dutta *et al* [6] proposed a heuristic method to significance prune search space and accurately identify candidate regions containing a maximum clique in dense graphs. Hao *et al* [17] studied the diversified top-k maximal clique detection problem based on FCA.

Generally, community detection focuses on discovering clusters in which nodes inside are tightly connected and separate from external nodes. Traditional community detection algorithms can be divided into the following categories: agglomeration [21] [22], division [23] [24], label propagation [25] and optimization [26]. In recent years, deep learning has been demonstrated to have great power on community detection. Compared with traditional community detection approaches, deep learning-based methods aim to identify community structures by creating more powerful representations of node attributes and community structures [27]. Concretely, depending on the used learning strategies, deep learning-based methods for finite and infinite community detection fall into five main categories: convolutional neural network (CNN)-based [28], auto-encoder-based [29], generative adversarial network-based [30], graph embedding-based [31] [32] and graph neural network (GNN)-based [33] [34]. Comprehensive surveys [27] [35] of community detection approaches are referred.

For detecting bridging structures, a great deal of recent approaches [9], [36]–[38] have been proposed. For example, Buccafurri *et al* [9] have deeply studied bridges are the most basic structure of a social networking scenario and argued that most understanding of structural characteristics of networks is based on the adequate knowledge of bridges. Zhang *et al* [38] proposed a novel algorithm to identify structural hole spanner in large social netwoeks based on community forest model and diminishing marginal utility.

However, identifying key structures from social networks in most studies have been performed independently, leading to huge computational overheads. Some literature [11]–[13] studies the possibility of jointly solving these detection tasks and trying to tackle them simultaneously. For example, Shen *et al* [11] studied exploring intrinsic structural regularities in networks using a general stochastic block model. However, this model is restricted to dividing network nodes into cohesive groups such that the members of each group have similar patterns of connections to other groups, and the group number needs to be specified. To jointly detect communities and structural hole spanners, He *et al* [13] proposed a harmonic modularity method using only the topological structure of the network. But, this model fails to have high flexibility when the detection task needs to be changed in practice. In this paper, based on our previous work [15] [17], we propose an efficient approach to identify cohesive structures and bridging structures simultaneously by learning concept interestingness measure under a unified and flexible FCA framework.

B. Concept Cognition and Concept Interestingness Measures

Formal concept analysis, a powerful computational intelligence methodology, is playing an important role in concept-cognitive learning [39], [40]. A concept is a cognitive unit, generally comprising its extent and intent parts, used to identify a real-world concrete entity or model a perceived-world abstract subject [41]. Basic fundamentals of FCA are shown in Section III-A. Because FCA only supports binary decision-making that considers accepting and rejecting two options, three-way concept analysis (3WCA) [42] combing FCA with three-way decisions has recently attracted many researchers. And many recent literature studies emerging three-way concept and its application in knowledge discovery [43] and cognitive learning [41], [44]. However, the existence of noisy concepts [18] decreases the quality of concept cognition, hence a crucial task is to improve concept readability or select useful concepts. Therefore, various concept interestingness measures, such as stability, separation and robustness, have been studied to tackle this task [45]. Among these measures, stability has been verified to be the more prominent in assessing the concept quality. Additionally, the stability of three-way concept was recently been proposed [46] and its potential applicability in natural language generation was also demonstrated.

A great deal of recent studies [20], [47]–[51] has focused on stability computation which has been proved NP-complete. To overcome such an computation task, Babin and Kuznetsov [47] utilized random Monte Carlo Sampling to develop an approximate algorithm of stability calculation. Subsequently, Buzmakov *et al* [48] provided the upper and lower bounds of stable values through the existing structure of concept lattice. Recently, Ibrahim and Rokia [49] explored variance reduction techniques including leverage stratification, low-discrepancy and hybridization, and then introduced an approach for estimating stability. For calculating accurate value of stability, Jay *et al* [50] presented that accurate stability can be accumulated by already calculated results of all sub-concepts. Recently, Mouakher *et al* [20] pioneered an algorithm called DFSP that calculate stability by pruning the search space as much as possible and smartly counting generators. However, DFSP does not fully consider the hidden knowledge between concepts, such as the equivalence relation between the maximal non-generators and its lower neighbor concepts. Therefore, on the basis of optimizing this algorithm, we also consider the uniqueness of social networks, and propose a concept interestingness calculation algorithm called CICal that is more suitable for social network scenarios.

III. PRELIMINARY

In this section, we briefly review the basic notions involving formal concept analysis and key structures of a social network. The major notations used throughout this paper are listed in Table I.

TABLE I
THE MAIN NOTATIONS USED THROUGHOUT THIS PAPER.

Notation	Descriptions
$K(O,M,I)$	the formal context K with object set O, attribute set M and binary relation set I
$f(A)$	the common attributes set of all objects of A
$g(B)$	the object set with all attributes from B
(A,B)	the formal concept with the extent set A and the intent set B
L	the concept lattice
$\sigma(A,B)$	the stability of a concept (A,B)
$\xi(A,B)$	the separation of a concept (A,B)
$G(V,E)$	the graph G with vertex set V and edge set E
$\mathcal{C}$	the set of all maximal cliques in a graph
$\mathcal{IC}$	the set of all isolated maximal cliques in a graph
$\mathcal{B}$	the set of all bridges in a graph
$\mathcal{H}$	the set of all structural hole spanners in a graph
$\mathbb{C}$	the set of all communities in a graph
$M^{'}$	the modified adjacency matrix

A. Basics of Formal Concept Analysis

Formal concept analysis, a powerful mathematical theory for data analysis and visualization [14], utilizes a formal context as input to extract concepts organized in a hierarchical, lattice-based structure, namely concept lattice. A formal context is a triple $K = (O, M, I)$, where O denotes a set of objects, M represents a set of attributes, and $I \subseteq O \times M$ is a binary relation. Each pair $(x, m) \in I$ is described as follows: the object $x \in O$ contains the attribute $m \in M$. Given a subset of objects $A \subseteq O$ and a subset of attributes $B \subseteq M$,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3107529, IEEE Transactions on Network Science and Engineering

the following derivation operators are defined:

$$\begin{aligned}
f(x) &= \{m \in M | (x, m) \in I\}, \\
g(m) &= \{x \in O | (x, m) \in I\}, \\
f(A) &= \{m \in M | \forall x \in A, (x, m) \in I\} = \bigcap_{x \in A} f(x), \\
g(B) &= \{x \in O | \forall m \in B, (x, m) \in I\} = \bigcap_{m \in B} g(m).
\end{aligned}$$

where $f(A)$ is the common attributes set of all objects of A and $g(B)$ is the set of objects with all attributes from B.

A formal concept is a pair (A, B), where $A \subseteq O$, $B \subseteq M$ and $f(A) = B$, $g(B) = A$. The sets A and B are called the extent and the intent of the concept (A, B). In other words, a concept is defined as a maximal set of objects sharing a maximal set of attributes. In a concept lattice, a partial order relation exists between two concepts $(A, B) \leq (C, D)$ if $A \subseteq C(D \subseteq B)$, a pair (A, B) is a subconcept of (C, D) which is a superconcept of (A, B). If $(A, B) \leq (C, D)$ and there is no (X, Y) satisfies $(A, B) \leq (X, Y) \leq (C, D)$, (A, B) is a lower neighbor of (C, D) and (C, D) is an upper neighbor of (A, B).

Example 1 *Table II shows a formal context K with $O = \{1, 2, 3, 4\}$ and $M = \{a, b, c, d, e\}$, in which "$\times$" indicates that there is a binary relation between the object and the attribute. Since the objects 1,2 and 3 share the common attributes $\{a, b\}$ and the attributes a and b have the common objects $\{1, 2, 3\}$. Thus, $(\{1, 2, 3\}, \{a, b\})$ is a concept. $\{1, 2, 3\}$ is the extent of the concept, $\{a, b\}$ is the intent of the concept. The corresponding concept lattice L is shown in Figure 1. Each blue node represents a concept. The upper label of the node represents the intent of the concept, and the lower label represents the extent of the concept.*

TABLE II
EXAMPLE FORMAL CONTEXT K.

K	a	b	c	d	e
1	×	×			×
2	×	×			×
3	×	×	×		
4			×	×	×

In practice, the noise contained in the dataset favours the existence of many similar but distinct concepts, which may excessively impair concept readability and concept cognitive learning. Next, we will introduce two interestingness measures used in this paper, namely stability and separation, used to select the most useful and interesting concepts.

Definition 1 ***(Generator)*** *Let $K = (O, M, I)$ be a formal context. Given a formal concept (A, B) of K, if there exists a subset $P \subseteq A$ which satisfies $f(P) = B$, then P is a generator of A.*

Definition 2 *[18]* ***(Concept Stability)*** *Let $K = (O, M, I)$ be a formal context. Given a formal concept (A, B) of K, the intentional stability index σ of (A, B) is defined as follows:*

$$\sigma(A, B) = \frac{|\{P \subseteq A \,| f(P) = B\}|}{2^{|A|}} = \frac{|Gen|}{2^{|A|}} \tag{1}$$

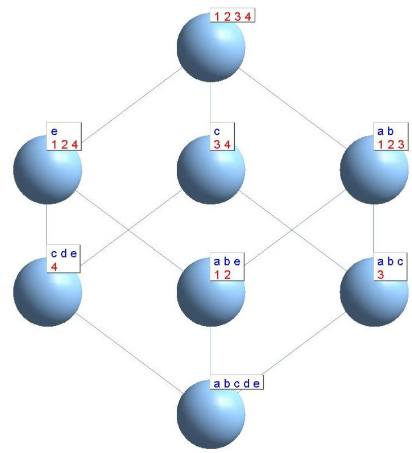

Fig. 1. The Concept Lattice L of K.

where $|Gen|$ is the total number of generators of A.

The stability describes the proportion of subsets of extent whose closure is equal to the intent [18]. In brief, given a formal concept (A, B), if some objects are removed from A, would the B remain the same? This measure quantifies the intentional degree of dependence on particular objects of the extent.

Definition 3 *[18]* ***(Separation)*** *Let $K = (O, M, I)$ be a formal context. Given a formal concept (A, B) of K, the separation index ξ of (A, B) is defined as follows:*

$$\xi(A, B) = \frac{|A||B|}{\sum_{a \in A} |f(a)| + \sum_{b \in B} |g(b)| - |A||B|} \tag{2}$$

The separation index is defined as the ratio between the area covered by a formal concept and the total area covered by its objects and attributes. It estimates the specificity of the object-attribute relation of a concept with respect to the formal context [45].

Example 2 *Continue the Example 1, the stability value of $(\{123\}, \{ab\})$ is equal to 3/8. Note that 8 stands for the cardinality of the power set of $\{123\}$ and the number of generators is 3. The separation value of $(\{123\}, \{ab\})$ is equal to 6/9. This is because the ratio between the area covered by $(\{123\}, \{ab\})$ and the total area covered by its objects and attributes is 6/9. Note that 6 is the area of the gray part of the table II, and 9 is the area of the grey part plus pink part.*

B. Key Structures of a Social Network

Definition 4 *A* ***clique*** *in an undirected graph $G = (V, E)$ is a subset of the vertices, such that every two distinct vertices are adjacent. A* ***maximal clique*** *is a clique that cannot be extended by including one more adjacent vertex. An* ***isolated maximal clique is*** *a special maximal clique in which no edge connects an object in it to any object outside of it.*

Definition 5 *A* ***community*** *is generally a cohesive subgraph in which nodes are more likely to be connected due to sharing common attributes or characteristics.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3107529, IEEE Transactions on Network Science and Engineering

Definition 6 *A **bridge** is an edge $\langle v_i, v_j \rangle$ of an undirected graph $G = (V, E)$, s.t., v_i and v_j are included in different connected components. Equivalently, an edge is a bridge if and only if it is not contained in any cycle (a cycle refers to a non-empty trail in which the only repeated vertices are the first and last vertices). A **structural hole spanner** is a bridging node between multiple connected components.*

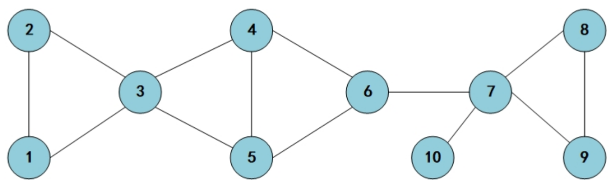

Fig. 2. A simple social network.

Example 3 *We use Figure 2 to clarify the definition of key structure more clearly. Note that the network structure is represented through a set of nodes without reference to the edge. Obviously, {1,2} is a 2-clique in Figure 2, but it is not a maximal clique. Because {1,2} is a subgraph of {1,2,3}. Since {1,2,3} does not exist exclusively within the vertex set of a larger clique, it is a maximal clique. In addition, {3,4,5} and {4,5,6} share 2 vertices, i.e., 4 and 5, thus {3,4,5,6} forms a 3-clique community. There are two bridge edges $\langle 6, 7\rangle$, $\langle 7, 10\rangle$ and three structural hole spanner 3, 6 and 7.*

IV. Problem Statement and Solution Framework

In this section, we first formulate the problem of key structures identification from a social network. Then, the solution framework of utilizing concept interestingness measures is elaborated.

Problem Statement: Given a social network $G = (V, E)$, joint key structures identification aims to extract all maximal cliques $\mathcal{C}$, isolated maximal cliques $\mathcal{IC}$, bridges $\mathcal{B}$, structural hole spanners $\mathcal{H}$ and communities $\mathbb{C}$ from G simultaneously.

Solution Framework: Figure 3 depicts a concept interestingness learning framework for identifying key structures from social networks. The framework is divided into four layers, namely, representation layer, concept layer, cognitive layer and application layer. The functions of each layer are as follows:

The representation layer is in charge of representing a social network with FCA's input, that is, the formal context. The concept layer generates a concept lattice from the formal context of a social network. In this step, network structures are represented as the concept form. The cognitive layer is responsible for learning the hidden knowledge of concepts for subsequent application. The center of the concept-cognitive circle is some intuitive knowledge, such as the extent and the intent. And the surroundings are some advanced concept-cognitive technologies, such as concept interestingness and other methods. The application layer applies the concept-cognitive results to various application fields, such as structural identification, social recommendation, etc.

Specifically, in this paper, in order to jointly identify key structures, a social network is first represented as the formal context that is the input of FCA. Then, the corresponding concept lattice is generated, and the topological structure is expressed in the concept form. Subsequently, concept interestingness, including stability and separation, are adopted to quantify the mapping relations between concepts and key structures. Finally, the required structures are identified from a social network according to the constructed mapping relations. In the next section, the proposed approach for identifying key structures from a social network by learning concept interestingness is presented in detail.

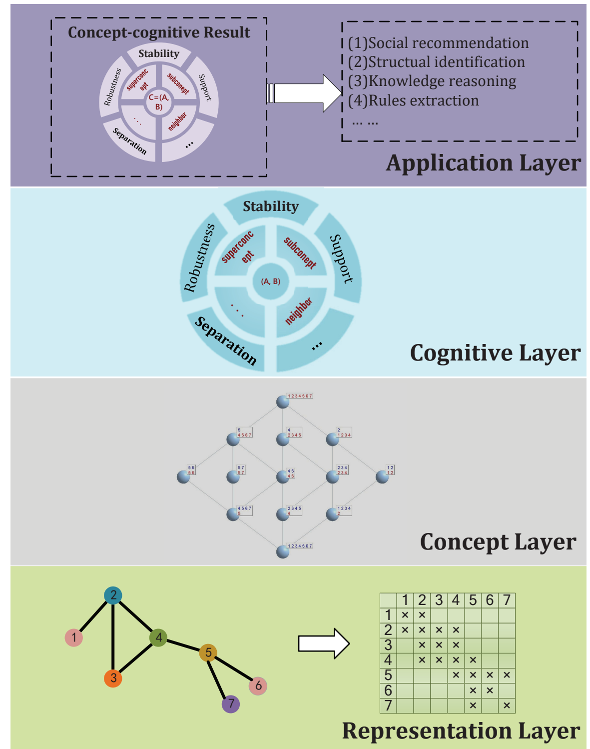

Fig. 3. Solution framework.

V. Identifying Key Structures from a Social Network

A. Constructing a Formal Context for a Social Network and Generating corresponding Concept Lattice

A social network is represented as a graph $G = (V, E)$ where the vertex set V represents individuals and the edge set E denotes the relationship between individuals. In Section III.A, we mentioned that the formal context $K = (O, M, I)$ is the input of FCA, which is very similar to the adjacency matrix of the graph representation. Each vertex v in G can be viewed as the object and attribute of the constructed formal context K. Hence, the formal context of a given social network can be formalized as $K = (V, V, I)$ by the following modified adjacency matrix, in which I is the binary relationship between vertices.

Definition 7 ***(Modified Adjacency Matrix)*** *[15] Let $G = (V, E)$ be a graph with vertices v_1 to v_n. The $n \times n$ matrix*

$M^{'}$ is called a modified adjacency matrix, in which

$$M^{'} = \begin{cases} a_{ij} = 1 & \textit{if } (v_i, v_j) \in E \\ a_{ij} = 1 & \textit{if } i = j \\ a_{ij} = 0 & \textit{otherwise} \end{cases} \tag{3}$$

Here, $K = (V, V, I)$ is equivalent to the modified adjacency matrix $M^{'}$. After the construction of formal context, we adopt our previous work [52] to generate the corresponding concept lattice, which is an incremental concept generation algorithm.

Example 4 *Figure 2 shows a social network composed of 10 users and their relationships. And the formal context of G and corresponding concept lattice is shown in Table III and Figure 4.*

TABLE III
A FORMAL CONTEXT K OF G.

G	1	2	3	4	5	6	7	8	9	10
1	×	×	×							
2	×	×	×							
3	×	×	×	×	×					
4			×	×	×	×				
5			×	×	×	×				
6				×	×	×	×			
7						×	×	×	×	×
8							×	×	×	
9							×	×	×	
10							×			×

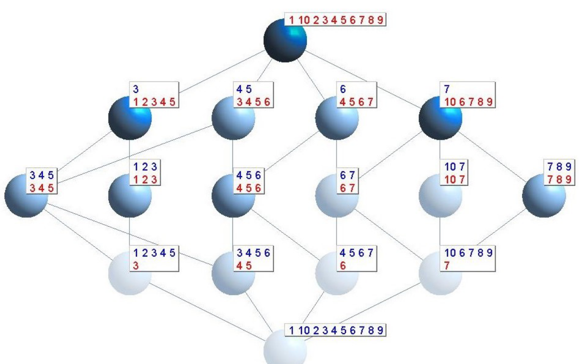

Fig. 4. The Concept Lattice of K.

B. Calculating Concept Interestingness for Social Network Scenarios

After generating all concepts, how to interpret network structures with concepts is becoming critical. In this paper, we adopt concept interestingness measures, namely stability and separation, to quantify hidden relations between concepts and key structures. This section introduces how to calculate concept interestingness efficiently, and the mapping relations between concepts and key structures are explored in the next section.

Although the separation of a concept (A, B) can be easily calculated according to Definition 3, stability computation is an NP-complete problem [19]. Most previous studies [47]–[49] focus on the approximate calculation to overcome such a computation task. However, we expect to compute accurate stability values to improve structure detection accuracy, which is verified to be right in later experimental results. It is noted that Mouakher *et al* [20] recent pioneered the exact stability calculation algorithm DFSP by pruning the search space as much as possible and fast counting generators. Specifically, DFSP mainly calculates the stability of (A, B) by pruning the maximal non-generators F and locating the minimal generators from $A - F$. The time complexity of DFSP is $O(T_{MaxNonG} + T_{MinG})$, where $T_{MaxNonG}$ and T_{MinG} represent the time complexity of maximal non-generator and minimal generator exploration, respectively. Thanks to DFSP providing the possibility to establish accurate relations between concepts and key structures, we naturally adopt DFSP and propose a Baseline algorithm (Algorithm 1) to calculate interestingness measures.

Algorithm 1: Baseline(L)

Require:
The concept lattice L
Ensure:
The stability and separation for each concept

```
begin
for each concept ∈ L do
    calculate separation according to Definition 3
    calculate stability by DFSP [20]
return the stability and separation for each concept
end
```

Although Baseline can accomplish the task of accurately calculating concept interestingness, it has two drawbacks:

- Memory consumption. Baseline has to store all concepts in memory. The number of formal concepts could be exponential to the number of entities (vertices), which makes Baseline cause memory consumption.
- Efficiency. Considering DFSP only process one concept at a time, and the number of concepts generated from social networks is generally large, Baseline is time-consuming and inefficient for social network scenario.

Revisiting Baseline, the root leading to its drawbacks discussed above is that it directly utilizes DFSP algorithms designed for the traditional formal context and does not take account of the uniqueness of social networks. Therefore, we need to develop new optimization strategies to overcome the drawbacks mentioned above.

Pre-Calculation Optimization: Based on Theorem 1, only about half of the number of generated concepts are required to calculate concept interestingness, thus the memory consumption will be reduced by around 50%. For example, in Figure 4, there is a concept $(\{3\}, \{1, 2, 3, 4, 5\})$ in the upper half-part of the concept lattice, so there must be a concept $(\{1, 2, 3, 4, 5\}, \{3\})$ in the lower half-part of the concept lattice. Therefore, only 10 out of 16 concepts are actually required to calculate concept interestingness.

Theorem 1 *Given a formal context $K = (V, V, I)$ of a social network and its corresponding concept lattice L, if a concept $(A, B) \in L$, there exists a concept $(B, A) \in L$.*

Proof *It has been proved in [53].* ∎

In-Calculation Optimization: We first present the definition of maximal non-generator. Next, in Theorem 2, we prove the equivalence between the lower neighbor concepts of a concept and its maximal non-generators. This means the maximal non-generators can be directly obtained by the lower neighbor concepts, and the time for detecting maximal non-generators can be saved. Based on Theorem 2, we advanced the procedure DFSP* which is presented in lines 7-11 of Algorithm 2 . The time complexity of DFSP* is $O(T_{MinG})$.

Definition 8 ***(Maximal Non-generator)*** *Let $K = (O, M, I)$ be a formal context. Given a formal concept (A, B) of K, if there exists a subset $P \subseteq A$ which satisfies $f(P) \neq B$, then P will be a maximal non-generator if there does not exist $P_1 \subseteq A$ and $P \subset P_1$ such that $f(P_1) \neq B$.*

Theorem 2 *Let $K = (O, M, I)$ be a formal context. Given a formal concept (A, B) of K, the maximal non-generators of (A, B) is equivalent to the lower neighbor concepts of (A, B).*

Proof *In order to prove the equivalence of lower neighbor concept and maximal non-generator, we need to prove from two directions: (1) suppose X is a maximal non-generator of A, then it can form a lower neighbor concept $(X, f(X))$ of (A, B); (2) if (X, Y) is a lower neighbor concept of (A, B), then X is a maximal non-generator of A.*

(1) Since X is a maximal non-generator of A, then $X \subset A \Rightarrow f(A) \subset f(X) \Rightarrow g(f(X)) \subset A$. In addition, $X \subseteq g(f(X))$, thus we can get $X \subseteq g(f(X)) \subset A$. Consider X is a maximal non-generator, thus $X = g(f(X))$. It means that there exists a lower neighbor concept $(X, f(X))$ of (A, B).

(2) Since (X, Y) is a lower neighbor concept of (A, B), then $B \subset Y, X \subset A, f(X) = Y$. So X is a non-generator of A. Suppose X is not maximal, there exists a maximal non-generator $\vec{X} \supset X$ of A. Based on Theorem 1, there should exist a concept $(\vec{X}, f(\vec{X}))$. This conclusion contradicts the initial conditions. Thus, X is a maximal non-generator of A. ∎

With the above optimization strategies, a concept interestingness calculation algorithm (CICal) is developed as shown in Algorithm 2. According to Theorem 1, only the upper half-part of concept lattice $\tilde{L}$ is the input of CICal. Besides, DFSP* is adopted to calculate concept stability based on Theorem 2. As the pseudocode is self-explained, we omit the description.

C. Exploring the Mapping Relationships between Concepts and Key Structures of a Social Network

In this section, the relations between concepts and key structures quantified using stability and separation are elaborated. Once network structures are represented in concept form, the combination of stability and separation is suitable for quantifying the strength of the connection between vertices in topology and the weakness of the connection between internal vertices and other remaining external vertices. For a

Algorithm 2: CICal($\tilde{L}$)

Require:
The upper half part of concept lattice $\tilde{L}$
Ensure:
The stability and separation for each concept
1: **begin**
2: **for each** concept $(A, B) \in \tilde{L}$ **do**
3: calculate separation according to Definition 3
4: calculate stability by DFSP* [20]
5: **return** the stability and separation for each concept
6: **end**
Procedure DFSP*(A, B)
7: obtain the maximal non-generators directly from the lower neighbor concepts (Theorem 2) and store into the set F
8: search the minimal generator from $A - F$
9: count generator N by minimal generator prefix
10: $\sigma = N/2^{|A|}$
11: **return** σ

concept (A, B), its stability captures the noise of a concept by estimating how its objects depend on removing each object, and its separation describes how these objects are strongly connected with other objects outside the concept [18]. So generally speaking, the higher the stability value of a concept, the more cohesive the structure corresponding to this concept; the higher the separation value, the less it connects with other objects outside.

Definition 9 ***(Equiconcept)*** *[15] A formal concept $c = (A, B)$ is called an equiconcept if $A = B$, i.e., its extent and intent are the same.*

In the following, we use $\hat{c}$ denoting an equiconcept (A, B). In addition, given a social graph G and its corresponding concept lattice $\mathbb{L}$, it is obvious that for any equiconcept $\hat{c} = (A, B) \in \mathbb{L}$, we have $|A| = |B|$.

Proposition 1 *Let $G = (V, E)$ be a social graph.*

1) *If $A \subseteq V$ is a clique, then $\forall v \in A$, $A \subseteq f(v)$ and $A \subseteq g(v)$;*
2) *If $A \subseteq V$ is a maximal clique, then (A, A) is an equiconcept of G;*
3) *If $A \subseteq V$ is an isolated maximal clique, then (A, A) is an equiconcept of G. Moreover, its stability σ and separation ξ satisfy the following constraints:*

$$\sigma(A, A) = \frac{2^{|A|} - 1}{2^{|A|}}, \quad (4)$$

$$\xi(A, A) = 1. \quad (5)$$

Proof *According to Definitions 4 and 7,*

1) A clique $A \subseteq V$ means that $\forall v \in A$ and any $v' \in A$, $(v, v') \in E$ and $(v', v) \in E$, hence $v' \in f(v)$ and $v' \in g(v)$, i.e., $A \subseteq f(v)$ and $A \subseteq g(v)$.

2) It has been proved in [15].

3) An isolated maximal clique $A \subseteq V$ means that (A, A) is an equiconcept of G. Suppose that there exists $v' \in A$ and

$f(v') \neq A$, then there exists $v_1 \in f(v')$ and $v_1 \notin A$, i.e., $(v_1, v') \in E$ and $v_1 \notin A$, this contradicts with "no edge connects an object in the isolated maximal clique $A \subseteq V$ to any object outside it". Hence for any $v \in A$, $f(v) = A$ and $g(v) = A$, i.e., $\forall v \in A$, v is a generator of A. According to Definition 2, $|Gen| = 2^{|A|} - 1$, $\sum_{v\in A} |f(v)| = |A|^2$ and $\sum_{v\in A} |g(v)| = |A|^2$, hence

$$\begin{aligned}\sigma(A, A) &= \frac{|Gen|}{2^{|A|}} = \frac{2^{|A|}-1}{2^{|A|}},\\ \xi(A, A) &= \frac{|A||A|}{\sum_{v\in A}|f(v)| + \sum_{v\in A}|g(v)| - |A||A|}\\ &= \frac{|A|^2}{|A|^2 + |A|^2 - |A|^2} = 1.\end{aligned}$$

■

Proposition 2 *Let G be a social graph and $\mathbb{L}$ be its corresponding concept lattice, an equiconcept concept $\hat{c} = (A, B) \in \mathbb{L}$ with $A = B = \{v_1, v_2\}$ represents a bridge of G iff $\hat{c}$ respectively has the following stability and separation:*

$$\sigma(A, B) = \begin{cases} 0.25 & d_1 \geq 2, d_2 \geq 2 \\ 0.5 & d_1 \geq 2, d_2 = 1 \end{cases} \quad (6)$$

where d_1, d_2 respectively denote the degree of v_1 and v_2.

$$\xi(A, B) \leq \begin{cases} 0.5 & d_1 \geq 2, d_2 \geq 2 \\ \frac{2}{3} & d_1 \geq 2, d_2 = 1 \end{cases} \quad (7)$$

Proof *We need to prove that: (1) the bridge is denoted by an equiconcept with its extent and intent contained by only two objects; (2) the stability and separation values of this equiconcept satisfy Eq.(6) and Eq.(7).*

(1) Let an edge $\{v_1, v_2\}$ be a bridge. Assuming that there exist edges $\{v_1, v_3\}$ and $\{v_2, v_3\}$, that is, v_1 and v_2 have a common attribute v_3, so there will be a cycle $v_1 \to v_2 \to v_3 \to v_1$. However, it contradicts Definition 6, an edge $\{v_1, v_2\}$ is a bridge if and only if it is not contained in any cycle. Thus, v_1 and v_2 have the common attributes only $\{v_1, v_2\}$. In other words, there is no node adjacent to nodes of v_1 and v_2 except themselves. From the definition of formal concept and Definition 7, it can be represented by an equiconcept $(\{v_1, v_2\}, \{v_1, v_2\})$.

(2) The powerset of $\hat{c}$ is $\{\emptyset, \{v_1\}, \{v_2\}, \{v_1, v_2\}\}$. When the degree of v_1 and v_2 are greater than or equal to 2, denoted by $d_1 \geq 2, d_2 \geq 2$, only one subset $\{v_1, v_2\}$ satisfies the stability condition in Eq.(1). This implies that the stability of the equiconcept $\hat{c}$ is $\frac{1}{4}$. For the calculation of separation, $|A| = |B| = 2, |f(v_1)| = |g(v_1)| = d_1 + 1, |f(v_2)| = |g(v_2)| = d_2 + 1$, thus the separation of $\hat{c}$ is $\frac{2}{d_1+d_2}$. Further, the separation of $\hat{c}$ is less than or equal to $\frac{1}{2}$. Similarly, when $d_1 \geq 2$ and $d_2 = 1$, it is easily derived that $f(v_2) = B, f(v_1, v_2) = B$. Therefore, the stability of the equiconcept $\hat{c}$ is 0.5 and the separation is equal to $\frac{2}{d_1+1}$. Further, the separation of $\hat{c}$ is less than or equal to $\frac{2}{3}$. ■

Proposition 3 *Let G be a social graph and $\mathbb{L}$ be its corresponding concept lattice, a concept $c = (A, B) \in \mathbb{L}$ with $|A| = 1$ represents a structural hole of G iff c respectively has the following stability and separation:*

$$\sigma(A, B) = 0.5 \quad (8)$$

$$\xi(A, B) \leq 0.5 \quad (9)$$

Proof *We need to prove that: (1) the structural hole is represented by a concept with its extent contained by only one object; (2) the stability and separation values of this concept satisfy Eq.(8) and Eq.(9), respectively.*

(1) Let node h be a structural hole between two non-overlapping subgraph parts P_1 and P_2 of G s.t. $h \in P_1 \wedge h \in P_2 \wedge P_1 \cap P_2 = \emptyset$. From Definition 6, two subparts are only indirectly connected through the node h. In this case, at least one node $n_1 \in P_1$, $n_2 \in P_2$ in each part is connected to h. Due to $h \in P_1 \wedge h \in P_2 \wedge P_1 \cap P_2 = \emptyset$, the common object of $\{n_1, n_2, h\}$ is h only, so there exists a concept $c = (\{h\}, \{n_1, n_2, h\})$.

(2) From Definition 2, it depicts a proportion of the subsets of A whose closure is equal to B. Since $|A| = 1$, thus $\varphi(A) = \{A, \emptyset\}$. That is, all subsets of A have only the empty set and itself. Due to $f(A) = B$ and $f(\emptyset) = \emptyset$, only one subset A satisfies the stability condition in Definition 2. So, the stability of (A, B) is equal to 0.5. For the calculation of separation, $|A| = 1, |f(A)| = |B|$, so the separation ξ is equal to $\frac{|B|}{\sum_{b\in B}|g(b)|}$, from the Definition 7, $\forall b \in B, |g(b)| \geq 2 \Rightarrow \sum_{b\in B}|g(b)| \geq 2|B|$. Thus, the separation of (A, B) is less than or equal to 0.5. ■

D. Key Structures Identification Algorithm

Based on the above propositions, we obtain the mapping relations between key structures and concepts as shown in Table IV.

TABLE IV
THE MAPPING RELATIONAL TABLE.

Proposition	Key Structure	Concept (A, B)
Proposition 1	maximal clique	$A = B$
Proposition 1	isolated maximal clique	$A = B, \sigma = \frac{2^{\lvert A\rvert}-1}{2^{\lvert A\rvert}}, \xi = 1$
Proposition 2	bridge	$A = B, \sigma = 0.25, \xi \leq 0.5$ $A = B, \sigma = 0.5, \xi \leq 2/3$
Proposition 3	structural hole	$\lvert A\rvert = 1$, $\sigma = 0.5$, $\xi \leq 0.5$

For detecting communities, an equiconcept with high stability and separation is likely to represent an independent community or the core part of a potential community. For example, the equiconcept on behalf of an isolated maximal clique has high stability (satisfy $\frac{2^{|A|}-1}{2^{|A|}}$) and high separation (equal to 1). Therefore, we first select equiconcepts that represent cliques and sort them in descending order by the product of stability and separation $\sigma \cdot \xi$. Then, we percolate the remaining cliques to get the final predicted communities.

Based on the above discussions, a key structures identification algorithm is presented in Algorithm 3. The input is the upper half part of concept lattice $\tilde{L}$. The algorithm starts by initializing the set of key structures to $\emptyset$ (line 1). The goal of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3107529, IEEE Transactions on Network Science and Engineering

Algorithm 3: Key Structures Identification Algorithm

Input: The upper half part of concept lattice $\tilde{L}$
Output: Set of maximal cliques $\mathcal{C}$, isolated maximal cliques $\mathcal{IC}$, bridges $\mathcal{B}$, structural holes $\mathcal{H}$ and communities $\mathbb{C}$

1 **Initialize** $\mathcal{C}$, $\mathcal{IC}$, $\mathcal{B}$, $\mathcal{H}$, $\mathbb{C} \leftarrow \emptyset$;
2 **foreach** *concept* $(A, B) \in \tilde{L}$ **do**
3 $\quad \sigma, \xi \leftarrow$ compute stability, separation by invoking CICal;
4 $\quad$ **if** $A = B$ **then**
5 $\quad\quad \mathcal{C} \leftarrow \mathcal{C} \cup \langle (A, B), \sigma \cdot \xi \rangle$;
6 $\quad\quad$ **if** $\sigma = 1 - 1/2^{|A|} \wedge \xi = 1$ **then**
7 $\quad\quad\quad \mathcal{IC} \leftarrow \mathcal{IC} \cup (A, B)$;
8 $\quad\quad$ **if** $|A| = 2 \wedge \sigma = 0.25 \wedge \xi \leq 0.5$ **then**
9 $\quad\quad\quad \mathcal{B} \leftarrow \mathcal{B} \cup (A, B)$;
10 $\quad\quad$ **if** $|A| = 2 \wedge \sigma = 0.5 \wedge \xi \leq \frac{2}{3}$ **then**
11 $\quad\quad\quad \mathcal{B} \leftarrow \mathcal{B} \cup (A, B)$;
12 $\quad$ **if** $|A| = 1 \wedge \sigma = 0.5 \wedge \xi \leq 0.5$ **then**
13 $\quad\quad \mathcal{H} \leftarrow \mathcal{H} \cup (A, B)$;
14 $\mathbb{C} \leftarrow Sort(\mathcal{C})$;
15 **for** $C_i = (A_i, B_i)$, $C_j = (A_j, B_j) \in \mathbb{C}$ **do**
16 $\quad$ **if** $|A_i \cap A_j| \geq min(|A_i|, |A_j|) - 1$ **then**
17 $\quad\quad C_{ij} \leftarrow (A_i \cup A_j, B_i \cup B_j)$;
18 $\quad\quad \mathbb{C} \leftarrow \mathbb{C} \setminus \{C_i, C_j\}$;
19 $\quad\quad \mathbb{C} \leftarrow \mathbb{C} \cup C_{ij}$;
20 **return** $\mathcal{C}$, $\mathcal{IC}$, $\mathcal{B}$, $\mathcal{H}$, $\mathbb{C}$

the for loop (lines 2-13) is to detect maximal cliques, isolated maximal cliques, bridges, structural holes and store them in the set. First, the stability and separation of each concept are calculated by invoking CICal (line 3). Then, lines 4-5 identify maximal cliques and store them into the set $\mathcal{C}$ (c.f. Proposition 1). Lines 6-7 detect isolated maximal cliques and store them in the set $\mathcal{IC}$ (c.f. Proposition 1). Line 8-9 identify bridges and stored them into the set $\mathcal{B}$ (c.f. Proposition 2). At last, the structural holes are detected and stored into the set $\mathcal{H}$ (lines 12-13, c.f. Proposition 3). The for loop (lines 14-19) aims to detect communities by aggregating maximal cliques. First, it sorts all maximal cliques in descending order by the value of $\sigma \cdot \xi$ and stores them into $\mathbb{C}$ (line 14). Next, it detects communities by iteratively merging every two neighboring cliques once they satisfy the aggregate constraint (lines 15-19). At last, the algorithm return the set of detected maximal cliques $\mathcal{C}$, isolated maximal cliques $\mathcal{IC}$, bridges $\mathcal{B}$, structural holes $\mathcal{H}$ and communities $\mathbb{C}$ (line 20).

VI. Experiments

In this section, we conduct comprehensive experiments to evaluate the performance of the proposed concept interestingness calculation algorithm and key structures identification algorithm. Experiment 1 evaluates the effectiveness of CICal algorithm, and Experiment 2 assesses the identification performance of our proposed key structures identification algorithm. All experiments are conducted on PC with Inter Core i7-8565U 1.80GHz 1.99GHz CPUs and 16GB RAM. All the data sets and source codes are publicly available online[1].

A. Experiment 1 - Concept Interestingness Computational Performance Evaluation

DataSets. In this experiments, we adopt three available network datasets, namely Football[2], Neural[2] and Email[3]. The critical statistics of the datasets are presented in Table V including the number of vertices, the number of edges and the number of generated concepts.

TABLE V
Statistics of Datasets.

Dataset	$\|V\|$	$\|E\|$	$\|C\|$	Description
Football	115	613	3271	Network of relations between football players
Neural	297	2148	17442	Neural network of C.Elegans
Email	1133	5451	23153	Email communication network of University Rovira i Virgili

Comparison Algorithms.

- **Jay** [50] is a stability calculation algorithm that requires browsing the entire concept lattice to calculate stability.
- **Baseline** is the baseline solution of concept interestingness calculation shown in Algorithm 1 that directly adopts DFSP [20] to calculate stability.
- **Baseline*** is an optimization algorithm of Baseline with in-calculation optimization only.
- **CICal** is our proposed Algorithm 2 with both pre-calculation optimization and in-calculation optimization.

For the sake of fairness, we only counted the time for each algorithm to calculate stability values. The difference between the last three algorithms is that Baseline directly uses DFSP [20] to calculate stability, while Baseline* uses in-calculation optimization to optimize Baseline and CICal adopts pre-calculation and in-calculation optimization.

Experimental results. Table VI reports the running time of the above four algorithms on the tested datasets.

TABLE VI
The running time of comparison algorithms on datasets.

Dataset	Jay(s)	Baseline(s)	Baseline*(s)	CICal(s)
Football	129.14	15.94	11.50	**6.38**
Neural	375.97	157.42	112.19	**62.96**
Email	549.01	190.56	130.39	**76.22**

As shown in Table VI, Jay algorithm consumes the most time among four algorithms on all datasets, since it needs to browse the entire concept lattice to calculate the stability value of each concept. It is noted that Baseline is faster than Jay algorithm on all datasets because it directly invoke the most efficient DFSP algorithm to calculate stability. Baseline*

[1]https://github.com/jiegao19/FCA4SNA
[2]http://www-personal.umich.edu/~mejn/netdata/
[3]http://konect.cc/networks/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3107529, IEEE Transactions on Network Science and Engineering

is further faster than Baseline on all datasets since it adopts an advanced DFSP* which leverages potential knowledge in concept lattice to speed up stability calculation. This reveals the effectiveness of our proposed in-calculation optimization strategy. Obviously, our CICal algorithm is about 7 times faster than Jay, 2.5 times faster than Baseline on Email datasets and is the most efficient on all datasets. Since all optimization strategies (pre-calculation optimization and in-calculation optimization) are adopted in our CICal algorithm, the search space has been significantly pruned, and the potential knowledge in the concept lattice has been fully considered. The experimental results demonstrate the effectiveness of pre-calculation optimization strategy and in-calculation optimization strategy.

B. Experiment 2 - Key Structures Identification Performance Evaluation

DataSets. We evaluate the performance of our algorithm on four real-network datasets in this experiments. Terrorist, Residence, Elegans and Email are publicly available at the following URL[3]. Engineering is a large benchmark dataset recently released in [34] for overlapping community detection. Table VII briefly summarizes the statistics of these datasets.

TABLE VII
DATASET STATISTICS. K STANDS FOR 1000.

Dataset	$\|V\|$	$\|E\|$	Description
Terrorist	64	243	Relation Network between suspected terrorists
Residence	217	2672	Friendship network between students at a residence hall
Elegans	453	4596	Metabolic network of the roundworm Caenorhabditis elegans
Email	1.1K	5.4K	Email communication network of University Rovira i Virgili
Engineering	14.9K	49.3K	Co-authorship network in Engineering constructed from Microsoft Academic

Comparison Algorithms.

- **CPM** [21] detects the communities by percolating k-cliques if they share k-1 object nodes.
- **GN** [24] is a decomposition algorithm that gradually removes the edges between communities and obtains a relatively cohesive community structure .
- **COIN** [16] adopts FCA's stability to identify and remove noisy bridges, and then detect communities by percolating the remaining cliques.
- **NOCD** [34] is a graph neural network model for overlapping community detection.

To ensure a fair comparison, we re-run all algorithms 20 times and report the average result. Besides, we used the authors released code for NOCD[4], and adopted the adjacency matrix as input of NOCD and their suggested setting for our experiments.

Evaluation measures. In this experiment, we use Normalized Mutual Information (NMI) [54] and F1-measures to evaluate the quality of the detected communities and bridge structures, respectively.

NMI is widely used to measure the accuracy of community detection by estimating the similarity between the ground-truth communities and the predicted communities [55]. Given the ground-truth communities $\mathbb{C}^*$ and the predicted communities $\mathbb{C}$, NMI is defined as:

$$NMI = \frac{2 * MI(\mathbb{C}, \mathbb{C}^*)}{H(\mathbb{C}) + H(\mathbb{C}^*)}$$

where MI and H represent the mutual information and entropy, respectively [54]. The larger the NMI value, the more similar the detected communities are to the ground-truth communities. The NMI equals 1 if and only if they are identical, whereas it has an expected value of 0 if they are independent.

F1-measure is defined as follows:

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

in which $precision$ is the ratio of the number of correctly detected to the total number of detection results, and $recall$ is the ratio of the number of correctly detected to the total number of ground-truth structures. Clearly, the larger the F1 value, the more accurate the predicted bridges.

Experimental Results. Table VIII shows the NMI values of the community detection algorithms on the experimental datasets.

TABLE VIII
COMPARISON OF NMI VALUES (IN %) OF THE COMMUNITY DETECTION ALGORITHMS ON THE TESTED SOCIAL NETWORKS.

Dataset	CPM	GN	COIN	NOCD	Our Approach
Terrorist	28.2	34.7	40.2	48.2	**52.5**
Residence Hall	28.8	29.6	34.3	37.4	**40.9**
Elegans	22.6	23.9	31.0	34.8	**37.2**
Email	17.5	13.8	29.4	35.3	**36.8**
Engineering	7.8	9.4	14.5	**17.9**	17.7

Obviously, our approach has achieved the highest score in the first four datasets, and its score on the Engineering dataset is slightly lower than NOCD. That is because the FCA-based approach detects communities based on the concepts that are the fine-grained knowledge representation of structure topology. It is worth noting that, as shown in Figure 5, our algorithm has a higher NMI value compared with another FCA-based COIN algorithm, especially with about +12% on Terrorist. This is because we leverage CICal to calculate accurate concept interestingness for improving structure detection accuracy, verifying that it is right to devise an accurate calculation algorithm tailored for social network scenarios.

For bridge structures detection, Figure 6 presents the performance of comparison algorithms on the tested datasets. Obviously, our approach outperforms the comparison algorithm in term of F1-measure. This is because of two reasons: (1) we invoke our proposed CICal to calculate the stability accurately. However, COIN adopts an approximate calculation algorithm, which will inevitably cause calculation errors and further reduce the accuracy of structure identification; (2)

[4]https://github.com/shchur/overlapping-community-detection

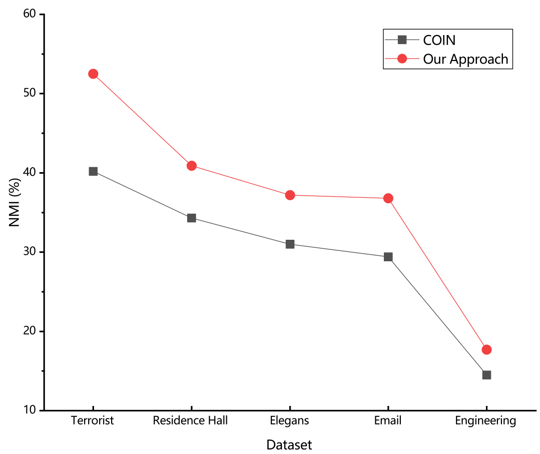

Fig. 5. NMI value of FCA-based community detection algorithms on the experimental datasets.

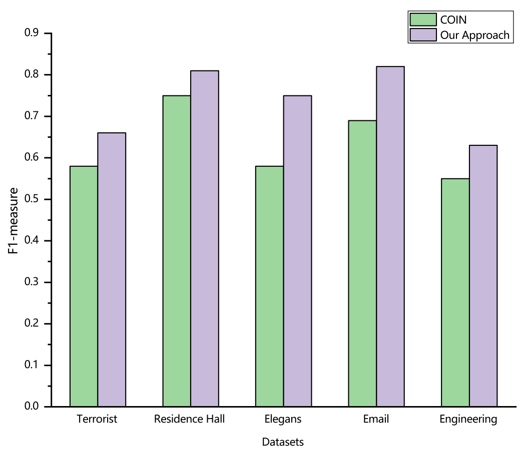

Fig. 6. F1-measure value of bridge detection algorithms on the tested datasets.

besides, we combine separation and stability to quantify the mapping relations between concept and bridge. This can weaken each measure's drawbacks, perfect the mapping relations and further improve the accuracy of structure detection.

C. Case Study

In this section, a case study conducted on a real network is introduced to illustrate the potential applications of the identified key structures.

Terrorist[3] is an undirected network that describes the organizational relationships between the 64 suspected terrorists involved in the Madrid train bombing on March 11, 2004. The dataset consists of 64 nodes and 243 edges, in which each node represents a terrorist and an edge represents a connection between two terrorists.

With our proposed algorithm, we detected 49 maximal cliques, 21 bridges, 23 structural hole spanners and 4 communities from this terrorist network. The visualization of detected communities of the terrorist network by our approach is shown in Figure 7. For example, <24, 49>, <45, 46>, <16, 33> are the bridges and 8, 17, 30, 34, 49 are the structural hole spanners bridging different communities. These bridging nodes, structural hole spanners in sociology, can better control the information spread among communities [13]. It implies that these individuals are most likely to be the senior leaders of the terrorist network. If the connections between these points can be cut off in time, the entire terrorist network can be split into several relatively independent small networks. This has guiding significance for combating terrorism and protecting public safety.

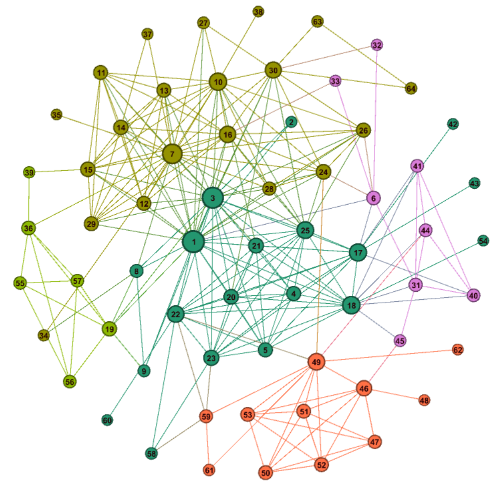

Fig. 7. The Detected Communities of Terrorists Network by Our Approach.

VII. Conclusions

In this paper, we propose an efficient approach to identify multiple key structures simultaneously under the unified FCA framework. We first implement the FCA-based representation of a social network and then generate concepts that characterize network structures. Then, an efficient concept interestingness calculation algorithm suitable for the social network scenarios is developed. Next, we adopt interestingness measures including stability and separation to discover hidden relations between particular concepts and key structures and further propose an algorithm for key structures identification from social networks. Experiments conducted on real social networks demonstrated that the efficiency and effectiveness of our proposed approach. The work for further exploration is as follows: (1) design an approach that simultaneously solves the problem of generating a concept lattice and calculating concept stability value; (2) conduct more practical applications of key structures identification, such as influence maximization, social recommendation and so forth; (3) explore the relationship between the stability of three-way concept and key structures in complex networks, such as balanced structure mining in signed network.

ACKNOWLEDGMENT

This work was funded in part by the National Natural Science Foundation of China (Grant No. 61702317), the Fundamental Research Funds for the Central Universities, China (GK202103080), and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 840922. This work reflects only the authors' view and the EU Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] D. Camacho, Á. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Information Fusion*, vol. 63, pp. 88–120, 2020.

[2] J. Ma, J. Wen, M. Zhong, L. Liu, C. Li, W. Chen, Y. Yang, H. Tu, and X. Li, "Dbrec: dual-bridging recommendation via discovering latent groups," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1513–1522.

[3] Y. Zhang, T. Lyu, and Y. Zhang, "Cosine: Community-preserving social network embedding from information diffusion cascades," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[4] J. Du, C. Jiang, K.-C. Chen, Y. Ren, and H. V. Poor, "Community-structured evolutionary game for privacy protection in social networks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 3, pp. 574–589, 2017.

[5] C. Lu, J. X. Yu, H. Wei, and Y. Zhang, "Finding the maximum clique in massive graphs," *Proceedings of the VLDB Endowment*, vol. 10, no. 11, pp. 1538–1549, 2017.

[6] S. Dutta and J. Lauri, "Finding a maximum clique in dense graphs via $\chi 2$ statistics," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 2421–2424.

[7] H. Molter, R. Niedermeier, and M. Renken, "Enumerating isolated cliques in temporal networks," in *International Conference on Complex Networks and Their Applications*. Springer, 2019, pp. 519–531.

[8] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky, "Community interaction and conflict on the web," in *Proceedings of the 2018 world wide web conference*, 2018, pp. 933–943.

[9] E. Corradini, A. Nocera, D. Ursino, and L. Virgili, "Defining and detecting k-bridges in a social network: The yelp case, and more," *Knowledge-Based Systems*, vol. 195, p. 105721, 2020.

[10] R. S. Burt, *Structural holes: The social structure of competition*. Harvard university press, 2009.

[11] H.-W. Shen, X.-Q. Cheng, and J.-F. Guo, "Exploring the structural regularities in networks," *Physical Review E*, vol. 84, no. 5, p. 056111, 2011.

[12] B. Yang, J. Liu, and D. Liu, "Characterizing and extracting multiplex patterns in complex networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 469–481, 2011.

[13] L. He, C.-T. Lu, J. Ma, J. Cao, L. Shen, and P. S. Yu, "Joint community and structural hole spanner detection via harmonic modularity," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 875–884.

[14] B. Ganter and R. Wille, *Formal concept analysis: mathematical foundations*. Springer Science & Business Media, 2012.

[15] 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.

[16] M.-H. Ibrahim, R. Missaoui, and A. Messaoudi, "Detecting communities in social networks using concept interestingness," in *Proceedings of the 28th Annual International Conference on Computer Science and Software Engineering*. IBM Corp., 2018, pp. 81–90.

[17] F. Hao, Z. Pei, and L. T. Yang, "Diversified top-k maximal clique detection in social internet of things," *Future Generation Computer Systems*, vol. 107, pp. 408–417, 2020.

[18] M. Klimushkin, S. Obiedkov, and C. Roth, "Approaches to the selection of relevant concepts in the case of noisy data," in *International Conference on Formal Concept Analysis*. Springer, 2010, pp. 255–266.

[19] S. O. Kuznetsov, "On stability of a formal concept," *Annals of Mathematics and Artificial Intelligence*, vol. 49, no. 1, pp. 101–115, 2007.

[20] A. Mouakher and S. B. Yahia, "On the efficient stability computation for the selection of interesting formal concepts," *Information Sciences*, vol. 472, pp. 15–34, 2019.

[21] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[22] S. Jabbour, N. Mhadhbi, B. Radaoui, and L. Sais, "Detecting highly overlapping community structure by model-based maximal clique expansion," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1031–1036.

[23] L. Fu, F. Li, and D. Li, "Community division algorithm based on node density and similarity," in *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE, 2020, pp. 739–743.

[24] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[25] G. Yang, W. Zheng, C. Che, and W. Wang, "Graph-based label propagation algorithm for community detection," *International Journal of Machine Learning and Cybernetics*, pp. 1–11, 2019.

[26] H.-J. Li, Z. Bu, A. Li, Z. Liu, and Y. Shi, "Fast and accurate mining the community structure: integrating center locating and membership optimization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2349–2362, 2016.

[27] F. Liu, S. Xue, J. Wu, C. Zhou, W. Hu, C. Paris, S. Nepal, J. Yang, and P. S. Yu, "Deep learning for community detection: progress, challenges and opportunities," *arXiv preprint arXiv:2005.08225*, 2020.

[28] G. Sperlí, "A deep learning based community detection approach," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2019, pp. 1107–1110.

[29] J. Cao, D. Jin, L. Yang, and J. Dang, "Incorporating network structure with node contents for community detection on large networks using deep learning," *Neurocomputing*, vol. 297, pp. 71–81, 2018.

[30] Y. Jia, Q. Zhang, W. Zhang, and X. Wang, "Communitygan: Community detection with generative adversarial nets," in *The World Wide Web Conference*, 2019, pp. 784–794.

[31] C. Tu, X. Zeng, H. Wang, Z. Zhang, Z. Liu, M. Sun, B. Zhang, and L. Lin, "A unified framework for community detection and network representation learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 6, pp. 1051–1065, 2018.

[32] S. Cavallari, E. Cambria, H. Cai, K. C.-C. Chang, and V. W. Zheng, "Embedding both finite and infinite communities on graphs," *IEEE Computational Intelligence Magazine*, vol. 14, no. 3, pp. 39–50, 2019.

[33] D. Jin, Z. Liu, W. Li, D. He, and W. Zhang, "Graph convolutional networks meet markov random fields: Semi-supervised community detection in attribute networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 152–159.

[34] O. Shchur and S. Günnemann, "Overlapping community detection with graph neural networks," *arXiv preprint arXiv:1909.12201*, 2019.

[35] D. Jin, Z. Yu, P. Jiao, S. Pan, P. S. Yu, and W. Zhang, "A survey of community detection approaches: From statistical modeling to deep learning," *arXiv preprint arXiv:2101.01669*, 2021.

[36] P. J. Jones, R. Ma, and R. J. McNally, "Bridge centrality: A network approach to understanding comorbidity," *Multivariate Behavioral Research*, pp. 1–15, 2019.

[37] W. Xu, M. Rezvani, W. Liang, J. X. Yu, and C. Liu, "Efficient algorithms for the identification of top-k structural hole spanners in large social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 5, pp. 1017–1030, 2017.

[38] Y. Zhang, H. Xu, Y. Xu, J. Deng, J. Gu, R. Ma, J. Lai, J. Hu, X. Yu, L. Hou *et al.*, "Finding structural hole spanners based on community forest model and diminishing marginal utility in large scale social networks," *Knowledge-Based Systems*, vol. 199, p. 105916, 2020.

[39] Y. Mi, Y. Shi, and J. Li, "A generalized concept-cognitive learning: A machine learning viewpoint," *arXiv preprint arXiv:1801.02334*, 2018.

[40] R. Shivhare, A. K. Cherukuri, and J. Li, "Establishment of cognitive relations based on cognitive informatics," *Cognitive Computation*, vol. 9, no. 5, pp. 721–729, 2017.

[41] J. Li, C. Huang, J. Qi, Y. Qian, and W. Liu, "Three-way cognitive concept learning via multi-granularity," *Information sciences*, vol. 378, pp. 244–263, 2017.

[42] J. Qi, L. Wei, and Y. Yao, "Three-way formal concept analysis," in *International conference on rough sets and knowledge technology*. Springer, 2014, pp. 732–741.

[43] F. Hao, Y. Yang, G. Min, and V. Loia, "Incremental construction of three-way concept lattice for knowledge discovery in social networks," *Information Sciences*, 2021.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3107529, IEEE Transactions on Network Science and Engineering

[44] R. Shivhare and A. K. Cherukuri, "Three-way conceptual approach for cognitive memory functionalities," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 21–34, 2017.

[45] S. O. Kuznetsov and T. Makhalova, "On interestingness measures of formal concepts," *Information Sciences*, vol. 442, pp. 202–219, 2018.

[46] F. Hao, J. Gao, C. Bisogni, G. Min, V. Loia, and C. De Maio, "Stability of three-way concepts and its application to natural language generation," *Pattern Recognition Letters*, 2021.

[47] M. A. Babin and S. O. Kuznetsov, "Approximating concept stability," in *International Conference on Formal Concept Analysis*. Springer, 2012, pp. 7–15.

[48] A. Buzmakov, S. O. Kuznetsov, and A. Napoli, "Scalable estimates of concept stability," in *International Conference on Formal Concept Analysis*. Springer, 2014, pp. 157–172.

[49] M.-H. Ibrahim and R. Missaoui, "Approximating concept stability using variance reduction techniques," *Discrete Applied Mathematics*, vol. 273, pp. 117–135, 2020.

[50] N. Jay, F. Kohler, and A. Napoli, "Analysis of social communities with iceberg and stability-based concept lattices," in *International conference on formal concept analysis*. Springer, 2008, pp. 258–272.

[51] J. Gao, F. Hao, and D.-S. Park, "On the computation of concept stability based on maximal non-generator for social networking services," *Applied Sciences*, vol. 10, no. 23, p. 8618, 2020.

[52] 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 *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)*. IEEE, 2019, pp. 2811–2816.

[53] F. Hao, Y. Yang, B. Pang, N. Y. Yen, and D.-S. Park, "A fast algorithm on generating concept lattice for symmetry formal context constructed from social networks," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–8, 2019.

[54] A. F. McDaid, D. Greene, and N. Hurley, "Normalized mutual information to evaluate overlapping community finding algorithms," *arXiv preprint arXiv:1110.2515*, 2011.

[55] E. Alinezhad, B. Teimourpour, M. M. Sepehri, and M. Kargari, "Community detection in attributed networks considering both structural and attribute similarities: two mathematical programming approaches," *Neural Computing and Applications*, vol. 32, no. 8, pp. 3203–3220, 2020.

Jie Gao received the B.Sc. degree in Computer Science and Technology from Qingdao University of Science and Technology, China, in 2018. He is currently working toward the M.Sc. degree at the School of Computer Science, Shaanxi Normal University, China. His research interests include social network analysis and formal concept analysis.

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. Since 2016, he has been with Shaanxi Normal University, Xi'an, China, where he is an Associate Professor. He is currently taking a Marie Sklodowska-Curie Individual Fellowship at the University of Exeter, Exeter, United Kingdom. His research interests include social computing, ubiquitous computing, big data analysis and processing and mobile cloud computing.

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.

Geyong Min received the BSc degree in computer science from the Huazhong University of Science and Technology, China, in 1995, and the PhD degree in computing science from the University of Glasgow, United Kingdom, in 2003. He is a Professor of High Performance Computing and Networking in the Department of Computer Science within the College of Engineering, Mathematics and Physical Sciences at the University of Exeter, United Kingdom. His research interests include next generation internet, wireless communications, multimedia systems, information security, ubiquitous computing, modelling, and performance engineering.