Skyline (λ, k)-cliques Identification from Fuzzy Attributed Social Networks

Fei Hao*, Jie Gao, Jianrui Chen, Aziz Nasridinov, Geyong Min

***Abstract*—Identifying the optimal groups of users that are closely connected and satisfy some ranking criteria from an attributed social network, attracts significant attention from both academia and industry. Skyline query processing, a multi-criteria decision-making optimized technique, is recently embedded into cohesive subgraphs mining in graphs/social networks. However, the existing studies cannot capture the fuzzy property of connections between users in social networks. To fill this gap, in this paper we formulate a novel model of the skyline (λ, k)-cliques over a fuzzy attributed social network and develop a Formal Concept Analysis (FCA) based skyline (λ, k)-cliques identification algorithm. Specifically, λ can be regarded as a quality control parameter for measuring the stability of the cohesive groups. Extensive experimental results conducted on three real-world datasets demonstrate the effectiveness of the skyline (λ, k)-clique model in a fuzzy attributed social network. Further, an illustrative example is executed for revealing the usefulness of our model. It is expected that our proposed skyline (λ, k)-clique model can be widely used in various graph-based computational social systems, such as optimal team formation in crowdsourcing, and group recommendation in social networks.**

***Index Terms*—Fuzzy Attributed Social Network, Clique, Skyline, Formal Concept Analysis**

I. Introduction

Motivation. As a fundamental research issue in social network analysis, cohesive subgraph computation, which identifies a group of highly connected vertices, has been applied in various fields, such as social recommendation, network routing, and knowledge graph [1]. A cohesive subgraph identification is critical to graph structure analysis and three types of cohesiveness measurement have been studied in recent years. The existing methods are mainly categorized as follows: (1) the social cohesiveness is measured by k-clique [2], k-core [3], [4], k-truss [5], [6], k-plex [7]; (2) the spatial cohesiveness [8] is measured by spatial distance between individuals in the networks; (3) both social cohesiveness and spatial cohesiveness are jointly considered [9]. These measures are successfully implemented in social networks or geo-social networks that ignore the attributes of nodes as well as the vague relations among them in the real world. Recent research efforts on skyline k-cliques identification [10] and skyline cohesive group detection [9] are achieved by taking both structural property and attributes of nodes into account.

However, the sharp increase on the scale of social networks, with diversified social interactions, is bringing significant challenges for key structures identification and its applications in social networks. Particularly, with the advancement of society, the relationship between users in social networks exhibits a lot of indistinct and vague features. For instance, the friendship between children, and the technological innovation cooperation relationship between enterprises are all fuzzy relations. The traditional binary relationships seem to be unable to describe the fuzzy characteristics among users. Therefore, these vague and uncertain relations among users in social networks greatly lead to many challenges on skyline group detection and narrow the range of applications, such as optimal team formation, product recommendation, and so forth.

Consider the following two hypothetical scenarios as our motivating examples.

- **Scenario 1:** An IT company will develop a project, they are going to build a team of engineers skilled in the following areas: R=$\{IR, AI, DM, CV\}$ (*IR: Information Retrieval, AI: Artificial Intelligence, D-M: Data Mining, CV: Computer Vision*). Let us suppose that the IT company has five candidates $\{Jack, Susan, John, Thomas, Jessie\}$ from a social network with a different set of skills and its corresponding proficiency as well as the execution cost (2-dimensional space) (as shown in Figure 1, the weights on the nodes and edges indicate the proficiency for the skills and the execution cost between users). For instance, *John* has two skills *AI* and *DM* with proficiency 0.94 and 0.92, while *Thomas* has only one skill *DM* with proficiency 0.86. Besides, the execution cost between them is 3. The social collaborative relationships among these candidates can be extracted from the collaboration networks such as DBLP and Google Scholar. How to make an optimal team for maximizing the proficiency as well as minimizing the cost is a critical issue. This problem is called team formation in social networks, and the process of selecting an optimal team should consider the vagues relationships among candidates as well as the attributes (*e.g.*, skill proficiency and cost) of themselves. For instance, the engineers $John, Thomas, Jessie$ can form an optimal team (as shown in the triangle area of Figure 1), *i.e.*, skyline (λ, k)-clique for executing the given project efficiently.
- **Scenario 2:** Let us assume the travel agency intend to recommend a candidate list of k hotels to a large

F. Hao and G. Min are with the Department of Computer Science, University of Exeter, Exeter EX4 4QF, U.K.

J. Gao and J. Chen 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.

A. Nasridinov is with the Department of Computer Science, Chungbuk National University, Cheongju 28644, Korea

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

Manuscript received****, ****; revised ****, ****.

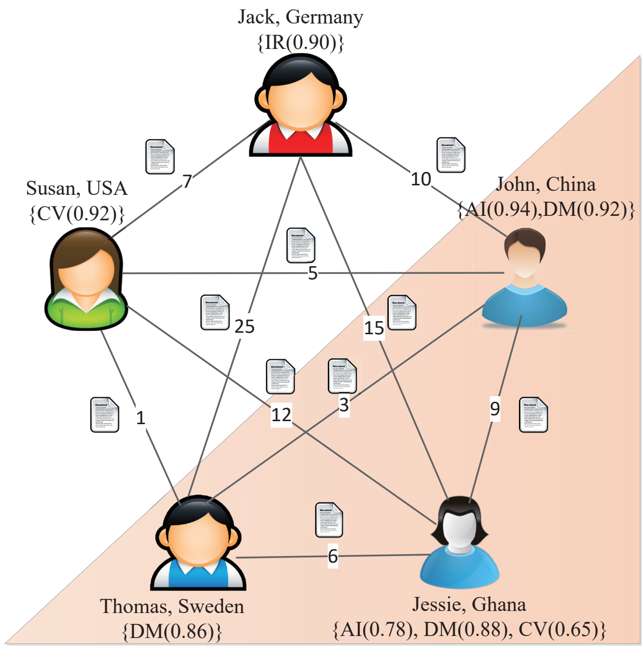


Fig. 1. Team Formation Application

number of guests who are attending a group traveling. For simplicity, the walking time between two hotels should be less than a given tolerant minutes. Mathematically, as depicted in Figure 2, candidate hotels can be viewed as the nodes in a fuzzy attributed social graph in which two nodes are connected if their walking distance is less than the given threshold. Besides, each hotel owns different attributes such as price and service quality. Recommending the optimal list of hotels is equivalent to identifying the skyline (λ, k)-cliques from a fuzzy attributed social networks. As shown in Figure 2, the three hotels and their mutual connections (green lines) labeled with a rounded rectangle form a skyline (λ, k)-clique.

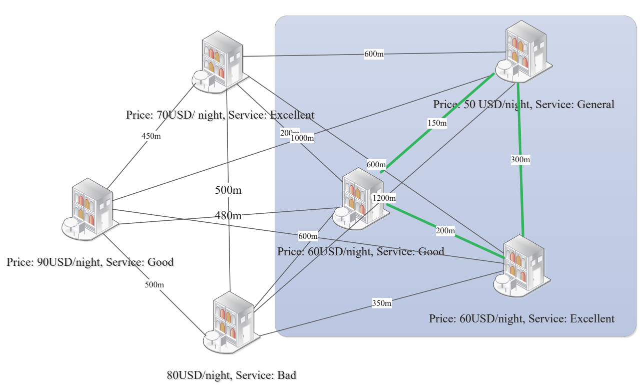


Fig. 2. Hotels Recommendation Application

From the theoretical modeling point of view, the above scenarios can be formulated as a novel skyline cohesive group modeling problem from a fuzzy attributed social network that is a fuzzy social network with the considerations of both structural property and attributes of nodes. Aiming to fill in this research gap and tackle the existing challenges, this work pioneers a novel problem on the skyline (λ, k)-cliques identification from a fuzzy attributed social network and proposes an FCA-based skyline (λ, k)-cliques identification approach.

Contributions. Different from the existing research about skyline k-cliques identification in a graph with the consideration of nodes' attributes only [10], our work considers the fuzzy feature of the edges as well as the attributes of nodes, and formulates a novel problem on skyline (λ, k)-cliques identification from a fuzzy attributed social network. Then, the corresponding identification approach is presented. In summary, this paper makes the following contributions.

- **Formulation of Skyline (λ, k)-cliques Identification from a Fuzzy Attributed Social Network:** We pioneer the formalization of skyline (λ, k)-cliques identification from a fuzzy attributed social network. When a fuzzy cut λ and an integer k are given, skyline (λ, k)-cliques identification returns the optimal (λ, k)-clique which can dominate other (λ, k)-cliques in d-dimensional space. In addition, we formally prove that the skyline (λ, k)-cliques identification in a fuzzy attributed social network is NP-hard.
- **FCA-based Skyline (λ, k)-cliques Identification Approach:** To make FCA applicable to our proposed problem, a fuzzy formal context is firstly constructed based on a fuzzy matrix to represent a fuzzy attributed social network. On one hand, considering the vertices in a fuzzy attributed social network with 2-dimensional attributes, a dominance formal context is established via the newly defined dominance matrix. Then, a directed skyline graph is constructed based on the obtained skyline layers. Given the fuzzy cut λ and the input parameter k, the (λ, k)-cliques identification approach based on FCA is presented. Finally, skyline (λ, k)-cliques can be discovered from a fuzzy attributed social network based on the dominance relationships among the (λ, k)-cliques.
- **Evaluation:** We conduct extensive experiments including an illustrative example on four real-world datasets. To be specific, the trend of processing time and size of (λ, k)-clique and skyline (λ, k)-clique are analysed by adjusting the fuzzy cut parameter λ and clique size k. Concretely, the number of (λ, k)-cliques decreases with the increase of k and λ since the number of k-cliques will decrease with large k. And, the number of skyline (λ, k)-cliques first increases and then decreases with the increase of k. However, the number of the skyline (λ, k)-cliques decreases with the increase of λ. The experimental results show that our proposed approach can better characterize the cohesive subgraphs (groups), especially the parameter λ can be viewed as a type of quality control parameter which is used for evaluating the stability of the groups.

Roadmap. The remainder of this paper is structured as follows. Section II overviews the related work on group skyline and skyline cliques. The preliminaries of this work as well as the problem formulation about skyline (λ, k)-cliques identification from a fuzzy attributed social network are provided in Section III. Section IV elaborates an FCA-based skyline

(λ, k)-cliques identification approach and the corresponding algorithms. The experimental results and analysis are reported in Section V. At last, Section VI concludes this paper including our future work.

II. Related Work

In this section, we will summarize the existing work on group skyline queries and cohesive subgraph detection from a network that are related to this work.

A. Group Skyline Queries

Skyline query processing is an important research issue in database field. Recent years have witnessed the development and extension of skyline definition. However, most of state-of-art literatures concentrate on traditional skyline query processing, such as skyline query over data stream [11], [12], and skyline query in the subspace [13]. In many real-world applications, we need to select multiple points, *i.e.*, a group of points instead of a single point. For example, in online sports competitive game industry, the players compete by selecting players from the real world to form their own luxury team, aiming to surpass other players. To this end, Zhang *et.al.* [14] extended the skyline and proposed the model of skyline groups that aimed to identify groups of points that are not dominated by other groups. Recently, Liu *et.al.* [15], proposed the g-skyline model and presented the relevant algorithms for group skyline problem including PointWise, UnitWise, and UnitWise+.

The common idea of the above algorithms is to generate a set enumeration tree including candidate point groups, and at the same time, prune the non-group skyline to enumerate the candidate group skyline. Further, the efficiency of g-skyline computation was investigated and advanced in [16]–[18]. In addition, the group skyline model has been applied over multi-valued attributed graphs [10], [19]. In [19], a skyline community model by using a k-core structure was presented for finding interesting communities from a multi-valued network. Recent work [10] formulated the novel model of skyline k-cliques over multi-valued attributed graphs by the virtue of skyline query and then developed efficient computational algorithms. In many real-world social networking applications, the social relationships are often uncertain, and fuzzy, that leads a huge challenge for community detection in a fuzzy social network. Unfortunately, the existing research on skyline communities or skyline cliques models cannot be directly used in such a fuzzy social network.

B. Cohesive Subgraph Detection

A cohesive subgraph is a primary vehicle for social networking analysis. There have been a large number of cohesive subgraph models, such as k-cliques [20], [21], k-clique community [20], maximal cliques [22], k-core [23], k-truss [24], and social-balanced densest subgraph [25] are emerging from complex networks/social networks. However, these models aim to process the topological structure of social networks only, and neglect the attributes of nodes. Li *et.al.* [26] investigated the problem of community detection in attributed graphs. They learned the associated attributes for underlying communities from the given node attributes and then proposed a novel community structure embedding method to encode inherent community structures for community detection purposes. To detect the cohesive subgraph from the attributed graph, Wang *et.al.* [27] adopted Non-negative Matrix Factorization (NMF) to combine network structure and node attributes. Xie *et.al.* [28] devised two influential community search algorithms by taking both the influence and node attributes into consideration. Concretely, their approach can efficiently identify the attributed pkd-truss community by maximizing the attribute and newly defined influence relevance scoring function. But these works consider a single attribute of the nodes only and thus the techniques are not applicable to the problem addressed in this paper. Regarding multi-valued attributed graph, both [19] and [10] are recent work to present a skyline community model/skyline clique model by combing skyline with cohesive graph models in multi-valued attributed graph.

Different from the previous group skyline models, this paper pioneers a novel group skyline model, termed Skyline (λ, k)-cliques, which characterizes both structural property and attributes of nodes together. Aiming to identify the skyline (λ, k)-cliques from a fuzzy attributed social network, FCA methodology is utilized for constructing the skyline layers and further generating the skyline (λ, k)-cliques.

III. Preliminary Knowledge and Problem Definition

In this section, the preliminary knowledge about this paper are first revisited. The problem of skyline (λ, k)-cliques identification from a Fuzzy Attributed Social Network is then formulated and its computational hardness is analyzed as well.

A. Preliminaries

Definition 1 ***(Dominance)*** *Given a set of points P in d-dimensional space, and p, $p^{'}$ are different nodes in P. If (1) $p.i \le p^{'}.i$ for all dimensions and (2) there exists at least one dimension such that $p.i < p^{'}.i$, $i \in [1, d]$. We say p dominates $p^{'}$, denoted as $p \preceq p^{'}$.*

Example 1 *A group of 2-dimensional points with group size 10 is shown in Table I.*

TABLE I
A SET OF 2-DIMENSIONAL DATA

P	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
x	5	5	14	33	26	10	38	21	30	15
y	450	400	360	340	300	250	200	150	120	80

Let us assume that the smaller the x and y, the better. For example, points $p_5 = (26, 300)$ and $p_6 = (10, 250)$, due to $p_6.x < p_5.x$ and $p_6.y < p_5.y$, p_6 dominates p_5, i.e., $p_6 \preceq p_5$.

Definition 2 ***(Strict Dominance)*** *Given a set of points P in d-dimensional space, and p, $p^{'}$ are different nodes in P. If*

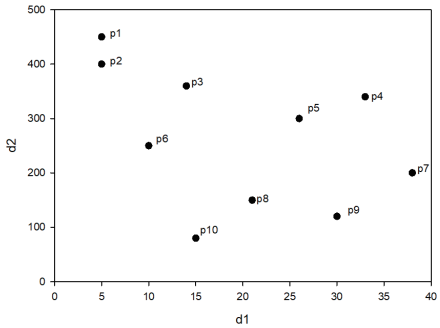


Fig. 3. A set of 2-dimensional Data Points

$p.i < p^{'}.i$ for all dimension, $i \in [1, d]$. We say p strictly dominates $p^{'}$, denoted as $p \prec p^{'}$.

Continue the above example, we can easily find that $p_6 \prec p_5$, $p_2 = (5, 400) \not\prec p_1 = (5, 450)$.

Definition 3 ***(Skyline)*** *Given a set of points P in d-dimensional space,* Skyline *refers to a set of points that are not dominated by other points.*

Definition 4 ***(Group Dominance)*** *Given a set of points P in d-dimensional space, and G=$\{p_1, p_2, \cdots, p_s\}$, $G^{'}$=$\{p^{'}_1, p^{'}_2, \cdots, p^{'}_s\}$ are two groups of points with group size s. If we can find two permutations of s points in G and $G^{'}$, i.e, G=$\{p_{u_1}, p_{u_2}, \cdots, p_{u_s}\}$, $G^{'}$=$\{p_{v_1}, p_{v_2}, \cdots, p_{v_s}\}$, so that $p_{u_i} \preceq p_{v_i}$ for all i ($1 \leq i \leq s$) and $p_{u_i} \prec p_{v_i}$ for at least one i, then G dominates $G^{'}$, termed $G \preceq_g G^{'}$.*

Definition 5 ***(Group Skyline)*** *A group skyline refers to a group of points that is not dominated by other groups of points.*

Example 2 *Let us continue Example 1, we also assume that the smaller the x and y, the better. Given two groups of points $G = \{p_8, p_9, p_{10}\}$ and $G^{'} = \{p_4, p_5, p_7\}$, we can find two permutations of G and $G^{'}$, i.e., $G = \{p_8, p_9, p_{10}\}$ and $G^{'} = \{p_5, p_4, p_7\}$, such that $p_8 \preceq p_5$, $p_9 \preceq p_4$, $p_{10} \preceq p_7$, therefore, $G \preceq_g G^{'}$.*

Definition 6 ***(Skyline Layer)*** *[15] Given a group of points P in d-dimensional space with group size n. The first Skyline Layer l_1 contains the skyline points for P, and we use S_1 to represent the set of points in l_1,* i.e., *$S_1 = Skyline(P)$; the second Skyline Layer l_2 contains the skyline points for $P - S_1$, and we use S_2 to represent the set of points in l_2, i.e., $S_2 = Skyline(P - S_1)$; and generally, the j_{th} Skyline Layer l_j contains the remaining skyline points except for the Skyline Layers from l_1 to l_{j-1}, denoted as $S_j = Skyline(P - \bigcup_{i=1}^{j-1} S_i)$.*

B. Problem Statement

This work studies the skyline (λ, k)-cliques over a fuzzy attributed social network. Thus, a fuzzy attributed social network with d-dimensional numeric attributes is represented as $\mathcal{G}$=$(V, E, \mathcal{A}, \sigma, \mu)$ where V indicates the set of nodes, E refers to the set of edges, $\mathcal{A} = \{\mathcal{A}_i\}$ denotes the attributes of nodes, σ is a fuzzy subset of a set V and μ: $E \rightarrow [0, 1]$ is a fuzzy relation on σ that assigns a degree of membership $\mu(e)$ to each edge $e \in E$. Each node $v \in V$ is linked with d-dimensional numeric attributes and the i_{th} dimensional value of v is represented as $v.D_i$. Without loss of generality, we assume smaller values are preferred. Note that this paper utilizes the terms node and point interchangeably.

Definition 7 *((λ, k)**-clique**) Given a fuzzy attributed social network $\mathcal{G}$, for a set of k nodes $C \subseteq V$, the clique's degree of membership of C, termed $cdm(C, \mathcal{G})$, is defined as the degree of membership in a graph sampled from $\mathcal{G}$, C is a k-clique. For a given fuzzy cut λ, C is called a (λ, k)-clique if $cdm(C, \mathcal{G}) \geq \lambda$.*

Definition 8 *(**Skyline** (λ, k)**-clique**) In a fuzzy attributed social network $\mathcal{G}$, if a (λ, k)-clique is not group dominated by other (λ, k)-cliques, then we call it as a skyline (λ, k)-clique.*

Problem Statement–(Skyline (λ, k)-cliques Identification from a Fuzzy Attributed Social Network). Given a fuzzy attributed social network $\mathcal{G}$, a fuzzy cut λ, and an integer k, each node $v \in V$ is associated with d-dimensional attributes, the goal of the proposed problem is to extract skyline (λ, k)-cliques from a fuzzy attributed social network $\mathcal{G}$.

Note that the above parameters λ and k are given by users empirically according to the requirements of applications.

Theorem 1 *The problem on the skyline (λ, k)-cliques identification from a fuzzy attributed social network is NP-hard.*

Proof *To prove the hardness of the above problem, we consider an instance: the d-dimensional attributes of all nodes share the same value in $\mathcal{G}$. That is to say, any (λ, k)-clique is also a skyline (λ, k)-clique as it is not group dominated by other (λ, k)-cliques. Hence, this problem is degraded as the (λ, k)-clique computation problem which is NP-hard.*

IV. Proposed Approach

In this section, we elaborate that how to use Formal Concept Analysis methodology for identifying the skyline k-cliques from a fuzzy attributed social network. First, a framework of FCA-based skyline k-cliques identification from a fuzzy attributed social network is provided. Then, the devised algorithm and its time complexity are presented in detail.

A. The Framework of FCA-based Skyline (λ, k)-cliques Identification from a Fuzzy Attributed Social Network

Figure 4 depicts a framework of FCA-based skyline (λ, k)-cliques identification from a fuzzy attributed social network. On the one hand, considering the nodes in a fuzzy attributed social network have 2-dimensional attributes, a dominance formal context is established via the newly defined dominance matrix. Then, a directed skyline graph is constructed based on the obtained skyline layers; on the other hand, a fuzzy formal context is constructed based on the fuzzy matrix of the given fuzzy attributed social network. Under the fuzzy cut λ and the input parameter k, the (λ, k)-cliques identification based on FCA is presented. Finally, skyline (λ, k)-cliques can be discovered from a fuzzy attributed social network based on the dominance relationships among the (λ, k)-cliques.

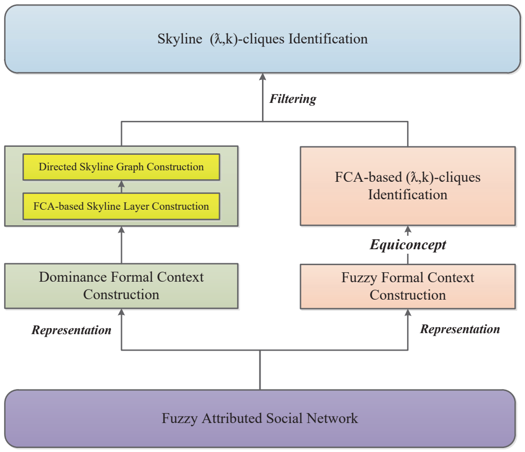


Fig. 4. The Framework of FCA-based Skyline (λ, k)-cliques Identification from a Fuzzy Attributed Social Network

B. Dominance Formal Context Construction

In order to represent the dominance relationships among nodes in d-dimensional space from a fuzzy attributed social network, we modify the traditional formal context and present the dominance formal context. To ensure the applicability of FCA to represent the dominance between nodes in a fuzzy attributed social network, the nodes are regarded as both objects and attributes in the constructed dominance formal context. Hence, a dominance formal context of a fuzzy attributed social network can be formalized as $K^{\mathcal{G}}_{\preceq}$=$(V, V, R)$ by the following dominance matrix, in which R is the dominance relationships among nodes.

Definition 9 ***(Dominance Matrix)*** *Let $\mathcal{G}$ be a fuzzy attributed social network with n nodes that are assumed to be ordered from v_1 to v_n. The $n \times n$ matrix D is called a dominance matrix, in which*

$$D = \begin{cases} d_{ij} = 1 & \textit{if} \quad (v_i.d_1 \leq v_j.d_1 \& v_i.d_2 \leq v_j.d_2), \\ d_{ij} = 1 & \textit{if} \quad i = j, \\ d_{ij} = 0 & \textit{otherwise.} \end{cases} \tag{1}$$

Hence, $K^{\mathcal{G}}_{\preceq}$ is equivalent to the dominance matrix of $\mathcal{G}$, *i.e.*, $K^{\mathcal{G}}_{\preceq} \equiv D$. Besides, a dominance formal context of a fuzzy attributed social network $K^{\mathcal{G}}_{\preceq}$ satisfies the following two properties.

- $K^{\mathcal{G}}_{\preceq}$ is a lower triangular matrix and asymmetry. The dominance relations between nodes are unilateral.
- All the diagonal elements are marked with "1".

Proof *For Property (1), we know that the dominance formal context stores the information about whether a node dominating another node. That is to say, if a node v_k dominates v_{k+1} in d-dimensional space (here d=2), then m_{ij} is marked with "1", otherwise "0". Obviously, these dominance relations form a lower triangular matrix and asymmetry. Naturally, it is also an asymmetry matrix.*
For Property (2), according to Definition 1, a given node v_i dominates itself, i.e., $v_i \preceq v_j$, *therefore the element $m_{ij} = 1 (i = j)$.*

Example 3 *Let us continue Example 1, we assume smaller values are preferred. Then, we assume that 10 points from a fuzzy attributed social network $\mathcal{G}$, are regarded as both objects and attributes. Their dominance relations among these points can be represented with the following dominance matrix.*

TABLE II
A DOMINANCE FORMAL CONTEXT $K^{\mathcal{G}}_{\preceq}$

V/V	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
p_1	1	0	0	0	0	0	0	0	0	0
p_2	1	1	0	0	0	0	0	0	0	0
p_3	0	0	1	0	0	0	0	0	0	0
p_4	0	0	0	1	0	0	0	0	0	0
p_5	0	0	0	1	1	0	0	0	0	0
p_6	0	0	1	1	1	1	0	0	0	0
p_7	0	0	0	0	0	0	1	0	0	0
p_8	0	0	0	1	1	0	1	1	0	0
p_9	0	0	0	1	0	0	1	0	1	0
p_{10}	0	0	0	1	1	0	1	1	1	1

Obviously, the above matrix is a lower triangular matrix with all the diagonal elements marked with "1". It is consistent with the properties of the dominance matrix.

C. Directed Skyline Graph Construction

A directed skyline graph is constructed with the skyline layers and the dominance relationships among them. Hence, skyline layers construction based on FCA is the first essential step to be provided. Then, a directed skyline graph is further presented by incorporating dominance relationships among skyline layers.

1) FCA-based Skyline Layers Construction: The basic idea of FCA-based skyline layers construction is to take the extents of i-extent concepts (i-extent concept is a special concept which contains i objects) as the i_{th} skyline layer. The steps for constructing skyline layers are as follows.

1) **Step 1**: By using the algorithm presented in [2], [20], we firstly generate the concept lattice $\mathcal{L}(K^{\mathcal{G}}_{\preceq})$ of the dominance formal context $K^{\mathcal{G}}_{\preceq}$ constructed from a given fuzzy attributed social network $\mathcal{G}$.

   Technically, Figure 5 shows the concept lattice generated from the dominance formal context as shown in Table II. Clearly, 13 formal concepts are obtained and they are organized as a Hasse diagram by a partial-order relation. Specifically, a formal concept, such as $(\{p_2\}, \{p_1, p_2\})$ in $\mathcal{L}(K^{\mathcal{G}}_{\preceq})$, is interpreted as "*the point p_2 dominates the points p_1, p_2*", *i.e.*, $p_1 \preceq p_1$ and $p_1 \preceq p_2$.
2) **Step 2**: Our aim at the first skyline layer is to find out some points that can dominate the other points as many as possible. That is to say, we should identify some special formal concepts so that the cardinality of intent is maximized and of the extent is minimized. Due to the extent dominating the intent in obtained formal concepts, we first find out the concepts which contain the smallest

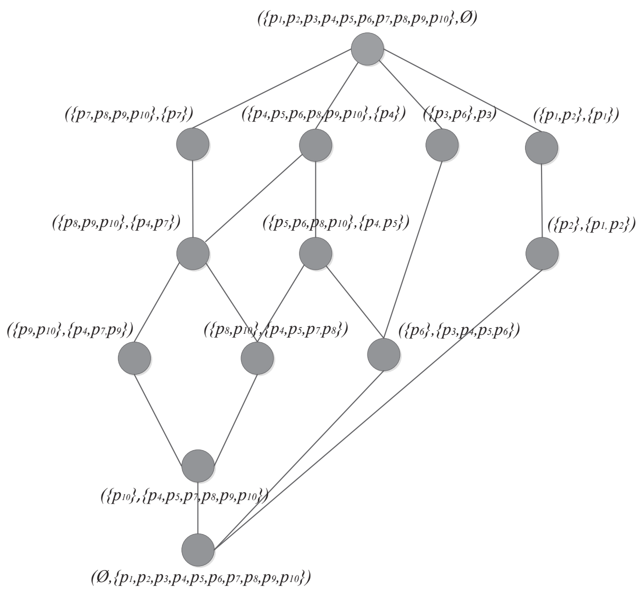


Fig. 5. Concept Lattice of Dominance Formal Context (as shown in Table II)

cardinality of extent, then combine the extents of these concepts and finally form the first skyline layer l_1.

$$S_1 := \bigcup_{i=1}^{m} \arg\min_{A_i} |A_i| \,. \tag{2}$$

As can be seen from Figure 5, the formal concepts $(\{p_2\},\{p_1,p_2\})$, $(\{p_6\},\{p_3,p_4,p_5,p_6\})$, and $(\{p_{10}\},\{p_4,p_5,p_7,p_8,p_9,p_{10}\})$ satisfy that the cardinality of extent is the smallest, thus the corresponding extents are stored as the first skyline layer l_1, *i.e.*, S_1=$\{p_2,p_6,p_{10}\}$.

3) **Step 3**: After constructing the first skyline layer l_1, we will construct other skyline layers, *i.e.*, $l_2, l_3, \cdots, l_j$. Regarding the construction of l_2, we first remove the points included in S_1 and then repeat Step 2. Simply, we will find out the concepts which contain the smallest cardinality of extent, then combine the extents of these concepts and finally form the second skyline layer l_2, *i.e.*, S_2=$\{p_1,p_3,p_8,p_9\}$. Therefore, this working process is formally described as follows.
   As for the construction of l_i, we first remove the points included in S_{i-1} and then repeat Step 2. Finally, the set of points in l_i is stored in S_i.

Eventually, the skyline layers are constructed as shown in Figure 6.

Clearly, four skyline layers are constructed, *i.e.*, S_1=$\{p_2,p_6,p_{10}\}$, S_2=$\{p_1,p_3,p_8,p_9\}$, S_3=$\{p_5,p_7\}$, and S_4=$\{p_4\}$.

2) Directed Skyline Graph Construction: Directed skyline graph, an important data structure for characterizing the points from the top j skyline layers as well as their dominance relationships, in order to identify skyline (λ, k)-cliques from a fuzzy attributed social network. This section focuses on how to construct a directed skyline graph by incorporating the dominance relationships among skyline layers.

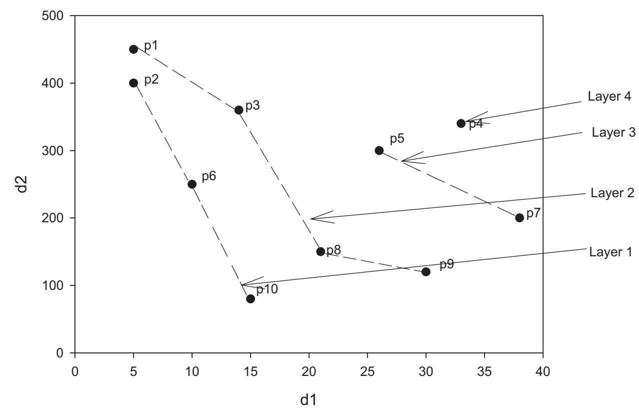


Fig. 6. The Skyline Layers

Definition 10 ***(Directed Skyline Graph)*** *[15] A directed skyline graph can be represented as a directed graph G=(V,E) with V indicating the set of points and E indicating the set of dominance relation between points. Each node has a structure as shown in Figure 7.*

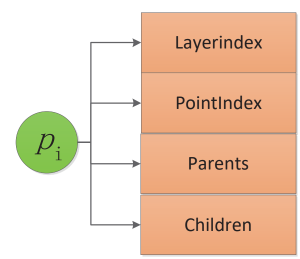


Fig. 7. Data Structure of a Point p_i

Note that the layer index ranging from 1 to j indicates the skyline layer that the point lies on, point index ranging from 0 to S_j-1 and S_j refers to the number of points in the first j skyline layers, parents include all points that dominate this point, and children include all the points that are dominated by the point.

According to Definition 10 and the obtained skyline layers in the previous section, the points' information is listed in Table III.

TABLE III
POINTS' INFORMATION

point	layer index	point index	parents	children
p_1	2	3	p_2	$\emptyset$
p_2	1	0	$\emptyset$	p_1
p_3	2	4	p_6	$\emptyset$
p_4	4	9	p_4	p_4
p_5	3	7	p_5, p_6, p_8, p_{10}	p_4, p_5
p_6	1	1	$\emptyset$	p_3, p_4, p_5, p_6
p_7	3	8	p_7, p_8, p_9, p_{10}	p_4, p_7
p_8	2	5	p_8, p_{10}	p_4, p_5, p_7, p_8
p_9	2	6	p_9, p_{10}	p_4, p_7, p_9
p_{10}	1	2	$\emptyset$	$p_4, p_5, p_7, p_8, p_9, p_{10}$

A directed skyline graph of the point set can be constructed as shown in Figure 8, in which the point index value of the node and transitive dominance relationship between points are omitted. For instance, $p_{10} \preceq p_7$ can be inferred based on the transitivity of dominance relationships, *i.e.*, $p_{10} \preceq p_9$ and $p_9 \preceq p_7$, thus $p_{10} \preceq p_7$.

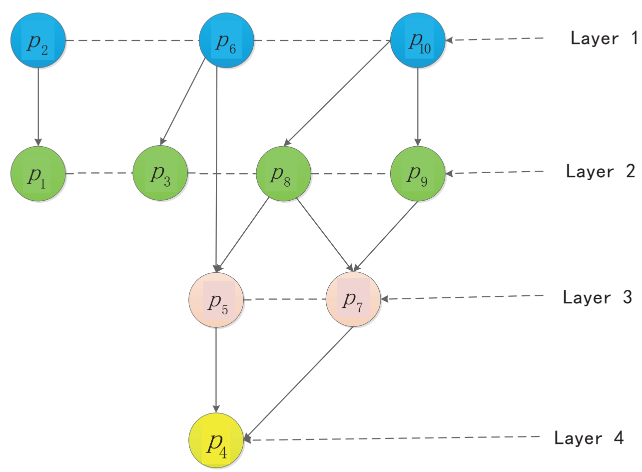


Fig. 8. Directed Skyline Graph

D. Skyline (λ, k)-cliques Identification

Skyline (λ, k)-cliques identification from a fuzzy attributed social network is composed of two technical parts. First, we should identify the (λ, k)-cliques from a given fuzzy attributed social network. Second, the skyline (λ, k)-cliques are extracted based on group dominance among these (λ, k)-cliques.

1) (λ, k)-cliques Identification from a Fuzzy Attributed Social Network: According to the problem statement, a fuzzy attributed social network $\mathcal{G}$ can be represented as a set of nodes in which some of them have fuzzy relationships. To describe this fuzzy relation between nodes, the nodes are viewed as both objects and attributes. Then, a fuzzy formal context of $\mathcal{G}$ can be constructed with the following fuzzy adjacency matrix R^*, denoted as $K(\mathcal{G})$=(V, V, R^*) [29].

Definition 11 ***(Fuzzy Adjacency Matrix)*** *A fuzzy adjacency matrix R^*=$(r_{ij})_{m\times n}$ is an $m \times n$ matrix if*

$$R^* = \begin{cases} r_{ij} = \mu(e_{ij}) & \textit{if} \quad (v_i, v_j) \in E, \\ r_{ij} = 1 & \textit{if} \quad i = j, \\ r_{ij} = 0 & \textit{otherwise.} \end{cases} \tag{3}$$

However, the above fuzzy formal context is not easy to handle. With the given fuzzy cut λ, the above fuzzy formal context can be simplified by modifying the r_{ij}=1 if $\mu(e_{ij}) \geq \lambda$, otherwise, r_{ij}=0. Intuitively, λ is a type of quality control parameter for evaluating the cohesiveness among nodes.

To detect the cliques from a fuzzy attributed social network, the above fuzzy attributed social network is firstly transformed into an undirected and unweighted social graph by removing the degree of membership from the edges. In our previous work [20], it is proved that there exist the one-to-one mappings from cliques is to equiconcepts from the concept lattice of the fuzzy attributed social network.

Example 4 *Figure 9 shows a typical fuzzy attributed social network including 10 nodes and 19 edges. The degrees of membership are assigned to each edge. Each node has 2-dimensional attributes as shown in Table I. For instance, there exists a fuzzy relation between p_2 and p_8 with a degree of membership of 0.33.*

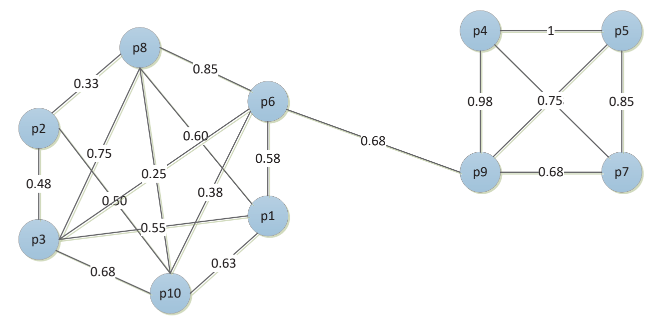


Fig. 9. A Fuzzy Attributed Social Network $\mathcal{G}$

With Definition 7, we suppose that the fuzzy cut λ is 0.35 and k is 3, then the subgraph g_1 formed by nodes $\{p_1, p_6, p_{10}\}$, is a (λ, k)-clique since $cdm(g_1, \mathcal{G})$=0.38$>\lambda$. As matter of fact, all (λ, k)-cliques can be easily obtained as follows: $\{p_1, p_3, p_8\}$, $\{p_1, p_3, p_{10}\}$, $\{p_1, p_6, p_8\}$, $\{p_1, p_6, p_{10}\}$,$\{p_2, p_3, p_{10}\}$.

2) Skyline (λ, k)-cliques Identification from a Fuzzy Attributed Social Network: Figure 8 derives the following observation: according to the definition of skyline layer, the points on the l_1 are skyline points of the whole point set P, which dominate the points on other layers except the l_1. The points on the l_2 are the skyline points of the subsets of P after removing the points from the l_1, that is, the points on the second layer dominate the points on the l_1 and other layers outside the l_2. The points on the l_3 are the skyline points of the subsets left by the set P to remove the points on the layers l_1 and l_2, that is, the points on the l_3 dominate the points except for the first three layers. We can easily get the following conclusion.

The points on the lower layer dominate that on the higher layer, and the values of some attributes on the points from the lower layer are not worse than those from the higher layer, and the values of at least one attribute on the points from the lower layer are better than those from the higher layer, that is to say, the points on the lower layer are better than those on the higher layer. Therefore, if the number of points from the lower level in a group of points is more, then the group of points will be better.

According to Definition 4, we define dominance relationship between two (λ, k)-cliques g and $g^{'}$ as follows.

Definition 12 *$(g \preceq g^{'})$ Given two (λ, k)-cliques g and $g^{'}$, we first compare the number of points from skyline layer l_1 in g and $g^{'}$, if they are the same, then continue to compare the number of points from other skyline layers $l_2, l_3, \cdots, l_m$ in g and $g^{'}$ until the number of points from skyline layer l_i in g is greater than that of $g^{'}$ or the number of points from all layers in g and $g^{'}$ are the same.*

Example 5 *Let us continue Example 4, we use $g_1, g_2, \cdots, g_5$ to represent the (λ, k)-cliques, $\{p_1, p_6, p_8\}$, $\{p_2, p_3, p_{10}\}$,*

$\{p_1, p_3, p_{10}\}$, $\{p_1, p_3, p_8\}$,$\{p_1, p_6, p_{10}\}$, respectively. According to Definition 12, the dominance relationships among all (λ, k)-cliques (λ=0.38, k=3) can be represented with the following dominance matrix.

TABLE IV
A DOMINANCE MATRIX BETWEEN (λ, k)-CLIQUES

$g/g^{'}$	g_1	g_2	g_3	g_4	g_5
g_1	1	0	1	1	0
g_2	1	1	1	1	1
g_3	1	0	1	1	0
g_4	0	0	0	1	0
g_5	1	1	1	1	1

Clearly, we notice that the all elements in the 2_{nd} and the 5_{th} rows of the above dominance matrix are "1". It implies that (λ, k)-cliques g_2 and g_5 dominate all other (λ, k)-cliques. Therefore, g_2 and g_4 are the skyline (λ, k)-cliques in this fuzzy attributed social network $\mathcal{G}$.

E. Algorithm Description

With the above-mentioned framework and approach, we develop the corresponding algorithms for (λ, k)-cliques identification from a fuzzy attributed social network. This section mainly presents 3 pseudo-code of our implementation procedures, including (1) (λ, k)-cliques mining from a fuzzy attributed social network; (2) skyline (λ, k)-cliques identification; (3) a sub-algorithm of Algorithm 2 is to obtain the number of points on the skyline layer l_i.

Algorithm 1: (λ, k)-cliques Mining from a Fuzzy Attributed Social Network

Input: A fuzzy attributed social network $\mathcal{G}$; A fuzzy cut λ and a parameter k
Output: A set of (λ, k)-cliques τ
1 Initialize τ=$\emptyset$
2 **begin**
3 Fuzzy formal context $K(\mathcal{G})$ construction via fuzzy adjacency matrix
4 $K(\mathcal{G})$ refinement by filtering out the membership values which are less than λ
5 Fuzzy concept lattice $\mathcal{L}(K(\mathcal{G}))$ building
6 **end**
7 **for** each concept $(X, B) \in \mathcal{L}(K(\mathcal{G}))$
8 **begin**
9 **if** X=B and $|X| = |B| = k$
10 $\tau \leftarrow \tau \cup (X, B)$
11 **end**
12 **if** X=B and $|X| = |B| > k$
13 **for** i=$k+1$ **to** M **do**
14 **begin**
15 $\tau \leftarrow \tau \cup Derived((X^i, B^{'}))$
16 **end**

Algorithm 1 proceeds as: First, the whole algorithm includes the inputs of a fuzzy attributed social network $\mathcal{G}$, a fuzzy cut λ and a parameter k; then, a set of (λ, k)-cliques with τ is initialized (Line 1). After that, it goes into the procedures of fuzzy formal context construction and concept lattice generation (Lines 2-6). Lines 7-11 insert the k-equiconcepts (X, B)(*i.e.*, explicit (λ, k)-cliques) into τ. The remaining set of (λ, k)-cliques can be derived from other high-order equiconcepts and finally be inserted into τ (Lines 12-16).

Based on the (λ, k)-cliques detected via Algorithm 1, Algorithm 2 is in charge of identifying skyline (λ, k)-cliques from a given fuzzy attributed social network $\mathcal{G}$.

Algorithm 2: Skyline (λ, k)-cliques Identification Algorithm

Input: Two (λ, k)-cliques g and $g^{'}$, layers
Output: 1 or 0
1 $result$=0, $layer$=1;
2 **while** $layer < layers.size$ **do**
3 **if** $getNumber(g) > getNumber(g^{'})$ **then**
4 *result*=1, **break**;
5 **else if** $getNumber(g) < getNumber(g^{'})$ **then**
6 **break**
7 **else if** $getNumber(g) == getNumber(g^{'})$ **then**
8 ++*layer*;
9 **return** *result*;

Its working process is described as follows. Given two (λ, k)-cliques g and $g^{'}$, we first initialize a variable $result$ and $layer$ (Line 1). Then, it traverses all the skyline layers and calculates the number of points in each layer by invoking Algorithm 3. If the number of points in l_i of g is greater than that of l_i in $g^{'}$, then $result \leftarrow 1$, otherwise, it will continue to traverse the other skyline layers until all layers are checked (Lines 2-8).

Algorithm 3: *GetNumber* Algorithm

Input: A (λ, k)-clique g, layers, l_i
Output: The number of points on the l_i
count=0
for each point p_i in g **do**
for each point p_i in l_i **do**
if $p_i == p_j$ **then**
++count;
break;
if *count*==0 **then**
count=1;
return *count*;

F. Time Complexity Analysis

This section focuses on discussion about the time complexity of skyline (λ, k)-cliques identification in fuzzy attributed social networks. First, the time complexity about Algorithm 1 for obtaining the (λ, k)-cliques is $\Theta(|V|^3)$ ($|V|$ is the number of vertices of $\mathcal{G}$) since this procedure is similar to the k-cliques detection in social networks [20]. Then, the time complexity of Algorithm 2 for identifying the skyline (λ, k)-cliques is composed of time complexity of Algorithm 3 and itself. We denote $|\tau|$ as the number of (λ, k)-cliques, $|l_i|$ as the number of points on the layer l_i. Hence, the operational complexity of Algorithm 3 is $\Theta(k \times |l_i|)$. Since Algorithm 2 traverses all layers, the operational complexity is $|\tau|^2|L|k\sum|l_i|$=$|\tau|^2|L|k^2$ (here, L is the number of layers). In summary, the time complexity of the Algorithm 1 is $\Theta(|V|^3 + |\tau|^2|L|k^2)$.

V. Experiments and Analysis

This section is devoted to evaluating the proposed approach via conducting empirical studies as well as an illustrative example on four real-world networks. The experiments aim to validate the effectiveness of the proposed approach for detecting the skyline (λ, k)-cliques from a fuzzy attributed social network.

A. Experiment Setup

Initially, the following three datasets of social networks are described.

Dataset I: This is a human contact social network, called Train bombing dataset [30]. It includes contacts between suspected terrorists involved in the train bombing of Madrid on March 11, 2004 as reconstructed from newspapers. This Dataset is a time series that treat specific attacks as endpoints and depict the evolution of relations between individuals indirectly and directly associated with the Madrid train bombing[1]. Concretely, a node represents a terrorist and an edge between two terrorists indicates that there was a contact between the two terrorists. The edge weights denote how "*strong*" the connection was. This connections contain friendship and co-participating in training camps or previous attacks.

Dataset II: This data set is composed of the carbon exchanges in the cypress wetlands of South Florida during the dry season [2], called Florida ecosystem dry dataset [31]. Nodes in this dataset represent taxon and an edge denotes that a taxon uses another taxon as food with a given trophic factor (feeding level). The field concerned with the network analysis of such structures is called trophic network.

Dataset III: It is a weighted, directed network representing the neural network of Caenorhabditis elegans, termed CEG dataset [32], [33]. It includes 297 nodes representing the neurons which are higer-level grouping of the brain of Caenorhabditis elegans and 2148 connections between neurons [3]. The field concerned with the network analysis of such structures is called network neuroscience.

Table V shows the critical statistics of these datasets. Note that d_{avg} denotes average degree, and d_{max} refers to the maximum degree of each dataset.

Note that the nodes in the experimental datasets do not own multi-valued numerical attributes, therefore we generate the multi-valued numerical attributes values for the nodes of these datasets by using the commonly used approach in skyline processing [34].

[1]http://en.wikipedia.org/wiki/2004_Madrid_train_bombings
[2]http://vlado.fmf.uni-lj.si/pub/networks/data/bio/foodweb/foodweb.htm
[3]https://www.cc.gatech.edu/dimacs10/archive/clustering.shtml

TABLE V
Statistics of Datasets.

Dataset	Vertices	Edges	d_{avg}	d_{max}
Train bombing	64	243	7.59	29
Florida ecosystem dry	128	2137	33.4	110
CEG	297	2148	7.896	134

We will evaluate and compare our algorithm with the *Baseline* approach which conducts the pair-wise dominance checking by Definition 4 to exhaustively enumerate all possible permutations of each group. Experimentally, we conduct various experiments by adjusting the parameters including fuzzy cut λ, clique size k on a default dimension d=2 of the datasets.

B. Results Analysis

All algorithms are implemented in JAVA language and are run on an Intel(R) Core (TM) i7-8565U @ 1.80GHz 1.99GHz, 20GB RAM computer. In this section, we conduct the various experiments by adjusting the parameters λ, k.

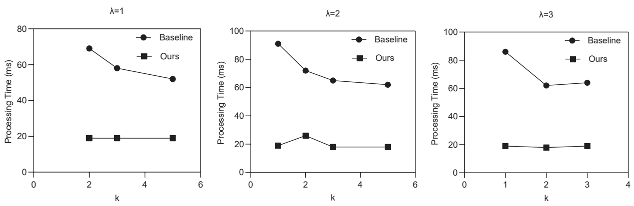
Experiment-1: Size of Skyline Layers. By considering the multi-valued attributes for the nodes in the experimental datasets, we constructed the directed skyline graphs for them. And, the size of the constructed skyline layers under different λ (*i.e.*, different network sizes) are shown in Table VI.

TABLE VI
Size of Skyline Layers.

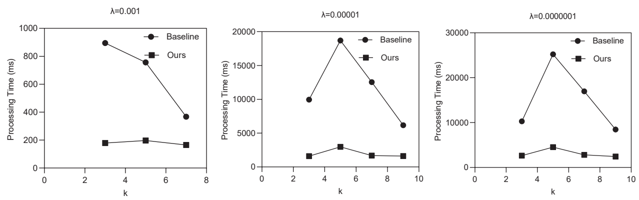
Dataset	λ	Network Size	Number of Skyline Layers
Dataset I	1	64*243	14
	2	64*29	
	3	64*8	
Dataset II	0.001	128*787	23
	0.00001	128*1866	
	0.0000001	128*2123	
Dataset III	1	297*2148	32
	2	297*1317	
	3	297*933	

It is observed that the number of skyline layers is not changed when the parameter λ is updated. In addition, the number of skyline layers is proportional to the network size.

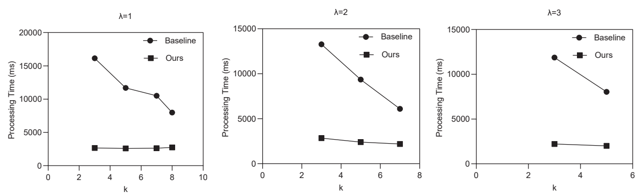
Experiment-2: Varying k and λ. Figures 10(a)-10(c) show the time cost comparison of our algorithm and baseline algorithm when varying the k from 1 to 5 and λ from 1 to 3 for Dataset I. Figures 10(d)-10(f) report the time cost comparison of our algorithm and baseline algorithm when varying the k from 3 to 9 and λ is 0.001, 0.00001, 0.0000001 for Dataset II. Figures 10(g)-10(i) demonstrate the time cost comparison of our algorithm and baseline algorithm when varying the k from 3 to 9 and λ from 1 to 3 for Dataset III. It is easily to find that our algorithm can significant reduce the processing time for identifying the skyline (λ, k)-cliques. The running time of identifying the skyline (λ, k)-cliques is depending on how many skyline (λ, k)-cliques are detected from a fuzzy attributed social network. For instance, the time cost is 4537 ms when k=5, λ=0.0000001, and the number of skyline (λ, k)-cliques reaches the largest value.



(a) Dataset I, λ=1 (b) Dataset I, λ=2 (c) Dataset I, λ=3



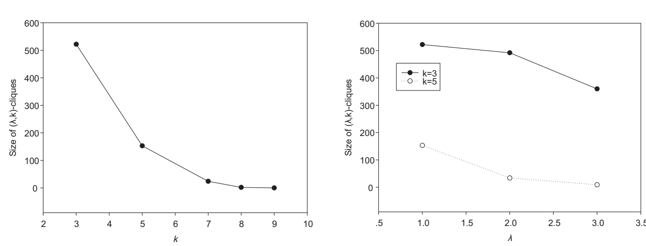
(d) Dataset II, λ=0.001 (e) Dataset II, λ=0.00001 (f) Dataset II, λ=0.0000001



(g) Dataset III, λ=1 (h) Dataset III, λ=2 (i) Dataset III, λ=3

Fig. 10. Processing Time of Compared Algorithms Varying k and λ.

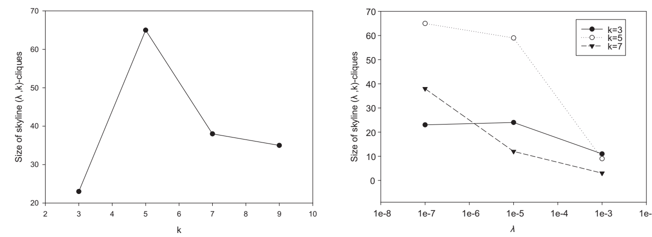
Experiment-3: Size of (λ, k)-cliques. We test the trends between the size of (λ, k)-cliques and parameters λ, k, respectively. Figure 11 reports the size of (λ, k)-cliques set. The dataset adopted here is *CEG* and similar trends can be captured in the other two datasets. Obviously, the number of (λ, k)-cliques decreases with the increase of k and λ since the number of k-cliques will decrease with large k.



(a) Varying k (b) Varying λ

Fig. 11. The Size of (λ, k)-cliques in CEG Dataset.

Experiment-4: Size of Skyline (λ, k)-cliques. Similarly, we also exam the trends between the size of the skyline (λ, k)-cliques and parameters λ, k, respectively. The size of the skyline (λ, k)-cliques set is shown in Figure 12. The dataset adopted here is *Florida ecosystem dry* and similar trends can be observed in the other two datasets. Obviously, the number of skyline (λ, k)-cliques first increases and then decreases with the increase of k. However, the number of the skyline (λ, k)-cliques decreases with the increase of λ.



(a) Varying k (b) Varying λ

Fig. 12. The Size of Skyline (λ, k)-cliques in Florida ecosystem dry Dataset.

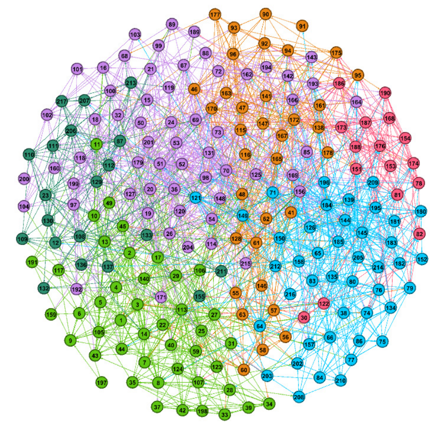


Fig. 13. The Visualization of Dataset in Illustrative Example

C. Illustrative Example

We use the friendship data among the 217 residents living at a residence hall located on the Australian National University campus [35] for our illustrative example. The dataset consists of 217 nodes and 2672 edges. A node represents a person and an edge shows that there was contact between the two persons. The edge weights indicate the strength of each friendship tie. In addition, two numerical attributes are assigned to each person. The average degree d_{avg} is 24.627 and the maximum degree d_{max} is 80. The visualization of this network is shown in Figure 13.

By using the proposed detection approach on the skyline (λ, k)-cliques, Tables VII-IX present the experimental results when a fuzzy cut λ ranging from 2 to 4, and $k \in \{3, 5, 7, 9\}$. Note that both the (λ, k)-cliques and skyline (λ, k)-cliques are not available in the network of the illustrative example.

TABLE VII
DETECTION RESULTS WHEN λ=2.

λ=2	(λ, k)-cliques	skyline (λ, k)-cliques	Time (ms)
k=3	275	28	2949
k=5	162	4	2580
k=7	42	4	2463
k=9	3	2	2700

Figure 14 shows four groups, *i.e.*, skyline (2,7)-cliques identified from the above network dataset. Their topology is as follows.

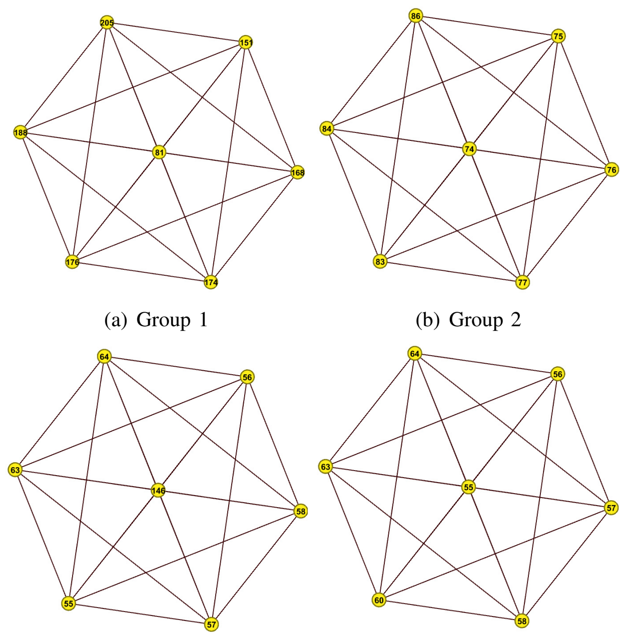
TABLE VIII
DETECTION RESULTS WHEN λ=3.

λ=3	(λ, k)-cliques	skyline (λ, k)-cliques	Time (ms)
k=3	267	27	2612
k=5	120	4	2632
k=7	30	3	2647
k=9	3	2	2223

TABLE IX
DETECTION RESULTS WHEN λ=4. ("-" INDICATES THE RESULTS ARE NOT AVAILABLE)

λ=4	(λ, k)-cliques	skyline (λ, k)-cliques	Time (ms)
k=3	138	8	2021
k=5	9	1	2638
k=7	-	-	-
k=9	-	-	-

- Group 1: $\{55, 56, 57, 58, 60, 63, 64\}$,
- Group 2: $\{74, 75, 76, 77, 83, 84, 86\}$,
- Group 3: $\{55, 56, 57, 58, 63, 64, 146\}$,
- Group 4: $\{81, 151, 168, 174, 176, 188, 205\}$.



(a) Group 1 (b) Group 2

(c) Group 3 (d) Group 4

Fig. 14. Skyline (2,7)-cliques.

Interestingly, Group 4 is removed from the set of skyline (2,7)-cliques when λ=3 since its clique's degree of membership is less than 3. The skyline (3,7)-cliques are listed as follows.

- Group 1: $\{55, 56, 57, 58, 60, 63, 64\}$,
- Group 2: $\{74, 75, 76, 77, 83, 84, 86\}$,
- Group 3: $\{55, 56, 57, 58, 63, 64, 146\}$.

However, the skyline (4,7)-cliques are not available from the network of illustrative example due to the higher value of the friendship quality control parameter λ. Therefore, when k is a constant value, the skyline (λ, k)-cliques are not changed if the friendship quality control parameter λ is updated. It implies that the constant skyline (λ, k)-cliques represents the stable groups.

VI. CONCLUSIONS

This paper proposed a novel skyline (λ, k)-clique model to identify cohesive subgraphs from a fuzzy attributed social network. An efficient FCA-based skyline (λ, k)-cliques identification approach is presented by representing the fuzzy attributed social network with a fuzzy formal context and constructing the dominance formal context as well as the directed skyline graph for describing the dominance relations between nodes. The algorithms on (λ, k)-cliques and skyline (λ, k)-cliques identification are developed, respectively. The effectiveness of the skyline (λ, k)-clique model is evaluated by experiments on three real datasets. From the practice point of view, an illustrative example was also conducted to reveal the usefulness of the proposed skyline (λ, k)-clique model in a fuzzy attributed social network. In the future, we plan to investigate the evolution of skyline (λ, k)-cliques in a fuzzy attributed social network.

ACKNOWLEDGMENT

This work was funded in part by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 840922, the National Natural Science Foundation of China (Grant No. 61702317), the Fundamental Research Funds for the Central Universities (GK202103080). 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] S. Hu, L. Zou, J. X. Yu, H. Wang, and D. Zhao, "Answering natural language questions by subgraph matching over knowledge graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 5, pp. 824–837, 2017.

[2] 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.

[3] A. E. Sarıyüce, B. Gedik, G. Jacques-Silva, K.-L. Wu, and Ü. V. Çatalyürek, "Incremental k-core decomposition: algorithms and evaluation," *The VLDB Journal*, vol. 25, no. 3, pp. 425–447, 2016.

[4] K. Burleson-Lesser, F. Morone, M. S. Tomassone, and H. A. Makse, "K-core robustness in ecological and financial networks," *Scientific reports*, vol. 10, no. 1, pp. 1–14, 2020.

[5] X. Huang, H. Cheng, L. Qin, W. Tian, and J. X. Yu, "Querying k-truss community in large and dynamic graphs," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014, pp. 1311–1322.

[6] E. Akbas and P. Zhao, "Truss-based community search: a truss-equivalence based indexing approach," *Proceedings of the VLDB Endowment*, vol. 10, no. 11, pp. 1298–1309, 2017.

[7] J. Zhu, B. Chen, and Y. Zeng, "Community detection based on modularity and k-plexes," *Information Sciences*, vol. 513, pp. 127–142, 2020.

[8] C. Zhang, Y. Zhang, W. Zhang, L. Qin, and J. Yang, "Efficient maximal spatial clique enumeration," in *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. IEEE, 2019, pp. 878–889.

[9] Q. Li, Y. Zhu, and J. X. Yu, "Skyline cohesive group queries in large road-social networks," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020, pp. 397–408.

[10] C. Zhang, W. Zhang, Y. Zhang, L. Qin, F. Zhang, and X. Lin, "Selecting the optimal groups: Efficiently computing skyline k-cliques," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1211–1220.

[11] L. Zhu, C. Li, and H. Chen, “Efficient computation of reverse skyline on data stream,” in *2009 International Joint Conference on Computational Sciences and Optimization*, vol. 1. IEEE, 2009, pp. 735–739.

[12] Z. Wang, J. Xin, L. Ding, J. Ba, and X. Gao, “ρ-dominant skyline computation on data stream,” *IEEE Access*, vol. 6, pp. 53 201–53 213, 2018.

[13] T. Jiang, B. Zhang, D. Lin, Y. Gao, and L. Qing, “Efficient column-oriented processing for mutual subspace skyline queries,” *Soft Computing*, pp. 1–19, 2020.

[14] N. Zhang, C. Li, N. Hassan, S. Rajasekaran, and G. Das, “On skyline groups,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 4, pp. 942–956, 2013.

[15] J. Liu, L. Xiong, J. Pei, J. Luo, and H. Zhang, “Finding pareto optimal groups: Group-based skyline,” *Proceedings of the VLDB Endowment*, vol. 8, no. 13, pp. 2086–2097, 2015.

[16] C. Wang, C. Wang, G. Guo, X. Ye, and S. Y. Philip, “Efficient computation of g-skyline groups,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 4, pp. 674–688, 2017.

[17] W. Yu, Z. Qin, J. Liu, L. Xiong, X. Chen, and H. Zhang, “Fast algorithms for pareto optimal group-based skyline,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 417–426.

[18] K. Li, Z. Yang, G. Xiao, K. Li *et al.*, “Progressive approaches for pareto optimal groups computation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 3, pp. 521–534, 2018.

[19] R.-H. Li, L. Qin, F. Ye, J. X. Yu, X. Xiao, N. Xiao, and Z. Zheng, “Skyline community search in multi-valued networks,” in *Proceedings of the 2018 international conference on management of data*, 2018, pp. 457–472.

[20] 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.

[21] F. Hao, Z. Pei, D.-S. Park, L. T. Yang, Y.-S. Jeong, and J.-H. Park, “Iceberg clique queries in large graphs,” *Neurocomputing*, vol. 256, pp. 101–110, 2017.

[22] Y. Yang, F. Hao, B. Pang, G. Min, and Y. Wu, “Dynamic maximal cliques detection and evolution management in social internet of things: A formal concept analysis approach,” *IEEE Transactions on Network Science and Engineering*, 2021.

[23] F. Hao, K. Xinchang, and D.-S. Park, “Fca-based θ-iceberg core decomposition in graphs,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–6, 2017.

[24] H. Kabir and K. Madduri, “Parallel k-truss decomposition on multicore systems,” in *2017 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2017, pp. 1–7.

[25] F. Hao, D.-S. Park, Z. Pei, H. Lee, and Y.-S. Jeong, “Identifying the social-balanced densest subgraph from signed social networks,” *The Journal of Supercomputing*, vol. 72, no. 7, pp. 2782–2795, 2016.

[26] Y. Li, C. Sha, X. Huang, and Y. Zhang, “Community detection in attributed graphs: An embedding approach,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[27] X. Wang, D. Jin, X. Cao, L. Yang, and W. Zhang, “Semantic community identification in large attribute networks,” in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[28] X. Xie, M. Song, C. Liu, J. Zhang, and J. Li, “Effective influential community search on attributed graph,” *Neurocomputing*, vol. 444, pp. 111–125, 2021.

[29] F. Hao, D.-S. Park, S. Li, and H. M. Lee, “Mining λ-maximal cliques from a fuzzy graph,” *Sustainability*, vol. 8, no. 6, p. 553, 2016.

[30] B. Hayes, “Connecting the dots,” *American Scientist*, vol. 94, no. 5, pp. 400–404, 2006.

[31] R. E. Ulanowicz and D. L. DeAngelis, “Network analysis of trophic dynamics in south florida ecosystems,” *US Geological Survey Program on the South Florida Ecosystem*, p. 114, 1999.

[32] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-worldnetworks,” *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[33] J. Kunegis, “Konect: the koblenz network collection,” in *Proceedings of the 22nd international conference on world wide web*, 2013, pp. 1343–1350.

[34] S. Borzsony, D. Kossmann, and K. Stocker, “The skyline operator,” in *Proceedings 17th international conference on data engineering*. IEEE, 2001, pp. 421–430.

[35] L. C. Freeman, C. M. Webster, and D. M. Kirke, “Exploring social structure using dynamic three-dimensional color images,” *Social networks*, vol. 20, no. 2, pp. 109–118, 1998.

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. 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.

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 computing and formal concept analysis.

Jianrui Chen received the Ph.D. degree in computer science and engineering from Xidian University, China, in 2012. She is currently Professor of School of Computer Science, Shaanxi Normal University, Xian, China. She has authored and coauthored lots of papers in top journals and conferences. Her research interests are broadly in the area of artificial intelligence, network analysis, data mining.

Aziz Nasridinov received the B.Sc. degree from the Tashkent University of Information Technologies in 2006 and the M.Sc. and Ph.D. degrees from Dongguk University in 2009 and 2012, respectively. He is currently an Associate Professor of computer science with Chungbuk National University. His research interests include databases, big data, data mining, and distributed processing system applications.

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.