Article

Mining λ-Maximal Cliques from a Fuzzy Graph

Fei Hao [1], Doo-Soon Park [1,*], Shuai Li [2] and Hwa Min Lee [1]

[1]   Department of Computer Software Engineering, Soonchunhyang University, 22 Soonchunhyang-ro, Shinchang-myeon, Asan 31538, Korea; fhao@sch.ac.kr (F.H.), leehm@sch.ac.kr (H.M.L.)
[2]   Department of Theoretical and Applied Sciences (DiSTA), University of Insubria, Via Mazzini 5, Varese 21100, Italy; shuaili.sli@gmail.com
*   Correspondence: parkds@sch.ac.kr; Tel.: +82-41-530-1317

Abstract: The depletion of natural resources in the last century now threatens our planet and the life of future generations. For the sake of sustainable development, this paper pioneers an interesting and practical problem of dense substructure (*i.e.*, maximal cliques) mining in a fuzzy graph where the edges are weighted by the degree of membership. For parameter $0 \le \lambda \le 1$ (also called fuzzy cut in fuzzy logic), a newly defined concept λ-maximal clique is introduced in a fuzzy graph. In order to detect the λ-maximal cliques from a fuzzy graph, an efficient mining algorithm based on Fuzzy Formal Concept Analysis (FFCA) is proposed. Extensive experimental evaluations are conducted for demonstrating the feasibility of the algorithm. In addition, a novel recommendation service based on an λ-maximal clique is provided for illustrating the sustainable usability of the problem addressed.

Keywords: sustainability; λ-maximal cliques; fuzzy graph; fuzzy concept analysis; degree of membership; fuzzy cut

1. Introduction

Future Sustainability Computing is the computational sustainable development that meets the needs and aspirations of the present without compromising the ability of future generations to meet their own needs. In particular, it utilizes the data mining and machine learning [1,2] (e.g., supervised and unsupervised learning), decision and optimization problems [3] (e.g., linear and integer programming, dynamic programming), sequential decision making under uncertainty (e.g., Markov decision processes and recommender systems) [4,5], and networks (e.g., fuzzy graphs and network algorithms) [6,7]. In this work, we design a novel approach on topological structure mining of social networks for future computational sustainability.

Background. Recent studies provide evidence of the importance of graph models used in many real world phenomena, such as social networks, biological networks and finance. A graph is a convenient way of representing information with relationships between objects [8]. The current complicated ubiquitous information enables the relationship between objects to become more vague. For example, (1) a social network can be established through sensoring data, and communications between individuals maybe missed or anonymized; (2) the relationships are vague in nature, such as one person influencing another in a social network; (3) and fuzzy trust relationships often occur in mobile social networks [9]. Therefore, it is natural to represent this kind of information using a Fuzzy Graph model [8].

The research on data mining based on a Fuzzy Graph model has attracted a lot of attention from the communities of data science and fuzzy logic. To emphasize the importance of λ-maximal clique mining in fuzzy graph, the following motivating example is provided.

Motivating Example. Considering the following scenario as a motivating example, a global research institute recently received funding for support of a big Information and Communications Technology (ICT) project which covers *Information retrieval (IR), Artificial Intelligence (AI), Data Mining (DM), and Computer Vision (CV)*. They will release this project to encourage social collaborative research among the scholars. Since the given project should be completed efficiently before a given deadline, it is expected to build a team of scientists who are good at the required research areas and have closer collaborative relationships.

Suppose Figure 1 shows a scholar social network including five scientists who come from different countries. However, they have some previous academic collaboration. The weights on the edges refer to the number of co-authored publications between two scientists. Intuitively, these weights describe the collaboration strength between them. Hence, this scholar social network can be viewed as a fuzzy graph with the weights indicating the membership values for representing the collaboration strength between scientists. Clearly, to complete the released ICT project, there are five possible research teams $\{Jack, John, Susan\}$, $\{Jack, Susan, Jessie\}$, $\{Jack, John, Jessie\}$, $\{Jack, Thomas, Jessie\}$, and $\{Jack, Jessie\}$ for taking the project.
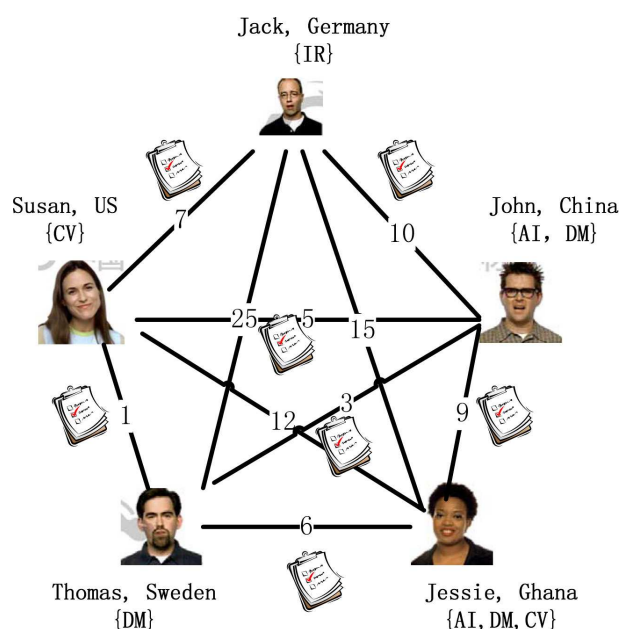

Figure 1. A motivating example.

From reliability and quality points of view, the research institute usually sets a parameter for controlling the quality, and this parameter corresponds to the λ in our problem. Overall, finding a team that can cover more countries and guarantee the product's quality is becoming a very interesting research issue and a practical problem.

Contributions. The major contributions of this paper are summarized as follows.

- A newly defined concept termed λ-maximal cliques is introduced. Furthermore, a novel problem of λ-maximal clique mining from a fuzzy graph is formalized;
- To address the proposed problem, an efficient fuzzy formal concept analysis based identification approach for finding the λ-maximal cliques is presented. The key idea is to prove the equivalence relation between λ-maximal clique and maximal fuzzy equiconcept;
- Extensive experiments are conducted with a real-world network dataset for evaluating the identification performance of the proposed approach. In addition, we present a promising recommendation service for illustrating the usability and extensibility of the proposed problem.

Roadmap. The rest of this paper is organized as follows. Section 2 mathematically describes the problem and targets of *λ-Maximal Clique Mining in a Fuzzy Graph*. To better represent and characterize the relationships between vertices of a fuzzy graph, the theory of fuzzy formal concept analysis is provided in Section 3. By incorporating the properties of fuzzy formal concept analysis, a novel solution for addressing the proposed problem based fuzzy formal concept analysis is presented in Section 4. Section 5 shows the experimental results with further discussions. An overview of related work is presented in Section 6. Section 7 concludes this paper.

2. Problem Definition

Any relation $R \subseteq S \times S$ on a set S can be characterized with a *graph* with node set S and edge set R. Similarly, any fuzzy relation $\mu : S \times S \to [0,1]$ can be viewed as a *fuzzy graph* in which each edge $e \in S \times S$ has a weight falling into $[0,1]$.

A fuzzy graph is a degree of membership over a deterministic graph. For simplicity, we consider the undirected graphs only since the edges can be regarded as unordered pairs of vertices. Mathematically, a fuzzy graph is defined as a 4-tuple $\mathcal{G} = (V, E, \sigma, \mu)$, where V is a set of vertices, $E \subseteq V \times V$ is a set of edges, σ is a fuzzy subset of a set V and $\mu : E \to [0,1]$ is a fuzzy relation on σ that assigns a degree of membership $\mu(e)$ to each edge $e \in E$.

First of all, let us recall two fundamental definitions in graphs about clique and maximal clique that are standard and apply not to fuzzy graphs but to deterministic graphs.

Definition 1. *[10] A clique in $G = (V, E)$ is a subset $S \subset V$ such that, for any two vertices $v_i, v_j \in S$, there exists an edge $(v_i, v_j) \in E$.*

Definition 2. *A set of vertices $M \subseteq V$ is a maximal clique in a graph $G = (V, E)$, if (1) M is a clique in G, and (2) there is no vertex $v \in V \setminus M$ such that $M \cup \{v\}$ is a clique in G.*

Definition 3. *In a fuzzy graph $\mathcal{G}$, for a set of vertices $C \subseteq V$, the clique's degree of membership of C, termed $cdm(C, \mathcal{G})$, is defined as the degree of membership in which, in a graph sampled from $\mathcal{G}$, C is a clique. For a given fuzzy cut λ, C is called a λ-clique if $cdm(C, \mathcal{G}) \geq \lambda$.*

For any set of vertices $C \subseteq V$, let E_C be the set of edges $\{e = \{u,v\}|e \in E, u, v \in C and u \neq v\}$, i.e., the set of edges connecting vertices in C.

Observation 1. *For any set of vertices $C \subseteq V$ in $\mathcal{G} = (V, E, \sigma, \mu)$, such that C is a clique in $G = (V, E)$, $cdm(C, \mathcal{G})=min\{e \in E_C|\mu(e)\}$.*

Proof. Let G be a graph sampled from $\mathcal{G}$. The set C will be a clique in G *iff* every edge in E_C is present in G. Since the event of selecting different edges is independent of each edge, the $cdm(C, \mathcal{G})$ is to catch the minimal degree of membership from the set of edges $e \in E_C$. Observation 1 holds. □

Similar to the definition of maximal clique, we define the λ-maximal clique in the fuzzy graph as follows.

Definition 4. *For a fuzzy graph $\mathcal{G} = (V, E, \sigma, \mu)$, and a fuzzy cut λ, a set $M \subseteq V$ is defined as a λ-maximal clique if*

- *M is a λ-clique in $\mathcal{G}$;*
- *There is no vertex $v \in (V \setminus M)$ such that $M \cup \{v\}$ is a λ-clique in $\mathcal{G}$.*

Problem. (*λ-Maximal Clique Mining*) *Given a fuzzy graph $\mathcal{G}$ and a fuzzy cut λ, the λ-Maximal Clique Mining problem is to find all vertex sets $M \in V$ such that M is a λ-maximal clique in $\mathcal{G}$.*

The extensibility of our proposed problem lies in offering the topological intelligence with the flexible controlling parameter λ. Furthermore, this type of intelligence can be applied to various sustainable applications, such as personalized recommendation systems, optimized team formation

with the quality parameter λ, and targeted marketing that guarantees the strength of reciprocal relationships greater than λ.

Based on Observation 1 and the problem statement, the following two important observations are easily derived.

Observation 2. *Suppose C is a λ-clique in $\mathcal{G}$. Then, for all edges $e \in E_C$, $\mu(e) \geq \lambda$ holds.*

Proof. Since a λ-clique satisfies a condition that the $min\{e \in E_C | \mu(e)\}$ should be greater than λ, it is straightforward to know that $\mu(e) \geq \lambda$. $\square$

Observation 3. *For any two sets of vertex A, B in $\mathcal{G}$, if $cdm(A, \mathcal{G}) \leq cdm(B, \mathcal{G})$, then we have $cdm(A \cup B, \mathcal{G})=cdm(A, \mathcal{G})$.*

Proof. Both $cdm(A, \mathcal{G})$ and $cdm(B, \mathcal{G})$ obtain the minimal degrees of membership, denoted with d_A, d_B, from edges $e \in E_A$ and $e \in E_B$, respectively. Therefore, $cdm(A \cup B, \mathcal{G})$ is equivalent to obtaining the minimal values from d_A, d_B, since $d_A \leq d_B$, $cdm(A \cup B, \mathcal{G})=cdm(A, \mathcal{G})$ holds. $\square$

In order to understand the above definitions and problem descriptions, an illustrative example is provided in this section as well.

Example 1. *Figure 2 presents an illustrative example for demonstrating how the proposed problem works. Clearly, Figure 2a is a fuzzy graph including seven vertices and degrees of membership on each edge. For example, the relationship between individuals C and D is a fuzzy relation with a degree of membership of 0.38. In particular, the given subgraphs g_1 and g_2 are two cliques since vertices are connected each other.*
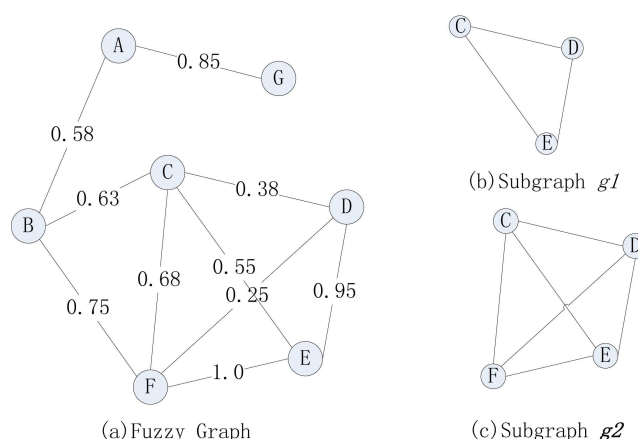

Figure 2. An illustrative example.

According to Definition 3, we assume the fuzzy cut λ is 0.2, then the subgraph g_1 is a λ-clique because $cdm(g_1, \mathcal{G})=min\{0.38, 0.55, 0.95\} = 0.38 > \lambda$. Here is a natural question. Are there any other cliques that satisfy the property of λ-clique in the fuzzy graph? Note that vertices in subgraph g_2 are more than the subgraph g_1 and $cdm(g_2, \mathcal{G}) = 0.25 > \lambda$. Therefore, the output of the λ-maximal cliques from this given fuzzy graph is subgraph g_2.

3. Fuzzy Formal Concept Analysis

This section presents the methodology about Fuzzy Formal Concept Analysis (FFCA) that we adopt in this paper. FFCA incorporates the fuzzy logic into Formal Concept Analysis (FCA) and handles the vague information represented by membership values [11]. Firstly, the definition of fuzzy formal context is formally provided. Then, the construction procedure of fuzzy concept lattice is presented.

Definition 5. *(Fuzzy Formal Context) A fuzzy formal context is formally represented as a triple $K = (O, A, R = \varphi(O \times A))$, where O is a set of objects, A is a set of attributes, and R is a fuzzy set on domain $O \times A$. Each relation $(o, a) \in R, o \in O, a \in A$ has a membership value $\mu(o, a)$ in $[0, 1]$. Therefore, $< o, a > \in I$ can be interpreted as: object o has attribute a with membership value $\mu(o, a)$ under the relation R.*

Remark: Note that a fuzzy formal context is represented as a cross-table with degrees of membership.

Example 2. *Table 1 shows a fuzzy formal context. The set of objects is composed of three documents D_1, D_2, D_3. This context has three attributes "Computer Vision (CV)", "Data Mining (DM)", "Big Data (BD)" representing three research topics. The relationship between an object and an attribute is represented by a membership value between 0 and 1.*

Table 1. A cross-table of a fuzzy formal context. **O**: Objectives; **A**: Attributes.

$O \backslash A$	CV	DM	BD
D_1	0.8	0.12	0.52
D_2	0.88	0.75	0.23
D_3	0.1	0.15	0.69

In real-world applications, a confidence threshold (also called fuzzy cut) λ is set for filtering out the lower membership values. Table 2 shows the cross-table of the refined fuzzy formal context given in Table 1 with $\lambda = 0.5$.

Table 2. The fuzzy formal context in Table 1 with $\lambda = 0.5$. **O**: Objectives; **A**: Attributes.

$O \backslash A$	CV	DM	BD
D_1	0.8	-	0.52
D_2	0.88	0.75	-
D_3	-	-	0.69

Definition 6. *(Fuzzy Concept) Suppose $K = (O, A, R)$ is a fuzzy formal context and λ is a confidence threshold. For $X \subseteq O$ and $Y \subseteq A$, we define the following operations:*

$$X^* = \{a \in A | \forall o \in X : \mu(o, a) \geq \lambda\}, \tag{1}$$

$$Y^* = \{o \in O | \forall a \in Y : \mu(o, a) \geq \lambda\}. \tag{2}$$

A fuzzy concept of a fuzzy formal context K with a fuzzy cut λ is a pair $(X_f = \varphi(X), Y)$, where $X \subseteq O$, $Y \subseteq A$, $X^* = Y$ and $Y^* = X$. Each object $o \in \varphi(X)$ has a membership μ_o defined as:

$$\mu_o = min_{a \in Y}\mu(o, a), \tag{3}$$

where $\mu(o, a)$ is the degree of membership between object o and attribute a that is defined in I. Particularly, if $Y = \{\}$, then $\mu_o = 1$ for every o.

Definition 7. *Let $C(K)$ be the set of all fuzzy concepts of the fuzzy formal context $K = (O, A, R)$. If $(X_1, Y_1), (X_2, Y_2) \in C(K)$, then let*

$$(\varphi(X_1), Y_1) \leq (\varphi(X_2), Y_2) \Leftrightarrow \varphi(X_1) \subseteq \varphi(X_2)(\Leftrightarrow Y_1 \supseteq Y_2), \tag{4}$$

where "$\leq$" is a partial relation of $C(K)$.

Definition 8. *(Fuzzy Concept Lattice) A concept lattice $L=(C(K), \leq)$ can be obtained by all fuzzy concepts $C(K)$ of a fuzzy formal context K with the partial order $\leq$. Its graphical representation is a Hasse diagram.*

Figure 3 illustrates the fuzzy concept lattice for the fuzzy formal context as shown in Table 1. Each circle denotes a fuzzy concept. The upper labels and lower labels of the circles represent intents and extents of the concepts, respectively. Besides, the information inside of the rectangles indicate the degrees of membership.
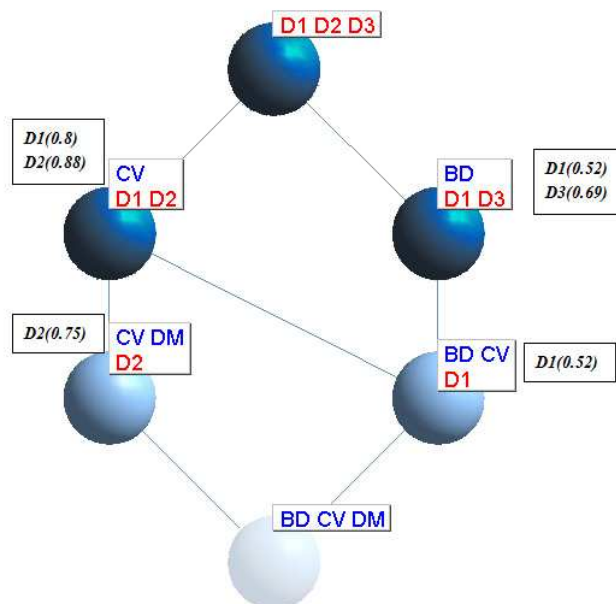

Figure 3. A fuzzy concept lattice of Table 1 with $\lambda = 0.5$.

4. λ-Maximal Clique Mining in Fuzzy Graphs

Aiming to discover all λ-maximal cliques from fuzzy graphs, this section firstly provides an overview of the proposed solution for addressing the problem. After that, the associated technical details contained in the solution are elaborated on separately.

4.1. An Overview of Solution

The novelty of the proposed solution is to mine all λ-maximal cliques by using the Fuzzy Formal Concept Analysis (FFCA) theory. Figure 4 shows the working principle of our proposed solution, which is composed of a fuzzy graph as the initial input, the set of all λ-maximal cliques as the final output, and three main technical modules: (1) constructing a fuzzy formal context for a given fuzzy graph according to the fuzzy matrix between vertex; (2) building the corresponding fuzzy concept lattice under the given fuzzy cut λ; and (3) exploring the equivalence relation between the newly defined concept, named *fuzzy equiconcept* with the maximal cardinality of extents/intents and λ-maximal cliques.

With these three technical modules, the following three subsections will separately discuss:

- how to construct a fuzzy formal context from a given fuzzy graph (see Section 4.2);
- how to extract the fuzzy concepts and build the corresponding fuzzy concept lattice (see Section 4.3);
- why there exists an equivalence relation between *fuzzy equiconcept* with the maximal cardinality of extents/intents and λ-maximal cliques (see Section 4.4).
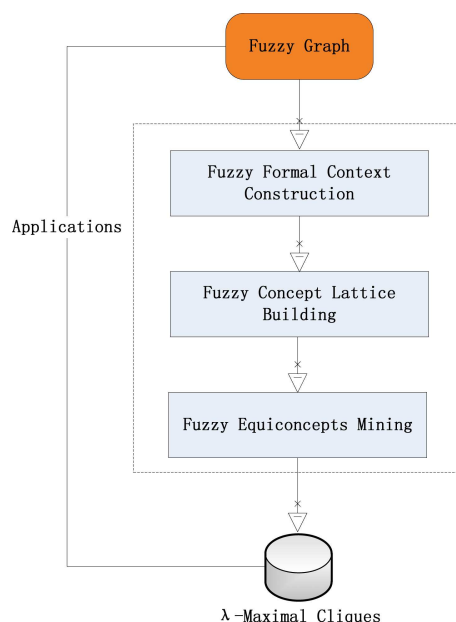
Figure 4. The framework of the proposed solution.

4.2. Fuzzy Formal Context Construction for a Fuzzy Graph

As mentioned in the problem definition, a fuzzy graph $\mathcal{G}$ can be modeled as a set of vertices, in which some of them have some relationships with others with the degrees of membership. To characterize this fuzzy relation I between vertices, we regard the vertices as both objects and attributes and construct a fuzzy formal context of the fuzzy graph $\mathcal{G}$ by using Fuzzy Matrix R^*, denoted as $FFC(\mathcal{G}) = (V, V, R^*)$.

Definition 9. *(Fuzzy Matrix) A fuzzy matrix $R^*=(r_{ij})_{m \times n}$ is an $m \times n$ matrix if:*

$$R^* = \begin{cases} r_{ij} = \mu(e_{ij}), & \text{if there exists an edge from } v_i \text{ to } v_j, \\ & \text{and } i \neq j, \\ r_{ij} = 1, & \text{if } i = j, \\ a_{ij} = 0, & \text{otherwise,} \end{cases} \tag{5}$$

where $r_{ij} = \mu(e_{ij})$, i.e., r_{ij} is the degree of membership between vertex v_i and v_j.

Hence, the $FFC(\mathcal{G})$ is equivalent to the Fuzzy Matrix of $\mathcal{G}$, *i.e.,* $FFC(\mathcal{G}) \equiv R^*$.

Property 1. *According to the properties of R^*, the fuzzy formal context $FFC(\mathcal{G})$ has the following important properties:*

- *The $FFC(\mathcal{G})$ is symmetric;*
- *All the diagonal elements are "1" in R^*.*

Here is an example for illustrating how to construct a fuzzy formal context for a given fuzzy graph.

Example 3. *Let us continue with Example 1. All the vertices appear $\{A, B, C, D, E, F, G\}$ in Figure 2a are viewed as objectives and attributes from the perspective of fuzzy formal context. Then, the weights (membership values) on the edges are the elements in the fuzzy formal context. Eventually, the fuzzy formal context of Figure 2a is constructed as follows,*

Table 3. Fuzzy formal context of Figure 2a. **O**: Objectives; **A**: Attributes.

O \ A	A	B	C	D	E	F	G
A	1	0.58	0	0	0	0	0.85
B	0.58	1	0.63	0	0	0.75	0
C	0	0.63	1	0.38	0.55	0.68	0
D	0	0	0.38	1	0.95	0.25	0
E	0	0	0.55	0.95	1	1.0	0
F	0	0.75	0.68	0.25	1.0	1	0
G	0.85	0	0	0	0	0	1

4.3. Fuzzy Concept Lattice Building

A fuzzy concept lattice is built upon a given fuzzy cut parameter λ that is used for removing the relations with the lower membership values. Therefore, before building the fuzzy concept lattice, we should refine the fuzzy formal context with the fuzzy cut λ in terms of different application scenarios. According to Definition 6, the fuzzy concepts can be easily obtained. Furthermore, the fuzzy concept lattice is built by Definition 8.

4.3.1. **Fuzzy Formal Context Reconstruction**

When $\lambda = 0.6$ in Example 3, then the membership values that are less than λ are filtered out from Table 3. Consequently, a refined fuzzy formal context is shown in Table 4.

Table 4. Fuzzy formal context of Figure 2a. **O**: Objectives; **A**: Attributes.

O \ A	A	B	C	D	E	F	G
A	1	0	0	0	0	0	0.85
B	0	1	0.63	0	0	0.75	0
C	0	0.63	1	0	0	0.68	0
D	0	0	0	1	0.95	0	0
E	0	0	0	0.95	1	1.0	0
F	0	0.75	0.68	0	1.0	1	0
G	0.85	0	0	0	0	0	1

4.3.2. **Fuzzy Concepts Extraction and Hasse Diagram Representation**

According to Definition 8, we can obtain a fuzzy concept lattice denoted as $L=(C(FFC(\mathcal{G})), \leq)$, as shown in Figure 5.

Clearly, we obtain 10 fuzzy concepts. For example, the nodes $(\{B, C, F\}, \{B, C, F\})$ and $(\{D, E, F\}, \{E\})$ in the Hasse diagram indicate the fuzzy concepts in Figure 5.

4.4. The Proposed Mining Approach of λ-Maximal Cliques

This section first defines two new concepts, *i.e., Fuzzy Equiconcept, k-Fuzzy Equiconcept, Maximal Fuzzy Equiconcept*, and then we present an efficient mining approach of λ-maximal cliques by finding the Fuzzy Equiconcepts that have the maximal number of extent/intent.

Definition 10. *(Fuzzy Equiconcept) For a fuzzy formal context $K = (O, A, R)$, if a pair (X, Y) satisfies $X^*=Y$, $Y^*=X$ and $X=Y$, then the pair (X, Y) is called a Fuzzy Equiconcept, where X and Y indicate the extent and intent of the fuzzy equiconcept, respectively.*
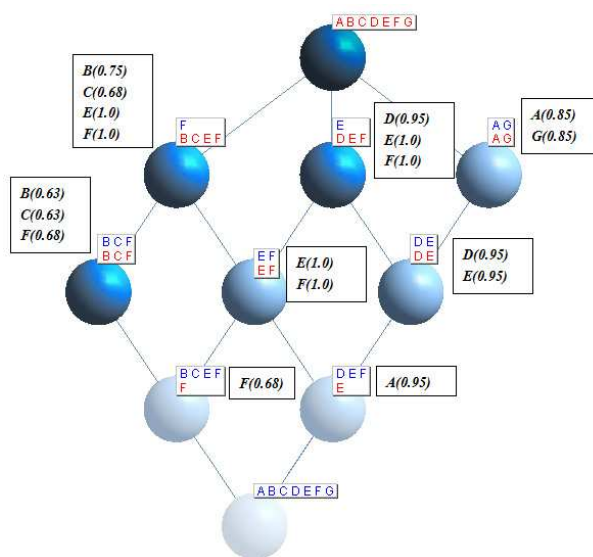
Definition 11. *(k-Fuzzy Equiconcept) For a fuzzy formal context $K = (O, A, R)$, if a pair (X, Y) satisfies $X^*=Y$, $Y^*=X$, $X=Y$ and $|X|=|Y|=k$, then the pair (X, Y) is called an $k-$Fuzzy Equiconcept, where X and Y indicate the extent and intent of the fuzzy equiconcept, respectively.*

Figure 5. Fuzzy concept lattice $L=(C(FFC(\mathcal{G})), \leq)$.

Definition 12. *(Maximal Fuzzy Equiconcept) For a fuzzy formal context $K = (O, A, R)$ of $\mathcal{G}$, if a pair (X, Y) satisfies $X^*=Y$, $Y^*=X$, $X=Y$, (X, Y) is defined as a Maximal Fuzzy Equiconcept if*

- *(X, Y) is a Fuzzy Equiconcept in $L=C(K, \leq)$;*
- *There is no vertex $v \in (V \setminus X)$ such that $X \cup \{v\}$ is a Fuzzy Equiconcept in $L=C(K, \leq)$.*

Based on Definitions 11 and 12, the following observation is easily obtained.

Observation 4. *Among all Fuzzy Equiconcepts (X_i, Y_i), $(i = 1, 2, \cdots, H)$ (H is the total number of Fuzzy Equiconcepts), the Maximal Fuzzy Equiconcept (X_{max}, Y_{max}) has the relationship with Fuzzy Equiconcepts:*

$$max = \arg\max_{i \in \mathcal{H}} |X_i|.$$

This observation studies the correlation between Maximal Fuzzy Equiconcept (X_{max}, Y_{max}) and Fuzzy Equiconcepts (X_i, Y_i) $(i = 1, 2, \cdots, H)$.

Proof. This proof is straightforward. That is to say, among all Fuzzy Equiconcepts, there exists at least one Fuzzy Equiconcept that has the maximal cardinality of extent or intent of the fuzzy concepts. □

Example 4. *We try to figure out these definitions and their correlations via this example. As can be seen from Figure 5, there exist four fuzzy equiconcepts, i.e., $(\{B, C, F\}, \{B, C, F\})$, $(\{E, F\}, \{E, F\})$, $(\{D, E\}, \{D, E\})$, and $(\{A, G\}, \{A, G\})$ since their extent is the same as intent in the fuzzy concepts. Intuitively, $(\{B, C, F\}, \{B, C, F\})$ is a three-fuzzy equiconcept, and $(\{E, F\}, \{E, F\})$, $(\{D, E\}, \{D, E\})$, and $(\{A, G\}, \{A, G\})$ are the two-fuzzy equiconcepts. Note that, among all fuzzy equiconcepts, the amount of extents of $(\{B, C, F\}, \{B, C, F\})$ is the maximum, thus $(\{B, C, F\}, \{B, C, F\})$ is regarded as the maximal fuzzy equiconcept.*

The following sections will investigate an equivalence relation between detection of λ-maximal cliques and maximal fuzzy equiconcepts. Based on this proposed equivalence relation, an efficient algorithm for mining the λ-maximal cliques from a fuzzy graph is presented.

Observation 5. *(Equivalence relation between λ-maximal clique and maximal fuzzy equiconcept) Suppose $P_{\lambda mc}$ is a mining problem of λ-maximal clique, P_{mfe} is an extraction problem of the maximal fuzzy equiconcept, and the following equivalence relation holds:*

$$P_{\lambda mc} \equiv P_{mfe}. \tag{6}$$

Proof. For a given fuzzy graph $\mathcal{G}$ with the fuzzy cut λ, the $P_{\lambda mc}$ is actually equivalent to finding the maximal cliques from the defuzzification graph after filtering out the edges on which the membership values are less than λ. From this point of view, $P_{\lambda mc}$ can be transformed into a fundamental problem "*maximal clique mining in graph*" [12,13] with the constraint that all membership values on the edges in this current graph should be greater than than λ. Clearly, the procedure of removing the edges from the fuzzy graph is the same as setting the elements in the fuzzy formal context as "0". Moreover, extracting the maximal fuzzy equiconcepts from the fuzzy concept lattice is the same as extracting the maximal equiconcepts (defined in [10]) from the graph with the fuzzy cut constraint. In this case, both new problems are degraded to the existing problems with the corresponding constraints.

Based on previous work [10], it is easy validating that $P_{\lambda mc} \Rightarrow P_{mfe}$ and $P_{\lambda mc} \Leftarrow P_{mfe}$ holds. Eventually, $P_{\lambda mc} \equiv P_{mfe}$. □

Inspired by the above equivalence relation between λ-maximal clique and maximal fuzzy equiconcept, a novel detection theorem of λ-maximal cliques based on fuzzy formal concept analysis is derived.

Theorem 1. *Given a fuzzy graph $\mathcal{G}$ and a fuzzy cut λ, the λ-maximal clique mining based on fuzzy formal concept analysis is to extract the maximal fuzzy equiconcepts.*

Proof. The proof of this theorem is straightforward. □

Lemma 1. *If $\lambda = 0$, the λ-maximal clique mining problem is degraded into a maximal clique mining in a fuzzy graph. The solution for addressing this problem is just to extract the equiconcepts (Note that the extent and intent of equiconcept is exactly same) which has the maximum number of extents/intents.*

4.5. Algorithm

We devise an algorithm based on the proposed mining theorem. Algorithm 1 depicts how does our solution works for mining the λ-maximal cliques from the fuzzy graph.

The working process of Algorithm 1 is described as follows: At the beginning, a fuzzy graph $\mathcal{G}$ and a fuzzy cut parameter λ are the inputs of the whole algorithm; Then, we initialize a set of λ-maximal cliques with Γ (Line 1). After the initialization of the algorithm, it goes into the fuzzy formal context construction and fuzzy concept generation codes section (Lines 26). Lines 7–13 return the index *max* of fuzzy equiconcepts that has the maximal cardinality of extents. After obtaining the index *max*, the algorithm outputs all λ-maximal cliques (Lines 14–18).

5. Empirical Study

We conduct experiments for various algorithms on a real-life scientist collaboration network. The goal of our experiments is to show that the proposed approach and algorithm can efficiently detect the λ-maximal cliques from fuzzy graphs. Additionally, the usability and extensibility are also illustrated via a concrete application on recommendation service.

5.1. Data Collection and Preprocessing

5.1.1. Data Source

The experimental dataset we adopted in this paper is a network of collaborations between scientists working on "*Networks*" [14,15] Figure 6 presents a visualization of the collaboration social network of scientists and a local zooming part of this network. Clearly, this collaboration social network contains 1589 scientists (vertices in the graph), 2742 collaborations (edges in the graph).

5.1.2. Data Preprocessing

This paper makes the fuzzification of the strength of collaborative relationship by a membership function, and a fuzzy graph is constructed.

Algorithm 1 Fuzzy Formal Concept Analysis Based λ-Maximal Clique Mining Algorithm

---

Input:

    A fuzzy graph: $\mathcal{G} = (V, E, \sigma, \mu)$;

    A fuzzy cut: λ;

Output:

    Set of λ-maximal cliques: Γ

  1: Initialize $\Gamma = \varnothing$

  2: *begin*

  3: Construct a fuzzy formal context $FFC(\mathcal{G})$ via fuzzy matrix

  4: Refine the fuzzy formal context $FFC(\mathcal{G})$ in the above step by filtering out the membership values which are less than λ

  5: Build a fuzzy concept lattice $L = (C(FFC(\mathcal{G})), \leq)$

  6: *end*

  7: *for* $i = 1$ **to** H **do**

  8: *begin*

  9:    *if* $((X_i, Y_i) \in C(FFC(\mathcal{G})))$ && $(X_i = Y_i)$

10:      *begin*

11:      $max = \arg\max_{i \in \mathcal{H}} |X_i|$

12:      *end*

13: *end*

14: *for* $i = 1$ **to** H **do**

15: *begin*

16:    *if* $(X_i, Y_i) \in C(FFC(\mathcal{G}))$ && $(X_i = Y_i)$ && $(i = max)$

17:      $\Gamma \leftarrow (X_i, Y_i)$

18: *end*

Definition 13. *(**Reciprocal Collaboration Variable**) In a scientist social network, let P_u be the set of publications held by user us. The reciprocal collaboration variable (RC) is defined as follows:*

$$RC(u, v) = |P_u \cap P_v|. \tag{7}$$

$RC(u, v)$ refers to the number of publications co-authored by scientists u and v.

To characterize the strength of the collaborative relationship between scientists *u* and *v*, we evaluate the degrees of membership according to the defined membership function, with $RC(u, v)$ as the input parameter. Generally, we set this membership function shaping like a semi-open trapezoid:

$$\mu_{Collaboration}(x) = \begin{cases} \frac{1}{a}x & x \in (0, a], \\ 1 & x > a. \end{cases} \tag{8}$$

According to domain knowledge, we set *a* = 3, *a* = 5, and *a* = 7 for instantiating the above membership functions as shown in Figure 7. For example, for two scientists *Bob* and *Alice*, if they co-authored for three publications, *i.e.*, $RC(Bob, Alice)$ = 3, then the degrees of membership on their collaborative relationship are 1, 0.6 and 0.43, respectively. According to the observation of the studied collaboration social network, the relationships can be characterized very well when *a* = 5.
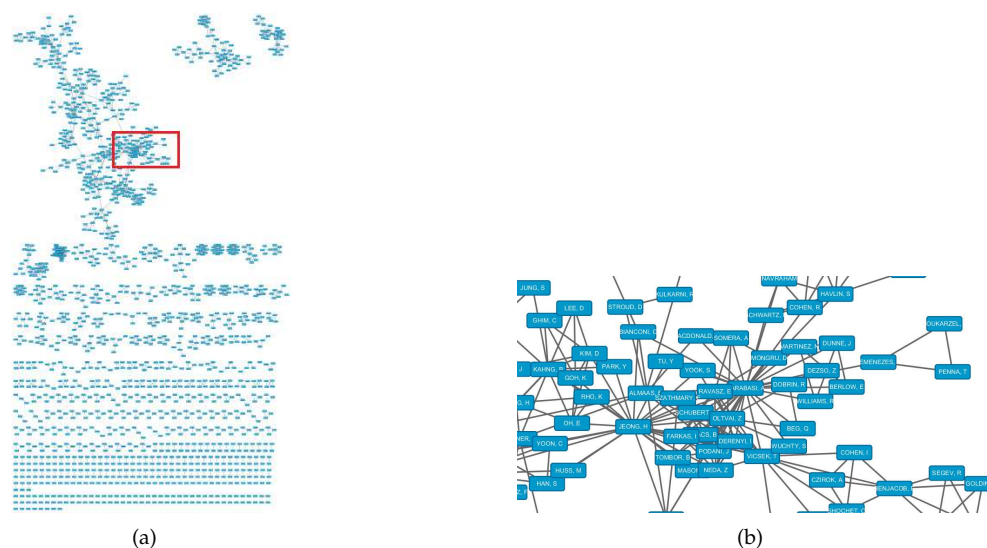
(a)　　　　　　　　　　　　　　　　　　　　(b)

Figure 6. Dataset visualization: (**a**) Visualization of Collaboration Network of Scientists; (**b**) Local Zooming Area (Red Rectangle in Figure 6a) of the Collaboration Network
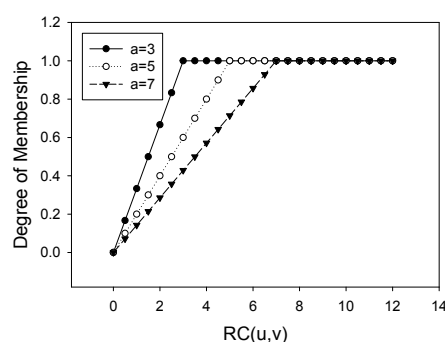

Figure 7. The membership function.

With the above data preprocessing approach, we can easily obtain a fuzzy graph that describes the strength of collaboration between two scientists. Figure 8 presents a distribution of the degree of membership on strength of relationships between scientists in the experimental dataset. Obviously, the degrees of membership on the strength of relationships fall into a range between 0.2 and 0.5. That is to say, most of the scientists co-authored 1–3 publications together.

5.2. Results and Discussions

To evaluate the performance of the proposed algorithm, we study the correlation between various λ and the number of vertices in detected λ-maximal cliques (shown in Table 5). Without loss of generality, the parameter λ is adjusted from 0 to 1 with the step length 0.2, *i.e.*, $\lambda =\{0, 0.2, 0.4, 0.6, 0.8, 1\}$.

Let us observe the topological structures of λ-maximal clique when the fuzzy cut $\lambda = 0.8$. As we know, three λ-maximal cliques can be detected in this case. Figure 9 shows the detected λ-maximal cliques with $\lambda = 0.8$, *i.e.*, $\{Hilgetag.C, Burns.G, Oneill.M, Young.M\}$, $\{Kashtan.N, Milo.R, Alon.U, Itzkovitz.S\}$, $\{Pastorsatorras.R, Vespignani.A, Moreno.Y, Vazquez.A\}$.
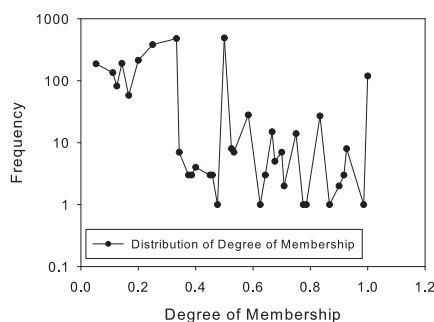
Figure 8. Distribution of degree of membership on strength of relationships.

Table 5. The correlation between various λ and the number of vertices in detected λ-maximal cliques.

λ	Number of vertices in λ-maximal clique	Number of λ-maximal cliques
0	20	1
0.2	6	20
0.4	6	1
0.6	5	1
0.8	4	3
1.0	4	1

Practically speaking, we can select the three collaborative research teams detected above including four scientists for completing the given project with the required quality assurance 80% (that corresponds to the λ in our problem).

5.3. Recommendation Services Based on λ-Maximal Cliques

In order to validate the usability and extensibility of the proposed problem, we now provide a concrete application in the setting of recommender systems. One solution suggested by [1] assumed that, in the recommendation systems for healthcare or medical treatments, they utilize the key concept from the traditional E-Commerce field, and they consider the geographical and temporal factors to build the recommender systems in the mobile social network. In another solution in [4], user similarity is reflected by the clustering network that the user formed, which proceeds in an online fashion. Then, they build the recommender system to satisfy the user's requirement on the fly. Finally, one recent piece of literature [2] proposed a spatial social union based model to construct the recommendation system via the available location database or information. However, our solution is completely different from above-mentioned methods. In fact, the physical location information might not be available, or user similarity might have a more general measure metric. Nevertheless, our strategy is not limited to the application scenario of mobile social networks only, which is more general and can be adapted to a single flexible component that can be easily integrated into the existing recommendation systems.

One illustration example is shown in Figure 10 below (for simplicity, the weight membership of edges in the user or item graph is omitted) at current timestamp when user 2 needs to serve, and the system might want to provide a recommendation for him/her.

First of all, based on previous or existing users' purchased history, we build the item's fuzzy graph, e.g., a user might buy a diaper and beer at the same time, then an edge whose weight reflects the closeness of two items is drawn between the diaper and beer, which naturally constructs the item graph that expresses the item's relationship of association. For a given fuzzy cut parameter or confidence threshold λ, we can easily mine the underlying λ-maximal cliques with the help of the machinery that we develop in this paper. For instance, for the current serving user 2, we follow the paradigm of Figure 4 to construct the item's λ-maximal cliques that are induced by (each) distinct

user. Then, the system can recommend the best clique with corresponding optimal λ, clearly the other cliques can be filtered out from the recommended pool of candidate items, e.g., in Figure 10 the derived optimal λ-maximal clique with respect to user 2 is consistent with items 4, 6 and 7. Then, we calculate the similarity regarding user 2 and the candidate items sitting inside the lambda maximal clique (in reality, the cardinality of this set is typically small), so that the recommended item can be smoothly pushed to that user by the ranking procedure of "Top-K score". In short, essentially we only need to maintain relatively small neighbor sets that are made up of different interrelated items with respect to the current serving user, which obviously can substantially reduce the computational complexity of running time or space requirement. In the meantime, as we can see in Figure 10, the prior timestamp the system served user 1 and the corresponding items fuzzy graph of clustering is the one denoted by green triangles in the item graph on the right-hand side, and a similar spirit also holds for the next timestamp when we serve user 3. However, it is worth pointing out that, even in the hardest environment, *i.e.*, a user that a system encounters in its history in which there are no previous records regarding him or her. In this case, we still can mine the user's λ-maximal cliques based on the user's social network composed by their friendship so that we can leverage the so-called "cold-start" problem in recommender systems as well. Therefore, when each user arrives, we make the fully personalized recommendation with the facility approach of our FFCA-based λ-maximal clique mining to efficiently discover the corresponding hidden structure from the fuzzy graph thereof. From the sustainable development point of view, this illustrative application incorporated the semantic descriptions of relationships between users and dynamically recommended the economical and suitable items to different users due to their personalized requirements that reduce both monetary cost and time cost of recommendation services providers (such as governments, companies and organizers). In another words, the computational resources are saved and recommendation efficiency has been improved.

Figure 9. Detected λ-maximal cliques with $\lambda = 0.8$.
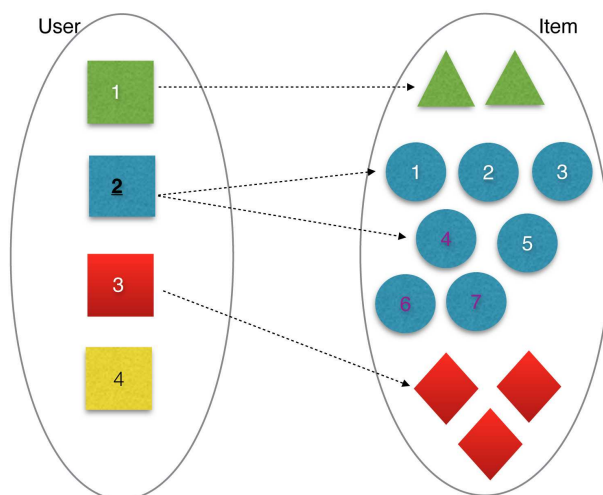

Figure 10. An illustrative example of recommendation service based on λ-maximal cliques.

6. Related Work

There has been a lot of recent work on exhibiting fuzzy or possibilistic clustering and communities detection using fuzzy logic. Reichardt [16] proposed a fast community detection algorithm based on *q*-state Potts model. Aiming to identify clusters from heterogeneous data and connect these clusters between the different node types, Blochl *et al.* [17] developed a fuzzy partitional clustering method based on a non-negative matrix factorization (NMF) model. Nepusz *et al.* [18] tackled the problem of fuzzy community detection in networks. Their approach allows each vertex to belong to multiple communities with different membership degrees and transforms the problem as a constrained optimization problem. Havens [19] recently presented a newly defined soft modularity function for detecting fuzzy communities in social networks. One previous work [20] is similar to our work, but it just presented some theoretical definitions without any investigation on topological structures mining in fuzzy graphs. Bandyopadhyay [21] firstly formalized the problem of finding maximum fuzzy cliques in fuzzy graphs. Then, the proposed problem was reduced to an unconstrained quadratic 0-1 programming problem. The maximum fuzzy clique in that paper is defined as a relaxed clique that can allow other vertices with higher membership degrees and merge them into cliques.

In addition, there exist some other works about clique mining from uncertain graphs. Zou *et al.* [22] present a problem of finding top-*k* cliques with the top-*k* highest probability of existence from an uncertain graph, which differs from [22], a recent research work focusing on mining the maximal cliques from an uncertain graph [23]. They also defined a new concept, named *α-maximal clique*, in an uncertain graph. Obviously, the weights on the edges of an uncertain graph are the probability of existence. However, the weights on the edges of a fuzzy graph refer to the degrees of membership. Inspired by those differences, this paper investigates a novel problem of mining *λ*-maximal cliques from fuzzy graphs.

7. Conclusions

Starting with the brand new concept of *λ*-maximal clique, this paper presents a theoretical study of the *λ*-maximal clique mining over fuzzy graphs from a computational sustainable point of view. To discover the *λ*-maximal cliques from a fuzzy graph, the fuzzy formal context methodology is employed. An equivalence relationship between the maximal fuzzy equiconcept and *λ*-maximal cliques is explored and proved. Based on this relationship, this paper also devises a corresponding algorithm. Experimental results indicate that our proposed algorithm achieves the *λ*-maximal cliques efficiently. In addition, a correlation between *λ* and the number of *λ*-maximal cliques is determined. This correlation reveals a fact that the size of *λ*-maximal clique decreases as the *λ* increases. Finally, a concrete use case application about recommendation service based on *λ*-maximal cliques is presented for showing the usability and extensibility of the addressed problem.

Author Contributions: Fei Hao proposed the topic and initial idea, performed experiments and wrote the paper; Doo-Soon Park improved the idea and designed the experiments; Shuai Li analyzed the data and wrote Section 5.3; HwaMin analyzed the experimental results and improved the presentation of the work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   Li, S.; Hao, F.; Li, M.; Kim, H. Medicine rating prediction and recommendation in mobile social networks. In Proceedings of the 8th International Conference on Grid and Pervasive Computing (GPC 2013), Seoul, Korea, 9–11 May 2013; pp. 216–223.

2. Hao, F.; Li, S.; Min, G.; Kim, H.C.; Yau, S.S.; Yang, L.T. An efficient approach to generating location-sensitive recommendations in ad-hoc social network environments. *IEEE Trans. Serv. Comput.* **2015**, *8*, 520–533.

3. Korda, N.; Szorenyi, B.; Li, S. Distributed clustering of linear bandits in peer to peer networks. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016.

4. Gentile, C.; Li, S.; Zappella, G. Online clustering of bandits. In Proceedings of the 31st International Conference on Machine Learning (ICML 2014), Beijing, China, 21–26 June 2014; pp. 757–765.

5. Jeong, W.H.; Kim, S.J.; Park, D.S.; Kwak, J. Performance improvement of a movie recommendation system based on personal propensity and secure collaborative filtering. *J. Inf. Process. Syst.* **2013**, *9*, 157–172.

6. Li, S.; Karatzoglou, A.; Gentile, C. Collaborative filtering bandits. In Proceedings of the 39th International ACM SIGIR Conference on Information Retrieval (SIGIR 2016), Pisa, Italy, 17–21 July 2016.

7. Guo, H.; Feng, Y.; Hao, F.; Zhong, S.; Li, S. Dynamic fuzzy logic control of genetic algorithm probabilities. *J. Comput.* **2014**, *9*, 22–27.

8. Sunitha, M.S.; Mathew, S. Fuzzy graph theory: A survey. *Ann. Pure Appl. Math.* **2013**, *4*, 92–110.

9. Hao, F.; Min, G.; Lin, M.; Luo, C.; Yang, L.T. MobiFuzzyTrust: An efficient fuzzy trust inference mechanism in mobile social networks. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *25*, 2944–2955.

10. Hao, F.; Min, G.; Pei, Z.; Park, D.S.; Yang, L.T. K-clique communities detection in social networks based on formal concept analysis. *IEEE Syst. J.* **2015**, *99*, 1–10, doi:10.1109/JSYST.2015.2433294.

11. Quan, T.T.; Hui, S.C.; Cao, T.H. A fuzzy FCA-based approach to conceptual clustering for automatic generation of concept hierarchy on uncertainty data. In Proceedings of the CLA 2004 International Workshop on Concept Lattices and their Applications, Ostrava, Czech Republic, 23–24 September 2004; pp. 1–12.

12. Ostergard, P.R. A fast algorithm for the maximum clique problem. *Discret. Appl. Math.* **2002**, *120*, 197–207.

13. Cheng, J.; Ke, Y.; Fu, A.W.C.; Yu, J.X.; Zhu, L. Finding maximal cliques in massive networks. *ACM Trans. Database Syst. (TODS)* **2011**, *36*, 21, doi: 10.1145/2043652.2043654.

14. Newman, M.E. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **2006**, *74*, 036104, doi: 10.1103/PhysRevE.74.036104.

15. Newman, M. Network data. Available online: http://www-personal.umich.edu/~mejn/netdata/ (accessed on 8 June 2016).

16. Reichardt, J.; Bornholdt, S. Detecting fuzzy community structures in complex networks with a Potts model. *Phys. Rev. Lett.* **2004**, *93*, 218701, doi: 10.1103/PhysRevLett.93.218701.

17. Blochl, F.; Hartsperger, M.L.; Stumpflen, V.; Theis, F.J. Uncovering the structure of heterogenous biological data: Fuzzy graph partitioning in the k-partite setting. In Proceedings of the German Conference on Bioinformatics 2010, Bielefeld, Germany, 20–22 September 2010; pp. 31–40.

18. Nepusz, T.; Petroczi, A.; Negyessy, L.; Bazso, F. Fuzzy communities and the concept of bridgeness in complex networks. *Phys. Rev. E* **2008**, *77*, 016107, doi: 10.1103/PhysRevE.77.016107.

19. Havens, T.C.; Bezdek, J.C.; Leckie, C.; Ramamohanarao, K.; Palaniswami, M. A soft modularity function for detecting fuzzy communities in social networks. *IEEE Trans. Fuzzy Syst.* **2013**, *21*, 1170–1175.

20. Nair, P.S.; Cheng, S.C. Cliques and fuzzy cliques in fuzzy graphs. In Proceedings of IFSA World Congress and 20th NAFIPS International Conference, Vancouver, BC, Canada, 25–28 July 2001; pp. 2277–2280.

21. Bandyopadhyay, S.; Bhattacharyya, M. A neuro-GA approach for the maximum fuzzy clique problem. In *Advances in Neuro-Information Processing*; Springer Berlin Heidelberg: Auckland, New Zealand, 2009; pp. 605–612.

22. Zou, Z.; Li, J.; Gao, H.; Zhang, S. Finding top-k maximal cliques in an uncertain graph. In Proceedings of the 26th IEEE International Conference on Data Engineering (ICDE 2010), Long Beach, CA, USA, 1–6 March 2010; pp. 649–652.

23. Mukherjee, A.P.; Xu, P.; Tirthapura, S. Mining maximal cliques from an uncertain graph. In Proceedings of the 31st IEEE International Conference on Data Engineering (ICDE 2015), Seoul, Korea, 13–16 April 2015; pp. 243–254.