Formal Concept Analysis applied in Knowledge Network: A survey

Xv Zhang[1], Weizhuo Li[12]*
School of Modern Posts, Nanjing University of Posts and Telecommunications, China.
State Key Laboratory for Novel Software Technology, Nanjing University, P.R. China
liweizhuo@amss.ac.cn

Abstract. With the rapid iterative development of artificial intelligence (AI) technologies, knowledge network is one of the core ways to express massive information, and the requirements of semantic analysis and knowledge mining of it are becoming more and more important. As a mathematical tool based on lattice theory, Formal Concept Analysis (FCA) has gradually attracted the attention of AI researchers in recent years because of its advantages of convenient modeling, strong interpretability and deep network node characteristics mining. As the summary of FCA's research works on knowledge network is lagging, we present a survey that first focuses on the research works of FCA for knowledge networks. We systematically summarize the existing methods from knowledge graph, social networks, semantic representations, machine learning and data mining. Finally, we give an outlook on the future directions of FCA applied in the knowledge network.

Keywords: Formal Concept Analysis; Knowledge Network; Knowledge Graph; Social Network; Semantic Representation.

1 Introduction

With the rapid development of artificial intelligence (AI) technologies such as Large Language Models, Deep Learning and Knowledge Graph, Knowledge Network, as one of the core ways to express massive information and knowledge, its expression mode and analysis technology become more and more important. Nevertheless, how to represent the complex semantic relations of knowledge network efficiently and mine its deep knowledge association efficiently, has become the common concern of academia and industry. Formal Concept Analysis (FCA), as a mathematical tool based on lattice theory, has gradually attracted the attention of researchers of knowledge representation and reasoning in recent years because of its unique advantages in modeling convenience, strong interpretability and deep network node characteristics mining [1-3].

FCA constructs concept lattice (CL) through formal context (FC), and then presents the implicit relationship in data in an intuitive and structured way. For knowledge graph completion, by analyzing the entities and their relationships in the knowledge graph, FCA can discover new semantic associations, supplement the lacking triples of the knowledge graph and improve its ability to reason [4,5]. In the field of social network analysis, FCA has shown its unique value applied in community detection and relationship prediction. It regards users and their interactions in social networks as a formal context. Then, it identifies users with similar interests or behavior patterns. Finally, it could predict potential social relationships and provide strong support for personalized recommendation and information dissemination analysis [6,7]. In the task of semantic representation, FCA transforms the complex semantic information into a structured concept lattice, which can enable the search engines to understand and process natural language texts more effectively, so that the performances of natural language processing tasks could be improved, such as information retrieval and text classification. In addition, it can help researchers and data analysts to discover hidden patterns and features in data [8-10]. Recently, the integration of FCA and Machine Learning technology provides a new idea that builds an intelligent system with strong interpretability and expressivity [11].

Combined with symbols reasoning and statistics reasoning has always become one of the most important research topics in AI. Although FCA has made several remarkable progresses in both theoretical research and practical application, the summary of research works about FCA applied in Knowledge Networks [12-14] is lagging. It is difficult to reflect on the latest developments and applications of FCA in knowledge networks. To fill this research gap, we try to collect the FCA's research works applied in the knowledge network in the past 20 years and conduct a comprehensive survey of them. Precisely, we systematically summarize the existing methods from knowledge graph, social networks, semantic representations, machine learning and data mining, and provide an outlook on future directions of FCA applied in knowledge network.

The rest of this paper is organized as follows. We present background knowledge in Section 2. We summarize existing works of FCA applied in knowledge network and give a comparison of them in Section 3. Section 4 shows the limitations of FCA and its Optimized works, followed by the conclusions and future direction in Section 5.

2 Preliminary

In this section, we first introduce basic notions of FCA, including formal context, formal concept and so on. Then, we display an example of FCA and related tool. Finally, we present Knowledge Network and its different representations.

2.1 Formal Concept Analysis

Formal Concept Analysis theory is mainly based on the groundbreaking work published by Ganter and Wille. FCA takes an input table, specifies a group of objects and their attributes, and finds the natural clusters of all attributes and all objects in the input data. Among them, the input table is called Formal Context, and the natural object cluster is a collection of all objects sharing a subset of common attributes, and also a collection of all attributes shared by one of the natural object clusters. At the same time, natural attribute clustering corresponds to natural object clustering one by one. Such a pair of natural object clusters and natural attribute clusters form a concept, which is considered the formal unit of human thinking [15]. It is based on a philosophical understanding that a concept consists of two parts: its extension belongs to all the objects of the concept, and its connotation consists of all the attributes shared by these objects. These concepts show an inductive hierarchy when organized into a lattice. A set of algorithms based on tools such as meaning and exploration are used to construct such a Formal Concept Lattice, which is helpful to further infer from the ideas clearly expressed as formal concepts.

The Formal Context [16] is given by a triple (G, M, I) and is denoted as K. This tuple consists of two sets G and M, as well as the relationship I between G and M. The elements of G are called the sets of objects, the elements of M are called the sets of attributes of the objects. I is defined as a binary relation $I \subseteq G \times M$, where (g, m) in I if and only if object g has attribute m. The tuples (A_i, B_i) are derived from the formal context (G, M, I), where A is the set of objects and B is the set of attributes. For each object in A that is not in A_i, there is an attribute in B_i that this object does not have. For each attribute in B that is not in B_i, there is an object in A_i that does not have this attribute.

For example, in the following text, we take four animals, Lion, Bee, Eagle, and Clownfish, as objects and take certain characteristics they possess as attributes. If the animal has the corresponding characteristics, then we associate the animal object with the attribute of that characteristic. The data in Table 1 is an example of a formal context, where the behavior objects are listed as attributes. Among them, the cell "×" indicates that the object has this attribute. For example, the animal Lion has the characteristic of "Gregarious".

Table 1. An example of a formal context

	Terrestrial	Has the Tail	Gregarious	Eats nectar	Can fly	Has 6 legs
Lion	×	×	×			
Bee	×		×	×	×	×
Eagle	×	×			×	
Clown-fish		×	×			

Formal concepts in FCA are defined as pairs (A, B), where $A \subseteq G$, $B \subseteq M$, and:

$$B = \{ m \in M \mid \forall g \in A: (g, m) \in I \}$$
$$A = \{ g \in G \mid \forall m \in B: (g, m) \in I \} \tag{1}$$

According to Formula (1), the formal concept satisfies the following conditions: A is the set of all objects with the attribute set B; B is the set of all common attributes of the object set A [16,17].

In the formal context, the relationship between objects and attributes is defined through the binary relation I. Specifically, for an object $g \in G$ and its property $m \in M$. If $(g, m) \in I$, it indicates that the object g has the property m. This kind of relationship can be deterministic (such as 1 and 0 in a binary matrix), or fuzzy (such as membership degree in a fuzzy relationship).

The Galois connection is a special correspondence between two partially ordered sets, that is, every connection of Galois languages will produce an isomorphic relationship. Galois duality is the Galois connection between two mutually dual partially ordered sets [18]. It is defined as follows:

Let $(A, \leq)$ and $(B, \leq)$ be two partially ordered sets, and the Galois connection between them is composed of two monotonic functions $F: A \rightarrow B$ and $G: B \rightarrow A$. Therefore, for all $a \in A$ and $b \in B$, we have:

$$F(a) \leq b \;\; if \;\; a \leq G(b), \tag{2}$$

When defining a concept as a pair (A, B), where A is the extension and B is the connotation, FCA considers a Galois duality. By minimizing one of A and B, we can always maximize the other, and vice versa.

The concept lattice is a lattice structure composed of all formal concepts according to the containment relationship, represented as $\boldsymbol{\mathcal{B}}(G, M, I)$. In the concept lattice, there are two fundamental relationships between concepts [19]:

①Order Relation [16]: (A_1, B_1) and (A_2, B_2) are formal concepts in the Formal Concept Lattice $\boldsymbol{\mathcal{B}}(G, M, I)$. The subconcept relationship represented by $\leq$ is as shown in Formula (3) :

$$\big((A_1, B_1), \leq, (A_2, B_2)\big) \;\leftrightarrow\; A_1 \subseteq A_2, \leftrightarrow B_2 \subseteq B_1, \tag{3}$$

②Join and Meet: The Join of two concepts (A_1, B_1) and (A_2, B_2) is their least upper bound, denoted as $(A_1, B_1) \vee (A_2, B_2)$; Their Meet is their maximum lower bound, denoted as $(A_1, B_1) \wedge (A_2, B_2)$.

The Galois connections between concepts can be visualized in the form of concept lattices. The formal concept cells in Table 1 can be visualized using the Conexp-clj tool. Each individual node represents the formal concept in the concept cell, as shown in Figure 1.

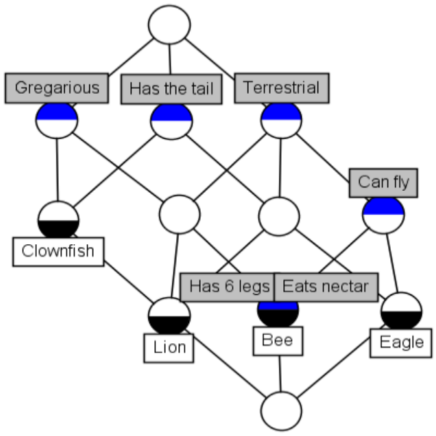

Fig. 1. The visualization of the formal concept grid in Table 1

Conexp-clj is an FCA tool based on the Clojure language, aiming to provide efficient computing performance and flexible scalability. Its core algorithm is based on the construction and operation of concept lattices and utilizes the functional programming characteristics of Clojure to achieve efficient processing of large-scale datasets. Conexp-clj supports multiple FCA algorithms, including the Close-by-One(CbO) algorithm and its variants [20]. These algorithms can enumerate the closure set in a polynomial delay manner, ensuring the efficiency and accuracy of the calculation process. The development and application of Conexp-clj have been verified in multiple studies. For instance, Hanika [21] et al. described in detail the implementation and performance evaluation of Conexp-clj in their research, demonstrating its efficiency and reliability when dealing with large-scale datasets.

The process of using Conexp-clj is as follows: First, convert the input data into a formal context, that is, a triple (G, M, I), where $I \subseteq G \times M$ is the binary relation between the object and the property. Then, by using the CbO algorithm, the concept lattice is gradually constructed from the formal context. The CbO algorithm gradually expands the closure set to ensure that the generation of each concept is based on the existing closures, thereby avoiding duplicate calculations. Conexp-clj supports various operations on concept lattices, including concept merging, splitting, filtering, etc., for further analysis and visualization.

Conexp-clj performs well when dealing with large-scale datasets, especially in the scenario of large-scale datasets, where it can efficiently handle datasets containing tens of thousands of objects and attributes. Meanwhile, it is applicable to processing data with multi-level structures and complex relationships, and supports incremental updates, capable of quickly updating the concept lattice when the data changes [20].

However, since this tool often processes by enumerating closure sets, in the worst case, the time complexity is $O(2^M)$ (where M is the attribute set), making it difficult to directly handle large-scale data. Moreover, recursive calls and intermediate closure storage may lead to memory bottlenecks, especially when there are a large number of attributes.

2.2 Knowledge Network

As one of the core ways to express massive information and knowledge, knowledge network has become more and more important in semantic analysis and knowledge mining. how to achieve above tasks has become the common concerns of academia and industry. In this section, we divide the knowledge network into three forms (i.e., directed graph, undirected graph and hypergraph), and exemplify them through knowledge graph, social network and complex events, respectively.

The directed graph is defined as $G = (V, E)$, where V is a set of nodes, which is a non-empty finite set, and E is an edge set, which is a directed edge. Each directed edge is composed of ordered the vertex pairs in vertex set V. The main representative of a directed graph is the knowledge graph (KG). As a semantic network based on graph structure, knowledge graph is used to represent entities and their semantic relationships, aiming at realizing the structural modeling of real-world knowledge. Among them, nodes represent objects in the real world, and have unique identities and attributes; The edge represents the semantic association between entities, which has directionality and types, such as attribute relationship and hierarchical relationship [22]. The basic unit of knowledge graph, is shaped like a triple of $(Entity\ 1, Relationship, Entity\ 2)$ or $(Entity, Attribute, Value)$, such as $(Einstein, Birthplace, Germany)$. Through the nodes and edges of the directed graph, the hierarchical structure and complex association of knowledge are expressed intuitively, which supports logical reasoning and semantic query.

The undirected graph is defined as $G = (V, E)$, where V is a node set and a non-empty finite set, and E is an edge set, in which each edge is composed of unordered vertex pairs in the vertex set V, and there is no direction. The main case of an undirected graph is the social network, whose graph structure is composed of nodes and edges, which is used to depict the relationship between social entities (such as people, organizations and groups) and emphasize the interaction between individuals, information dissemination and social structure. A Social network is usually represented by an undirected graph, which is suitable for the symmetrical relationship between entities and can be weighted or labeled. Nodes represent social entities, such as individuals, enterprises and communities. Social networks can analyze social behaviors, such as influence dissemination and community discovery, through node centrality and clustering. Common applications are reflected in social platform analysis, information dissemination, recommendation systems and public opinion monitoring [23]. It should be noted that the above Formal Concept Lattice is a special undirected graph, and the nodes of extent and intent have a partial order relationship in the lattice.

The Hypergraph is a triple $H = (V, E, L)$, where V is the node set, E is the edge set, and L is the label set of the edge. Each edge $e \in E$ is a subset of a node, representing a multivariate relationship [24]. Hypergraph can be expressed as a Formal Context, in which nodes are objects, edges are attributes, and labels are extensions of attributes. The application of hypergraph embodies the modeling of complex events: for example, in the "earthquake-rescue" event, there are many entity associations such as earthquake intensity, disaster-stricken area, rescue team, and material allocation, among which Vertex represents entities (such as concepts, objects, events, etc.) in the knowledge system and can carry attributes (such as name, type, description, etc.); Hyperedge: connecting multiple nodes, indicating the pluralistic relationship between nodes (such as "meeting" involving multiple participants and "teamwork" involving multi-person collaboration). Super edge can contain attributes (such as relationship type, timestamp, weight, etc.); Hierarchical structure: Hypergraphs can be nested with super edges to form a hierarchical knowledge structure (for example, a "project" consists of multiple "task" super edges, and each task contains subtask nodes). Generally speaking, the super edge can integrate the dynamic relationship of these event nodes, and can also be used to model the multivariate relationship. For the storage of this kind of hypergraph data information, NoSQL database is usually selected (such as a document database, key value database, graph database and column family database), and the specific selection should be based on the structural characteristics of the hypergraph.

3 The works of FCA applied in Knowledge Network

3.1 Systematic Literature Survey

We collected a large number of documents by using the academic literature website, and when we included them in the literature reference, we gave priority to the documents published in recent years, while excluding those from older years. We made an intensive reading analysis on the papers that put forward novel methods, integrated their methods, and summarized their applicability, advantages and disadvantages. Finally, we have collected 30 works related to FCA published in conferences and periodicals in the past ten years, subdivided them according to the research fields and task scenarios of FCA in the knowledge network, and summarized the maturity of each task with scenario solutions in different research fields, as shown in Table 2.

According to the evaluation results shown in the table, under the current research background, the application of FCA in Knowledge Graph, Social Networks, Semantic Representations, Combined with Machine Learning (ML) & Date Mining (DM) and other research fields has attracted attention. It also reflects the research enthusiasm in the task scenarios such as information retrieval and association, data classification and prediction, theoretical (complexity) analysis, community detection and Data fusion.

Among them, in the task scenario of information retrieval and association, Knowledge Base Representation, Knowledge Extraction, Knowledge Evolution, Association Rule Mining of Semantic Representations, Classification and Regression model of combined with ML& DM, Data Visualization, Granular concept analysis all show high maturity, but the research on Community Analysis, Social Internet of Things and Semantic Representations of Social Networks needs to be further explored.

In the data classification and prediction task, the association rule mining of Semantic Representations, the research on classification and regression model, Data Visualization combined with ML& DM is very mature. However, the research on Knowledge Representation of Knowledge Graph, Knowledge extraction, Social Internet of Things of Social Networks and Semantic Representations needs to be further explored.

Under the task scenario of theoretical (complexity) analysis, Knowledge Graph representation, Knowledge Evolution, Association Rule Mining of Semantic Representations, Classification and Regression model of Combined with ML& DM and Granular concept analysis are highly mature.

In the Community detection task, the research on Social Networks' Community Analysis, Social Internet of Things and Granular concept analysis is very mature. However, the Community Analysis of Social Networks, Association Rule Mining of Semantic Representations, classification and regression model Combined with ML& DM need to be further studied.

In the Data fusion task, the research on the base representation of the Knowledge Graph is very mature, while the research on Knowledge extraction of the Knowledge Graph, Knowledge Social Internet of Things of Social Networks needs to be further deepened. In other fields, there are many gaps in the corresponding task scenarios, which may be a new research trend in the near future and still need to be paid more attention in the future.

Table 2. Maturity of FCA's solution for each task scenario in different research fields: ● indicates that the research is relatively mature; ▲ indicates that there exist the room for improvement; ○ means that further research is needed to bridge the gap.

Research Field		Task scenario									
		Information retrieval and association		Data classification and prediction		Theoretical(complexity) analysis		Community detection		Data fusion	
		Maturity	References	Maturity	References	Maturity	References	Maturity	References	Maturity	References
Knowledge Graph	Knowledge Base Representation	●	[4][5][27] [42][41]	▲	[4]	●	[24][41]	○		●	[24][27] [42]
	Knowledge Extraction	●	[19][26]	▲	[19]	○		○		▲	[26]
	Knowledge Evolution	●	[29][43]	○		●	[29] [43]	○		○	
Social Networks	Community Analysis	▲	[6]	○		○		▲	[6]	○	
	Social Internet of Things	▲	[7]	▲	[28]	○		●	[7][28]	▲	[7]
Semantic Representations	Knowledge Representation	▲	[21]	▲	[21]	○		○		○	
	Association Rule Mining	●	[10] [18][25] [29][40]	●	[10] [23][29]	●	[10] [23][29]	▲	[23]	○	
Combined with ML& DM	Classification and Regression Model	●	[32][33]	●	[17][31] [32][34]	●	[17][33]	▲	[32]	○	
	Data Visualization	●	[35][36]	●	[35][36]	○		○		○	
Granular concept analysis		●	[14][16] [22]	●	[14][16] [22]	●	[14][16]	●	[14]	○	

3.2 FCA applied in Knowledge Graph

In the field of Knowledge Graph, with the open-source knowledge base of Wiki data, Hanika et al. [25] put forward an effective method to identify the comprehensible meaning implied in the data, and overcome the complexity problem that Wiki data cannot be directly modeled by extracting the context representation of Wiki data in a systematic way. Knowledge Graph describes entities through a large number of Resource Description Framework data (subject-predicate-object triplets), while scholars such as Yang et al. [26] have defined the concept lattice of RDF graphs. Since RDF data is essentially a triple graph, its connotation can be described by a graphic pattern. The graphic pattern closure is defined as the connected component of the graph through the simple form of graph realization. These graphic patterns correspond to join queries and can support the generalization of attributes, thus formalizing the concept map.

RDF data exists in independent and scattered resources, and needs to be centralized, navigated and searched to meet the application requirements of specific fields. Mehwis et al. [27] put forward a data integration method of FCA to create a navigation space on the semantic web data. This method extends FCA and provides centralized access to data generated by multiple resources by using RDF triples and RDF schemas from multiple independent sources. Then, SPARQL query can be put forward in this navigation space, so as to access these distributed resources from a platform and realize the purpose of information retrieval.

Ferré and Peggy Cellier [23] and others described the algorithm and performance evaluation of Graph-FCA in detail in their research, which proved its effectiveness in processing hypergraph data. A Pattern usually refers to a specific substructure or topological feature that appears repeatedly in a graph. Pattern basis is a set of basic patterns used to generate all concepts in a hypergraph. By using the generation algorithm, we extract the key patterns in the hypergraph and reduce the computational complexity [5]. After the concept lattice is generated by using the pattern base, the semantic analysis and Knowledge Representation in the data can be found by showing the conceptual relationship in the data through the hierarchical structure. Finally, the complexity of the concept lattice is reduced by the optimization algorithm, and the visualization and analysis efficiency are improved.

3.3 FCA applied in Social Networks

The research progress of FCA technology in the identification and abstraction of core data structures in Social Networks has maintained a certain room for improvement. At present, scholars such as Hao et al. [6] have innovatively applied the FCA method to Social Network key structure detection, dynamic evolution management and entity abstract generation of Knowledge Graph in many papers, which fully demonstrated the powerful ability and unique advantages of FCA in dealing with complex network data.

In the identification of key structures of Social Networks, Hao et al. [28] put forward a unified framework method based on FCA. By constructing the Formal Context of Social Networks and generating concept lattices, the association between concepts and network structures is quantified by using concept interesting measures (such as stability and separability), and the joint detection of key structures such as maximum clique, bridge structure and structural hole is realized. This method effectively solves the problem of the high computational cost of traditional independent detection tasks, and verifies its efficiency and accuracy on several real network data sets. In the Social Internet of Things (SIoT) scenario, the research team further explored the problem of detecting diversified top-k maximum cliques. By proving the equivalence relationship between the concepts of maximum clique and equality, an intelligent greedy algorithm based on a stack data structure was designed, which significantly improved the maximum clique coverage under the premise of ensuring the diversity of results, showing the feasibility and superiority of the FCA method in dealing with large-scale complex networks.

Aiming at the challenge of dynamic changes in Social Networks, Yang et al. [7] proposed a dynamic maximum clique detection and evolution management method. By developing Add-FCA and Dec-FCA algorithms that are suitable for users to join and leave the scene, the efficient response to the dynamic changes of the Formal Context of Social Networks is realized, and by using the equivalence theorem between the concepts of equality and the maximum clique, the evolution modes of the maximum clique, such as invariance, change, addition and disappearance, are captured. The experimental results reveal that the maximum number of clusters decreases with the increase of parameter k, and verify the effectiveness of the algorithm in quantitative relationship analysis. Also, the increasing RDF description of entities leads to information overload.

In this context, automatic entity summarization has received extensive attention in recent years. Selecting the most concise and typical facts from the lengthy RDF data and briefly describing an entity has become the main goal. Yang et al. [26] and others proposed an incremental entity summarization method IES-FCA based on FCA. This method constructs concept lattice incrementally that considers the importance, redundancy and uniqueness of triples in combination with the improved ranking algorithm. Compared with the non-incremental methods, this method can reduce the time consumption, better than the existing algorithms in F1 measurement, MAP and NDCG.

3.4 FCA applied in Semantic Representations

In order to break the limitations of data processing scale, Knowledge Representation and cross-domain applications, a new method is provided for knowledge discovery and semantic understanding of complex data. In addition to the current mature research fields, FCA has also carried out further research in the field of Representation Learning.

The FCA2VEC method proposed by Dürrschnabel et al. [21] aims to combine FCA with word embedding technology to solve the problem of efficient processing and interpretability of large-scale data sets. This method initializes semantic information by pre-training word vectors, and optimizes entity vectors by FCA embedding technology, thus preserving conceptual structure in a low-dimensional space. The core of the research is to realize the calculation of the covering relation and canonical basis of the concept lattice through the nonlinear embedding (Closure2Vec) and object/attribute vectorization (Object2Vec and Attribute2Vec) of the closure operator. Experiments show that this method is significantly superior to traditional methods in link prediction and attribute clustering tasks, especially in 3D embedding. For example, on Wiki44k data set, the average F1 score of FCA2VEC is nearly 30% higher than the random baseline, which proves its effectiveness in embedding the Knowledge Graph [29].

The DeepFCA framework proposed by Li et al. [30] focuses on the task of biomedical ontology matching. By combining FCA with pre-trained word vectors (such as word vectors on PubMed and Wikipedia), the entity vectors are optimized by synonyms and antonyms. By constructing a Formal Context based on tokens, this method extracts positive sample pairs from the concept lattice, and designs a loss function to narrow the vector distance of semantically similar entities. The experimental results show that it can obtain better performances than traditional FCA methods on OAEI data sets.

3.5 FCA applied in Combined with ML& DM

Traditional FCA mainly deals with binary relational data, while Graph-FCA can deal with multivariate relational data by introducing the concept of n-ary, which has unique advantages in complex network analysis. The concept of n-ary is a binary group (X, Y), where x is a subset of nodes, y is a subset of edges, and it is satisfied that x is a closure of y and y is a closure of X. Closure operators are used to generate closures of nodes and edges, ensuring that every concept is generated based on existing closures.

Ferré et al. [5] described the algorithm and performance evaluation of Graph-FCA in their research. The core idea of Graph-FCA [4] is to expand the Formal Context into a hypergraph, and reveal the hierarchical structure and implicit relationship in the data through the generation of a schema base and the construction of a concept lattice.

Baixeries et al. [24] constructed a formal background, and associated an acyclic hypergraph with its symmetric dependency set, which proved that the formal background described the closure of acyclic hypergraphs accurately. This achievement not only supplements the FCA representation of AJD, but also shows the modular advantages of FCA in depicting RDBM dependency, which provides a new theoretical tool for database design, standardization and constraint reasoning.

Although FCA can generate a structured view of data and discover the implication relationship between data, it faces scalability problems when dealing with large-scale data, and it is difficult to integrate into neural network processes. Marquer et al. [31] and other scholars have studied the combination of deep learning and FCA, and put forward a framework called LatticeNN [32-34]. Through the deep generation model GraphRNN, the data-independent embedding model Bag of Attributes (BoA) has been developed, which opens up a new way for the realization of deep learning of FCA, is innovative in the embedding representation of formal concepts, and provides theoretical support and practical guidance for subsequent research [35,36].

4 Optimization for FCA

4.1 Parallelization improvement of the CbO algorithm

To overcome the high complexity of FCA, Ganter et al. proposed the Close-by-One (CbO) algorithm [37]. It has become the core algorithm of FCA because of the completeness of its concept generation and it can enumerate closure sets by polynomial delay to ensure the efficiency and accuracy of the calculation process. CbO algorithm ensures that every concept is generated based on the existing closures by gradually expanding the closure set, thus avoiding repeated calculation.

In recent years, with the explosive growth of data scale, the traditional serial CbO algorithm faces a serious performance bottleneck when dealing with large-scale Formal Context because of the high reasoning complexity of FCA. To solve this problem, many scholars have proposed a variety of parallelization improvement schemes, mainly including GPU-based acceleration models and distributed memory architecture.

Distributed Close-by-One (DCbO) model is suitable for very large-scale Formal Context, which alleviates the performance bottleneck of the traditional serial CbO algorithm. It decomposes large-scale context into manageable chunks by horizontal and vertical partitions, which avoids the bottleneck of single-machine memory and enables the algorithm to handle larger data sets. By using the auxiliary structure of group ID and Cuckoo filter, the communication times and traffic between machines are significantly reduced, and the efficiency of distributed computing is improved [38]. Through MPI communication, the algorithm uses sparse matrix transmission to exchange only non-zero closure differences, which ensures correctness and non-repetition. However, its network communication cost accounts for 35% of the total time, which is a big challenge, and it is still necessary to further compress the transmitted data. Zou et al. [39] propose a new parallel algorithm for computing formal concepts, which is composed of two parallel phases. The new algorithm parallelizes both the computations of the top L recursion levels and the workload distribution, which decouples worker threads from the main thread so as to improve the efficiency of the algorithm. For DCbO, reducing communication overhead through advanced data compression techniques or optimizing the communication protocol could significantly enhance its scalability. In comparison, Zou et al.'s approach focuses more on parallelizing computations and workload distribution, which may offer better performance in scenarios with less communication-intensive tasks. However, in highly distributed environments where communication latency is significant, the trade-off between computation and communication needs to be carefully balanced to achieve optimal performance.

4.2 Incremental CbO algorithm

The traditional serial CbO algorithm is suitable for static Formal Context, that is, the data will not change during the execution of the algorithm. When dealing with large-scale data, if the data changes dynamically, such as adding or deleting. The incremental Close-by-One (IncCbO) algorithm is an efficient concept lattice updating method for dynamic Formal Context, and its core goal is to calculate only the affected concept closures when the data increment arrives to avoid the overhead of total reconstruction.

In order to solve the problem of formal concept increment, Yang [26] et al. proposed the dynamic difference matrix model (IncCbO) algorithm. Only the concept of the changed part is recalculated, and the new concept is inserted into the original lattice structure, which greatly reduces the occupied memory, improves the FCA efficiency in a dynamic environment, and is suitable for the deployment of edge devices [40].

Marquer et al. [31] discussed how to combine these lattices through related concepts and concept alignment. It is emphasized that there are interactions between dimensions of related concepts, such as concept alignment and filtering, which may trigger more research on the application of FCA in neuroscience and concept learning, especially in different types of learning and concept structure modification.

Chen et al. [27] put forward the KG-CbO model algorithm, which maps the RDF triples in the Knowledge Graph to the Formal Context, and then uses the CbO algorithm to generate the ontology hierarchy, thus realizing the structural and hierarchical processing of the knowledge in the KG. It combines the rich Semantic Representation ability of Knowledge Graph and the deep mining ability of FCA to data structure [41].

Jens et al. [42] put forward the Neuro-CbO model, which uses graph neural network (GNN) to model the Formal Context, and predicts the priority of closure calculation by learning the embedded representation of nodes, thus optimizing the execution order of the CbO algorithm and improving the efficiency of concept lattice construction.

In order to improve the complexity of the algorithm, Zhu et al. [43] put forward knowledge state transfer methods based on attribute-oriented concept lattice for the conjunctive model and ability model respectively. This method has good performance when the skills cover a wide range, and the number of questions is large.

5 Conclusion and future direction

FCA plays an irreplaceable role in the research of knowledge representation and reasoning. This paper focused on the research works of FCA applied in knowledge network, which systematically summarized the existing methods in view of knowledge graph, social network, semantic representation, ML&DM. The value of FCA applied in knowledge networks has been indicated in the recent works of complex semantic representation and knowledge mining. Nevertheless, there still exist some problems to be solved which are listed as the future directions of FCA.

1. **FCA modeling based on Representation Learning**: As Deep Learning is good at feature extraction and representation of unstructured knowledge, FCA is good at finding structured tables and displaying them in the form of lattices. The combination of them could deeply mine domain knowledge. For example, the pre-training models of Deep Learning are utilized to provide the initial vectors with rich semantics for FCA. Then, with the help of FCA's concept lattice construction ability, knowledge can be further layered and systematized. It can improve the ability of domain knowledge reasoning and decision.
2. **Incremental FCA algorithm for Real-time Data**: With the dynamic data environment, the incremental FCA algorithm is still the key to enhancing the application value of FCA. Although the current algorithms can realize local updates, there still exist problems of cascade updates and data discarding. Therefore, it is necessary to develop an efficient incremental mechanism and reduce the computational complexity so that FCA can adapt to special scenarios with high real-time reasoning requirements.

3. **Visual and interpretable FCA for Analysis and Decision**: Due to the rise of large language models, it is important for the interpretability of AI. It is necessary to develop interactive visualization tools so that users can intuitively understand the hierarchical structure and semantic association of the concept lattice. Furthermore, the interpretation method of formal concept can be further explored, which can provide strong support for complex data analysis and decision-making.

To sum up, FCA has broad application prospects in theoretical research and actual data analysis of knowledge networks. The improvement and solution of the above problems can help FCA play an irreplaceable role in various tasks and help the innovation and development of AI.

References

1. Wenjie C, Yufeng W, Ruheng L, et al. An efficient location recommendation scheme based on clustering and data fusion. Comput. Electr. Eng., 2019, 77: 289-299.
2. Yu G, Yong Y, Jiang C, et al. Formal concept analysis assisted large-scale global optimization and its application to cloud task scheduling. Complex Intell. Syst, 2025, 11(6): 245-245.
3. Huang B, Li J, Li Q, et al. Competence-based knowledge space theory from the perspective of formal concept analysis. Int. J. Mach. Learn. Cybern. 2025:1-13.
4. Ferré. Graph-FCA Meets Pattern Structures. In: ICFCA, 2023, pp. 33-48
5. Ferré. A Proposal for Extending Formal Concept Analysis to Knowledge Graphs Lecture Notes in Computer Science, 2015, pp. 271-286
6. Fei H, Geyong M, Zheng P. K-clique Communities Detection in Social Networks based on Formal Concept Analysis. IEEE Systems Journal, 2015,11(1), 250-259.
7. Yixuan Y, Fei H, Beibei P, et al. Dynamic Maximal Cliques Detection and Evolution Management in Social Internet of Things: A Formal Concept Analysis Approach. IEEE Trans. Netw. Sci. Eng., 2022, 9(3): 1020-1032.
8. Qiang B and Guangqing T. Analysis of the Progress and Hotspots in Applied Research of FCA and Concept Lattice Theory Abroad. Data Analysis and Knowledge Discovery, 2010 11: 17-23.
9. Yunzhong Z, Di L and Yuanming Z. Research Trend of Knowledge Discovery Based on Formal Concept Analysis. Inf. Sci, 2018, 36(9): 153-158.
10. Chao Z, Zhiyu R, Wenchao W. Semantic Roles Mining Algorithms Based on Formal Concept Analysis. Computer Science, 2018, 45(12): 117-122+129.
11. Achyanta Kumar S, Shyamanta M. H, Smriti Kumar S. Formal concept analysis: current trends and directions. Artif. Intell. Rev, 2013, 44(1): 47-86
12. Jonas P, Dmitry I. I, Sergei O. K, et al. Formal concept analysis in knowledge processing: A survey on applications. Expert Syst. Appl, 2013, 40(16): 6538-6560.
13. Jonas P, Sergei O. K, Dmitry I. I, et al. Formal Concept Analysis in knowledge processing: A survey on models and techniques. Expert Syst. Appl, 2013,40(16): 6601-6623.
14. Jinhai L, Ling W, Zhuo, et al. Concept Lattice Theory and Method and Their Research Prospect. Pattern Recognition and Artificial Intelligence, 2020, 33(07): 619-642.
15. László K, Conceptual clustering with application on FCA context. Expert Syst. Appl, 2024, 245: 123013
16. Jinhai L, Yufei L, Yunlong M, et al. Meso-Granularity Labeled Method for Multi-Granularity Formal Concept Analysis. Journal of Computer Research and Development, 2020, 57(2): 447-458.
17. Lianbo M, Fengrong C, Huanxi Z. Formal Concept Analysis based Grouping Co-

Evolutionary Optimization Algorithms for arge-Scale Global Optimization. Chinese Journal of Computers, 2021, 44(7): 1310-1325.

18. Guanghua G, Yuyao C, Dong C, et al. Object Image Annotation Based on Formal Concept Analysis and Semantic Association Rules. Acta Automatica Sinica, 2020, 46(4): 767-781.
19. Liwei S, Zheng Y, Hongping L, et al. Knowledge Discovery and Rule Extraction Based on Heterogeneous Network Linguistic Formal Context. Pattern Recognition and Artificial Intelligence, 2024, 37(5): 469-478.
20. Jan K and Petr K. Systematic categorization and evaluation of CbO-based algorithms in FCA. Inf. Sci., 2021,575: 265-288.
21. Dominik D, Tom H, Maximilian S. FCA2VEC Embedding Techniques for Formal Concept Analysis. Complex Data Analytics with Formal Concept Analysis, 2022: 47-74.
22. Qian H, Zhong Y, Keyun Q, et al. A novel outlier detection approach based on formal concept analysis. Knowl. Based Syst, 2023, 268: 110486.
23. Jie G, Fei H, Zheng P, et al. Learning Concept Interestingness for Identifying Key Structures from Social Networks. IEEE Trans. Netw. Sci. Eng., 2021, 8(4): 3220-3232
24. Jaume B. A Formal Context for Closures of Acyclic Hypergraphs. In: ICFCA, 2019, pp. 259-273
25. Hanika T, Marx M, Stumme G. Discovering implicational knowledge in Wiki data. In: ICFCA, 2019, pp. 315-323
26. Erhe Y, Fei H, Yixuan Y, et al. Incremental Entity Summarization with Formal Concept Analysis. IEEE Trans. Serv. Comput. 2022, 15(6): 3289-3303
27. Mehwish A, Amedeo N. An Approach Towards Classifying and Navigating RDF data based on Pattern Structures. In: ICFCA, 2015, pp.33-48
28. Fei H, Zheng P, Laurence T. Y. Diversified top-k maximal clique detection in Social Internet of Things. Future Gener. Comput. Syst. 2020, 107: 408-417.
29. Steffffen S. Concepts in Application Context. In: ICFCA, 2019, pp. 45-52
30. Guoxuan L. DeepFCA: Matching Biomedical Ontologies Using Formal Concept Analysis Embedding Techniques. In: ICMHI, 2020, pp.259-265.
31. Esteban M. LatticeNN Deep Learning and Formal Concept Analysis. Ph.D thesis, 2020
32. Egor D and Sergei O. K. Decision Concept Lattice vs. Decision Trees and Random Forests. In: ICFCA, 2021, pp. 252–260
33. Doudou G and Weihua X. R-FCCL: An Approach of Fuzzy-Based Concept-Cognitive Learning with Robustness for High-Dimensional Data. Journal of Computer Research and Development, 2025, 62(2):383-396.
34. Francisco J. Valverde-A, Carmen Pelaez-M. Leveraging Formal Concept Analysis to Improve n-fold validation in Multilabel Classification. In: ICFCA, 2021, pp. 44-51
35. Peter E, Jean V. A Survey of Hybrid Representations of Concept Lattices in Conceptual Knowledge Processing. In: ICFCA, 2010, pp. 296-311
36. Uta P. Representing Concept Lattices with Euler Diagrams. In: ICFCA, 2023, pp. 183-197
37. Jan K, Petr K. Systematic categorization and evaluation of CbO-based algorithms in FCA. Inf. Sci. 2021, 575: 265-288.
38. Muneeswaran P, Jyoti, Sriram K. A Hybrid Partitioning Strategy for Distributed FCA. In: CLA, 2020, pp. 71-82
39. Ligeng Z, Tingting H, Jianhua D. A new parallel algorithm for computing formal concepts based on two parallel stages. Inf. Sci. 2022, 586: 514-524
40. Uta P. Conceptual Alignment with Formal Concept Analysis. In: ICFCA, 2019, pp. 14-27
41. Hua M, Chang L, Xiaolei Y, et al. Knowledge level extraction method based on concept lattice and its application. J. Huazhong Univ. Sci. Technol., 2024, 52(11): 37-42.
42. Jens K. Concept Lattices of RDF Graphs. Computer Science, 2015: 81-91
43. Yongfan Z, Qingyuan X, Yinfeng Z, et al. Knowledge-state Transfer Method Based on Attribute-oriented Concept Lattices. J. Chin. Comput. Syst., 2024, 45(11): 2647-2655.