Interval-Valued Intuitionistic Fuzzy Decision With Graph Pattern in Big Graph

Lei Li, *Senior Member, IEEE*, Lan Jiang, Chenyang Bu, *Member, IEEE*, Yi Zhu, and Xindong Wu, *Fellow, IEEE*

***Abstract*—Graph pattern matching (GPM) in big graph has been widely used in decision making, such as expert finding, social group discovery, etc. However, these existing works consider neither the preference of the decision maker (DM), nor the subjectivity of constraints during the process of GPM. Therefore, this paper proposes an interval-valued intuitionistic fuzzy decision (IVIFD) with graph pattern in big graph. As traditionally IVIFD can maximally reduce the uncertainty of decision making and it is only valid for small datasets, which makes it impossible to be applied in big graph. In this paper, GPM is adopted to prune the searching space, which makes it possible to process IVIFD under the preference of the DM later. Technically, firstly, each DM selects the preferred vertices and/or edges in big graph, and the interval-valued intuitionistic fuzzy preference (IVIFP) is calculated and used to form the contextual constraints to conduct GPM. Secondly, a probability-certainty density function is introduced to capture the subjective probability of the contextual preference of subgraphs via the bijection from rating space of the context to preference space of the context, which leads to an interval-valued intuitionistic fuzzy set (IVIFS). In addition to the IVIFS, the IVIFD is made through interval-valued intuitionistic fuzzy cross entropy and grey relation degree. Moreover, the weight problem between different contexts is taken into account and handled respectively as three cases. Finally, numerical experiments and perturbation analysis validate the effectiveness and stability of our proposed method, and verify its necessity and efficiency through ablation experiments.**

***Index Terms*—Interval-valued intuitionistic fuzzy decision, graph pattern matching, big graph, preference.**

I. Introduction

When facing with complex decision making, decision makers (DMs) usually have to consider many contextual factors [20], [26]. For example, when selecting a team of experts, the DMs need to consider not only the experts' expertise, but also the trustworthiness between experts in order to efficiently complete the target task. However, due to the uncertainty and the ambiguity of human thinking, it is difficult for DMs to evaluate these contextual factors with crisp numbers [32]. Thus, Atanassov [1] proposes the concept of an intuitionistic fuzzy set (IFS), which takes the concepts of membership, non-membership and hesitancy degree together, and can simultaneously express the evidence of support, opposition and uncertainty. Subsequently, Atanassov and Gargov [2] extend an IFS to an interval-valued intuitionistic fuzzy set (IVIFS). Compared with IFSs, IVIFSs are more flexible and effective in expressing fuzziness, uncertainty and hesitancy in real-world decision making problems [2], [15].

In recent years, with the significant increase in the interactions and complex relations of Big Data, many scholars have paid more attention to big graph. One of the most important research areas about big graph is graph pattern matching (GPM) [18], [19], [22], which can be used in a variety of big graphs, such as social networks, biological networks, and communication networks. During GPM, with the designed pattern graphs that show specific and personalized requirements, subgraphs can be matched from big graphs to satisfy the required patterns. These matched subgraphs will complete the required tasks, such as expert finding [6]. Mathematically, given a query which contains a pattern graph G_P and a data graph G_D, a GPM algorithm finds those subgraphs G_M that match G_P in G_D [19]. For example, as shown in Fig. 1, it is a small domain expert-oriented job search social network. G_D in Fig. 1(b) is a data graph, where each vertex $v_i \in V$ is associated with the context about professional ability, to illustrate the professional ability of expertise of v_i. Such as $\rho_{N_{B_1}} = (0.6, 0.8, 0.5, 0.7)$ are ratings for vertex N_{B_1} with the context about professional ability. Moreover, each edge (v_i, v_j) is associated with the context about social trust, and the context about social intimacy, to illustrate trustworthiness and intimacy social relationships between experts [21], [29], [38]. Such as $T_{N_{B_1},N_{C_2}} = (0.6, 0.7, 0.9, 0.4)$, $R_{N_{B_1},N_{C_2}} = (0.6, 0.5, 0.8, 0.4)$ are ratings for edge (N_{B_1}, N_{C_2}) with the context about social trust and the

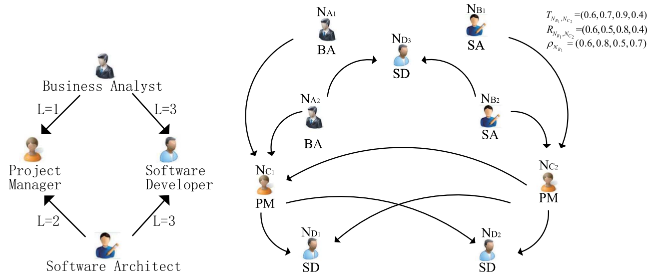
Manuscript received 22 May 2021; revised 20 September 2021; accepted 6 December 2021. Date of publication 25 January 2022; date of current version 23 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grants 62076087, 91746209, and 61906059, in part by the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT) of the Ministry of Education of China under Grant IRT17R32, and in part by the Engineering Research Center for Software Testing and Evaluation of Fujian Province under Grant C2015004. *(Corresponding author: Chenyang Bu.)*

Lei Li, Lan Jiang, and Chenyang Bu are with the Key Laboratory of Knowledge Engineering with Big Data, Hefei University of Technology, Ministry of Education, Hefei 230601, China, and with the Intelligent Interconnected Systems Laboratory of Anhui Province, Hefei University of Technology, Hefei 230601, China, and also with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China (e-mail: lilei@hfut.edu.cn; jl18726094859@163.com; chenyangbu@hfut.edu.cn).

Yi Zhu is with the School of Information Engineering, Yangzhou University, Yangzhou 225009, China, and also with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China (e-mail: zhuyi@yzu.edu.cn).

Xindong Wu is with the Mininglamp Academy of Sciences, Mininglamp Technologies, Beijing 100083, China (e-mail: wuxindong@mininglamp.com).

Digital Object Identifier 10.1109/TETCI.2022.3141062



(a) Pattern Graph G_P (b) Data Graph G_D

Fig. 1. Data graph and pattern graph in a social network.

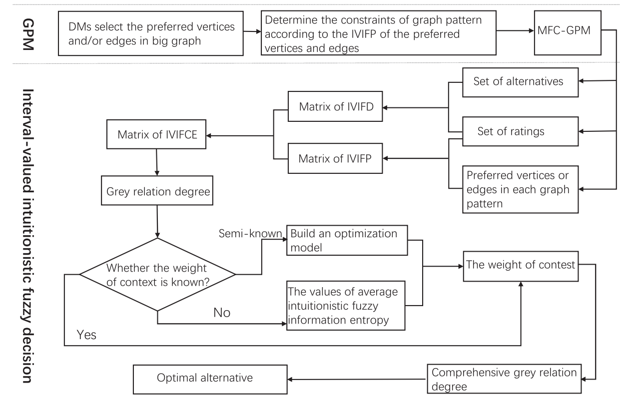


Fig. 2. The framework of our proposed method.

context about social intimacy respectively. As for GPM-based expert location in social networks, Fan *et al.* [6] identify top-K experts in a social network by supporting bounded simulation of graph patterns. However, this bounded simulation hasn't take the rich contextual information on the vertices and edges in big graph into account to get better query results. In [18], multi-fuzzy-constrained graph pattern matching (MFC-GPM) is proposed to take into account both the contextual information on the vertices and edges in big graph and the fuzziness of information. In this literature, although there are some existing works [18] which have taken into account fuzziness into GPM, they only consider the fuzziness of contextual constraints. As an Interval-Valued Intuitionistic Fuzzy Decision (IVIFD) can maximally reduce the uncertainty of decision making, with the personal preference of DM an interval-valued intuitionistic fuzzy preference (IVIFP) can be introduced in this paper.

As traditionally IVIFD is only valid for small datasets, which makes it impossible to be applied in big graph. However, as the volume of information swells, DMs frequently need to make decisions in vast amounts of data. For example, DMs need to find an expert team in job search social networks with complex relationships, and there may be thousands of alternatives. Thus, in order to be able to make IVIFD in big graph, this paper proposes to use GPM for pruning. As each subgraph matched by the requested graph pattern may have many vertices and edges with different contextual information, the contextual preference of subgraphs can be aggregated from the contextual ratings of vertices and edges in the subgraph. As the degree of belief that individuals have to perform the contextual behavior satisfactorily [12], the certainty of contextual preference can be taken as the subjective probability about the context from the aggregated contextual ratings. Following [12] and [34], we can introduce a probability-certainty density function (PCDF) to capture the subjective probability of the contextual preference of subgraphs via the bijection from rating space of the context to preference space of the context, which leads to an IVIFS, including interval-valued positive/negative preference, and interval-valued uncertainty. In fact, Pang *et al.* [35] also apply graph theory to fuzzy modeling to solve the problem of a non-monotone fuzzy rule in Takagi-Sugeno-Kang Fuzzy Inference System. In contrast, our work focuses on solving problems that traditionally IVIFD cannot be applied in big graph and addressing the aggregation of contextual preference of subgraphs. Generally, in this work, we firstly use the determined contextual constraints for GPM, and then process the IVIFP to arrive at decision-making results that match the preferences of DMs, which has been also illustrated in Fig. 2.

The theoretical and technical contributions in this paper can be analyzed and summarized as follows:

1) As traditionally IVIFD can maximally reduce the uncertainty of decision making and it is only valid for small datasets, which makes it impossible to be applied in big graph. In this paper, MFC-GPM is adopted to prune the searching space, which makes it possible to process IVIFD under the preference of the DM later. With these two steps, we can not only reduce the uncertainty of decision making but also handle it effectively and efficiently.
2) We introduce a PCDF to capture the subjective probability of the contextual preference of subgraphs via the bijection from rating space of the context to preference space of the context, which leads to an IVIFS, including interval-valued positive/negative preference, and interval-valued uncertainty.
3) A numerical experiment shows the effectiveness of our proposed IVIFD in big graph. The perturbation analysis and ablation experiments of three large-scale real-world social graphs show the stability and efficiency of our proposed method, and the comparison with the benchmark data in the ablation experiment also proves the necessity of steps in IVIFD under the preference of the DM.

The rest of this paper is organized as follows. In Section II, some related works on IVIFD and GPM are reviewed. Then, the contextual preference of subgraphs can be estimated as an IVIFS by a PCDF. In Section IV, the related definitions about IVIFD are introduced and evaluated. Section V introduces the steps of IVIFD in big graph, and discusses the related weight problems during this process. Section VI introduces grey relation degree to perturbation analysis and ablation experiment to illustrate the stability and effectiveness of the proposed method. Section VII summarizes this paper and presents the future work.

II. Related Work

Since a crisp number cannot describe uncertainty and imprecision properly, Zadeh [39] proposes a fuzzy set. Afterwards, many extended theories and methods are proposed to handle different cases and applications. An interval-valued fuzzy set (IFS) presented by Zadeh [40], generalizes the membership degree from a crisp number to an interval number. In addition, Atanassov [1] simultaneously considers the membership degree and non-membership degree of the element to an IFS. Hesitant fuzzy sets are defined and introduced by Torra [31] in terms of a function in addition to membership degree and non-membership degree for each element in the domain. Atanassov and Gargov [2] extend IFS to an IVIFS, which describes membership degree and non-membership degree via interval numbers and provides a more reasonable mathematical model to deal with uncertain events and fuzzy information. Due to an appropriate description of uncertainty and imprecision, fuzzy theories are widely used in decision making [15]. However, while traditional IVIFD minimizes uncertainty of decision making, it is only effective for small datasets. For example, an investment company selects one of the four candidate companies to invest in [16], and DM make talent selection among several candidates [17]. In fact, with the advancement of information technology, DMs always need to make decisions in hundreds or tens of thousands of alternatives. Thus, we introduce graph pattern matching for pruning and PCDF to capture the subjective probability of the contextual of subgraphs.

The research in the field of GPM can be divided into isomorphism-based GPM and simulation-based GPM. Isomorphism-based GPM requires that the topological structure of the matched subgraph should be exactly the same as the pattern graph. Representative algorithms are DDST algorithm [28], IncBMatch algorithm [4], G-Ray algorithm [30], R-join algorithm [3], etc. Isomorphism-based GPM is more common to be applied in areas with strict structural requirements such as protein structure matching [11]. Compared with isomorphism-based GPM, simulation-based GPM is more flexible and is suitable for the detection and/or analysis of interrelationships between vertices. Henzinger *et al.* [9] propose an algorithm for calculating similarity relationships on labeled graphs, and further introduce the graph simulation, which requires that vertices in the data graph maintain the successor relationship with the corresponding vertices in the pattern graph. Moreover, Fan *et al.* [5] propose bounded simulation, which matches vertices with the same label and the path whose length is not greater than the bounded length in the data graph, that is, a path in the data graph matches an edge of the pattern graph. Ma *et al.* [24] propose strong simulation based on bounded simulation. However, all these existing works have failed to use the contextual information of vertices and/or edges contained in the big graph data. Shemshadi *et al.* [27] consider multi-label information on vertices but without the constraints on the edges. Liu *et al.* [22] propose MC-GPM, which is an extension of bounded simulation, where constraints are set on vertices and edges to match more accurate results. Subsequently, Liu *et al.* [23] proposed the concept of strong social graph to improve efficiency of MC-GPM. However, there is still no fuzziness mentioned in existing MC-GPM methods. With fuzzy attributes, Liu *et al.* [18] propose MFC-GPM to prevent the loss of subgraphs during matching.

III. Certainty-Based Contextual Preference on Subgraphs

As each subgraph matched by the requested graph pattern may have many vertices and edges with different contextual information, the contextual preference of subgraphs can be aggregated from the contextual ratings of vertices and edges in the subgraph. As the degree of belief that individuals have to perform the contextual behavior satisfactorily [12], the certainty of contextual preference can be taken as the subjective probability about the context from the aggregated contextual ratings. Following [12] and [34], we can introduce a PCDF to capture the subjective probability of the contextual preference of subgraphs via the bijection from rating space of the context to preference space of the context, which leads to an IVIFS, including interval-valued positive/negative preference, and interval-valued uncertainty.

A. Rating Space of Contexts

In most existing rating systems[1],[2],[3], the ratings are usually represented by a series of fixed discrete numbers. For example, the ratings at eBay[1] are in the set of $\{-1, 0, 1\}$. At Epinions[2], each rating is an integer in $\{1, 2, 3, 4, 5\}$. At YouTube[3], each rating is in $\{-10, -9, \ldots, 10\}$. In order to analyze these ratings conveniently, they should be normalized to the range of [0,1], and then it can be partitioned into k mutually exclusive ratings, say $r_1, r_2, \ldots,$ and r_k, where $0 \le r_{i-1} < r_i \le 1$. For example, at Epinions[2], after normalization, the ratings are in $\{0, 0.25, 0.5, 0.75, 1\}$. Let $p_i^{(A^{C_j})} = P(r_i, A^{C_j})$ be the probability to obtain rating r_i on context C_j in subgraph A $(1 \le i \le k)$, and then

$$V^{(A^{C_j})} = \{p_1^{(A^{C_j})}, p_2^{(A^{C_j})}, \ldots, p_k^{(A^{C_j})}\}$$

where $\sum_{i=1}^{k} p_i^{(A^{C_j})} = 1$. The number of occurrences $x_i^{(A^{C_j})}$ of r_i on C_j in A conforms to a multinomial distribution [10]. That is due to the fact that if r_i has exactly one of k possible outcomes with probability $p_i^{(A^{C_j})}$, the number of occurrences $x_i^{(A^{C_j})}$ of r_i must follow a multinomial distribution [10], i.e.,

$$f(X^{(A^{C_j})}|V^{(A^{C_j})}) = \frac{\left(\sum_{i=1}^{k} x_i^{(A^{C_j})}\right)!}{\prod_{i=1}^{k}(x_i^{(A^{C_j})}!)} \prod_{i=1}^{k} \left(p_i^{(A^{C_j})}\right)^{x_i^{(A^{C_j})}}, \tag{1}$$

Definition 1: The rating space X of contexts consists of the occurrences of all the ratings on contexts $\{C_j\}$ in subgraph A as

[1][Online]. Available: http://www.eBay.com/
[2][Online]. Available: http://www.epinions.com/
[3][Online]. Available: http://www.youtube.com/

follows

$$X=\left\{\begin{array}{c}X^{(A^{C_1})}\\X^{(A^{C_2})}\\\vdots\\X^{(A^{C_m})}\end{array}\right\}=\left\{\begin{array}{cccc}x_1^{(A^{C_1})}&x_2^{(A^{C_1})}&\dots&x_k^{(A^{C_1})}\\x_1^{(A^{C_2})}&x_2^{(A^{C_2})}&\dots&x_k^{(A^{C_2})}\\\vdots&\vdots&\vdots&\vdots\\x_1^{(A^{C_m})}&x_2^{(A^{C_m})}&\dots&x_k^{(A^{C_m})}\end{array}\right\}$$

where $x_i^{(A^{C_j})}\geq 0$, and $x_i^{(A^{C_j})}\in\mathbb{R}, i=1,2,\dots,k$.

B. Preference Space of Contexts

According to the trisecting and acting models of human cognitive behaviors [13], [36], the preference space of contexts consists of three-way preference on contexts $\{C_j\}$ in subgraph A as follows, which includes positive preference $\widetilde{\mu}^{(A^{C_j})}$, negative preference $\widetilde{\nu}^{(A^{C_j})}$, and uncertainty $\widetilde{\pi}^{(A^{C_j})}$.

Definition 2: The IFS $\widetilde{A}^{(C_j)}$ for the preference on context C_j in subgraph A can be represented as

$$\widetilde{A}^{(C_j)}=\{\widetilde{\mu}^{(A^{C_j})},\widetilde{\nu}^{(A^{C_j})},\widetilde{\pi}^{(A^{C_j})}\}\tag{2}$$

where $\widetilde{\mu}^{(A^{C_j})}\subseteq[0,1]$, $\widetilde{\nu}^{(A^{C_j})}\subseteq[0,1]$, and $\widetilde{\pi}^{(A^{C_j})}\subseteq[0,1]$.

Definition 3: Preference space P of contexts can be defined as

$$P=\left\{\begin{array}{c}\widetilde{A}^{(C_1)}\\\widetilde{A}^{(C_2)}\\\vdots\\\widetilde{A}^{(C_m)}\end{array}\right\}=\left\{\begin{array}{ccc}\widetilde{\mu}^{(A^{C_1})}&\widetilde{\nu}^{(A^{C_1})}&\widetilde{\pi}^{(A^{C_1})}\\\widetilde{\mu}^{(A^{C_2})}&\widetilde{\nu}^{(A^{C_2})}&\widetilde{\pi}^{(A^{C_2})}\\\vdots&\vdots&\vdots\\\widetilde{\mu}^{(A^{C_m})}&\widetilde{\nu}^{(A^{C_m})}&\widetilde{\pi}^{(A^{C_m})}\end{array}\right\}$$

C. Certainty of Preference on Contexts

The Bayesian inference is adopted to update the probability with the available information about the preferences, i.e., update the posterior distribution $f(V^{(A^{C_j})}|X^{(A^{C_j})})$ from the prior distribution, i.e., a uniform distribution without additional information [8], [10]. As the cumulative probability of a distribution within [0,1] equals 1, the density of a PCDF has the mean value 1 within [0,1], and this makes $f(X^{(A^{C_j})}|V^{(A^{C_j})})=1$. Then, the certainty can be determined by the deviations of posterior distribution from the prior distribution, i.e., a uniform distribution. Hence, we have the following definition about certainty [12], [34].

Definition 4: The certainty $\Phi^{(A^{C_j})}$ of preference on context C_j in subgraph A can be estimated from ratings $X^{(A^{C_j})}$ as

$$\begin{aligned}&\Phi^{(A^{C_j})}(X^{(A^{C_j})})\\&=\frac{1}{2}\int_\Omega|f(V^{(A^{C_j})}|X^{(A^{C_j})})-f(X^{(A^{C_j})}|V^{(A^{C_j})})|\mathrm{d}V^{(A^{C_j})}\\&=\frac{1}{2}\int_\Omega|\frac{\prod_{i=1}^k\left(p_i^{(A^{C_j})}\right)^{x_i^{(A^{C_j})}}}{\int_\Omega\prod_{i=1}^k\left(p_i^{(A^{C_j})}\right)^{x_i^{(A^{C_j})}}\mathrm{d}V^{(A^{C_j})}}-1|\mathrm{d}V^{(A^{C_j})},\end{aligned}$$

where $\frac{1}{2}$ is to remove the double counting of the deviations.

Since $\frac{1}{2}$ is the middle point of the range of ratings [0,1], which represents the neutral preference between positiveness and negativeness, the ratings in $(\frac{1}{2},1]$ can indicate positive preference and the ratings in $[0,\frac{1}{2})$ can indicate negative preference. Hence, we can have the following definition.

Definition 5: The *interval-valued positiveness* $[\underline{\alpha}^{(A^{C_j})}(X^{(A^{C_j})}),\overline{\alpha}^{(A^{C_j})}(X^{(A^{C_j})})]$ can be defined to be the interval of degree for ratings on context C_j in subgraph A to be positive

$$\underline{\alpha}^{(A^{C_j})}(X^{(A^{C_j})})=\frac{\sum_{r_i>\frac{1}{2}}(2r_i-1)x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}},\tag{3}$$

$$\overline{\alpha}^{(A^{C_j})}(X^{(A^{C_j})})=\frac{\sum_{r_i>\frac{1}{2}}x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\tag{4}$$

and the *interval-valued negativeness* $[\underline{\beta}^{(A^{C_j})}(X^{(A^{C_j})}),\overline{\beta}^{(A^{C_j})}(X^{(A^{C_j})})]$ can be defined by

$$\underline{\beta}^{(A^{C_j})}(X^{(A^{C_j})})=\frac{\sum_{r_i<\frac{1}{2}}(1-2r_i)x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\tag{5}$$

$$\overline{\beta}^{(A^{C_j})}(X^{(A^{C_j})})=\frac{\sum_{r_i<\frac{1}{2}}x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\tag{6}$$

D. Bijection From Rating Space to Preference Space

With Definitions 4 and 5, we can have the following definition.

Definition 6: The bijection from rating space X of contexts to preference space P of contexts, i.e., from ratings $X^{(A^{C_j})}$ to an *interval-valued intuitionistic fuzzy set* (IVIFS) of contextual preference

$$\widetilde{A}^{(C_j)}=\{\widetilde{\mu}^{(A^{C_j})},\widetilde{\nu}^{(A^{C_j})},\widetilde{\pi}^{(A^{C_j})}\},\tag{7}$$

where *interval-valued positive preference* $\widetilde{\mu}^{(A^{C_j})}=[\underline{\mu}^{(A^{C_j})},\overline{\mu}^{(A^{C_j})}]$, *interval-valued negative preference* $\widetilde{\nu}^{(A^{C_j})}=[\underline{\nu}^{(A^{C_j})},\overline{\nu}^{(A^{C_j})}]$, and *interval-valued uncertainty* $\widetilde{\pi}^{(A^{C_j})}=[\underline{\pi}^{(A^{C_j})},\overline{\pi}^{(A^{C_j})}]$, can be estimated as

$$\begin{aligned}\underline{\mu}^{(A^{C_j})}(X^{(A^{C_j})})&=\underline{\alpha}^{(A^{C_j})}(X^{(A^{C_j})})\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\\&=\frac{\sum_{r_i>\frac{1}{2}}(2r_i-1)x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\end{aligned}\tag{8}$$

$$\begin{aligned}\overline{\mu}^{(A^{C_j})}(X^{(A^{C_j})})&=\overline{\alpha}^{(A^{C_j})}(X^{(A^{C_j})})\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\\&=\frac{\sum_{r_i>\frac{1}{2}}x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\end{aligned}\tag{9}$$

$$\begin{aligned}\underline{\nu}^{(A^{C_j})}(X^{(A^{C_j})})&=\underline{\beta}^{(A^{C_j})}(X^{(A^{C_j})})\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\\&=\frac{\sum_{r_i<\frac{1}{2}}(1-2r_i)x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\end{aligned}\tag{10}$$

$$\begin{aligned}\overline{\nu}^{(A^{C_j})}(X^{(A^{C_j})})&=\overline{\beta}^{(A^{C_j})}(X^{(A^{C_j})})\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\\&=\frac{\sum_{r_i<\frac{1}{2}}x_i^{(A^{C_j})}}{\sum_{i=1}^k x_i^{(A^{C_j})}}\Phi^{(A^{C_j})}(X^{(A^{C_j})}),\end{aligned}\tag{11}$$

$$\underline{\pi}^{(A^{C_j})} = 1 - \overline{\mu}^{(A^{C_j})}(X^{(A^{C_j})}) - \overline{\nu}^{(A^{C_j})}(X^{(A^{C_j})}), \tag{12}$$

$$\overline{\pi}^{(A^{C_j})} = 1 - \underline{\mu}^{(A^{C_j})}(X^{(A^{C_j})}) - \underline{\nu}^{(A^{C_j})}(X^{(A^{C_j})}). \tag{13}$$

Hence, the contextual preference of subgraphs can be estimated as an IVIFS by a PCDF, which includes interval-valued positive/negative preference, and interval-valued uncertainty.

IV. Interval-Valued Intuitionistic Fuzzy Decision

To evaluate the difference between two IVIFSs, with Definition 6, the interval-valued intuitionistic fuzzy cross entropy (IVIFCE) [14], [37] is introduced.

Definition 7: With IVIFSs $\widetilde{A}^{(C_{j_1})} = \{\widetilde{\mu}^{(A^{C_{j_1}})}, \widetilde{\nu}^{(A^{C_{j_1}})}, \widetilde{\pi}^{(A^{C_{j_1}})}\}$ and $\widetilde{B}^{(C_{j_2})} = \{\widetilde{\mu}^{(B^{C_{j_2}})}, \widetilde{\nu}^{(B^{C_{j_2}})}, \widetilde{\pi}^{(B^{C_{j_2}})}\}$, the IVIFCE between $\widetilde{A}^{(C_{j_1})}$ and $\widetilde{B}^{(C_{j_2})}$ can be estimated as

$$\begin{aligned}\widetilde{CE}(\widetilde{A}^{(C_{j_1})}, \widetilde{B}^{(C_{j_2})}) &= \sum_{i=1}^{n}\left(\frac{2+\mu^{*(A^{C_{j_1}})}-\nu^{*(A^{C_{j_1}})}}{4}\right)\\ &\times \log_2 \frac{2(2+\mu^{*(A^{C_{j_1}})}-\nu^{*(A^{C_{j_1}})})}{4+\mu^{*(A^{C_{j_1}})}-\nu^{*(A^{C_{j_1}})}+\mu^{*(B^{C_{j_2}})}-\nu^{*(B^{C_{j_2}})}}\\ &+\sum_{i=1}^{n}\left(\frac{2-\mu^{*(A^{C_{j_1}})}+\nu^{*(A^{C_{j_1}})}}{4}\right)\\ &\times \log_2 \frac{2(2-\mu^{*(A^{C_{j_1}})}+\nu^{*(A^{C_{j_1}})})}{4-\mu^{*(A^{C_{j_1}})}+\nu^{*(A^{C_{j_1}})}-\mu^{*(B^{C_{j_2}})}+\nu^{*(B^{C_{j_2}})}}\end{aligned}$$

where

$$\mu^{*(A^{C_j})} = \underline{\mu}^{(A^{C_j})} + \overline{\mu}^{(A^{C_j})}$$

and

$$\nu^{*(A^{C_j})} = \underline{\nu}^{(A^{C_j})} + \overline{\nu}^{(A^{C_j})}$$

From Definition 7, it is easy to show that $\widetilde{CE}$ is asymmetrical. In order to improve this, we have the following definition.

Definition 8: The *improved IVIFCE* can be evaluated as follows

$$\begin{aligned}\widetilde{CE}^{*}(\widetilde{A}^{(C_{j_1})}, \widetilde{B}^{(C_{j_2})}) &= \widetilde{CE}(\widetilde{A}^{(C_{j_1})}, \widetilde{B}^{(C_{j_2})})\\ &+ \widetilde{CE}(\widetilde{B}^{(C_{j_2})}, \widetilde{A}^{(C_{j_1})})\end{aligned}$$

After the DM selects the preferred vertices or edges with certain context in big graph which satisfy the personal requirements to complete certain task, his/her interval-valued intuitionistic fuzzy preference (IVIFP) can be represented as an IVIFS as follows.

Definition 9: The IVIFP can be estimated as an IVIFS

$$\widetilde{o}^{(A^{C_j})} = \{\widetilde{\mu}^{(A_o^{C_j})}, \widetilde{\nu}^{(A_o^{C_j})}, \widetilde{\pi}^{(A_o^{C_j})}\} \tag{14}$$

$$\widetilde{\mu}^{(A_o^{C_j})} = [\min \underline{\mu}^{(\widehat{A}^{C_j})}, \max \overline{\mu}^{(\widehat{A}^{C_j})}], \tag{15}$$

$$\widetilde{\nu}^{(A_o^{C_j})} = [\min \underline{\nu}^{(\widehat{A}^{C_j})}, \max \overline{\nu}^{(\widehat{A}^{C_j})}], \tag{16}$$

$$\widetilde{\pi}^{(A_o^{C_j})} = [1 - \max \overline{\mu}^{(\widehat{A}^{C_j})} - \max \overline{\nu}^{(\widehat{A}^{C_j})}, \tag{17}$$

$$1 - \min \underline{\mu}^{(\widehat{A}^{C_j})} - \min \underline{\nu}^{(\widehat{A}^{C_j})}], \tag{18}$$

where $\widehat{A}^{C_j} \in A^{C_j}$ are the preferred vertices or edges in subgraph A with context C_j which satisfy the personal requirements to complete certain task decided by a DM.

With Definitions 8 and 9, the similarity between a subgraph and the IVIFP can be evaluated as follows [7], [33].

Definition 10: With the IVIFCE, from the aspect of the IVIFP, the satisfaction of a subgraph can be evaluated according to a *grey relation degree* as follows,

$$\widetilde{\gamma}(\widetilde{A}_l^{(C_j)}, \widetilde{o}^{(A_l^{C_j})}) = \frac{\min_l \min_j \widetilde{CE}_{lj}^{*} + \xi \max_l \max_j \widetilde{CE}_{lj}^{*}}{\widetilde{CE}_{lj}^{*} + \xi \max_l \max_j \widetilde{CE}_{lj}^{*}} \tag{19}$$

where $\widetilde{CE}_{lj}^{*} = \widetilde{CE}^{*}(\widetilde{A}_l^{(C_j)}, \widetilde{o}^{(A_l^{C_j})})$, ξ is the grey distinguishing parameter to control the difference of IVIFSs $\widetilde{A}_l^{(C_j)}$ and $\widetilde{o}^{(A_l^{C_j})}$, and it can be set as $\xi = \frac{1}{2}$ here.

When the weights for different contexts $\{C_j\}$ are known, as for all contexts $\{C_j\}$, the comprehensive grey relation degree of subgraph A_l can be evaluated as follows.

Definition 11: With subgraph A_l ($l \in [1, q]$) and the weight ω_j of context C_j, their *comprehensive grey relation degree* can be aggregated from all context $\{C_j\}$ as

$$\widetilde{\gamma}^{*}(\widetilde{A}_l) = \sum_{j=1}^{m} \omega_j \widetilde{\gamma}(\widetilde{A}_l^{(C_j)}, \widetilde{o}^{(A_l^{C_j})}) \tag{20}$$

where $\sum_{j=1}^{m} \omega_j = 1$.

The larger $\widetilde{\gamma}^{*}$, the closer subgraph A_l to IVIFP, which is preferable.

Definition 12: With the improved IVIFCE from the IVIFP, the *average intuitionistic fuzzy information entropy* $E(C_j)$ for context $\{C_j\}$ can be evaluated as

$$E(C_j) = -\frac{1}{\ln(\psi)} \sum_{l=1}^{q} \left[\frac{\widetilde{CE}_{lj}^{*}}{\sum_{l=1}^{q} \widetilde{CE}_{lj}^{*}} \ln\left(\frac{\widetilde{CE}_{lj}^{*}}{\sum_{l=1}^{q} \widetilde{CE}_{lj}^{*}}\right)\right] \tag{21}$$

where $\widetilde{CE}_{lj}^{*} = \widetilde{CE}^{*}(\widetilde{A}_l^{(C_j)}, \widetilde{o}^{(A_l^{C_j})})$, ψ is the parameter to make $E(C_j) = 0$ with $\widetilde{CE}_{lj}^{*} = 0$ and $E(C_j) = 1$ with $\widehat{A}^{C_j} = A^{C_j}$.

V. IVIFD in Big Graph

Our proposed *Interval-Valued Intuitionistic Fuzzy Decision (IVIFD) in Big Graph* mainly consists of two steps, which are Multi-Fuzzy-Constrained Graph Pattern Matching (MFC-GPM) and IVIFD. As traditionally IVIFD is only valid for small datasets, which makes it impossible to be applied in big graph. In this paper, MFC-GPM is adopted to prune the searching space, which makes it possible to process IVIFD later. The formal framework for this method is shown in Fig. 2, and the detailed process is illustrated as follows.

Step 1: Firstly, let's conduct the MFC-GPM, with the framework shown in Fig. 3, where the vertices with the same color represent the vertices with the same label.

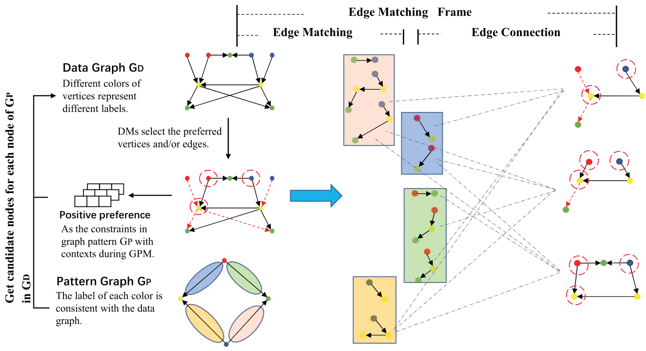


Fig. 3. The framework of GPM.

Step 1.1: A reputation system stores all ratings for vertices and edges $\widehat{A}^{C_j}$ with certain context C_j; Meanwhile, DMs select the preferred vertices or edges $\widehat{A}^{C_j}$ in big graph under context C_j, which represent the satisfied requirements to complete certain task preferred by DMs, which leads to IVIFP $\widetilde{o}^{(A^{C_j})}$.

Step 1.2: During GPM, the constraints in matched subgraph A with context C_j can be determined by $\{\theta_j^{(1)}\widetilde{\mu}^{(A_o^{C_j})}, \theta_j^{(2)}\widetilde{\nu}^{(A_o^{C_j})}, \theta_j^{(3)}\widetilde{\pi}^{(A_o^{C_j})}\}$, where $\theta_j = \{\theta_j^{(1)}, \theta_j^{(2)}, \theta_j^{(3)}\}$ are the parameters to control the degree of relativity of IVIFP. Then the exploration-based method based on edge topology is adopted to improve the efficiency of edge connection, and the breadth-first bounded search is used for edge matching instead of the shortest path query between two vertices as in [18]. Generally, MFC-GPM is a process of continuous edge matching and edge connection to obtain subgraphs matched by graph patterns.

Step 2: With the above obtained subgraphs and their $\widetilde{A}_l$ $(l \in [1, q])$, we can process the IVIFD.

Step 2.1: With Definition 6, IVIFS matrix for matched subgraphs $\{A\}$ with different contexts $\{C_j\}$ can be evaluated. Meanwhile, IVIFP with Definition 9, an improved IVIFCE with Definition 8, and a grey relation degree with Definition 10 can be obtained.

Step 2.2: According to the situation of weights $\{\omega_j\}$, we have the following three cases.

Case 1: If the weight ω_j of context C_j are known, from Definition 11, the comprehensive grey relation degree $\widetilde{\gamma}^*(\widetilde{A}_l)$ can be computed for all obtained subgraph $\widetilde{A}_l$ and $l \in [1, q]$, and then with these $\widetilde{\gamma}^*$ all obtained subgraphs can be ranked, i.e., the larger $\widetilde{\gamma}^*$, the better subgraph. Hence, the IVIFD can be conducted.

Case 2: If the weight ω_j of context C_j are unknown, with the average intuitionistic fuzzy information entropy $E(C_j)$ introduced in Definition 12, without loss of generality, the weight ω_j of context C_j can be evaluated by

$$\omega_j = \frac{1 - E(C_j)}{\sum_{j=1}^{m}(1 - E(C_j))}, \tag{22}$$

as the smaller entropy, the better. Then, it is possible to conduct the process as in Case 1, and make the IVIFD.

TABLE I
THE DETAIL INFORMATION OF THREE REAL-WORLD DATA GRAPH

Dataset	Vertices	Edges	Description
wiki-Vote	7,115	103,689	Wikipedia who-votes-on-whom network
Email-Enron	36,692	183,831	Email communication network from Enron
Epinions	75,879	508,837	Who-trusts-whom network of Epinions.com

Case 3: If the weight ω_j of context C_j are semi-known, i.e., the range of ω_j are known and $\omega_j \in W$, we can have the following l programming models.

$$\max \sum_{l=1}^{q}\sum_{j=1}^{m} \omega_j \widetilde{\gamma}(\widetilde{A}_l^{(C_j)}, \widetilde{o}^{(A_l^{C_j})}) \tag{23}$$

$$s.t. \begin{cases} \omega_j \in W \\ \sum_{j=1}^{m} \omega_j = 1 \\ \omega_j \geq 0 \\ j = 1, 2, \ldots, m \end{cases} \tag{24}$$

As they are linear programming problems, it is easy to get ω_j. Then, it is possible to conduct the process as in Case 1, and make the IVIFD.

VI. EXPERIMENTS

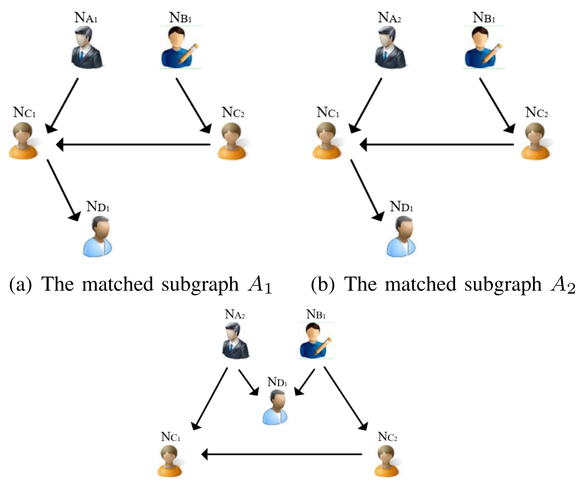
In this section, in addition to providing an illustrative example, we also conduct perturbation analysis and ablation experiments about our proposed method on three real-world social graphs, with the details shown in Table I. However, the adopted datasets contain only the network structure about vertices and edges. The contexts of the vertices and edges and their ratings mentioned in our proposed method can be mined and evaluated from the existing social network methods, which is another very challenging problem [25], but they are out of the scope of this work. In our experiments, without loss of generality, we randomly generate them with function *rand()* in *SQL*.

A. Illustrative Example

A technology company needs to find an expert team to complete a software development project from the job search social graph G_D in Fig. 1(b). This project is required to be composed of experts in at least four roles: project manager (PM), business analyst (BA), software architect (SA), and software developer (SD), and the expert team needs to conform to graph pattern G_P shown in Fig. 1(a). Data graph G_D stores contextual ratings about professional ability C_1, social trust C_2, and social intimacy C_3. The IVIFD process in big graph for this example can be illustrated as follow.

Step 1.1: In this step, as the DMs select the preferred vertices and edges in data graph G_D, here the preferred vertex set is $\{N_{A_2}, N_{C_1}, N_{B_1}\}$, and the preferred edge set is $\{(N_{A_1}, N_{C_1}), (N_{B_2}, N_{C_2}), (N_{C_1}, N_{D_1}), (N_{C_1}, N_{D_2})\}$. According to the contextual ratings of vertices and edges, the IVIFP of G_D can be calculated by Definitions 6 & 9 as

$$\widetilde{o}_1^{(G_D^{C_j})} = \{[0.067, 0.278], [0.047, 0.417], [0.583, 1.0]\}$$



(a) The matched subgraph A_1 (b) The matched subgraph A_2

(c) The matched subgraph A_3

Fig. 4. The matched subgraphs obtained by GPM.

TABLE II
IVIFS MATRIX

	C_1	C_2	C_3
A_1	$[(0.275, 0.417), (0.0, 0.0), (0.583, 0.45)]$	$[(0.171, 0.313), (0.035, 0.104), (0.583, 0.658)]$	$[(0.275, 0.417), (0.0, 0.0), (0.583, 0.45)]$
A_2	$[(0.171, 0.313), (0.035, 0.104), (0.583, 0.658)]$	$[(0.204, 0.417), (0.0, 0.0), (0.583, 0.592)]$	$[(0.275, 0.417), (0.0, 0.0), (0.583, 0.45)]$
A_3	$[(0.242, 0.313), (0.035, 0.104), (0.583, 0.516)]$	$[(0.137, 0.25), (0.057, 0.167), (0.583, 0.726)]$	$[(0.22, 0.333), (0.028, 0.083), (0.584, 0.56)]$

TABLE III
IVIFP MATRIX

	C_1	C_2	C_3
O_1	$[(0.133, 0.417), (0.0, 0.0), (0.583, 0.733)]$	$[(0.067, 0.208), (0.071, 0.417), (0.583, 1.0)]$	$[(0.133, 0.417), (0.0, 0.0), (0.583, 0.733)]$
O_2	$[(0.067, 0.278), (0.047, 0.417), (0.583, 1.0)]$	$[(0.133, 0.417), (0.0, 0.0), (0.583, 0.733)]$	$[(0.133, 0.417), (0.0, 0.0), (0.583, 0.733)]$
O_3	$[(0.067, 0.278), (0.047, 0.417), (0.583, 1.0)]$	$[(0.0, 0.0), (0.0, 0.0), (0.0, 0.0)]$	$[(0.0, 0.0), (0.0, 0.0), (0.0, 0.0)]$

$$\widetilde{o_2}^{(G_D^{C_j})} = \{[0.067, 0.417], [0.071, 0.208], [0.583, 0.867]\}$$

$$\widetilde{o_3}^{(G_D^{C_j})} = \{[0.067, 0.208], [0.071, 0.417], [0.583, 1.0]\}$$

In GPM, positive preference is usually taken into account rather than together with negative preference and uncertainty. Here without loss of generality, we have the IVIFP of G_D as follows.

$$\widetilde{o_1}^{(G_D^{C_j})} = \{[0.067, 0.278]\}$$

$$\widetilde{o_2}^{(G_D^{C_j})} = \{[0.067, 0.417]\}$$

$$\widetilde{o_3}^{(G_D^{C_j})} = \{[0.067, 0.208]\}$$

Step 1.2: During GPM, the constraints in graph pattern G_P with contexts C_1, C_2 and C_3 are $3.1\widetilde{o}^{(G_D^{C_1})}$, $3.8\widetilde{o}^{(G_D^{C_2})}$, $2.1\widetilde{o}^{(G_D^{C_3})}$, respectively. With MFC-GPM, the three matched subgraphs are shown in Fig. 4.

Step 2.1: With Definition 6, the IVIFS matrix of matched subgraphs for different contexts can be obtained respectively and shown in Table II. In addition, with Definition 9, IVIFP matrix of each subgraph is shown in Table III. Moreover, with Definition 8, the improved IVIFCE between the subjective evaluation value of each alternative and the subjective preference value of the DMs is calculated to form the distance matrix $\widetilde{CE}^*$:

$$\widetilde{CE}_{33}^* = \begin{bmatrix} 0.0020 & 0.0282 & 0.0020 \\ 0.0195 & 0.0005 & 0.0020 \\ 0.0260 & 0.0024 & 0.0179 \end{bmatrix}$$

Furthermore, with Definition 10, the grey relation degree between a subgraph and the IVIFP as follows:

$$\widetilde{\gamma}_{33} = \begin{bmatrix} 0.9059 & 0.3451 & 0.9059 \\ 0.4342 & 1.0 & 0.9059 \\ 0.3637 & 0.8847 & 0.4565 \end{bmatrix}$$

Step 2.2: Weight analysis.

Case 1: If the weight $\omega_j (j = 1, 2, 3)$ of context C_j are 0.20, 0.30, and 0.21 respectively, the comprehensive grey relation degree between the matched subgraphs and the IVIFP can be obtained according to Definition 11.

$$\widetilde{\gamma}_1^* = 0.475, \widetilde{\gamma}_2^* = 0.577, \widetilde{\gamma}_3^* = 0.434$$

Hence, we have $\widetilde{\gamma}_2^* > \widetilde{\gamma}_1^* > \widetilde{\gamma}_3^*$, and $A_2 > A_1 > A_3$, where a higher $\widetilde{\gamma}^*$ leads to more preferable from the viewpoint of IVIFP, that is, the expert team selected by the technology company is A_2.

Case 2: If the weight ω_j of context C_j are unknown, the average intuitionistic fuzzy information entropy $E(C_j)$ introduced in Definition 12 can be calculated as

$$E(C_1) = 0.755, E(C_2) = 0.321, E(C_3) = 0.55$$

Then the weight ω_j of C_j can be evaluated by (22)

$$\omega_1 = 0.178, \omega_2 = 0.494, \omega_3 = 0.328$$

Then, it is possible to conduct the process as in Case 1. Specifically, the comprehensive grey relation degree $\widetilde{\gamma}^*$ of each candidate is

$$\widetilde{\gamma}_1^* = 0.629, \widetilde{\gamma}_2^* = 0.868, \widetilde{\gamma}_3^* = 0.652$$

Hence, $\widetilde{\gamma}_2^* > \widetilde{\gamma}_3^* > \widetilde{\gamma}_1^*$, and $A_2 > A_3 > A_1$. Therefore, the expert team selected by the technology company is A_2.

Case 3: If the weight ω_j of context C_j are semi-known, such as $0.15 \leq \omega_1 \leq 0.23$, $0.20 \leq \omega_2 \leq 0.46$, and $0.18 \leq \omega_3 \leq 0.43$. By (23) & (24), it is easy to construct a mathematical programming model with the goal of maximizing the comprehensive grey relation degree, and then obtain the weight ω_j

$$\omega_1 = 0.15, \omega_2 = 0.42, \omega_3 = 0.43.$$

Then, it is possible to conduct the process as in Case 1, and we can obtain the comprehensive grey relation degree

$$\widetilde{\gamma}_1^* = 0.67, \widetilde{\gamma}_2^* = 0.875, \widetilde{\gamma}_3^* = 0.622,$$

which leads to $\widetilde{\gamma}_2^* > \widetilde{\gamma}_1^* > \widetilde{\gamma}_3^*$. So the expert team selected by the technology company is A_2.

Without considering the contextual constraints in graph pattern A, there are 9 matched subgraphs. After IVIFD, the expert team selected by the technology company is A_2, which is the optimal alternative. With the proper constraints, the non-optimal subgraphs will be pruned, and this will effectively reduce the searching cost and remain the optimal alternative.

TABLE IV
THE DECISION-MAKING RESULTS UNDER DIFFERENT GREY RESOLUTIONS ON EPINIONS

Rank	$\xi = 0.05$		$\xi = 0.20$		$\xi = 0.35$		$\xi = 0.50$		$\xi = 0.65$		$\xi = 0.80$		$\xi = 0.95$	
	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives
1	0.9462	A_{1013}	0.9781	A_{1013}	0.9862	A_{1013}	0.9899	A_{1013}	0.9921	A_{1013}	0.9935	A_{1013}	0.9944	A_{1013}
2	0.9309	A_{1191}	0.9721	A_{359}	0.9838	A_{359}	0.9886	A_{359}	0.9912	A_{359}	0.9928	A_{359}	0.9939	A_{359}
3	0.9282	A_{1374}	0.9646	A_{1374}	0.9762	A_{1374}	0.9820	A_{1374}	0.9856	A_{1374}	0.9880	A_{1374}	0.9897	A_{639}
4	0.9265	A_{1129}	0.9582	A_{1191}	0.9731	A_{639}	0.9809	A_{639}	0.9851	A_{639}	0.9879	A_{639}	0.9896	A_{1374}
5	0.9240	A_{670}	0.9571	A_{1129}	0.9708	A_{1069}	0.9788	A_{1069}	0.9834	A_{1069}	0.9863	A_{1069}	0.9884	A_{1069}
6	0.9213	A_{1176}	0.9547	A_{639}	0.9700	A_{1191}	0.9766	A_{1191}	0.9810	A_{1417}	0.9843	A_{1417}	0.9867	A_{1417}
7	0.9191	A_{1358}	0.9533	A_{670}	0.9694	A_{1129}	0.9762	A_{1129}	0.9809	A_{1191}	0.9842	A_{1043}	0.9866	A_{1043}
8	0.9029	A_{601}	0.9527	A_{1069}	0.9667	A_{952}	0.9758	A_{1417}	0.9808	A_{1043}	0.9838	A_{1191}	0.9859	A_{1191}
9	0.9013	A_{999}	0.9500	A_{1176}	0.9665	A_{1417}	0.9756	A_{1043}	0.9805	A_{1129}	0.9835	A_{952}	0.9859	A_{952}
10	0.8998	A_{359}	0.9492	A_{952}	0.9664	A_{1043}	0.9752	A_{952}	0.9802	A_{952}	0.9835	A_{1129}	0.9857	A_{1129}

TABLE V
THE DECISION-MAKING RESULTS UNDER DIFFERENT GREY RESOLUTIONS ON WIKI-VOTE

Rank	$\xi = 0.05$		$\xi = 0.20$		$\xi = 0.35$		$\xi = 0.50$		$\xi = 0.65$		$\xi = 0.80$		$\xi = 0.95$	
	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives
1	0.5603	A_{27}	0.7020	A_{27}	0.7704	A_{27}	0.8123	A_{27}	0.8408	A_{27}	0.8616	A_{27}	0.8775	A_{27}
2	0.5288	A_{23}	0.6437	A_{8}	0.7232	A_{8}	0.7729	A_{8}	0.8070	A_{8}	0.8320	A_{8}	0.8511	A_{8}
3	0.4914	A_{8}	0.6343	A_{23}	0.6991	A_{23}	0.7579	A_{1}	0.7978	A_{1}	0.8259	A_{1}	0.8468	A_{1}
4	0.2924	A_{1}	0.5827	A_{1}	0.6958	A_{1}	0.7440	A_{23}	0.7770	A_{23}	0.8024	A_{23}	0.8240	A_{30}
5	0.2488	A_{30}	0.5335	A_{30}	0.6547	A_{30}	0.7235	A_{30}	0.7683	A_{30}	0.8002	A_{30}	0.8226	A_{23}
6	0.2366	A_{5}	0.5149	A_{19}	0.6409	A_{3}	0.7135	A_{3}	0.7611	A_{3}	0.7948	A_{3}	0.8200	A_{3}
7	0.2335	A_{19}	0.5146	A_{3}	0.6388	A_{19}	0.7098	A_{19}	0.7565	A_{19}	0.7897	A_{19}	0.8147	A_{19}
8	0.2322	A_{18}	0.5112	A_{18}	0.6338	A_{18}	0.7042	A_{18}	0.7505	A_{18}	0.7836	A_{18}	0.8086	A_{18}
9	0.2310	A_{3}	0.4817	A_{7}	0.6101	A_{7}	0.6861	A_{7}	0.7366	A_{7}	0.7728	A_{7}	0.8001	A_{7}
10	0.2086	A_{7}	0.4771	A_{17}	0.6043	A_{17}	0.6796	A_{17}	0.7300	A_{17}	0.7662	A_{17}	0.7935	A_{17}

TABLE VI
THE DECISION-MAKING RESULTS UNDER DIFFERENT GREY RESOLUTIONS ON EMAIL-ENRON

Rank	$\xi = 0.05$		$\xi = 0.20$		$\xi = 0.35$		$\xi = 0.50$		$\xi = 0.65$		$\xi = 0.80$		$\xi = 0.95$	
	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives	γ_i	alternaltives
1	0.9603	A_{893}	0.9897	A_{893}	0.9941	A_{893}	0.9959	A_{893}	0.9968	A_{893}	0.9974	A_{893}	0.9978	A_{893}
2	0.9362	A_{1004}	0.9830	A_{1004}	0.9902	A_{1004}	0.9931	A_{1004}	0.9947	A_{1004}	0.9957	A_{1004}	0.9964	A_{1004}
3	0.9336	A_{688}	0.9813	A_{688}	0.9891	A_{688}	0.9923	A_{688}	0.9941	A_{688}	0.9952	A_{688}	0.9959	A_{688}
4	0.9141	A_{1018}	0.9723	A_{1018}	0.9835	A_{1018}	0.9882	A_{1018}	0.9908	A_{1018}	0.9925	A_{1018}	0.9937	A_{1018}
5	0.8907	A_{269}	0.9662	A_{269}	0.9800	A_{269}	0.9858	A_{269}	0.9889	A_{269}	0.9910	A_{269}	0.9924	A_{269}
6	0.8907	A_{270}	0.9662	A_{270}	0.9800	A_{270}	0.9858	A_{270}	0.9889	A_{270}	0.9910	A_{270}	0.9924	A_{270}
7	0.8907	A_{273}	0.9662	A_{273}	0.9800	A_{273}	0.9858	A_{273}	0.9889	A_{273}	0.9910	A_{273}	0.9924	A_{273}
8	0.8715	A_{639}	0.9609	A_{639}	0.9770	A_{639}	0.9836	A_{639}	0.9873	A_{639}	0.9896	A_{639}	0.9912	A_{639}
9	0.8692	A_{114}	0.9596	A_{937}	0.9765	A_{937}	0.9834	A_{937}	0.9872	A_{937}	0.9896	A_{937}	0.9912	A_{937}
10	0.8690	A_{214}	0.9580	A_{185}	0.9753	A_{185}	0.9825	A_{185}	0.9864	A_{185}	0.9890	A_{185}	0.9906	A_{185}

B. *Perturbation Analysis*

In order to study the stability of our proposed IVIFD in big graph in this paper and validate its feasibility, we perform perturbation analysis on the three data sets according to the changes of the grey distinguishing parameter ξ introduced in the grey relation degree in (19), such as taking $\xi = 0.05$, 0.20, 0.35, 0.50, 0.65, 0.80, 0.95, respectively.

As shown in Tables IV & V, different colors represent different subgraphs, where the subgraphs with cool-color rank higher, and the subgraphs with warm-color rank lower. With the perturbation analysis results of Epinions and Wiki-Vote, the ranking of some subgraphs fluctuates, but generally they show a certain degree of stability. More specifically, the subgraphs with the largest grey relation degree are always A_{1013} and A_{27} respectively, indicating that the slight fluctuation of the

TABLE VII
COMPARISON OF P-METHOD AND C-METHOD UNDER DIFFERENT ATTRIBUTES

Attribute	Epinions		Wiki-Vote		Email-Enron	
	P-Method	**C-Method**	P-Method	**C-Method**	P-Method	**C-Method**
Times (min)	39.40	**8.32**	73.93	**17.09**	47.61	**7.40**
Matched subgraphs	18,395	**60**	23,221	**615**	25,400	**199**
Optimal subgraph	A_{43}	A_{43}	A_{513}	A_{513}	A_{181}	A_{181}

ranking does not affect the decision making. In contrast, Table VI shows that the ranking result does not change with the grey distinguishing parameter ξ, which represents a certain degree of stability. As the grey relation degree calculated by IVIFCE is relatively low sensitive to the grey distinguishing parameter, the decision result is relatively stable, and the weight setting is relatively reasonable.

In conclusion, the IVIFD in big graph constructed in this paper is stable and reliable.

C. Ablation Study

In this section, experiments on three real-world social graphs are conducted to evaluate (1) the effectiveness of our proposed IVIFD in big graph; and (2) the superiority of our method in improving the efficiency of decision making. Specifically, we mainly study the step of determining constraints in graph pattern with contexts based on the preferred vertices and edges selected by DMs.

First, the effectiveness and efficiency of our method can be verified by pruning the step of determining the context constraints. Specifically, the pruned method (P-Method) is compared with the complete method (C-Method) on the running time, the number of matched subgraphs and the optimal subgraph, respectively. The comparison results are shown in Table VII. (1) The running time of P-Method on the datasets Epinions, Wiki-Vote and Email-Enron is 4.74, 4.33 and 6.43 times that of C-Method, respectively. This shows that our method (C-Method) greatly reduces the running time and improves time efficiency. (2) Judging from the number of subgraphs obtained by GPM, the number of subgraphs obtained by C-Method is 0.33%, 2.65%, and 0.78% of P-Method respectively. This shows that our method largely prunes the matched results. (3) On the three datasets, the optimal subgraphs of P-Method and C-method IVIFD are A_{43}, A_{513}, and A_{181}, respectively. This shows that our method does not prune the optimal solution. This proves that our method can preserve the optimal subgraph while pruning the invalid subgraphs, greatly reducing our decision making calculations, and improving time efficiency while ensuring effectiveness.

Then, the effectiveness and efficiency of our method can be verified by changing the parameters θ_j $(j \in [1, k])$ in the step of determining the context constraints. In the grey relation analysis experiment, we have $\theta_1 = 3.1$, $\theta_2 = 3.8$, and $\theta_3 = 2.1$. In the illustrative example, the validity of the parameters can be verified, that is, invalid subgraphs can be pruned. In order to further illustrate the effectiveness of parameter settings, we set

TABLE VIII
VARIABLES ADOPTED IN THE PAPER

Symbol	Description
A^{C_j}	graph pattern A with context C_j
r_i	rating on vertices and edges in A^{C_j}
k	the number of context C_j $(1 \leq j \leq k)$
q	the number of graph pattern $\widetilde{A_l}(l \in [1, q])$
$P_i^{(A^{C_j})}$	the probability to obtain the rating r_i on vertices and edges in A^{C_j}
$x_i^{(A^{C_j})}$	the number of occurrences of r_i on vertices and edges in A^{C_j}
$X^{(A^{C_j})}$	the occurrences of all the ratings on context C_j in subgraph A
$\widetilde{\mu}^{(A^{C_j})}$	positive preference
$\widetilde{\nu}^{(A^{C_j})}$	negative preference
$\widetilde{\pi}^{(A^{C_j})}$	uncertainty preference
$\Phi^{(A^{C_j})}$	the certainty of preference on context C_j in subgraph A
$\widetilde{A}^{(C_j)}$	the intuitionistic fuzzy set for the preference on vertices and edges in A^{C_j}, $\widetilde{A}^{(C_j)} = \{< \widetilde{\mu}^{(A^{C_j})}, \widetilde{\nu}^{(A^{C_j})}, \widetilde{\pi}^{(A^{C_j})} >\}$
$\widehat{A}^{(C_j)}$	the preferred vertices or edges in A^{C_j}
$\widetilde{CE}$	the interval-valued intuitionistic fuzzy cross entropy
$\widetilde{CE}^*$	the improved interval-valued intuitionistic fuzzy cross entropy
$\widetilde{\partial}^{(A^{C_j})}$	the interval-valued intuitionistic fuzzy preference
$\widetilde{\gamma}$	grey relation degree
$\widetilde{\gamma}^*$	comprehensive grey relation degree
$E(C_j)$	the average intuitionistic fuzzy information entropy
ω_j	the weight of context C_j

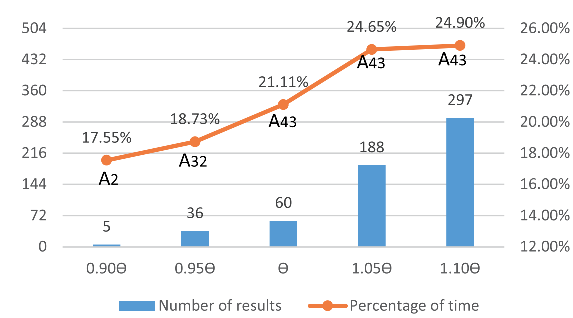


Fig. 5. IVIFD results under different θ on Epinions.

$\theta_j = 0.90\theta_j, 0.95\theta_j, \theta_j, 1.05\theta_j, 1.10\theta_j$, respectively, to conduct experiments on three datasets.

As shown in Figs. 5–7, (1) Time efficiency has an inverse relationship with decision quality. When taking 0.90θ and 0.95θ, the matched subgraphs are less than that with θ, where θ is the initial value set by our method. However, when 0.90θ and 0.95θ are taken, the optimal subgraph cannot be obtained. This is because the reduction of the constraint interval causes the optimal subgraph to be pruned. (2) On the datasets Epinions and Email-Enron, both 1.05θ and 1.10θ can determine the correct optimal subgraph, but the running time is between 3% and 9% longer than that with θ. (3) On the dataset Wiki-Vote, 1.10θ cannot obtain the optimal subgraph. This is because when the left and right intervals expand at the same time, there is a situation where the interval shrinks at the end, resulting in the optimal

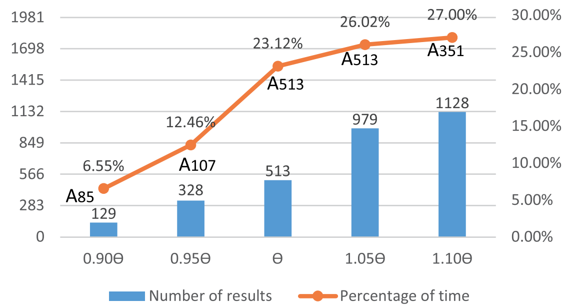


Fig. 6. IVIFD results under different θ on Wiki-Vote.

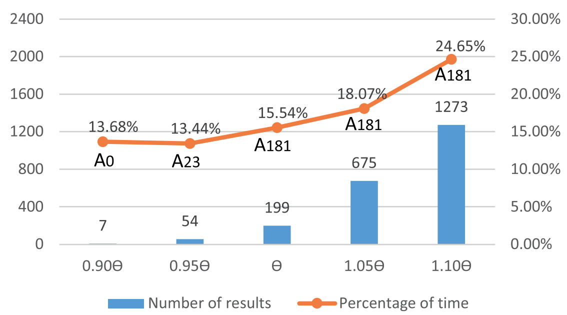


Fig. 7. IVIFD results under different θ on Email-Enron.

subgraph being pruned. In conclusion, our method can make the optimal decision in the least time.

VII. Conclusion

Our proposed Interval-Valued Intuitionistic Fuzzy Decision (IVIFD) in Big Graph mainly consists of two steps, which are MFC-GPM and IVIFD. As traditionally IVIFD can maximally reduce the uncertainty of decision making and it is only valid for small datasets, which makes it impossible to be applied in big graph. In this paper, MFC-GPM is adopted to prune the searching space, which makes it possible to process IVIFD later. The perturbation analysis and ablation experiments of three large-scale real-world social graphs show the stability and efficiency of our IVIFD.

However, the risk preference factors of DMs may also lead to the problem of conflict between decision results and DMs' intuitive judgments, which will be discussed in future. In addition, applying our IVIFD to more scenarios is also an interesting issue that deserves to be studied.

References

[1] K. T. Atanassov, "Intuitionistic fuzzy sets," in *Fuzzy Sets and Systems*, Amsterdam, The Netherlands: Elsevier, vol. 20, 1986, pp. 87–96.

[2] K. T. Atanassov and G. Gargov, "Interval-valued intuitionistic fuzzy sets," in *Fuzzy Sets and Systems*, vol. 31, Amsterdam, The Netherlands: Elsevier, 1989, pp. 343–349.

[3] J. Cheng, J. X. Yu, B. Ding, P. S. Yu, and H. Wang, "Fast graph pattern matching," in *Proc. 24th Int. Conf. Data Eng.*, Cancún, Mexico, 2008, pp. 913–922.

[4] W. Fan, J. Li, J. Luo, Z. Tan, X. Wang, and Y. Wu, "Incremental graph pattern matching," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Athens, Greece, 2011, pp. 925–936.

[5] W. Fan, J. Li, S. Ma, N. Tang, Y. Wu, and Y. Wu, "Graph pattern matching: From intractable to polynomial time," *Proc. VLDB Endowment*, vol. 3, no. 1, pp. 264–275, 2010.

[6] W. Fan, X. Wang, and Y. Wu, "ExpFinder: Finding experts by graph pattern matching," in *Proc. 29th IEEE Int. Conf. Data Eng.*, Brisbane, Australia, 2011, pp. 1316–1319.

[7] Z. Gong, X. Tan, and Y. Yang, "Optimal weighting models based on linear uncertain constraints in intuitionistic fuzzy preference relations," *J. Oper. Res. Soc.*, vol. 70, no. 8, pp. 1296–1307, 2019.

[8] M. S. Hamada, A. G. Wilson, C. S. Reese, and H. F. Martz, *Bayesian Rel.* New York, NY, USA: Springer, 2008.

[9] M. R. Henzinger, T. A. Henzinger, and P. W. Kopke, "Computing simulations on finite and infinite graphs," in *Proc. 36th Annu. Symp. Found. Comput. Sci.*, Milwaukee, WI, USA, 1995, pp. 453–462.

[10] W. W. Hines, D. C. Montgomery, D. M. Goldsman, and C. M. Borror, *Probability and Statistics in Engineering*. New York, NY, USA: Wiley, 2003.

[11] J. Hu and A. L. Ferguson, "Global graph matching using diffusion maps," *Intell. Data Anal.*, vol. 20, no. 3, pp. 637–654, 2016.

[12] L. Li and Y. Wang, "Subjective trust inference in composite services," in *Proc. 20th AAAI Conf. Artif. Intell.*, Atlanta, Georgia, USA, 2010, pp. 1377–1384.

[13] L. Li, X. Wu, H. Chen, C. Zhou, G. Liu, and Y. Jiang, "Weighted partial order oriented three-way decisions under score-based common voting rules," *Int. J. Approx. Reasoning*, vol. 123, pp. 41–54, 2020.

[14] Y. Li, Y. Cheng, Q. Mou, and S. Xian, "Novel cross-entropy based on multi-attribute group decision-making with unknown experts' weights under interval-valued intuitionistic fuzzy environment," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 1295–1304, 2020.

[15] H. Liao, Z. Xu, X. Zeng, and J. M. Merigó, "Framework of group decision making with intuitionistic fuzzy preference information," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 1211–1227, Aug. 2015.

[16] J. Ye, "Multicriteria fuzzy decision-making method based on a novel accuracy function under interval-valued intuitionistic fuzzy environment," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6899–6902, 2009.

[17] X. Ji, W. Lei, and H. Xue, "Interval intuitionistic fuzzy decision model with abnormal information and its application in talent selection," *Math. Problems Eng.*, vol. 116, pp. 1–10, 2021.

[18] G. Liu, L. Li, and X. Wu, "Multi-fuzzy-constrained graph pattern matching with big graph data," *Intell. Data Anal.*, vol. 24, no. 4, pp. 941–958, 2020.

[19] G. Liu *et al.*, "MCS-GPM: Multi-constrained simulation based graph pattern matching in contextual social graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1050–1064, Jun. 2018.

[20] G. Liu, Y. Wang, and M. A. Orgun, "Quality of trust for social trust path selection in complex social networks," in *Proc. 9th Int. Conf. Auton. Agents Multiagent Syst.*, Toronto, Canada, 2010, pp. 1575–1576.

[21] G. Liu, Y. Wang, and M. A. Orgun, "Trust transitivity in complex social networks," in *Proc. 20th AAAI Conf. Artif. Intell.*, San Francisco, California, USA, 2011, pp. 1222–1229.

[22] G. Liu *et al.*, "Multi-constrained graph pattern matching in large-scale contextual social graphs," in *31st IEEE Int. Conf. Data Eng.*, Seoul, South Korea, 2015, pp. 351–362.

[23] G. Liu, Y. Wang, B. Zheng, Z. Li, and K. Zheng, "Strong social graph based trust-oriented graph pattern matching with multiple constraints," in *Proc. 31st IEEE Trans. Emerg, Top. Comput. Intell.*, 2020, pp. 1–11.

[24] S. Ma, Y. Cao, W. Fan, J. Huai, and T. Wo, "Strong simulation: Capturing topology in graph pattern matching," *ACM Trans. Database Syst.*, vol. 39, no. 1, pp. 1–46, 2014.

[25] J. Mao, W. Tian, Y. Yang, and J. Liu, "An efficient social attribute inference scheme based on social links and attribute relevance," *IEEE Access*, vol. 7, pp. 153074–153085, 2019.

[26] V. L. G. Nayagam, S. Jeevaraj, and P. Dhanasekaran, "An intuitionistic fuzzy multi-criteria decision-making method based on non-hesitance score for interval-valued intuitionistic fuzzy sets," *Soft Comput.*, vol. 21, no. 23, pp. 7077–7082, 2017.

[27] A. Shemshadi, Q. Z. Sheng, and Y. Qin, "Efficient pattern matching for graphs with multi-labeled nodes," *Knowl. Based Syst.*, vol. 109, pp. 256–265, 2016.

[28] C. Song, T. Ge, C. X. Chen, and J. Wang, "Event pattern matching over graph streams," *Proc. VLDB Endowment*, vol. 8, no. 4, pp. 413–424, 2014.

[29] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "ArnetMiner: Extraction and mining of academic social networks," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2008, pp. 990–998.

[30] H. Tong, C. Faloutsos, B. Gallagher, and T. Eliassi-Rad, "Fast best-effort pattern matching in large attributed graphs," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, San Jose, California, USA, pp. 737–746, 2007.

[31] V. Torra, "Hesitant fuzzy sets," *Int. J. Intell. Syst.*, vol. 25, no. 6, pp. 529–539, 2010.

[32] S. Wan, F. Wang, and J. Dong, "A three-phase method for group decision making with interval-valued intuitionistic fuzzy preference relations," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 998–1010, Apr. 2018.

[33] F. Wang and S. Wan, "A comprehensive group decision-making method with interval-valued intuitionistic fuzzy preference relations," *Soft. Comput.*, vol. 25, no. 1, pp. 343–362, 2021.

[34] Y. Wang and M. P. Singh, "Evidence-based trust: A mathematical model geared for multiagent systems," *ACM Trans. Auton. Adapt. Syst.*, vol. 5, no. 4, pp. 1–28, 2010.

[35] L. Pang, K. Tay, C. Lim, and H. Ishibuchi, "A new monotone fuzzy rule relabeling framework with application to failure mode and effect analysis methodology," *IEEE Access*, vol. 8, pp. 144908–144930, 2020.

[36] Y. Yao, "Three-way decision: An interpretation of rules in rough set theory," in *Proc. 4th Int. Conf. Rough Sets Knowl. Technol.*, Gold Coast, Australia, 2009, pp. 642–649.

[37] J. Ye, "Fuzzy cross entropy of interval-valued intuitionistic fuzzy sets and its optimal decision-making method based on the weights of alternatives," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 6179–6183, 2011.

[38] S. Yoo, Y. Yang, F. Lin, and I. Moon, "Mining social networks for personalized email prioritization," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, Paris, France, 2009, pp. 967–976.

[39] L. A. Zadeh, "Fuzzy sets," *Inf. Control.*, vol. 8, no. 3, pp. 338–353, 1965.

[40] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning - I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.

Lei Li (Senior Member, IEEE) received the bachelor's degree in information and computational science from Jilin University, Changchun, China, in 2004, the master's degree in applied mathematics from the Memorial University of Newfoundland, St. John's, NL, Canada, in 2006, and the Ph.D. degree in computing from Macquarie University, Sydney, NSW, Australia, in 2012.

He is currently an Associate Professor of computer science and technology with the Key Laboratory of Knowledge Engineering with Big Data of Ministry of Education, Hefei University of Technology, Hefei, China, with Intelligent Interconnected Systems Laboratory of Anhui Province, Hefei University of Technology, and with the School of Computer Science and Information Engineering, Hefei University of Technology. His research interests include data mining, social computing, and graph computing.

Lan Jiang received the bachelor's degree from Hefei Normal University, Hefei, China, in 2020. She is currently working toward the master's degree with the Key Laboratory of Knowledge Engineering with Big Data of Ministry of Education, Hefei University of Technology, Hefei, China, Intelligent Interconnected Systems Laboratory of Anhui Province, Hefei University of Technology, and School of Computer Science and Information Engineering, Hefei University of Technology. Her current research interests include data mining and graph computing.

Chenyang Bu (Member, IEEE) received the B.E. degree from the Hefei University of Technology, Hefei, China, in 2012, and the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2017.

He is currently an Assistant Professor with the Hefei University of Technology. His main research interests include evolutionary algorithms and their applications in areas, such as knowledge graphs, educational data mining, and power systems. He is a reviewer of several international journals, including IEEE TEVC, IEEE TNNLS, IEEE TCYB, IEEE TETCI and IEEE TII.

Yi Zhu received the bachelor's degree in electronic science and technology from Anhui University, Hefei, China, in 2006, the master's degree in pattern recognition and intelligence systems from the University of Science and Technology of China, Hefei, China, in 2012, and the Ph.D. degree in software engineering from the Hefei University of Technology, Hefei, China, in 2018.

He is currently an Assistant Professor with the School of information Engineering, Yangzhou University, Yangzhou, China. His research interests include data mining, knowledge engineering, and recommendation systems.

Xindong Wu (Fellow, IEEE) received the Ph.D. degree in artificial intelligence from the University of Edinburgh, Edinburgh, U.K. He is the Chief Scientist with the Mininglamp Academy of Sciences, Mininglamp Technologies, Beijing, China, and a Chang Jiang Scholar with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China. His research interests include data mining, knowledge-based systems, and web information exploration. He is the Steering Committee Chair of the IEEE International Conference on Data Mining (ICDM), the Editor-in-Chief of the *Knowledge and Information Systems*, and Editor-in-Chief of the *Springer book series, Advanced Information and Knowledge Processing* (AIKP). He is Fellow of AAAS.