Multi-fuzzy-objective graph pattern matching in big graph environments with reliability, trust and social relationship

Lei Li [1] · Fang Zhang [1] · Zan Zhang [1] · Peipei Li [1] · Chenyang Bu [1]

Abstract

With the advent of the era of big data, the scale of data has grown dramatically, and there is a close correlation between massive multi-source heterogeneous data, which can be visually depicted by a big graph. Big graph, especially from Web data, social networks, or biometric data, has attracted more and more attention from researchers, which usually contains complex relationships and multiple attributes. How to perform efficient query and matching on big graph data is the basic problem on analyzing big graph. Using multi-constrained graph pattern matching, we can design patterns that meet our specific requirements, and find matched subgraphs to locate the required patterns to accomplish specific tasks. So how to find matched subgraphs with good attributes in big graph becomes the key problem on big graph research. Considering the possibility that a node in a subgraph may fail due to reliability, in order to select more and better matched subgraphs, in this paper, we introduce fuzziness and reliability into multi-objective graph pattern matching, and then use a multi-objective genetic algorithm NSGA-II to find the subgraphs with higher reliability and better attributes including social trust and social relationship. Finally, a reliability-based multi-fuzzy-objective graph pattern matching method (named as RMFO-GPM) is proposed. The experimental results on real data sets show the effectiveness of the proposed RMFO-GPM method comparing with other state-of-art methods.

Keywords Reliability · Graph Pattern Matching · Multi-Fuzzy-Objective · NSGA-II Algorithm

---

This article belongs to the Topical Collention: *Special Issue on Trust, Privacy, and Security in Crowdsourcing Computing*
Guest Editors: An Liu, Guanfeng Liu, Mehmet A. Orgun, and Qing Li

✉ Zan Zhang
zhangzan99@163.com

1   Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology) and School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, China

1 Introduction

With the rapid development of the Internet and information technology, the data scale has grown rapidly, and the types of data have been increasing as well [36]. Meanwhile, there is a close correlation between massive multi-source heterogeneous data, and the relationship between entities in different fields is usually complicated, which can be visually depicted by the graph, such as social networks [32], Web data, or biological data [15]. In these big data, there are not only the information about each data entity (i.e., the nodes in the graph), but also the related information among the data entities (i.e., the edges connected between the nodes in the graph) [20]. We can call these complex and large-scale data as *Big Graph* data. The relevant theories and technologies on big graph can be used for the detection of the relationship between various groups and people [15], the detection of traffic accidents in road networks, task assignment in crowdsourcing [21, 33, 34], or the detection of proteins in biological data. Graph pattern matching, as an important method for efficient query on big graph, is widely used in the above mentioned various fields [1, 35]. Specifically, how to obtain accurate and efficient matching queries in big graph has attracted more and more attention from researchers [22, 23].

In social network analysis, entities and relationships between entities can be represented in the form of big graphs [18]. The nodes in the graph represent participants, and the edges represent social relationships between them, which can be subordinates, domestic affection, friendship, or colleagues [16]. Querying a group or a person with a specific relationship on social graph can be converted to a graph pattern matching problem in which a matched subgraph or a specific node is located in a big graph [17], such as expert finding [5], travel planning [30], important role detecting [2], or learning group selecting, etc. For example, in social security analysis, an application system that exploits the law of crime and assists in solving crimes [29] can be developed through certain graph pattern matching based technology. Meanwhile, graph matching technology can be used to mine the shopping behavior of online shopping users and find out the most favorite shopping patterns of different users. Fan et al., [9] can recommend the content by users who share the same hobbies based on the mining of the big graph relationship on YouTube. However, due to the NP-complete time complexity, it is hard to apply graph pattern matching in big graph environments. In a specific field, how to find a matched subgraph that meets their specific requirements and has a high reliability in a big graph environment efficiently and effectively becomes a key issue.

The main contributions of this paper are summarized as follows:

1. In the traditional multi-constrained graph pattern matching, there may be a case in the data graph, where a certain attribute value of a subgraph is slightly lower than the given threshold value, but other attribute values are significantly better than the given threshold values; the corresponding graph may be a better subgraph, but it is not selected by the traditional multi-constrained pattern matching. Therefore, we introduce fuzzy numbers to include these good subgraphs.
2. In the traditional multi-constrained graph pattern matching, there may be a failure of a node in the matched subgraph which leads to the failure of the whole matched subgraph to accomplish a specific task. Therefore, we introduce two attributes including the total times of experiments and the times of successes to each node in the graph to evaluate the probability of a subgraph working normally, and the theory of reliability is introduced to evaluate the reliability of a matched subgraph.

3. In big graph environments, there may be thousands of matched subgraphs for a given graph pattern. How to quickly select multiple targets from many matched subgraphs which is a better matched subgraph cannot be solved by a simple single objective optimization algorithm. In this paper, multi-objective genetic algorithm NSGA-II is used for multi-objective optimization, and the better subgraphs in matched subgraphs can be selected. Specifically, a reliability-based multi-fuzzy-objective graph pattern matching method (named as RMFO-GPM) is proposed. The experimental results on real data sets show the effectiveness of the proposed RMFO-GPM method compared with other existing methods.

The rest structure of this paper is as follows. We describe the progress of graph pattern matching and multi-objective genetic algorithm NSGA-II in section 2, and then introduce the main ideas and steps of the reliability-based pattern matching algorithm RMFO-GPM for multi-fuzzy-objective graphs in section 3. The experiments on real data sets show the effectiveness of the proposed RMFO-GPM method in section 4. Finally, we summarize our proposed work in section 5.

2 Related work

2.1 Graph pattern matching

This section introduces the current research status of graph pattern matching, more specifically two most important research points including isomorphism-based graph pattern matching and simulation-based graph pattern matching.

(1) Isomorphism-based graph pattern matching: Isomorphic matching requires a bijective function between a pattern graph and a data graph. The isomorphic matching requires that the matched subgraph's topology is exactly the same as the pattern graph. This matching method is mainly used in the application of graph data with strict structural requirements, such as network abnormal behavior detection [6] or protein-molecule interaction.

Tong et al., [30] proposed a fast G-Ray method for finding matched subgraphs, which introduced a goodness score function to measure the matching degree between the obtained subgraphs and query patterns, and returned the optimal K subgraphs in all matched subgraphs. Cheng et al., [4] proposed that the graph pattern matching should be regarded as accessibility query in graph data table. Based on the join index of clustered graph, an R-join (where R stood for accessibility) algorithm including filtering and acquisition steps was proposed. Then an optimization method was proposed to optimize R-join/R-semijoins sequence, which improved the efficiency of graph pattern matching. In addition, Cheng et al., [5] proposed a method to construct a timely ranking list based on the spanning tree of the cyclic query graph, and use the timely ranking list to answer the multidimensional representation of the Top-K matched subgraphs. With this representation, a cost model was proposed to estimate the minimum number of spanning trees consumed in each ranking list for a given Top-K matching subgraph query. Then, the matched subgraphs were sorted based on the number of spanning trees, and the optimal K matched subgraphs can be returned. However, this graph indexing method was inefficient for processing large-scale graph queries. Therefore, Sun et al., [28] proposed an

efficient graph exploration and large-scale parallel computing to improve the efficiency of graph matching, which can be validated by experiments efficiently on Web data graphs.

Isomorphism-based graph pattern matching is very popular in many applications, such as 3D object matching or protein structure matching. Indexing, parallel and distributed methods are effective ways to improve the efficiency of graph pattern matching. However, such graph pattern matching still suffers from high computational costs because it is NP-complete. Isomorphism-based graph pattern matching is too harsh for the applications of social networks which do not require strict matching accuracy. Therefore, many scholars have turned to study simulation-based graph pattern matching.

(2)   Simulation-based pattern matching: Henzinger [11] proposed the concept of graph simulation for the first time, which required that the nodes in the matched subgraph maintained the same succession relationship with the corresponding nodes in the pattern graph. Fan et al., [8] extended the graph simulation by proposing the bounded simulation, in which the label of each node of the graph was not unique, and the bounded length can be specified on the side of the query graph. In bounded simulation, it is unnecessary to match each node and each edge exactly like isomorphic matching, but to match any node with the same label and a path whose length is not greater than the bounded length in the data graph, that is, a path in the data graph is matched with an edge in the pattern graph, and the length of the path is less than or equal to the boundary length defined by the corresponding edge in the pattern graph. Fan et al., [10] proposed a method to find Top-K matches for specific node patterns, in which some patterns containing important nodes and edges had high priority for matching in query graphs. In order to solve the problem that the topological structure of the matched subgraph and the topology of the pattern graph were different in bounded simulation, Ma et al., [24] proposed a strong simulation method based on the bounded simulation. Strong simulation required matching the nodes of the subgraph with the corresponding nodes in the pattern graph to be searched to maintain the same predecessor relationship and the radius of the matched nodes in the matched subgraph was less than or equal to the radius of the pattern graph. Considering that the graph would contain attribute information of both nodes and edges, Liu et al., [19] further proposed the multi-constrained simulation, which was an extension of bounded simulation. In this model, multiple attribute information can be defined on the edge and the node of data graph, and the aggregated attribute information of each path in data graph should be larger than the minimum value of given attribute information of nodes and edges in pattern graph while satisfying bounded simulation. In order to add attribute information of nodes and edges to graph pattern matching, a graph pattern matching method based on multi-constrained simulation was proposed. In order to solve the efficiency and validity of multi-constrained Top-K graph pattern matching in large-scale social graphs, Shi et al., [26] proposed an index HB-Tree, which can index the label and degree of nodes in data graphs and effectively obtain candidates for designated query nodes v_0. Then, a multi-constrained Top-K graph pattern matching method named MTK was proposed, which can effectively and efficiently recognize Top-K matching of v_0.

Simulated matching is more flexible than isomorphic matching, and it can find more useful matched subgraphs. Simulated matching is mainly used to detect the relationship between groups (e.g., drug-related trading network), which focuses on the application analysis of graph data mining on the relationship between nodes.

2.2 NSGA-II genetic algorithm

NSGA-II algorithm [7] is improved by Kalyanmoy Deb on the basis of NSGA [27], which is a genetic algorithm based on non-dominated sorting, and can be used to solve multi-objective optimization problems.

In order to overcome the shortcomings of high computational complexity, including the lack of elite mechanism and the requirement to specify shared parameters in NSGA, NSGA-II mainly improves the following three aspects:

(1) A fast non-dominated sorting algorithm is proposed to reduce the time cost of the algorithm from $O(MN^3)$ to $O(MN^2)$, where M is the number of objective functions and N is the population size.
(2) The elite strategy is introduced to merge the parents and offspring of the population into a hierarchical ranking, which ensures that some excellent individuals of the population will not be discarded in the evolutionary process, thus improving the accuracy of the optimization results.
(3) The comparison operator of congestion degree is used to avoid the difficulty of preset sharing parameters.

NSGA-II is one of the most popular multi-objective genetic algorithms, which has been widely used in all walks of life to find the best solution. For example, Li et al., [14] used an NSGA-II-based method to predict multi-label, which can greatly save the time consumption and memory occupancy of label prediction and improve the accuracy of label prediction to a certain extent. Jiang et al., [31] proposed a multi-objective time-of-use pricing optimization method based on NGSA-II, which can better achieve the effect of peak cutting and valley filling. NSGA-II has a good effect in finding multi-objective optimal solutions, so this paper uses NSGA-II method to find better subgraphs in multi-constrained matched subgraphs.

3 Reliability-based multi-fuzzy-objective graph pattern matching

3.1 Preliminaries

3.1.1 Reliability

Reliability is the probability of an entity to perform its functions under certain conditions at certain time. Please refer to Table 1 for all symbols and notations adopted to introduce and describe reliability in this section.

The reliability of a system is related to the reliability of components of the system and its various combinations. From the perspective of reliability, each component in the system can be status of normal or malfunction, in which the reliability can be evaluated as the probability of the function of the component is normal. Here, in big graph, we can take a subgraph as a system, and the reliability of a subgraph is determined by the reliability of its component nodes and their connections. Obviously, the subgraph S functions well only when all its component nodes perform well. Set the reliability of component node i as $R_i^{(S)}(t) = P\left(T_i^{(S)} > t\right), i = 1, 2, \ldots m$, where $T_i^{(S)}$ is the lifetime of component node i in

Table 1 The symbols and notations used in Section 3.1.1

S	subgraph
i	component node
$R_i^{(S)}(t)$	the reliability of component node i
$T_i^{(S)}$	the lifetime of component node i in subgraph S
$T^{(S)}$	the lifetime of the subgraph
$R^{(S)}$	the reliability of the subgraph
m	the total number of component nodes
k	some of component nodes
ϕ	a known function
$n_i^{(S)}$	the number of test for component node i
$s_i^{(S)}$	the number of succeed test for component node i
$f_i^{(S)}$	the number of failed test for component node i
$Z^{(S)}$	the number of succeed test for the subgraph
$\theta^{(S)}$	the number of failed test for the subgraph
l	$\prod_{k=1}^{m}(n_k+1)$
$R_L^{(S)}(Z^{(S)})$	the lower confidence with $1-\alpha$ for reliability $R^{(S)}$
$\hat{R}^{(S)}$	the maximum likelihood estimation on $R^{(S)}$
$F_{1-\alpha}(m, n)$	F distribution with quantile $1-\alpha$

subgraph S, which is the time from $t=0$ to the time when the component node becomes malfunction. Assume that the malfunction of a component node is independent to each other, i.e., $T_1^{(S)}, T_2^{(S)}, ..., T_m^{(S)}$ share mutual independence. Suppose that all component nodes in the subgraph S begin to function at $t=0$, and the lifetime of the subgraph should be

$$T^{(S)} = \min\left(T_1^{(S)}, T_2^{(S)}, ..., T_m^{(S)}\right) \tag{1}$$

Then the reliability of the subgraph is

$$R^{(S)}(t) = P\left(T^{(S)} > t\right) = \prod_{i=1}^{m} R_i^{(S)}(t) \tag{2}$$

However, from time to time, although there are m component nodes in the subgraph, it can function well if and only if there are at least k component nodes which function well, where $k < m$. For example, a composite service can perform well only when some key components function well to satisfy the specific requirements; a small community can be selected only when some community members can perform the functionality well. Mathematically, if the lifetime of component nodes in the subgraph S are $T_1^{(S)}, T_2^{(S)}, ..., T_m^{(S)}$, which are independent and conform to the same distribution, and the reliability of component nodes in the subgraph S are denoted as $R_i^{(S)}(t) = P\left(T_i^{(S)} > t\right)$, then the reliability of the subgraph is

$$R^{(S)} = \sum_{j=k}^{m} \binom{m}{j}\left(R_i^{(S)}(t)\right)^j\left(1-R_i^{(S)}(t)\right)^{m-j} \tag{3}$$

The reliability of the subgraph S is

$$R^{(S)} = \phi\left(R_1^{(S)}, R_2^{(S)}, ..., R_m^{(S)}\right) \tag{4}$$

where ϕ is a known function. From the history record, as for component node i there are $n_i^{(S)}$ tests, where it successes $s_i^{(S)}$ times, and fails $f_i^{(S)}$ times, $n_i^{(S)} = s_i^{(S)} + f_i^{(S)}$, $n_i^{(S)} \geq 1$, $s_i^{(S)} \geq 0$, $f_i^{(S)} \geq 0$. Denote

$$Z^{(S)} = \left(s_1^{(S)}, s_2^{(S)}, \ldots, s_m^{(S)} \right) \tag{5}$$

$$\theta^{(S)} = \left(R_1^{(S)}, R_2^{(S)}, \ldots, R_m^{(S)} \right) \tag{6}$$

Assume each test is independent to each other. Obviously, the distribution of $Z^{(S)}$ is dependent on $\theta^{(S)}$,

$$P_{\theta^{(S)}} \left(Z^{(S)} = \left(i_1^{(S)}, i_2^{(S)}, \ldots, i_m^{(S)} \right) \right) = \prod_{k=1}^{m} \binom{n_k^{(S)}}{i_k^{(S)}} \left(R_k^{(S)} \right)^{i_k^{(S)}} \left(1 - R_k^{(S)} \right)^{n_k^{(S)} - i_k^{(S)}} \tag{7}$$

Then it is possible to sort $Z^{(S)}$ with certain rule

$$z_1 \succ z_2 \succ \ldots z_l \tag{8}$$

where

$$l = \prod_{k=1}^{m} (n_k + 1) \tag{9}$$

Let

$$G^{(S)} \left(z, \theta^{(S)} \right) = \sum_{z_k \geq z} P_{\theta^{(S)}} \left(Z^{(S)} = z_k \right) \tag{10}$$

$$R_L^{(S)}(z) = \inf \left\{ \phi \left(\theta^{(S)} \right) : G^{(S)} \left(z, \theta^{(S)} \right) > \alpha \right\} \tag{11}$$

With Theorem 2.1 in Chapter 8 of [3], we have

$$P_\theta \left(R^{(S)} \geq R_L^{(S)} \left(Z^{(S)} \right) \right) \geq 1 - \alpha \tag{12}$$

where $0 < \alpha < 1$, and $R_L^{(S)} \left(Z^{(S)} \right)$ is the lower confidence with $1 - \alpha$ for reliability $R^{(S)}$.

With Eq. (2), we have the maximum likelihood estimation on $R^{(S)}$ as

$$R^{(S)} = \prod_{j=1}^{m} \frac{s_j^{(S)}}{n_j^{(S)}} \tag{13}$$

Theorem 1: If there is no malfunction in the subgraph S, i.e., $\max(f_i) = 0$, then the confidence lower limits of the reliability of the subgraph S with confidence level 1-α is

$$R_L^{(S)} \left(n_1^{(S)}, \ldots, n_m^{(S)} \right) = \alpha^{\frac{1}{n_*^{(S)}}} \tag{14}$$

where $n_*^{(S)} = \min\left(n_1^{(S)}, \ldots, n_m^{(S)}\right)$.

Proof: From Eq. (11), we have

$$R_L^{(S)}\left(n_1^{(S)}, \ldots, n_m^{(S)}\right) = \inf\left\{\prod_{i=1}^m R_i^{(S)} : \prod_{i=1}^m R_i^{n_i^{(S)}} = \alpha\right\} \tag{15}$$

Then

$$\alpha = \prod_{i=1}^m R_i^{n_i^{(S)}} = \left(R^{(S)}\right)^{n_*^{(S)}} \prod_{j=1}^m R_j^{n_j^{(S)} - n_*^{(S)}} \tag{16}$$

i.e.,

$$R^{(S)} = \left[\frac{\alpha}{\prod_{j=1}^m R_j^{n_j^{(S)} - n_*^{(S)}}}\right]^{\frac{1}{n_*^{(S)}}} \tag{17}$$

Obviously,

$$R_L^{(S)}\left(n_1^{(S)}, \ldots, n_m^{(S)}\right) \geq \alpha^{\frac{1}{n_*^{(S)}}} \tag{18}$$

i.e., the confidence lower limits of the reliability of the subgraph S with confidence level $1 - \alpha$ is

$$R_L^{(S)}\left(n_1^{(S)}, \ldots, n_m^{(S)}\right) = \alpha^{\frac{1}{n_*^{(S)}}} \tag{19}$$

It is easy to calculate the reliability of the subgraph S with no malfunction at all based on the above theorem. However, it is difficult to compute the reliability of the subgraph without any constraint. Here we adopt an interpolation method to approximately calculate the confidence lower limits of the reliability of the subgraph S with confidence level $1 - \alpha$ as follows. With

$$n_*^{(S)} = \min\left(n_1^{(S)}, \ldots, n_m^{(S)}\right) \tag{20}$$

and

$$s_*^{(S)} = n_*^{(S)} \prod_{i=1}^m \frac{s_i^{(S)}}{n_i^{(S)}} \tag{21}$$

Let $R_{LM}^{(S)(1)}$ and $R_{LM}^{(S)(2)}$ satisfy.

$$\sum_{x=\left[s_*^{(S)}\right]}^{n_*^{(S)}} \binom{n_*^{(S)}}{x} \left(R_{LM}^{(S)(1)}\right)^x \left(1 - R_{LM}^{(S)(1)}\right)^{n_*^{(S)} - x} = \alpha,$$

$$\sum_{x=\left[s_*^{(S)}\right]+1}^{n_*^{(S)}} \binom{n_*^{(S)}}{x} \left(R_{LM}^{(S)(2)}\right)^x \left(1 - R_{LM}^{(S)(2)}\right)^{n_*^{(S)} - x} = \alpha.$$

Then the approximate confidence lower limits of the reliability of the subgraph S with confidence level $1 - \alpha$ is

$$R_{LM}^{(S)} = R_{LM}^{(S)(1)} + \left(s_*^{(S)} - \left[s_*^{(S)}\right]\right)\left(R_{LM}^{(S)(2)} - R_{LM}^{(S)(1)}\right) \tag{22}$$

With the Lindstrom-Madden approach from [3], we can have

$$R_{LM}^{(S)(1)} = \cfrac{1}{\left[1 + \cfrac{n_*^{(S)} - \left[s_*^{(S)}\right] + 1}{\left[s_*^{(S)}\right]} F_{1-\alpha}\left(2\left(n_*^{(S)} - \left[s_*^{(S)}\right] + 1\right), 2\left[s_*^{(S)}\right]\right)\right]}$$

$$R_{LM}^{(S)(2)} = \cfrac{1}{\left[1 + \cfrac{n_*^{(S)} - \left[s_*^{(S)}\right]}{\left[s_*^{(S)}\right] + 1} F_{1-\alpha}\left(2\left(n_*^{(S)} - \left[s_*^{(S)}\right]\right), 2\left(\left[s_*^{(S)}\right] + 1\right)\right)\right]}$$

where $F_{1-\alpha}(m, n)$ is F distribution with quantile $1 - \alpha$.

3.1.2 Multi-fuzzy-objective simulation

As for reliability, each node v_i in the data graph has two attributes including the total number of experiments n_i and the number of successes s_i, so we can define the probability that node v_i works normally $p_i = {}^{s_i}/_{n_i}$. In the data graph, the edge between nodes has attribute information, such as social trust T and social intimacy r. The premise that a path in the graph works normally is that a certain number of nodes on the path should work normally, so the probability of normal operation of the path should be calculated by the multiplication of the probability of normal operation of its components, which is the same as the method when calculating social trust or social intimacy proposed by Liu et al., in [19].

On the one hand, in the traditional multi-constrained graph pattern matching, there may be a case that in the data graph, where a certain attribute value of a subgraph is slightly lower than the corresponding given threshold value, but other attribute values are obviously better than the corresponding given threshold values. The corresponding subgraphs should be better subgraphs, but they are excluded by traditional multi-constrained graph pattern matching. Therefore, we can introduce fuzzy numbers to obtain this kind of good subgraphs. On the other hand, because social trust and social intimacy reflect the subjective consciousness of participants, and using numerical values to express trust and intimacy between participants has some ambiguity, so we introduce fuzzy parameters [12, 13] to appropriately expand the range of trust and social intimacy. For example, when the minimum value of social trust required in the pattern graph is t, we introduce a fuzzy parameter $\gamma(0 < \gamma \le 1)$ to modify the minimum value of social trust to $\gamma \bullet t$.

Definition 1: **Data graph** $G_D = \left(V, E, f_V^D, f_E^D\right)$ is
a labeled directed graph, where – V is a set of nodes;
- E is a set of edges, and $(v_i, v_j) \in E$ is a directed edge from node $v_i \in V$ to node $v_j \in V$;
- f_V is a function defined on the node set V such that for each node $v \in V$, $f_V^D(v)$ is a set of attributes associated with v, represented by a set of labels for v.

– f_E^D is a function defined on edge set E such that for each edge $e \in E$, f_E^D is the set of attributes associated with e, represented by a set of labels for e.

Definition 2: Pattern Graph $G_P = \left(V_P, E_P, f_V^P, f_E^P\right)$ is a directed graph with labels, where – V_P and E_P are node sets and directed edge sets, respectively;

– f_V^P is a function defined on the node set V_P such that $f_V^P(v)$ is the attributes associated with v for each node $v \in V$;

– f_E^P is a function defined on the edge set E_P such that $f_E^P(e)$ is the attributes associated with e for each edge $e \in E$;

Definition 3: Multi-Fuzzy-Objective Simulation (MFOS): Given the data graph $G_D = \left(V, E, f_V^D, f_E^D\right)$ and the pattern graph $G_P = \left(V_P, E_P, f_V^P, f_E^P\right)$, G_D simulates the matched G_P through the multi-fuzzy-objective, that is, the multi-fuzzy-objective simulation map G_D to G_P, defined as $G_P \lhd_S^{MFO} G_D$ if there is a binary relationship $S \subseteq V_P \times V$ such that

- For all $u \in V_P$, there exists $v \in V$ such that $(u, v) \in S$;
- For each pair $(u, v) \in S$,
- For each edge (u, u') in E_P, there exists a non-empty path ρ from v to v' in G_D such that $(u', v') \in S$, and the length of path $len(\rho)$ should be minimized fuzzily.
- The attributes of f_V^D or f_E^D should be optimized fuzzily.

In this paper, when multiple matched subgraphs with good attributes are required to be selected, this decision-making is affected by the following six factors:

- *Objective 1: Social trust value (T):* This value represents the degree of trust between two participants in a social network. The bigger, the better.

$$\max f_1 = T$$

- *Objective 2: Social relationship value (R):* This value represents the degree of intimacy between two participants in a social network. The bigger, the better.

$$\max f_2 = R$$

- *Objective 3: Reliability value (R_{LM}):* This value indicates the possibility that the matched subgraph will work normally. The bigger, the better.

$$\max f_3 = R_{LM}$$

- *Objective 4: The membership degree of social relationship*: This value indicates the degree of fuzziness of social relationship value. The closer the degree of social relationship to 1, i.e., the closer the value of social relationship to the specified threshold value on the pattern graph, the better.

$$\max f_4 = \begin{cases} \dfrac{T-B}{A\lambda_T-B} & \left(\dfrac{T-B}{A\lambda_T-B} < 1\right) \\ 1 & otherwise \end{cases}$$

- *Objective 5: The membership value of reliability*: This value represents the degree of fuzziness of reliability. The closer to 1, the less ambiguous.

$$\max f_5 = \begin{cases} \dfrac{R_{LM}^{(S)}-C}{A\lambda_{R_{LM}^{(S)}}-C} & \left(\dfrac{R_{LM}^{(S)}-C}{A\lambda_{R_{LM}^{(S)}}-C} < 1\right) 1 \qquad otherwise \end{cases}$$

- *Objective 6: The path length of matched subgraph*: The larger the path length of matched subgraph, the smaller the aggregated attribute value of the path of matched subgraph. Hence, the smaller the value, the better.

$$\min f_6 = Pathlength$$

where $A\lambda_T$ and $A\lambda_{R_{LM}^{(S)}}$ represent aggregated values of social trust and reliability in pattern graph; T and R represent the social trust value and social relationship value of matched subgraphs; The calculation methods of these values are mentioned in [19]; the reliability values are obtained by the above Eq. (22); B and C are parameters set according to the personal requirements of users.

3.2 Reliability-based multi-fuzzy-objective graph pattern matching algorithm

The most classical graph pattern matching is isomorphism-based graph pattern matching, which is to find whether there is a matched subgraph in the data graph with the structure pattern matching in the query graph. Since this subgraph simulation is an NP-complete problem, it is difficult to locate the matched subgraph directly. Therefore, we here need first find a strong subgraph in the big graph, then compress the strong subgraph, calculate the index of the compressed graph, and finally propose a heuristic algorithm based on the simulation of graph pattern matching. Considering the reliability of nodes in the subgraph matching and the selection of more and better subgraphs, we propose a reliability-based multi-fuzzy-objective graph pattern matching (shorted as RMFO-GPM) method, with the detailed process as follows:

Step 1:   *Find the strong subgraph*: The strong subgraph is a closely related subgraph, that is, all attribute values in the subgraph are good enough. Formally, we have the following definition about the strong graph.

Definition 4: A strong graph is a strongly connected graph where each node associated with a high reliability in a specific domain which is connected with the edges associated with strong trust and strong social relationships Obviously, the use of graph pattern matching in the strong subgraph can greatly improve the searching efficiency. Based on the theory from social psychology [25], social structure and social relations between persons usually remain stable for a long time, such as the one in [19]. In order to facilitate the calculation, we put the attribute information of a node about the probability when the node performs normally on an edge of the graph, as adopted by [19].

Step 2:   *Compress the strong subgraph*: As for nodes in the strong subgraph, we perform the accessible compression, the subgraph pattern compression, and the subgraph attribute compression.

Step 2.1:   *Accessible compression*: If two nodes in a strong subgraph have the same ancestor and can reach each other's descendant nodes, then the two nodes can be compressed into one node.

Step 2.2:   *Subgraph pattern compression*: If two nodes in a strong subgraph have the same label, the same descendant node and the same ancestor node, the two nodes can be compressed into one node.

Step 2.3:   *Subgraph attribute compression*: If two nodes in a strong subgraph have the same label, the same descendant node, the same ancestor node and one path with one node dominates the other path with another node, the two nodes can be compressed into one node and the aggregation value of the attribute is the attribute value of the dominated node [19].

Step 3:   *Index the compressed graph of strong subgraph*: As for the nodes of strong subgraph, evaluate the accessible index value, the subgraph pattern index value, the subgraph attribute index value and the aggregated attribute value.

Step 3.1:   *Accessible index*: Records the ancestors and precursors that the node can access.

Step 3.2:   *Subgraph pattern index*: Record the shortest path length between any two nodes in the strong subgraph.

Step 3.3:   *Subgraph attribute indexes and aggregated attribute values*: These values are related to the shortest path in the subgraph pattern index.

Step 4:   *Subgraph pattern matching heuristic algorithm*: Dijkstra algorithm can be used to get the minimum of the maximal aggregated values of node attributes. Usually the smaller the node attribute value, the greater the probability of satisfying the maximum requirement. Each matched subgraph contains four attribute values, including the value of social trust and social intimacy on the edge, the value of reliability of matched subgraph obtained from the total number of experiments and the number of successes of nodes in matched subgraph, and the path length of the whole matched subgraph. The matched subgraph set can be used in NSGA-II algorithm to make decision according to the six objective functions mentioned above, and the better partial matched subgraphs can be found. NSGA-II is a multi-objective genetic algorithm, which has a good effect in

finding Pareto frontier quickly and maintaining population diversity. The algorithm can obtain a set of Pareto solutions satisfying multi-objective requirements in all feasible solutions. Because NSGA-II can search Pareto solutions well in academia and industry, this paper uses NSGA-II to select better matched subgraphs from the matched subgraph set. Hence, a reliability-based multi-fuzzy-objective graph pattern matching method (named as RMFO-GPM) is proposed.

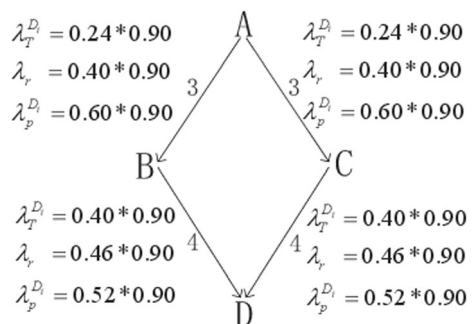# 4 Experiments

4.1 Experimental settings

4.1.1 Dataset

This paper adopts datasets from the Epinions dataset (please refer to http://snap.stanford.edu) shared by Stanford University. The Epinions data set is a trust network, including the user trust relationship, the user's score information for the item, and the comment information. Statistically, this data set has 75,879 nodes and 508,837 edges.

4.1.2 Parameter settings

As mentioned in [19], the mining of the social factor values through social networks is beyond the scope of our work. The value of the real world attribute is usually different from each other, which may be low or high. Therefore, in the experiment we also use the rand() function in MYSQL to generate a random number between [0, 1] to simulate the value of the attribute. In the experiment, we randomly generate three sets of random numbers using the rand() function and discover that the experimental results are similar to each other. Therefore, in this paper we choose to present the results of only a set of data. For each node's number of experimental successes and the total number of experiments, we use a heuristic algorithm simulated by the above subgraph to conduct 10 sets of experiments firstly, and then count the total number of experiments and the number of successful experiments. In addition, set the confidence levels in the F distribution function to 0.10, 0.20, 0.30, 0.40, 0.50, 0.60,

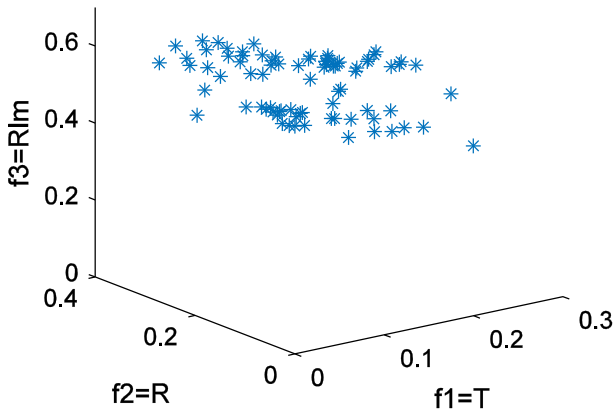Figure 1 The pattern graph adopted in the experiments

$$\lambda_T^{D_i} = 0.24 * 0.90 \qquad A \qquad \lambda_T^{D_i} = 0.24 * 0.90$$
$$\lambda_r = 0.40 * 0.90 \qquad \lambda_r = 0.40 * 0.90$$
$$\lambda_p^{D_i} = 0.60 * 0.90 \quad 3 \quad \diagup \quad 3 \quad \lambda_p^{D_i} = 0.60 * 0.90$$

$$B \qquad\qquad C$$

$$\lambda_T^{D_i} = 0.40 * 0.90 \qquad \lambda_T^{D_i} = 0.40 * 0.90$$
$$\lambda_r = 0.46 * 0.90 \quad 4 \quad\quad 4 \quad \lambda_r = 0.46 * 0.90$$
$$\lambda_p^{D_i} = 0.52 * 0.90 \qquad \lambda_p^{D_i} = 0.52 * 0.90$$

$$D$$

Figure 2 The results of the optimized matched subgraphs by NGSA-II when pop = 800.70, 0.80, and 0.90, respectively

Figure 1 is the pattern graph inputted in the experiment, and 0.90 in Figure 1 is the membership parameter. We have conducted many experiments on the membership parameter, and the experiment shows that when the membership parameter is 0.90, the effect is the best, so the membership parameter is set as $\gamma = 0.90$ in the experiment. The first multiplier of the equation in Figure 1 is the minimum of social trust, social intimacy and probability of performing normally.

Since the matched results of subgraph simulation already contain the attribute values of each matched subgraph, we only need to select the better matched subgraphs. Therefore, the decision variables in the NSGA-II algorithm in this paper are the binary representation of the sequence number of matched subgraph. If crossover operation is used in NSGA-II algorithm, many repetitive individuals will be generated in the descendant population, so mutation operation is only used in this experiment and the mutation efficiency is 0.10. Experiments on population sizes of 80, 90, 100, 110 and 120 show that the population size as 100 and the generation number as 4000 are the best. Figures 2, 3, 4 are the results of the matched subgraphs obtained by the Epinions dataset with a confidence level of 0.10 in the RMFO-GPM. The population size of the multi-objective genetic algorithm is 80, 100, and 120, respectively, and generation number is 4000. From Figures 2, 3, 4, we can see that the Pareto solution with population size as 100 and generation number as 4000 (as in Figure 3) is evenly distributed,
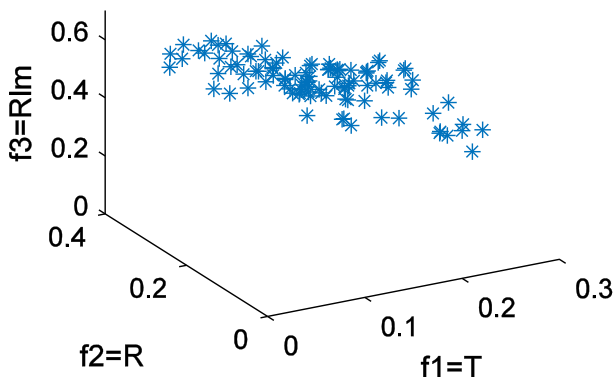

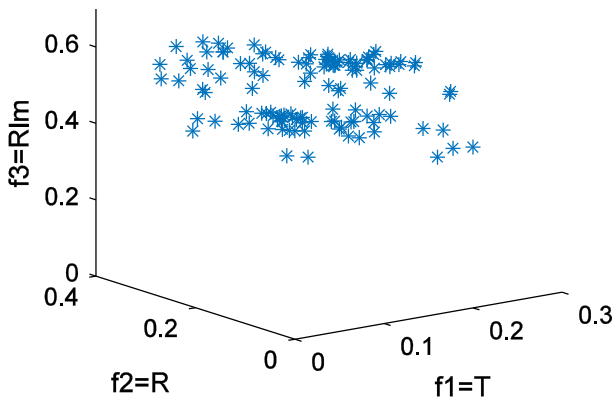**Figure 3** The results of the optimized matched subgraphs by NGSA-II when pop = 100

Figure 4 The results of the optimized matched subgraphs by NGSA-II when pop = 120

and the corresponding effect is better. The coordinate axes of the three-dimensional graph are the values of the objective functions f_1, f_2, f_3.

4.2 Experimental results and analysis

On the one hand, there is no existing method to consider that one attribute value of a subgraph is slightly lower than the given corresponding threshold value, but the other attribute values are obviously better than the given corresponding threshold value; this subgraph may be a better subgraph, so this paper compares our proposed one with the no fuzzy graph pattern matching (i.e., reliability-based multi-objective graph pattern matching, named as RMO-GPM) in a comparative experiment. On the other hand, as there is no existing method to consider the failure state of some nodes in matched subgraph, this paper sets up an unreliable graph pattern matching algorithm (named as MFO-GPM) to conduct a comparative experiment. The experimental results validate the effectiveness of the proposed RMFO-GPM algorithm.

The uncertainty of the direction of graph search results in that the results of matched subgraphs are not exactly the same each time. Hence, for the sake of generality, we conduct
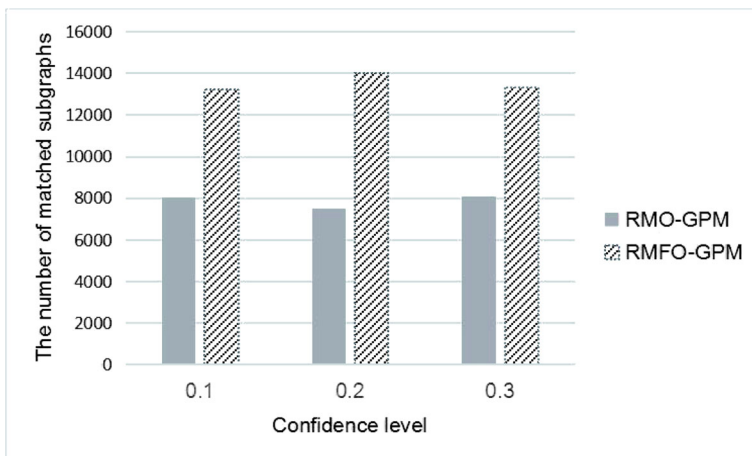

Figure 5 The number of matched subgraphs under three different confidence levels

Table 2 The optimized attribute values of the matched subgraphs with confidence level 0.10 in NGSA-II

	The number of matched subgraphs	T	R	R_{LM}
RMFO-GPM	13,962	0.1519	0.2406	0.4952
	14,027	0.1440	0.2525	0.5240
	11,661	0.1451	0.2450	0.5147
RMO-GPM	8603	0.0878	0.1804	0.4949
	7721	0.0943	0.1823	0.5128
	7791	0.0963	0.1816	0.5052

three groups of experiments with the same confidence level. The information of matched subgraphs obtained in RMFO-GPM and RMO-GPM algorithms is shown in Figure 5 and Table 2. Specifically, Figure 5 shows the average number of matched subgraphs with confidence level of 0.10, 0.20 and 0.30 respectively under RMFO-GPM and RMO-GPM algorithms. In addition, we can observe that RMFO-GPM algorithm with fuzzy numbers can select more matched subgraphs. Table 2 lists the attribute information of the matched subgraph obtained under RMFO-GPM and RMO-GPM when the confidence level is 0.1, wherethe second column in Table 2 shows the number of matched subgraphs. By analyzing Table 2, we can see that although the number of matched subgraphs is different under the same conditions, the attributes of matched subgraphs are basically the same. Therefore, we can determine the relationship between the confidence level and the reliability value R_{LM} by discussing the change of the average value of the best matched subgraph in NSGA-II. It can also be seen from Table 3 that each attribute value of the better matched subgraph obtained under the RMFO-GPM algorithm is larger than the corresponding attribute value obtained by the RMO-GPM. Therefore, by analyzing Figure 5, Table 2, and Table 3, it can be concluded that RMFO-GPM can get more matched subgraphs and the matched subgraphs have better attribute values, which means that RMFO-GPM can select a certain attribute value is slightly lower for a given corresponding threshold value, but other attribute values are significantly better than the matched subgraph of the given corresponding threshold value .

Table 3 reflects the attribute values obtained under three different graph pattern matching algorithms with confidence levels of 0.10, 0.20, and 0.30, respectively. According to the analysis of RMFO-GPM and MFO-GPM in Table 3, the reliability value of the matched subgraph obtained by RMFO-GPM is larger. Because MFO-GPM does not consider the reliability of matched subgraphs, subgraphs are selected with less reliable matched subgraphs. The matched subgraph obtained by MFO-GPM needs to optimize

Table 3 The attribute values of matched subgraphs under different matching algorithms

|	$\alpha = 0.10$			$\alpha = 0.20$			$\alpha = 0.30$		
	T	R	R_{LM}	T	R	R_{LM}	T	R	R_{LM}
RMFO-GPM	0.1470	0.2460	0.5113	0.1417	0.2518	0.4656	0.1451	0.2496	0.4604
MFO-GPM	0.0785	0.1576	0.3679	0.0682	0.1448	0.3949	0.0963	0.1735	0.4006
Increment (%) of RMFO-GPM from MFO-GPM	87.37	56.06	38.98	107.72	73.86	17.90	50.69	43.88	14.93
RMO-GPM	0.0928	0.1814	0.5043	0.0931	0.1857	0.4550	0.0943	0.1861	0.4404
Increment (%) of RMFO-GPM from RMO-GPM	58.44	35.59	1.39	52.14	35.59	2.33	53.85	34.13	4.54
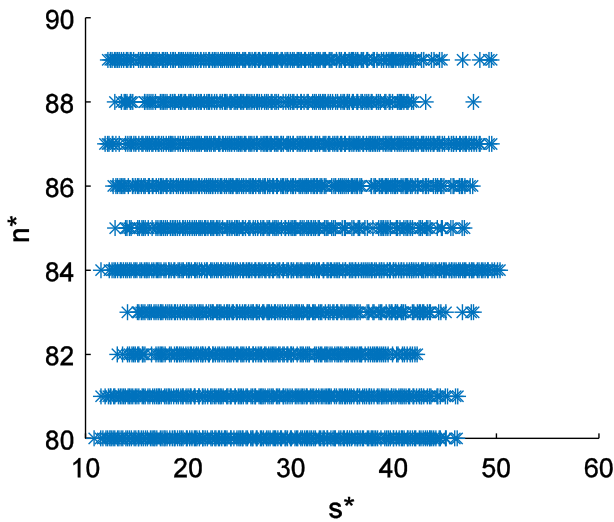
Figure 6 The relationship between s_* and n_* in Epinions1–1

multiple objectives in NSGA-II, and the value of reliability is large, but other attribute values are not good or other attributes are good enough but reliability is not good, so each attribute value of the matched subgraph is smaller than the attribute value of the preferred matched subgraph of RMFO-GPM. This experiment fully demonstrates the effectiveness of reliability.

Epinions 1–1 in Figure 6 represents the results of the first matched subgraph set obtained at a confidence level of 0.10. Specifically, Figure 6 shows the relationship between s_* and n_* in Epinions1–1, where n_* is the minimum integer value of the total number of experiments for all nodes in a matched subgraph, and s_* is the product of the probability of a matched subgraph working normally P and n_*. Because there are many n_* values in many matched subgraphs, and the probability of a matched subgraph working normally is not unique, which leads to the existence of several n_* values corresponding to the value of s_*. The larger the value of n_*, the larger the value of s_*, while the probability of normal operation of the matched subgraph remains unchanged (Figure 7).
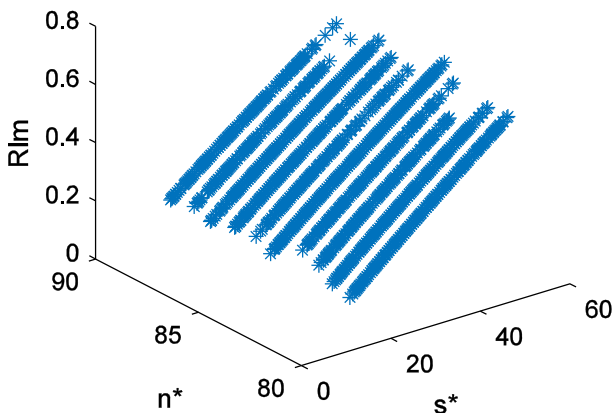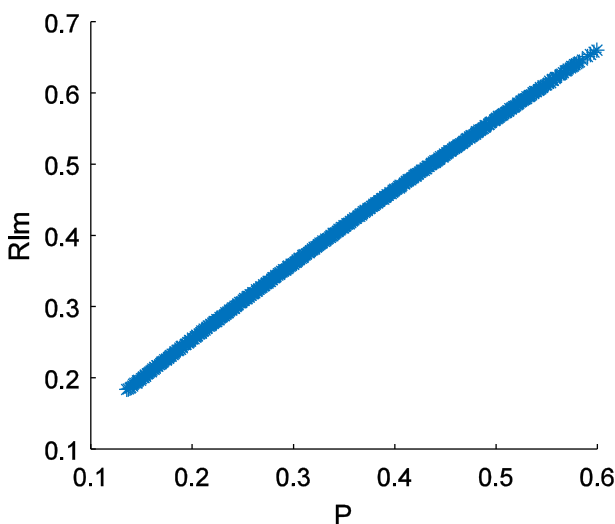

Figure 7 The relationship between s_*, n_* and R_{LM} in Epinions1–1

Table 4 The attribute values of matched subgraphs with different confidence levels

	T	R	s_*	n_*	R_{LM}	Pathlength	$R_{LM}^{(S)}$
0.1	0.1470	0.2460	38.0385	84.5500	0.5113	6.0135	0.5111
0.2	0.1417	0.2518	36.2438	84.8867	0.4656	5.9941	0.4661
0.3	0.1451	0.2496	36.9023	84.2433	0.4604	5.9941	0.4620
0.4	0.1538	0.2500	36.7307	83.8100	0.4461	5.9844	0.4453
0.5	0.1526	0.2484	36.2546	84.1700	0.4256	6.0103	0.4260
0.6	0.1503	0.2398	36.8792	84.9367	0.4155	6.0071	0.4148
0.7	0.1465	0.2389	35.6614	84.6600	0.3894	6.0523	0.3863
0.8	0.1483	0.2461	35.9511	84.2733	0.3782	6.0232	0.3777
0.9	0.1425	0.2409	37.6161	84.5600	0.3736	6.0329	0.3693

For the sake of generality, we conduct three experiments with RMFO-GPM algorithm at confidence of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9, respectively, and then select the best subgraphs for the matched subgraphs with NSGA-II. The average attribute values of the best subgraphs obtained from the three experiments are shown in Table 4, where the last column $R_{LM}^{(s)}$ denotes the true values that should be taken when the parameters are rounded to s_* and n_*. Through the analysis of Table 4, we can see that the values of T, R and Pathlength of each matched subgraph are basically the same, and the values of s_* and n_* of the selected better matched subgraph are not much different. Under the same confidence level, the mean value of reliability represented by R_{LM} is basically consistent with the true value represented by $R_{LM}^{(S)}$. Therefore, we can regard the mean attribute value of the better matched subgraph listed in Table 4 as the results of the same matched subgraph under different confidence levels. Figure 8 describes the relationship between confidence and matched subgraphs. Particularly, Figure 8 shows that the reliability of matched subgraph decreases as the increase of confidence. In the same subgraph, the values of s_*, n_*, T, R and Pathlength should be fixed, and the reliability value of matched subgraph decreases as the decrease of confidence (Figure 9).

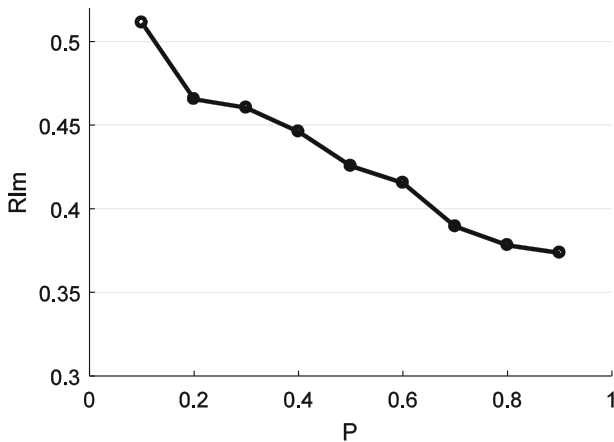**Figure 8** The relationship between P and R_{LM} in Epinions1–1

Figure 9 The relationship between confidence level and R_{LM}

5 Conclusion

In this paper, with the definition and principle of reliability, we can calculate the reliability of matched subgraphs, which is the probability that matched subgraphs work normally. The better matched subgraphs can affect people's decision-making attitude to a certain extent. We can evaluate the probability of completing a specific task according to the attributes of subgraphs, including reliability, trust and social relationship. In this paper, a multi-objective genetic algorithm NSGA-II is used to solve the problem of how to select the best matched subgraph considering reliability, trust and social relationship, in many sets of matched subgraphs. Finally, a reliability-based multi-fuzzy-objective graph pattern matching (named as RMFO-GPM) is proposed. The experimental results show that the proposed RMFO-GPM is effective comparing with other state-of-art methods.

References

1. Bai, Y., Wang, C., Ying, X.: Para-G: Path Pattern Query Processing on Large Graphs. World Wide Web Journal. **20**(3), 515–541 (2017)
2. Brynielsson, J., Högberg, J., Kaati, L., Mårtenson, C., Svenson, P.: Detecting Social Positions using Simulation. International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2010). pp. 48–55 (2010)
3. Chen, J.: Survival Analysis and Reliability (in Chinese). Peking University Press (2005)
4. Cheng, J., Yu, J. X., Ding, B., Yu, P. S., Wang, H.: Fast Graph Pattern Matching. IEEE International Conference on Data Engineering (ICDE 08). pp. 913–922 (2008)
5. Cheng, J., Zeng, X., & Yu, J.X.: Top-K Graph Pattern Matching Over Large Graphs. 2013 IEEE 29th International Conference on Data Engineering (ICDE 2013). pp. 1033–1044 (2013)

6.   Choudhury, S., Holder, L., Chin, G., Ray, A., Beus, S., Feo, J.: StreamWorks - A system for Dynamic Graph Search. ACM SIGMOD International Conference on Management of Data (SIGMOD 13). pp. 1101–1104 (2013)

7.   Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)

8.   Fan, W., Li, J., Ma, S., Tang, N., Wu, Y., Wu, Y.: Graph Pattern Matching: from Intractable to Polynomial Time. Proceedings of the VLDB Endowment. **3**(1–2), 264–275 (2010)

9.   Fan, W., Li, J., Luo, J., Tan, Z., Wang, X., Wu, Y.: Incremental Graph Pattern Matching. ACM SIGMOD International Conference on Management of Data (SIGMOD 11). pp. 925–936 (2011)

10. Fan, W., Wang, X., Wu, Y.: Diversified Top-K Graph Pattern Matching. Proceedings of the VLDB Endowment. **6**(13), 1510–1521 (2013)

11. Henzinger, M.R.: Computing Simulations on Finite and Infinite Graphs. 36th Annual Symposium on Foundations of Computer Science (FOCS 95). pp. 453 (2002)

12. Li, L., Wang, Y.: Context Based Trust Normalization in Service-Oriented Environments. International Conference on Autonomic and Trusted Computing (ATC 2010). pp. 122–138 (2010)

13. Li, L., Wang, Y., Varadharajan, V.: Fuzzy Regression Based Trust Prediction in Service-Oriented Applications. International Conference on Autonomic and Trusted Computing (ATC-09). pp. 221–235 (2009)

14. Li, L., Chu, Y., Wang, M., Han, L., Wu, X.: NSGA-II-based Multi-label Seed Node Selection in Network Environments (in Chinese). J Electron Inf Technol. **39**(9), 2040–2047 (2017)

15. Li, L., Zhou, C., He, J., Wang, J., Li, X., Wu, X.: Collective Semantic Behavior Extraction in Social Networks. J Comput Sci. **28**(2018), 236–244 (2018)

16. Li, L., Li, A., Wang, M., Wu, X.: Social Network based Behavior Analysis and Mining, Science Publishing, 2018

17. Liu, G., Wang, Y., Orgun, M.A.: Optimal Social Trust Path Selection in Complex Social Networks, Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI10), pp. 1391–1398 (2010)

18. Liu, G., Wang, Y., Orgun, M.A.: Finding K Optimal Social Trust Paths for the Selection of Trustworthy Service Providers in Complex Social Networks. IEEE Trans. Serv. Comput. **6**(2), 152–167 (2013)

19. Liu, G., Zheng, K., Wang, Y., Orgun, M. A., Liu, A., Zhao, L., Zhou, X. F.: Multi-constrained Graph Pattern Matching in Large-scale Contextual Social Graphs. IEEE International Conference on Data Engineering (ICDE 2015). pp.351–362 (2015)

20. Liu, G., Liu, Y., Zheng, K., Liu, A., Li, Z., Wang, Y., Zhou, X.: MCS-GPM: Multi-Constrained Simulation Based Graph Pattern Matching in Contextual Social Graphs. IEEE Trans. Knowl. Data Eng. **30**(6), 1050–1064 (2018)

21. Liu, A., Wang, W., Shang, S., Li, Q., Zhang, X.: Efficient Task Assignment in Spatial Crowdsourcing with Worker and Task Privacy Protection. GeoInformatica. **22**(2), 335–362 (2018)

22. Liu, G., Shi, Q., Zheng, K., Liu, A., Li, Z., Zhou, X.: An Efficient Method for Top-k Graph based Node Matching, World Wide Web Journal, In Press (2019)

23. Liu, G., Shi, Q., Zheng, K., Li, Z., Liu, A., Xu, J.: Context-aware Graph Pattern based Top-k Designated Nodes Finding in Social Graphs. World Wide Web Journal. **22**(2), 751–770 (2019)

24. Ma, S., Cao, Y., Fan, W., Huai, J., Wo, T.: Strong Simulation: Capturing Topology in Graph Pattern Matching. ACM Trans. Database Syst. **39**(1), 1–46 (2014)

25. Parton, N.: Reflections on the Social Construction of Reality: a Treatise in the Sociology of Knowledge. Gynecol. Oncol. **131**(2), 400–403 (2009)

26. Shi, Q., Liu, G., Zheng, K., Liu, A., Li, Z., & Zhao, L., et al.: Multi-constrained Top-K Graph Pattern Matching in Contextual Social Graphs. 2017 IEEE International Conference on Web Services (ICWS 2017). pp. 588–595 (2017)

27. Srinivas, N., & Deb, K.: Multi-objective Optimization Using Non-dominated Sorting in Genetic Algorithms. MIT Press (1994)

28. Sun, Z., Wang, H., Wang, H., Shao, B., Li, J.: Efficient Subgraph Matching on Billion Node Graphs. Proceedings of the VLDB Endowment. **5**(9), 788–799 (2012)

29. Tang, D., Zhang, Y., He, Y.: Research and Application on Crime Rule based on Graph Data Mining Algorithm (in Chinese). J Comput Technol Dev. **21**(11), 89–91 (2011)

30. Tong, H., Faloutsos, C., Gallagher, B., & Eliassi-Rad, T.: Fast Best-effort Pattern Matching in Large Attributed Graphs. ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 07). pp. 737–746 (2007)

31. Wang, X., Lv, Z., Tang, Z.: Multi-objective Dynamic Optimal Dispatching of Grid-connected Microgrid based on Tou Power Price Mechanism (in Chinese). Power System Protection & Control. **45**(4), 9–18 (2017)

32. Xiao, M., Wu, J., Huang, L., Cheng, R., Wang, Y.: Online Task Assignment for Crowdsensing in Predictable Mobile Social Networks. IEEE Trans. Mob. Comput. **16**(8), 2306–2320 (2017)

33. Xiao, M., Ma, K., Liu, A., Zhao, H., Li, Z., Zheng, K., Zhou, X.: SRA: Secure Reverse Auction for Task Assignment in Spatial Crowdsourcing. IEEE Trans. Knowl. Data Eng. (2019). https://doi.org/10.1109/TKDE.2019.2893240

34. Zhai, D., Sun, Y., Liu, A., Li, Z., Liu, G., Zhao, L., Zheng, K.: Towards Secure and Truthful Task Assignment in Spatial Crowdsourcing. World Wide Web Journal. 1–24 (2019). https://doi.org/10.1007/s11280-018-0638-2

35. Zhang, T., Gao, Y., Qiu, L., Chen, L., Linghu, Q., Pu, S.: Distributed Time-respecting Flow Graph Pattern Matching on Temporal Graphs, World Wide Web Journal, In Press (2019)

36. Zhu, J., Jiang, W., Liu, A., Liu, G., Zhao, L.: Effective and Efficient Trajectory Outlier Detection based on Time-dependent Popular Route. World Wide Web Journal. **20**(1), 111–134 (2017)

Publisher's note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.