

Uncertain Attribute Graph Sub-Graph Isomorphism and its Determination Algorithm

Yanjun Zhao*, Chunying Zhang

College of Science, Hebei United University

46 Xinhua Road, Tangshan 063009, Hebei, China Tel: 0315-2592480

*Corresponding author, e-mail: teacher_jsj@126.com

Abstract

The expectative sub-graph isomorphism of uncertain attribute graph is based on the analysis of complex network structure and the characteristic of uncertain attribute graph. The expectative sub-graph isomorphism of uncertain attribute graph is only one threshold value as constraint conditions. The method is simple, but the computation is large amount. Therefore, this paper brings in the definition of $\alpha - \beta$ sub-graph isomorphism of uncertain attribute graph, and explains the concept. Then, this paper designs and comes true the algorithm of $\alpha - \beta$ sub-graph isomorphism. Finally, through the experiments proves that $\alpha - \beta$ sub-graph isomorphic is better than expectative sub-graph, and it analyzes the variation in the different threshold cases. The research of $\alpha - \beta$ sub-graph isomorphism algorithm laid the foundation for uncertain attribute graph sub-graph query and community mining.

Keywords: expectative sub-graph isomorphism, $\alpha - \beta$ sub-graph isomorphism, uncertain attribute graph

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

As a generic data structure, graph can modeling and express the real world all kinds of complex data entity and the relationship between the entities. The whole social relationship can be abstracted as a graph in the field of social relation network [1-3]. Attribute graph [4-5] is the graph structure considering the vertices and edges attributes and relationships between attributes in the traditional map based on. It can be better described various nodes and the relationship between attributes in the social network, and easier to analyze the effects of these properties. However, vertices and vertices relations and their attributes are probability in a social network or other complex networks. Uncertain attribute graph is presented considering to this factor, and it further describes the social network uncertainty.

Sub-graph isomorphism is the classification of sub-graph to network with the same characteristics, and research the sub-graph isomorphism of data graph is a more profound study on the social networks. Now, about the sub-graph isomorphism research work is following. Dong Anguo [6] given algorithms for sub-graph isomorphism in graph pattern mining. The algorithms are based on algebra theory by making use of degree sequence of a vertex and eigenvalue of adjacency matrix. Xie Chunxin [7] given OES: sub-graph isomorphism verification algorithm. Which find a sub-graph isomorphism graph by searching edge by edge, and it can raise the verification efficiency by adjusting the edge order. Liu Bo [8] given design and implementation of sub-graph isomorphism detection algorithm based on relational model. He proposes a new sub-graph isomorphism algorithm named Relational Graph Decomposition Index (RGDI), and it has higher detection efficiency. Fred DePiero [9] given an algorithm using lengthier paths to approximate subgraph isomorphism. These sub-graph isomorphism algorithm design are based on traditional graph or uncertain graph. But in the real complex networks, it is often used to describe by uncertain attribute graph [10]. However, how to design the algorithm of uncertain attribute graph for uncertain attribute sub-graph isomorphism, it is an important issues.

Attributes figure is the expansion of traditional graph, and it is considered the vertex and edge attributes and the relationship of attribute between the structure of the graph. And without considering the attribute, attribute graph degenerate into traditional figure. So the

traditional figure is a special case of tradition graph. The research on the attribute graph structure, its various characteristics and operation is the theory complement and innovation of traditional graph. On uncertain graph sub-graph isomorphism has some research. With the uncertain attribute graph is put forward, and on uncertain attribute graph sub-graph isomorphism problem is urgent. The uncertainty in the data is everywhere, so it is the particularly important that research uncertain attribute graph. This paper presents two kinds definitions of sub-graph isomorphism on uncertainty attribute graph data. In probability, uncertainty attribute graph expectations sub-graph isomorphism is uncertain graph sub-graph isomorphism definition direct extension. The probabilistic significance of expectations sub-graph isomorphism of uncertain attribute graph is very intuitive, but it is huge computational cost and matches results describe complex. So, this paper puts forward uncertainty attribute sub-graph isomorphism definition and judgment algorithm. Uncertain attribute graph isomorphism is using two threshold values to replace uncertainty attribute graph expectative sub-graph isomorphism the single limit threshold value.

This paper structure mainly divided into five parts following.

2. The Uncertain Attribute Graph

DEFINITION 2.1. (Uncertain attribute graph) uncertain attribute graph I and uncertain attribute graph II are called uncertain attribute graph. It is denoted as $GA_p = ((V(VA, LV), E(EA, LE)), P) \cdot (V(VA, LV), E(EA, LE))$ is an attribute graph; P is the probability function of edge, vertex and their attribute. (P includes $P_E : E \rightarrow [0,1]$, $P_V : V \rightarrow [0,1]$, $P_{EA} : EA \rightarrow [0,1]$ and $P_{VA} : VA \rightarrow [0,1]$).

3. Expectative Sub-graph Isomorphism of Uncertain Attribute Graph I

Because of the uncertain attribute graph is divided into two aspects to discuss, we still discuss the uncertain attribute graph expectative sub-graph isomorphism following two aspects.

3.1. Expectative the Sub-graph Isomorphism of Uncertain Attribute Graph I

First, we discuss the uncertain attribute graph I:

NATURE 1. According to whether P the value is 1, the uncertain attribute graph I divided into:

$$GA_1^{UC} = \{GA_1 \mid GA_1 \subseteq GA_1, 0 < P(e_i) < 1 \text{ or } 0 < P(v_i) < 1 \text{ or } (0 < P(e_i) < 1 \text{ and } 0 < P(v_i) < 1), e_i \in E, v_i \in V\} \quad \text{and}$$

$$GA_1^C = \{GA_1 \mid GA_1 \subseteq GA_1, P(e_i) = 1, P(v_i) = 1, e_i \in E, v_i \in V\}$$

GA_1^C is the certain attribute graph set of GA_1 . GA_1^{UC} is the certain attribute graph set of GA_1 . Two set to meet $GA_1^C \cap GA_1^{UC} = \Phi$ and $GA_1^C \cup GA_1^{UC} = GA_1$.

$GA_{1c}(V_1(VA, LV), E_1(EA, LE))$ is a possible world graph the uncertain attribute graph I $GA_1 = ((V(VA, LV), E(EA, LE)), P_E, P_V)$, and $V_1(VA, LV) \subseteq V(VA, LV)$, $E_1(EA, LE) \subseteq E(EA, LE)$.

GA_1 contains GA_{1c} , and $\text{sign } GA_1 \Rightarrow GA_{1c}$. So the probability of “ GA_1 contains GA_{1c} ” computational formula is:

$$P(GA_1 \Rightarrow GA_{1c}) = \prod_{GA_{1c}} P(e_i) \cdot P(v_i) \prod_{GA_1^{UC} - GA_{1c}} (1 - P(e_i))(1 - P(v_i)) \quad (1)$$

The probability of “ GA_1 sub-graph isomorphic to GA_{1c} ” is the following formula:

$$P(GA_{11} \subseteq GA_{12})$$

$$= \sum_{GA_1 \cap GA_2} P(GA_{11} \Rightarrow GA_1) * P(GA_{12} \Rightarrow GA_2) * \phi(GA_1, GA_2) \quad (2)$$

Function $\varphi(GA_1, GA_2)$ values in $\{0,1\}$. If GA_1 and GA_2 sub-graph isomorphism, $\varphi(GA_1, GA_2)$ value is 1, or 0. Apparently, $P(GA_{II} \subseteq GA_{I2})$ is the probability expectation of $\varphi(GA_1, GA_2)$. According to the expectative value we can be defined for the expectation sub-graph isomorphism of uncertain attribute graph I.

DEFINITION3.1. Known uncertainty attribute graph GA_{II} and GA_{I2} , and expectative threshold value $\delta \in (0,1]$, if and only if $P(GA_{II} \subseteq GA_{I2}) \geq \delta$, GA_{II} is sub-graph isomorphic to GA_{I2} . Sign $GA_{II} \subseteq_{\delta} GA_{I2}$.

3.2 The expectative sub-graph isomorphic of uncertain attribute graph II

Similarly, we can define the expectative sub-graph isomorphism of uncertain attribute graph II.

NATURE 2. According to the a value whether P is 1, the uncertain attribute graph divides into mutually disjoint subsets of

$$GA_{II}^{UC} = \{GA_{II}^{\cdot} \mid GA_{II}^{\cdot} \subseteq GA_{II}, 0 < P(ea_i) < 1 \text{ or } 0 < P(va_i) < 1$$

$$\text{or } (0 < P(ea_i) < 1 \text{ and } 0 < P(va_i) < 1), ea_i \in EA, va_i \in VA\}$$

and

$GA_{II}^C = \{GA_{II}^{\cdot} \mid GA_{II}^{\cdot} \subseteq GA_{II}, P(ea_i) = 1, P(va_i) = 1, ea_i \in EA, va_i \in VA\}$. Where GA_{II}^C is the certain attribute graph of GA_{II} , and GA_{II}^{UC} is the uncertain attribute graph of GA_{II} , and Two set to meet $GA_{II}^C \cap GA_{II}^{UC} = \Phi$ and $GA_{II}^C \cup GA_{II}^{UC} = GA_{II}$.

$GA_{IIC}(V_1(VA, LV), E_1(EA, LE))$ is a possible world graph of uncertain attribute graph II, and $V_1(VA, LV) \subseteq V(VA, LV)$, $E_1(EA, LE) \subseteq E(EA, LE)$. GA_{II} contains GA_{IIC} , and signs $GA_{II} \Rightarrow GA_{IIC}$. So the probability of " GA_{II} contains GA_{IIC} " is the following formula:

$$P(GA_{II} \Rightarrow GA_{IIC}) = \prod_{GA_{IIC}} P(ea_i) \cdot P(va_i) \prod_{GA_{II}^{UC} - GA_{IIC}} (1 - P(ea_i)) \cdot (1 - P(va_i)) \quad (3)$$

The probability of " GA_{II} sub-graph isomorphic to GA_{IIC} " is the following formula:

$$P(GA_{III} \subseteq GA_{I12}) = \sum_{GA_1 \cap GA_2} P(GA_{III} \Rightarrow GA_1) * P(GA_{I12} \Rightarrow GA_2) * \varphi(GA_1, GA_2). \quad (4)$$

Function $\varphi(GA_1, GA_2)$ values in $\{0,1\}$. If GA_1 and GA_2 sub-graph isomorphism, $\varphi(GA_1, GA_2)$ value is 1, or 0. Apparently, $P(GA_{III} \subseteq GA_{I12})$ is the probability expectation of $\varphi(GA_1, GA_2)$. According to the expectative value we can be defined for the expectation sub-graph isomorphism of uncertain attribute graph II.

DEFINITION 3.2. Known uncertainty attribute graph GA_{III} and GA_{I12} , and expectative threshold value $\delta \in (0,1]$. If and only if $P(GA_{III} \subseteq GA_{I12}) \geq \delta$, GA_{III} is sub-graph isomorphic to GA_{I12} . Sign $GA_{III} \subseteq_{\delta} GA_{I12}$.

3.3. The Expectative Sub-graph Isomorphic of Uncertain Attribute Graph

DEFINITION 3.3. (the expectative sub-graph isomorphism of uncertain attribute graph) GA_{P1} , GA_{P2} and expectation threshold $\delta \in (0,1)$ are known quantity. If and only if $\sqrt{P(GA_{III} \subseteq GA_{I12})P(GA_{II} \subseteq GA_{I2})} \geq \delta$, GA_{P1} is sub-graph isomorphic to GA_{P2} . Sign $GA_{P1} \subseteq_{\delta} GA_{P2}$.

The algorithm of expectative sub-graph isomorphism judging is costly. Expectative sub-graph isomorphism still exist the problem of matching results describe difficulty in the application. These problems limit the application of expectative sub-graph isomorphism.

4. The α - β Sub-graph Isomorphism of Uncertain Attribute

4.1. The α - β Sub-graph Isomorphism of Uncertain Attribute Graph I

DEFINITION 4.1. The uncertain attributes set of uncertain attribute graph I GA_1 is. Definition $\alpha: 2^{GA_1^{UC}} \rightarrow (0,1]$ is error function. For the any sub-graph GA_1' of GA_1^{UC} , error function value is the following:

$$\alpha(GA_1') = \begin{cases} 1 - \prod_{GA_1' \subseteq GA_1} (1 - P(e_i))(1 - P(v_i)) & GA_1' \neq \Phi \\ 0 & GA_1' = \Phi \end{cases} \quad (5)$$

DEFINITION 4.2. The uncertain attributes set of uncertain attribute graph I GA_1^{UC} is know. $GA_{1\Delta}$ is any sub-graph of GA_1 . Definition $\beta: 2^{GA_{1\Delta}} \rightarrow (0,1]$ is strength function. For the any sub-graph $GA_{1\Delta}$ of GA_1 , strength function is the following:

$$\beta(GA_{1\Delta}) = \sqrt[|GA_{1\Delta}|]{\prod_{GA_{1\Delta} \in GA_1} P(v_i) \cdot P(e_i)} \quad (6)$$

DEFINITION 4.3. Uncertain attribute graph I GA_{11} and GA_{12} are known quantity. Hypothesis the uncertain attribute of GA_{11} is GA_{11}^{UC} , if the presence $GA_1' \subseteq GA_1^{UC}$ and mapping function f , while satisfying the following:

- (1) $I_{\max}(GA_{11}) - GA_1'$ sub-graph isomorphic to $I_{\max}(GA_{12})$ and f is mapping function;
- (2) $\alpha(GA_1') \leq \alpha_{\max}$, Constant α_{\max} called error threshold value, $\alpha_{\max} \in [0,1]$;
- (3) $\beta(f(E_1)f(V_1)) \geq \beta_{\min}$, Constant β_{\min} called strength threshold value, $\beta \in [0,1]$

So uncertain attribute graph GA_{11} α - β sub-graph isomorphic to uncertain attribute graph GA_{12} . Sign $GA_{11} \subseteq_{\alpha, \beta} GA_{12}$.

THEOREM 1. If the uncertain attribute graph I GA_{11} is isomorphism to the uncertain attribute graph I GA_{12} in limiting condition α_{\max} and β_{\min} , then $P(GA_{11} \subseteq GA_{12}) \geq \delta(\alpha_{\max}, \beta_{\min})$, and $\delta(\alpha_{\max}, \beta_{\min}) = (1 - \alpha_{\max}) * \beta_{\min}$.

If the uncertain attribute graph I GA_{11} is α - β sub-graph isomorphism to uncertain attribute graph I GA_{12} , then the probability is $P(GA_{11} \subseteq GA_{12}) \geq \delta(\alpha_{\max}, \beta_{\min})$ that the uncertain attribute graph I GA_{11} is α - β sub-graph isomorphism to uncertain attribute graph I GA_{12} in the real world. As showing in definition 3.1, expectative sub-graph isomorphism uses a single constant as the limit threshold value, and α - β sub-graph isomorphism uses two constants α_{\max} and β_{\min} as the limit threshold value to replace δ . Sub-graph isomorphism has probability significance. It can be expressed as following. If the uncertain attribute graph I GA_{11} is sub-graph isomorphic to a the uncertain attribute graph I GA_{12} and $\delta(\alpha_{\max}, \beta_{\min})$ is not less than the desired threshold δ , then, uncertain attribute graph I GA_{11} is expectative sub-graph isomorphic to uncertain attribute graph I GA_{12} .

4.2. The α - β Sub-graph Isomorphism of Uncertain Attribute Graph II

DEFINITION 4.4. The uncertain attributes set of uncertain attribute graph II GA_{II} is $GA_{II}^{UC} = \{GA \mid GA \subseteq GA_{II}, 0 < P_{va} < 1, 0 < P_{ea} < 1\}$. Define $\alpha: 2^{GA_{II}^{UC}} \rightarrow (0,1]$ is error function. For the any sub-graph $GA_{II\Delta}$ of GA_{II}^{UC} , error function value is the following:

$$\alpha((GA_{II\Delta})) = \begin{cases} 1 - \prod_{GA_{II\Delta} \subseteq GA_{II}} (1 - P(ea_i))(1 - P(va_i)) & GA_{II\Delta} \neq \Phi \\ 0 & GA_{II\Delta} = \Phi \end{cases} \quad (7)$$

DEFINITION 4.5. The uncertain attributes set of uncertain attribute graph II GA_{II} is known. $GA_{II\Delta}$ is any sub-graph of GA_{II} . Definition $\beta: 2^{GA_{II\Delta}} \rightarrow (0,1]$ is strength function. For the any sub-graph $GA_{II\Delta}$ of GA_{II} , strength function is the following:

$$\beta(GA_{II\Delta}) = |^{GA_{II\Delta}} \sqrt{\prod_{GA_{II\Delta} \in GA_{IIUC}} p(va_i) \cdot P(ea_i)} \quad (8)$$

DEFINITION 4.6. Uncertain attribute graph II GA_{II1} and GA_{II2} are known quantity. Hypothesis the uncertain attribute of GA_{II1} is GA_{II1}^{UC} , if the presence $GA_{II1}' \subseteq GA_{II1}^{UC}$ and mapping function f , while satisfying the following:

- (1) $I_{\max}(GA_{II1}) - GA_{II1}'$ sub-graph isomorphic to $I_{\max}(GA_{II2})$ and f is mapping function;
- (2) $\alpha(GA_{II1}') \leq \alpha_{\max}$, Constant α_{\max} called error threshold value, $\alpha_{\max} \in [0,1]$;
- (3) $\beta(f(EA_1)f(VA_1)) \geq \beta_{\min}$, Constant β_{\min} called strength threshold value, $\beta \in [0,1]$

So uncertain attribute graph GA_{II1} α - β sub-graph isomorphic to uncertain attribute graph GA_{II2} . Sign $GA_{II1} \subseteq_{\alpha,\beta} GA_{II2}$.

4.3. The Semantics of α - β Sub-graph Isomorphism of Uncertain Attribute Graph

DEFINITION 4.7. Uncertain attribute graph GA_{P1} and GA_{P2} are known quantity. Hypothesis the uncertain attribute of GA_{P1} is GA_{P1}^{UC} , if the presence $GA_{P1}' \subseteq GA_{P1}^{UC}$ and mapping function f , while satisfying the following:

- (1) $I_{\max}(GA_{P1}) - GA_{P1}'$ sub-graph isomorphic to $I_{\max}(GA_{P2})$ and f is mapping function;
- (2) $\sqrt{\alpha(GA_{P1}')\alpha(GA_{P1})} \leq \alpha_{\max}$, Constant α_{\max} called error threshold value, $\alpha_{\max} \in [0,1]$;
- (3) $\sqrt{\beta(f(EA_1)f(VA_1)f(E_1)f(V_1))} \geq \beta_{\min}$, Constant β_{\min} called strength threshold value, $\beta \in [0,1]$

So uncertain attribute graph GA_{P1} α - β sub-graph isomorphic to uncertain attribute graph GA_{P2} . Sign $GA_{P1} \subseteq_{\alpha,\beta} GA_{P2}$.

5. Experimental Analysis

5.1. The Experimental Data and the Parameter Setting

Uncertain attribute graph data includes the query graph and uncertain attribute graph. The experiment uses 4 groups query graph GA5, GA8, GA11 and GA14, and each graph number of vertices is n . the α - β sub-graph isomorphism of uncertain attribute graph data used parameters mainly includes $(\alpha_{\max}, \beta_{\max})$, and setting up three of parameters are (0.2,0.7), (0.4,0.7) and (0.4,0.3). It mainly observes the error threshold and the time of calculation how to change in the same strength function circumstances, and the strength threshold and the time of calculation how to change in the same error threshold circumstances.

5.2. Analysis of Experimental Results

Through the experiment, it can be seen that the error threshold is smaller and shorter calculation time in the same strength threshold circumstances from the analytical Figure 1; the intensity threshold increases and the time of computational implementation the longer in the same error threshold under circumstances. Figure 2 is average execution time of before and after optimization sub-graph isomorphism judgment algorithm. The parameter of experiment is the second groups because it is a representative sample. The execution time of decision algorithm shortens in the search order optimizing after.

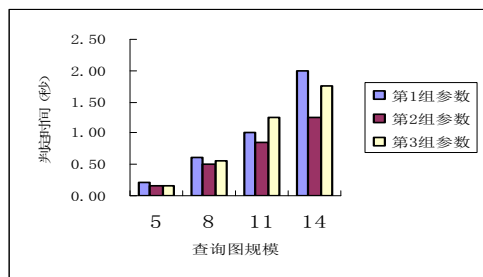


Figure 1. The Time of Computational Implementation under Threshold Circumstances

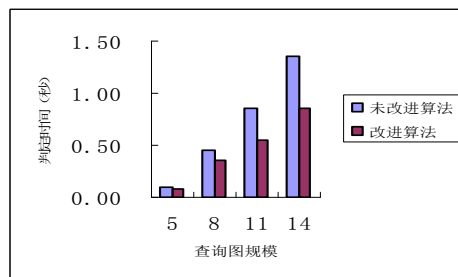


Figure 2. Average Execution Time of before and after Optimization Sub-graph Isomorphism Judgement Algorithm

6. Results and Prospects

Uncertain attribute graph expectations sub-graph isomorphism is a kind of ideal state. But it is often difficult to achieve in the practical problems. This paper focuses on defined of uncertain attribute graph $\alpha - \beta$ sub-graph isomorphism and its determination method. In setting the two threshold value, $\alpha - \beta$ sub-graph isomorphism is more close to the actual isomorphism. The determining algorithm efficiency has been greatly improved in the premise that threshold can be modulated by. $\alpha - \beta$ Sub-graph isomorphism lay a foundation for sub-graph query and community mining work in-depth development in complex network (Social Network). However, how to set the threshold to make sub-graph isomorphism is more reasonable and the actual situation more matching, it is the further research problems.

References

- [1] Zhang Shuo, Gao Hong, Li Jianzhong. Efficient Query Processing on Uncertain Graph Databases. *Chinese Journal of Computers*. 2009; 32(10): 2066-2079.
- [2] Zhang Yanan, Zou Zhaonian, Li Jianzhong. Algorithm for $\alpha - \beta$ sub-graph Isomorphism Problem on Uncertain Graph. *Intelligent Computer and Applications*. 2011; 10: 1-3.
- [3] Zhang, Chunying, Guo, Jingfeng, Chen, Xiao. *Research on random walk rough matching algorithm of attribute sub-graph*. International Conference on Advanced Materials and Computer Science, ICAMCS. 2011; 5: 297-302.
- [4] JK Cheng, TS Huang. A subgraph isomorphism algorithm using resolution original research article. *Pattern Recognition*. 1981; 13(5): 371-379.
- [5] Dong Anguo, Gao Lin, Zhao Jianbang. Algorithms for sub-graph isomorphism in graph pattern Mining. *Mathematics in Practice and Theory*. 2011; 7: 105-111.
- [6] Xie Chunxin, Wang Wei. OES: sub-graph isomorphism verification algorithm. *Computer Engineering*. 2011; 3: 46-51.
- [7] Liu Bo, Fang Bin, Zhang Shiyong, Li Zhilin. Design and implementation of subgraph isomorphism detection algorithm based on relational model. *Computer Engineering*. 2011; 11: 39-43.
- [8] Fred DePiero, David Krout. An algorithm using length-r paths to approximate subgraph isomorphism. *Pattern Recognition Letters*. 2003; 24(1-3): 33-46.
- [9] Zhou Aoying, Jin Cheqing, Wang Guoren. A Survey on the Management of Uncertain Data. *Chinese Journal of Computers*. 2009; 1: 1-16.