Ashwini. V. Urade *et al*, International Journal of Computer Science and Mobile Applications,
Vol.2 Issue. 11, November- 2014, pg. 155-158 **ISSN: 2321-8363**

Enhanced Approach for Keyword Based Search on Uncertain Graph Data: A Review

Ashwini. V. Urade
III sem M.Tech CSE, Vidarbha Institute of Technology, Nagpur, India
rupaliurade88@gmail.com

Prof. Pravin Kulurkar
Assistant Professor, Vidarbha Institute of Technology, Nagpur, India
pravin.kulurkar@gmail.com

Abstract: In many real applications, graph data is subjected to uncertainties due to incompleteness and imprecision of data. However, there is no work on keyword search over uncertain graph data even though the uncertain graphs have been widely used in real applications, such as modelling road networks, influential detection in social networks etc. Mining sub graphs is the ultimate goal of research on uncertain graph data management to retrieve the useful data from uncertain graph data. A keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. Approximate mining algorithms can be used to form sub graph from uncertain graph data based on scores at the level of keywords, data elements, element sets, and sub graphs that connect these elements. To retrieve the efficient keyword from sub graph keyword matching algorithm can be used for uncertain graph data. The objective of propose technique is to reduce the high cost of processing keyword search queries on uncertain graph data and improve the performance of keyword search, without compromising its result quality. Also o reduce processing time for keyword search in uncertain graph data.

Keywords: Database, algorithm, uncertain data, graph data

1. INTRODUCTION

In recent years, the study of keyword search technology based on Graph data has become a hot spot, and it is generally applied to the field of information retrieval. In the field of traditional graph database, the research on keyword search has already gained some achievement, but in the field of uncertain graph data, the study on keyword search has just started. , However, all graphs in the database are assumed to be certain or accurate, and in real-life applications, this assumption is often invalid. For example, RDF data can be highly unreliably due to errors in the web data or data expiration.

In the application of the data integration, it is needed to incorporate such RDF data from various data sources into an integrated database. In this case, uncertainties/inconsistencies often exist. Like In social networks, each link between any two persons is often associated with a probability that represents the uncertainty of the link or the strength of influence a person has over another person in viral marketing. In XML data (a tree or graph structure), uncertainties are incorporated in XML documents known as probabilistic XML document (p-document). Keyword searching in RDF data, social networks and XML data has many important applications.

Therefore, it is necessary to relax the strict assumption of Deterministic or well certain graphs and study keyword search over uncertain graphs. Keyword Query Analysis and Mining sub-graph pattern is the ultimate goal of research on uncertain graph data management to retrieve the useful data from uncertain graph data. The

keyword routing method can be used to route keywords only to relevant sources to reduce the high cost of processing keyword search queries over all sources. A keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism can be used for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and sub graphs that connect these elements.

To overcome these issues, we propose a new technique for searching keyword on uncertain graph data. For this we use mining algorithm for creating sub-graph from uncertain graph.

2. RELATED WORK

In literature, we study most of the recent mining and sampling techniques that have been developed in data mining domain. The filtering-and-verification framework to answer the query on uncertain graph data. In the filtering phase, we perform existence, path-based and tree-based probabilistic pruning phases, which filter out most false sub trees. In the verification, we propose a sampling algorithm to verify the candidates. To support the effective probabilistic pruning rules, we propose a probabilistic keyword index that integrates discriminative structural information with probabilistic features of uncertain graphs. [1].

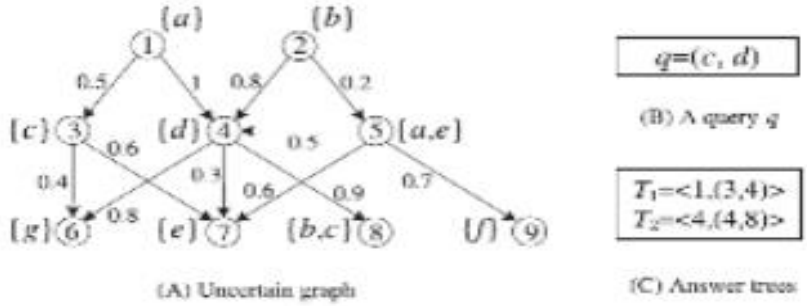

Fig.1. Example of query and answers.

A method that uses an index of the uncertain graph database to reduce the number of comparisons needed to find frequent subgraph patterns. The proposed algorithm relies on the apriori property for enumerating candidate subgraph patterns efficiently. Then, the index is used to reduce the number of comparisons required for computing the expected support of each candidate pattern. It also enables additional optimizations with respect to scheduling and early termination, that further increase the efficiency of the method. [2].

A novel method for computing top-k routing plans based on their potentials to contain results for a given keyword query. Also propose to route keywords only to relevant sources to reduce the high cost of processing keyword search queries over all sources. Employ a keyword-element relationship summary that compactly represents relationships between keywords and the data elements mentioning them. A multilevel scoring mechanism is proposed for computing the relevance of routing plans based on scores at the level of keywords, data elements, element sets, and sub-graphs that connect that connect data elements [3].

An optimized algorithm DMPUTop-k for processing most probable uncertain Top-k queries in the distributed environment. [4].

An efficient approximation algorithm to determine whether a sub-graph pattern can be output or not. This paper propose also propose efficient sampling algorithm for matching the keyword in sub-graph. [5].

All these techniques tried to cover different issues maintaining the cost of implementation but it requires more time and the high cost of processing keyword search queries on uncertain graph data.

3. PROBLEM DEFINITION

Uptill now, keyword searching is done only in certain graph database but in real application, there is uncertain graph data. However, so far, there is no work on keyword search in uncertain graph data. For keyword searching in uncertain graph database, two phases were used which are filtering and verification. For filtering purpose, there were also sub phases which are existence probabilistic prune, path based probabilistic prune and tee based probabilistic phase which consumed more time for filtering and finally verification is applied. This procedure consumed much more time so it is necessary to reduce processing time for that a new approach can be used which will also reduce the high cost of processing keyword search query over uncertain graph data. This approach greatly helps to improve the performance of keyword search, without compromising its result quality.

4. PROJECT OBJECTIVES

The objective of proposed techniques is
- To search keyword over Uncertain graph data.
-To reduce the high cost of processing keyword search queries on uncertain graph data.
-To improve the performance of keyword search, without compromising its result quality.
-To reduce processing time for keyword search in uncertain graph data.

5. INVESTIGATIONAL OUTCOME

To achieve the objective of this project, we have proposed following techniques;
-While creating the sub graph over uncertain graph data, we are going to find an approximate set of frequent sub graph patterns in graph database by using Approximate mining algorithm like MUSE(Mining uncertain Sub-graph Pattern)
-Then we find the efficient keyword in this subgraph by using Efficient sampling algorithm. This algorithm is used to verify the candidate

6. CONCLUSION

This review paper proposes a technique to perform keyword based search on uncertain graph using mining and sampling algorithms. It also improves the performance of keyword search, without compromising its result quality.

REFERENCES

[1] Ye Yuan, Guoren Wang, Lei Chen, and Haixun Wang," Efficient Keyword Search on Uncertain Graph Data", IEEE Transactions On Knowledge And Data Engineering, Vol. 25, No. 12, December 2013.

[2] O. Papapetrou, E. Ioannou, and D. Skoutas, "Efficient Discovery of Frequent Sub-graph patterns in Uncertain Graph Database" Proc. 14th Int'1 conf. Extending Database Technology (EDBT), 2011.

[3] Thanh Tran And Lei Zhang," Keyword Query Routing", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 2, February 2014.

[4] Zhao Zhibin, Yu Yang, Bao Yubin, Yu Ge," Optimizing Distributed Top-k Queries on Uncertain Data", IEEE, 2013.

[5] Z. Zou, H. Gao, J. Li, and S. Zhang, "Mining Frequent Subgraph Patterns from Uncertain Graph Data," IEEE Trans. Knowledge and Data Eng., vol. 22, no. 9, pp. 1203-1218, Sept. 2010.

[6] K. Yi, F. Li, D. Srivastava, and G. Kollios, "Efficient Processing of Top-K Queries in Uncertain Databases," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE), 2008.

[7] G. Kollios, M. Potamias, and E. Terzi, "Clustering Large Probabilistic Graphs," IEEE Trans. Knowledge and Data Eng ,Feb. 2013.

[8] E. Adar and C. Re, "Managing Uncertainty in Social Networks," IEEE Data Eng. Bull., vol. 30, no. 2, pp. 15-22, June 2007

[9] M. Potamias, F. Bonchi, A. Gionis, and G. Kollios,"K-Nearest Neighbours' in Uncertain Graph, " Proc. VLDB Endowment, vol. 3, pp. 997-1008,2010

[10] Mayssam Sayyadian, Hieu LeKhac, AnHai Doan, Luis Gravano," Efficient Keyword Search Across Heterogeneous Relational Databases", IEEE 2007.

[11] H. He, H. Wang, J. Yang, and P.S. Yu, "Blinks: Ranked Keyword Searches on Graph" Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 305-316, 2007.

[12] B. B. Dalvi, M.Kshirsagar, and S. Sudarshan, "Keyword Search on External Memory Data Graph," Poc. VLDB Endowment, vol. 1, pp. 1189-1204, 2008.

[13] B.K.K. Golenberg and Y. Sagiv, "Keyword Proximity Search in Complex Data Graph," Proc. ACM SIGMOD Int'l Conf. Management of Data, 2008.

