

Retrieval of Relevant Graphs

¹Dipti S. Gupta & ²Prof. Dr. S. M. Kamalapur

¹M.E. Student, ²Professor

¹Department of Computer Engineering

¹K. K. Wagh Institute of Engineering Education & Research, Savitribai Phule Pune University,
Nashik, India.

Received: December 07, 2018

Accepted: January 17, 2019

ABSTRACT: Graph mining is one of the novel approaches for mining the graph structure. There are different graph mining approaches which can be used to solve the graph retrieval problem. The graph retrieval is the process of finding relevant graphs, according to the user requests, from a collection of graphs. A swarm intelligence approach, which is a form of artificial intelligence based on the collective behavior of decentralized and self-organized systems, can be used in the graph retrieval problem. Many complex and real problems can be solved using swarm optimization, which is a bio-inspired method. The power of graph mining techniques can be exploited for extracting relevant knowledge, which can be used later by the swarms. Then, the swarms can use the extracted knowledge to explore the solution's space. The performance mainly depends on the run-time and the quality of returned graphs.

Key Words: Artificial intelligence, bio-inspired method, graph mining, graph retrieval, swarm intelligence, swarm optimization.

I. INTRODUCTION

A graph is a general tool for modeling the structural relationships between different data objects. It has been prevalently used in a wide range of application domains. In numerous applications, such as chemical compound structures in chemistry, attributed graphs in image processing, electrical circuits in electricity, food chains in ecology, protein interaction networks in biology, road networks in transport, and topological networks on the web, data are usually represented by graphs. With the increasing popularity of graph databases in various applications, graph mining has become an active and most significant research area.

Graph mining is one of the novel approaches for mining the graph structure. In recent years, the graph retrieval problem has attracted much attention. The graph retrieval is the process of finding relevant graphs, according to the user requests, from a collection of graphs. The graph retrieval problem, when directly dealing with a large number of graphs, has extremely high runtime. So, to handle the graph retrieval problem, bio-inspired approaches can be used.

The bio-inspired approaches like evolutionary-based approaches and the swarm intelligence approaches transform the graph retrieval problem into an optimization problem and consider the collection of graphs as a space of solutions. Optimization is one of the most challenging research areas spanning across different fields such as computer science, operations research and engineering. It is the process by which an optimum is achieved. The term swarm intelligence comes from the field of artificial intelligence, in which bees, ants, insects or bird behavior are analyzed. This natural behavior of different swarms is adopted in the field of computer science in order to solve various optimization problems. In particular, swarm intelligence algorithms refer to the branch of optimization algorithms that simulate and model the imprecision, randomized, and stochastic features of these physical, chemical, or biological elements in arriving at marvelous solutions. Some of the known bio-inspired techniques used in solving optimization problems are bees swarm optimization (BSO), ant colony optimization (ACO), particle swarm optimization (PSO), firefly optimization (FFA), honey bees mating optimization for TSP (HBMOTSP), African buffalo optimization (ABO), bat algorithm (BA), genetic algorithm (GA), adaptive simulated annealing with greedy search (ASA-GS). The intensification and diversification strategies of these approaches allow finding an approximate subset of graphs in a reasonable computational time. Moreover, the communication between the swarms permits to find high-quality solutions compared to the evolutionary ones. Nevertheless, the process is still stochastic, and when the space of solutions is large, the swarms are disoriented, which degrades the quality of the final subset of graphs returned by the swarms. Thus, there is a need to decompose the initial problem into many subproblems, where each of which could be solved independently.

Motivated by the success of BSO in dealing with many optimization problems, the aim is to investigate a new BSO algorithm which is guided by graph mining techniques for solving the graph retrieval problem.

The organization of the paper is as follows. Section II discusses the related work for the graph retrieval. In Section III, performance evaluations are presented. Section IV discusses the conclusions. Finally, Section V presents acknowledgement.

RELATEDWORK

Different approaches have been used in the literature for retrieval of relevant graphs. There are several methods which can be used to solve the graph retrieval problem. The aim of these approaches is to provide approximate solutions in a reasonable time complexity. The approaches to retrieve the relevant graphs are presented here.

A. Method based on maximum common subgraph:

Zhu et al. (2012) [1] developed algorithms for finding top-k similar graphs in graph databases. The problem of finding top-k graphs in a graph database that are most similar to a query graph has been studied and a similarity measure based on maximum common subgraph (MCS) has been introduced. But it reduces the number of MCS computations rather than making an MCS computation faster.

B. Greedy approach:

Ranu et al. (2014) [2] formulated the problem of top-k representative queries on graph databases. A greedy constant factor approximation of the optimal answer set was designed. To achieve scalability, an index structure called NB-Index was designed to facilitate fast answering of queries.

C. Multi-layered indexing approach:

Liang et al. (2017) [3] considered the similarity search problem that retrieves relevant graphs from a graph database under the well-known graph edit distance (GED) constraint. A parameterized, partition-based GED lower bound was proposed that can be instantiated into a series of tight lower bounds towards synergistically pruning false-positive graphs from G before costly GED computation is performed. An efficient, selectivity-aware algorithm was designed to partition graphs of G into highly selective subgraphs. They were further incorporated in a cost-effective, multi-layered indexing structure, MLIndex (Multi-Layered Index), for GED lower bound crosschecking and false-positive graph filtering with theoretical performance guarantees.

D. Partition-based approach:

Zhao et al. (2017) [4] developed a partition-based approach for efficient structure similarity searches. A framework involving a novel filtering scheme was proposed based on variable-size non-overlapping partitions. A dynamic partitioning technique was devised with enhanced matching condition and mismatching partition recycling to strengthen the pruning power. Moreover, an extension-based verification algorithm leveraging matching partitions was conceived to expedite the final verification. For index construction, a cost-aware graph partitioning method was proposed to optimize the index. In addition, the index was extended to form a hierarchical inverted index, in order to support top-k similarity queries. Based on it, tailored search procedure with look-ahead and computation-sharing strategies were devised.

E. Discrete particle swarm optimization approach:

Gong et al. (2016) [5] introduced a discrete particle swarm optimization algorithm for resolving high-order graph matching problems, which incorporates several re-defined operations, a problem-specific initialization method based on heuristic information, and a problem-specific local search procedure.

F. Bees swarm optimization approach:

Djnouri et al. (2018) [6] developed an algorithm for bees swarm optimization guided by data mining techniques for document information retrieval. Useful knowledge is discovered by using data mining techniques to solve the Document Information Retrieval problem, and then swarms use this knowledge to explore the whole space of documents intelligently. This approach is limited to only document information retrieval whereas it can be used for other optimization problems such as graph problems.

III. PERFORMANCE EVALUATION

In the existing graph retrieval methods, the time complexity is high. However, using bees swarm optimization can reduce the run-time. The performance of graph retrieval problem is evaluated through a number of experiments in terms of runtime performance and the quality of returned graphs. Regarding the evaluation measure, F-measure is used. It is based on Recall and Precision and it is the well-known measure for the retrieval problem:

- Recall: It is the ratio of the number of relevant graphs retrieved to the total number of all relevant

graphs.

$$Recall = \frac{|RGR|}{|ARG|} \quad (1)$$

where, RGR is the set of the Relevant Graphs Retrieved and ARG is the set of All Relevant Graphs.

- Precision: It is the ratio of the number of relevant graphs retrieved to the total number of returned graphs.

$$Precision = \frac{|RGR|}{|RG|} \quad (2)$$

where, RG is the set of all returned graphs.

- F-measure: This measure allows to combine precision and recall measures, which is defined as follows:

$$F\text{-measure} = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (3)$$

IV. CONCLUSIONS

In this paper, study of different graph retrieval approaches, their methods and performance evaluation measures has been presented. The main focus is on the study of the bees swarm intelligence approach for retrieval of relevant graphs from the graph database according to user request. The swarms explore the space of graphs using knowledge discovery by two graph mining techniques which are K-means clustering algorithm and frequent subgraph mining algorithm. These two techniques allow dividing a collection of graphs into several sub-collections, each of which is featured by the set of frequent subgraphs between the graphs belonging to it. The swarms then use this knowledge to explore each cluster deeply for any user's request. The best top-k relevant graphs are retrieved after applying bees swarm optimization.

V. ACKNOWLEDGEMENT

I would like to express my gratitude to my guide Prof. Dr. S. M. Kamalapur, Professor, Department of Computer Engineering, K.K.W.I.E.E.R., Nashik for giving me a moral support, valuable guidance and encouragement in making this literature survey. A special thanks to Prof. Dr. K. N. Nandurkar, Principal, K.K.W.I.E.E.R., Nashik and Prof. Dr. S. S. Sane, Head of Department of Computer Engineering, K.K.W.I.E.E.R., Nashik for their kind support and suggestions. It would not have been possible without the kind support. I would like to extend my sincere thanks to all the faculty members of the department of computer for their help.

REFERENCES

1. Y. Zhu, L. Qin, J. X. Yu, and H. Cheng. Finding Top-K Similar Graphs in Graph Databases. In EDBT, pages 456–467, 2012.
2. Ranu, Sayan, Minh Hoang, and Ambuj Singh. "Answering top-k representative queries on graph databases." Proceedings of the 2014 ACM SIGMOD international conference on Management of data. ACM, 2014.
3. Liang, Yongjiang, and Peixiang Zhao. "Similarity search in graph databases: A multi-layered indexing approach." Data Engineering (ICDE), 2017 IEEE 33rd International Conference on. IEEE, 2017.
4. Zhao, Xiang, et al. "Efficient structure similarity searches: a partition-based approach." The VLDB Journal—The International Journal on Very Large Data Bases 27.1 (2018): 53-78.
5. Gong, Maoguo, et al. "Discrete particle swarm optimization for high-order graph matching." Information Sciences 328 (2016): 158-171.
6. Djenouri, Youcef, Asma Belhadi, and Riadh Belkebir. "Bees swarm optimization guided by data mining techniques for document information retrieval." Expert Systems With Applications 94 (2018): 126-136.