香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. ZHU Xuliang
Date	31 July 2023 (Monday)
Time:	10:30 am – 12:30 nn (35 mins presentation and 15 mins Q & A)
Venue:	1) DLB637, 6/F, David C Lam Building, Shaw Campus 2) ZOOM (Meeting ID: 916 5509 9624) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 6:00 pm, 30 July 2023)

Interactive Search and Summarization on Hierarchical Graphs

Abstract

The hierarchical graph is a data structure to represent entities and general-certain relationships between entities, which has been widely used in real applications, such as ImageNet, disease ontology, Wikipedia categories, ACM computing classification system, and so on. Due to massive terminologies and complex structures in a large hierarchical graph, it is challenging to resolve hierarchical graph analytics problems algorithmically, even with the help of leveraging human intelligence, such as object categorization, library classification, labelling, etc. In this thesis, we study two important problems of hierarchical graph analytics: data summarization and interactive search. Data summarization uses a small-sized answer to give a direct and human-friendly overview of the hierarchical graph data being analysed, which is useful for understanding and visualization. Interactive search leverages human intelligence to categorize target labels in a hierarchy, which has applications in image classification, product categorization, and database search.

First, we study the data summarization on hierarchical tree. We motivate and formulate our kWTS-problem as selecting a diverse set of k nodes to summarize a hierarchical tree T with weighted terminologies. To depict diverse summarization and important vertices, we design a summary score function for capturing vertices' diversity coverage and structure correlation. To efficiently tackle it, we first propose an efficient greedy algorithm GTS and an optimal algorithm OTS. In addition, we propose a useful optimization technique of tree reduction to remove useless nodes with zero weights and shrink the tree into a smaller one, which ensures the efficiency acceleration of both GTS and OTS in real-world datasets.

Next, we study the data summarization on hierarchical DAGs. Similar to kWTS-problem, we propose a problem of finding k representative vertices to summarize a hierarchical DAG called kDAG-problem. The studied kDAG-problem is theoretically proven to be NP-hard. To efficiently tackle it, we propose a greedy algorithm with an approximation guarantee, which iteratively adds vertices with the large summary contributions into answers. To further improve the effectiveness and efficiency, we propose a subtree extraction based method EXT-Greedy and a scalable algorithm k-PCGS based on candidate pruning and DAG compression. Extensive experimental results on real-world datasets show the effectiveness and efficiency of our proposed algorithms on both hierarchical tree and DAG datasets.

Last, we study the interactive search on hierarchical graphs. We propose a new problem of budget constrained interactive graph search for multiple targets called kBM-IGS problem. To tackle the kBM-IGS problem, we develop a novel framework to ask questions using the best vertex with the largest expected gain, which provides a balanced trade-off between target probability and benefit gain. Based on the framework, we propose an efficient algorithm STBIS to handle the SingleTarget problem and a dynamic programming based method kBM-DP to tackle the MultipleTargets problem. Experiments on large real-world datasets with ground-truth targets verify both the effectiveness and efficiency of our proposed methods.

***** ALL INTERESTED ARE WELCOME *****