香港浸會大學理學院

HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. LIAO Xuankun
Date	23 August 2024 (Friday)
Time:	2:30 pm – 4:30 pm (35 mins presentation and 15 mins Q & A)
Venue:	ZOOM (Meeting ID: 933 4798 3243) (The password and direct link will only be provided to registrants)
Registration:	https://bit.ly/bucs-reg (Deadline: 12:00 nn, 22 August 2024)

Cohesive Subgraph Computation over Large-Scale Directed Graphs

Abstract

Directed graphs are widely used data structures for representing large-scale entities and the relationships among them in real-world applications, including online social networks, biological networks, and financial networks. Computing cohesive subgraphs from directed graphs has gained significant attention. Despite extensive research efforts in this field, existing algorithms face certain limitations. First of all, the sheer size of the graph data necessitates huge memory space, which often exceeds the memory capacity of a single machine. Existing centralized algorithms cannot handle such huge graphs. Besides, directed graphs are typically highly dynamic, with frequent insertions and deletions of vertices and edges. Existing algorithms suffer from efficiency issues when handling frequent graph updates. To address these limitations, in this dissertation, we conduct a comprehensive exploration of efficient techniques for computing cohesive subgraphs over large-scale directed graphs.

First, we study the problem of distributed D-core decomposition over large directed graphs. Given a directed graph G and integers k and l, a D-core is the maximal subgraph $H \subseteq G$ such that for every vertex of H, its in-degree and out-degree are no smaller than k and l, respectively. The problem of distributed D-core decomposition is to compute all possible D-cores using a collection of machines. To tackle this problem, we first introduce the concept of anchored coreness, based on which a novel H-index-based algorithm is proposed. We further devise a novel concept of skyline coreness and design an elegant D-index to compute skyline coreness. These algorithms are implemented under both block-centric and vertex-centric distributed graph processing frameworks. The proposed models and algorithms are validated through extensive theoretical analysis and empirical evaluation.

Second, we study the maintenance of D-core in dynamic directed graphs. We start by introducing our theoretical findings to find D-cores that require updates. Based on these theoretical results, we propose a local-search-based algorithm with non-trivial optimizations for efficiently processing single-edge deletions and insertions. We further present an H-index-based algorithm to handle batch updates. The comprehensive empirical studies over real-world graphs demonstrate the efficiency of our proposed algorithms.

Furthermore, many applications involve continuously generated streaming graphs with unbounded edges arriving at a high speed. Hence, we finally explore the problem of community search over streaming directed graphs utilizing the model of D-truss. Given two integers k_c and k_f, the D-truss model, also denoted as (k_c k_f)-truss, considers the cycle triangle and flow triangle and ensures that each edge can form cycle triangles with k_c vertices and flow triangles with k_f vertices, respectively. To efficiently retrieve D-truss-based communities in streaming directed graphs, we present a peeling-based algorithm that iteratively removes edges failing to meet the support constraints. To improve the peeling-based algorithm's performance, three optimization strategies have been proposed. Moreover, we introduce the concept of D-truss peeling order. Based on this, we propose a novel order-based algorithm that preserves the community by maintaining the deletion order of edges in the peeling algorithm. The effectiveness of our proposed algorithms is demonstrated through experiments conducted on real-world graphs.

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