

TÓM TẮT THÔNG TIN VỀ LUẬN ÁN

Tên đề tài luận án: Khai thác đồ thị con trên đồ thị có trọng số

Ngành: Khoa học máy tính

Mã số ngành: 62480101

Họ tên nghiên cứu sinh: Lê Thị Ngọc Thảo

Khóa đào tạo: 2015

Người hướng dẫn khoa học: GS.TS. Lê Hoài Bắc và PGS.TS. Võ Đình Bảy

Cơ sở đào tạo: Trường Đại học Khoa học Tự nhiên - ĐHQG.HCM

1. TÓM TẮT NỘI DUNG LUẬN ÁN:

- Trên cơ sở phân tích và đánh giá những ưu nhược điểm của các phương pháp trước đây thường được dùng để giải bài toán khai thác đồ thị con phổ biến trên đồ thị có trọng số, mục tiêu của luận án là nghiên cứu và đề xuất những kết quả lý thuyết, phương pháp tiếp cận và thuật toán mới nhằm giải quyết hiệu quả bài toán khai thác tập đồ thị con phổ biến trên một đồ thị có trọng số đỉnh.
- Để đạt được mục tiêu, luận án đề xuất một phương pháp tiếp cận dựa trên việc kế thừa thuật toán GraMi được đề xuất bởi Elseidy và các đồng sự. Phương pháp tiếp cận được đưa ra dựa trên ý tưởng sử dụng mô hình CSP (Constraint Satisfaction Problem) để biểu diễn dữ liệu đồ thị và sử dụng lại thuật toán GraMi để giải quyết bài toán khai thác đồ thị con phổ biến trên một đồ thị, dựa trên cơ sở bổ sung thêm yếu tố về trọng số vào mô hình này.

2. NHỮNG KẾT QUẢ MỚI CỦA LUẬN ÁN:

- Thuật toán WeGraMi và OWGraMi sử dụng độ đo MaxMin thỏa tính chất DCP (Downward Closure Property) để tính trọng số của đồ thị con và áp dụng chiến lược cắt tỉa không gian tìm kiếm dựa trên ngưỡng trọng số để khai thác hiệu quả các đồ thị con phổ biến có trọng số.
- Các chiến lược hiệu quả đã được đề xuất để tỉa sớm các cạnh không thể đạt đến ngưỡng trọng số đã cho và tái sử dụng lại trọng số của các đồ thị con đã tính nhằm tăng tốc khai thác đồ thị con phổ biến có trọng số trong thuật toán OWGraMi.
- Để mở rộng phương pháp giải bài toán khai thác đồ thị con phổ biến có trọng số với hướng tiếp cận sử dụng một độ đo khác với độ đo MaxMin, luận án đã đề xuất thuật toán AWeGraMi sử dụng độ đo trung bình (AveMin) để tính trọng số của đồ thị con. Vì độ đo AveMin không thỏa tính chất DCP, luận án đồng thời sử

dùng MaxMin làm chặn trên (upper bound) nhằm tĩa sớm các đồ thị con phổ biến không thỏa ngưỡng trọng số.

3. CÁC ỨNG DỤNG/ KHẢ NĂNG ỨNG DỤNG TRONG THỰC TIỄN HAY NHỮNG VẤN ĐỀ CÒN BỎ NGỎ CẦN TIẾP TỤC NGHIÊN CỨU

- Phát triển các phương pháp khai thác đồ thị con có trọng số phổ biến bằng cách sử dụng các phương pháp tiếp cận khác cũng như nghiên cứu sử dụng các độ đo khác để tính trọng số của đồ thị con sao cho đảm bảo tính chất DCP nhằm phục vụ cho mục tiêu cắt tĩa sớm không gian tìm kiếm khi khai thác đồ thị con phổ biến.
- Mở rộng đối tượng nghiên cứu trọng số của đồ thị sang trọng số trên cạnh hoặc kết hợp giữa trọng số cạnh và trọng số đỉnh hoặc phát triển mô hình đồ thị đa trọng số nhằm phục vụ cho mô phỏng việc thể hiện các dữ liệu trọng số đa dạng trong thực tiễn.
- Hướng đến phát triển các phiên bản song song của các thuật toán đã được luận án đề xuất, sử dụng các hệ thống tính toán hiệu năng cao để triển khai trên môi trường phân tán nhằm áp dụng hiệu quả cho việc khai thác dữ liệu lớn.

CÁN BỘ HƯỚNG DẪN

NGHIÊN CỨU SINH
(Ký tên, họ tên)

GS.TS. Lê Hoài Bắc

Lê Thị Ngọc Thảo

PGS.TS. Võ Đình Bảy

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO
HIỆU TRƯỞNG

THESIS INFORMATION

Thesis title: Frequent subgraph mining in a weighted graph

Specialty: Computer Science

Code: 62480101

Name of PhD Student: Le Thi Ngoc Thao

Academic year: 2015

Supervisor: Prof. Le Hoai Bac, Assoc. Prof. Vo Dinh Bay

At: UNIVERSITY OF SCIENCE, VNU-HCMC

1. SUMMARY

- Based on the analysis and assessment of the advantages and disadvantages of previous methods used to solve the problem of mining frequent subgraphs on a weighted graph, the aim of the thesis is to research and propose new theoretical foundations, approaches and algorithms to effectively solve the problem of mining common subgraphs on a graph with weighted nodes.
- The thesis proposes an approach based on inheriting the GraMi algorithm proposed by Elseidy et al. The proposed approach uses the CSP (Constraint Satisfaction Problem) model to represent graph data and applies the GraMi algorithm to solve the frequent subgraph mining problem on a weighted graph, by adding a weight factor to this model.

2. NOVELTY OF THESIS

- The WeGraMi and OWGraMi algorithms use the MaxMin measure satisfying the DCP (Downward Closure Property) to calculate the weights of the subgraph and apply a search space pruning strategy based on the weight threshold in the process of mining the weighted frequent subgraph.
- Effective strategies have been proposed for early pruning of edges that cannot reach the weight threshold and reusing the weights of subgraphs computed in previous steps to speed up weighted frequent subgraph mining in the OWGraMi algorithm.
- In order to extend the method of solving the weighted frequent subgraph mining problem with another measure, the thesis has proposed the AWeGraMi algorithm using the average measure (AveMin) to calculate the weight of the subgraph. Because the AveMin measure does not satisfy the DCP, the thesis has simultaneously used MaxMin as the upper bound to early pruning frequent subgraphs that do not satisfy the weight threshold.

3. APPLICATIONS/ APPLICABILITY/PERSPECTIVE

- Developing a weighted frequent subgraph mining method by using other approaches as well as researching the use of other measures to calculate the weight of the subgraph so that the DCP is satisfied, to be able to early prune the search space when mining frequent subgraphs.
- Expanding the researched object from the vertex weights of the graph to the weights on the edge, or a combination of edge weights and vertex weights or multi-weight graph models to serve to simulate data representation of diversity weights in real life.

This work moves towards developing parallel processing versions of the algorithms proposed by the thesis, using high-performance computing systems to deploy in a distributed environment to effectively carry out big data mining.

SUPERVISORS

Prof. Le Hoai Bac

Assoc. Prof. Vo Dinh Bay

PhD STUDENT

Le Thi Ngoc Thao

**CONFIRMATION OF THE UNIVERSITY OF SCIENCE
PRESIDENT**