



Subgraph isomorphism: identifying a given subgraph in a larger graph

Hứa sẽ qua môn

Ho Chi Minh City University of Technology

2022.12

Outline

1 Cơ sở lý thuyết

2 Hướng tiếp cận

3 Thực nghiệm

4 Kết luận



1. Đồ thị

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- Đồ thị: Gồm một tập các đỉnh không rỗng được nối với nhau bằng các cạnh, được chia thành các loại như sau:
 - Đồ thị vô hướng
 - Đồ thị có hướng
 - Đồ thị lai
 - Đa đồ thị
 - Đơn đồ thị



2. Biểu diễn đồ thị

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- Biểu diễn đồ thị bằng ma trận kề
- Biểu diễn đồ thị bằng ma trận liên thuộc
- Biểu diễn đồ thị bằng danh sách cạnh (cung)
- Biểu diễn đồ thị bằng danh sách kề
- Biểu diễn đồ thị bằng hình học



4. Đồ thị đẳng cấu và đồ thị con đẳng cấu

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- Hai đồ thị G_1 và G_2 được gọi là đẳng cấu với nhau, ký hiệu là $G_1 \approx G_2$, nếu tồn tại một hàm song ánh $f: V(G_1) \rightarrow V(G_2)$ sao cho mỗi quan hệ liên kề được bảo toàn.
- Đồ thị $G=(V,E)$ là đồ thị con đẳng cấu của đồ thị $G'=(V',E')$ nếu $\exists S$ sao cho $S \subseteq G'$ và $S \approx G$. Khi đó, đồ thị G' được gọi là đồ thị đầu vào và đồ thị G là đồ thị mẫu (model graph).



5. Tam giác, k-truss, và ý nghĩa của chúng đối với đồ thị bạn bè

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- Tam giác: Một tập hợp gồm ba đỉnh, trong đó mỗi đỉnh có mối quan hệ kề với tất cả các đỉnh khác. Các ứng dụng của việc đếm tam giác có thể kể đến tính toán hệ số phân cụm (cục bộ và toàn cục), tính bắc cầu và phát hiện các cộng đồng.
- K-truss: Một đồ thị con mà mỗi cạnh của tập con đó thuộc ít nhất $(k - 2)$ tam giác được tạo từ các điểm của truss. Ý nghĩa của Ktruss đối với mạng lưới xã hội:
 - Phân loại cộng đồng và mối quan hệ
 - Phân tích sự lan truyền thông tin
 - Ứng dụng trong chiến dịch quảng cáo và tiếp thị
 - Dự đoán tương tác tương lai
 - Tích hợp với chiến lược xã hội



6. Ý nghĩa bài toán đếm tam giác và phân rã truss đối với bài toán đồ thị đẳng cấu

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- Việc đếm tam giác có thể được coi là một trường hợp đặc biệt của bài toán đẳng cấu đồ thị con với tam giác là đồ thị con. Các thuật toán tìm hình tam giác không tìm thấy các đồ thị con trùng khớp chung nhưng một phần của các thuật toán có thể tìm được.
- Các thuật toán tìm k-trusses có thể được sử dụng để mô tả đặc tính của đồ thị (con). Nếu đặc điểm k-truss của hai đồ thị (phụ) không nhất quán thì điều này chứng tỏ rằng các đồ thị (con) này không đẳng cấu. Đặc điểm k-truss nhất quán chỉ ra rằng cần phải kiểm tra bổ sung để xác định xem có tồn tại đẳng cấu hay không.



1. Đếm tam giác

a. Các hướng tiếp cận thông dụng

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- Tính toán dựa trên việc nhân ma trận
- Đếm và liệt kê tam giác
- Đếm xấp xỉ tam giác (Approximate triangle counting)



1. Đếm tam giác

b. Thuật toán Wolf,et. al

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

Đầu vào: Ma trận kề A và ma trận liên thuộc E của đồ thị lớn G .

Đầu ra: Số lượng tam giác n_T có trong đồ thị G .

Mã giả

Algorithm 1 Triangle counting

```
C = A × E
nb = find(C == 2)
 $n_T = \frac{nb}{3}$ 
return  $n_T$ 
```



2. Phân rã Truss

a. Các hướng tiếp cận thông dụng

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

- **Phân rã truss:** quá trình chia nhỏ đồ thị thành các truss nhỏ hơn. Mục tiêu là tìm cách tối ưu hóa cấu trúc truss để hiểu rõ hơn về mối quan hệ và tính chất của các phần trong đồ thị.
 - Thuật toán I/O-efficient
 - Thuật toán In-Memory



2. Phân rã Truss

b. Thuật toán triển khai trên đồ thị

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

Đầu vào: Ma trận kề A và ma trận liên thuộc E của đồ thị lớn G .

Đầu ra: Ma trận liên thuộc của K -truss tối đa.

Mã giả

Algorithm 2 Array based implementation of k -Truss algorithm

```
R =  $A \times E$ 
s =  $1 \times (\mathbf{R} == 2)$ 
x = find( $s < k-2$ )
while x is not empty do
     $E_x = E(:, x)$ 
     $pos \leftarrow$  nonzero indices each column of  $E_x$ 
     $A(pos(:, 0)pos(:, 1)) = A(pos(:, 1)pos(:, 0)) = 0$  {Cập nhật ma trận  $A$ }
     $E = E(:, x_c)$  { $x_c$  là phần bù của  $x$  trong tập cạnh của đồ thị}
     $R = A \times E$ 
     $s = 1 \times (\mathbf{R} == 2)$ 
    x = find( $s < k-2$ )
end while
return  $E$ 
```



3. Xử lí dữ liệu

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

Trong phạm vi assignment này, đầu vào là đơn đồ thị vô hướng với cạnh có trọng số mặc định bằng 1 được cho với 2 tệp dữ liệu biểu diễn ma trận kề và ma trận liên thuộc và lưu dưới định dạng tệp .tsv

2	1	1
3	1	1
4	1	1
15	1	1
26	1	1
37	1	1
48	1	1
59	1	1
70	1	1
81	1	1
92	1	1
103	1	1
114	1	1

(a) File adj

1	1	1
2	1	1
1	2	1
3	2	1
1	3	1
4	3	1
8	4	1
9	4	1
1	5	1
15	5	1

(b) File inc



1. Thực nghiệm kiểm thử tính đúng đắn của thuật toán trên dataset có số lượng nhỏ

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

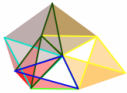



Tổng quan dataset: Để kiểm thử tính đúng đắn của thuật toán, một số dataset nhỏ với kích thước như sau được chạy thử.

Dataset	Số đỉnh	Số cạnh
1	9	22
2	11	20
3	19	27
4	28	43
5	38	75



1. Thực nghiệm kiểm thử tính đúng đắn của thuật toán trên dataset có số lượng nhỏ

Thuật toán đếm tam giác

Dataset	Kết quả thuật toán	Kiểm thử
1	Number of triangle: 19.0	
2	Number of triangle: 7.0	
3	Number of triangle: 2.0	
4	Number of triangle: 5.0	

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm



Kết luận



1. Thực nghiệm kiểm thử tính đúng đắn của thuật toán trên dataset có số lượng nhỏ

Thuật toán phân rã k-truss

Kết quả chạy thử trên tập dataset 1,2 và 5 với giá trị $k=4$:

Dataset	Kết quả thuật toán
1	
2	
5	Không tồn tại 4-truss



2. Áp dụng thuật toán trên dataset thực tế từ mạng xã hội Facebook

Tổng quan dataset

- Dataset này biểu diễn đồ thị cấu thành từ tập hợp các vòng tròn bạn bè trên Facebook, và được thu thập dữ liệu từ người dùng ẩn danh thông qua khảo sát.



2. Áp dụng thuật toán trên dataset thực tế từ mạng xã hội Facebook

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

Bài toán đếm tam giác

Kết quả từ thuật toán:

```
Number of triangle:  
1612010.0
```



2. Áp dụng thuật toán trên dataset thực tế từ mạng xã hội Facebook

Cơ sở lý thuyết

Hướng tiếp cận

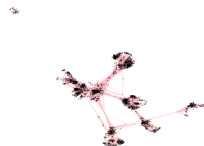
Thực nghiệm

Kết luận

■ $k=3$



■ $k=5$



2. Áp dụng thuật toán trên dataset thực tế từ mạng xã hội Facebook

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

■ $k=30$



■ $k=50$



2. Áp dụng thuật toán trên dataset thực tế từ mạng xã hội Facebook

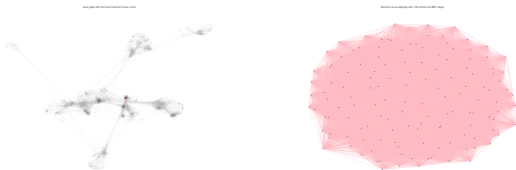
Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

■ $k=97$



■ $k=98$



5. Kết luận

Cơ sở lý thuyết

Hướng tiếp cận

Thực nghiệm

Kết luận

Thuật toán được sử dụng trong bài nghiên cứu này áp dụng chủ yếu các phép tính toán đại số tuyến tính, chủ yếu là nhân ma trận, vì thế ngoài ưu điểm là dễ hiện thực hóa thuật toán và tinh chỉnh cho phù hợp thì thời gian thực hiện của thuật toán khá lâu cũng như sẽ tốn nhiều dữ liệu cho việc tính toán và lưu trữ ma trận. Nhằm khắc phục điều này, những hướng đi sau được đề xuất cho các bài nghiên cứu sắp tới, bao gồm:

- Tối ưu hóa định dạng lưu trữ cho các đồ thị thưa
- Sử dụng mô hình khác để phù hợp với các đồ thị mạng lưới xã hội thực tế với kích thước lớn
- Tối ưu hóa việc trực quan hóa đồ thị để làm rõ hơn mối quan hệ giữa cá nhân với cá nhân và giữa cộng đồng với cộng đồng



Thank You!