Tạp chí Tin học và Điều khiển học, T. 18, S. 3 (2002), 231–236

CÁC LỚP SIÊU ĐỒ THỊ PHI CHU TRÌNH VÀ CÁC THUẬT TOÁN ĐOÁN NHẬN CHÚNG

NGUYỄN VĂN ĐỊNH

Abstract. In [1] and [4], θ-acyclic hypergraphs were introduced, where θ is Berge, α, β, or γ. Several desirable properties of these corresponding acyclic database schemes have been established.

In this paper we give a simple and clear presentation of the above classes of acyclic hypergraphs and the algorithms for their recognition.

Tóm tắt. Trong [1] và [4] đã định nghĩa một loạt các kiểu siêu đồ thị phi chu trình, bao gồm: *Berge-phi chu trình, α-phi chu trình, β-phi chu trình, γ-phi chu trình.* Nhiều tính chất mong muốn của các lược đồ CSDL phi chu trình tương ứng cũng đã được nhiều tác giả nghiên cứu và chứng minh.

Trong bài báo này, các loại siêu đồ thị phi chu trình sẽ được giới thiệu một cách ngắn gọn và rõ ràng cùng với các thuật toán đoán nhận chúng.

1. SIÊU ĐỒ THỊ PHI CHU TRÌNH VÀ LƯỢC ĐỒ CSDL PHI CHU TRÌNH

Định nghĩa 1. Trong mô hình CSDL quan hệ, một lược đồ CSDL $\mathbf{R}$ trên một tập thuộc tính U là một tập các lược đồ quan hệ, ký hiệu $\mathbf{R} = \{R_1, R_2, ..., R_m\}$ sao cho $R_i \subseteq U,\ i = 1, ..., m$, $\bigcup\limits_{1 \le i \le m} R_i = U$. Một CSDL quan hệ $\mathbf{r}$ trên $\mathbf{R}$ là tập các quan hệ $\{r_1, r_2, ..., r_m\}$ trong đó mỗi r_i là một thể hiện của lược đồ quan hệ R_i.

Định nghĩa 2. Một siêu đồ thị là một cặp $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ trong đó $\mathcal{N}$ là một tập hữu hạn các đỉnh và $\mathcal{E} \subseteq 2^{\mathcal{N}} - \{\emptyset\}$ là một tập các siêu cạnh mà mỗi siêu cạnh (để đơn giản ta cũng gọi là cạnh) là một tập con khác rỗng của $\mathcal{N}$.

Như vậy, một lược đồ CSDL $\mathbf{R}= \{R_1, R_2, ..., R_m\}$ trên tập thuộc tính U hiển nhiên được xem như một siêu đồ thị với tập các đỉnh là tập thuộc tính U, còn các lược đồ quan hệ con $R_1, R_2, ..., R_m$ là các siêu cạnh của siêu đồ thị này. Ta gọi $\mathcal{H} = (U, \mathbf{R})$ là siêu đồ thị liên kết với lược đồ CSDL $\mathbf{R}$ $= \{R_1, R_2, ..., R_m\}$.

Định nghĩa 3. Cho $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ là một siêu đồ thị, các đỉnh $s, t \in \mathcal{N}$. Một *đường đi* từ đỉnh s tới đỉnh t là một dãy gồm k cạnh $(k \ge 1)$ $E_1, E_2, ..., E_k$ sao cho:

(a) $s \in E_1$,

(b) $t \in E_k$,

(c) $E_i \cap E_{i+1} \ne \emptyset$ nếu $1 \le i < k$.

Ta cũng nói dãy trên là một đường đi từ E_1 tới E_k.

Hai đỉnh (hay hai cạnh) được gọi là *liên thông* nếu như có một đường đi từ đỉnh (cạnh) này tới đỉnh (cạnh) kia.

Một tập đỉnh (cạnh) là liên thông nếu mọi cặp đỉnh (cặp cạnh) đều liên thông. Một *thành phần liên thông* là một tập liên thông cực đại các cạnh.

Không làm mất tính tổng quát, trong bài báo này, ta luôn giả thiết siêu đồ thị $\mathcal{H} = (U, \mathbf{R})$ liên kết với lược đồ CSDL $\mathbf{R} = \{R_1, R_2, ..., R_m\}$ là liên thông.

Định nghĩa 4. [2] Siêu đồ thị $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ được gọi là *Berge-phi chu trình* nếu nó không chứa một *Berge-chu trình*, là một dãy có dạng:

$$(E_1, x_1, E_2, x_2, ..., E_m, x_m, E_{m+1}),$$

trong đó:

(a) $x_1, x_2, ..., x_m$ là các đỉnh phân biệt của $\mathcal{H}$,
(b) $E_1, E_2, ..., E_m$ là các cạnh phân biệt của $\mathcal{H}$ và $E_{m+1} = E_1$,
(c) $m \geq 2$, có nghĩa là có ít nhất hai cạnh tham gia,
(d) $x_i \in E_i$ và E_{i+1} $(1 \leq i \leq m)$.

Nhận xét 1. Theo định nghĩa 4, dễ thấy rằng $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ là Berge-chu trình nếu tồn tại 2 cạnh $E_i, E_j \in \mathcal{E}$ với $|E_i \cap E_j| \geq 2$.

Ví dụ 1. Xét siêu đồ thị ở hình 1.

Hình 1

Siêu đồ thị này có hai cạnh $E_1 = AEF$ và $E_2 = AEC$ làm thành một *Berge-chu trình* (E_1, E, E_2, A, E_1) như ở hình 2.

E_1 E_2

Hình 2.

Định nghĩa 5. [4] Một siêu đồ thị $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ được gọi là $\gamma-$phi chu trình nếu nó không chứa một $\gamma-$chu trình là một dãy có dạng:

$$(E_1, x_1, E_2, x_2, ..., E_m, x_m, E_{m+1}),$$

trong đó:
(a) $x_1, x_2, ..., x_m$ là các đỉnh phân biệt của $\mathcal{H}$,
(b) $E_1, E_2, ..., E_m$ là các cạnh phân biệt của $\mathcal{H}$ và $E_{m+1} = E_1$,
(c) $m \geq 3$, có nghĩa là có ít nhất ba cạnh tham gia,
(d) $x_i \in E_i$ và E_{i+1} $(1 \leq i \leq m)$, và
(e) nếu $1 \leq i \leq m$ thì x_i không thuộc E_j nào trừ E_i và E_{i+1}.

Cho $\mathbf{R} = \{R_1, R_2, ..., R_m\}$ là một lược đồ CSDL. Thuật toán Graham [5] áp dụng một số hữu hạn lần hai quy tắc (thao tác) sau lên $\mathbf{R} = \{R_1, R_2, ..., R_m\}$ cho tới khi không thể áp dụng được quy tắc nào thêm nữa:
(a) Nếu A là một thuộc tính chỉ xuất hiện đúng trong một R_i, thì loại A khỏi R_i,
(b) Loại R_i khỏi lược đồ CSDL $\mathbf{R}$, nếu có một $R_j, i \neq j$, sao cho $R_i \subseteq R_j$.

Ta nói thuật toán Graham là *thành công* nếu với đầu vào $\mathbf{R}$ nó kết thúc với một tập rỗng; ngược lại thuật toán là *thất bại.*

Chú ý rằng, đương nhiên thuật toán Graham có thể được thực hiện trên một siêu đồ thị. Khi đó thao tác (a) là loại bỏ một đỉnh chỉ xuất hiện đúng trong một cạnh, còn thao tác (b) là loại bỏ một cạnh bị chứa trong một cạnh khác.

Định nghĩa 6. Cho $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ là một siêu đồ thị. Khi đó:
(a) $\mathcal{H}$ được gọi là *α-phi chu trình* (thường chỉ gọi là *phi chu trình*) nếu thuật toán Graham là thành công với đầu vào là $\mathcal{H}$.
(b) $\mathcal{H}$ được gọi là *β-phi chu trình* nếu mọi siêu đồ thị con của $\mathcal{H}$ (là một tập con của $\mathcal{E}$) đều là

α-phi chu trình.

Trong [4], Fagin đã đưa ra năm định nghĩa tương đương của tính β-phi chu trình, bốn định nghĩa tương đương của tính γ-phi chu trình, và đã chứng minh định lý sau về mối quan hệ giữa các loại siêu đồ thị phi chu trình.

Định lý 1. [4] *Đối với các lớp siêu đồ thị phi chu trình, tính phi chu trình được kéo theo bởi trật tự sau: Berge-phi chu trình $\Rightarrow$ γ-phi chu trình $\Rightarrow$ β-phi chu trình $\Rightarrow$ α-phi chu trình. Không có một phép kéo theo ngược nào được thỏa mãn.*

Từ định lý trên, dễ dàng chứng minh hệ quả sau về sự liên hệ giữa các kiểu siêu đồ thị chu trình.

Hệ quả 1 *Nếu siêu đồ thị $\mathcal{H}$ là chu trình, ta có trật tự các phép kéo theo như sau: α-chu trình $\Rightarrow$ β-chu trình $\Rightarrow$ γ-chu trình $\Rightarrow$ Berge-chu trình.*

Ta nói một lược đồ CSDL $\mathbf{R}$ là *θ-phi chu trình* nếu siêu đồ thị liên kết với nó $\mathcal{H} = (U, \mathbf{R})$ là *θ-phi chu trình* (với $\theta =$ Berge, α, β hay γ).

2. CÁC THUẬT TOÁN ĐOÁN NHẬN CÁC LỚP SIÊU ĐỒ THỊ PHI CHU TRÌNH

2.1. Thuật toán đoán nhận tính Berge-phi chu trình

Trong trường hợp tổng quát, có thể mở rộng thuật toán tìm kiếm *"trước tiên theo chiều rộng" (Breadth-first)* hoặc *"trước tiên theo chiều sâu" (Depth-first)* để xác định xem một đồ thị vô hướng có chu trình hay không cho trường hợp của siêu đồ thị. Theo [8] thuật toán rõ ràng là có thời gian đa thức.

Trong bài báo này, chúng tôi đưa ra một thuật toán mới, dựa trên nhận xét 1 rằng mọi siêu đồ thị $\mathcal{H} = (\mathcal{N}, \mathcal{E})$ có $|E_i \cap E_j| \geq 2,\ E_i, E_j \in \mathcal{E}, i \neq j$, đều là Berge-chu trình. Nếu $|E_i \cap E_j| < 2, i \neq j$, khi đó ta có thể áp dụng các đặc trưng định lượng cho các siêu đồ thị ω-phi chu trình [7]. Từ đó ta xây dựng được thuật toán đa thức sau:

Thuật toán 1.

Input: $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, với $|\mathcal{N}| = n, |\mathcal{E}| = m$.

Output: Quyết định $\mathcal{H}$ là Berge-phi chu trình?

Procedure: Var i, j : integer

1. for $i := 1$ to $m - 1$ do
2.   for $j := i + 1$ to m do
3.   if $|E_i \cap E_j| \geq 2$ then go to 7;
4.   end;
5. end;
6. if $\sum\limits_{1 \leq i \leq m} (|R_i| - 1) = n - 1$ then goto 8;
7. print "$\mathcal{H}$ is Berge-cyclic"; exit;
8. print "$\mathcal{H}$ is Berge-acyclic";
9. End.

2.2. Thuật toán đoán nhận tính α-phi chu trình

Định nghĩa 6 đã dùng thuật toán Graham để xác định tính α-phi chu trình của một siêu đồ thị. Thực ra, thuật toán Graham được mô tả ở trên mới chỉ là các quy tắc khử đỉnh và khử cạnh, thể hiện các quy tắc này bằng ngôn ngữ thuật toán không phải là đơn giản. Chúng tôi đã "lượng hoá" hai quy tắc của thuật toán Graham và trình bày thuật toán dưới dạng một thủ tục. Trong [6] đã cài đặt thuật toán này, và đã chứng minh rằng thuật toán Graham là Church-Rosser, theo nghĩa siêu đồ thị kết quả là duy nhất, không phụ thuộc vào dãy các thao tác được thực hiện. Rõ ràng thuật toán là đa thức theo $m + n$.

Thuật toán 2. Input: $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, với $|\mathcal{N}| = n, |\mathcal{E}| = m$.

```
Output: Quyết định H là α-phi chu trình?
Procedure:
Var COUNT_ALL, COUNT_OLD, COUNT_EDGE : integer;
COUNT_ALL := Σ_{1≤i≤m} |E_i| ;
Repeat
    1. COUNT_OLD := COUNT_ALL;
    2. for i := 1 to m do    /* loại các cạnh bị chứa trong cạnh khác */
          for j := 1 to m do
             if (i ≠ j) and E_i ≠ ∅ and (E_i ⊆ E_j) then
                begin
                   COUNT_ALL : = COUNT_ALL - |E_i|;
                   E_i := ∅;
                end;
          end;
    end;
    3. for i := 1 to n do    /* đếm số cạnh chứa đỉnh x_i để loại đỉnh cô lập */
          COUNT_EDGE := 0;
          for j := 1 to m do
             if x_i ∈ E_j then COUNT_EDGE := COUNT_EDGE + 1;
          end;
          if COUNT_EDGE = 1 then
                begin
                   COUNT_ALL := COUNT_ALL - 1;
                   for j := 1 to m do
                   E_j := E_j − {x_i};
                   end;
                end;
    end;
Until COUNT_ALL = COUNT_OLD;
If COUNT_ALL = 0    /* khi đó siêu đồ thị kết quả là rỗng */
    Then print "H is α-acyclic";
    Else print "H is α-cyclic";
END.
```

2.3. Thuật toán đoán nhận tính β-phi chu trình

Trước hết, ta trình bày một vài khái niệm liên quan đến tính β-chu trình của một siêu đồ thị.

Cho $(S_1, ..., S_m, S_{m+1})$ là một dãy các tập, trong đó các tập $S_1, ..., S_m$ là phân biệt, và $S_{m+1} = S_1$. Ta gọi S_i và S_{i+1} là *láng giềng* $(1 \leq i \leq m)$. Dãy $(S_1, ..., S_m, S_{m+1})$ nói trên được gọi là một *chu trình thuần túy* nếu $m \geq 3$ và với $i \neq j$ thì $S_i \cap S_j \neq \emptyset$ nếu và chỉ nếu S_i và S_j là láng giềng của nhau.

Một *β-chu trình* trong siêu đồ thị $\mathcal{H}$ là một dãy cạnh $(S_1, ..., S_m, S_{m+1})$ sao cho nếu $X = S_1 \cap ... \cap S_m$, và $S'_i = S_i - X$ $(1 \leq i \leq m)$ thì $(S'_1, ..., S'_m, S'_{m+1})$ là một chu trình thuần tuý.

Nói cách khác, β-chu trình là một dãy cạnh có dạng $(S'_1 \cap X, ..., S'_m \cap X, S'_{m+1} \cap X)$ trong đó $(S'_1, ..., S'_m, S'_{m+1})$ là một chu trình thuầnúy. Hiển nhiên là một chu trình thuần túy cũng là β-chu trình, ứng với trường hợp $X = \emptyset$.

Trong [4] đã chứng minh sự tương đương của hai định nghĩa sau:

(i) Một siêu đồ thị là β-chu trình nếu nó có một β-chu trình.

(ii) Một siêu đồ thị là β-chu trình nếu nó có một siêu đồ thị con là α-chu trình.

Nếu $(S_1, ..., S_m, S_{m+1})$ là một β-chu trình, ta nói (S_1, S_2, S_3) bắt đầu một β-chu trình. Gọi $\mathcal{S} = (S_1, S_2, S_3)$ là một bộ ba cạnh phân biệt của $\mathcal{H}$. Đặt $X = S_1 \cap S_2 \cap S_3$ và $S_i' = S_i - X$, với $i = 1, 2, 3$.

Dễ thấy rằng, theo định nghĩa β-chu trình, nếu $S_1' \cap S_2' = \emptyset$ hoặc $S_1' \cap S_3' = \emptyset$ thì $\mathcal{S}$ không thể bắt đầu một β-chu trình nào của $\mathcal{H}$. Vậy với mục đích xác định $\mathcal{S} = (S_1, S_2, S_3)$ có bắt đầu một β-chu trình không, có thể giả thiết có đồng thời $S_1' \cap S_2' \neq \emptyset$ và $S_2' \cap S_3' \neq \emptyset$.

Gọi $T = \{E \in \mathcal{E} : (E = S_1)$ hoặc $(E = S_3)$ hoặc $(X \subset E$ và $E \cap S_2' = \emptyset)\}$.

Dễ thấy là $S_2 \notin T$.

Đặt $T' = \{E - X : E \in T\}$, rõ ràng là $S_1' \in T'$ và $S_3' \in T'$. Trong [4] đã chứng minh rằng S_1' và S_3' thuộc cùng một thành phần liên thông của siêu đồ thị T' nếu và chỉ nếu $\mathcal{S} = (S_1, S_2, S_3)$ bắt đầu một β-chu trình của siêu đồ thị $\mathcal{H}$.

Từ đó ta có thuật toán thời gian đa thức sau đây để đoán nhận tính β-phi chu trình của siêu đồ thị $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, thuật toán này là sự mở rộng tự nhiên của thuật toán xác định các thành phần liên thông của một đồ thị vô hướng thông thường [8].

Thuật toán 3.

Input: $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, với $|\mathcal{N}| = n, |\mathcal{E}| = m$.
Output: Quyết định $\mathcal{H}$ là β-phi chu trình?

1. for each $\mathcal{S} = (S_1, S_2, S_3)$ of $\mathcal{H}$ do
   $X := S_1 \cap S_2 \cap S_3$;
   $S_1' := S_1 - X$;
   $S_2' := S_2 - X$;
   $S_3' := S_3 - X$;
   If $S_1' \cap S_2' = \emptyset$ or $S_2' \cap S_3 = \emptyset$ then go to 1;
2. $T := \{E \in \mathcal{E} : (E = S_1)$ or $(E = S_3)$ or $(X \subset E$ and $E \cap S_2' = \emptyset)\}$;
   $T' := \{E - X : E \in T\}$;
3. Kiểm tra xem S_1' và S_3' có thuộc cùng một thành phần liên thông trong T' hay không ? Nếu đúng thì kết luận siêu đồ thị $\mathcal{H}$ là β-chu trình, thuật toán kết thúc. Còn nếu sai thì quay lại bước 1 nếu vẫn còn một bộ ba mới chưa được xét. Ngược lại, nếu mọi bộ ba đã được xét mà không có bộ ba nào có thể là bắt đầu một β-chu trình, thì siêu đồ thị $\mathcal{H}$ là phi chu trình. Thuật toán kết thúc.

2.4. Thuật toán đoán nhận tính γ-phi chu trình

Trong phần này, chúng ta giới thiệu một thuật toán đa thức của D'Atri và Moscarini [3] để đoán nhận tính γ-phi chu trình. Thuật toán bao gồm việc áp dụng một số hữu hạn lần các quy tắc sau đây, theo một thứ tự bất kỳ cho đến khi không thể áp dụng được thêm một quy tắc nào nữa.

Thuật toán 4.

Input: $\mathcal{H} = (\mathcal{N}, \mathcal{E})$, với $|\mathcal{N}| = n, |\mathcal{E}| = m$.
Output: Quyết định $\mathcal{H}$ là γ-phi chu trình?

1. Nếu một đỉnh là *cô lập* (isolate - chỉ thuộc đúng một cạnh) thì loại bỏ đỉnh đó;
2. Nếu một cạnh chỉ chứa đúng một đỉnh (*singleton*) thì loại bỏ cạnh đó (nhưng không loại bỏ đỉnh đó khỏi các cạnh khác chứa nó);
3. Nếu một cạnh là *rỗng* (empty - không chứa đỉnh nào) thì loại bỏ nó;
4. Nếu hai cạnh chứa cùng các đỉnh như nhau thì loại bỏ một trong chúng;
5. Nếu hai đỉnh là *tương đương cạnh* (edge equivalent - tức chúng thuộc cùng một số cạnh như nhau) thì loại bỏ một trong hai đỉnh đó khỏi mọi cạnh có chứa nó.
6. Rõ ràng thuật toán luôn kết thúc, nếu nó kết thúc với siêu đồ thị kết quả là một tập rỗng thì $\mathcal{H}$ là γ-phi chu trình, trái lại thì $\mathcal{H}$ là γ-chu trình.

Tính đúng đắn của thuật toán này đã được Fagin chứng minh trong [4]. Trong một bài báo khác, chúng tôi sẽ cài đặt và chứng minh thuật toán là Church–Rosser, tức là siêu đồ thị kết quả sau khi áp dụng thuật toán là duy nhất, không phụ thuộc vào dãy các thao tác được thực hiện.

Ví dụ 2. Áp dụng thuật toán 4 cho siêu đồ thị cho ở hình 3.
Các cạnh của siêu đồ thị là:

A B C
B C D E
C
D E F

Hình 3

Vì đỉnh A, F là cô lập và cạnh {C} chỉ chứa đúng một đỉnh, nên chúng bị loại theo quy tắc (1) và (2), siêu đồ thị còn lại là:

B C
B C D E
D E

Ta thấy D và E là tương đương cạnh, B và C cũng vậy, theo quy tắc (5), ta loại E và C khỏi cả hai cạnh chứa nó. Siêu đồ thị còn lại là:

B
B D
D

Bây giờ, cạnh thứ nhất và thứ ba chỉ chứa một đỉnh, loại bỏ các cạnh đó, còn lại:

B D

Cuối cùng, cả hai đỉnh B và D đều là chỉ có trong đúng một cạnh, loại chúng theo quy tắc (1), còn lại một cạnh không chứa đỉnh nào, ta loại cạnh rỗng này theo quy tắc (3), kết quả cuối cùng là tập rỗng các cạnh. Vậy R là γ-phi chu trình.

Ví dụ 3. Xét lược đồ CSDL $\mathbf{R} = \{AB\ , BC\ , ABC\ \}$ có siêu đồ thị liên kết cho trong hình 4. Dễ thấy lược đồ **R** là Berge-chu trình (theo nhận xét 1) nhưng **R** là α-phi chu trình (dễ kiểm chứng rằng thuật toán Graham là thành công trên **R**). **R** cũng là β-phi chu trình, vì mọi siêu đồ thị con của **R** đều là α-phi chu trình. Nhưng có thể thấy **R** là γ-chu trình vì không có quy tắc nào của thuật toán 4 áp dụng được cho **R**.

Hình 4

TÀI LIỆU THAM KHẢO

[1] Beeri C. et al., On the desirability of acyclic database schemes, *J. ACM.* **30** (1983) 479–513.

[2] Berge C., *Graphs and Hypergraphs*, North-Holland, NewYork, 1973.

[3] D'Atri A. and Moscarini M., *Acyclic hypergraphs; Their recognition and top-down versus bottom-up generation*. Yech. Rep. R.29, consiglio Nazionale delle Richerche, Instituto di Analist dei Sistemt ed Informatica, 1982.

[4] Fagin, R., Degrees of acyclicity for hypergraphs and relational database schemes, *J. ACM* **30** (1983) 514–550.

[5] Graham M.H., *On the Universal Relation*, Tech Rep, Univ. of Toronto, Toronto, Ont., Can., Sept. 1979.

[6] Hồ Thuần và Nguyễn Văn Định, Biểu diễn biểu thức kết nối của nhiều quan hệ dưới dạng siêu đồ thị và việc xác định trình rút gọn đầy đủ, *Báo cáo khoa học tại Hội thảo quốc gia về Công nghệ thông tin*, Hải Phòng, 6-2001.

[7] Nguyen Van Dinh, On the desirability of ω-acyclic database schemes, *Journal of Computer Science and Cybernetics* **17** (3) (2001) 53–59.

8] Seggewick R., *Algorithms*, Addison-Wesley Publishing Company, 2nd edition, 1988.

[9] Tarjan R. and Yannakakis M., Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. Tech. Rep., Bell Labs, Murray Hill, N.J., Mar. 1982.

Nhận bài ngày 15 - 2 - 2002

Trường Đại học Nông nghiệp I, Hà Nội.