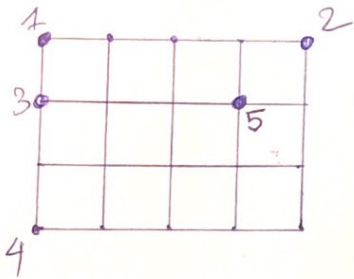


### Bài 3. Phân cụm



Khoảng cách city-block:

	1	2	3	4	5
1	0	4	1	3	4
2	4	0	5	<del>7</del>	<del>2</del>
3			0	2	3
4				0	5
5					0

⇒ kết hợp nút 1 và 3 thành nút 6.

$$d(6, 2) = \max\{d(1, 2), d(3, 2)\} = 5$$

$$d(6, 4) = \max\{d(1, 4), d(3, 4)\} = 3$$

$$d(6, 5) = \max\{d(1, 5), d(3, 5)\} = 4$$

	2	4	5	6
2	0	<del>7</del>	2	5
4		0	5	3
5			0	4
6				0

⇒ kết hợp nút 2 và 5 thành nút 7

$$d(4, 7) = \max\{d(5, 4), d(2, 4)\} = 7$$

$$d(6, 7) = \max\left\{d(1, 2), d(3, 2), d(1, 5), d(3, 5)\right\}$$

	4	6	7
5	0	<del>5</del>	<del>7</del>
6		0	5
7			0

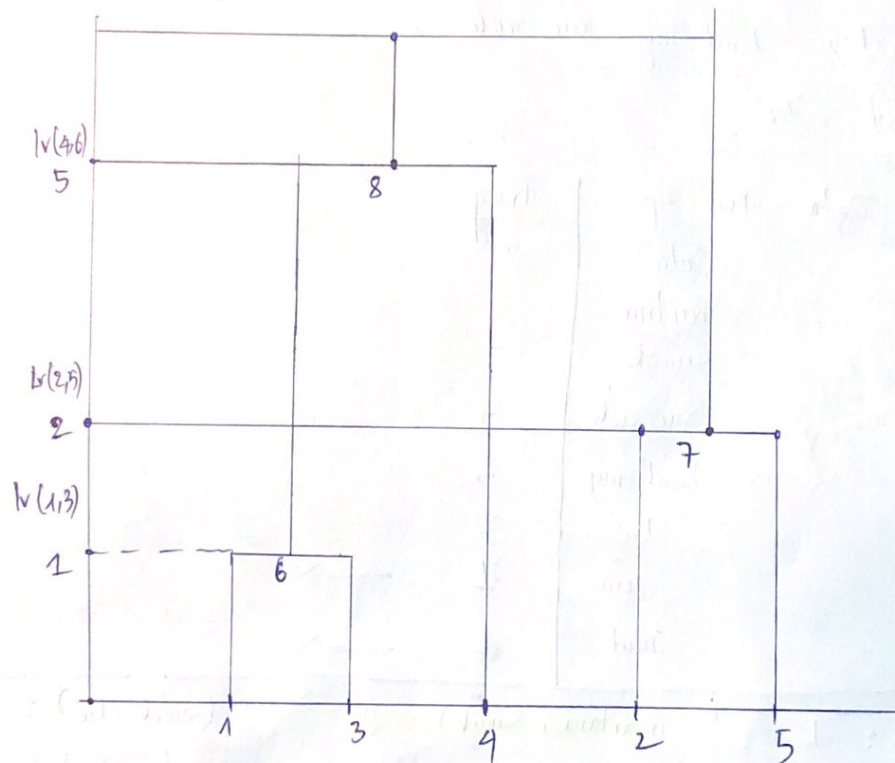
⇒ kết hợp {4, 6} thành nút 8.

$$d(7, 8) = \max \left\{ d(2, 1), d(2, 3), d(2, 4), d(5, 1), d(5, 3), d(5, 4) \right\}$$

	7	8
7	0	7
8		0

Kết thúc phân cụm.

Ta thu được cây phân cấp kết nhập.



## Bài 2. Giao dịch.

1. Cola, martini, snack, xúc xích
2. martini, snack
3. bánh mì, bơ, cola
4. xúc xích, martini
5. bánh mì, sữa, bơ
6. snack, bánh mì, xúc xích
7. mít, bơ

Ta có:	<del>Cola</del>	Tên SP	freq:
		Cola	2
		martini	3
		snack	3
		xúc xích	3
		bánh mì	3
		bơ	3
		sữa	1 → X
		mít	1 → X

- (Cola, martini) : 1
- (Cola, snack) : 1
- (Cola, xúc xích) : **1**
- (Cola, bánh mì) : 1
- (Cola, bơ) : 1

- (martini, snack) : **2**
- (martini, xúc xích) : **2**
- (martini, bánh mì) : 0
- (martini, bơ) : 0
- (snack, xúc xích) : **2**
- (snack, bánh mì) : 1

- (snack, bơ) : 0
- (xúc xích, bơ) : 0
- (xúc xích, bánh mì) : 1
- (xúc xích, bơ) : 0
- (bánh mì, bơ) : **2**



### Tập mục 2

(martini, snack, xúc xích) : 1.

Vậy tập các tập mục thường xuyên là:

• Tập mục 1: Cola, martini, snack, xúc xích, bánh mì, bơ

• Tập mục 2: (Cola, xúc xích), (martini, snack), (martini, xúc xích)  
(snack, xúc xích), (bánh mì, bơ)

④ Vì tập mục thường xuyên không có tập mục có 3 phần tử nên các luật kết hợp với ngưỡng độ hỗ trợ  $s = 2$  và độ tin cậy  $c = 0.65$  có dạng:  $\{a\} \rightarrow \{b\}$  trong đó  $a, b$  là các item đơn.

để ý ~~chỉ có~~:  $\frac{1}{2} < c < \frac{2}{3}$  nên các luật kết hợp là:

1) martini  $\rightarrow$  snack

2) snack  $\rightarrow$  martini

3) martini  $\rightarrow$  xúc xích

4) xúc xích  $\rightarrow$  martini

5) snack  $\rightarrow$  xúc xích

6) xúc xích  $\rightarrow$  snack

8) bánh mì  $\rightarrow$  bơ

9) bơ  $\rightarrow$  bánh mì.

Bài tập phân cụm bằng k-means,  $k=2$  và khoảng cách City-block.

$$w_1 (1, 0, 1, 0)$$

$$w_2 (2, 0, 1, 0)$$

$$w_3 (1, 1, 1, 0)$$

$$w_4 (1, 0, 0, 0)$$

$$w_5 (4, 1, 2, 0)$$

B1: Chọn 2 tâm  $C_0 = w_1$ ,  $C_1 = w_3$ .

• Ta có:  $D(w_2, w_1) = 1$ ,  $D(w_2, w_3) = 2$   
 $\rightarrow w_2$  vào nhóm tâm  $C_0$ .

•  $D(w_4, w_1) = 1$ ,  $D(w_4, w_3) = 2$   
 $\rightarrow w_4 \rightarrow C_0$ .

•  $D(w_5, w_1) = 5$ ,  $D(w_5, w_3) = 4$

$w_5 \rightarrow C_1$ .

Nhóm 1:  $w_1, w_2, w_4 \rightarrow$  tâm nhóm:  $C_0 = \frac{w_1 + w_2 + w_4}{3} = (1, 0, \frac{2}{3}, 0)$

Nhóm 2:  $w_3, w_5 \rightarrow$  tâm:  $C_1 = \frac{w_3 + w_5}{2} = (\frac{5}{2}, 1, \frac{3}{2}, 0)$ .

B2: Ta có:

•  $D(w_1, C_0) = \frac{1}{3}$ ,  $D(w_1, C_1) = 3$

$\Rightarrow w_1$  vào nhóm  $C_0$ .

•  $D(w_2, C_0) = \frac{5}{3}$ ,  $D(w_2, C_1) = 2$

$\Rightarrow w_2$  vào nhóm  $C_0$ .

•  $D(w_3, C_0) = \frac{4}{3}$ ,  $D(w_3, C_1) = 2$

$\Rightarrow w_3$  vào nhóm  $C_0$ .

•  $D(w_4, C_0) = \frac{2}{3}$ ,  $D(w_4, C_1) = 4$

$\Rightarrow w_4$  vào nhóm  $C_0$ .

•  $D(w_5, C_0) = \frac{16}{3}$ ,  $D(w_5, C_1) = 2$

$\Rightarrow w_5$  vào nhóm  $C_1$ .



Cấp nhất:

Nhóm  $C_0$ :  $w_1, w_2, w_3, w_4 \rightarrow$  tần  $C_0 = (\frac{5}{4}, \frac{1}{4}, \frac{3}{4}, 0)$

Nhóm  $C_1$ :  $w_5 \rightarrow$  tần  $C_1 = (4, 1, 2, 0)$

Bước 3)

•  $D(w_1, C_0) = \frac{3}{4}$ ,  $D(w_1, C_1) = 5 \Rightarrow w_1$  vào  $C_0$ .

•  $D(w_2, C_0) = \frac{5}{4}$ ,  $D(w_2, C_1) = 4 \Rightarrow w_2$  vào  $C_0$ .

•  $D(w_3, C_0) = \frac{5}{4}$ ,  $D(w_3, C_1) = 4 \Rightarrow w_3$  vào  $C_0$ .

•  $D(w_4, C_0) = \frac{5}{4}$ ,  $D(w_4, C_1) = 6 \Rightarrow w_4$  vào  $C_0$ .

Vậy  $w_1, w_2, w_3, w_4$  nhóm  $C_0$ .

$w_5$  nhóm  $C_1$ .

$\Rightarrow$  Thuật toán dừng với hai cụm là:

$\{w_1, w_2, w_3, w_4\}$  và  $\{w_5\}$ .

Phân cụm phân cấp

	$w_1$	$w_2$	$w_3$	$w_4$	$w_5$
$w_1$	0	1	1	1	5
$w_2$		0	2	2	4
$w_3$			0	2	4
$w_4$				0	6
$w_5$					0

$\Rightarrow$  kết hợp:  $\{w_1, w_2\}$  thành  $w_6$ .

② ~~thứ~~

$$d(w_6, w_3) = \min(d(w_1, w_3), d(w_2, w_3)) = 1$$

	$w_3$	$w_4$	$w_5$	$w_6$
$w_3$	0	2	4	1
$w_4$		0	6	1
$w_5$			0	4
$w_6$				0

$$d(w_6, w_4) = \min(d(w_1, w_4), d(w_2, w_4)) = 1$$

$$d(w_6, w_5) = \min(d(w_1, w_5), d(w_2, w_5)) = 4$$

⇒ kết hợp  $w_3$  và  $w_6$  thành  $w_7$ .

$$d(w_7, w_4) = \min\{d(w_3, w_4), d(w_6, w_4)\} = 1$$

	$w_4$	$w_5$	$w_7$
$w_4$	0	6	1
$w_5$		0	4
$w_7$			0

$$d(w_7, w_5) = \min\{d(w_3, w_5), d(w_6, w_5)\} = 4$$

• kết hợp  $w_4$  và  $w_7$  thành  $w_8$ .

$$d(w_8, w_5) = \min\{d(w_4, w_5), d(w_7, w_5)\} = 4$$

	$w_5$	$w_8$
$w_5$	0	4
$w_8$		0

kết thúc phân cấp.

Cây phân cấp

