

Distributed Algorithm for

Computing Formal Concepts Using Map-Reduce Framework

by Petr Krajca and Vilem Vychodil

Presentation by Johannes Gosda and Stefan Wehrmeyer

Advanced M/R Seminar, HPI

17.11.2009

Challenge

Finding groups of similar articles.

A T T R I B U T E S

A R T I C L E S

	Drehkreuz	Studenten	Abschluss	Gründung	Lehrstühle
Lufthansa	X			X	
IFI		X	X		X
GermanWings	X			X	
HPI		X	X	X	X

Solution: Maximal Rectangles

Maximal rectangles define patterns in the matrix.

These maximal rectangles are equivalent to Formal Concepts.

But how can you find maximal rectangles?

	Dre		0	3	2	1	4	Lehrstühle
Lufthansa		L	X	X				
IFI		G	X	X				X
GermanWings		I			X	X	X	
HPI		H		X	X	X	X	X

Operators \downarrow and \uparrow

$B\downarrow$ = Set of all **objects** sharing all attributes from B.

$A\uparrow$ = Set of all **attributes** shared by all objects from A.

Examples:

$\{\}\downarrow = \{L, I, G, H\}$

$\{1\}\downarrow = \{I, H\}$

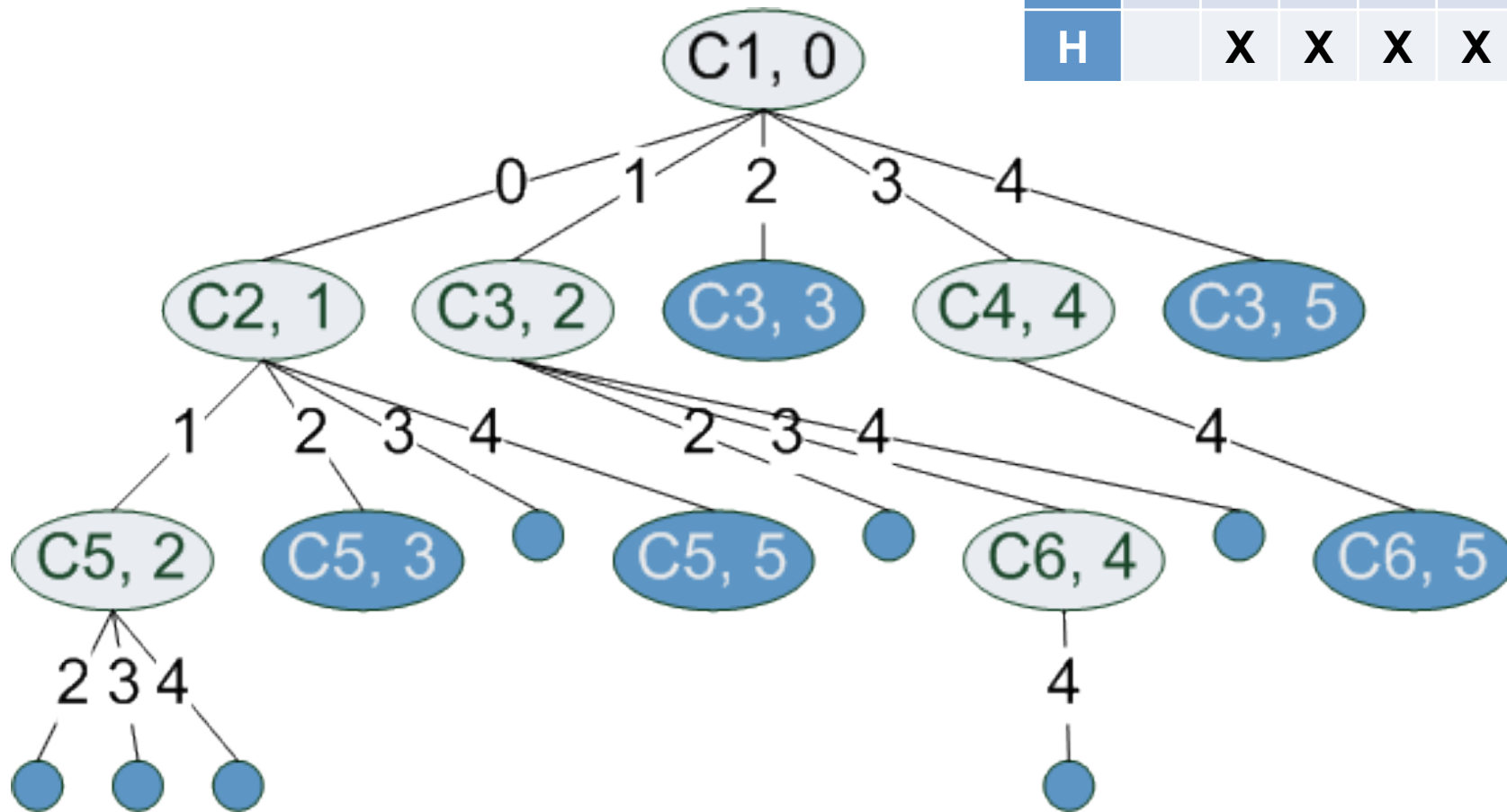
$\{I, H\}\uparrow = \{1, 2, 4\}$

$(\{I, H\}, \{1, 2, 4\})$

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X

C1 = Initial Formal Concept = ($\{\downarrow\}$, $\{\downarrow\uparrow\}$)

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X



Step by Step

Calculate initial Formal Concept: $C1 = (\{\downarrow, \{\downarrow\uparrow\})$

Map $(\{L, I, G, H\}, \{\})$, 0

$(\{L, I, G, H\} \cap \{0\}\downarrow, (\{\downarrow \cap \{0\}\downarrow)\uparrow) = C2$

$(\{L, I, G, H\} \cap \{1\}\downarrow, (\{\downarrow \cap \{1\}\downarrow)\uparrow) = C3$

$(\{L, I, G, H\} \cap \{2\}\downarrow, (\{\downarrow \cap \{2\}\downarrow)\uparrow) = C3$

$(\{L, I, G, H\} \cap \{3\}\downarrow, (\{\downarrow \cap \{3\}\downarrow)\uparrow) = C4$

$(\{L, I, G, H\} \cap \{4\}\downarrow, (\{\downarrow \cap \{4\}\downarrow)\uparrow) = C3$

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X

All Formal Concepts:

$C1 = (\{L, I, G, H\}, \{\}) = (\{\downarrow, \{\downarrow\uparrow\})$

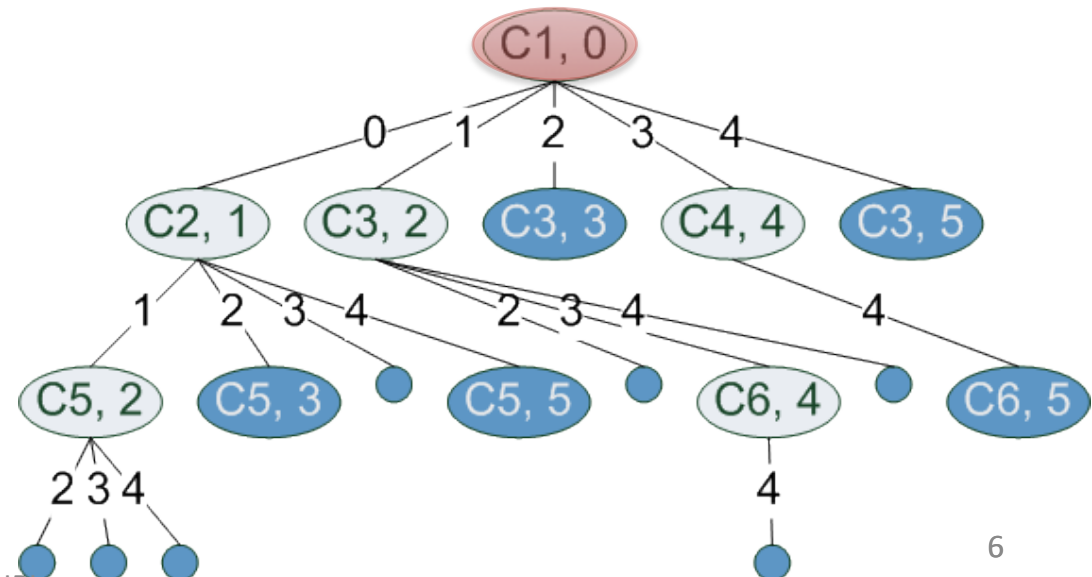
$C2 = (\{L, G\}, \{0, 3\})$

$C3 = (\{I, H\}, \{1, 2, 4\})$

$C4 = (\{L, G, H\}, \{3\})$

$C5 = (\{\}, \{0, 1, 2, 3, 4\})$

$C6 = (\{H\}, \{1, 2, 3, 4\})$



Step by Step

Calculate initial Formal Concept: $C1 = (\{\downarrow, \{\downarrow\uparrow\})$

Reduce $(C2,1), (C3,2), (C3,3), (C4,4), (C3,5)$

$$\{\downarrow\} \cap Y0 = \{0,3\} \cap Y0$$

$$\{\downarrow\} \cap Y1 = \{1,2,4\} \cap Y1$$

$$\{\downarrow\} \cap Y2 \neq \{1,2,4\} \cap Y2$$

$$\{\downarrow\} \cap Y3 = \{3\} \cap Y3$$

$$\{\downarrow\} \cap Y4 \neq \{1,2,4\} \cap Y4$$

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X

$Y0 = \{\}, Y1 = \{0\}, Y2 = \{0,1\}, \text{etc.}$

All Formal Concepts:

$$C1 = (\{L,I,G,H\}, \{\downarrow\}) = (\{\downarrow, \{\downarrow\uparrow\})$$

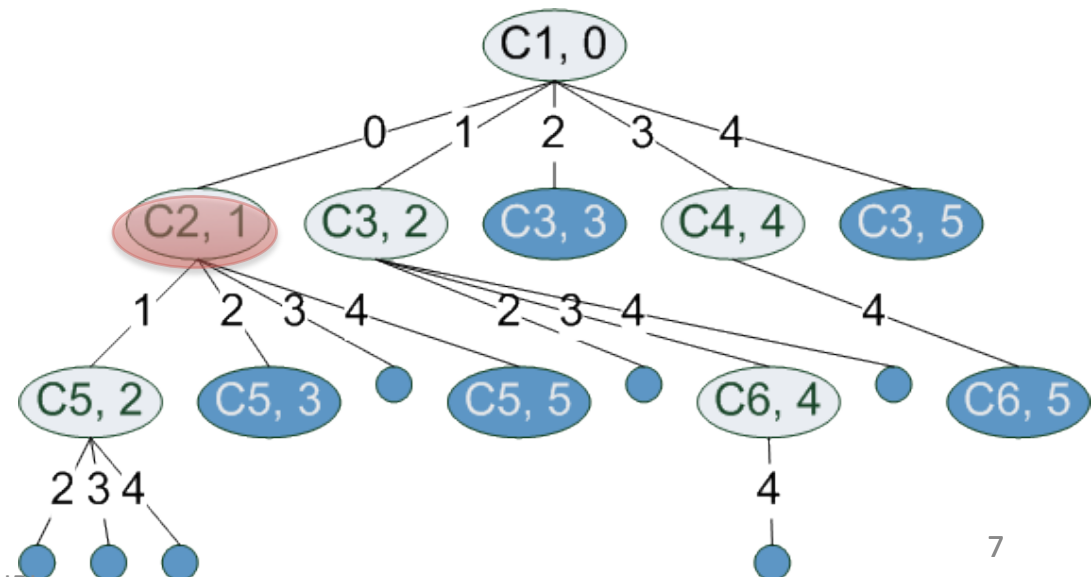
$$C2 = (\{L,G\}, \{0,3\})$$

$$C3 = (\{I,H\}, \{1,2,4\})$$

$$C4 = (\{L,G,H\}, \{3\})$$

$$C5 = (\{\}, \{0,1,2,3,4\})$$

$$C6 = (\{H\}, \{1,2,3,4\})$$



Step by Step

input Formal Concept: (C2,1), (C3,2), (C4,4)

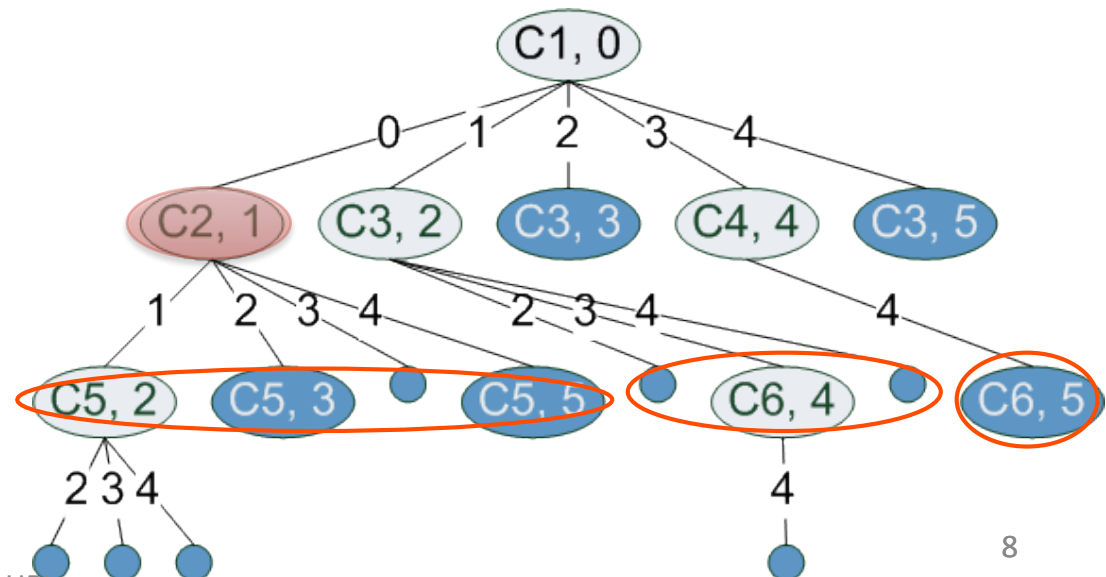
Map iteration 2

- $(\{L,G\} \cap \{1\} \downarrow, (\{L,G\} \cap \{1\} \downarrow) \uparrow) = C5$
- $(\{L,G\} \cap \{2\} \downarrow, (\{L,G\} \cap \{2\} \downarrow) \uparrow) = C5$
- $(\{L,G\} \cap \{4\} \downarrow, (\{L,G\} \cap \{4\} \downarrow) \uparrow) = C5$
- $(\{I,H\} \cap 3 \downarrow, (\{I,H\} \cap 3 \downarrow) \uparrow) = C6$
- $(\{L,G,H\} \cap 4 \downarrow, (\{L,G,H\} \cap 4 \downarrow) \uparrow) = C6$

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X

All Formal Concepts:

- $C1 = (\{L,I,G,H\}, \{ \}) = (\{ \} \downarrow, \{ \} \uparrow)$
- $C2 = (\{L,G\}, \{0,3\})$
- $C3 = (\{I,H\}, \{1,2,4\})$
- $C4 = (\{L,G,H\}, \{3\})$
- $C5 = (\{ \}, \{0,1,2,3,4\})$
- $C6 = (\{H\}, \{1,2,3,4\})$



Step by Step

input Formal Concept: (C2,1), (C3,2), (C4,4)

Reduce iteration 2

$$\{0,3\} \cap Y1 = \{0,1,2,3,4\} \cap Y1$$

$$\{0,3\} \cap Y2 \neq \{0,1,2,3,4\} \cap Y2$$

$$\{0,3\} \cap Y4 \neq \{0,1,2,3,4\} \cap Y4$$

$$\{1,2,4\} \cap Y3 = \{1,2,3,4\} \cap Y3$$

$$\{3\} \cap Y4 \neq \{1,2,3,4\} \cap Y4$$

	0	1	2	3	4
L	X			X	
I		X	X		X
G	X			X	
H		X	X	X	X

$Y0 = \{\}, Y1 = \{0\}, Y2 = \{0,1\}, \text{etc.}$

All Formal Concepts:

$$C1 = (\{L,I,G,H\}, \{\}) = (\{\}\downarrow, \{\}\uparrow)$$

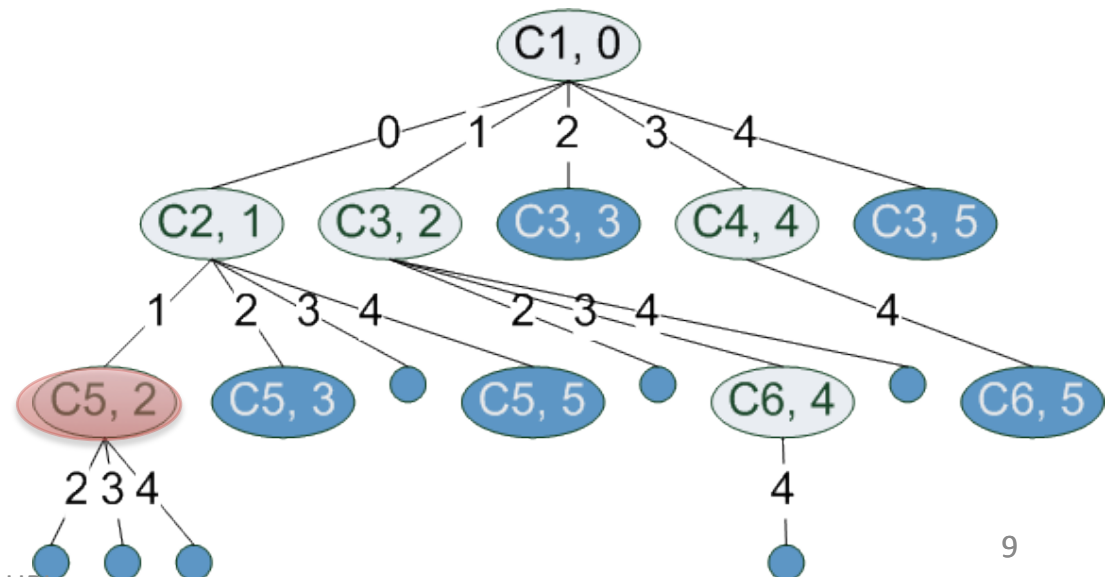
$$C2 = (\{L,G\}, \{0,3\})$$

$$C3 = (\{I,H\}, \{1,2,4\})$$

$$C4 = (\{L,G,H\}, \{3\})$$

$$C5 = (\{\}, \{0,1,2,3,4\})$$

$$C6 = (\{H\}, \{1,2,3,4\})$$



The Algorithm

- $\text{MapConcepts}(\langle B_0, y \rangle, \langle C, D \rangle) \rightarrow \langle D, j \rangle, \langle E, F \rangle$
 - Generates new Formal Concepts by applying \downarrow and \uparrow operators
 - New Objects: Set of old objects which share the new attribute
 - New Attributes: those shared by all new objects.
- $\text{ReduceConcepts}(\langle D, j \rangle, \langle E, F \rangle) \rightarrow \langle D, j + 1 \rangle, \langle E, F \rangle$
 - Ensures that each Formal Concept is emitted exactly once
- Use emitted Formal Concepts as new input to MapConcepts

```
// Y = set of all attributes
for j = y to |Y| do
  if j ∈ D then continue;
  set E to C ∩ {j}↓
  set F to E↑
  emit ⟨⟨B0, j⟩, ⟨E, F⟩⟩
end
```

```
Yj = { all attributes
        smaller than j }
if D ∩ Yj == F ∩ Yj
  emit ⟨⟨D, j+1⟩, ⟨E, F⟩⟩
end
```

Plans

- Reproduce measurements from paper
- Find ideal number of nodes for data sets
- Evaluate possible performance improvements

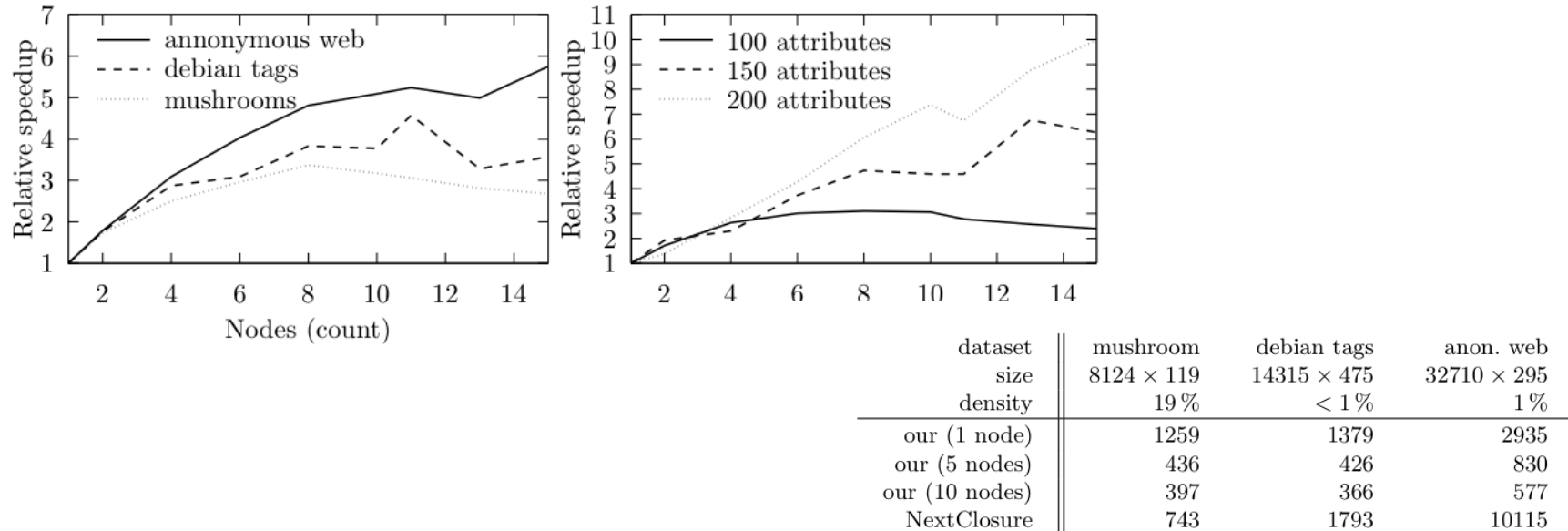


Fig. 4. Time needed to compute all formal concepts for various datasets (in seconds)