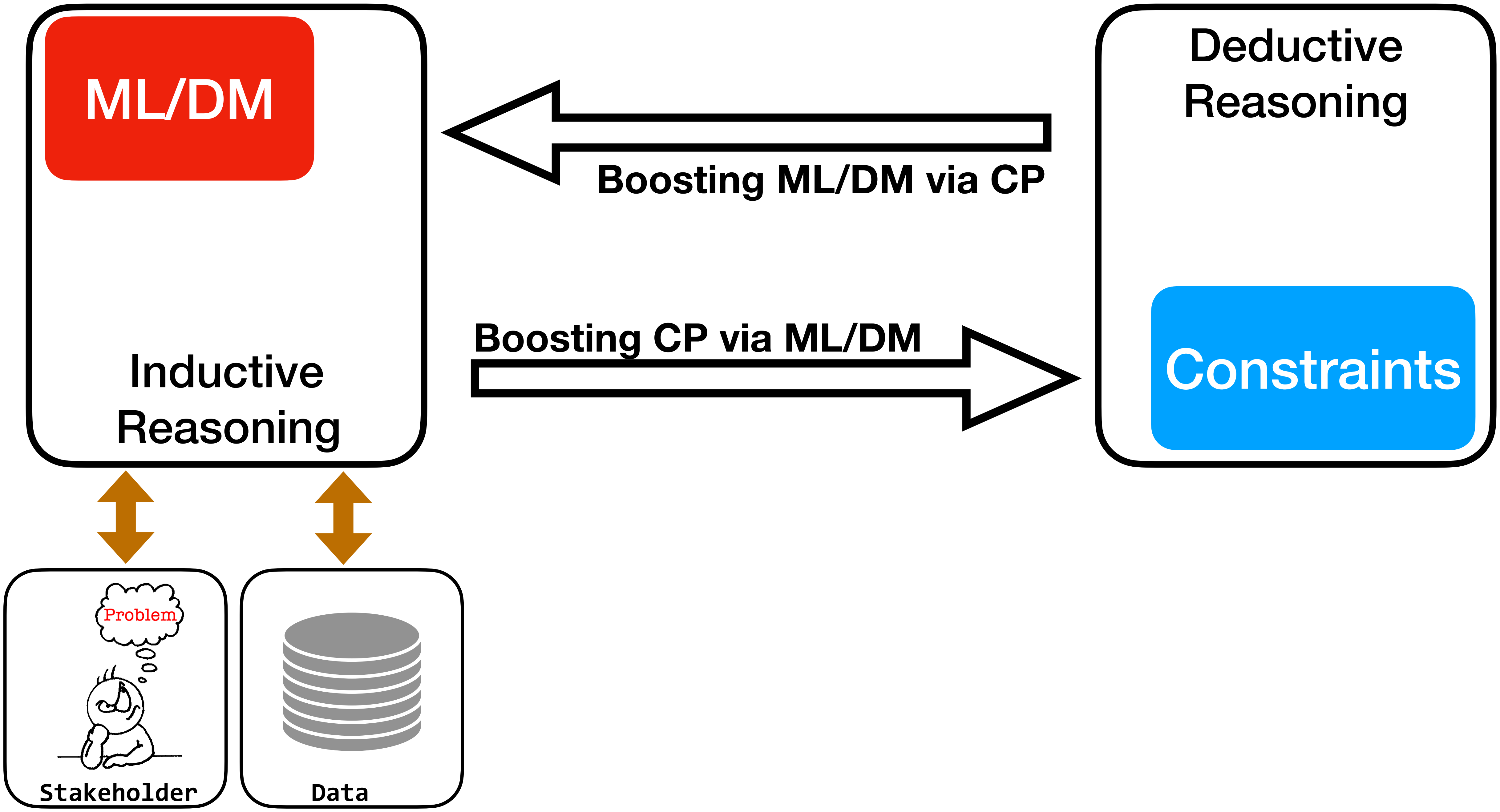
Declarative Itemset Mining

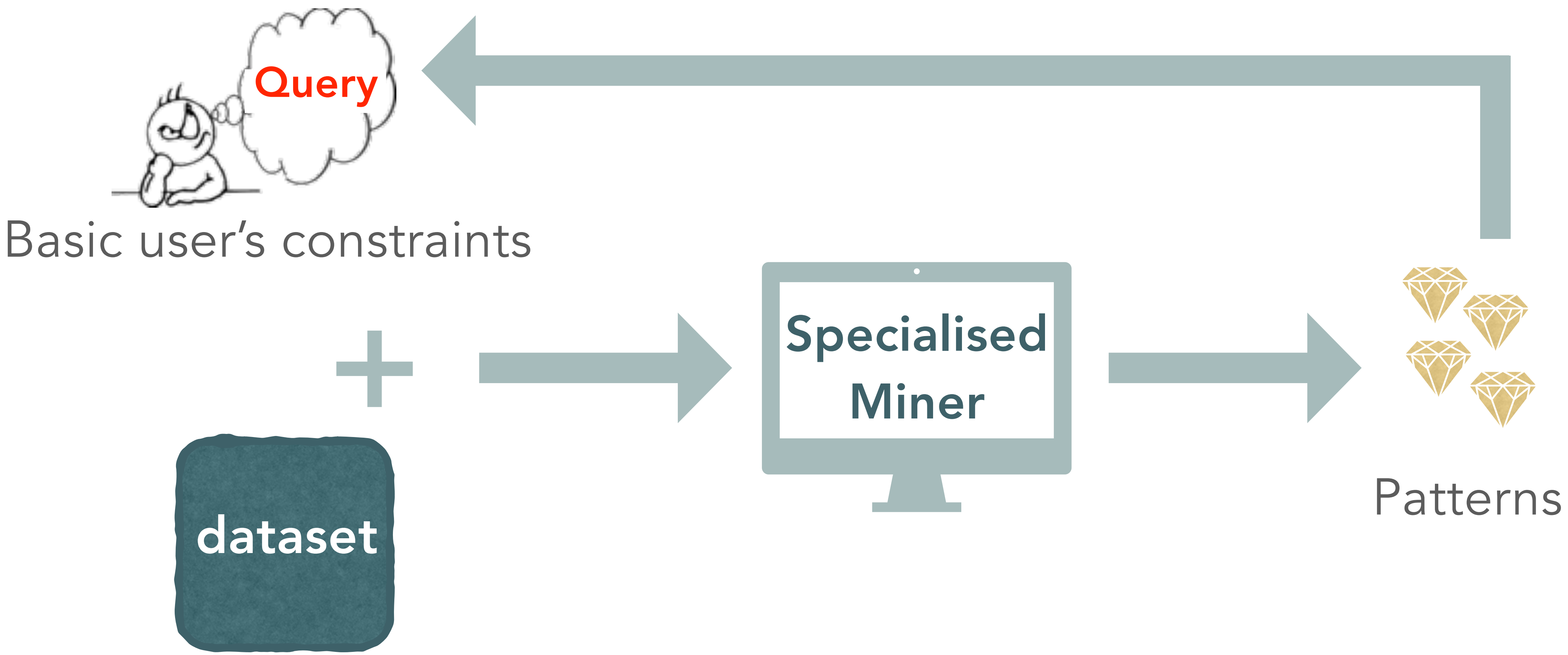
Master 1 - IDS

Nadjib Lazaar
Ing - Phd - HDR - Professor - Paris-Saclay University - LISN - LaHDAK
lazaar@lisn.fr https://perso.lisn.upsaclay.fr/lazaar/
03/02/2025

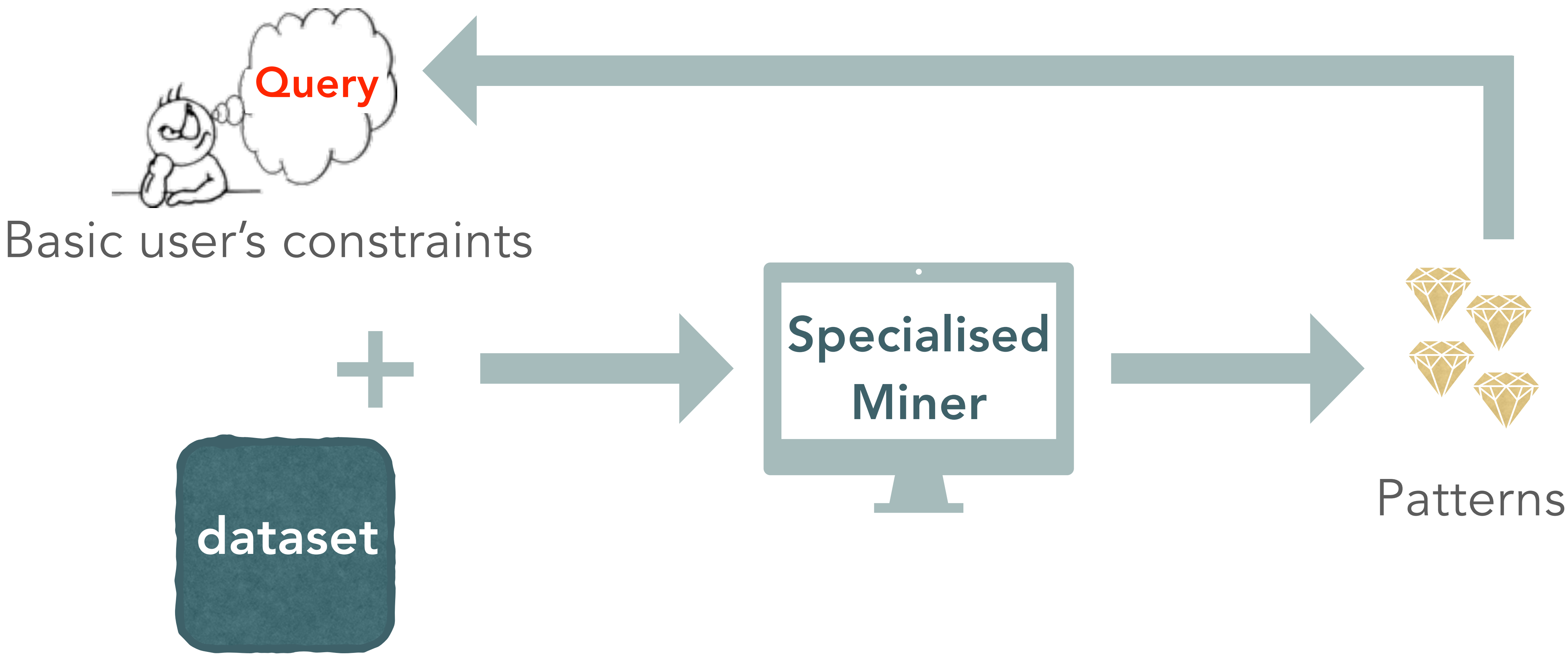
CONTEXT


ML/DM
Inductive Reasoning
Boosting ML/DM via CP
Boosting CP via ML/DM
Deductive Reasoning
Constraints
Problem
Stakeholder
Data

SPECIALISED VS DECLARATIVE PATTERN MINING

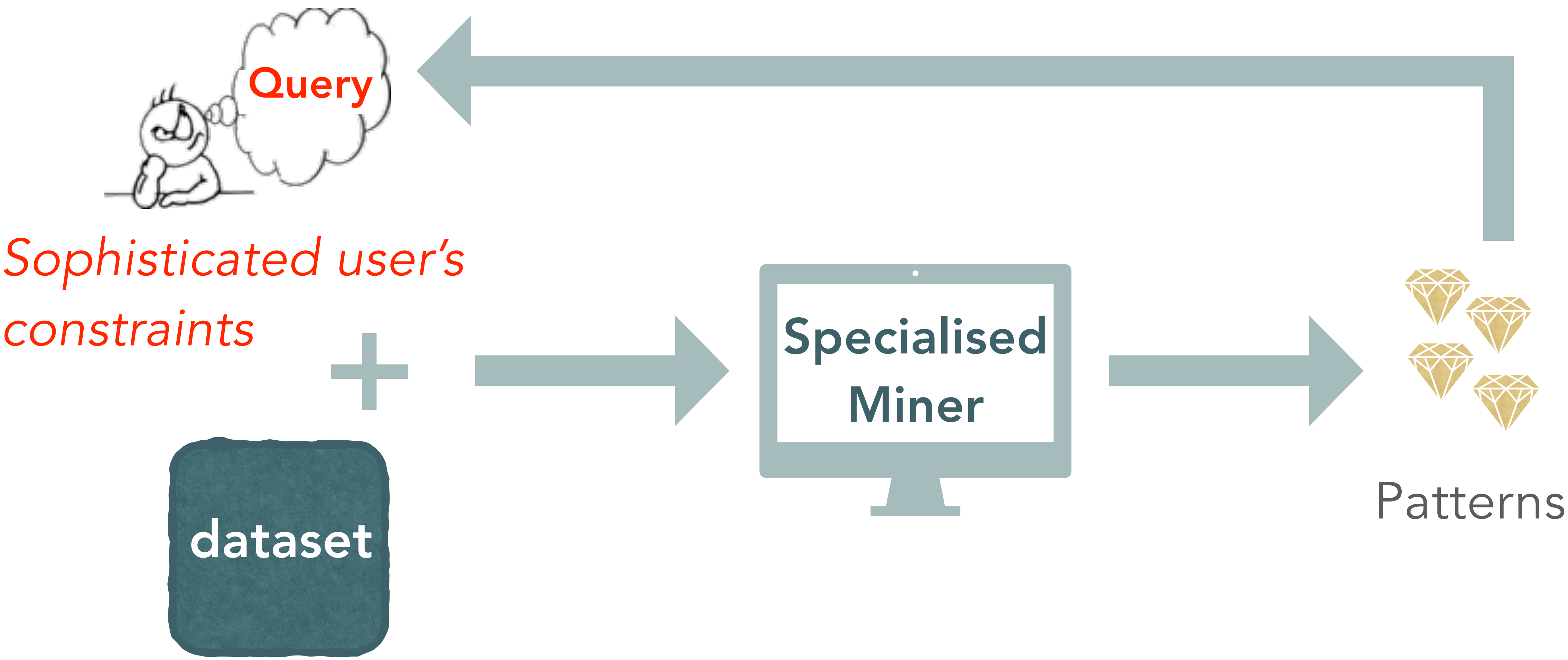


SPECIALISED VS DECLARATIVE PATTERN MINING



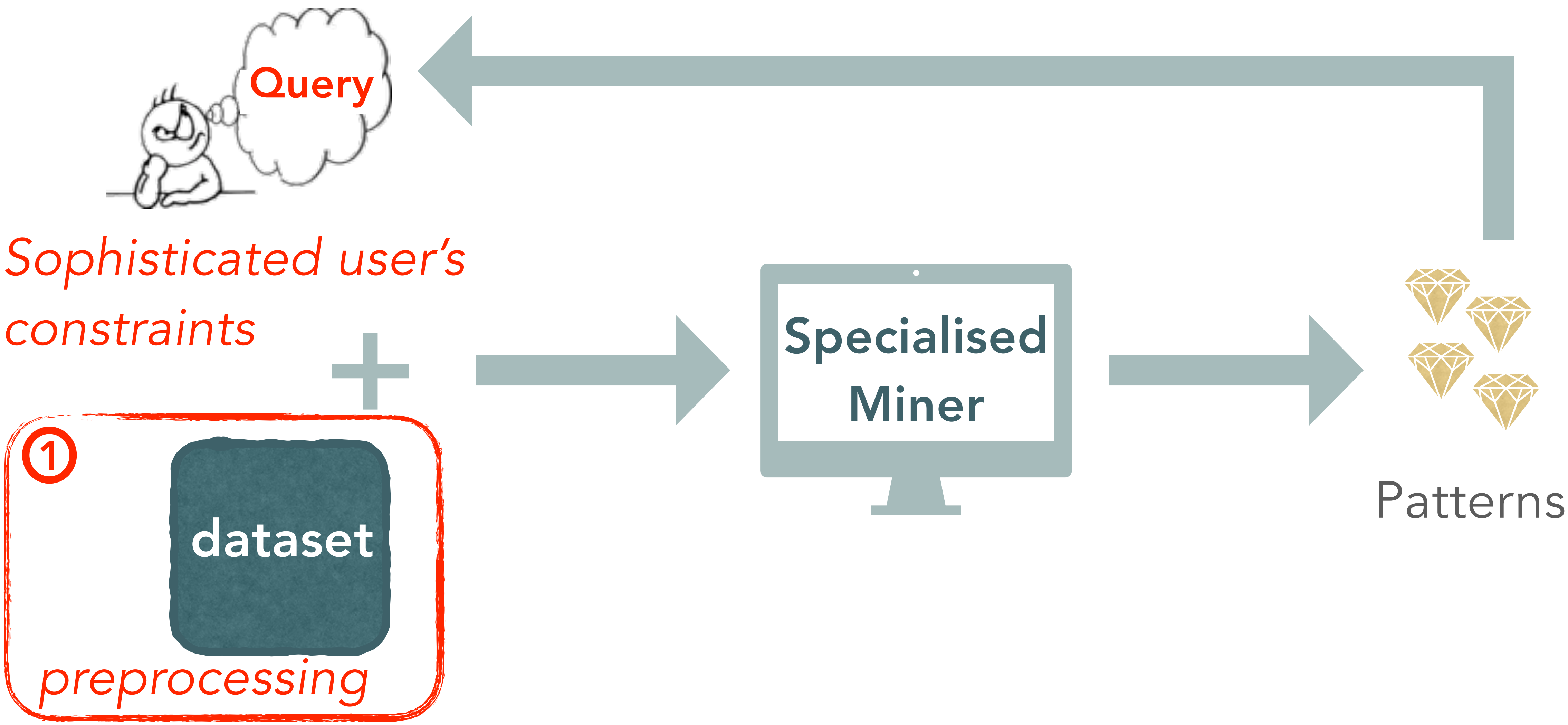
Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

SPECIALISED VS DECLARATIVE PATTERN MINING



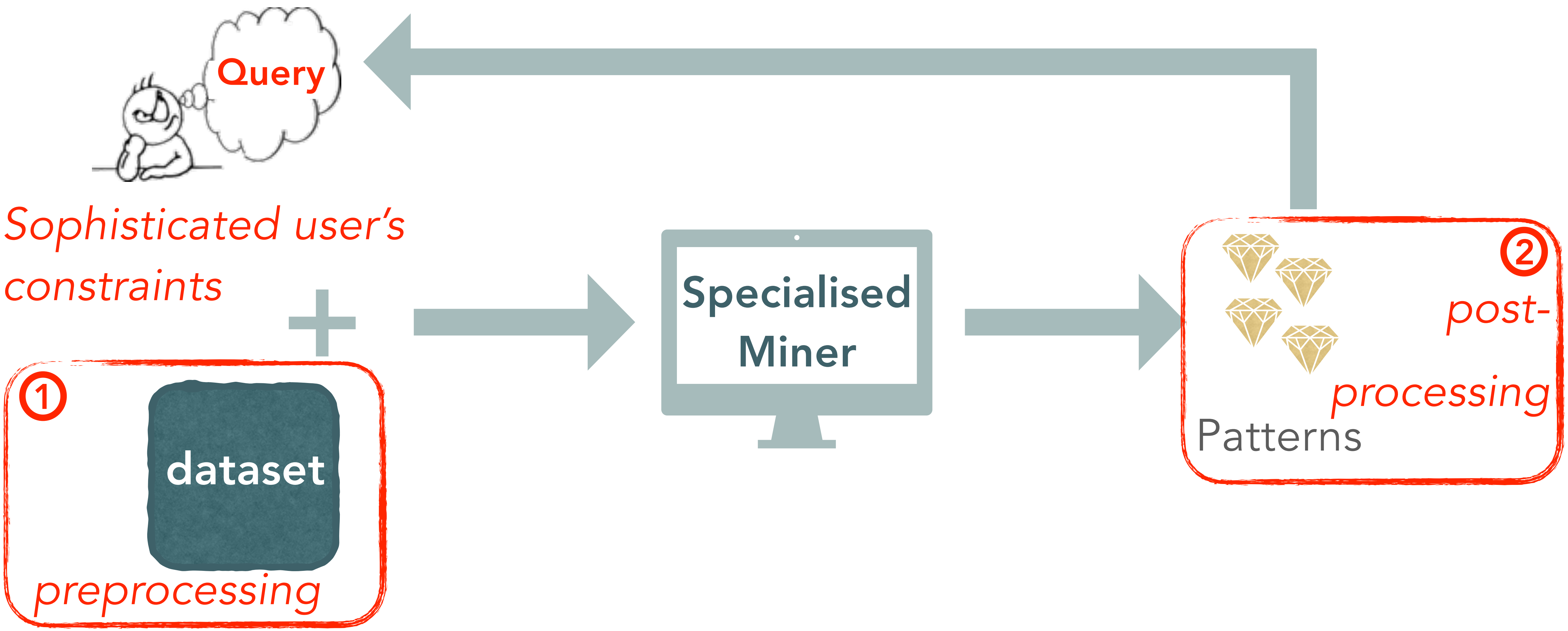
Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

SPECIALISED VS DECLARATIVE PATTERN MINING



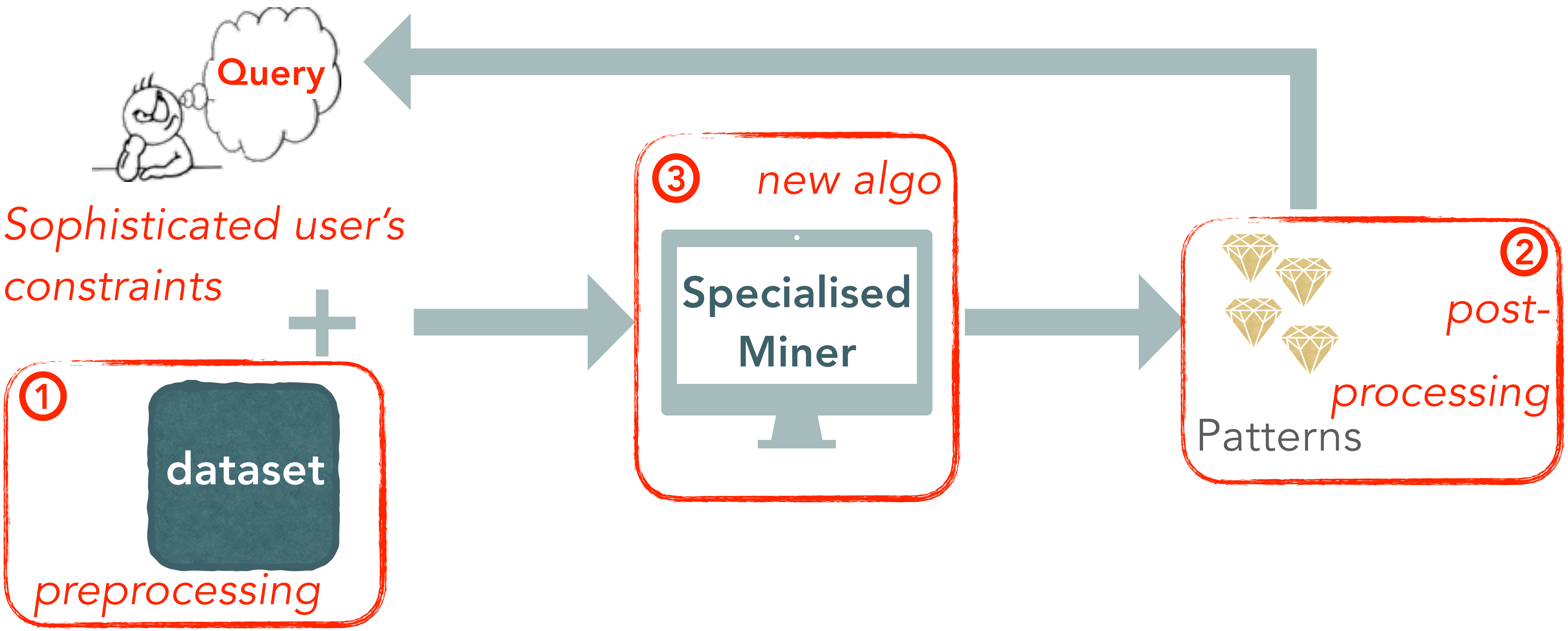
Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

SPECIALISED VS DECLARATIVE PATTERN MINING



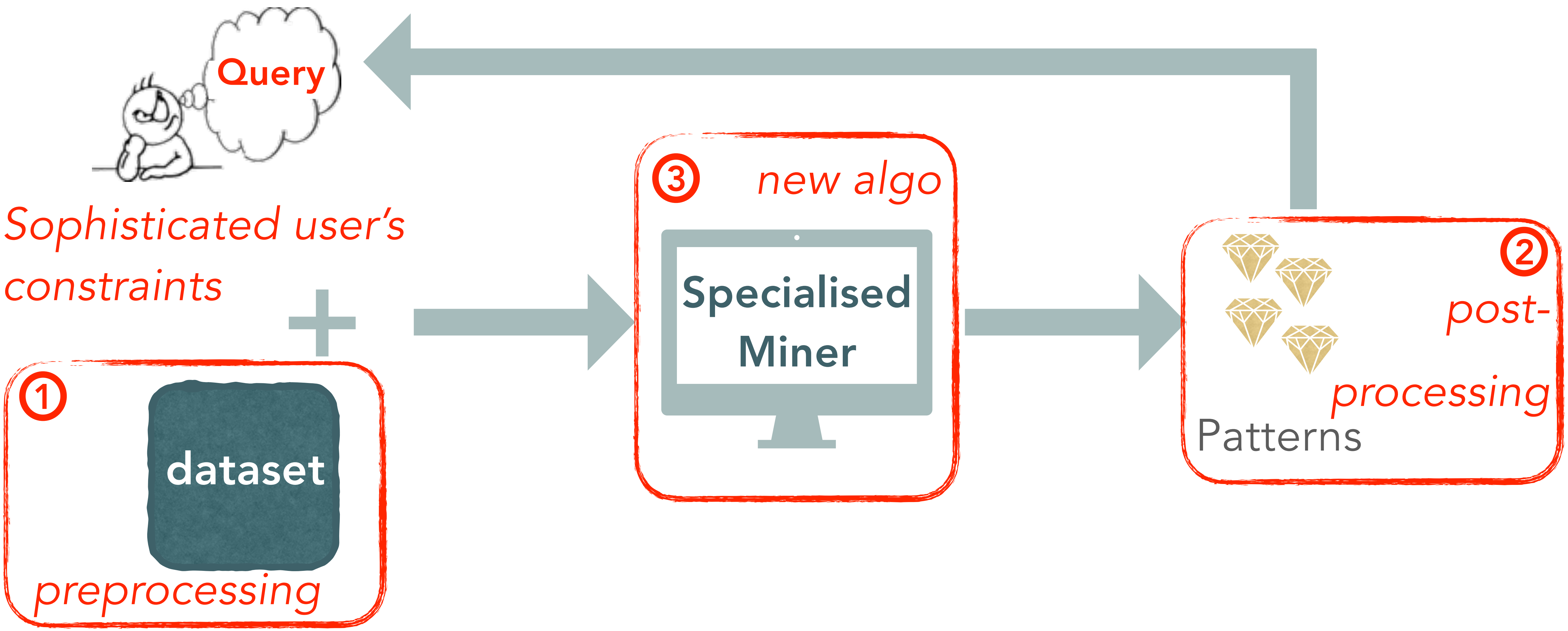
Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

SPECIALISED VS DECLARATIVE PATTERN MINING



Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

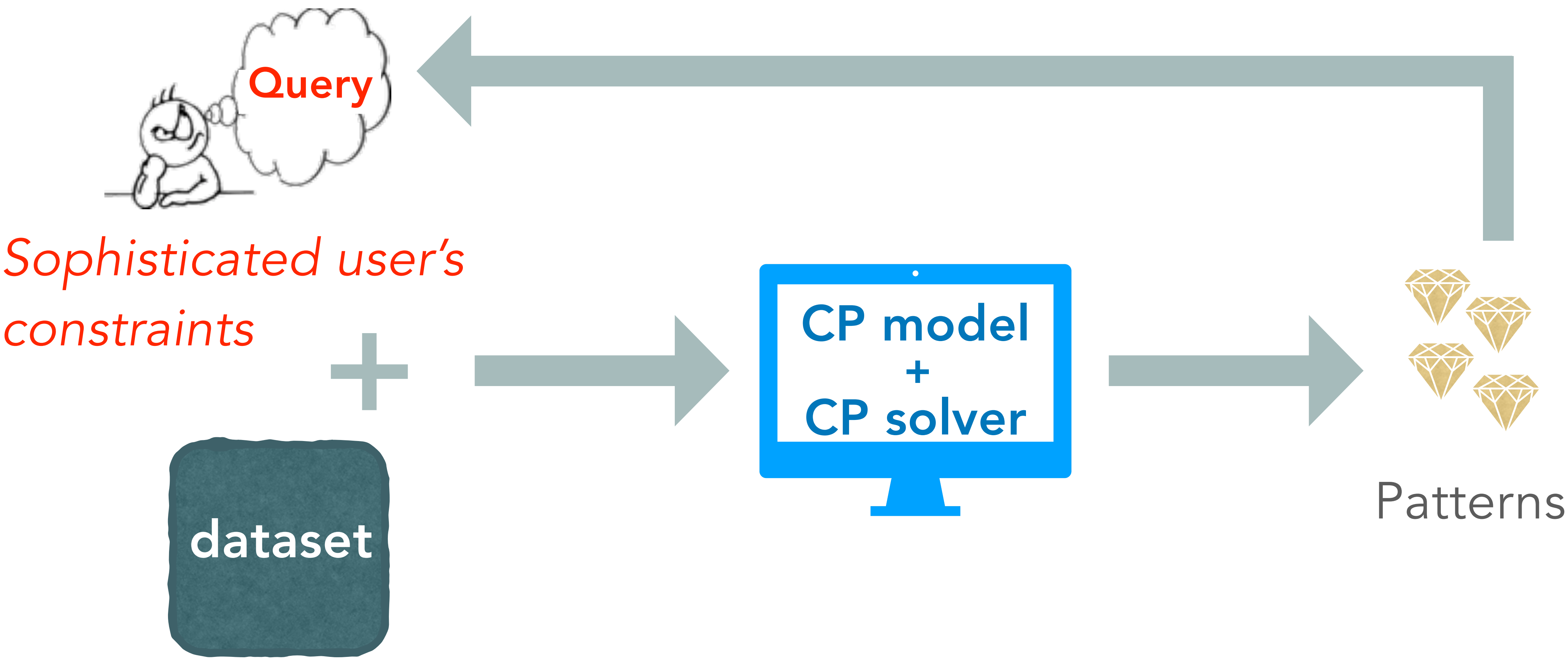
SPECIALISED VS DECLARATIVE PATTERN MINING



Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

Need: *Declarative way to deal with more complex queries and iterative process*

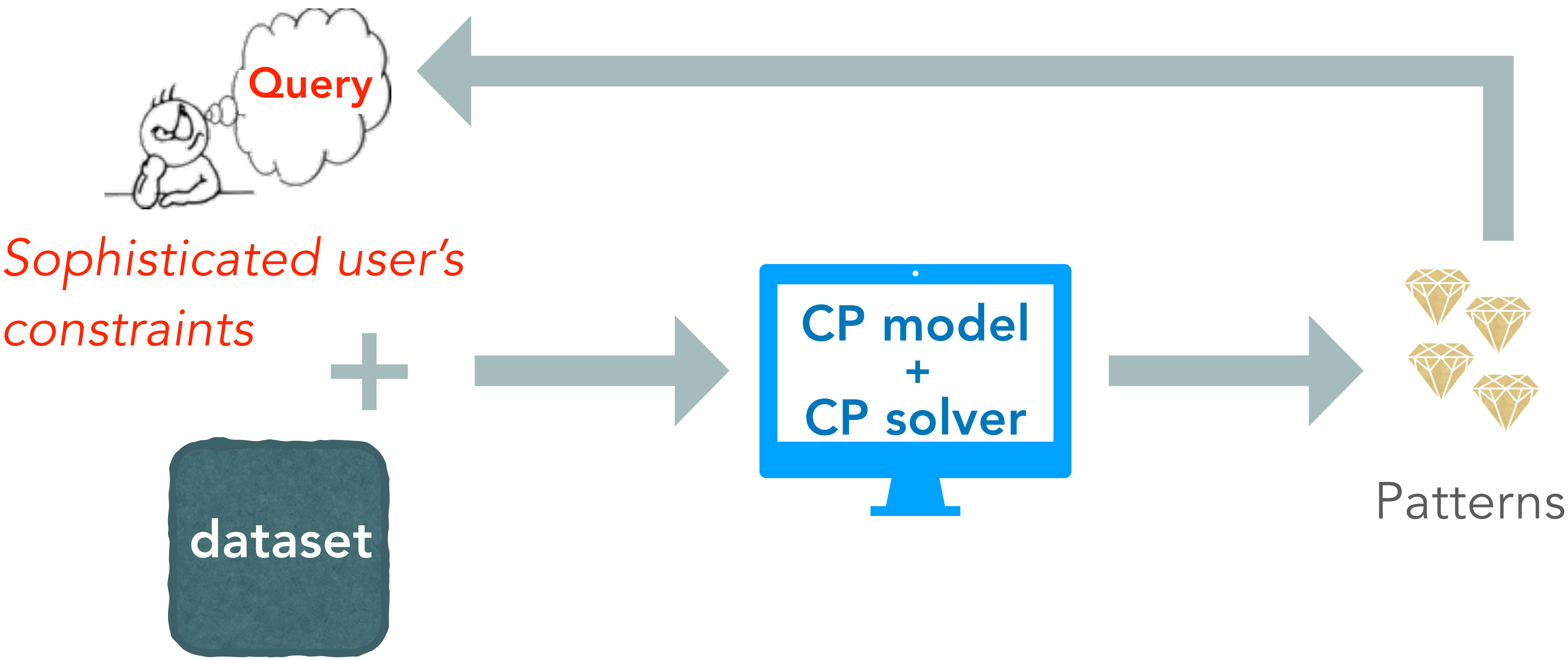
SPECIALISED VS DECLARATIVE PATTERN MINING



Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

Need: *Declarative way to deal with more complex queries and iterative process*

SPECIALISED VS DECLARATIVE PATTERN MINING

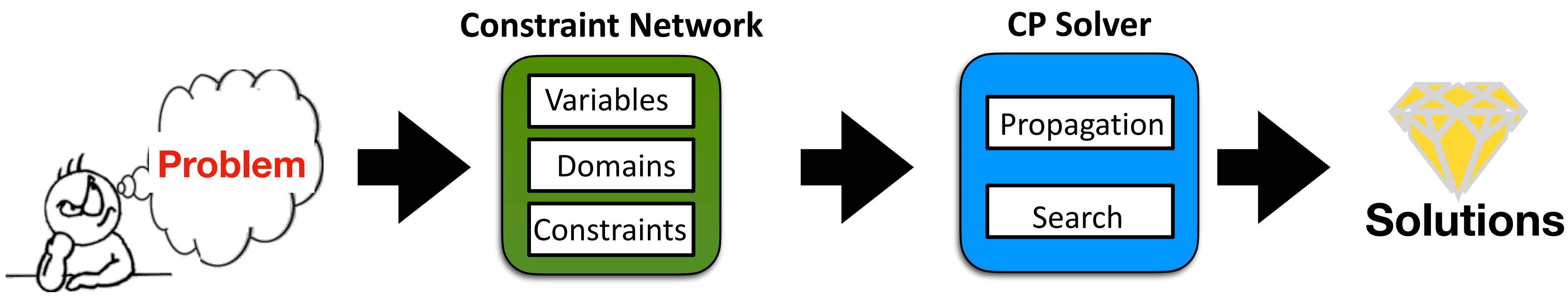


Limitations: Dealing with sophisticated user's constraints **[Wojciechowski and Zakrzewicz, 02]**

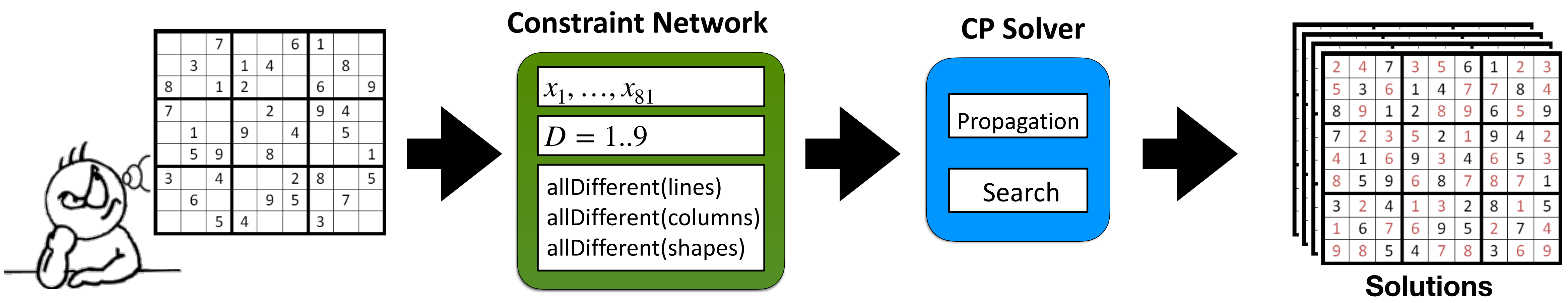
Need: *Declarative way to deal with more complex queries and iterative process*

CONSTRAINT PROGRAMMING

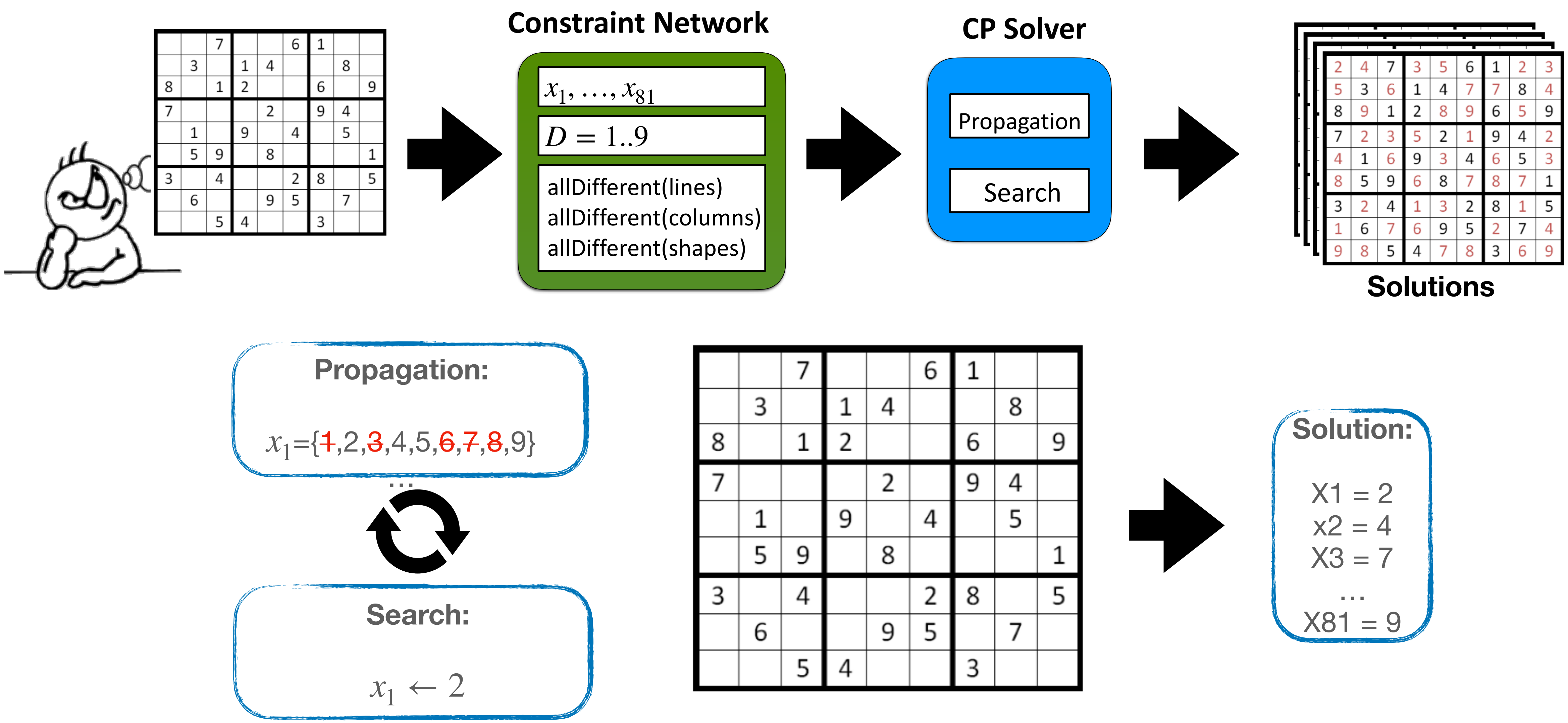
[ROSSI ET AL, CP HANDBOOK 06]



CONSTRAINT PROGRAMMING

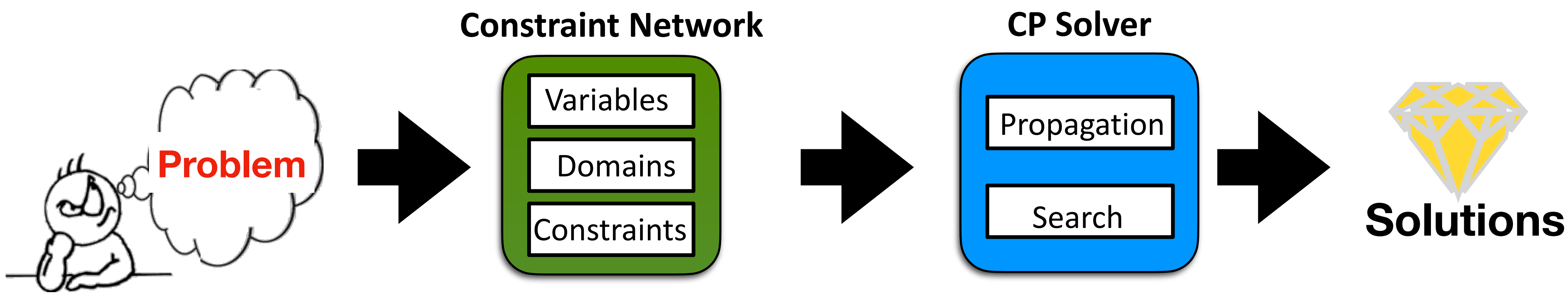


CONSTRAINT PROGRAMMING



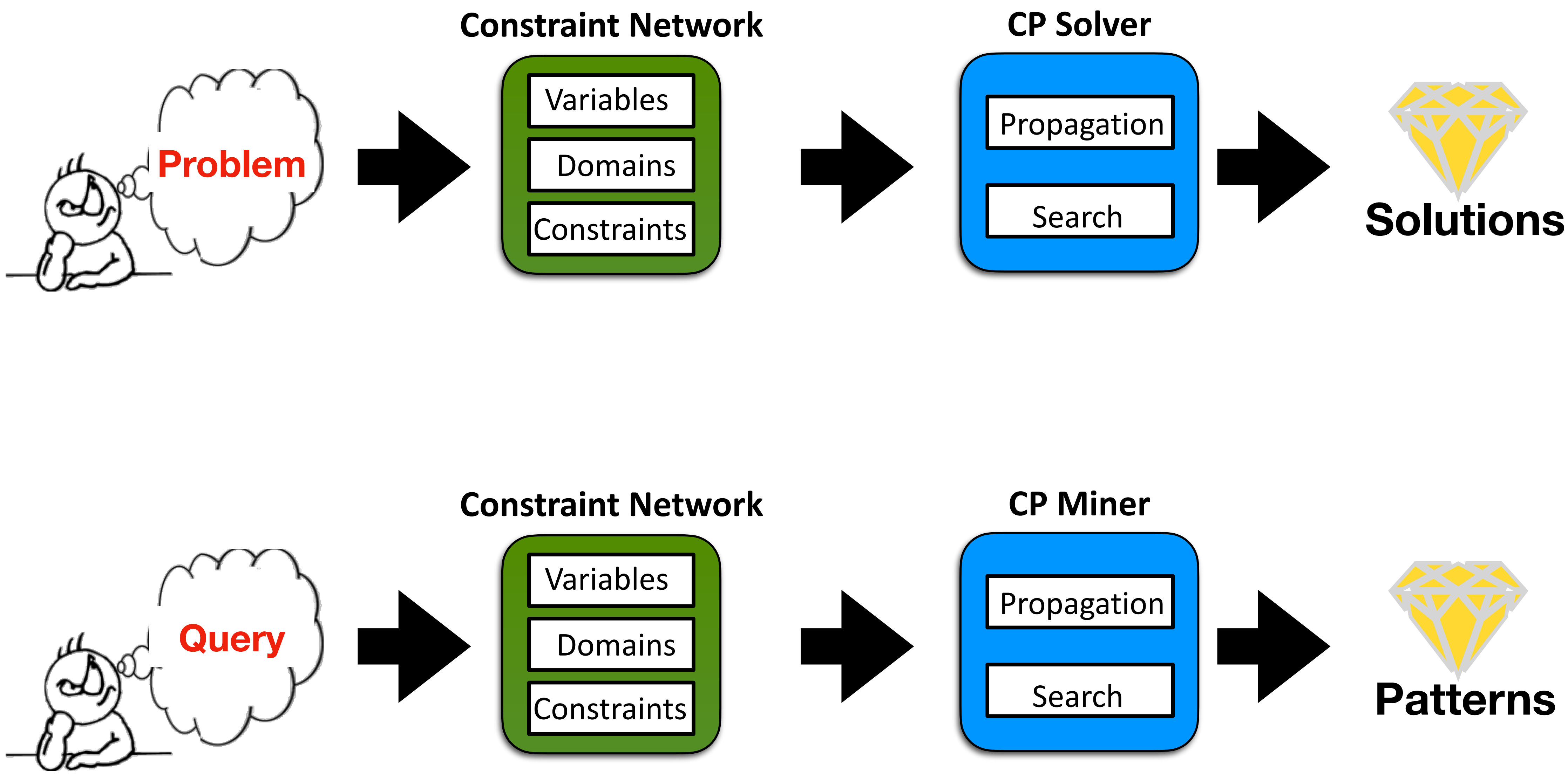
CONSTRAINT PROGRAMMING

[ROSSI ET AL, CP HANDBOOK 06]



CONSTRAINT PROGRAMMING

[ROSSI ET AL, CP HANDBOOK 06]



FREQUENT ITEMSET MINING (FIM)

- Aims at finding regularities in transactional databases
- **Given:**
  - A set of items $I = \{i_1, \ldots, i_n\}$
  - A set of transactions over the items $T = \{t_1, \ldots, t_m\}$
  - A minimum frequency threshold α
- **The need:**
  - The set of itemset $P \subseteq I$: $\quad freq(P) \geq \alpha$

EXAMPLE

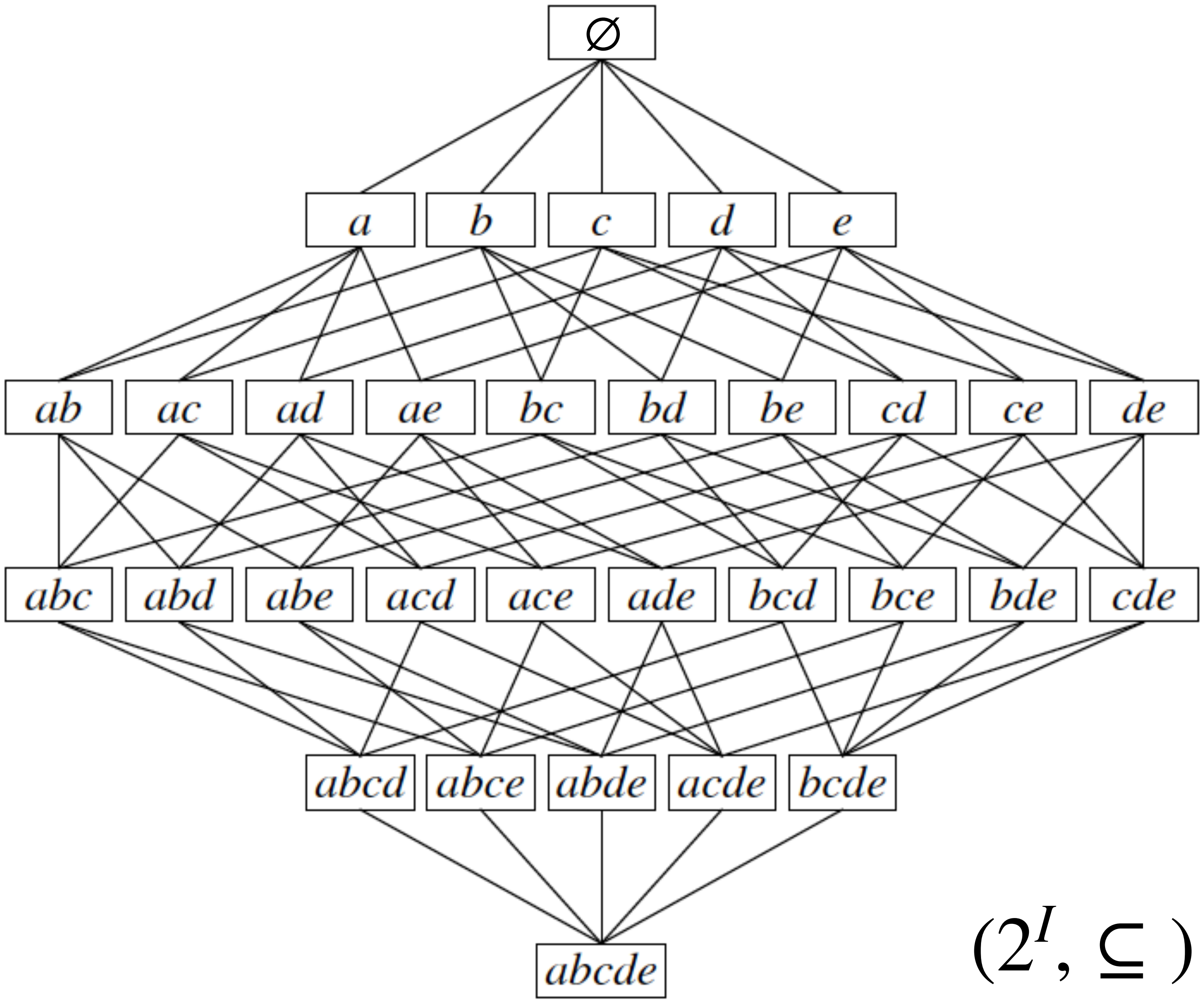
	a	b	c	d	e
t_1:	a			d	e
t_2:		b	c	d	
t_3:	a		c		e
t_4:	a		c	d	e
t_5:	a				e
t_6:	a		c	d	
t_7:		b	c		
t_8:	a		c	d	e
t_9:		b	c		e
t_{10}:	a			d	e

EXAMPLE

t_1:	a			d	e
t_2:		b	c	d	
t_3:	a		c		e
t_4:	a		c	d	e
t_5:	a				e
t_6:	a		c	d	
t_7:		b	c		
t_8:	a		c	d	e
t_9:		b	c		e
t_{10}:	a			d	e

Query: « Mining itemsets of a minimum frequency of $\alpha = 30\,\%$ »

EXAMPLE

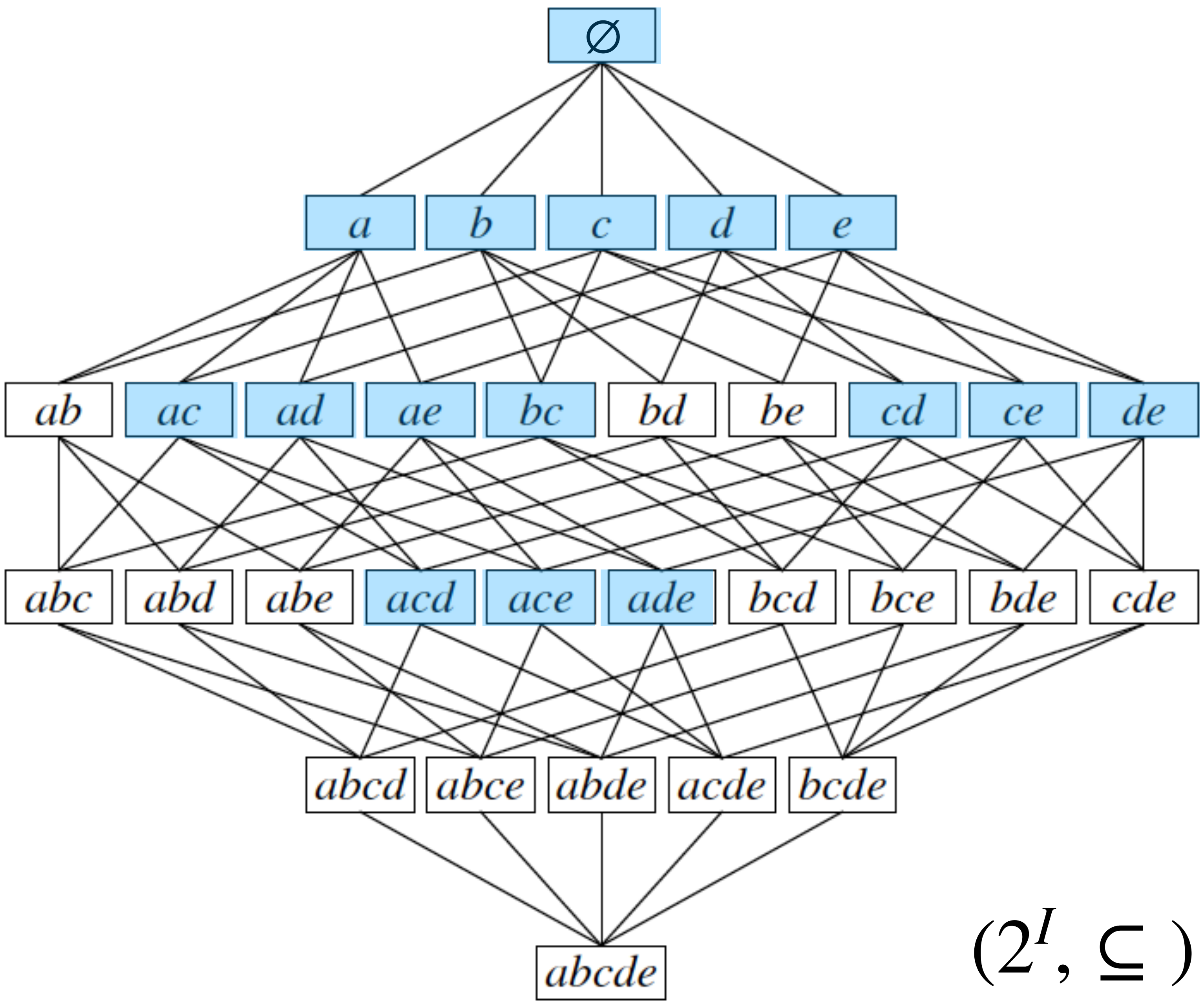


$(2^I, \subseteq)$

t_1:	a			d	e
t_2:		b	c	d	
t_3:	a		c		e
t_4:	a		c	d	e
t_5:	a				e
t_6:	a		c	d	
t_7:		b	c		
t_8:	a		c	d	e
t_9:		b	c		e
t_{10}:	a			d	e

Query: « Mining itemsets of a minimum frequency of $\alpha = 30\,\%$ »

EXAMPLE

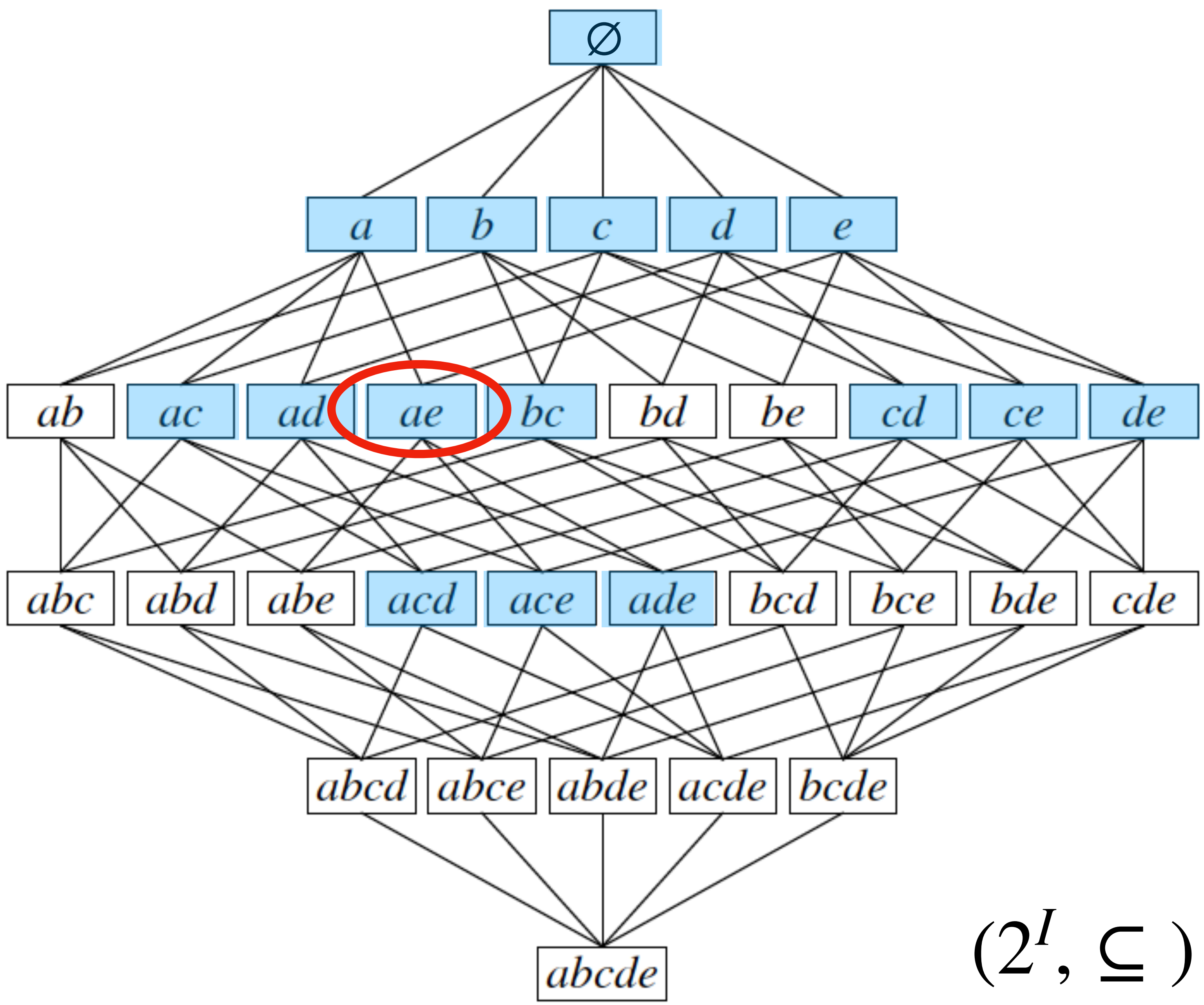


$(2^I, \subseteq)$

t_1:	a			d	e
t_2:		b	c	d	
t_3:	a		c		e
t_4:	a		c	d	e
t_5:	a				e
t_6:	a		c	d	
t_7:		b	c		
t_8:	a		c	d	e
t_9:		b	c		e
t_{10}:	a			d	e

Query: « Mining itemsets of a minimum frequency of $\alpha = 30\,\%$ »

EXAMPLE

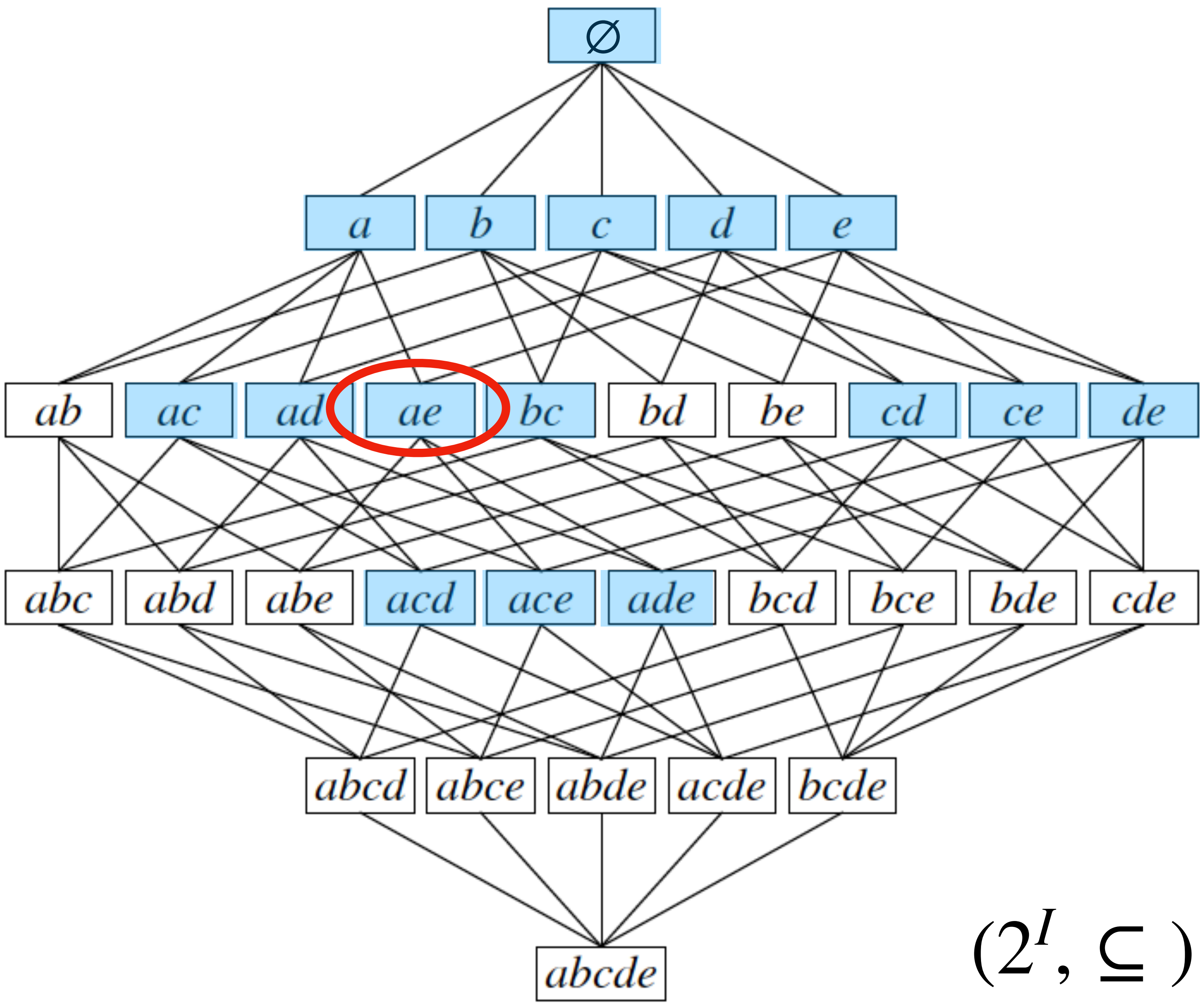


$(2^I, \subseteq)$

t_1:	a			d	e
t_2:		b	c	d	
t_3:	a		c		e
t_4:	a		c	d	e
t_5:	a				e
t_6:	a		c	d	
t_7:		b	c		
t_8:	a		c	d	e
t_9:		b	c		e
t_{10}:	a			d	e

Query: « Mining itemsets of a minimum frequency of $\alpha = 30\,\%$ »

EXAMPLE

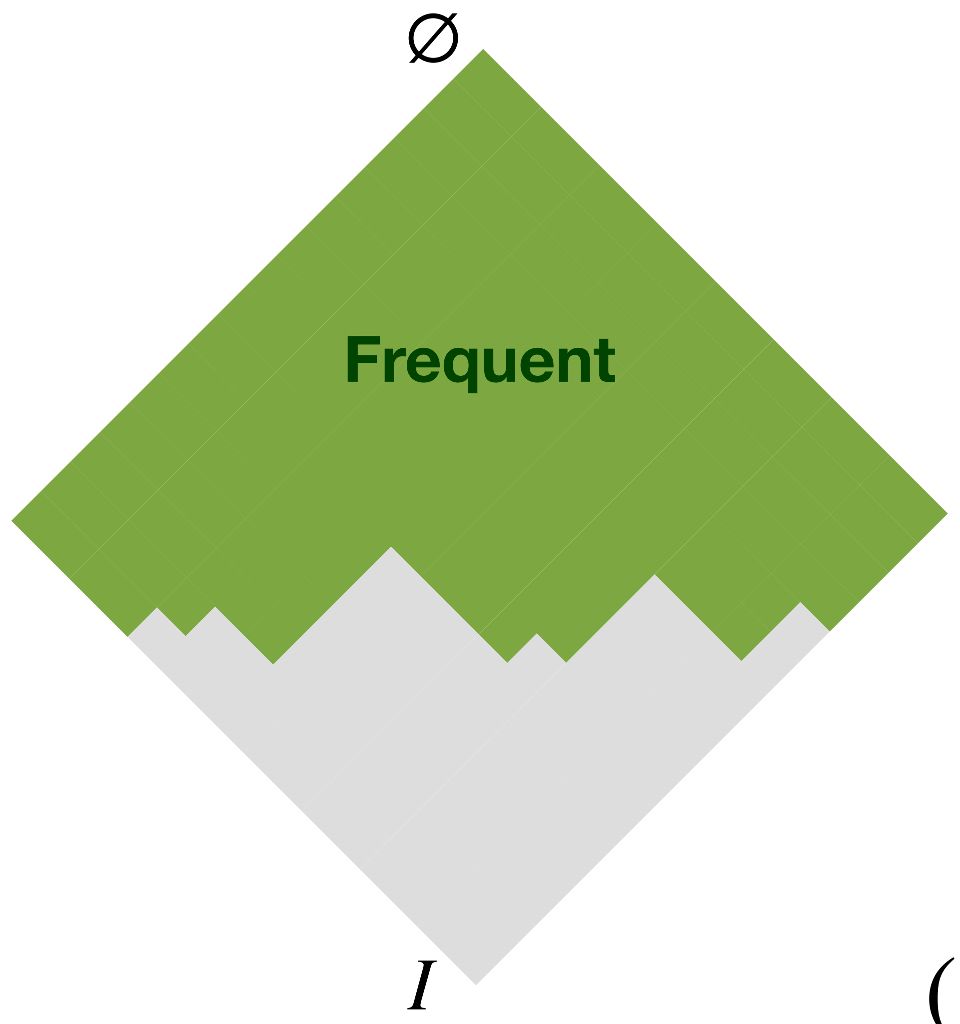


$(2^I, \subseteq)$

t_1:	a			d	e
t_2:		b	c	d	
t_3:	a		c		e
t_4:	a		c	d	e
t_5:	a				e
t_6:	a		c	d	
t_7:		b	c		
t_8:	a		c	d	e
t_9:		b	c		e
t_{10}:	a			d	e

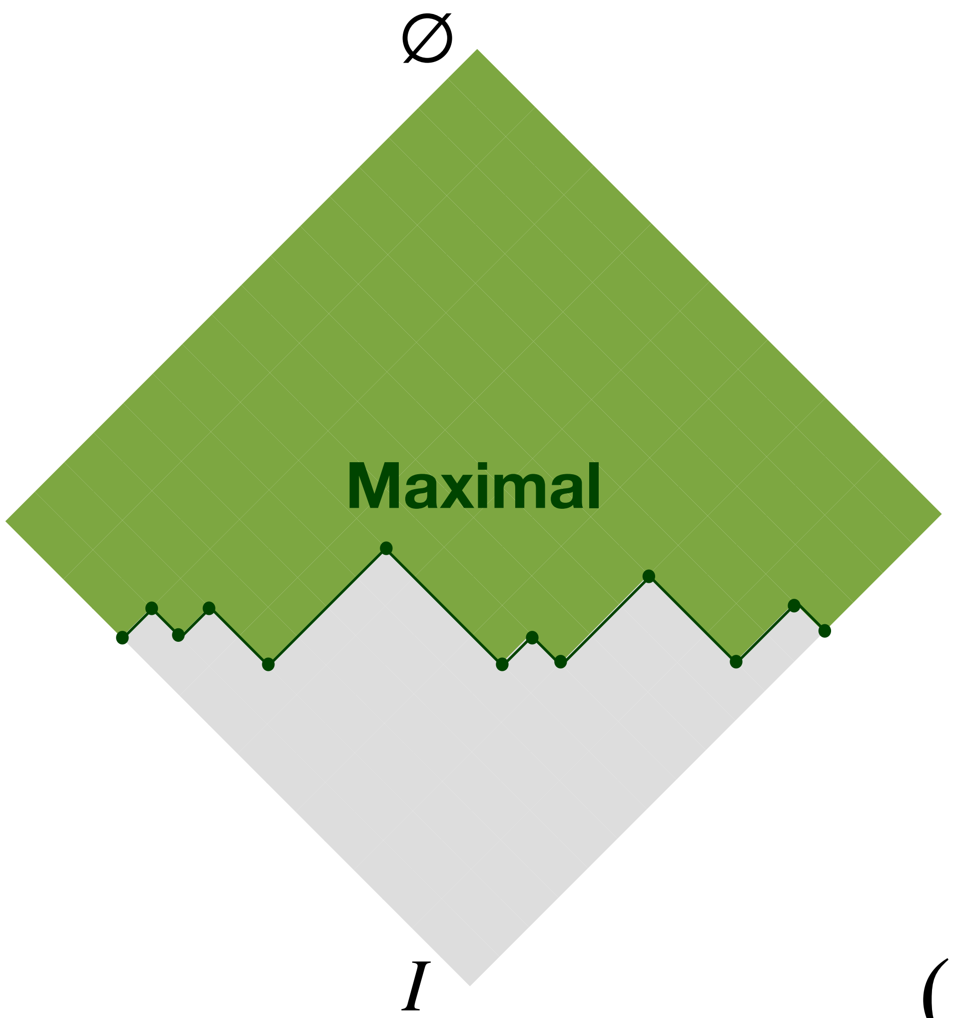
Query: « Mining itemsets of a minimum frequency of $\alpha = 30\,\%$ »

CONDENSED REPRESENTATIONS



$(2^I, \subseteq)$

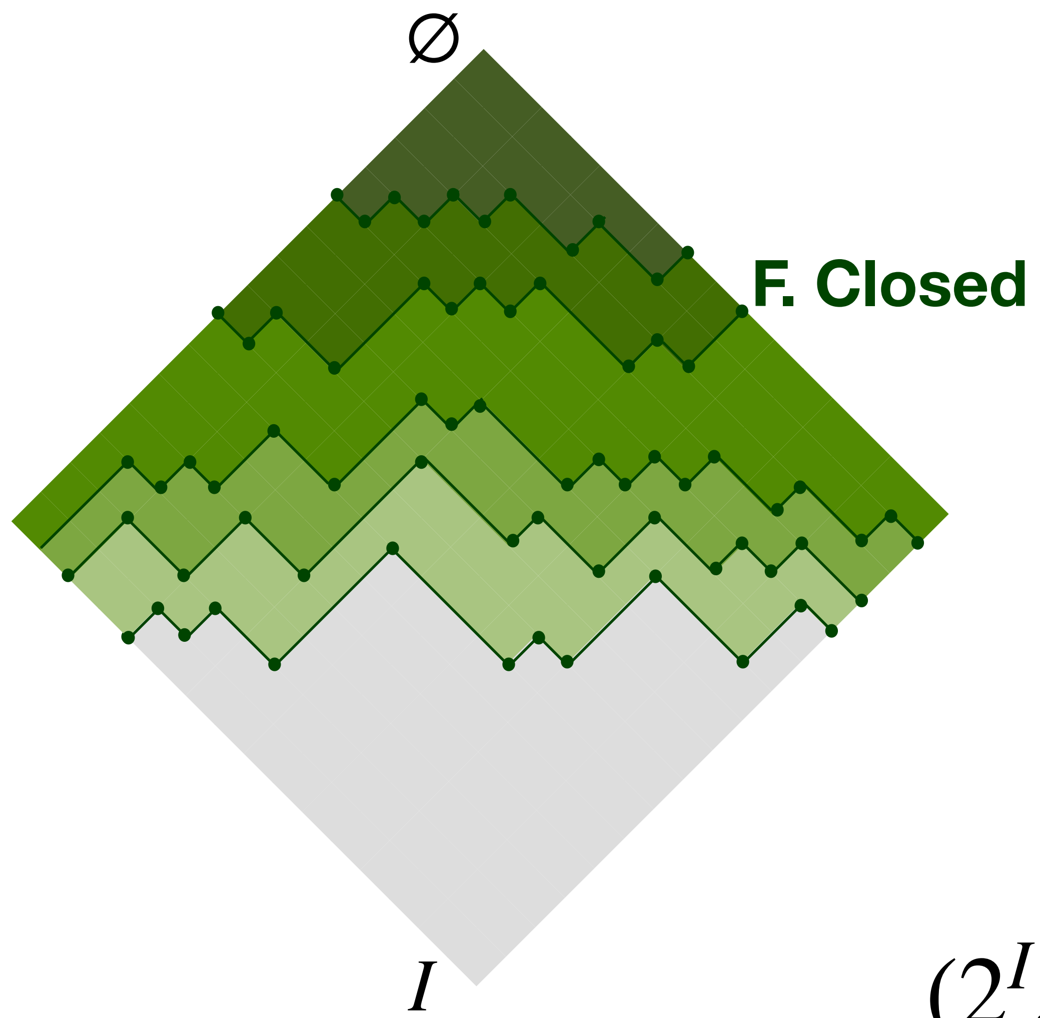
CONDENSED REPRESENTATIONS



$(2^I, \subseteq)$

$$Max_\alpha = \{P \subset I \mid freq(P) \geq \alpha \wedge \forall P' \supset P : freq(P') < \alpha\}$$

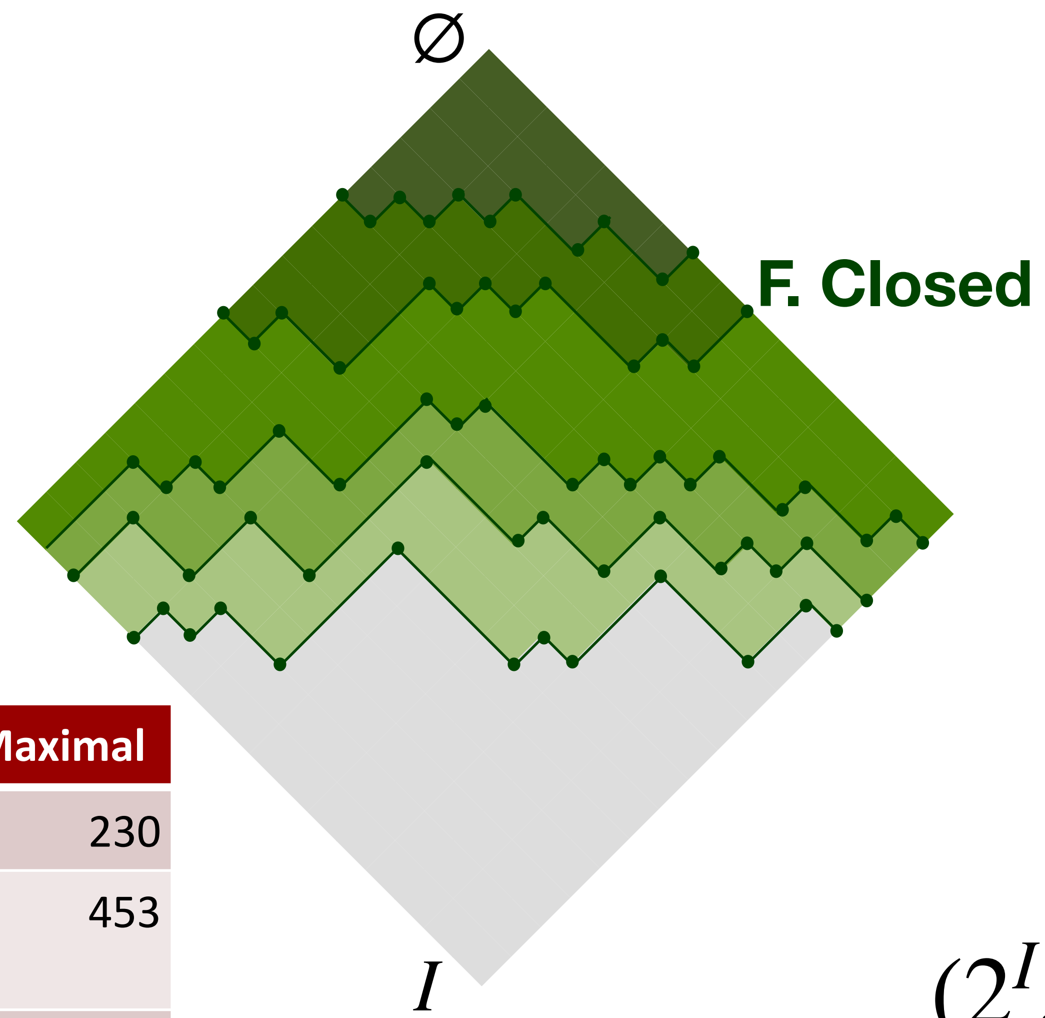
CONDENSED REPRESENTATIONS



$(2^I, \subseteq)$

$$C_\alpha = \{P \subset I \mid freq(P) \geq \alpha \wedge \forall P' \supset P : freq(P') < freq(P)\}$$

CONDENSED REPRESENTATIONS

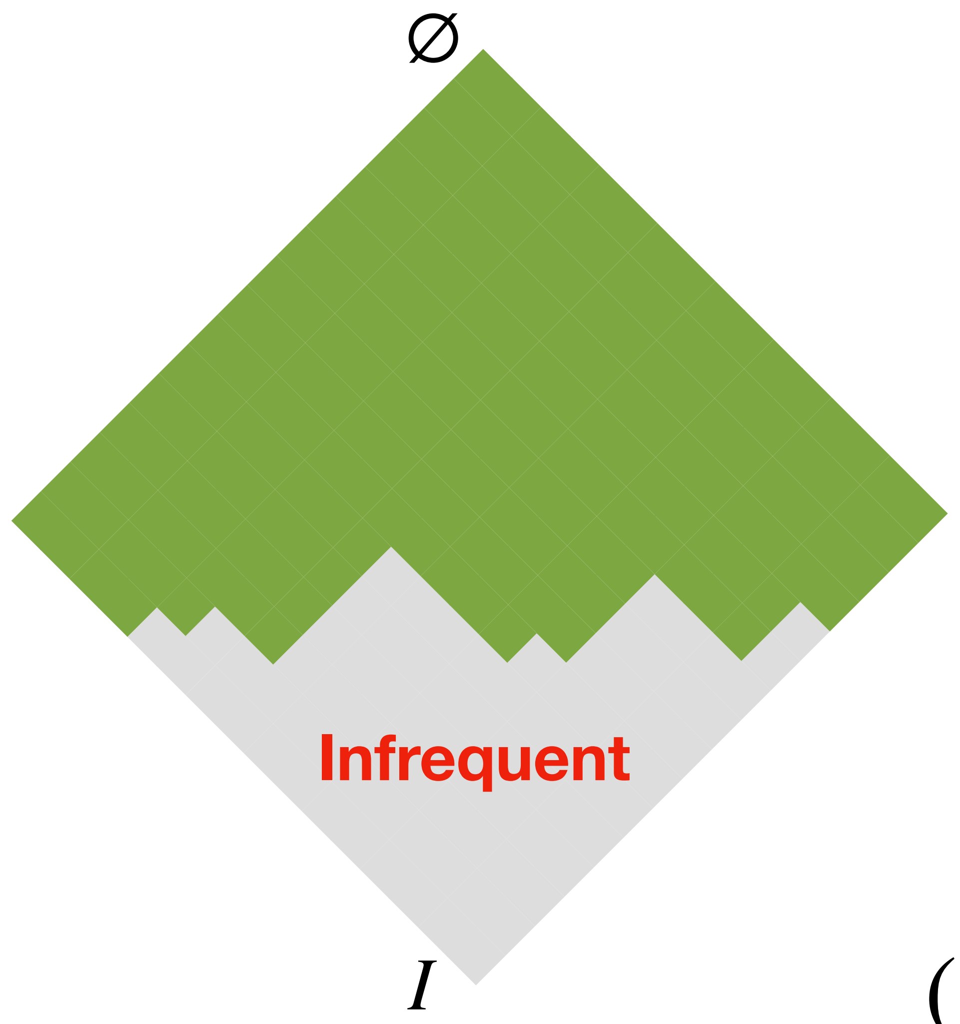


$(2^I, \subseteq)$

Dataset	#Frequent	#Closed	#Maximal
Zoo-1	151 807	3 292	230
Mushroom	155 734	3 287	453
Lymph	9 967 402	46 802	5 191
Hepatitis	27 . 10⁷+	1 827 264	189 205

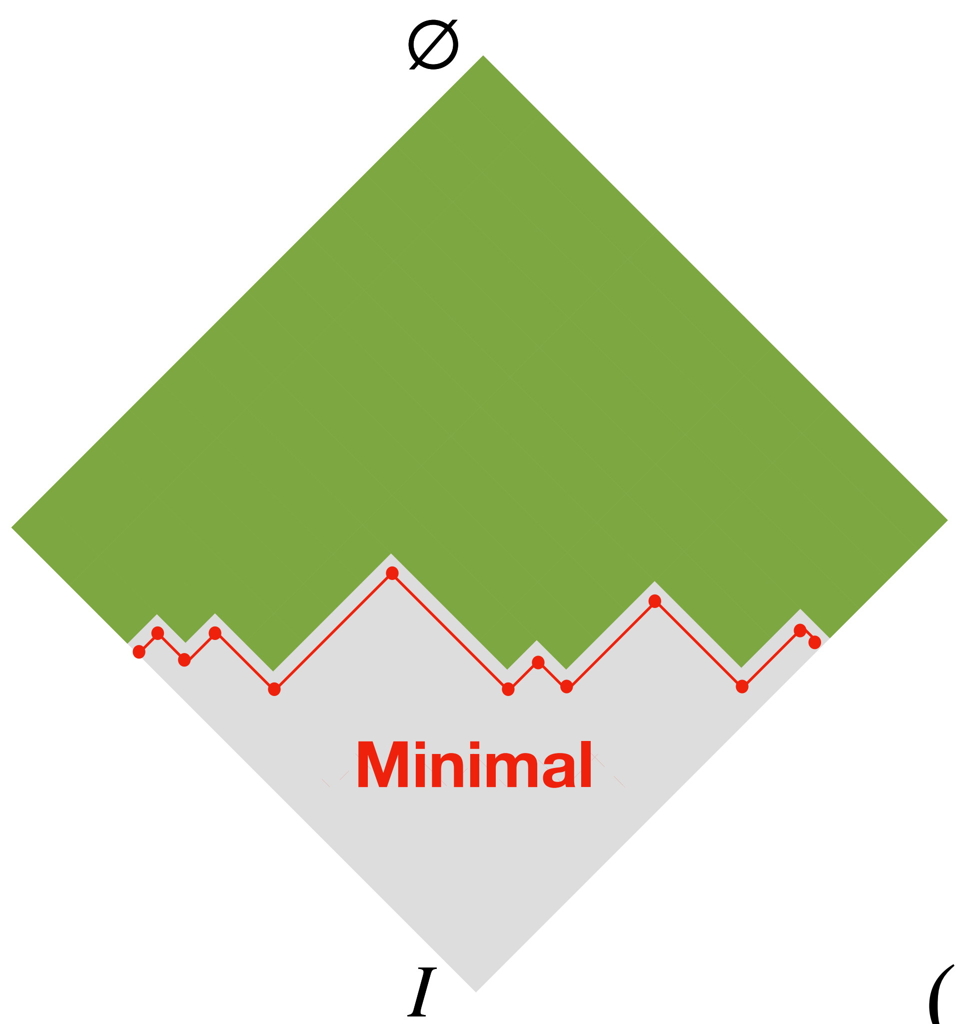
$$C_\alpha = \{P \subset I \mid freq(P) \geq \alpha \wedge \forall P' \supset P : freq(P') < freq(P)\}$$

CONDENSED REPRESENTATIONS



$(2^I, \subseteq)$

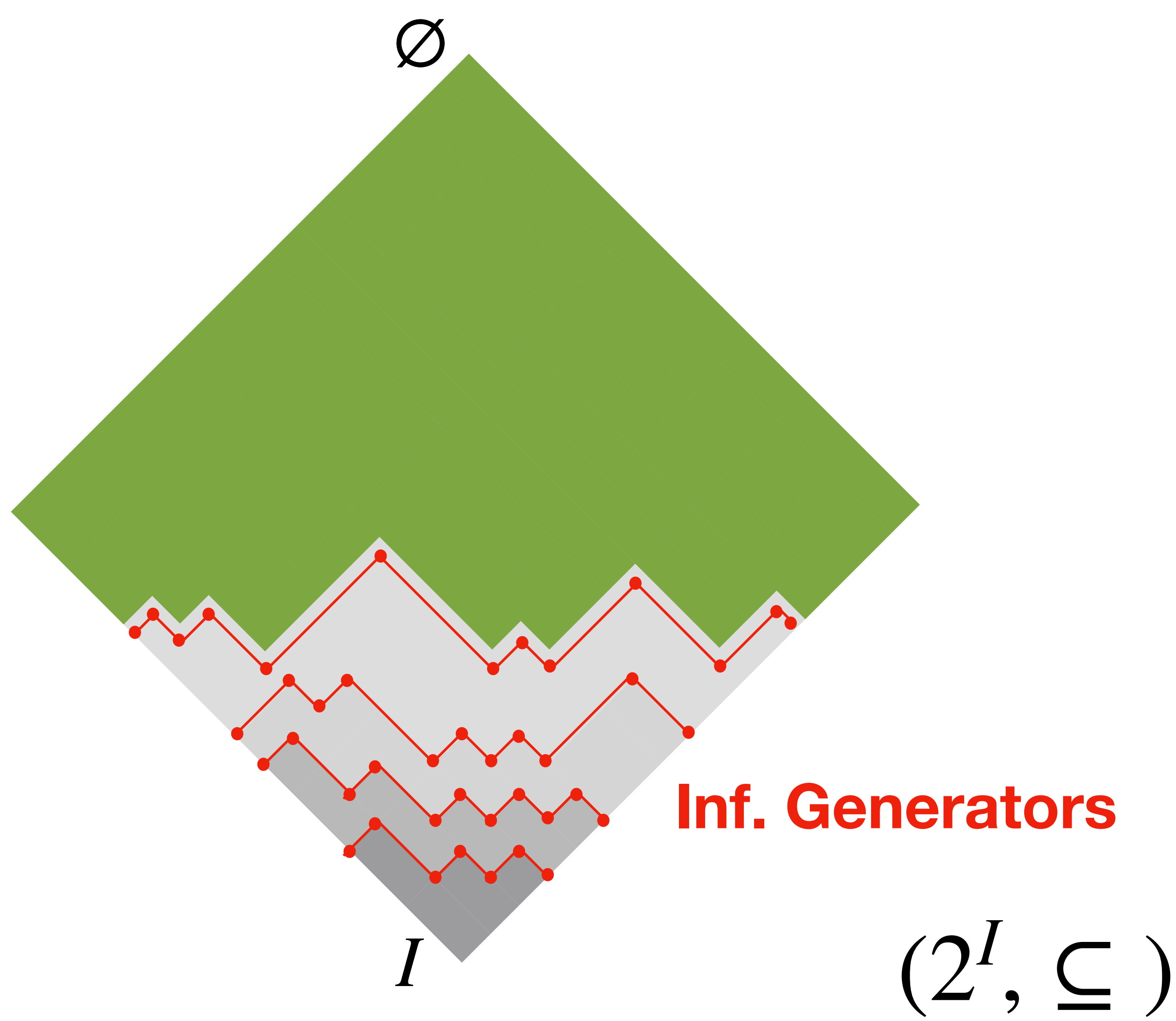
CONDENSED REPRESENTATIONS



$(2^I, \subseteq)$

$$Min_\alpha = \{P \subset I \,|\, freq(P) < \alpha \wedge \forall P' \subset P : freq(P') \geq \theta\}$$

CONDENSED REPRESENTATIONS



$$G_\alpha = \{P \subset I \,|\, freq(P) < \alpha \wedge \forall P' \subset P : freq(P') > freq(P)\}$$

USER'S CONSTRAINT TAXONOMY

[CP18]

- **User's constraints on patterns**
  - frequent, closed, maximal, size, price…
- **User's constraints on items**
- **User's constraints on transactions**

USER'S CONSTRAINT TAXONOMY

[CP18]

- User's constraints on patterns
  - frequent, closed, maximal, size, price…
- User's constraints on items
- User's constraints on transactions

	food	electronics	cleaning	…
M				
T				
W				
Th				
F				
S				
Su				

USER'S CONSTRAINT TAXONOMY

[CP18]

- User's constraints on patterns
  - frequent, closed, maximal, size, price…

What

- User's constraints on items
- User's constraints on transactions

	food	electronics	cleaning	…
M				
T				
W				
Th				
F				
S				
Su				

USER'S CONSTRAINT TAXONOMY

[CP18]

- User's constraints on patterns
  - frequent, closed, maximal, size, price…

What

- User's constraints on items
- User's constraints on transactions

Where

	food	electronics	cleaning	…
M				
T				
W				
Th				
F				
S				
Su				

USER'S CONSTRAINT TAXONOMY

[CP18]

- User's constraints on patterns
  - frequent, closed, maximal, size, price…

What

- User's constraints on items
- User's constraints on transactions

Where

	food	electronics	cleaning	…
M				
T				
W				
Th				
F	sub-dataset			
S				
Su		Patterns		

A GENERAL CP MODEL FOR ITEMSET MINING

[DE READT ET AL, KDD08]

	food	electroni	cleaning	…
M				
T				
W				
Th				
F				
S				
Su				

A GENERAL CP MODEL FOR ITEMSET MINING

[DE READT ET AL, KDD08]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

P

T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

A GENERAL CP MODEL FOR ITEMSET MINING

[DE READT ET AL, KDD08]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

What

P

T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

A GENERAL CP MODEL FOR ITEMSET MINING

[DE READT ET AL, KDD08]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

What

P

T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

➤ **Reified CP Model (CP4IM)** [De readt et al, KDD08]**:**

Coverage constraints: $$\forall t \in T : (T_t = 1) \leftrightarrow \sum_{i \in I} P_i(1 - D_{ti}) = 0$$

Frequent constraints: $$\forall i \in I : (P_i = 1) \rightarrow \sum_{t \in T} T_t D_{ti} \geq \alpha$$

Closeness constraints: $$\forall i \in I : (P_i = 1) \leftrightarrow \sum_{t \in T} T_t(1 - D_{ti}) = 0$$

EXAMPLE (FREQUENT ITEMSET)

EXAMPLE (FREQUENT ITEMSET)

t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)

T	P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
0/1	t1:		B	C		E	F	G	H
0/1	t2:	A			D			G	
0/1	t3:	A		C	D				H
0/1	t4:	A				E	F		
0/1	t5:		B			E	F		
0/1	t6:		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)

T	P	0	1	0	0	1	0	0	0
1	t1:		B	C		E	F	G	H
0	t2:	A			D			G	
0	t3:	A		C	D				H
0	t4:	A				E	F		
1	t5:		B			E	F		
1	t6:		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)

T	P	0	1	0	0	1	0	0	0
1	t1:		B	C		E	F	G	H
0	t2:	A			D			G	
0	t3:	A		C	D				H
0	t4:	A				E	F		
1	t5:		B			E	F		
1	t6:		B			E	F	G	

$$cover(BE) = \{t_1, t_5, t_6\}$$

$$freq(BE) = 50\,\%$$

A GENERAL CP MODEL FOR ITEMSET MINING

[CP18]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

What

P

T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

A GENERAL CP MODEL FOR ITEMSET MINING

[CP18]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

What

$$H = \langle H_1, \ldots, H_n \rangle$$
$$V = \langle V_1, \ldots, V_m \rangle$$

H
P
V T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

A GENERAL CP MODEL FOR ITEMSET MINING

[CP18]

Boolean variables:

$P = \langle P_1, \ldots, P_n \rangle$

$T = \langle T_1, \ldots, T_m \rangle$

What

$H = \langle H_1, \ldots, H_n \rangle$

$V = \langle V_1, \ldots, V_m \rangle$

Where

H

P

V T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

EXAMPLE (FREQUENT ITEMSET)

T	P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
0/1	t1:		B	C		E	F	G	H
0/1	t2:	A			D			G	
0/1	t3:	A		C	D				H
0/1	t4:	A				E	F		
0/1	t5:		B			E	F		
0/1	t6:		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)

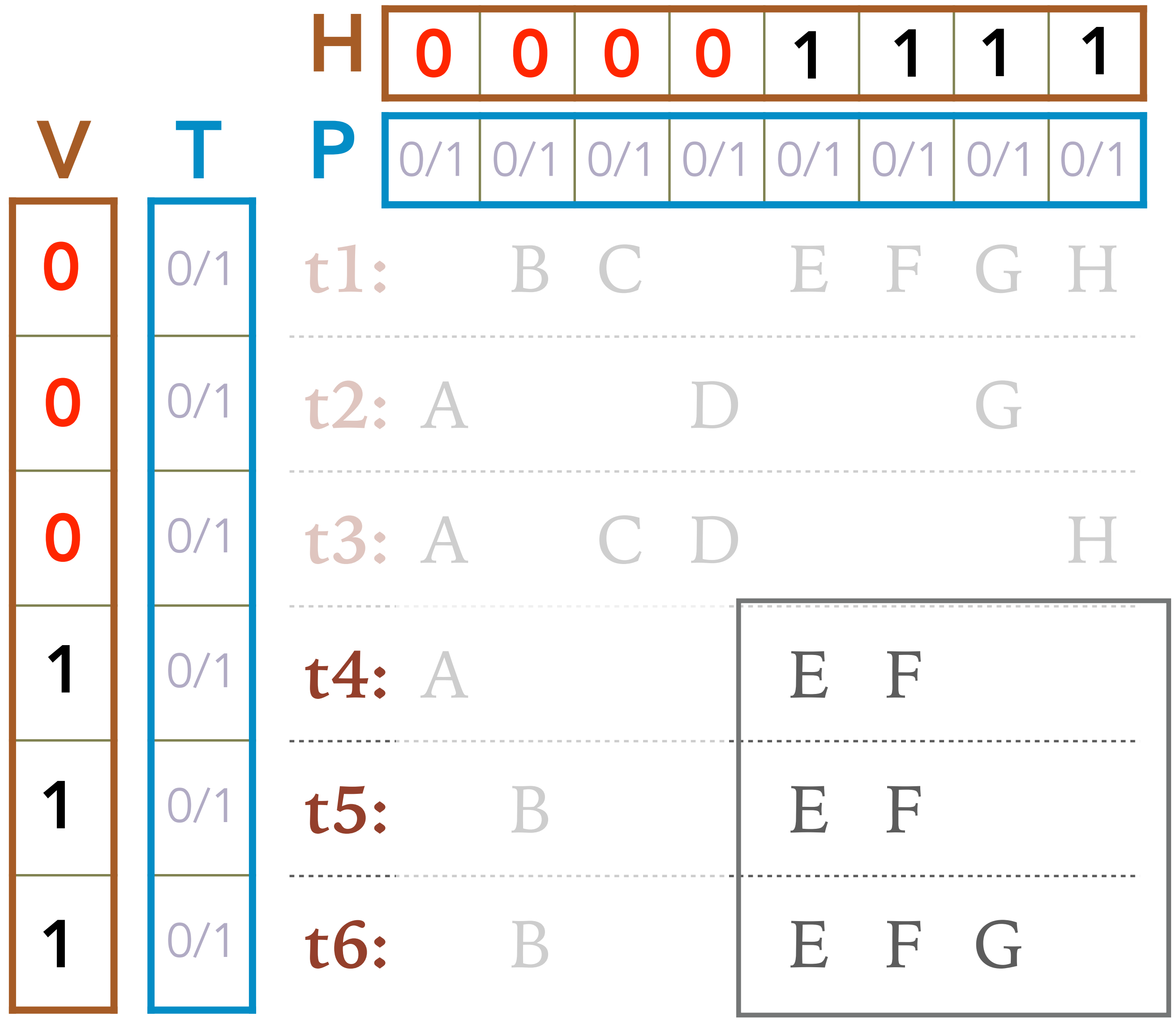
V	T									
		H	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
		P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
0/1	0/1	**t1:**		B	C		E	F	G	H
0/1	0/1	**t2:**	A			D			G	
0/1	0/1	**t3:**	A		C	D				H
0/1	0/1	**t4:**	A				E	F		
0/1	0/1	**t5:**		B			E	F		
0/1	0/1	**t6:**		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)

H	0	0	0	0	1	1	1	1
P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1

V	T									
0	0/1	t1:		B	C		E	F	G	H
0	0/1	t2:	A			D			G	
0	0/1	t3:	A		C	D				H
1	0/1	t4:	A				E	F		
1	0/1	t5:		B			E	F		
1	0/1	t6:		B			E	F	G	

EXAMPLE (FREQUENT ITEMSET)



A GENERAL CP MODEL FOR ITEMSET MINING

[CP18]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

What

$$H = \langle H_1, \ldots, H_n \rangle$$
$$V = \langle V_1, \ldots, V_m \rangle$$

Where

H

P

V T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

A GENERAL CP MODEL FOR ITEMSET MINING

[CP18]

Boolean variables:

$P = \langle P_1, \ldots, P_n \rangle$

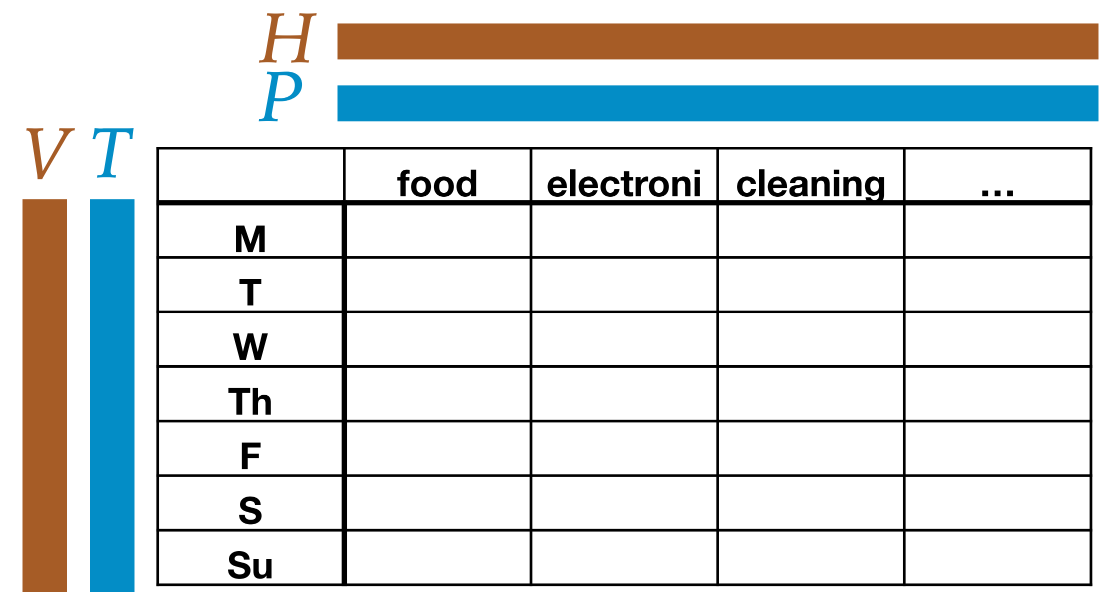
$T = \langle T_1, \ldots, T_m \rangle$

What

$H = \langle H_1, \ldots, H_n \rangle$

$V = \langle V_1, \ldots, V_m \rangle$

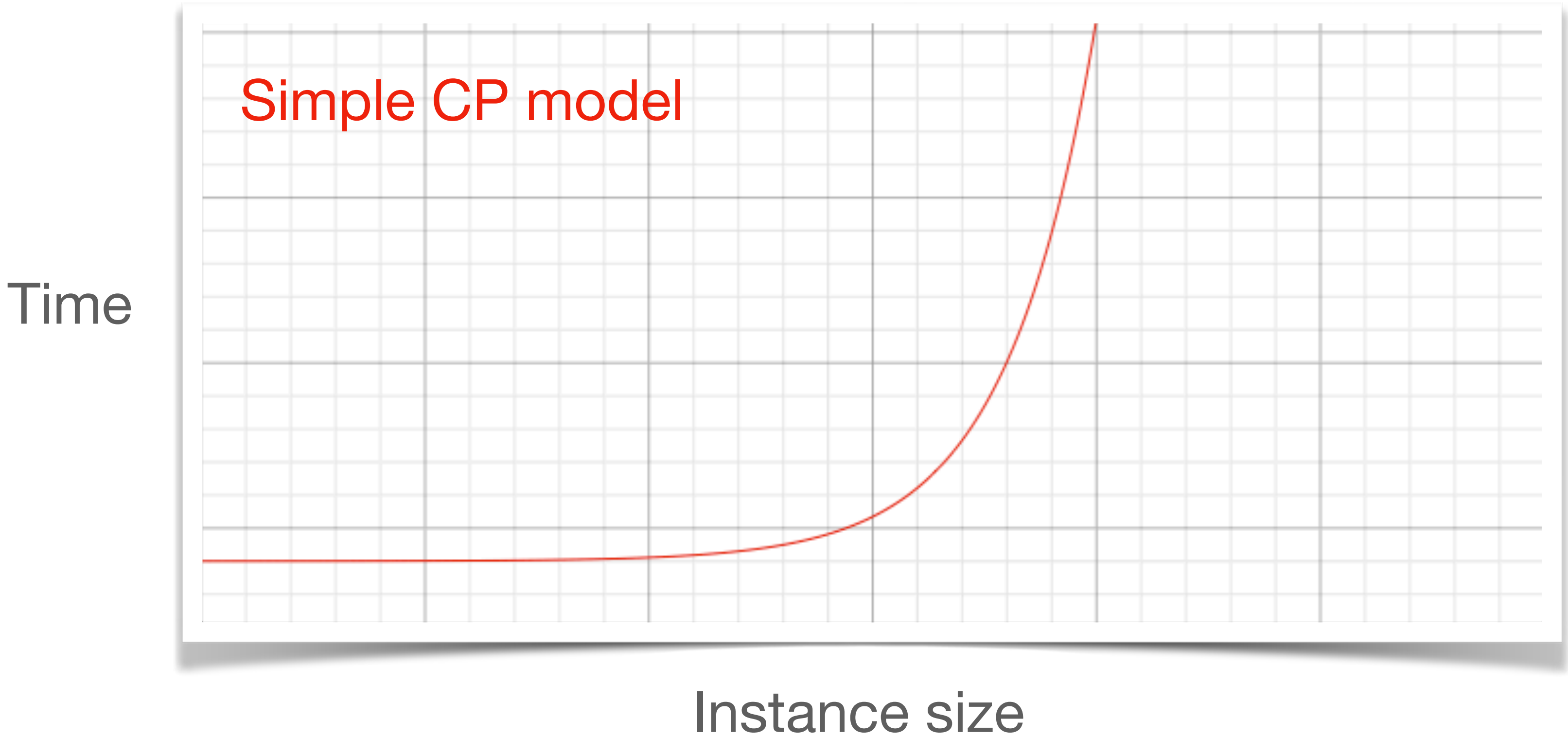
Where



	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

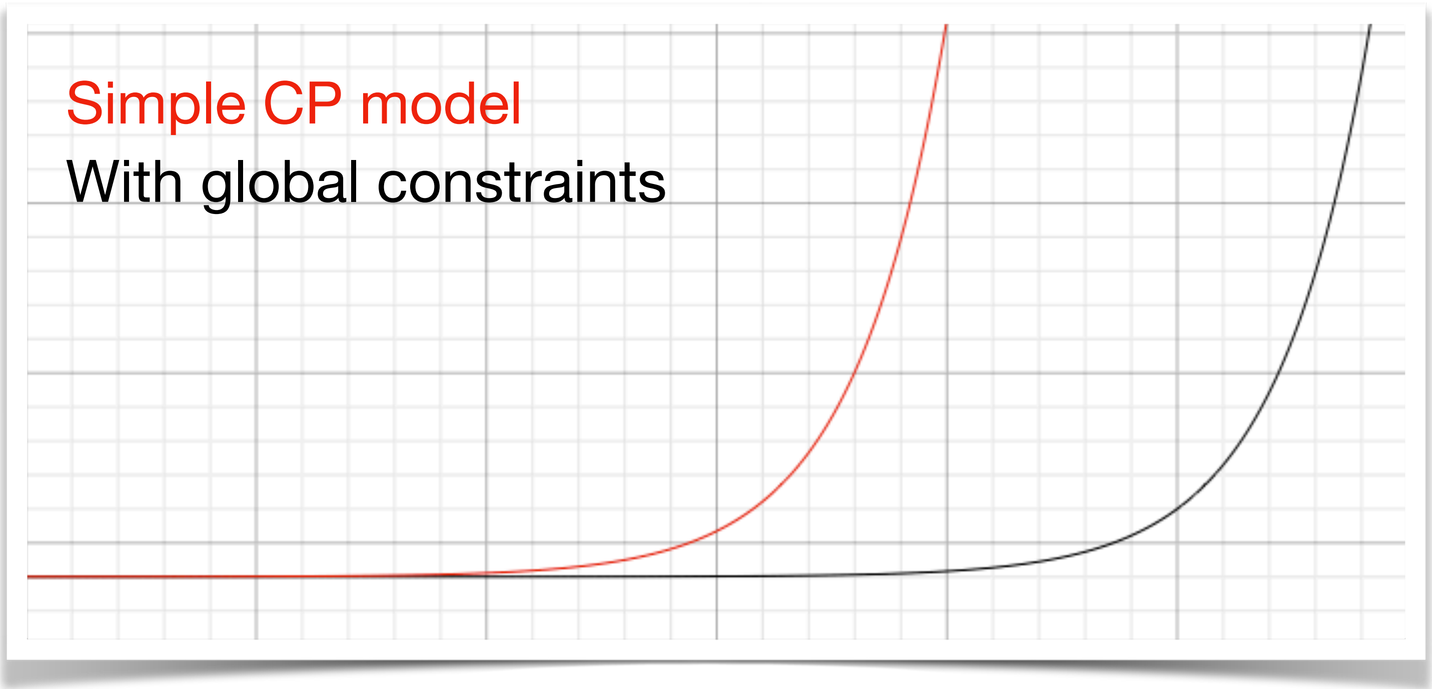
- **Large Number of Auxiliary Variables**
- **High number of reified constraints with significant arity**

GLOBAL CONSTRAINTS



GLOBAL CONSTRAINTS

Time



Instance size

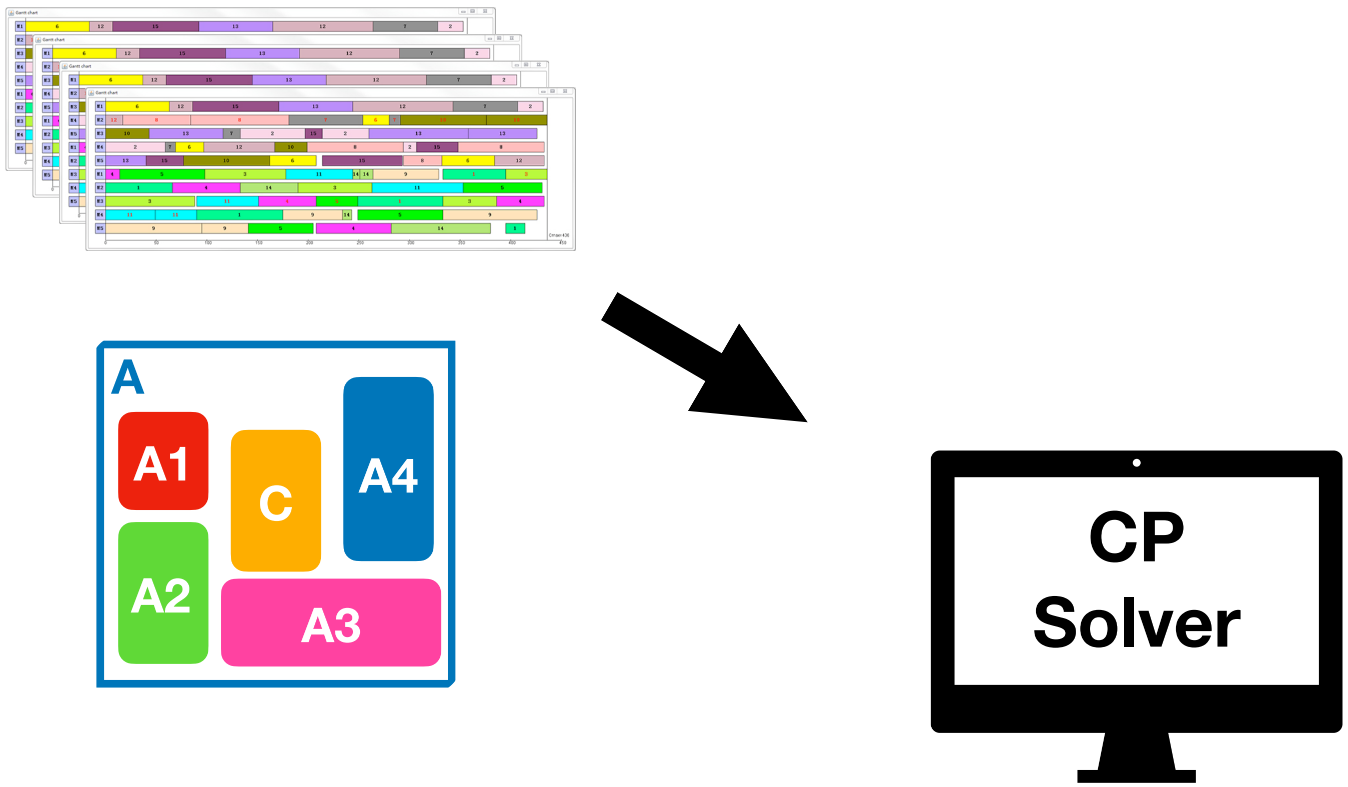
GLOBAL CONSTRAINTS

[GLOBAL CONSTRAINT CATALOG]

GLOBAL CONSTRAINTS

[GLOBAL CONSTRAINT CATALOG]

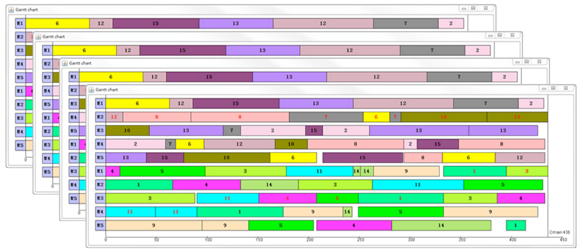
Scheduling instances

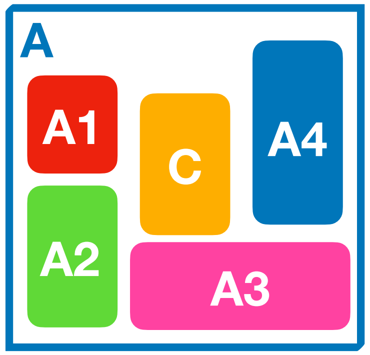


GLOBAL CONSTRAINTS

[GLOBAL CONSTRAINT CATALOG]

Scheduling instances

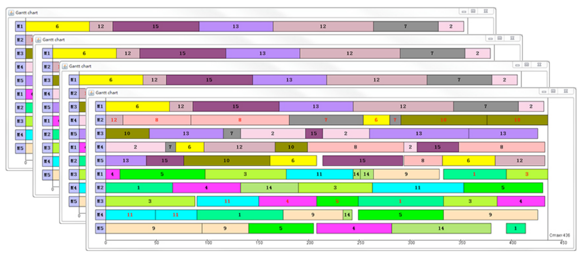


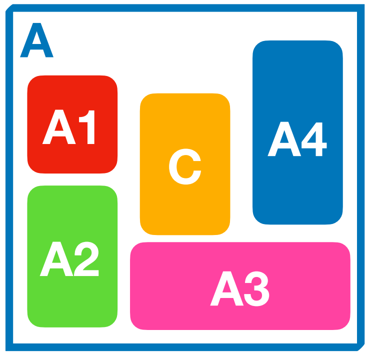


Cumulative

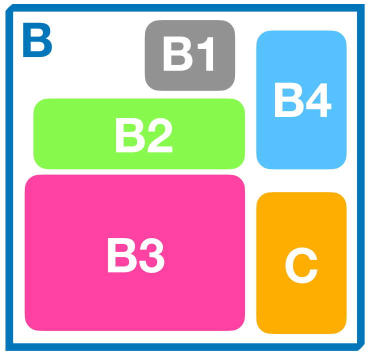
GLOBAL CONSTRAINTS

Scheduling instances





Vehicule routing instances



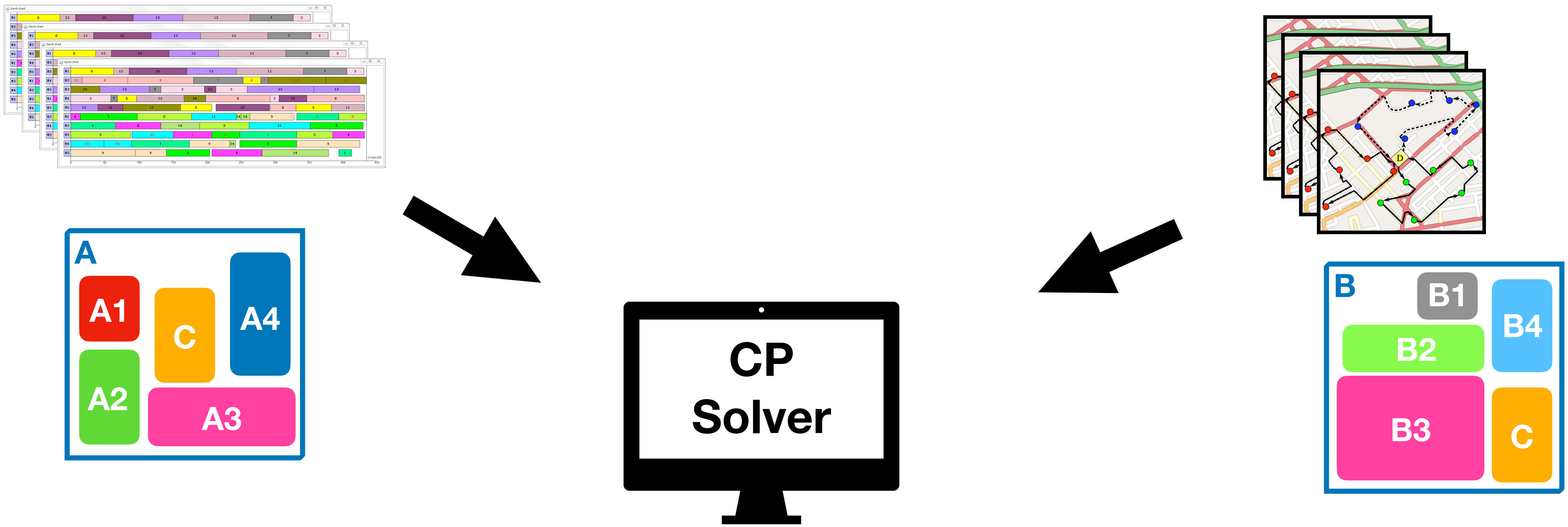
`Cumulative`

GLOBAL CONSTRAINTS

[GLOBAL CONSTRAINT CATALOG]

Scheduling instances

Vehicule routing instances



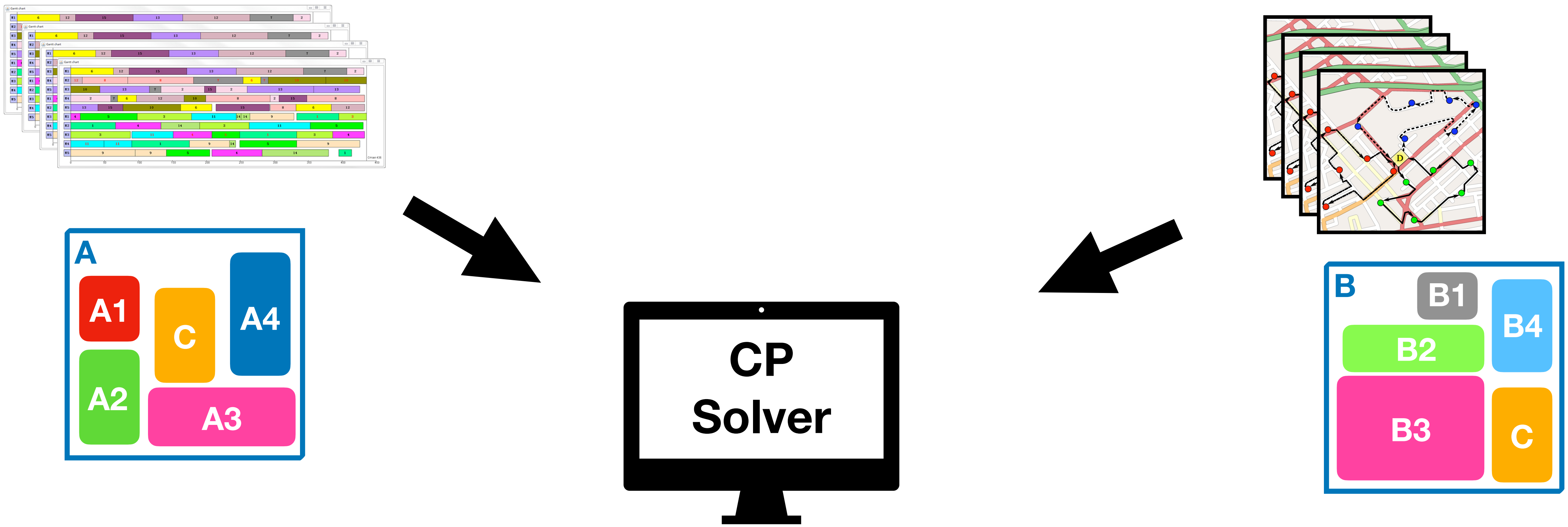
Cumulative
Alldifferent

GLOBAL CONSTRAINTS

[GLOBAL CONSTRAINT CATALOG]

Scheduling instances

Vehicule routing instances



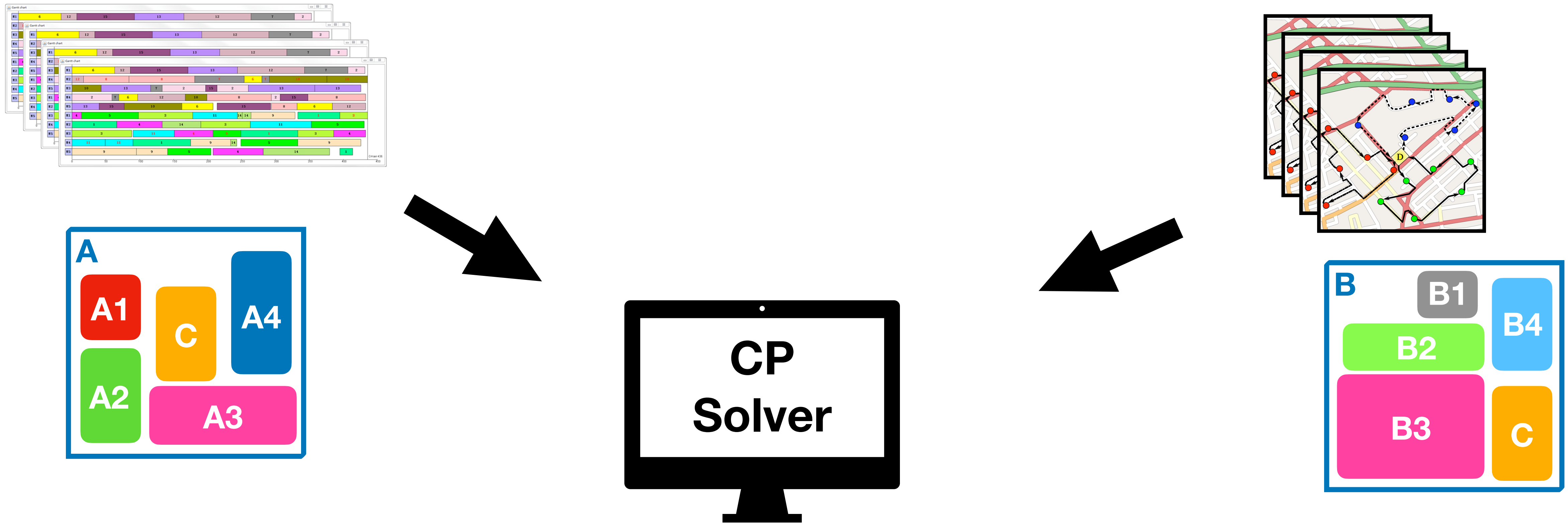
Cumulative
Alldifferent

GLOBAL CONSTRAINTS

[GLOBAL CONSTRAINT CATALOG]

Scheduling instances

Vehicule routing instances

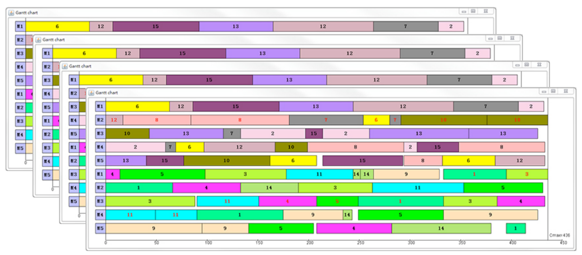


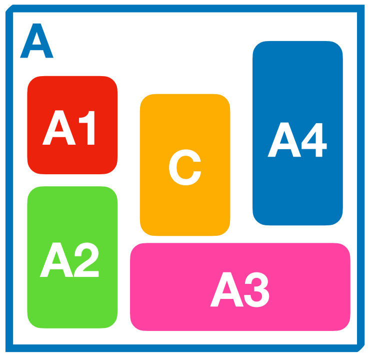

```
Cumulative
Alldifferent
Sum
knapsack
Element
GlobalCardinality
Regular
…
```

GLOBAL CONSTRAINTS

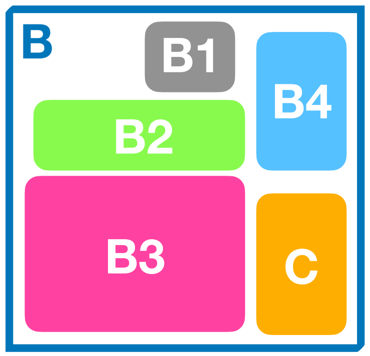
[GLOBAL CONSTRAINT CATALOG]

Scheduling instances





Vehicule routing instances

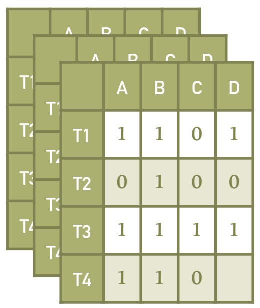


toolbox

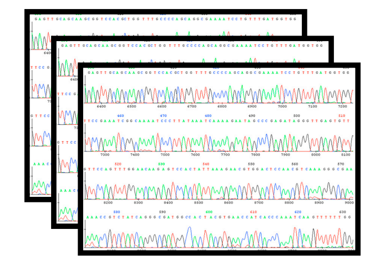
```
Cumulative
Alldifferent
Sum
knapsack
Element
GlobalCardinality
Regular
…
```

GLOBAL CONSTRAINTS IN PATTERN MINING

Itemset Mining Tasks



Association rules Mining Tasks

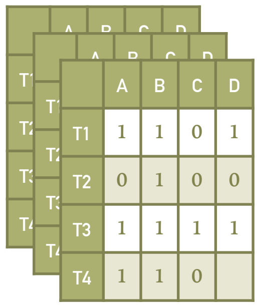


toolbox

```
Cumulative
Alldifferent
Sum
knapsack
Element
GlobalCardinality
Regular
…
```

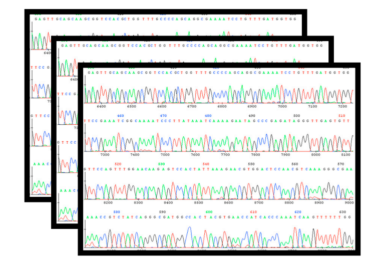
GLOBAL CONSTRAINTS IN PATTERN MINING

Itemset Mining Tasks



	A	B	C	D
T1	1	1	0	1
T2	0	1	0	0
T3	1	1	1	1
T4	1	1	0	

Association rules Mining Tasks



•
•
•

toolbox

```
Cumulative
Alldifferent
Sum
knapsack
Element
GlobalCardinality
Regular
…
```

DM toolbox

```
ClosedPattern
CoverSize
Generator
FrequentSubs
InfrequentSupers
Confident
…
```

Efficiency:

“Fast Solutions with Efficient Propagators”

A GENERAL CP MODEL FOR ITEMSET MINING

[DE READT ET AL, KDD08]

Boolean variables:

$$P = \langle P_1, \ldots, P_n \rangle$$
$$T = \langle T_1, \ldots, T_m \rangle$$

What

P

T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

➤ Reified CP Model (CP4IM):

Coverage constraints: $$\forall t \in T : (T_t = 1) \leftrightarrow \sum_{i \in I} P_i(1 - D_{ti}) = 0$$

Frequent constraints: $$\forall i \in I : (P_i = 1) \rightarrow \sum_{t \in T} T_t D_{ti} \geq \alpha$$

Closeness constraints: $$\forall i \in I : (P_i = 1) \leftrightarrow \sum_{t \in T} T_t(1 - D_{ti}) = 0$$

Boolean variables:

$P = \langle P_1, \ldots, P_n \rangle$

$T = \langle T_1, \ldots, T_m \rangle$ What

P

T

	food	electroni	cleaning	...
M				
T				
W				
Th				
F				
S				
Su				

➤ **Reified CP Model (CP4IM):**

Coverage constraints: $\forall t \in T : (T_t = 1) \leftrightarrow \sum P_i(1 - D_{ti}) = 0$

Frequ

- m **auxiliary variables**
- $(2n + m)$ **reified constraints of arity** $(n+1)$ **and** $(m+1)$

Closeness constraints: $\forall i \in I : (P_i = 1) \leftrightarrow \sum_{t \in T} T_t(1 - D_{ti}) = 0$

CLOSEDPATTERN CONSTRAINT

[CP16]

➤ **Definition 1:**

Let P be a vector of boolean variables, D be a dataset and θ a minimum support. Given a complete σ assignment on P, $ClosedPattern_{D,\alpha}(\sigma)$ holds iff $freq(\sigma^+) \geq \alpha$ and σ^+ is closed.

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

> ➤ **Definition 1:**
>
> Let P be a vector of boolean variables, D be a dataset and θ a minimum support. Given a complete σ assignment on P, $ClosedPattern_{D,\alpha}(\sigma)$ holds iff $freq(\sigma^+) \geq \alpha$ and σ^+ is closed.

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

- One boolean vector: P of items (n vars)
- ClosedPattern propagator enforces domain consistency in time $O(n^2m)$ with a space complexity of $O(nm)$
- CPU time speed-up factor of ranged between 10 and 200 comparing to reified model
- Dealing with large datasets

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 3$

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

P	1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

$\alpha = 3$

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

P	1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

$\alpha = 3$

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

P	1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

$\alpha = 3$

Rule 1:
If A is present, adding any other item can only decrease the frequency below α

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

P	1	0	0	0	0	0	0	0
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

$\alpha = 3$

Rule 1:
If A is present, adding any other item can only decrease the frequency below α

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 3$

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 3$

	A	B	C	D	E	F	G	H
P	0/1	0/1	0/1	0/1	1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

P	0/1	0/1	0/1	0/1	1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

$\alpha = 3$

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

P	0/1	0/1	0/1	0/1	1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

$\alpha = 3$

Rule 2:
If E is present, then F must also be present

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\alpha \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 3$

P	0/1	0/1	0/1	0/1	1	1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

Rule 2:
If E is present, then F must also be present

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\theta \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 2$

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:					E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\theta \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 2$

P	0/1	**0**	0/1	0/1	**1**	**1**	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:					E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\theta \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 2$

P	0/1	0	0/1	0/1	1	1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:					E	F		
t6:		B			E	F	G	

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\theta \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 2$

P	0/1	0	0/1	0/1	1	1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:					E	F		
t6:		B			E	F	G	

Rule 3:
If B is absent, then G cannot be present

CLOSEDPATTERN CONSTRAINT

[CP16]

- **Given a partial assignment, we have three filtering rules:**
  - **Rule 1:** if item i decreases frequency under $\theta \Rightarrow$ remove 1 from $dom(P_j)$
  - **Rule 2:** if item i is an extension $\Rightarrow$ remove 0 from $dom(P_j)$
  - **Rule 3:** if item i is always with an absent item $\Rightarrow$ remove 1 from $dom(P_j)$

$\alpha = 2$

P	0/1	0	0/1	0/1	1	1	0	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:					E	F		
t6:		B			E	F	G	

Rule 3:
If B is absent, then G cannot be present

CLOSEDPATTERN CONSTRAINT (SOME RESULTS)

[CP16]

$\mathcal{D}$	θ (%)	$\#C$ ($\approx$)	Time (s) ClosedPattern	CP4IM	lcm
chess	50	10^5	3.21	10.97	**.32**
	40	10^6	12.27	40.85	**.44**
	30	10^6	45.92	136.31	**.07**
	20	10^7	187.89	467.52	**7.55**
	10	10^8	969.40	1 950.51	**41.55**
splice1	20	10^2	0.59	22.59	**.04**
	10	10^3	0.14	25.54	**.07**
	5	10^4	3.23	138.54	**.46**
	1	10^7	400.10	1 652.41	**3.59**
connect	90	10^3	0.92	7.10	**.22**
	80	10^4	1.65	16.57	**.31**
	70	10^4	4.09	33.72	**.40**
	60	10^5	7.30	45.73	**.39**
	50	10^5	14.53	110.19	**.52**
	40	10^5	27.32	153.39	**.83**
	30	10^5	49.97	304.52	**.37**
	20	10^6	157.40	712.68	**.37**
	10	10^7	760.71	2 597.89	**7.70**
T40*	10	10^2	1.13	OOM	**.43**
	5	10^2	1.78	OOM	**.31**
	1	10^5	25.78	OOM	**1.32**
	0.5	10^6	953.58	OOM	**3.31**
retail	10	10	2.55	OOM	**.06**
	1	10^2	4.02	OOM	**.10**
	0.5	10^3	12.73	OOM	**.32**
	0.1	10^4	796.82	OOM	**.80**
	0.05	10^4	2 645.06	OOM	**.07**

PATTERN MINING GLOBAL CONSTRAINTS

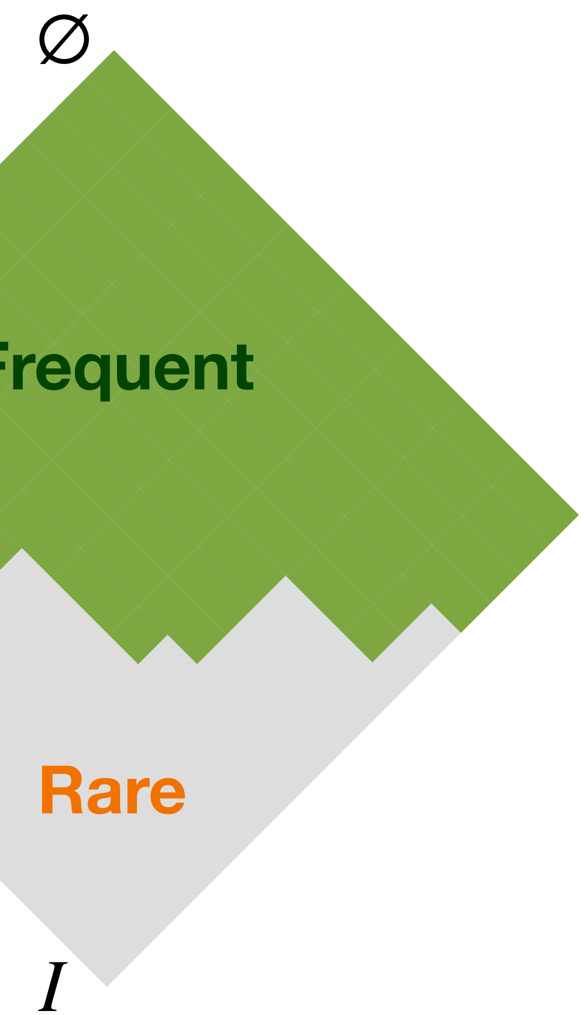
Constraint	Pattern	DC	Ref
Prefix-Projection	Sequence Mining	(Yes)	[Kemmar et al., CP15]
ClosedPattern	Closed Itemsets	Yes	[Lazaar et al., CP16]
Coversize	Closed Itemsets	Yes	[Schaus et al., CP17]
Generator	Generator Itemsets	Yes	[Belaid et al., SDM19]
Confident	Rules Confidence	No	[Belaid et al., SDM19]
FrequentSubs	Itemset Patterns	Yes	[Belaid et al., IJCAI19]
InfrequentSupers	Itemset Patterns	Yes	[Belaid et al., IJCAI19]
FreqRare	Multiple Item Supports	(Yes)	[Belaid&Lazaar, ICTAI21]
GC4CIP	Closed Interval Patterns	No	[Bekkoucha et al., CPAIOR24]

Declarativity:

"High-Level Modeling in Constraint-Based Data Mining"

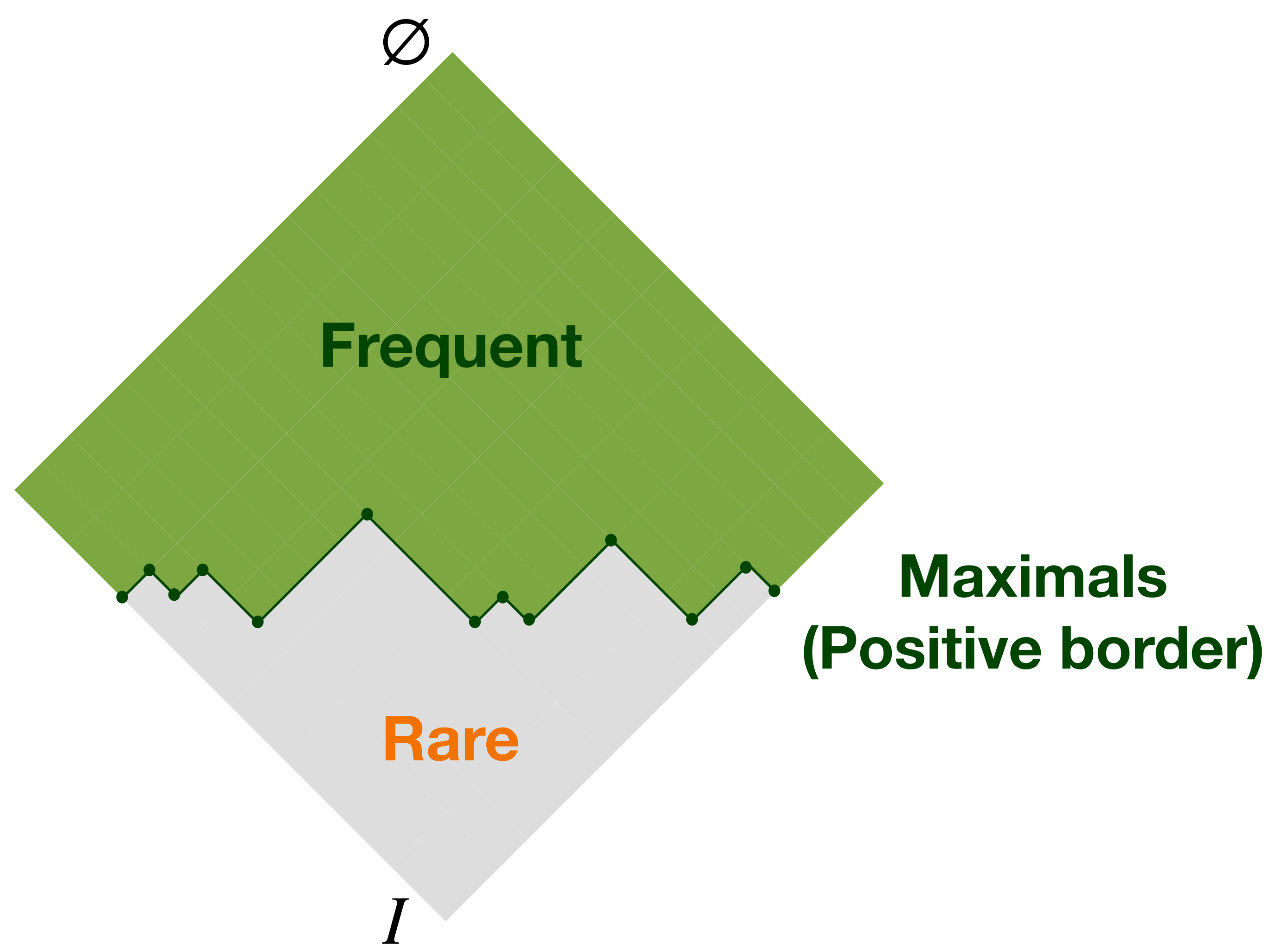
MINING MINIMALS AND/OR MAXIMALS

[MANNILA_TOIVONEN, 97]



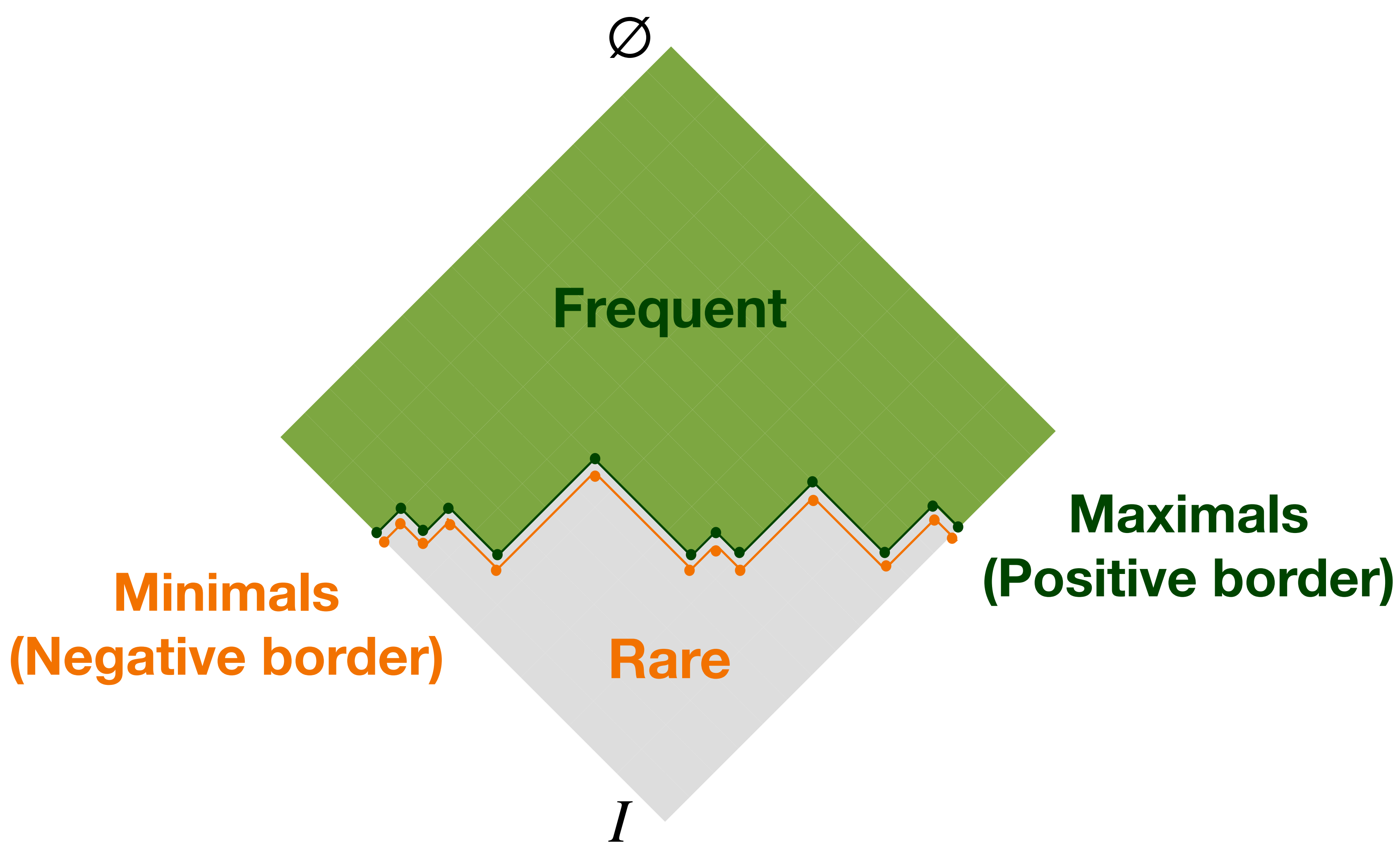
MINING MINIMALS AND/OR MAXIMALS

[MANNILA_TOIVONEN, 97]

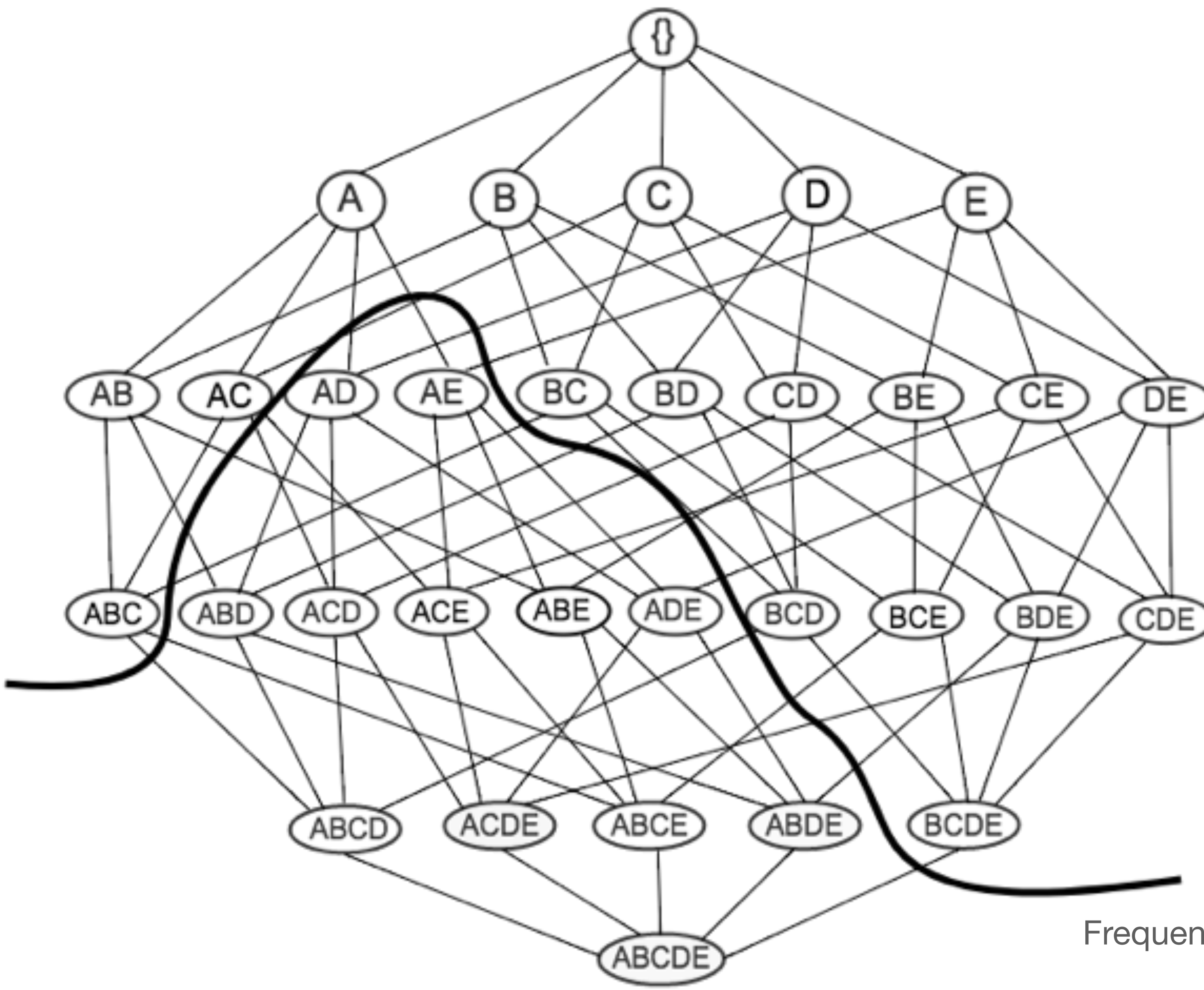


MINING MINIMALS AND/OR MAXIMALS

[MANNILA_TOIVONEN, 97]

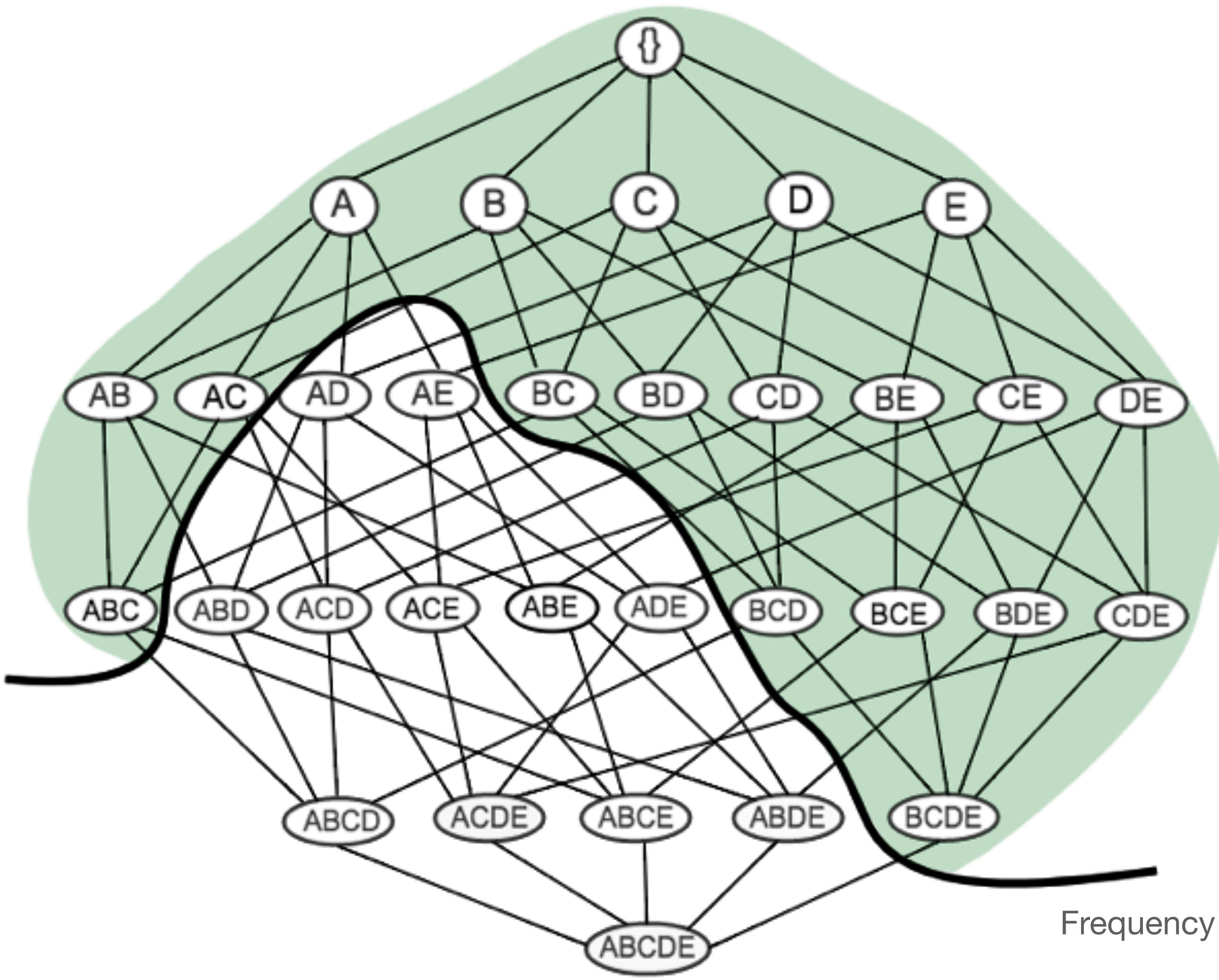


EXAMPLE



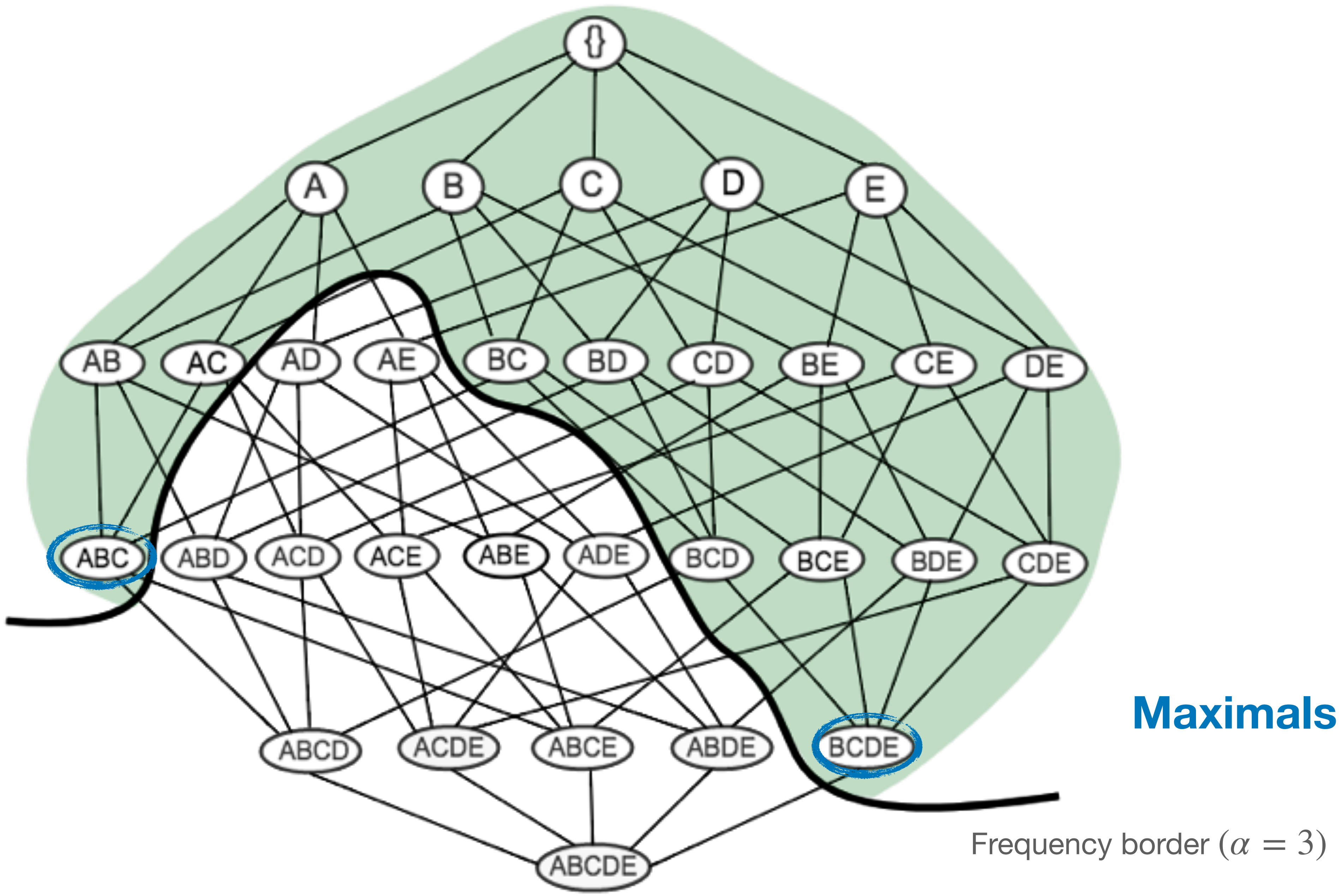
Frequency border ($\alpha = 3$)

EXAMPLE



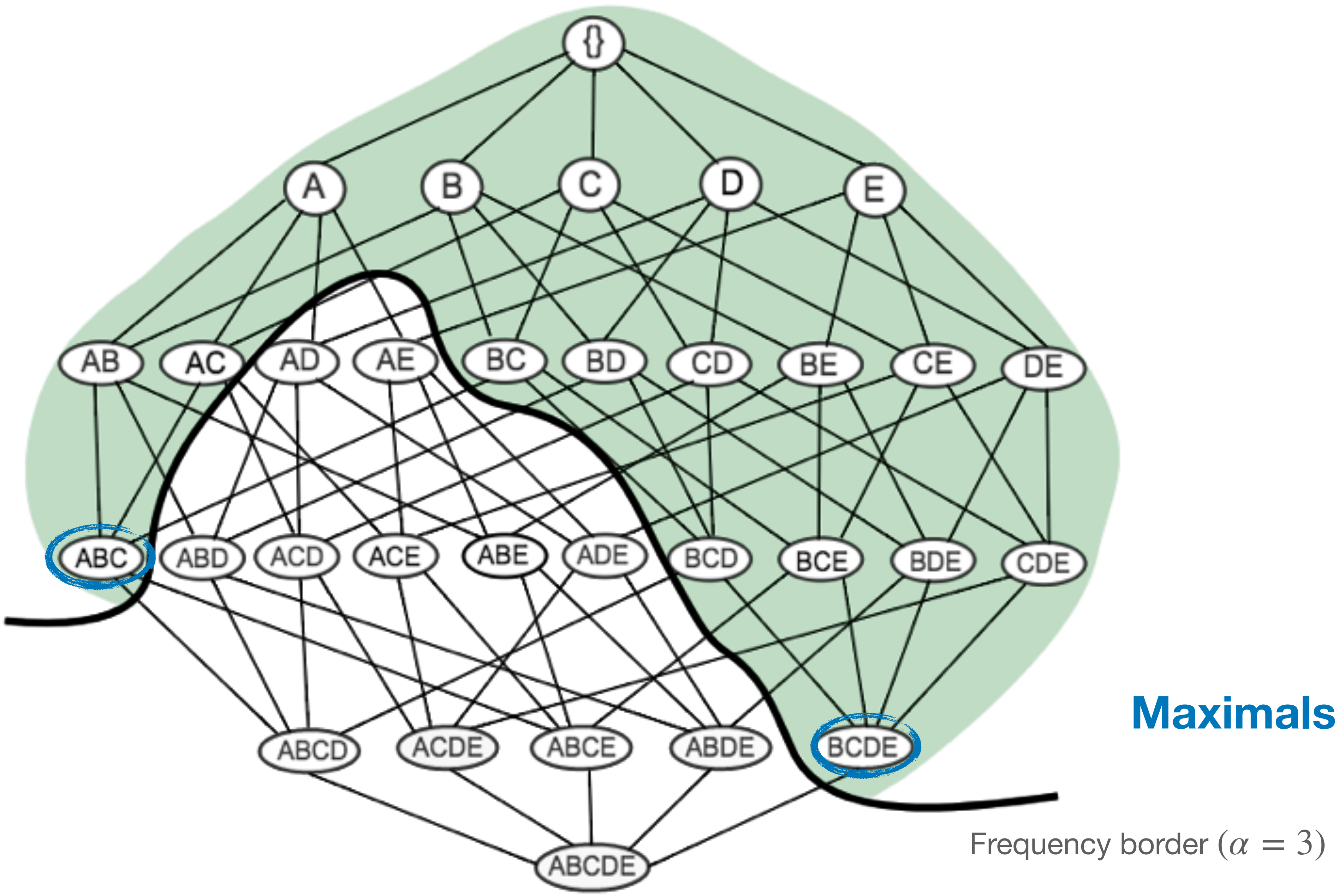
Frequency border ($\alpha = 3$)

EXAMPLE



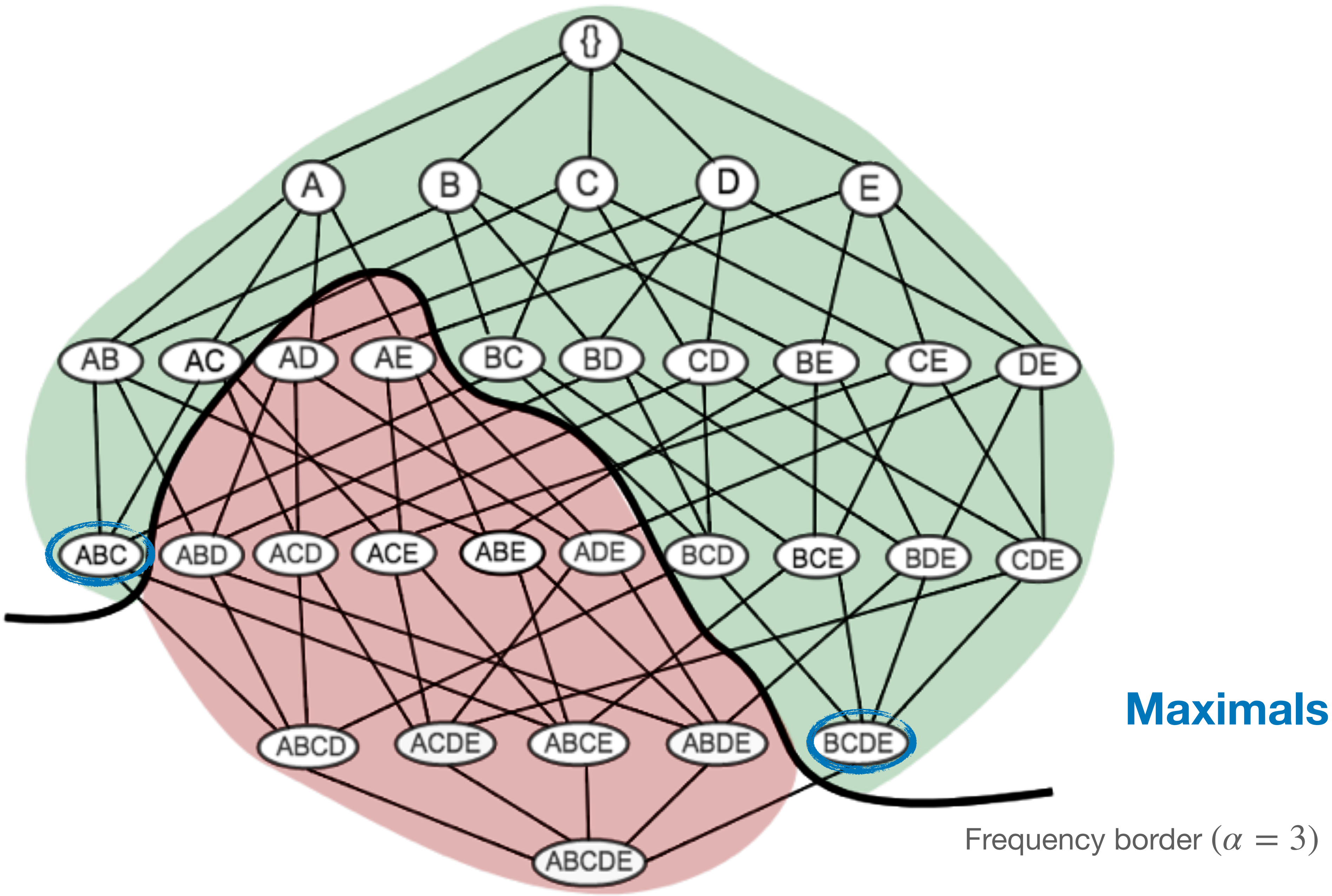
{}
A
B
C
D
E
AB
AC
AD
AE
BC
BD
CD
BE
CE
DE
ABC
ABD
ACD
ACE
ABE
ADE
BCD
BCE
BDE
CDE
ABCD
ACDE
ABCE
ABDE
BCDE
ABCDE
Maximals
Frequency border ($\alpha = 3$)

EXAMPLE



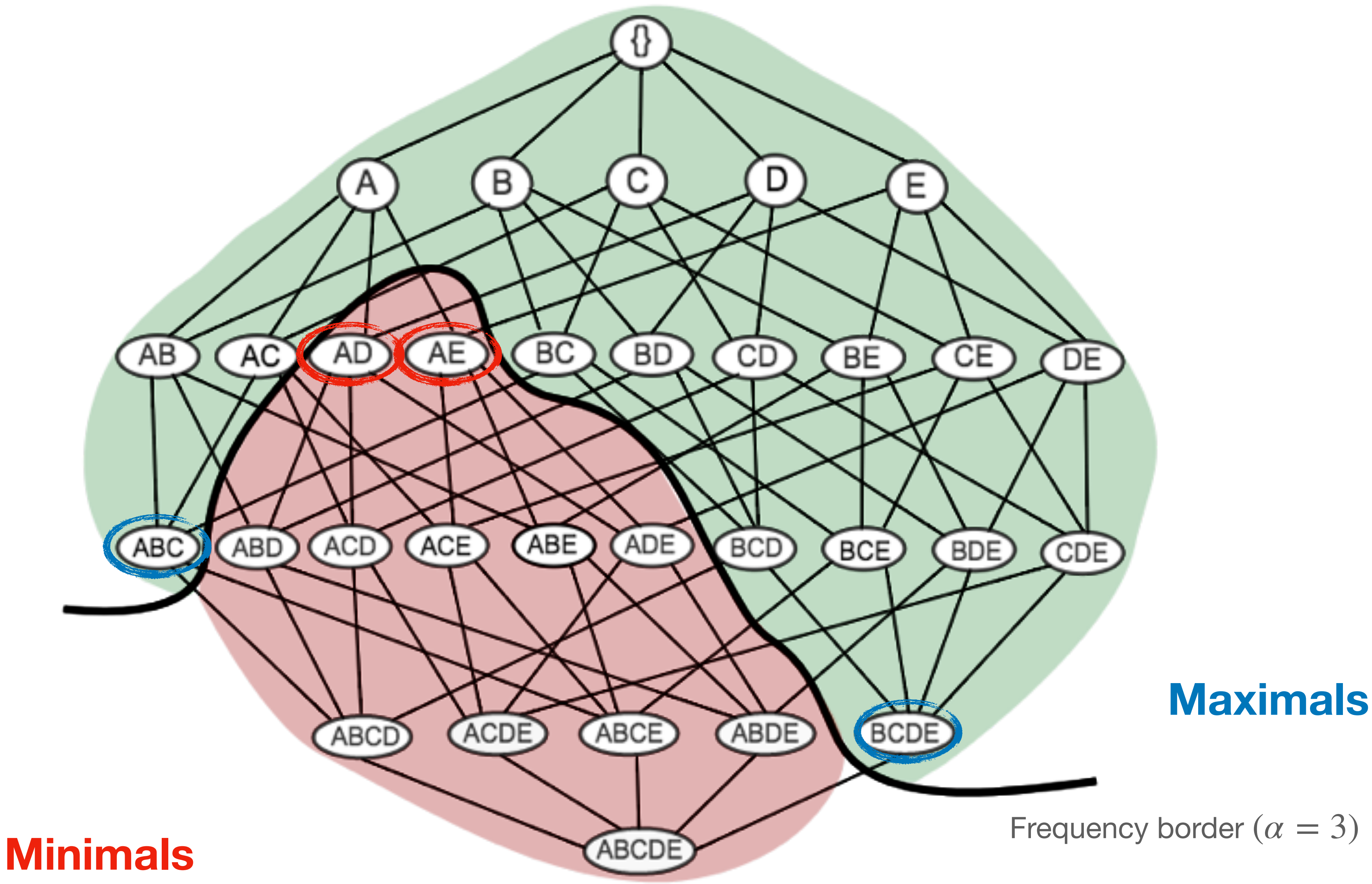
{}
A
B
C
D
E
AB
AC
AD
AE
BC
BD
CD
BE
CE
DE
ABC
ABD
ACD
ACE
ABE
ADE
BCD
BCE
BDE
CDE
ABCD
ACDE
ABCE
ABDE
BCDE
ABCDE
Maximals
Frequency border ($\alpha = 3$)

EXAMPLE



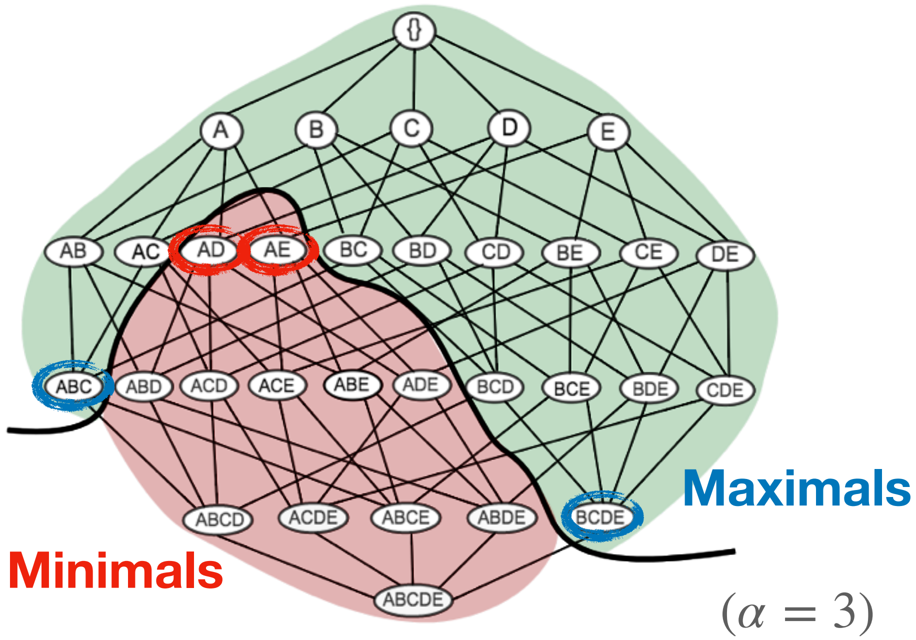
{}
A
B
C
D
E
AB
AC
AD
AE
BC
BD
CD
BE
CE
DE
ABC
ABD
ACD
ACE
ABE
ADE
BCD
BCE
BDE
CDE
ABCD
ACDE
ABCE
ABDE
BCDE
ABCDE
Maximals
Frequency border ($\alpha = 3$)

EXAMPLE



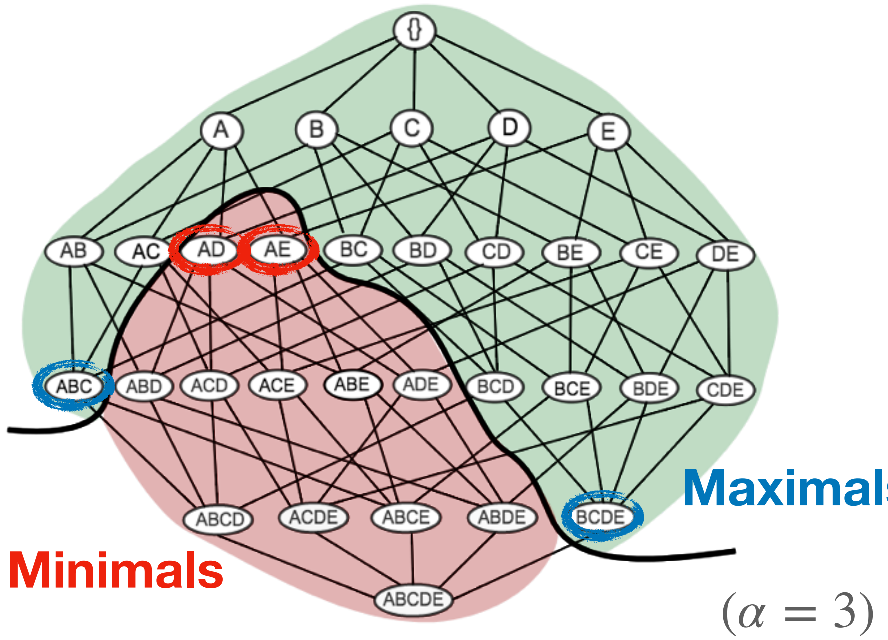
{}
A
B
C
D
E
AB
AC
AD
AE
BC
BD
CD
BE
CE
DE
ABC
ABD
ACD
ACE
ABE
ADE
BCD
BCE
BDE
CDE
ABCD
ACDE
ABCE
ABDE
BCDE
ABCDE
Maximals
Minimals
Frequency border ($\alpha = 3$)

MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]

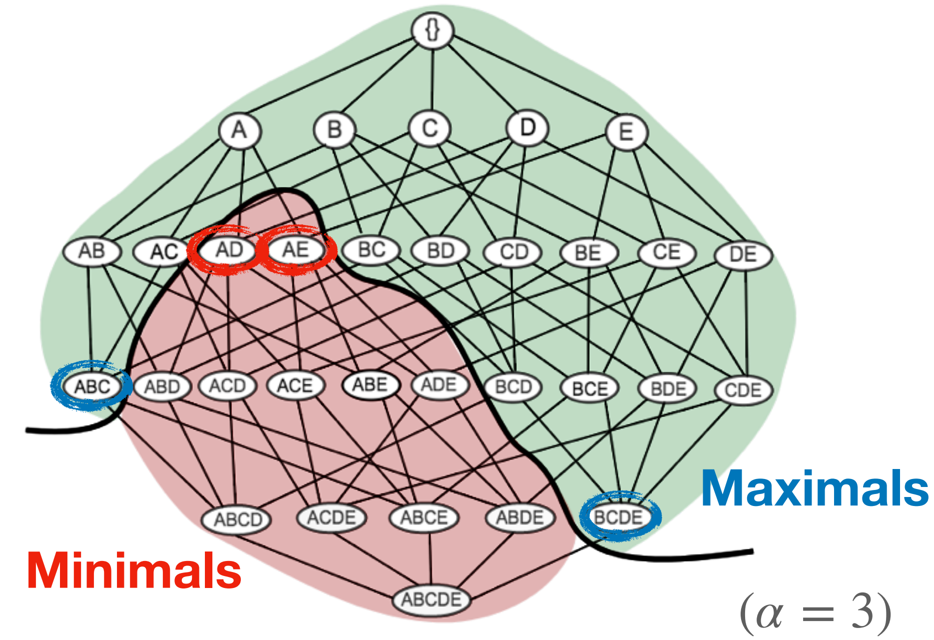


MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]

- A:[Frequent itemsets]



MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]

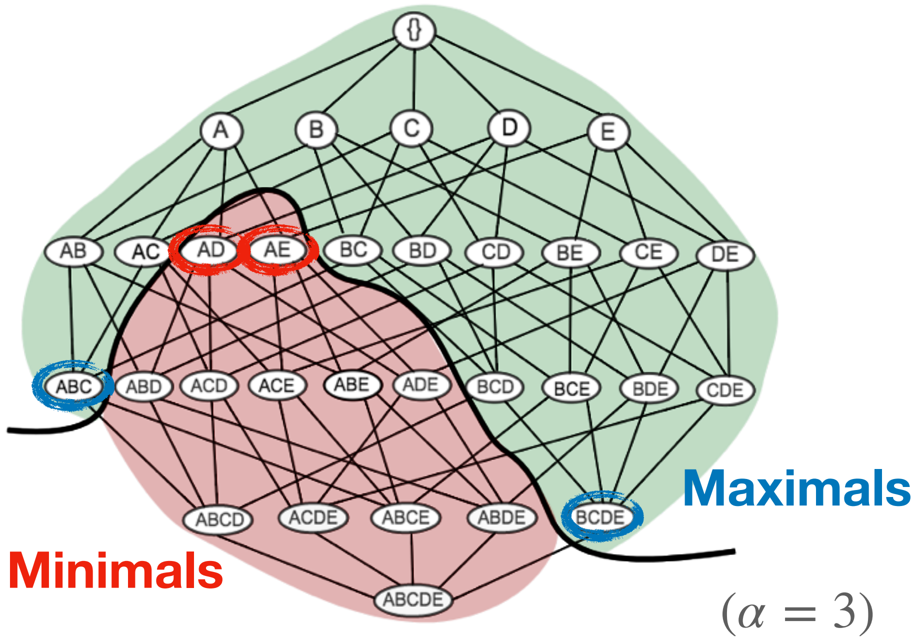


- A:[Frequent itemsets]

MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]

- A:[Frequent itemsets]

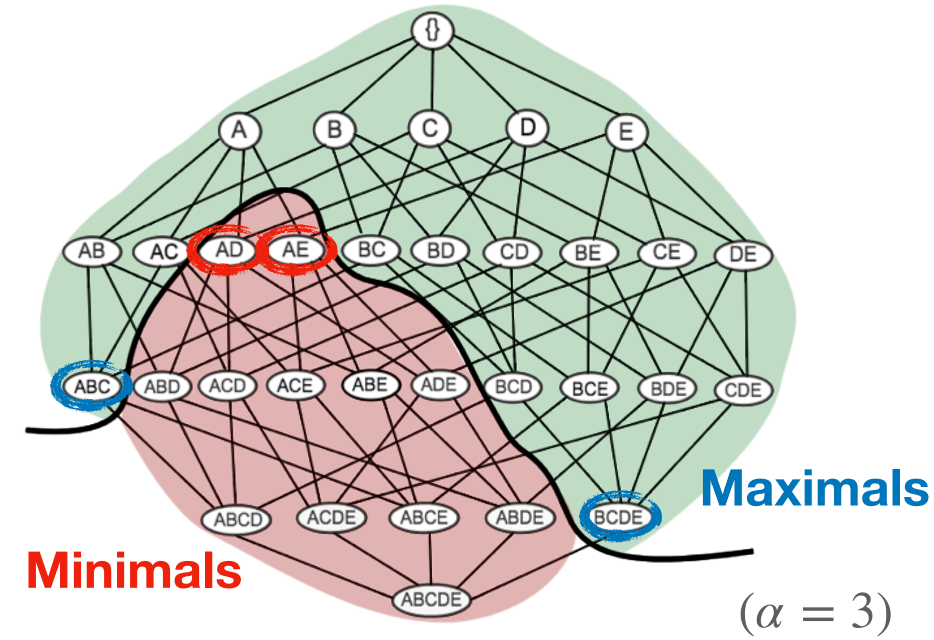
- $\neg A$:[Infrequent itemsets]



MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]

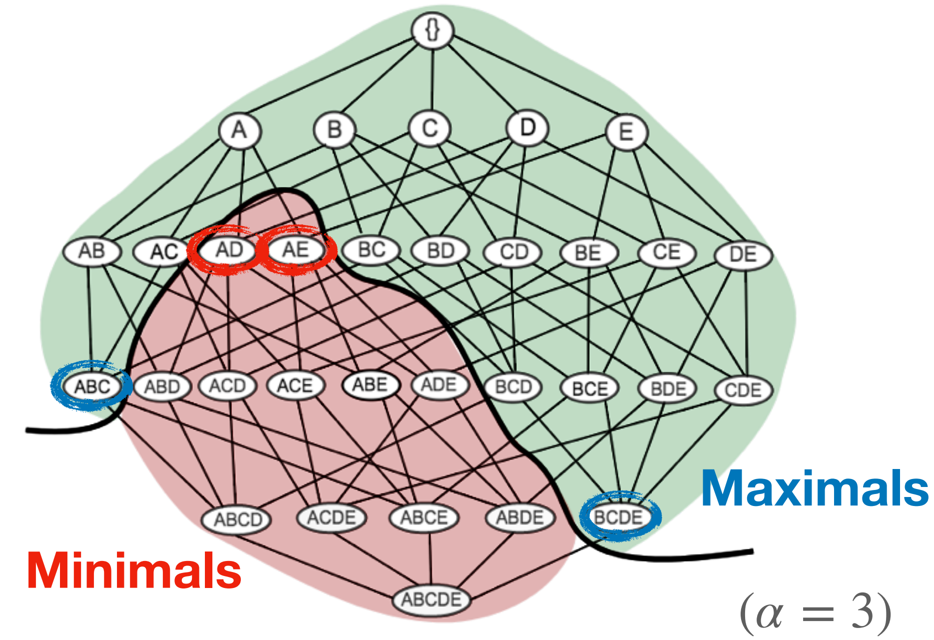
- A:[Frequent itemsets]

- $\neg A$:[Infrequent itemsets]



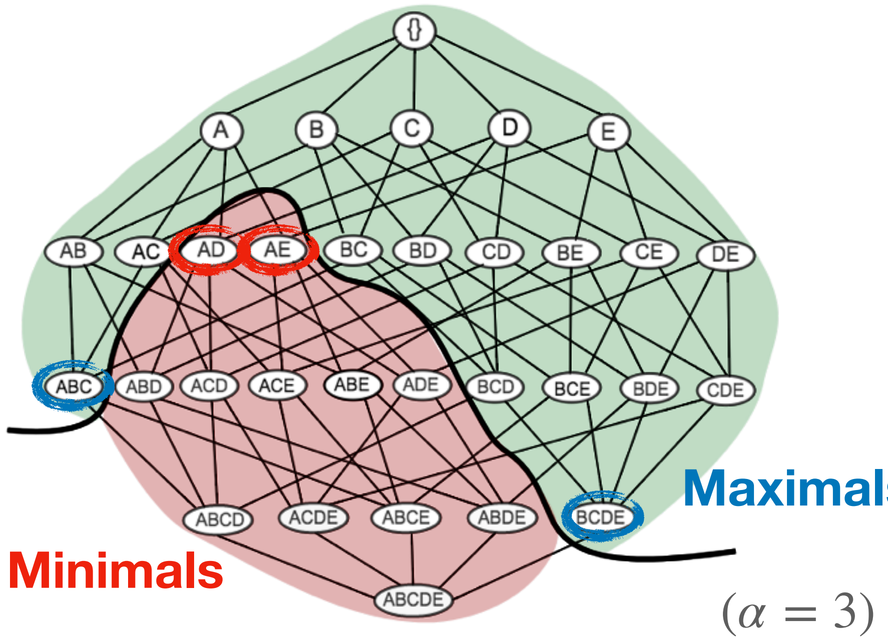
$(\alpha = 3)$

MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]



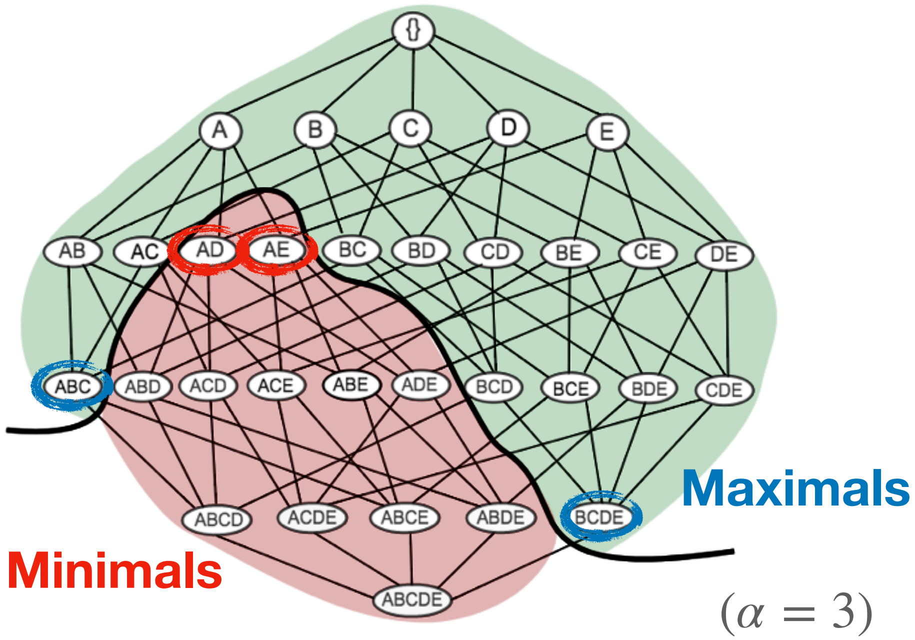
- A:[Frequent itemsets]
- $\neg A$:[Infrequent itemsets]
- **Borders (Maximals + Minimals):** patterns B:[with only frequent subsets] B and C:[with only infrequent supersets]

MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]



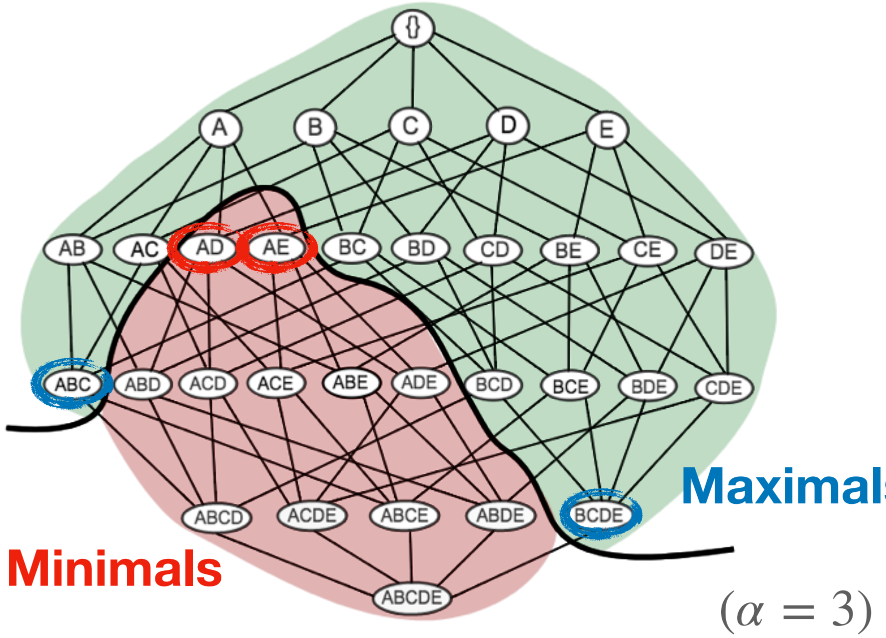
- A:[Frequent itemsets]
- $\neg A$:[Infrequent itemsets]
- **Borders (Maximals + Minimals):** patterns B:[with only frequent subsets] B and C:[with only infrequent supersets]

MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]



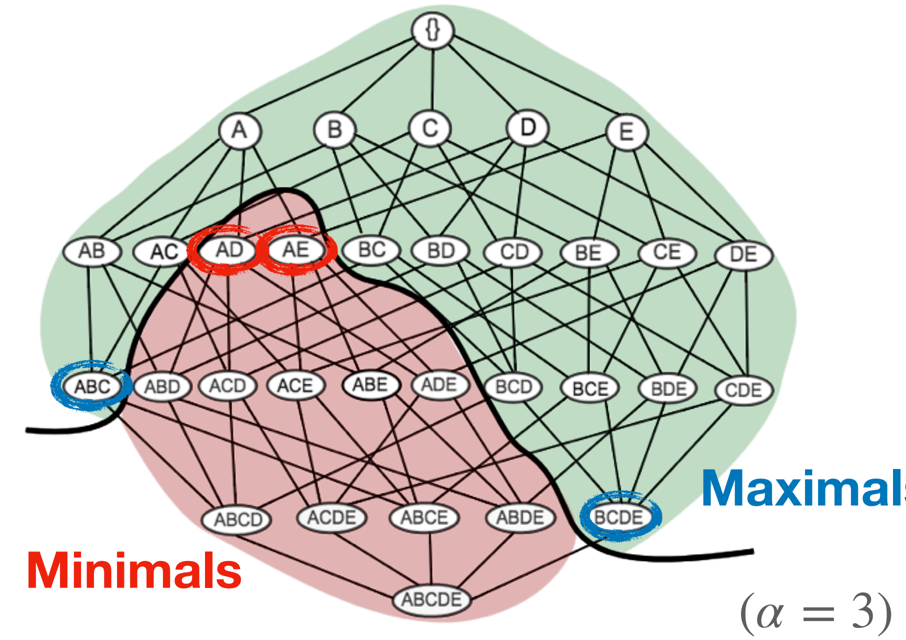
- A:[Frequent itemsets]
- $\neg A$:[Infrequent itemsets]
- **Borders (Maximals + Minimals):** patterns B:[with only frequent subsets] B and C:[with only infrequent supersets]
- **Thus,**

MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]



- A:[Frequent itemsets]
- $\neg A$:[Infrequent itemsets]
- **Borders (Maximals + Minimals):** patterns B:[with only frequent subsets] B and C:[with only infrequent supersets]
- **Thus,**
  - **Maximals + ~~Minimals~~:** B and C and A

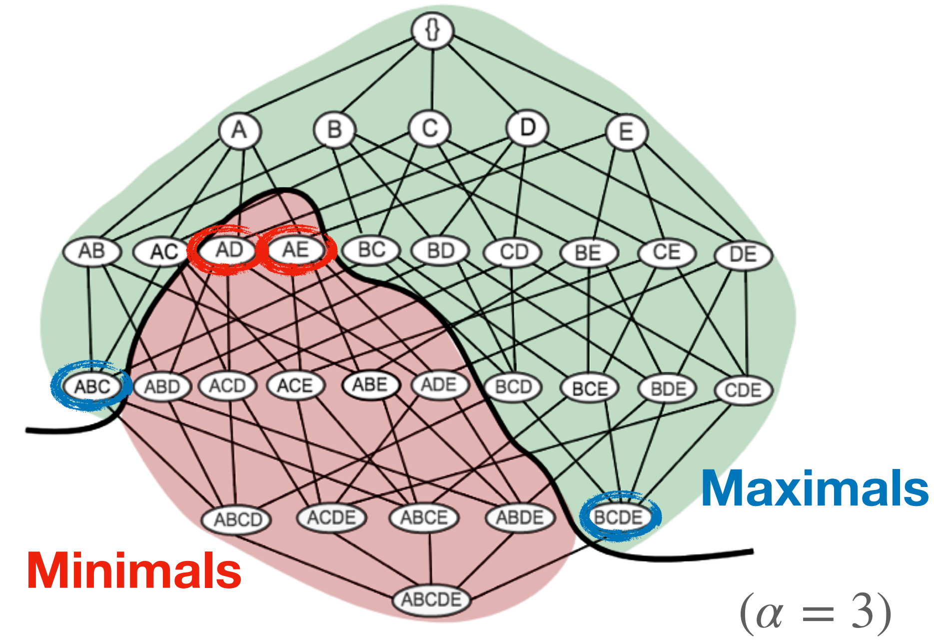
MINING MINIMALS AND/OR MAXIMALS [BELAID_BESSIERE_LAZAAR, IJCAI19]



- A:[Frequent itemsets]
- $\neg A$:[Infrequent itemsets]
- **Borders (Maximals + Minimals):** patterns B:[with only frequent subsets] B and C:[with only infrequent supersets]
- **Thus,**
  - **Maximals + ~~Minimals~~:** B and C and A
  - **~~Maximals~~ + Minimals:** B and C and $\neg A$

GENERIC CP MODEL FOR MINING BORDERS [BELAID_BESSIERE_LAZAAR, IJCAI19]

Borders(b):

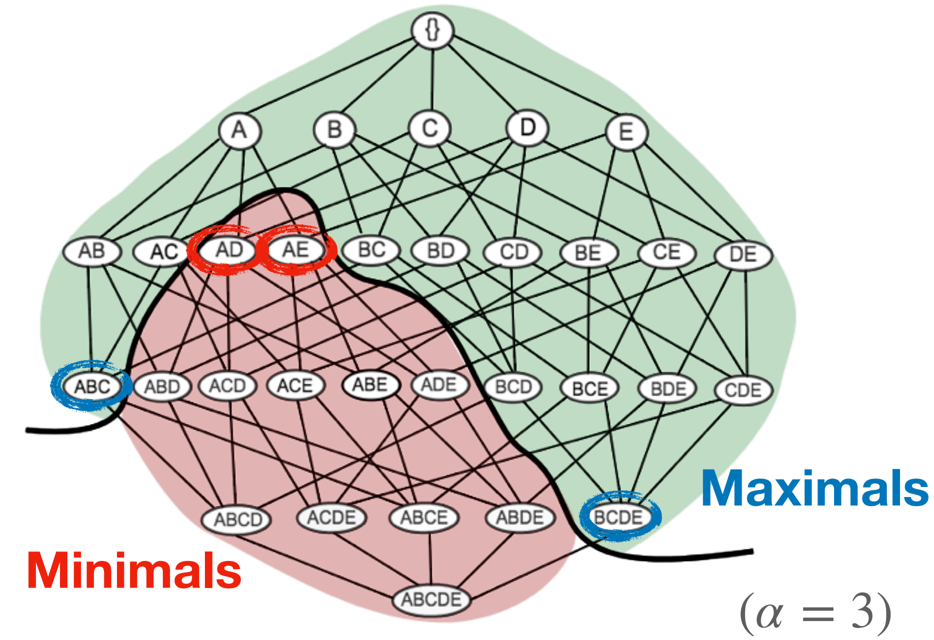


P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

GENERIC CP MODEL FOR MINING BORDERS [BELAID_BESSIERE_LAZAAR, IJCAI19]

Borders(b):

- ➤ $b \Leftrightarrow Frequent(P)$

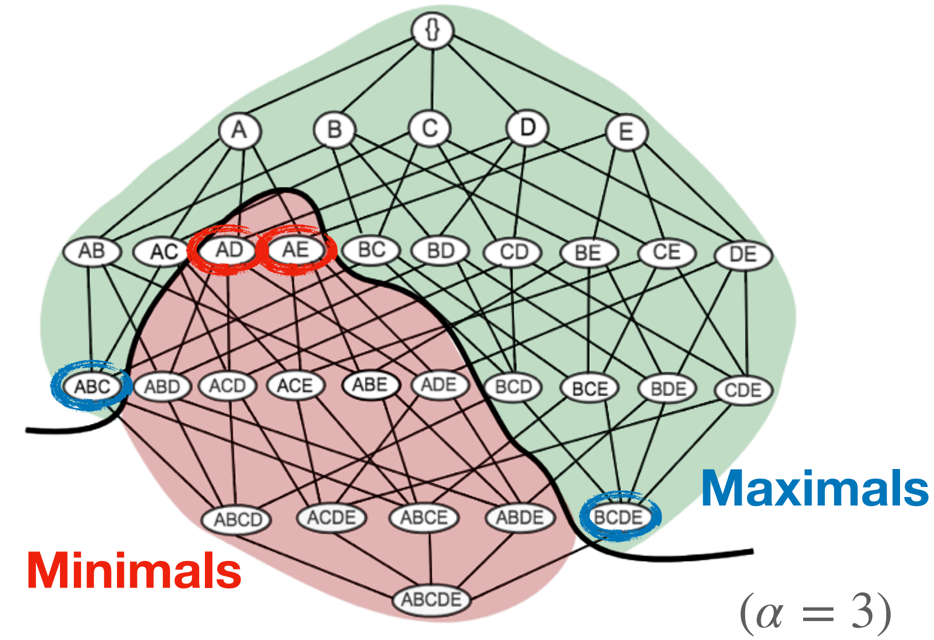


P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

GENERIC CP MODEL FOR MINING BORDERS [BELAID_BESSIERE_LAZAAR, IJCAI19]

`Borders(b):`

- ➤ $b \Leftrightarrow Frequent(P)$
- ➤ $FreqSubs(P)$

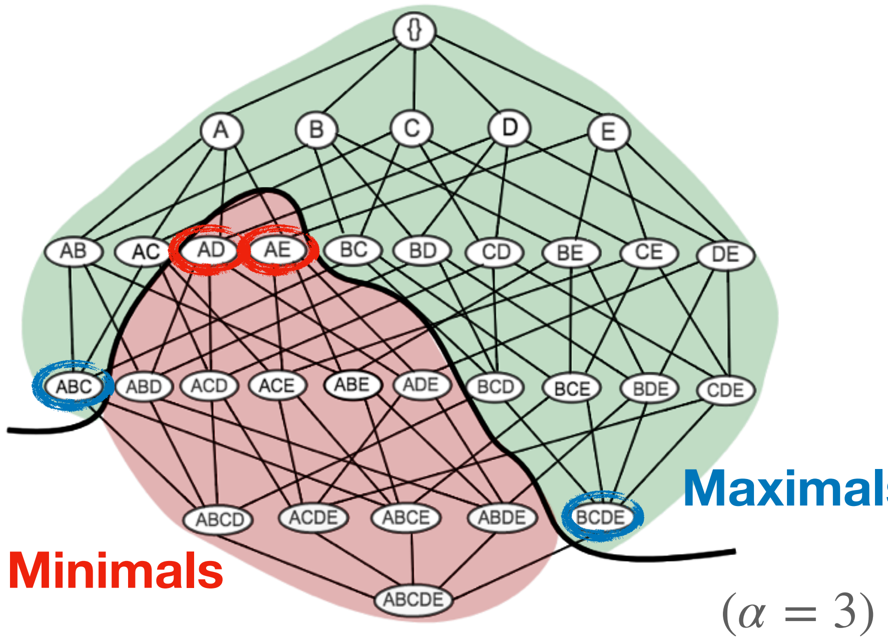


P								
	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

GENERIC CP MODEL FOR MINING BORDERS [BELAID_BESSIERE_LAZAAR, IJCAI19]

Borders(b):

- $b \Leftrightarrow Frequent(P)$
- $FreqSubs(P)$
- $InfreqSupers(P)$



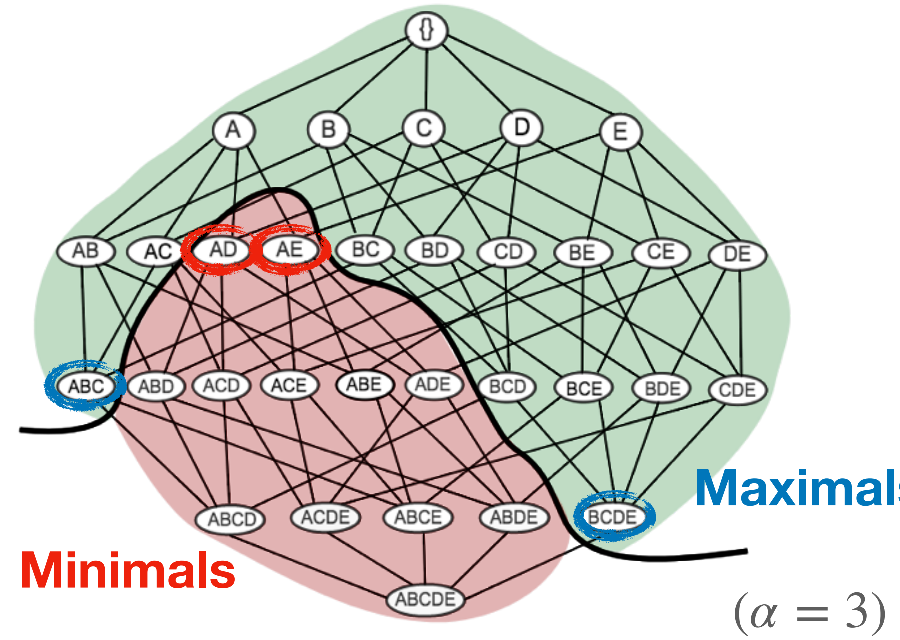
$(\alpha = 3)$

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

GENERIC CP MODEL FOR MINING BORDERS [BELAID_BESSIERE_LAZAAR, IJCAI19]

`Borders(b):`

- $b \Leftrightarrow \mathit{Frequent}(P)$
- $\mathit{FreqSubs}(P)$
- $\mathit{InfreqSupers}(P)$
  - $b = \mathit{true}$: **Maximals!**



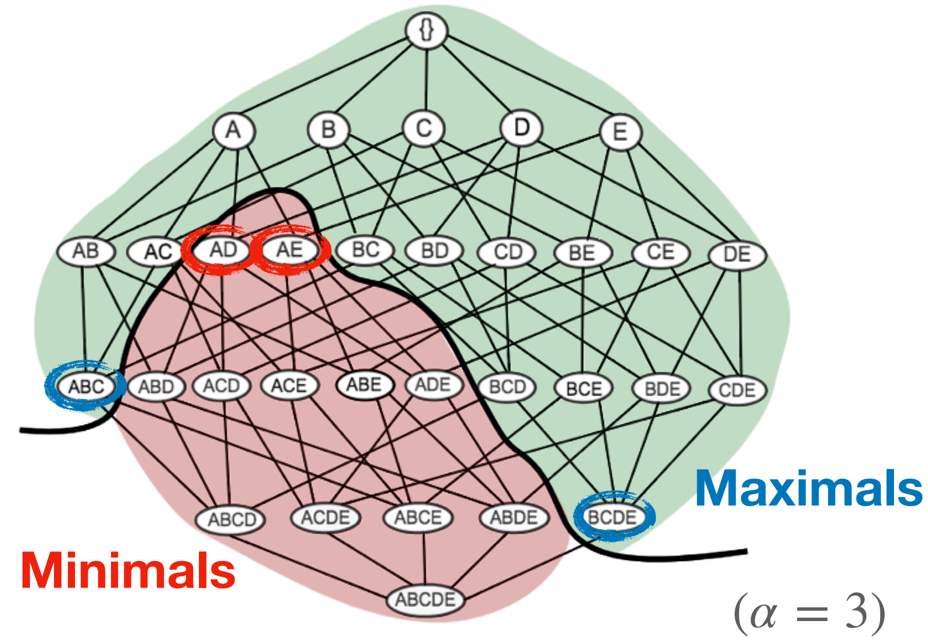
$(\alpha = 3)$

P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

GENERIC CP MODEL FOR MINING BORDERS [BELAID_BESSIERE_LAZAAR, IJCAI19]

```
Borders(b):
```

- $b \Leftrightarrow Frequent(P)$
- $FreqSubs(P)$
- $InfreqSupers(P)$
  - $b = true$: **Maximals!**
  - $b = false$: **Minimals!**



P	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
t1:		B	C		E	F	G	H
t2:	A			D			G	
t3:	A		C	D				H
t4:	A				E	F		
t5:		B			E	F		
t6:		B			E	F	G	

Comparative Analysis:

Declarative vs Specialized Methods

When Does Each Approach Excel?

SOME RESULTS

[IJCAI19]

SOME RESULTS

[IJCAI19]

Basic Data Mining Task

Instances	Eclat-Z	CP	#Rules
Zoo_5	**24.99**	43.45	30,792,317
Vote_5	**5.66**	6.43	2,075,212
Anneal_80	**93.96**	208.87	84,589,753
Chess_60	**15.98**	68.74	17,522,446
Mushroom_10	**27.81**	50.67	14,331,056
Connect_90	**3.03**	125.48	3,640,704
T10_0.02	**15.12**	63.35	1,303,932
T40_0.1	TO	TO	$> 7.10^9$
Pumsb_80	**34.35**	1567.72	19,749,382

TO: **timeout of one hour**

SOME RESULTS

[IJCAI19]

Basic Data Mining Task

Instances	ECLAT-Z	CP	#RULES
Zoo_5	**24.99**	43.45	30,792,317
Vote_5	**5.66**	6.43	2,075,212
Anneal_80	**93.96**	208.87	84,589,753
Chess_60	**15.98**	68.74	17,522,446
Mushroom_10	**27.81**	50.67	14,331,056
Connect_90	**3.03**	125.48	3,640,704
T10_0.02	**15.12**	63.35	1,303,932
T40_0.1	TO	TO	$> 7.10^9$
Pumsb_80	**34.35**	1567.72	19,749,382

TO: **timeout of one hour**

Specialised approach wins!

SOME RESULTS

[IJCAI19]

Basic Data Mining Task

Instances	ECLAT-Z	CP	#RULES
Zoo_5	**24.99**	43.45	30,792,317
Vote_5	**5.66**	6.43	2,075,212
Anneal_80	**93.96**	208.87	84,589,753
Chess_60	**15.98**	68.74	17,522,446
Mushroom_10	**27.81**	50.67	14,331,056
Connect_90	**3.03**	125.48	3,640,704
T10_0.02	**15.12**	63.35	1,303,932
T40_0.1	TO	TO	$> 7.10^9$
Pumsb_80	**34.35**	1567.72	19,749,382

TO: timeout of one hour

First solution: 1.25s 390M solutions before TO

Specialised approach wins!

SOME RESULTS

[IJCAI19]

Basic Data Mining Task

Instances	ECLAT-Z	CP	#RULES
Zoo_5	**24.99**	43.45	30,792,317
Vote_5	**5.66**	6.43	2,075,212
Anneal_80	**93.96**	208.87	84,589,753
Chess_60	**15.98**	68.74	17,522,446
Mushroom_10	**27.81**	50.67	14,331,056
Connect_90	**3.03**	125.48	3,640,704
T10_0.02	**15.12**	63.35	1,303,932
T40_0.1	TO	TO	$> 7.10^9$
Pumsb_80	**34.35**	1567.72	19,749,382

TO: timeout of one hour

Specialised approach wins!

Complex Data Mining Task

Instances	*ub*	*lb*	ECLAT-Z-PP	CP	#RULES
Zoo_5	1	9	491.48	**0.06**	12
Vote_5	1	2	38.49	**0.05**	23
Anneal_80	1	12	1622.19	**0.15**	73
Chess_60	1	8	284.22	**0.08**	24
Mushroom_10	1	11	249.00	**0.07**	14
Connect_90	1	11	61.80	**0.26**	12
T10_0.02	1	11	84.47	**5.44**	0
T40_0.1	1	11	TO	**8.33**	39
Pumsb_80	1	12	741.49	**0.34**	32

TO: timeout of one hour

SOME RESULTS

[IJCAI19]

Basic Data Mining Task

Instances	ECLAT-Z	CP	#RULES
Zoo_5	**24.99**	43.45	30,792,317
Vote_5	**5.66**	6.43	2,075,212
Anneal_80	**93.96**	208.87	84,589,753
Chess_60	**15.98**	68.74	17,522,446
Mushroom_10	**27.81**	50.67	14,331,056
Connect_90	**3.03**	125.48	3,640,704
T10_0.02	**15.12**	63.35	1,303,932
T40_0.1	TO	TO	$> 7.10^9$
Pumsb_80	**34.35**	1567.72	19,749,382

TO: timeout of one hour

Specialised approach wins!

Complex Data Mining Task

Instances	*ub*	*lb*	ECLAT-Z-PP	CP	#RULES
Zoo_5	1	9	491.48	**0.06**	12
Vote_5	1	2	38.49	**0.05**	23
Anneal_80	1	12	1622.19	**0.15**	73
Chess_60	1	8	284.22	**0.08**	24
Mushroom_10	1	11	249.00	**0.07**	14
Connect_90	1	11	61.80	**0.26**	12
T10_0.02	1	11	84.47	**5.44**	0
T40_0.1	1	11	TO	**8.33**	39
Pumsb_80	1	12	741.49	**0.34**	32

TO: timeout of one hour

Declarative approach wins!

COMPLEX QUERY

[CP18]

COMPLEX QUERY

[CP18]

- **QUERY:**
  - Frequent closed itemset (cstrs on itemsets)
  - from at least lb_I and at most ub_I item categories (cstrs on items)
  - and at least lb_T and at most ub_T transaction categories (cstrs on trans.)

COMPLEX QUERY

[CP18]

- **QUERY:**
  - Frequent closed itemset (cstrs on itemsets)
  - from at least lb_I and at most ub_I item categories (cstrs on items)
  - and at least lb_T and at most ub_T transaction categories (cstrs on trans.)

Instances	$\#\mathcal{I}_i$	$\#\mathcal{T}_i$	(lb_I, ub_I)	(lb_T, ub_T)	$\#D$	#FCIs	PP-LCM	CP-ITEMSET
Zoo_70_6	6	10	(2,3)	(2,3)	5,775	8	39.69	**1.75**
Zoo_50_11	6	10	(3,4)	(3,4)	11,550	9	88.66	**3.36**
Zoo_85_5	6	10	(2,6)	(2,10)	57,741	8	521.89	**31.86**
Primary_82_5	3	12	(2,3)	(2,10)	16,280	8	199.58	**36.13**
Vote_70_6	6	29	(2,3)	(2,3)	142,100	2	TO	**118.67**
Vote_72_5	8	29	(2,3)	(2,3)	341,040	2	TO	**201.79**
Mushroom_80_5	17	12	(2,2)	(2,2)	8,976	10	446.42	**102.68**
Mushroom_82_5	17	12	(2,2)	(3,3)	29,920	7	TO	**455.19**
Chess_90_16	5	34	(2,3)	(2,2)	11,220	3	286.42	**87.22**

Time in seconds

COMPLEX QUERY

[CP18]

- QUERY:
  - Frequent closed itemset (cstrs on itemsets)
  - from at least lb_I and at most ub_I item categories (cstrs on items)
  - and at least lb_T and at most ub_T transaction categories (cstrs on trans.)

Instances	$\#\mathcal{I}_i$	$\#\mathcal{T}_i$	(lb_I, ub_I)	(lb_T, ub_T)	$\#D$	#FCIs	PP-LCM	CP-ITEMSET
Zoo_70_6	6	10	(2,3)	(2,3)	5,775	8	39.69	**1.75**
Zoo_50_11	6	10	(3,4)	(3,4)	11,550	9	88.66	**3.36**
Zoo_85_5	6	10	(2,6)	(2,10)	57,741	8	521.89	**31.86**
Primary_82_5	3	12	(2,3)	(2,10)	16,280	8	199.58	**36.13**
Vote_70_6	6	29	(2,3)	(2,3)	142,100	2	TO	**118.67**
Vote_72_5	8	29	(2,3)	(2,3)	341,040	2	TO	**201.79**
Mushroom_80_5	17	12	(2,2)	(2,2)	8,976	10	446.42	**102.68**
Mushroom_82_5	17	12	(2,2)	(3,3)	29,920	7	TO	**455.19**
Chess_90_16	5	34	(2,3)	(2,2)	11,220	3	286.42	**87.22**

Time in seconds

- CP-ItemSet is 4 to 26 times faster than PP-LCM
- PP is about 90% of the total time

TAKE-AWAY MESSAGE

- Specialised methods are suitable for:
  - Enumerating (millions of) Patterns
  - Handling classic constraints (simple queries)
- Declarative methods are suitable for:
  - Taking into account user's constraints (complex queries)
  - Supporting iterative and interactive data mining processes

- Advantages of using CP : compactness, expressiveness and (often) higher efficiency

➤ Ability to Equip CP solvers with DM features (e.g., advanced data structures, global constraints, hybrid and adaptive reasoning mechanisms)

Declarative Itemset Mining

Master 1 - IDS

Nadjib Lazaar
Ing - Phd - HDR - Professor - Paris-Saclay University - LISN - LaHDAK
lazaar@lisn.fr https://perso.lisn.upsaclay.fr/lazaar/
03/02/2025