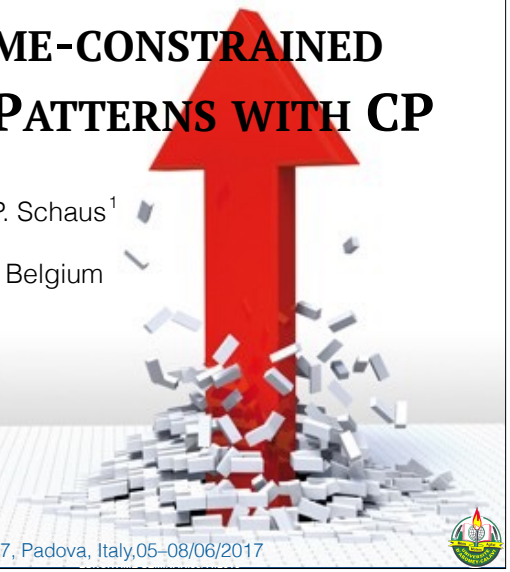
Mining Time-constrained Sequential Patterns with CP

J. AOGA[1], T. Guns[2], P. Schaus[1]

[1]UCLouvain, [2]VUB — Belgium



CPAIOR 2017, Padova, Italy,05–08/06/2017

General Overview

Sequential Pattern Mining (SPM)

General Overview

Telecom
- Network analysis
- People behavior
- Inter-Connection

Sequential Pattern Mining (SPM)

General Overview

Telecom
- Network analysis
- People behavior
- Inter-Connection

Bioinformatics
- DNA Analysis
- Disease Analysis
- Gene Analysis

Sequential Pattern Mining (SPM)

General Overview

Telecom
- Network analysis
- People behavior
- Inter-Connection

Bioinformatics
- DNA Analysis
- Disease Analysis
- Gene Analysis

Sequential Pattern Mining (SPM)

Recommendation
- Purchase analysis
- web usage mining
- Comments mining

SPM Problem

Client1	Milk	Coffee	Sugar	Coffee	Sugar
Client2	Coffee	Milk	Coffee	Sugar	
Client3	Milk	Coffee			
Client4	Coffee	Sugar	Egg		

Sequence Database (SDB)

SPM Problem

Sequence

Client1	Milk	Coffee	Sugar	Coffee	Sugar
Client2	Coffee	Milk	Coffee	Sugar	
Client3	Milk	Coffee			
Client4	Coffee	Sugar	Egg		

Sequence Database (SDB)

- Sequence : < Milk Coffee Sugar Coffee Sugar>

SPM Problem

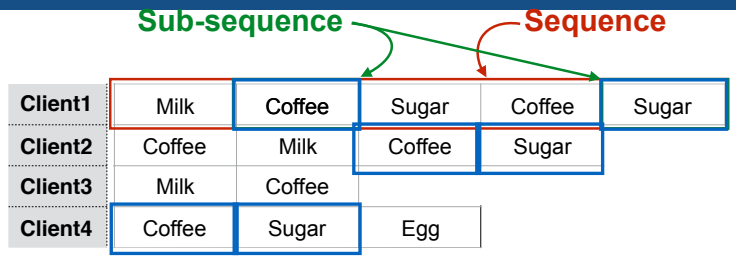
Sub-sequence Sequence

Client1	Milk	Coffee	Sugar	Coffee	Sugar
Client2	Coffee	Milk	Coffee	Sugar	
Client3	Milk	Coffee			
Client4	Coffee	Sugar	Egg		

Sequence Database (SDB)

- Sequence : < Milk Coffee Sugar Coffee Sugar>
- Sub-sequence : <Coffee Sugar>

SPM Problem

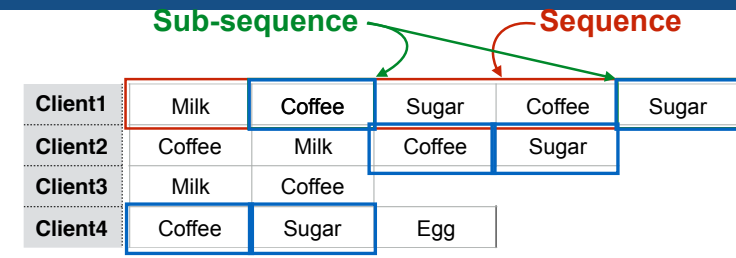


Client1	Milk	**Coffee**	Sugar	Coffee	Sugar
Client2	Coffee	Milk	Coffee	Sugar	
Client3	Milk	Coffee			
Client4	Coffee	Sugar	Egg		

Sequence Database (SDB)

- Sequence : < Milk Coffee Sugar Coffee Sugar>
- Sub-sequence : <Coffee Sugar>
- Support (<Coffee Sugar>) = 3

SPM Problem



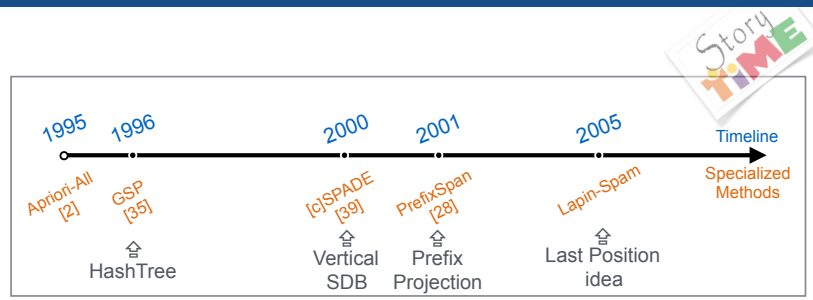
Client1	Milk	**Coffee**	Sugar	Coffee	Sugar
Client2	Coffee	Milk	Coffee	Sugar	
Client3	Milk	Coffee			
Client4	Coffee	Sugar	Egg		

Sequence Database (SDB)

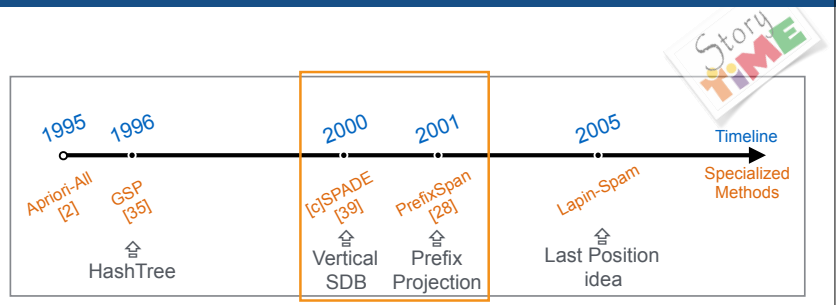
- Sequence : < Milk Coffee Sugar Coffee Sugar>
- Sub-sequence : <Coffee Sugar>
- Support (<Coffee Sugar>) = 3

Problem : Find all subsequences with support ≥ Given Threshold

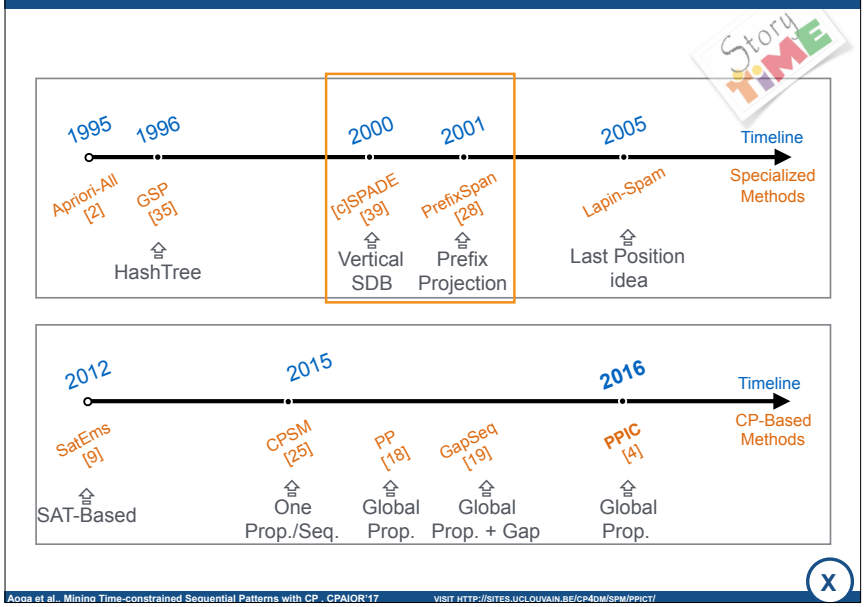
Related Work



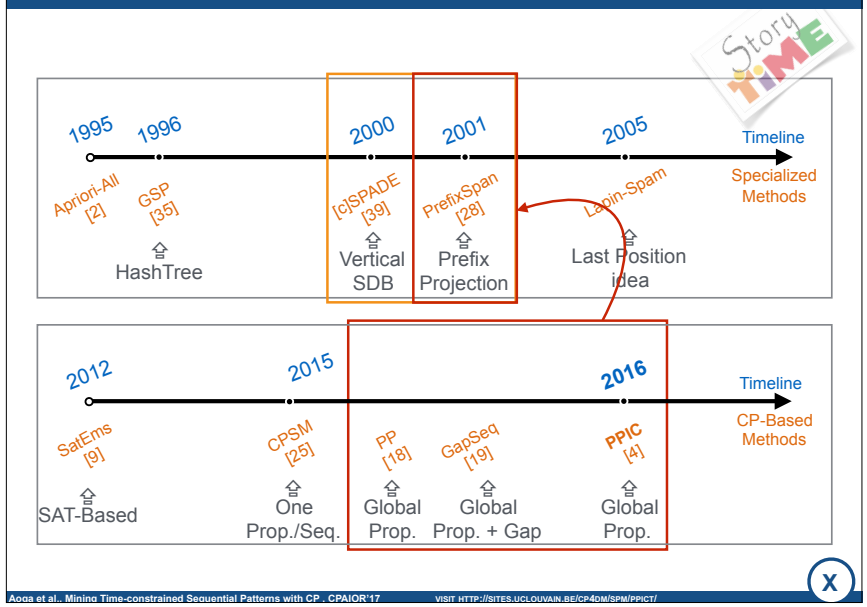
Related Work



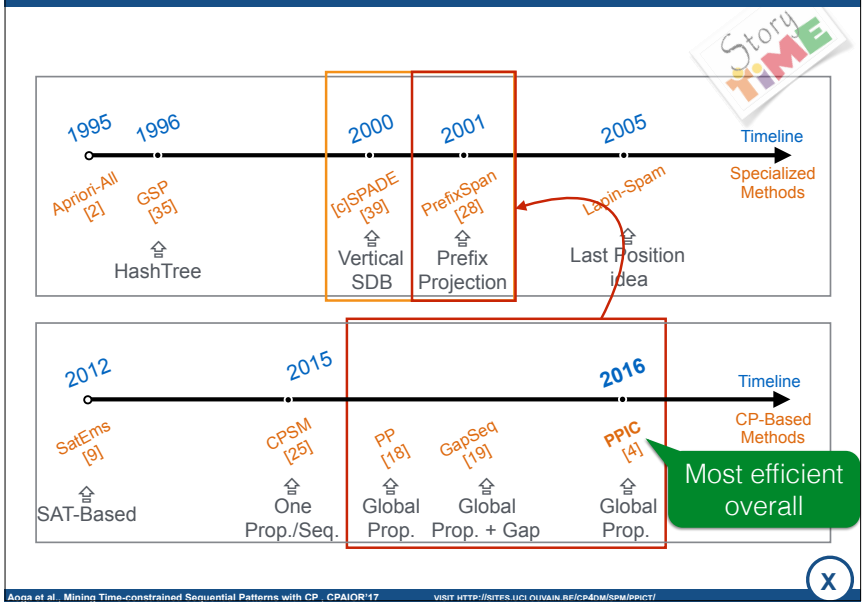
Related Work



Related Work



Related Work



PPICT : CONTRIBUTION

Goal: Capture the most common time-related constraints: namely timed events, minimum/maximum gap and span

☑ *Adapt* trailed-based data structure to efficiently capture **all** valid embeddings (previously only smallest needed)

☑ Algorithmic improvements to avoid scanning overlapping time windows, and to efficiently compute the frequency of symbols

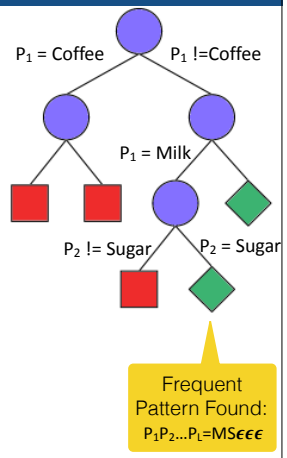
☑ Can be combined with many other contraints: Regular/ Grammar, Gcc, Among, …

CP : Filtering + DFSearch

Vi	P_1	P_2	P_3	P_4	P_5
Di	ϵ	ϵ	ϵ	ϵ	ϵ
	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg

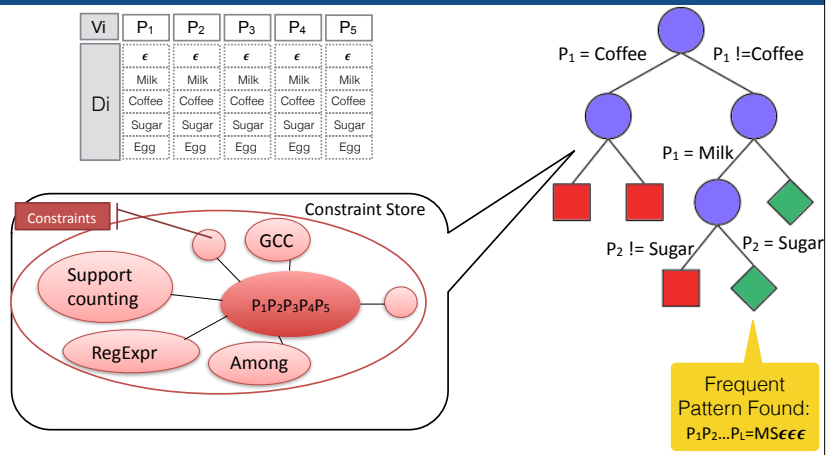
CP : Filtering + DFSearch

Vi	P_1	P_2	P_3	P_4	P_5
Di	ϵ	ϵ	ϵ	ϵ	ϵ
	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg



CP : Filtering + DFSearch

Vi	P_1	P_2	P_3	P_4	P_5
Di	ϵ	ϵ	ϵ	ϵ	ϵ
	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg



CP : Filtering + DFSearch

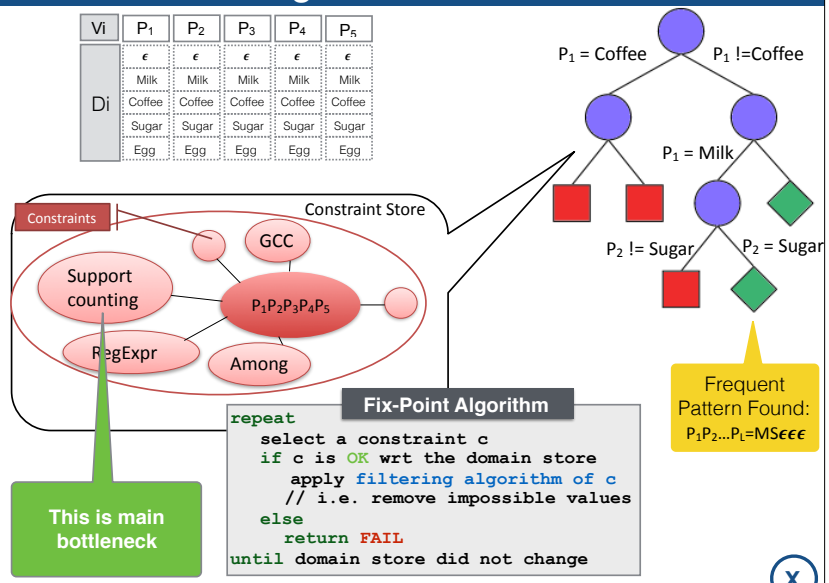
Vi	P_1	P_2	P_3	P_4	P_5
Di	ϵ	ϵ	ϵ	ϵ	ϵ
	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg

Fix-Point Algorithm

```
repeat
   select a constraint c
   if c is OK wrt the domain store
      apply filtering algorithm of c
      // i.e. remove impossible values
   else
      return FAIL
until domain store did not change
```

CP : Filtering + DFSearch

Vi	P_1	P_2	P_3	P_4	P_5
Di	ϵ	ϵ	ϵ	ϵ	ϵ
	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg



MinSup=3 (75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

MinSup=3 (75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

P_1	P_2	P_3	P_4	P_5
	ϵ	ϵ	ϵ	ϵ
Milk	Milk	Milk	Milk	Milk
Coffee	Coffee	Coffee	Coffee	Coffee
Sugar	Sugar	Sugar	Sugar	Sugar
Egg	Egg	Egg	Egg	Egg

MinSup=3 (75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

Supports M :

P_1	P_2	P_3	P_4	P_5
	ϵ	ϵ	ϵ	ϵ
Milk	Milk	Milk	Milk	Milk
Coffee	Coffee	Coffee	Coffee	Coffee
Sugar	Sugar	Sugar	Sugar	Sugar
Egg	Egg	Egg	Egg	Egg

MinSup=3
(75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

Supports
M : 3
C : 4
S : 3
~~E : 1~~

P_1	P_2	P_3	P_4	P_5
	ϵ	ϵ	ϵ	ϵ
Milk	Milk	Milk	Milk	Milk
Coffee	Coffee	Coffee	Coffee	Coffee
Sugar	Sugar	Sugar	Sugar	Sugar
Egg	Egg	Egg	Egg	Egg

MinSup=3
(75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

Supports
M : 3
C : 4
S : 3
~~E : 1~~

P_1	P_2	P_3	P_4	P_5
	ϵ	ϵ	ϵ	ϵ
Milk	Milk	Milk	Milk	Milk
Coffee	Coffee	Coffee	Coffee	Coffee
Sugar	Sugar	Sugar	Sugar	Sugar
Egg	Egg	Egg	Egg	Egg

Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17 VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

MinSup=3
(75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

Supports
M : 3
C : 4
S : 3
~~E : 1~~

P_1	P_2	P_3	P_4	P_5
	ϵ	ϵ	ϵ	ϵ
Milk	Milk	Milk	Milk	Milk
Coffee	Coffee	Coffee	Coffee	Coffee
Sugar	Sugar	Sugar	Sugar	Sugar
Egg	Egg	Egg	Egg	Egg

Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17 VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

MinSup=3
(75%)

	0	1	2	3	4
1	M	C	S	C	S
2	C	M	C	S	
3	M	C			
4	C	S	E		

Supports
M : 3
C : 4
S : 3
~~E : 1~~

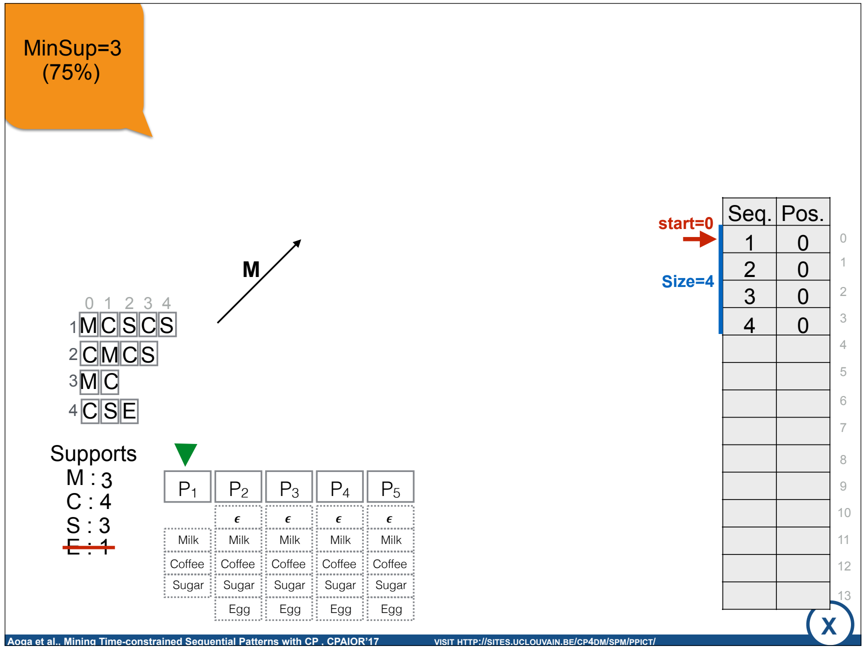
P_1	P_2	P_3	P_4	P_5
	ϵ	ϵ	ϵ	ϵ
Milk	Milk	Milk	Milk	Milk
Coffee	Coffee	Coffee	Coffee	Coffee
Sugar	Sugar	Sugar	Sugar	Sugar
Egg	Egg	Egg	Egg	Egg

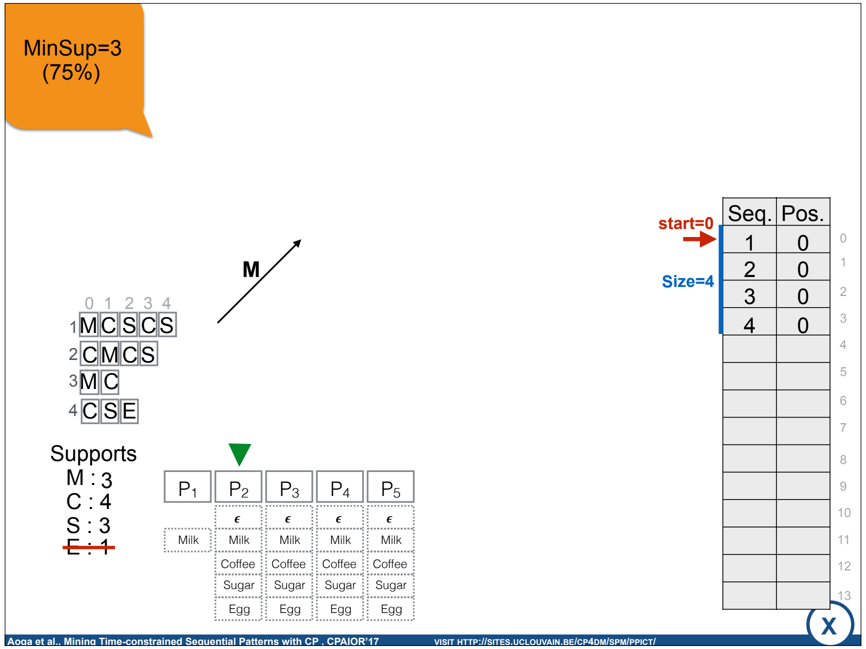
start=0

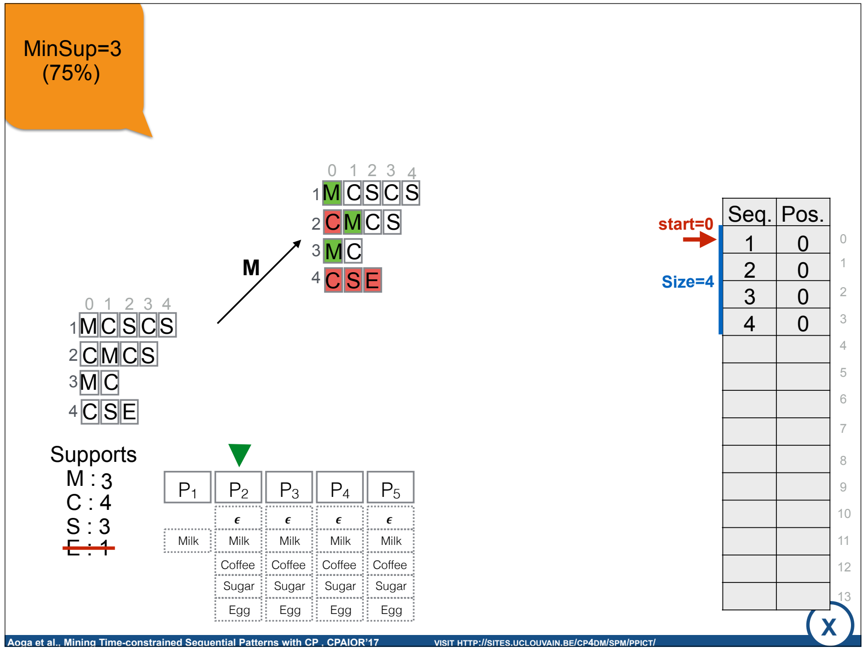
Size=4

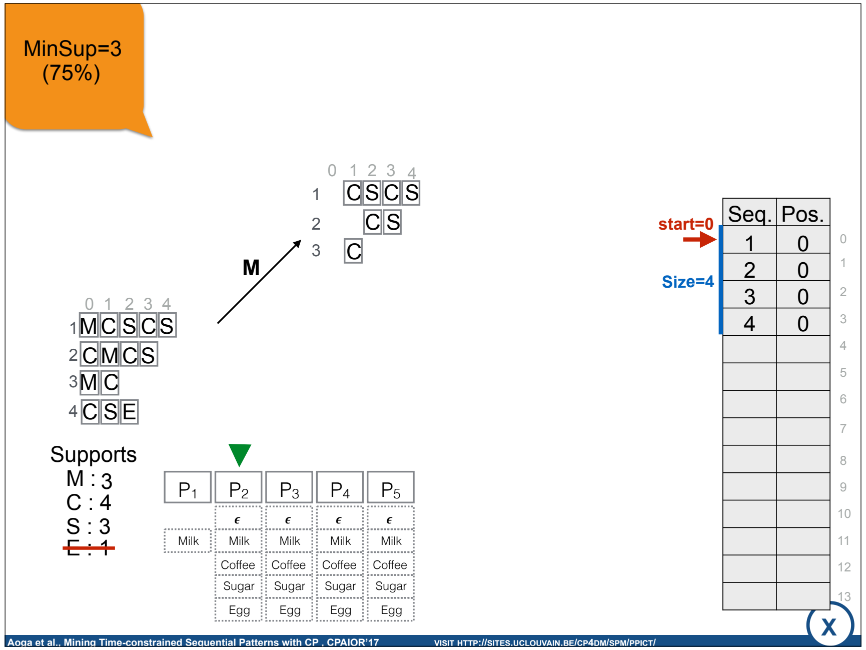
Seq.	Pos.
1	0
2	0
3	0
4	0

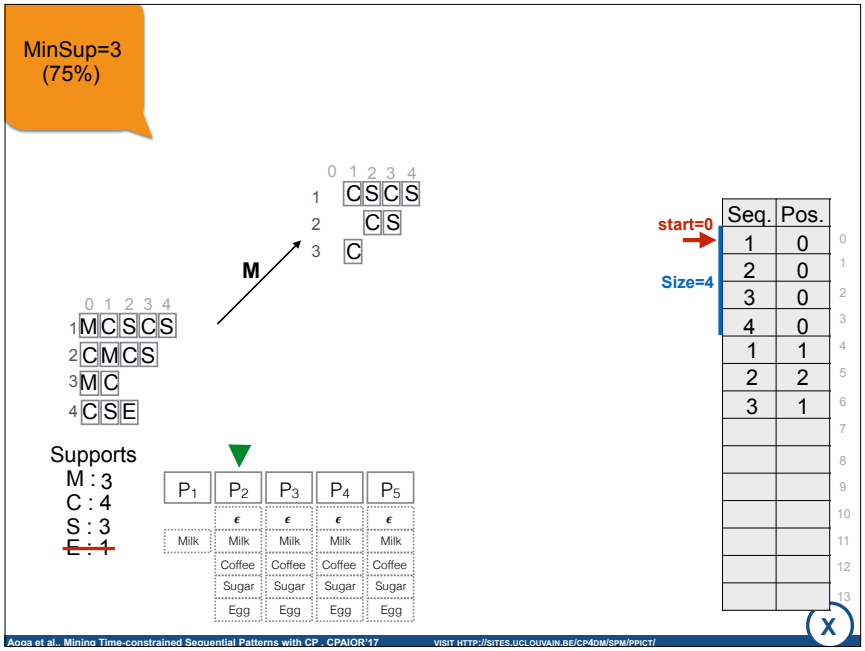
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17 VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

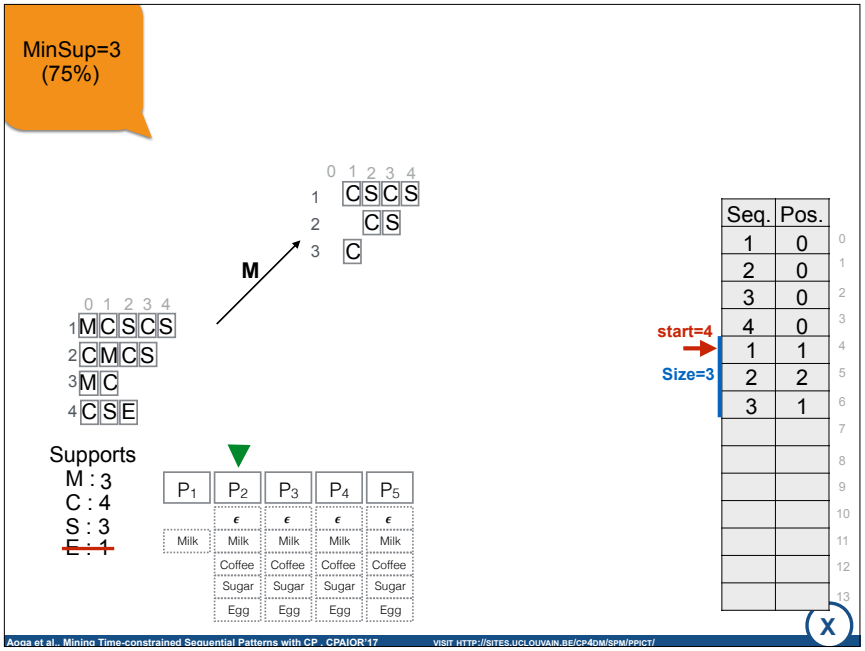

MinSup=3 (75%)
M
1 MCSCS
2 CMCS
3 MC
4 CSE
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
start=0
Size=4
Seq. Pos.
1 0
2 0
3 0
4 0
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

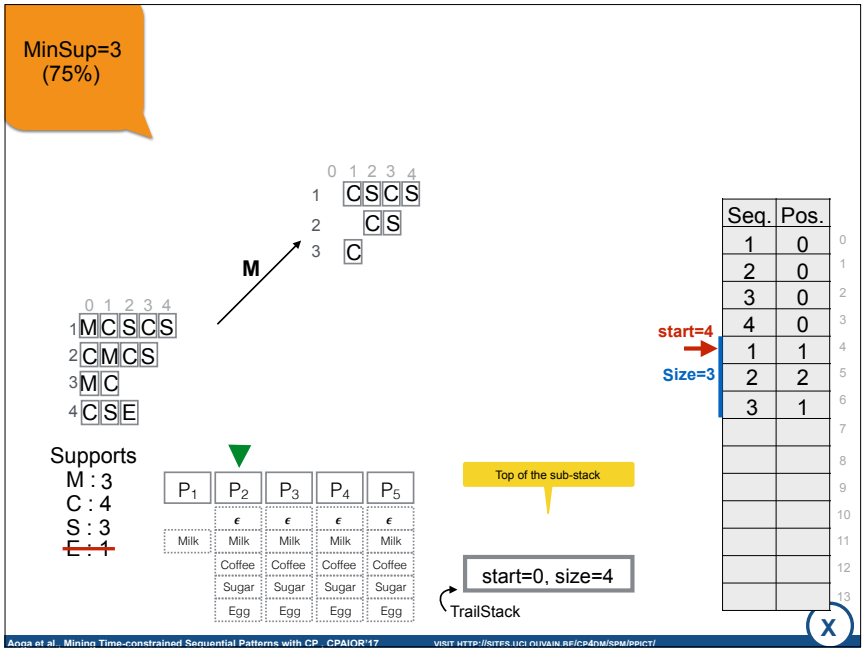

MinSup=3 (75%)
M
1 MCSCS
2 CMCS
3 MC
4 CSE
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
start=0
Size=4
Seq. Pos.
1 0
2 0
3 0
4 0
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

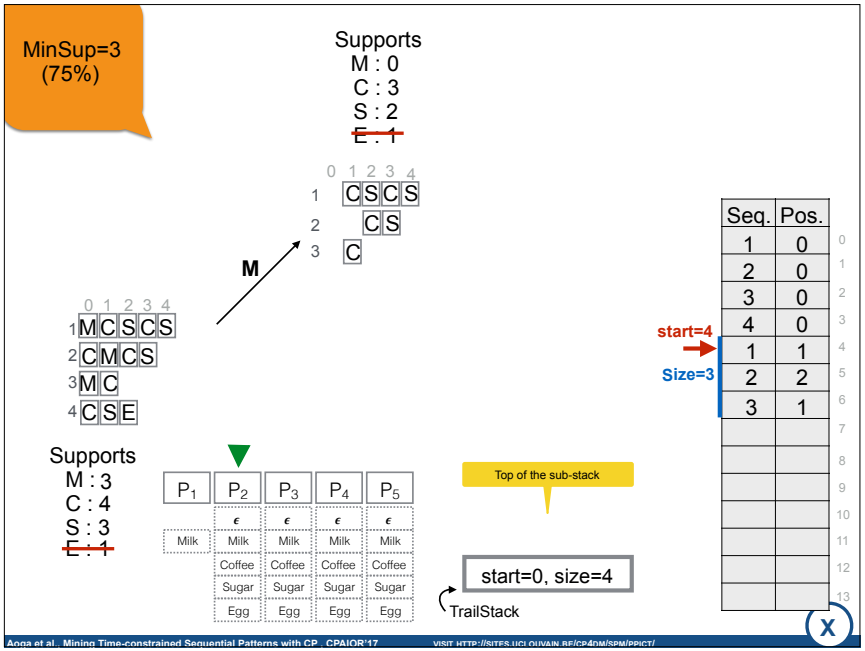

MinSup=3 (75%)
M
1 MCSCS
2 CMCS
3 MC
4 CSE
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
start=0
Size=4
Seq. Pos.
1 0
2 0
3 0
4 0
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

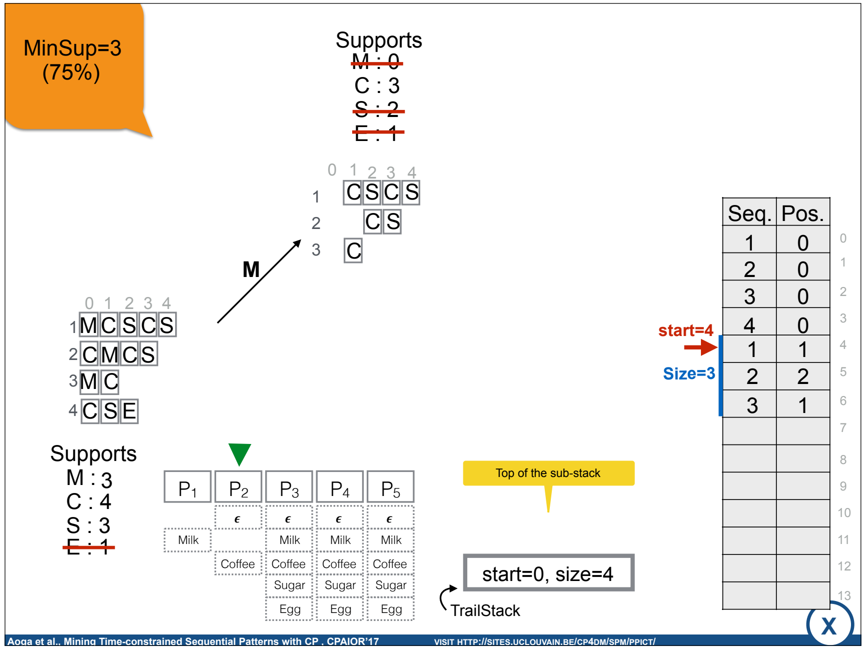

MinSup=3 (75%)
M
1 CSCS
2 CS
3 C
1 MCSCS
2 CMCS
3 MC
4 CSE
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
start=0
Size=4
Seq. Pos.
1 0
2 0
3 0
4 0
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

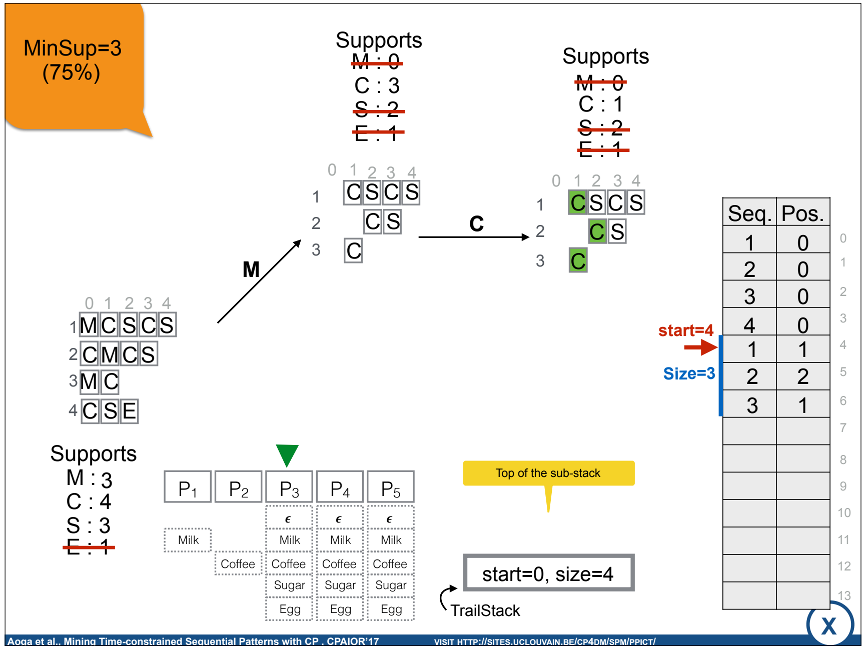

MinSup=3
(75%)
M
Supports
M : 3
C : 4
S : 3
E : 1
start=0
Size=4
Seq.
Pos.
Milk
Coffee
Sugar
Egg
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

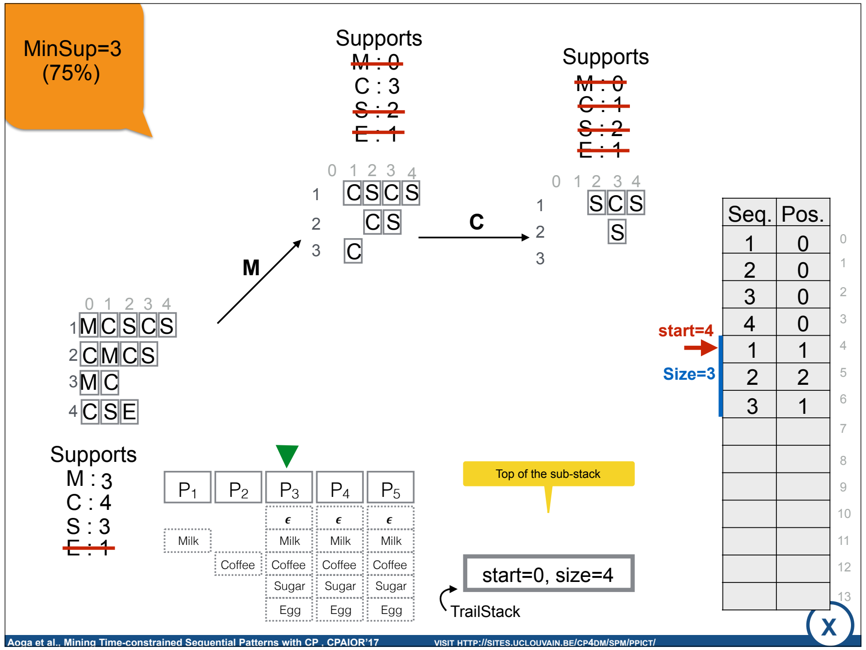

MinSup=3
(75%)
M
Supports
M : 3
C : 4
S : 3
E : 1
start=4
Size=3
Seq.
Pos.
Milk
Coffee
Sugar
Egg
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

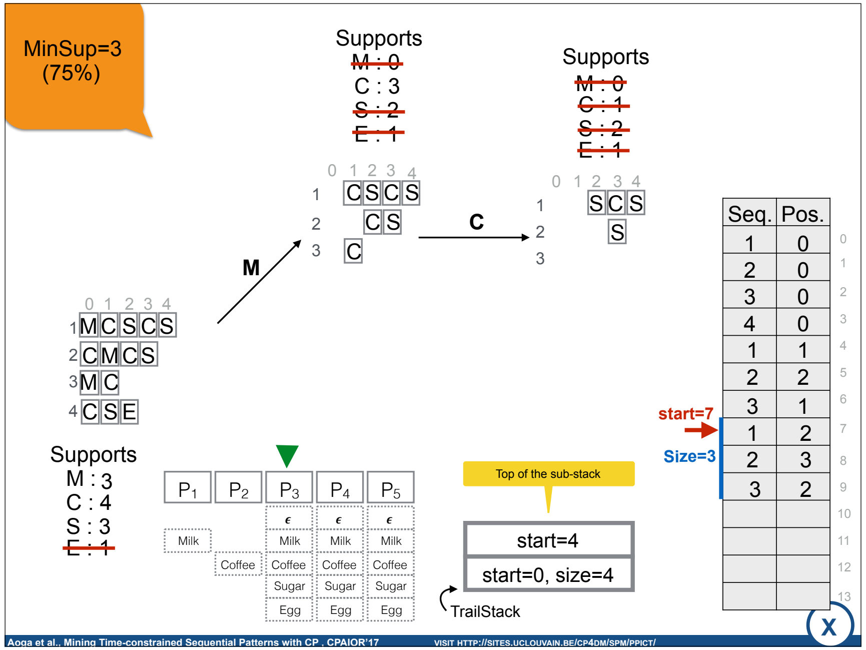

MinSup=3
(75%)
M
Supports
M : 3
C : 4
S : 3
E : 1
start=4
Size=3
Top of the sub-stack
start=0, size=4
TrailStack
Seq.
Pos.
Milk
Coffee
Sugar
Egg
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

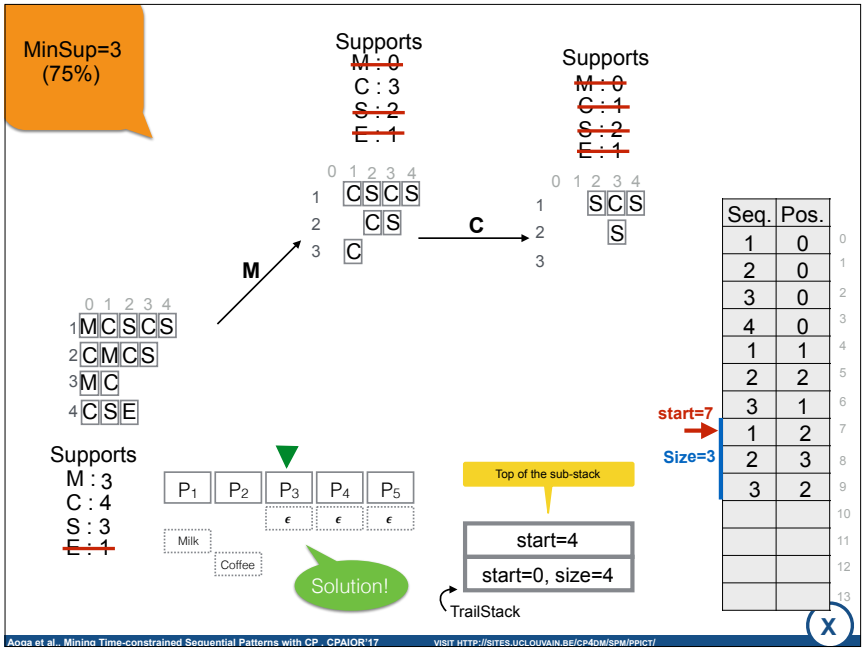

MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
M
Supports
M : 3
C : 4
S : 3
E : 1
start=4
Size=3
Top of the sub-stack
start=0, size=4
TrailStack
Seq.
Pos.
Milk
Coffee
Sugar
Egg
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

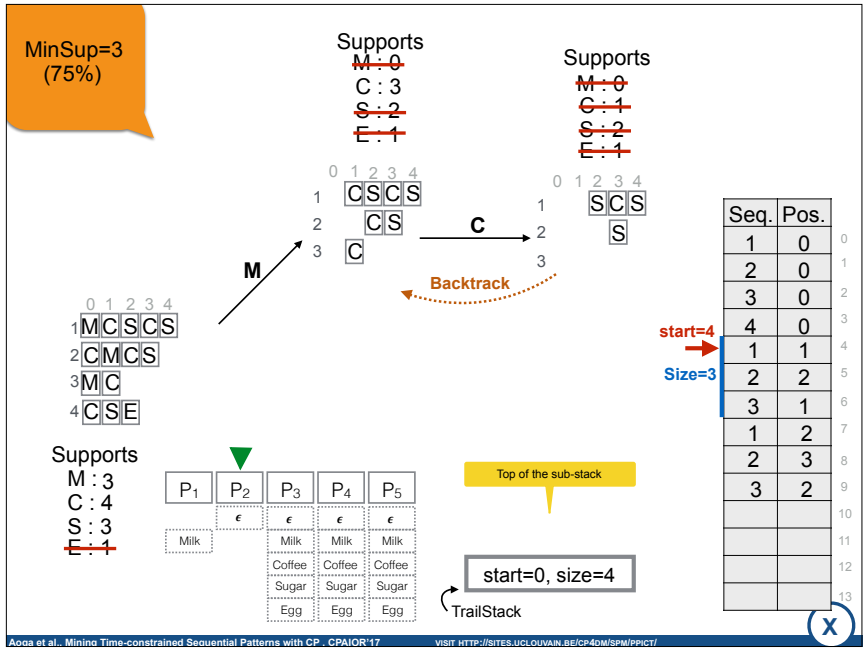

MinSup=3 (75%)
Supports
M : 0
C : 3
S : 2
E : 1
M
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Sugar
Egg
Top of the sub-stack
start=0, size=4
TrailStack
Seq. Pos.
start=4
Size=3
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

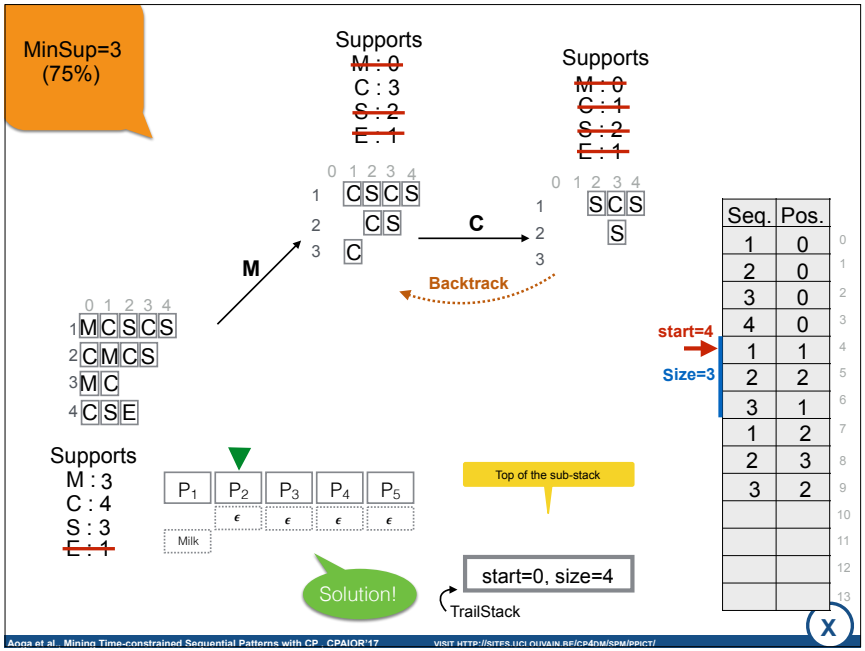

MinSup=3 (75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Sugar
Egg
Top of the sub-stack
start=0, size=4
TrailStack
Seq. Pos.
start=4
Size=3
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

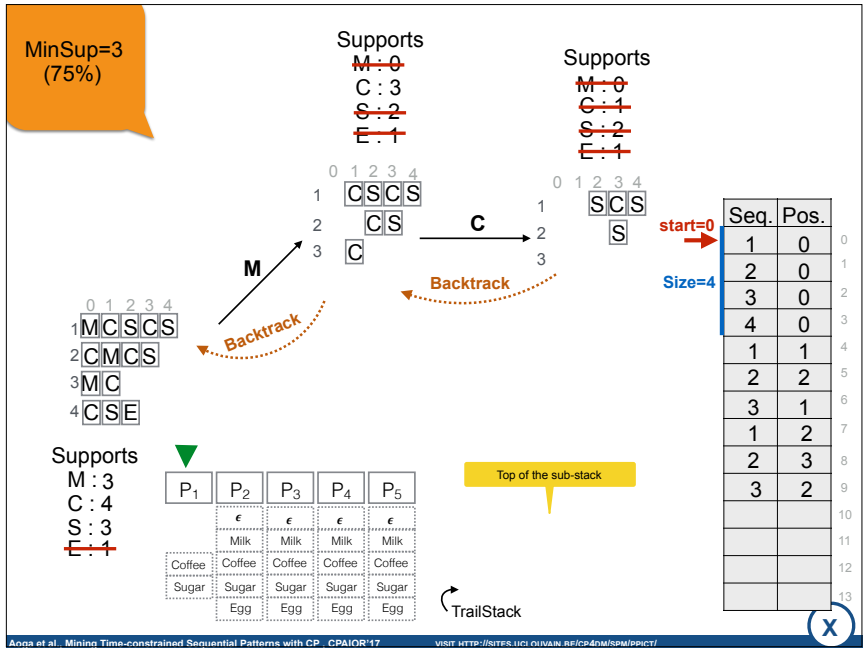

MinSup=3 (75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Sugar
Egg
Top of the sub-stack
start=0, size=4
TrailStack
Seq. Pos.
start=4
Size=3
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

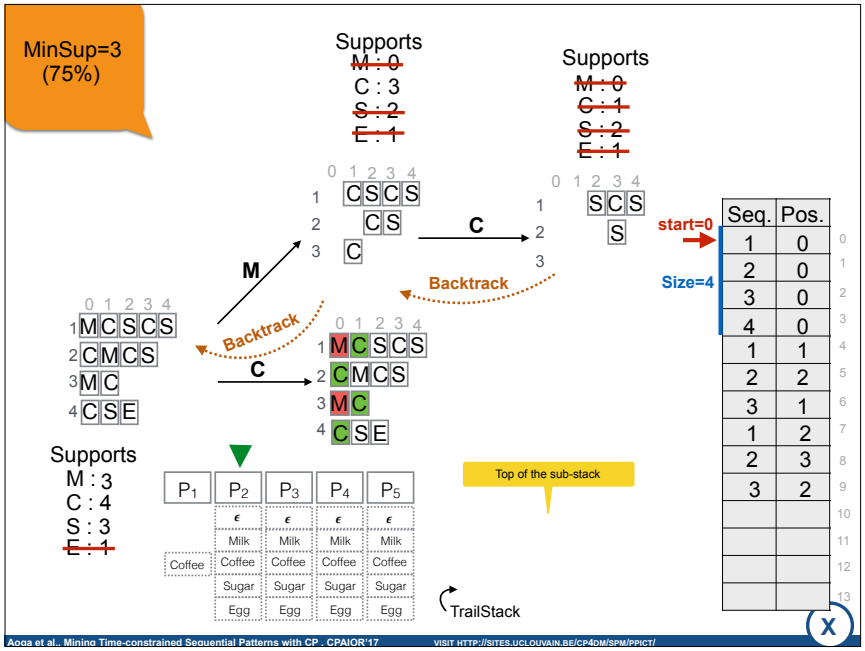

MinSup=3 (75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Sugar
Egg
Top of the sub-stack
start=4
start=0, size=4
TrailStack
Seq. Pos.
start=7
Size=3
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

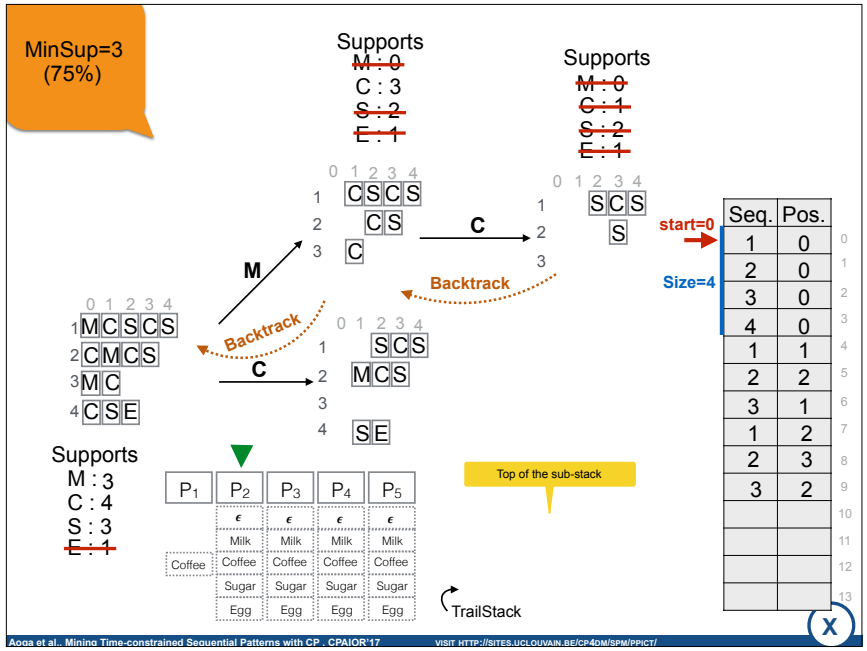

MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Seq. Pos.
start=7
Size=3
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Solution!
Top of the sub-stack
start=4
start=0, size=4
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

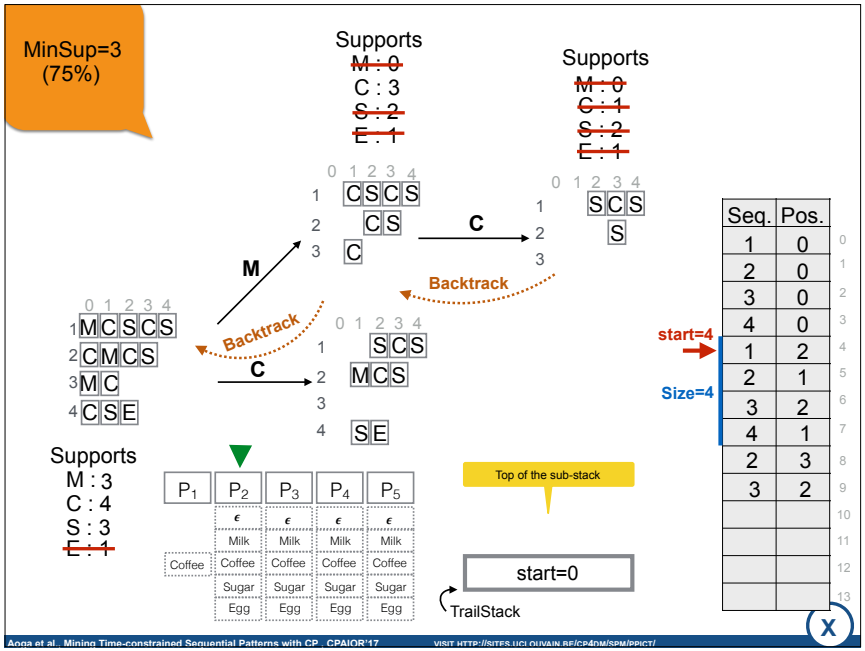

MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Backtrack
Seq. Pos.
start=4
Size=3
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Sugar
Egg
Top of the sub-stack
start=0, size=4
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

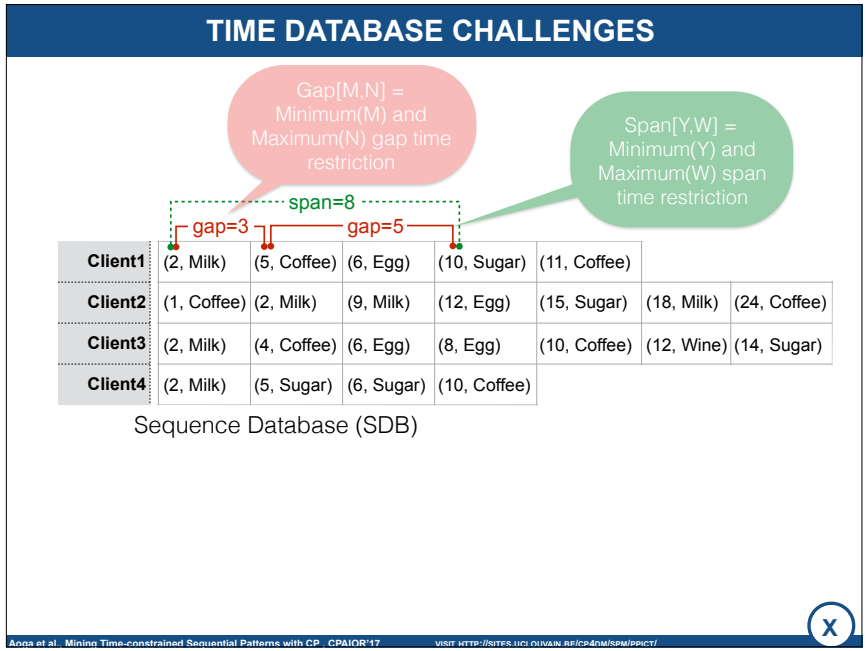

MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Backtrack
Seq. Pos.
start=4
Size=3
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Solution!
Top of the sub-stack
start=0, size=4
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
M
C
Backtrack
Backtrack
Seq. Pos.
start=0
Size=4
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Milk
Coffee
Sugar
Egg
Top of the sub-stack
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

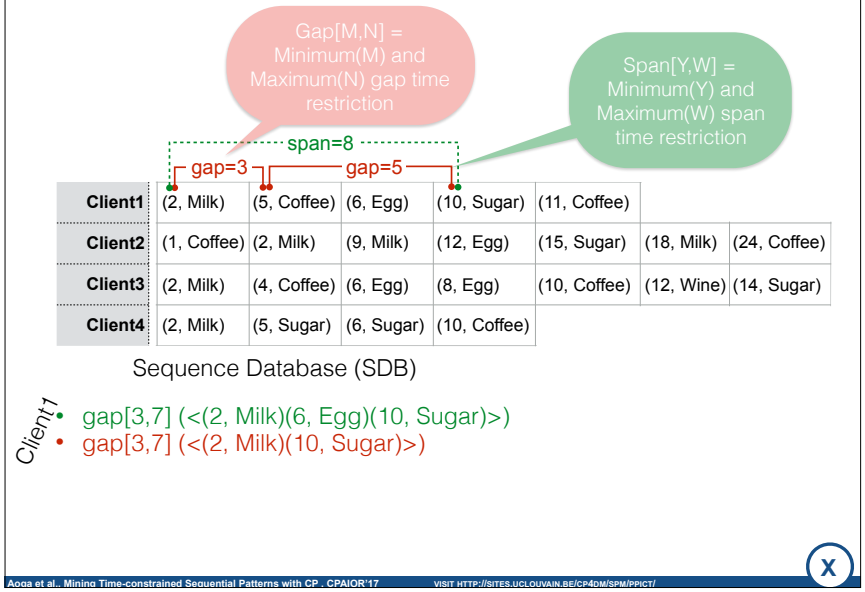

MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
C
M
Backtrack
Backtrack
C
start=0
Size=4
Seq. Pos.
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Coffee
Milk
Sugar
Egg
Top of the sub-stack
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
C
M
Backtrack
Backtrack
C
start=0
Size=4
Seq. Pos.
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Coffee
Milk
Sugar
Egg
Top of the sub-stack
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
(75%)
Supports
M : 0
C : 3
S : 2
E : 1
Supports
M : 0
C : 1
S : 2
E : 1
C
M
Backtrack
Backtrack
C
start=4
Size=4
Seq. Pos.
Supports
M : 3
C : 4
S : 3
E : 1
P1 P2 P3 P4 P5
Coffee
Milk
Sugar
Egg
Top of the sub-stack
start=0
TrailStack
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


TIME DATABASE CHALLENGES
Gap[M,N] = Minimum(M) and Maximum(N) gap time restriction
Span[Y,W] = Minimum(Y) and Maximum(W) span time restriction
span=8
gap=3
gap=5
Client1 (2, Milk) (5, Coffee) (6, Egg) (10, Sugar) (11, Coffee)
Client2 (1, Coffee) (2, Milk) (9, Milk) (12, Egg) (15, Sugar) (18, Milk) (24, Coffee)
Client3 (2, Milk) (4, Coffee) (6, Egg) (8, Egg) (10, Coffee) (12, Wine) (14, Sugar)
Client4 (2, Milk) (5, Sugar) (6, Sugar) (10, Coffee)
Sequence Database (SDB)
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

TIME DATABASE CHALLENGES

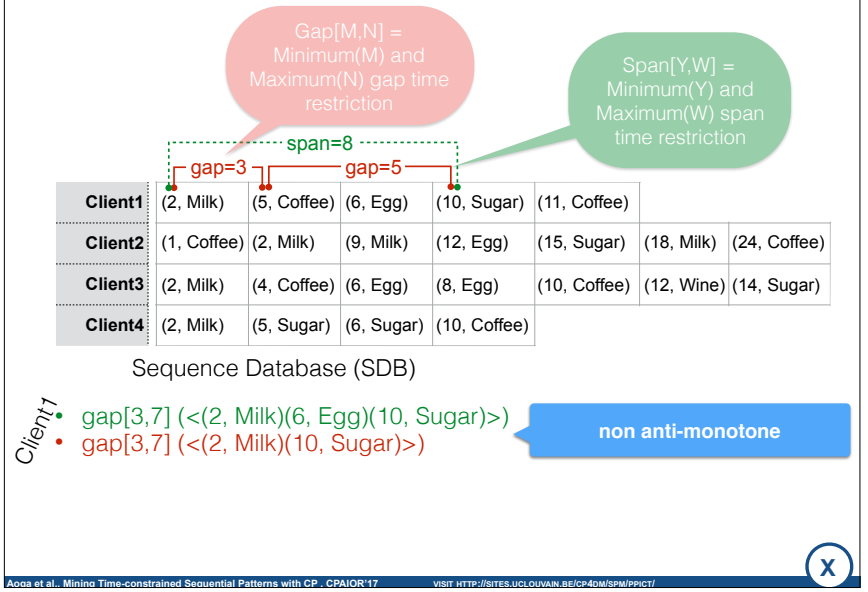


Client1	(2, Milk)	(5, Coffee)	(6, Egg)	(10, Sugar)	(11, Coffee)	
Client2	(1, Coffee)	(2, Milk)	(9, Milk)	(12, Egg)	(15, Sugar)	(18, Milk)	(24, Coffee)
Client3	(2, Milk)	(4, Coffee)	(6, Egg)	(8, Egg)	(10, Coffee)	(12, Wine)	(14, Sugar)
Client4	(2, Milk)	(5, Sugar)	(6, Sugar)	(10, Coffee)		

Sequence Database (SDB)

Client1
- gap[3,7] (<(2, Milk)(6, Egg)(10, Sugar)>)
- gap[3,7] (<(2, Milk)(10, Sugar)>)

TIME DATABASE CHALLENGES



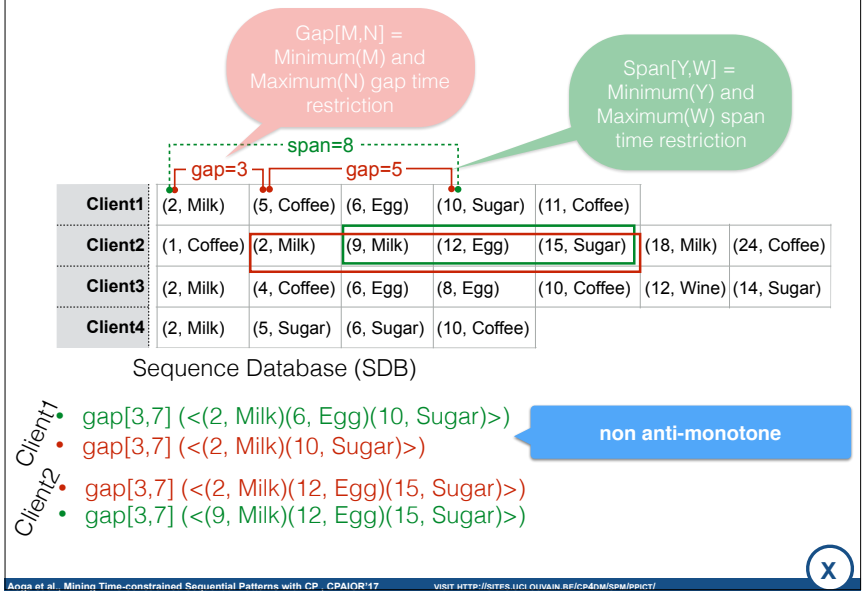
Client1	(2, Milk)	(5, Coffee)	(6, Egg)	(10, Sugar)	(11, Coffee)	
Client2	(1, Coffee)	(2, Milk)	(9, Milk)	(12, Egg)	(15, Sugar)	(18, Milk)	(24, Coffee)
Client3	(2, Milk)	(4, Coffee)	(6, Egg)	(8, Egg)	(10, Coffee)	(12, Wine)	(14, Sugar)
Client4	(2, Milk)	(5, Sugar)	(6, Sugar)	(10, Coffee)		

Sequence Database (SDB)

Client1
- gap[3,7] (<(2, Milk)(6, Egg)(10, Sugar)>)
- gap[3,7] (<(2, Milk)(10, Sugar)>)

non anti-monotone

TIME DATABASE CHALLENGES



Client1	(2, Milk)	(5, Coffee)	(6, Egg)	(10, Sugar)	(11, Coffee)	
Client2	(1, Coffee)	(2, Milk)	(9, Milk)	(12, Egg)	(15, Sugar)	(18, Milk)	(24, Coffee)
Client3	(2, Milk)	(4, Coffee)	(6, Egg)	(8, Egg)	(10, Coffee)	(12, Wine)	(14, Sugar)
Client4	(2, Milk)	(5, Sugar)	(6, Sugar)	(10, Coffee)		

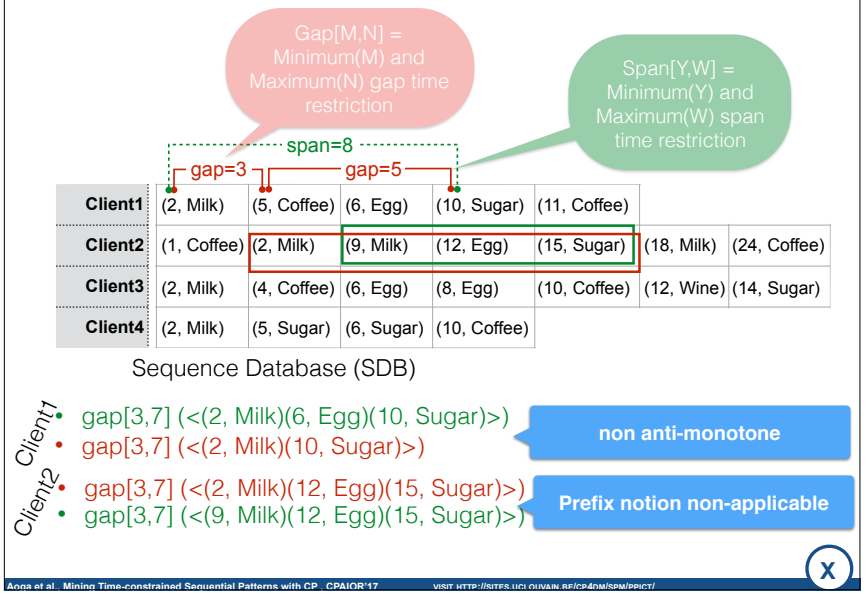
Sequence Database (SDB)

Client1
- gap[3,7] (<(2, Milk)(6, Egg)(10, Sugar)>)
- gap[3,7] (<(2, Milk)(10, Sugar)>)

non anti-monotone

Client2
- gap[3,7] (<(2, Milk)(12, Egg)(15, Sugar)>)
- gap[3,7] (<(9, Milk)(12, Egg)(15, Sugar)>)

TIME DATABASE CHALLENGES



Client1	(2, Milk)	(5, Coffee)	(6, Egg)	(10, Sugar)	(11, Coffee)	
Client2	(1, Coffee)	(2, Milk)	(9, Milk)	(12, Egg)	(15, Sugar)	(18, Milk)	(24, Coffee)
Client3	(2, Milk)	(4, Coffee)	(6, Egg)	(8, Egg)	(10, Coffee)	(12, Wine)	(14, Sugar)
Client4	(2, Milk)	(5, Sugar)	(6, Sugar)	(10, Coffee)		

Sequence Database (SDB)

Client1
- gap[3,7] (<(2, Milk)(6, Egg)(10, Sugar)>)
- gap[3,7] (<(2, Milk)(10, Sugar)>)

non anti-monotone

Client2
- gap[3,7] (<(2, Milk)(12, Egg)(15, Sugar)>)
- gap[3,7] (<(9, Milk)(12, Egg)(15, Sugar)>)

Prefix notion non-applicable

MinSup=3
Gap[3,7] t.u.

	1	2	3	4	5	6	7
1	2:**M**	5:C	6:E	10:S	11:C		
2	1:C	2:M	9:M	12:E	15:S	18:M	24:C
3	2:M	4:C	6:E	8:E	10:C	12:W	14:S
4	1:M	2:S	3:S	4:C			

Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di	€	€	€	€	€	€	€
	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
	Wine	Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (**start:end**)

MinSup=3
Gap[3,7] t.u.

	1	2	3	4	5	6	7
1	2:**M**	5:C	6:E	10:S	11:C		
2	1:C	2:M	9:M	12:E	15:S	18:M	24:C
3	2:M	4:C	6:E	8:E	10:C	12:W	14:S
4	1:M	2:S	3:S	4:C			

Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di	€	€	€	€	€	€	€
	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
	Wine	Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (**start:end**)

MinSup=3
Gap[3,7] t.u.

	1	2	3	4	5	6	7
1	2:**M**	5:C	6:E	10:S	11:C		
2	1:C	2:M	9:M	12:E	15:S	18:M	24:C
3	2:M	4:C	6:E	8:E	10:C	12:W	14:S
4	1:M	2:S	3:S	4:C			

Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di		€	€	€	€	€	€
	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
	Wine	Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (**start:end**)

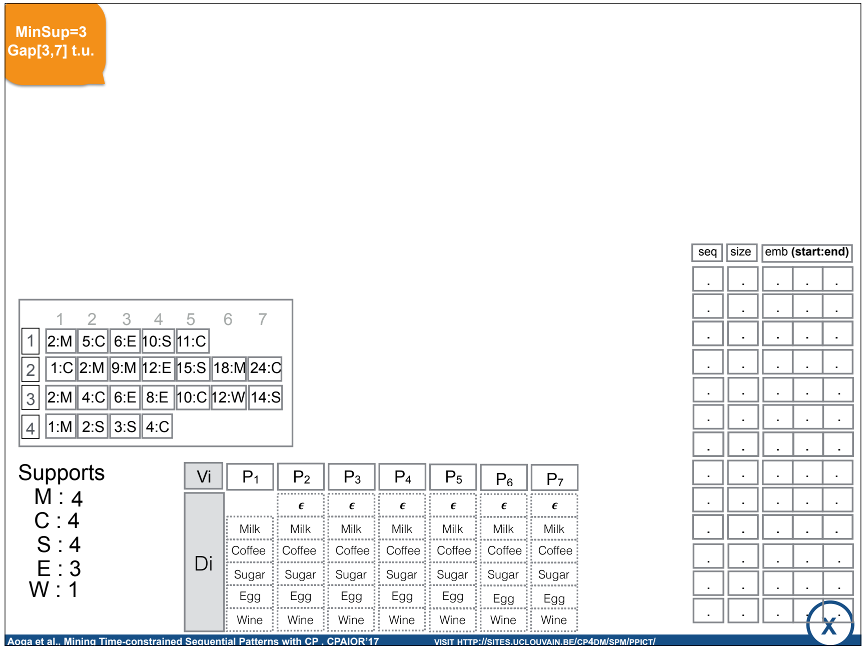
MinSup=3
Gap[3,7] t.u.

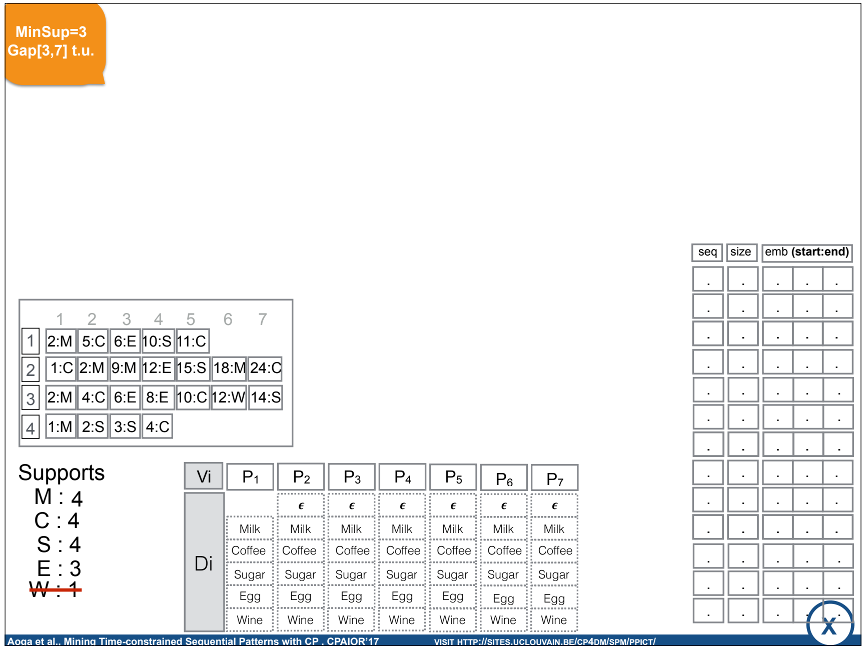
	1	2	3	4	5	6	7
1	2:**M**	5:C	6:E	10:S	11:C		
2	1:C	2:**M**	9:M	12:E	15:S	18:M	24:C
3	2:**M**	4:C	6:E	8:E	10:C	12:W	14:S
4	1:**M**	2:S	3:S	4:C			

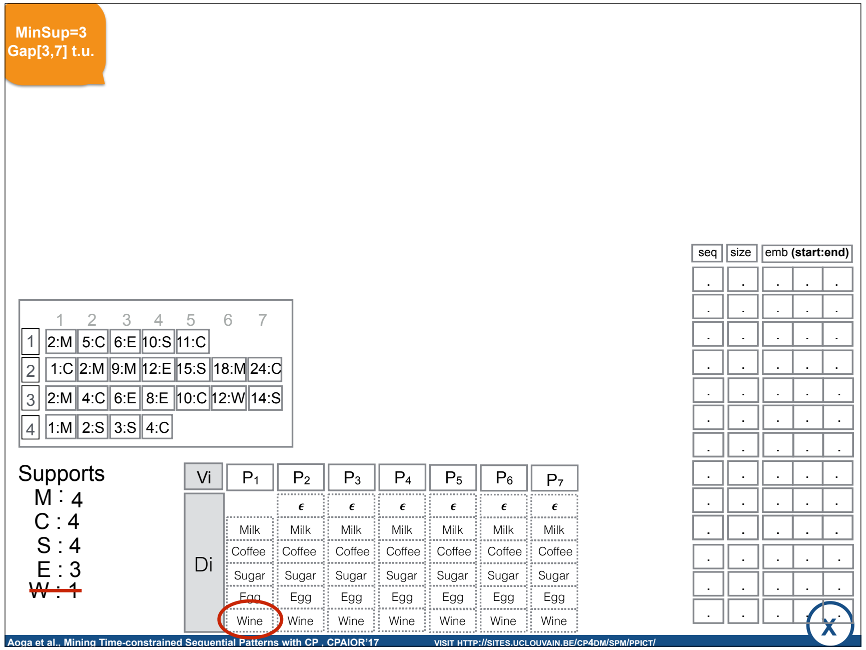
Supports
M :

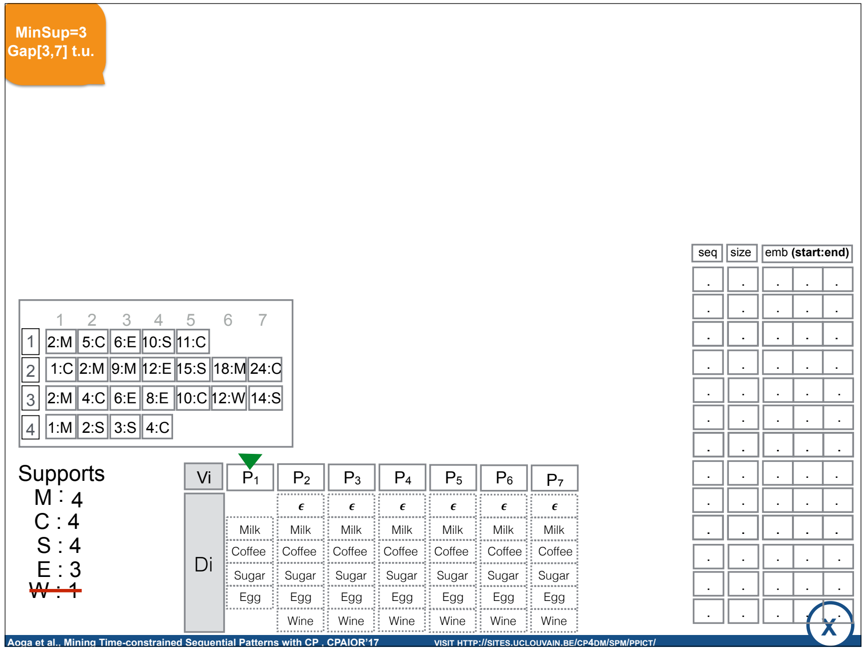
Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di		€	€	€	€	€	€
	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
	Wine	Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (**start:end**)


MinSup=3
Gap[3,7] t.u.
seq | size | emb (start:end)
1 2 3 4 5 6 7
1 | 2:M | 5:C | 6:E | 10:S | 11:C
2 | 1:C | 2:M | 9:M | 12:E | 15:S | 18:M | 24:C
3 | 2:M | 4:C | 6:E | 8:E | 10:C | 12:W | 14:S
4 | 1:M | 2:S | 3:S | 4:C
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Vi | P1 | P2 | P3 | P4 | P5 | P6 | P7
Di
P1: Milk Coffee Sugar Egg Wine
P2–P7: ϵ Milk Coffee Sugar Egg Wine
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
seq | size | emb (start:end)
1 2 3 4 5 6 7
1 | 2:M | 5:C | 6:E | 10:S | 11:C
2 | 1:C | 2:M | 9:M | 12:E | 15:S | 18:M | 24:C
3 | 2:M | 4:C | 6:E | 8:E | 10:C | 12:W | 14:S
4 | 1:M | 2:S | 3:S | 4:C
Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~
Vi | P1 | P2 | P3 | P4 | P5 | P6 | P7
Di
P1: Milk Coffee Sugar Egg Wine
P2–P7: ϵ Milk Coffee Sugar Egg Wine
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
seq | size | emb (start:end)
1 2 3 4 5 6 7
1 | 2:M | 5:C | 6:E | 10:S | 11:C
2 | 1:C | 2:M | 9:M | 12:E | 15:S | 18:M | 24:C
3 | 2:M | 4:C | 6:E | 8:E | 10:C | 12:W | 14:S
4 | 1:M | 2:S | 3:S | 4:C
Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~
Vi | P1 | P2 | P3 | P4 | P5 | P6 | P7
Di
P1: Milk Coffee Sugar Egg Wine
P2–P7: ϵ Milk Coffee Sugar Egg Wine
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
seq | size | emb (start:end)
1 2 3 4 5 6 7
1 | 2:M | 5:C | 6:E | 10:S | 11:C
2 | 1:C | 2:M | 9:M | 12:E | 15:S | 18:M | 24:C
3 | 2:M | 4:C | 6:E | 8:E | 10:C | 12:W | 14:S
4 | 1:M | 2:S | 3:S | 4:C
Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~
Vi | P1 | P2 | P3 | P4 | P5 | P6 | P7
Di
P1: Milk Coffee Sugar Egg
P2–P7: ϵ Milk Coffee Sugar Egg Wine
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

MinSup=3
Gap[3,7] t.u.

Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~

Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
		Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (start:end)

MinSup=3
Gap[3,7] t.u.

Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~

Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
		Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (start:end)

Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17 VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

MinSup=3
Gap[3,7] t.u.

Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~

Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
		Wine	Wine	Wine	Wine	Wine	Wine

seq	size	emb (start:end)

Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17 VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

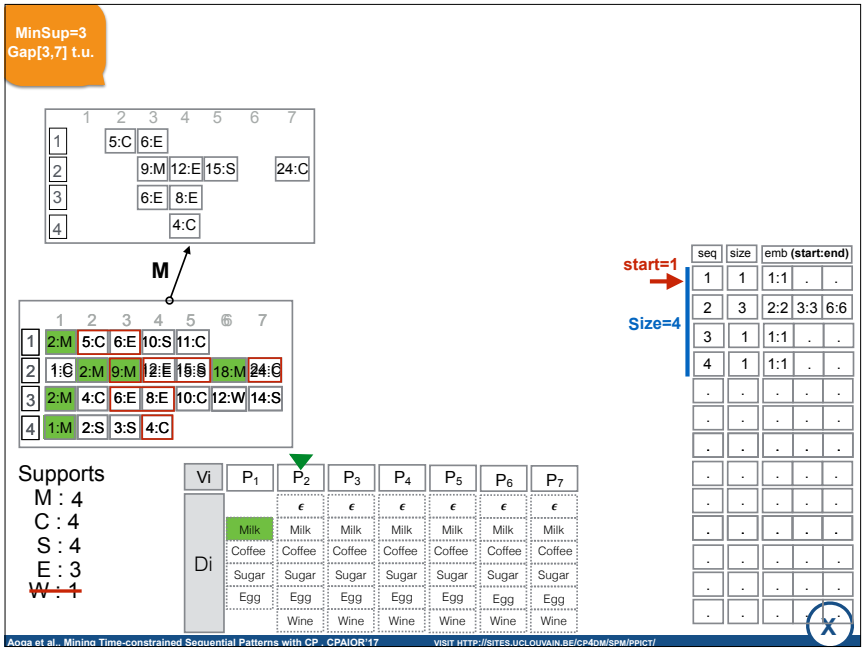
MinSup=3
Gap[3,7] t.u.

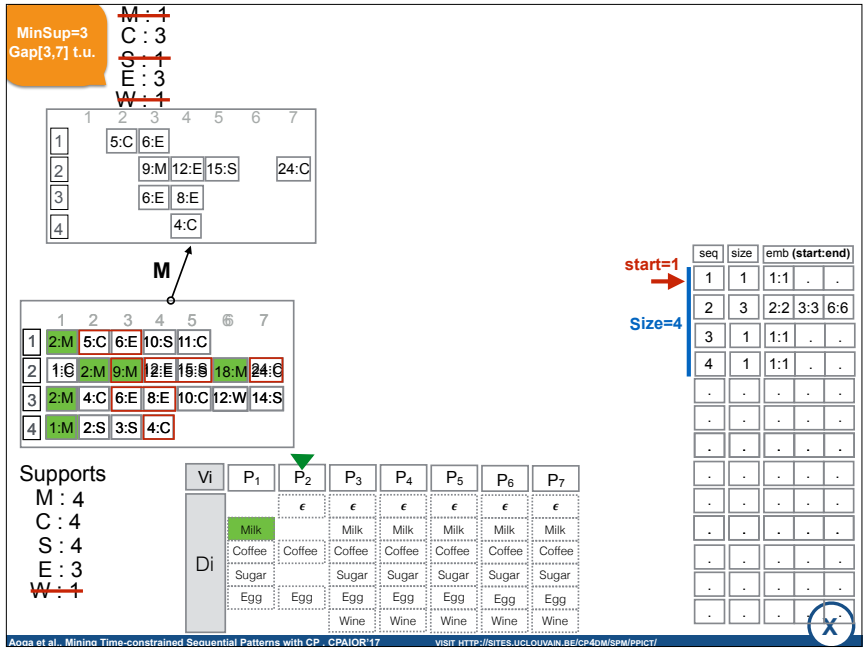
Supports
M : 4
C : 4
S : 4
E : 3
~~W : 1~~

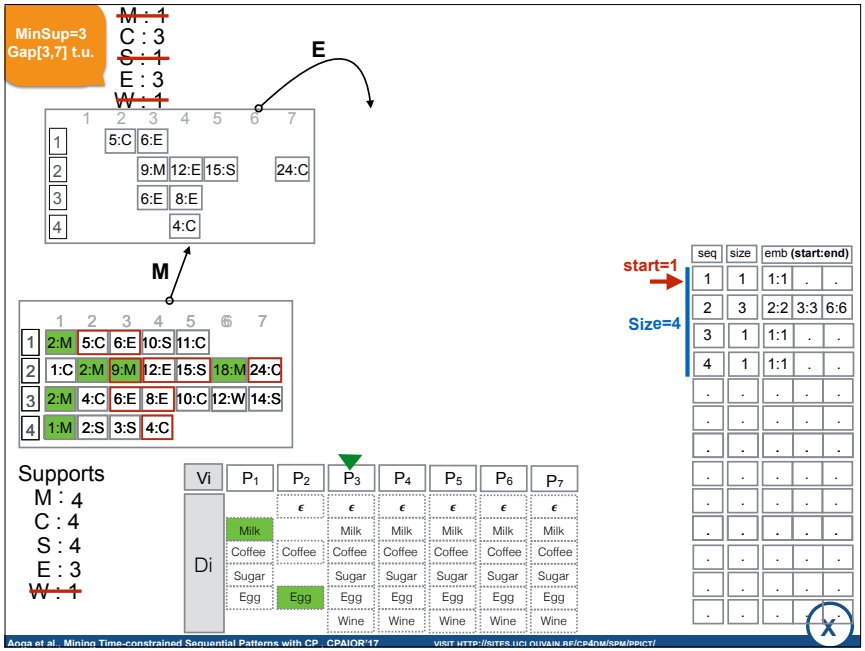
Vi	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Di	Milk	Milk	Milk	Milk	Milk	Milk	Milk
	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee	Coffee
	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar	Sugar
	Egg	Egg	Egg	Egg	Egg	Egg	Egg
		Wine	Wine	Wine	Wine	Wine	Wine

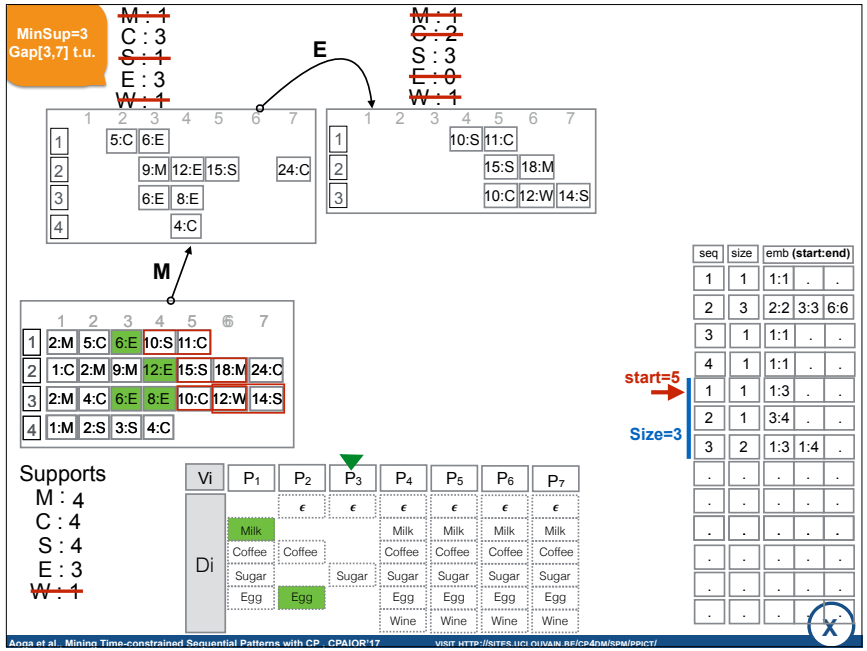
seq	size	emb (start:end)

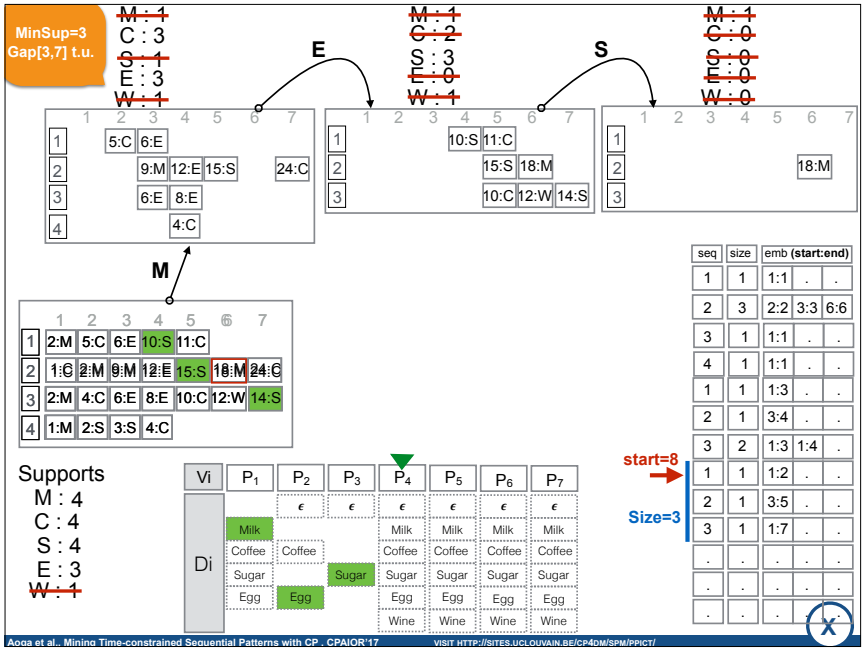
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17 VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

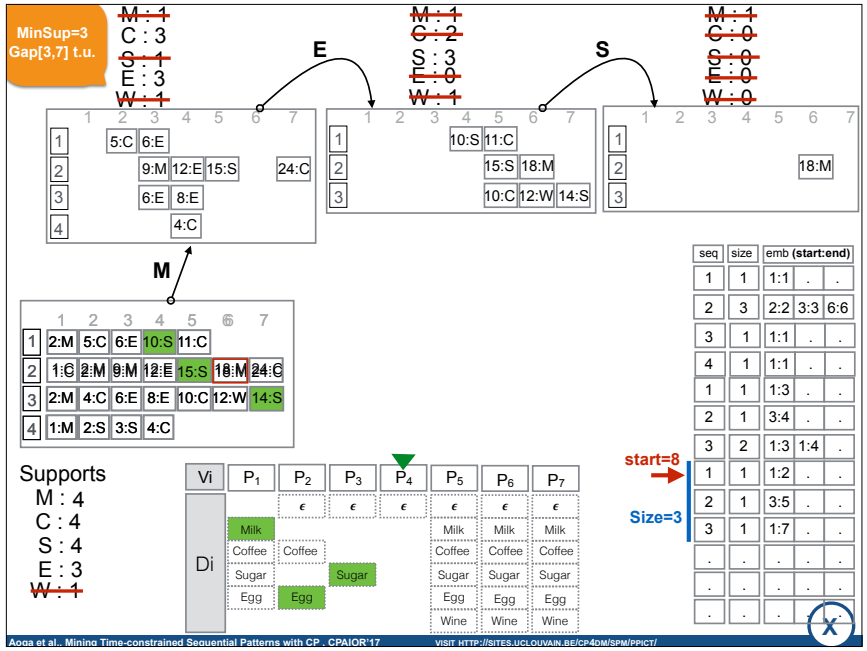

MinSup=3
Gap[3,7] t.u.
M
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
start=1
Size=4
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

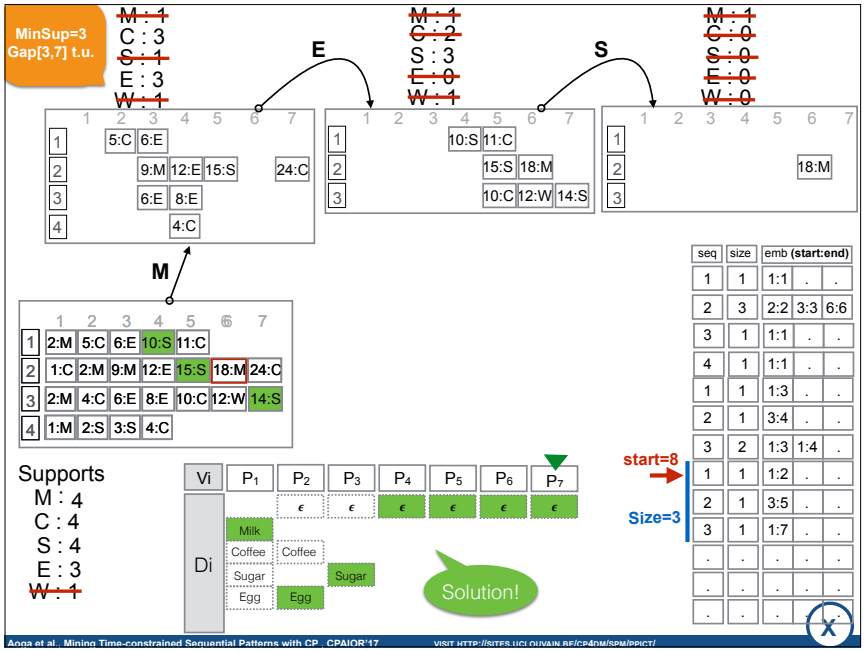

MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
M
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
start=1
Size=4
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

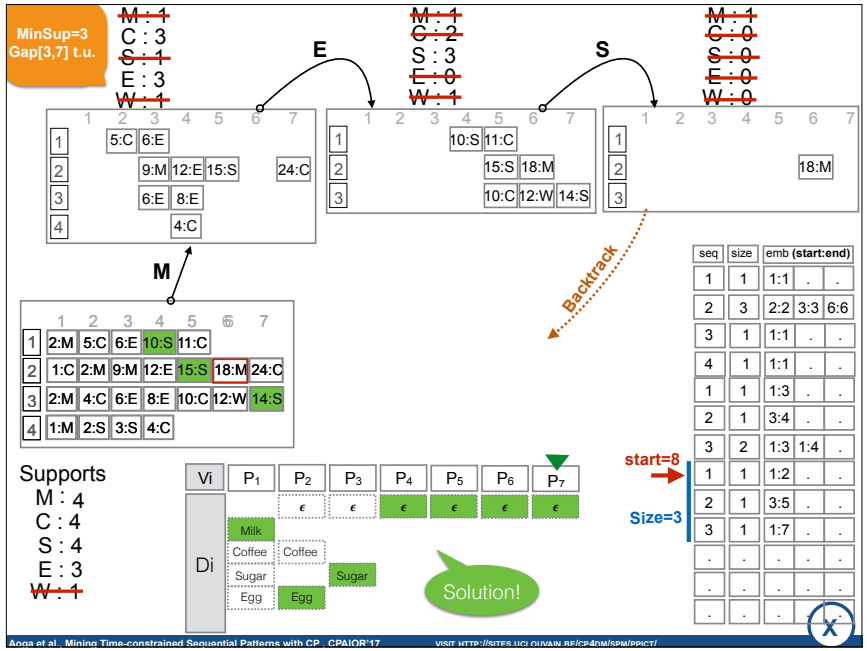

MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
E
M
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
start=1
Size=4
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

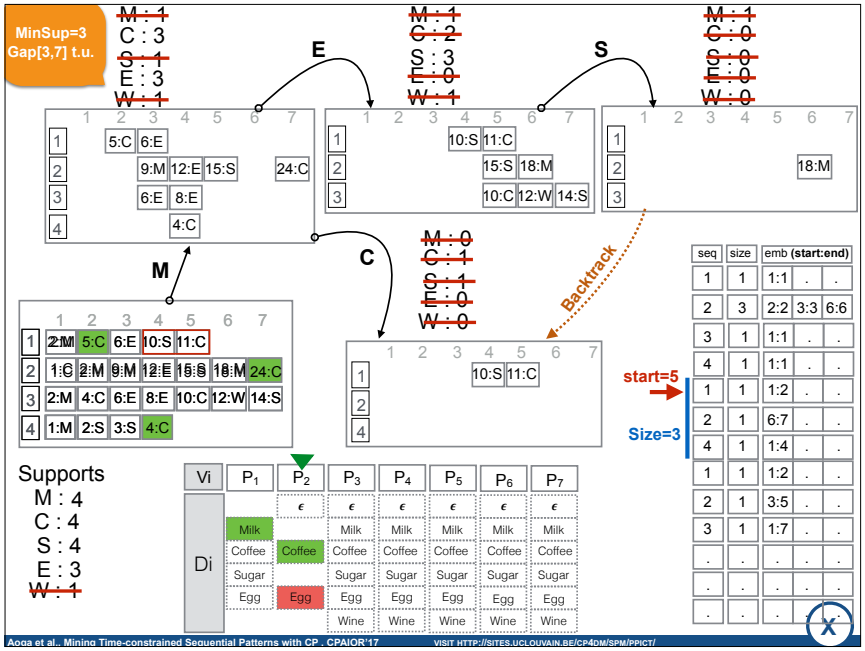

MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
E
M : 1
C : 2
S : 3
E : 0
W : 1
M
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
start=5
Size=3
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

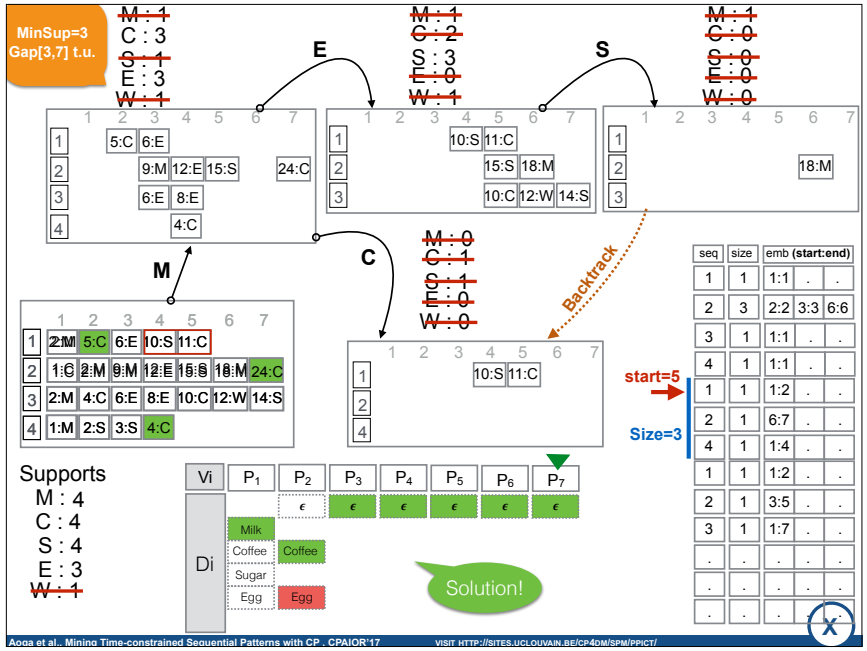

MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
E
M : 1
C : 2
S : 3
E : 0
W : 1
S
M : 1
C : 0
S : 0
E : 0
W : 0
5:C 6:E
9:M 12:E 15:S 24:C
6:E 8:E
4:C
10:S 11:C
15:S 18:M
10:C 12:W 14:S
18:M
M
2:M 5:C 6:E 10:S 11:C
1:C 2:M 9:M 12:E 15:S 18:M 24:C
2:M 4:C 6:E 8:E 10:C 12:W 14:S
1:M 2:S 3:S 4:C
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Vi
P1 P2 P3 P4 P5 P6 P7
Di
Milk
Coffee
Sugar
Egg
Wine
start=8
Size=3
seq size emb (start:end)
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
E
M : 1
C : 2
S : 3
E : 0
W : 1
S
M : 1
C : 0
S : 0
E : 0
W : 0
5:C 6:E
9:M 12:E 15:S 24:C
6:E 8:E
4:C
10:S 11:C
15:S 18:M
10:C 12:W 14:S
18:M
M
2:M 5:C 6:E 10:S 11:C
1:C 2:M 9:M 12:E 15:S 18:M 24:C
2:M 4:C 6:E 8:E 10:C 12:W 14:S
1:M 2:S 3:S 4:C
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Vi
P1 P2 P3 P4 P5 P6 P7
Di
Milk
Coffee
Sugar
Egg
Wine
start=8
Size=3
seq size emb (start:end)
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

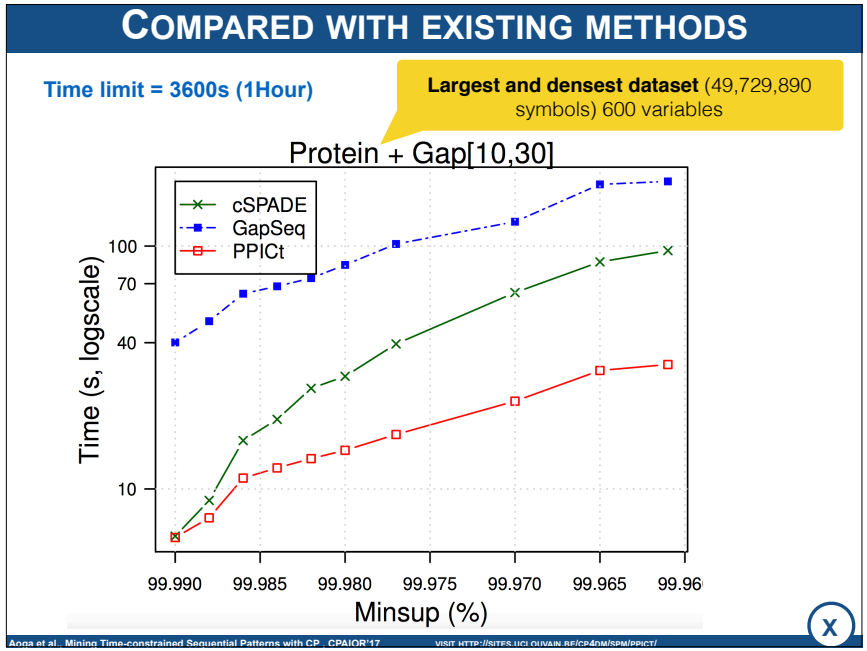

MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
E
M : 1
C : 2
S : 3
E : 0
W : 1
S
M : 1
C : 0
S : 0
E : 0
W : 0
5:C 6:E
9:M 12:E 15:S 24:C
6:E 8:E
4:C
10:S 11:C
15:S 18:M
10:C 12:W 14:S
18:M
M
2:M 5:C 6:E 10:S 11:C
1:C 2:M 9:M 12:E 15:S 18:M 24:C
2:M 4:C 6:E 8:E 10:C 12:W 14:S
1:M 2:S 3:S 4:C
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Vi
P1 P2 P3 P4 P5 P6 P7
Di
Milk
Coffee
Sugar
Egg
Solution!
start=8
Size=3
seq size emb (start:end)
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
M : 1
C : 3
S : 1
E : 3
W : 1
E
M : 1
C : 2
S : 3
E : 0
W : 1
S
M : 1
C : 0
S : 0
E : 0
W : 0
5:C 6:E
9:M 12:E 15:S 24:C
6:E 8:E
4:C
10:S 11:C
15:S 18:M
10:C 12:W 14:S
18:M
Backtrack
M
2:M 5:C 6:E 10:S 11:C
1:C 2:M 9:M 12:E 15:S 18:M 24:C
2:M 4:C 6:E 8:E 10:C 12:W 14:S
1:M 2:S 3:S 4:C
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Vi
P1 P2 P3 P4 P5 P6 P7
Di
Milk
Coffee
Sugar
Egg
Solution!
start=8
Size=3
seq size emb (start:end)
Aoga et al., Mining Time-constrained Sequential Patterns with CP . CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
E
S
M
C
Backtrack
start=5
Size=3
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Vi
Di
Milk
Coffee
Sugar
Egg
Wine
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/


MinSup=3
Gap[3,7] t.u.
E
S
M
C
Backtrack
start=5
Size=3
Supports
M : 4
C : 4
S : 4
E : 3
W : 1
Solution!
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

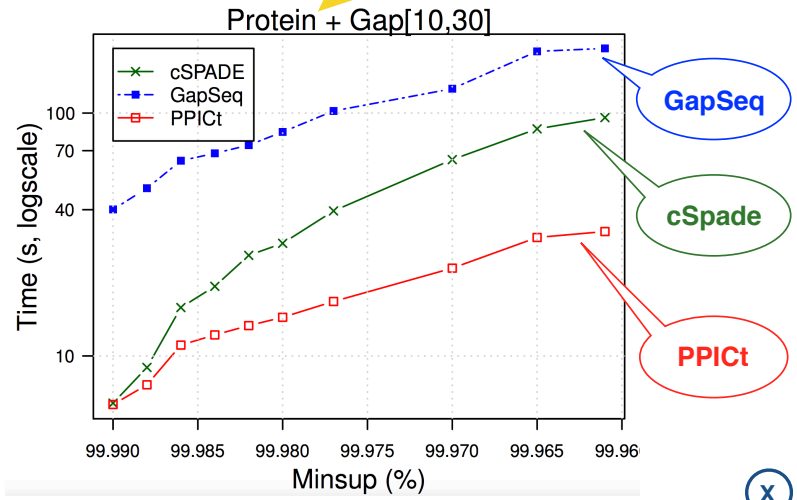

EXPERIMENTS
OSCAR
Scala
www.oscarlib.org
I'm cleverer than Obelix
I'm stronger than Asterix


COMPARED WITH EXISTING METHODS
Time limit = 3600s (1Hour)
Largest and densest dataset (49,729,890 symbols) 600 variables
Protein + Gap[10,30]
cSPADE
GapSeq
PPICt
Time (s, logscale)
100
70
40
10
99.990
99.985
99.980
99.975
99.970
99.965
99.96
Minsup (%)
Aoga et al., Mining Time-constrained Sequential Patterns with CP , CPAIOR'17
VISIT HTTP://SITES.UCLOUVAIN.BE/CP4DM/SPM/PPICT/

COMPARED WITH EXISTING METHODS

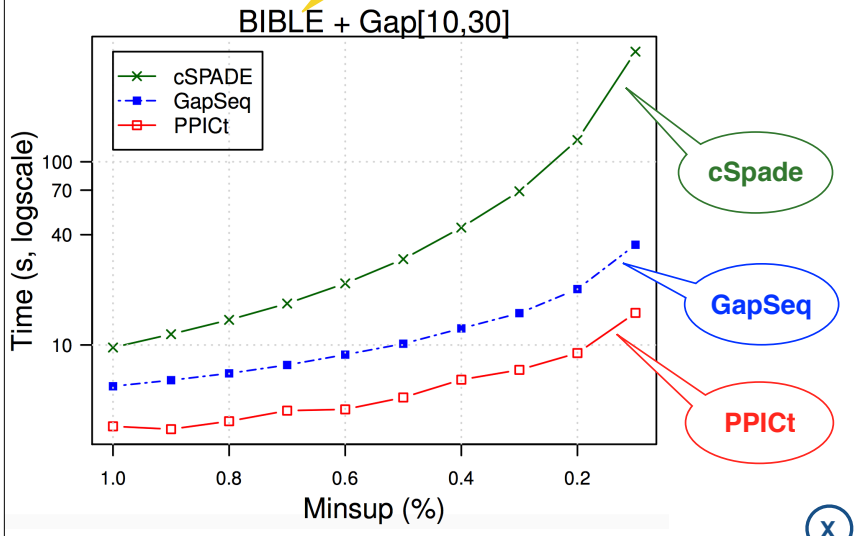
Time limit = 3600s (1Hour)

Largest and densest dataset (49,729,890 symbols) 600 variables



COMPARED WITH EXISTING METHODS

sparse dataset (787,066 symbols) 100 variables



Handling of different additional constraints

Methods	Frequency	Gap	Span	Regular/ Grammar	Among/ Gcc	Length
PPICt	x	x	x	x	x	x
GapSeq	x	x*			x	x
cSPADE	x	x	x**			x

Combining constraint over **Bible** dataset (13,905 symbols, 36,369 sequences)

+Gap

nSols : 32 307
Time(s) : 46

+Length+Gcc+Regular

nSols : 8
Time(s) : 0.19

Take-Away message

- Combining both SPM and CP techniques can lead to very efficient, modular and flexible approaches.
- Many kind of existing modules (in CP-Solvers) are reusable for free
- **Efficient memory using Trail-based backtracking aware data structure** really speed up search in DFSearch (not only for data mining)
- Code, data and apps are open http://sites.uclouvain.be/cp4dm/spm/


thank you!