



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH



---

## CONFORMANCE CHECKING IN THE LARGE: PARTITIONING AND TOPOLOGY

**Jorge Munoz-Gama, Josep Carmona and Wil M.P. van der Aalst**

# Motivation

*REALITY REFLECTION*



# Community Motivates Itself

**BPI Workshop 2013**

**BPM 2013**

**Process Mining**

**Decomposition**

**RPST**

**Conformance Dimensions**

**Alignment based Conformance**

**Conformance Checking Applications**

**Conformance Checking**

**Big Data**

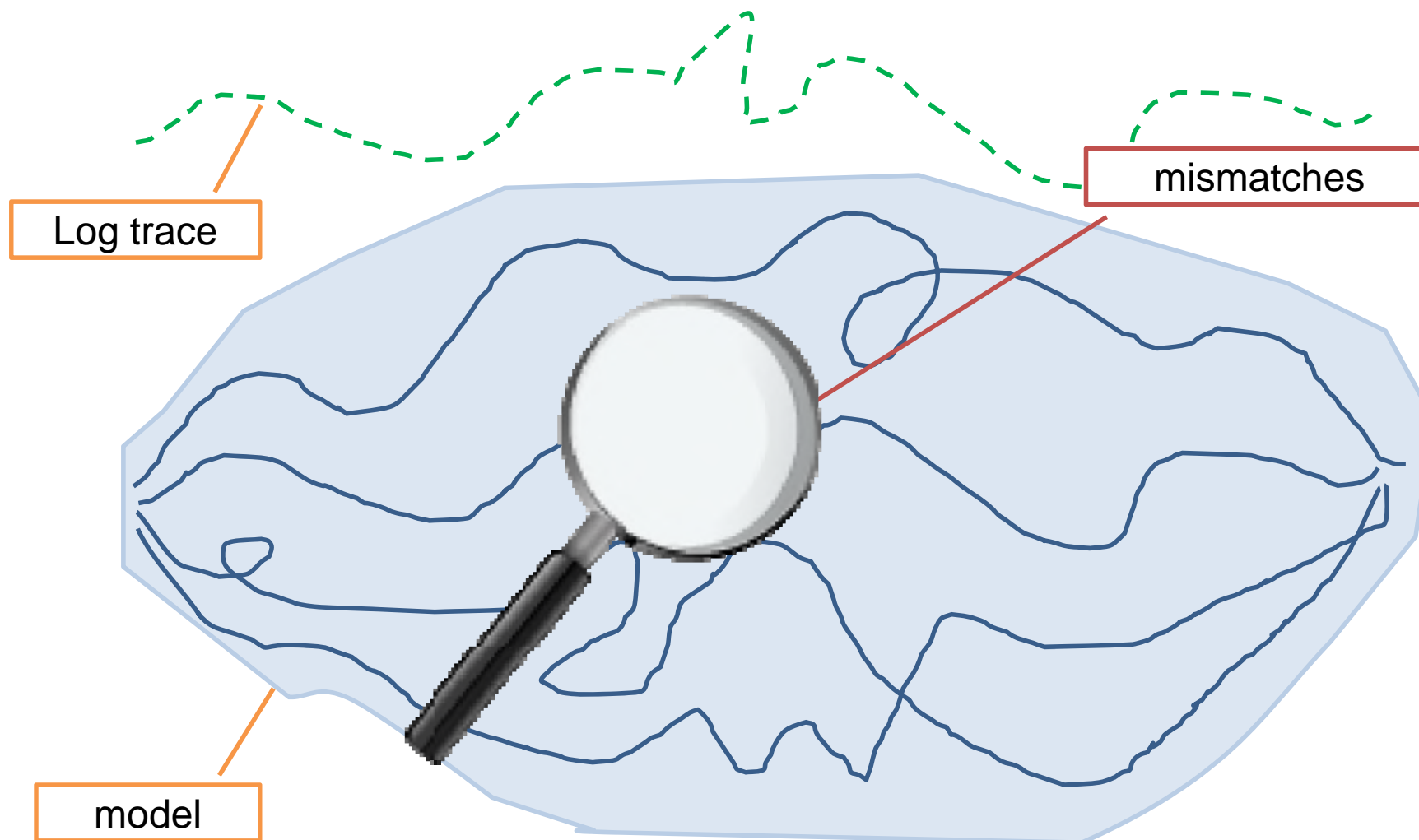
**BPM 2013 (Wednesday)**

26 AUGUST	
08:00 - 09:00	Opening
09:00 - 10:30	Keynote
10:30 - 11:30	Discrepancy and Multi-Step Calculation of Process Models using Multiple Query Databases
11:00 - 11:30	Discovery of Business Process Models from Event Logs
11:30 - 12:00	Discovery of Unusual Conditions in Event Logs
12:00 - 13:00	Lunch
13:00 - 13:30	Alignment of the Relationship between Process Models and Process Data
13:30 - 14:00	Process Mining: A Survey

27 August CONFERENCE TUESDAY	
CONFERENCE SESSIONS 1 - 3 GARDEN HALL	
08:30 - 09:00	Welcome Message from the General Chair, Message from the Program Committee Chair, Message from the Steering Committee Chair
09:00 - 10:00	KEYNOTE: <b>Business Process Mining: The New Frontier</b> (Chairman: Florin Danilă)
10:00 - 10:30	CONFERENCE BREAK
10:30 - 11:00	CONFERENCE SESSION 1: <b>Process Mining</b> Bridging Abstract Layers in Process Mining by Automated Matching of Events and Activities
11:00 - 11:30	Using Configurable Process Models from Collections of Eventlogs
11:30 - 12:00	Slice, Mine and Dice: Complexity aware Automated Discovery of Business Process Models
12:00 - 12:30	CONFERENCE SESSION 2: <b>Process Data</b> Verification of Query Completeness over Process Models and Extracting Complex Data Dependencies
12:30 - 13:00	Modeling Units in Business Processes
13:00 - 13:30	Industry Case Study: Methodology for Academic Dream and Career Pathways Completion
13:30 - 14:00	Industry Case Study: Creative Case Management Workflows
14:00 - 14:30	Industry Case Study: VisionWaves: Aligning Business Process Management and Performance Management to Achieve Business (Process) Excellence
14:30 - 15:00	Industry Case Study: iODIS - Discovering Performance Activity Affiliation
15:00 - 15:30	Industry Case Study: Navigating Knowledge from XDD-based Workflows

28 August CONFERENCE WEDNESDAY	
CONFERENCE SESSION 4 - 7 GARDEN HALL	
10:30 - 09:00	KEYNOTE: <b>Business Process Mining: The New Frontier</b> (Chairman: Florin Danilă)
CONFERENCE BREAK	
10:30 - 11:00	SCOPION CHAIR: <b>Chun Guzman</b>
11:00 - 11:30	Masahito de Leon and Willem Van der Aalst
11:30 - 12:00	Jorge Muñoz-Gama, Josep Domingo and Willem Van der Aalst
12:00 - 12:30	Laura Klapáček, Vlad Hrișcu, Florin Danilă and Petruș W. Golik

# Conformance General Idea



# Conformance in a Nutshell

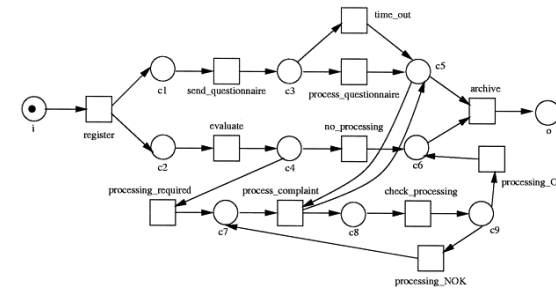
*Log*



**A B B C E**

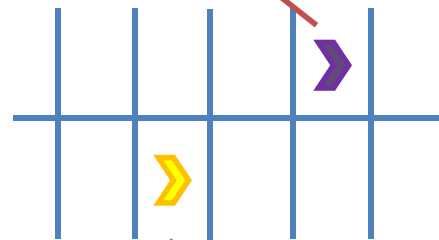
- Conformance mismatch on the Log

*Model*



**A B C D E**

*Alignment*



*Fitness*

How much behavior of the log is captured by the model?

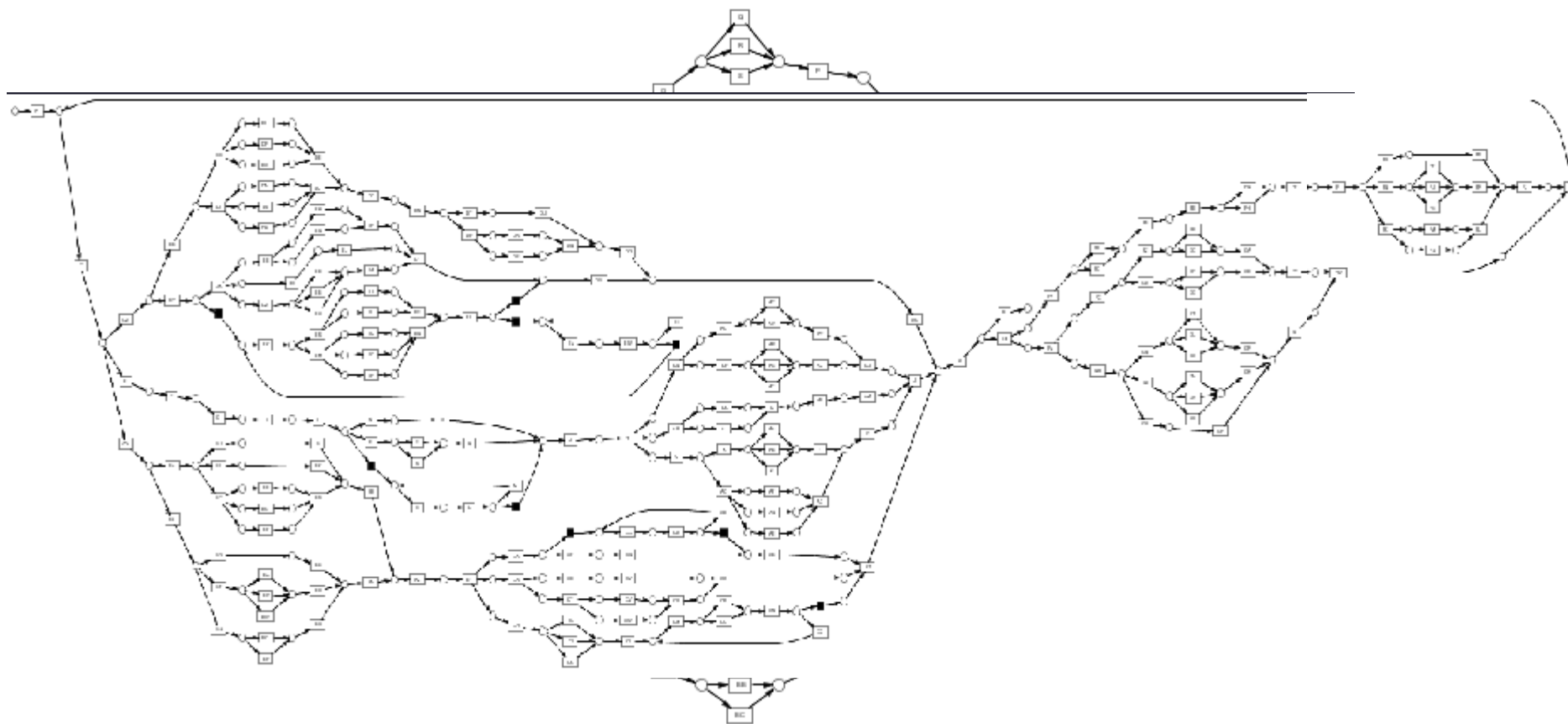
- Conformance mismatch on the Model

*Precision*

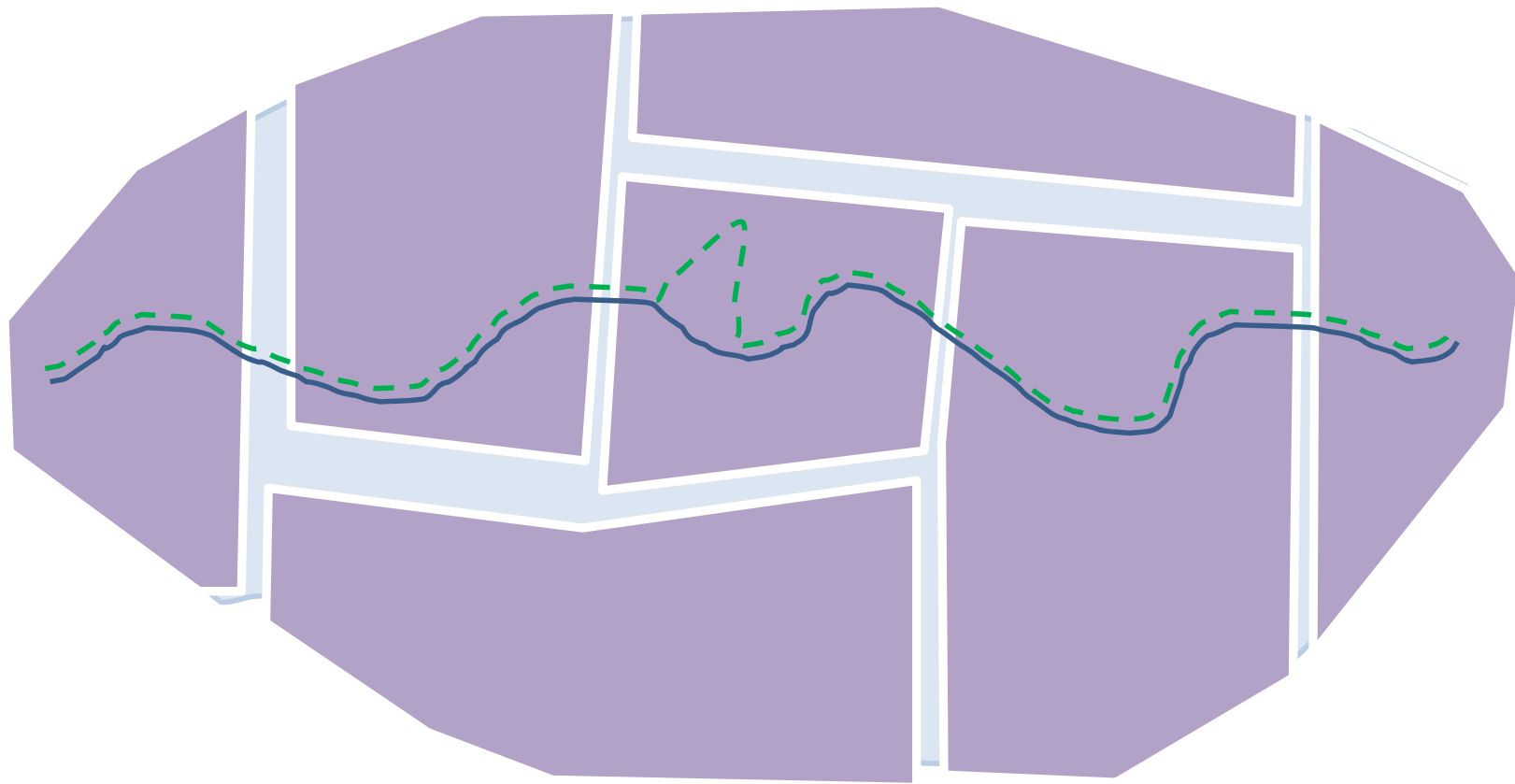
How accurate is the model describing the log?

# Conformance in the Large

- How easy is to diagnose a conformance problem here?
- How much time it takes?



# General Idea: Decomposition



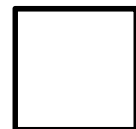
# The 4 Challenges



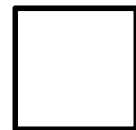
Comprehensible



Guaranties



Fast



Diagnosis



# SESE and RPST for decomposing

## SESE

Single Entry  
Single Exit  
components

## RPST

Refined Process  
Structure Tree

- Based on graph decomposition

\* Hopcroft, J., Tarjan, R.E.: **Dividing a graph into triconnected components.** *SIAM J. Comput.* 2(3), 1973

\* *Artem Polyvyanyy: Structuring Process Models. PhD Thesis. University of Potsdam (Germany), January 2012*

# Interior, Boundary, Entry, and Exit nodes

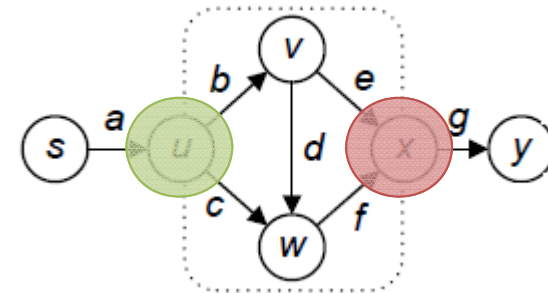
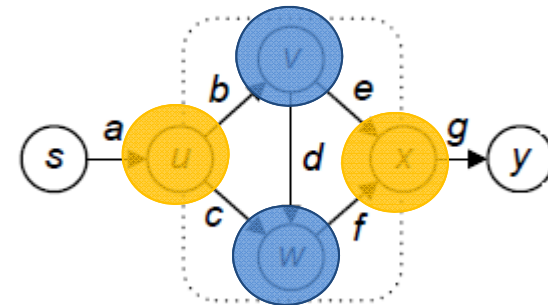
- Given a subgraph and a node of it:

- **Interior node**: connected only to nodes of the subgraph.

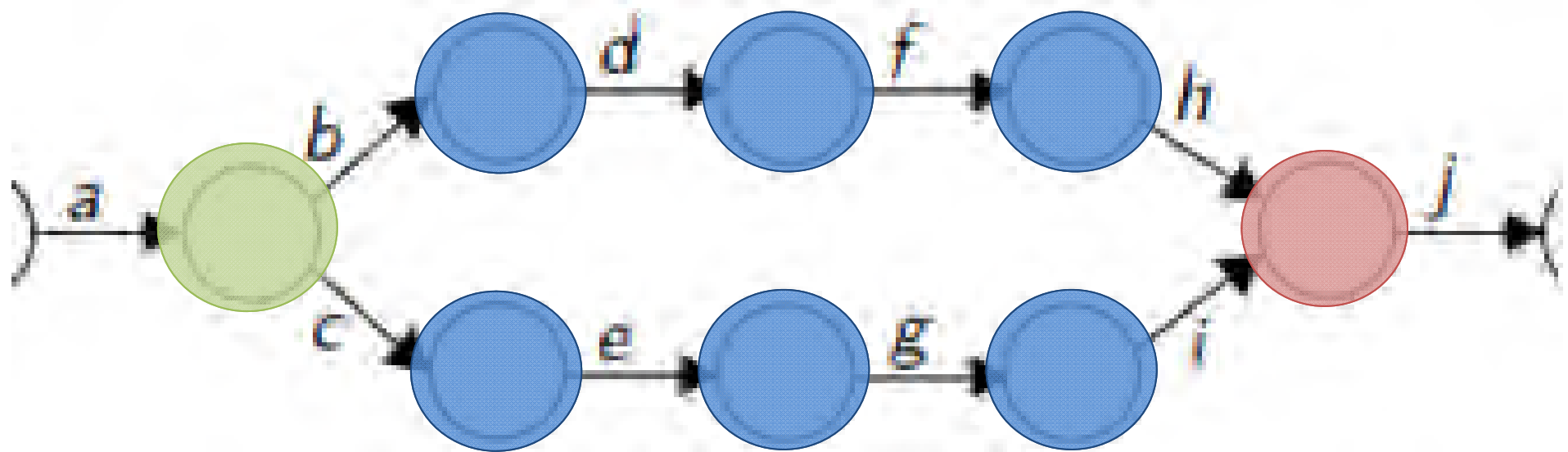
- **Boundary node**: not interior

- **Entry node**: boundary where
  - no incoming edge in subgraph
  - or all outgoing edges in

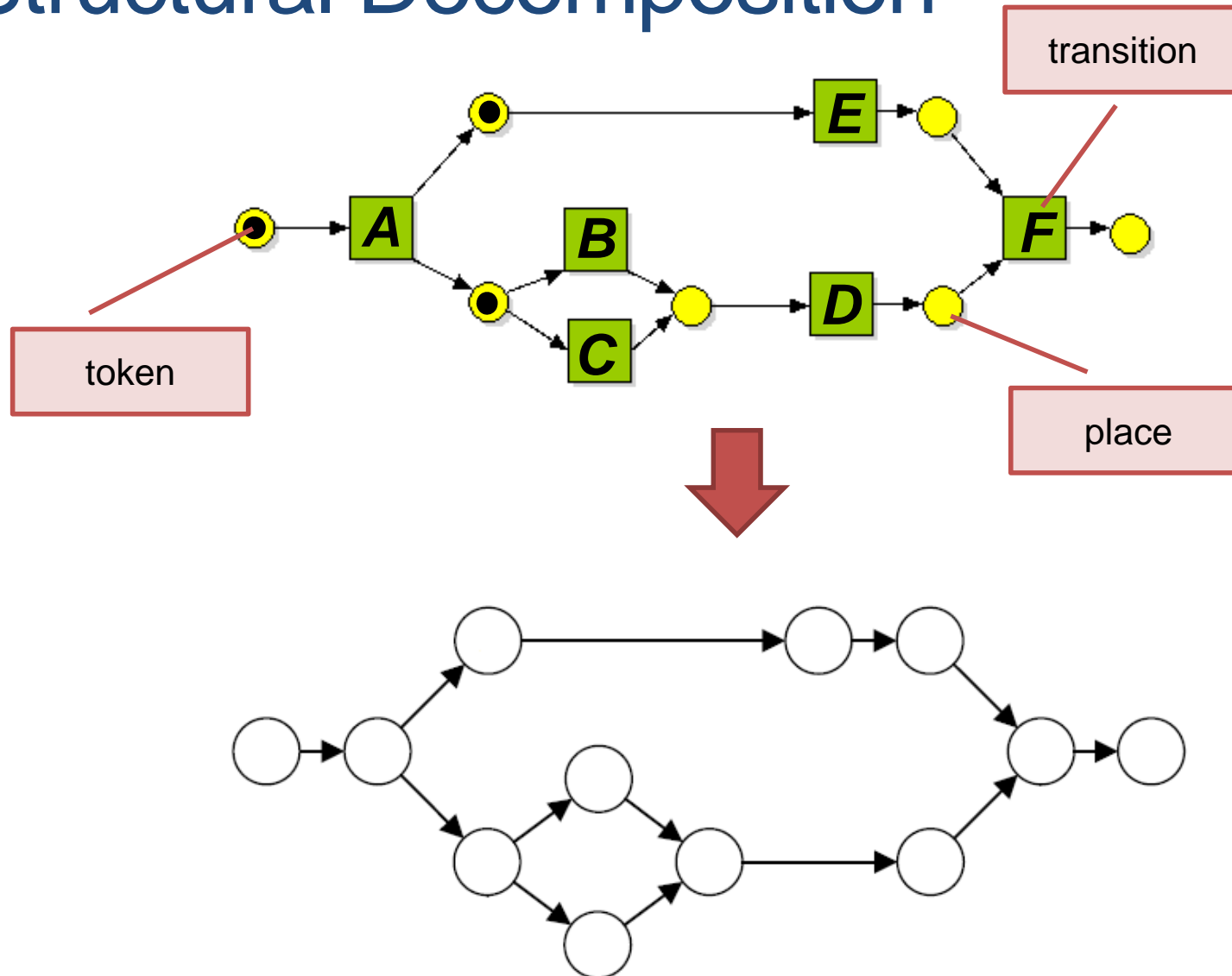
- **Exit node**: boundary where
  - no outgoing edge in subgraph
  - or all incoming edges in



# Nodes examples

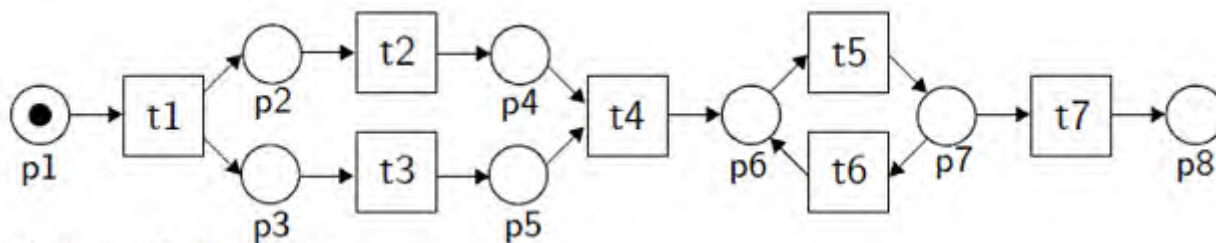


# Structural Decomposition

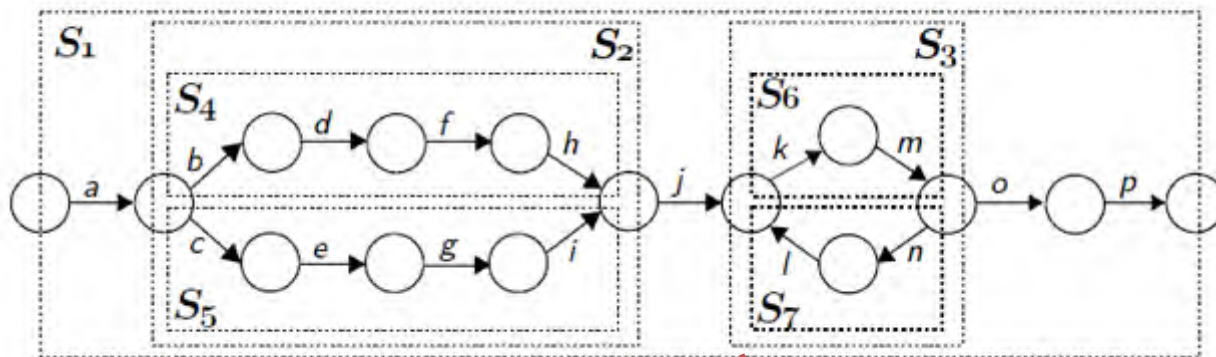


# Example of SESE and RPST

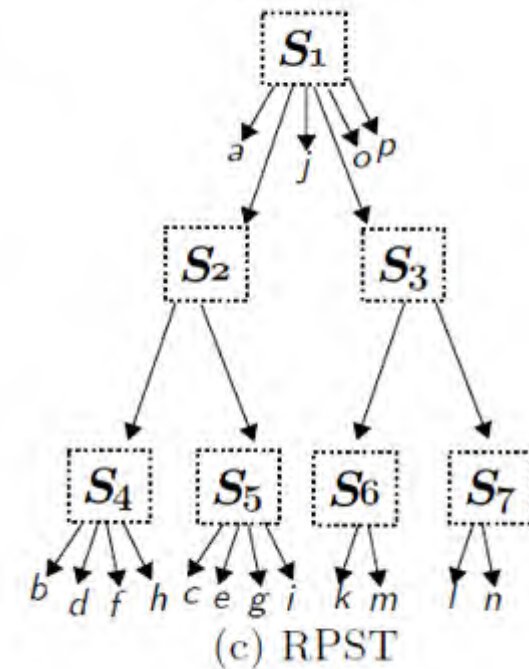
- Unique
- Modular
- Linear Time



(a) workflow net



(b) workflow graph and SESE decomposition



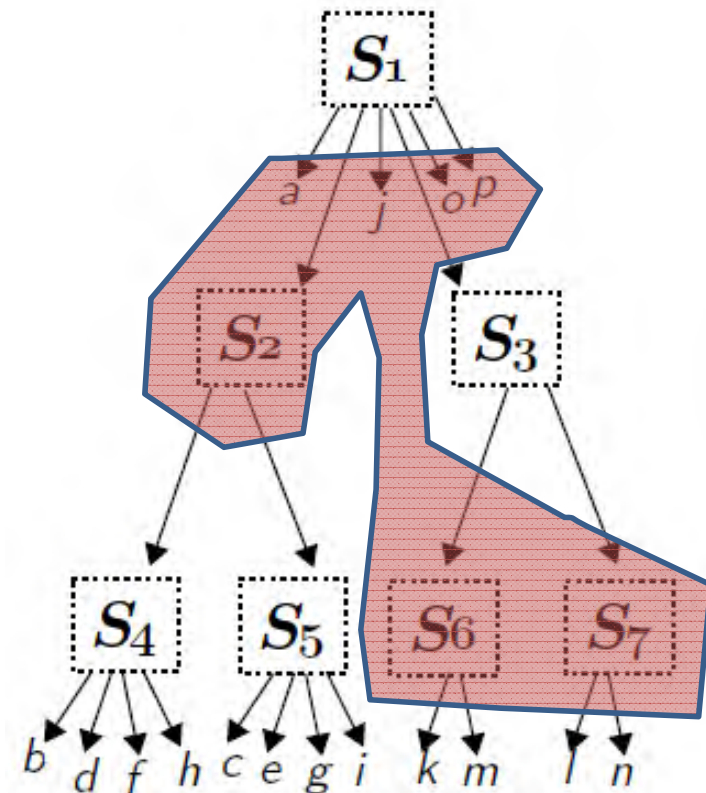
(c) RPST

SESE: set of edges which graph has a **S**ingle **E**nter node and a **S**ingle **E**xit node

Refined Process Structure Tree (RPST) containing non overlapping SESEs

# Why SESE and RPST?

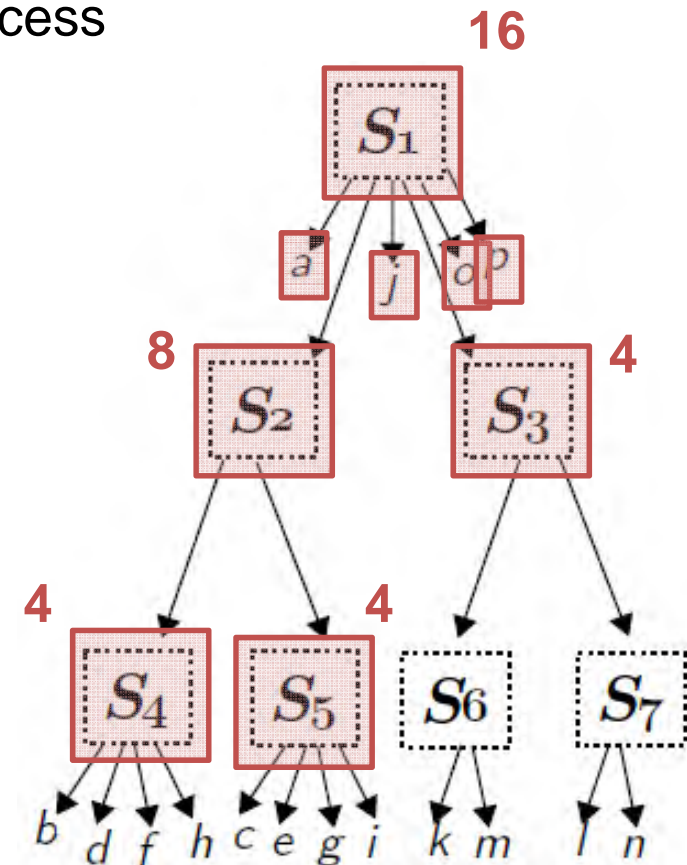
- Why SESE?
  - Only one entry; only one exit
  - Represent subprocesses within the process
  - Intuitive for conformance diagnosis
- Why RPST?
  - Partitioning over the RPST
  - Any cut is a partitioning
  - Algorithm to partitioning by size (k)



# Why SESE and RPST?

$K < 5$

- Why SESE?
  - Only one entry; only one exit
  - Represent subprocesses within the process
  - Intuitive for conformance diagnosis
- Why RPST?
  - Partitioning over the RPST
  - Any cut is a partitioning
  - Algorithm to partitioning by size (k)



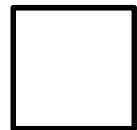
# The 4 Challenges



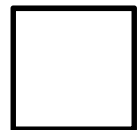
Comprehensible



Guaranties



Fast



Diagnosis



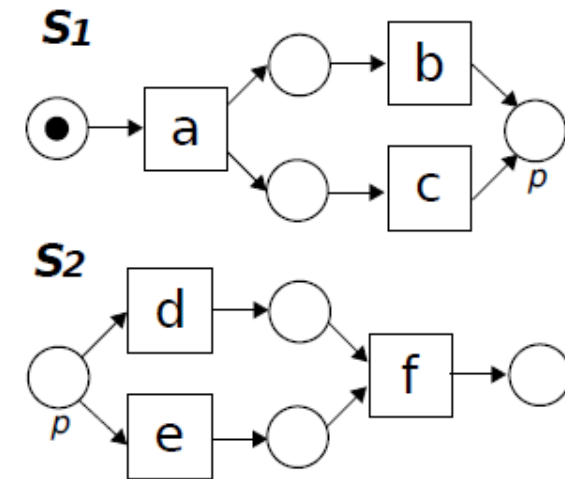
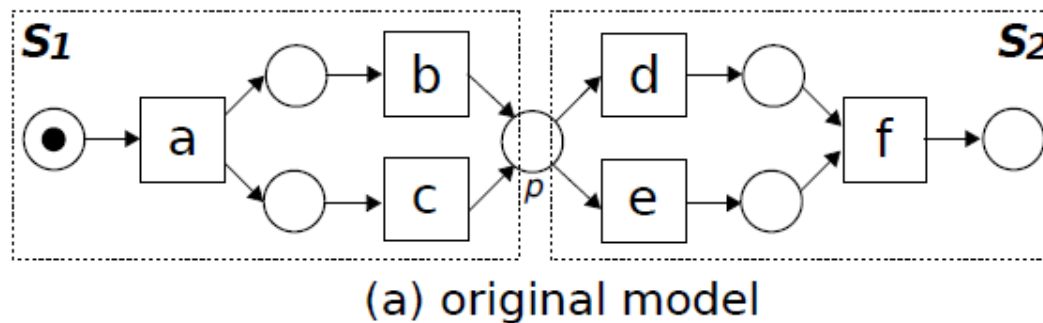
# Properties of the Partitioning

- What about the guaranties in conformance checking?
- *Decomposed Perfectly Fitting Checking*: A model/log is perfectly fitting if and only if all the components are perfectly fitting

\* *W.M.P. van der Aalst* : **Decomposing Petri nets for process mining: A generic approach**. *Distributed and Parallel Databases*, 2013

# SESE and Decomposed Perfectly Fitting

- SESEs (per se) do not satisfy the *Decomposed Perfectly Fitting Checking* property



- 1 token in  $p$   $\Rightarrow$   $abcdef$  fits **S** but not **S2**
- 2 tokens in  $p$   $\Rightarrow$   $abdecf$  fits **S1** and **S2** but not **S**

# Valid Decomposition

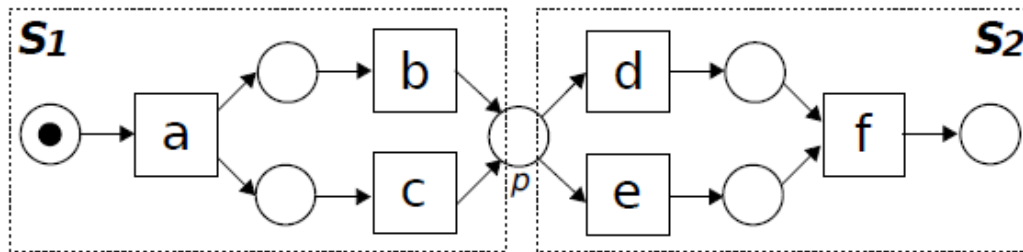
- The problem is in the boundary places
  - No reflection on the log
- A partition with only transitions shared among components (no places neither arcs)
  - Transitions have reflect on the log
  - Use that reflection to sync the components
  - This is known as a *valid decomposition*
- *Details in:*

\* *W.M.P. van der Aalst* : **Decomposing Petri nets for process mining: A generic approach.** *Distributed and Parallel Databases*, 2013

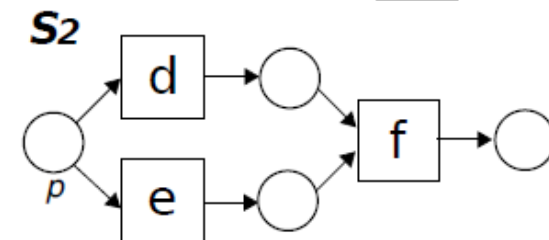
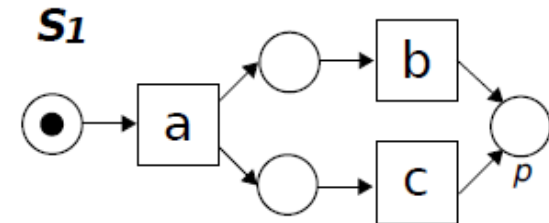
\* *J. Munoz-Gama, J. Carmona, and W.M.P. van der Aalst* : **Conformance checking in the large: partitioning and topology.** *BPM 2013*

# SESE to Valid Decomposition

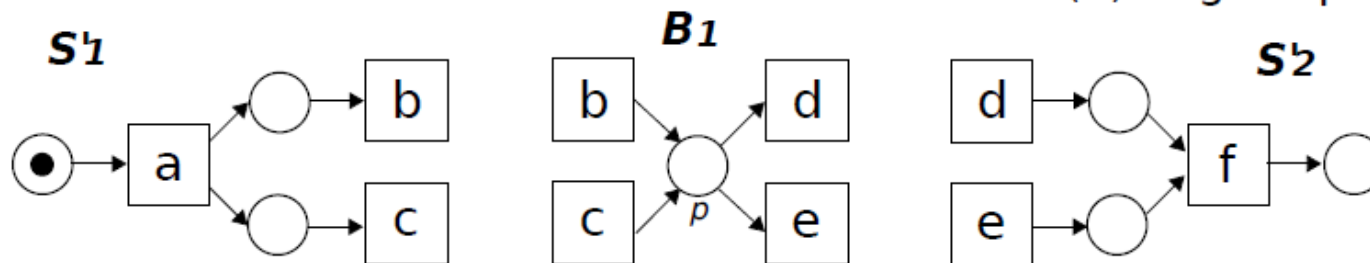
- Create a 'bridge' for each shared place



(a) original model



(b) original partitioning



(c) extended partitioning

# Results

Few with non negligible size

			[9,11]		k = 50			k = 100			k = 200					
	P	T	f	t	S/B	>5	nf	t	S/B	>5	nf	t	S/B	>5	nf	t
prAm6	347	363	0.92	75	129/57	29	7(3%)	423	62/27	14	1(9%)	323	27/12	7	1(10%)	180
prBm6	317	317	1	88	93/38	22	0(0%)	608	66/29	14	0(0%)	318	36/16	8	0(0%)	114
prCm6	317	317	0.57	2743	93/38	22	58(92%)	189	66/29	14	41(94%)	185	36/16	8	22(96%)	502
prDm6	529	429	-	-	105/34	33	5(8%)	1386	60/23	18	4(14%)	986	33/15	9	4(23%)	1284
prEm6	277	275	0.97	3566	82/35	20	2(5%)	529	35/13	11	2(5%)	343	15/7	5	2(6%)	211
prFm6	362	299	-	-	108/43	28	2(6%)	1667	57/23	15	2(21%)	863	21/9	5	1(23%)	562
prGm6	357	335	-	-	94/37	25	2(8%)	867	67/31	15	2(8%)	850	51/25	11	2(8%)	474

SESEs and Bridges

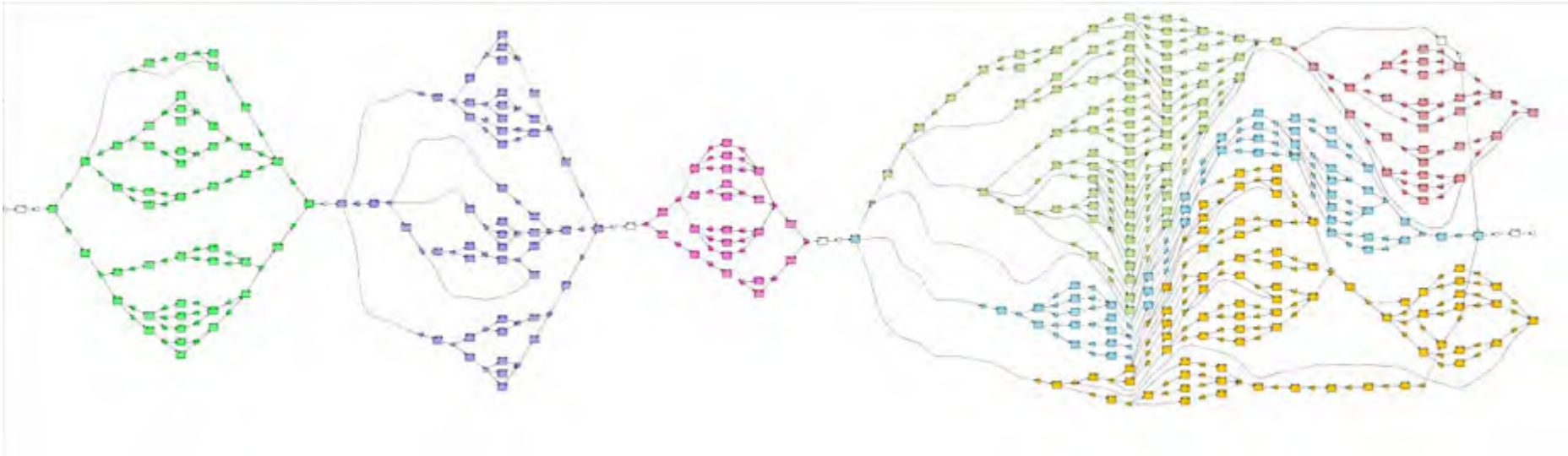
# SESE + Bridging Theorem

- **Theorem:** SESE decomposition with Bridging post-processing satisfies the *Decomposed Perfectly Fitting Checking*

# The 4 Challenges

- Comprehensible
- Guaranties
- ? Fast
- Diagnosis

# Results



- 1 Net – 1h 15min
- 7 Subnets – 2min



# Results

Not always faster: short traces, fitting.  
Overhead of the decomposition

	<i>P</i>	<i>T</i>	[9,11]		<i>k</i> = 50				<i>k</i> = 100				<i>k</i> = 200			
			<i>f</i>	<i>t</i>	<i>S/B</i>	>5	nf	<i>t</i>	<i>S/B</i>	>5	nf	<i>t</i>	<i>S/B</i>	>5	nf	<i>t</i>
prAm6	347	363	0.92	75	129/57	29	7(3%)	423	62/27	14	1(9%)	323	27/12	7	1(10%)	180
prBm6	317	317	1	88	93/38	22	0(0%)	608	66/29	14	0(0%)	318	36/16	8	0(0%)	114
prCm6	317	317	0.57	2743	93/38	22	58(92%)	189	66/29	14	41(94%)	185	36/16	8	22(96%)	502
prDm6	529	429	-	-	105/34	33	5(8%)	1386	60/23	18	4(14%)	986	33/15	9	4(23%)	1284
prEm6	277	275	0.97	3566	82/35	20	2(5%)	529	35/13	11	2(5%)	343	15/7	5	2(6%)	211
prFm6	362	299	-	-	108/43	28	2(6%)	1667	57/23	15	2(21%)	863	21/9	5	1(23%)	562
prGm6	357	335	-	-	94/37	25	2(8%)	867	67/31	15	2(8%)	850	51/25	11	2(8%)	474

Results even in cases that the original  
approach can handle

**Better performance with large, concurrency, long traces and  
concentrated conformance problems**

# The 4 Challenges

- Comprehensible
- Guaranties
- Fast
- ? Diagnosis

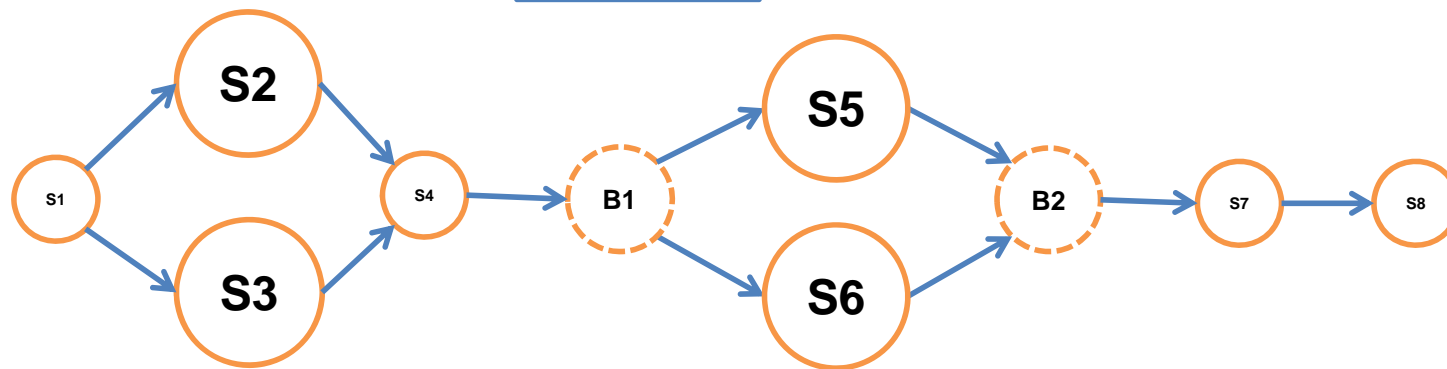
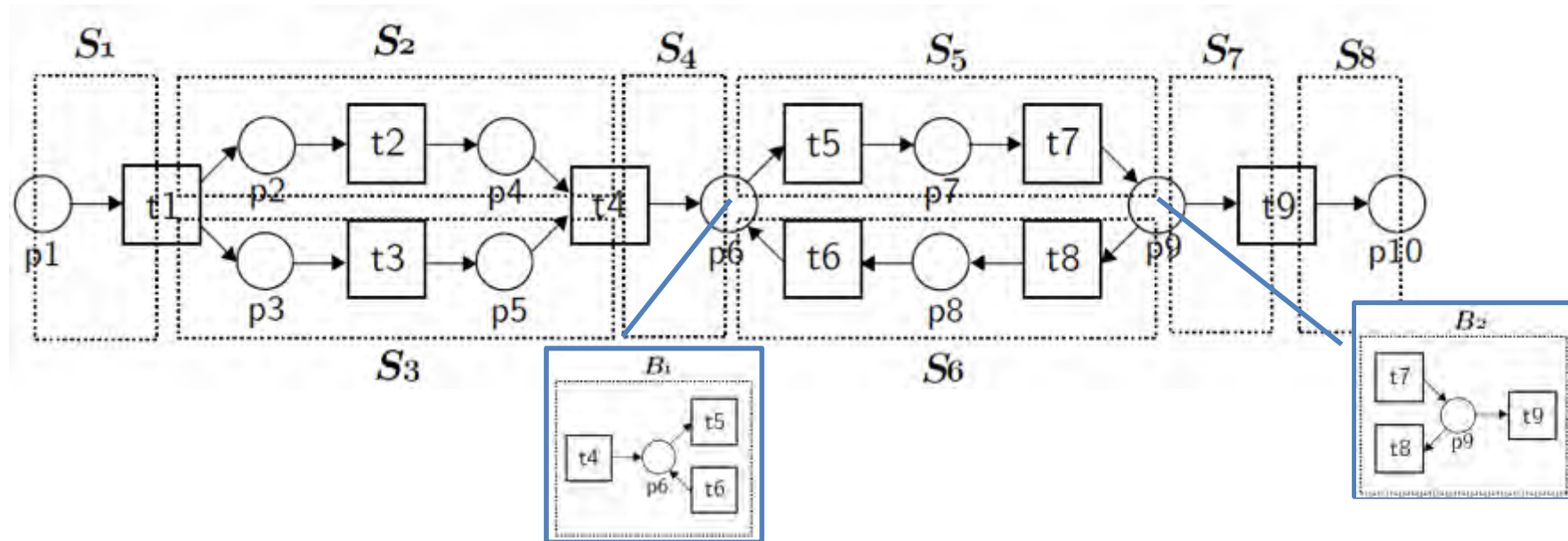
# Results

Conformance problems spread

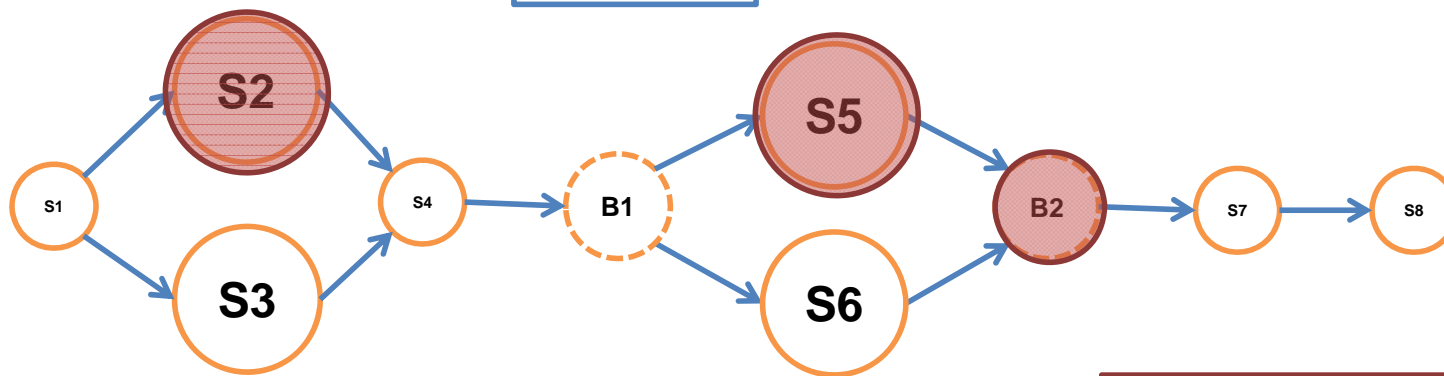
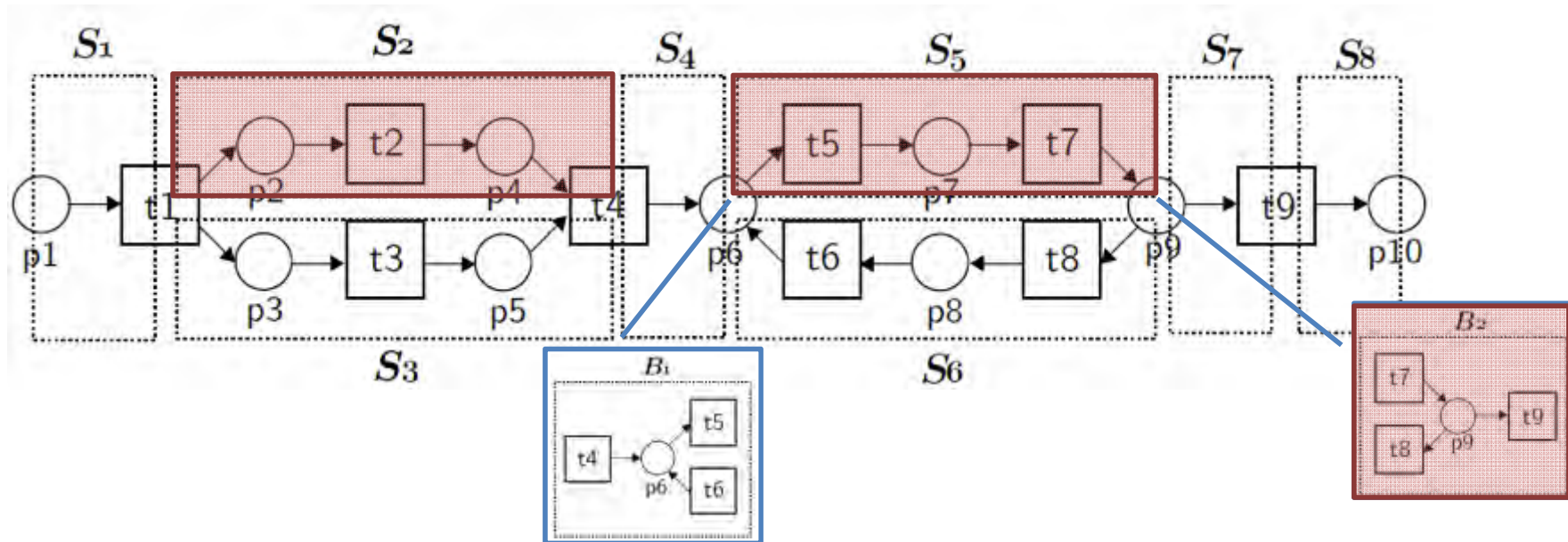
	<i>P</i>	<i>T</i>	<i>[9,11]</i>		<i>k = 50</i>			<i>k = 100</i>			<i>k = 200</i>					
			<i>f</i>	<i>t</i>	<i>S/B</i>	>5	nf	<i>t</i>	<i>S/B</i>	>5	nf	<i>t</i>	<i>S/B</i>	>5	nf	<i>t</i>
prAm6	347	363	0.92	75	129/57	29	7(3%)	423	62/27	14	1(9%)	323	27/12	7	1(10%)	180
prBm6	317	317	1	88	93/38	22	0(0%)	608	66/29	14	0(0%)	318	36/16	8	0(0%)	114
prCm6	317	317	0.57	2743	93/38	22	58(92%)	189	66/29	14	41(94%)	185	36/16	8	22(96%)	502
prDm6	529	429	-	-	105/34	33	5(8%)	1386	60/23	18	4(14%)	986	33/15	9	4(23%)	1284
prEm6	277	275	0.97	3566	82/35	20	2(5%)	529	35/13	11	2(5%)	343	15/7	5	2(6%)	211
prFm6	362	299	-	-	108/43	28	2(6%)	1667	57/23	15	2(21%)	863	21/9	5	1(23%)	562
prGm6	357	335	-	-	94/37	25	2(8%)	867	67/31	15	2(8%)	850	51/25	11	2(8%)	474

Conformance problems concentrated

# Topology

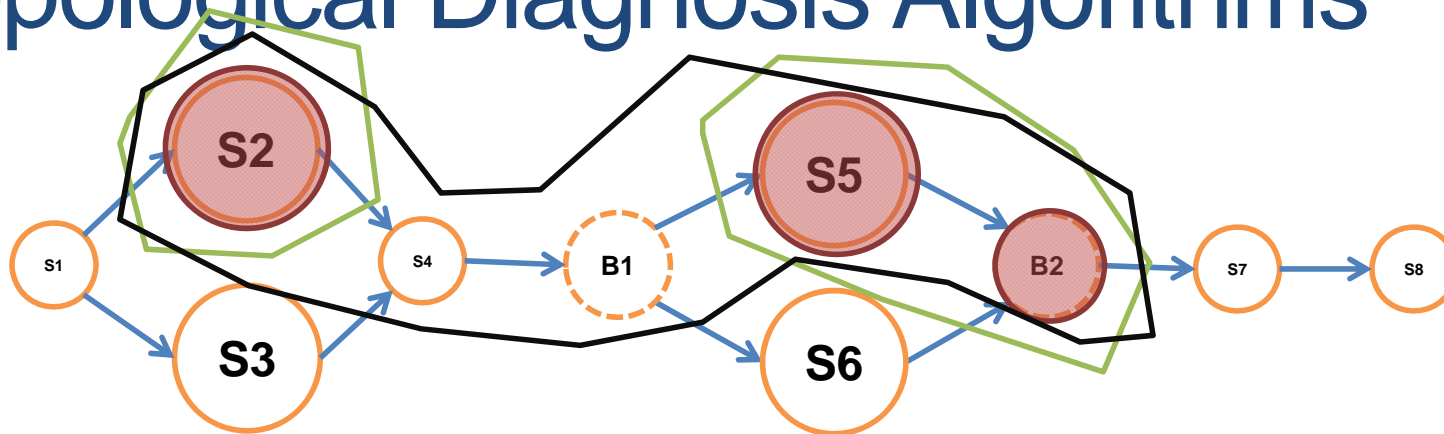


# Topology Enhancement

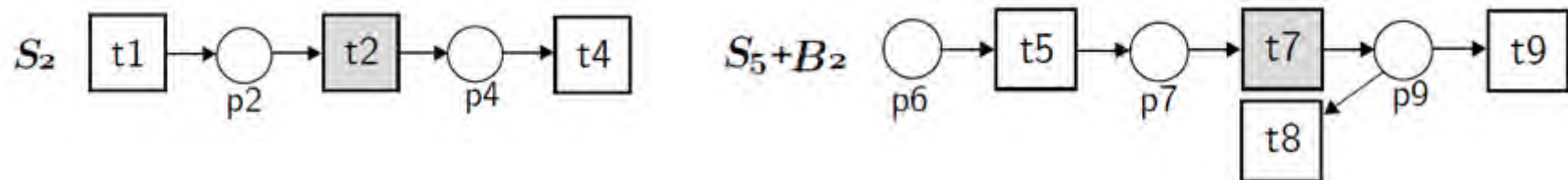


t1,t3,t4,t5,t7,t7,t9

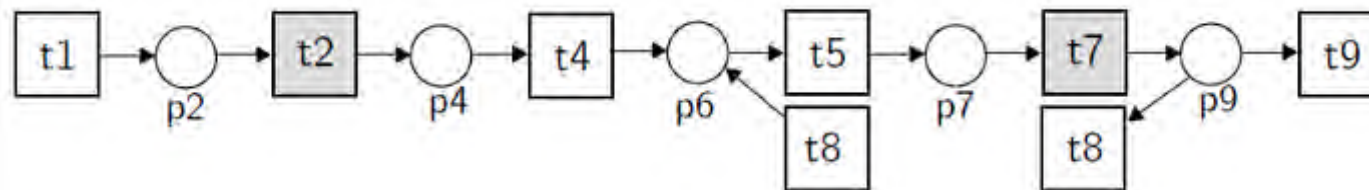
# Topological Diagnosis Algorithms



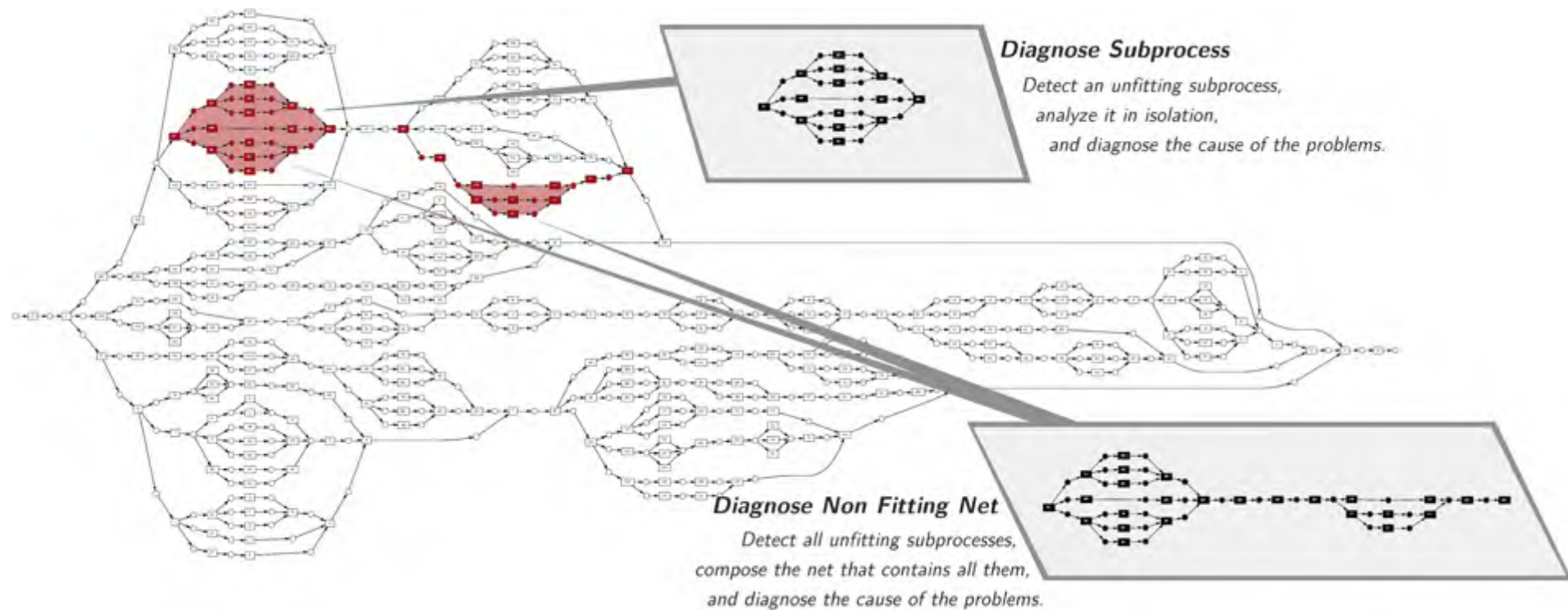
- Non-Fitting (Weakly) Connected Components



- Non-Fitting Subnet



# Topological Diagnosis in Large



# Topological Results

			[9,11]		k = 50				k = 100				k = 200			
	P	T	f	t	S/B	>5	nf	t	S/B	>5	nf	t	S/B	>5	nf	t
prAm6	347	363	0.92	75	129/57	29	7(3%)	423	62/27	14	1(9%)	323	27/12	7	1(10%)	180
prBm6	317	317	1	88	93/38	22	0(0%)	608	66/29	14	0(0%)	318	36/16	8	0(0%)	114
prCm6	317	317	0.57	2743	93/38	22	58(92%)	189	66/29	14	41(94%)	185	36/16	8	22(96%)	502
prDm6	529	429	-	-	105/34	33	5(8%)	1386	60/23	18	4(14%)	986	33/15	9	4(23%)	1284
prEm6	277	275	0.97	3566	82/35	20	2(5%)	529	35/13	11	2(5%)	343	15/7	5	2(6%)	211
prFm6	362	299	-	-	108/43	28	2(6%)	1667	57/23	15	2(21%)	863	21/9	5	1(23%)	562
prGm6	357	335	-	-	94/37	25	2(8%)	867	67/31	15	2(8%)	850	51/25	11	2(8%)	474

From almost 700  
nodes ...

			NFWCC				NFN		
	P	T	C <sub>c</sub>	V	P	T	V	P	T
prAm6	317	317	7	1	2.1	3	14	15	14
prCm6	317	317	38	1.5	8.2	9.5	113	315	317
prDm6	529	429	5	1	9.4	9.4	31	55	52
prEm6	277	275	2	1	1	2	31	29	40
prFm6	362	299	2	1	13	11	7	27	25
prGm6	357	335	2	1	16.5	14.5	5	34	29

... to 70



# The 4 Challenges

- Comprehensible
- Guaranties
- Fast
- Diagnosis

# Future Work

- Estimate fitness
  - Not decisional but metric
- Divide-and-Conquer Alignment Algorithms
  - Reconstruct the alignment
- New decompositions
  - Less trivial components
- New conformance dimensions
  - Precision, generalization, ...

# Conclusions

- Partitioning Technique for Conformance Checking
- Based on SESE and RPST
- May be faster, distributed, and help on the diagnosis
- Topology for diagnosis
- Implemented in ProM framework.



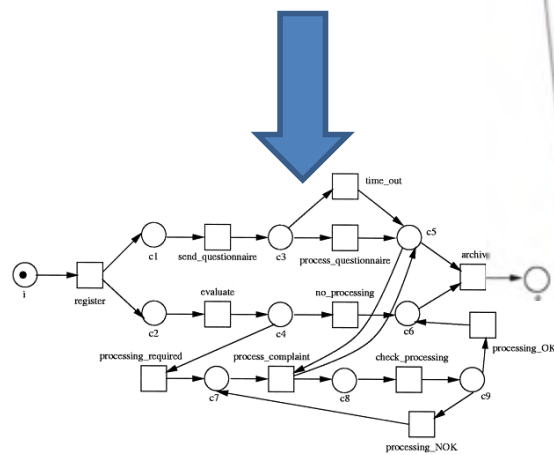
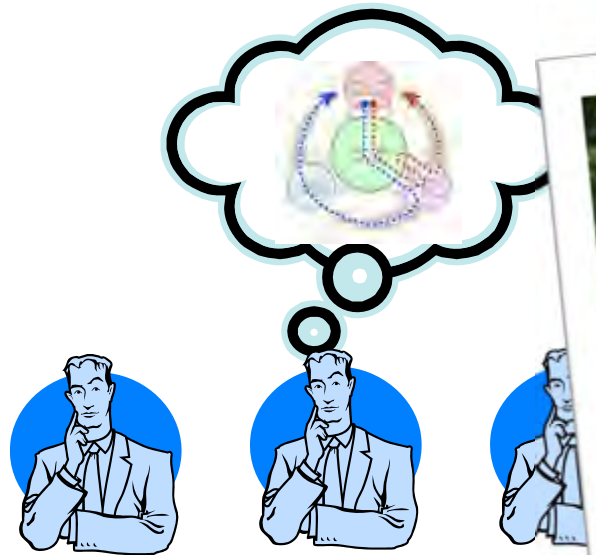
UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

**TU/e** Technische Universiteit  
Eindhoven  
University of Technology  
Where innovation starts

---

**Thank You**

# Process Mining in a Nutshell



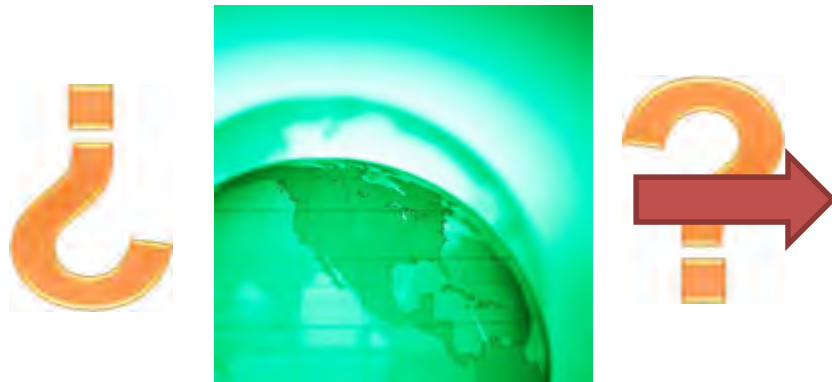
**THEORY**



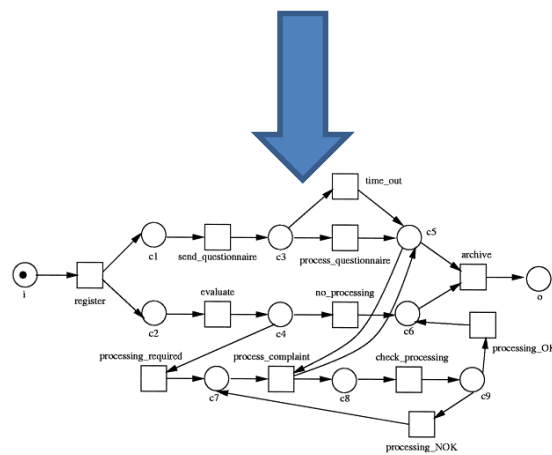
**REALITY**

# Process Mining in a Nutshell

*REALITY REFLECTION*



**LOGS**



**THEORY**



**REALITY**