Conformance Checking: Token-Based Replay and Alignments

Wil van der Aalst

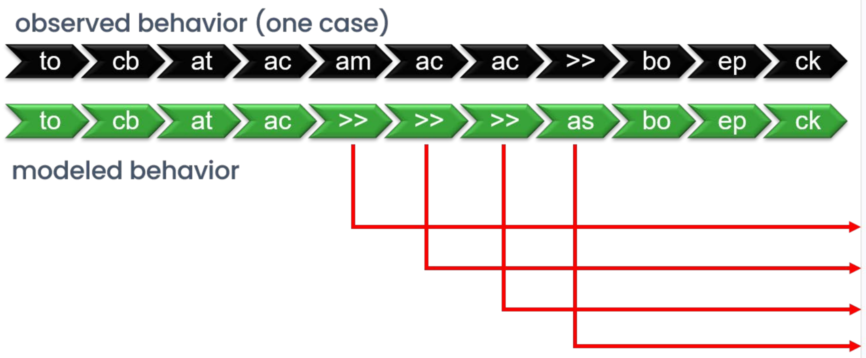
Process and Data Science @ RWTH Aachen University & Celonis

www.vdaalst.com, @wvdaalst

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

About this third lecture

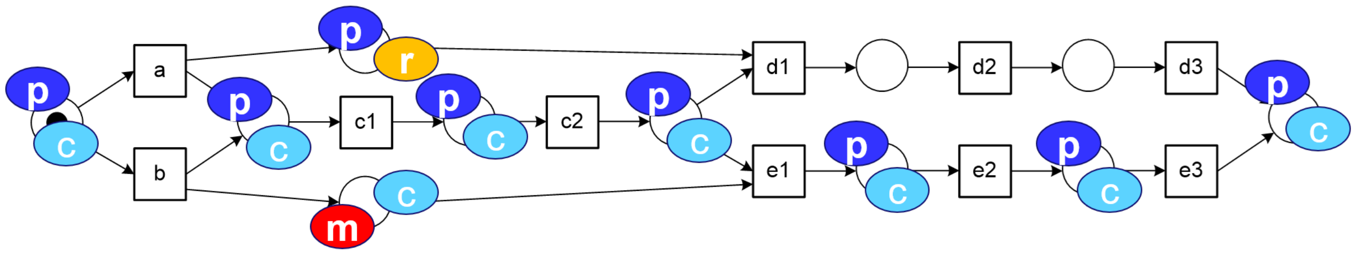


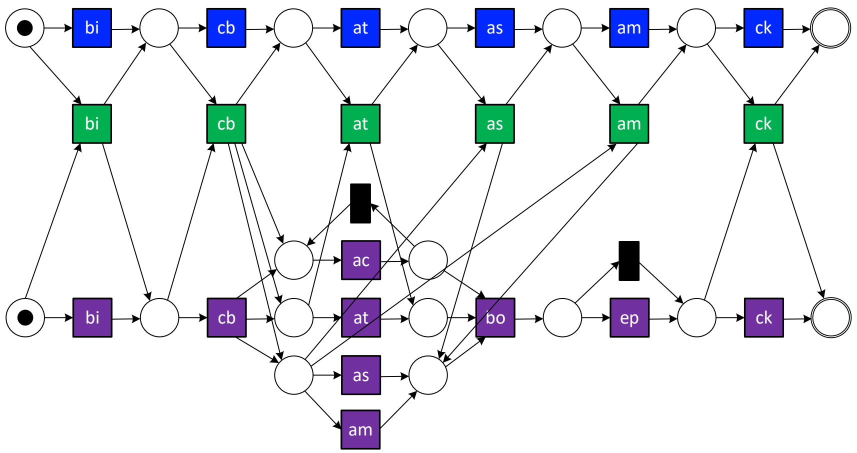
Variant detail

Event sequence	Deviations	Time delta
take order		0
create base		8 minutes
add tomato		1 hours
add cheese		21 seconds
add mushrooms	Occurred unexpectedly	15 seconds
add cheese	Occurred too often	26 seconds
add cheese	Occurred too often	38 seconds
add salami	Missing event	-
bake in oven		2 minutes
eat pizza		7 minutes
clean kitchen		8 minutes

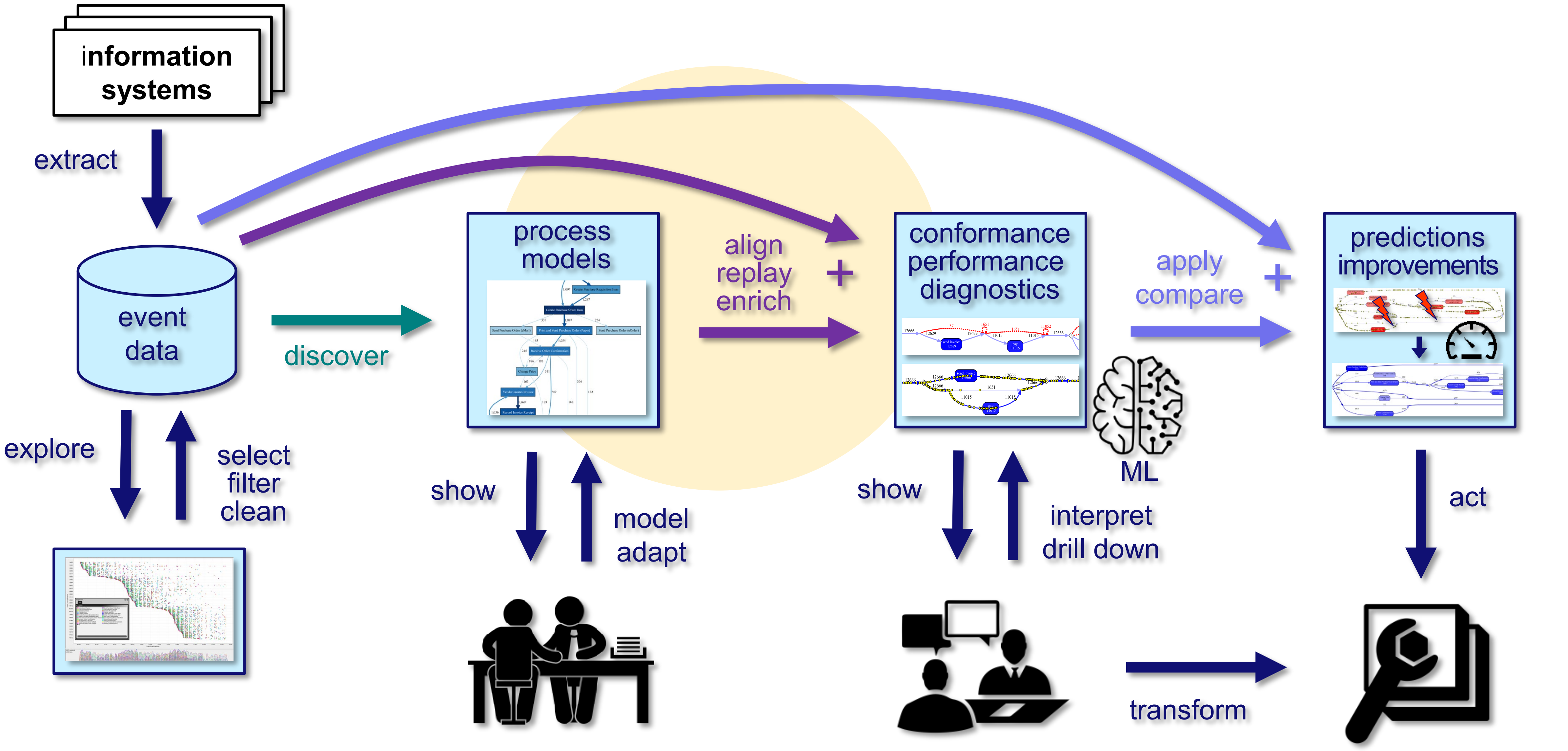
- ❑ Conformance
- ❑ Alignments informal
- ❑ Token-based replay
- ❑ Alignments revisited

⟨a,c1,c2,e1,e2,e3⟩



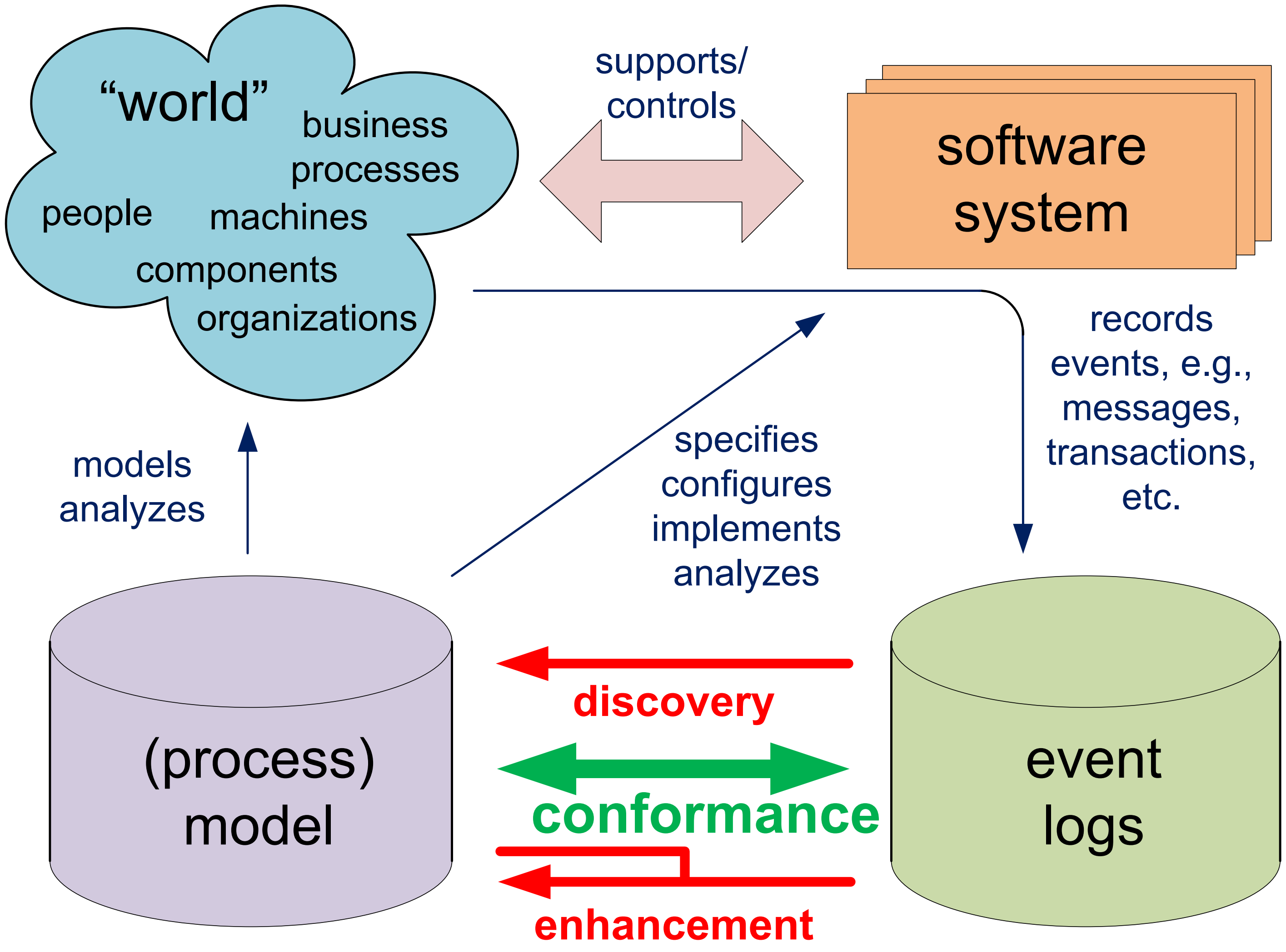


Chair of Process and Data Science



Chair of Process and Data Science

Conformance checking



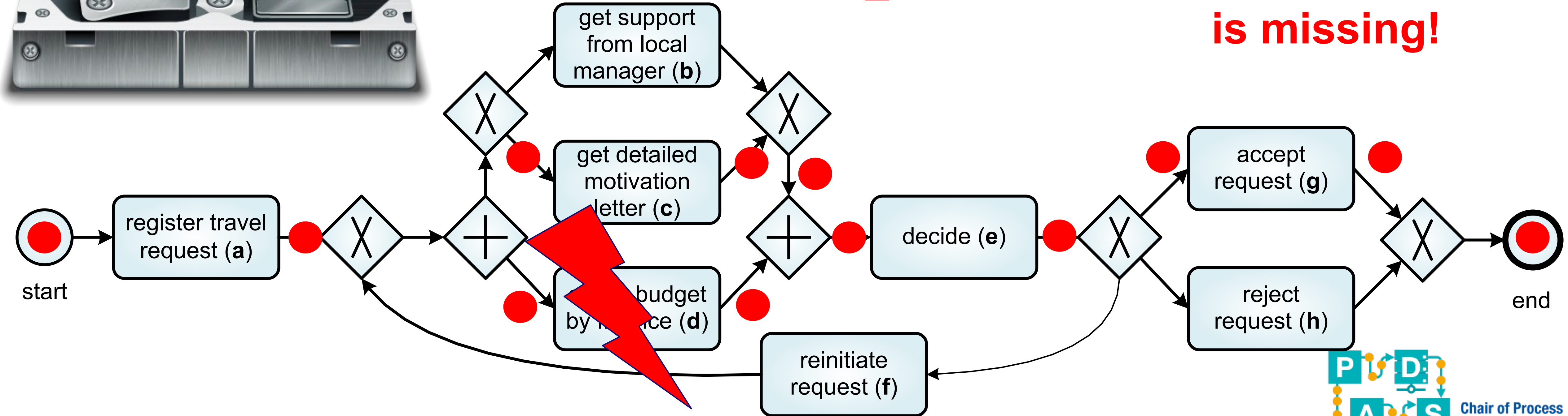
Chair of Process and Data Science

Replay example used before

a c e g

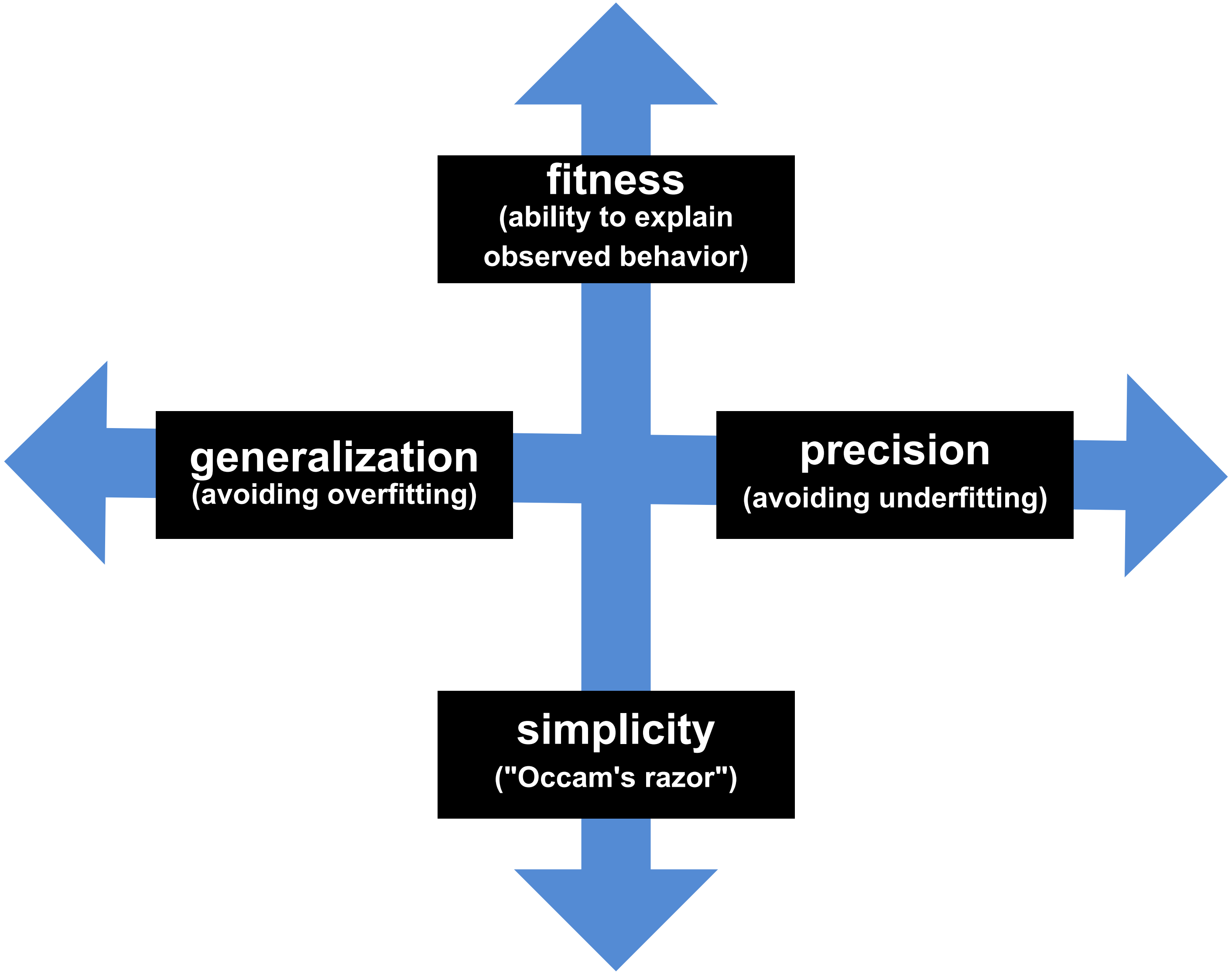
?

check budget (d) is missing!



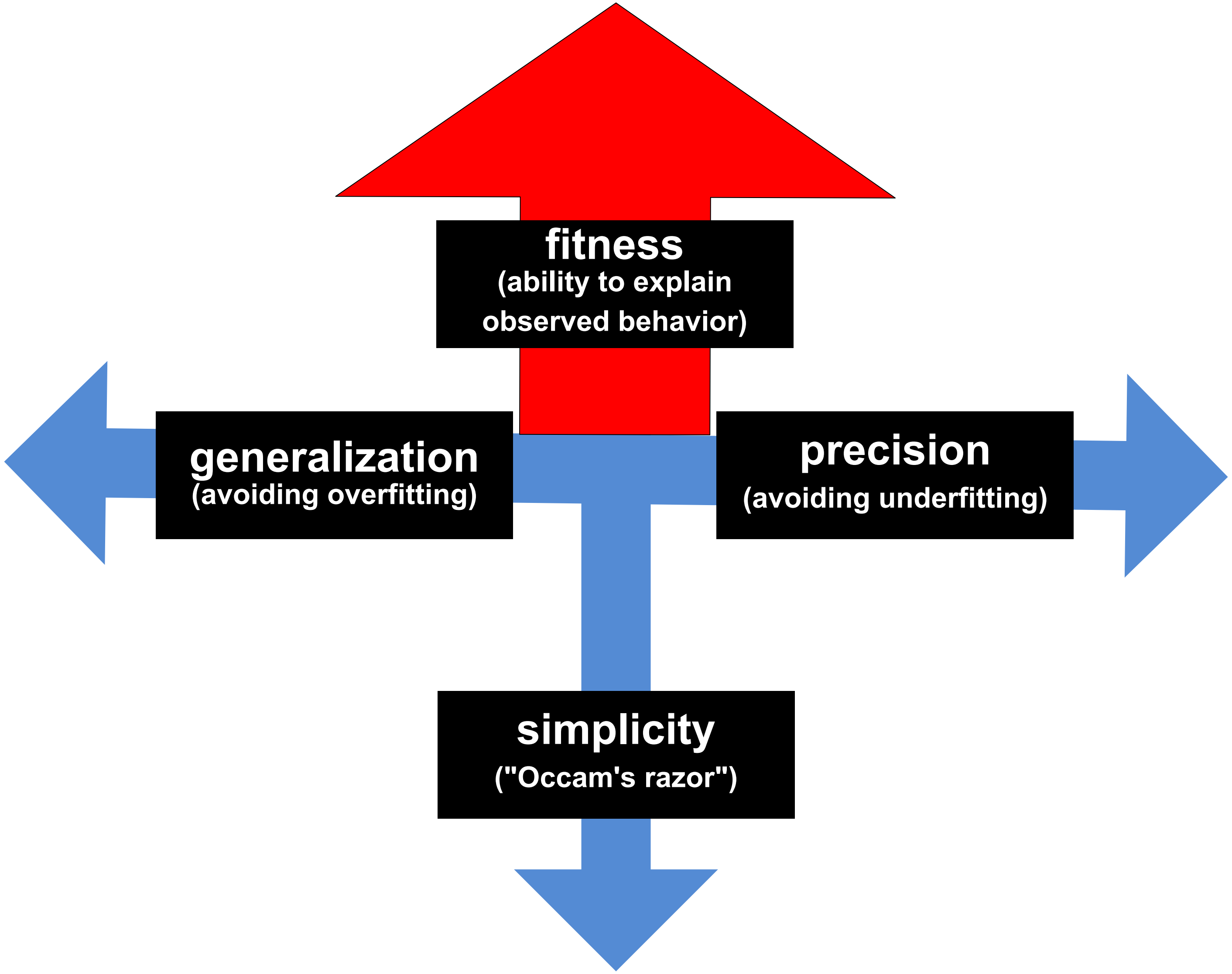
Chair of Process and Data Science

Four dimensions to compare log and model

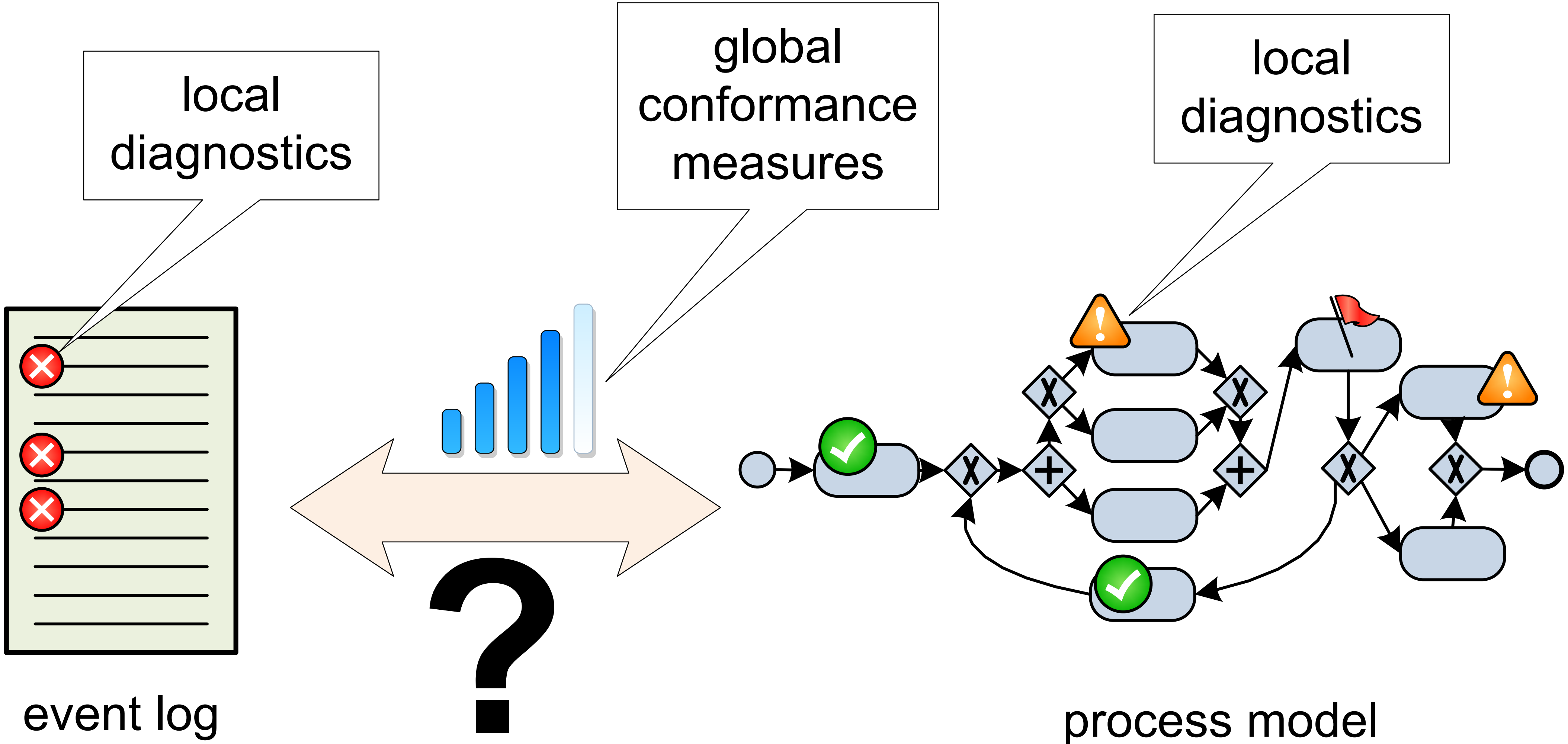


Chair of Process and Data Science

Replay fitness is dominant



Conformance diagnostics and measures



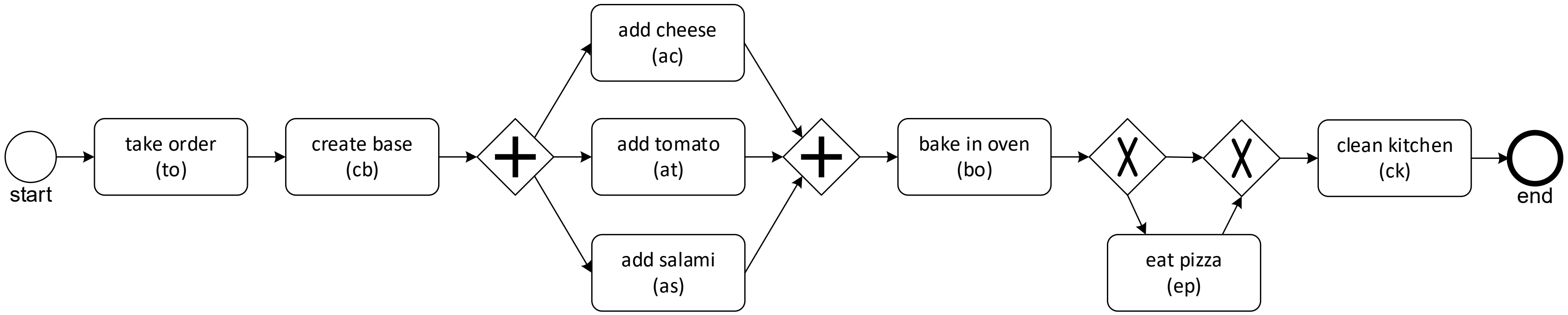
Chair of Process
and Data Science

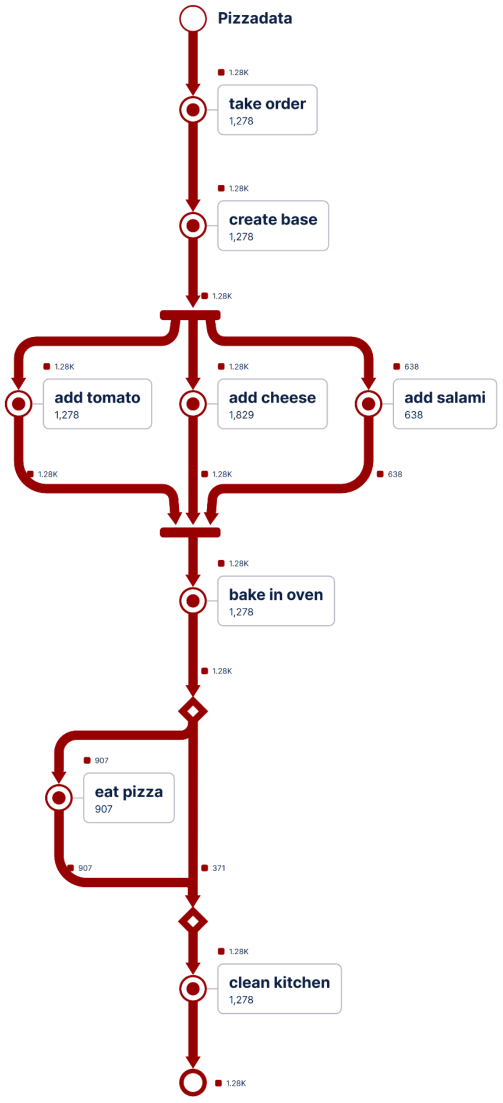
Alignments (informal)

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

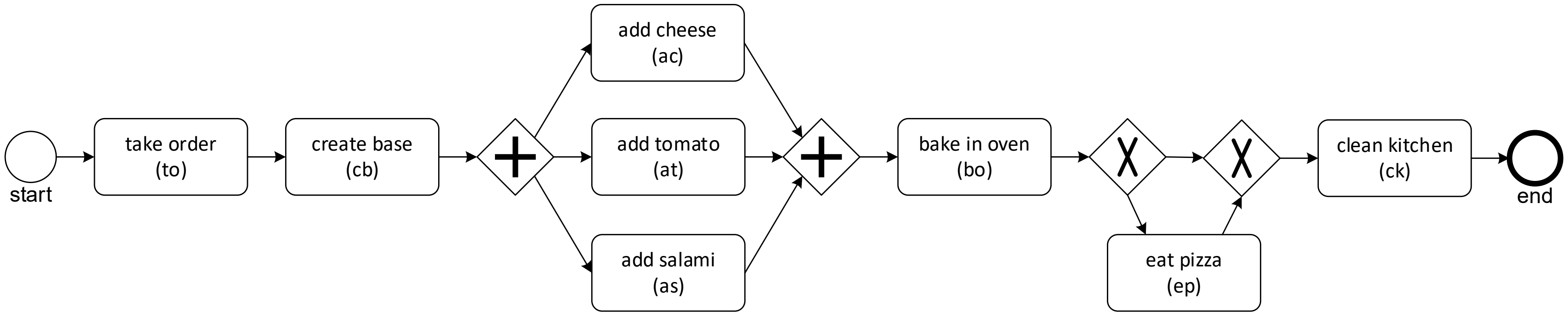
For conformance checking we need a process model and event data





For conformance checking we need a process model and event data

Just a single
serving of cheese!

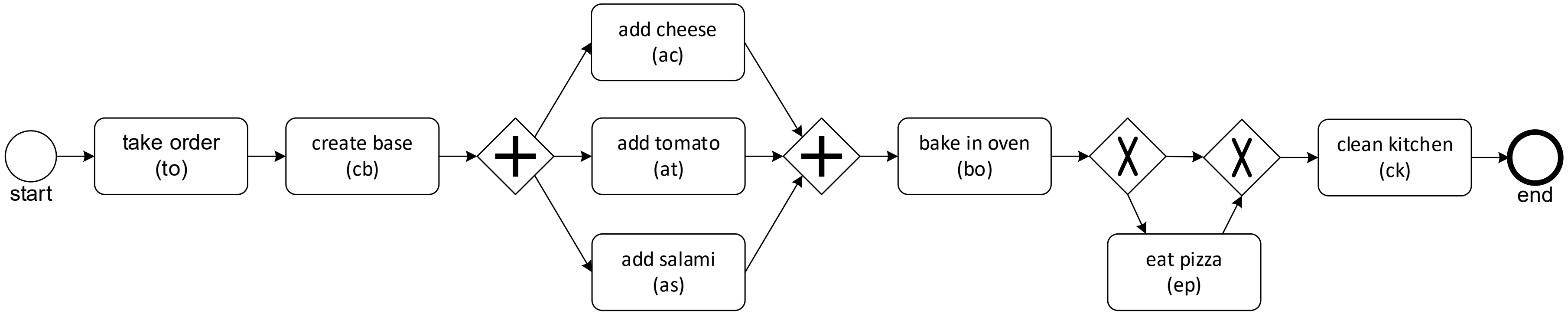


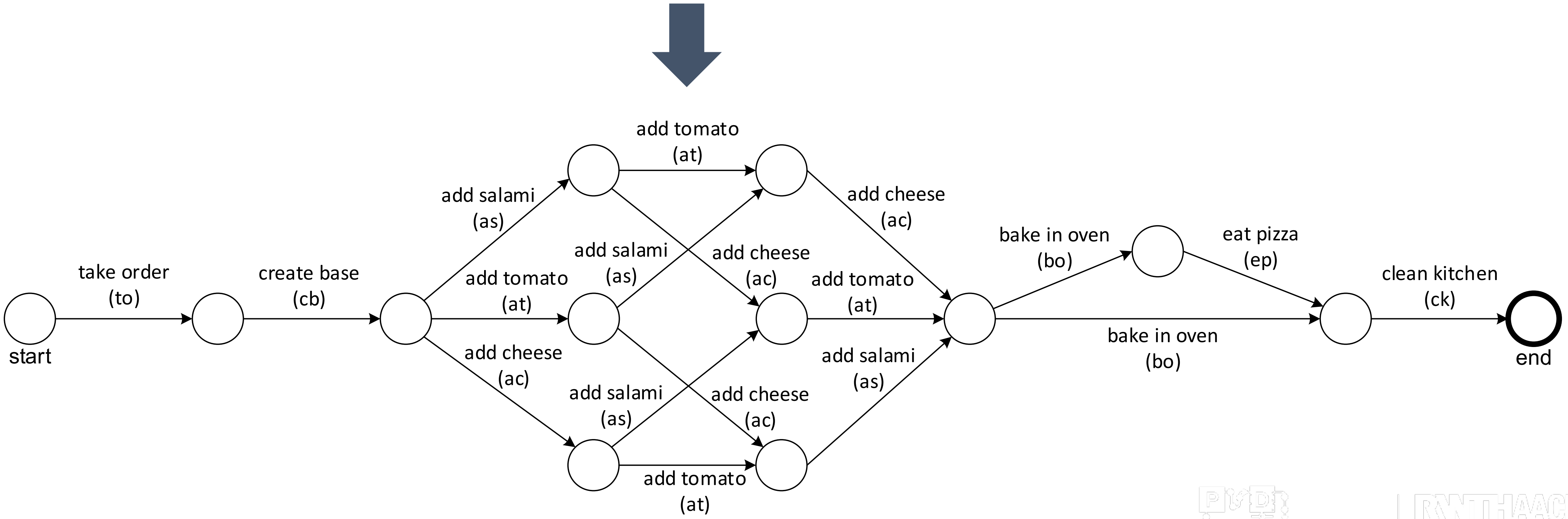
Only salami, no
mushrooms allowed.

to cb am at ac ac bo ep ck

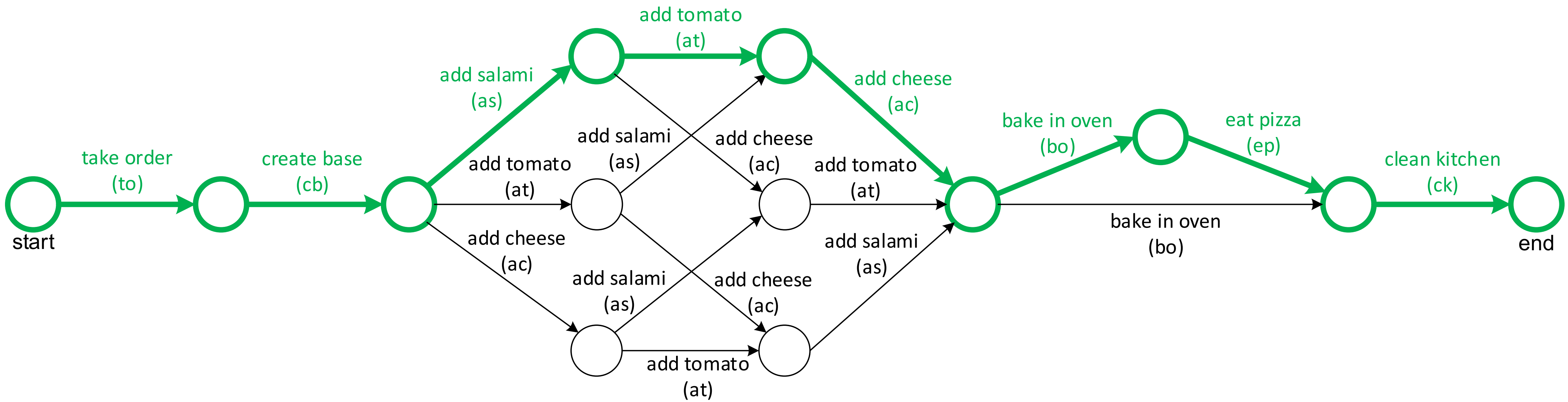
A variant appearing 15 times in the original event data.

The state space of the BPMN model (normally much larger and with loops)

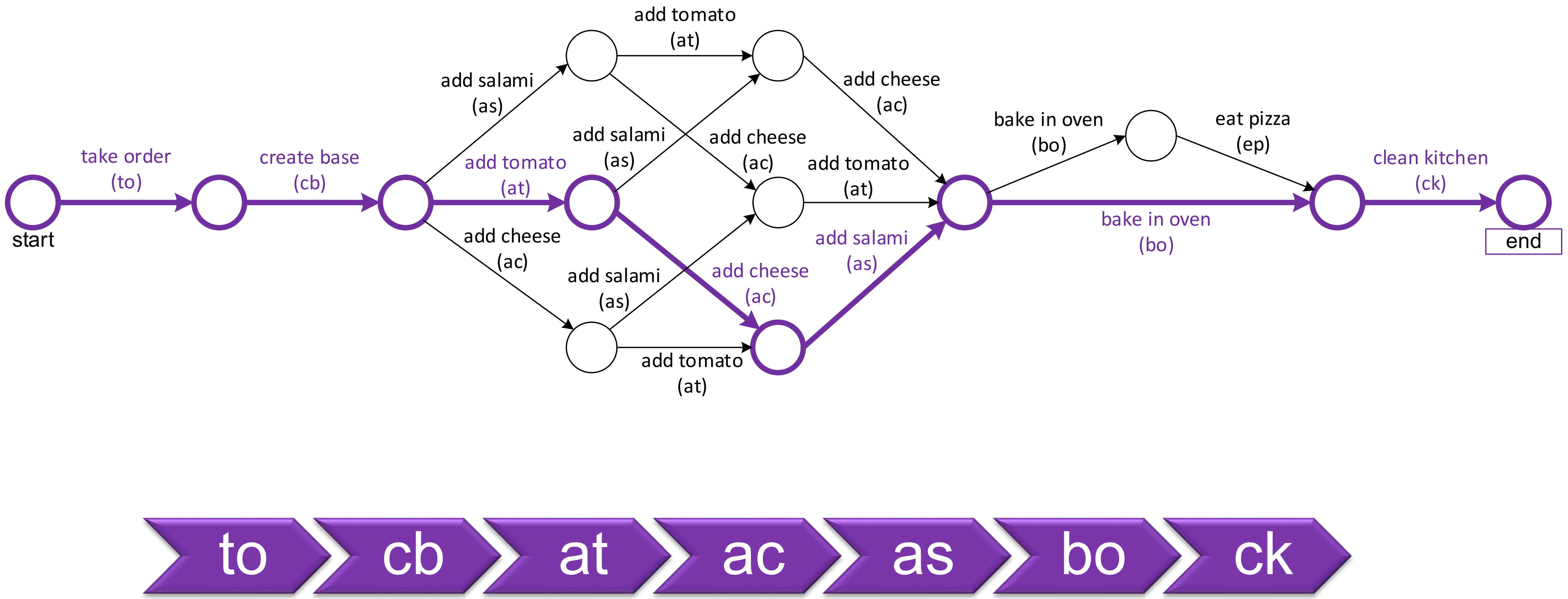




Behavior of BPMN model is any path from start to end

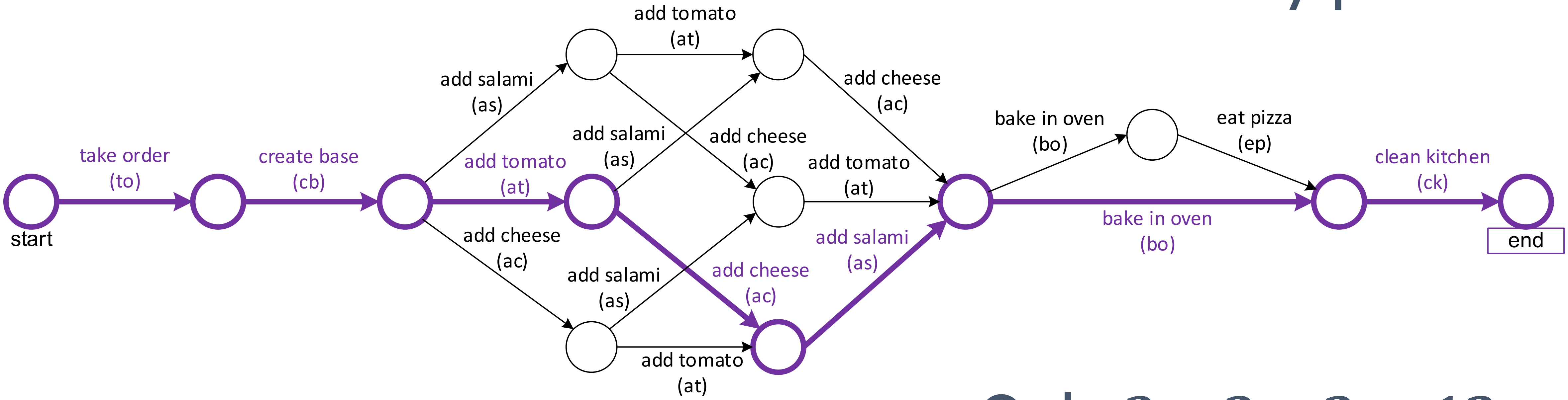


Another path



Another path

How many paths?



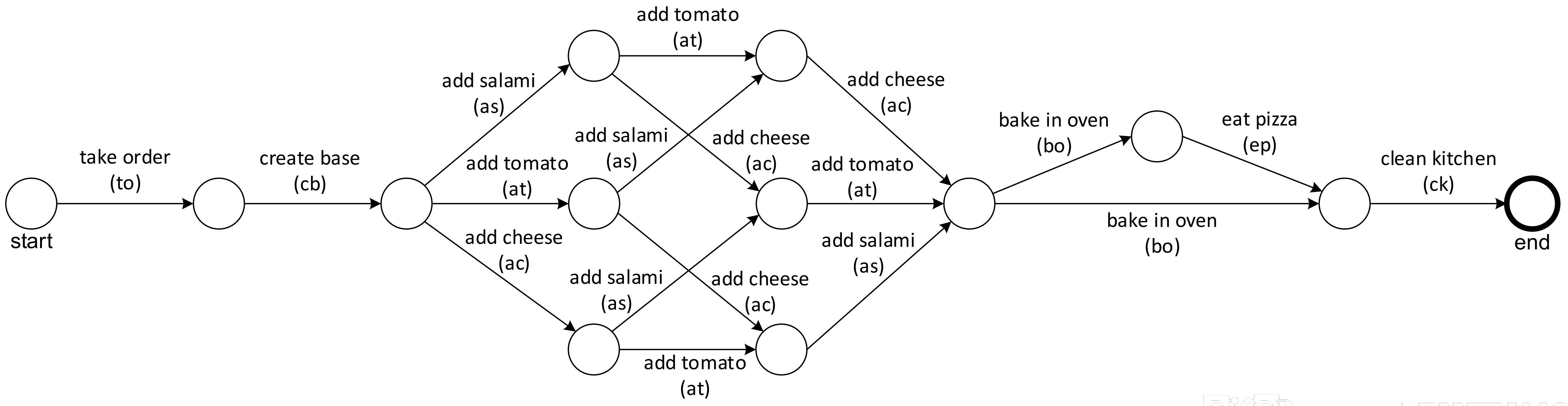
Only $3 \times 2 \times 2 = 12$

Find a path closest to the observed sequence

to › cb › am › at › ac › ac › bo › ep › ck

observed behavior

?

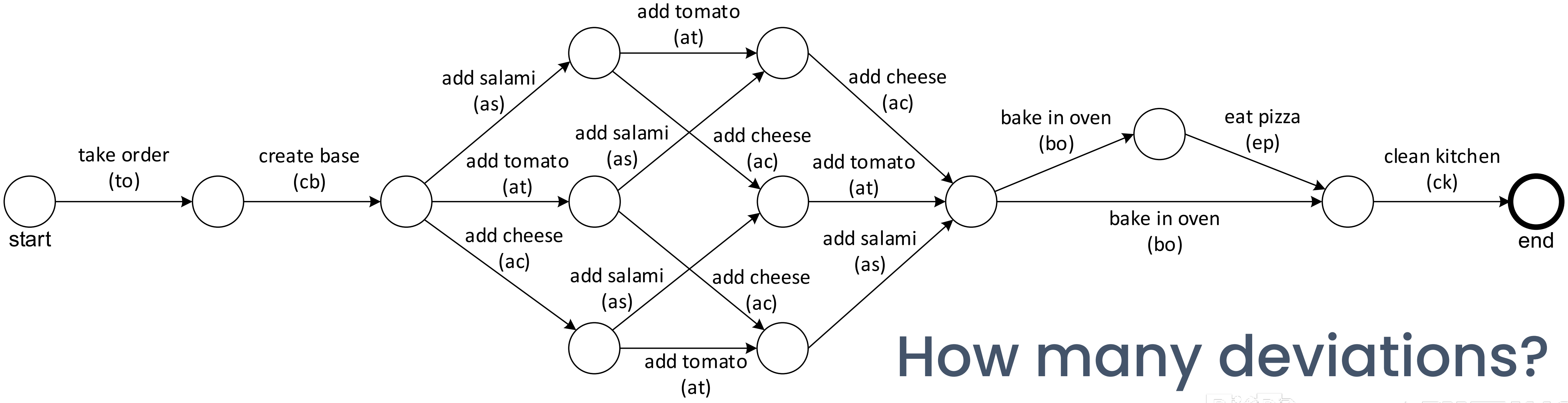


Find a path closest to the observed sequence

observed behavior

?

What is the path closest to the observed sequence?

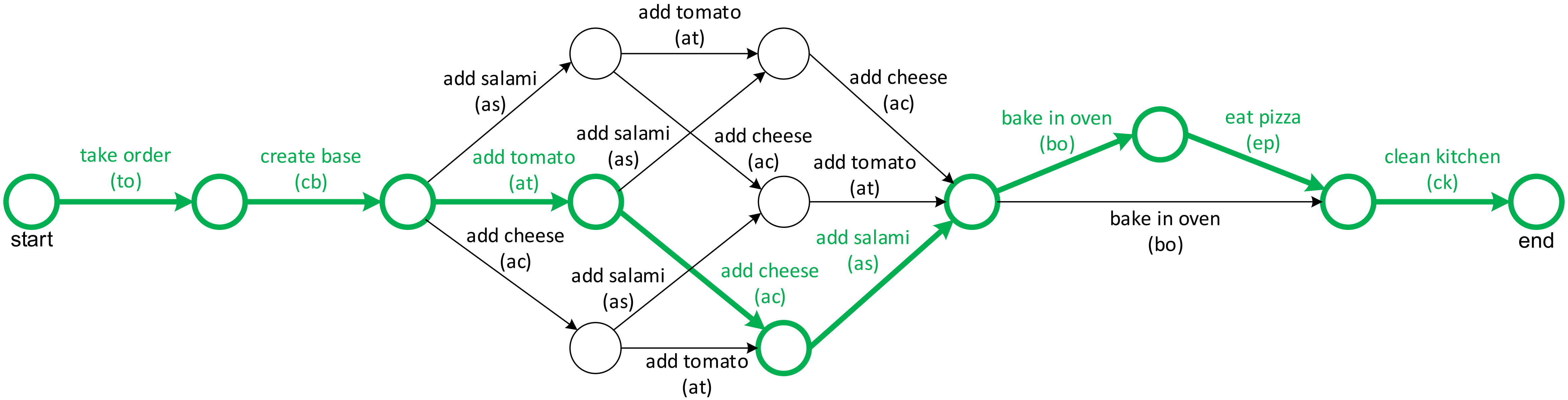


How many deviations?

Find a path closest to the observed sequence

to cb am at ac ac bo ep ck

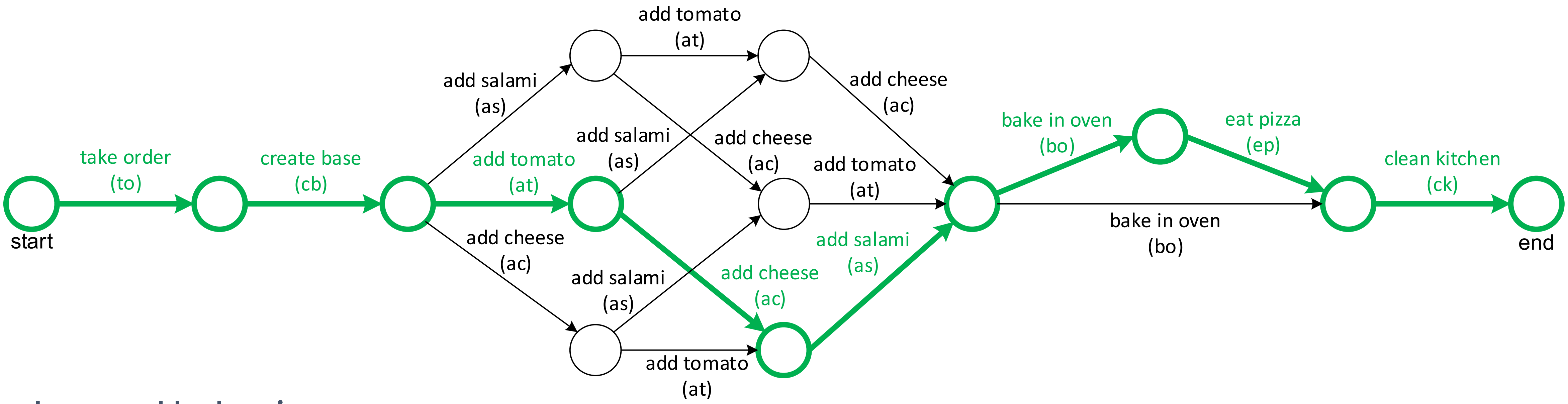
observed behavior



to cb at ac as bo ep ck

modeled behavior

This is a path closest to the observed sequence (three deviations)



observed behavior

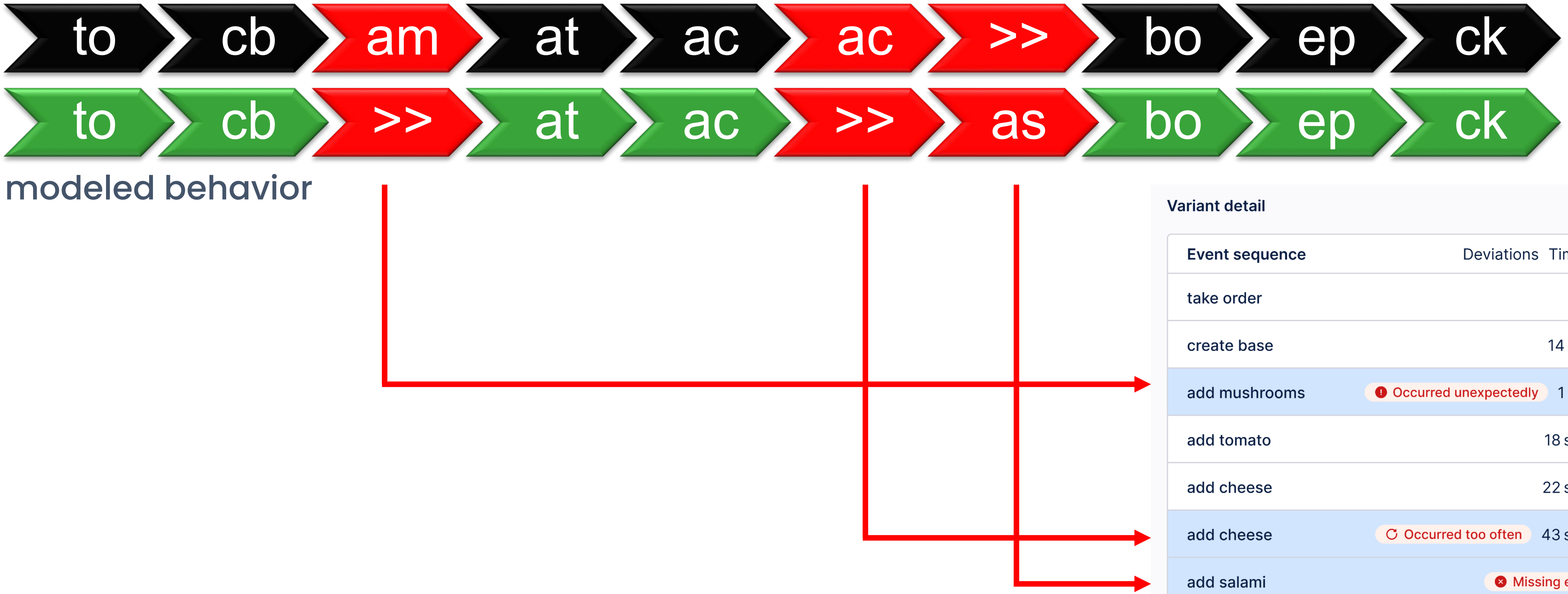
to	cb	am	at	ac	ac	>>	bo	ep	ck

modeled behavior

to	cb	>>	at	ac	>>	as	bo	ep	ck

Three deviations!

observed behavior (15 cases)

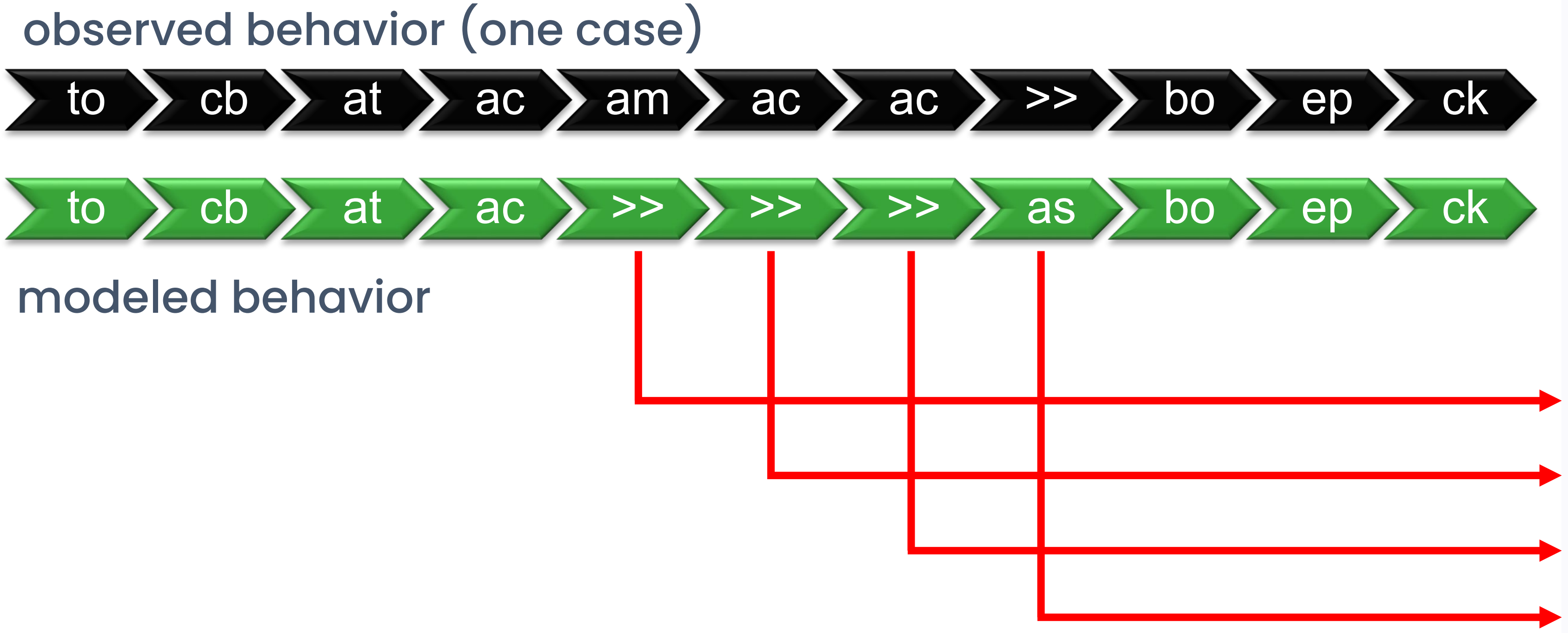


modeled behavior

Variant detail

Event sequence	Deviations	Time delta
take order		0
create base		14 minutes
add mushrooms	Occurred unexpectedly	1 minutes
add tomato		18 seconds
add cheese		22 seconds
add cheese	Occurred too often	43 seconds
add salami	Missing event	-
bake in oven		1 minutes
eat pizza		20 minutes
clean kitchen		22 minutes

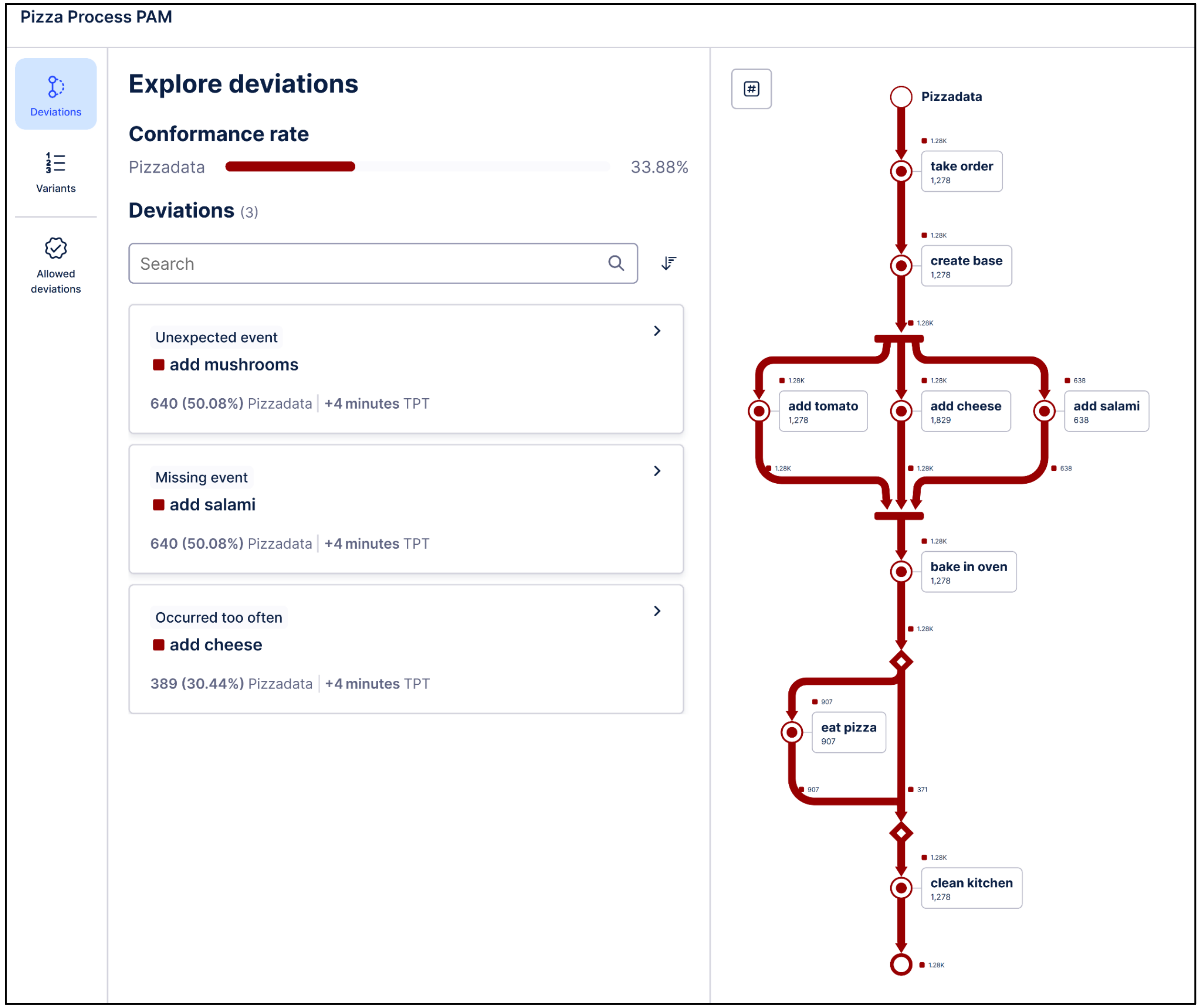
Another example with four deviations



Variant detail

Event sequence	Deviations	Time delta
take order		0
create base		8 minutes
add tomato		1 hours
add cheese		21 seconds
add mushrooms	Occurred unexpectedly	15 seconds
add cheese	Occurred too often	26 seconds
add cheese	Occurred too often	38 seconds
add salami	Missing event	-
bake in oven		2 minutes
eat pizza		7 minutes
clean kitchen		8 minutes

Notoriously difficult problem: Requires state-of-the-art techniques



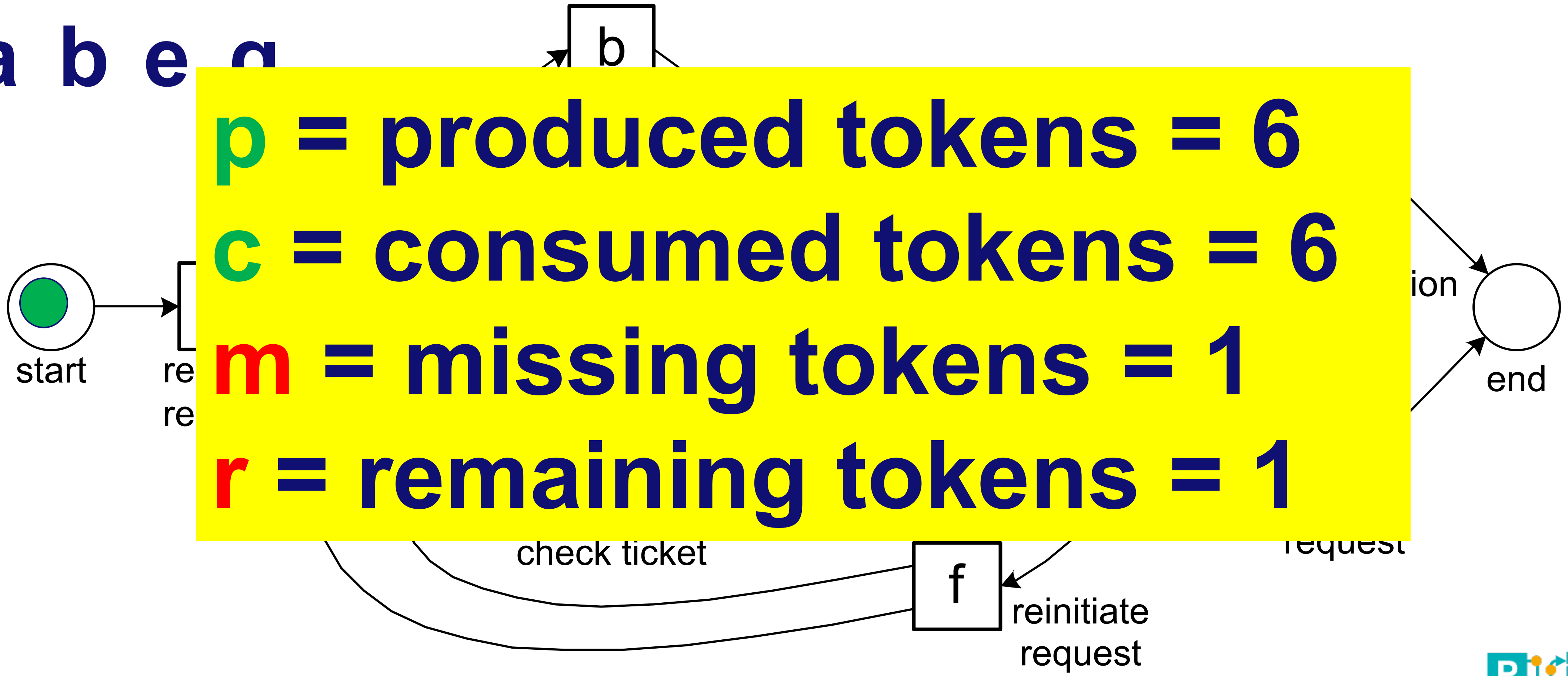
- Alignments are the “gold standard” in conformance checking.
- Requires solving an optimization problem for each case.
- Also needed for performance analysis.
- Made scalable using state-of-the-art techniques.

Token-based replay

Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

Counting tokens while replaying

a b e g



p = produced tokens = 6

c = consumed tokens = 6

m = missing tokens = 1

r = remaining tokens = 1

Quantifying fitness at the trace level

$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.83333$$

p = produced tokens = 6

c = consumed tokens = 6

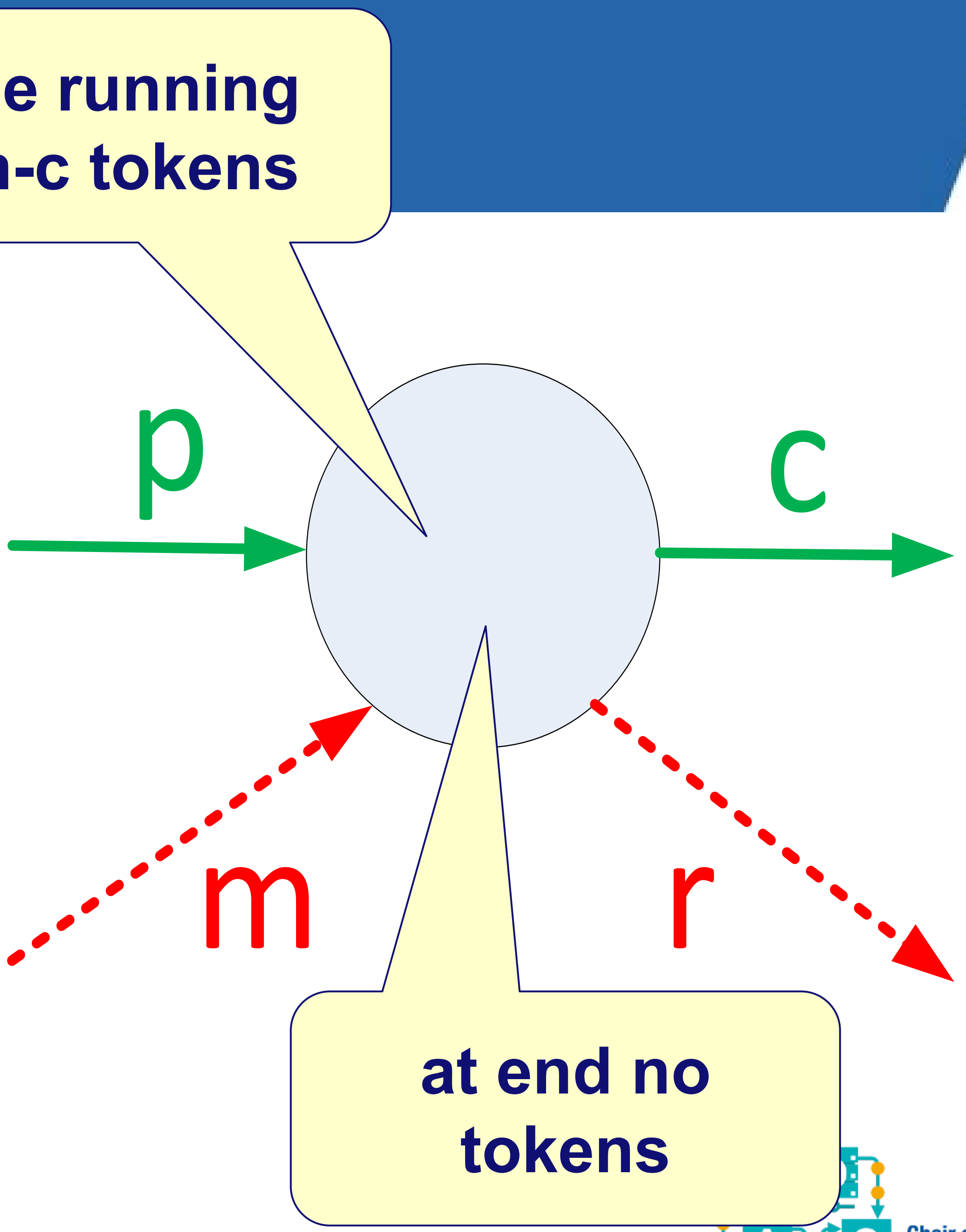
m = missing tokens = 1

r = remaining tokens = 1

Approach (1/3)

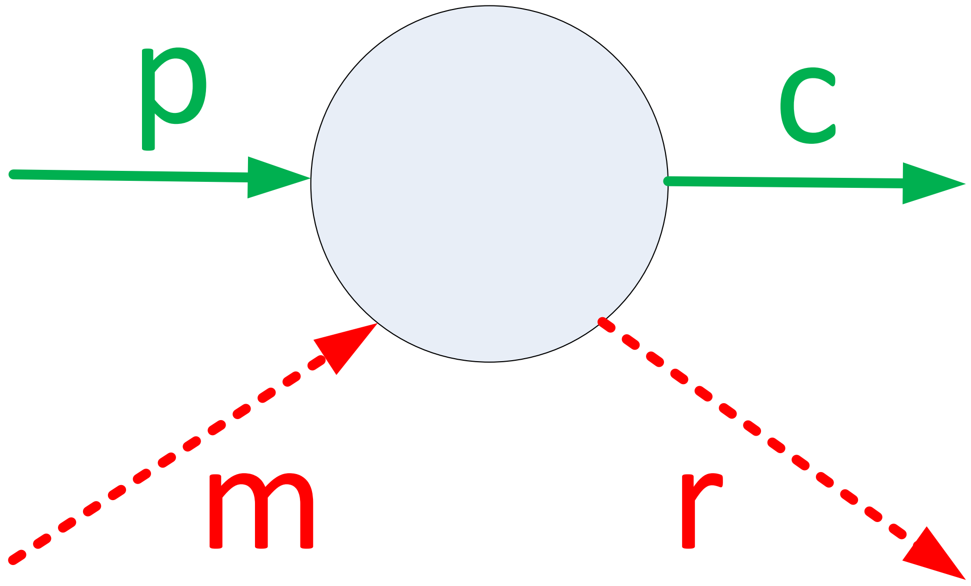
Use four counters:

- p = produced tokens
- c = consumed tokens
- m = missing tokens (consumed while not there)
- r = remaining tokens (produced but not consumed)



Chair of Process and Data Science

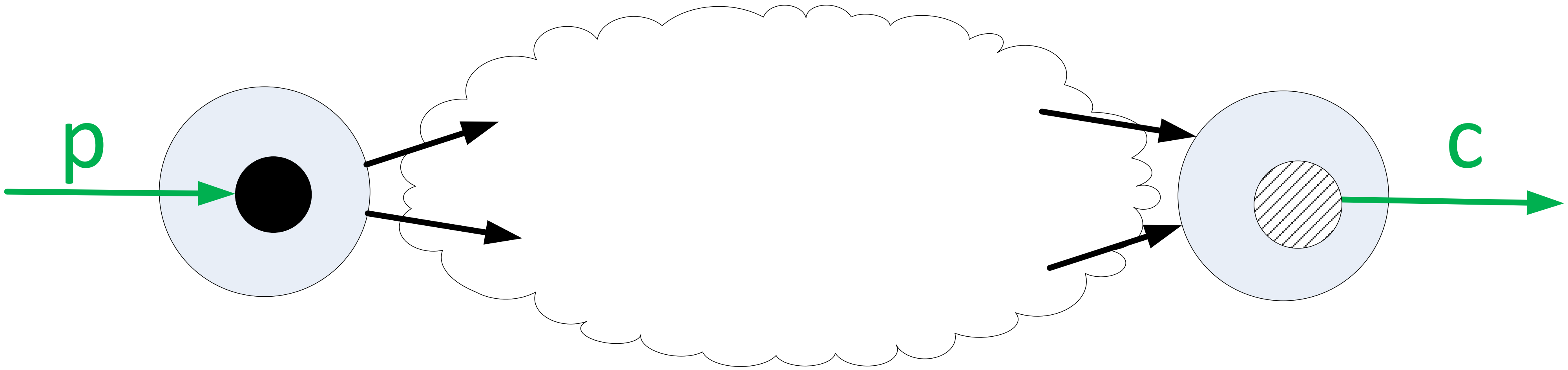
Approach (2/3)



- **Invariants**
  - At any time: $p+m \geq c \geq m$ (also per place)
  - At the end: $r = p + m - c$ (also per place)

Chair of Process and Data Science

Approach (3/3)



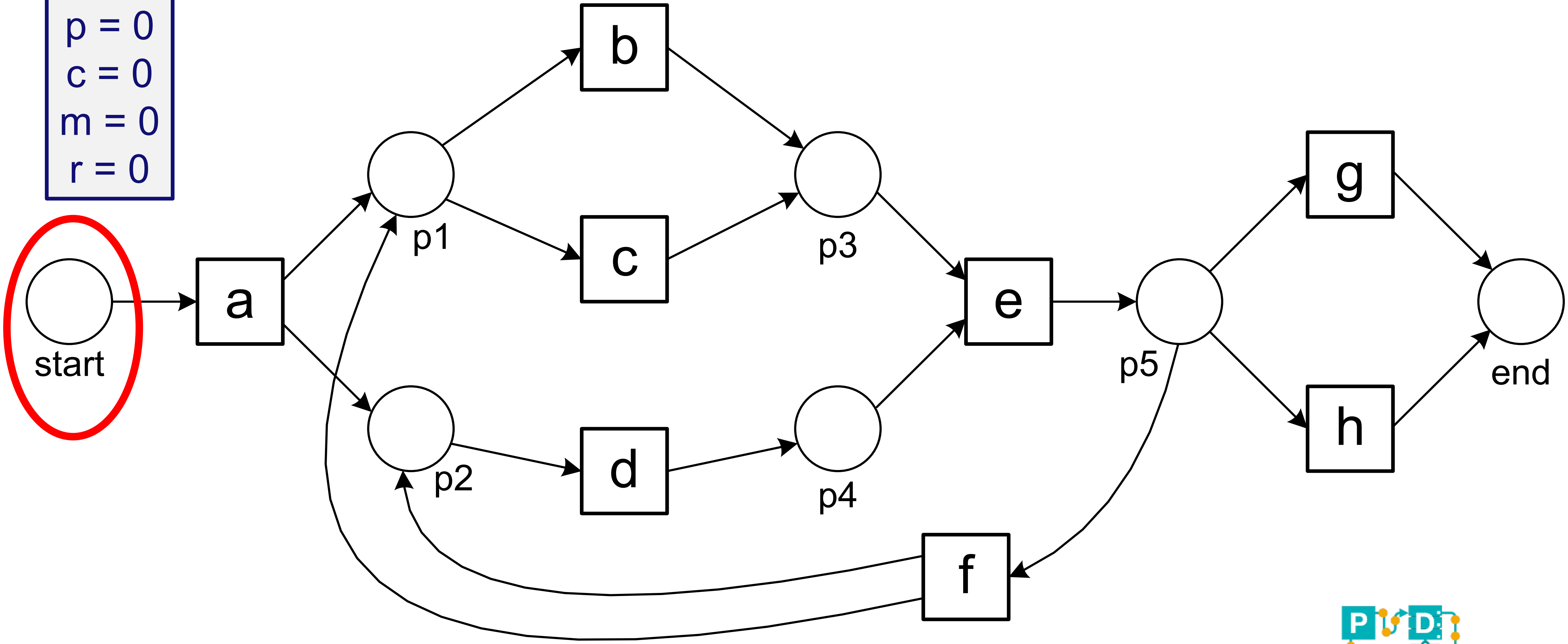
Initialization and finalization:

- In the beginning a token is **produced** for the source place: p = 1.
- At the end a token is **consumed** from the sink place (also if not there): c' = c + 1.

Chair of Process and Data Science

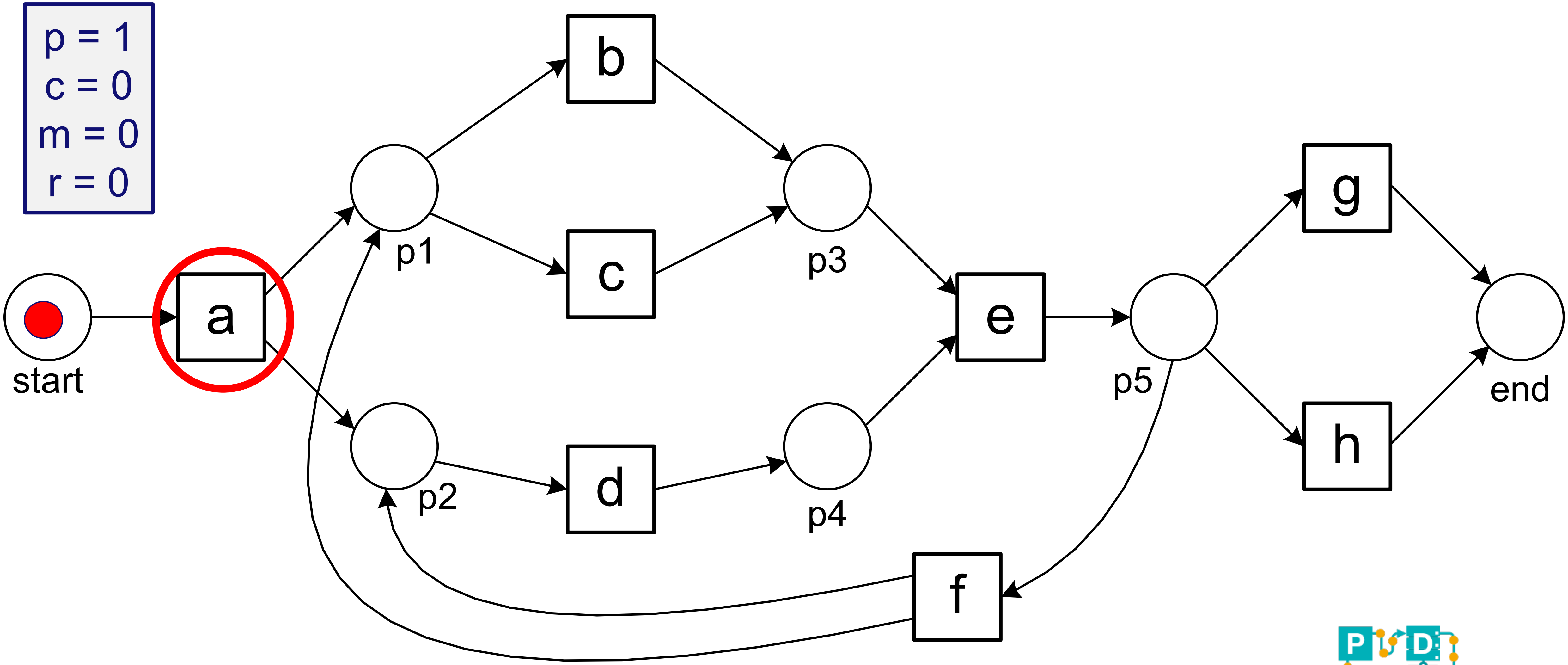
Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

p = 0
c = 0
m = 0
r = 0



Chair of Process and Data Science

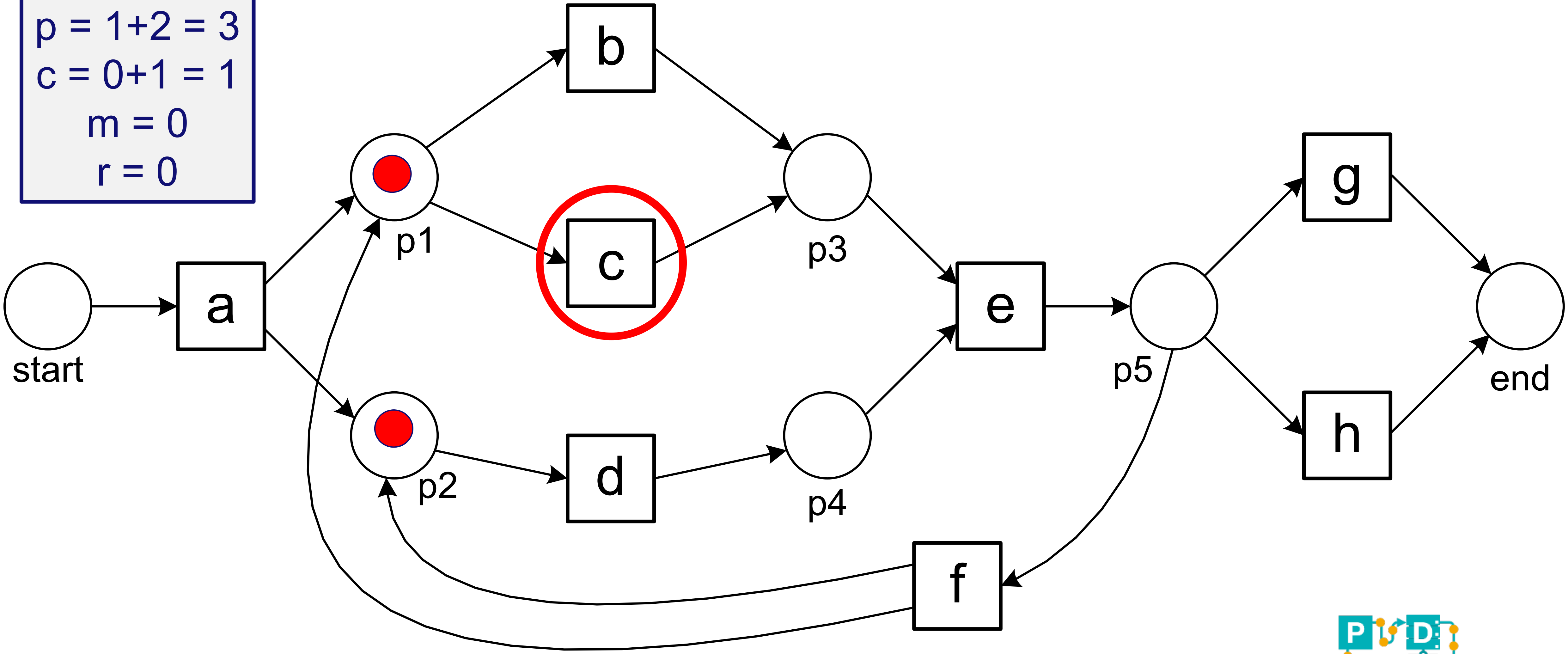
Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$



Chair of Process and Data Science

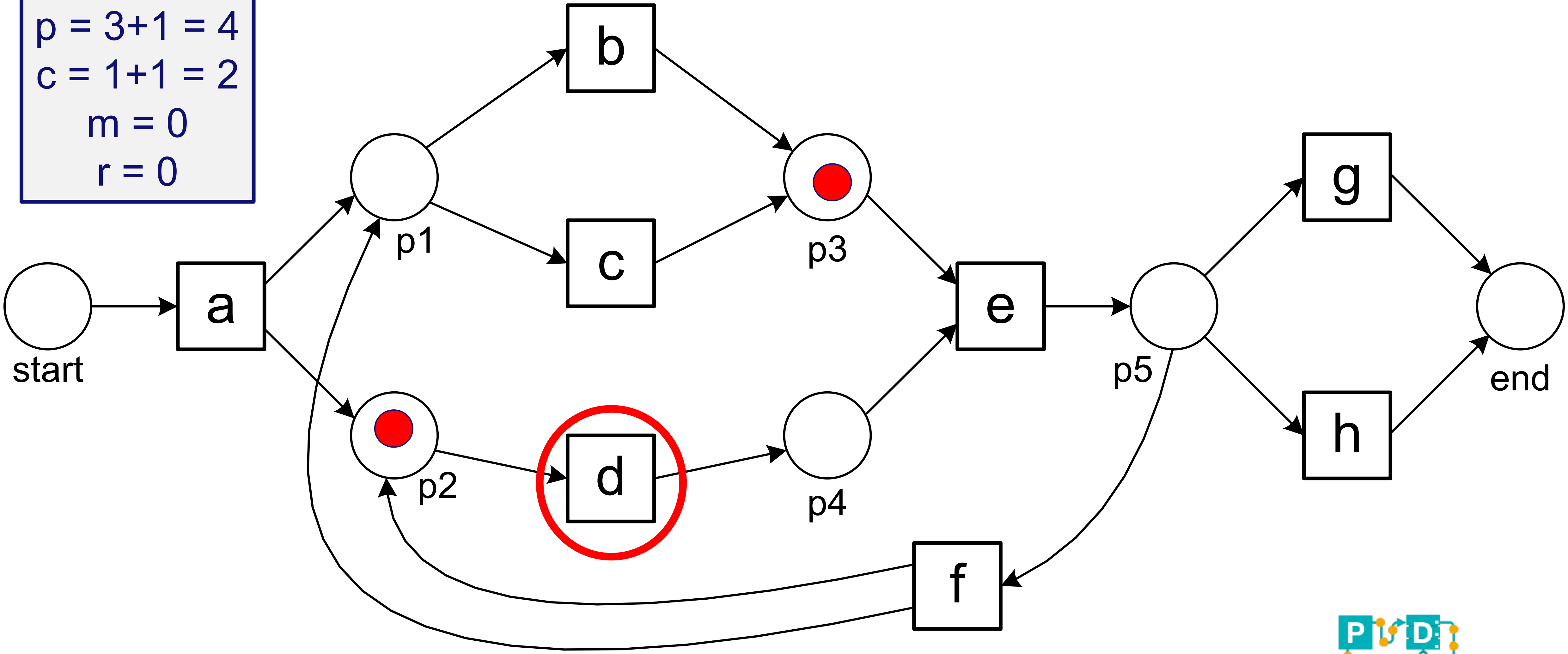
Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

p = 1+2 = 3
c = 0+1 = 1
m = 0
r = 0



Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

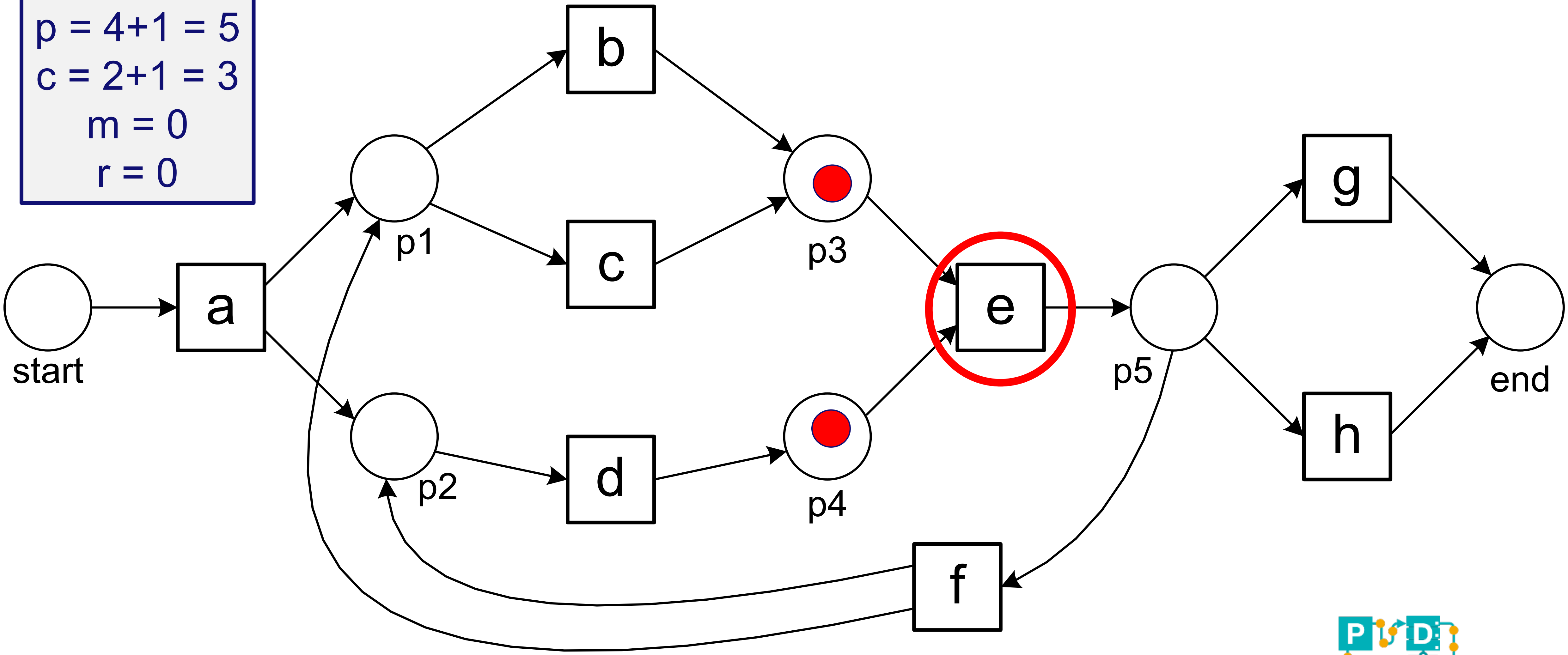
p = 3+1 = 4
c = 1+1 = 2
m = 0
r = 0



Chair of Process and Data Science

Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

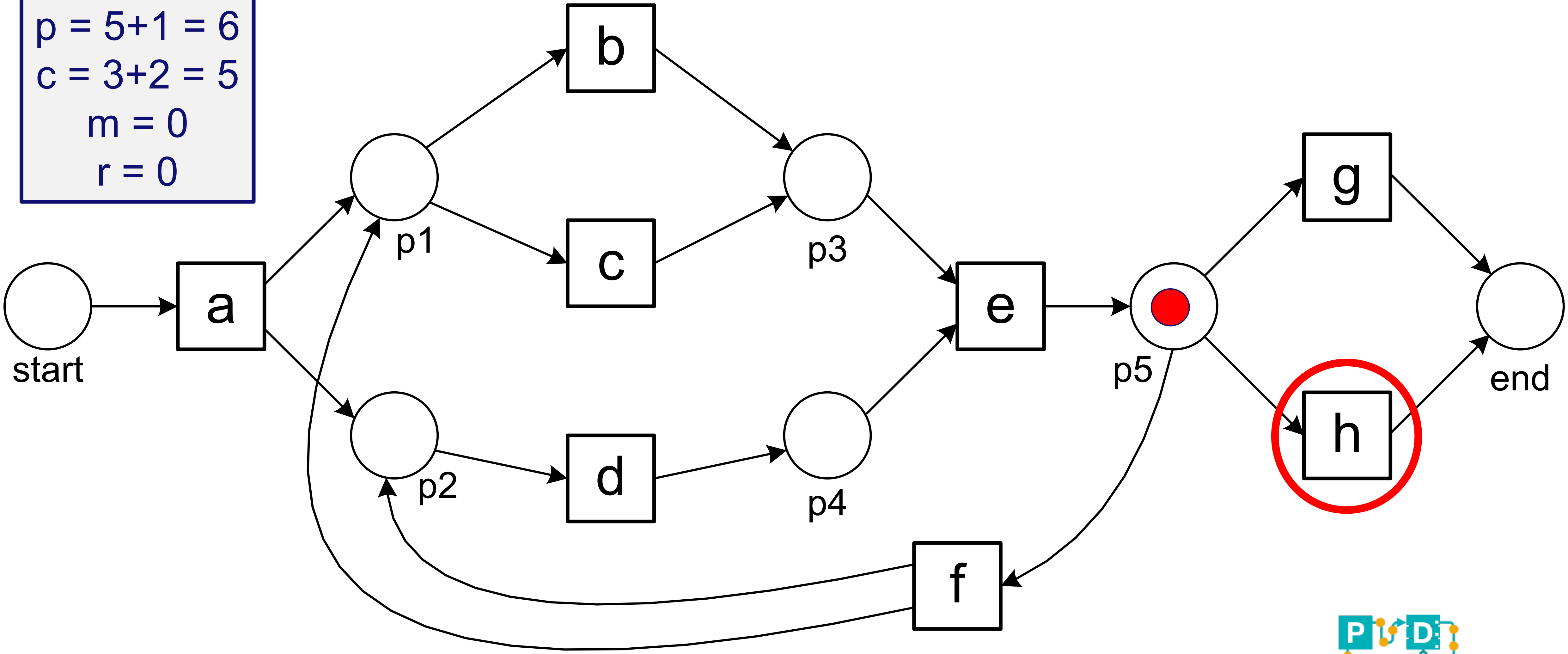
p = 4+1 = 5
c = 2+1 = 3
m = 0
r = 0



Chair of Process and Data Science

Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

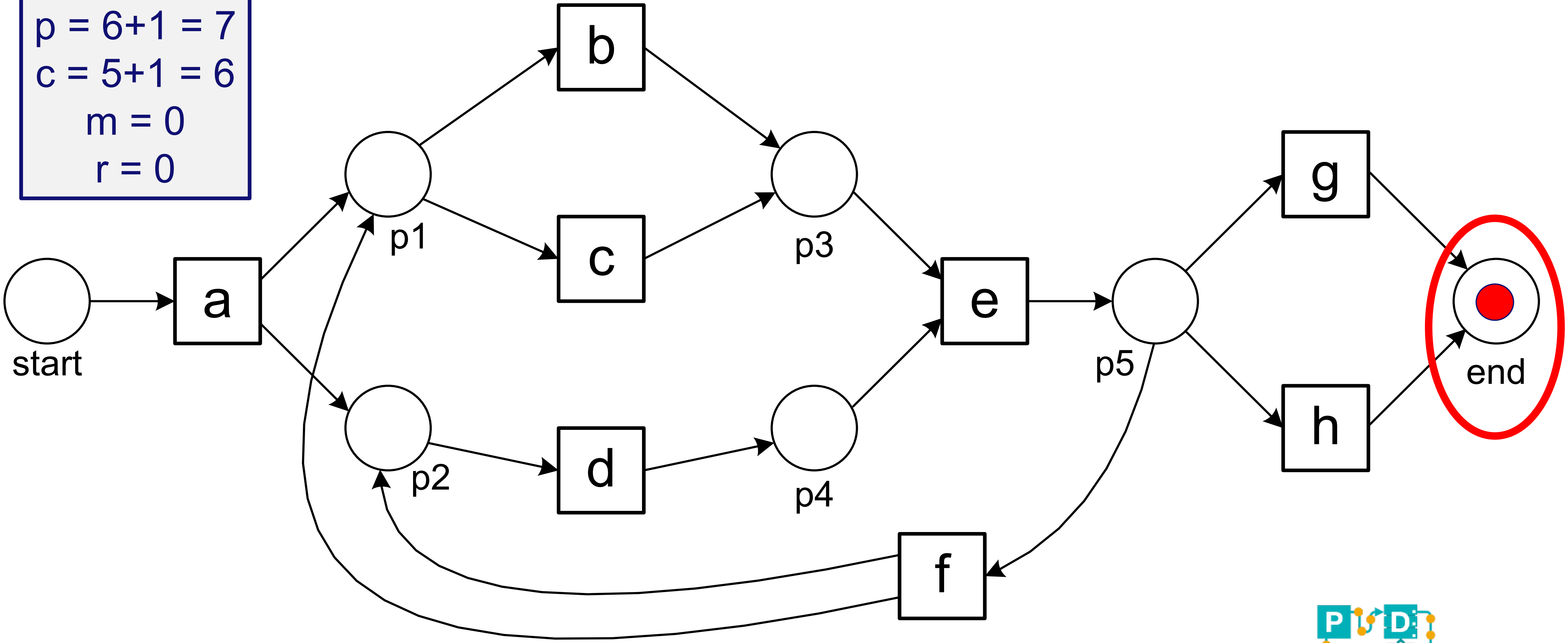
p = 5+1 = 6
c = 3+2 = 5
m = 0
r = 0



Chair of Process and Data Science

Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

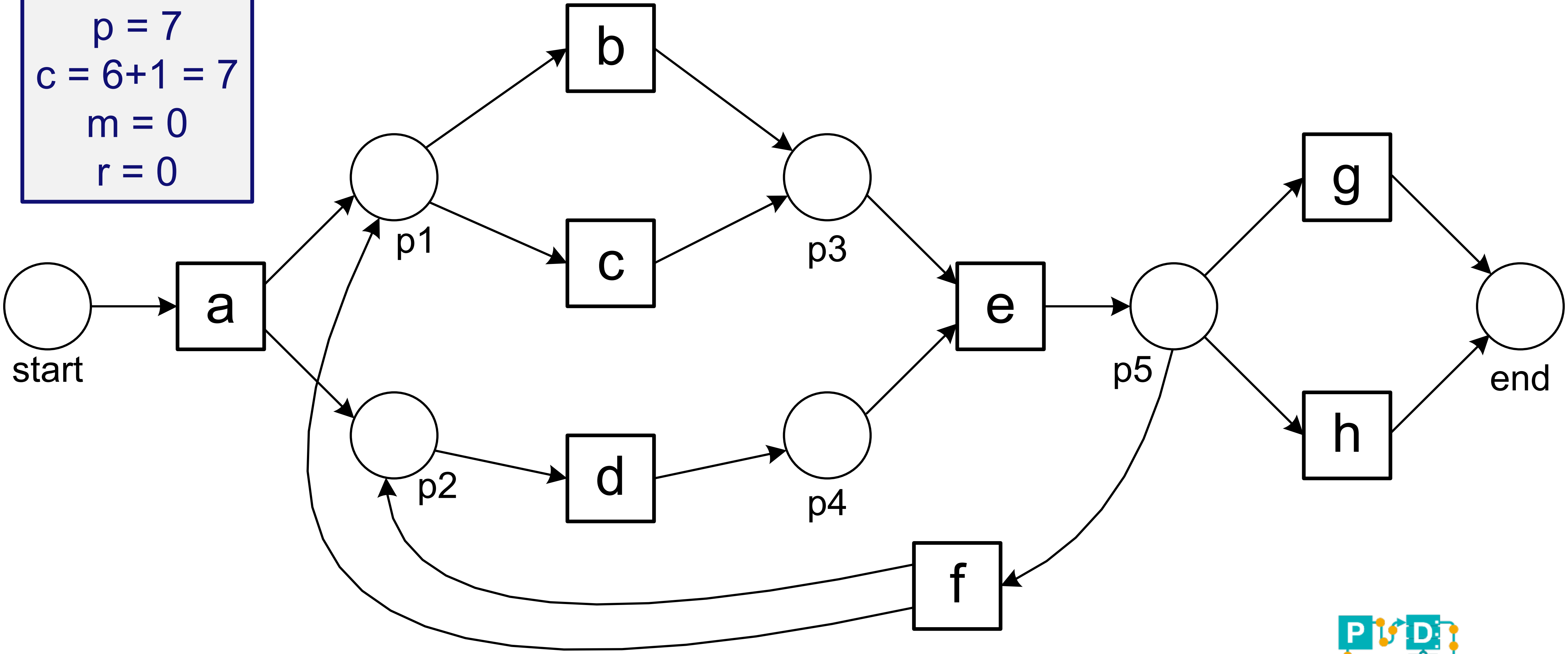
p = 6+1 = 7
c = 5+1 = 6
m = 0
r = 0



Chair of Process and Data Science

Replaying $\sigma_1 = \langle a, c, d, e, h \rangle$

p = 7
c = 6+1 = 7
m = 0
r = 0

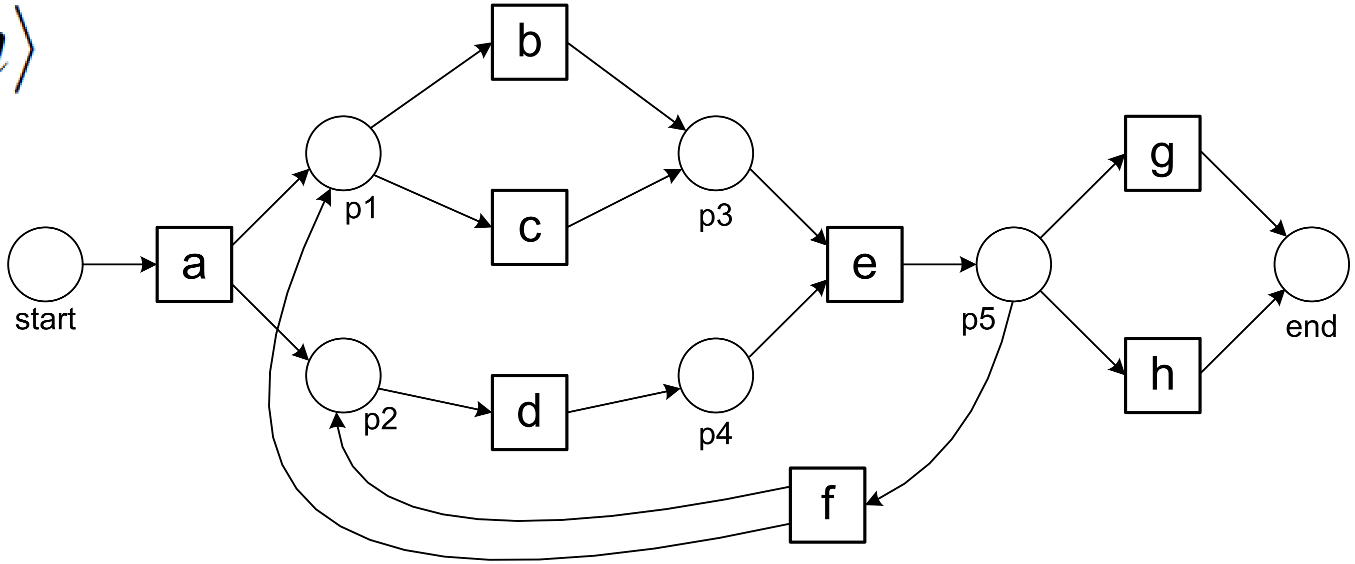


Chair of Process and Data Science

Quantifying fitness at the trace level

p = 7
c = 7
m = 0
r = 0

$$\sigma_1 = \langle a, c, d, e, h \rangle$$

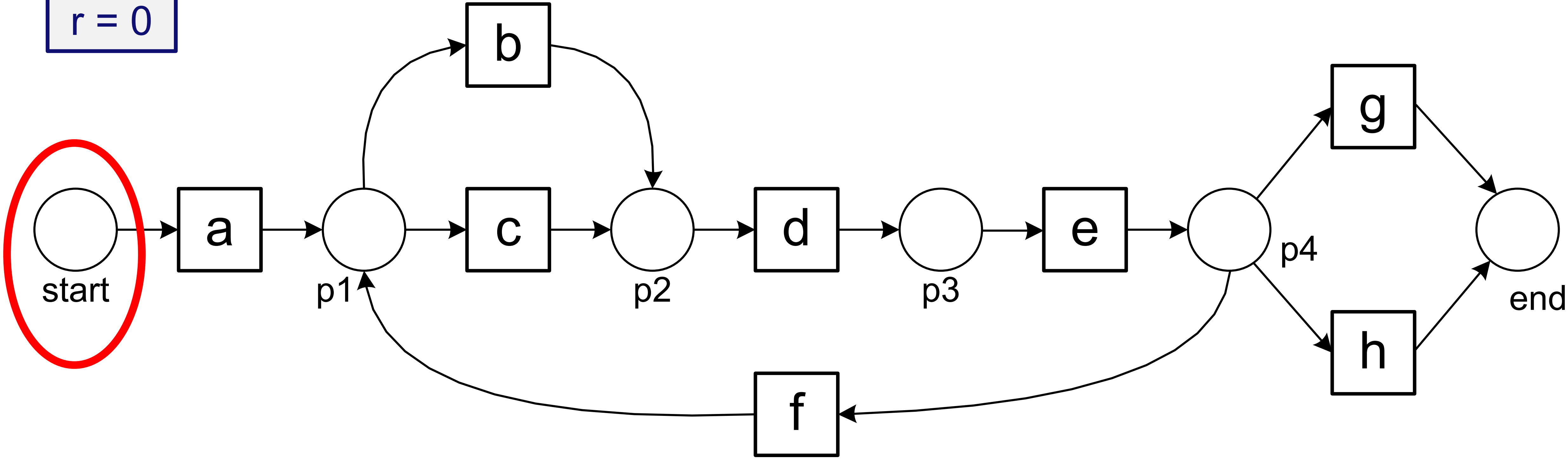


$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{0}{7}\right) + \frac{1}{2}\left(1 - \frac{0}{7}\right) = 1$$

Replaying

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

p = 0
c = 0
m = 0
r = 0

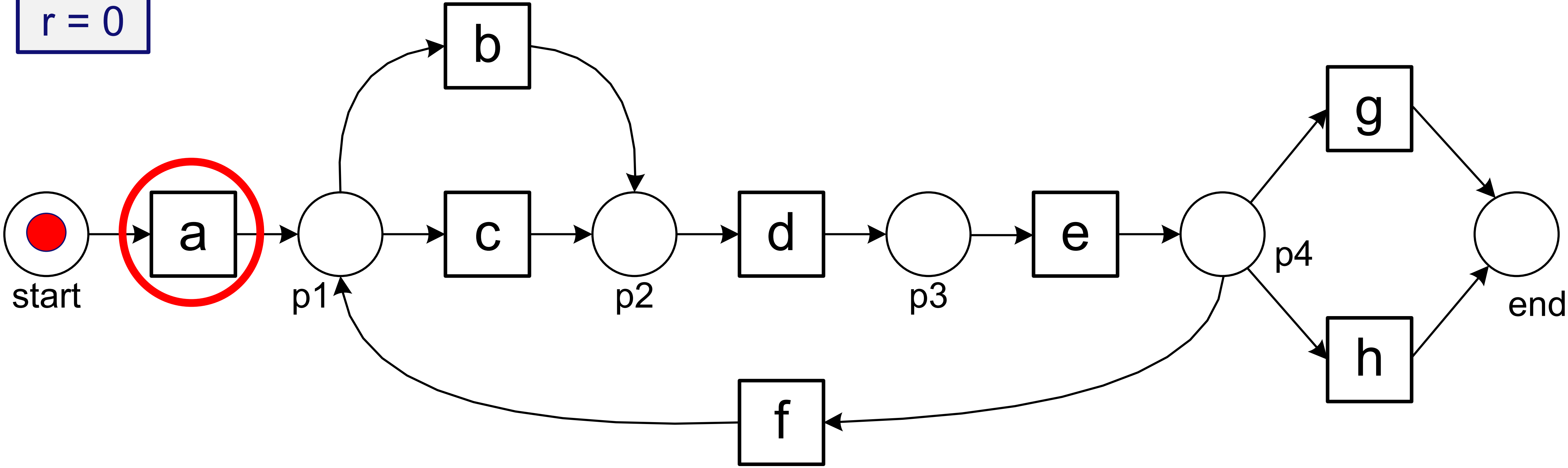


Chair of Process and Data Science

Replaying

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

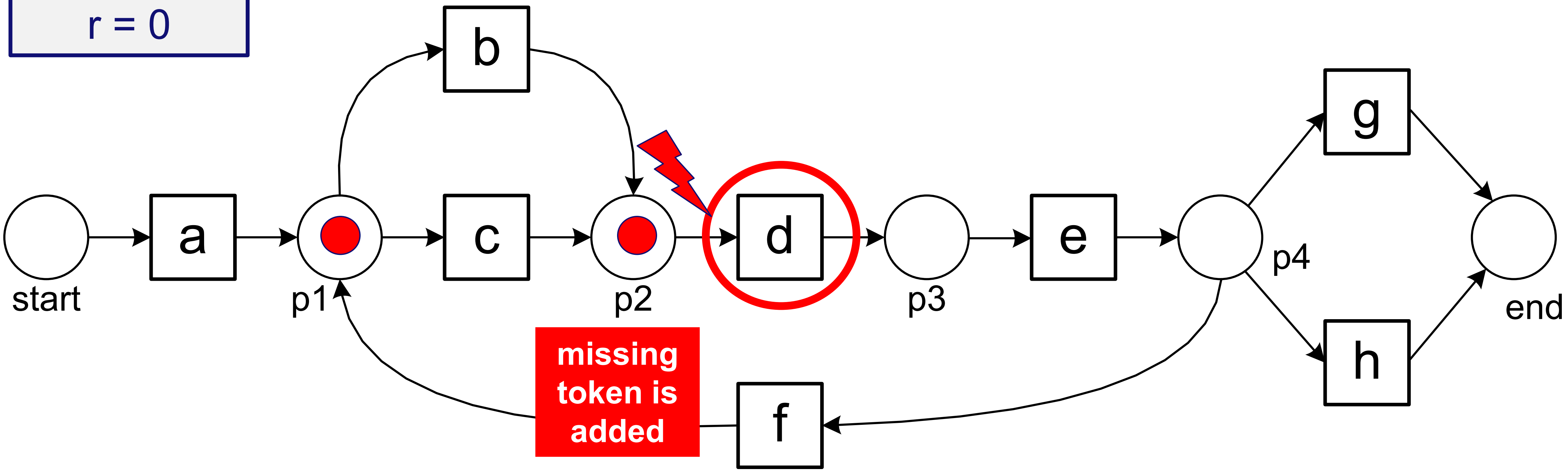
p = 1
c = 0
m = 0
r = 0



Replaying

$\sigma_3 = \langle a, d, c, e, h \rangle$

p = 1+1 = 2
c = 0+1 = 1
m = 0
r = 0

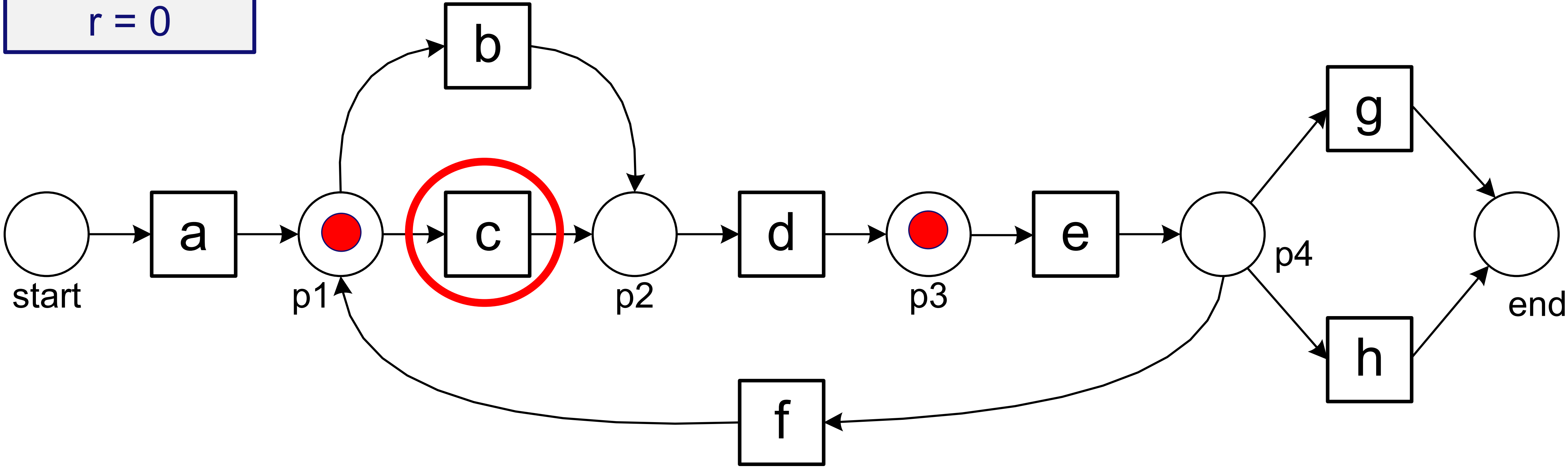


Chair of Process and Data Science

Replaying

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

p = 2+1 = 3
c = 1+1 = 2
m = 0+1 = 1
r = 0

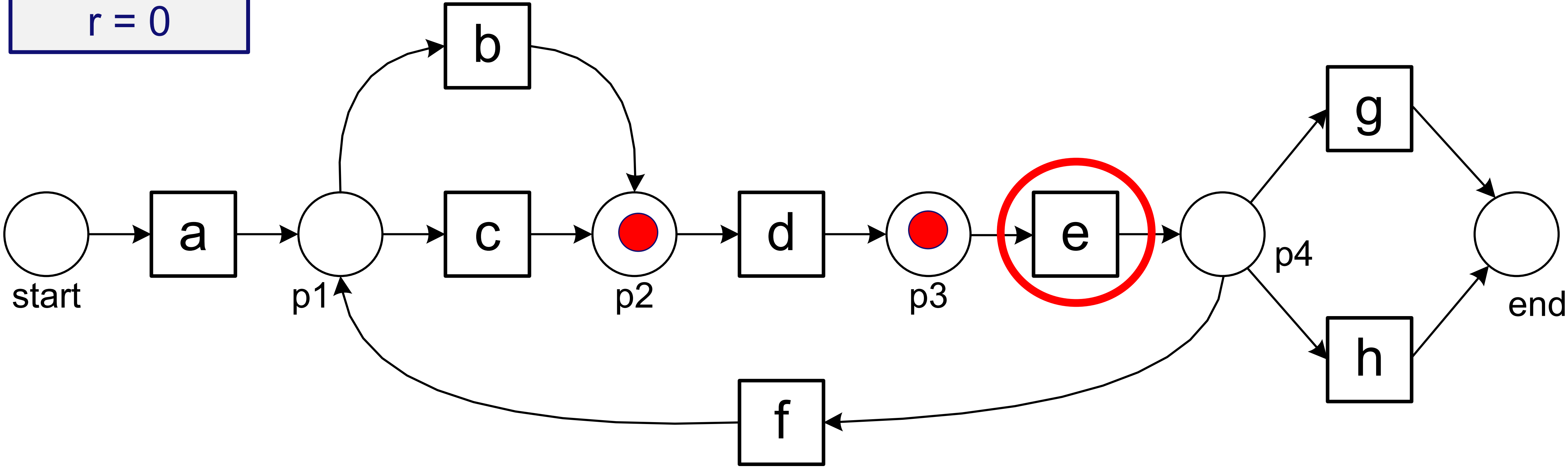


Chair of Process and Data Science

Replaying

$\sigma_3 = \langle a, d, c, e, h \rangle$

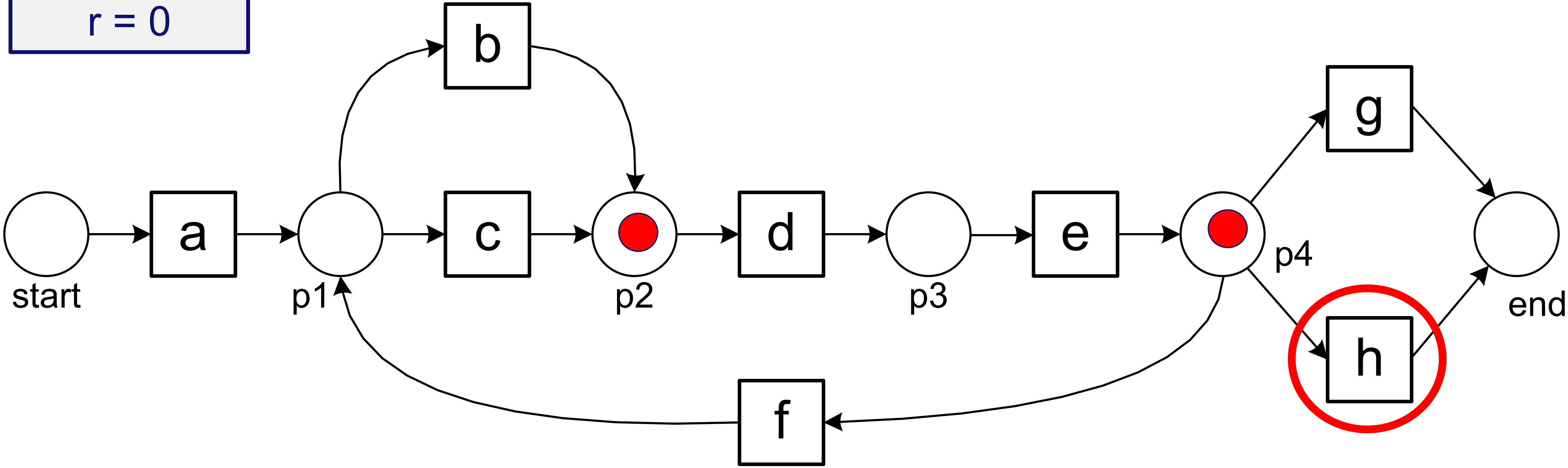
p = 3+1 = 4
c = 2+1 = 3
m = 1
r = 0



Replaying

$\sigma_3 = \langle a, d, c, e, h \rangle$

p = 4+1 = 5
c = 3+1 = 4
m = 1
r = 0

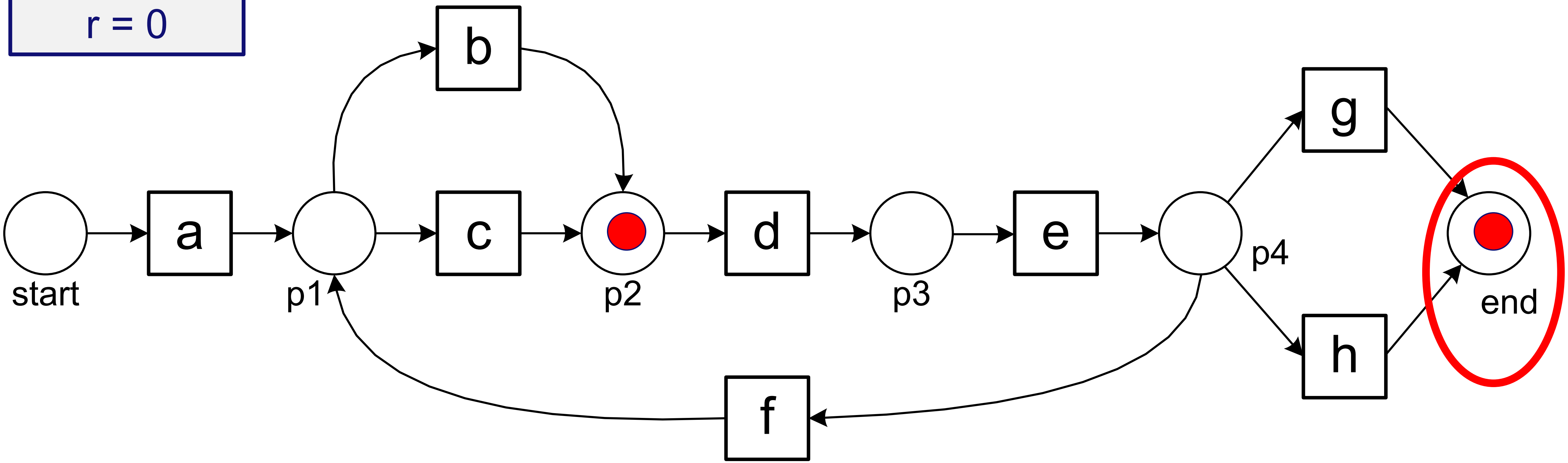


Chair of Process and Data Science

Replaying

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

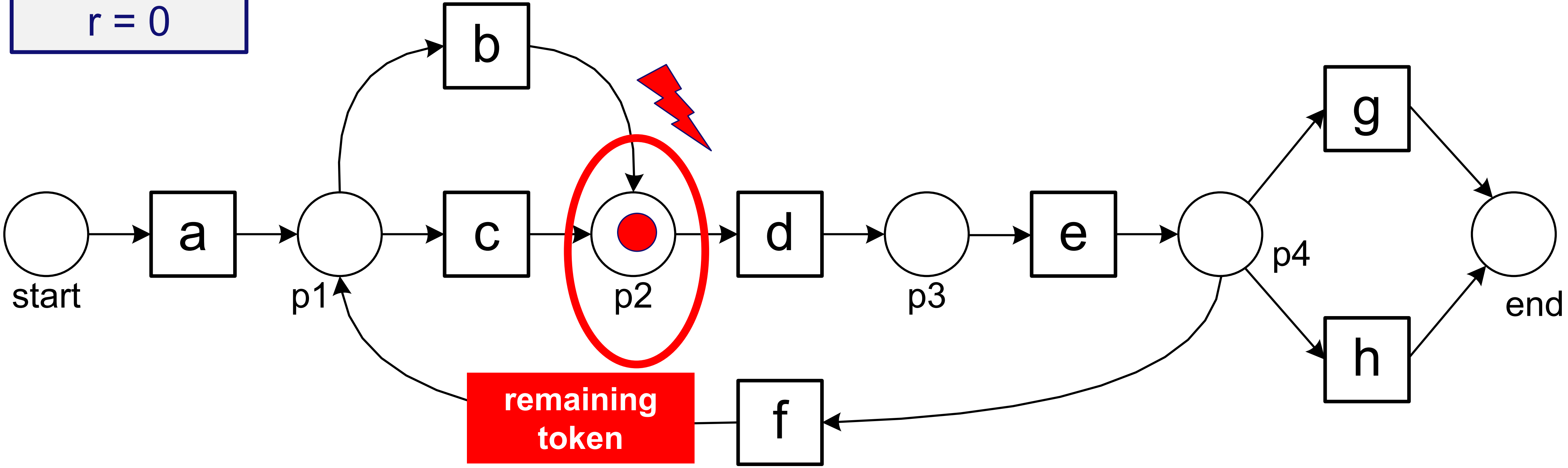
p = 5+1 = 6
c = 4+1 = 5
m = 1
r = 0



Chair of Process and Data Science

Replaying $\sigma_3 = \langle a, d, c, e, h \rangle$

p = 6
c = 5+1 = 6
m = 1
r = 0

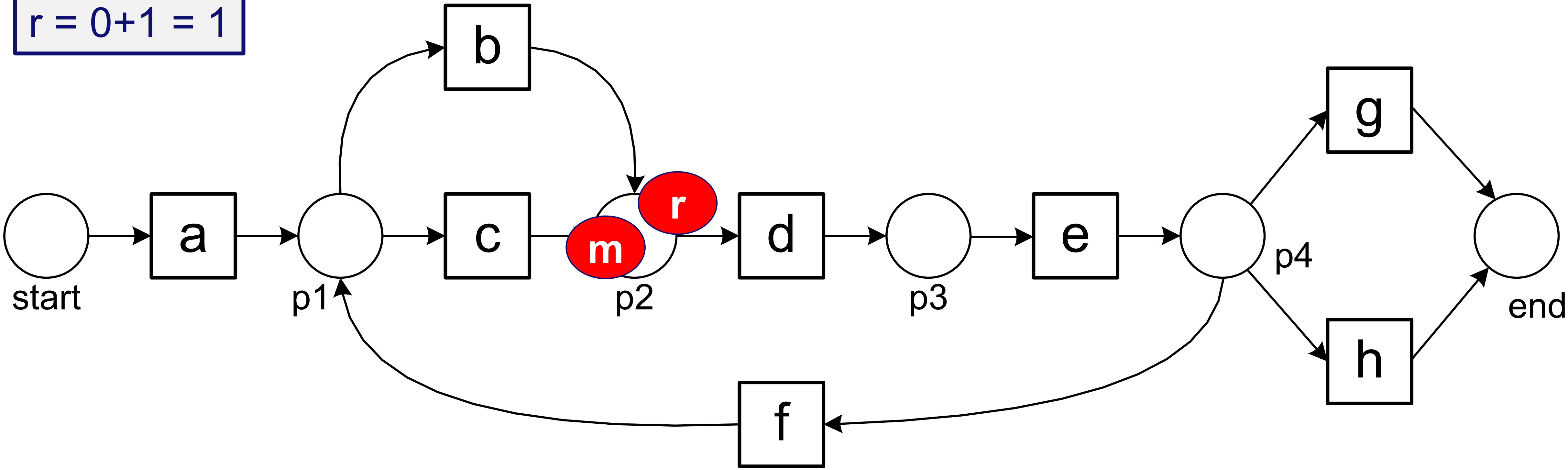


Chair of Process and Data Science

Replaying

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

p = 6
c = 6
m = 1
r = 0+1 = 1

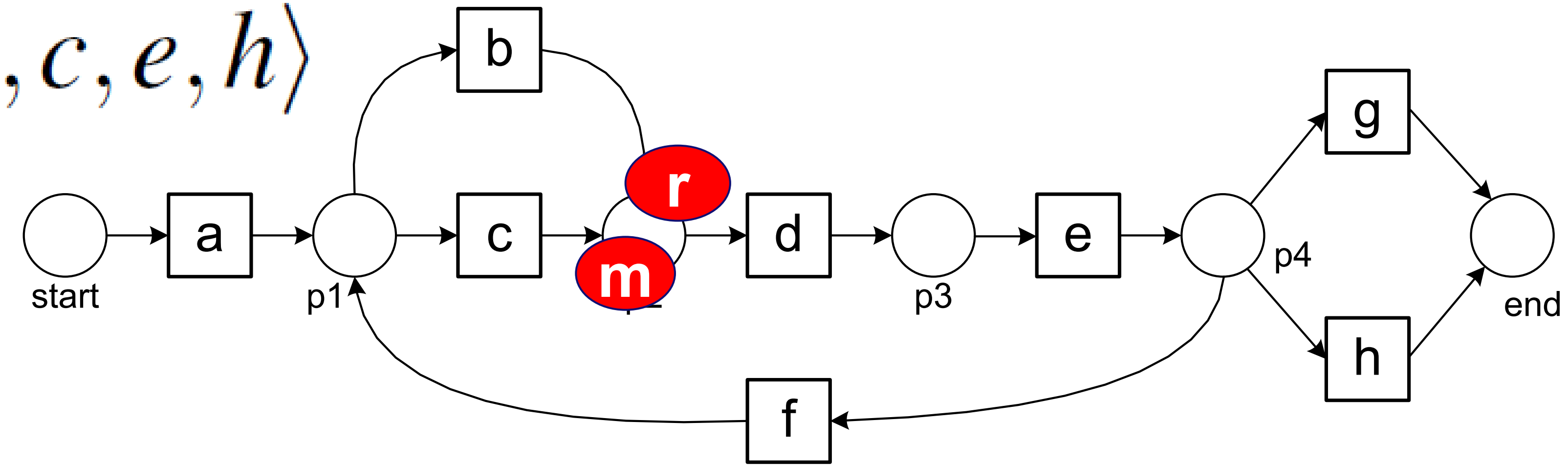


Chair of Process and Data Science

Quantifying fitness at the trace level

p = 6
c = 6
m = 1
r = 1

$$\sigma_3 = \langle a, d, c, e, h \rangle$$



$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.8333$$

Fitness at the log level

$$fitness(L,N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}}\right)$$

missing tokens

consumed tokens

remaining tokens

produced tokens

Looks scar
just needs t
sums of p, c, m, and r
over the multiset of
traces in de

Chair of Process and Data Science

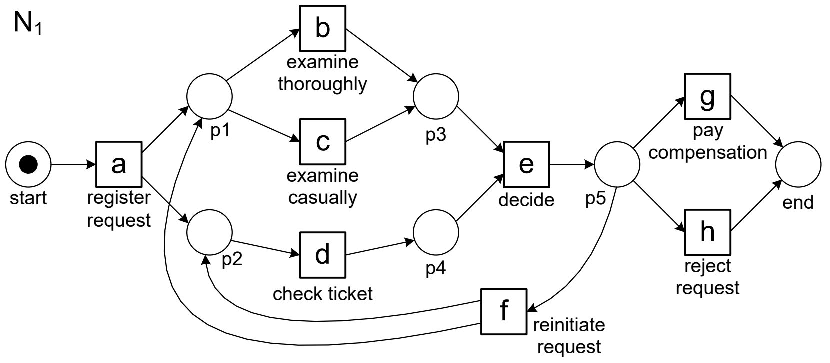
#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

?

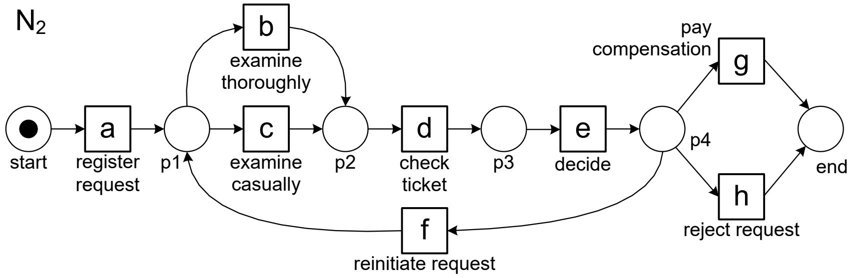
$$fitness(L,N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}}\right)$$

$$fitness(L_{full}, N_1) = 1$$
$$fitness(L_{full}, N_2) = 0.9504$$
$$fitness(L_{full}, N_3) = 0.8797$$
$$fitness(L_{full}, N_4) = 1$$

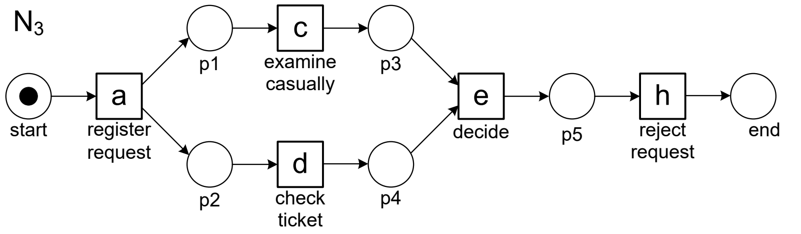
N_1



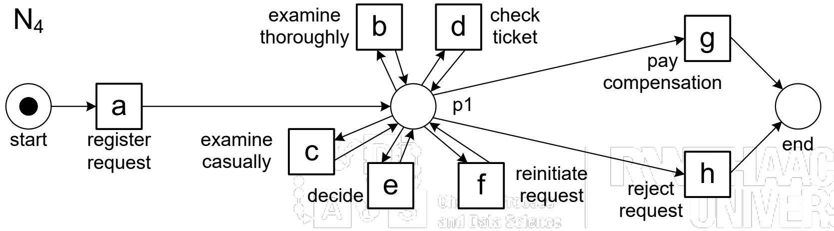
N_2



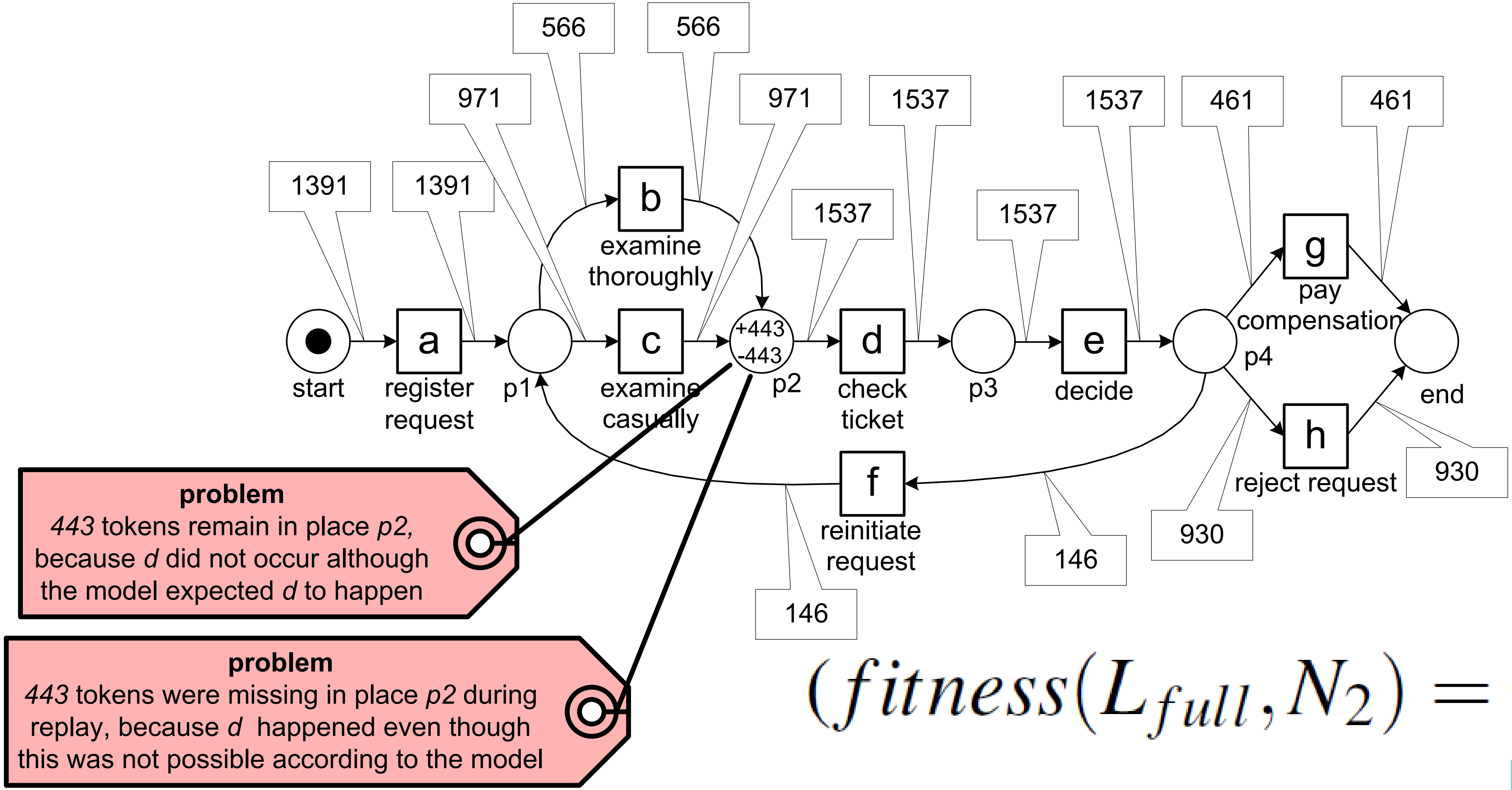
N_3



N_4

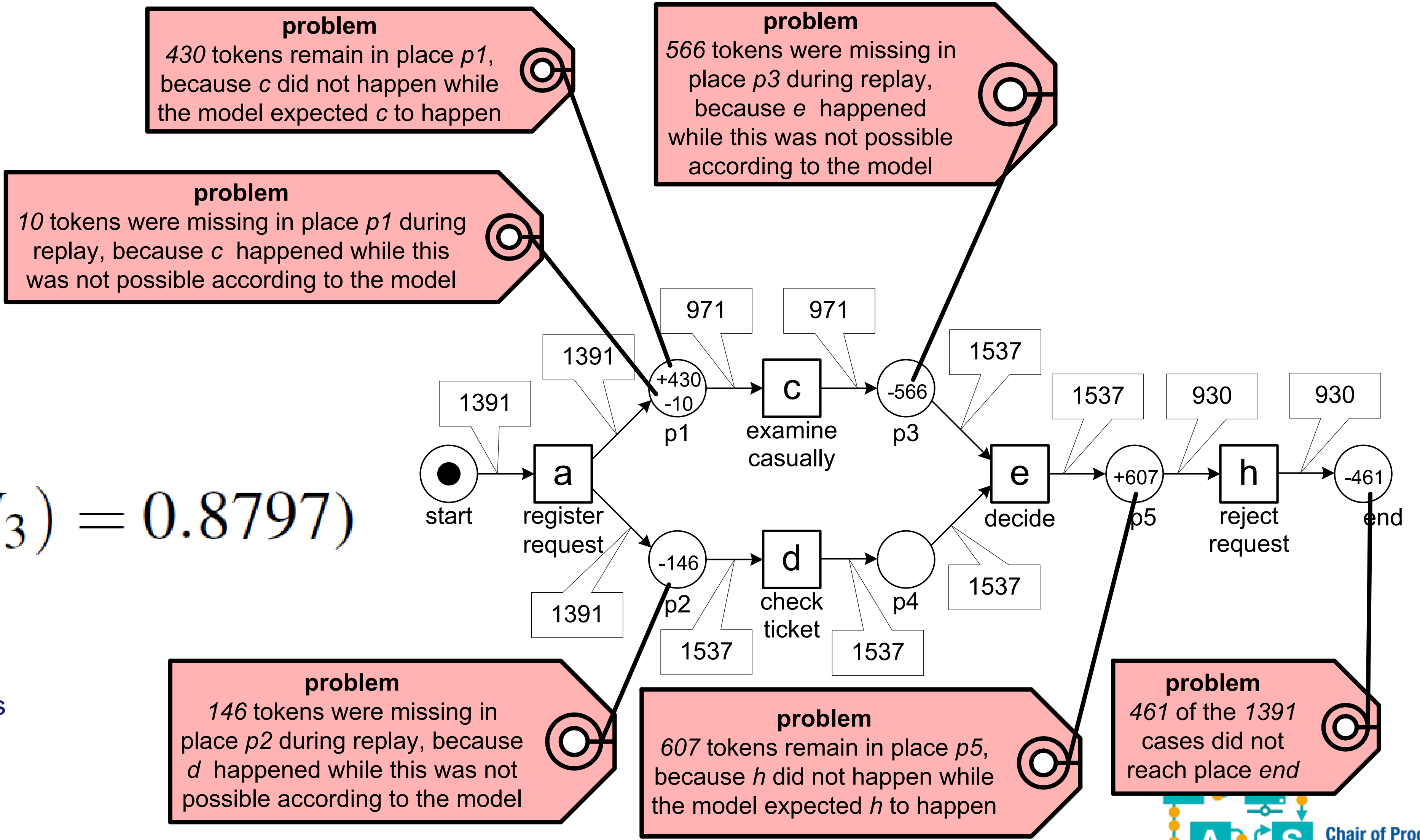


Diagnostics



$$(fitness(L_{full}, N_2) = 0.9504)$$

Diagnostics



$$(fitness(L_{full}, N_3) = 0.8797)$$

Remark: If event log and model consider different sets of activities, this should be addressed first. Activities in log, but not in model, are simply ignored.

Limitations

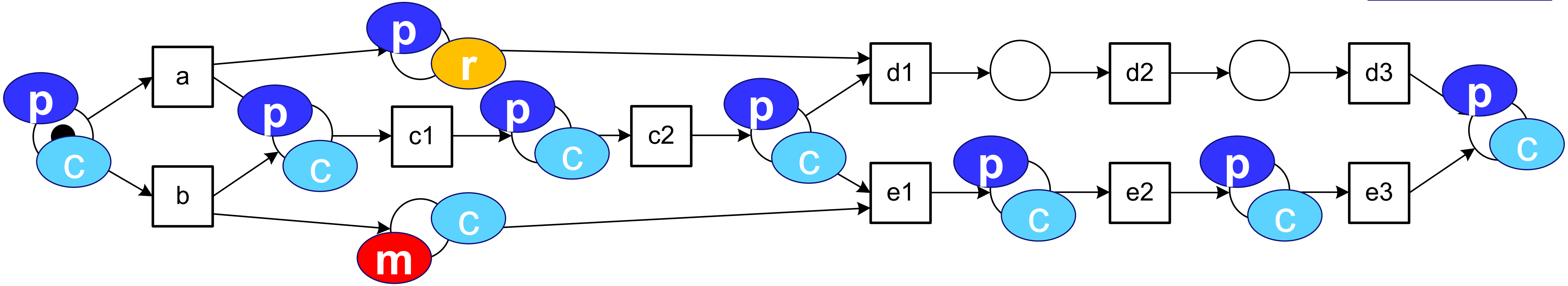
- Basic replay approach assumes **visible & uniquely labeled** transitions.
- Implementations need to use **heuristics** to deal with silent transitions and multiple transitions having the same label.
- Conformance values sometimes **too optimistic** due to "token flooding".
- Local decision making may lead to misleading results.

Chair of Process and Data Science

Local decision making is not enough ...

p = 8
c = 8
m = 1
r = 1
f = 0.875

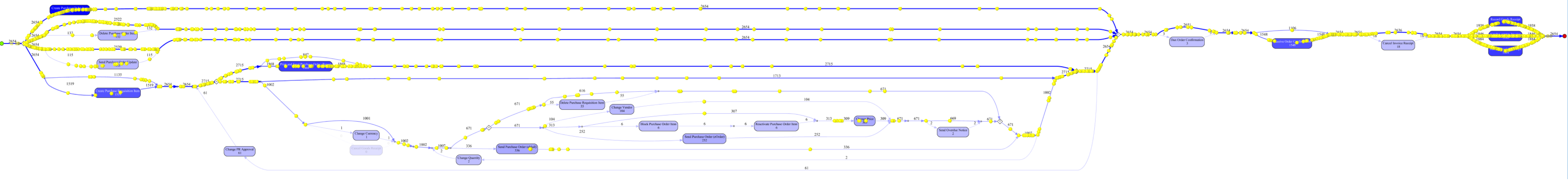
⟨a,c1,c2,e1,e2,e3⟩



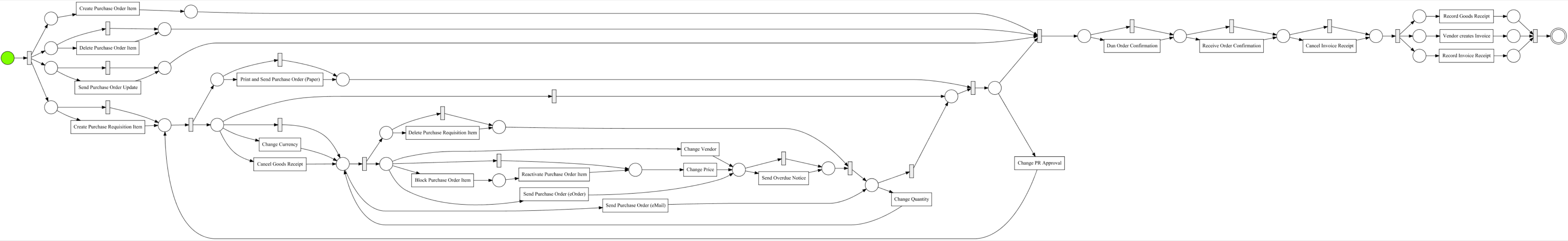
- Replay technique does **not** provide a corresponding path through the model (vital for conformance/performance analysis and other diagnostics).
- We would like to see the "closest path", i.e., ⟨b,c1,c2,e1,e2,e3⟩.

Chair of Process and Data Science

Challenge: Silent and duplicate transitions

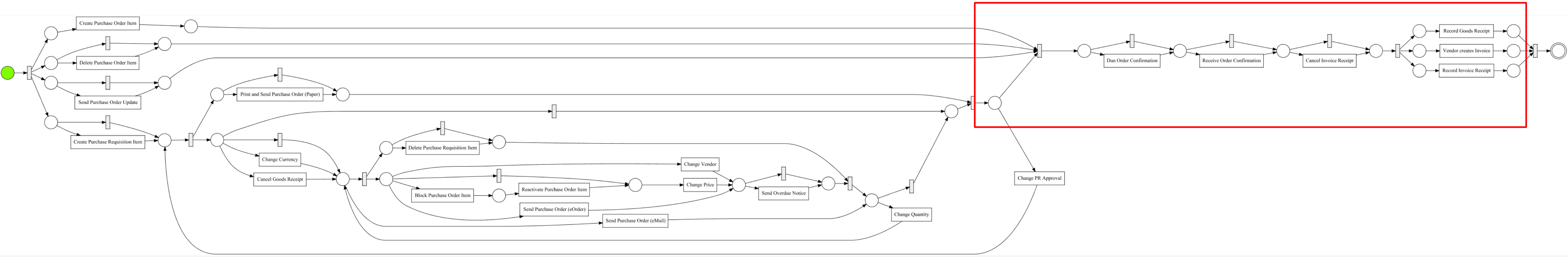


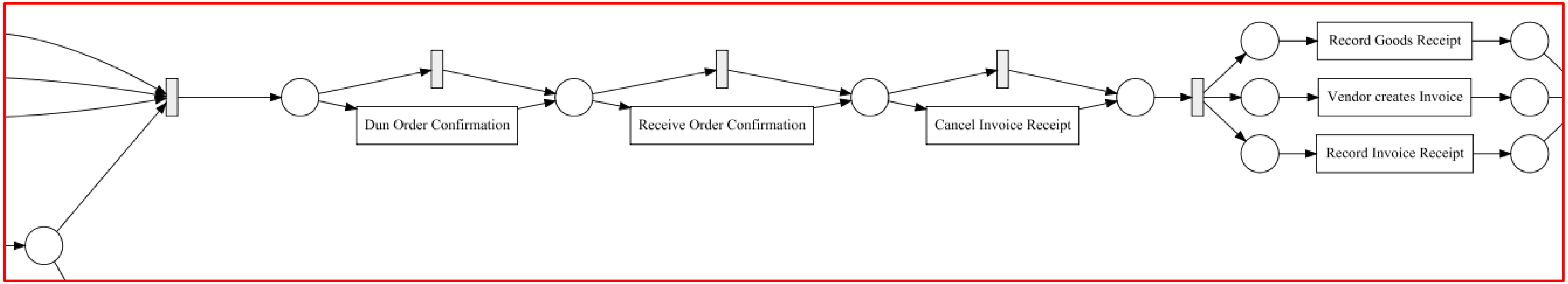
The Inductive Miner may produce many silent transitions …



Chair of Process and Data Science

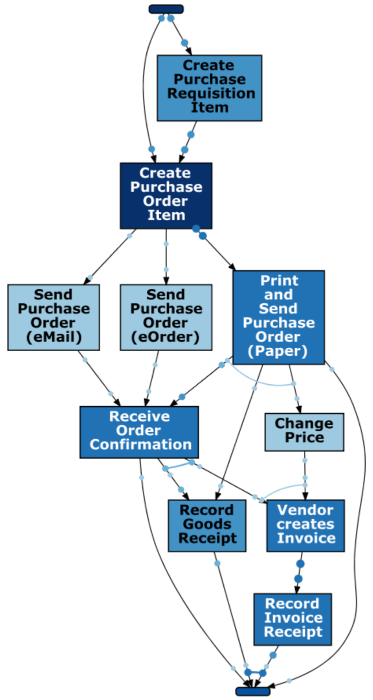
Challenge: Silent and duplicate transitions



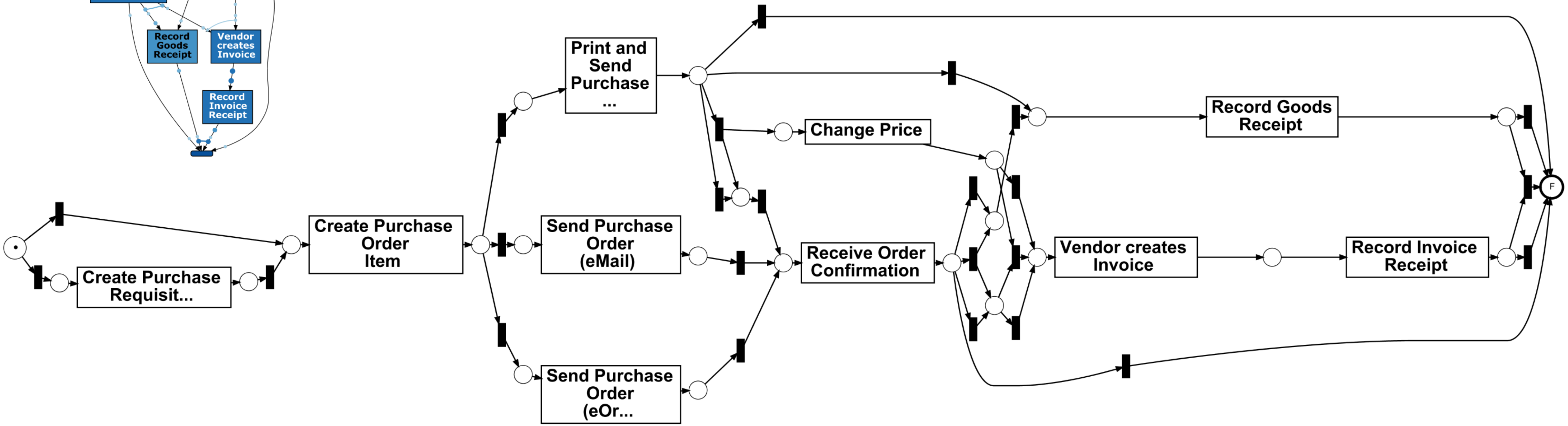


Chair of Process and Data Science

Challenge: Silent and duplicate transitions



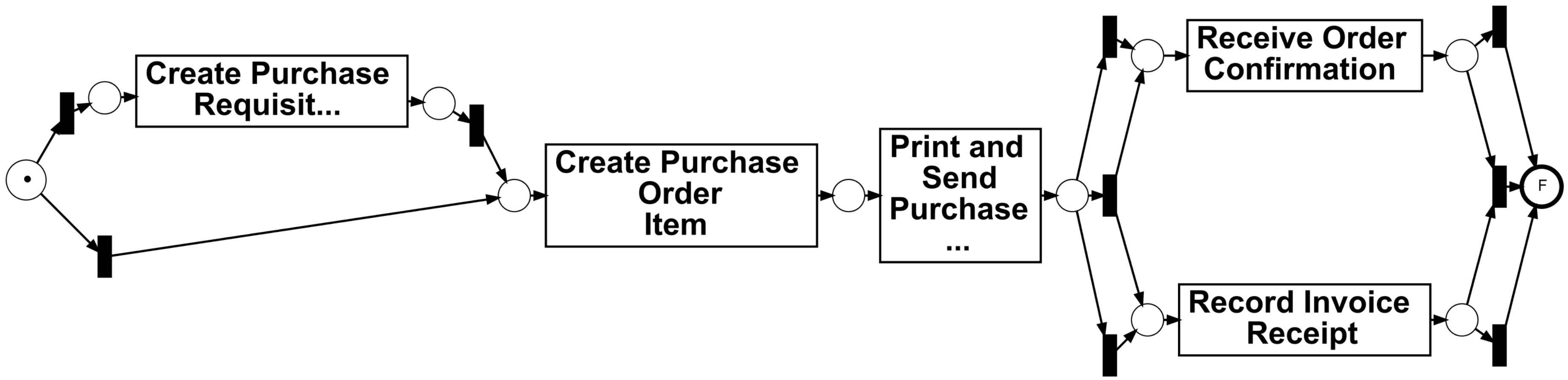
The Heuristic Miner may also produce many silent transitions



©

Challenge: Silent and duplicate transitions

- Some form of state-space exploration is needed to decide which silent transitions to fire.
- If a trace is non-fitting, the set of possibilities grows further.



Alignments

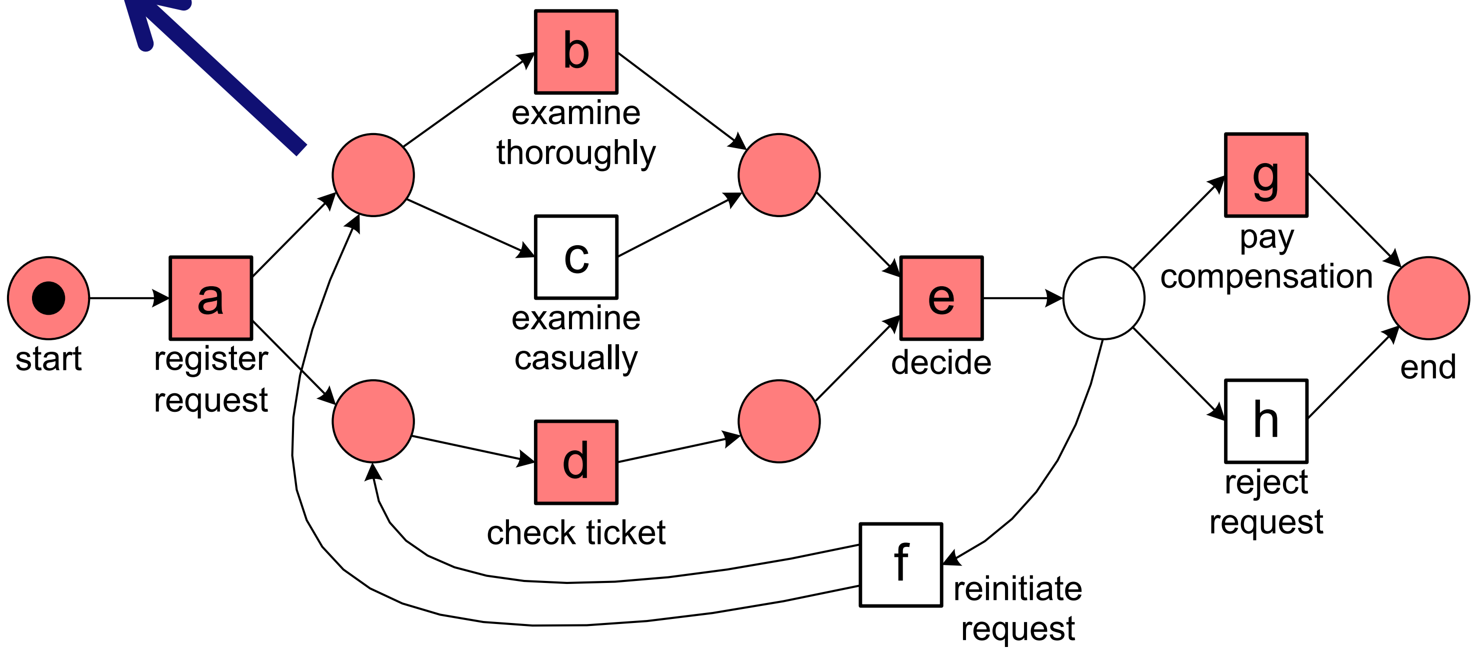
Requirements

- Conformance checking should not impose restrictions on the process notation (e.g., silent transitions and two transitions with the same label should be possible).
- Two semantically equivalent models should have the same conformance value. No heuristics!
- Should provide a "closest matching path" through the process model for any trace in the event log.
  - Also required for performance analysis!
  - Beyond the analysis of replay fitness (advanced diagnostics, precision, generalization, etc.).

Alignments

a	b	d	e	g
a	b	d	e	g

⟨a,b,d,e,g⟩

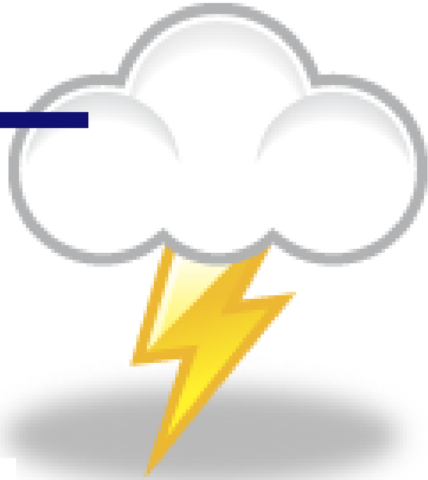


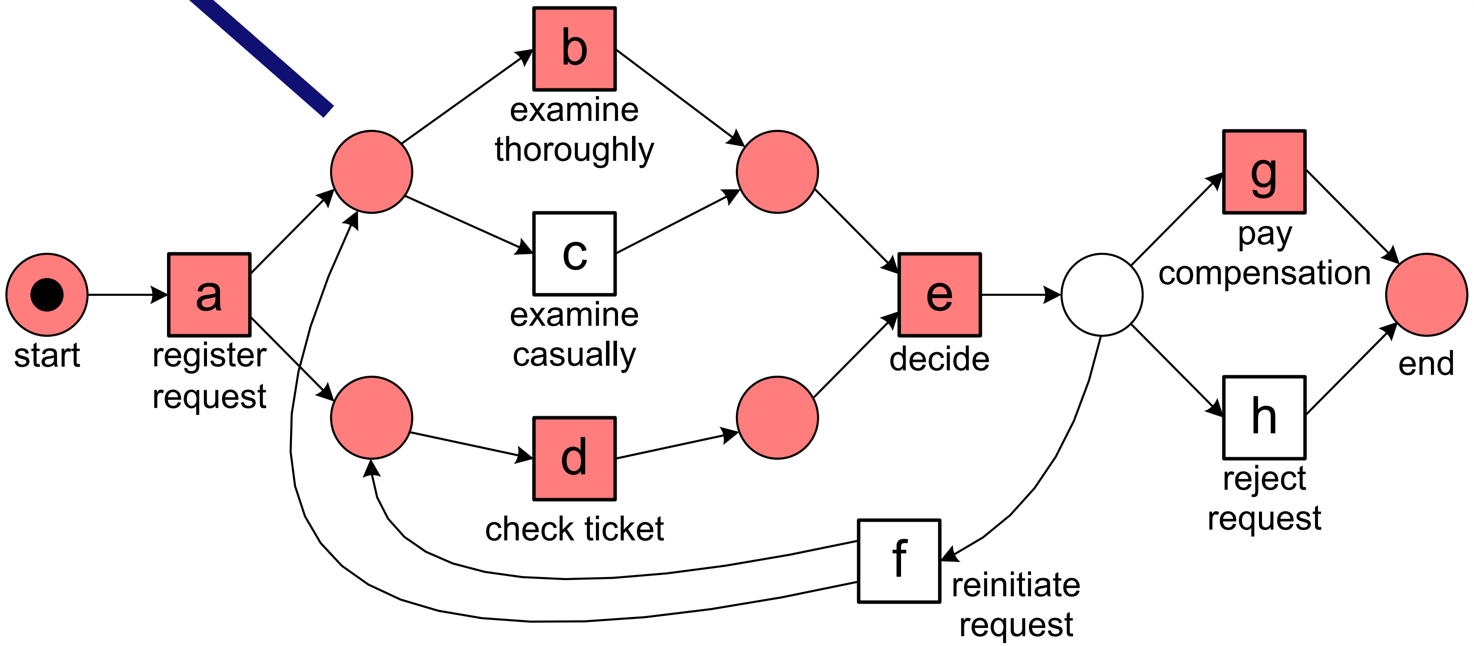
#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

Alignments

a	»	d	e	g	h
a	b	d	e	g	»

⟨a,d,e,g,h⟩





Terminology

alignment (sequence of moves)

move in log only

move in model

...nous move (move in both)

a	[illegible]	[illegible]	d	e	g
[illegible]	»	c	d	e	g

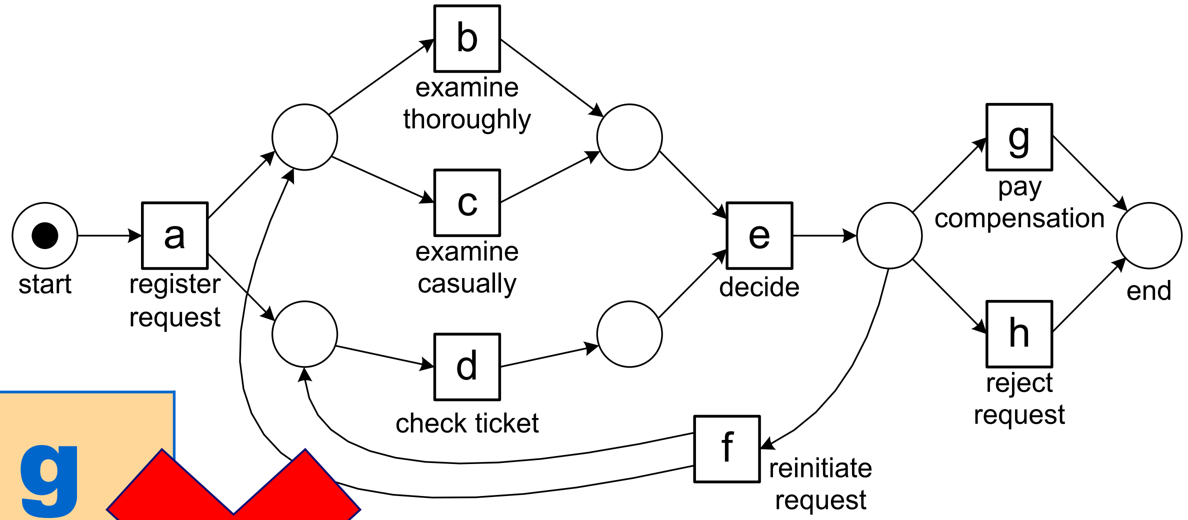
Independent of process model notation!

- Projectio... row (remove "no moves") corresponds to th... the event log.
- ...ion on bottom row (remove "no moves") ...rresponds to a run of the model.

Chair of Process and Data Science

Optimal alignment for ⟨a,b,d,e,g⟩

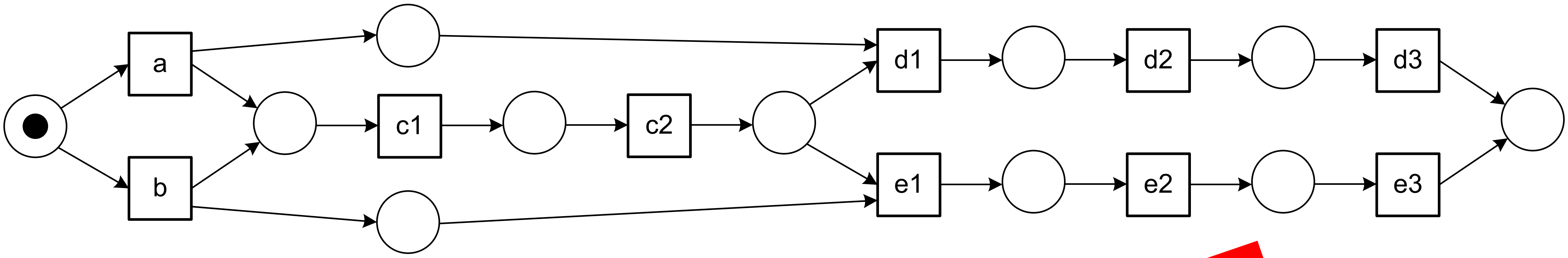
a	b	d	e	g
a	b	d	e	g



a	b	»	d	e	g
a	»	c	d	e	g

a	b	d	e	g	»	»	»	»	»
»	»	»	»	»	a	c	d	e	g

Optimal alignment for ⟨a,c1,c2,e1,e2,e3⟩ ?



a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	[illegible]	[illegible]

a	c1	c2	[illegible]	[illegible]	»	e1	e2	e3
a	c[illegible]	[illegible]	d1	d2	d3	»	»	»

Depends on cost function!

Chair of Process and Data Science

Standard cost function

count »'s in alignment

1

...	a	...
...	»	...

1

...	»	...
...	a	...

0

...	a	...
...	a	...

inf

...	a	...
...	b	...

Chair of Process and Data Science

Using the standard cost f

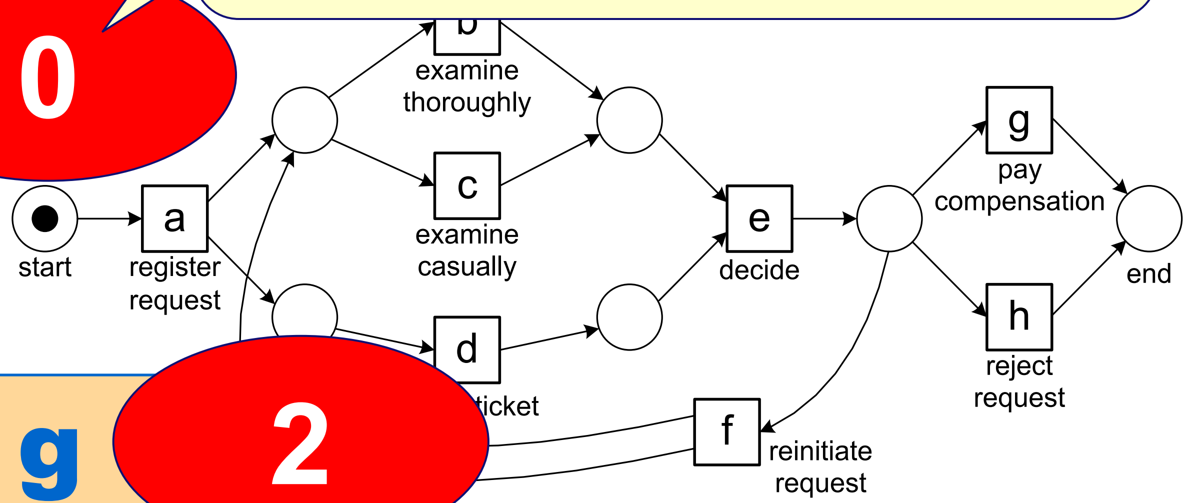
optimal
there is no other alignment that has lower costs

a	b	d	e	g
a	b	d	e	g

0

a	b	»	d	e	g
a	»	c	d	e	g

2



a	b	d	e	g	»	»	»	»	»
»	»	»	»	»	a	c	d	e	g

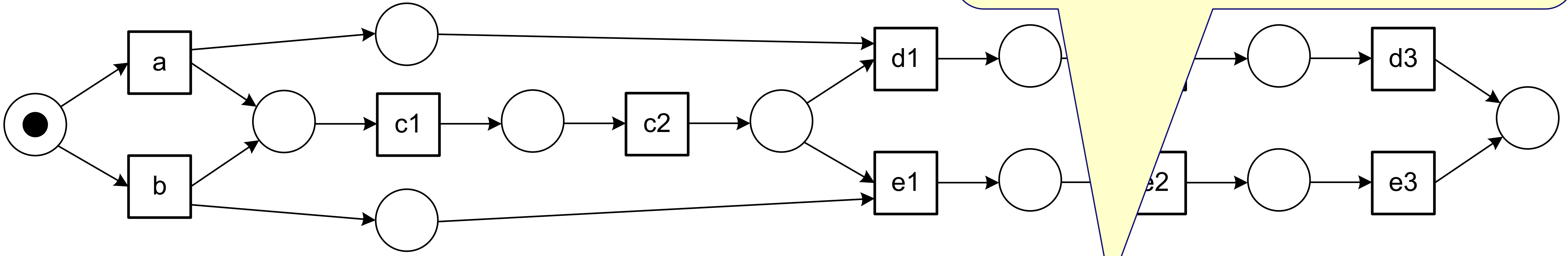
10

Using the standard cost fu

trace in log: ⟨a,c1,c2,e1,e2,e3⟩

optimal

there is no other alignment that has lower costs



a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

2

a	c1	c2	»	»	»	e1	e2	e3
a	c1	c2	d1	d2	d3	»	»	»

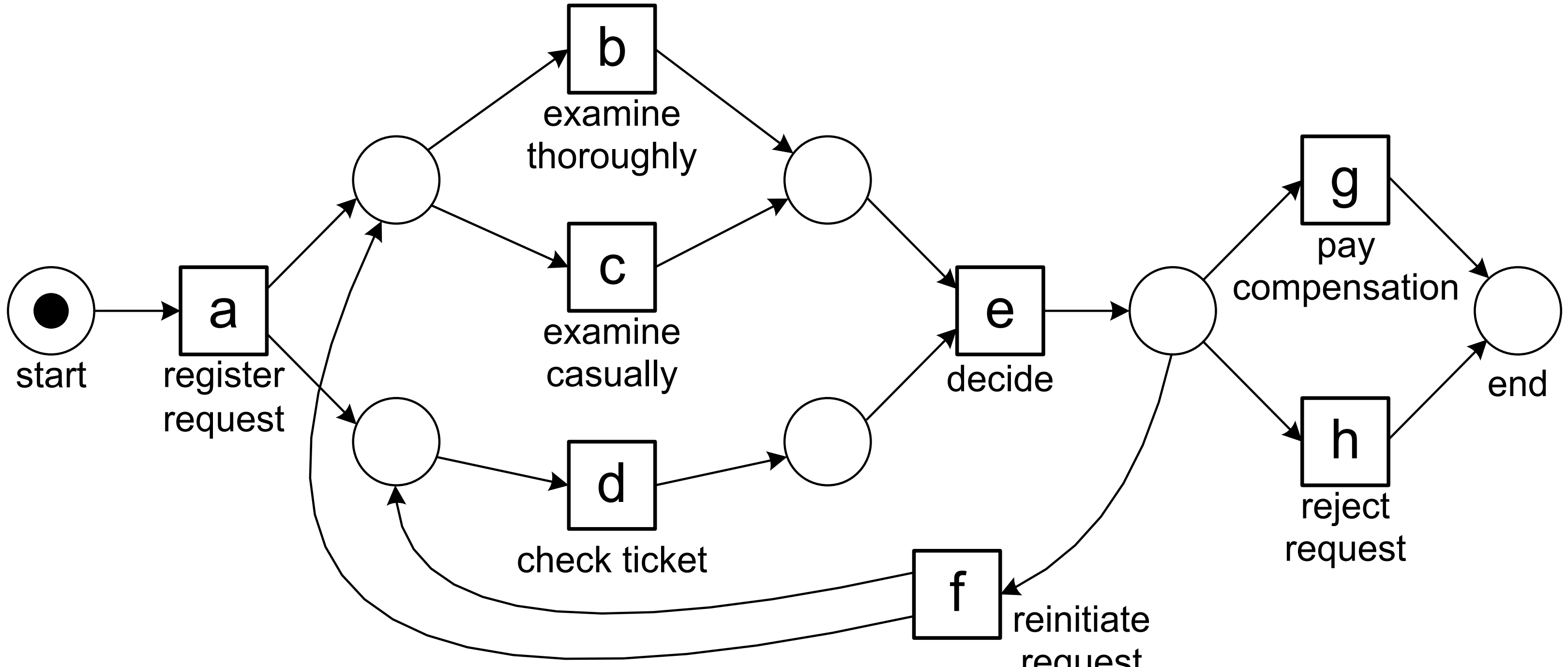
6

Chair of Process and Data Science

Optimal alignment for ⟨a,b,e,f,d,e,g⟩ (1/2)

a	b	e	f	d	e	g
a	b	»	»	d	e	g

2



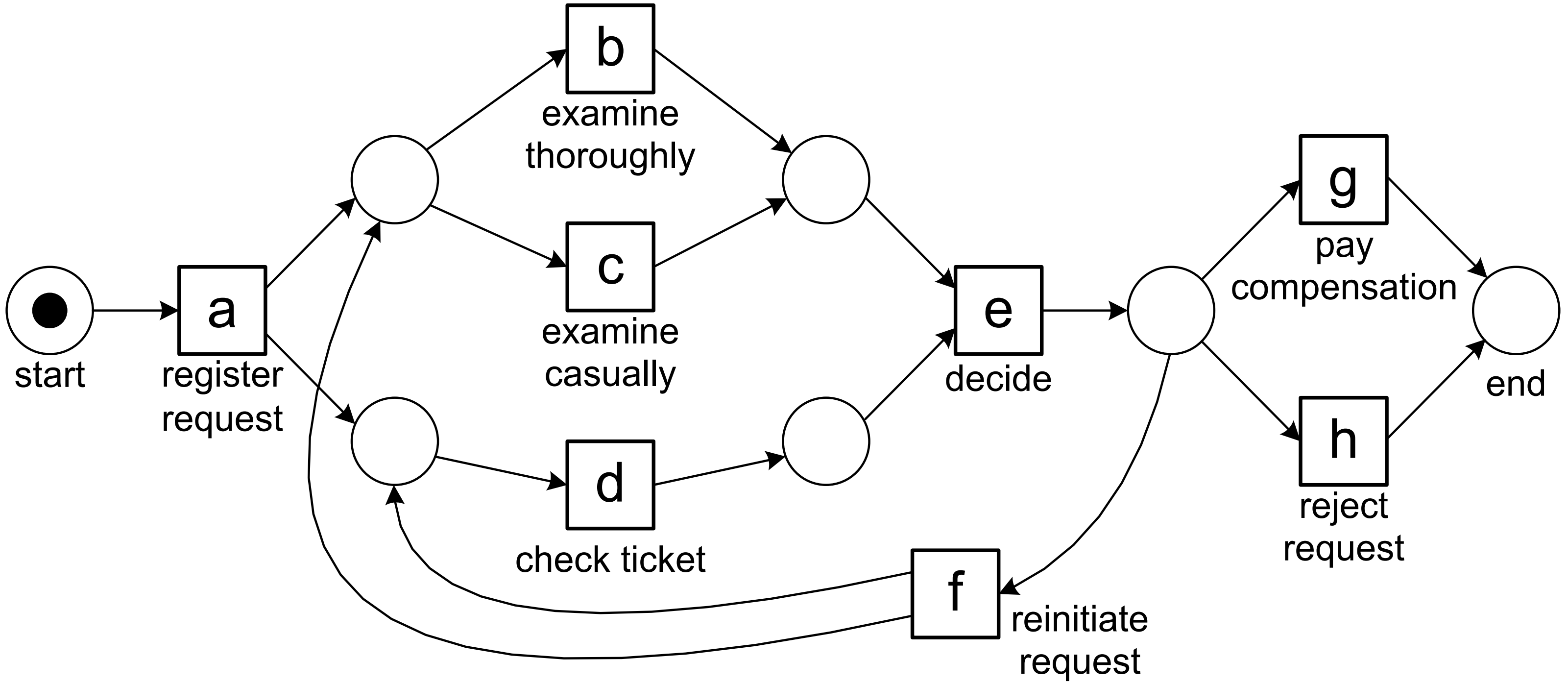
loop is not taken: e and f in event log are discarded

Chair of Process and Data Science

Optimal alignment for ⟨a,b,e,f,d,e,g⟩ (2/2)

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	b	e	g

2



loop is taken: d and b are missing in event log

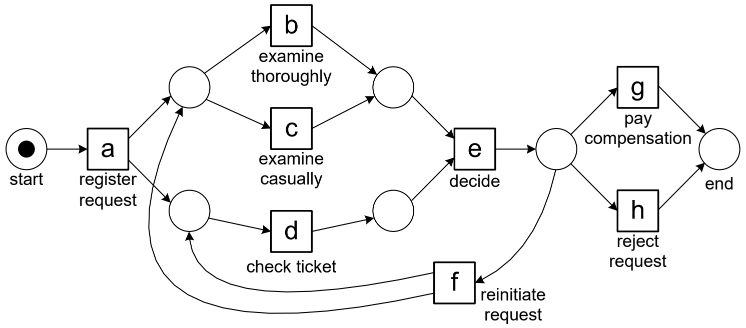
Chair of Process and Data Science

Not one unique optimal alignment

a	b	»	e	f	d	»	e	g	2
a	b	d	e	f	d	b	e	g	

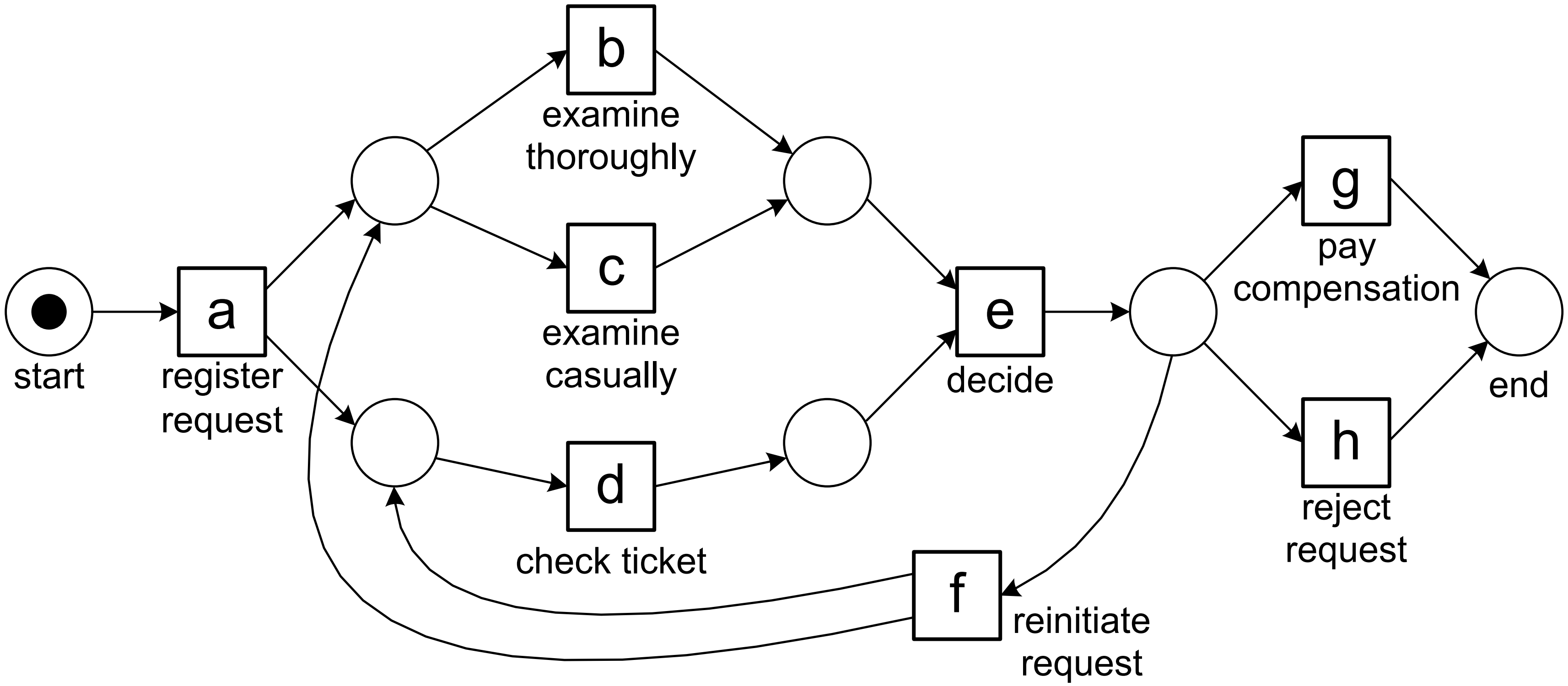
a	b	e	f	d	e	g	2
a	b	»	»	d	e	g	

...



Chair of Process and Data Science

Question: How many optimal alignments are there for ⟨a,b,e,f,d,e,g⟩ ?



⟨a,b,e,f,d,e,g⟩

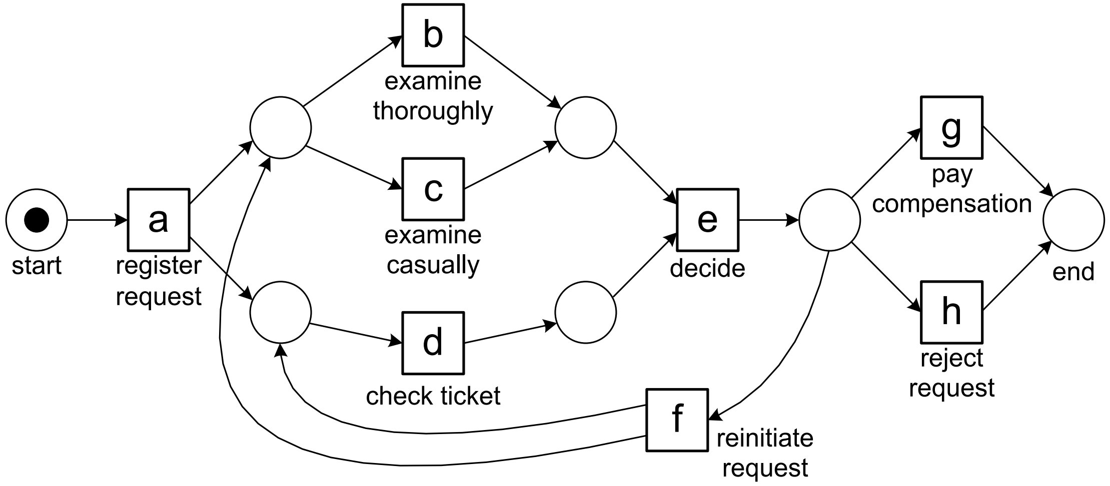
Chair of Process and Data Science

Answer: 9

1+(2x2)+(2x2)=9 optimal alignments having cost 2

a	b	e	f	d	e	g
a	b	»	»	d	e	g

1x



move in model can be reordered in concurrent part

4x

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	b	e	g

4x

a	b	»	e	f	d	»	e	g
a	b	d	e	f	d	c	e	g

Chair of Process and Data Science

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

Chair of Process and Data Science

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

similar activities (lower costs for related activities)

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$50	...

resource-related conformance costs
(done by someone that does or does not have the specified role)

Chair of Process and Data Science

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

time-related conformance costs (activity should happen within a preset deadline)

Chair of Process and Data Science

Any cost structure is possible

...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

data-related conformance costs (routing condition is violated, e.g., path only for more valuable orders)

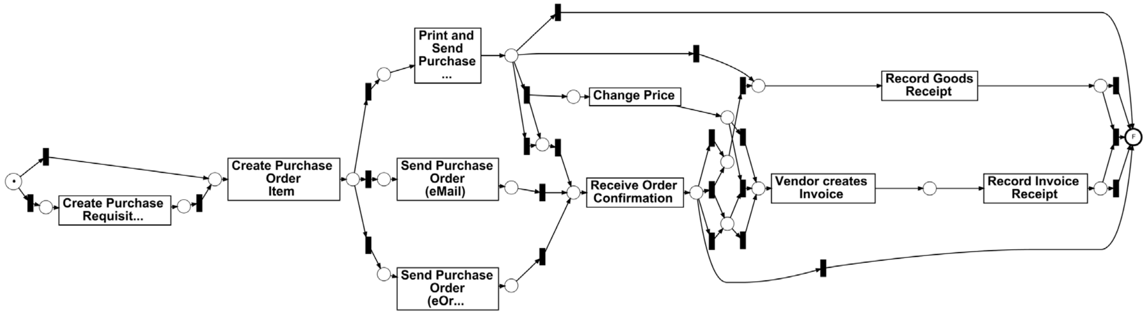
Chair of Process and Data Science

Any cost structure is possible

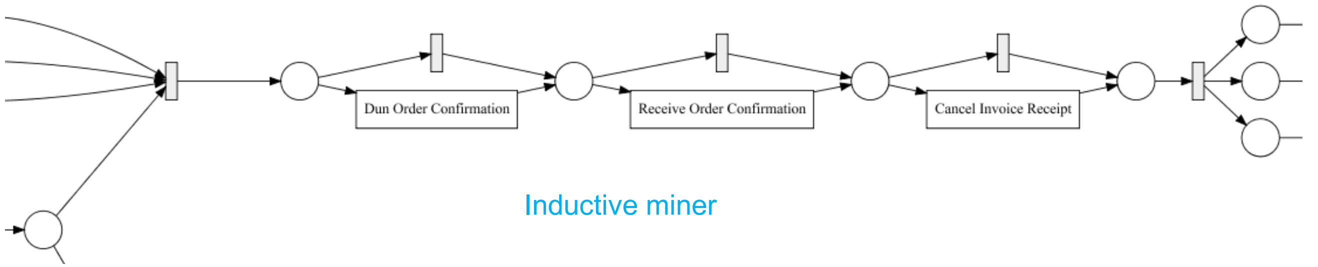
...	send-letter(John,4 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

risk-related conformance costs, context-dependent conformance costs, …

Dealing With Silent transitions



Heuristic miner



Inductive miner

- **Two options:**
  - **Remove these transitions when considering the model's behavior** **(Model reduction: any Nondeterministic Finite Automaton (NFA) with silent activities, can be translated into an NFA without silent activities, which, in turn, can be translated into a Deterministic Finite Automaton (DFA).)**
  - **The corresponding moves on model have cost 0** **(or a very small epsilon in case of loops)**

Chair of Process and Data Science

Computing fitness

C… n optimal alignment with a

…nd s…

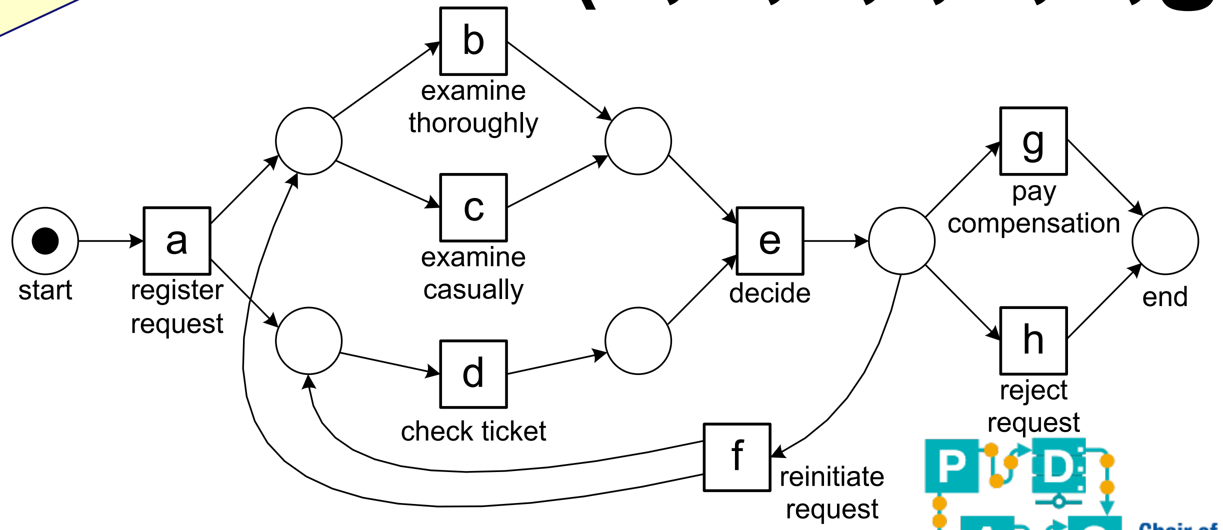
mo… model.

all events cause a move in log only

shortest path from initial state to final state

⟨a,b,e,f,d,e,g⟩

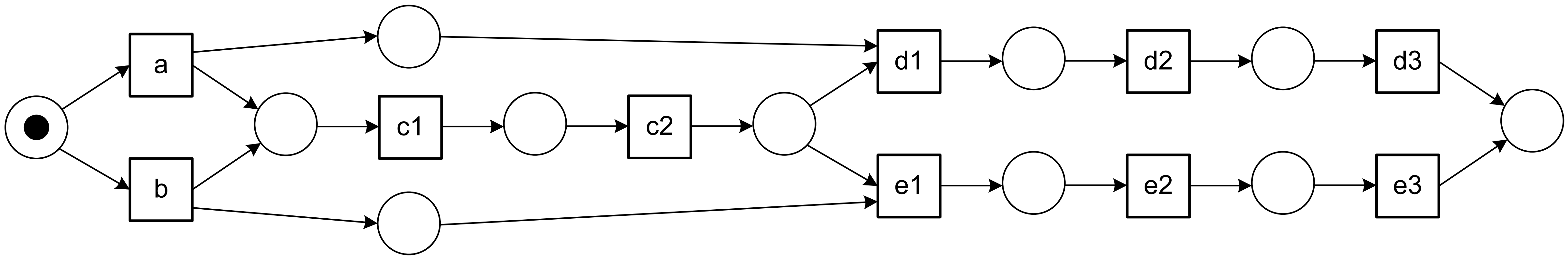
$$1 - \frac{2}{7+5} = 0.833$$



Chair of Process and Data Science

Question: Compute alignment-based fitness

⟨a,c1,c2,e1,e2,e3⟩



Answer

⟨a,c1,c2,e1,e2,e3⟩

cost of optimal alignment = 2

all ev... move in log only

worst-case scenario

...rtest path from initial state to final state

a		c1	c2		e2	
»		c1	c2	e	e2	e

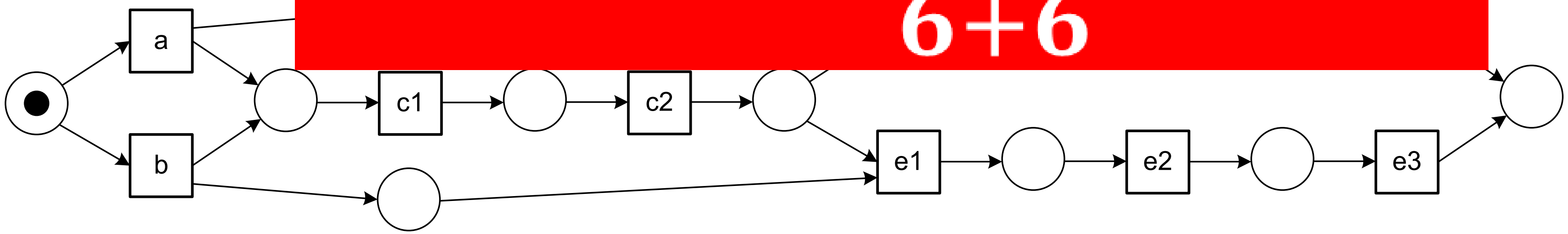
a	c1	c2	e1	e2	e3	»	»	»	»	»	»
»	»	»	»	»	»	a	c1	c2	d1	d2	d3

Chair of Process and Data Science

Answer

$$\text{fitness} = 1 - \frac{2}{6+6} = 0.833$$

⟨a,c1,c2,e1,e2,e3⟩



2

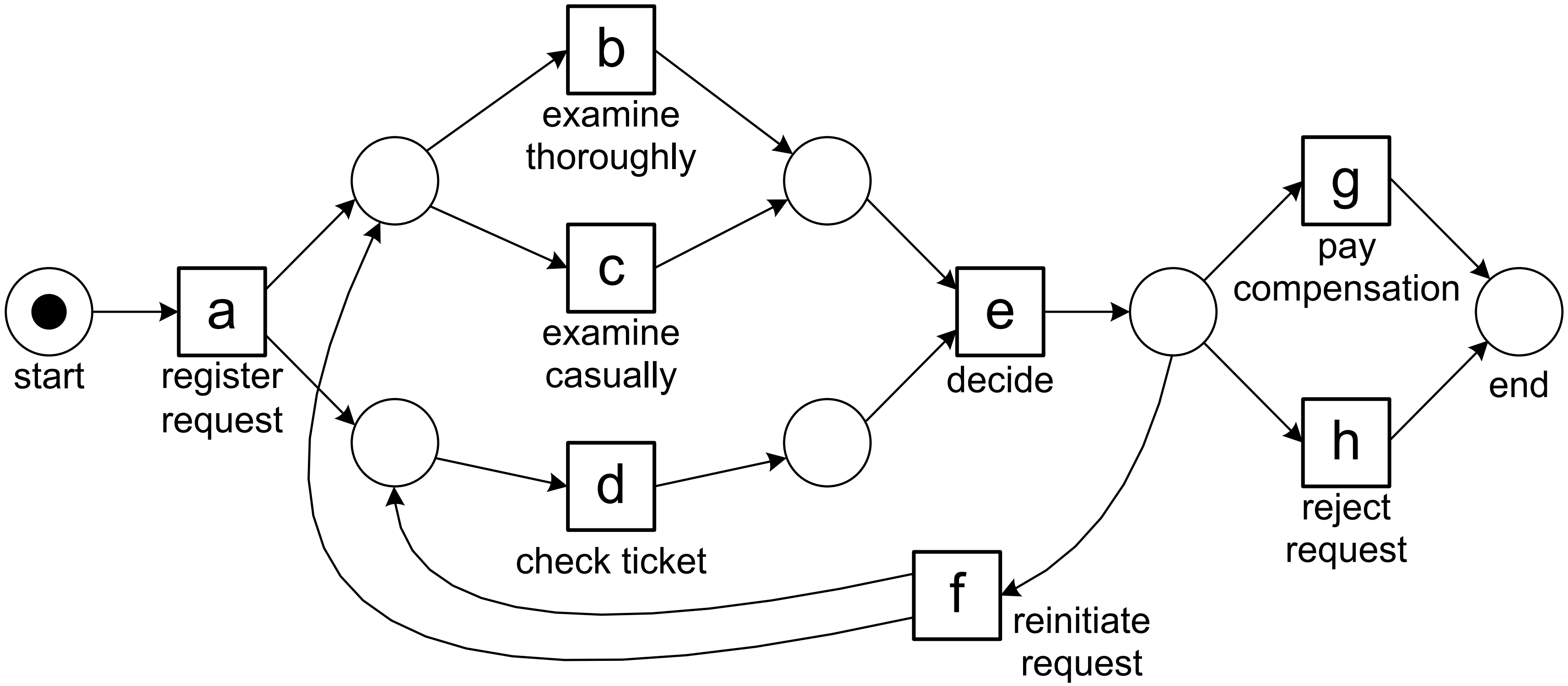
a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

6

a	c1	c2	e1	e2	e3	»	»	»	»	»	»
»	»	»	»	»	»	a	c1	c2	d1	d2	d3

6

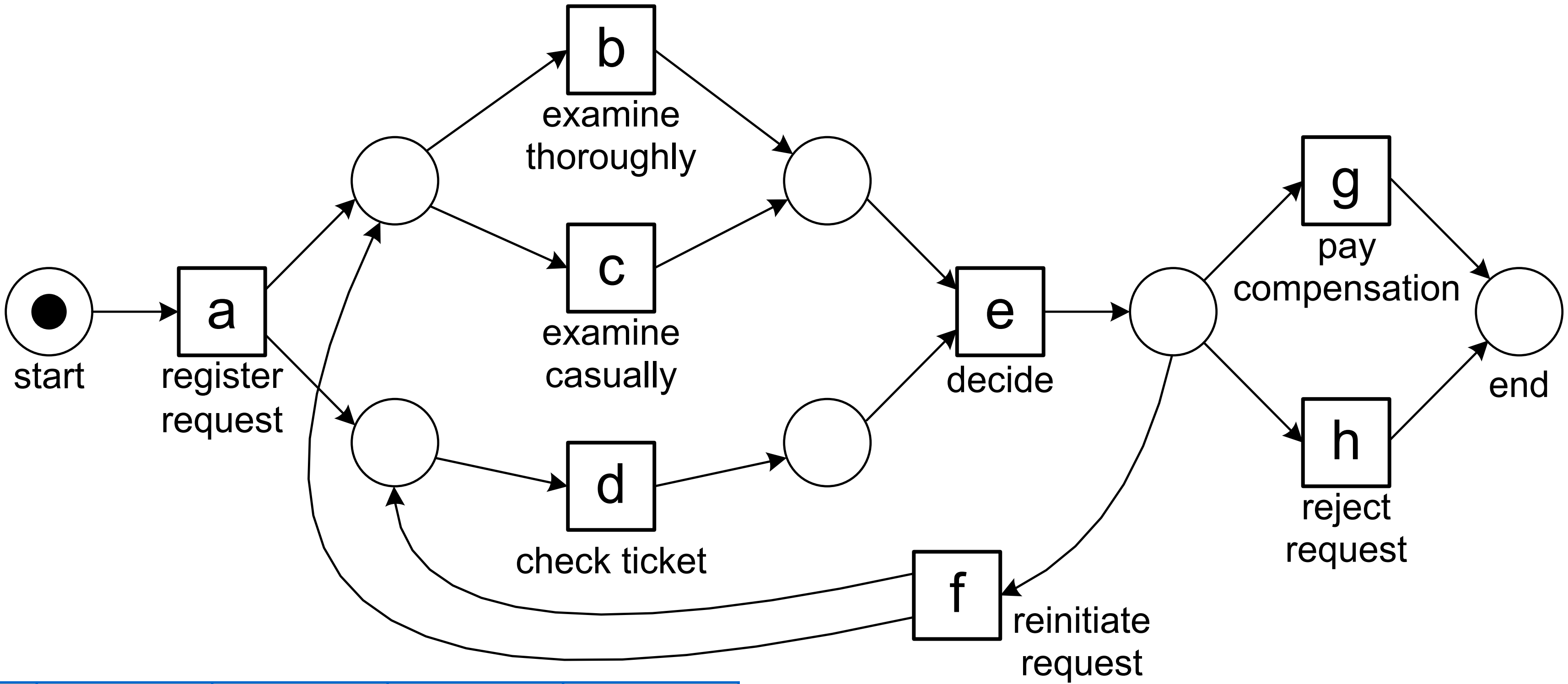
Another example: What is the alignment-based fitness of the trace ⟨b,c⟩ ?



⟨b,c⟩

Chair of Process and Data Science

What is the alignment-based fitness of the trace ⟨b,c⟩ ?



»	b	c	»	»	»
a	b	»	d	e	h

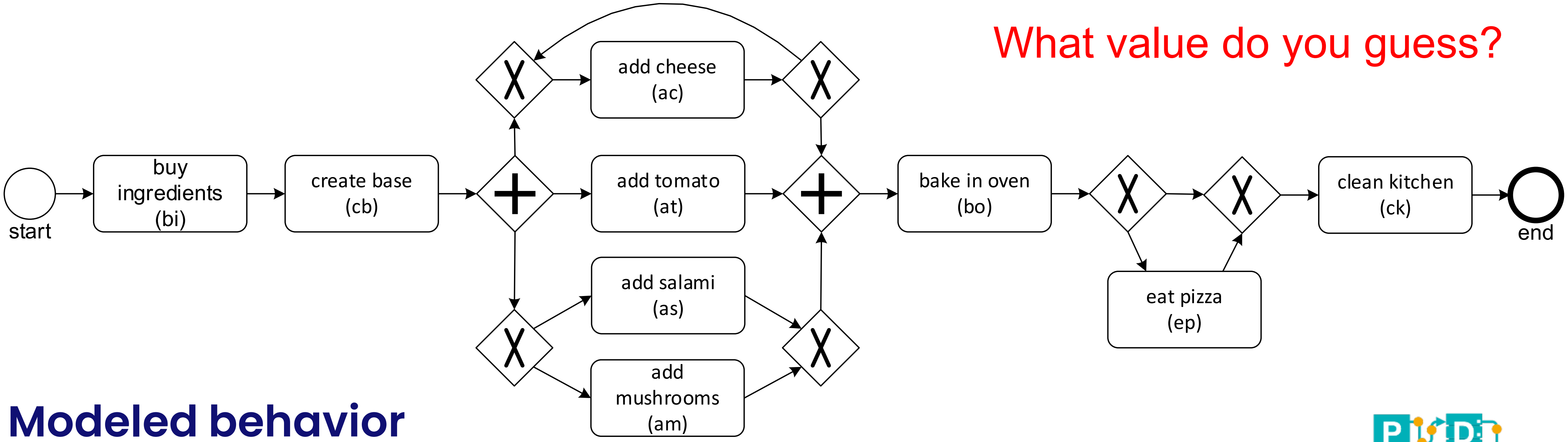
$$1-\frac{5}{2+5}=\frac{2}{7}=0.286$$

Alignments are notation independent

Actual behavior

bi → cb → at → as → am → ck

What value do you guess?



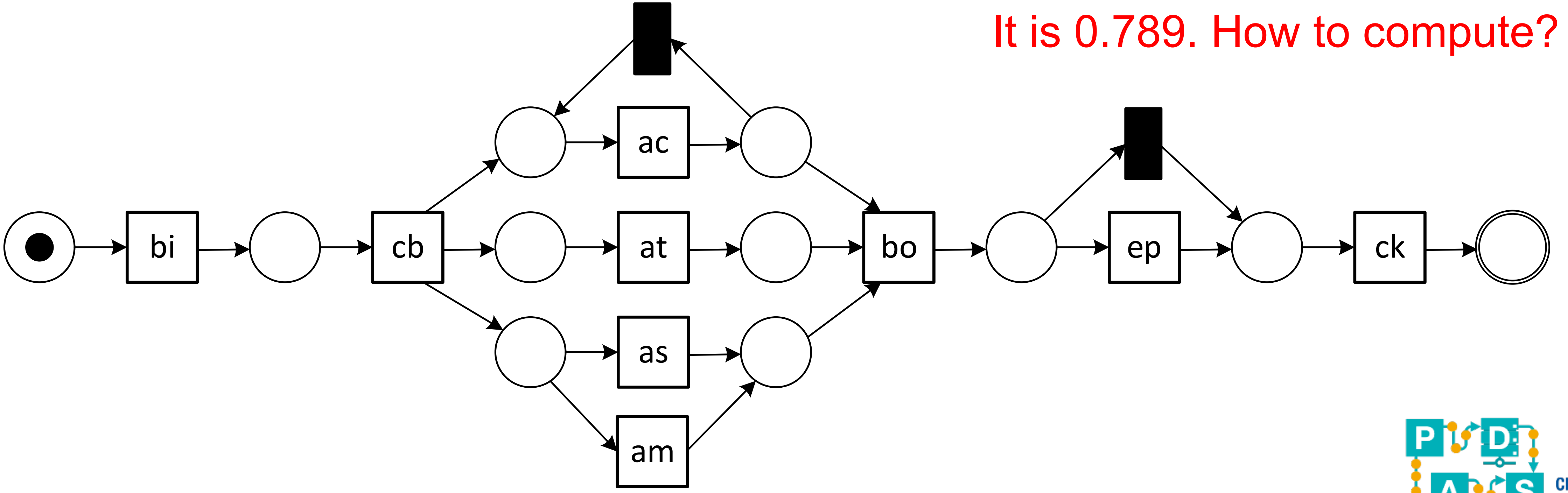
Modeled behavior

Chair of Process and Data Science

Alignments are notation independent

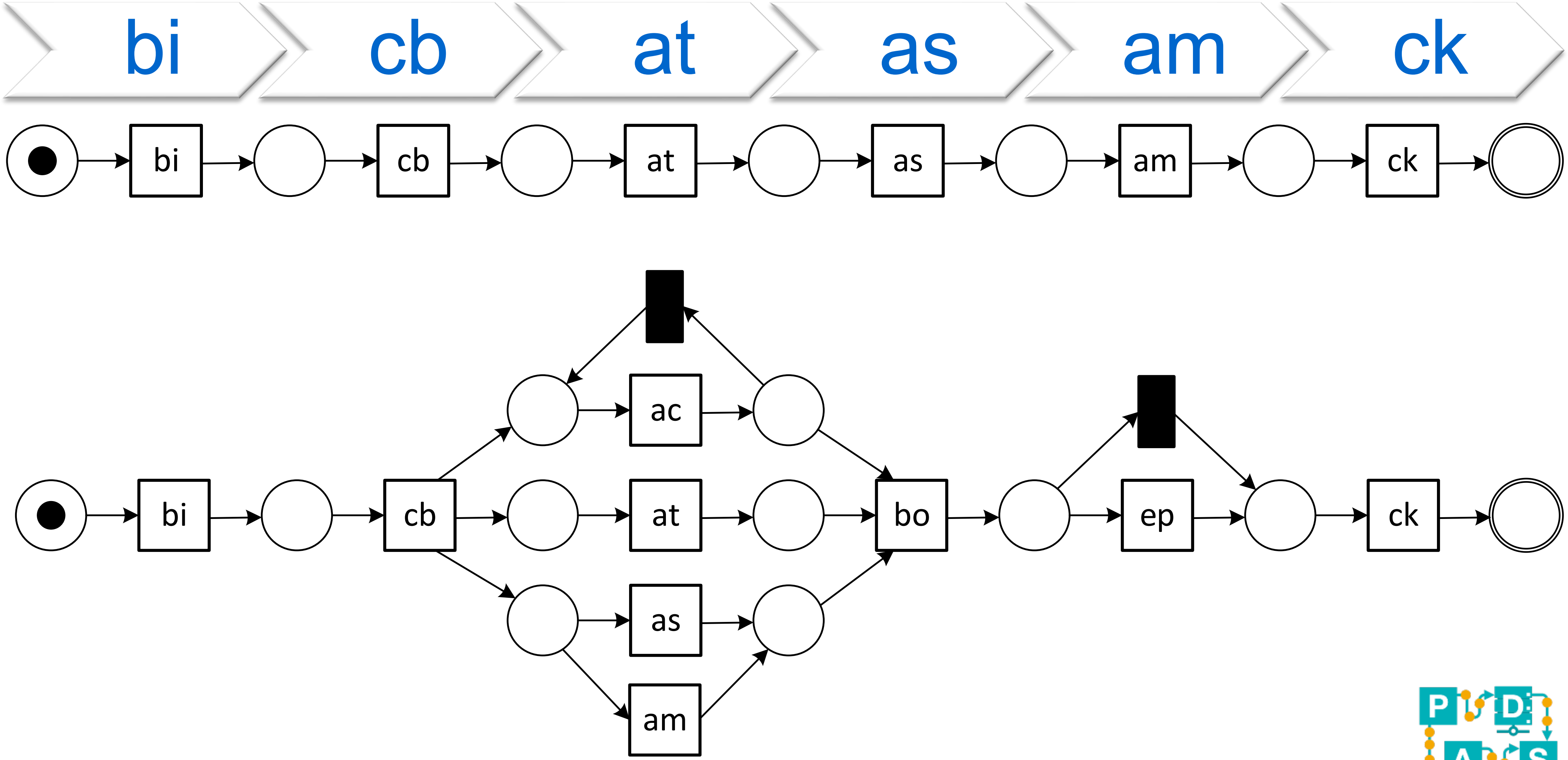
What value do you guess?

It is 0.789. How to compute?



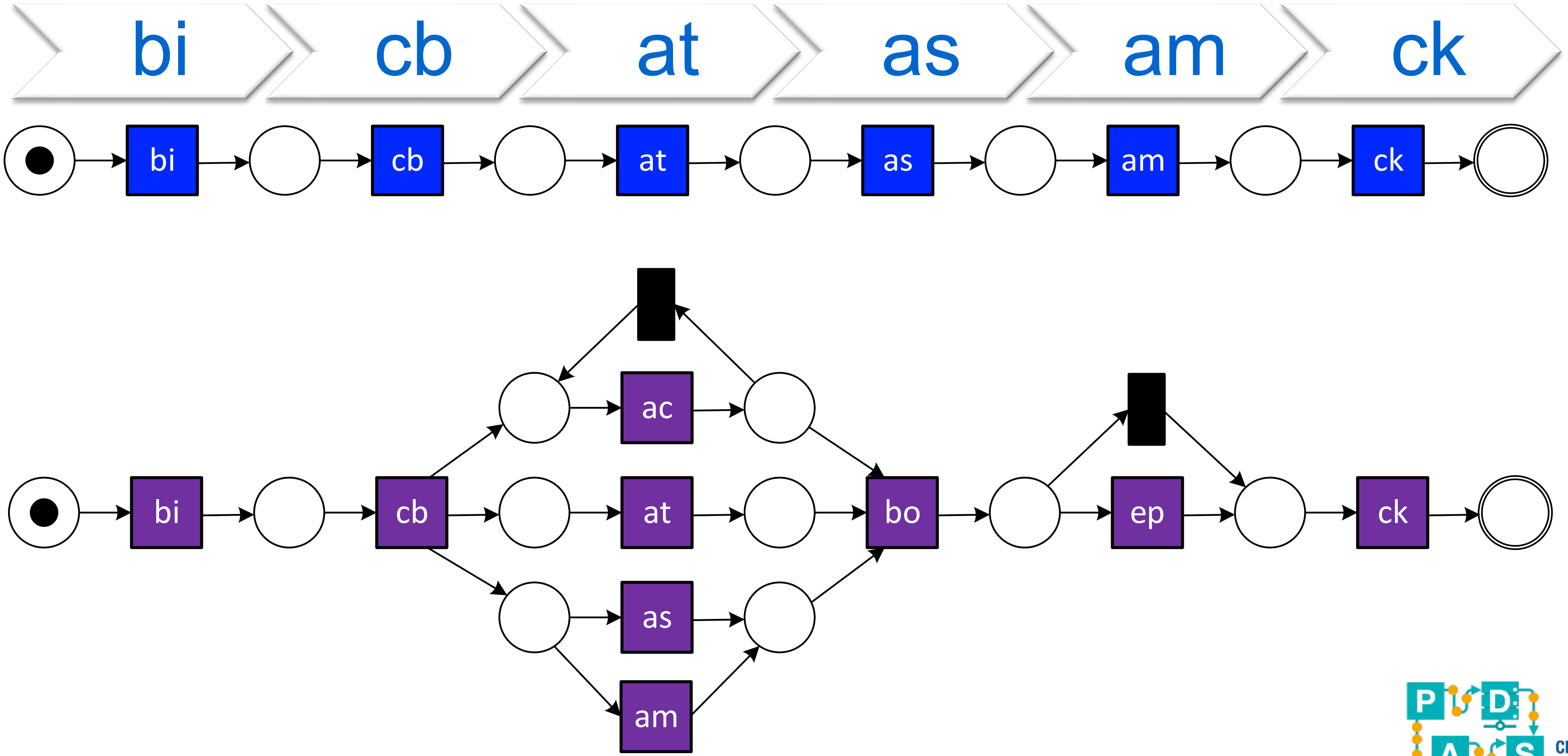
Chair of Process and Data Science

Conformance Checking: Model and trace are mapped onto Petri nets

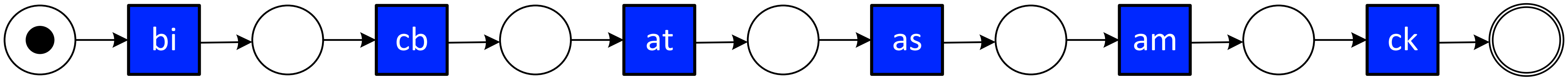


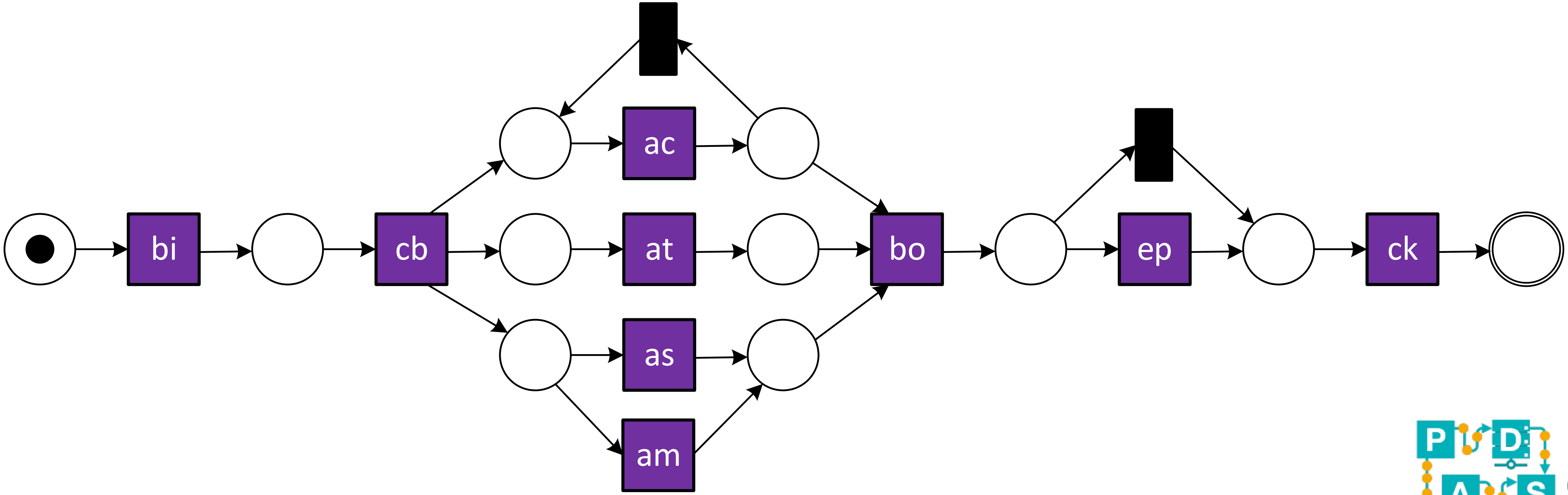
Chair of Process and Data Science

Conformance Checking: Model and trace are mapped onto Petri nets

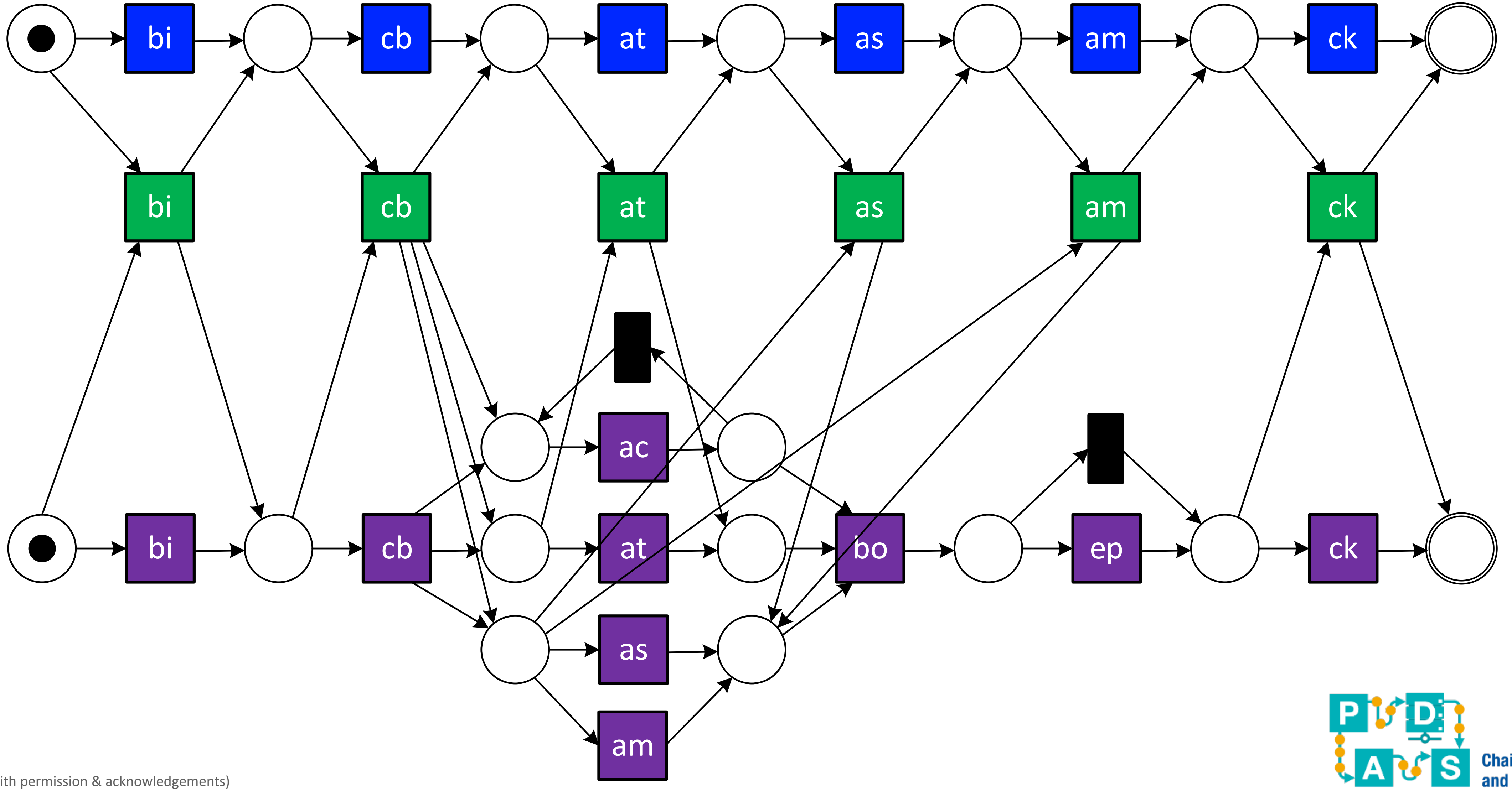


Conformance Checking: Model and trace are mapped onto Petri nets



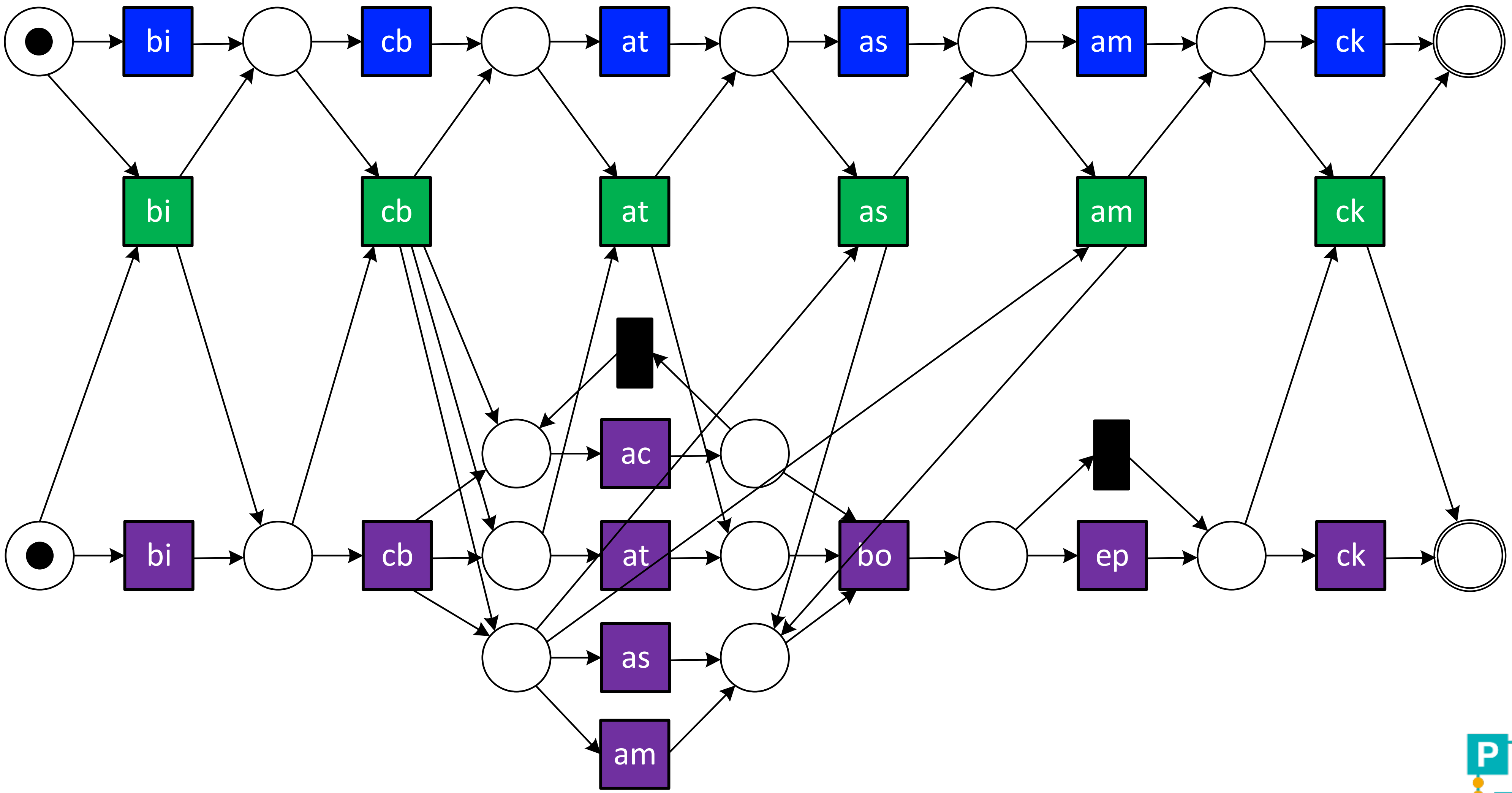


Conformance Checking: Synchronous product



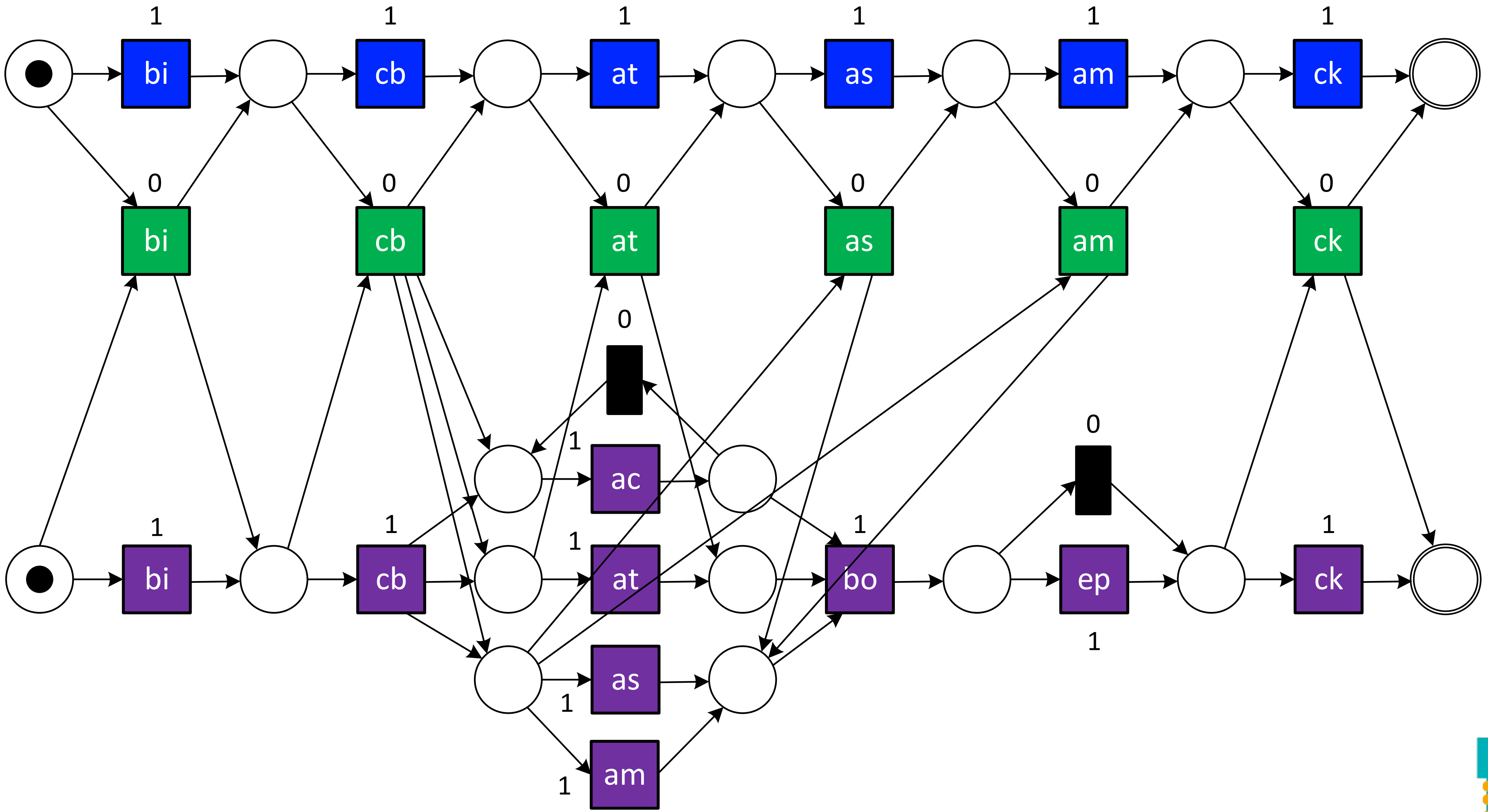
Chair of Process and Data Science

Conformance Checking: Compute a shortest path

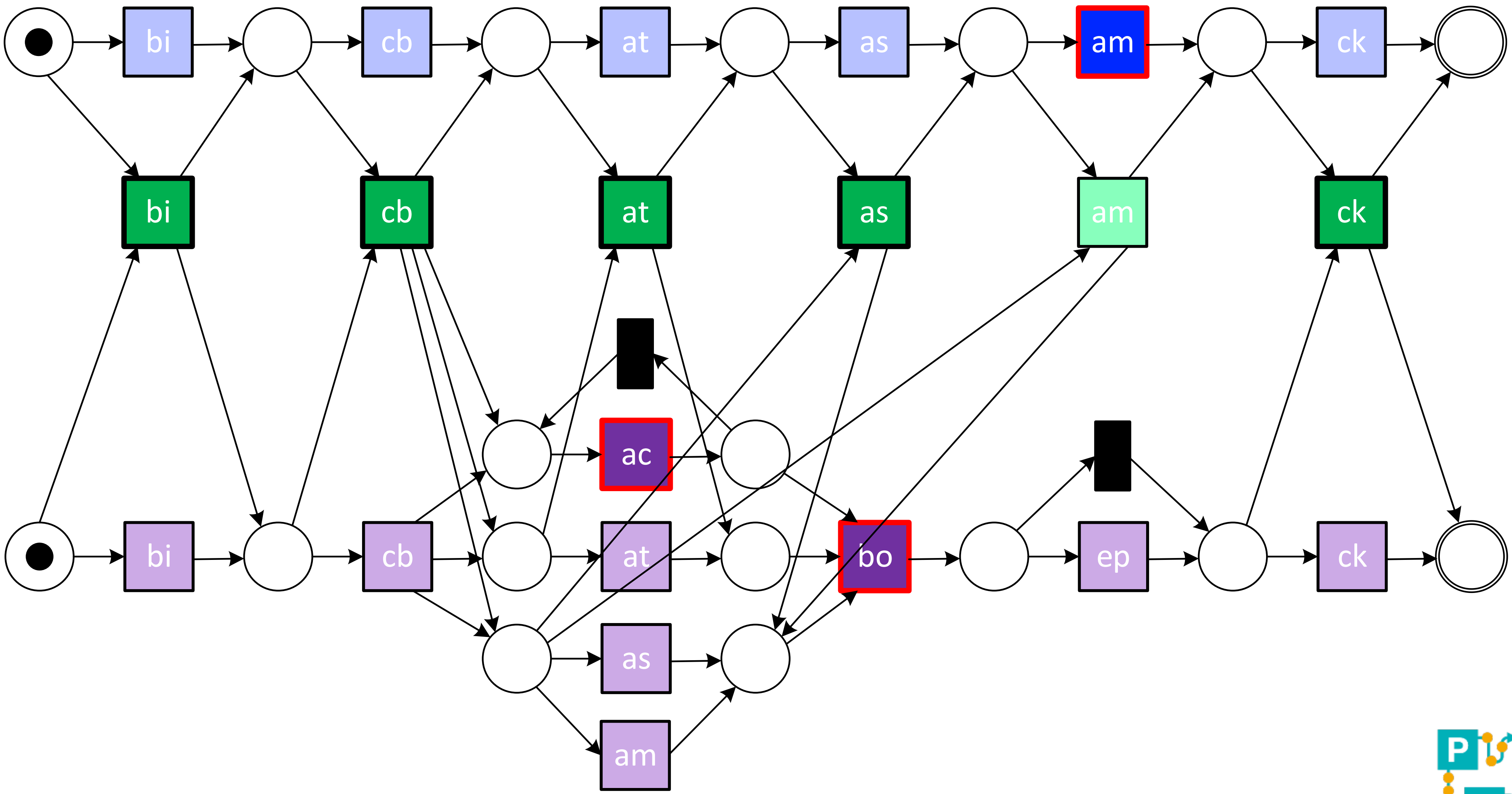


Chair of Process and Data Science

Conformance Checking: Compute a shortest path

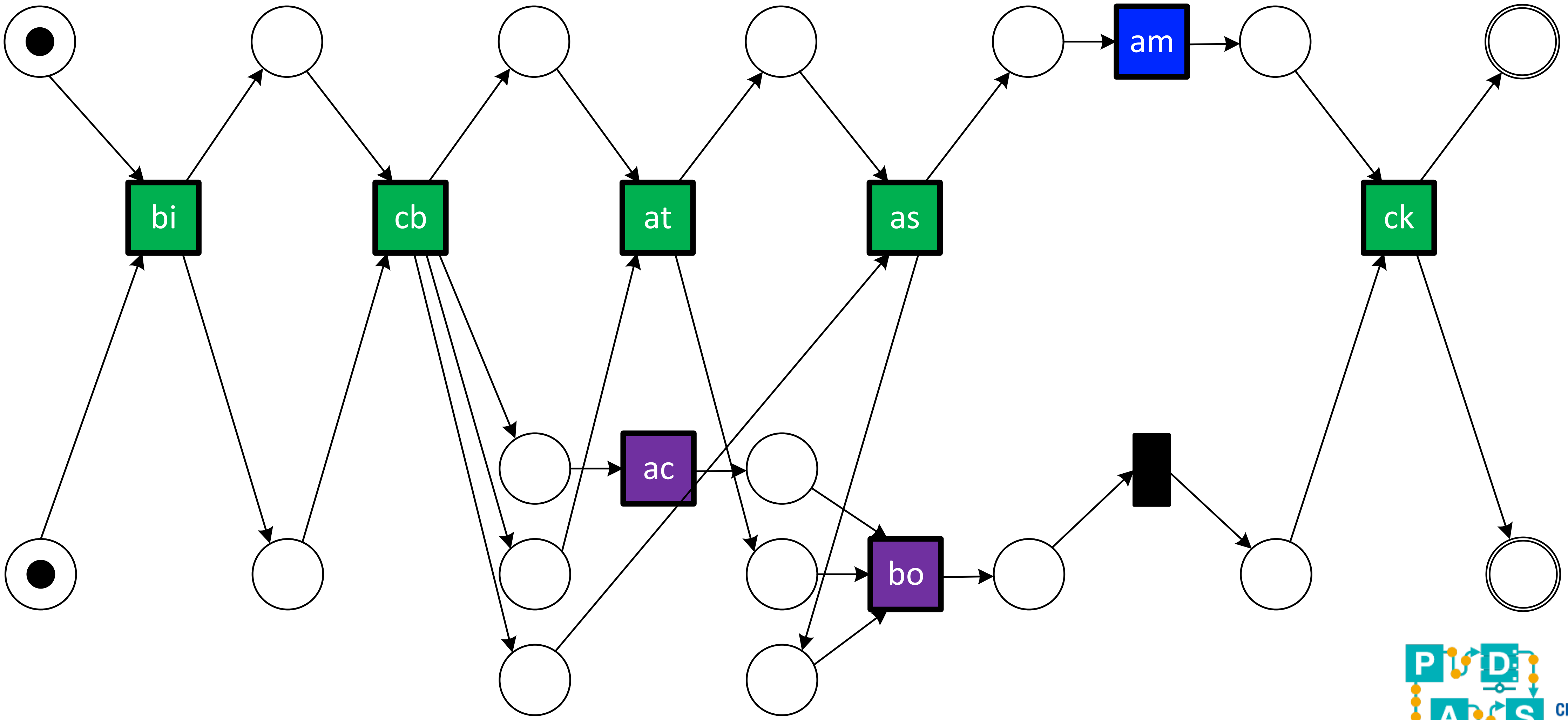


Conformance Checking: Compute a shortest path (cost = 3)



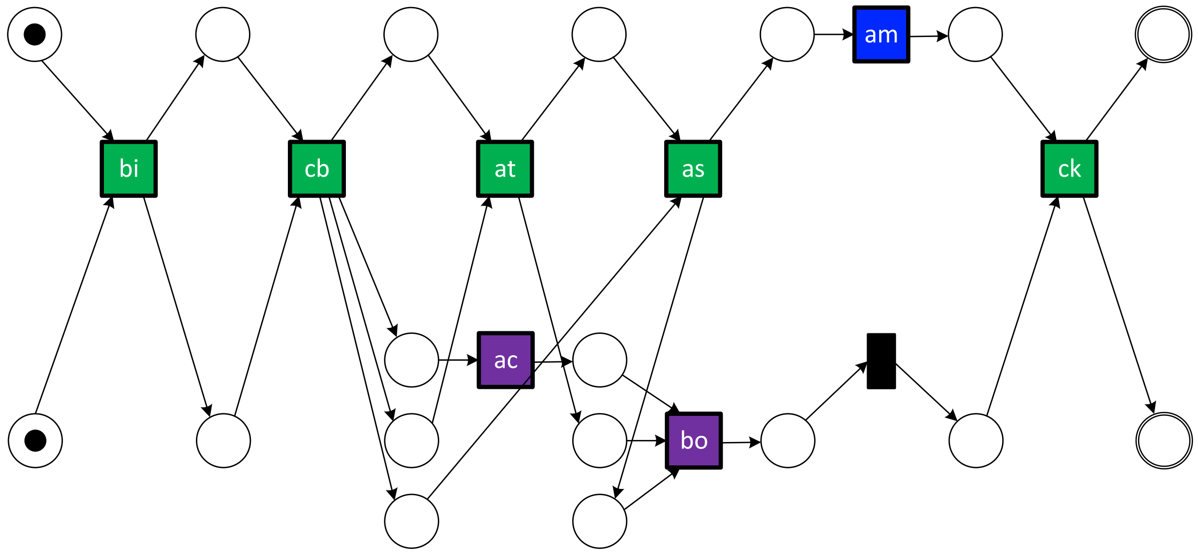
PADS Chair of Process and Data Science

Conformance Checking: Does the trace fit the model?



Chair of Process and Data Science

Conformance Checking: Does the trace fit the model?



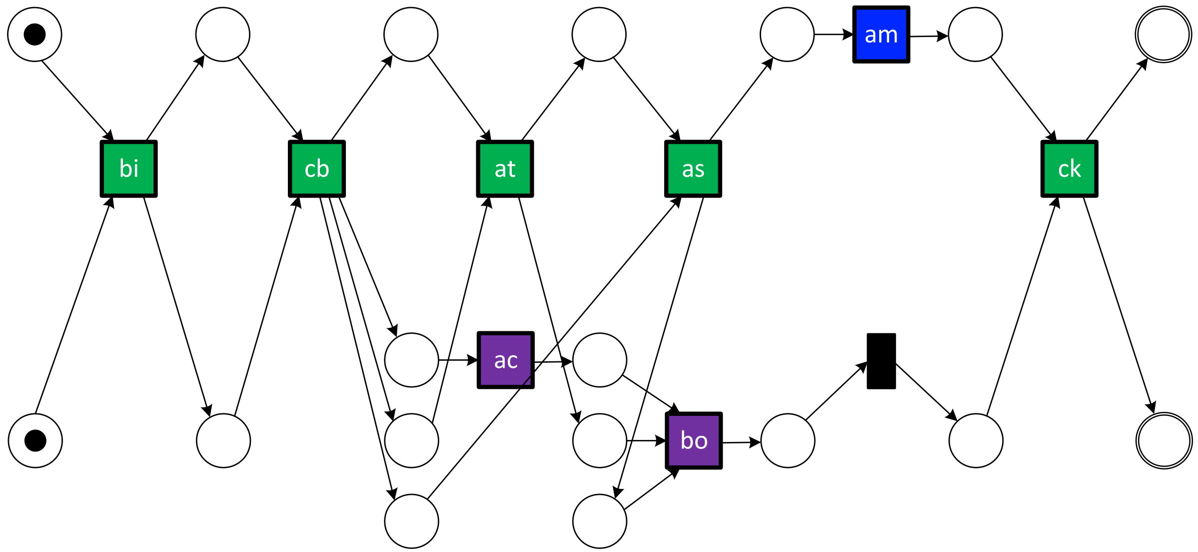
Actual behavior

bi	cb	at	as	>>	>>	am	>>	ck
bi	cb	at	as	ac	bo	>>	τ	ck

Modeled behavior

Chair of Process and Data Science

Conformance Checking: Does the trace fit the model?



Actual behavior

bi	cb	at	as	>>	>>	am	>>	ck
bi	cb	at	as	ac	bo	>>	τ	ck

Modeled behavior

Chair of Process and Data Science

Conformance Checking: Does the trace fit the model?

bi	cb	at	as	>>	>>	am	>>	ck
bi	cb	at	as	ac	bo	>>	τ	ck

We should have added cheese (ac) and backed the pizza in the oven (bo). Moreover, should not have added mushrooms (am) because we already added salami.

Alignment-based fitness:

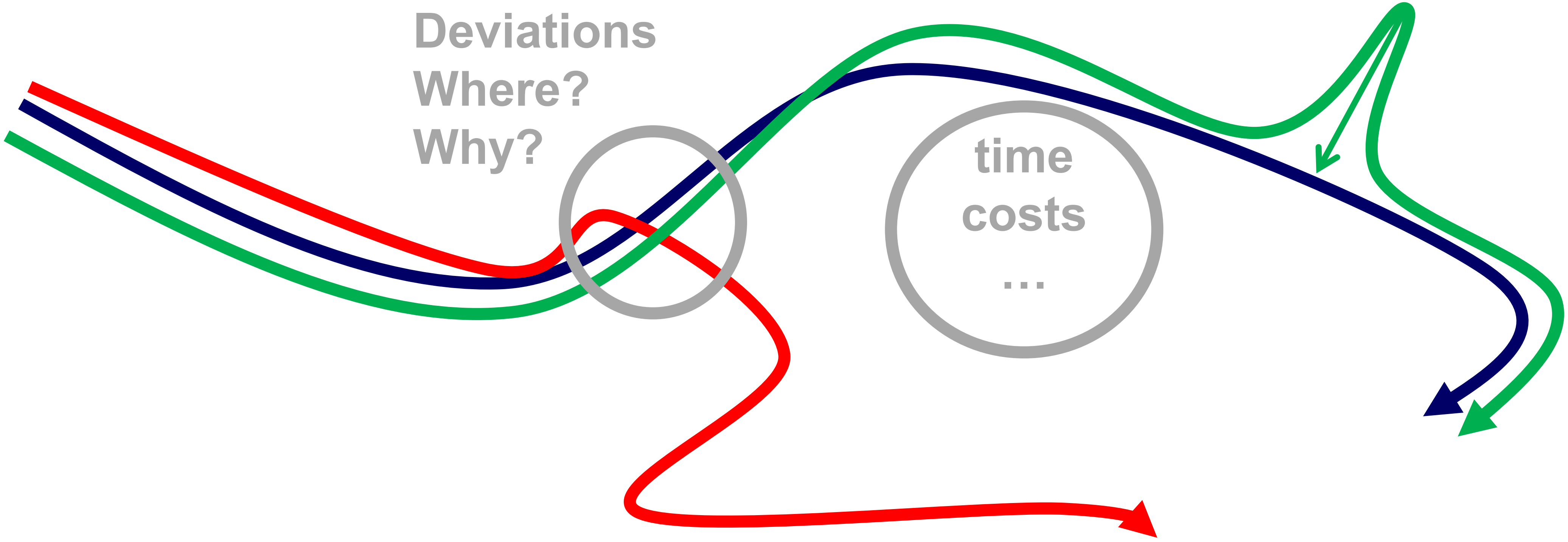
- Optimal alignment: cost 3
- Length of trace: 6
- Shortest visible model trace: 7

$$1 - \frac{3}{6+7} = \frac{10}{13} = 0.769$$

Chair of Process and Data Science

Advantages of aligning log and model

- Observed behavior is directly related to modeled behavior.
- Very flexible (any cost structure).
- Detailed diagnostics.
- After aligning log and model, other quality dimensions can be investigated (separation of concerns).

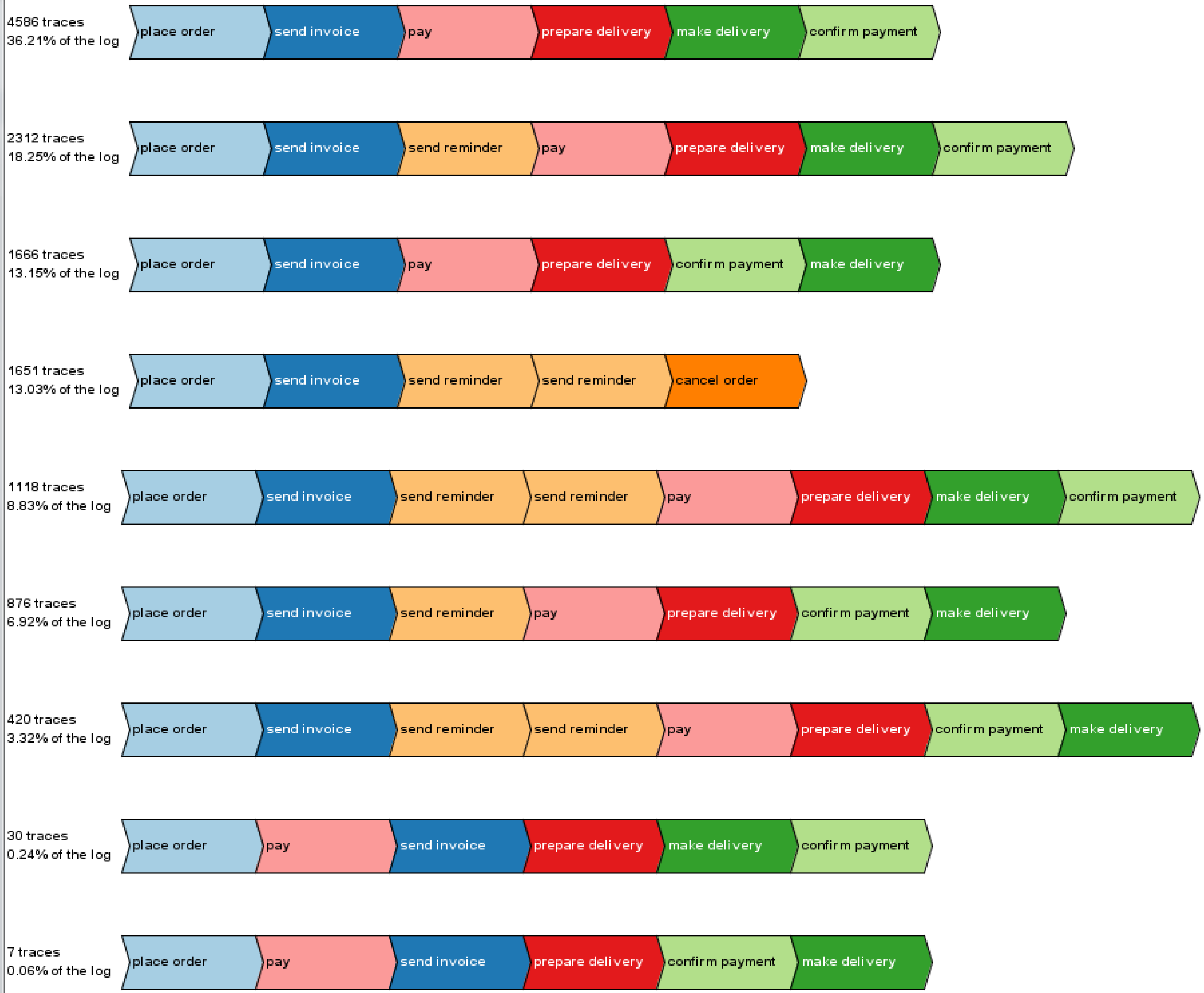


Chair of Process and Data Science

Conformance checking in ProM

ProM: Load Event Log

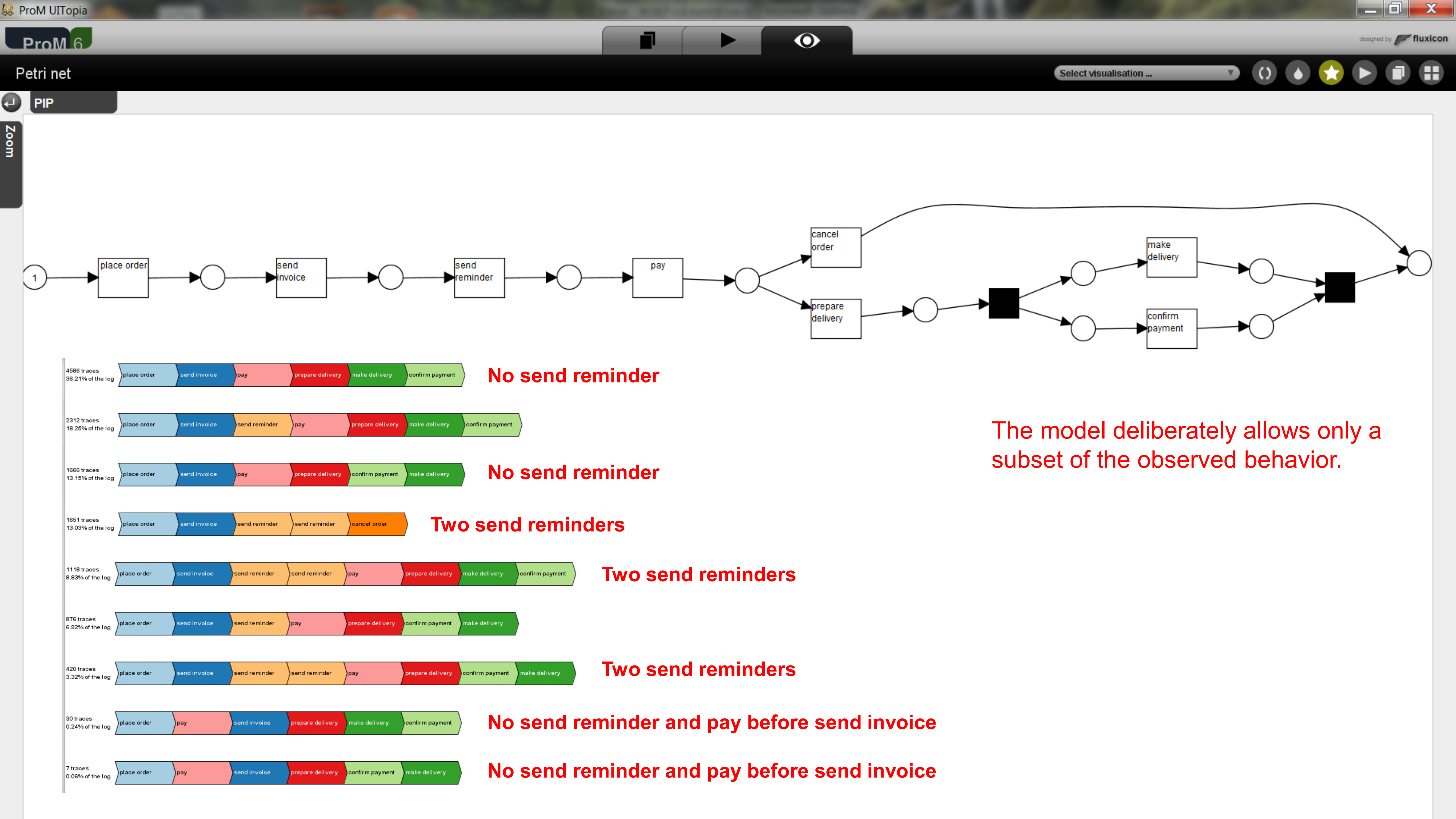
Chair of Process and Data Science

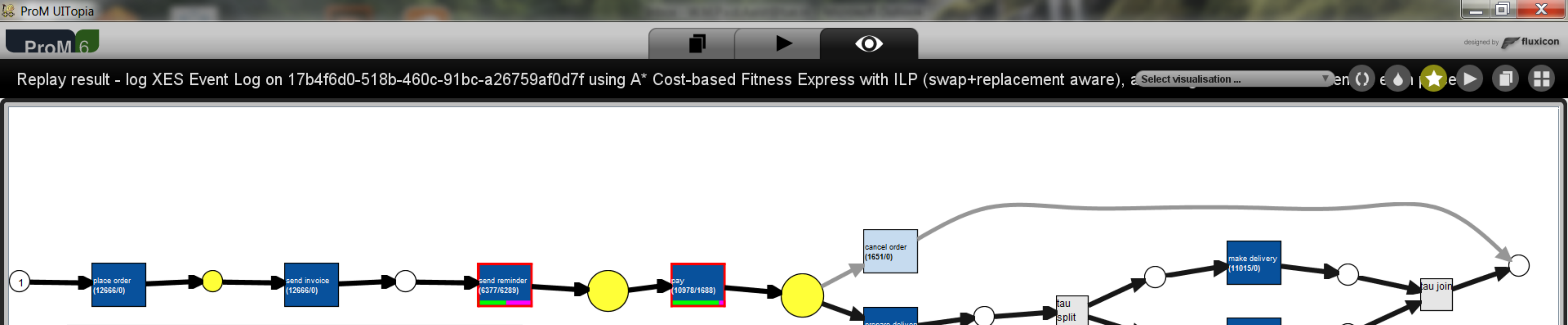


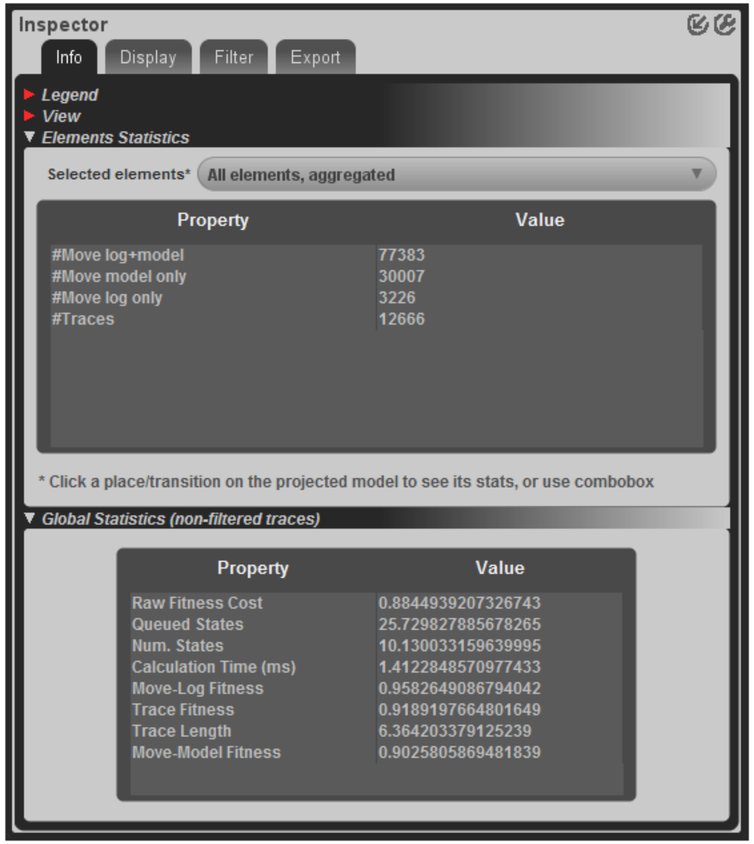
Event log

- **12,666 cases**
- **80,609 events**
- **8 unique activities**

Chair of Process
and Data Science


ProM UITopia
ProM 6
Petri net
Select visualisation ...
designed by fluxicon
PIP
Zoom
1
place order
send invoice
send reminder
pay
cancel order
prepare delivery
make delivery
confirm payment
4586 traces 36.21% of the log
place order send invoice pay prepare delivery make delivery confirm payment
No send reminder
2312 traces 18.25% of the log
place order send invoice send reminder pay prepare delivery make delivery confirm payment
1666 traces 13.15% of the log
place order send invoice pay prepare delivery confirm payment make delivery
No send reminder
1651 traces 13.03% of the log
place order send invoice send reminder send reminder cancel order
Two send reminders
1118 traces 8.83% of the log
place order send invoice send reminder send reminder pay prepare delivery make delivery confirm payment
Two send reminders
876 traces 6.92% of the log
place order send invoice send reminder pay prepare delivery confirm payment make delivery
420 traces 3.32% of the log
place order send invoice send reminder send reminder pay prepare delivery confirm payment make delivery
Two send reminders
30 traces 0.24% of the log
place order pay send invoice prepare delivery make delivery confirm payment
No send reminder and pay before send invoice
7 traces 0.06% of the log
place order pay send invoice prepare delivery confirm payment make delivery
No send reminder and pay before send invoice
The model deliberately allows only a subset of the observed behavior.





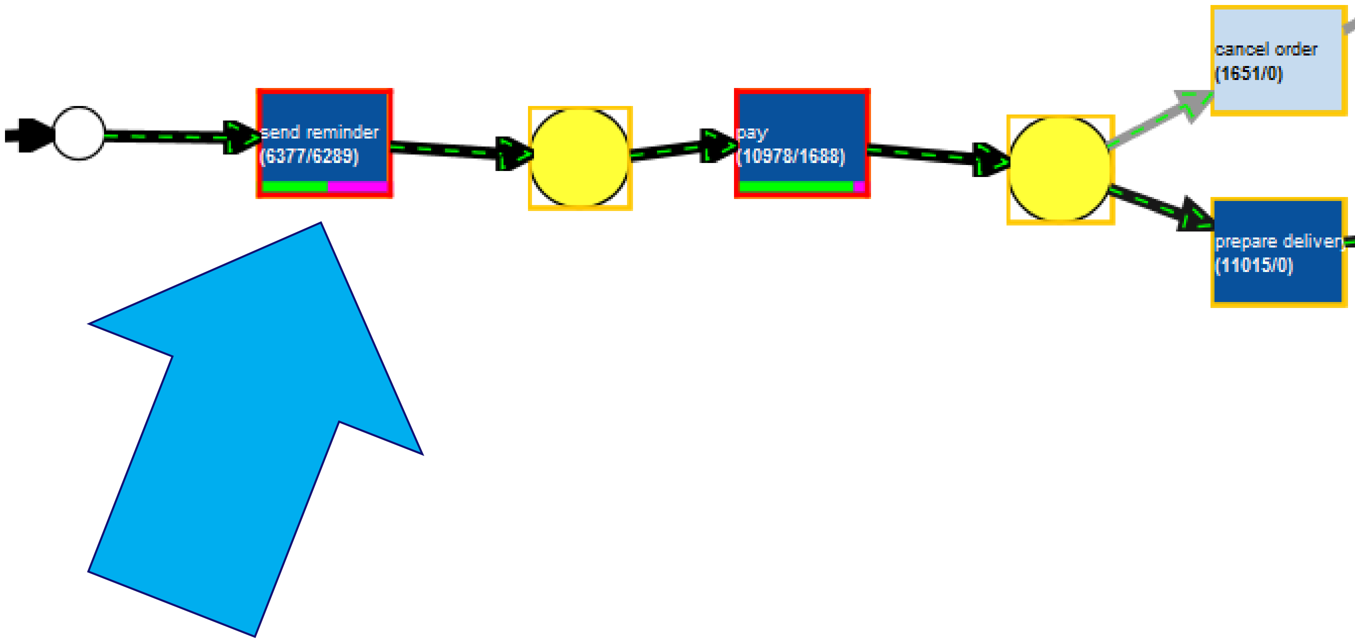
3226 move on log only (yellow)

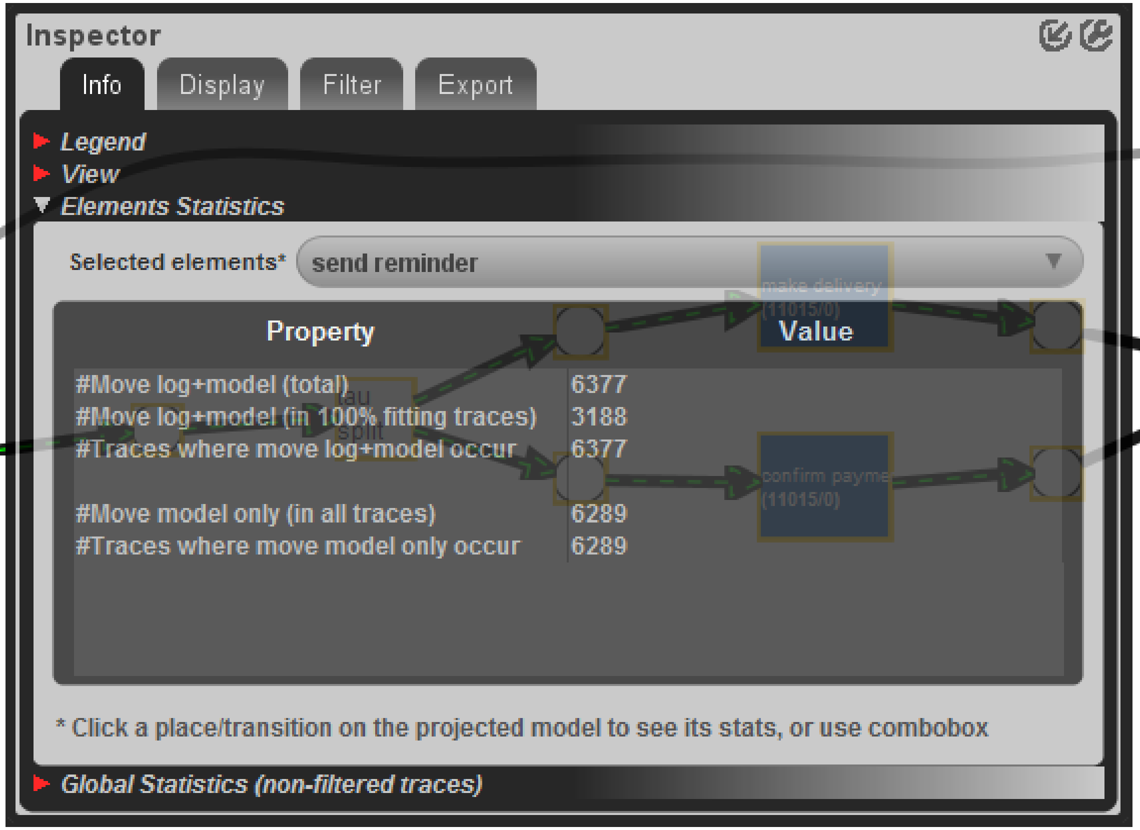
30007 move on model only

- **7977 real moves on model only (purple)**
- **22030 occurrences of silent activities**

30007=6289+1688+11015+11015 move on model only

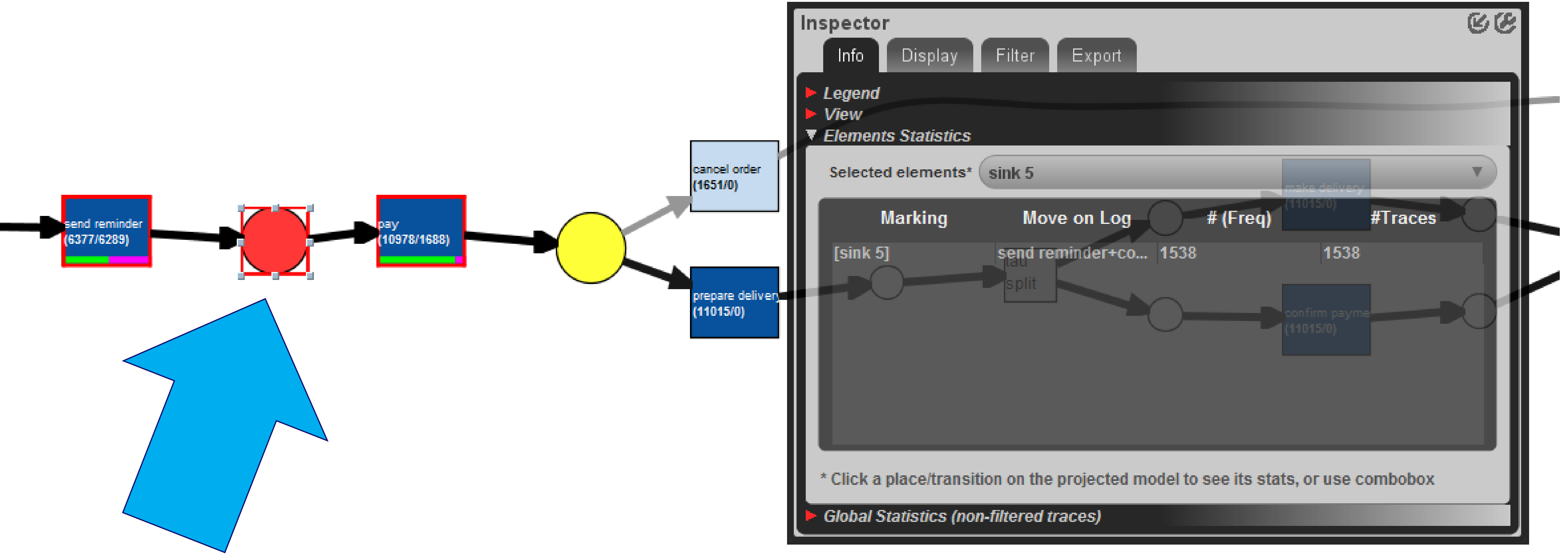
Send reminder is often skipped (6289 times)



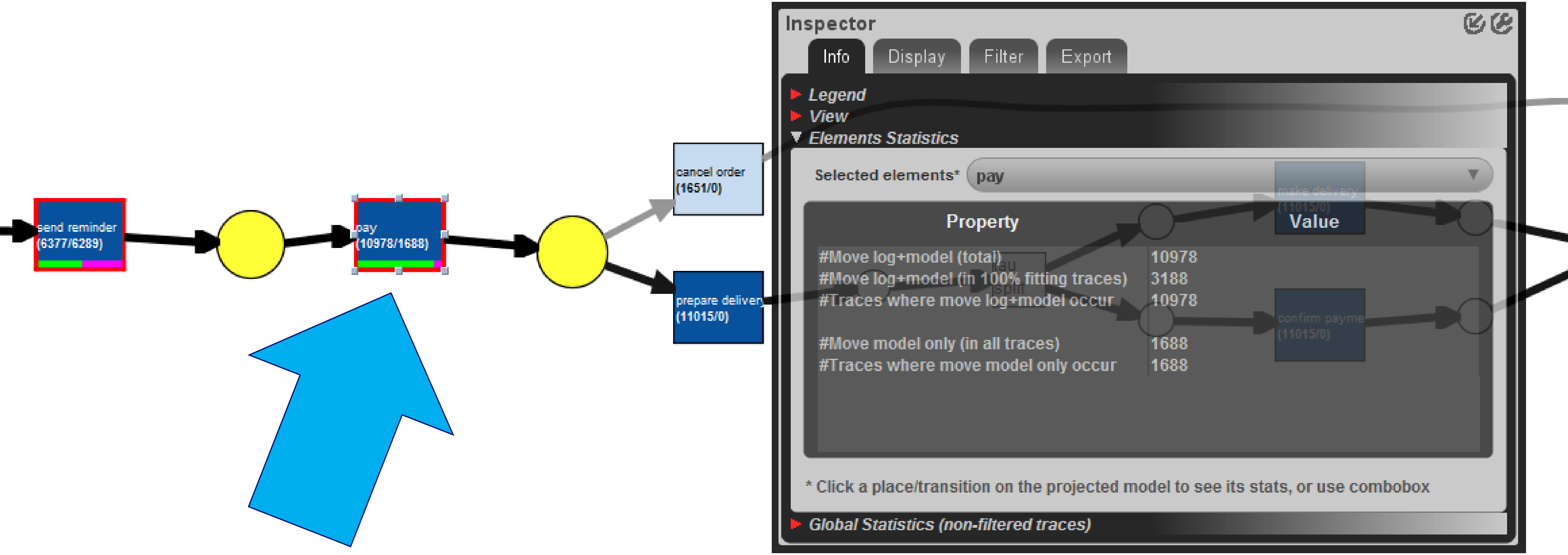


Chair of Process and Data Science

Additional send reminder (1538 times in this place)

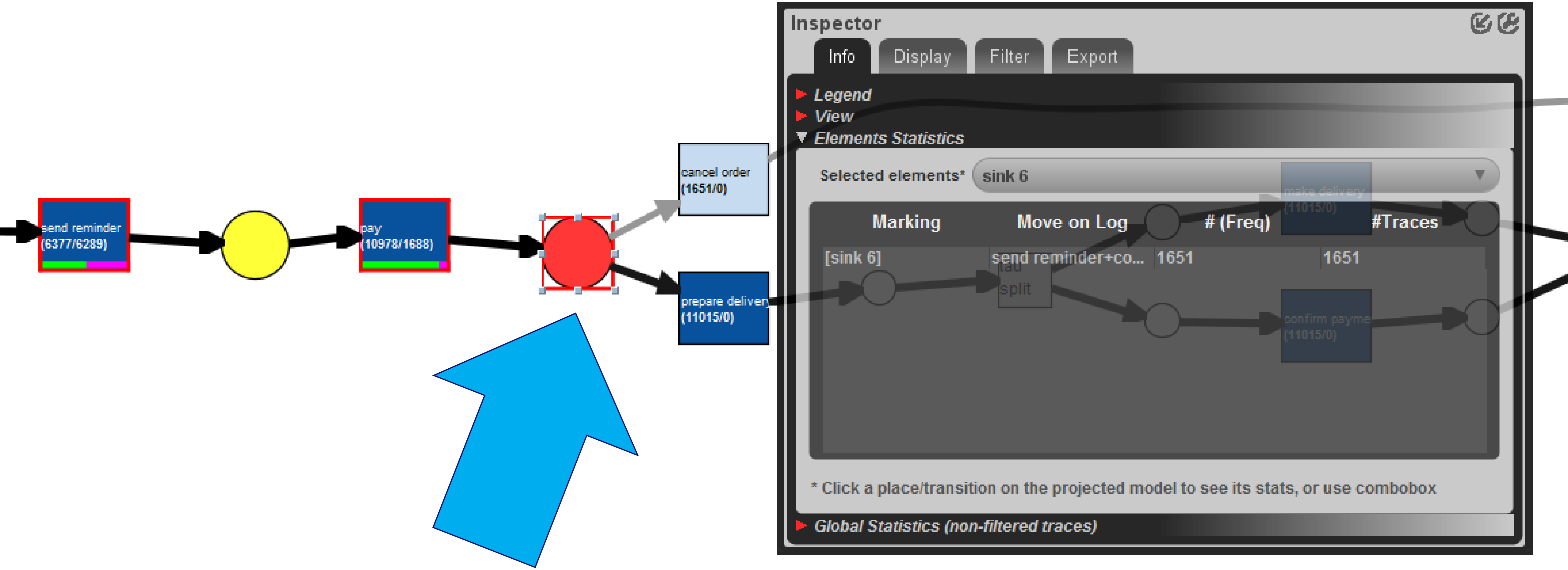


Pay is sometimes skipped (1688 times)

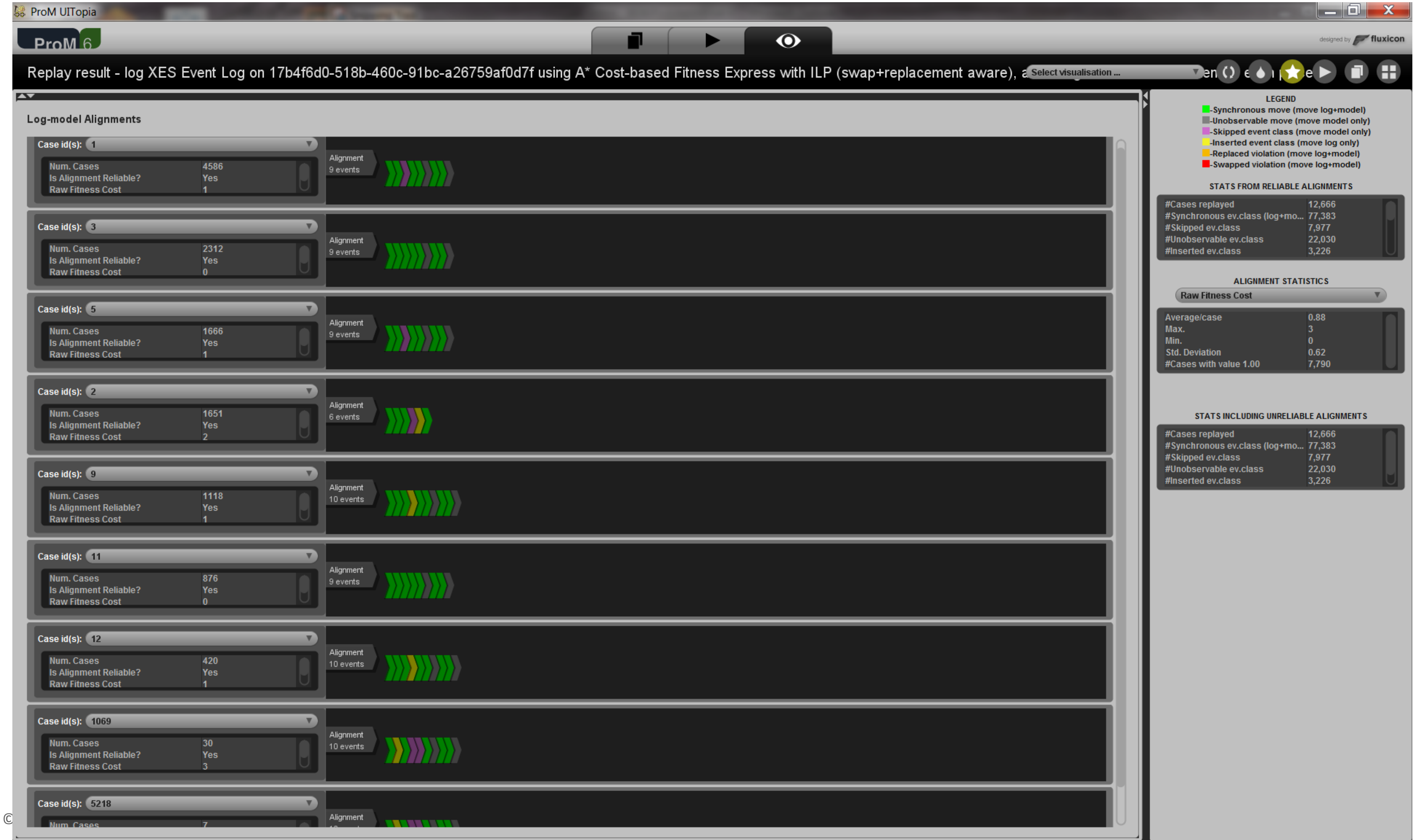


Chair of Process and Data Science

Additional send reminder (1651 times in this place)



Log view



synchronous move

move on model
(required activity was skipped in event log)

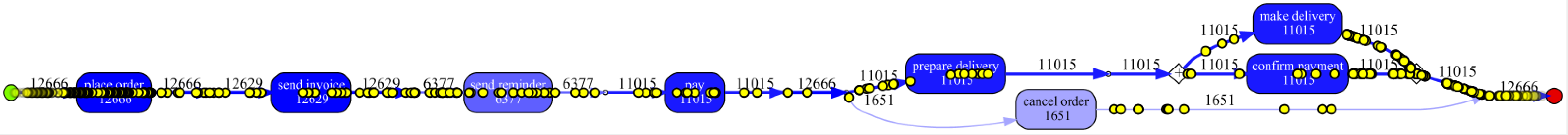
move on model
(silent transition in model was fired)

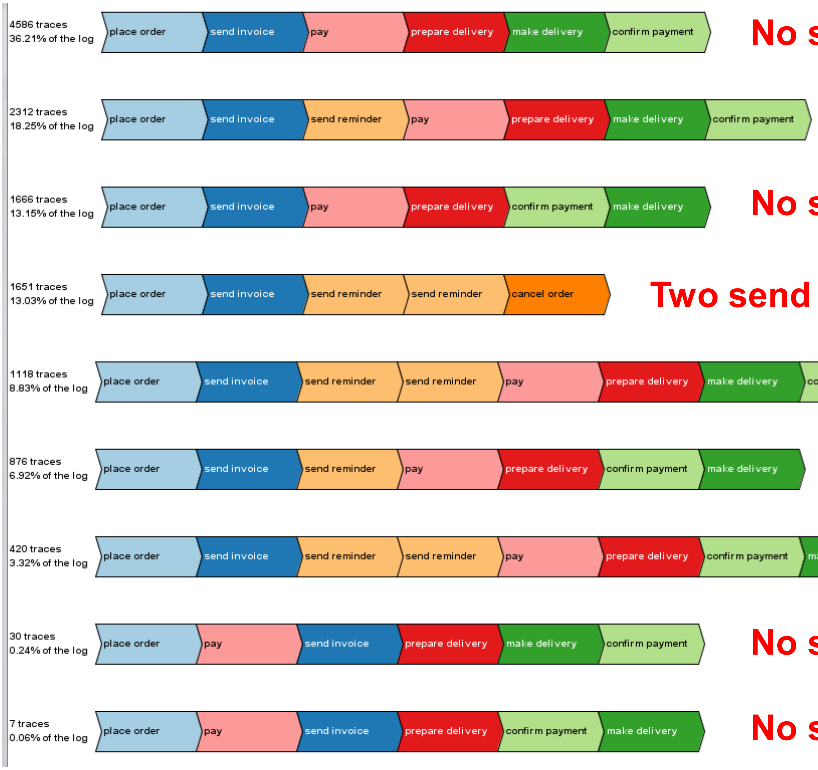
move on log
(activity in log was not possible in model)

Chair of Process and Data Science

Many more plug-ins exploring conformance (based on alignments or not)

Inductive miner





No send reminder

No send reminder

Two send reminders

Two send reminders

Two send reminders

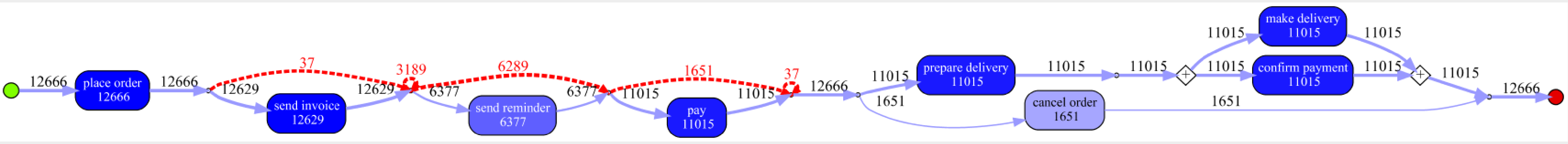
No send reminder and pay before send invoice

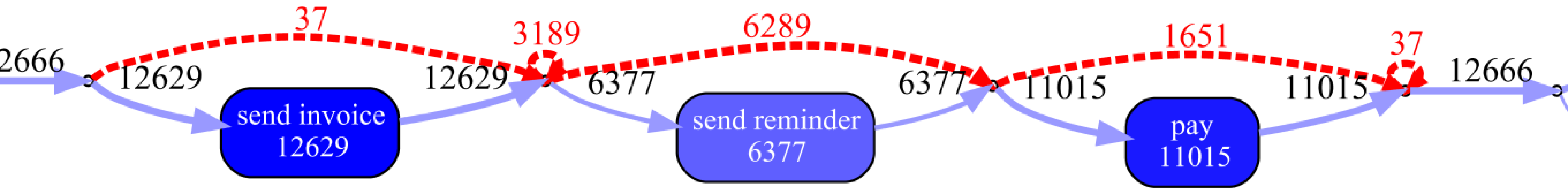
No send reminder and pay before send invoice

Again, the model deliberately allows only a subset of the observed behavior.

Chair of Process and Data Science

Deviations in inductive miner





Note that some numbers are different than before because of different alignments.

Chair of Process and Data Science

Drill down on deviations

ProM UITopia

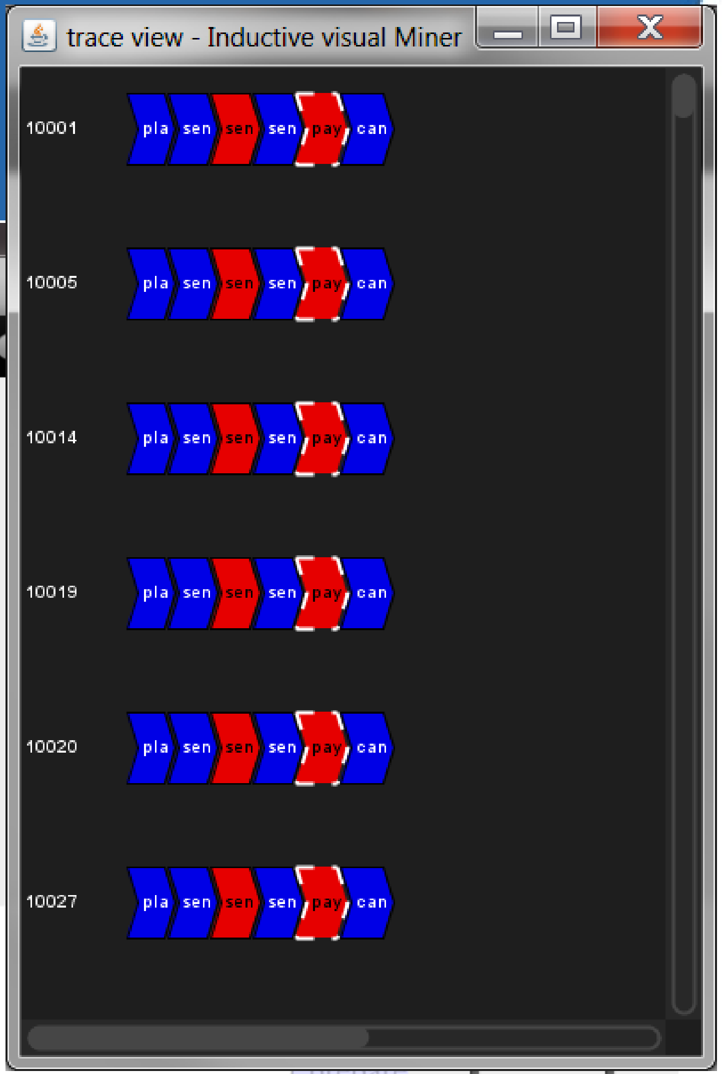
Inductive visual Miner

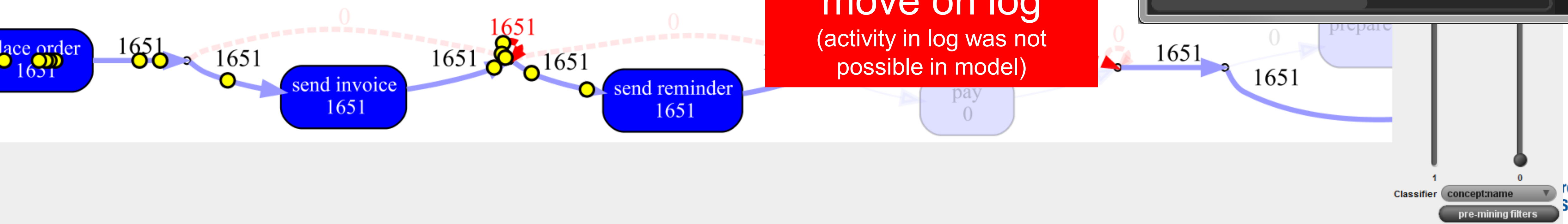
Pay is skipped 1651 times and all of these cases had an extra send reminder (these are the cases that were cancelled).

synchronous move

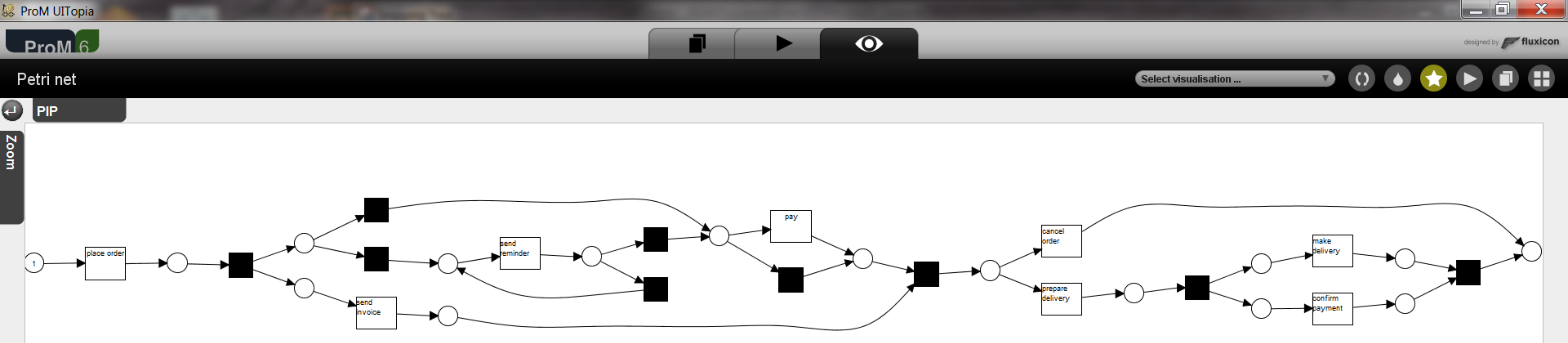
move on model
(required activity was skipped in event log)

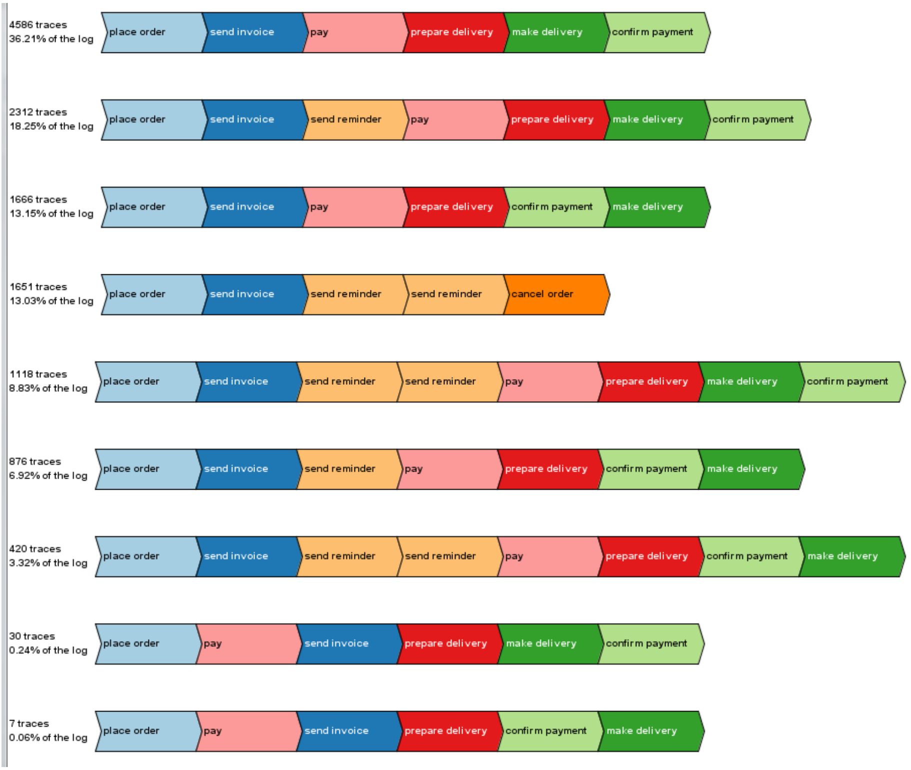
move on log
(activity in log was not possible in model)





Same event log and different (underfitting) model

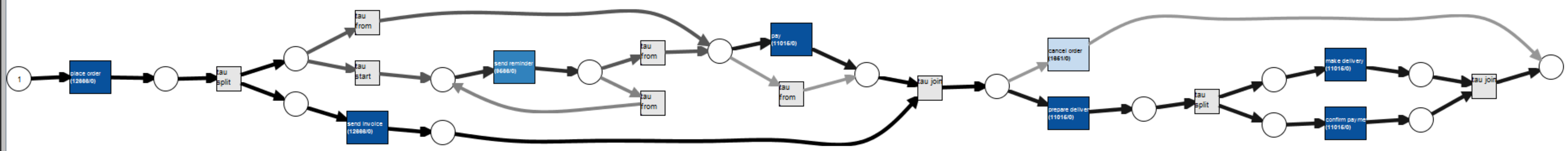




Perfect replay fitness (can reproduce all traces), but underfitting (i.e., not precise enough).

ProM UITopia

Replay result - log XES Event Log on ad14ea2b-9e58-4d4c-b840-02be6fb3dff3 using A* Cost-based Fitness Express with ILP (swap+replacement aware), a



Inspector

Info | Display | Filter | Export

- Legend
- View
- Elements Statistics

Selected elements* All elements, aggregated

Property	Value
#Move log+model	80609
#Move model only	71245
#Move log only	0
#Traces	12666

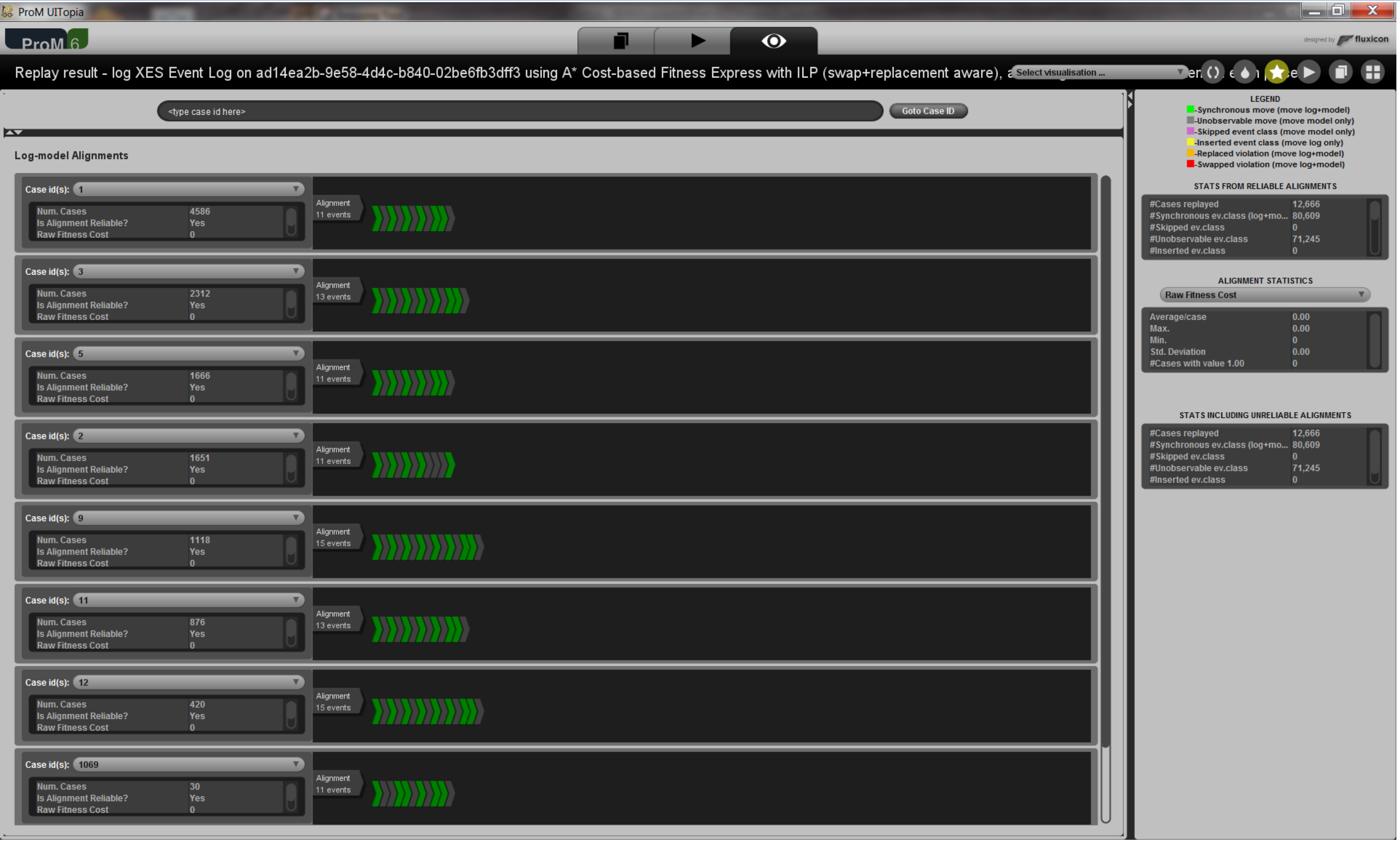
* Click a place/transition on the projected model to see its stats, or use combobox

- Global Statistics (non-filtered traces)

Property	Value
Raw Fitness Cost	0.0
Queued States	45.91386388757287
Num. States	15.896652455392376
Calculation Time (ms)	8.713721774830239
Move-Log Fitness	1.0
Trace Fitness	1.0
Trace Length	6.364203379125239
Move-Model Fitness	1.0

No problems, only moves on model for silent steps.

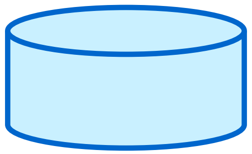
Log view



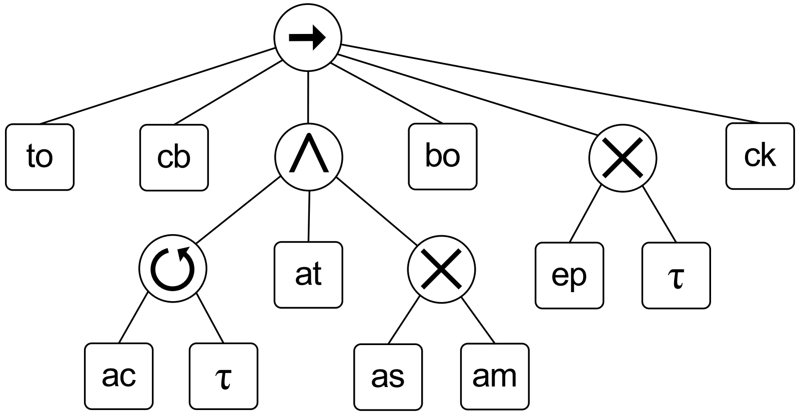
Chair of Process and Data Science

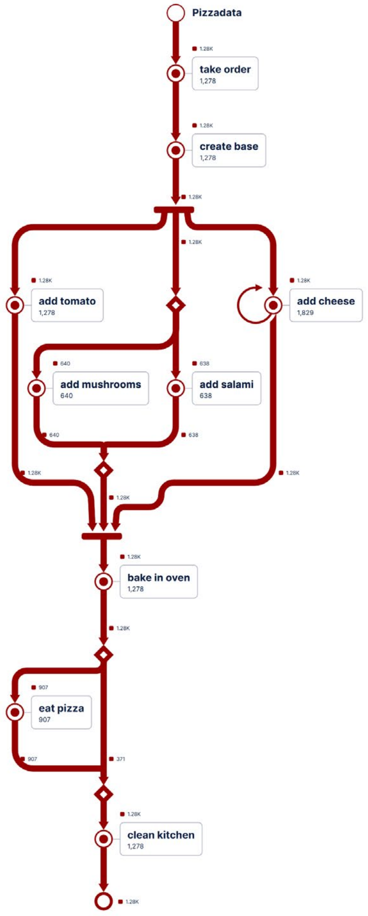
Conformance checking in Celonis

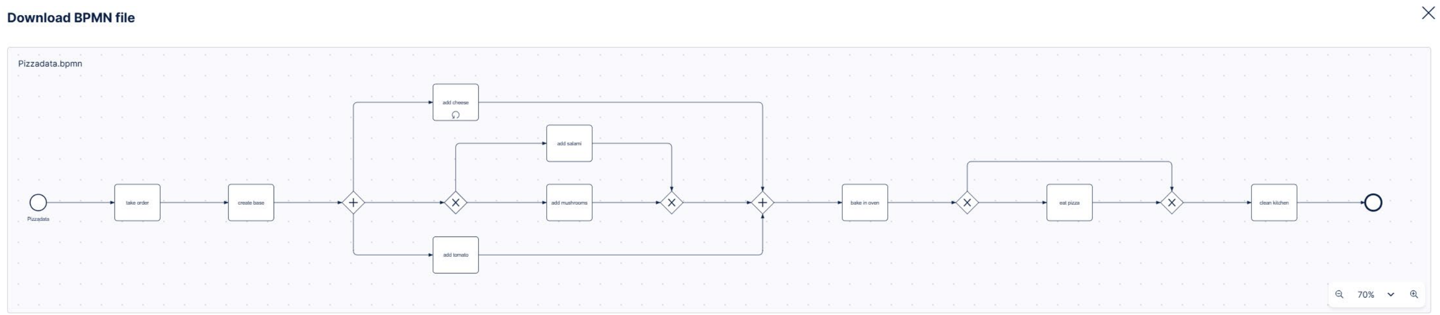
Discovered BPMN model in Celonis



1278 cases/pizzas



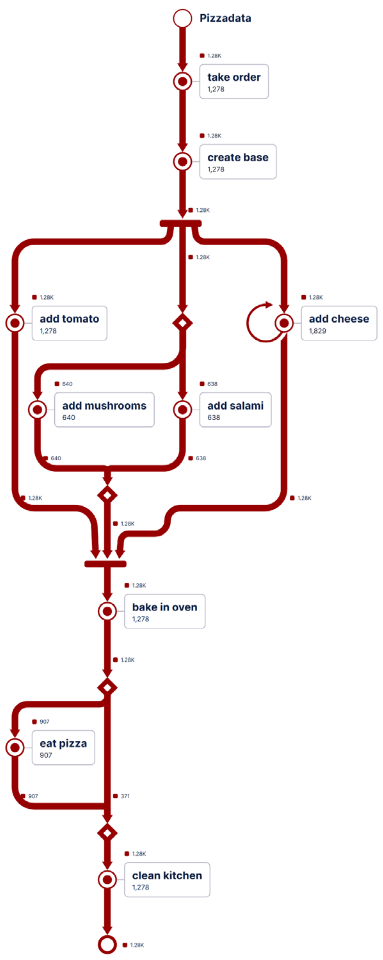


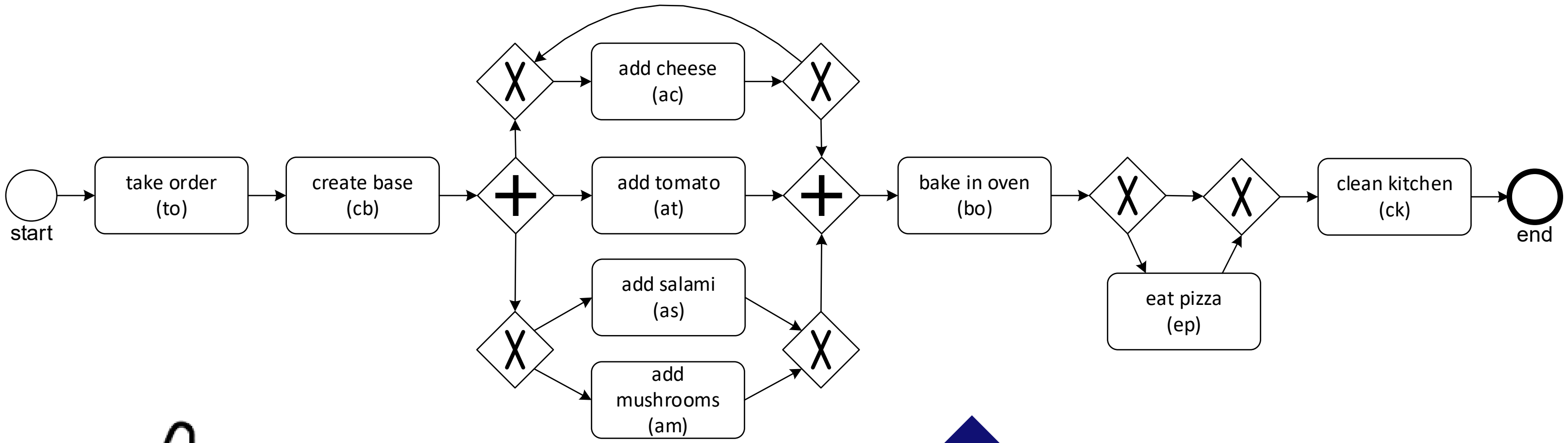


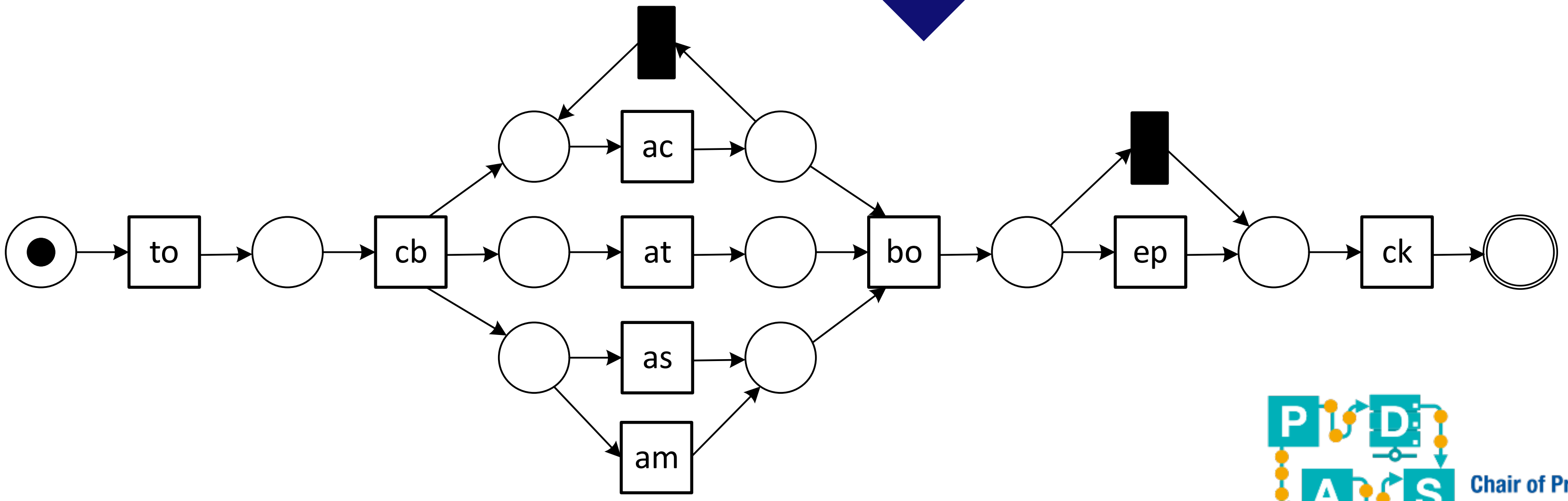
Discovered using IM

Chair of Process and Data Science

This is the de facto behavior (3x)





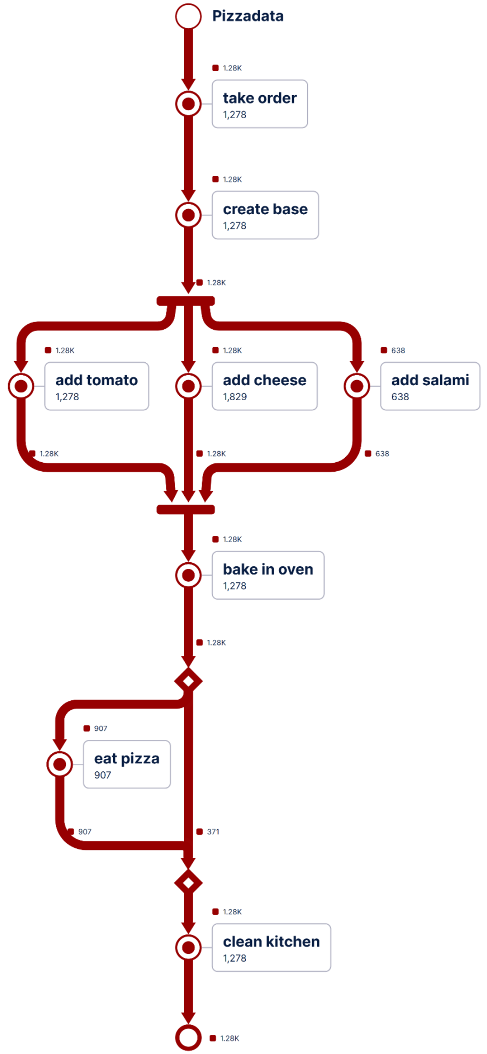


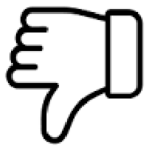
Chair of Process and Data Science

Let’s modify the process model

- Salami is mandatory.
- No mushrooms allowed.
- Just a single serving of cheese.

Assume this is the normative behavior!

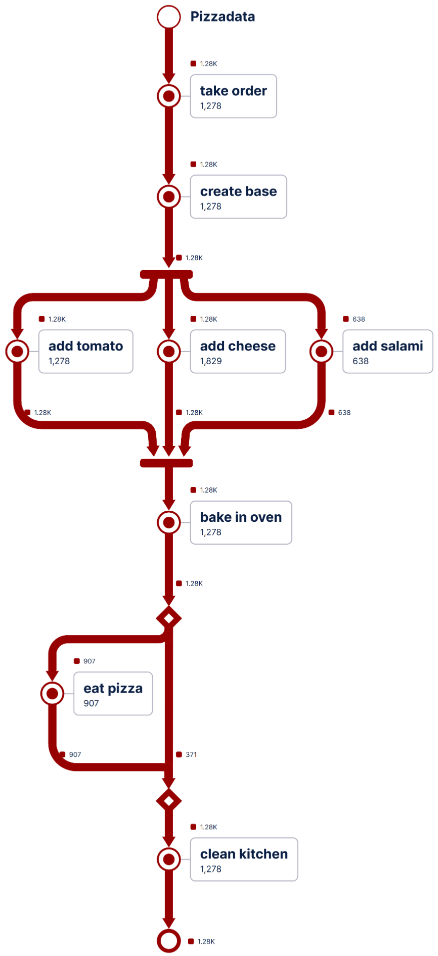


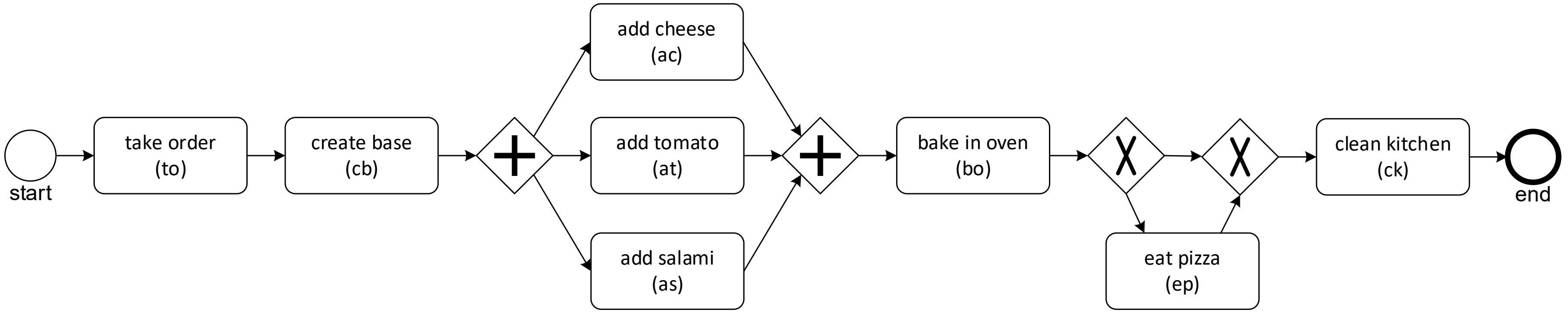


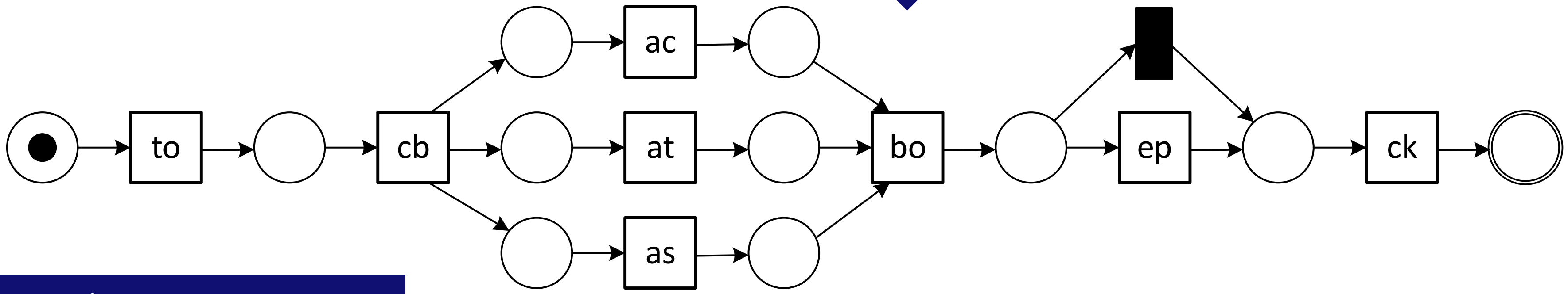
Chair of Process and Data Science

Let's modify the process model …

This is the normative behavior (3x)







- Salami is mandatory.
- No mushrooms allowed.
- Just a single serving of cheese.

Chair of Process and Data Science

Conformance checking using the Celonis Process Adherence Manager

Conformance rate

Pizzadata 33.88%

Deviations (3)

Search

Unexpected event
add mushrooms
640 (50.08%) Pizzadata | +4 minutes TPT

Missing event
add salami
640 (50.08%) Pizzadata | +4 minutes TPT

Occurred too often
add cheese
389 (30.44%) Pizzadata | +4 minutes TPT

There were 640 pizzas with mushrooms instead of salami (50.08%) and 389 pizzas with multiple servings of cheese (30.44%).

Total: 1278 cases/pizzas

Chair of Process and Data Science

Studio > Celosphere25 > Pizza Process

Search CTRL /

1

6 Publish

Pizza Process PAM

Deviations

Variants

Allowed deviations

Explore deviations

Conformance rate

Pizzadata 0%

Deviations (3)

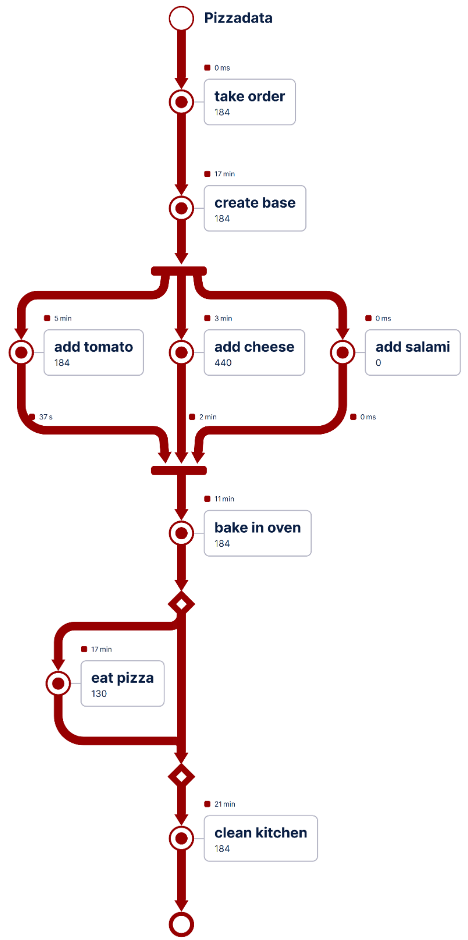
Search

Occurred too often
add cheese
184 (100%) Pizzadata | 0 milliseconds TPT

Unexpected event
add mushrooms
184 (100%) Pizzadata | 0 milliseconds TPT

Missing event
add salami
184 (100%) Pizzadata | 0 milliseconds TPT

184 cases (14.4%) have all three deviations



Filters

Select the data you want to compare to the target model.

Pizzadata
add salami is missing

Pizzadata
add cheese occurred too often

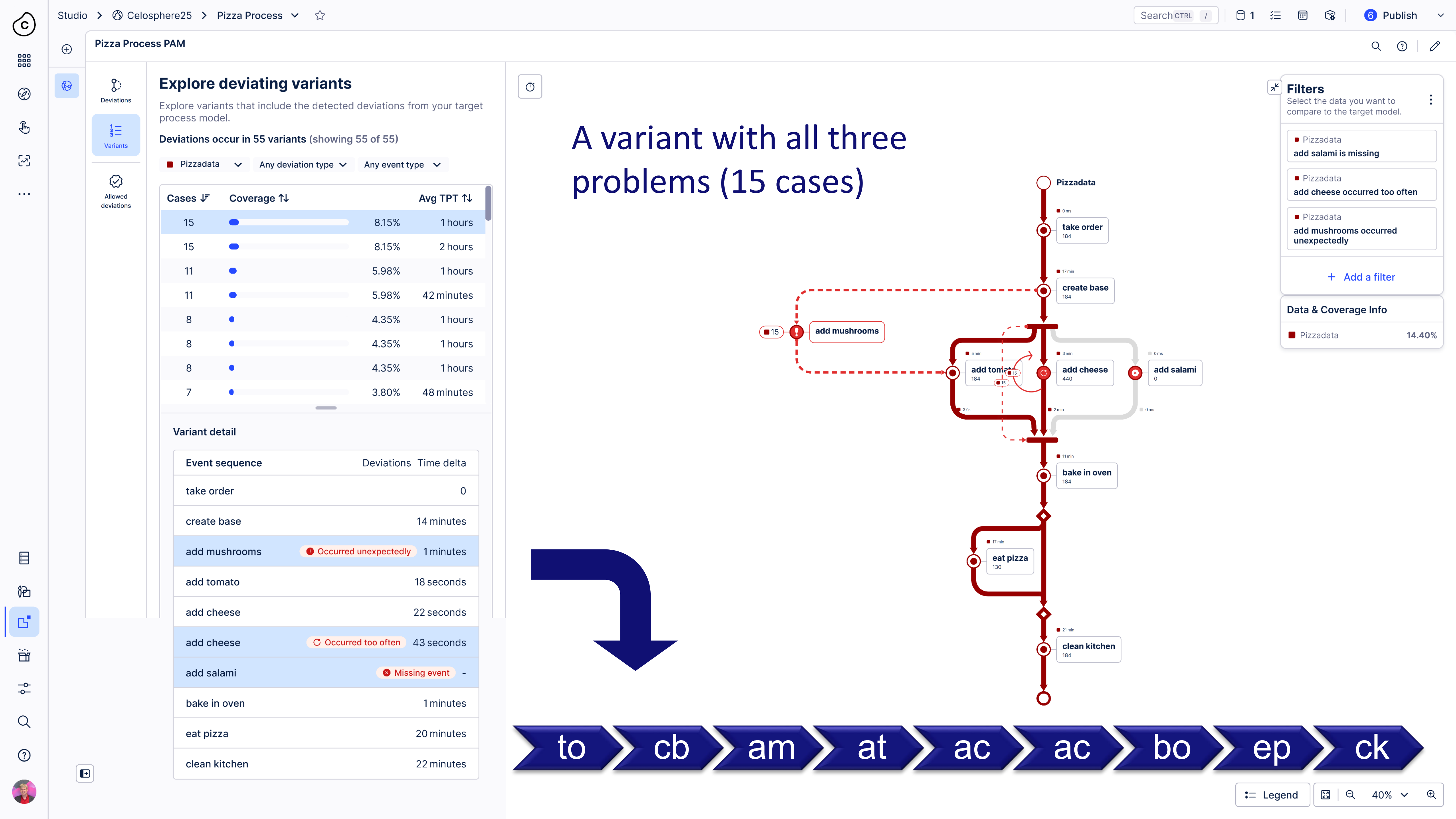
Pizzadata
add mushrooms occurred unexpectedly

+ Add a filter

Data & Coverage Info

Pizzadata 14.40%

Legend 40%



Studio
Celosphere25
Pizza Process
Publish
Pizza Process PAM
Deviations
Variants
Allowed deviations
Explore deviating variants
Explore variants that include the detected deviations from your target process model.
Deviations occur in 55 variants (showing 55 of 55)
Pizzadata
Any deviation type
Any event type
Cases
Coverage
Avg TPT
15 8.15% 1 hours
15 8.15% 2 hours
11 5.98% 1 hours
11 5.98% 42 minutes
8 4.35% 1 hours
8 4.35% 1 hours
8 4.35% 1 hours
7 3.80% 48 minutes
Variant detail
Event sequence
Deviations
Time delta
take order 0
create base 14 minutes
add mushrooms Occurred unexpectedly 1 minutes
add tomato 18 seconds
add cheese 22 seconds
add cheese Occurred too often 43 seconds
add salami Missing event -
bake in oven 1 minutes
eat pizza 20 minutes
clean kitchen 22 minutes
A variant with all three problems (15 cases)
Pizzadata
take order
create base
add mushrooms
add cheese
add salami
bake in oven
eat pizza
clean kitchen
Filters
Select the data you want to compare to the target model.
Pizzadata
add salami is missing
Pizzadata
add cheese occurred too often
Pizzadata
add mushrooms occurred unexpectedly
Add a filter
Data & Coverage Info
Pizzadata 14.40%
to
cb
am
at
ac
ac
bo
ep
ck
Legend
40%

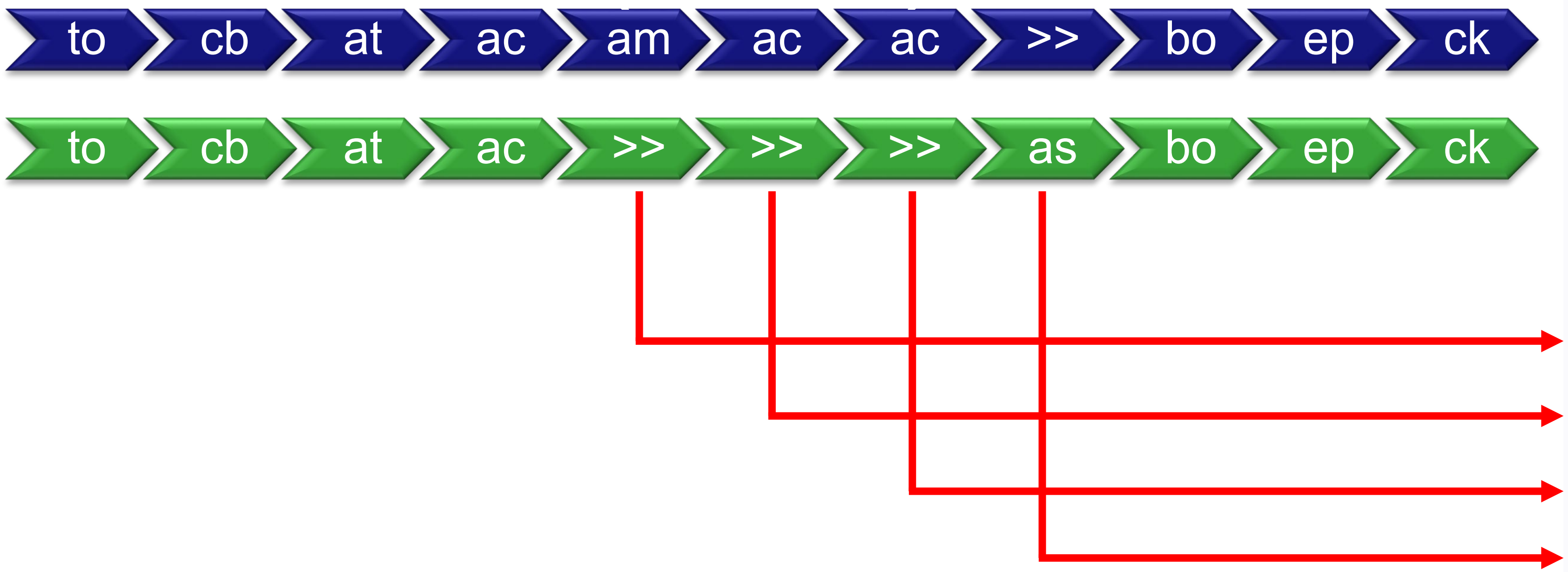
Three deviations!

to	cb	am	at	ac	ac	>>	bo	ep	ck
to	cb	>>	at	ac	>>	as	bo	ep	ck

Variant detail

Event sequence	Deviations	Time delta
take order		0
create base		14 minutes
add mushrooms	Occurred unexpectedly	1 minutes
add tomato		18 seconds
add cheese		22 seconds
add cheese	Occurred too often	43 seconds
add salami	Missing event	-
bake in oven		1 minutes
eat pizza		20 minutes
clean kitchen		22 minutes

Another example with four deviations

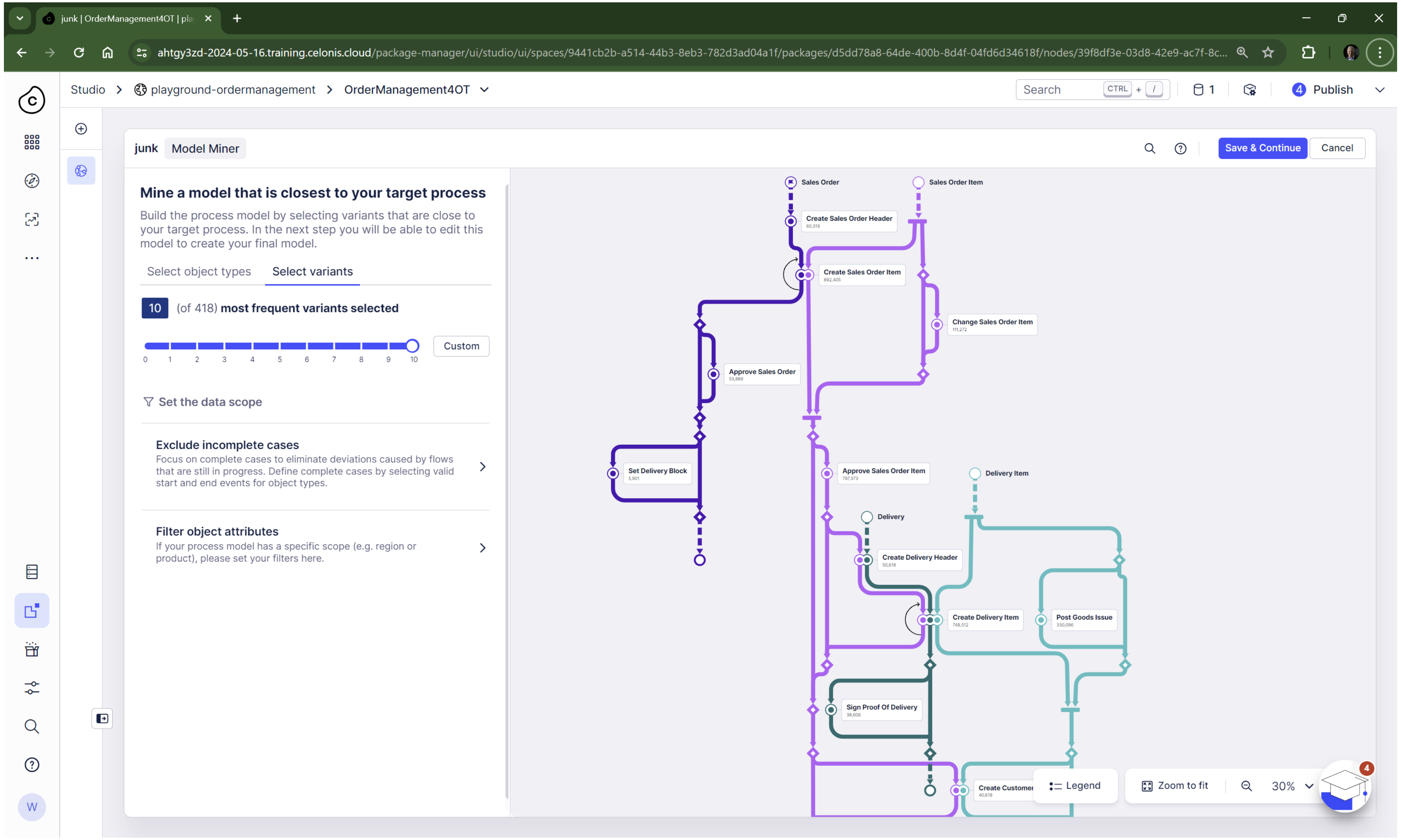


Variant detail

Event sequence	Deviations	Time delta
take order		0
create base		8 minutes
add tomato		1 hours
add cheese		21 seconds
add mushrooms	Occurred unexpectedly	15 seconds
add cheese	Occurred too often	26 seconds
add cheese	Occurred too often	38 seconds
add salami	Missing event	-
bake in oven		2 minutes
eat pizza		7 minutes
clean kitchen		8 minutes

Chair of Process and Data Science

Celonis also supports object-centric conformance checking

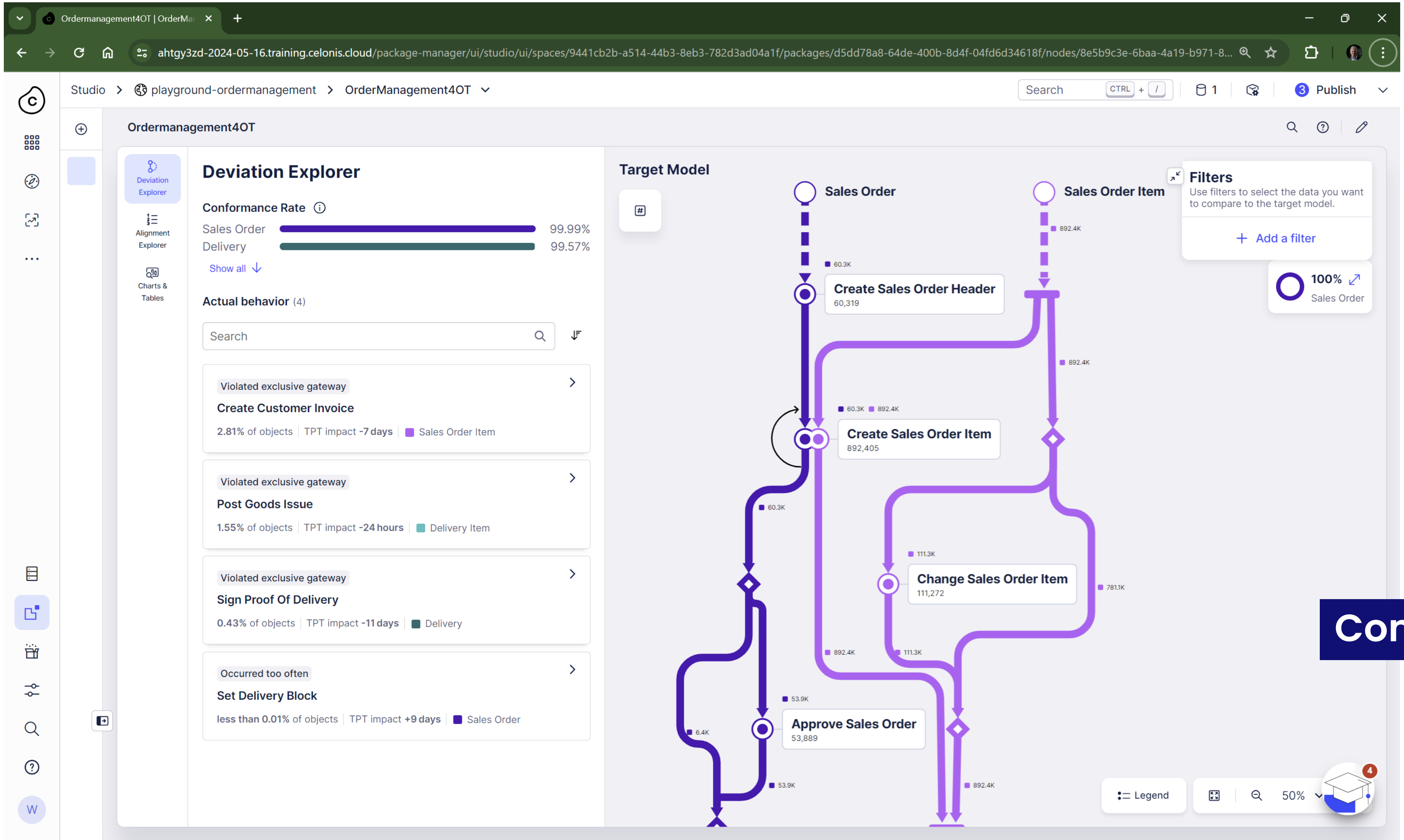


Recall: Object-centric process discovery using an extension of the Inductive Mining (IM) algorithm.

Select a subset of variants before applying IM.

Chair of Process and Data Science

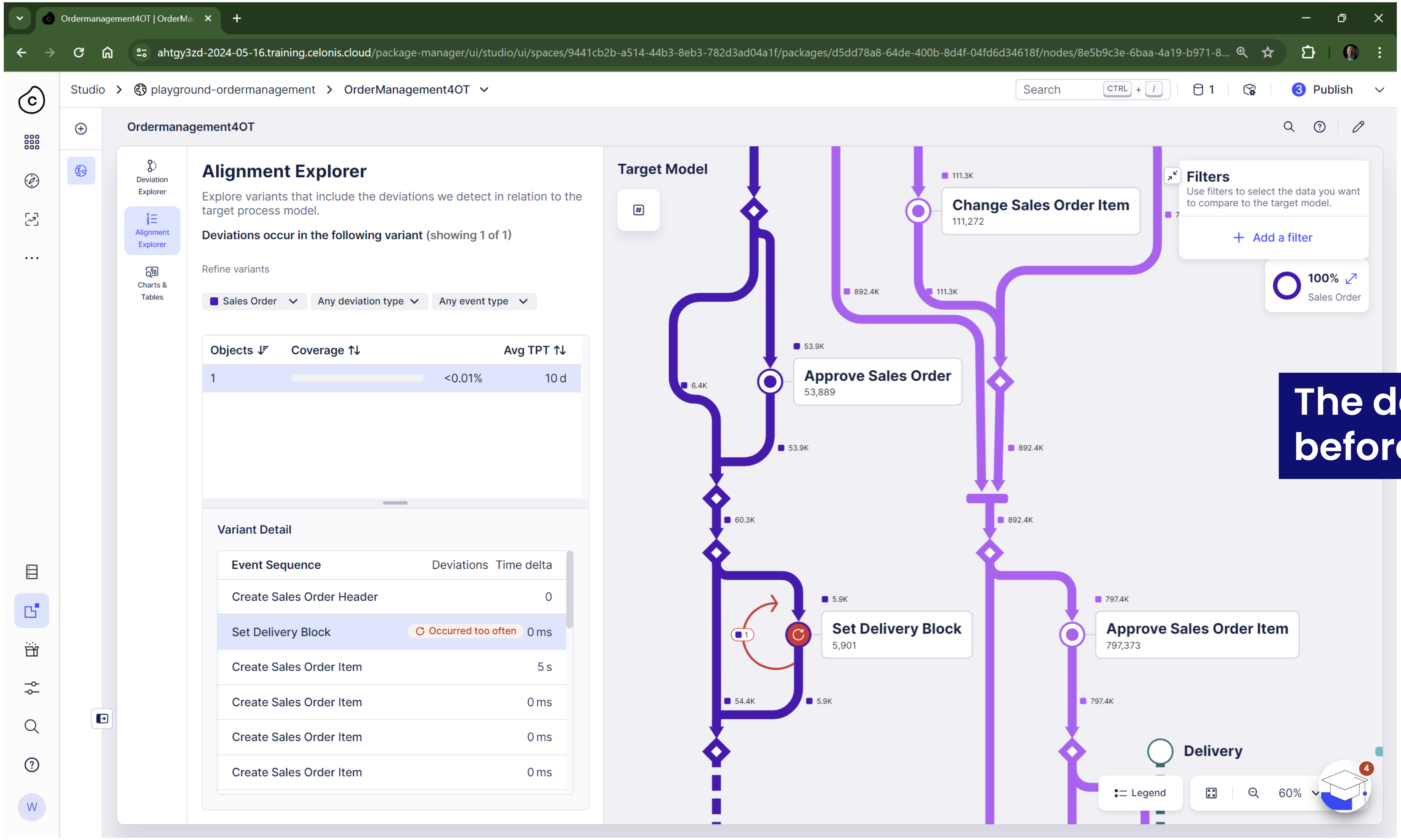
Process Adherence Management (PAM): Check Conformance



Computes alignments

Chair of Process and Data Science

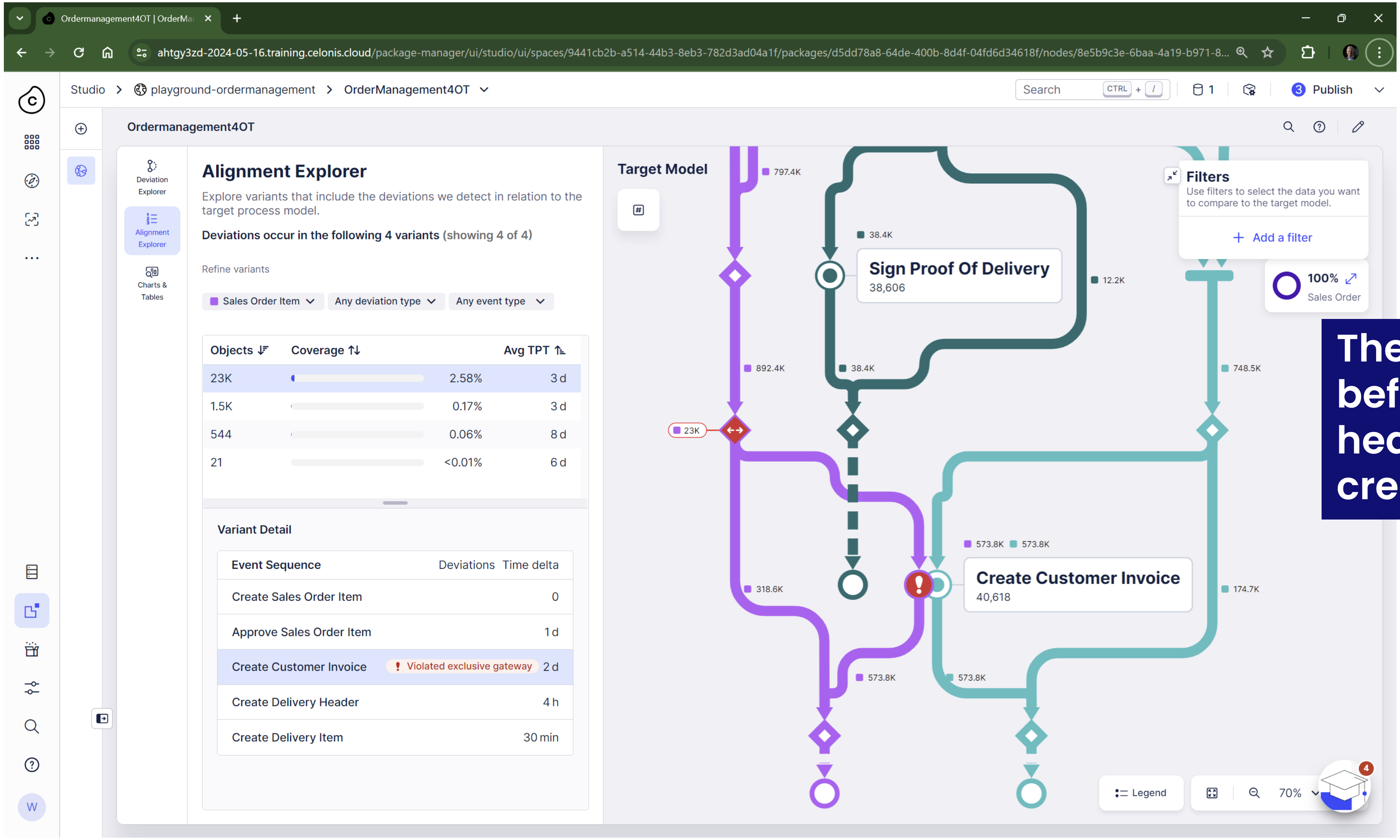
Process Adherence Management (PAM): Show Alignments



The delivery block is set before items are created

Chair of Process and Data Science

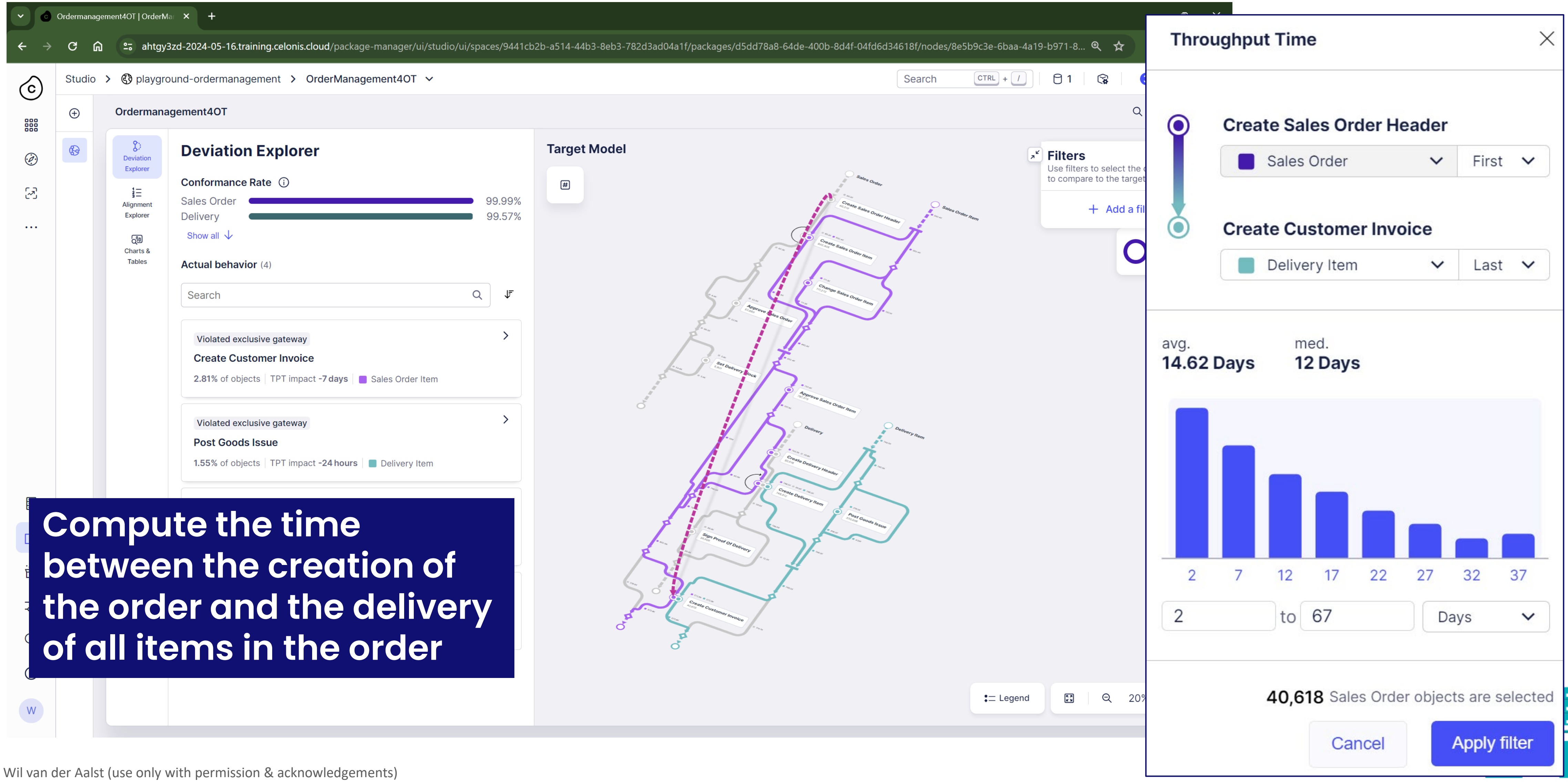
Process Adherence Management (PAM): Show Alignments



The invoice is created before the delivery header and item are created.

Chair of Process and Data Science

Process Adherence Management (PAM): Analyze Performance

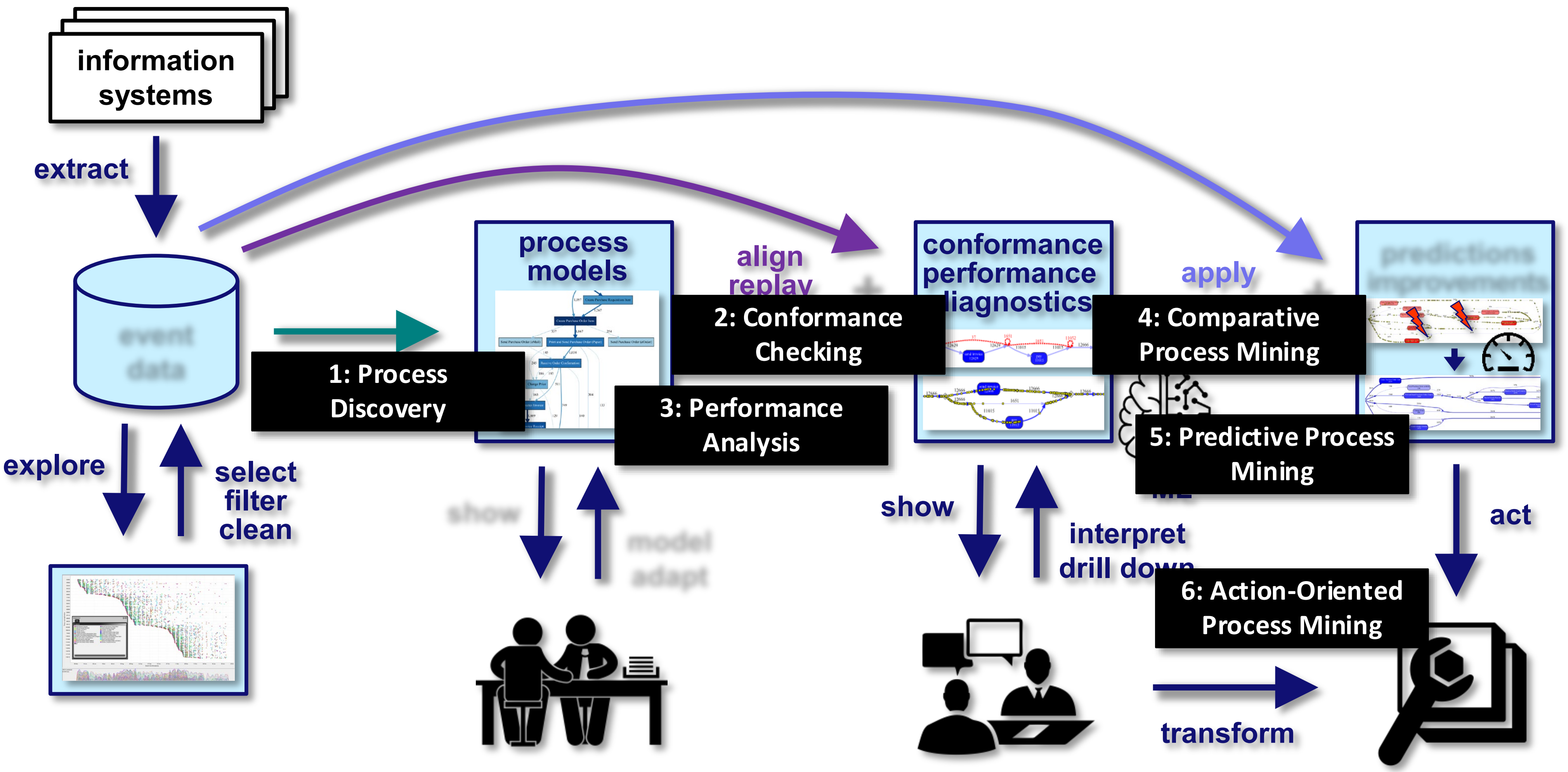


Compute the time between the creation of the order and the delivery of all items in the order

Chair of Process and Data Science

Conclusion

There is much more …



Chair of Process and Data Science

Websites

- www.processmining.org
- www.process-mining-summer-school.org
- www.tf-pm.org
- www.promtools.org
- www.celonis.com/academic-signup
- xes-standard.org
- ocel-standard.org
- www.pads.rwth-aachen.de
- www.vdaalst.com

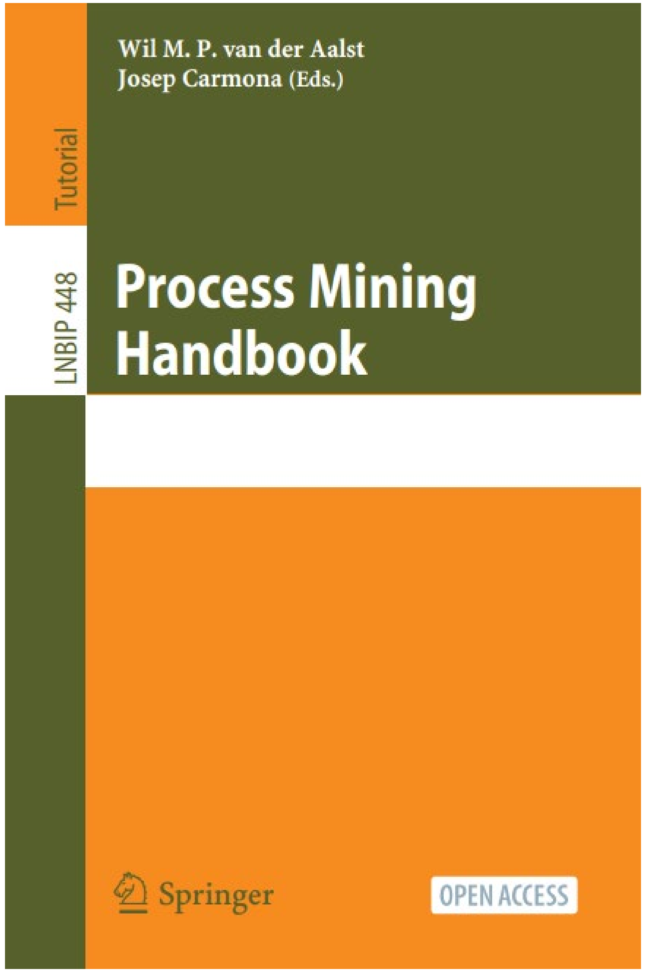
Chair of Process and Data Science

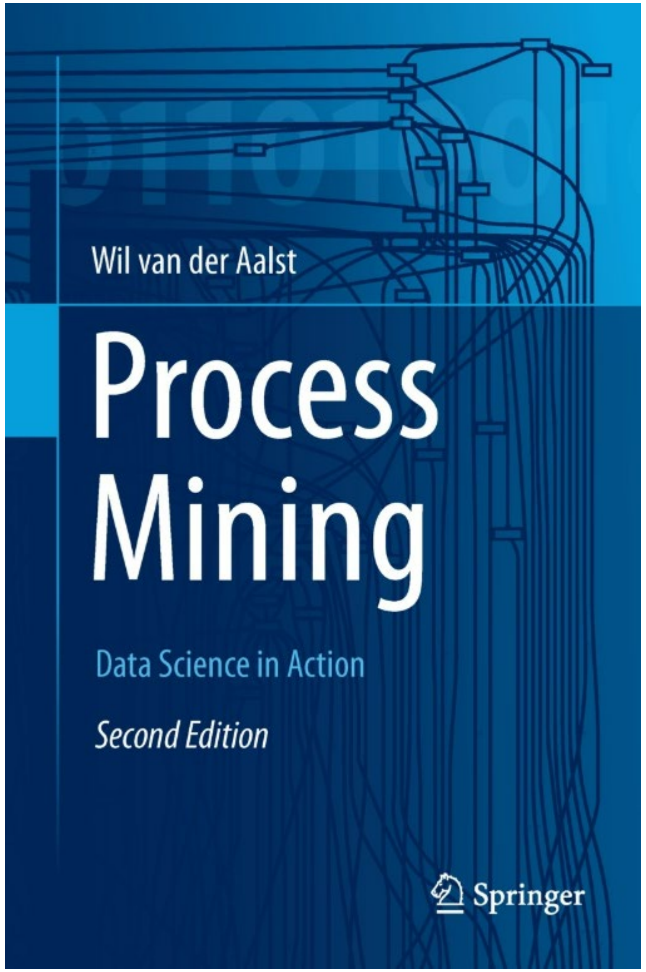
Online courses

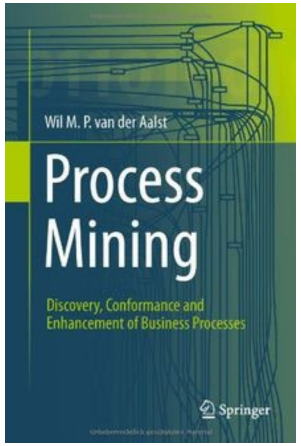
- **Coursera course "Process Mining: Data science in Action"**
  Register via coursera.org/learn/process-mining (>175.000 participants since 2015).
- **Celonis/RWTH course "Process Mining: From Theory to Execution"**
  Register via www.celonis.com/wils-process-mining-class.
- **edX course "A Hands-on Introduction to Process Mining"**
  Register via RWTHx: www.edx.org/school/rwthx.
- **edX course "Object Centric Process Mining"**
  Register via RWTHx: www.edx.org/school/rwthx.

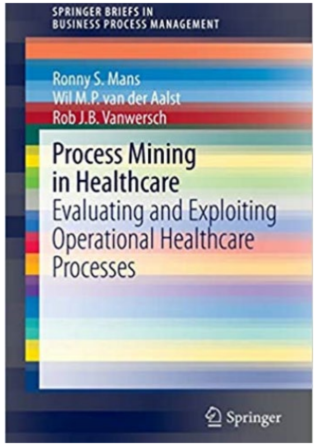
Chair of Process and Data Science

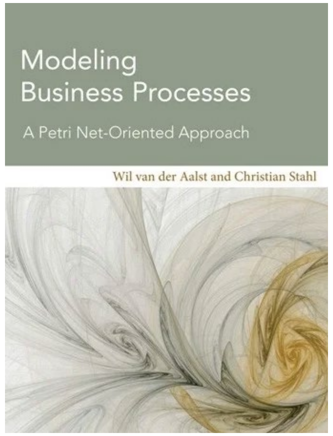
Books (not intended to be complete)

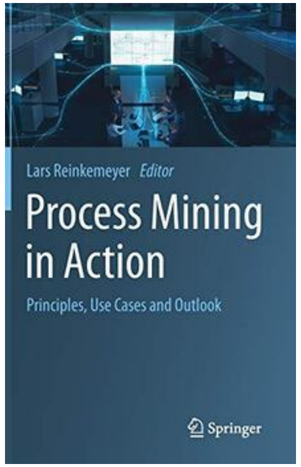


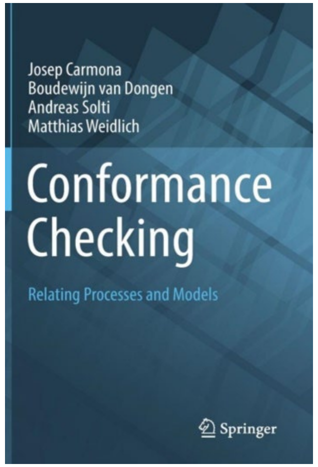












Chair of Process and Data Science