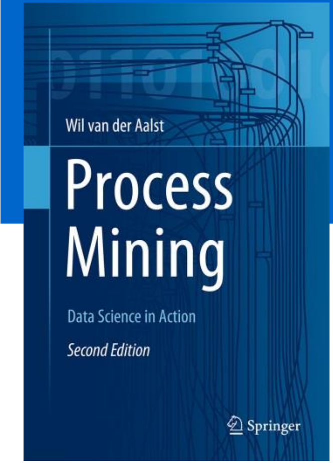
Learning Hybrid Process Models from Events

Process Mining for the Real World

Smart Data Analytics (SDA) research group,
University of Bonn, 29-11-2017

SMART
DATA
ANALYTICS
FROM DATA TO KNOWLEDGE



Wil van der Aalst

Process Mining

Data Science in Action

Second Edition

Springer

Wil van der Aalst

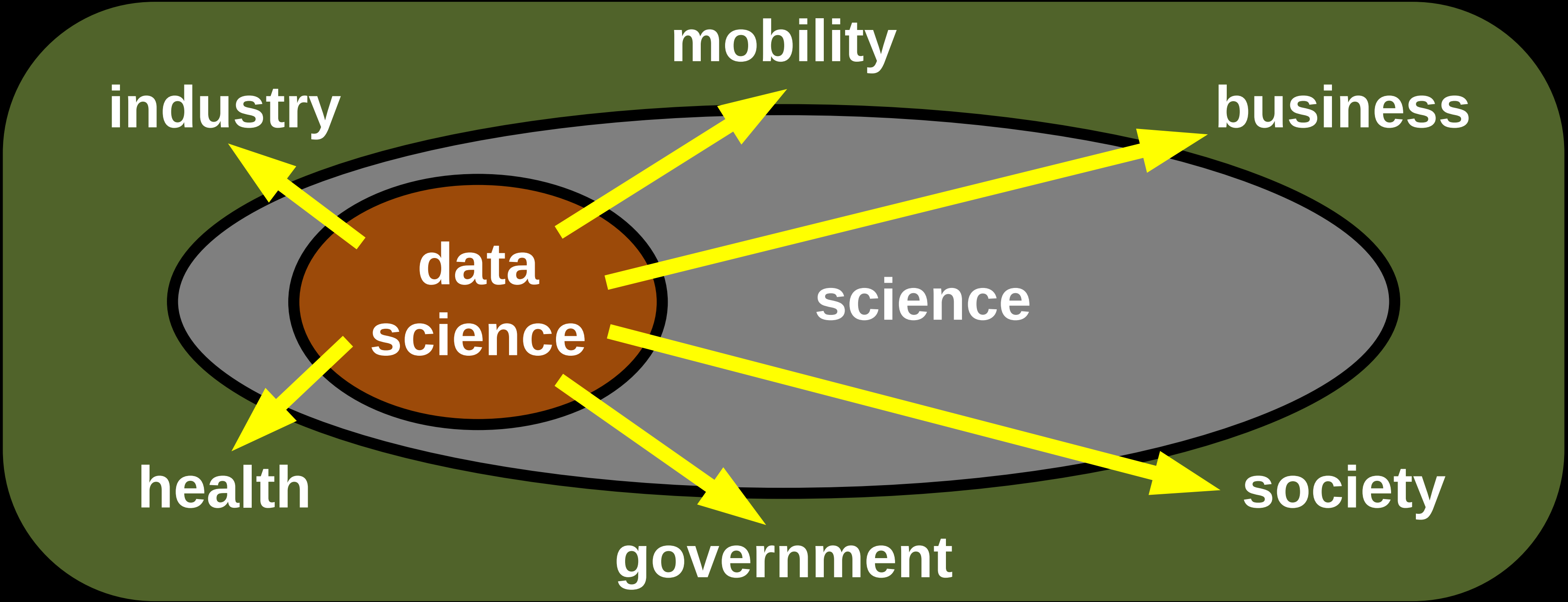
www.vdaalst.com | @wvdaalst

TU/e

TU/e
DSC/e

process
mining
hybrid
process
discovery

Positioning PM

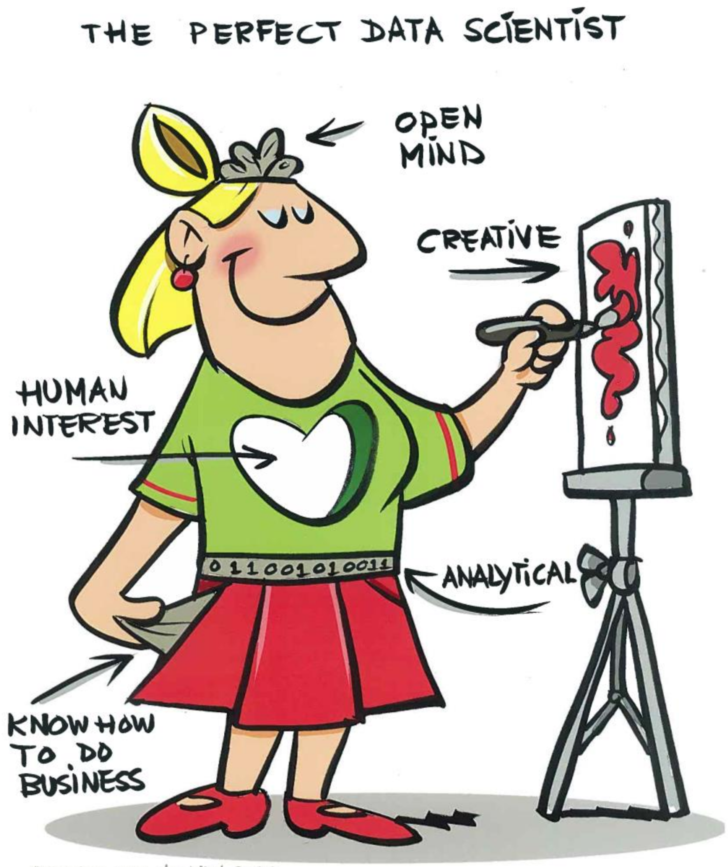

mobility
industry
business
data science
science
health
society
government

03809
01334

Launched in 2013

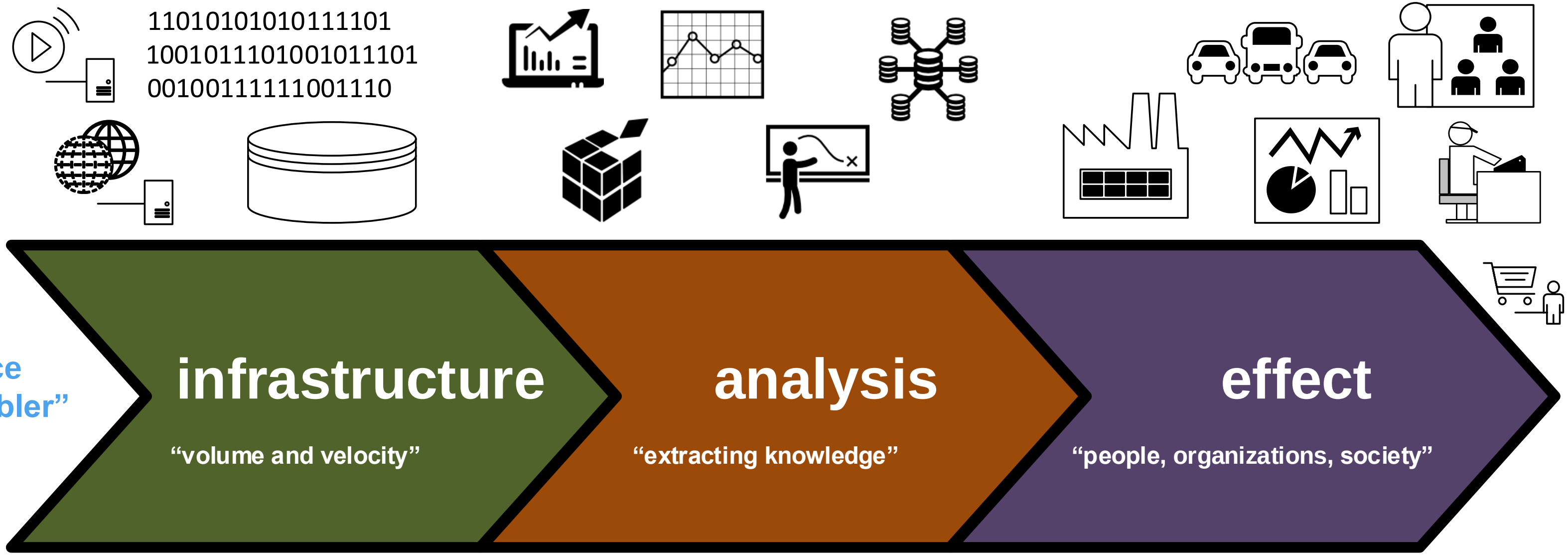
DSC/e
Data Science Center Eindhoven
Turning Data into Value
Wil van der Aalst
Scientific Director DSC/e
TU/e Technische Universiteit Eindhoven University of Technology
Technische
Eindhoven

DSC/e Technische Universiteit Eindhoven University of Technology



©Marion van de Wiel 2014

Data Science as "the enabler"



infrastructure

"volume and velocity"

analysis

"extracting knowledge"

effect

"people, organizations, society"

- instrumentation
- big data infrastructures and distributed systems
- databases and data management
- programming
- security
- ...

- statistics
- data/process mining
- machine learning/artificial intelligence
- operations research
- algorithms
- visualization
- ...

- ethics & privacy
- IT law
- human technology interaction
- operations management
- business models
- entrepreneurship
- ...

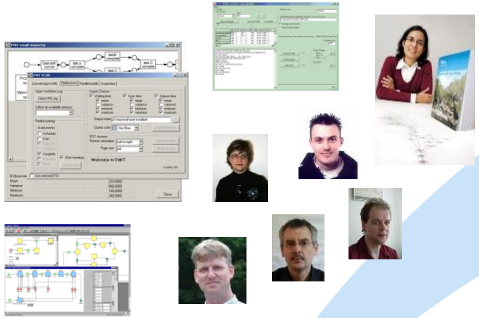
the data science pipeline

JADS

Jheronimus
Academy
of Data Science

☯

Uptake of process mining



Start of process mining at TU/e (1999)

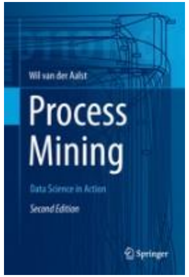
Alpha miner & Heuristic miner (2000-2002)

First version of ProM (2004)

Conformance checking, other perspectives, prediction, etc. (2005 -)

Process mining book, courses, uptake commercial tools, etc. (2011 -)

coursera



Wil van der Aalst

Process Mining

Data Science in Action

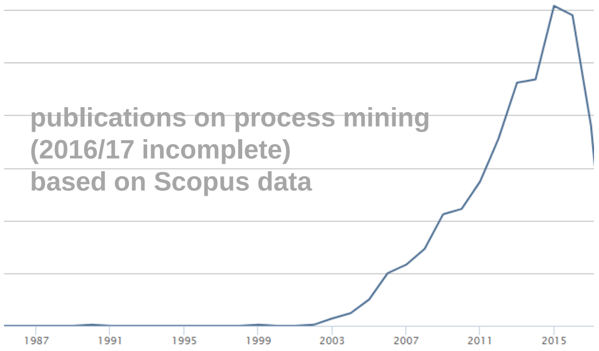
Second Edition

Springer

ProM

process mining workbench

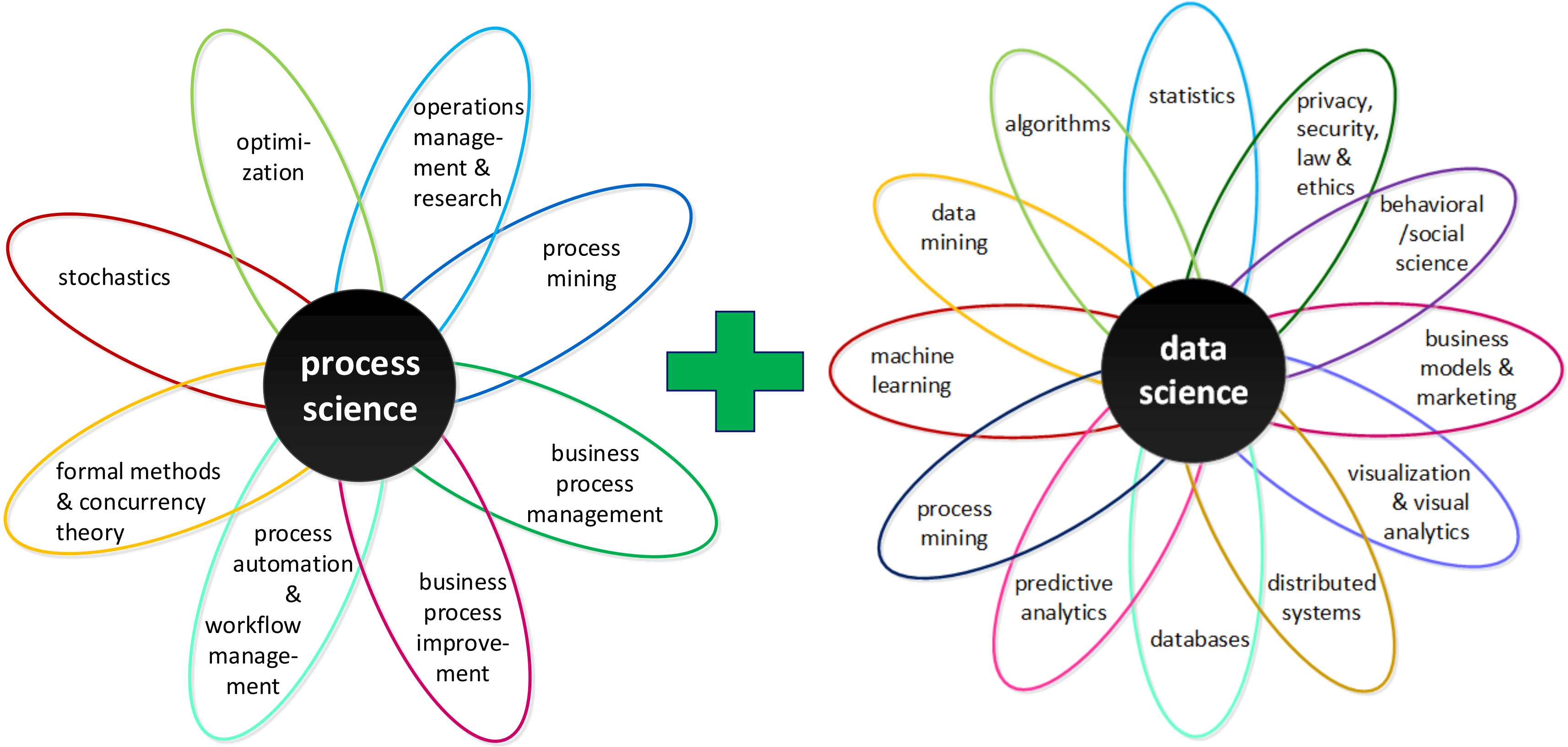
Evolved from 29 plug-ins in 2004 to 274 plug-ins 2009 (ProM 5.2) to over 1500 plug-ins today.



publications on process mining (2016/17 incomplete) based on Scopus data

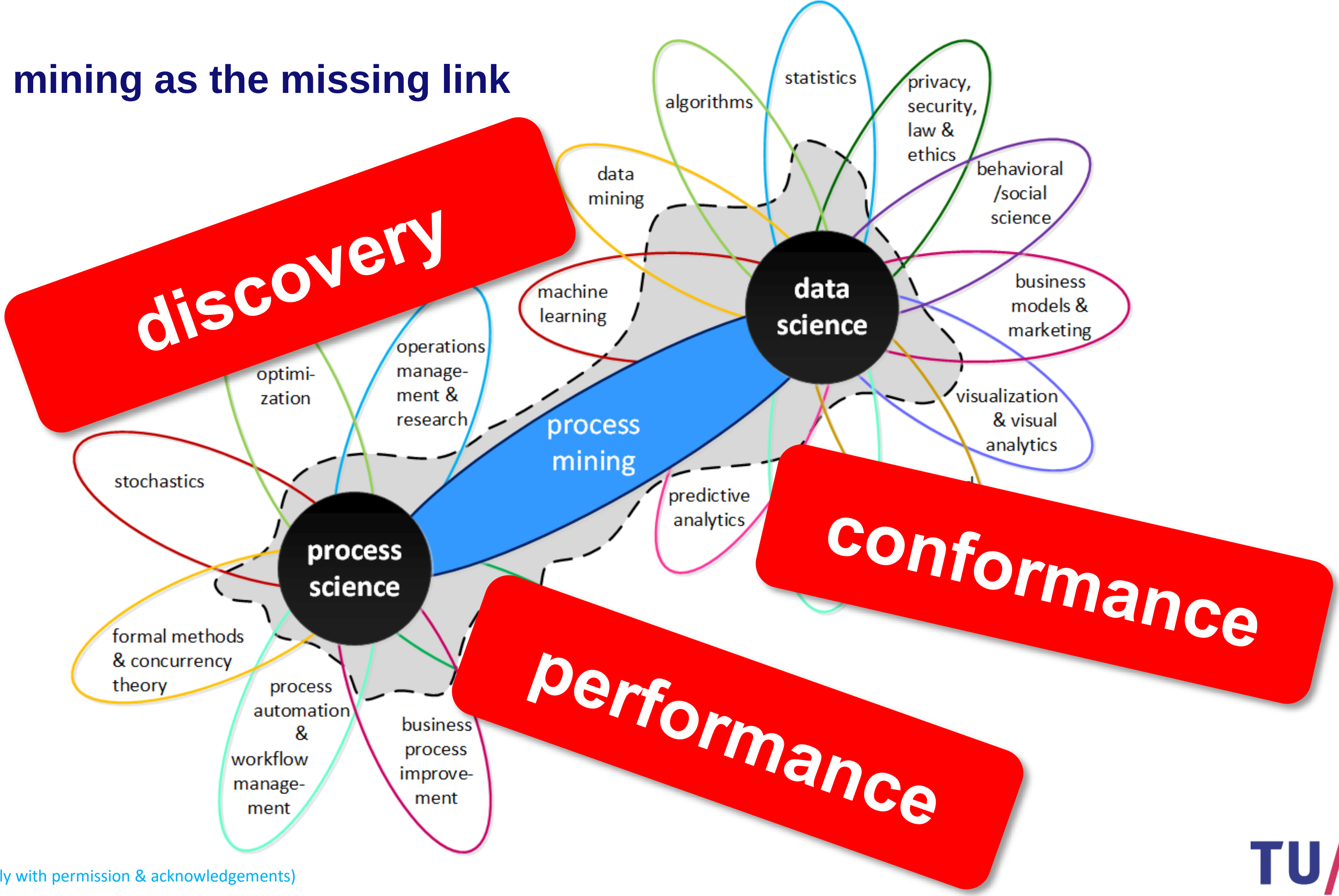
©Wil van der Aalst (use only with permission & acknowledgements)

TU/e

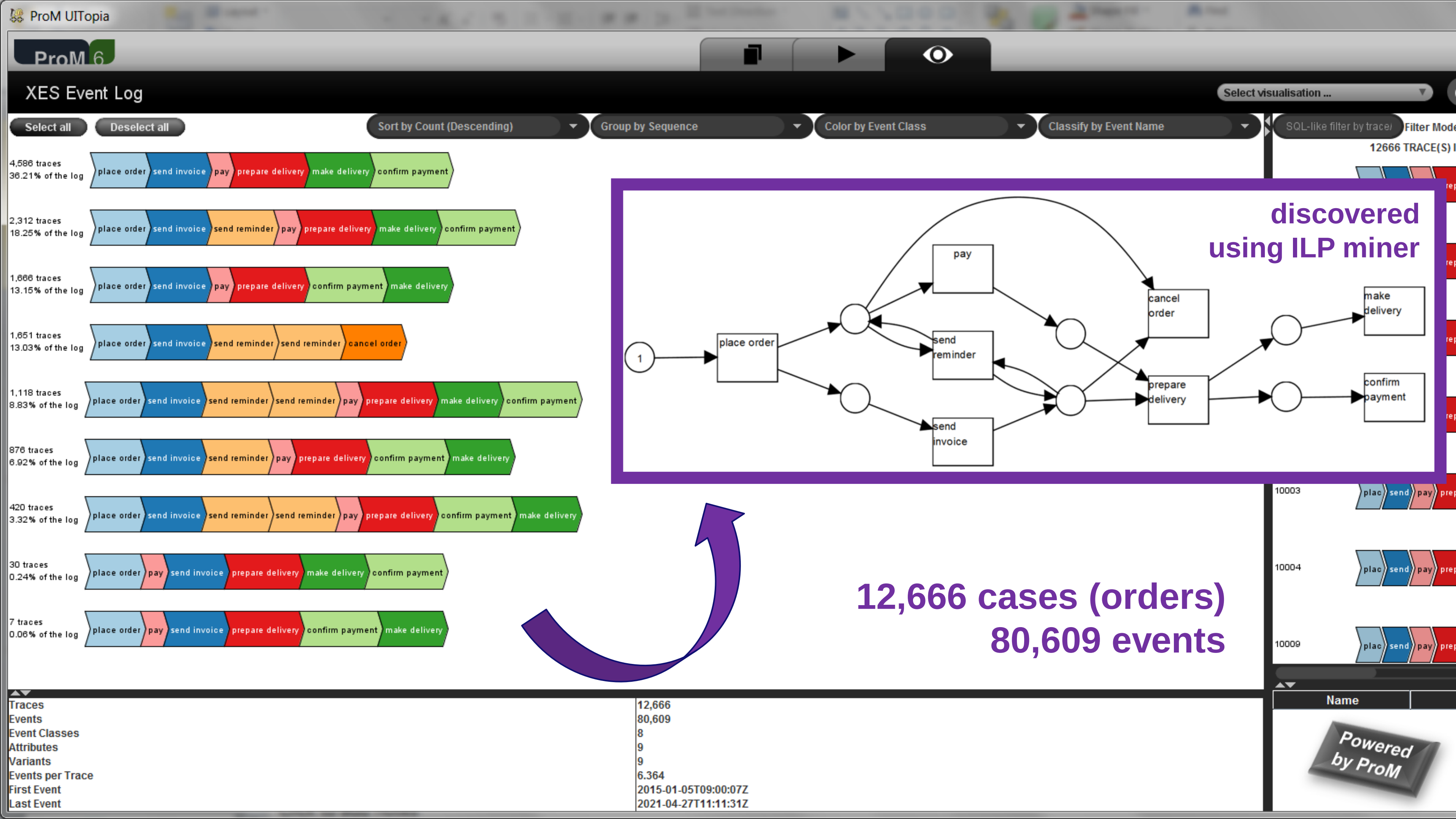

process science
optimi-zation
operations manage-ment & research
process mining
business process management
business process improve-ment
process automation & workflow manage-ment
formal methods & concurrency theory
stochastics
+
data science
statistics
privacy, security, law & ethics
behavioral /social science
business models & marketing
visualization & visual analytics
distributed systems
databases
predictive analytics
process mining
machine learning
data mining
algorithms

TU/e

process mining as the missing link


discovery
conformance
performance
process mining
data science
process science
algorithms
statistics
privacy, security, law & ethics
data mining
behavioral /social science
machine learning
business models & marketing
visualization & visual analytics
predictive analytics
optimization
operations management & research
stochastics
formal methods & concurrency theory
process automation & workflow management
business process improvement

©Wil van der Aalst (use only with permission & acknowledgements)



4,586 traces 36.21% of the log: place order, send invoice, pay, prepare delivery, make delivery, confirm payment

2,312 traces 18.25% of the log: place order, send invoice, send reminder, pay, prepare delivery, make delivery, confirm payment

1,666 traces 13.15% of the log: place order, send invoice, pay, prepare delivery, confirm payment, make delivery

1,651 traces 13.03% of the log: place order, send invoice, send reminder, send reminder, cancel order

1,118 traces 8.83% of the log: place order, send invoice, send reminder, send reminder, pay, prepare delivery, make delivery, confirm payment

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

30 traces 0.24% of the log: place order, pay, send invoice, prepare delivery, make delivery, confirm payment

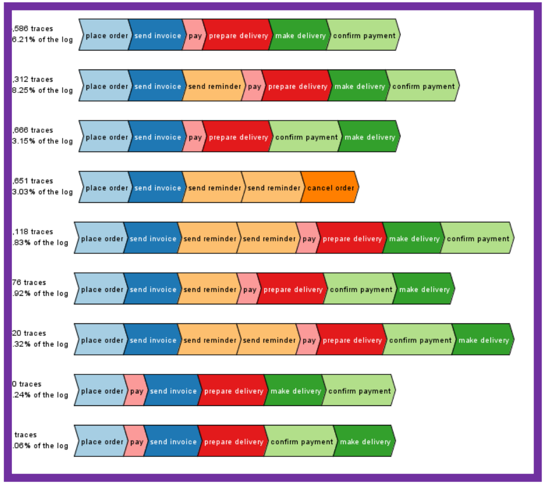
7 traces 0.06% of the log: place order, pay, send invoice, prepare delivery, confirm payment, make delivery

Traces	12,666
Events	80,609
Event Classes	8
Attributes	9
Variants	9
Events per Trace	6.364
First Event	2015-01-05T09:00:07Z
Last Event	2021-04-27T11:11:31Z

discovered using ILP miner

12,666 cases (orders)
80,609 events

Taxonomy: Not just CF discovery!



,586 traces 6.21% of the log — place order, send invoice, pay, prepare delivery, make delivery, confirm payment

,312 traces 8.25% of the log — place order, send invoice, send reminder, pay, prepare delivery, make delivery, confirm payment

,666 traces 3.15% of the log — place order, send invoice, pay, prepare delivery, confirm payment, make delivery

,651 traces 3.03% of the log — place order, send invoice, send reminder, send reminder, cancel order

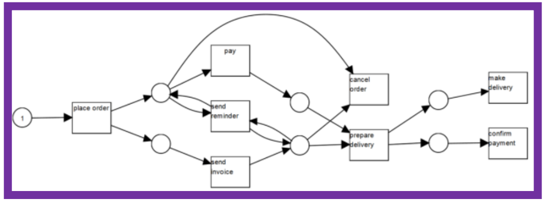
,118 traces .83% of the log — place order, send invoice, send reminder, send reminder, pay, prepare delivery, make delivery, confirm payment

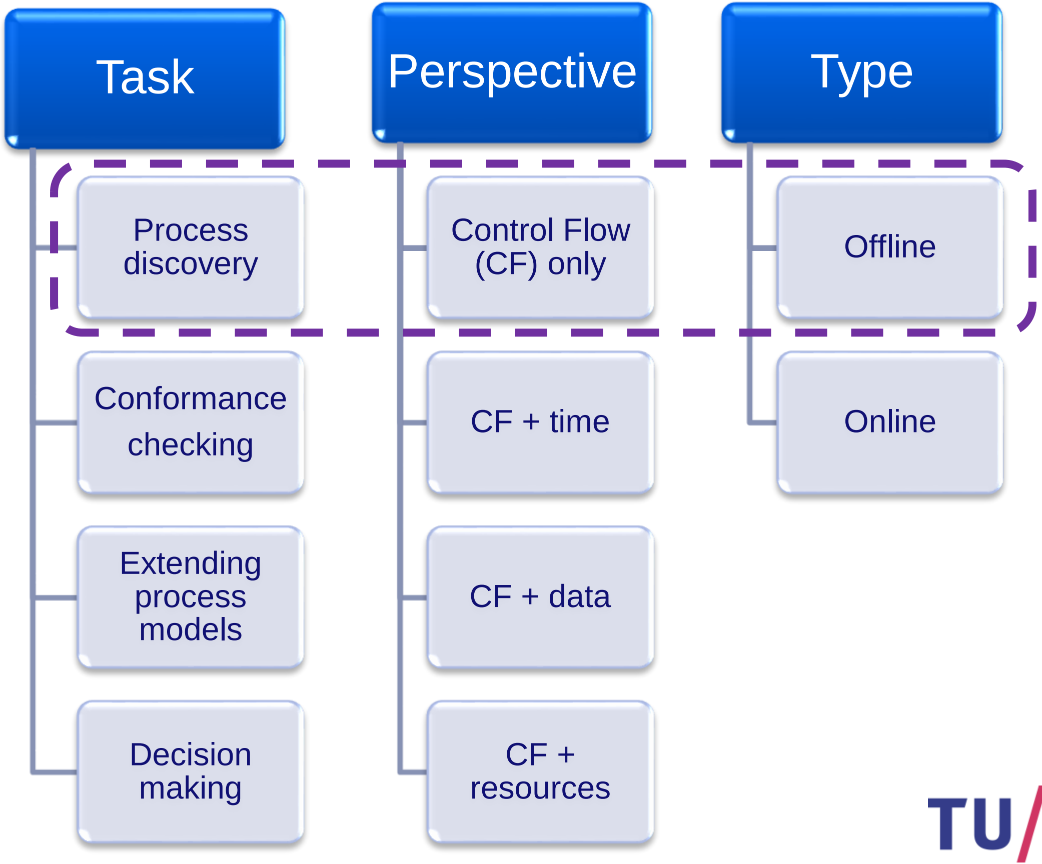
76 traces .92% of the log — place order, send invoice, send reminder, pay, prepare delivery, confirm payment, make delivery

20 traces .32% of the log — place order, send invoice, send reminder, send reminder, pay, prepare delivery, confirm payment, make delivery

0 traces .24% of the log — place order, pay, send invoice, prepare delivery, make delivery, confirm payment

traces .06% of the log — place order, pay, send invoice, prepare delivery, confirm payment, make delivery





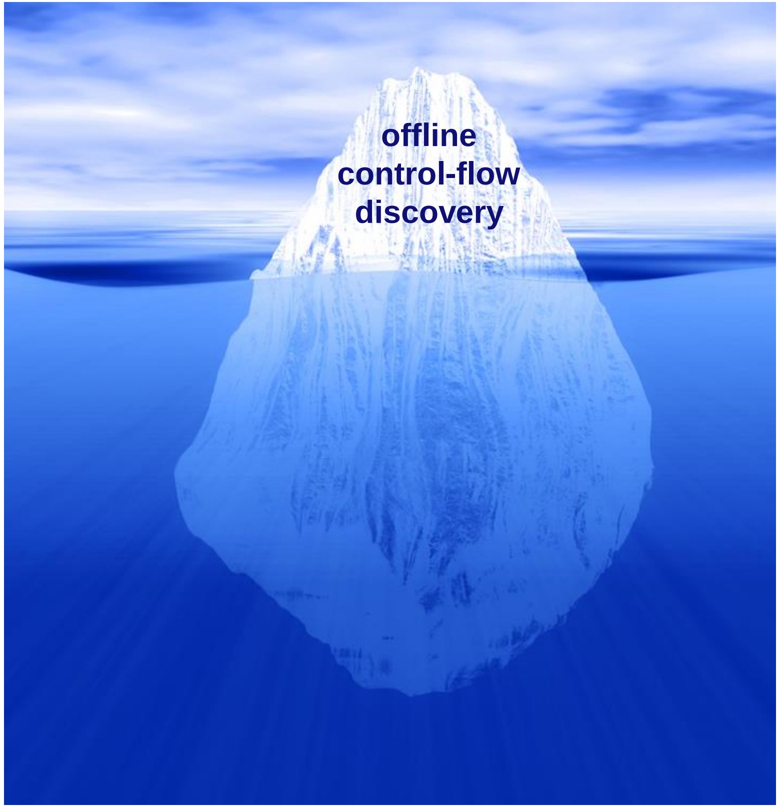
Task
- Process discovery
- Conformance checking
- Extending process models
- Decision making

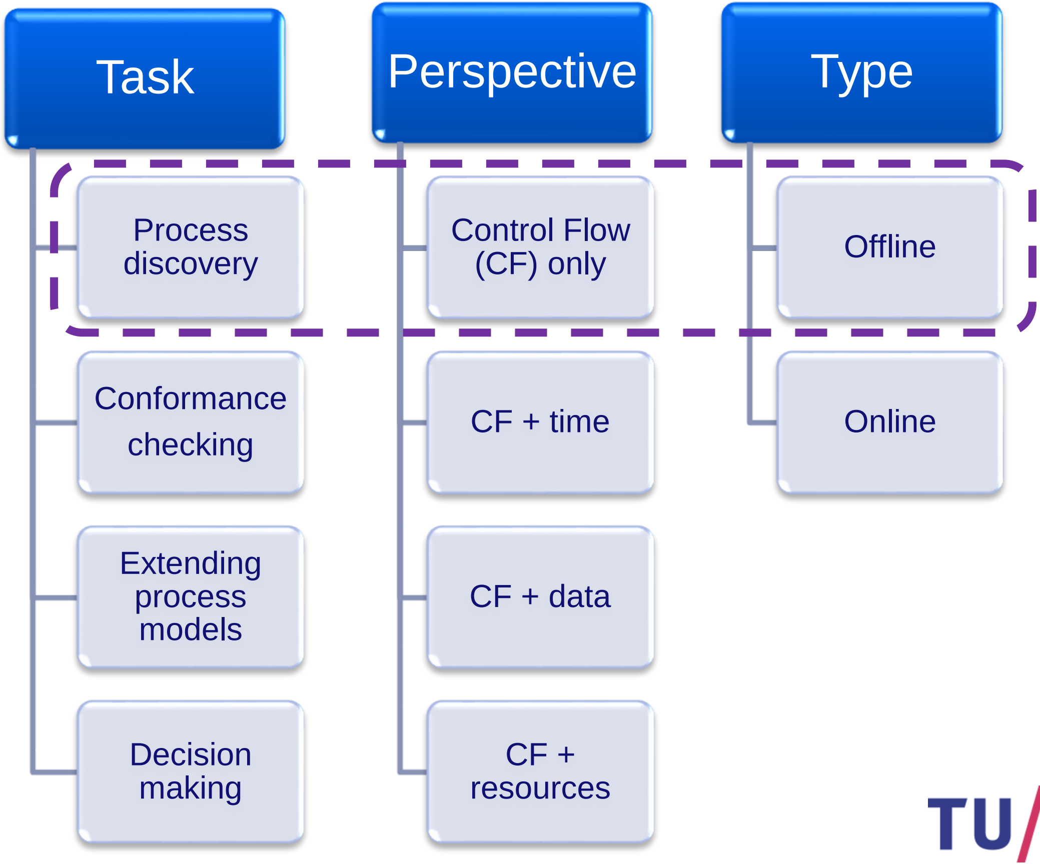
Perspective
- Control Flow (CF) only
- CF + time
- CF + data
- CF + resources

Type
- Offline
- Online

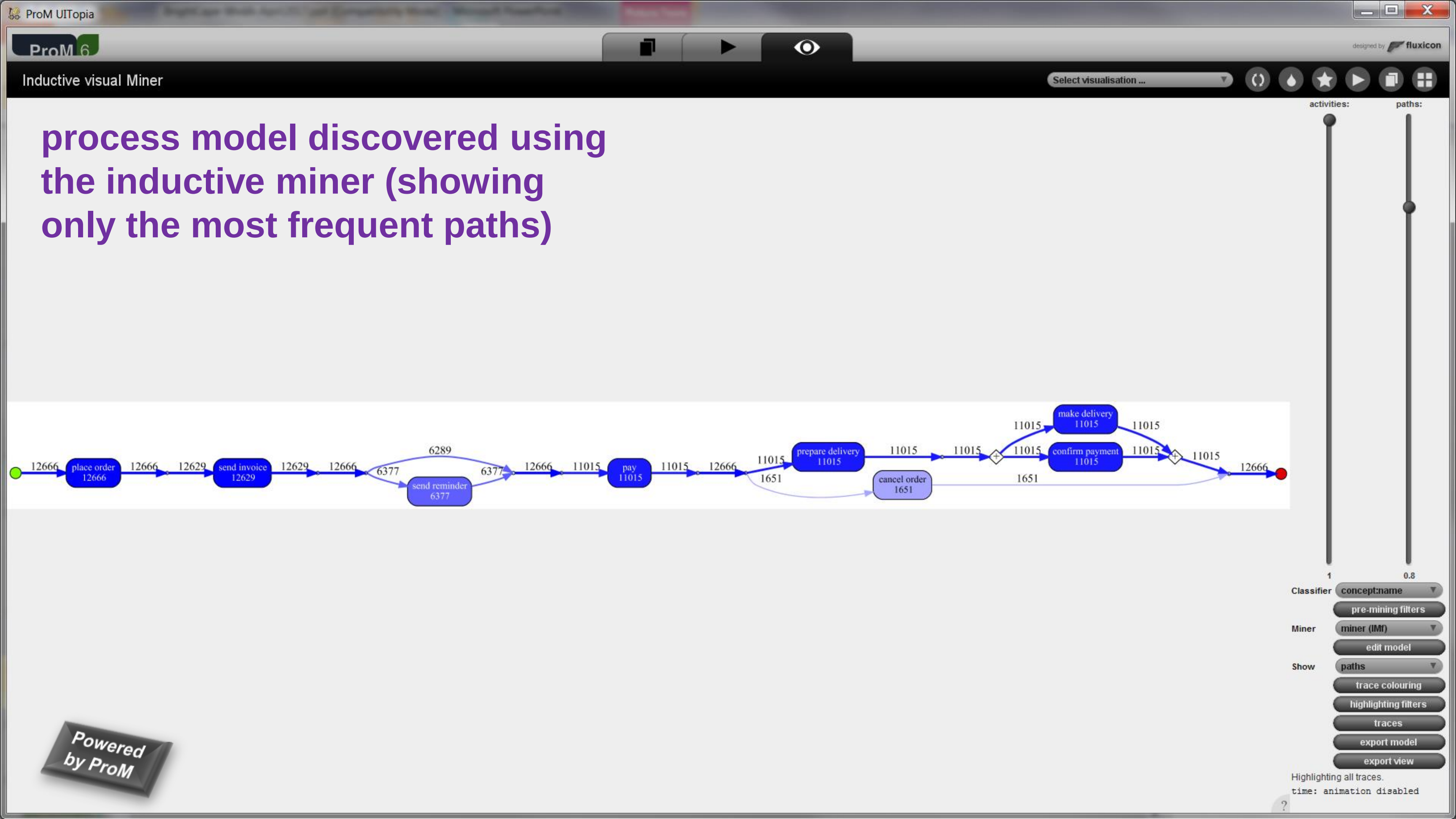
©Wil van der Aalst (use only with permission & acknowledgements)

Taxonomy: Not just CF discovery!





process model discovered using the inductive miner (showing only the most frequent paths)



Classifier concept:name

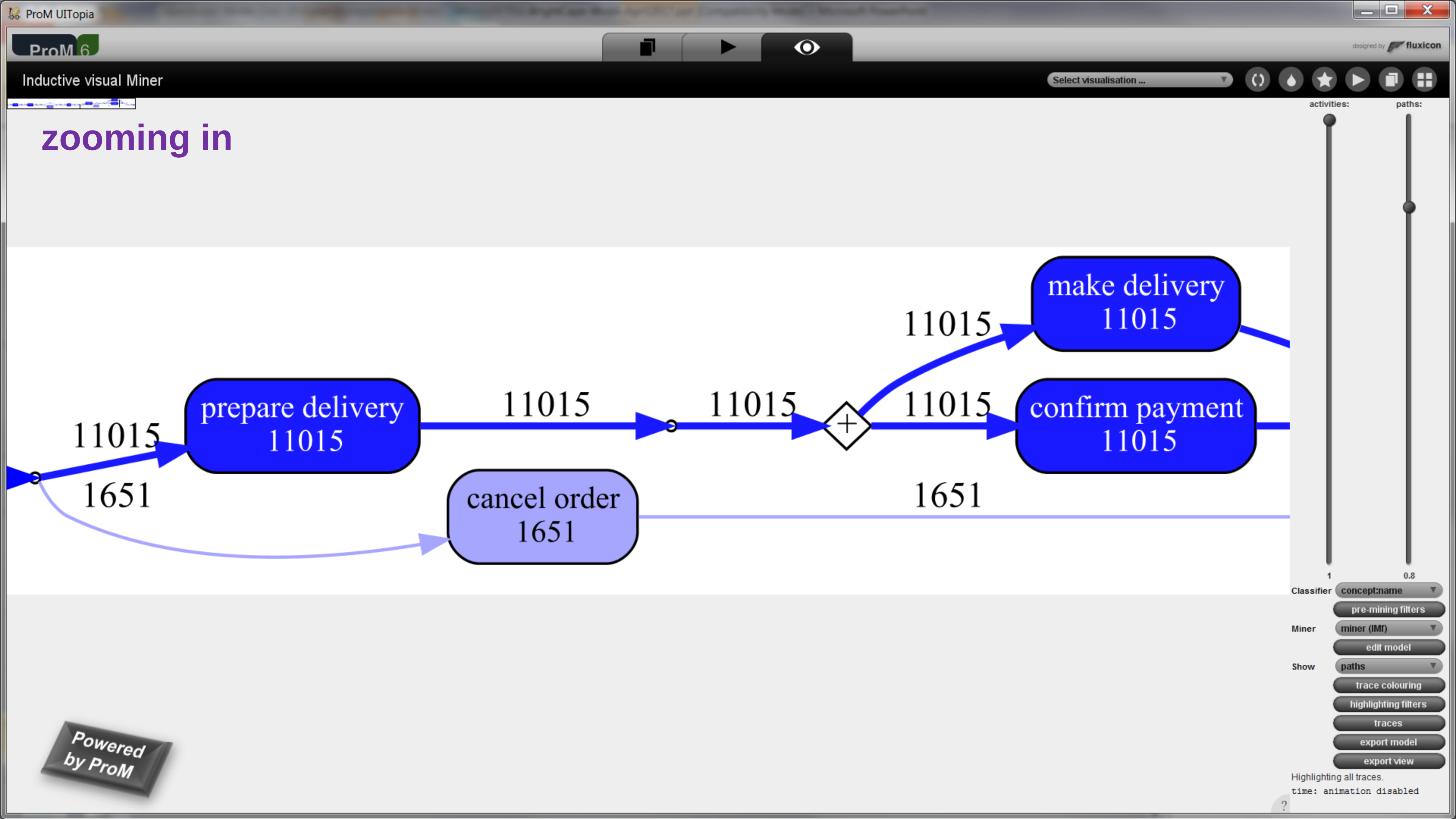
Miner miner (IMf)

Show paths

Highlighting all traces.

time: animation disabled

zooming in



ProM UITopia

Inductive visual Miner

Classifier: concept:name

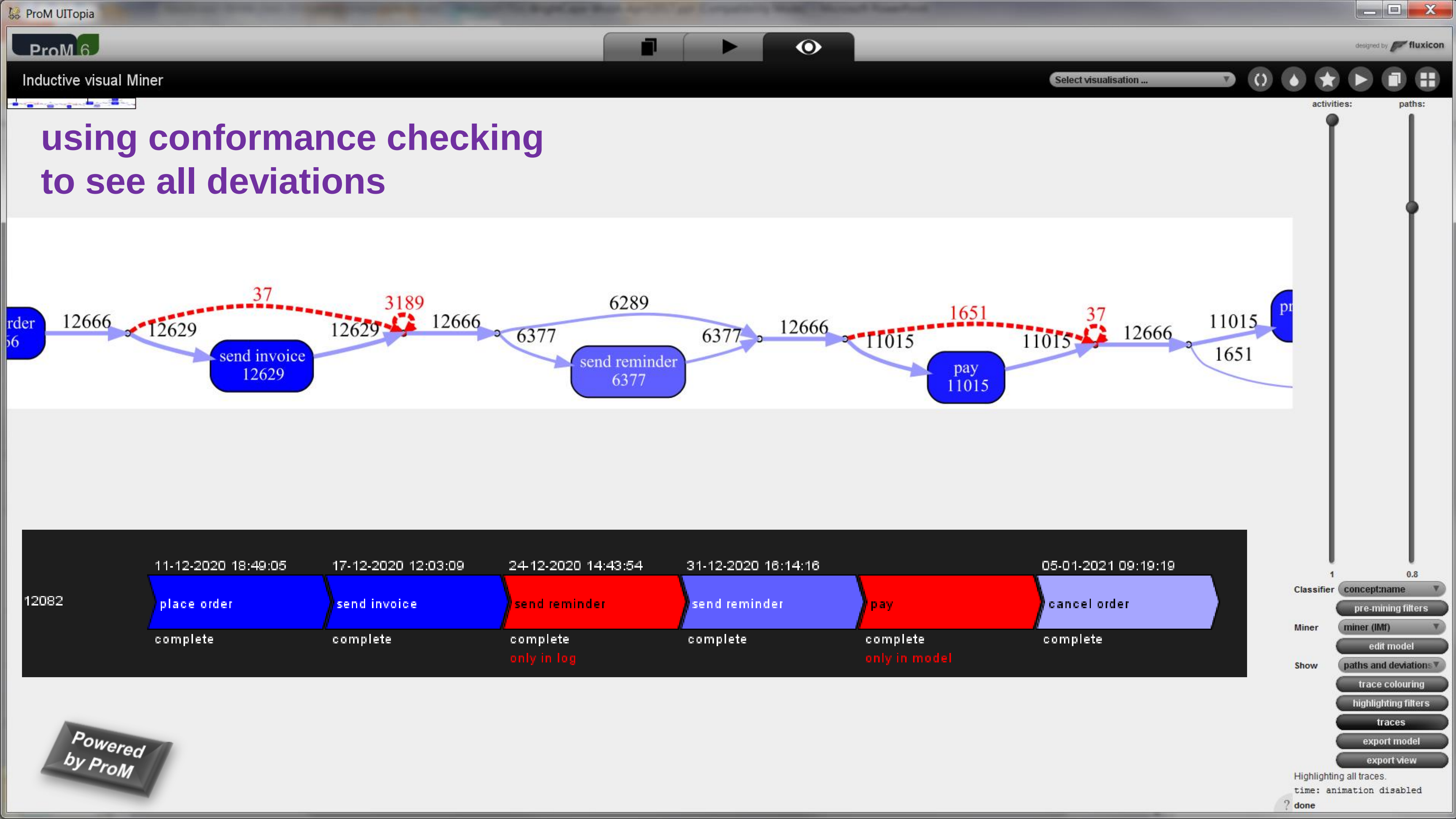
Miner: miner (IMf)

Show: paths

Highlighting all traces.

```
time: animation disabled
```

using conformance checking to see all deviations



12082

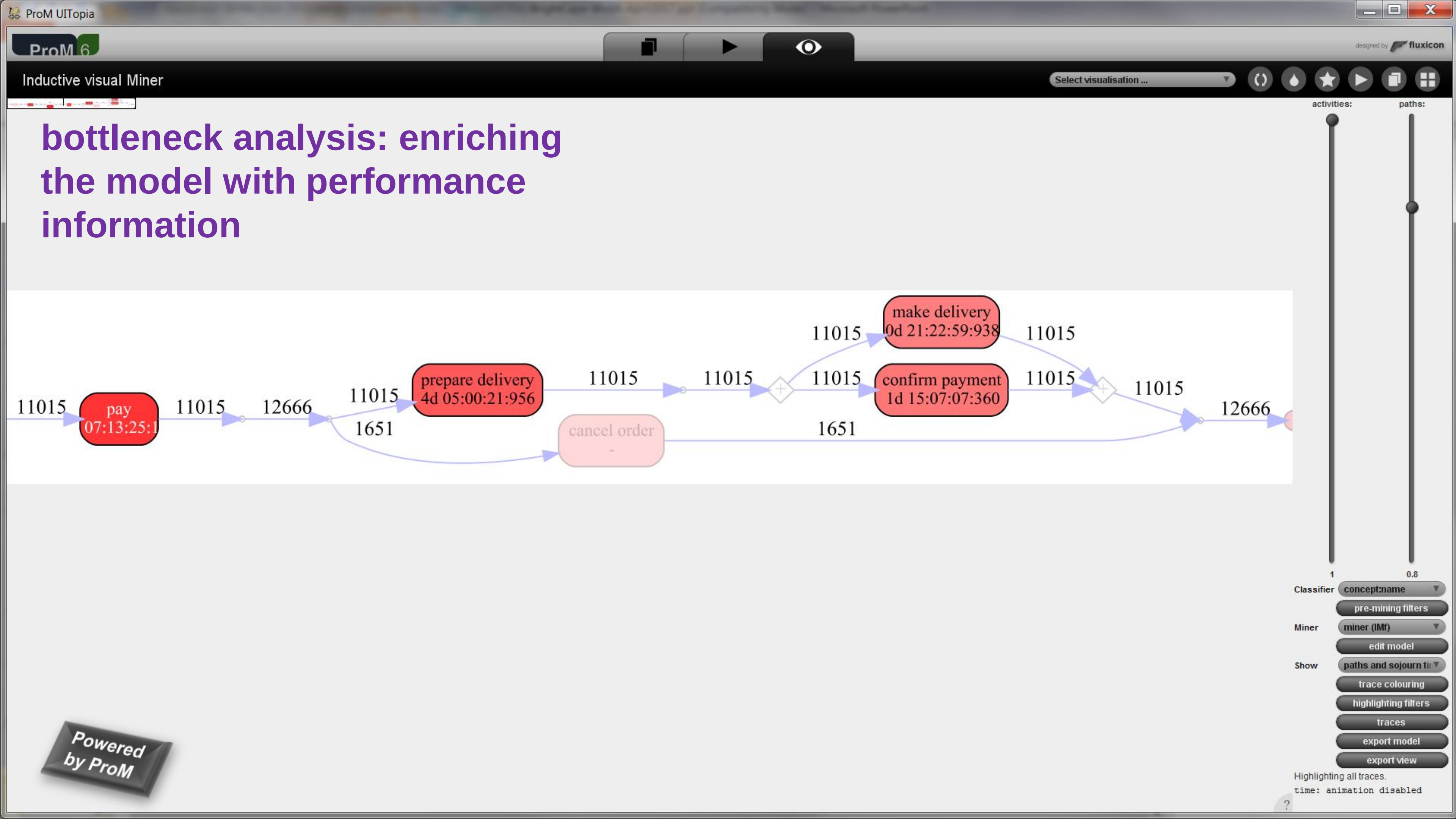
11-12-2020 18:49:05	17-12-2020 12:03:09	24-12-2020 14:43:54	31-12-2020 16:14:16		05-01-2021 09:19:19
place order	send invoice	send reminder	send reminder	pay	cancel order
complete	complete	complete	complete	complete	complete
		only in log		only in model	

Highlighting all traces.

time: animation disabled

done

bottleneck analysis: enriching the model with performance information



ProM UITopia

Inductive visual Miner

Classifier: concept:name

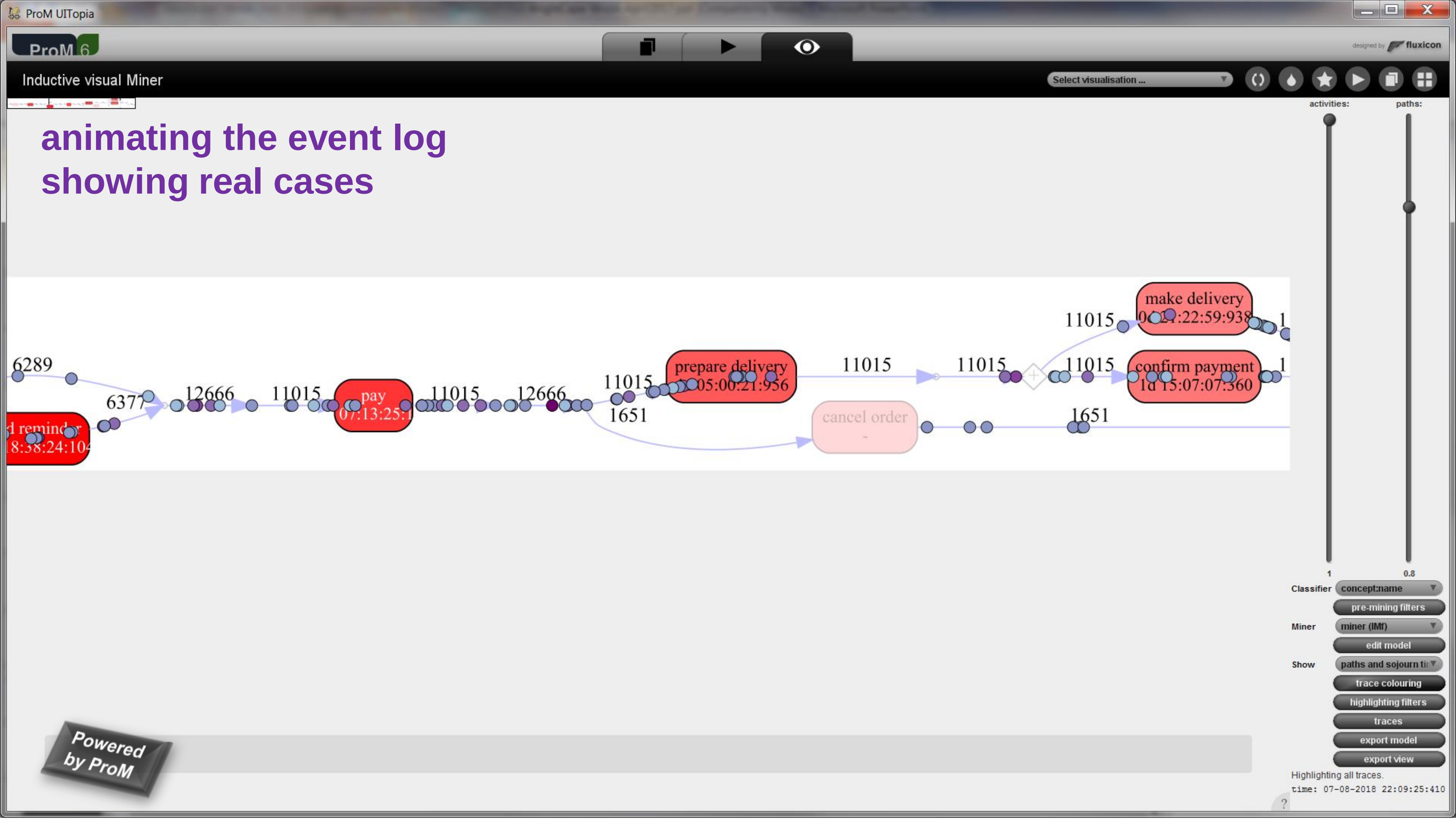
Miner: miner (IMf)

Show: paths and sojourn ti

Highlighting all traces.

```
time: animation disabled
```

animating the event log showing real cases



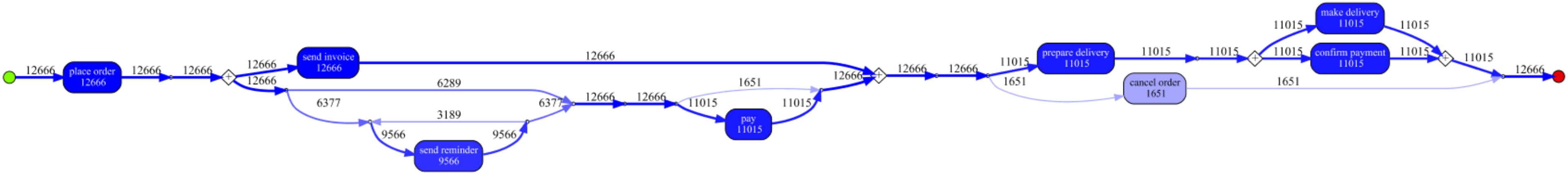
Classifier: concept:name

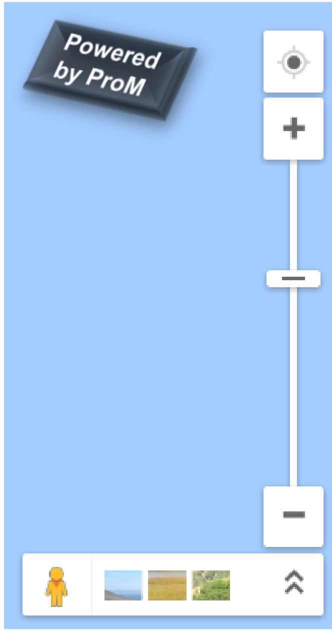
Miner: miner (IMf)

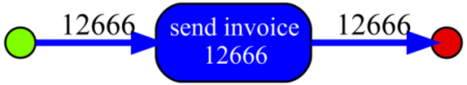
Show: paths and sojourn ti

Highlighting all traces.

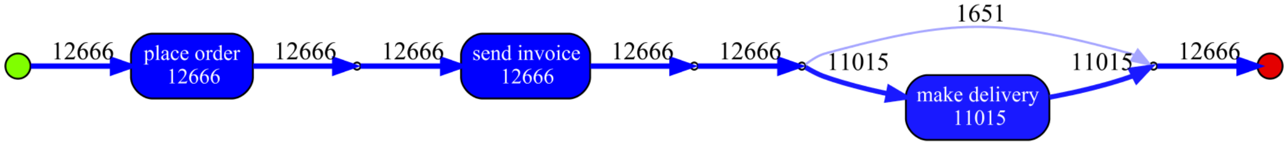
```
time: 07-08-2018 22:09:25:410
```

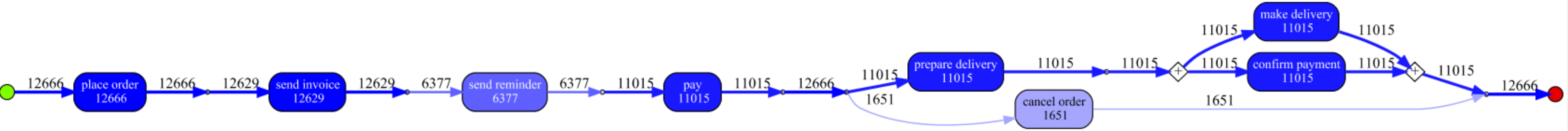




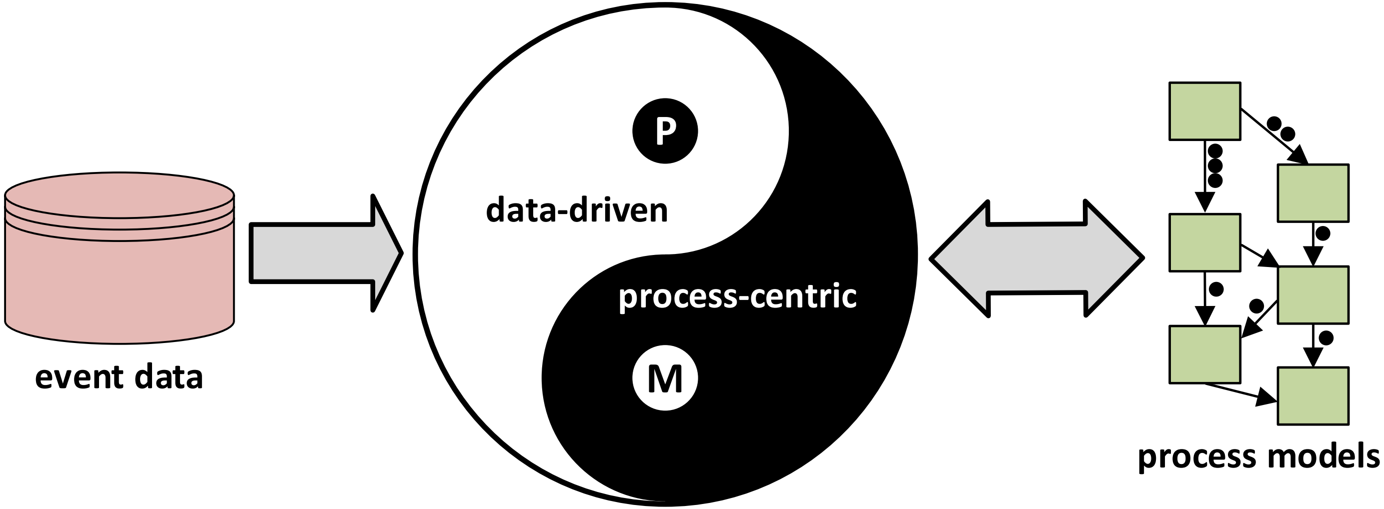


seamless abstraction: one log many views





Data-driven & process-centric

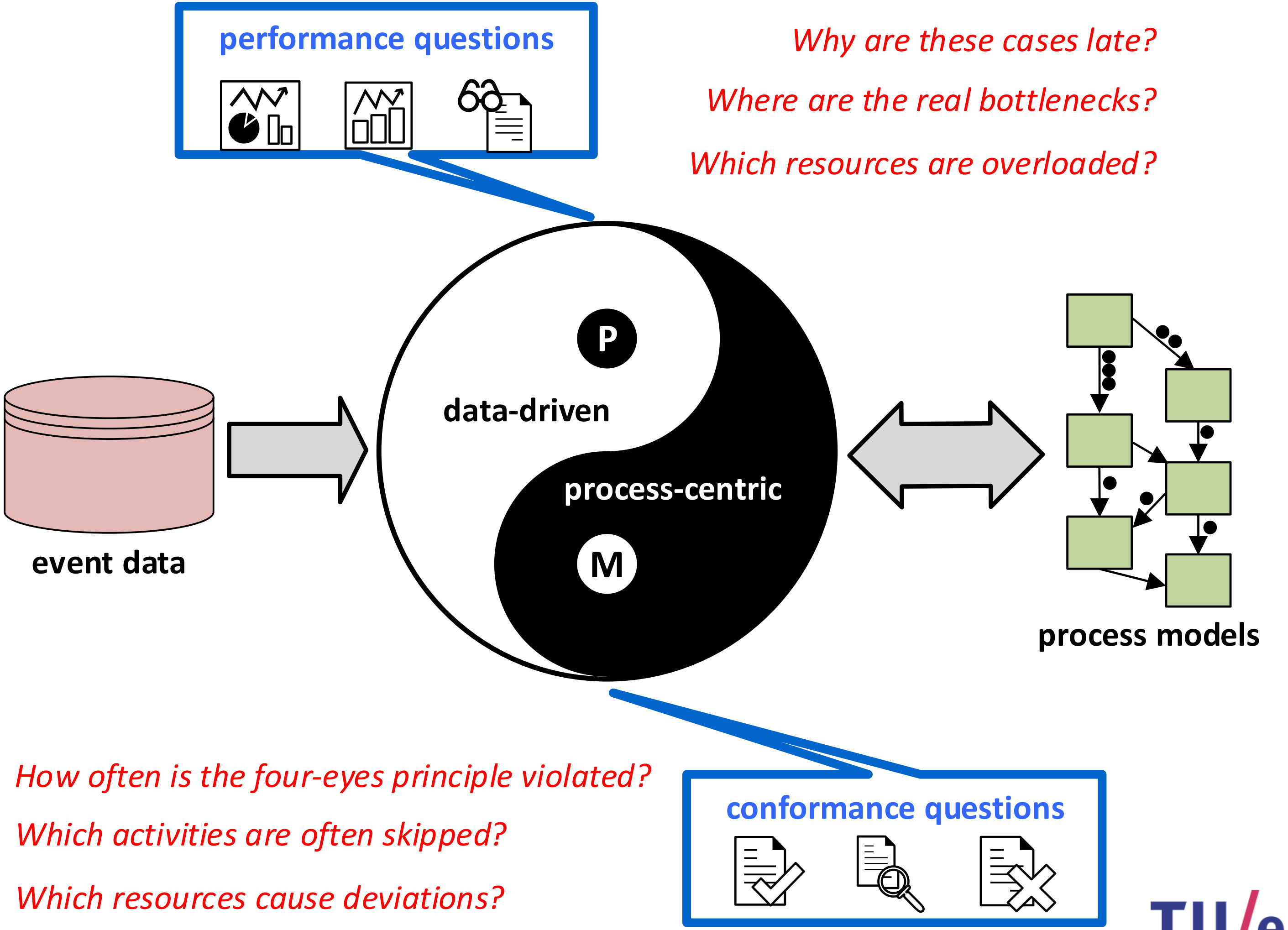


Answering two types of questions

Why are these cases late?

Where are the real bottlenecks?

Which resources are overloaded?



How often is the four-eyes principle violated?

Which activities are often skipped?

Which resources cause deviations?

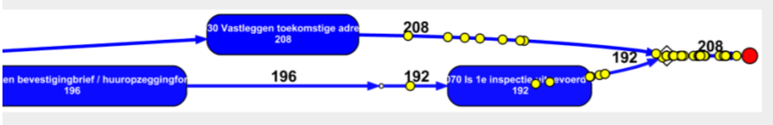
Process mining results may hurt and trigger resistance, but this only supports the need for it.

Process Mining Software

ProM
process mining workbench

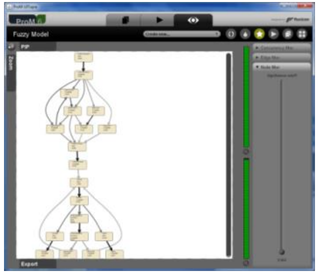
rapid ProM

TU/e


30 Vastleggen toekomstige adre
208
208
192
208
en bevestigingbrief / huuropzeggingfor
196
196
192
070 Is 1e inspectie uitgevoerd
192

1500+ plug-ins available covering the whole process mining spectrum

- Merge two Event Logs using AIS algorithm
- Mine for a Fuzzy Model
- Mine for a Handover-of-Work Social Network
- Mine for a Petri Net using ILP
- Mine for a Reassignment Social Network
- Mine for a Similar-Task Social Network



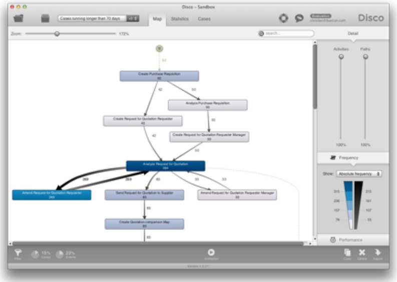
>130k downloads

ProM process mining workbench

Disco

PROCESSGOLD

myinvenio



celonis
process mining

minit

QPR
Quality. Processes. Results.

ICARO
TECH

perceptivesoftware
a Lexmark company

XMPRO
GET BETTER AT GETTING WORK DONE

StereoLogic

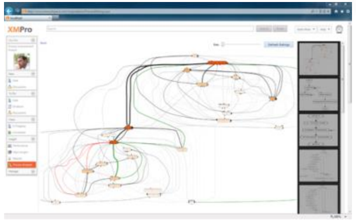
software AG

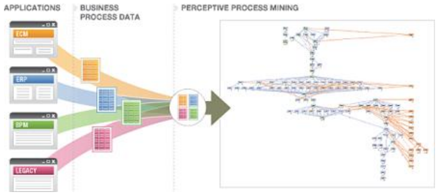
ARIS
Process Performance Manager

Rialto
by exeura

LANA
LABS

FUJITSU

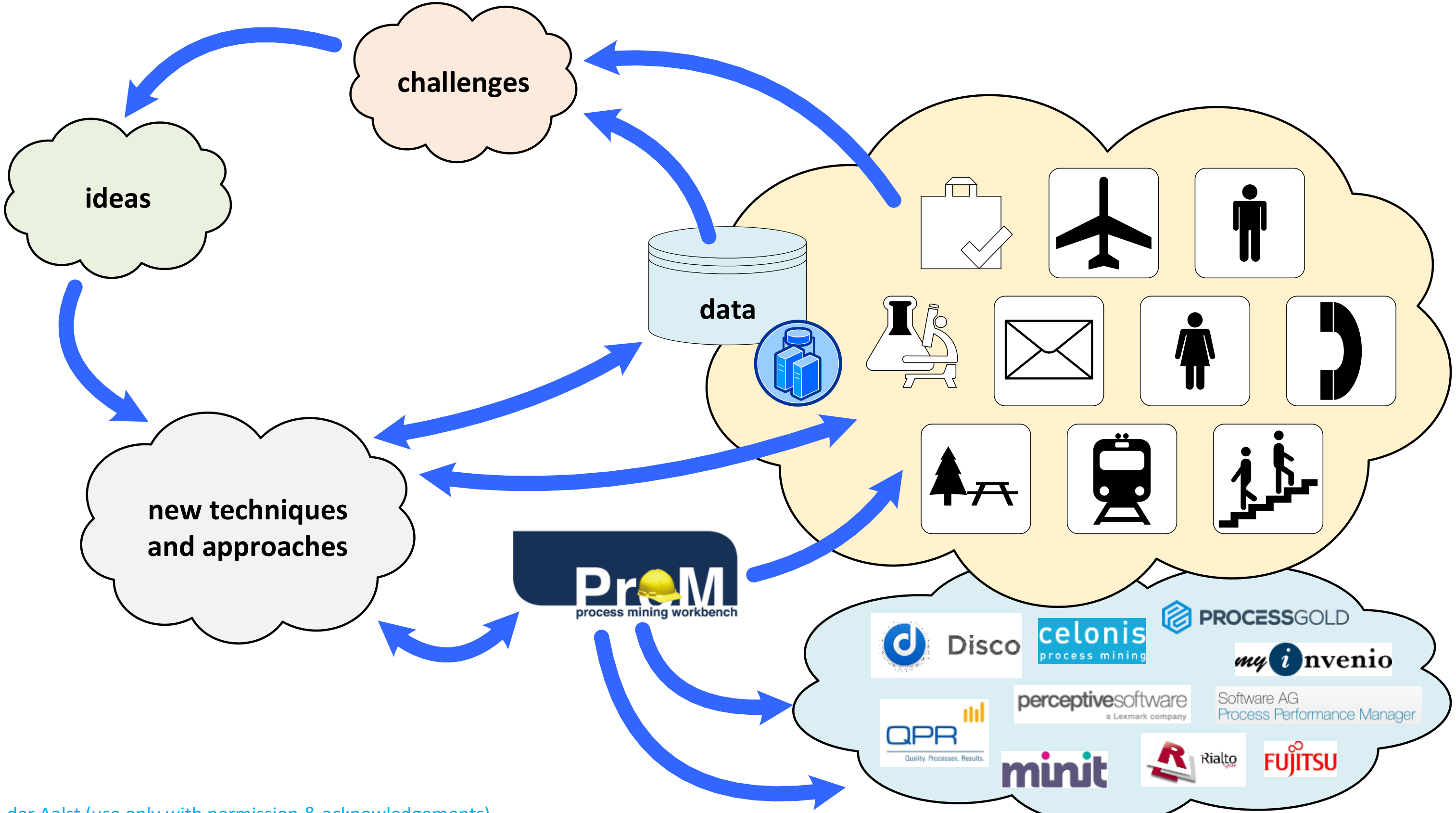




SNP The Transformation Company

TU/e

Interaction with industry



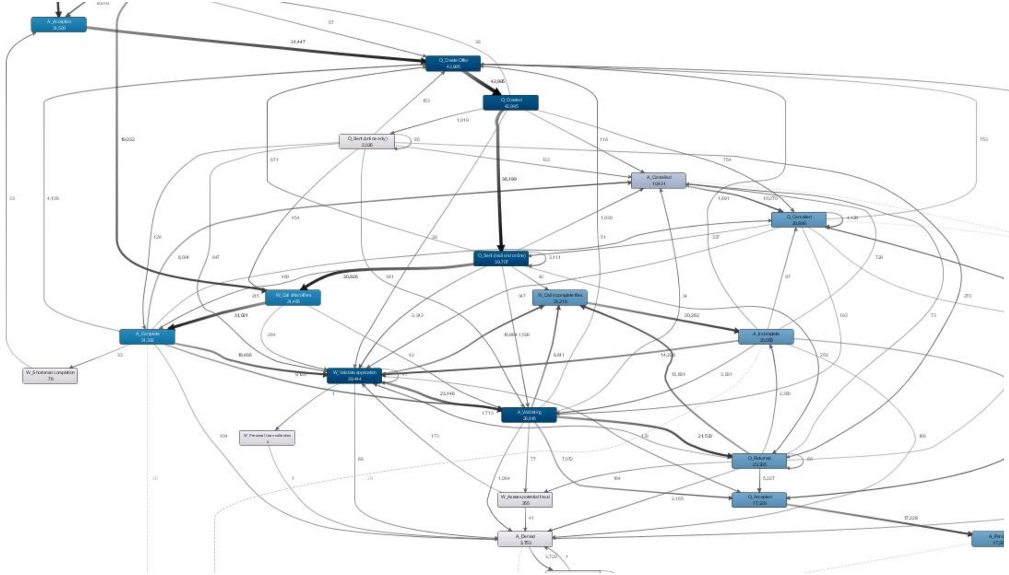
©Wil van der Aalst (use only with permission & acknowledgements)

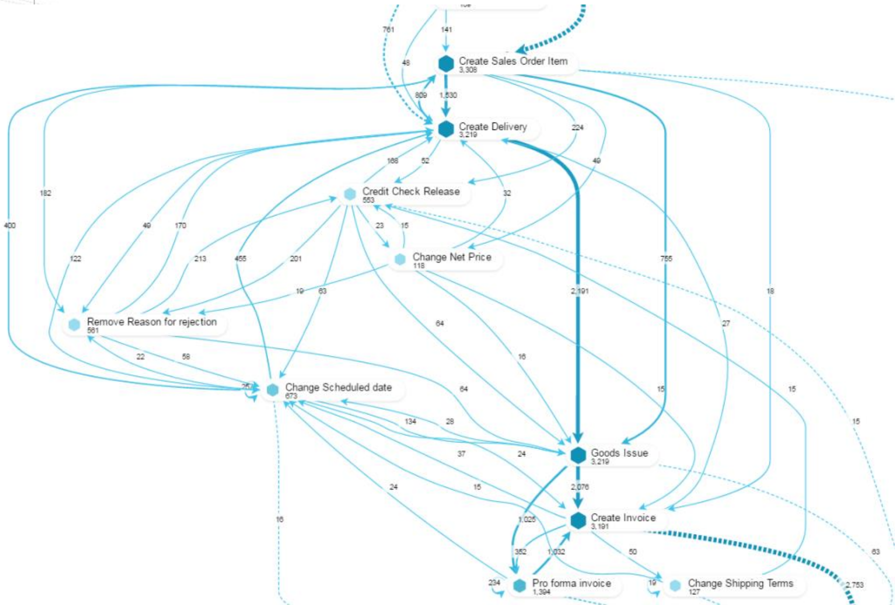
Job done?

Not really …

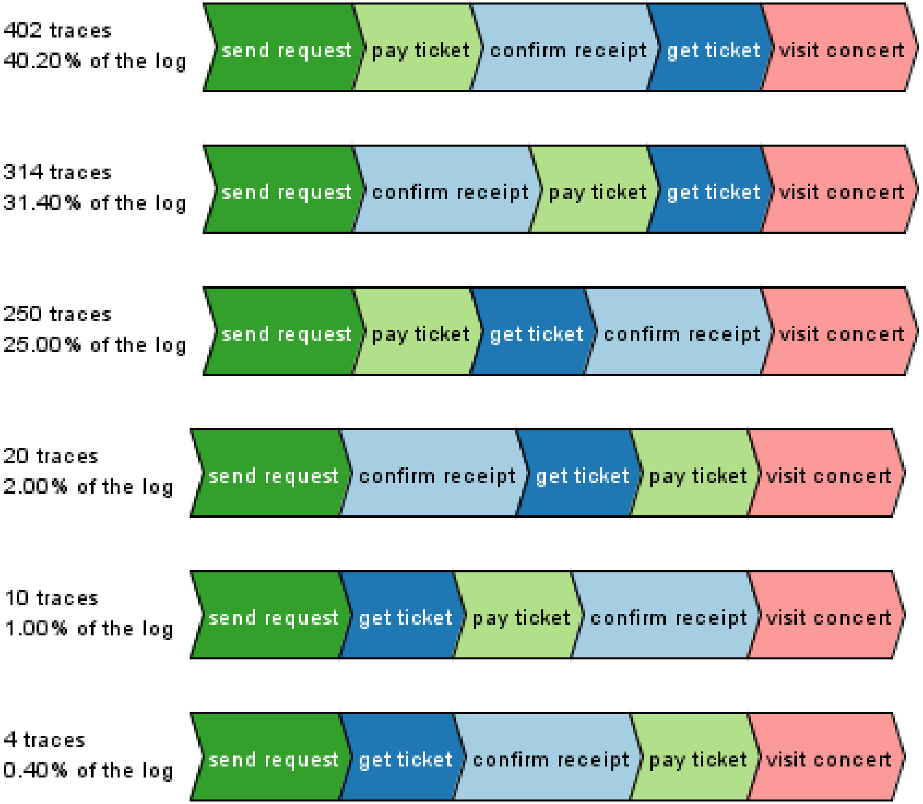
Concurrency and Semantics

Boxes and arrows: What do they mean?





An example log and “proper models”



402 traces
40.20% of the log
send request → pay ticket → confirm receipt → get ticket → visit concert

314 traces
31.40% of the log
send request → confirm receipt → pay ticket → get ticket → visit concert

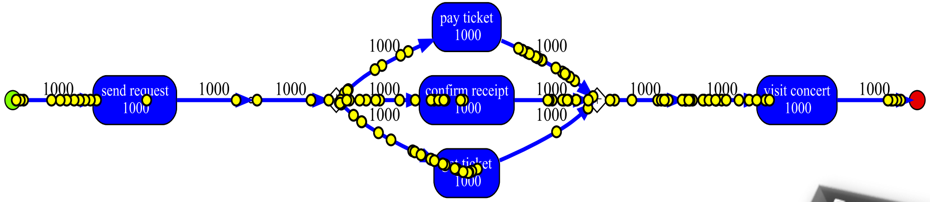
250 traces
25.00% of the log
send request → pay ticket → get ticket → confirm receipt → visit concert

20 traces
2.00% of the log
send request → confirm receipt → get ticket → pay ticket → visit concert

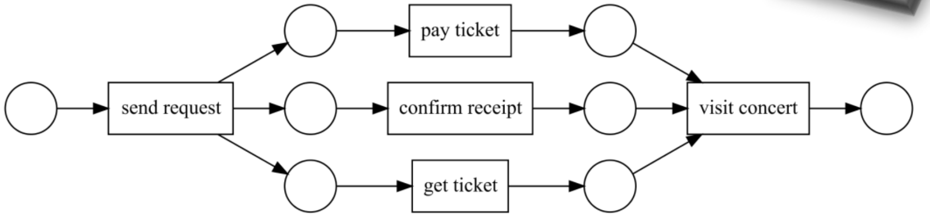
10 traces
1.00% of the log
send request → get ticket → pay ticket → confirm receipt → visit concert

4 traces
0.40% of the log
send request → get ticket → confirm receipt → pay ticket → visit concert

1000 cases and 5000 events

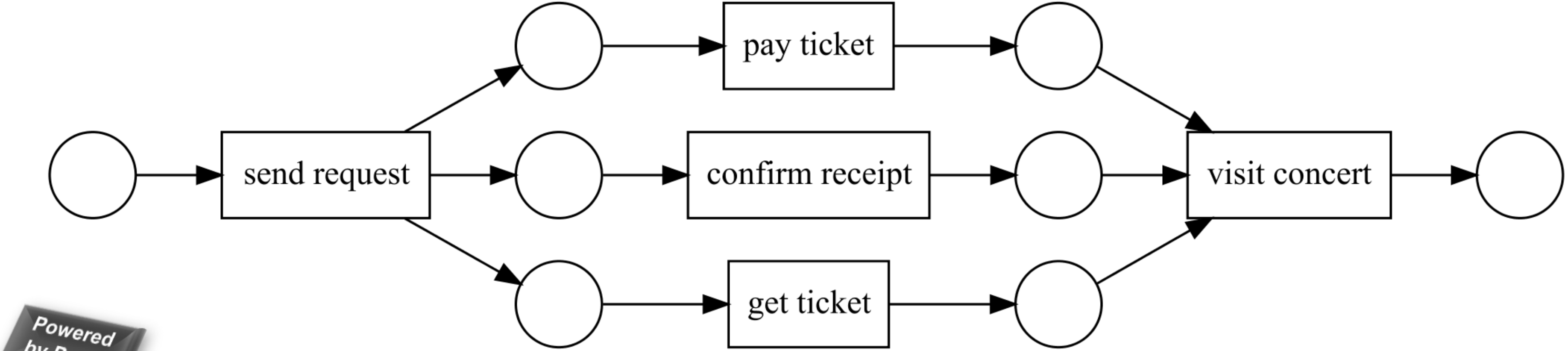


Powered by ProM

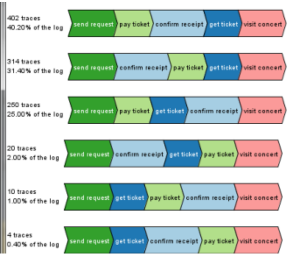


Inductive miner, ILP miner, Alpha miner, etc.

Model with 3 concurrent activities



Powered by ProM



402 traces
40.20% of the log
send request | pay ticket | confirm receipt | get ticket | visit concert

314 traces
31.40% of the log
send request | confirm receipt | pay ticket | get ticket | visit concert

250 traces
25.00% of the log
send request | pay ticket | get ticket | confirm receipt | visit concert

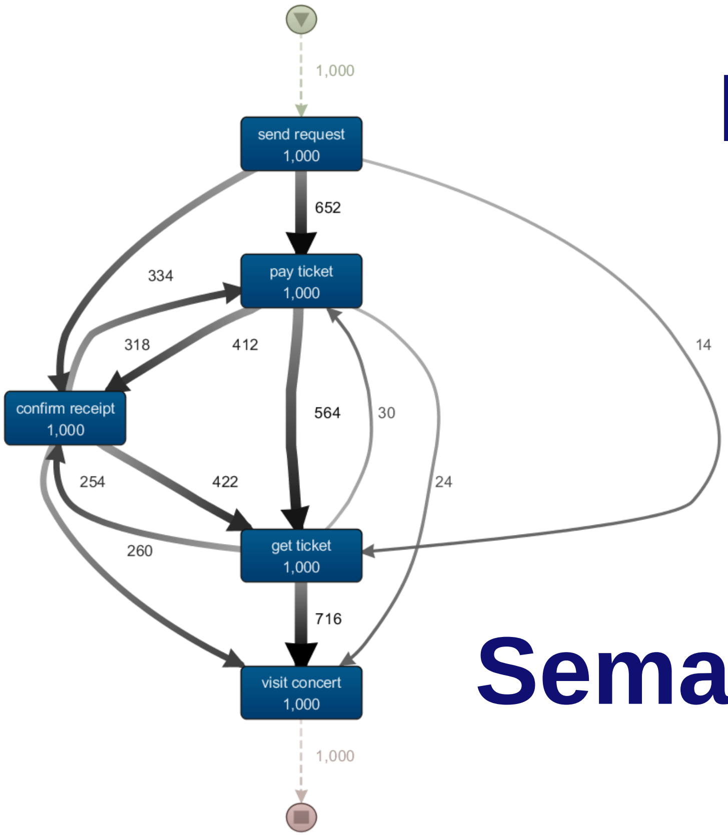
20 traces
2.00% of the log
send request | confirm receipt | get ticket | pay ticket | visit concert

10 traces
1.00% of the log
send request | get ticket | pay ticket | confirm receipt | visit concert

4 traces
0.40% of the log
send request | get ticket | confirm receipt | pay ticket | visit concert

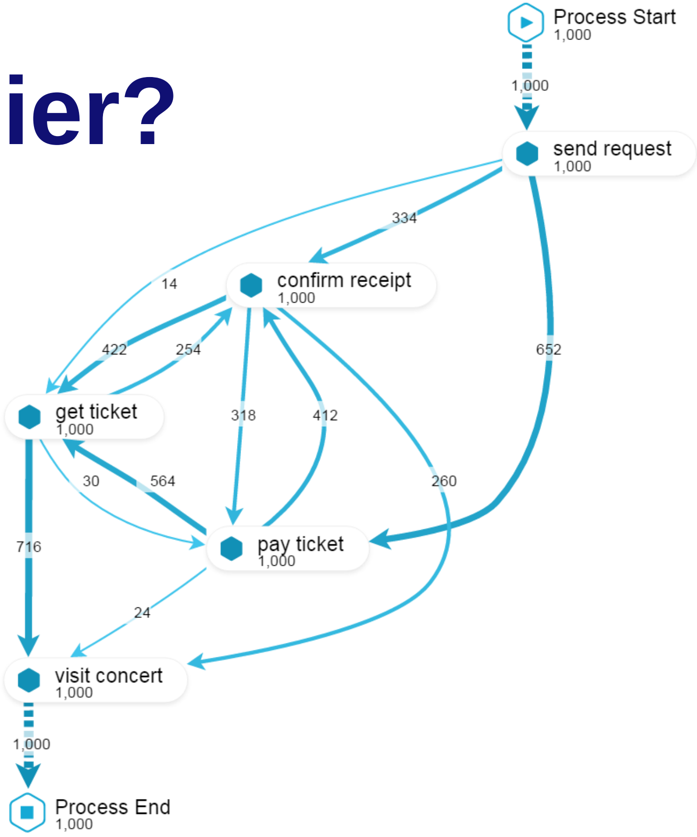
Too difficult ☺?

Concurrency & Semantics

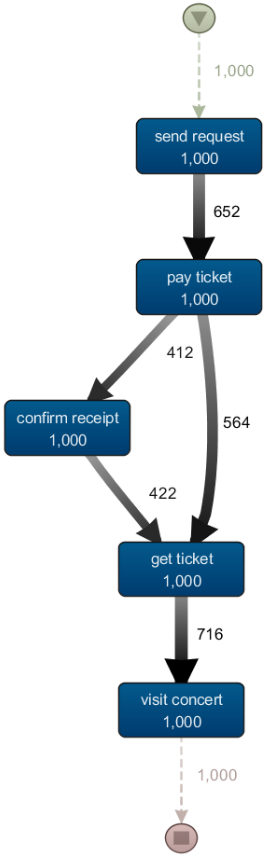


Much easier?

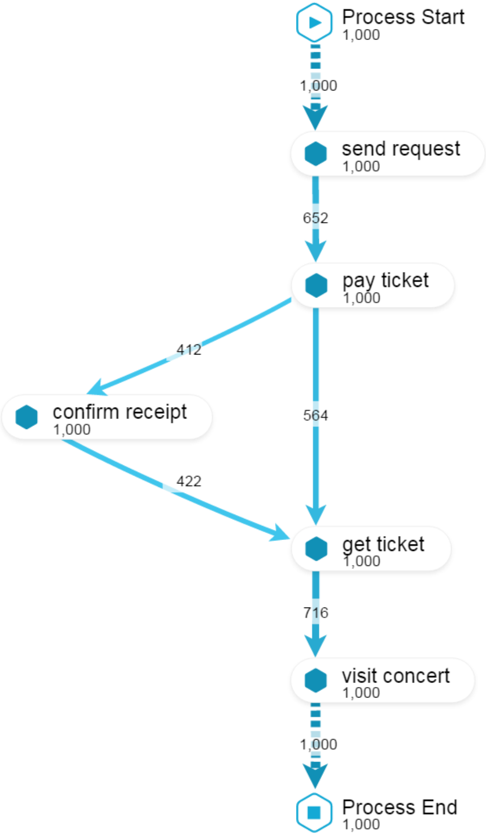
Semantics?



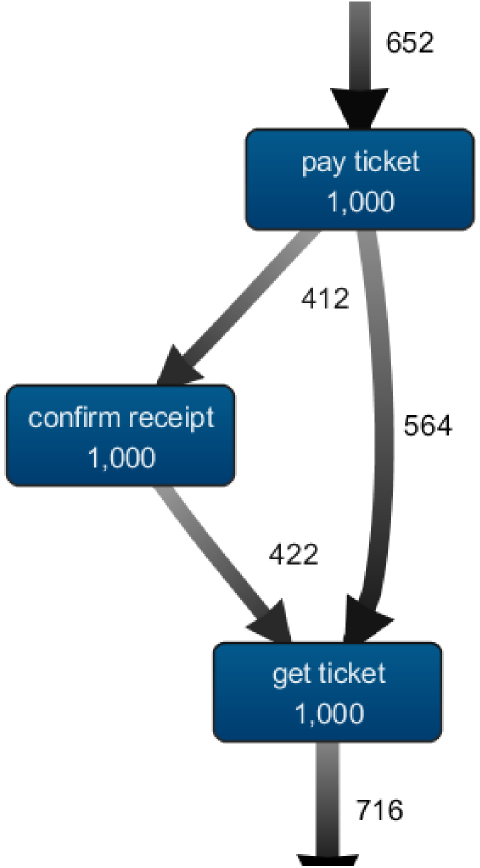
Concurrency & Semantics



Do frequencies make sense?

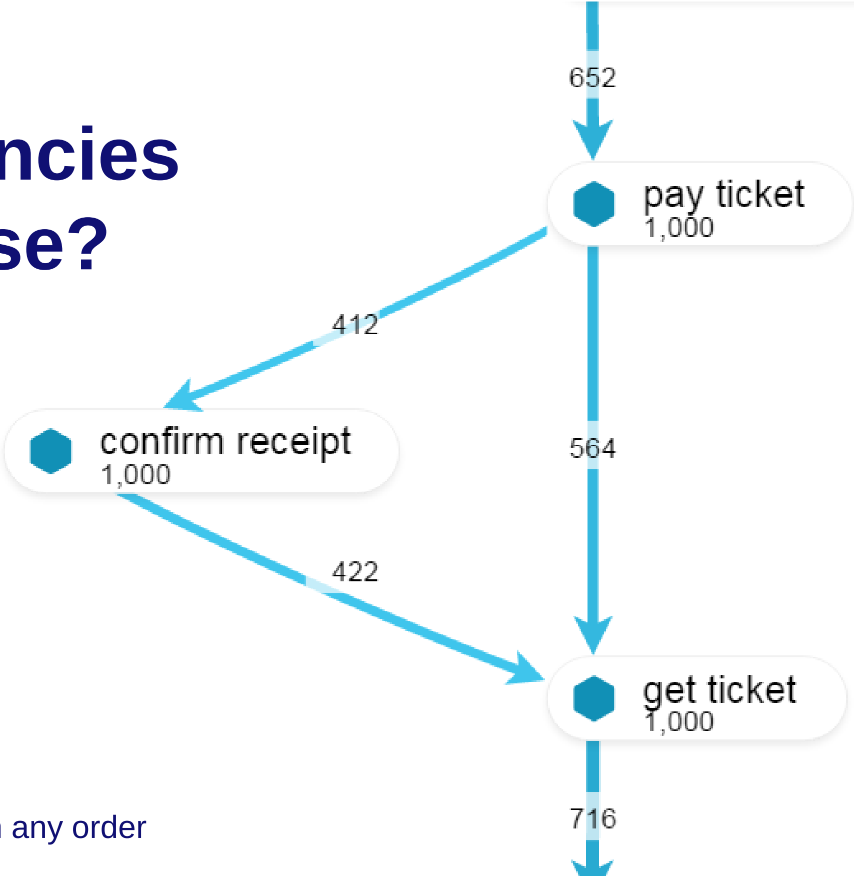


Concurrency & Semantics

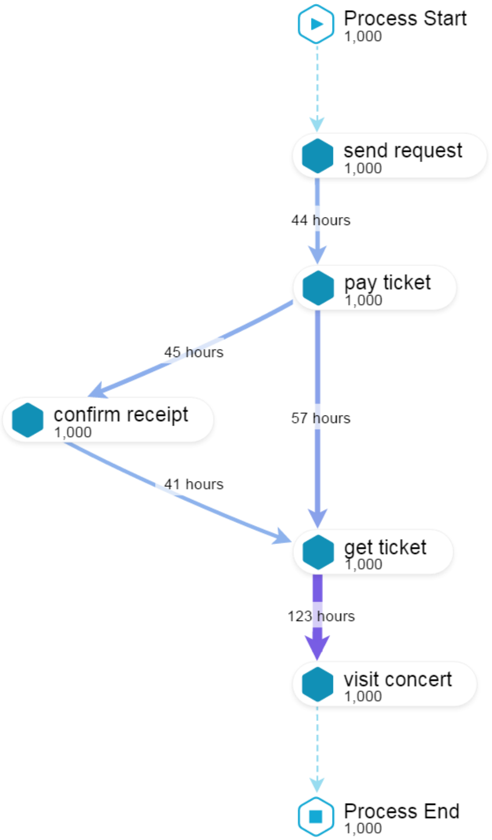


Do frequencies make sense?

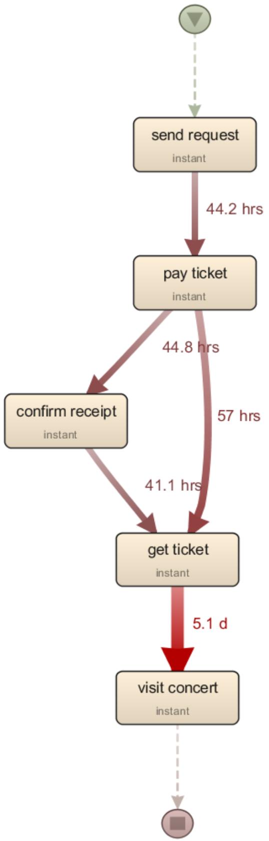
- ➢ do not add up
- ➢ were performed in any order



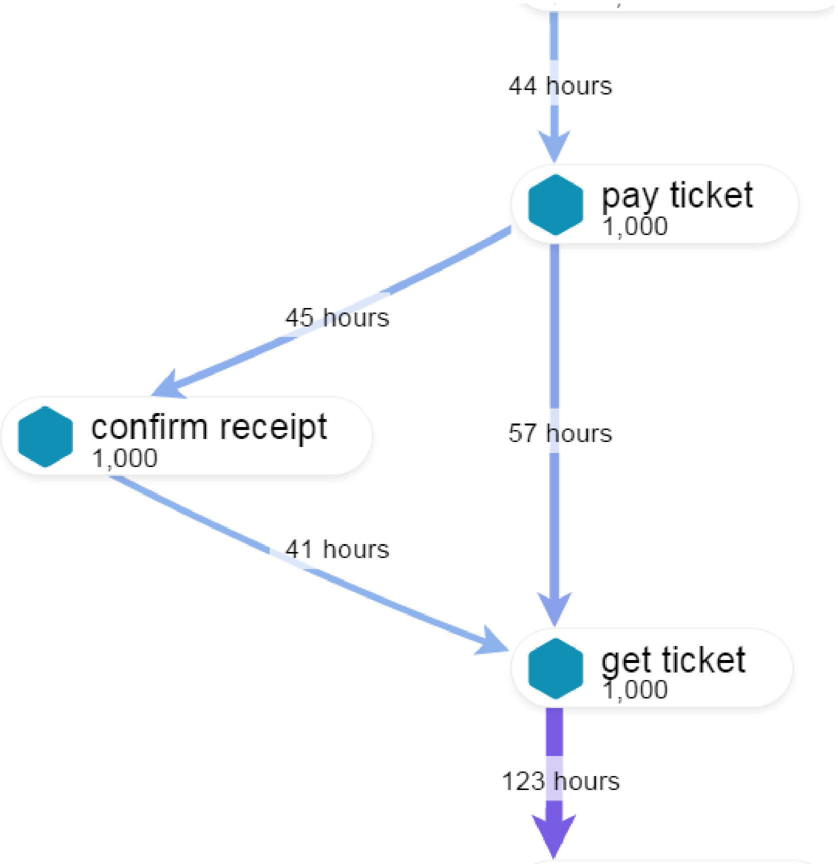
Concurrency & Semantics



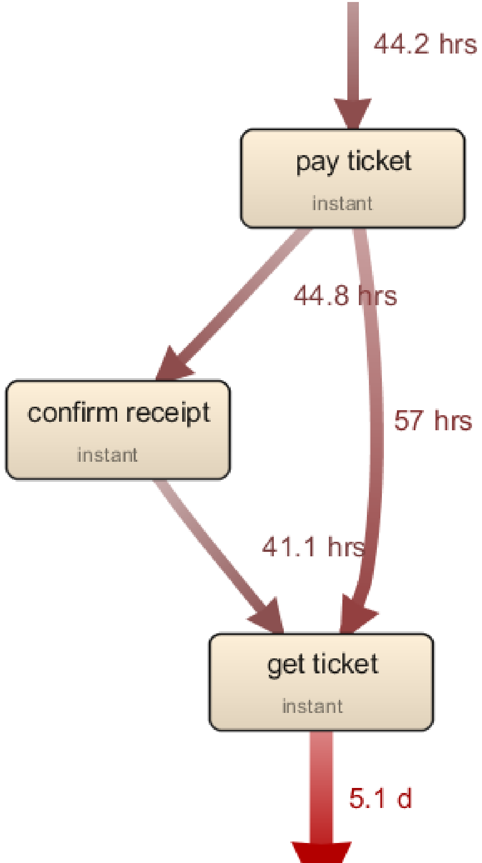
Do times make sense?



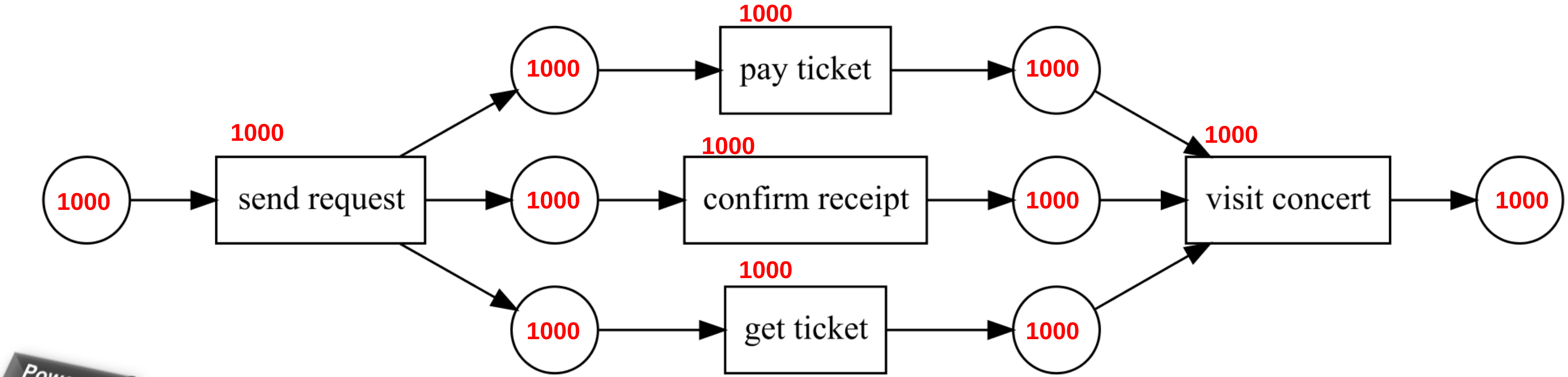
Concurrency & Semantics

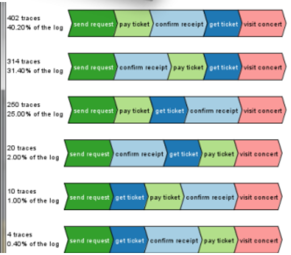


Do times make sense?



Concurrency & Semantics





402 traces
40.20% of the log
send request | pay ticket | confirm receipt | get ticket | visit concert

314 traces
31.40% of the log
send request | confirm receipt | pay ticket | get ticket | visit concert

250 traces
25.00% of the log
send request | pay ticket | get ticket | confirm receipt | visit concert

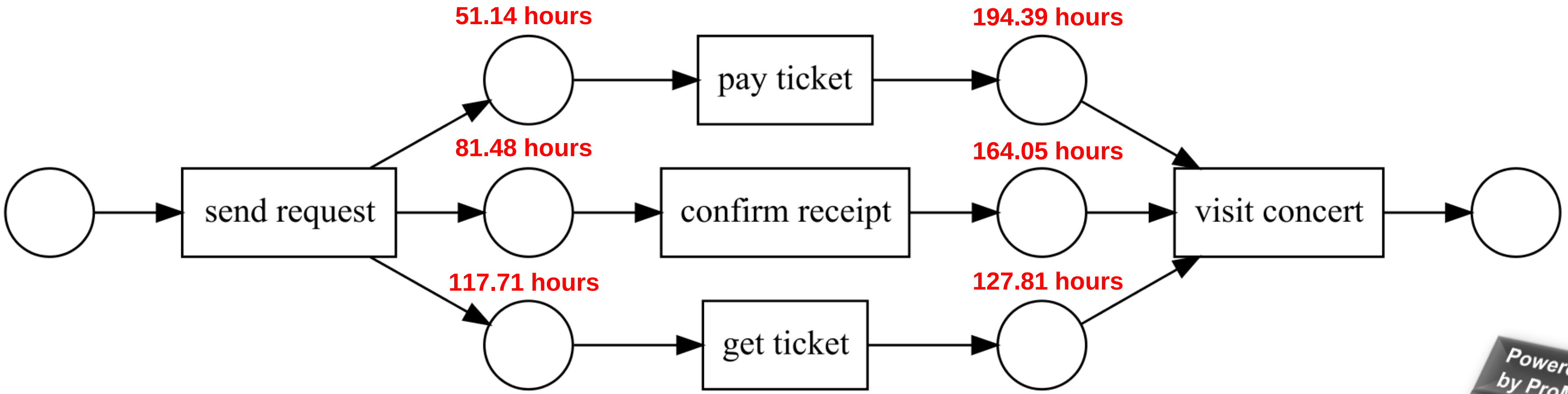
20 traces
2.00% of the log
send request | confirm receipt | get ticket | pay ticket | visit concert

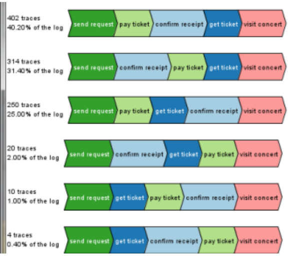
10 traces
1.00% of the log
send request | get ticket | pay ticket | confirm receipt | visit concert

4 traces
0.40% of the log
send request | get ticket | confirm receipt | pay ticket | visit concert

Correct numbers

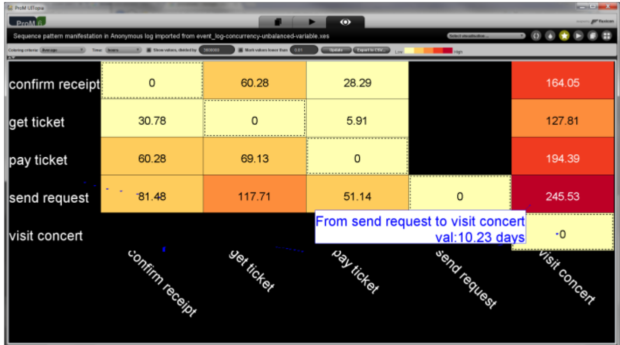
Concurrency & Semantics





402 traces 40.20% of the log	send request	pay ticket	confirm receipt	get ticket	visit concert
314 traces 31.40% of the log	send request	confirm receipt	pay ticket	get ticket	visit concert
250 traces 25.00% of the log	send request	pay ticket	get ticket	confirm receipt	visit concert
20 traces 2.00% of the log	send request	confirm receipt	get ticket	pay ticket	visit concert
10 traces 1.00% of the log	send request	get ticket	pay ticket	confirm receipt	visit concert
4 traces 0.40% of the log	send request	get ticket	confirm receipt	pay ticket	visit concert

Correct times



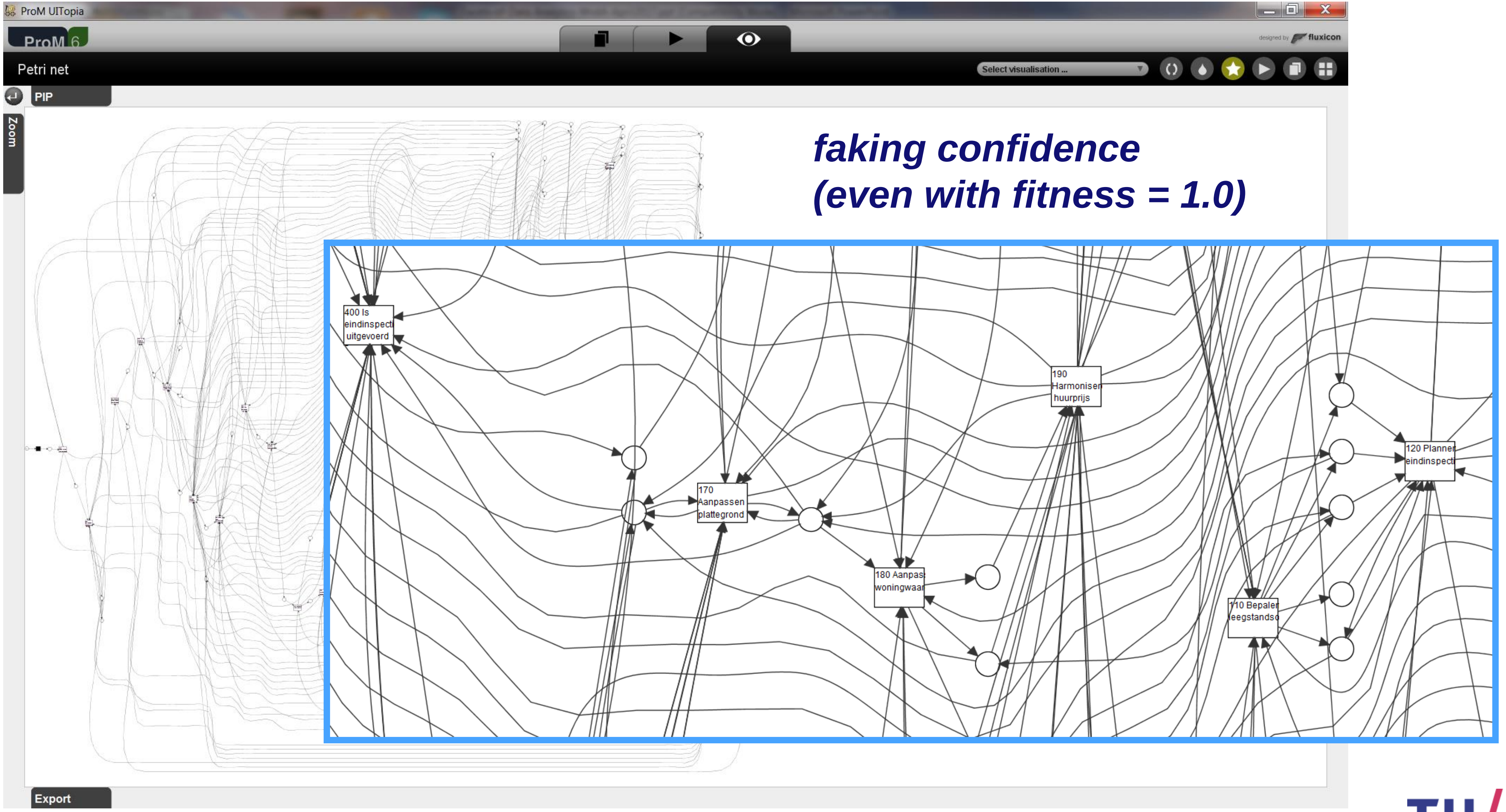
Sequence pattern manifestation in Anonymous log imported from event_log-concurrency-unbalanced-variable.xes

	confirm receipt	get ticket	pay ticket	send request	visit concert
confirm receipt	0	60.28	28.29		164.05
get ticket	30.78	0	5.91		127.81
pay ticket	60.28	69.13	0		194.39
send request	81.48	117.71	51.14	0	245.53
visit concert					0

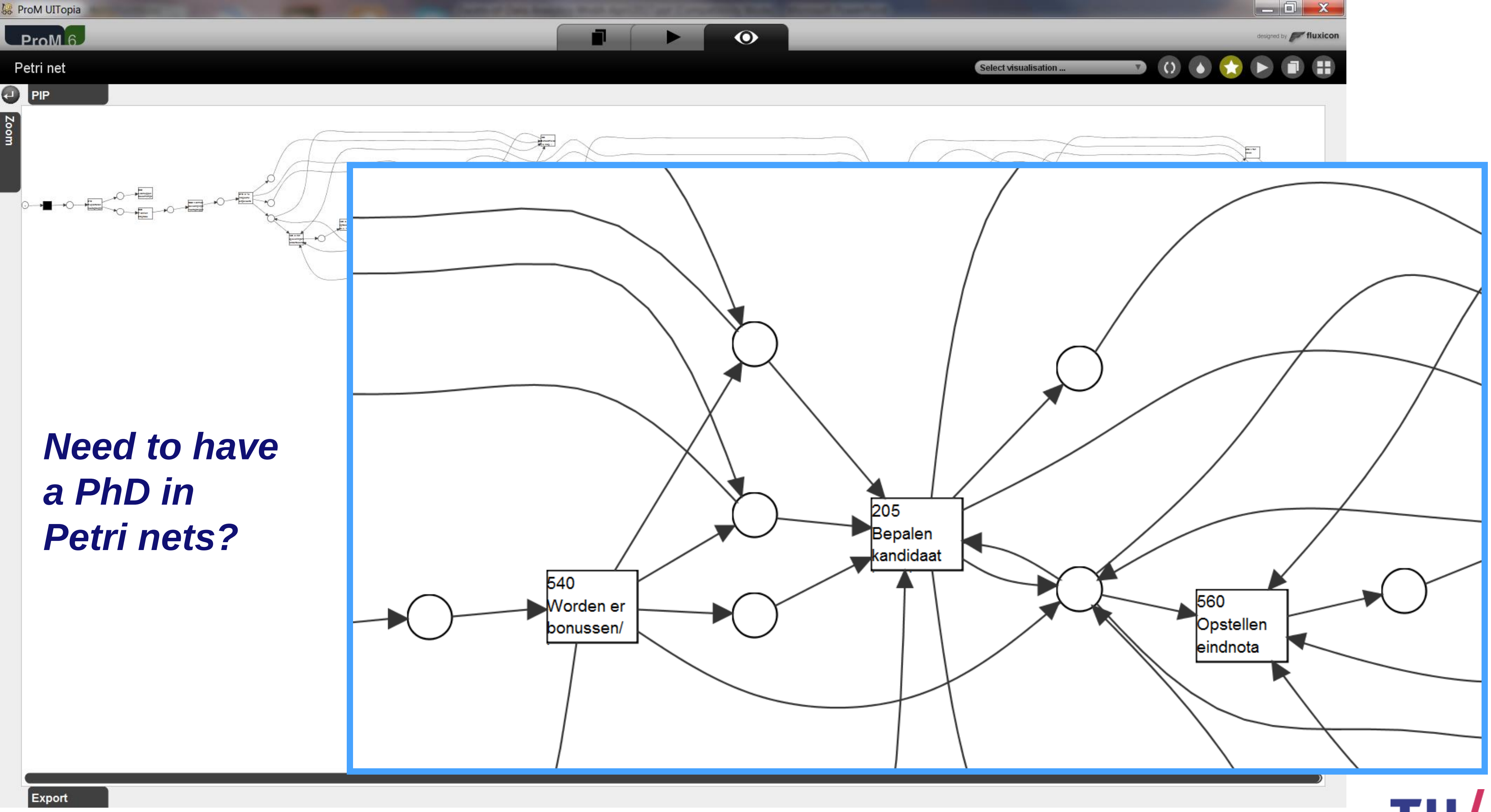
From send request to visit concert
val:10.23 days

©Wil van der Aalst (use only with permission & acknowledgements)

But …



faking confidence
(even with fitness = 1.0)

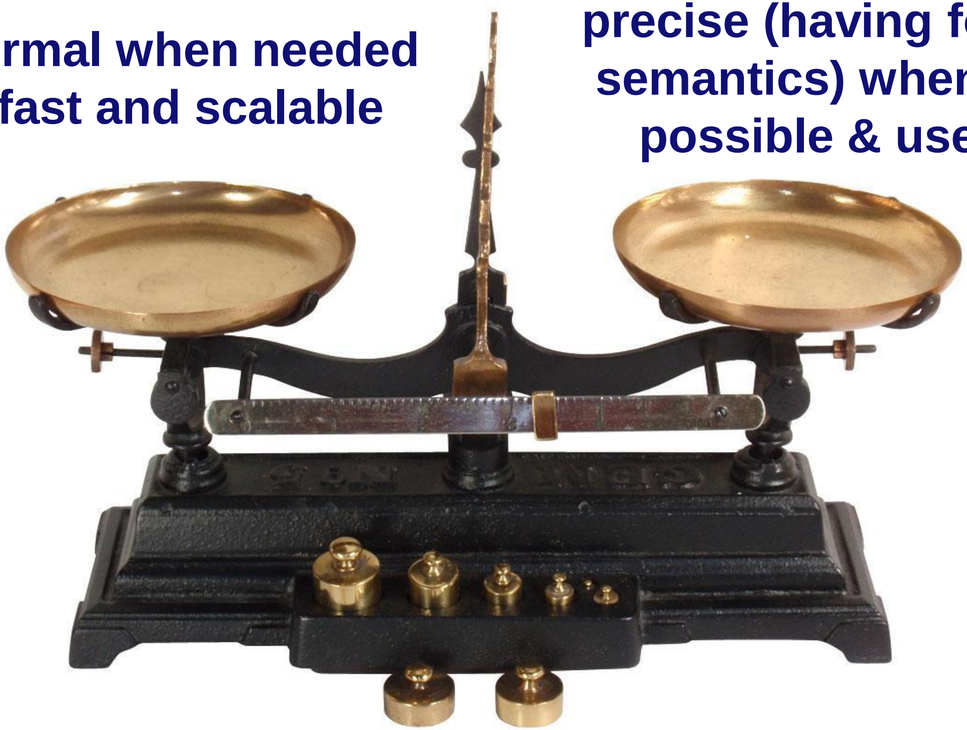


Need to have a PhD in Petri nets?

How to combine the best of both worlds?

informal when needed & fast and scalable

precise (having formal semantics) whenever possible & useful

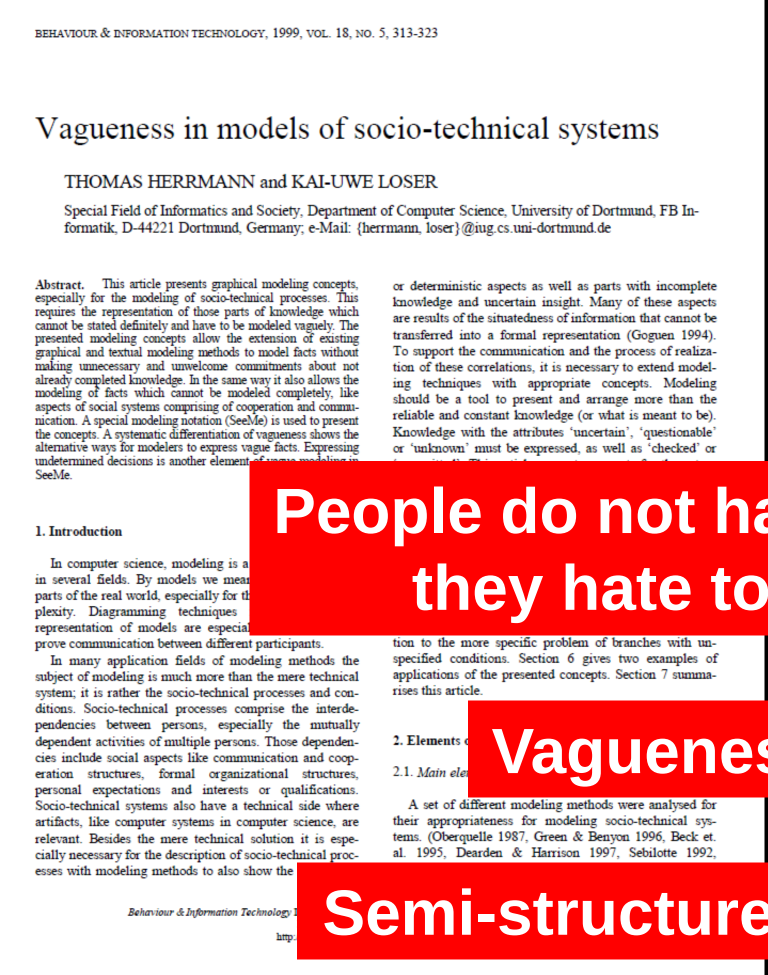


- commercial tools
- heuristic miner
- fuzzy miner
- etc.

- basic inductive miner
- ILP miner
- other region-based approaches
- etc.

Idea: Hybrid process models

Joint work with Riccardo De Masellis, Chiara Di Francescomarino, Chiara Ghidini



BEHAVIOUR & INFORMATION TECHNOLOGY, 1999, VOL. 18, NO. 5, 313-323

Vagueness in models of socio-technical systems

THOMAS HERRMANN and KAI-UWE LOSER

Special Field of Informatics and Society, Department of Computer Science, University of Dortmund, FB Informatik, D-44221 Dortmund, Germany; e-Mail: {herrmann, loser}@iug.cs.uni-dortmund.de

Abstract. This article presents graphical modeling concepts, especially for the modeling of socio-technical processes. This requires the representation of those parts of knowledge which cannot be stated definitely and have to be modeled vaguely. The presented modeling concepts allow the extension of existing graphical and textual modeling methods to model facts without making unnecessary and unwelcome commitments about not already completed knowledge. In the same way it also allows the modeling of facts which cannot be modeled completely, like aspects of social systems comprising of cooperation and communication. A special modeling notation (SeeMe) is used to present the concepts. A systematic differentiation of vagueness shows the alternative ways for modelers to express vague facts. Expressing undetermined decisions is another element SeeMe.

1. Introduction

In computer science, modeling is a in several fields. By models we mea parts of the real world, especially for th plexity. Diagramming techniques representation of models are especia prove communication between different participants.

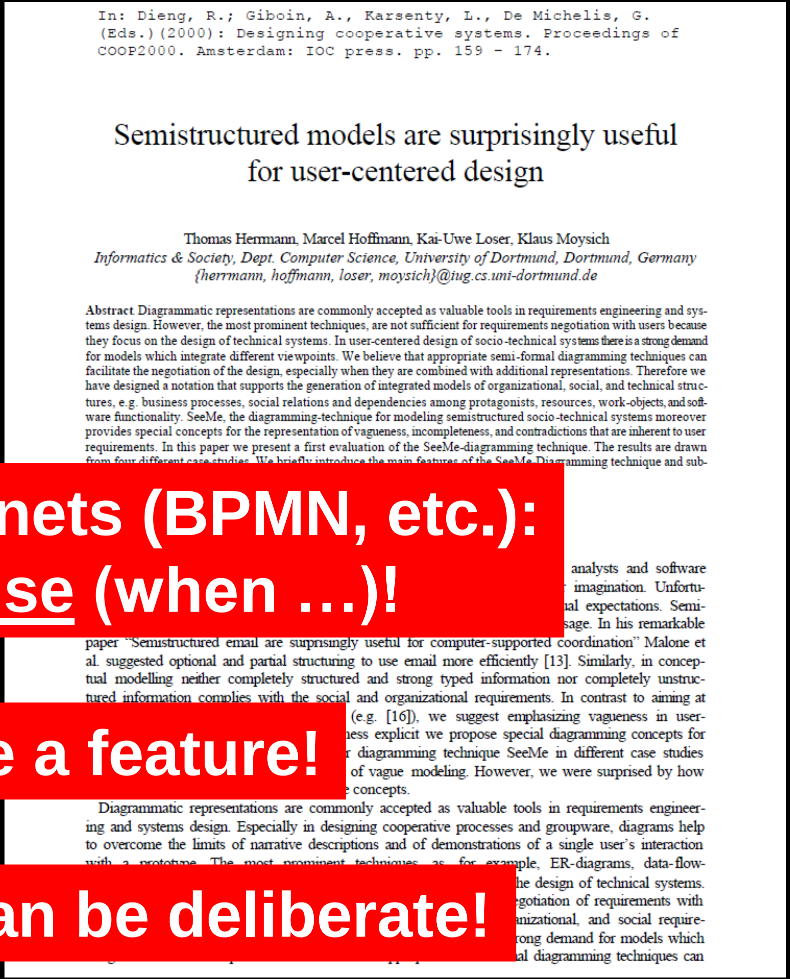
In many application fields of modeling methods the subject of modeling is much more than the mere technical system; it is rather the socio-technical processes and conditions. Socio-technical processes comprise the interdependencies between persons, especially the mutually dependent activities of multiple persons. Those dependencies include social aspects like communication and cooperation structures, formal organizational structures, personal expectations and interests or qualifications. Socio-technical systems also have a technical side where artifacts, like computer systems in computer science, are relevant. Besides the mere technical solution it is especially necessary for the description of socio-technical processes with modeling methods to also show the or deterministic aspects as well as parts with incomplete knowledge and uncertain insight. Many of these aspects are results of the situatedness of information that cannot be transferred into a formal representation (Goguen 1994). To support the communication and the process of realization of these correlations, it is necessary to extend modeling techniques with appropriate concepts. Modeling should be a tool to present and arrange more than the reliable and constant knowledge (or what is meant to be). Knowledge with the attributes 'uncertain', 'questionable' or 'unknown' must be expressed, as well as 'checked' or

tion to the more specific problem of branches with unspecified conditions. Section 6 gives two examples of applications of the presented concepts. Section 7 summarises this article.

2. Elements

2.1. *Main ele*

A set of different modeling methods were analysed for their appropriateness for modeling socio-technical systems. (Oberquelle 1987, Green & Benyon 1996, Beck et. al. 1995, Dearden & Harrison 1997, Sebilotte 1992,



In: Dieng, R.; Giboin, A., Karsenty, L., De Michelis, G. (Eds.)(2000): Designing cooperative systems. Proceedings of COOP2000. Amsterdam: IOC press. pp. 159 - 174.

Semistructured models are surprisingly useful for user-centered design

Thomas Herrmann, Marcel Hoffmann, Kai-Uwe Loser, Klaus Moysich
Informatics & Society, Dept. Computer Science, University of Dortmund, Dortmund, Germany
{herrmann, hoffmann, loser, moysich}@iug.cs.uni-dortmund.de

Abstract. Diagrammatic representations are commonly accepted as valuable tools in requirements engineering and systems design. However, the most prominent techniques, are not sufficient for requirements negotiation with users because they focus on the design of technical systems. In user-centered design of socio-technical systems there is a strong demand for models which integrate different viewpoints. We believe that appropriate semi-formal diagramming techniques can facilitate the negotiation of the design, especially when they are combined with additional representations. Therefore we have designed a notation that supports the generation of integrated models of organizational, social, and technical structures, e.g. business processes, social relations and dependencies among protagonists, resources, work-objects, and software functionality. SeeMe, the diagramming-technique for modeling semistructured socio-technical systems moreover provides special concepts for the representation of vagueness, incompleteness, and contradictions that are inherent to user requirements. In this paper we present a first evaluation of the SeeMe-diagramming technique. The results are drawn from four different case-studies. We briefly introduce the main features of the SeeMe-Diagramming technique and sub-

analysts and software imagination. Unfortu- al expectations. Semi- sage. In his remarkable paper "Semistructured email are surprisingly useful for computer-supported coordination" Malone et al. suggested optional and partial structuring to use email more efficiently [13]. Similarly, in conceptual modelling neither completely structured and strong typed information nor completely unstructured information complies with the social and organizational requirements. In contrast to aiming at (e.g. [16]), we suggest emphasizing vagueness in user- ess explicit we propose special diagramming concepts for r diagramming technique SeeMe in different case studies of vague modeling. However, we were surprised by how concepts.

Diagrammatic representations are commonly accepted as valuable tools in requirements engineering and systems design. Especially in designing cooperative processes and groupware, diagrams help to overcome the limits of narrative descriptions and of demonstrations of a single user's interaction with a prototype. The most prominent techniques, as, for example, ER-diagrams, data-flow- he design of technical systems. egotiation of requirements with nizational, and social require- rong demand for models which al diagramming techniques can

People do not hate Petri nets (BPMN, etc.): they hate to be precise (when …)!

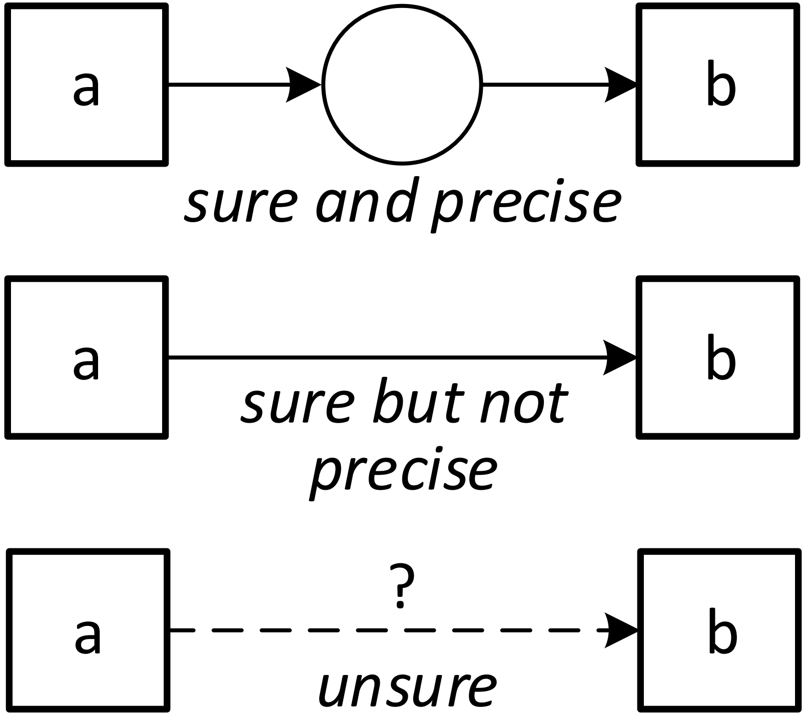
Vagueness can be a feature!

Semi-structuredness can be deliberate!

Wil van der Aalst, Riccardo De Masellis, Chiara Di Francescomarino, Chiara Ghidini: Learning Hybrid Process Models From Events: Process Discovery Without Faking Confidence. BPM 2017 https://doi.org/10.1007/978-3-319-65000-5_4

July 2016

Hybrid Petri nets have three types of arcs



sure and precise

sure but not precise

unsure

strong causality

logic can be captured*

determined based on thresholds

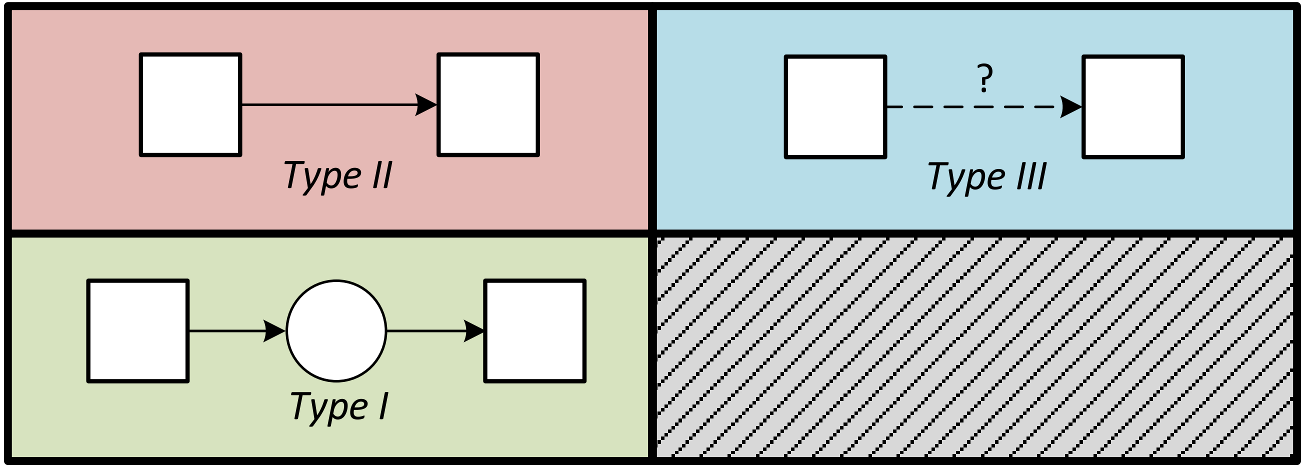
logic cannot be captured*

determined based on thresholds

weak causality

* = easily / of a certain quality / within the representational bias

Hybrid Petri nets have three types of arcs



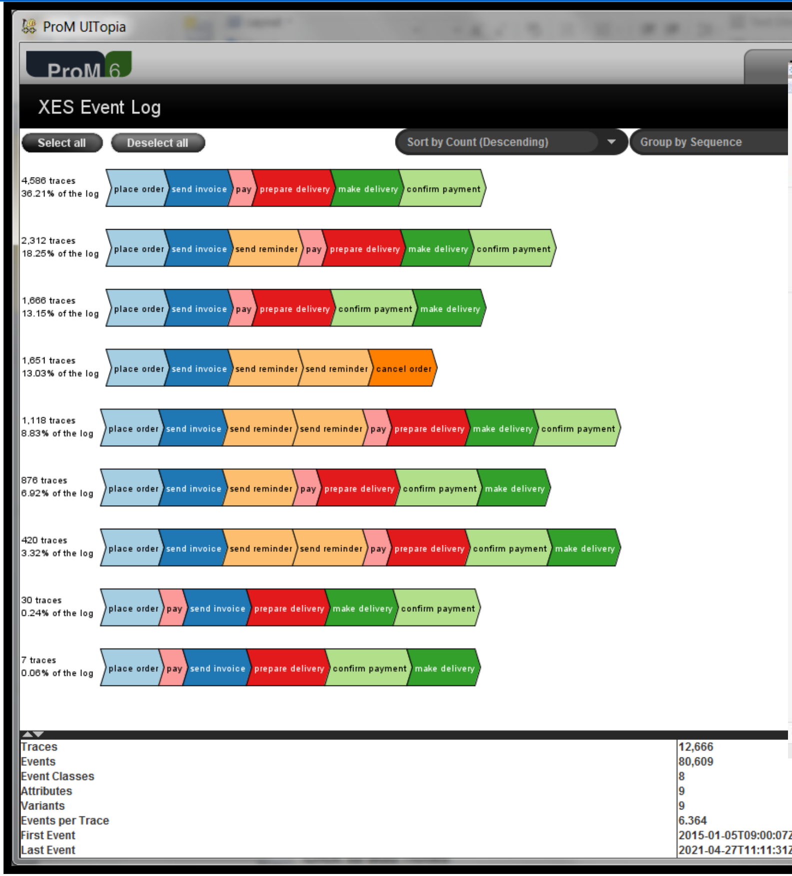
informal
(annotations that are deliberately vague)

formal
(firm statements about the inclusion or exclusion of traces)

strong causality
("sure")

weak causality
("unsure")

Phase 0: Get data



4,586 traces
36.21% of the log
place order → send invoice → pay → prepare delivery → make delivery → confirm payment

2,312 traces
18.25% of the log
place order → send invoice → send reminder → pay → prepare delivery → make delivery → confirm payment

1,666 traces
13.15% of the log
place order → send invoice → pay → prepare delivery → confirm payment → make delivery

1,651 traces
13.03% of the log
place order → send invoice → send reminder → send reminder → cancel order

1,118 traces
8.83% of the log
place order → send invoice → send reminder → send reminder → pay → prepare delivery → make delivery → confirm payment

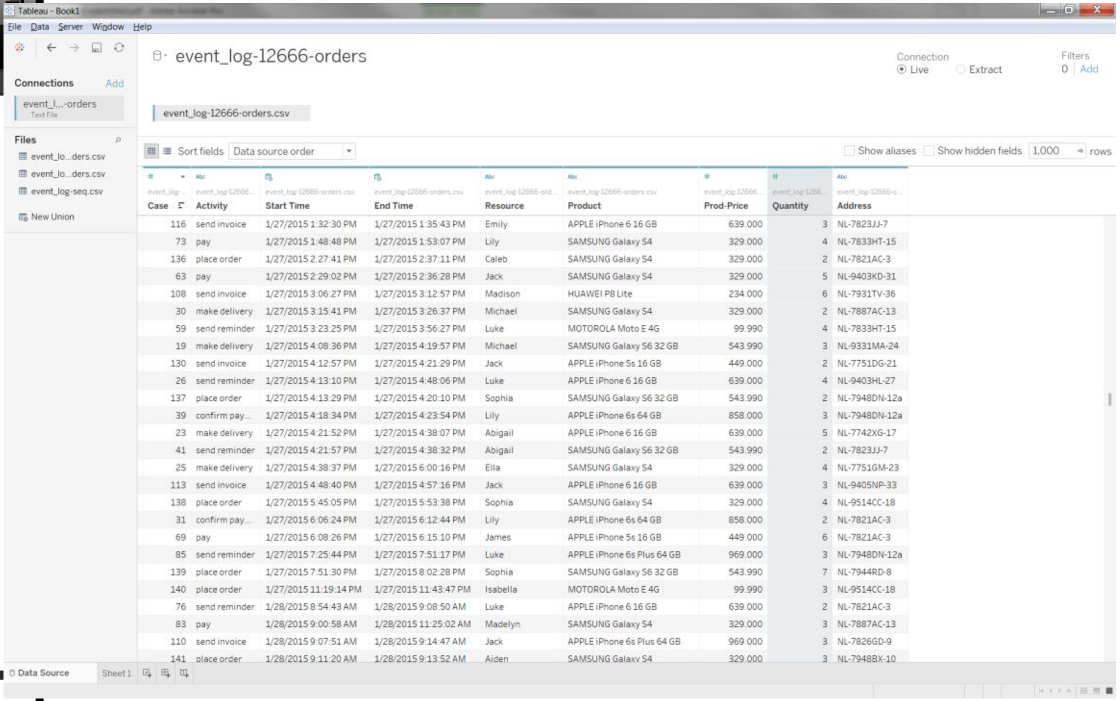
876 traces
6.92% of the log
place order → send invoice → send reminder → pay → prepare delivery → confirm payment → make delivery

420 traces
3.32% of the log
place order → send invoice → send reminder → send reminder → pay → prepare delivery → confirm payment → make delivery

30 traces
0.24% of the log
place order → pay → send invoice → prepare delivery → make delivery → confirm payment

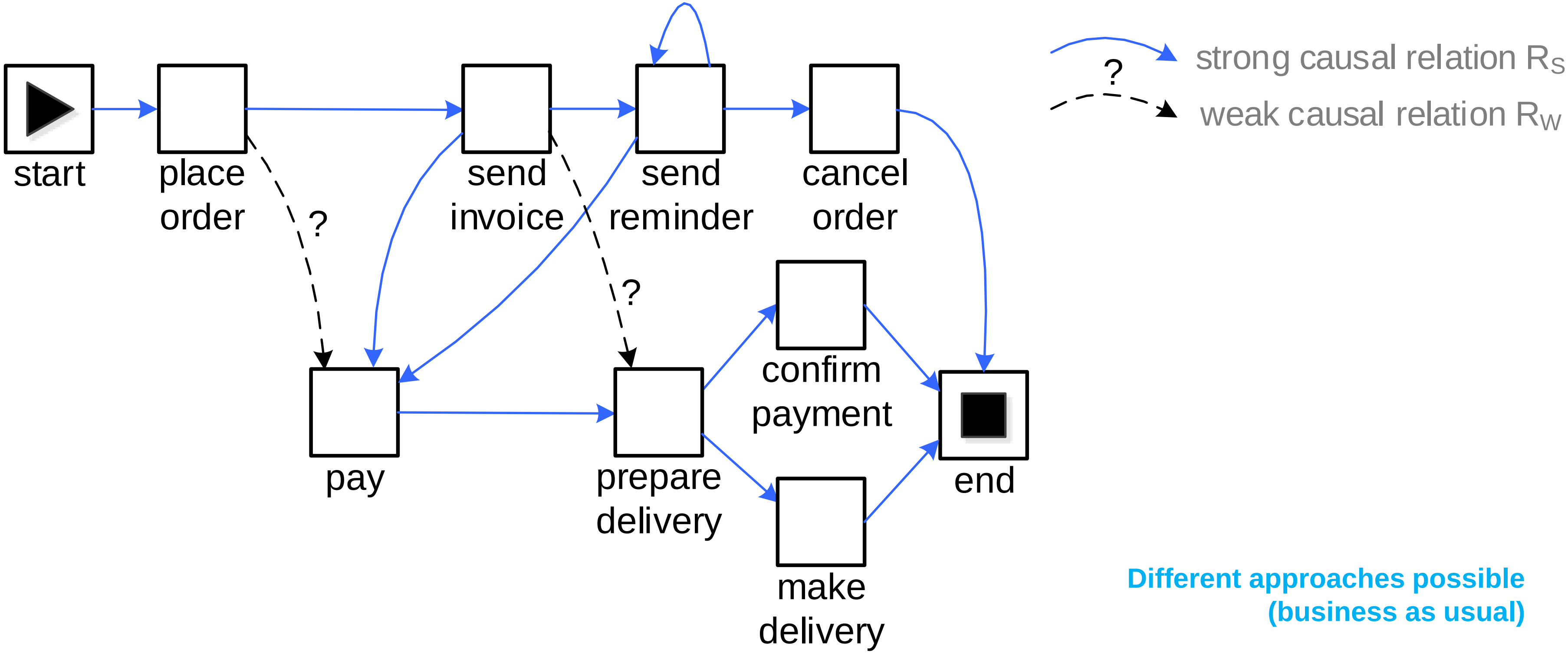
7 traces
0.06% of the log
place order → pay → send invoice → prepare delivery → confirm payment → make delivery

Traces	12,666
Events	80,609
Event Classes	8
Attributes	9
Variants	9
Events per Trace	6.364
First Event	2015-01-05T09:00:07Z
Last Event	2021-04-27T11:11:31Z



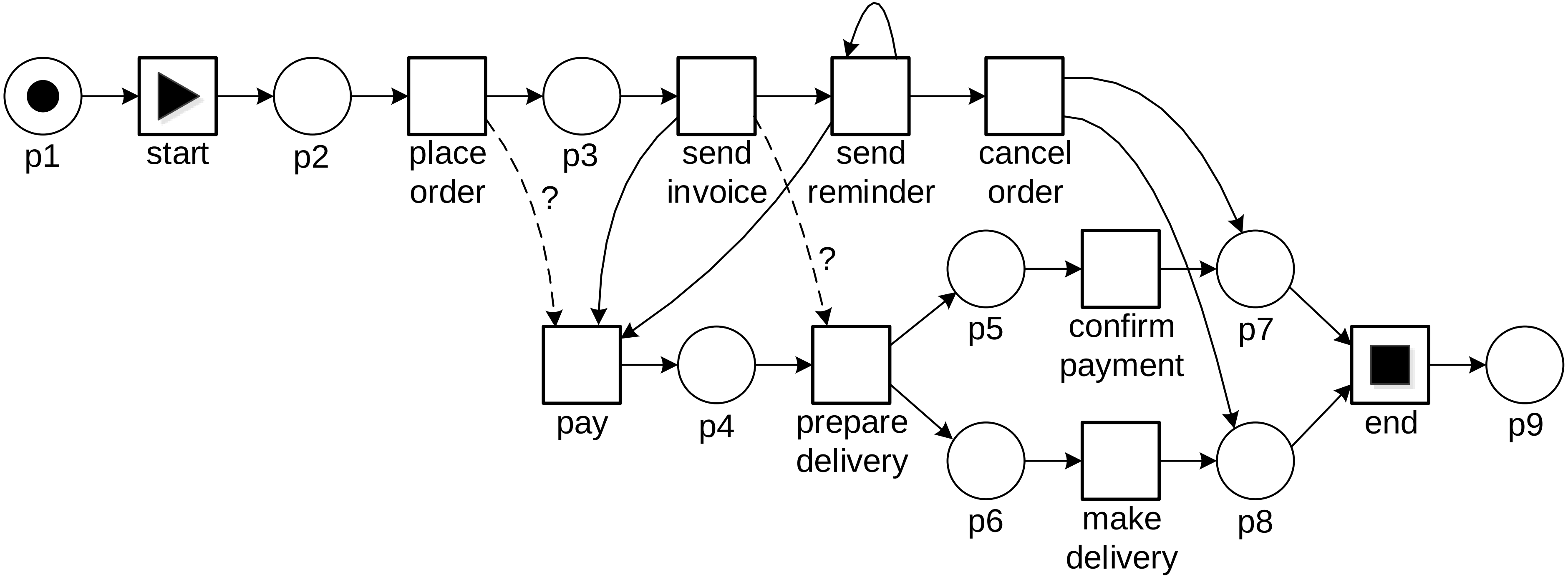
Case	Activity	Start Time	End Time	Resource	Product	Prod-Price	Quantity	Address
116	send invoice	1/27/2015 1:32:30 PM	1/27/2015 1:35:43 PM	Emily	APPLE iPhone 6 16 GB	639.000	3	NL-7823JJ-7
73	pay	1/27/2015 1:48:48 PM	1/27/2015 1:53:07 PM	Lily	SAMSUNG Galaxy S4	329.000	4	NL-7833HT-15
136	place order	1/27/2015 2:27:41 PM	1/27/2015 2:37:11 PM	Caleb	SAMSUNG Galaxy S4	329.000	2	NL-7821AC-3
63	pay	1/27/2015 2:29:02 PM	1/27/2015 2:36:28 PM	Jack	SAMSUNG Galaxy S4	329.000	5	NL-9403KD-31
108	send invoice	1/27/2015 3:06:27 PM	1/27/2015 3:12:57 PM	Madison	HUAWEI P8 Lite	234.000	6	NL-7931TV-36
30	make delivery	1/27/2015 3:15:41 PM	1/27/2015 3:26:37 PM	Michael	SAMSUNG Galaxy S4	329.000	2	NL-7887AC-13
59	send reminder	1/27/2015 3:23:25 PM	1/27/2015 3:56:27 PM	Luke	MOTOROLA Moto E 4G	99.990	4	NL-7833HT-15
19	make delivery	1/27/2015 4:08:36 PM	1/27/2015 4:19:57 PM	Michael	SAMSUNG Galaxy S6 32 GB	543.990	3	NL-9331MA-24
130	send invoice	1/27/2015 4:12:57 PM	1/27/2015 4:21:29 PM	Jack	APPLE iPhone 5s 16 GB	449.000	2	NL-7751DG-21
26	send reminder	1/27/2015 4:13:10 PM	1/27/2015 4:48:06 PM	Luke	APPLE iPhone 6 16 GB	639.000	4	NL-9403HL-27
137	place order	1/27/2015 4:13:29 PM	1/27/2015 4:20:10 PM	Sophia	SAMSUNG Galaxy S6 32 GB	543.990	2	NL-7948DN-12a
39	confirm pay...	1/27/2015 4:18:34 PM	1/27/2015 4:23:54 PM	Lily	APPLE iPhone 6s 64 GB	858.000	3	NL-7948DN-12a
23	make delivery	1/27/2015 4:21:52 PM	1/27/2015 4:38:07 PM	Abigail	APPLE iPhone 6 16 GB	639.000	5	NL-7742XG-17
41	send reminder	1/27/2015 4:21:57 PM	1/27/2015 4:38:32 PM	Abigail	SAMSUNG Galaxy S6 32 GB	543.990	2	NL-7823JJ-7
25	make delivery	1/27/2015 4:38:37 PM	1/27/2015 6:00:16 PM	Ella	SAMSUNG Galaxy S4	329.000	4	NL-7751GM-23
113	send invoice	1/27/2015 4:48:40 PM	1/27/2015 4:57:16 PM	Jack	APPLE iPhone 6 16 GB	639.000	3	NL-9405NP-33
138	place order	1/27/2015 5:45:05 PM	1/27/2015 5:53:38 PM	Sophia	SAMSUNG Galaxy S4	329.000	4	NL-9514CC-18
31	confirm pay...	1/27/2015 6:06:24 PM	1/27/2015 6:12:44 PM	Lily	APPLE iPhone 6s 64 GB	858.000	2	NL-7821AC-3
69	pay	1/27/2015 6:08:26 PM	1/27/2015 6:15:10 PM	James	APPLE iPhone 5s 16 GB	449.000	6	NL-7821AC-3
85	send reminder	1/27/2015 7:25:44 PM	1/27/2015 7:51:17 PM	Luke	APPLE iPhone 6s Plus 64 GB	969.000	3	NL-7948DN-12a
139	place order	1/27/2015 7:51:30 PM	1/27/2015 8:02:28 PM	Sophia	SAMSUNG Galaxy S6 32 GB	543.990	7	NL-7944RD-8
140	place order	1/27/2015 11:19:14 PM	1/27/2015 11:43:47 PM	Isabella	MOTOROLA Moto E 4G	99.990	3	NL-9514CC-18
76	send reminder	1/28/2015 8:54:43 AM	1/28/2015 9:08:50 AM	Luke	APPLE iPhone 6 16 GB	639.000	2	NL-7821AC-3
83	pay	1/28/2015 9:00:58 AM	1/28/2015 11:25:02 AM	Madelyn	SAMSUNG Galaxy S4	329.000	3	NL-7887AC-13
110	send invoice	1/28/2015 9:07:51 AM	1/28/2015 9:14:47 AM	Jack	APPLE iPhone 6s Plus 64 GB	969.000	3	NL-7826GD-9
141	place order	1/28/2015 9:11:20 AM	1/28/2015 9:13:52 AM	Aiden	SAMSUNG Galaxy S4	329.000	3	NL-7948BX-10

Phase 1: Learn a Causal Graph

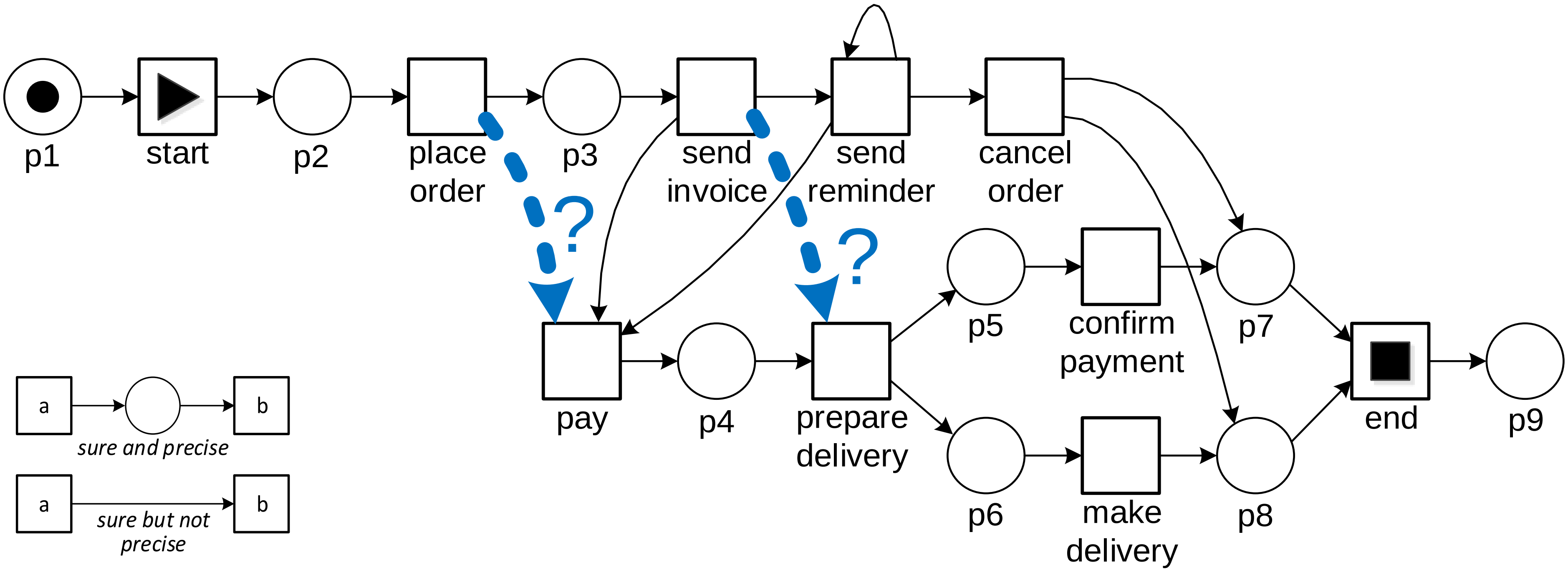


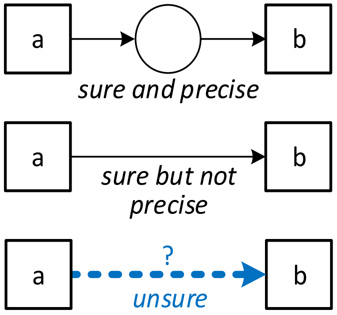
Different approaches possible (business as usual)

Phase 2: Learn a Hybrid Petri Net

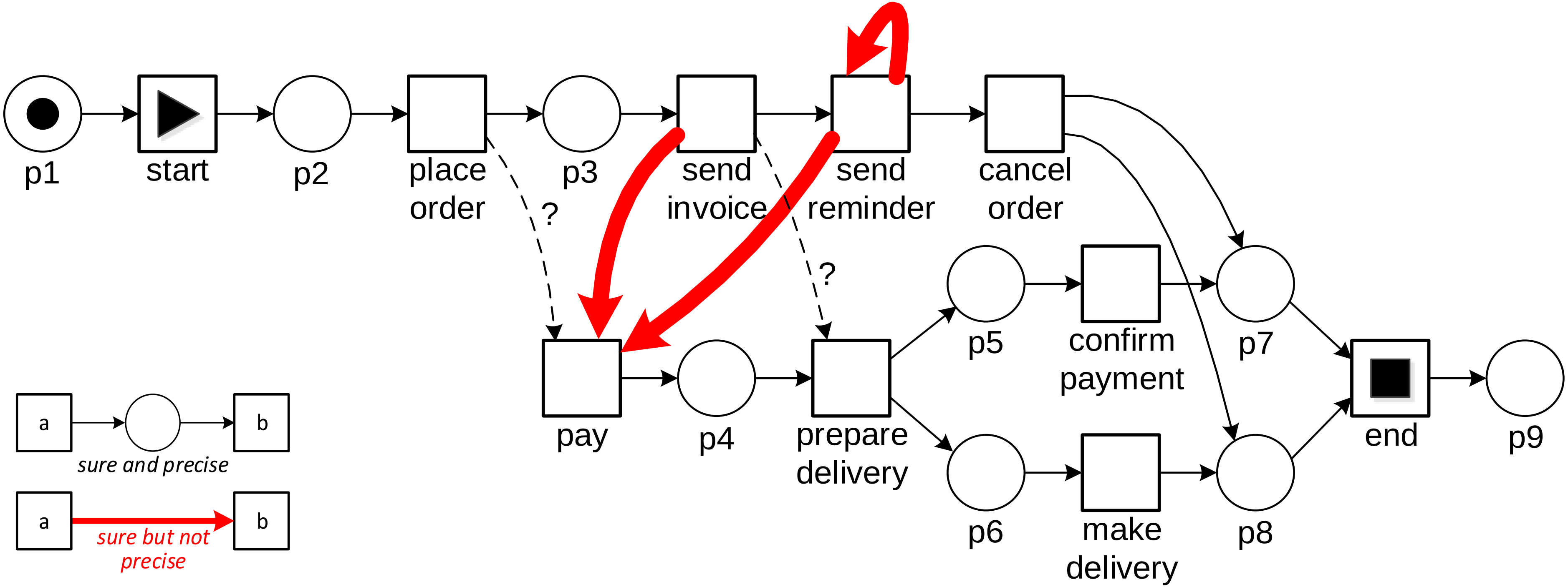


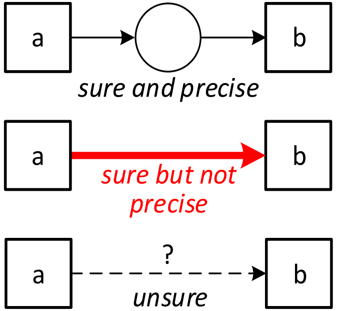
Phase 2: Learn a Hybrid Petri Net



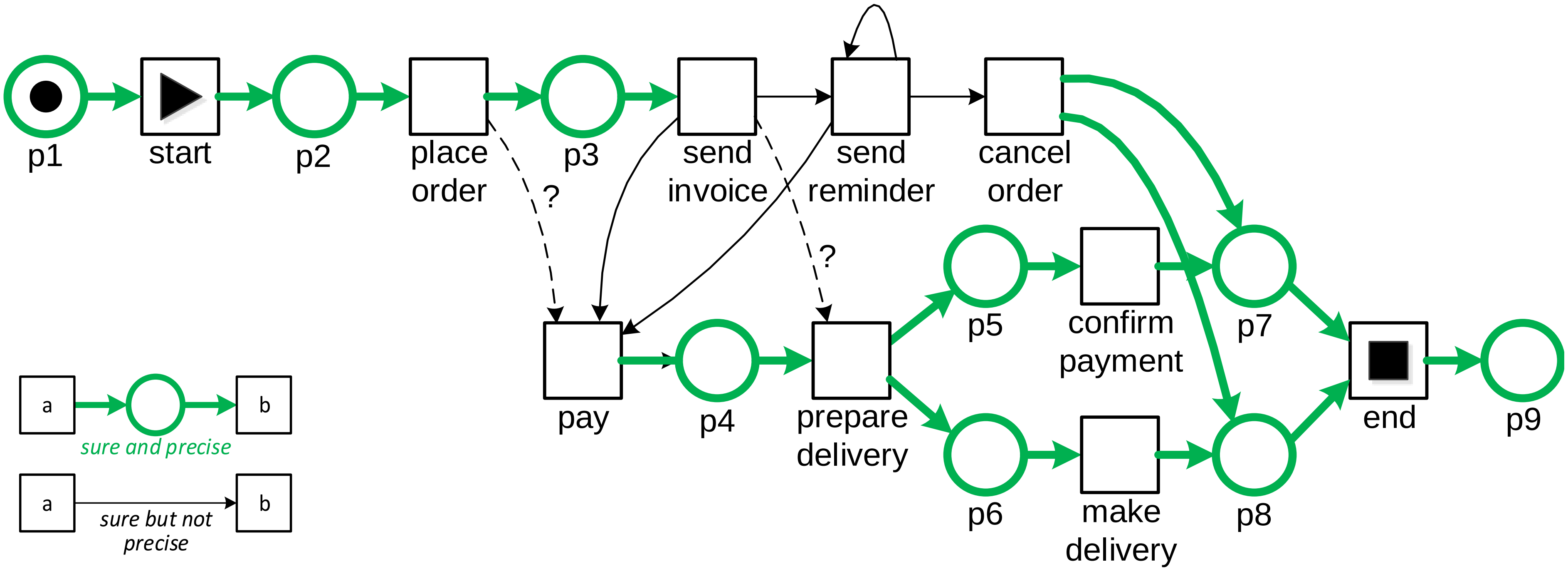


Phase 2: Learn a Hybrid Petri Net





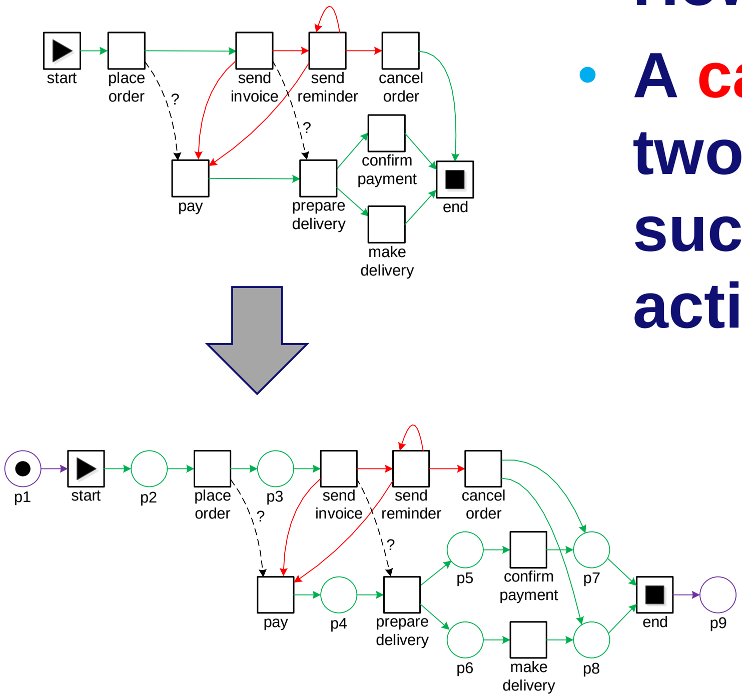
Phase 2: Learn a Hybrid Petri Net

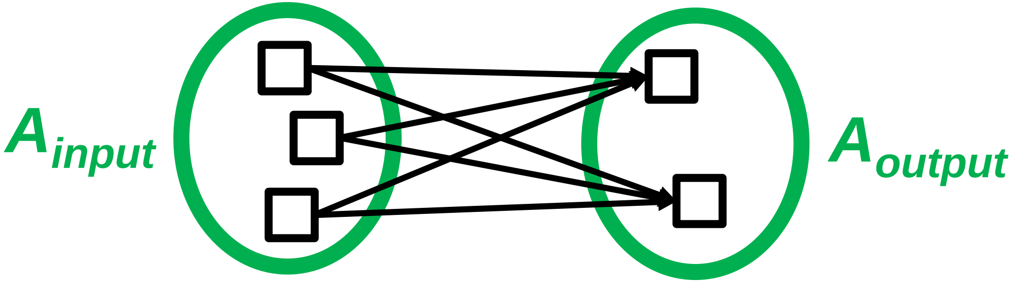


©Wil van der Aalst (use only with permission & acknowledgements)

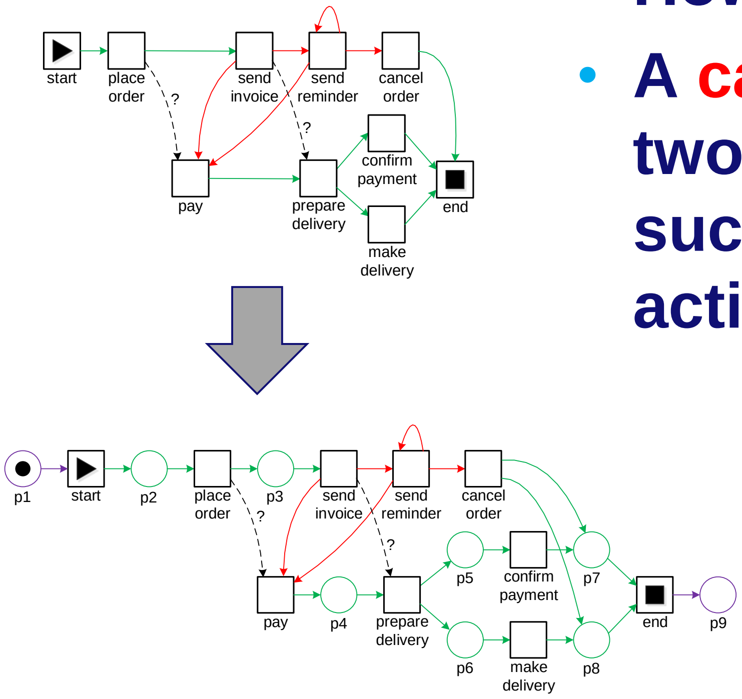
Phase 2: Learn a Hybrid Petri Net

- **How? ►Generate candidate places**
- **A candidate place is characterized by two sets of activities A_{input} and A_{output} such that sure arcs are connecting any activity in A_{input} to any activity in A_{output}**

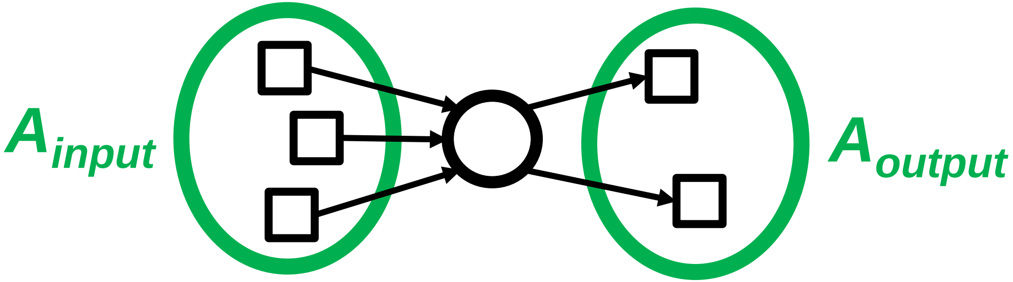




Phase 2: Learn a Hybrid Petri Net

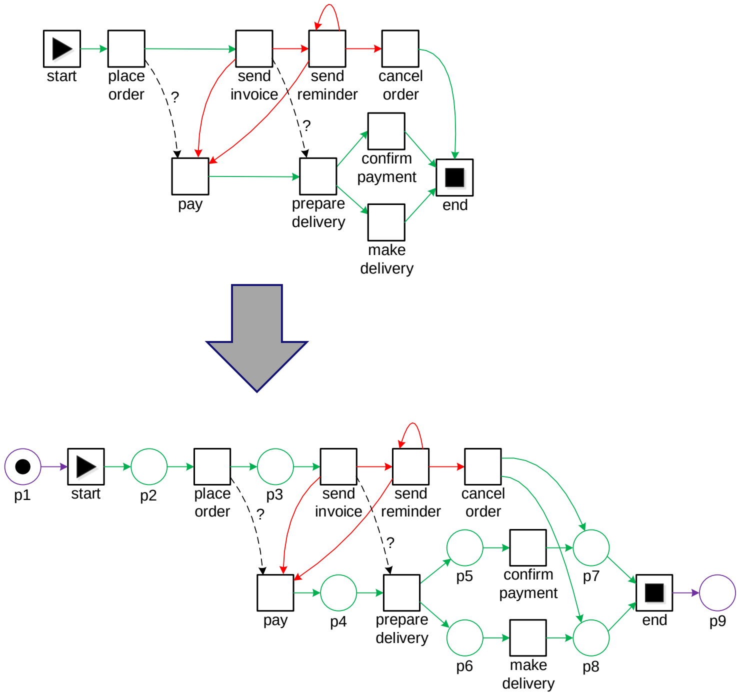


- How? ▶Generate candidate places
- A candidate place is characterized by two sets of activities A_{input} and A_{output} such that sure arcs are connecting any activity in A_{input} to any activity in A_{output}

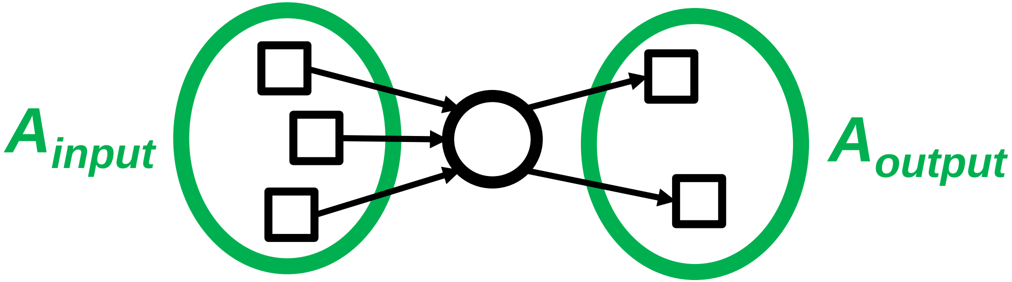


©Wil van der Aalst (use only with permission & acknowledgements)

Phase 2: Learn a Hybrid Petri Net

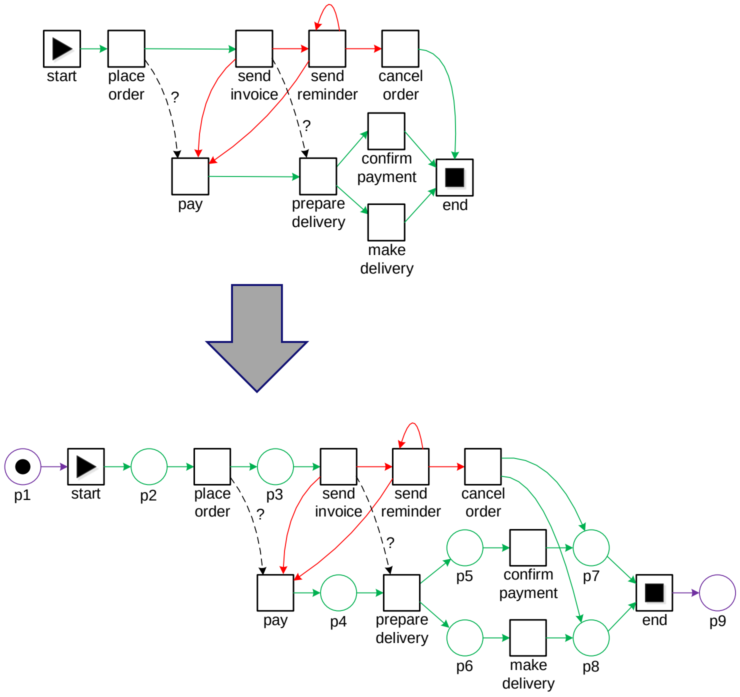


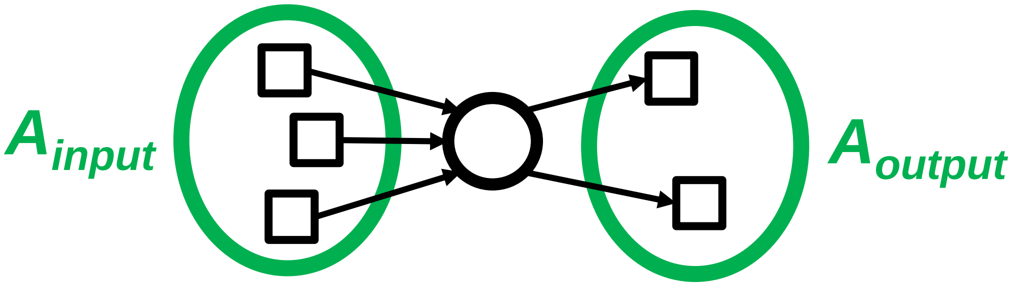
- **Determine the quality of each candidate place $p=(A_{input},A_{output})$**



- **Ideally: start empty, finish empty, not negative (compare ILP miner)**

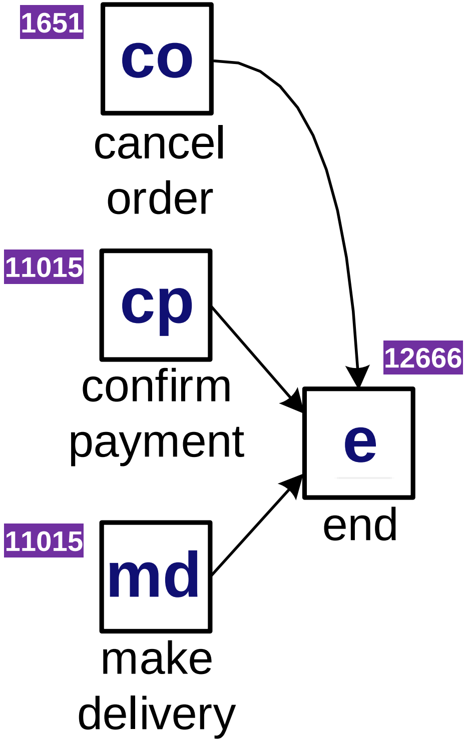
Phase 2: Learn a Hybrid Petri Net





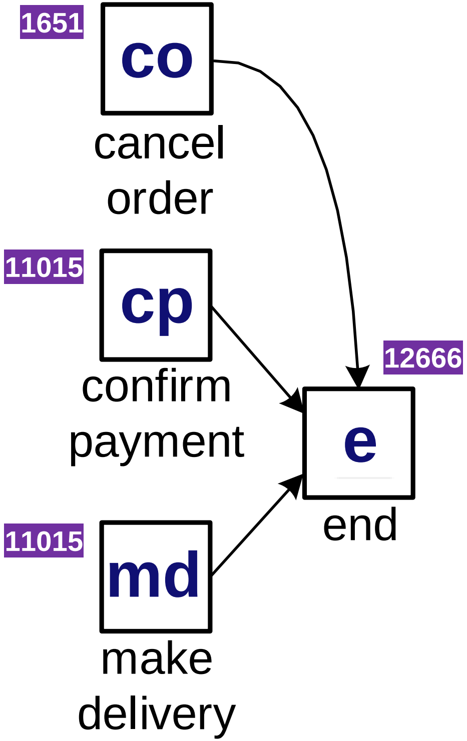
- **Three possible scoring functions:**
  a) **Fraction of cases perfectly fitting**
  b) **Fraction of relevant cases perfectly fitting**
  c) **Global score (extremely efficient)**

Generate candidate places and evaluate



({co},{e})
({cp},{e})
({md},{e})
({co,cp},{e})
({co,md},{e})
({cp,md},{e})
({co,cp,md},{e})

Generate candidate places and evaluate



~~({co},{e})~~

~~({cp},{e})~~

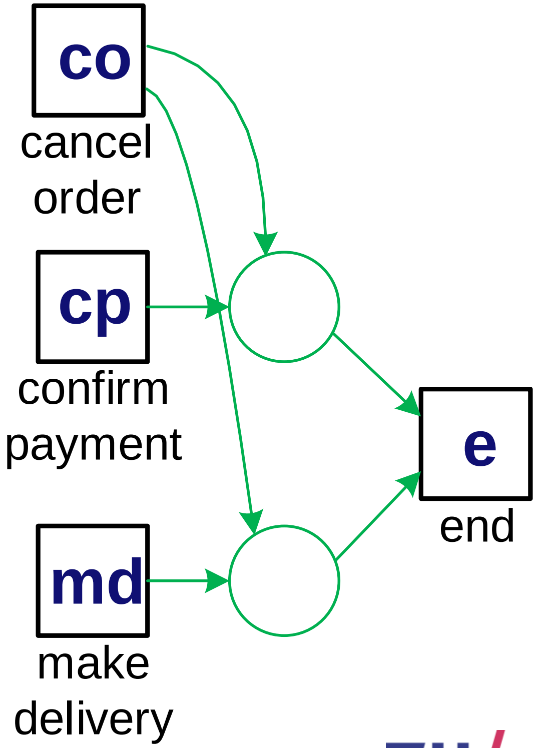
~~({md},{e})~~

({co,cp},{e})

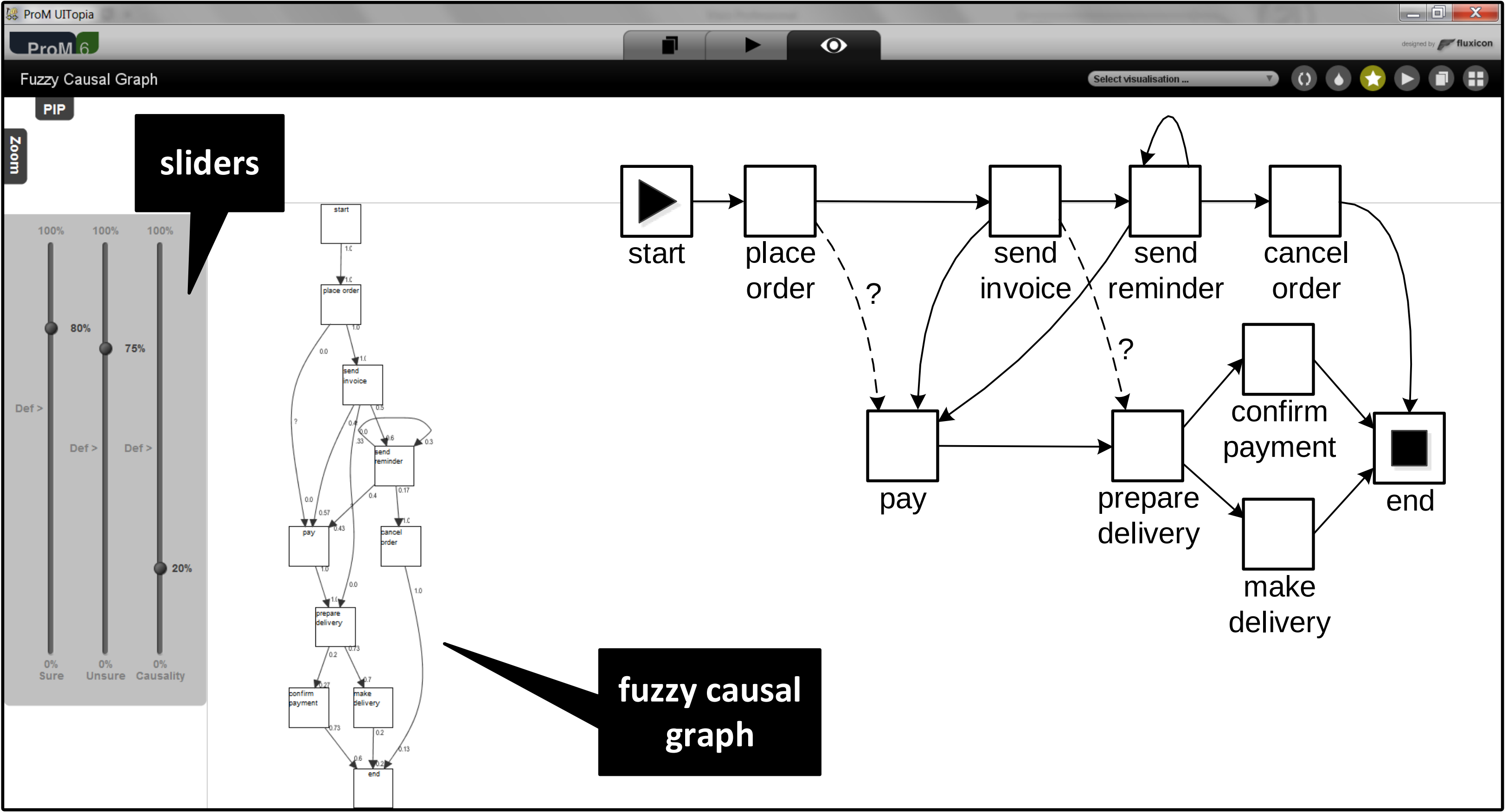
({co,md},{e})

~~({cp,md},{e})~~

~~({co,cp,md},{e})~~



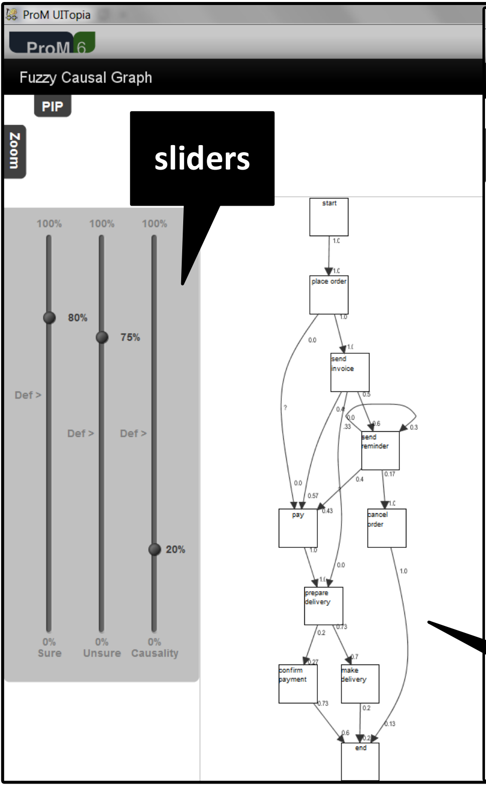
ProM Implementation



sliders

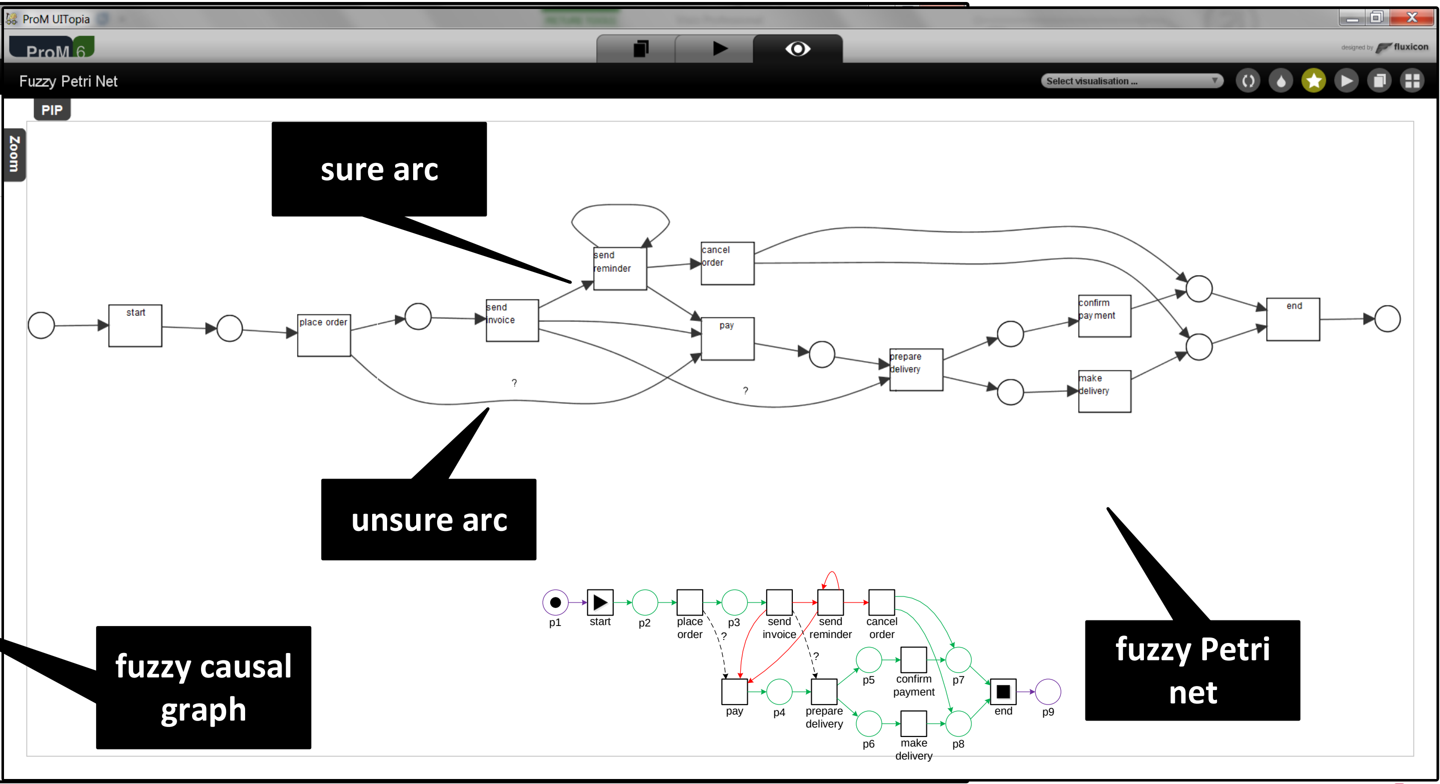
fuzzy causal graph

ProM Implementation



sliders

fuzzy causal graph



sure arc

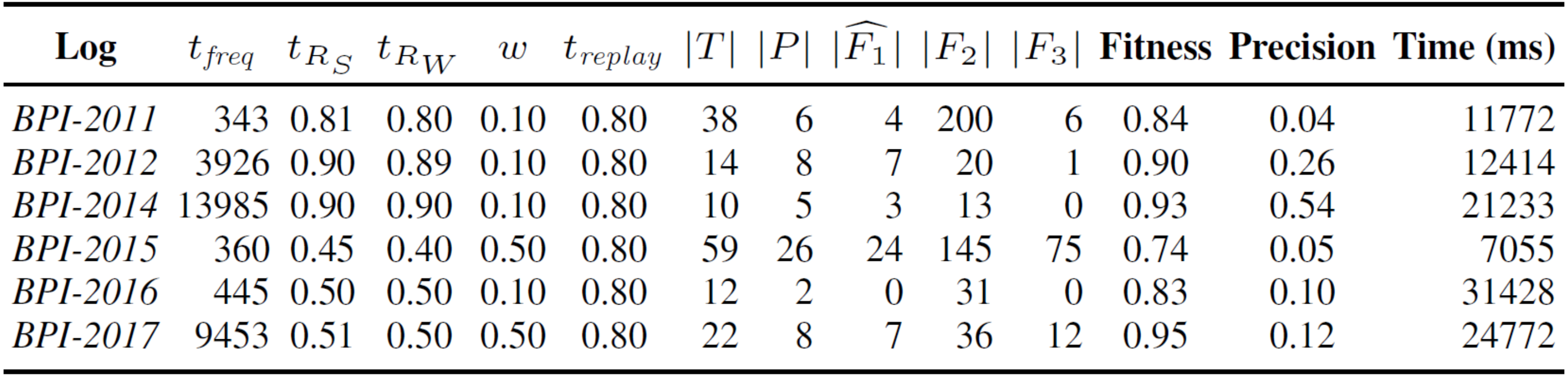
unsure arc

fuzzy Petri net

Parameters

- Threshold for activity frequency (t_{freq})
- Parameters used to compute strength of relations taking into account concurrency and loops (*c* and *w*)
- Thresholds for strong and weak causalities (t_{R_S} and $t_{R_W}$with $t_{R_S} > t_{R_W}$)
- Threshold for place quality (t_{replay})

Evaluation (see paper and technical report for more details)



Log	t_{freq}	t_{R_S}	t_{R_W}	w	t_{replay}	$\lvert T\rvert$	$\lvert P\rvert$	$\lvert\widehat{F_1}\rvert$	$\lvert F_2\rvert$	$\lvert F_3\rvert$	**Fitness**	**Precision**	**Time (ms)**
BPI-2011	343	0.81	0.80	0.10	0.80	38	6	4	200	6	0.84	0.04	11772
BPI-2012	3926	0.90	0.89	0.10	0.80	14	8	7	20	1	0.90	0.26	12414
BPI-2014	13985	0.90	0.90	0.10	0.80	10	5	3	13	0	0.93	0.54	21233
BPI-2015	360	0.45	0.40	0.50	0.80	59	26	24	145	75	0.74	0.05	7055
BPI-2016	445	0.50	0.50	0.10	0.80	12	2	0	31	0	0.83	0.10	31428
BPI-2017	9453	0.51	0.50	0.50	0.80	22	8	7	36	12	0.95	0.12	24772

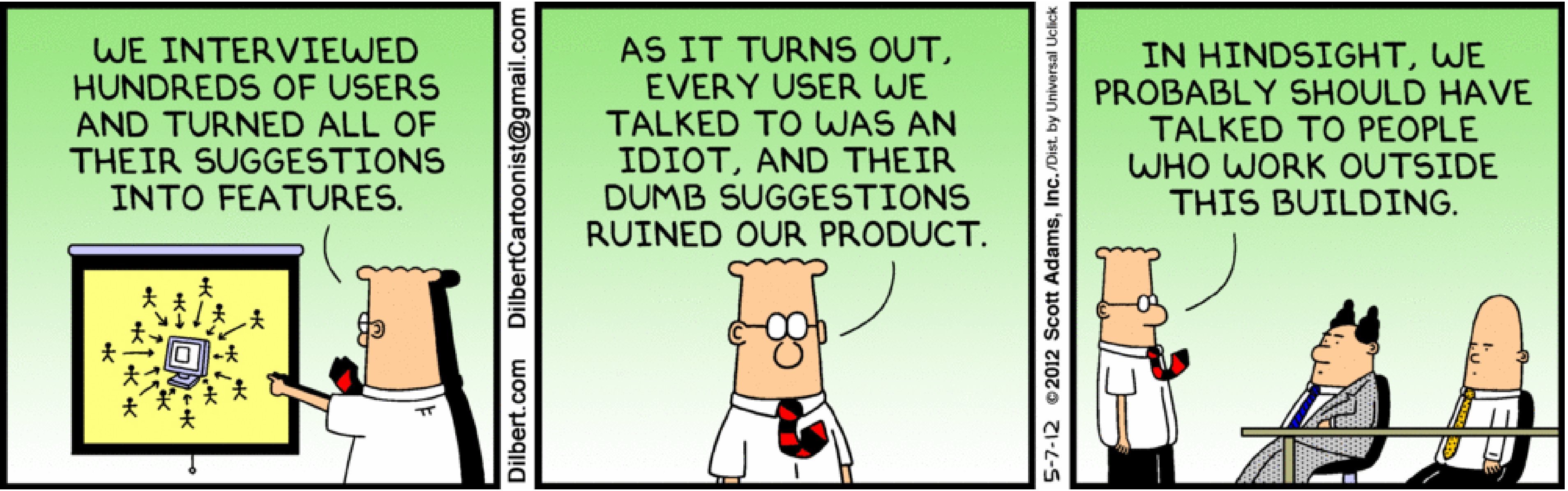
- **Behaves as expected, e.g., when t_{replay} goes up fitness goes up and precision goes down**

arxiv.org/abs/1703.06125

Performance

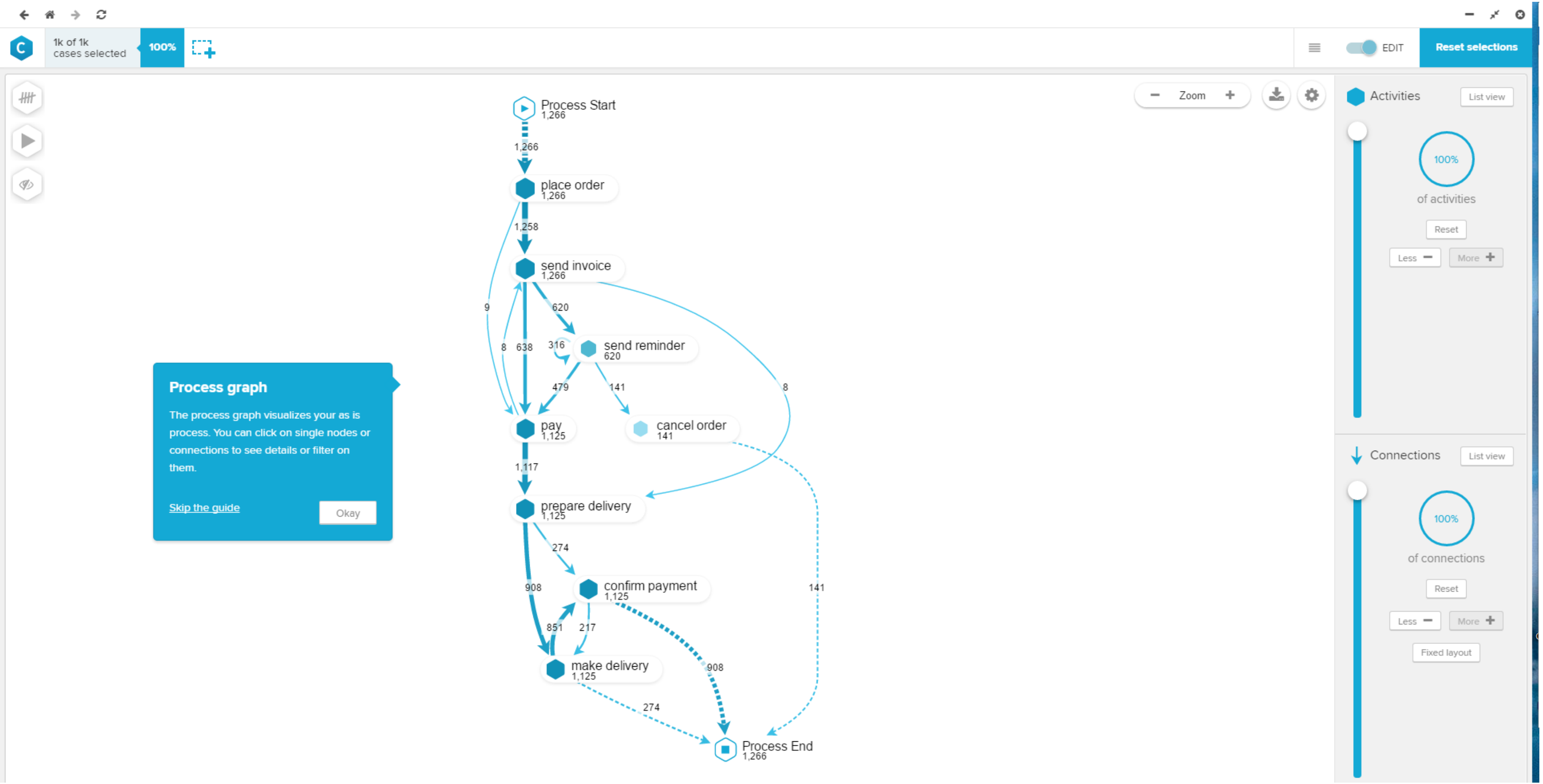
- Good, but room for improvement
- Smartly pruning the set of candidate places (avoid conflicting or less informative places)
- Lazy place evaluation
- Distribution/decomposition using e.g. Spark (see joint work with Long Cheng and Boudewijn van Dongen in a slightly different setting)

Evaluation is not easy


WE INTERVIEWED HUNDREDS OF USERS AND TURNED ALL OF THEIR SUGGESTIONS INTO FEATURES.
AS IT TURNS OUT, EVERY USER WE TALKED TO WAS AN IDIOT, AND THEIR DUMB SUGGESTIONS RUINED OUR PRODUCT.
IN HINDSIGHT, WE PROBABLY SHOULD HAVE TALKED TO PEOPLE WHO WORK OUTSIDE THIS BUILDING.
Dilbert.com DilbertCartoonist@gmail.com
5-7-12 ©2012 Scott Adams, Inc./Dist. by Universal Uclick

©Wil van der Aalst (use only with permission & acknowledgements)

But, the need is obvious



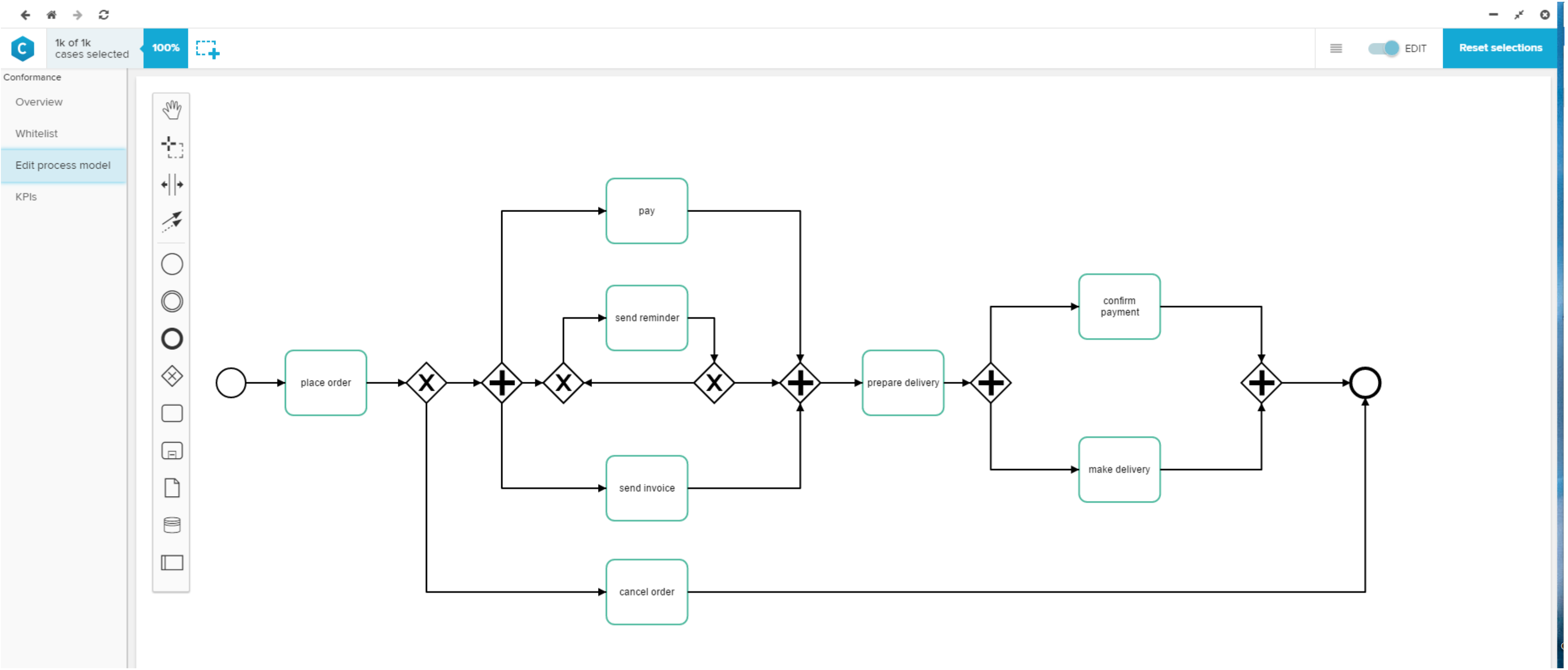
Process graph

The process graph visualizes your as is process. You can click on single nodes or connections to see details or filter on them.

Skip the guide

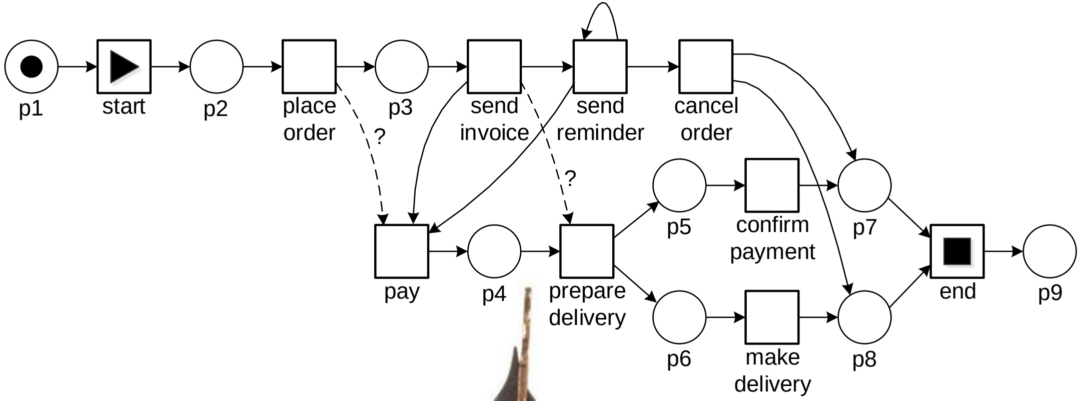
Okay

But, the need is obvious



Conclusion

process
mining
hybrid
process
discovery



Learning Hybrid Process Models from Events

Process Discovery Without Faking Confidence

Thanks!

Looking for talented PhDs and Postdocs !!

RWTH AACHEN UNIVERSITY