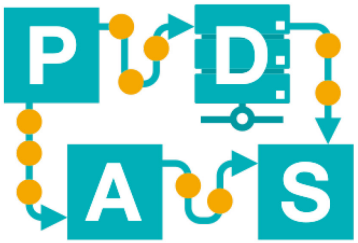
Process Mining Unsupervised

Lecture 14

IDS-L14

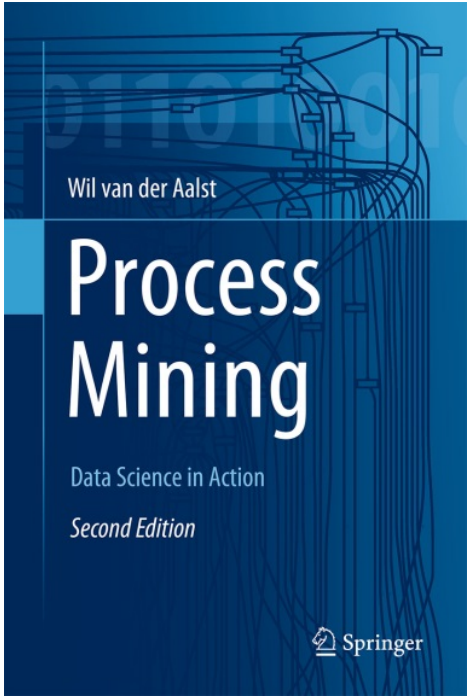


Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Outline of Today's Lecture

- Positioning process mining
- Event data (as described before)
- Process discovery (overall challenges)
- Process discovery: Bottom-up (learning places)
- Process discovery: Top-down (inductive mining)
- Tooling and applications

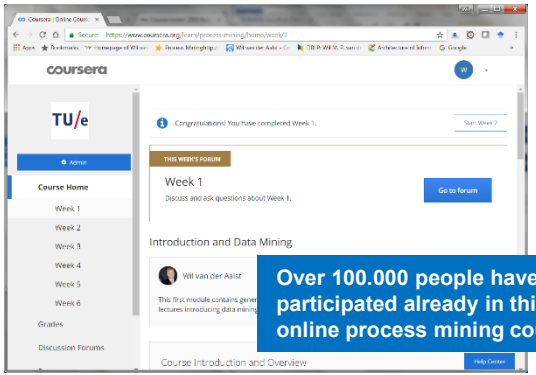


W. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, Berlin, 2016" (http://springer.com/9783662498507)

Chair of Process and Data Science

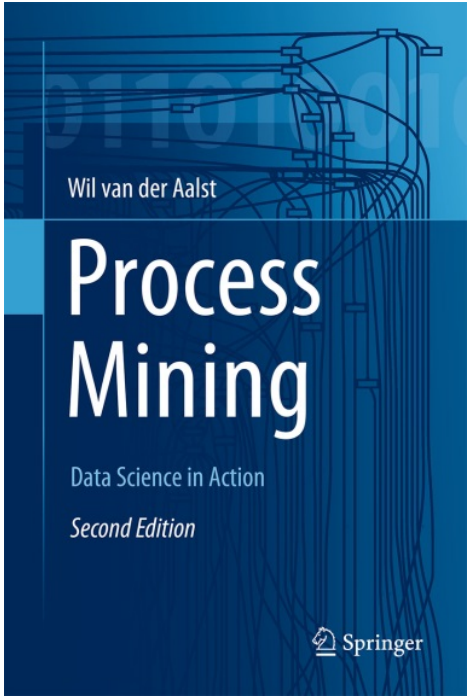
Material (only as background information or if you want to dive deeper)

- Chapters 2 (PM), 5 (event data), 6 (process discovery), 7.5 (inductive miner) of W. van der Aalst. Process Mining: Data Science in Action. Springer-Verlag, Berlin, 2016 (http://springer.com/9783662498507)
- Coursera Process Mining Course https://www.coursera.org/course/procmin



Over 100.000 people have participated already in this online process mining course.

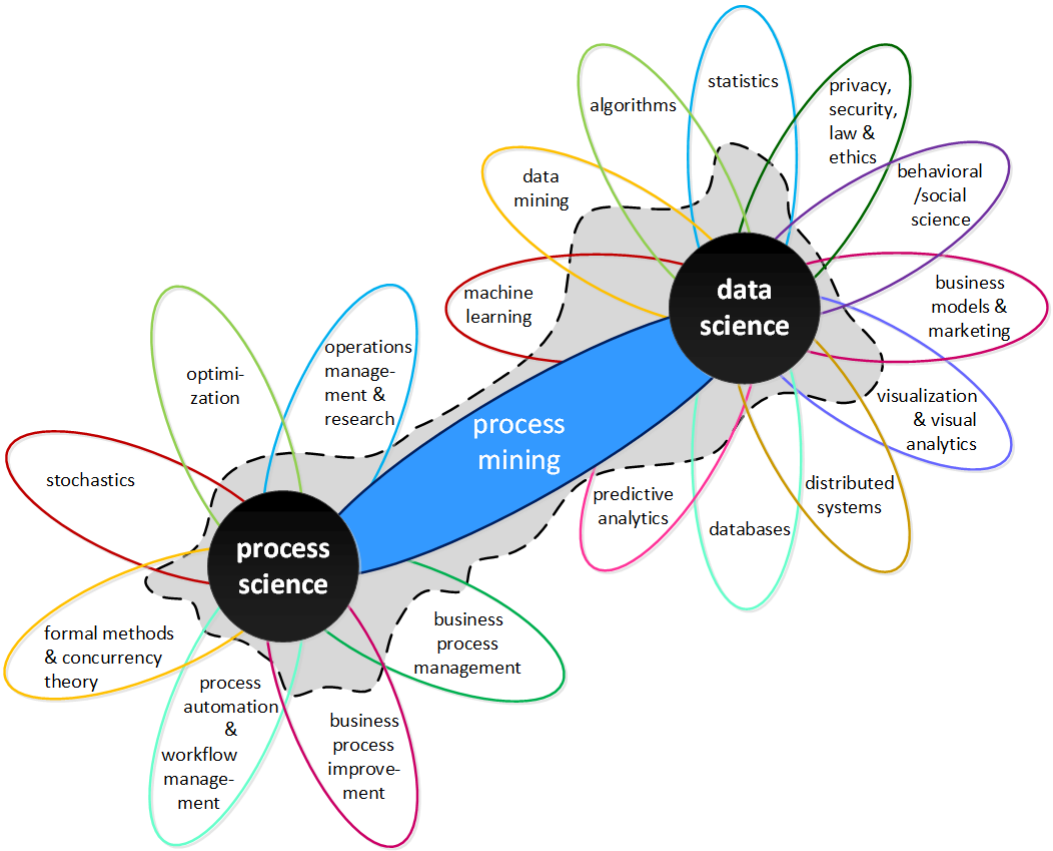
coursera
education for everyone



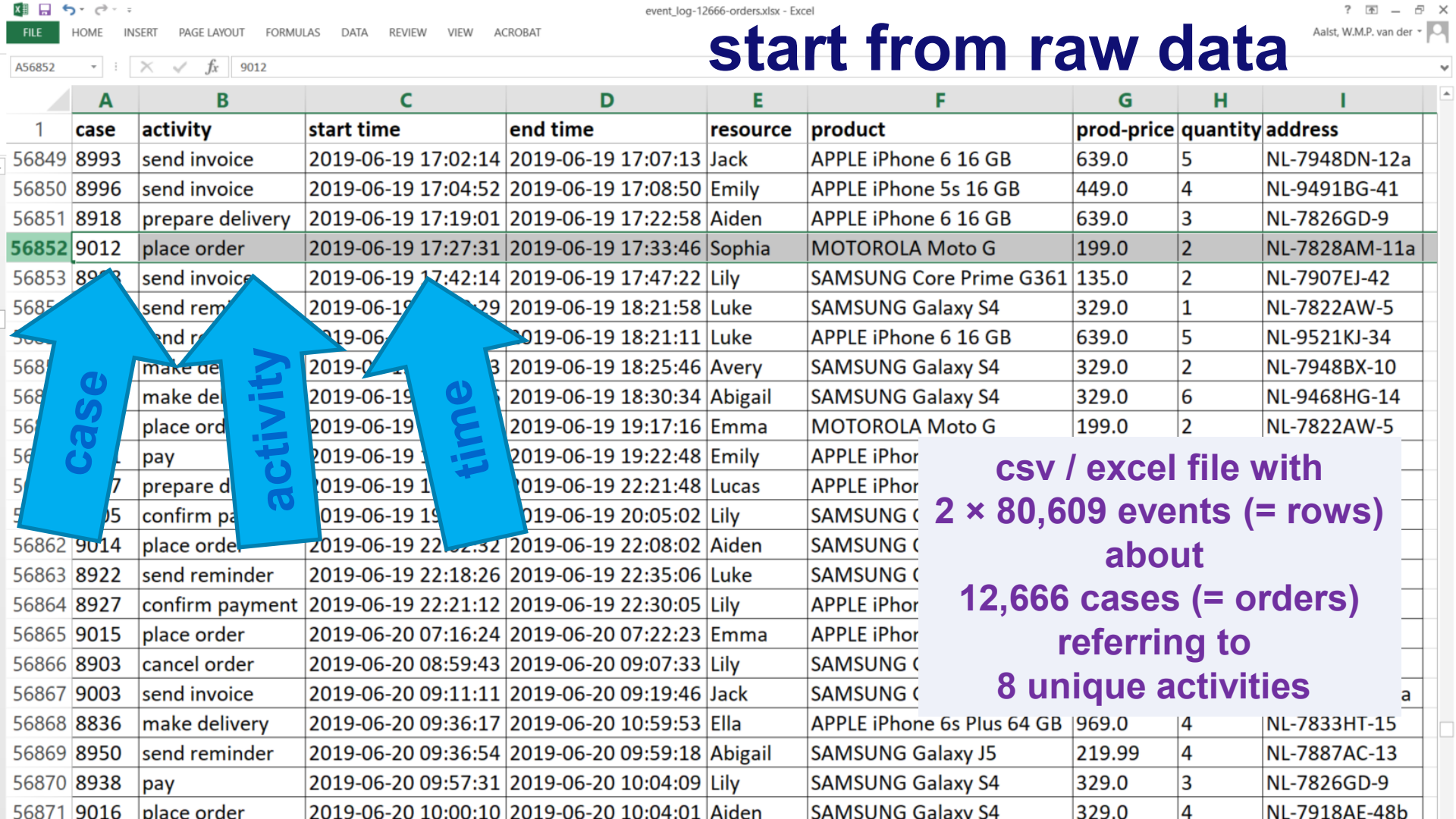
Chair of Process and Data Science

Positioning process mining

Positioning process mining



Chair of Process and Data Science



	A	B	C	D	E	F	G	H	I
1	case	activity	start time	end time	resource	product	prod-price	quantity	address
56849	8993	send invoice	2019-06-19 17:02:14	2019-06-19 17:07:13	Jack	APPLE iPhone 6 16 GB	639.0	5	NL-7948DN-12a
56850	8996	send invoice	2019-06-19 17:04:52	2019-06-19 17:08:50	Emily	APPLE iPhone 5s 16 GB	449.0	4	NL-9491BG-41
56851	8918	prepare delivery	2019-06-19 17:19:01	2019-06-19 17:22:58	Aiden	APPLE iPhone 6 16 GB	639.0	3	NL-7826GD-9
56852	9012	place order	2019-06-19 17:27:31	2019-06-19 17:33:46	Sophia	MOTOROLA Moto G	199.0	2	NL-7828AM-11a
56853	89[illegible]	send invoice	2019-06-19 17:42:14	2019-06-19 17:47:22	Lily	SAMSUNG Core Prime G361	135.0	2	NL-7907EJ-42
5685[illegible]		send rem[illegible]	2019-06-19[illegible]29	2019-06-19 18:21:58	Luke	SAMSUNG Galaxy S4	329.0	1	NL-7822AW-5
56[illegible]		[illegible]nd r[illegible]	2019-06[illegible]	2019-06-19 18:21:11	Luke	APPLE iPhone 6 16 GB	639.0	5	NL-9521KJ-34
568[illegible]		make de[illegible]	2019-0[illegible]3	2019-06-19 18:25:46	Avery	SAMSUNG Galaxy S4	329.0	2	NL-7948BX-10
568[illegible]		make del[illegible]	2019-06-19[illegible]	2019-06-19 18:30:34	Abigail	SAMSUNG Galaxy S4	329.0	6	NL-9468HG-14
56[illegible]		place ord[illegible]	2019-06-19[illegible]	2019-06-19 19:17:16	Emma	MOTOROLA Moto G	199.0	2	NL-7822AW-5
56[illegible]		pay	2019-06-19[illegible]	2019-06-19 19:22:48	Emily	APPLE iPhor[illegible]			
5[illegible]		prepare d[illegible]	2019-06-19 1[illegible]	2019-06-19 22:21:48	Lucas	APPLE iPhor[illegible]			
5[illegible]		confirm pa[illegible]	2019-06-19 19[illegible]	2019-06-19 20:05:02	Lily	SAMSUNG G[illegible]			
56862	9014	place orde[illegible]	2019-06-19 22[illegible]:32	2019-06-19 22:08:02	Aiden	SAMSUNG G[illegible]			
56863	8922	send reminder	2019-06-19 22:18:26	2019-06-19 22:35:06	Luke	SAMSUNG G[illegible]			
56864	8927	confirm payment	2019-06-19 22:21:12	2019-06-19 22:30:05	Lily	APPLE iPhor[illegible]			
56865	9015	place order	2019-06-20 07:16:24	2019-06-20 07:22:23	Emma	APPLE iPhor[illegible]			
56866	8903	cancel order	2019-06-20 08:59:43	2019-06-20 09:07:33	Lily	SAMSUNG G[illegible]			
56867	9003	send invoice	2019-06-20 09:11:11	2019-06-20 09:19:46	Jack	SAMSUNG G[illegible]			[illegible]a
56868	8836	make delivery	2019-06-20 09:36:17	2019-06-20 10:59:53	Ella	APPLE iPhone 6s Plus 64 GB	969.0	4	NL-7833HT-15
56869	8950	send reminder	2019-06-20 09:36:54	2019-06-20 09:59:18	Abigail	SAMSUNG Galaxy J5	219.99	4	NL-7887AC-13
56870	8938	pay	2019-06-20 09:57:31	2019-06-20 10:04:09	Lily	SAMSUNG Galaxy S4	329.0	3	NL-7826GD-9
56871	9016	place order	2019-06-20 10:00:10	2019-06-20 10:04:01	Aiden	SAMSUNG Galaxy S4	329.0	4	NL-7918AE-48b

Let’s focus on the events of order 9012

1	case	activity	start time	resource
56852	9012	place order	2019-06-19 17:27:31	Sophia
56946	9012	send invoice	2019-06-21 12:27:02	Jack
57302	9012	pay	2019-07-02 09:41:51	Emily
57439	9012	prepare delivery	2019-07-05 16:16:26	Olivia
57470	9012	confirm payment	2019-07-08 10:42:33	Lily
57506	9012	make delivery	2019-07-09 08:30:15	Ella

ProM UITopia

ProM 6

XES Event Log

Select all | Deselect all

4,586 traces
36.21% of the log — place order, send invoice

2,312 traces
18.25% of the log — place order, send invoice

1,666 traces
13.15% of the log — place order, send invoice

1,651 traces
13.03% of the log — place order, send invoice

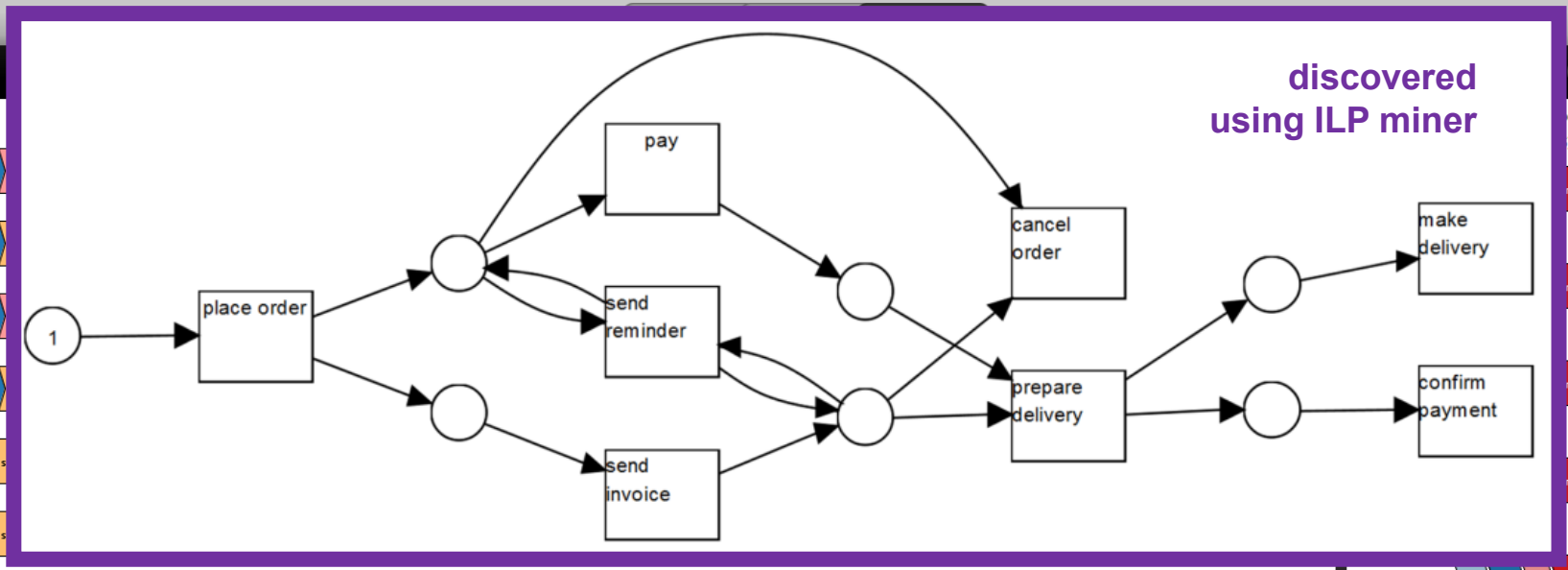
1,118 traces
8.83% of the log — place order, send invoice

876 traces
6.92% of the log — place order, send invoice

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



12,666 cases (orders)
80,609 events (only
using completes)

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

10003 — plac, send, pay, prep

10004 — plac, send, pay, prep

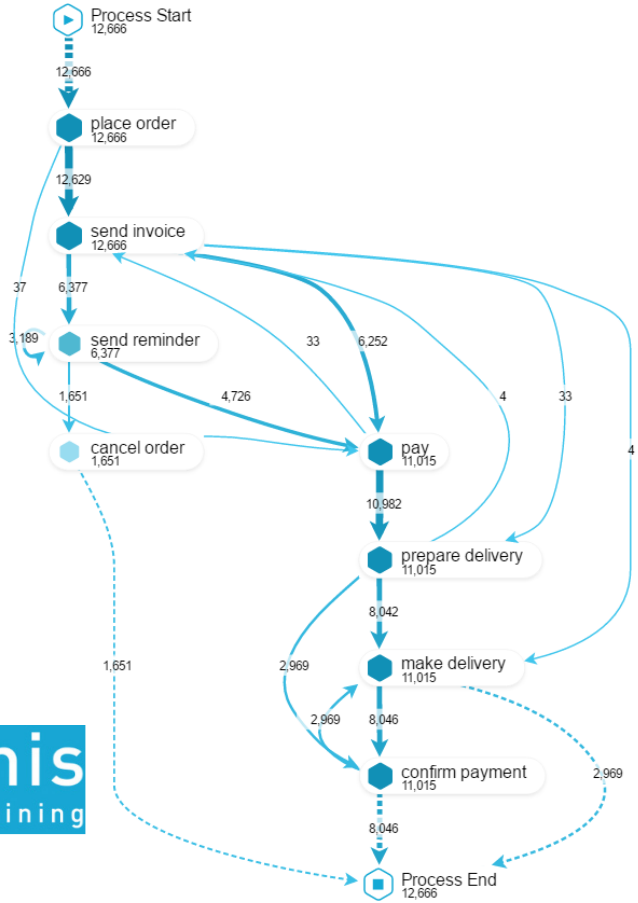
10009 — plac, send, pay, prep

Name

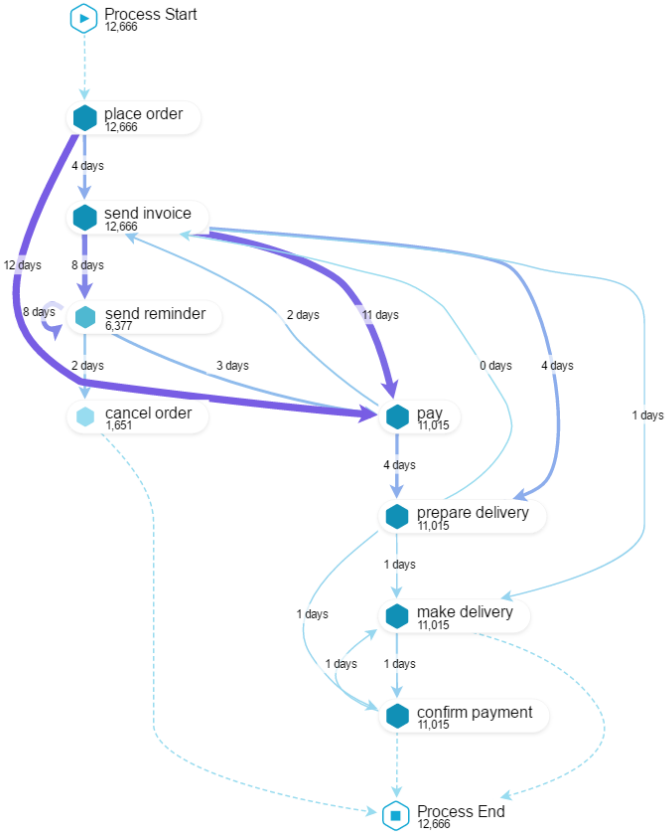
Powered by ProM

“happy”

“freq”



“time”



celonis process mining

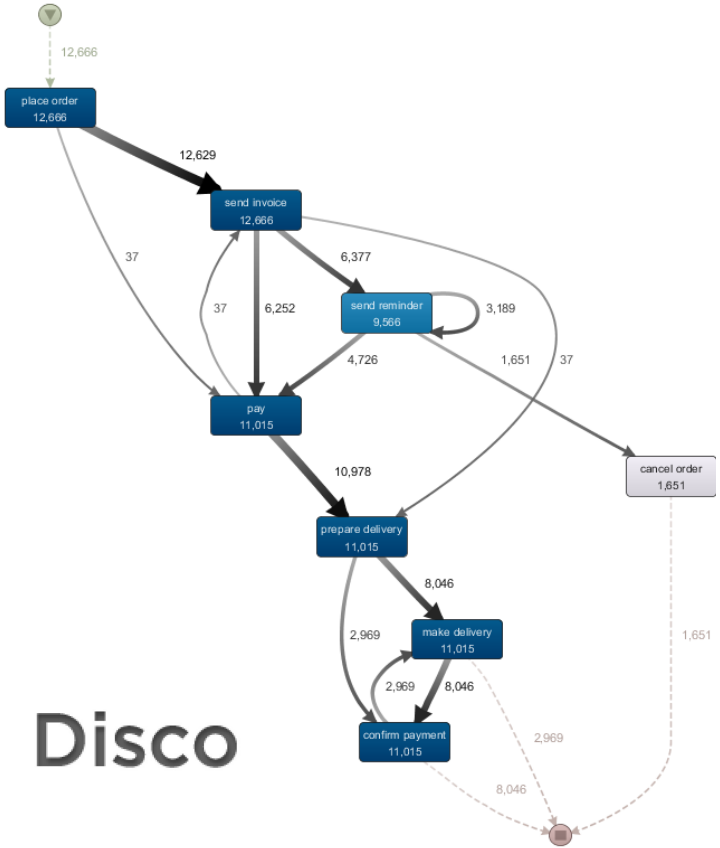
Chair of Process and Data Science

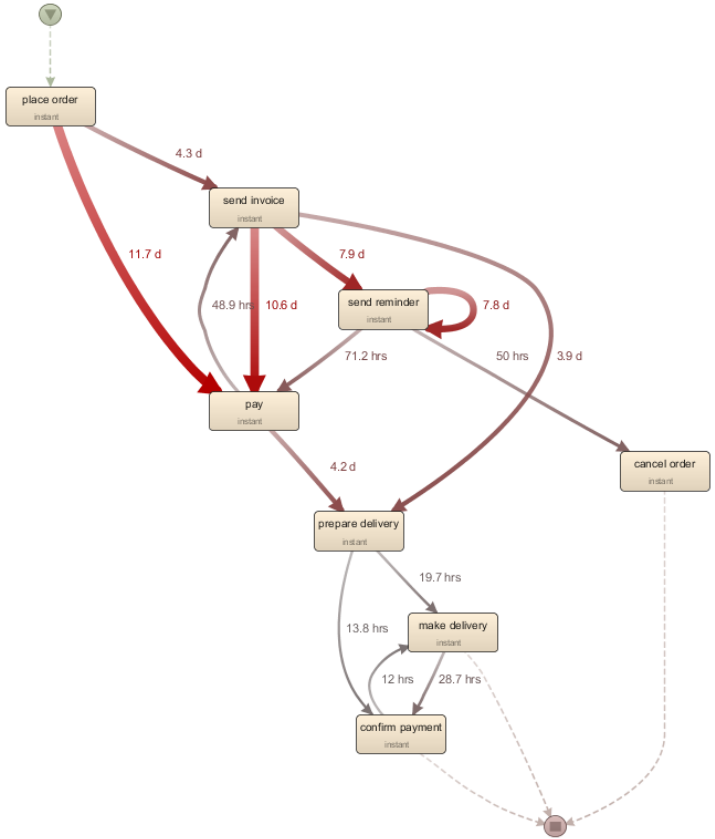
“happy”

“freq”

“time”

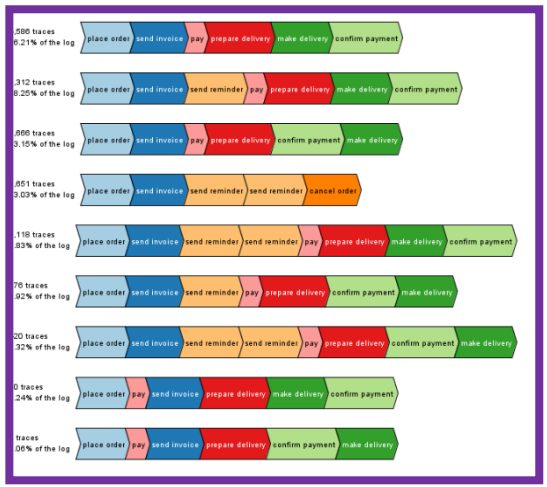
Disco

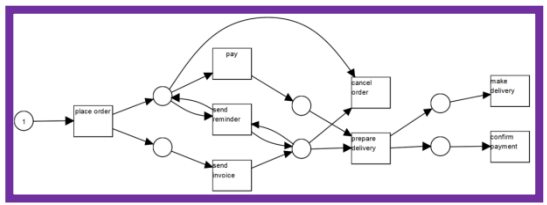


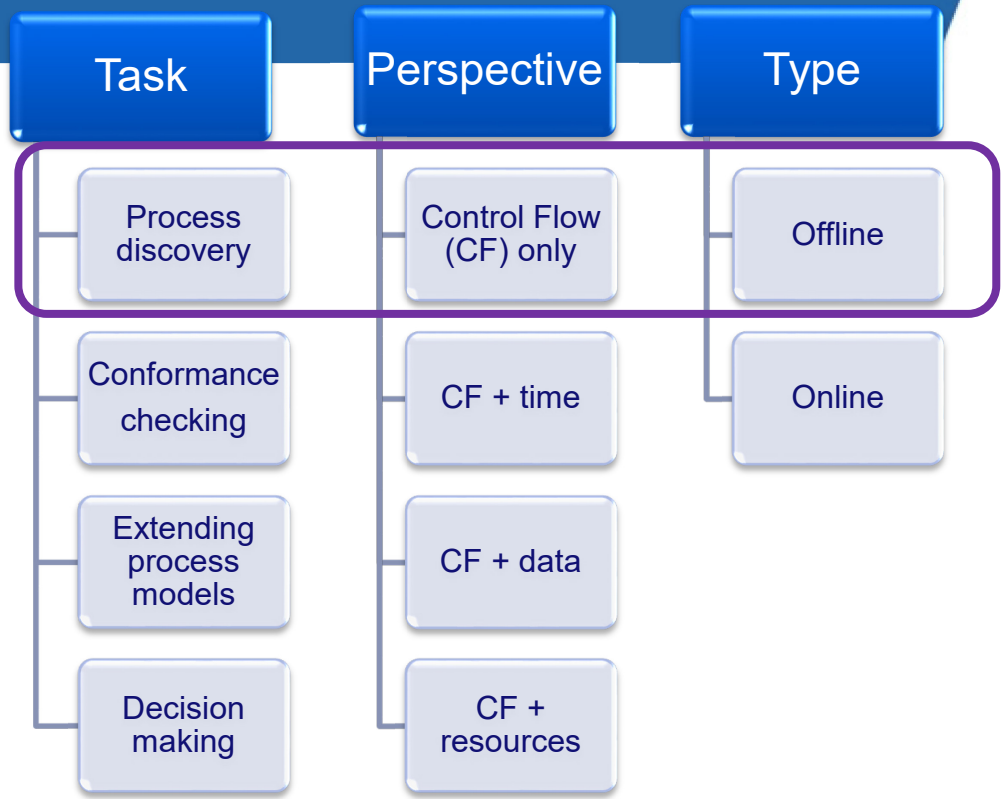


Chair of Process and Data Science

Taxonomy: Not just CF discovery!







Chair of Process and Data Science

ProM UITopia

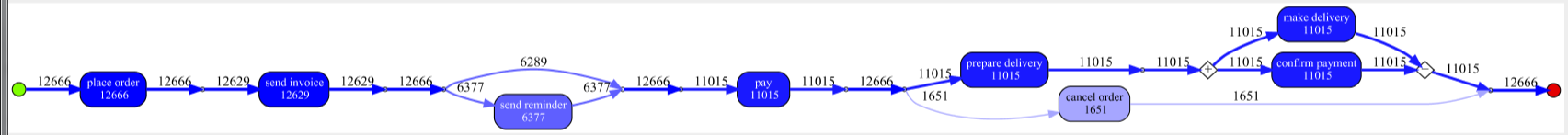
ProM 6

designed by fluxicon

Inductive visual Miner

Select visualisation ...

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



activities: paths:

1 0.8

Classifier concept:name

pre-mining filters

Miner miner (IMf)

edit model

Show paths

trace colouring

highlighting filters

traces

export model

export view

Highlighting all traces.

time: animation disabled

Powered by ProM

ProM UITopia

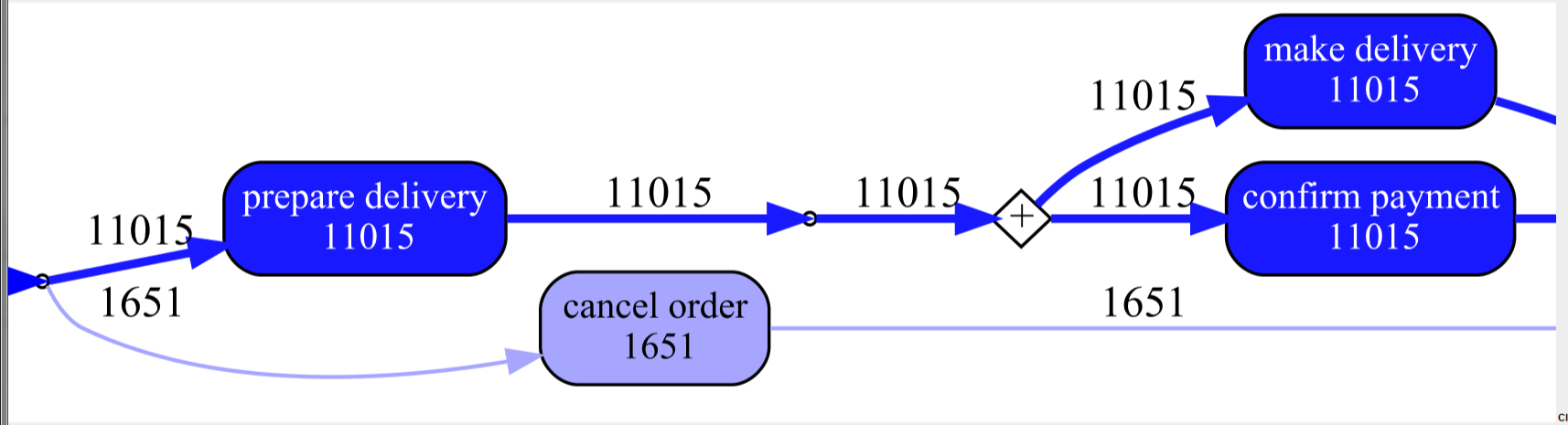
ProM 6

Inductive visual Miner

Select visualisation ...

designed by fluxicon

zooming in



activities: paths:

1 0.8

Classifier concept:name

pre-mining filters

Miner miner (IMf)

edit model

Show paths

trace colouring

highlighting filters

traces

export model

export view

Highlighting all traces.

time: animation disabled

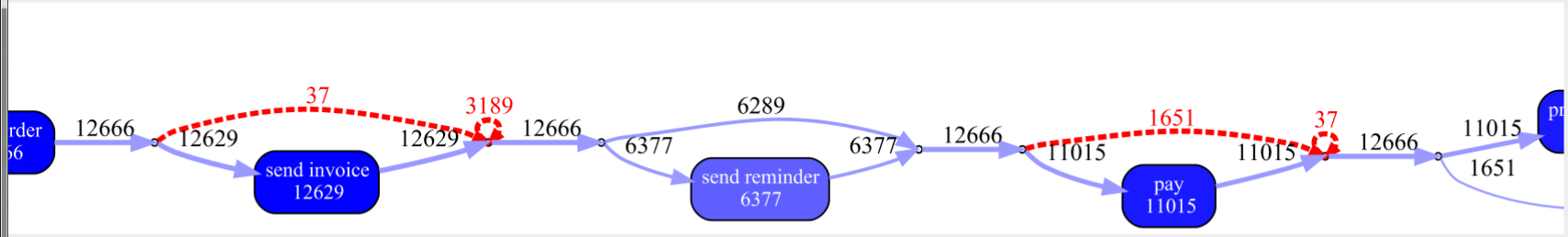
Powered by ProM

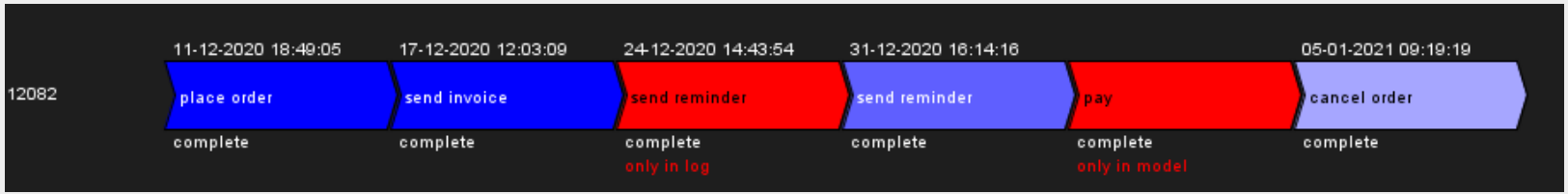
ProM UITopia

ProM 6

Inductive visual Miner

using conformance checking to see all deviations





happened in reality but not allowed by the model

required by the model but did not happen

Powered by ProM

activities: paths:

Classifier concept:name

pre-mining filters

Miner miner (IMf)

edit model

Show paths and deviations

trace colouring

highlighting filters

traces

export model

export view

Highlighting all traces.

time: animation disabled

done

ProM UITopia

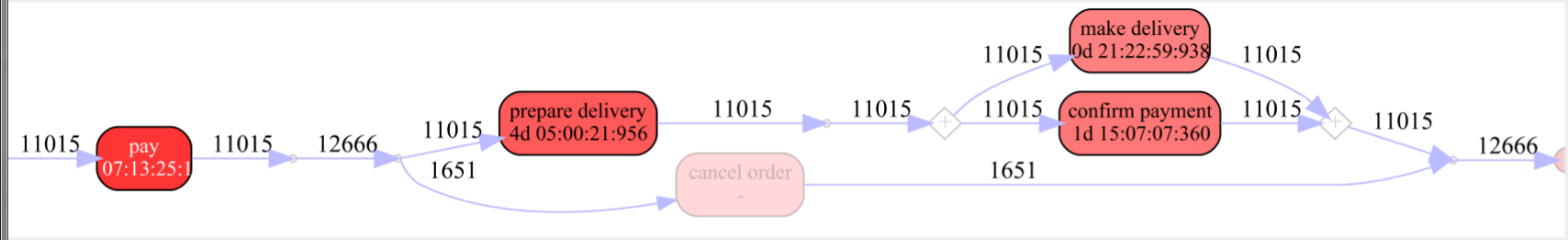
ProM 6

designed by fluxicon

Inductive visual Miner

Select visualisation ...

bottleneck analysis: enriching the model with performance information



activities: paths:

1 0.8

Classifier concept:name

pre-mining filters

Miner miner (IMf)

edit model

Show paths and sojourn ti

trace colouring

highlighting filters

traces

export model

export view

Highlighting all traces.

time: animation disabled

Powered by ProM

ProM UITopia

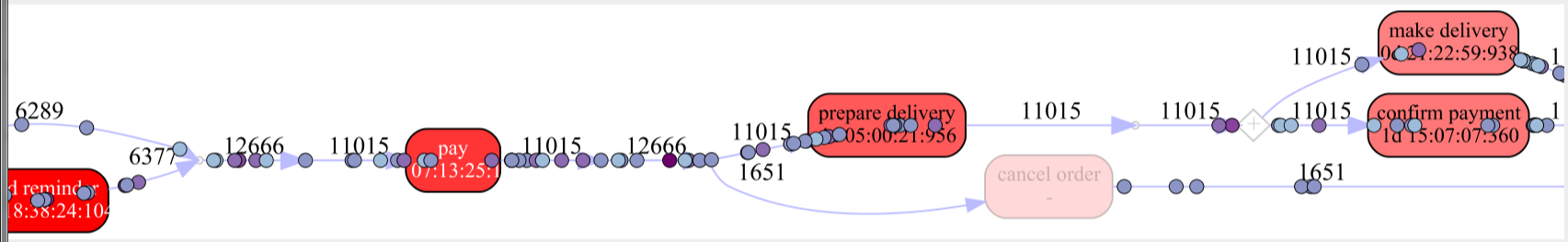
ProM 6

designed by fluxicon

Inductive visual Miner

Select visualisation ...

animating the event log showing real cases



activities:

paths:

1

0.8

Classifier concept:name

pre-mining filters

Miner miner (IMf)

edit model

Show paths and sojourn ti

trace colouring

highlighting filters

traces

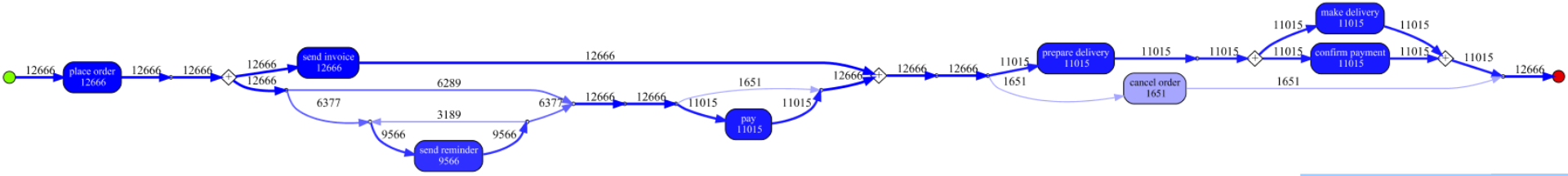
export model

export view

Highlighting all traces.

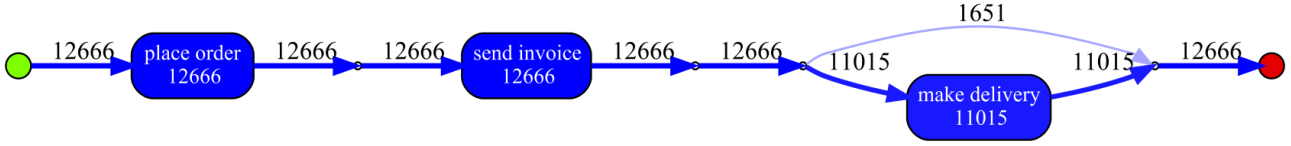
time: 07-08-2018 22:09:25:410

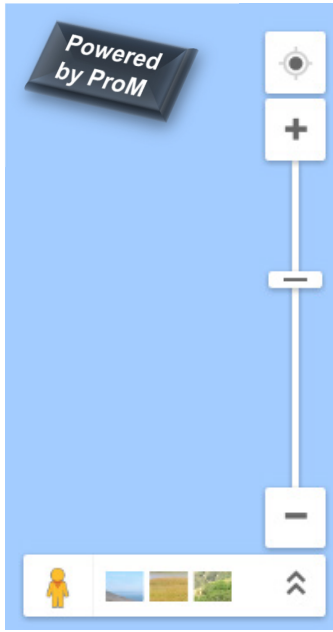
Powered by ProM

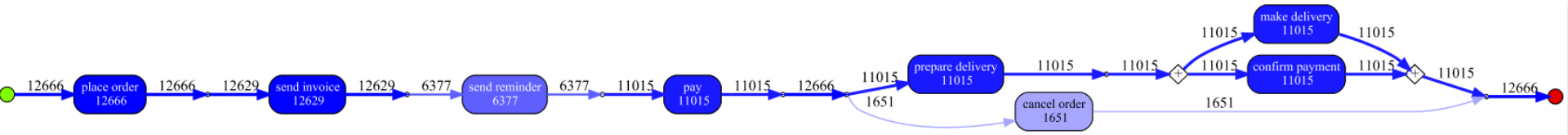


seamless abstraction

▶ one log many possible views

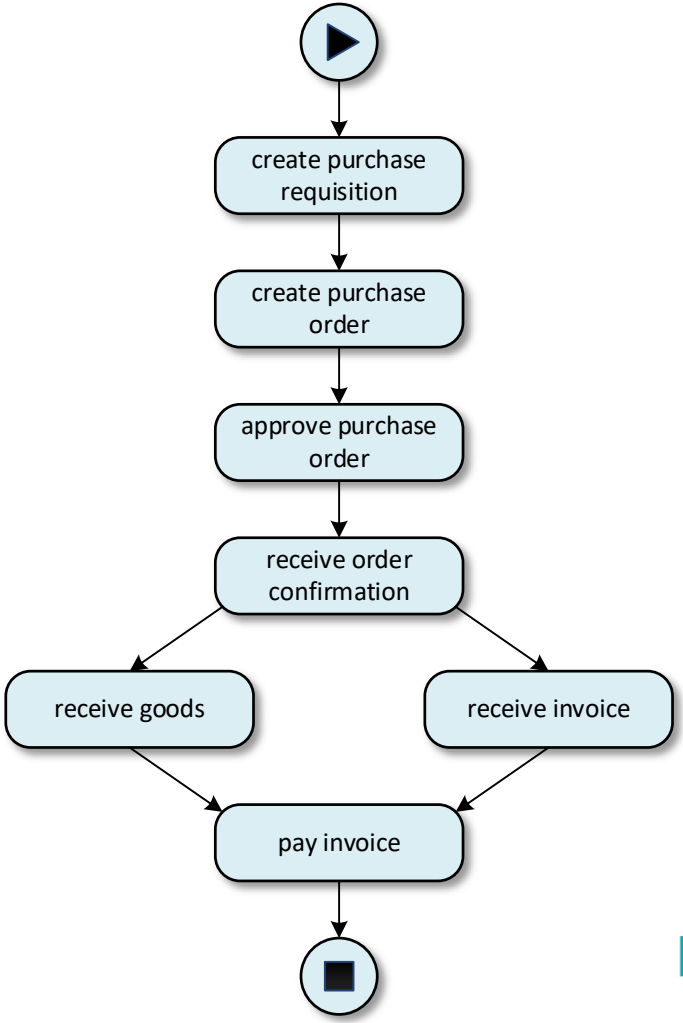






Purchase-to-Pay

- **Simple process found in almost any organization.**
- **Data available in e.g. SAP.**
- **Most cases follow the so-called “happy path”.**
- **80/20 rule applies.**



Chair of Process and Data Science

67k of 67k cases selected

100%

EDIT

Reset selections

Zoom

Activities

List view

100%

of activities

Reset

Less

More

Connections

List view

100%

of connections

Reset

Less

More

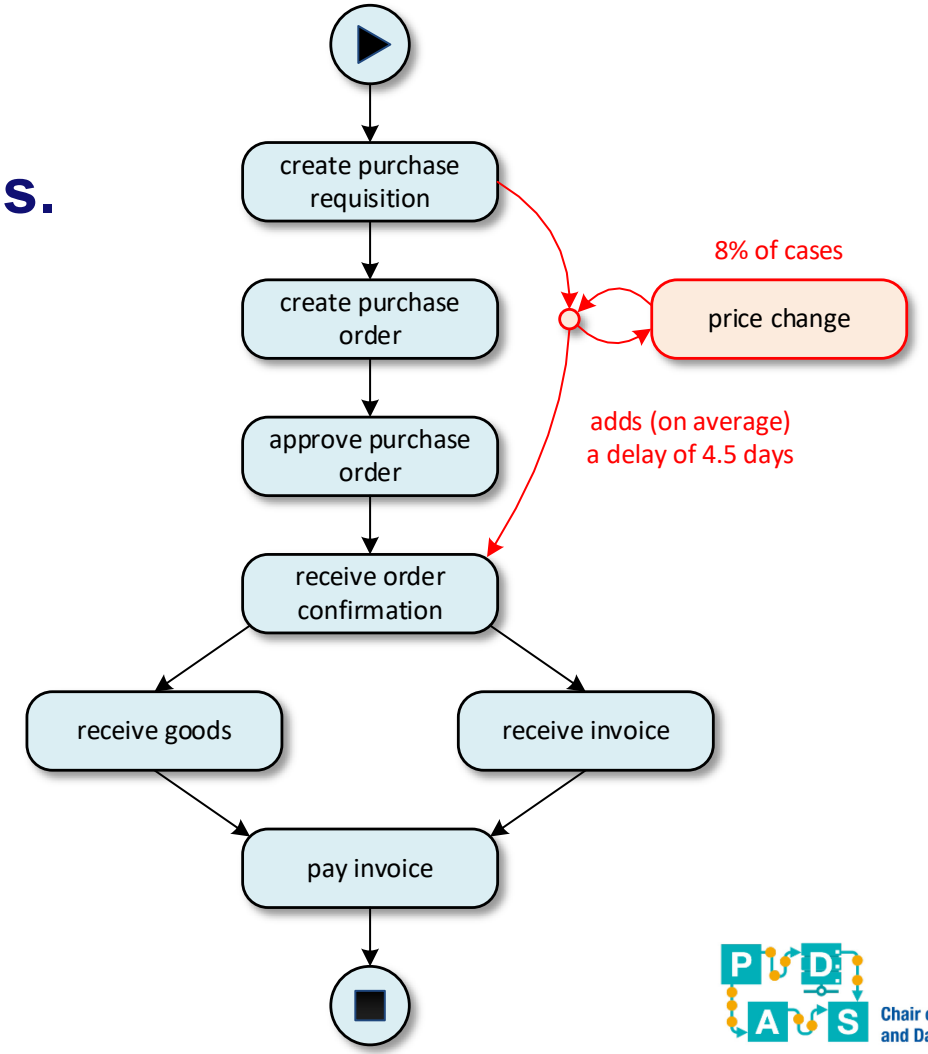
Fixed layout

Real process may look like this:
700,000 cases may exhibit 7,000 unique variants

Process Explorer | Case Explorer | Process overview | Purchasing Overview | Throughput Times | Changes | New App 7

Price changes

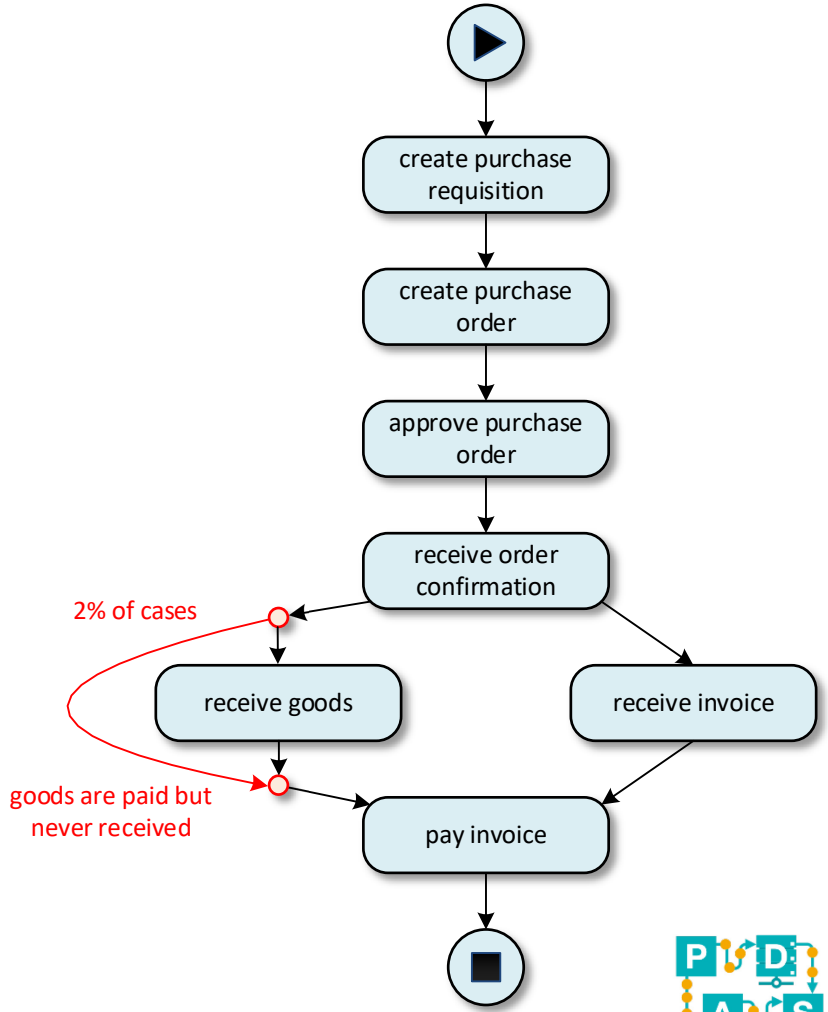
- **One of the many variations.**
- **Changing prices result in lots of extra work and significant delays.**



Chair of Process and Data Science

Pay before receipt

- **Goods are paid before they have been received.**
- **Goods arrived too late or not at all.**
- **May indicate fraud.**

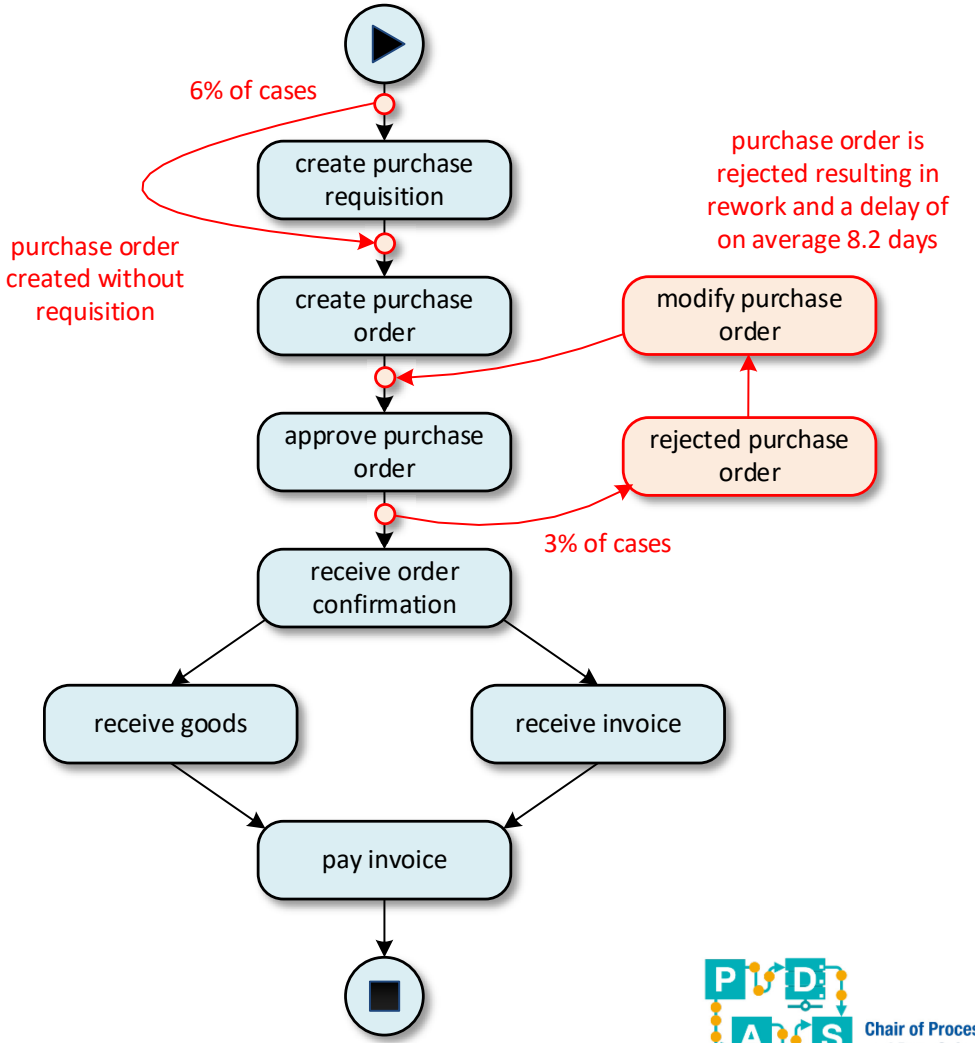


Chair of Process and Data Science

Two additional variations

- **Orders created without requisition.**
- **Rejected orders generating rework.**

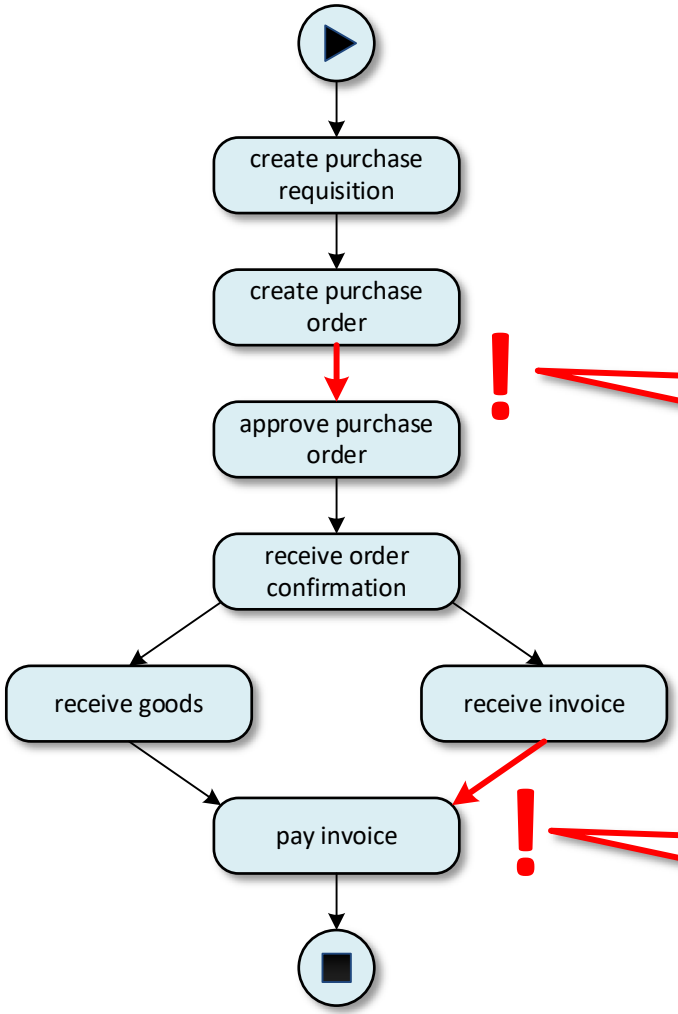
- **7000-4 = 6996 variants to go …**
- **Can be sorted based on frequency or impact.**



Chair of Process and Data Science

Performance problems

- **Delays inside the process.**
- **Excessive flow times.**
- **Not meeting Service Level Agreements (SLAs).**



Drill down to event data and uncover the root causes.

The approval of purchase orders takes too long (28% more than 10 days).

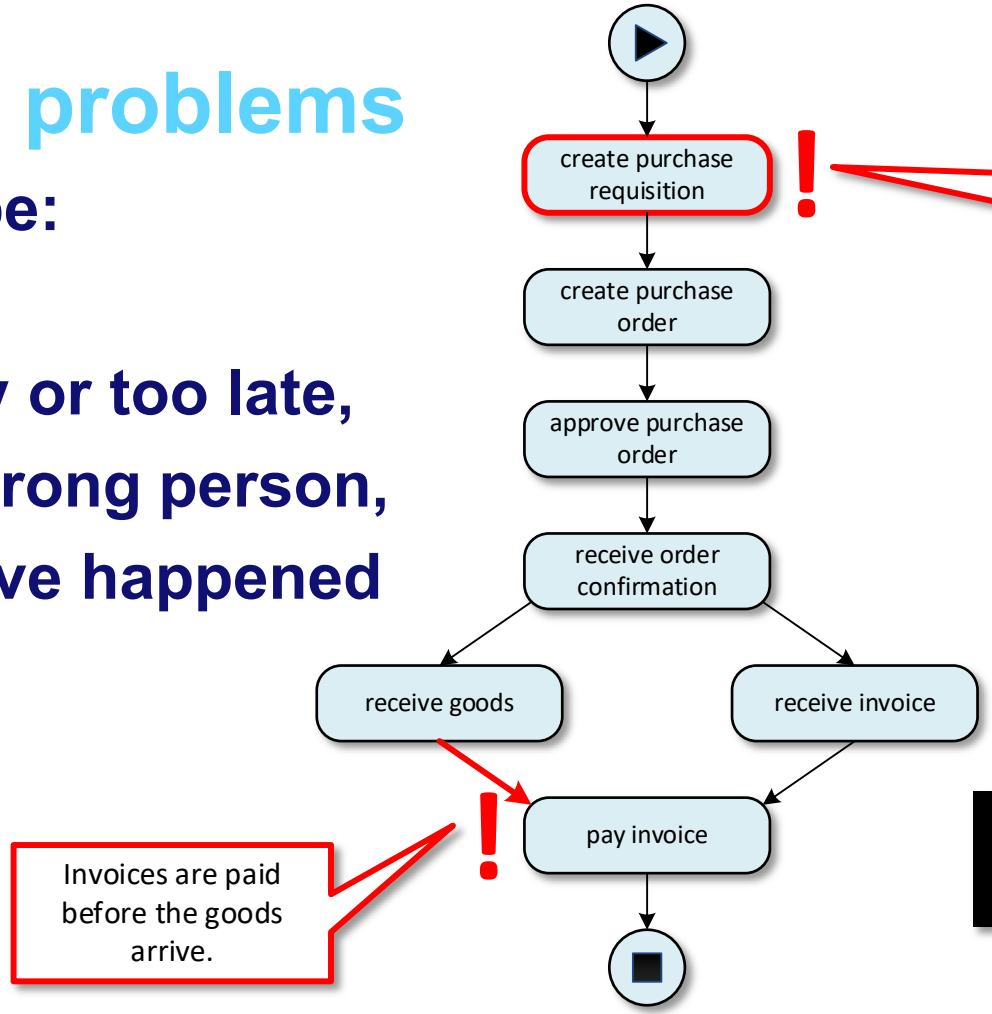
It takes to long to pay invoices resulting in complaints and fines (15% more than 3 weeks).

Chair of Process and Data Science

Compliance problems

Activities may be:

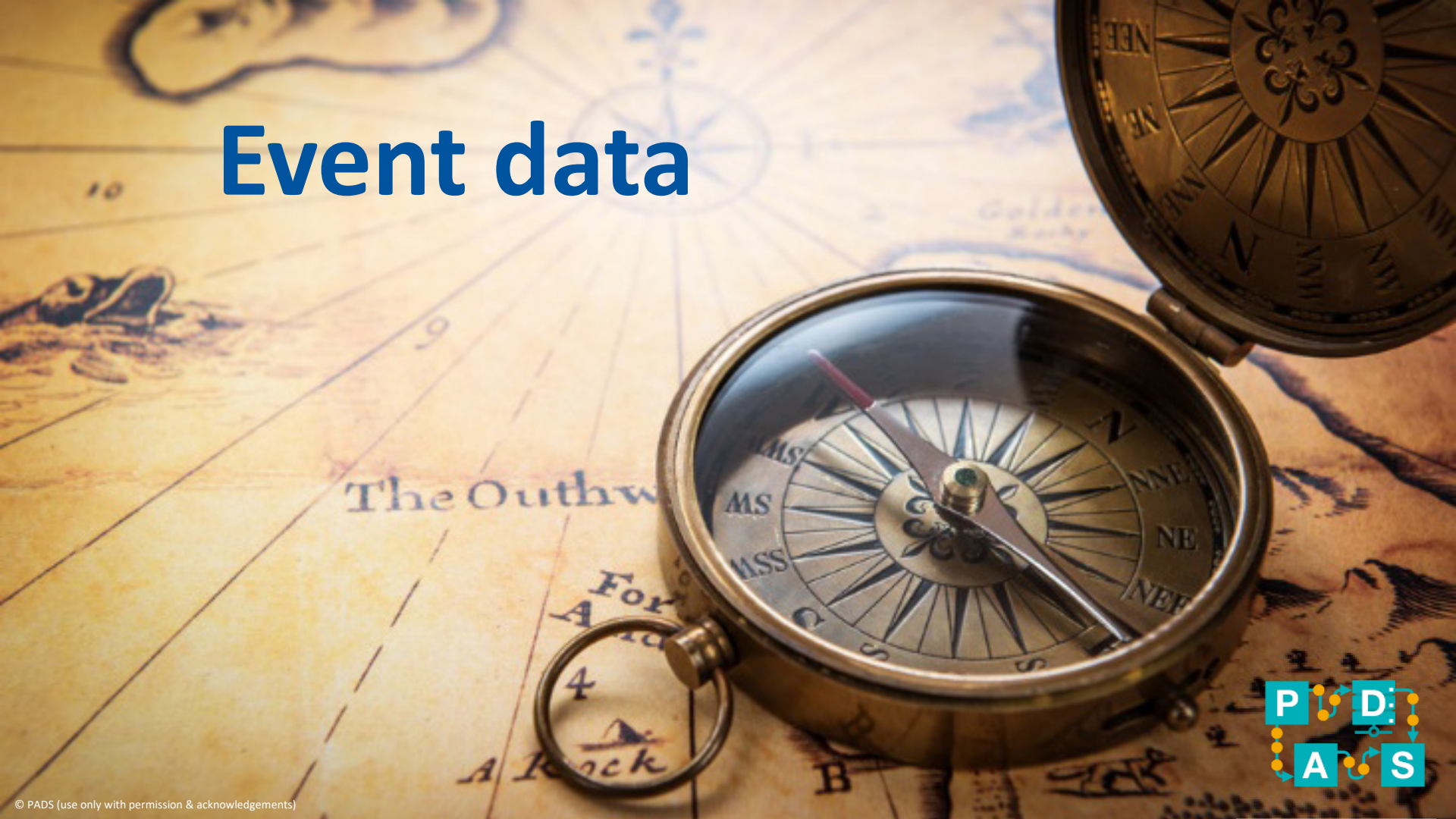
- **skipped,**
- **done too early or too late,**
- **done by the wrong person,**
- **should not have happened at all.**



Drill down to event data and uncover the root causes.

Chair of Process and Data Science

Event data



Example 1

every row is an event (here: an exam attempt)

student name	course name	exam date	mark
Peter Jones	Business Information systems	16-1-2014	8
Sandy Scott	Business Information systems	16-1-2014	5
Bridget White	Business Information systems	16-1-2014	9
John Anderson	Business Information systems	16-1-2014	8
Sandy Scott	BPM Systems	17-1-2014	7
Bridget White	BPM Systems	17-1-2014	8
Sandy Scott	Process Mining	20-1-2014	5
Bridget White	Process Mining	20-1-2014	9
John Anderson	Process Mining	20-1-2014	8
…	…	…	…

case id | **activity name** | **timestamp** | **other data**

Example 2

order number	activity	timestamp	user	product	quantity
9901	register order	22-1-2014@09.15	Sara Jones	iPhone5S	1
9902	register order	22-1-2014@09.18	Sara Jones	iPhone5S	2
9903	register order	22-1-2014@09.27	Sara Jones	iPhone4S	1
9901	check stock	22-1-2014@09.49	Pete Scott	iPhone5S	1
9901	ship order	22-1-2014@10.11	Sue Fox	iPhone5S	1
9903	check stock	22-1-2014@10.34	Pete Scott	iPhone4S	1
9901	handle payment	22-1-2014@10.41	Carol Hope	iPhone5S	1
9902	check stock	22-1-2014@10.57	Pete Scott	iPhone5S	2
9902	cancel order	22-1-2014@11.08	Carol Hope	iPhone5S	2
…	…	…	…	…	…

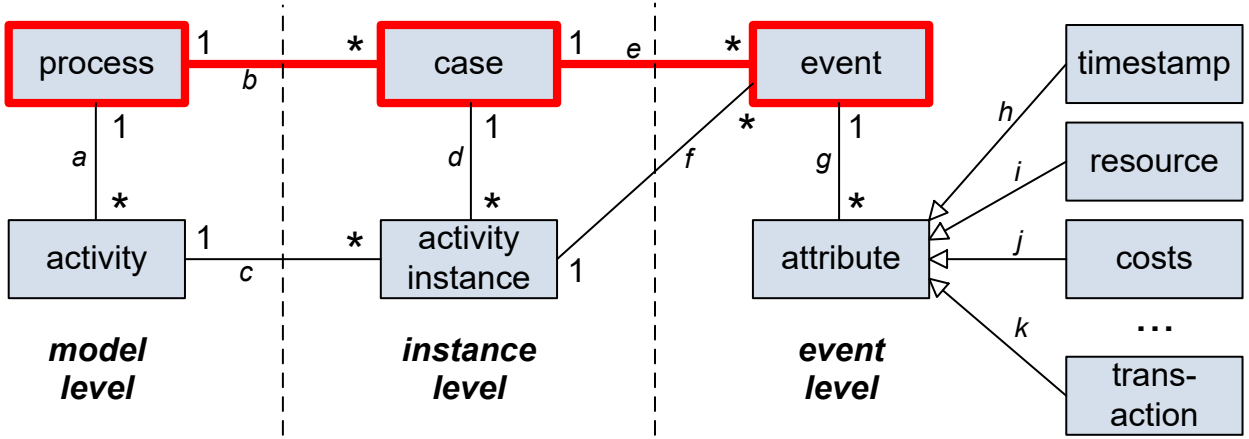
case id **activity name** **timestamp** **resource** **other data**

Example 3

patient	activity	timestamp	doctor	age	cost
5781	make X-ray	23-1-2014@10.30	Dr. Jones	45	70.00
5541	blood test	23-1-2014@10.18	Dr. Scott	61	40.00
5833	blood test	23-1-2014@10.27	Dr. Scott	24	40.00
5781	blood test	23-1-2014@10.49	Dr. Scott	45	40.00
5781	CT scan	23-1-2014@11.10	Dr. Fox	45	1200.00
5833	surgery	23-1-2014@12.34	Dr. Scott	24	2300.00
5781	handle payment	23-1-2014@12.41	Carol Hope	45	0.00
5541	radiation therapy	23-1-2014@13.57	Dr. Jones	61	140.00
5541	radiation therapy	23-1-2014@13.08	Dr. Jones	61	140.00
…	…	…	…	…	…

case id | **activity name** | **timestamp** | **resource** | **other data**

Process – case - event



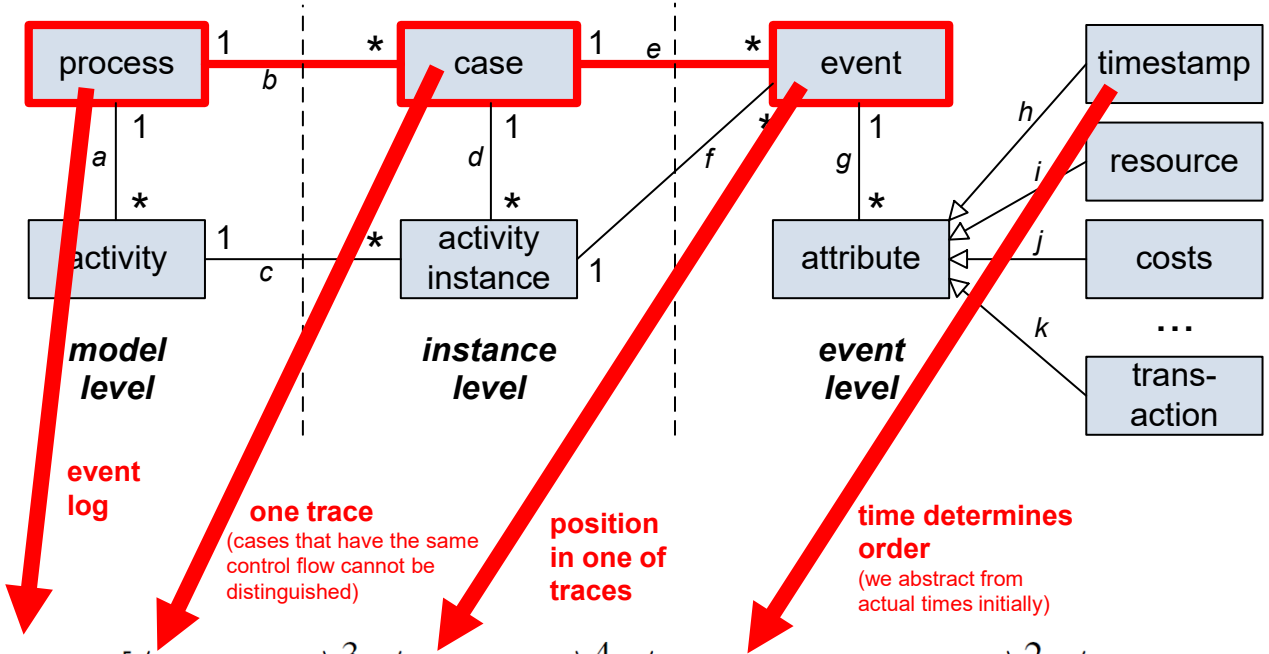
XES
Extensible Event Stream

$$L \in \mathfrak{B}(A^*)$$

$$L_2 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^4, \langle a,b,c,e,f,b,c,d\rangle^2, \langle a,b,c,e,f,c,b,d\rangle,$$
$$\langle a,c,b,e,f,b,c,d\rangle^2, \langle a,c,b,e,f,b,c,e,f,c,b,d\rangle]$$

Chair of Process and Data Science

Process – case - event



XES
Extensible Event Stream

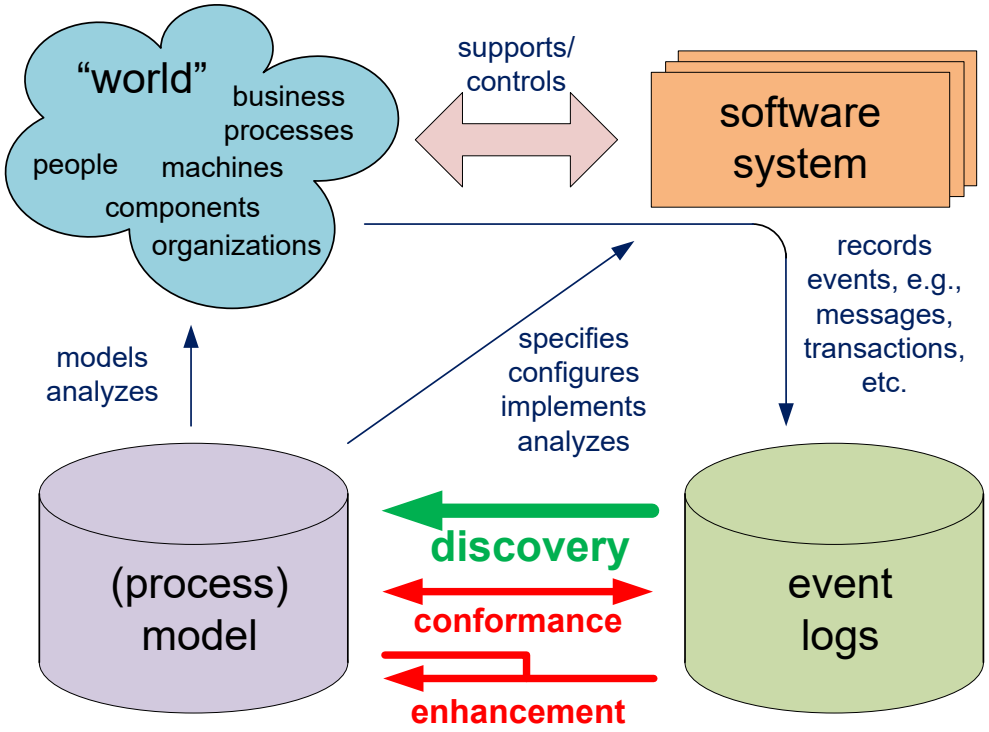
$$L \in \mathfrak{B}(A^*)$$

$$L_2 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^4, \langle a,b,c,e,f,b,c,d\rangle^2, \langle a,b,c,e,f,c,b,d\rangle,$$
$$\langle a,c,b,e,f,b,c,d\rangle^2, \langle a,c,b,e,f,b,c,e,f,c,b,d\rangle]$$

Chair of Process and Data Science

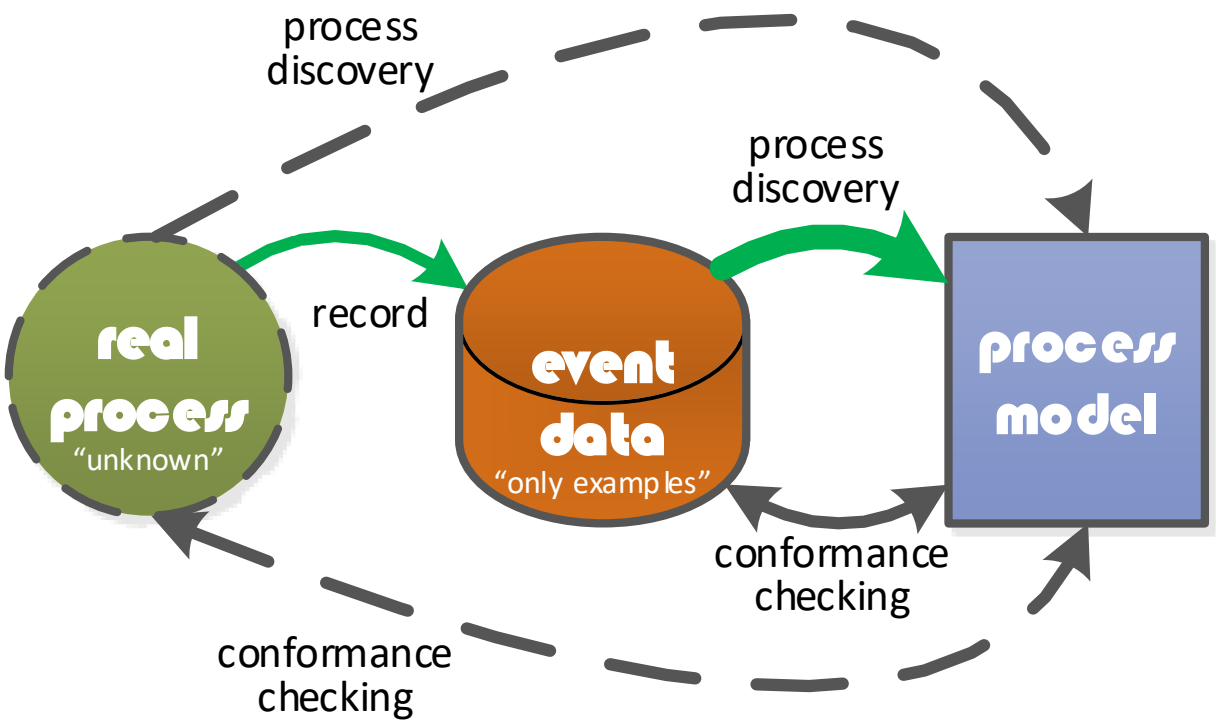
Process discovery

Process discovery



Chair of Process and Data Science

Process discovery



No negative examples
(cannot see what cannot happen)

Almost vs poorly fitting traces

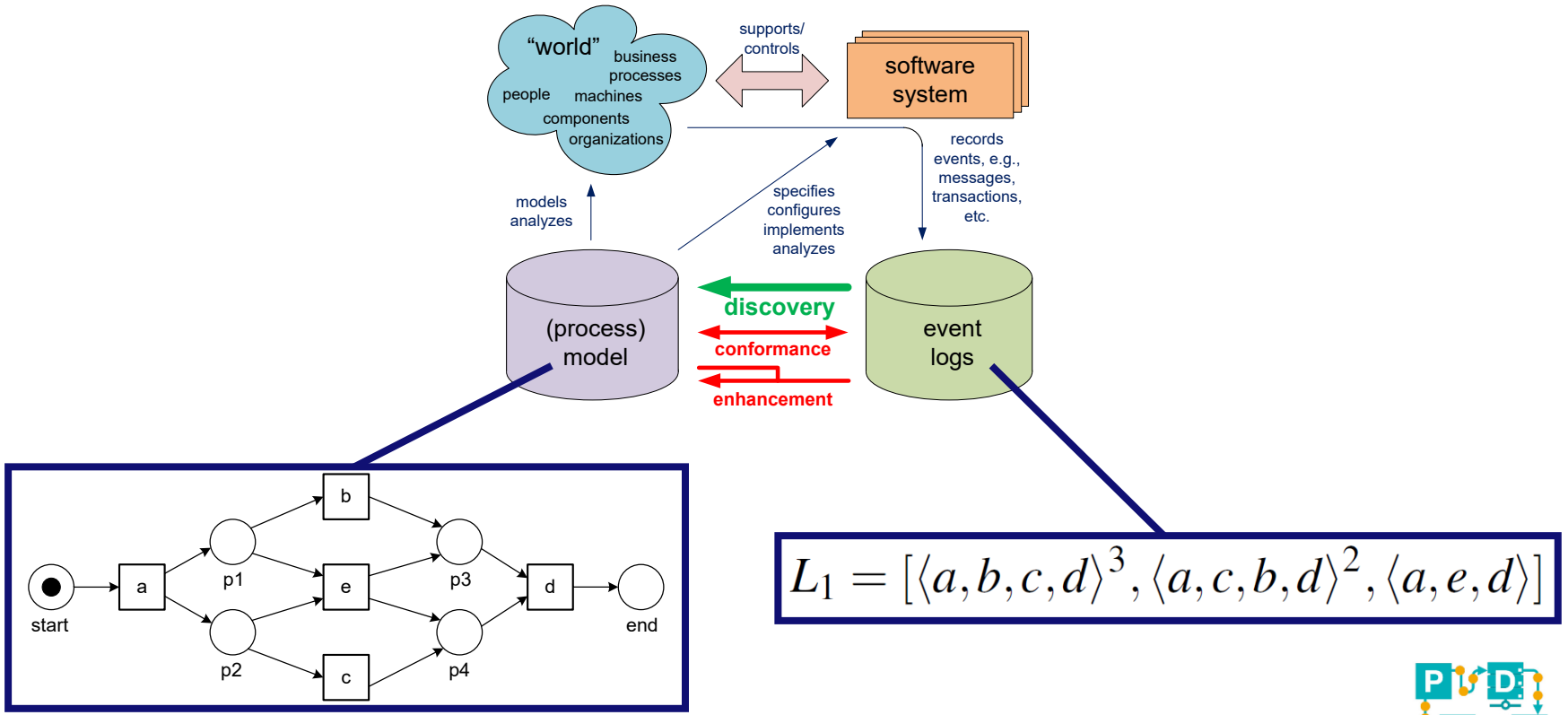
Log contains only a fraction of possible traces

In case of loops often infinitely many possible traces

Murphy's law for process mining
(anything is possible, so probabilities matter)

Chair of Process and Data Science

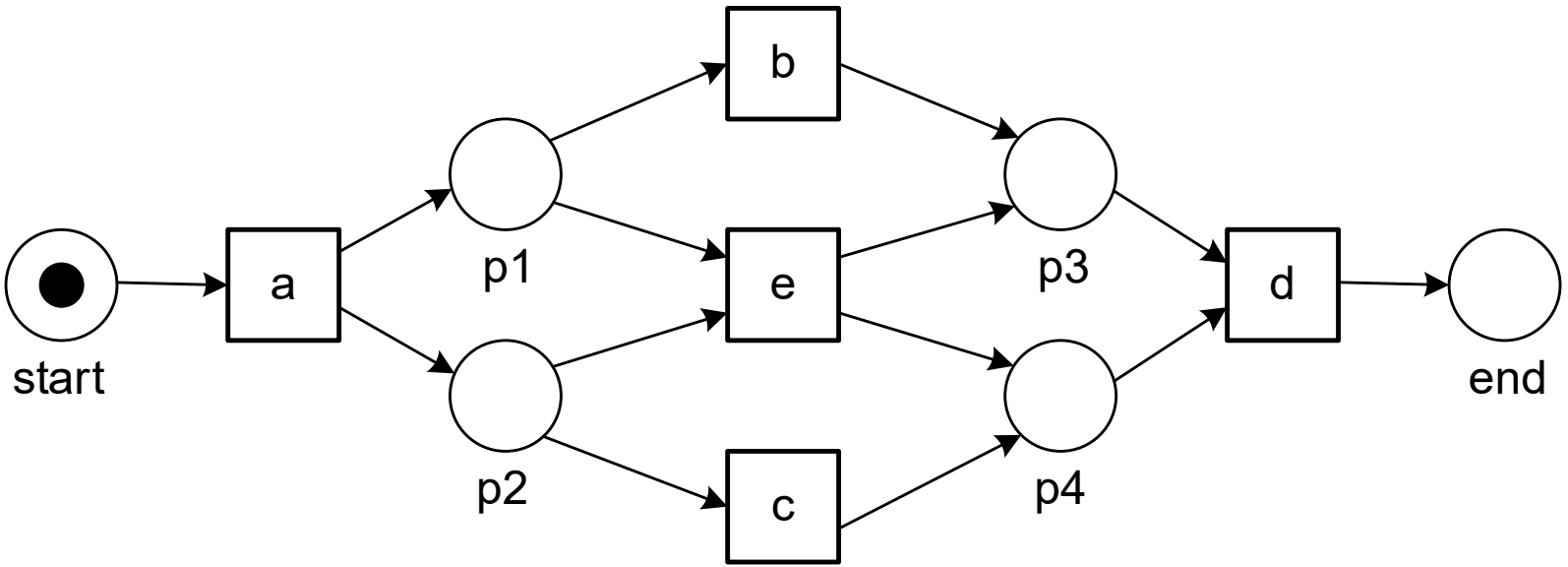
Let’s consider the simplest setting possible



Chair of Process and Data Science

Process discovery: Bottom-up

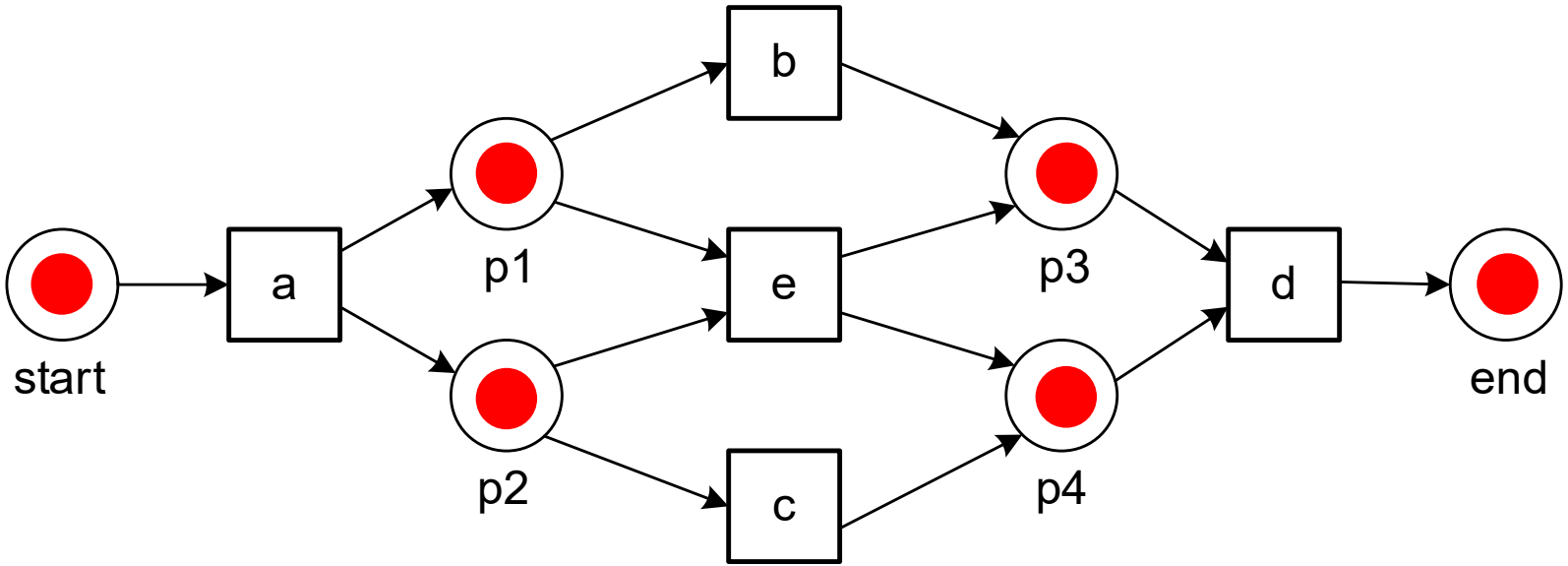
Let's use accepting Petri nets to explain bottom-up approaches



Initial marking [start], final marking [end].

Chair of Process and Data Science

Example trace (1/3)

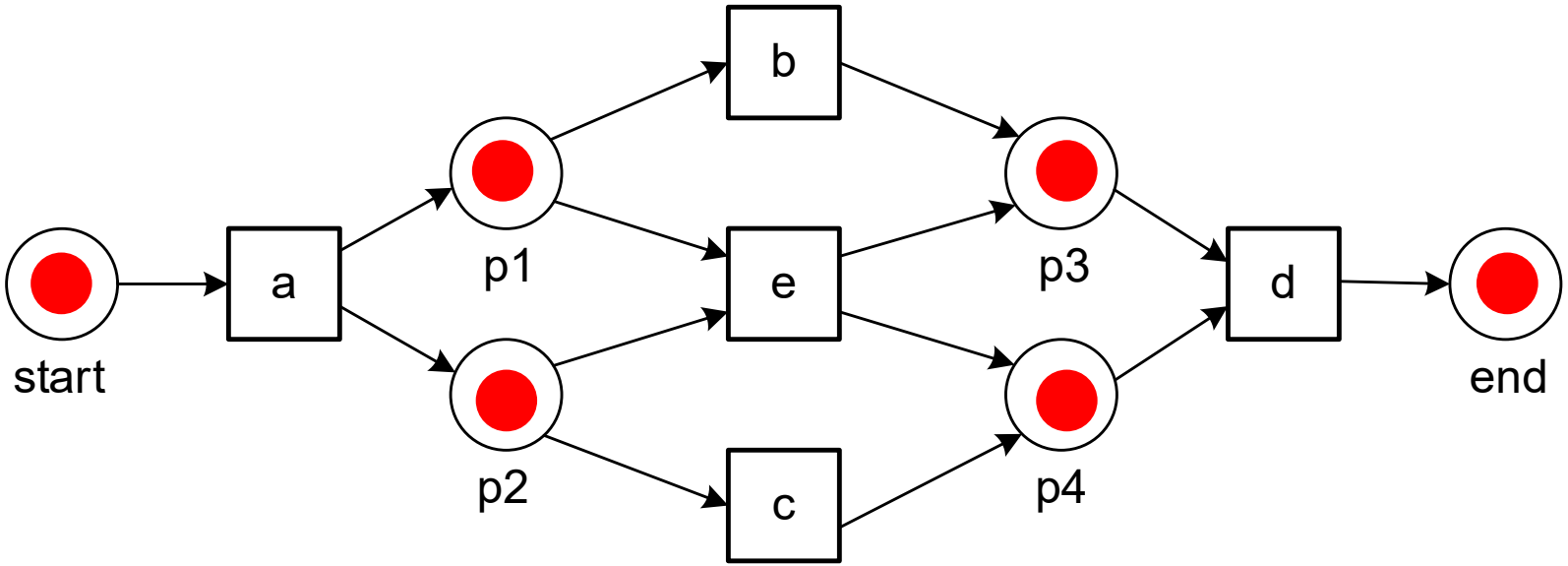


a b c d

Initial marking [start], final marking [end].

Chair of Process and Data Science

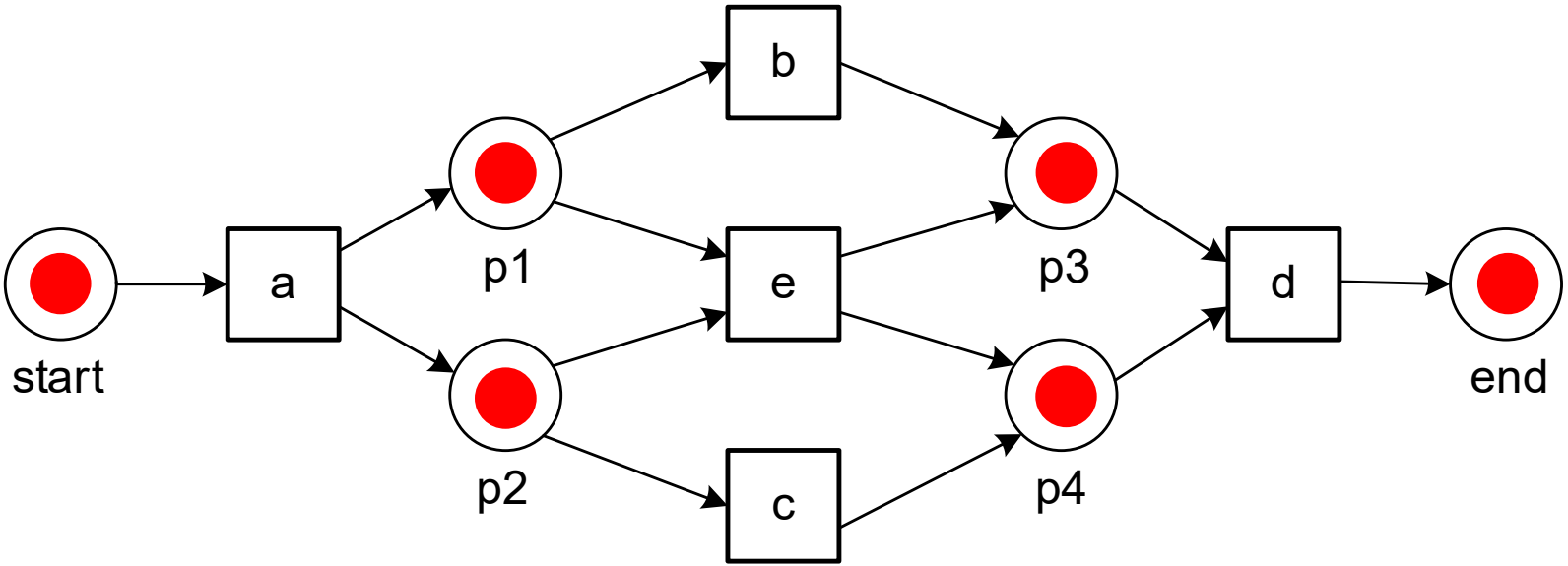
Example trace (2/3)



a c b d

Initial marking [start], final marking [end].

Example trace (3/3)



a e d

Initial marking [start], final marking [end].

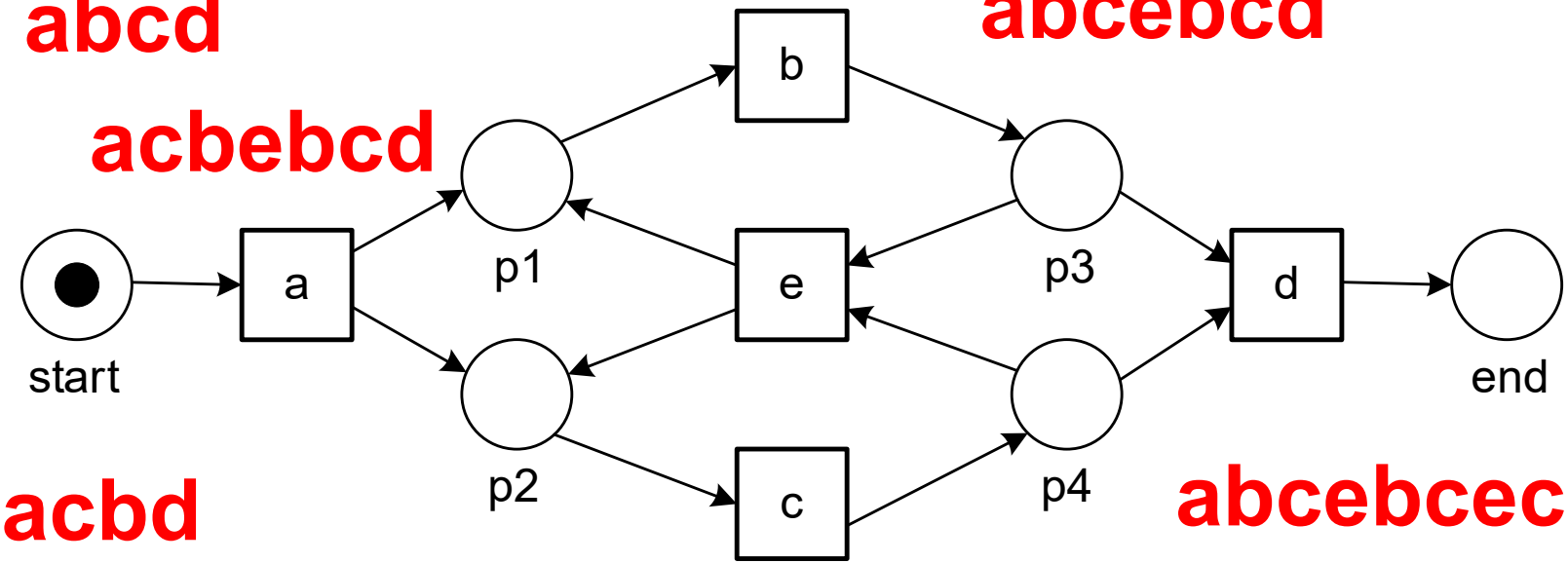
Chair of Process and Data Science

Another example (infinitely many possible traces)

abcd

abcebcd

acbebcd



acbd

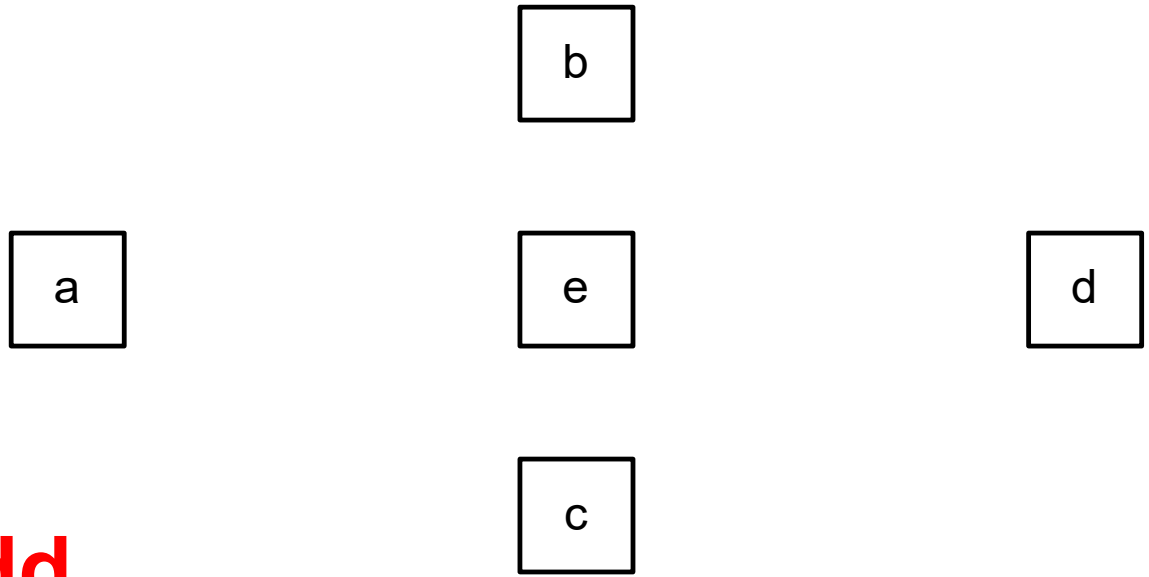
abcebcecbd

acbebcebcebcecbd

Initial marking [start], final marking [end].

Chair of Process and Data Science

The essence of Petri nets: Model allows for any trace!



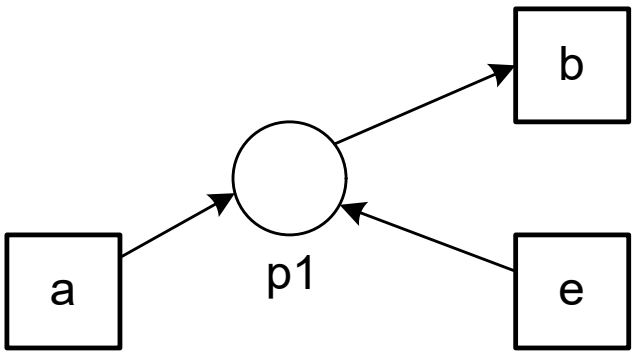
aaaddd

eceaa

Initial marking [ ], final marking [ ].

Places are constraints

- Place cannot “go negative”.
- Should end up with the right number of tokens.



d

c

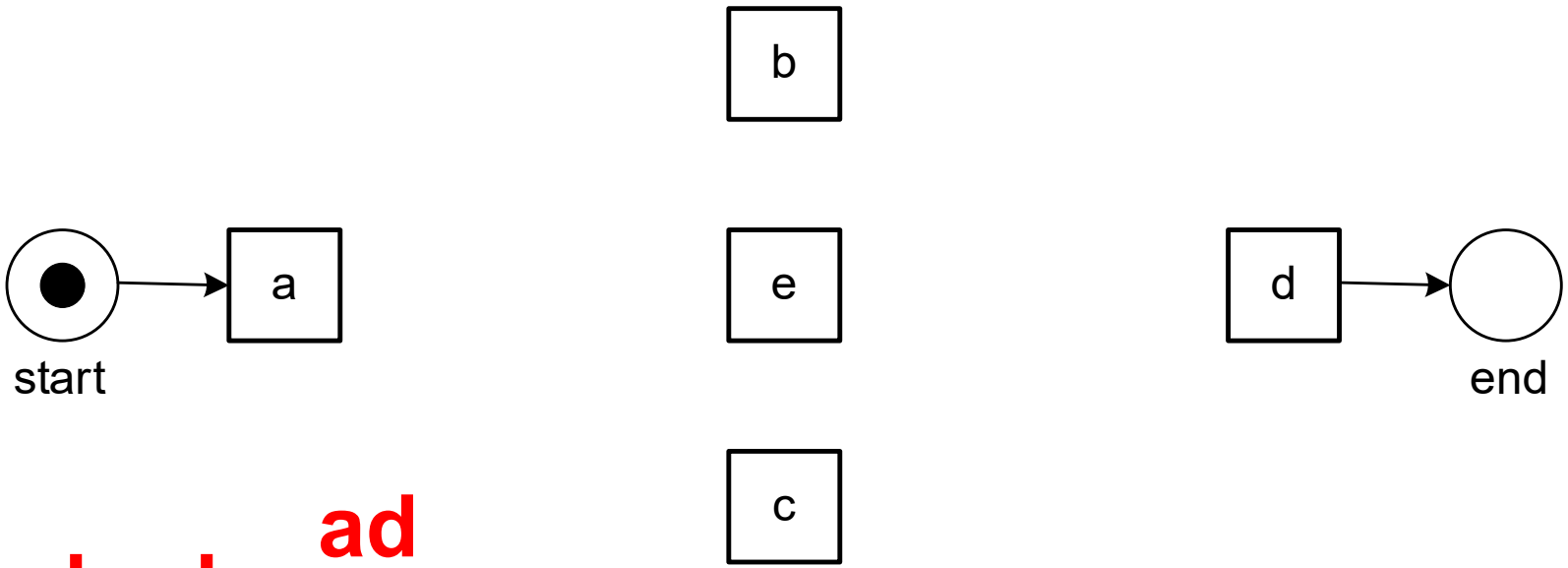
dcd

abc

abebd

Initial marking [ ], final marking [ ].

Places are constraints



abcd

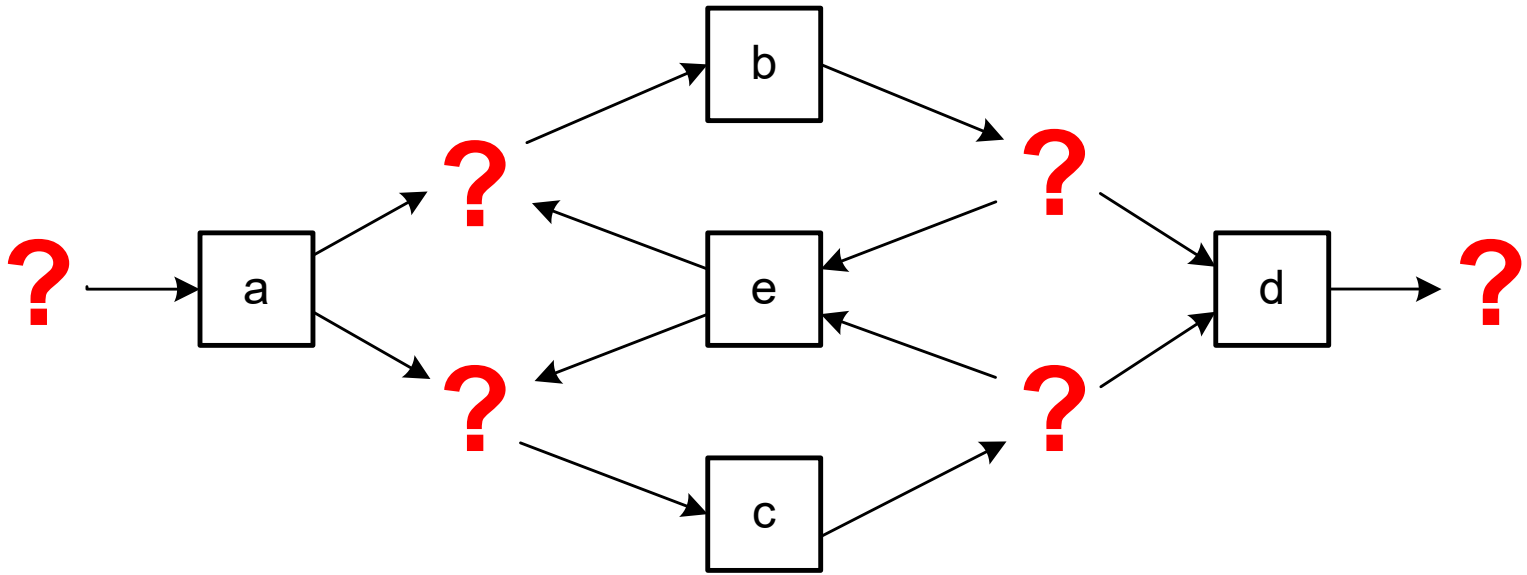
ad

da

dccca

Initial marking [start], final marking [end].

Hence, process discovery is just finding places

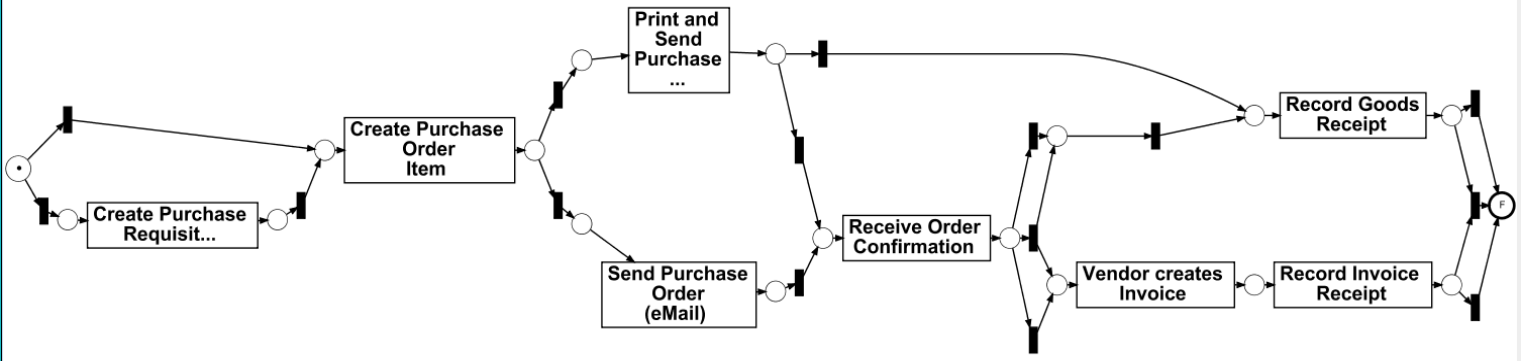


Initial marking [ ], final marking [ ].

Many approaches possible

- **Heuristics** that provide only guarantees for limited classes of model (e.g., Alpha algorithm and heuristic miner).
- Approaches that **formally guarantee perfect replayability** of the event log (e.g., state-based regions).
- **Genetic** and other **evolutionary** approaches (very flexible).
- **Optimization**-based approaches that turn discovery into an optimization problem (e.g., ILP miner).
- Brute-force approaches that exploit **monotonicity** properties (apriory-style algorithms).

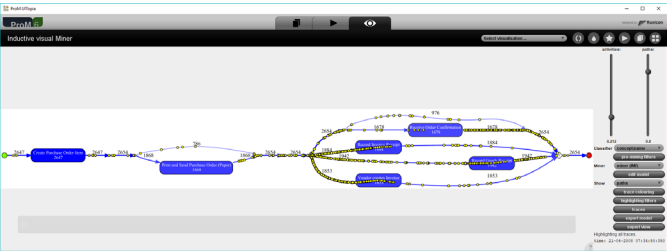
Example: Heuristic miner applied to SAP data



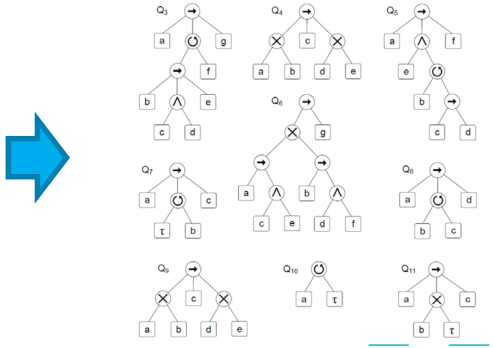
Process discovery: Top-down

Example of a top-down discovery approach

- **Inductive mining** (work done by Sander Leemans and extended by Maikel Leemans).
- **Family** of approaches with different **guarantees** and **scalability** characteristics (all can ensure replayability of the whole event log).



$$
\begin{aligned}
L_3 &= [\langle a,b,c,d,e,f,b,d,c,e,g\rangle, \langle a,b,d,c,e,g\rangle^2, \\
&\quad \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g\rangle] \\
L_4 &= [\langle a,c,d\rangle^{45}, \langle b,c,d\rangle^{42}, \langle a,c,e\rangle^{38}, \langle b,c,e\rangle^{22}] \\
L_5 &= [\langle a,b,e,f\rangle^2, \langle a,b,e,c,d,b,f\rangle^3, \langle a,b,c,e,d,b,f\rangle^2, \langle a,b,c,d,e,b,f\rangle^4, \\
&\quad \langle a,e,b,c,d,b,f\rangle^3] \\
L_6 &= [\langle a,c,e,g\rangle^2, \langle a,e,c,g\rangle^3, \langle b,d,f,g\rangle^2, \langle b,f,d,g\rangle^4] \\
L_7 &= [\langle a,c\rangle^2, \langle a,b,c\rangle^3, \langle a,b,b,c\rangle^2, \langle a,b,b,b,b,c\rangle] \\
L_8 &= [\langle a,b,d\rangle^3, \langle a,b,c,b,d\rangle^2, \langle a,b,c,b,c,b,d\rangle] \\
L_9 &= [\langle a,c,d\rangle^{45}, \langle b,c,e\rangle^{42}] \\
L_{10} &= [\langle a,a\rangle^{55}] \\
L_{11} &= [\langle a,b,c\rangle^{20}, \langle a,c\rangle^{30}]
\end{aligned}
$$

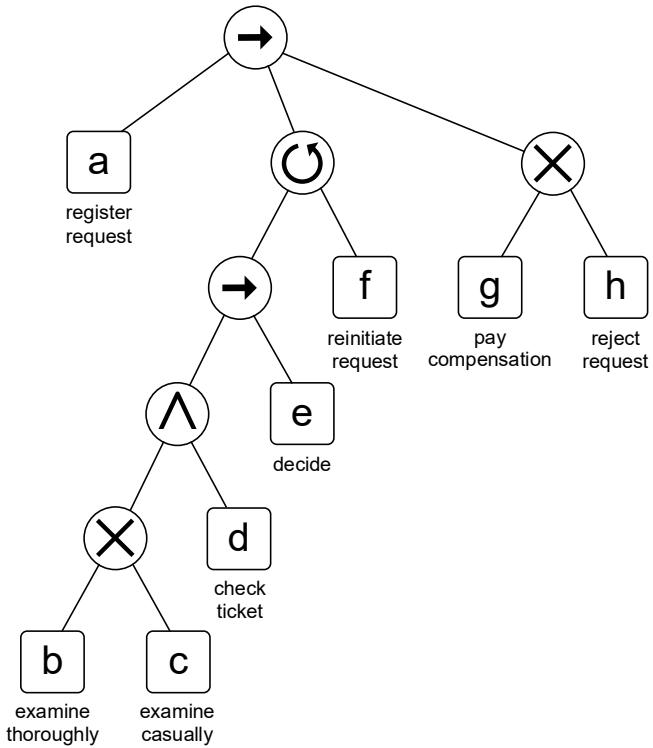


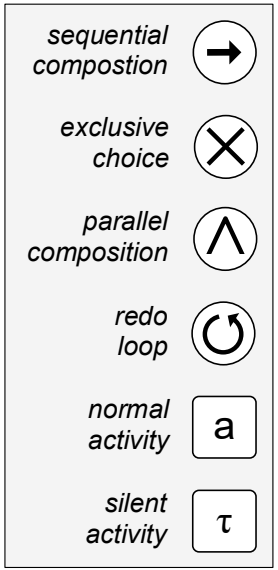
Input: Event log (simplified)

a b c d
a b c d
a b c d
a c b d
a c b d
a c b d
a c b d
a b c e f b c d
a b c e f b c d
a c b e f b c d
a c b e f b c d
a b c e f c b d
a c b e f b c e f c b d

3x a b c d
4x a c b d
2x a b c e f b c d
2x a c b e f b c d
1x a b c e f c b d
1x a c b e f b c e f c b d

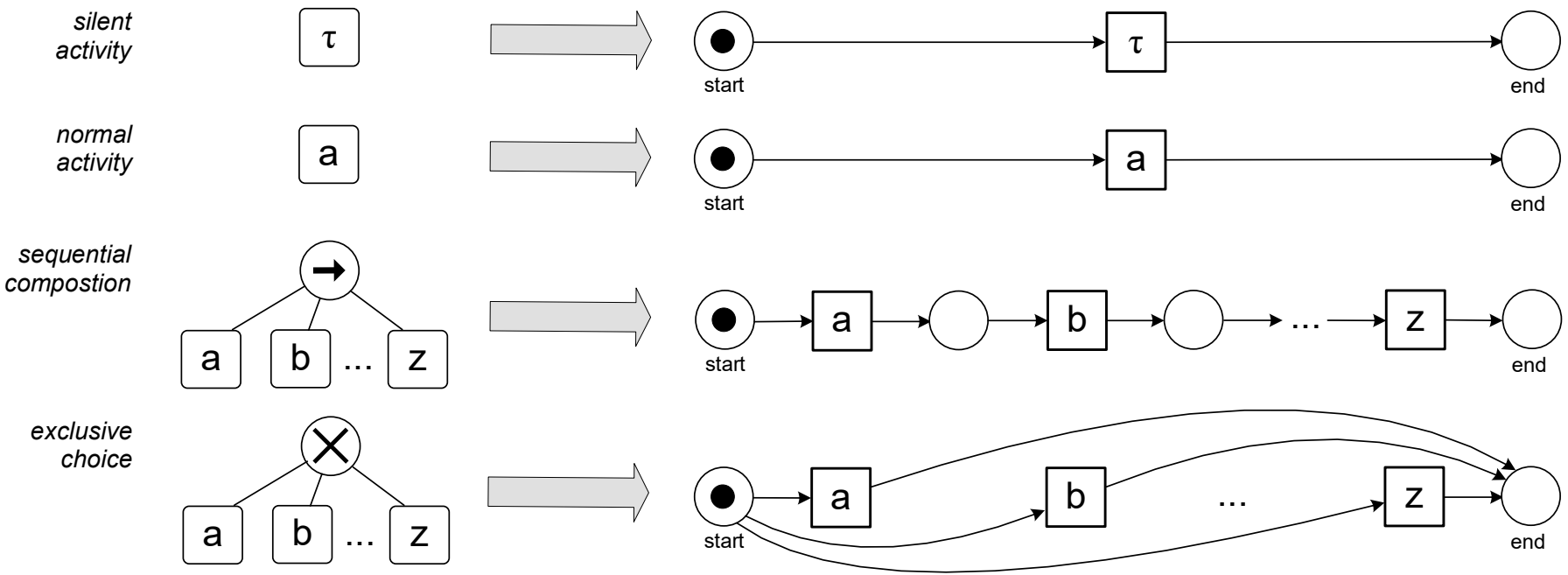
Output: Process Tree





Chair of Process and Data Science

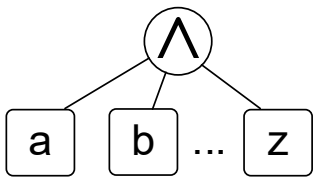
Output: Process Tree

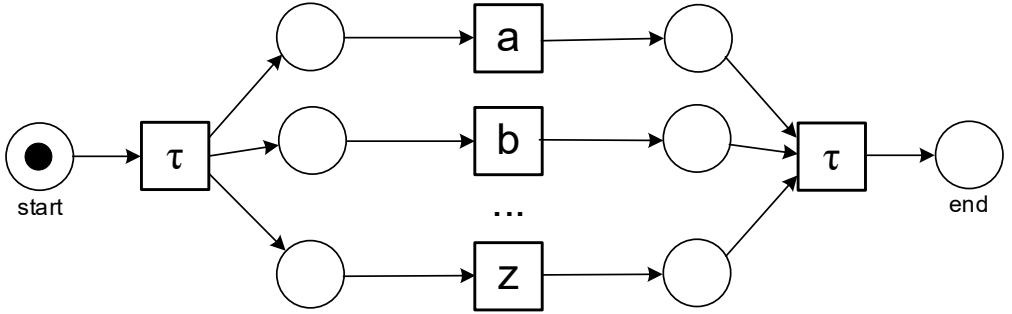


Chair of Process and Data Science

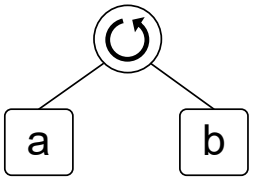
Output: Process Tree

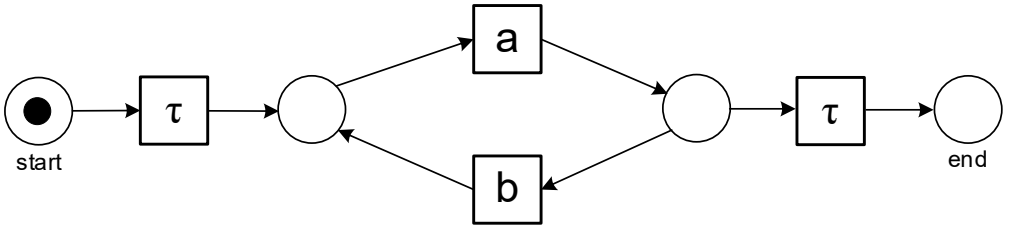
parallel composition





redo loop





Chair of Process and Data Science

Directly-follows graph based on event log

3x a b c d

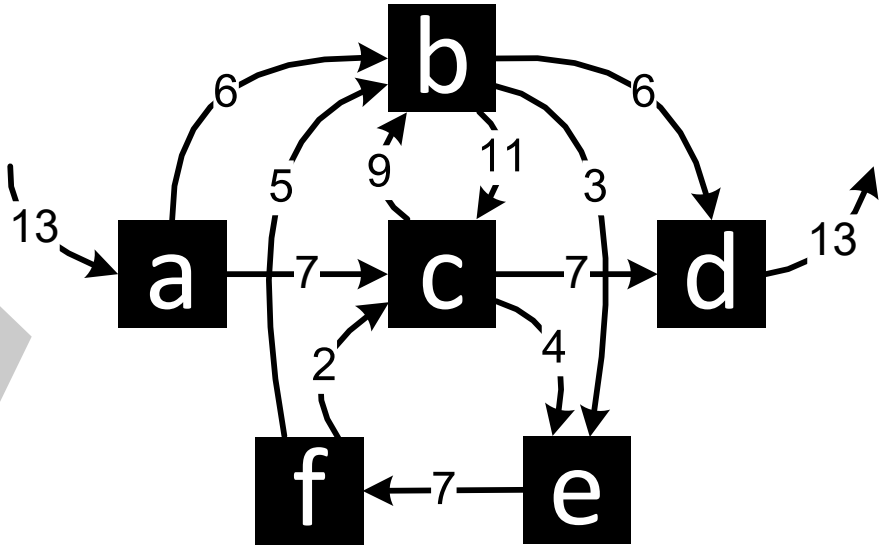
4x a c b d

2x a b c e f b c d

2x a c b e f b c d

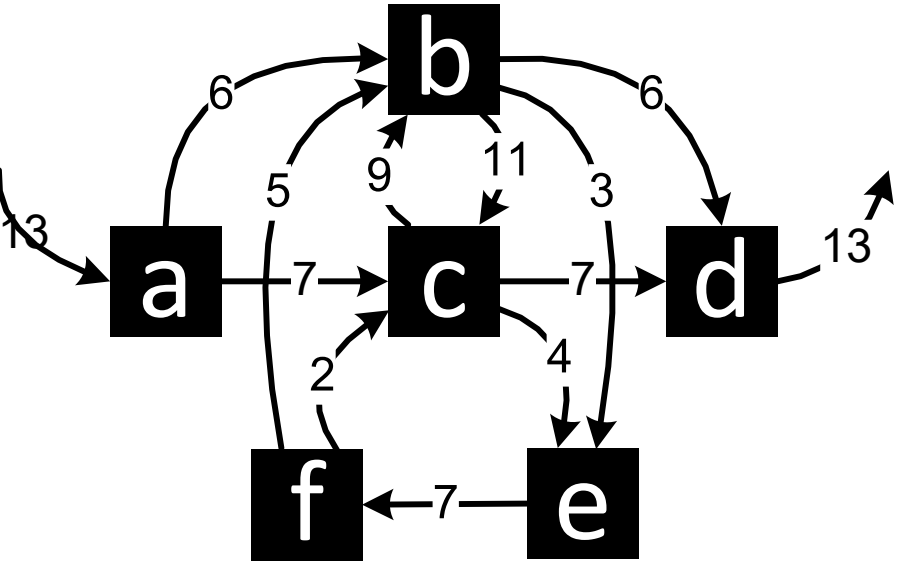
1x a b c e f c b d

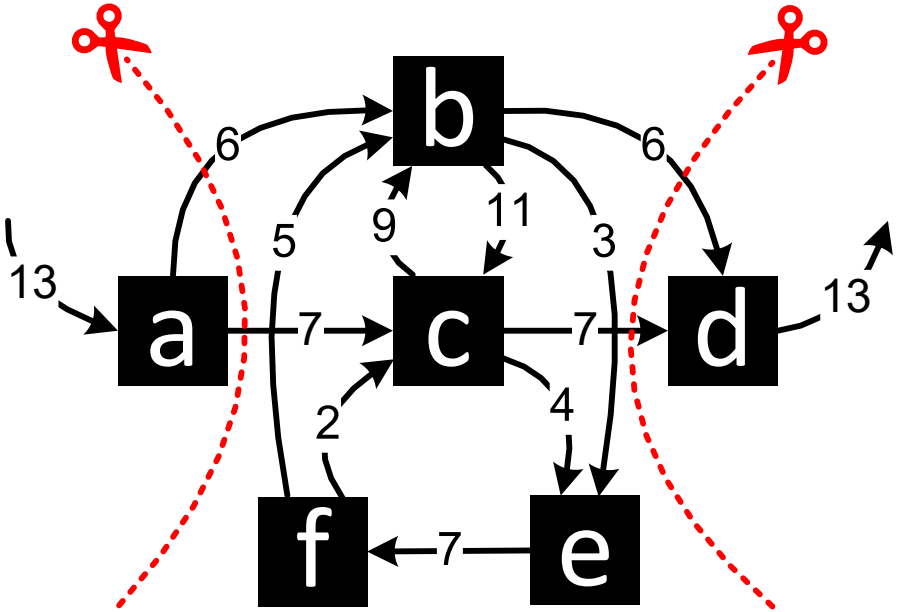
1x a c b e f b c e f c b d



Chair of Process and Data Science

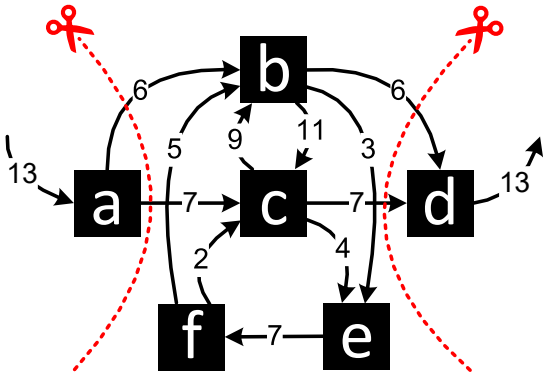
Sequence cut

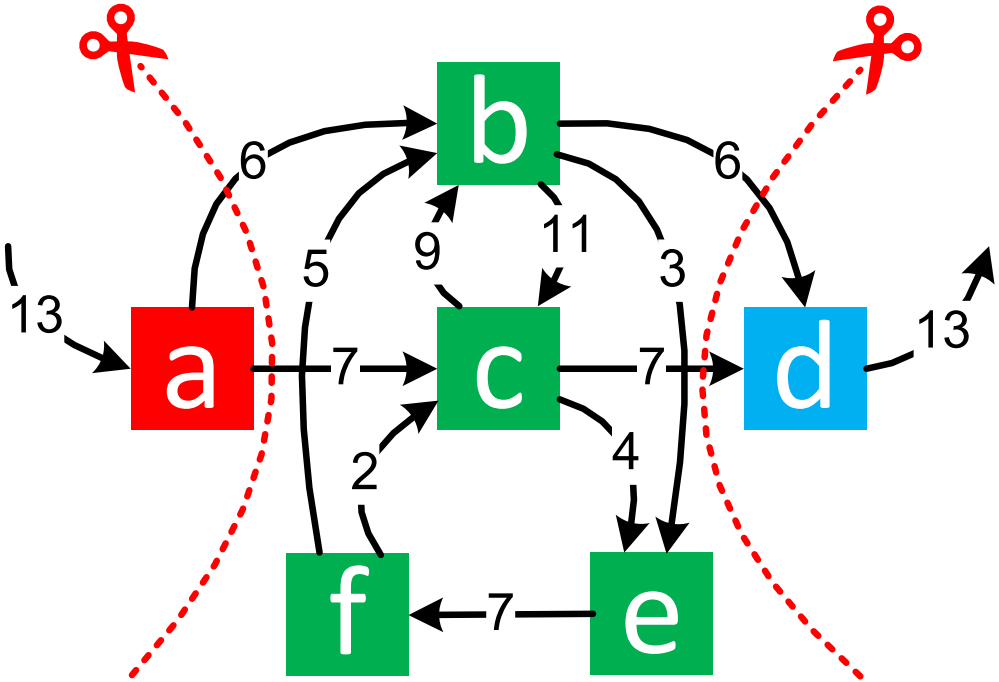




Chair of Process and Data Science

Partition activities based on sequence cut

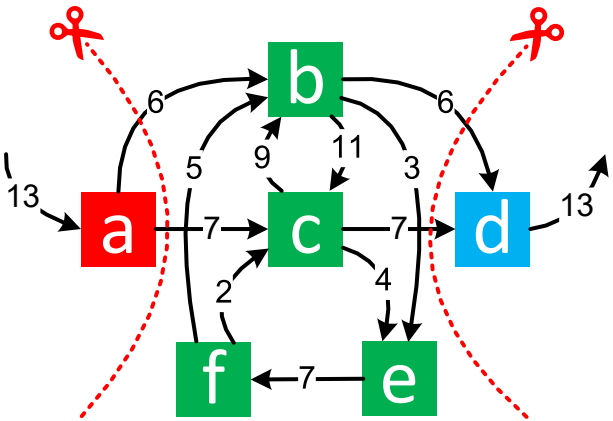




{a} , {b,c,e,f} , {d}

Chair of Process and Data Science

Partition events based on sequence cut



3x a b c d

4x a c b d

2x a b c e f b c d

2x a c b e f b c d

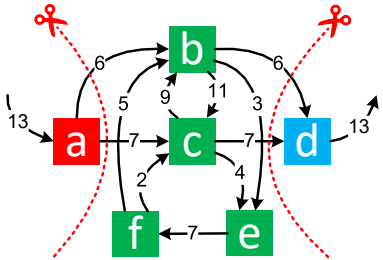
1x a b c e f c b d

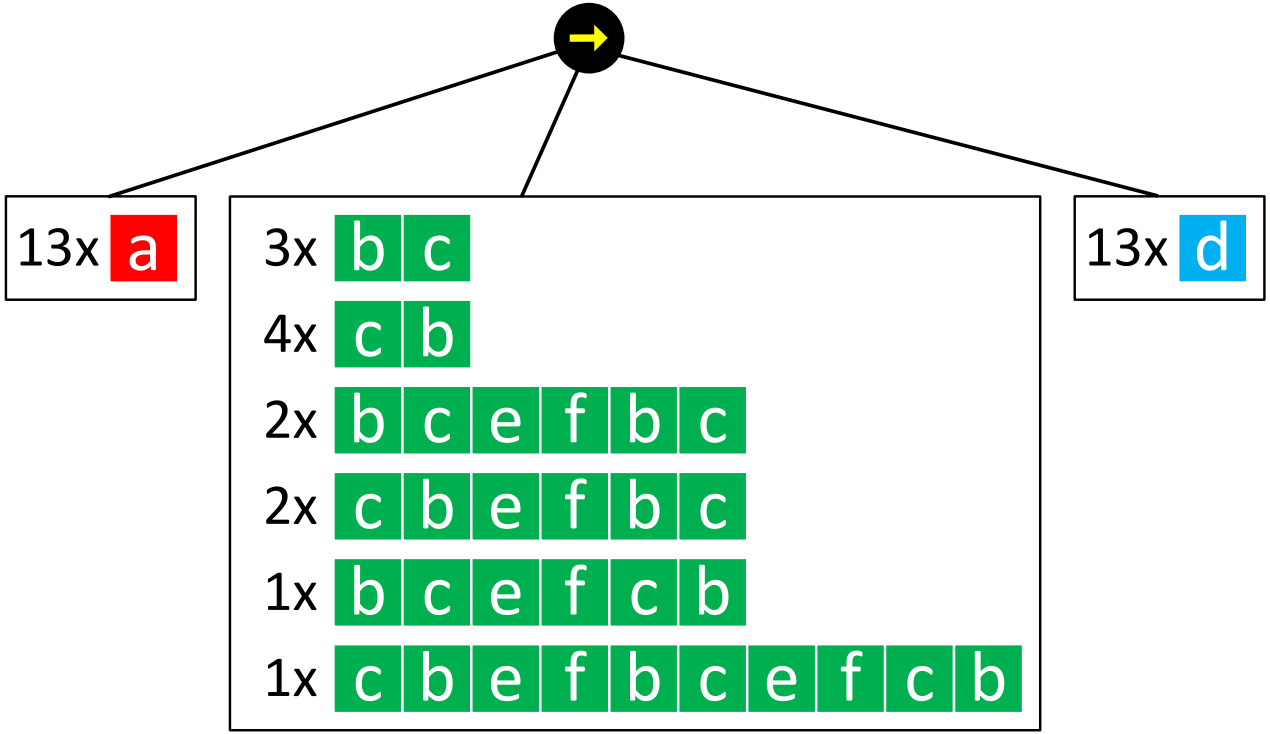
1x a c b e f b c e f c b d

{a} , {b,c,e,f} , {d}

Chair of Process and Data Science

Partition events based on sequence cut

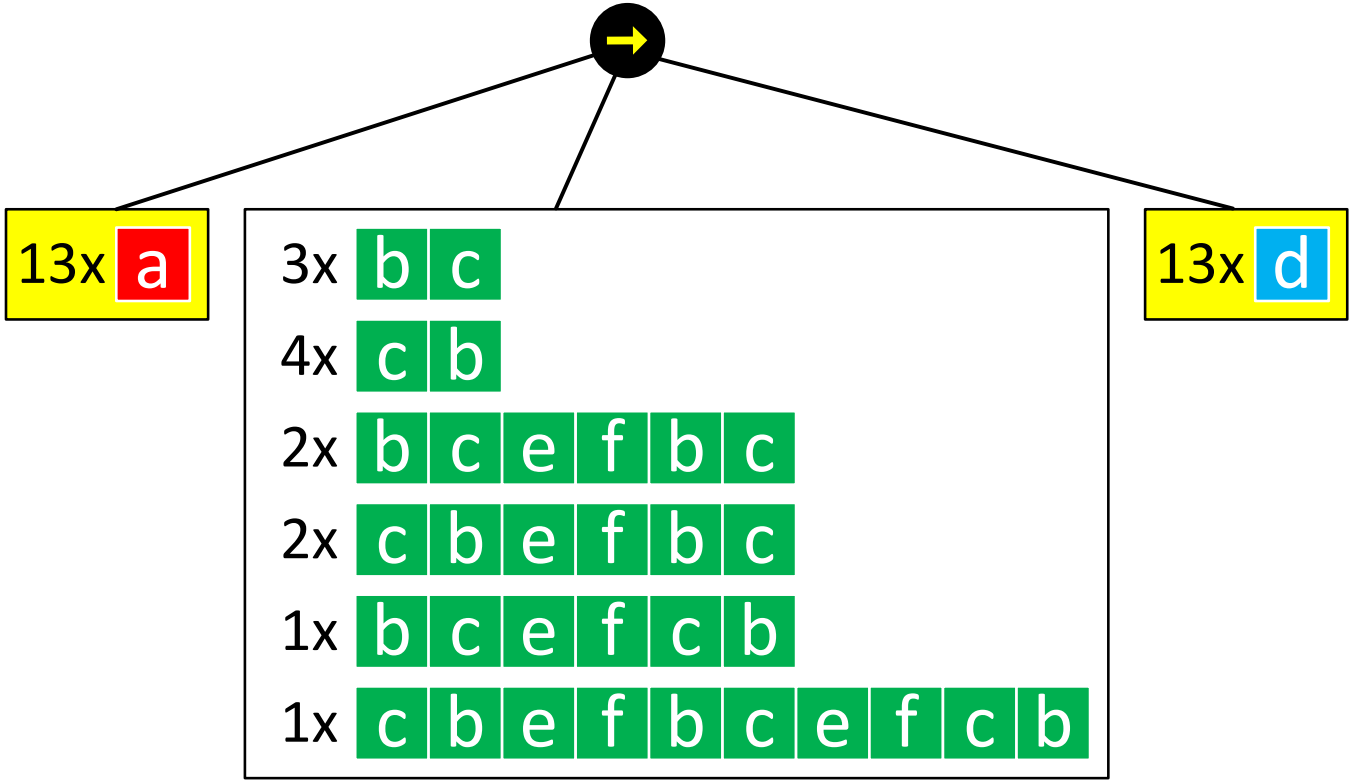




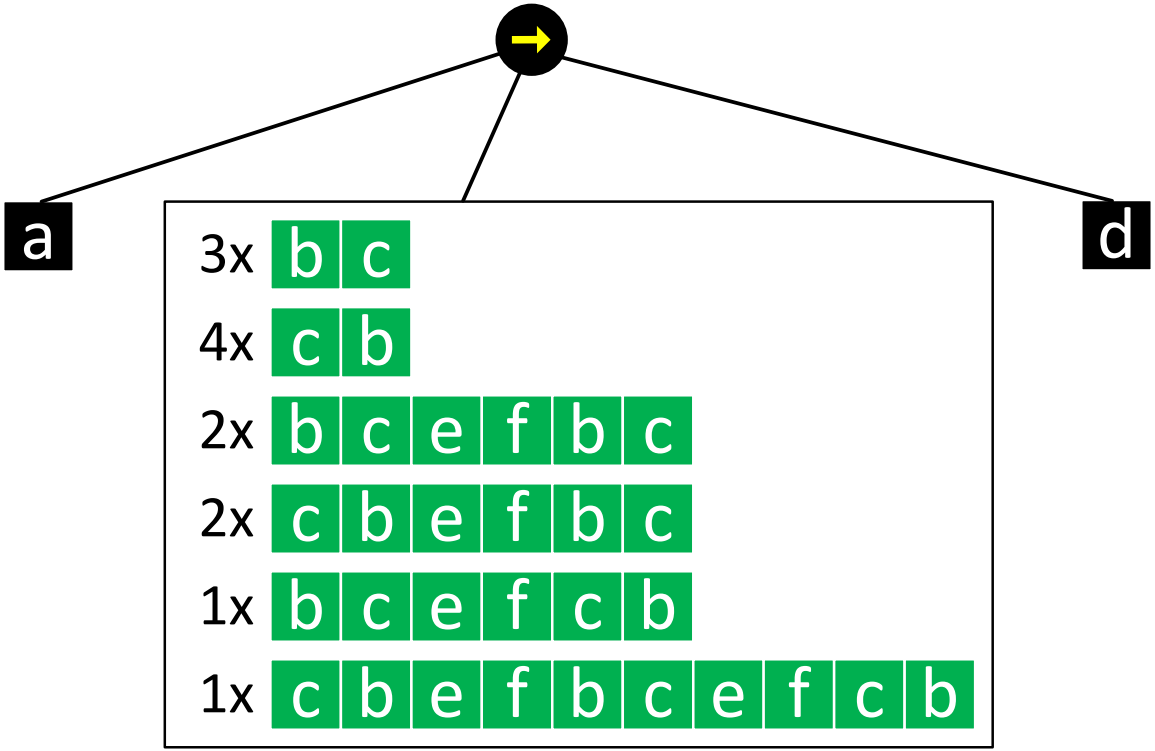
3x a b c d
4x a c b d
2x a b c e f b c d
2x a c b e f b c d
1x a b c e f c b d
1x a c b e f b c e f c b d

Chair of Process and Data Science

Handle base cases

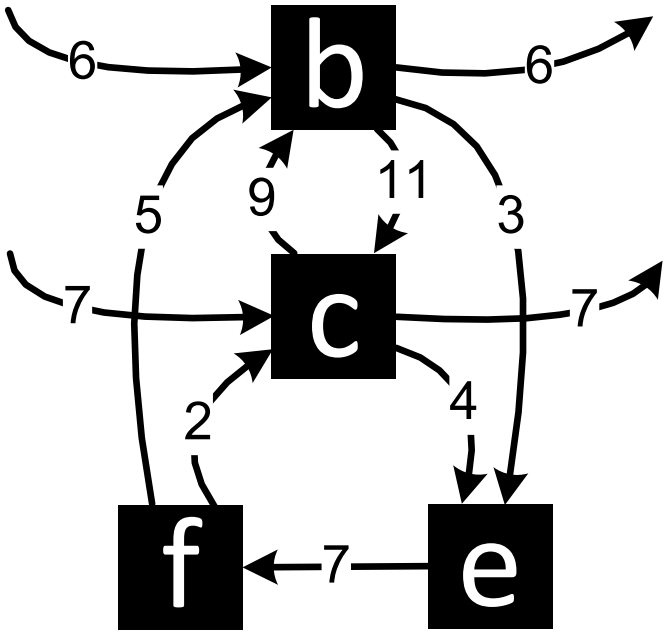


Recurse on non-base cases



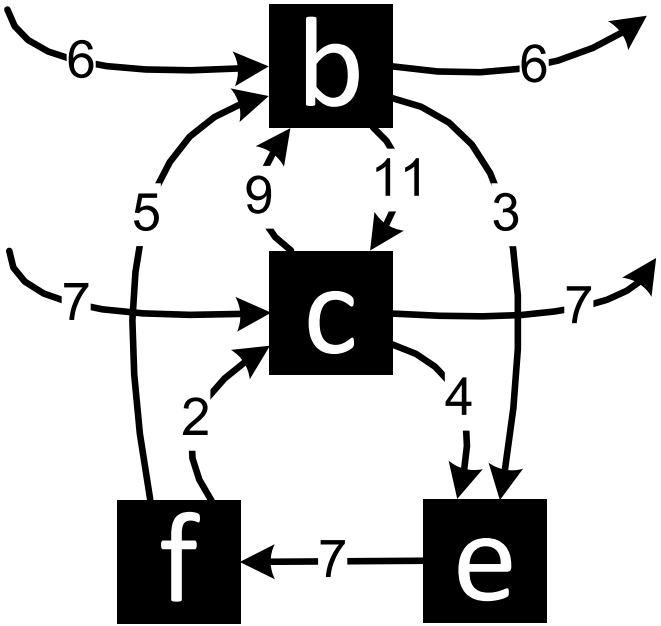
Directly-follows graph based on sublog

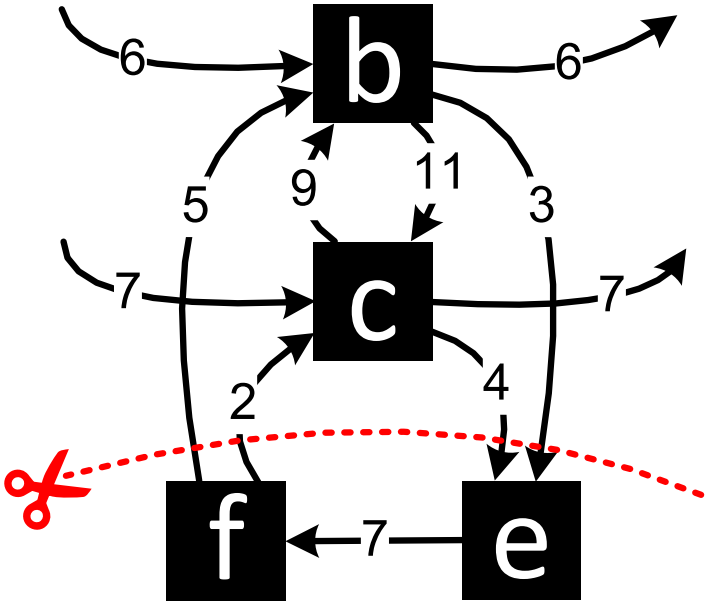
3x b c
4x c b
2x b c e f b c
2x c b e f b c
1x b c e f c b
1x c b e f b c e f c b



Chair of Process and Data Science

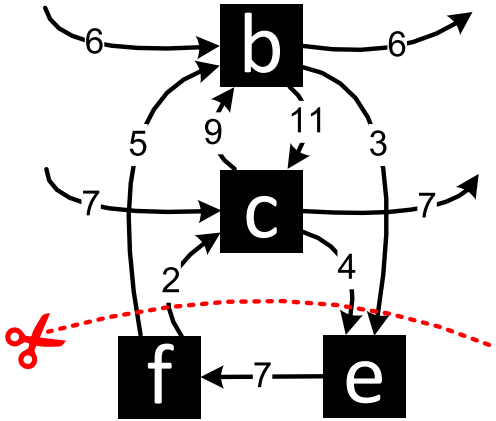
Loop cut



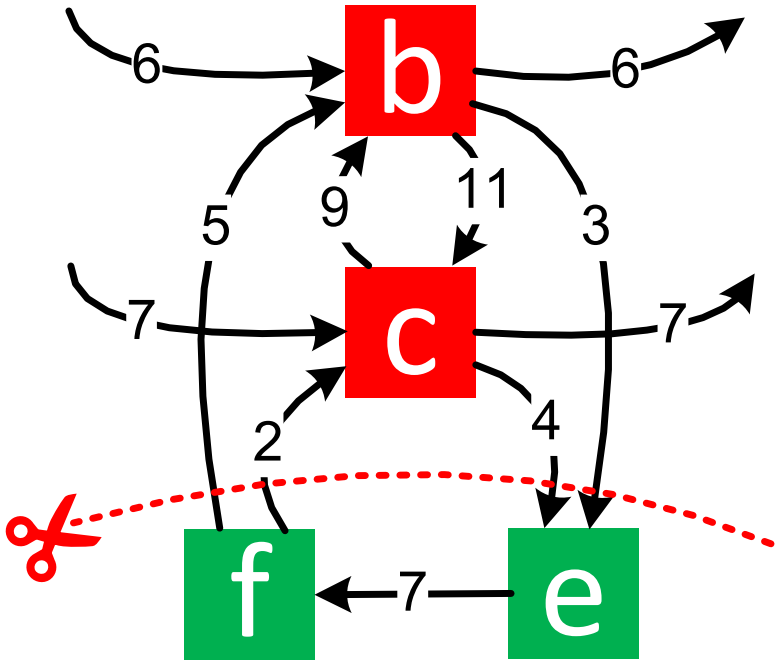


Chair of Process and Data Science

Partition activities based on loop cut

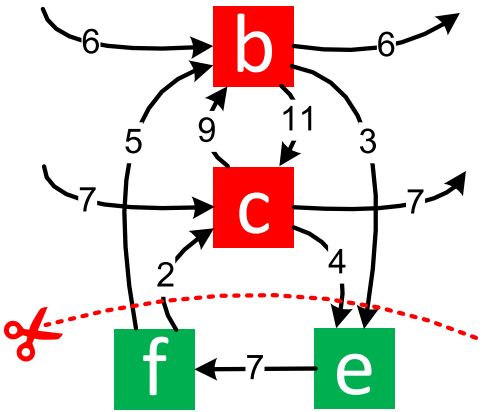


{b,c} , {e,f}



Chair of Process and Data Science

Partition events based on loop cut



{b,c} ,
{e,f}

3x b c

4x c b

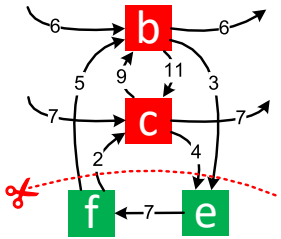
2x b c e f b c

2x c b e f b c

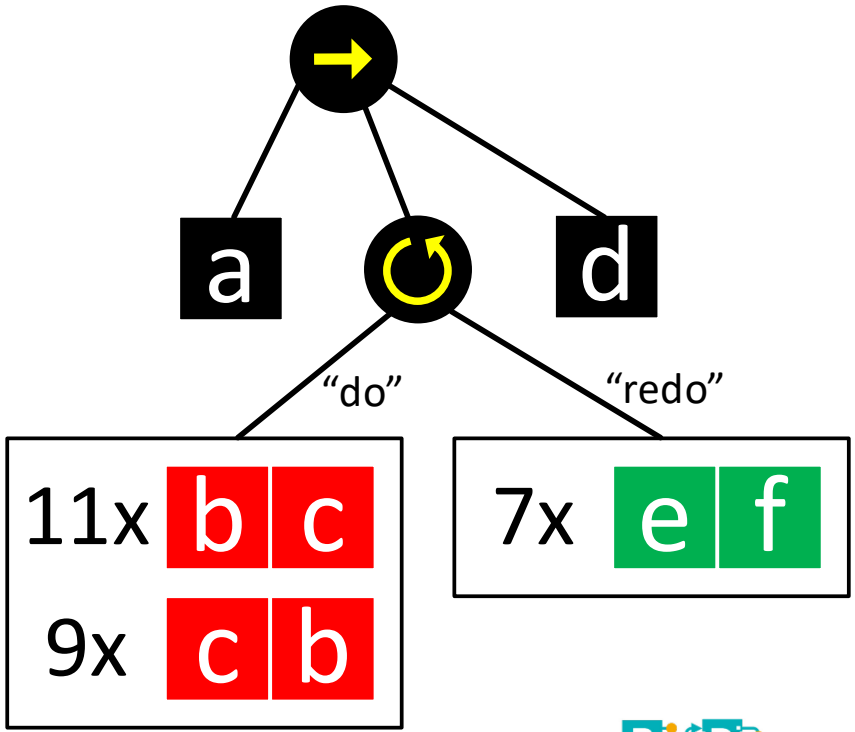
1x b c e f c b

1x c b e f b c e f c b

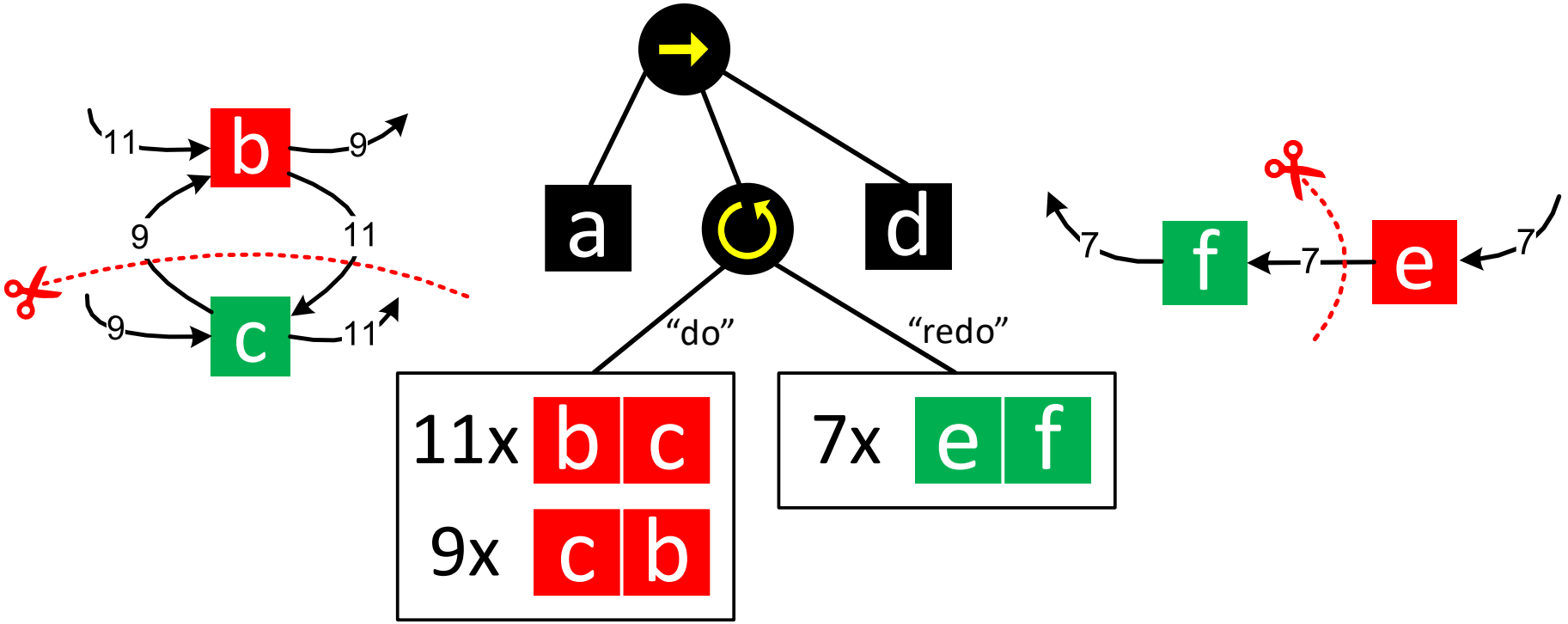
Partition events based on loop cut



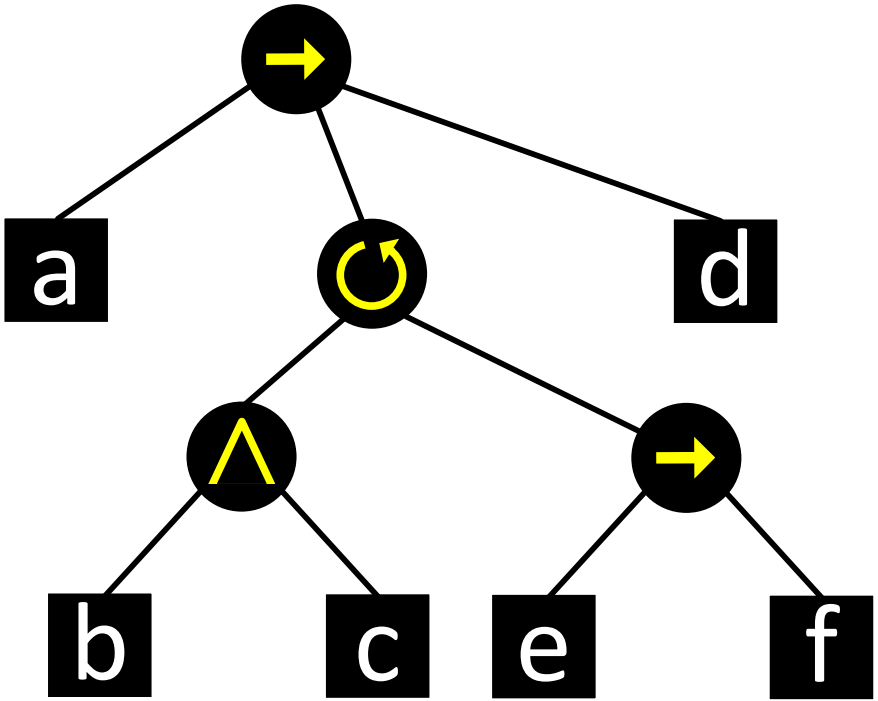
3x b c
4x c b
2x b c e f b c
2x c b e f b c
1x b c e f c b
1x c b e f b c e f c b



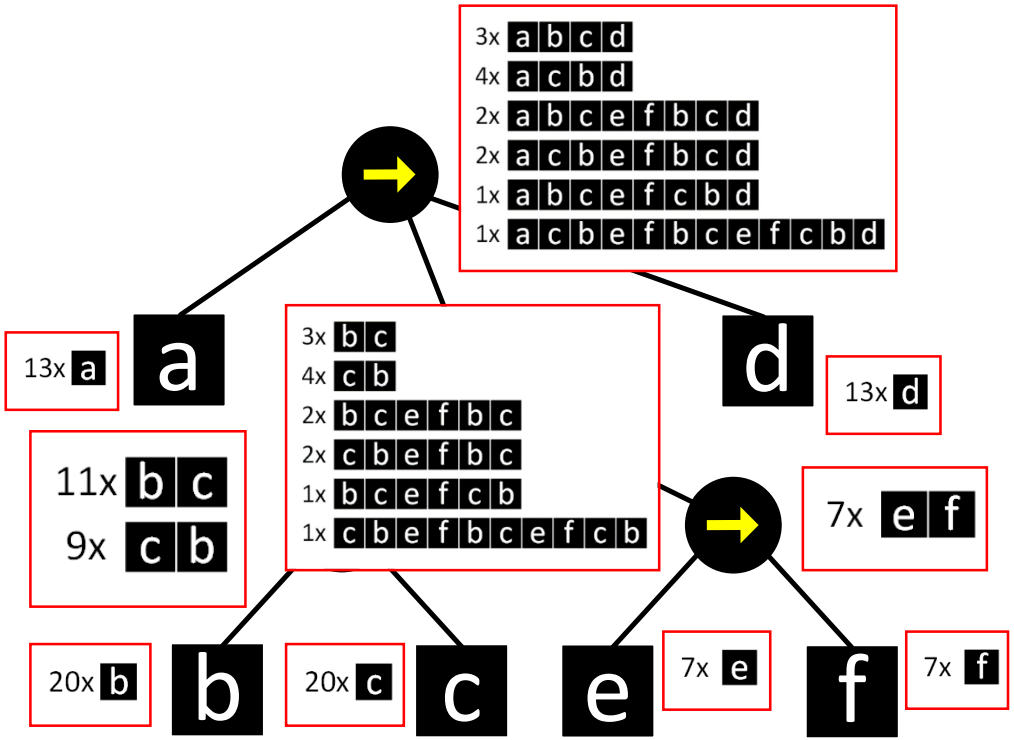
Recurse on the two sublogs



Final model

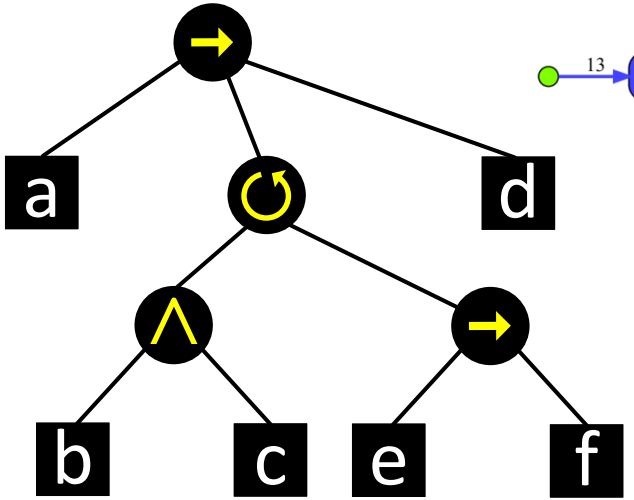


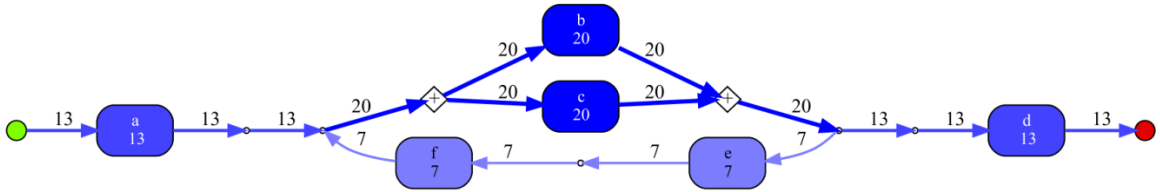
Top-down process

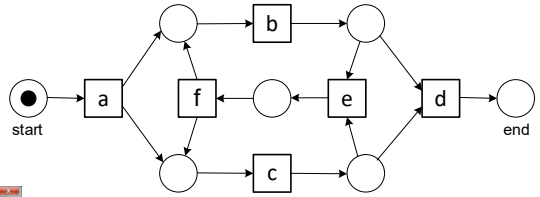


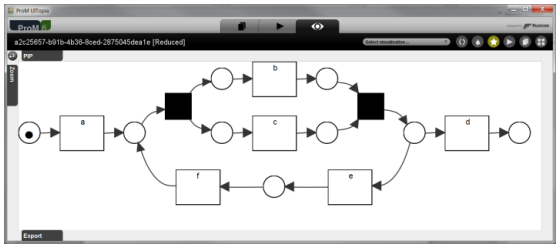
Chair of Process and Data Science

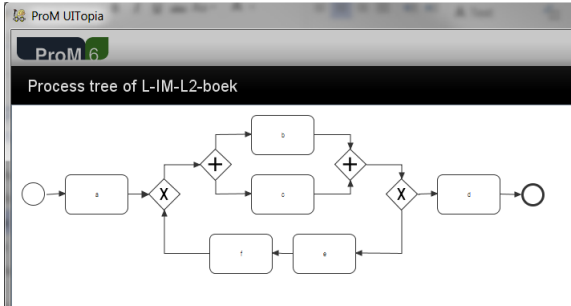
Alternative notations











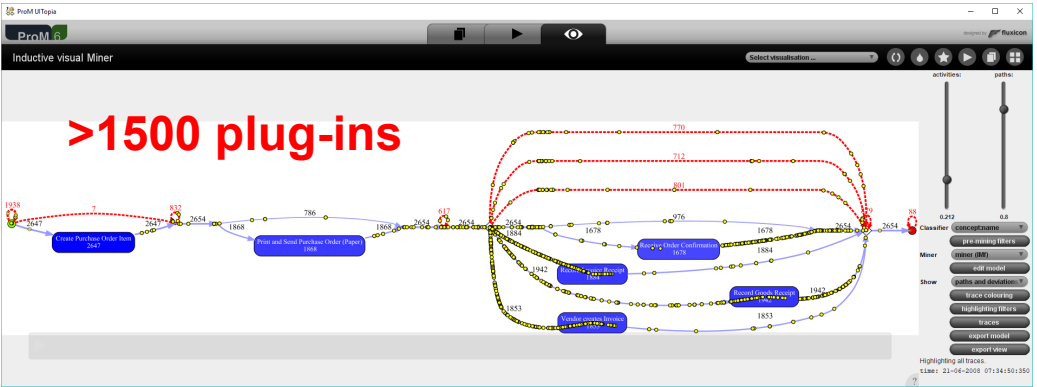
Properties

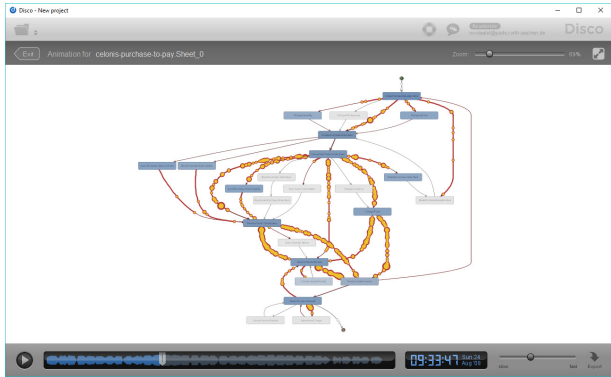
- Basic algorithm formally guarantees that the original event log can be replayed. Models are always sound.
- If the event log was generated from a basic process tree (no duplication of labels), then an equivalent model will be found.
- Extensions to deal with infrequent behavior and incomplete event logs.
- Highly scalable. Dealing with billions of events, millions of cases, and thousands of unique activities.
- Allows for distribution, streaming, etc.

Tooling and applications

Tooling

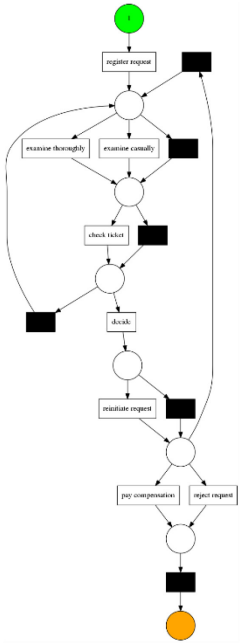
- **ProM** is the de facto standard in the scientific world.
- Ideas initially developed in ProM have been adopted by commercial vendors.
- Currently, more than 25 commercial vendors offering process mining software (Celonis, Fluxicon, ProcessGold, QPR, etc.).

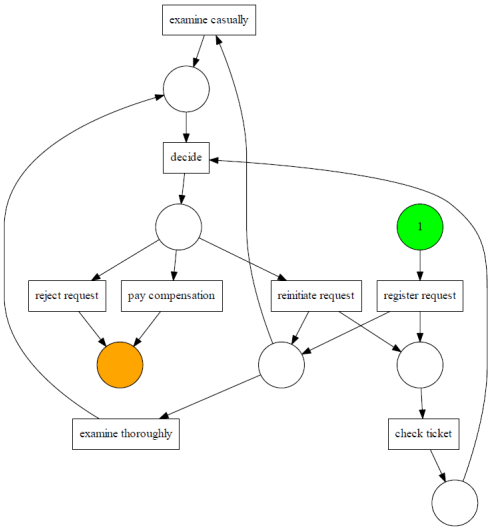




pm4py

- Simple python library
- New initiative (opportunities for thesis projects, HiWi jobs, etc.)
- Please read the instructions
- Install before Friday **(installing the Microsoft Visual Studio 2017 compiler takes time).**





Chair of Process and Data Science

Applications: Always about compliance and performance

- Processes supported by ERP and CRM systems (e.g., SAP).
- Healthcare (range of hospitals).
- Logistics and production (e.g., with Vanderlande).
- E-learning (e.g., based on Coursera).
- E-government (see CoSeLog project).
- Smart homes / quantified self (with Philips).
- High-tech systems.
- Auditing.
- Fraud detection.
- Etc.

Conclusion

Short summary of lecture

- Process discovery as a form of unsupervised process mining.
- Starts from event data.
- Goal: End-to-end models

Sequence mining versus process mining

Sequence mining	Process mining
Focus on local patterns	Focus on end-to-end processes
Focus on itemsets	Focus on activities
Low-level patterns	Higher-level behaviors (e.g., concurrency and choice).
Specific event data (itemset based)	General event data (case, activity, resource, lifecycle, etc.)
…	…

#	Lecture	date	day
Lecture 1	Introduction	10/10/2018	Wednesday
Instruction 1			
Instruction 2			
Instruction 3			
Instruction 4			
Instruction 5			
Lecture 12	Association rules	21/11/2018	Wednesday
Lecture 13	Sequence mining	22/11/2018	Thursday
Instruction 6	*Clustering, frequent items sets, association rules*	*23/11/2018*	*Friday*
Lecture 14	Process mining (unsupervised)	28/11/2018	Wednesday
Lecture 15	Process mining (supervised)	29/11/2018	Thursday
Instruction 7	*Process mining and sequence mining*	*30/11/2018*	*Friday*
Lecture 16	Text mining (1/2)	05/12/2018	Wednesday
Instruction 8	*Text mining and process mining*	*06/12/2018*	*Thursday !!*
Lecture 17	Text mining (2/2)	12/12/2018	Wednesday
Lecture 18	Data preprocessing, data quality, binning, etc.	13/12/2018	Thursday
Lecture 19	Visual analytics & information visualization	19/12/2018	Wednesday
backup		20/12/2018	Thursday
Instruction 9	*Text mining, preprocessing and visualization*	*21/12/2018*	*Friday*
Lecture 20	Responsible data science (1/2)	09/01/2019	Wednesday
Lecture 21	Responsible data science (2/2)	10/01/2019	Thursday
Instruction 10	*Responsible data science*	*11/01/2019*	*Friday*
Lecture 22	Big data (1/2)	16/01/2019	Wednesday
Lecture 23	Big data (2/2)	17/01/2019	Thursday
Instruction 11	*Big data*	*18/01/2019*	*Friday*
Lecture 24	Closing	23/01/2019	Wednesday
backup		24/01/2019	Thursday
Instruction 12	*Example exam questions*	*25/01/2018*	*Friday*
backup		30/01/2019	Wednesday
backup		31/01/2019	Thursday
extra	*Question hour*	*01/02/2019*	*Friday*

The deadline for the first assignment is Sunday 09/12/2018 23:59.

Introduction to Data Science - Lecture
Participants
Grades
Sections
General
Introduction
Crash Course in Python
Basic data visualisation/exploration
Decision trees
Regression
Support Vector Machines
Neural Networks
Evaluation of Supervised Learning Problems
Assignment 1
Clustering
Frequent Item Sets
Association Rules

Introduction to Data Science - Lecture

Dashboard / My courses / Introduction to Data Science -... / Sections / Assignment 1 / Assi

Assignment 1

This assignment guides you through the analysis of a real-life data set using the techniques and tools provided in the course.

- Please read the pdf file carefully, use two training and test dataset correctly and complete the provided Jupyter notebook.
- As an answer, you should only upload the provided Jupiter notebook template in this section.
- Please note after the mentioned deadline the upload option will be closed automatically.
- The deadline for the assignment is Sunday 09/12/2018 23:59.

Assignment 1 Answer Template.ipynb
Assignment 1 IDS.pdf
data-set-assignment1.zip

Grading summary

11 days

Participants	299
Submitted	4
Needs grading	4
Due date	Sunday, 9 December 2018, 11:59 PM
Time remaining	11 days 16 hours

View all submissions | Grade