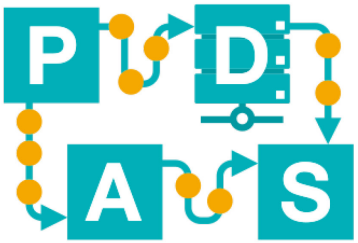
Process Mining Supervised

Lecture 15

IDS-L15

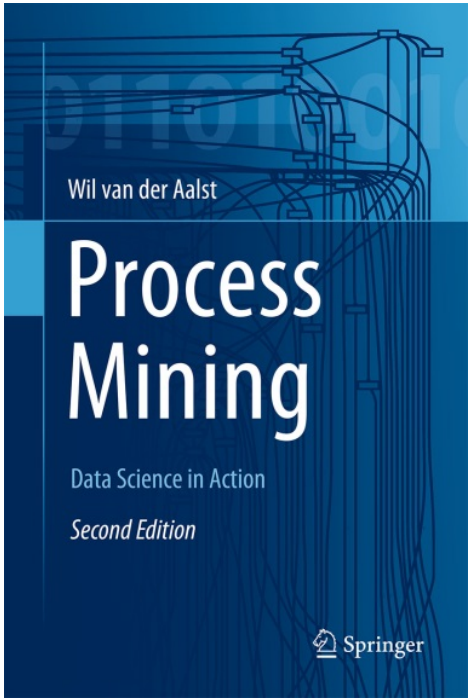


Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Outline of Today's Lecture

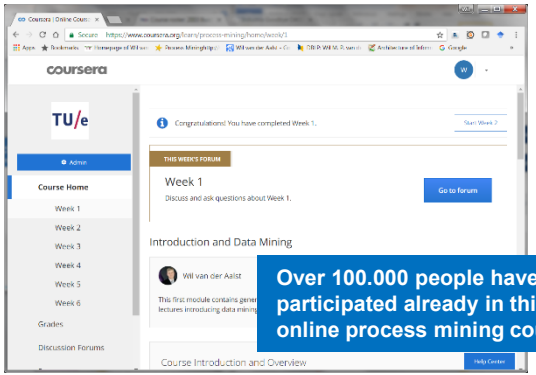
- Short recap
- Two main questions
- Conformance checking
- Generating supervised learning problems
- Tooling & Ecosystem



Chair of Process and Data Science

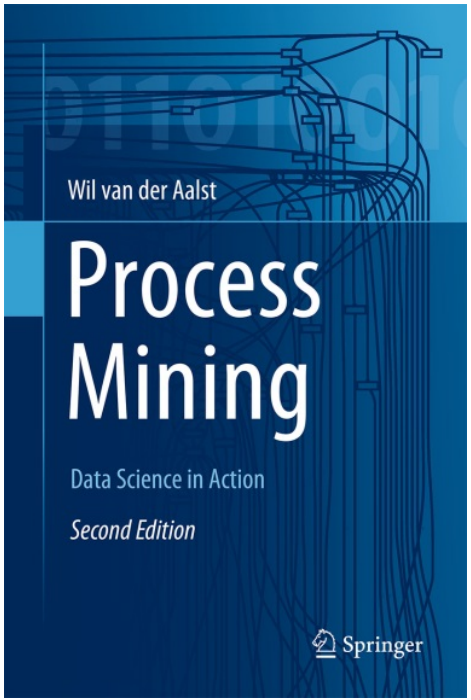
Material (only as background information or if you want to dive deeper)

- **Chapters 8-10 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.

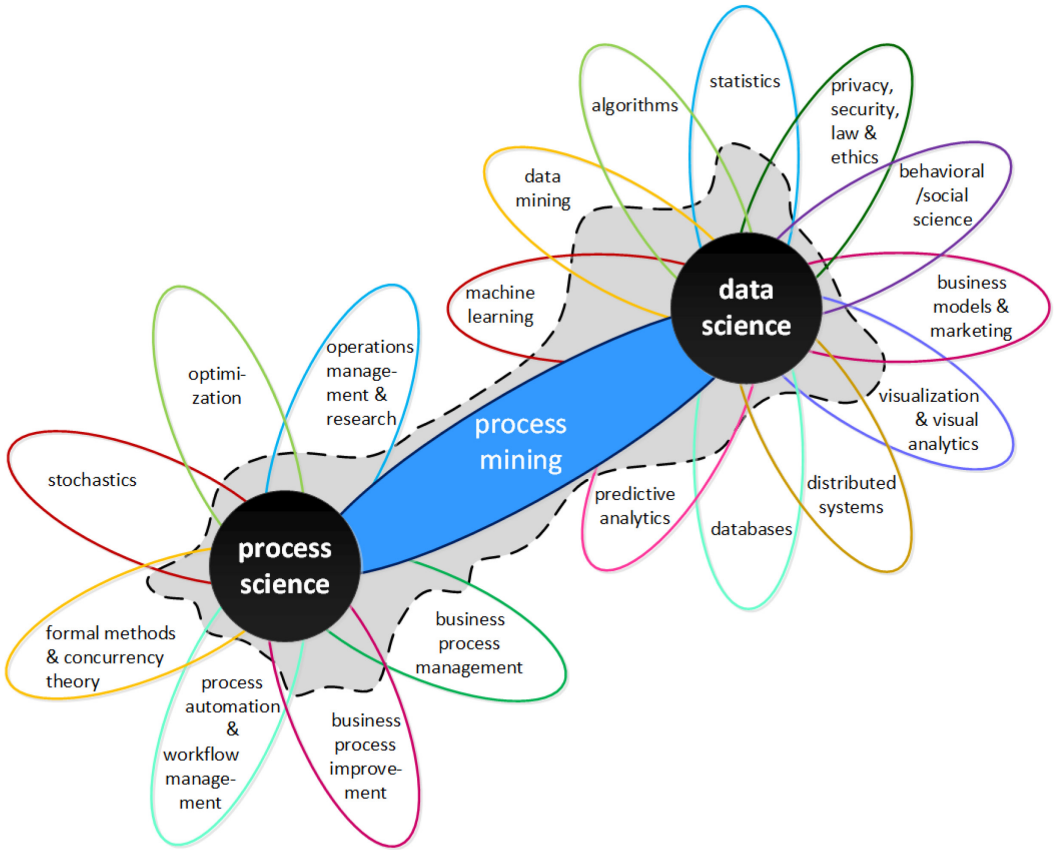
coursera
education for everyone



Chair of Process and Data Science

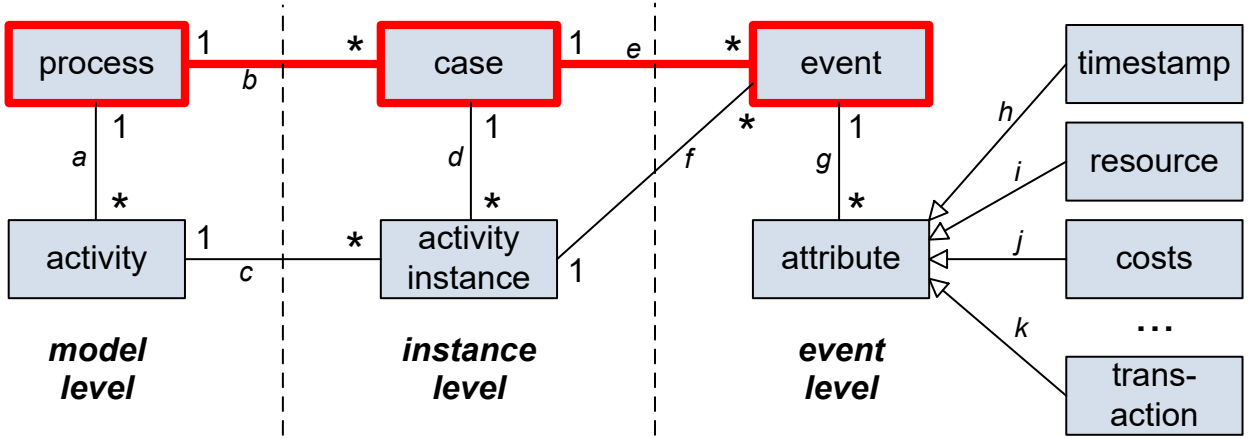
Short recap

Positioning process mining



Chair of Process and Data Science

Input: Event data

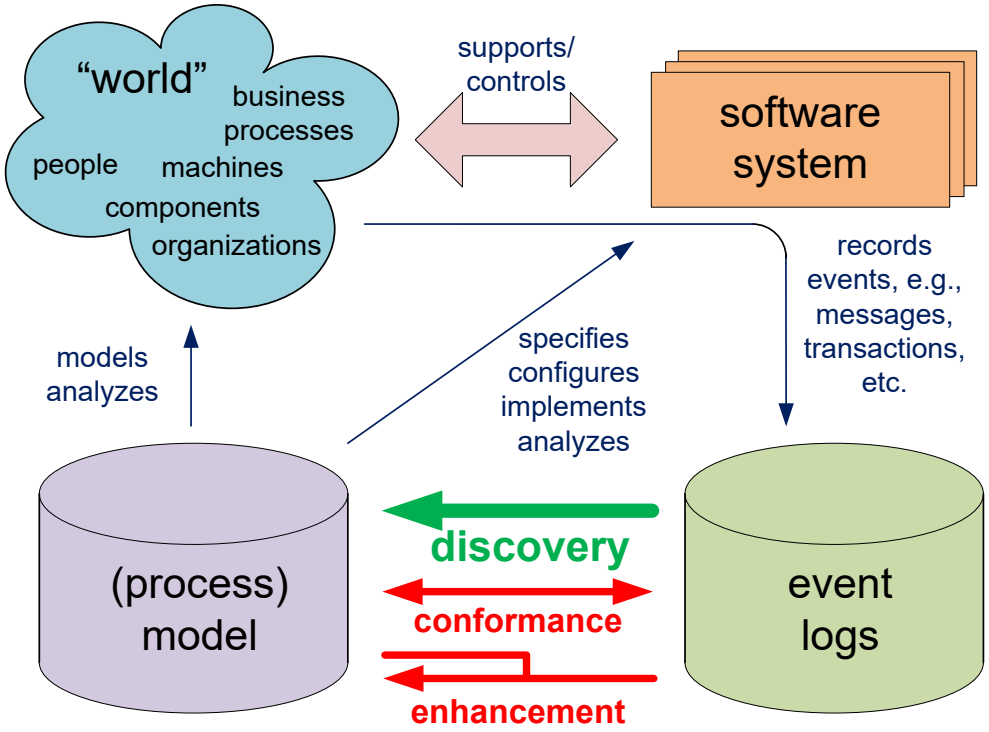


XES
Extensible Event Stream

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

Chair of Process and Data Science

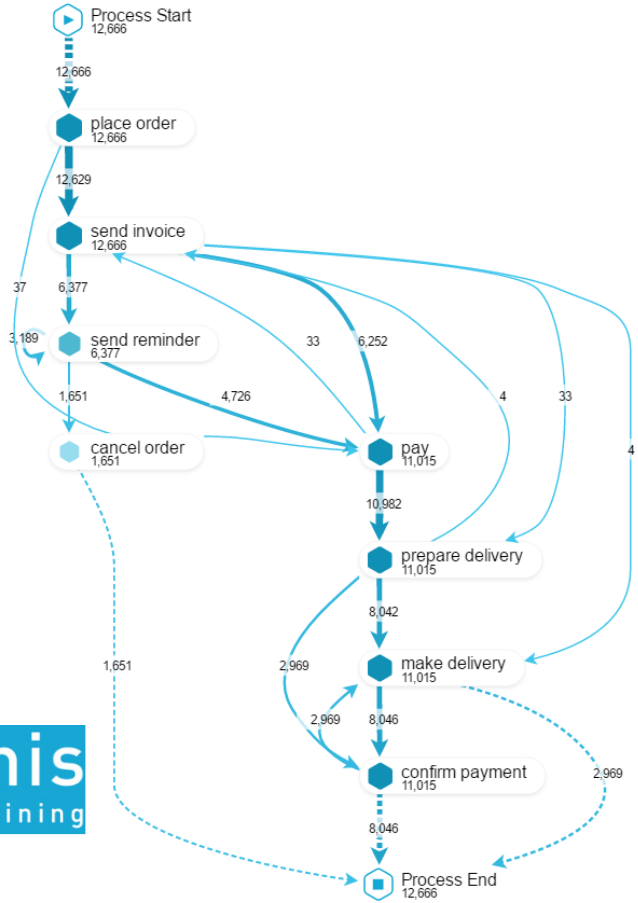
Process discovery



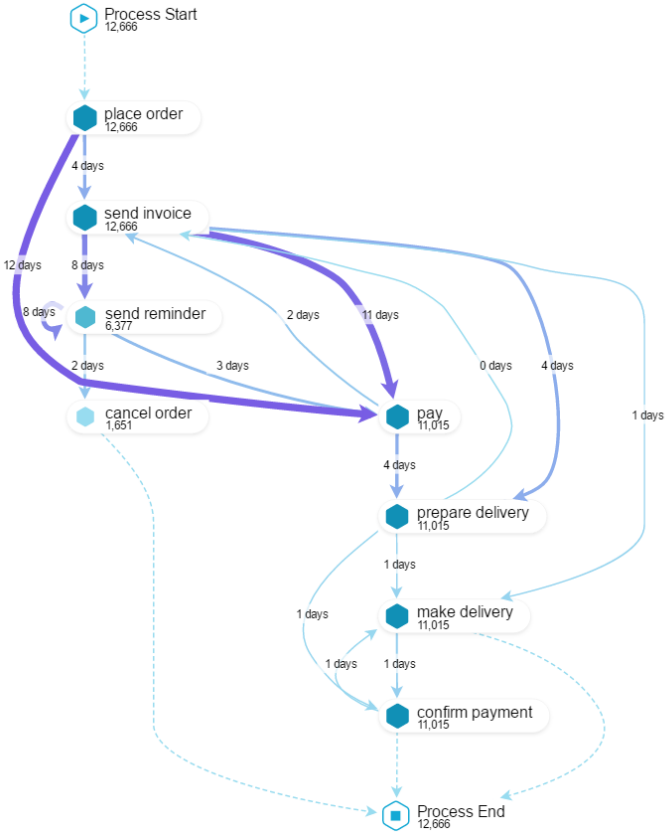
Chair of Process and Data Science

“happy”

“freq”



“time”



celonis process mining

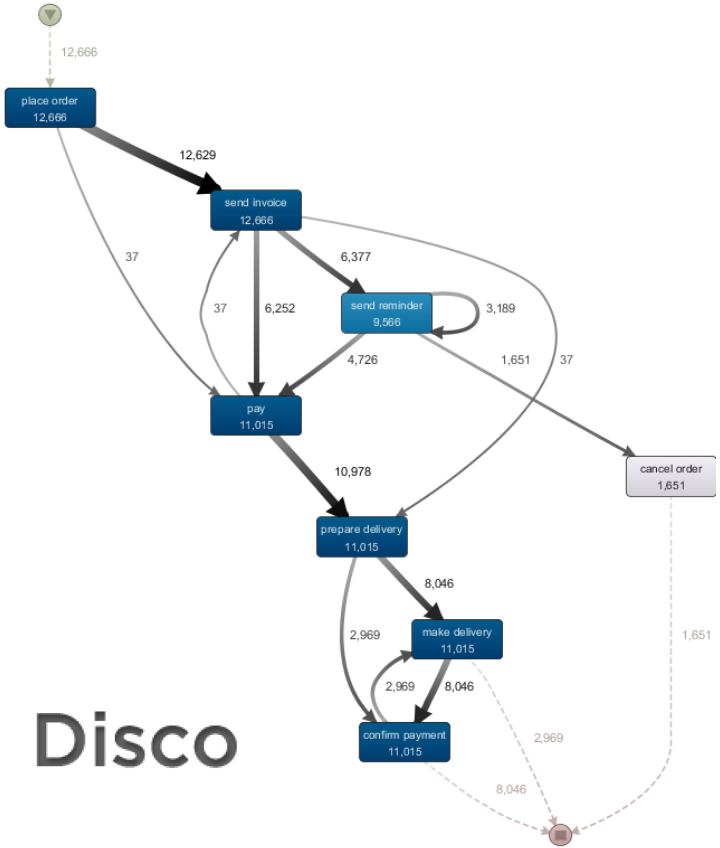
Chair of Process and Data Science

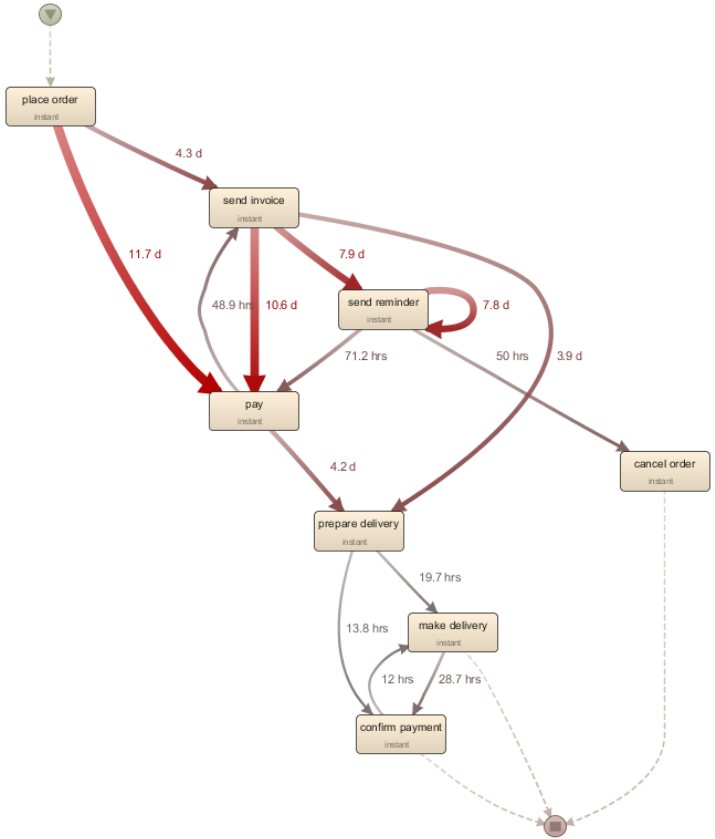
“happy”

“freq”

“time”

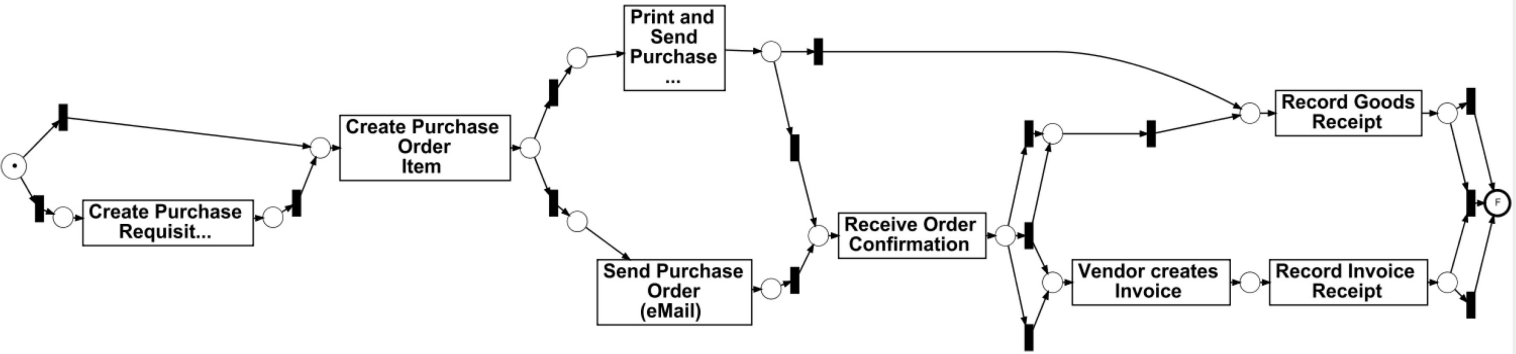
Disco

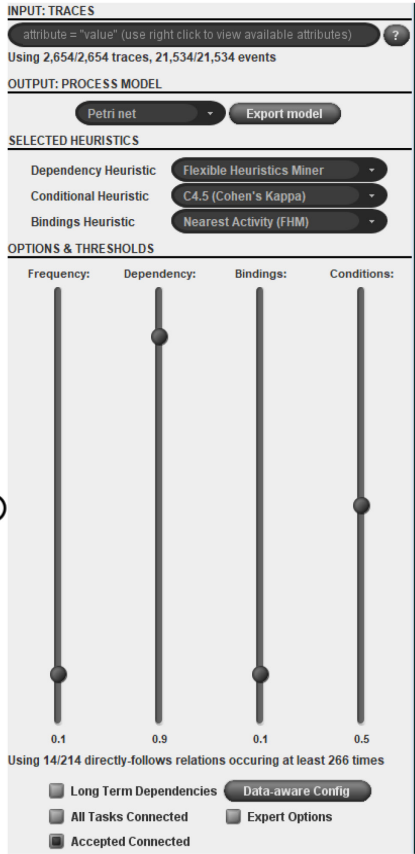




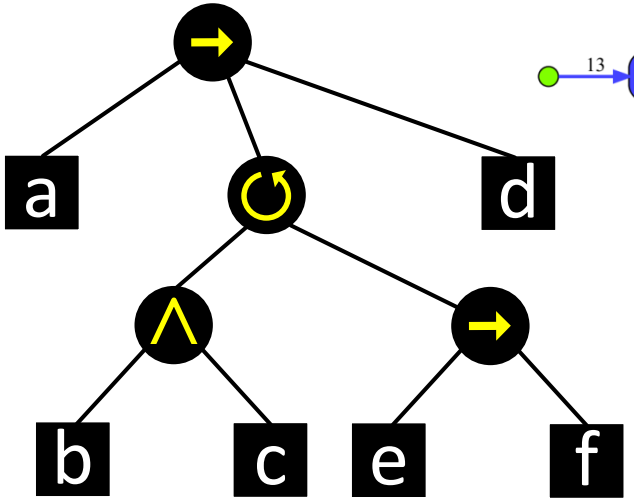
Chair of Process and Data Science

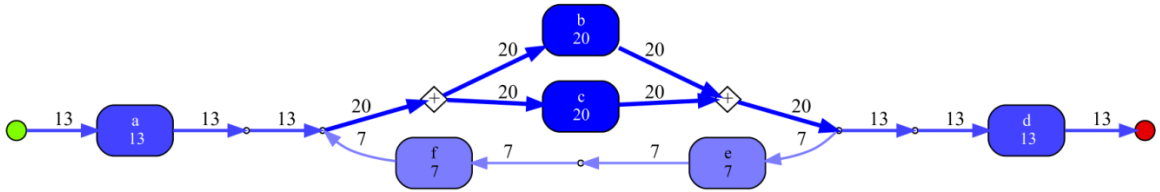
Bottom-up process discovery

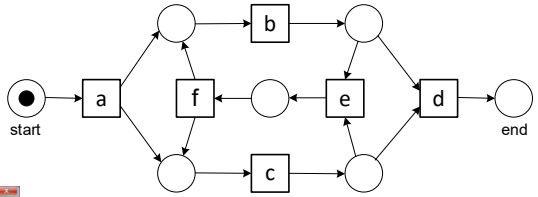


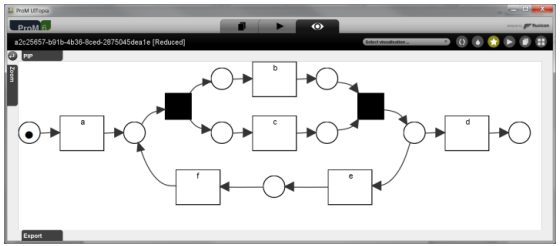


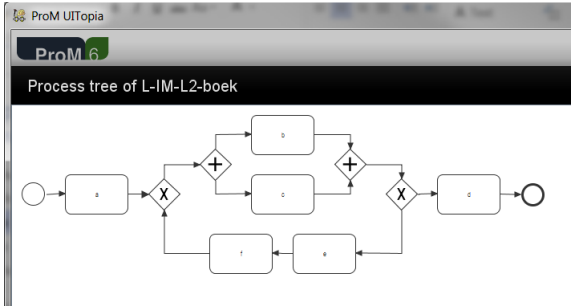
Top-down process discovery in ProM











Two main questions


event
data
process
mining
process
models


Reporting: What happened?
Diagnosis: Why did it happen?
performance problems
Prediction: What will happen?
Recommendation: What is the best that can happen?

Heathrow

Making every journey better

- Why do bags miss a plane?
- Why do I need to wait so long for my bags?
- When and why does the system break down?
- Am I using the available capacity properly?

UNIKLINIK RWTH AACHEN

- How long do patients have to wait for the first appointment?
- Why are there always long queues at the X-rays dept. between 11.00-13.30 ?
- How often do we need to refuse patients?

- When and why are we unable to deliver on the planned date?
- How quickly can we answer questions?
- What is causing late payments?

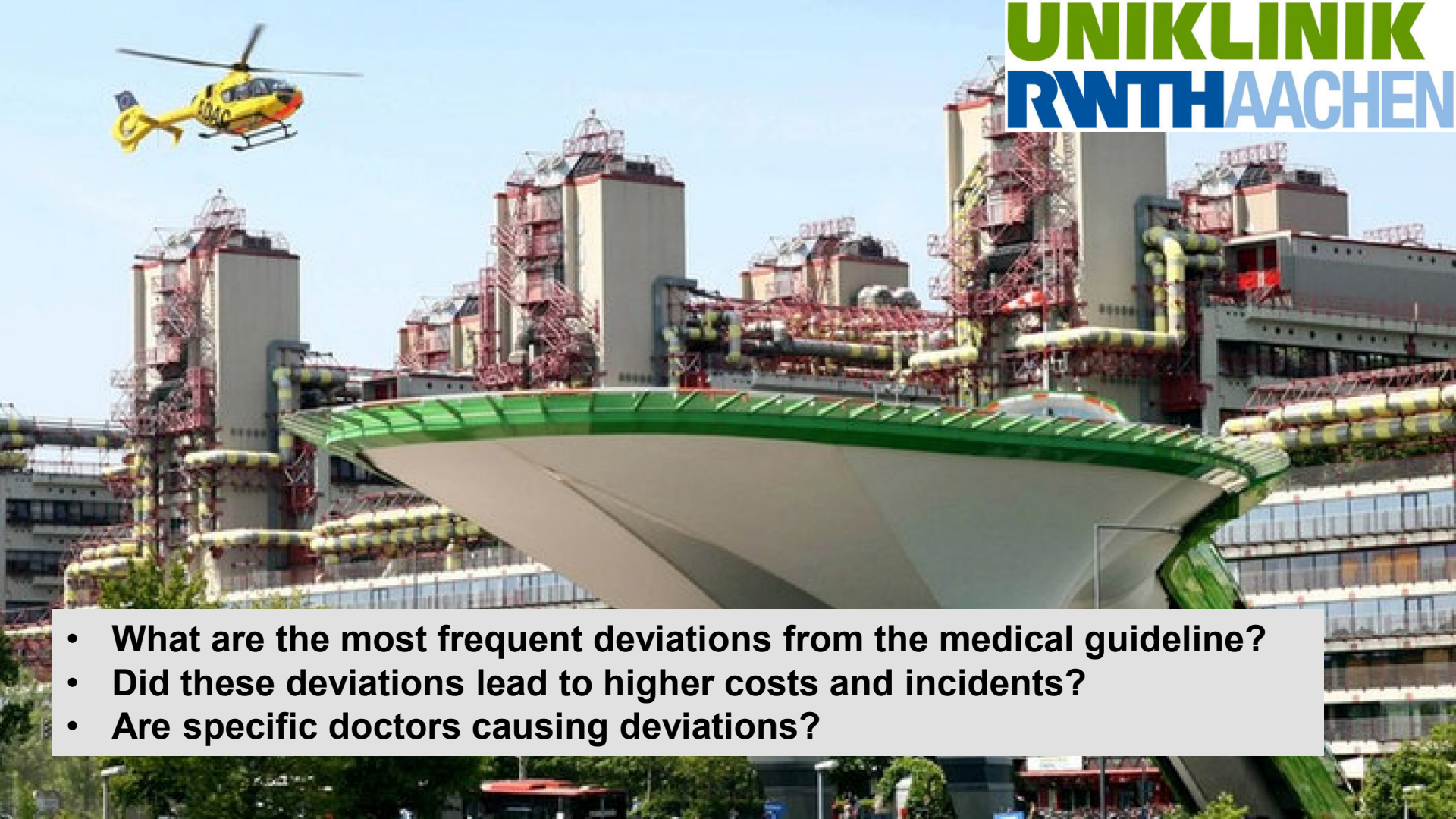
- How can we help organizations to save costs?
- How can we help organizations to serve customers better?
- What are best practices?


Reporting: What happened?
Diagnosis: Why did it happen?
compliance problems
Prediction: What will happen?
Recommendation: What is the best that can happen?

Heathrow

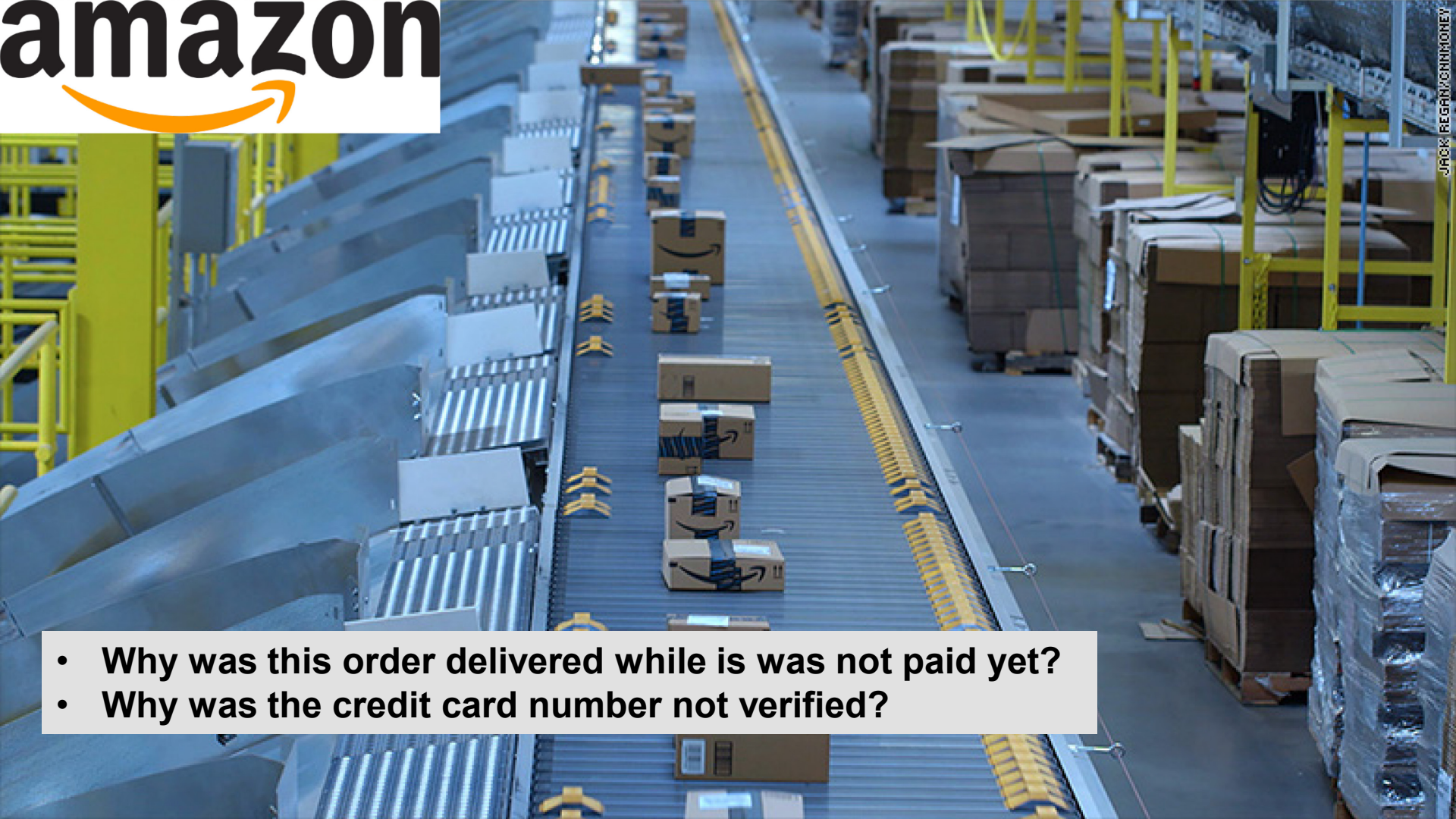
Making every journey better

- **Why did a bag take a different route?**
- **Why was this bag not scanned?**
- **Why did the order change?**



UNIKLINIK RWTH AACHEN

- What are the most frequent deviations from the medical guideline?
- Did these deviations lead to higher costs and incidents?
- Are specific doctors causing deviations?

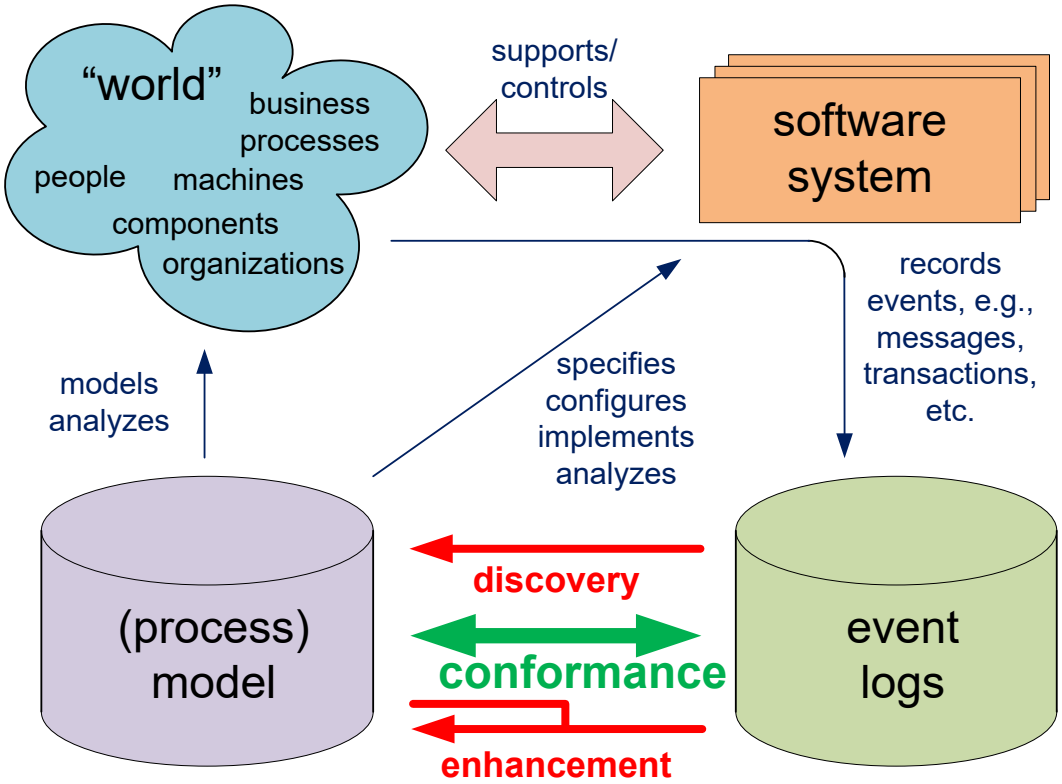


- Why was this order delivered while is was not paid yet?
- Why was the credit card number not verified?

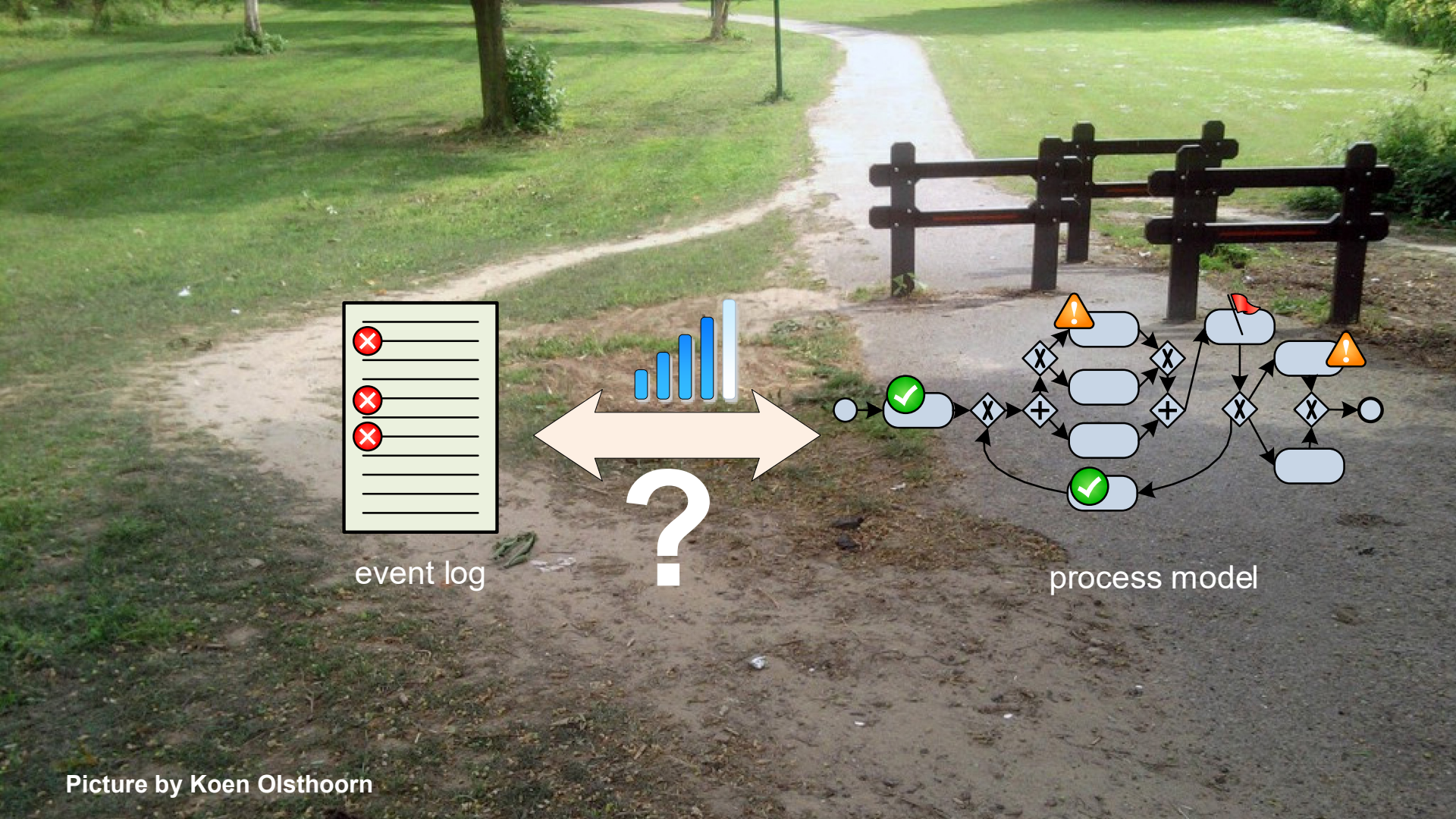
- How can we help organizations to detect fraud?
- How can we help organizations to audit their key processes?

Conformance checking

Conformance checking

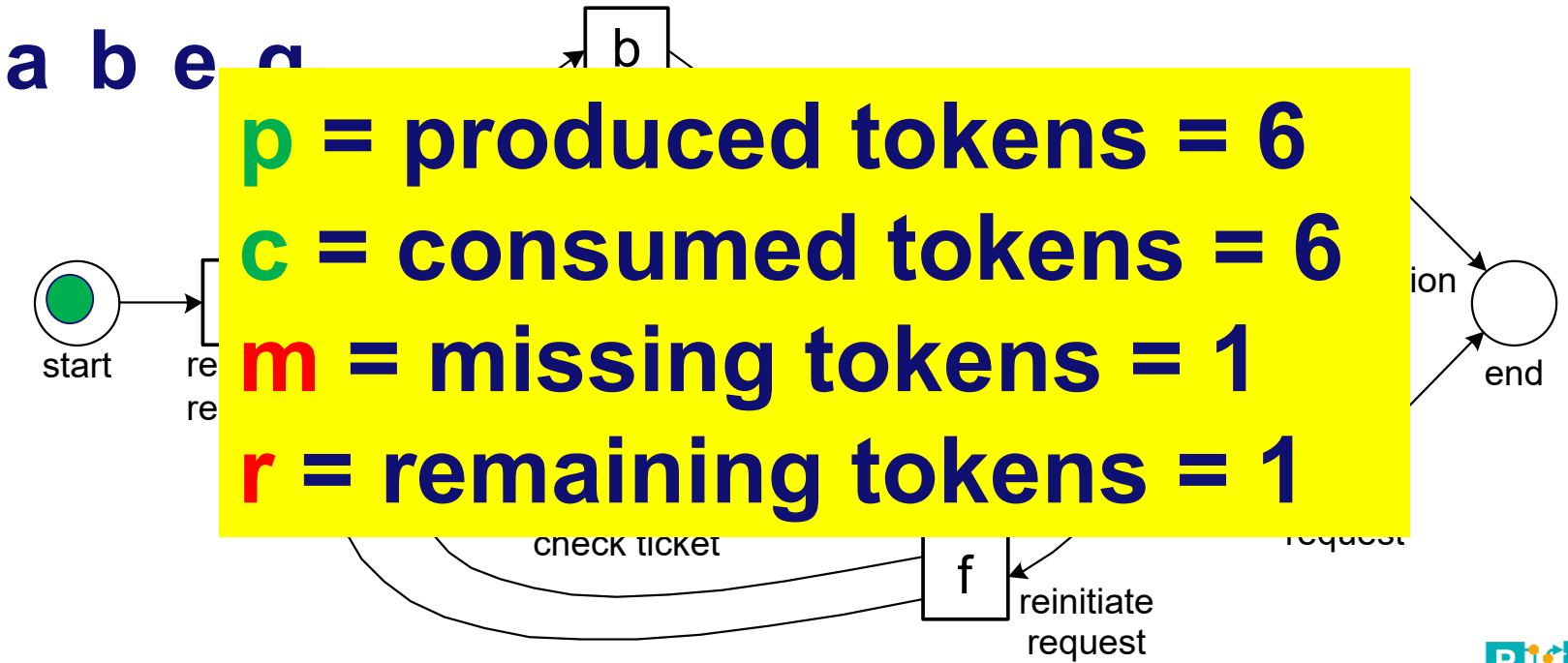


Chair of Process and Data Science



Picture by Koen Olsthoorn

Counting tokens while replaying



Chair of Process and Data Science

Quantifying fitness at the trace level

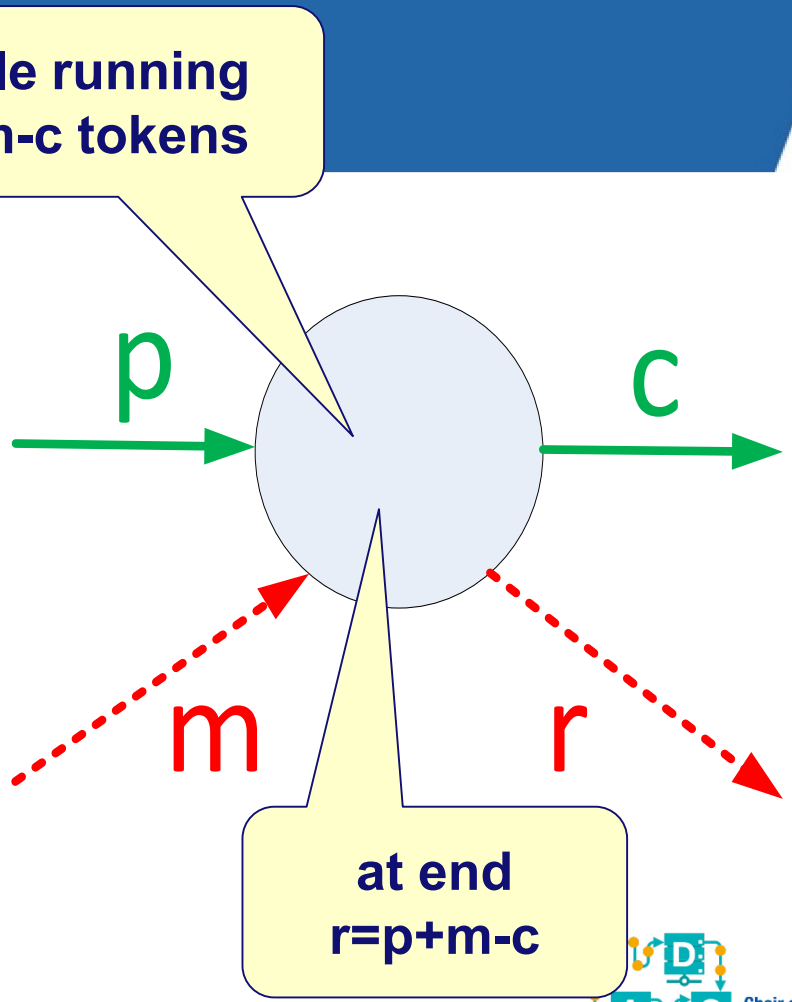
$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.83333$$

p = produced tokens = 6
c = consumed tokens = 6
m = missing tokens = 1
r = remaining tokens = 1

Approach (1/3)

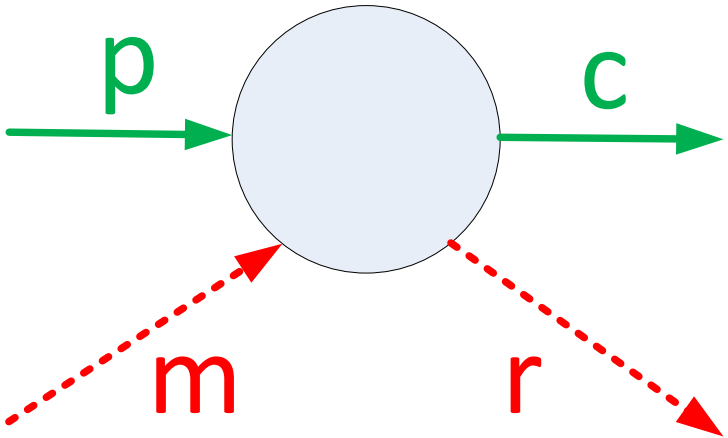
Use four counters:

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



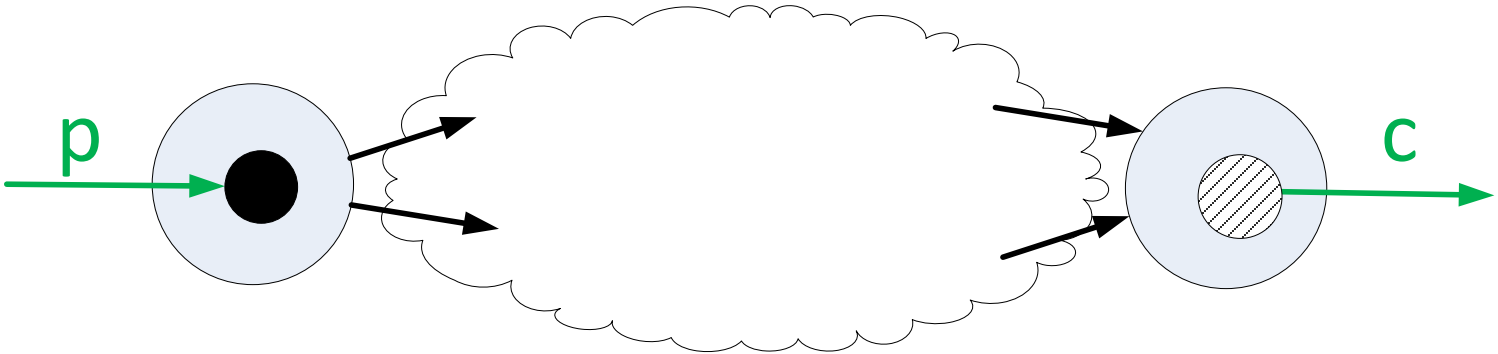
Chair of Process and Data Science

Approach (2/3)



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

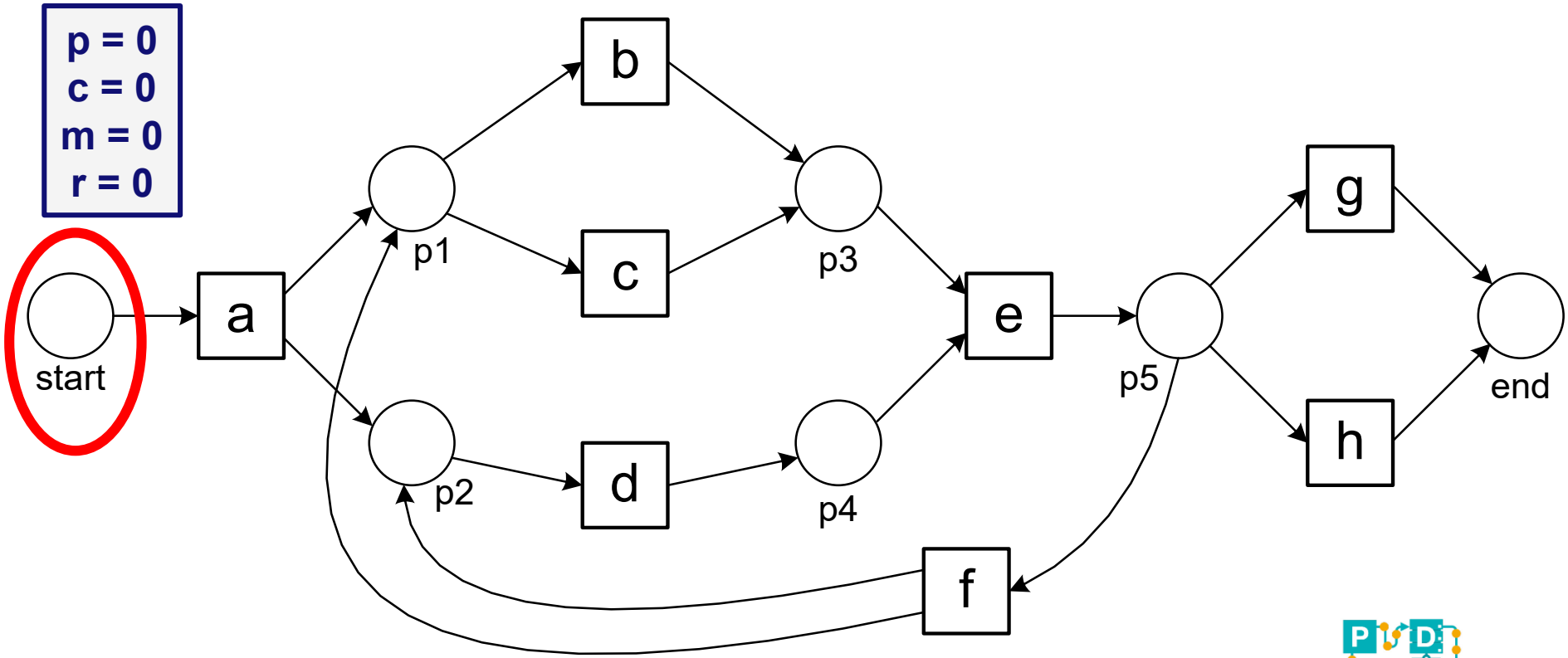
Approach (3/3)



Initialization and finalization:

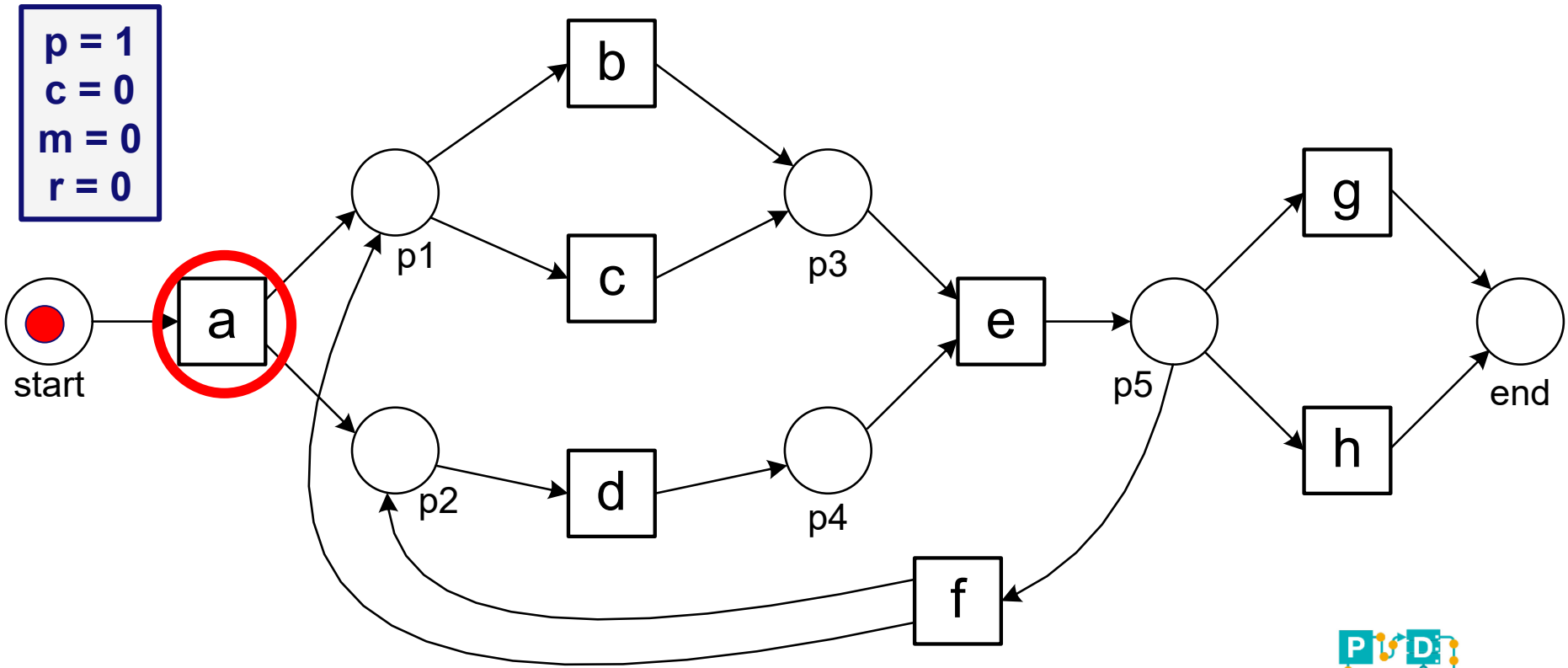
- **In the beginning a token is produced for the source place: p = 1.**
- **At the end a token is consumed from the sink place (also if not there): c' = c + 1.**

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



Chair of Process and Data Science

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

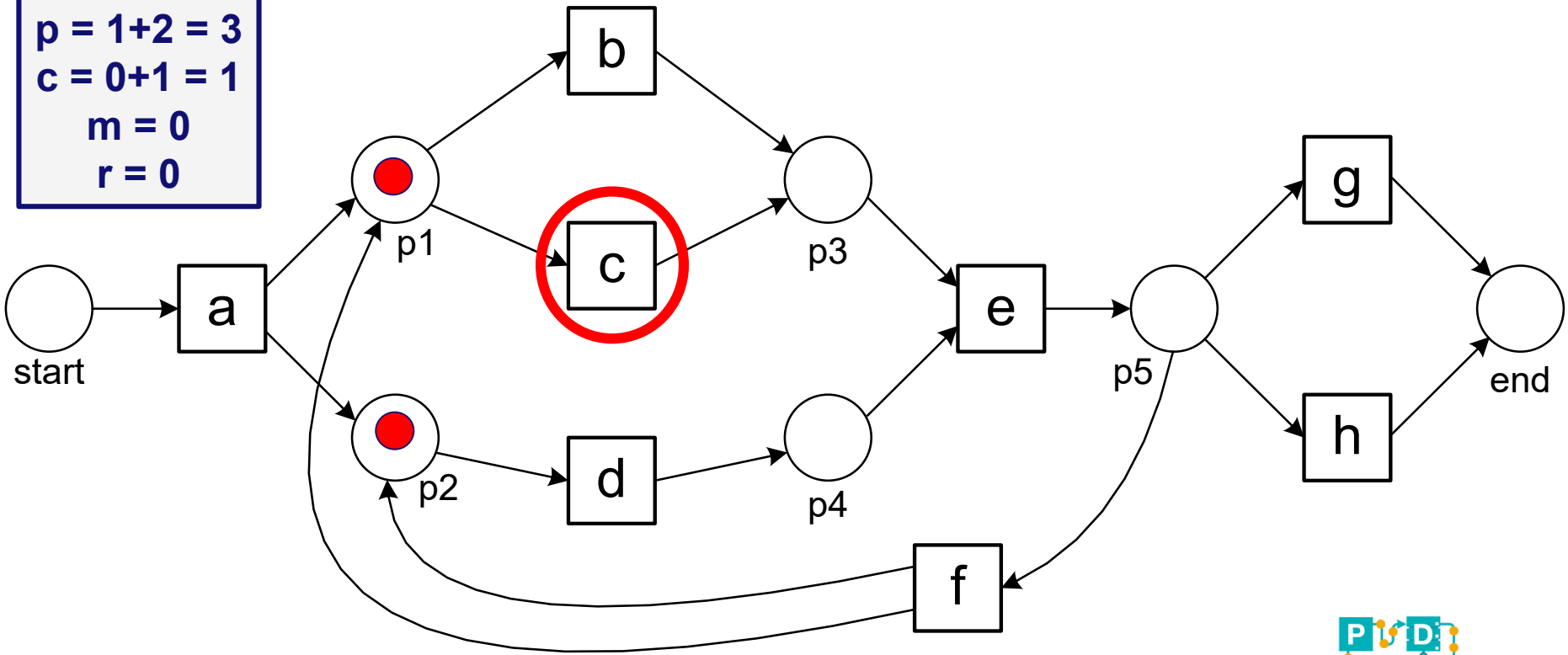


Chair of Process and Data Science

Replaying

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

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

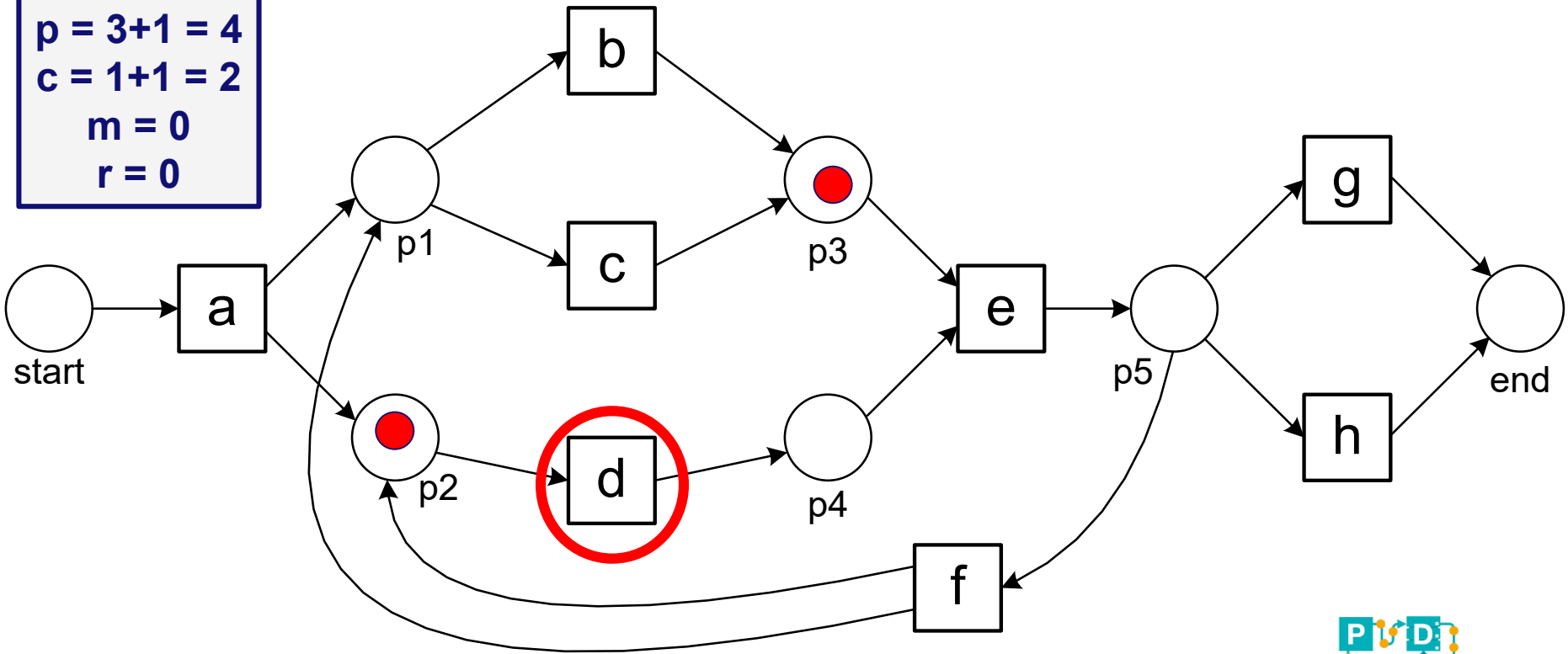


Chair of Process and Data Science

Replaying

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

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

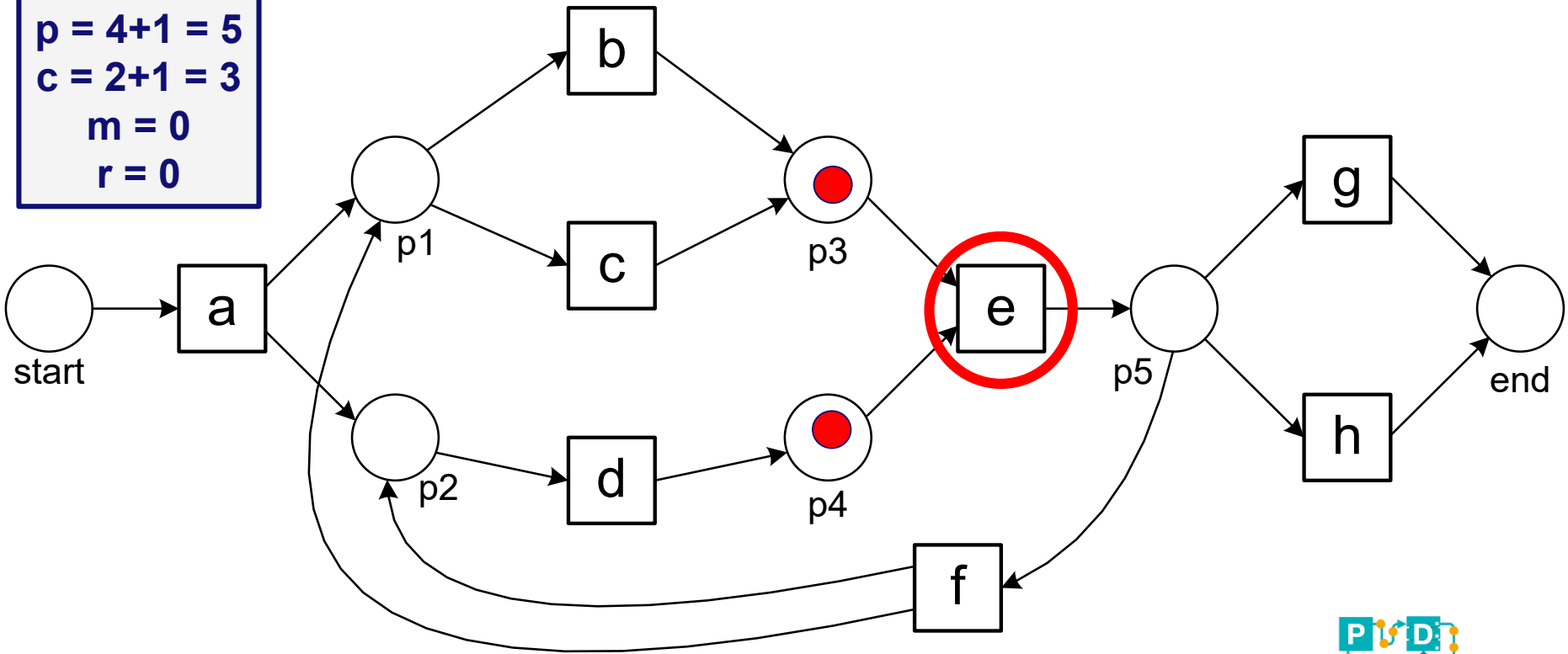


Chair of Process and Data Science

Replaying

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

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

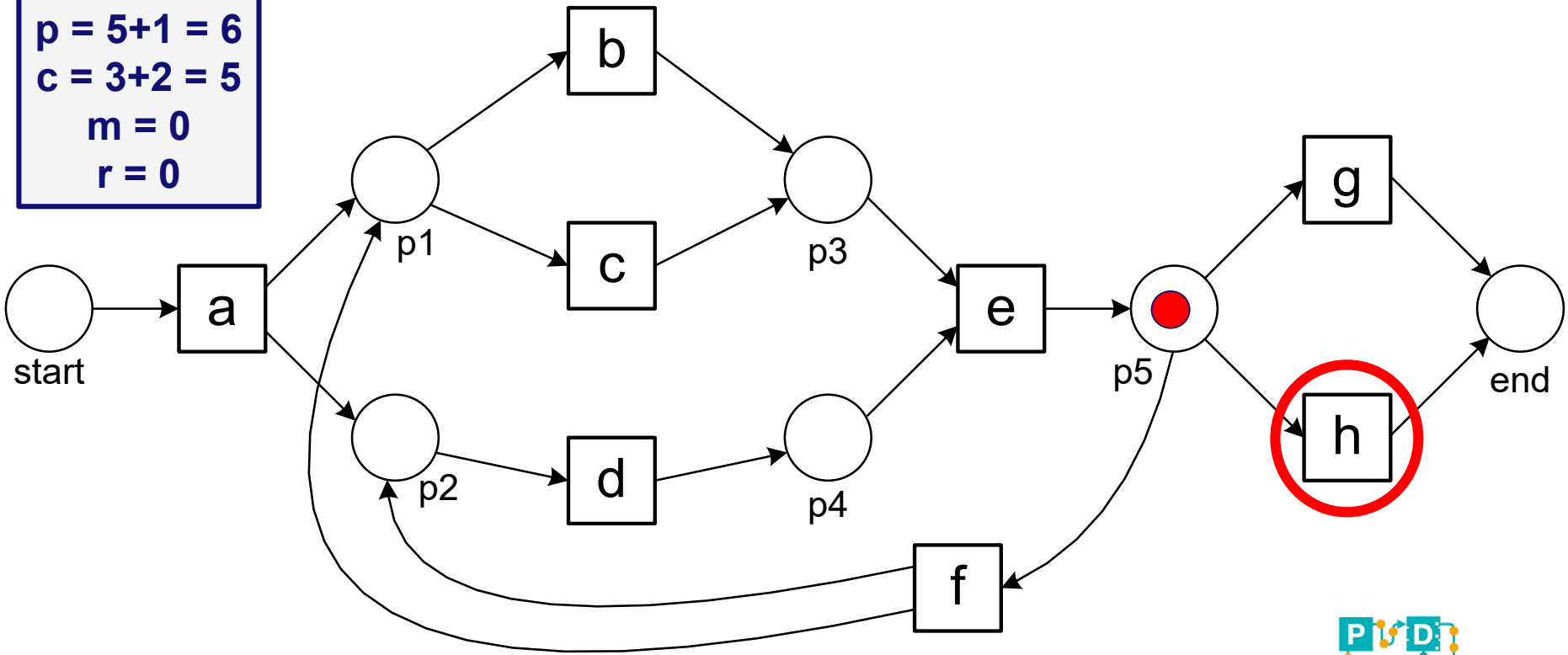


Chair of Process and Data Science

Replaying

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

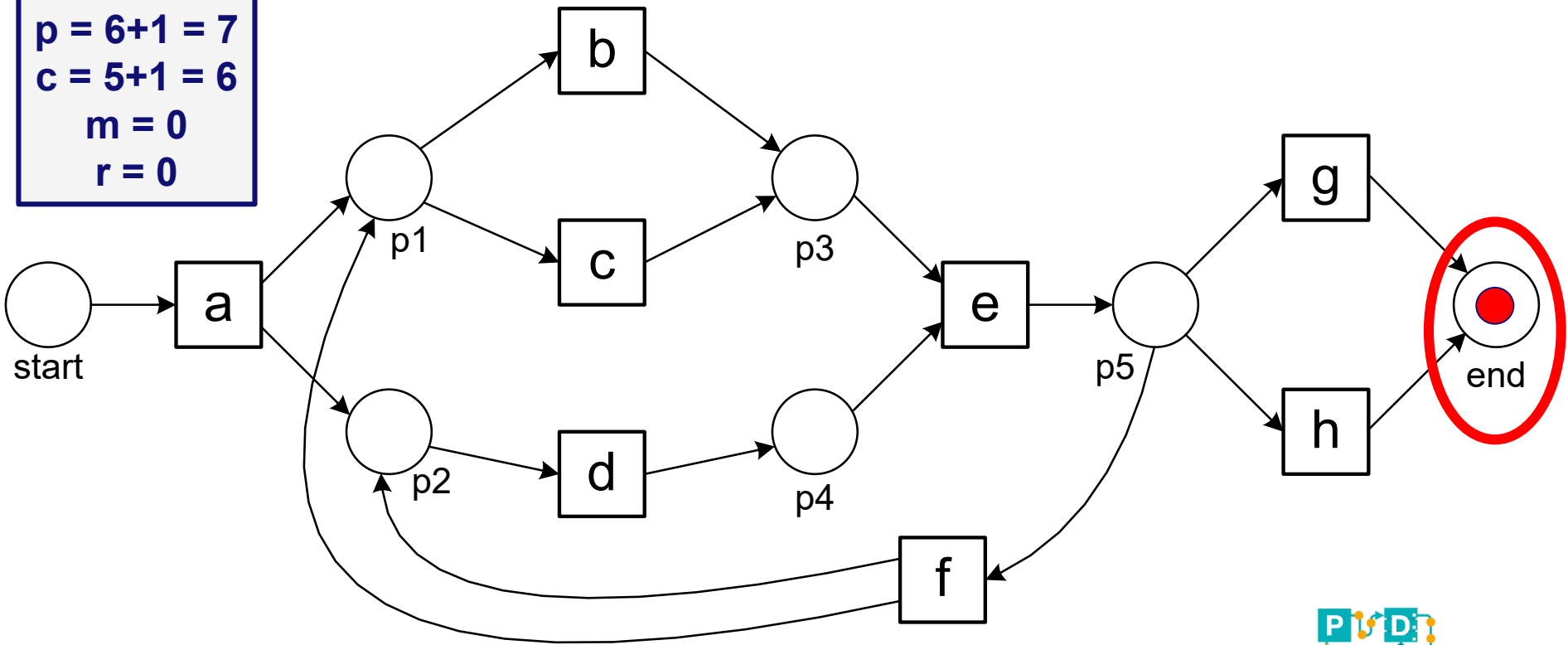
p = 5+1 = 6
c = 3+2 = 5
m = 0
r = 0



Chair of Process and Data Science

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

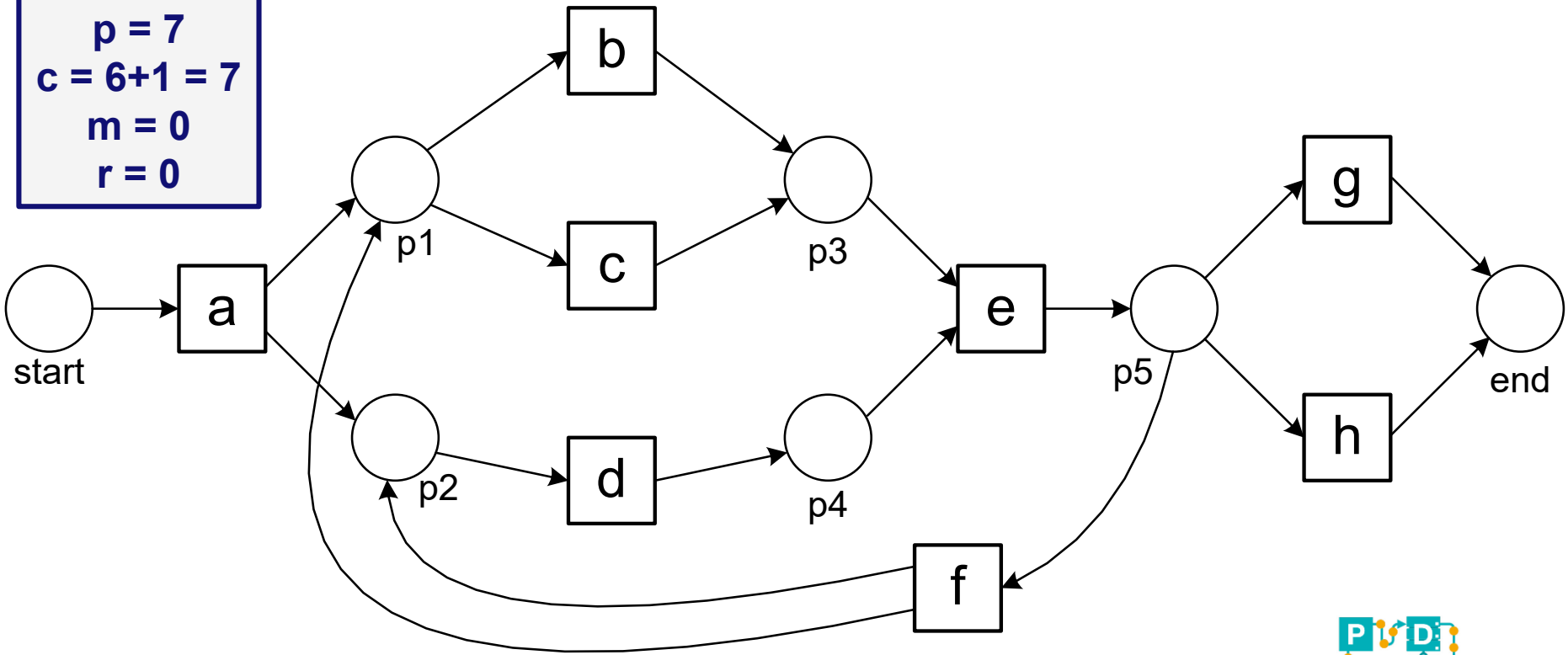
p = 6+1 = 7
c = 5+1 = 6
m = 0
r = 0



Chair of Process and Data Science

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

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

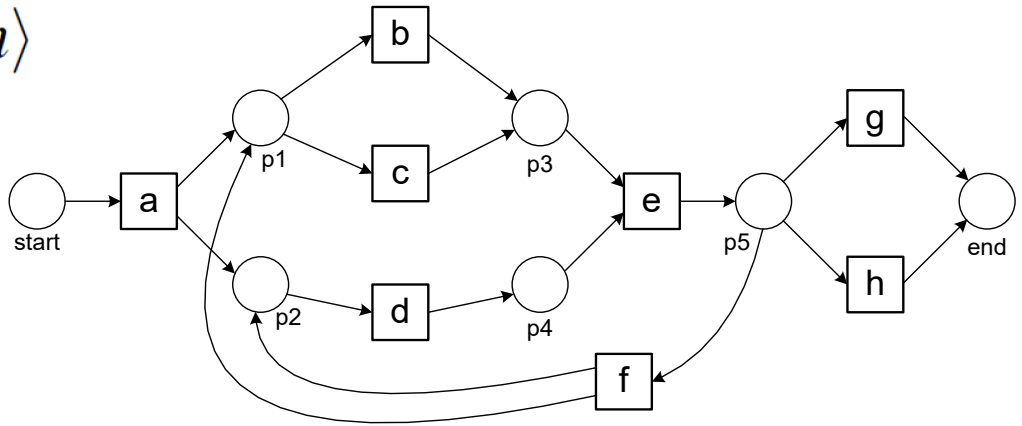


Chair of Process and Data Science

Quantifying fitness at the trace level

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

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



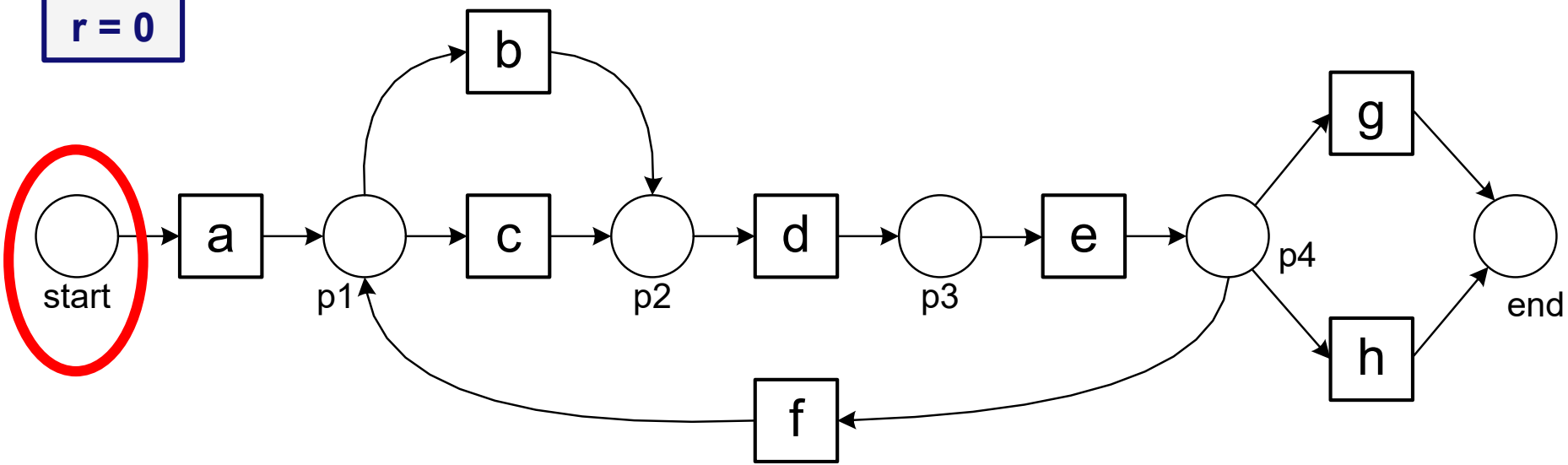
$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{0}{7}\right) + \frac{1}{2}\left(1 - \frac{0}{7}\right) = 1$$

Chair of Process and Data Science

Replaying

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

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

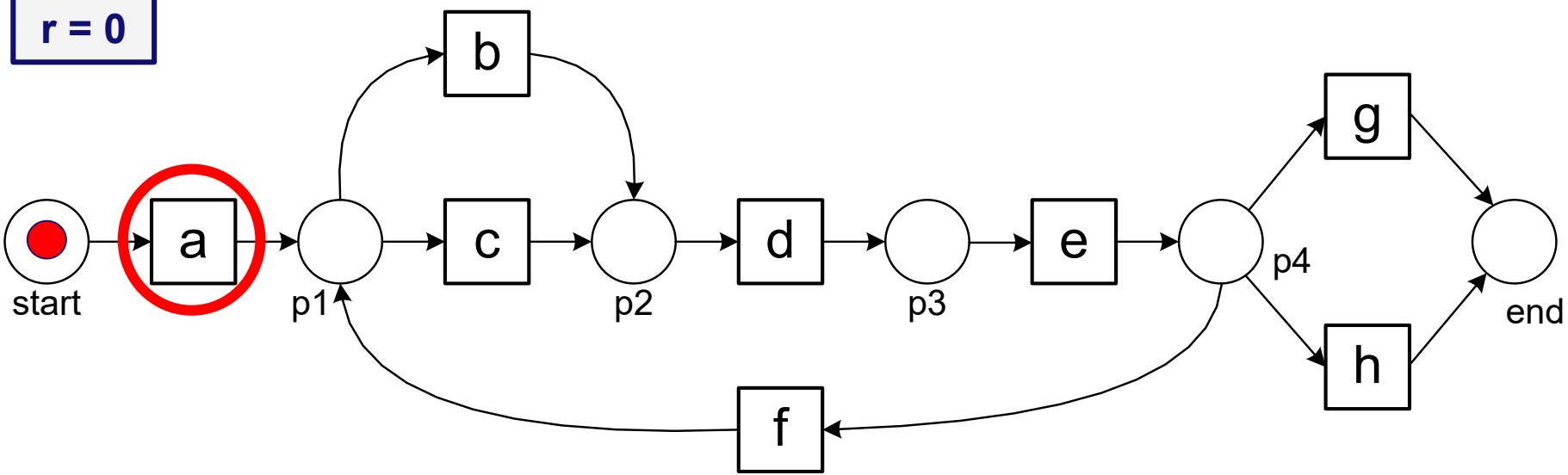


Chair of Process and Data Science

Replaying

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

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

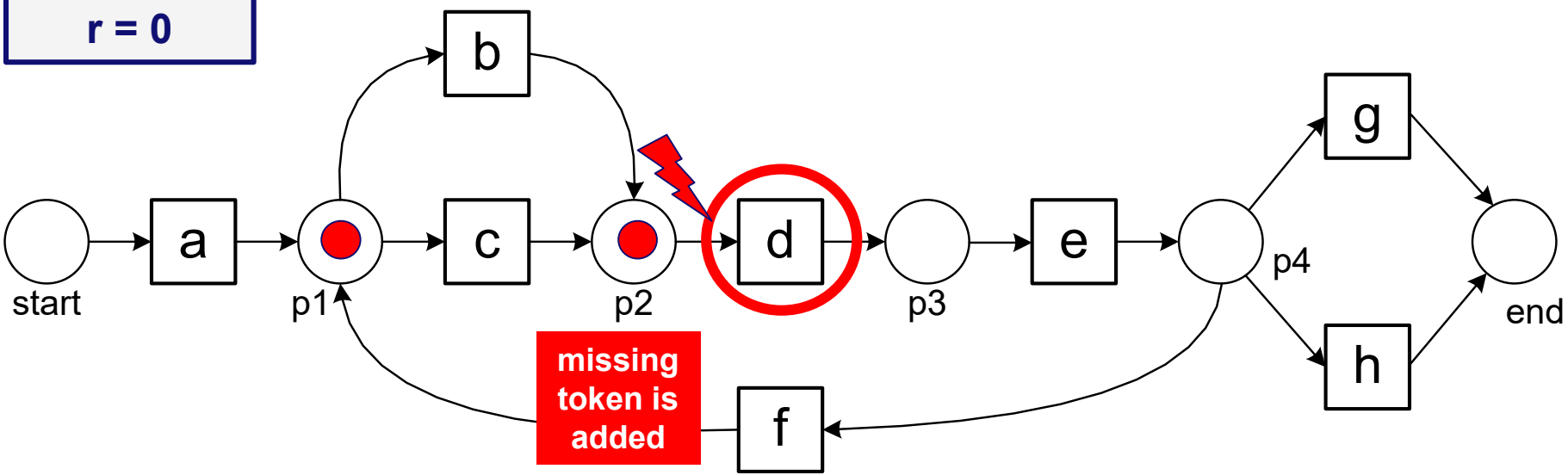


Chair of Process and Data Science

Replaying

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

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

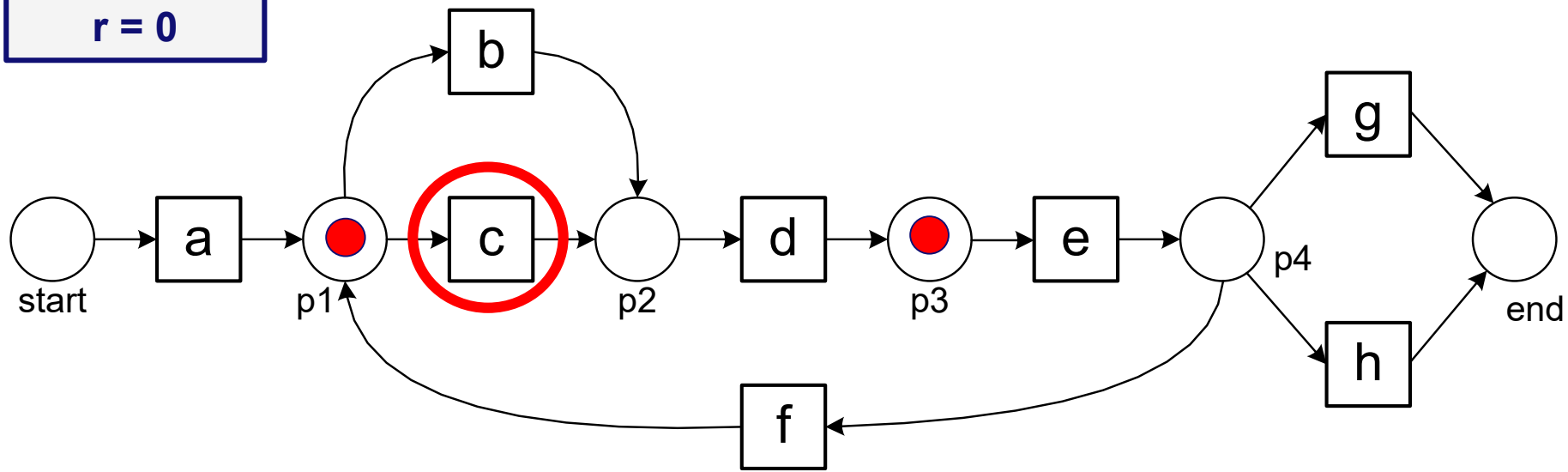


Chair of Process and Data Science

Replaying

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

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

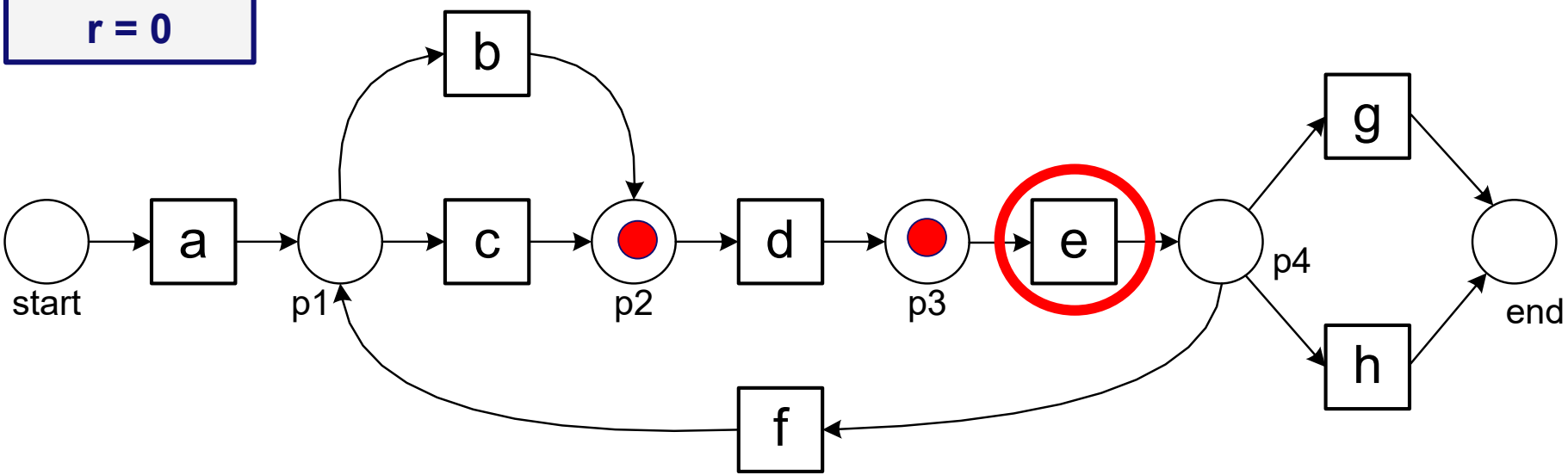


Chair of Process and Data Science

Replaying

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

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

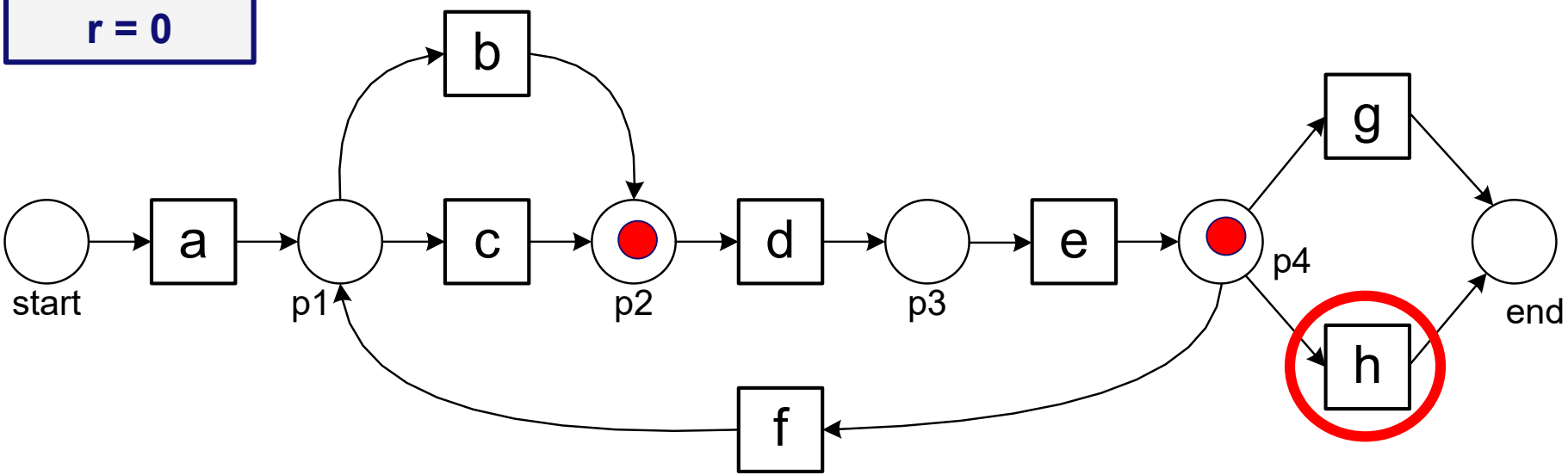


Chair of Process and Data Science

Replaying

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

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

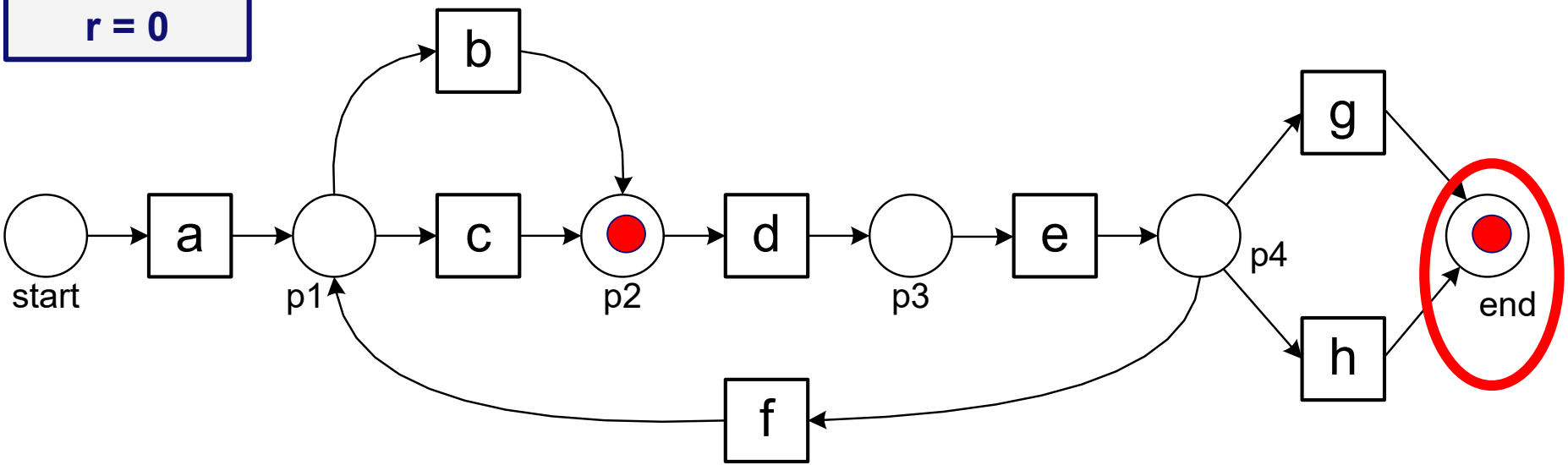


Chair of Process and Data Science

Replaying

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

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

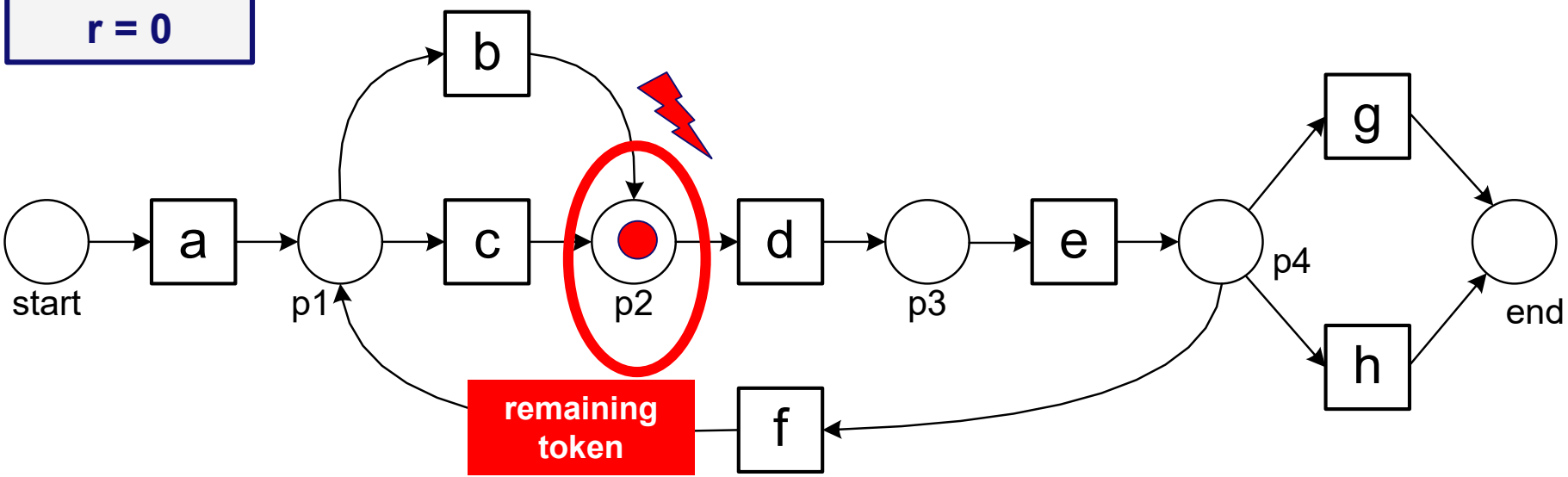


Chair of Process and Data Science

Replaying

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

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

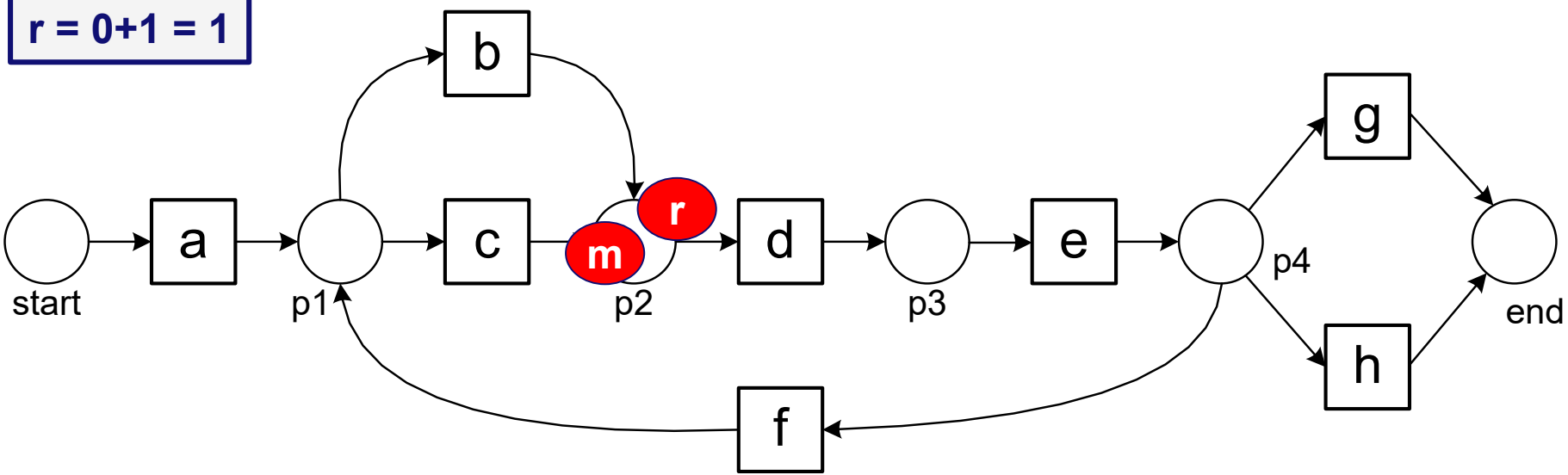


Chair of Process and Data Science

Replaying

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

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

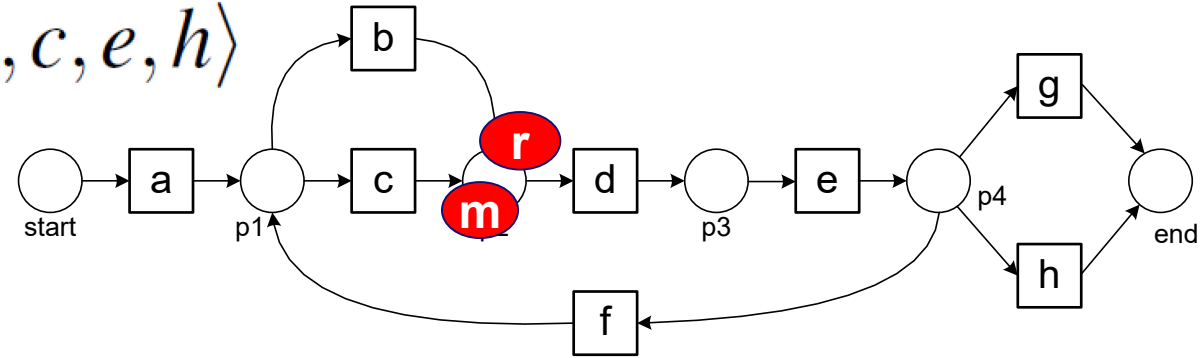


Chair of Process and Data Science

Quantifying fitness at the trace level

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

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



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

Chair of Process and Data Science

Fitness at the log level

missing tokens

consumed tokens

remaining tokens

produced tokens

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

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

Chair of Process and Data Science

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

?

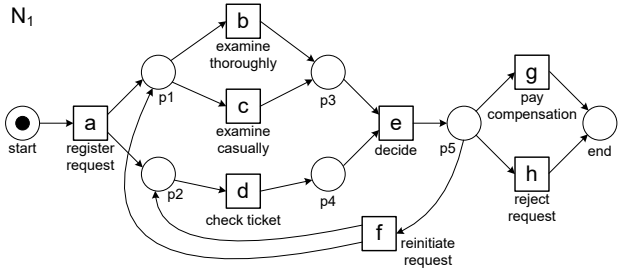
$$fitness(L,N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}}\right)$$

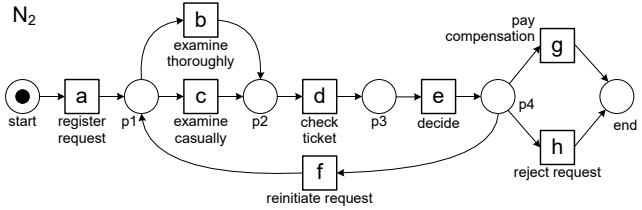
$$fitness(L_{full}, N_1) = 1$$

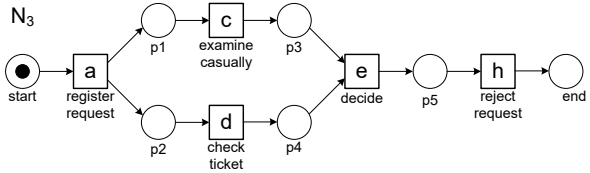
$$fitness(L_{full}, N_2) = 0.9504$$

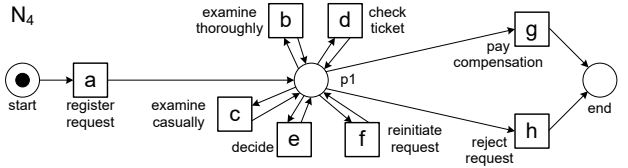
$$fitness(L_{full}, N_3) = 0.8797$$

$$fitness(L_{full}, N_4) = 1$$

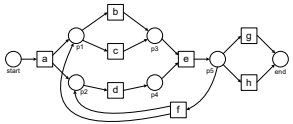








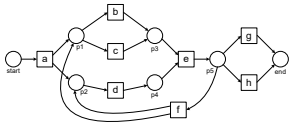
Diagnostics



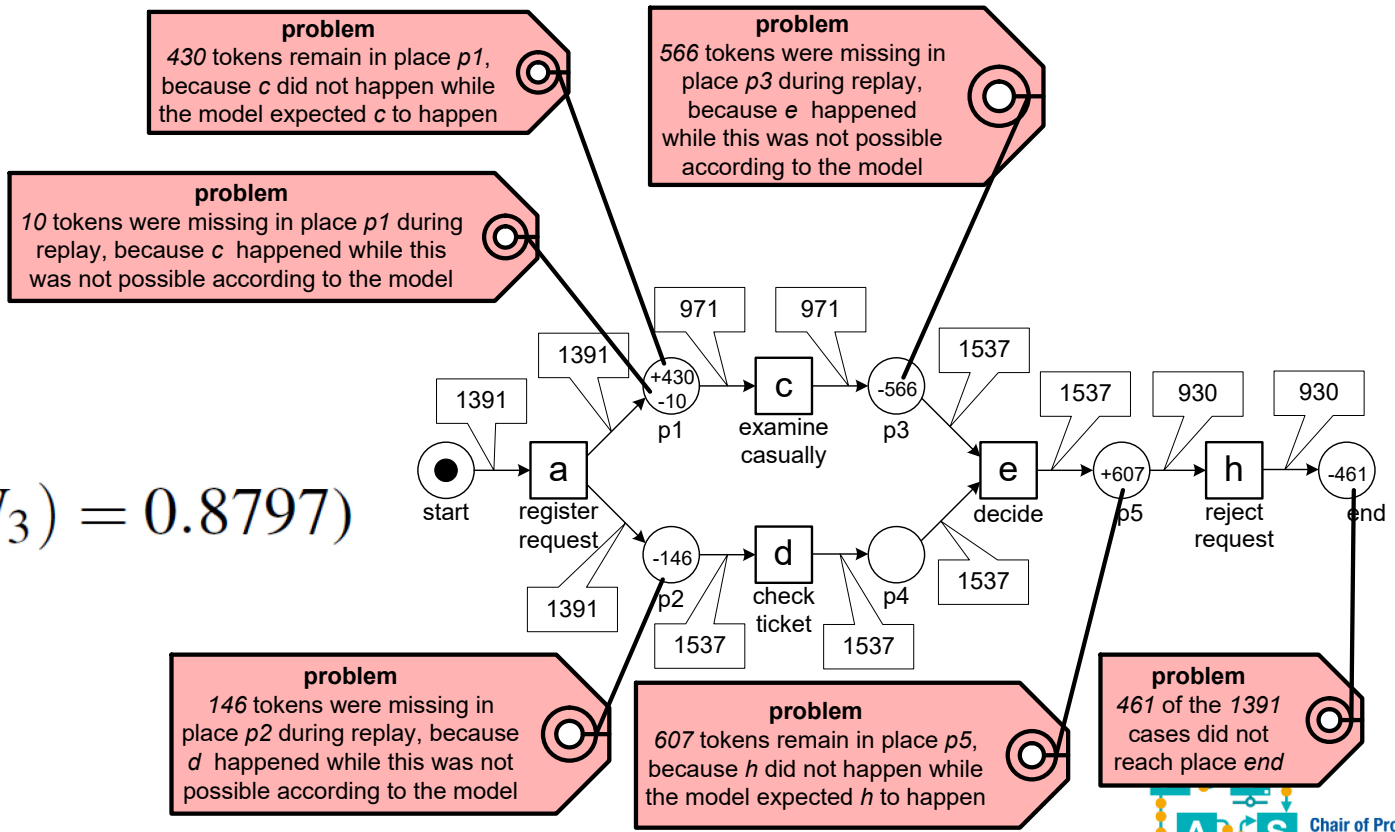
$$(fitness(L_{full}, N_2) = 0.9504)$$

Chair of Process and Data Science

Diagnostics



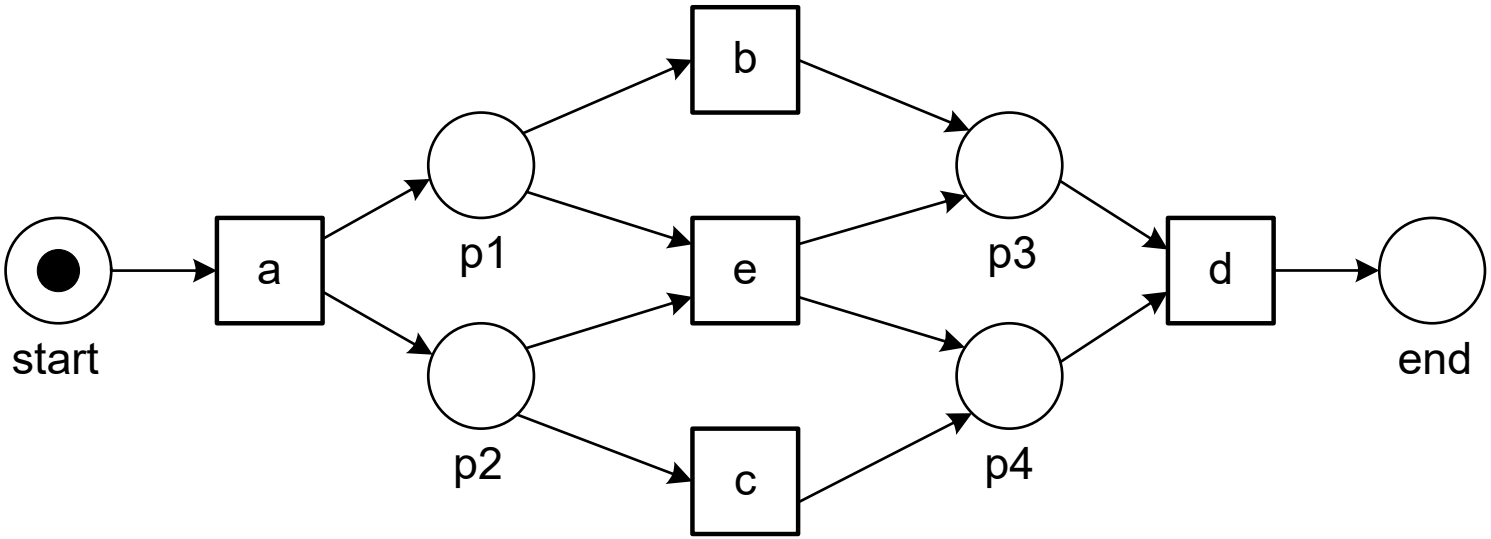
$$(fitness(L_{full}, N_3) = 0.8797)$$



Question (may take some time)

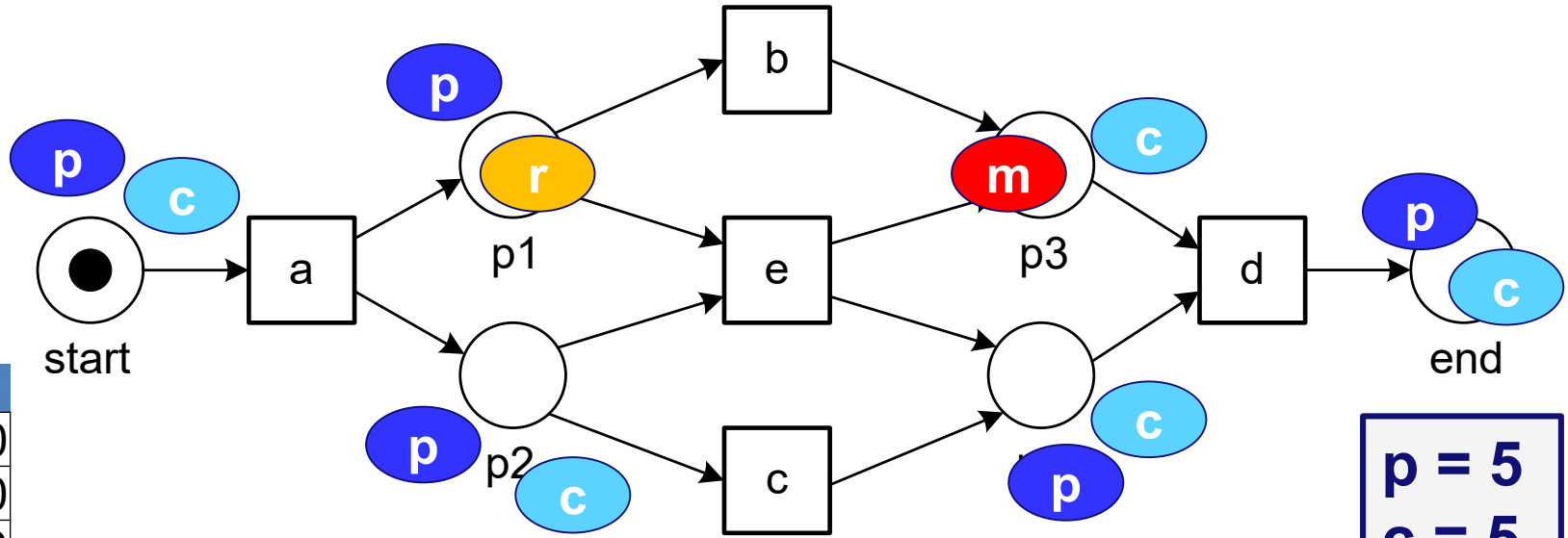
Compute fitness using missing and remaining tokens

trace	frequency
abcd	10
acbd	10
aed	10
abd	2
acd	1
ad	1
abbd	1



- **Consider the event log containing 35 cases.**
- **What is the fitness?**

Let us pick one trace: acd



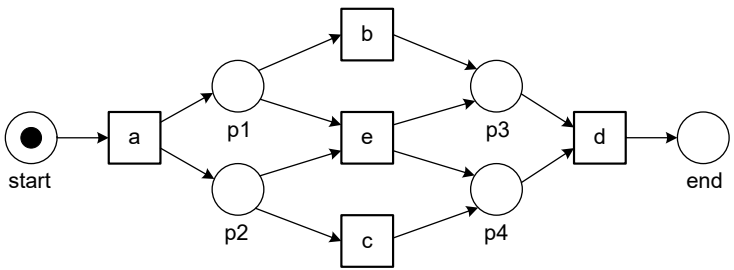
trace	frequency
abcd	10
acbd	10
aed	10
abd	2
acd	1
ad	1
abbd	1

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

p = 5
c = 5
m = 1
r = 1

Fitness = 0.9658

trace	frequency	produced tokens (p)	remaining tokens (r)	consumed tokens (c)	missing tokens (m)	produced tokens (all)	remaining tokens (all)	consumed tokens (all)	missing tokens (all)
abcd	10	6	0	6	0	60	0	60	0
acbd	10	6	0	6	0	60	0	60	0
aed	10	6	0	6	0	60	0	60	0
abd	2	5	1	5	1	10	2	10	2
acd	1	5	1	5	1	5	1	5	1
ad	1	4	2	4	2	4	2	4	2
abbd	1	6	2	6	2	6	2	6	2



205	7	205	7
sum p	sum r	sum c	sum m

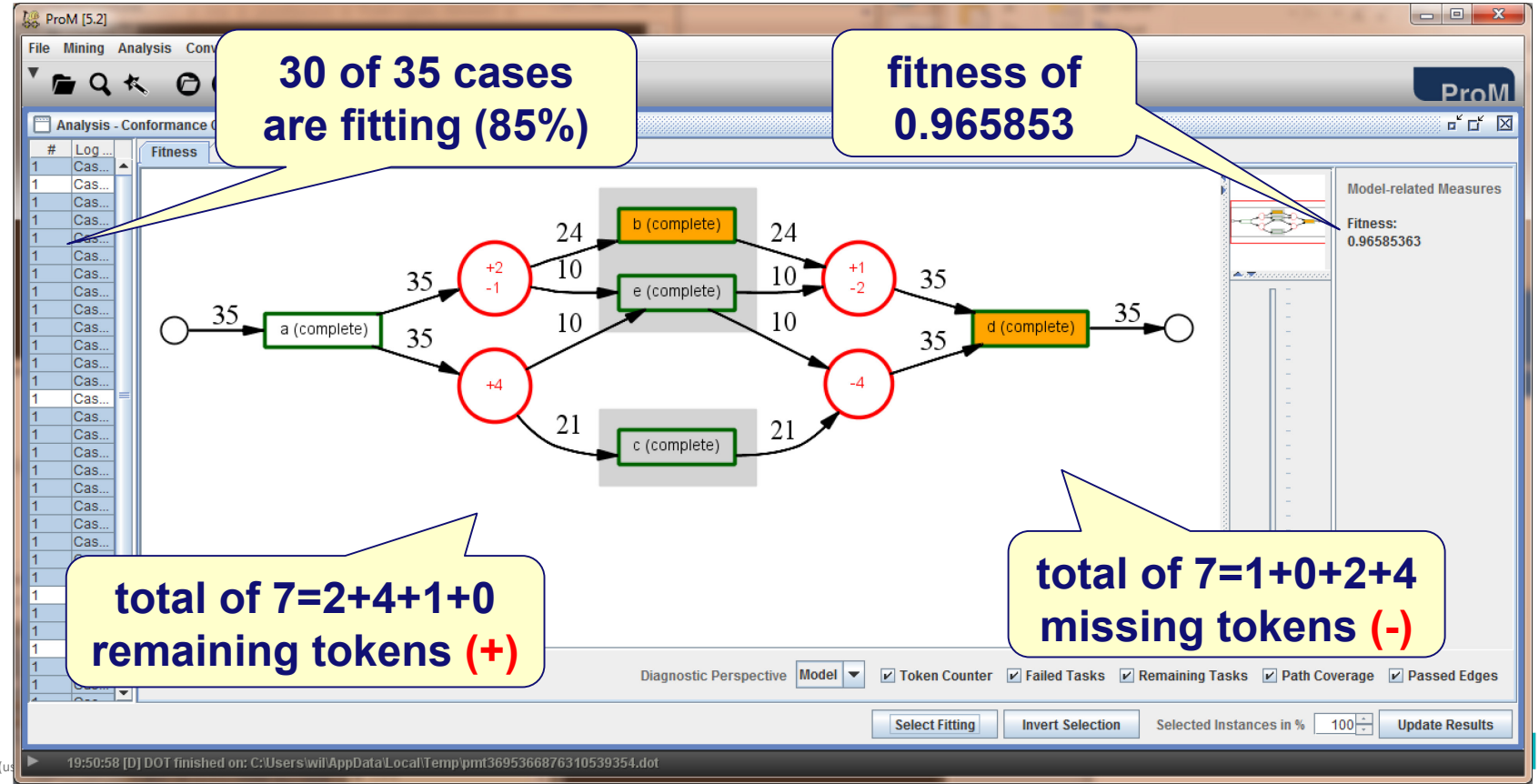
fitness	0.965853659

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

Chair of Process and Data Science

ProM 5.2 output

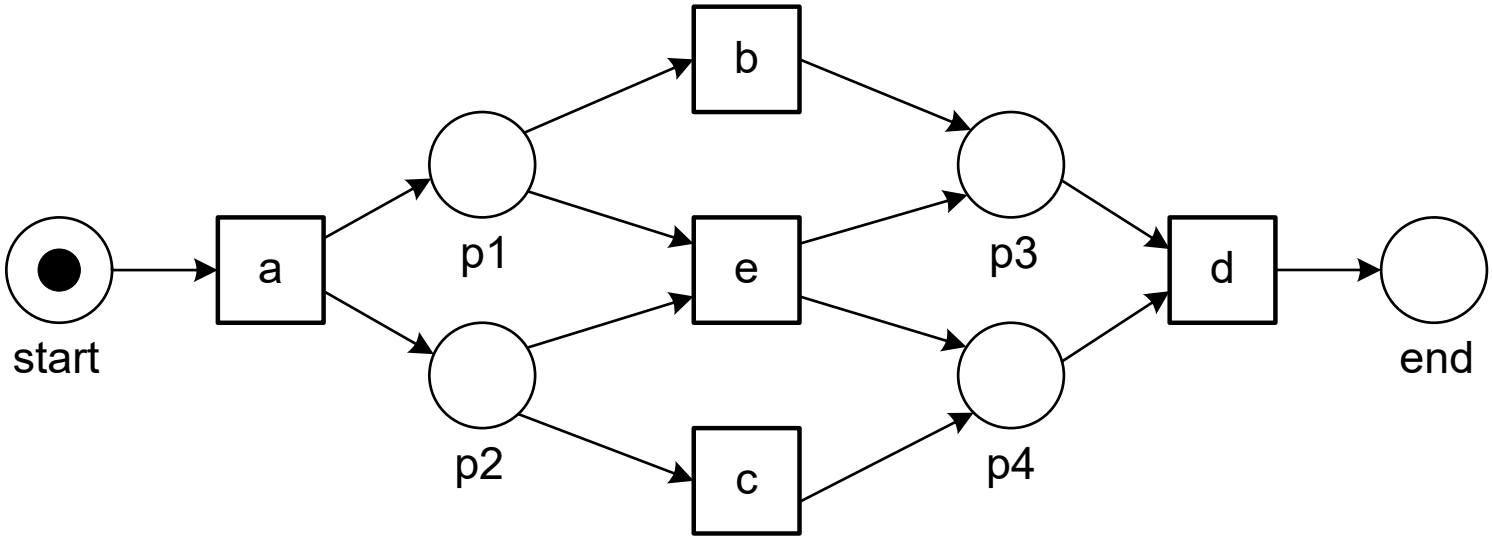
(ProM 6 only supports more advanced conformance checking techniques)



Question

Compute fitness using missing and remaining tokens

trace	frequency
e	1

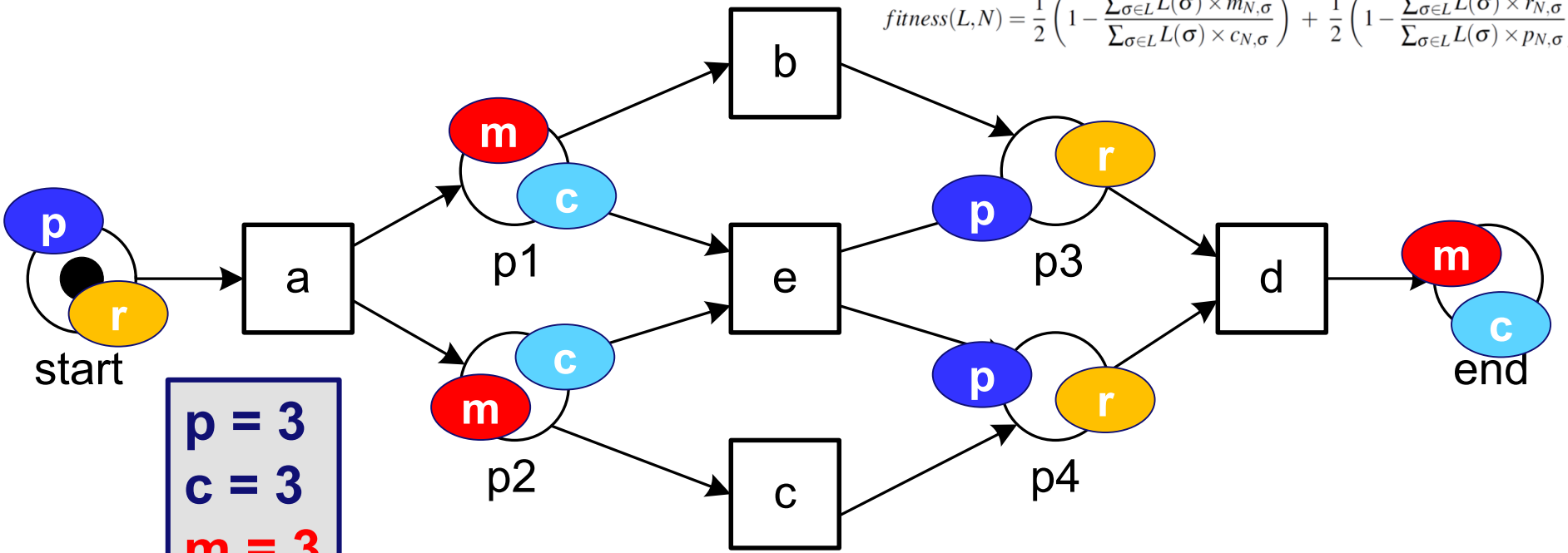


- **Consider the event log containing just one case: L = [⟨e⟩].**
- **What is the fitness (using token-based replay)?**

Chair of Process and Data Science

Answer obtained by replaying ⟨e⟩

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



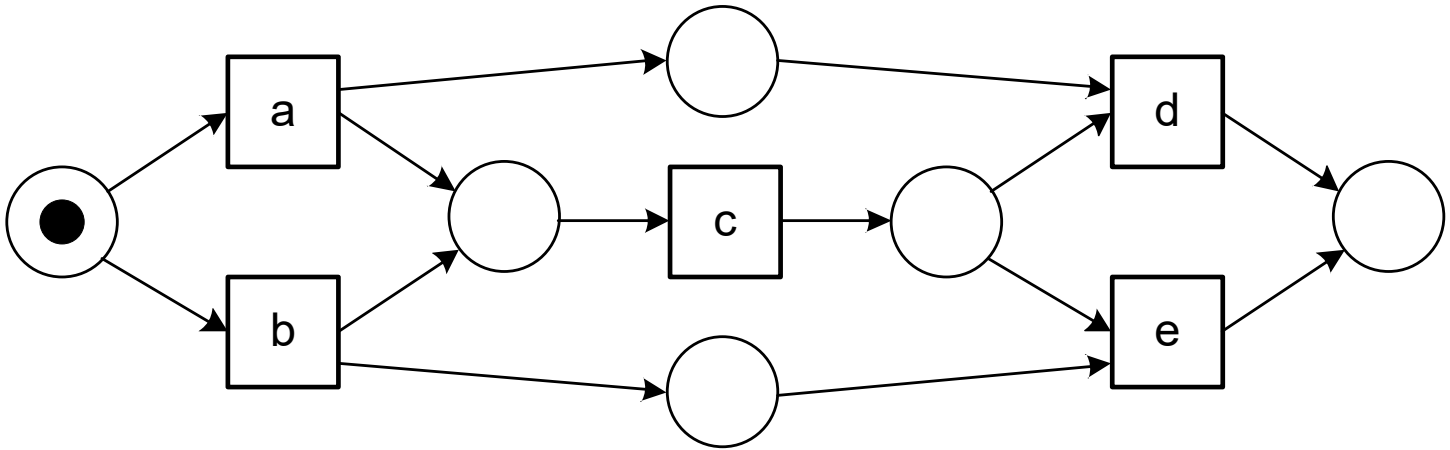
p = 3
c = 3
m = 3
r = 3

fitness = 0 !

Chair of Process and Data Science

Another example

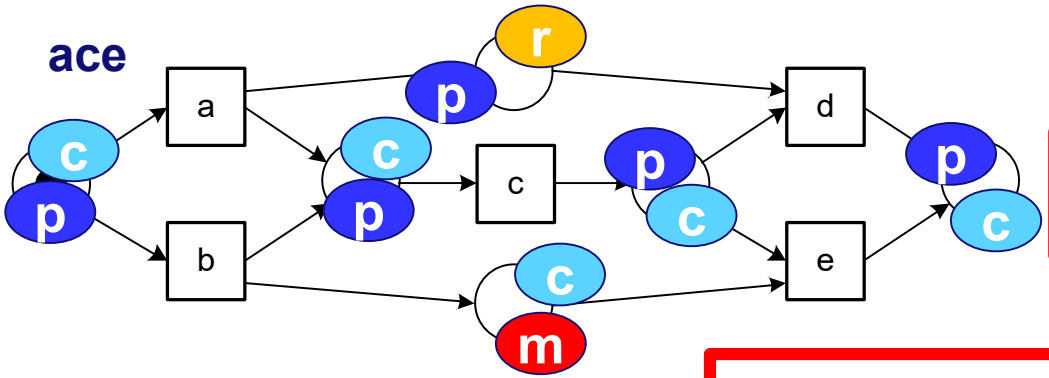
trace	frequency
acd	10
bce	10
ace	5
bcd	5
dca	1
abd	1
d	1



- **Consider the event log containing 33 cases.**
- **What is the fitness?**

Fitness = 0.895705521

trace	frequency	produced tokens (p)	remaining tokens (r)	consumed tokens (c)	missing tokens (m)	produced tokens (all)	remaining tokens (all)	consumed tokens (all)	missing tokens (all)
acd	10	5	0	5	0	50	0	50	0
bce	10	5	0	5	0	50	0	50	0
ace	5	5	1	5	1	25	5	25	5
bcd	5	5	1	5	1	25	5	25	5
dca	1	5	3	5	3	5	3	5	3
abd	1	6	3	5	2	6	3	5	2
d	1	2	1	3	2	2	1	3	2

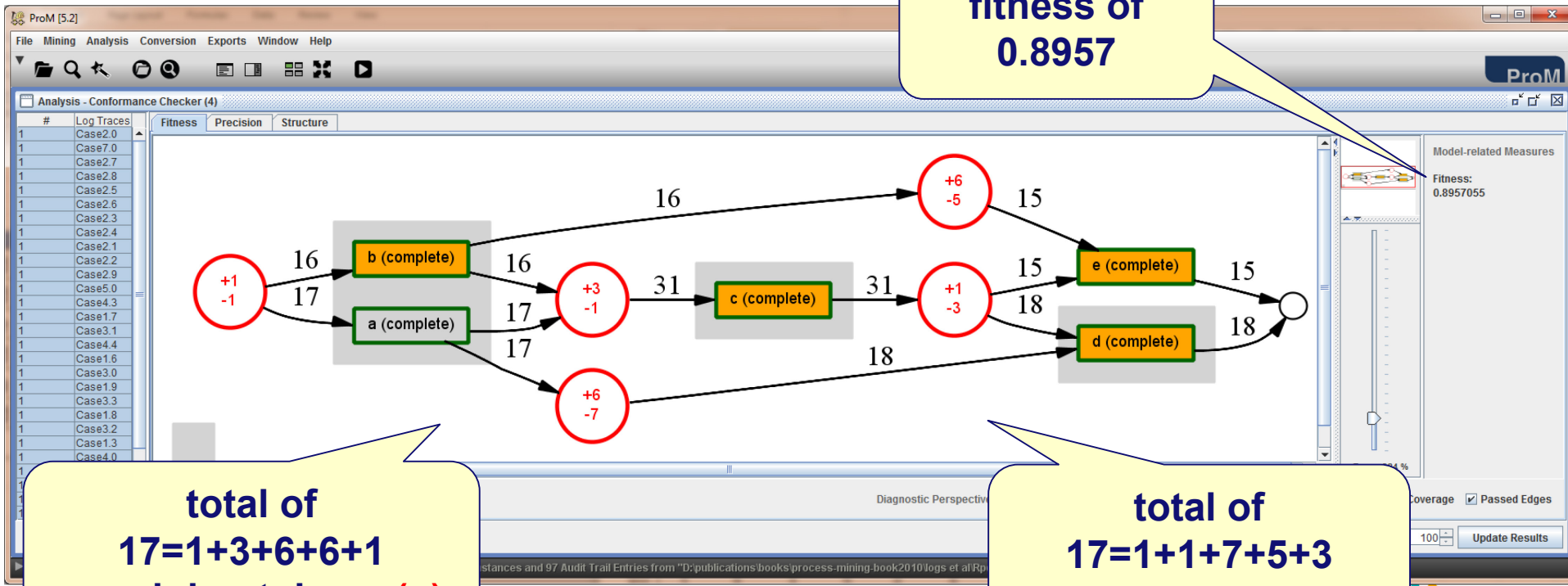


163	17	163	17
sum p	sum r	sum c	sum m

fitness **0.895705521**

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

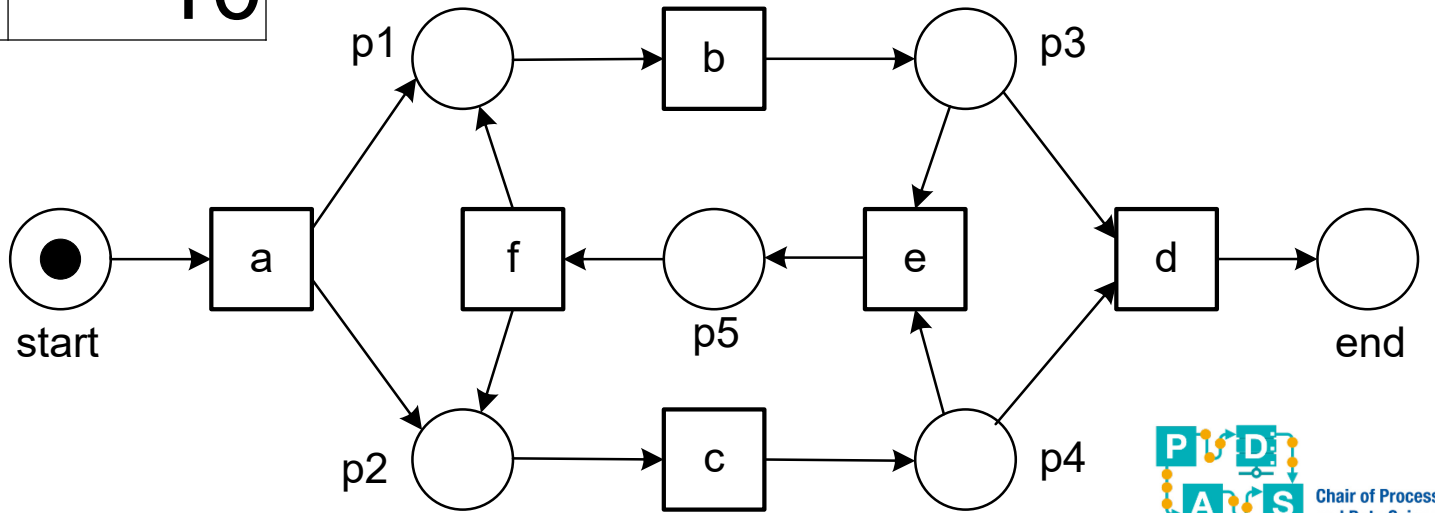
ProM 5.2 diagnostics



Another example

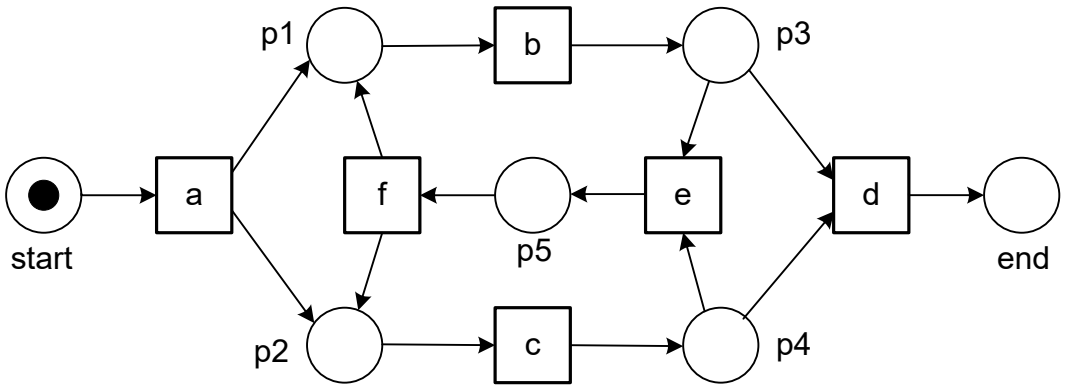
trace	frequency
abefcd	10
abbefccd	10

- **Consider the event log containing 20 cases.**
- **What is the fitness?**



Fitness = 0.8

trace	frequency	produced tokens (p)	remaining tokens (r)	consumed tokens (c)	missing tokens (m)	produced tokens (all)	remaining tokens (all)	consumed tokens (all)	missing tokens (all)
abefcd	10	9	2	9	2	90	20	90	20
abbefccd	10	11	2	11	2	110	20	110	20



200	40	200	40
sum p	sum r	sum c	sum m

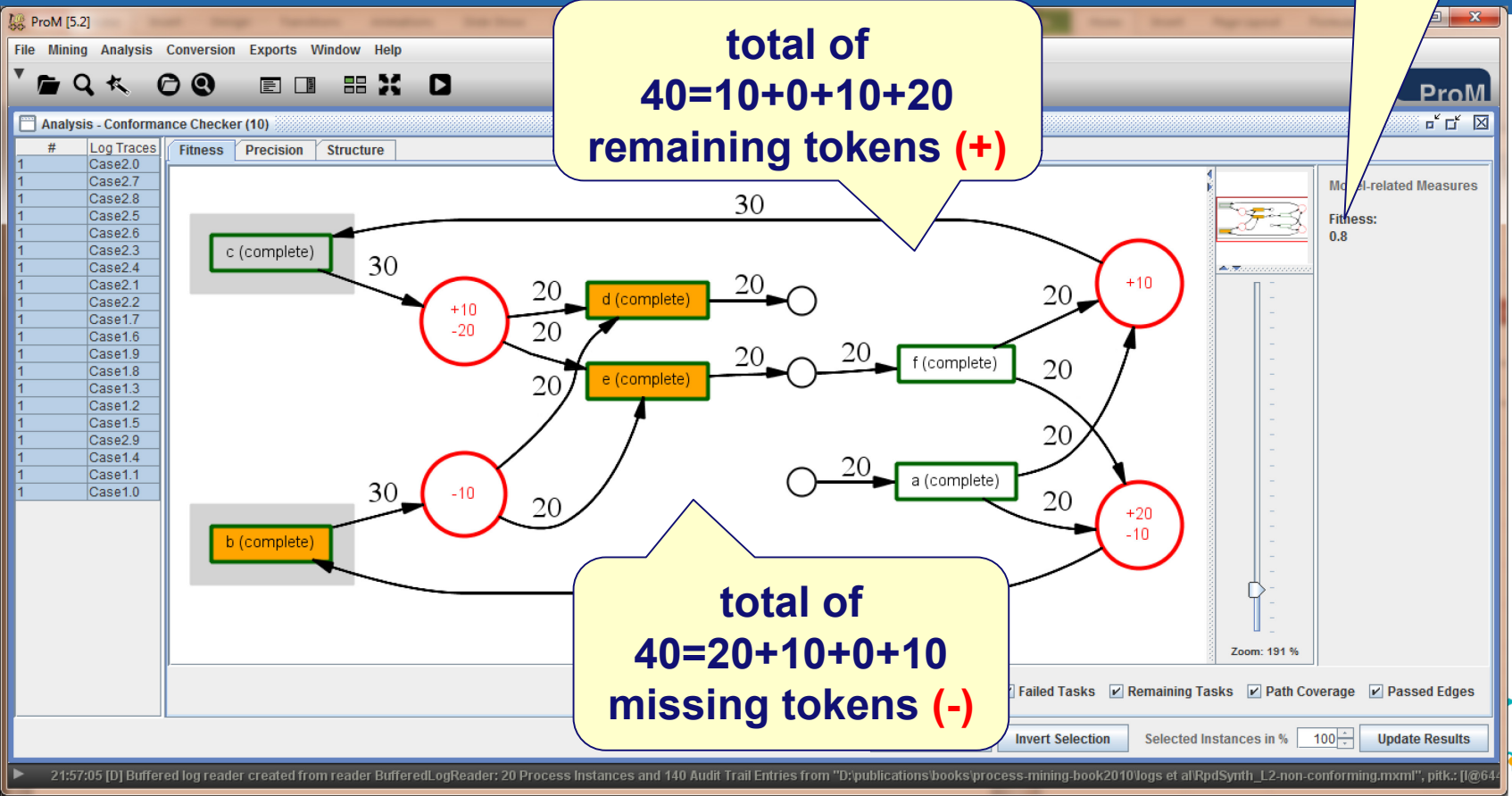
fitness **0.8**

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

Chair of Process and Data Science

ProM 5.2 diagnostics

fitness of 0.8



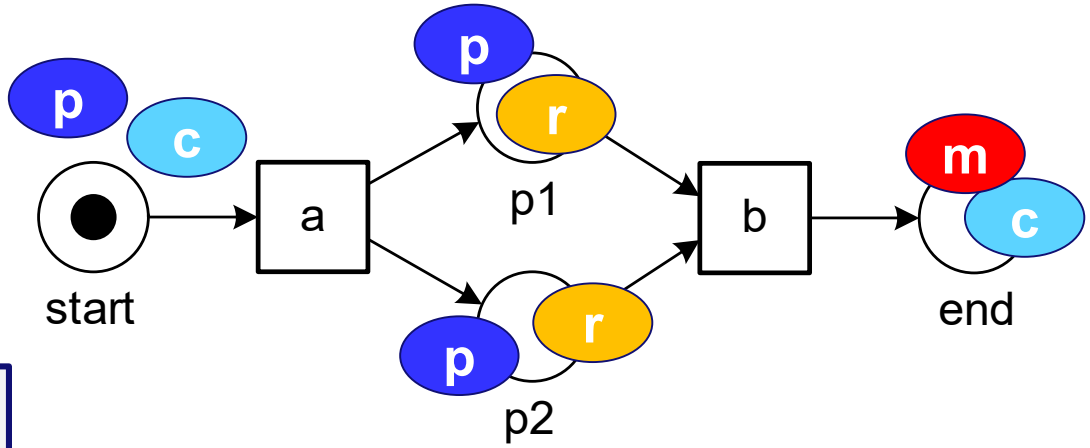
Thus far c=p and m=r. Always?

- **Provide if possible a log and model such that c≠p and m≠r at end.**

- **Hint: Recall that r = p+m-c**

Event log $[\langle a \rangle]$

r = p+m-c



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

$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{1}{2}\right) + \frac{1}{2}\left(1 - \frac{2}{3}\right) = 0.4166$$

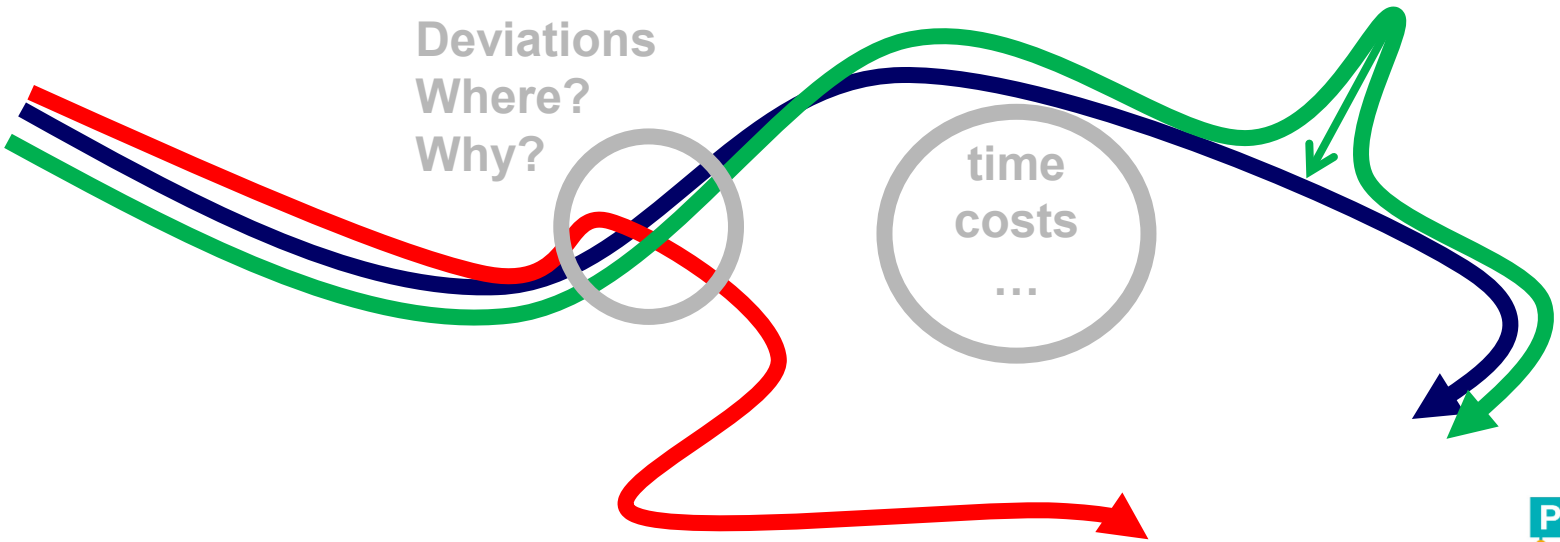
Chair of Process and Data Science

Limitations

- Basic replay approach assumes visible & uniquely labeled transitions.
- ProM / pm4py implementation uses heuristics to deal with silent transitions and multiple transitions having the same label.
- Conformance values sometimes too optimistic due to "token flooding".
- Local decision making may lead to misleading results.

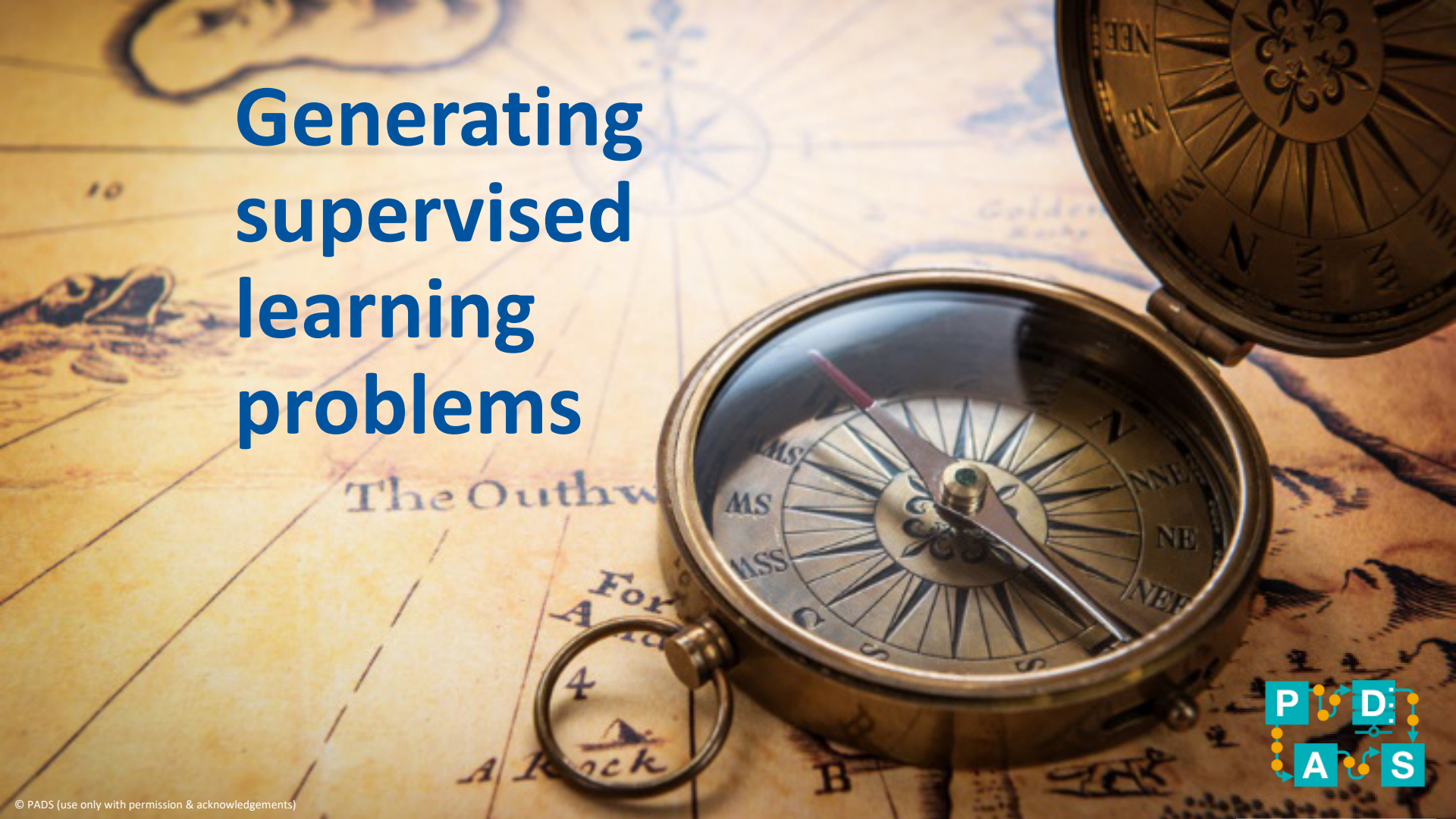
Alignments

- **Outside scope of this course.**
- **Find the “closest path” in the model.**

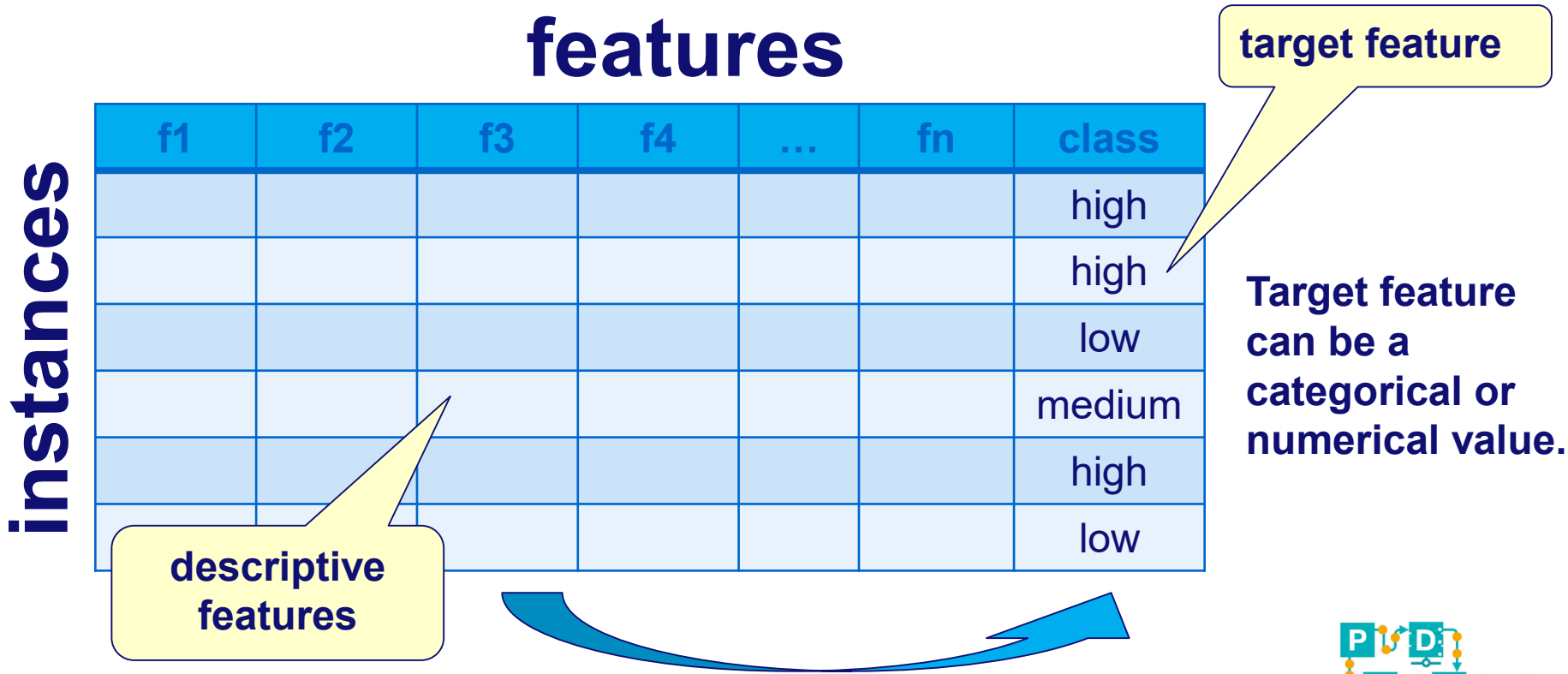


Chair of Process and Data Science

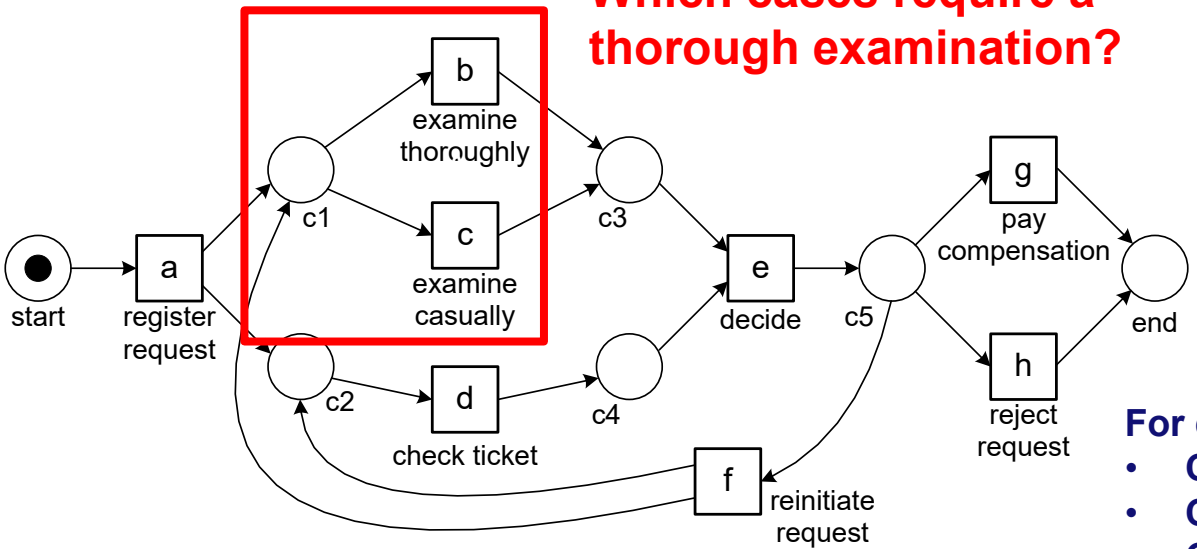
Generating supervised learning problems



Supervised learning problems



Decision mining



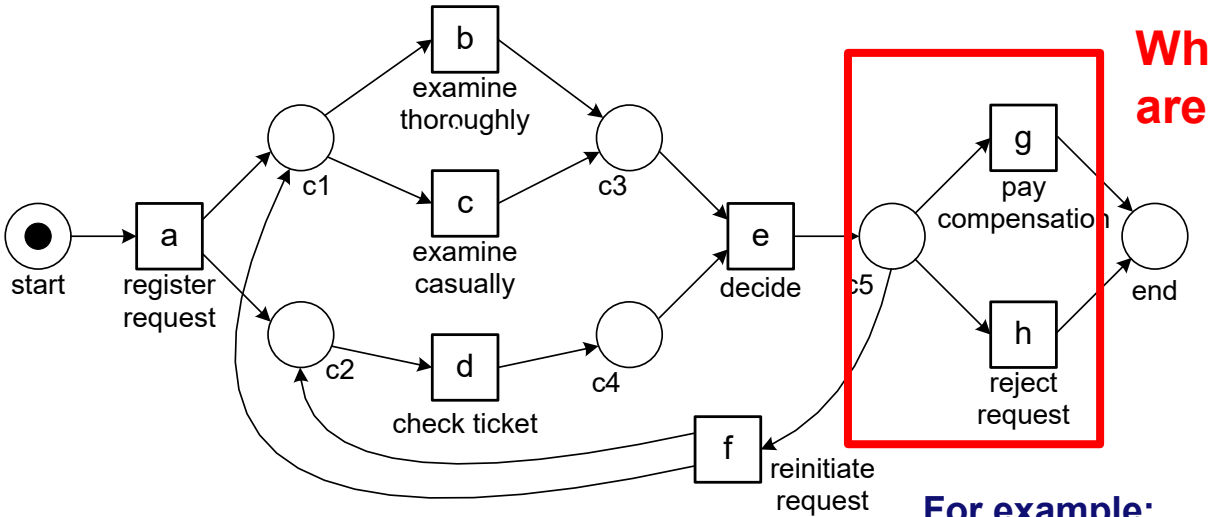
Which cases require a thorough examination?

For example:

- **Cases handled by John.**
- **Cases handled in January.**
- **Cases that were submitted late.**
- **Cases of new customers.**
- **…**

Chair of Process and Data Science

Decision mining



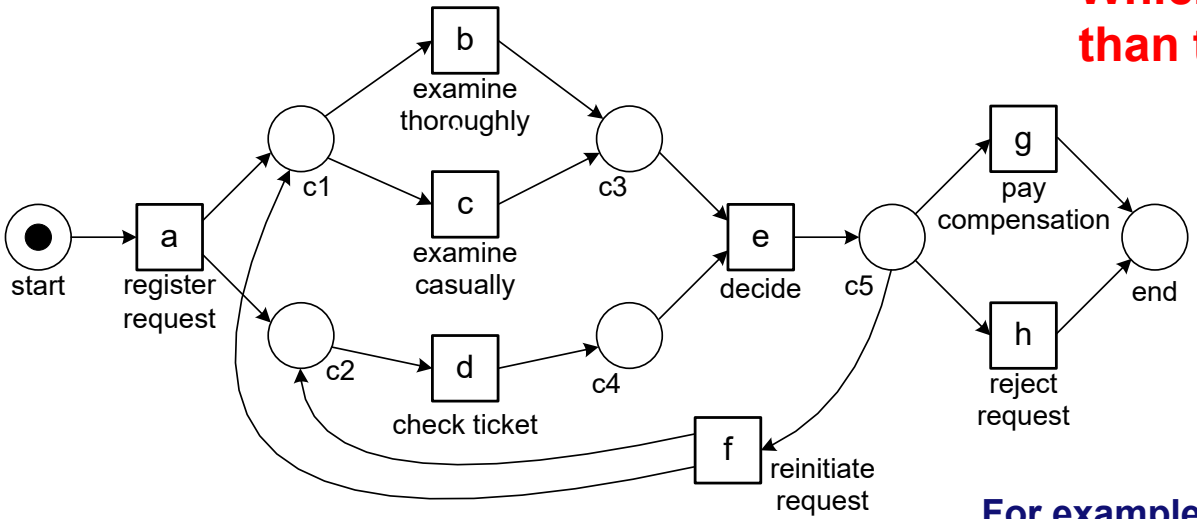
Which cases are rejected?

For example:

- **Cases above €500.**
- **Cases that required multiple checks.**
- **Cases that got delayed.**
- **…**

Chair of Process and Data Science

Performance mining



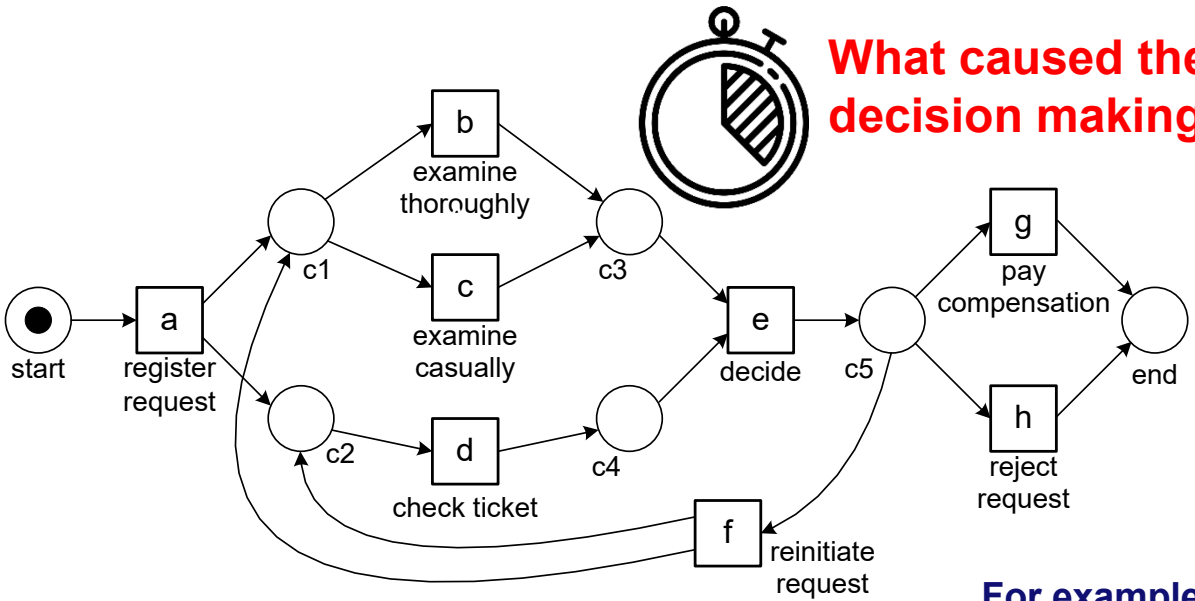
Which cases took more than two months?

For example:

- **Cases handled by Mary.**
- **Cases that required multiple checks.**
- **…**

Chair of Process and Data Science

Performance mining



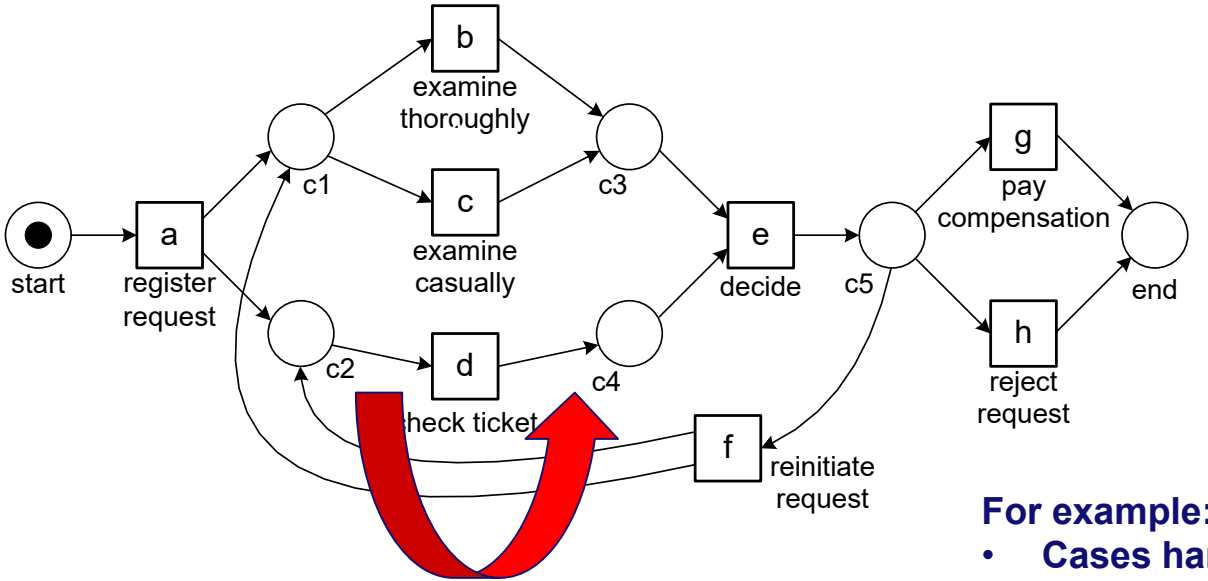
What caused the delays in decision making in May?

For example:

- **A lack of resources due to illness.**
- **An unusual percentage or rework.**
- **…**

Chair of Process and Data Science

Deviation mining



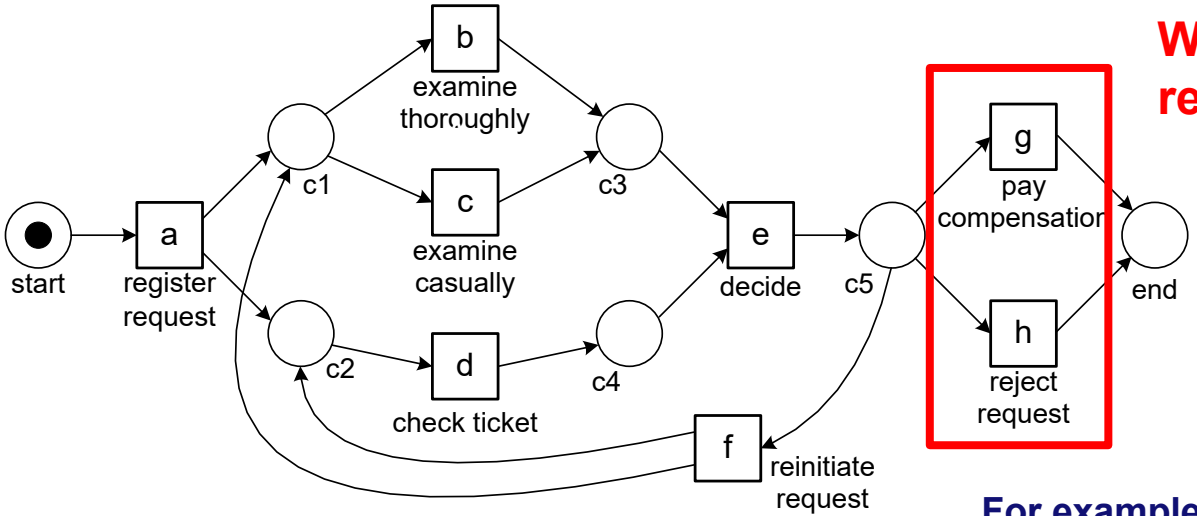
For which cases was the ticket not checked?

For example:

- **Cases handled by Mary.**
- **Cases initiated by the downtown office.**
- **…**

Chair of Process and Data Science

Deviation mining

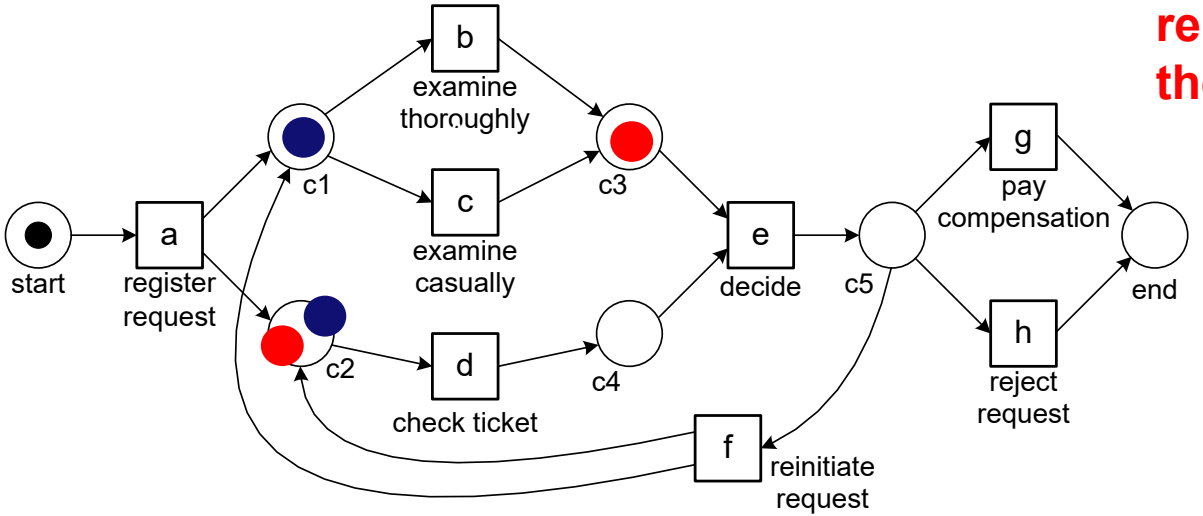


Which cases were rejected an also paid?

For example:

- **Cases handled by Pete.**
- **Cases that arrived in June.**
- **…**

Operational support mining

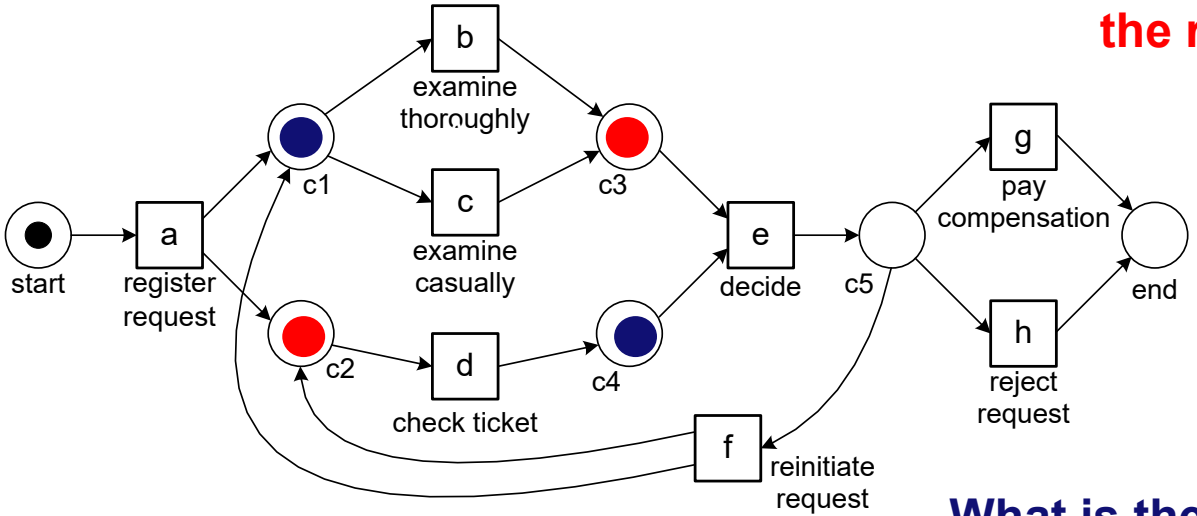


What is the expected remaining flow time of the red case?

What is the expected remaining flow time of the blue case?

Chair of Process and Data Science

Operational support mining



What is the probability that the red case will be rejected?

What is the probability that the blue case will need two decisions?

Chair of Process and Data Science

General pattern

Chair of Process and Data Science

Reminder: Data science is like a making cocktail. Mix the proper ingredients in the right way!

Python
Visualization
Decision trees
Regression
Support vector machines
Neural networks
Evaluation
Clustering
Frequent items sets
Association rules
Sequence mining
Process mining
Process discovery
Conformance checking
Text mining
Preprocessing
Visual analytics
Encryption
Anonymization
Big data infra
Distribution

Decision trees

Conformance checking

Process discovery

Preprocessing

Python
Visualization
Decision trees
Regression
Support vector machines
Neural networks
Evaluation
Clustering
Frequent items sets
Association rules
Sequence mining
Process mining
Process discovery
Conformance checking
Text mining
Preprocessing
Visual analytics
Encryption
Anonymization
Big data infra
Distribution

Process discovery

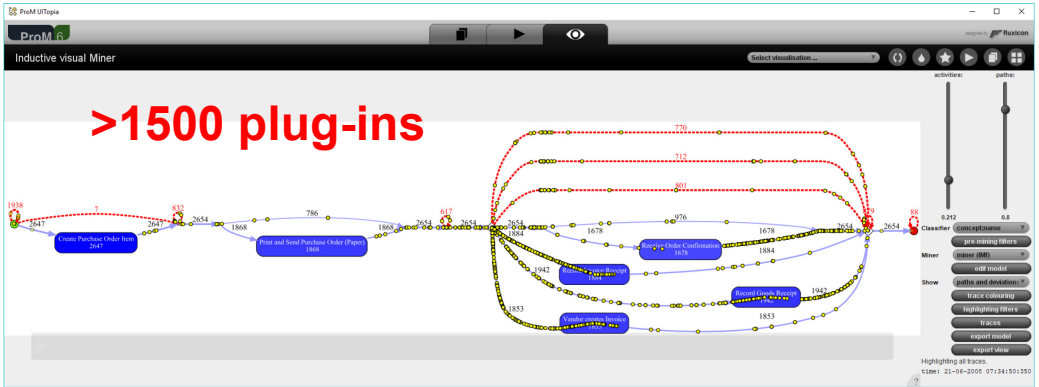
Clustering

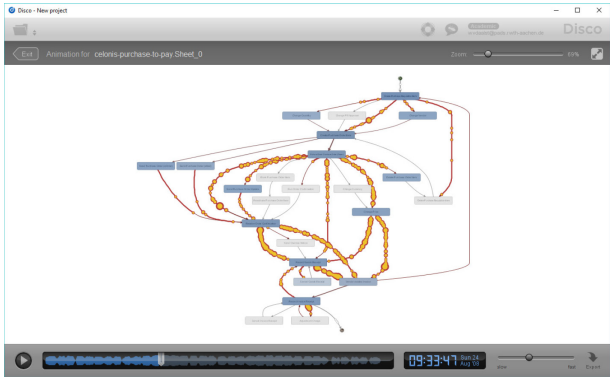
Preprocessing

Tooling & Ecosystem

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.).

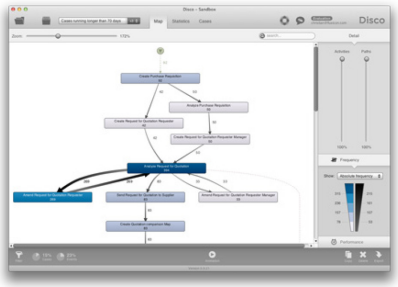




Disco

PROCESSGOLD

MPM
QLIK® INSIDE



celonis
process mining

Rialto
by Exura

minit

OnBase
by Hyland

perceptivesoftware
a Lexmark company

QPR
Quality. Processes. Results.

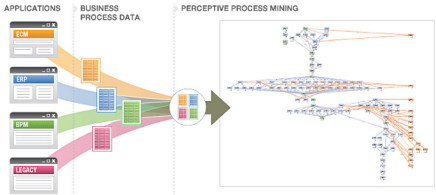
software AG

StereoLogic

ICARO
TECH

paf
now

FUJITSU



myinvenio

LANA Process Mining

Disco - New project

event_log_10000_cases +17

Map | Statistics | Cases

Academic
w.m.p.v.d.aalst@tue.nl

Disco

Statistics views

- Overview — Global statistics
- Activity — Activity classes
- product — Other attribute
- prod-price — Other attribute

Overview
Global statistics

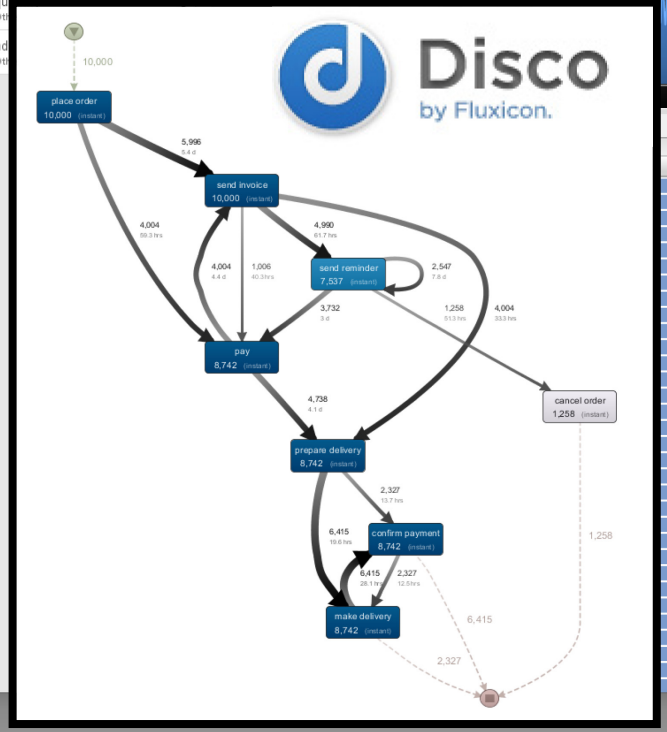
Events over time
Active cases over time
Case variants
Events per case
Case duration

Log timeline

Events	63,763
Cases	10,000
Activities	8
Median case duration	13.9 d
Mean case duration	14.9 d
Start	05.01.2015 09:00:07
End	31.12.2019 14:46:02

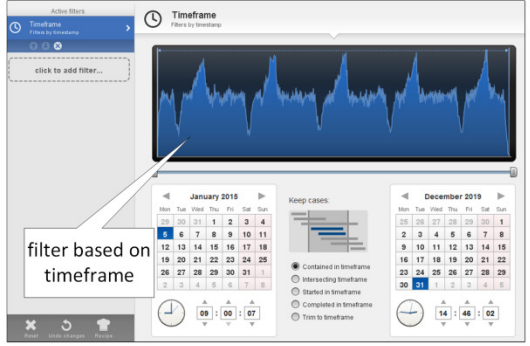
Cases (10000) | Variants (9)

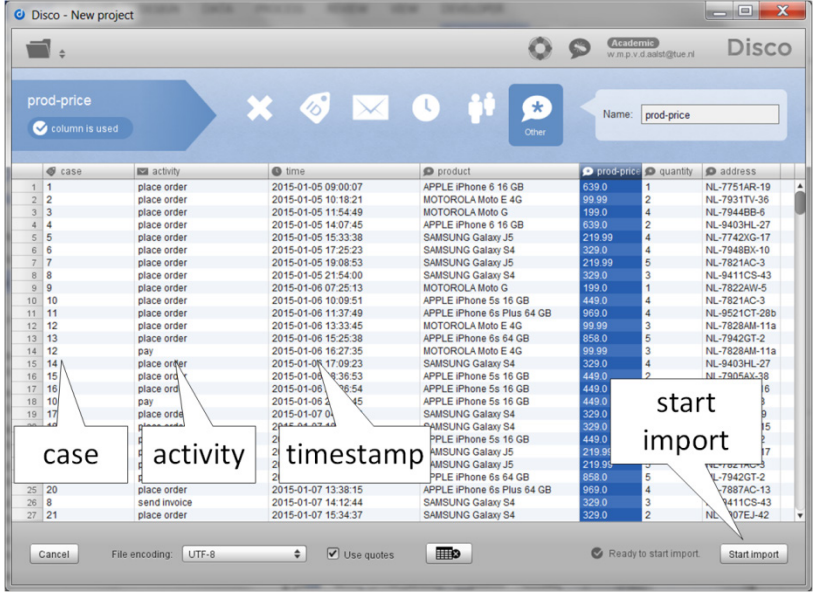
Variant	Started	Finished	Duration
7	05.01.2015 09:00:07	26.01.2015 16:42:28	21 days, 7 hours
1	05.01.2015 10:18:21	15.01.2015 15:52:30	10 days, 5 hours
4	05.01.2015 11:54:49	09.01.2015 18:38:58	4 days, 6 hours
3	05.01.2015 14:07:45	22.01.2015 13:18:30	16 days, 23 hours
1	05.01.2015 15:33:38	12.01.2015 17:27:36	7 days, 1 hour
5	05.01.2015 17:25:23	02.02.2015 12:31:09	27 days, 19 hours
4	05.01.2015 19:08:53	15.01.2015 14:56:54	9 days, 19 hours
9	05.01.2015 21:54:00	13.01.2015 15:49:53	7 days, 17 hours
4	06.01.2015 07:25:13	15.01.2015 11:27:50	9 days, 4 hours
1	06.01.2015 10:09:51	15.01.2015 19:15:18	9 days, 9 hours
1	06.01.2015 11:37:49	14.01.2015 09:14:28	7 days, 21 hours
4	06.01.2015 13:33:45	14.01.2015 11:30:05	7 days, 21 hours
4	06.01.2015 15:25:38	13.01.2015 12:25:34	6 days, 20 hours
2	06.01.2015 17:09:23	22.01.2015 18:59:10	16 days, 1 hour
3	06.01.2015 18:36:53	22.01.2015 14:39:39	15 days, 20 hours
8	06.01.2015 21:26:54	26.01.2015 17:16:02	19 days, 19 hours
1	07.01.2015 04:42:36	16.01.2015 10:17:14	9 days, 5 hours
3	07.01.2015 10:10:58	21.01.2015 17:31:29	14 days, 7 hours
8	07.01.2015 11:40:04	28.01.2015 10:27:12	20 days, 22 hours
9	07.01.2015 13:38:15	13.01.2015 13:22:15	5 days, 23 hours
1	07.01.2015 15:34:37	19.01.2015 09:11:23	11 days, 17 hours
1	07.01.2015 17:27:21	16.01.2015 09:09:25	8 days, 15 hours
5	07.01.2015 19:12:50	03.02.2015 14:34:33	26 days, 19 hours
6	07.01.2015 22:01:54	19.01.2015 13:15:02	11 days, 15 hours
8	08.01.2015 07:12:36	28.01.2015 10:41:14	20 days, 3 hours
3	08.01.2015 09:55:59	26.01.2015 15:52:42	18 days, 5 hours
6	08.01.2015 12:10:05	15.01.2015 13:54:59	7 days, 1 hour
1	08.01.2015 13:38:17	14.01.2015 12:30:26	5 days, 22 hours
5	08.01.2015 15:34:42	02.02.2015 14:10:36	24 days, 22 hours
2	08.01.2015 17:27:31	29.01.2015 11:26:06	20 days, 17 hours
3	08.01.2015 19:13:09	26.01.2015 14:16:02	17 days, 19 hours
4	08.01.2015 22:02:32	20.01.2015 10:35:40	11 days, 12 hours

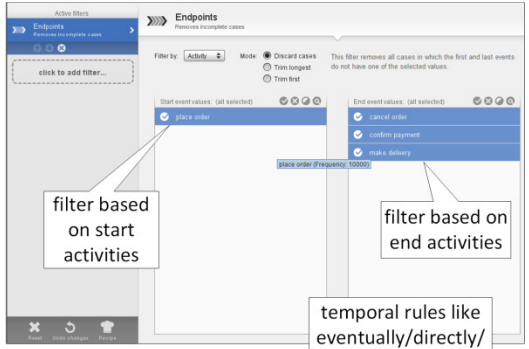


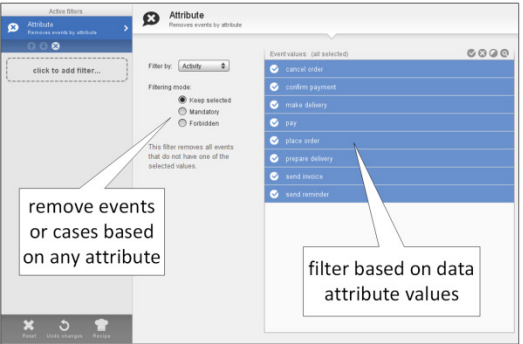
Filter | Copy | Delete | Export

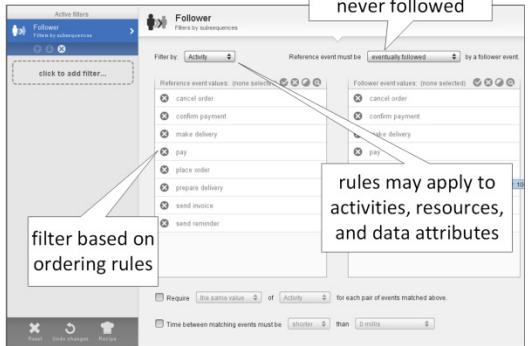
Version 1.9.1











Simple loading and filtering

Chair of Process and Data Science

Disco - New project

event_log_10000_cases +17

Map | Statistics | Cases

Academic
w.m.p.v.d.aalst@tue.nl

Disco

Zoom: 159%

search...

Detail

Activities 100%

Paths 100%

Frequency

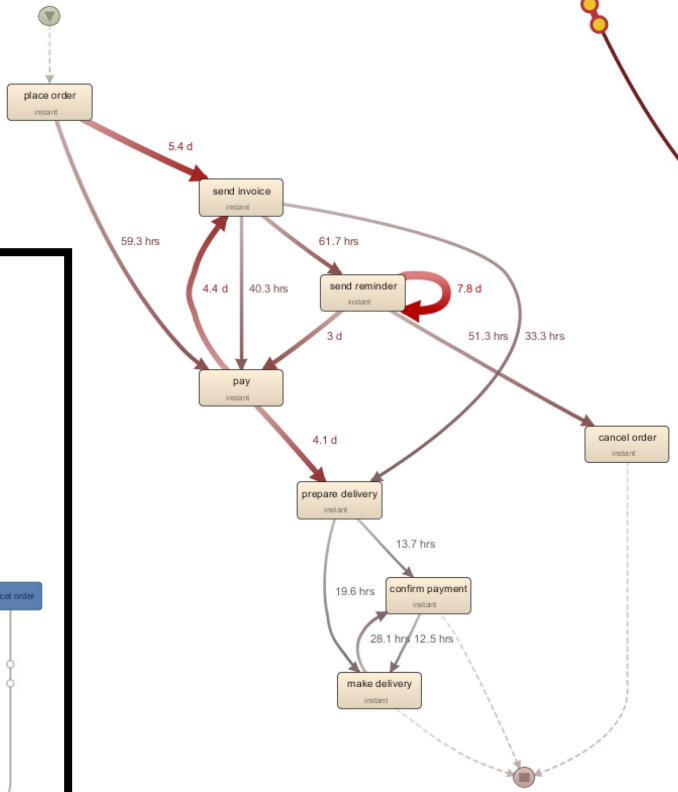
Performance

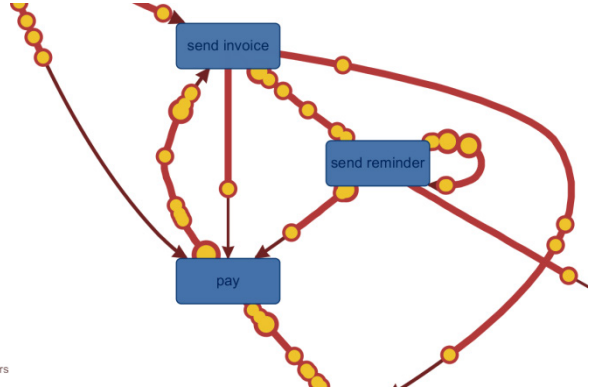
Show: Mean duration

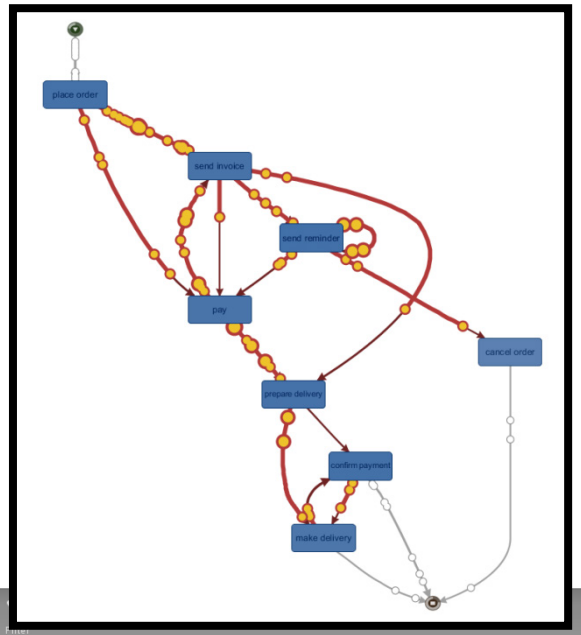
instant 6.3 d
instant 4.7 d
instant 3.1 d
instant 37.6 hrs

Add secondary metrics

Disco
by Fluxicon.







Animation

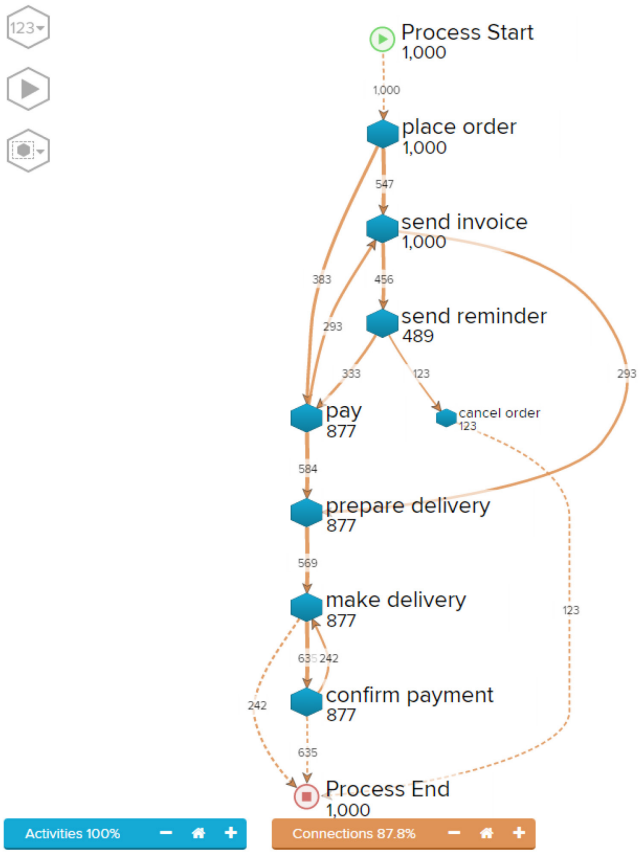
Copy Delete Export

Version 1.9.1

Celonis Process Mining

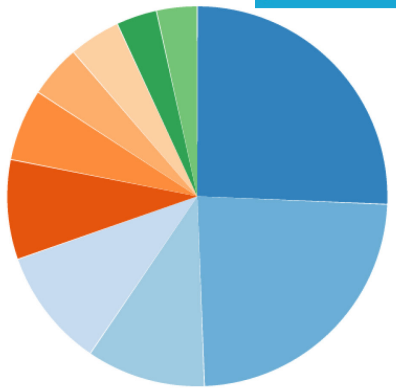
Artificial Logs Original / Spreadsheet-1000-cases
Last save on 2016-03-04 at 15:38

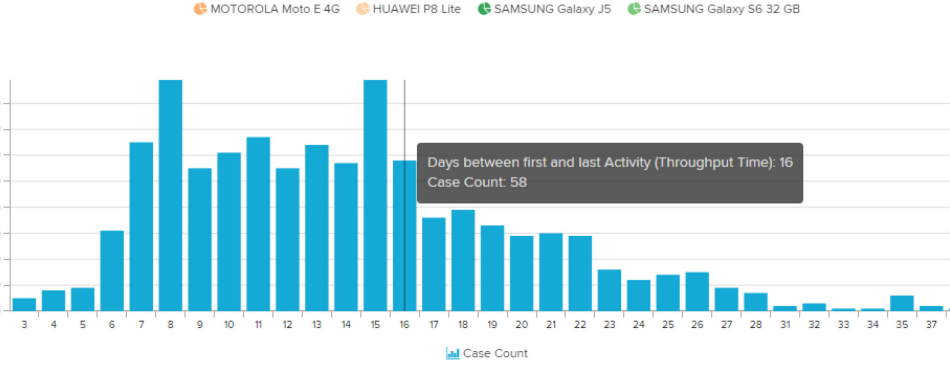
100% 1000 of 1000 cases selected



Activities 100%

Connections 87.8%



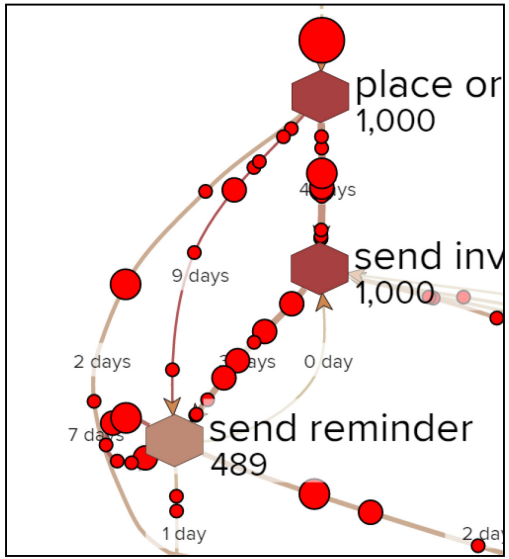


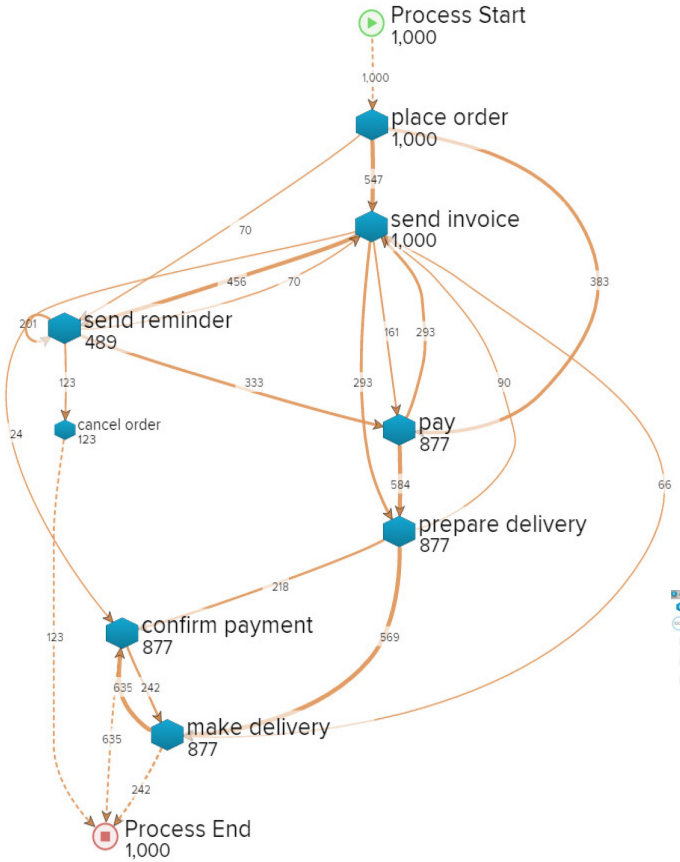
modelguidedabstraction.tex updated
"modelguidedabstraction.tex" was updated to the latest version (click to view).

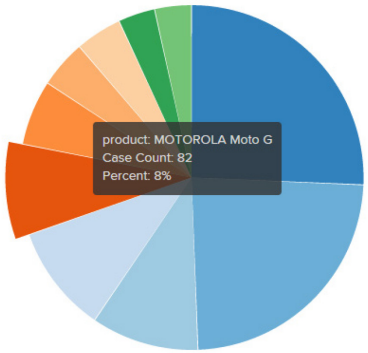
Process | Cases | Sheet 1

celonis
process mining

STORIES HELP Edit Refresh

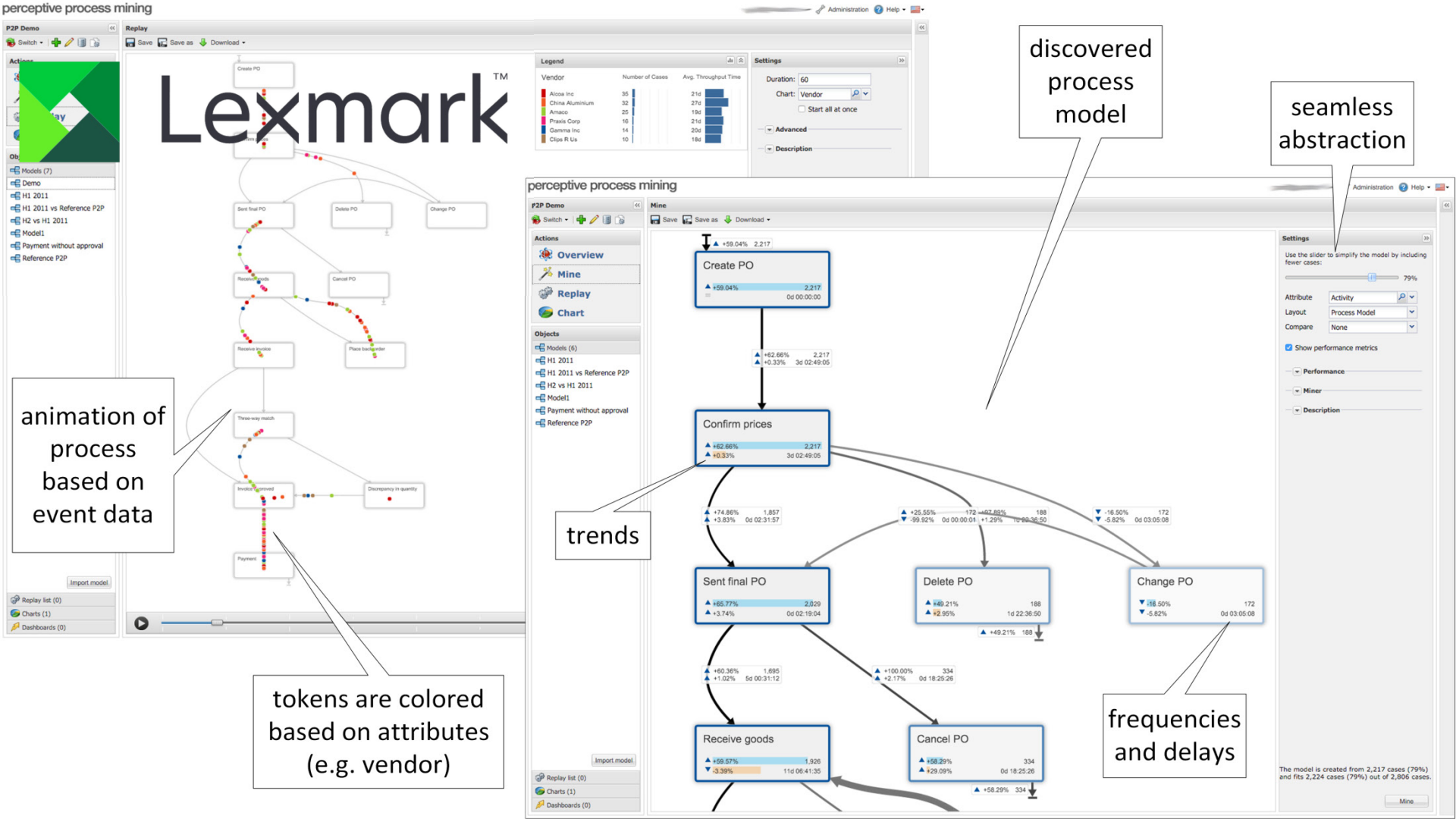


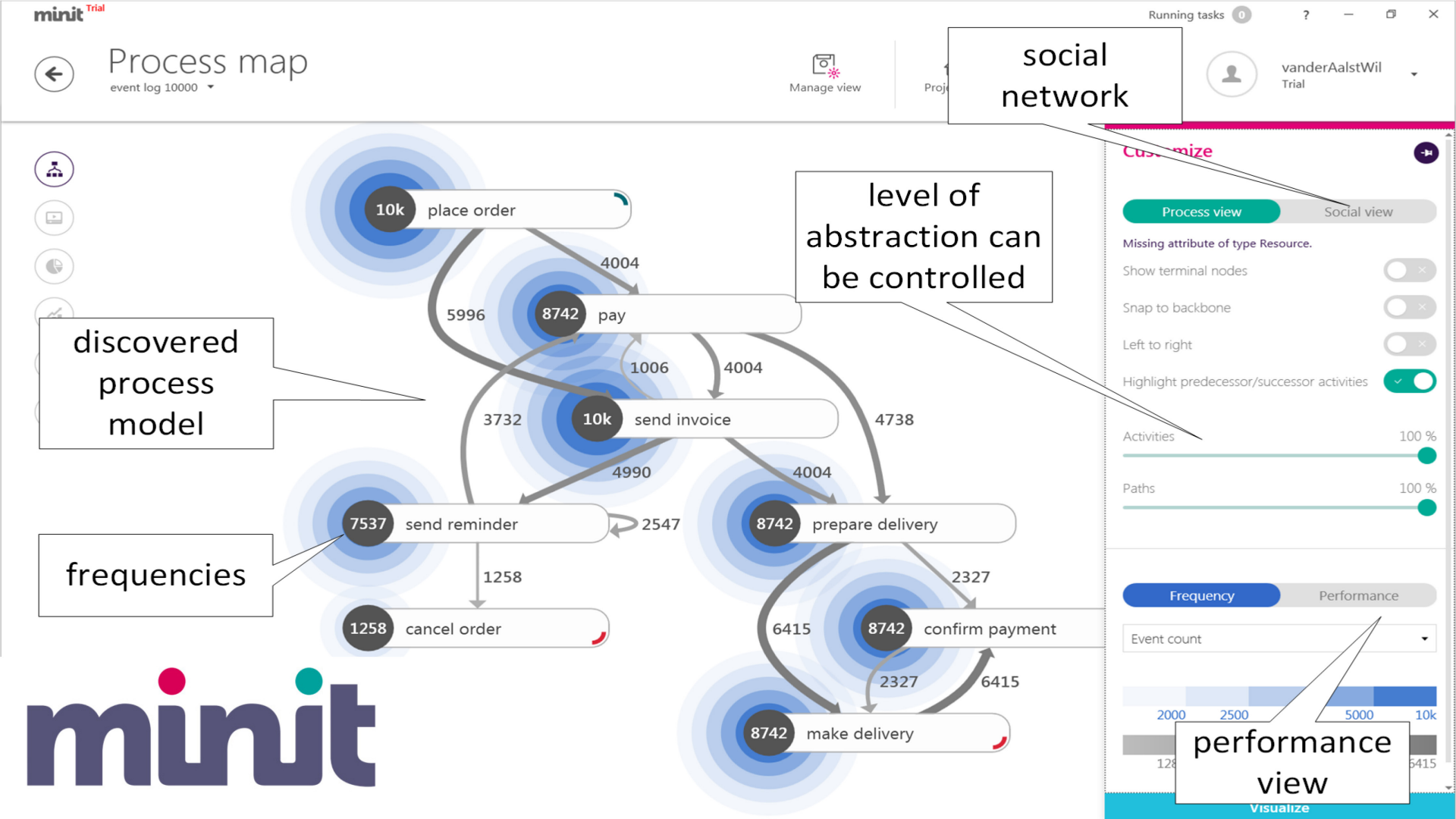


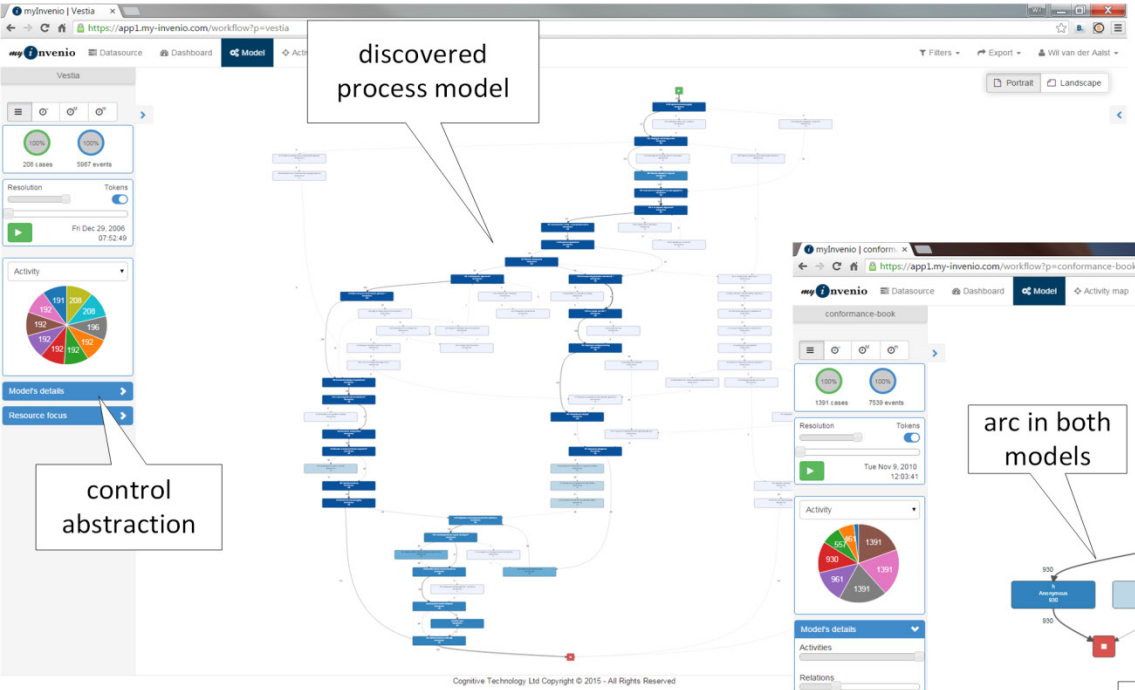


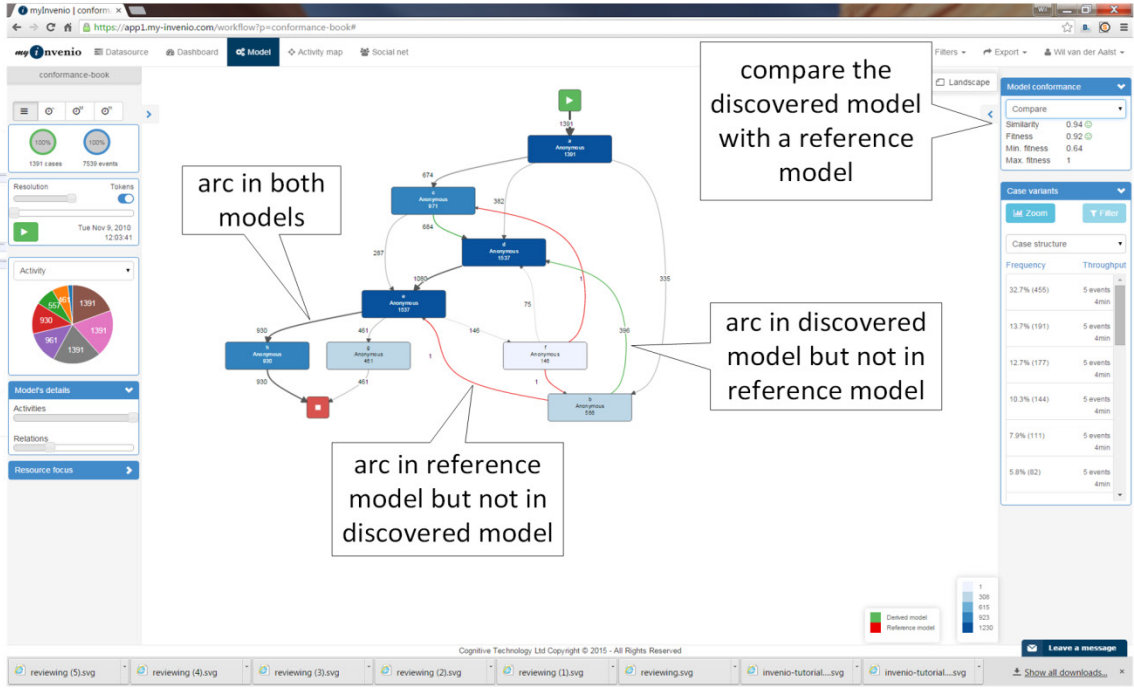
Activities 100% Connections 100%

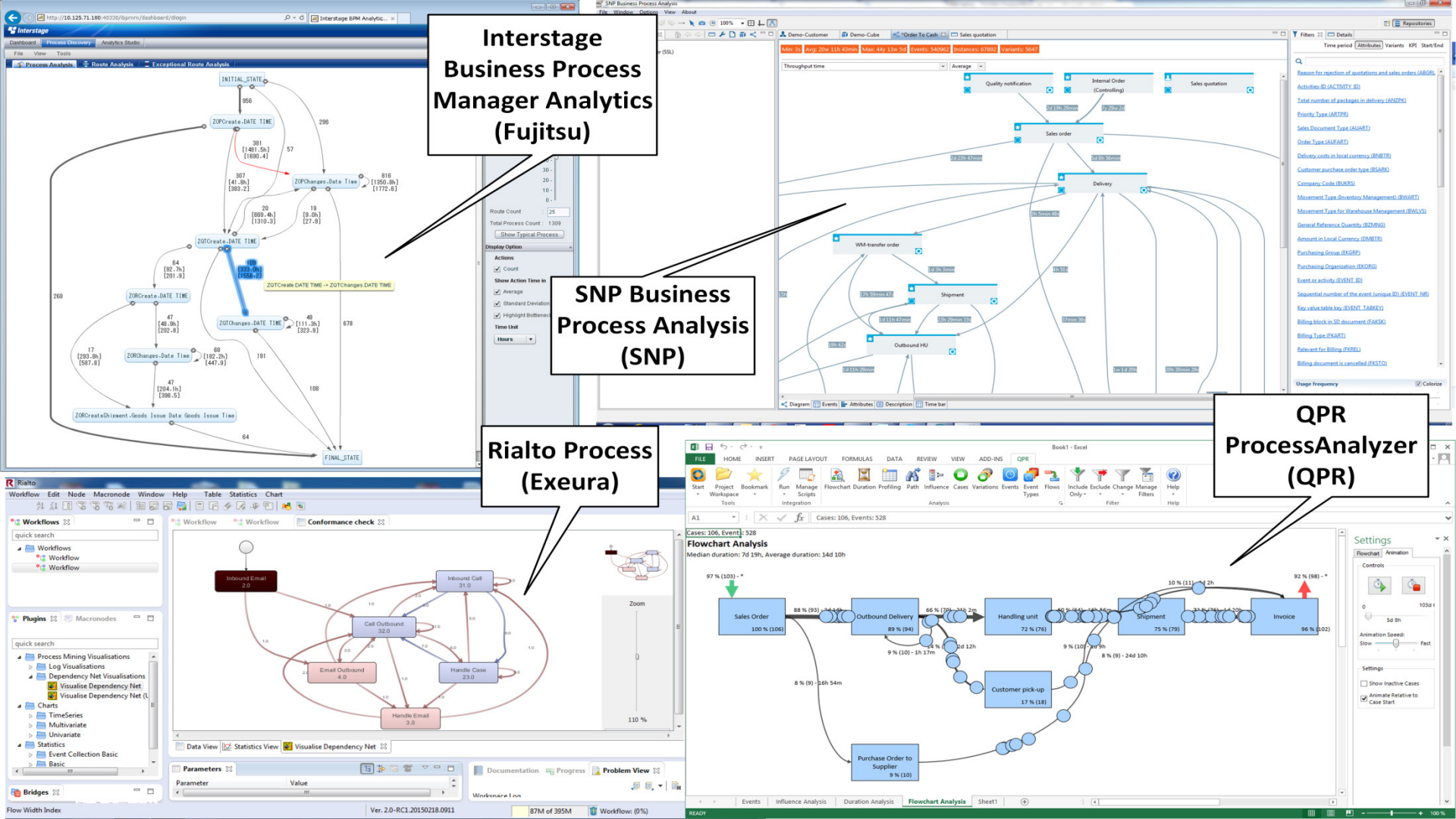
Process Cases Sheet 1


perceptive process mining
Lexmark™
discovered process model
seamless abstraction
animation of process based on event data
trends
tokens are colored based on attributes (e.g. vendor)
frequencies and delays
Create PO
Confirm prices
Sent final PO
Delete PO
Change PO
Receive goods
Cancel PO
Payment
Overview
Mine
Replay
Chart
Objects
Settings
Legend
Vendor
Number of Cases
Avg. Throughput Time
Duration: 60
Chart: Vendor
Start all at once
Advanced
Description
Models (7)
Demo
H1 2011
H1 2011 vs Reference P2P
H2 vs H1 2011
Model1
Payment without approval
Reference P2P
Import model
Replay list (0)
Charts (1)
Dashboards (0)
Use the slider to simplify the model by including fewer cases:
Attribute
Activity
Layout
Process Model
Compare
None
Show performance metrics
Performance
Miner
Description
The model is created from 2,217 cases (79%) and fits 2,224 cases (79%) out of 2,806 cases.

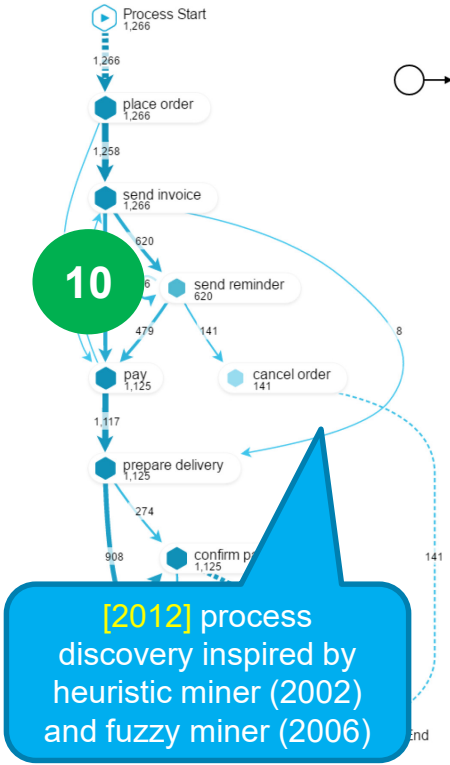

minit Trial
Running tasks 0
Process map
event log 10000
Manage view
vanderAalstWil
Trial
social network
level of abstraction can be controlled
discovered process model
frequencies
performance view
10k place order
4004
5996
8742 pay
1006
4004
3732
10k send invoice
4738
4990
4004
7537 send reminder
2547
8742 prepare delivery
1258
2327
1258 cancel order
6415
8742 confirm payment
2327
6415
8742 make delivery
Process view
Social view
Missing attribute of type Resource.
Show terminal nodes
Snap to backbone
Left to right
Highlight predecessor/successor activities
Activities 100 %
Paths 100 %
Frequency
Performance
Event count
2000
2500
5000
10k
Visualize
minit


discovered process model
control abstraction
Cognitive Technology Ltd Copyright © 2015 - All Rights Reserved


compare the discovered model with a reference model
arc in both models
arc in discovered model but not in reference model
arc in reference model but not in discovered model
Cognitive Technology Ltd Copyright © 2015 - All Rights Reserved

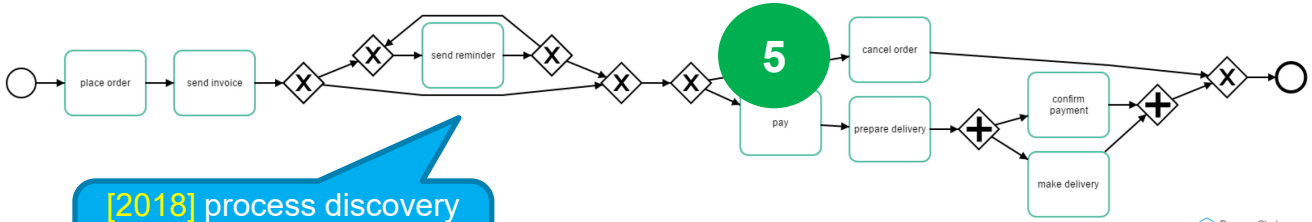

Interstage Business Process Manager Analytics (Fujitsu)
SNP Business Process Analysis (SNP)
Rialto Process (Exeura)
QPR ProcessAnalyzer (QPR)

Technology transfer (e.g., Celonis)



10

[2012] process discovery inspired by heuristic miner (2002) and fuzzy miner (2006)



5

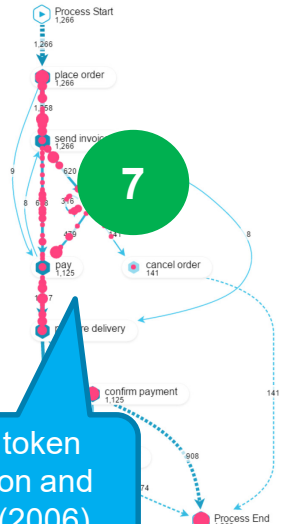
[2018] process discovery based on the inductive miner (2013)

11

[2017] process-based root cause analysis (2006)

12

[2017] token-based conformance checking (2005)



7

[2013] token animation and sliders (2006)

Chair of Process and Data Science

Adoption & Practical Relevance

Identified as a new market segment by Gartner. (April 2018)

"Celonis Process Mining operates on one of the largest SAP HANA In-Memory database installations in the world. More than 70 SAP ERP systems are simultaneously connected to SIEMENS' global HANA landscape and can be accessed on the fly."

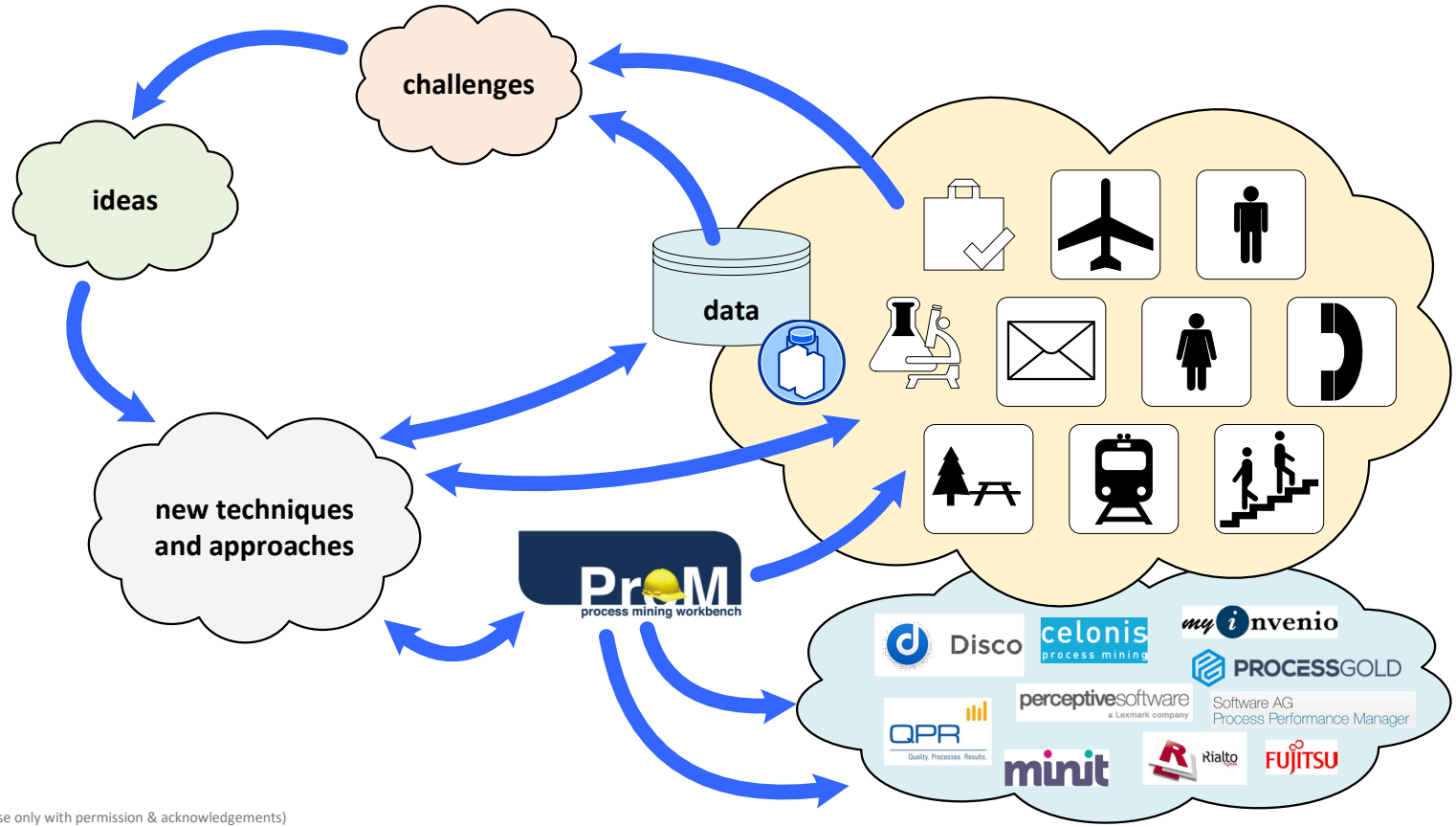
Celonis is one of the 5 German unicorns (valued over 1bln)

"Within Siemens AG there are now more than 2,500 active users of Process Dash worldwide."

"Siemens was able to achieve savings of double-digit millions as a result of the worldwide application of process mining."

*See process mining case studies at the IEEE Task Force on process Mining website

Interaction with industry



Chair of Process and Data Science


warning
advertisement ahead …

We welcome students to join the team!

- If you want to specialize in process mining (i.e., a combination of data science and process science) and you can provide "evidence" of this, then we are very interested in your application for:
  - Bachelor thesis projects
  - Master thesis projects
  - HiWi jobs (only if you can show experience, long-term commitment possible)
- Also check out our courses (BPI and APM courses in next semester), seminars, and practical assignments.

Chair of Process and Data Science

Praktikant / Werkstudent Data Science

(m/f/x) full-time, Munich

Are you ready for a new challenge?

You are looking for a position that is analytical but also involves customer contact? Working for a fast-growing company in a dynamic industry has always been your dream? As a part of our Data Science team, you help make digital business processes more transparent and harmonize our customers' gigantic data flows by actively using our Celonis Process Mining Technology and by applying the latest extraction and transformation methods. In doing so, you are involved in projects with process data of varying extent and complexity. After successful evaluation and preparation of the process data you connect the respective ERP-/Cloud systems with our software. Over the whole course of the implementation project, you proficiently handle our customers' individual needs and actively participate in customer workshops.

YOU...

... are a student of mathematics, business informatics or a comparable degree program with outstanding performance

... have already gained substantial know-how in Business Intelligence and data analytics

... enjoy working with customers and have awesome presentation and communication skills

... have ideally gained first experiences in programming, working with databases and ETL

... are proficient in SQL

... count demonstrating initiative, working independently, and thinking out of the box to the main characteristics you feature and seek in a job

... fancy discussing hot new topics and technology with colleagues over a beer and a game of table football

... have very good skills in verbal and written English, optimally also in German

If you are interested in a working student position, you ideally bring enough time to do an internship beforehand.

WE...

... are visionary and one of the fastest growing technology-unicorn in the world

... offer the world's most powerful tool for analyzing and optimizing IT-supported business processes and data volumes

... are pioneers and market leader in the area of Process Mining

... are distinguished by an unique combination of innovative start-up atmosphere combined with great professionalism and self-responsible work

Interested? Apply now!

Lisa Stibi | Junior Talent Acquisition Manager
Theresienstraße 6 | 80333 München
l.stibi@celonis.de
+49 162 7954708

www.celonis.com/careers/

Praktikant / Werkstudent Data Science & Product Management

(m/f/x) full-time, Munich

Are you ready for a new challenge?

You are looking for a job to work independently and put your ideas into action? Static structures and endless hierarchy levels bore you? Then Celonis is the place to be for you!
As a member of our content store team you work on scalable business solutions for Celonis Process Mining which we provide to our customers worldwide. Working at the intersection of data science and our internal development, you identify and implement new use cases and leverage existing solutions according to the demands of our customers or specific industries. In addition, you collect digital data from unknown IT Systems and bring transparency into underlying business processes. You implement smart algorithms and develop innovative applications that generate a value for our customers. For this, you combine your Data Science skills with your business know-how in order to take our process mining technology to a new level.

YOU...

... are a student of business informatics / computer engineering / business administration with focus on IT or a comparable degree program

... find it easy to understand business processes of any kind and feel comfortable depicting their structure and relevant KPIs

... can manage programming with SQL, even blindfolded

... are familiar with data structures such as those of SAP, Oracle or Salesforce

... are driven by topics such as Big Data, Data Science and Business Intelligence, and ideally, you have already gained first experiences working in these areas

... approach your work strategically and proactively

... convince with excellent verbal and written English skills

If you are interested in a working student position, you ideally bring enough time to do an internship beforehand.

WE...

...are visionary and one of the fastest growing Tech-Unicorns in the world

...offer our customers the Intelligent Business Cloud which is the world's most powerful tool for analyzing, optimizing, and transforming all IT-supported business processes

...are pioneers and market leader in the area of Process Mining

...distinguish ourselves through a unique combination of innovative start-up atmosphere paired with great professionalism and self-responsible work

Interested? Apply now!

Lisa Stibi | Junior Talent Acquisition Manager
Theresienstraße 6 | 80333 München
l.stibi@celonis-de
+49 162 7954708

www.celonis.com/careers/

Example: internships

Data Scientist

(m/f/x) full-time, Munich

Are you ready for a new challenge?

Actively drive Celonis' expansion and work in project teams to kick-start our customers' process mining journey. You make digital business processes transparent and harmonize our customers' gigantic data flows by using our Celonis Process Mining Technology and by applying the most up-to-date extraction and transformation methods. You are involved in implementation projects with process data of varying degrees and complexity and for customers across industries.
After the successful evaluation and preparation of the process data, you connect the respective on-premise/ Cloud systems with our software. You extract and transform customers' data and design process- and customer-specific analyses.
Over the course of our projects, you expertly handle our customers' individual needs and actively participate in customer workshops.

YOU...

...have successfully completed your studies in Business Informatics, Computer Science / Mathematics / Physics or a comparable degree program

...have gained prior knowledge in Business Intelligence and data analysis

...are already experienced in programming, ETL and working with databases

... have excellent analytical skills and are always well-organized and known for being a quick learner

...enjoy evaluating complex data and working with complicated processes

...are excited by Big Data, Data Mining and Process Mining seek continuous improvement of your know-how

...are very dedicated, visionary and looking for a product that you can develop with passion and determination

...have very good English and German skills, other languages are an advantage

WE...

...are visionary and one of the fastest growing Tech-Unicorns in the world

...offer our customers the Intelligent Business Cloud which is the world's most powerful tool for analyzing, optimizing, and transforming all IT-supported business processes

...are pioneers and market leader in the area of Process Mining

...distinguish ourselves through a unique combination of innovative start-up atmosphere paired with great professionalism and self-responsible work

Interested? Apply now!

Laura Nagl | Talent Acquisition Manager
Theresienstraße 6 | 80333 Munich
l.nagl@celonis.de
+49 152 0911 4918

www.celonis.com/careers/

Business Process Analyst

(m/f/x) full-time, Munich

Are you ready for a new challenge?

As a member of our content store team you work on scalable business solutions for Celonis Process Mining, which we provide to our customers worldwide. As a business analyst, you develop new use cases and leverage existing solutions according to the demands of our customers or specific industries. In addition, you collect digital data from unknown IT Systems and bring transparency into underlying business processes. You implement smart algorithms and develop innovative applications that generate a value for our customers. For this, you combine your Data Science Skills with your business know-how in order to take our process mining technology to a new level.

YOU...

... possess an above-average university degree in Economic Computer Science / Information-oriented Business Administration / Mathematics or equivalent

... have a sound understanding of all kinds of business processes and have many ideas how to improve them

... have a good knowledge of SQL and are familiar with the relational databases

... are enthused by Big Data, Data Science or Business Intelligence and have ideally already gained experience in this field

... have an analytical mind, work in a structured manner and are a quick learner

... want to have an impact and are looking for a product that you can develop with passion and drive

... have very good English skills, German is a plus

WE...

... are visionary and one of the fastest growing technology-unicorn in the world

... offer the world's most powerful tool for analyzing and optimizing IT-supported business processes and data volumes

... are pioneers and market leader in the area of Process Mining

... are distinguished by an unique combination of innovative start-up atmosphere combined with great professionalism and self-responsible work

Example: jobs

Interested? Apply now!

Alexandra Haberkern | Senior Talent Acquisition Manager
Theresienstraße 6 | 80333 Munich
a.haberkern@celonis.de
+49 89 4161596-742

www.celonis.com/careers/

Conclusion

Short summary of lecture

- Process Mining is great!
- Unsupervised: Process Discovery
- Supervised: Performance and conformance analysis
- Opportunities do dive deeper!

#	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*

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

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

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

10 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