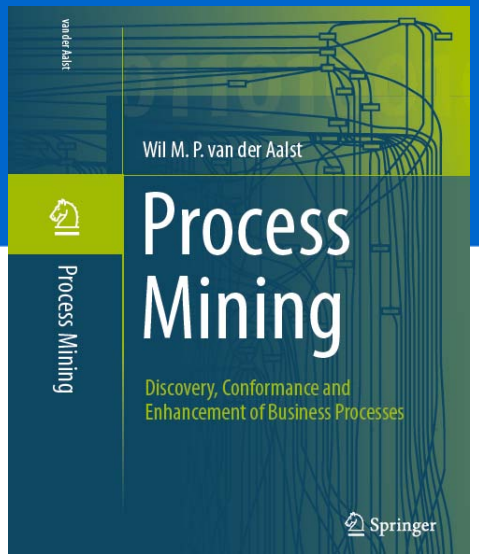
Process Mining

Enabling Data-Driven Process Discovery and Analysis Using ProM

Keynote SIMPDA 2011 – Campione d'Italia, June 29, 2011

prof.dr.ir. Wil van der Aalst

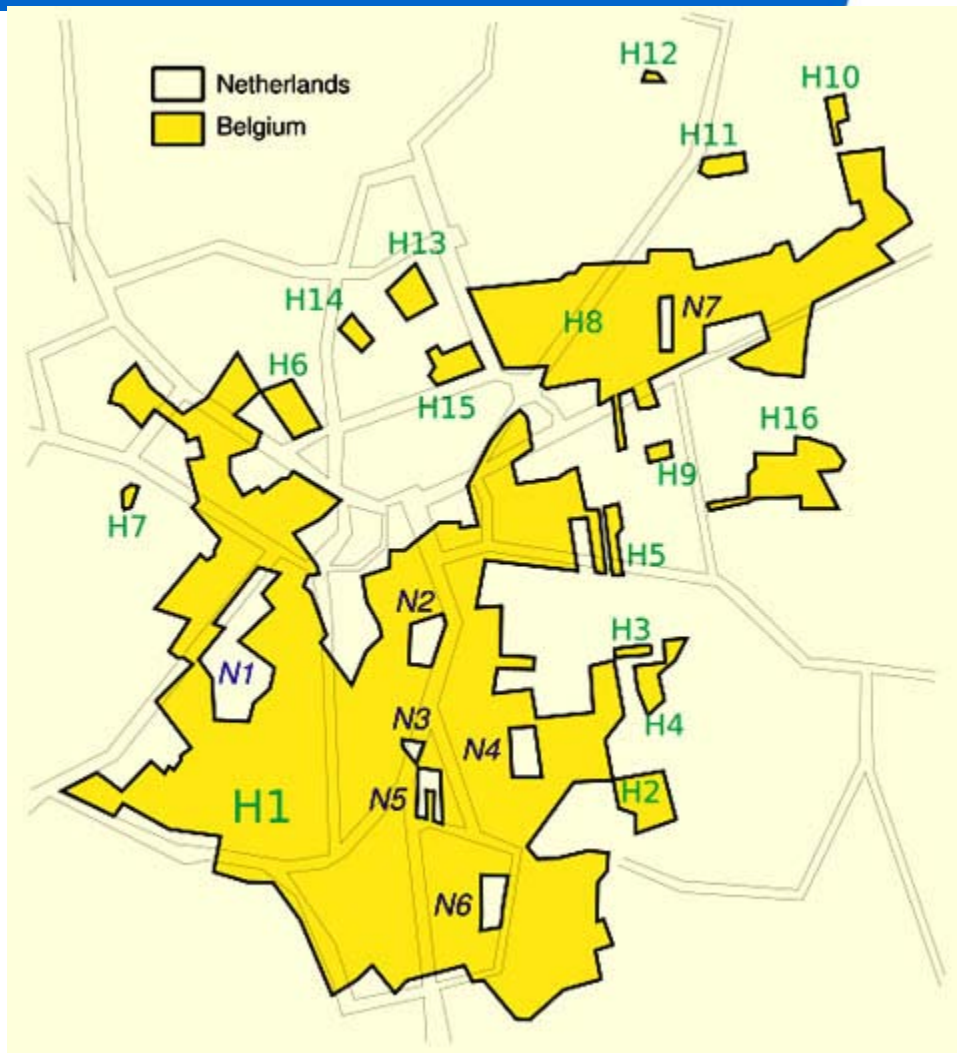
www.processmining.org



TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

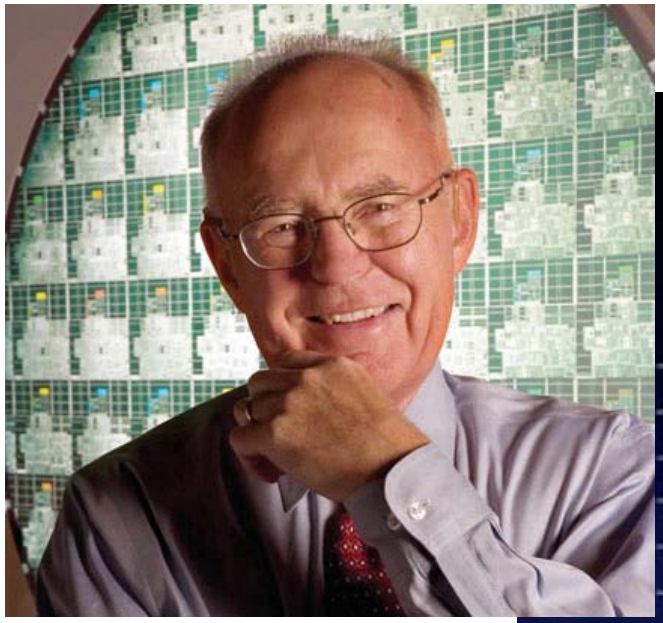
Baarle-Nassau (NL) /Baarle-Hertog (B)



22 enclaves in NL, 7 enclaves in B

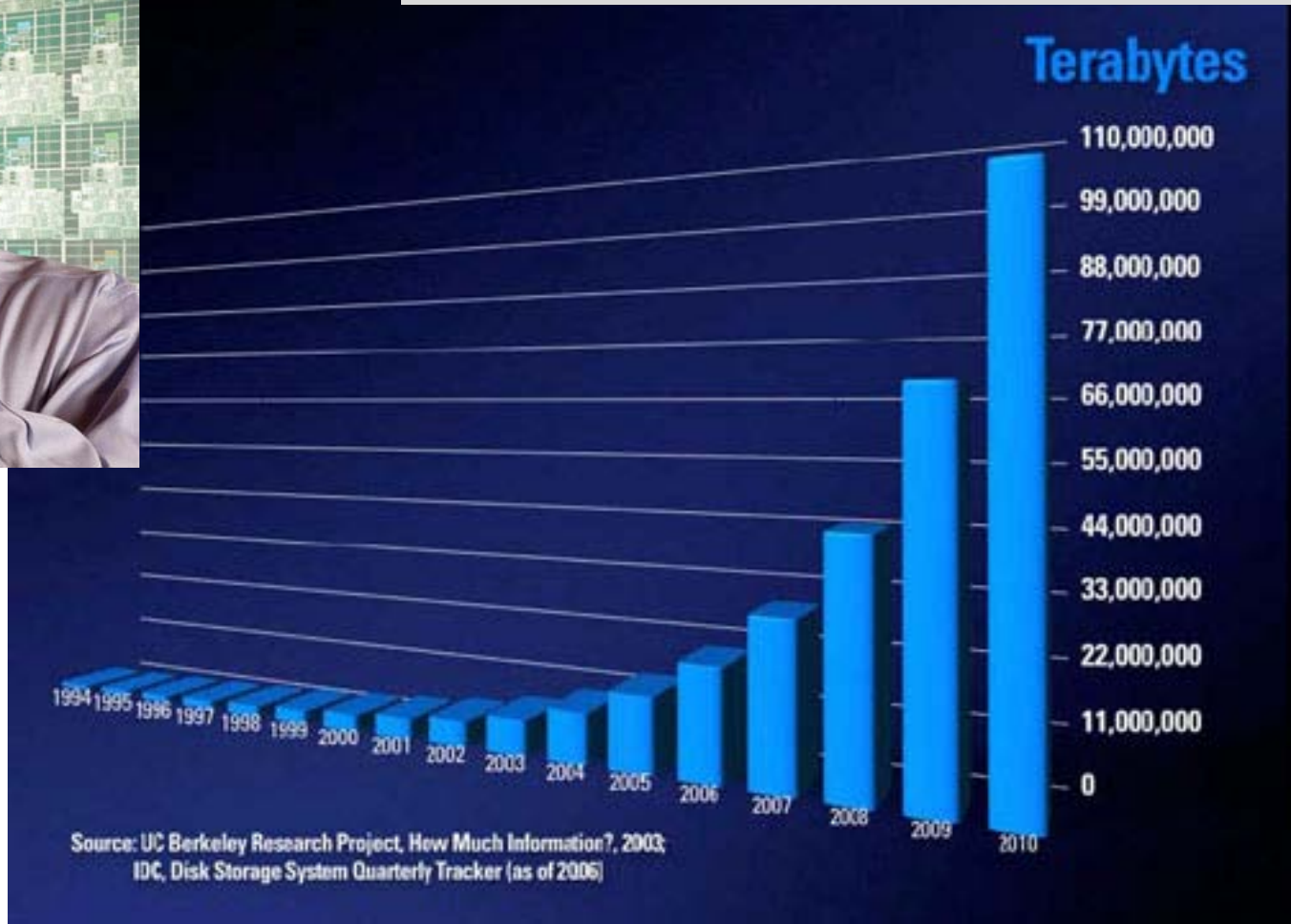
Due to medieval treaties (1198), agreements, land-swaps and sales between the Lords of Breda (NL) and the Dukes of Brabant (B)

Data explosion

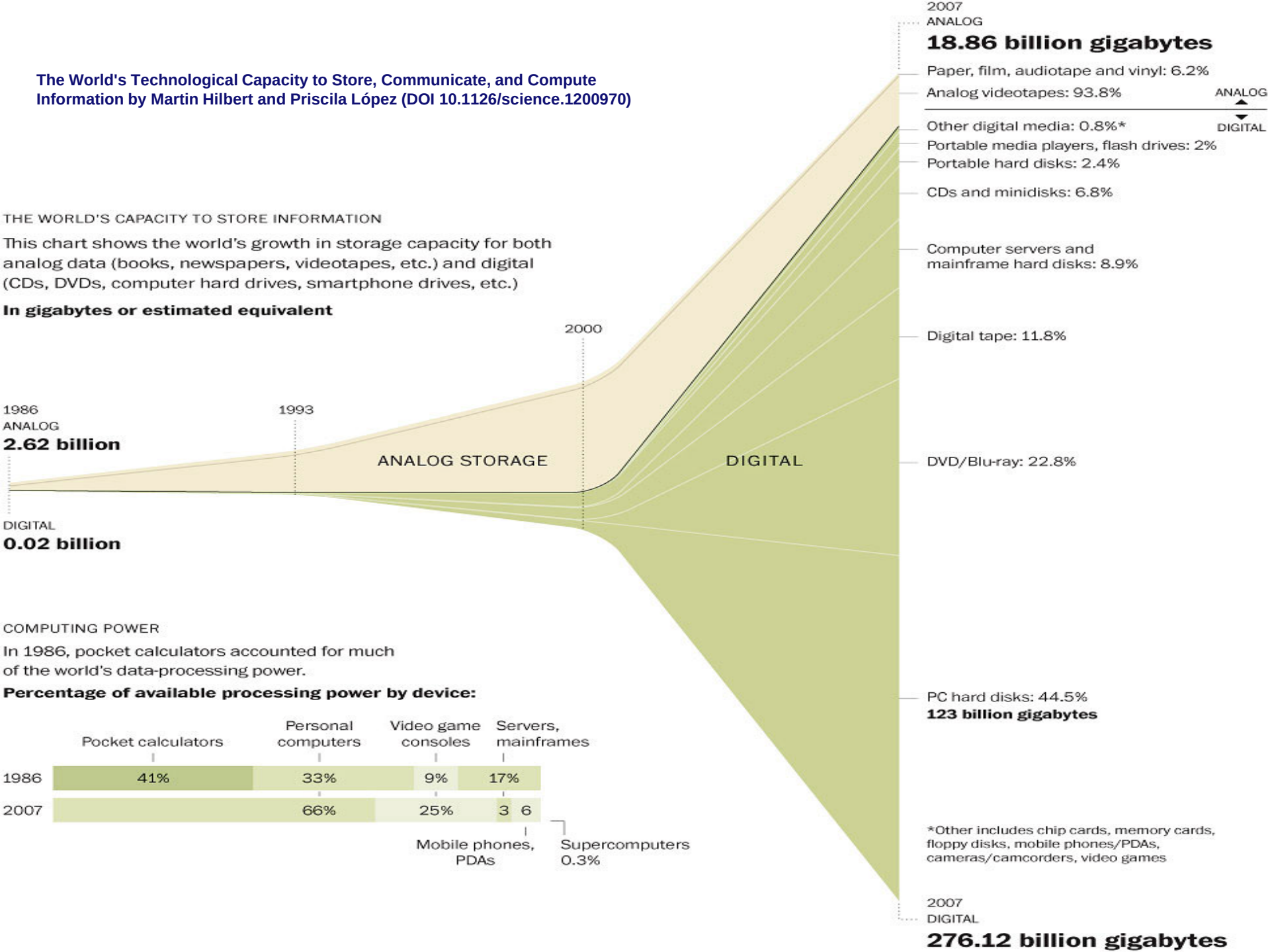


From Bits to Zettabytes

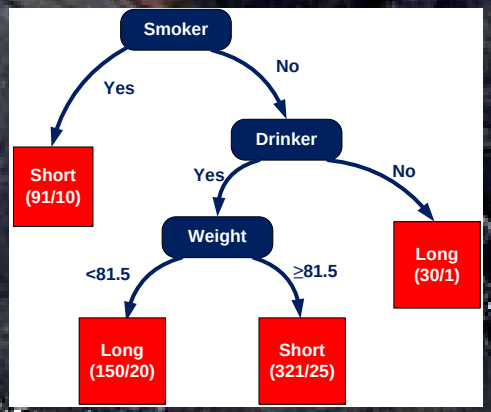
A "bit" is the smallest unit of information possible. One bit has two possible values: 1 (on) and 0 (off). A "byte" is composed of 8 bits and can represent $2^8 = 256$ values. To talk about larger amounts of data, multiples of 1000 are used: 1 Kilobyte (KB) equals 1000 bytes, 1 Megabyte (MB) equals 1000 KB, 1 Gigabyte (GB) equals 1000 MB, 1 Terabyte (TB) equals 1000 GB, 1 Petabyte (PB) equals 1000 TB, 1 Exabyte (EB) equals 1000 PB, and 1 Zettabyte (ZB) equals 1000 EB. Hence, 1 Zettabyte is 10^{21} = 1,000,000,000,000,000,000,000 bytes. Note that here we used the International System of Units (SI) set of unit prefixes, also known as SI prefixes, rather than binary prefixes. If we assume binary prefixes, then 1 Kilobyte is $2^{10} = 1024$ bytes, 1 Megabyte is $2^{20} = 1048576$ bytes, and 1 Zettabyte is $2^{70} \approx 1.18 \times 10^{21}$ bytes.



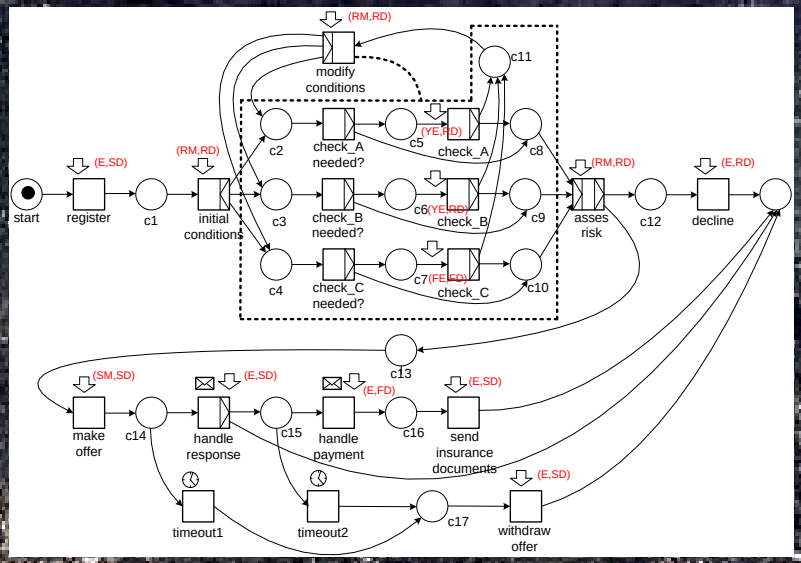
The World's Technological Capacity to Store, Communicate, and Compute Information by Martin Hilbert and Priscila López (DOI 10.1126/science.1200970)



Process Mining =

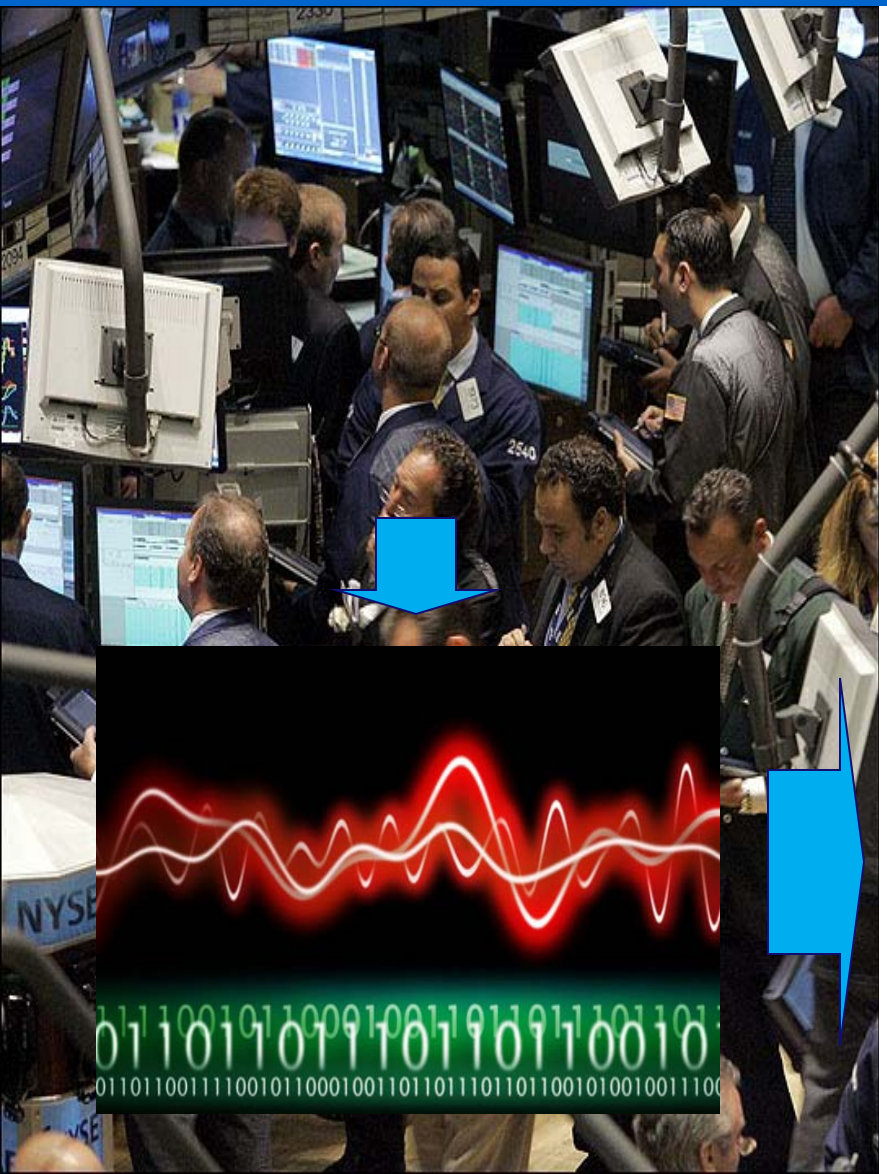


+



Data Mining Process Analysis

Process Mining

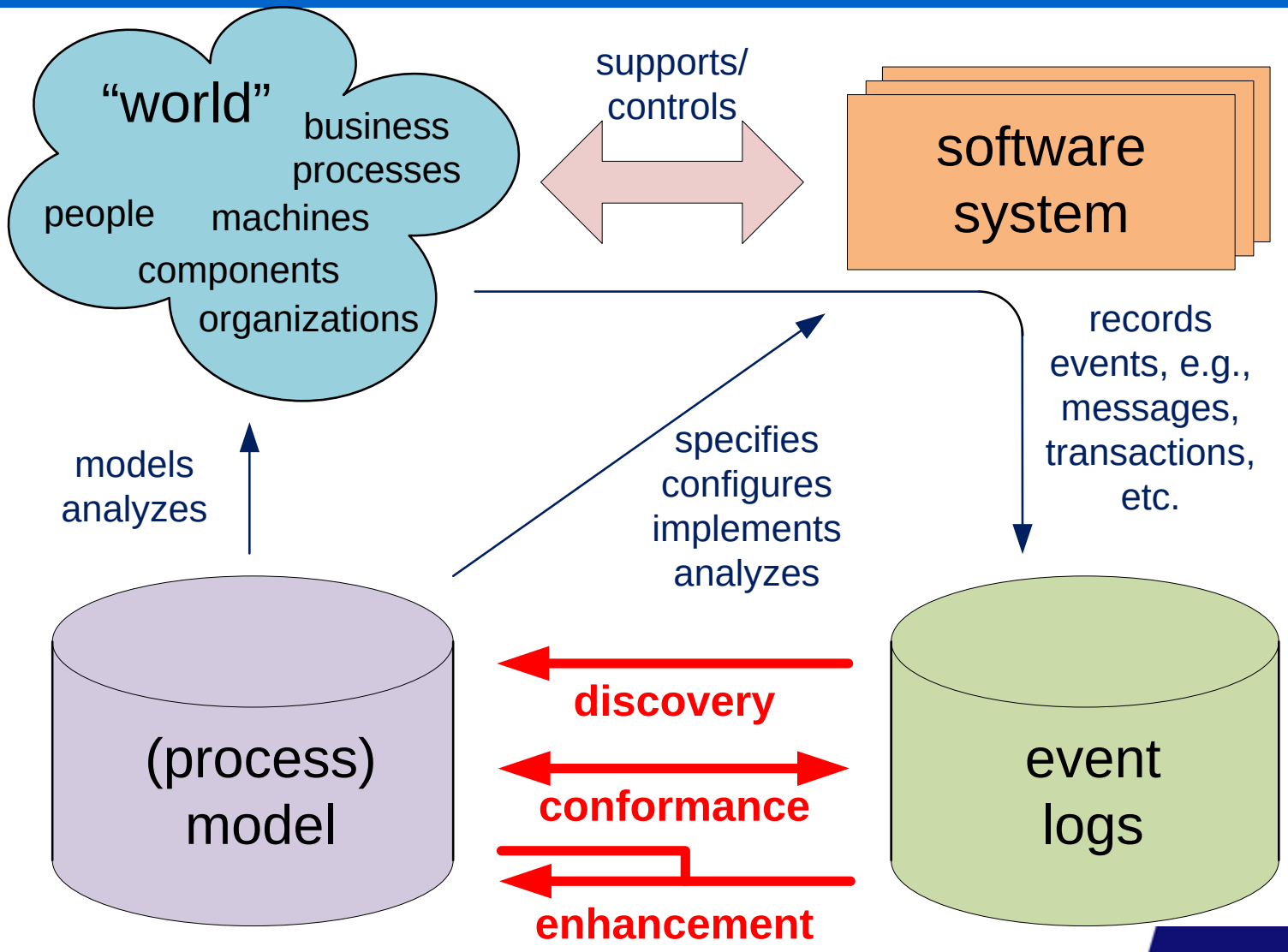


- Process discovery: "What is really happening?"
- Conformance checking: "Do we do what was agreed upon?"
- Performance analysis: "Where are the bottlenecks?"
- Process prediction: "Will this case be late?"
- Process improvement: "How to redesign this process?"
- Etc.

We applied ProM in >100 organizations

- **Municipalities** (e.g., Alkmaar, Heusden, Harderwijk, etc.)
- **Government agencies** (e.g., Rijkswaterstaat, Centraal Justitieel Incasso Bureau, Justice department)
- **Insurance related agencies** (e.g., UWV)
- **Banks** (e.g., ING Bank)
- **Hospitals** (e.g., AMC hospital, Catharina hospital)
- **Multinationals** (e.g., DSM, Deloitte)
- **High-tech system manufacturers and their customers** (e.g., Philips Healthcare, ASML, Ricoh, Thales)
- **Media companies** (e.g. Winkwaves)
- **...**

Process Mining



Starting point: event log

case id	event id	properties				
		timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	...
3	35654521	30-12-2010:14.32	register request			
	35654522	30-12-2010:15.06	examine casually			
	35654524	30-12-2010:16.34	check ticket			
	35654525	06-01-2011:09.18	decide			
	35654526	06-01-2011:12.18	reinitiate request			
	35654527	06-01-2011:13.06	examine thoroughly			
	35654530	08-01-2011:11.43	check ticket			
	35654531	09-01-2011:09.55	decide			
	35654533	15-01-2011:10.45	pay compensation			
4	35654641	06-01-2011:15.02	register request			
	35654643	07-01-2011:12.06	check ticket			
	35654644	08-01-2011:14.43	examine thoroughly			
	35654645	09-01-2011:12.02	decide			
	35654647	12-01-2011:15.44	reject request			
5	35654711	06-01-2011:09.02	register request			
	35654712	07-01-2011:10.16	examine casually			
	35654714	08-01-2011:11.22	check ticket			
	35654715	10-01-2011:13.28	decide			
	35654716	11-01-2011:16.18	reinitiate request			
	35654718	14-01-2011:14.33	check ticket			
	35654719	16-01-2011:15.50	examine casually			
	35654720	19-01-2011:11.18	decide	Sara	200	...
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	...
	35654722	21-01-2011:09.06	examine casually	Sue	400	...
	35654724	21-01-2011:11.34	check ticket	Pete	100	...
	35654725	23-01-2011:13.12	decide	Sara	200	...
	35654726	24-01-2011:14.56	reject request	Mike	200	...
6	35654871	06-01-2011:15.02	register request	Mike	50	...
	35654873	06-01-2011:16.06	examine casually	Ellen	400	...
	35654874	07-01-2011:16.22	check ticket	Mike	100	...
	35654875	07-01-2011:16.52	decide	Sara	200	...
	35654877	16-01-2011:11.47	pay compensation	Mike	200	...
...	...	...	...	...	...	...

case id	event id	properties				
		timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	...

XES, MXML, SA-MXML, CSV, etc.

Simplified event log

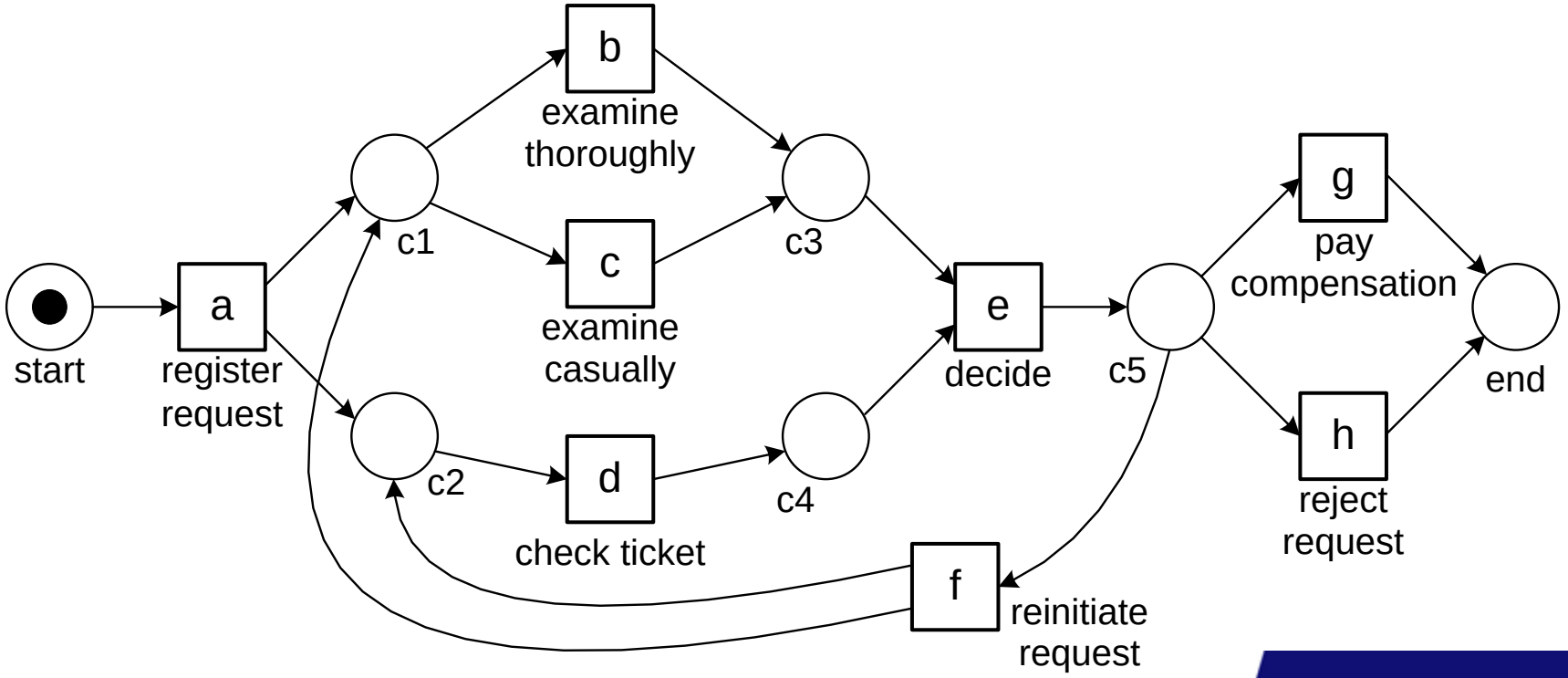
case id	event id	properties				
		timestamp	activity	resource		
1	35654423	30-12-2010:11.02	register request	Pete		
	35654424	31-12-2010:10.06	examine thoroughly	Sue		
	35654425	05-01-2011:15.12	check ticket	Mike		
	35654426	06-01-2011:11.18	decide	Sara		
	35654427	07-01-2011:14.24	reject request	Pete		
2	35654483	30-12-2010:11.32	register request	Mike		
	35654485	30-12-2010:12.12	check ticket	Mike		
	35654487	30-12-2010:14.16	examine casually	Pete		
	35654488	05-01-2011:11.22	decide	Sara		
	35654489	08-01-2011:12.05	pay compensation	Ellen		
3	35654521	30-12-2010:14.32	register request	Pete		
	35654522	30-12-2010:15.06	examine casually	Mike		
	35654524	30-12-2010:16.34	check ticket	Ellen		
	35654525	06-01-2011:09.18	decide	Sara		
	35654526	06-01-2011:12.18	reinitiate request	Sara		
	35654527	06-01-2011:13.06	examine thoroughly	Sean		
	35654530	08-01-2011:11.43	check ticket	Pete		
	35654531	09-01-2011:09.55	decide	Sara		
	35654533	15-01-2011:10.45	pay compensation	Ellen		
4	35654641	06-01-2011:15.02	register request	Pete	50	…
	35654643	07-01-2011:12.06	check ticket	Mike	100	…
	35654644	08-01-2011:14.43	examine thoroughly	Sean	400	…
	35654645	09-01-2011:12.02	decide	Sara	200	…
	35654647	12-01-2011:15.44	reject request	Ellen	200	…
5	35654711	06-01-2011:09.02	register request	Ellen	50	…
	35654712	07-01-2011:10.16	examine casually	Mike	400	…
	35654714	08-01-2011:11.22	check ticket	Pete	100	…
	35654715	10-01-2011:13.28	decide	Sara	200	…
	35654716	11-01-2011:16.18	reinitiate request	Sara	200	…
	35654718	14-01-2011:14.33	check ticket	Ellen	100	…
	35654719	16-01-2011:15.50	examine casually	Mike	400	…
	35654720	19-01-2011:11.18	decide	Sara	200	…
	35654721	20-01-2011:12.48	reinitiate request	Sara	200	…
	35654722	21-01-2011:09.06	examine casually	Sue	400	…
	35654724	21-01-2011:11.34	check ticket	Pete	100	…
	35654725	23-01-2011:13.12	decide	Sara	200	…
	35654726	24-01-2011:14.56	reject request	Mike	200	…
6	35654871	06-01-2011:15.02	register request	Mike	50	…
	35654873	06-01-2011:16.06	examine casually	Ellen	400	…
	35654874	07-01-2011:16.22	check ticket	Mike	100	…
	35654875	07-01-2011:16.52	decide	Sara	200	…
	35654877	16-01-2011:11.47	pay compensation	Mike	200	…
…	…	…	…	…	…	…

case id	trace
1	$\langle a, b, d, e, h \rangle$
2	$\langle a, d, c, e, g \rangle$
3	$\langle a, c, d, e, f, b, d, e, g \rangle$
4	$\langle a, d, b, e, h \rangle$
5	$\langle a, c, d, e, f, d, c, e, f, c, d, e, h \rangle$
6	$\langle a, c, d, e, g \rangle$
…	…

a = register request,
b = examine thoroughly,
c = examine casually,
d = check ticket,
e = decide,
f = reinitiate request,
g = pay compensation,
and h = reject request

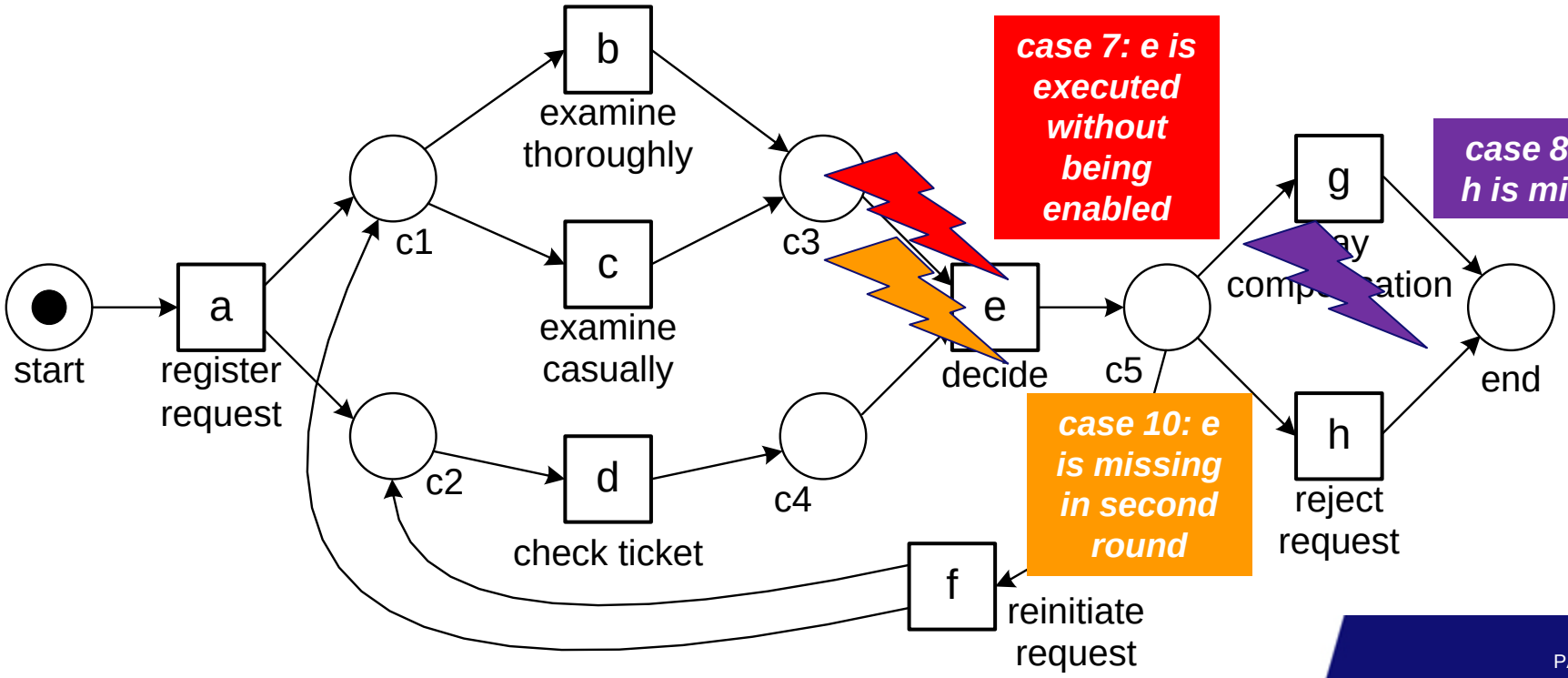
Process discovery

case id	trace
1	$\langle a,b,d,e,h \rangle$
2	$\langle a,d,c,e,g \rangle$
3	$\langle a,c,d,e,f,b,d,e,g \rangle$
4	$\langle a,d,b,e,h \rangle$
5	$\langle a,c,d,e,f,d,c,e,f,c,d,e,h \rangle$
6	$\langle a,c,d,e,g \rangle$
...	...

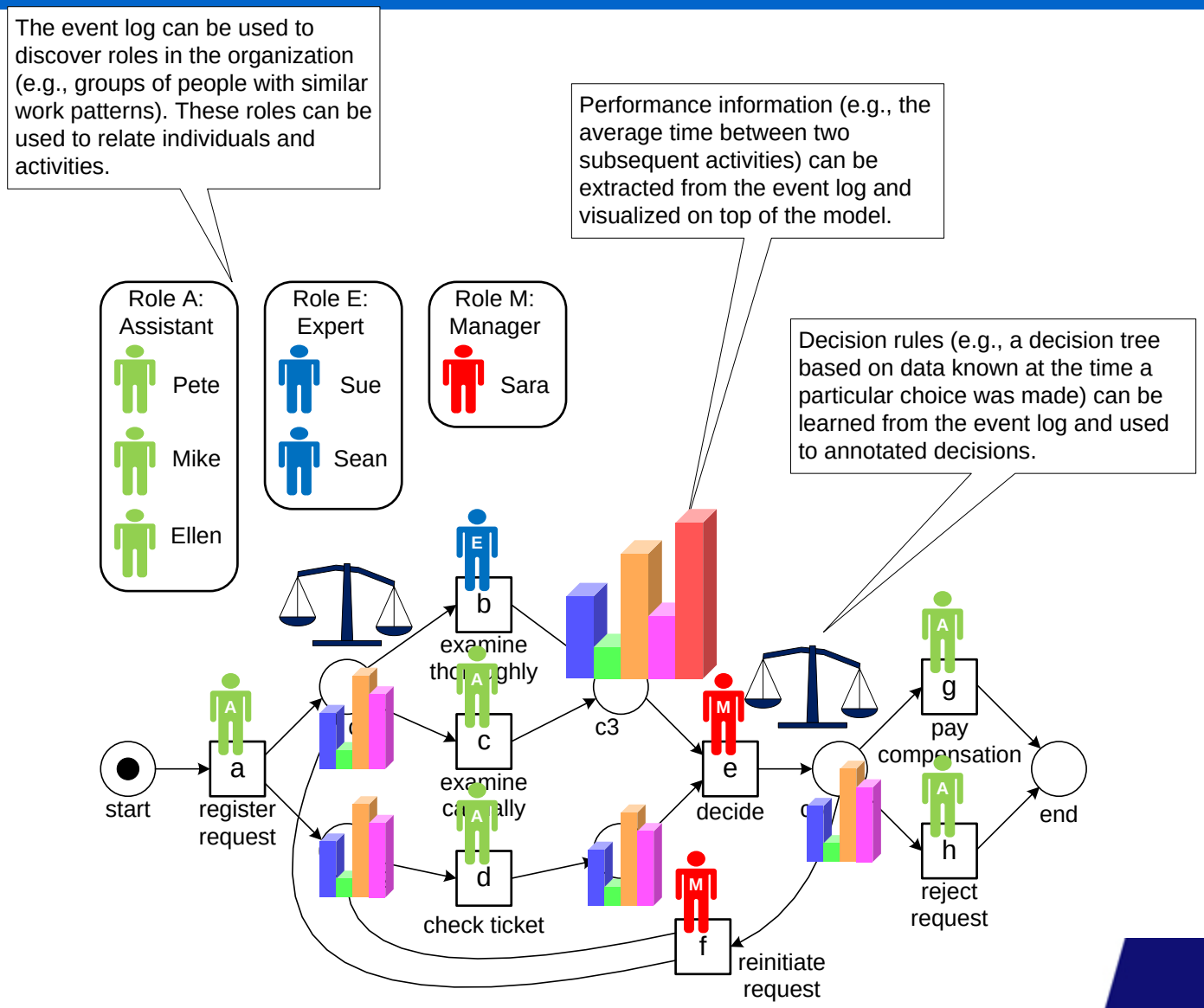


Conformance checking

case id	trace
1	$\langle a,b,d,e,h\rangle$
2	$\langle a,d,c,e,g\rangle$
3	$\langle a,c,d,e,f,b,d,e,g\rangle$
4	$\langle a,d,b,e,h\rangle$
5	$\langle a,c,d,e,f,d,c,e,f,c,d,e,h\rangle$
6	$\langle a,c,d,e,g\rangle$
7	$\langle \mathbf{a},\mathbf{b},\mathbf{e},\mathbf{g}\rangle$
8	$\langle \mathbf{a},\mathbf{b},\mathbf{d},\mathbf{e}\rangle$
9	$\langle a,d,c,e,f,d,c,e,f,b,d,e,h\rangle$
10	$\langle \mathbf{a},\mathbf{c},\mathbf{d},\mathbf{e},\mathbf{f},\mathbf{b},\mathbf{d},\mathbf{g}\rangle$



Extension: Adding perspectives to model based on event log



All supported by …

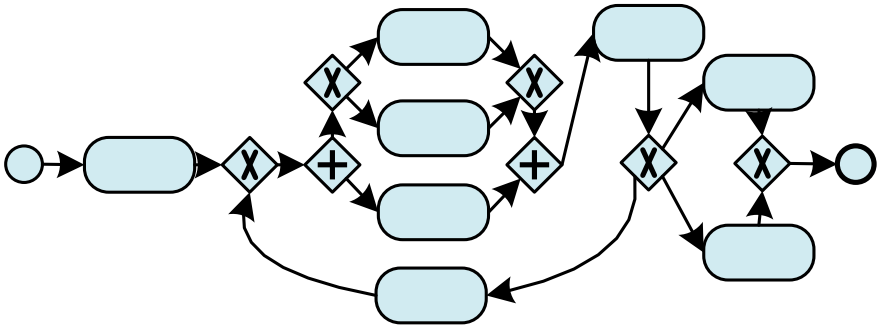
Let us play …

Play-Out

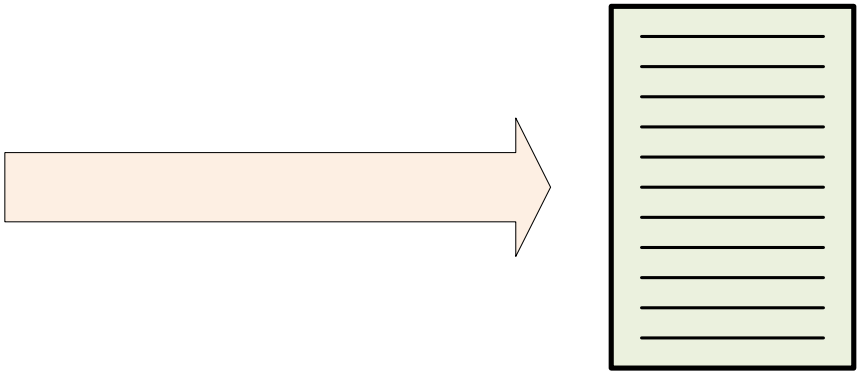
Play-In

Replay

Play-Out

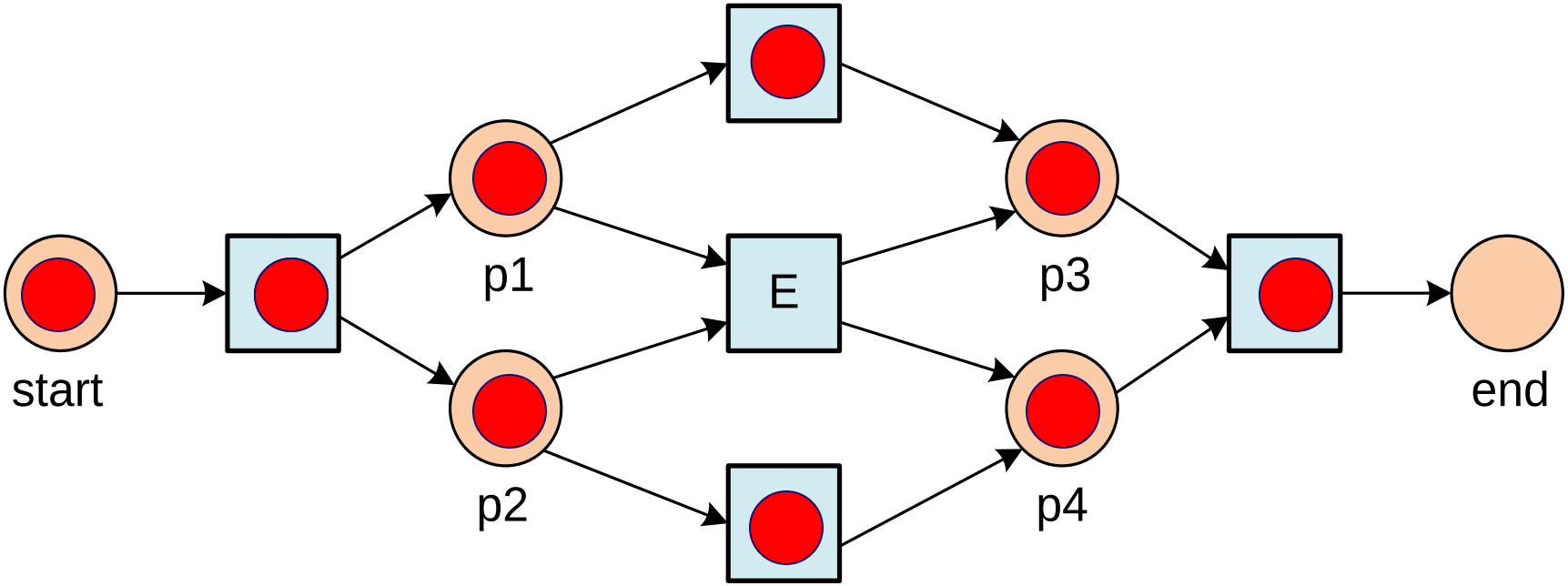


process model



event log

Play-Out (Classical use of models)

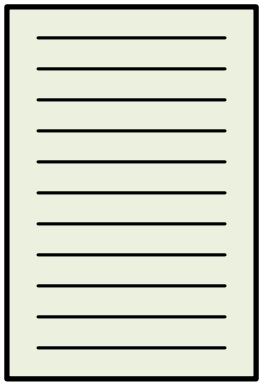


A B C D A E D A E D

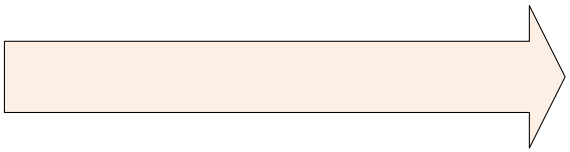
A B C D A C B D

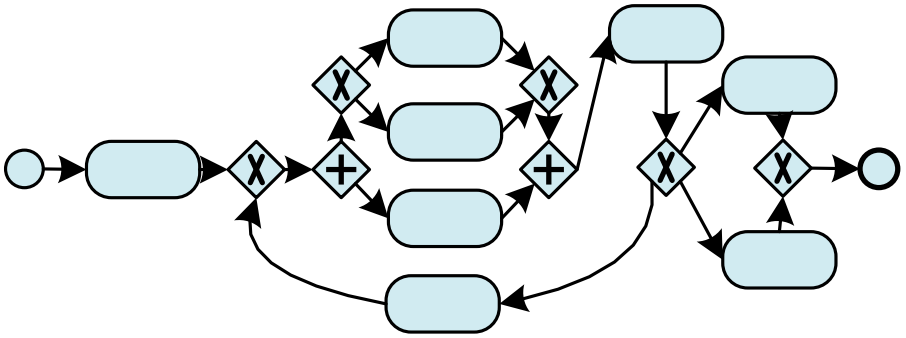
A C B D A E D A C B D

Play-In



event log





process model

Play-In

A B C D

A E D

A E D

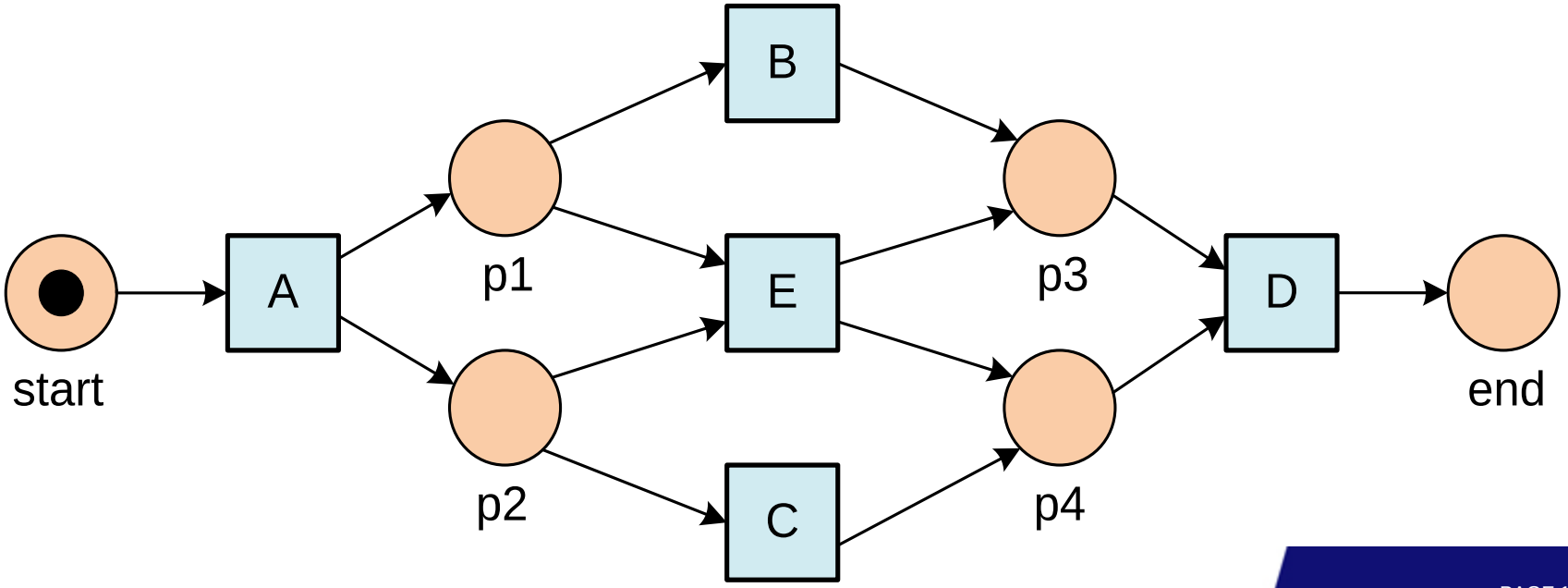
A B C D

A C B D

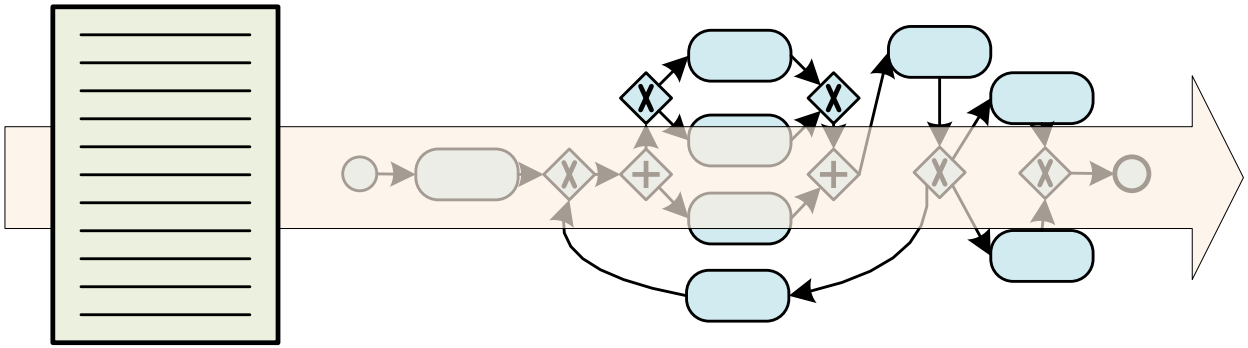
A C B D

A E D

A C B D



Replay



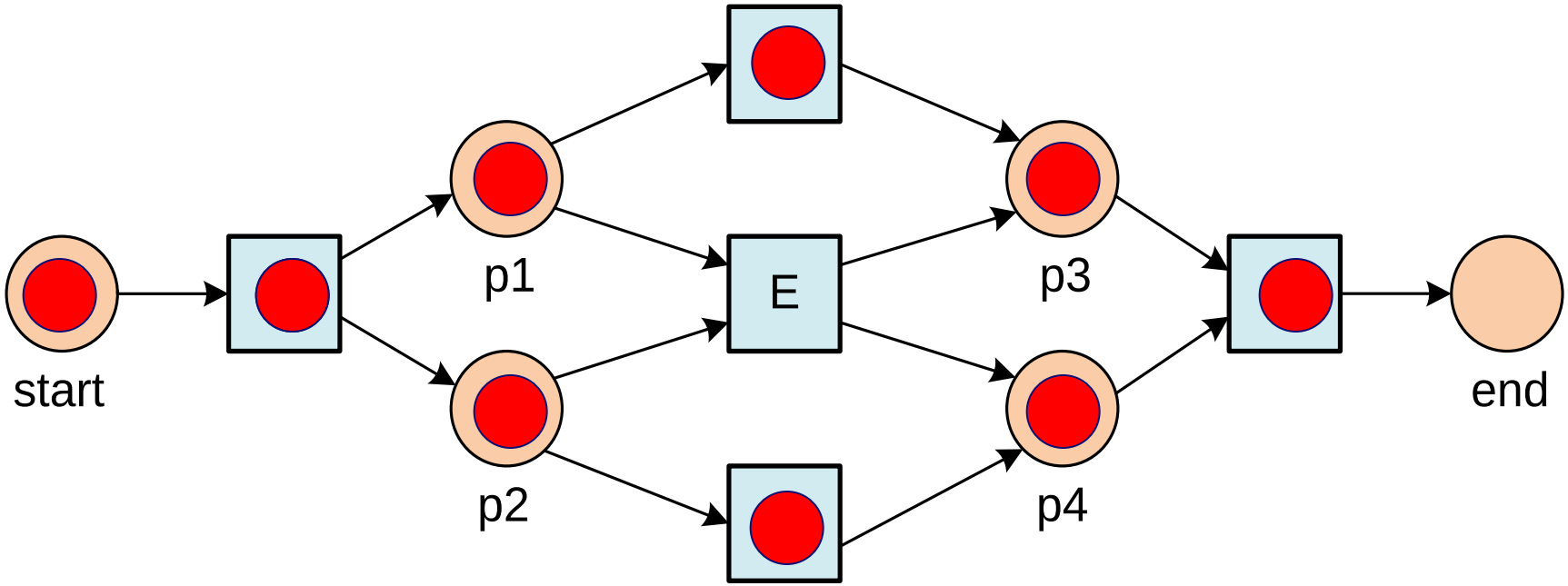
event log

process model

- extended model showing times, frequencies, etc.
- diagnostics
- predictions
- recommendations

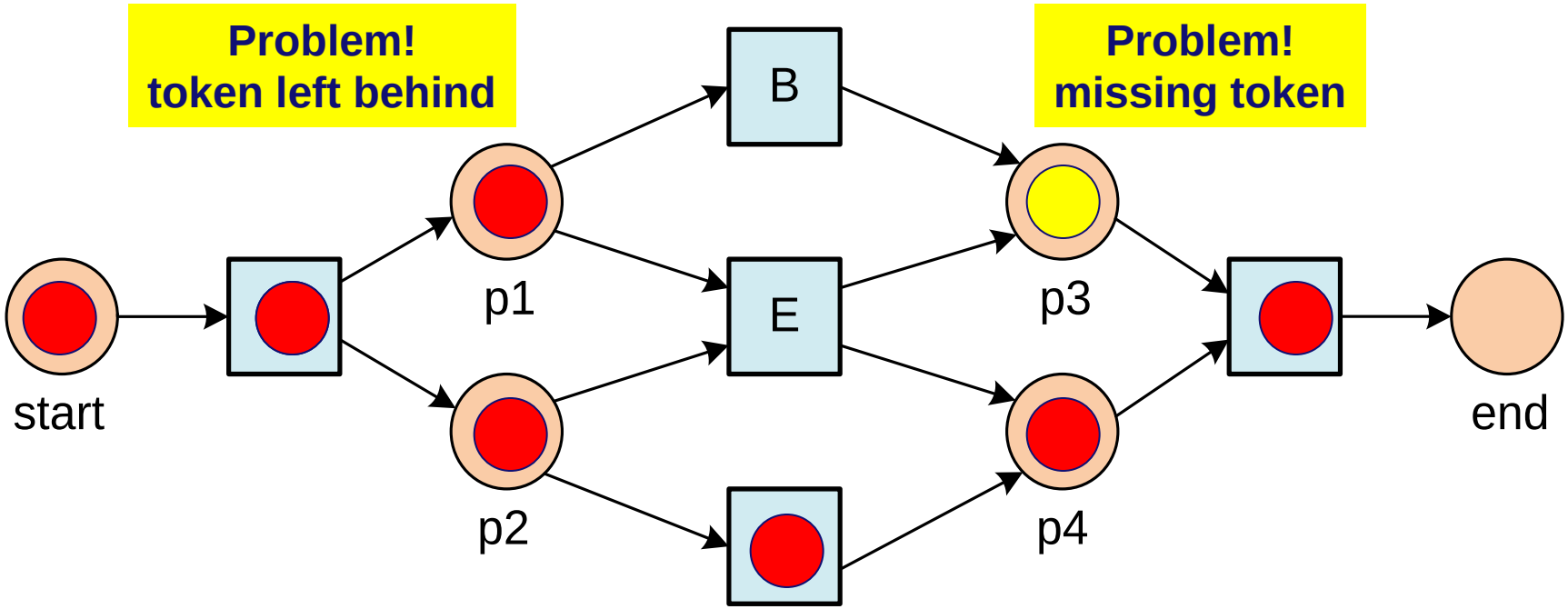
Replay

A B C D



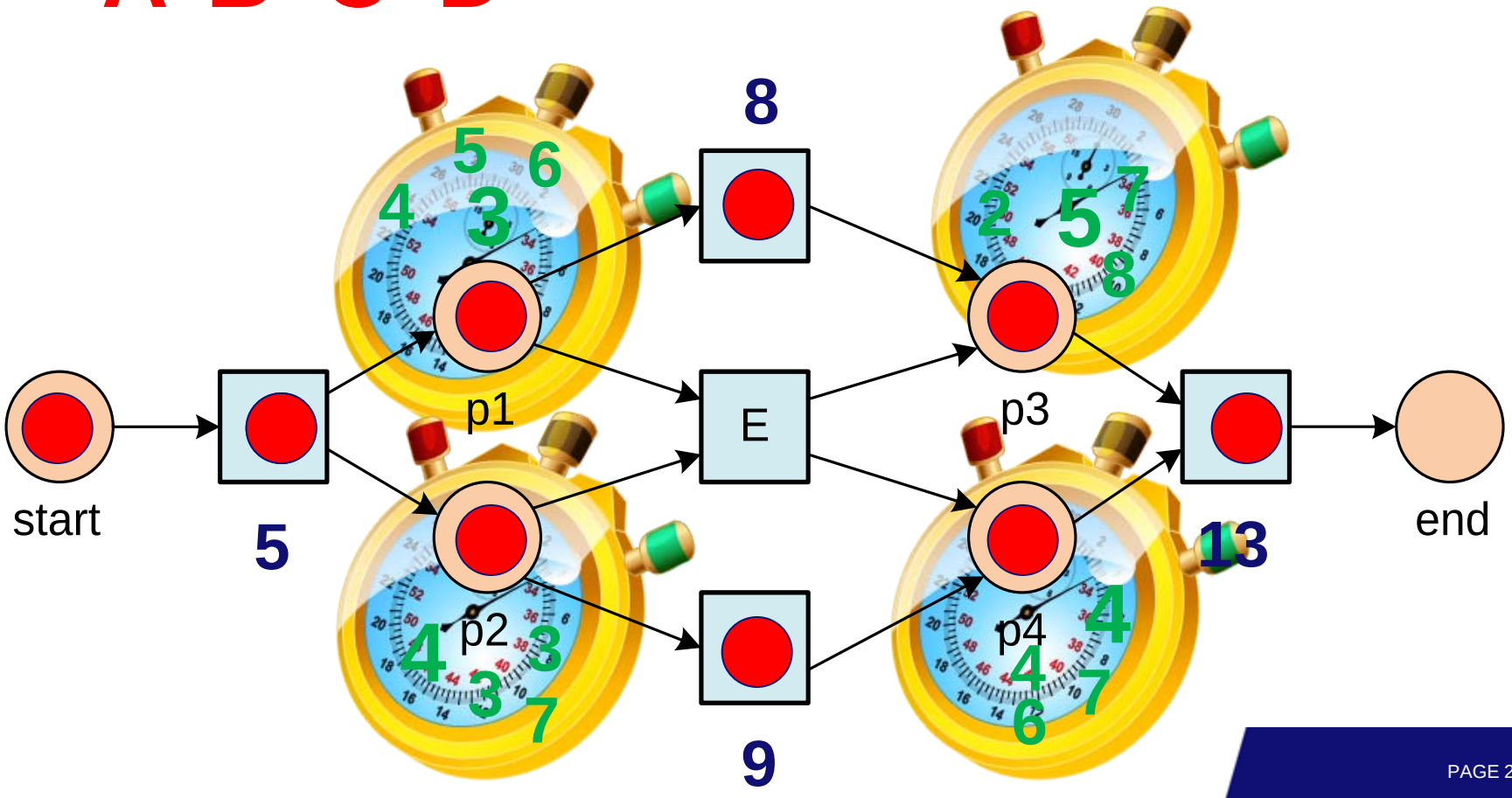
Replay can detect problems

A C D

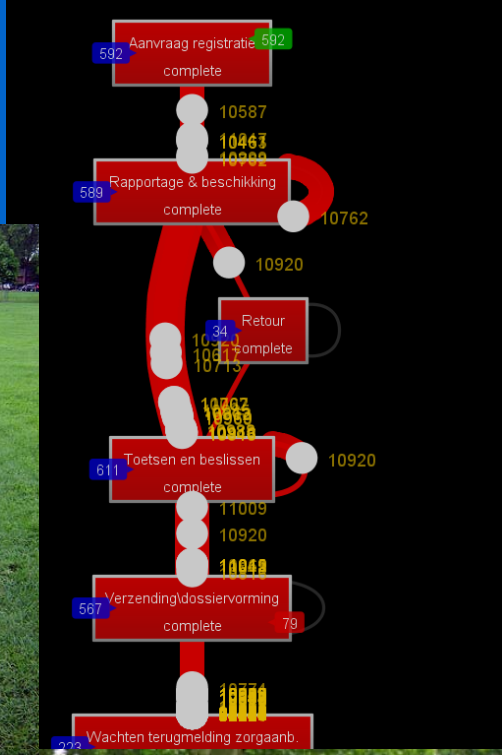


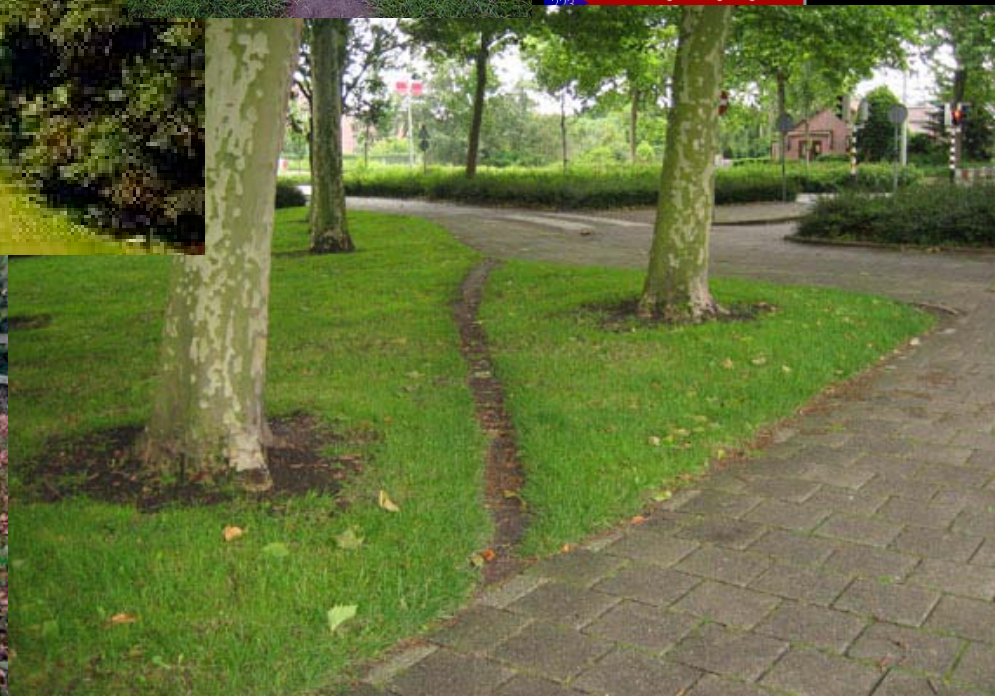
Replay can extract timing information

$A^5 B^8 C^9 D^{13}$



Desire lines in process models





Process Discovery (small selection)

automata-based learning

distributed genetic mining

language-based regions

heuristic mining

state-based regions

partial-order based mining

genetic mining

LTL mining

stochastic task graphs

neural networks

fuzzy mining

mining block structures

hidden Markov models

α algorithm

multi-phase mining

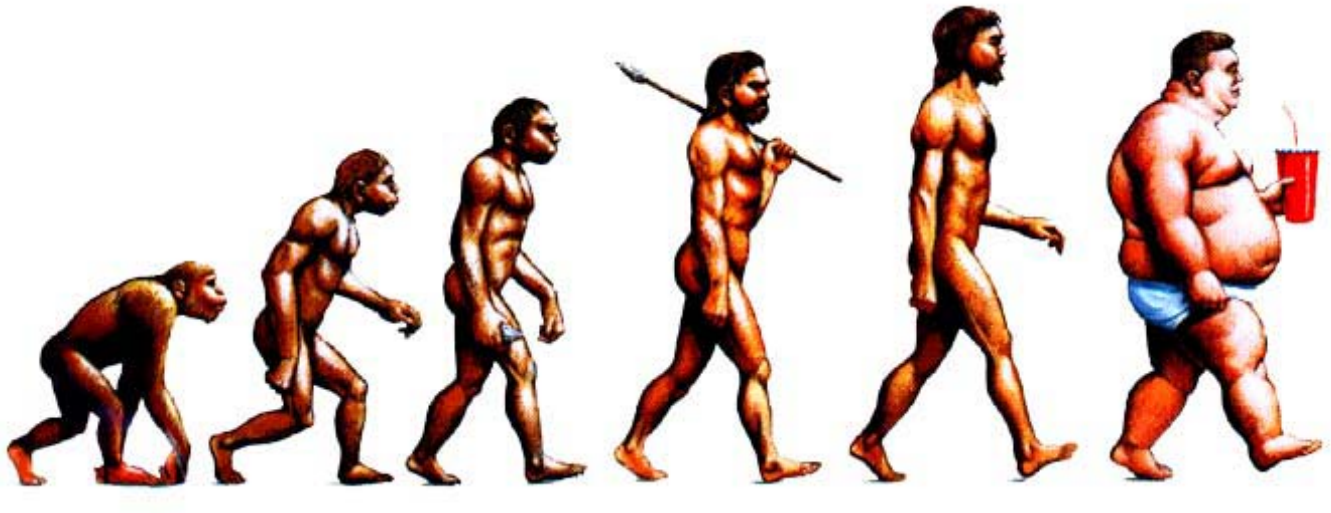
conformal process graph

α# algorithm

ILP mining

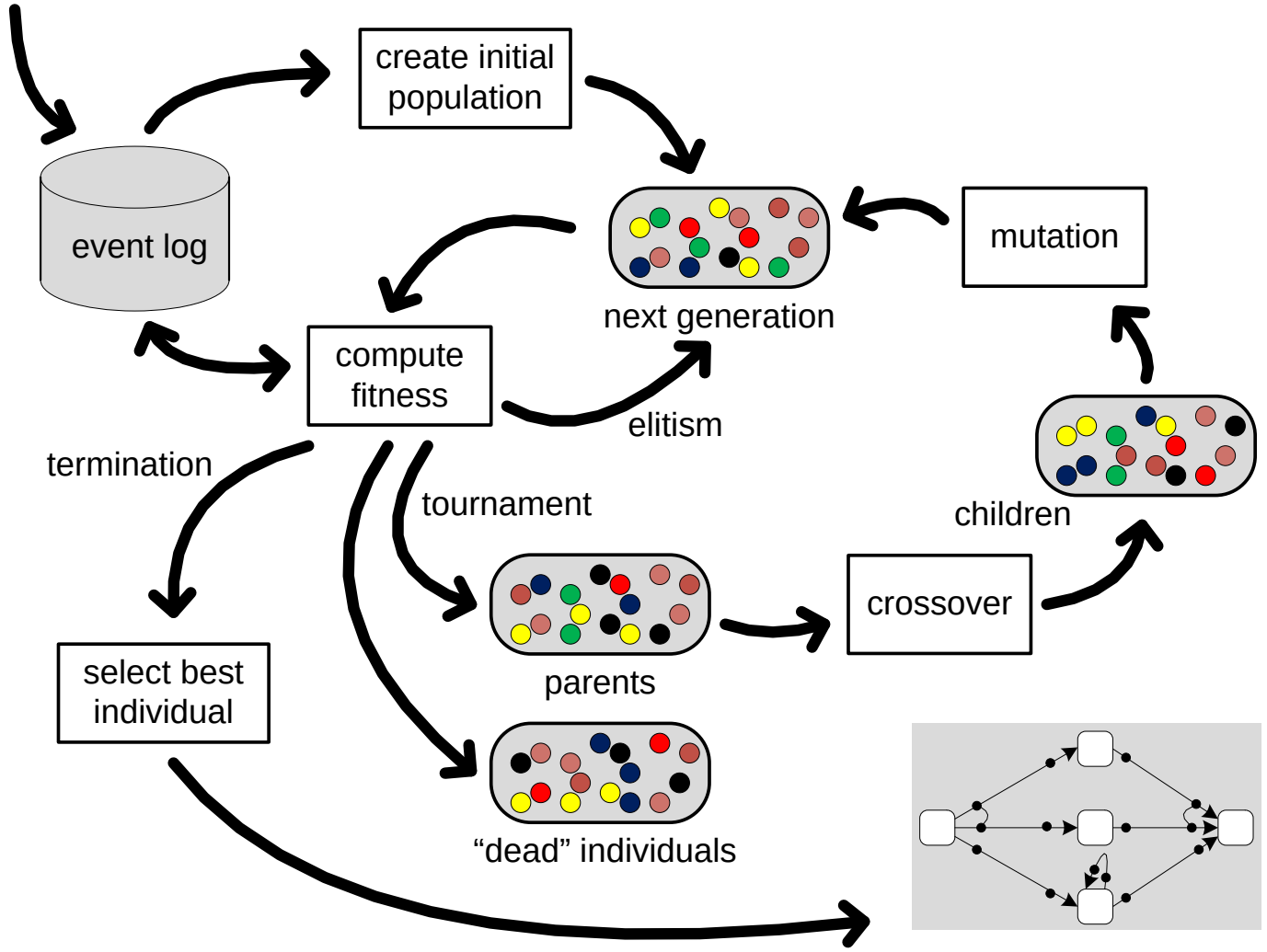
α++ algorithm

Example of a process discovery technique: Genetic Mining

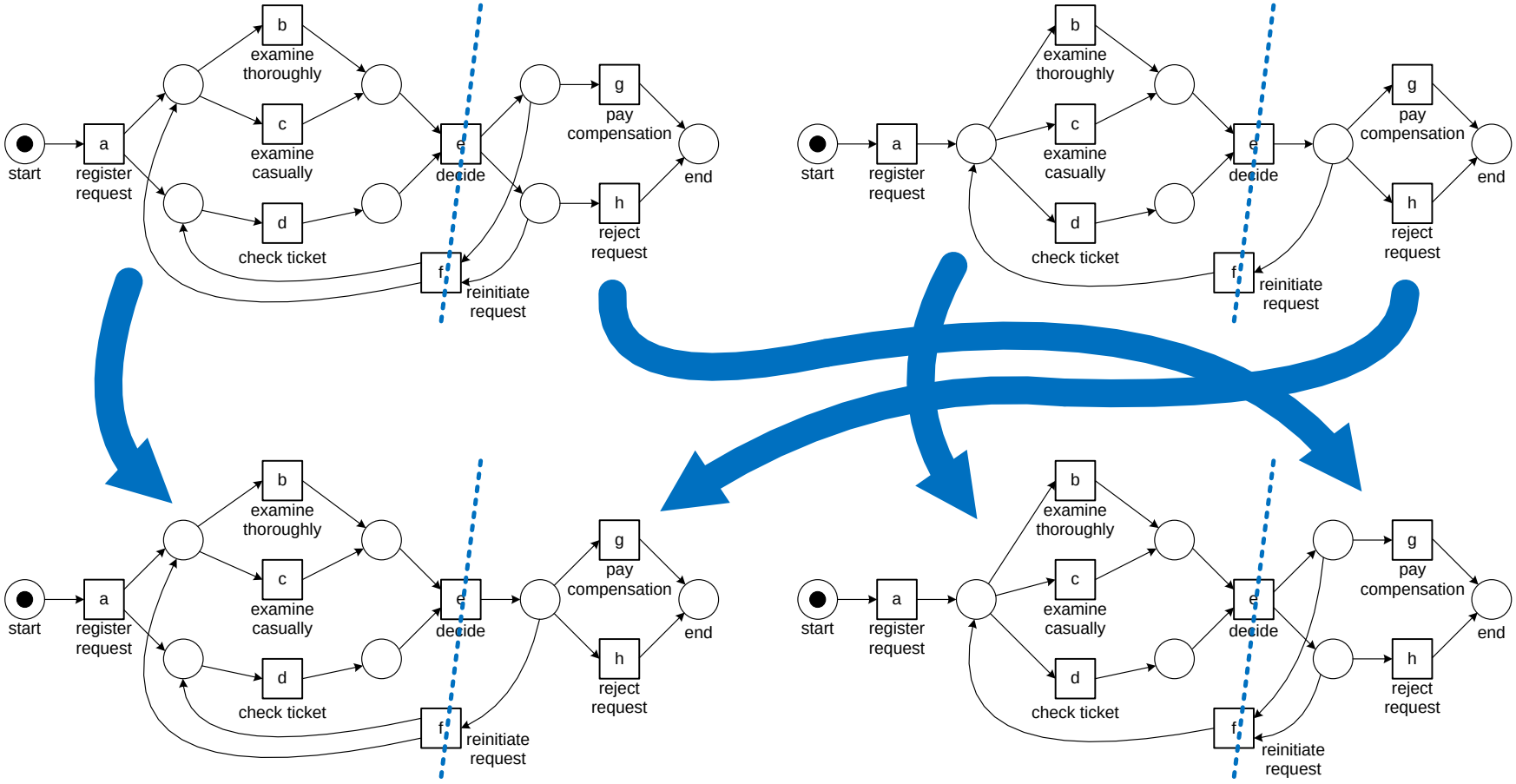


- **Discovery algorithms have become readily available.**
- **Genetic mining:**
  - **requires a lot of computing power, but can be distributed easily,**
  - **can deal with noise, infrequent behavior, duplicate tasks, invisible tasks,**
  - **allows for incremental improvement and combinations with other approaches (heuristics post-optimization, etc.).**

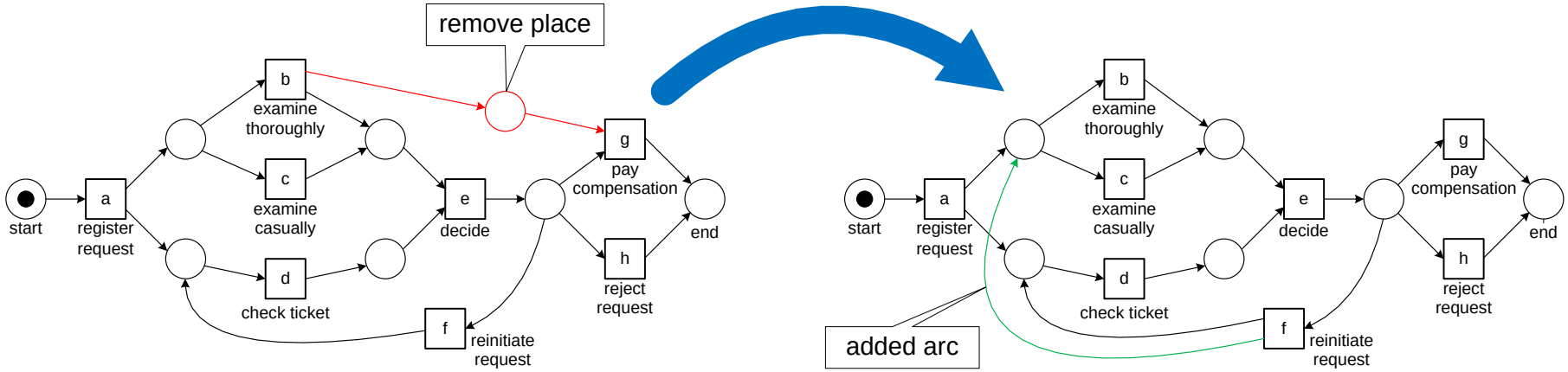
Genetic process mining: Overview



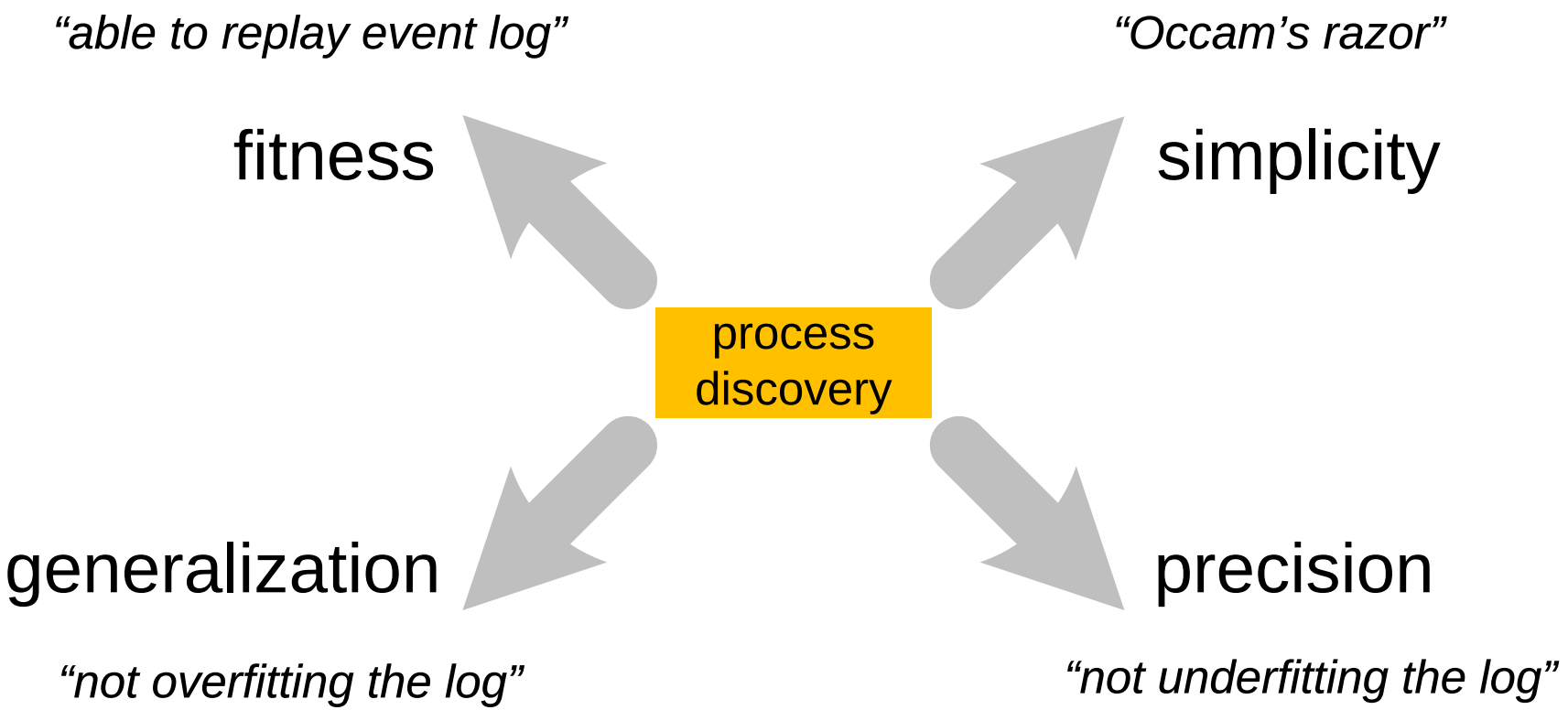
Example: crossover



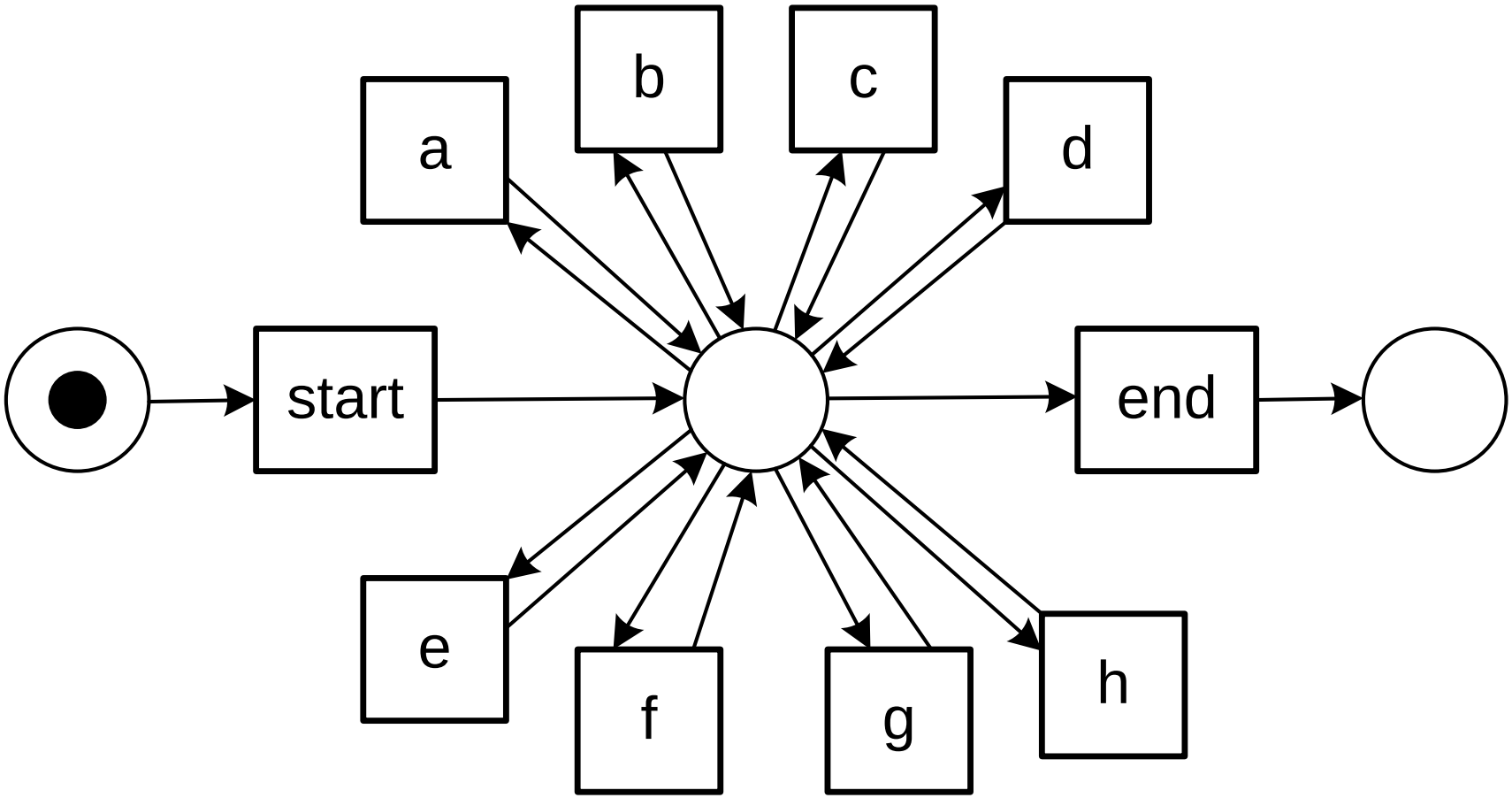
Example: mutation



Challenge: four competing quality criteria

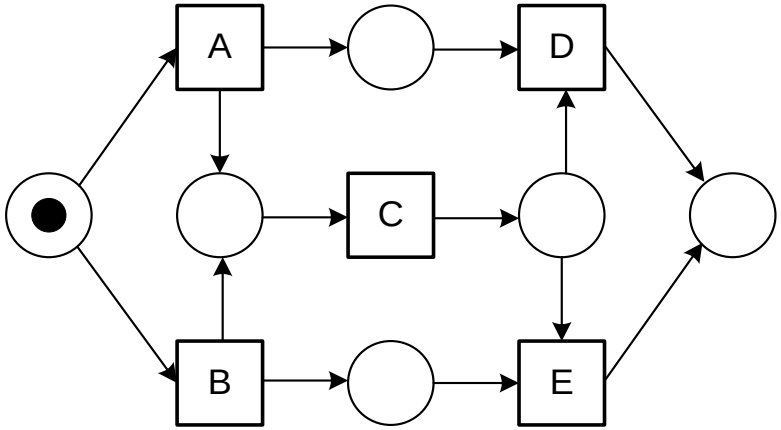


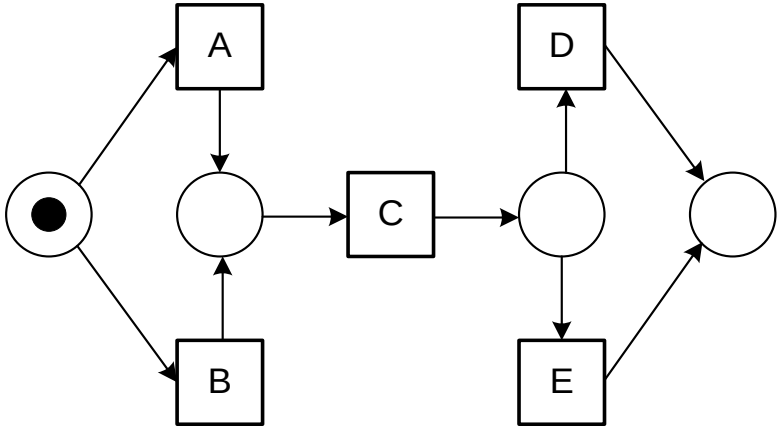
Flower model



What is the best model?

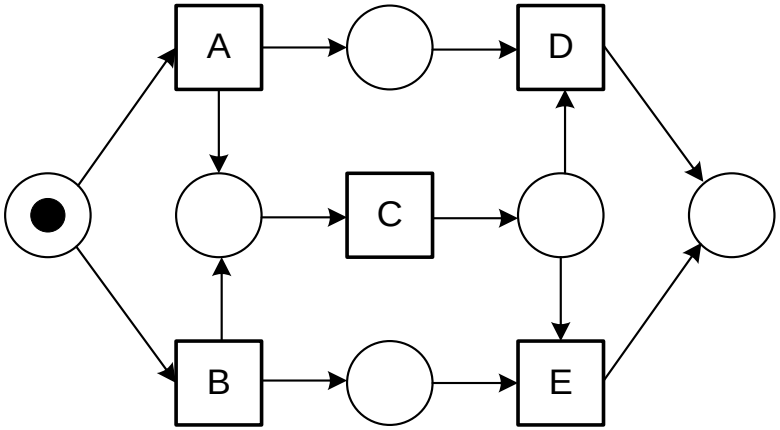
ACD	**99**
ACE	**0**
BCE	**85**
BCD	**0**

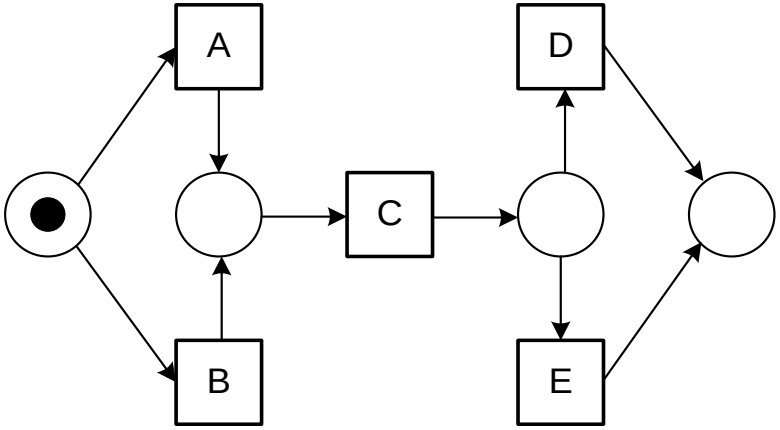




What is the best model?

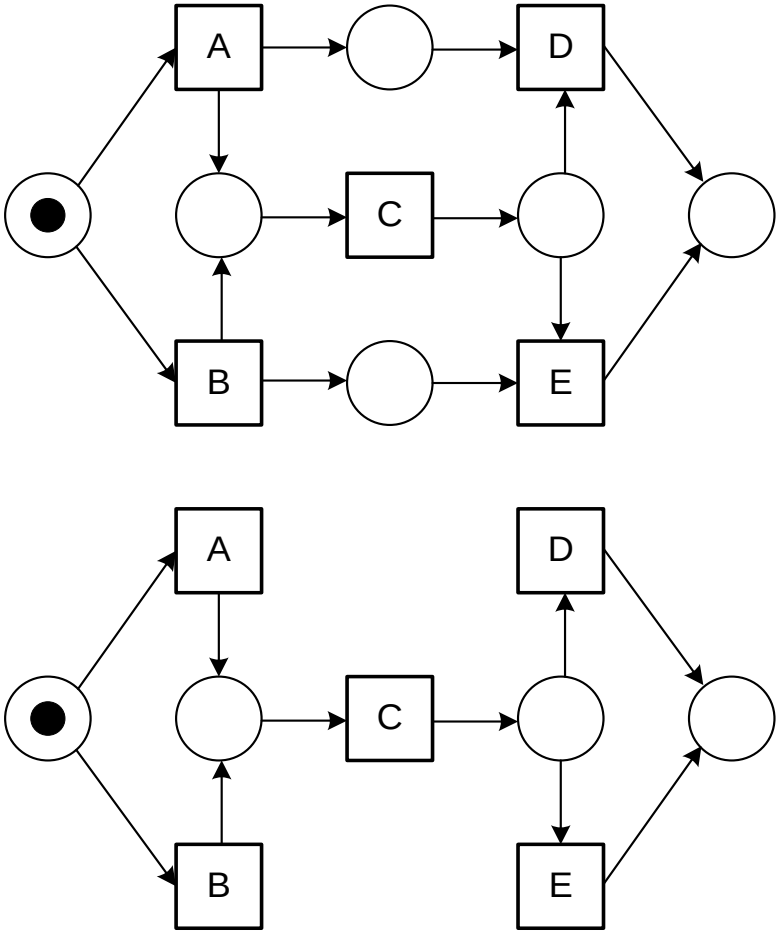
ACD	99
ACE	88
BCE	85
BCD	78



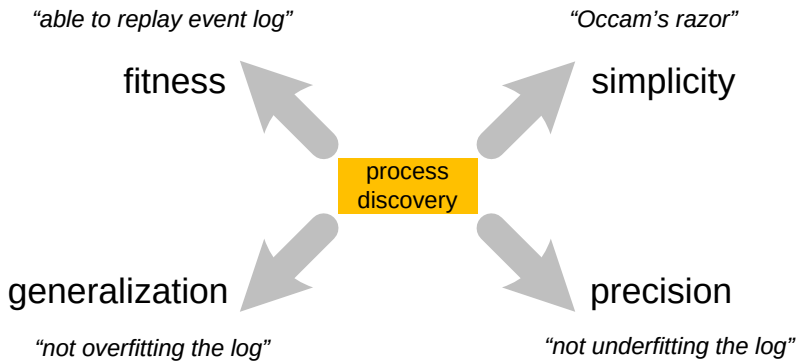


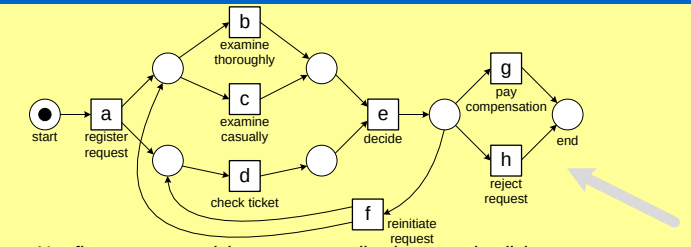
What is the best model?

ACD	**99**
ACE	**2**
BCE	**85**
BCD	**3**

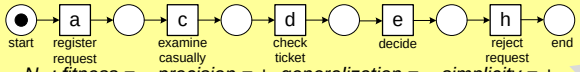


Example: one log four models

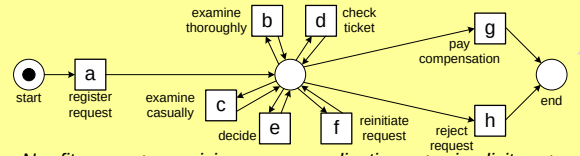




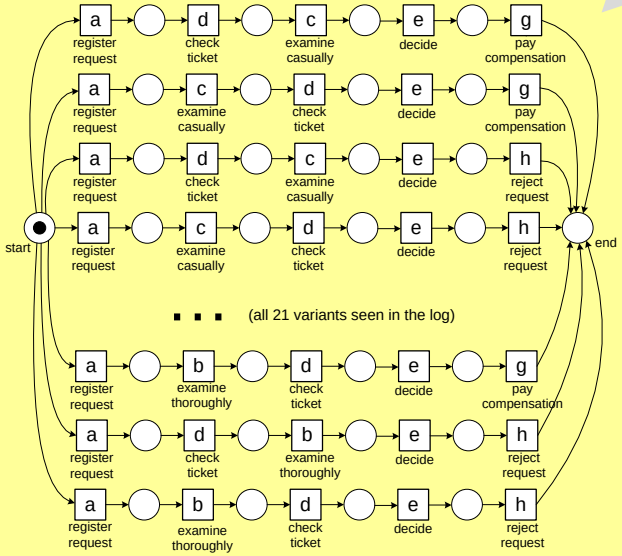
N_1 : *fitness* = +, *precision* = +, *generalization* = +, *simplicity* = +



N_2 : *fitness* = -, *precision* = +, *generalization* = -, *simplicity* = +



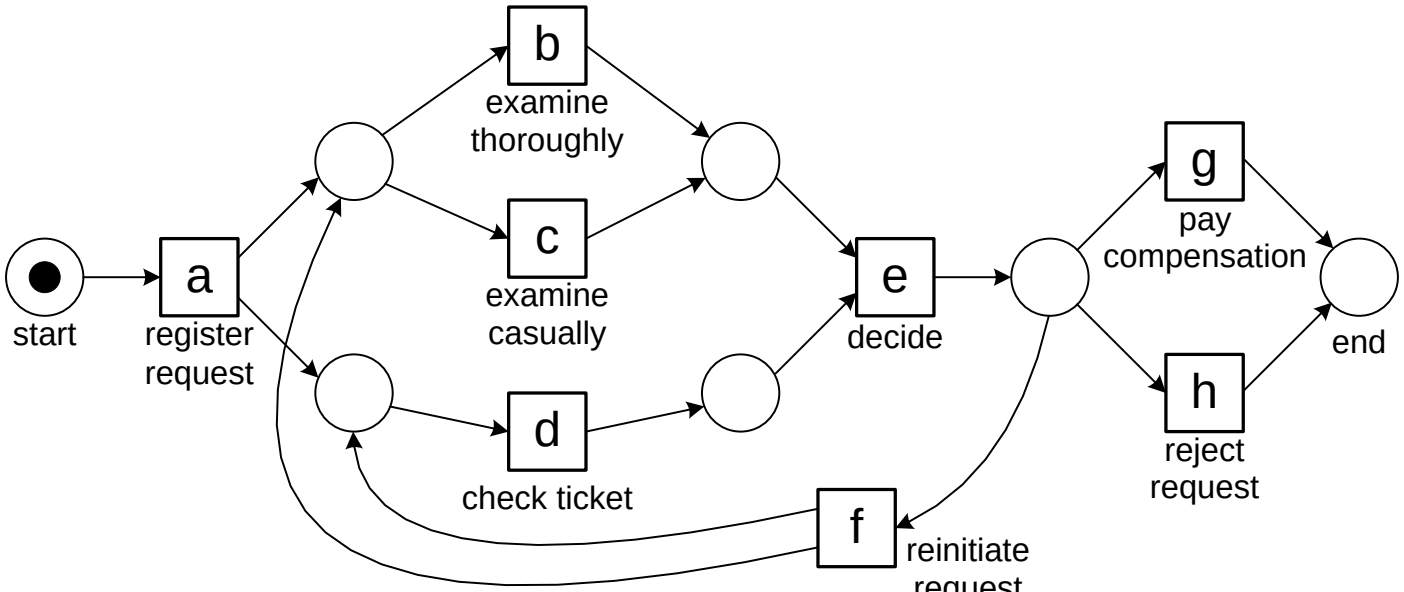
N_3 : *fitness* = +, *precision* = -, *generalization* = +, *simplicity* = +



N_4 : *fitness* = +, *precision* = +, *generalization* = -, *simplicity* = -

#	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	

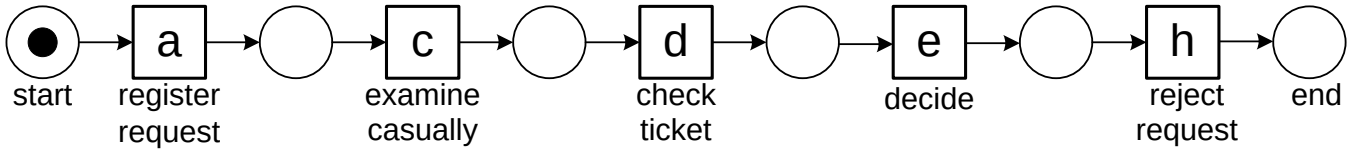
Model N_1



N_1 : fitness = +, precision = +, generalization = +, simplicity = +

#	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	

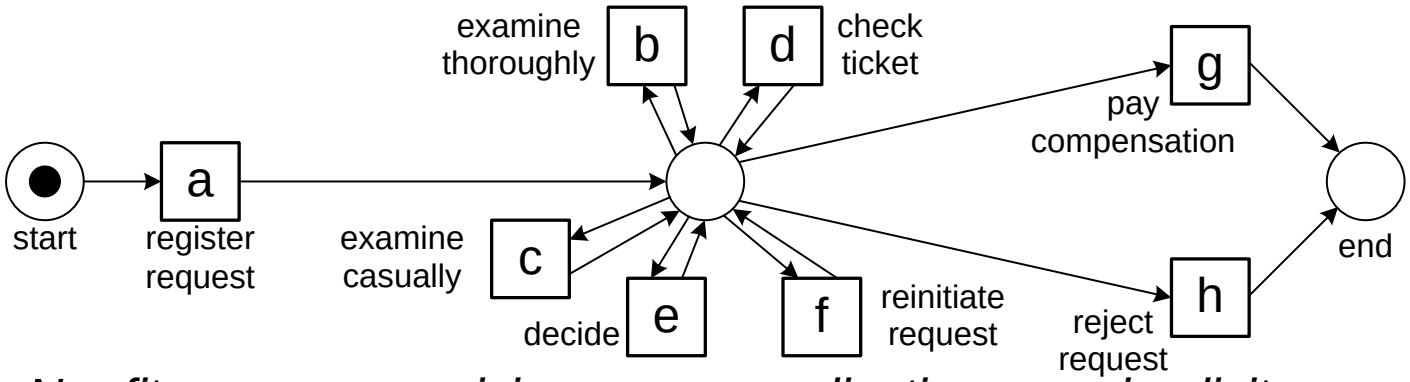
Model N_2



N_2 : fitness = -, precision = +, generalization = -, simplicity = +

#	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	

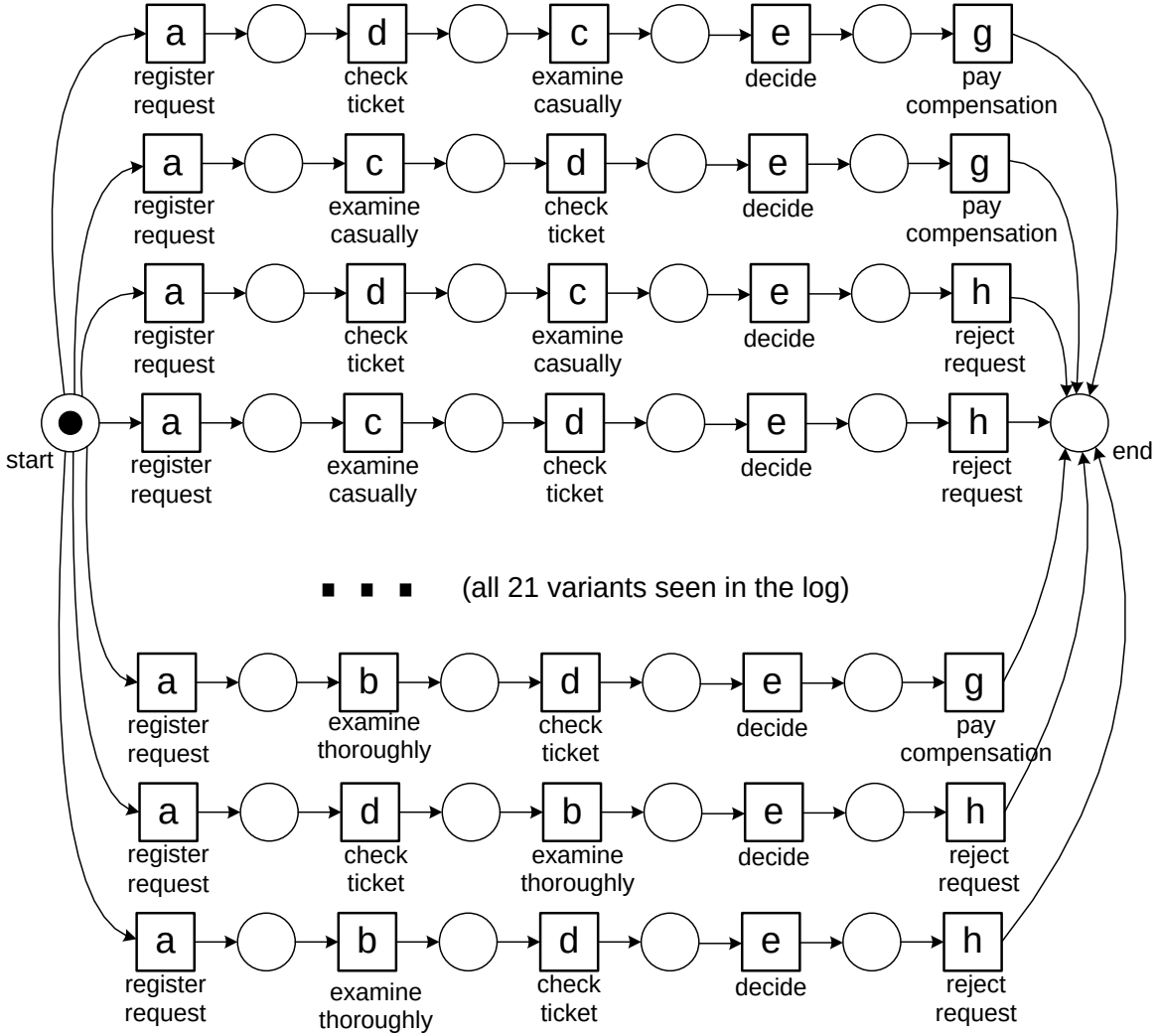
Model N_3



N_3 : fitness = +, precision = -, generalization = +, simplicity = +

#	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	

Model N_4

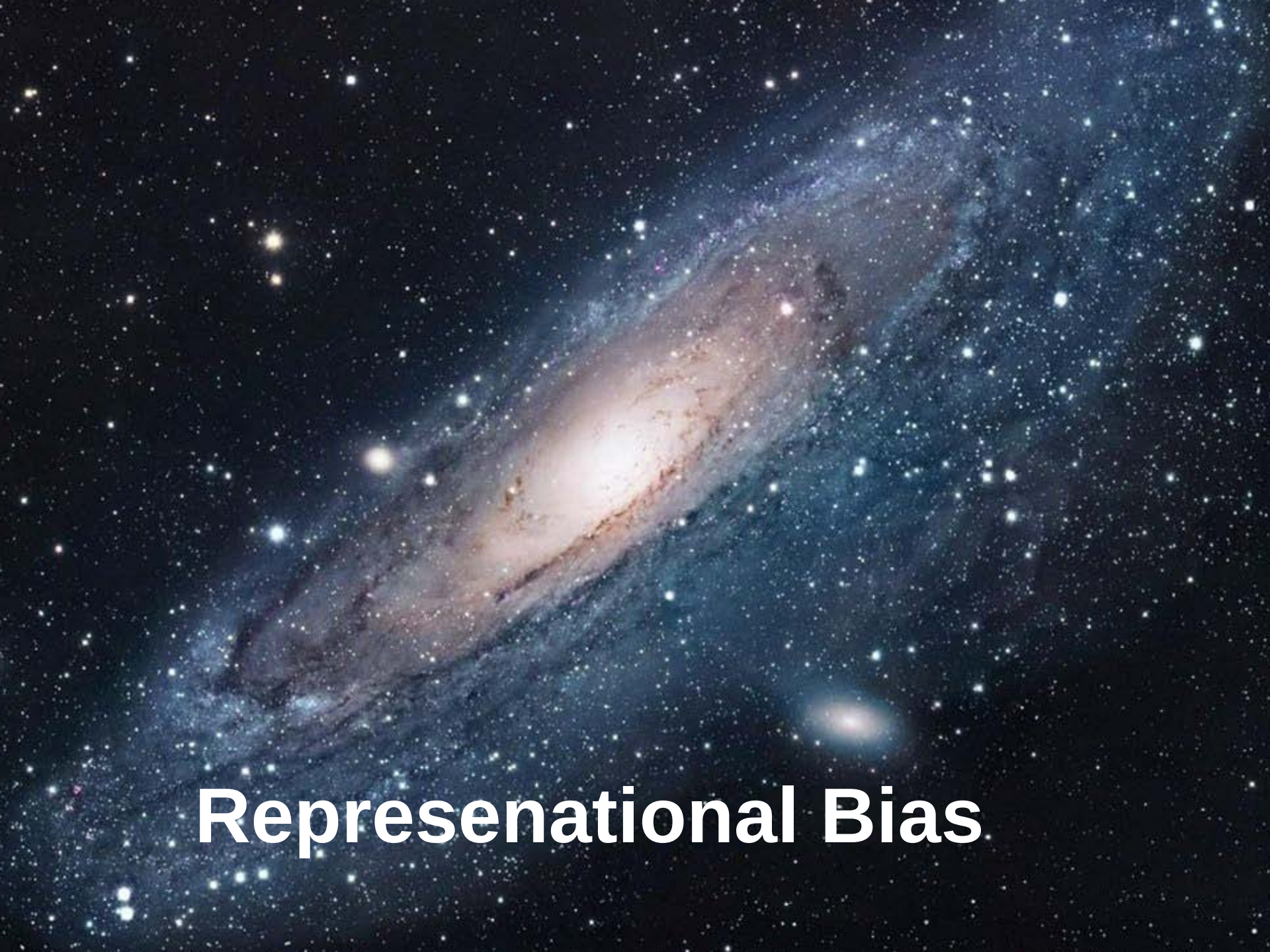


N_4 : fitness = +, precision = +, generalization = -, simplicity = -

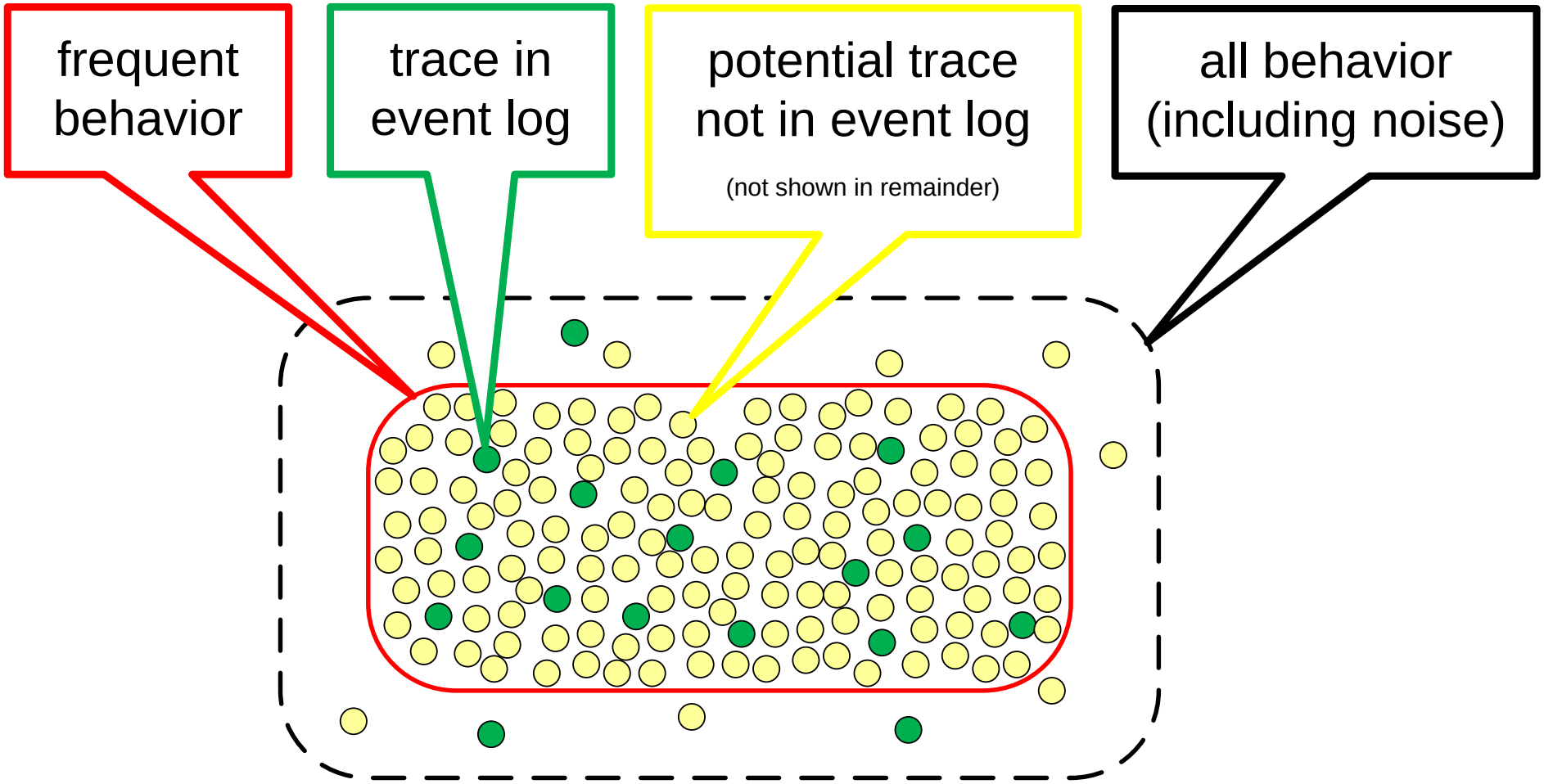
#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

Why is process mining such a difficult problem?

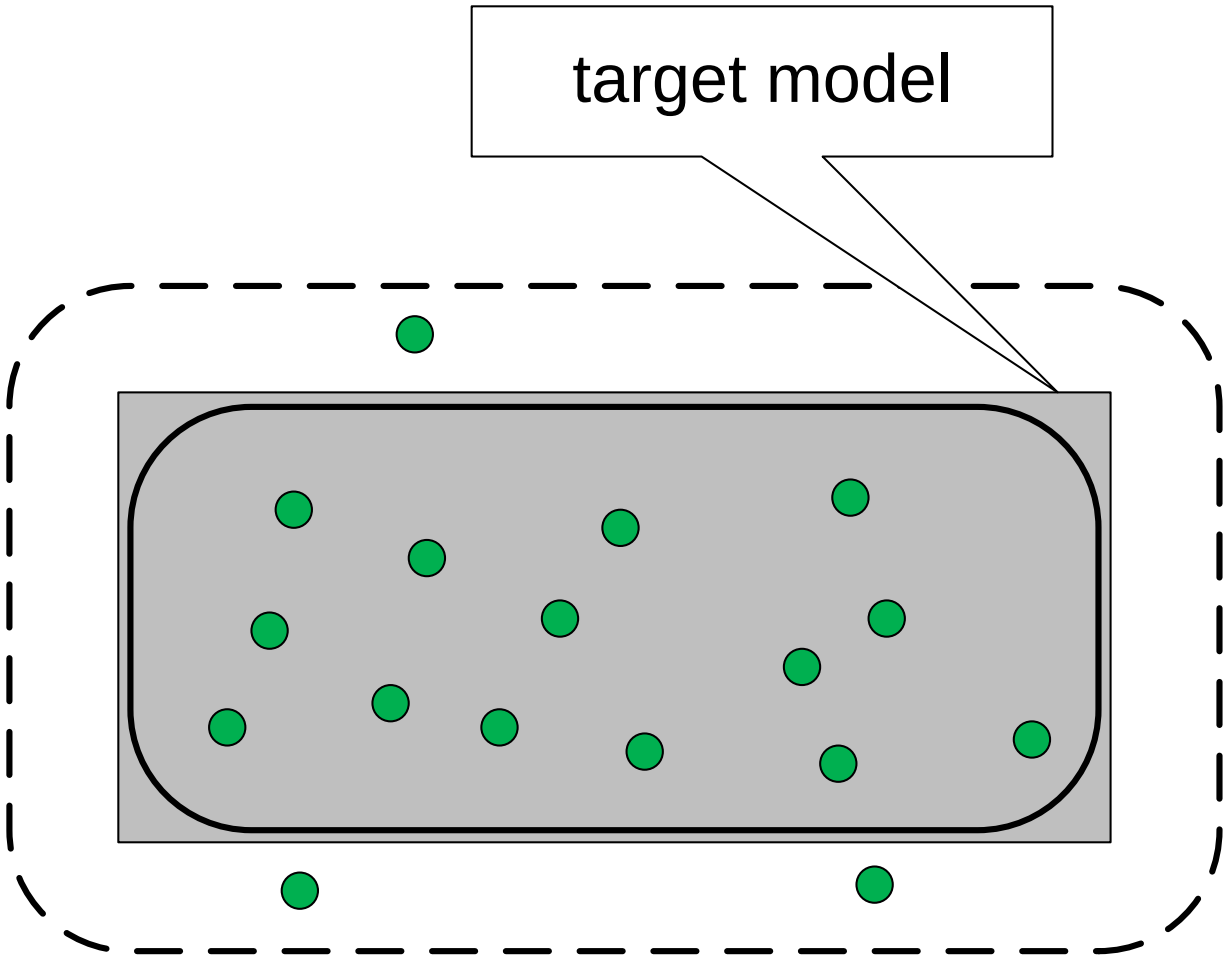
- There are no negative examples (i.e., a log shows what has happened but does not show what could not happen).
- Due to concurrency, loops, and choices the search space has a complex structure and the log typically contains only a fraction of all possible behaviors.
- There is no clear relation between the size of a model and its behavior (i.e., a smaller model may generate more or less behavior although classical analysis and evaluation methods typically assume some monotonicity property).


Represenational Bias

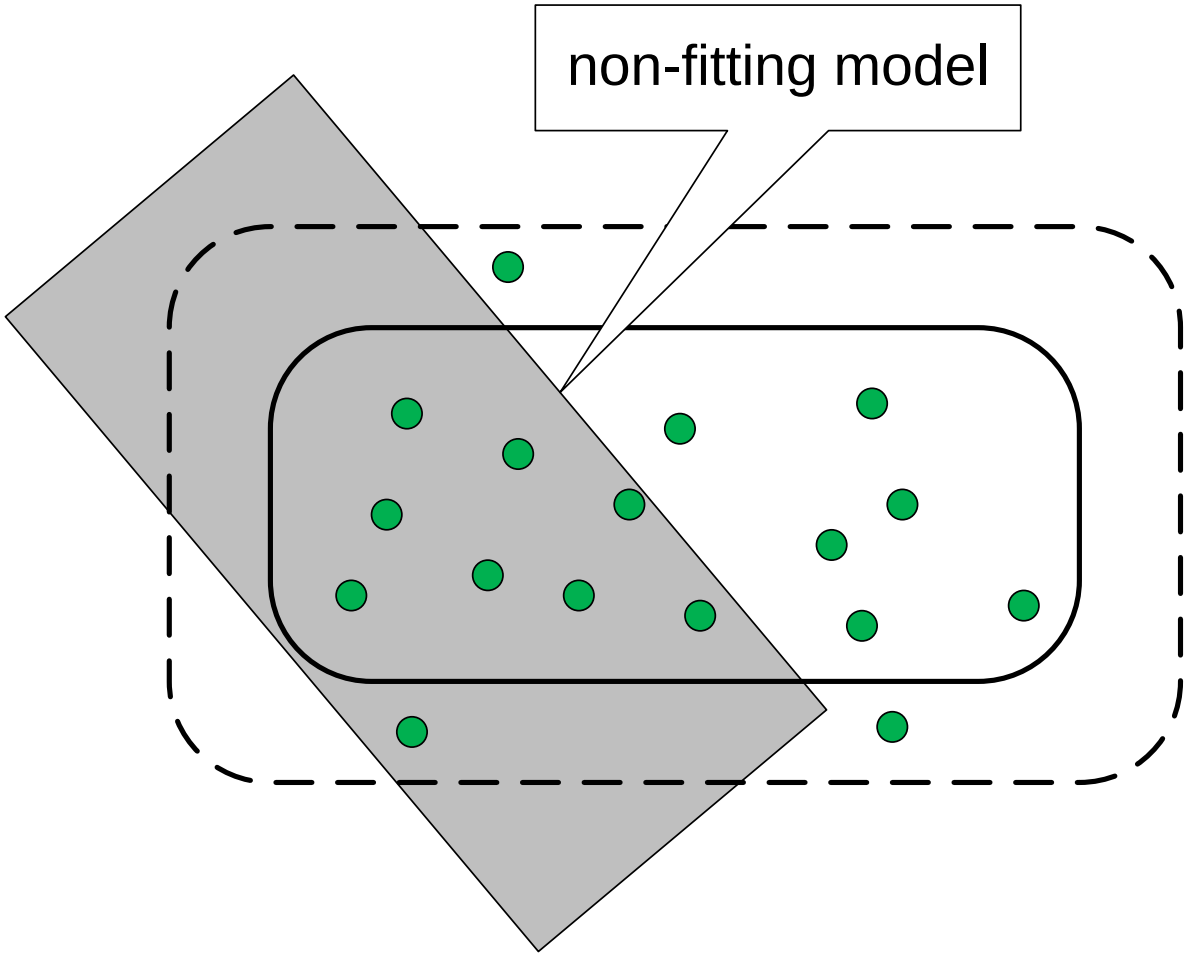
Finding the right model based on *example* traces in event log



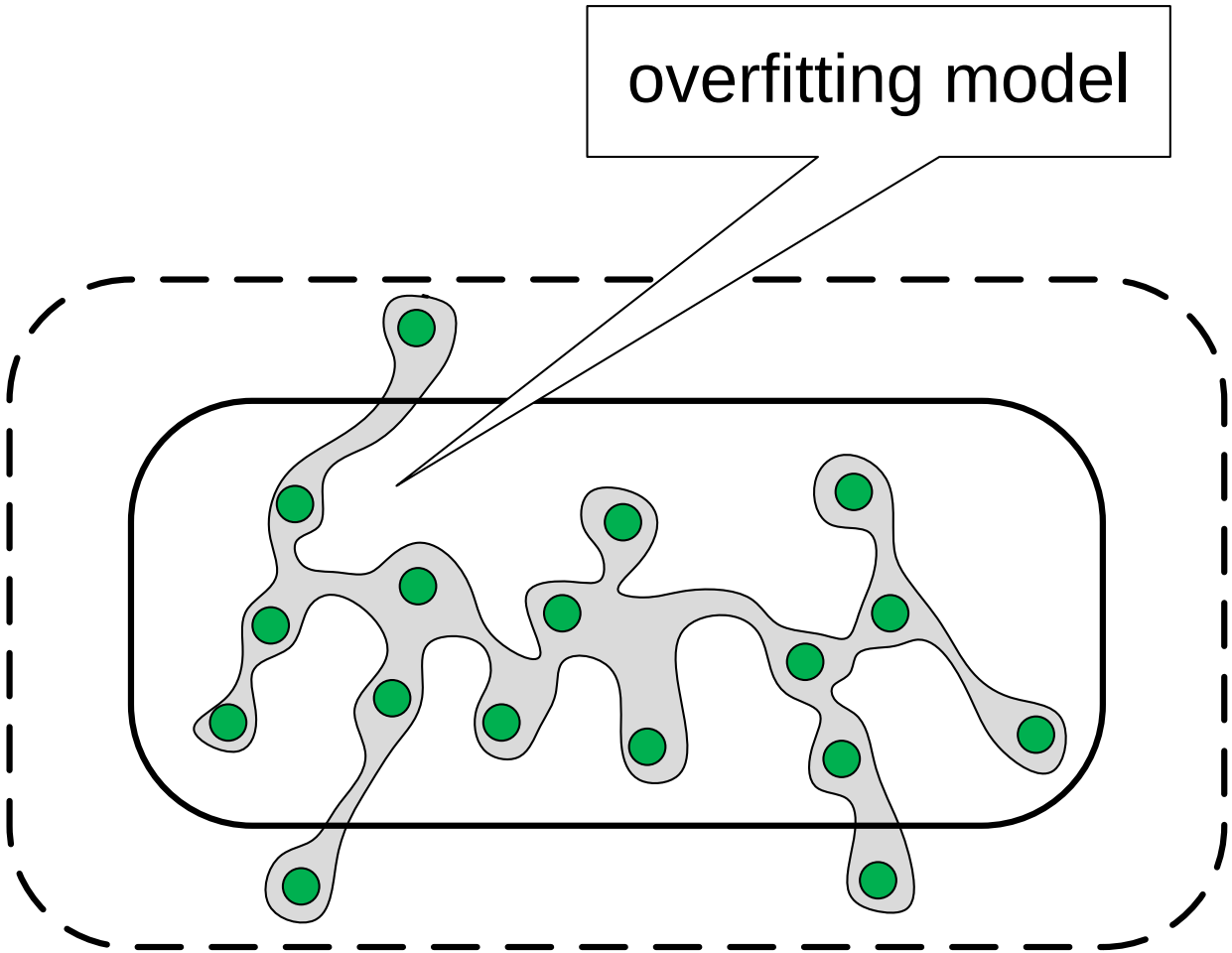
Goal



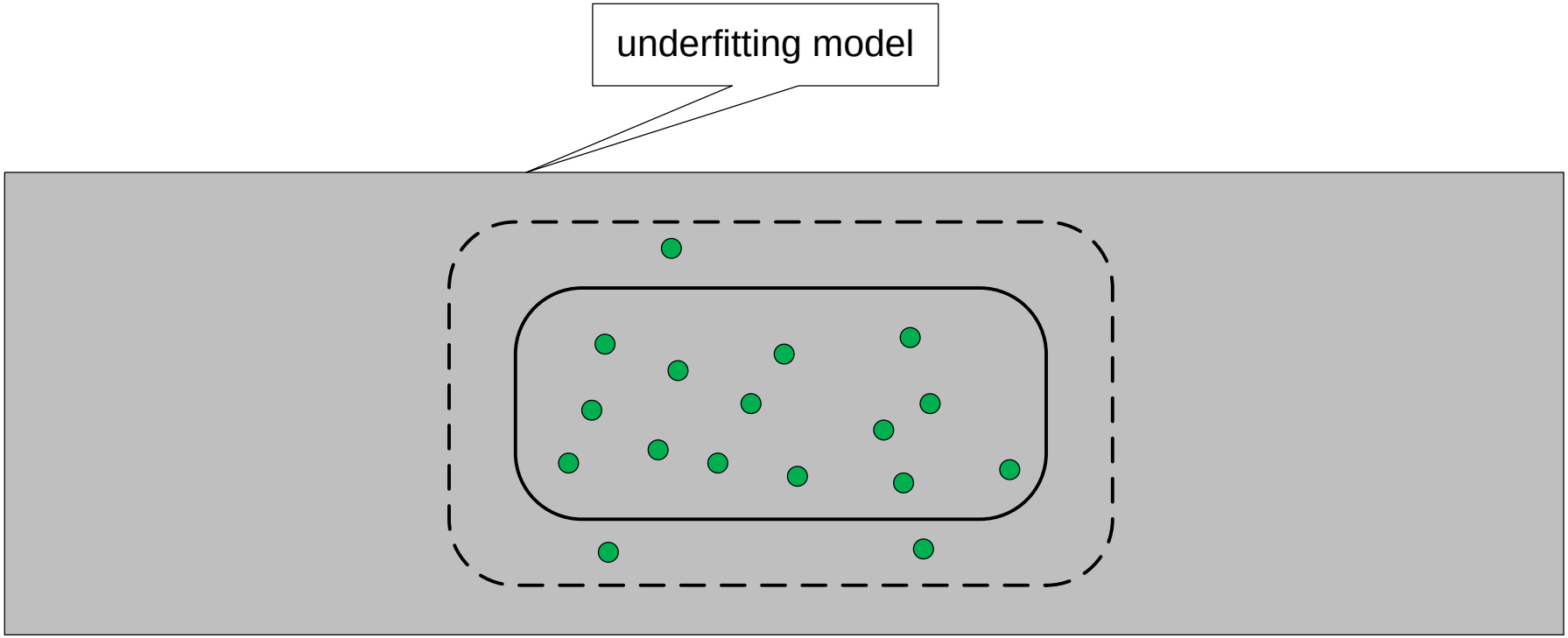
Problem 1: Lack of fitness



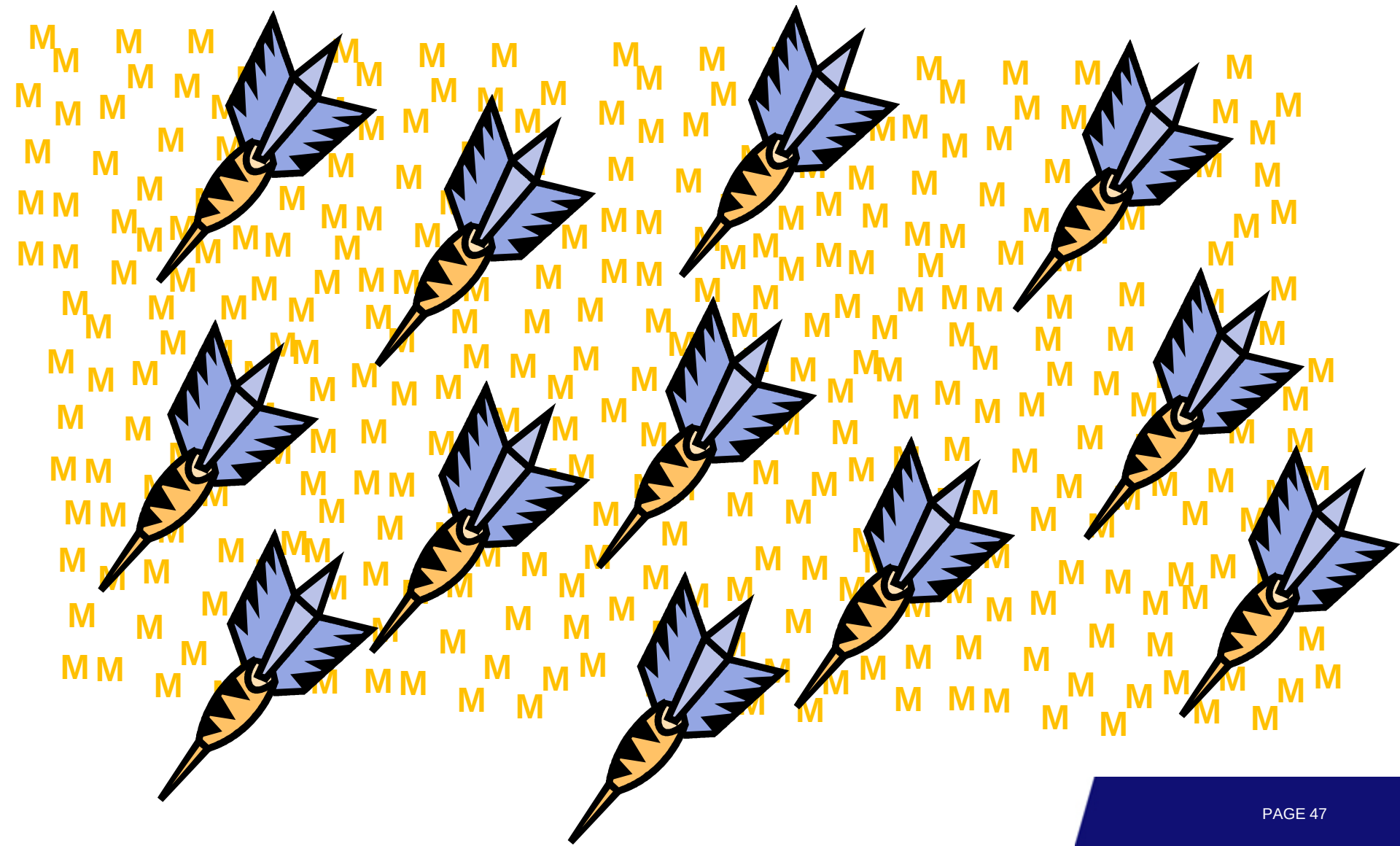
Problem 2: Overfitting



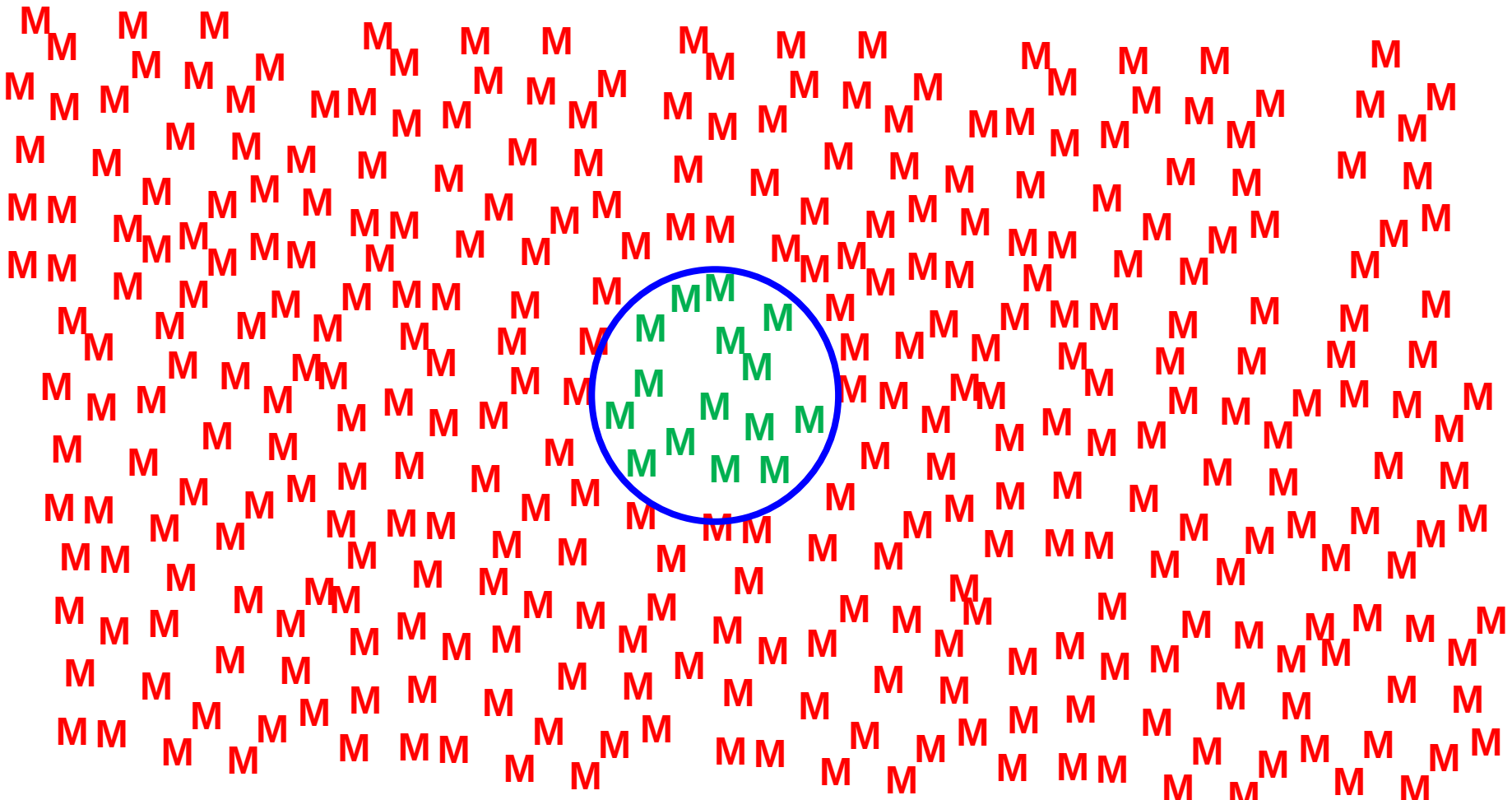
Problem 3: Underfitting



Search space

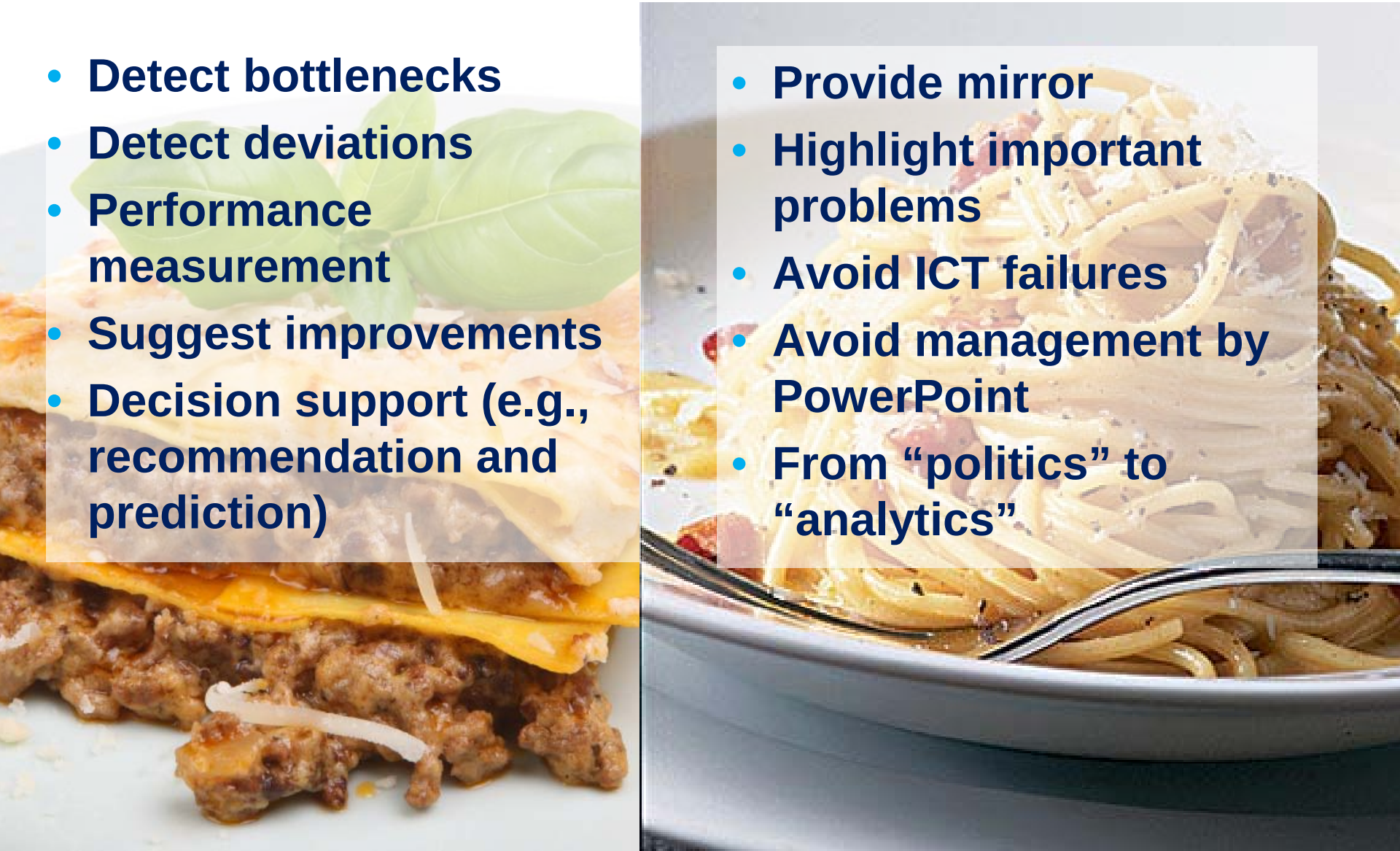


Problem



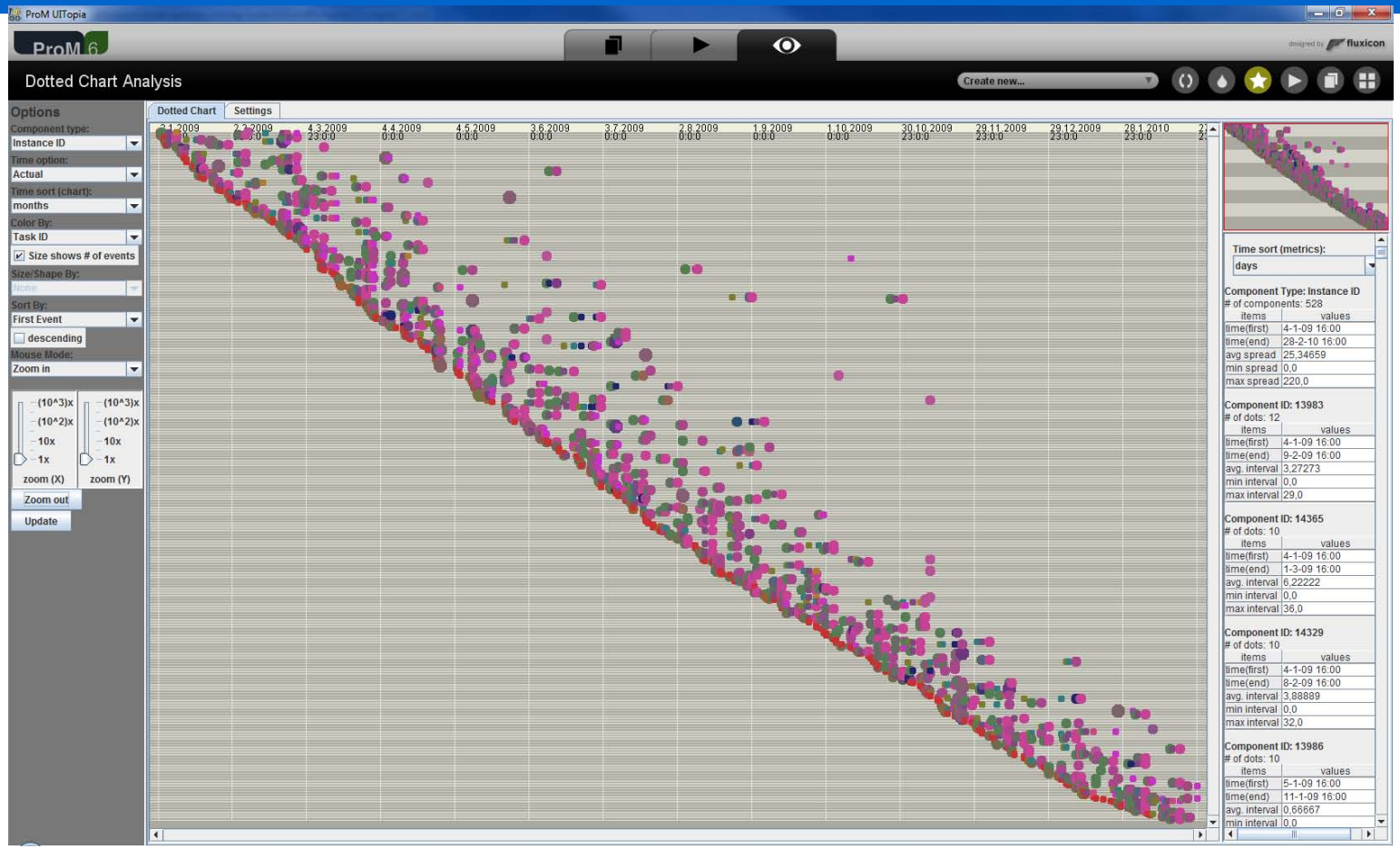
How can process mining help?

- Detect bottlenecks
- Detect deviations
- Performance measurement
- Suggest improvements
- Decision support (e.g., recommendation and prediction)



- Provide mirror
- Highlight important problems
- Avoid ICT failures
- Avoid management by PowerPoint
- From “politics” to “analytics”

Example of a Lasagna process: WMO process of a Dutch municipality

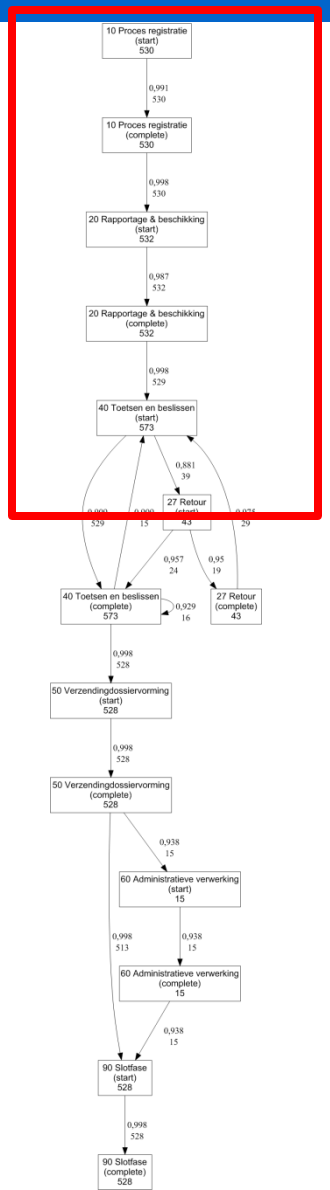


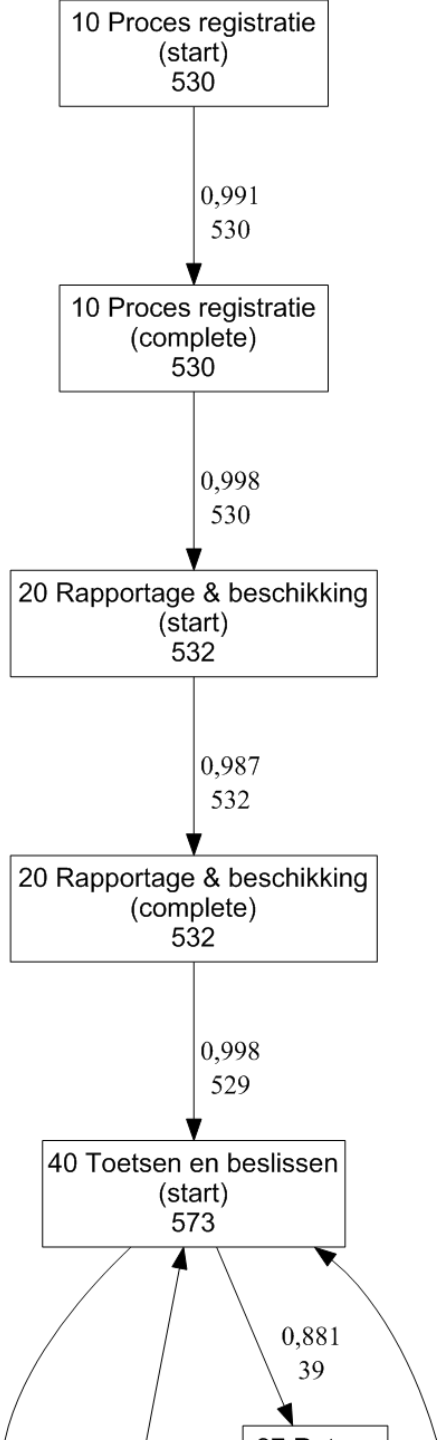
Each line corresponds to one of the 528 requests that were handled in the period from 4-1-2009 until 28-2-2010. In total there are 5498 events represented as dots. The mean time needed to handled a case is approximately 25 days.

WMO process
(Wet Maatschappelijke Ondersteuning)

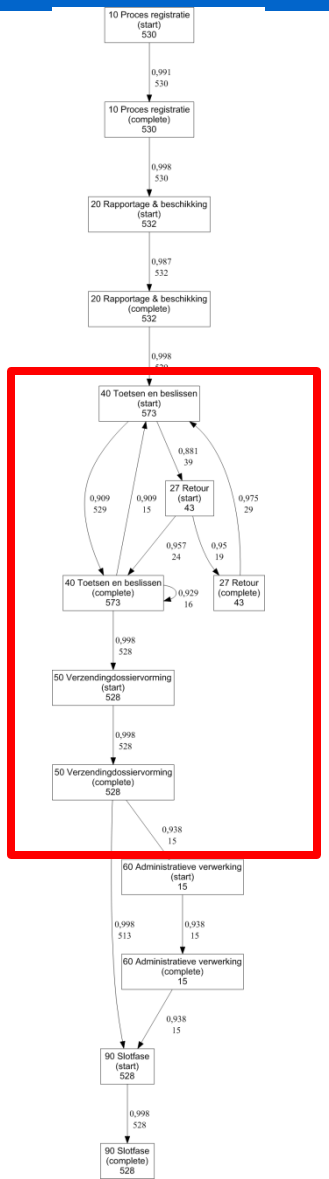
- WMO refers to the social support act that came into force in The Netherlands on January 1st, 2007.
- The aim of this act is to assist people with disabilities and impairments. Under the act, local authorities are required to give support to those who need it, e.g., household help, providing wheelchairs and scootmobiles, and adaptations to homes.
- There are different processes for the different kinds of help. We focus on the process for handling requests for household help.
- In a period of about one year, 528 requests for household WMO support were received.
- These 528 requests generated 5498 events.

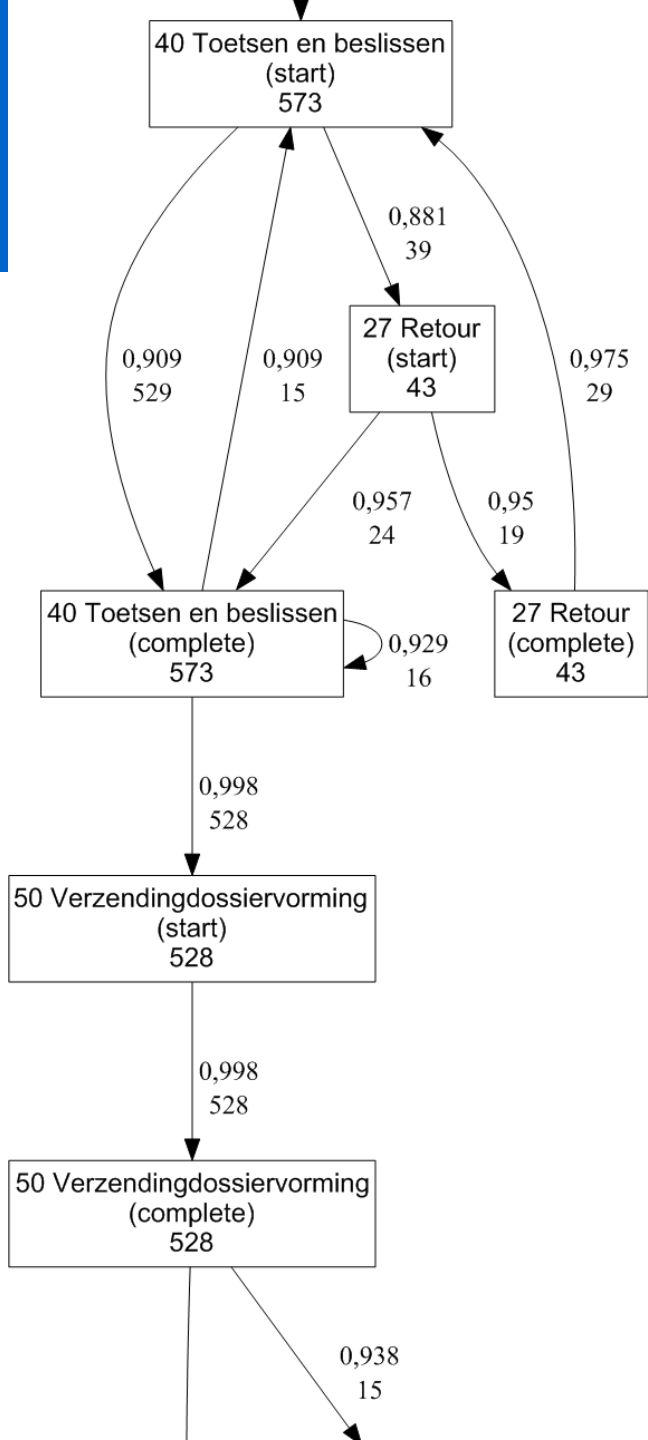
C-net discovered using heuristic miner (1/3)



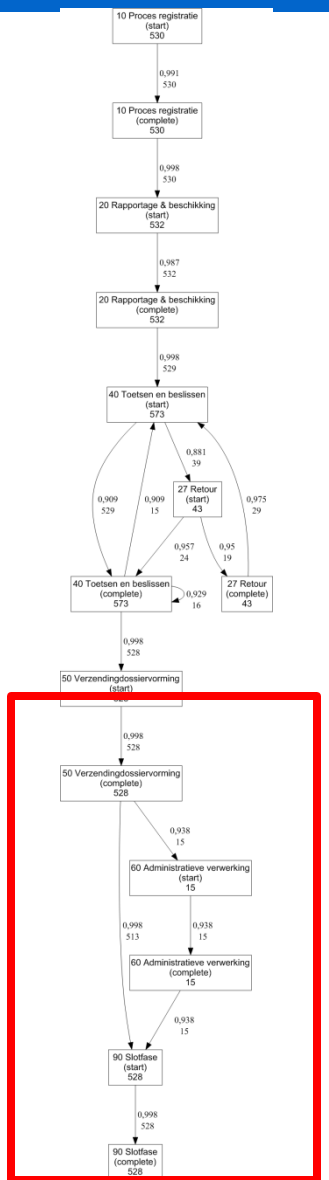


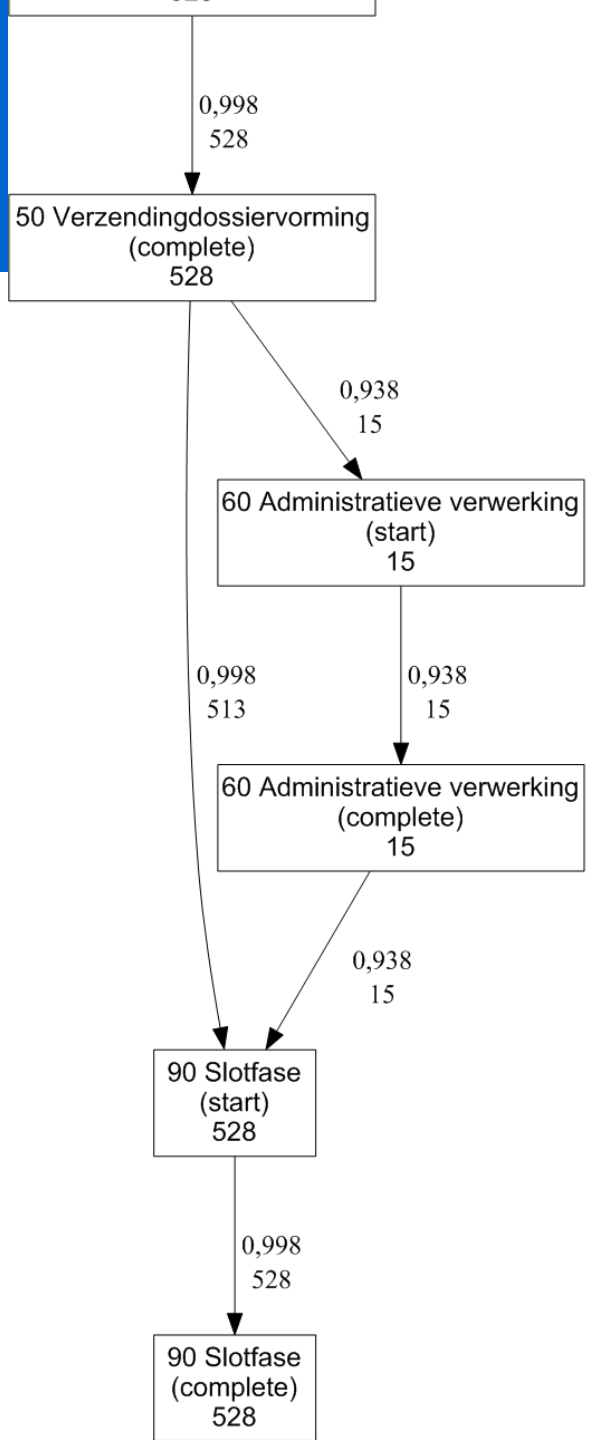
C-net discovered using heuristic miner (2/3)



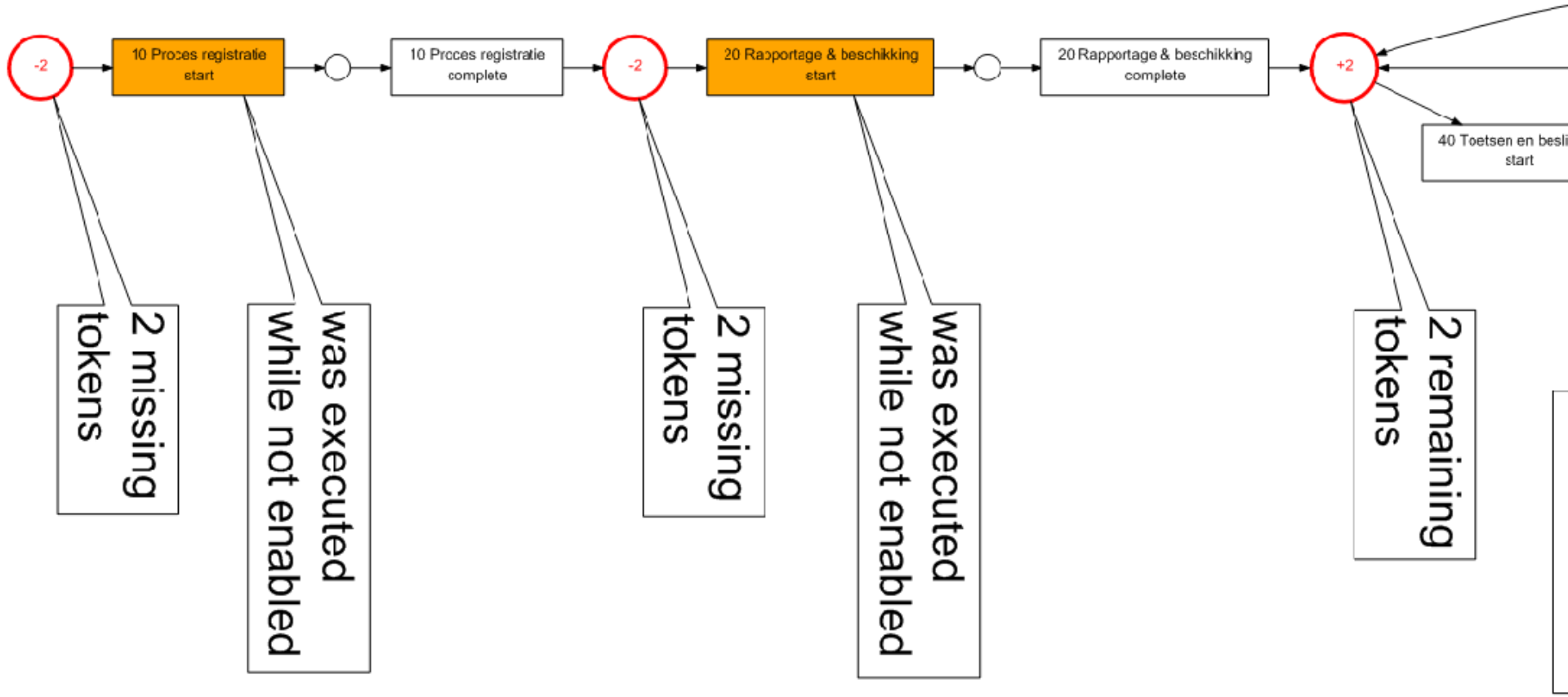


C-net discovered using heuristic miner (3/3)

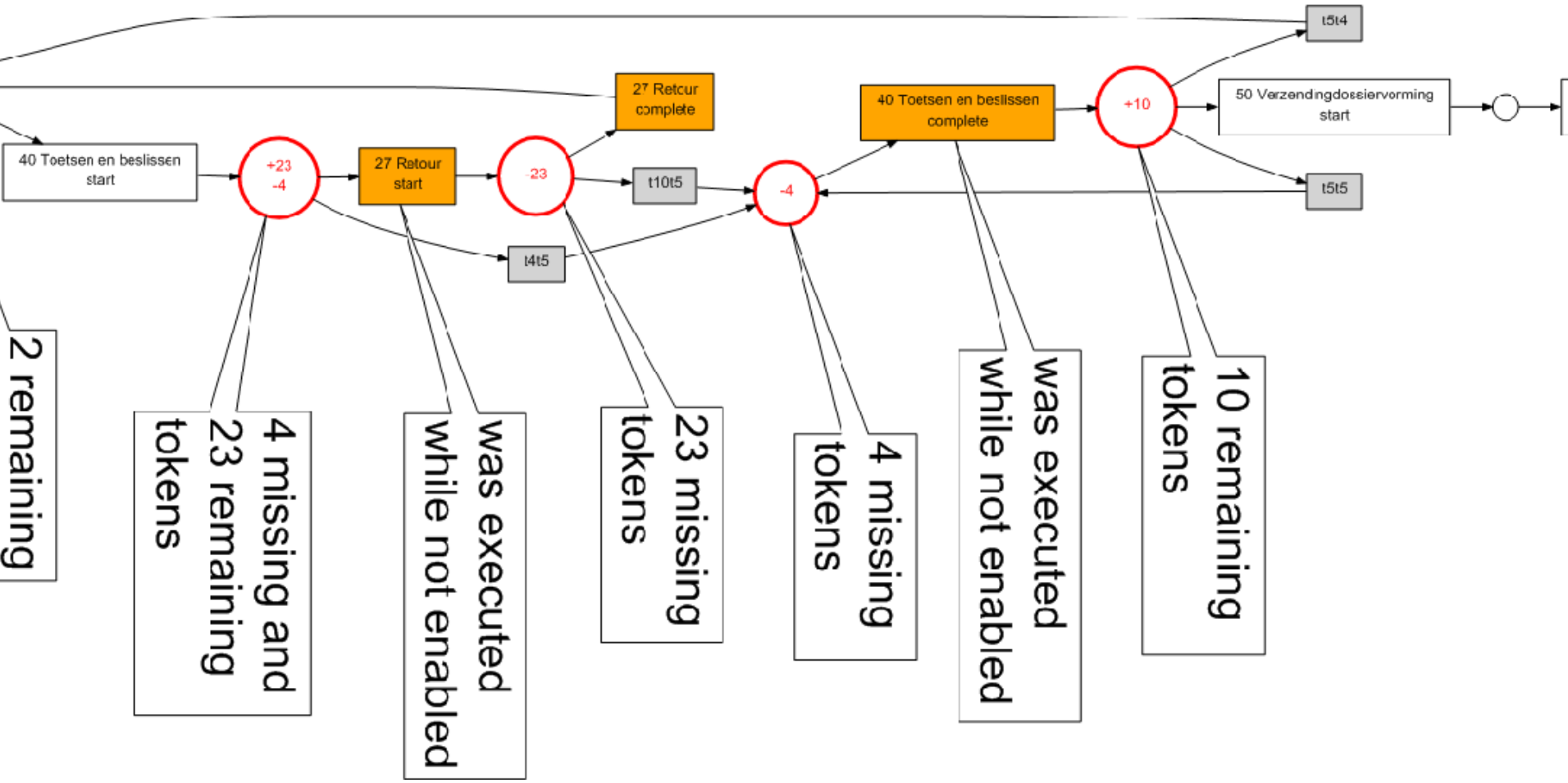




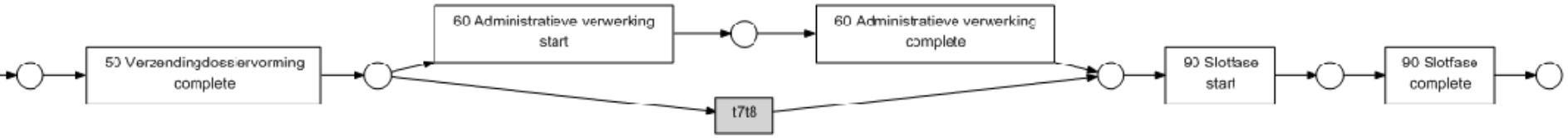
Conformance check WMO process (1/3)



Conformance check WMO process (2/3)

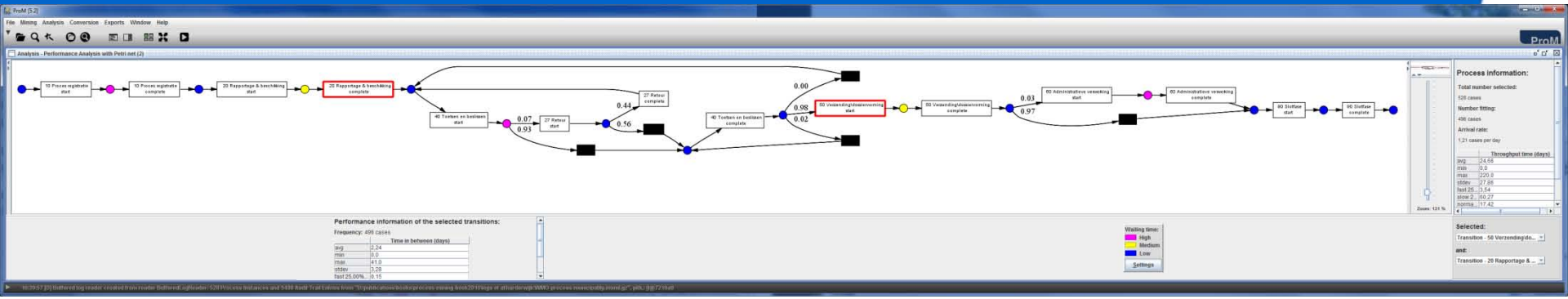


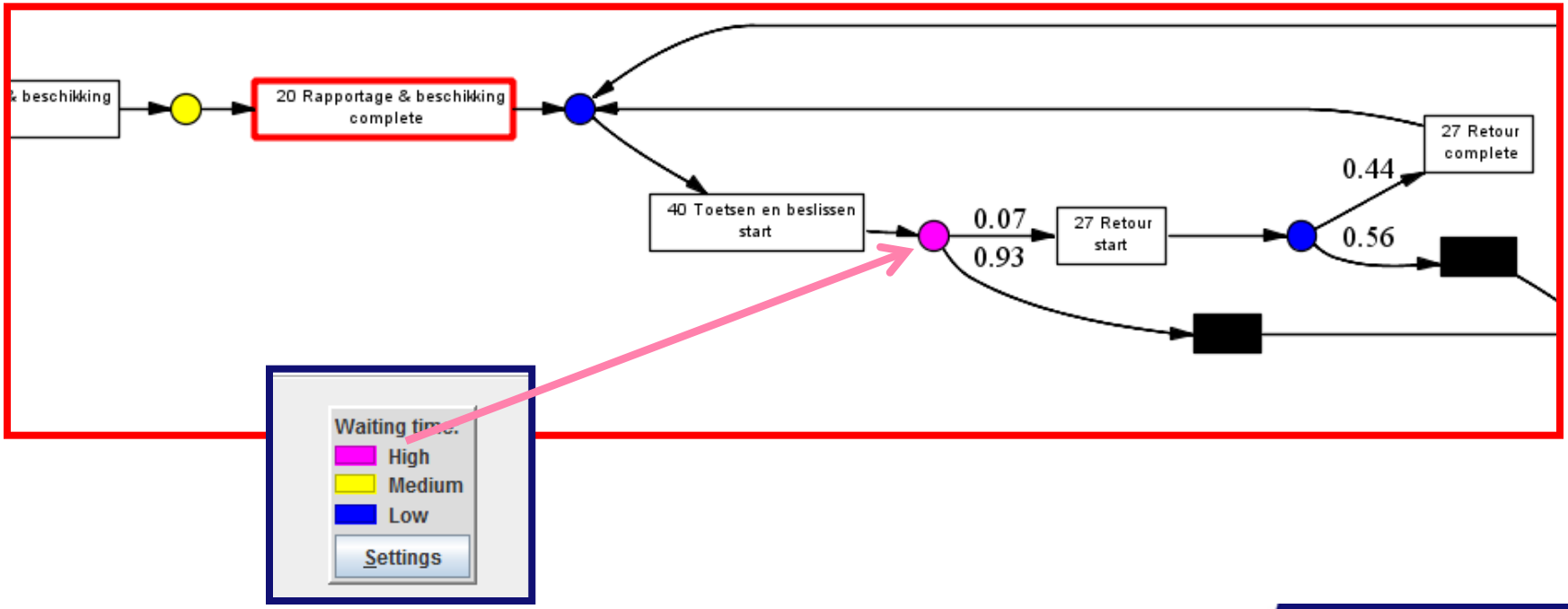
Conformance check WMO process (3/3)



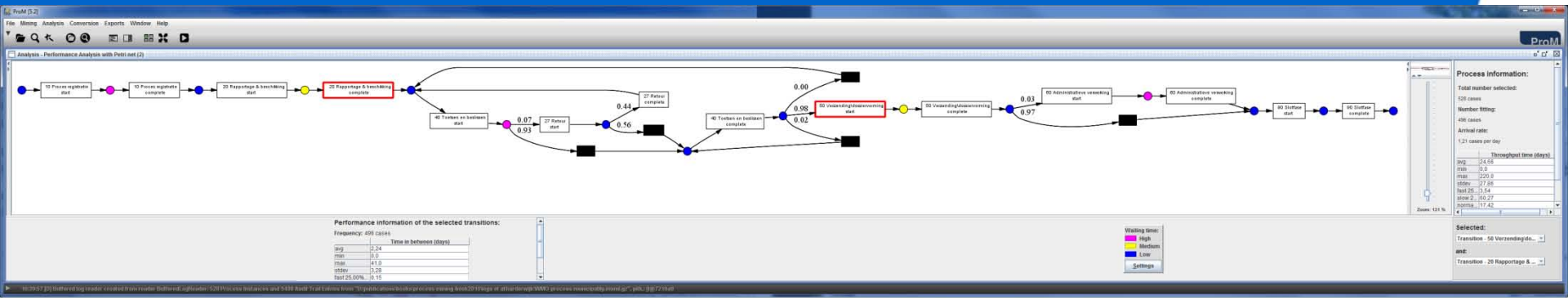
The fitness of the discovered process is 0.99521667. Of the 528 cases, 496 cases fit perfectly whereas for 32 cases there are missing or remaining tokens.

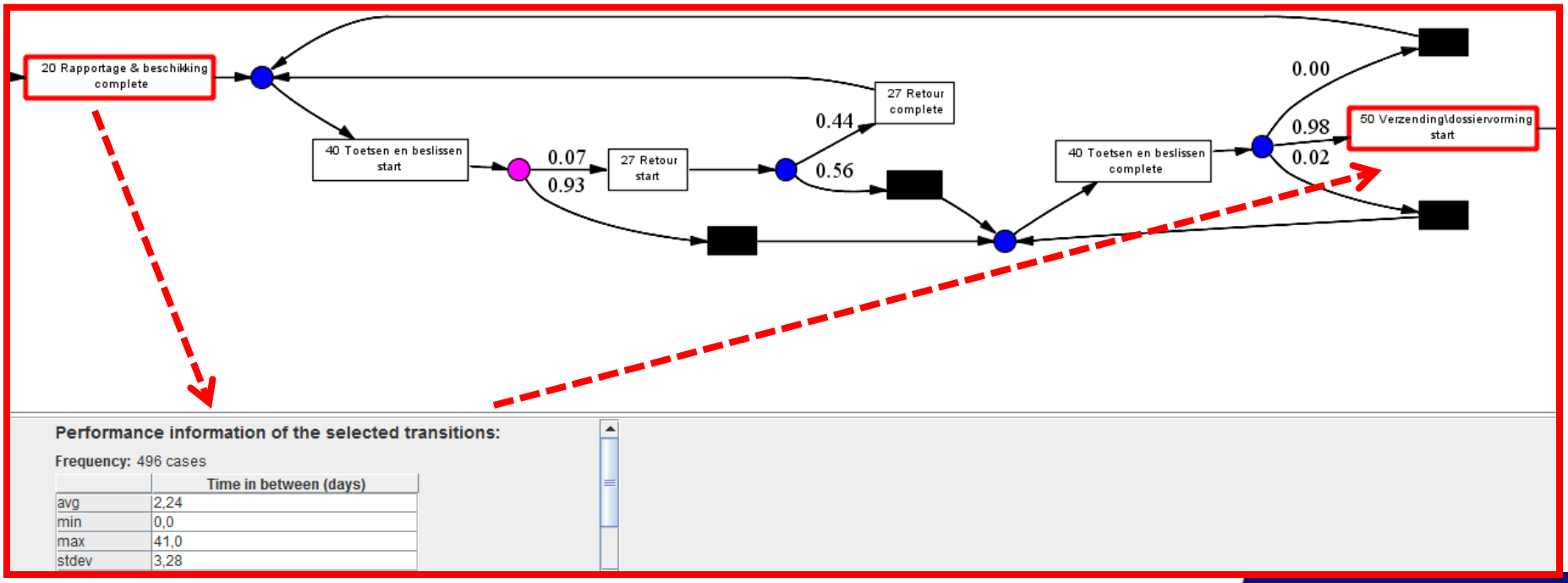
Bottleneck analysis WMO process (1/3)





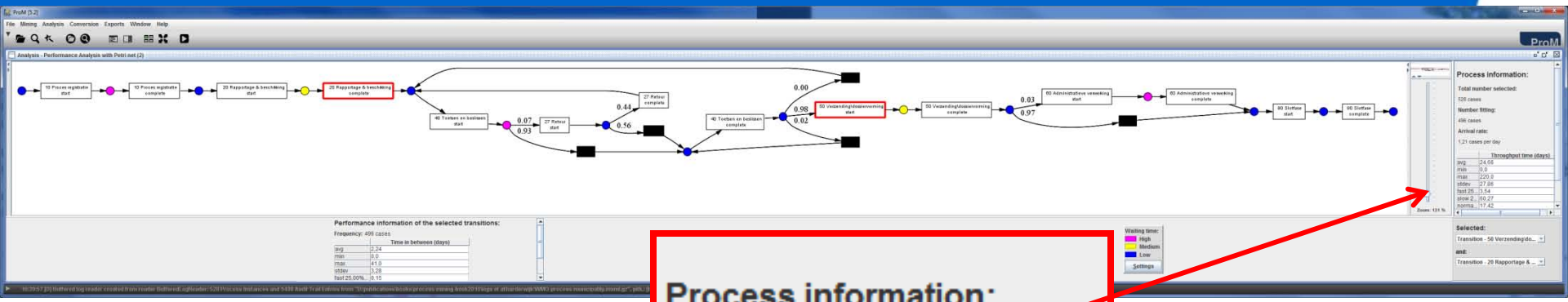
Bottleneck analysis WMO process (2/3)





	Time in between (days)
avg	2,24
min	0,0
max	41,0
stdev	3,28

Bottleneck analysis WMO process (3/3)



flow time of approx. 25 days with a standard deviation of approx. 28

Process information:

Total number selected:

528 cases

Number fitting:

496 cases

Arrival rate:

1,21 cases per day

	Throughput time (days)
avg	24,66
min	0,0
max	220,0
stdev	27,86
fast 25...	3,54
slow 2...	60,27
norma...	17,42

Two additional Lasagna processes

RWS ("Rijkswaterstaat") process

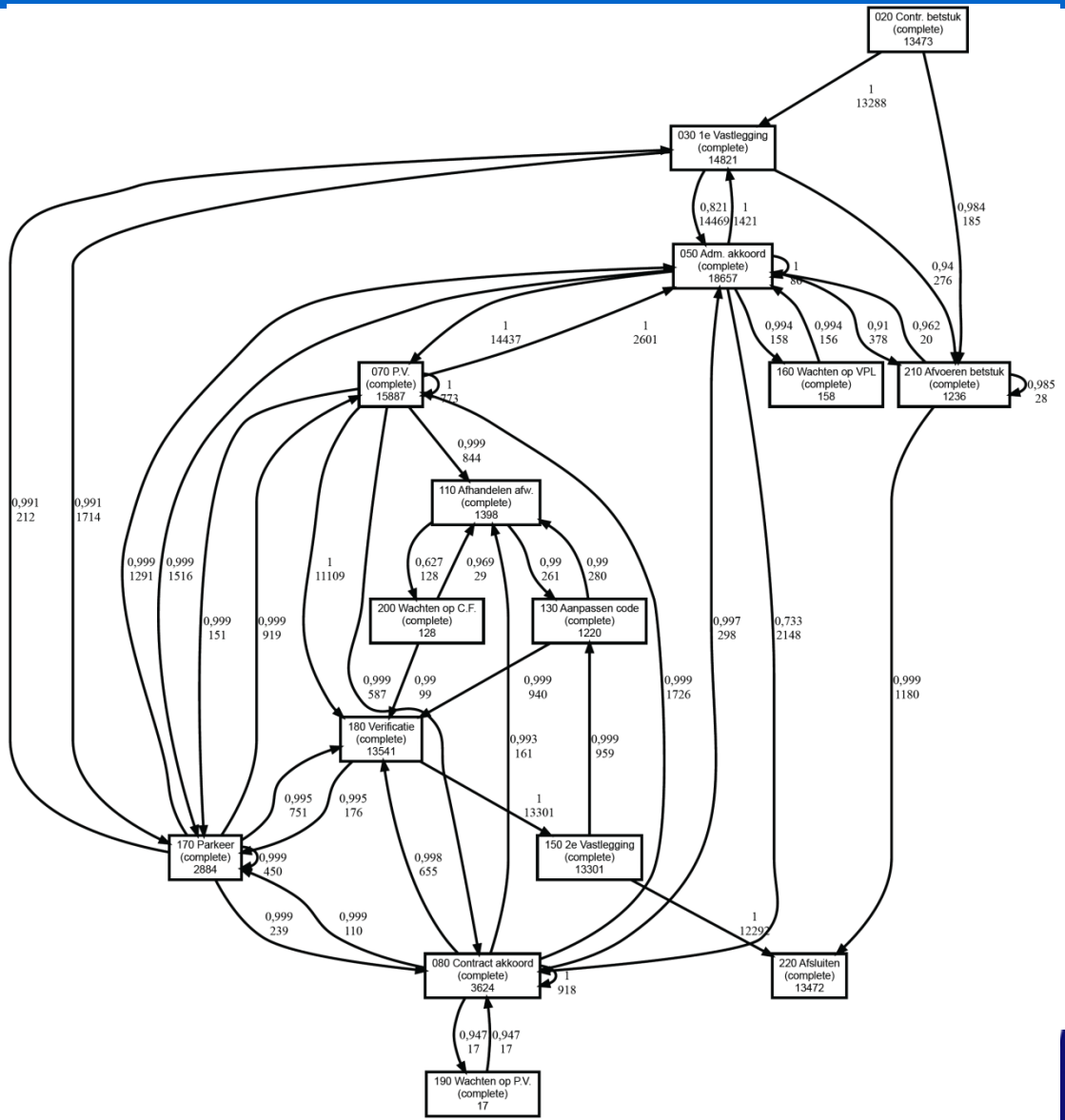
WOZ ("Waardering Onroerende Zaken") process

RWS Process

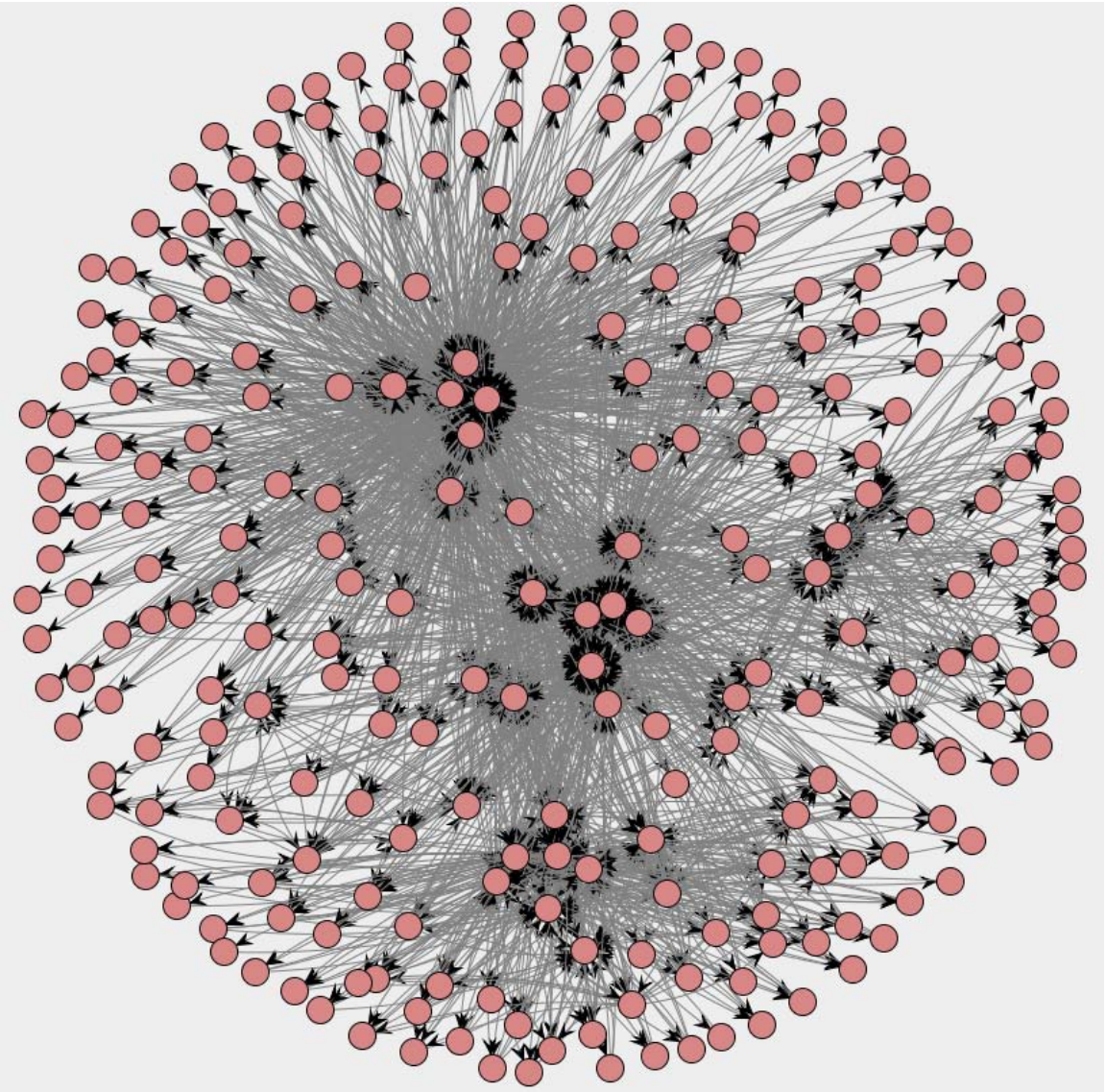
Rijkswaterstaat

- The Dutch national public works department, called “Rijkswaterstaat” (RWS), has twelve provincial offices. We analyzed the handling of invoices in one of these offices.
- The office employs about 1,000 civil servants and is primarily responsible for the construction and maintenance of the road and water infrastructure in its province.
- To perform its functions, the RWS office subcontracts various parties such as road construction companies, cleaning companies, and environmental bureaus. Also, it purchases services and products to support its construction, maintenance, and administrative activities.

C-net discovered using heuristic miner

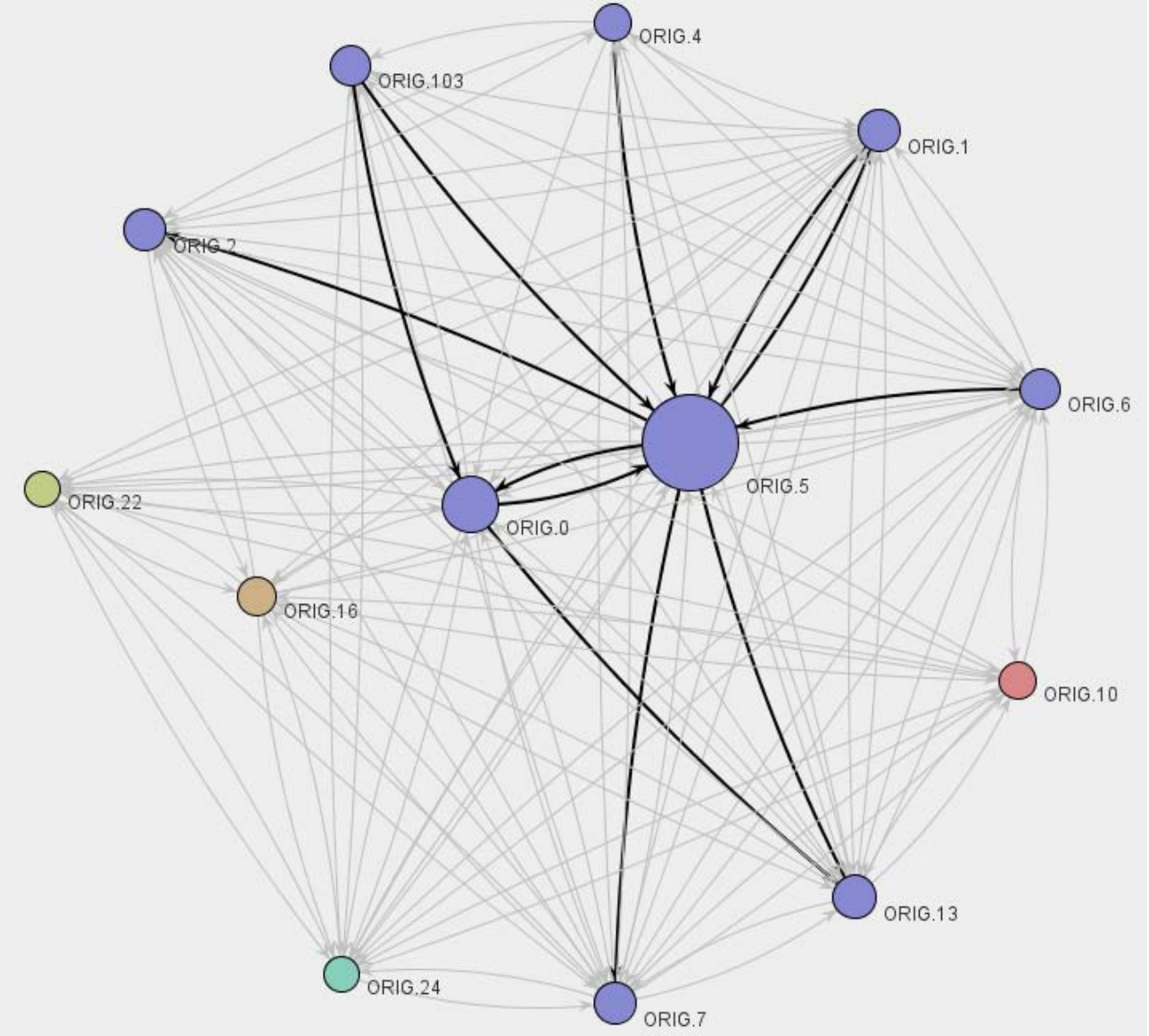


Social network constructed based on handovers of work



Each of the 271 nodes corresponds to a civil servant. Two civil servants are connected if one executed an activity causally following an activity executed by the other civil servant

Social network consisting of civil servants that executed more than 2000 activities in a 9 month period.

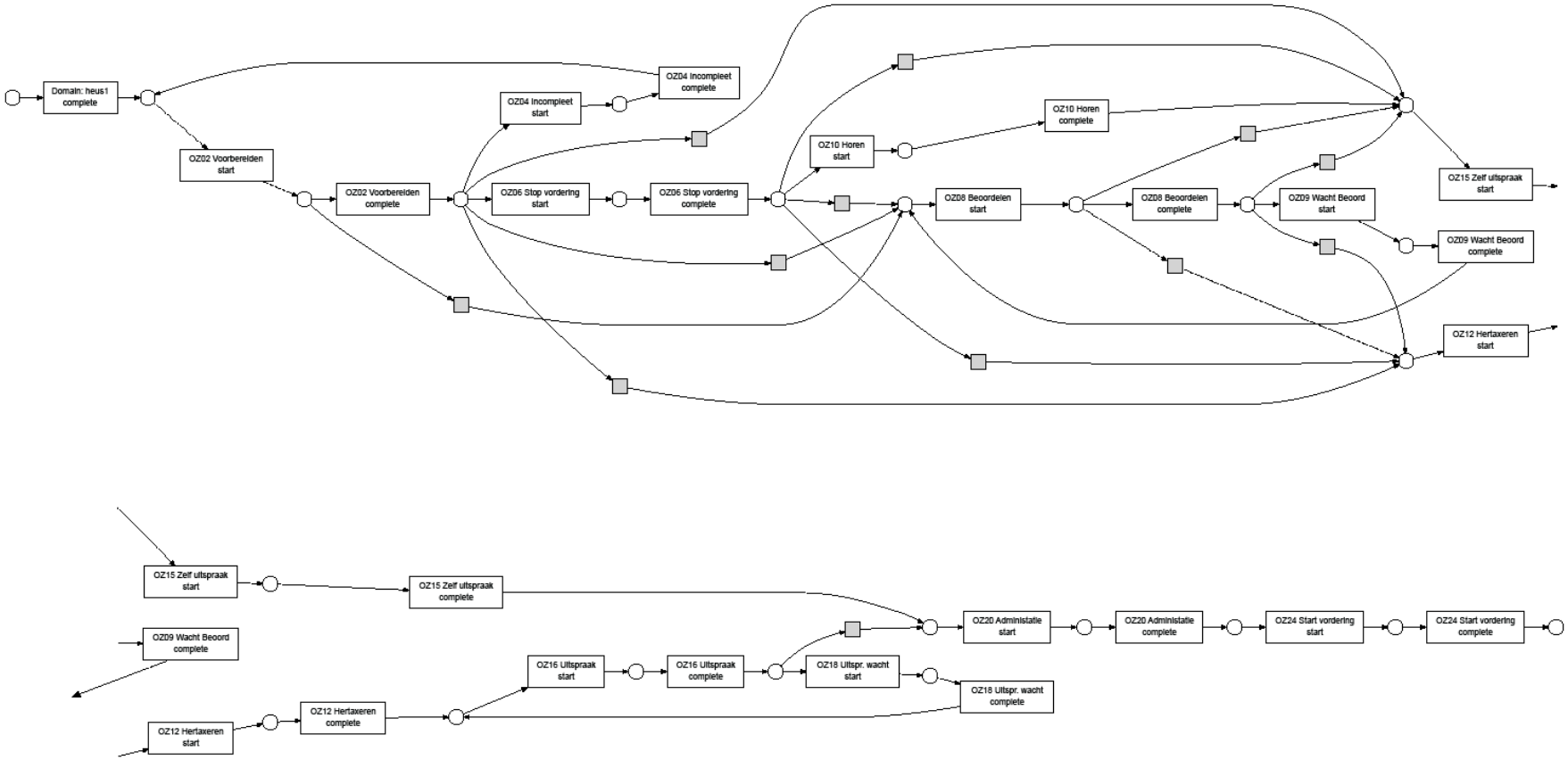


The darker arcs indicate the strongest relationships in the social network. Nodes having the same color belong to the same clique.

WOZ process

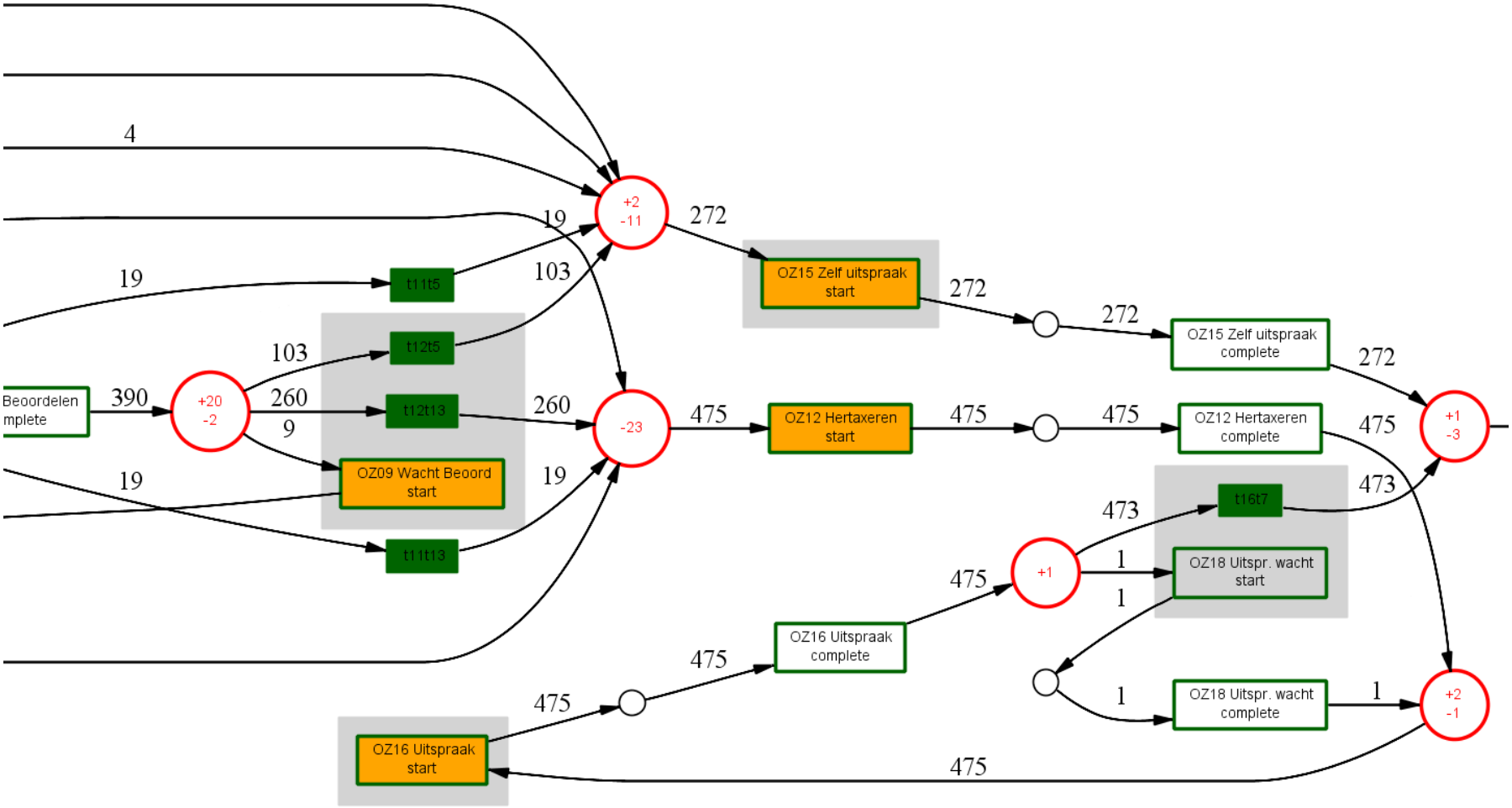
- Event log containing information about 745 objections against the so-called WOZ (“Waardering Onroerende Zaken”) valuation.
- Dutch municipalities need to estimate the value of houses and apartments. The WOZ value is used as a basis for determining the real-estate property tax.
- The higher the WOZ value, the more tax the owner needs to pay. Therefore, there are many objections (i.e., appeals) of citizens that assert that the WOZ value is too high.
- “WOZ process” discovered for another municipality (i.e., different from the one for which we analyzed the WMO process).

Discovered process model

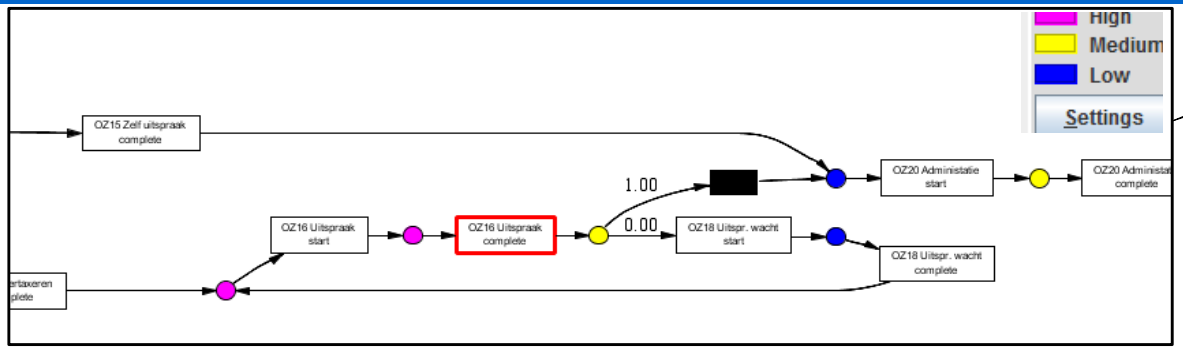


The log contains events related to 745 objections against the so-called WOZ valuation. These 745 objections generated 9583 events. There are 13 activities. For 12 of these activities both start and complete events are recorded. Hence, the WF-net has 25 transitions.

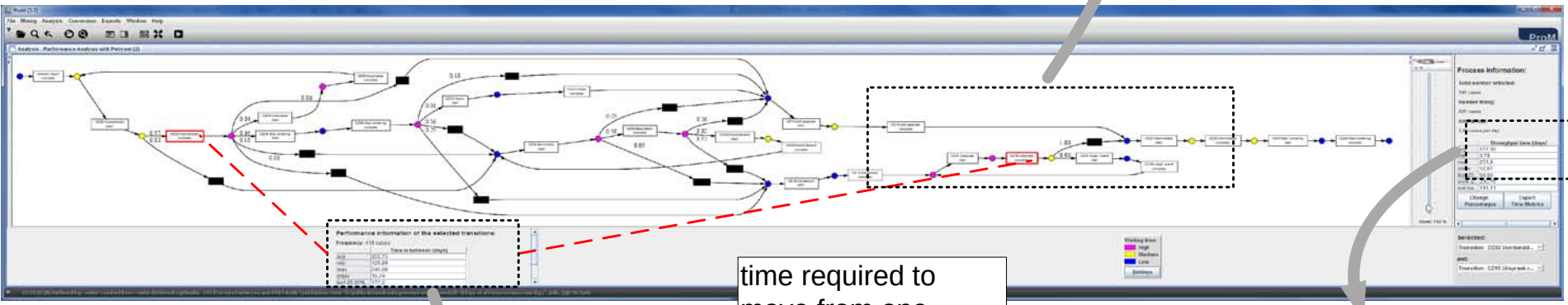
Conformance checker:
(fitness is 0.98876214)



Performance analysis



bottleneck detection: places are colored based on average durations



time required to move from one activity to another

Performance information of the selected transitions:

Frequency: 416 cases

	Time in between (days)
avg	202,73
min	126,89
max	245,98
stdev	19,74
fast 25,00%...	177,2

information on total flow time

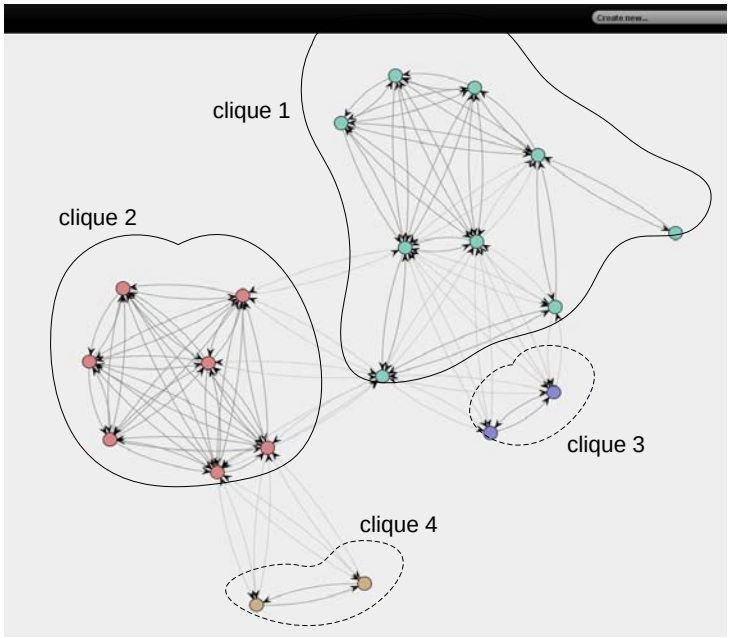
2,85 cases per day

	Throughput time (days)
avg	177,99
min	3,78
max	251,9
stdev	52,87
fast 25...	98,98
slow 2...	230,76
norma...	191,11

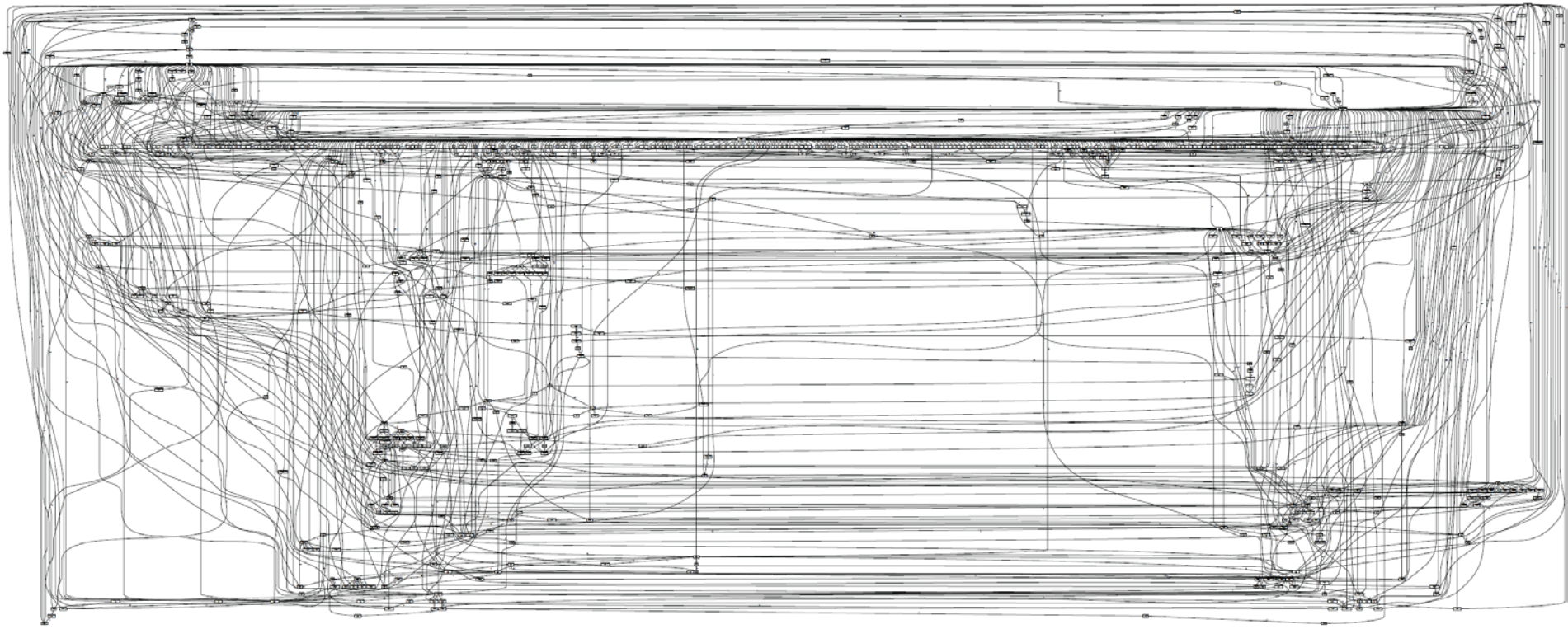
Change | Export

Resource-activity matrix (four groups discovered)

user	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
user 1	0	0	51	0	0	0	0	0	0	0	0	0	0
user 2	1	2	0	0	2	0	0	0	0	38	0	69	0
user 3	0	9	0	0	0	0	0	0	0	0	0	0	0
user 4	2	0	0	0	0	0	0	0	0	0	0	0	0
user 5	117	0	4	0	3	0	0	0	0	1	0	20	6
user 6	172	6	14	0	7	3	0	0	1	2	0	48	53
user 7	1	41	8	14	275	8	8	865	55	180	0	128	5
user 8	2	868	7	6	105	0	0	79	266	441	0	844	3
user 9	90	0	2	0	1	2	0	0	1	2	0	27	28
user 10	0	0	0	899	0	0	0	0	0	0	0	0	1019
user 11	336	1	3	1	4	2	0	0	0	1	0	18	23
user 12	1	645	13	21	419	3	0	3	217	281	1	334	9
user 13	0	1	0	0	0	0	0	0	0	0	0	0	0
user 14	0	0	0	0	0	0	0	0	0	1	0	0	0
user 15	0	0	0	0	0	0	0	2	2	0	0	2	0
user 16	1	3	3	2	1	0	0	1	2	3	1	0	0
user 17	0	4	0	0	0	0	0	0	0	0	0	0	0
user 18	9	0	0	0	0	0	0	0	0	0	0	0	0
user 19	13	1	0	0	1	0	0	0	0	0	0	4	0
user 20	0	0	0	21	0	0	0	0	0	0	0	0	258



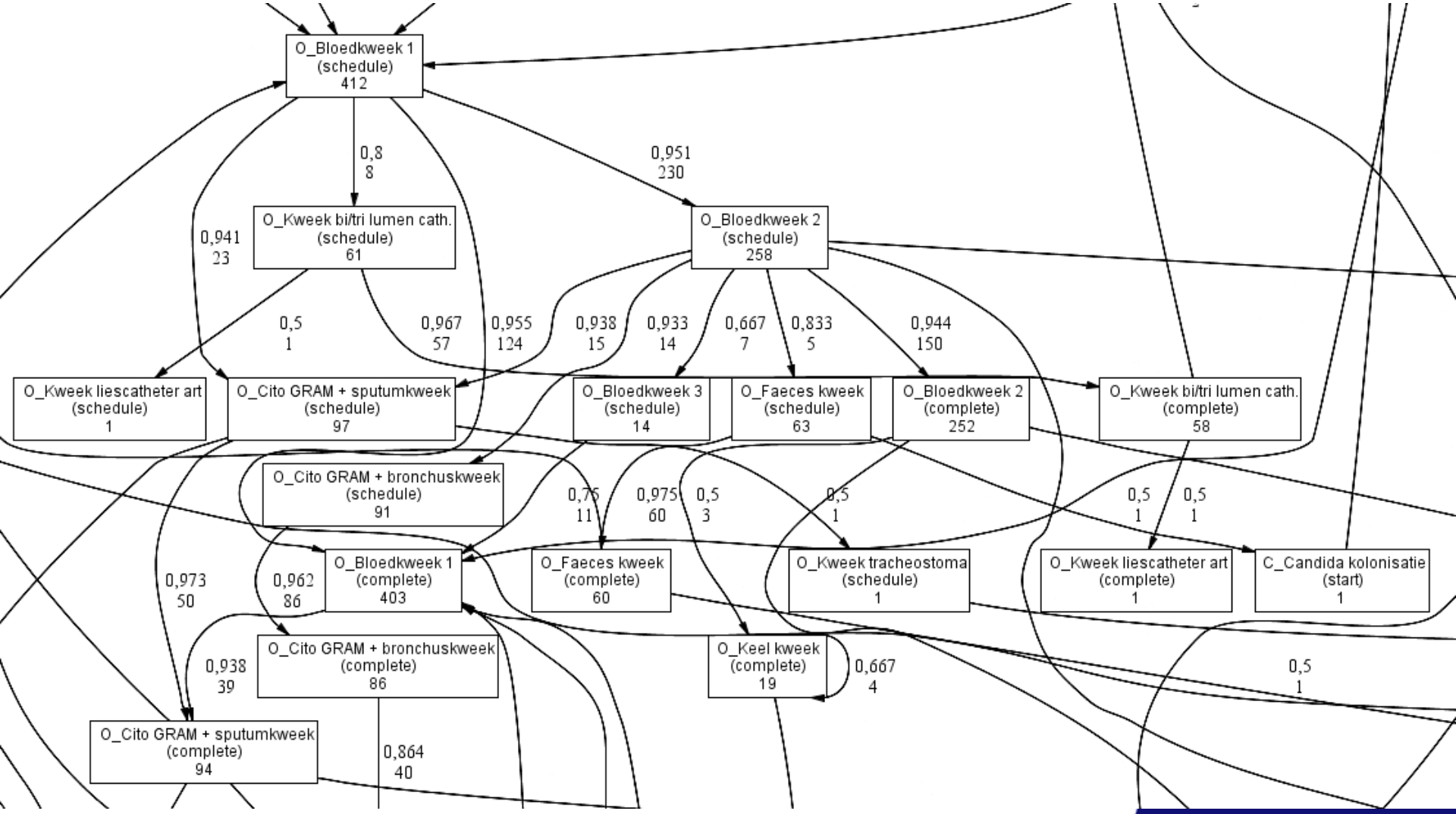
Example of a Spaghetti process



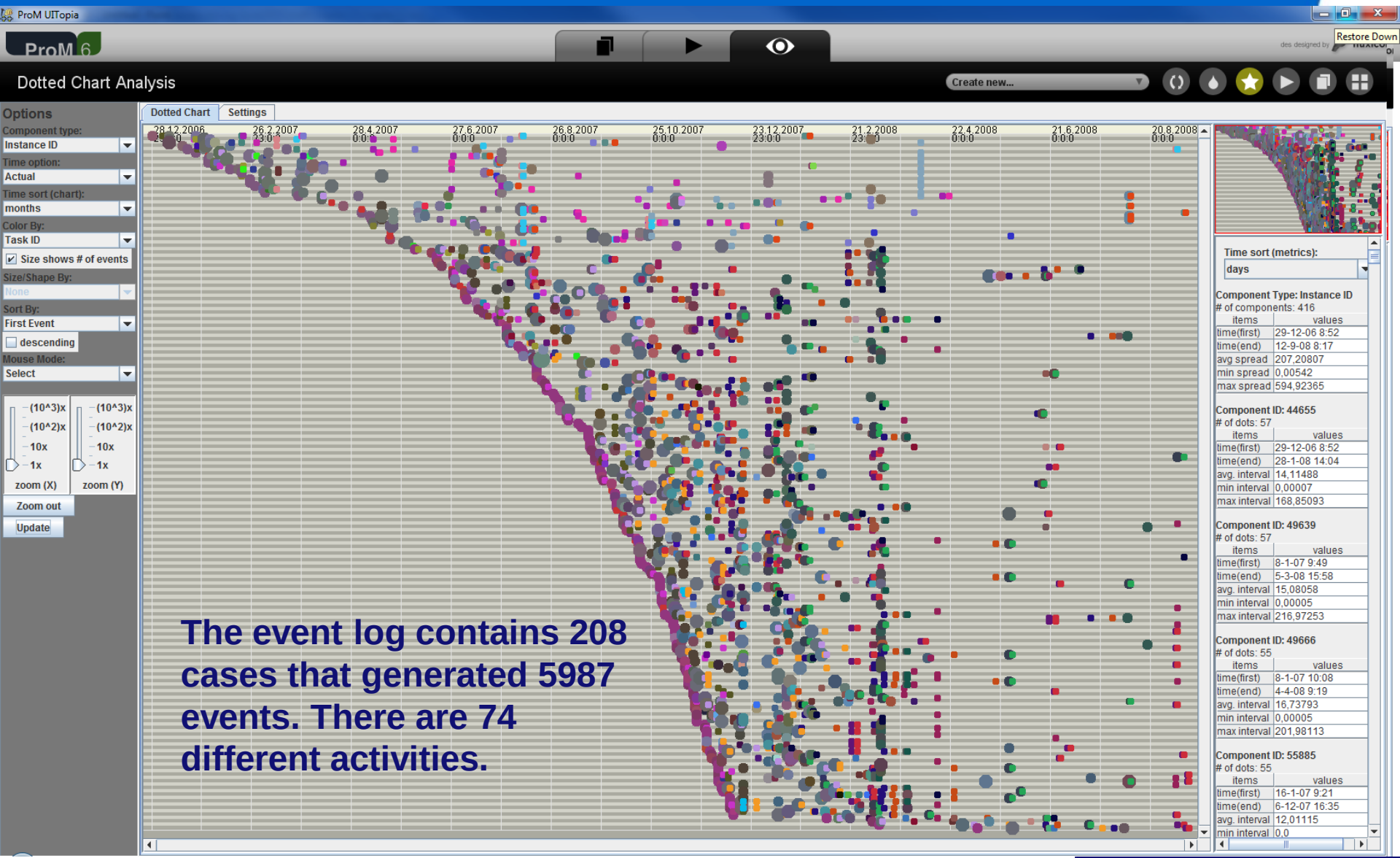
Spaghetti process describing the diagnosis and treatment of 2765 patients in a Dutch hospital. The process model was constructed based on an event log containing 114,592 events. There are 619 different activities (taking event types into account) executed by 266 different individuals (doctors, nurses, etc.).

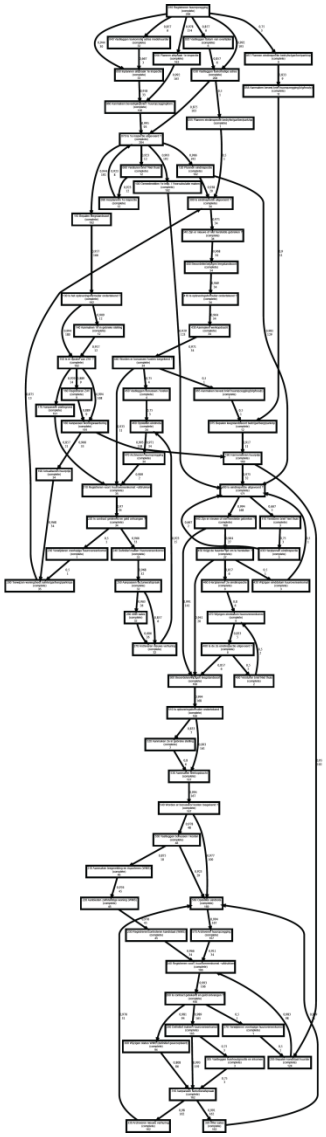
Fragment

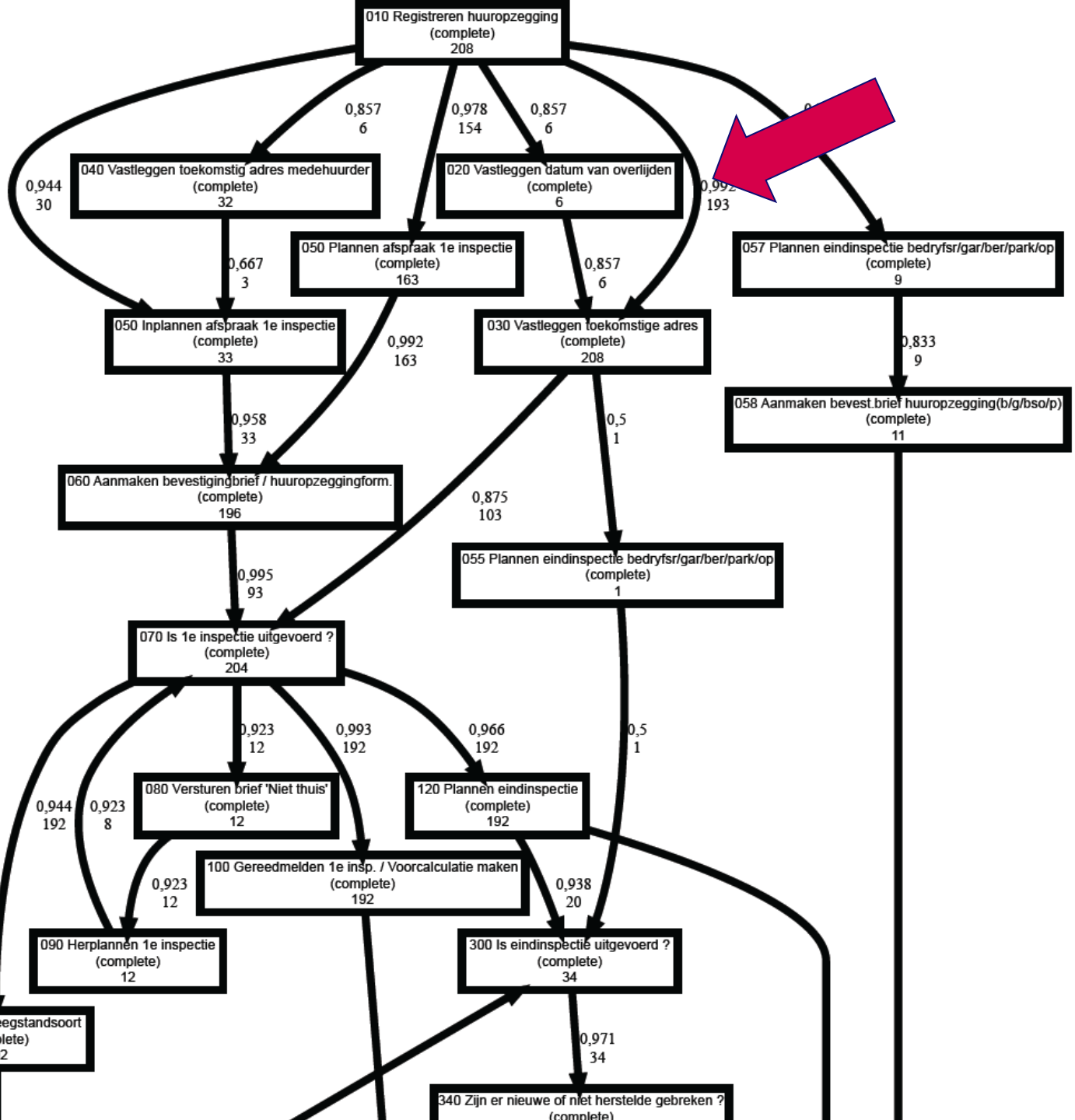
18 activities of the 619 activities (2.9%)



Another example
(event log of Dutch housing agency)






010 Registreren huuropzegging
(complete)
208
0,857
6
0,978
154
0,857
6
0,944
30
040 Vastleggen toekomstig adres medehuurder
(complete)
32
020 Vastleggen datum van overlijden
(complete)
6
0,992
193
057 Plannen eindinspectie bedryfsr/gar/ber/park/op
(complete)
9
0,667
3
050 Plannen afspraak 1e inspectie
(complete)
163
0,857
6
050 Inplannen afspraak 1e inspectie
(complete)
33
0,992
163
030 Vastleggen toekomstige adres
(complete)
208
0,833
9
058 Aanmaken bevest.brief huuropzegging(b/g/bso/p)
(complete)
11
0,958
33
0,5
1
060 Aanmaken bevestigingbrief / huuropzeggingform.
(complete)
196
0,875
103
055 Plannen eindinspectie bedryfsr/gar/ber/park/op
(complete)
1
0,995
93
070 Is 1e inspectie uitgevoerd ?
(complete)
204
0,923
12
0,993
192
0,966
192
0,5
1
0,944
192
0,923
8
080 Versturen brief 'Niet thuis'
(complete)
12
120 Plannen eindinspectie
(complete)
192
0,923
12
100 Gereedmelden 1e insp. / Voorcalculatie maken
(complete)
192
0,938
20
090 Herplannen 1e inspectie
(complete)
12
300 Is eindinspectie uitgevoerd ?
(complete)
34
110 Bepalen leegstandsoort
(complete)
192
0,971
34
340 Zijn er nieuwe of niet herstelde gebreken ?
(complete)

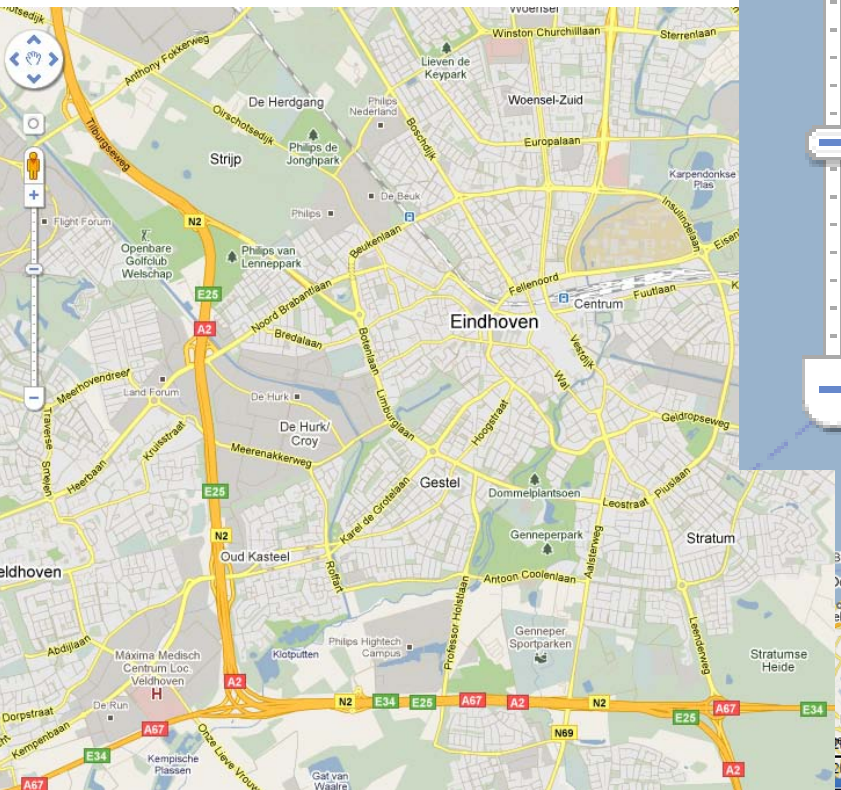
processes
maps
navigation

Example of a map

Road map of The Netherlands. The map abstracts from smaller cities and less significant roads; only the bigger cities, highways, and other important roads are shown. Moreover, cities aggregate local roads and local districts. Also note the use of color, size, etc. to emphasize important things.

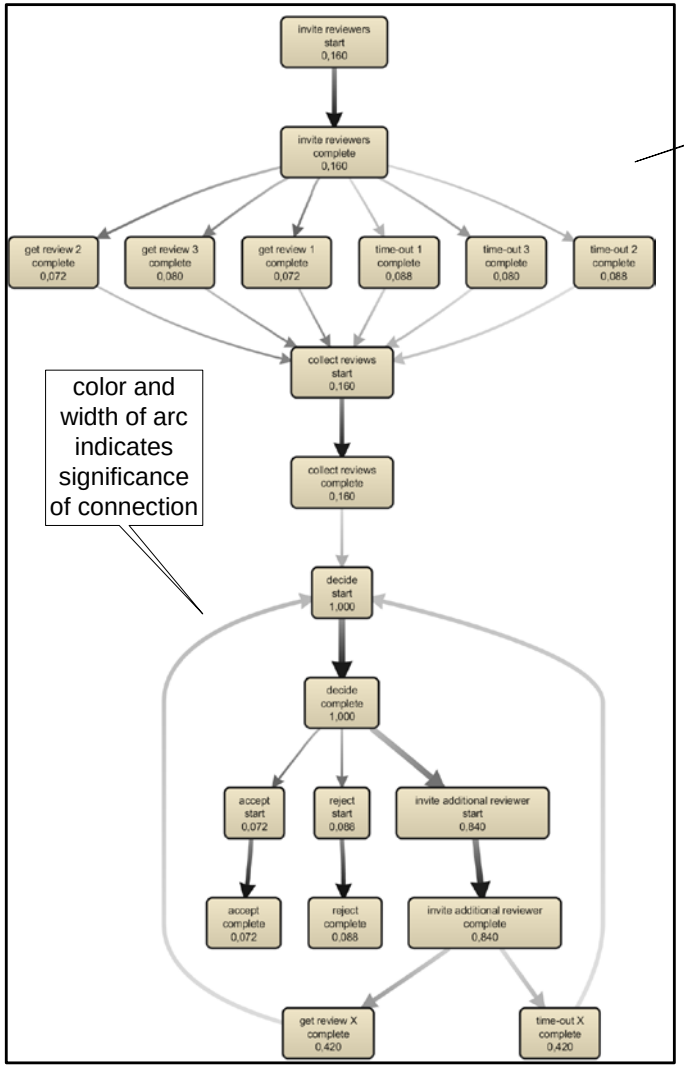
Example: Slider in Google Maps

Google maps Nederland



Amsterdam
Rotterdam
Den Haag
Utrecht
Eindhoven
Groningen
Leeuwarden
Zwolle
Arnhem
Nijmegen
Breda
Tilburg
's-Hertogenbosch
Antwerpen
Brugge
Gent
Bruxelles
Brussel
Leuven
Maastricht
Hasselt
Aachen
Haarlem
Zaanstad
Leiden
Delft
Dordrecht
Amersfoort
Apeldoorn
Enschede
Deventer
Nijmegen
Eindhoven
Venlo
Roermond
Sittard-Geleen
Heerlen
Genk
Mechelen
Sint-Niklaas
Middelburg
Vlissingen
Terneuzen
Zeebrugge
Oostende
Knokke-Heist
Duisburg
Essen
Krefeld
Düsseldorf
Mönchengladbach
Wuppertal
Solingen
Leverkusen
Köln
Bergisch Gladbach
Bonn
Aachen
Maastricht
Hasselt
Leuven
Bruxelles
Brussel
Antwerpen
Gent
Brugge
Zeebrugge
Oostende
Middelburg
Vlissingen
Terneuzen
Roosendaal
Bergen op Zoom
Breda
Tilburg
's-Hertogenbosch
Oosterhout
Dordrecht
Rotterdam
Den Haag
Gouda
Utrecht
Amersfoort
Apeldoorn
Deventer
Enschede
Zwolle
Arnhem

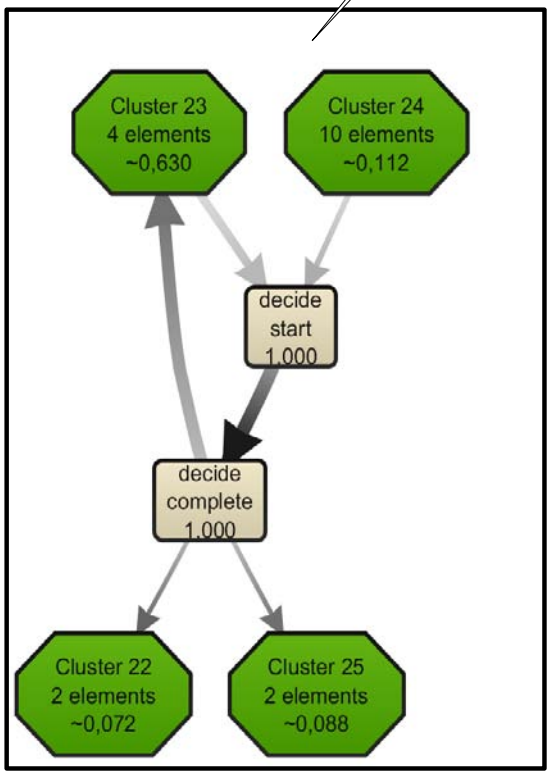
Fuzzy miner: one process - two views
(100 cases generating 3730 events)



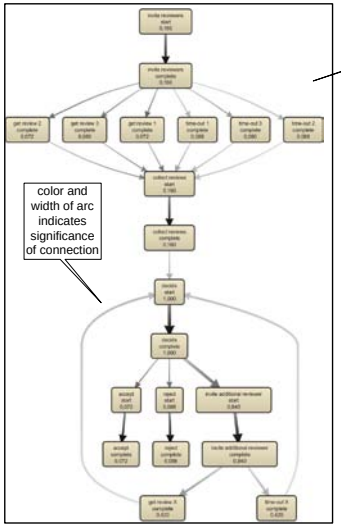
fuzzy model showing all activities

color and width of arc indicates significance of connection

fuzzy model showing only two activities

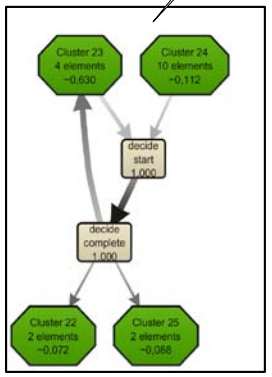


Balancing between both extremes



fuzzy model showing all activities

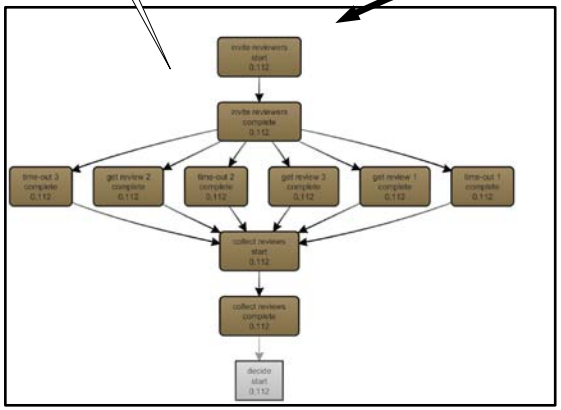
color and width of arc indicates significance of connection

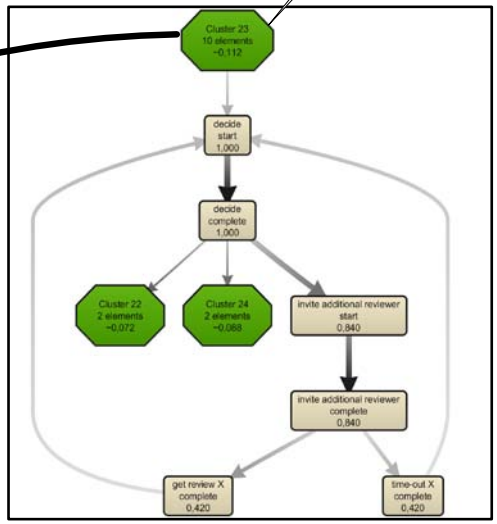


fuzzy model showing only two activities

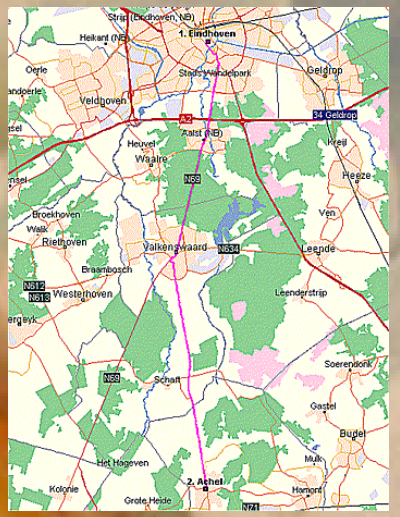
aggregated node containing 10 activities

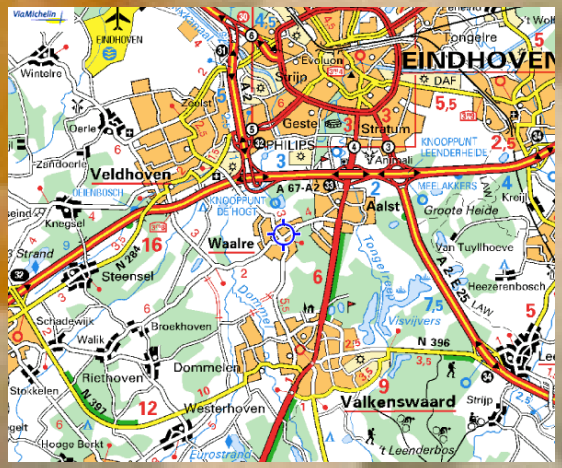
inner structure of aggregated node



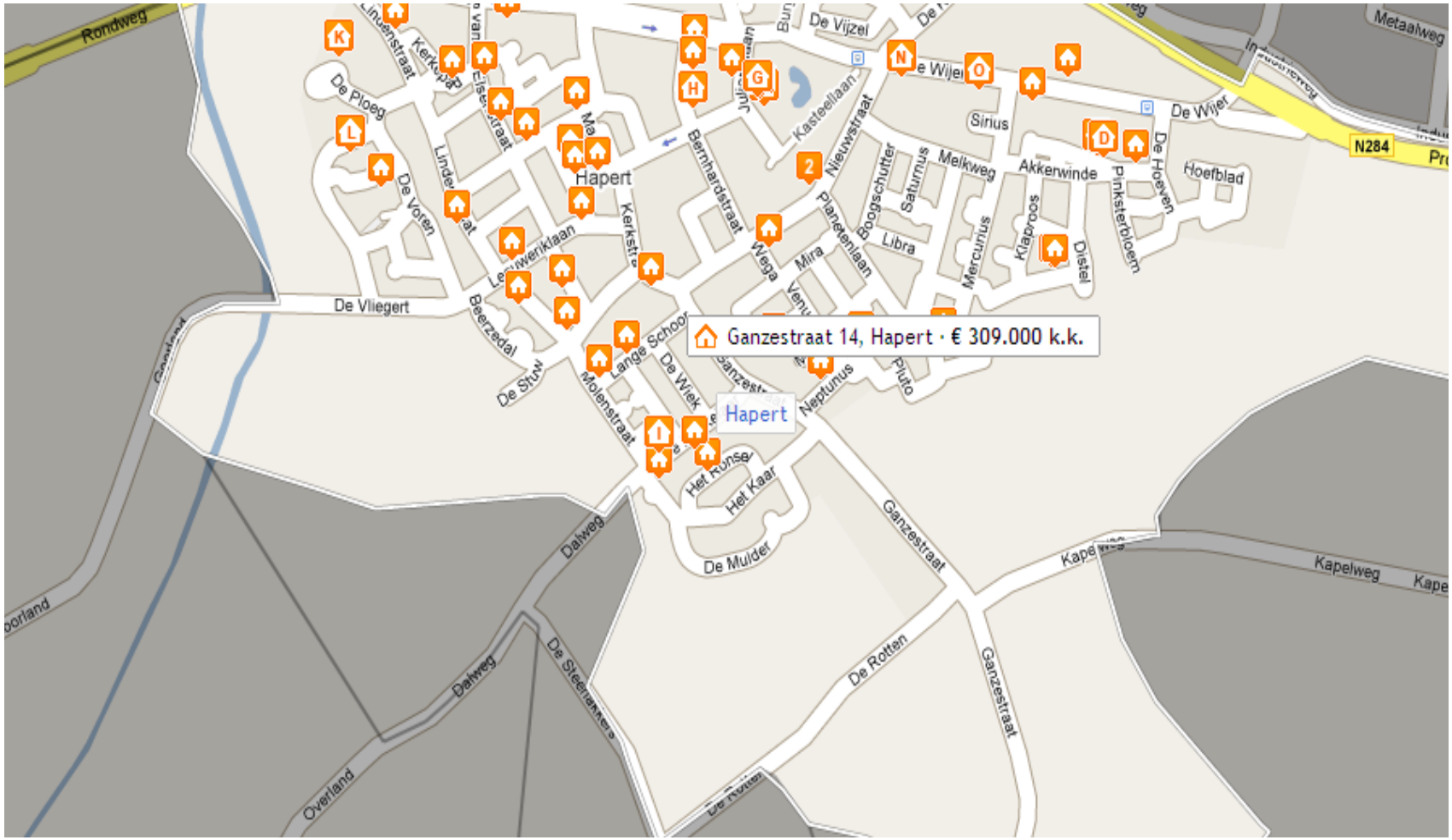


Not a single map!

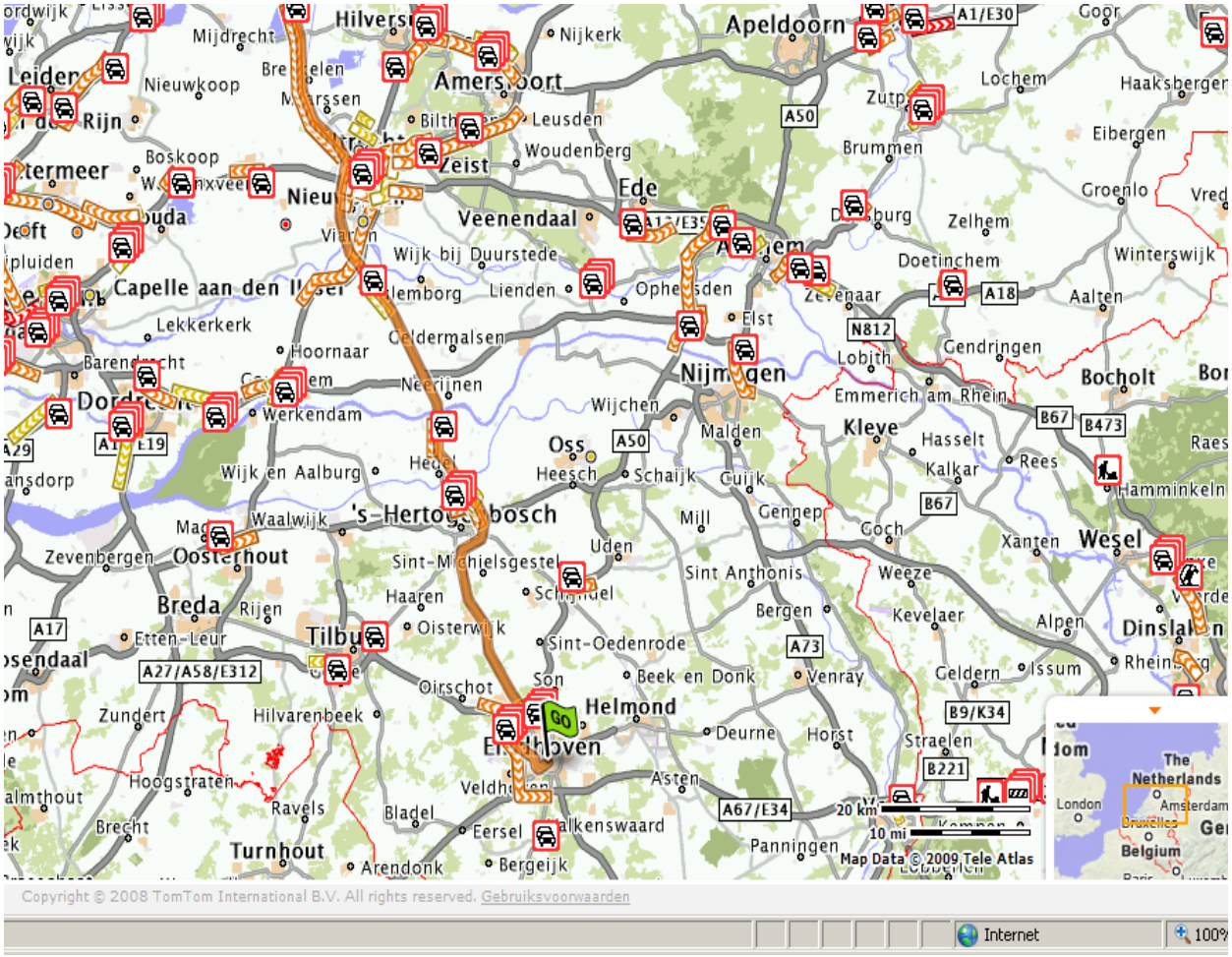




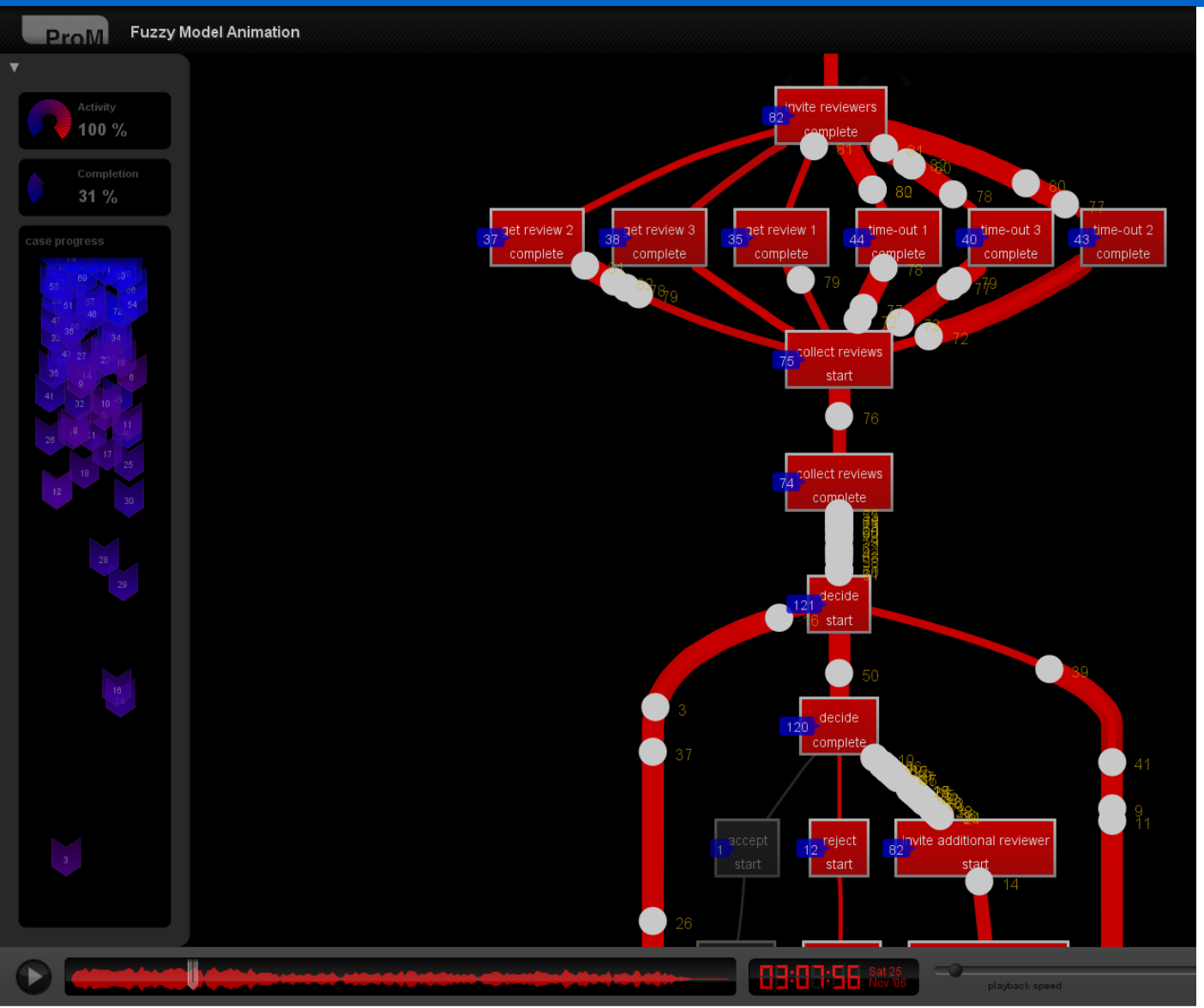
Projecting dynamic information on business process maps



Projecting traffic jams on maps



Business process movies



Navigation

- Whereas a TomTom device is continuously showing the expected arrival time, users of today's information systems are often left clueless about likely outcomes of the cases they are working on.
- Car navigation systems provide directions and guidance without controlling the driver. The driver is still in control, but, given a goal (e.g. to get from A to B as fast as possible), the navigation system recommends the next action to be taken.
- Operational support provides TomTom functionality for business processes.

Recommend: How to get home ASAP? Take a left turn!

Detect: You drive too fast!

Predict: When will I be home? At 11.26!

Conclusion: two types of processes

PrOM
process mining workbench

Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

More and more information about business processes is recorded by information systems in the form of so-called "event logs". Despite the omnipresence of such data, most organizations diagnose problems based on fiction rather than facts. Process mining is an emerging discipline based on process model-driven approaches and data mining. It not only allows organizations to fully benefit from the information stored in their systems, but it can also be used to check the conformance of processes, detect bottlenecks, and predict execution problems.

Wil van der Aalst delivers the first book on process mining. It aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. In Part I, the author provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Part II focuses on process discovery as the most important process mining task. Part III moves beyond discovering the control flow of processes and highlights conformance checking, and organizational and time perspectives. Part IV guides the reader in successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM. Finally, Part V takes a step back, reflecting on the material presented and the key open challenges.

Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

Features and Benefits:

- First book on process mining, bridging the gap between business process modeling and business intelligence.
- Written by one of the most influential and most-cited computer scientists and the best-known BPM researcher.
- Self-contained and comprehensive overview for a broad audience in academia and industry.
- The reader can put process mining into practice immediately due to the applicability of the techniques and the availability of the open-source process mining software ProM.

Computer Science

ISBN 978-3-642-19344-6

springer.com

van der Aalst Process Mining

www.processmining.org

www.win.tue.nl/ieeetfpm/

Springer