

Process mining in Collaborative Networks: Discovering Real Processes and Interactions

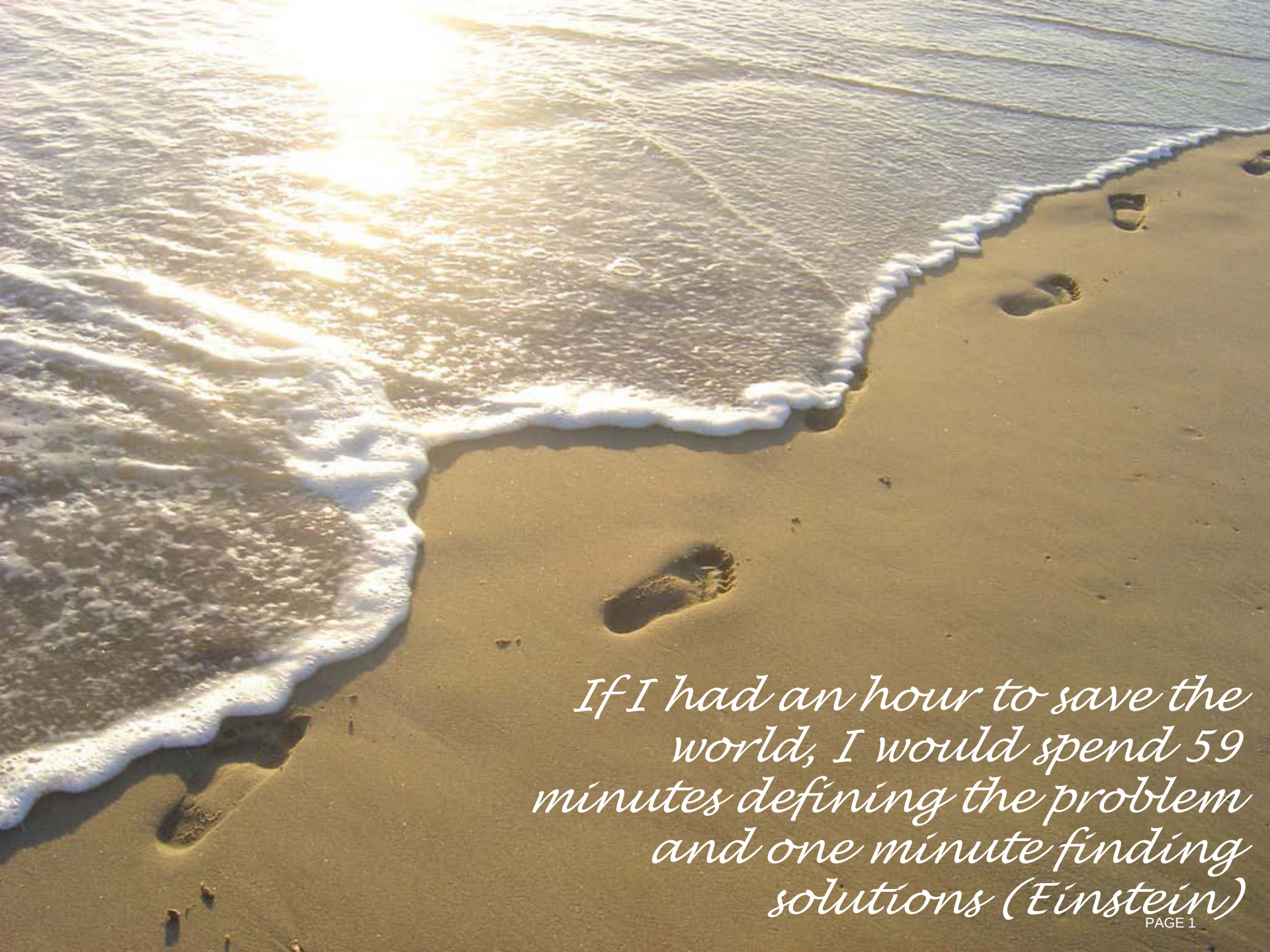
prof.dr.ir. Wil van der Aalst
www.processmining.org



TU **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

A photograph of a beach at sunset. The sun is low on the horizon, creating a bright, golden glow over the water and sand. Waves are breaking onto the shore, with white foam visible. Several footprints are visible in the wet sand, leading from the water towards the viewer. The overall mood is peaceful and contemplative.

*If I had an hour to save the
world, I would spend 59
minutes defining the problem
and one minute finding
solutions (Einstein)*

PROCESS MINING

virtual extended enterprises

(DYNAMIC) VIRTUAL ORGANIZATIONS

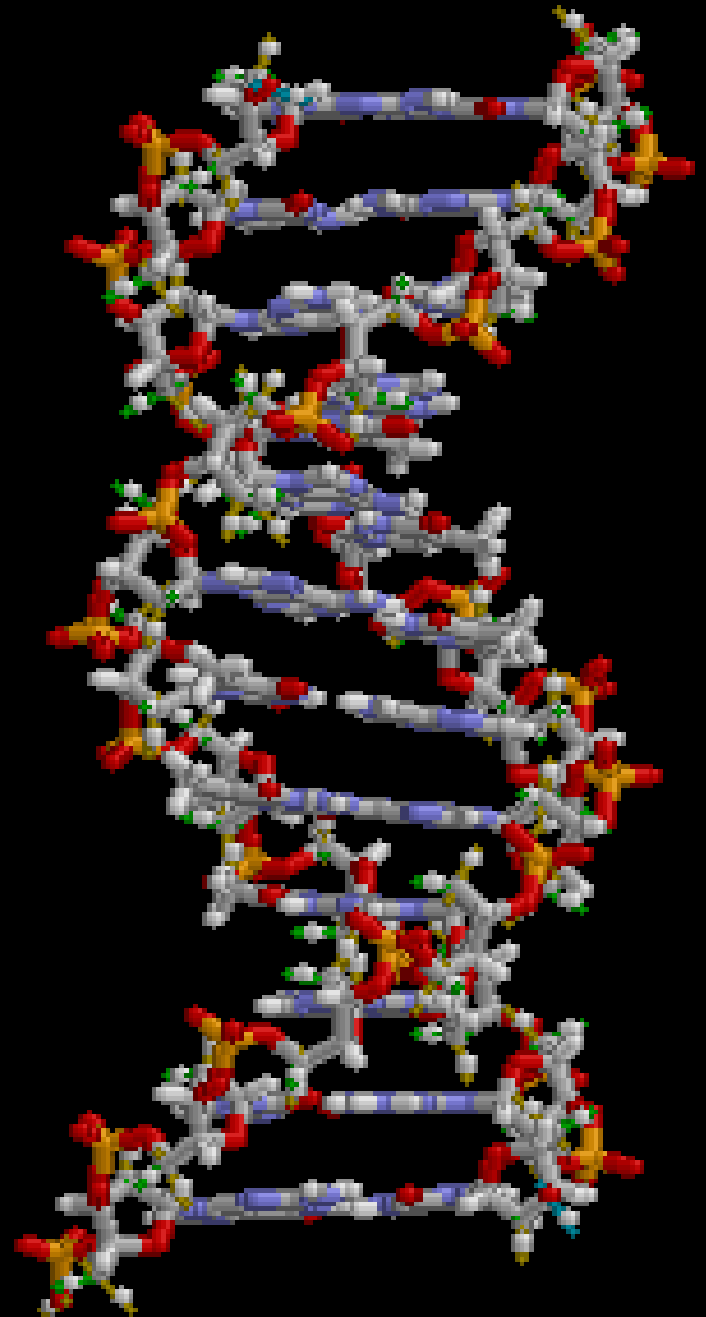
(professional) virtual communities

service oriented computing

internet of things/future internet

Process Mining: Why?

- physicists study matter and its motion through spacetime,
- chemists study the composition, structure, and properties of matter,
- biologists study evolution, organisms, cells, and genes,
- management science?
- computer science?
- social science?



Process Mining



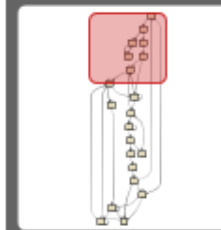
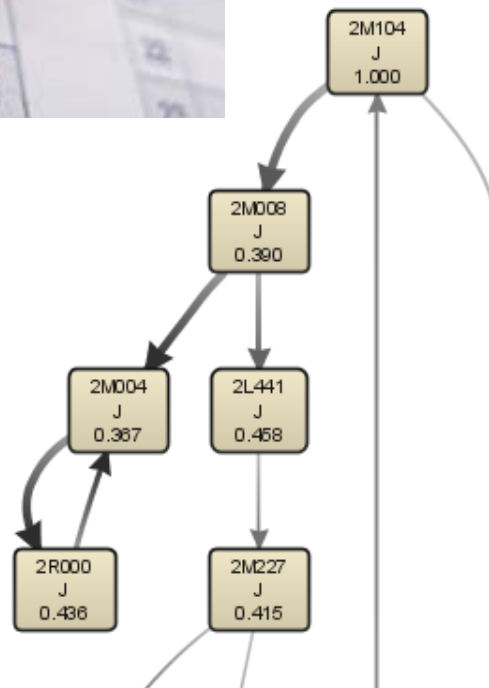
- **Process discovery:** "What is really happening?"
- **Conformance checking:** "Do we do what was agreed upon?"
- **Performance analysis:** "Where are the bottlenecks?"
- **Process prediction:** "Will this case be late?"
- **Process improvement:** "How to redesign this process?"
- **Etc.**



ter) (2)

Animation

ProM



Zoom

► Concurrency filter

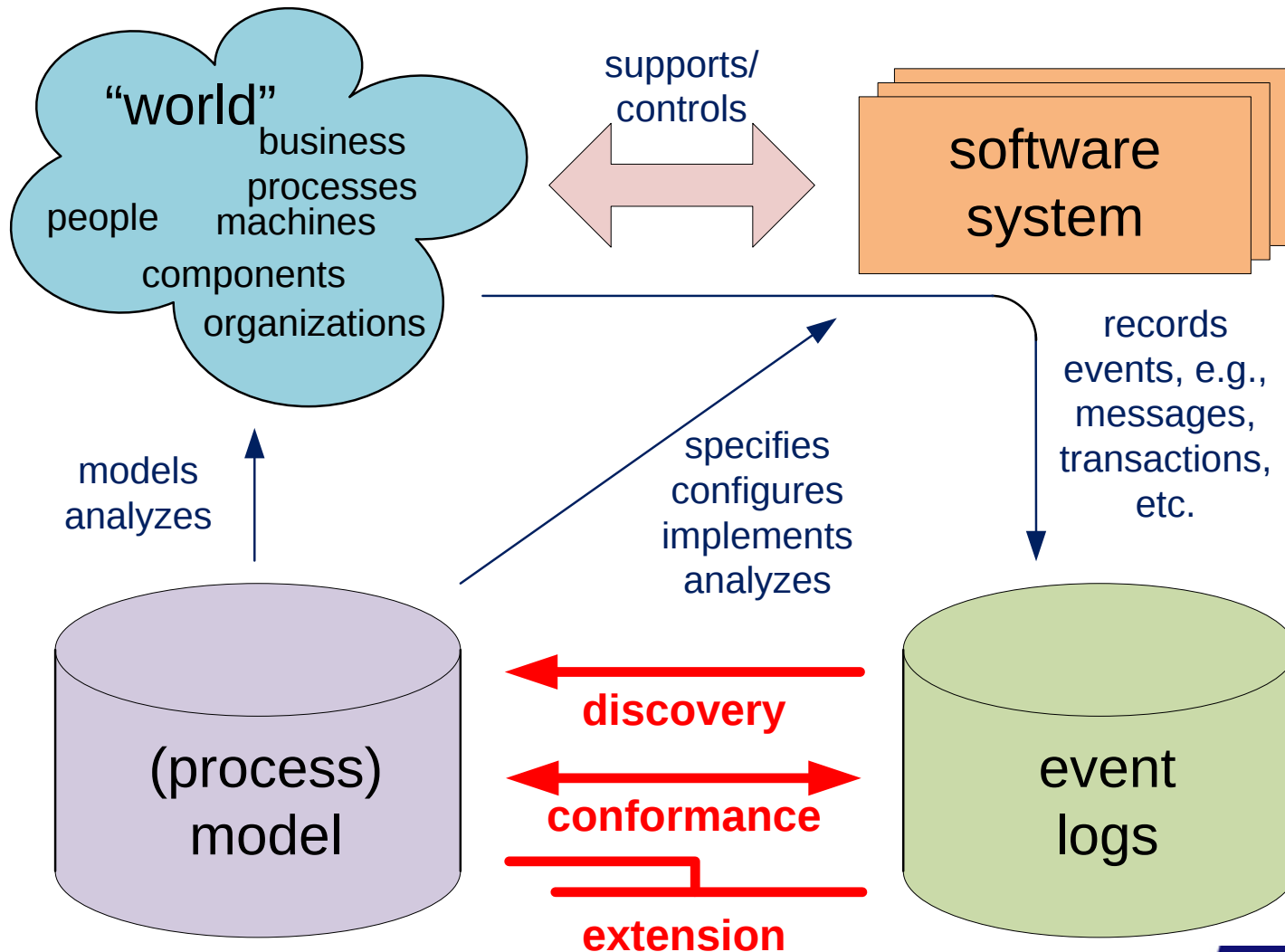
► Edge filter

▼ Node filter

Significance cutoff

- Process discovery: "What is the real curriculum?"
- Conformance checking: "Do students meet the prerequisites?"
- Performance analysis: "Where are the bottlenecks?"
- Process prediction: "Will a student complete his studies (in time)?"
- Process improvement: "How to redesign the curriculum?"

Process mining: Linking events to models



Process mining as a mirror ...



Where did we apply process mining?

- **Municipalities** (e.g., Alkmaar, Heusden, 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, Thales)
- **Media companies** (e.g. Winkwaves)
- ...

Wet Maatschappelijke Ondersteuning (WMO) Harderwijk

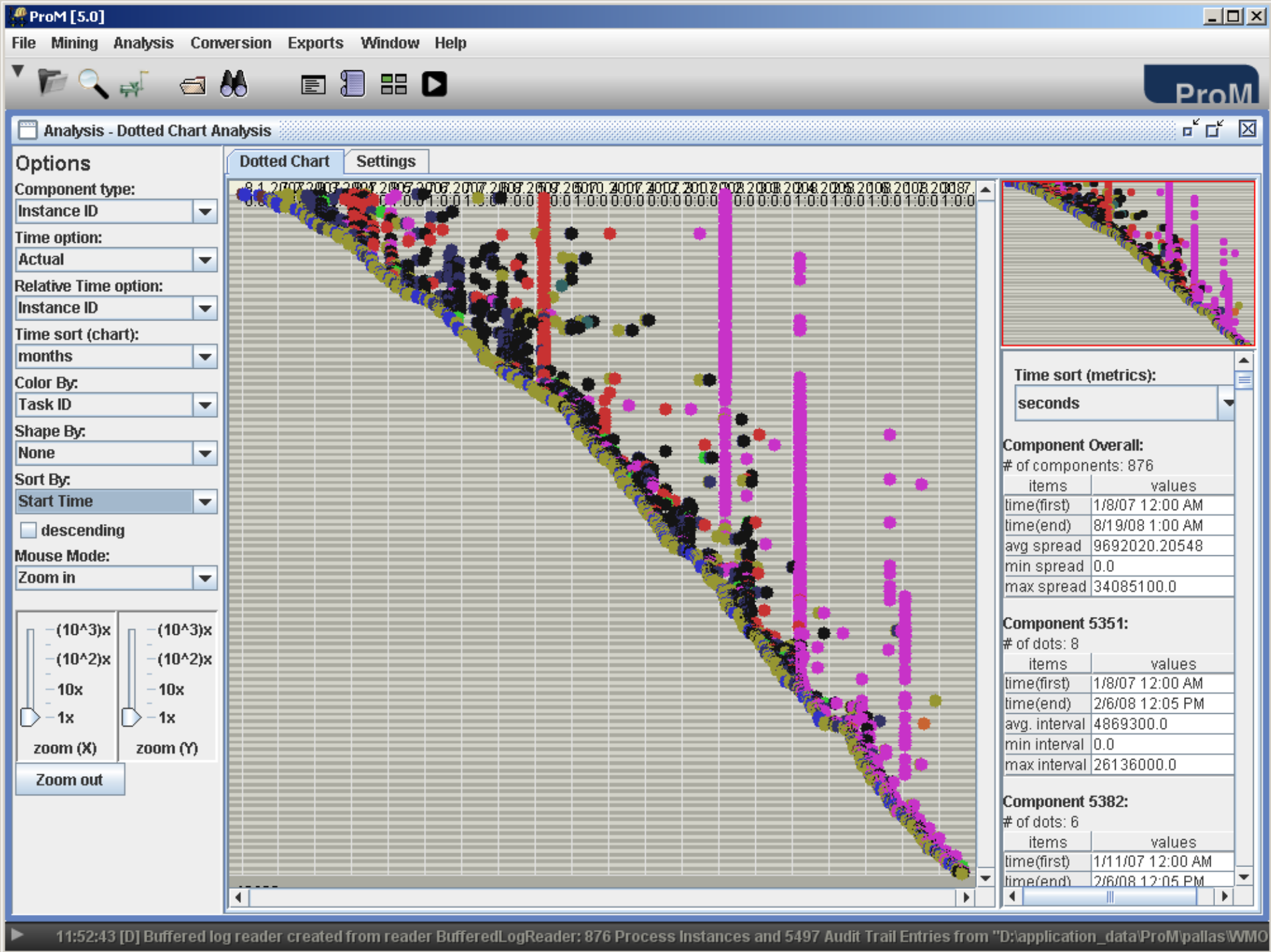
- **WMO: supporting citizens of municipalities (illness, handicaps, elderly, etc.).**
- **Examples:**
 - wheelchair, scootmobiel, ...
 - adaptation of house (elevator), ...
 - household help, ...
- **Event log:**
 - CSV format (Pallas Athena/Futura Technology)
 - 876 process instances
 - 5497 events
 - January 2007- August 2008

Original log



All elements

Model element	Event type	Occurrences (absolute)	Occurrences (relative)
Toetsen en beslissen	complete	910	16.554%
Rapportage & beschikking	complete	872	15.863%
Verzending\dossiervorming	complete	844	15.354%
Aanvraag registratie	complete	842	15.317%
Slotfase	complete	828	15.063%
Administratieve verwerking	complete	616	11.206%
Wachten terugmelding zorgaanb.	complete	515	9.369%
Retour	complete	60	1.092%
Eigen onderzoek	complete	6	0.109%
Onderzoek CIZ	complete	4	0.073%





Settings

[illegible]

Time sort (metrics):

seconds

Component Overall:

```
# of components: 876
```

items	values
time(first)	1/8/07 12:00 AM
time(end)	8/19/08 1:00 AM
avg spread	9692020.20548
min spread	0.0
max spread	34085100.0

Component 5757:

```
# of dots: 2
```

items	values
time(first)	2/1/07 12:00 AM
time(end)	2/1/07 12:00 AM
avg. interval	0.0
min interval	0.0
max interval	0.0

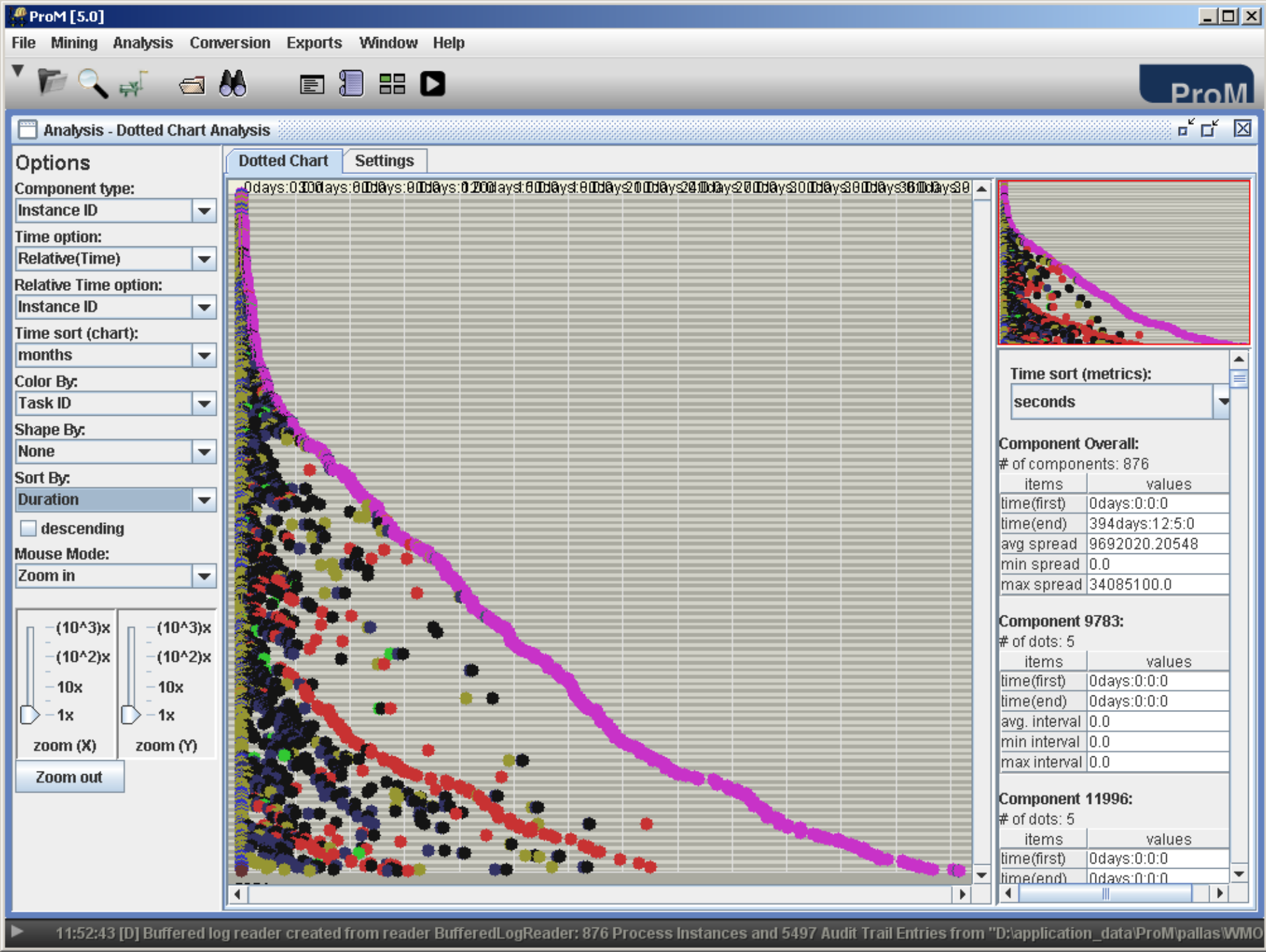
Component 5768:

```
# of dots: 1
```

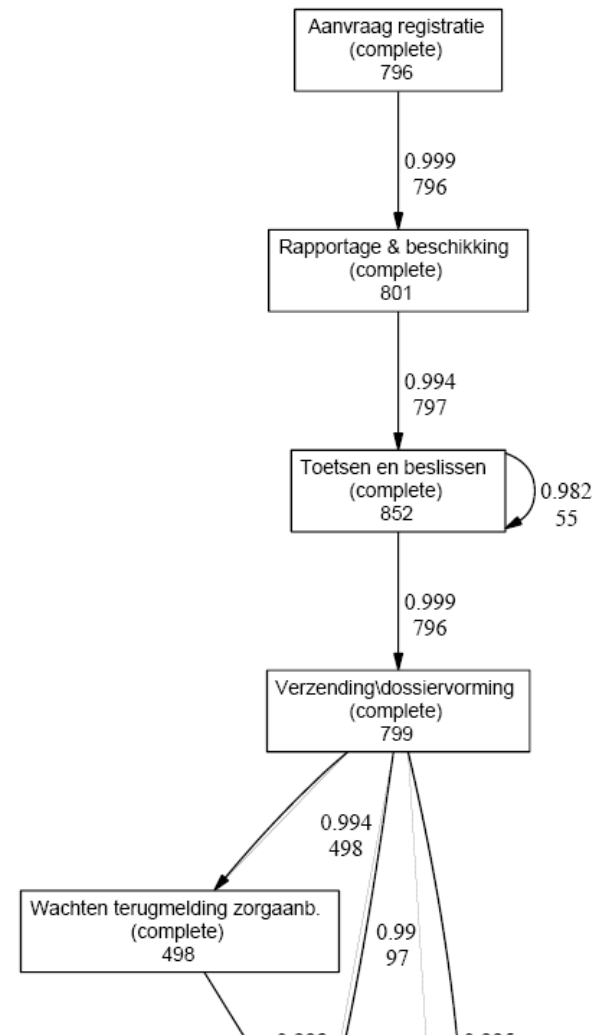
items	values
time(first)	2/5/07 12:00 AM
time(end)	2/5/07 12:00 AM

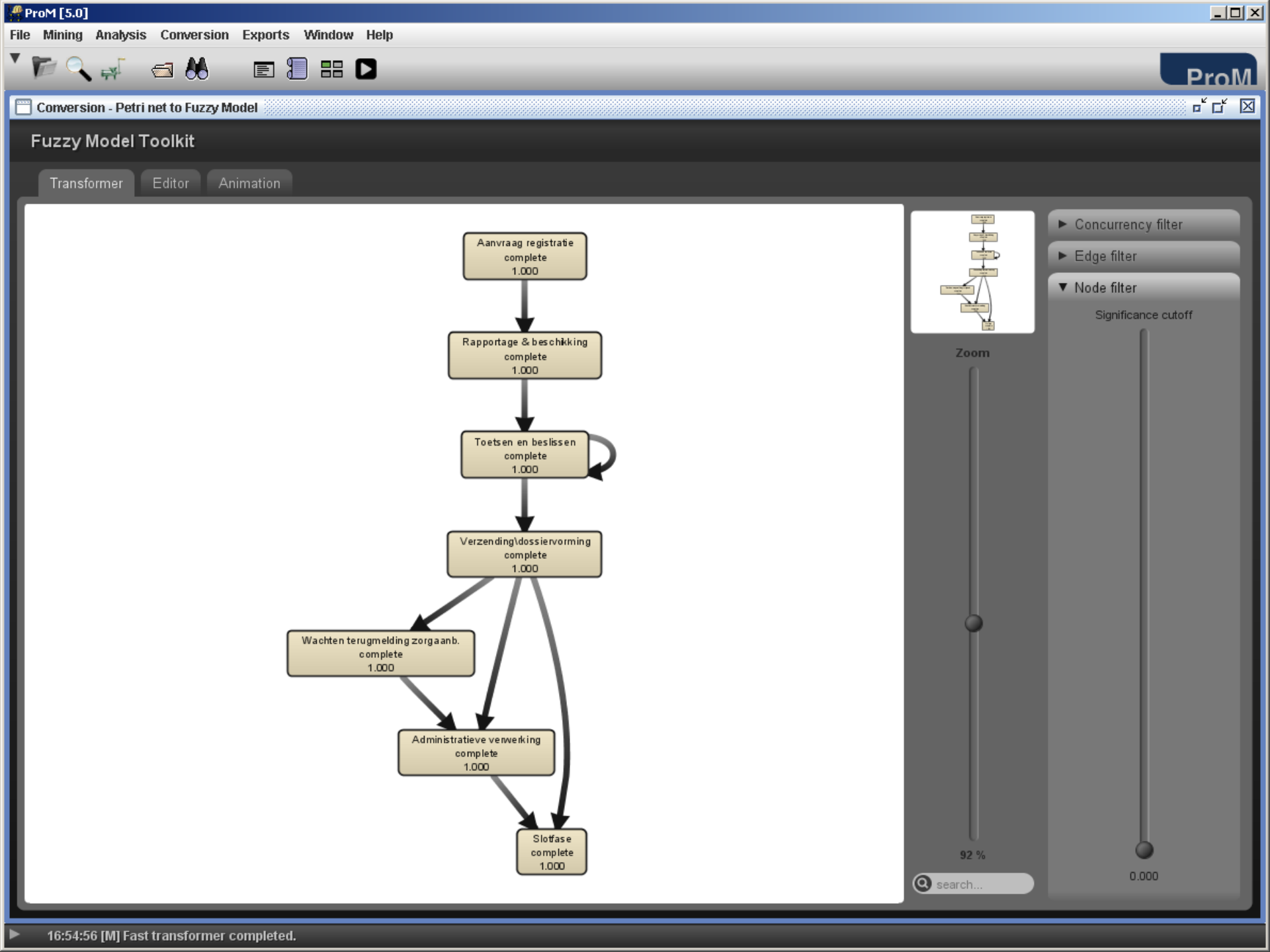
"D:\application_data\ProM\pal

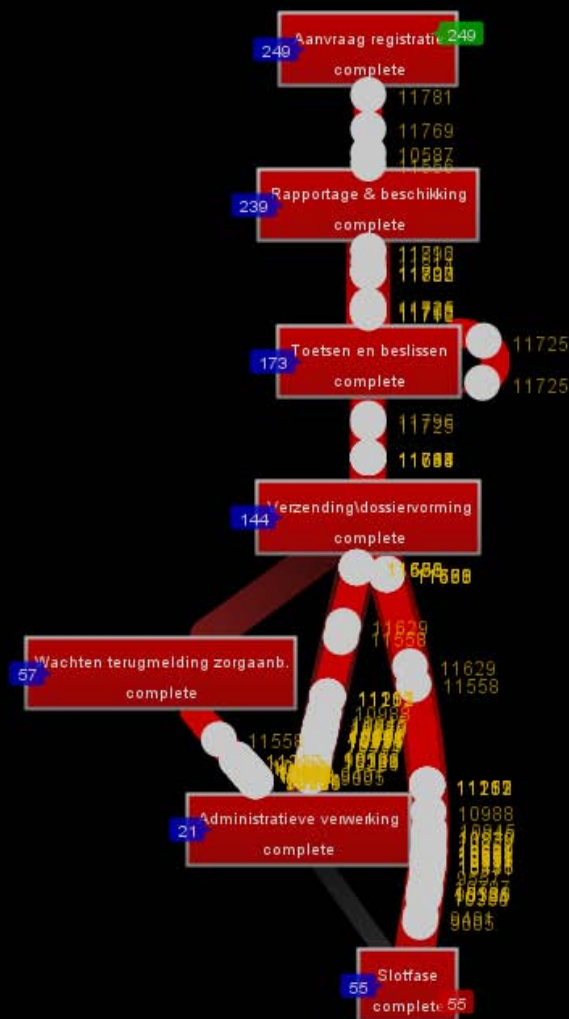
Zoom out



Discovered process

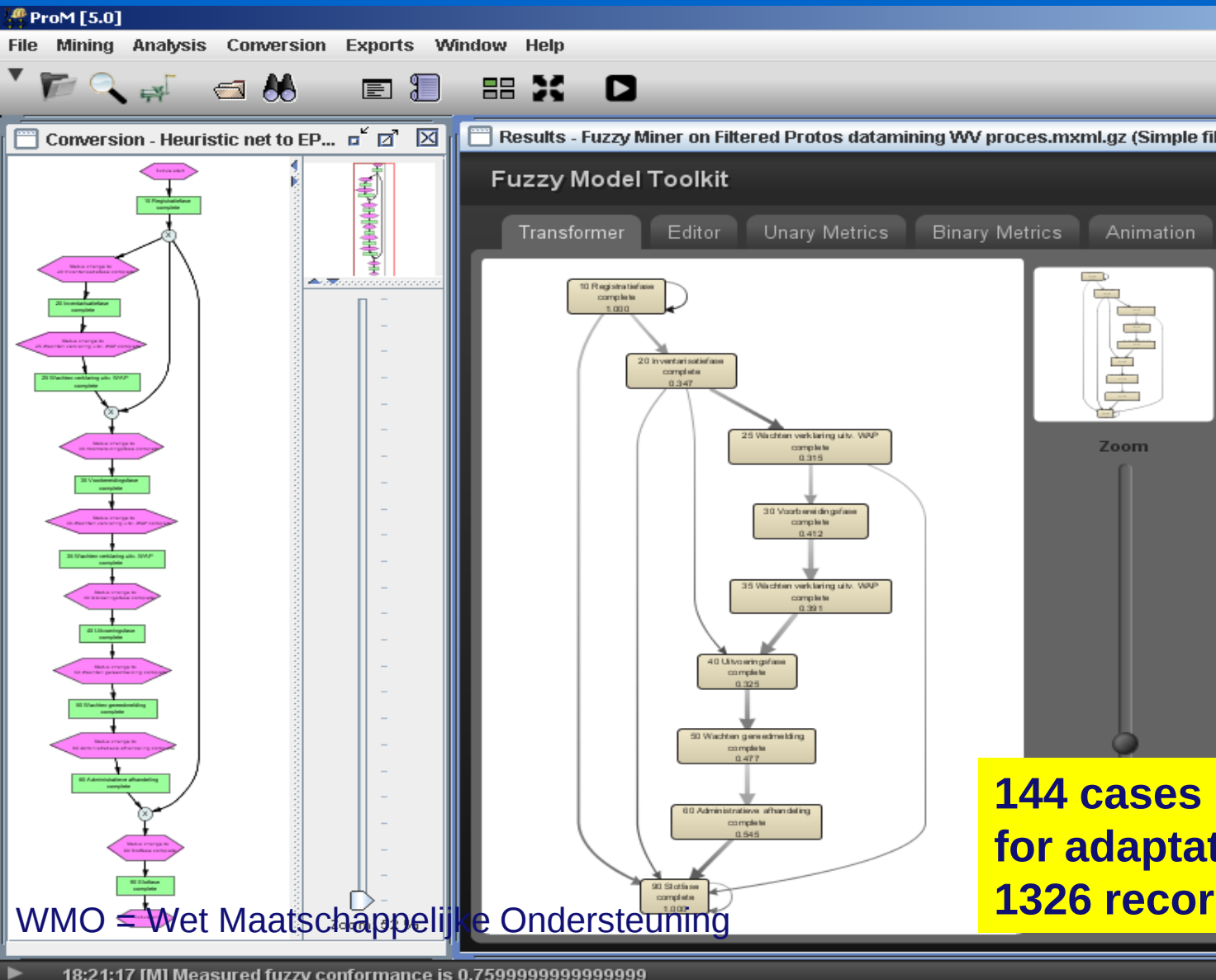






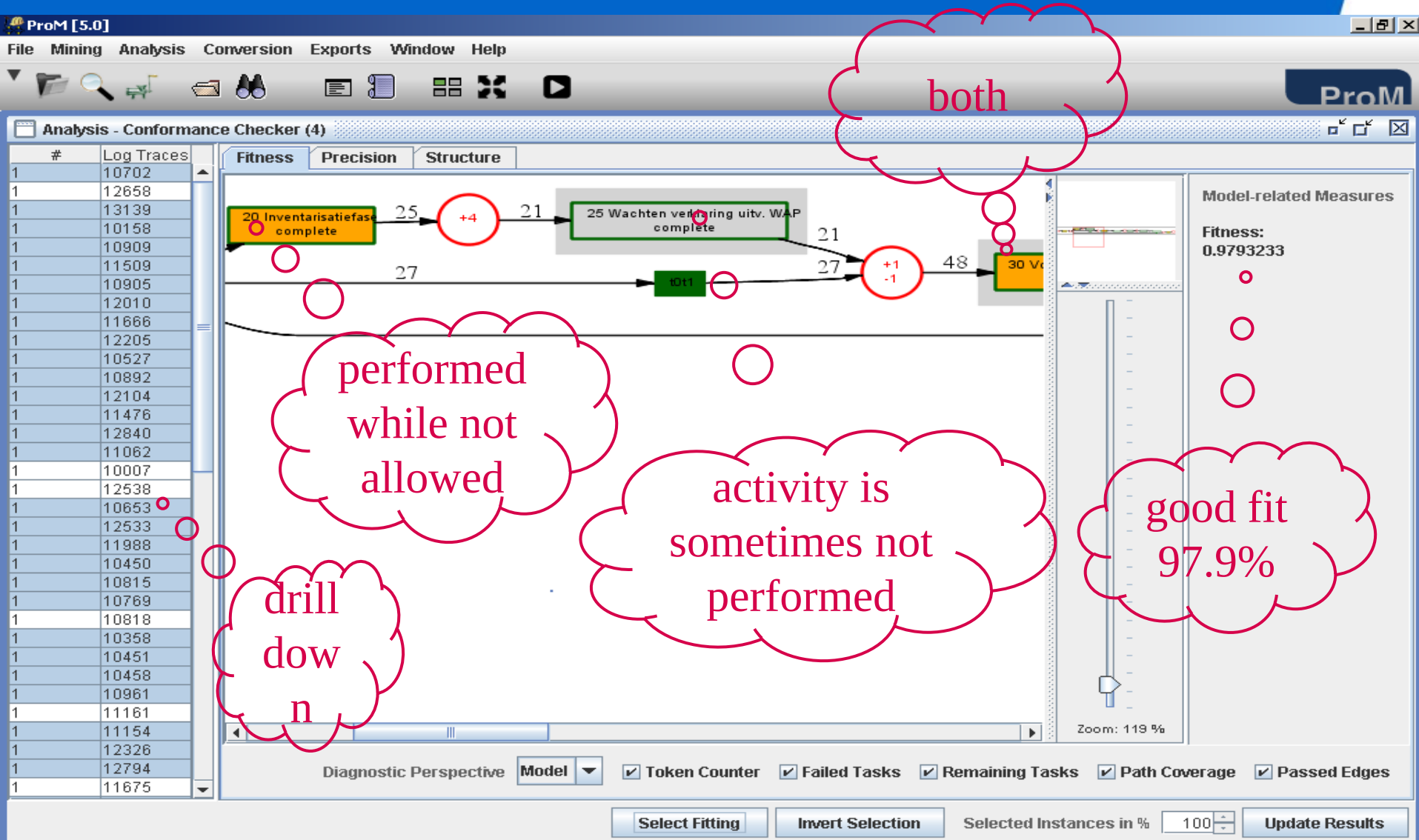
18:14:30 Tue 3 Jun '08

Discovered WMO process

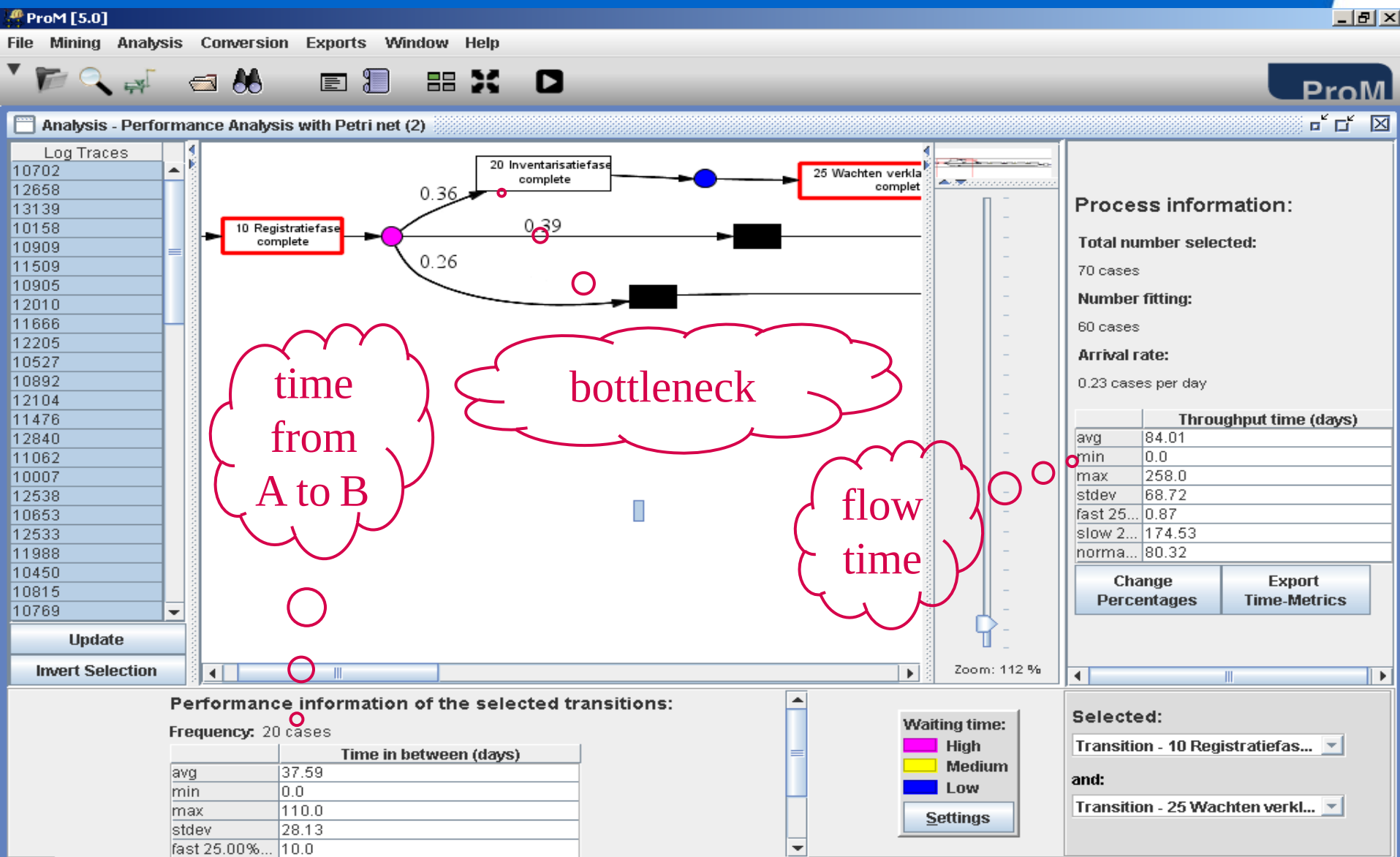


144 cases (i.e., requests for adaptation of house)
1326 recorded events

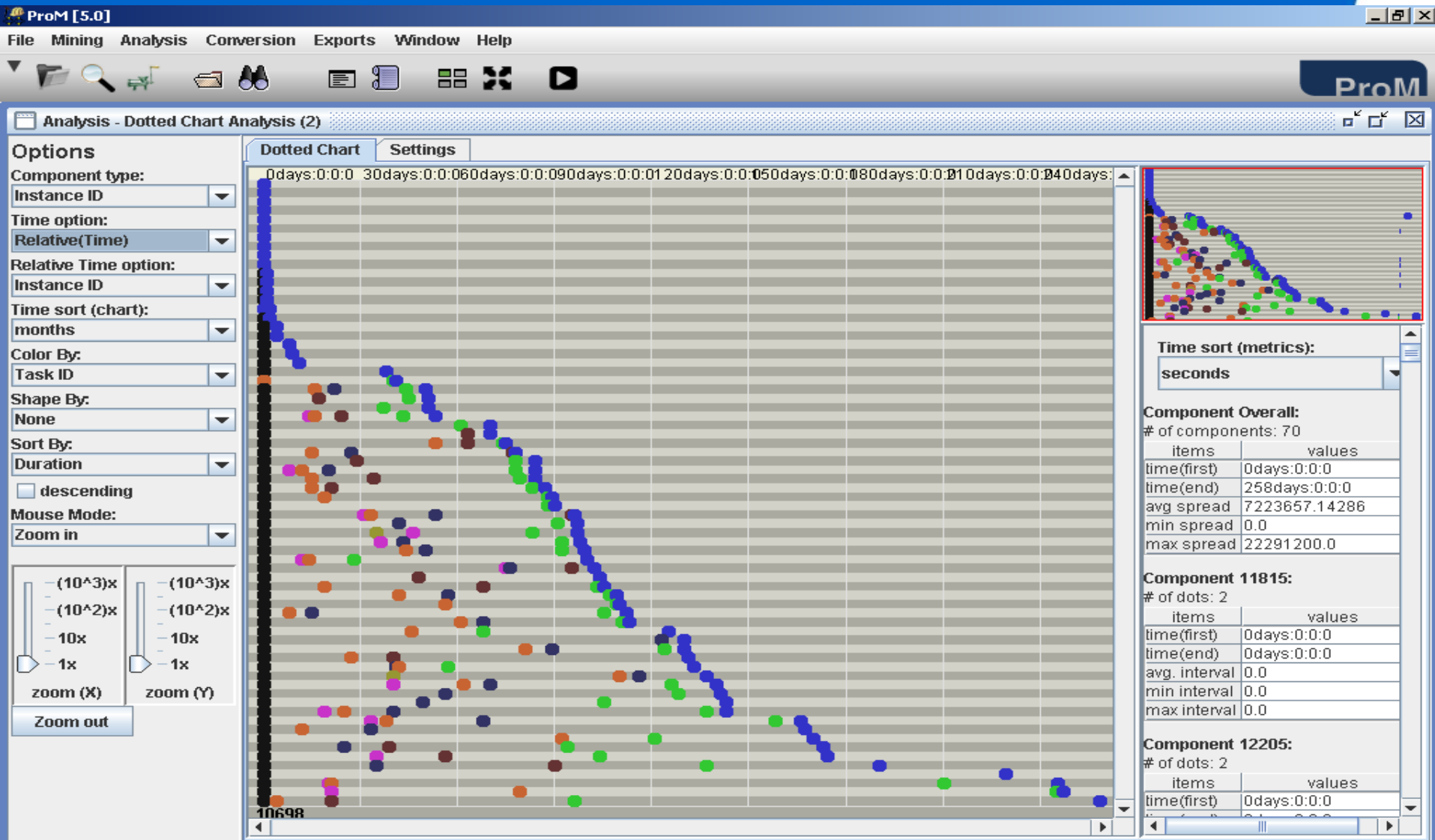
Conformance check of discovered model



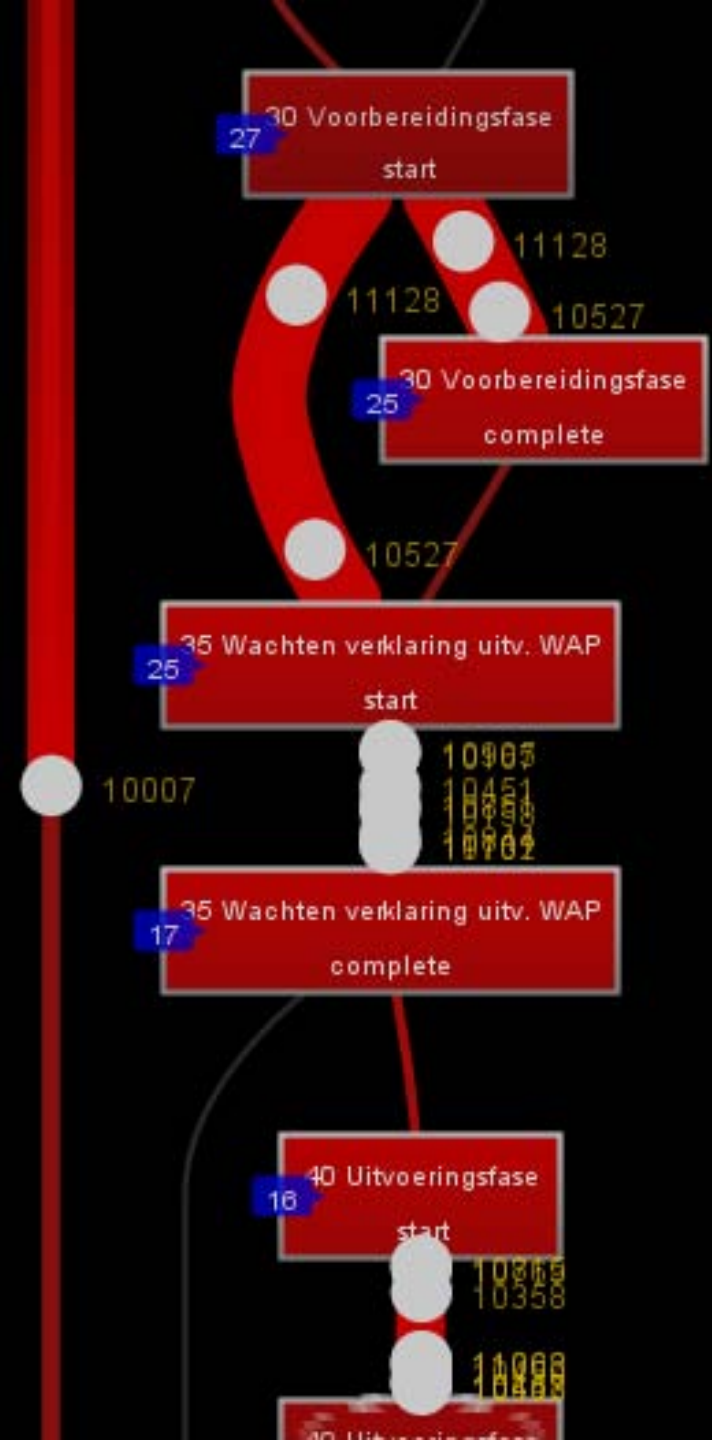
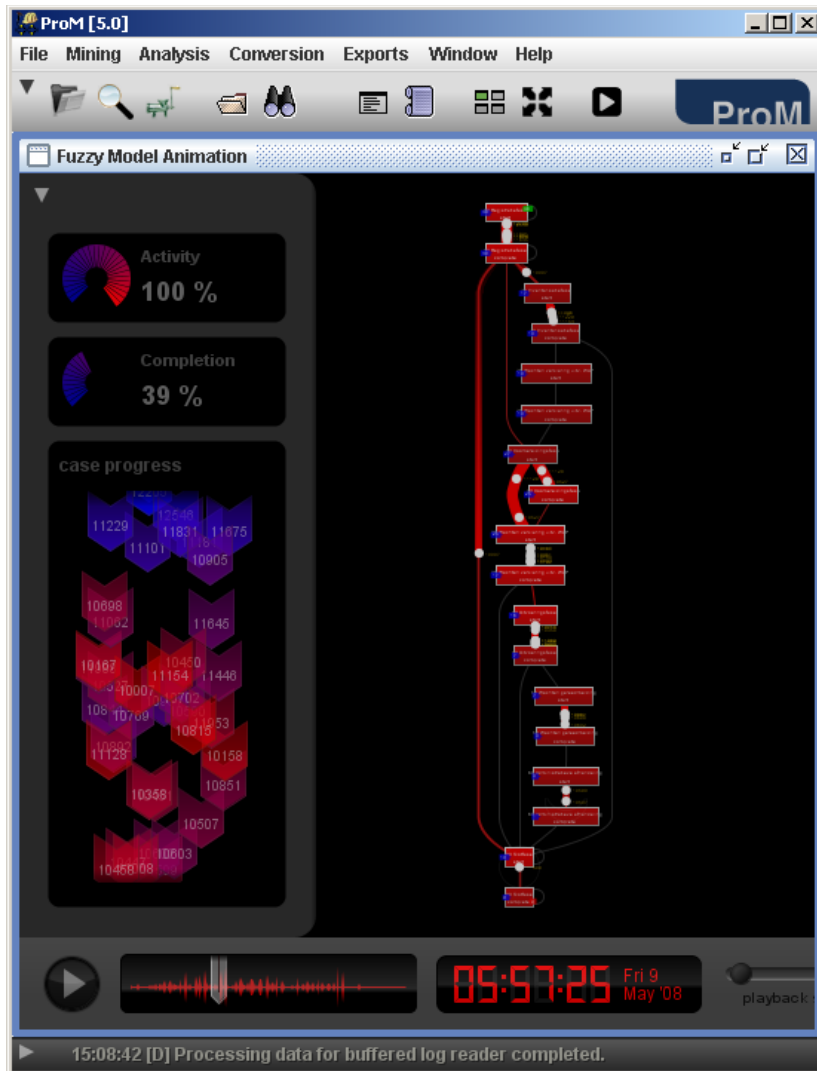
Performance analysis



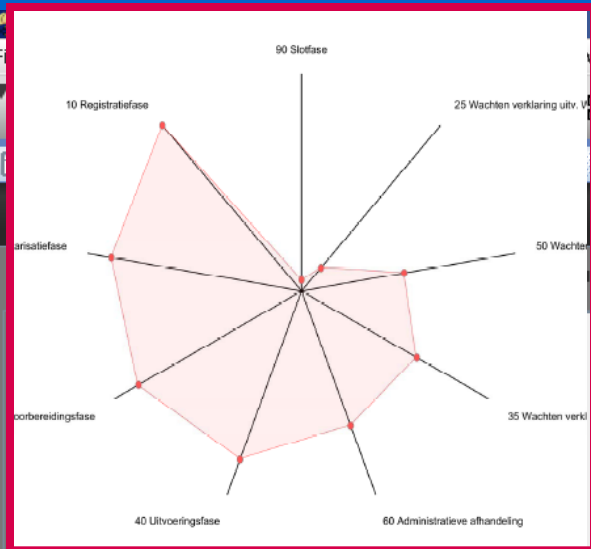
Events sorted by duration



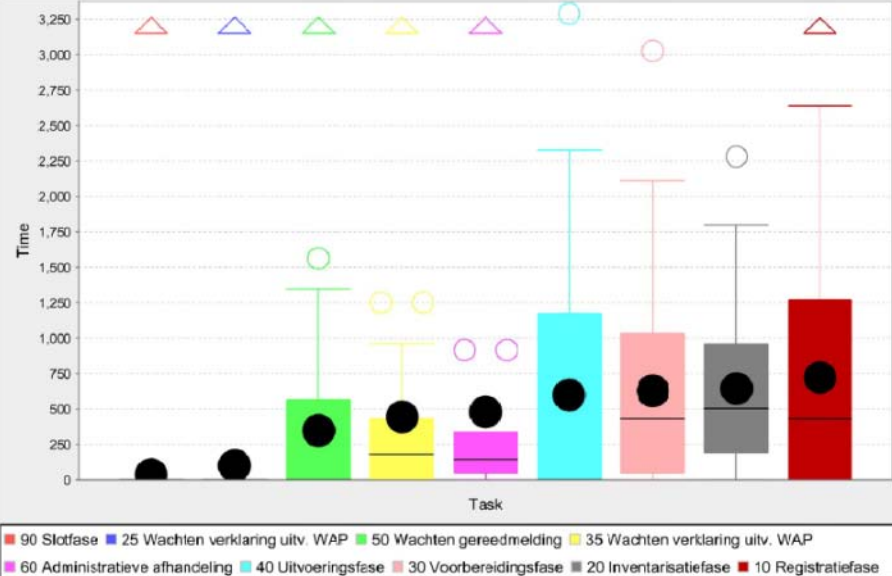
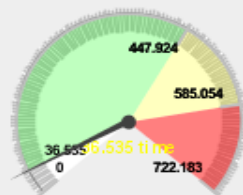
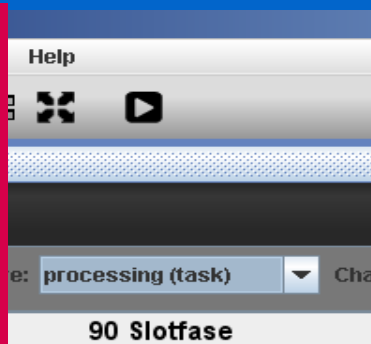
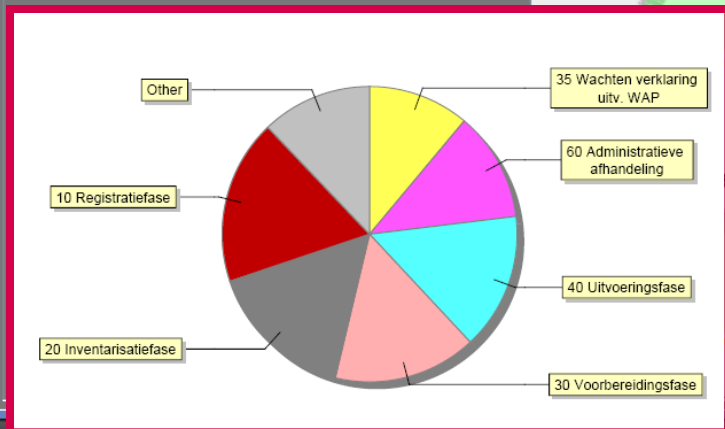
"Real" animation



And of course ...



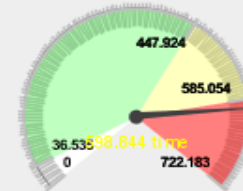
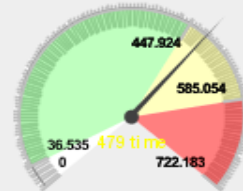
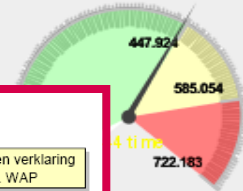
Upper value (normal) 447.9239436619718
 Upper value (warning) 585.0535211267606
 Update



35 Wachten verklaring uitv. WAP

60 Administratieve afhandeling

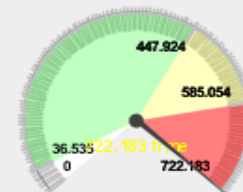
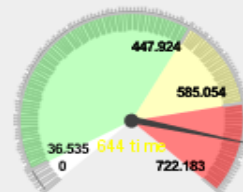
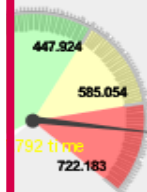
40 Uitvoeringsfase



30 Voorbereidingsfase

20 Inventarisatiefase

10 Registratiefase



Process Discovery

α



TU **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

$>, \rightarrow, ||, \#$ relations

- Direct succession: $x > y$ iff for some case x is directly followed by y .
- Causality: $x \rightarrow y$ iff $x > y$ and not $y > x$.
- Parallel: $x || y$ iff $x > y$ and $y > x$
- Choice: $x \# y$ iff not $x > y$ and not $y > x$.

case 1 : task A
case 2 : task A
case 3 : task A
case 3 : task B
case 1 : task B
case 1 : task C
case 2 : task C
case 4 : task A
case 2 : task B
case 2 : task D
case 5 : task E
case 4 : task C
case 1 : task D
case 3 : task C
case 3 : task D
case 4 : task B
case 5 : task F
case 4 : task D



ABCD
ACBD
EF

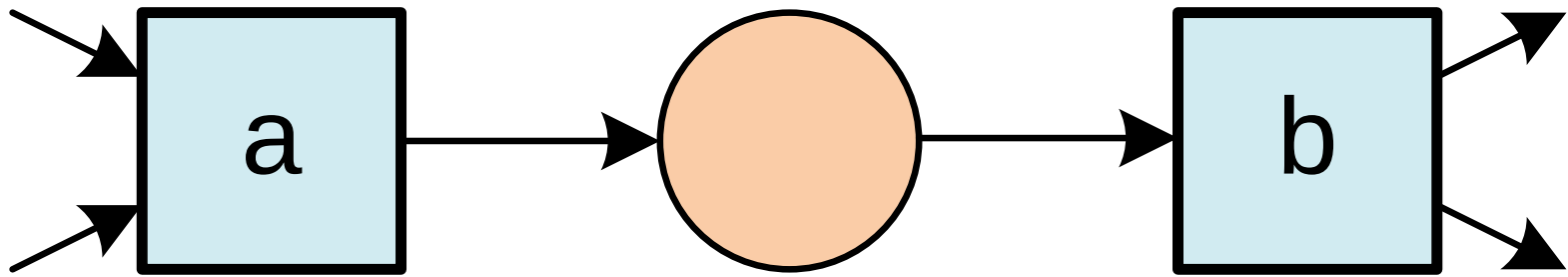


A > B
A > C
B > C
B > D
C > B
C > D
E > F

A → B
A → C
B → D
C → D
E → F

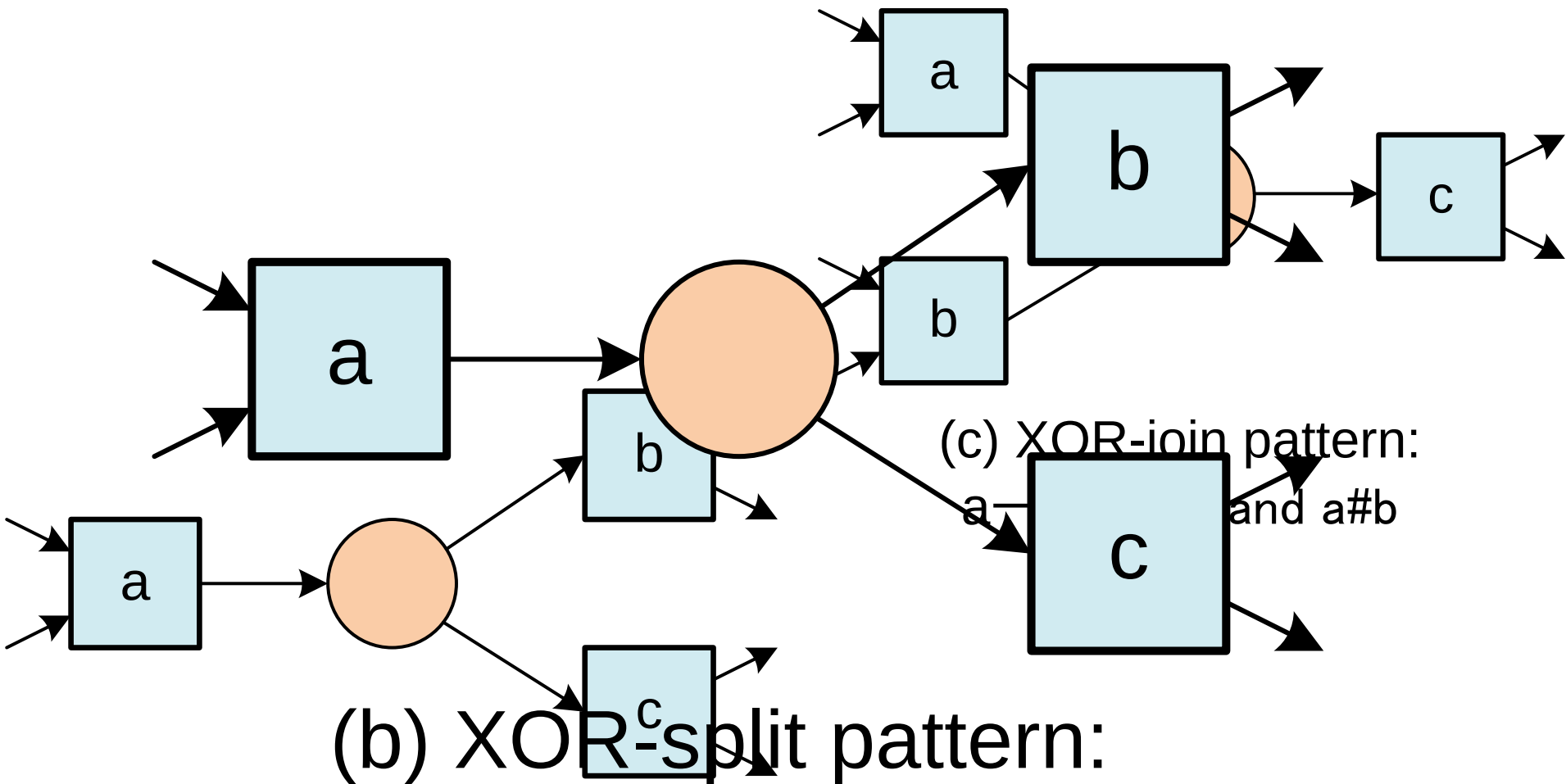
B || C
C || B

Basic Idea Used by α Algorithm (1)



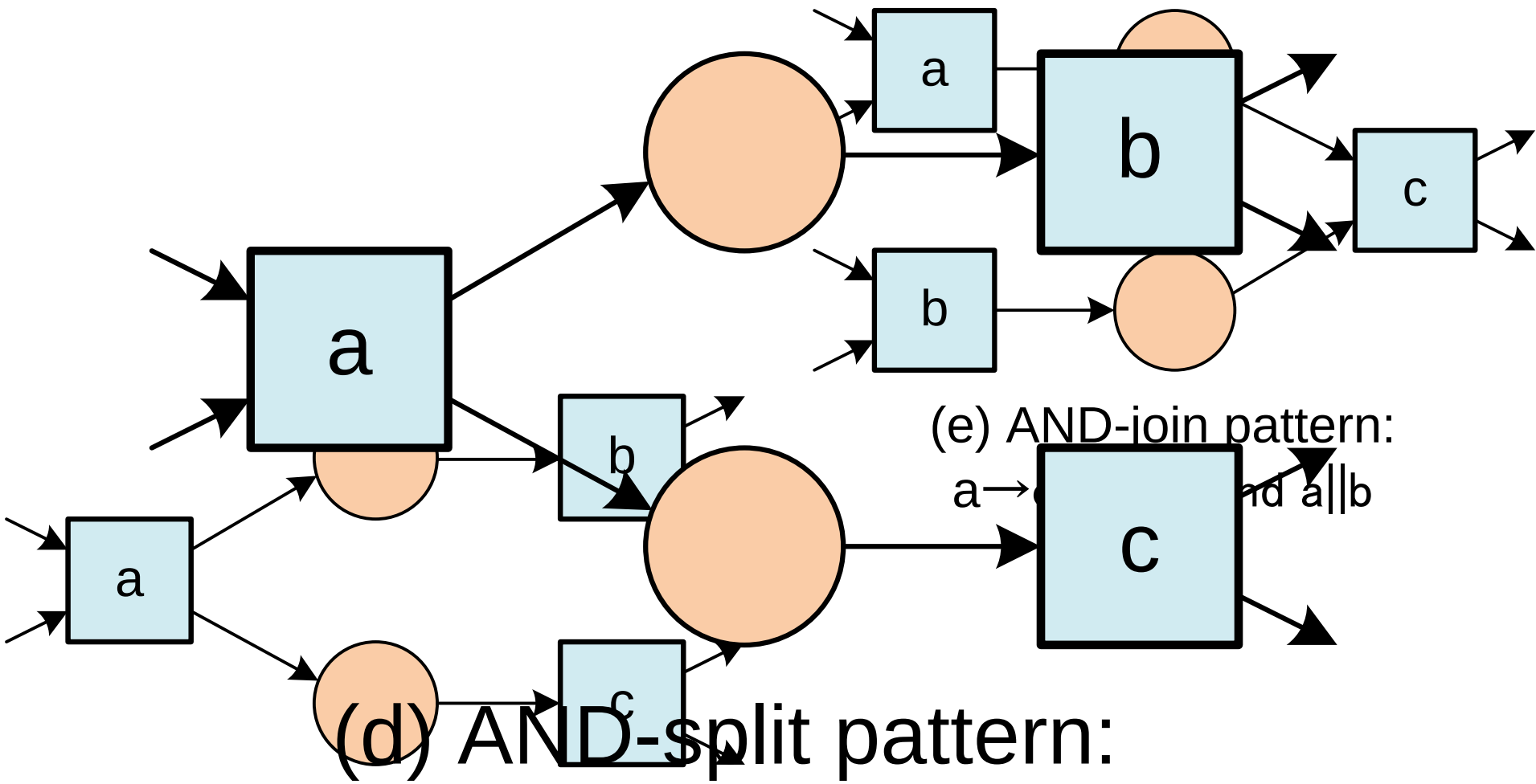
(a) sequence pattern: $a \rightarrow b$

Basic Idea Used by α Algorithm (2)



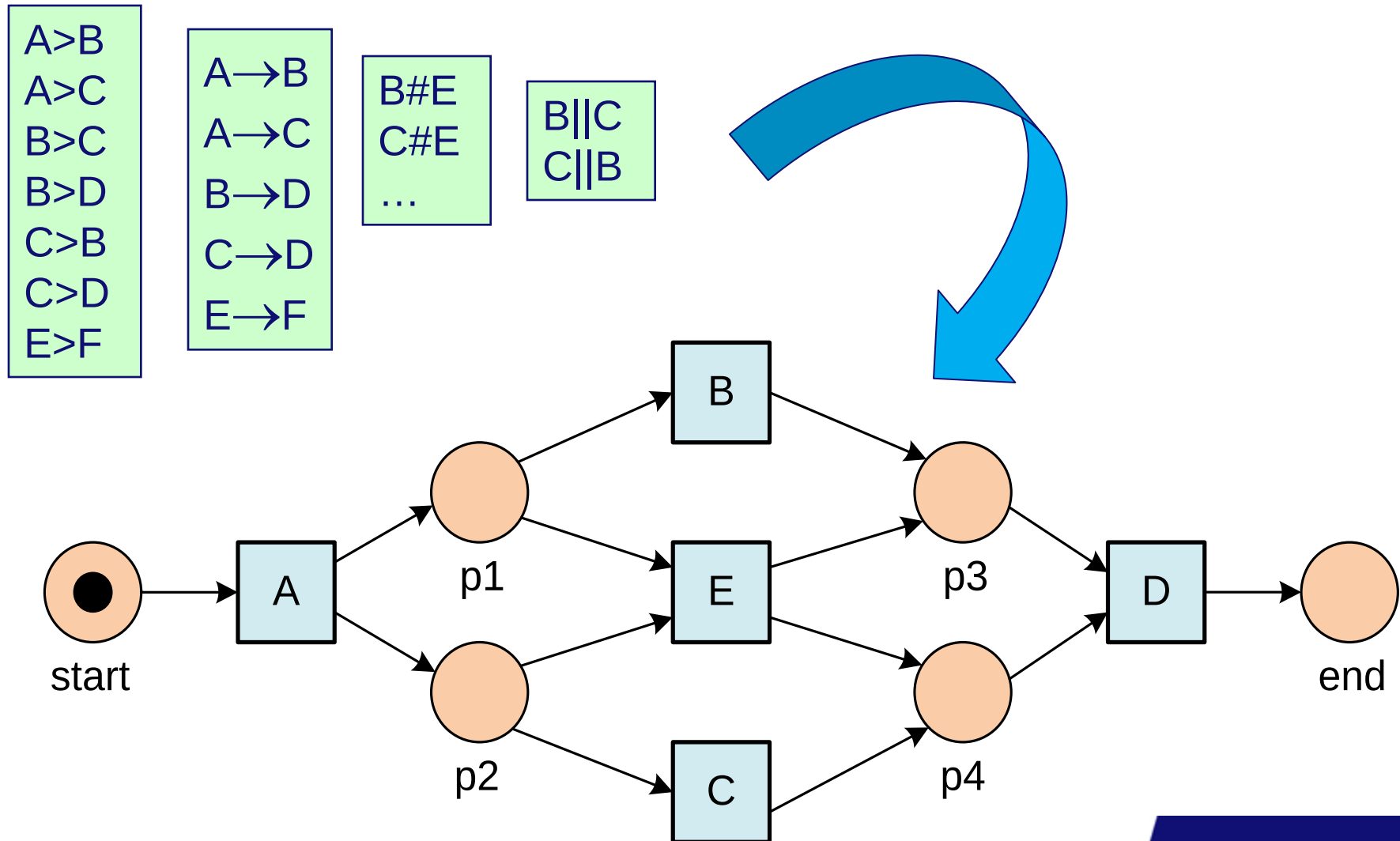
(b) XOR-split pattern:
 $a \rightarrow b$, $a \rightarrow c$, and $b \# c$
 $a \rightarrow b$, $a \rightarrow c$, and $b \# c$

Basic Idea Used by α Algorithm (3)



(d) AND-split pattern:
 $a \rightarrow b$, $a \rightarrow c$, and $b \parallel c$

Example Revisited



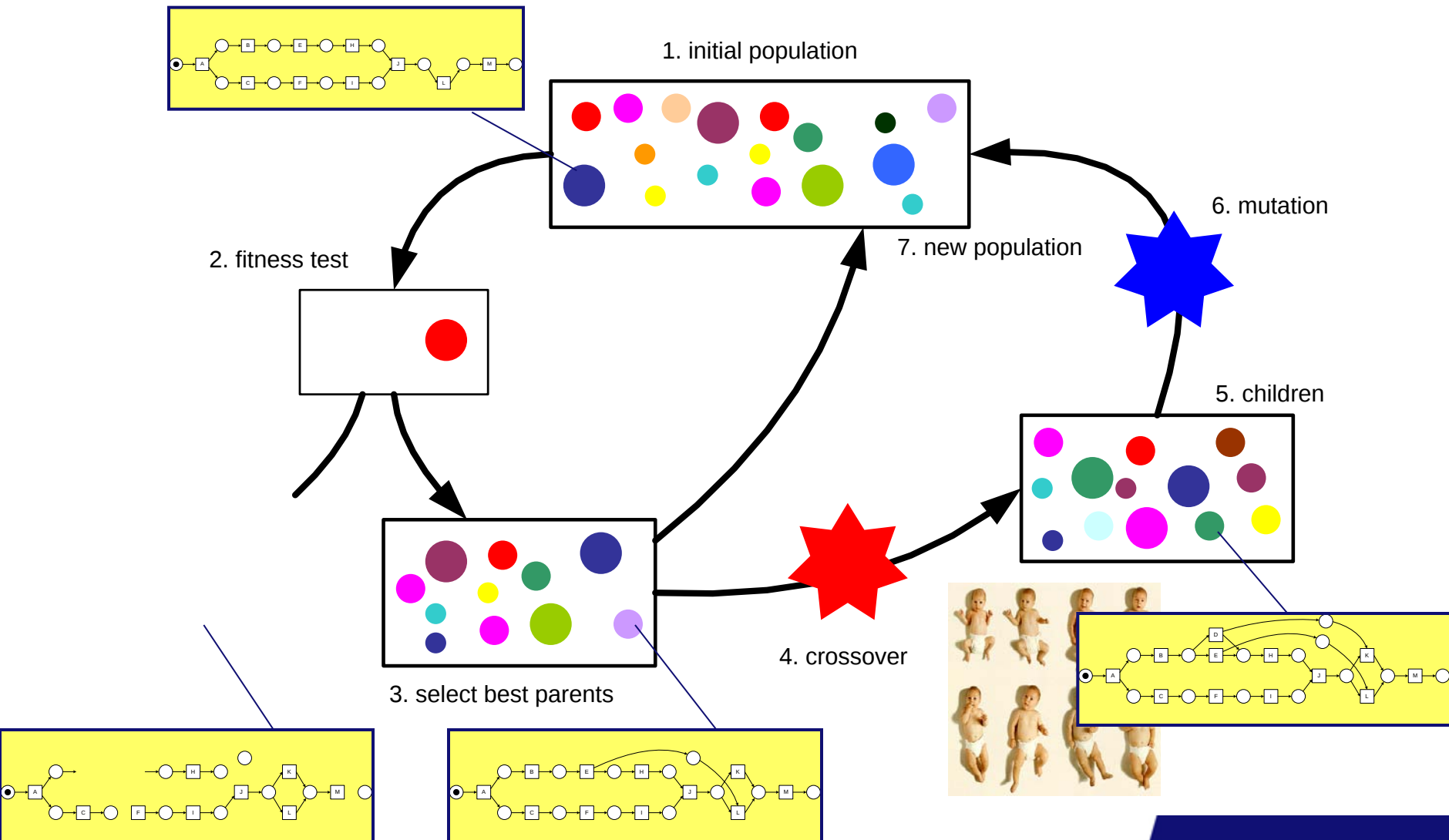
Overview of process discovery techniques

- **Classical techniques (e.g., learning state machines and the theory of regions): cannot handle concurrency and/or do not generalize (i.e., if it did not happen, it cannot happen).**
- **Algorithmic techniques**
 - Alpha miner
 - Alpha+, Alpha++, Alpha#
 - Heuristic miner
 - Multi phase miner
 - ...
- **Genetic process mining**
- **Region-based process mining**
 - State-based regions
 - Language based regions



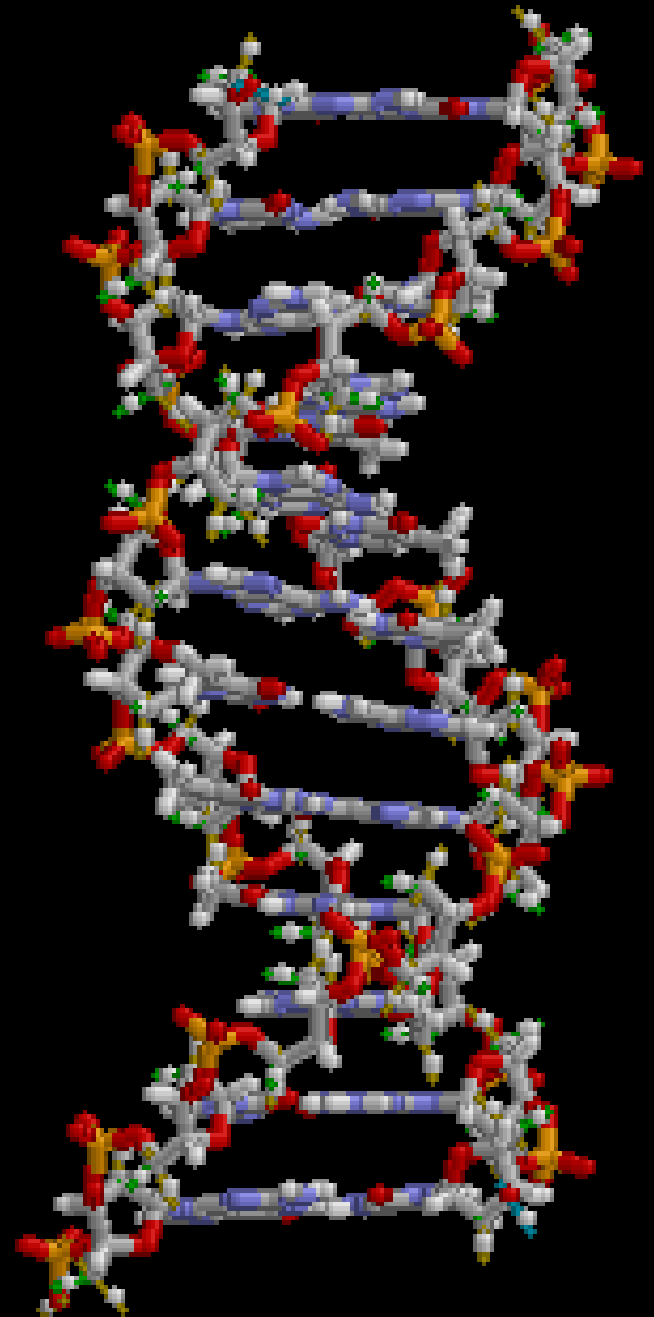
Genetic Mining

(Ana Karla Alves de Medeiros et al.)



Challenges

- Dealing with semi-structured processes
- Balancing between overfitting and underfitting
- $365!/(365^{365}) \approx 1.454955\text{E-}157 \approx 0$
($1\text{E}79$ is the estimated number of atoms in the universe)
- Suitable notations/visualizations (executable?, declarative?, "zoomable", "abstractable", etc.)
- Data explosion (mining TB logs)



Process Mining Software



Futura Reflect



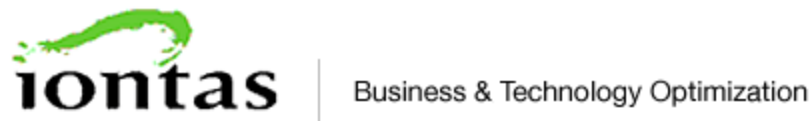
Comprehend



ARIS Process Performance Manager



Interstage Automated Business Process Discovery & Visualization

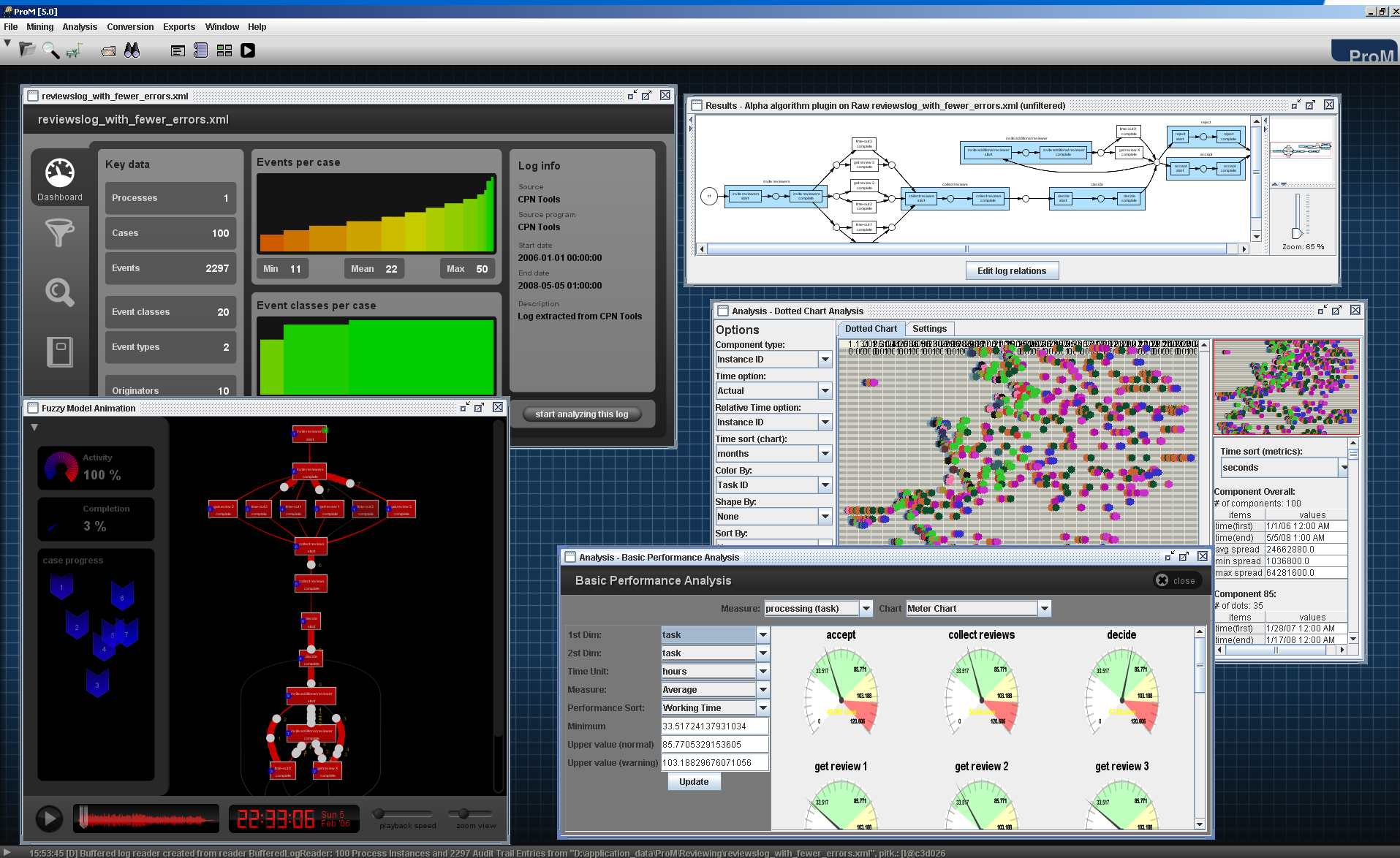


Process Discovery Focus

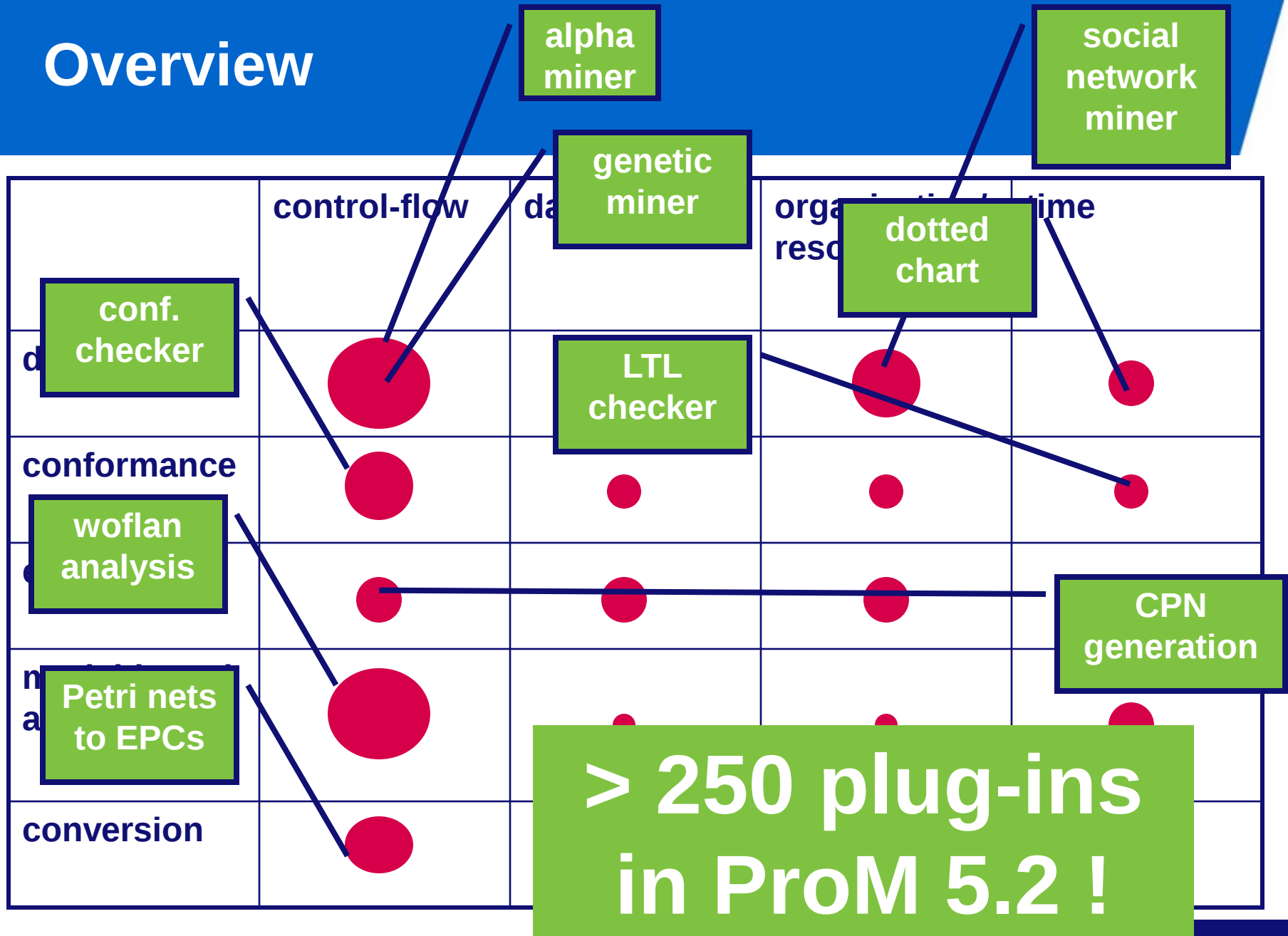


Enterprise Visualization Suite

Screenshot of ProM 5.0



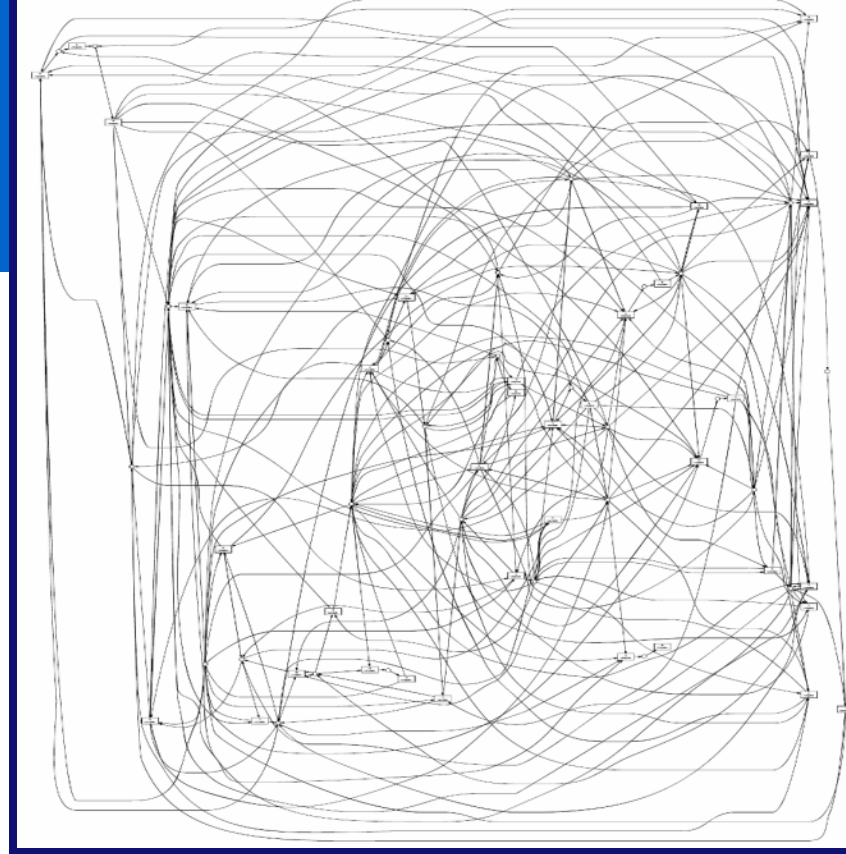
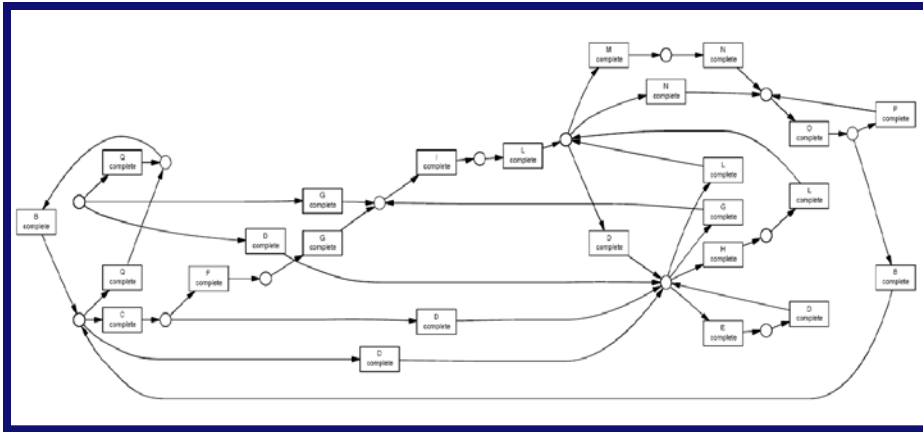
Overview



Reality $\neq$ PowerPoint (or Visio)



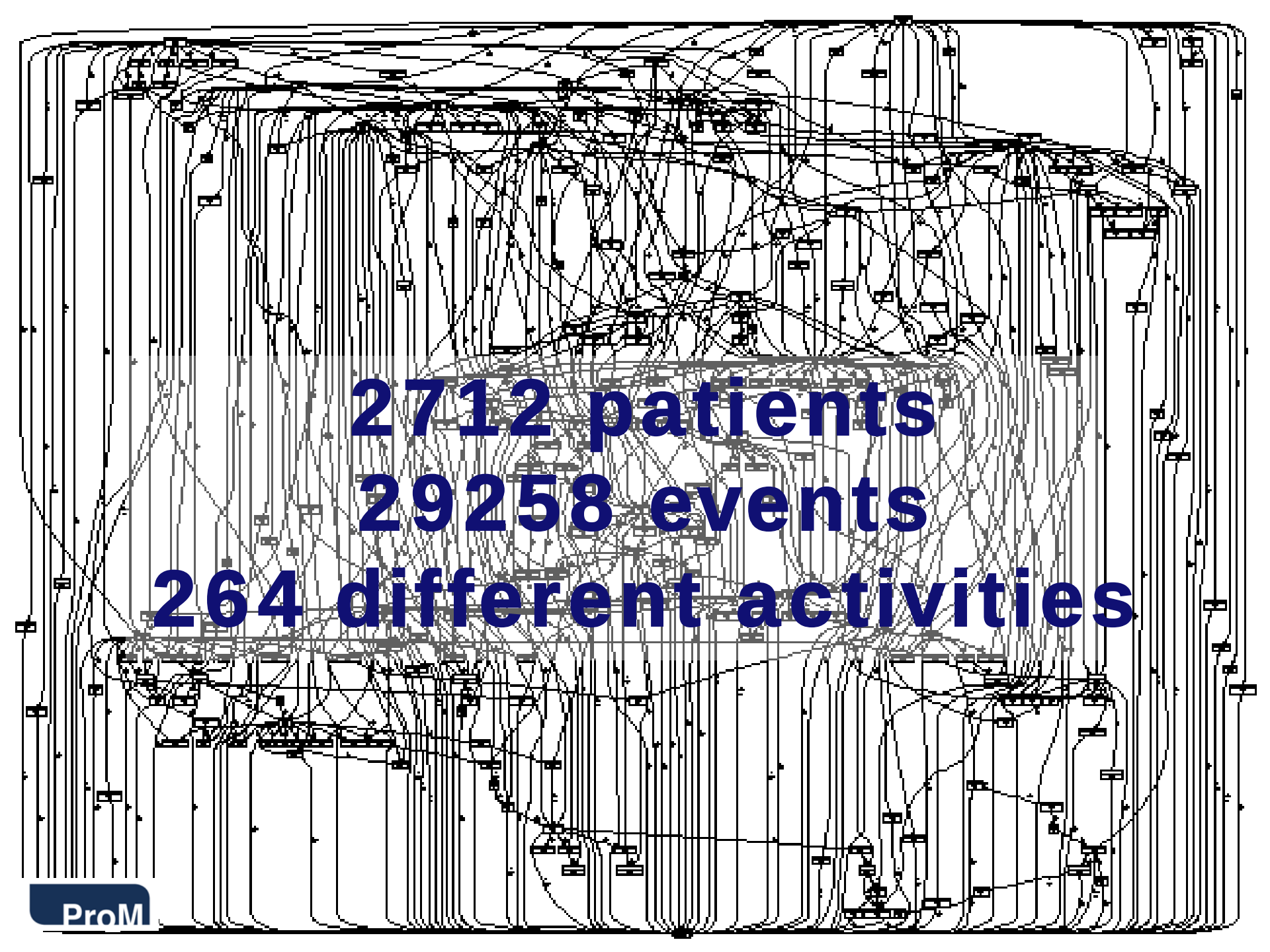
Process spectrum



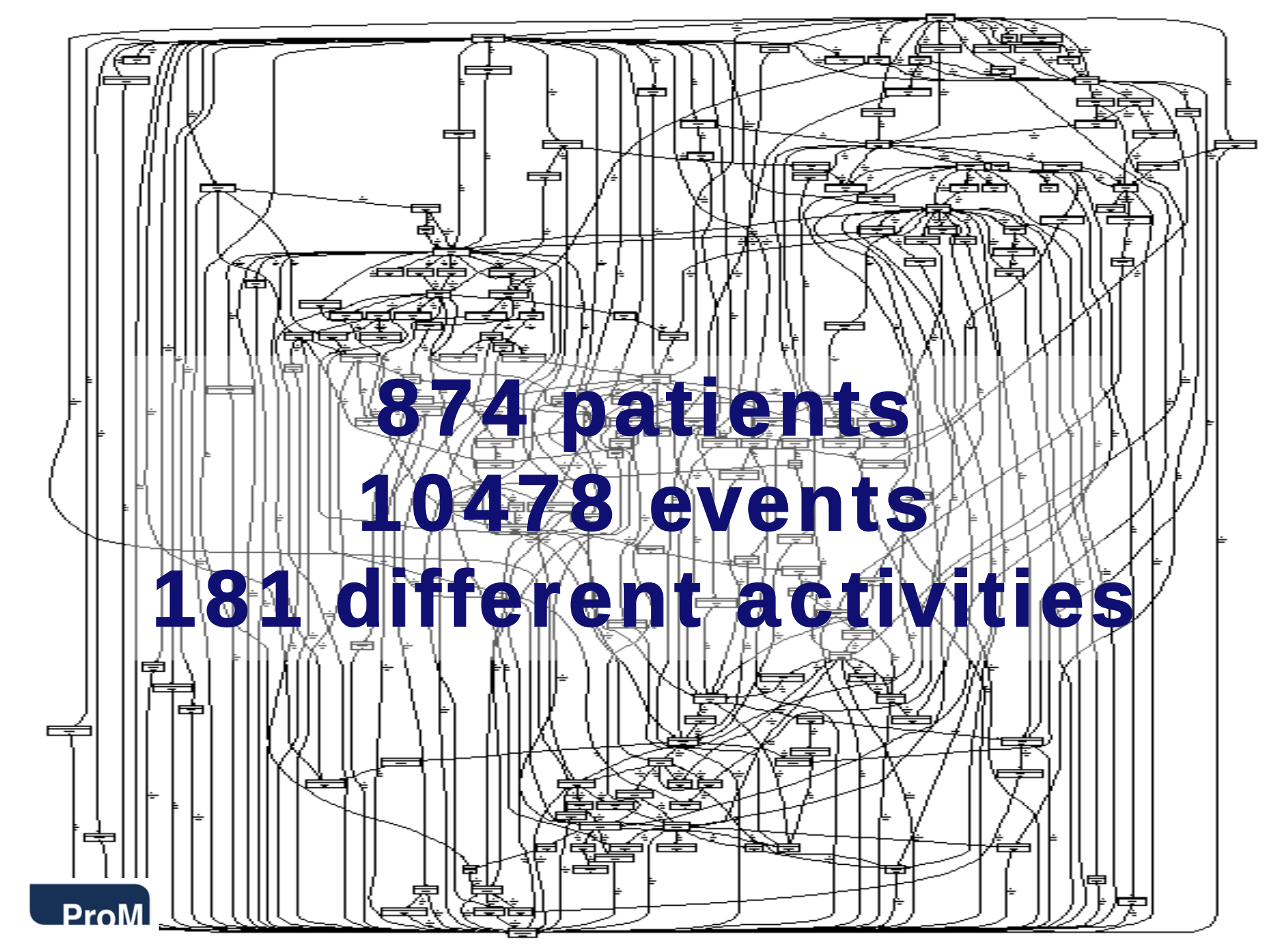
**structured
(Lasagna)**

**unstructured
(Spaghetti)**

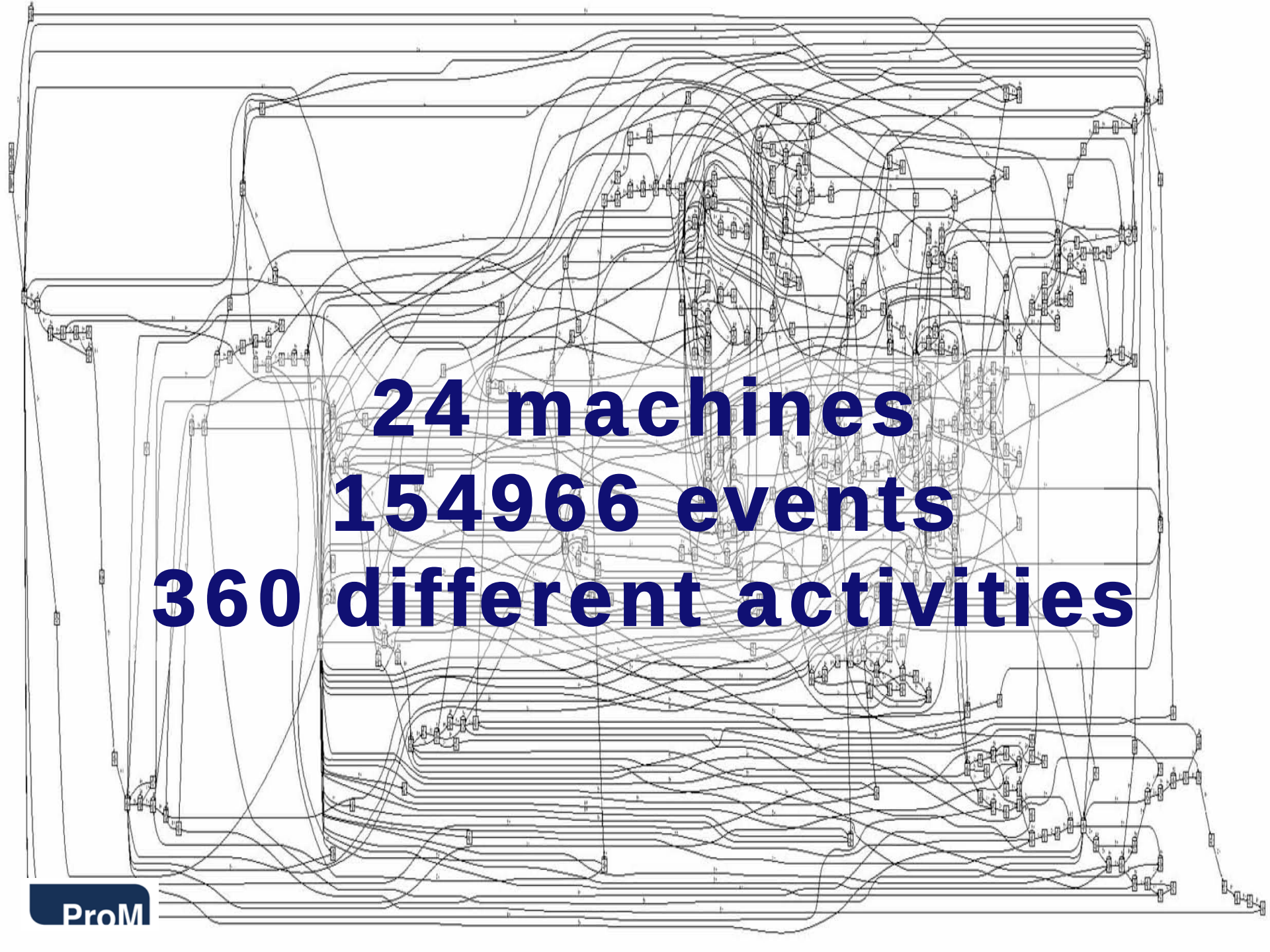
375 houses
18640 events
82 different activities



2712 patients
29258 events
264 different activities



874 patients
10478 events
181 different activities

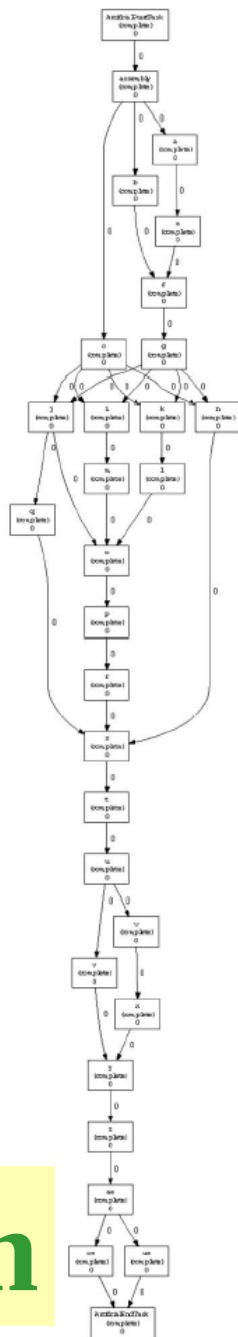


24 machines
154966 events
360 different activities

#	Log Traces
1	0431
1	0278
1	0185
1	0466
1	0391
1	1722
1	1694
1	1256
1	1343
1	1981
1	1754
1	1662
1	1453
1	1298
1	1876
1	1656
1	1099
1	1919
1	1348
1	1596
1	1164
1	1032
1	1794
1	1160

100

a complete



92

89

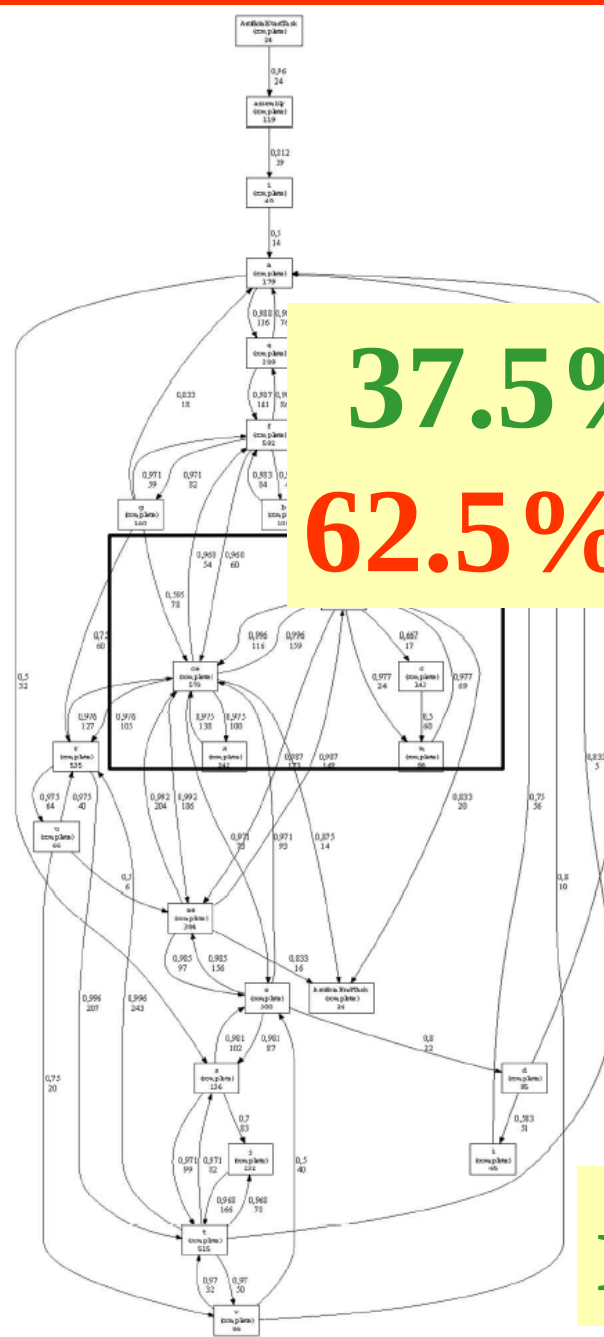
e complete

instances

Diagnostic Perspective

design

ProM



Fitness:
0.37501124

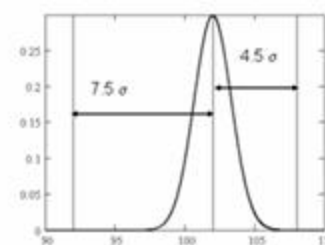
37.5% OK
62.5% NOK

with Coverage ☒ Passed Edges

100 Update Results

reality

Sigma niveau	Percentage goede producten	Aantal fouten per miljoen producten
1	31%	690.000
2	69,1%	309.000
3	93,32%	67.000
4	99,379%	6.200
5	99,9767%	230
6	99,99967%	3,4



6σ?

1σ!

Process Mining: TomTom for Business Processes



TU **e**

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

How can process mining help?



- Good maps?
- Navigation by PowerPoints?
- Traffic information?
- Where is the next fuel station?
- Who is in charge?
- Seamless zoom?
- Customizable views?
- When will the destination be reached?



ProM [5.0]

File Mining Analysis Conversion Exports Window Help

ProM

Results - Fuzzy Miner on Raw MiningBronbestand_Goed_Ontology.mxml (unfiltered)

Fuzzy Model Toolkit

Transformer Editor Unary Metrics Binary Metrics

ProM

Results - Fuzzy Miner on Raw MiningBronbestand_Goed_Ontology.mxml (unfiltered)

Fuzzy Model Toolkit

Transformer Editor Unary Metrics Binary Metrics

Concurrency filter

Edge filter

Node filter

Significance cutoff

Zoom

41 %

search...

0.485

08:49:29 [M] Measured fuzzy conformance is 0.6839424087761667



ProM

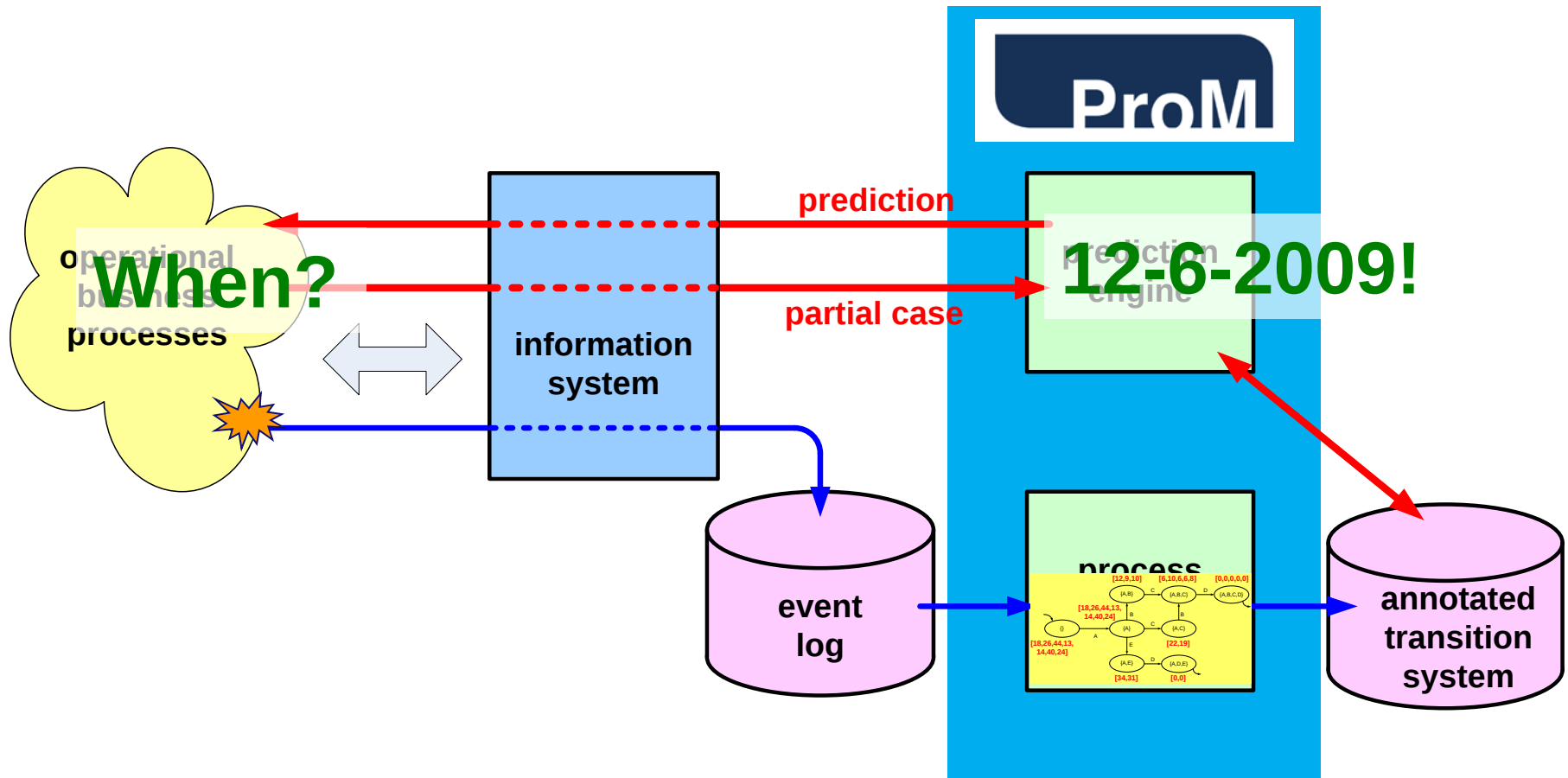
19:40:35 [D] Processing data for buffered log reader completed.

18:36:05 Fri 9 Nov '07

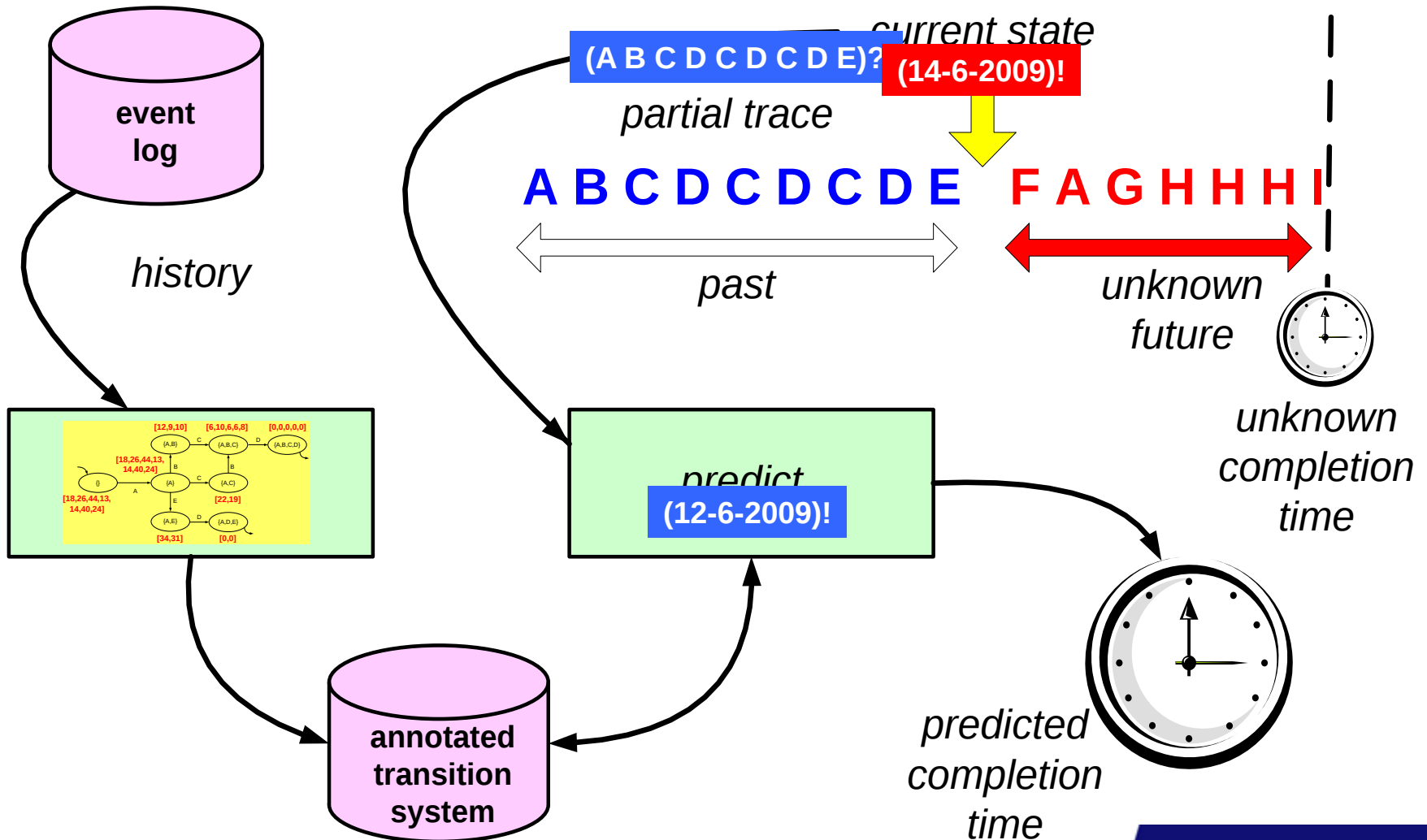
playback speed



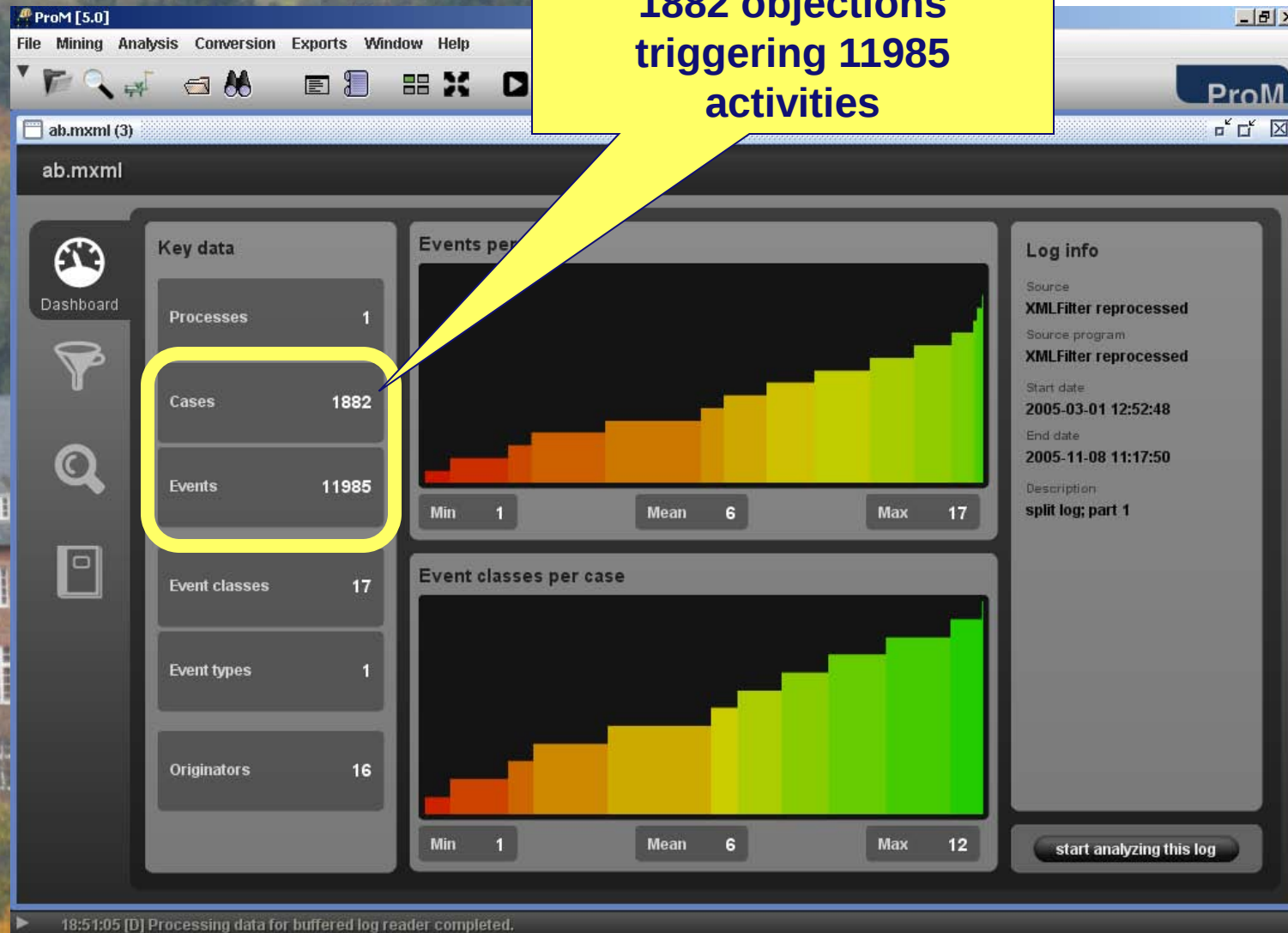
Approach



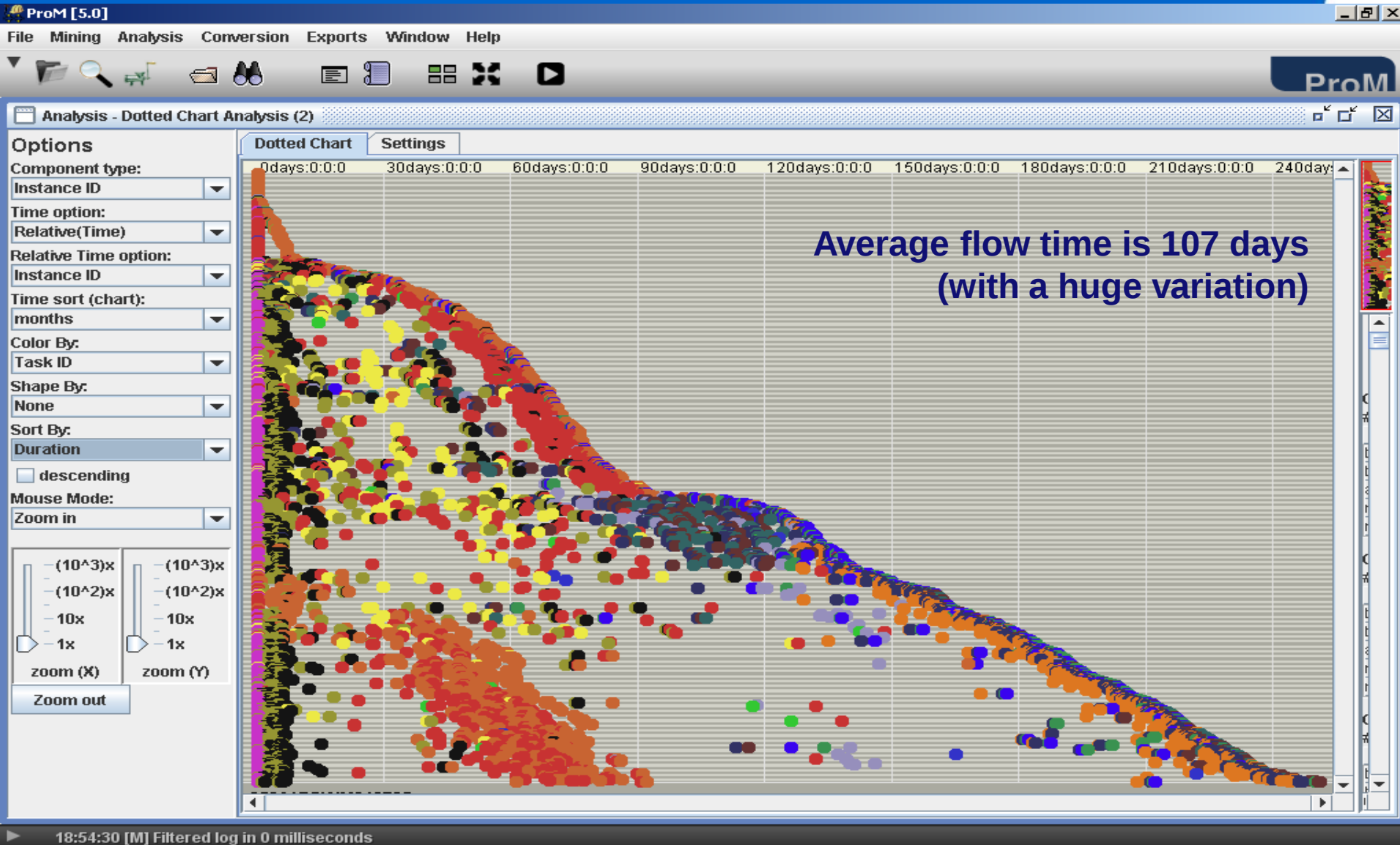
Input: partial trace and historic information



Example: WOZ process in Dutch Municipality



All 11985 events at a glance



ProM [5.0]

File Mining Analysis Conversion Exports Window Help

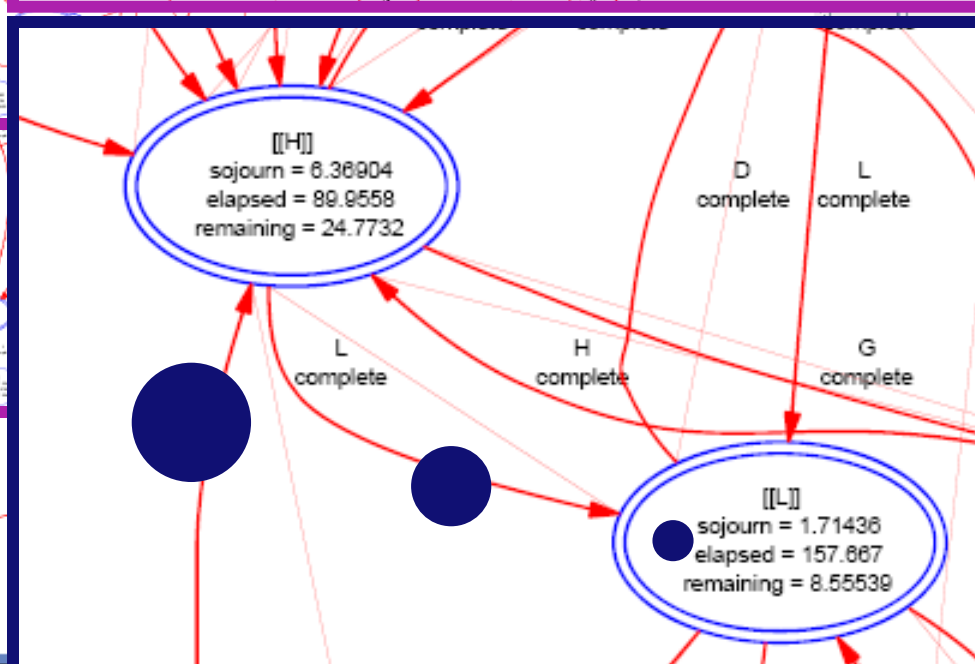
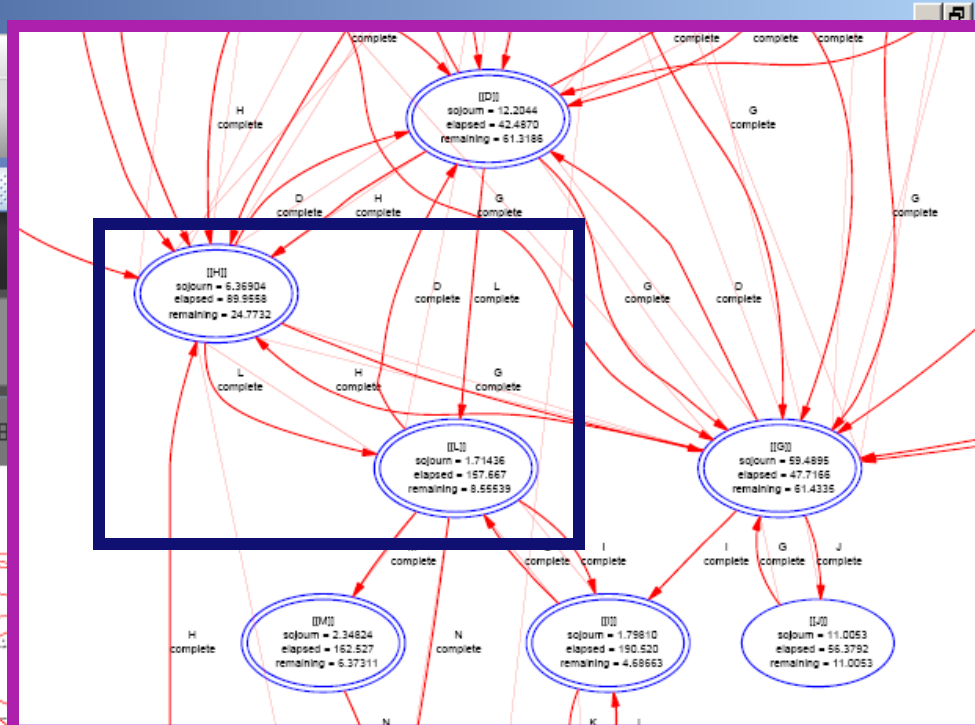
Analysis - FSM analyzer (16)

FSM based Performance Analysis

Event Log :

Time Unit: Measure: Color B

For partial traces corresponding to this state the estimated time until completion is 8.5 days



Parts Window Help

Mean Average Error (MAE)

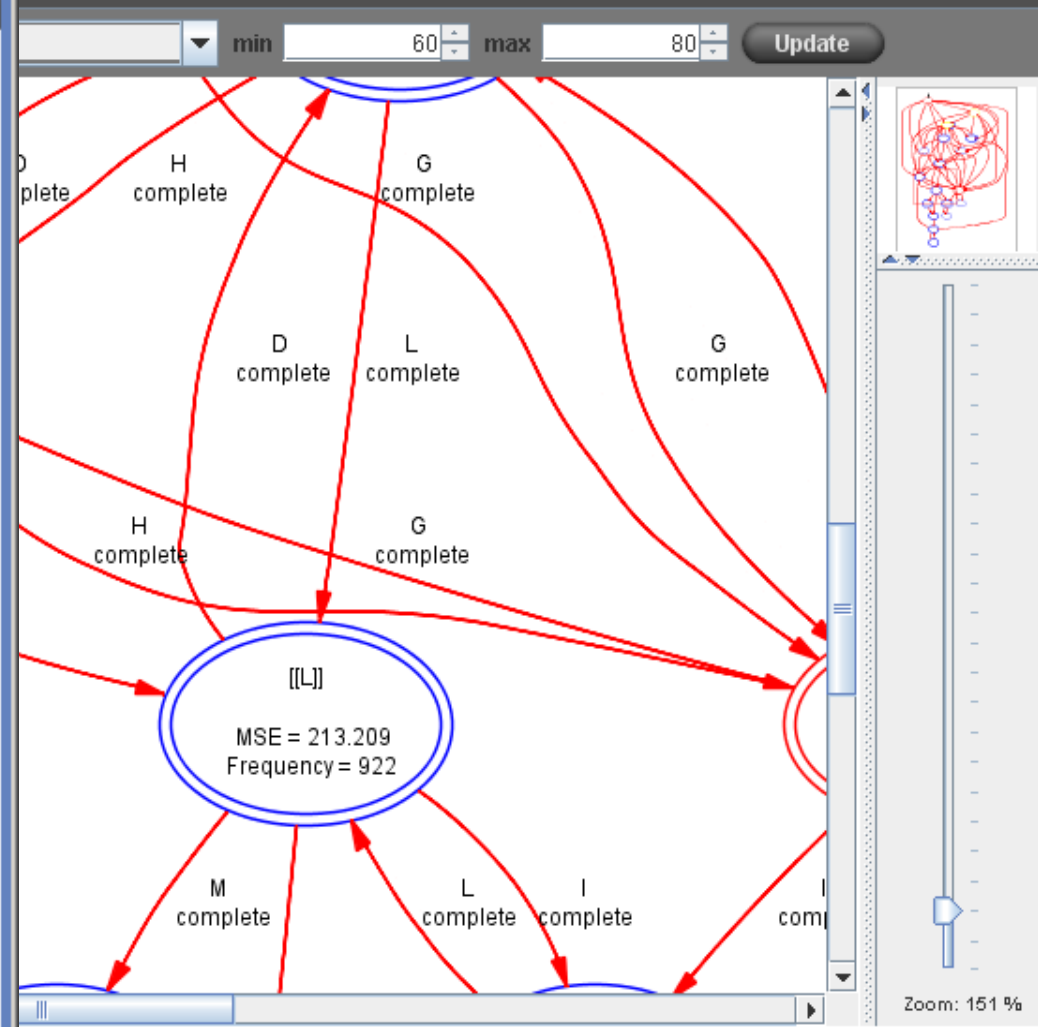
rooted MSE

MAPE

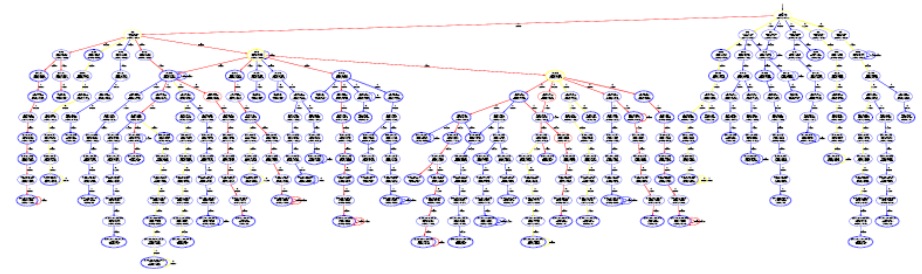
graph view text view

[[B]]	62.3442	70.5320	702.14	1488
[[C]]	64.8881	72.8452	406.781	896
[[D]]	63.1935	69.4784	51727.3	1240
[[G]]	72.8231	74.6510	24.3339	1546
[[I]]	3.25089	5.59291	103.264	629
[[L]]	9.82483	14.6016	406.304	922
[[M]]	8.89983	13.6616	1331.10	866
[[N]]	7.74203	12.2609	1478.89	799
[[O]]	3.17161	9.07740	1351.93	932
[[P]]	2.98513	8.61544	554.336	709
[[F]]	34.0397	44.4393	52294.3	16
[[H]]	16.2393	22.9282	447.301	326
[[E]]	46.4377	55.1125	1232.75	27
[[Q]]	36.6227	47.2489	133.769	238
[[J]]	2.67055	2.67055	25.7842	2
[[K]]	0.0	0.0	0.0	1
overall(mean)	31.3298	37.1198	11359.0	18
overall(aggregated)				

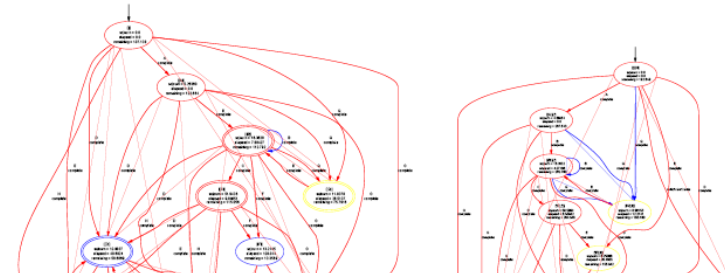
Cross validation: training set and test set



Some results



(a) Set abstraction based on all activities



Abstraction	MAE	RMSE	MAPE
Set abstraction based on all activities (cf. Figure 17(a))	41.648	47.513	1505.07
Set abstraction based on last activity (cf. Figure 17(b))	43.080	49.666	1818.49
Set abstraction based on last activity and additional information related to the occurrence of “I” (cf. Figure 17(c))	17.129	23.550	900.07
Complete abstraction (cf. Figure 17(d))	63.391	74.965	7169.55
Simple heuristic: half of average total flow time (53.57 days)	61.750	75.505	6188.04

Process mining in Collaborative Networks: An Outlook



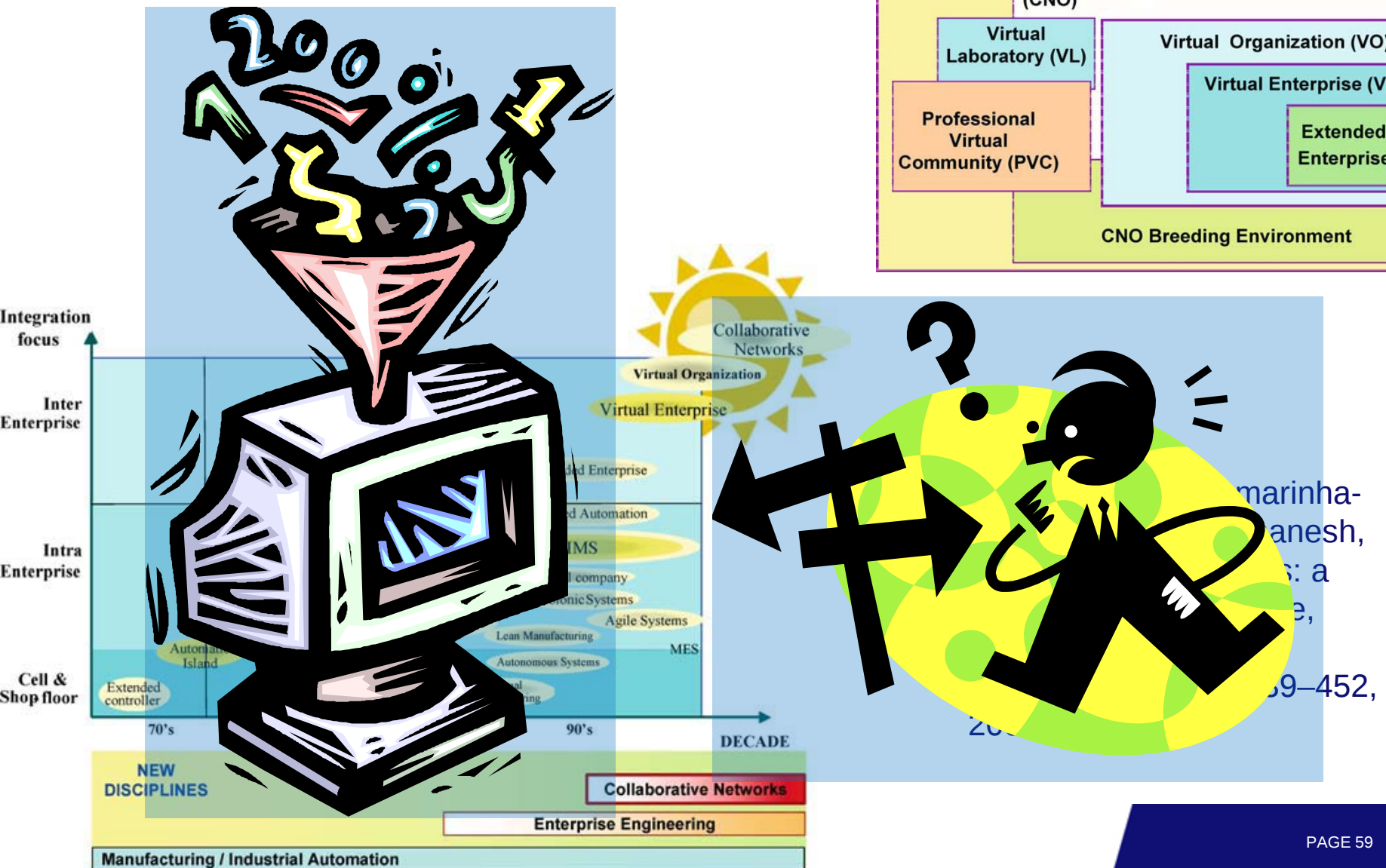
TU **e**

Technische Universiteit
Eindhoven
University of Technology

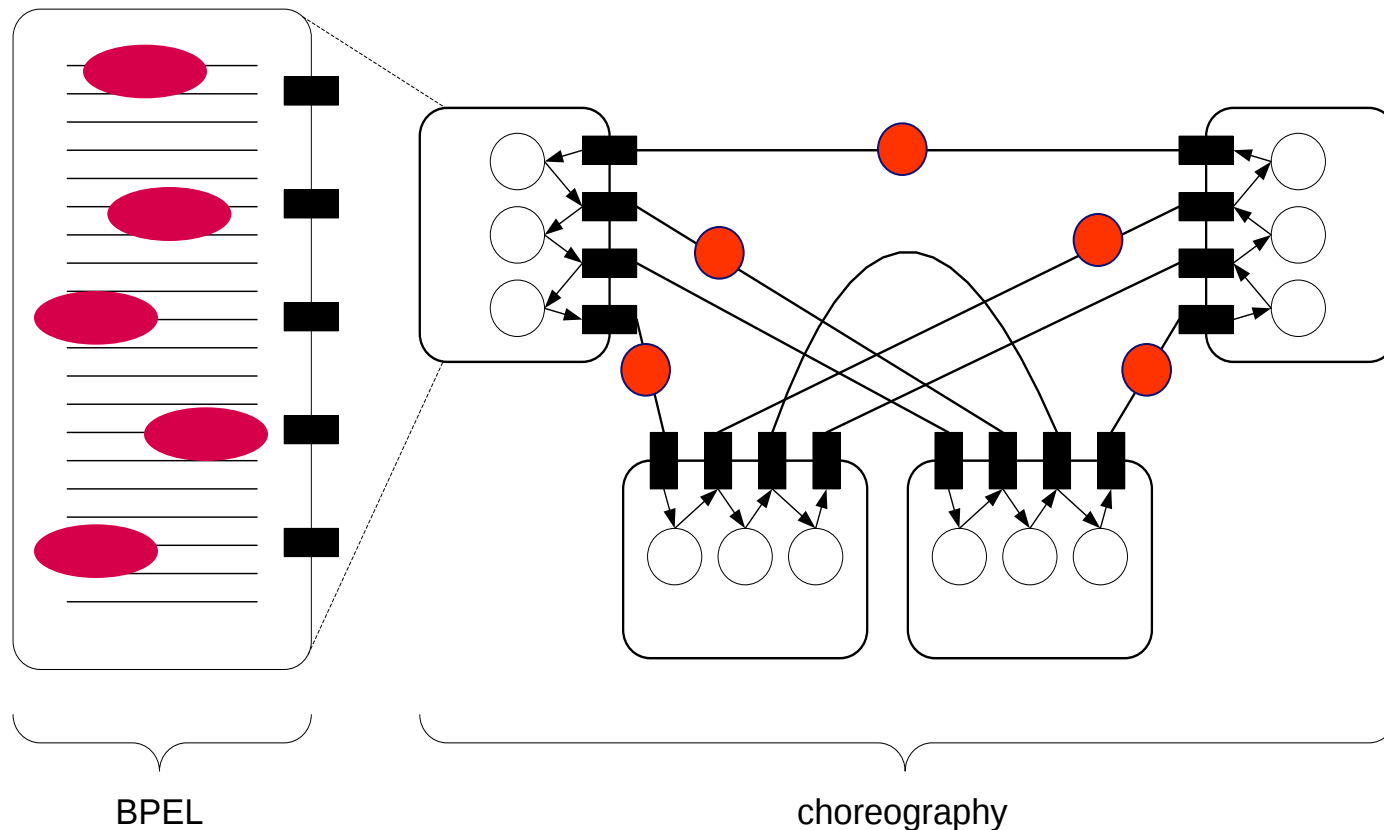
Where innovation starts



Collaborative Networks and Process Mining?



Technological Perspective: Process mining and SOA



Wat vinden wij over

Mijn groepen

Expertpro

Personality Analyzer



Users

Not Individuals

Details (table)



Groups

Cluster 0 (# 41)
Cluster 1 (# 19)
Cluster 2 (# 173)
Cluster 3 (# 24)
Cluster 4 (# 101)
Cluster 5 (# 20)
Cluster 6 (# 33)
Cluster 7 (# 32)
Cluster 8 (# 97)
Cluster 9 (# 14)

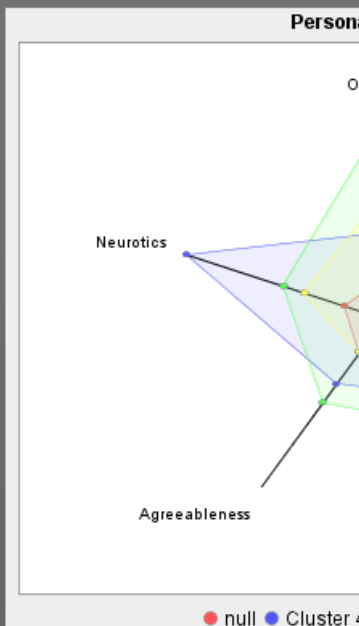
Cluster 6

Neurotics

Agreeableness

● null ● Cluster 4

Persona



User	Openness	Conscientiousness	Extraversion	Agreeableness	Neurotics
U_1	25,0 %	50,0 %	33,2 %	8,2 %	37,5 %

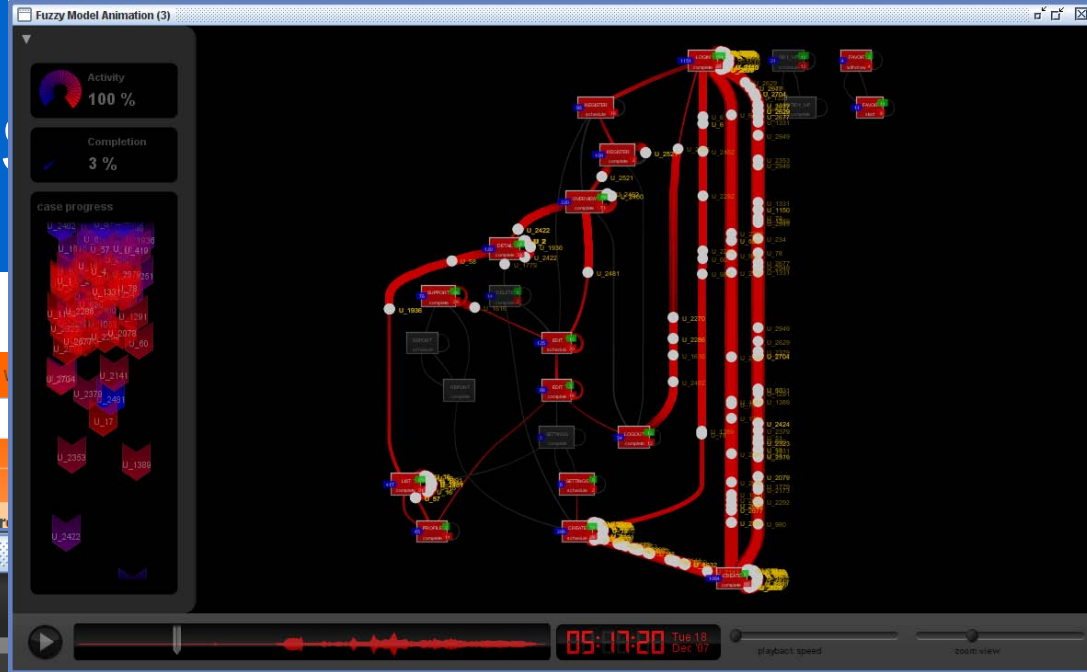
Openness	Hava A Rich Vocabulary Full of Ideas	Excellent Ideas
	25,0 %	50,0 %
		0,0 %

Conscientiousness	Order And Regularity	Pay Attention to Details	Tidy Up
	75,0 %	50,0 %	25,0 %

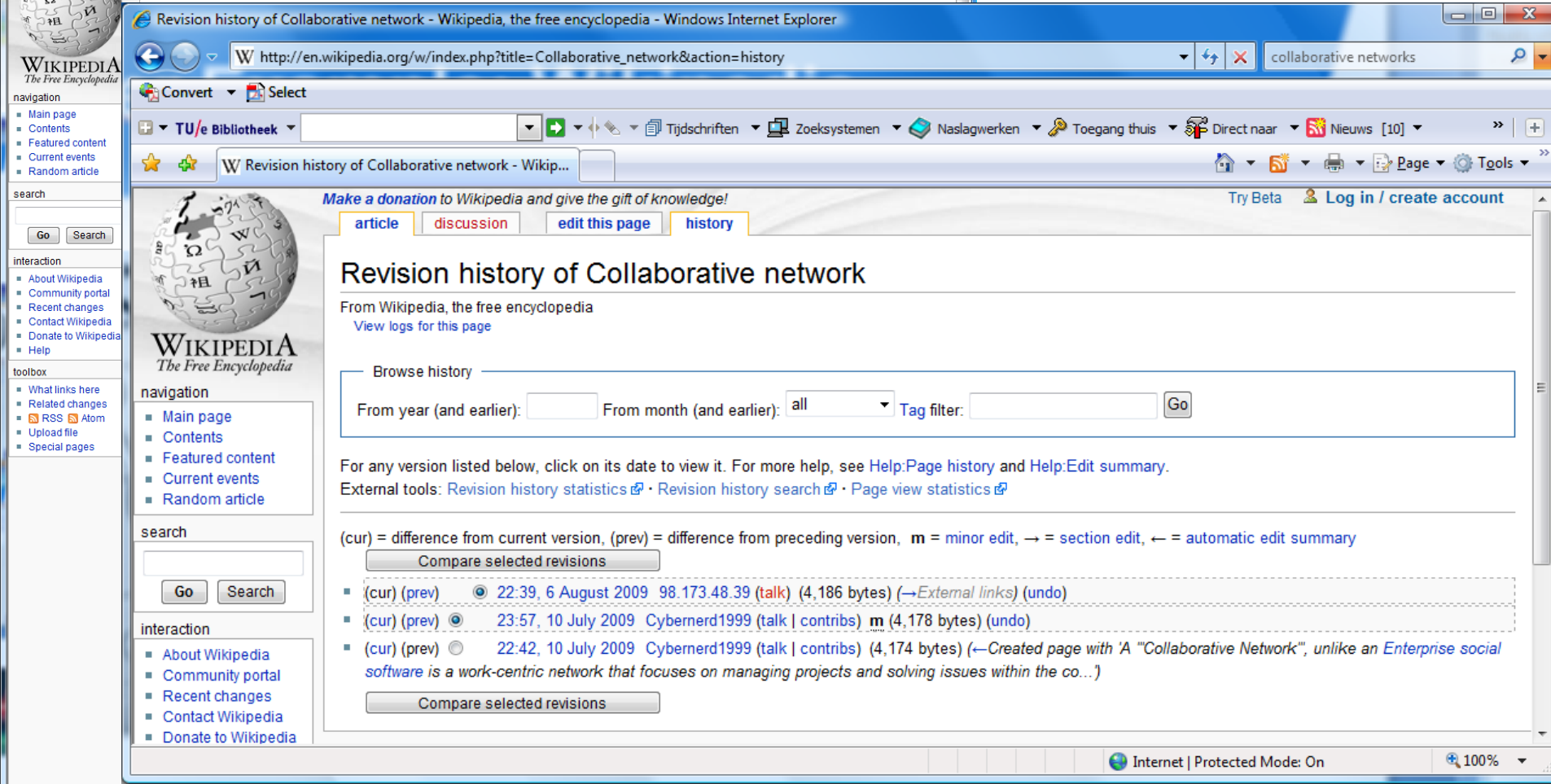
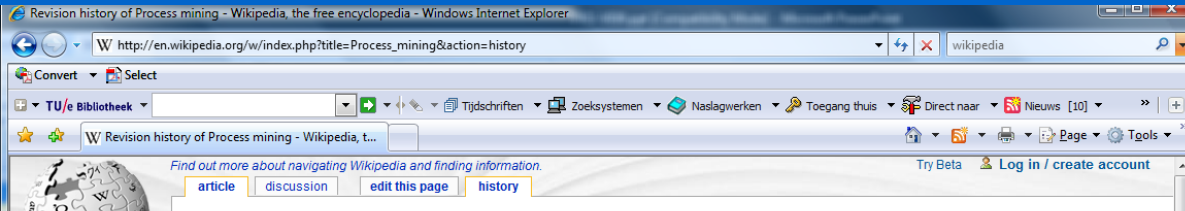
Extraversion	Moderating Groups	Public Person	Start Conversations
	50,0 %	25,0 %	25,0 %

Agreeableness	Interested in Others	On Good Terms	Show Gratitude
	25,0 %	0,0 %	0,0 %

Neuroticism	Irritated Easily	Overwhelmed by Emotions
	0.0 %	75.0 %



Another Example: Wikipedia



Another Example: LinkedIn



Discussions | LinkedIn - Windows Internet Explorer

http://www.linkedin.com/groupAnswers?viewQuestions=&gid=1915049&foru

File Edit View Favorites Tools Help

Convert Select

TU/e Bibliotheek

Tijdschriften

Discussions | LinkedIn

LinkedIn People Jobs Answers Companies

Account & Settings | Help | Sign Out | Language

Explore People Search: Engineer at IBM - Internet - Senior Consultant

Search Inbox

Inbox

Inbox (5) Sent Archived

Compose Message

Received

Archive Mark Read Mark Unread

From	Subject	Status	Date
Chantal de Graaf	Chantal de Graaf invites you to join Dutch Business Process Intelligence Network on LinkedIn		10/04/2009
Mitra Heravizadeh	Join my network on LinkedIn	Accepted	10/01/2009
Dimka Karastoyanova	Join my network on LinkedIn	Accepted	9/30/2009
Peep Kungas	Join my network on LinkedIn	Accepted	9/28/2009
Emilian Pascualu	Invitation to connect on LinkedIn	Accepted	9/24/2009
Paul Hofmann	Join my network on LinkedIn	Accepted	9/22/2009
Wim Nuijten	Join my network on LinkedIn	Accepted	9/15/2009
Rania Khalaf	Join my network on LinkedIn	Accepted	9/15/2009
Michel Reniers	Join my network on LinkedIn	Accepted	9/07/2009
Erik Nooijen	Invitation to connect on LinkedIn	Accepted	9/05/2009
Barbara Weber	Join my network on LinkedIn	Accepted	9/04/2009
Piotr Biernacki	Join my network on LinkedIn	Accepted	8/26/2009
Georg Singer	Invitation to connect on LinkedIn	Accepted	8/24/2009
Genady Grabarnik	Join my network on LinkedIn	Accepted	8/20/2009
Marcello La Rosa	Marcello La Rosa invites you to join Business Process Management research group on LinkedIn		8/12/2009
Marcello La Rosa	Marcello La Rosa invites you to join YAWL on LinkedIn		8/12/2009
W.J.A.M. vdn Heuvel	Join my network on LinkedIn	Accepted	7/23/2009
Gero Decker	Join my network on LinkedIn	Accepted	7/22/2009
Florian Rosenberg	Join my network on LinkedIn	Accepted	7/20/2009

Home Groups Bakkenist Alumni BPM Group BPM Guru See all » Profile Edit My Profile View My Profile Recommendations Contacts Inbox (5) Applications Add Connections

Groups

Process Mining

Overview Discussions News Jobs

Discussions: Recent Act

Featured Discussion

BPTrends 2009 Business

Posted 1 month ago by Paul Harmon Methodologist at BPTrends Associat Follow | Add comment »

Featured Discussion

Process Mining starts with what are their experience

Posted 5 months ago by George Van Mining Following Stop | 8 comments »

ProM 5.2 released

Posted 1 month ago by Christian W. Follow | Move to Jobs | 4 comments

Will van der Aalst

full professor at Eindhoven University of Technology

http://www.linkedin.com/groups?home=&gid=1915049&trk=anet Ug_hm&goback=%2Egdr_1254

Internet | Protected Mode: On 100%

Conclusion



TU/e

Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

Conclusion

- The abundance of event data enables a wide variety of **process mining** techniques ranging from process discovery to conformance checking.
- A **new tool** for analyzing **collaborative networks**.
- Check out **ProM** with its 250+ plug-ins.
- Contribute: case studies, plug-ins, etc.

Thanks!

cf. www.processmining.org

- Wil van der Aalst
- Peter van den Brand
- Boudewijn van Dongen
- Christian Günther
- Eric Verbeek
- Ana Karla Alves de Medeiros
- Anne Rozinat
- Minseok Song
- Ton Weijters
- Remco Dijkman
- Gianluigi Greco
- Antonella Guzzo
- Kristian Bisgaard Lassen
- Ronny Mans
- Jan Mendling
- Vladimir Rubin
- Nikola Trcka
- Irene Vanderfeesten
- Barbara Weber
- Lijie Wen
- Mercy Amiyo
- Carmen Bratosin
- Toon Calders
- Jorge Cardoso
- Ronald Crooy
- Florian Gottschalk
- Monique Jansen-Vullers
- Peter Khisa Wakholi
- Nicolas Knaak
- Sven Lambrechts
- Joyce Nakatumba
- Mariska Netjes
- Mykola Pechenizkiy
- Maja Pesic
- Hajo Reijers
- Stefanie Rinderle
- Domenico Saccà
- Helen Schonenberg
- Marc Voorhoeve
- Jianmin Wang
- Jan Martijn van der Werf
- Martin van Wingerden
- Jianhong Ye
- Huub de Beer
- Elena Casares
- Alina Chipaila
- Walid Gaaloul
- Martijn van Giessel
- Shaifali Gupta
- Thomas Hoffmann
- Peter Hornix
- René Kerstjens
- Ralf Kramer
- Wouter Kunst
- Laura Maruster
- Andriy Nikolov
- Adarsh Ramesh
- Jo Theunissen
- Kenny van Uden
- ...

Relevant WWW sites



- <http://www.processmining.org>
- <http://promimport.sourceforge.net>
- <http://prom.sourceforge.net>
- <http://www.workflowpatterns.com>
- <http://www.workflowcourse.com>
- <http://www.vdaalst.com>

“Not everything that counts can be counted, and not everything that can be counted counts” (Einstein)



<http://www.senternovem.nl/innovatievouchers>
MKB 2.500 – 7.500 euro

