TU/e technische universiteit eindhoven

Process Mining in the Context  of Web Services

Prof.dr.ir. Wil van der Aalst

Eindhoven University of Technology,

P.O. Box 513, 5600 MB Eindhoven, The Netherlands

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

Outline

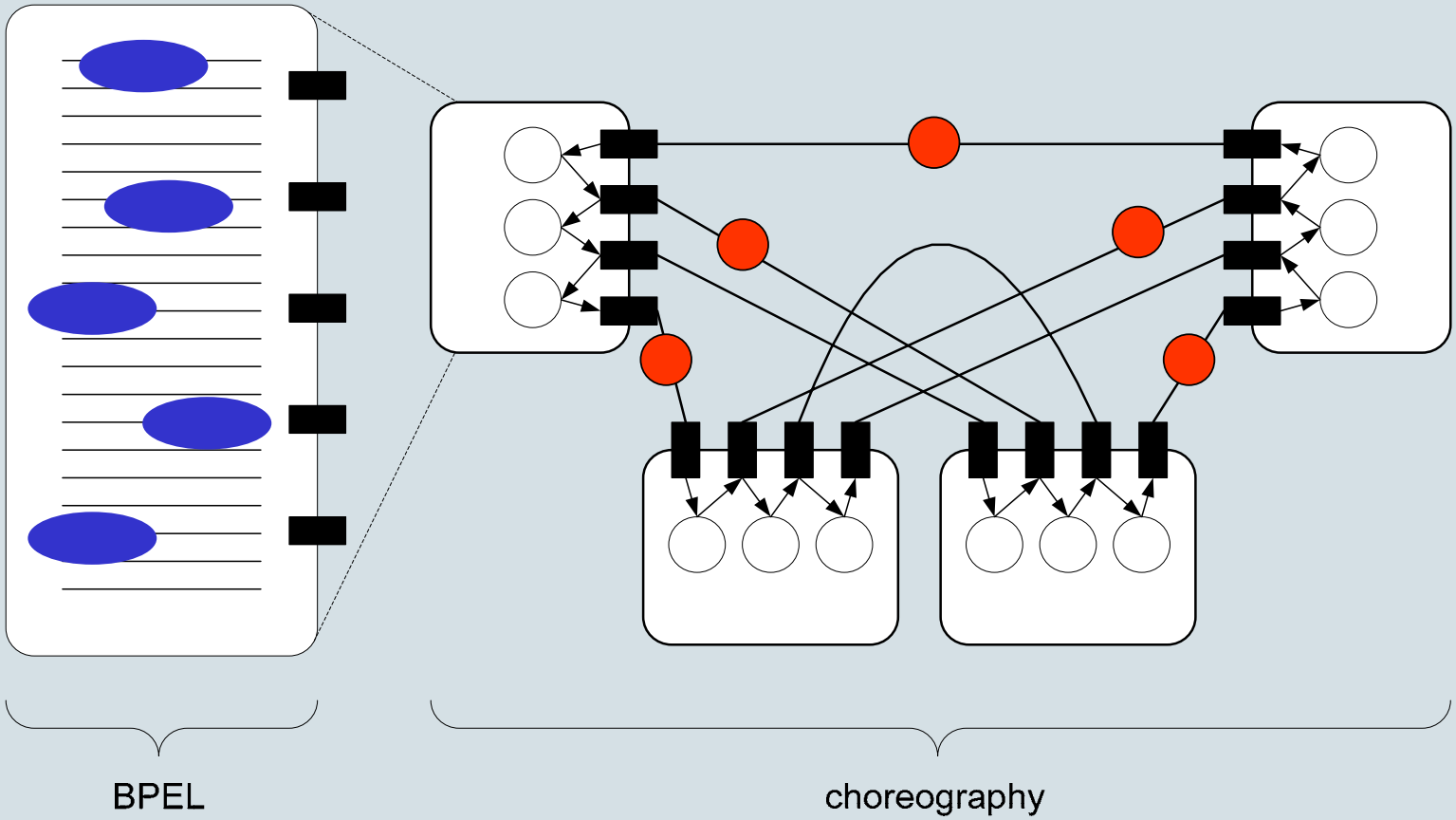
- Web services monitoring
- Process Mining
- Running example
- Discovery
- Conformance checking
- Reality Check
- Conclusion

The work of many people!

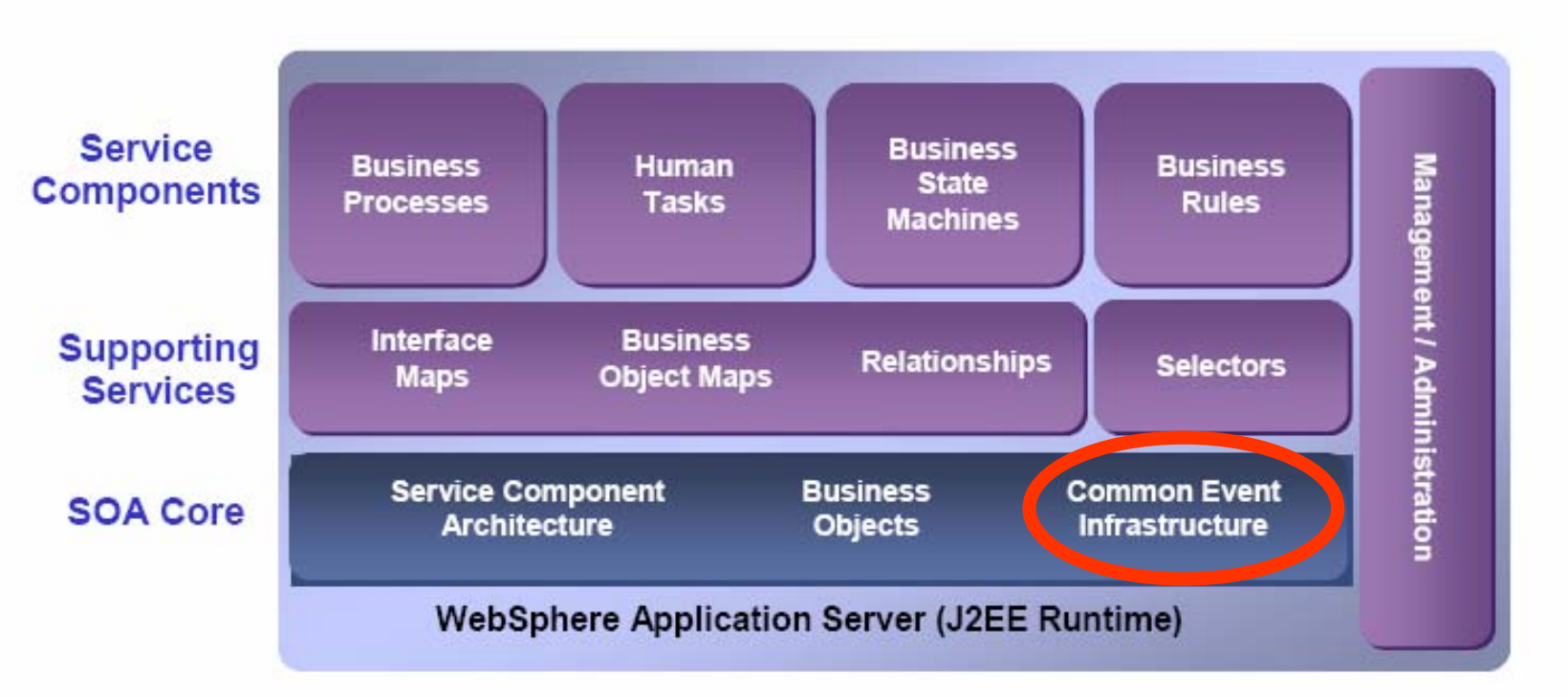
Thanks to Ton Weijters, Boudewijn van Dongen, Ana Karla Alves de Medeiros, ***Anne Rozinat****, Christian Günter,* ***Eric Verbeek****, Ronny Mans, Minseok Song, Laura Maruster, Huub de Beer, Peter van den Brand, Jan Mendling, Andriy Nikolov, Jianmin Wang, Lijie Wen, Irene Vanderfeesten, Mariska Netjes, Steffi Rinderle, Walid Gaaloul, Gianluigi Greco, Antonella Guzzo, etc. etc.*

Web Services Monitoring

Setting: Services, composition, and choreography



Example: IBM's *WebSphere Process Server architecture*

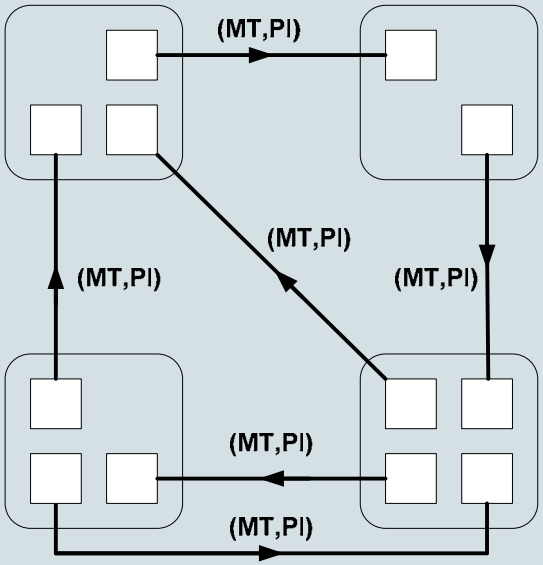


- **Common event infrastructure (CEI)**

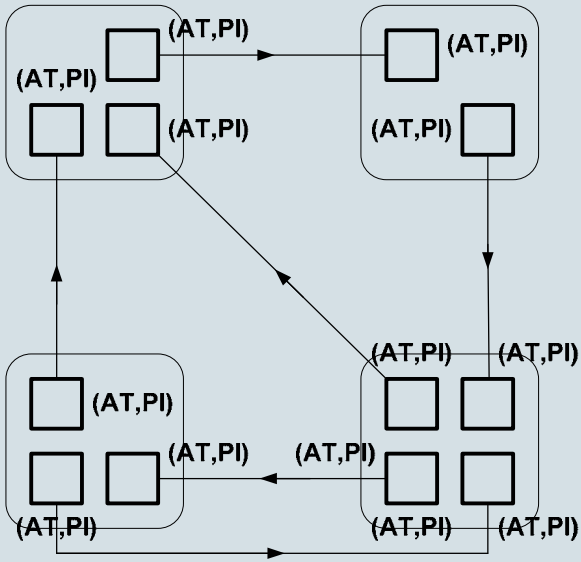
Logging events

- local/global
- messages/activities

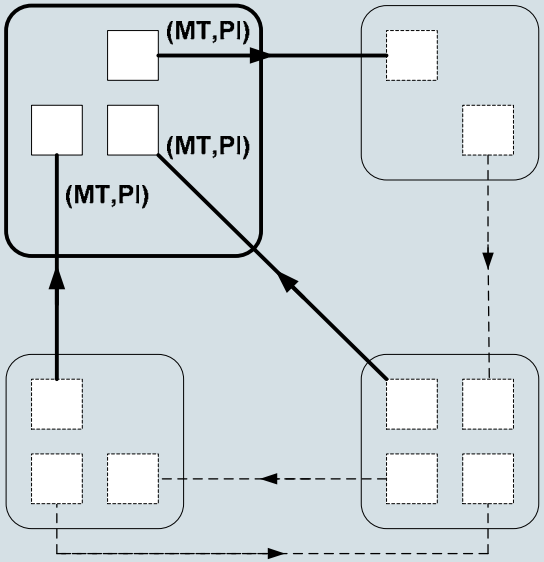
Services use BPEL or not, may have a model or not, are known or not, and may deviate from what is expected or not.



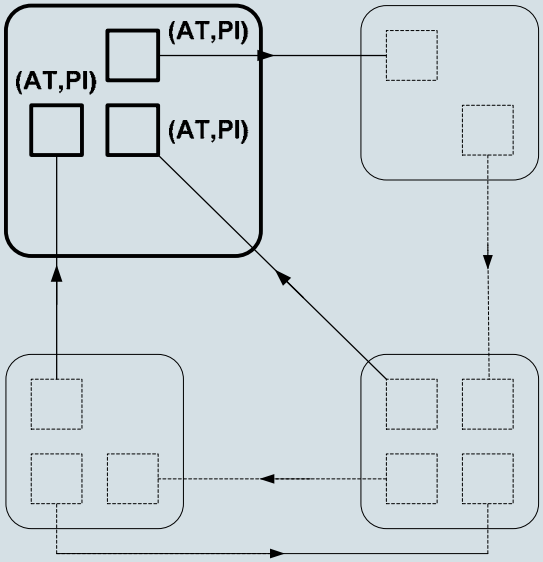
(a) Global message observer



(b) Global activity observer



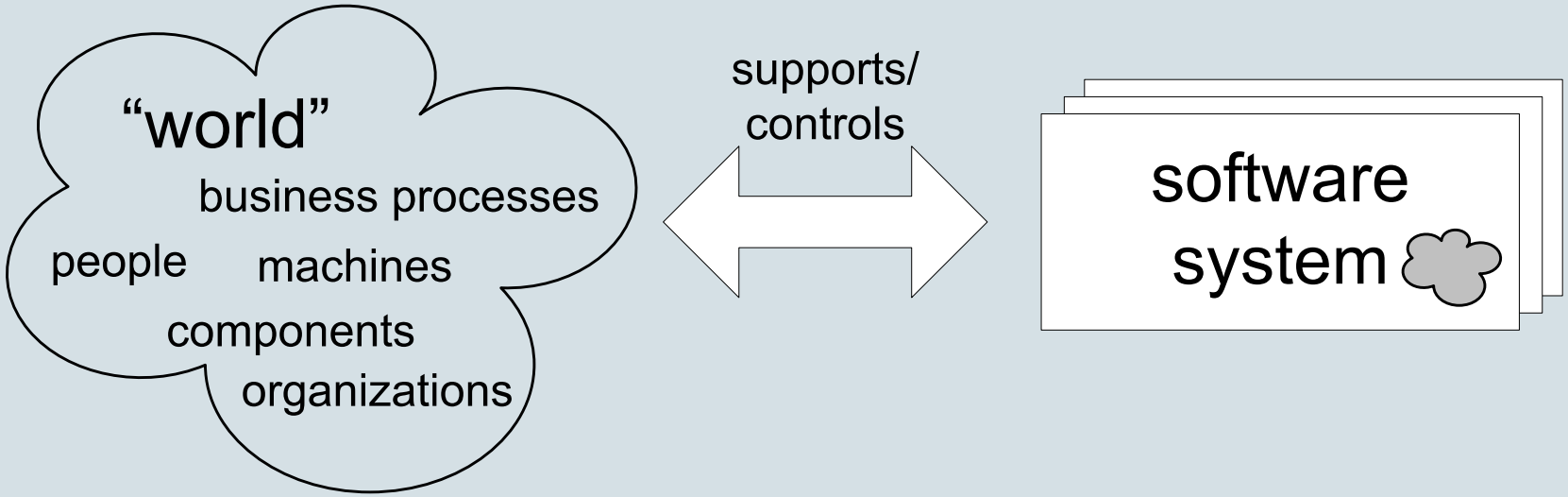
(c) Local message observer



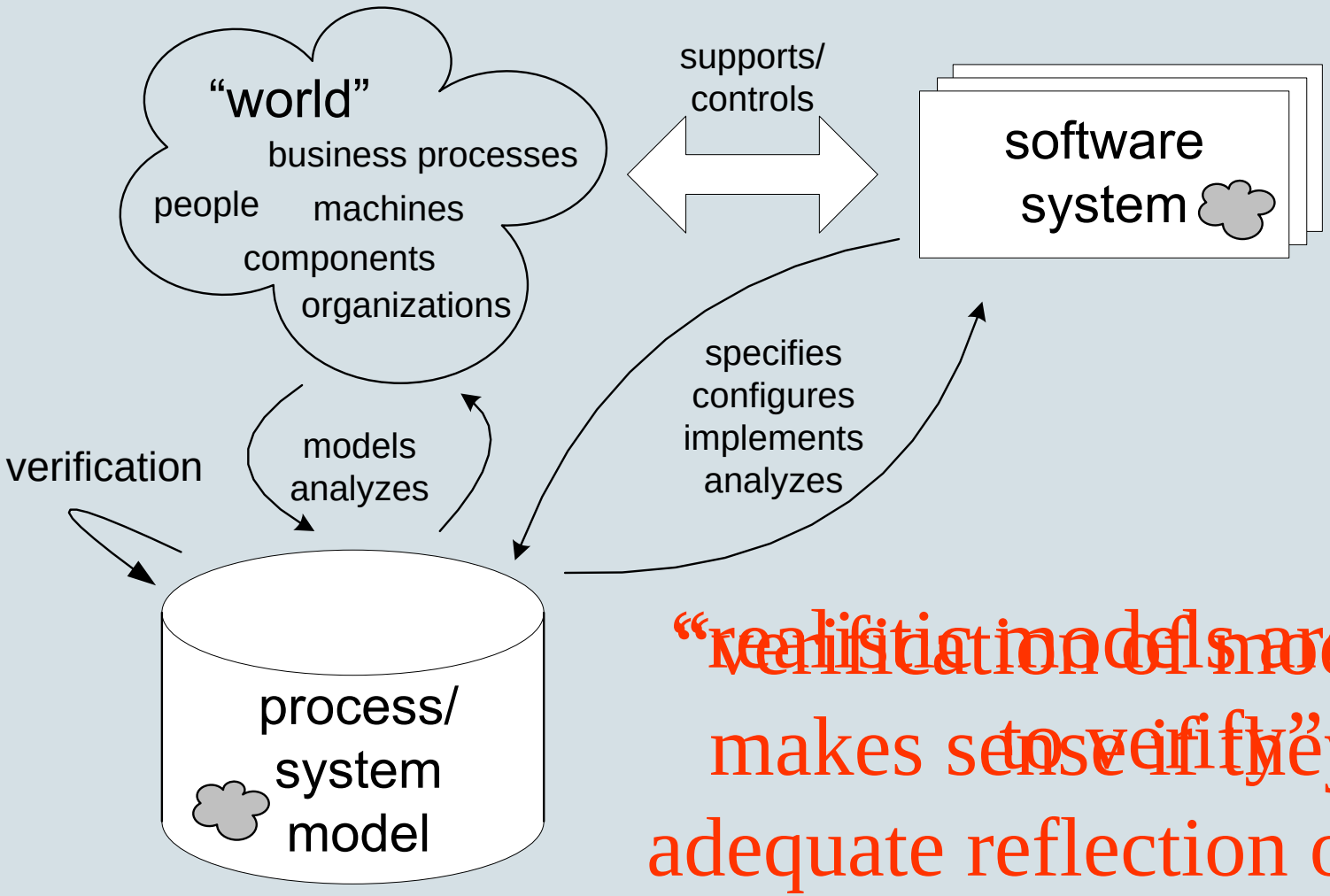
(b) Local activity observer

Overview Process Mining

Software systems are the mirror image of the “world”



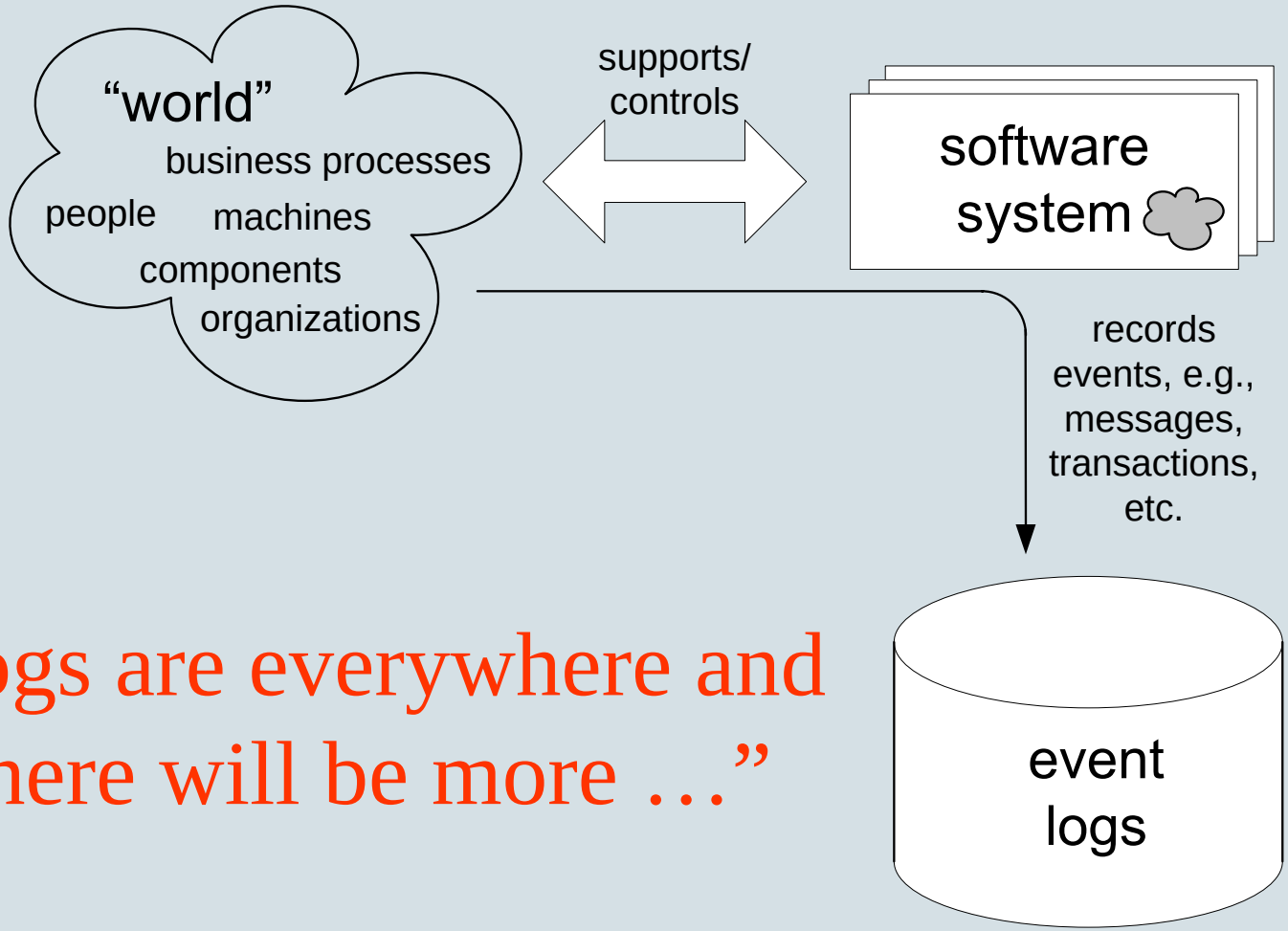
Dual role of process models



"verification models are difficult to verify"

"verification models are difficult makes sense if they are an adequate reflection of reality"

Event logs are a reflection of reality



"logs are everywhere and there will be more …"

Examples:

staffware. THE POWER OF PROCESS

IBM

océ

FileNet An IBM Company

Pallas Athena PROCESS MANAGEMENT

PHILIPS sense and simplicity

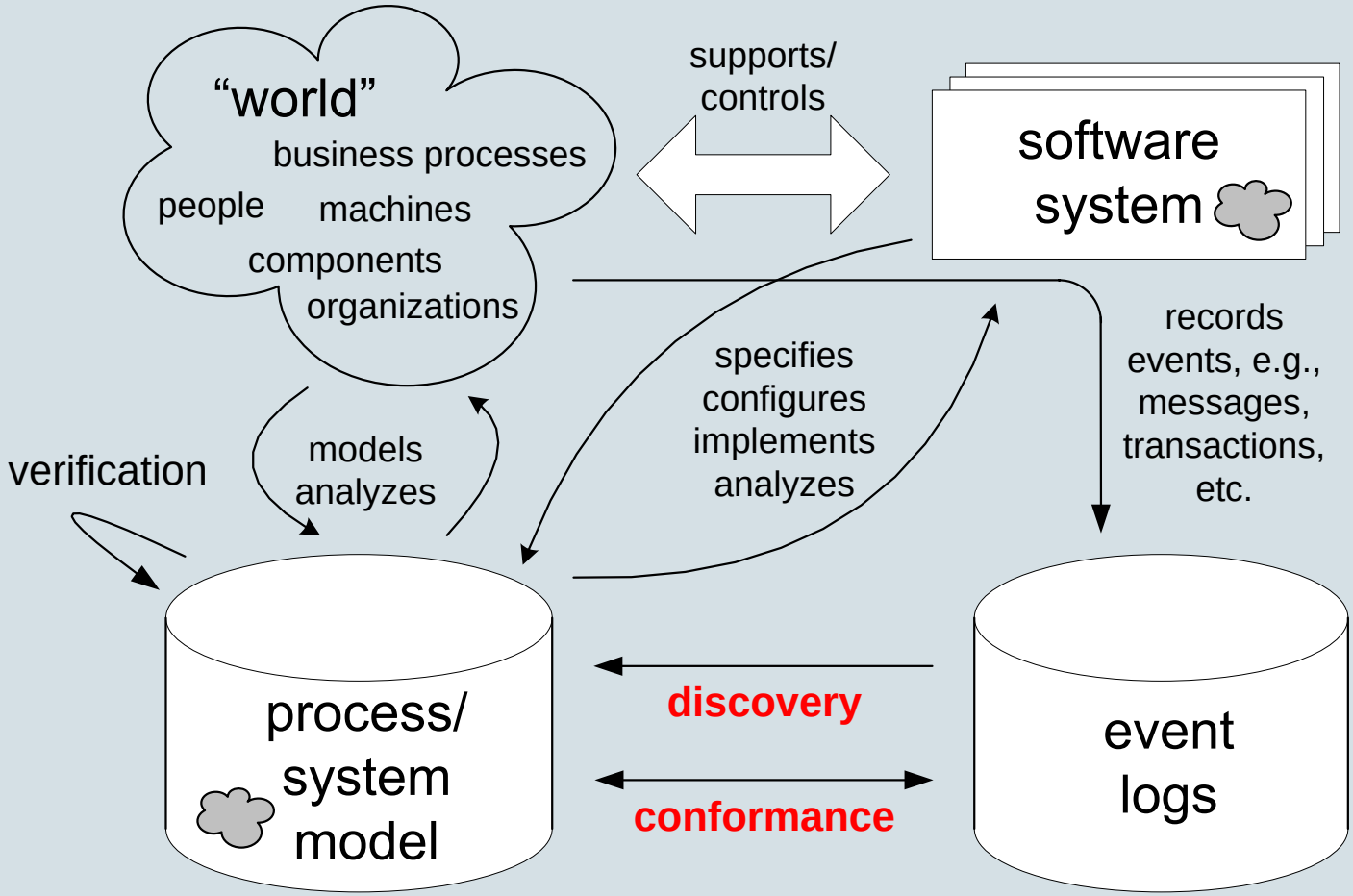
ORACLE

SAP

ASML

WebSphere. software

Process mining: Linking events to models



Toy example to explain basic idea:

Reviewing of papers for IPA workshop

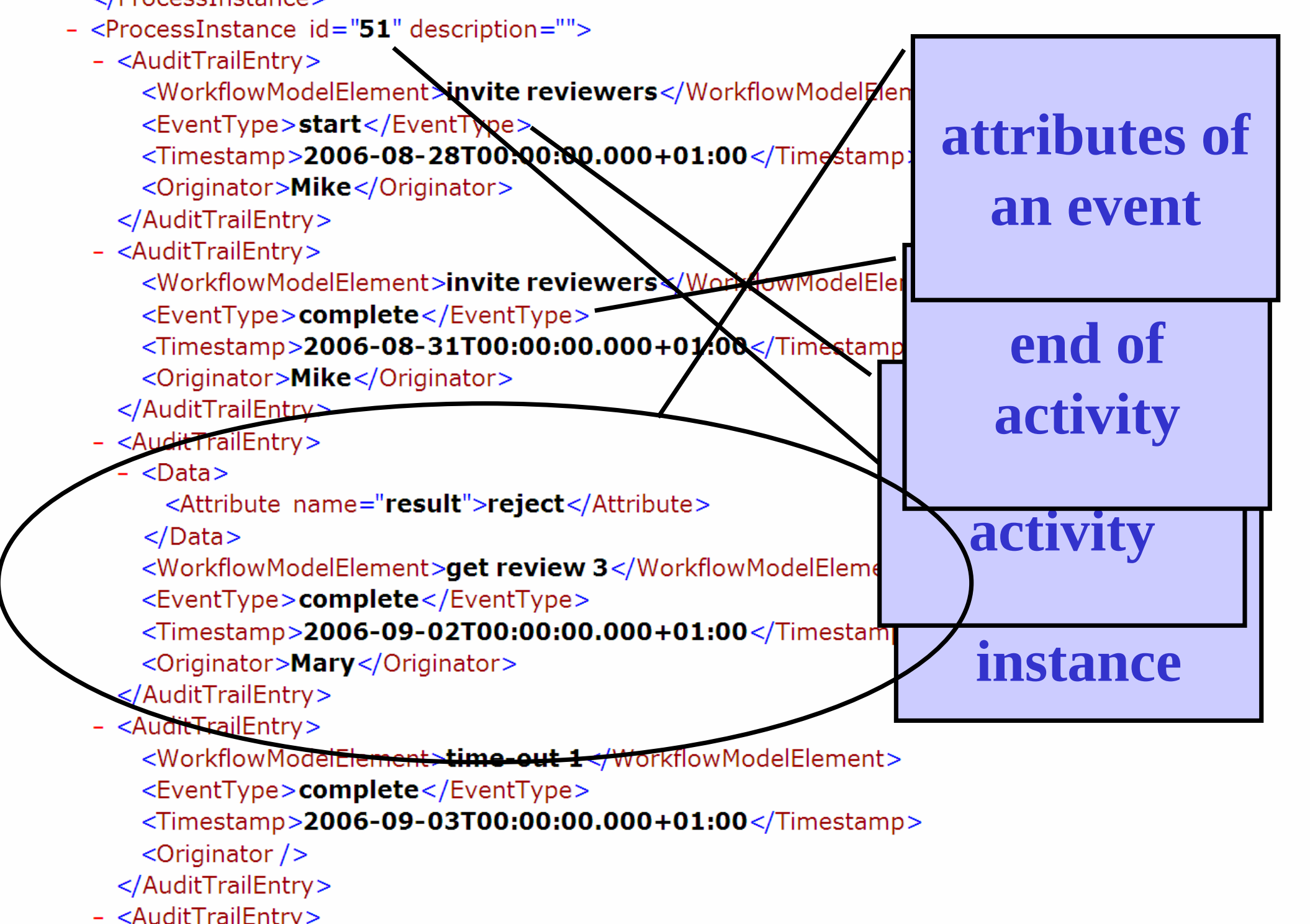
☺

Event log:

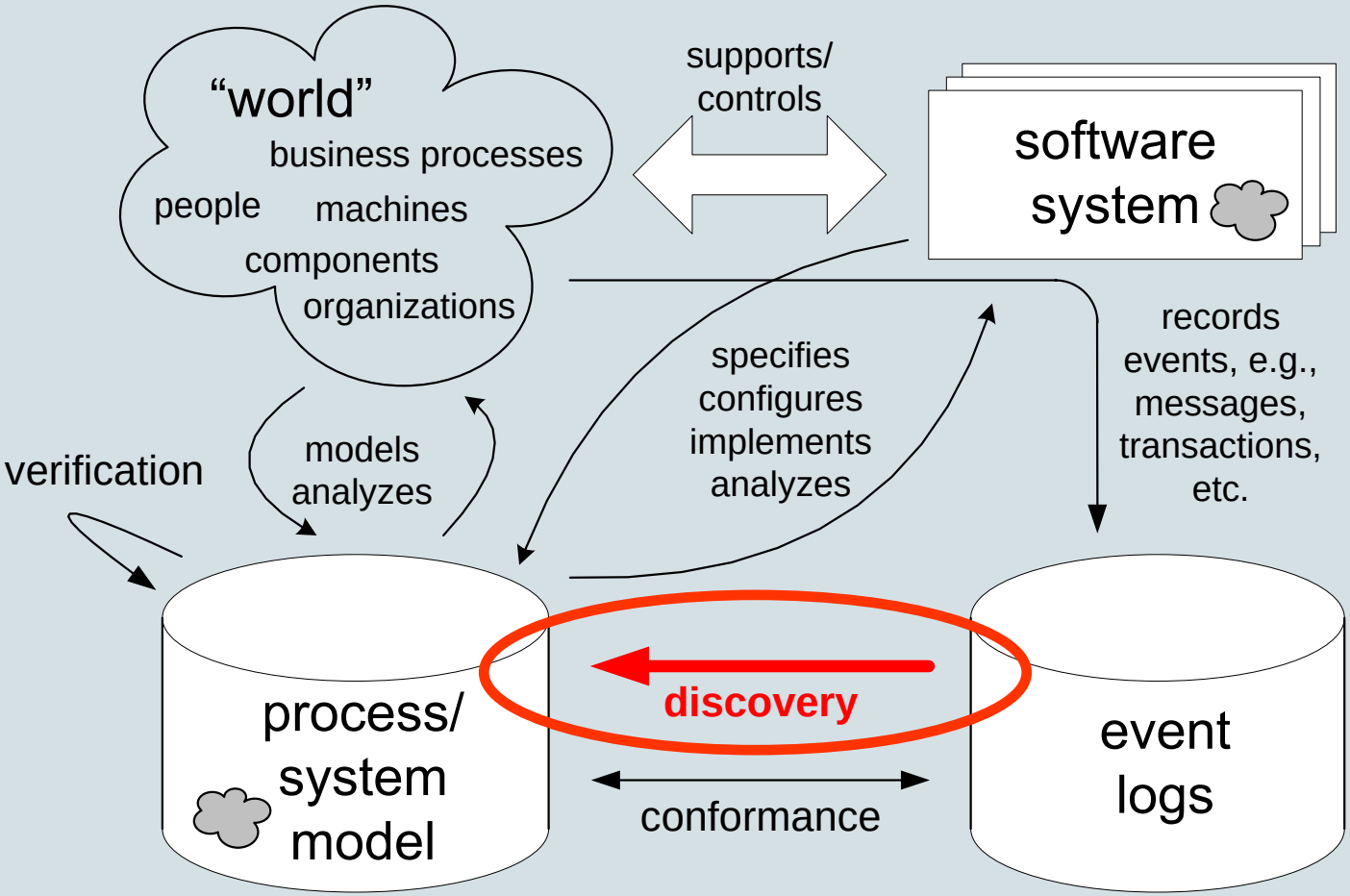
- processes
  - process instances
    - events

Per event:

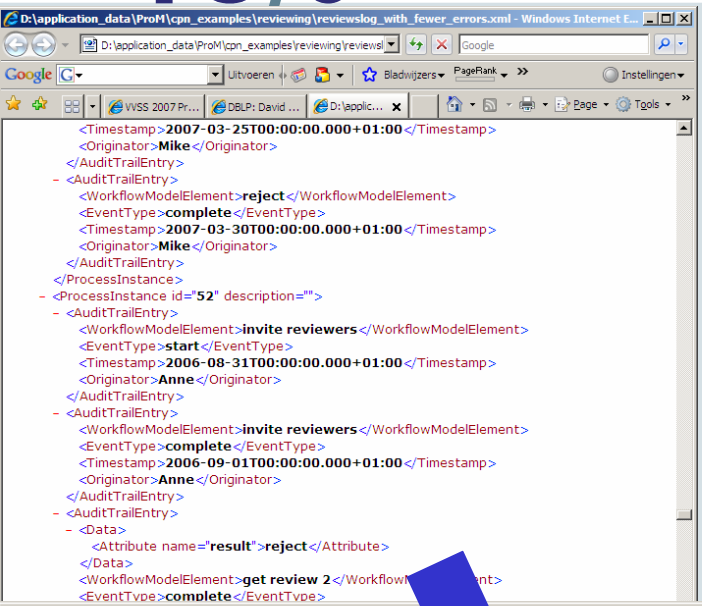
- activity name
- (event type)
- (originator)
- (timestamp)
- (data)

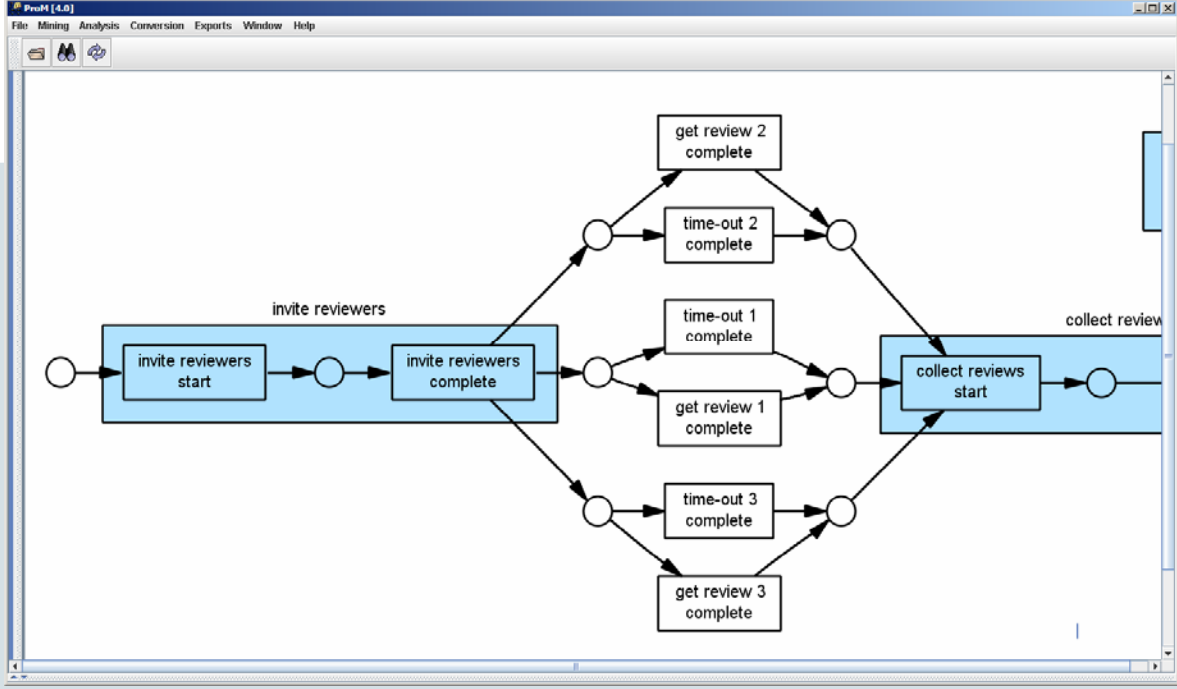

- <ProcessInstance id="51" description="">
- <AuditTrailEntry>
<WorkflowModelElement>invite reviewers</WorkflowModelEle
<EventType>start</EventType>
<Timestamp>2006-08-28T00:00:00.000+01:00</Timestamp
<Originator>Mike</Originator>
</AuditTrailEntry>
- <AuditTrailEntry>
<WorkflowModelElement>invite reviewers</WorkflowModelEle
<EventType>complete</EventType>
<Timestamp>2006-08-31T00:00:00.000+01:00</Timestamp
<Originator>Mike</Originator>
</AuditTrailEntry>
- <AuditTrailEntry>
- <Data>
<Attribute name="result">reject</Attribute>
</Data>
<WorkflowModelElement>get review 3</WorkflowModelEleme
<EventType>complete</EventType>
<Timestamp>2006-09-02T00:00:00.000+01:00</Timestam
<Originator>Mary</Originator>
</AuditTrailEntry>
- <AuditTrailEntry>
<WorkflowModelElement>time-out 1</WorkflowModelElement>
<EventType>complete</EventType>
<Timestamp>2006-09-03T00:00:00.000+01:00</Timestamp>
<Originator />
</AuditTrailEntry>
- <AuditTrailEntry>
attributes of an event
end of activity
activity
instance

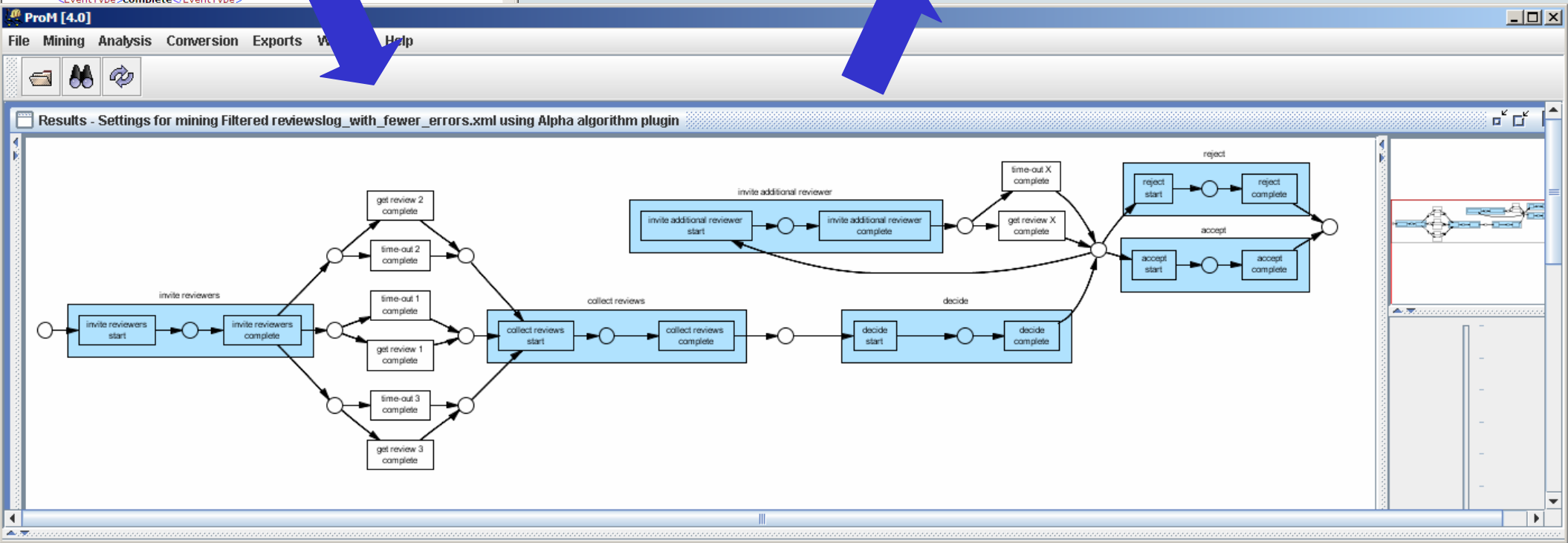
Discovery


"world"
business processes
people
machines
components
organizations
supports/
controls
software
system
records
events, e.g.,
messages,
transactions,
etc.
specifies
configures
implements
analyzes
models
analyzes
verification
process/
system
model
discovery
conformance
event
logs

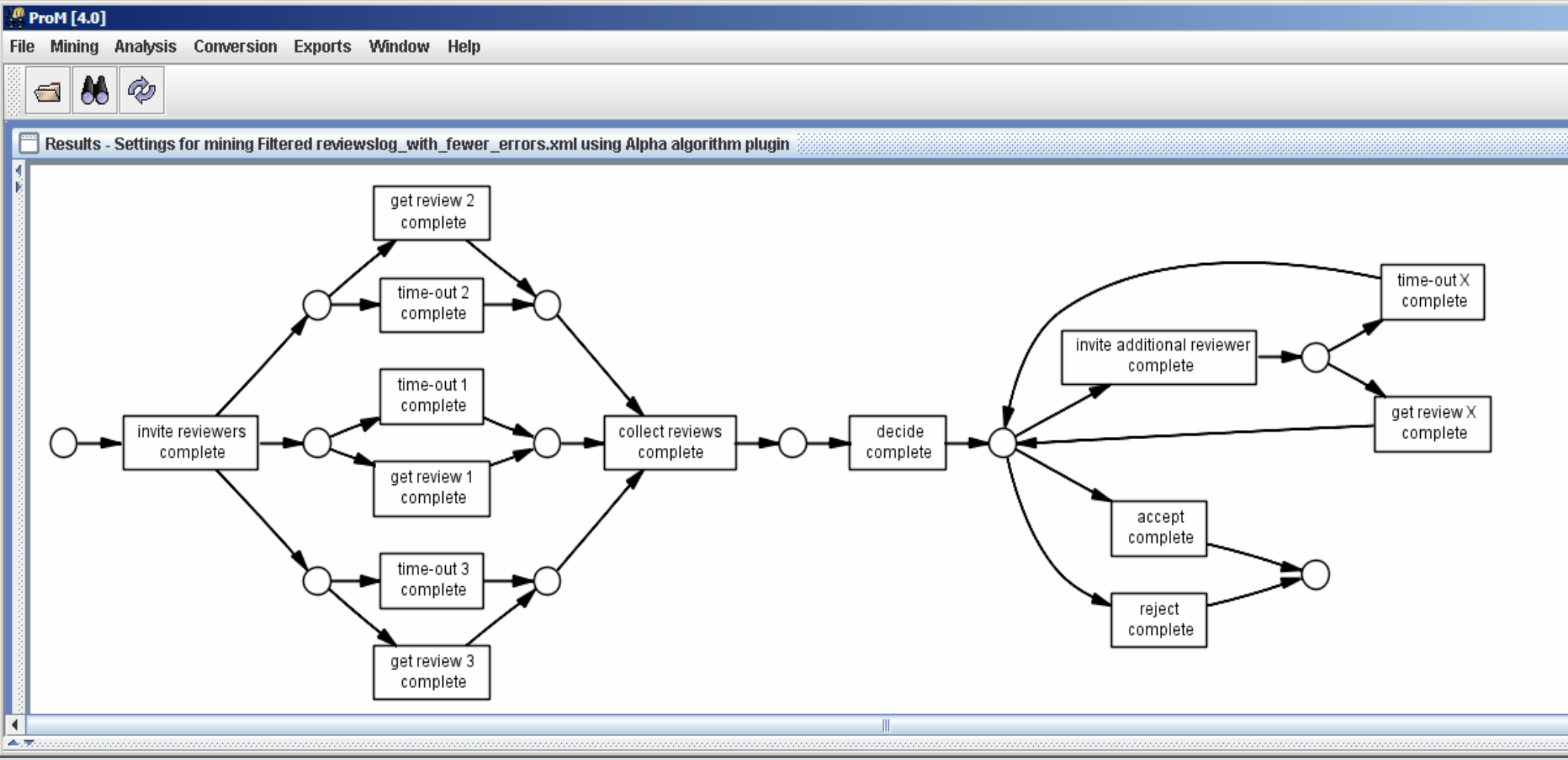
TU/e technische universiteit eindhoven







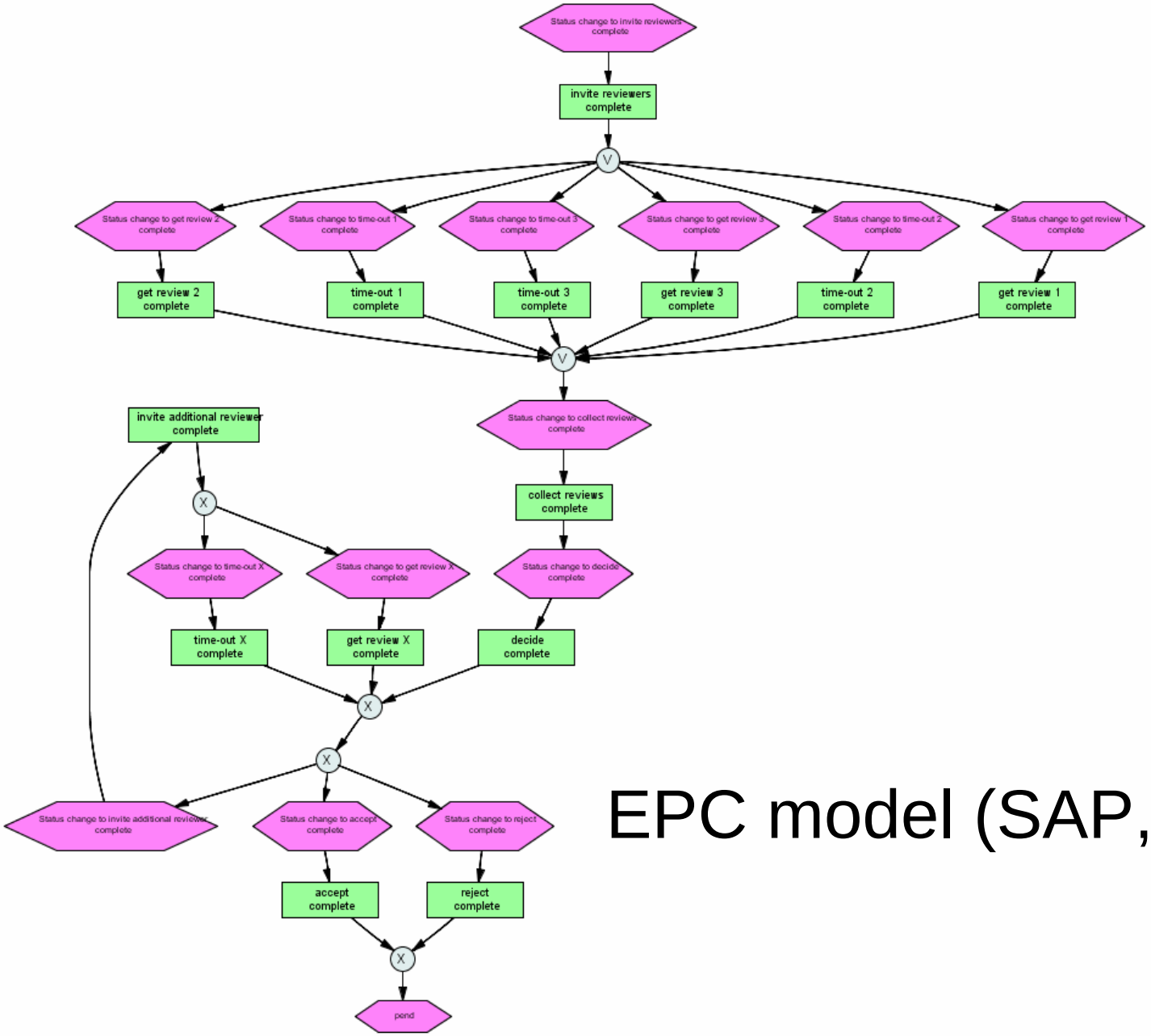
No transactional information



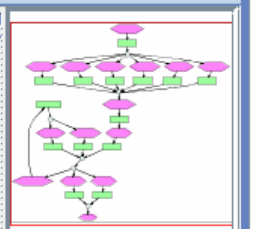
ProM [4.0]

File Mining Analysis Conversion Exports Window Help

Conversion - Labeled WF net to EPC

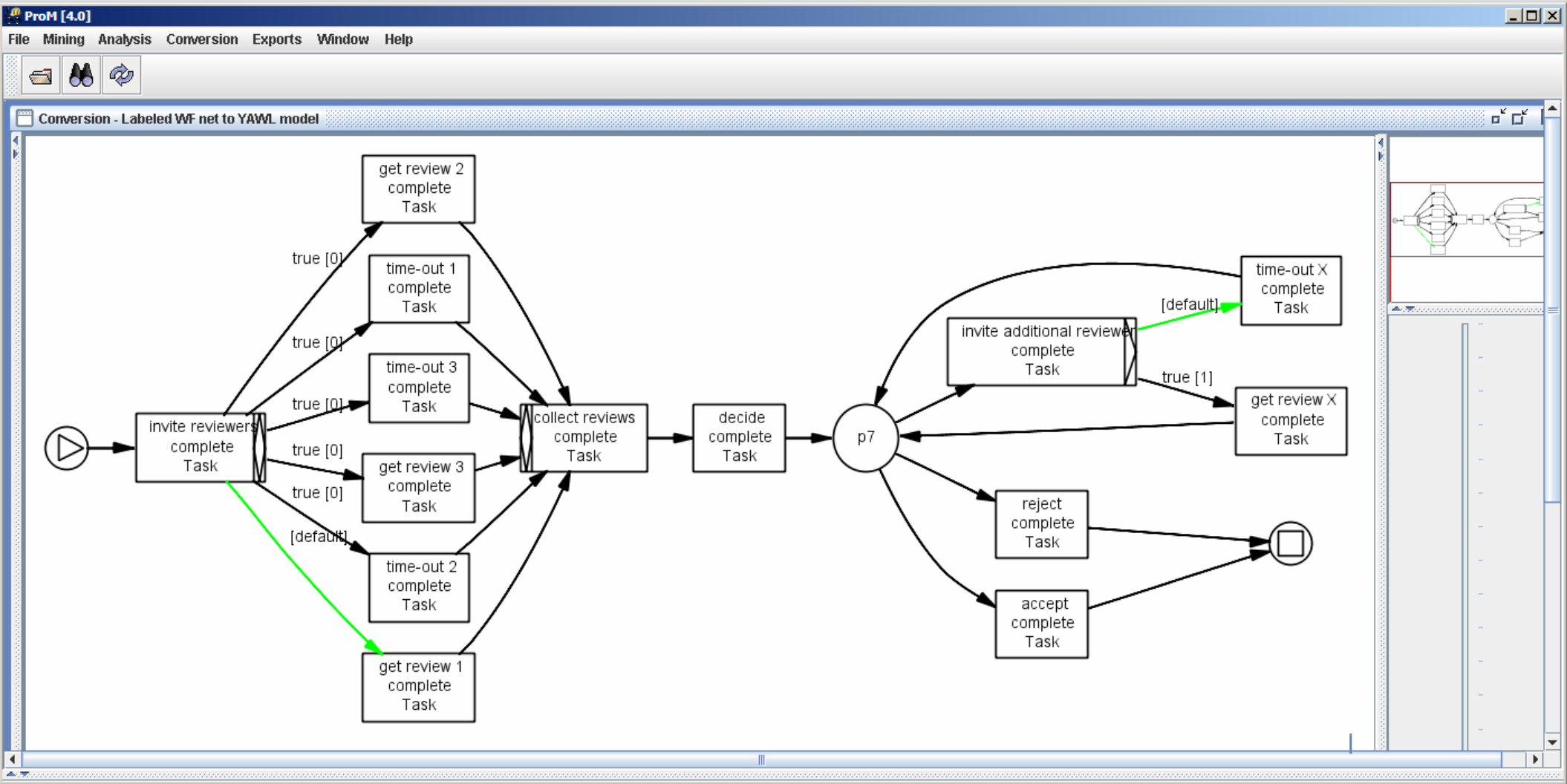


EPC model (SAP,ARIS, etc)



Zoom: 110 %

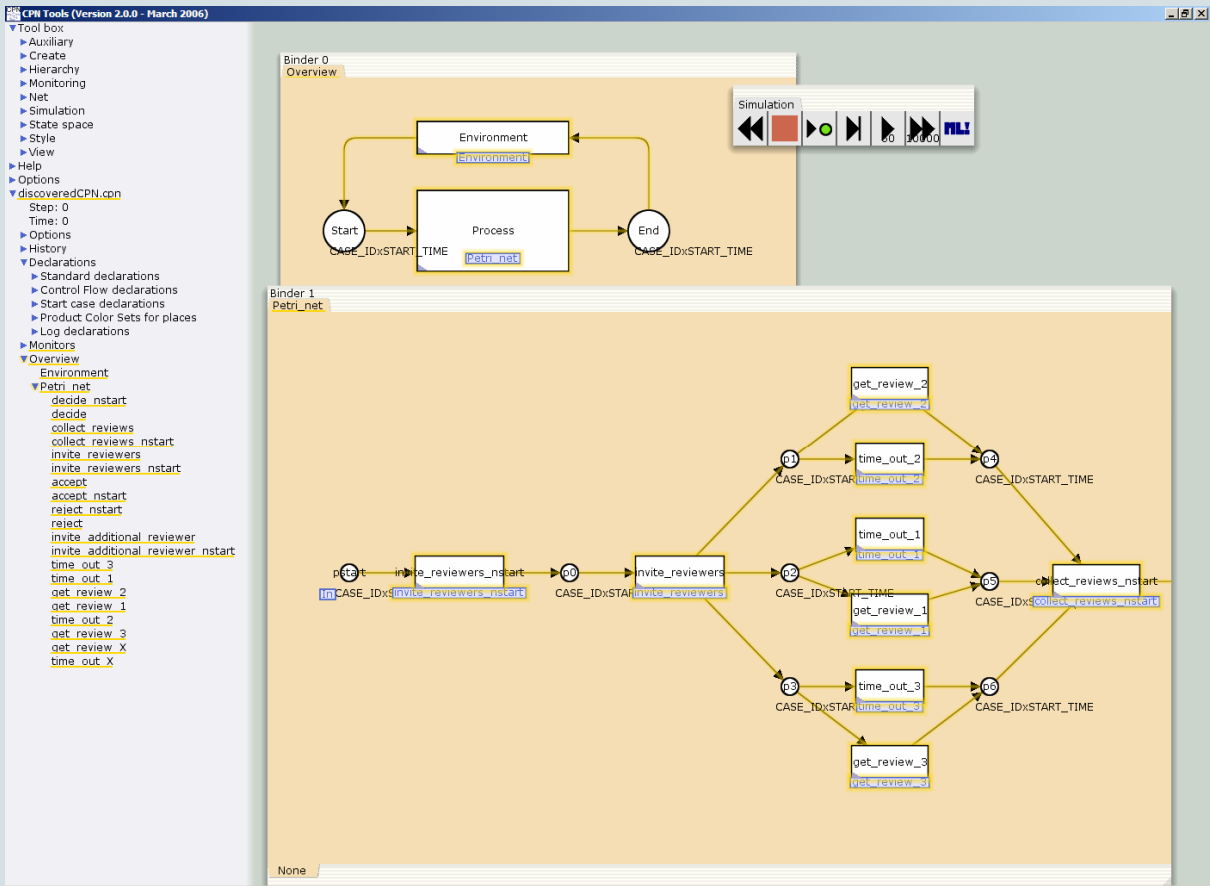
YAWL model (executable workflow model)

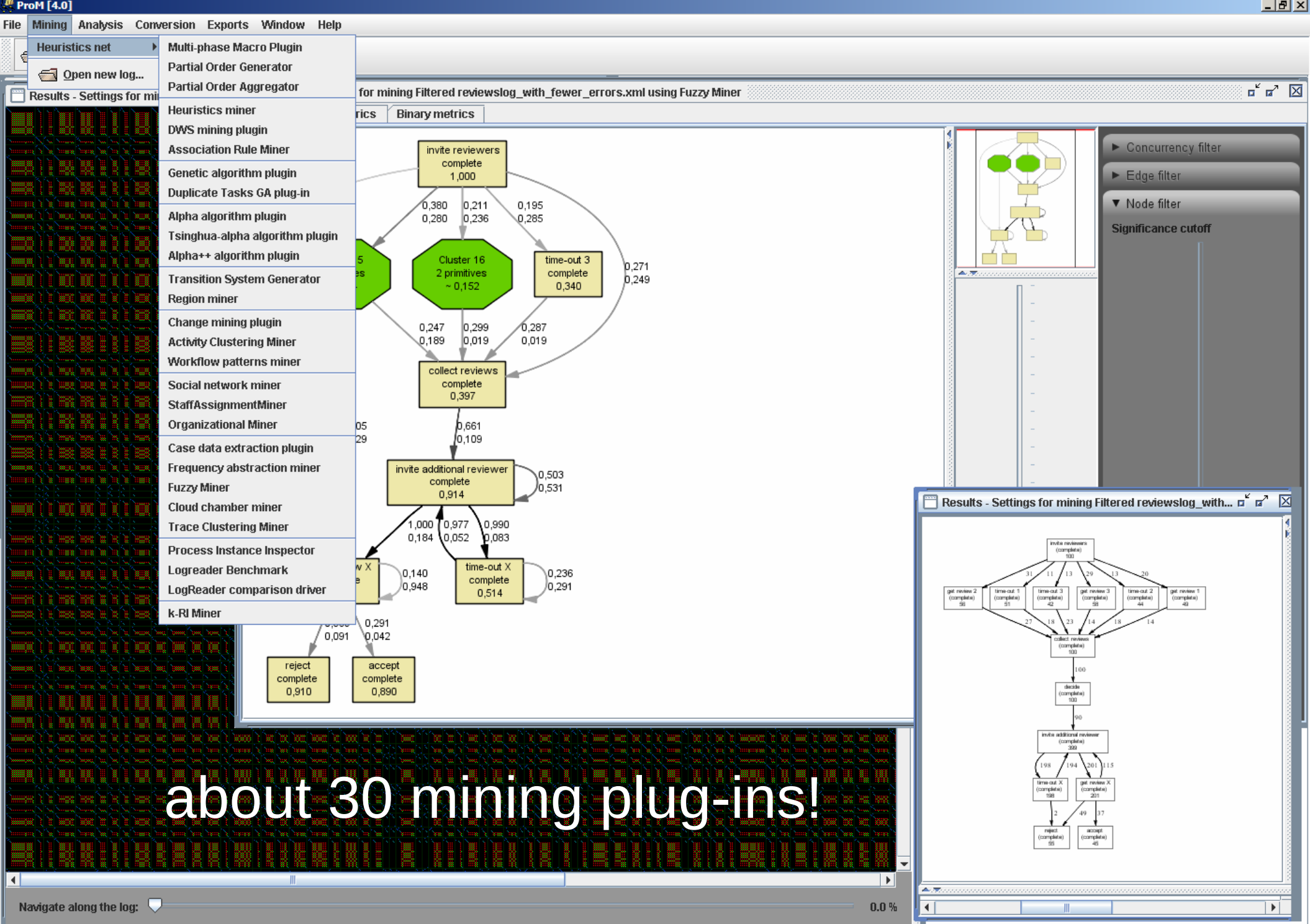


Conversions/exports/imports

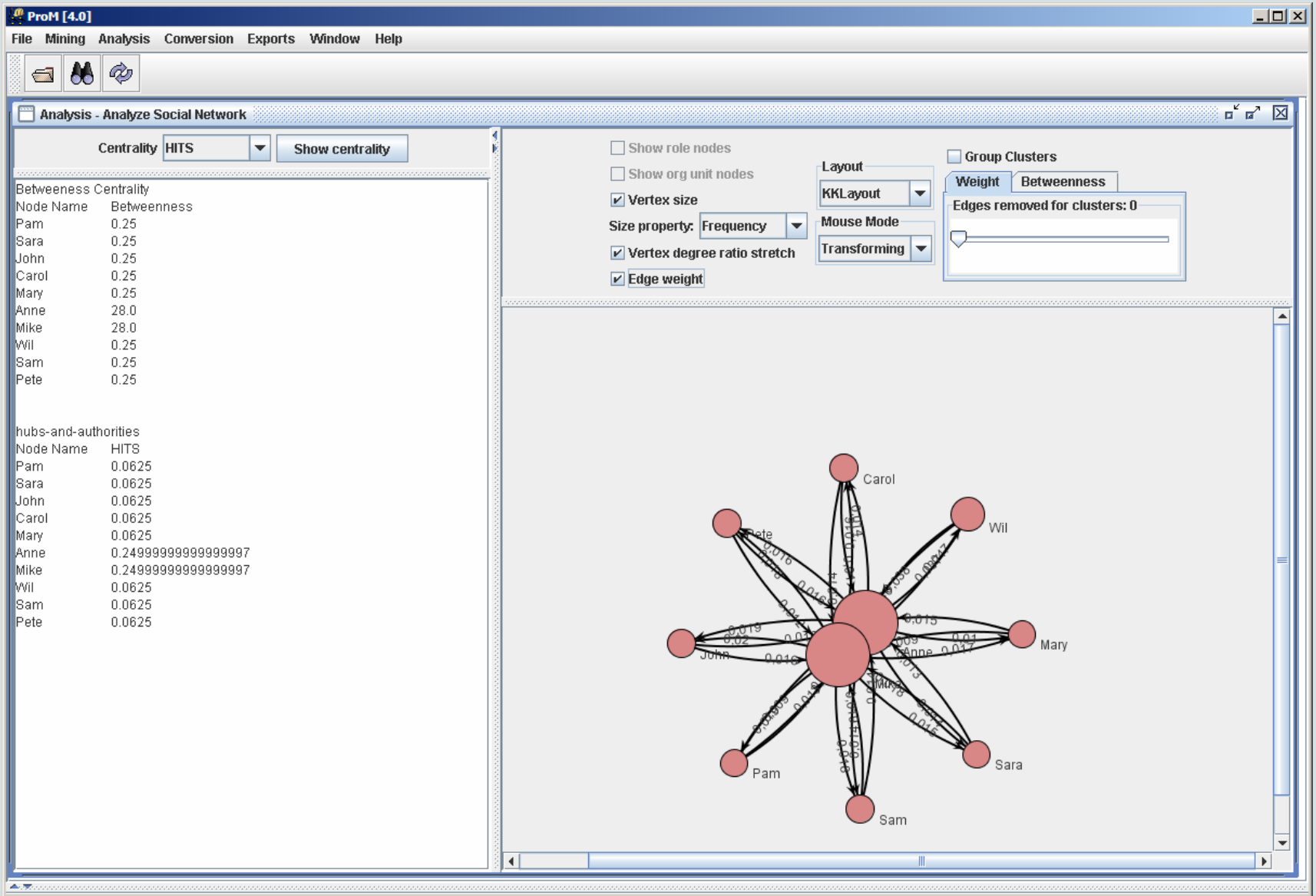
- ARIS – ARIS PPM
- **BPEL 1.1 (WebSphere/Oracle)**
- YAWL
- CPN Tools
- Petrify
- Woflan
- Heuristics nets
- …

link to Eric

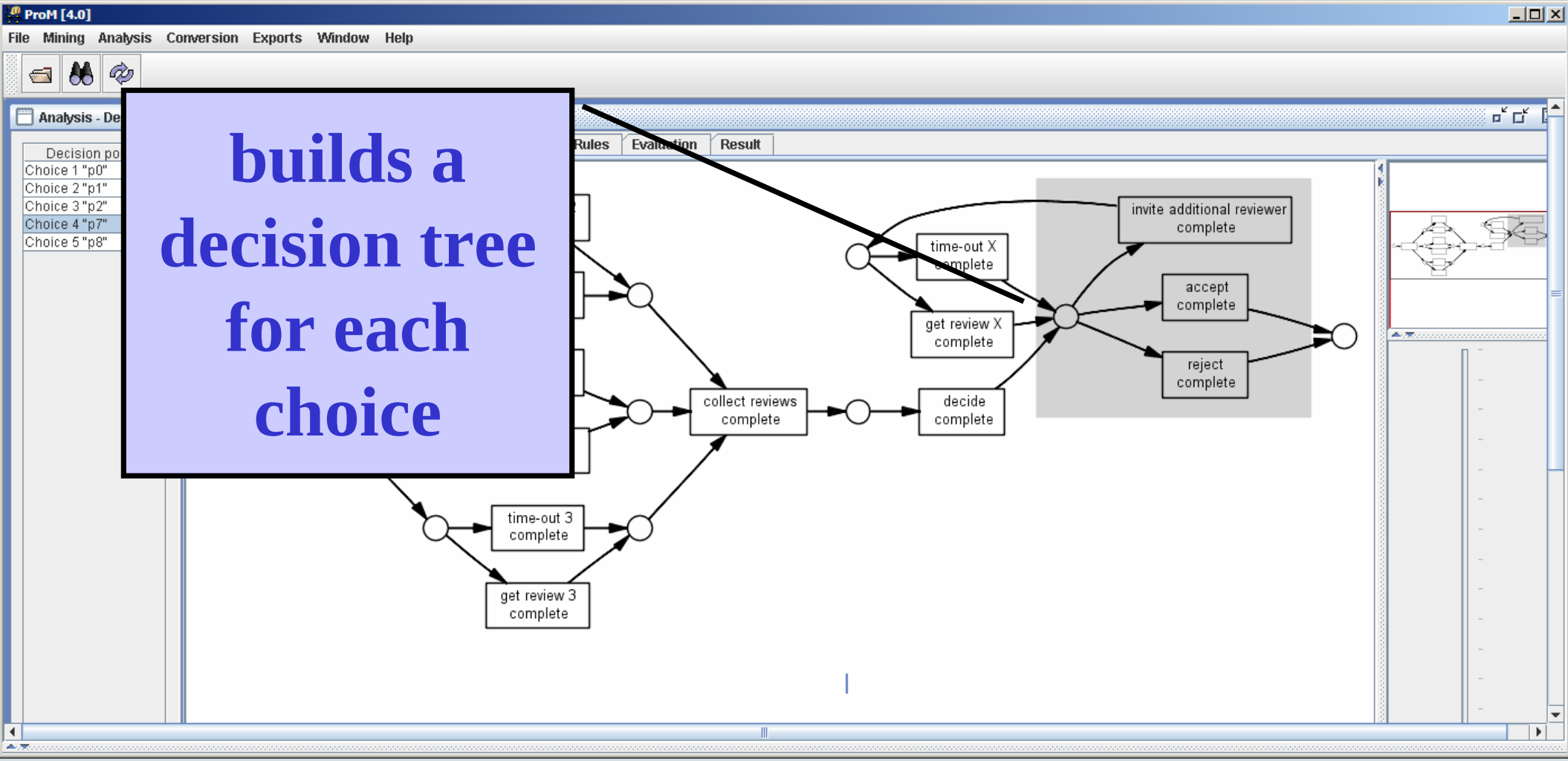



ProM [4.0]
File
Mining
Analysis
Conversion
Exports
Window
Help
Heuristics net
Open new log...
Multi-phase Macro Plugin
Partial Order Generator
Partial Order Aggregator
Heuristics miner
DWS mining plugin
Association Rule Miner
Genetic algorithm plugin
Duplicate Tasks GA plug-in
Alpha algorithm plugin
Tsinghua-alpha algorithm plugin
Alpha++ algorithm plugin
Transition System Generator
Region miner
Change mining plugin
Activity Clustering Miner
Workflow patterns miner
Social network miner
StaffAssignmentMiner
Organizational Miner
Case data extraction plugin
Frequency abstraction miner
Fuzzy Miner
Cloud chamber miner
Trace Clustering Miner
Process Instance Inspector
Logreader Benchmark
LogReader comparison driver
k-RI Miner
Results - Settings for mi
for mining Filtered reviewslog_with_fewer_errors.xml using Fuzzy Miner
Binary metrics
invite reviewers complete 1,000
Cluster 16 2 primitives ~ 0,152
time-out 3 complete 0,340
collect reviews complete 0,397
invite additional reviewer complete 0,914
time-out X complete 0,514
reject complete 0,910
accept complete 0,890
Concurrency filter
Edge filter
Node filter
Significance cutoff
Results - Settings for mining Filtered reviewslog_with...
about 30 mining plug-ins!
Navigate along the log:
0.0 %

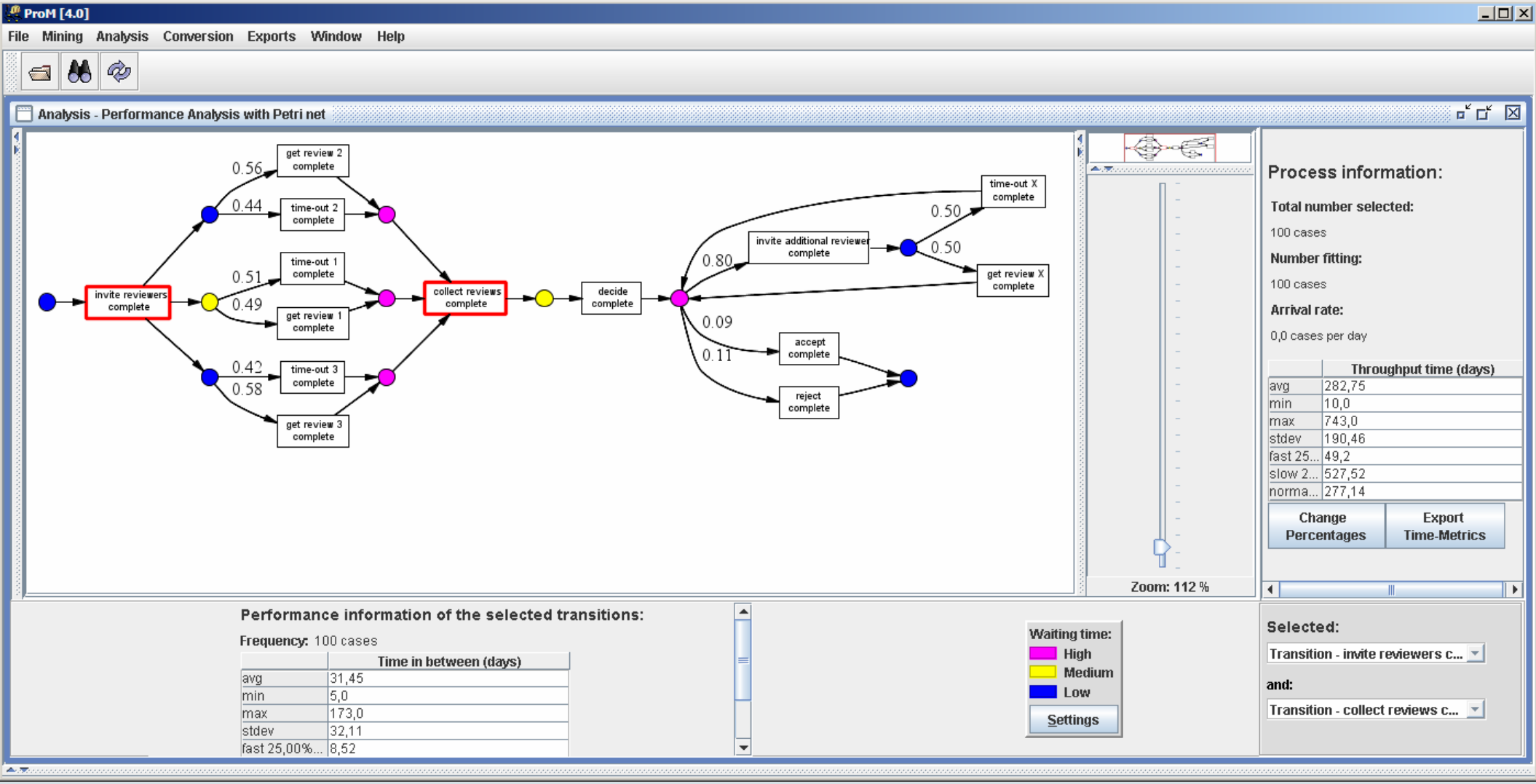
Social network analysis



Decision point analysis

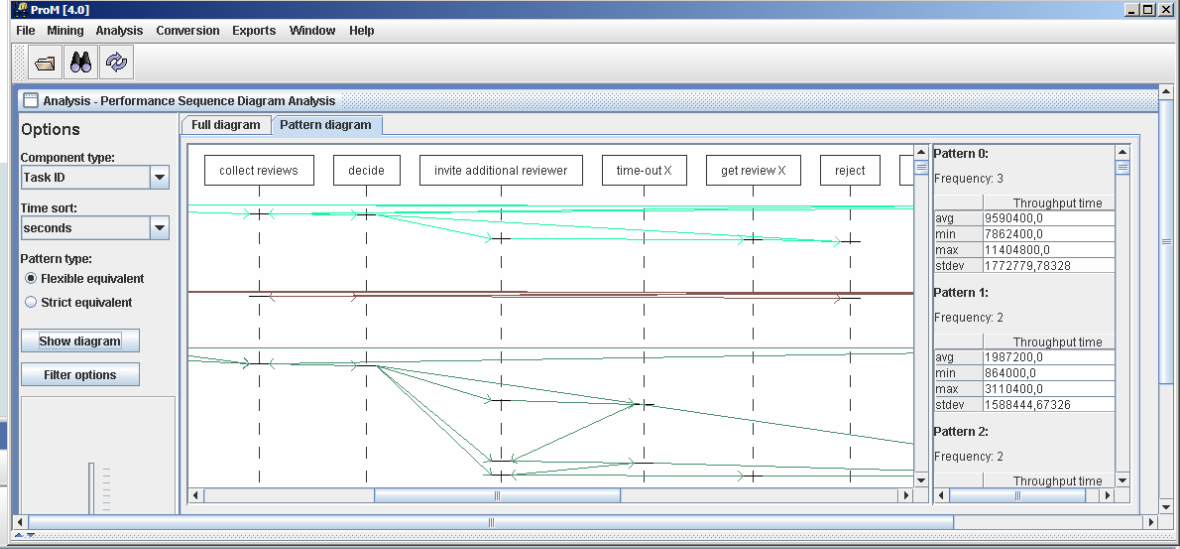


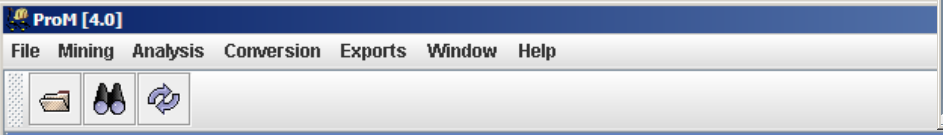
Performance analysis

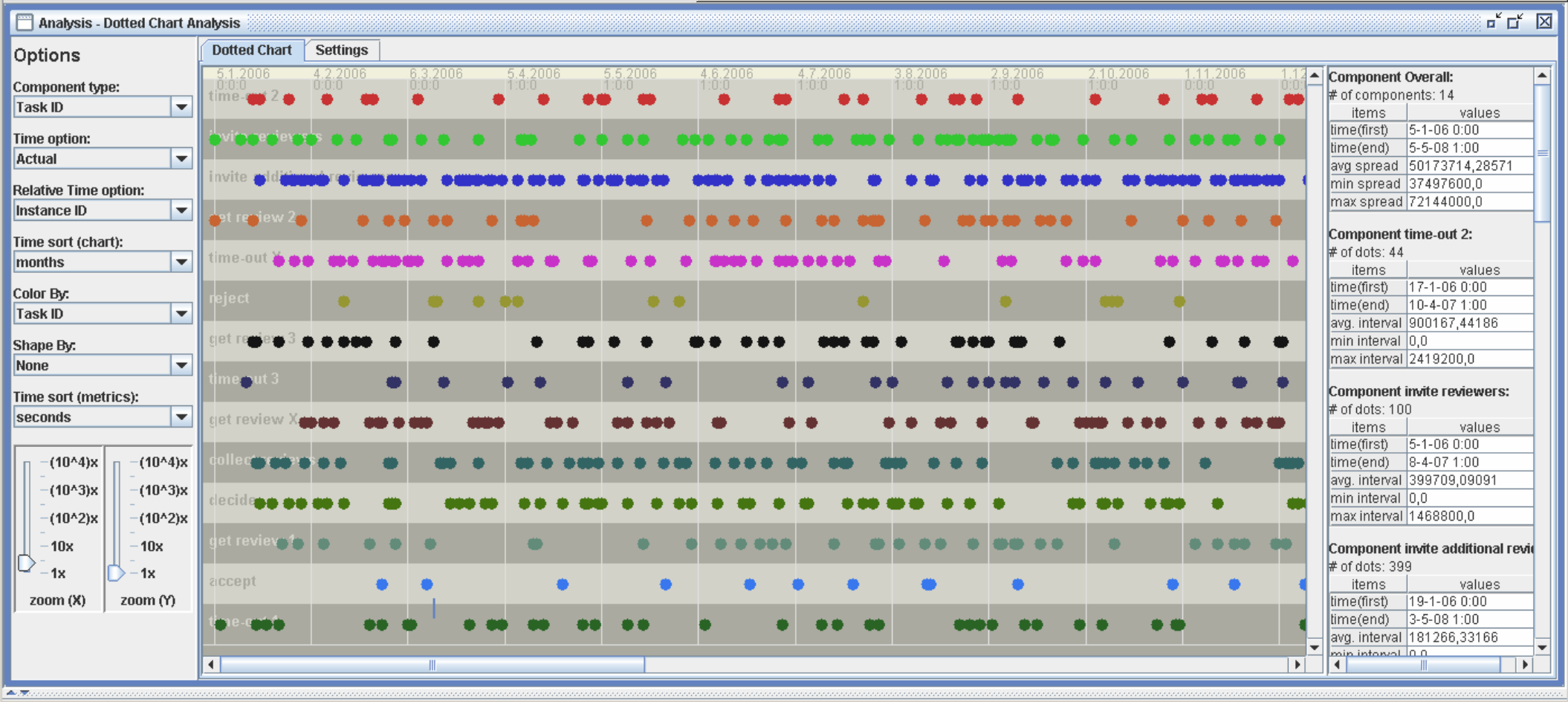


TU/e technische universiteit eindhoven

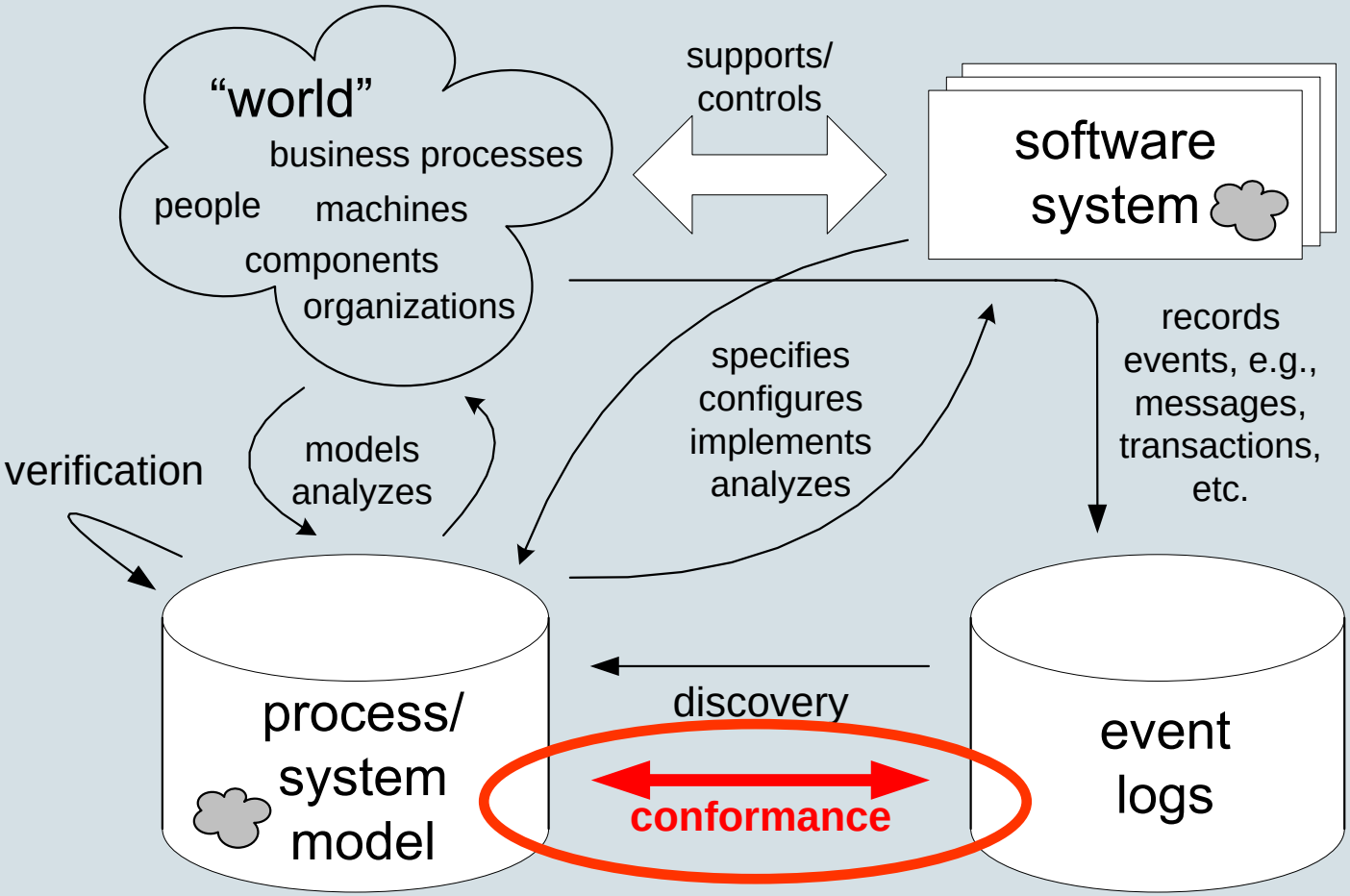
Discovering patterns



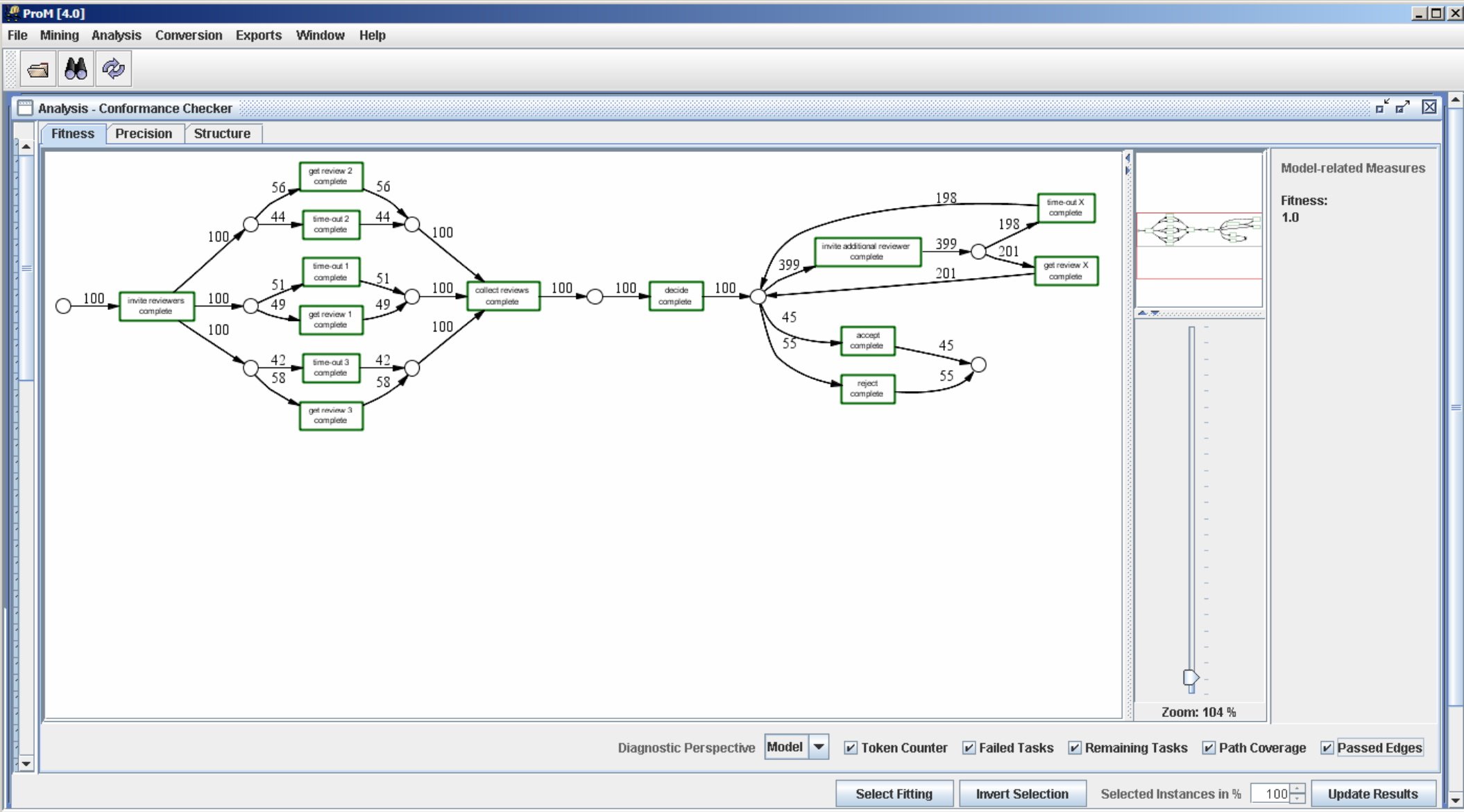




Conformance Checking

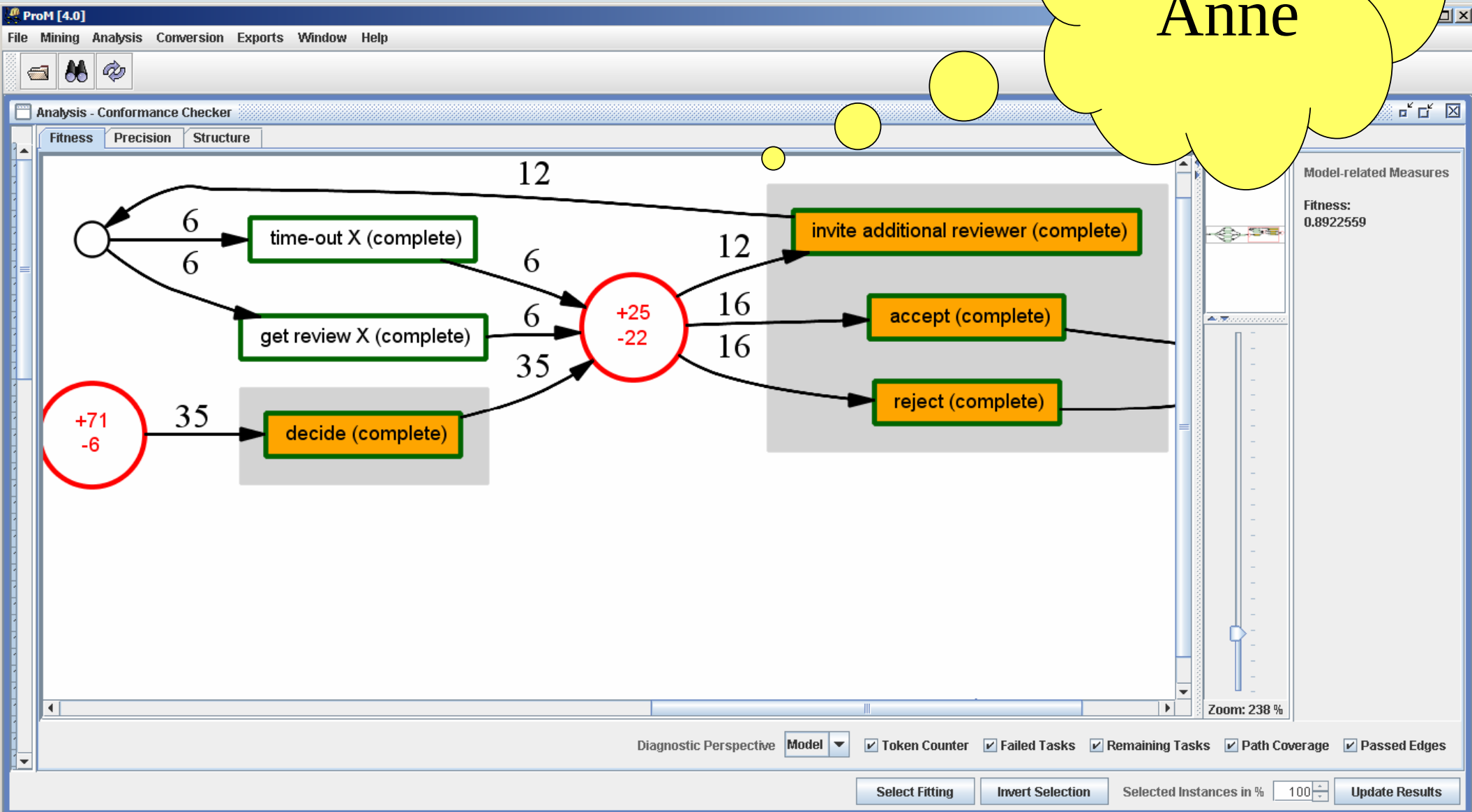


Comparing the discovered model with the log (f=1)

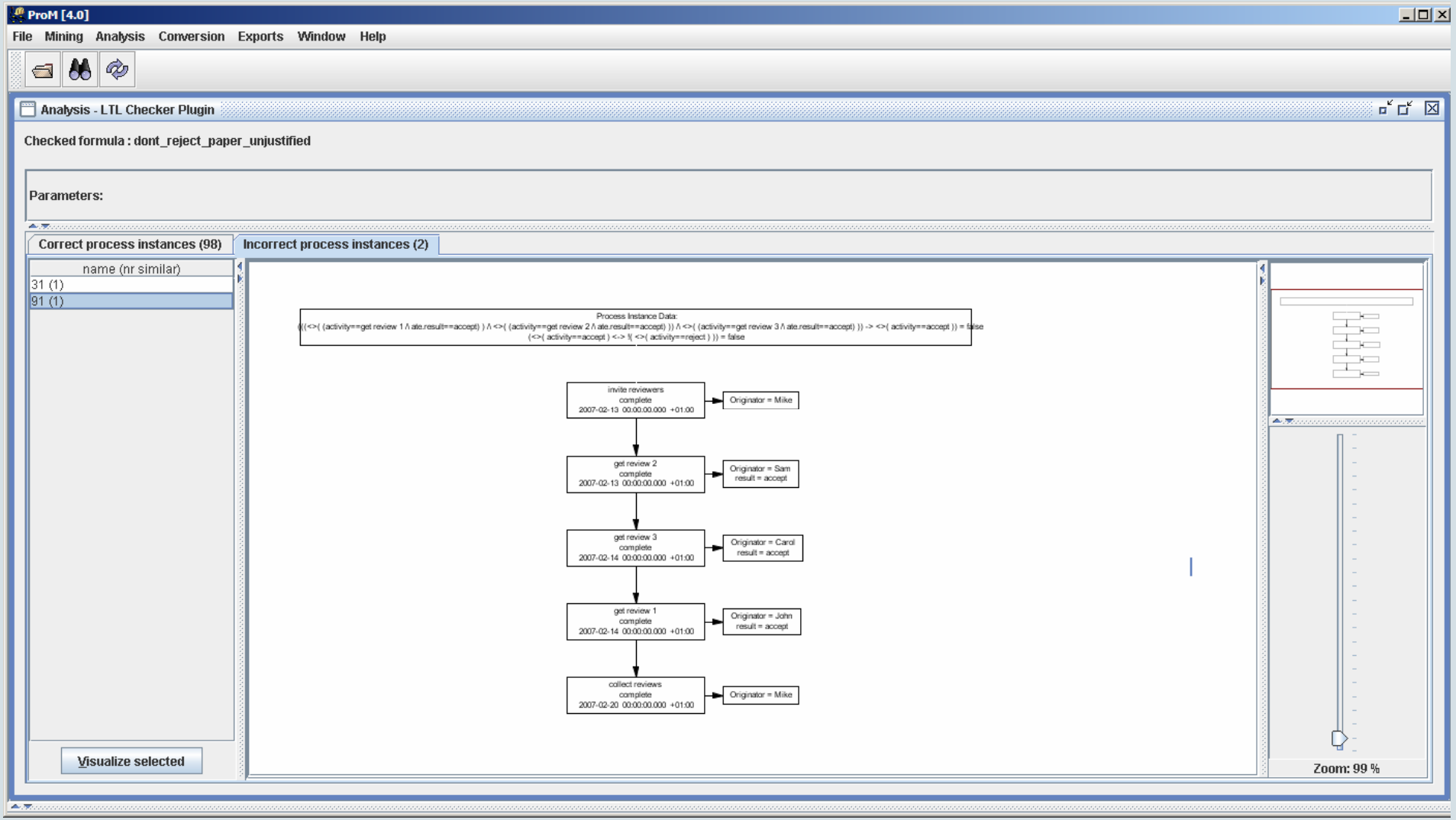


Adding deviations to the log (f=0.89)

link to Anne



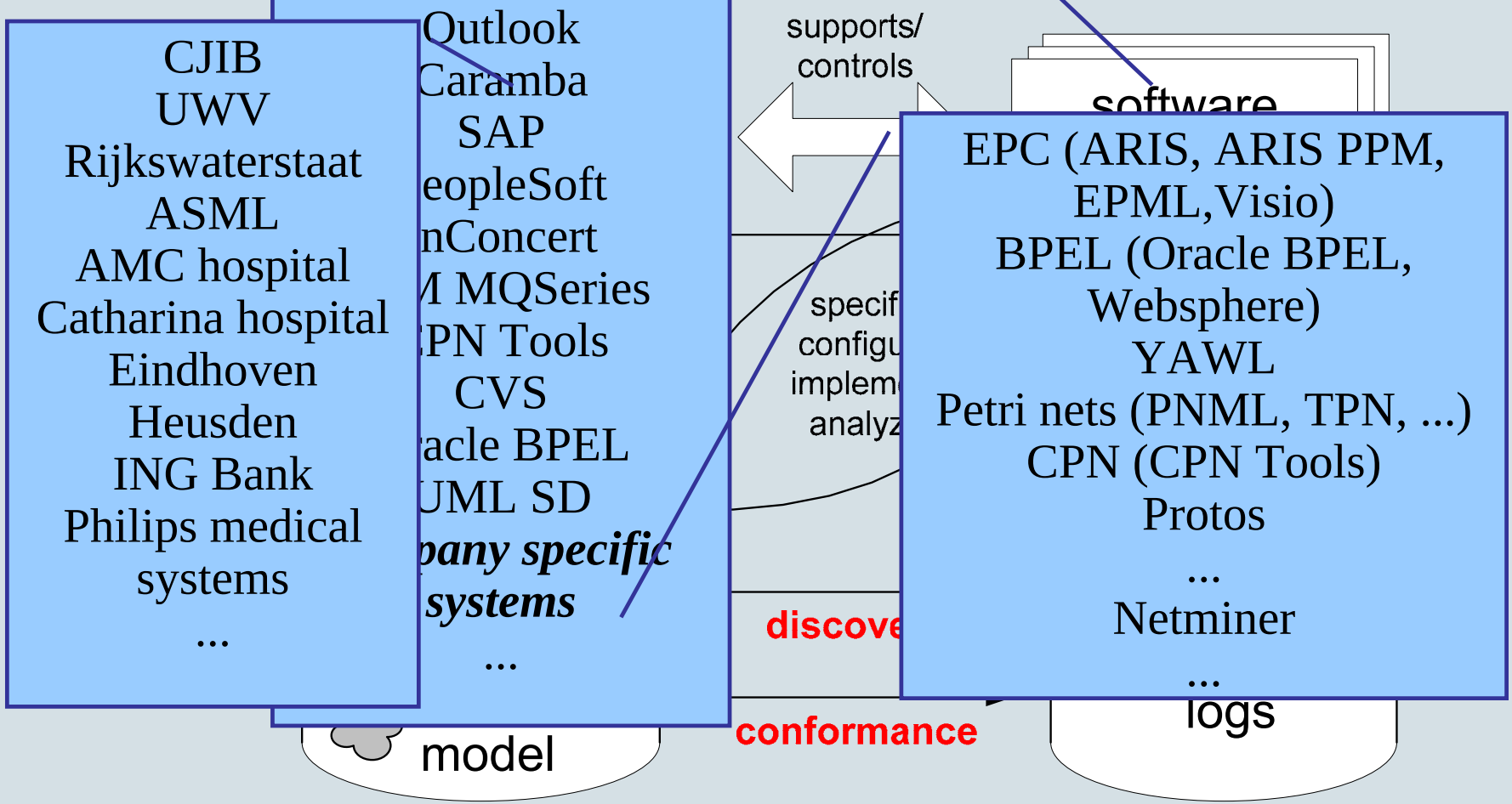
LTL checker plug-in




REALITY
CHECK
AHEAD

TU/e

Goal c mplete support

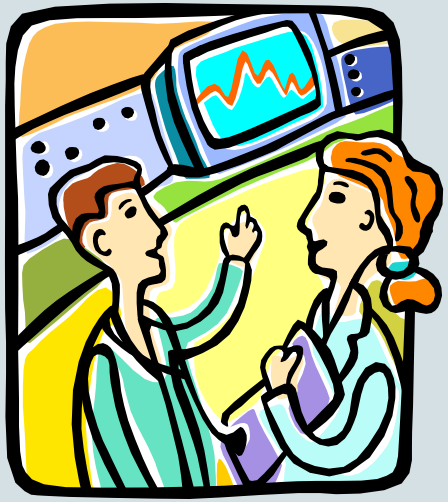


Reality Check

- Process mining on structured/administrative workflow-like logs is relatively easy.
- However, let us look at two ***extreme*** logs:
  - A log from a hospital with information on treatments, complications, and diagnoses.
  - A log from a manufacturer of professional systems with information on system tests.

First example: Hospital data

- Information on treatment, complication, and diagnosis events.
- Data:
  - 2712 cases (all unique)
  - 29258 events
  - +/- 10.8 events per case
  - 264 different events (activities)

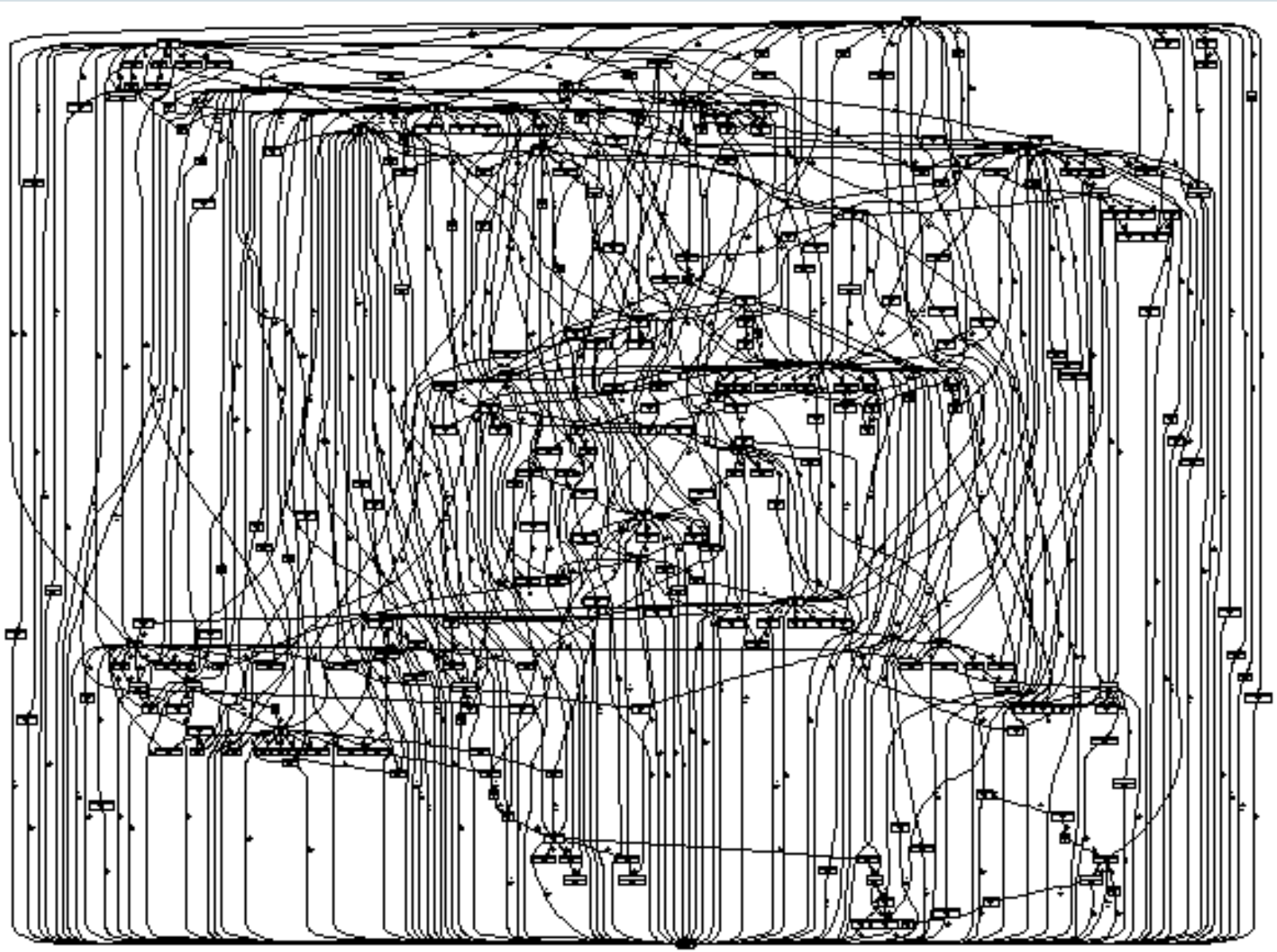


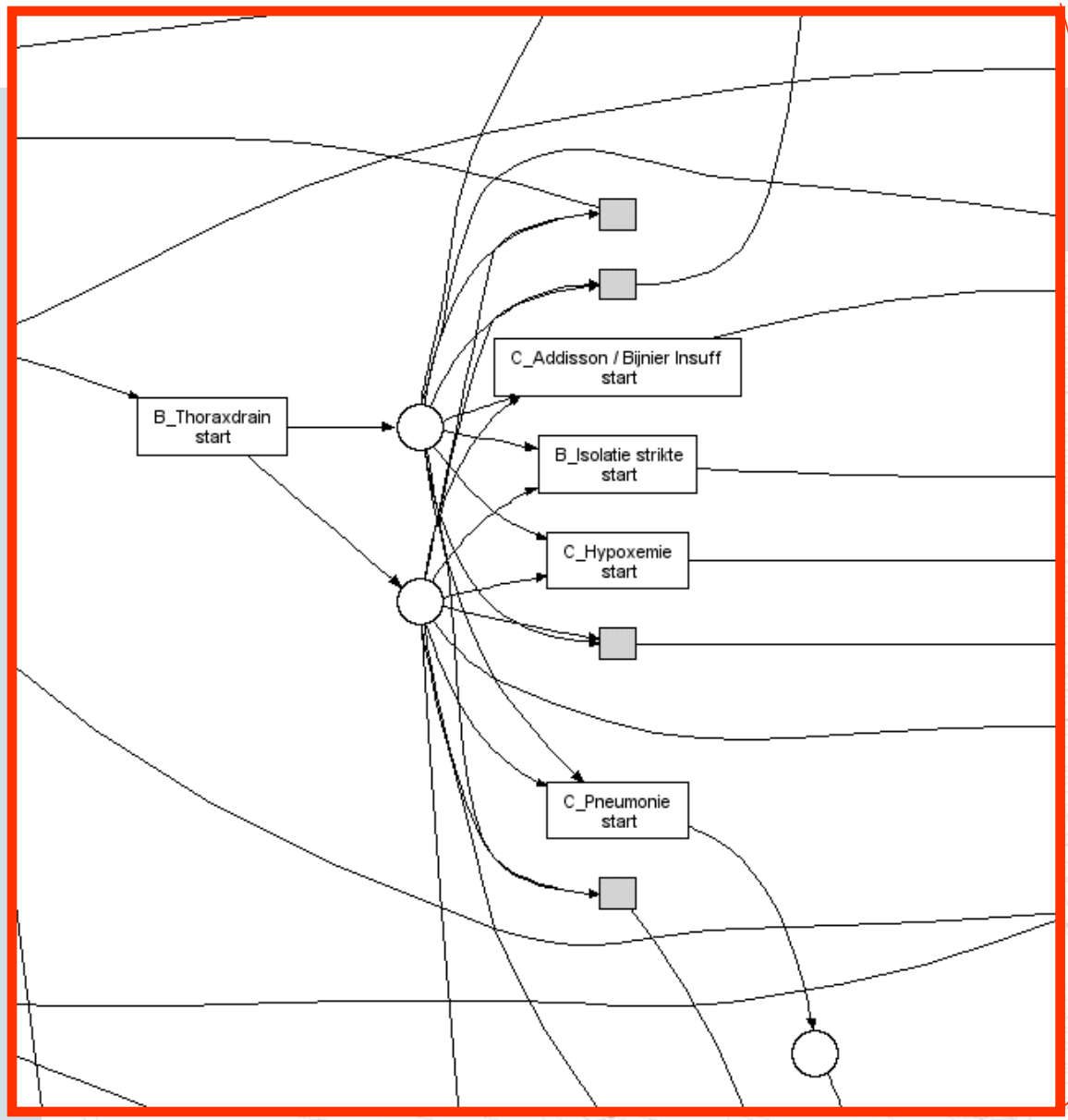
Frequency of activities

Model element	Event type	Occurrences (absolute)	Occurrences (relative)
B_Perifeer infuus	start	2837	9,696%
B_Maagsonde	start	2430	8,305%
B_Beademing	start	2187	7,475%
B_Catheter a Demeure	start	2096	7,164%
B_Basiszorg	start	2010	6,87%
B_Arterie lijn op OK	start	2002	6,843%
B_O2 masker/slang	start	1954	6,679%
B_Thoraxdrain	start	1863	6,367%

C_N Phrenicus Paralyse	start	1	0,003%
C_TIA	start	1	0,003%
B_Horizontaal	start	1	0,003%
C_Cholecystitis, acalc	start	1	0,003%
C_Decubitus hak st. 3a	start	1	0,003%
C_Druk necrose elders	start	1	0,003%
B_Decubitus zorg stadium 3b	start	1	0,003%
C_Haemolyse	start	1	0,003%
B_Decubitus zorg stadium 4b	start	1	0,003%
B_Isolatie Beschermend	start	1	0,003%
B_Donor Weefsel	start	1	0,003%
C_Polyurie (>40ml/kg/24u)	start	1	0,003%
C_Decubitus overig st. 3a	start	1	0,003%
C_Intra-peritoneaal Abces	start	1	0,003%

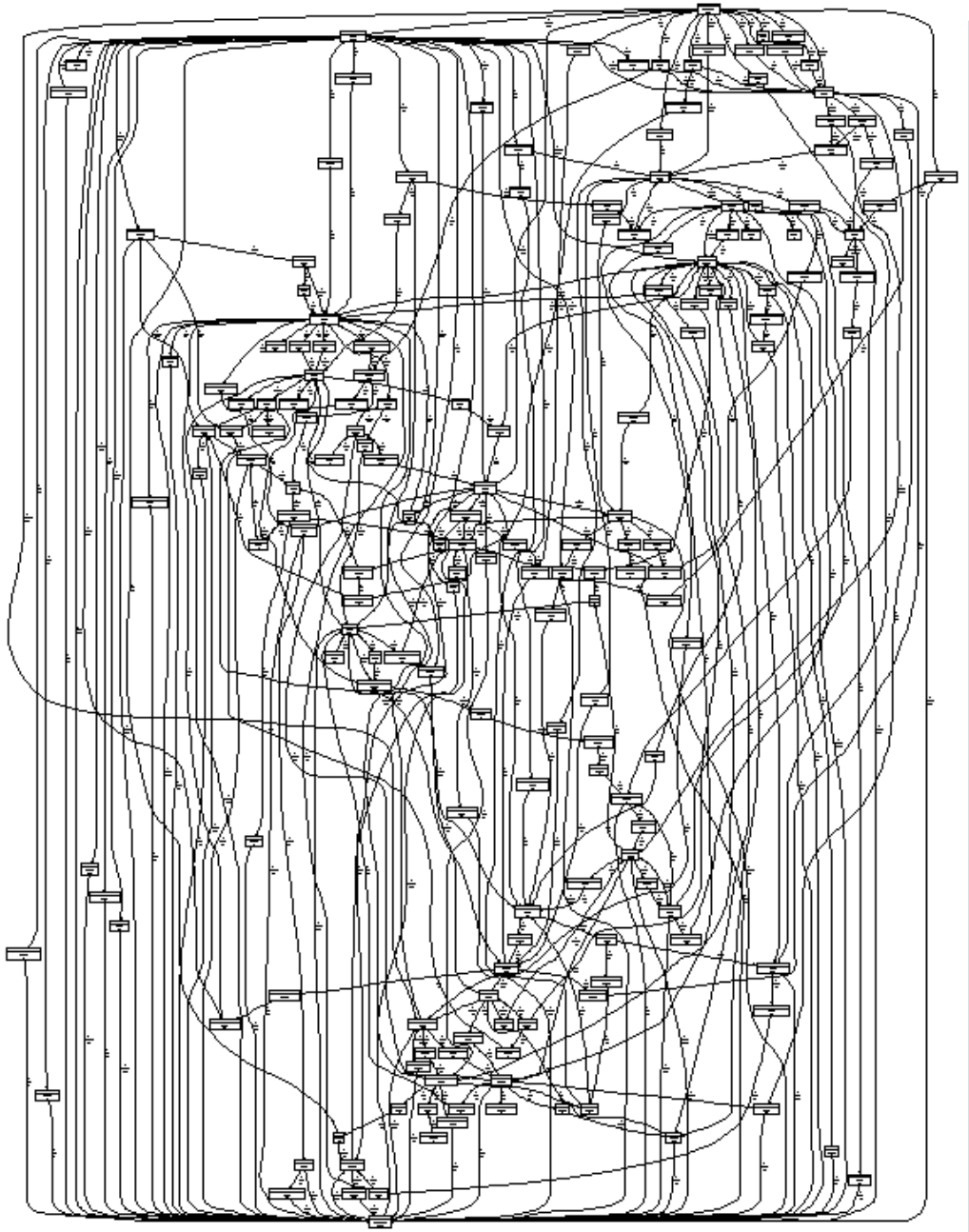
Heuristics miner




C_Addisson / Bijnier Insuff start
B_Thoraxdrain start
B_Isolatie strikte start
C_Hypoxemie start
C_Pneumonie start

Selection: Care after hart surgery

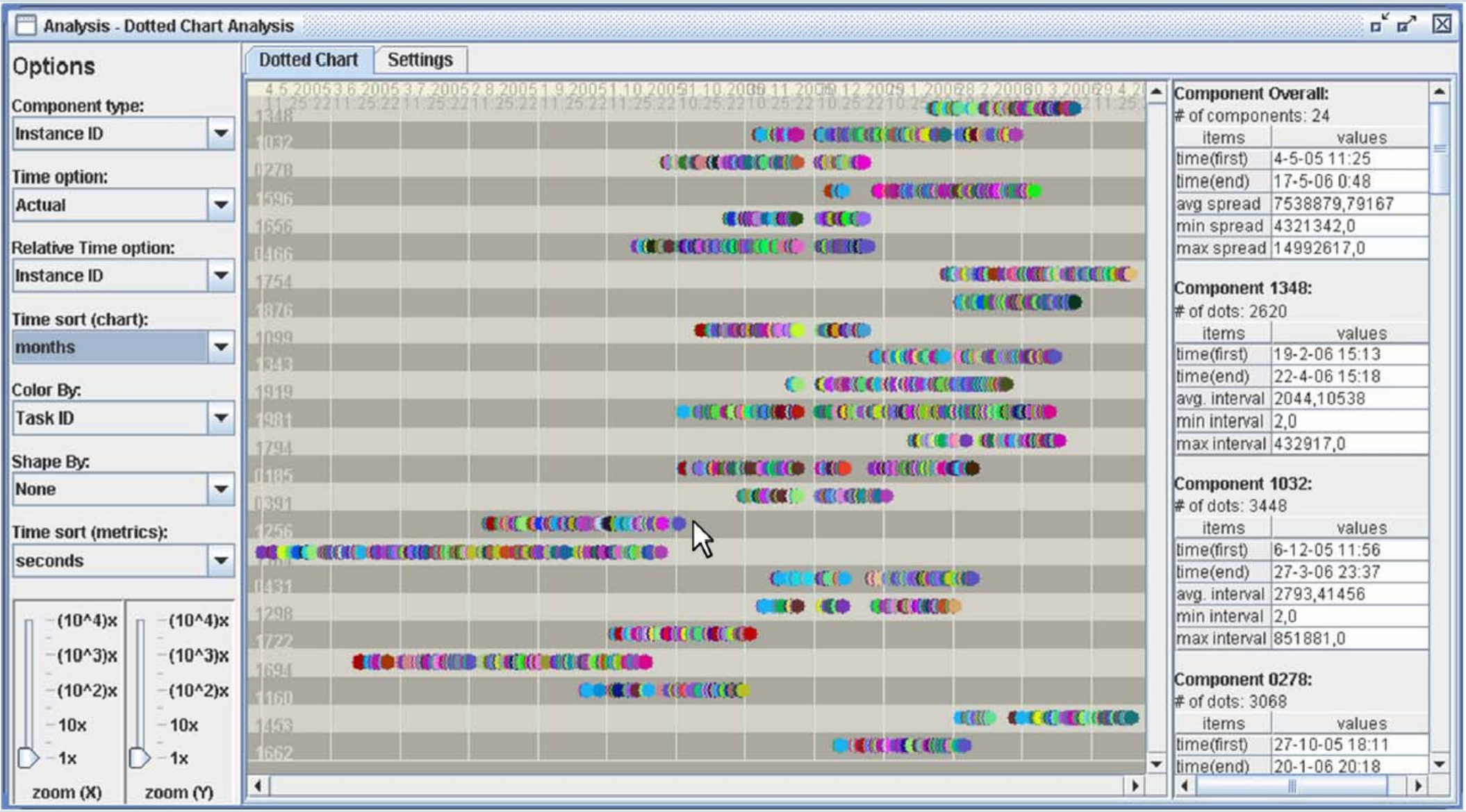
- Data
  - 874 cases (all unique)
  - 10478 events
  - 181 different events (activities)



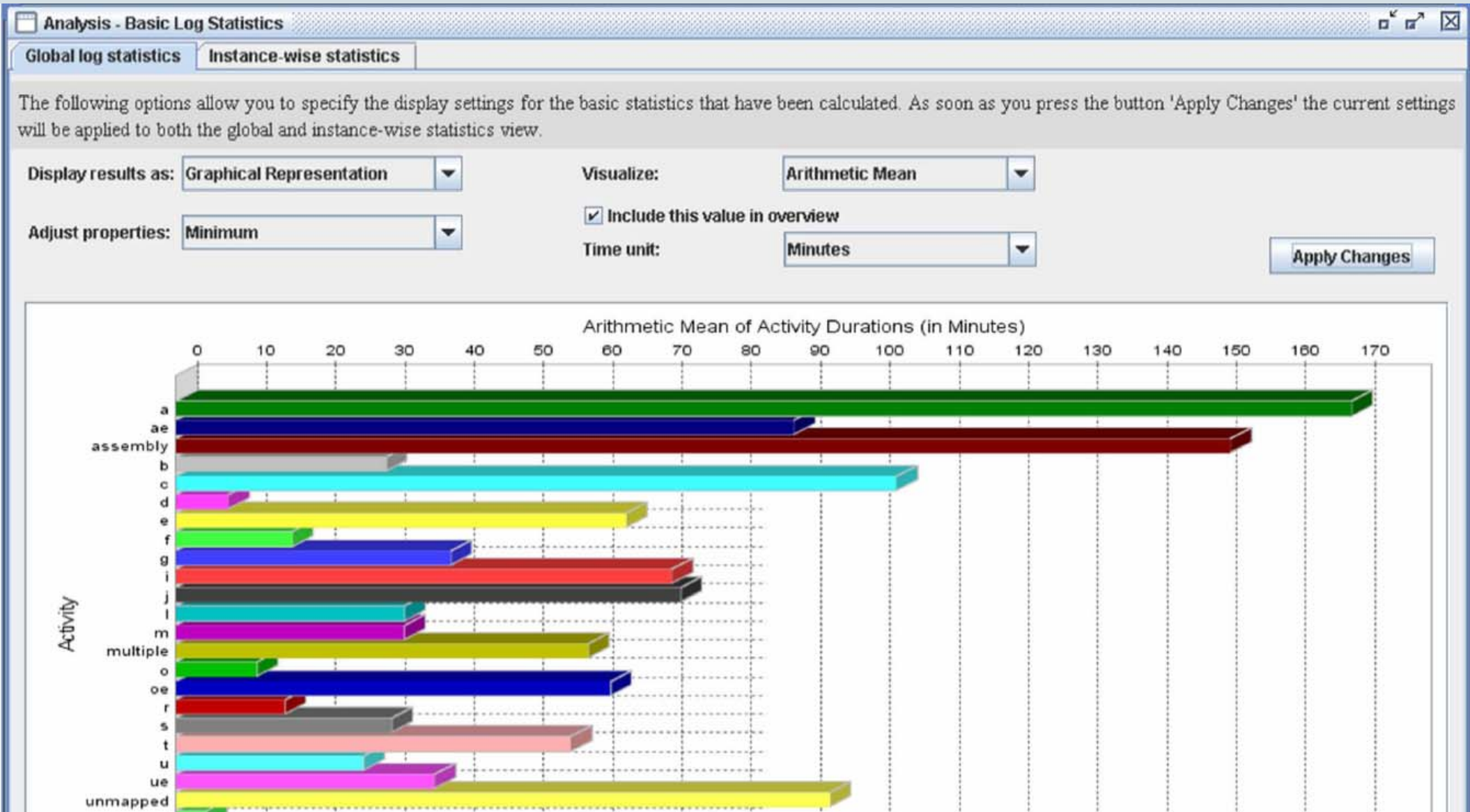
Second example: Test data from high-tech system manufacturer

- Information on testing process of complex professional machines
- Data:
  - 24 comparable cases
  - 154966 events
  - +/- 6450 events per case
  - between 2820 and 16250 events per machine
  - 720 different events (start/complete activities)

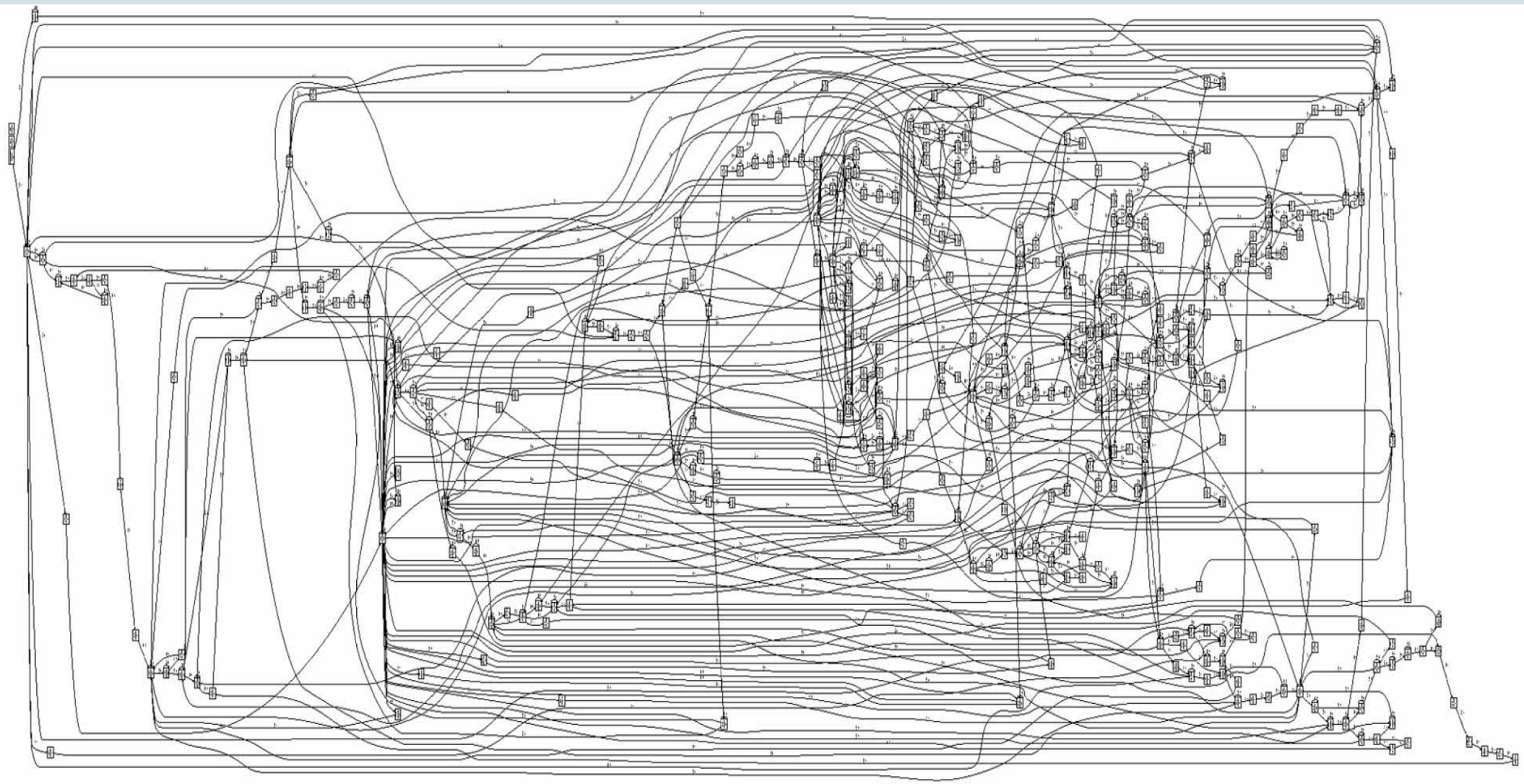
Helicopter view



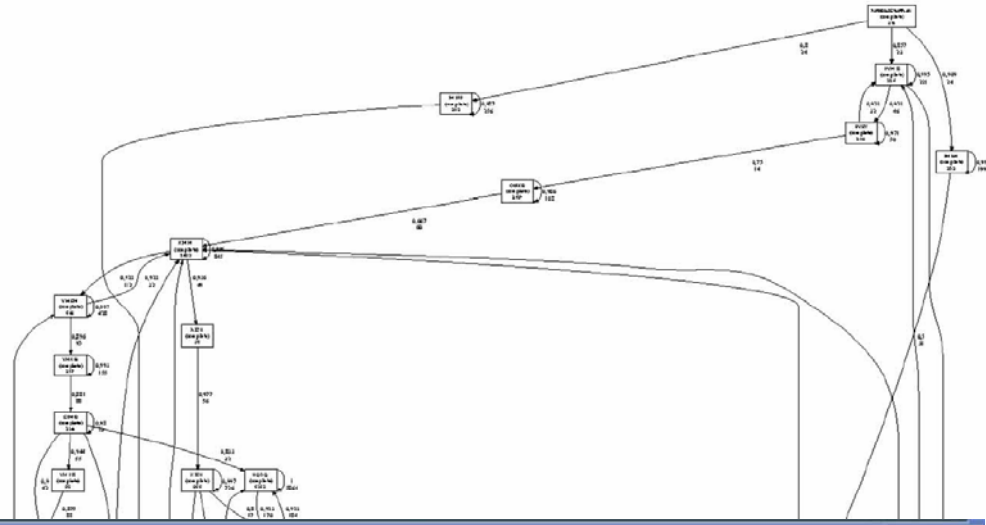
Average time spent in job-steps (aggregated events)

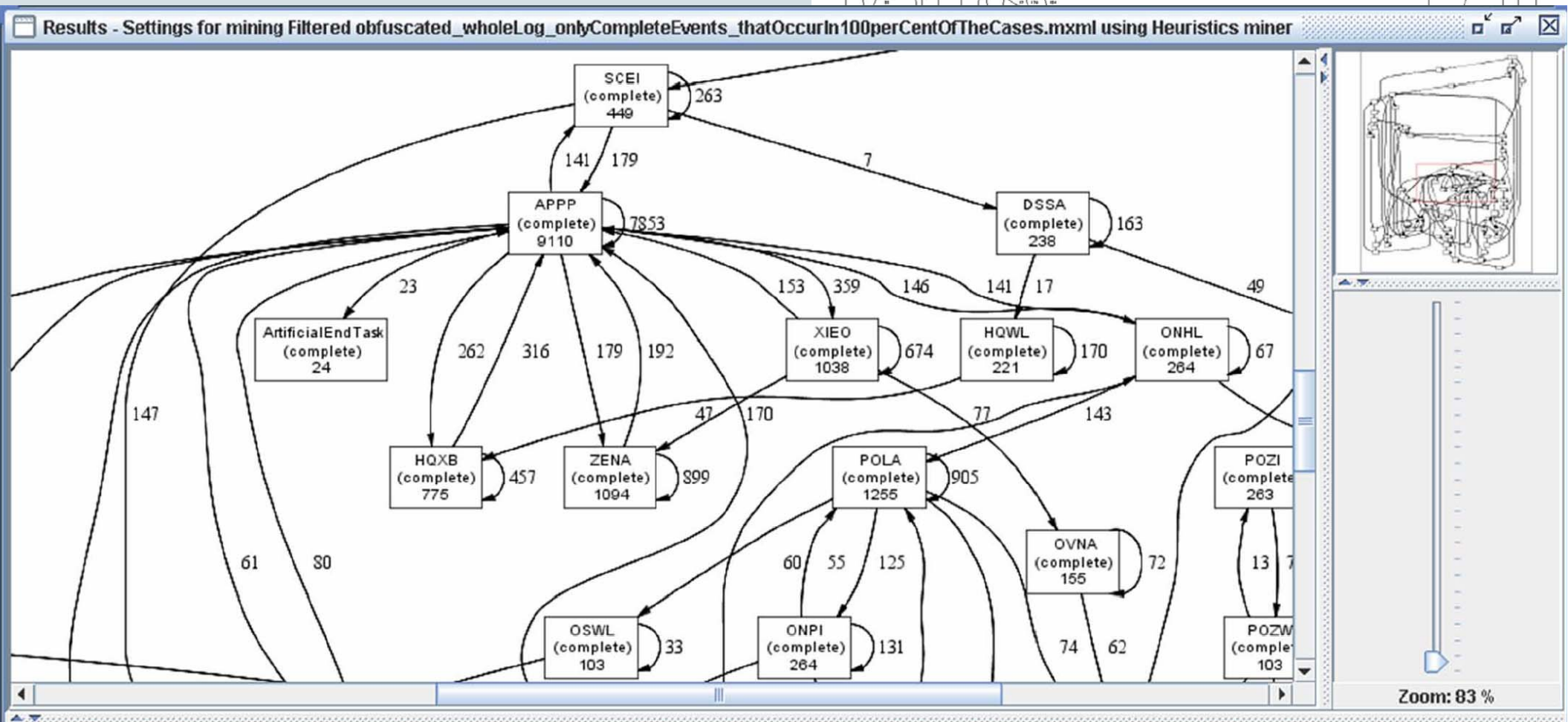


Mining just the complete events (# 360)…

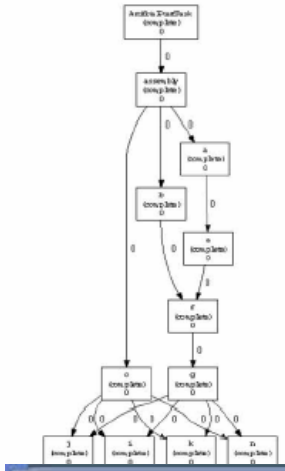


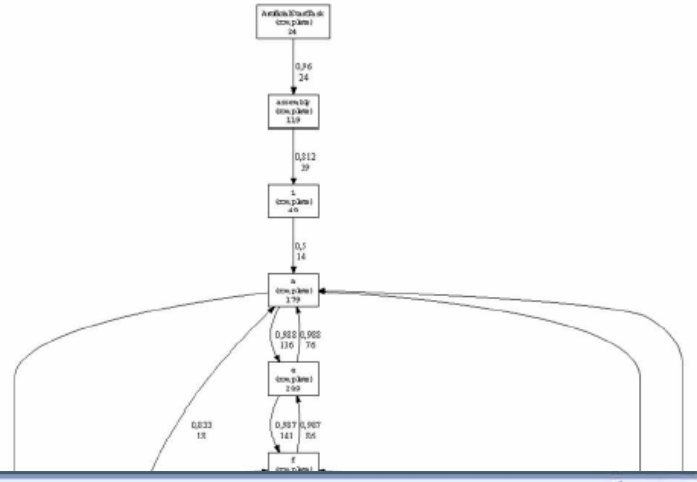
Common activities (#70)

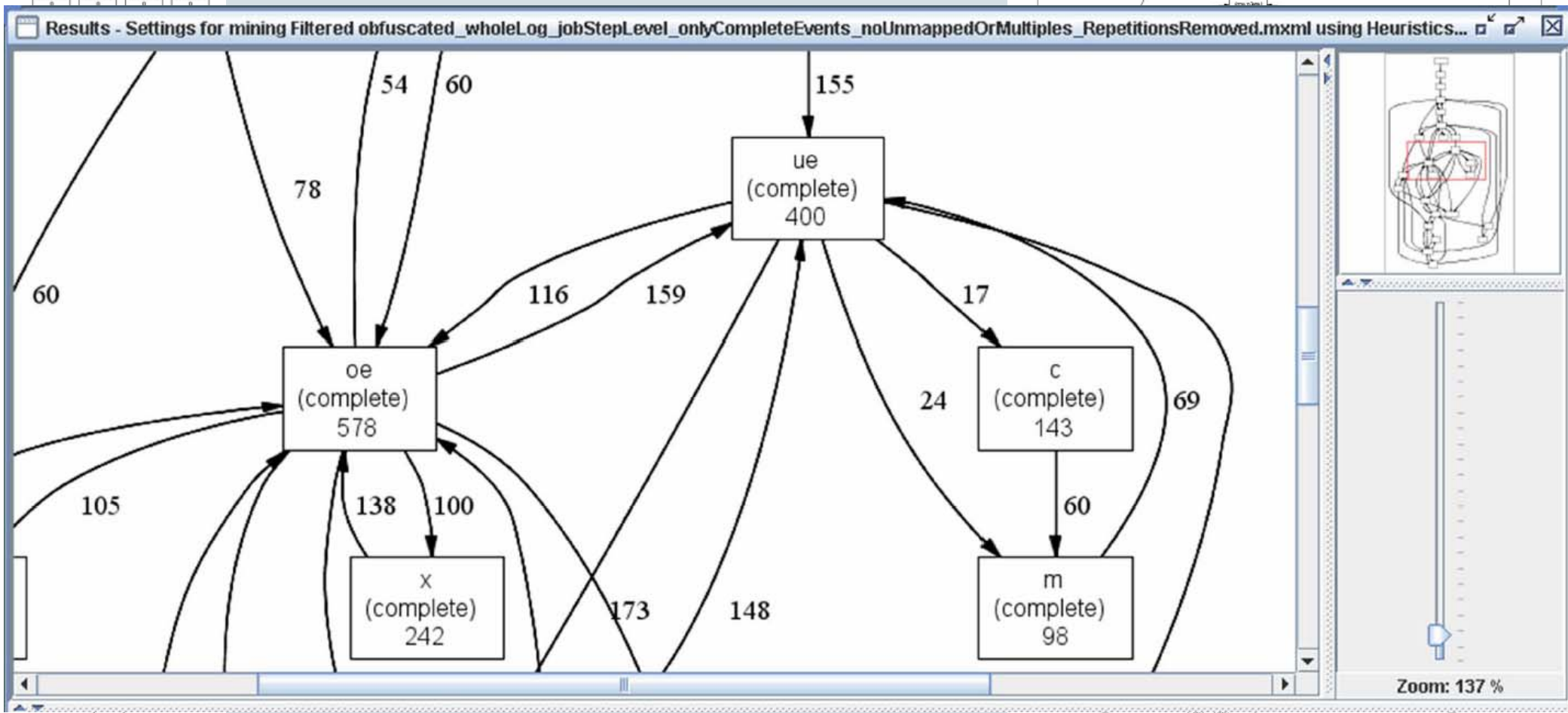




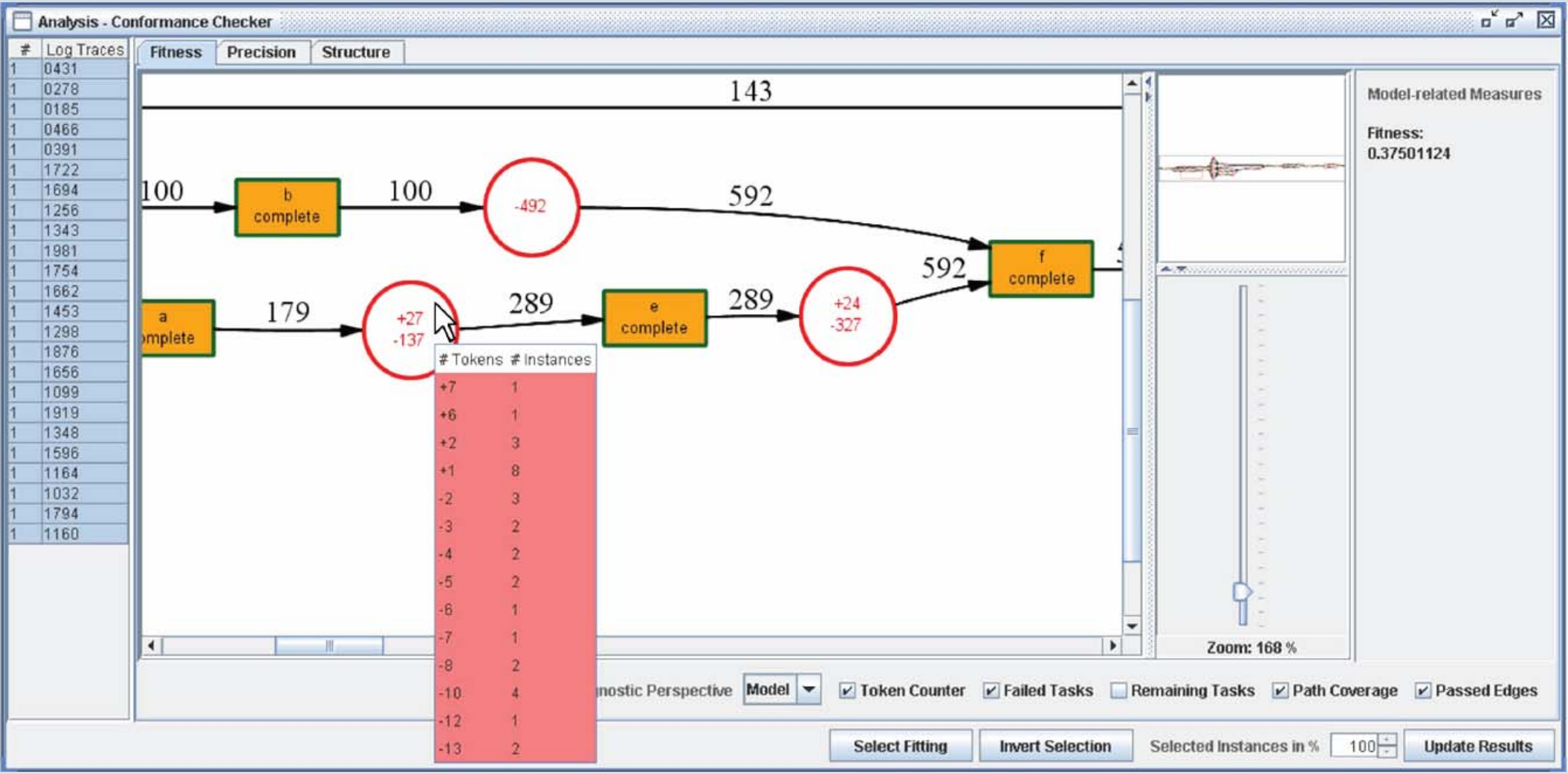
Job step level







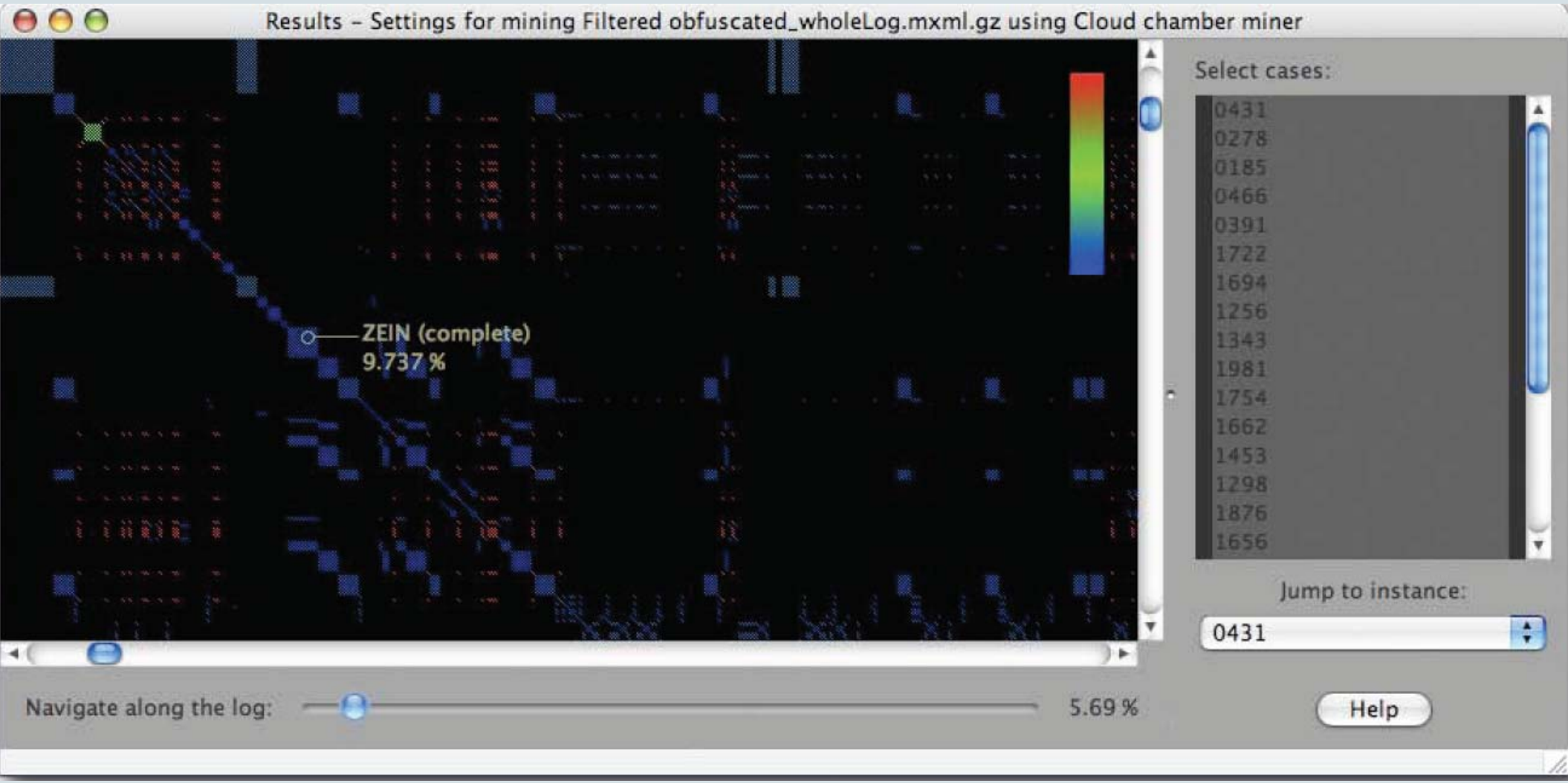
Conformance checker (reference model – job steps)



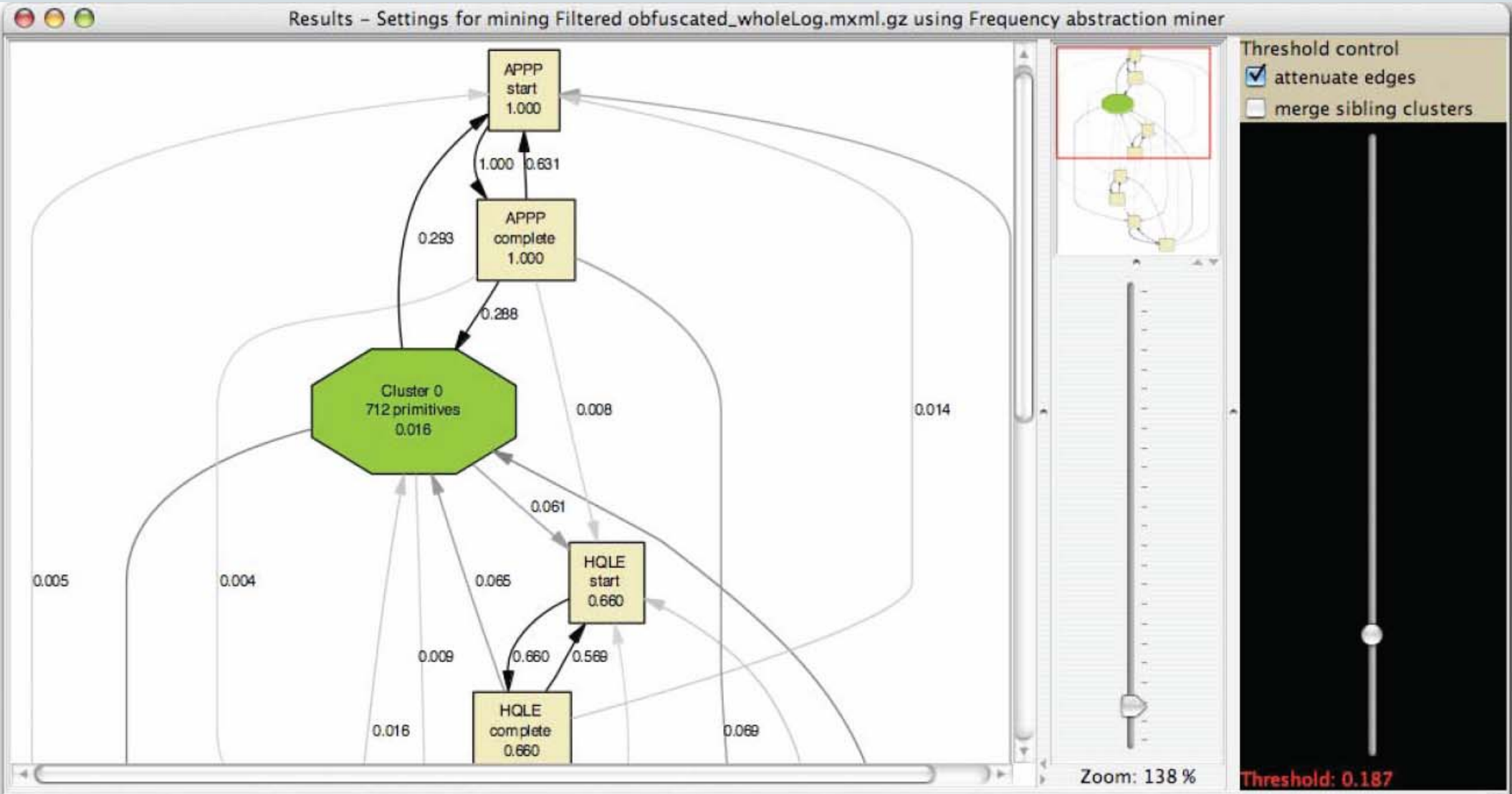
Discovered models fit better than reference model

Machine ID	*Fitness with respect to the reference process model*	*Fitness with respect to the discovered process model*
0431	$f = 0.30895045$	$f = 0.75113416$
0278	$f = 0.38491702$	$f = 0.82790506$
0185	$f = 0.37574032$	$f = 0.7171429$
0466	$f = 0.35643995$	$f = 0.74496365$
0391	$f = 0.38410193$	$f = 0.72710335$
1722	$f = 0.33359805$	$f = 0.7599035$
1694	$f = 0.39662793$	$f = 0.7821204$
1256	$f = 0.40988705$	$f = 0.7436544$
1343	$f = 0.3985147$	$f = 0.70074475$
1981	$f = 0.35723096$	$f = 0.6668054$
1754	$f = 0.401513$	$f = 0.77557445$
1662	$f = 0.4138763$	$f = 0.76897943$
1453	$f = 0.40458512$	$f = 0.5956065$
1298	$f = 0.37758428$	$f = 0.42357332$
1876	$f = 0.35556892$	$f = 0.7529762$
1656	$f = 0.36802232$	$f = 0.65629774$
1099	$f = 0.424476$	$f = 0.67167395$
1919	$f = 0.33690846$	$f = 0.72738254$
1348	$f = 0.41031277$	$f = 0.63753587$
1596	$f = 0.41015995$	$f = 0.58138484$
1164	$f = 0.37603533$	$f = 0.67173433$
1032	$f = 0.32361075$	$f = 0.7062931$
1794	$f = 0.39387232$	$f = 0.7337234$
1160	$f = 0.40484217$	$f = 0.7697767$

ProM's cloud chamber miner

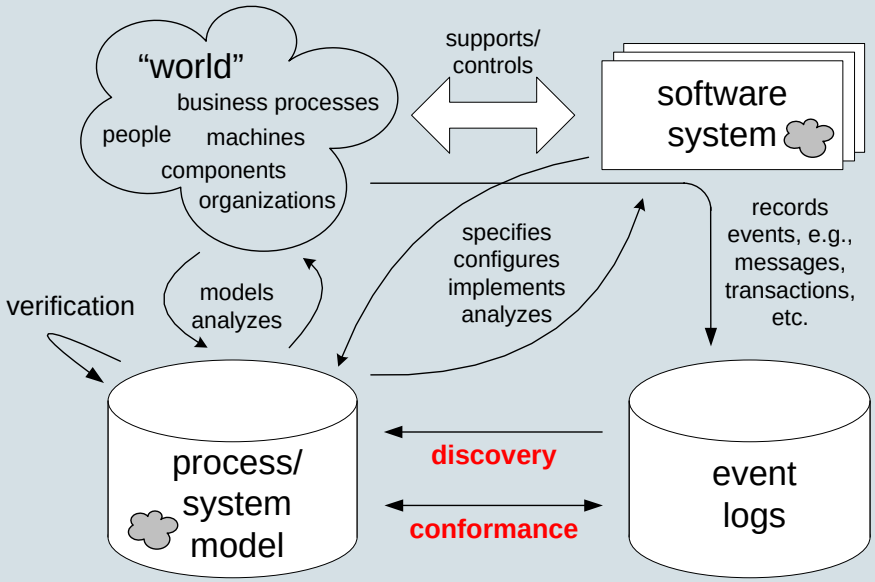


ProM's Frequency abstraction miner



Conclusion

- Reality is different from models!
- The existence of event data enables a wide variety of process mining techniques: discovery and conformance.
- In the context of services there many event logs around!
- ProM supports this (150 plug-ins)
- Although quite successful for "structured processes", "spaghetti processes" remain a challenge (two examples were given).
- Research should aim to address this challenge.



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.win.tue.nl/is/
- http://is.tm.tue.nl/staff/wvdaalst