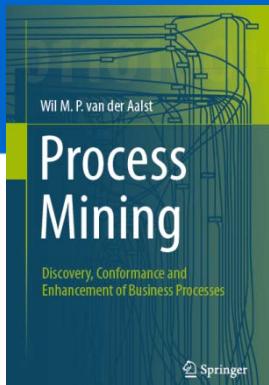


Process Mining: Data Science in Action

Process Mining Software

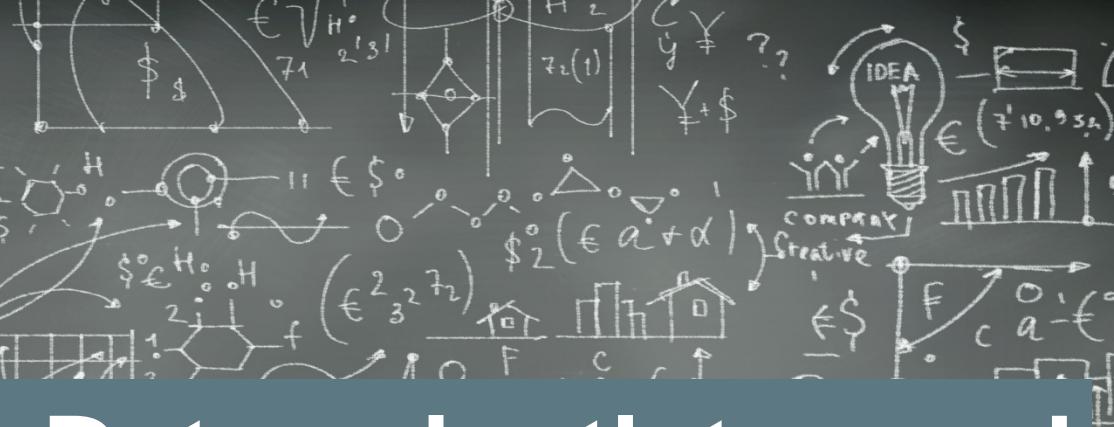
prof.dr.ir. Wil van der Aalst
www.processmining.org



TU/e

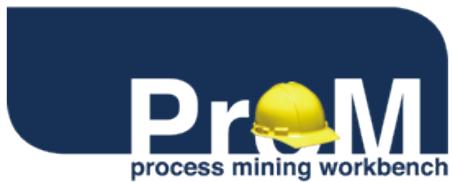
Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

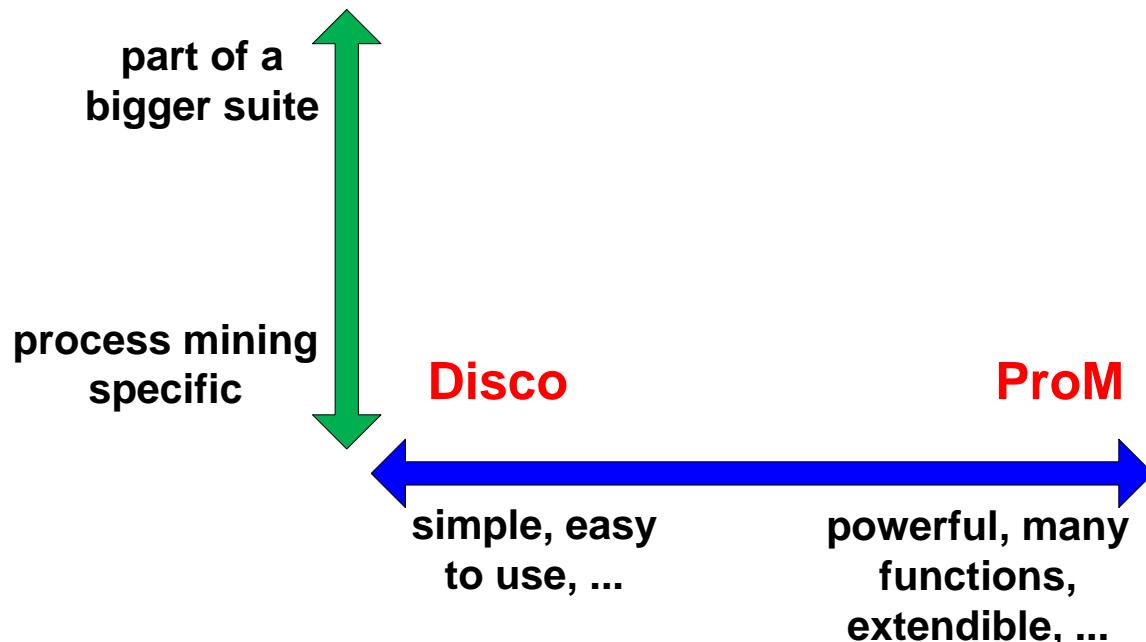


Data scientists need
to have data (see
previous lectures),
tools, and preferably
also questions.

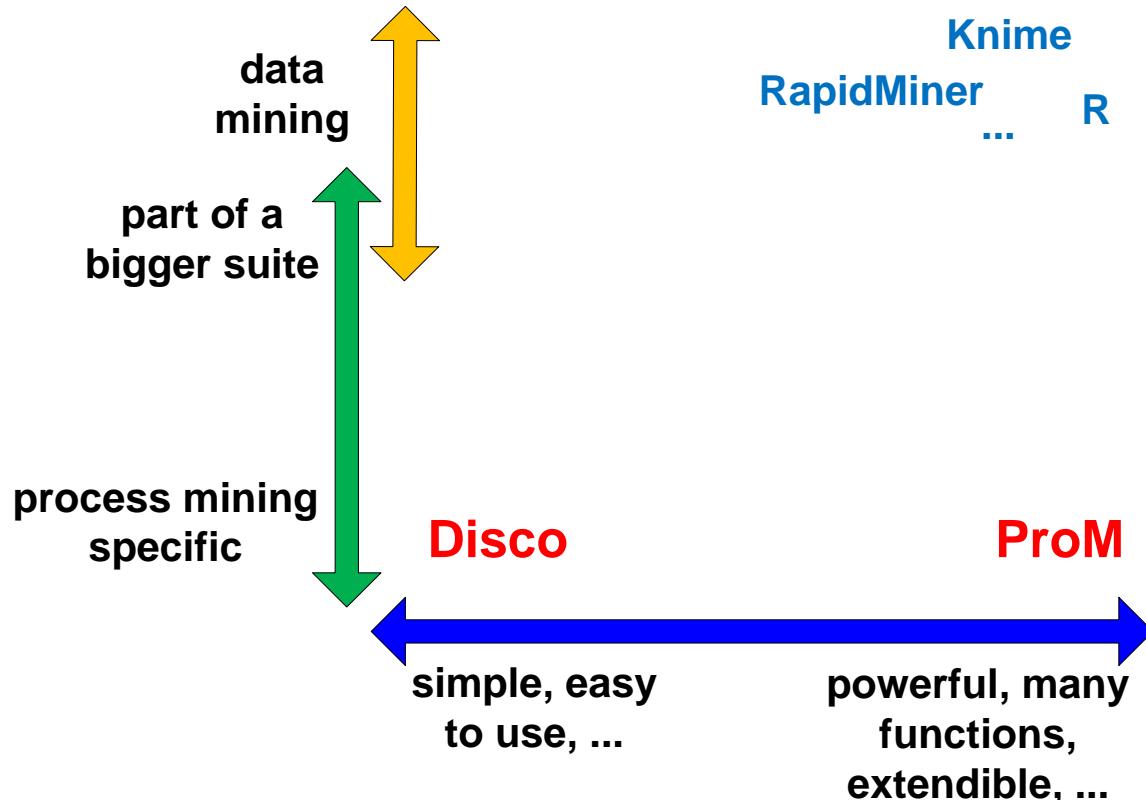
Process mining toolbox



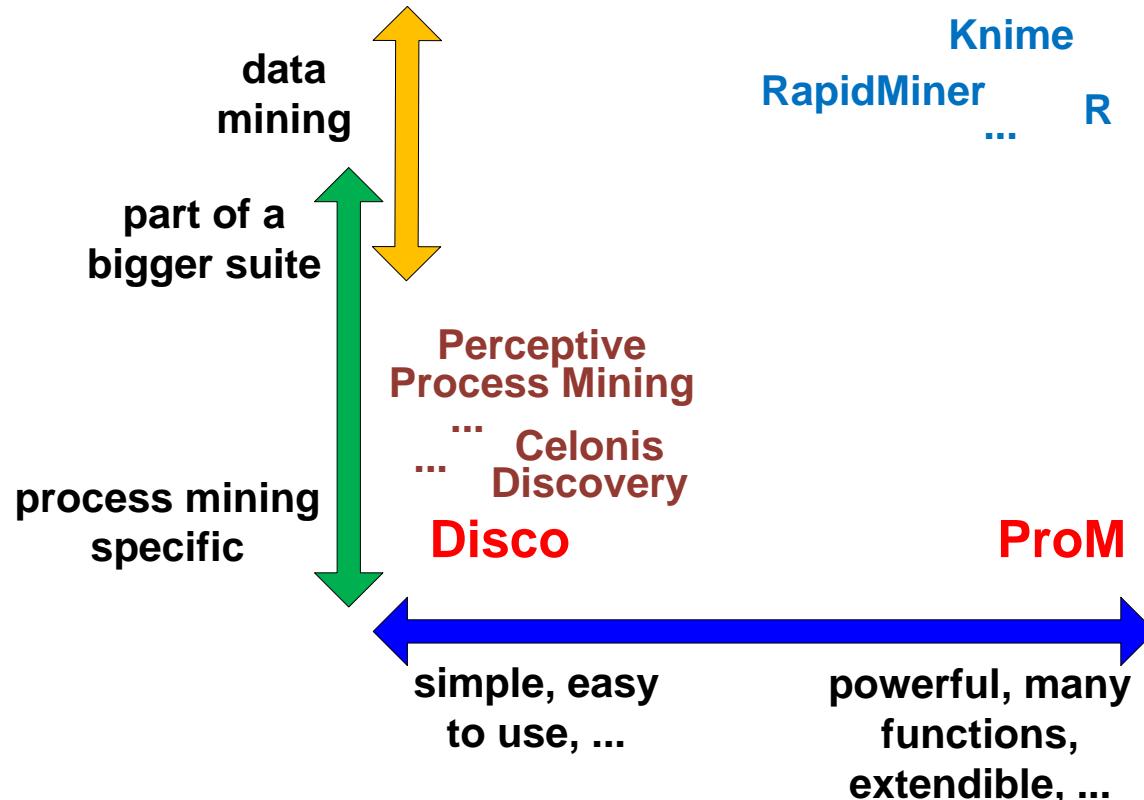
Spectrum of tools



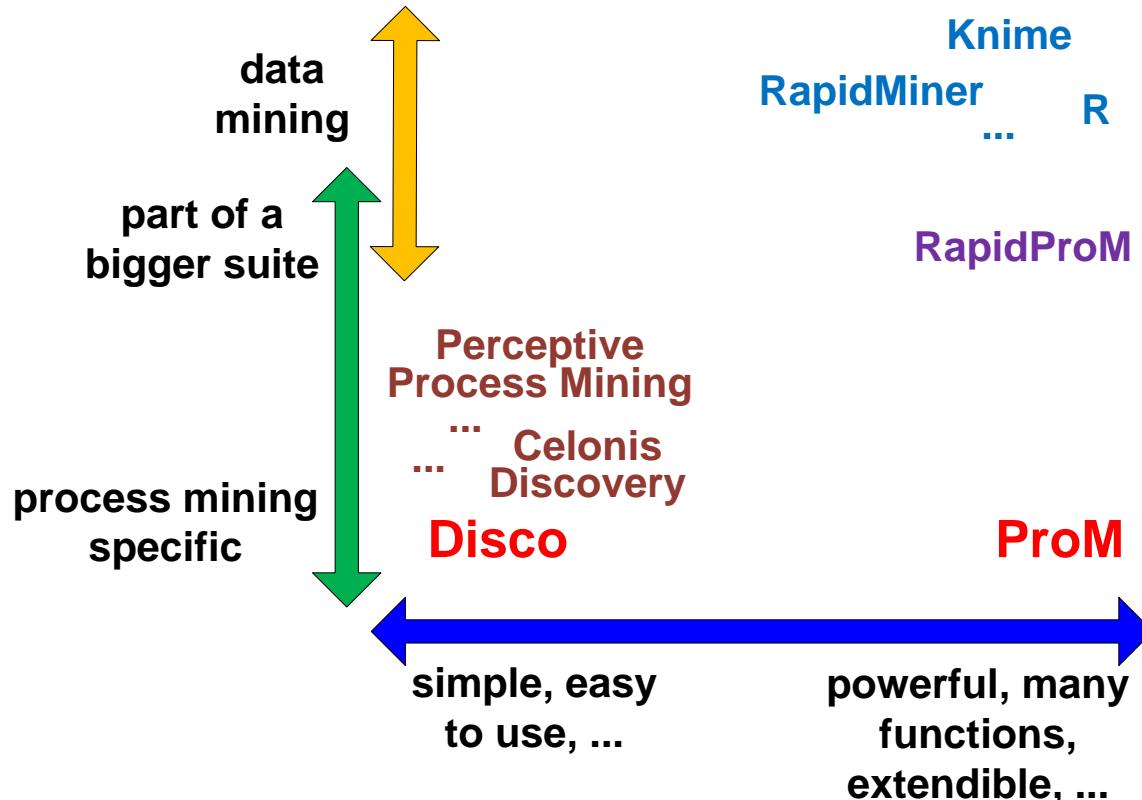
Spectrum of tools



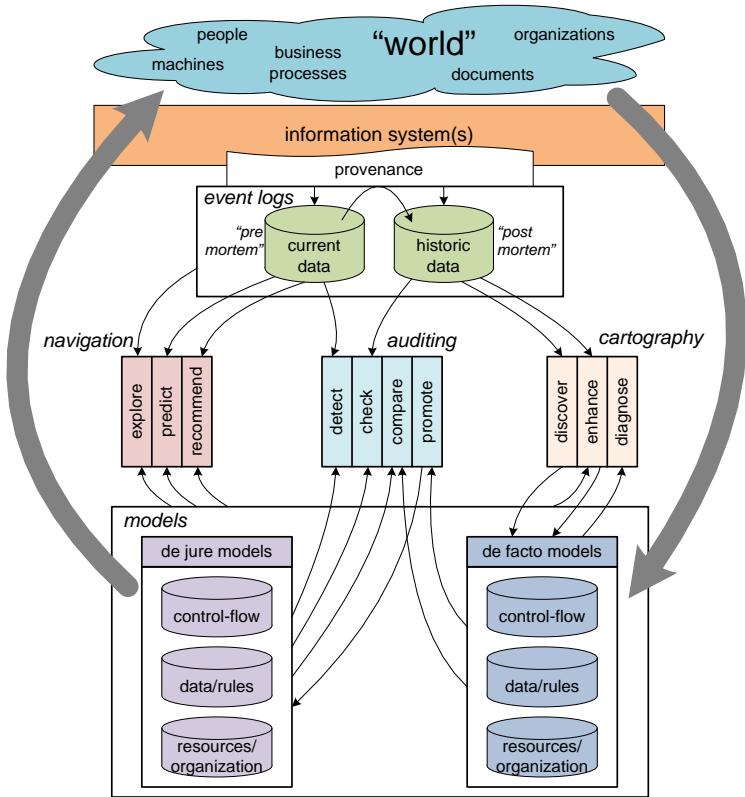
Spectrum of tools



Spectrum of tools

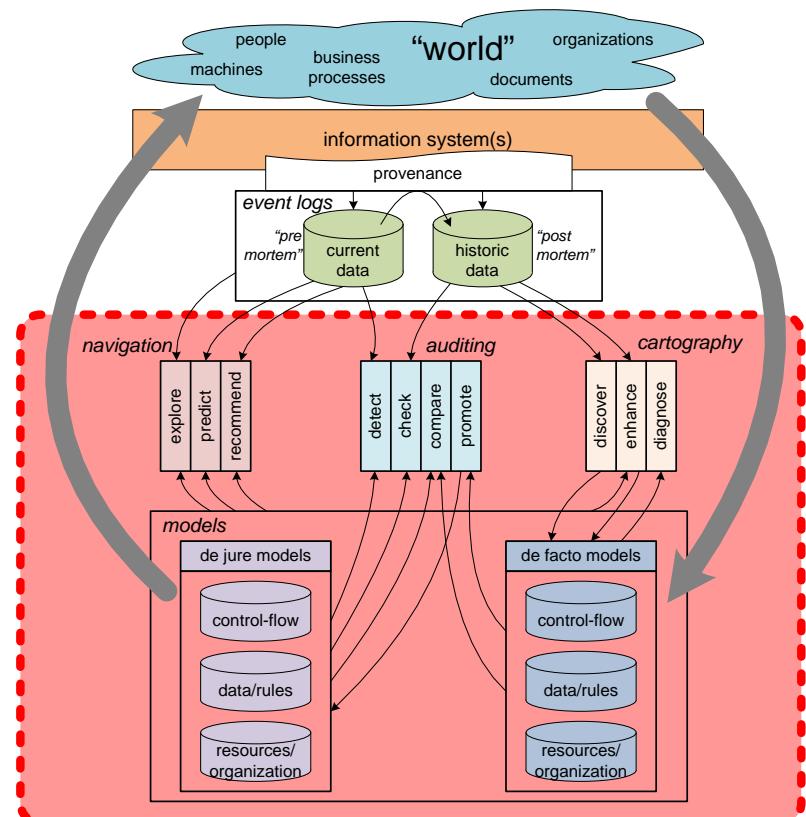


Checklist for a process mining tool



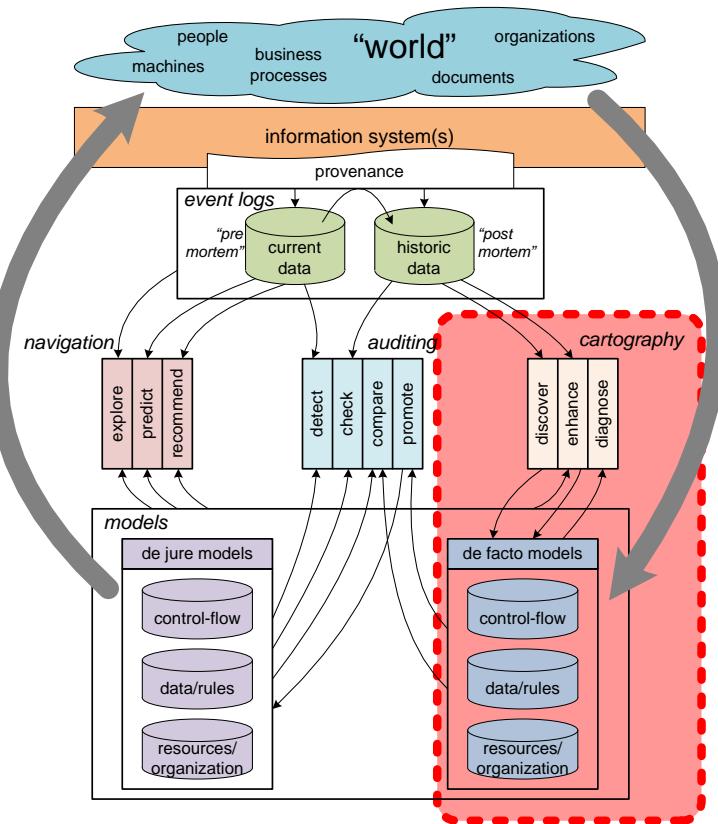
- **Which of the 10 process mining activities are supported?**
- **Which perspectives are supported?**
- **What modeling notations are supported?**
- **What import and export formats are supported (event logs, models, alignments, etc.)?**

Example: ProM



- **Aims to cover the whole process mining spectrum.**
- **Notations supported: Petri nets (many types), BPMN, C-nets, fuzzy models, transition systems, Declare, etc.**
- **Also supports conformance checking and operational support.**
- **Many plug-ins are experimental prototypes and not user friendly.**

Example: Disco



- Focus on discovery and performance analysis (including animation).
- Powerful filtering capabilities for comparative process mining and ad-hoc checking of patterns.
- Uses a variant of fuzzy models, etc.
- Does not support conformance checking and operational support.
- Easy to use and excellent performance.

Other commercial tools (1/2)

- **Perceptive Process Mining** (before Futura Reflect and BPM|one) (Perceptive Software)
- **ARIS Process Performance Manager** (Software AG)
- **QPR ProcessAnalyzer** (QPR)
- **Celonis Discovery** (Celonis)
- **Interstage Process Discovery** (Fujitsu)
- **Discovery Analyst** (StereoLOGIC)
- **XMAalyzer** (XMPro)
- ...

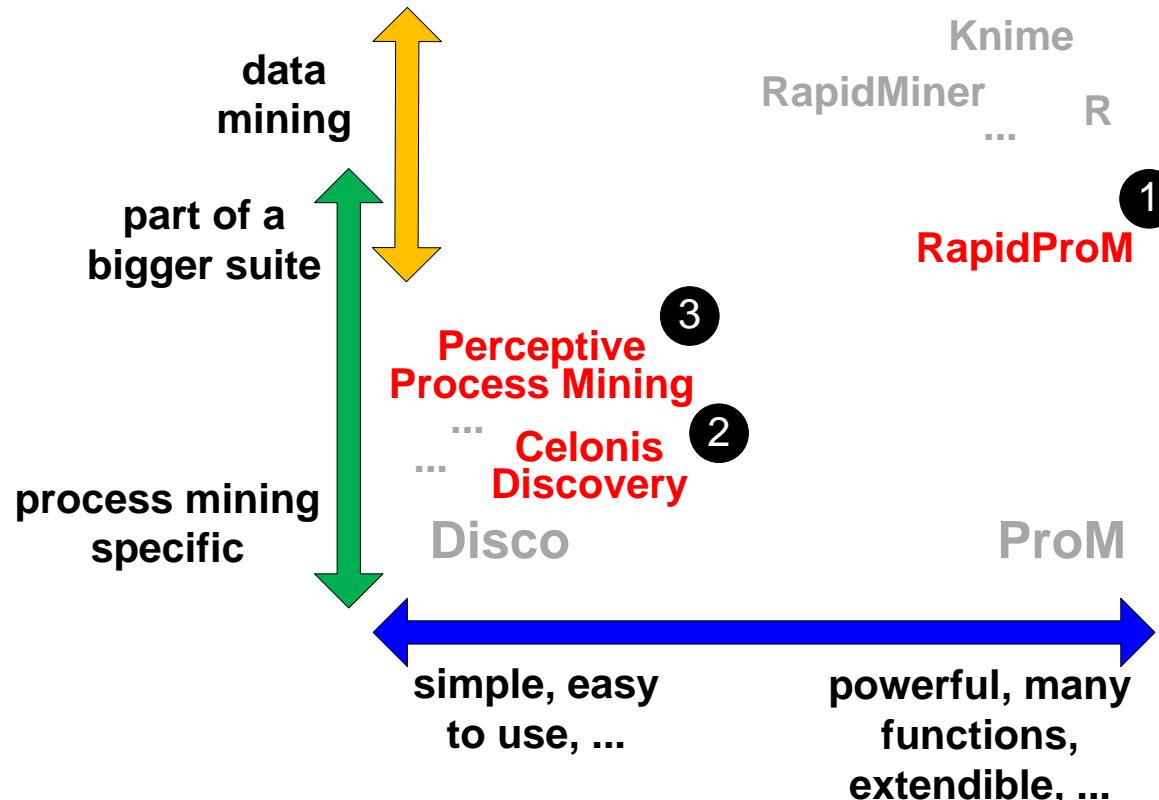


Other commercial tools (2/2)

- Most tools are similar to Disco in terms of scope (but often less user-friendly and less performant).
- Typical coverage:

| | control-flow only | control-flow and ... | | | |
|--|----------------------|----------------------|-----------|------|------|
| | | time | resources | data | |
| discovery  | ✓ | ✓ | ✗ | ✗ | ✗ |
| conformance  | ✗ | ✗ | ✗ | ✗ | ✗ |
| enhancement  | ✗ | ✓ | ✗ | ✗ | ✗ |

Three examples (next to ProM and Disco)



RapidProM

Rapid-I Market Place

https://marketplace.rapid-i.com/UpdateServer/faces/product_details.xhtml?productId=rmx_prom

Apps Homepage of Wil van... Process Mininghttp... Wil van der Aalst - G... DBLP: Wil M. P. v... Architecture of Infor... Google

Marketplace

Home Categories Contact Help Login

RapidProM – ProM Framework Extension

RapidMiner World 2014

RapidMiner Conference

RapidMiner World 2014

The ProM framework and toolset has become the de facto standard for process mining. ProM is a plug-able environment for process mining using XMMML, SA-XMMML, or XES as input format. Now the ProM framework and the RapidMiner data analysis solution are connected. As such any discovery, conformance, or extension algorithm of ProM can be used within a RapidMiner analysis process or a dedicated process mining analysis can be constructed. For more information see <http://rapidprom.org>.

Vendor: **ronny mans**
 Last Update: **7/22/14 1:54 PM**
 Product web site <http://rapidprom.org/>
 Category: **Operators**
 Product ID: **rmx_prom**
 Namespace: **prom**

Today's Information Systems (ISs) record huge amounts of data about the business processes they support. These data can be used for process mining. This way we can analyse the operational processes within an organization based on facts rather than fiction. Examples of these processes are the handling of a loan application within a bank or the treatment of a patient suffering from colorectal cancer within a hospital.

Currently, process mining is gaining more and more attention both in industry and practice. As such, the number of process mining tools is steadily increasing. However, none of these tools allow for composing and executing analysis workflows consisting of multiple process mining algorithms. As a result, the analyst needs to perform repetitive process mining tasks manually and scientific process experiments are extremely labor intensive.

To this end, we have connected RapidMiner, which allows for the definition and execution of analysis workflows, with the process mining framework ProM 6. As such any discovery, conformance, or extension algorithm of ProM 6 can be used within a RapidMiner analysis process thus supporting process mining workflows.

For more information see <http://rapidprom.org/> and <http://processmining.org/>.

Download

| Platform | Version | Release date | File size | License |
|----------|---------|-----------------|-----------|---------|
| ANY | 1.0.10 | 7/22/14 1:54 PM | 19 MB | AGPL |

[Download statistics](#)
 Total: 1005; This week: 44; Today: 8

 14 times bookmarked

Top Links

Latest Updates

www.rapidprom.org

< new process > RapidMiner 5.3.015 @nbin1027

File Edit Process Tools View Help

Operators Search Process XML

Main Process

Process

XML

Search

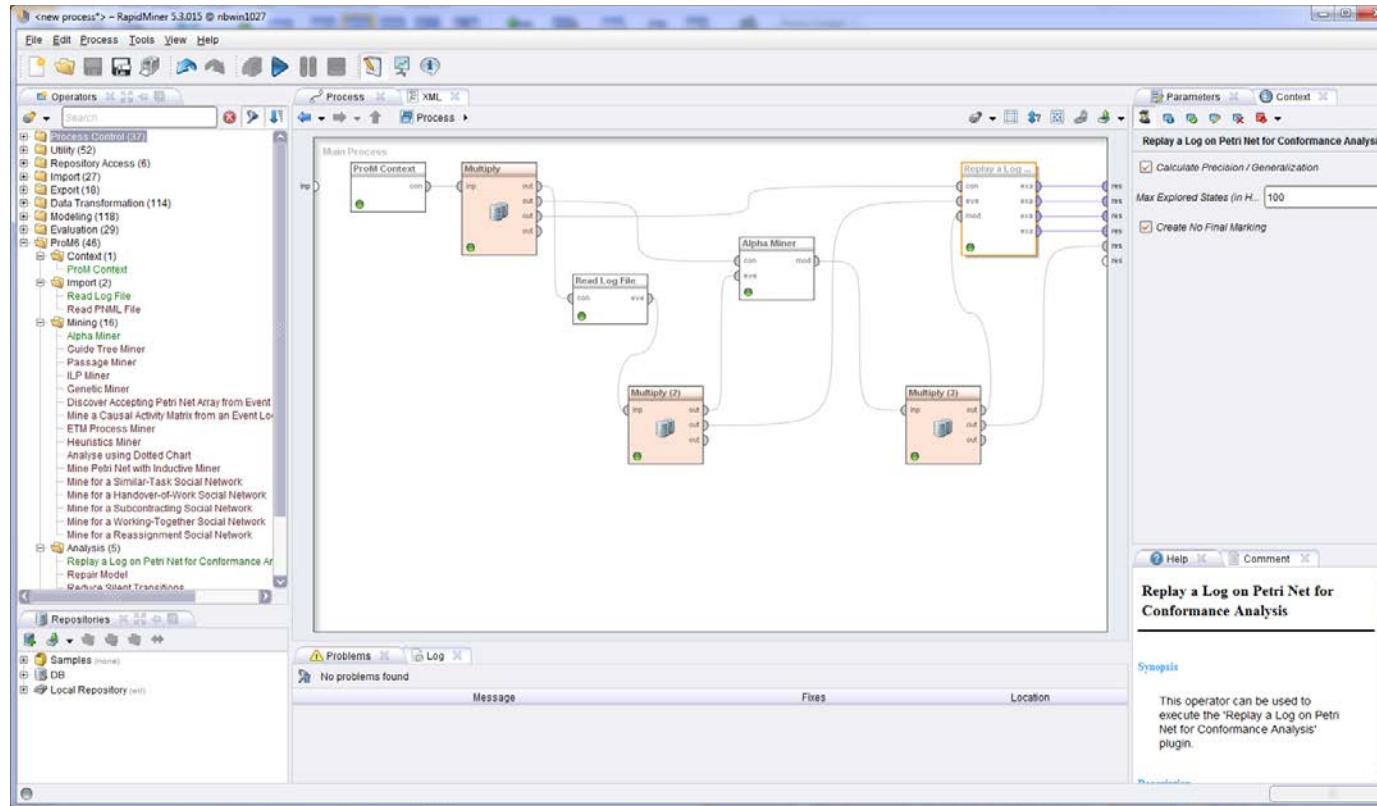
- Process Control (37)
- Utility (52)
- Repository Access (6)
- Import (27)
- Export (18)
- Data Transformation (114)
- Modeling (118)
- Evaluation (29)
- ProM6 (46)
 - Context (1)
 - Import (2)
 - Read Log File
 - Read PNMML File
 - Mining (16)
 - Alpha Miner
 - Guide Tree Miner
 - Passage Miner
 - ILP Miner
 - Genetic Miner
 - Discover Accepting Petri Net Array from Event Log Array
 - Mine a Causal Activity Matrix from an Event Log
 - ETM Process Miner
 - Heuristics Miner
 - Analyse using Dotted Chart
 - Mine Petri Net with Inductive Miner
 - Mine for a Similar-Task Social Network
 - Mine for a Handover-of-Work Social Network
 - Mine for a Subcontracting Social Network
 - Mine for a Working-Together Social Network
 - Mine for a Reassignment Social Network
 - Analysis (5)
 - Replay a Log on Petri Net for Conformance Analysis
 - Repair Model
 - Reduce Silent Transitions
 - Replay an Event Log Array on an Accepting Petri Net Array
 - Replay a Log on Petri Net for Performance / Conformance Analysis
 - Export (2)
 - Filtering (3)
 - Conversion (17)
 - Convert ProM Log into Exampleset
 - Convert Exampleset into Log
 - Convert Logs into Event Log Array
 - Convert Heuristics Net into Petri Net
 - Convert Causal Activity Graph to an Activity Cluster Array
 - Convert Causal Activity Matrix to an Causal Activity Graph
 - Convert Petri Net to an Accepting Petri Net
 - Decompose Accepting Petri Net using an Activity Cluster Array
 - using an Activity Cluster Array
 - array from Accepting Petri Net Array
 - Net Array into an Accepting Petri Net

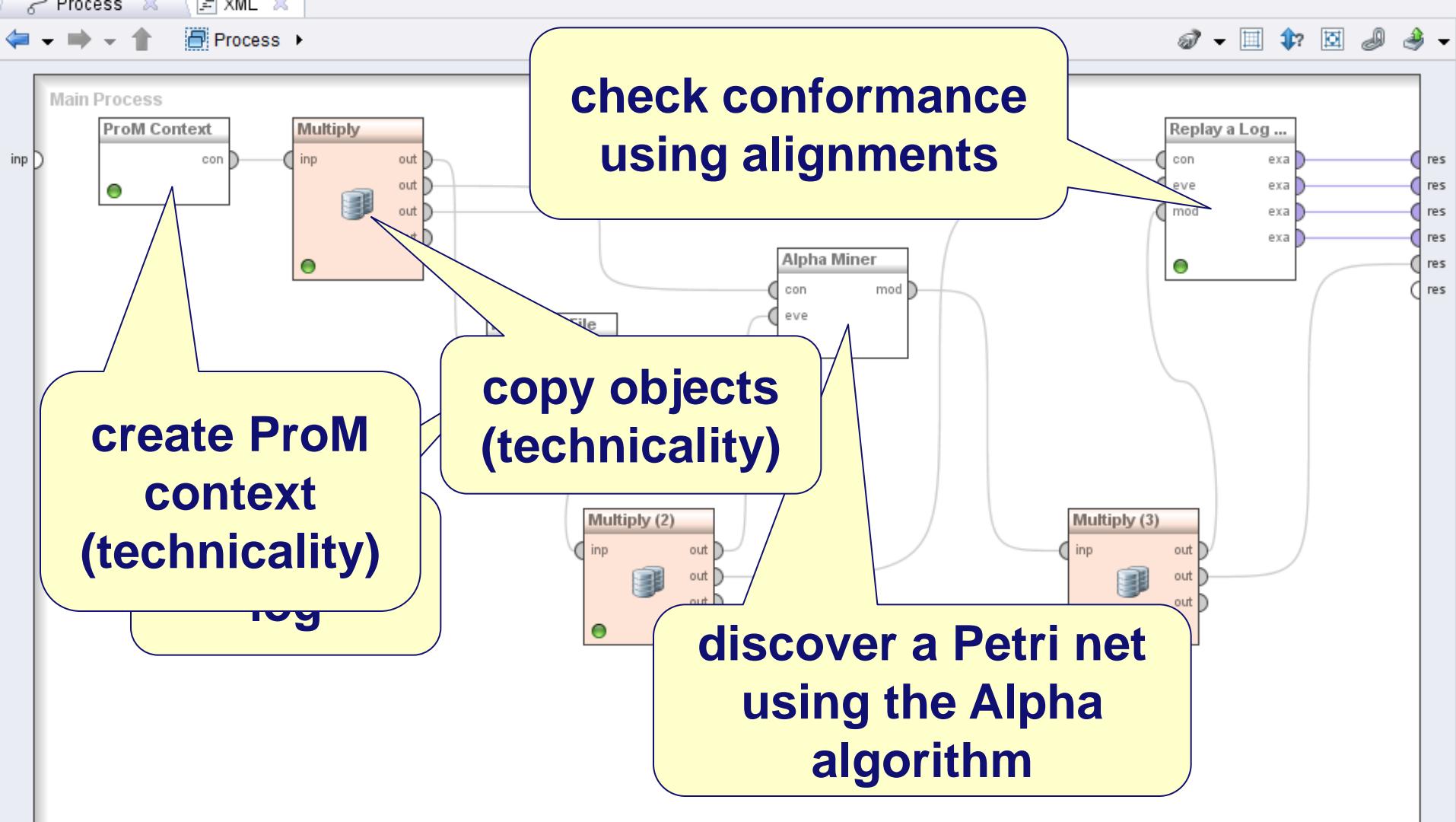
Problems Log

No problems found

Message

Create a process mining workflow





<new process*> - RapidMiner 5.3.015 @ nbwin1027

File Edit Process Tools View Help

PIP Zoom

```

graph LR
    Start(( )) --> Invite[invite reviewers+com]
    Invite --> Invited(( ))
    Invited --> R1(( ))
    Invited --> R2(( ))
    Invited --> R3(( ))
    R1 --> Get1[get review 1+complete]
    R1 --> TimeOut1[time-out 1+complete]
    Get1 --> Collect[collect reviews+com]
    TimeOut1 --> Collect
    R2 --> Get2[get review 2+complete]
    R2 --> TimeOut2[time-out 2+complete]
    Get2 --> Collect
    TimeOut2 --> Collect
    R3 --> Get3[get review 3+complete]
    R3 --> TimeOut3[time-out 3+complete]
    Get3 --> Collect
    TimeOut3 --> Collect
    Collect --> Decide[decide+com]
    Decide --> End(( ))
  
```

ExampleSet (Replay a Log on Petri Net for Conformance Analysis)

Result Overview

PetriNetOOObject (Alpha Miner)

Repositories

Samples (none)

DB

Local Repository (wi)

<new process*> - RapidMiner 5.3.015 @ nbwin1027

File Edit Process Tools View Help

PIP Zoom

ExampleSet (Replay a Log on Petri Net for Conformance Analysis)

Result Overview

ExampleSet (Replay a Log on Petri Net for Conformance Analysis)

Data View (selected)

Meta Data View

Plot View

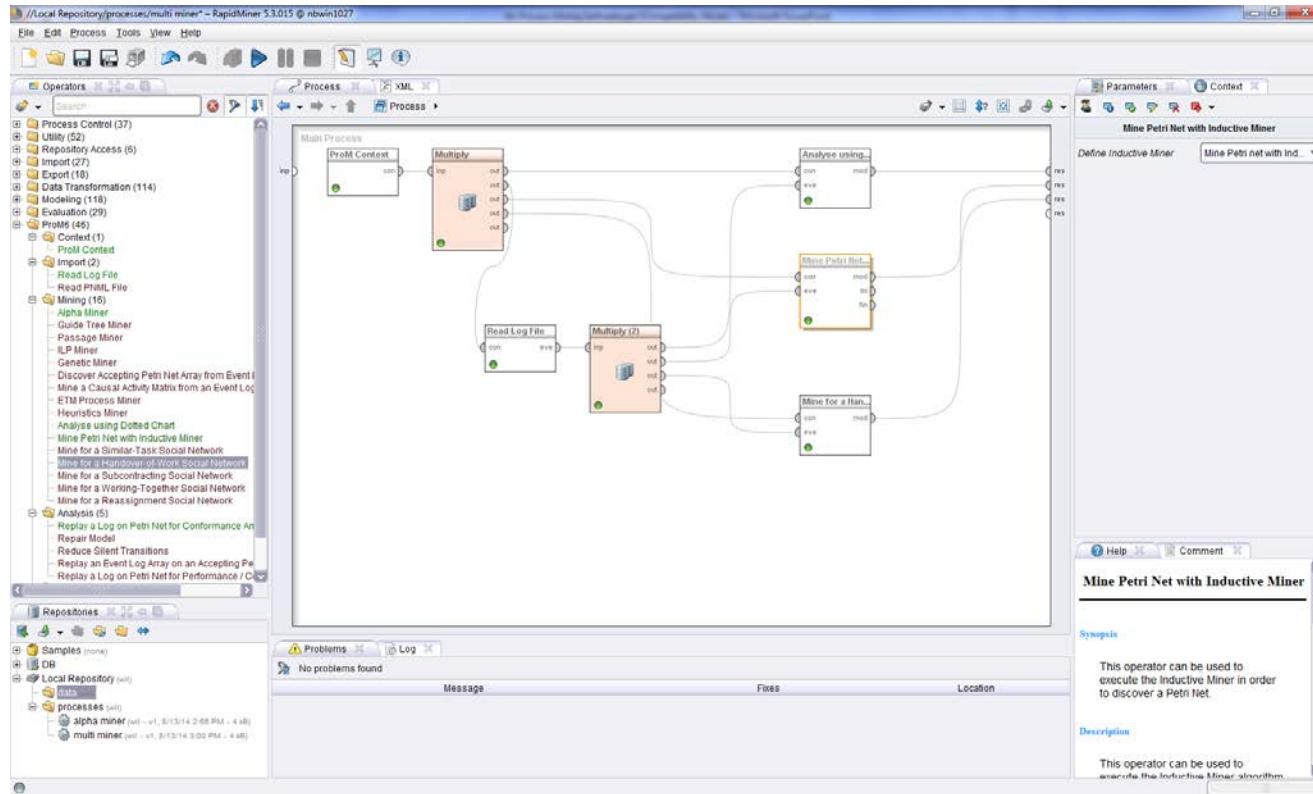
Advanced Charts

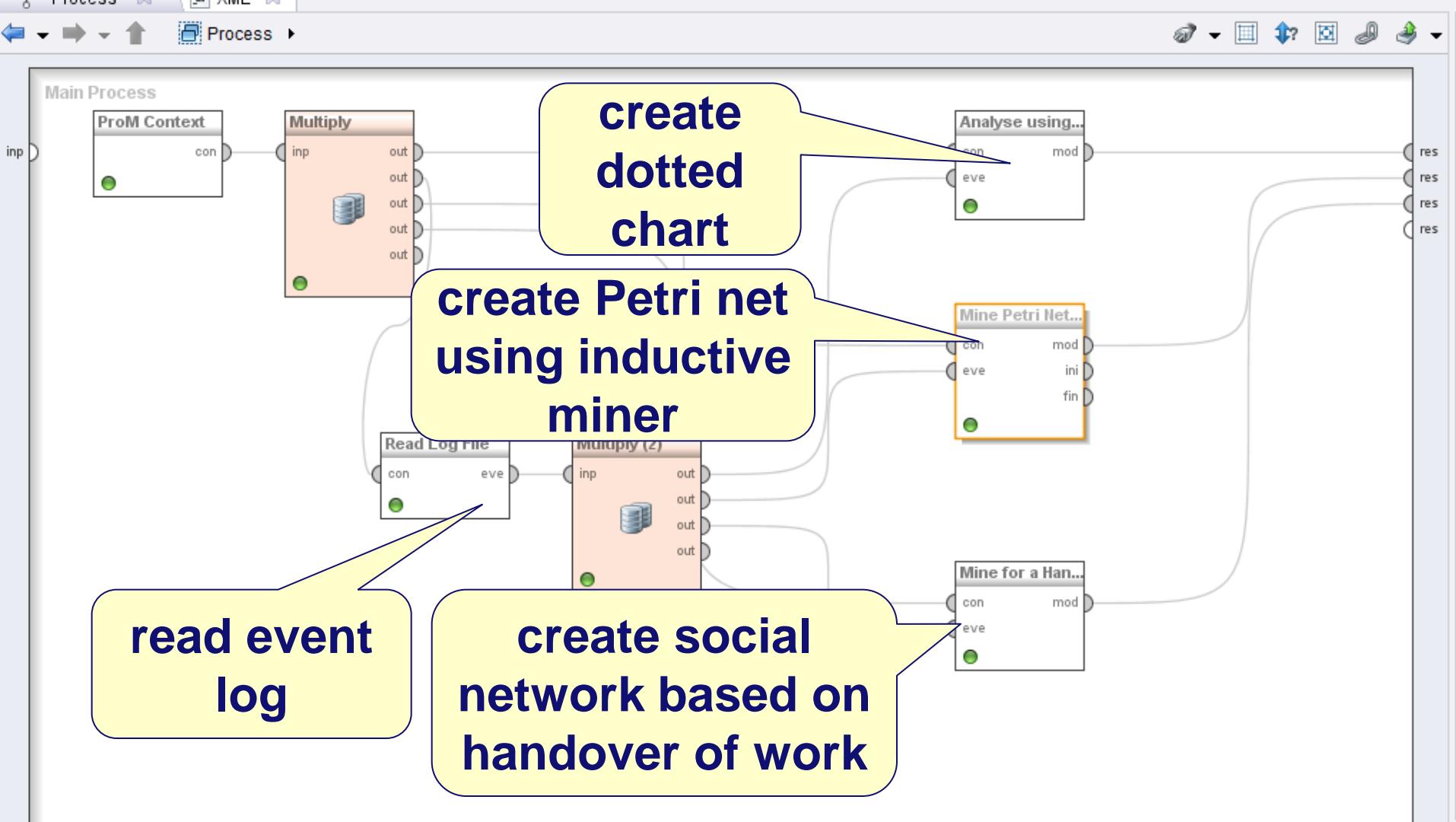
ExampleSet (8 examples, 0 special attributes, 2 regular attributes)

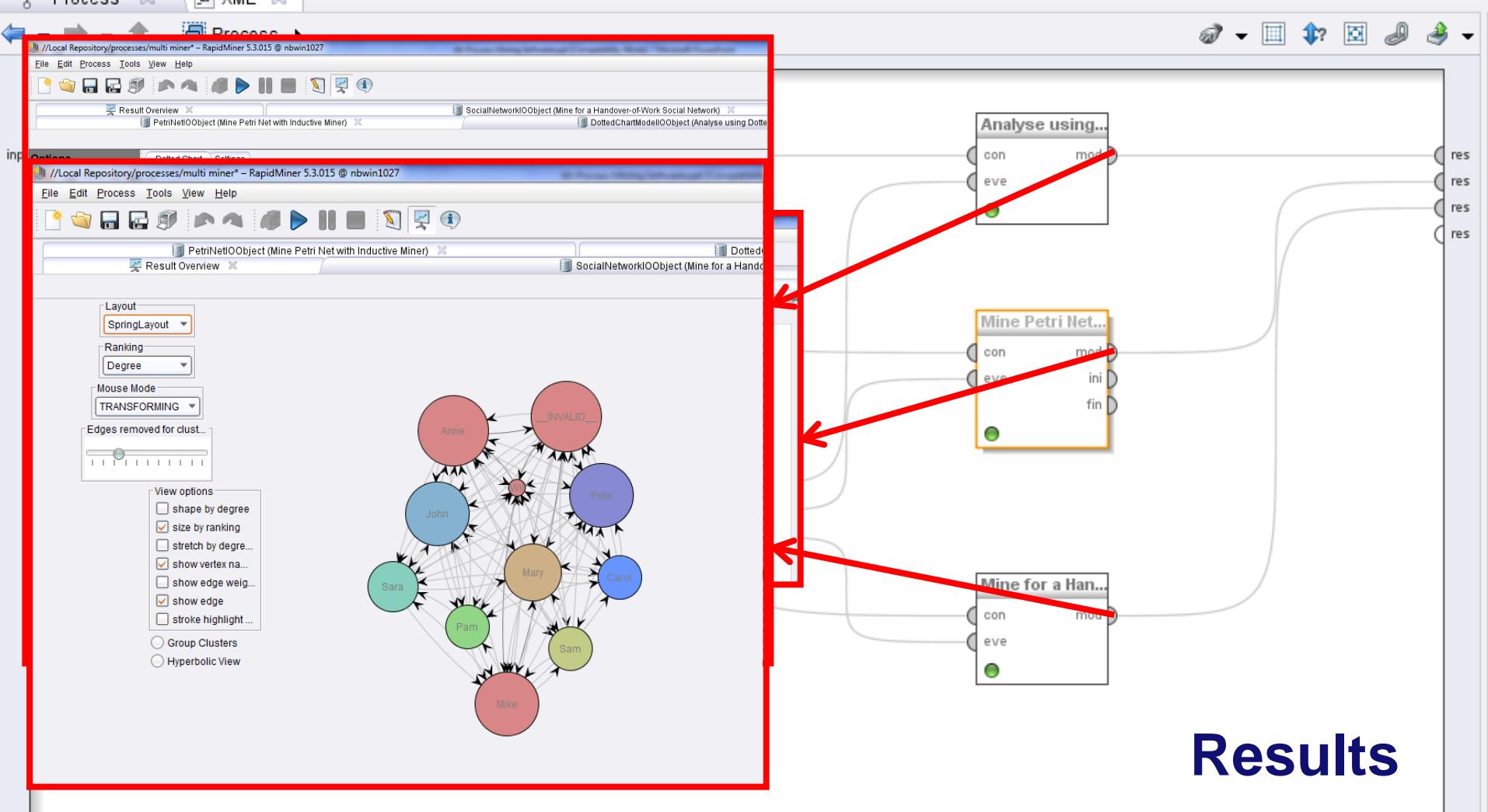
| Row No. | Name | Value |
|---------|--------------------|--------|
| 1 | Trace Fitness | 1 |
| 2 | Move-Log Fitness | 1 |
| 3 | Move-Model Fitness | 1 |
| 4 | Raw Fitness Cost | 0 |
| 5 | Num. States | 23.780 |
| 6 | Queued States | 78.600 |
| 7 | Generalization | 0.999 |
| 8 | Precision | 0.727 |

Export

Another process mining workflow







RapidProM

- Can be used to repeat process mining analyses (scientific experiments, reusable macros, ...).
- Can be used to combine process mining with data mining, text mining, etc.



Celonis Discovery

Process view (total number of cases: 100)

Coverage:

88



Align Process Export Open Process Analysis Save Save As Animate Root cause analysis

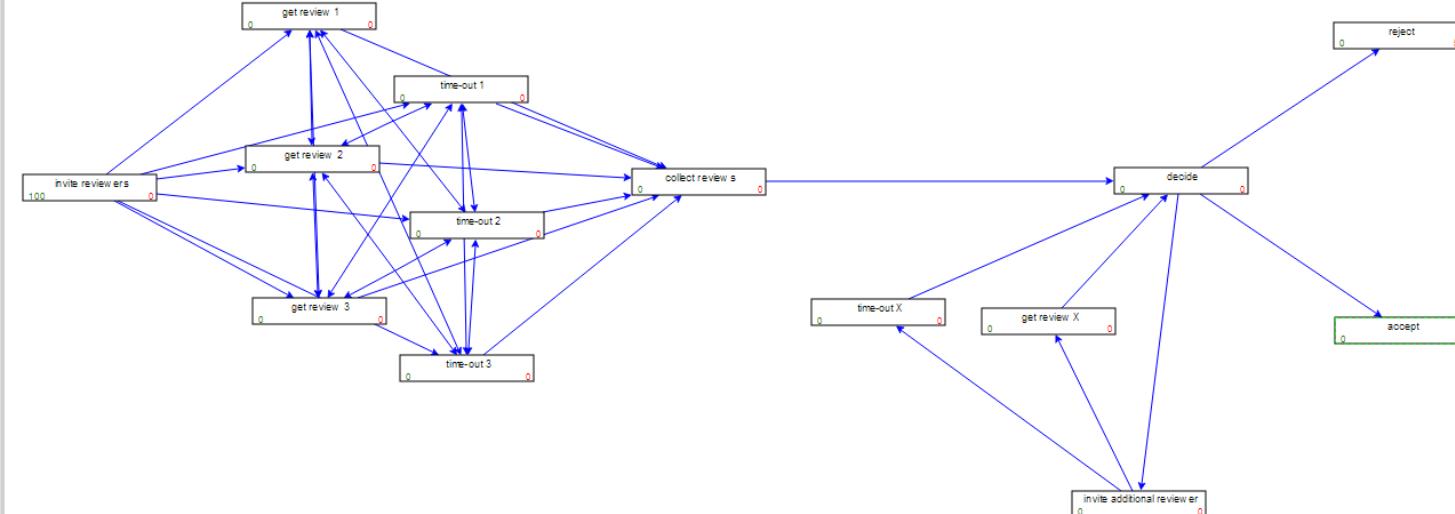
Process settings

General settings KPIs Notepad

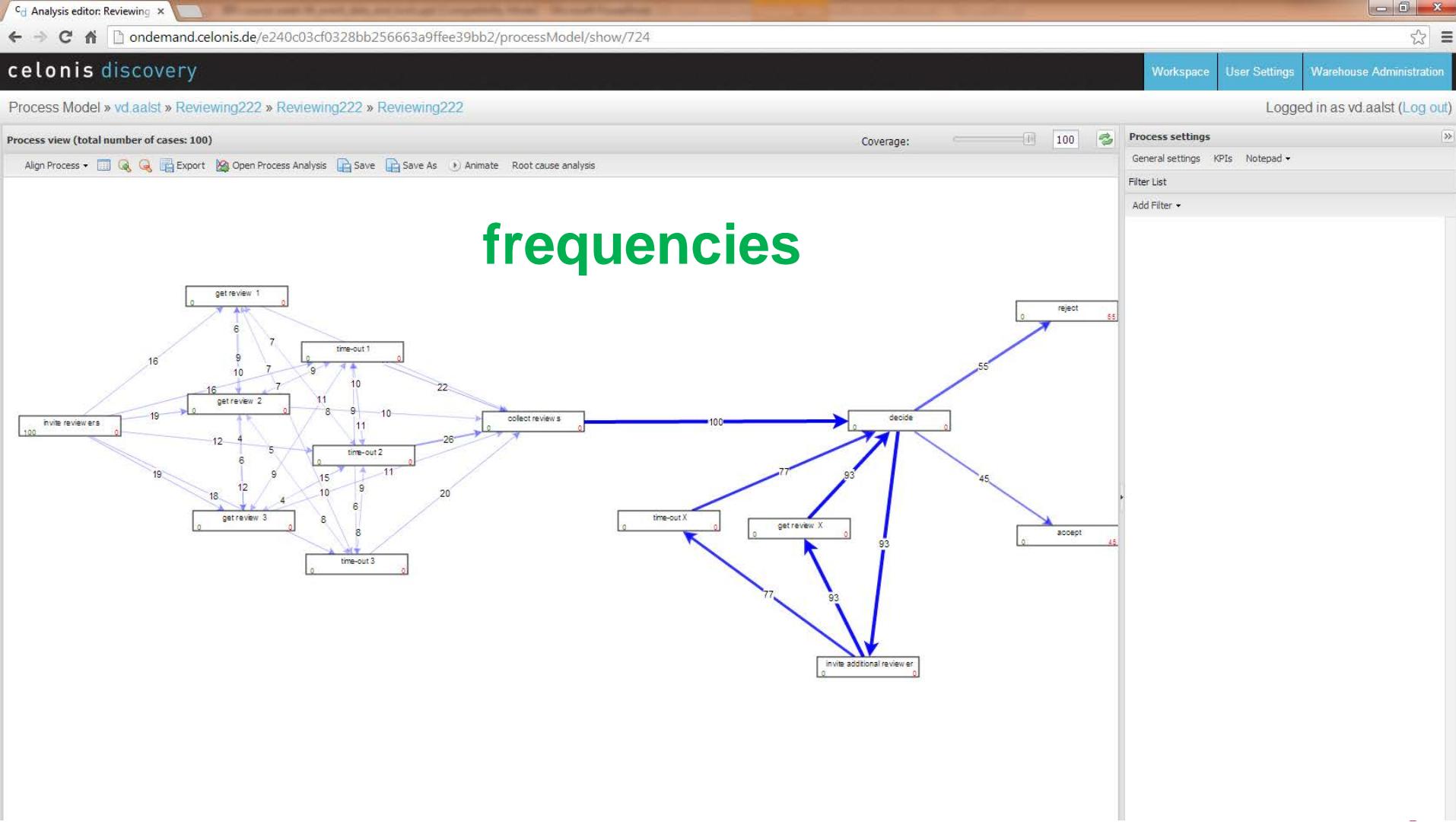
Filter List

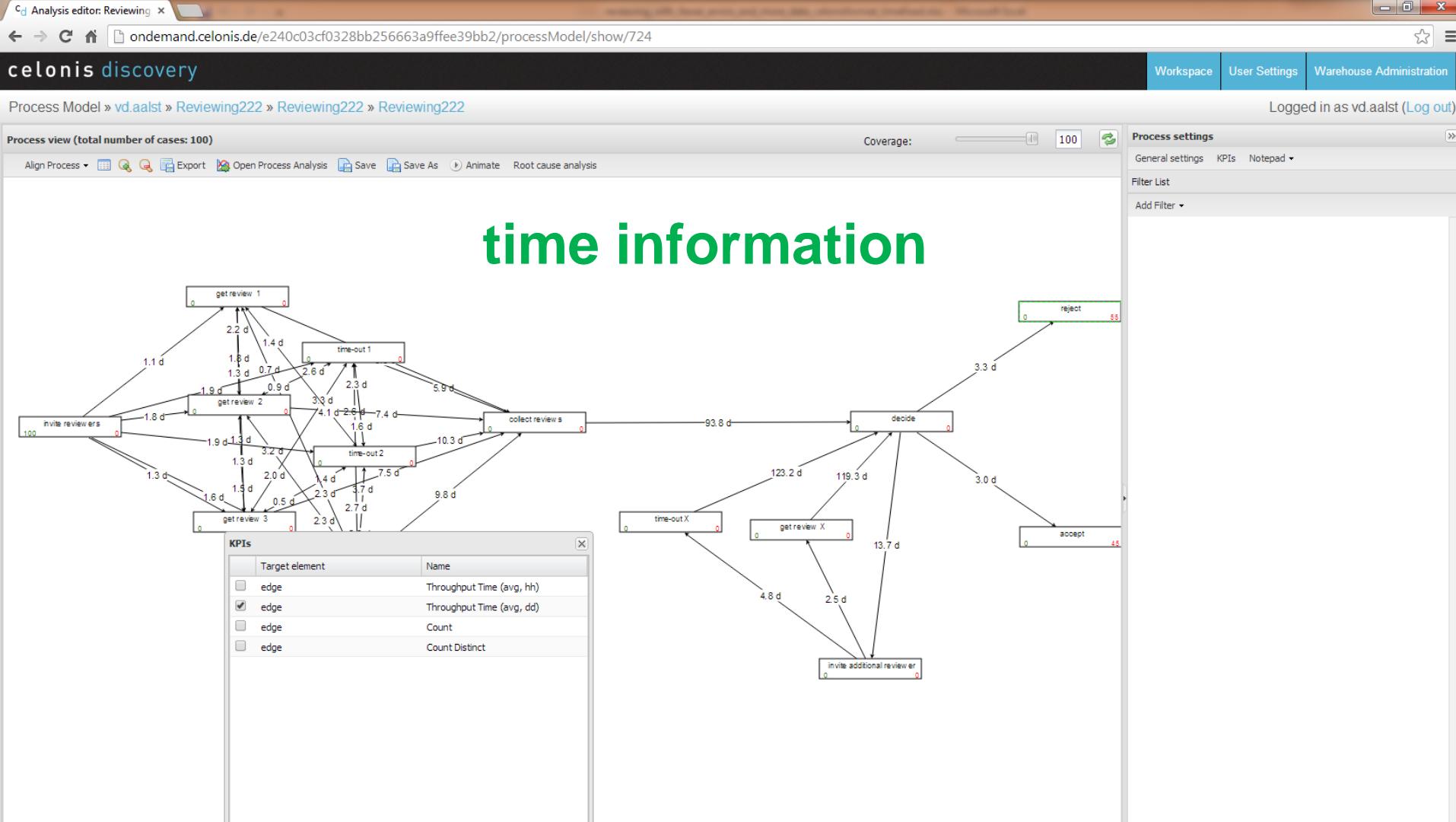
Add Filter

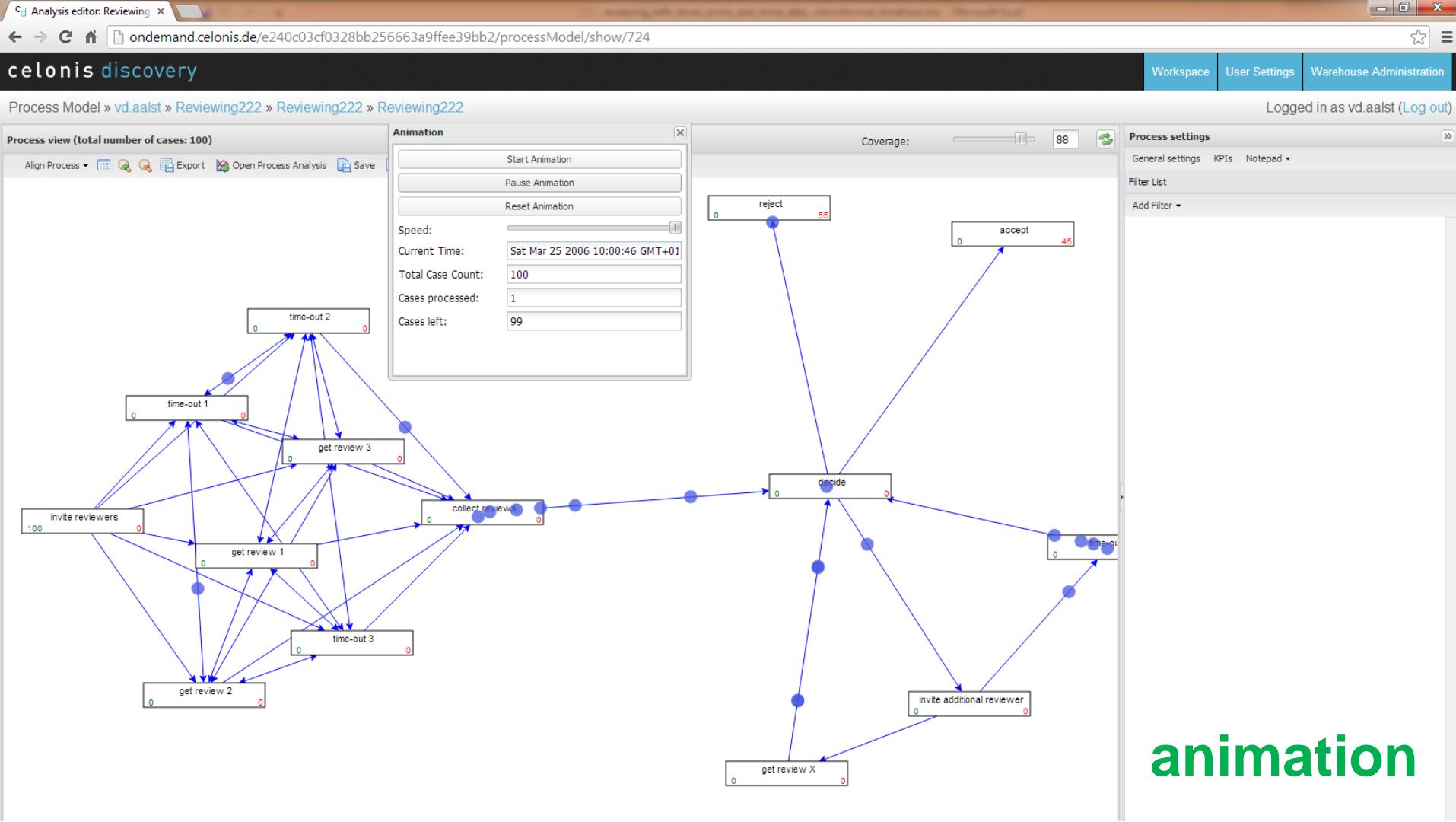
web-based

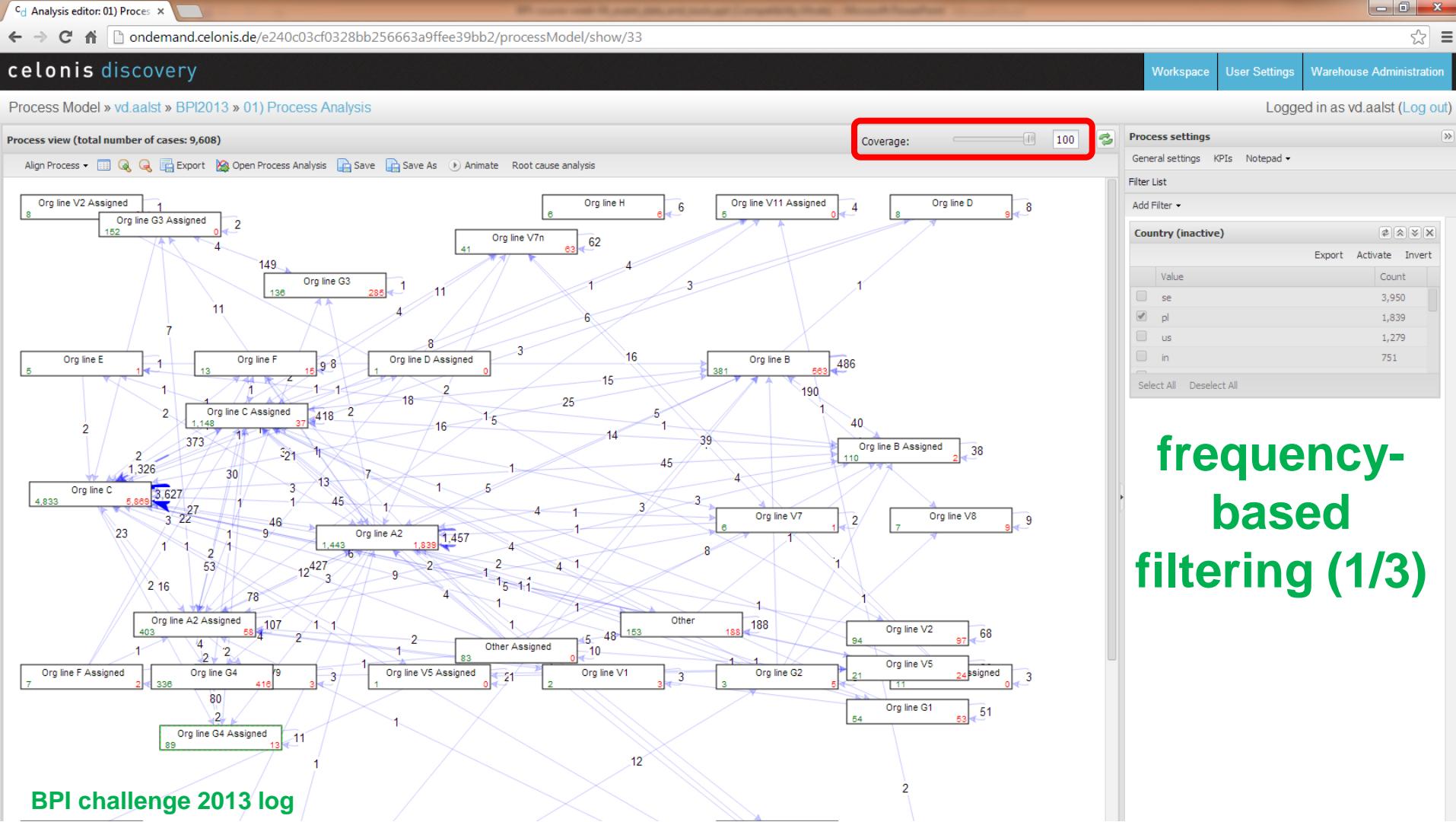


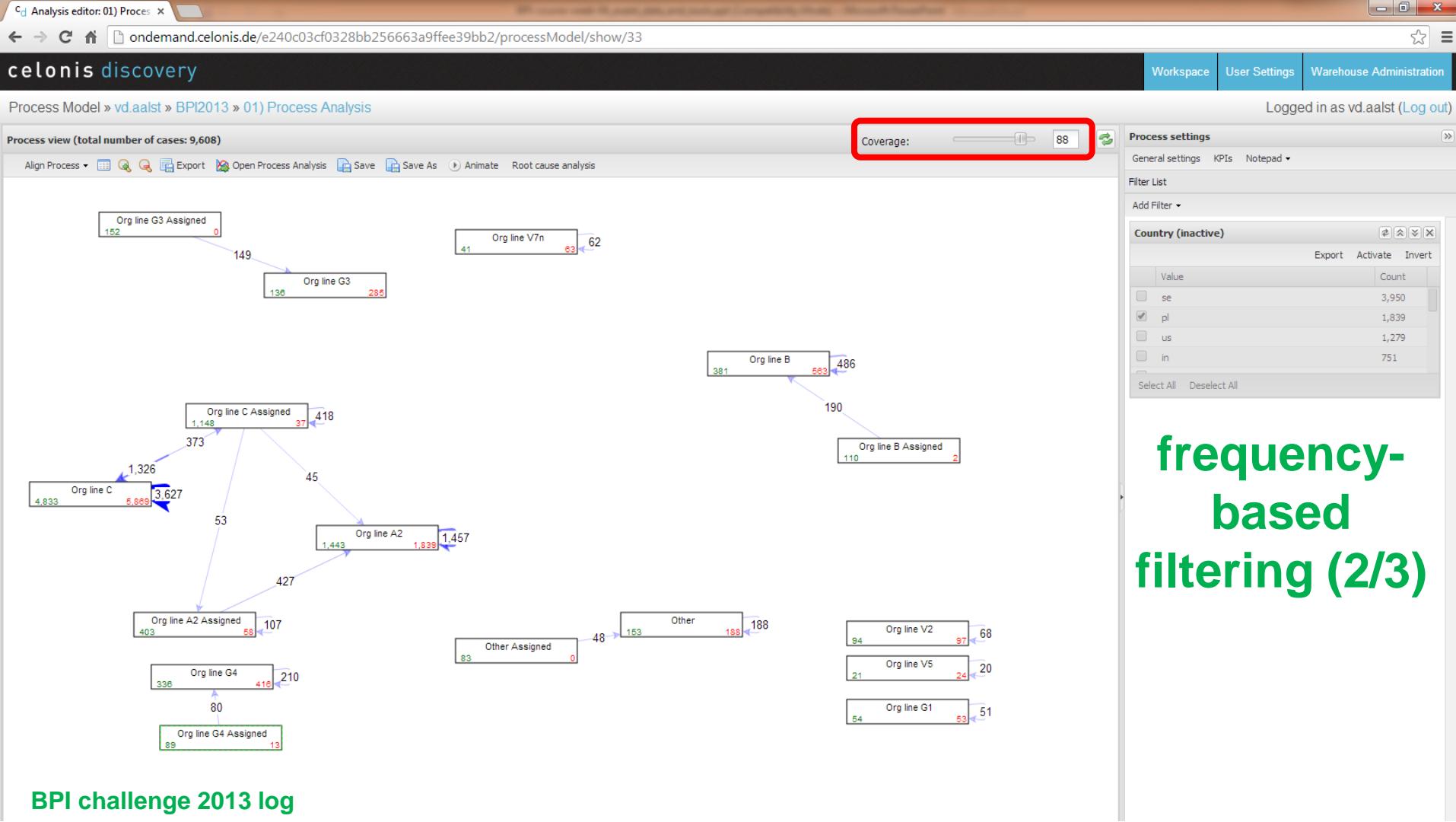
supports process discovery (but no concurrency)











Analysis editor: 01) Proces x

ondemand.celonis.de/e240c03cf0328bb256663a9ffee39bb2/processModel/show/33

celonis discovery

Process Model » vd.aalst » BPI2013 » 01 Process Analysis

Logged in as vd.aalst (Log out)

Process view (total number of cases: 9,608)

Coverage: 68

Align Process Export Open Process Analysis Save Save As Animate Root cause analysis

Org line G3: 136 285

Org line B: 381 486

Org line C Assigned: 1,148 37

Org line C: 4,833 5,627

Org line A2: 1,443 1,457

Org line G4: 336 210

Process settings

General settings KPIs Notepad

Filter List

Add Filter

Country (inactive)

| Value | Count |
|-------|-------|
| se | 3,950 |
| pl | 1,839 |
| us | 1,279 |
| in | 751 |

Export Activate Invert

Select All Deselect All

frequency-based filtering (3/3)

BPI challenge 2013 log

celonis discovery

Process Model » vd.aalst » BPI2013 » 01 Process Analysis

Logged in as vd.aalst (Log out)

Process view (total number of cases: 9,608)

Coverage: 68

Process settings

General settings KPIs Notepad Filter List

Process Analysis

Add OLAP component Refresh Open in Fullscreen Root cause analysis

Distribution by Country

| Country | count |
|---------|-------|
| se | 3,950 |
| pl | 1,839 |
| us | 1,279 |
| in | 751 |
| be | 572 |
| fr | 396 |
| br | 336 |
| cn | 160 |
| gb | 139 |
| kr | 69 |
| ca | 69 |
| nl | 63 |

Latest_Impact

| Latest_Impact | Average Processing Time [min] |
|---------------|-------------------------------|
| High | 52,734 |
| Medium | 45,415 |
| Major | 40,242 |
| Low | 19,295 |

Distribution of Latest_Impact

Major: 191 (1.94%)

Low: 3,600 (36.51%)

Medium: 5,306 (53.81%)

High: 763 (7.74%)

Legend: Medium (Dark Red), High (Orange), Low (Green), Major (Dark Green)

Distribution by Number of different Users

Count

| Number of Users | Count |
|-----------------|-------|
| 1 | 3,500 |
| 2 | 3,500 |
| 3 | 1,200 |
| 4 | 400 |
| 5 | 100 |
| 6 | 50 |
| 7 | 20 |
| 8 | 10 |
| 9 | 5 |
| 10 | 2 |
| 11 | 1 |
| 12 | 1 |
| 13 | 1 |
| 14 | 1 |
| 15 | 1 |
| 16 | 1 |
| 17 | 1 |
| NULL | 100 |

performance-related dashboards

Perceptive Process Mining

perceptive process mining

w.m.p.v.d.aalst@tue.... ▾ ? Help ▾

Reviewing222

Switch ▾ + Pencil

Actions

Overview

Mine

Replay

Chart

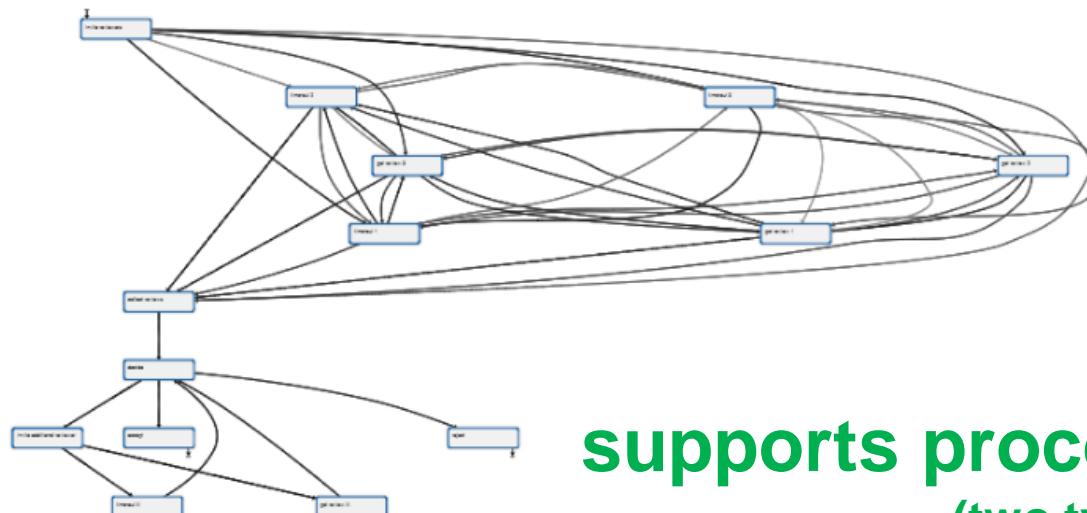
Objects

Models (0)

Mine

Save Save as Download as Flash Download path information

web-based



Settings

Use the slider to simplify the model by including fewer cases:

100%

Attribute Activity

Layout Process Model

Compare None

Performance

Miner

 Fast

Uses a miner that is very fast, but does not mine parallel execution of tasks.

 Thorough

Uses a miner that can mine parallelism, but takes longer to run.

Description

supports process discovery
(two types)

Actions

 Overview

 Mine

 Replay

 Chart

Objects

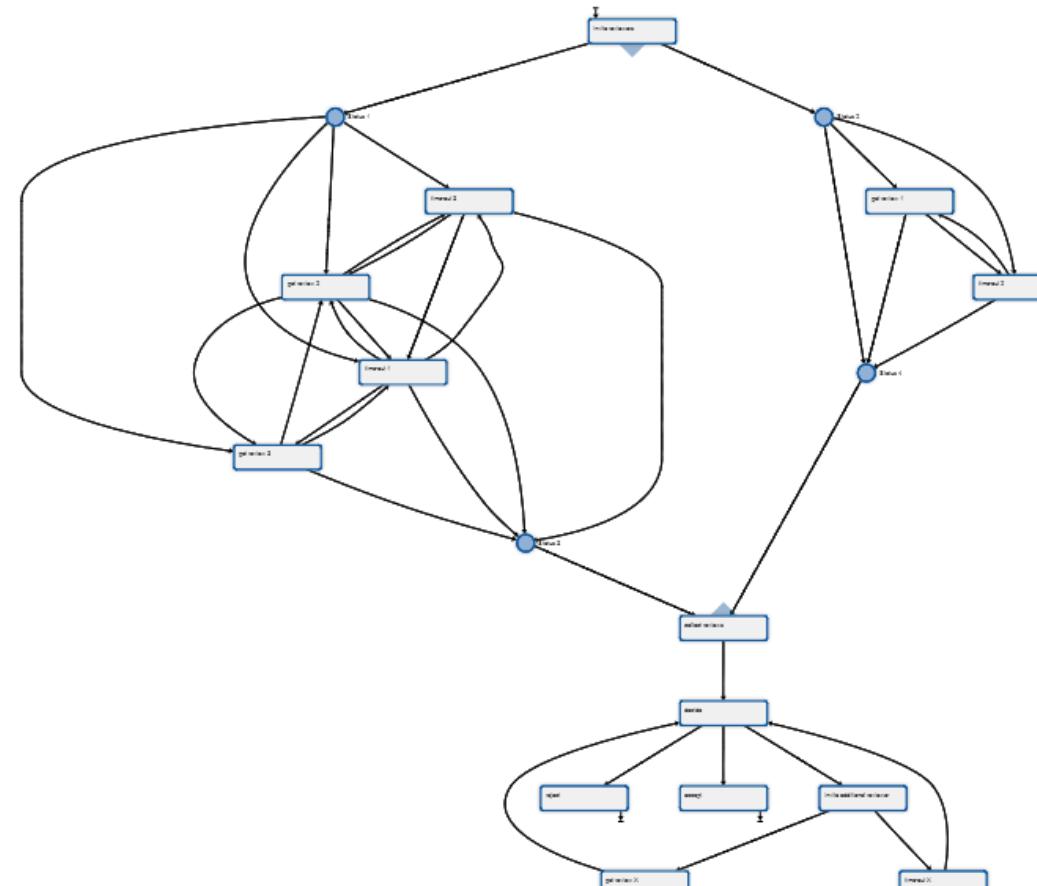
 Models (0)

Import model

 Replay list (0)

 Charts (0)

 Dashboards (0)



Completeness: 100.00%

Settings

Use the slider to simplify the model by including fewer cases:

100%

Attribute

Activity



Layout

Process Model

Miner

Fast

Uses a miner that is very fast, but does not mine parallel execution of tasks.

Thorough

Uses a miner that can mine parallelism, but takes longer to run.

Description

Stop mining

perceptive process mining

Reviewing222

Mine

Save Save as Download as Flash Download path information

Actions

Overview

Mine

Replay

Chart

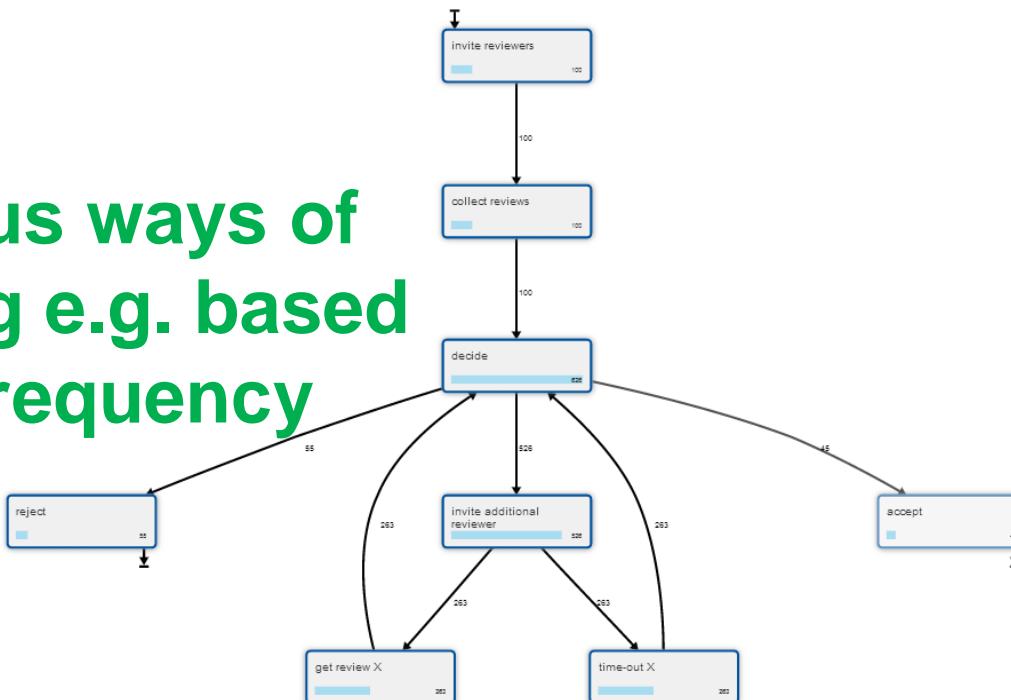
Objects

Models (2)

Reviewing222-par

Reviewing222seq

various ways of filtering e.g. based on frequency



Import model

Reply list (0)

Settings

Use the slider to simplify the model by including fewer cases:

Attribute Activity

Layout Process Model

Compare None

Performance

 Show performance metrics

Per Activity

Count Activity

Do not us Activity

Per arrow

Count Wait + Proc. Til

Do not us Wait + Proc. Til

Arrow width Use second me

Miner

Description

Filtering

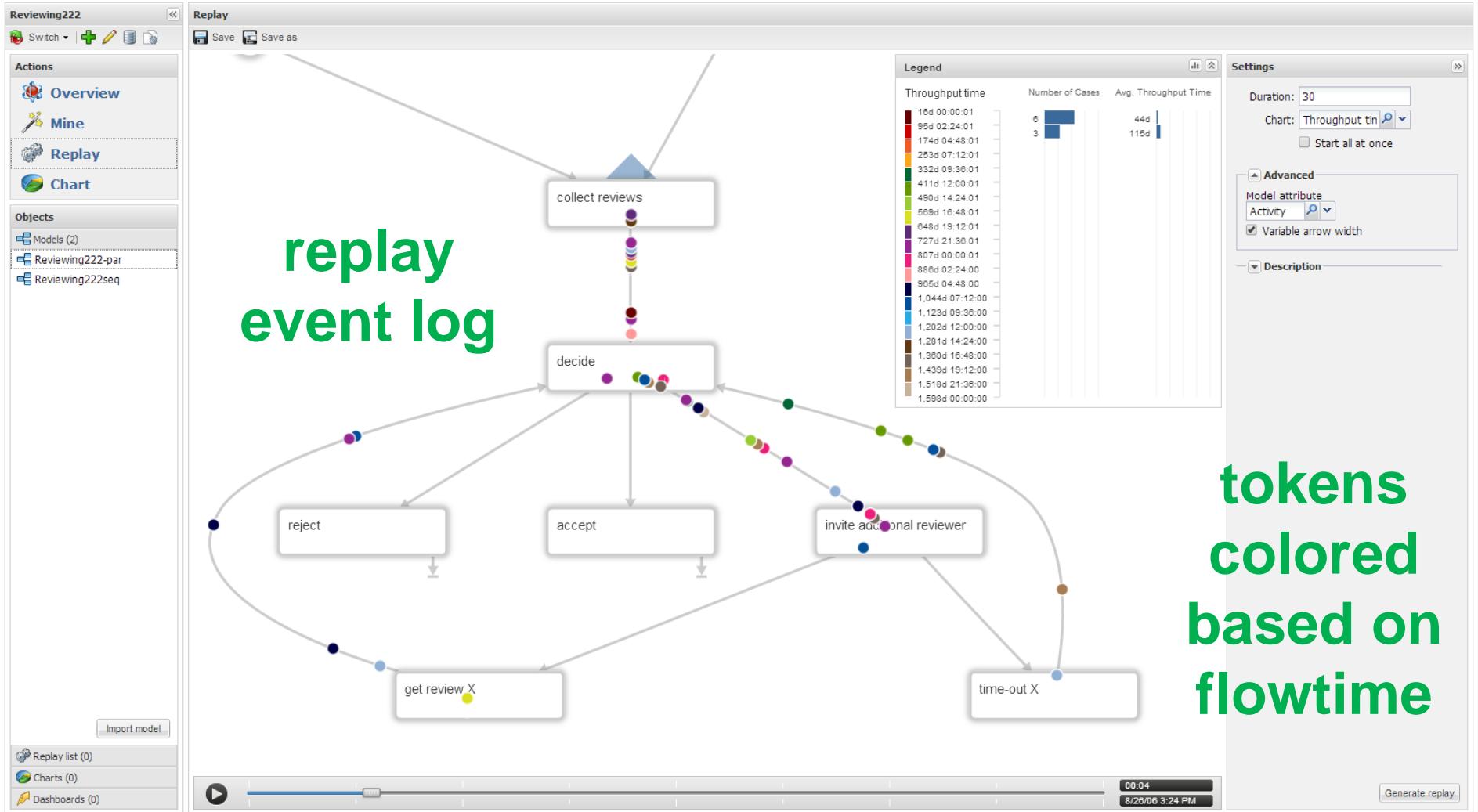
Switch

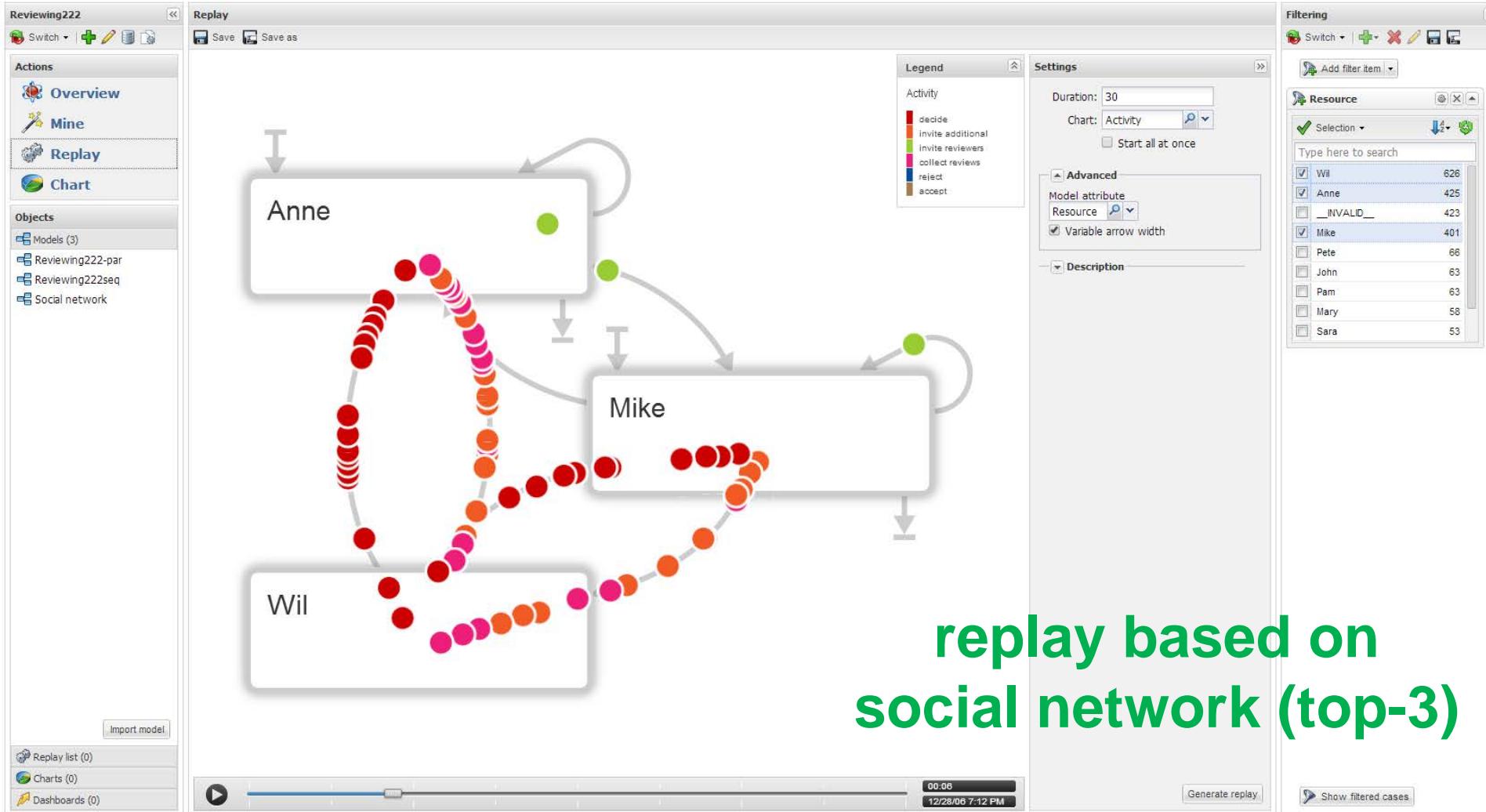
Add filter item

Activity

Type here to search

| | | |
|-------------------------------------|------------------|-----|
| <input checked="" type="checkbox"/> | invite reviewers | 100 |
| <input checked="" type="checkbox"/> | reject | 55 |
| <input type="checkbox"/> | time-out 1 | 55 |
| <input type="checkbox"/> | time-out 2 | 55 |
| <input type="checkbox"/> | get review 3 | 50 |
| <input type="checkbox"/> | time-out 3 | 50 |
| <input checked="" type="checkbox"/> | accept | 45 |
| <input type="checkbox"/> | get review 1 | 45 |
| <input type="checkbox"/> | get review 2 | 45 |







Actions

Overview

Mine

Replay

Chart

Objects

Models (0)

Import model

Replay list (0)

Charts (0)

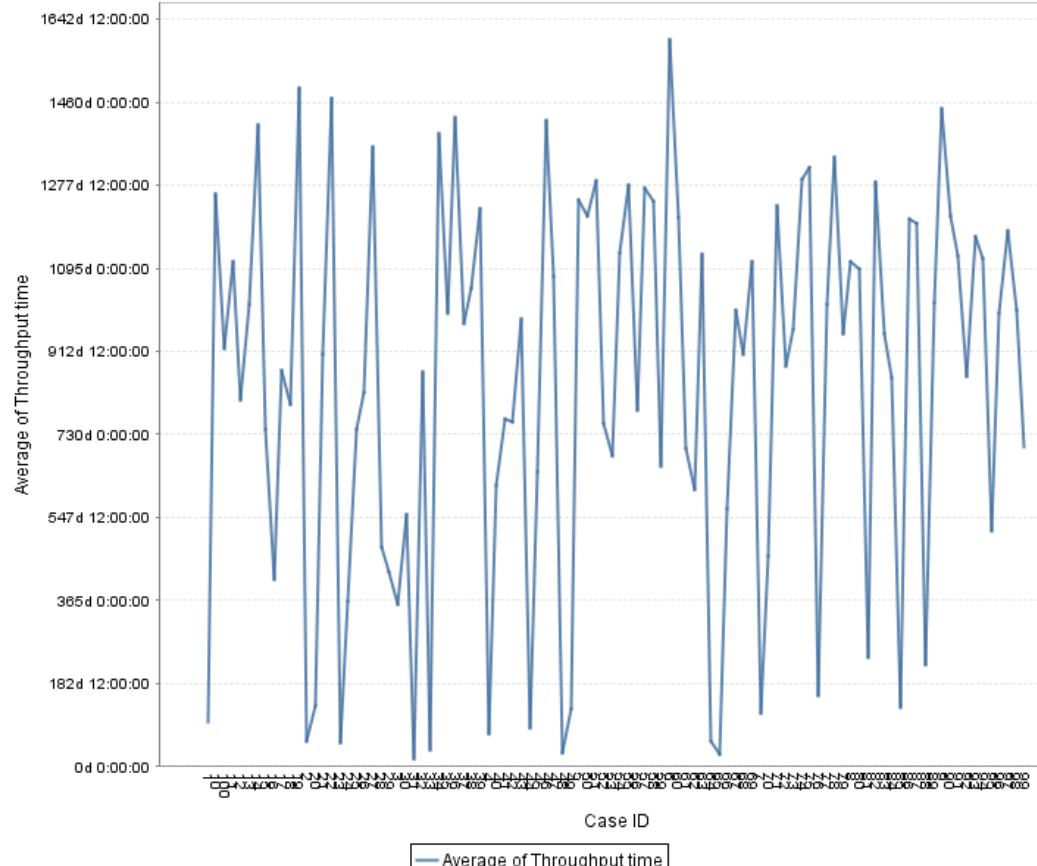
Dashboards (0)

Chart

Save | Save as | Download

[Pie Chart](#) [Line Chart](#) [Bar Chart](#) [Area Chart](#) [Stacked Areas](#)

Flow time per case



X-Axis

Attribute

Split on attribute

Data Series

Average

Chart

Chart title

Data levels: Cases

Update

performance-related charts

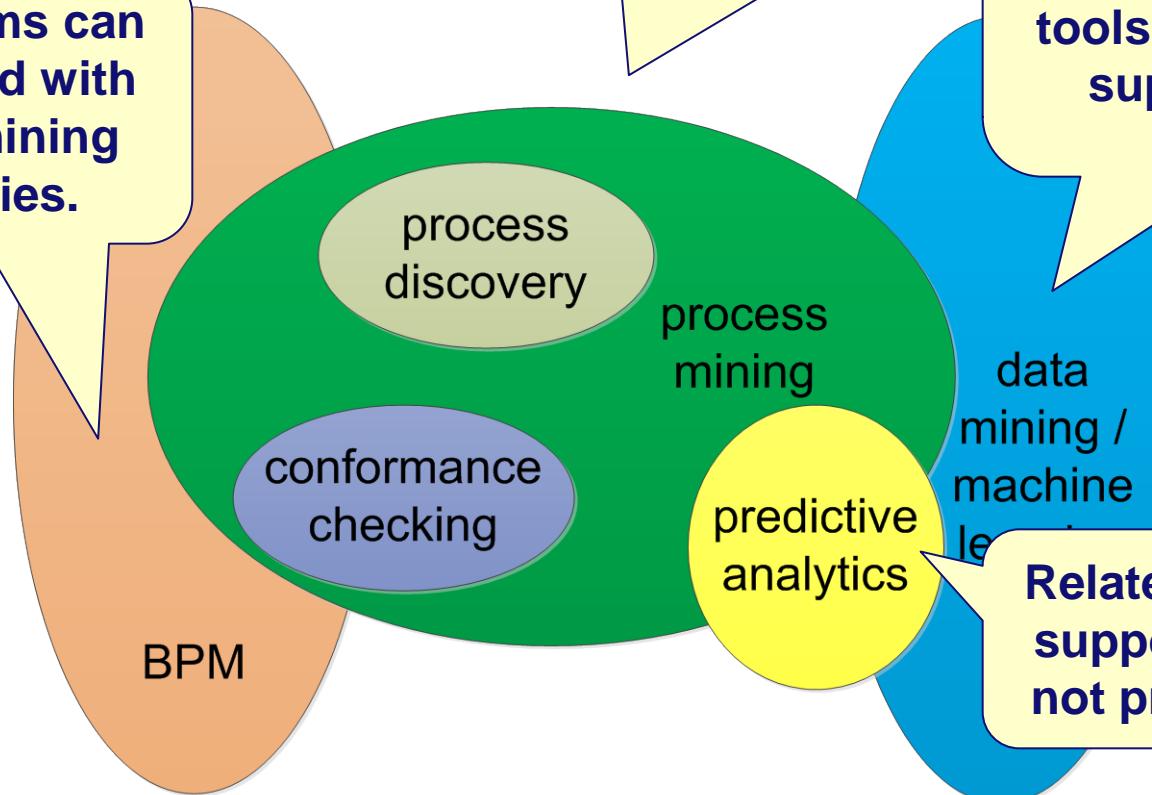
Related tools

(beyond pure process mining to)

BPM systems can be extended with process mining capabilities.

How about BI tools?

Although related, classical data mining and machine learning tools typically do not support process mining.

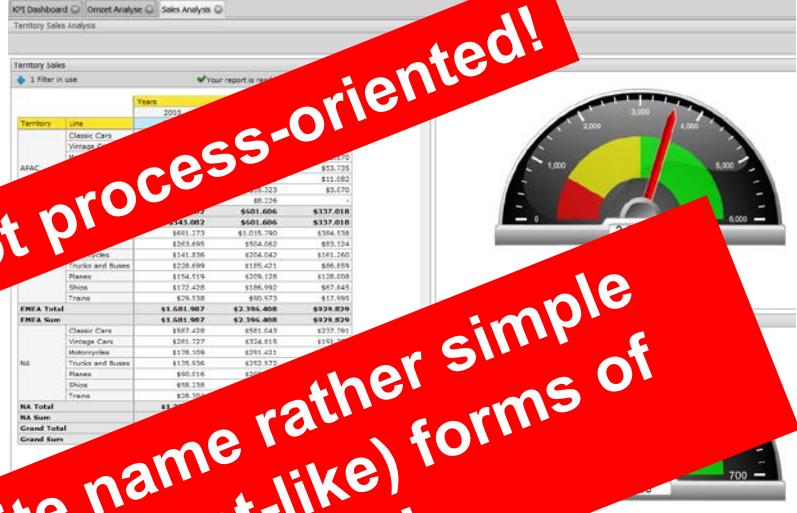


Business Intelligence (BI) tools

- Possible definition (one of many): "BI is a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision making." [Boris Evelson, Forrester Research 2010]
- Many products, e.g., **IBM Cognos Business Intelligence (IBM)**, **Oracle Business Intelligence (Oracle)**, **SAP BusinessObjects (SAP)**, **WebFOCUS (Information Builders)**, **MS SQL Server (Microsoft)**, **MicroStrategy (MicroStrategy)**, **NovaView (Panorama Software)**, **QlikView (QlikTech)**, **SAS Enterprise Business Intelligence (SAS)**, **TIBCO Spotfire Analytics (TIBCO)**, **Jaspersoft (Jaspersoft)**, and **Pentaho BI Suite (Pentaho)**.

Typical functionality

- ETL (Extract, Transform, and Load).
- Ad-hoc querying.
- Reporting.
- Interactive dashboards.
- Alert generation.



Despite name rather simple (spreadsheet-like) forms of analysis!

A photograph of a climber rappelling down a dark, textured rock face. The climber is silhouetted against a bright sky. In the background, a large, rounded mountain silhouette is visible against a sunset or sunrise sky. The water of a bay is in the foreground.

Tools are available, but process mining is still a relatively young discipline.

New tools will appear in coming years and process mining functionality will be embedded in more BI/BPM/DM suites.

Avoid the trap:

- process mining \subseteq data mining
 - process mining \subseteq BI
 - process mining \subseteq ...
- Use checklists!**

Part I: Preliminaries

Chapter 1

Introduction

Chapter 2

Process Modeling and Analysis

Chapter 3

Data Mining

Part II: From Event Logs to Process Models

Chapter 4

Getting the Data

Chapter 5

Process Discovery: An Introduction

Chapter 6

Advanced Process Discovery Techniques

Part III: Beyond Process Discovery

Chapter 7

Conformance Checking

Chapter 8

Mining Additional Perspectives

Chapter 9

Operational Support

Part IV: Putting Process Mining to Work

Chapter 10

Tool Support

Chapter 11

Analyzing “Lasagna Processes”

Chapter 12

Analyzing “Spaghetti Processes”

Part V: Reflection

Chapter 13

Cartography and Navigation

Chapter 14

Epilogue

Wil M. P. van der Aalst
Process Mining
Discovery, Conformance and Enhancement of Business Processes

