UNIVERSITY OF TARTU
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
Institute of Computer Science
Masters of Software Engineering

Musie Kebede

Comparative Evaluation of Process Mining Tools

Master's Thesis (30 ECTS)

Supervisor: Marlon Dumas

Tartu 2015

Comparative Evaluation of Process Mining Tools

Abstract:
Process mining is relatively young research area that meets the gap between businesses processes and various IT systems. Event logs are the primary sources for a process mining project and they are captured by different data sources including databases, ERP systems, CRM systems, audit trails, hospital information systems, bank transaction logs, etc. The extracted knowledge from this log enable us to discover the actual process and existing process model for further analysis, evaluation and continuous improvement in their quality. This way, various process mining tools have been developed in the market. Nevertheless, there is a lack of sufficent and comprehensive evaluation frameworks that assist users in selecting the right tool.

This thesis proposes a framework that enables the comparison of process mining tools in terms of their functional features. The proposed operations are linked to typical problems reported in existing process mining use cases. Using this framework, the thesis compares three process mining tools, namely ProM, Disco and Celonis The comparison shows that while these tools provide comparable functionality they differ in terms of the way the functionality is provided.

Keywords:
Process mining, Evaluation framework, process mining tools

Analüüsiva tarkvara protsessi võrdlev hinnang

Resümee:

Analüüsiv protsess on üsna uus uurimisvaldkond, mis avaldub äriprotsesside ja erinevate IT-süsteemide vahepealses lüngas. Logid on esmased allikad analüüsiva tarkvara projekti jaoks ja need on salvestatud erinevatesse andmekogumisallikatesse, seal hulgas andmebaasid, EPR süsteemid, CRM süsteemid, jälitusketid, haigla infosüsteemid, pangaülekannete logid ja nii edasi. Laiem teadmine sellest logist võimaldab meil avastada tõelist protsessi ja eksisteerivaid töötlusmudeleid edasiseks analüüsiks, hindamiseks ja nende kvaliteedi jätkuvaks arendamiseks. Sel viisil on erinevad protsessi analüüsivad tarkvarad arenenumad turul. Sellest hoolimata on puudus piisavast ja kõikehõlmavast hinnangu raamistikust, mis aitaks kasutajatel valida õiget tarkvara.

See teema esitab raamistiku, mis võimaldab protsessi analüüsivate tarkvarade võrdlust silmas pidades nende funktsionaalseid omadusi. Esitatud operatsioonid on ühendatud tüüpiliste probleemidega, millest on raporteeritud olemasolevate protsessi analüüsivate tarkvarade kasutamisel. Kasutades seda raamistikku, see tees võrdleb kolme protsessi analüüsivat tarkvara, nimeliselt ProM, Disco ja Celonis. Võrldus näitab, et need vahendid tagavad võrldevust funktsionaalselt, ent erinevad sel moel, mil nende funktsionaalsus on tagatud.

Märksõnad:

Protsessi analüüs, hinnangu raamistik, protsessi tarkvara

ACKNOWLEDGEMENTS

I want to show my gratitude to my supervisor Marlon Dumas for his constant follow up, help and valuable advices that helped bringing this Master's thesis this level. It would be unimagined to reach at this point without his support.

Besides my advisor, I would like to thank the Estonian Education Ministry for supporting my studies in giving a need-based scholarships for the last one year.

Last but not the least, I would like to thank my father who has helped me to persist and reach to this point in life.

Table of Contents

1. Introduction

Various companies want to understand how their processes operate in the real world (Tiwari et al., 2008). Using process mining, data stored in different process management systems can be utilized for a better overview of the actual process. That means, deviations can be analyzed and continuous improvements can be made (Van der Aalst, 2011). For a long time business intelligence tools have been already implemented and supported the industry to analyze large amounts of data in terms of OLAP analysis, adhoc reporting, “what if“ analysis and with other additional reporting features. Nevertheless, the need to discover insights into the process itself rather than assuming the underlying process as if it works fine, paved the way to the development of various process mining tools. As a result of this, various tools that support a wide range of mining algorithm techniques [1,2] and extensions has been developed and are under development both in the open source and commercial market.

Although every vendor has its own standards and conventions, the functionality of process mining will not change but rather grows. Thus, the more these tools are available for the public, the more the industry that uses them searches for evaluation framework that plays a main role in decision making while selecting those tools.

1.1. Goal

The intent of this thesis is to bridge the gap between the tool vendors and the potential user by providing an evaluation framework that is derived on a list of key business process problems mapped to the respective process mining operations.

The final aim of this thesis is to compare and analyze three selected tools including ProM, Disco and Celonis using the framework as a foundation. Moreover, the same framework can be used for other tools available in the market.

1.2. Research questions

The main research question addresses an approach for selecting the right process mining tool, among several emerging, that can fit in to the underlying business use case.

In this work, we identify evaluation criteria that would answer the research question

1. What are the main problems a given process mining tool user can confront?
2. What are the respective process mining operations which can be performed to solve the problem?
3. Which tool can support the operation and additional functionalities that are important with regard to process mining?

1.3. Approach

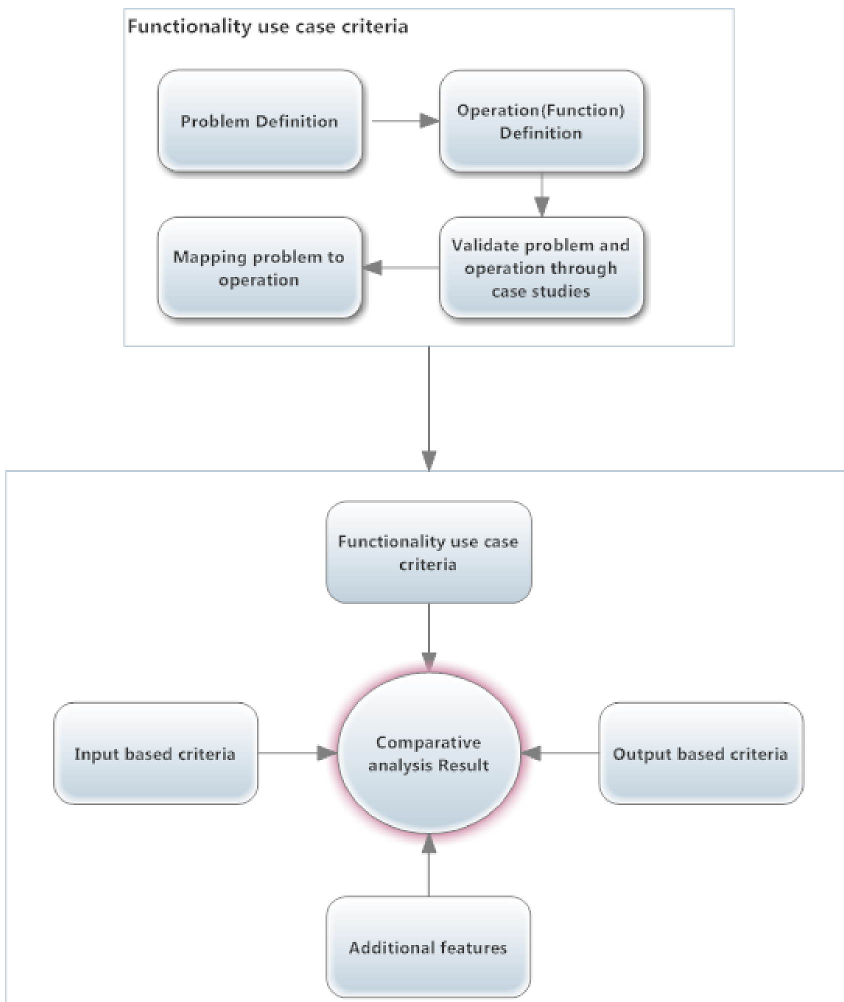


Figure 1.1: Approach to a comparative analysis of process mining tools

Our approach follows two interconnected steps which is shown in Figure 1.1. Since there is no standard reference for process mining functionalities, we followed analytical approach to illustrate prominent problems, extract the possible operations primarily and then acquiring the possible list of functional use cases which are capable to compare the selected tools.

On the second phase, process mining framework is applied to guide the overall comparison analysis [3] that assures the representation of the subject itself. There can be many questions from the user perspective while using this framework as a reference for making decisions. Thus, by this approach we will make sure that the evaluation framework can comply with the existing process mining framework.

1.4. Outline

In the following chapters, we will discuss the concepts and guidelines that is used to approach the underlying research question. Chapter 2 gives an overview of process mining tools including the criteria for the selection of the tools for comparison. In Chapter 3, we propose an evaluation framework that is based on a problem-operation approach. The first section of chapter 3 presents a list of problem types, classification framework and classification category as input, output and function mapping with the indicated problems in the beginning of the chapter. In Chapter 4, after a short introduction of the evaluated systems, we evaluate each system based on the outlined framework in the former chapter and draw a summary of the main findings. Finally, in chapter 5 a conclusion is provided to summarize the overall work.

2. Related Work

Since process mining is a recent research subject, the focus tends in developing new techniques and algorithms rather than developing a standard use case and attribute that allows to compare various tools.

According to Turner, Chris J., et al. [4] comparison of process mining tools was made in the UK market, which elaborates the practice of business process mining along with the method of analysis. The paper mentions types of business process mining, process models and process mining algorithms as a ground for comparing 7 process mining tools: Futura Reflect, Fluxicon, Comprehend, ARIS, BPM One, Iontas Focus Suite and Fujitsu Automated Process discovery Service.

On the other hand, Irina Ailenei, Anne Rozinat, Albert Eckert and Wil M. P. Van der Aalst [5] mention: Unlike, assessing the quality of the discovered process model [6-7], a focus on the different functionalities related to the process mining use case context (role in this case) has been illustrated in the research work.

A recent master's dissertation work provides an exploratory research on the use of process mining tools by answering the questions [8]

- Why users choose a certain process mining tools
- What are the criteria to choose those process mining tools

The questions accordingly were addressed using in depth interviews and online surveys. In reference to the final findings: usability, visualization, integration and functionality criteria's including import functionality, export functionality and the presence of advanced filter mechanisms by the tools was considered as an answer to research question.

Nevertheless, unlike other related works, we adopt analytical methodology for capturing the possible problems while mining a business processes in comparison to the operations (techniques) used to answer those problems. Furthermore, using process mining framework as a means to set the overall evaluation framework assures the dependability of the solution on process mining techniques. The thesis in general addresses issues related to process mining tool selection that in turn enables to create transparent and analysis based decision for the consumer.

3. Evaluation Framework

The purpose of process mining is to collect information from event logs from various perspective such as control flow, data flow, organizational perspective and performance perspective [9, 10]. Based on the information, one can construct a process model presented as a Petri net notation in order to visualize how the real world process works. While working with different level of event log data that requires in-depth analysis and optimization it is required to consider a framework. This framework later defines the evaluation criteria for the tools.

3.1. Process mining problem types

A. Inspecting and cleaning event data

Data quality is an important factor in the result of a process mining analysis project. Therefore, some of the preliminary steps to avoid unwanted data is checking the current log and identify whether it can represent the underlying process structure. An event log can consist of an outlier and it is important to define and detect them before cleaning the logs. Logs can also contain different level of granularity both timestamp wise and logs extracted from information systems of the same nature.

Based on the main finding from the process mining manifesto, treating event data as a first class citizen is crucial for inspection [11]. This means it should be worth to assume that the given data has actually driven from an event that already occurred and the attributes of the event are also correct. For instance, we cannot calculate the duration of cases if the timestamp attribute misses out of the event log.

B.Understanding the business process model

The lack of understanding the processes as they are: such as where it starts and ends, what is included and excluded, what the actual process is working, is another problem that can be addressed by using process mining. Once we have a clear picture of the workflow of the process, it is possible to solve the problem related to efficiency of underlying models. According to a study made based on an online survey explores the ability of model checker to find how correctly the user can draw the right conclusion from a set of process model. Given two groups of analysis, the first with an abstract activity and second with illustration one can draw the conclusion that understandability resides in personal knowledge, structure of the model and the content [12]. Empirical study provides the fact that larger and real world process models have more deadlocks than a simpler model [13, 14].

C. Performance issues

The presence of timestamps in an event log enables the discovery of bottlenecks. Bottleneck exposes performance issues and process mining tools detect where and why it occurs. Performance is related to the degree to which an organization achieves its objective in respect to various tasks. The indicators for a progress in an organization are called key process indicators (KPI) which acquire the aspect of time, cost and quality related measures. The following rule applies before finding bottlenecks:

- Discover process models
- Align log and model
- Replay the alignment

D. Understanding deviance and variance

Processes might not be handled properly all the way through due to deviation of cases. The more cases deviate from the appropriate line, the harder likely to inspect them case by case in a later stage. Rare deviations happen usually but should have to be evaluated for a change. Deviations can be expected or unexpected leading to either to high process performance or low performance respectively. Furthermore, unlike deviance, process variance relates to cases that are detached with respect to others which introduces a misalignment in the flow of activity within the process.

E. Measuring business alignment

Comparing the real behavior of an information system or its users with the intended or expected behavior is crucial for maintaining the fit between business processes and the respective information system. Business alignment is a way in which events relate to model elements and vice versa. In various organizations new processes are developing and existing processes are changing through time. For example, ERP systems record every transaction performed every day and store the event logs. A case study in a Dutch institution [15] which collects fines that must be paid: Each fine was considered, independent case as soon as the fine is paid and the process ends. Among many cases performed using the system, selected 130,136 cases, showed that there is divergence between the predefined process and the actual process. From business alignment point of view, this scenario might indicate a misalignment between information system (predefined process) and actual process.

Case_id	Date_def	Order	Act_code	Planned_date	Exec_date	Trigger_date
1877402	13-10-2001	1	2	30-09-1992	07-10-1992	16-12-1992
1877402	13-10-2001	2	6	16-12-1992	23-12-1992	06-02-1993
2688373	27-10-2001	1	2	20-01-1993	27-01-1993	06-04-1993
1877402	13-10-2001	3	33	06-02-1993	06-02-1993	11-02-1993
1877402	13-10-2001	4	52	11-02-1993	15-02-1993	24-02-1993
1877402	13-10-2001	5	53	24-02-1993	22-02-1993	19-05-1993
2688373	27-10-2001	2	6	06-04-1993	14-04-1993	29-05-1993
1877402	13-10-2001	6	73	19-05-1993	19-05-1993	27-05-1993
1877402	13-10-2001	7	71	27-05-1993	27-05-1993	11-06-1993
2688373	27-10-2001	3	7	29-05-1993	05-06-1993	20-07-1993
1877402	13-10-2001	8	12	11-06-1993	12-06-1993	13-07-1993
1877402	13-10-2001	9	71	13-07-1993	13-07-1993	13-07-1993
1877402	13-10-2001	10	12	13-07-1993	13-07-1993	03-08-1993
2688373	27-10-2001	4	12	20-07-1993	21-07-1993	26-07-1993
2688373	27-10-2001	5	32	26-07-1993	26-07-1993	27-07-1993
2688373	27-10-2001	6	33	27-07-1993	27-07-1993	28-07-1993
2688373	27-10-2001	7	52	28-07-1993	03-08-1993	18-08-1993
1877402	13-10-2001	11	99	03-08-1993	03-08-1993	28-12-1993
2688373	27-10-2001	8	53	18-08-1993	10-08-1993	14-07-1995
4645624	20-10-2001	1	2	08-11-1993	15-11-1993	26-01-1994
4838702	20-10-2001	1	2	01-12-1993	08-12-1993	16-02-1994
1877402	13-10-2001	12	16	28-12-1993	28-12-1993	04-01-1994
1877402	13-10-2001	13	2	04-01-1994	12-01-1994	23-03-1994
4645624	20-10-2001	2	6	26-01-1994	02-02-1994	19-03-1994
5368603	20-10-2001	1	2	15-02-1994	22-02-1994	05-05-1994
[illegible]	[illegible]	2	6	[illegible]	[illegible]	[illegible]

Figure 3.1. A fragment of the log [15]

3.2. **Process mining operations**

3.2.1. Classification framework

Every process mining tool meant to support process mining techniques. Nevertheless, based on the vendor the tool can either support partial or extended features. Therefore, finding a good evaluation framework among those tools can totally depend on the definition of process mining itself. Once having a process mining framework, it allows to identify and select specific use case.

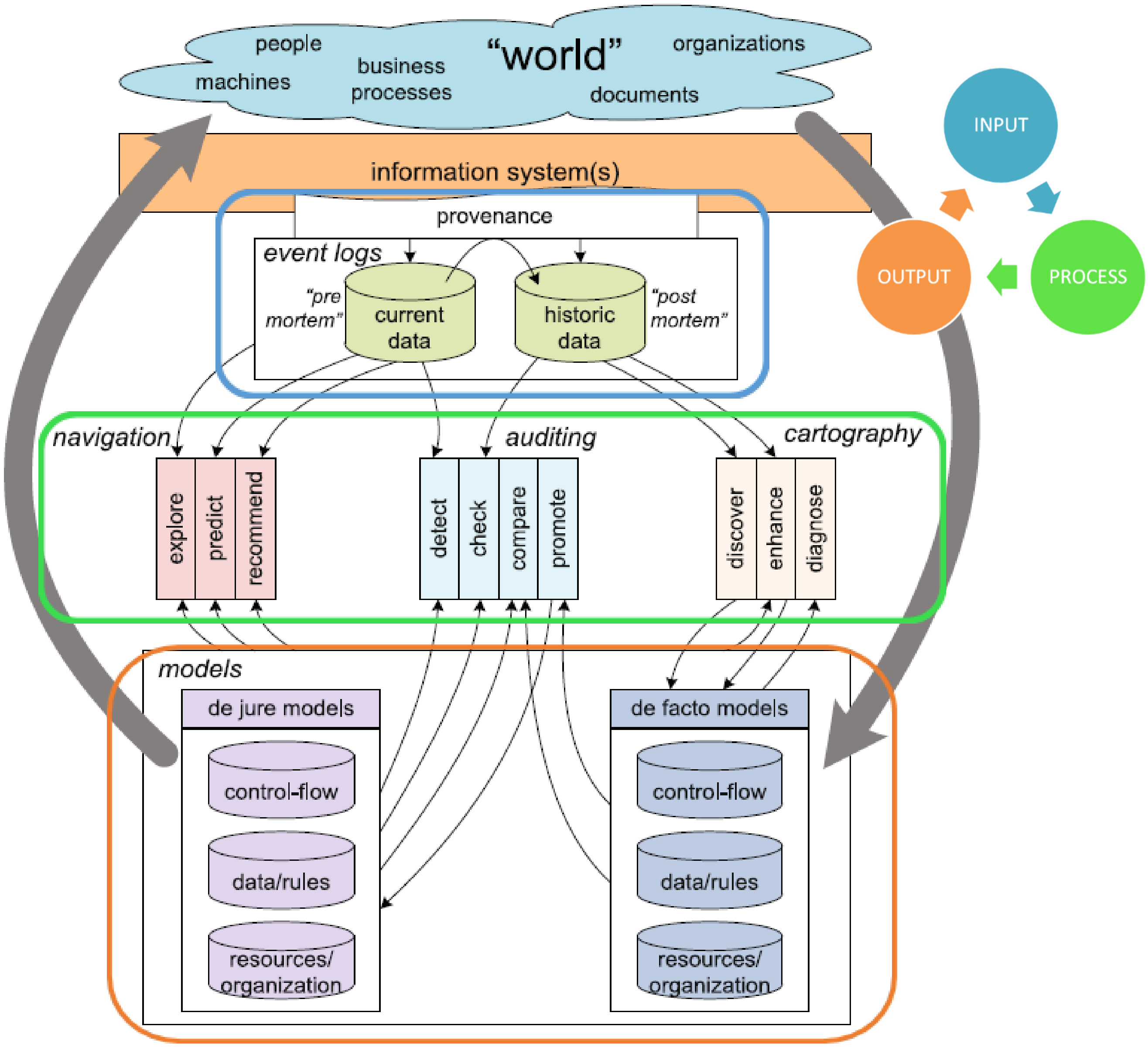


Figure. 3.2. Process mining framework from (van der Aalst et, 2011 p.262)

3.2.2. Classification of inputs and outputs

A. Input based criteria

Input data for process mining is available as an event log. Nevertheless it requires preprocessing before feeding to process mining systems so that it can achieve the targeted analysis smoothly. Among several challenges many people face include locating the relevant data, identifying major process instances, flattening the event data and scoping the problem at hand.

Event logs categorized into pre mortem and post mortem to indicate whether the cases of the event data are not completed or completed respectively [x]. In analyzing mining tools since a real time uncompleted data is not used, we focus on historic event data.

Type: Defines the format type(s) used by the specific process mining tool.
Size: How many event cases the tool can process

B. Output based criteria

Model Notation Support: Based on the end user requirement a variety of model notations are available including BPMN, WF, Petri nets, EPCs, transition systems, heuristics and other formats.

C. Additional criteria

Supported Platform: which infrastructure every tool supports

3.2.3 Classification of functions

A. Filtering the data

Before extracted event log is used it requires filtering. One of the main reason includes Information systems are not error free, either data do not reflect real transaction data or the program can malfunction. Therefore, process mining tools provide different level of filtering.

B. Process Discovery

Process mining techniques supporting discovery do not assume an a-priori model, i.e., based on an event log, some model is constructed. From the model, a process mining tool can be evaluated based on the support for the following criteria's.

- Structure of the process: Analyze discovery capability for unknown or existing process models.

- Routing options: Get a detail view for one or more paths within the process flow.
- Frequent path: Identifying which path most cases follow.
- The distribution of cases over path: Getting a visualization in to and identify what occurs in the process flow when different cases are dispersed through a given path.
- Mining Loops: In a given process, it is possible to find same task executed multiple times that can occur in a number of instances introducing loops to the process model. For instance, in an event log $L1 = [(a, c)^2, (a, b, c)^3, (a, b, b, c)^2, (a, b, b, b, b, c)^1]$ a loop of length one occurs at task b. Process mining tools apply different algorithms to deal with loops.

C. Conformance Checking

Conformance checking signals the need for a better control of the process. Other discrepancies may reveal desirable deviations. For instance, employees may deviate to serve the customers better or to handle circumstances not foreseen by the process model. In fact, flexibility and non-conformance often correlate positively. When checking conformance, it is important to view deviations from two angles: (a) the model is “wrong” and does not reflect reality (“How to improve the model?”), and (b) cases deviate from the model and corrective actions are needed (“How to improve control to enforce a better conformance?”). Conformance checking techniques supports both viewpoints.

By comparing what really happened (as seen in the data) and what should have happened (as captured by organizational process models or business rules), process mining can detect (un-)desirable deviations in the actual execution of business.

The key factor to conformance checking include

- Exception from normal path: detect outliers from the deviating process path.
- The degree to which the rule holds: checks whether the referencing process model fits in discovered model.
- Compliance checks: comparison ability between the real and ideal process.

D. Trace Clustering techniques

Trace clustering operation enables to split unstructured process into homogeneous subsets and for each subset, a process model is created. Most real life processes rules are not strict rather flexible for example those in healthcare and customer service to mention. The analysis result for process model extraction are similarly unstructured One of the major factors to unstructured output is having a diversity of an event logs. Nevertheless, dividing in to structured process variant, it is possible to create homogenous subsets of cases.

E. Social Network mining

Given a log with resources information, process mining can discover relationships between resources (such as delegation of work and clusters of resource networks).

F. Decision rule mining

Machine learning algorithms are applied on top of business process models in order to give insight into a process model. Starting from a discovered process model (i.e., a model discovered by conventional process mining algorithms), it is possible to enhance models by including patterns, i.e., each choice is analyzed and linked to properties of individual cases and activities if it is applicable.

G. Performance reporting

Performance measures (KPI) is possible because of events in the log have been coupled to model elements through replay and include the following actions:

- Waiting and service times: Statistics such as the average waiting time for an activity can be projected onto the process model. Activities with a high variation in service time could be highlighted in the model.
- Cycle time analysis: i.e., the time spent in the process from start to finish by a case or group of cases. Primarily measured in time units. This figures are available as statistics in process mining tools
- Bottleneck detection and analysis: The multi-set of durations attached to each place can be used to discover and analyze bottlenecks. The places where most time is spent can be highlighted. Moreover, cases that spend a long time in a particular place can be further investigated.
- Flow time analysis: the overall flow time can be computed. One can also point to two arbitrary points in the process, say a and c, and compute how many times a case flows from a to b. The multi-set of durations to go from a to b can be used to

compute all kinds of statistics, e.g., the average flow time between a and b or the fraction of cases taking more than some preset form.

- Utilization of resources: While replaying the model, times and frequencies are collected that can show how a resource performs.

The resulting process model will consist of all the above information in order to answer key performance issues.

H. Process visualization

By replaying recorded data, process mining can animate past process executions in various forms with the goal of enabling effective extraction of information (i.e. visual analytics).

I. Discriminative Rule Mining

Based on the study indicates that process deviation can be analyzed through the following steps [16]:

1. inspecting the deviation using process mining tool
2. applying association rule mining
3. Discussion with the business domain experts

In our case, based on the tool we compare, it is possible to apply deviation rules and methods in advance. Moreover, In order to find the root cause it is important to ask the following questions: Where does the process deviate? Which cases? Which locations in the process? How does the process deviate? What is going “wrong”?

K. Delta Analysis

Delta analysis compares the derived process model with respect to a reference model which is either the workflow model or the model originally used to configure the system. This analysis is important to answer the problems related to business alignment as long as events are logged and a control over the process is maintained. For instance, to illustrate how delta analysis works, suppose Figure. 1.4. (a) Represents perceived process model by a management and Figure 1.4(b) process model discovered.

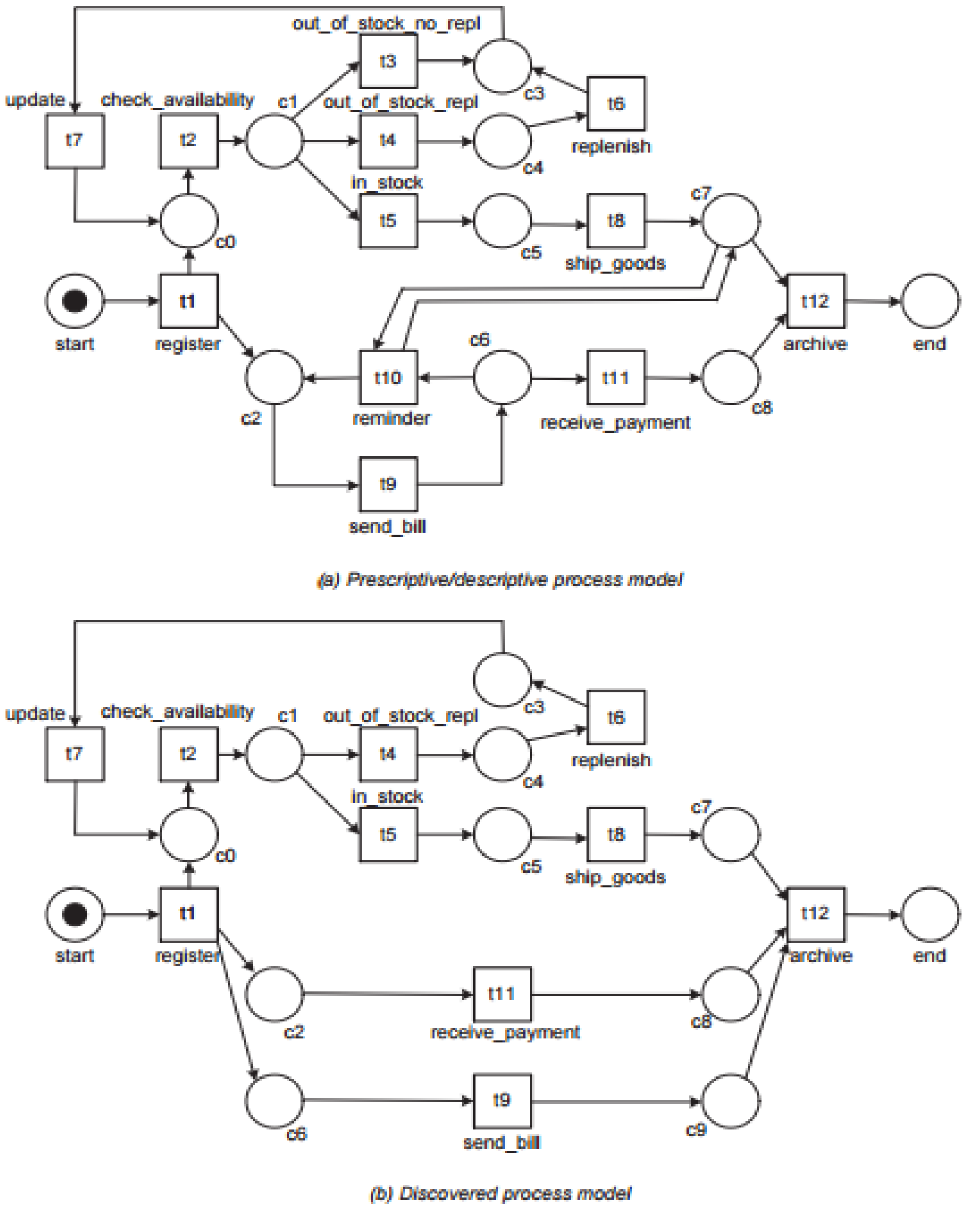


Figure. 2.3. Comparing discovered order processing model with predefined process model (adapted from [15])

While comparing the two models, it is obvious to discover four differences.

- In the discovered model out-of-stock always results in a replenishment order.
- In the discovered model receive_payment and send_bill occur parallel while sometimes customers can pay before acquiring the bill.

- In the discovered model it is impossible to send multiple bills as a reminder unlike the perceived model.
- In the discovered model it is possible to send the bill before shipping the good.

Thus, it was possible to identify the difference between the perceived process model and the discovered process model. Nevertheless, it requires decision to which side people must incline because both ways can be justified based on requirements.

3.3. Mapping problems to operations based on case studies

In a given process mining problem, there is a respective process mining operation(s) that can be applied to answer a process related question. Neverthless, mapping them ease the method of implementing the operation to a certain process mining related task. Therefore, in the following section , a list of process mining case studies are specified inorder to classify and link related problems to a list of operations.

Case studies:

A. **Description**: application of process mining in the industry at one of the offices of the Dutch National Public Works Department, in charge of the construction and repair of road and water infrastructure [17].
History: Process mining implemented in one of the 12 offices, 1000 civil servants which subcontracts work to various parties such as road construction companies, cleaning companies, and environmental bureaus.
Aim: Analyze processing of invoices sent by the various subcontractors and suppliers
Tool: ProM
Log: based on a log of the process of handling invoices
Problem: There are 12 offices, in the year 2001 these offices maintained 18 process versions to handle invoices. At the end the performance of all these different process versions failed. In response to these issues, the national RWS management decided to unify invoice processing in search of efficiency gains.

Payment duration (days)	Norm (%)	Actual (%)|
0–31	90	70
32–62	5	22
63	5	8

In 2002, the WfMS was implemented at the RWS office involved in the case study. The RWS event log (or ‘‘RWS log’’ for short) contains 14,279 cases. The total number of logged events is 147,579 and 487 employees participated in the process execution. Nevertheless, there has been problem during implementation such as noise and incompleteness.

Operations:

1. Process model discovery which unveiled the fact that many cases will go through loops during their life-cycle
2. Frequency analysis : activities with a frequency below 1% of the total number of events
3. Adding resource to control-flow model : Mining the organizational perspective that shows the performers responsible for 070_PV are the project leaders, who normally operate outside the main RWS building where the invoice processing takes place
4. Case analysis: ‘‘Is there a relation between the time that an invoice is being paid and the amount of money being involved with the invoice?

B. **Description**: a case study based on a log of inventory processes. Inventory process were performed in a Polish company producing mattresses [18].

History: 500 employees with warehouse management system support

Aim: Analyze Product management process

Tools: ProM and Disco

Log: On the basis of captured 554,745 events associated with 87,660 instances captured from January 1st, 2013 to May 25th, 2013 that involved 55 people. Some of the activities in the process cycle include product, rest, shipment approved, on fork, Shipped, Deleted and each activity recorded in WMS.

Problem: Conformance to model, level of conformance to a FIFO policy for various families of products and was able to schedule action aiming to eliminate discrepancies

Operations:

1. Model to Log conformance checking test

C. **Description**: applicability of process mining using a real case of a gynecological oncology process in a Dutch hospital [19].

Aim: to provide new insights that facilitate the improvement of existing care flows

Tool: ProM

Log: taken from service data of 627 gynecological oncology patients treated between 2005 and 2006

Problem: scattered information systems with billing and treatment measuring systems were operating independently. Later, after aggregating the logs a spaghetti process model discovered.

Operations:

1. Preprocessing of logs : to detect a representative forth lower level activities and aggregation
2. Process Discovery
3. Log clustering for spaghetti model (Self Organizing Map)
4. Discovery of social networks from process logs
5. Using dotted chart to show overall events and performance information of the log

D. Description: Copenhagen Airports Process Mining Case Story[1]

Aim: needs to meet several performance KPIs

Log: from bag tag log that consists the data about Check-in, destination and other parameters.

Problem: Check baggage process for bottlenecks

Operations:

1. Process discovery with a purpose of identifying bottlenecks

E. Description: Six months study in one of the Australian insurance firm called Suncorp on the application of process mining [20].

Aim: to provide new insights for applying process mining from the event logs recorded in claims management system and enhance time taking claims.

Tool: ProM, Disco

Log: claims log from sub systems

Problem: Defining performance distributions, frequent path taken, identify how the process is performed.

Operations:

1. Filtering of logs: to detect a representative for the lower level activities and aggregation
2. Representative Process Discovery

[1] source: http://fluxicon.com/blog/2014/12/process-mining-case-story-copenhagen-airports-as/

F. Description: mining clinical care pathways followed by its effect on patients [21]
Aim: to provide insights from the healthcare data
Tool: SaaS
Log: EMR data
Problem: how clinical activities impact patients in their care journeys
Operations:
The following figure consists of possible operations taken during the analysis.

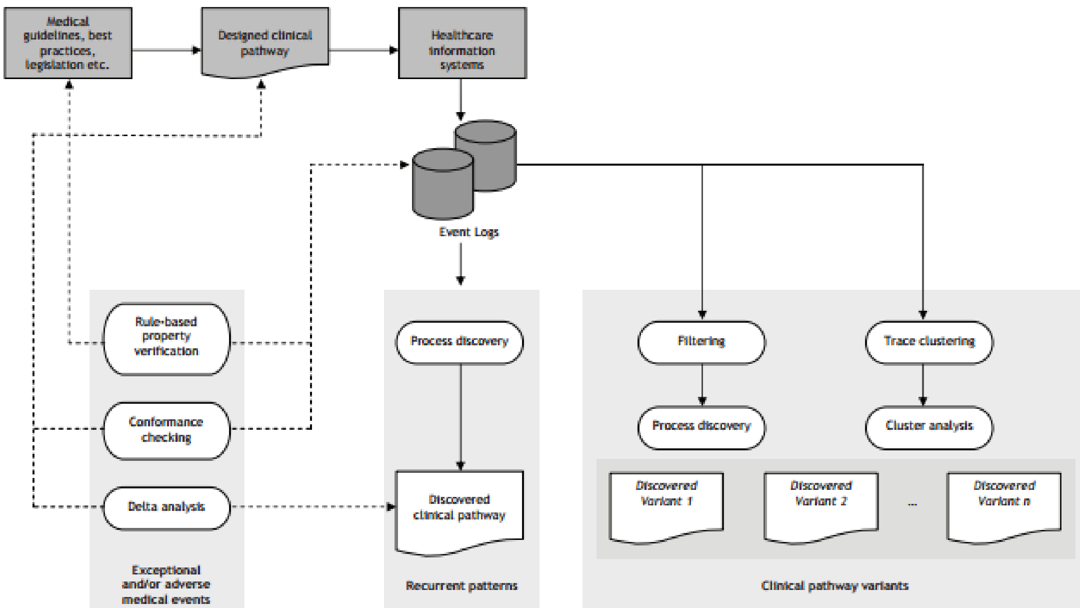


Figure. 1.5. Process mining architecture for clinical pathway analysis. [21]

G. Description: analyzing a refund service process for an electronic manufacturer[2]
Aim: to assist in optimizing service delivery
Tool: Disco
Log: Logs from IT system
Problem: a severe fallback on the flow of the process that led to customer complain and time consumption while inspecting the problem
Operations:

1. Detect bottlenecks
2. Detect deviation from the required process

[2] source: http://fluxicon.com/disco/files/Fluxicon-Disco-Case-Study-Refund-Process.pdf

H. Description: analyzing deviation of processes from the specified business rule[3]
Aim: partial automation of the analysis of deviations in a process by clustering deviating cases into corresponding business rules through use of process mining and association rule mining techniques
Tool: Disco
Log: Logs from IT system
Problem: complicated deviated cases
Operations:
1. Find the main deviations

I. Description: analyze real case of private dental procedures [22]
Aim: gaining in depth understanding of dental “single crown on implants" process
Tool: ProM
Log: Logs from IT system
Problem: complicated deviated cases
Operations:
1. Preprocessing
2. Control flow perspective for the paths patients follow (process discovery)
3. Resource perspective (Social network mining)
4. Discovery of performance related information

Overall, these case studies specify the usage of process mining techniques and tools for a variety of problems, as listed in the problem section of each case studies.

	Problems	Possible Operations	Case studies
1	Inspecting and cleaning event data	• Filtering the data	C,E,F,I
2	Understanding the business process	• Process discovery • Conformance checking • Social network mining • Decision rule mining • Process visualization	A, B,C,D, E,G, I

[3] source: http://essay.utwente.nl/62633/1/MSc_JJ_Stoop.pdf

3	Performance issues	• Performance reporting	G
4	Understanding deviance and variance	• Discriminative rule mining • Trace clustering	C,F, H
5	Measuring business alignment	• Delta analysis	

Table 1. Mapping business process mining with respective possible operations and case studies that discuss the operations

4. Evaluation

In the first section of this chapter, we discuss and analyze an event log taken from a company that repair telephones. [4] The main goal of providing this sample log is to illustrate whether or not the process mining tools in action can support the operation mentioned in the previous chapter. At the end of the chapter, we will summarize the result and reflect the outcome of this thesis work.

Overview: The Company repairs three different kinds of phones. The process begins with registering a device with defect as requested by the customer. After registration, problem detection office analyses the problem and the defect is labeled. There are a total of 10 categories of defects the company repairs. Once problem categorized, device is sent to repair department and a notification is dispatched to the customer. The repair department has two groups in which the simple defects and complex defects are repaired respectively. Nevertheless, some of the repairs can be done by both group. Once the device repair is finished, an employee from Quality assurance checks whether the defect is fixed or not. If the device is assured as fixed, the process would be archived and the telephone is sent to the customer. If not, Quality Assurance employee returns the device back again to the repair department. In order to avoid any throughput time, the company fixes the same device for a short number of times. When a defect is not fixed, it would be archived and new device is dispatched to customer.

4.1. Evaluated systems

Among several available tools three of them are selected for further analysis at this stage. This includes: ProM 6, Fluxicon (Disco) and Celonis. For the commercial tools, the selection was based on availability of the tool for academic evaluation purpose.

ProM (Process Mining) is toolset maintained by Eindhoven University of Technology, which collects prototypes developed in several research projects. The project is open source and aims largely the academic and research group. The plugins added on demand enables to solve complex process exploration. ProM imports event logs compliant with the MXML or XES formats and can load process model definitions in different standards. Some of the main features of ProM are: discovering the control-flow perspective of a process, social network analysis, analyzing the resource and performance perspective of

[4] source: http://www.processmining.org/_media/tutorial/repairexample.zip

a process, discovering events based on decision rules and conformance checking with a variety of algorithms. ProM provides several export formats such as CSV and PNG.

Disco is a commercial process mining tool developed by Fluxicon, which can run on top of Windows or Mac Desktops. It supports a wide range of event log import formats including CSV, MS Excel, MXML, XES, FXL Disco Logs and DSC Disco project files. Some of the features include automated process discovery, animation of process maps, event log filtering with various parameters, project management and detailed statistics.

Celonis is a commercial company that provides yet another software as a service process mining tool. Among the features for this tool include Automated Integration of source data, real-time surveillance of all business transactions, execution of process analyses, various filtering mechanisms and process reporting.

4.2. ProM

Filtering Data

Step 1: After importing the event log directly into the ProM

In ProM, from the *Log summary* we can find basic information about the log including starting and ending events, events per case, the number of total events including absolute and relative occurances. In addition, we can *inspect* every single process instances througly using the Inspector for ProM.

Inorder to filter data, we use the Action button and the action labeled with green (which can take log as input) and actions with yellow are actions that require more than a log. There are various options to filter a log in ProM including Filtering log by events, Filtering log by attribute, Filtering log by event attribute values, Filter log on trace attribute values , filter log using prefix closed language, Filter log on first six months per trace and Filter Log using Simple Herustics .

Applied on the Log:

Filter Log using Simple Herustics: Using this filter enables us to select events that we would like to keep or avoid. If a class has aborted cases, we can discard the instance. The next step in using this filter keeps those events that only start with the tasks selected in step 1.

Figure. 4.1. Log filtered using simple herustics

Filter Log by adding artifical events that automatically adds a starting and ending event for the process model.

Process discovery

The process model structure depiction depends on the audience.For example control flow, Petri net variants, fuzzy model , BPMN representation. Fuzzy models are not executable like Petri nets, but they allow for a simpler representation , addresing problems related to unstructured and large amount of activities. In a tool with more than 600 + plugins we could not discover entirely, neverthless for the expermential log , we used fuzzy miner, herustics miner and Multiphase miner to show the resulting process model. Input for a fuzzy miner is filtered log and its output is a fuzzy model.

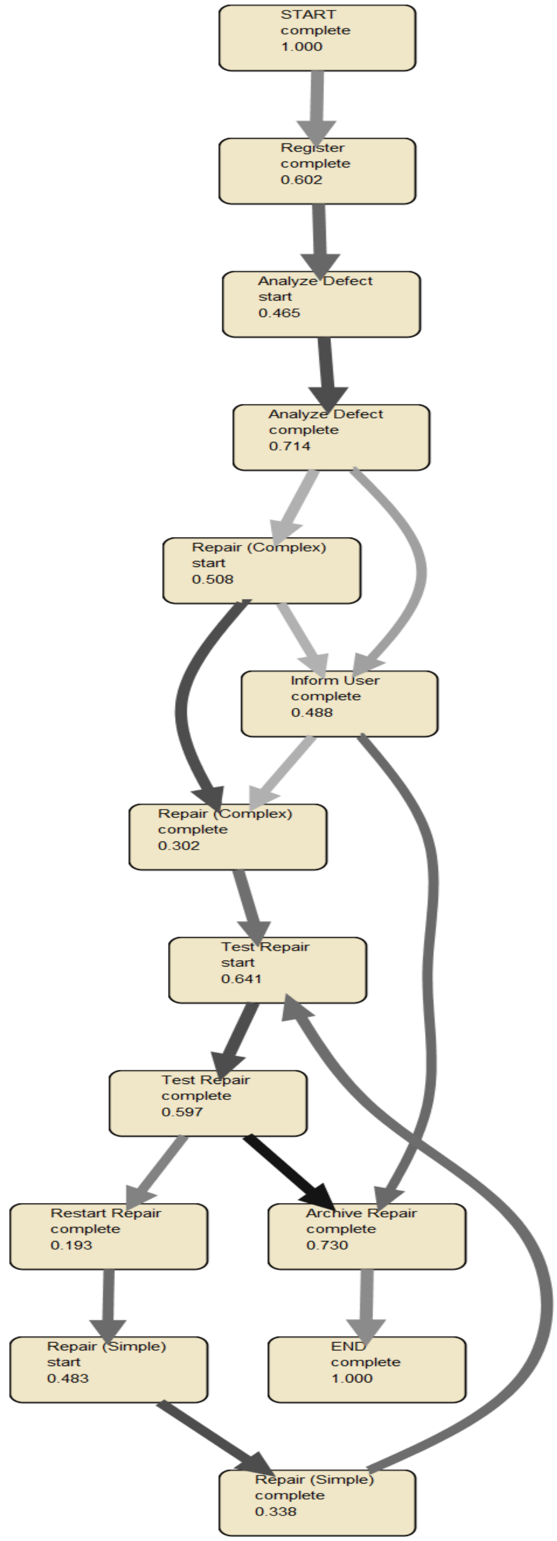

START
complete
1.000
Register
complete
0.602
Analyze Defect
start
0.465
Analyze Defect
complete
0.714
Repair (Complex)
start
0.508
Inform User
complete
0.488
Repair (Complex)
complete
0.302
Test Repair
start
0.641
Test Repair
complete
0.597
Restart Repair
complete
0.193
Archive Repair
complete
0.730
Repair (Simple)
start
0.483
END
complete
1.000
Repair (Simple)
complete
0.338

Figure. 4.2. Discovered process model for telephone repair Company using Fuzzy miner for ProM

Herustics miner was the second algorithm derived after the alpha algorithm which potentially addresses noise and concurrent activities. Herustics miner takes filtered log and outputs an Herustics net. This miner is important at times a Petri net is required for further analysis and when having less amount of events are represented in a real life log.

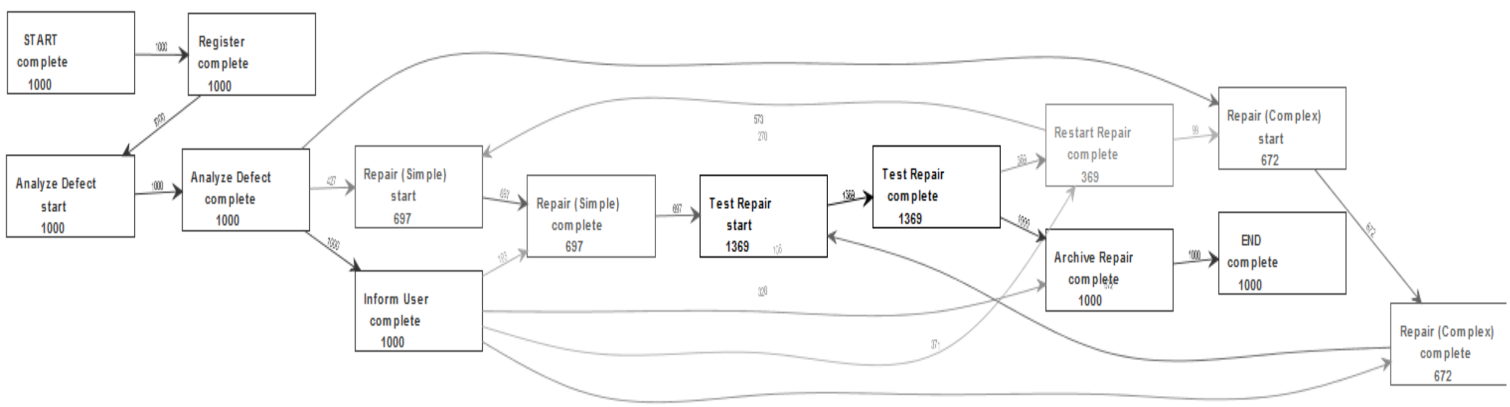


Figure. 4.3. Discovered process model for telephone repair Company using Heuristics miner for ProM

In ProM, questions such as which are the most frequent paths , loop pattern, routing options and distribution of paths can be found using pattern abstraction visualizer.

The pattern abstraction viewer discovers that Restart repair has a loop value of 2 one with repair (simple) and other repair (complex) as depicted in Figure 1.8.

Conformance checking

- The plug-in in ProM requires the process model in Petri net format(a .tpn or .pnml file) and it can be read with a format like in EPC or a YAWL
- Log should be preprocessed so that instances with similar property can be aggregated for the same of saving calculation time for conformance metrics.

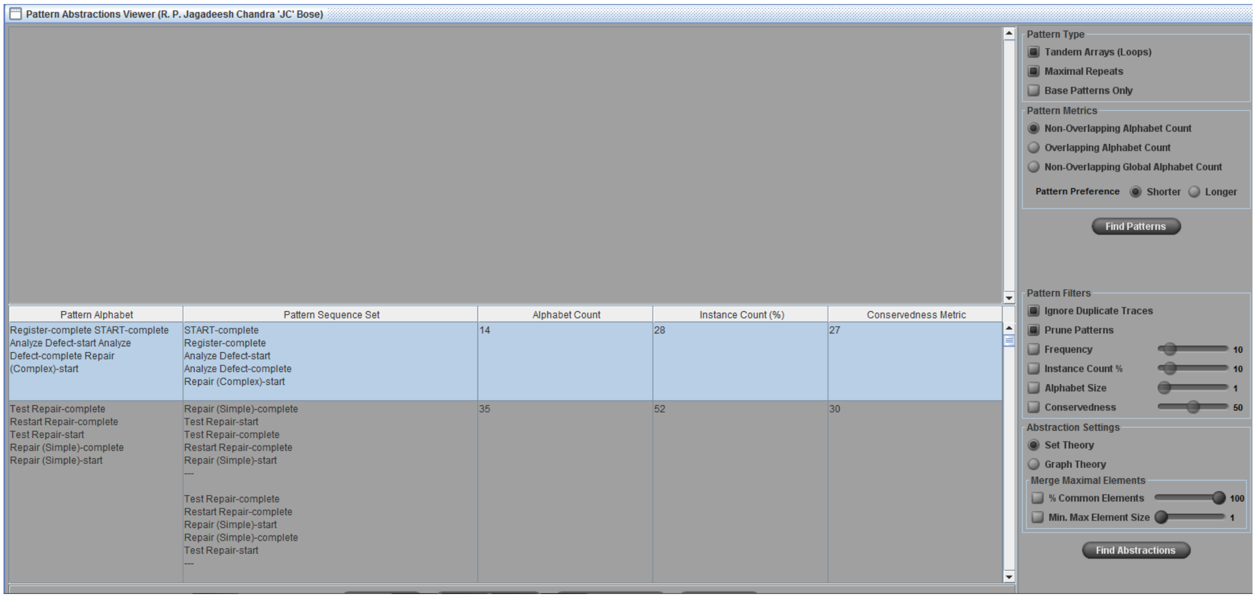


Figure. 4.4. Loop patterns using pattern visualizer for ProM

The conformance checker reveals the conformance of the log with mined process model has a fitness of 0.89. The green label indicates the frequency of synchronous move whereas the pink reflects the move in the model.

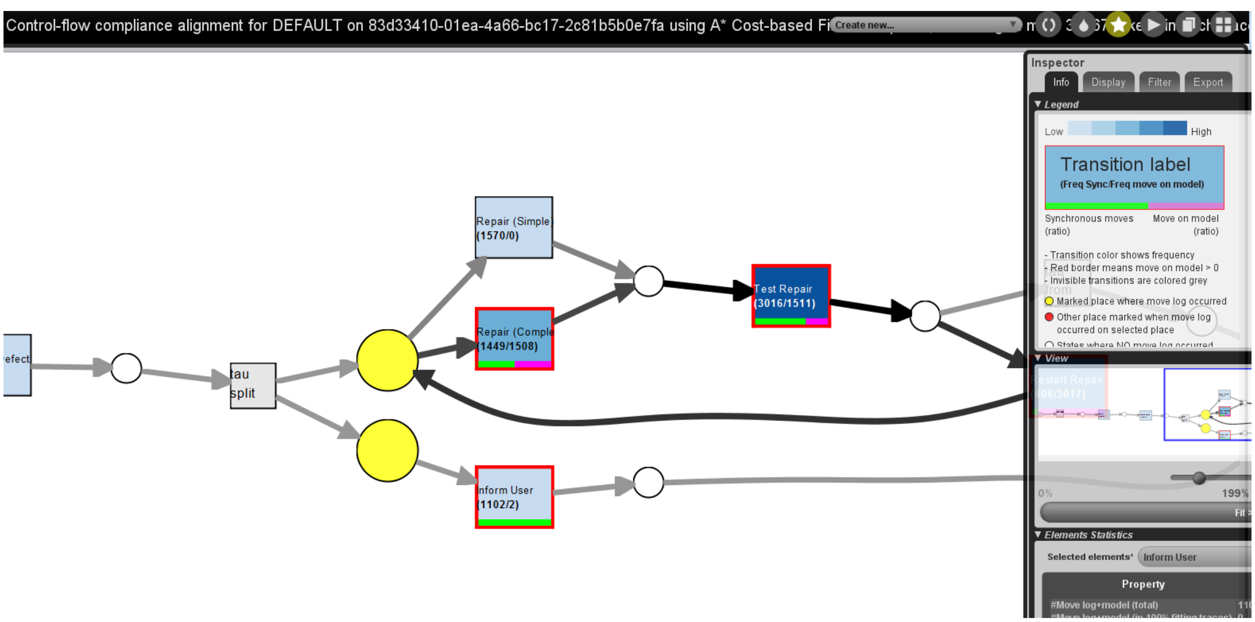


Figure. 4.5. Conformance checker in ProM (a)

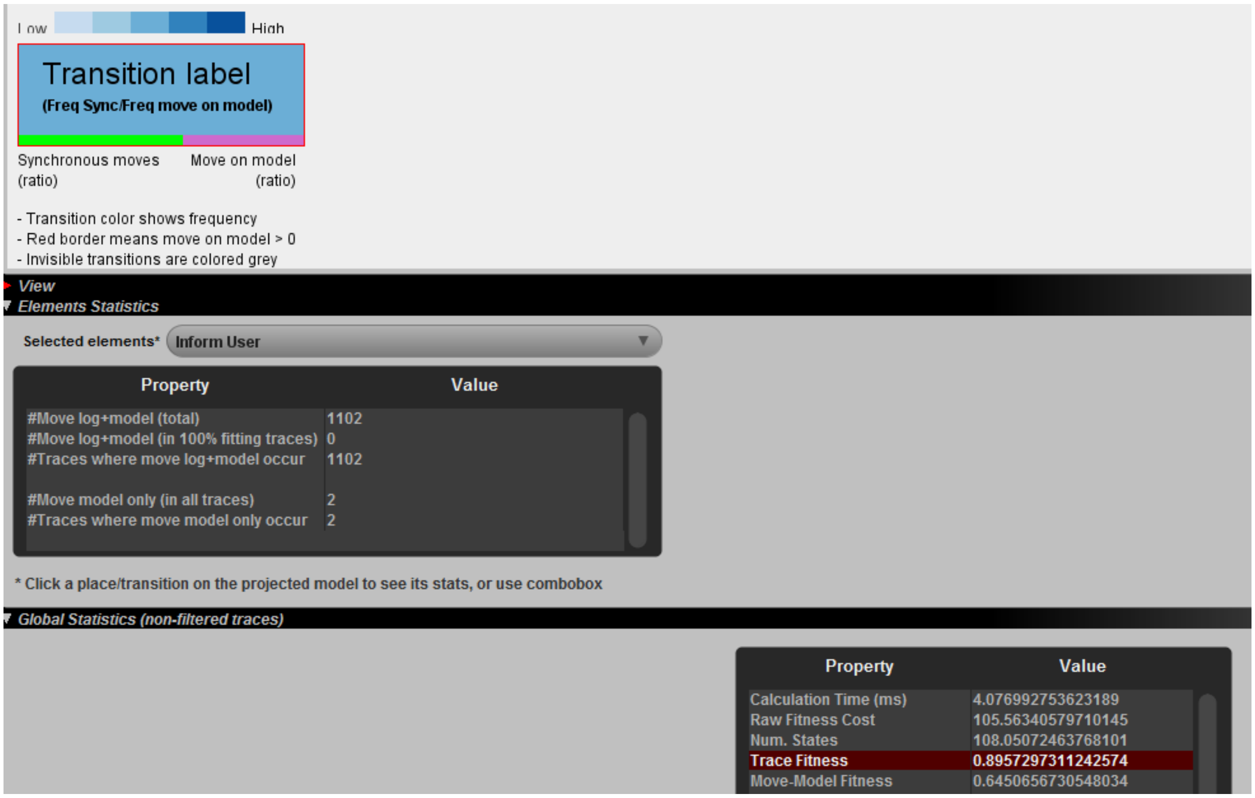


Figure. 4.6. Conformance checker in ProM (b, statistics view)

Social Network Mining

The social network miner provides an organizational perspective of processes with an option of mining for a handover-of-work social network, a reassignment social network, similar-task social network, subcontracting social network and working-together social network. For instance, from the diagram below tester2 handovers tasks to the system, Solver c1, solver c2, solver c3, solver s1, solver s2 and solver s2.

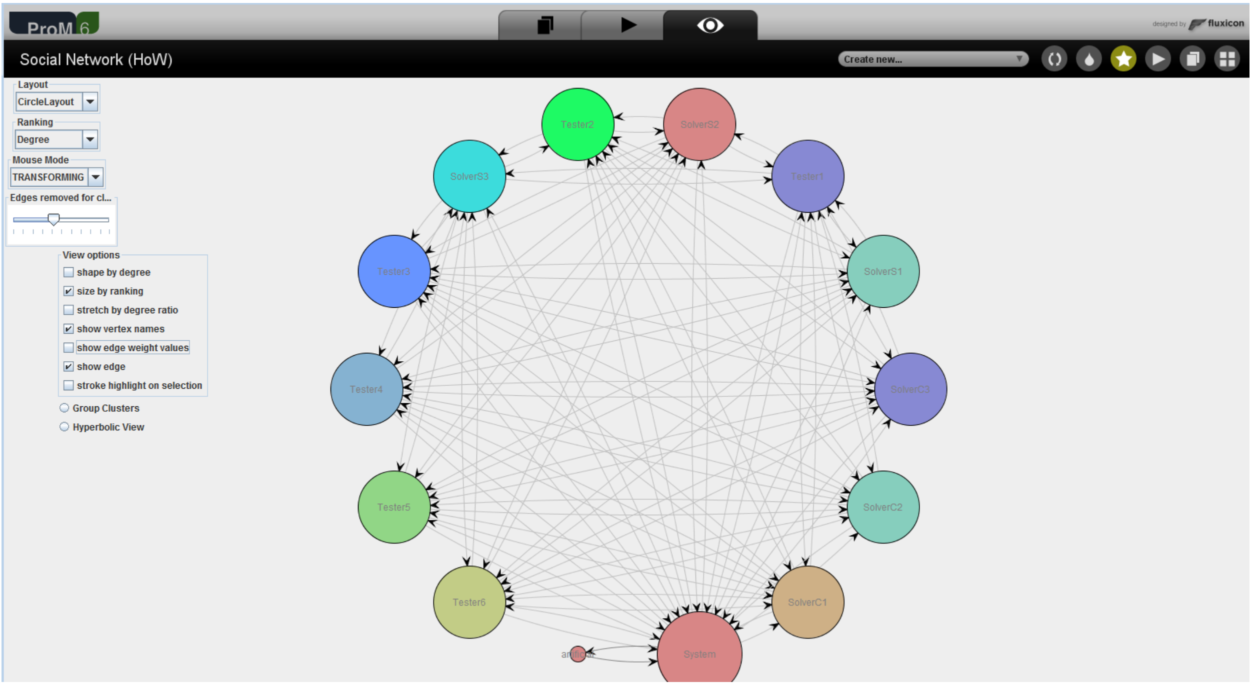
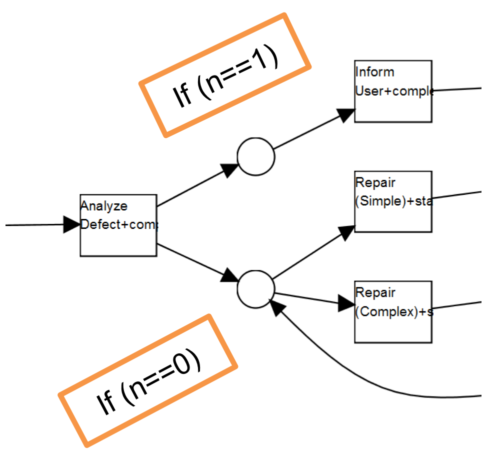


Figure. 4.7. Mining for a handover-of-work social network

Decision rule mining

The decision miner plugin in ProM determines the decision points in a Petri nets model and deals with hidden and duplicate activities.



N: repaired/not

Figure. 4.8. The model view provides a visualization of each decision point

Process visualization

The capabality of ProM to animate the log on top of a fuzzy mined model allows for searching bootlnecks and discover what hinders the actual process from running smoothly. The time spent in inspecting repaired phones and informing user takes much time.

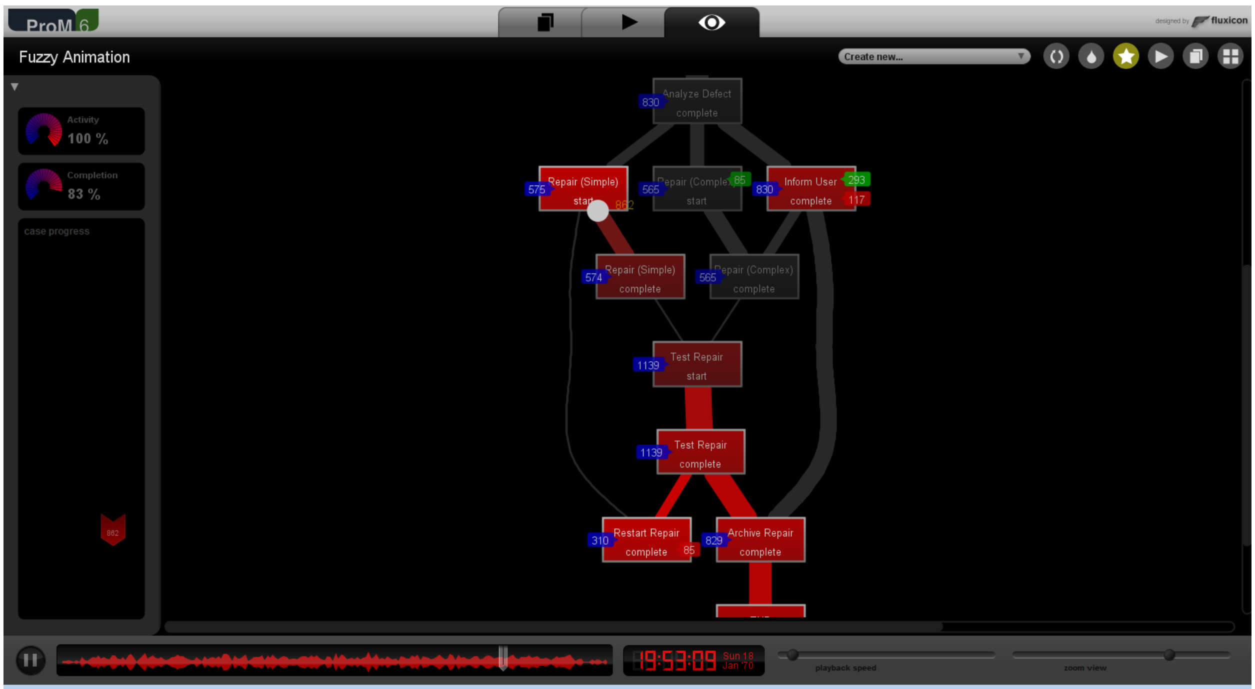


Figure. 4.9. Log animated on top of a fuzzy model in ProM

Performance reporting

In ProM, the plugin for analysis of performance from an event log is named as Replaying log on pertrinet for performance analysis which is depicted in Figure. 1.13.

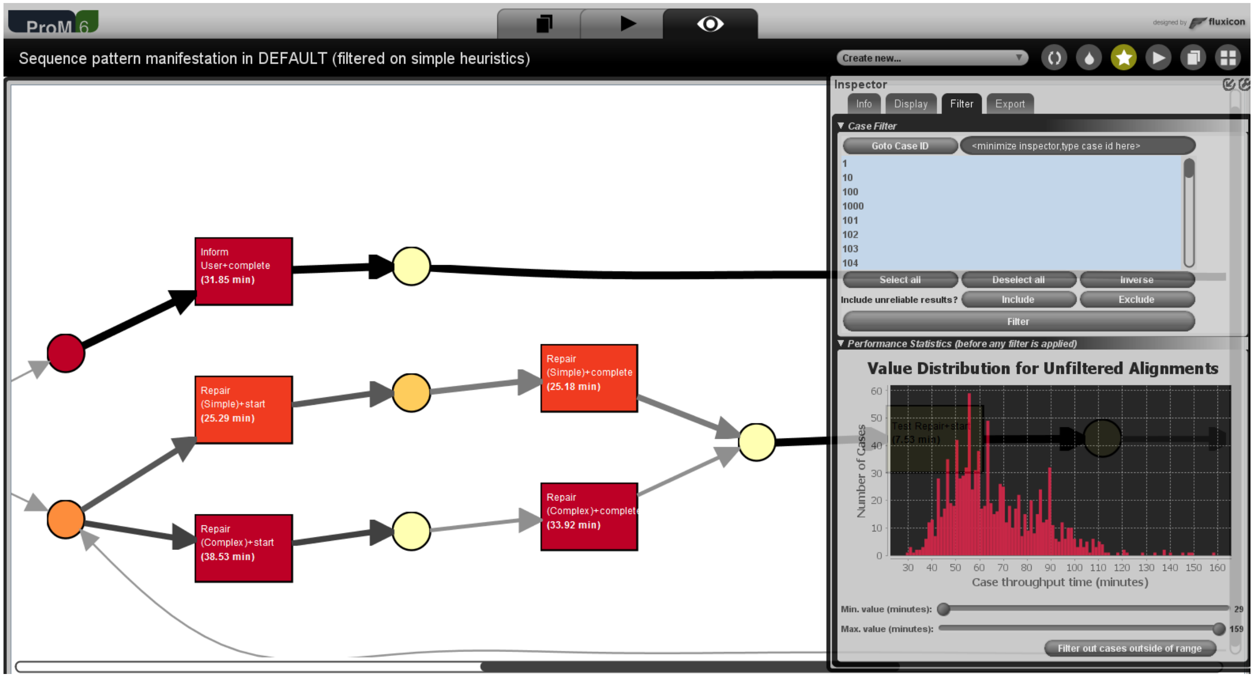


Figure. 4.10. Log performance analysis using ProM

4.3. Disco

Filtering Data

Before using Disco, the event log should fulfill the minimum requirement for analysis, including case ID, Activity and Timestamp. Furthermore, Disco reads pre-configured files in various standard formats:

- .mxml and .mxml.gz
- .xes and .xes.gz
- .fxl
- .dsc

Disco provides an automatic process discovery mapping once the event log is imported in to the system. This enables to easily filter and inspect attributes in parallel with the discovered fuzzy model as depicted in Figure 1.14 (a,b).

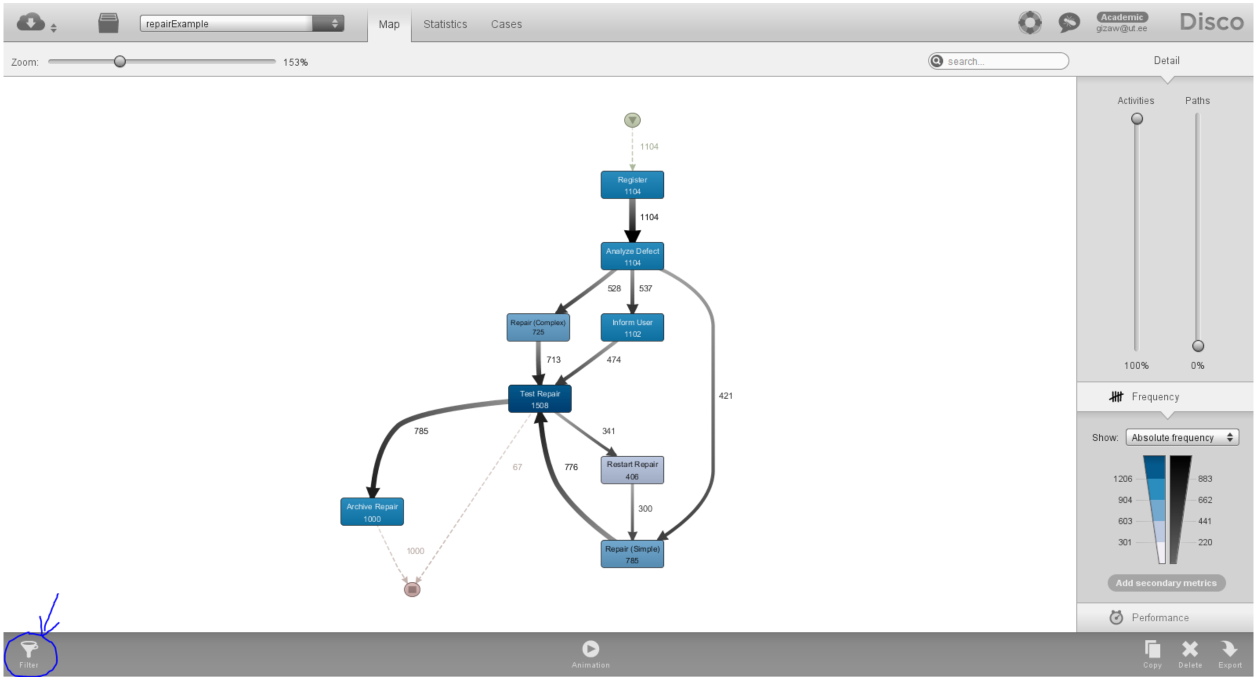


Figure. 4.11. Log Filter option with a process model in Disco (a)

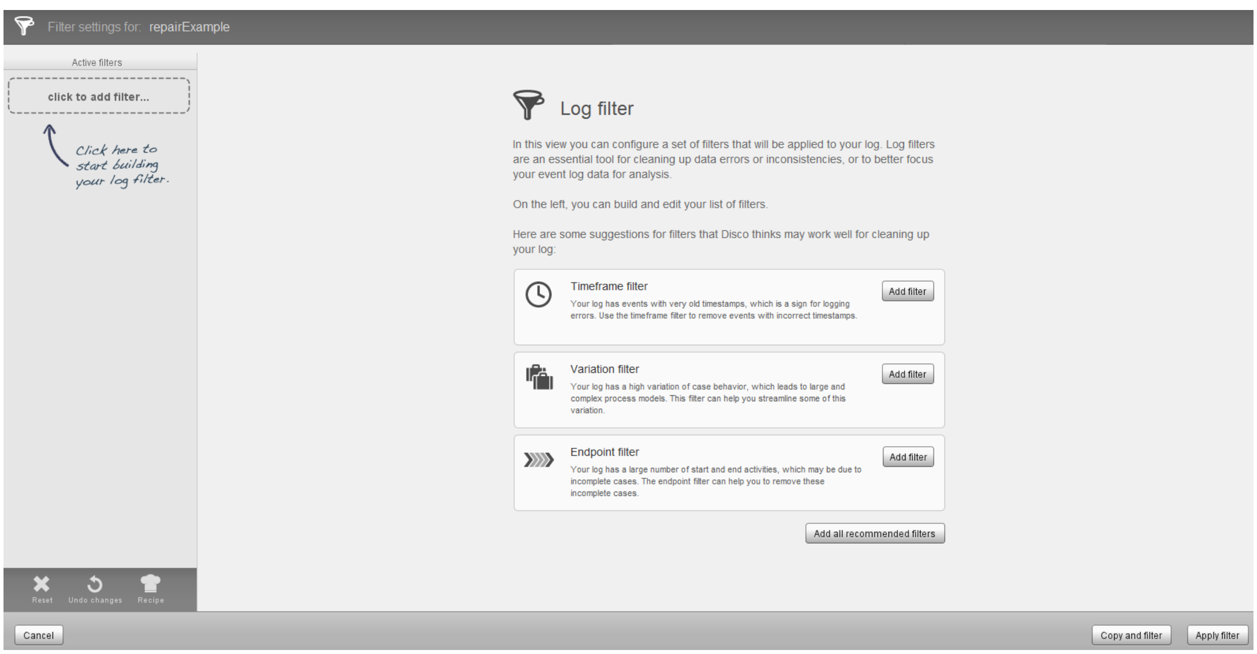


Figure. 4.11. Log Filter option with a process model in Disco (b)

There are six types of filters available in Disco:

1. *The time frame filter*: This filter enables to focus on part of a log executed in a certain period of time. For instance, in a given log taken over the years, if the user requires to analyses only the process for the latest six months the time frame filter provides flexible selection of the time frame.

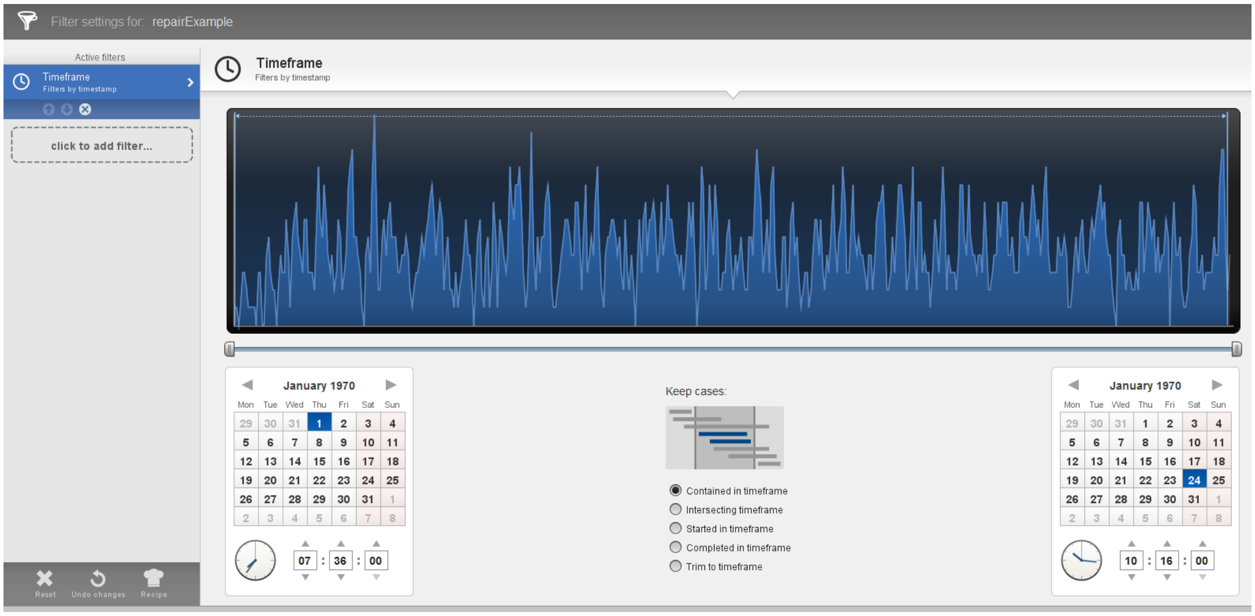


Figure. 4.12. Log Timeframe filter in Disco

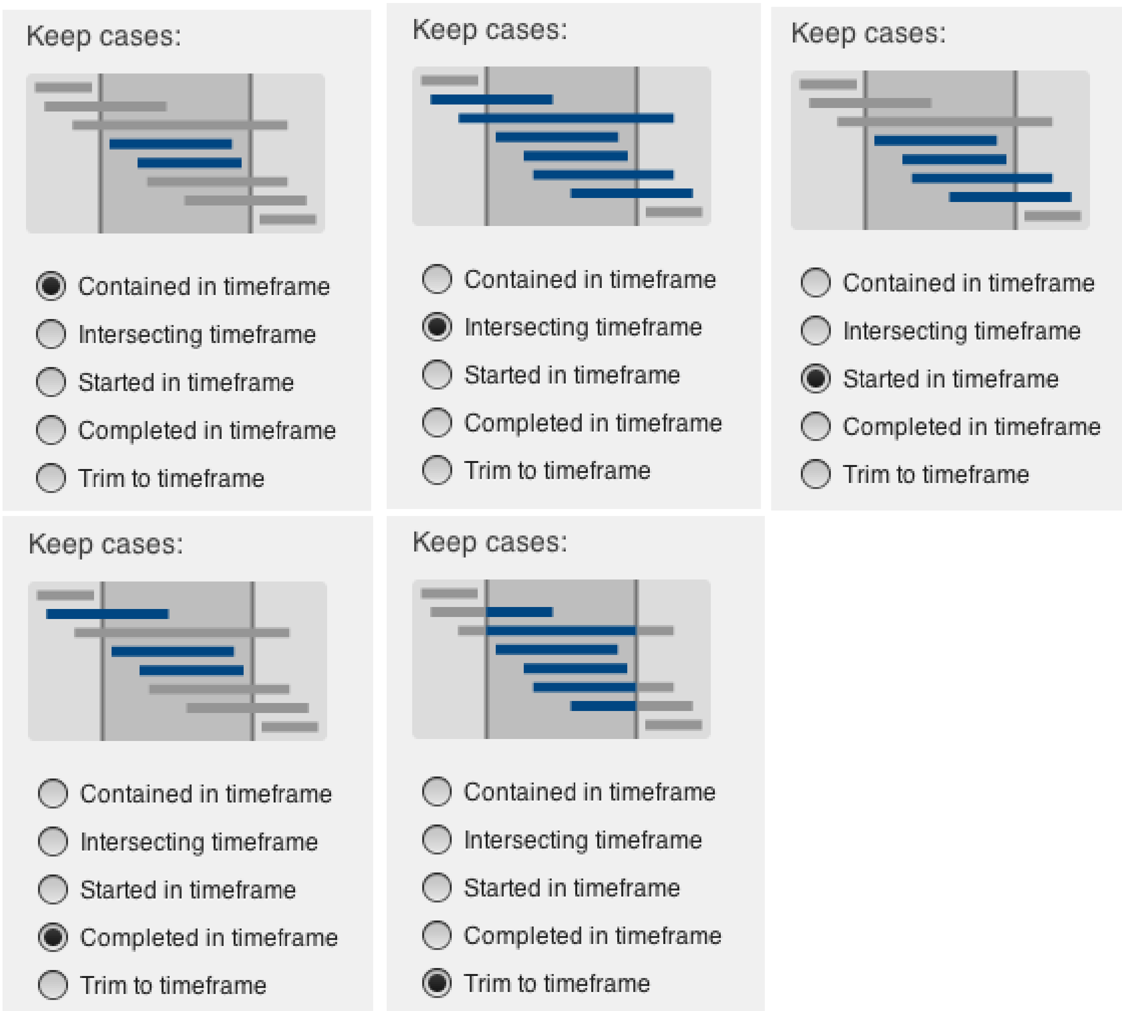


Figure 4.13. Cases to keep based on user preference

Contained in timeframe: cases which started and finished on the given boundary of time are kept.
Intersecting timeframe: The reference point for selecting an intersecting time frame is the earliest and latest timestamp for every individual case.
Started in timeframe: keep cases that are within a selected start timeframe.
Completed in timeframe: keep cases that are within a selected end timeframe.
Trim to timeframe: enables to arrange the cases in a given timeframe and select the cases that match a specific timeframe with start and end.

Activity	Frequency	Relative frequency	Median duration
Test Repair	1508	19.5 %	8 mins
Register	1104	14.27 %	0 millis
Analyze Defect	1104	14.27 %	8 mins
Inform User	1102	14.25 %	0 millis
Archive Repair	1000	12.93 %	0 millis
Repair (Simple)	785	10.15 %	9 mins
Repair (Complex)	725	9.37 %	17 mins
Restart Repair	406	5.25 %	0 millis

Figure 4.13. Log filtered based on completion in the timeframe (activity view).

2. *The variation filter*

The variation filter provides a method to filter either the most common or exceptional behavior in a process. When the process is structured, the variation filter is important to select behaviors. Nevertheless, as the process gets complex the variation filter is of less functionality. In Figure. 4.14, one can see there are a total of 60 variants and variant 1 covers almost 30% of all the cases in the log. Moreover, the variation filter provides a coverage feedback to get the overview of how many percent of cases are covered by the current selection. In this case the coverage for common logs is 64%.

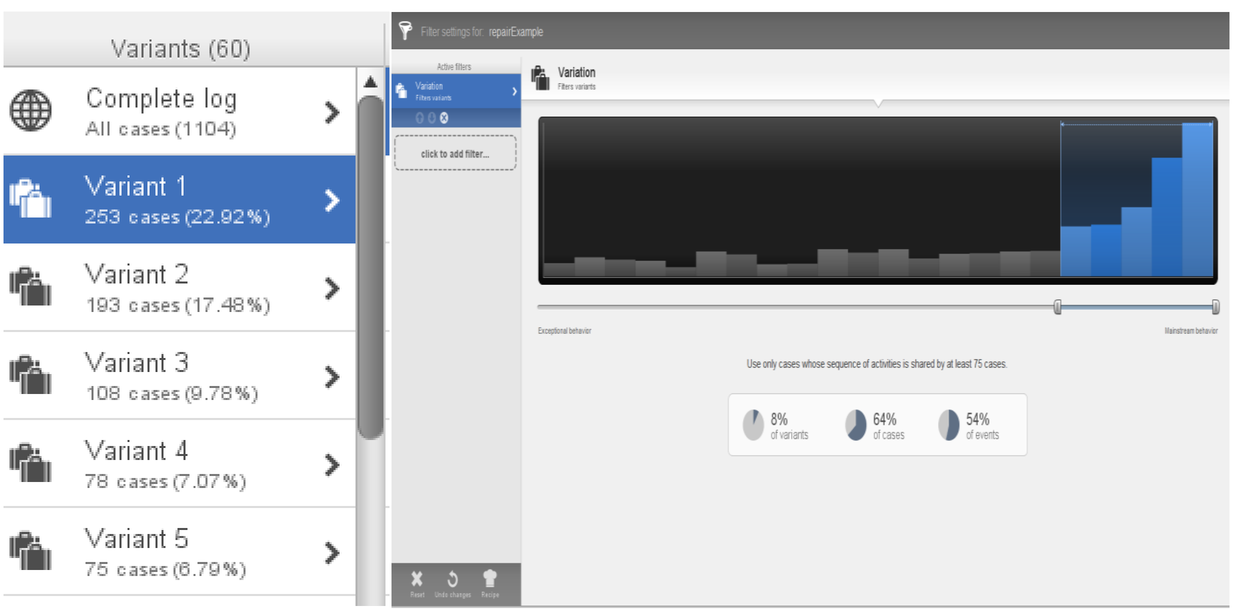


Figure 4.14. The variation filter for the telephone repair example for common behavior in the process.

3. *The performance filter*

The performance filter is another case filter in relation to a specific set of performance criteria. In case of we want to see all cases that are below or above a given throughput time, the performance filter allows us to do that exactly.

The following option enables the use of performance filter much easier in Disco.

- *Filter cases by case duration*: From case duration filter, one can discover that 87% of the cases took a time of 39 minutes to an hour and half.

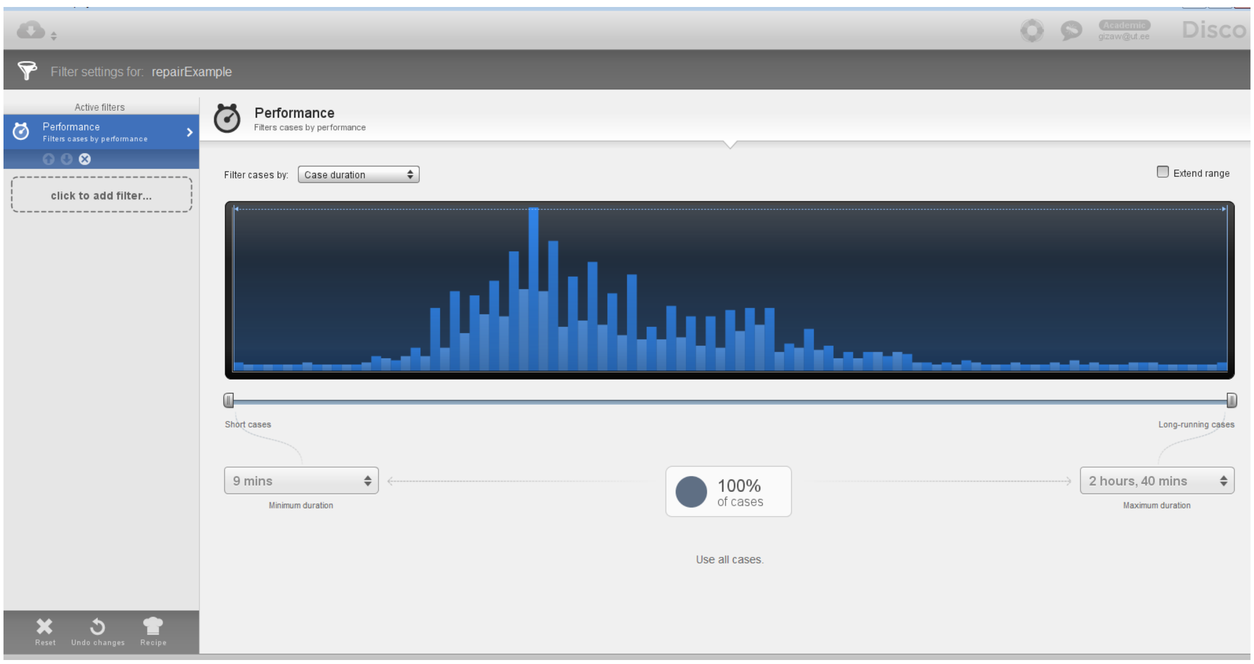


Figure 4.15. The performance filter by case duration

- *Filter cases by number of events*: From all cases, minimum number of events in a given case is 3 while the maximum number of event ranges to 16.

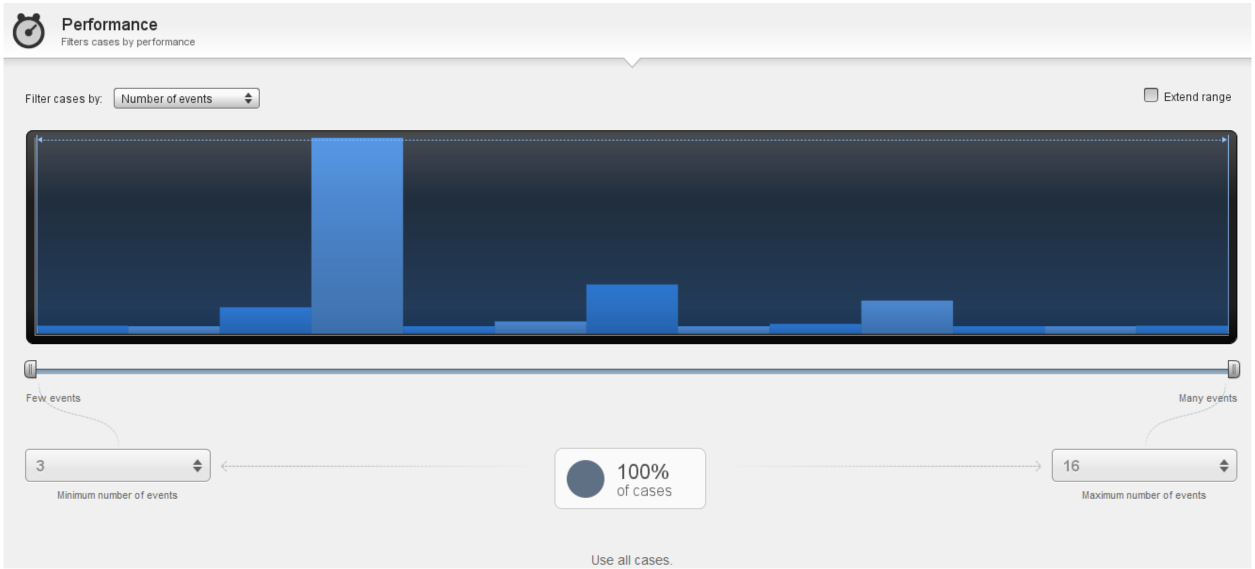


Figure 4.16. The performance filter by number of events

- *Filter cases by case utilization:* Case utilization is the ratio active vs waiting time. If case utilization ranges to 100%, it indicates that the process is running actively and there is no waiting times. The lower the utilization rate, the larger the waiting time. In our specific sample log, Figure. 4.17. Shows that there are cases with longer waiting time (32%) to performing actively (99%)

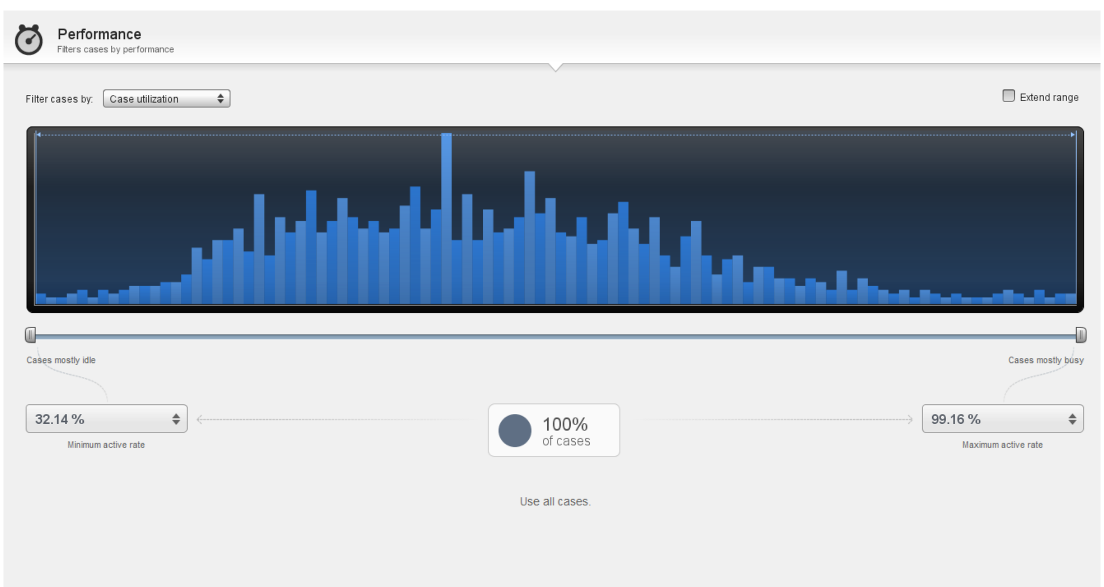


Figure 4.17. The performance filter by number of events

- *Filter cases by total active time:* indicates active time for the whole case

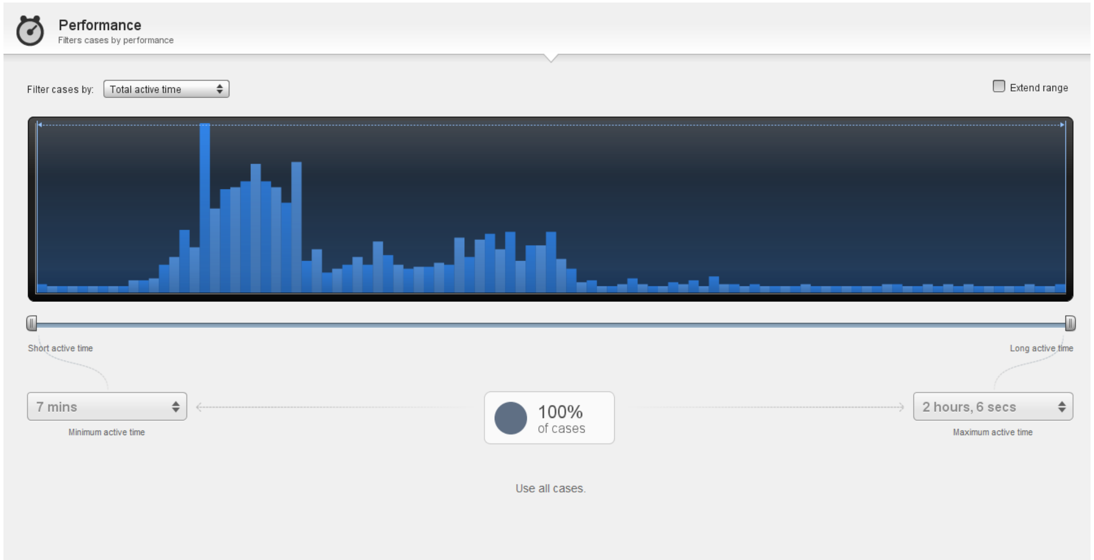


Figure 4.18. The performance filter by number of events

- *Filter cases by maximal active time*: indicates the average time block in active cases which ranges from 3 min to 33 min.

- *Filter cases by total waiting time (mean, maximal)*: indicates the sum of idle time for the whole case

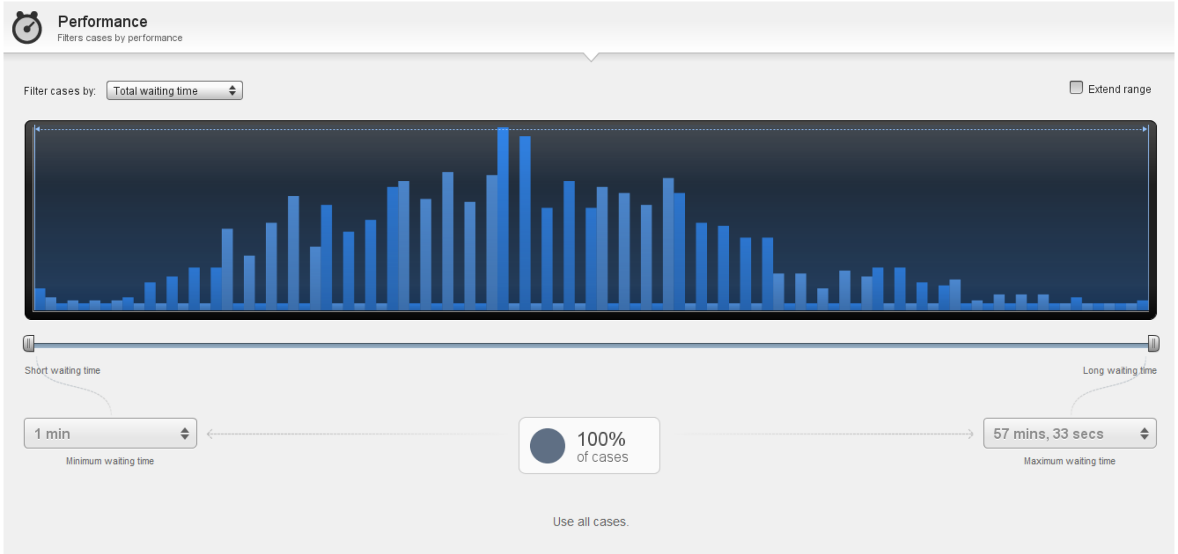


Figure 4.19. The performance filter by waiting times

4. *The endpoint filter*

The endpoint filter trims out events that are not in the boundary of the starting activity and end activity. In the trim filter, we can either focus on a specific events to start and end.

5. *The attribute filter*

The attribute filter allows to select events based on activity and resources. In the sample log, the attribute filer supports to select cases in which the repair for the telephone based on phones that have been fixed or not.

Process discovery

In Disco, after the necssary filter is applied , one can discover a fuzzy model with all the required information attached. For instance, for a given activity the notification bar display the respective frequency and performance.

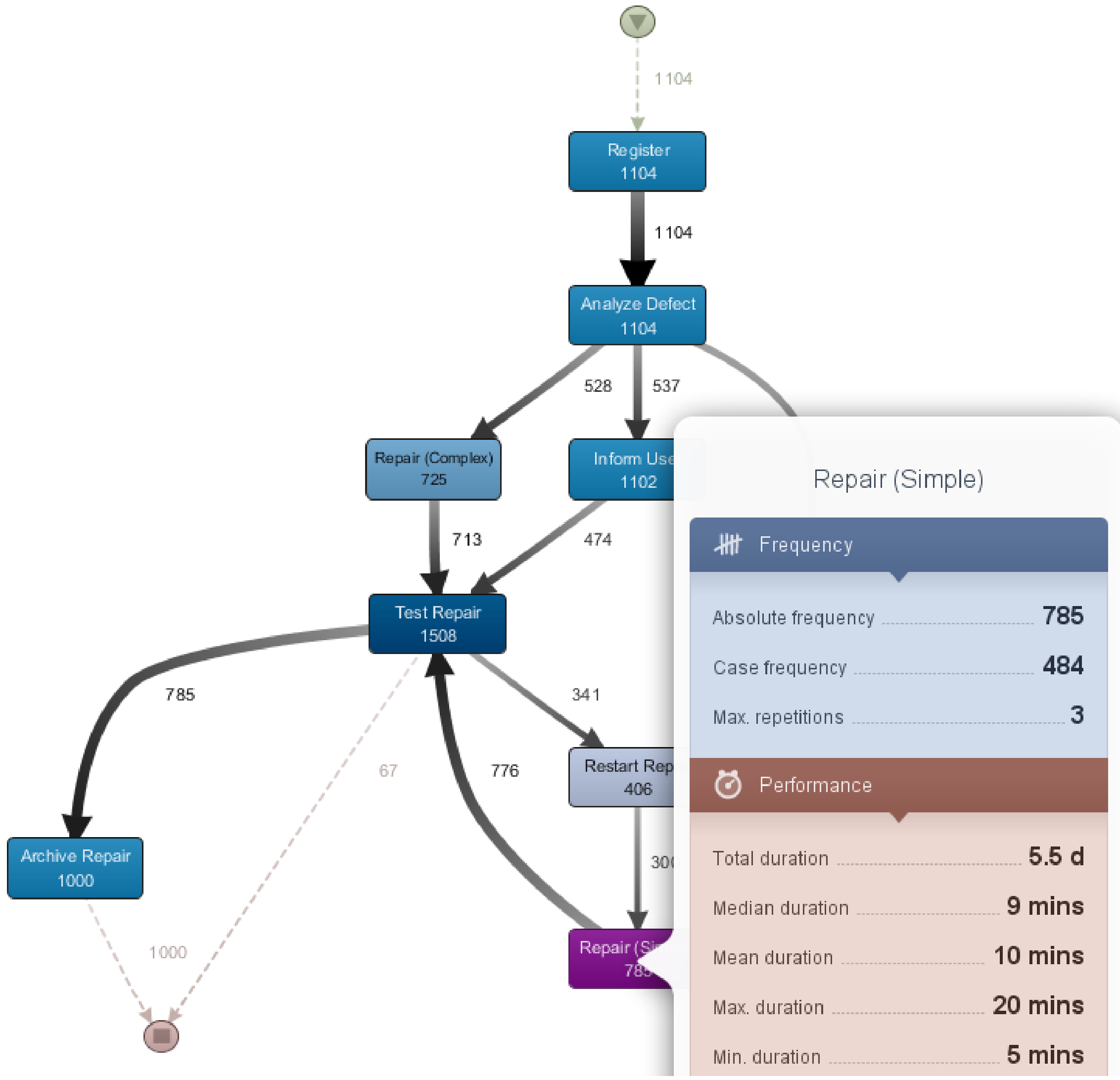


Figure 4.20. Automatic process discovery in Disco for the telephone repair process

Process visualisation

This feature of Disco to animate processes, makes it easy to spot bottlenecks on the fly. Moreover, it provides a capability to export the animation in video format.

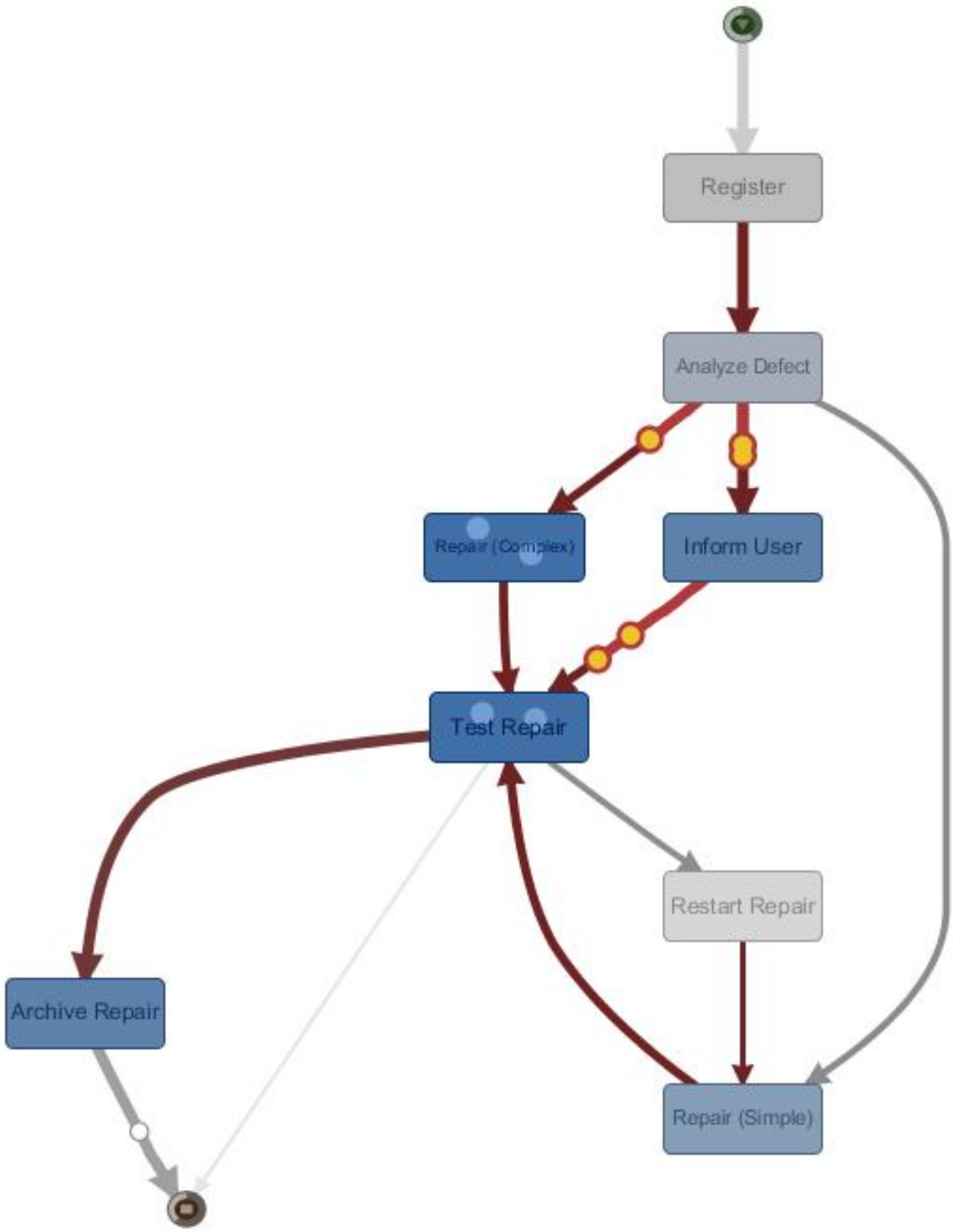


Figure 4.21. Process map animation in Disco for the telephone repair

4.4. Celonis

The Celonis process mining tool presents real time process discovery technique while information systems are in operation. It supports SAP databases and other various

RDBMS. In order to import event log in to the Celonis, we need to establish data source connection with the dedicated PostgreSQL. Celonis can only intake a time stamp that is calculated (In the telephone repair case, complete time – start time).

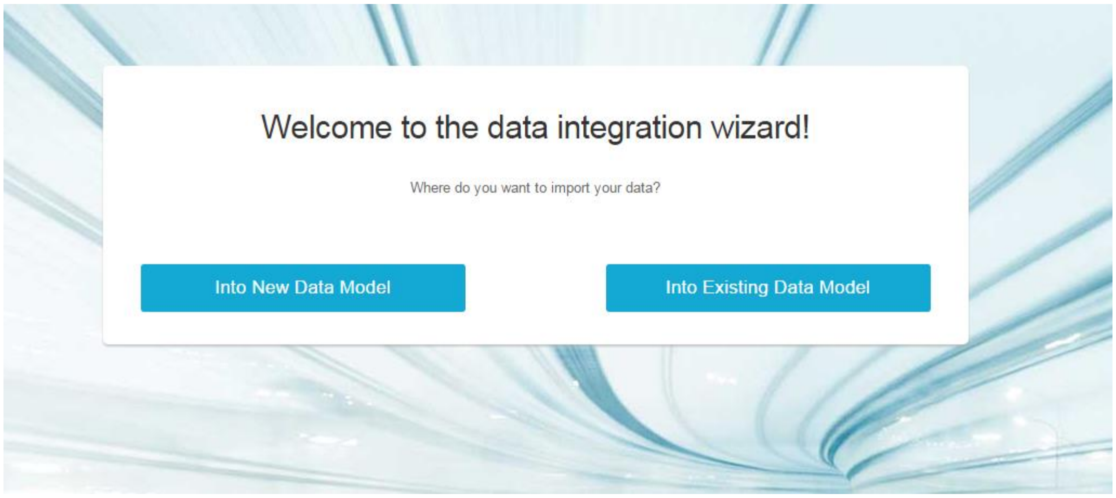


Figure 4.22. Log Data import wizard for Celonis (a)

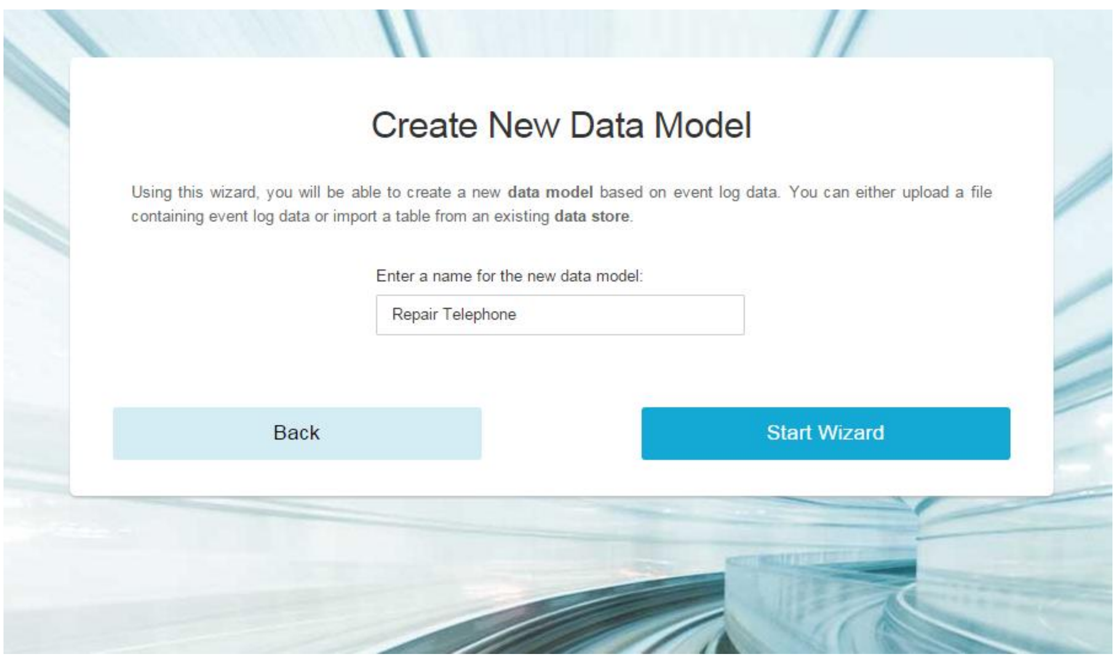


Figure 4.22. Log Data import wizard for Celonis (b)

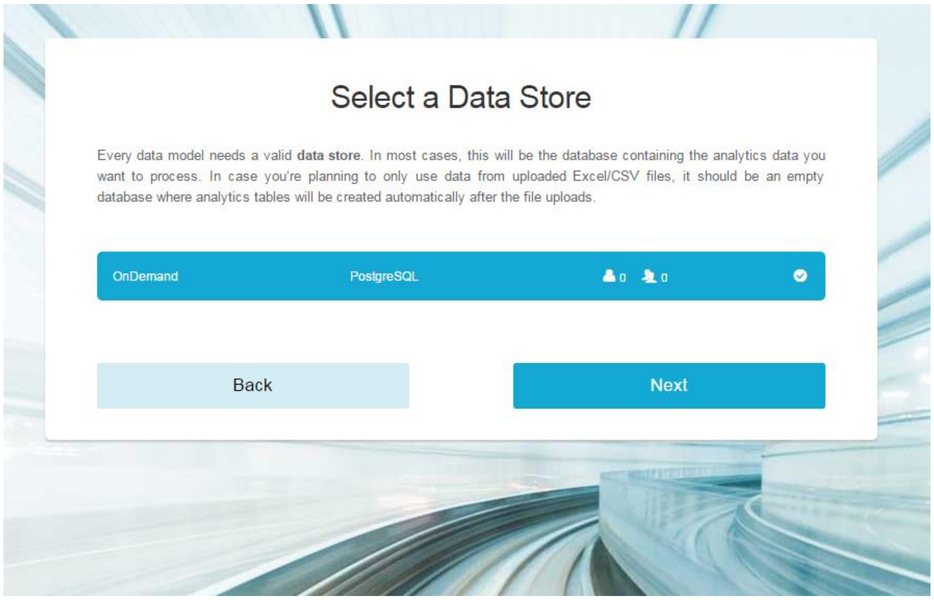


Figure 4.22. Log Data import wizard for Celonis (c)

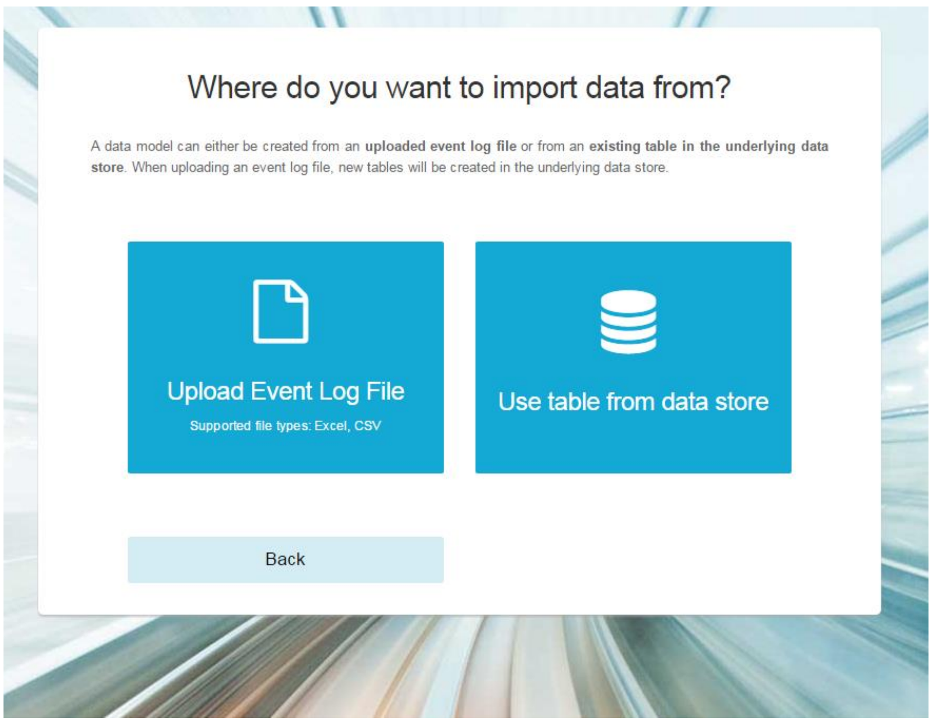


Figure 4.22. Log Data import wizard for Celonis (d)

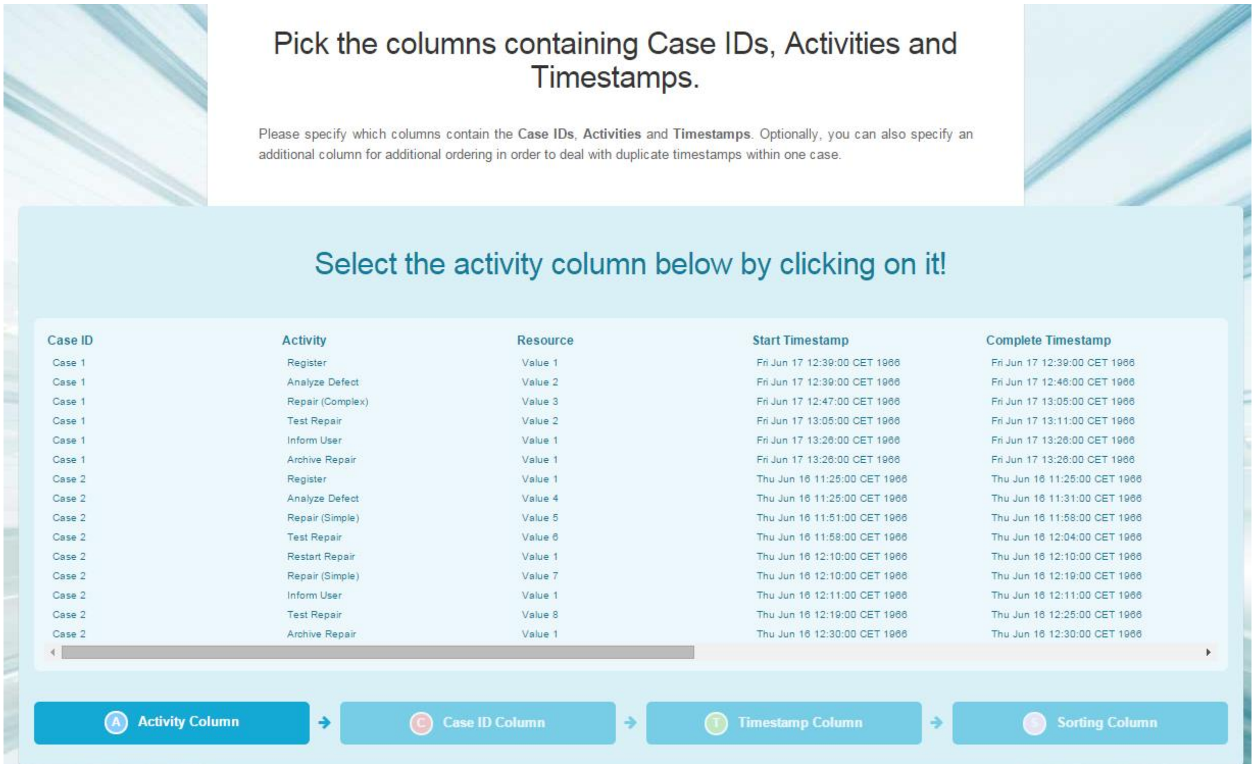


Figure 4.22. Log Data import wizard for Celonis (e)

Once the log is imported, we create a document which is configured as a database table which we can apply component functions to the underlying log (Figure 4.23).

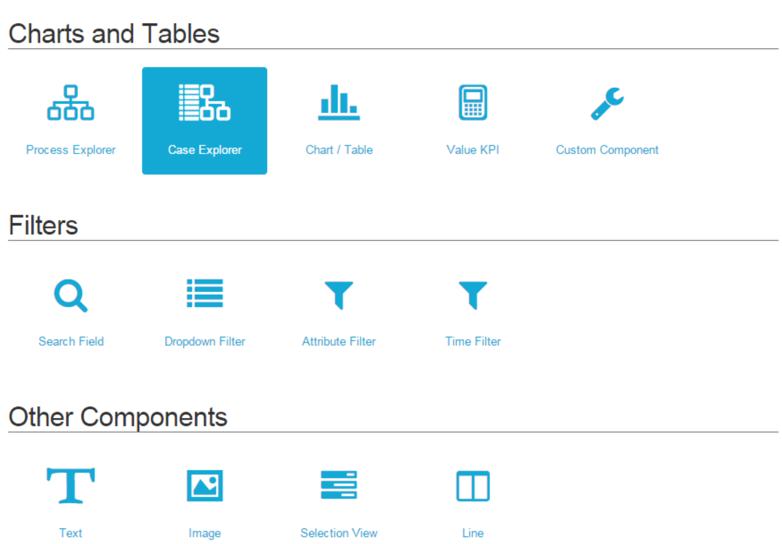


Figure. 4.23 Component functions in Celonis

Filtering Data

Attribute filter: Under the use of several attribute filters Celonis process mining chooses the cases fitting all attribute filters.

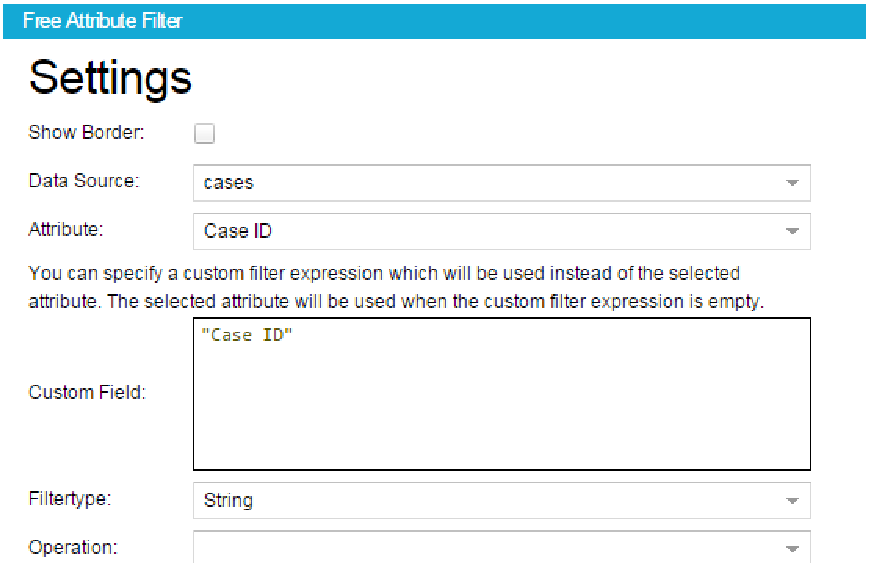


Figure. 4.24 Attribute filter in Celonis

Variant filter:

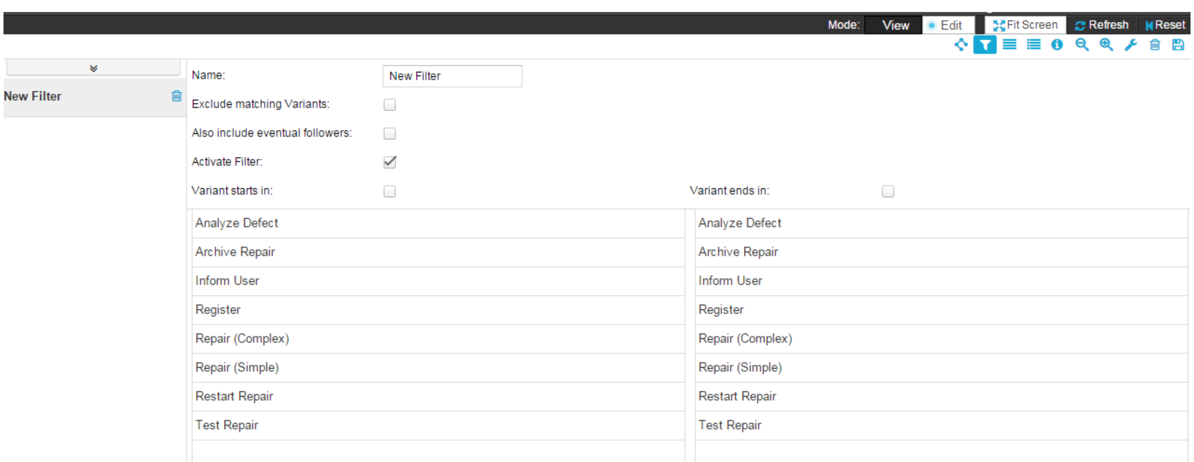


Figure. 4.25 Variant filter in Celonis

PQL Filter: Uses the specifically designed process query language to filter and analyze process patterns regardless of complexity.

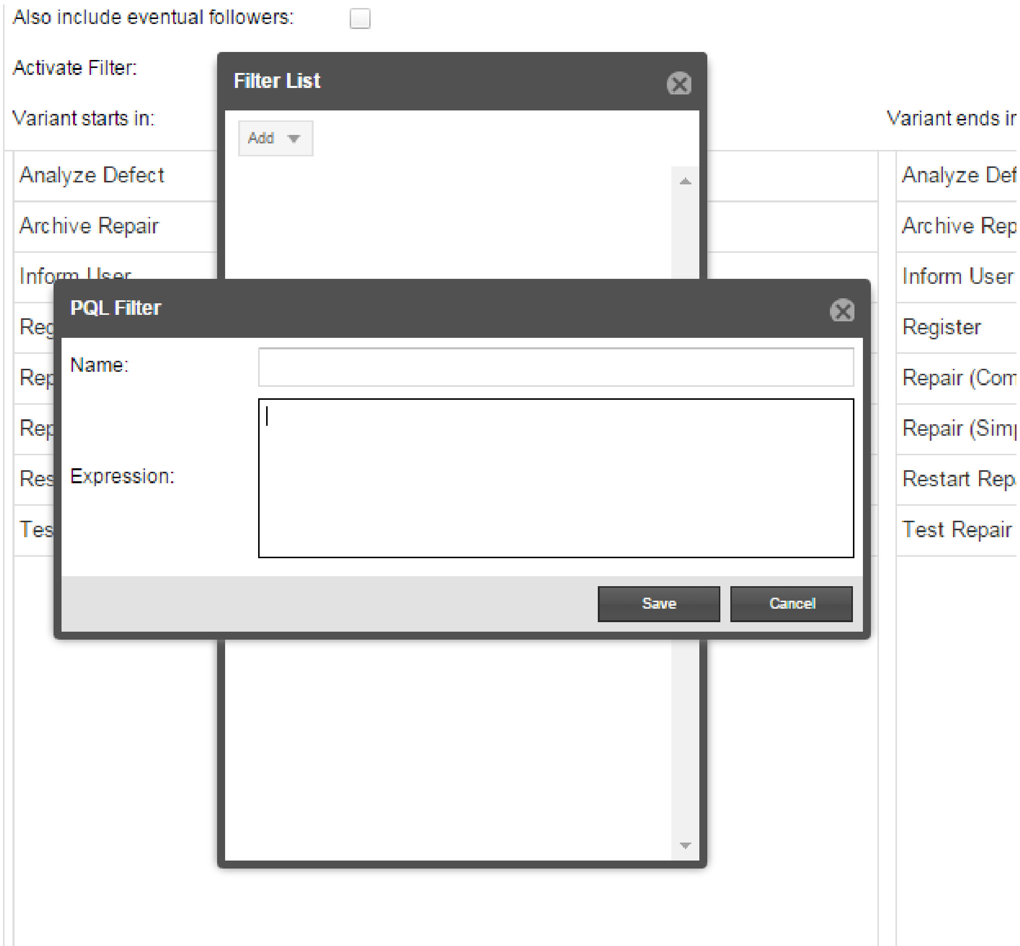


Figure. 4.26 PQL filter in Celonis

Process discovery

A process discovery in Celonis works in similar way as that of Disco except, while in Celonis the path that is most frequent is can be either expanded or narrowed below as depicted below in Figure. 4.27

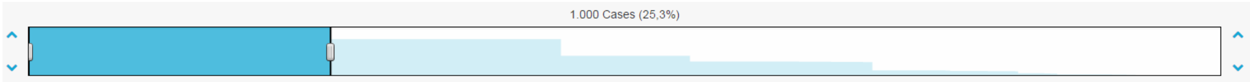


Figure. 4.27 Detail viewer in Celonis

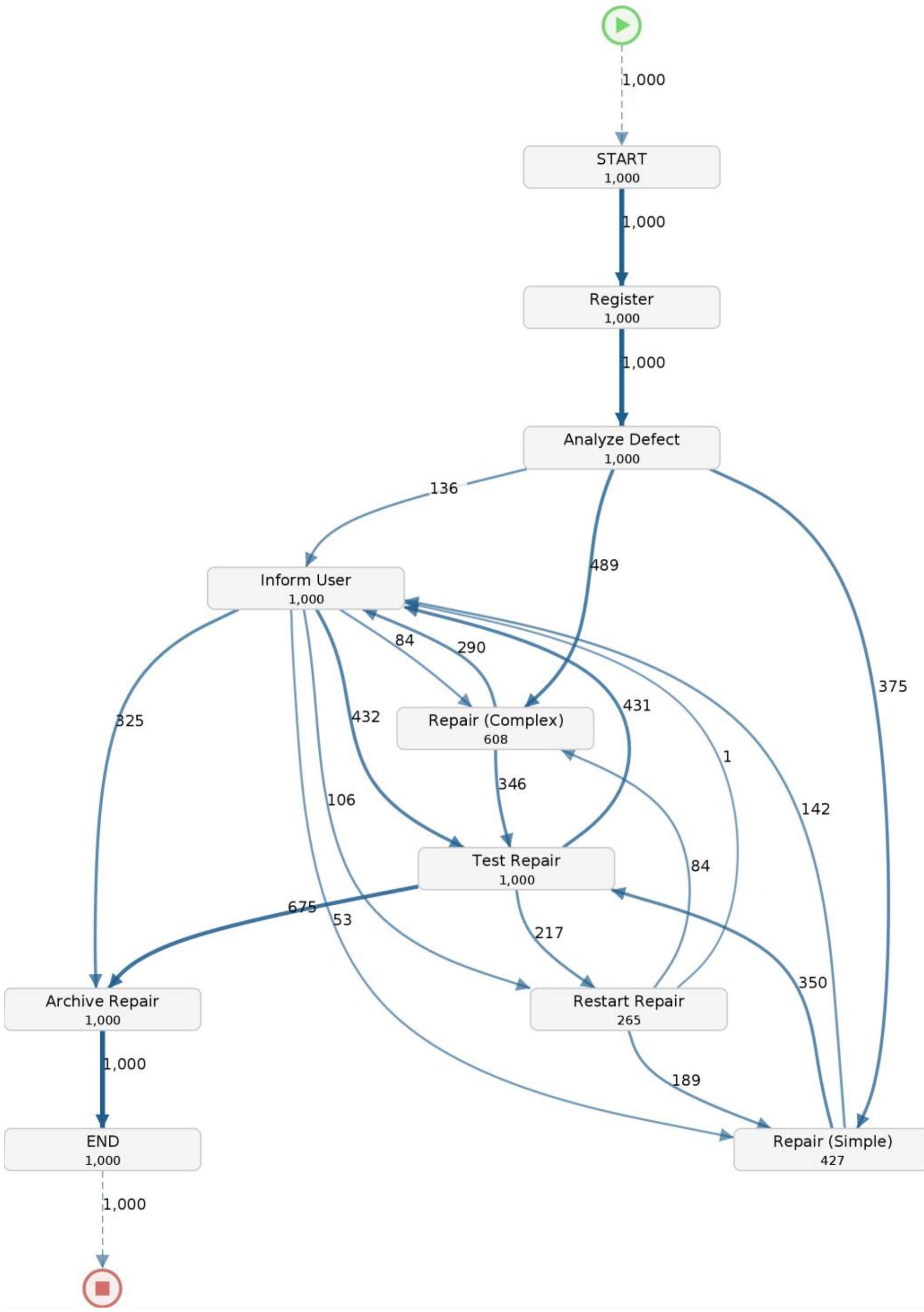


Figure. 4.28 Process discovery for the telephone repair process in Celonis

4.5. Summary and discussion

The following section presents a summary of the findings as depicted in Figure. 4.29. Furthermore, the support of a tool for a given feature is labeled by the mark ✓ and features that are not supported are labeled by the mark ×

Features	**ProM (V. 6.4.1)**	**Disco (V. 1.8.2)**	**Celonis**
Import Type Support	Mxml,xes	Csv,xls, mxml, xes and fxl	Csv,xls
Import Log Size Capacity	unlimited	up to 5 million events	Based on the quota of database
license	Open source	Evaluation / commercial	Evaluation / commercial
Output model notation	BPMN, WF, Petri nets, EPCs, transition systems, heuristics	Fuzzy model	Fuzzy model and support of charts
Supported platform	Standalone desktop version	Standalone desktop version	On premise/ Software as a service version
Filtering data	✓	✓	✓
Process discovery	✓	✓	✓
Conformance checking	✓	×	×
Social network mining	✓	×	×
Decision rule mining	✓	×	×
Process visualization	✓	✓	✓
Performance reporting	✓	✓	✓
Discriminative rule mining	✓	×	×
Trace clustering	✓	×	×
Delta analysis	✓	✓	✓

Figure. 4.29. Summary for comparison of process mining tools (ProM, Disco and Celonis)

Disco provides a filtering mechanism that is more transparent and clearly represented compared to the other tools. ProM has a variety of filtering options compared to the other tools. Process discovery is one of the common operations that can be performed by any of the tools. It is possible to present the discovered model in various formats based on the end user preference. In case of ProM, the discovered model can be converted to BPMN diagram, Petri net , herustics model, transition system, inductive mined model and fuzzy model format. but for Disco and Celonis the output uses fuzzy mining model to display the discovered model.

On the other hand, delta analysis can be performed regardless of the tools used since it is dependent on either the workflow model or the model originally used to configure the system initially. The usage of the process mining tools at this stage is for discovering the model that aligns together with the former model which is required to make delta analysis.

In this study, we have addressed the research question by discovering possible problems that can be answered using process mining , followed by discovery of a list of process mining operations that solve the problem itself. Inorder to select core problems and operations, mapping aligned to a set of case studies was used. The same mapping is used to compare the tools. Based on the results we have acquired , ProM can support all core operations. Disco and Celonis cover partial operations. Disco is more suitable for both a beginner or expert users due to its simplicity to use and fast processing of event logs.

Neverthless, in the Disco point of view , Celonis can do similar tasks except Celonis provides live analysis of process and a visualization. On the other hand, Celonis convert event logs to database elements before analaysing the data. Moreover, the timestamp must include a full set of activity timespan rather than starting and ending time in case of a given activity. The data backed up in an online or on premise database server. Based on overall observation, if there is a data that must be monitored on regular basis, Celonis is preferred. Neverthless for a usage in projects that are either small or big with a requirement of fast,efficent and easy end user experiecnce , Disco is preferable. Moreover, any of the tools can be used together while making the process discovery, bootlneck analysis and later can be exported as xes format in to ProM for further analysis. Regarding ProM , it can support all the operations specified and can be used with a variety of available mining algorithms. The results can be used by the practisioners and other users that are interested in using one of these 3 process mining tools.

5. Conclusion

The main goal of this thesis was to compare and analyze process mining tools based on a well-defined framework. In order to approach the problem, a list of process mining problems with possible operations were acquired in the first round. Later, the mapping of problem to operation was categorized based on a case studies in the field of process mining. The criteria for the selection of the cases studies was based on its relevance to the categorization in to a scenario having a specific problem and implemented process mining operation. In another words, the studies mention a certain problem and utilize process mining tool in order to solve the problem. With the summary of the problem, possible operations and case studies, a framework has been developed.

The developed framework select and focus mainly on the operations that use process mining tool. Since, our focus was to approach the comparative analysis on the basis of the supported/unsupported operations. Therefore, we were able to compare ProM, Disco and Celonis using a telephone repair event log sample to show whether the operation is supported or not in the respective tools.

Despite the lack of in depth analysis on a real life log, the framework can be used to understand process mining tools from a different perspective unlike similar works. The approach of analyzing process mining tools based on process mining techniques/operations is much more precise for a practitioner expecting to understand which specific operations the tools can perform.

Although the research has reached its aim, there were some unavoidable limitations. First, we evaluated the tool using one specific scenario which is based on core techniques of process mining tools as implemented by case studies. Therefore, it was not possible to generalize the results. Second, the fact that we analyzed the tools using the documentation that is insufficient in availability so that the interpretation of the findings might be affected. Finally, the conclusions of the comparison might be different if the evaluation has been performed by expert users of the respective tools or by the tool vendors themselves.

References

[1] Jiafei Li, Dayou Liu, Bo Yang: Process Mining: Extending alpha -Algorithm to Mine Duplicate Tasks in Process Logs. APWeb/WAIM Workshops 2007:396-407

[2] K. G. Srinivasa, M. Jagadish, K. R. Venugopal, Lalit M. Patnaik: Data Mining based Query Processing using Rough Sets and Genetic Algorithms. CIDM 2007:275-282

[3] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, vol. 136. 2011, 285 p.

[4] Turner, Chris J., et al. "Process mining: from theory to practice." Business Process Management Journal 18.3 (2012): 493-512.

[5] A.K. Alves de Medeiros and C.W. G¨unther. Process mining: Using CPN tools to create test logs for mining algorithms. Proceedings of the Sixth Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools, 2005.

[6] S. Goedertier, D. Martens, J. Vanthienen, and B. Baesens. Robust Process Discovery with Artificial Negative Events. Journal of Machine Learning Research, 10:1305–1340, 2009.

[7] J. Munoz-Gama and J. Carmona. A Fresh Look at Precision in Process Conformance. In R. Hull, J. Mendling, and S. Tai, editors, Business Process Management (BPM 2010), volume 6336, pages 211–226, 2010.

[8] Diederik Verstraete. Process mining in practice: comparative study of process mining software (Masters Dissertation). Ghent University, Belgium, 2013
Available from http://lib.ugent.be/fulltxt/RUG01/002/165/042/RUG01-002165042_2014_0001_AC.pdf

[9] W. M. P. Van der Aalst, B. F. Van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters, “Workflow mining: A survey of issues and approaches,” Data and Knowledge Engineering, vol. 47, no. 2. pp. 237–267, 2003.

[10] A. Rozinat, A. K. A. de Medeiros, C. W. Guenther, A. J. M. M. Weijters, and W. M. P. van der Aalst, “The need for a process mining evaluation framework in research and practice - Position paper,” in business process management workshops, 2008, vol. 4928, pp. 84–89.

[11] IEEE Task Force on Process Mining. Process Mining Manifesto. In BPM Workshops, Lecture Notes in Business Information Processing. Springer-Verlag, Berlin, 2011.

[12] J. Mendling and M. Strembeck, “Influence Factors of Understanding Business Process Models,” in 11th International Conference on Business Information Systems, 2008, pp. 142–153.

[13] Mendling, J., Moser, M., Neumann, G., Verbeek, H., Dongen, B., Aalst, W.: Faulty EPCs in the SAP Reference Model. In S. Dustdar, J.F., Sheth, A., eds.: BPM 2006, Proceedings. LNCS 4102. (2006) 451457

[14] Mendling, J., Verbeek, H., Dongen, B., Aalst, W., Neumann, G.: Detection and

Prediction of Errors in EPCs of the SAP Reference Model. Data & Knowl. Eng. (2007) accepted for publication.
[15] W. M. P. van der Aalst, “Business alignment: Using process mining as a tool for Delta analysis and conformance testing,” *Requir. Eng.*, vol. 10, no. 3, pp. 198–211, 2005.
[16] J. Swinnen, B. Depaire, M. Jans, and K. Vanhoof, “A Process Deviation Analysis – A Case Study,” in *Business Process Management Workshops SE - 8*, vol. 99, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Springer Berlin Heidelberg, 2012, pp. 87–98.
[17] W. M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. Alves de Medeiros, M. Song, and H. M. W. Verbeek, “Business process mining: An industrial application,” *Inf. Syst.*, vol. 32, no. 5, pp. 713–732, 2007.
[18] Z. Paszkiewicz, “Process Mining Techniques in Conformance Testing of Inventory Processes: An Industrial Application,” in *Business Information Systems Workshops SE - 28*, vol. 160, W. Abramowicz, Ed. Springer Berlin Heidelberg, 2013, pp. 302–313.
[19] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. Van der Aalst, and P. J. M. Bakker, “PROCESS MINING IN HEALTHCARE - A Case Study,” in *Communications in Computer and Information Science, volume 25*, 2009, pp. 425–438.
[20] S. Suriadi, M. T. Wynn, C. Ouyang, A. H. M. Ter Hofstede, and N. J. Van Dijk, “Understanding process behaviours in a large insurance company in Australia: A case study,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7908 LNCS, pp. 449–464.
[21] F. Caron, J. Vanthienen, K. Vanhaecht, E. Van Limbergen, J. Deweerdt, and B. Baesens, “A process mining-based investigation of adverse events in care processes,” *Heal. Inf. Manag. J.*, vol. 43, no. 1, pp. 16–25, 2014.
[22] R. Mans, H. Reijers, M. van Genuchten, and D. Wismeijer, “Mining processes in dentistry,” *Proc. 2nd ACM SIGHIT Symp. Int. Heal. informatics - IHI ’12*, p. 379, 2012.
[23] Wil M. P. van der Aalst: Process Mining - Discovery, Conformance and Enhancement of Business Processes, Springer 2011

Appendix

I. License

Non-exclusive licence to reproduce thesis and make thesis public

I, **Musie Kebede Gizaw** (date of birth: 03.29.1988),
(*author's name*)

1. Here with grant the University of Tartu a free permit (non-exclusive licence) to:
   1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and
   1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

of my thesis
Comparative Evaluation of Process Mining Tools,
(title of thesis)
supervised by Marlon Dumas,
(supervisor's name)

2. I am aware of the fact that the author retains these rights.
3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **25.05.2015**