

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

CÁSSIO DE ABREU RAMOS

**Estudo Comparativo entre Ferramentas de
Mineração de Processos de Negócio**

Monografia apresentada como requisito parcial
para a obtenção do grau de Bacharel em Ciência
da Computação

Orientador: Prof. Lucineia Heloisa Thom

Porto Alegre
2019

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof^a. Jane Fraga Tutikian

Pró-Reitor de Graduação: Prof. Wladimir Pinheiro do Nascimento

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Sergio Luis Cechin

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

RESUMO

Este estudo realiza um comparativo entre as ferramentas de mineração de processos de negócio mais populares existentes em âmbito tanto acadêmico quanto industrial. Procura determinar diferenças e semelhanças entre as ferramentas, organizadas em duas categorias - *open-source* e comerciais - de forma a sugerir qual atende melhor o usuário de mineração de processos conforme seu perfil e objetivos pretendidos. Foi realizado um estudo de caso utilizando um *log* de eventos real interpretado em duas dessas ferramentas, uma de cada categoria, comparando suas funcionalidades e detectando diferenças e semelhanças. Ferramentas *open-source* tendem a ser mais completas, buscando abordar todas as características contidas na mineração de processos, enquanto ferramentas comerciais tendem a apresentar interfaces mais familiares a usuários iniciantes e mais alternativas para filtragem e exibição de dados. Este estudo busca auxiliar tanto o usuário iniciante quanto o avançado a encontrar a ferramenta que mais preenche suas expectativas no momento em que estão inseridos na área da mineração de processos de negócio.

Palavras-chave: Mineração de processos; ferramentas de mineração; processos de negócio.

Comparative Study Among Business Process Mining Tools

ABSTRACT

This is a comparative study among the most popular business process mining tools available at both academic and industry scope. The study aims to determine differences and similarities among the tools, arranged in two categories - *open-source* and commercial – in order to suggest which one best suits a process mining user, according to his profile and intended goals. A case study was performed using a real-life event log interpreted by two of these tools, one of each category, comparing features and detecting differences and similarities among them. *Open-source* tools tend to be more comprehensive, seeking to address all characteristics contained in process mining, while commercial tools tend to provide more familiar interfaces to novice users and more alternatives for filtering and displaying data. This study intends to help both the beginner and the advanced user to identify the tool that most fulfills their expectations at the time these users are inserted in the area of business process mining.

Keywords: Process mining; process mining tools; business processes.

LISTA DE FIGURAS

| | | |
|-------------|------------------------------------------------------------------------------|----|
| Figura 2.1 | Ciclo de vida de BPM..... | 12 |
| Figura 2.2 | Visão geral das fases de mineração de processos | 15 |
| Figura 3.1 | Importação de um <i>log</i> de eventos em ProM | 22 |
| Figura 3.2 | Ações possíveis em ProM após a importação do <i>log</i> de eventos..... | 23 |
| Figura 3.3 | Filtragem de <i>Log</i> de eventos por heurística simples..... | 23 |
| Figura 3.4 | Visão geral dos <i>Plug-ins</i> de ProM e seus tipos..... | 25 |
| Figura 3.5 | Análise de rede social em ProM | 26 |
| Figura 3.6 | Modelo de descoberta de processos em Apromore | 28 |
| Figura 3.7 | Modelo baseado em frequência em Apromore..... | 28 |
| Figura 3.8 | Modelo de processos em Apromore em formato BPMN | 29 |
| Figura 3.9 | Modelo de processos agrupado em estágios em Apromore | 30 |
| Figura 3.10 | Checagem de conformidade em Apromore | 30 |
| Figura 3.11 | Monitor preditivo em Apromore..... | 31 |
| Figura 3.12 | Filtro por período de tempo em bupaR..... | 33 |
| Figura 3.13 | Modelo de processo em bupaR..... | 34 |
| Figura 3.14 | Exemplo de checagem de conformidade em bupaR..... | 35 |
| Figura 3.15 | <i>Logs</i> de eventos exibidos em Disco | 39 |
| Figura 3.16 | Filtros disponíveis para utilização em Disco | 40 |
| Figura 3.17 | Informações sobre a ferramenta Celonis | 41 |
| Figura 3.18 | Importação de <i>logs</i> de eventos em Celonis..... | 42 |
| Figura 3.19 | Descoberta de processos em Celonis..... | 43 |
| Figura 3.20 | Checagem de conformidade em Celonis | 43 |
| Figura 3.21 | Análise de rede social em Celonis | 44 |
| Figura 3.22 | Descoberta de processos em ARIS PPM..... | 45 |
| Figura 3.23 | Checagem de conformidade em ARIS PPM | 46 |
| Figura 3.24 | QPR Process Analyzer Presentation Object | 47 |
| Figura 4.1 | <i>Log</i> de eventos analisado | 53 |
| Figura 4.2 | Modelo de processo em ProM | 53 |
| Figura 4.3 | Modelo de processo agrupado em ProM | 54 |
| Figura 4.4 | Modelo de processo em Disco..... | 55 |
| Figura 4.5 | Modelo de processo com duas métricas em Disco | 56 |
| Figura 4.6 | Informações adicionais sobre um caminho no modelo de processo em Disco | 56 |
| Figura 4.7 | <i>Log</i> de eventos filtrado por atributos de evento em ProM | 57 |
| Figura 4.8 | Análise por instância de processo em ProM..... | 58 |
| Figura 4.9 | Dados gerais interpretados após o <i>log</i> em ProM | 58 |
| Figura 4.10 | Informações gerais do processo em Disco | 59 |
| Figura 4.11 | Atividades do processo em Disco..... | 59 |
| Figura 4.12 | Análise das variantes do processo em Disco | 60 |
| Figura 4.13 | Filtragem por atributos e ponto final em Disco | 61 |
| Figura 4.14 | Checagem de conformidade em ProM | 62 |
| Figura 4.15 | Obtenção de regras de decisão em ProM..... | 62 |
| Figura 4.16 | Modelo visualizado em animação em ProM | 63 |
| Figura 4.17 | Modelo visualizado em animação em Disco | 64 |
| Figura 4.18 | Análise de performance em ProM | 64 |

LISTA DE TABELAS

| | |
|---------------------------------------------------------------------------------|----|
| Tabela 2.1 Exemplo de informações encontradas em um <i>log</i> de eventos | 16 |
| Tabela 2.2 Comparativo entre Trabalhos Relacionados | 20 |
| Tabela 3.1 Comparativo entre Ferramentas <i>Open-Source</i> | 36 |
| Tabela 3.2 Comparativo entre Ferramentas Comerciais | 48 |
| Tabela 4.1 Tabela comparativa entre ferramentas estudadas no Capítulo 4..... | 65 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|------|---------------------------------------------------|
| BPMN | Business Process Management Notation |
| YAWL | Yet Another Workflow Language |
| MXML | Mining eXtensible Markup Language |
| PNML | Petri Net Markup Language |
| XPDL | XML Process Definition Language |
| IEEE | Institute of Electrical and Electronics Engineers |
| CRM | Customer Relationship Management |
| ERP | Enterprise Resource Planning |
| EPC | Event-Driven Process Chains |
| XES | eXtensible Event Stream |
| CSV | Comma Separated Values |
| PNG | Portable Network Graphics |
| AML | ARIS Markup Language |
| CTL | Computation Tree Logic |
| BPM | Business Process Management |
| SVG | Scalable Vector Graphics |
| PDF | Portable Document Format |
| PQL | Process Query Language |
| BI | Business Intelligence |

SUMÁRIO

| | |
|-------------------------------------------------------------------|-----------|
| 1 INTRODUÇÃO | 9 |
| 1.1 Principais Contribuições | 9 |
| 1.2 Organização do Texto | 10 |
| 2 FUNDAMENTOS SOBRE BPM E MINERAÇÃO DE PROCESSOS | 11 |
| 2.1 Fundamentos sobre Gerenciamento de Processos de Negócio | 11 |
| 2.2 Conceitos básicos de BPM | 13 |
| 2.3 Mineração de Processos de Negócio | 13 |
| 2.4 Trabalhos Relacionados | 16 |
| 2.5 Comparativo entre Trabalhos Relacionados | 18 |
| 2.6 Resumo do Capítulo | 19 |
| 3 ANÁLISE DAS FERRAMENTAS | 21 |
| 3.1 Ferramentas Open-Source | 21 |
| 3.1.1 ProM | 21 |
| 3.1.2 Apomore | 26 |
| 3.1.3 bupaR | 31 |
| 3.1.4 Comparativo Entre as Ferramentas <i>Open-Source</i> | 35 |
| 3.2 Ferramentas Comerciais | 37 |
| 3.2.1 Disco | 38 |
| 3.2.2 Celonis | 41 |
| 3.2.3 ARIS PPM | 44 |
| 3.2.4 QPR Process Analyzer | 46 |
| 3.2.5 Comparativo Entre as Ferramentas Comerciais | 48 |
| 3.3 Resumo do Capítulo | 50 |
| 4 ESTUDO DE CASO | 52 |
| 4.1 Modelo de descoberta | 52 |
| 4.2 Filtragem de dados | 57 |
| 4.3 Checagem de conformidade | 60 |
| 4.4 Regras de decisão | 61 |
| 4.5 Visualização de processos | 62 |
| 4.6 Análise de performance | 63 |
| 4.7 Análise dos resultados obtidos | 64 |
| 4.8 Resumo do capítulo | 66 |
| 5 CONCLUSÃO | 67 |
| REFERÊNCIAS | 69 |

1 INTRODUÇÃO

Nos últimos anos, a área de mineração de processos de negócio tem atraído a atenção de diversos pesquisadores e profissionais, buscando compreender as diferentes características da área. As pesquisas na área de mineração têm sido motivadas pela necessidade de inúmeras organizações, que desejavam entender como seus processos são executados no mundo real (KEBEDE, 2015). Antigamente, na tentativa de melhorar os processos de suas organizações, os responsáveis por elas apenas “desenhavam caixas e setas” como forma de representar modelos de processos, mas raramente esses modelos eram conectados de fato aos dados manipulados nessa organização (AALST, 2018). A mineração de processos apresentou-se com o objetivo de conectar modelos de processos com dados de eventos (AALST, 2018), capturados por sistemas de informação vinculados a essas organizações. Objetivando suprir essas necessidades, surgiram diversas ferramentas visando interpretar dados conforme as fases e objetivos da mineração de processos. Entretanto, a diversidade de ferramentas disponíveis, tanto em âmbito acadêmico quanto industrial (KEBEDE, 2015), pode resultar em uma escolha não tão adequada da ferramenta que melhor atenda todas as expectativas de um usuário ou de uma organização.

Neste contexto, o presente trabalho consiste em traçar um comparativo entre as ferramentas de mineração de processos de negócio mais populares nos ambientes acadêmico e industrial baseado em critérios e parâmetros existentes nessa área. São comparadas sete ferramentas de mineração existentes no mercado: ProM (AALST et al., 2009), Apromore (ROSA et al., 2011), BupaR (JANSSENSWILLEN; DEPAIRE, 2017), Disco (GÜNTHER; ROZINAT, 2012), Celonis (CELONIS, 2019), ARIS PPM (SOFTWARE AG, 2019) e QPR Process Analyzer (QPR, 2019). Tais ferramentas são classificadas em duas categorias: *open-sources* e comerciais. Essas categorias são comparadas através de um estudo de caso. No contexto deste estudo de caso, foi analisada uma ferramenta de cada categoria, de forma a estabelecer diferenças e semelhanças em seus comportamentos, buscando sugerir qual ferramenta melhor atende às expectativas de um usuário de mineração de processos, conforme seu perfil (iniciante ou avançado).

1.1 Principais Contribuições

A principal contribuição deste estudo está em estabelecer parâmetros de comparação entre as ferramentas de mineração de processos. Esses parâmetros são estabelecidos

com o objetivo de auxiliar o usuário na tarefa de escolher a ferramenta ideal que atenda às necessidades desse usuário interessado em utilizar mineração de processos, inserido em um ambiente individual ou organizacional. Além disso, este trabalho apresenta como contribuições:

- a contextualização sobre mineração de processos, fundamentos de gerenciamento de processos de negócio (*BPM - Business Process Management*), assim como a relação entre os dois conceitos;
- o detalhamento de funcionalidades contidas em ferramentas de mineração e como essas ferramentas tratam os dados de entrada, contidos nos *logs* de eventos, e como (e se) processam as fases da mineração de processos;
- a comparação entre as ferramentas de mineração, tendo como base características que remetem às fases da mineração e a categoria a qual pertencem, buscando vantagens e desvantagens, diferenças e similaridades entre cada uma;
- a identificação, através de estudo de caso, das principais diferenças e similaridades entre as categorias de ferramentas de mineração de processos.

1.2 Organização do Texto

Este texto está organizado da seguinte forma. O Capítulo 2 apresenta conceitos fundamentais em BPM e o ciclo de vida de BPM. Além disso, discute a relação de mineração de processos de negócio com o ciclo de vida de BPM, assim como suas características, seus objetivos e suas fases. Por fim, o capítulo descreve trabalhos relacionados ao presente estudo.

O Capítulo 3 descreve as ferramentas de mineração de processos, organizadas entre *open-sources* e comerciais, assim como suas principais funcionalidades. O capítulo também traz comparações entre as ferramentas, com base em suas diferenças e semelhanças.

O Capítulo 4 traz um estudo de caso, onde um caso real de *log* de eventos é interpretado por uma ferramenta de cada categoria contida no Capítulo 3 e os resultados obtidos em cada ferramenta são discutidos, com o apoio de uma tabela comparativa entre as características avaliadas nesse capítulo.

O Capítulo 5 conclui o trabalho, discute limitações encontradas ao longo do desenvolvimento do trabalho e perspectivas futuras de pesquisa.

2 FUNDAMENTOS SOBRE BPM E MINERAÇÃO DE PROCESSOS

Este capítulo apresenta os conceitos fundamentais em gerenciamento de processos de negócio e mineração de processos, necessários para o entendimento do presente trabalho. Além disso, o capítulo discute trabalhos relacionados ao tema apresentado neste trabalho.

2.1 Fundamentos sobre Gerenciamento de Processos de Negócio

BPM é definido como uma ciência que tem como objetivos supervisionar como o trabalho é realizado em uma organização, garantir resultados consistentes e aproveitar oportunidades de melhoria como, por exemplo, reduzindo custos, tempo de execução e taxas de erro de processo (DUMAS et al., 2013). *BPM* não se trata de melhorar a execução de atividades individualmente, e sim de melhorar a execução de um conjunto de eventos relacionados entre si, atividades e pontos de decisão, agregando valor à organização e seus clientes (DUMAS et al., 2013). Esse conjunto ganha o nome de *processo de negócio*, ou apenas processo.

O ciclo de vida de *BPM* inclui diferentes fases (DUMAS et al., 2013), ilustradas na Figura 2.1. Na fase de *Identificação do processo*, um problema de negócio é detectado, e os processos referentes a esse problema são identificados, delimitados e relacionados entre si. O resultado dessa fase é uma arquitetura de processos - nova ou atualizada - que traz uma visão geral dos processos realizados em uma organização e as relações entre eles.

Na fase de *Descoberta do processo* (ou modelagem do processo *as-is*), após sua identificação, os processos são documentados (modelados), em seu estado atual, em sua forma *as-is*. O objetivo desta fase é descobrir e modelar os processos através de notações como, por exemplo, BPMN - *Business Process Management Notation*, Redes de Petri, YAWL - *Yet Another Workflow Language*, EPC - *Event-Driven Process Chains*, etc.

A fase de *Análise do processo* busca identificar, documentar, analisar e quantificar problemas detectados na fase de descoberta através de métricas de medição de desempenho. Tem como resultado uma coleção estruturada desses problemas.

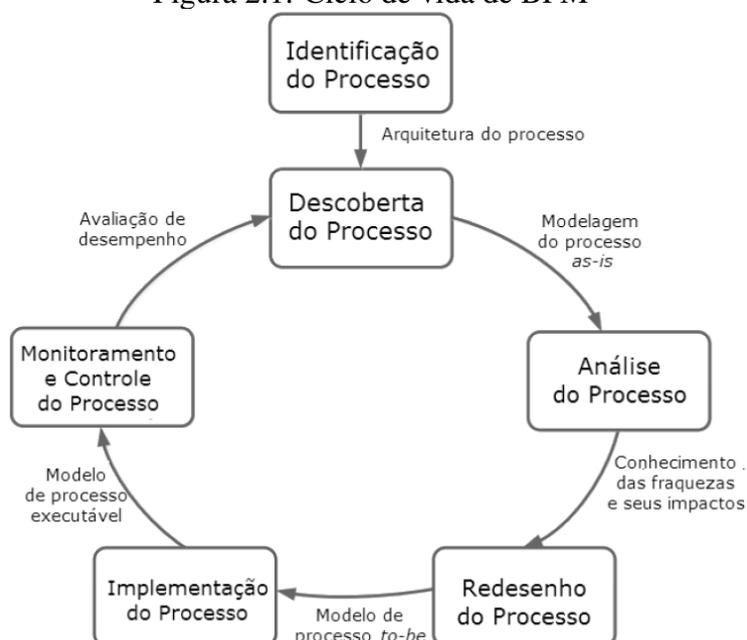
A fase de *Redesenho do processo* (ou melhoria do processo) tem como objetivo identificar alterações no processo que auxiliem a resolver os problemas identificados na fase de análise e apresentam como resultado o processo *to-be*, um modelo que visa pre-

encher os requisitos do processo que a organização tem como alvo e abordar possíveis problemas do processo atual (WANG; SUN, 2010).

Na fase de *Implementação do processo*, as modificações necessárias para evolução do modelo *as-is* para o modelo *to-be* são preparadas e efetuadas, visando abranger dois aspectos: gerenciamento de mudanças organizacionais e automação de processos

Finalmente, na fase de *Monitoramento e controle do processo*, dados relevantes são coletados nesta fase e analisados para verificação de desempenho do processo. Gargalos, erros recorrentes e desvios são identificados e ações de correção são realizadas.

Figura 2.1: Ciclo de vida de BPM



Fonte: (DUMAS et al., 2013)

Potencialmente, a mineração de processos desempenha um papel em cada uma das fases do ciclo de vida de BPM, com exceção da fase de *implementação* (AALST et al., 2011) (VERSTRAETE, 2014). É um instrumento valioso para as fases descritas na Figura 2.1. As técnicas de mineração de processos - que são detalhadas na Seção 2.3 - podem ser utilizadas para suporte operacional, para influenciar casos em execução com o apoio de modelos aprendidos usando informações históricas, para ajustar processos e guiar sua (re)configuração com o apoio de suporte à decisão (AALST et al., 2011).

2.2 Conceitos básicos de BPM

Como visto neste capítulo, um *processo de negócio* é composto por eventos, atividades e pontos de decisão. Um *evento* representa uma ação ocorrida instantaneamente (DUMAS et al., 2013) como, por exemplo, o recebimento de um pedido.

Uma *atividade* representa unidades de trabalho com uma duração (DUMAS et al., 2013) (o registro de um pedido, por exemplo). Um evento pode disparar a execução de uma série de atividades relacionadas, que geralmente ocorrem consecutivamente. São chamados *pontos de decisão* instantes no tempo em que as decisões tomadas nesses instantes afetam o modo de execução de um processo (DUMAS et al., 2013) (a decisão entre prosseguir ou não o andamento de um pedido, por exemplo).

Dentre diversos outros conceitos inerentes a BPM, faz-se necessário para o acompanhamento deste estudo a definição de alguns deles. Atividades de um processo são conectadas por *controles de fluxo*, que realizam a execução sequencial ou concorrente dessas atividades (DUMAS et al., 2013). Por exemplo, a equipe de segurança de um aeroporto pode, simultaneamente, checar a existência de objetos impróprios na mala de um passageiro e checar os metais que o mesmo passageiro carrega.

Um *subprocesso* é uma atividade normalmente utilizada em um modelo de processos para representar um processo coordenado por outro processo. Por exemplo, o processo de aquisição de materiais para atender um pedido ocorre nesse e em vários outros processos. Subprocesso pode ser confundido com *tarefa*. Enquanto *tarefa* trata-se de um passo atômico de trabalho em um processo (podendo ser manual ou automatizado) e contido em uma atividade específica, subprocessos podem ser definidos como um processo próprio (com início e fim) e reutilizado em outros processos (DUMAS et al., 2013).

2.3 Mineração de Processos de Negócio

Mineração de processos de negócio consiste em um conjunto de técnicas que combinam aspectos de modelagem e análise de processos de negócio com mineração de dados e *machine learning* (AILENEI et al., 2011). O objetivo da mineração de processos é extrair informações relacionadas a um processo utilizando o conhecimento contido em *logs* de eventos, obtidos a partir dos dados registrados por sistemas de informação (AALST, 2011).

Surgiu com a intenção de preencher a lacuna existente entre BI - *Business Inteli-*

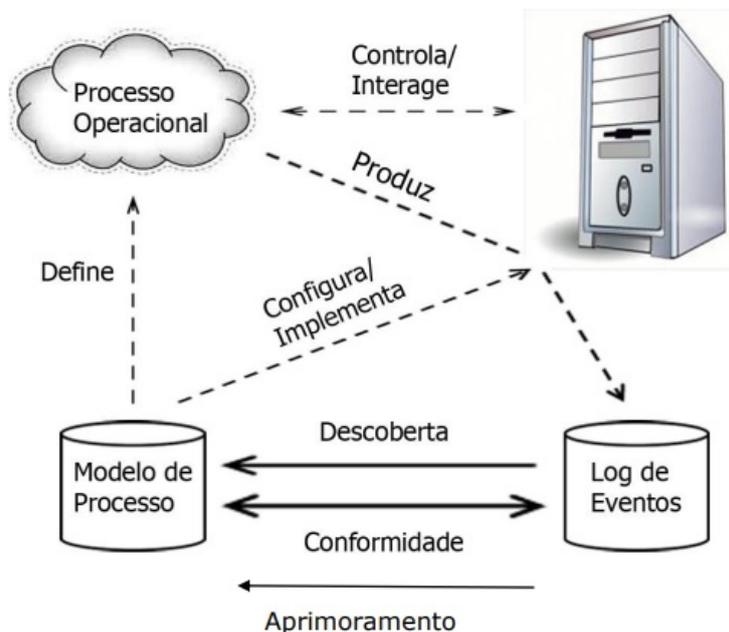
gence e BPM (AALST, 2011). BI refere-se a ferramentas e técnicas que utilizam dados de eventos objetivando auxiliar em tomadas de decisões, porém, as poucas ferramentas de BI que oferecem recursos maduros de mineração de dados o fazem com o foco apenas nos dados e em tomadas de decisão locais, e não nos processos de ponta a ponta (AALST, 2011). BPM, como visto na Seção 2, baseia-se em modelos de processos. Contudo, esses modelos podem apresentar uma desconexão dos dados reais do evento, focando-se mais em um modelo *idealizado* de realidade do que em fatos observados (AALST, 2011). Essa combinação entre modelos de processos e dados de eventos possibilitam novas formas de análise focada em processos.

A partir das informações obtidas, a mineração de processos permite analisar ocorrências do passado, o motivo da ocorrência e prováveis ocorrências do futuro. Também permite descobrir o motivo e o momento em que organizações ou pessoas desviam-se do objetivo final de um certo processo e possibilitar um melhor controle do processo e aprimorá-lo, objetivando aumentar seu desempenho (AALST, 2011).

A mineração de processos pode ser dividida em três fases (AALST, 2011): *descoberta*, *conformidade* e *aprimoramento*. A fase de *descoberta* analisa o modelo inicial do processo, que pode ser obtido manualmente, usando uma notação para modelagem de processos (Redes de Petri, BPMN, YAWL, EPC, etc. (AALST et al., 2011)) ou através da análise de um *log* de eventos, minerando todos os possíveis caminhos de resolução. A fase de *conformidade* busca medir o nível de alinhamento entre o modelo de descoberta do processo e as informações obtidas no *log* de eventos e verificar a existência de possíveis desvios entre eles. Também nessa etapa, o modelo pode ser estendido, incluindo, por exemplo, tempos e frequências. Assim, é possível detectar problemas, mediante uso de mineração de decisão (ROZINAT; AALST, 2006) e extrair informações de tempo de duração dos processos executados. Já a fase de *aprimoramento* possibilita utilizar informações adicionais contidas no *log* de eventos (AALST, 2011), de forma que, ao final do processo, se tenha um modelo integrado de processos que permite ao usuário visualizar múltiplas perspectivas e possibilidades (AALST, 2011). A Figura 2.2 mostra uma visão geral das fases da mineração de processos.

As informações contidas em um *log* de eventos são representadas da seguinte forma: um processo é dividido em *casos*, que representam instâncias de um processo (AALST, 2011). Um caso contém eventos, que estão relacionados apenas a um único caso, geralmente são ordenados e podem conter *atributos*, como atividade, *recurso* responsável por essa atividade (como uma pessoa ou *software*), *timestamp*, dentre outros

Figura 2.2: Visão geral das fases de mineração de processos



Fonte: (PESTANA, 2017).

possíveis (AALST, 2011). Normalmente, casos e eventos contém identificadores que os diferenciam ao longo do conteúdo do *log*. A Tabela 2.1 representa um exemplo de um *log* de eventos e seus dados.

Logs de eventos são a principal fonte de um projeto de mineração de processos, sendo obtidos através de diferentes fontes, tais como bases de dados, sistemas ERP - *Enterprise Resource Planning* (AALST; WEIJTERS; MARUSTER, 2004), sistemas CRM - *Customer Relationship Management* (DEVI, 2006), sistema de informação de um hospital (AALST et al., 2011), *logs* de transações bancárias, etc (CELIK UFUK E AKÇETIN, 2018). O conhecimento extraído desses *logs* permite descobrir o processo que ocorre na prática e seu modelo obtido após as fases de mineração para posterior análise, avaliação e melhoria contínua na qualidade dos processos tratados. Sendo assim, já existem no mercado diversas ferramentas de mineração de processos visando a extração de todas as informações possíveis durante e após a execução de um processo. Porém, pode não ser uma tarefa fácil para o usuário a escolha da ferramenta ideal para alcance de seus objetivos.

Tabela 2.1: Exemplo de informações encontradas em um *log* de eventos

| <i>ID do evento</i> | <i>Atividade</i> | <i>Ator da atividade</i> | <i>Timestamp</i> |
|---------------------|------------------|--------------------------|------------------|
| 10010 | Atividade A | João | 01.01.2019 10:05 |
| 10011 | Atividade A | Pedro | 01.01.2019 10:12 |
| 10012 | Atividade A | Maria | 01.01.2019 11:22 |
| 10013 | Atividade B | João | 02.01.2019 12:25 |
| 10014 | Atividade B | Maria | 02.01.2019 13:45 |
| 10015 | Atividade C | Pedro | 05.01.2019 14:05 |
| 10016 | Atividade B | Pedro | 06.01.2019 14:05 |
| 10017 | Atividade C | João | 06.01.2019 14:55 |
| 10018 | Atividade D | Maria | 07.01.2019 10:05 |
| 10019 | Atividade D | João | 07.01.2019 11:29 |

Fonte: O Autor.

2.4 Trabalhos Relacionados

Apesar da existência de diversos trabalhos, a área de mineração de processos pode ser considerada ainda emergente (AALST, 2018). Os trabalhos desenvolvidos por Wil Van der Aalst e seus colaboradores deram início a pesquisas mais detalhadas sobre essa área. Van der Aalst (AALST, 2011) detalhou as características da mineração de processos, dividindo-a em três fases: *descoberta*, *conformidade* e *aprimoramento*. Este trabalho deu origem a diversos outros, tendo como objetivo detalhar o funcionamento de cada uma das fases, analisando os algoritmos utilizados, o formato ideal de um *log* de eventos, dentre outras técnicas. Atualmente, o foco de pesquisa da área está voltado ao desenvolvimento de novas técnicas, novos algoritmos e novas ferramentas (AILENEI et al., 2011), tanto no âmbito acadêmico quanto industrial, bem como estudar as ferramentas existentes, entender como processam de cada uma das fases e compará-las, que é exatamente o objetivo do presente trabalho.

Aruna Devi (DEVI, 2006) realizou um estudo comparativo e analítico sobre as ferramentas ProM, Disco e Celonis, trazendo uma breve explicação sobre mineração de processos, seus benefícios e desvantagens e comparando as ferramentas referidas sob alguns aspectos de: formatos de importação e de exportação dos dados contidos nos *logs* de eventos, existência (ou não) de filtragem de dados, checagem de conformidade entre o *log* e o modelo de processos, dentre outros.

As mesmas ferramentas são analisadas e comparadas no trabalho de Musie Kebede (KEBEDE, 2015) usando uma tabela comparativa semelhante a de Devi (2006). Adicionalmente, Kebede utiliza casos de uso para mapear problemas e identificar as técnicas de mineração de processos necessárias para suas resoluções. Neste trabalho, o autor

procura responder três questões de pesquisa: primeiro, quais são os principais problemas enfrentados por usuários de ferramentas de mineração de processos; segundo, quais as operações contidas na mineração de processos podem ser utilizadas para resolver um certo problema; e, terceiro, qual ferramenta de mineração de processos pode realizar essas operações.

Irina Ailenei (AILENEI et al., 2011) realizou uma análise comparativa escolhendo a ferramenta *open-source* ProM e as comerciais ARIS, Fourspark, Futura Reflect e QPR Process Analyzer, contando com o apoio de diversos casos de uso contidos dentro de cada fase da mineração de processos. Esses casos foram validados com o auxílio de entrevistas com usuários e pessoas com experiência no domínio. Também foi realizada uma pesquisa com a participação de analistas e gerentes de processos, pesquisadores e consultores, com o objetivo de descobrir qual aspecto abordado nos casos de uso é mais importante e que mais corresponde às necessidades dos usuários que têm contato com a área.

Turner (TURNER et al., 2012) realiza uma análise comparativa entre ferramentas comerciais (ARIS, Fluxicon, Fujitsu Automated Process Discovery Service (APDS), Futura Reflect, Iontas Focus Suite, Comprehend e BPM One) e entre técnicas de mineração de processos de negócio desenvolvidas tanto acadêmica quanto comercialmente.

Verstraete (VERSTRAETE, 2014) constatou, na prática, a necessidade de mais informações sobre *softwares* de mineração de processos, o que o motivou a desenvolver seu trabalho. Estabeleceu duas questões de pesquisa: primeiro, quais os principais critérios a serem considerados para escolher um *software* de mineração de processos; e, segundo, quais ferramentas de mineração de processos são utilizadas e por quem. Essas questões foram respondidas com o auxílio de uma pesquisa *online* realizada com todos aqueles que estivessem envolvidos com mineração de processos como um usuário, não restringindo a pesquisa a usuários em ambiente industrial ou a usuários em ambiente acadêmico. A pesquisa questionou aos usuários quais fatores consideravam para escolher uma ferramenta de mineração (usabilidade, visualizações, preço, funcionalidades, dentre outras), quão importante avaliavam certos conceitos (de automatização, animação, integração e criação de redes sociais), o principal objetivo buscado pelo usuário (otimização de processos, análise de performance, monitoramento, etc.) e quais ferramentas eram utilizadas por esses usuários, sendo ProM e Disco consideradas as mais populares.

O artigo de Agarwal (AGARWAL NEHA E SINGH, 2014) apresenta, além de uma descrição sobre mineração de processos e suas fases, uma análise comparativa sobre as ferramentas ProM, Emit, Fluxicon e ARIS.

O estudo de Celik (CELIK UFUK E AKÇETIN, 2018) faz um comparativo entre as ferramentas ProM, Disco, Celonis e My-Invenio, uma ferramenta comercial que se diferencia das demais por ser utilizada via plataforma *web*, permitindo que usuários possam acessá-la por qualquer dispositivo (*desktop*, *smartphone* ou *tablet*). Contudo, a utilização da ferramenta requer conexão à Internet e limita a capacidade do *log* de eventos importado a 30 MB, o que pode limitar a experiência do usuário, conforme sua necessidade e em relação às demais ferramentas disponíveis comercialmente.

Já Claes (CLAES JAN E BRU, 2018) desenvolveu um questionário destinado a usuários finais de ferramentas de mineração. O objetivo foi proporcionar aos usuários finais a possibilidade de responder sobre suas áreas de atuação (analistas de processos, pesquisadores, estudantes, etc.), sobre quais ferramentas utilizam (Disco, ProM e Celonis foram apresentadas como as mais populares), o que consideram mais importante em uma ferramenta de mineração, suas preferências de ambiente (plataforma *web* ou aplicativo instalado no computador do usuário - *standalone program*) e os benefícios do ambiente escolhido, considerando seus objetivos finais.

2.5 Comparativo entre Trabalhos Relacionados

A Tabela 2.2 compara diretamente os trabalhos relacionados trazidos na Seção 2.4, elencando suas principais características. A entrada [9] da tabela é reservada às características deste estudo, objetivando auxiliar na comparação com os demais trabalhos. Este estudo comparativo descreve o ciclo de vida de BPM, assim como o estudo de Verstraete (2014) [5], que descreve também as características contidas nas fases da mineração de processos.

As fases da mineração de processos também são descritas no presente trabalho e nas abordagens de Devi (2006) [1], Ailenei et al. (2011) [3], Agarwal Neha e Singh (2014) [6] e Celik Ufuk e Akçetin (2018) [7]. Esses trabalhos também trazem um comparativo entre as ferramentas de mineração existentes no mercado, tal como os de Kebede (2015) [2] e Turner et al. (2012) [4]. As características utilizadas nesses trabalhos para comparação entre as ferramentas também são utilizadas para comparar as ferramentas estudadas no presente trabalho. O questionário realizado por Claes Jan e Bru (2018) [8] obteve as ferramentas mais populares entre os usuários da área de mineração de processos e seus objetivos ao utilizar essas ferramentas, informações que são descritas no presente trabalho.

2.6 Resumo do Capítulo

Neste capítulo, são apresentados fundamentos sobre BPM e seu ciclo de vida, composto pelas fases de identificação, descoberta, análise, redesenho, implementação, monitoramento e controle de processo. Alguns conceitos básicos sobre BPM também são apresentados de forma a contribuir no acompanhamento dos capítulos do presente trabalho. As características da mineração de processos, assim como seus objetivos e suas fases (descoberta, conformidade e aprimoramento) são detalhadas. Finalmente, os trabalhos relacionados ao presente trabalho são analisados e um paralelo entre suas principais características e as deste trabalho são estabelecidos, mediante uso de uma tabela comparativa entre eles.

Tabela 2.2: Comparativo entre Trabalhos Relacionados

| <i>Id.</i> | <i>Trabalho</i> | <i>Principais características</i> |
|------------|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [1] | Devi (2006) | <ul style="list-style-type: none"> • Mineração de processos • Comparativo entre ferramentas |
| [2] | Kebede (2015) | <ul style="list-style-type: none"> • Casos de uso para mapear problemas e identificar técnicas para soluções • Questões de pesquisa • Comparativo entre ferramentas |
| [3] | Ailenei et al. (2011) | <ul style="list-style-type: none"> • Questões de pesquisa • Mineração de processos • Comparativo entre ferramentas • Casos de uso elencados entre fases da mineração e validados mediante pesquisa com usuários |
| [4] | Turner et al. (2012) | <ul style="list-style-type: none"> • Comparativo entre técnicas de mineração • Comparativo entre ferramentas |
| [5] | Verstraete (2014) | <ul style="list-style-type: none"> • Questões de pesquisa • Ciclo de vida de BPM • Mineração de processos • Questionário com usuários sobre aspectos e ferramentas de mineração |
| [6] | Agarwal Neha e Singh (2014) | <ul style="list-style-type: none"> • Mineração de processos • Comparativo entre ferramentas |
| [7] | Celik Ufuk e Akçetin (2018) | <ul style="list-style-type: none"> • Mineração de processos • Comparativo entre ferramentas |
| [8] | Claes Jan e Bru (2018) | <ul style="list-style-type: none"> • Questionário com usuários sobre aspectos e ferramentas de mineração |
| [9] | Este estudo | <ul style="list-style-type: none"> • Ciclo de vida de BPM • Mineração de processos • Comparativo entre ferramentas • Estudo de caso |

Fonte: O Autor.

3 ANÁLISE DAS FERRAMENTAS

Neste capítulo, são comparadas sete ferramentas de mineração de processos. As ferramentas foram escolhidas com base em sua popularidade e uso tanto acadêmico quanto industrial. A análise comparativa é realizada classificando as ferramentas em duas categorias: *open-source* e comerciais (proprietárias). As ferramentas *open-source* analisadas são ProM, Apromore e BupaR e as de uso comercial analisadas são Disco, Celonis, ARIS PPM e QPR Process Analyzer. Elas são comparadas dentro de suas categorias com o apoio de uma tabela comparativa abordando diversos critérios, conforme a categoria em que estão inseridas, destacando suas diferenças e semelhanças. A comparação entre as ferramentas é realizada com o uso das características comparadas nos trabalhos relacionados estudados no Capítulo 2.

3.1 Ferramentas Open-Source

Esta Seção traz o comparativo entre as ferramentas *open-source* ProM, Apromore e BupaR. As funcionalidades contidas em cada uma delas são detalhadas e, após, comparadas mediante uso de uma tabela comparativa, que foca nas seguintes características: formatos de importação de dados e exportação de modelos, versão corrente da ferramenta, plataformas de execução e se realizam filtragem de dados, descoberta de processos, checagem de conformidade, mineração de rede social, mineração de regras de decisão, visualização de processos e análise de performance.

3.1.1 ProM

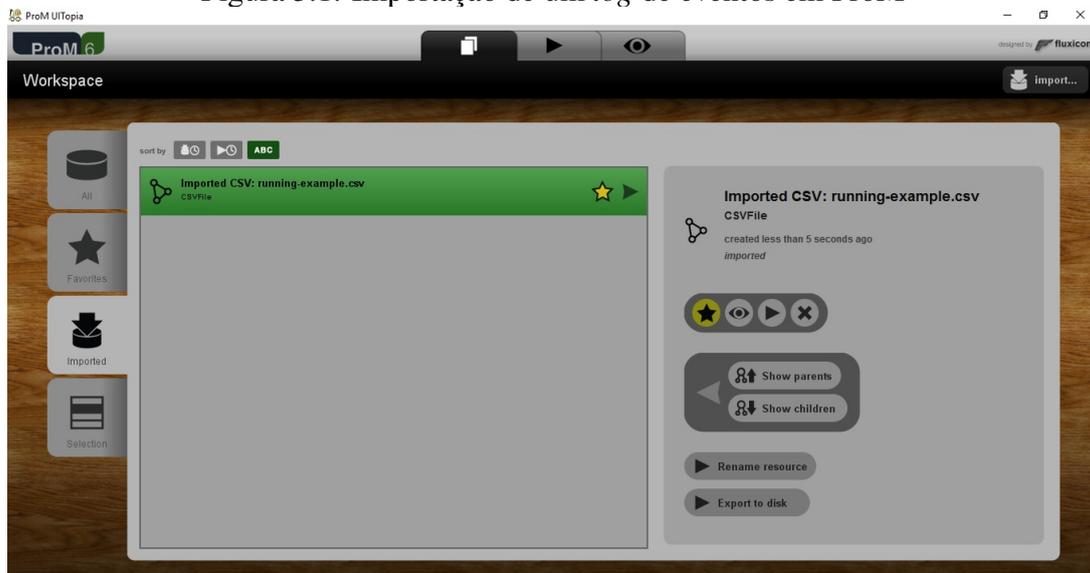
ProM (abreviatura de *Process Mining*) é uma ferramenta *open-source* desenvolvida na *Eindhoven Technical University*, na Holanda (AALST, 2011). Foi desenvolvida pelo grupo de pesquisa liderado por Wil Van der Aalst (AILENEI et al., 2011) e recebe a colaboração de diversos pesquisadores na área de mineração de processos de negócio. É uma ferramenta expansível, de fácil uso que suporta uma ampla variedade de técnicas de mineração de processos com o auxílio de *plug-ins* (EINDHOVEN UNIVERSITY OF TECHNOLOGY, 2019a). É desenvolvida em Java e independente de plataforma, podendo ser baixado gratuitamente em <http://www.promtools.org>. Desenvolvedores e pesquisado-

res podem colaborar com a expansão dessa ferramenta, implementando novos *plug-ins* ou auxiliando na manutenção dos já existentes (EINDHOVEN UNIVERSITY OF TECHNOLOGY, 2019b).

Algumas das principais funcionalidades que ProM disponibiliza são: descoberta do fluxo de controle de um processo, análise de rede social, análise de recursos e de performance contidas em um processo, descoberta de eventos baseada em regras de decisão e checagem de conformidade usando uma série de algoritmos.

ProM é capaz de importar *logs* de eventos compatíveis com os formatos MXML - *Mining eXtensible Markup Language* e XES - *eXtensible Event Stream* (KEBEDE, 2015) e permite exibir o modelo de processo usando diferentes notações, como BPMN, Redes de Petri, sistema de transição, modelo heurístico, modelo minerado indutivo e modelo *fuzzy* (DEVI, 2006). A ferramenta disponibiliza vários formatos de exportação dos dados, tais como CSV - *Comma Separated Values*, PNG - *Portable Network Graphics* (que permite uma representação visual dos resultados), MXML e XES (AILENEI et al., 2011). A Figura 3.1 refere-se a tela inicial de ProM e traz exemplo de importação de um *log* de eventos em formato CSV. A partir dessa importação, os dados contidos no *log* podem ser visualizados ou submetidos à análise dos vários *plug-ins* disponíveis na ferramenta, conforme Figura 3.2.

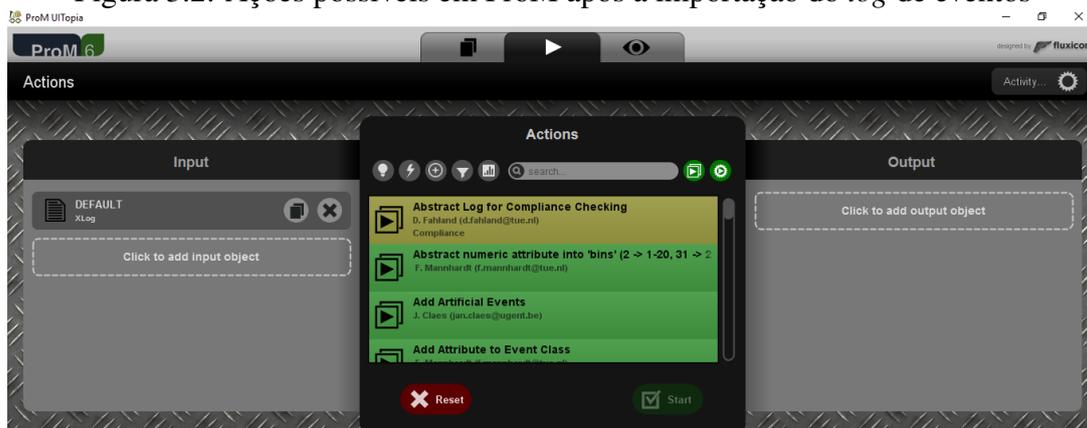
Figura 3.1: Importação de um *log* de eventos em ProM



Fonte: O Autor.

Após a importação de um *log* de eventos para o ambiente do ProM e a execução das ações, a guia *Summary* traz informações obtidas a partir do processamento do *log*, como eventos de início e de fim, eventos por caso e o número total de eventos, como é

Figura 3.2: Ações possíveis em ProM após a importação do *log* de eventos



Fonte: O Autor.

possível visualizar na Figura 3.3. A guia *Inspector* permite inspecionar cada instância do processo em questão. Já a guia *Dashboard* traz informações como número de processos, número de casos e número de eventos, bem como número de classes de eventos e de tipos de eventos.

Para filtragem dos dados, o botão *Action* disponibiliza diversas opções como, por exemplo, filtrar *logs* por eventos, por atributos, por valores de atributos de um evento, por valores de atributos de um traço, bem como extrair traços organizados por índice e ordenar eventos em traços baseados em seus *timestamps*, dentre outros filtros possíveis. Também existe a opção *Filter log using simple heuristics*, que permite ao usuário selecionar eventos que queira manter ou omitir na exibição final. A Figura 3.3 exibe o resultado de um *log* filtrado por heurística simples.

Figura 3.3: Filtragem de *Log* de eventos por heurística simples



Fonte: O Autor.

Como visto nesta subseção, o modelo de descoberta em ProM pode ser representado em diversos formatos e é obtido com o auxílio de inúmeros *plug-ins*, que recebem como entrada um *log* filtrado por determinado atributo e devolvem um modelo no formato desejado, permitindo também a conversão entre formatos (Redes de Petri para BPMN, modelo minerado indutivo para modelo *fuzzy*, etc). Dúvidas sobre quais são os caminhos mais frequentes em um processo, padrões de *loops* durante a execução e distribuição de caminhos podem ser sanadas utilizando o visualizador de abstração de padrões.

Para realizar a checagem de conformidade em ProM, o *plug-in* responsável por essa funcionalidade requer o modelo de processo no formato de Redes de Petri (arquivo .tpn ou .pnml) e seu resultado é apresentado no formato EPC ou YAWL (KEBEDE, 2015). O *log* é pré-processado e as instâncias com propriedades similares são agrupadas com o objetivo de otimizar o tempo de execução das métricas de conformidade. As métricas de conformidade realizam os seguintes métodos de análise (DONGEN et al., 2005): *Análise espacial de estado*, com o apoio de um grafo de cobertura originada do modelo de processo que é percorrido com o objetivo de detectar *loops*, relação entre tarefas no modelo de perspectiva e tarefas duplicadas; *Análise estrutural do modelo de processo*, com o objetivo de detectar tarefas invisíveis e redundantes e; *Análise de repetição de log*, onde o *log* é replicado de maneira não-bloqueante e em perspectiva de *log*, com o objetivo tanto de medir a quantidade de *tokens* consumidos, produzidos ou ausentes quanto de medir o número médio de transações ativas.

Conforme Rozinat (ROZINAT, 2010), existem três dimensões conhecidas de conformidade: *fitness*, precisão/generalização e estrutura. A dimensão mais dominante é a de *fitness* (ROZINAT, 2010), que consiste em medir quanto do comportamento observado em um *log* pode ser visto (o quanto ele se encaixa - “*fit*”) em um modelo de processos. É justamente essa dimensão que é utilizada pelo ProM para realizar a checagem de conformidade.

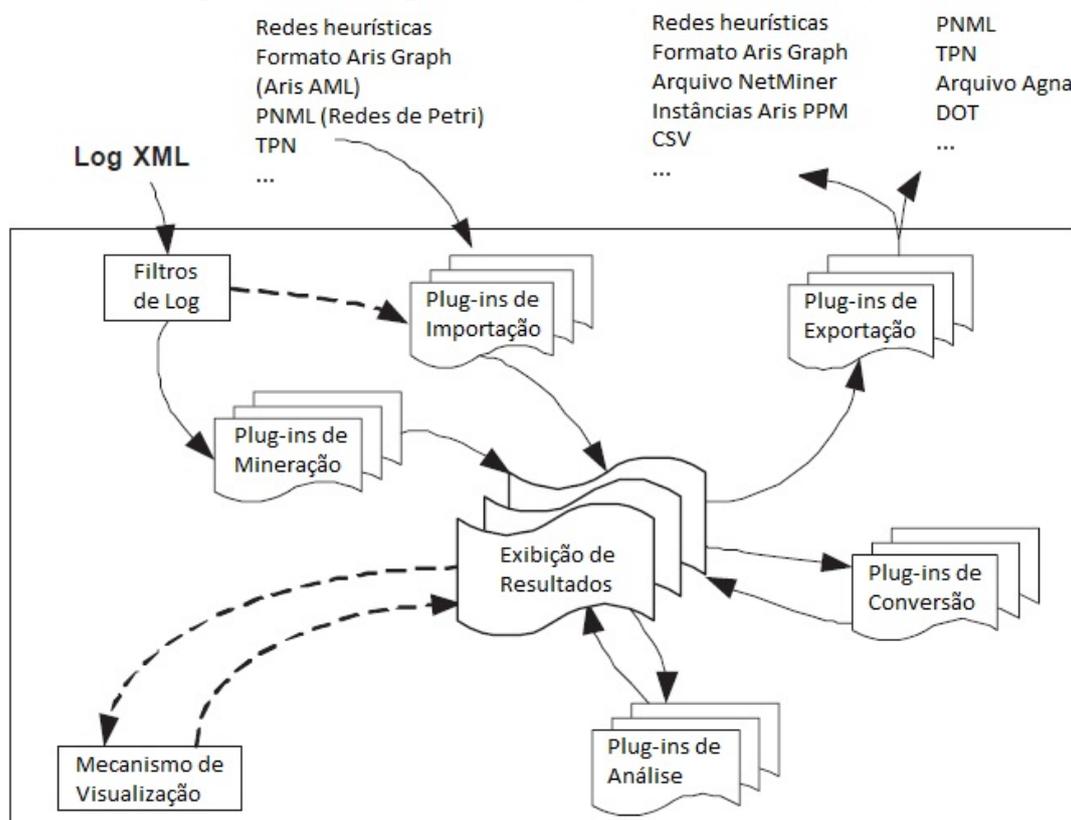
Os *plug-ins* utilizados em ProM são separados em cinco tipos, conforme Figura 3.4 (DONGEN et al., 2005):

- *Plug-ins de Mineração*, responsáveis por implementar algoritmos de mineração capazes de construir, por exemplo, uma Rede de Petri baseada em um *log* de eventos;
- *Plug-ins de Exportação*, responsáveis por implementar funcionalidades do tipo “Salvar como” para alguns objetos como, por exemplo, “Salvar como EPCs”, “Salvar como Rede de Petri”, dentre outros;
- *Plug-ins de Importação*, os quais implementam funcionalidades para exportação de

objetos, tais como instâncias em EPCs para CSV;

- *Plug-ins de Análise*, que realizam uma análise das propriedades do resultado de uma mineração. Analisando uma Rede de Petri, por exemplo, há um *plug-in* que constrói invariantes de lugar e de transição, bem como um grafo de cobertura. Para cada formato de entrada, existem um ou mais *plug-ins* que extraem diversas informações possíveis (*Log Summary*, para informações gerais acerca do *log* importado; *Analyze Social Network*, para análise de rede social, dentre outras (AALST et al., 2009));
- *Plug-ins de Conversão*, que realizam conversões entre diferentes formatos de dados como, por exemplo, de EPCs para Redes de Petri.

Figura 3.4: Visão geral dos *Plug-ins* de ProM e seus tipos

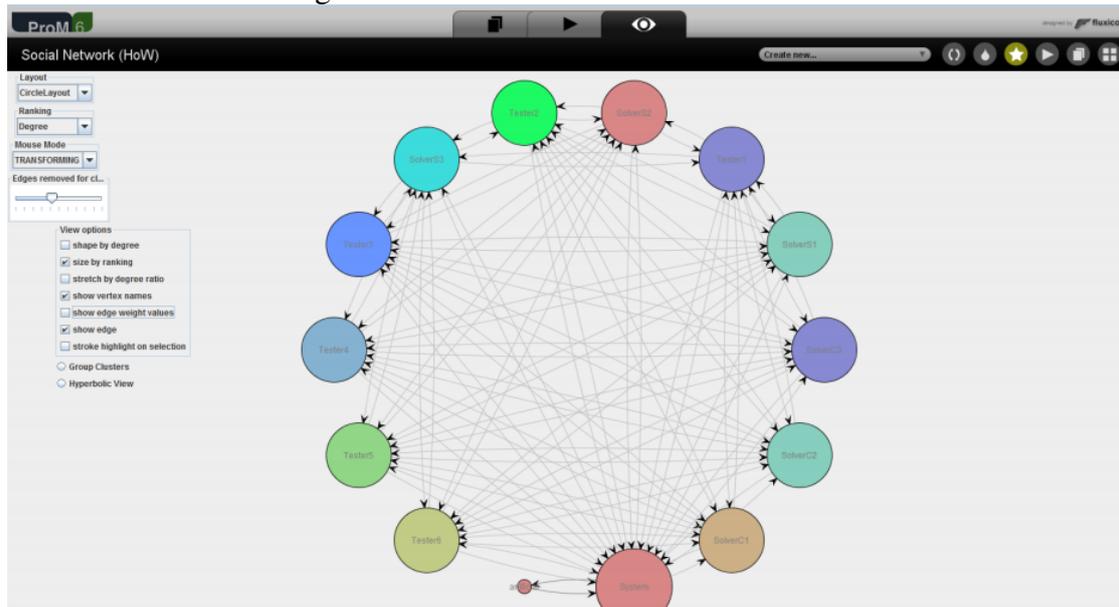


Fonte: Adaptado de (DONGEN et al., 2005)

A análise de rede social possibilita a visualização de um processo mediante uma perspectiva organizacional, permitindo opções como mineração para uma rede social de entrega de trabalho (*handover-of-work*), como apresenta a Figura 3.5, que permite visualizar a entrega de trabalho entre dois indivíduos i e j e duas atividades subsequentes, onde a primeira é completada por i e a segunda por j (EINDHOVEN UNIVERSITY OF TECHNOLOGY, 2019a). ProM também permite análise de rede social mediante uso de outras

métricas (KEBEDE, 2015): *redesignação*, que identifica relações hierárquicas entre participantes; *tarefa similar*, que identifica a relação entre indivíduos que realizam as mesmas atividades; *subcontratação*, que identifica o número de vezes em que j executou uma atividade entre duas atividades executadas por i ; ou *trabalho em conjunto*, que identifica a relação entre indivíduos trabalhando em um mesmo caso (EINDHOVEN UNIVERSITY OF TECHNOLOGY, 2019a).

Figura 3.5: Análise de rede social em ProM



Fonte: (KEBEDE, 2015)

A mineração por regras de decisão permite encontrar os pontos de decisão de um processo e tratar atividades escondidas ou duplicadas (ROZINAT; AALST, 2006). Em ProM, a visualização de processos em formato de animação permite ao usuário detectar gargalos e impedimentos à execução normal desse processo. ProM também possui um *plug-in* que permite ao usuário analisar a performance do processo após a interpretação do *log* de eventos (KEBEDE, 2015).

3.1.2 Apromore

Apromore é uma abreviação de *Advanced PROcess MOdel REpository* (ROSA et al., 2011). É uma ferramenta *open-source* criada na *University of Melbourne*, na Austrália (THE UNIVERSITY OF MELBOURNE, 2019). É resultado de mais de dez anos de pesquisas e desenvolvimento, dos quais várias outras universidades e colaboradores individuais fizeram parte, um processo que contou com patrocínios tanto governamentais

quanto de setores privados da sociedade (THE UNIVERSITY OF MELBOURNE, 2019).

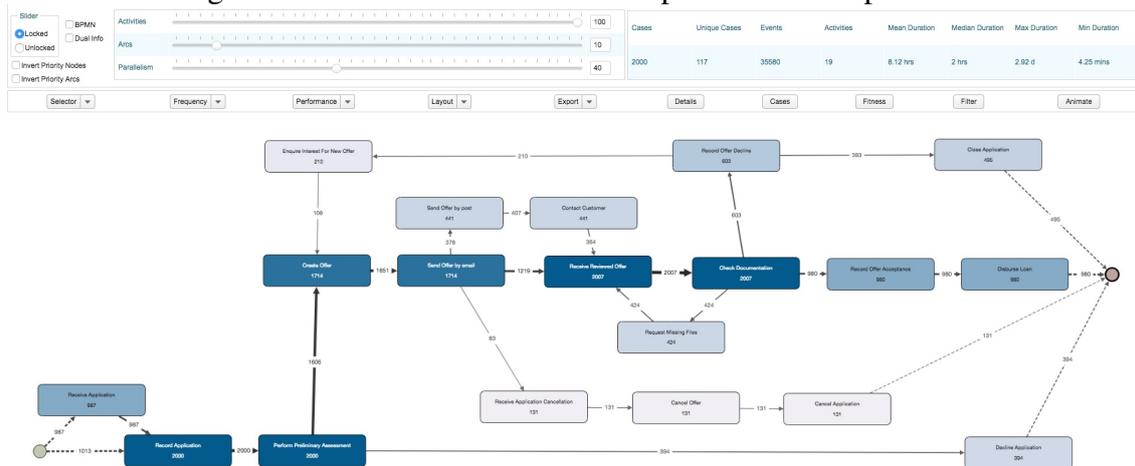
Apromore tem usuários empresas, tais como Intesa Sanpaolo Banking Group, o maior banco da Itália, Nool e P4I – *Partners4Innovation* (THE UNIVERSITY OF MELBOURNE, 2019). O projeto é apoiado pelas universidades de Melbourne (Austrália), Tartu (Estônia) e Camerino (Itália). A ferramenta pode ser acessada na nuvem em *releases* públicos, abertos desenvolvidos por essas universidades que permitem a qualquer usuário criar ou deletar modelos ou *logs*. É recomendado ao usuário acessar o *nodo* mais próximo de onde está localizado. Outra forma de acessar Apromore é realizando o *download* para execução na máquina do usuário, com o auxílio da ferramenta de virtualização do tipo contêiner *Docker* em máquinas com sistema operacional Windows, Mac ou Linux. Atualmente a versão disponível para *download* é a versão 7.0.1.

A ferramenta permite a importação de *logs* de eventos no formato XES ou MXML e trabalha com modelos de diversos formatos, como BPMN, XPD - *XML Process Definition Language*, Redes de Petri (PNML - *Petri Net Markup Language*), AML - *ARIS Markup Language* e YAWL (CONFORTI et al., 2015). Também permite a conversão entre modelos em suas formas canônicas (ROSA et al., 2011). O formato canônico traz uma representação comum e não-ambígua de um processo obtida independente de sua notação e/ou nível de abstração, representando apenas as características estruturais desse processo, facilitando a comparação entre diferentes modelos (ROSA et al., 2011).

Apromore foi a primeira ferramenta de mineração de processos a realizar a descoberta de processos de forma automática com o uso de modelos em BPMN e a checagem de conformidade entre *logs* de eventos e modelos BPMN (THE UNIVERSITY OF MELBOURNE, 2019). Oferece uma série de funcionalidades além da manipulação de grandes modelos de processos, como descoberta automática de processos, mineração de processos, checagem de conformidade, análise de variantes e monitoramento de processos. Essas funcionalidades são organizadas de acordo com o ciclo de vida de BPM (VERENICH et al., 2018).

A Figura 3.6 exibe um exemplo da descoberta de processos em Apromore, que permite o ajuste da complexidade do modelo de descoberta, aumentando ou reduzindo a frequência dos nodos ou arcos exibidos na visualização, de forma que o usuário possa visualizar tanto os caminhos mais frequentes que um processo realiza quanto apenas os menos frequentes, selecionando as caixas *Invert Priority Nodes* e/ou *Invert Priority Arcs*. O modelo exibe estatísticas como, por exemplo, o total, mínimo ou máximo número de vezes que uma atividade foi executada. Essas informações podem ser visualizadas sob um

Figura 3.6: Modelo de descoberta de processos em Apromore

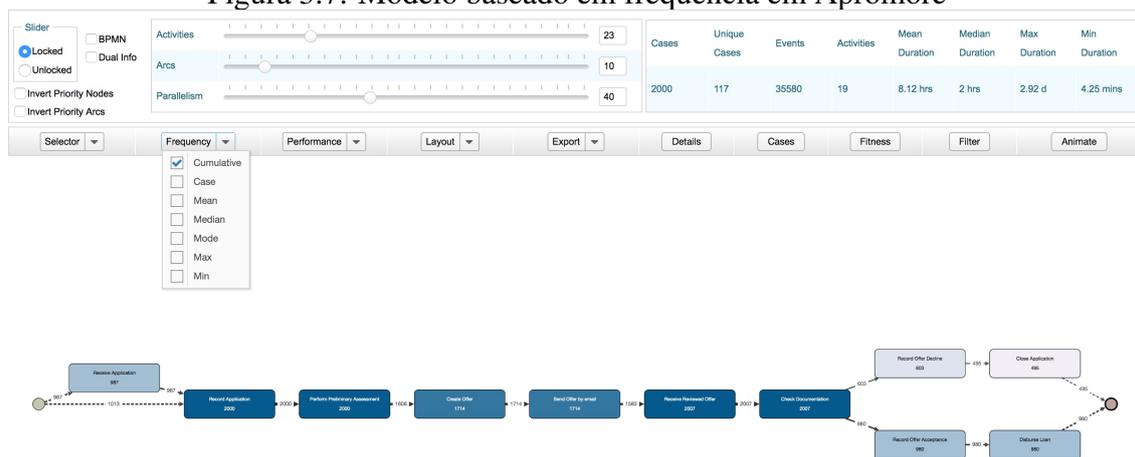


Fonte: apromore.org

esquema de cores em relação às atividades do processo (quanto mais escura a cor azul, mais vezes a atividade ocorre) e em relação à sequência das atividades (quanto mais fino é o arco que liga as atividades, menor é a frequência que ocorre a sequência).

A ferramenta também permite exibir um modelo baseado na duração de uma atividade ou no tempo entre a execução de cada atividade do processo (total, médio, mediano, mínimo ou máximo), visualizadas sob o mesmo esquema de cores e linhas da exibição exemplificada na Figura 3.7 (THE UNIVERSITY OF MELBOURNE, 2019).

Figura 3.7: Modelo baseado em frequência em Apromore



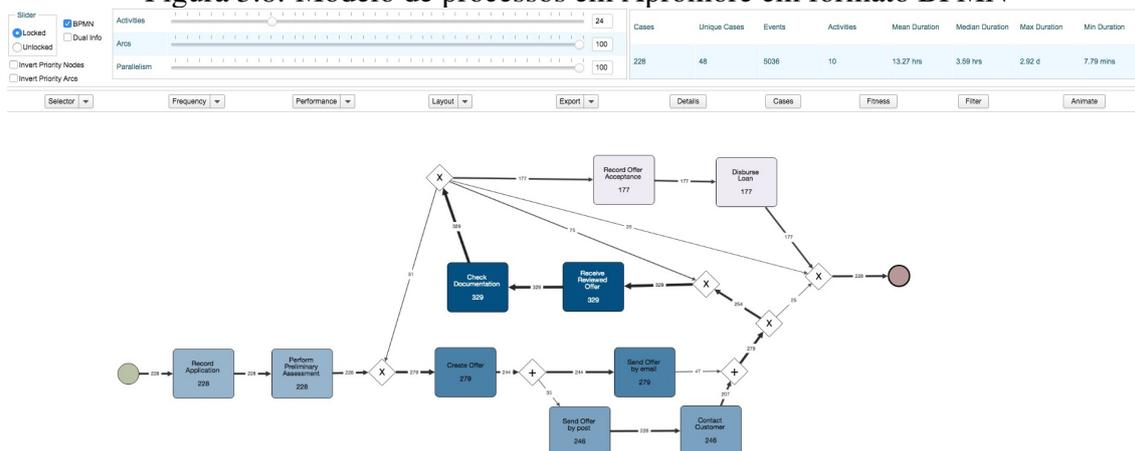
Fonte: apromore.org

Os botões de opção *Locked* ou *Unlocked* permitem (ou não) a alteração das barras deslizantes que representam três métricas de exibição de um modelo: atividades, arcos e paralelismo. O usuário, ao selecionar o botão *Selector*, decide qual dos atributos contidos no *log* será o foco do modelo de processo.

Um *log* de eventos pode ser filtrado por diversos critérios. Por exemplo, é possível selecionar eventos que contenham uma atividade específica, que utilizem um certo recurso, que contenham um valor específico para um certo atributo ou que sejam executados dentro de um período de tempo informado. Também é possível filtrar por dependência direta entre atividades, o que pode ser representado pela origem e destino do arco que liga essas atividades em um modelo de descoberta (THE UNIVERSITY OF MELBOURNE, 2019). Para medir a qualidade de um modelo de descoberta de processos em relação a um *log* de eventos, deve-se acessar o botão *Fitness*.

Essas mesmas funcionalidades são adicionadas a uma análise avançada de relações comportamentais entre as atividades com o auxílio da opção *Discover -> Discover BPMN model (advanced)*, bem como a possibilidade de analisar subprocessos existentes. É possível visualizar o resultado na Figura 3.8.

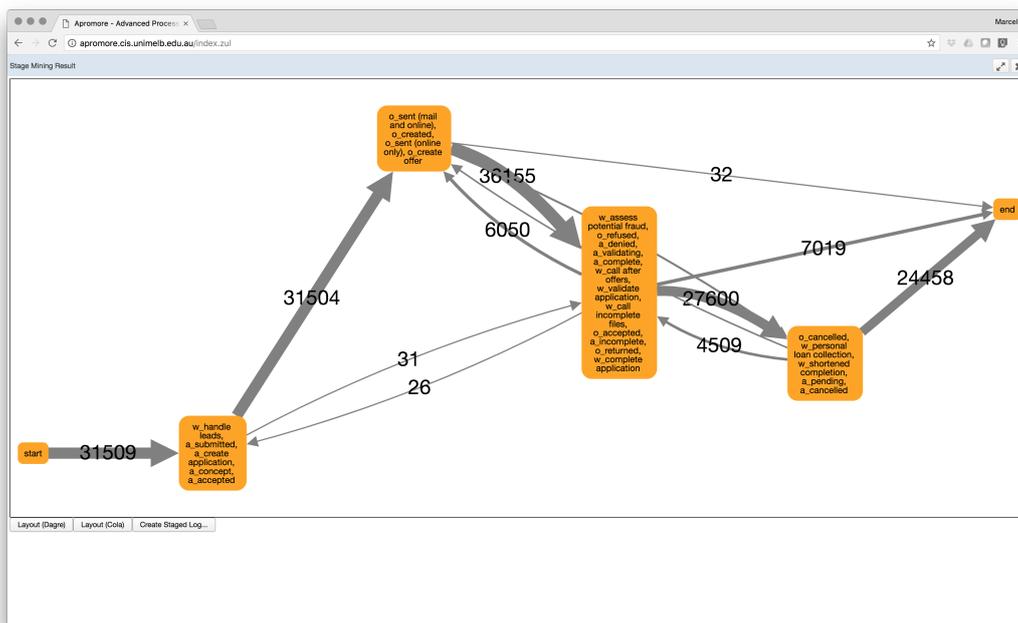
Figura 3.8: Modelo de processos em Apromore em formato BPMN



Fonte: apromore.org

Ainda em relação à descoberta de processos, Apromore possibilita a extração de regras de negócio em CTL - *Computation Tree Logic* (GROEFSEMA; BEEST; ARMAS-CERVANTES, 2017), descrevendo restrições que identificam o comportamento dos modelos de processos usados como entrada. Também disponibiliza a funcionalidade de agrupar atividades por estágios de processamento. Estágios são fragmentos de eventos que contém atividades com alto número de dependências diretas entre si e um baixo número de dependências entre outros possíveis estágios. Dessa forma, traz uma visão mais macro do processo que desejamos descobrir, o que é visto na Figura 3.9. A análise de rede social em Apromore pode ser feita selecionando o guia *Selector* e depois a opção *org:group*, exibindo a entrega de trabalho em um grupo e mapeando o papel organizacional de cada ator envolvido no processo (THE UNIVERSITY OF MELBOURNE, 2019).

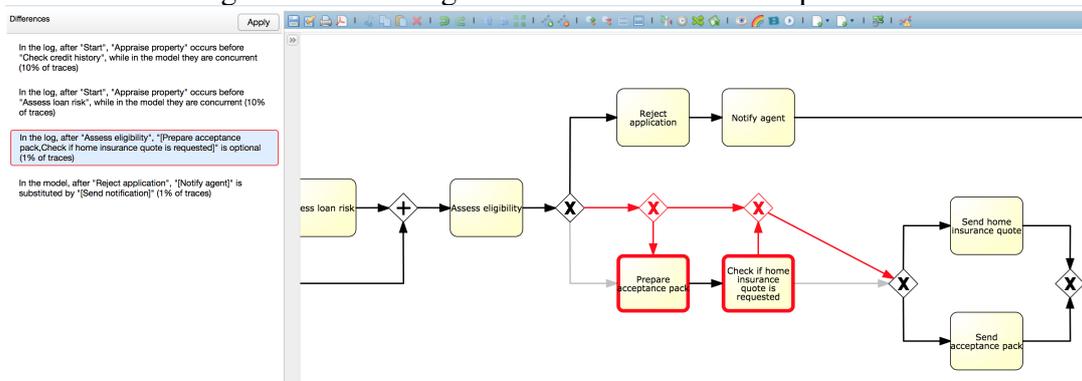
Figura 3.9: Modelo de processos agrupado em estágios em Apromore



Fonte: apromore.org

A checagem de conformidade nessa ferramenta permite identificar desvios entre o *log* de eventos e o modelo de processos comparando o modelo em BPMN com o *log* carregado no repositório da ferramenta (VERENICH et al., 2018). Os desvios são detectados e exibidos em linguagem natural e, ao selecionar um certo desvio, é exibido ao lado a representação visual do modelo de processo e indicado no modelo onde o desvio ocorre. Uma extensão dessa funcionalidade permite a automática correção desses desvios, devolvendo ao usuário o modelo corrigido, como visto na Figura 3.10.

Figura 3.10: Checagem de conformidade em Apromore



Fonte: apromore.org

Funcionalidades adicionais de Apromore reforçam o diferencial dessa ferramenta, que é a manipulação e integração com BPM. Algumas delas dizem respeito à criação e

edição de modelos de processo em formato BPMN, disponível em duas versões: *bpmn.io* e *Signavio*, que permitem salvar o modelo em formato BPMN ou PDF - *Portable Document Format* (THE UNIVERSITY OF MELBOURNE, 2019). Outra funcionalidade permite mostrar animações que simulam a execução de um processo em seu formato modelado, inclusive para modelos BPMN. Ferramentas externas, como BEBoP (FORNARI et al., 2017) e BeProVe, são integradas à Apromore para verificação de diretrizes de modelagem e corretude sintática, respectivamente.

Apromore possibilita analisar variantes entre duas versões de *logs* de eventos ou duas versões de modelos de processo. No que se refere ao monitoramento de processos, o usuário pode gerar modelos preditivos (VERENICH et al., 2018), que são capazes de prever propriedades genéricas de processo como, por exemplo, tempo restante até a conclusão de um caso ou a próxima atividade mais provável a ser executada. Acessando a opção *Predictively Monitor Log*, um *log* é selecionado, adicionado a um novo monitor preditivo e, selecionando a opção *Stream log to dataflow*, são criados fluxos de previsões, visualizados em um painel que exibe uma lista com casos finalizados ou em andamento, conforme visto na Figura 3.11 (THE UNIVERSITY OF MELBOURNE, 2019). É possível exportar os dados do painel em formato CSV para uso futuro.

Figura 3.11: Monitor preditivo em Apromore

| Running cases | | Completed cases | | Completed events | | Events per completed case | | Average case duration | | | | |
|---------------|----------------|-------------------|-------------------|--------------------|-------------------|---------------------------|---------------|-----------------------|------------------|-------------|-----------------------|----------------------|
| 188 | | 0 | | 499 | | - | | - | | | | |
| All events | Outcomes | Case duration | Remaining time | Case length | | | | | | | | |
| Case | Events elapsed | Start time | Latest event time | Target supply date | Supplier Location | Delivery Type | Sector | Line Total Cost | Delay Rank | Late Supply | Next Activity | Predicted Completion |
| 191315 | 4 | 2017-Aug-16 20:24 | 2017-Aug-22 19:54 | 2017-Nov-16 | International | Sea | Manufacturing | \$2,048.00 | 97% Just in time | 4% | 91% Supply Date Reco | 2017-Nov-09 21:15 |
| 191472 | 2 | 2017-Aug-20 23:53 | 2017-Aug-22 18:23 | 2017-Oct-28 | International | Sea | Manufacturing | \$7,183.00 | 80% Just in time | 34% | 72% Order Confirmed | 2017-Oct-26 07:25 |
| 182222 | 23 | 2017-Feb-10 01:27 | 2017-Aug-22 00:51 | 2017-Feb-12 | International | Courier | Hi Tech | \$73,868.00 | 91% Severe | 100% | 79% Goods Supplied | 2017-Aug-27 18:05 |
| 190823 | 4 | 2017-Aug-09 23:36 | 2017-Aug-21 19:53 | 2017-Sep-19 | International | Sea | Manufacturing | \$5,297.00 | 81% Just in time | 29% | 49% Delivered to Ship | 2017-Sep-24 01:02 |
| 190822 | 5 | 2017-Aug-09 23:36 | 2017-Aug-21 19:53 | 2017-Sep-20 | International | Sea | Manufacturing | \$3,559.00 | 93% Just in time | 12% | 47% Delivered to Ship | 2017-Sep-28 15:16 |
| 190822 | 4 | 2017-Aug-09 23:36 | 2017-Aug-21 19:53 | 2017-Sep-20 | International | Sea | Manufacturing | \$3,559.00 | 88% Just in time | 14% | 50% Delivered to Ship | 2017-Sep-24 13:34 |
| 190873 | 2 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-09 | International | Sea | Hi Tech | \$883.00 | 88% Just in time | 15% | 75% Supply Date Reco | 2017-Dec-06 18:39 |
| 190873 | 1 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-09 | International | Sea | Hi Tech | \$883.00 | 89% Just in time | 18% | 77% Order Confirmed | 2017-Dec-06 09:08 |
| 190872 | 2 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-03 | International | Sea | Hi Tech | \$1,086.00 | 80% Just in time | 27% | 57% Order Confirmed | 2017-Nov-10 05:42 |
| 190872 | 1 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-03 | International | Sea | Hi Tech | \$1,086.00 | 80% Just in time | 32% | 76% Order Confirmed | 2017-Nov-23 11:49 |
| 190871 | 2 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-01 | International | Sea | Hi Tech | \$1,881.00 | 83% Just in time | 22% | 74% Supply Date Reco | 2017-Nov-21 20:35 |
| 190871 | 1 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-01 | International | Sea | Hi Tech | \$1,881.00 | 84% Just in time | 27% | 77% Order Confirmed | 2017-Nov-24 10:46 |
| 190870 | 2 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Oct-29 | International | Sea | Hi Tech | \$2,162.00 | 62% Just in time | 30% | 78% Supply Date Reco | 2017-Nov-21 07:21 |
| 190870 | 1 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Oct-29 | International | Sea | Hi Tech | \$2,162.00 | 63% Just in time | 35% | 78% Order Confirmed | 2017-Nov-23 17:25 |
| 190869 | 2 | 2017-Aug-21 00:06 | 2017-Aug-21 00:06 | 2017-Nov-04 | International | Sea | Hi Tech | \$3,957.00 | 86% Just in time | 20% | 73% Supply Date Reco | 2017-Dec-03 18:22 |

Fonte: apromore.org

3.1.3 bupaR

BupaR - *Business Process Analysis in R* - trata-se de uma ferramenta *open-source* desenvolvida pela Universidade de Hasselt, na Bélgica, e mantida pela equipe liderada por

Gert Janssenswillen (HASSELT UNIVERSITY, 2019). É uma ferramenta que contém um pacote integrado de nove pacotes desenvolvidos em linguagem R (JANSSENSWILLEN; DEPAIRE, 2017) que têm como objetivo manusear e analisar dados contidos em processos de negócios (HASSELT UNIVERSITY, 2019). Os pacotes estão disponíveis em repositórios na plataforma de hospedagem de código-fonte *GitHub* e usuários finais podem contribuir na atualização desses repositórios, além da equipe desenvolvedora. Para começar, deve-se utilizar a ferramenta CRAN, que opera com códigos e documentação em linguagem R, para instalação dos pacotes na máquina do usuário. A última versão disponível em <https://gertjanssenswillen.github.io/bupaR/index.html> é a versão 0.4.0.9000.

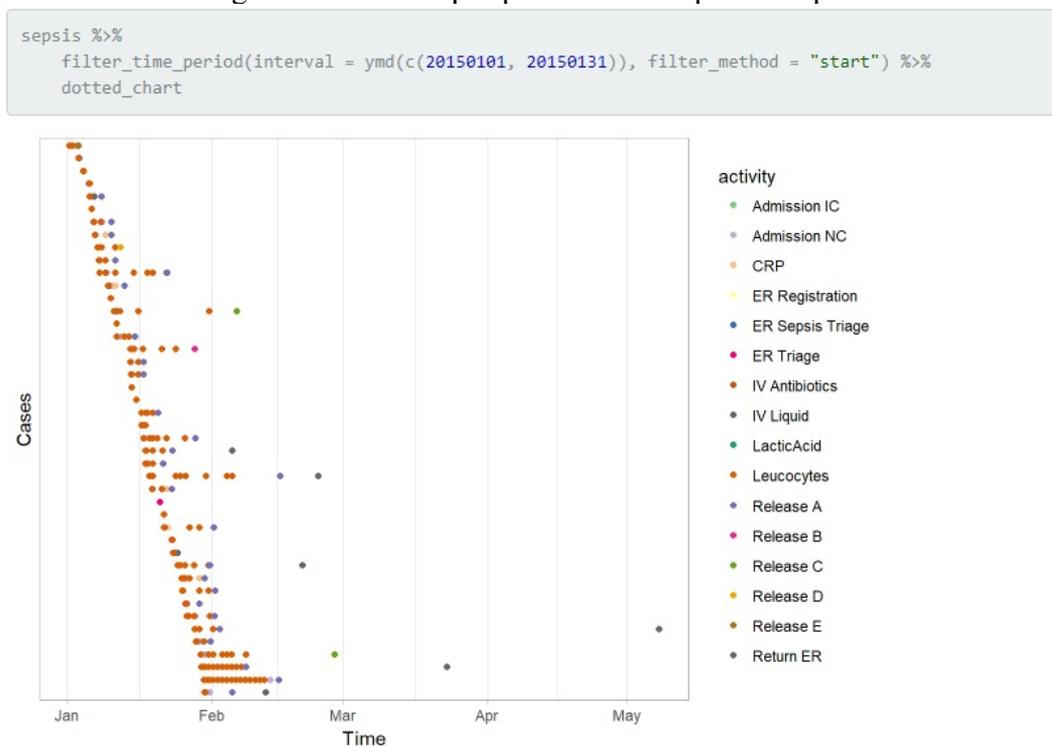
A ferramenta interpreta *logs* de eventos com o auxílio de dois de seus pacotes. O pacote central *bupaR* permite ao usuário criar objetos de *logs* de eventos, por meio de um *dataset* que contém características contidas em *logs* de eventos como, por exemplo, ID de caso, de atividade, *timestamp*, etc (JANSSENSWILLEN; DEPAIRE, 2017). O pacote também é capaz de ler e interpretar arquivos de formato CSV. Já o pacote *xesreadR* é capaz de interpretar arquivos de formato XES (JANSSENSWILLEN et al., 2019). *BupaR* disponibiliza também um pacote denominado *eventdataR*, que contém diversos exemplos de *logs* de eventos prontos para uso, tanto artificiais quanto representando eventos reais (JANSSENSWILLEN et al., 2019). O modelo de exportação dos dados é representado em Redes de Petri (em formato PNML), com o auxílio do pacote *petrinetR*, que tem como função construir, visualizar, exportar e reproduzir Redes de Petri (JANSSENSWILLEN et al., 2019).

A filtragem de dados em *BupaR* pode ser realizada tanto para casos quanto para eventos. A filtragem por casos é feita com o auxílio do pacote *edeaR* e pode ser baseada em performance, em características do fluxo de controle ou período de tempo (HASSELT UNIVERSITY, 2019):

- *Baseadas em performance*: filtra casos por tempo de processamento (*throughput time* ou *processing time*) ou por tamanho do caminho percorrido - *trace length*);
- *Baseadas em características do fluxo de controle*: filtra casos por presença (ou ausência) de uma certa atividade (ou mais de uma ao mesmo tempo), casos que tenham início (ou fim) em uma certa atividade, por precedência entre duas ou mais atividades ou por frequência em um caminho;
- *Baseadas em período de tempo*: filtra por casos iniciados, completados ou contidos em um certo intervalo informado ou que o intersectem.

A filtragem por eventos permite filtrar atividades (ou recursos) por nome e por frequência que essas atividades ou recursos ocorram em um processo – por porcentagem de ocorrência ou por intervalo em um atributo. O filtro especial (*trim filter*), que permite filtrar por atividades de início e de término (filtrado a partir da primeira aparição da atividade de início até a última aparição da atividade de término) com o uso da função *filter-trim* e também por intervalos de tempo, com o uso da função *filter-time-period*, que possui dois argumentos de entrada (*interval* e *filter-method*) (HASSELT UNIVERSITY, 2019). Este segundo argumento pode atuar também como um filtro de casos tal como o funcionamento da filtragem por casos baseada em períodos de tempo. A Figura 3.12 traz um exemplo de filtro por período de tempo exibindo um gráfico onde os pontos representam as atividades executadas.

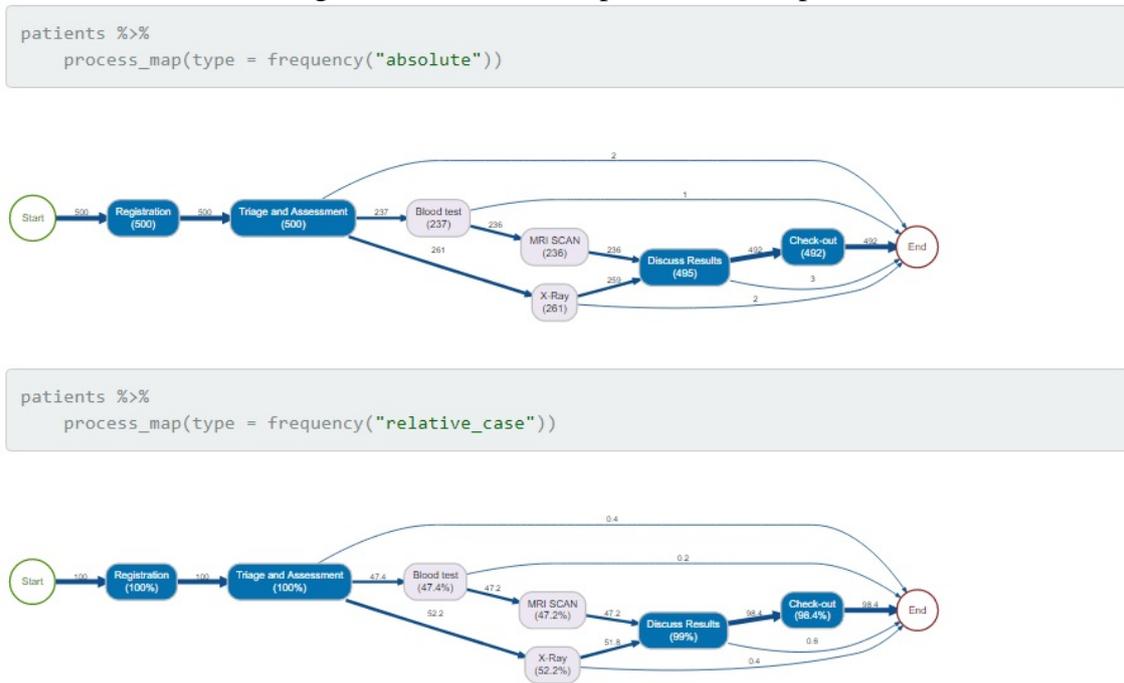
Figura 3.12: Filtro por período de tempo em bupaR



Fonte: www.bupar.net

O pacote *processmapR* é responsável por criar o modelo de descoberta dos processos a partir de seus *logs* de eventos, como visto na Figura 3.13 (HASSELT UNIVERSITY, 2019). Por *default*, exhibe o comportamento do processo por frequências de atividades ou fluxos, em números absolutos ou relativos. Também permite ao usuário analisar o comportamento por performance, com foco no tempo de processamento das atividades realizadas. Caso não seja suficiente, o usuário pode customizar a análise utilizando atributos numéricos combinados com funções de agregação (média, soma, máximo, mínimo,

Figura 3.13: Modelo de processo em bupaR



Fonte: www.bupar.net

etc.) ou também combinar mais de um atributo com o apoio das funções *type-edges* e *type-nodes*. A função *resource-map* contida nesse pacote pode criar uma rede de entrega de trabalho (*handover-of-work*), que pode ser representada tanto por uma Rede de Petri quanto por uma matriz de precedência de recursos. Já o pacote *processanimateR* é uma extensão de *processmapR* (JANSSENSWILLEN et al., 2019) e possibilita exibir a execução de um processo em animação em formato SVG - *Scalable Vector Graphics*, tornando a visualização de seu funcionamento mais fácil para o usuário.

O pacote *processmonitR* disponibiliza ao usuário gráficos de monitoramento de processos por diferentes perspectivas: focado em atividades, em recursos, em retrabalho e em performance (HASSELT UNIVERSITY, 2019).

Mais recentemente, foi adicionado o pacote *processcheckR* que tem como objetivo realizar a checagem de conformidade de um processo trabalhado na ferramenta. A checagem é baseada em regras é feita com o apoio da função *check-rule*, que recebe dois argumentos: o primeiro se refere a uma regra (início, término, precedência, sucessão, dentre outras) e o segundo corresponde a um nome possível de atributo referido no primeiro argumento (GERT JANSSENSWILLEN, 2019), como exemplificado na Figura 3.14.

Figura 3.14: Exemplo de checagem de conformidade em bupaR

```
sepsis %>%
  # check if cases starts with "ER Registration"
  check_rule(starts("ER Registration"), label = "r1") %>%
  # check if activities "CRP" and "LacticAcid" occur together
  check_rule(and("CRP", "LacticAcid"), label = "r2") %>%
  group_by(r1, r2) %>%
  n_cases()
#> # A tibble: 4 x 3
#> # Groups:   r1 [?]
#>   r1     r2     n_cases
#>   <lgl> <lgl>     <int>
#> 1 FALSE FALSE       10
#> 2 FALSE  TRUE       45
#> 3  TRUE  FALSE      137
#> 4  TRUE  TRUE       858
```

Fonte: www.github.com/gertjanssenswillen

3.1.4 Comparativo Entre as Ferramentas *Open-Source*

A Tabela 3.1 traz a comparação entre as ferramentas *open-source* e suas características. ProM é uma ferramenta desenvolvida com a missão de ser a plataforma padrão de mineração de processos de negócio no mundo acadêmico, o que pode ser visto estudando suas funcionalidades, que visam abordar todas as características contidas nas fases de descoberta, conformidade e aprimoramento. A existência de inúmeros *plug-ins*, com a capacidade de realizar diferentes tarefas no contexto da mineração, tornam a ferramenta expansível e de fácil manutenção. ProM disponibiliza diversas opções de formatos para exportação do modelo de processos. A checagem de conformidade contida na ferramenta obedece as métricas de análise espacial de estado, estrutural do modelo e de repetição de *log*, além de utilizar a dimensão de *fitness* (ROZINAT, 2010).

Apromore, além de ser uma ferramenta difundida no meio acadêmico, também tem como clientes grandes empresas que utilizam seus serviços de mineração de processos (THE UNIVERSITY OF MELBOURNE, 2019). Uma alternativa de acesso à ferramenta é a possibilidade de acessá-la na nuvem e manipular modelos e *logs* livremente, conforme o *nodo* mais próximo geograficamente de quem utiliza, uma cortesia das universidades parceiras e que hospedam esses nodos. Tem como diferencial em comparação às ferramentas analisadas nesse estudo a adaptação e a integração com BPM e seu ciclo de vida (THE UNIVERSITY OF MELBOURNE, 2019), e também a presença da análise de modelos BPMN em todas as fases da mineração de processos que realiza. Possui uma

Tabela 3.1: Comparativo entre Ferramentas *Open-Source*

| <i>Características</i> | <i>ProM</i> | <i>Apromore</i> | <i>bupaR</i> |
|---------------------------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------|
| Formato de importação suportado | MXML, XES | MXML, XES | CSV, XES, inserção diretamente na ferramenta |
| Modelo de exportação | BPMN, Redes de Petri, sistema de transição, modelos <i>fuzzy</i> , heurístico e minerado indutivo | BPMN, Redes de Petri, YAWL, AML, XPDL | Redes de Petri |
| Versão corrente | 6.8 | 7.0.1 | 0.4.0.900 |
| Plataforma de execução | Máquina do usuário (<i>stand-alone</i>) | Máquina do usuário (<i>stand-alone</i>) ou nuvem (<i>Cloud Nodes</i>) | Máquina do usuário (<i>stand-alone</i>) |
| Filtragem de dados | Sim | Sim | Sim |
| Descoberta de processos | Sim | Sim | Sim |
| Checagem de conformidade | Sim | Sim | Sim |
| Mineração de rede social | Sim | Sim | Sim |
| Mineração de regras de decisão | Sim | Sim | Não |
| Visualização de processos | Sim | Sim | Sim |
| Análise de performance | Sim | Sim | Sim |

Fonte: O Autor.

descoberta de processos bastante eficiente, onde o usuário tem a capacidade de extrair diversas informações tanto em relação às atividades de um processo quanto às conexões entre elas. Disponibiliza também uma descoberta mais avançada para modelos BPMN. Dentre várias outras funcionalidades adicionais, permite um monitoramento preditivo dos *logs* de eventos importados (THE UNIVERSITY OF MELBOURNE, 2019).

A ferramenta BupaR é composta por nove pacotes integrados, cada um respondendo a uma ou mais funcionalidades que são responsáveis por diferentes estágios da mineração de processos, desde um pacote para interpretar os *logs* de eventos até um pacote para criação dos modelos de exportação em formato de Redes de Petri (JANSSENSWILLEN; DEPAIRE, 2017). A modularidade entre pacotes é uma facilidade que auxilia não só na manutenção das funcionalidades existentes, como também no desenvolvimento de novas funcionalidades, como a checagem de conformidade, desenvolvida recentemente (GERT JANSSENSWILLEN, 2019). A filtragem de dados, por casos ou eventos, permite

ao usuário analisar o processo por diferentes perspectivas, requisitando comandos simples de entrada por parte do usuário.

As três ferramentas *open-source* estudadas buscam abordar todas - ou praticamente todas - as características contidas nas fases da mineração de processos, principalmente ProM e Apromore. Possuem uma filtragem de dados e uma fase de descoberta de processos eficiente, permitindo ao usuário final analisar um processo por várias perspectivas, diferenciando-se apenas no modo como são implementadas (*plug-ins*, *features* ou separada em pacotes).

São várias as opções de formatos que ProM disponibiliza ao usuário para exportação do modelo de descoberta de processos, enquanto BupaR tem apenas como opção o formato de Redes de Petri. A existência de diversos *plug-ins* realizando diversas funcionalidades em ProM tornam a ferramenta de fácil expansão e manutenção, o que também pode ser visto em BupaR de forma análoga, no que diz respeito a seus pacotes. Essa variedade de plugins torna ProM a ferramenta mais robusta dentre as analisadas nesse trabalho, assim como o fato de ter sido a primeira ferramenta a implementar a checagem de conformidade (ROZINAT, 2010), usando métricas específicas, mais próximas do que é visto na teoria, enquanto BupaR faz uma checagem mais simples, baseada em regras (GERT JANSSENSWILLEN, 2019). ProM, assim como Apromore, pode extrair regras de decisão após interpretar *logs* de eventos.

Apromore possui alguns diferenciais em relação às outras ferramentas *open-source* estudadas. Pode ser executada na nuvem e possui alta integração com o ciclo de vida de BPM, o que não é visto em ProM e BupaR. É uma ferramenta não tão robusta quanto ProM, mas possui uma interface mais amigável a um usuário iniciante na área de mineração de processos. BupaR é uma ferramenta menos robusta que ProM e Apromore, mas não se afasta tanto delas no que diz respeito à quantidade de funcionalidades que apresenta. Uma diferença de BupaR é a execução dessas funcionalidades por meio de funções, o que pode ser mais simples comparado às outras duas ferramentas, apesar de não apresentar o mesmo alto nível de abstração delas.

3.2 Ferramentas Comerciais

Esta Seção traz o comparativo entre as ferramentas comerciais Disco, Celonis, ARIS PPM e QPR Process Analyzer. As funcionalidades contidas em cada uma delas são detalhadas e comparadas mediante uso de uma tabela elaborada neste trabalho, com base

nas seguintes características: formatos de importação de dados e exportação de modelos, capacidade do *log* importado, plataforma em que é possível executá-las e se realizam filtragem de dados, descoberta de processos, checagem de conformidade, mineração de rede social, visualização de processos e análise de performance.

3.2.1 Disco

Disco é uma ferramenta de mineração de processos comercial desenvolvida pela empresa holandesa Fluxicon pelo time integrado pelos Doutores Anne Rozinat e Christian W. Gunther, além de Rudi Niks, um dos pioneiros nessa área (FLUXICON, 2019). Pode ser obtida mediante *download* em <https://fluxicon.com/disco/>, exigindo alguns requisitos do usuário interessado em seu uso, como Windows XP ou superior, MAC OS X 10.7 ou superior, Java 7.0 ou superior, dentre outras características operacionais. Também é permitido ao usuário, conforme sua necessidade, comprar o *software*, entrando em contato diretamente com a equipe de desenvolvimento, podendo oferecer diversas outras soluções além das disponíveis na versão gratuita. A Fluxicon permite que o usuário utilize a versão *demo* da ferramenta pelo tempo que tiver interesse, no entanto, essa versão estabelece limite de importação de dados a 100 eventos.

A empresa mantém parceria com diversas universidades, para as quais disponibiliza uma licença acadêmica, para que professores, pesquisadores e estudantes delas possam utilizar a ferramenta em sua totalidade. Entre as universidades parceiras estão as brasileiras UFPE (Universidade Federal de Pernambuco), UFRA (Universidade Federal Rural da Amazônia), UFPA (Universidade Federal do Pará), UFBA (Universidade Federal da Bahia), UNB (Universidade de Brasília), UFRJ (Universidade Federal do Rio de Janeiro), Unirio (Universidade Federal do Rio de Janeiro), Unigranrio (Universidade do Grande Rio), UFF (Universidade Federal Fluminense), PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro), USP (Universidade de São Paulo), UFMG (Universidade Federal de Minas Gerais), UEL (Universidade Estadual de Londrina), UFPR (Universidade Federal do Paraná), PUCPR (Pontifícia Universidade Católica do Paraná), IFES (Instituto Federal do Espírito Santo), Católica SC (Católica de Santa Catarina), UDESC (Universidade do Estado de Santa Catarina), FURG (Universidade Federal do Rio Grande), Unisinos (Universidade do Vale do Rio dos Sinos) e UFRGS (Universidade Federal do Rio Grande do Sul) (FLUXICON, 2019).

Disco permite importar diversos formatos de *logs* de eventos, tais como CSV,

MXML, MXML.GZ (formatos encontrados a partir da versão 5 de ProM), XES, XES.GZ (encontrados a partir da versão 6 de ProM), XLS, XLSX (obtidos a partir de dados salvos com o auxílio do Microsoft Excel), além dos formatos gerados pelo próprio Disco, tais como FXL Disco Logs e DSC Disco (FLUXICON, 2019). No entanto, os *logs* de eventos fornecidos à ferramenta devem preencher certos requisitos mínimos para serem analisados, tais como ID do caso, atividade e *timestamp* (GÜNTHER; ROZINAT, 2012), como exemplificado na Figura 3.15. A ferramenta exporta os dados no formato de modelo *fuzzy*.

Figura 3.15: Logs de eventos exibidos em Disco

| Case ID | Activity | Start Date | End Date | Agent Position | Customer ID | Product | Service Type | Resource | |
|---------|----------|---------------|---------------|----------------|-------------|-------------|--------------|----------------------|---------|
| 1 | Case 1 | Inbound Call | 9.3.10 9:05 | 9.3.10 9:10 | FL | Customer 1 | MacBook Pro | Referred to Servicer | Helen |
| 2 | Case 1 | Handle Case | 11.3.10 10:30 | 11.3.10 10:32 | FL | Customer 1 | MacBook Pro | Referred to Servicer | Helen |
| 3 | Case 1 | Call Outbound | 11.3.10 11:45 | 11.3.10 11:52 | FL | Customer 1 | MacBook Pro | Referred to Servicer | Henk |
| 4 | Case 2 | Inbound Call | 4.3.10 11:43 | 4.3.10 11:46 | FL | Customer 2 | MacBook Pro | Referred to Servicer | Susi |
| 5 | Case 3 | Inbound Call | 25.3.10 9:32 | 25.3.10 9:33 | FL | Customer 3 | MacBook Pro | Referred to Servicer | Mary |
| 6 | Case 4 | Inbound Call | 6.3.10 11:41 | 6.3.10 11:51 | FL | Customer 4 | iPhone | Referred to Servicer | Fred |
| 7 | Case 5 | Inbound Call | 18.3.10 10:54 | 18.3.10 11:01 | FL | Customer 5 | MacBook Pro | Product Assistance | Kenny |
| 8 | Case 6 | Inbound Call | 25.3.10 17:09 | 25.3.10 17:13 | FL | Customer 6 | MacBook Pro | Referred to Servicer | Harold |
| 9 | Case 6 | Inbound Call | 25.3.10 17:16 | 25.3.10 17:18 | FL | Customer 6 | MacBook Pro | Referred to Servicer | Nancy |
| 10 | Case 6 | Inbound Call | 26.3.10 8:36 | 26.3.10 8:40 | FL | Customer 6 | MacBook Pro | Referred to Servicer | Elena |
| 11 | Case 7 | Inbound Call | 18.3.10 11:49 | 18.3.10 11:50 | FL | Customer 7 | MacBook Pro | Product Assistance | Karen |
| 12 | Case 8 | Inbound Call | 11.3.10 9:20 | 11.3.10 9:23 | FL | Customer 8 | MacBook Pro | Referred to Servicer | Karen |
| 13 | Case 9 | Inbound Email | 19.3.10 19:47 | 21.3.10 8:17 | FL | Customer 9 | MacBook Pro | Product Assistance | Samuil |
| 14 | Case 9 | Call Outbound | 21.3.10 8:32 | 21.3.10 8:33 | FL | Customer 9 | MacBook Pro | Product Assistance | Samuil |
| 15 | Case 9 | Handle Email | 21.3.10 8:33 | 21.3.10 8:33 | FL | Customer 9 | MacBook Pro | Product Assistance | Samuil |
| 16 | Case 10 | Handle Email | 27.3.10 11:29 | 27.3.10 11:30 | FL | Customer 10 | iPhone | Product Assistance | Jochem |
| 17 | Case 11 | Inbound Call | 27.3.10 8:09 | 27.3.10 8:11 | FL | Customer 11 | iPhone | Product Assistance | Irena |
| 18 | Case 12 | Inbound Call | 29.3.10 9:28 | 29.3.10 9:29 | FL | Customer 3 | MacBook Pro | Product Assistance | Wil |
| 19 | Case 13 | Inbound Call | 5.3.10 10:13 | 5.3.10 10:15 | FL | Customer 12 | MacBook Pro | Product Assistance | Ton |
| 20 | Case 14 | Inbound Call | 4.3.10 7:49 | 4.3.10 7:50 | FL | Customer 3 | iPhone | Product Assistance | Wil |
| 21 | Case 15 | Inbound Call | 7.3.10 8:06 | 7.3.10 8:13 | FL | Customer 13 | iPhone | Product Assistance | Erik |
| 22 | Case 16 | Inbound Call | 25.3.10 10:26 | 25.3.10 10:34 | FL | Customer 14 | MacBook Pro | Referred to Servicer | Helen |
| 23 | Case 17 | Inbound Call | 4.3.10 7:35 | 4.3.10 7:46 | FL | Customer 15 | MacBook Pro | Referred to Servicer | Kenny |
| 24 | Case 17 | Handle Case | 4.3.10 7:53 | 4.3.10 7:55 | FL | Customer 15 | MacBook Pro | Referred to Servicer | Kenny |
| 25 | Case 17 | Handle Case | 9.3.10 11:16 | 9.3.10 11:18 | FL | Customer 15 | MacBook Pro | Referred to Servicer | Kenny |
| 26 | Case 17 | Handle Case | 11.3.10 11:15 | 11.3.10 11:19 | FL | Customer 15 | MacBook Pro | Referred to Servicer | Kenny |
| 27 | Case 17 | Inbound Call | 14.3.10 17:53 | 14.3.10 17:56 | FL | Customer 15 | MacBook Pro | Referred to Servicer | Monique |

Fonte: O Autor.

Disco tem como ponto positivo fornecer ao usuário um mapeamento automático de descoberta de processos tão logo o *log* de eventos é importado (GÜNTHER; ROZINAT, 2012). Esse mapeamento permite uma filtragem e inspeção de atributos facilitada, executando em paralelo à obtenção do modelo *fuzzy* descoberto.

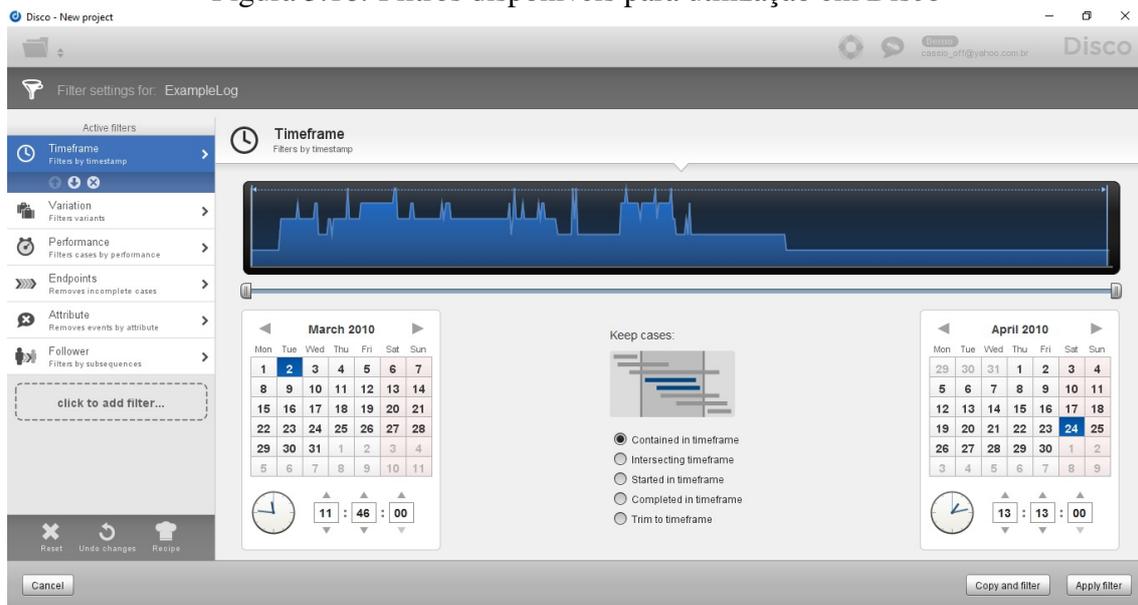
Disco possibilita analisar a execução de *logs* mediante o uso de seis tipos de filtros (GÜNTHER; ROZINAT, 2012), exibidos na Figura 3.16:

- **Filtro de prazo (timeframe):** O filtro de prazo permite ao usuário selecionar o *log* executado apenas em um intervalo de tempo. Também permite selecionar casos que tenham ocorrido estritamente dentro desse intervalo, casos intersectando o intervalo (que estivessem ocorrendo no início ou no final), casos que iniciaram – ou que terminaram – nesse intervalo e casos que tenham ocorrido dentro do intervalo dado;
- **Filtro de variação:** É importante para selecionar comportamentos de um processo, detectando quais são os mais (ou menos) comuns. Porém, à medida que o processo

estudado torna-se robusto, o filtro torna-se menos efetivo;

- *Filtro de performance*: Permite filtrar casos por um critério específico como, por exemplo, por sua duração, por número de eventos, por utilização do caso, por tempo total de execução, por tempo máximo de execução e por tempo total ocioso;
- *Filtro de ponto final (endpoint filter)*: Exclui casos que estejam em execução tanto no início de um dado intervalo quanto no final deste, permitindo filtrar apenas os casos que ocorrem dentro do intervalo;
- *Filtro de atributos*: Permite selecionar casos filtrando-os por atividade realizada ou por recursos obtidos;
- *Filtro de segmento (follower filter)*: tem como principal função identificar segregação de funções na execução do processo, também permitindo adicionar padrões de segmento.

Figura 3.16: Filtros disponíveis para utilização em Disco



Fonte: O Autor.

Após a aplicação de um determinado filtro, a ferramenta devolve o modelo *fuzzy* contendo as informações desejadas pelo usuário, como exemplificado na Figura 4.4, no Capítulo 4.

Disco também disponibiliza uma funcionalidade de animação (GÜNTHER; ROZINAT, 2012) que permite ao usuário visualizar a existência de gargalos simulando a execução do processo analisado, bem como permite que tal animação seja exportada em formato de vídeo.

3.2.2 Celonis

Celonis é uma ferramenta comercial de mineração de processos que começou a ser desenvolvida em 2011 pela empresa homônima alemã, pelo time integrado por Bastian Nominacher, Martin Klenk e Alexander Rinke. Possui como objetivo “analisar os processos de hoje para tornar o mundo de amanhã mais eficiente” (CELONIS, 2019). Pode ser obtida mediante *download* em <https://www.celonis.com/try-or-buy/>, disponibilizando uma versão *trial* válida por 30 dias para novos usuários, de forma que apenas após esse período de adaptação à ferramenta, o usuário sinta-se pronto para entrar em contato com a organização e adquiri-la. A ferramenta roda em ambiente Windows ou Mac OsX. Celonis tem como clientes diversas empresas mundialmente conhecidas, como a Siemens, Vodafone, Cisco, Uber, dentre outras. É uma das ferramentas líderes de mercado no segmento de mineração de processos (CELONIS, 2019). A versão mais recente da ferramenta é a versão 4.4. Informações gerais sobre Celonis estão contidas na Figura 3.17 e serão detalhadas nesta Seção.

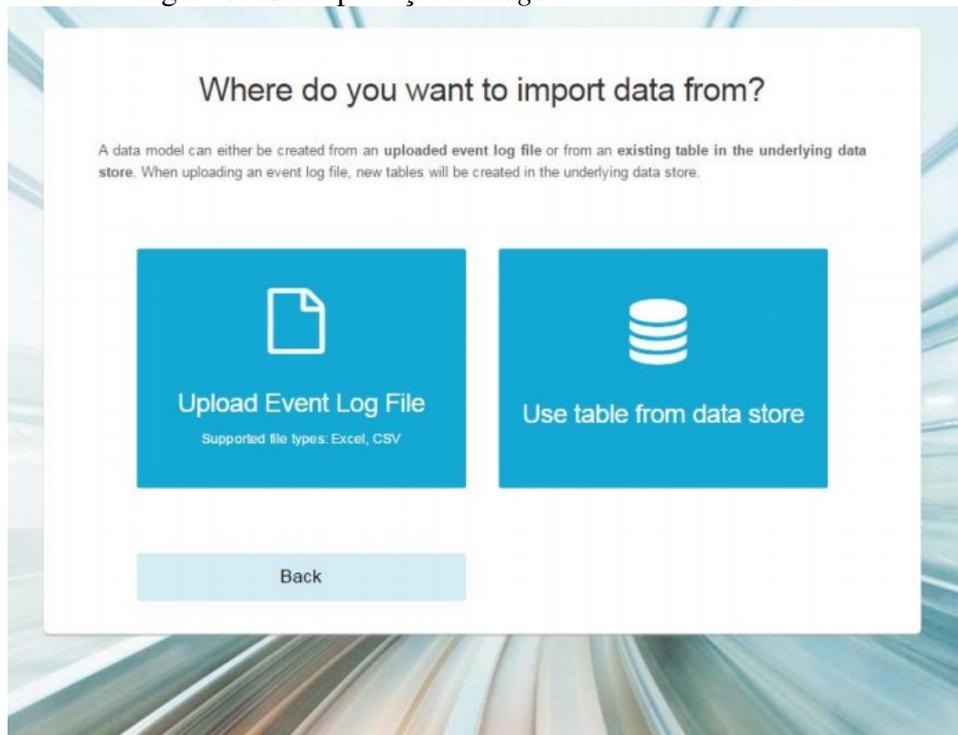
Figura 3.17: Informações sobre a ferramenta Celonis

The screenshot shows the Celonis website's 'Process mining: How does it work?' page. The header includes the Celonis logo and navigation links: 'Intelligent Business Cloud', 'Solutions', 'Customers', 'Explore', and 'Company', along with a 'GET STARTED' button. The main heading is 'Process mining: How does it work?'. Below this, a paragraph explains that most operational systems record digital footprints, and process mining captures these to show the journey to complete a process. A link 'what is process mining' is provided for getting started. A list of questions is also shown, such as 'What are the "stages" or main components of process mining?'. To the right, a grid of six cards highlights key features: 'New App' (A new app waiting to be built), 'Process AI' (Detect and analyze deviations from the most common path), 'Process Overview' (Get the main insights on your process), 'Process Explorer' (Analyze and understand your process), 'Conformance' (Compare the real process to your target process), and 'Social' (Understand how your team is working).

Fonte: www.celonis.com

A empresa alemã SAP, uma das maiores do ramo de desenvolvimento de *software* no mundo, estabeleceu parceria com a Celonis para desenvolvimento da sua própria ferramenta (*SAP Process Mining by Celonis*), que executa a mineração de processos na nuvem, permitindo a seus usuários a análise de grandes volumes de dados contidos nos *logs* de eventos (CELONIS, 2019).

Celonis tem como diferencial permitir a integração contínua em tempo real com sistemas de banco de dados relacionais (RDBMS) (KEBEDE, 2015) como os da SAP,

Figura 3.18: Importação de *logs* de eventos em Celonis

Fonte: (KEBEDE, 2015)

Salesforce, Oracle, Microsoft, entre diversas outras. Outro diferencial está no fato da empresa oferecer tanto a opção da plataforma via nuvem (SaaS – *Software as a Service*) quanto por instalação local na máquina do usuário (*on-premise*).

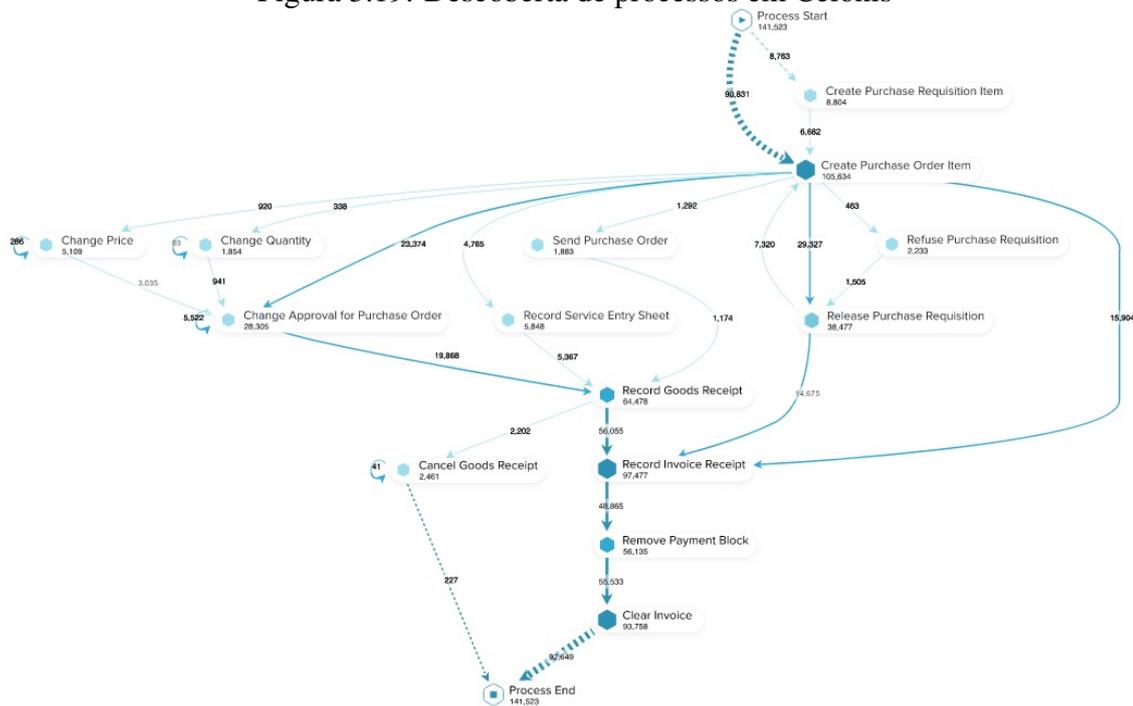
Celonis pode importar *logs* de eventos em formato CSV ou XLS, de tamanho limitado apenas pela cota do banco de dados utilizado, como exibido na Figura 3.18. A ferramenta exporta os dados interpretados no formato de modelo *fuzzy*.

Para importação dos *logs* de eventos para o Celonis, é preciso estabelecer conexão com o sistema dedicado PostgreSQL. O modelo de dados pode ser criado carregando um arquivo CSV ou XLS para a ferramenta ou uma tabela armazenada no banco de dados integrado à ferramenta. Após a importação, é configurada uma tabela de banco de dados, e é sobre ela que são aplicadas as funções componentes do *log*.

Existem três tipos de filtragem de dados em Celonis (KEBEDE, 2015):

- *Filtro de atributos*: Usado para selecionar casos baseados em atributos específicos de um certo caso;
- *Filtro de variantes*: Permite explorar o processo em questão baseado nas variantes mais ou menos comuns encontradas;
- *Filtro PQL - Process Query Language*: Usado para analisar padrões de processo.

Figura 3.19: Descoberta de processos em Celonis



Fonte: Adaptado de www.celonis.com

A descoberta de processos em Celonis permite ao usuário encontrar o chamado “caminho feliz”, que se trata da obtenção do caminho mais comum realizado por um processo de negócio, visualizar as variantes desse processo e detectar suas vulnerabilidades, caso existam, como pode-se visualizar na Figura 3.19 (CELONIS, 2019).

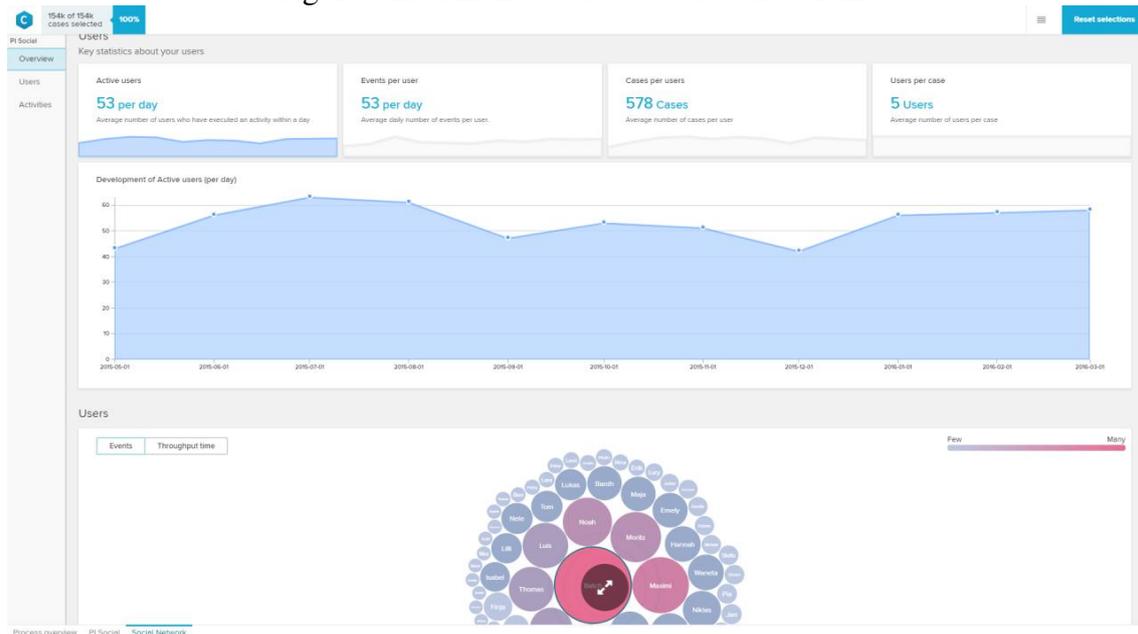
Figura 3.20: Checagem de conformidade em Celonis



Fonte: Adaptado de www.celonis.com

Celonis também realiza a análise dos resultados obtidos após a descoberta e a checagem de conformidade, que ocorre de forma automatizada, por *machine learning*, que checa a performance do processo em comparação ao seu modelo de descoberta (CELONIS, 2019). A ferramenta traz estatísticas como porcentagem de casos em conformidade, porcentagem de violações a ela e análise do impacto dessas violações no tempo de execução ou nos passos do processo, como visto na Figura 3.20.

Figura 3.21: Análise de rede social em Celonis



Fonte: Adaptado de www.celonis.com

A Figura 3.21 traz a análise de rede social em Celonis, que permite detectar a atuação de usuários, membros de um time ou assemelhados de forma a maximizar sua contribuição nos objetivos da organização a qual pertencem, possibilitando redistribuição de tarefas e otimização de recursos (CELONIS, 2019).

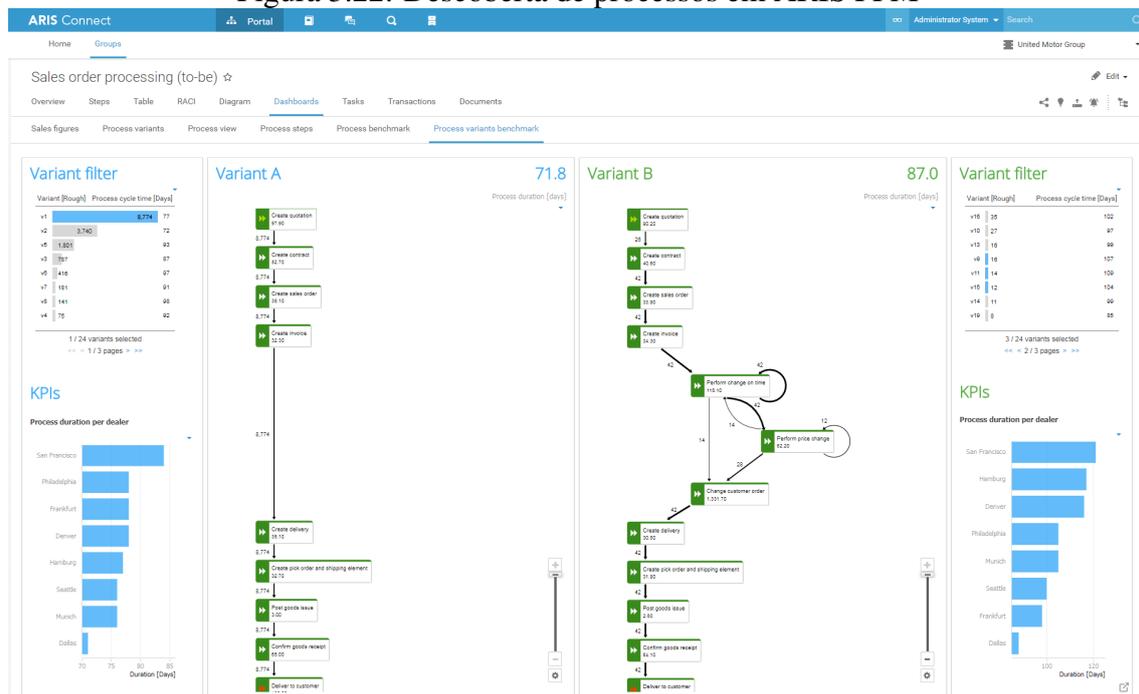
3.2.3 ARIS PPM

ARIS PPM (*Process Performance Manager*) é uma ferramenta comercial de mineração de processos de negócio desenvolvida pela *Software AG* e tem como clientes conhecidos organizações como Coca-Cola, Philips, Mapfre Seguros, PricewaterhouseCoopers, entre outros (SOFTWARE AG, 2019). Disponibiliza ao usuário interessado em seu uso uma versão *trial* da ferramenta com validade de 30 dias, uma forma de ambientar os novos usuários à plataforma antes de adquiri-la entrando em contato com a organização

desenvolvedora. A última versão lançada pela *Software AG* é a ARIS 10 SR8.

O formato CSV é aceito pela ferramenta para processamento dos *logs* de eventos (AILENEI et al., 2011). Também permite exportar dados de sistemas de banco de dados como os da SAP, Salesforce e Oracle, por exemplo. O modelo de exportação pode ser representado tanto em EPC quanto em BPMN, este último adicionado à versão mais recente da ferramenta (SOFTWARE AG, 2019).

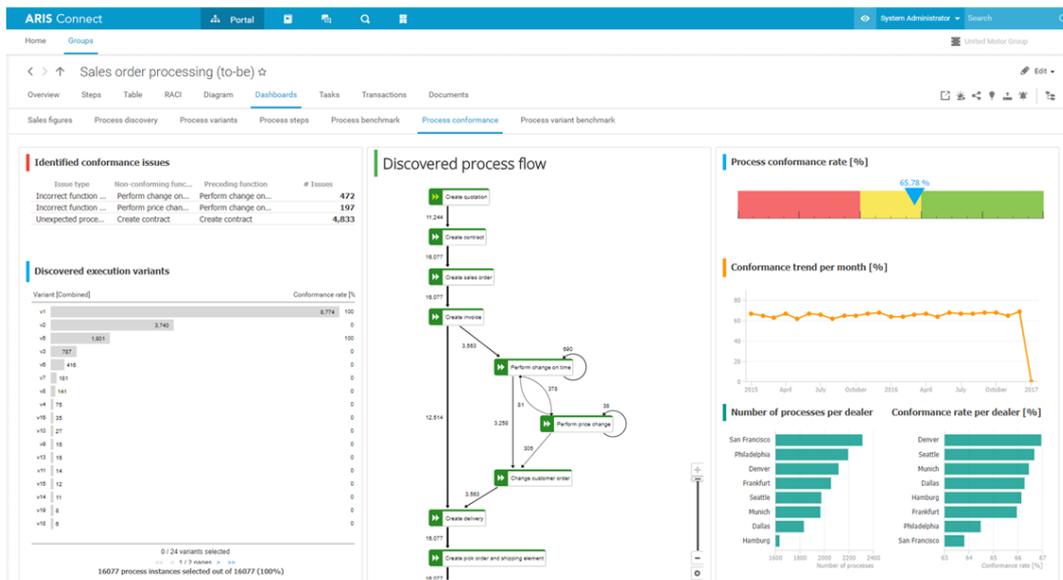
Figura 3.22: Descoberta de processos em ARIS PPM



Fonte: www.softwareag.com

A Figura 3.22 representa a descoberta de processos em ARIS PPM, que realiza dois passos necessários: primeiro, os eventos pertencentes a mesma instância de um processo são agrupados e, segundo, esses eventos são organizados em seu fluxo correto (AILENEI et al., 2011). Possibilita identificar variantes, partindo do fluxo mais comum seguido em um processo até a visualização completa de todas as operações executadas. Após essa identificação, é possível ter acesso às causas de possíveis gargalos, relacionados a custo ou tempo alto de execução, desvios do fluxo mais comum e onde ocorrem possíveis perdas de performance (SOFTWARE AG, 2019). Também em sua versão mais recente, ARIS PPM trouxe a checagem de conformidade baseada em modelos no formato BPMN, complementando a que já existia em formato EPC que é exibida na Figura 3.23.

Figura 3.23: Checagem de conformidade em ARIS PPM



Fonte: www.softwareag.com

3.2.4 QPR Process Analyzer

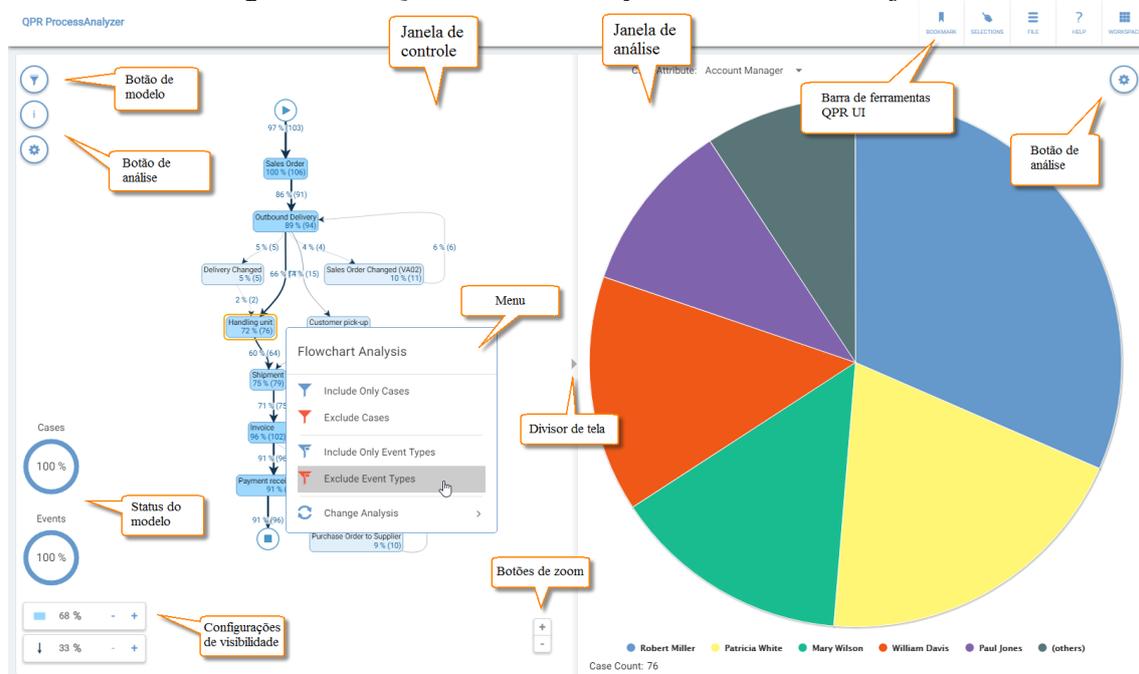
QPR Process Analyzer é uma ferramenta comercial desenvolvida pela empresa *QPR Software Oyj*, sediada em Helsinque, Finlândia (QPR, 2019), para mineração de processos de negócio. Tem como clientes conhecidos organizações como Nokia e Banco Espírito Santo e parceiros como SAP, Oracle, Microsoft, entre outros. QPR Process Analyzer fornece ao usuário três possibilidades de ambiente: *Xpress*, instalada na máquina do usuário (*stand-alone*); *Pro*, com o auxílio de serviços *web*; ou *Database*, com conexão direta a um servidor de banco de dados do tipo SQL Server (QPR, 2019). É necessário obter uma licença mensal para uso da ferramenta com o custo de 1000 euros mensais (QPR, 2019).

A ferramenta suporta *logs* de eventos importados em formato CSV, XES, PACM (formato de compressão de dados da própria ferramenta) e SQL Database (QPR, 2019). Contudo, é recomendado ao usuário que o arquivo de importação siga requisitos como: um cabeçalho na primeira linha, as três primeiras colunas reservadas para informações de ID do evento, nome do evento e *timestamp*, que deve estar em formato *data/horário*. Os dados podem ser exportados em formato PACM, contendo atributos de caso e evento, marcadores para os quais o usuário que deseja exportar tenha ao menos direito de visualização dos dados, além dos direitos de usuário específicos do modelo importado (QPR,

2019).

A descoberta de processos na ferramenta é realizada com o uso do objeto de apresentação PAPO (*QPR Process Analyzer Presentation Object*), que permite a filtragem dos dados e realização de análises sobre eles. A exibição dos resultados é dividida em duas janelas: de controle e de análise. A janela de controle exibe informações acerca do filtro que o usuário deseje aplicar, enquanto a janela de análise exibe informações acerca da análise dos itens selecionados na janela de controle. Os filtros disponíveis para seleção permitem filtrar apenas casos, excluir certos casos, apenas eventos ou excluir certos eventos. PAPO disponibiliza também análises por caso, duração, evento, tipo de evento, fluxograma, variações, entre outras, como visto na Figura 3.24.

Figura 3.24: QPR Process Analyzer Presentation Object



Fonte: Adaptado de <https://devnet.onqpr.com/>

QPR Process Analyzer realiza a checagem de conformidade com o uso da funcionalidade *Conformance Analysis View*, que permite a visualização tanto das violações à conformidade quanto sugere possíveis causas da não-conformidade (QPR, 2019). Inclui em sua visualização as violações agrupadas entre eventos indesejados, eventos que ocorrem fora da ordem do modelo e casos incompletos. Também exibe a porcentagem de casos que apresentam violação, o impacto da violação na duração de um caso e o impacto no número de eventos a serem executados até o término da execução de um processo. Quanto às possíveis causas de não-conformidade, o botão *Show Root Cases* exibe o número total de casos que possuem um certo valor de atributo informado de um caso, a

porcentagem de casos em não-conformidade e a porcentagem de contribuição desse valor de atributo informado para a não-conformidade dos casos (QPR, 2019). O modelo de visualização para checagem de conformidade é definido usando a notação BPMN 2.0.

3.2.5 Comparativo Entre as Ferramentas Comerciais

Tabela 3.2: Comparativo entre Ferramentas Comerciais

| <i>Características</i> | <i>Disco</i> | <i>Celonis</i> | <i>ARIS PPM</i> | <i>QPR Process Analyzer</i> |
|------------------------------------|------------------------------------------------------------|----------------------------------------------------------|-------------------------------------------|----------------------------------------------------------------------------------------------|
| Formato de importação suportado | CSV, MXML, MXML.GZ, XES, XES.GZ, XLS, XLSX, FXL, DSC | CSV, XLS | CSV | CSV, XES, PACM, SQL Database |
| Modelo de exportação | Modelo <i>fuzzy</i> | Modelo <i>fuzzy</i> com suporte gráfico | EPC, BPMN | PACM |
| Capacidade do <i>log</i> importado | 100 eventos (versão demo), 5 milhões de eventos (completa) | Limitado pelo tamanho do banco de dados | Limitado pelo tamanho do banco de dados | 1 milhão de eventos (Excel) ou limitado pelo tamanho do banco de dados |
| Plataforma de execução | Máquina do usuário (<i>stand-alone</i>) | Máquina do usuário (<i>on-premise</i>) ou nuvem (SaaS) | Máquina do usuário (<i>stand-alone</i>) | Máquina do usuário (<i>Xpress</i>), Web (<i>Pro</i>) ou Database (conexão ao SQL Server) |
| Filtragem de dados | Sim | Sim | Sim | Sim |
| Descoberta de processos | Sim | Sim | Sim | Sim |
| Checagem de conformidade | Não | Sim | Sim | Sim |
| Mineração de rede social | Não | Sim | Não | Não |
| Visualização de processos | Sim | Sim | Sim | Sim |
| Análise de performance | Sim | Sim | Sim | Sim |

Fonte: O Autor.

A Tabela 3.2 traz a comparação entre as ferramentas comerciais avaliadas neste

estudo e suas características. Disco possui uma interface bastante simples e familiar ao usuário iniciante em mineração de processos. Apesar de ser uma ferramenta desenvolvida para uso comercial, as parcerias estabelecidas com inúmeras universidades distribuídas por todo o mundo trazem como resultado a ampla utilização também no meio acadêmico. Interpreta diversos formatos de *logs* de eventos, incluindo as mais populares (CSV, MXML, XLS), além de formatos da própria ferramenta (FXL, DSC). Ainda que a capacidade do *log* de eventos importado seja limitada a 5 milhões de eventos, esse número pode ser considerado suficientemente grande para atender satisfatoriamente seus usuários. Tem uma filtragem de dados bastante eficiente e que permite análises por diversas perspectivas a quem trabalha com a ferramenta. O mapeamento automático dos processos executados, bem como a exibição em formato de animação, simulando uma execução, são outras vantagens que podem trazer novos usuários à ferramenta desenvolvida pela Fluxicon.

Celonis conta com a confiança de diversas organizações multinacionais, o que suporta a posição da ferramenta entre as líderes no segmento de mineração de processos de negócio. Uma delas é a alemã SAP, que pôde desenvolver sua própria ferramenta. A integração com sistemas RDBMS é um diferencial de Celonis, assim como a opção de execução na nuvem (no modelo SaaS – *Software as a Service*). Adicionalmente, a ferramenta disponibiliza a análise de rede social, permitindo a seus usuários uma análise mais detalhada no que diz respeito a atuação de atores em uma organização.

ARIS PPM, em comparação às ferramentas analisadas nesse estudo, é a que possui maior integração com BPMN, permitindo ao usuário utilizar esse formato para exportar um modelo de processos e checar sua conformidade, que pode ser checada também em EPC.

QPR Process Analyzer pode ser considerada uma ferramenta multi ambiente - *stand-alone*, *web* ou *Database*. Seu objeto de apresentação - PAPO - facilita ao usuário visualizar o modelo de processos, filtrar dados que ache necessário e analisá-los. A checagem de conformidade da ferramenta, sendo possível, sugere possíveis causas da não-conformidade (QPR, 2019).

O acesso a Disco é facilitado, em comparação às demais ferramentas comerciais, disponibilizando uma versão *demo* sem data de expiração, diferentemente de Celonis e ARIS PPM, além de dispor a usuários acadêmicos uma versão ilimitada. Por outro lado, o acesso a QPR Process Analyzer possivelmente é o mais limitado, devido ao alto custo financeiro de sua licença mensal. Disco também é a ferramenta que interpreta o maior número de formatos de *logs* de eventos existentes, apenas com a ressalva de exigir certos

requisitos mínimos de formatação. Em contrapartida, Celonis e ARIS PPM suportam apenas o formato CSV (Celonis também suporta o formato XLS). A filtragem de dados em Disco é bastante eficiente e de fácil visualização por seu usuário final, tendo como desafiante nesse quesito QPR Process Analyser e seu objeto PAPO.

A ferramenta QPR Process Analyzer é a que disponibiliza mais alternativas de ambiente para que um usuário utilize seus serviços, ao contrário de Disco e ARIS PPM, que operam somente em instalação local na máquina do usuário (*stand-alone*).

A checagem de conformidade é uma fase da mineração de processos que aos poucos vai sendo integrada às ferramentas existentes na indústria. ARIS PPM e QPR Process Analyzer acompanharam essa tendência e realizam essa fase baseada em modelos BPMN, assim como exibem seus resultados de forma mais detalhada em relação às demais ferramentas comerciais estudadas. Celonis também possui essa funcionalidade, ao contrário da ferramenta Disco.

Dentre as quatro ferramentas comerciais estudadas, Celonis é a única que tem implementada a análise de rede social, buscando atingir os objetivos de uma organização maximizando a contribuição dos membros lotados nela, distribuindo tarefas entre esses membros e otimizando o uso dos recursos disponíveis. Assim sendo, podemos concluir que Disco e Celonis são as ferramentas mais completas, com interface mais amigável ao usuário, maior facilidade de acesso e que permitem um maior número de possibilidades a aqueles que desejam utilizá-las com o objetivo de minerar seus processos de negócio.

3.3 Resumo do Capítulo

Neste capítulo, são detalhadas as funcionalidades de sete ferramentas de mineração de processos de negócio: ProM, Apromore, BupaR (*open-sources*), Disco, Celonis, ARIS PPM e QPR Process Analyzer (comerciais/proprietárias). Também são construídas tabelas comparativas dentro das categorias, buscando comparar as funcionalidades das ferramentas tendo como base as seguintes características: formatos de importação de dados e exportação de modelos, plataformas de execução e a existência ou não de filtragem de dados, descoberta de processos, checagem de conformidade, mineração de rede social, mineração de regras de decisão, visualização de processos e análise de performance. A partir da construção das tabelas em cada categoria, é possível detectar semelhanças entre as ferramentas, bem como as diferenças entre elas. As tabelas também permitem apresentar vantagens e desvantagens de cada uma das ferramentas, conforme as expectativas

e necessidades de um usuário da área de mineração de processos.

4 ESTUDO DE CASO

Neste capítulo, é apresentado um estudo de caso envolvendo as duas ferramentas mais populares e de mais fácil acesso, ou seja, que não requerem investimentos financeiros elevados para utilização, conforme visto no Capítulo 3. As ferramentas estudadas são ProM, representando as ferramentas *open-source*, e Disco, representando as ferramentas comerciais/proprietárias. Para estudar o comportamento delas, foi escolhido um exemplo de *log* de eventos contendo informações reais contido no repositório da *4TU.Datacenterum* (<https://data.4tu.nl/repository/>), de livre acesso. O *log* escolhido contém eventos de casos de sepse em um hospital e tem como colaborador o pesquisador Felix Mannhardt (MANNHARDT, 2016). Uma sepse é uma condição de saúde com potencial risco de vida geralmente causada por uma infecção (SINGER et al., 2016).

A máquina em que as ferramentas foram instaladas e o *log* foi interpretado possui as seguintes características: *Notebook* com processador *Intel Core i5* com 2.5 GHz (*gigahertz*) de frequência; memória RAM de 4 GB (*gigabytes*) de capacidade; sistema operacional Windows 10 de 64 bits; disco rígido com 500 GB de capacidade.

Um *caso* representa o caminho que um paciente percorre pelo hospital. Os eventos foram gravados pelo sistema de ERP de um certo hospital. O *log* contém por volta de 1000 casos, com um total de 15000 eventos registrados em 16 diferentes atividades. Foram registrados 39 atributos de dados como, por exemplo, o grupo responsável por uma atividade, o resultado de um teste e informações de *checklists*. Eventos e valores de atributos foram suprimidos do *log* e os *timestamps* dos eventos foram randomizados, mas o tempo entre eventos em um caminho não foi alterado. O *log* está em formato XES, portanto, compatível com ambas as ferramentas que são estudadas neste capítulo. Apenas para efeitos de visualização, foi exportado o *log* de eventos em formato CSV para geração da Figura 4.1, onde é possível visualizar informações contidas no *log*.

4.1 Modelo de descoberta

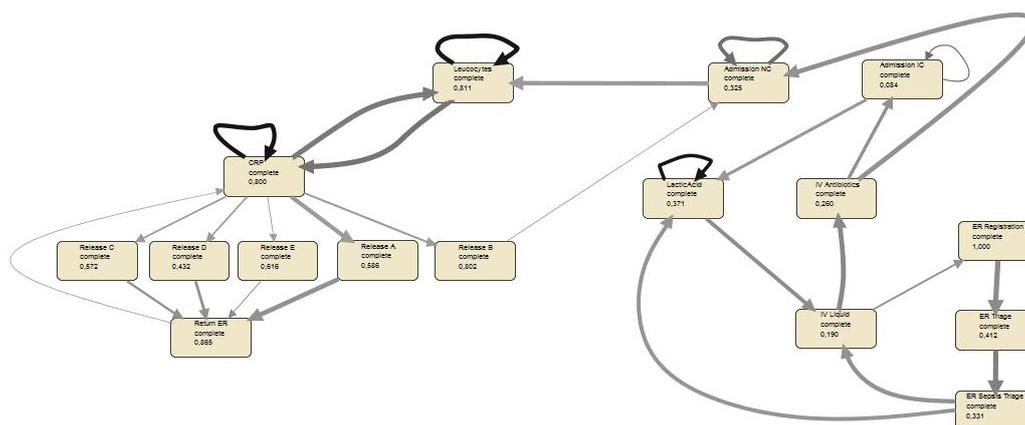
Após a leitura do *log* de eventos, foram gerados, em ambas as ferramentas, os modelos de descoberta do processo. Como a ferramenta Disco permite exibir o modelo de processos apenas em modelo *fuzzy*, escolhe-se a mesma opção de exibição em ProM para efeitos de comparação entre as ferramentas. A Figura 4.2 exibe o modelo de processo em ProM e a Figura 4.4 exibe o modelo de processo em Disco.

Figura 4.1: Log de eventos analisado

| Case ID | Activity | Complete Timestamp | Variant | Variant Index | Age | CRP | Diagnose | DiagnosticAnt | DiagnosticAstrup | DiagnosticBlood | DiagnosticECG | DiagnosticLacticAcid | DiagnosticLiquor | DiagnosticOther | DiagnosticSputum | Diagnost |
|---------|----------|--------------------|-------------------------|---------------|-----|-------|----------|---------------|------------------|-----------------|---------------|----------------------|------------------|-----------------|------------------|----------|
| 2 | A | ER Registration | 2014/10/22 07:15:41.000 | Variant 63 | 63 | 85 | A | true | true | true | true | true | false | false | false | true |
| 3 | A | Leucocytes | 2014/10/22 07:27:00.000 | Variant 63 | 63 | 21.0 | | | | | | | | | | |
| 4 | A | CRP | 2014/10/22 07:27:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 5 | A | LacticAcid | 2014/10/22 07:27:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 6 | A | ER Triage | 2014/10/22 07:33:37.000 | Variant 63 | 63 | | | | | | | | | | | |
| 7 | A | ER Sepsis Triage | 2014/10/22 07:34:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 8 | A | IV Liquid | 2014/10/22 10:03:47.000 | Variant 63 | 63 | | | | | | | | | | | |
| 9 | A | IV Antibiotics | 2014/10/22 10:03:47.000 | Variant 63 | 63 | | | | | | | | | | | |
| 10 | A | Admission NC | 2014/10/22 10:13:19.000 | Variant 63 | 63 | | | | | | | | | | | |
| 11 | A | CRP | 2014/10/24 05:00:00.000 | Variant 63 | 63 | 109.0 | | | | | | | | | | |
| 12 | A | Leucocytes | 2014/10/24 05:00:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 13 | A | Leucocytes | 2014/10/26 05:00:00.000 | Variant 63 | 63 | 47.0 | | | | | | | | | | |
| 14 | A | CRP | 2014/10/26 05:00:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 15 | A | Leucocytes | 2014/10/28 05:00:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 16 | A | CRP | 2014/10/28 05:00:00.000 | Variant 63 | 63 | 15.0 | | | | | | | | | | |
| 17 | A | CRP | 2014/10/30 05:00:00.000 | Variant 63 | 63 | 9.0 | | | | | | | | | | |
| 18 | A | Leucocytes | 2014/10/30 05:00:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 19 | A | Leucocytes | 2014/10/31 05:00:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 20 | A | CRP | 2014/10/31 05:00:00.000 | Variant 63 | 63 | 9.0 | | | | | | | | | | |
| 21 | A | CRP | 2014/11/02 05:00:00.000 | Variant 63 | 63 | 6.0 | | | | | | | | | | |
| 22 | A | Leucocytes | 2014/11/02 05:00:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 23 | A | Release A | 2014/11/02 12:15:00.000 | Variant 63 | 63 | | | | | | | | | | | |
| 24 | B | ER Registration | 2014/12/21 08:04:24.000 | Variant 64 | 64 | 45 | B | false | true | true | true | true | false | false | true | true |
| 25 | B | ER Triage | 2014/12/21 08:17:19.000 | Variant 64 | 64 | | | | | | | | | | | |
| 26 | B | CRP | 2014/12/21 08:36:00.000 | Variant 64 | 64 | 240.0 | | | | | | | | | | |
| 27 | B | LacticAcid | 2014/12/21 08:36:00.000 | Variant 64 | 64 | | | | | | | | | | | |
| 28 | B | Leucocytes | 2014/12/21 08:36:00.000 | Variant 64 | 64 | | | | | | | | | | | |
| 29 | B | ER Sepsis Triage | 2014/12/21 09:15:45.000 | Variant 64 | 64 | | | | | | | | | | | |
| 30 | R | | | | 64 | | | | | | | | | | | |

Fonte: O Autor.

Figura 4.2: Modelo de processo em ProM

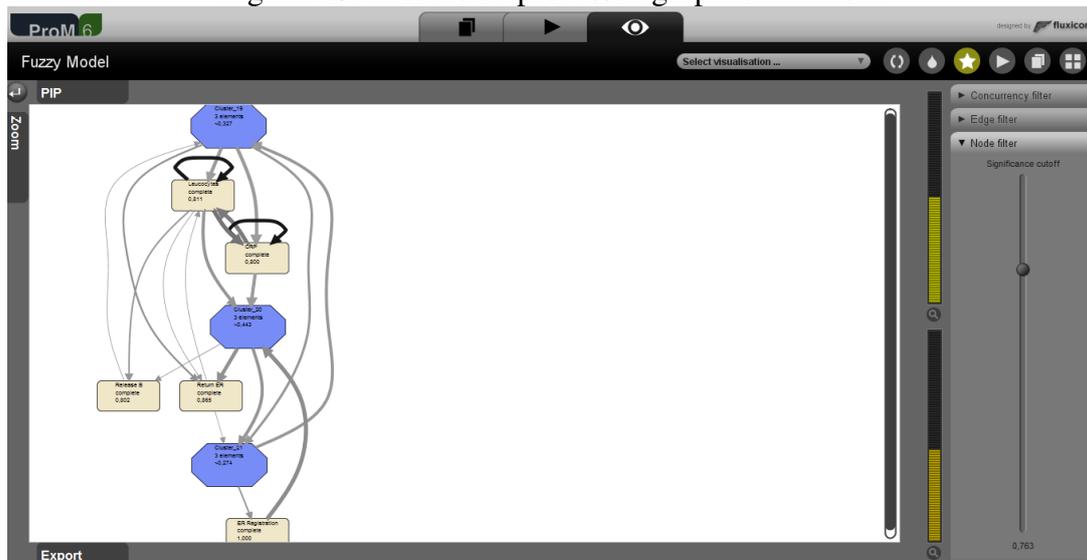


Fonte: O Autor.

Em ambas as ferramentas, o modelo pode ser visualizado em diferentes níveis de detalhamento. Em ProM, esse detalhamento é ajustável com o uso de três filtros: *nós* (correspondente às atividades), *arestas* (correspondente aos caminhos) ou por *concorrência*. Ao lado esquerdo do filtro constam duas barras: a barra *superior* que indica o nível de detalhamento do modelo e a *inferior* que indica o nível de conformidade do log em relação ao modelo exibido. A Figura 4.3 apresenta o modelo de processo com detalhamento de 50,58% e nível de conformidade de 66,08%. Quanto maior o detalhamento do modelo, maior é a porcentagem representada na barra superior. Reduzir o detalhamento impacta no nível de conformidade, o que explica o nível alcançado na Figura 4.3.

O filtro por nós contém uma barra de rolagem onde o extremo inferior refere-se ao nó menos significativo e o superior refere-se ao nó mais significativo. Dessa forma, o

Figura 4.3: Modelo de processo agrupado em ProM

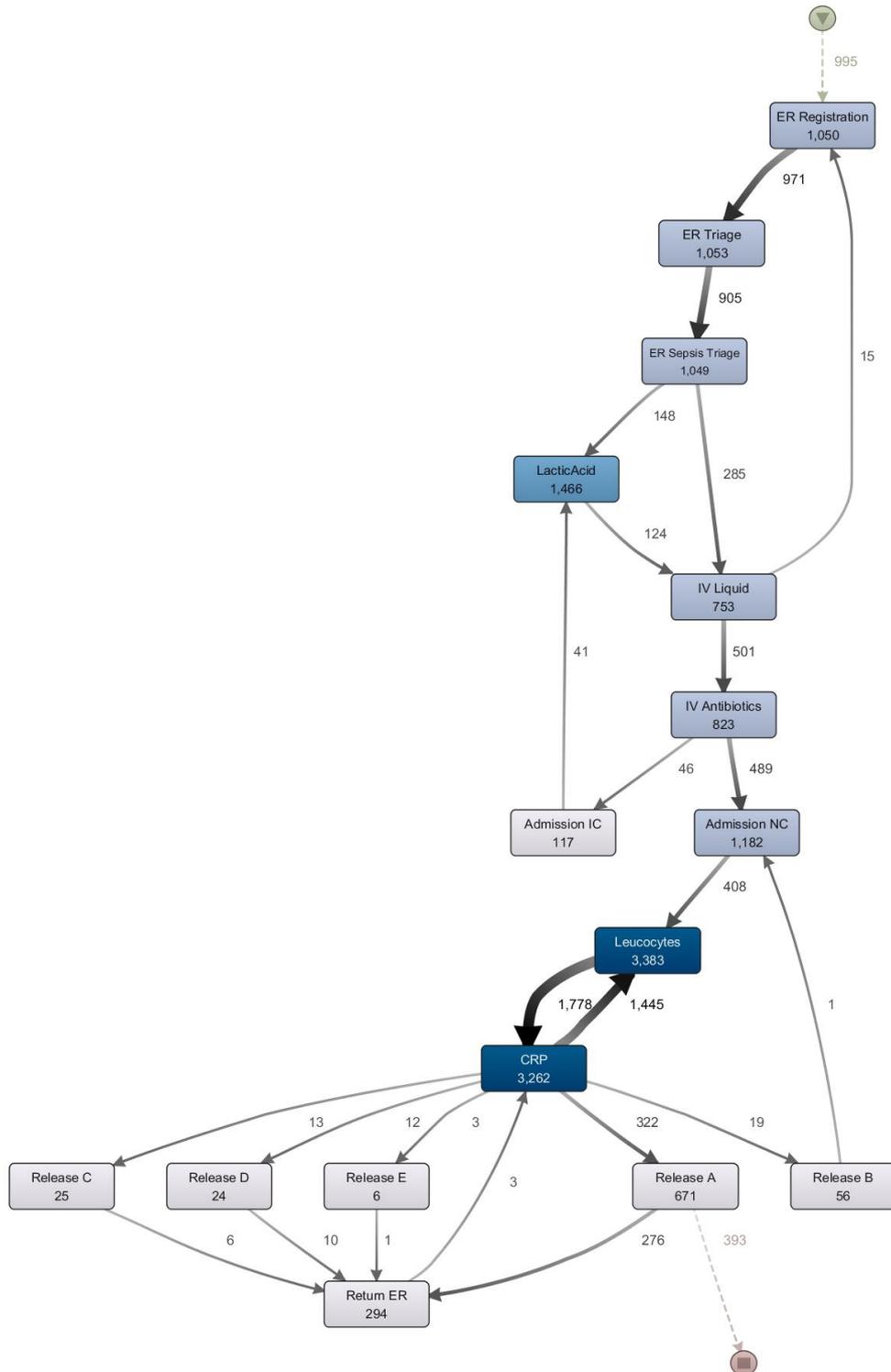


Fonte: O Autor.

usuário pode visualizar um modelo apenas com os nós que realizam as atividades mais relevantes de um processo, desde que a barra alcance o extremo superior. As demais atividades não exibidas nesse modelo são agrupadas em *clusters* (ROZINAT, 2010), com a indicação de quantas atividades estão contidas dentro deles, conforme Figura 4.3, onde *Cluster-19* contém as atividades *Admission IC*, *Admission NC* e *Release C*; *Cluster-20* contém as atividades *ER Triage*, *ER Sepsis Triage* e *Release A*; e *Cluster-21* contém as atividades *IV Liquid*, *IV Antibiotics* e *LacticAcid*. O filtro por arestas permite mostrar apenas as arestas mais significativas ou todas as existentes. Já o filtro de concorrência habilita/desabilita a exibição de relações conflitantes no modelo, sequencializando ou paralelizando essas relações.

A Figura 4.4 exibe o modelo de processos exportado em Disco. O detalhamento do modelo é ajustável por duas barras de rolagem: uma referindo-se às atividades e outra aos caminhos percorridos pelo processo. Os extremos inferiores das barras limitam a exibição do modelo apenas às atividades e/ou caminhos mais significativos do processo, enquanto, por sua vez, os superiores exibem todas as atividades e/ou todos os caminhos. Também é possível gerar o modelo contendo informações acerca da frequência em que ocorrem atividades e caminhos e/ou da *performance* em relação ao tempo em que elas executam dentro de um processo. A Figura 4.5 exibe um modelo contendo informações de tempo médio de *performance* e também de frequência absoluta. No que diz respeito à *performance*, também podem ser exibidos no modelo o tempo mínimo, máximo, mediano ou total de uma atividade ou caminho.

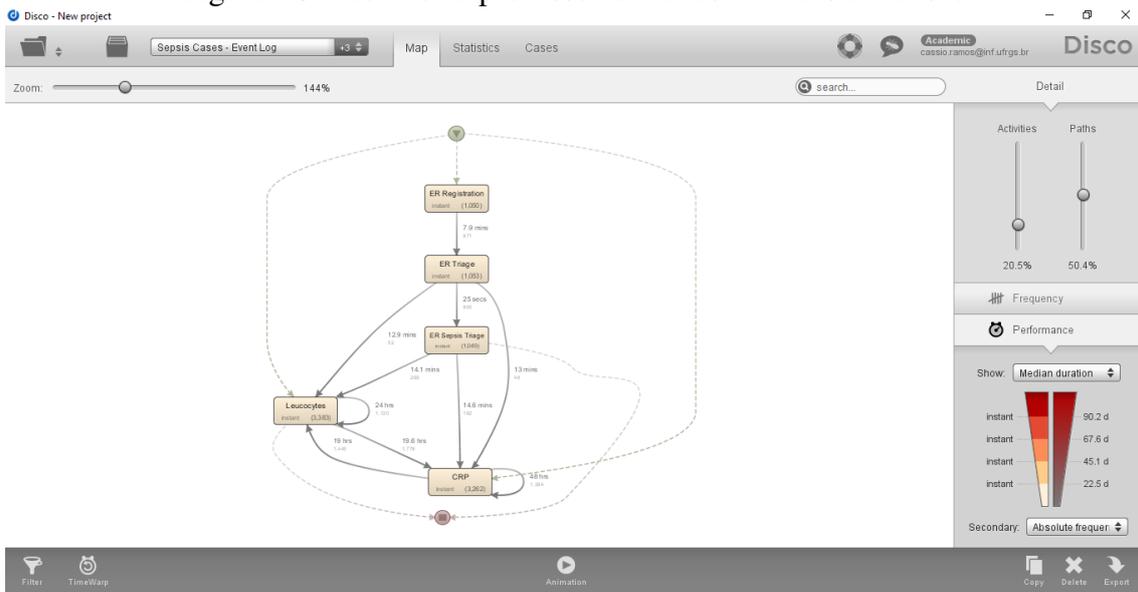
Figura 4.4: Modelo de processo em Disco



Fonte: O Autor.

Quanto à frequência, também pode ser exibida a frequência relativa de ocorrência de uma atividade ou caminho e o máximo de repetições ocorridas. Na Figura 4.6, selecio-

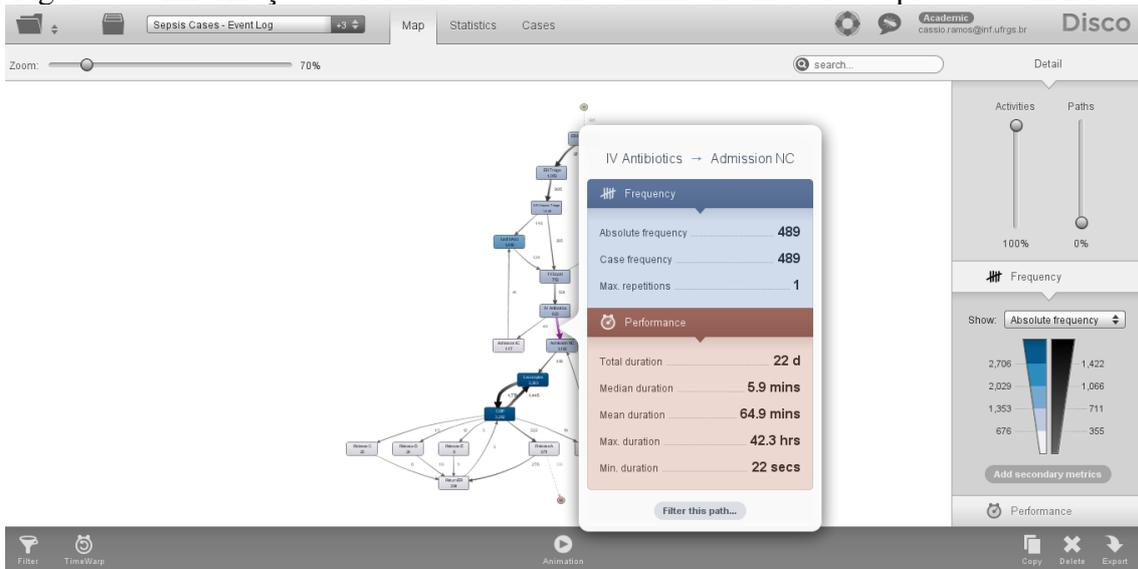
Figura 4.5: Modelo de processo com duas métricas em Disco



Fonte: O Autor.

nando uma atividade - ou um caminho - no modelo, são exibidas todas essas informações simultaneamente como, por exemplo, clicando no caminho *IV Antibiotics -> Admission NC*.

Figura 4.6: Informações adicionais sobre um caminho no modelo de processo em Disco



Fonte: O Autor.

4.2 Filtragem de dados

Como visto na Seção 3.1.1, existem algumas opções de filtragem de dados em ProM. Para fins de análise dos dados contidos no *log* de eventos estudado nesta Seção, foi utilizado o filtro de atributos de evento (*Filter on Event Attributes*). Na Figura 4.7 é possível visualizar o número total de casos e de eventos selecionando a aba *Dashboard*, além de um gráfico de eventos por caso, onde é possível deduzir que um caso contido no *log* de eventos estudado contém no mínimo 3 eventos e no máximo 185 eventos. Abaixo desse gráfico, é apresentado outro, de atividades por caso, onde é visto que um caso contido no *log* de eventos estudado contém entre 3 e 12 atividades. A aba *Inspector* apresenta uma análise por instância do processo, onde os eventos são representados por triângulos e os de cor verde correspondem aos eventos que ocorrem mais frequentemente, enquanto os de cor vermelha ocorrem com uma frequência menor, como na Figura 4.8.

Figura 4.7: Log de eventos filtrado por atributos de evento em ProM

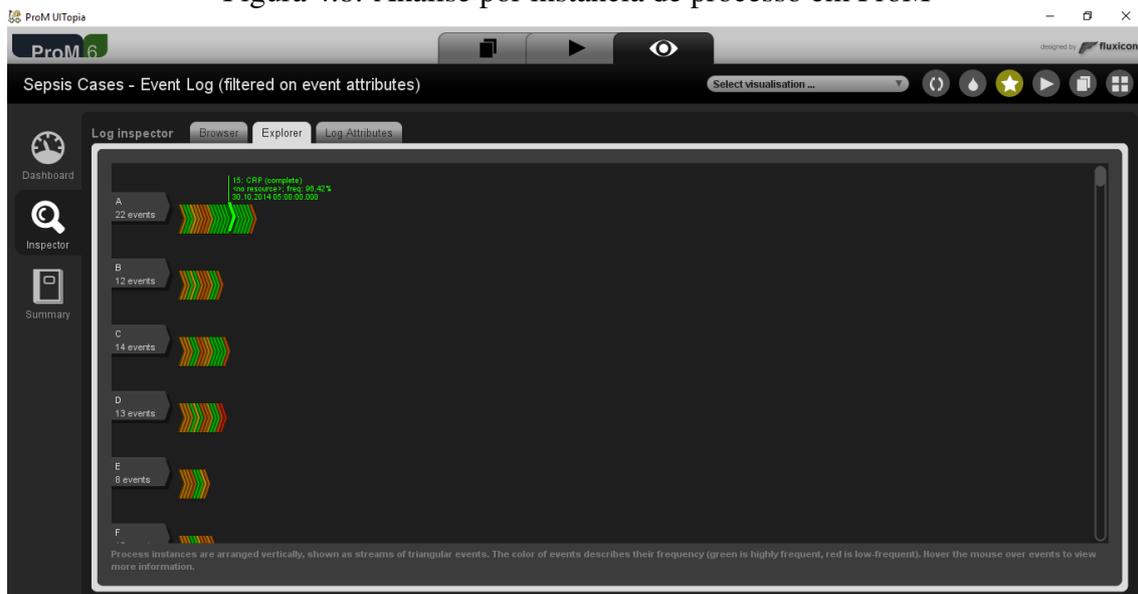


Fonte: O Autor.

Já a aba *Summary* apresenta dados gerais interpretados após a leitura do *log* como, por exemplo, o número de ocorrências de uma atividade, tanto em números absolutos quanto relativos (em porcentagem), como é visível na Figura 4.9, em que as atividades *Leucocytes* e *CRP*, com 3383 e 3262 ocorrências, respectivamente, constam na maioria dos casos, conforme informações contidas no *log* estudado.

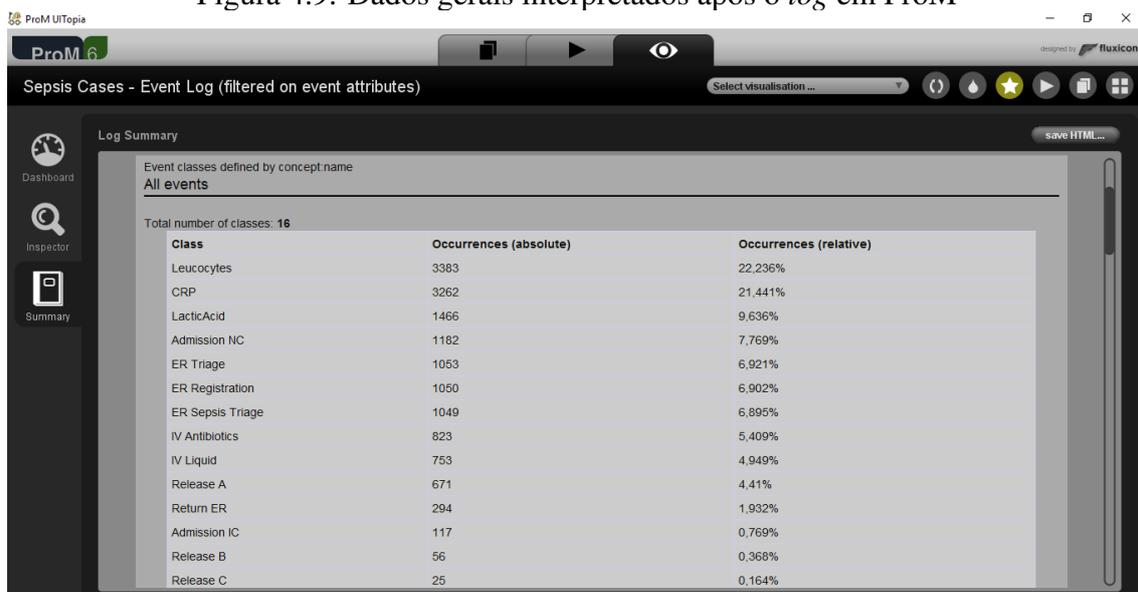
A filtragem de dados em Disco permite ao usuário diversas possibilidades de análise dos dados obtidos junto ao *log*, como visto na Seção 3.2.1. Na Figura 4.10, é possível visualizar informações gerais acerca do processo. Clicando em *Overview*, são exibidas

Figura 4.8: Análise por instância de processo em ProM



Fonte: O Autor.

Figura 4.9: Dados gerais interpretados após o log em ProM

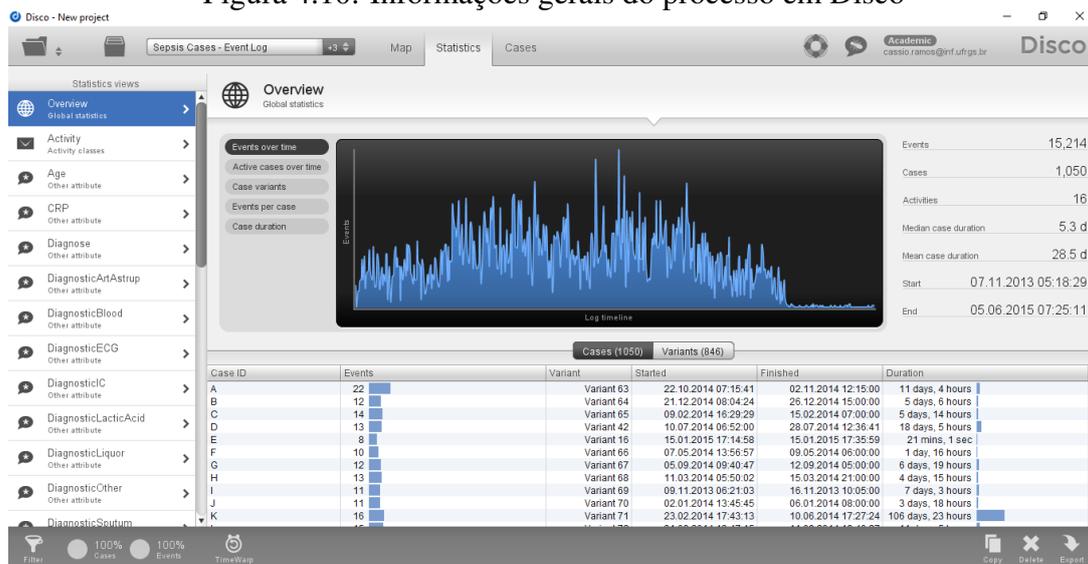


Fonte: O Autor.

estatísticas globais, como número total de eventos, de casos, de atividades, tempo médio de duração de um caso e primeiro e último *timestamps* registrados no *log*.

Na parte central da tela, a ferramenta traz cinco tipos de gráficos: eventos executados através do tempo, casos ativos em um período de tempo, variantes de caso, eventos por caso e duração de casos. Abaixo do gráfico escolhido, são exibidas as informações interpretadas para sua exibição, podendo o usuário optar entre casos ou variantes. Além de *Overview* pode-se selecionar a visualização de estatísticas por atividades ou por qual-

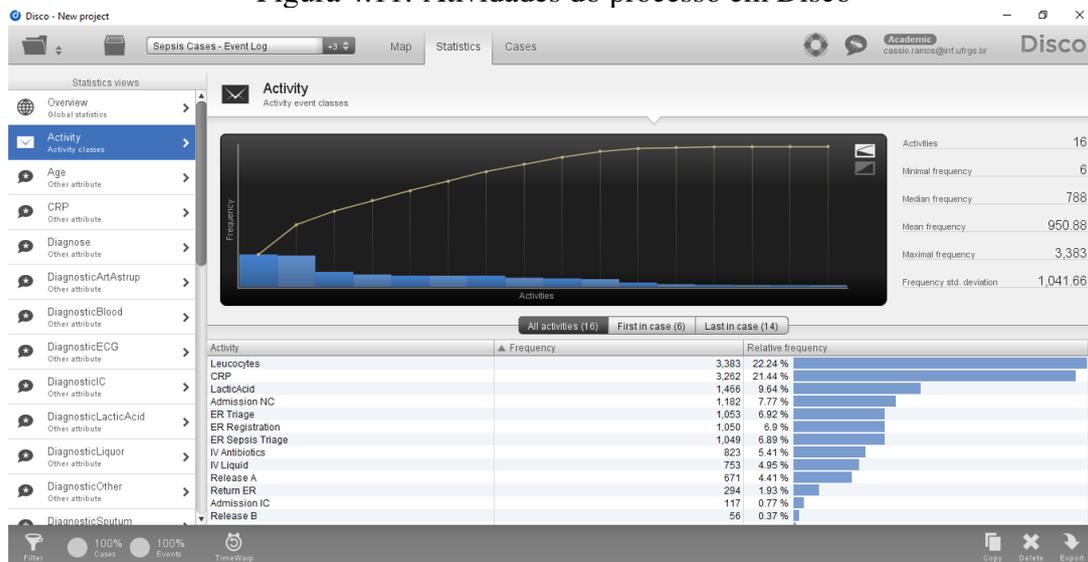
Figura 4.10: Informações gerais do processo em Disco



Fonte: O Autor.

quer das informações correspondentes às colunas no *log*, como idade ou diagnóstico, no exemplo do *log* de casos de sepse. A Figura 4.11 representa a análise por atividades e suas informações gerais, com o apoio de um gráfico que ordena de forma crescente ou decrescente as atividades mais frequentes encontradas na execução do processo, apresentando resultados semelhantes aos obtidos na filtragem de dados de ProM.

Figura 4.11: Atividades do processo em Disco

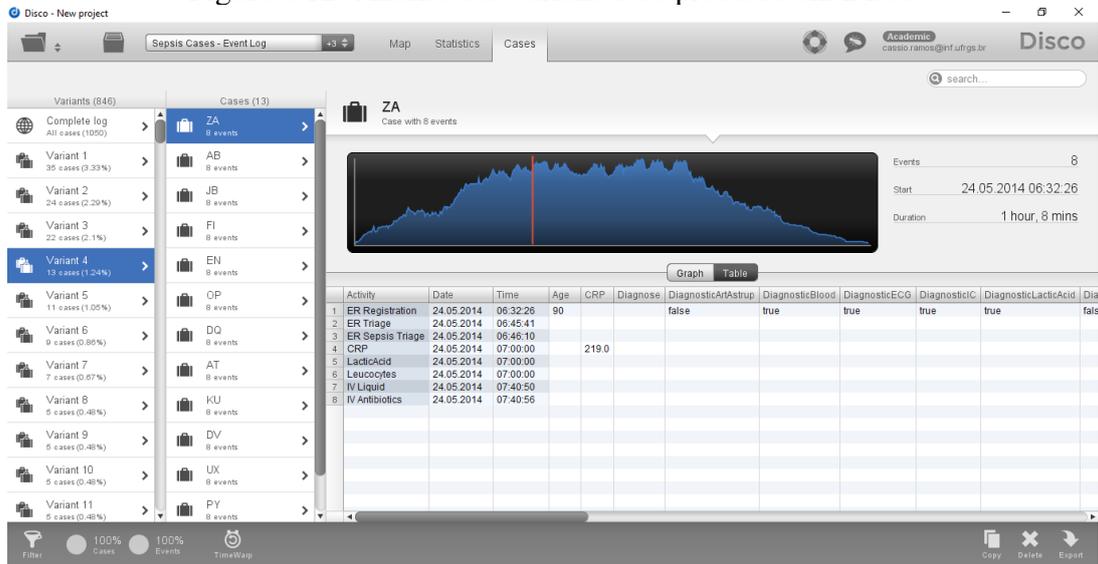


Fonte: O Autor.

A análise das variantes do processo em Disco pode ser feita selecionando qualquer um dos tipos de variantes encontradas na execução do processo, também com apoio gráfico. Por exemplo, analisando a Figura 4.12, a variante 4 contém 13 casos e, selecionando

a variante, têm-se à disposição o detalhamento de cada um desses casos. Considerando, como exemplo o caso ZA, que contém 8 eventos, é possível visualizar todos os eventos executados, tal como os valores em cada coluna interpretada e informações sobre início e duração do caso - o caso ZA tem duração de 1 hora e 8 minutos.

Figura 4.12: Análise das variantes do processo em Disco



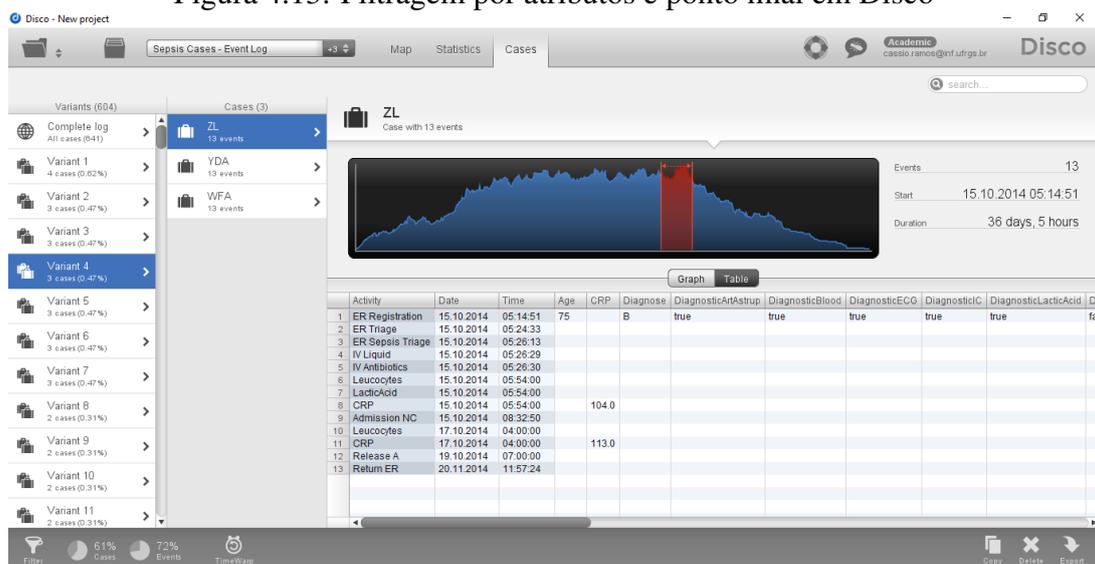
Fonte: O Autor.

A Seção 3.2.1 detalha os filtros existentes em Disco e suas características. Há a possibilidade da utilização de mais de um filtro para analisar casos de forma mais específica. A Figura 4.13 é resultado do uso de dois filtros: o filtro de ponto final (*endpoint filter*), usado para exibir apenas eventos cuja idade do paciente seja igual ou superior a 60 anos; e o filtro de atributos, usado para exibir apenas casos cujo término ocorra exclusivamente nas atividades *Release A*, *Release B*, *Release C*, *Release D*, *Release E* e *Return ER*. Foram filtrados 641 casos (61% do total) que contém 72% dos eventos realizados ao longo de todo o processo.

4.3 Checagem de conformidade

O *plug-in* disponível em ProM que faz a checagem de conformidade, como visto na Seção 3.1.1, requer como entrada o modelo de processos no formato de Redes de Petri. Para o estudo do *log* de casos de sepse, foi utilizando o *plug-in* de mineração indutiva para geração do modelo nesse formato. A checagem foi realizada com o auxílio do *plug-in Conformance Checking of DPN* (LEONI et al., 2014), que além do modelo em Redes de Petri recebe o próprio *log* como entrada.

Figura 4.13: Filtragem por atributos e ponto final em Disco



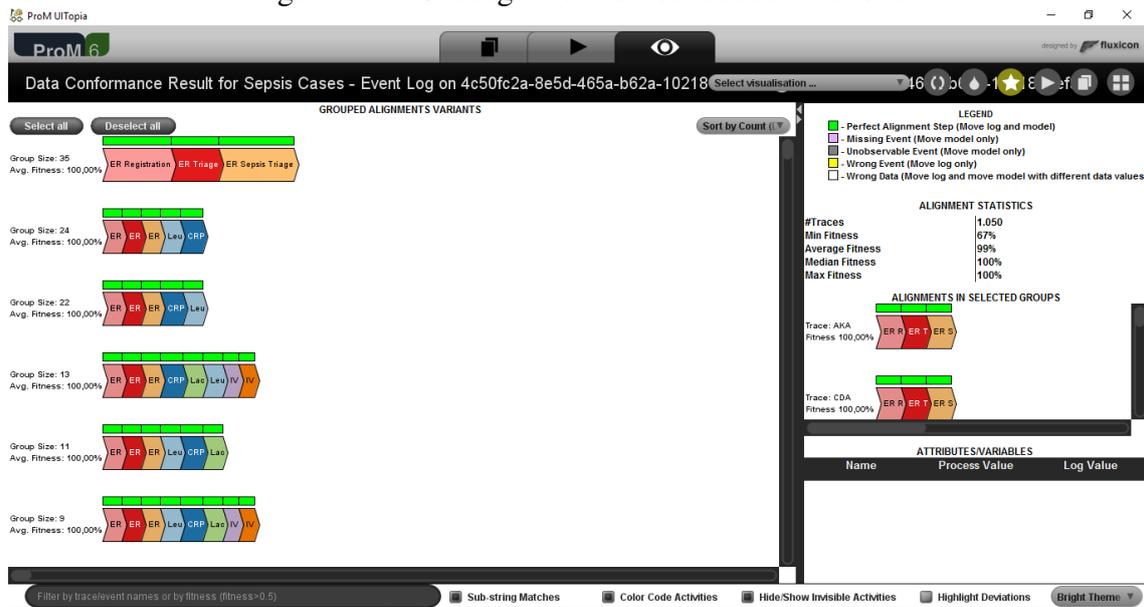
Fonte: O Autor.

O resultado da checagem de conformidade gera uma projeção de modelo de processos, bem como dados obtidos durante a fase de checagem como, por exemplo, uma análise dos caminhos executados e seu agrupamento conforme as variantes entre si, como pode ser visualizado na Figura 4.14, onde a exibição está ordenada por contagem de um grupo de atividades. Por exemplo, o caminho *ER Registration - ER Triage - ER Sepsis Triage* foi detectado 35 vezes em análise aos logs. Também é possível ordenar essas variantes por tamanho do caminho ou pela medida de *fitness* de um caminho. Ao lado direito dessa exibição, constam estatísticas gerais de conformidade do log, onde é mostrado que o *fitness* do log em conformidade com o modelo de processos gerado é de 99%.

4.4 Regras de decisão

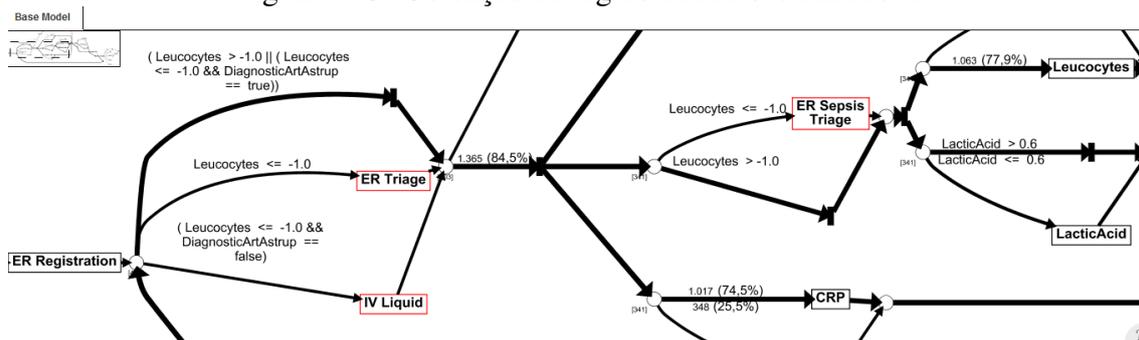
Conforme visto na Seção 3.1.1, a obtenção de regras de decisão em ProM é realizada de forma a encontrar pontos de decisão na execução de um processo. Para análise do log de casos de sepse (MANNHARDT, 2016), foi gerado neste trabalho um modelo de descoberta representado na Figura 4.15 contendo tanto regras referentes a atividades do processo quanto a valores de atributos representados pelas colunas do log. Na mesma figura, é possível visualizar regras referentes a valores numéricos contidos nas atividades *Leucocytes* e *LacticAcid*, bem como da coluna *DiagnosticArtAstrup*, com valor *booleano*.

Figura 4.14: Checagem de conformidade em ProM



Fonte: O Autor.

Figura 4.15: Obtenção de regras de decisão em ProM



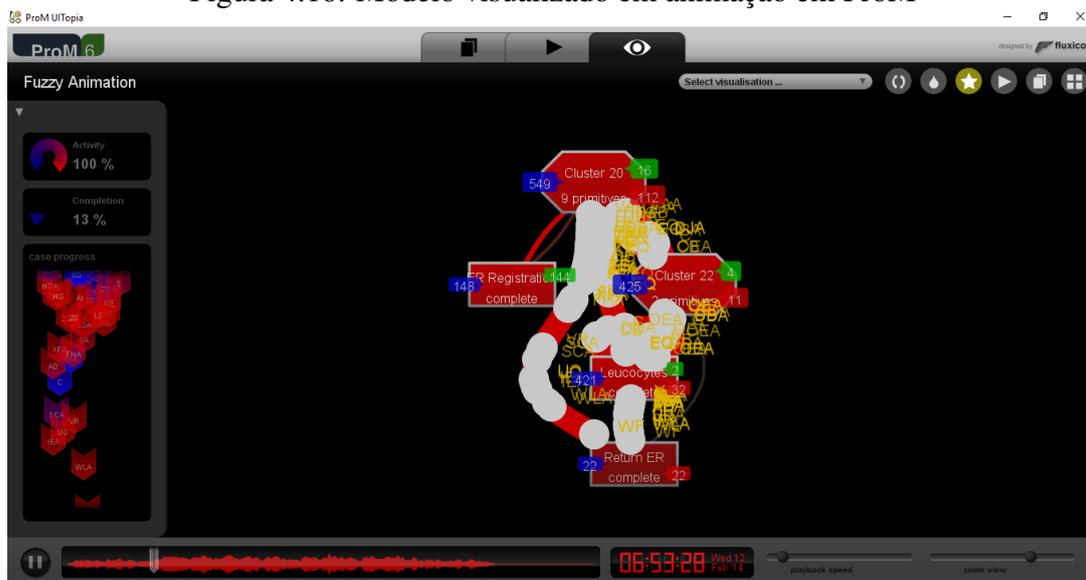
Fonte: O Autor.

4.5 Visualização de processos

Ambas as ferramentas estudadas neste Capítulo disponibilizam ao usuário a funcionalidade da visualização de processos em formato de animação, simulando uma execução. A animação exibida tem como principal objetivo a detecção de gargalos (quando o processamento dos dados de entrada em um ponto do processo prejudicam seu andamento normal) e outros possíveis impedimentos ao andamento normal de um processo. Tanto ProM quanto Disco exibem essa animação em um modelo *fuzzy*.

ProM possui um *plug-in* responsável por essa funcionalidade que devolve ao usuário um modelo *fuzzy* completo do modelo de processos ou agrupado (em *clusters*), como ocorrido no exemplo de casos de sepsis estudado nesta Seção. Além do modelo *clusterizado*, a Figura 4.16 exibe uma medida de porcentagem de atividades e de casos executados

Figura 4.16: Modelo visualizado em animação em ProM



Fonte: O Autor.

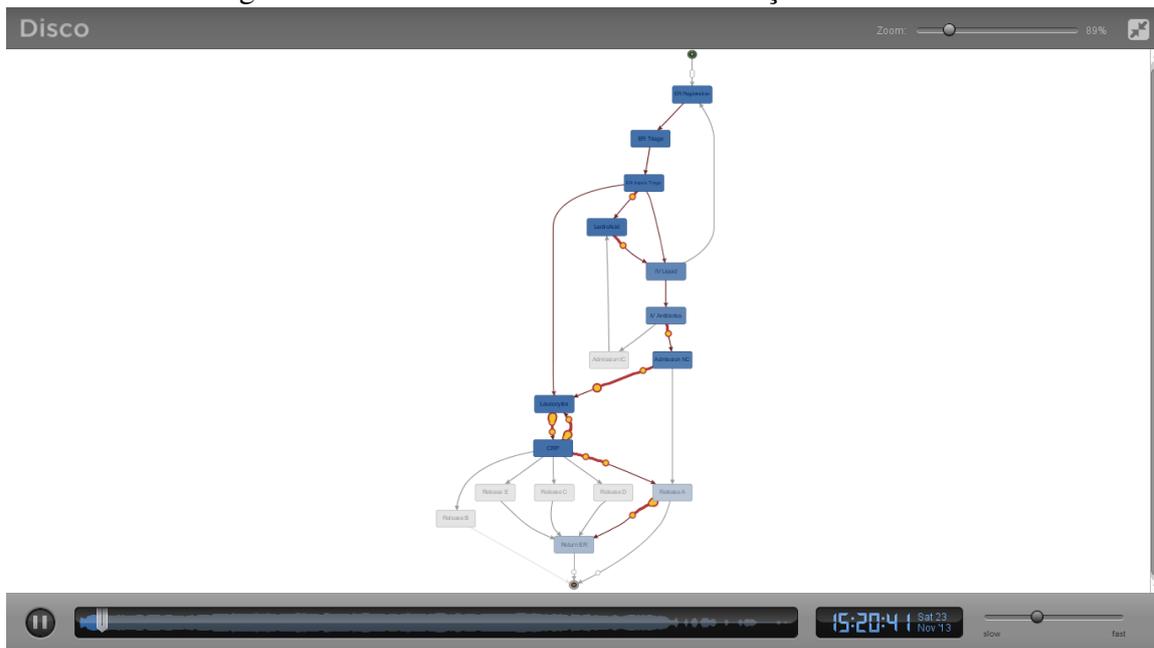
na exibição, o caminho que os casos percorrem, uma progressão de tempo considerando os *timesteps* contidos nos eventos, uma barra para aumentar ou diminuir a velocidade de exibição da animação e, finalmente, uma barra de *zoom* para visualização do modelo.

A Figura 4.17 mostra a visualização do processo de casos de sepse em Disco. Exibe o caminho que os casos percorrem, sendo representados por círculos de cor amarela. Na parte superior da tela, há uma barra de *zoom* que permite aproximar a visualização ou torná-la mais abrangente. Na parte inferior, exibe a progressão de tempo considerando os *timesteps* dos eventos e uma barra que torna mais rápida ou mais devagar a animação.

4.6 Análise de performance

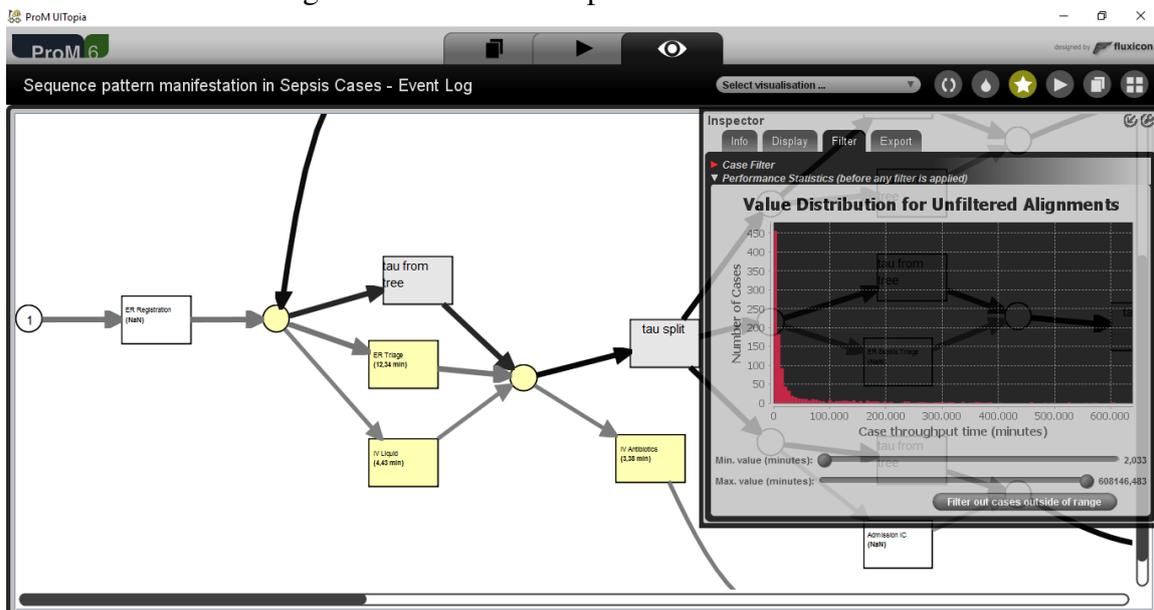
O *plug-in* responsável pela análise de performance em ProM exibe informações como tempo médio de execução de uma certa atividade e um gráfico com distribuição de valores, que mostra a razão entre o tempo médio de um caso pelo número de casos, conforme demonstrado na Figura 4.18, onde são mostrados os tempos médios das atividades *ER Triage* (12,34 minutos), *IV Liquid* (4,43 minutos) e *IV Antibiotics* (3,38 minutos). As atividades cujo tempo de execução seja desprezível, não impactando no tempo final de execução do processo, são identificadas com a descrição NaN (*Not a Number*), uma codificação usada pela IEEE (*Institute of Electrical and Electronics Engineers*) para representar números não-ordinários (ORUP, 1999).

Figura 4.17: Modelo visualizado em animação em Disco



Fonte: O Autor.

Figura 4.18: Análise de performance em ProM



Fonte: O Autor.

4.7 Análise dos resultados obtidos

Nas subseções 3.1.1 e 3.1.4 é visto que ProM procura abordar todas as características contidas nas fases da mineração de processos de negócio. Isto fica ainda mais claro após a realização do estudo dos casos de sepsis, que permite uma análise detalhada do processo sob a ótica de todas as fases da mineração. A fase de descoberta é abor-

dada na Seção 4.1 com a obtenção do modelo de descoberta e a possível filtragem de dados posterior. A fase de checagem de conformidade mostrou-se robusta, permitindo diferentes tipos de análises da conformidade do modelo. Também em relação a fase de conformidade, a ferramenta realiza mineração de regras de decisão. Já a presença da fase de aprimoramento é vista na possibilidade da visualização de processos, de análises de rede social e de performance.

Disco é uma ferramenta que não aborda todas as fases da mineração de processos, no entanto, permitiu estudar as características contidas no *log* de casos de sepsis de forma bastante detalhada, com o auxílio de sua filtragem de dados, no que diz respeito à fase de descoberta. A fase de aprimoramento é notada na visualização de uma simulação de execução de um processo, assim como na análise de performance desse processo.

Tabela 4.1: Tabela comparativa entre ferramentas estudadas no Capítulo 4

| <i>Características</i> | <i>ProM</i> | <i>Disco</i> |
|------------------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| Formato de importação suportado | MXML, XES | CSV, MXML, MXML.GZ, XES, XES.GZ, XLS, XLSX, FXL, DSC |
| Modelo de exportação | BPMN, Redes de Petri, sistema de transição, modelos <i>fuzzy</i> , heurístico e minerado indutivo | Modelo <i>fuzzy</i> |
| Capacidade do <i>log</i> importado | Ilimitada | 100 eventos (versão demo), 5 milhões de eventos (completa) |
| Plataforma de execução | Máquina do usuário (<i>stand-alone</i>) | Máquina do usuário (<i>stand-alone</i>) |
| Filtragem de dados | Sim | Sim |
| Descoberta de processos | Sim | Sim |
| Checagem de conformidade | Sim | Não |
| Mineração de regras de decisão | Sim | Não |
| Mineração de rede social | Sim | Não |
| Visualização de processos | Sim | Sim |
| Análise de performance | Sim | Sim |

Fonte: O Autor.

Ambas as ferramentas permitiram exibir o modelo de processos com diferentes níveis de detalhamento e filtrar os dados extraindo todo o conhecimento possível do *log* de casos de sepsis. Disco, contudo, apresenta os dados em uma interface mais familiar,

de mais fácil entendimento para um usuário iniciante na área de mineração de processos, bem como permitiu uma análise mais profunda dos dados contidos no *log*. Por outro lado, ProM é uma ferramenta mais completa, extraindo também dados sobre conformidade e regras de decisão. A Tabela comparativa 4.1 resume as informações trazidas nesta Seção.

4.8 Resumo do capítulo

Neste capítulo, é realizado um estudo de caso envolvendo as ferramentas ProM e Disco, representando uma de cada categoria estudada no Capítulo 3. Consiste em interpretar um *log* de eventos com informações reais, contendo casos de sepse em um hospital, em ambas as ferramentas, de forma a estabelecer um comparativo entre elas tendo como base diversas características presentes (ou não) em ambas. As características estudadas são a de modelo de descoberta, filtragem de dados, checagem de conformidade, mineração de regras de decisão, visualização de processos e análise de performance. A partir da obtenção dos resultados em cada ferramenta, é possível comparar suas diferenças e traçar semelhanças entre elas, com o apoio de uma tabela comparativa.

5 CONCLUSÃO

O presente trabalho comparou as principais ferramentas de mineração de processos de negócio existentes nos âmbitos acadêmico e industrial, classificando-as em duas categorias: *open-source* e comerciais/proprietárias. A mineração de processos é uma área com domínio de pesquisa ainda emergente, como visto na Seção referente aos trabalhos relacionados a este estudo, e as ferramentas de mineração vieram ao mercado para ampliar o entendimento sobre a área, abordando todas – ou praticamente todas – as características de cada fase da mineração de processos.

As ferramentas *open-source* detalhadas neste estudo possuem algumas diferenças entre elas no que diz respeito ao número de funcionalidades implementadas e nos formatos de modelos e *logs* que interpretam. No entanto, se assemelham no fato de que são ferramentas desenvolvidas no ambiente acadêmico e que buscam ao máximo seguir as características das fases de mineração.

As quatro ferramentas comerciais comparadas neste trabalho, apesar das diferenças em relação ao número de funcionalidades e em tipos de formatos de modelos e *logs* pelos quais processam os dados que recebem como entrada, são similares no que diz respeito às suas interfaces, mais simples para usuários iniciantes na área de mineração. Apesar de não possuírem tantas funcionalidades quanto as ferramentas *open-source*, trazem resultados bastante satisfatórios ao que se propõem.

Este trabalho apresenta algumas limitações no que se refere à comparação entre ferramentas de diferentes categorias: enquanto as ferramentas *open-source* têm suas licenças de uso gratuitas, as ferramentas comerciais disponibilizam apenas versões *demo* (algumas não apresentando a totalidade das funcionalidades disponíveis) ou licenças de custo financeiro alto. Essa situação pode prejudicar uma comparação mais direta entre as categorias, assim como dificultar a utilização em um estudo de caso das ferramentas que apresentem essa limitação, como o estudo de caso apresentado no Capítulo 4. Outras limitações do trabalho dizem respeito à análise e comparação das ferramentas estudadas nos Capítulos 3 e 4, que são realizadas por uma única pessoa, e também dizem respeito às informações sobre os algoritmos suportados por cada ferramenta, pois nem todas as ferramentas estudadas neste trabalho disponibilizam essa informação para indivíduos que não estejam envolvidos em seu desenvolvimento.

O estudo de caso apresentado no Capítulo 4 evidencia a diferença principal entre ferramentas acadêmicas/*open-source* e industriais/comerciais, ou seja, ter à disposição to-

das as funcionalidades possíveis representando todas as fases da mineração de processos, ou priorizar uma interface mais simples e um tratamento de dados mais robusto. Essa questão pode definir o perfil do usuário – iniciante ou avançado – de mineração de processos, assim como a ferramenta mais apropriada para seu objetivo final.

Como trabalho futuro, sugere-se o desenvolvimento das características de outras ferramentas de mineração de processos existentes no mercado, possibilitando mais opções de comparação entre ferramentas. Também é sugerida a comparação entre ferramentas organizando-as em outras categorias como, por exemplo, em plataformas de execução (*stand-alone* ou nuvem), a comparação entre ferramentas estabelecendo o foco em alguma característica específica (filtragem de dados, modelo de descoberta, conformidade, etc.) ou em alguma fase específica da mineração de processos, ou a comparação entre os algoritmos suportados por cada ferramenta a ser analisada. Também sugere-se nessas comparações o envolvimento de mais pessoas, que estejam inseridas na área de mineração de processos, por meio de questionários – onde podem ser extraídas as necessidades desses usuários e o que esperam de ferramentas de mineração – ou tarefas delegadas a esses usuários, para que eles executem essas tarefas nas ferramentas. Finalmente, também é possível a realização de uma análise comparativa de desempenho entre ferramentas de mineração, levando em conta a presença de *bugs* e seus impactos na execução, assim como a frequência em que essas ferramentas de mineração de processos de negócio são atualizadas.

REFERÊNCIAS

- AALST, W. M. Van der. Using process mining to bridge the gap between bi and bpm. **IEEE Computer**, v. 44, n. 12, p. 77–80, 2011.
- AALST, W. M. van der. Process discovery from event data: Relating models and logs through abstractions. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 3, p. e1244, 2018.
- AALST, W. M. Van der et al. Prom: The process mining toolkit. **BPM (Demos)**, v. 489, n. 31, p. 2, 2009.
- AALST, W. V. D. **Process mining: discovery, conformance and enhancement of business processes**. [S.l.]: Springer, 2011.
- AALST, W. V. D. et al. Process mining manifesto. In: SPRINGER. **International Conference on Business Process Management**. [S.l.], 2011. p. 169–194.
- AALST, W. Van der; WEIJTERS, T.; MARUSTER, L. Workflow mining: Discovering process models from event logs. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 16, n. 9, p. 1128–1142, 2004.
- AGARWAL NEHA E SINGH, L. Process mining tools: A comparative analysis and review. **Advances in Computer Science and Information Technology (ACSIT)**, v. 1, n. 2, p. 26–29, 2014.
- AILENEI, I.-M. et al. Process mining tools: A comparative analysis. Citeseer, 2011.
- CELIK UFUK E AKÇETIN, E. Process mining tools comparison. v. 9, p. 97–104, 11 2018.
- CELONIS. **Celonis - Process Mining: How Does It Work?** 2019. Available from Internet: <<https://www.celonis.com/process-mining/how-does-process-mining-work>>. Accessed in: 22 mai. 2019.
- CLAES JAN E BRU, F. The perceived quality of process discovery tools. 2018.
- CONFORTI, R. et al. Analysis of business process variants in apromore. In: **Proceedings of the Demo Track of the 13th International Conference on Business Process Management (BPM'15)**. [S.l.: s.n.], 2015.
- DEVI, A. T. An informative and comparative study of process mining tools. **Int. J. Sci. Eng. Res**, v. 8, n. 5, p. 8–10, 2006.
- DONGEN, B. F. V. et al. The prom framework: A new era in process mining tool support. In: SPRINGER. **International conference on application and theory of petri nets**. [S.l.], 2005. p. 444–454.
- DUMAS, M. et al. **Fundamentals of business process management**. [S.l.]: Springer, 2013.
- EINDHOVEN UNIVERSITY OF TECHNOLOGY. **Process Mining.org**. 2019. Available from Internet: <<http://www.processmining.org/prom/start>>. Accessed in: 23 mar. 2019.

- EINDHOVEN UNIVERSITY OF TECHNOLOGY. **ProM - Process Mining Workbench**. 2019. Available from Internet: <<http://www.promtools.org/doku.php>>. Accessed in: 23 mar. 2019.
- FLUXICON. **Process Mining and Automated Process Discovery Software for Professionals - Disco**. 2019. Available from Internet: <<https://fluxicon.com/disco/>>. Accessed in: 29 mai. 2019.
- FORNARI, F. et al. Checking business process modeling guidelines in apromore. In: **BPM (Demos)**. [S.l.: s.n.], 2017.
- GERT JANSSENSWILLEN. **Gert Janssenswillen (GitHub)**. 2019. Available from Internet: <<https://github.com/gertjanssenswillen>>. Accessed in: 20 mai. 2019.
- GROEFSEMA, H.; BEEST, N. van; ARMAS-CERVANTES, A. Automated compliance verification of business processes in apromore. In: **BPM (Demos)**. [S.l.: s.n.], 2017.
- GÜNTHER, C. W.; ROZINAT, A. Disco: Discover your processes. **BPM (Demos)**, Citeseer, v. 940, p. 40–44, 2012.
- HASSELT UNIVERSITY. **bupaR: Business Process Analytics in R**. 2019. Available from Internet: <<http://bupar.net/index.html>>. Accessed in: 19 mai. 2019.
- JANSSENSWILLEN, G.; DEPAIRE, B. Bupar: business process analysis in r. 2017.
- JANSSENSWILLEN, G. et al. bupar: Enabling reproducible business process analysis. **Knowledge-Based Systems**, Elsevier, v. 163, p. 927–930, 2019.
- KEBEDE, M. Comparative evaluation of process mining tools. **University of Tartu**, 2015.
- LEONI, M. D. et al. Decomposing conformance checking on petri nets with data. **BPM Center Report BPM-14-06**, 2014.
- MANNHARDT, F. **Sepsis cases-event log. Eindhoven University of Technology (2016)**. 2016.
- ORUP, H. **Floating point NaN comparison**. [S.l.]: Google Patents, 1999. US Patent 5,931,943.
- PESTANA, L. **Aplicação do Process Mining na Auditoria de Processos Governamentais**. Dissertation (Master) — Universidade Federal do Espírito Santo, 2017.
- QPR. **QPR Process Analyzer Wiki**. 2019. Available from Internet: <https://devnet.onqpr.com/pawiki/index.php/QPR_ProcessAnalyzer_Wiki>. Accessed in: 23 mai. 2019.
- ROSA, M. L. et al. Apromore: An advanced process model repository. **Expert Systems with Applications**, Elsevier, v. 38, n. 6, p. 7029–7040, 2011.
- ROZINAT, A. Process mining: conformance and extension. Technische Universiteit Eindhoven, 2010.

ROZINAT, A.; AALST, W. M. van der. Decision mining in prom. In: SPRINGER. **International Conference on Business Process Management**. [S.l.], 2006. p. 420–425.

SINGER, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). **Jama**, American Medical Association, v. 315, n. 8, p. 801–810, 2016.

SOFTWARE AG. **ARIS PPM - Process Mining Software, Process Inteligente, Software AG**. 2019. Available from Internet: <https://www.softwareag.com/corporate/products/process/process_mining/default.html>. Accessed in: 24 mai. 2019.

THE UNIVERSITY OF MELBOURNE. **Apromore - Advanced Process Analytics Platform**. 2019. Available from Internet: <<http://apromore.org/>>. Accessed in: 10 mai. 2019.

TURNER, C. J. et al. Process mining: from theory to practice. **Business Process Management Journal**, Emerald Group Publishing Limited, v. 18, n. 3, p. 493–512, 2012.

VERENICH, I. et al. Predictive process monitoring in apromore. In: SPRINGER. **International Conference on Advanced Information Systems Engineering**. [S.l.], 2018. p. 244–253.

VERSTRAETE, D. **Process mining in practice: Comparative study of process mining software**. Thesis (PhD) — MS thesis, Faculty of Economics and Business Administration, Ghent . . . , 2014.

WANG, M.; SUN, Z. **Handbook of Research on Complex Dynamic Process Management: Techniques for Adaptability in Turbulent Environment**. [S.l.: s.n.], 2010.