UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIA DA COMPUTAÇÃO

RAPHAEL JOSÉ D'CASTRO

PRÉ-PROCESSAMENTO PARA MINERAÇÃO DE PROCESSOS:

TÉCNICAS PARA SIMPLIFICAÇÃO AUTOMÁTICA DE LOGS DE EVENTOS

Recife
2020

RAPHAEL JOSÉ D'CASTRO

PRÉ-PROCESSAMENTO PARA MINERAÇÃO DE PROCESSOS:

TÉCNICAS PARA SIMPLIFICAÇÃO AUTOMÁTICA DE LOGS DE EVENTOS

Tese apresentada ao Programa de Pós-Graduação em Ciências da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciências da Computação.

Área de concentração: Inteligência Computacional.

Orientador: Prof. Dr. Adriano Lorena Inácio de Oliveira.

Coorientador: Prof. Dr. Ricardo Massa Ferreira Lima.

Recife

2020

Catalogação na fonte
Bibliotecária Mariana de Souza Alves CRB4-2105

D277p D'Castro, Raphael José
Pré-processamento para mineração de processos: técnicas para simplificação automática de logs de eventos/ Raphael José D'Castro. – 2020.
189f., il., fig., tab.

Orientador: Adriano Lorena Inácio de Oliveira.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2020.
Inclui referências e apêndices.

1. Inteligência Computacional. 2. Mineração de processos. 3. Pré-processamento de logs de eventos. 4. Simplificação de Log de eventos. I. Oliveira, Adriano Lorena Inácio de. (orientador) II. Título.

006.31 CDD (22. ed.) UFPE-CCEN 2021-36

Raphael José D'Castro

"Pré-processamento para Mineração de Processos: Técnicas para Simplificação Automática de Logs de Eventos"

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Aprovado em: 04/11/2020.

_____

Orientador: Prof. Dr. Adriano Lorena Inácio de Oliveira

BANCA EXAMINADORA

_____

Prof. Dr. Cleber Zanchettin
Centro de Informática / UFPE

_____

Profa. Dra. Cristine Martins Gomes de Gusmão
Departamento de Engenharia Biomédica / UFPE

_____

Prof. Dr. George Augusto Valença Santos
Departamento de Computação / UFRPE

_____

Prof. Dr. Byron Leite Dantas Bezerra
Escola Politécnica de Pernambuco / UPE

_____

Prof. Dr. Edson Emílio Scalabrin
Departamento de Informática / PUC/PR

Ao meu pai, *“In Memoriam”*, minha mãe, minha esposa e ao meu filho que, com muito carinho e apoio, possibilitaram que chegasse até esta etapa de minha vida.

AGRADECIMENTOS

Ao final dessa jornada, posso olhar para trás e enxergar com clareza as diversas lições aprendidas durante a realização deste doutorado. Nesse caminho, repleto de incertezas e obstáculos, pude contar com o incentivo e apoio de várias pessoas.

Em primeiro lugar, agradeço a meus pais pelos esforços e carinho empreendidos na minha educação, criando as condições para que eu chegasse até aqui.

Agradeço aos meus orientadores Adriano Lorena Inácio de Oliveira e Ricardo Massa Ferreira Lima pela oportunidade, apoio e confiança em mim depositada.

Agradeço aos meus colegas do TJPE pelo incentivo e apoio valioso nos momentos críticos para que eu pudesse me dedicar a este trabalho. Em especial, à Juliana Neiva, pela amizade, suporte incondicional e pelas longas conversas que influenciaram importantes decisões que permitiram a continuidade desse trabalho.

Agradecemos ao Centro de Informática da Universidade Federal de Pernambuco pelo apoio institucional e por proporcionar um ambiente adequado para o desenvolvimento da pesquisa.

Gostaria de agradecer ao magistrado Bráulio Gusmão pelo entusiasmo com que abraçou o tema, revigorando o ânimo da pesquisa, bem como criando as condições para que esse trabalho “ganhe vida” e possa contribuir com a melhoria das instituições públicas.

Agradeço especialmente à Marina Tôrres Leal D’Castro, meu amor, minha esposa, minha parceira e amiga, pela paciência, colaboração e suporte. Estar ao lado dela me deu forças para seguir em frente e não desistir.

Por fim, agradeço ao meu filho, Vinícius Leal D’Castro, que já em suas primeiras semanas de vida estava ao meu lado em madrugadas produtivas e felizes. Desde então, preenche minha vida com alegria, energia e esperança.

RESUMO

Nos últimos anos, desenvolveu-se uma disciplina denominada mineração de processos, cujo objetivo é ajudar a descobrir e analisar processos de negócios através da exploração de informações não triviais oriundas de sistemas de informações (logs de eventos). A mineração de processos vem avançando rapidamente, mas ainda existem importantes desafios que influenciam no seu resultado. Um dos desafios é lidar com diferentes níveis de granularidade nos eventos registrados. Em certas circunstâncias, deseja-se visualizar o processo em um nível de abstração diferente do modelado. Contudo, no atual cenário da mineração de processos, essa transformação precisa ser realizada no pré-processamento dos dados, sendo custosa e exigindo conhecimento abrangente sobre o negócio. Conduzimos estudos exploratórios que nos permitiram identificar meios de alterar a granularidade dos eventos e propusemos duas abordagens para este fim. As abordagens propostas transformam automaticamente os logs de eventos através do agrupamento de eventos de atividades afins. Com isso, proporcionam uma visão alternativa para o processo e uma significativa redução de tamanho dos logs de eventos. Outro problema tratado nesta tese diz respeito à incidência de atividades recorrentes nos logs de eventos. Estas atividades ocorrem em diversas fases do processo (contextos de negócio), propiciando modelos mais difíceis de interpretar. Propusemos uma abordagem de pré-processamento de log de eventos que identifica e trata as atividades recorrentes. O tratamento consiste em desmembrar as atividades recorrentes a partir dos contextos de negócios aos quais as atividades estão vinculadas. Os estudos conduzidos mostraram que os modelos de processos descoberto a partir de logs de eventos transformados pelas abordagens propostas apresentaram qualidade superior no tocante ao *fitness*, bem como em algumas medidas de simplicidade. Cabe ressaltar que todas as abordagens dessa tese de doutorado foram avaliadas através de estudos com logs de eventos reais.

Palavras-chave: Mineração de processos. Pré-processamento de logs de eventos. Simplificação de Log de eventos.

ABSTRACT

Process mining is an emerging discipline that helps discover and analyze business processes by exploring non-trivial information from information systems records (event logs). Process mining is advancing rapidly, but there are still critical challenges that influence its results. One of the challenges is dealing with different levels of granularity in registered events. In certain circumstances, we want to see the process at a different abstraction level than in the process modeled. However, in the current process mining scenario, this transformation must occur in the data pre-processing, being costly, and requiring knowledge about the business. Exploratory studies allowed us to identify ways to change the granularity of events and proposed two approaches for this purpose. Our techniques automatically transform event logs by grouping events from related activities, providing an alternative view of the process and a significant reduction in the event logs' size. Another problem addressed in this thesis concerns the incidence of recurring activities in the event logs. These activities occur at different stages of the process (business contexts), providing models that are more difficult to interpret. We proposed an event log pre-processing approach that identifies and handles recurring activities. The treatment consists of breaking up recurring activities from the business contexts to which the activities are linked. The studies showed that the processes models discovered from event logs transformed by the approaches proposed have superior fitness and better simplicity measures. The studies and evaluations performed in this Ph.D. thesis used real event logs.

Keywords: Process mining. Event logs pre-processing. Event log simplification.

LISTA DE FIGURAS

LISTA DE ABREVIATURAS E SIGLAS

BPM	*Business Process Management*
BPMN	*Business Process Model and Notation*
BPMS	*Business Process Management System*
CIS	*Computational Intelligence Society*
CMMN	*Case Management Modelling and Notation*
CNJ	Conselho Nacional de Justiça
CSV	*Comma-separated values*
DCR-Graphs	*Declare, Dynamic Condition Response-Graphs*
DMTC	*Data Mining Technical Committee*
DPIL	*Declarative Process Intermediate Language*
ECaM	*Extended Cardoso Metric*
ECyM	*Extended Cyclomatic Metric*
EPC	*Driven Process Chain*
ERP	*Enterprise Resource Planning*
ETL	*Extract, Transform and Load*
IEEE	*Instituto de Engenheiros Elétricos e Eletrônicos*
MXML	*MXML - Mining eXtensible Markup Language*
OMG	*Object Management Group*
PAIS	*Process-aware Information systems*
PJe	Processo Judicial Eletrônico
PM4Py	*Python Mining for Python*
PMS	*Process management system*
UML	*Unified Modeling Language*
WfMS	*Workflow Management Systems*
XES	*eXtensible Event Stream*
XML	*Extensible Markup Language*
YAWL	*Yet Another Workflow Language*

SUMÁRIO

1 INTRODUÇÃO

Vivemos em um mundo que está se tornando mais complexo e as transformações vêm ocorrendo mais rapidamente (TAIT, 2019). Segundo Hannes Apitzsch, executivo da Siemens AG, a afirmação que o mundo está se tornando cada vez mais complexo se tornou um mantra ouvido diariamente (BOENNER, 2020). Essa percepção não é novidade no mundo corporativo, nem tão pouco se restringe ao universo das instituições privadas. Organizações governamentais também enfrentam cotidianamente o desafio de transformar o serviço público, tornando-o cada vez mais ágil e eficiente. Nesse contexto, a gestão dos processos de negócios assume um papel de destaque nas organizações e, na última década, têm recebido uma atenção considerável dos pesquisadores (DAVIS; BRABANDER, 2007; DUMAS; VAN DER AALST; TER HOFSTEDE, 2005). Mathias Weske define os processos de negócios da seguinte forma:

> *Um processo de negócios consiste em um conjunto de atividades que são executadas em coordenação em um ambiente organizacional e técnico. Essas atividades colaboram para atingir conjuntamente um objetivo de negócio. Cada processo de negócios é executado por uma única organização, mas pode interagir com processos de negócios realizados por outras organizações.* (WESKE, 2012)

As organizações vêm gradualmente transferindo para os sistemas de informações o controle da execução de seus processos de negócios (DUMAS; VAN DER AALST; TER HOFSTEDE, 2005). Os sistemas que possuem uma noção explícita de processo são conhecidos através das seguintes denominações: *Process-aware Information systems (PAIS)*, *Business process management system (BPMS)* e *Process management system (PMS).* Em geral, estes sistemas registram informações sobre a execução dos processos aos quais controla, tais como: quando uma atividade iniciou, quando uma atividade foi concluída, quem realizou a atividade e outras. Portanto, uma enorme quantidade de dados sobre a execução dos processos de negócios se encontra “escondida” em bancos de dados ou arquivos de logs destes sistemas. Esses dados podem oferecer uma grande oportunidade para descoberta de conhecimento relevante sobre os processos de negócios, contribuindo para uma operação mais eficiente para a organização.

Nos últimos anos, se desenvolveu uma disciplina de pesquisa denominada mineração de processos de negócios, ou, simplesmente mineração de processos, com o objetivo de apoiar as organizações a descobrir e analisar seus processos de negócios através da extração de informações não triviais dos registros da execução dos processos de negócios em seus respectivos sistemas de informações (ROZINAT et al., 2009). A mineração de processos foi caracterizada como a interseção da modelagem de processos de negócios com a mineração de dados (REMBERT; ELLIS, 2009). Nos últimos anos a mineração de processos vem sendo enxergada como uma ponte que une a ciência de dados à ciência de processos, uma vez que a primeira desconhece a noção de processo enquanto a outra tende a ser orientada a modelos idealizados que ignoram os dados sobre a execução real (VAN DER AALST, 2016).

Um marco importante para a mineração de processos foi a publicação do "Manifesto da mineração de processos" (VAN DER AALST et al., 2012) em 2012. O documento, elaborado por uma força-tarefa criada em 2009 pelo IEEE, apresentou princípios orientadores e principais desafios da mineração de processos. Desde então, é crescente o interesse no tema, como evidência disso o curso online[1] oferecido pelo Professor Dr. Wil van der Aalst já contou com mais de 115.000 inscritos de mais de 190 países. Segundo relatório publicado pela Gartner (MARC KERREMANS, 2019), existem 19 fornecedores de soluções de mineração de processos, demonstrando que mineração tem se consolidado como importante mercado de TIC (MARC KERREMANS, 2019).

1.1 Motivação

Na mineração de processos, os dados sobre a execução de atividades em um processo de negócio são conhecidos como log de eventos (ver Seção 2.4). Qualquer abordagem de mineração de processos requer um log de eventos do processo que se deseja explorar. Através destes as técnicas de mineração de processos são capazes de descobrir o processo real (AS-IS), comparar o processo modelado (TO-BE) com o processo real (AS-IS), identificar gargalos, verificar a incidência de retrabalho e oferecer insights para o aprimoramento de processos de negócios.

A mineração de processos vem avançando rapidamente, mas ainda existem desafios e oportunidades de evolução na área. Dentre os desafios listados em (VAN

[1] Curso online disponibilizado no Coursera (https://www.coursera.org/learn/process-mining)

DER AALST et al., 2012), o primeiro diz respeito ao esforço considerável que se emprega para obtenção de dados adequados à mineração de processos (logs de eventos). Questões como a incidência de dados discrepantes *(outliers)* ou diferentes níveis de granularidade nos eventos registrados influenciam no resultado da mineração de processos.

A questão da granularidade das atividades de um processo é relevante, pois, em certas circunstâncias, deseja-se visualizar o processo em um nível de abstração diferente do que o processo foi modelado. Em alguns momentos, desejamos reunir um conjunto de atividades como sendo uma única atividade (macro atividade), em outros queremos detalhes sobre as etapas intermediárias de uma atividade. Contudo, promover essa transformação é algo trabalhoso e requer conhecimento abrangente sobre o negócio.

O cenário atual da mineração de processos nos mostra que há uma carência por ferramentas para o tratamento de problemas comuns aos ambientes reais, sobretudo no tocante ao pré-processamento dos dados (logs de eventos). Cabe ressaltar que o pré-processamento dos dados é uma etapa custosa. Por outro lado, a qualidade dos dados impacta na qualidade dos modelos de processos descobertos *(garbage in garbage out)*. Portanto, ferramentas de pré-processamento de logs de eventos podem agilizar uma etapa fundamental da mineração de processos, bem como pode propiciar a descoberta de modelos de processos de maior qualidade.

1.2 Definição do problema

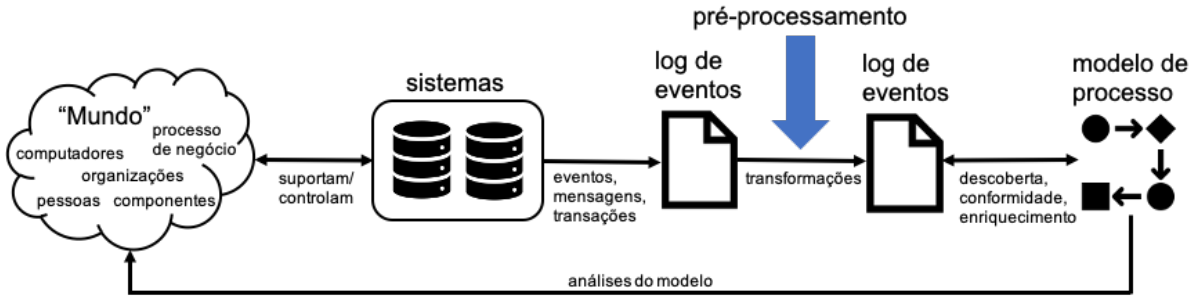
A mineração de processos explora o conhecimento registrado nos logs de eventos. Sendo que, nem sempre os eventos registrados se encontram no nível de granularidade que se deseja enxergar. Em alguns casos os registros são demasiadamente detalhados, prejudicando a extração de conhecimento relevante. Em outros, não há distinção entre etapas intermediárias de uma atividade.

Outro problema clássico da mineração de processos é lidar com processos pouco estruturados; ou seja, processos de negócios que permitem um grau mais alto de liberdade em sua execução. Os logs de eventos oriundos destes ambientes inviabilizam a aplicação de algumas técnicas de mineração de processos (VAN DER AALST, 2011); uma vez que, parte das abordagens não está preparada para lidar com a complexidade inerente a este tipo de processos. Segundo van der Aalst (2011), para

habilitar abordagens mais avançadas em processos pouco estruturados, por exemplo previsões e recomendações baseadas em logs de eventos, é recomendável “estruturar mais o processo”. Em outras palavras, quando diante de processos pouco estruturados, faz-se necessário manipulação do log de eventos para que se permita a exploração de todo o potencial da mineração de processos. Em contrapartida, os processos pouco estruturados oferecem um maior potencial para realização de melhorias (VAN DER AALST, 2011).

Alguns algoritmos de mineração de processos foram concebidos para lidar com a complexidade inerente aos processos pouco estruturados. Contudo, não resolvem todos os desafios encontrados nos processos pouco estruturados (D’CASTRO; OLIVEIRA; TERRA, 2018). Por outro lado, a descoberta de modelos complexos torna difícil compreender o funcionamento do processo descoberto. As abordagens para lidar com a complexidade diferem no ponto em que algumas técnicas buscam reduzir a complexidade durante a descoberta de processos, enquanto outras atuam no pré-processamento de logs de eventos. A Figura 1, adaptação da Figura 2.5 em (VAN DER AALST, 2016), mostra uma visão geral da mineração de processos. Como pode ser observado, os fenômenos do “mundo”, registrados nos sistemas de informação, subsidiam a aquisição dos dados para mineração de processos (log de eventos). O pré-processamento consiste em uma etapa de transformação dos dados (log de eventos) com o intuito de favorecer as etapas subsequentes da mineração de processos.

Figura 1 – Visão geral da mineração de processos



Coletar e preparar dados são as atividades que consomem mais tempo em projetos de mineração de dados (MUNSON, 2012). Na mineração de processos a realidade não é muito diferente (VAN DER AALST et al., 2012). Ainda assim, a literatura sobre o pré-processamento de logs de eventos na mineração de processos é escassa.

Outro aspecto a ser considerado em relação ao tamanho dos logs de eventos, diz respeito aos custos computacionais e financeiros. Algumas ferramentas limitam o tamanho do log de eventos, outras usam a quantidade de eventos como parâmetro para licenciamento. Portanto, quanto maiores os logs de eventos, mais caro é a mineração de processos.

1.3 Objetivos

Esta pesquisa tem como objetivo principal propor técnicas para transformação de logs de eventos que contribuam para o aumento da qualidade dos modelos de processos descobertas através da mineração de processos. Para tanto, buscamos compreender os fatores que influenciam na qualidade dos modelos de processos para propor soluções que ajudem a melhorar estes aspectos. As abordagens propostas nesta tese buscam oferecer visões alternativas coerentes com a realidade através da transformação dos logs de eventos. Também faz parte dos objetivos a avaliação das abordagens propostas através de experimentos com dados reais.

A seguir listamos os principais objetivos específicos perseguidos nesta tese:

- Investigar e classificar comportamentos que impactam no aumento da complexidade em modelos de processos descobertos através da mineração de processos em ambientes reais;
- Analisar e propor métricas e indicadores capazes de identificar em logs de eventos comportamentos que impactam no aumento da complexidade em modelos de processos descobertos através da mineração de processos;
- Desenvolver técnicas de pré-processamento que sejam capazes de: simplificar os logs de eventos através da mudança da granularidade dos eventos e contribuir para a qualificação dos modelos de processos descobertos através destes logs;
- Desenvolver técnicas de pré-processamento de logs de eventos que proporcionem visões alternativas coerentes para os processos de negócios.

1.4 Contribuições

A principal contribuição deste trabalho é oferecer técnicas automáticas para transformação de logs de eventos, que é um dos principais desafios para mineração de processos, caracterizado no “Manifesto da mineração de processos” (VAN DER AALST et al., 2012) como (C1) *Finding, Merging, and Cleaning Event Data*. Cabe ressaltar que o desafio é bastante abrangente, abarcando diversos aspectos da preparação dos dados para mineração de processos. O escopo dessa pesquisa foca em aspectos relacionados aos níveis de granularidade dos eventos e ao comportamento excepcional.

Também esperamos contribuir com a redução da barreira de implantação da mineração de processos nas organizações, uma vez que a complexidade inerente aos processos reais exige esforço humano substancial na etapa de pré-processamento dos dados. Esperamos que as abordagens propostas nesta tese contribuam para redução de custos, tornando a mineração de processos viável nestes ambientes.

Apenas uma pequena parcela dos estudos em mineração de processos é voltada para sua aplicação em organizações públicas. Em relação a aplicações no judiciário a literatura é bastante escassa. Contudo, enxergamos que o poder público, sobretudo o judiciário, é o destinatário ideal para esse tipo de tecnologia, uma vez que oferecem ambientes fortemente orientados a processos e dispõem dos elementos essenciais para a mineração de processos. Além disso, o setor público é ávido por soluções que contribuam para o incremento da sua eficiência.

O estudo apresentado no Capítulo 4 desta tese subsidiou o artigo *Process Mining Discovery Techniques in a low-structured Process Works?* (D’CASTRO; OLIVEIRA; TERRA, 2018), apresentado no *7th Brazilian Conference on Intelligent Systems (BRACIS)*. O estudo mostrou que as ferramentas analisadas apresentam limitações frente às aos processos flexíveis. As limitações observadas serviram de ponto de partida para as abordagens propostas nesta tese.

A abordagem proposta no Capítulo 5 levou ao artigo *Process Mining Pre-processing Technique for Event Log Simplification Based on Merging of Events*, submetido ao Decision Support Systems, Elsevier, 2020. Juntamente com as outras duas abordagens para transformação de logs de eventos apresentadas nesta tese, esperamos que contribuam para avanço do estado da arte no pré-processamento de logs de eventos.

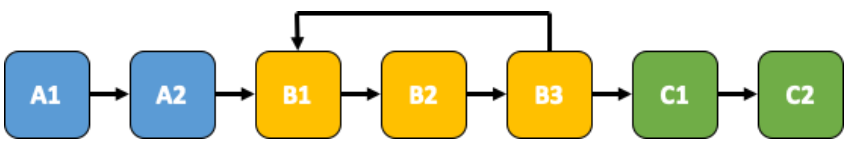
1.5 Metodologia

Nesta tese, realizamos uma pesquisa aplicada, uma vez que a ênfase é na solução de problemas específicos relacionados a complexidade e a qualidade dos modelos de processos descobertos a partir da mineração de processos complexos. A lógica da pesquisa é indutiva, pois foca na compreensão dos fatores geradores de complexidade em modelos de processos descobertos através da mineração de processos, visando analisar seu impacto, gerando conclusões e soluções mais gerais. O objetivo da pesquisa é exploratório e descritivo, pois inicialmente visa aprofundar um tema pouco explorado e, ao final, realiza estudos experimentais sobre uma proposta elaborada na etapa exploratória.

A pesquisa utilizou uma abordagem majoritariamente baseada em métodos quantitativos, embora possua elementos de métodos qualitativos para coleta de feedback de especialistas. Também utilizamos uma abordagem mista na metodologia de pesquisa, uma vez que se tratou de uma investigação que empregou mais de um método de coleta de dados de diferentes fontes. Nesse sentido, foram utilizados quatro métodos de coleta de dados: estudo de caso exploratório, revisão da literatura, experimentos e entrevistas. O estudo de caso exploratório possibilitou a identificação de limitações e oportunidades para mineração de processos complexos. A revisão da literatura forneceu as bases para criação dos indicadores utilizados nas abordagens propostas. Os experimentos subsidiaram a avaliação do impacto das abordagens propostas na qualidade e simplicidade dos modelos de processos. Por fim, realizamos entrevistas com especialistas para obter feedback dos resultados obtidos.

O trabalho foi organizado nas seguintes fases: (A) estudos exploratórios, (B) desenvolvimento e (C) avaliação das abordagens. A Figura 2 apresenta as atividades e fases correspondentes na metodologia empregada.

Figura 2 – Fase e atividades da metodologia empregada



A seguir detalhamos as fases e atividades apresentadas na Figura 2.

(A) Estudos exploratórios: realização de estudo de caso exploratório com intuito de explorar a mineração de processos em ambientes a fim de identificar limitações e oportunidades de melhorias. A fase é dividida em duas etapas:

(A1) Mineração de processos em ambientes de processos complexos: investigação sobre a eficácia da mineração de processos quando aplicada em ambientes de processos complexos.

(A2) Identificação e classificação de fatores geradores de complexidade em modelos de processos: investigação para identificação e classificação de padrões de comportamento observados em processos que influenciam na complexidade dos modelos descobertos pelas técnicas de mineração de processos.

(B) Desenvolvimento de abordagens: realização de revisão da literatura, elaboração teórica, desenvolvimento prático e análise preliminar de abordagem. A fase corresponde a um ciclo de três atividades executado para cada abordagem proposta.

(B1) Desenvolvimento teórico: revisão da literatura, realização de estudos para elaboração teórica de abordagem para tratamento dos problemas elencados na etapa A2.

(B2) Desenvolvimento prático: implementação da abordagem concebida em B1.

(B3) Análise Preliminar: execução com massa de teste e análise dos resultados observados com abordagem implementada em B2. As limitações e oportunidades de melhorias são insumos para uma nova rodada de desenvolvimento teórico (B1).

(C) Avaliação das abordagens: realização de estudos contemplando análise a partir de logs de eventos de ambientes reais.

(C1) Análises: realização de experimentos utilizando dados reais para avaliação da eficácia das abordagens propostas através de métricas de qualidade consagradas na literatura de mineração de processo, bem como a coleta de feedback de especialistas.

(C2) Interpretação dos resultados: análise dos resultados observados na atividade A2, inclusive, recorrendo ao apoio de especialista no negócio (magistrados e analistas de sistemas) quando necessário.

1.6 Trabalhos Relacionados

Esta tese aborda uma diversidade de temas relevantes para a mineração de processos, de forma que se relaciona a diversos trabalhos. Organizamos essa seção nos seguintes tópicos: Seção 1.6.1 contempla os trabalhos relacionados que abordam o agrupamento de atividades em modelos de processos descobertos pela mineração de processos; Seção 1.6.2 aborda os trabalhos relacionados ao tema eliminação de atividade infrequente; e por fim, a Seção 1.6.3 reúne os demais trabalhos relacionados.

1.6.1 Agrupamento de atividades

Em (GÜNTHER; VAN DER AALST, 2007) foi apresentada uma abordagem para descoberta de processos baseada na experiência dos seus idealizadores em mineração de processos reais. Uma das características da abordagem proposta consiste no agrupamento (*clustering*) de atividades. A ideia apresentada no trabalho de Günther e van der Aalst (2007) difere de qualquer abordagem proposta nesta tese, uma vez que atuamos na fase de pré-processamento, transformando os registros de eventos em vez de criar novas visualizações para o processo.

Suriadi et al. (SURIADI et al., 2017) apresentam um estudo que analisa imperfeições observadas em logs de eventos de ambientes reais que comprometem a qualidade da mineração de processos. Onze classes de imperfeições foram apresentadas. Dentre estes, os autores denominaram os eventos que ocorrem em um curto período de tempo como eventos colaterais. Semelhante ao nosso trabalho, a estratégia sugerida para tratamento de eventos colaterais foi a fusão dos eventos das atividades colaterais. Segundo os autores, o sucesso da técnica depende do conhecimento de especialistas no domínio da aplicação e também da habilidade dos analistas de processos. Cabe ressaltar que os autores não oferecerem uma proposta para identificação e tratamento do que chamou de atividades colaterais, sugerindo apenas que a questão seja tratada por especialistas. Nosso trabalho aborda a questão da granularidade, mas o nosso enfoque não é em imperfeições de dados. Além disso,

propusemos técnicas para tratamentos das situações observadas.

Em (FOLINO; GUARASCIO; PONTIERI, 2015) é proposto um método automatizado para descoberta de modelos de processo consistindo em três partes: um modelo lógico de agrupamento de eventos, para abstrair eventos de baixo nível em classes; um modelo de agrupamento de rastreamento lógico, para discriminar entre as variantes do processo; e um conjunto de esquemas de fluxo de trabalho, cada um descrevendo uma variante em termos dos clusters de eventos descobertos. A abordagem de agrupamento de eventos se baseia em modelos preditivos *(Predictive Clustering Trees).* A abordagem apresentada em (FOLINO; GUARASCIO; PONTIERI, 2015) se dirige para os mesmos problemas desta tese; contudo, difere em relação a estratégia e técnicas propostas seguem caminhos distintos.

PRODEL et al. (PRODEL et al., 2018) propuseram uma abordagem para visualização do processo em um nível de abstração mais alto baseado no agrupamento de atividades de acordo com uma estrutura hierárquica de atividades pré-definida. Nosso trabalho difere do trabalho apresentado em (PRODEL et al., 2018) na medida em que não utilizamos qualquer conhecimento ou artefato além do log de eventos. As abordagens também diferem no ponto em que o referido trabalho propõe uma técnica para descoberta de processos, enquanto propusemos abordagens de pré-processamento de logs de eventos.

Em (REHSE; FETTKE, 2019) é apresentada uma prova de conceito para uma nova abordagem de descoberta de componentes (subprocessos) para modelo de referência. Para tanto, as atividades são agrupadas hierarquicamente com base em sua proximidade no log de eventos. O objetivo da abordagem proposta em (REHSE; FETTKE, 2019) é descobrir subprocessos para favorecer o reuso de componentes na etapa de modelagem de processos.

1.6.2 Pré-processamento de comportamento infrequente

Conforti, Rosa e Hofstede (CONFORTI; ROSA; HOFSTEDE, 2017) propõem uma solução para eliminação de comportamento infrequente através do pré-processamento de logs de eventos. Esta tese se propõe a realizar o pré-processamento de logs de eventos e contempla a eliminação de comportamento infrequente. Contudo, os objetivos e métodos empreendidos diferem. Diferentemente do nosso trabalho, a abordagem de Conforti, Rosa e Hofstede (2017) tem foco na eliminação de comportamento ruidoso do log de eventos, enquanto adaptamos a

eliminação de comportamento infrequente de forma complementar a abordagem de agrupamento de atividades afins. Inclusive, no Capítulo 5 apresentamos um estudo no qual avaliamos uma das métricas utilizadas na abordagem em (CONFORTI; ROSA; HOFSTEDE, 2017) que se mostrou pouco eficaz para nossos objetivos.

Outros três trabalhos (CHAPELA-CAMPA; MUCIENTES; LAMA, 2019; FANI SANI; VAN ZELST; VAN DER AALST, 2018a; SUN et al., 2019) apresentam abordagem para tratamento de comportamento infrequente através do pré-processamento do log de eventos.

1.6.3 Outros trabalhos relacionados

Tax, Sidorova e van der Aalst (TAX; SIDOROVA; VAN DER AALST, 2018) apresentaram abordagem para eliminação de eventos de atividades caóticas através do pré-processamento do log de eventos. Atividades caóticas foram definidas como atividades que podem ocorrer a qualquer momento, independentemente do estado em que o processo se encontre. O trabalho apresenta algumas semelhanças com a abordagem proposta no Capítulo 7 desta tese, inclusive, parte das métricas utilizadas em nossa tese se baseou em métricas adotadas em (TAX; SIDOROVA; VAN DER AALST, 2018). Contudo, a despeito da afinidade no tocante às métricas, as abordagens diferem completamente na forma de tratamento para os problemas abordados. Esta questão foi abordada com mais detalhes no Capítulo 7.

Recentemente, Sani (SANI, 2020) apresentou ideias gerais para pré-processamento de log de eventos para lidar com o intuito de diminuir o tamanho e a complexidade dos dados de eventos. O objetivo do trabalho está bem alinhado com os propósitos de nossa pesquisa. Contudo, o trabalho aborda apenas superficialmente o tema, apresentando resultados preliminares e indicando o desenvolvimento de métodos para pré-processamento de logs de eventos.

1.7 Organização do Documento

Os próximos capítulos estão organizados da seguinte forma: o Capítulo 2 apresenta os conceitos necessários para a compreensão desta tese, bem como das abordagens nela propostas; o Capítulo 3 apresenta os objetivos e desafios da mineração de processos, bem como seus principais algoritmos e ferramentas; o Capítulo 4 discute a mineração de processo complexos, analisando fatores geradores

de complexidade em modelos de processos e destacando perfis de atividades que influenciam na produção de modelos de processos complexos; o Capítulo 5 apresenta uma abordagem para agrupamento de atividades afins e o Capítulo 6 uma extensão dessa abordagem incorporando a eliminação de comportamento infrequente; o Capítulo 7 aborda a problemática das atividades recorrentes e seus reflexos na descoberta de processos, bem como a proposição de uma abordagem para seu tratamento; e finalmente, o Capítulo 8 apresenta uma breve conclusão sumarizando as conclusões relatadas nos capítulos anteriores, além de indicar as limitações e possíveis trabalhos futuros.

2 PRELIMINARES

Neste capítulo, apresentamos a definição formal e a notação dos conceitos fundamentais utilizados nesta tese. A Seção 2.1 apresenta o conceito de processo de negócio e a Seção 2.2 mostra suas formas de representação (modelos de processos). A Seção 2.3 mostra definições matemáticas dos conceitos básicos que suportam as técnicas de mineração de processos. A Seção 2.4 aborda os logs de eventos, que são ponto de partida para a mineração de processos. As duas seções seguintes apresentam técnicas complementares utilizadas: árvore de decisão (Seção 2.5) e medidas de similaridades entre textos (Seção 2.6).

2.1 Processos de negócio

Um processo de negócio, ou simplesmente processo[2], pode ser definido por um conjunto de atividades que recebem uma ou mais entradas, consome recursos e produz uma saída que tem valor para os clientes (HAMMER; CHAMPY, 1993). Mesmo quem não convive com processos formais em seu ambiente de trabalho, convive com processos cotidianamente, mesmo que não se dê conta. Por exemplo, quando necessitamos levar o carro para manutenção. Considerando que se trata de uma revisão regular em oficina autorizada, primeiro precisamos escolher uma oficina no site da montadora. Em seguida, agendamos um atendimento, geralmente indicando o tipo de serviço desejado. Antes de entregar o carro temos que aguardar uma vistoria e posteriormente assinar um termo confirmando as informações da vistoria. Outras atividades seguem até que o serviço seja aprovado, pago e aceito pelo cliente. Observando esse exemplo à luz da definição de HAMMER e CHAMPY, podemos identificar que se trata de um processo. A Tabela 1 apresenta algumas etapas do processo (hipotético) de manutenção veicular.

Podemos encontrar exemplos semelhantes quando interagimos com as mais variadas organizações (Universidades, Hospitais, Planos de Saúde, Operadores de telefonia, Poder Judiciário, Prefeitura, Delegacia e outras). Cada vez mais as organizações usam sistemas de informações que suportam esses processos. O nível de suporte fornecido por esses sistemas varia de acordo com a aplicação. Por

[2] No restante desta tese, usamos o termo processo, mas assumimos implicitamente que processos devem ser executados no contexto de organizações profissionais, ou seja, processos que descrevem como os casos são tratados com um ponto inicial e final bem definido.

exemplo, o processo judicial é altamente regulado, logo o seu sistema de suporte deve ser sensível ao processo para garantir um maior controle sobre sua execução. Por outro lado, um atendimento hospitalar tende a ser mais flexível. Portanto, em geral, os sistemas de informações destes estabelecimentos não se prestam a controlar o processo, apenas registram as informações necessárias para fins de cobrança e manutenção de histórico do paciente.

Tabela 1 – Etapas iniciais do agendamento de serviço de manutenção veicular

ENTRADA	**ATIVIDADE (RECURSO)**	**SAÍDA**
Tipo de serviço e Região	Escolha da oficina (Site)	Indicação de oficinas habilitadas para o serviço, bem como os canais de agendamentos disponíveis
Solicitação de atendimento (Identificação do solicitante, do veículo e tipo de serviço)	Agendamento (Atendente)	Agendamento realizado (Agendamento da vistoria e identificação do consultor)
Informações sobre o estado do veículo e o próprio veículo	Vistoria (Consultor)	Termo de vistoria indicando o estado que o veículo se encontra

A despeito do nível de sensibilidade dos sistemas de informações, todos deixam "pegadas" que indicam o que aconteceu e quando. Essas pegadas são chamadas de logs de eventos (ver seção 2.4) e geralmente são armazenadas em bancos de dados ou em arquivos de log. Por conter dados factuais sobre a execução dos processos, estamos falando de uma fonte valiosa de informações que passa despercebida por grande parte das organizações. Muitas vezes os proprietários dos processos têm uma percepção baseada exclusivamente na intuição sobre a forma como os seus processos são executados.

Business Process Management (BPM) ou Gestão de Processos de Negócios é uma disciplina que combina abordagens que permeiam todo o ciclo de vida de um processo, passando pelas etapas de: planejamento, análise, desenho, implementação, monitoramento e refino. Até recentemente, havia poucas conexões entre os dados produzidos durante a execução do processo e o design do processo real. A mineração de processos oferece a possibilidade de realmente “fechar” o ciclo de vida do BPM, uma vez que os dados registrados por sistemas de informação podem ser usados para fornecer uma melhor visão dos processos reais, contribuindo diretamente para melhoria da qualidade dos processos (VAN DER AALST, 2016).

Vale ressaltar que a mineração de processos não se limita ao BPM, sendo acessível a qualquer processo que tenha os eventos registrados em algum sistema de informação.

2.2 Modelos de Processos

Um aspecto importante dos processos diz respeito a sua forma de representação. Podemos distinguir dois tipos de linguagens de modelagem de processos (SCHÖNIG; JABLONSKI, 2016):

- Procedimentais: voltados para processos de rotina bem estruturados com fluxo de controle exatamente predeterminado
- Declarativos: adequado para processos ágeis com fluxo de controle que evolui no tempo de execução sem ser exatamente definido a priori.

Em outras palavras, os modelos procedimentais definem o comportamento que é permitido e os modelos declarativos definem o comportamento que não é permitido (DUNZER et al., 2019). O primeiro tipo inclui linguagens de modelagem populares, como linguagem de modelagem unificada (UML), notação de modelo de processos de negócios (BPMN) e redes de Petri. São exemplos de linguagens de processos declarativas (SCHÖNIG; JABLONSKI, 2016): *Declare*, *Dynamic Condition Response-Graphs* (DCR-*Graphs*), *Case Management Modelling and Notation* (CMMN) e *Declarative Process Intermediate Language* (DPIL).

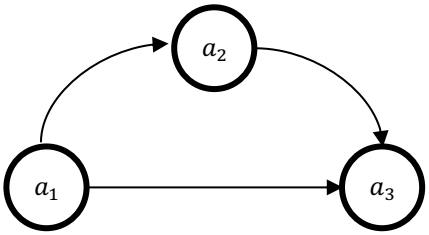
Os processos ágeis são comuns na área da saúde, onde, por exemplo, os processos de diagnóstico e tratamento do paciente requerem flexibilidade para lidar com circunstâncias imprevistas. Quanto mais restrições forem adicionadas ao modelo, menos alternativas possíveis de execução permanecem (SCHÖNIG; JABLONSKI, 2016). Por exemplo, a abordagem *Declare*, que em vez de definir explicitamente a ordem das atividades nos modelos, conta com restrições para determinar implicitamente a possível ordem das atividades (qualquer ordem que não viole restrições é permitida) (VAN DER AALST; PESIC; SCHONENBERG, 2009).

Em se tratando de modelos procedimentais, geralmente são representados através de uma notação gráfica. A Figura 3 ilustra um log de eventos A e um modelo de um processo em uma notação simplificada gerado a partir de A. Na Figura 3, as

atividades do processo são representadas por círculos e as transições de uma atividade para outra por setas conectando as atividades envolvidas na transição.

Figura 3 – Exemplo de modelo de processo em notação simplificada

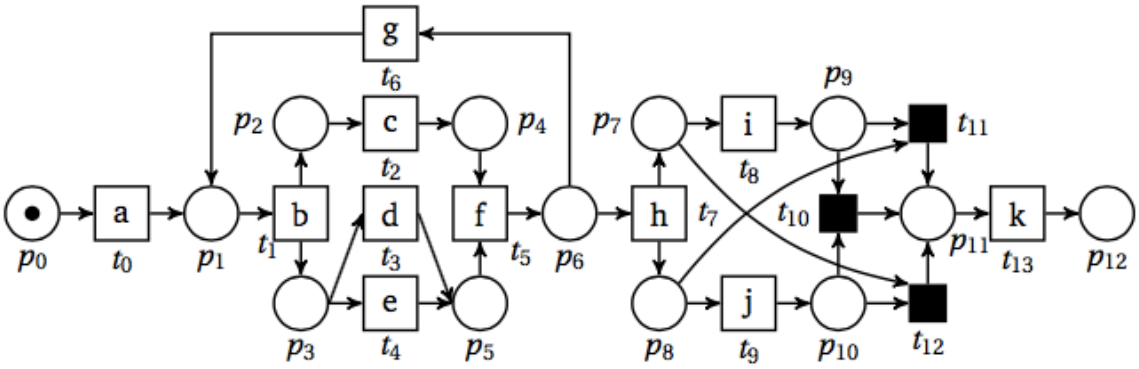
$$A = \{\langle a_1, a_2\rangle, \langle a_1, a_3\rangle, \langle a_2, a_3\rangle, \langle a_1, a_2, a_3\rangle\}$$



Diversas notações de modelos de processos são utilizadas pelas ferramentas de mineração de processos para representação dos modelos descobertos. A seguir apresentamos de forma sucinta as principais notações para modelagem utilizadas pelas abordagens de mineração de processos.

Uma das linguagens de modelagem de processos mais utilizada na mineração de processos são as redes de Petri. Elas se baseiam em uma notação muito simples com círculos representando lugares, quadrados representando transições e as setas conectando-as de forma bipartida (MURATA, 1989). As transições podem representar uma tarefa que quando executada consome um *token*, representados por pontos pretos que são movidos entre os lugares. A distribuição dos *tokens* sobre os lugares, chamada de marcação, indica o estado de uma instância do processo. Um exemplo de um modelo de rede Petri é mostrado na Figura 4.

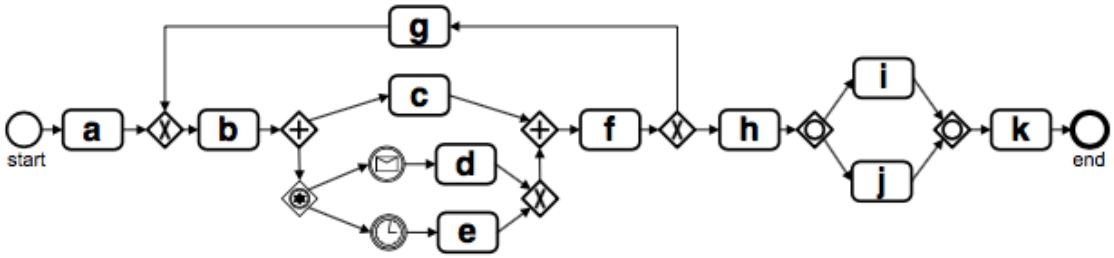
Figura 4 – Exemplo de Rede de Petri (AALST, 2011)



O BPMN *(Business Process Model and Notation)* é uma notação de modelagem de processos de negócios amplamente utilizada na indústria (OMG, 2014). A notação

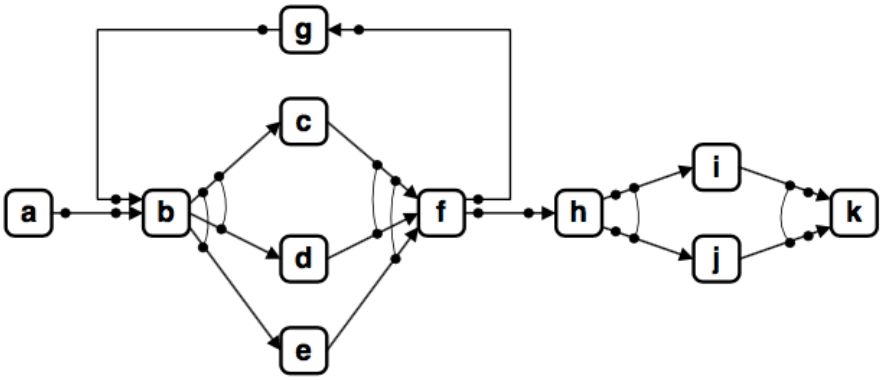
BPMN foi padronizada pelo OMG (*Object Management Group*), sendo atualmente suportada por diversas ferramentas. A notação BPMN é bastante extensa com diferentes tipos de atividades, eventos, *gateways* (escolhas) e outros elementos, resultando em uma notação complexa. As abordagens de mineração de processos tendem a utilizar apenas um subconjunto de elementos da notação. Poucas ferramentas comerciais de mineração de processos descobrem modelos em BPMN. No ProM (ver Seção 3.2) existem *plugins* que realizam a conversão de outros formatos para BPMN e vice-versa. A Figura 5 apresenta um modelo de processo em BPMN.

Figura 5 – Exemplo de modelo em BPMN (AALST, 2011)



Uma rede Causal (ou C-*net*) é uma notação de modelagem de processo adaptada à descoberta do processo (VAN DER AALST; ADRIANSYAH; VAN DONGEN, 2011). A C-*net* consiste em um grafo onde os nós representam atividades e os arcos representam dependências causais. Além disso, cada atividade possui um conjunto de possíveis ligações de entrada e um conjunto de possíveis ligações de saída. Um exemplo de uma rede causal é mostrado na Figura 6, que descreve o mesmo processo que os exemplos das notações de modelagem anteriores.

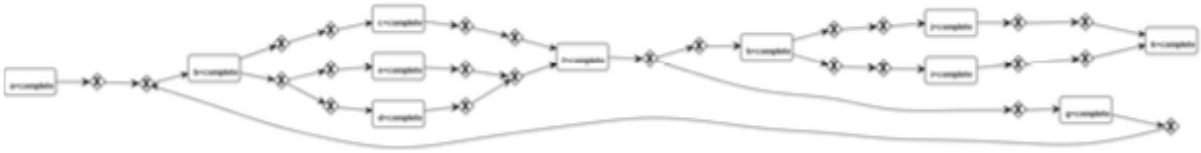
Figura 6 – Exemplo de modelo em C-net (AALST, 2011)



A rede Heurística (ou *Heuristic Net)* é uma notação gerada pelo algoritmo *Heuristic Miner* (WEIJTERS; RIBEIRO, 2011) e consiste em uma representação em

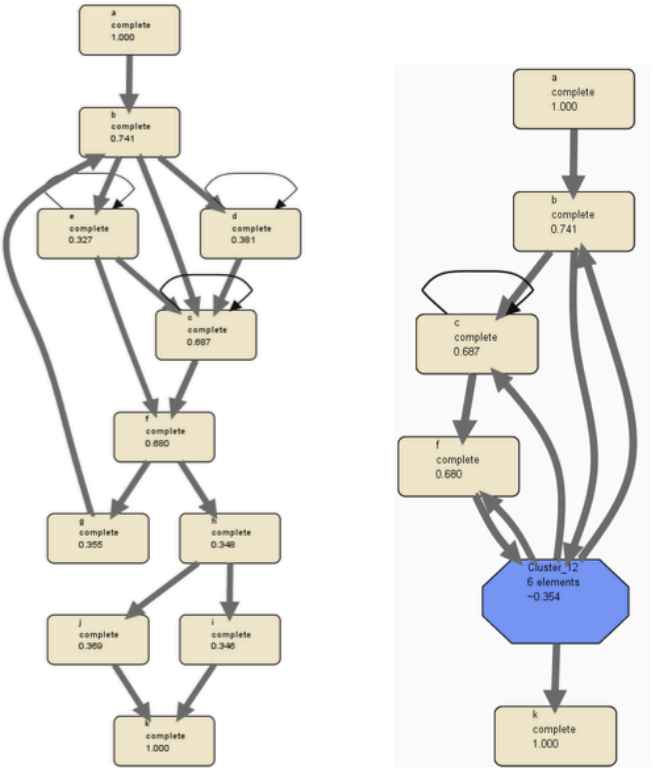
que as ligações de atividades podem ser visualizadas como escolhas da notação BPMN. Um exemplo de uma rede Heurística é mostrado na Figura 7.

Figura 7 – Exemplo de modelo de uma rede Heurística



O modelo *Fuzzy* corresponde a notação de modelagem de processo adotada pelo algoritmo *Fuzzy Miner* (GÜNTHER; VAN DER AALST, 2007). Uma das características dessa notação é a possibilidade de representar agrupamentos de atividades. A Figura 8 apresenta dois modelos gerados a partir de um mesmo processo. No primeiro, um modelo *Fuzzy* não há atividades agrupadas. No segundo, algumas atividades do processo foram agrupadas, sendo representado por um octógono azul.

Figura 8 – Exemplos de modelos Fuzzy



Existem outras representações de modelos de processos que não foram apresentadas nessa seção, tais como: YAWL (*Yet Another Workflow Language*), *Driven Process Chains* (EPCs) ou cadeias de Markov (VAN DER AALST, 2016).

2.3 Conceitos e Notações Básicas

2.3.1 Conjuntos

Definição 1 (Conjunto). Um conjunto é uma coleção não ordenada de objetos únicos, chamados de elementos do conjunto. Um conjunto pode ser representado listando seus elementos entre chaves. Outra maneira alternativa de definir um conjunto é através da declaração de uma propriedade (ou predicado) válido somente para os elementos do conjunto.

Exemplo: $A = \{a_1, a_2, a_3, \dots, a_n\}$ e $B = \{x \in \mathbb{N} \mid x \text{ é divisível por } 2\}$.

O símbolo ∈ expressa que um elemento pertence a um conjunto e o ∉ expressa que um elemento não pertence a um conjunto. Ou seja, se $a_1 \in A$ e $b_1 \notin A$, temos que a_1 pertence ao conjunto A e b_1 não pertence a conjunto A. Cabe ressaltar que a ordem dos elementos de um conjunto é irrelevante, portanto, $\{a, b, c\} = \{b, c, a\}$.

Nesta tese, utilizamos as seguintes notações padrões para conjuntos:

- $\mathbb{N}$ é o conjunto de números naturais $\{1, 2, 3 \dots\}$
- $\mathbb{N}_0$ é o conjunto de números naturais incluindo zero $\{0, 1, 2, \dots\}$
- $\mathbb{R}$ é conjunto de números reais, sendo $\mathbb{R}^+$ e $\mathbb{R}^-$ respectivamente os conjuntos de números positivos e negativos. $\mathbb{R}_0^+$ é o conjunto de números reais positivos incluindo zero, ou seja, $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$.

Um conjunto é $finito$ quando possui uma quantidade finita de elementos. Caso contrário é chamado de $infinito$. Um conjunto sem elementos é chamado de conjunto vazio, sendo representado por ∅ ou {}.

Definição 2 (Cardinalidade de um conjunto). A cardinalidade de um conjunto finito A consiste na quantidade de elementos de A, sendo representada por $|A|$.

Exemplo: dado $A = \{a, b, c\}$, então $|A| = 3$.

Definição 3 (Subconjunto, subconjunto próprio e igualdade). O conjunto A é um subconjunto de B se cada elemento A também pertencer a B. Este tipo de relação é

denotado por $A \subseteq B$. Caso todos os elementos dos conjuntos A e B sejam coincidentes, ou seja, $A \subseteq B$ e $B \subseteq A$, existe a igualdade entre os conjuntos, denotado por $A = B$.

Exemplo: dado $A = \{a, b, c\}$, $B = \{a, b, c, d, e\}$ e $C = \{c, b, a\}$, então: $A \subseteq B$, $C \subseteq B$, $A \neq B$, $A = C$.

Definição 4 (Conjunto disjunto). Dois conjuntos, A e B, são $disjuntos$ ou $mutuamente\ exclusivos$, se não há elementos comuns entre ambos.

2.3.2 Operações, Funções e Relações

Definição 5 (União, Intersecção e Diferença). A união entre dois conjuntos A e B, denotada por $A \cup B$, consiste no conjunto com todos os elementos de A, B. A interseção entre dois conjuntos A e B, denotada por $A \cap B$, consiste no conjunto de elementos coincidentes em A e B. A diferença entre dois conjuntos A e B, denotada por $A - B$, consiste em todos os elementos presentes em A que não pertencem a B.

Exemplo: dado $A = \{a, b, c, d\}$ e $B = \{c, d, e, f\}$, então: $A \cup B = \{a, b, c, d, e, f\}$, $A \cap B = \{c, d\}$ e $A - B = \{a, b\}$.

Definição 6 (Particionamento). O particionamento de um conjunto A é uma coleção $\mathcal{C}$ de subconjuntos não vazios não sobrepostos de A cuja união é o conjunto A. Ou seja, $P_1, P_2 \in \mathcal{C}$ se $P_1 \neq P_2 \wedge P_1 \cap P_2 = \emptyset \wedge A = \emptyset \cup P_1 \cup P_2 \cup P_n$.

Exemplo: dado $A = \{a, b, c, d, e, f\}$, então podemos ter $\mathcal{C} = \{\{a, b, c\}, \{d\}, \{e, f\}\}$.

Definição 7 (Par ordenado e produto cartesiano). O $produto\ cartesiano$ entre dois conjuntos A e B, denotado por $A \times B$, consiste no conjunto de todos os $pares\ ordenados$ (a, b), tal que $a \in A$ e $b \in B$. Ou seja, $A \times B = \{(a, b) | a \in A \text{ e } b \in B\}$. Além disso, $(a, b) \neq (b, a)$, a não ser que $a = b$.

Exemplo: dado $A = \{a_1, a_2, a_3, \}$ e $B = \{b_1, b_2\}$, então temos que $A \times B = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2), (a_3, b_1), (a_3, b_2)\}$.

Definição 8 (Relação). A relação binária $\mathcal{R}$ de um conjunto A consiste no conjunto de pares ordenados definidos através de alguma regra, ou seja, $\mathcal{R} \subseteq A \times B$. Alternativamente, pode-se escrever uma relação com $a\ \mathcal{R}\ b = (a, b) \in \mathcal{R}$.

Exemplo: dado $A = \{1, 2, 3, 4\}$ e a relação $\mathcal{R}$ "metade de" em A, temos que $\mathcal{R} = \{(1,2), (2,4)\}$.

Definição 9 (Função). Uma função f de um conjunto A para um conjunto B, denotada por $f: A \rightarrow B$, consiste em um mapeamento que atribui a cada elemento de A um valor $f(a) \in B$.

2.3.3 Multiconjunto, Tuplas e Sequências

Definição 10 (Multiconjunto). Uma $multiconjunto$ consiste em uma coleção não ordenada que permite a repetição de elementos.

Exemplo: $multiconjunto\ A = [a, a, a, b, b, c]$. Alternativamente, pode ser representado por $A = [a^3, b^2, c^1]$.

Definição 11 (Tupla ou Lista). Uma $tupla$ ou $lista$ consiste em uma coleção ordenada de elementos. Uma $n - upla$ é uma lista ordenada de n elementos.

Exemplo: $A = \langle a, a, b, c \rangle$ ou $A = \langle b, c \rangle$.

Definição 12 (Sequência). Uma $sequência$ é uma lista ordenada de símbolos. Para uma sequência $\boldsymbol{s} = \langle s_1, s_2, \dots, s_n \rangle$ temos as seguintes características:

- $|\boldsymbol{s}|$ é o tamanho da sequência $\boldsymbol{s}$
- $\boldsymbol{s}(i)$ representa o símbolo da posição i
- $\boldsymbol{s}(i, j)$ representa a subsequência contínua de $\boldsymbol{s}$ que inicia na posição i e termina na posição j

Exemplo: dado $A = \langle a, a, b, c \rangle$, temos $|\boldsymbol{s}| = 4$, $\boldsymbol{s}(3) = b$ e $\boldsymbol{s}(2,4) = \langle a, b, c \rangle$.

2.4 Log de Eventos

Logs de eventos são o ponto de partida para a mineração de processos (ROZINAT et al., 2009). Para os objetivos dessa tese, faz-se necessário caracterizar apropriadamente a noção de log de eventos, bem como os conceitos correlatos. Esta seção abrange a definição formal de log de eventos e seus elementos, como também oferece uma visão prática sobre sua aquisição e uso. Esta seção foi organizada como segue: a Seção 2.4.1 define os principais conceitos relacionados aos logs de eventos, a Seção 2.4.2 apresenta os formatos de armazenamento dos logs de eventos e, finalmente, a Seção 2.4.3 aborda aspectos práticos acerca da manipulação dos logs de eventos nas ferramentas de mineração de processos.

2.4.1 Conceitos

No contexto da mineração de processos, um log de eventos armazena dados sobre a ocorrência de atividades capturadas pelos sistemas de informações enquanto suportam a execução de um processo de negócio. A cada execução de uma instância do processo temos uma sequência de eventos (VAN DER AALST, 2016), que podem ser caracterizados por vários atributos. Por exemplo, um evento pode ter os seguintes atributos: a atividade executada, o início da execução da atividade (data/hora), um indicativo do responsável pela execução ou outros. A seguir apresentamos a definição formal de eventos e seus atributos, contudo, antes cabe esclarecer que o conteúdo apresentado nessa seção se restringe às características necessárias para os objetivos dessa tese, não tendo qualquer pretensão de ser exaustivo e completo. O conceito de logs de eventos foi concebido para abarcar as mais diversas realidades e objetivos de mineração de processos. Logo, a apresentação exaustiva de todos os detalhes nos desviaria dos objetivos.

Definição 13 (Evento e Atributo (VAN DER AALST, 2016)). Seja $\mathcal{E}$ o universo de eventos e AN o conjunto de todos os atributos. Para qualquer evento $e \in \mathcal{E}$ que tenha um atributo $a \in AN$, $\#_a$ denota o valor do atributo a. Em outras palavras, $\#_a$ é o valor do atributo a para o evento e. Caso o evento e não tenha um determinado atributo n, então $\#_n(e) = \perp$ (nulo).

Exemplo: Tomando por base um evento e com três atributos, sendo eles: atividade (o

que foi executado), tempo (quando foi executado) e recurso (por quem foi executado). Seja $\mathcal{A}, \mathcal{T}, \mathcal{R}$ respectivamente os domínios de atividade, tempo e recurso; ou seja, o conjunto de valores possíveis para cada atributo. Então, $\#_{atividade}(e) \in \mathcal{A}$ indica o nome da atividade associada ao evento e, $\#_{tempo}(e) \in \mathcal{T}$ indica a data/hora de execução registrada no evento e e $\#_{atividade}(e) \in \mathcal{R}$ indica o responsável pela realização da atividade registrada no evento e.

Definição 14 (Caminho). Um caminho compreende uma sequência finita de eventos distintos que seguem uma ordem temporal não decrescente em relação ao seu atributo tempo. Logo, seja $\boldsymbol{t} = \langle e_1, e_i, e_j, e_{|\boldsymbol{t}|} \rangle$ um caminho e $e_i, e_j \in \boldsymbol{t}$ eventos, então $\#_{tempo}(e_i) \leq \#_{tempo}(e_i)$.

Definição 15 (Log de Eventos (VAN DER AALST, 2016)). Um log de eventos $\mathcal{L}$ é formalmente definido através da quádrupla (4-tupla) $\mathcal{L} = (\mathrm{E}, \Sigma, \#, \mathcal{E})$, na qual:

- E é um conjunto não vazio de identificadores únicos de eventos,
- Σ é um conjunto finito não vazio de nomes de atividades,
- $\#$ o valor atribuído para os atributos em um evento,
- $\mathcal{E} \subseteq \mathrm{E}^*$ consiste no conjunto finito de caminhos em E.

2.4.2 Log de eventos na mineração de processos

O livro *Process Mining: Data Science in Action* (VAN DER AALST, 2016) apresenta um panorama do papel do log de eventos para a mineração de processos:

> *A mineração de processos é impossível sem logs de eventos adequados. (...) Dependendo da técnica de mineração de processos usada, esses requisitos podem variar. O desafio é extrair esses dados de uma variedade de fontes de dados, por exemplo, bancos de dados, arquivos simples, logs de mensagens, logs de transações, sistemas ERP e sistemas de gerenciamento de documentos. Ao combinar e extrair dados, a sintaxe e a semântica desempenham um papel importante. Além disso, dependendo das perguntas que se procura responder, são necessárias visões diferentes sobre os dados disponíveis. A mineração de processos, como qualquer outra abordagem de análise orientada a dados, precisa lidar com problemas de qualidade de dados.*

Do fragmento textual apresentado fica evidente que não há mineração de processos sem log de eventos, mas a sua extração ou montagem oferece desafios. Essa ideia é corroborada no livro *Process Mining in Action Principles, Use Cases and Outlook* (BOENNER, 2020).

> *(...) Embora os logs de eventos relevantes sejam determinados pelo objetivo, ou seja, quais atividades devem ser consideradas, a identificação dos logs de eventos nos sistemas de origem pode ser o maior fator de esforço. Não apenas requer uma compreensão clara de qual log está armazenado na base de dados, mas também o acesso a todos os dados necessários. Em grandes organizações globais, trilhas digitais podem ser distribuídas por várias fontes de TI diferentes, incluindo sistemas ERP e não ERP.*

Os principais desafios para a extração de logs de eventos foram sumarizados da seguinte forma (VAN DER AALST, 2016):

1. Correlação: Eventos em um log de eventos são agrupados por caso. Esse requisito simples pode ser bastante desafiador caso os eventos estejam espalhados em sistemas diferentes. Neste caso, relacionar eventos registrados em sistemas distintos pode não ser trivial.
2. *Timestamp*: Os eventos precisam ser ordenados por caso. Em princípio, não é exigido registros de data e hora. No entanto, ao combinar dados de fontes diferentes, normalmente é necessário depender de registros de data e hora para classificar eventos (em ordem de ocorrência).
3. *Snapshot*: Os casos podem ter uma duração que se estende além do período gravado, por exemplo, um caso foi iniciado antes do início do log de eventos ou ainda estava em execução quando a gravação parou. Portanto, é importante perceber que os logs de eventos geralmente fornecem apenas uma foto de um processo que se encontra em execução.
4. Escopo: Os sistemas de informações corporativas podem ter milhares de tabelas com dados relevantes para os negócios. Como decidir quais tabelas incorporar? É necessário conhecimento de domínio para localizar os dados necessários. Obviamente, o escopo desejado depende dos dados disponíveis e das perguntas que precisam ser respondidas.

5. Granularidade: Em muitos sistemas, os eventos no log de eventos têm um nível de granularidade diferente das atividades relevantes para os usuários finais. Alguns sistemas produzem eventos de baixo nível que são detalhados demais para serem apresentados às partes interessadas em gerenciar ou melhorar o processo. Nessa tese oferecemos uma proposta para lidar com o problema da granularidade.

O fato é que as informações que se prestam à mineração de processos muitas vezes se encontram “escondidas” e/ou espalhadas nos sistemas de informações. Assim, em muitos ambientes tecnológicos, a mineração de processos requer um trabalho prévio de ETL[3] (*Extract, Transform and Load)* para aquisição de dados adequados.

Por outro lado, a despeito da maior ou menor complexidade para se obter o log de eventos, em geral, as informações necessárias para construir logs de eventos básicos (suficiente para grande parte das técnicas de mineração) estão disponíveis nos sistemas de informações. Assim, existe um “custo inicial” que pode ser alto, mas o risco de o trabalho ser infrutífero é baixo. Além disso, uma vez “desbravado o caminho” a extração de novos logs de eventos se torna trivial para as análises subsequentes.

A qualidade dos dados no log de eventos é uma questão relevante para a mineração de processos que está intimamente ligada aos dados disponíveis para construção do log de eventos. Como a questão da qualidade dos logs de eventos corresponde ao objeto desta tese, abordaremos esta questão mais detalhadamente nos próximos capítulos. Neste ponto vamos avançar na apresentação dos elementos que constituem os logs de eventos e a sua representação.

Como apresentado, um log de eventos consiste em um conjunto de eventos vinculado às instâncias do processo. Uma instância de processo também é chamada de caso. Nesse contexto, os eventos consistem no conjunto de dados que caracterizam a realização de uma determinada atividade do processo.

Não existe um padrão que determine as informações que um evento deve dispor, essa questão está relacionada às perguntas que se deseja responder. Contudo, é de

[3] ETL é definido no contexto de Business Intelligence ou mineração de dados como o processo que compreende três etapas: extração de dados em fontes externas, transformação dos dados para atender aos objetivos e carga dos dados no sistema de destino.

se esperar que um log de eventos ao menos indique a instância do processo que se refere, a atividade realizada e quando foi executada (data/hora da execução). A identificação da instância do processo é essencial para que se possa agrupar os eventos em caminhos e a data/hora da execução para se estabelecer uma noção de ordem destes. A identificação da atividade não é obrigatória, embora seja incomum um log de eventos que não indique as atividades realizadas. Na ausência da identificação da atividade, o evento deve dispor de um atributo alternativo (ex.: pessoa que executou ou unidade organizacional) para que se tenha algum contexto sobre sua execução. Do contrário, tem-se um conjunto inútil de informações para fins de mineração de processos. Portanto, vamos aqui considerar como atributos básicos de um evento a identificação do caso e da atividade, bem como a data/hora de execução da atividade. É importante ter em mente que um log de eventos básico habilita realizar a mineração processos para alguns objetivos, contudo logs de eventos mais ricos podem oferecer outras perspectivas de análise, conforme veremos no capítulo seguinte.

Um aspecto que vem ganhando cada vez mais relevância diz respeito à proteção de dados pessoais. Em 25/05/2018, entrou em vigor na União Europeia o “Regulamento Geral de Proteção de Dados”, conhecido como GPDR, sua sigla em inglês, que estabelece regras sobre como as empresas e os órgãos públicos devem lidar com os dados pessoais. Inspirado na GPDR, foi criada no Brasil a Lei Geral de Proteção de Dados (LGPD, Lei nº 13.709/2018). Portanto, a adequação aos preceitos legais é um novo aspecto que deve ser observado em projetos de mineração de processos que utilizem dados pessoais.

A Tabela 2 apresenta um fragmento de um log de eventos. É possível observar que cada linha apresenta um evento e que as colunas correspondem aos atributos do evento. Podemos perceber que os atributos básicos estão presentes (identificação da instância/caso, atividade e data/hora de início da execução), bem como a existência de atributos adicionais (responsável pela execução e a classe do evento).

Existem padrões para armazenamento de logs de eventos. O primeiro padrão estabelecido para log de eventos foi o MXML - Mining eXtensible Markup Language, que surgiu em 2003 e foi posteriormente adotado pela ferramenta de mineração de processos ProM. O XES - eXtensible Event Stream (IEEE COMPUTATIONAL INTELLIGENCE SOCIETY, 2016; VERBEEK et al., 2010) é o atual padrão de para log de evento. Sucessor do MXML, foi pensado a partir das limitações observadas nas

experiências práticas com o MXML. Esse novo formato foi projetado para ser menos restritivo e verdadeiramente extensível. A maior parte das ferramentas de mineração de processos suporta o formato XES.

Tabela 2 – Fragmento de log de eventos

Caso	Atividade	Início	Responsável	Classe
657797	Citar (Inicial)	2016-06-01 07:25:09	56234	319
657797	Finalizar Ato	2016-12-21 09:46:32	7199	319
657797	Arquivo definitivo	2016-12-21 09:46:34	NULO	319

O XES suporta cinco tipos de dados principais: $String$, $Date$, Int, $Float$ e $Boolean$, correspondentes aos tipos de dados XML padrão. A semântica dos atributos é especificada por meio de extensões. O XES define cinco extensões padrão:

- $Concept$ (conceito): definida tanto para caminhos quanto para eventos, captura seus nomes. Para caminhos, normalmente representa o identificador de caso, enquanto para eventos representa o nome da atividade.
- $Time$ (tempo): definida para eventos, captura sua data/hora, correspondendo ao $\#_{tempo}(e)$ para todo $e \in \mathcal{E}$.
- $Organization$ (organização): definida para eventos, captura a perspectiva organizacional de um processo. É composta pelos atributos recurso, papel e grupo; sendo respectivamente, o responsável pelo evento, seu papel na organização e o grupo ao qual pertence dentro da organização, ex.: um departamento.
- $Lifecycle$ (ciclo de vida): definida para eventos, indica o *status* do evento, ex.: inicial, completo, suspenso.
- $Semantic$ (semântica): define um modelo de referência para todos os elementos do log de eventos.

A despeito da existência de um formato próprio para armazenamento de log de eventos reconhecido pelas principais ferramentas de mineração de processos, logs de eventos no formato de arquivos separados por vírgula (CSV) ou até em planilhas do Excel também são amplamente aceitos como entrada de dados. A Figura 9

apresenta um fragmento de log de eventos no formato XES.

Figura 9 – Fragmento de log de evento no formato XES

```
<?xml version="1.0" encoding="UTF-8"?>
<log xes.version="1.0" xmlns="http://www.xes-standard.org">
   <extension name="Concept" prefix="concept" uri="http://.../concept.xesext"/>
   <extension name="Lifecycle" prefix="lifecycle" uri="http://.../lifecycle.xesext"/>
   <extension name="Time" prefix="time" uri="http://.../time.xesext"/>
   <extension name="Organizational" prefix="org" uri="http://.../org.xesext"/>
   <global scope="trace">
      <string key="concept:name" value="name"/>
   </global>
   <global scope="event">
      <string key="concept:name" value="name"/>
      <string key="lifecycle:transition" value="transition"/>
      <string key="org:resource" value="resource"/>
      <date key="time:timestamp" value="2018-01-07T12:44:30.415-03:00"/>
      <string key="ID_CLASSE" value="string"/>
   </global>
   <classifier name="Activity" keys="NAME_TASK_INSTANCE"/>
   <trace>
      <string key="concept:name" value="657797"/>
      <event>
         <string key="concept:name" value="Citar (Inicial)"/>
         <string key="lifecycle:transition" value="start"/>
         <string key="org:resource" value="56234"/>
         <date key="time:timestamp" value="2016-06-01T07:25:09.000-03:00"/>
         <string key="ID_CLASSE" value="319"/>
      </event>
      <event>
         <string key="concept:name" value="Finalizar Ato"/>
         <string key="lifecycle:transition" value="complete"/>
         <string key="org:resource" value="7199"/>
         <date key="time:timestamp" value="2016-12-21T09:46:32.652-03:00"/>
         <string key="ID_CLASSE" value="319"/>
      </event>
      <event>
         <string key="concept:name" value="Arquivo definitivo"/>
         <string key="lifecycle:transition" value="start"/>
         <string key="org:resource" value="NOT_SET"/>
         <date key="time:timestamp" value="2016-12-21T09:46:34.000-03:00"/>
         <string key="ID_CLASSE" value="319"/>
      </event>
        ⋮
   </trace>
      ⋮
</log>
```

2.5 Árvore de Decisão

Árvore de Decisão é uma técnica, utilizada nesta tese, que usa uma abordagem de "dividir e conquistar" para aprender com um conjunto de instâncias independentes (WITTEN; FRANK, 2005). Ou seja, consiste em uma abordagem de aprendizagem de máquina baseada em treinamento supervisionado.

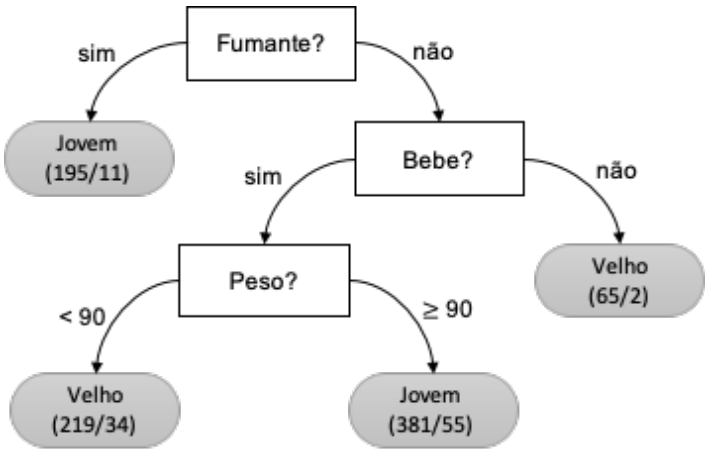
Uma árvore de decisão é composta por nós e folhas. Os nós testam um atributo específico sobre um valor de referência para o atributo examinado. As folhas informam a classificação que se aplica a todas as instâncias que as alcançam. Uma nova instância é classificada percorrendo-se a árvore a partir de sua parte superior testando sucessivamente os nós até atingir uma folha. A instância será classificada de acordo

com a classe atribuída pela folha à qual alcançou.

Os principais algoritmos usados para gerar árvores de decisão são: ID3 *(Dichotomizer Iterative),* criado por J. Ross Quinlan entre o final da década de 1970 e o início da década de 1980; e C4.5, também criado por Quinlan como sucessor do ID3, que se tornou o algoritmo padrão para comparação com novos algoritmos de indução de árvore de decisão (HAN; KAMBER; HARDCOVER, 2011).

A Figura 10 apresenta uma árvore de decisão derivada de uma base de dados para estudo dos efeitos do consumo de álcool, tabagismo e peso corporal na expectativa de vida. A base de dados utilizada contempla informações de 860 pessoas falecidas. As que faleceram antes dos 70 anos de idade são classificadas como “Jovem” e as que faleceram com 70 anos ou mais são classificadas como “Velho”. (VAN DER AALST, 2016)

Figura 10 – Árvore de decisão (VAN DER AALST, 2016)



Usualmente as árvores de decisão são usadas na tarefa de classificação, mas, nessa tese, a técnica foi empregada para estimar os parâmetros da abordagem apresentada no Capítulo 5. A escolha da abordagem foi motivada pelo fato da abordagem fornecer um resultado explicável, algo importante para os propósitos desta tese.

2.6 Métricas de similaridades entre strings[4]

A similaridade de Jaro-Winkler é uma medida de semelhança entre duas strings. Nesta tese, a métrica é utilizada para comparar os nomes de atividades do processo.

[4] Strings são cadeias de caracteres

Nessa seção apresentaremos a similaridade de Jaro-Winkler, bem como sua predecessora, similaridade de Jaro.

2.6.1 Jaro e Jaro-Winkler

A similaridade de Jaro é uma métrica para quantificar a similaridade entre duas *strings*. Seu funcionamento consiste em identificar os caracteres iguais nas strings comparadas, diferenciando as que ocupam a mesma posição das que estão deslocadas. Já a similaridade Jaro-Winkler (WINKLER, 1990) é uma extensão da similaridade de Jaro. Winkler observou que os erros de digitação ocorrem mais frequentemente no meio ou final da palavra, mas raramente no início (DRESSLER; NGOMO, 2017). Então, propôs a incorporação dessa ideia na abordagem de Jaro. As medidas de similaridade de Jaro e Jaro-Winkler são definidas como segue:

Definição 16 (Similaridade de Jaro). Seja Σ^* o conjunto de todas as strings possíveis em Σ. Seja s_i uma string, $|s_i|$ denota o tamanho da string s_i. A similaridade de Jaro $sim_j = \ \Sigma^* \times \Sigma^* \rightarrow [0,1]$ é definida assim:

$$sim_j = \begin{cases} 0, & se\ m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right), & caso\ contrário \end{cases}$$

- $|s_1|, |s_2|$ é o tamanho das strings que estão sendo comparadas (s_1 e s_2).
- m é quantidade de caracteres que combina, ou seja, são iguais e que ocupam a mesma posição ou uma posição diferente dentro da distância máxima (w).
- t é a metade do número de transposições, que consiste na existência de caracteres iguais em posições diferentes (dentro da distância máxima w).

A distância máxima (w) entre dois caracteres para a transposição seja considerada válida é definida como segue:

$$w = \left\lfloor \frac{max(|s_1|, |s_2|)}{2} \right\rfloor - 1$$

Definição 17 (Similaridade de Jaro-Winkler). Seja ℓ o tamanho do prefixo e p o fator peso, ambos parâmetros operacionais. A similaridade de Jaro-Winkler sim_w é definida da seguinte forma:

$$sim_w = sim_j + \ell p(1 - sim_j)$$

O exemplo a seguir demonstra o passo-a-passo do algoritmo Jaro-Winkler para cálculo da similaridade entre as strings “PROCESSO” e “PROCEDIMENTO”. Usaremos os valores usuais $\ell = 0{,}1$ e $p = 4$.

A Tabela 3 mostra a posição dos caracteres em cada string na primeira linha, nas duas linhas seguintes as strings s_1 e s_2 e na última linha o resultado para cada par de caracteres.

Tabela 3 – Comparação entre strings

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$\boldsymbol{s_1}$	P	R	O	C	E	S	S	O				
$\boldsymbol{s_2}$	P	R	O	C	E	D	I	M	E	N	T	O
	=	=	=	=	=	≠	≠	t_{8-12}	≠	≠	≠	t_{8-12}

Os possíveis resultados para cada par de caractere avaliados são:

- (=) quando os caracteres que ocupam a mesma posição nas respectivas *strings* são iguais;
- (t_{i-j}) quando existem caracteres iguais que ocupam posições diferentes e estão a uma distância menor que w. Sendo i a posição do caractere na string s_1 e j a posição do caractere na string s_2;
- (≠) quando não existe correspondência entre o caractere de s_1 como algum caractere de s_2 dentro da distância w;

As strings s_1, s_2 têm respectivamente os tamanhos $|s_1| = 8$ e $|s_2| = 12$. De acordo com o resultado observado na última linha da Tabela 3, existem 5 pares de caracteres iguais que ocupam a mesma posição (de 1 até 5) e existe um par de caracteres iguais, mas ocupam posições diferentes (caractere “O” nas posições 8 e

12). Antes de considerar uma transposição faz-se necessário verificar se a distância entre os dois caracteres está dentro da distância máxima (w).

$$w = \left\lfloor \frac{max(|s_1|, |s_2|)}{2} \right\rfloor - 1 = \left\lfloor \frac{max(8, 12)}{2} \right\rfloor - 1 = \left\lfloor \frac{12}{2} \right\rfloor - 1 = 5$$

Como a distância entre os caracteres (8 para 12) é menor que a distância máxima ($w = 5$), podemos considerar a transposição t_{8-12} como válida. Assim, temos todos os elementos necessários para calcular o valor da similaridade de Jaro: $m = 6$ e $t = 0{,}5$ (metade das transposições). Logo:

$$sim_j = \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m - t}{m}\right) = \frac{1}{3}\left(\frac{6}{8} + \frac{6}{12} + \frac{6 - 0{,}5}{6}\right) = 0{,}72$$

Com base no valor obtido para a similaridade Jaro e os valores dos parâmetros ℓ e p, o valor da similaridade Jaro-Winkler é calculado como segue:

$$sim_w = sim_j + \ell p(1 - sim_j) = 0{,}72 + 0{,}1 * 4 * (1 - 0{,}72) = 0{,}83$$

Logo, a medida de similaridade (Jaro-Winkler) entre as palavras “PROCESSO” e “PROCEDIMENTO” tem o valor de 0,83; indicando uma alta similaridade entre as palavras.

3 MINERAÇÃO DE PROCESSOS

Neste capítulo apresentaremos uma visão abrangente sobre a mineração de processos, contemplando seus objetivos, técnicas e ferramentas. Para tanto, o capítulo foi organizado como segue. Na Seção 3.1 são apresentados os objetivos da mineração de processos, bem como as principais técnicas nos diferentes objetivos. Na Seção 3.2 é apresentado o arsenal de ferramentas disponíveis para o usuário da mineração de processos. Na Seção 3.3 são abordados os aspectos relacionados à qualidade dos modelos descobertos através da mineração de processos. Por fim, na Seção 3.4, os princípios norteadores e principais desafios da mineração de processos que se relacionam com essa tese são introduzidos.

3.1 Objetivos da mineração de processos

Como visto no capítulo anterior, os logs de eventos são o ponto de partida para as técnicas de mineração de processos, que utilizam esses dados factuais de execução para analisar a execução real dos processos (VAN DER AALST et al., 2012). Usando técnicas de mineração de processos, é possível: (1) extrair automaticamente modelos dos fluxos reais do processo através das “pegadas” nos sistemas de suporte aos processos; (2) detectar desvios dos procedimentos documentados; e (3) enriquecer os modelos existentes com informações relevantes. A mineração de processos possui uma ampla variedade de técnicas que podem ser classificadas em três categorias (VERBEEK et al., 2010):

- Descoberta, que usa logs de eventos registrados por um ou mais sistemas de informações para descobrir o comportamento real do processo.
- Conformidade, que compara o modelo de processo formal com o processo descoberto por meio de logs de eventos gerados durante a execução de diferentes instâncias do processo. Essa comparação possibilita mostrar onde o modelo formal de processo falha ao representar o processo real. Ou seja, ele pode determinar se a realidade, conforme registrada nos logs, está em conformidade com o modelo.

- Aperfeiçoamento, que usa informações registradas pelo processo real nos logs de eventos para melhorar o próprio processo.

As seções seguintes apresentam as principais técnicas para cada uma das categorias da mineração de processos.

3.1.1 Descoberta

Esse tipo de mineração de processo é o precursor e o mais explorado (GARCIA et al., 2019). O desenvolvimento de algoritmo para descoberta automática de processos, ou seja, a descoberta de modelos de processos baseados em eventos observados, iniciou na década de 90 (AGRAWAL; GUNOPULOS; LEYMANN, 1998; COOK; WOLF, 1998). Em 2012, foi publicado uma pesquisa (CLAES; POELS, 2012) indicando os algoritmos *α*/*α*++, *Heuristics miner*, *Genetic Miner* e *Fuzzy miner* como os principais algoritmos de descoberta utilizados[5] no início desta década. A ordem dos algoritmos aqui apresentada segue apenas uma lógica cronológica, não considerando o grau de popularidade das técnicas.

Em geral, a notação e o algoritmo de descoberta de processo são fortemente acoplados (WEIJTERS; VAN DER AALST; MEDEIROS, 2006), pois, dependendo da abordagem, algumas anotações internas têm mais sentido do que outras. Além disso, a capacidade de um modelo de processo para expressar determinado comportamento só é útil se o algoritmo de descoberta do processo for capaz detectar tal comportamento. Portanto, uma notação de modelagem de processo com menos poder expressivo não é uma má ideia se o algoritmo de descoberta de processo não possuir capacidade para perceber comportamentos mais complexos. Ou seja, a notação de modelagem de processo escolhida deve ser adequada para o algoritmo de descoberta de processo e vice-versa. A seguir apresentamos uma visão geral sobre os principais algoritmos.

O primeiro algoritmo a possibilitar a descoberta automática de processos tratando a questão da concorrência entre as atividades foi algoritmo *α*, descrito pela primeira vez em (AALST; WEIJTERS; MARUSTER, 2004). O *α* é um dos algoritmos de descoberta de processos mais básicos, não possuindo parâmetros de operação.

[5] A pesquisa identificou os principais plug-ins utilizados no ProM (ver Seção 3.3) que foi apontado como principal ferramenta de mineração de processos no período da pesquisa.

O seu resultado é representado na forma de uma rede de Petri. Algumas variações foram propostas para tornar o algoritmo mais robusto, dentre elas o *α++.*

Atualmente, o algoritmo não se destina a aplicações práticas da mineração de processos, pois tem problemas na presença de comportamento pouco frequente ou modelagem de comportamentos complexos. No entanto, este algoritmo fornece uma excelente oportunidade para entender como os algoritmos de descoberta de processos operam, uma vez que é muito simples e grande parte das suas ideias foram incorporadas em técnicas mais modernas.

O algoritmo explora o log de eventos buscando padrões específicos; para tanto, baseadas na ordem de execução, classifica as relações entre as atividades em quatro tipos, como definidas a seguir.

Seja $a, b \in \mathcal{L}$ atividades do log de eventos $\mathcal{L}$, $a >_L b$ denota uma relação de ordem onde a atividade a é sucedida pela atividade b. As relações entre as atividades são definidas da seguinte forma:

- $a \rightarrow b$, sse $\exists a >_L b \wedge \nexists b >_L a$
- $a \leftarrow b$, sse $\nexists a >_L b \wedge \exists b >_L a$
- $a \# b$, sse $\exists a >_L b \wedge \exists b >_L a$
- $a \parallel b$, sse $\nexists a >_L b \wedge \nexists b >_L a$

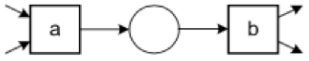
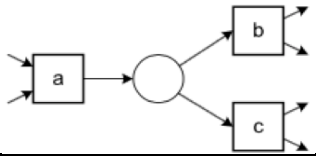
Com base nas relações de ordem entre as atividades, o algoritmo monta uma matriz de “pegadas”. A Tabela 4 mostra uma matriz de “pegadas” construída a partir do log de eventos $\mathcal{L} = [\langle a, b, c, d\rangle^3, \langle a, c, d, b\rangle^4, \langle a, b, c, e, f, b, c, d\rangle^2, \langle a, b, c, e, f, c, b, d\rangle, \langle a, c, b, e, f, b, c, d\rangle^2, \langle a, c, b, e, f, b, c, e, f, c, b, d\rangle]$.

Tabela 4 – Exemplo de matriz de pegadas

	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{c}$	$\boldsymbol{d}$	$\boldsymbol{e}$	$\boldsymbol{f}$
$\boldsymbol{a}$	#	→	→	#	#	#
$\boldsymbol{b}$	←	#	‖	→	→	←
$\boldsymbol{c}$	←	‖	#	→	→	←
$\boldsymbol{d}$	#	←	←	#	#	#
$\boldsymbol{e}$	#	←	←	#	#	→
$\boldsymbol{f}$	#	→	→	#	←	#

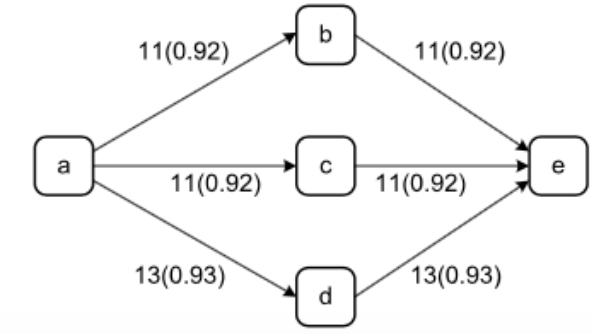
A partir da matriz de “pegadas” o algoritmo α associa os padrões encontrados a estruturas pré-definidas de modelo do processo. A Tabela 5 mostra duas estruturas pré-definidas de modelo do processo utilizadas pelo algoritmo. Após identificar as estruturas individuais, o algoritmo as conecta de forma a obter um modelo único que represente todo o processo.

Tabela 5 – Exemplo de estruturas pré-definidas do algoritmo α

PADRÃO	ESTRUTURA PRÉ-DEFINIDA
$< a \rightarrow b >$	
$< a \rightarrow b + a \rightarrow c + b \nparallel c >$	

Posteriormente, um novo algoritmo denominado de *Heuristic Miner*, projetado para tratar melhor que seus predecessores o comportamento pouco frequente encontrado nos logs de eventos, foi apresentado em (WEIJTERS; VAN DER AALST, 2003). O algoritmo se baseia na construção de um grafo de dependências entre atividades, conforme apresentado na Figura 11. As relações de dependências fracas (valores inferiores a um *threshold*) não são inseridas no grafo de dependência, proporcionando assim a eliminação do comportamento infrequente do modelo de processo resultante.

Figura 11 – Grafo de dependência entre atividades no *Heuristic* Miner



O grafo de dependências é montado através de uma matriz com as medidas de dependências entre as atividades de um log de eventos. A medida de dependência entre atividades no *Heuristic Miner* é calculada conforme Definição 18.

Definição 18 (Medida de dependência entre atividades no *Heuristic Miner*). Seja $a, b \in \mathcal{L}$ atividades do log de eventos $\mathcal{L}$. Seja $|a \succ b|$ a quantidade de transições entre as atividades a e b. $|a \Rightarrow_L b|$ denota a medida de dependência entre atividades a e b, sendo definida como segue:

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a \succ b| - |b \succ a|}{|a \succ b| + |b \succ a| + 1}, & se\ a \neq b \\ \dfrac{|a \succ a|}{|a \succ a| + 1}, & se\ a = b \end{cases}$$

A representação convencional para modelos de processos descobertos através do *Heuristic Miner* é o C-Net (Figura 6). Contudo, existem implementações alternativas que geram modelos em outros formatos, como por exemplo redes de Petri.

O *Genetic Miner* (DE MEDEIROS; WEIJTERS; VAN DER AALST, 2007) consiste em uma abordagem para descoberta de processos baseadas em algoritmos genéticos visando reduzir a limitação dos algoritmos quando em face de estruturas não-triviais e ruídos nos dados. A ideia da abordagem é aproveitar a robustez "natural" a ruídos, bem como outros aspectos dos algoritmos genéticos (DE MEDEIROS; WEIJTERS; VAN DER AALST, 2007). Assim como os demais algoritmos apresentados nessa seção, o *Genetic Miner* foi implementado como um *plug-in* do ProM. Este algoritmo foi popular no ProM 5, mas perdeu popularidade no ProM 6 com a chegada dos novos algoritmos de descoberta de processos (CLAES; POELS, 2012).

O *Fuzzy Miner* (GÜNTHER; VAN DER AALST, 2007) foi criado para atacar os problema da aplicação da mineração de processos em ambientes reais, para tanto, foi projetado sob as seguintes premissas: (1) nem todos os logs são confiáveis e (2) nem sempre existe um processo exato que é refletido no log. A construção do modelo de processo no *Fuzzy Miner* é semelhante ao *Heuristic* Miner. As diferenças entre as abordagens se concentram na forma de interação com o usuário na representação do modelo de processo. O *Fuzzy Miner* propõe uma abordagem exploratória, na qual o usuário pode enxergar o processo em diferentes níveis de abstração de acordo com os parâmetros escolhidos. Essa forma de interação com usuário se tornou o padrão para as ferramentas comerciais de mineração de processos.

Novas abordagens para descoberta de processos foram propostas, a exemplo da descoberta de processos baseada em regiões, do algoritmo *Inductive Mining* e do

algoritmo *Fodina* (evolução do *Heristic Miner)*. Em 2017, uma revisão da literatura (AUGUSTO et al., 2017) encontrou 86 estudos abordando técnicas para descoberta de processos no período compreendido entre os anos de 2011 e 2017. Muitos dos estudos identificados se referiam aos mesmos métodos de descoberta de processo (extensões, otimização, preliminares ou generalização de outro estudo). Diante disso, os autores agruparam os estudos e com isso chegaram a 35 grupos principais de algoritmos de descoberta.

Outro estudo recente apresentou um mapeamento da literatura (GARCIA et al., 2019) na área. Além das técnicas, o estudo mapeou as principais áreas de aplicações da mineração de processos. Dos 572 estudos identificados, a maior parte (28,03%) se concentra área da saúde, seguida por: TIC (16,44%), Indústria (13,32%), Educação (10,55%), Finança (6,40%), Logística (4,67%) e Setor público (4%).

3.1.2 Verificação de Conformidade

Atualmente, grande parte das organizações documenta seus processos de alguma forma, quer seja para fins de atendimento a normas (regulação ou certificação) ou mesmo treinamento dos colaboradores. As técnicas de verificação conformidade permitem confrontar o modelo descoberto através da mineração de processos com o modelo documentado para identificar divergências, podendo ser usada para (VAN DER AALST, 2012):

- Avaliar se os processos documentados descrevem a realidade com precisão;
- Identificar casos divergentes e entender o que eles têm em comum;
- Apoiar a atividade de auditoria através da identificação das partes do processo em que há maior incidência de desvios;
- Avaliar a qualidade de um modelo de processo descoberto.

Existem duas famílias de abordagens de verificação de conformidade: baseada em reprodução baseada em tokens e alinhamento de caminhos. Em uma reprodução baseada em *tokens* cada instância do processo contida em um log de eventos é submetida a uma rede de Petri e uma métrica de conformidade é aplicada. Por exemplo, em (ROZINAT; WILL M.P. VAN DER, 2008) durante a reprodução de uma instância do processo, caso se chegue a uma situação em que não se consegue

avançar, um *token* adicional é gerado para que se possa seguir para o próximo estado. Ao final da execução são computados os tokens que sobraram, bem como os que foram artificialmente adicionados para possibilitar a execução completa. Existem outras abordagens para realização de verificação de conformidade baseada na reprodução de tokens. Uma importante limitação dessas abordagens é que requer um modelo na forma de uma rede de Petri, exigindo a conversão para os casos em que o modelo esteja representado em outra notação.

A abordagem de alinhamentos foi introduzida para oferecer uma alternativa mais flexível à reprodução baseada em *tokens*. A Figura 12 apresenta dois exemplos (γ_1 e γ_2) nos quais os caminhos extraídos de um log de eventos (parte superior) são comparados com um caminho ajustado para o modelo do processo (parte inferior). No primeiro exemplo (γ_1) podemos observar que existe um alinhamento entre o caminho executado e o modelo, uma vez que não foi necessário qualquer ajuste. Já no segundo podemos identificar discrepâncias denotadas por ≫. Ao computar as discrepâncias é possível obter insights sobre a conformidade do modelo (DUNZER et al., 2019). Cabe registrar que as abordagens baseadas em alinhamentos geram alto custo computacional.

Figura 12 – Exemplo de alinhamento de caminhos (DUNZER et al., 2019)

$$\gamma_1 = \begin{array}{|c|c|c|c|c|c|} \hline a & g & c & f & e & h \\ \hline a & g & c & f & e & h \\ \hline \end{array}$$

$$\gamma_2 = \begin{array}{|c|c|c|c|c|c|} \hline a & \gg & d & b & e & h \\ \hline a & b & d & \gg & e & h \\ \hline \end{array}$$

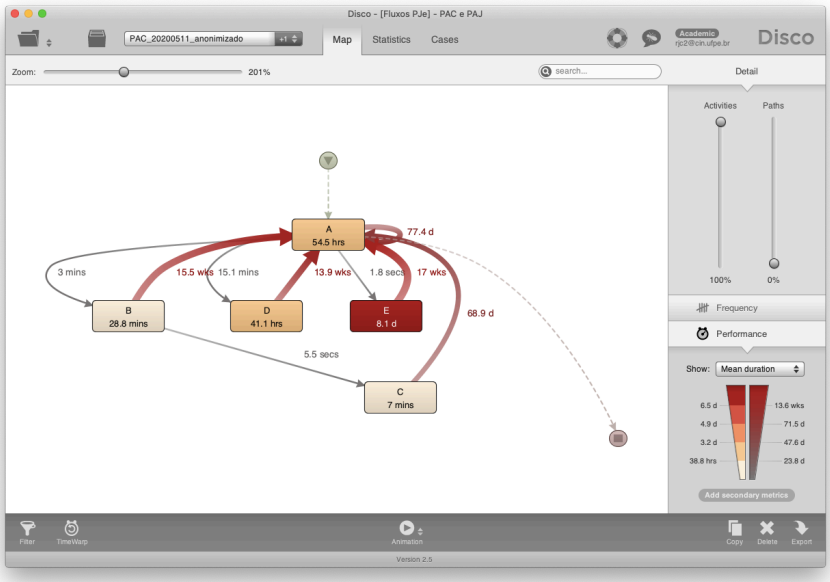
Em geral, as técnicas de verificação de conformidade oferecem como saída uma medida de qualidade do modelo, tais como: fitness, precisão ou outras. Na seção 3.3 abordaremos as métricas de qualidade de modelos de processos.

3.1.3 Aprimoramento

Esse tipo de atividade de mineração de processo visa em estender o modelo de processo descoberto com informações relevantes. Em analogia aos aplicativos móveis de GPS que combinam informações *online* sobre tráfego com os mapas, dando ênfase às ruas congestionadas, a mineração de processos pode combinar os

dados de data e hora dos eventos com o modelo de processo para oferecer previsões usando estatísticas ou aprendizado de máquina (GARCIA et al., 2019). Esse tipo de mineração de processos pode dar suporte a questões sobre a duração de um caso, identificar as atividades e transições mais demoradas (gargalos), quais são os recursos críticos e outras. Em geral, as ferramentas de mineração de processos incorporam essas informações aos modelos descobertos realçando em cores as atividades ou apresentando os arcos que representam as transições com espessura variável a partir de alguma métrica escolhida. A Figura 13 mostra a tela da ferramenta comercial Disco. Pode-se observar a diferença de tons de vermelho nas atividades (retângulos) e as diferentes espessuras de arcos (transições) indicando faixas de duração das atividades e transições.

Figura 13 – Exemplo de mineração de dados do tipo aprimoramento no Disco



3.2 *Ferramentas*

As ferramentas de mineração de processos amadureceram ao longo da última década. Atualmente, existem diversas ferramentas acadêmicas e comerciais que disponibilizam uma gama de funcionalidades de mineração de processos. Em 2012, uma pesquisa posicionou o ProM como a ferramenta de mineração de processos mais popular tanto para pesquisa quanto para uso em aplicações práticas (CLAES; POELS, 2012). Desde então, muitas ferramentas surgiram, mas o ProM permanece popular, sobretudo pela sua abrangência.

O ProM, desenvolvido pelo *Process Mining Group*, é um *framework* extensível que suporta uma ampla variedade de técnicas para mineração de processos na forma de plug-ins (VAN DONGEN et al., 2005). É um software de código aberto que dispõe de grande parte das técnicas de mineração de processos descritas na literatura. A primeira versão totalmente funcional do ProM (v1.1) foi lançada em 2004 contendo 29 *plugins*. Em 2006, o ProM 4.0 já contemplava 142 *plugins* (AALST, 2011). Desde então, a quantidade de plug-ins saltou para quase 300 na versão 5.2 (lançada em 2009) e passou para a casa dos 400 na versão v6.6. Esse crescimento comprova a importância para a comunidade de mineração de processos. O ProM não é a única ferramenta de código aberto disponível, existe a Apromore (ROSA et al., 2011) também com viés acadêmico que suporta modelos em diferentes notações.

Recentemente as abordagens de mineração de processos estão sendo disponibilizadas em plataformas que favorecem uma maior integração com as demais abordagens das ciências de dados. No final de 2019 foi lançada uma biblioteca em *Python* para mineração de processos denominada de *Process Mining for Python* (PM4Py). A ideia consiste em oferecer uma abordagem algorítmica para a mineração de processos integrada às bibliotecas de ciência de dados mais modernas (BERTI et al., 2019). De maneira similar foi disponibilizada uma extensão comunitária para a plataforma KNIME[6], denominada PM4KNIME, que contempla alguns dos principais plug-ins do ProM.

Em relação a ferramentas comerciais, nos últimos anos o mercado entrou em franca ascensão, conforme pode ser observado no relatório publicado pelo Gartner em 2019 (MARC KERREMANS, 2019):

> *Em 2018, a estimativa do mercado de mineração de processos do Gartner para receita de licença e manutenção de novos produtos estava chegando a 160 milhões de dólares. O mercado dos EUA entrou no jogo em 2018. Muitas organizações tomaram conhecimento dos benefícios da mineração de processos e, no ano passado, projetamos que esse mercado poderia facilmente triplicar ou quadruplicar de tamanho durante os próximos dois anos. No entanto, como a maioria dos fornecedores não está alcançando rápido o suficiente com a enorme demanda, esperamos um atraso na realização desse crescimento.*

[6] https://www.knime.com/

O referido relatório apontou a existência de 19 fornecedores de soluções de mineração de processos (incluindo os mantenedores do Apromore e ProM). Dentre as ferramentas comerciais podemos destacar as seguintes: Disco[7], MyInvenio[8], Celonis Snap[9] e Everflow[10]. Esta última desenvolvida por uma empresa brasileira.

As ferramentas comerciais se apresentam como boas alternativas para reduzir a barreira de implantação da mineração de processos nas organizações, uma vez que trabalham melhor aspectos como usabilidade e desempenho. Contudo, é importante frisar que estas ferramentas não oferecem ganhos do ponto de vista funcional, apenas implementam *plugins* existentes em ferramentas como ProM tornando-os mais adequados para uso profissional.

3.3 Qualidade dos modelos de processos

Os algoritmos de descoberta de processos têm como objetivo descobrir modelos de processos a partir de logs de eventos que reflitam com a maior fidedignidade possível o comportamento registrado. Contudo, avaliar a qualidade de um modelo de processo descoberto através da mineração de processo não é uma questão trivial. Podemos analisar o problema a partir de dois pontos de vista: (1) a qualidade dos dados utilizados para mineração de dados ou (2) a capacidade do modelo representar a realidade. Ambos os pontos de vista serão abordados no Capítulo 4 por se tratar de desafios clássicos da mineração de processos. Portanto, o enfoque desta seção será oferecer uma visão geral sobre as questões relativas à qualidade dos modelos, bem como as principais métricas para avaliação de modelos de processos.

Dado um log de eventos, não há um modelo perfeito de processo a ser descoberto (GÜNTHER; VAN DER AALST, 2007). Muitos fatores estão em jogo quando se pretende identificar qual o melhor modelo para um determinado processo. Inclusive, como mostrado na Seção 3.1, alguns algoritmos permitem alguma parametrização, resultando em diferentes modelos de processos, mas esses algoritmos não comunicam claramente qual é o efeito dos parâmetros na qualidade do modelo de processo resultante.

[7] https://fluxicon.com/disco/
[8] https://www.my-invenio.com/
[9] https://www.celonis.com/snap-signup
[10] https://everflow.ai/pt/process-mining-pt/

A Figura 14 mostra as quatro dimensões ortogonais de qualidade que normalmente são consideradas ao avaliar resultado obtido através da descoberta de processo (AALST, 2011; VAN DER AALST, 2013; VAN DER AALST; ADRIANSYAH; VAN DONGEN, 2012). É preciso fazer a ressalva que, atualmente, não existem métricas padrões para nenhum tipo de modelo de processo (PRODEL et al., 2018).

A dimensão do *fitness* quantifica a fração do log de eventos suportado pelo modelo de processo. A precisão quantifica a parcela do comportamento descrito pelo modelo de processo que não é observado no log de eventos. A dimensão da generalização quantifica a probabilidade de um comportamento não visto anteriormente, mas permitido, seja suportado pelo modelo de processo. Finalmente, a simplicidade avalia a complexidade do modelo de processo, uma vez que os modelos mais simples são preferidos em relação aos mais complexos de acordo com o princípio conhecido como Navalha de Occam (*Occam's Razor*) (BERLINGERIO et al., 2009).

Figura 14 – Dimensões de qualidade na descoberta de processos

Em mineração de processos, assim como em qualquer abordagem baseada em dados, os algoritmos dispõem apenas de um conjunto limitado de dados para representar toda a realidade. Isso significa que nem todos os comportamentos permitidos possam estar presentes no log de eventos disponível. Assim, um comportamento que não foi observado no log de eventos não implica necessariamente em um comportamento proibido. A Figura 15 apresenta as diferentes possibilidades

de manifestações dos comportamentos em um processo. Inicialmente, devemos considerar que um comportamento pode ser observado nas seguintes entidades:

- Sistema (S), compreende os comportamentos permitidos pelo sistema de suporte ao processo. Ou seja, consiste no universo de possibilidades disponíveis para os usuários do sistema.
- Log de Eventos (L), contempla os comportamentos capturados pelo log de eventos. Ou seja, os comportamentos realizados que foram devidamente registrados no log de eventos.
- Modelo do Processo (M), consiste no universo de todos os comportamentos permitidos pelo modelo do processo.

Figura 15 – Diagrama relacionando os comportamentos observados no sistema, log de eventos e modelo de processo (BUIJS, 2014)

Considerando que temos um log de eventos extraído de um sistema qualquer, e, a partir deste, descoberto um modelo de processo através de alguma técnica de mineração de processo, espera-se que um comportamento observado no modelo descoberto corresponda a um comportamento registrado no log, que por sua vez tenha correspondência com um comportamento permitido no sistema. Contudo, nem sempre existe o alinhamento esperado. Na Figura 15 observamos situações em que existe uma sobreposição parcial ou sequer há sobreposição dos comportamentos observados nessas entidades.

O diagrama da Figura 15 mostra sete áreas descritas da seguinte forma:

1. $S \cap L \cap M$ – Alinhamento completo: O comportamento é permitido pelo sistema, está presente no log de eventos e está adequadamente previsto no modelo do processo.
2. $S \setminus (L \setminus M)$ – Ruído: Comportamento não previsto no sistema e nem no modelo, mas observado no log de eventos. Estes casos podem indicar imperfeições no log de eventos ou falha no sistema.
3. $S \setminus (L \cap M)$ – Ruídos modelados: Semelhante a região 2, mas neste caso o ruído foi incluído no modelo do processo.
4. $(S \setminus M) \setminus L$ – Generalização inadequada: Neste caso o comportamento é observado exclusivamente no modelo do processo. Possivelmente, trata-se de comportamento oriundo da capacidade de generalização do algoritmo de descoberta de processo.
5. $(S \cap M) \setminus L$ – Generalização: Semelhante ao 4, contudo, nesse caso o comportamento generalizado pelo algoritmo de descoberta encontra respaldo no comportamento permitido pelo sistema, embora não tenha sido observado no log de eventos.
6. $(S \setminus L) \setminus M$ – Comportamento ignorado: o comportamento previsto no sistema não foi observado no log de eventos e, consequentemente, não foi incorporado ao modelo do processo.
7. $(S \cap L) \setminus M$ – Comportamento abstraído: o comportamento está previsto no sistema e possui correspondência no log de eventos, mas não foi incorporado ao modelo do processo. Possivelmente, como resultado de uma abordagem filtragem de comportamento infrequente.

É importante ter em mente que muitas vezes não há como descrever precisamente todos os comportamentos aceitos pelo sistema, em primeiro lugar porque as possibilidades tendem a ser infinitas, mas também por conta da presença de comportamentos imprevistos em qualquer sistema do mundo real. Além disso, os sistemas tendem a mudar ao longo do tempo. Entretanto, a mineração de processos se presta a encontrar um modelo de processo que descreva o sistema com a maior

precisão possível, usando nada além do comportamento observado no *log* (BUIJS, 2014).

A partir da Figura 15 podemos definir três dentre as quatro principais métricas de qualidade, conforme apresentado na Tabela 6. Seja S, L, M respectivamente o conjunto de todos os comportamentos observados no sistema, log de eventos e modelo do processo; $|*|$ denota a quantidade de comportamento capturado em $*$. Por exemplo, $|L \cap M|$ denota a quantidade de comportamento observado no log de eventos e no modelo.

Tabela 6 – Medidas de qualidade vs. comportamentos observados

MÉTRICA DE QUALIDADE	**COMPORTAMENTO OBSERVADO**
Fitness	$\frac{\lvert L \cap M \rvert}{\lvert L \rvert}$
Precisão	$\frac{\lvert L \cap M \rvert}{\lvert M \rvert}$
Generalização	$\frac{\lvert S \cap M \rvert}{\lvert S \rvert}$
Simplicidade	Não se aplica

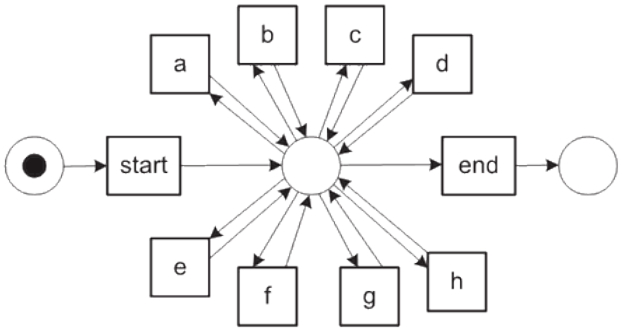
Existem várias abordagens para mensurar de forma prática as quatro dimensões de qualidade apresentadas. Para o *fitness*, o desafio está em como relacionar os eventos observados nos casos (instâncias do processo no *log* de eventos) com os nós no modelo de processo, especialmente em situações em que o modelo de processo e o caso estão em desacordo. A forma como essa questão é avaliada é crucial para o resultado obtido. Por exemplo, dada uma instância do processo $L_i = \langle a, b, c, d, e \rangle$ que não possui previsão no modelo, contudo neste modelo existe um caminho $C_j = \langle a, b, d, e \rangle$, ou seja, existe um desacordo entre a instância e o modelo em apenas uma atividade (c). Diante dessa situação, pode-se simplesmente considerar que a instância está completamente em desacordo com o modelo ou pode-se optar por uma abordagem mais elaborada, buscando medir o nível de desacordo entre o comportamento observado e o modelo. Conforme introduzido na Seção 3.3.2, temos uma abordagem robusta para a medição do *fitness* através da reprodução das instâncias sobre o modelo como a repetição baseada em tokens (ROZINAT; WILL

M.P. VAN DER, 2008) e reprodução das instâncias do processo em um modelo para mensurar o nível de “alinhamento” destas em relação ao modelo (ADRIANSYAH; BUIJS, 2013; VAN DER AALST; ADRIANSYAH; VAN DONGEN, 2012). Além disso, outras abordagens foram propostas: eventos negativos artificiais (DE WEERDT et al., 2011; GOEDERTIER et al., 2009) e comparação de fluxos de eventos com fluxos modelo (COOK; WOLF, 1999).

Um modelo com um bom *fitness* suporta a maior parte do comportamento observado no log de eventos. Ou seja, um modelo tem o *fitness* máximo se todos os casos observados no log de eventos podem ser reproduzidos integralmente (desde o seu início até o final). Embora seja um critério de qualidade importante, por si só o *fitness* não é suficiente para avaliar a qualidade de um modelo de processo. Pois, é muito fácil construir um modelo extremamente simples que seja capaz de reproduzir todos os casos de um log de eventos, mas que por outro lado também permite indiscriminadamente qualquer comportamento não previsto.

A Figura 16 mostra um “modelo em flor” representado na forma de uma rede de Petri. O modelo possui *fitness* perfeito quando confrontado com qualquer *log* de eventos que contenha as atividades $\{a, b, c, d, e, f, g, h\}$. Contudo, por razões óbvias, não é desejável obter um modelo que permita qualquer comportamento. Um modelo com essas características não possui uma boa precisão.

Figura 16 – Rede de Petri de um “modelo em flor”



A precisão é uma métrica que indica a quantidade de comportamento adicional permitido pelo modelo de processo, ou seja, comportamento presente no modelo, mas não observado no log de eventos. O comportamento cíclico (*loop*) pode gerar comportamentos potencialmente infinitos, complicando o cálculo dessa dimensão de qualidade (BUIJS; VAN DONGEN; VAN DER AALST, 2014). Muitas métricas de precisão existentes não levam explicitamente em consideração possíveis desvios

entre o comportamento observado no log de eventos e o comportamento modelado nos modelos, enquanto muitos estudos de caso mostram que esses desvios geralmente ocorrem na prática (ADRIANSYAH et al., 2014).

Um modelo de processo também não deve refletir exatamente o comportamento observado no registro de eventos, uma vez que o registro de eventos é apenas uma amostra dos comportamentos possíveis. O fato de um comportamento não ter sido observado não significa que ele seja indesejado. Portanto, espera-se de um bom modelo a capacidade de generalizar o comportamento observado de forma a prever comportamentos que embora não tenham sido observados sejam admitidos pelo sistema. Contudo, a generalização é a dimensão de qualidade mais difícil de se avaliar. Uma abordagem estatística é proposta em (VAN DER AALST; ADRIANSYAH; VAN DONGEN, 2012), mas esbarra em questões relacionadas ao paralelismo de atividades. Em (ROZINAT; WILL M.P. VAN DER, 2008), uma métrica de generalização é proposta assumindo ser o inverso da precisão, embora seja uma aproximação aceitável para algumas aplicações práticas, do ponto de vista teórico não está correto. Além disso, outras abordagens com métricas de generalização são propostas em (DONGEN; DIJKMAN; MENDLING, 2008; VAN DONGEN; MENDLING; VAN DER AALST, 2006).

A geração de um modelo com baixa precisão é caracterizada como *underfitting*, resultante de um problema de generalização excessiva do comportamento observado no log de eventos. A geração de um modelo que não generaliza suficientemente é caracterizada como *overfitting*, resultante da geração de um modelo muito específico.

A quarta dimensão de qualidade é a simplicidade, que segundo o princípio da Navalha de Occam (*Occam's razor*) não se deve aumentar além do necessário o número de entidades para explicar qualquer coisa. Ou seja, o melhor modelo de processo é o modelo mais simples que consiga explicar o que é observado no log de eventos (VAN DER AALST; ADRIANSYAH; VAN DONGEN, 2012). Em (MENDLING et al., 2008) temos que o tamanho do modelo de processo é a medida mais usual de simplicidade. Assim, a simplicidade do modelo de processo é definida, em geral, em termos das quantidades dos seus elementos (PRODEL et al., 2018). Contudo, métricas mais sofisticadas também podem ser usadas, por exemplo, métricas que levam em conta a "estrutura" ou "entropia" do modelo (VAN DER AALST; ADRIANSYAH; VAN DONGEN, 2012). Em (MENDLING; NEUMANN; VAN DER AALST, 2007) são apresentadas várias métricas de simplicidade. Três métricas de

complexidade especificamente projetadas para o domínio da mineração de processos foram proposta em (LASSEN; VAN DER AALST, 2009), sendo elas:

1. *Extended Cardoso Metric* (ECaM) é uma adaptação para redes de Petri da métrica de Cardoso. A métrica de Cardoso (CARDOSO, 2005) se baseia na contagem de gateways (XOR, OR e AND) do modelo, enquanto a ECaM se baseia em locais e transição de uma rede de Petri.
2. *Extended Cyclomatic Metric* (ECyM) é uma adaptação da métrica *McCabe's Cyclomatic Number* (MCCABE, 1976) que foi concebida para medir a complexidade de uma máquina de estados (rede de *workflow*) de softwares. A métrica ECyM leva em conta os possíveis estados do processo e as transições que podem ocorrer em uma rede de Petri através do seu gráfico de alcançabilidade.
3. *Structuredness metric* é uma métrica de simplicidade que reconhece e pontua estruturas (sequências, escolhas, iterações e etc.), tendo seu valor obtido através da soma dos valores atribuídos às estruturas identificadas. A métrica foi criada com o intuito de superar limitações das métricas ECaM e ECyM; uma vez que, a primeira se baseia apenas sintaxe do modelo, ignorando a complexidade do comportamento que este permite, e, a segunda, segue caminho oposto, oferecendo uma medida de complexidade de comportamento que ignora a sintaxe do modelo.

Balancear *fitness*, precisão, generalização e simplicidade é um dos grandes desafios da mineração de processos (VAN DER AALST et al., 2012). Esta é a razão pela qual as técnicas de descoberta de processos mais robustas oferecem parâmetros para que o usuário possa explorar os modelos até encontrar um que considere adequado aos seus objetivos. Na Seção 3.4 abordaremos esse e outros desafios da mineração de processos.

3.4 Princípios norteadores e desafios

Em virtude de sua condição de disciplina emergente, a mineração de processos enfrenta desafios fundamentais. Visando promover uma agenda para desenvolvimento da mineração de processos, o *IEEE Task Force on Process Mining*,

grupo ligado ao *Data Mining Technical Committee (DMTC*) e Computational Intelligence Society (CIS) do *Institute of Electrical and Electronic Engineers (IEEE)*, publicou o manifesto da mineração processos (VAN DER AALST et al., 2012), listando princípios norteadores e desafios a serem perseguidos pela comunidade acadêmica.

Desde então, novas ferramentas e abordagens trouxeram progressos significativos para os tópicos propostos no manifesto. Contudo, os princípios norteadores elencados se perpetuam, bem como grande parte dos desafios não foram completamente superados. Nesta seção apresentaremos os princípios norteadores e desafios do manifesto da mineração de processos que se relacionam com esta tese.

O primeiro princípio elencado no manifesto da mineração de processos (GP1) preconiza que os logs de eventos devem ser tratados como informações de primeira classe, uma vez que a qualidade de um resultado de mineração de processos depende muito dos dados fornecidos. Infelizmente, os logs de eventos registrados nos sistemas não são criados para fins analíticos, sendo apenas um "subproduto" criado para *debug* ou gerenciamento interno dos WfMS[11] *(Workflow Management Systems).* Esta tese não aborda diretamente o problema da qualidade do registro dos eventos nos sistemas de informações, mas entendemos que contribui nesta temática, ainda que indiretamente, por evidenciar a importância da qualidade no registro dos eventos.

Outro princípio (GP5) afirma que: “os modelos devem ser tratados como abstrações propositais da realidade” (VAN DER AALST et al., 2012). Dado um log de eventos, pode haver várias visualizações úteis, bem como os usuários podem desejar visões diferentes. Esta tese se alinha perfeitamente a essa perspectiva, uma vez que propõe uma abordagem para oferecer visualizações alternativas para um processo através do pré-processamento de logs de eventos.

Dentre os desafios podemos destacar o (C1), Encontrando, agrupando e limpando eventos. Esse desafio aborda, entre outras coisas, o nível de granularidade de um evento, que pode variar de uma atividade simples a um procedimento complexo. Nesta tese, oferecemos uma abordagem para tratar aspectos relacionados a granularidade em logs de eventos. Este tema será abordado em mais detalhes na Seção 4.3.

O desafio C2 do manifesto da mineração de processos, Lidando com logs complexos de eventos com características diversas, alerta que alguns logs contém

[11] WfMS (*Workflow Management Systems*) são sistemas que podem ser configurados para reproduzir o comportamento definido em um processo de negócio.

eventos em um nível de abstração muito baixo, sendo estes de pouco interesse para os usuários, devendo assim agregá-los para oferecer uma visão com nível mais alto de abstração (VAN DER AALST et al., 2012). Com a abordagem proposta buscamos justamente favorecer a visualização dos modelos dos processos em diferentes níveis de abstração através da mudança na granularidade dos eventos. Portanto, entendemos que essa tese está diretamente relacionada ao desafio C2 do manifesto da mineração de processos.

4 PRÉ-PROCESSAMENTO DE LOGS DE EVENTOS

Os processos de negócios ou simplesmente processos estão no cotidiano das pessoas e das organizações cumprindo as mais variadas funções (compras de insumos, gerenciamento de pessoal, contabilidade e assim por diante). É comum que processos sejam suportados por algum tipo de sistema de informação. Os sistemas de informações sensíveis aos processos são conhecidos como *Workflow Management Systems* (WfMS) ou apenas sistemas de fluxos. Esses sistemas se baseiam na noção de processos rígidos e bem definidos, no qual, quando adequadamente configurados, são capazes de controlar eficientemente a correta execução dos processos. Pois, os usuários só terão acesso às atividades que são autorizados e que estejam habilitadas para a fase em que o processo se encontra.

Por outro lado, o cenário atual é de grande dinamismo e competitividade, exigindo que as organizações respondam rapidamente às necessidades de mudanças do ambiente. Essa situação é potencialmente problemática para os sistemas de fluxos, uma vez que estes não estão preparados para lidar com exceções ou improvisos, exigindo em qualquer contexto a execução estrita do comportamento prescrito. Diante disso, é comum adoção por funcionários, ou mesmo toda organização, da prática de *by-pass*[12] do sistema (GÜNTHER, 2009).

A despeito da disseminação dos sistemas de fluxos, muitos processos de negócios são controlados por sistemas de informações que não exercem qualquer controle sobre o processo de negócios que apoiam. Nessas condições, mesmo que haja um processo bem definido, existe alta probabilidade de haver uma grande variabilidade na sua execução.

A flexibilidade na execução de um processo é um importante fator de aumento na complexidade dos modelos de processos descobertos com a mineração de processo, independentemente de sua causa. Neste capítulo abordamos as especificidades dos processos flexíveis, frequentemente encontrados em ambientes reais. Também apresentamos um estudo aplicando ferramentas de mineração de processos em um ambiente real, bem como identificamos e classificamos padrões de comportamentos em atividades que contribuem para o aumento da complexidade dos modelos de processos.

[12] By-pass consiste na execução de ação deliberada do usuário com vistas a "enganar" o sistema para realizar um comportamento não previsto por este.

O capítulo está organizado em três seções. A primeira, Seção 4.1, apresenta as características, limitações e alternativas para mineração de processos quando aplicada em processos flexíveis. A segunda, Seção 4.2, apresenta um estudo no qual foram aplicadas e avaliadas, em um ambiente real, ferramentas (comercial e acadêmica) de descoberta de processos concebidas para lidar com processos complexos. Na Seção 4.3 abordamos a questão do pré-processamento de logs de eventos para lidar com desafios apresentados pela mineração de processos. E por fim, na Seção 4.4, são apresentadas as conclusões obtidas nos estudos apresentados neste capítulo.

4.1 Processos complexos

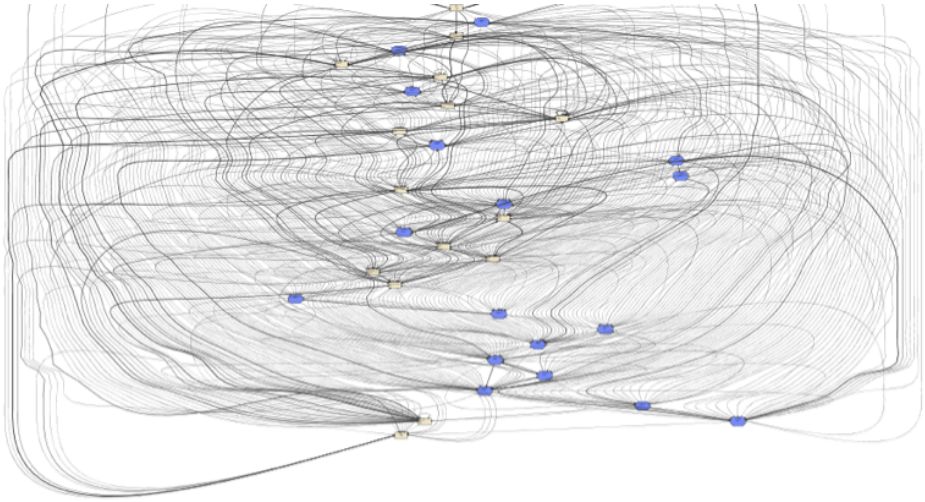
Algumas organizações possuem processos de negócios bem estruturados, nos quais as atividades são repetíveis e contam com entradas e saídas bem definidas. Contudo, é comum a incidência de processos pouco estruturados no seio das organizações. Os processos pouco estruturados são mais flexíveis, permitindo uma maior liberdade aos indivíduos na execução das atividades. Nesses processos, a experiência, intuição ou preferências pessoais são determinantes no rumo das ações (VAN DER AALST, 2010).

Muitos processos são documentados sob a forma de modelos normativos ou descritivos, no entanto, esses modelos idealizados não necessariamente correspondem ao processo real. Não podemos perder de vista que os processos emergem do comportamento humano (GÜNTHER; VAN DER AALST, 2007), mesmo quando controlados por sistemas de informações. Como resultado, pode haver uma desconexão entre processo desenhado (TO-BE) e a realidade (AS-IS). A experiência com a mineração de processos tem mostrado que a disparidade entre o esperado e a realidade é, frequentemente, maior do que se imagina.

Os processos quando bem estruturados produzem modelos conhecidos como “lasanha”, pois permitem sua visualização através de camadas (VAN DER AALST, 2011). Já os processos pouco estruturados tendem a produzir modelos conhecidos como “espaguete”, devido ao emaranhado de relações que as atividades apresentam produzindo a aparência do prato homônimo (VAN DER AALST, 2011).. A Figura 17 mostra um modelo de processo do tipo “espaguete”, gerado no ProM (*plug-in Mine for Fuzzy Model*) a partir de um log de eventos com pouco mais de setenta mil eventos.

Podemos ver que mesmo a partir de um log de eventos relativamente pequeno pode-se obter um modelo de processo incompreensível. Cabe ressaltar que a ferramenta utilizada para gerar o modelo apresentado na Figura 17 é dotada de mecanismos para simplificação do modelo de processo, mas adotamos parâmetros que restringiram esse mecanismo com vistas a mostrar o quão complexo o modelo do processo pode ficar quando diante de dados reais sem que se promova alguma simplificação.

Figura 17 – Exemplo de modelo de processo “espaguete”



Lidar com a complexidade inerente aos processos flexíveis é um dos grandes desafios da mineração de processos. Por outro lado, os processos complexos são muito atraentes do ponto de vista da mineração de processos, pois oferecem maior oportunidade de melhoria. Trabalhar com processos bem organizados é muito mais fácil, mas não há muito a melhorar. (VAN DER AALST et al., 2012)

O principal caminho para viabilizar a exploração de processos complexos é promover a sua simplificação (VAN DER AALST, 2011). Para tanto, adota-se duas estratégias: (1) filtragem de comportamento durante a descoberta dos modelos de processos e (2) pré-processamento dos logs de eventos.

Conforme introduzido na Seção 3.4, o *Fuzzy Miner* (GÜNTHER; VAN DER AALST, 2007) é um dos primeiros e mais importantes algoritmos a adotar a filtragem de comportamento infrequente na descoberta automática de processos. Outros algoritmos de descoberta de processos foram adaptados para incorporar características do *Fuzzy Miner*. Por exemplo, o *Heurístic Miner* ganhou uma extensão chamada de *Fodina*, contemplando mais robustez através da eliminação de

comportamento ruidoso, além de disponibilizar ao usuário a possibilidade de determinar a sensibilidade do algoritmo (GARCIA et al., 2019).

Os idealizadores do Fuzzy Miner criaram uma ferramenta comercial batizada de Disco com algumas das inovações trazidas pelo *Fuzzy Miner* que se consolidaram como padrão de mercado. O *Fuzzy Miner* e Disco, respectivamente, ferramentas acadêmicas e comerciais, são referências na mineração de processos flexíveis. Diante disso, promovemos um estudo para avaliar a eficácia de ambas em um ambiente real. O estudo foi conduzido com a finalidade de identificar eventuais limitações e oportunidades de melhorias quando aplicadas em um domínio pouco explorado pela mineração de processos (judiciário).

4.2 Mineração de processos complexos

O estudo apresentado nessa seção serviu de base para o artigo intitulado *Process Mining Discovery Techniques in a low-structured Process Works?* (D'CASTRO; OLIVEIRA; TERRA, 2018), apresentado no *7th Brazilian Conference on Intelligent Systems (BRACIS)*, realizado na cidade de São Paulo em 2018. O objetivo foi avaliar a eficácia de ferramentas de referência na descoberta de processos em ambientes flexíveis. O estudo foi conduzido sobre um domínio pouco explorado pela mineração de processos: processos judiciais de um tribunal de justiça brasileiro. A seguir apresentamos uma contextualização do ambiente no qual o estudo foi aplicado.

4.2.1 Ambiente do Estudo

A Emenda Constitucional no. 45, de 31 de dezembro de 2004, chamada de "Reforma do Poder Judiciário", introduzindo o inciso LXXVIII, no artigo 5°**.** da Constituição Federal, aduzindo que "a todos, no âmbito judicial e administrativo, são assegurados a razoável duração do processo e os meios que garantem a celeridade de sua tramitação". Essa emenda foi proposta com o intuito de acabar com a morosidade no judiciário. (BOFF; HASSE, 2017)

Seguindo na mesma direção, o Poder Judiciário adotou o processo judicial eletrônico, regulamentado por meio da Lei n°**.** 11.419/2006, de 19 de dezembro de 2006 (Lei do Processo Eletrônico). O artigo 8° da referida lei diz que "os órgãos do Poder Judiciário poderão desenvolver sistemas eletrônicos de processamento de ações judiciais por meio de autos total ou parcialmente digitais, utilizando,

preferencialmente, a rede mundial de computadores e acesso por meio de redes internas e externas" (BRASIL, 2006).

Em meados de 2011 o Conselho Nacional de Justiça (CNJ) deu um novo passo na informatização do judiciário e lançou o sistema nacional PJe (Processo Judicial Eletrônico). Em 2014 foi determinada a adoção do PJe em todos os Tribunais do país. Uma característica importante desse sistema é que se trata de um sistema de fluxo (WfMS) baseado em BPM. Em Cartilha veiculada no IV Encontro Nacional de Judiciário (CNJ, 2010), o CNJ esclarece aos tribunais do país o potencial de um sistema de fluxos para o judiciário conforme segue:

> *"É possível atribuir um fluxo diferente para cada uma das classes processuais existentes. Quanto mais específico o fluxo, mais fácil será automatizar tarefas de gabinete e secretaria. À primeira vista, pode ser que pensemos que essa é uma característica dispensável. A experiência mostra, no entanto, que ela é essencial. Com honrosas exceções, a grande maioria dos sistemas processuais trabalha em dois extremos no que concerne à tramitação ou ao acompanhamento da tramitação dos processos judiciais. De um lado, temos o engessamento total: o sistema tem em seu código os passos passíveis de serem praticados e alteração dessa via reclama reescrever o programa em algum grau. Do outro lado, temos a liberdade absoluta: o sistema permite que o usuário pratique qualquer ato. Não há limites e, em razão disso, surge o problema dos erros reiterados: sem freio, uma desatenção momentânea pode fazer com que um processo siga um tortuoso caminho, inclusive com a possibilidade da anulação da decisão. Mais que isso, a liberdade total não vem sem outro custo: uma imensa dificuldade em automatizar procedimentos, já que sempre é necessária uma intervenção humana para, fazendo uso da inteligência, informar à máquina qual deve ser o próximo passo. O PJe, com seus fluxos configuráveis, fica entre esses dois extremos. Embora se possa definir caminhos mais rígidos se isso for conveniente ou necessário, a alteração dos fluxos não depende da reescrita do sistema ou do pessoal da TI, mas da atuação de alguém que conhece processo judicial, muito provavelmente um servidor especialista do tribunal. Além disso, esses caminhos rígidos podem levar à automatização de tarefas repetitivas. Finalmente, pode-se definir caminhos tão amplos que estaríamos simulando a situação da liberdade absoluta. Tudo depende de como se quer ver o sistema funcionar"*.

Ainda na mesma cartilha, o CNJ sugere que os benefícios oriundos da adoção de um sistema de fluxos tendem a aumentar com a experiência, a partir da otimização dos fluxos processuais. Também reforça que o sistema mantém registro abrangente de logs para fins de auditorias. Logo, como antevisto e evidenciado pelo CNJ, o judiciário, através da adoção do sistema PJe se tornou terreno fértil para mineração de processos.

4.2.2 Análise Preliminar

Obtidos junto ao Tribunal de Justiça de Pernambuco, os dados utilizados no estudo apresentado neste capítulo correspondem a execução de processo judiciais similares (processos de uma mesma classe[13]) no sistema PJe. Para realização do estudo, foram coletados todos os eventos associados a uma mesma classe processual no período de 13 meses (01/06/2016 a 04/07/2017), resultando em 73.412 eventos de 3.359 casos (processos judiciais). A Tabela 7 sumariza as características básicas do log de eventos utilizado.

Tabela 7 – Características do log de eventos no estudo

EVENTOS	CASOS	ATIVIDADES	PERÍODO
73.412	3.359	76	01/06/2016 a 04/07/2017

Em análise preliminar identificamos as seguintes características sobre o conjunto de dados disponibilizados para o estudo:

- Média de 16 eventos por caso com alta variabilidade (mínimo de 6 e máximo de 45 eventos por caso);
- Grande variação nas frequências das atividades. Para facilitar o estudo agrupamo-las em três faixas: alta (mais de 1500 ocorrências) contemplando 13 atividades, intermediária (entre 400 e 1500 ocorrências) contemplando 19 atividades e baixa (até 400 ocorrências) com 44 atividades;

[13] O Conselho Nacional de Justiça mantém uma classificação de processos judiciais (disponível em https://www.cnj.jus.br/sgt/consulta_publica_classes.php).

- Grande quantidade (mais de 2000) de "caminhos" diferentes (variantes) foram observados no log de eventos. Mais da metade dos casos (52%) percorreu um caminho "inédito".

Os resultados da análise preliminar demonstram claramente se tratar de um processo altamente flexível, sobretudo quando observamos que poucos casos seguiram caminhos semelhantes. A seguir descrevemos os estudos exploratórios realizados com as ferramentas: *Mine for Fuzzy Model (ProM 6.6)* e *Disco*.

4.2.3 Estudos exploratórios

O *plugin Mine for Fuzzy Model* dispõe de vários parâmetros de inicialização. Explorando a ferramenta observamos que algumas configurações comprometeram a integridade dos modelos e as demais produziram pouco ou nenhum efeito nos modelos. Dentre estes apresentamos os resultados observados com os parâmetros de inicialização denominado $distância\ máxima\ do\ evento$. Este parâmetro é responsável por estabelecer uma distância limite para considerar que há relação entre duas atividades. Por exemplo, dada uma sequência de atividades $s = \langle a, b, c, d, e \rangle$ e $distância\ máxima\ do\ evento = 3$, temos que a se relaciona com d, mas não com e. Analisamos o impacto na medida significância (relevância) através da variação da $distância\ máxima\ do\ evento$ em três configurações distintas: (1) mínimo, (2) padrão e (3) máximo. A Tabela 8 apresenta os valores observados para a significância em relação às cinco atividades mais frequentes no log de eventos. A significância consiste em uma métrica de importância relativa do comportamento baseada na frequência das relações de precedência entre as atividades do log de eventos (GÜNTHER; VAN DER AALST, 2007).

A primeira coluna da Tabela 8 mostra um identificador das atividades. A coluna seguinte (frequência), refere-se ao número de vezes que a atividade foi observada no log de eventos. A coluna cobertura indica o percentual de caso no log de eventos em se observou ao menos uma vez a incidência da respectiva atividade. As últimas três colunas apresentam os valores de significância obtidos para cada uma das configurações avaliadas para o parâmetro $distância\ máxima\ do\ evento$. Também pode-se observar na Tabela 8 que não houve alteração relevante na significância ao variarmos a $distância\ máxima\ do\ evento$. Logo, os modelos de processos resultantes

destas configurações demonstraram diferença mínima.

Tabela 8 – Análise do impacto do parâmetro $distância\ máxima\ do\ evento$

ATIVIDADE	FREQUÊNCIA	COBERTURA	SIGNIFICÂNCIA		
			(1)	(2)	(3)
AS	7.350	76,9%	0,726	0,738	0,739
AJ-F	5.990	99,6%	0,584	0,581	0,581
CC	5.926	98,6%	0,558	0,561	0,562
AUD-C	5.310	96,9%	0,544	0,550	0,551
AUD-FC	5.308	96,9%	0,523	0,524	0,524

Os efeitos observados explorando os parâmetros de inicialização da ferramenta nos levaram a concluir pela adoção de valores padrão na inicialização para realização do estudo exploratório.

Além dos já abordados, a ferramenta possui mais cinco parâmetros operacionais divididos em três categorias, sendo eles: *Node filter* ($significance\ cutoff$), *Edge filter* ($cutoff$ *e* $utility\ rate$) *e Concurrency filter* ($preserve$ *e* $ratio$). Os parâmetros da categoria *Node filter* impactam nas atividades e os da categoria *Edge filter* nas transições entre as atividades. Por fim, temos os parâmetros de *Concurrency filter* que determinam o comportamento do algoritmo em face de atividades concorrentes. Realizamos experimentos variando cada um dos parâmetros seguindo três configurações (baixa, intermediária e alta) enquanto os demais parâmetros foram mantidos com valores padrão.

O parâmetro $significance\ cutoff$ determina a intensidade na qual o algoritmo buscará promover abstração (eliminação de atividades) ou agregação (criação de *clusters* com atividades afins) no modelo do processo. Com o valor de $significance\ cutoff = 0.000$, todas as atividades foram incluídas no modelo de processo resultante e nenhum cluster foi criado. Com o valor de $significance\ cutoff = 0.250$, observou-se a eliminação das cinco atividades menos frequentes no modelo de processo. Além disso, surgiram oito *clusters*, sendo o maior composto por 26 atividades e o menor com apenas duas. Com valor de $significance\ cutoff = 0.750$, o algoritmo reuniu todas as atividades em um único

cluster. Com isso, observamos que o $significance\ cutoff$ é bem sensível, sendo recomendável a adoção de valores baixos para este parâmetro. A Tabela 9 apresenta os valores obtidos para as duas medidas de qualidade do modelo apresentado pela ferramenta. A primeira coluna indica o valor utilizado para o parâmetro $significance\ cutoff$. A segunda coluna mostra uma medida do nível de detalhe do modelo resultante *(model detail).* Esta medida indica o percentual do comportamento observado no log de eventos que foi incorporado ao modelo de processo resultante. A terceira coluna apresenta uma medida de conformidade entre o log de eventos e o modelo resultante *(log conformance).* Como pode ser observado, quanto menor o nível de detalhe do modelo maior foi a conformidade com o log.

Tabela 9 – Análise do parâmetro $significance\ cutoff$ no *Mine for Fuzzy Model*

VALOR	MODEL DETAIL	LOG CONFORMANCE
0,000	100%	78,21%
0,250	78,2%	85,91%
0,500	48,53%	86,75%
0,750	17,82%	100%

A utilidade ($utility\ rate$) é uma métrica de relevância para transição no *Mine for Fuzzy Model*, resultante da combinação de duas outras métricas: significância e correlação. O parâmetro $utility\ rate$ determina o grau de influência da significância e da correlação no cálculo da utilidade de uma transição. Valores de $utility\ rate$ baixos determinam a predominância da significância sobre a correlação, consequentemente valores altos fazem prevalecer a correlação no cálculo da utilidade. Por sua vez, o parâmetro c$utoff$ estabelece o nível de sensibilidade do algoritmo para eliminar as transições do modelo. O valor padrão para o parâmetro $cutoff$ adotado pela ferramenta é 0.2. Observamos que valores acima do padrão adotado pela ferramenta produziram modelos demasiadamente complexos devido à quantidade excessiva de conexões apresentadas.

A variação nos parâmetros de *Concurrency filter* ($preserve$ e $ratio$) não causou alterações relevantes nos modelos resultantes, fato que pode ser explicado pela baixa

incidência de atividades concorrentes no log de eventos analisado.

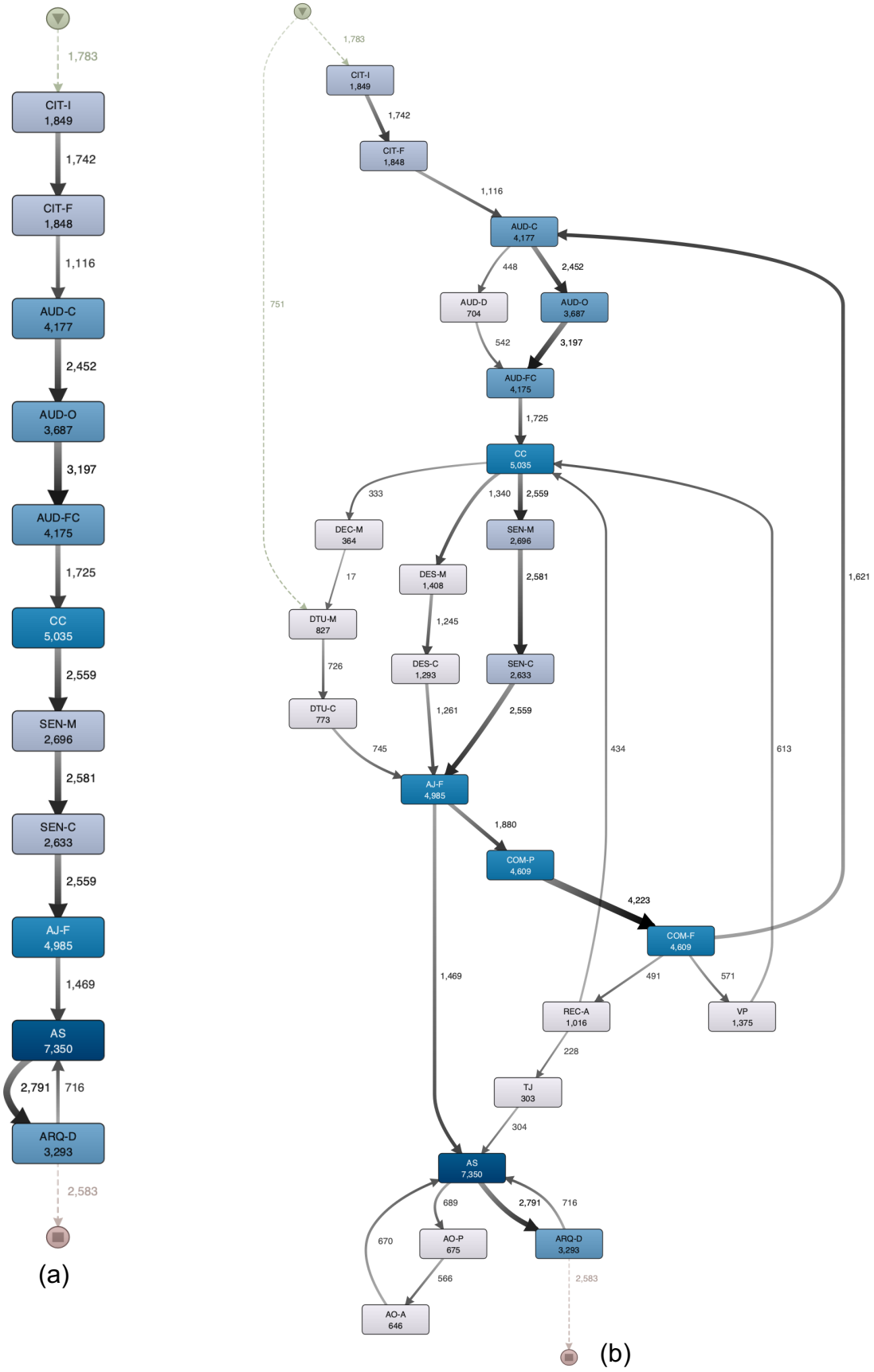
O estudo exploratório com a ferramenta comercial *Disco* seguiram a mesma estratégia dos realizados para o *Mine for Fuzzy Model*, além disso foram realizados com os mesmos dados. Como o Disco não apresenta medidas de relevância atribuídas aos elementos do modelo, bem como não oferece qualquer medida de qualidade do modelo descoberto, realizamos uma análise qualitativa dos modelos resultantes. A ferramenta analisada possui apenas dois parâmetros operacionais: $activities$ e $paths$. O primeiro determina o nível de sensibilidade quanto às atividades e o segundo em relação às conexões, ambos podendo variar entre 0.0% e 100.0%. Os experimentos mostraram que o valor do parâmetro $paths$ deve ser baixo para evitar modelos demasiadamente complexos. Em relação ao parâmetro $activities$, observamos que mesmo quando diante de configurações agressivas (valores altos) o modelo do processo contemplou dez atividades mais frequentes. A Figura 18 mostra dois modelos de processo gerado pelo *Disco* com os seguintes parâmetros: (a) $activities = 0.0\%, paths = 0.0\%$ e (b) $activities = 20.0\%, paths = 0.0\%$.

Os modelos obtidos foram submetidos à apreciação por especialistas no processo para obter um feedback em relação à coerência com a realidade e simplicidade dos modelos. Dois especialistas de negócios apoiaram na avaliação dos modelos de processos. As conclusões foram sintetizadas em dois pontos:

- Os modelos mais simples omitiram atividades relevantes devido a alta variabilidade na frequência das atividades.
- Nenhum dos modelos produzidos no *Mine for Fuzzy Model* apresentou agrupamentos coerentes. Ou seja, atividades sem afinidades do ponto de vista de negócio foram reunidas gerando modelos mais difíceis de interpretar

Além do feedback dos especialistas, através deste estudo foi possível verificar que, embora sejam baseadas no mesmo algoritmo, existe uma diferença significativa entre as duas ferramentas. O *plugin* implementado no ProM é rico em detalhes e parâmetros, oferecendo maior controle ao usuário na geração do modelo. Por outro lado, sua operação é muito complexa. Já o *Disco* possui uma operação bastante simplificada, mas, para isso, omite informações importantes, por exemplo, métricas de qualidade do modelo resultante.

Figura 18 – Modelos gerados no Disco (a) *activities*=0,0% e (b) *activities*=20,0%



O estudo realizado mostrou que ao aplicar os mecanismos de simplificação de modelos de processos oferecidos pelas ferramentas avaliadas houve perda de informação relevante e geração de agrupamentos incoerentes. Associamos os efeitos indesejados observados nos modelos de processos descobertos à características comuns a processos complexos, tais como: aspectos relacionados a granularidade das atividades e a incidência de comportamento recorrente. Assim, concluímos ser necessário a realização de pré-processamento nos logs de eventos para atacar diretamente os problemas identificados, uma vez que os mecanismos oferecidos pelas ferramentas não se mostram eficazes para estas circunstâncias.

4.3 Alternativas para pré-processamento de logs de eventos

Um dos grandes desafios para mineração de processos diz respeito aos diferentes níveis de granularidade do eventos (VAN DER AALST et al., 2012) no log, pois nem sempre existe um alinhamento entre o objetivo da mineração de processos e como os dados da execução dos processos são armazenados. Por exemplo, os sistemas podem gerar vários registros para uma mesma atividade, gerando com isso informações inúteis do ponto de vista do negócio. Por outro lado, a maioria das técnicas de mineração de processos pressupõe que os eventos do log estão no nível adequado de granularidade, ou seja, correspondem aos conceitos conhecidos no nível de negócios (VAN ZELST et al., 2020).

Também precisamos considerar a problemática da granularidade dos eventos sob a perspectiva das atividades do processo. Atividade é um termo genérico que pode se referir a uma tarefa atômica ou a um conjunto de tarefas. Em certas circunstâncias, deseja-se visualizar o processo em um nível de abstração mais alto do que o processo foi modelado. Nesses casos, desejamos reunir um conjunto de atividades como sendo uma única atividade (macroatividade). Nessa linha, PRODEL et al. propôs uma abordagem para visualização do processo em um nível de abstração mais alta baseado em uma estrutura hierárquica de classes de eventos (PRODEL et al., 2018) preexistente. Como grande parte dos trabalhos em mineração de processos, a abordagem foi criada no domínio da saúde, que conta com uma taxonomia de procedimentos, logo há uma estrutura hierárquica definida previamente. Contudo, para muitos domínios o problema reside justamente na criação de uma classificação hierárquica dos eventos, pois é algo trabalhoso e requer conhecimento abrangente

sobre o negócio.

As ferramentas de descoberta de modelos de processos não oferecem meios eficazes de alterar a granularidade dos eventos e atividades. Como alternativa, algumas ferramentas oferecem funcionalidade para filtragem do log de eventos, na qual o usuário pode eliminar eventos a partir de critérios definidos sobre os atributos do evento (nome, classe, duração, regra de precedência e outras). As ferramentas de filtro são úteis para eliminar comportamento indesejado, mas inúteis para promover outras manipulações nos logs de eventos, tais como agrupar ou dividir eventos de forma a representar melhor os eventos. Para estes casos, resta a manipulação "manual" de dados que exigem esforço e conhecimento técnico considerável.

Nossa investigação nos levou a concluir que existem problemas para realização de mineração em processos complexos não superados completamente pelas técnicas existentes que seriam minimizados com uma solução de pré-processamento no log de eventos. Além disso, entendemos que a abordagem de pré-processamento deve ser realizada sem informação a priori sobre as relações entre as atividades dos processos. Ou seja, deve-se contar apenas com os dados existentes em um log de eventos básico. Esta premissa é justificada pelo fato de que as informações a priori são produzidas por "especialistas no negócio", demandando esforço considerável e conhecimento a priori do processo.

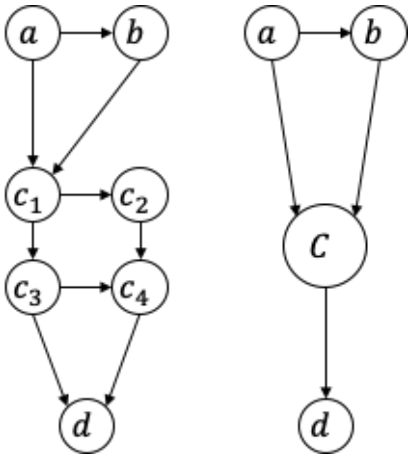
Mesmo em um processo flexível, é de se esperar que algum padrão emerja do relacionamento entre as atividades. Percebemos em nossos estudos exploratórios a existência de dois perfis de atividades:

- Afins: são atividades que se relacionam prioritariamente com um conjunto reduzido de atividades e seus rótulos guardam similaridades;
- Recorrentes: se relacionam com atividades variadas e em diversas fases do processo.

As atividades afins e recorrentes impactam de forma distinta na complexidade dos processos. As atividades afins podem ser agrupadas gerando uma atividade em um nível de abstração mais alto sem comprometer a integridade do modelo de processo. Já na direção oposta, observamos que as atividades recorrentes podem ser desmembradas de acordo com a fase do processo se encontram gerando indiretamente modelos de processos mais fáceis de serem interpretados.

A Figura 19 ilustra o agrupamento de atividades afins, na qual pode ser observado que as atividades c_1, c_2, c_3, c_4 do modelo original (Figura 19a) foram agrupadas em uma única atividade C no modelo apresentado na Figura 19b. No Capítulo 5 apresentamos uma abordagem que explora as ideias aqui apresentadas para agrupamento de atividades Afins através do pré-processamento do log de eventos.

Figura 19 – Modelo de processo (a) sem atividades agrupadas (b) com atividades agrupadas



Em relação às atividades recorrentes, observamos que se trata de importante fator de complexidade nos modelos de processos, pois sua ocorrência em momentos distintos provoca o aparecimento de loops prejudicando sobremaneira a “legibilidade” do modelo de processo. Além do mais, por se relacionarem com uma grande quantidade de atividades, incrementam a quantidade de transições das atividades e consequentemente a complexidade do modelo. As abordagens usuais não são eficazes no tratamento desse tipo de atividade. O fato de não apresentarem um contexto claro para sua execução dificulta a aplicação de uma abordagem de agregação. Por outro lado, não podemos supor que as atividades recorrentes são necessariamente infrequentes; ao contrário, por ocorrerem em vários contextos de negócios distintos tendem a apresentar uma alta incidência, tornando as abordagens de filtragem de comportamento infrequente inócua para estes casos. No Capítulo 7 apresentamos uma abordagem para tratamento das atividades recorrentes.

4.4 Conclusão do capítulo

Os estudos realizados nos permitiram conhecer melhor a realidade dos ambientes de processos flexíveis e as limitações das ferramentas de mineração de processos quando em face destes. O fato de as ferramentas de mineração de processos atuarem nos logs de eventos tal qual lhe é apresentado, oferecendo no máximo a filtragem de eventos, limita a qualidade dos modelos descobertos e onera os usuários com uma carga substancial de trabalho para manipulação “manual” do log de eventos.

Investigamos logs de eventos de processos reais e identificamos dois perfis de atividades (afins e recorrentes) que impactam diretamente na complexidade dos modelos de processos derivados destes. Para o primeiro (atividades afins) vislumbramos que o agrupamento destas atividades proporciona uma visão diferente (granularidade mais alta), bem como modelos melhores e menores. Já para o segundo, vislumbramos o oposto, consideramos que o desmembramento das atividades é o caminho para uma visão alternativa relevante para o modelo de processo.

As conclusões serviram de inspiração e base teórica e concepção das abordagens de pré-processamento dos logs de eventos visando superar as limitações observadas. Os Capítulos 5 e 6 apresentam abordagens para agrupamento de atividades afins e o Capítulo 7 apresenta uma abordagem para desmembramento de atividades recorrentes.

5 AGRUPAMENTO DE ATIVIDADES AFINS

Neste capítulo apresentamos uma abordagem para simplificação de logs de eventos através do agrupamento de atividades baseada em um novo indicador de afinidade denominado $activity\ fuzzy\ match$. O indicador proposto utiliza uma nova métricas de similaridade entre rótulos das atividades denominada de a $activity\ fuzzy\ similarity\ (afs)$ associada a uma nova métrica de relações de dependência entre as atividades denominada $highest\ direct\ relation\ ratio\ (hdrr)$.

O capítulo está organizado nas seguintes seções: A Seção 5.1 mostra a visão geral da abordagem proposta neste capítulo. Nas Seção 5.2 e Seção 5.3 apresentamos as métricas que suportam o $activity\ fuzzy\ match$. A Seção 5.4 detalha o indicador $activity\ fuzzy\ match$ e seus parâmetros. Na Seção 5.5 apresentamos a abordagem de transformação do log de eventos baseada no $activity\ fuzzy\ match.$ A Seção 5.6 mostra os resultados obtidos com a aplicação da abordagem em um ambiente real. Por fim, apresentamos as conclusões na Seção 5.7.

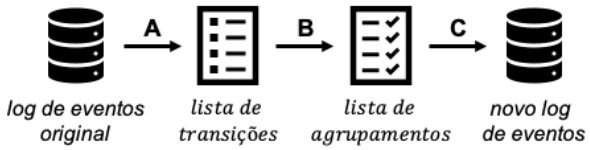
5.1 Visão geral da abordagem

A abordagem proposta neste capítulo tem por objetivo proporcionar uma mudança na granularidade nos logs de eventos sem comprometer sua capacidade de representar a realidade observada. Para tanto, buscamos identificar e agrupar automaticamente atividades afins (que façam parte um mesmo contexto de negócios) propiciando a descoberta de processos em um nível de abstração mais alto, contando apenas com um log de eventos básico[14]. Partimos da premissa que se uma atividade é frequentemente precedida por outra e existe similaridade em seus rótulos existe alta propensão dessas atividades possuírem uma afinidade tal que sua fusão é coerente do ponto de vista do negócio. Seguindo nessa linha, propomos um indicador, denominado de $activity\ fuzzy\ match$, que captura a relação de precedência e a similaridade entre os rótulos das atividades, bem como um procedimento para transformação do log de eventos baseado nesse indicador.

[14] Consideramos como básico o log de eventos que contêm apenas os atributos mínimos para realização da mineração de processos, conforme abordado na Seção 2.2.2.

A Figura 20 mostra as etapas da abordagem proposta neste capítulo. Como pode ser visto, a primeira etapa (A) consiste na geração da $lista\ de\ transições$, que consiste em uma listagem com todas as transições existentes no log de eventos com sua respectiva frequência.

Figura 20 – Visão geral da abordagem de transformação do log de eventos



Na mineração de processos, uma transição consiste em um par de eventos indicando que duas atividades foram executadas em sequência. A frequência de uma transição consiste no número de ocorrências da transição observado no log de eventos. As transições podem ser formalmente definidas como segue:

Definição 19 (Transição). Seja $\mathcal{D}_A$ o domínio de atividades no log de eventos $\mathcal{L}$ (conjunto de todas as atividades existentes em $\mathcal{L}$). Seja $\{a_1, \dots, a_n\} \in \mathcal{D}_A,\ A = \langle a_1, \dots, a_n \rangle, 1 \leq j \leq |A|$. Dizemos que há $transição$ entre duas atividades a_j e a_{j+1}, denotado por $a_j \succ a_{j+1}$ sse $\exists a_j, a_{j+1}, a_k \in \mathcal{D}_A | \ a_j \not\succ a_k$. A inexistência de transição entre duas atividades é denotada por $\not\succ$. Em outras palavras, existe uma transição entre duas atividades quando há no log de eventos a incidência de uma ou mais sequências de eventos dessas atividades. Por exemplo, dada uma sequência de eventos $A = \langle a, b, c \rangle$, observamos uma transição entre a e b ($a \succ b$). Contudo, não há transição entre a e c, logo $a \not\succ c$.

A criação da lista de transições consiste em um procedimento de percorrer todo o log de eventos identificando as transições existentes e computando a frequência das transições identificadas. A etapa seguinte (B) consiste em identificar afinidade entre as atividades; para tanto, aplica-se o indicador $activity\ fuzzy\ match$ sobre cada uma das transições da $lista\ de\ transições$. As transições consideradas afins passam a compor a $lista\ de\ agrupamentos$. Cada transição incluída na $lista\ de\ agrupamentos$ recebe o nome do respectivo nome do agrupamento.

Por fim, na terceira etapa (C), temos a transformação do log de eventos. Esta etapa consiste em percorrer novamente o log de eventos avaliando as transições. Caso as atividades enviadas na transição avaliada integrem a lista de agrupamentos, o par de eventos será agrupado em um único evento.

Na seção seguinte abordamos as métricas adotadas no $activity\ fuzzy\ match$ para capturar a relação de precedência entre atividades de um log de eventos.

5.2 Relação de precedência entre atividades

As métricas sobre a relação de precedência *(direct-follows)* entre atividades são largamente utilizadas pelos algoritmos de descoberta automática de processos e também pelos algoritmos voltados à eliminação de comportamento infrequente nos logs de eventos. Uma medida usual de relação de precedência entre atividades é conhecida como $direct - follows\ relation$ e consiste na frequência absoluta das transições entre duas atividades (TAX; SIDOROVA; VAN DER AALST, 2018). Assim, o $direct - follows\ relation$ entre as atividades a e b, denotado por $|a \succ b|$, representa a quantidade de vezes em que se observa uma transição entre as duas atividades.

Por se tratar de uma medida global, o $direct - follows\ relation$ não captura adequadamente a relevância da relação de precedência no contexto das atividades envolvidas. Visando contornar essa limitação, em (CONFORTI; ROSA; HOFSTEDE, 2017) foi adotada uma métrica relativa para relação de precedência denominada de $frequência\ relativa\ (fr)$. A seguir, apresentamos a sua definição.

Definição 20 (fr) (CONFORTI; ROSA; HOFSTEDE, 2017). Seja $\mathcal{D}_A$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Dadas as atividades $a, b \in \mathcal{D}_A$, seja $|a|$ a frequência absoluta de a, $|b|$ a frequência absoluta de b e $|a \succ b|$ é a quantidade de transições de a para b. Assim, a $frequência\ relativa\ (fr)$ entre a e b, denotada por $fr(a, b)$, é definida como segue:

$$fr(a,b) = \frac{2 * |a \succ b|}{|a| + |b|}$$

Por exemplo, dado o log de eventos $\mathcal{L} = [\langle a, b, c\rangle^{10}, \langle a, c, b\rangle^{4}, \langle a, b, a\rangle^{2}]$, temos que $|a| = 18$, $|b| = 16$ e $|a \succ b| = 12$. Assim, temos que $fr(a, b) \cong 0{,}70$.

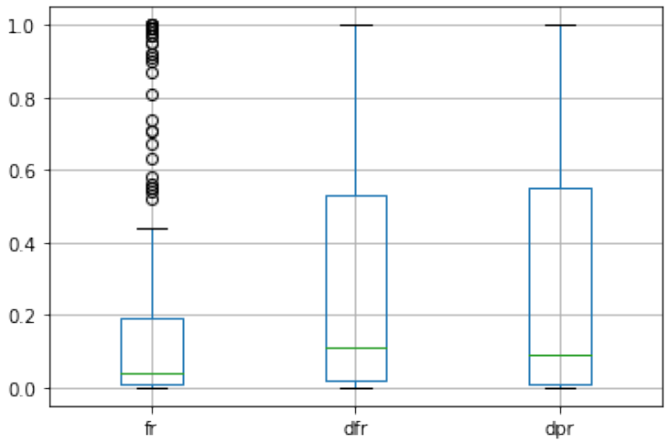
A $frequência\ relativa$ conforme apresentada (CONFORTI; ROSA; HOFSTEDE, 2017) é similar a métrica $directly - follows\ ratio$ apresentada em (TAX; SIDOROVA; VAN DER AALST, 2018). O trabalho de Tax, Sidorova e van der Aalst (2018) também apresentou uma métrica para a relação de precedência entre as atividades no sentido inverso, denominada de $directly - precedes\ ratio$. As métricas $directly - follows\ ratio\ (dfr)$ e $directly - precedes\ ratio\ (dpr)$ são definidas formalmente como segue:

Definição 21 (dfr e dpr) (TAX; SIDOROVA; VAN DER AALST, 2018). Seja $\mathcal{D}_A$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Dadas as atividades $a, b \in \mathcal{D}_A$, $dfr(a, b)$ e $dpr(a, b)$ são assim definidos:

$$dfr(a, b) = \frac{|a \succ b|}{|a|} \qquad dpr(a, b) = \frac{|b \succ a|}{|a|}$$

Visando compreender melhor as diferenças entre as três métricas (fr, dfr e dpr), conduzimos um estudo comparando os valores obtidos para 249 transições de um log de eventos real (mesmo log de eventos utilizado no estudo apresentado no Capítulo 4). A Figura 21 mostra um gráfico de caixa gerado a partir dos valores obtidos para as métricas analisadas.

Figura 21 – Gráfico de caixa com as métricas *fr, dfr* e *dpr*



Como podemos observar na Figura 21, os valores com a métrica fr diferem significativamente dos valores observados para com as métricas dfr e dpr. Diante desta constatação, analisamos individualmente as transições com valores discrepantes entre as métricas. Desta análise, concluímos que o fr não é uma métrica adequada para nossos propósitos por ser frágil quando há uma disparidade muito grande entre a frequência absoluta das atividades envolvidas na transição. Além disso, observamos que parte das transições afins foram mais bem identificadas pela dfr e outra parte pelo dpr, considerando para tanto o feedback de especialistas. Diante destas observações, propomos uma nova métrica denominada de $highest\ direct\ relation\ ratio\ (hdrr)$, que combina as métricas dfr e dpr. A métrica $hdrr$ é definida como segue:

Definição 22 ($hdrr$). Seja $\mathcal{D}_A$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Dadas as atividades $a, b \in \mathcal{D}_A$, O operador $hdrr$, denotado por $\overset{+}{\rightarrow}$, é assim definido:

$$x \overset{+}{\rightarrow} y = \max\left(dfr(a,b), dpr(a,b)\right)$$

Por exemplo, dado o log de eventos $\mathcal{L} = [\langle a,b,c\rangle^{10}, \langle a,c,b\rangle^{4}, \langle a,b,a\rangle^{2}]$, temos que $dfr(a,b) = 0{,}67$ e $dpr(a,b) = 0{,}75$. Logo, $a \overset{+}{\rightarrow} b = 0{,}75$.

5.3 Similaridade entre nome de atividades

Na modelagem de processos de negócios, a definição dos rótulos para as atividades é uma questão relevante, uma vez que o rótulo atribuído a uma atividade deve ser capaz de exprimir a ação e o contexto desta. Do contrário, a legibilidade do processo ficará comprometida. Em (MENDLING; HESSE; OBERWEIS, 2008) os autores avaliaram como o estilo de rotulagem das atividades impacta na qualidade do modelo, concluindo que os melhores resultados são observados pela convenção verbo/objeto (MALONE; CROWSTON; HERMAN, 2003; SHARP; MCDERMOTT, 2009).

Independente do estilo adotado, o fato é que seguir um padrão consistente para rotular as atividades em um modelo de processo é uma boa prática reconhecida. Considerando que o rótulo da atividade exprime a ação realizada nesta, partimos da

premissa que rótulos similares sugerem a existência de afinidade entre as atividades. Quer seja por indicar que as atividades correspondem a ações similares ou por estarem inseridas em um mesmo contexto de negócio. Com esta ideia em mente propusemos uma nova métrica para capturar a similaridade entre nomes de atividades.

Muitas aplicações exigem uma medida de similaridade entre objetos representados através de texto (JEH; WIDOM, 2002). A literatura dispõe de várias abordagens voltadas a mensurar a similaridade entre *strings* (DOWLING; HALL, 1980; DRESSLER; NGOMO, 2017; JEH; WIDOM, 2002; LEVENSHTEIN, 1966). Entendemos que a comparação direta das *strings* com os nomes das atividades não é uma maneira eficaz de identificar se as atividades são similares, pois é sensível ao tamanho das *strings*, a ordem das palavras dentre outros aspectos. Diante disso, propomos uma nova métrica baseada na medida de similaridade de Jaro-Winkler (JWS) (ver Seção 2.4.1) denominada de $activity\ fuzzy\ similarity\ (afs)$. A ideia é oferecer uma medida de similaridade adequada ao contexto dos nomes de atividades em processos de negócios.

O ponto de partida do afs consiste em obter para cada *string* um conjunto contendo as palavras com tamanho superior a φ_{WL}, sendo φ_{WL} um parâmetro que determina o tamanho mínimo de uma palavra. Este parâmetro auxilia no tratamento das *stopwords*[15], *TAG's* ou códigos que possam constar nos nomes das atividades.

Com base nas listas de palavras podemos calcular o $activity\ fuzzy\ similarity$ conforme definido a seguir:

Definição 23 (afs). Seja $\mathcal{D}_A$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Dada a atividade $a, b \in \mathcal{D}_A$, seja $S_a = \langle \mathrm{w}_a^1, \sqcup^1, \dots, w_a^i \rangle$ e $S_b = \langle \mathrm{w}_b^1, \sqcup^1, w_b^2, \sqcup^2, \dots, w_b^i \rangle$ as *strings* com o nome das atividades, na qual $\mathrm{w}_x^1, \dots, \mathrm{w}_x^i$ são *substrings* que compreendem palavras e $\sqcup^1, \sqcup^2, \dots, \sqcup^{i-1}$ *substrings* compreendendo os separadores de palavras (espaço, barras, vírgula, ponto e vírgula e outros). Então, seja $W_a = \{\mathrm{w}_a^1, \dots, \mathrm{w}_a^i\}$ o conjunto de palavras de a e $W_b = \{\mathrm{w}_b^1, \dots, \mathrm{w}_a^i\}$ o conjunto de palavras de a, no qual $w_x^i \subset S_x \wedge |w_x^i| \geq \varphi_{WL} | \forall x \in \mathcal{D}_A, 1 < i \leq |W_a|$. Seja $jw: \boldsymbol{s} \times \boldsymbol{s} \to \mathbb{R}^+$ a medida JWS (definida na Seção 2.4), o $conjunto\ de\ palavras\ similares\ (WM)$ é definida como segue:

[15] *Stopword* são palavras que não contribuem para o significado do texto, por exemplo, artigos ou preposições

$$WM = \{(a,b) \in W_a \times W_b \mid jw(x,y) \geq \varphi_{WS}\}$$

O $conjunto\ de\ palavras\ similares\ (WM)$ consiste em um conjunto de pares ordenados oriundos do produto cartesiano dos conjuntos dos nomes das atividades (W_a e W_b) que obtiveram uma medida de JWS superior ao valor do parâmetro φ_{WS}. Seja, $|WM|$ a quantidade de elementos em WM e $|W_a|$ a quantidade de elementos em W_a. Definimos o $activity\ fuzzy\ similarity$, denotado por $afs: W_a \times W_b \rightarrow \mathbb{R}^+$, como segue:

$$afs(a,b) = \frac{|WM|}{|W_a|}$$

O Algoritmo 1 descreve o procedimento para cálculo do afs.

Algoritmo 1
Activity fuzzy similarity

```
     entrada: String a, b; Threashold φ_wl, φ_ws;
     saída: wm | valor de afs(a,b)
1    # linhas 2 a 4: Cria a lista de palavras para a primeira atividade (W_a)
2    para todo s em {a,b} faça
2       para todo w_s^i ∈ s | seja w_s^i uma palavra de s faça
3          se tamanho(w_s^i) > φ_wl então
4             ListaDePalavras W_s ⇐ w_s^i;
5    # linhas 6 a 21: calcula a JWS (ws) para os pares de W_A × W_B e gera a lista de palavras similares (WM)
6    para todo w_a ∈ W_a faça
7       para todo w_b ∈ W_b faça
8          ws ⇐ calculaJWS(w_a, w_b);
9             se ws > φ_ws então
10               ListaDePalavrasSimilares WM ⇐ (w_a, w_b);
11   wn ⇐ conta elementos em WM; # obtém a quantidade de elementos na lista de palavra similares
12   wna ⇐ conta elemento em W_A; # obtém a quantidade de elementos na lista de palavras de a
13   # linhas 14: calcula a proporção entre a quantidade de palavras similares e a quantidade de palavras em A
14   wm ⇐ wn / wna;
15   retorna wm; # retorna o valor calculado para afs(a,b)
```

A seguir, apresentamos um exemplo de cálculo do $activity\ fuzzy\ similarity$. Para tanto, seja $S_a = 'designar\ de\ audiência'$ a *string* com o nome da atividade a e $\boldsymbol{S_b} = 'controle\ de\ audiência'$ a *string* com o nome da atividade b. Seja $\varphi_{wl} = 3$ o valor do parâmetro que estabelece tamanho mínimo da palavra e $\varphi_{ws} = 0{,}8$ o valor de referência para considerar duas palavras similares.

Considerando $\varphi_{wl} = 3$, as palavras com menos de três caracteres são descartadas. Então, temos os seguintes conjuntos de palavras:

$$\mathrm{W}_a = \{"designar",\ "audiência"\} \text{ e } \mathrm{W}_b = \{"controle",\ "audiência"\}$$

O próximo passo é calcular a JWS para os pares de palavras sobre $W_a \times W_b$. A Tabela 10 apresenta o resultado obtido para cada par de palavras. Considerando o valor definido para φ_{ws}, duas palavras só serão consideradas similares se o JWS for igual ou superior a 0,8.

Tabela 10 – Resultado da similaridade entre as palavras

PALAVRAS EM W_a	PALAVRAS EM W_b	JWS	RESULTADO
designar	controle	0,50	DISSIMILAR
designar	audiência	0,65	DISSIMILAR
audiência	controle	0,41	DISSIMILAR
audiência	audiência	1,00	SIMILAR

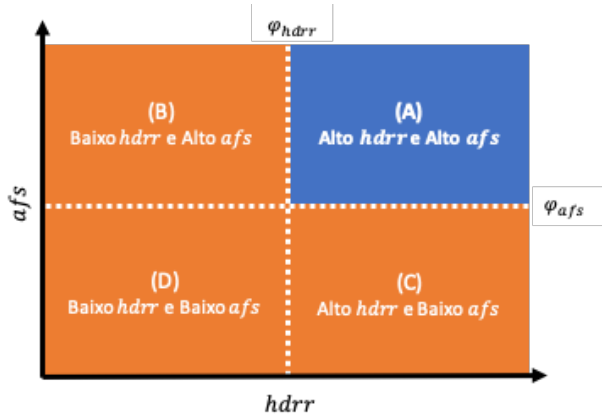
A Tabela 9 mostra que existe apenas uma palavra similar (audiência) em W_a e W_b. Portanto, temos $|WM| = 1$ e $|N_A| = 2$, resultando em $afs(a, b) = 0{,}5$.

5.4 Activity Fuzzy Match

Conforme abordado anteriormente, o objetivo da abordagem proposta neste capítulo é agrupar atividades afins promovendo assim uma mudança de granularidade do log de eventos. Para tanto, nossa abordagem precisa ser dotada da capacidade de identificar no log de eventos a existência de relações entre as atividades tais que seja coerente sua reunião em uma atividade mais abrangente preservando a consistência do processo. Partimos da premissa que as relações de precedência e a similaridade entre os nomes das atividades podem sugerir uma afinidade entre atividades; portanto, propusemos as métricas $highest\ direct\ relation\ ratio\ (hdrr)$ e $activity\ fuzzy\ similarity\ (afs)$. Quando tomadas individualmente, essas métricas não são suficientes para indicar se uma atividade pode ser agrupada com outra. Por exemplo, duas atividades podem ter similaridade entre os rótulos, mas participarem de contextos distintos no processo de negócio. O agrupamento destas atividades pela simples similaridade entre rótulos não faz sentido do ponto de vista de negócio. Por outro lado, quando combinadas, estas métricas podem indicar o tipo de afinidade que buscamos. A Figura 22 mostra quatro quadrantes obtidos ao combinar as métricas

$hdrr$ (eixo x) e afs (eixo y). Os quadrantes são delimitados por *thresholds* (φ_{hdrr} e φ_{afs}). O quadrante azul (A) contempla o cenário de alta propensão de afinidade entre atividades, pois apresenta valores altos para $hdrr$ e afs. Os demais quadrantes (B, C e D) correspondem aos cenários de baixa propensão para afinidade entre as atividades.

Figura 22 – Combinação das métricas *hdrr* e *afs*



Com base na ideia ilustrada na Figura 22, definimos um indicador, denominado $activity\ fuzzy\ match$. O indicador $activity\ fuzzy\ match$ é definido como segue:

Definição 24 ($activity\ fuzzy\ match$). Dado um log de eventos $\mathcal{L}$, seja $\mathcal{D}_a$ o domínio de atividades em $\mathcal{L}$. Seja $\varphi_{hdrr}, \varphi_{afs} \in \mathbb{R}[0..1]$ *thresholds* de $hdrr$ e afs respectivamente. Dadas as atividades $a, b \in \mathcal{D}_a$. O operador $activity\ fuzzy\ match$, denotado por $\bowtie$, é definido por:

$$a \bowtie b = \begin{cases} 1, & a \overset{+}{\rightarrow} b \geq \varphi_{hdrr} \wedge afs(a,b) \geq \varphi_{afs} \\ 0, & caso\ contrário \end{cases}$$

Como pode ser observado, o $activity\ fuzzy\ match$ conta com dois parâmetros: φ_{hdrr} e φ_{afs}. Para obter bons resultados com o uso do $activity\ fuzzy\ match$ é importante que os parâmetros estejam alinhados com as características dos logs de eventos. Infelizmente, não temos uma regra para determinar os parâmetros ideais, uma vez que diversos fatores podem influenciar no resultado. Por exemplo, o nível de estruturação do processo tende a impactar no valor ideal para φ_{hdrr}, já que quanto mais estruturado um processo menos variabilidade no comportamento será

observado, resultando em um aumento do valor médio do $hdrr$. Por outro lado, fatores como a convenção de nomes e o idioma podem afetar os valores médios de afs.

Visando estimar os parâmetros adequados para o domínio de aplicação dessa tese, realizamos um estudo utilizando uma técnica de aprendizagem supervisionada de máquina (árvore de decisão) para estimar os valores para φ_{hdrr} e φ_{afs}. Por se tratar de uma abordagem supervisionada, elaboramos, com apoio de especialistas, uma base rotulada a partir de um log de eventos real. Para este estudo também utilizamos o log de eventos do estudo preliminar (Seção 4.2) e do comparativo entre métricas apresentadas na Seção 5.2. A base em questão contemplou cinco atributos: (i) rótulo da atividade A, (ii) rótulo da atividade B, (iii) valor de $hdrr$, (iv) valor do afs e (v) classificação do especialista quanto a existência de afinidade entre as atividades. A Tabela 11 mostra um fragmento da base de dados rotulada.

Tabela 11 – Fragmento da base de dados rotulada utilizada na estimativa dos valores de φ_{hdrr} e φ_{afs}

ATIVIDADE A	**ATIVIDADE B**	$\boldsymbol{A \overset{+}{\rightarrow} B}$	$\boldsymbol{afs(A,B)}$	**AFIM?**
Citar (Inicial)	Finalizar Citar (Inicial)	0,98	1,00	SIM
Finalizar Citar (Inicial)	Controle de Audiência	0,81	0,00	NÃO
Controle de Audiência	Operações da Audiência	0,87	0,50	SIM
Operações da Audiência	Finalizar Controle de Audiência	0,96	0,50	SIM

Por fim, submetemos a base ao algoritmo C4.5, usando os valores de $A \overset{+}{\rightarrow} B$ e $afs(A,B)$ como variáveis independentes e a classificação atribuída pelo especialista como alvo. A Figura 23 mostra a árvore de decisão obtida. A partir desta consideramos os seguintes parâmetros para o todos os experimentos realizados nesta tese: $\varphi_{hdrr} = 0{,}4477$ e $\varphi_{afs} = 0{,}45$.

A Figura 24 mostra um gráfico de dispersão com a distribuição dos valores de $hdrr$ e afs para o log de eventos. Cada ponto no gráfico corresponde a uma transição identificada no log de eventos e o tamanho do ponto indica a frequência da transição. Assim, quanto maior o ponto, mais eventos correspondendo a transição são observados no log de eventos. Além disso, a cor indica o resultado do $activity\ fuzzy\ match$, sendo: laranja para $a \bowtie b = 0$ e azul para $a \bowtie b = 1$.

Na Figura 24 podemos observar que a uma quantidade significativa de atividades se encontram na zona de afinidade definida pelo $activity\ fuzzy\ match$. Além disso, observamos que os maiores pontos (transições) estão na “zona de afinidade” (região de pontos azuis), sugerindo que o agrupamento dessas atividades teria um impacto relevante no log de eventos.

A Seção 5.5 mostra o procedimento de transformação do log de a partir do $activity\ fuzzy\ match$.

Figura 23 – Árvore de decisão com valores estimados para os parâmetros do *activity fuzzy match*

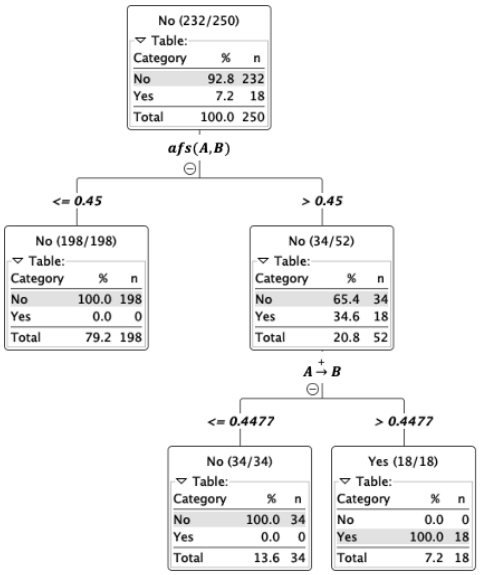
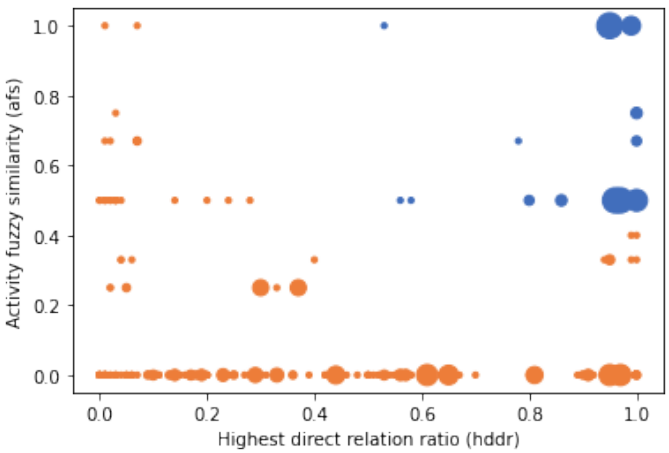


Figura 24 – Gráfico de dispersão de *hdrr* e *afs*



5.5 Transformação do Log de Eventos

Nesta seção apresentamos todos os procedimentos previstos na abordagem proposta neste capítulo. Conforme mostrado na Figura 20, a abordagem produz os seguintes artefatos: $lista\ de\ transições$, $lista\ de\ agrupamentos$ e $novo\ log\ de\ eventos$. A seguir apresentamos os algoritmos para geração de cada um destes artefatos.

O Algoritmo 2 apresenta o procedimento para geração da $lista\ de\ transições$ a partir do log de eventos.

Algoritmo 2
Geração da $lista\ de\ transições$

```
    entrada: LogDeEventos ℒ
    saída: ListaDeTransições τ
1   para todo e_i ∈ ℒ faça
2       se #_c(e_i) = #_c(e_{i+1}) | seja e_i, e_{i+1} eventos de ℒ e #_c o caso relacionado ao evento então
3           se (#_a(e_i), #_a(e_{i+1})) não existe em τ | seja #_a o atributo nome da atividade então
4               τ ⇐ (#_a(e_i), #_a(e_{i+1}));
5   retorna τ
```

O Algoritmo 3 apresenta o procedimento de geração $lista\ de\ agrupamentos$ a partir da $lista\ de\ transições$. Como pode ser visto no Algoritmo 3, cada elemento da $lista\ de\ atividades\ afins$ contém, além dos nomes das duas atividades, uma string composta pelas palavras que combinaram nos nomes das atividades. Essa informação será útil para denominar o agrupamento das duas atividades afins; portanto, a denominamos de $nome\ do\ agrupamento$.

Algoritmo 3
Geração da $lista\ de\ agrupamentos$.

```
    entrada: ListaDeTransições τ
    saída: ListaDeAtividadesAfins μ
1   para todo t_i^a, t_i^b ∈ τ faça
2       se t_i^a ⋈ t_i^b = 1 | seja t_i^a, t_i^b nomes das atividades em t_i então
3           nN ⇐ t_i^a ∩ t_i^b;  # nN recebe as palavras que combinaram entre t_i^a e t_i^b
4           μ ⇐ (t_i^a, t_i^b, nN);
7   retorna μ
```

De posse da $lista\ de\ agrupamentos$, temos duas opções: (1) criar um novo log de eventos substituindo o nome das atividades registradas nos eventos ou (2) inserir um novo atributo no log de eventos original registrando o $nome\ do\ agrupamento$. O benefício da primeira opção é a redução do log de eventos. Por outro lado, a segunda

opção oferece mais flexibilidade, possibilitando duas visões diferentes no mesmo log de eventos. Nesta tese adotamos a estratégia de criação de um novo log de eventos.

O procedimento de criação do $novo\ log\ de\ eventos$ consiste em percorrer todo o log de eventos original observando os eventos em pares sucessivos. Caso o par de eventos corresponda ao mesmo caso e as suas atividades integrem a $lista\ de\ agrupamentos$, então, será criado no $novo\ log\ de\ eventos$ um único evento resultante da fusão dos eventos originais. A fusão dos eventos se dá da seguinte forma: (i) o atributo atividade do novo evento passa a ser o $nome\ do\ agrupamento$ conforme indicado na $lista\ de\ agrupamentos$, o atributo de data/hora de início recebe o valor correspondente no primeiro evento e a data/hora de finalização recebe o valor correspondente no segundo evento. Se o caso registrado nos eventos for diferente ou as atividades correspondentes não integrarem a $lista\ de\ agrupamentos$, o primeiro evento do par é incorporado ao novo log de eventos tal qual registrado no log de eventos original. O procedimento de criação do log de eventos é apresentado no Algoritmo 4.

Algoritmo 4
Geração do novo log de eventos.

entrada: $ListaDeAgrupamentos\ \mu$, $LogDeEventos\ \mathcal{L}$
saída: $LogDeEventos\ \mathcal{L}^*$ | seja $\mathcal{L}^*$ o log de eventos transformado
1 **para todo** $e_i \in \mathcal{L}$ **faça**
2 **se** $\#_c(e_i) = \#_c(e_{i+1})$ | seja e_i, e_{i+1} eventos de $\mathcal{L}$ e $\#_c$ o atributo caso do evento **então**
3 **se** $\left(\#_a(e_i), \#_a(e_{i+1})\right)$ **existe em** μ | seja $\#_a$ o atributo nome da atividade **então**
4 $agrup \Leftarrow \mu\left(\#_a(e_i), \#_a(e_{i+1})\right)$ # recupera o nome do agrupamento registrado em μ
5 $\mathcal{L}^* \Leftarrow \left(\#_c(e_i), agrup, \#_i(e_i), \#_f(e_{i+1})\right)$ | seja $\#_i, \#_f$ atributos de data/hora inicial e final
6 **senão**
7 $\mathcal{L}^* \Leftarrow (e_i)$;
8 **retorna** $\mathcal{L}^*$

Cabe destacar que nossa abordagem preconiza uma intervenção em nível de evento e não de atividades. Essa opção se justifica pela necessidade de se preservar a coerência com a realidade. Em termos práticos, significa que ao identificar que duas atividades são afins não atualizamos todos os eventos relativos a estas atividades, mas apenas os que ocorrem em sequência.

A próxima seção apresenta um estudo avaliando os resultados obtidos com a abordagem proposta. Para tanto, utilizamos três logs de eventos reais (um já utilizado nos estudos apresentados anteriormente acrescido de outros dois logs de eventos).

5.6 Avaliação da abordagem de agrupamento de atividades

Nesta seção apresentamos experimentos para avaliação da abordagem proposta neste capítulo. Os três logs de eventos utilizados foram obtidos junto a dois tribunais de justiça brasileiros. A Tabela 12 apresenta as características dos logs de eventos utilizados em nossos experimentos.

Tabela 12 – Logs de eventos utilizados nos experimentos da tese

LOG DE EVENTOS	TRIBUNAL	CASOS	EVENTOS	ATIVIDADES	PERÍODO
1	1	3.359	73.412	76	13 meses
2	1	3.798	51.518	117	12 meses
3	2	4.229	50.718	102	6 meses

O primeiro log de eventos foi utilizado nos estudos preliminares (Seção 4.2) e no comparativo entre métricas de relação de precedência (Seção 5.2) e na estimativa dos parâmetros para o $activity\ fuzzy\ match$ (Seção 5.4). Em relação ao log de eventos 2, cabe registrar que, originalmente, continha 293.992 eventos e 136.851 casos. No entanto, uma inspeção nos dados mostrou que mais de 43% dos casos possuíam até duas atividades. Essa situação decorre do fato que o log de eventos contemplou uma janela com muitos casos recém iniciados. Para não comprometer a qualidade dos estudos, eliminamos do log casos com menos de dez eventos, restando 3.798 casos com 51.518 eventos. O log de evento 3 não demandou qualquer intervenção prévia. Cabe esclarecer que os três logs de eventos são pequenos para o contexto de dados reais, mas são bem representativos para o problema investigado, uma vez que os modelos de processos descobertos a partir deles são bastante complexos.

Como vimos nas seções anteriores, a métrica $activity\ fuzzy\ similarity$ requer os parâmetros φ_{wl} e φ_{ws}, sendo o primeiro responsável por estabelecer o tamanho mínimo de uma palavra no rótulo da atividade e o segundo responsável por estabelecer o limiar de similaridade entre palavras. Conforme mostrado na Seção 5.5, o $activity\ fuzzy\ match$ também conta com dois parâmetros para operar, consistindo em um *threshold* para cada uma das métricas utilizadas pelo indicador, ou seja, φ_{hdrr} e φ_{afs}. A Tabela 13 mostra: os parâmetros, as técnicas aos quais os parâmetros se

referem e os valores adotados nos experimentos realizados (a obtenção dos valores dos dois últimos parâmetros foi discutida na Seção 5.5).

Tabela 13 – Parâmetros utilizados nos experimentos para avaliação da abordagem de agrupamento de atividades

PARÂMETRO	MÉTRICA/INDICADOR	VALOR
$\boldsymbol{\varphi_{wl}}$	$activity\ fuzzy\ similarity$	4
$\boldsymbol{\varphi_{ws}}$	$activity\ fuzzy\ similarity$	0,8
$\boldsymbol{\varphi_{hdrr}}$	$activity\ fuzzy\ match$	0,4477
$\boldsymbol{\varphi_{afs}}$	$activity\ fuzzy\ match$	0,45

5.6.1 Análise do impacto da abordagem

Esta seção apresenta os principais resultados observados através de estudo realizados aplicando a abordagem de modificação de log de eventos baseada no indicador $activity\ fuzzy\ match$. Para avaliar o impacto da abordagem no log de eventos, comparamos os originais com os transformados pela abordagem proposta. A Tabela 14 mostra um comparativo entre os logs de eventos.

Tabela 14 – Impactos da abordagem de agregação de atividades afins

CATEGORIA OBSERVADA	LOG DE EVENTOS		
	1	2	3
(A) Redução na quantidade de eventos	38%	30%	42%
(B) Percentual de atividades completamente absorvidas pelos agrupamentos criados	30%	34%	35%
(C) Percentual de atividades parcialmente absorvidas pelos agrupamentos criados	29%	40%	41%
(D) Percentual de atividades que não sofreram impacto com os agrupamentos criados	41%	26%	24%
(E) Redução na quantidade de atividades (considerando as atividades eliminadas e os agrupamentos criados	9%	15%	14%
(F) Percentual de agrupamentos no novo log de eventos	23%	23%	25%

Na linha (A) da Tabela 14 podemos observar que houve uma redução de 30% a 42% no tamanho dos logs de eventos (quantidade de eventos). Isso representa uma redução significativa no tamanho dos arquivos que armazenam os logs de eventos, contribuindo assim para a economia de recursos computacionais (armazenamento e processamento). É importante destacar que os logs de eventos, em processos reais, tendem a ser muito grandes, podendo chegar facilmente à casa dos milhões de registros. Por outro lado, fatores como capacidade de processamento e regras de licenciamento limitam o tamanho do log de eventos nas ferramentas de mineração de processos. Assim, a redução do log de eventos pode ser útil ao lidar com grandes volumes de dados, ajudando a contornar as limitações impostas pelas ferramentas de mineração de processos ou infraestrutura computacional.

As demais linhas (B-F) da Tabela 14 mostram o impacto da abordagem nas atividades do processo. Cabe ressaltar que a quantidade de atividades não influencia diretamente no tamanho do log de eventos, mas repercute na complexidade dos modelos de processos proveniente destes. Pois, como vimos na Seção 3.3, quanto mais atividades, mais complexo um modelo de processo tende a ser.

A linha (B) da Tabela 14 indica o percentual de atividade que foram completamente absorvidas pelos agrupamentos. Estas atividades tiveram todos os seus eventos agrupados. Observamos a absorção completa dos eventos em 30% a 35% das atividades presentes nos logs de eventos analisados.

Na linha (C) da Tabela 14 temos o percentual de atividades parcialmente absorvidas por agrupamentos. A absorção parcial de atividades por agrupamentos ficou entre 29% (log de eventos 1) e 41% (log de eventos 3). Como vimos no Algoritmo 4 da Seção 5.5, a abordagem proposta nesta tese atua em nível de evento. Ocorre que uma atividade pode apresentar afinidade com outra, mas ainda assim se relacionar diretamente com uma terceira. Nessas circunstâncias, teremos parte dos eventos absorvidos pelo log de eventos e parte permanece conforme o original. Nesses casos, o efeito na simplificação do processo é reduzido, mas não podemos abrir mão da coerência do modelo do processo em nome de uma redução. Além do mais, a tendência é que o comportamento não absorvido seja um comportamento residual, estando, portanto, mais exposto à ação de ferramentas de eliminação de comportamento excepcional. No Capítulo 6 apresentamos uma abordagem aplicando a eliminação de comportamento infrequente associada à nossa abordagem de agregação.

Na linha (D) da Tabela 14 temos o percentual de atividades que não sofreram impacto com nossa abordagem. Ou seja, a abordagem não identificou afinidade entre estas atividades e qualquer outra no log de eventos, logo seus eventos permaneceram tal qual no log original. Como pode ser visto, os logs de eventos 2 e 3 apresentaram resultados semelhantes, respectivamente 26% e 24% das atividades não tiveram afinidade descoberta. Já no log de eventos 1 esse valor saltou para 41% das atividades sem afinidade com as demais.

Na linha (E) da Tabela 14 temos uma comparação entre a quantidade de atividades de eventos nos logs de eventos original e transformado. Como pode ser observado, houve redução de atividades da ordem de 9% a 15% nos logs de eventos analisados. Por fim, na linha (F) da Tabela 14 temos a parcela de atividades no log de eventos transformada oriunda de eventos agrupados. Como podemos observar os resultados nos três logs de eventos foram similares, variando entre 23% e 25%.

Além da análise dos efeitos nos logs de eventos proporcionados pela abordagem de agrupamento de atividades afins, conduzimos experimentos com o intuito de comparar os logs de eventos em relação à qualidade dos modelos de processos descobertos a partir destes. Para tanto, geramos vários modelos de processos com *Heuristic Miner* utilizando a biblioteca PM4Py (BERTI et al., 2019). A escolha do algoritmo *Heuristic Miner* se justifica por se tratar de algoritmo consagrado e a conversão do modelo resultante para redes de Petri é simples e eficaz.

De posse dos modelos, computamos o *fitness* e a precisão para cada modelo utilizando a biblioteca PM4Py. As métricas de simplicidade dos modelos foram avaliadas no *plugin Show Petri-net Metrics* do ProM, uma vez que a biblioteca PM4Py não oferece algoritmo para cálculo das métricas desejadas.

A seguir detalharemos como se deu a escolha dos parâmetros para geração dos modelos de processos. Conforme implementado no PM4Py, o *Heuristic Miner* conta com seis parâmetros que possibilitam definir o nível de sensibilidade para comportamento ruidoso. Geramos 64 configurações diferentes de parâmetros e aplicamos sobre cada um dos três logs de eventos originais ($\mathcal{L}^1$, $\mathcal{L}^2$, $\mathcal{L}^3$) e transformados ($\mathcal{L}_A^1$, $\mathcal{L}_A^2$, $\mathcal{L}_A^3$), resultando em 384 modelos de processos avaliados. Das 64 configurações adotadas 54 tiveram valores escolhidos através da combinação de valores baixos, intermediários e altos de forma a abranger um amplo espectro. As outras 10 configurações foram obtidas através de geração de valores aleatórios para os parâmetros. Os resultados detalhados para todos os modelos estão disponíveis no

Apêndice B (Métricas de Qualidade) desta tese.

A Tabela 15 mostra o valor médio obtido para quatro medidas de qualidade disponibilizadas pelo PM4Py, sendo as três primeira medidas de *fitness* (BERTI; VAN DER AALST, 2019) e a última uma medida de precisão (MUÑOZ-GAMA; CARMONA, 2010), todas baseadas em abordagem de reprodução de tokens.

A medida de qualidade MF1 *(perc_fit_traces)* captura o percentual de instâncias do processo contidas no log de eventos que são reproduzidas perfeitamente no modelo de processo. Podemos observar evolução nesta métrica quando da comparação dos logs de eventos transformados $\mathcal{L}_A^2$ e $\mathcal{L}_A^3$ com seus correspondentes originais $\mathcal{L}^2$ e $\mathcal{L}^3$.

Tabela 15 – Comparativo entre *fitness* e precisão dos modelos de processos descobertos para os logs de eventos ($\mathcal{L}^1$, $\mathcal{L}_A^1$, $\mathcal{L}^2$, $\mathcal{L}_A^2$, $\mathcal{L}^3$, $\mathcal{L}_A^3$)

MÉTRICA	$\mathcal{L}^1$	$\mathcal{L}_A^1$	$\mathcal{L}^2$	$\mathcal{L}_A^2$	$\mathcal{L}^3$	$\mathcal{L}_A^3$
MF1 *(perc_fit_traces[16])*	< 0,01	0,02	< 0,01	0,15	0,40	0,60
MF2 *(average_trace_fitness)*	0,97	0,94	0,93	0,94	0,97	0,98
MF3 *(log_fitness)*	0,97	0,94	0,93	0,94	0,97	0,98
MP *(precision)*	0,56	0,50	0,84	0,77	0,64	0,81

Os valores de MF2 *(average_trace_fitness)* e MF3 *(log_fitness),* representam, respectivamente, a média de *fitness* das instâncias do processo e uma medida de *fitness* do log de eventos. Podemos observar que os valores médios são similares para todos os logs de processos, indicando que a abordagem de agrupamento de atividades afins não interfere na qualidade do modelo em relação a dimensão *fitness*.

Em relação a métrica MP *(precision),* que mede a precisão do modelo do processo, podemos observar que houve uma leve redução quando da comparação dos logs de eventos transformados $\mathcal{L}_A^1$ e $\mathcal{L}_A^2$ com seus correspondentes originais $\mathcal{L}^1$ e $\mathcal{L}^2$. Já em relação a $\mathcal{L}_A^3$ e comparação com $\mathcal{L}^3$ observamos uma evolução.

Analisando a causa da redução, ainda que leve, observada na precisão dos logs de eventos $\mathcal{L}_A^1$ e $\mathcal{L}_A^2$, observamos que a presença de comportamento altamente

[16] Os valores de perc_fit_traces foram convertidos para o intervalo entre 0 e 1 para manter compatibilidade de escala com os demais valores da tabela.

infrequente prejudica a identificação de afinidade entre atividades. Sendo, portanto, necessário um tratamento prévio para sua eliminação do log de eventos.

Os estudos apresentados nessa seção foram realizados em um Dual-Core Intel i5 (1,8 GHz) com 8GB de RAM. O tempo aproximado de processamento das métricas MF1, MF2, MF3 e MP para os 384 modelos foi de 12 horas (3,5 horas para $\mathcal{L}^1$, 3 horas para $\mathcal{L}_A^1$, 1 hora para $\mathcal{L}^2$, 1 horas para $\mathcal{L}_A^2$, 3 horas para $\mathcal{L}^3$ e 1 hora para $\mathcal{L}_A^3$).

Também avaliamos o impacto da abordagem em relação a simplicidade dos modelos resultantes. Considerando que a variação na configuração dos parâmetros teve pouca influência na qualidade dos modelos resultantes, selecionamos apenas um modelo (configuração padrão do *Heuristic Miner* no PM4Py) oriundo de cada log de eventos para análise comparativa da simplicidade. A Tabela 16 apresenta os valores obtidos os logs de eventos ($\mathcal{L}^1$, $\mathcal{L}_A^1$, $\mathcal{L}^2$, $\mathcal{L}_A^2$, $\mathcal{L}^3$, $\mathcal{L}_A^3$) de três métricas de simplicidade (*Extended Cardoso metric* - ECaM, *Extended Cyclomatic metric* - ECyM e *Structuredness metric*) propostas em (LASSEN; VAN DER AALST, 2009). Os valores foram obtidos através do *plug-in Show Petri-net Metrics* do ProM 6.9.

Tabela 16 – Comparativo entre métricas de simplicidade dos modelos de processos descobertos para os logs de eventos ($\mathcal{L}^1, \mathcal{L}_A^1, \mathcal{L}^2, \mathcal{L}_A^2, \mathcal{L}^3, \mathcal{L}_A^3$)

MÉTRICA	$\mathcal{L}^1$	$\mathcal{L}_A^1$	$\mathcal{L}^2$	$\mathcal{L}_A^2$	$\mathcal{L}^3$	$\mathcal{L}_A^3$
ECaM	135	143	240	216	244	255
ECyM	N/A	N/A	N/A	95	133	140
Structuredness	88627	N/A	228480	187050	N/A	N/A

Cabe esclarecer que a métrica ECyM requer um modelo de processo com um gráfico de alcançabilidade finito e a métrica *Structuredness* é restrita a redes de *workflow*. Observamos que alguns modelos analisados não atendem aos requisitos para as métricas, inviabilizando a análise da simplicidade para estes casos. Os modelos de processos nessas condições foram indicados com “N/A” na Tabela 16.

Analisando os resultados apresentados na Tabela 16 não foi possível concluir se a abordagem de agrupamento de atividades afins impactou positiva ou negativamente na simplicidade dos modelos de processos descobertos. Pois, observamos uma discreta oscilação na métrica ECaM, quando comparados os logs

originais em relação aos seus correspondentes transformados. Além disso, as métricas ECyM e *Structuredness* foram prejudicadas pelo fato de parte dos modelos descobertos não atenderem aos seus requisitos.

Diante dos resultados observados consideramos análise de simplicidade dos modelos de processos inconclusiva, o que nos levou a realizar uma análise qualitativa visando avaliar outros aspectos da abordagem proposta neste capítulo.

5.6.2 Análise dos agrupamentos

Uma análise dos agrupamentos produzidos pela abordagem foi conduzida com o apoio de três especialistas no processo de negócio, sendo dois magistrados e um analista de negócio. Os magistrados selecionados atuam em unidades judiciárias representadas nos logs de eventos. Já o analista de negócio possui experiência de mais de cinco anos com o sistema PJe e com os processos de negócios judiciais.

Os especialistas contribuíram com uma análise comparativa entre os modelos de processos descobertos a partir dos logs de eventos originais em relação aos logs de eventos transformados pela abordagem apresentada neste capítulo.

A ferramenta Disco foi utilizada para descoberta dos modelos de processos. As Figuras 25 a 27 mostram fragmentos de modelos de processos nos quais atividades afins foram agrupadas pela abordagem apresentada neste capítulo. Cada figura mostra o fragmento original e o fragmento transformado pelo agrupamento de atividades afins. As atividades impactadas (agrupadas) são identificadas com rótulos em vermelho com a finalidade de facilitar a identificação da transformação proporcionada.

A Figura 25 mostra dois fragmentos de processos: o da esquerda oriundo do log de eventos original 1 e o da direita oriundo do log de eventos transformado 1. Como podemos ver no fragmento de modelo de processo apresentado, nossa abordagem propiciou dois agrupamentos $A = \{A_1, A_2\}$ e $B = \{B_1, B_2, B_3, B_4\}$, contemplando duas e quatro atividades respectivamente.

A Figura 26 mostra fragmentos equivalentes do modelo de processos original (lado esquerdo) e simplificado (lado direito) do log de eventos 2. Como pode ser observado, nesse fragmento há dois agrupamentos contemplando duas atividades cada $A = \{A_1, A_2\}$ e $B = \{B_1, B_2\}$.

A Figura 27 mostra fragmentos equivalentes dos processos original (lado esquerdo) e simplificado (lado direito) do log de eventos 3. No fragmento de modelo de processo mostrado pode-se observar a incidência de três agrupamentos (A, B, C) com duas atividades cada.

Os agrupamentos observados nos modelos de processos descobertos foram submetidos a análise dos especialistas para observação sob duas perspectivas: coerência e abrangência. Em relação à coerência, buscamos avaliar se o modelo de processo se manteve consistente do ponto de vista do negócio após a transformação no log de eventos. Em relação a abrangência, buscamos avaliar se o agrupamento contemplou todas as atividades que deveria. Todos os agrupamentos gerados nos três logs de eventos foram classificados nas seguintes categorias: (1) coerente e abrangente, (2) coerente e incompleto e (3) incoerente.

A Figura 28 apresenta um gráfico de barra com o resultado da avaliação de especialistas sobre os agrupamentos em cada um dos três logs de eventos. Entre 69% (logs de eventos 1 e 2) e 73% (log de eventos 3) dos presentes nos logs de eventos foram classificados como coerentes e abrangentes. Isso significa que a maior parte dos agrupamentos realizados pela abordagem proposta foi coerente do ponto de vista de negócio e contemplou a totalidade (ou próximo disso) das atividades que poderiam ser reunidas. Além disso, entre 23% (log de eventos 3) e 31% (logs de eventos 1 e 2) dos agrupamentos foram classificados como coerentes e parciais. Nestes casos, segundo os especialistas, a abordagem proposta promoveu agrupamentos coerentes, mas deixou de fora alguma atividade que poderia ser incorporada ao agrupamento. Analisando estes casos observamos que na maior parte deles a abordagem foi incapaz de reunir todas as atividades em único agrupamento, fragmentando-o em duas ou mais partes. Um exemplo deste cenário pode ser observado na Figura 26, no qual, do ponto de vista dos especialistas, as quatro atividades ao invés de formarem dois agrupamentos poderiam ser reunidas em um único.

Por fim, apenas um agrupamento foi classificado como incoerente. Analisando o caso observamos se tratar de efeito colateral de comportamento altamente infrequente (a atividade possuía apenas um evento). Portanto, novamente nos deparamos com uma situação que pode ser contornada com uma prévia eliminação de comportamento infrequente.

Figura 25 – Fragmentos de modelos de processo original (a) e simplificado (b) oriundos do log de eventos 1

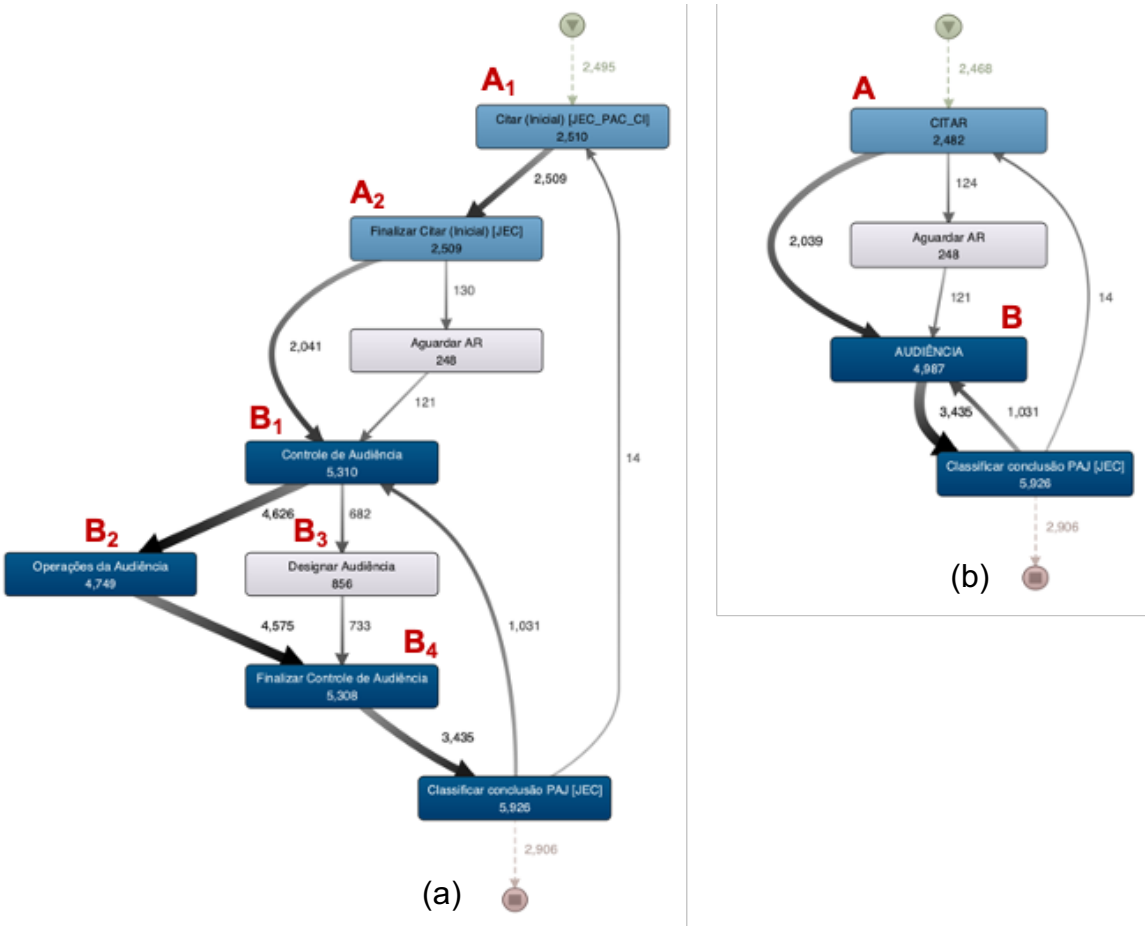


Figura 26 – Fragmentos de modelos de processo original (a) e simplificado (b) oriundos do log de eventos 2

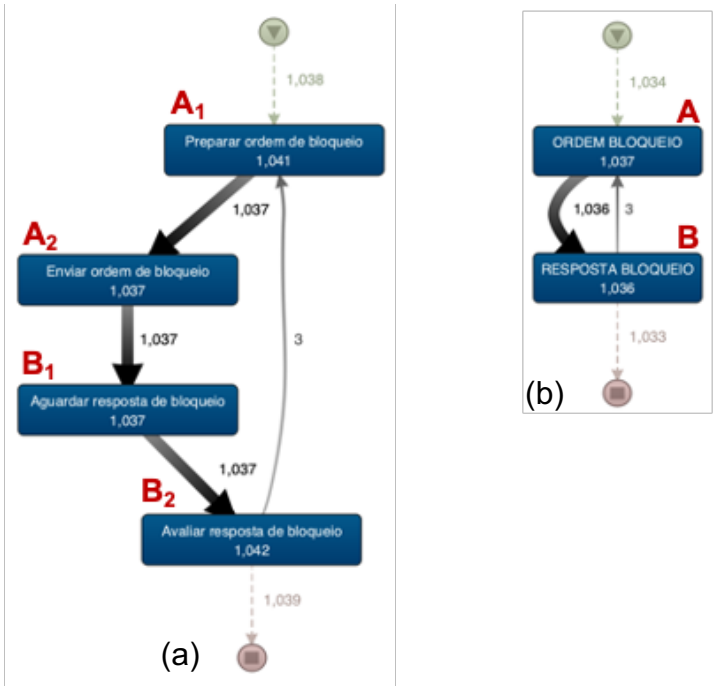


Figura 27 – Fragmentos de modelos de processo original (a) e simplificado (b) oriundos do log de eventos 3

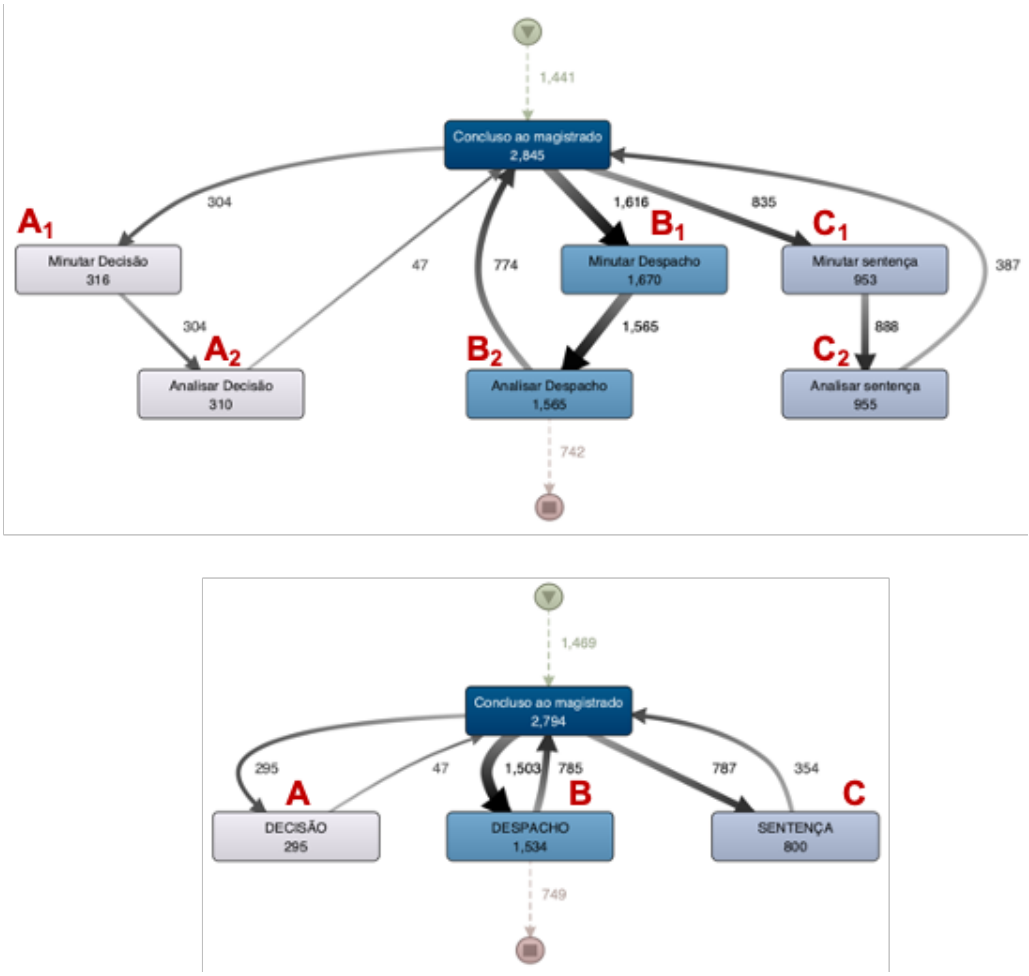
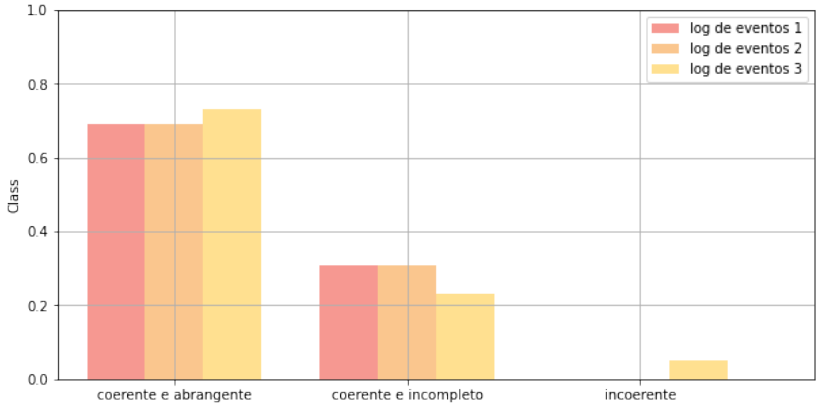


Figura 28 – Análise dos agrupamentos produzidos pela abordagem



5.7 *Conclusão do capítulo*

Este capítulo apresentou uma nova abordagem para transformação de logs de eventos através do agrupamento de eventos de atividades afins. Duas métricas e um indicador foram propostos para identificação da afinidade entre atividades. Além disso,

desenvolvemos um algoritmo para transformação do log de eventos comprometido com a preservação do comportamento real do processo. Como consequência, conseguimos reduzir os logs de eventos e oferecer uma visão alternativa do processo através da mudança na granularidade dos eventos.

A qualidade da abordagem foi avaliada por meio de um estudo usando três logs de eventos reais de dois tribunais de justiça brasileiros e contou com apoio de especialistas no negócio. Em relação à qualidade dos modelos resultantes dos logs de eventos transformados observamos uma melhoria discreta no *fitness* e uma oscilação na precisão. Já a análise da simplicidade foi prejudicada principalmente pelo fato de os modelos resultantes não atenderem aos requisitos da métricas. O *feedback* coletado junto aos especialistas no negócio sugere que a abordagem produz agrupamentos consistentes do ponto de vista do negócio e que estes reúnem todas as atividades que deveriam.

Com os resultados obtidos, consideramos que a abordagem apresentada neste capítulo pode contribuir em dois aspectos: (1) reduzir o tamanho do log de eventos e (2) oferecer uma visão alternativa dos processos através de uma abordagem automática para alteração da granularidade dos eventos de um log de eventos. A redução no tamanho dos logs de eventos contribui para a economia de recursos computacionais. Além disso, muitas ferramentas de mineração de processos têm limitações no número de eventos que podem manipular, portanto, a redução de logs de eventos pode ser útil ao lidar com grandes logs de eventos.

Observamos ainda que o comportamento extremamente infrequente reduz a eficácia da abordagem proposta, deixando comportamento residual atrelado às atividades originais ao invés de vincular a um agrupamento. Também creditamos ao comportamento residual o prejuízo na avaliação da simplicidade dos modelos de processos. Portanto, concluímos que a incorporação de tratamento de comportamento altamente infrequente na abordagem de agrupamento de atividades afins proposta poderia gerar modelos de processos melhores e mais simples. No Capítulo 6 apresentamos uma extensão da abordagem de agrupamento de atividades afins que incorpora a filtragem de comportamento infrequente.

6 COMPORTAMENTO INFREQUENTE

O comportamento infrequente está presente em processos reais. Também denominados de *ruídos* ou *outliers*, esses comportamentos podem resultar de imperfeições nos logs de eventos (SURIADI et al., 2017) ou, conforme abordado na Seção 4.1, resultarem em excepcionalidades capturadas nos log de eventos. Processos flexíveis permitem uma maior liberdade de execução, consequentemente, tendem a apresentar uma maior incidência de comportamento infrequente. Em geral, deseja-se a eliminação desse comportamento do modelo de processos descobertos (VAN DER AALST, 2016), pois, sua incorporação resulta em modelos complexos e incompreensíveis, ocultando o comportamento correto ou relevante do processo subjacente (SANI; VAN ZELST; VAN DER AALST, 2017). Por outro lado, nem todo comportamento excepcional é irrelevante. Em certos domínios de aplicação, compreender o impacto do comportamento infrequente pode ser de grande valor, tais como na melhoria dos fluxos ou mesmo na realização de auditorias. Distinguir a sua causa é uma tarefa desafiadora, fazendo com que a quase totalidade das abordagens propostas tratem o comportamento infrequente como ruído e prescrevem como solução a sua eliminação do log de eventos.

No Capítulo 5, observamos que o comportamento infrequente reduz a capacidade de identificação de afinidade entre as atividades pela abordagem proposta. Os resultados obtidos nos experimentos nos permitiram concluir que incorporar uma etapa prévia de filtragem de comportamento infrequente traria melhores resultados. Então, promovemos estudos para mapeamento das características das principais técnicas de pré-processamento de log de eventos voltadas a filtragem de comportamento infrequente. Em seguida, formulamos uma abordagem para filtragem de comportamento infrequente integrada a abordagens de agregação de atividades afins. Também repetimos os estudos apresentados no Capítulo 5 para comparar o resultado da abordagem de agregação de atividades afins com e sem a filtragem de comportamento infrequente.

Este capítulo está organizado nas seguintes seções: A Seção 6.1 apresenta as técnicas de pré-processamento de log de eventos para eliminação do comportamento infrequente descritas na literatura. A Seção 6.2 mostra a abordagem proposta para filtragem de comportamento infrequente. A Seção 6.3 apresenta a avaliação da

abordagem proposta. Por fim, a Seção 6.4 contempla a conclusão, apresentando os resultados obtidos com incorporação da técnica de filtragem de comportamento infrequente em nossa abordagem de agrupamento de atividades afins.

6.1 Técnicas de tratamento de comportamento infrequente

Duas estratégias são adotadas no tratamento de comportamento infrequente: (1) abstração de comportamento infrequente e (2) filtragem de comportamento infrequente no log de eventos. No primeiro, a ideia é omitir do modelo de processos descoberto o comportamento infrequente sem alterar o log de eventos. Já na segunda, a estratégia consiste em eliminar os registros de comportamento infrequente dos logs de eventos. Ou seja, no primeiro o comportamento é ignorado na construção do modelo de processo e no segundo a informação é eliminada do conjunto de dados.

A filtragem de comportamento infrequente no log de eventos tem maior afinidade com o escopo dessa tese, uma vez que é uma ação tipicamente realizada na etapa de pré-processamento. Porém, essa não é uma tarefa trivial, uma vez que pode ser custosa do ponto de vista computacional e não há garantia de eficácia, especialmente no contexto de processos complexos (SURIADI et al., 2017). Segundo Conforti, Rosa e Hofstede (CONFORTI; ROSA; HOFSTEDE, 2017), a literatura na área de filtragem de comportamento infrequente em log de eventos é muito escassa e, em geral, as abordagens são simplistas. Em recente revisão da literatura, Van Zelst et al. (VAN ZELST et al., 2020) destacaram algumas técnicas mais refinadas para tratamento de comportamentos infrequentes através do pré-processamento de logs de eventos (CHAPELA-CAMPA; MUCIENTES; LAMA, 2019; CONFORTI; ROSA; HOFSTEDE, 2017; FANI SANI; VAN ZELST; VAN DER AALST, 2018a; SANI; VAN ZELST; VAN DER AALST, 2017; SUN et al., 2019).

Em (CONFORTI; ROSA; HOFSTEDE, 2017) temos uma técnica automatizada para filtrar sistematicamente comportamentos pouco frequentes de logs de eventos. Para tanto, cria um autômato (grafo direcionado) que captura as relações de dependências (direta) entre os eventos do log. Em seguida, é calculada uma medida de frequência relativa entre transições do autômato e as que obtêm valor inferior ao *threshold* estabelecido são removidas do autômato. Finalmente, o log original é reproduzido no autômato e o comportamento não alinhado é removido. Cabe ressaltar que a abordagem proposta prevê uma etapa na qual o usuário pode indicar

comportamentos obrigatórios para manutenção da coerência do modelo resultante. Ou seja, comportamentos que devem ser preservados no log de eventos mesmo que sejam infrequentes, por exemplo, atividades iniciais e finais. Essa etapa requer que o usuário indique explicitamente aos algoritmos quais transições devem ser preservadas. Os autores afirmam que a abordagem foi implementada no ProM, contudo constatamos que a mesma não está mais disponível. Apenas uma versão *standalone* implementada em Java pode ser encontrada no site de um dos autores do trabalho.

Em (SANI; VAN ZELST; VAN DER AALST, 2017) temos uma abordagem de filtragem de comportamento infrequente genérica baseada em probabilidades condicionais entre sequências de atividades. A técnica recebe como entrada um log de eventos e um parâmetro (limite) e retorna um log de eventos filtrado. Para tanto, usa-se a probabilidade condicional da ocorrência de uma atividade após uma determinada sequência de atividades. Se essa probabilidade for menor que o limite especificado, o evento será descartado.

Em (FANI SANI; VAN ZELST; VAN DER AALST, 2018a) temos uma extensão de (FANI SANI; VAN ZELST; VAN DER AALST, 2018b) que propõe um método de "correção" de comportamento infrequente. A abordagem usa um método probabilístico para identificação do comportamento infrequente que é substituído por um comportamento mais provável (comportamento frequente mais próximo do comportamento observado).

Em (CHAPELA-CAMPA; MUCIENTES; LAMA, 2019) temos o algoritmo WoSimp, que busca simplificar processos abstraindo o comportamento pouco frequente do log. O algoritmo busca preservar não apenas atividades frequentes, mas também subprocessos frequentes identificados.

Em (SUN et al., 2019) temos uma abordagem que parte da premissa que eventos fortemente relacionados a outros correspondem ao comportamento "normal", enquanto que eventos com relacionamentos fracos têm mais probabilidade de se tratar de ruído. O método associa dois níveis de dependência (local e global) e usa uma abordagem probabilística para realização da filtragem de eventos. Além disso, a abordagem realiza uma filtragem dupla contemplando dois níveis de granularidade.

Nos cinco métodos descritos observamos a adoção de uma medida da relação de precedência entre as atividades com um *thresholds* associado para identificação de comportamento infrequente, seguido por uma etapa de transformação do log de

eventos para eliminação do comportamento infrequente. Além disso, essas técnicas se baseiam nas premissas de que a maior parte do comportamento infrequente de um log de eventos decorre de erros nos registros dos eventos e que é muito difícil separar o comportamento infrequente proveniente de ruídos nos dados de uma excepcionalidade intencional na execução do processo. No domínio em que aplicamos a abordagem, observamos que os dados são consistentes, sendo raras as imperfeições nos registros dos eventos; logo, o comportamento infrequente é predominantemente fruto de excepcionalidades intencionais ou erros na operação do sistema. Além disso, nosso objetivo com a filtragem de comportamento infrequente é apenas eliminar o comportamento extremamente infrequente para aumentar a qualidade dos agrupamentos gerados. Ou seja, não almejamos propor uma nova abordagem de filtragem de comportamento, buscamos uma abordagem que possa melhorar os resultados da abordagem proposta para agrupamento de atividades afins através da redução do comportamento residual produzido por esta. Assim, analisamos as características das técnicas disponíveis na literatura e propusemos uma abordagem simplificada para filtragem de comportamento infrequente.

6.1.1 Identificação de comportamento infrequente

O primeiro passo para a eliminação de comportamento infrequente é a identificação deste. Na Seção 5.2 apresentamos uma métrica, denominada de $highest\ direct\ relation\ ratio\ (hdrr)$ para aferir o nível da relação de precedência entre duas atividades. A métrica, que captura o maior valor entre as métricas dfr e dpr, se mostrou eficaz para identificação de afinidade entre as atividades. Partimos da premissa que uma pequena adaptação seria suficiente para possibilitar a identificação de transições pouco relevantes. Assim, propomos a métrica denominada de $lowest\ direct\ relation\ ratio\ (ldrr)$. A definição formal a métrica $ldrr$ é a seguinte:

Definição 25 ($ldrr$). Seja $\mathcal{D}_A$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Dadas as atividades $a, b \in \mathcal{D}_A \wedge |a| > 1 \wedge |b| > 1$. O operador $ldrr$, denotado por $\bar{\rightarrow}$, é definido como segue:

$$x \bar{\rightarrow} y = \begin{cases} 0, & |a| = 1 \vee |b| = 1 \\ \min\left(dfr(a,b), dpr(a,b)\right), & caso\ contrário \end{cases}$$

Por exemplo, dado o log de eventos $\mathcal{L} = [\langle a, b, c\rangle^{10}, \langle a, c, b\rangle^{4}, \langle a, b, a\rangle^{2}]$, temos que $dfr(a, b) = 0{,}67$ e $dpr(a, b) = 0{,}75$. Logo, $a \bar{\rightarrow} b = 0{,}67$. Cabe observar que adicionamos uma exceção para a condição em que uma das atividades da transição possua apenas um evento.

Comparamos a métrica proposta $ldrr$ com a métrica $frequência\ relativa\ (fr)$ (CONFORTI; ROSA; HOFSTEDE, 2017) no tocante a identificação de comportamento extremamente infrequente. Novamente tomamos por base o log de eventos 1 (Tabela 12). Selecionamos um subconjunto de transições com frequência absoluta muito baixa (transições observadas cinco vezes ou menos no log de eventos). Em seguida, analisamos os valores obtidos para ambas as métricas usando um *thresholds* (φ) com valor baixo ($\varphi = 0{,}1$). A escolha de um *thresholds* baixo se justifica pelo fato de buscarmos a identificação de comportamento extremamente infrequente. A Tabela 17 mostra o resultado obtido com ambas as métricas em transições que foram observadas dez vezes ou menos no log de eventos, ou seja, transições com frequência absoluta menor ou igual a 10. Como podemos ver na Tabela 17, temos uma quantidade maior de transições abaixo do *thresholds* com a métrica $ldrr$ (90%) em comparação com a métrica fr (82%).

Tabela 17 – Transições classificadas a partir das métricas *fr* e *ldrr*

TRANSIÇÕES COM BAIXA FREQUÊNCIA ABSOLUTA		
$\boldsymbol{fr(a, b) < 0{,}1}$	$\boldsymbol{a \bar{\rightarrow} b < 0{,}1}$	**TOTAL**
82%	90%	116

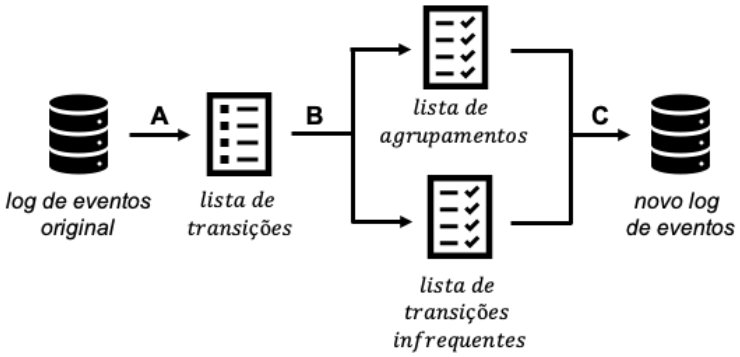
Analisando individualmente as transições do log de eventos, observamos que 92,2% (107 de 116) destas obtiveram valor inferior a $0{,}1$ para ambas as métricas e 7,8% (9 de 116) obtiveram valor discrepante de $ldrr$ e fr. Analisando as classificações díspares com o apoio de especialistas no negócio consideramos a métrica $ldrr$ mais adequada por dois aspectos: (1) compatibilidade entre as métricas usadas na abordagem de agrupamento de atividades afins e filtragem de comportamento infrequente e (2) maior capacidade de identificação de comportamento infrequente.

A próxima seção apresenta a abordagem para filtragem de comportamento infrequente integrada à abordagem de agrupamento de atividades afins.

6.1.2 Filtragem de comportamento infrequente

Nesta seção apresentamos a nossa abordagem para filtragem de log de eventos integrada a abordagem de agregação de atividades afins. A Figura 29 apresenta uma visão geral da abordagem.

Figura 29 – Abordagem de filtragem de comportamento infrequente integrada à abordagem de agregação de atividades afins



Em comparação com a abordagem apresentada no Capítulo 5 dessa tese, podemos observar a inclusão de uma lista de $lista\ de\ transições\ infrequentes$ no processo de transformação do log de eventos. A identificação das atividades infrequentes se dá através de métrica $ldrr$ aliada a um *thresholds* (φ_{ldrr}). O procedimento de geração da $lista\ de\ transições\ infrequentes$ é apresentado no Algoritmo 5.

Algoritmo 5
Geração da $lista\ transições\ infrequentes$

```
     entrada: ListaDeTransições τ, Threshold φ_ldrr;
     saída: ListaDeTransiçõesInfrequentes ϑ
1    para todo t_i^a, t_i^b ∈ τ faça
2        se t_i^a → t_i^b ≥ φ_ldrr então
3            ϑ ⇐ (t_i^a, t_i^b);
4    retorna ϑ
```

O procedimento de criação do novo log de eventos é bastante similar ao apresentado no Algoritmo 4. A diferença é que há uma verificação se a transição possui valor de $ldrr$ superior ao *thresholds* φ_{ldrr}. Se não for, o evento original é eliminado. O procedimento de transformação do log de eventos através da abordagem

de agrupamento de atividades afins com filtragem de comportamento infrequente é apresentado no Algoritmo 6.

Algoritmo 6
Geração do novo log de eventos.

```
    entrada: ListaDeAgrupamentos μ, ListaDeTransiçõesInfrequentes ϑ LogDeEventos ℒ
    saída: LogDeEventos ℒ* | seja ℒ* o log de eventos transformado
1   para todo e_i ∈ ℒ faça
2     se #_c(e_i) = #_c(e_i+1) | seja e_i, e_i+1 eventos de ℒ e #_c o atributo caso do evento então
3     se (#_a(e_i), #_a(e_i+1)) não existe em ϑ | seja #_a o atributo nome da atividade então
4           se (#_a(e_i), #_a(e_i+1)) existe em μ | seja #_a o atributo nome da atividade então
5             agrup ⇐ μ(#_a(e_i), #_a(e_i+1))  # recupera o nome do agrupamento registrado em μ
6             ℒ* ⇐ (#_c(e_i), agrup, #_i(e_i), #_f(e_i+1)) | seja #_i, #_f atributos de data/hora inicial e final
7           senão
8             ℒ* ⇐ (e_i) ;
9   retorna ℒ*
```

A Tabela 18 mostra um comparativo entre os resultados observados com a abordagem de agrupamento de atividades afins sem (abordagem 1) e com (abordagem 2) eliminação de comportamento infrequente. O comparativo foi realizado sobre o log de eventos 1 (Tabela 12).

Tabela 18 – Comparativo entre as abordagens sem (abordagem 1) e com (abordagem 2) eliminação de comportamento infrequente

CATEGORIA OBSERVADA	ABORDAGEM 1	ABORDAGEM 2
(A) Redução na quantidade de eventos (impacto no tamanho do log de eventos)	38%	46%
(B) Percentual de atividades completamente absorvidas em agrupamentos ou filtrada por comportamento infrequente	30%	70%
(C) Percentual de atividades parcialmente absorvidas pelos agrupamentos criados	29%	26%
(D) Percentual de atividades que não sofreram impacto com os agrupamentos criados	41%	4%
(E) Redução na quantidade de atividades	9%	50%
(F) Percentual de agrupamentos no novo log de eventos	23%	39%

Na linha (A) da Tabela 18 podemos observar que a filtragem de comportamento infrequente proporcionou um incremento moderado na redução de eventos do log. Contudo, no tocante às atividades observamos que houve um incremento substancial

entre a abordagem 1 e 2 (de 9% para 50%). Na linha (B) podemos observar que a filtragem de comportamento infrequente potencializou o agrupamento de atividades, passando a absorver completamente 70% das atividades do log. Como resultado, na linha (F) vemos que a quantidade de agrupamentos no log de eventos saltou de 23% no log de eventos gerado pela abordagem 1 para 39% no log de eventos gerado pela abordagem 2. Além disso, na linha (D) temos que a quantidade de atividades que não sofreram qualquer transformação foi reduzida de 41% para apenas 4%. Os resultados apresentados na Tabela 18 demonstram que o log de eventos após transformado pela abordagem 2 proporciona modelos de processos mais simples que a abordagem 1.

Na próxima seção apresentamos uma comparação entre os resultados observados com as abordagens de filtragem de comportamento infrequente aliada ao agrupamento de atividades afins em outros logs de eventos.

6.2 Avaliação da abordagem

Nesta seção apresentamos os resultados dos experimentos realizados com abordagem proposta neste capítulo (agregação de atividades afins com filtragem de comportamento infrequente). Para fins de comparação, repetimos os experimentos com os mesmos três logs de eventos detalhados na Tabela 12 da Seção 5.6. Além disso, mantivemos os mesmos parâmetros utilizados nos experimentos anteriores.

6.2.1 Análise do impacto da abordagem

Conforme já indicado na seção anterior, o agrupamento de atividades afins conforme apresentado no Capítulo 5 será denominado de abordagem 1 e a abordagem de agrupamento de atividades afins com a filtragem de comportamento infrequente será denominado de abordagem 2. Os estudos conduzidos sobre as duas abordagens utilizaram os mesmos logs de eventos e parâmetros. Em relação ao parâmetro específico da filtragem de comportamento infrequente foi adotado o valor de $\varphi_{ldrr} = 0{,}1$.

A Tabela 19 mostra os efeitos observados no log de eventos transformado com abordagem 2 quando comparados com os logs de eventos originais. Podemos observar que os resultados das seis categorias analisadas foram semelhantes nos três logs de eventos analisados, demonstrando consistência da abordagem.

Tabela 19 – Impactos da abordagem de agregação de atividades afins com filtragem de comportamento infrequente

CATEGORIA OBSERVADA	LOG DE EVENTOS		
	1	2	3
(A) Redução na quantidade de eventos (impacto no tamanho do log de eventos)	46%	45%	53%
(B) Percentual de atividades completamente absorvidas em agrupamentos ou filtrada por comportamento infrequente	70%	62%	64%
(C) Percentual de atividades parcialmente absorvidas pelos agrupamentos criados	26%	38%	28%
(D) Percentual de atividades que não sofreram impacto com os agrupamentos criados	4%	4%	8%
(E) Redução na quantidade de atividades (considerando as atividades eliminadas e os agrupamentos criados)	50%	45%	46%
(F) Percentual de agrupamentos no novo log de eventos	39%	23%	33%

A seguir apresentamos gráficos (Figuras 30 a 32) comparando os resultados obtidos com as abordagens 1 e 2. Os dados para elaboração destes gráficos foram extraídos da Tabela 14 (abordagem 1) e Tabela 19 (abordagem 2).

Na Figura 30 temos o comparativo das abordagens 1 e 2 para o log de eventos 1. Como pode ser visto, houve melhoria em todos os aspectos avaliados. Essa redução foi proporcionada pelo incremento substancial na absorção total de atividades por agrupamentos ou filtrada por comportamento infrequente. Em relação a absorção parcial de atividades por agrupamentos (C) apresentou uma redução modesta.

Figura 30 – Gráfico para comparação entre as abordagens (log de eventos 1)

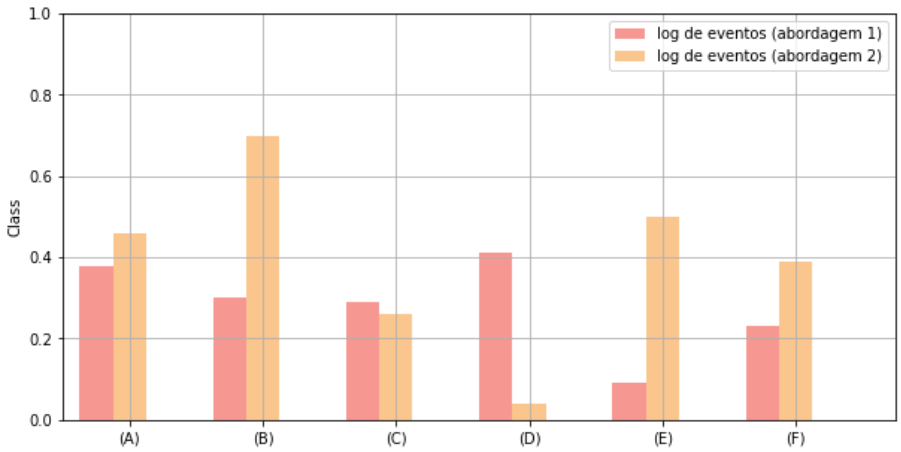


Figura 31 – Gráfico para comparação entre as abordagens (log de eventos 2)

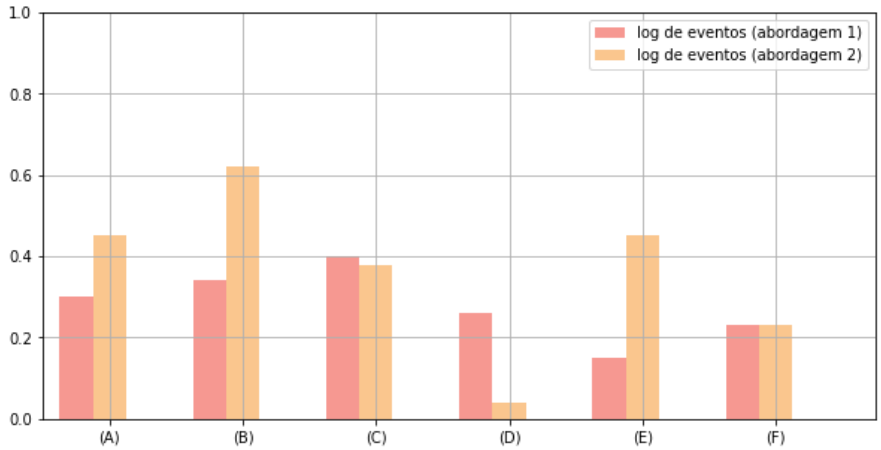
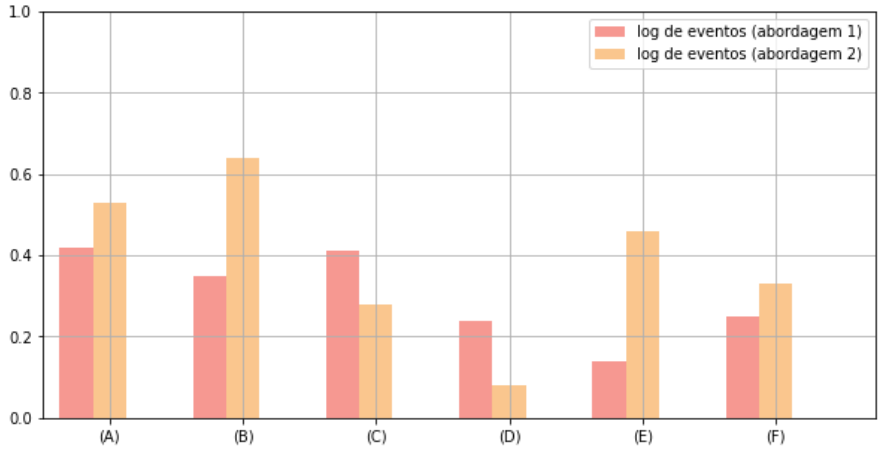


Figura 32 – Gráfico para comparação entre as abordagens (log de eventos 3)



Nas Figuras 31 e 32 podemos perceber resultados similares ao observado na Figura 30. Podemos destacar que, em relação ao log de eventos 2 (Figura 30), a categoria (F) foi igual para ambas as abordagens, indicando que o efeito no aumento nos agrupamentos proporcionado pela eliminação de comportamento foi menor que o observado no log de eventos 1.

Também realizamos estudo para avaliar se houve incremento na qualidade dos modelos descobertos a partir dos logs de eventos transformados com a abordagem 2 em comparação com os modelos obtidos com o log de eventos transformados pela abordagem 1. A Tabela 20 estende a Tabela 15 da Seção 5.6.1 adicionando os valores médios para os logs de eventos ($\mathcal{L}_F^1, \mathcal{L}_F^2, \mathcal{L}_F^3$) obtidos para as quatro medidas de qualidade: MF1 *(perc_fit_traces)*, MF2 *(average_trace_fitness)*, MF3 *(log_fitness)* e MP *(precision)* disponibilizadas no PM4Py.

Tabela 20 – Comparativo entre métricas de qualidade dos modelos de processos descobertos para os logs de eventos ($\mathcal{L}^1, \mathcal{L}_A^1, \mathcal{L}_F^1, \mathcal{L}^2, \mathcal{L}_A^2, \mathcal{L}_F^2, \mathcal{L}^3, \mathcal{L}_A^3, \mathcal{L}_F^3$)

MÉTRICA	$\mathcal{L}^1$	$\mathcal{L}_A^1$	$\mathcal{L}_F^1$	$\mathcal{L}^2$	$\mathcal{L}_A^2$	$\mathcal{L}_F^2$	$\mathcal{L}^3$	$\mathcal{L}_A^3$	$\mathcal{L}_F^3$
MF1	< 0,01	0,02	0,01	< 0,01	0,15	0,44	0,40	0,60	0,77
MF2	0,97	0,94	0,96	0,93	0,94	0,95	0,97	0,98	0,99
MF3	0,97	0,94	0,96	0,93	0,94	0,95	0,97	0,98	0,99
MP	0,56	0,50	0,66	0,84	0,77	0,85	0,64	0,81	0,87

Na primeira linha da Tabela 20 podemos observar que a abordagem 2 proporcionou uma evolução em relação à quantidade de instâncias que foram reproduzidas integralmente nos modelos de processos derivados dos logs de eventos 2 e 3. Nas linhas 2, 3 e 4 da Tabela 20 podemos observar que as medidas globais de *fitness* e precisão obtidos com os log de eventos transformados pela abordagem 2 ($\mathcal{L}_F^1, \mathcal{L}_F^2, \mathcal{L}_F^3$) são superiores aos obtidos com os logs das abordagens 1 ($\mathcal{L}_A^1, \mathcal{L}_A^2, \mathcal{L}_A^1$) e pelos logs de eventos originais ($\mathcal{L}^1, \mathcal{L}^2, \mathcal{L}^3$).

Os experimentos foram realizados em um Dual-Core Intel i5 (1,8 GHz) com 8GB de RAM. O tempo aproximado de processamento das métricas MF1, MF2, MF3 e MP para os 64 novos modelos foi de aproximadamente 2 horas de processamento (1 hora e 20 minutos para $\mathcal{L}_F^1$, 20 minutos para $\mathcal{L}_F^2$ e 20 minutos para $\mathcal{L}_F^3$). Os resultados detalhados se encontram no Apêndice B (Métricas de Qualidade) desta tese.

Em relação a simplicidade, realizamos análise semelhante a conduzida na Seção 5.6.1, na qual os modelos de referência (configuração padrão do *Heuristic Miner* no PM4Py) foram comparados através das métricas de simplicidade (*Extended Cardoso metric* - ECaM, *Extended Cyclomatic metric* - ECyM e *Structuredness metric*) proposta em (LASSEN; VAN DER AALST, 2009). A Tabela 21 apresenta os valores obtidos os logs de eventos ($\mathcal{L}^1$, $\mathcal{L}_A^1$, $\mathcal{L}_F^1$, $\mathcal{L}^2$, $\mathcal{L}_A^2$, $\mathcal{L}_F^2$, $\mathcal{L}^3$, $\mathcal{L}_A^3$ e $\mathcal{L}_F^3$) através do *plug-in Show Petri-net Metrics* do ProM 6.9.

Como podemos ver na Tabela 21 alguns modelos avaliando para as métricas ECyM e *Structuredness* ainda não atenderam os requisitos das métricas (identificados como “N/A”). Contudo, vislumbramos que os resultados já permitem avaliar os modelos em relação à simplicidade dos modelos de processos.

Tabela 21 – Comparativo entre métricas de simplicidade dos modelos de processos descobertos para os logs de eventos ($\mathcal{L}^1, \mathcal{L}_A^1, \mathcal{L}_F^1, \mathcal{L}^2, \mathcal{L}_A^2, \mathcal{L}_F^2, \mathcal{L}^3, \mathcal{L}_A^3, \mathcal{L}_F^3$)

MÉTRICA	$\mathcal{L}^1$	$\mathcal{L}_A^1$	$\mathcal{L}_F^1$	$\mathcal{L}^2$	$\mathcal{L}_A^2$	$\mathcal{L}_F^2$	$\mathcal{L}^3$	$\mathcal{L}_A^3$	$\mathcal{L}_F^3$
ECaM	135	143	128	240	216	128	244	255	185
ECyM	N/A	N/A	N/A	N/A	95	64	133	140	108
Struct.	88627	N/A	62500	228480	187050	512	N/A	N/A	N/A

A partir da linha 1 na Tabela 21 (métrica ECaM) observamos uma redução consistente nos valores para os modelos de processos oriundos dos logs de eventos da abordagem 2 ($\mathcal{L}_F^1$, $\mathcal{L}_F^2$ e $\mathcal{L}_F^3$) quando comparado com os demais, indicando que, em relação à sintaxe, os modelos oriundos de log de eventos transformados pela abordagem 2 são mais simples que os demais. Em relação a métrica ECyM, na segunda linha da Tabela 21, podemos observar que o modelo resultante do log de eventos $\mathcal{L}_F^2$ é mais simples que o modelo oriundo de $\mathcal{L}_A^2$ e que o modelo oriundo do log de eventos $\mathcal{L}_F^3$ é mais simples que os modelos de processos oriundos dos logs de eventos $\mathcal{L}^3$ e $\mathcal{L}_A^3$. Logo, do ponto de vista do comportamento permitido, os modelos resultantes de logs de eventos transformados pela abordagem 2 são mais simples que os demais. Em relação a métrica *Structuredness*, terceira linha da Tabela 21, observamos uma redução quando comparamos o modelo resultante do log de eventos $\mathcal{L}_F^1$ quando comparado a $\mathcal{L}^1$, bem como em $\mathcal{L}_F^2$ quando comparados a $\mathcal{L}^2$ e $\mathcal{L}_A^2$.

Diante dos resultados observados, consideramos que os modelos resultantes dos logs de eventos gerados pela abordagem 2 são mais simples que os modelos de processos oriundos dos logs de eventos originais e dos logs de eventos gerados pela abordagem 1.

6.2.2 Análise dos agrupamentos gerados

No Capítulo 5 vimos que a abordagem de agregação de atividades afins produziu 16 agrupamentos, destes 11 foram considerados “coerente/abrangente” e 5 “coerente/incompleto” pelos especialistas. Com a abordagem de filtragem de comportamento infrequente (abordagem 2) observamos que foram produzidos 14 agrupamentos, sendo 10 considerados “coerente/abrangente” e 4 “coerente/incompleto”.

No log de eventos 2 tivemos 23 agrupamentos gerados com a primeira abordagem. Esse número foi reduzido para 20 agrupamentos com a incorporação da filtragem de comportamento infrequente, sendo 13 agrupamentos “coerente/abrangente” e 7 “coerente/incompleto”.

A abordagem 1 produziu 22 agrupamentos com o log de eventos 3, já a abordagem 2 gerou 17 agrupamentos para esse log de eventos. Cabe ressaltar que a abordagem 1 havia gerado um cluster incoerente, problema resolvido com a abordagem 2. Dos 17 agrupamentos gerados temos 15 agrupamentos “coerente/abrangente” e 2 “coerente/incompleto”.

Como podemos observar, embora tenha havido uma redução na quantidade de agrupamentos produzidos, houve uma melhoria na qualidade destes quando gerados pela abordagem 2, uma vez que observamos menos agrupamentos incompletos. Além disso, a única ocorrência de agrupamento incoerente foi prevenida com a filtragem de comportamento infrequente.

6.3 Conclusão do capítulo

Este capítulo apresentou uma extensão da abordagem de agrupamento de eventos relacionados a atividades afins na qual foi incorporada uma abordagem para filtragem de comportamento infrequente. Estudos comparativos (quantitativo e qualitativo) entre as duas abordagens foram conduzidos e os seguintes resultados foram observados:

- Ampliação significativa da capacidade de redução do log de eventos;
- Melhoria na qualidade dos agrupamentos através da redução de agrupamentos incompletos;
- Melhoria nas métricas de qualidade (*fitness*, precisão e simplicidade) dos modelos de processos.

Com isso, consideramos que a abordagem de agrupamento de eventos de atividades afins associada à filtragem de comportamento infrequente potencializou a abordagem original em todos os aspectos analisados.

Cabe ressaltar que nem toda fonte de complexidade é oriunda de comportamento infrequente. Observamos que o comportamento recorrente também é uma fonte de complexidade em modelos de processos e não pode ser tratado através de técnicas de filtragem de comportamento infrequente, como veremos no Capítulo 7.

7 ATIVIDADES RECORRENTES

No Capítulo 4 introduzimos a ideia de atividade recorrente, definida como atividade de propósito amplo que não se vincula a um contexto de negócio. Enquadramos as atividades recorrentes como um fator que afeta a qualidade dos modelos de processos descobertos através da mineração de processos. Sua presença em logs de eventos provoca o surgimento de *loops* que prejudicam a qualidade dos modelos de processos, bem como dificultam a sua interpretação.

Neste capítulo vamos abordar a problemática das atividades recorrentes, comparar o conceito de atividade recorrente com outros similares encontrados na literatura e apresentar uma abordagem para tratamento de atividades recorrentes visando a melhoria na qualidade dos modelos de processos descobertos. O capítulo está organizado nas seguintes seções: Seção 7.1, apresenta um comparativo entre as atividades recorrentes e o conceito de atividades caóticas proposto por Tax, Sidorova e van der Aalst (2018); Seção 7.2, apresenta um comparativo entre os eventos de atividades recorrentes com a ideia de eventos de atividades espúrias introduzida por van Zelst et al. (2018); Seção 7.3, aborda a identificação de contexto de negócios; Seção 7.4 mostra a abordagem proposta para tratamento de atividades recorrentes e a Seção 7.5 apresenta a conclusão do capítulo.

7.1 Atividades recorrentes e atividades caóticas

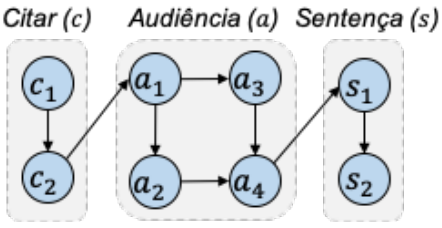
Em (TAX; SIDOROVA; VAN DER AALST, 2018) é apresentado o conceito de atividades caóticas que é semelhante ao de atividades recorrentes. Atividades caóticas são atividades que podem ocorrer em momentos arbitrários no tempo, ou seja, independem do estado do processo. Os conceitos têm em comum o fato destas atividades não se vincularem a uma fase do processo. Contudo, cabe esclarecer que as atividades caóticas, conforme definidas por TAX, SIDOROVA e VAN DER AALST (2018), podem ocorrer em momentos arbitrários, enquanto as atividades recorrentes possuem uma menor liberdade. Além disso, as atividades caóticas são vistas como um comportamento indesejado por comprometer a qualidade do modelo de processo. Já as atividades recorrentes, dizem respeito a comportamento relevante para o negócio, portanto, não devendo ser desprezadas.

Dois exemplos retirados do domínio de negócio judicial podem nos ajudar a distinguir atividades caóticas de atividades recorrentes: (1) petição e (2) comunicação. As *petições* consistem em demandas formuladas por atores externos à organização (partes e advogados) em um processo judicial em curso. Essas demandas podem ser formuladas a qualquer tempo e exigem uma ação por parte dos agentes internos da organização. A apreciação da petição ocorre em paralelo à tramitação do processo principal de forma que não afetam diretamente o seu fluxo do processo. Essas características indicam que as petições podem ser enquadradas como atividades caóticas, uma vez que podem ocorrem a qualquer tempo durante a tramitação de um processo judicial. Por outro lado, a *comunicação* consiste em uma ou mais atividades com a finalidade de promover a comunicação formal dos atos judiciais. Ou seja, consiste em um ato de comunicação formal dos atores internos da organização para os atores externos. Em geral, um processo contempla diversos atos judiciais, resultando na realização das atividades de comunicação várias vezes no curso do processo. Logo, trata-se de atividade recorrente, pois sua incidência se dá em diversas fases do processo. Contudo, mesmo ocorrendo em diversos momentos, a comunicação não ocorre a qualquer tempo.

A Figura 33 mostra um log de eventos $\mathcal{L}$ e um modelo de processos derivado deste. Podemos observar na Figura 33 que $\mathcal{L}$ é composto por duas variantes, cada uma contendo dez instâncias. Pode-se observar ainda que o modelo de processo não contempla atividades caóticas ou recorrentes, uma vez que não há repetição das atividades nas variantes apresentadas.

Figura 33 – Modelo de processo sem atividades caóticas ou recorrentes

$$\mathcal{L} = \{\langle c_1, c_2, a_1, a_2, a_4, s_1, s_2 \rangle^{10} \langle c_1, c_2, a_1, a_3, a_4, s_1, s_2 \rangle^{10}\}$$

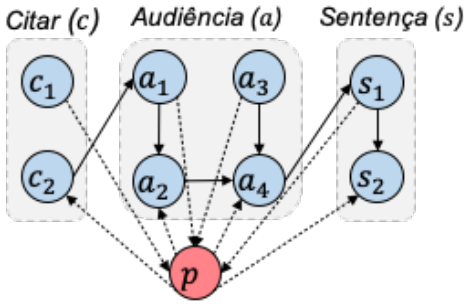


A Figura 34 mostra o log de eventos $\mathcal{L}$ com a inclusão da atividade caótica petição (p), resultando no log de eventos $\mathcal{L} + \{p\}$. A Figura 34 também mostra o modelo de processo decorrente do log de eventos $\mathcal{L} + \{p\}$. Podemos observar dois efeitos da inclusão da atividade caótica: (1) aumento de variantes do processo e (2) modelo de processo mais complexo em comparação à Figura 33. Além disso,

podemos observar que algumas relações de procedência ($c_1 \succ c_2$ e $a_1 \succ a_3$) foram ofuscadas pela presença da atividade caótica p, sendo o comportamento observado em $\mathcal{L}$ substituído em $\mathcal{L} + \{p\}$, respectivamente, por $c_1 \succ p \succ c_2$ e $a_1 \succ p \succ a_3$.

Figura 34 – Modelo de processo com atividade caótica petição (p)

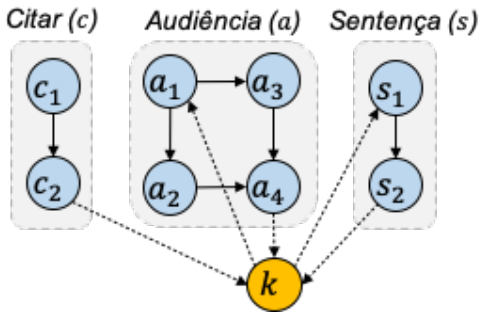
$$\mathcal{L} + \{p\} = \{\langle c_1, c_2, a_1, a_2, a_4, s_1, s_2 \rangle^2, \langle c_1, p, c_2, a_1, a_2, p, a_4, s_1, s_2 \rangle^5, \langle c_1, p, c_2, a_1, p, a_2, a_4, s_1, s_2 \rangle^3, \langle c_1, c_2, a_1, a_3, p, a_4, s_1, s_2 \rangle^2, \langle c_1, c_2, a_1, a_3, p, a_4, s_1, p, s_2 \rangle^2, \langle c_1, p, c_2, a_1, p, a_3, a_4, s_1, s_2 \rangle^4, \langle c_1, p, c_2, a_1, p, a_3, a_4, s_1, p, s_2 \rangle^2\}$$



A Figura 35 mostra o log de eventos $\mathcal{L}$ com a inclusão da atividade recorrente comunicação (k), resultando no log de eventos $\mathcal{L} + \{k\}$, bem como o modelo de processo decorrente. No exemplo apresentado podemos observar que também há um aumento das variantes no processo e da complexidade quando comparados com o log de eventos $\mathcal{L}$.

Figura 35 – Modelo de processo com atividade recorrente comunicação (k)

$$\mathcal{L} + \{k\} = \{\langle c_1, c_2, a_1, a_2, a_4, k, s_1, s_2, k \rangle^5, \langle c_1, c_2, k, a_1, a_3, a_4, s_1, s_2, k \rangle^5\} \langle c_1, c_2, k, a_1, a_3, a_4, k, s_1, s_2, k \rangle^{10}\}$$



O comportamento recorrente é bastante flexível, mas não pode ser confundido como um comportamento arbitrário. Pois, ainda que com bastante liberdade, existem pré-condições para sua realização. Contudo, temos que ter em mente que tomando por base apenas o log de eventos fica muito difícil distinguir o comportamento recorrente de um comportamento caótico. Os exemplos apresentados nas Figuras 34 e 35 mostram que o efeito da presença de ambos em um log de eventos é similar. Portanto, concluímos que a estratégia de identificação de atividades caóticas no log

de eventos pode ser aplicada para identificação de atividades recorrentes. Contudo, entendemos que o tratamento deve ser diferente para ambos os casos.

Um efeito importante provocado pelas atividades recorrentes e caóticas diz respeito ao impacto na qualidade do modelo de processos descoberto. Na Figura 34, podemos observar que o modelo de processo contempla todo o comportamento observado em $\mathcal{L}$. Além disso, não há comportamento previsto no modelo de processo que não seja observado em $\mathcal{L}$. Assim, o modelo de processo exemplificado na Figura 34 possui tanto *fitness* quanto precisão máxima. No exemplo da Figura 35 pode-se observar que alguns comportamentos registrados em $\mathcal{L} + \{k\}$ não foram contemplados no modelo do processo, por exemplo, as transições $c_1 \succ c_2$ e $a_1 \succ a_3$. Além disso, o modelo dá margem para comportamentos que não são observados no log de eventos no qual foi baseado. Por exemplo, os comportamentos $c_1 \succ p \succ a_2$, $c_1 \succ p \succ a_4$ e $c_1 \succ p \succ s_2$ são admitidos no modelo de processos, mas não existem no log de eventos. Logo, o modelo da Figura 35 possui *fitness* e precisão menor que o modelo da Figura 34. Na Figura 35 podemos observar que a inclusão da atividade recorrente k no log de eventos $\mathcal{L}$ também provocou impacto no *fitness* e precisão, uma vez que as transições $c_2 \succ a_1$ e $a_4 \succ s_1$ observadas em $\mathcal{L} + \{k\}$ não foram contempladas no modelo do processo correspondente. Além disso, os comportamentos $a_4 \succ k \succ a_1$ e $s_2 \succ k \succ a_1$, previstos no modelo de processo, não são observados em $\mathcal{L} + \{k\}$. Em relação à simplicidade dos modelos, é fácil perceber que o modelo apresentado na Figura 33 (sem atividades recorrentes ou caóticas) é mais simples que os modelos apresentados nas Figuras 34 (com atividades caóticas) e Figura 35 (com atividades recorrentes).

A despeito do prejuízo na qualidade dos modelos de processo, entendemos que as atividades recorrentes não devem ser descartadas dos logs de eventos por representar comportamento potencialmente relevante para o negócio. Portanto, propomos uma abordagem diferenciada para tratamento destas que será apresentada na Seção 7.3. Antes, analisaremos o conceito de eventos espúrios apresentado por van Zelst et al. (2018).

7.2 Eventos espúrios

O trabalho de van Zelst et al. (2018) aborda uma problemática no âmbito da mineração de processos *online*. A ampla maioria das técnicas de mineração de

processos atuam em modo *off-line*, ou seja, demandam um procedimento prévio de extração, tratamento e carga dos dados. Na mineração de processos *online* os eventos são capturados através de fluxos de dados (*streams*) e processados em tempo real. A mineração de processos *online* avança sobretudo com foco no suporte operacional, que consiste na capacidade de contribuir em tempo de execução do processo através da recomendação de comportamento para os usuários baseadas nos logs de eventos ou mesmo oferecendo previsões de acordo com o *status* da instância do processo.

Ao lidar com o tratamento de dados *online* de eventos, foi observado em (VAN ZELST et al., 2018) a incidência de eventos "fora de contexto". Ou seja, eventos no qual sua incidência é improvável em uma determinada fase do processo. Estes eventos foram denominados de espúrios e tratados como *outliers*, frutos de atrasos ou imperfeições nos registros dos eventos. A despeito de diversas diferenças entre os trabalhos de Tax, Sidorova e van der Aalst (2018) e van Zelst et al. (2018), encontramos afinidade entre os conceitos de atividades caóticas e eventos espúrios. Esta afinidade decorre do fato de que ambos podem ocorrer em qualquer momento do processo e são considerados pelos autores como comportamento ruidoso que deve ser eliminado. Contudo, os autores seguem estratégias distintas para a identificação e eliminação do comportamento indesejado.

Em (TAX; SIDOROVA; VAN DER AALST, 2018) temos uma abordagem estatística para filtragem das atividades caóticas baseada em medidas de entropia obtidas através das métricas $directly-follows\ ratio$ e $directly-precedes\ ratio$. As métricas $directly-follows\ ratio$ e $directly-precedes\ ratio$ foram abordadas no Capítulo 5 desta tese. Já em (VAN ZELST et al., 2018) temos uma abordagem baseada em autômatos probabilísticos, que representam o comportamento recente observado no fluxo de dados. Os autômatos são utilizados para determinar se um novo evento corresponde ao contexto do processo em um determinado momento.

A despeito de reconhecermos a afinidade entre os conceitos de atividades caóticas e eventos espúrios, além de considerar viável adotar a estratégia de identificação de atividades caóticas para identificação das atividades recorrentes, entendemos que a estratégia proposta para identificação de eventos espúrios não contribui para identificação de atividades recorrentes. Pois a abordagem proposta por van Zelst et al. (2018) foi concebida para atuar frente às restrições impostas pelo domínio da mineração de processos *online*, qual seja lidar com uma janela de dados

reduzida. Nessa tese propomos uma abordagem de mineração de processo *off-line*, de forma que não há necessidade da abordagem proposta se limitar a um conjunto de dados restrito. Portanto, no tocante a identificação de atividades recorrentes, consideramos uma adaptação da abordagem proposta por Tax, Sidorova e van der Aalst (2018) mais adequada.

Por outro lado, entendemos que a estratégia para identificação de contextos de negócios proposta por van Zelst et al. (2018) pode contribuir para o tratamento das atividades recorrentes. A seção seguinte apresenta uma visão geral sobre nossa abordagem para tratamento de atividades recorrentes.

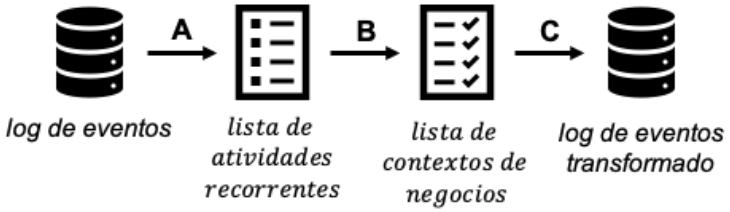
7.3 Desmembramento de atividades recorrentes

Como vimos na Seção 7.1 desta tese, a presença de atividades recorrentes no log de eventos prejudica a qualidade dos modelos descobertos. Por outro lado, a filtragem de atividades pode excluir comportamento relevante do modelo, comprometendo sua capacidade de representar a realidade de forma consistente. Propomos uma nova abordagem que consiste em desmembrar as atividades recorrentes de acordo com os contextos aos quais estão inseridas para que possamos obter modelos mais fáceis de interpretar e com maior qualidade.

A ideia geral consiste em identificar as atividades recorrentes e em seguida identificar contextos de negócios relevantes relacionados a estas. Então, os rótulos dos eventos inseridos em contextos de negócios relevantes de atividades recorrentes recebem a indicação do contexto de negócio ao qual fazem parte. Na prática, a inclusão do nome do contexto de negócio no rótulo da atividade resulta em um desmembramento das atividades representadas por estes eventos. A Figura 36 apresenta uma visão geral da abordagem proposta, que consiste em três etapas: (A) criação de uma $lista\ de\ atividades\ recorrentes$, (B) criação de uma lista de $contextos\ de\ negócios$ (relevantes) e (C) transformação do log de eventos para desmembramento contextual das atividades recorrentes. A etapa (A) preconiza uma abordagem para identificação de atividades recorrentes que será apresentada na Seção 7.3.1. A etapa (B) prevê uma abordagem para identificação de contextos de negócios relevantes relacionados a atividades recorrentes, que será mostrada na Seção 7.3.2. Por fim, na Seção 7.3.3 apresentamos a etapa (C), que consiste na

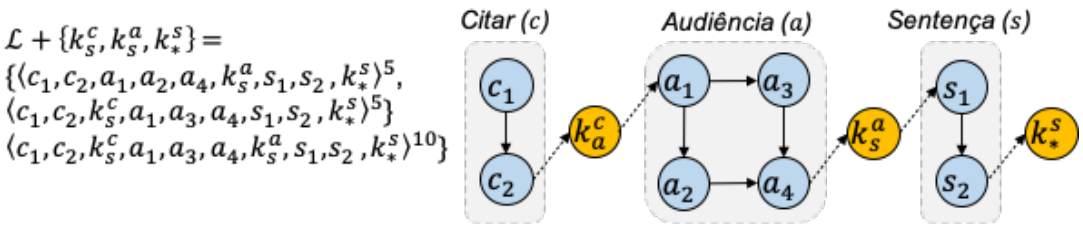
abordagem para transformação do log de eventos através do desmembramento das atividades recorrentes.

Figura 36 – Abordagem para desmembramento de atividades recorrentes



A Figura 37 ilustra o efeito esperado da abordagem proposta neste capítulo quando aplicada ao log de eventos $\mathcal{L} + \{k\}$ (Figura 35) e correspondente modelo de processos. Observando o modelo de processos da Figura 35 podemos perceber que a atividade recorrente k se relaciona com três contextos de negócios distintos: citar (c), audiência (a) e sentença (s). Aplicando a ideia de desmembramento contextual, podemos desmembrar a atividade recorrente k em três atividades correspondentes aos contextos identificados, resultando nas atividades k_a^c, k_s^a, k_*^s, sendo $k_s^c \cup k_s^a \cup k_*^s = k$.

Figura 37 – Modelo de processo com atividade desmembrada



Na Figura 37 ilustramos o efeito esperado no log de eventos ($\mathcal{L} + \{k_a^c, k_s^a, k_*^s\}$) e modelo de processo correspondente após o desmembramento da atividade k. No modelo de processo mostrado na Figura 37 podemos observar que a atividade k foi desmembrada nas seguintes: k_a^c, que corresponde à atividade k quando precedida pelo contexto de negócio citar (c) e sucedida pelo contexto de negócio audiência (a); k_s^a, que corresponde à atividade k quando precedida pelo contexto de negócio audiência (a) e sucedida pelo contexto de negócio sentença (s); e k_*^s, que corresponde à atividade k quando precedida pelo contexto de negócio sentença (s) e sucedida por outro contexto de negócio qualquer.

O primeiro aspecto que podemos vislumbrar com o desmembramento da atividade recorrente k é a redução da complexidade no modelo de processo. Mesmo com o aumento de atividades (9 atividades em $\mathcal{L} + \{k\}$ contra 11 atividades em $\mathcal{L} + \{k_a^c, k_s^a, k_*^s\}$) o modelo de processo mostrado na Figura 37 é mais fácil de interpretar que o modelo de processo da Figura 35. Além disso, podemos vislumbrar um incremento da qualidade do modelo de processo através do desmembramento das atividades, uma vez que há uma maior correspondência entre o modelo e seu log de eventos, impactando assim nas métricas de qualidade *fitness* e precisão para estes.

Nas próximas seções apresentamos a abordagem para identificação de atividades recorrentes, dos seus contextos de negócios e a transformação do log de eventos.

7.3.1 Identificação de atividades recorrentes

Na Seção 7.1 vimos que as atividades caóticas, conforme definidas em (TAX; SIDOROVA; VAN DER AALST, 2018), apresentam comportamento similar ao das atividades recorrentes. Tax, Sidorova e van der Aalst (2018) apresentam uma medida de entropia baseada na função para distribuição de probabilidade categórica em termos das métricas $directly-follows\ ratio\ (dfr)$ e $directly-precedes\ ratio\ (dpr)$ (Definição 21). A referida medida de entropia é definida como segue:

Definição 26 $(entropia)$ (TAX; SIDOROVA; VAN DER AALST, 2018). Seja $\mathcal{D}_A$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Dadas a atividade $a \in \mathcal{D}_A$, seja $dfr(a)$ o vetor contendo os valores de $dfr(a,z) \mid \forall z \in \mathcal{D}_A$ e $dpr(a)$ o vetor contendo os valores de $dpr(a,z) \mid \forall z \in \mathcal{D}_A$. Seja a entropia de a, denotada por $H(a)$, definida como $H(a) = H(dfr(a)) + H(dpr(a))$, sendo $H(dfr(a))$ e $H(dpr(a))$ assim definidos:

$$H\big(dfr(a)\big) = -\sum_{x \in X} dfr(a) \log_2(dfr(a))$$

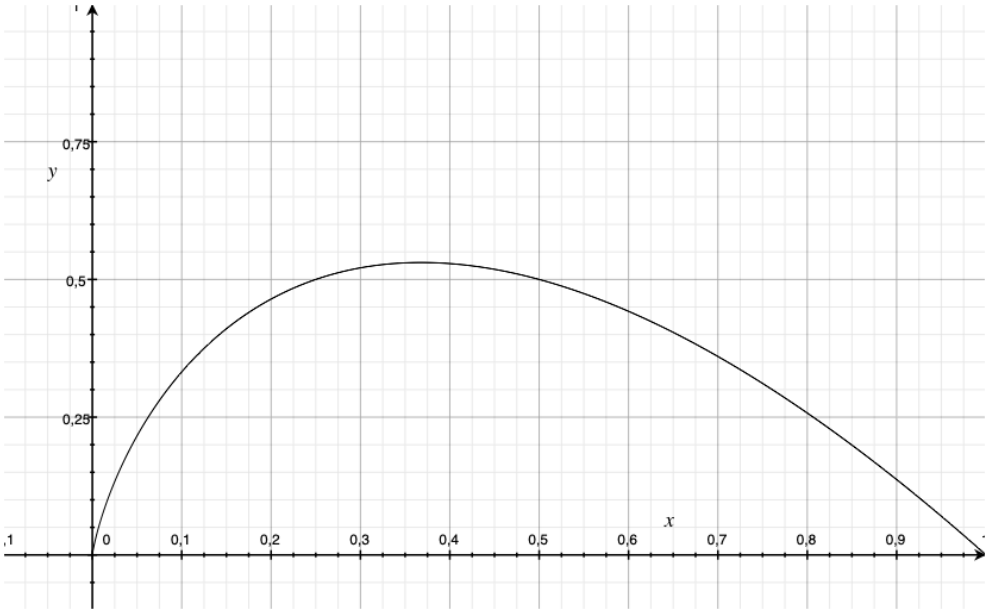
$$H\big(dpr(a)\big) = -\sum_{x \in X} dpr(a) \log_2(dpr(a))$$

Por exemplo, seja o log de eventos $\mathcal{L} = [\langle a,b,c\rangle^{10}, \langle a,c,b\rangle^{4}, \langle a,b,a\rangle^{2}]$, temos que $dfr(a) = \langle \frac{|a>a|}{|a|}, \frac{|a>b|}{|a|}, \frac{|a>c|}{|a|}, \frac{|a>\emptyset|}{|a|} \rangle$ e $dpr(a) = \langle \frac{|a>a|}{|a|}, \frac{|b>a|}{|a|}, \frac{|c>a|}{|a|}, \frac{|\emptyset>a|}{|a|} \rangle$, logo $dfr(a) =$

$\langle 0, \frac{12}{18}, \frac{4}{18}, \frac{2}{18} \rangle$ e $dpr(a) = \langle 0, \frac{2}{18}, 0, \frac{16}{18} \rangle$. Então, $H\big(dfr(a)\big) = 0 - \frac{12}{18}\log_2\left(\frac{12}{18}\right) - \frac{4}{18}\log_2\left(\frac{4}{18}\right) - \frac{2}{18}\log_2\left(\frac{2}{18}\right)$ e $H\big(dpr(a)\big) = 0 - \frac{2}{18}\log_2\left(\frac{2}{18}\right) - 0 - \frac{16}{18}\log_2\left(\frac{16}{18}\right)$. Por fim, temos que $H(a) = 1{,}22 + 0{,}50 \cong 1{,}72$.

A Figura 38 mostra um gráfico com a distribuição dos valores da função de entropia $f(x) = -x\log_2 x \mid x \in \mathbb{R}^+ [0 \leq x \leq 1]$, sendo $x = dfr(a)$ ou $x = dpr(a)$, adotada em (TAX; SIDOROVA; VAN DER AALST, 2018).

Figura 38 – Função de entropia (TAX; SIDOROVA; VAN DER AALST, 2018)



A partir do gráfico mostrado na Figura 38, podemos observar que os valores extremos de x ($dfr(a)$ ou $dpr(a)$) resultam em entropia mais baixa que valores intermediários. O máximo da função é atingido quando $x \cong 0{,}37$. A consequência disso é que, caso haja uma forte relação de precedência entre duas atividades a e b, tal que $dfr(a,b) \gg dfr(a,w) | w \neq b, \forall w \in \mathcal{D}_A$, então teremos o vetor $dfr(a)$ composto por valores extremos, resultando em uma entropia baixa. Por outro lado, se a atividade a não tiver forte relação de precedência (ou sucessão) com outra, o vetor $dfr(a)$ apresentará valores intermediários, resultando em uma alta entropia para a. Por exemplo, considerando uma situação hipotética em que uma atividade a, tal que $dfr(a) = \langle 0, \frac{8}{10}, \frac{2}{10}, 0 \rangle$; e uma atividade b, tal que $dfr(b) = \langle 0, \frac{3}{10}, \frac{4}{10}, 0 \rangle$. Nesse caso, temos que $H\big(dfr(a)\big) \cong 0{,}92$ e $H\big(dfr(b)\big) \cong 1{,}57$. Como pode ser observado, a atividade a apresentou uma entropia mais baixa que a atividade b. Isso se deve ao

fato da primeira ter uma forte relação de precedência com outra atividade enquanto a segunda não.

Partindo da premissa que as atividades recorrentes não firmam relações fortes com outras atividades, consideramos a medida de entropia apresentada em (TAX; SIDOROVA; VAN DER AALST, 2018) como adequada para identificação de atividades recorrentes. Além disso, a medida de entropia analisada se baseia em duas métricas similares adotadas em nossa abordagem para agrupamento de atividades afins e filtragem de comportamento infrequente, tornando-a compatível com as demais abordagens propostas. Assim, propomos uma abordagem para identificação de atividades recorrentes baseada na métrica proposta em (TAX; SIDOROVA; VAN DER AALST, 2018).

Nossa abordagem para identificação das atividades recorrentes consiste na métrica já apresentada na Definição 26 aliada a um parâmetro $\rho \in \mathbb{R} \mid 0 < \rho \leq 1$ definido pelo usuário. O parâmetro ρ tem a função de estabelecer um *thresholds* (φ_H) de entropia. As atividades com entropia maior que φ_H serão consideradas recorrentes. A seguir apresentamos a definição formal de φ_H.

Definição 27 (φ_H). Seja $\mathcal{D}_A = \{a_1, \dots, a_n\}$ o conjunto de todas as atividades do log de eventos $\mathcal{L}$. Seja H o vetor com o valor da entropia de todas as atividades $\mathcal{D}_A$, tal que $H_{\mathcal{D}_A} = \{H(a_1), \dots, H(a_n)\}$. Seja $H(a_n)$ a entropia da atividade a_n conforme a Definição 26. Seja $\max\left(H_{\mathcal{D}_A}\right)$ e $\min\left(H_{\mathcal{D}_A}\right)$, respectivamente, o valor mais alto e mais baixo de entropia em $H_{\mathcal{D}_A}$. O *thresholds* φ_H é definido por:

$$\varphi_H = \rho \max\left(H_{\mathcal{D}_A}\right) + (1 - \rho) \min\left(H_{\mathcal{D}_A}\right)$$

Por exemplo, dado um log de evento $\mathcal{L}$ com $\mathcal{D}_A = \{a, b, c, d\}$ sendo $H(a) = 1{,}75$, $H(b) = 0{,}6$, $H(c) = 2{,}4$ e $H(d) = 0{,}8$. Então, $H_{\mathcal{D}_A} = \{1{,}75;\ 0{,}60; 2{,}40;\ 0{,}85\}$ e $\max\left(H_{\mathcal{D}_A}\right) = 2{,}40$ e $\min\left(H_{\mathcal{D}_A}\right) = 0{,}60$. Dado o parâmetro $\rho = 0{,}8$, temos que $\varphi_H = 0{,}8 * 2{,}4 + (1 - 0{,}8) * 0{,}6$, logo $\varphi_H = 2{,}04$. Assim, avaliando as atividades de $\mathcal{D}_A$, temos que a atividade c é considerada recorrente, pois $H(c) > \varphi_H$. As demais atividades possuem entropia menor que φ_H, então não são consideradas recorrentes.

O procedimento de geração da $lista\ de\ atividades\ recorrentes$ é apresentado no Algoritmo 7.

Algoritmo 7
Geração da $lista\ de\ atividades\ recorrentes$

```
    entrada: LogDeEventos ℒ, Parâmetro ρ
    saída: ListaDeAtividadesRecorrentes ϖ
1   α ⇐ ListaDeAtividades(ℒ)
2   para todo a_i ∈ α faça
3       H_α[a_i] ⇐ H(a_i) | seja a_i atividade de ℒ, H(a_i) a entropia de a_i e H_α o vetor entropia de α
4   φ_H ⇐ ρ * max(H_α) + (1 − ρ) * min(H_α)
5   para todo a_i ∈ α faça
6     se H_α[a_i] > φ_H então
7         ϖ ⇐ append(a_i)
8   retorna ϖ
```

7.3.2 Identificação de contextos de negócios relevantes

A etapa seguinte da nossa abordagem de desmembramento de atividades recorrentes consiste na identificação dos contextos de negócios relevantes aos quais as atividades podem estar vinculadas. Em (FANI SANI; VAN ZELST; VAN DER AALST, 2018a) foi apresentada uma abordagem que explora a ideia de comportamento contextual. Os autores caracterizaram o comportamento contextual como sequências de atividades que ocorrem antes e depois de um comportamento observado. O trabalho apresentou uma métrica denominada de $frequência\ relativa\ contextual$ que é obtida pela média de ocorrências do contexto no log de eventos. A $frequência\ relativa\ contextual$, conforme definida em (FANI SANI; VAN ZELST; VAN DER AALST, 2018a), não é capaz de indicar a relevância do contexto em relação a uma atividade, uma vez que a medida leva em conta a frequência em relação a todo o log de eventos. Até onde sabemos, na literatura não há outras métricas que se alinhem ao nosso propósito. Portanto, propusemos um indicador de relevância contextual relativo às atividades recorrentes, formalmente definido como segue.

Definição 28 ($indicador\ de\ relevância\ contextual$). Dado um log de eventos $\mathcal{L}$, seja $\mathcal{D}_a$ o domínio de atividades em $\mathcal{L}$. Seja $\varphi_c \in \mathbb{R}[0..1]$ *thresholds* de relevância contextual. Dadas as atividades $a_{i-1}, a, a_{i+1} \in \mathcal{D}_a \mid a_{i-1} \succ a \succ a_{i+1}$, seja a_{i-1}, a_{i+1} contexto de a. O indicador de relevância de a_{i-1}, a_{i+1} em relação a atividade a, denotado por $\mathcal{C}_{a_{i-1}}^{a_{i+1}}(a)$, é definido por:

$$\mathcal{C}_{a_{i-1}}^{a_{i+1}}(a) = \begin{cases} 1, & \frac{|a_{i-1} \succ a \succ a_{i+1}|}{|a|} \geq \varphi_c \\ 0, & \frac{|a_{i-1} \succ a \succ a_{i+1}|}{|a|} < \varphi_c \end{cases}$$

Por exemplo, dado o log de eventos $\mathcal{L} = [\langle a, b, c\rangle^{10}, \langle a, c, b\rangle^{4}, \langle a, b, a\rangle^{2}]$ e $\varphi_c = 0{,}2$, temos que $|b| = 16$ e $|a \succ b \succ c| = 10$. Logo, $\frac{|a_{i-1} \succ a \succ a_{i+1}|}{|a|} = 0{,}625 > \varphi_c$, então, temos que $\mathcal{C}_a^c(b) = 1$.

O $indicador\ de\ relevância\ contextual$ conforme apresentado é útil para elaboração da $lista\ de\ contextos\ recorrentes$ conforme apresentado na Figura 35. A $lista\ de\ contextos\ recorrentes$ é composta dos contextos de negócios relevantes relativos à atividade recorrentes. O procedimento para geração da $lista\ de\ contextos\ de\ negócios$ é apresentado no Algoritmo 8.

Algoritmo 8
Geração da $lista\ de\ contextos\ recorrentes$

```
    entrada: LogDeEventos ℒ, ListaDeAtividadesRecorrentes ϖ
    saída: ListaDeContextosDeNegocios γ
1   para todo e_i ∈ ℒ | seja e_i evento de ℒ faça
2     se #_c(e_{i-1}) = #_c(e_i) e #_c(e_i) = #_c(e_{i+1}) | seja #_c o atributo caso do evento então
3       se #_a(e_i) existe em ϖ | seja #_a o atributo atividade do evento então
4         γ* ⇐ append(#_a(e_i), (#_a(e_{i-1}), #_a(e_{i+1})))
5   para todo a_i ∈ ϖ faça
6     se C^{a_{i+1}}_{a_{i-1}}(a_i) então
7       i = i + 1
8       γ ⇐ i, γ*
9   retorna γ
```

7.3.3 Desmembramento de atividades no log de eventos

Conforme apresentado no Figura 36, a $lista\ de\ contextos\ recorrentes$ é o *input* para o desmembramento das atividades recorrentes. O procedimento de transformação do log de eventos através do desmembramento de atividades recorrentes, consiste em percorrer todo o log de eventos buscando contextos de negócios presente na lista de $lista\ de\ contextos\ recorrentes$. Caso seja identificado um contexto recorrente, será adicionado ao rótulo original da atividade um identificador (i). O procedimento de geração de novo log de eventos a partir do desmembramento de atividades recorrentes é mostrado no Algoritmo 9.

Algoritmo 9
Transformação de log de eventos a partir do desmembramento de atividades

entrada: $LogDeEventos\ \mathcal{L}, ListaDeContextosDeNegocios\ \gamma$
saída: $LogDeEventos\ \mathcal{L}^*$ | seja $\mathcal{L}^*$ o log de eventos transformado
1 **para todo** $e_i \in \mathcal{L}$ | seja e_i evento de $\mathcal{L}$ **faça**
2 **se** $\#_c(e_{i-1}) = \#_c(e_i)$ **e** $\#_c(e_i) = \#_c(e_{i+1})$ | e $\#_c$ o atributo caso do evento **então**
3 **se** $\left(\#_a(e_i), \left(\#_a(e_{i-1}), \#_a(e_{i+1})\right)\right)$ **existe em** γ | seja $\#_a$ o atributo nome da atividade **então**
4 $ativ \Leftarrow \#_a(e_i) + i$
5 $\mathcal{L}^* \Leftarrow \left(\#_c(e_i), ativ, \#_i(e_i), \#_f(e_i)\right)$ | seja $\#_i, \#_f$ atributos de data/hora inicial e final
6 **senão**
7 $\mathcal{L}^* \Leftarrow (e_i)$;
8 **retorna** $\mathcal{L}^*$

7.4 Avaliação

Aplicamos o desmembramento de atividades recorrentes nos três logs de eventos resultantes da abordagem de agrupamento de atividades afins com filtragem de comportamento infrequente. A abordagem proposta neste capítulo foi aplicada sobre os logs de eventos $\mathcal{L}_F^1, \mathcal{L}_F^2, \mathcal{L}_F^2$ considerando o *threshold* $\varphi_c = 0{,}1$. A Tabela 22 mostra resultados observados.

Tabela 22 – Impactos observados no novo log de eventos

CATEGORIA OBSERVADA	LOGS DE EVENTOS		
	$\mathcal{L}_F^1$	$\mathcal{L}_F^2$	$\mathcal{L}_F^3$
(A) Percentual de atividades desmembradas	10,8%	10,9%	11,1%
(B) Percentual de eventos impactados	12,2%	2,4%	6%

Na primeira linha da Tabela 22 podemos observar que o percentual de atividades consideradas recorrentes foi bastante similar para os três logs de eventos analisados. Contudo, em relação aos eventos impactados observamos uma variação significativa (linha 2 da Tabela 22) nos resultados. No log de eventos 1 tivemos a maior proporção de eventos alterados (mais de 12%), indicando que no processo de negócio representado por este log de eventos os contextos recorrentes são mais relevantes que nos processos de negócios representados pelos demais logs de eventos. No log de eventos 2 observamos que apenas 2,4% dos eventos foram impactados. Esse resultado indica que as atividades desmembradas terão baixa frequência absoluta. Com isso, o desmembramento de atividades favorece indiretamente a eliminação de

atividades infrequentes nesses logs, uma vez que, em geral, os algoritmos de mineração de processos realizam uma filtragem de comportamento infrequente. No log de eventos 3 temos uma situação intermediária, na qual parte das atividades desmembradas são relevantes e parte estará exposta a ação dos filtros de atividades infrequentes.

Assim como realizado com as abordagens apresentadas nos Capítulos 5 e 6, realizamos experimentos para avaliar a análise da qualidade dos modelos gerados a partir da abordagem apresentada neste capítulo (abordagem 3). Cabe esclarecer que apresentamos comparativo apenas com a abordagem 2, pois, os logs de eventos utilizados nos experimentos são oriundos desta. Além disso, a abordagem 2 é uma extensão da abordagem 1, sendo infrutífero um estudo comparativo entre as abordagens 1 e 3.

A Tabela 23 mostra o comparativo entre os modelos gerados a partir dos logs de eventos resultantes da abordagem 3 ($\mathcal{L}_D^1, \mathcal{L}_D^2, \mathcal{L}_D^3$) com os da abordagem 2 ($\mathcal{L}_F^1, \mathcal{L}_F^2, \mathcal{L}_F^3$). Os experimentos seguiram o mesmo padrão dos realizados nas Seções 5.6.1 e 6.2.1, nos quais cada log de eventos foi submetido a descoberta de processos com o *Heuristic Miner* (PM4Py) com 64 configurações distintas. Os 64 modelos resultantes de cada log de eventos foram avaliados a partir das medidas de qualidade: MF1 *(perc_fit_traces)*, MF2 *(average_trace_fitness)*, MF3 *(log_fitness)* e MP *(precision)* através de abordagens baseadas na reprodução de tokens (BERTI; VAN DER AALST, 2019; MUÑOZ-GAMA; CARMONA, 2010).

Na primeira linha da Tabela 23, podemos observar que a abordagem 3 contribuiu significativamente para a melhoria dos modelos de processos no tocante a métrica MF1 *(perc_fit_traces)*. Nas métricas MF2 *(average_trace_fitness)* e MF3 *(log_fitness)*, respectivamente linhas 2 e 3 da Tabela 23, observamos uma melhoria discreta nas medidas globais de *fitness*. Em contrapartida, observamos uma redução discreta na precisão dos modelos do log de eventos 3 e moderada no log de eventos 2. Cruzando os resultados mostrados nas Tabelas 22 e 23 podemos perceber uma correlação negativa entre o impacto nos eventos e o impacto na precisão, sugerindo que o desmembramento de atividades recorrentes com baixa frequência absoluta impacta negativamente na precisão. Entendemos que essa questão pode ser contornada com a escolha de valores adequados para φ_c.

Tabela 23 – Comparativo entre métricas de qualidade dos modelos de processos descobertos para os logs de eventos ($\mathcal{L}_F^1, \mathcal{L}_D^1, \mathcal{L}_F^2, \mathcal{L}_D^2, \mathcal{L}_F^3, \mathcal{L}_D^3$)

MÉTRICA	$\mathcal{L}_F^1$	$\mathcal{L}_D^1$	$\mathcal{L}_F^2$	$\mathcal{L}_D^2$	$\mathcal{L}_F^3$	$\mathcal{L}_D^3$
MF1	0,01	0,40	0,44	0,54	0,77	0,77
MF2	0,96	0,97	0,95	0,98	0,99	0,92
MF3	0,96	0,96	0,95	0,98	0,99	0,98
MP	0,66	0,66	0,85	0,80	0,87	0,85

Os experimentos foram realizados em um Dual-Core Intel Core i5 (1,8 GHz) com 8GB de RAM. O tempo aproximado de processamento das métricas MF1, MF2, MF3 e MP para os 64 modelos de cada log de eventos foi de: 1 hora e 15 minutos para $\mathcal{L}_D^1$, 20 minutos para $\mathcal{L}_D^2$, 25 minutos para $\mathcal{L}_D^3$, totalizando 2 horas de processamento. Os resultados detalhados se encontram nos apêndices desta tese.

Em relação à simplicidade, realizamos análise semelhante à conduzida nas Seções 5.6.1 e 6.2.1, na qual os modelos de referência (configuração padrão do *Heuristic Miner* no PM4Py) gerados pela abordagem apresentada neste capítulo foram comparados com os modelos gerados pela abordagem estendida de agrupamento de atividades afins apresentado no Capítulo 6. Para tanto, foram usadas as métricas de simplicidade (*Extended Cardoso metric* - ECaM, *Extended Cyclomatic metric* - ECyM e *Structuredness metric*) apresentadas em (LASSEN; VAN DER AALST, 2009).

A Tabela 24 apresenta os valores obtidos para os logs de eventos ($\mathcal{L}_F^1$, $\mathcal{L}_D^1$, $\mathcal{L}_F^2$, $\mathcal{L}_D^2$, $\mathcal{L}_F^3$ e $\mathcal{L}_D^3$) através do *plug-in Show Petri-net Metrics* do ProM 6.9. Novamente não foi possível obter resultados para a métrica ECyM em relação aos modelos de processos oriundos do log de eventos 1 e para a métrica *Structuredness* em relação aos modelos de processos derivados do log de eventos 3 (identificados com “N/A”).

Em relação à métrica ECaM, primeira linha da Tabela 24, observamos que houve um aumento da complexidade para os modelos derivados do log de eventos 1 e uma redução para os demais modelos de processo. Em relação à métrica ECyM, na segunda linha da Tabela 24, observamos uma leve redução para os modelos de processos derivados dos logs de eventos 2 e 3. Contudo, na direção contrária a métrica *Structuredness* apresentou um crescimento para os modelos de processos derivados dos logs de eventos 1 e 2.

Tabela 24 – Comparativo entre métricas de simplicidade dos modelos de processos descobertos para os logs de eventos ($\mathcal{L}_F^1, \mathcal{L}_D^1, \mathcal{L}_F^2, \mathcal{L}_D^2, \mathcal{L}_F^3, \mathcal{L}_D^3$)

MÉTRICA	$\mathcal{L}_F^1$	$\mathcal{L}_D^1$	$\mathcal{L}_F^2$	$\mathcal{L}_D^2$	$\mathcal{L}_F^3$	$\mathcal{L}_D^3$
ECaM	128	149	128	126	185	177
ECyM	N/A	N/A	64	62	108	100
Structuredness	62500	166440	512	756	N/A	N/A

A análise dos resultados observados pela métricas de simplicidade nos indica que a abordagem possibilita a simplificação indireta do modelo de processo em relação ao comportamento permitido por estes, gerando assim valores mais baixos para a métrica ECyM. Além disso, a leve melhoria das métricas ECaM observada em dois dos três modelos observados sugere que a abordagem também contribui com a simplificação sintática do modelo, embora para tanto depende das características inerentes ao log de eventos. Por outro lado, o desmembramento de atividades adiciona um novo comportamento ao modelo de processo, algo que foi capturado pela métrica *Structuredness*.

7.5 Conclusão do capítulo

Neste capítulo apresentamos uma abordagem para desmembramento de atividades recorrentes através da transformação nos logs de eventos. Realizamos análise qualitativa dos impactos gerados pela abordagem tanto nos logs de eventos quanto nos modelos de processos decorrentes destes.

Os experimentos realizados mostraram que apesar da quantidade de atividades consideradas recorrentes nos três logs de eventos ser similar, a quantidade de eventos impactados variou significativamente nos logs de eventos analisados, evidenciando a diferença entre os processos analisados. Cabe ressaltar que a abordagem de desmembramento não influencia diretamente no tamanho do log de eventos, uma vez que não há criação ou eliminação de eventos, apenas o rótulo de eventos é alterado. O impacto da abordagem se manifesta nos modelos de processos descobertos a partir do log de eventos transformado pela abordagem proposta.

A análise comparativa dos modelos de processos descobertos a partir dos logs de eventos transformados pela abordagem de desmembramento de atividades recorrentes quando comparado a abordagem apresentada no Capítulo 6 nos mostrou que houve um incremento substancial do *fitness* associado a uma leve redução da precisão. Em relação à simplicidade também observamos um *trade-off*, uma vez que houve redução nos valores para as métricas ECaM[17] e ECyM e aumento para a métrica *Structuredness*.

Os resultados obtidos mostram que a abordagem proposta neste capítulo tem a capacidade de melhorar os modelos de processos sob alguns aspectos, não sem prejuízo de outros. Entendemos que os modelos de processos transformados por esta abordagem produzem modelos mais fáceis de serem interpretados, mas existe uma forte subjetividade neste aspecto. Assim, consideramos adequado que a abordagem proposta seja adotada como ferramenta de recomendação de transformação do log de eventos ficando a cargo do usuário decidir qual o melhor modelo para seus propósitos.

[17] Dos três modelos analisados houve redução em dois e incremento em um modelo

8 CONCLUSÃO

Os sistemas de fluxos oferecem às organizações flexibilidade para “moldar” a ferramenta para suas especificidades e com isso proporcionar uma gestão e execução mais eficiente dos seus processos de negócios. Para tanto, requer que o processo seja adequadamente modelado, bem como atualizado periodicamente. Por outro lado, a modelagem de processos pode ser desafiadora e custosa, uma vez que exige o mapeamento abrangente das situações vivenciadas no cotidiano das organizações. Além disso, é inevitável que o processo modelado reflita a visão das pessoas que o modelaram. Muitas vezes a visão dos envolvidos na modelagem do processo não corresponde à visão de outros atores do processo. Portanto, faz-se necessário analisar periodicamente como o processo tem sido executado para validar as premissas adotadas na modelagem, bem como identificar oportunidades de melhorias. A mineração de processos surge justamente para contribuir nesse contexto, servindo de ponte entre as ciências de processos e as ciências de dados (BERTI et al., 2019).

A literatura delimita bem os avanços e limitações da mineração de processos. O estudo preliminar apresentado nessa tese corrobora a existência de limitações importantes em ferramentas de mineração de processos quando lidam com processos complexos. A alternativa usual frente às limitações das ferramentas é o pré-processamento dos logs de eventos. Contudo, essa atividade tende a ser onerosa, pois exige esforço e conhecimento especializado, uma vez que impactam diretamente na qualidade dos modelos de processos descobertos.

Neste trabalho apresentamos três abordagens para pré-processamento de logs de eventos. As abordagens propostas nesta tese buscaram oferecer alternativas para transformação automática dos logs de eventos, que contribuam para os seguintes objetivos: 1) redução do log de eventos; 2) melhoria da qualidade dos modelos de processos derivados dos logs de eventos transformados; e 3) oferecimento de uma visão alternativa para o processo.

Para atingir os objetivos desejados recorremos à literatura para identificação de métricas existentes, as quais foram devidamente avaliadas posteriormente. Novas métricas e indicadores foram propostos para os casos em que não foram encontradas alternativas na literatura. Também foram propostos algoritmos para transformação dos

logs de eventos desenhados para realizar as transformações pretendidas sem prejudicar o comportamento original.

Os logs de eventos transformados, bem como os modelos de processos descobertos através destes, foram avaliados através de experimentos conduzidos com log de eventos extraídos de ambientes reais. Os experimentos foram conduzidos usando as ferramentas relevantes nas esferas comerciais e acadêmicas. Além disso, quando necessário, especialistas foram consultados para avaliar os resultados observados.

Os resultados mostraram que a abordagem de agregação de eventos de atividades afins, proposta no Capítulo 5, possibilita uma visão alternativa para o processo com um log de eventos mais enxuto, sem comprometer a qualidade dos modelos de processos derivados destes. A análise dos resultados com esta abordagem indicou a oportunidade de melhoria da qualidade dos modelos derivados através da incorporação de uma etapa de filtragem de comportamento infrequente. Diante disso, a questão do comportamento infrequente residual foi abordada no Capítulo 6, bem como uma nova abordagem foi proposta. Os experimentos conduzidos mostraram que a abordagem estendida de agrupamento de eventos de atividades afins apresentou melhoria nas diversas métricas de qualidade analisadas.

Com o estudo dos modelos de processos foi possível observar a incidência de comportamento recorrente e o seu impacto negativo na compreensão dos modelos de processos. Diante disso, propusemos uma abordagem para desmembramento das atividades recorrentes. Novos experimentos foram conduzidos e destes concluímos que a abordagem proporciona a simplificação dos modelos de processos. Contudo, pelo fato de o desmembramento de atividades resultar no incremento de elementos no modelo do processo, algumas métricas capturaram esse fato como aumento de complexidade apresentando valores mais altos. Ainda assim, nossa avaliação é que os modelos de processos derivados do desmembramento são de mais fácil compreensão que os que preservam comportamento recorrente. Mas, por haver uma carga de subjetividade nessa interpretação, entendemos que a abordagem deve ser adotada como uma ferramenta de recomendação, ficando a cargo do usuário definir a forma que melhor lhe convier.

8.1 Limitações e ameaças à validade

Apesar dos estudos conduzidos indicarem a eficácia das abordagens propostas, é importante destacar as limitações dessa pesquisa para que possamos identificar pontos de melhorias para trabalhos futuros nesta área. Em geral, podemos identificar as seguintes ameaças à validade:

- Os resultados descritos nesta tese foram extraídos de apenas três logs de eventos. A baixa quantidade de logs de eventos envolvida nos estudos é explicada pela dificuldade em se adquirir dados reais, bem como a disponibilidade de especialistas nos negócios. Contudo, amparados pelos resultados observados e pela experiência adquirida no domínio de negócio, estamos seguros de que os resultados observados são consistentes;
- Embora os estudos tenham sido conduzidos com logs de eventos de duas instituições distintas, as abordagens propostas foram aplicadas e avaliadas em um único domínio (poder judiciário). A escolha do domínio de aplicação se deu pela experiência dos pesquisadores na área, permitindo vislumbrar um forte alinhamento com os objetivos da mineração de processos. Entendemos que a utilização de logs de eventos do mesmo domínio confere mais robustez às comparações, bem como reforça a validade dos resultados. Contudo, é provável que outros problemas sejam identificados quando aplicados em outros domínios. Portanto, a adoção em outros domínios pode apresentar resultados diferentes dos obtidos nesta tese.
- A análise dos resultados de ferramentas de mineração de processos exige conhecimento multidisciplinar (conhecimento de negócio, conhecimento de gestão de processos e conhecimento de mineração de processos). No domínio de aplicação da pesquisa há escassez de especialistas que dominem todos os conhecimentos necessários, resultando em um conhecimento compartimentado em profissionais de diferentes visões. Para minimizar o impacto dessa questão, contamos com dois perfis de especialistas (magistrado e analista de sistema) com visões distintas sobre o problema abordado.

Em relação às limitações das abordagens propostas nesta tese podemos elencar os seguintes pontos:

- O agrupamento de atividades afins se baseia, dentre outras coisas, na similaridade entre os rótulos das atividades. Portanto, espera-se que haja uma uniformização mínima na denominação das atividades para que a abordagem gere valor.
- A escolha dos parâmetros operacionais adequados para as abordagens propostas nesta tese depende de características do processo analisado, de forma que, requer realização de análise prévia para identificação dos parâmetros ideais para o domínio.
- As abordagens propostas reduzem parte do esforço com o tratamento de dados, mas podem gerar inconsistências, logo não eliminam a necessidade de um especialista no negócio que possa apontar eventuais distorções para correção.
- Os estudos conduzidos abordaram os impactos das transformações dos logs de eventos nas técnicas de mineração de processos voltadas para verificação de conformidade.

8.2 Trabalhos futuros

As abordagens desenvolvidas nesta tese apontam para várias possibilidades de trabalhos futuros. Agrupamos tais possibilidades nas seguintes categorias:

- Desenvolvimento de algoritmos e ferramentas: implementação, na forma de bibliotecas em Python, das técnicas propostas nesta tese para utilização conjunta com outras abordagens algorítmicas de mineração de processos, como o PM4Py. Além disso, as abordagens aqui propostas podem ser implementadas na forma de *node* para utilização no KNIME conjuntamente com a extensão PM4KNIME ou mesmo na forma de *plug-in* para o ProM;
- Expansão do conjunto de métricas: a expansão do conjunto de métricas utilizadas nas abordagens, bem como das métricas usadas para a avaliação dos modelos de processos, pode proporcionar modelos mais qualificados e/ou *insights* para melhoria da abordagem;

- Definição de parâmetros ideais: a escolha de parâmetros adequados pode resultar em modelos melhores. O estabelecimento de valores de referência ou de uma abordagem para otimização de parâmetros para um modelo de processos deve contribuir significativamente para a obtenção de resultados melhores;

- Aplicação e adaptação a outros domínios: como os estudos conduzidos nesta tese foram aplicados em um domínio específico, entendemos ser relevante aplicar e avaliar em outros domínios para verificar se os resultados são similares;

- Avaliação do impacto no custo/esforço com o uso das abordagens propostas em relação ao pré-processamento manual dos logs de eventos.

REFERÊNCIAS

AALST, W. M. P. VAN DER. **Process Mining: Discovery, Conformance and Enhancement of Business Processes**. [s.l.] Springer Heidelberg Dordrecht London New York, 2011.

AALST, W. M. P. VAN DER; WEIJTERS, A. J. M. M.; MARUSTER, L. Workflow Mining: Discovering process models from event logs. **IEEE Transactions on Knowledge and Data Engineering**, n. 16(9), p. 1128–1142, 2004.

ADRIANSYAH, A. et al. Measuring precision of modeled behavior. **Information Systems and e-Business Management**, v. 13, n. 1, 2014.

ADRIANSYAH, A.; BUIJS, J. C. A. M. Mining process performance from event logs. **Lecture Notes in Business Information Processing**, v. 132 LNBIP, p. 217–218, 2013.

AGRAWAL, R.; GUNOPULOS, D.; LEYMANN, F. **Mining Process Models from Workflow Logs**. Proceedings of the 6th International Conference on Extending Database Technology: Advances in Database Technology (EDBT ’98). **Anais**...1998

AUGUSTO, A. et al. Automated Discovery of Process Models from Event Logs: Review and Benchmark. **IEEE Transactions on Knowledge and Data Engineering**, v. 31, n. 4, p. 686–705, 2017.

BERLINGERIO, M. et al. Temporal mining for interactive workflow data analysis. **Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’09**, p. 109, 2009.

BERTI, A. et al. Process mining for python (PM4py): Bridging the gap between process- And data science. **CEUR Workshop Proceedings**, v. 2374, p. 13–16, 2019.

BERTI, A.; VAN DER AALST, W. Reviving token-based replay: Increasing speed while improving diagnostics. **CEUR Workshop Proceedings**, v. 2371, p. 87–103, 2019.

BOENNER, A. **Process Mining in Action Principles, Use Cases and Outlook**. [s.l: s.n.].

BOFF, L. O.; HASSE, F. IMPLICAÇÕES DO USO DAS TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO (TIC’S) E DA SOCIEDADE DIGITAL NO ACESSO À JUSTIÇA NO PROCESSO JUDICIAL ELETRÔNICO– PJe. **Revista Jurídica - CCJ**, v. 21, n. 44, p. 161–183, 2017.

BRASIL. **Lei n° 11.419 de 19 de dezembro de 2006**, 2006.

BUIJS, J. C. A. M. **Flexible Evolutionary Algorithms for Mining Structured Process Models**. [s.l: s.n.].

BUIJS, J. C. A. M.; VAN DONGEN, B. F.; VAN DER AALST, W. M. P. Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. **International Journal of Cooperative Information Systems**, v. 23, n. 1, 2014.

CARDOSO, J. Control-flow Complexity Measurement of Processes and Weyuker's Properties. **6th International Conference on Enformatika**, v. 8, p. 213–218, 2005.

CHAPELA-CAMPA, D.; MUCIENTES, M.; LAMA, M. Simplification of Complex Process Models by Abstracting Infrequent Behaviour. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 11895 LNCS, p. 415–430, 2019.

CLAES, J.; POELS, G. Process Mining and the ProM Framework: A Survey. **Proc. BPM '12 Workshops**, v. 28, n. question 10, p. 187–198, 2012.

CNJ. **PJe · Processo Judicial Eletrônico**Brasília, DFCNJ, , 2010. Disponível em: <http://www.cnj.jus.br/tecnologia-da-informacao/processo-judicial-eletronico-pje>

CONFORTI, R.; ROSA, M. LA; HOFSTEDE, A. H. M. TER. Filtering out Infrequent Behavior from Business Process Event Logs. **IEEE Transactions on Knowledge and Data Engineering**, v. 29, n. 2, p. 300–314, 2017.

COOK, J. E.; WOLF, A. L. Discovering models of software processes from event-based data. **ACM Transactions on Software Engineering and Methodology**, v. 7, n. 3, p. 215–249, 1998.

COOK, J. E.; WOLF, A. L. Software process validation: quantitatively measuring the correspondence of a process to a model. **ACM Transactions on Software Engineering and Methodology**, v. 8, n. 2, p. 147–176, 1999.

D'CASTRO, R. J.; OLIVEIRA, A. L. I.; TERRA, A. H. Process mining discovery techniques in a low-structured process works? **Proceedings - 2018 Brazilian Conference on Intelligent Systems, BRACIS 2018**, p. 200–205, 2018.

DAVIS, R.; BRABANDER, E. **Aris Design : Getting Started with BPM**. [s.l.] Springer Science & Business Media, 2007.

DE MEDEIROS, A. K. A.; WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P. Genetic process mining: An experimental evaluation. **Data Mining and Knowledge Discovery**, v. 14, n. 2, p. 245–304, 2007.

DE WEERDT, J. et al. A robust F-measure for evaluating discovered process models. **IEEE SSCI 2011: Symposium Series on Computational Intelligence - CIDM 2011: 2011 IEEE Symposium on Computational Intelligence and Data Mining**, p. 148–155, 2011.

DONGEN, B. VAN; DIJKMAN, R.; MENDLING, J. Measuring similarity between business process models. **Advanced Information Systems Engineering**, v. 5074, p. 450–464, 2008.

DOWLING, G. R.; HALL, P. Approximate string matching. **ACM Comput. Surveys**,

v. 12, p. 381–402, 1980.

DRESSLER, K.; NGOMO, A. C. N. On the efficient execution of bounded Jaro-Winkler distances. **Semantic Web**, v. 8, n. 2, p. 185–196, 2017.

DUMAS, M.; VAN DER AALST, W. M. P.; TER HOFSTEDE, A. H. M. **Process-Aware Information Systems: Bridging People and Software through Process Technology**. [s.l.] John Wiley & Sons, 2005.

DUNZER, S. et al. Conformance checking: A state-of-the-art literature review. **ACM International Conference Proceeding Series**, 2019.

FANI SANI, M.; VAN ZELST, S. J.; VAN DER AALST, W. M. P. Repairing Outlier Behavior in Event Logs using Contextual Behavior. **Enterprise Modelling and Information Systems Architectures**, v. 14, n. 5, p. 115–131, 2018a.

FANI SANI, M.; VAN ZELST, S. J.; VAN DER AALST, W. M. P. Repairing outlier behaviour in event logs. **Lecture Notes in Business Information Processing**, v. 320, p. 115–131, 2018b.

FOLINO, F.; GUARASCIO, M.; PONTIERI, L. Mining multi-variant process models from low-level logs low-level logs. **Business information systems**, p. 165–177, 2015.

GARCIA, C. DOS S. et al. Process mining techniques and applications – A systematic mapping study. **Expert Systems with Applications**, v. 133, p. 260–295, 2019.

GOEDERTIER, S. et al. Robust Process Discovery with Artificial Negative Events. **Journal of Machine Learning Research**, v. 10, p. 1305--1340, 2009.

GÜNTHER, C. W. **Process Mining in Flexible Environments**. [s.l.] Technische Universiteit Eindhoven, 2009.

GÜNTHER, C. W.; VAN DER AALST, W. M. P. **Fuzzy Mining - Adaptive Process Simplification Based on Multi-perspective Metrics**. Proceedings of the 5th International Conference on Business Process Management (BPM 2007). **Anais**...2007Disponível em: <http://link.springer.com/10.1007/978-3-540-75183-0>

HAMMER, M.; CHAMPY, J. **Reengineering the Corporation A Manifesto for Business Revolution**. New York: [s.n.].

HAN, J.; KAMBER, M.; HARDCOVER, J. P. **Data Mining: Concepts and Techniques**. 3a. ed. [s.l.] Morgan Kaufmann, 2011.

IEEE COMPUTATIONAL INTELLIGENCE SOCIETY. **Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams**. [s.l: s.n.].

JEH, G.; WIDOM, J. **SimRank: A Measure of Structural-Context Similarity**. Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining KDD. **Anais**...2002

LASSEN, K. B.; VAN DER AALST, W. M. P. Complexity metrics for Workflow nets. **Information and Software Technology**, v. 51, n. 3, p. 610–626, 2009.

LEVENSHTEIN, V. Binary codes capable of correcting deletions, insertions and reversals. **Soviet Physics Doklady**, v. 10, n. 8, p. 707–710, 1966.

MALONE, T. W.; CROWSTON, K.; HERMAN, G. A. **Organizing business knowledge: The MIT process handbook**. [s.l.] MIT press, 2003.

MARC KERREMANS. **Market Guide for Process MiningGartner**. [s.l: s.n.]. Disponível em: <https://www.gartner.com/doc/reprints?id=1-SBXXPQO&ct=190625&st=sb>.

MCCABE, J. A complexity Measure. **IEEE Transactions on Software Engineering**, v. SE-2, n. 4, p. 308–320, 1976.

MENDLING, J. et al. Detection and prediction of errors in EPCs of the SAP reference model. **Data and Knowledge Engineering**, v. 64, n. 1, p. 312–329, 2008.

MENDLING, J.; HESSE, W.; OBERWEIS, A. The impact of activity labeling styles on process model quality. **Lecture Notes in Informatics**, v. P-129, n. May, 2008.

MENDLING, J.; NEUMANN, G.; VAN DER AALST, W. M. P. **Understanding the Occurrence of Errors in Process Models based on Metrics**. Proceedings of the 2007 OTM Confederated international conference on On the move to meaningful internet systems: CoopIS, DOA, ODBASE, GADA, and IS. **Anais**...2007

MUÑOZ-GAMA, J.; CARMONA, J. A fresh look at precision in process conformance. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6336 LNCS, p. 211–226, 2010.

MUNSON, M. A. A study on the importance of and time spent on different modeling steps. **ACM SIGKDD Explorations Newsletter**, v. 13, n. 2, p. 65–71, 2012.

MURATA, T. **Petri nets: Properties, analysis and applications.** IEEE. **Anais**...1989

OMG. **Business Process Model and Notation (bpmn) version 2.0**. Disponível em: <http://www.omg.org/spec/BPMN/2.0.2/>. Acesso em: 1 fev. 2018.

PRODEL, M. et al. Optimal Process Mining for Large and Complex Event Logs. **IEEE Transactions on Automation Science and Engineering**, v. 15, n. 3, p. 1309–1325, 2018.

REHSE, J. R.; FETTKE, P. Clustering Business Process Activities for Identifying Reference Model Components. **Lecture Notes in Business Information Processing**, v. 342, p. 5–17, 2019.

REMBERT, A. J.; ELLIS, C. **An initial approach to mining multiple perspectives of a business process**. Proceedings of the Richard Tapia Celebration of Diversity in Computing Conference 2009: Intellect, Initiatives, Insight, and Innovations.

Anais...2009

ROSA, L. et al. APROMORE: An advanced process model repository. **Expert Systems with Applications**, v. 38, n. 6, p. 7029–7040, 2011.

ROZINAT, A. et al. Discovering simulation models. **Information Systems**, v. 34, n. 3, p. 305–327, 2009.

ROZINAT, A.; WILL M.P. VAN DER, A. Conformance Checking of Processes Based on Monitoring Real Behavior. **Information Systems**, v. 33, n. 1, p. 64–95, 2008.

SANI, M. F. **Preprocessing Event Data in Process Mining**. CAiSE - International Conference on Advanced Information Systems Engineering. **Anais**...2020

SANI, M. F.; VAN ZELST, S. J.; VAN DER AALST, W. M. P. **Improving process discovery results by filtering outliers using conditional behavioural probabilities**. Business process management workshops - BPM 2017 international workshops, Barcelona, Spain,10–11 Sept 2017. **Anais**...2017

SCHÖNIG, S.; JABLONSKI, S. Comparing declarative process modelling languages from the organisational perspective. **Lecture Notes in Business Information Processing**, v. 256, p. 17–29, 2016.

SHARP, A.; MCDERMOTT, P. **Workflow modeling: tools for process improvement and applications development**. [s.l.] Artech House, 2009.

SUN, X. et al. Filtering out noise logs for process modelling based on event dependency. **Proceedings - 2019 IEEE International Conference on Web Services, ICWS 2019 - Part of the 2019 IEEE World Congress on Services**, p. 388–392, 2019.

SURIADI, S. et al. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. **Information Systems**, v. 64, p. 132–150, 2017.

TAIT, D. Our world is increasingly complex and fast moving - global organizations must optimize to survive in the face of disruption. **Insight for business leaders - FUJITSU BLOG**, 2019.

TAX, N.; SIDOROVA, N.; VAN DER AALST, W. M. P. Discovering more precise process models from event logs by filtering out chaotic activities. **Journal of Intelligent Information Systems**, v. 52, n. 1, p. 107–139, 2018.

VAN DER AALST, W. Process mining: Overview and opportunities. **ACM Transactions on Management Information Systems**, v. 3, n. 2, p. 1–17, 2012.

VAN DER AALST, W. M. P. Process Discovery: Capturing the Invisible. **IEEE Computational Intelligence Magazine**, n. February, p. 28–41, 2010.

VAN DER AALST, W. M. P. Process mining: discovering and improving Spaghetti and Lasagna processes. **2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)**, n. c, p. 1–7, 2011.

VAN DER AALST, W. M. P. et al. **Process mining manifesto**. Lecture Notes in Business Information Processing. **Anais**...2012

VAN DER AALST, W. M. P. Mediating between modeled and observed behavior: The quest for the “right” process: Keynote. **Proceedings - International Conference on Research Challenges in Information Science**, 2013.

VAN DER AALST, W. M. P. **Process mining: Data science in action**. 2. ed. [s.l.] Springer Heidelberg New York Dordrecht London, 2016.

VAN DER AALST, W. M. P.; ADRIANSYAH, A.; VAN DONGEN, B. Causal nets: A modeling language tailored towards process discovery. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6901 LNCS, p. 28–42, 2011.

VAN DER AALST, W. M. P.; ADRIANSYAH, A.; VAN DONGEN, B. Replaying history on process models for conformance checking and performance analysis. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 2, n. 2, p. 182–192, 2012.

VAN DER AALST, W. M. P.; PESIC, M.; SCHONENBERG, H. Declarative workflows: Balancing between flexibility and support. **Computer Science - Research and Development**, v. 23, n. 2, p. 99–113, 2009.

VAN DONGEN, B. F. et al. The ProM framework: A new era in process mining tool support. **Application and Theory of Petri Nets**, v. 3536, n. i, p. 444–454, 2005.

VAN DONGEN, B. F.; MENDLING, J.; VAN DER AALST, W. M. P. Structural Patterns for Soundness of Business Process Models. **10th IEEE International Enterprise Distributed Object Computing Conference (EDOC’06)**, p. 116–128, 2006.

VAN ZELST, S. J. et al. Filtering spurious events from event streams of business processes. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 10816 LNCS, p. 35–52, 2018.

VAN ZELST, S. J. et al. Event abstraction in process mining: literature review and taxonomy. **Granular Computing**, v. 2, 2020.

VERBEEK, H. M. W. et al. XES, XESame, and ProM 6. **Information Systems Evolution**, p. 60–75, 2010.

WEIJTERS, A. J. M. M.; RIBEIRO, J. T. S. **Flexible Heuristics Miner (FHM)**. CIDM. **Anais**...IEEE, 2011

WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P. Rediscovering Workflow Models from Event-Based Data using Little Thumb. **Integrated ComputerAided Engineering**, v. 10, p. 151–162, 2003.

WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P.; MEDEIROS, A. K. A. DE. Process Mining with the Heuristics Miner Algorithm. **Technische Universiteit**

Eindhoven, Tech. Rep. WP, v. 166, p. 1–34, 2006.

WESKE, M. **Business Process Management: Concepts, Languages, Architectures**. 2. ed. [s.l.] Springer-Verlag Berlin Heidelberg, 2012.

WINKLER, W. E. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. **Computing Science and Statistics**, p. 561–565, 1990.

WITTEN, I.; FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2nd Editio ed. [s.l.] Morgan Kaufmann, 2005.

APÊNDICE A – MODELOS DE PROCESSOS

Modelos de processos descobertos com *plug-in* do ProM *Mine for Fuzzy Model* utilizados no estudo apresentado na Seção 4.2.

Modelo 1 (Figura 39)

Parâmetros:

Métricas unárias:

$frequency\ significance\ metric = 1.0$
$Routing\ Siginficance = 1.0$

Métricas binárias:

$frequency\ significance = 1.0$
$distance\ significance = 1.0$
$proximity\ correlation = 1.0$
$endpoint\ correlation = 1.0$

Node filter:

$significance\ cutoff = 0.000$

Edge Filter:

$cutoff = 0.200$
$utility\ rate = 0.750$
$edge\ transformer = 'fuzzy\ edges'$
$ignore\ self-loops = checked$
$interpret\ absolute = checked$

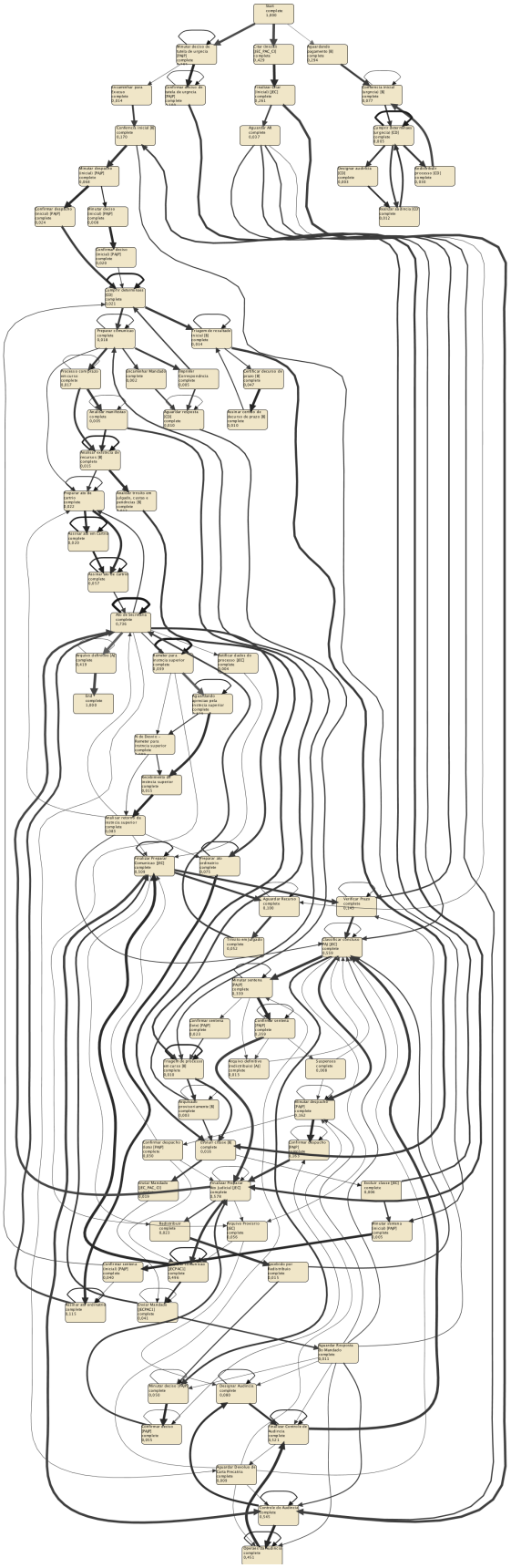
Concurrency filter:

$preserve = 0.600$
$ratio = 0.700$
$filter\ concurrency = checked$

Métricas sobre o modelo:

Model Detail: 100%
Log Conformance: 78,21%

Figura 39 – Modelo 1



Modelo 2 (Figuras 40 a 49)

Parâmetros:

Métricas unárias:

$frequency\ significance\ metric = 1.0$
$Routing\ Siginficance = 1.0$

Métricas binárias:

$frequency\ significance = 1.0$
$distance\ significance = 1.0$
$proximity\ correlation = 1.0$
$endpoint\ correlation = 1.0$

Node filter:

$significance\ cutoff = 0.250$

Edge Filter:

$cutoff = 0.200$
$utility\ rate = 0.750$
$edge\ transformer = 'fuzzy\ edges'$
$ignore\ self - loops = checked$
$interpret\ absolute = checked$

Concurrency filter:

$preserve = 0.600$
$ratio = 0.700$
$filter\ concurrency = checked$

Métricas sobre o modelo:

Model Detail: 78,2%
Log Conformance: 85,91%

Figura 40 – Modelo 2

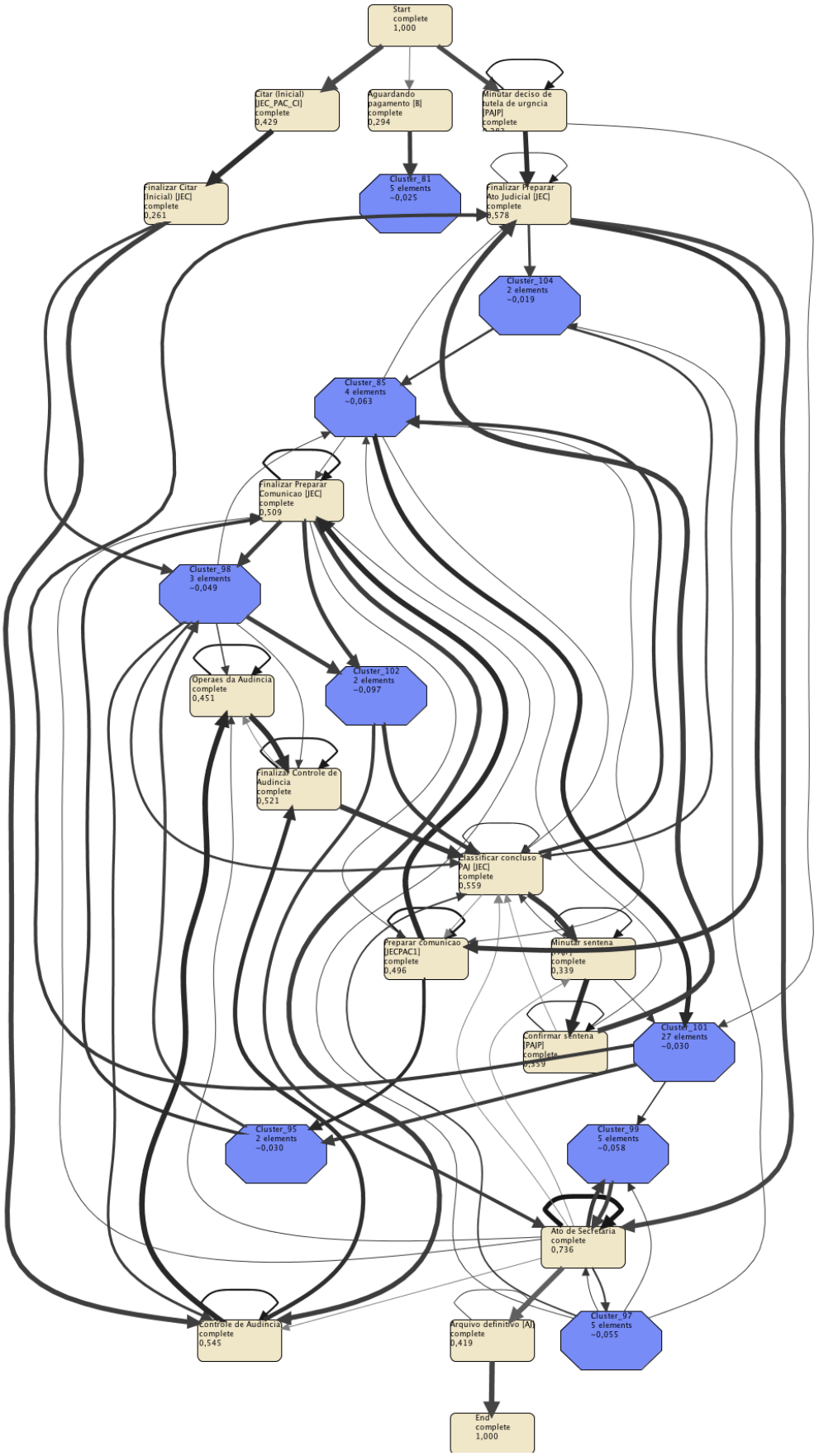


Figura 41 – Cluster 81 (Modelo 2)

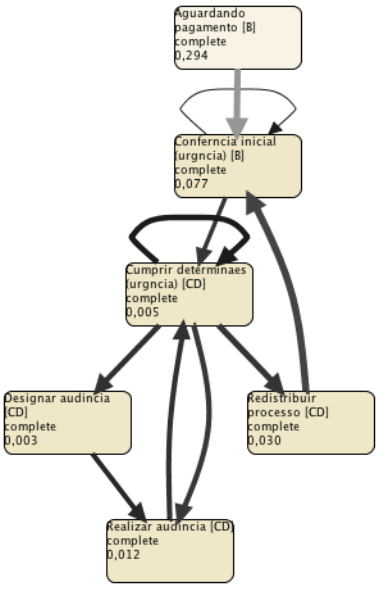


Figura 42 – Cluster 85 (Modelo 2)

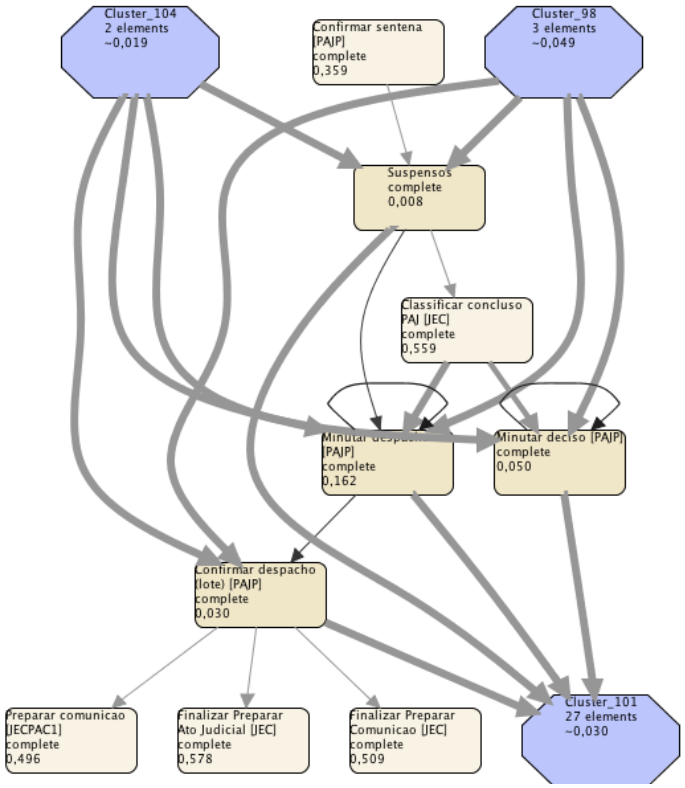


Figura 43 – Cluster 95 (Modelo 2)

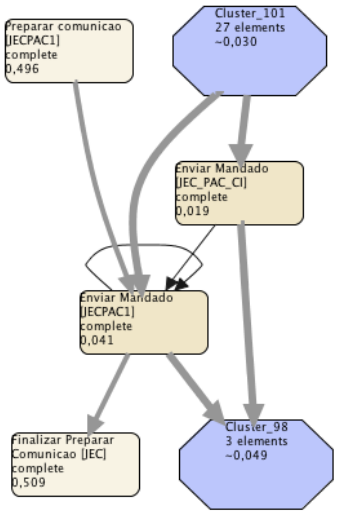


Figura 44 – Cluster 97 (Modelo 2)

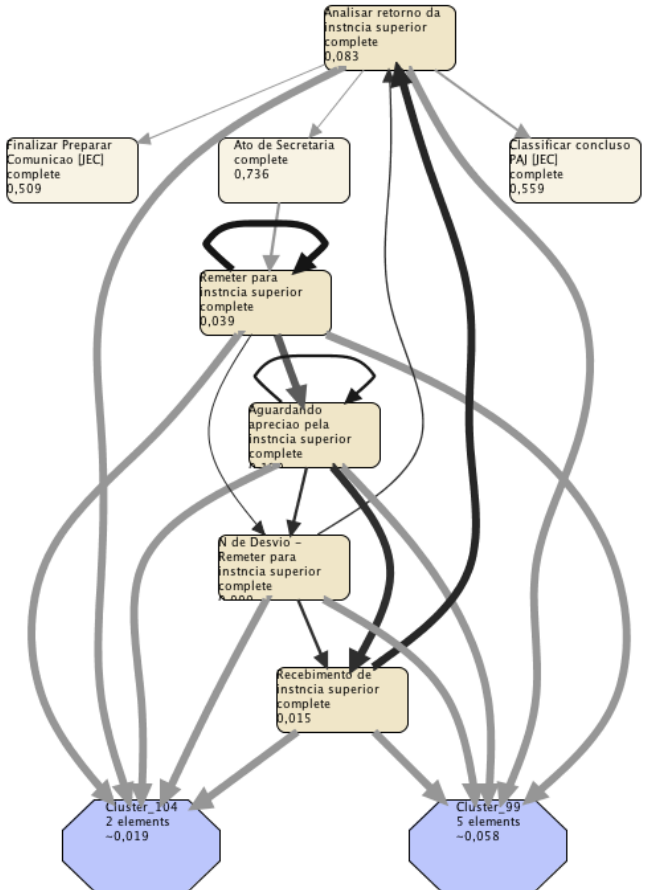


Figura 45 – Cluster 98 (Modelo 2)

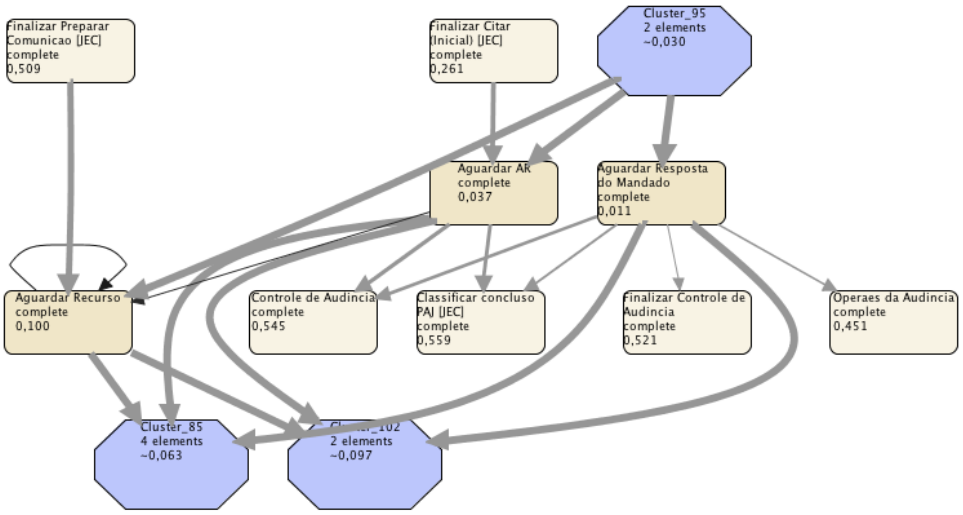


Figura 46 – Cluster 99 (Modelo 2)

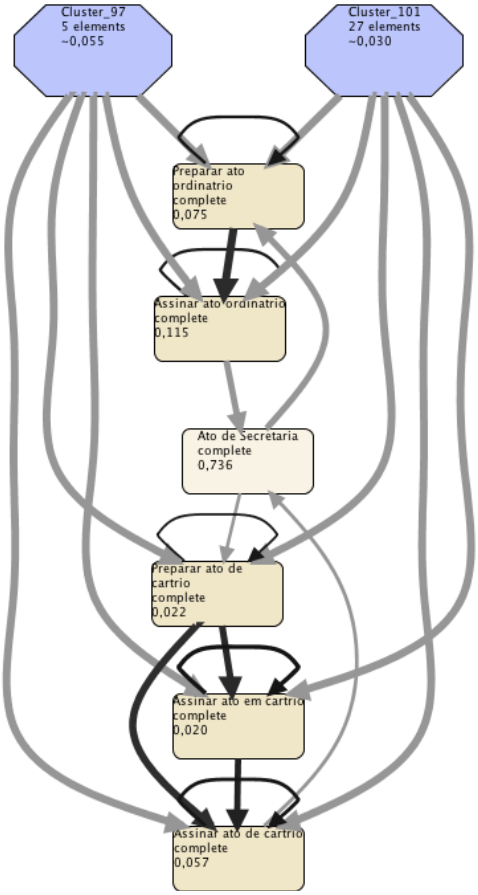


Figura 47 – Cluster 101 (Modelo 2)

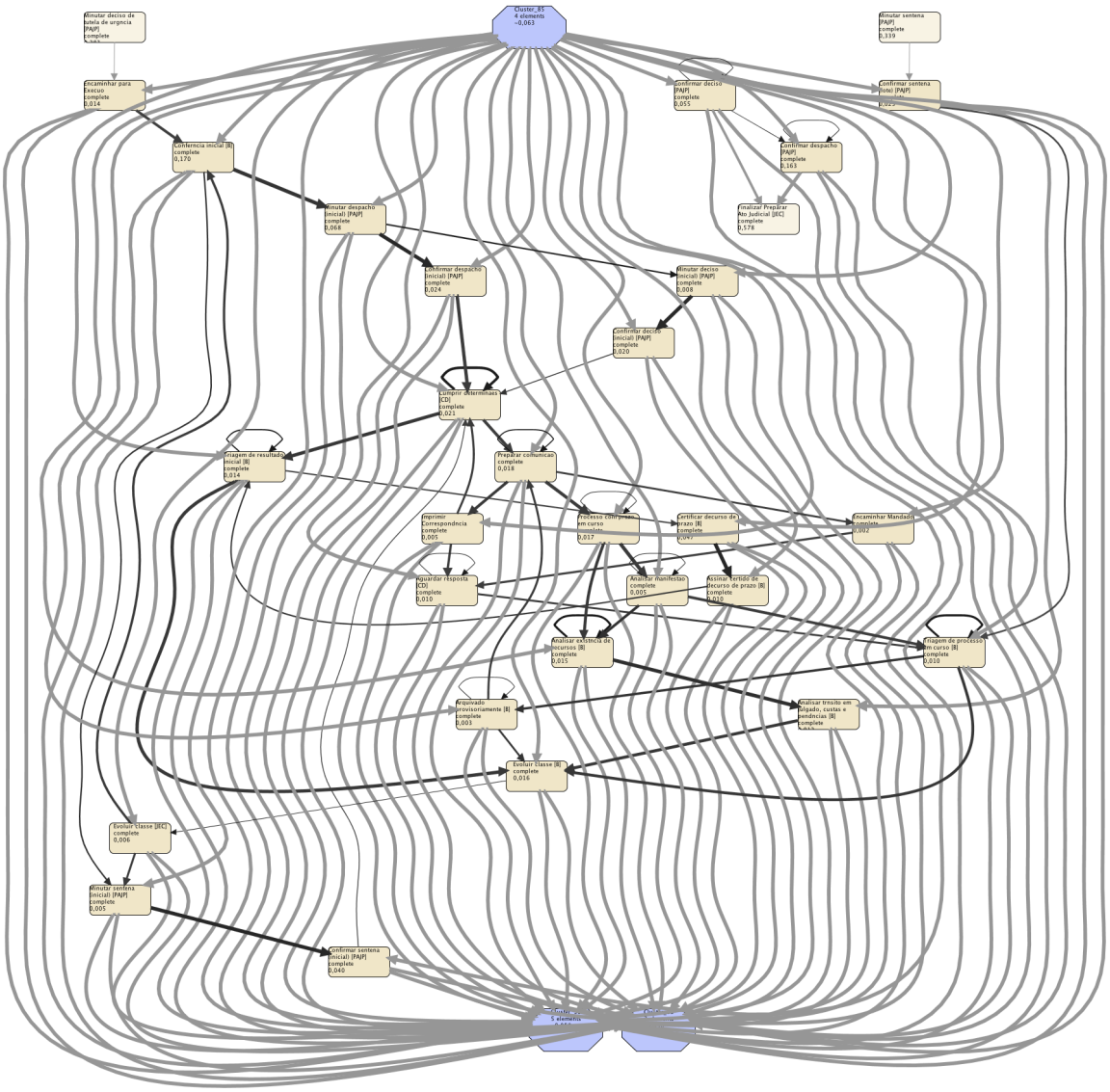


Figura 48 – Cluster 102 (Modelo 2)

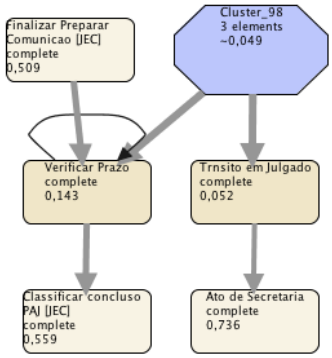
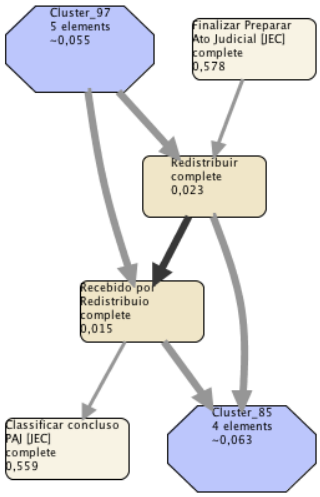


Figura 49 – Cluster 104 (Modelo 2)



Modelo 3 (Figura 50)

Parâmetros:

Métricas unárias:

$frequency\ significance\ metric = 1.0$
$Routing\ Siginficance = 1.0$

Métricas binárias:

$frequency\ significance = 1.0$
$distance\ significance = 1.0$
$proximity\ correlation = 1.0$
$endpoint\ correlation = 1.0$

Node filter:

$significance\ cutoff = 0.500$

Edge Filter:

$cutoff = 0.200$
$utility\ rate = 0.750$
$edge\ transformer = 'fuzzy\ edges'$
$ignore\ self - loops = checked$
$interpret\ absolute = checked$

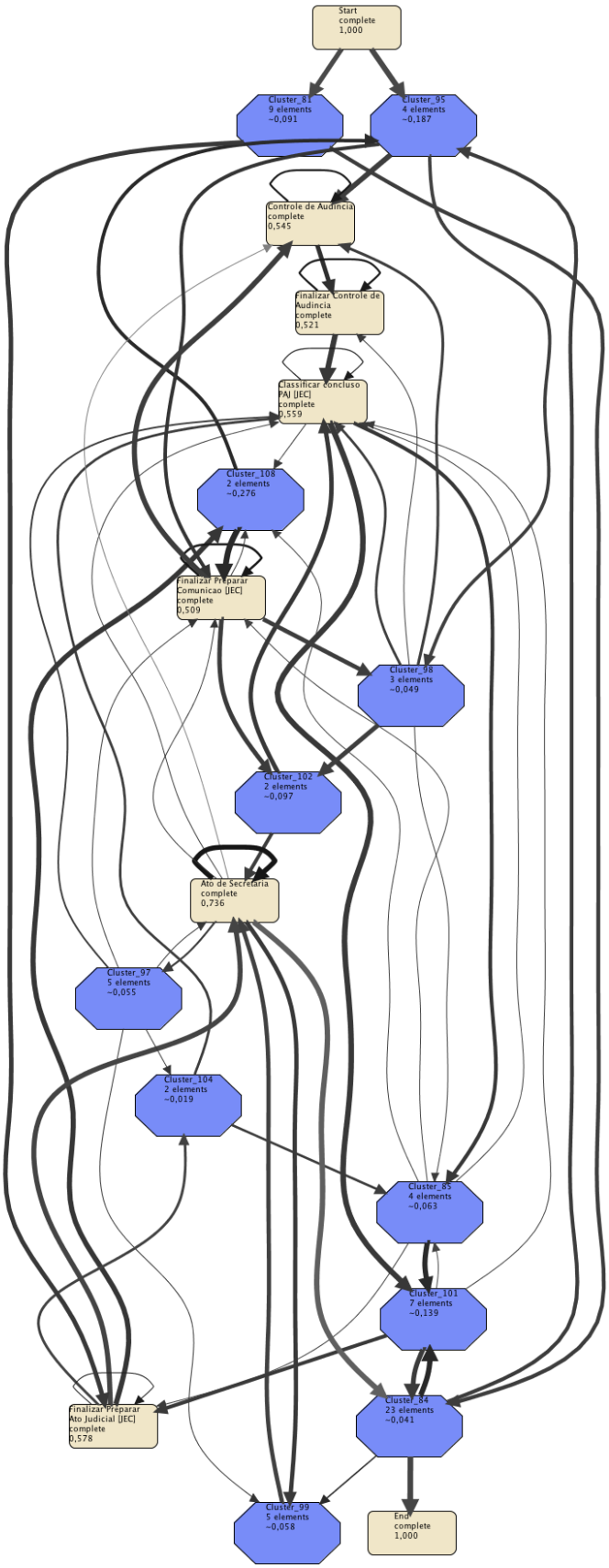
Concurrency filter:

$preserve = 0.600$
$ratio = 0.700$
$filter\ concurrency = checked$

Métricas sobre o modelo:

Model Detail: 48,53%
Log Conformance: 86,75%

Figura 50 – Modelo 3



Modelo 4 (Figuras 51 e 52)

Parâmetros:

Métricas unárias:

$frequency\ significance\ metric = 1.0$
$Routing\ Siginficance = 1.0$

Métricas binárias:

$frequency\ significance = 1.0$
$distance\ significance = 1.0$
$proximity\ correlation = 1.0$
$endpoint\ correlation = 1.0$

Node filter:

$significance\ cutoff = 0.750$

Edge Filter:

$cutoff = 0.200$
$utility\ rate = 0.250$
$edge\ transformer = 'fuzzy\ edges'$
$ignore\ self-loops = checked$
$interpret\ absolute = checked$

Concurrency filter:

$preserve = 0.600$
$ratio = 0.700$
$filter\ concurrency = checked$

Métricas sobre o modelo:

Model Detail: 17,82%
Log Conformance: 100%

Figura 51 – Modelo 4

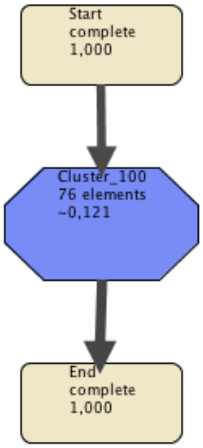
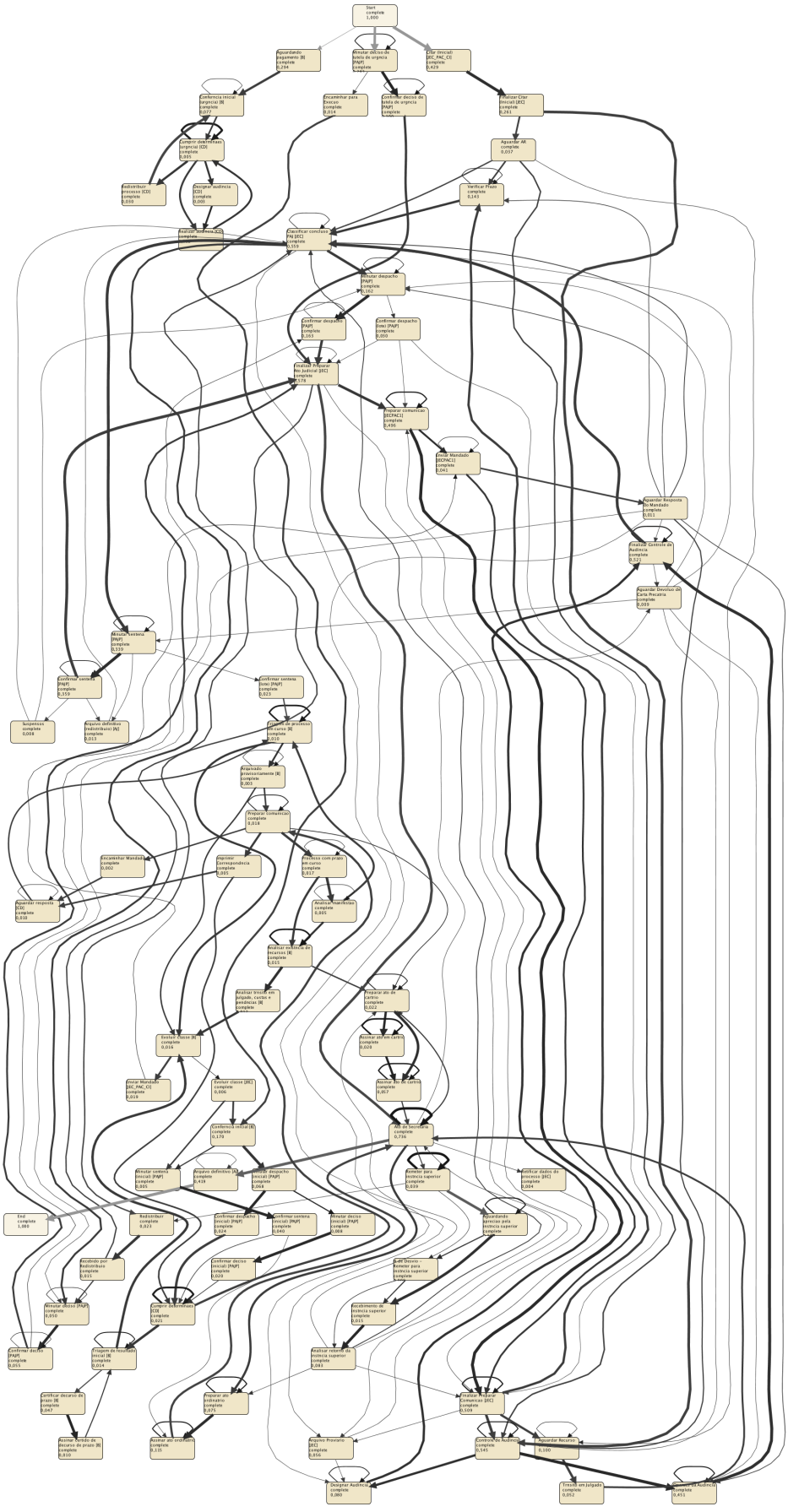
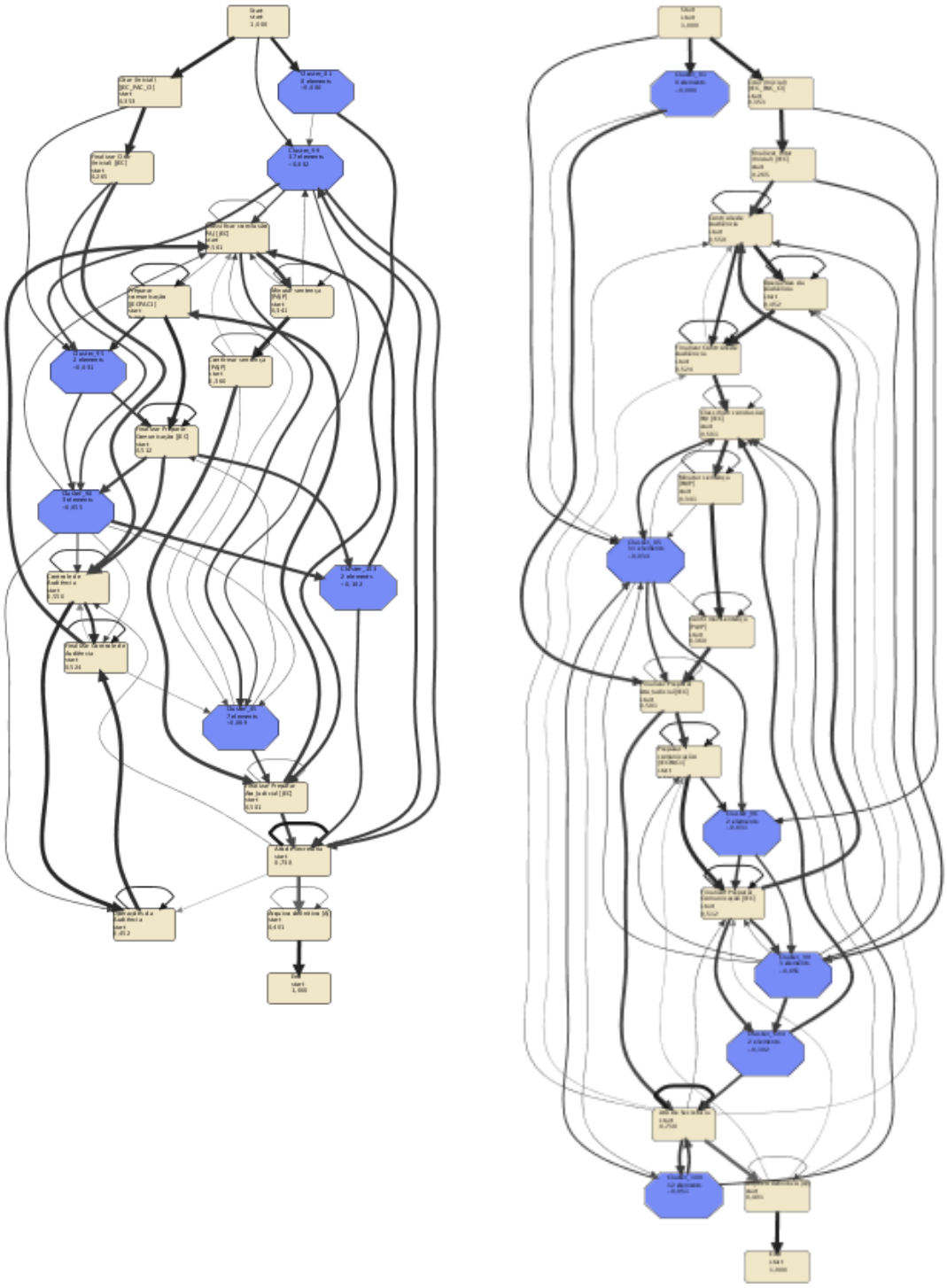


Figura 52 – Cluster 100 (Modelo 4)



A Figura 53 mostra dois modelos de processos com parâmetro $preserve = 1.000$ (Figura 52a) e $preserve = 0.000$ (Figura 52b).

Figura 53 – Modelos de processos (a) *preserve*=1.000 e (b) *preserve*=0.000



As Figuras 54 a 56 mostram, respectivamente, modelos de processos com os parâmetros $utility\ rate = 0.250$, $utility\ rate = 0.500$ e $utility\ rate = 0.750$.

Figura 54 – Modelo de processo com *utility rate*=0.250

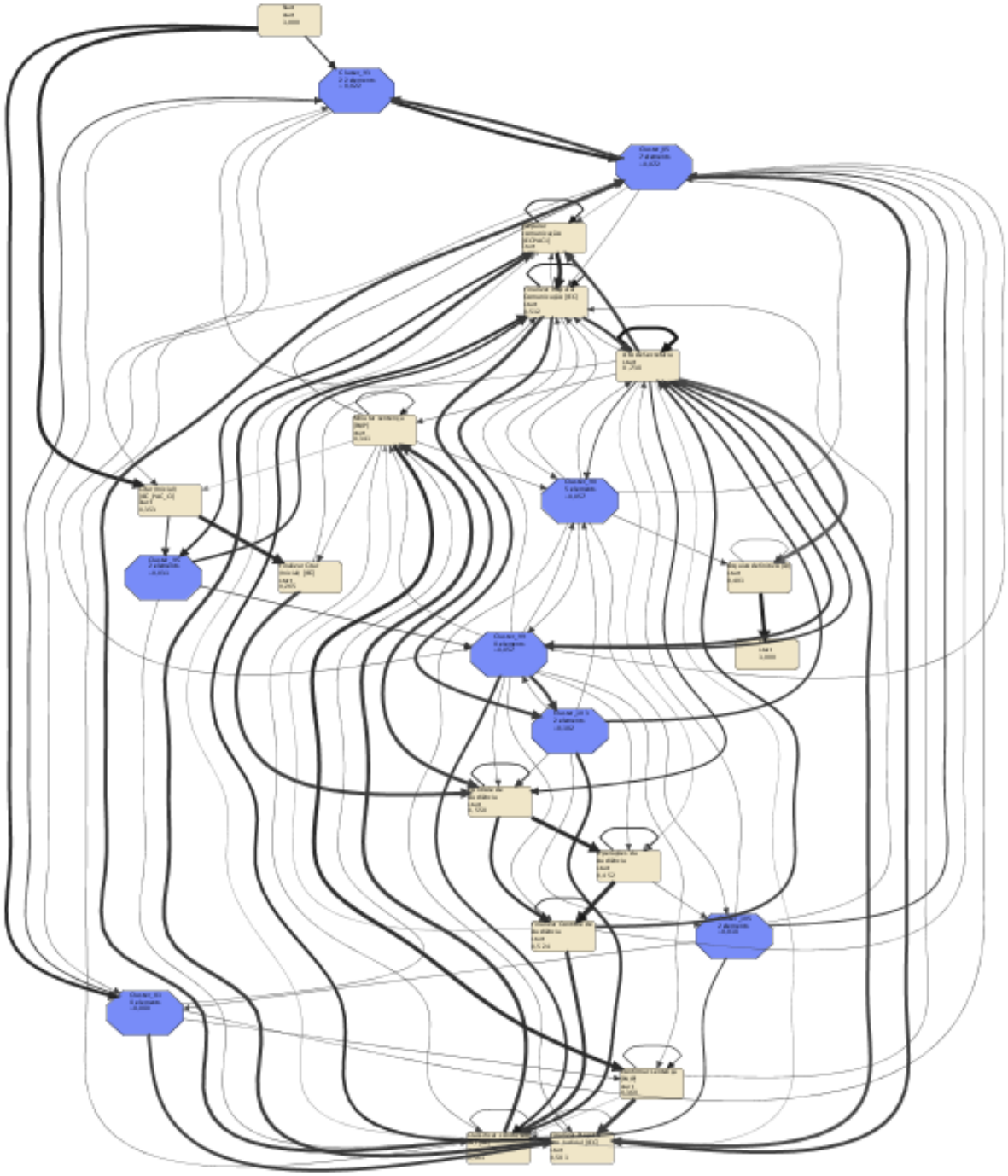


Figura 55 – Modelo de processo com *utility rate*=0.500

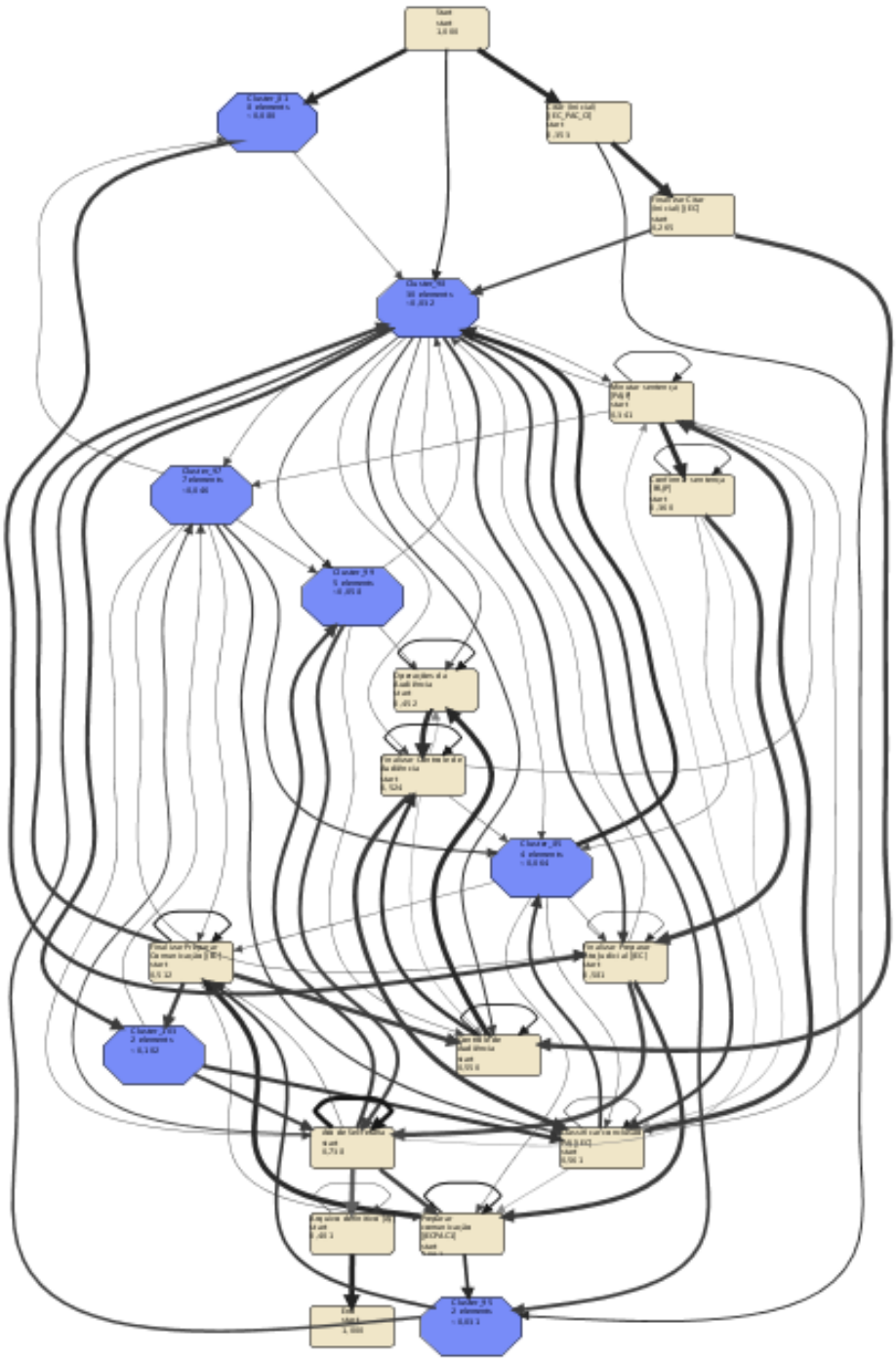
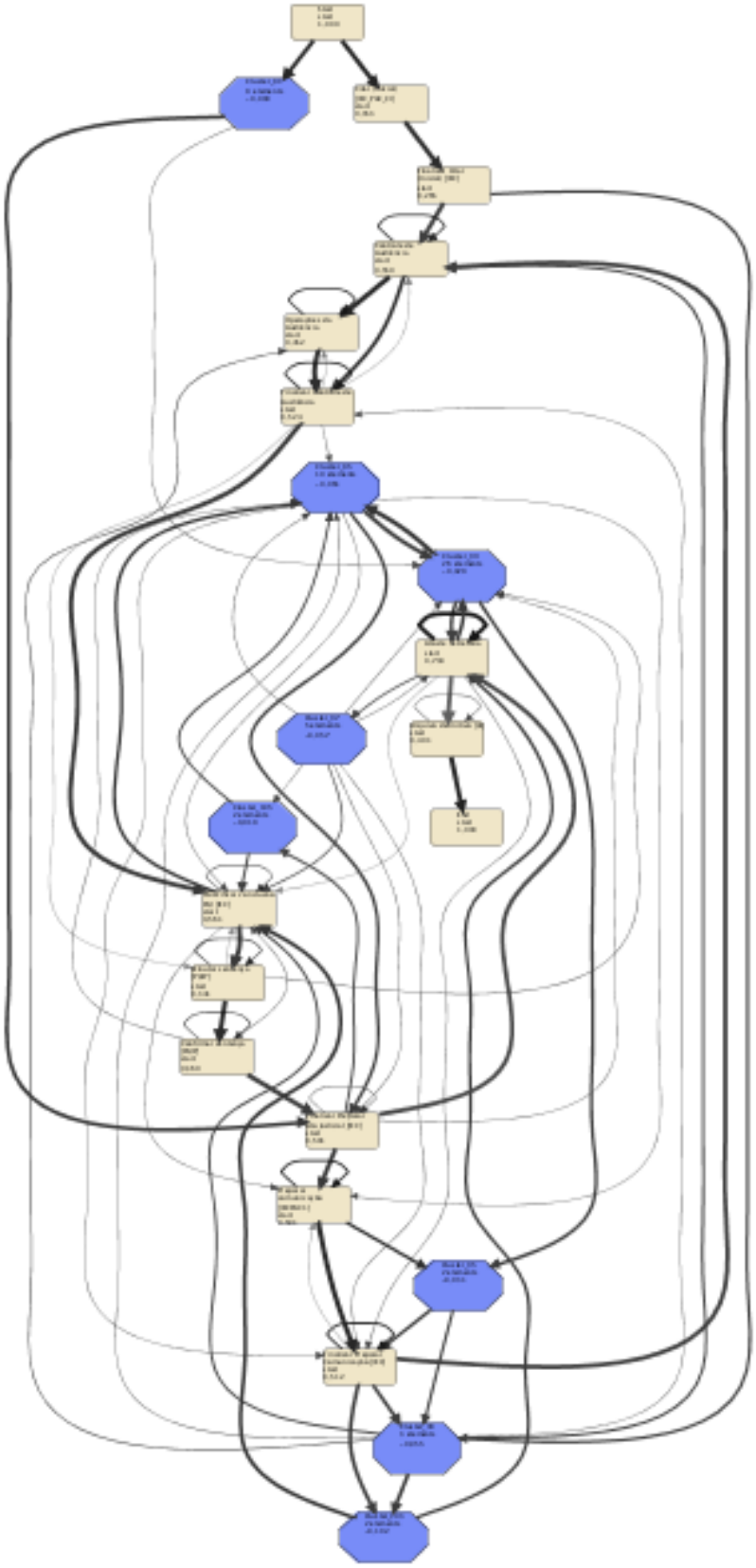


Figura 56 – Modelo de processo com *utility rate*=0.750



APÊNDICE B – MÉTRICAS DE QUALIDADE

Neste apêndice mostramos os valores obtidos para as métricas de qualidade dos experimentos realizados com as três abordagens apresentadas nesta tese sobre os três logs de eventos apresentados. A identificação dos logs de eventos é denotada por $\mathcal{L}_M^n$, tal que $n = \{1,2,3\}$ é o log de eventos de referência (Tabela 12) e $M = \{\emptyset, A, F, D\}$ a abordagem utilizada, sendo:

- $M = \emptyset$, log de eventos original (sem transformação)
- $M = A$, log de eventos transformado pela abordagem de agrupamento de atividades afins (Capítulo 5)
- $M = F$, log de eventos transformado pela abordagem de agrupamento de atividades com eliminação de comportamento infrequente (Capítulo 6)
- $M = D$, log de eventos transformado pela abordagem de desmembramento de atividades recorrentes (Capítulo 7)

Por exemplo, temos que o log de eventos $\mathcal{L}_A^2$ foi gerado através da transformação do log de eventos 2 (Tabela 12) a partir da abordagem de agrupamento de atividades afins (Capítulo 5). As Tabelas 25 a 36 mostram os parâmetros (P1 a P5) utilizados nas 64 combinações diferentes para o algoritmo *Heristic Miner* (PM4PY), sendo:

- DEPENDENCY_THRESH (P1)
- AND_MEASURE_THRESH (P2)
- MIN_ACT_COUNT (P3)
- MIN_DFG_OCCURRENCES (P4)
- DFG_PRE_CLEANING_NOISE_THRESH (P5)

Além disso as tabelas contemplam as seguintes métricas de qualidade (PM4PY) para os modelos gerados:

- perc_fit_traces (MF1)
- average_trace_fitness (MF2)
- log_fitness (MF3)
- precision (MP)

Por fim, a coluna T(s) indica a duração em segundos para cálculo das métricas de qualidade do modelo.

Tabela 25 – Métricas de qualidade log de eventos $\mathcal{L}^1$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	0,148853826	0,96568424	0,96392491	0,56110972	241,34
2	0,5	0,65	1	100	0,05	0,148853826	0,965854768	0,96421048	0,56084764	244,71
3	0,5	0,65	100	1	0,05	0,148853826	0,965814692	0,96405567	0,56057417	231,32
4	0,5	0,65	100	100	0,05	0,148853826	0,965829283	0,96407859	0,55989194	231,36
5	0,5	0,65	500	1	0,05	0,148853826	0,965799831	0,96403726	0,5599657	232,08
6	0,5	0,65	500	100	0,05	0,148853826	0,965970023	0,96432114	0,56008245	235,39
7	0,5	0,35	1	1	0,05	0,148853826	0,965977379	0,9643278	0,56026386	236,03
8	0,5	0,35	1	100	0,05	0,148853826	0,96598282	0,96433266	0,56014004	237,7
9	0,5	0,35	100	1	0,05	0,148853826	0,965812482	0,96404901	0,56048733	179
10	0,5	0,35	100	100	0,05	0,148853826	0,965680572	0,96391825	0,56031127	185,81
11	0,5	0,35	500	1	0,05	0,148853826	0,965819571	0,96406042	0,55994118	177,03
12	0,5	0,35	500	100	0,05	0,148853826	0,965796503	0,96403028	0,56037152	183,18
13	0,5	0,85	1	1	0,05	0,148853826	0,96598282	0,96433266	0,56019576	190,01
14	0,5	0,85	1	100	0,05	0,148853826	0,965982842	0,96433699	0,55871069	198,64
15	0,5	0,85	100	1	0,05	0,148853826	0,964523983	0,96247301	0,56153426	189,39
16	0,5	0,85	100	100	0,05	0,148853826	0,965688984	0,96392647	0,56145205	181,9
17	0,5	0,85	500	1	0,05	0,148853826	0,965824755	0,96406773	0,5607147	184,79
18	0,5	0,85	500	100	0,05	0,148853826	0,965846113	0,96419997	0,55943869	183,8
19	0,75	0,65	1	1	0,05	0,148853826	0,965855314	0,96421073	0,55965689	189,79
20	0,75	0,65	1	100	0,05	0,148853826	0,967893765	0,96674086	0,56466356	165,33
21	0,75	0,65	100	1	0,05	0,148853826	0,96598304	0,96433453	0,55950653	188,17
22	0,75	0,65	100	100	0,05	0,148853826	0,965798827	0,96403348	0,55859228	194,32
23	0,75	0,65	500	1	0,05	0,148853826	0,965701256	0,96394519	0,56106482	185,34
24	0,75	0,65	500	100	0,05	0,148853826	0,965989908	0,96434409	0,56002841	188,96
25	0,75	0,35	1	1	0,05	0,148853826	0,965976451	0,96432702	0,55956692	185,21
26	0,75	0,35	1	100	0,05	0,148853826	0,964634459	0,9626735	0,56176559	179,6
27	0,75	0,35	100	1	0,05	0,148853826	0,965862182	0,96422029	0,56109097	187,88
28	0,75	0,35	100	100	0,05	0,148853826	0,965916725	0,96419028	0,56047004	187,3
29	0,75	0,35	500	1	0,05	0,148853826	0,965805015	0,96404457	0,56051199	176,85
30	0,75	0,35	500	100	0,05	0,148853826	0,965688984	0,96392647	0,56135611	180,44
31	0,75	0,85	1	1	0,05	0,148853826	0,965867708	0,96422381	0,5609048	185,83
32	0,75	0,85	1	100	0,05	0,148853826	0,965799572	0,96403539	0,56056885	184,71
33	0,75	0,85	100	1	0,05	0,148853826	0,96598304	0,96433453	0,55914374	194,5

34	0,75	0,85	100	100	0,05	0,148853826	0,965967911	0,96431536	0,56033588	189,27
35	0,75	0,85	500	1	0,05	0,148853826	0,965824755	0,96406773	0,56059305	179,29
36	0,75	0,85	500	100	0,05	0,148853826	0,965812482	0,96404901	0,55987167	191,88
37	0,25	0,65	1	1	0,05	0,148853826	0,965679259	0,96391387	0,56200001	183,37
38	0,25	0,65	1	100	0,05	0,148853826	0,965681938	0,96391862	0,56066416	184,03
39	0,25	0,65	100	1	0,05	0,148853826	0,965696293	0,96394178	0,56186087	175,83
40	0,25	0,65	100	100	0,05	0,148853826	0,965844739	0,96419244	0,56076459	188,15
41	0,25	0,65	500	1	0,05	0,148853826	0,96584087	0,96419265	0,5612589	186,3
42	0,25	0,65	500	100	0,05	0,148853826	0,965984208	0,96433736	0,56006623	186,3
43	0,25	0,35	1	1	0,05	0,148853826	0,965963155	0,96431159	0,55874439	192,62
44	0,25	0,35	1	100	0,05	0,148853826	0,965837535	0,96418566	0,56139677	185,66
45	0,25	0,35	100	1	0,05	0,148853826	0,965970023	0,96432114	0,56013928	192,37
46	0,25	0,35	100	100	0,05	0,148853826	0,965817945	0,96405819	0,55980716	187,39
47	0,25	0,35	500	1	0,05	0,148853826	0,965988283	0,96434185	0,56026867	191,46
48	0,25	0,35	500	100	0,05	0,148853826	0,965981894	0,96433621	0,55964342	192,73
49	0,25	0,85	1	1	0,05	0,148853826	0,965995093	0,96435141	0,55997123	184,15
50	0,25	0,85	1	100	0,05	0,148853826	0,965696293	0,96394178	0,56130232	184,8
51	0,25	0,85	100	1	0,05	0,148853826	0,965967911	0,96431536	0,55899692	187,86
52	0,25	0,85	100	100	0,05	0,148853826	0,965804071	0,96404079	0,56007545	177,58
53	0,25	0,85	500	1	0,05	0,148853826	0,96597351	0,9643204	0,55924915	184,14
54	0,25	0,85	500	100	0,05	0,148853826	0,965975085	0,96432665	0,55936426	186,81
55	0.52	0.46	296	72	0.06	0,148853826	0,96588363	0,96418668	0,55955902	255,5
56	0.64	0.60	500	13	0.02	0,148853826	0,96579798	0,96403336	0,56049362	245,55
57	0.46	0.37	231	200	0.07	0,148853826	0,965797721	0,96403149	0,55996053	254,85
58	0.33	0.75	290	37	0.08	0,148853826	0,965819311	0,96405855	0,56000295	245,6
59	0.78	0.56	174	121	0.09	0,148853826	0,965813326	0,96405531	0,56052522	239,8
60	0.67	0.27	282	109	0.05	0,148853826	0,965873151	0,964233	0,56031879	249,22
61	0.69	0.29	43	82	0.02	0,148853826	0,965694447	0,96393565	0,55971168	247,51
62	0.69	0.24	274	58	0.05	0,148853826	0,96596817	0,96431723	0,55959211	249,16
63	0.75	0.22	280	78	0.08	0,148853826	0,9656689	0,96390276	0,56184147	237,54
64	0.52	0.29	472	161	0.09	0,148853826	0,965675729	0,96391231	0,56065815	243,83

Tabela 26 – Métricas de qualidade log de eventos $\mathcal{L}^2$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	0,210637177	0,931055122	0,93117479	0,83876686	57,05
2	0,5	0,65	1	100	0,05	0,210637177	0,931055122	0,93117479	0,83946029	56,1
3	0,5	0,65	100	1	0,05	0,210637177	0,931055122	0,93117479	0,83945702	55,82
4	0,5	0,65	100	100	0,05	0,263296472	0,931067908	0,93121068	0,83906705	56,86
5	0,5	0,65	500	1	0,05	0,263296472	0,931067908	0,93121068	0,83975395	55,95
6	0,5	0,65	500	100	0,05	0,210637177	0,931055122	0,93117479	0,83876033	57,4
7	0,5	0,35	1	1	0,05	0,263296472	0,931067908	0,93121068	0,83972842	54,61
8	0,5	0,35	1	100	0,05	0,263296472	0,931067908	0,93121068	0,83977683	56,01
9	0,5	0,35	100	1	0,05	0,210637177	0,931055122	0,93117479	0,83949623	55,54
10	0,5	0,35	100	100	0,05	0,263296472	0,931067908	0,93121068	0,83904359	57,65
11	0,5	0,35	500	1	0,05	0,210637177	0,931055122	0,93117479	0,83880272	58,01
12	0,5	0,35	500	100	0,05	0,263296472	0,931067908	0,93121068	0,83907621	54,53
13	0,5	0,85	1	1	0,05	0,263296472	0,931067908	0,93121068	0,83906705	54,93
14	0,5	0,85	1	100	0,05	0,263296472	0,931067908	0,93121068	0,83908927	55,78
15	0,5	0,85	100	1	0,05	0,263296472	0,931067908	0,93121068	0,83977419	56,45
16	0,5	0,85	100	100	0,05	0,210637177	0,931055122	0,93117479	0,83949358	56,02
17	0,5	0,85	500	1	0,05	0,263296472	0,931067908	0,93121068	0,83977029	56,14
18	0,5	0,85	500	100	0,05	0,210637177	0,931055122	0,93117479	0,83880335	55,9
19	0,75	0,65	1	1	0,05	0,263296472	0,931067908	0,93121068	0,83975722	55,91
20	0,75	0,65	1	100	0,05	0,263296472	0,931067908	0,93121068	0,83905074	56,14
21	0,75	0,65	100	1	0,05	0,210637177	0,931055122	0,93117479	0,83876686	55,36
22	0,75	0,65	100	100	0,05	0,263296472	0,931067908	0,93121068	0,83905337	56,13
23	0,75	0,65	500	1	0,05	0,263296472	0,931067908	0,93121068	0,8390599	53,87
24	0,75	0,65	500	100	0,05	0,210637177	0,931055122	0,93117479	0,83945375	56,54
25	0,75	0,35	1	1	0,05	0,210637177	0,931055122	0,93117479	0,83876422	57,05
26	0,75	0,35	1	100	0,05	0,210637177	0,931055122	0,93117479	0,8387636	56,13
27	0,75	0,35	100	1	0,05	0,210637177	0,931055122	0,93117479	0,83880661	56,05
28	0,75	0,35	100	100	0,05	0,210637177	0,931055122	0,93117479	0,83874077	57,38
29	0,75	0,35	500	1	0,05	0,210637177	0,931055122	0,93117479	0,83945375	57,48
30	0,75	0,35	500	100	0,05	0,263296472	0,931067908	0,93121068	0,83905664	56,69
31	0,75	0,85	1	1	0,05	0,210637177	0,93106771	0,93119151	0,83892991	57,25
32	0,75	0,85	1	100	0,05	0,210637177	0,931055122	0,93117479	0,83877338	56,37
33	0,75	0,85	100	1	0,05	0,263296472	0,931067908	0,93121068	0,83905337	56,33

34	0,75	0,85	100	100	0,05	0,263296472	0,931067908	0,93121068	0,83906643	56,54
35	0,75	0,85	500	1	0,05	0,210637177	0,931055122	0,93117479	0,83877338	57,21
36	0,75	0,85	500	100	0,05	0,210637177	0,931055122	0,93117479	0,83879294	55,41
37	0,25	0,65	1	1	0,05	0,210637177	0,931055122	0,93117479	0,83879946	55,84
38	0,25	0,65	1	100	0,05	0,210637177	0,931055122	0,93117479	0,83948643	56,39
39	0,25	0,65	100	1	0,05	0,210637177	0,931055122	0,93117479	0,83880272	56,22
40	0,25	0,65	100	100	0,05	0,210637177	0,931055122	0,93117479	0,83880924	57,53
41	0,25	0,65	500	1	0,05	0,210637177	0,931055122	0,93117479	0,83876686	56,59
42	0,25	0,65	500	100	0,05	0,210637177	0,931055122	0,93117479	0,83945111	56,13
43	0,25	0,35	1	1	0,05	0,342285413	0,931097992	0,93125798	0,84058163	56,2
44	0,25	0,35	1	100	0,05	0,210637177	0,931057729	0,93117804	0,83949623	66,61
45	0,25	0,35	100	1	0,05	0,210637177	0,931055122	0,93117479	0,83877012	70,39
46	0,25	0,35	100	100	0,05	0,263296472	0,931067908	0,93121068	0,83906969	66,58
47	0,25	0,35	500	1	0,05	0,210637177	0,931055122	0,93117479	0,83945438	57,93
48	0,25	0,35	500	100	0,05	0,263296472	0,931067908	0,93121068	0,83908274	55,89
49	0,25	0,85	1	1	0,05	0,210637177	0,931055122	0,93117479	0,83880272	58,22
50	0,25	0,85	1	100	0,05	0,210637177	0,931055122	0,93117479	0,83880272	57,54
51	0,25	0,85	100	1	0,05	0,263296472	0,931067908	0,93121068	0,83906705	72,07
52	0,25	0,85	100	100	0,05	0,263296472	0,931067908	0,93121068	0,83906316	56,87
53	0,25	0,85	500	1	0,05	0,210637177	0,931055122	0,93117479	0,83949032	55,92
54	0,25	0,85	500	100	0,05	0,263296472	0,931067908	0,93121068	0,83975395	56,87
55	0.52	0.46	296	72	0.06	0,263296472	0,931067908	0,93121068	0,83906969	69,88
56	0.64	0.60	500	13	0.02	0,210637177	0,931055122	0,93117479	0,83945375	75,45
57	0.46	0.37	231	200	0.07	0,263296472	0,931067908	0,93121068	0,83977029	70,59
58	0.33	0.75	290	37	0.08	0,263296472	0,931067908	0,93121068	0,83908927	76,29
59	0.78	0.56	174	121	0.09	0,210637177	0,931055122	0,93117479	0,8387962	71,1
60	0.67	0.27	282	109	0.05	0,263296472	0,931067908	0,93121068	0,83977683	71,02
61	0.69	0.29	43	82	0.02	0,263296472	0,931067908	0,93121068	0,83972779	71,5
62	0.69	0.24	274	58	0.05	0,263296472	0,931067908	0,93121068	0,83977356	71,8
63	0.75	0.22	280	78	0.08	0,210637177	0,931055122	0,93117479	0,8387799	71,96
64	0.52	0.29	472	161	0.09	0,210637177	0,931055122	0,93117479	0,83946682	71,6

Tabela 27 – Métricas de qualidade log de eventos $\mathcal{L}^3$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	39,84393474	0,970106398	0,97334318	0,63821813	161,17
2	0,5	0,65	1	100	0,05	39,962166	0,970147286	0,97336664	0,6383424	161,59
3	0,5	0,65	100	1	0,05	39,84393474	0,970117386	0,97335702	0,63780924	160,88
4	0,5	0,65	100	100	0,05	39,84393474	0,970101805	0,97333709	0,63832781	158,71
5	0,5	0,65	500	1	0,05	39,84393474	0,970101805	0,97333709	0,63807256	163,2
6	0,5	0,65	500	100	0,05	39,84393474	0,970122352	0,97336311	0,63770315	159,13
7	0,5	0,35	1	1	0,05	39,84393474	0,970122352	0,97336311	0,63781109	159,23
8	0,5	0,35	1	100	0,05	39,84393474	0,970117386	0,97335702	0,6378105	155,84
9	0,5	0,35	100	1	0,05	39,84393474	0,97015612	0,97343382	0,63831177	161,32
10	0,5	0,35	100	100	0,05	39,84393474	0,970101805	0,97333709	0,63832781	160,22
11	0,5	0,35	500	1	0,05	39,84393474	0,970117386	0,97335702	0,63771975	159,73
12	0,5	0,35	500	100	0,05	39,84393474	0,970117386	0,97335702	0,6377843	162,48
13	0,5	0,85	1	1	0,05	39,84393474	0,970122352	0,97336311	0,63781341	158,93
14	0,5	0,85	1	100	0,05	39,84393474	0,970117386	0,97335702	0,63780176	158,08
15	0,5	0,85	100	1	0,05	40,52967605	0,970620913	0,97367665	0,63992525	157,41
16	0,5	0,85	100	100	0,05	39,84393474	0,970117386	0,97335702	0,63780692	160,3
17	0,5	0,85	500	1	0,05	39,84393474	0,970156205	0,97343763	0,63781341	158,48
18	0,5	0,85	500	100	0,05	39,84393474	0,970122352	0,97336311	0,63765494	165,05
19	0,75	0,65	1	1	0,05	39,84393474	0,970122352	0,97336311	0,63781341	161,36
20	0,75	0,65	1	100	0,05	39,98581225	0,970798584	0,97404408	0,63832781	159,2
21	0,75	0,65	100	1	0,05	39,84393474	0,970099479	0,97333339	0,63846783	160,92
22	0,75	0,65	100	100	0,05	39,84393474	0,970122352	0,97336311	0,63780924	159,63
23	0,75	0,65	500	1	0,05	39,84393474	0,970393255	0,97372403	0,63780467	156,91
24	0,75	0,65	500	100	0,05	39,84393474	0,970110804	0,97334963	0,63835552	160,34
25	0,75	0,35	1	1	0,05	39,84393474	0,970122352	0,97336311	0,63756801	161,22
26	0,75	0,35	1	100	0,05	39,84393474	0,970101805	0,97333709	0,63831197	158,08
27	0,75	0,35	100	1	0,05	39,84393474	0,970117386	0,97335702	0,63779653	160,54
28	0,75	0,35	100	100	0,05	39,84393474	0,970101805	0,97333709	0,63832548	158,67
29	0,75	0,35	500	1	0,05	39,84393474	0,970101805	0,97333709	0,63830905	159,41
30	0,75	0,35	500	100	0,05	39,962166	0,970142692	0,97336055	0,63823922	159,75
31	0,75	0,85	1	1	0,05	39,84393474	0,970117386	0,97335702	0,63781341	159,67
32	0,75	0,85	1	100	0,05	39,84393474	0,970122352	0,97336311	0,6378105	160,82
33	0,75	0,85	100	1	0,05	39,84393474	0,970300465	0,97365016	0,6383424	159,5

34	0,75	0,85	100	100	0,05	40,08039726	0,970153353	0,97338999	0,63885326	160,98
35	0,75	0,85	500	1	0,05	39,84393474	0,970122352	0,97336311	0,63781341	160,9
36	0,75	0,85	500	100	0,05	39,84393474	0,970117386	0,97335702	0,63781341	159,92
37	0,25	0,65	1	1	0,05	39,89122724	0,970121868	0,97336828	0,63780322	161,33
38	0,25	0,65	1	100	0,05	39,84393474	0,970106398	0,97334318	0,63831615	158,4
39	0,25	0,65	100	1	0,05	39,84393474	0,970106398	0,97334318	0,63831615	160,12
40	0,25	0,65	100	100	0,05	39,84393474	0,970106398	0,97334318	0,63831323	158,66
41	0,25	0,65	500	1	0,05	39,84393474	0,970117386	0,97335702	0,63781341	160,75
42	0,25	0,65	500	100	0,05	39,84393474	0,970211677	0,97345281	0,63822318	161,07
43	0,25	0,35	1	1	0,05	39,84393474	0,970229835	0,97353188	0,6378733	161,23
44	0,25	0,35	1	100	0,05	39,84393474	0,970117386	0,97335702	0,63781341	158,96
45	0,25	0,35	100	1	0,05	39,84393474	0,970122352	0,97336311	0,63781341	162,31
46	0,25	0,35	100	100	0,05	39,84393474	0,970117386	0,97335702	0,63779944	161,77
47	0,25	0,35	500	1	0,05	39,84393474	0,970117386	0,97335702	0,63777867	161,7
48	0,25	0,35	500	100	0,05	39,84393474	0,970143858	0,97340592	0,63786437	157,26
49	0,25	0,85	1	1	0,05	39,84393474	0,970122352	0,97336311	0,63780692	176,12
50	0,25	0,85	1	100	0,05	39,84393474	0,970161171	0,97344372	0,63772207	158,95
51	0,25	0,85	100	1	0,05	39,84393474	0,970101805	0,97333709	0,63832781	160,52
52	0,25	0,85	100	100	0,05	39,84393474	0,970122352	0,97336311	0,63770752	161,79
53	0,25	0,85	500	1	0,05	39,84393474	0,970106398	0,97334318	0,63830234	160,06
54	0,25	0,85	500	100	0,05	39,84393474	0,970101805	0,97333709	0,63832781	157,46
55	0.52	0.46	296	72	0.06	39,84393474	0,970122352	0,97336311	0,63781341	227,17
56	0.64	0.60	500	13	0.02	39,84393474	0,970117386	0,97335702	0,63780031	217,35
57	0.46	0.37	231	200	0.07	39,84393474	0,970101805	0,97333709	0,63832363	216,43
58	0.33	0.75	290	37	0.08	39,84393474	0,970117386	0,97335702	0,63763893	220,39
59	0.78	0.56	174	121	0.09	39,84393474	0,970122352	0,97336311	0,63779944	223,63
60	0.67	0.27	282	109	0.05	39,84393474	0,970106398	0,97334318	0,63830905	222,07
61	0.69	0.29	43	82	0.02	39,84393474	0,970117386	0,97335702	0,63781341	217,17
62	0.69	0.24	274	58	0.05	39,84393474	0,970117386	0,97335702	0,63765348	219,11
63	0.75	0.22	280	78	0.08	39,84393474	0,970122352	0,97336311	0,63779012	218,76
64	0.52	0.29	472	161	0.09	39,84393474	0,970122352	0,97336311	0,63781341	220,28

Tabela 28 – Métricas de qualidade log de eventos $\mathcal{L}_A^1$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	1,78624591	0,945553635	0,94329636	0,47378058	232,85
2	0,5	0,65	1	100	0,05	1,72670438	0,94309363	0,94024822	0,49777523	193,56
3	0,5	0,65	100	1	0,05	1,54807979	0,941828948	0,93962645	0,47977724	210,49
4	0,5	0,65	100	100	0,05	1,60762132	0,942200875	0,9402235	0,47633394	230,16
5	0,5	0,65	500	1	0,05	1,54807979	0,940085598	0,93765358	0,50071824	196,96
6	0,5	0,65	500	100	0,05	1,54807979	0,939837113	0,93729735	0,50227882	208,07
7	0,5	0,35	1	1	0,05	1,57785055	0,9421031	0,94010034	0,47666613	220,27
8	0,5	0,35	1	100	0,05	1,60762132	0,942200238	0,94022606	0,47580378	236,73
9	0,5	0,35	100	1	0,05	1,69693361	0,943234662	0,94045119	0,497469	153,75
10	0,5	0,35	100	100	0,05	1,57785055	0,942315675	0,94040638	0,47665475	168,01
11	0,5	0,35	500	1	0,05	1,69693361	0,943241187	0,94046821	0,49760612	153,59
12	0,5	0,35	500	100	0,05	1,51830902	0,941923862	0,93975376	0,4792378	171,14
13	0,5	0,85	1	1	0,05	1,66716285	0,942831777	0,93978466	0,50079828	148,7
14	0,5	0,85	1	100	0,05	1,69693361	0,94301207	0,94013875	0,49806026	155,16
15	0,5	0,85	100	1	0,05	1,57785055	0,939938462	0,93744525	0,50148896	157,66
16	0,5	0,85	100	100	0,05	1,72670438	0,943300886	0,94053273	0,49756897	151,44
17	0,5	0,85	500	1	0,05	1,54807979	0,939540417	0,93678338	0,50450846	147,6
18	0,5	0,85	500	100	0,05	1,75647514	0,945264125	0,94290025	0,47456516	175,71
19	0,75	0,65	1	1	0,05	1,75647514	0,945266748	0,94290679	0,4747638	179,31
20	0,75	0,65	1	100	0,05	1,66716285	0,942829519	0,93977668	0,50089786	143,55
21	0,75	0,65	100	1	0,05	1,60762132	0,942192549	0,94020404	0,47661059	183,33
22	0,75	0,65	100	100	0,05	1,57785055	0,940043241	0,93753529	0,50142156	156,8
23	0,75	0,65	500	1	0,05	1,78624591	0,945566731	0,94333409	0,47379387	179,51
24	0,75	0,65	500	100	0,05	1,75647514	0,945271013	0,94291297	0,47431974	169,62
25	0,75	0,35	1	1	0,05	1,69693361	0,942954741	0,93998448	0,50065761	145,31
26	0,75	0,35	1	100	0,05	1,69693361	0,943032276	0,94017648	0,49801697	154,69
27	0,75	0,35	100	1	0,05	1,69693361	0,94490556	0,9423374	0,47736442	164,89
28	0,75	0,35	100	100	0,05	1,54807979	0,939558278	0,93681912	0,50375828	150,1
29	0,75	0,35	500	1	0,05	1,60762132	0,942181033	0,94018304	0,47657207	176,27
30	0,75	0,35	500	100	0,05	1,69693361	0,942998276	0,94010478	0,49760093	143,27
31	0,75	0,85	1	1	0,05	1,54807979	0,939786793	0,93713345	0,50388739	153,51
32	0,75	0,85	1	100	0,05	1,72670438	0,943109163	0,94027709	0,49789723	150,34
33	0,75	0,85	100	1	0,05	1,60762132	0,942186694	0,94019021	0,47716106	173,86

34	0,75	0,85	100	100	0,05	1,60762132	0,942206293	0,94023613	0,47618122	166,15
35	0,75	0,85	500	1	0,05	1,78624591	0,945553323	0,94329595	0,47352967	173,09
36	0,75	0,85	500	100	0,05	1,75647514	0,945278996	0,94293486	0,47311097	168,84
37	0,25	0,65	1	1	0,05	1,60762132	0,942198298	0,9402193	0,4766028	216,53
38	0,25	0,65	1	100	0,05	1,54807979	0,941831368	0,93962815	0,47974463	176
39	0,25	0,65	100	1	0,05	1,54807979	0,939547634	0,9368021	0,50453817	150,76
40	0,25	0,65	100	100	0,05	1,72670438	0,943102595	0,94026933	0,49780833	144,99
41	0,25	0,65	500	1	0,05	1,57785055	0,940172824	0,93777145	0,49945155	162,18
42	0,25	0,65	500	100	0,05	1,66716285	0,94263049	0,9395168	0,50198423	144,83
43	0,25	0,35	1	1	0,05	1,66716285	0,94284904	0,93982582	0,49980027	146,57
44	0,25	0,35	1	100	0,05	1,54807979	0,939546963	0,93679465	0,50481118	148,35
45	0,25	0,35	100	1	0,05	1,66716285	0,942646965	0,93955324	0,50145599	149,44
46	0,25	0,35	100	100	0,05	1,72670438	0,943097146	0,94025671	0,4979059	150,62
47	0,25	0,35	500	1	0,05	1,72670438	0,943099975	0,94026445	0,49781031	162,55
48	0,25	0,35	500	100	0,05	1,69693361	0,942523439	0,94045576	0,47989097	166,8
49	0,25	0,85	1	1	0,05	1,72670438	0,943077445	0,94020799	0,49819129	147,9
50	0,25	0,85	1	100	0,05	1,51830902	0,939481667	0,93671679	0,50500757	148,72
51	0,25	0,85	100	1	0,05	1,60762132	0,942393316	0,94049502	0,47509974	185,25
52	0,25	0,85	100	100	0,05	1,66716285	0,942628369	0,93950928	0,50219496	145,27
53	0,25	0,85	500	1	0,05	1,75647514	0,945257261	0,94288249	0,47382693	177,61
54	0,25	0,85	500	100	0,05	1,57785055	0,942097268	0,94008532	0,4768693	168,05
55	0.52	0.46	296	72	0.06	1,57785055	0,940179569	0,93778681	0,50058107	202,16
56	0.64	0.60	500	13	0.02	1,57785055	0,940166599	0,93775942	0,50056621	202,29
57	0.46	0.37	231	200	0.07	1,54807979	0,93955414	0,93681305	0,50449949	190,08
58	0.33	0.75	290	37	0.08	1,72670438	0,943106126	0,94026993	0,49819446	204,85
59	0.78	0.56	174	121	0.09	1,57785055	0,93993695	0,93743696	0,50115002	194,46
60	0.67	0.27	282	109	0.05	1,54807979	0,939812656	0,93716037	0,50401059	185,89
61	0.69	0.29	43	82	0.02	1,69693361	0,943224959	0,94043362	0,49754078	196,75
62	0.69	0.24	274	58	0.05	1,57785055	0,939933088	0,93742821	0,5011317	192,51
63	0.75	0.22	280	78	0.08	1,57785055	0,940148256	0,93772912	0,50138151	207,89
64	0.52	0.29	472	161	0.09	1,72670438	0,943309775	0,94055214	0,49735635	202,58

Tabela 29 – Métricas de qualidade log de eventos $\mathcal{L}_A^2$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	15,7977883	0,942459302	0,94078682	0,76730843	60,4
2	0,5	0,65	1	100	0,05	15,7977883	0,942501217	0,94082844	0,76717521	65,87
3	0,5	0,65	100	1	0,05	15,7977883	0,942495337	0,94081792	0,76699633	61,74
4	0,5	0,65	100	100	0,05	15,7977883	0,942468593	0,94080844	0,76711561	64,66
5	0,5	0,65	500	1	0,05	15,7977883	0,942503896	0,9408302	0,76718631	61,43
6	0,5	0,65	500	100	0,05	15,7977883	0,942501217	0,94082844	0,76718076	62,76
7	0,5	0,35	1	1	0,05	15,7977883	0,942498016	0,94081968	0,76699078	64,65
8	0,5	0,35	1	100	0,05	15,7977883	0,942504262	0,94083487	0,76700188	63,47
9	0,5	0,35	100	1	0,05	15,7977883	0,942461981	0,94078858	0,7673001	63,69
10	0,5	0,35	100	100	0,05	15,7977883	0,942500661	0,94082494	0,76701574	62,92
11	0,5	0,35	500	1	0,05	15,7977883	0,942494971	0,94081326	0,76718354	61,89
12	0,5	0,35	500	100	0,05	15,7977883	0,942464992	0,94079851	0,76714058	62,81
13	0,5	0,85	1	1	0,05	15,7977883	0,942461981	0,94078858	0,76729455	64,91
14	0,5	0,85	1	100	0,05	15,7977883	0,942462347	0,94079325	0,76712116	64,36
15	0,5	0,85	100	1	0,05	15,7977883	0,942498877	0,94082259	0,76700188	58,89
16	0,5	0,85	100	100	0,05	15,7977883	0,942464992	0,94079851	0,76713503	60,04
17	0,5	0,85	500	1	0,05	15,7977883	0,942494971	0,94081326	0,76718631	64,67
18	0,5	0,85	500	100	0,05	15,7977883	0,942471238	0,9408137	0,76714613	61,46
19	0,75	0,65	1	1	0,05	15,7977883	0,942492326	0,940808	0,76715856	60,43
20	0,75	0,65	1	100	0,05	15,7977883	0,942459302	0,94078682	0,76731121	63,09
21	0,75	0,65	100	1	0,05	15,7977883	0,942461981	0,94078858	0,76730565	62,96
22	0,75	0,65	100	100	0,05	15,7977883	0,94249765	0,94081502	0,76718354	62,12
23	0,75	0,65	500	1	0,05	15,7977883	0,942461981	0,94078858	0,76730565	61,11
24	0,75	0,65	500	100	0,05	15,7977883	0,942471238	0,9408137	0,76714058	64,49
25	0,75	0,35	1	1	0,05	15,7977883	0,942504229	0,94083837	0,76701297	63,54
26	0,75	0,35	1	100	0,05	15,7977883	0,942495005	0,94080976	0,76716689	65,19
27	0,75	0,35	100	1	0,05	15,7977883	0,942462347	0,94079325	0,76711838	64,51
28	0,75	0,35	100	100	0,05	15,7977883	0,942501583	0,9408331	0,76700188	65,24
29	0,75	0,35	500	1	0,05	15,7977883	0,942492326	0,940808	0,76716134	61,4
30	0,75	0,35	500	100	0,05	15,7977883	0,942462347	0,94079325	0,76712671	62
31	0,75	0,85	1	1	0,05	15,7977883	0,942461981	0,94078858	0,7673001	61,09
32	0,75	0,85	1	100	0,05	15,7977883	0,942465582	0,94079851	0,76729178	64,52
33	0,75	0,85	100	1	0,05	15,7977883	0,942462903	0,94079675	0,76729178	63,36

34	0,75	0,85	100	100	0,05	15,7977883	0,942462347	0,94079325	0,76712116	64,41
35	0,75	0,85	500	1	0,05	15,7977883	0,942459336	0,94078332	0,76729178	60,65
36	0,75	0,85	500	100	0,05	15,7977883	0,942501583	0,9408331	0,76699633	62,37
37	0,25	0,65	1	1	0,05	15,7977883	0,942501251	0,94082494	0,76715579	61,85
38	0,25	0,65	1	100	0,05	15,7977883	0,942500661	0,94082494	0,76701852	62,33
39	0,25	0,65	100	1	0,05	15,7977883	0,942494971	0,94081326	0,76718354	60,69
40	0,25	0,65	100	100	0,05	15,7977883	0,942465915	0,94080667	0,76712116	66,56
41	0,25	0,65	500	1	0,05	15,7977883	0,942462347	0,94079325	0,76712393	66,32
42	0,25	0,65	500	100	0,05	15,7977883	0,942495005	0,94080976	0,76716134	63,97
43	0,25	0,35	1	1	0,05	15,7977883	0,942500661	0,94082494	0,76701852	59,2
44	0,25	0,35	1	100	0,05	15,7977883	0,942501895	0,94084643	0,76715445	86,33
45	0,25	0,35	100	1	0,05	15,7977883	0,942468593	0,94080844	0,76712116	83,81
46	0,25	0,35	100	100	0,05	15,7977883	0,942471238	0,9408137	0,76714058	86,02
47	0,25	0,35	500	1	0,05	15,7977883	0,942459302	0,94078682	0,76730565	62,26
48	0,25	0,35	500	100	0,05	15,7977883	0,942492326	0,940808	0,76716134	64,77
49	0,25	0,85	1	1	0,05	15,7977883	0,942465582	0,94079851	0,76728345	63,84
50	0,25	0,85	1	100	0,05	15,7977883	0,942468593	0,94080844	0,76712393	64,11
51	0,25	0,85	100	1	0,05	15,7977883	0,942498016	0,94081968	0,76699633	73,75
52	0,25	0,85	100	100	0,05	15,7977883	0,942558001	0,94090381	0,76701297	63,91
53	0,25	0,85	500	1	0,05	15,7977883	0,942498572	0,94082318	0,76716134	61,7
54	0,25	0,85	500	100	0,05	15,7977883	0,942501217	0,94082844	0,76718076	61,86
55	0.52	0.46	296	72	0.06	15,7977883	0,942459336	0,94078332	0,76729178	77,71
56	0.64	0.60	500	13	0.02	15,7977883	0,94249765	0,94081502	0,76718631	80,16
57	0.46	0.37	231	200	0.07	15,7977883	0,942498016	0,94081968	0,76699633	76,72
58	0.33	0.75	290	37	0.08	15,7977883	0,942501583	0,9408331	0,76699633	82,07
59	0.78	0.56	174	121	0.09	15,7977883	0,94249765	0,94081502	0,76718631	79,6
60	0.67	0.27	282	109	0.05	15,7977883	0,942461981	0,94078858	0,7673001	78,01
61	0.69	0.29	43	82	0.02	15,7977883	0,942503896	0,9408302	0,76718631	78,99
62	0.69	0.24	274	58	0.05	15,7977883	0,942495337	0,94081792	0,76699356	86,41
63	0.75	0.22	280	78	0.08	15,7977883	0,942495337	0,94081792	0,76699078	86,5
64	0.52	0.29	472	161	0.09	15,7977883	0,94249765	0,94081502	0,76718631	82,99

Tabela 30 – Métricas de qualidade log de eventos $\mathcal{L}_A^3$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,78
2	0,5	0,65	1	100	0,05	60,1679817	0,977104422	0,97624745	0,81135969	49,17
3	0,5	0,65	100	1	0,05	60,1679817	0,977104422	0,97624745	0,81138852	50,37
4	0,5	0,65	100	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,12
5	0,5	0,65	500	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,09
6	0,5	0,65	500	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,32
7	0,5	0,35	1	1	0,05	60,1679817	0,977169674	0,97632613	0,8113298	50,13
8	0,5	0,35	1	100	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,43
9	0,5	0,35	100	1	0,05	60,1679817	0,977104422	0,97624745	0,81135969	49,9
10	0,5	0,35	100	100	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,36
11	0,5	0,35	500	1	0,05	60,1679817	0,977139009	0,97629318	0,81136405	49,82
12	0,5	0,35	500	100	0,05	60,1679817	0,977104422	0,97624745	0,81134047	49,82
13	0,5	0,85	1	1	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,45
14	0,5	0,85	1	100	0,05	60,1679817	0,977104422	0,97624745	0,81138852	50,03
15	0,5	0,85	100	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,41
16	0,5	0,85	100	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,87
17	0,5	0,85	500	1	0,05	60,1679817	0,977139009	0,97629318	0,81136405	50,39
18	0,5	0,85	500	100	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,15
19	0,75	0,65	1	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,61
20	0,75	0,65	1	100	0,05	60,1679817	0,977104422	0,97624745	0,81134047	49,7
21	0,75	0,65	100	1	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,03
22	0,75	0,65	100	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,68
23	0,75	0,65	500	1	0,05	60,1679817	0,977104422	0,97624745	0,81135969	49,98
24	0,75	0,65	500	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,77
25	0,75	0,35	1	1	0,05	60,1679817	0,977104422	0,97624745	0,81137472	49,89
26	0,75	0,35	1	100	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,05
27	0,75	0,35	100	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,58
28	0,75	0,35	100	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,25
29	0,75	0,35	500	1	0,05	60,1679817	0,977104422	0,97624745	0,81134047	50,38
30	0,75	0,35	500	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,43
31	0,75	0,85	1	1	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,39
32	0,75	0,85	1	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,46
33	0,75	0,85	100	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,95

34	0,75	0,85	100	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,07
35	0,75	0,85	500	1	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,13
36	0,75	0,85	500	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,16
37	0,25	0,65	1	1	0,05	60,1679817	0,977104422	0,97624745	0,81138852	50,62
38	0,25	0,65	1	100	0,05	60,1679817	0,977104422	0,97624745	0,81134589	50,01
39	0,25	0,65	100	1	0,05	60,6770171	0,977056581	0,97614603	0,81255672	49,57
40	0,25	0,65	100	100	0,05	60,1679817	0,977104422	0,97624745	0,81138852	50,59
41	0,25	0,65	500	1	0,05	60,1679817	0,977139009	0,97629318	0,81135863	53,13
42	0,25	0,65	500	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,22
43	0,25	0,35	1	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,45
44	0,25	0,35	1	100	0,05	60,1679817	0,977139009	0,97629318	0,81135863	50,46
45	0,25	0,35	100	1	0,05	60,1679817	0,977104422	0,97624745	0,81134047	49,39
46	0,25	0,35	100	100	0,05	60,1679817	0,977104422	0,97624745	0,81138852	50,2
47	0,25	0,35	500	1	0,05	60,1679817	0,977104422	0,97624745	0,81135969	50,19
48	0,25	0,35	500	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	49,94
49	0,25	0,85	1	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	64,71
50	0,25	0,85	1	100	0,05	60,1679817	0,977143386	0,97629872	0,81135863	50,29
51	0,25	0,85	100	1	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,31
52	0,25	0,85	100	100	0,05	60,1679817	0,977104422	0,97624745	0,8113693	50,13
53	0,25	0,85	500	1	0,05	60,6770171	0,977056581	0,97614603	0,812576	48,63
54	0,25	0,85	500	100	0,05	60,1679817	0,977139009	0,97629318	0,81136405	50,01
55	0.52	0.46	296	72	0.06	60,1679817	0,977104422	0,97624745	0,81138852	65,9
56	0.64	0.60	500	13	0.02	60,1679817	0,977139009	0,97629318	0,8113298	68,09
57	0.46	0.37	231	200	0.07	60,1679817	0,977139009	0,97629318	0,8113298	67,99
58	0.33	0.75	290	37	0.08	60,1679817	0,977104422	0,97624745	0,81139394	66,98
59	0.78	0.56	174	121	0.09	60,1679817	0,977104422	0,97624745	0,81134047	66,56
60	0.67	0.27	282	109	0.05	60,1679817	0,977139009	0,97629318	0,81133522	68,11
61	0.69	0.29	43	82	0.02	60,1679817	0,977139009	0,97629318	0,81135863	69,23
62	0.69	0.24	274	58	0.05	60,1679817	0,977104422	0,97624745	0,8113693	70,4
63	0.75	0.22	280	78	0.08	60,1679817	0,977139009	0,97629318	0,8113298	69,43
64	0.52	0.29	472	161	0.09	60,1679817	0,977139009	0,97629318	0,81135863	68,75

Tabela 31 – Métricas de qualidade log de eventos $\mathcal{L}_F^1$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	1,220601369	0,961374563	0,95787165	0,63538424	119,36
2	0,5	0,65	1	100	0,05	1,309913665	0,959188172	0,95506902	0,68008354	92,66
3	0,5	0,65	100	1	0,05	1,309913665	0,960747329	0,95701314	0,65575128	115,91
4	0,5	0,65	100	100	0,05	1,2801429	0,960208108	0,95657628	0,65691768	98,53
5	0,5	0,65	500	1	0,05	1,33968443	0,959341715	0,95529457	0,68034776	88,64
6	0,5	0,65	500	100	0,05	1,2801429	0,960219073	0,95653251	0,65642075	98,66
7	0,5	0,35	1	1	0,05	1,309913665	0,960898034	0,95723378	0,65553921	119,4
8	0,5	0,35	1	100	0,05	1,309913665	0,960911828	0,95729646	0,65483586	113,39
9	0,5	0,35	100	1	0,05	1,220601369	0,961520464	0,95810977	0,63471308	96,65
10	0,5	0,35	100	100	0,05	1,33968443	0,95937047	0,95534148	0,67998721	74,45
11	0,5	0,35	500	1	0,05	1,2801429	0,96006848	0,95632658	0,65679729	77,5
12	0,5	0,35	500	100	0,05	1,250372135	0,961707997	0,95840873	0,63342797	95,26
13	0,5	0,85	1	1	0,05	1,309913665	0,960753515	0,95706996	0,65408784	91,1
14	0,5	0,85	1	100	0,05	1,2801429	0,960065143	0,95631762	0,65699103	77,36
15	0,5	0,85	100	1	0,05	1,33968443	0,959175605	0,95501932	0,68016614	74,52
16	0,5	0,85	100	100	0,05	1,309913665	0,959036792	0,95477182	0,68065309	70,6
17	0,5	0,85	500	1	0,05	1,33968443	0,95936692	0,95533205	0,68001512	73,26
18	0,5	0,85	500	100	0,05	1,309913665	0,960945058	0,9573459	0,6548944	93,78
19	0,75	0,65	1	1	0,05	1,250372135	0,961653327	0,95828929	0,6349746	94,52
20	0,75	0,65	1	100	0,05	1,220601369	0,961473345	0,95808429	0,63484287	95,5
21	0,75	0,65	100	1	0,05	1,250372135	0,961503904	0,95808193	0,63518654	95,75
22	0,75	0,65	100	100	0,05	1,2801429	0,960287439	0,95663316	0,65678516	78,57
23	0,75	0,65	500	1	0,05	1,309913665	0,959227493	0,95508438	0,68053235	72,36
24	0,75	0,65	500	100	0,05	1,309913665	0,959159805	0,95504708	0,6802011	72,67
25	0,75	0,35	1	1	0,05	1,250372135	0,961688311	0,95831317	0,63474083	95,43
26	0,75	0,35	1	100	0,05	1,250372135	0,961673404	0,95836147	0,63391397	100,08
27	0,75	0,35	100	1	0,05	1,33968443	0,959180147	0,95504374	0,68038807	73,98
28	0,75	0,35	100	100	0,05	1,220601369	0,961324961	0,95780841	0,63560917	95,07
29	0,75	0,35	500	1	0,05	1,2801429	0,960927481	0,95731442	0,65637762	91,21
30	0,75	0,35	500	100	0,05	1,33968443	0,959158452	0,9549936	0,68052993	72,06
31	0,75	0,85	1	1	0,05	1,309913665	0,959208013	0,95503639	0,68064047	73,65
32	0,75	0,85	1	100	0,05	1,33968443	0,959334854	0,95532957	0,68001605	74,78
33	0,75	0,85	100	1	0,05	1,33968443	0,959189784	0,95506412	0,68011214	71,6

34	0,75	0,85	100	100	0,05	1,33968443	0,959169375	0,95503044	0,6805735	70,83
35	0,75	0,85	500	1	0,05	1,309913665	0,960927946	0,9572973	0,65499785	91,1
36	0,75	0,85	500	100	0,05	1,309913665	0,960743966	0,95702979	0,65616865	89,02
37	0,25	0,65	1	1	0,05	1,250372135	0,961508603	0,95805251	0,63520026	101,82
38	0,25	0,65	1	100	0,05	1,33968443	0,959179481	0,955043	0,68026277	78,51
39	0,25	0,65	100	1	0,05	1,220601369	0,961337963	0,95784331	0,634126	96,32
40	0,25	0,65	100	100	0,05	1,220601369	0,961350807	0,95782052	0,6350372	96,45
41	0,25	0,65	500	1	0,05	1,309913665	0,959163754	0,95505884	0,68023323	74,19
42	0,25	0,65	500	100	0,05	1,2801429	0,960573865	0,95680321	0,656152	90,66
43	0,25	0,35	1	1	0,05	1,250372135	0,961562537	0,95816806	0,6344142	94,11
44	0,25	0,35	1	100	0,05	1,309913665	0,959223961	0,95505834	0,68006398	74,23
45	0,25	0,35	100	1	0,05	1,2801429	0,960089951	0,95631279	0,65743497	77,96
46	0,25	0,35	100	100	0,05	1,250372135	0,960090225	0,95633209	0,65692321	81,06
47	0,25	0,35	500	1	0,05	1,2801429	0,960575483	0,95680978	0,65610444	92,15
48	0,25	0,35	500	100	0,05	1,2801429	0,960220021	0,95655293	0,65720654	77,77
49	0,25	0,85	1	1	0,05	1,309913665	0,960766438	0,95705894	0,65547976	90
50	0,25	0,85	1	100	0,05	1,309913665	0,96075391	0,95707272	0,65552896	94,17
51	0,25	0,85	100	1	0,05	1,309913665	0,960905402	0,95729276	0,65491581	93,11
52	0,25	0,85	100	100	0,05	1,220601369	0,961667938	0,95834494	0,63370235	93,45
53	0,25	0,85	500	1	0,05	1,2801429	0,960068558	0,95630363	0,65759085	78,67
54	0,25	0,85	500	100	0,05	1,309913665	0,960908547	0,95724733	0,65548155	92,66
55	0.52	0.46	296	72	0.06	1,33968443	0,959361834	0,95532493	0,68032892	98,74
56	0.64	0.60	500	13	0.02	1,309913665	0,959202173	0,95506847	0,6805123	99,65
57	0.46	0.37	231	200	0.07	1,250372135	0,961523749	0,95807468	0,63362545	130
58	0.33	0.75	290	37	0.08	1,2801429	0,960205557	0,95655646	0,65667387	106,17
59	0.78	0.56	174	121	0.09	1,220601369	0,961350769	0,95782126	0,63475931	127,9
60	0.67	0.27	282	109	0.05	1,33968443	0,959183199	0,95505437	0,68046073	98,26
61	0.69	0.29	43	82	0.02	1,250372135	0,961649179	0,95833659	0,63456862	128,75
62	0.69	0.24	274	58	0.05	1,250372135	0,961531275	0,95814314	0,63508685	131,84
63	0.75	0.22	280	78	0.08	1,309913665	0,960890477	0,95727038	0,65468739	125,32
64	0.52	0.29	472	161	0.09	1,309913665	0,960842355	0,9572216	0,65396999	120,79

Tabela 32 – Métricas de qualidade log de eventos $\mathcal{L}_F^2$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	44,1548183	0,951793007	0,95378723	0,85255322	13,11
2	0,5	0,65	1	100	0,05	44,1548183	0,951696535	0,95371241	0,85259931	13,58
3	0,5	0,65	100	1	0,05	44,1548183	0,95167602	0,95376882	0,85264281	13,23
4	0,5	0,65	100	100	0,05	44,1548183	0,952009001	0,95411888	0,85255976	13,35
5	0,5	0,65	500	1	0,05	44,1548183	0,952028592	0,95413317	0,85256113	13,29
6	0,5	0,65	500	100	0,05	44,1548183	0,951381435	0,95342712	0,85261117	13,08
7	0,5	0,35	1	1	0,05	44,1548183	0,951355939	0,95341642	0,85261117	12,97
8	0,5	0,35	1	100	0,05	44,1548183	0,953752887	0,95588392	0,85255976	13,48
9	0,5	0,35	100	1	0,05	44,1548183	0,951200185	0,95321516	0,85254531	12,98
10	0,5	0,35	100	100	0,05	44,1548183	0,951826069	0,95389275	0,85254395	13,44
11	0,5	0,35	500	1	0,05	44,1548183	0,951447325	0,95342896	0,85260326	13,44
12	0,5	0,35	500	100	0,05	44,1548183	0,951392421	0,95352091	0,85255581	12,93
13	0,5	0,85	1	1	0,05	44,1548183	0,953678992	0,95575842	0,8524781	13,8
14	0,5	0,85	1	100	0,05	44,1548183	0,951672569	0,95368671	0,85253604	13,46
15	0,5	0,85	100	1	0,05	44,1548183	0,951353914	0,95339144	0,85254395	13,19
16	0,5	0,85	100	100	0,05	44,1548183	0,951251867	0,95325848	0,85260722	13,1
17	0,5	0,85	500	1	0,05	44,1548183	0,951839791	0,95390109	0,85255581	13,56
18	0,5	0,85	500	100	0,05	44,1548183	0,951249767	0,95317373	0,85260326	13,06
19	0,75	0,65	1	1	0,05	44,1548183	0,951475086	0,95342466	0,85247415	13,11
20	0,75	0,65	1	100	0,05	44,1548183	0,951501605	0,95359601	0,85247019	13,47
21	0,75	0,65	100	1	0,05	44,1548183	0,951521099	0,95352062	0,85261512	13,28
22	0,75	0,65	100	100	0,05	44,1548183	0,951570923	0,9535911	0,85248996	13,2
23	0,75	0,65	500	1	0,05	44,1548183	0,951633122	0,95365026	0,85256113	13,35
24	0,75	0,65	500	100	0,05	44,1548183	0,951551662	0,95369336	0,85261908	13,29
25	0,75	0,35	1	1	0,05	44,1548183	0,951707237	0,95372004	0,85247019	13,45
26	0,75	0,35	1	100	0,05	44,1548183	0,951714645	0,9537881	0,8525479	13,31
27	0,75	0,35	100	1	0,05	44,1548183	0,951394432	0,95342427	0,85254136	13,07
28	0,75	0,35	100	100	0,05	44,1548183	0,951180947	0,9531993	0,85255581	13,06
29	0,75	0,35	500	1	0,05	44,1548183	0,951451735	0,95340496	0,85253208	13,21
30	0,75	0,35	500	100	0,05	44,1548183	0,951770473	0,9538243	0,85261908	13,4
31	0,75	0,85	1	1	0,05	44,1548183	0,95177655	0,9537712	0,85261512	12,8
32	0,75	0,85	1	100	0,05	44,1548183	0,951288912	0,95334059	0,85254927	13,1
33	0,75	0,85	100	1	0,05	44,1548183	0,951555822	0,95369203	0,85255185	13,33

34	0,75	0,85	100	100	0,05	44,1548183	0,951381727	0,95342147	0,85248205	13,22
35	0,75	0,85	500	1	0,05	44,1548183	0,95175499	0,95394256	0,85255581	13,49
36	0,75	0,85	500	100	0,05	44,1548183	0,951844142	0,95390548	0,85255185	13,2
37	0,25	0,65	1	1	0,05	44,1548183	0,951268759	0,95332121	0,85259931	13,15
38	0,25	0,65	1	100	0,05	44,1548183	0,951632426	0,95365025	0,85248996	13,42
39	0,25	0,65	100	1	0,05	44,1548183	0,951312482	0,95321828	0,8525374	13,15
40	0,25	0,65	100	100	0,05	44,1548183	0,951364718	0,95340544	0,85260722	13,21
41	0,25	0,65	500	1	0,05	44,1548183	0,951560522	0,9536137	0,8524781	13,58
42	0,25	0,65	500	100	0,05	44,1548183	0,951557455	0,95354299	0,85262303	12,88
43	0,25	0,35	1	1	0,05	44,1548183	0,951385499	0,95342147	0,85248205	13,37
44	0,25	0,35	1	100	0,05	44,1548183	0,951624405	0,95358769	0,85261908	18,38
45	0,25	0,35	100	1	0,05	44,1548183	0,951256598	0,95316507	0,85260326	15
46	0,25	0,35	100	100	0,05	44,1548183	0,951606442	0,95362508	0,85253345	18,8
47	0,25	0,35	500	1	0,05	44,2338073	0,951380164	0,95339036	0,85275257	12,83
48	0,25	0,35	500	100	0,05	44,1548183	0,951702689	0,95370293	0,85255185	12,98
49	0,25	0,85	1	1	0,05	44,1548183	0,951636383	0,95360536	0,85262303	13,03
50	0,25	0,85	1	100	0,05	44,1548183	0,951445981	0,95354179	0,85254136	14,04
51	0,25	0,85	100	1	0,05	44,1548183	0,951827251	0,9538823	0,85255976	13,63
52	0,25	0,85	100	100	0,05	44,1548183	0,951624944	0,95359217	0,85261117	13,08
53	0,25	0,85	500	1	0,05	44,1548183	0,951531643	0,9536366	0,85247019	13,42
54	0,25	0,85	500	100	0,05	44,1548183	0,951614414	0,95361978	0,85256113	13,1
55	0.52	0.46	296	72	0.06	44,1548183	0,951783026	0,9537761	0,85248996	17,45
56	0.64	0.60	500	13	0.02	44,1548183	0,951666769	0,95368059	0,85259535	18,3
57	0.46	0.37	231	200	0.07	44,1548183	0,951245338	0,95315528	0,85253345	17,52
58	0.33	0.75	290	37	0.08	44,1548183	0,95148996	0,95347945	0,85255185	17,26
59	0.78	0.56	174	121	0.09	44,1548183	0,951535618	0,95361792	0,85261512	17,23
60	0.67	0.27	282	109	0.05	44,1548183	0,951790113	0,95378163	0,85256113	17,58
61	0.69	0.29	43	82	0.02	44,1548183	0,951711587	0,95372282	0,85254136	17,99
62	0.69	0.24	274	58	0.05	44,1548183	0,951145419	0,95316137	0,85254927	16,98
63	0.75	0.22	280	78	0.08	44,1548183	0,951564113	0,95355509	0,85248996	17,32
64	0.52	0.29	472	161	0.09	44,1548183	0,951805008	0,95380155	0,85255322	17,18

Tabela 33 – Métricas de qualidade log de eventos $\mathcal{L}_F^3$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,6
2	0,5	0,65	1	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,61
3	0,5	0,65	100	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	19,25
4	0,5	0,65	100	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,79
5	0,5	0,65	500	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,76
6	0,5	0,65	500	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,7
7	0,5	0,35	1	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,73
8	0,5	0,35	1	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,68
9	0,5	0,35	100	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,79
10	0,5	0,35	100	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,57
11	0,5	0,35	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,75
12	0,5	0,35	500	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,62
13	0,5	0,85	1	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	20,25
14	0,5	0,85	1	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,49
15	0,5	0,85	100	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,66
16	0,5	0,85	100	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,74
17	0,5	0,85	500	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,64
18	0,5	0,85	500	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,6
19	0,75	0,65	1	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,77
20	0,75	0,65	1	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,85
21	0,75	0,65	100	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,57
22	0,75	0,65	100	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,84
23	0,75	0,65	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,66
24	0,75	0,65	500	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,79
25	0,75	0,35	1	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,68
26	0,75	0,35	1	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	21,5
27	0,75	0,35	100	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,71
28	0,75	0,35	100	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,63
29	0,75	0,35	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,74
30	0,75	0,35	500	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,75
31	0,75	0,85	1	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,92
32	0,75	0,85	1	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,72
33	0,75	0,85	100	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,7

34	0,75	0,85	100	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,79
35	0,75	0,85	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,73
36	0,75	0,85	500	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,73
37	0,25	0,65	1	1	0,05	77,3856209	0,987499222	0,98747615	0,87169229	18,92
38	0,25	0,65	1	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,82
39	0,25	0,65	100	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,6
40	0,25	0,65	100	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,85
41	0,25	0,65	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,69
42	0,25	0,65	500	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,77
43	0,25	0,35	1	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,73
44	0,25	0,35	1	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,75
45	0,25	0,35	100	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,69
46	0,25	0,35	100	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,82
47	0,25	0,35	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,68
48	0,25	0,35	500	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,7
49	0,25	0,85	1	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	27,39
50	0,25	0,85	1	100	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,71
51	0,25	0,85	100	1	0,05	77,3856209	0,987361662	0,98737874	0,87169229	18,92
52	0,25	0,85	100	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,8
53	0,25	0,85	500	1	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,96
54	0,25	0,85	500	100	0,05	77,3856209	0,987327112	0,98737571	0,87169229	18,8
55	0.52	0.46	296	72	0.06	77,3856209	0,987327112	0,98737571	0,87169229	24,64
56	0.64	0.60	500	13	0.02	77,3856209	0,987361662	0,98737874	0,87169229	25,41
57	0.46	0.37	231	200	0.07	77,3856209	0,987327112	0,98737571	0,87169229	25,33
58	0.33	0.75	290	37	0.08	77,3856209	0,987327112	0,98737571	0,87169229	25,2
59	0.78	0.56	174	121	0.09	77,3856209	0,987327112	0,98737571	0,87169229	26,17
60	0.67	0.27	282	109	0.05	77,3856209	0,987361662	0,98737874	0,87169229	25,75
61	0.69	0.29	43	82	0.02	77,3856209	0,987327112	0,98737571	0,87169229	26,35
62	0.69	0.24	274	58	0.05	77,3856209	0,987327112	0,98737571	0,87169229	25,74
63	0.75	0.22	280	78	0.08	77,3856209	0,987361662	0,98737874	0,87169229	25,5
64	0.52	0.29	472	161	0.09	77,3856209	0,987327112	0,98737571	0,87169229	25,25

Tabela 34 – Métricas de qualidade log de eventos $\mathcal{L}_D^1$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	40,66686514	0,975332493	0,96963199	0,66462435	80,8
2	0,5	0,65	1	100	0,05	40,1905329	0,975058164	0,96931622	0,66432276	83,83
3	0,5	0,65	100	1	0,05	40,1905329	0,97501173	0,96922437	0,66464808	81,63
4	0,5	0,65	100	100	0,05	40,1905329	0,975194079	0,969513	0,66457181	83,36
5	0,5	0,65	500	1	0,05	40,1905329	0,974909713	0,96909964	0,66457843	82,72
6	0,5	0,65	500	100	0,05	40,1905329	0,97514643	0,96946246	0,65994432	87,03
7	0,5	0,35	1	1	0,05	40,1905329	0,975120934	0,96942544	0,65959158	83,8
8	0,5	0,35	1	100	0,05	40,16076213	0,974965425	0,96919207	0,66021291	68,51
9	0,5	0,35	100	1	0,05	40,1905329	0,975086194	0,96940151	0,66065302	66,3
10	0,5	0,35	100	100	0,05	40,16076213	0,974757513	0,96885665	0,66511094	64,32
11	0,5	0,35	500	1	0,05	40,16076213	0,97479945	0,96894215	0,66048835	68,94
12	0,5	0,35	500	100	0,05	40,1905329	0,974701891	0,9687376	0,66536404	64,11
13	0,5	0,85	1	1	0,05	40,1905329	0,975074471	0,96937992	0,66064967	66,73
14	0,5	0,85	1	100	0,05	40,16076213	0,974751909	0,96886612	0,66465453	65,56
15	0,5	0,85	100	1	0,05	40,16076213	0,975117029	0,96940789	0,65967416	69,29
16	0,5	0,85	100	100	0,05	40,1905329	0,974959717	0,96917595	0,66502461	64,71
17	0,5	0,85	500	1	0,05	40,1905329	0,974983983	0,96920508	0,66446345	65,45
18	0,5	0,85	500	100	0,05	40,1905329	0,974976556	0,96921042	0,66495169	64,35
19	0,75	0,65	1	1	0,05	40,16076213	0,974833293	0,96897522	0,66041808	67,49
20	0,75	0,65	1	100	0,05	40,1905329	0,975003721	0,96921189	0,66454644	63,75
21	0,75	0,65	100	1	0,05	40,16076213	0,974730299	0,96884088	0,66510923	68,45
22	0,75	0,65	100	100	0,05	40,1905329	0,974981467	0,96919861	0,66442624	64,02
23	0,75	0,65	500	1	0,05	40,1905329	0,974862447	0,96899149	0,66473779	63,59
24	0,75	0,65	500	100	0,05	40,22030366	0,975105616	0,9693861	0,66479386	65,62
25	0,75	0,35	1	1	0,05	40,16076213	0,974883001	0,96906372	0,66482928	65,34
26	0,75	0,35	1	100	0,05	40,1905329	0,97503653	0,9693022	0,66470399	64,81
27	0,75	0,35	100	1	0,05	40,1905329	0,975012536	0,96922801	0,66460404	64,2
28	0,75	0,35	100	100	0,05	40,1905329	0,97503749	0,96930184	0,66407735	65,16
29	0,75	0,35	500	1	0,05	40,16076213	0,97493941	0,96915966	0,66011444	67,02
30	0,75	0,35	500	100	0,05	40,1905329	0,975113964	0,9694459	0,65958975	68,97
31	0,75	0,85	1	1	0,05	40,1905329	0,97503904	0,96929625	0,66475313	66,11
32	0,75	0,85	1	100	0,05	40,1905329	0,975113844	0,96945812	0,65959642	67,3
33	0,75	0,85	100	1	0,05	40,1905329	0,975071967	0,96937546	0,66063963	67,08

34	0,75	0,85	100	100	0,05	40,1905329	0,974887861	0,9690178	0,66479207	65,63
35	0,75	0,85	500	1	0,05	40,16076213	0,974940522	0,9691706	0,66068261	66,22
36	0,75	0,85	500	100	0,05	40,1905329	0,975009462	0,96924202	0,66384431	65,85
37	0,25	0,65	1	1	0,05	40,33938672	0,97497174	0,96918357	0,66008283	68,17
38	0,25	0,65	1	100	0,05	40,16076213	0,974883308	0,96907201	0,66437504	66,25
39	0,25	0,65	100	1	0,05	40,16076213	0,974962744	0,96917998	0,66023268	69,2
40	0,25	0,65	100	100	0,05	34,3852337	0,971089664	0,96553247	0,68101127	60,61
41	0,25	0,65	500	1	0,05	40,1905329	0,974997455	0,96923537	0,66451422	65,3
42	0,25	0,65	500	100	0,05	40,1905329	0,975092856	0,96939732	0,66017825	67,69
43	0,25	0,35	1	1	0,05	40,1905329	0,975280222	0,96961897	0,66055776	66,74
44	0,25	0,35	1	100	0,05	40,1905329	0,975129249	0,96947933	0,65935305	67,96
45	0,25	0,35	100	1	0,05	40,16076213	0,974832608	0,96895937	0,6604178	67,24
46	0,25	0,35	100	100	0,05	40,16076213	0,974964878	0,96919179	0,66028953	67,22
47	0,25	0,35	500	1	0,05	40,1905329	0,975124028	0,96942106	0,66020165	69,09
48	0,25	0,35	500	100	0,05	40,1905329	0,975113266	0,96941034	0,66017166	67,22
49	0,25	0,85	1	1	0,05	43,88210777	0,975973123	0,97032281	0,66100096	66,06
50	0,25	0,85	1	100	0,05	40,16076213	0,974862719	0,96900277	0,66471065	63,32
51	0,25	0,85	100	1	0,05	40,1905329	0,975048644	0,9693145	0,66432953	64,74
52	0,25	0,85	100	100	0,05	40,1905329	0,975036554	0,96929098	0,66477012	64,71
53	0,25	0,85	500	1	0,05	40,33938672	0,975035164	0,96928084	0,66495485	65,01
54	0,25	0,85	500	100	0,05	40,1905329	0,975036315	0,969299	0,66417553	65,24
55	0.52	0.46	296	72	0.06	40,16076213	0,974864226	0,96905353	0,66482591	88,09
56	0.64	0.60	500	13	0.02	40,1905329	0,975028284	0,96928565	0,65951937	91,92
57	0.46	0.37	231	200	0.07	40,1905329	0,975121214	0,96943846	0,65966327	92,34
58	0.33	0.75	290	37	0.08	40,16076213	0,974878816	0,96907147	0,6644055	85,71
59	0.78	0.56	174	121	0.09	40,16076213	0,974982946	0,96925549	0,65983358	88,57
60	0.67	0.27	282	109	0.05	40,1905329	0,975135692	0,96947534	0,65947838	89,84
61	0.69	0.29	43	82	0.02	40,16076213	0,974752782	0,96886412	0,66466806	84,6
62	0.69	0.24	274	58	0.05	40,1905329	0,97486427	0,96902027	0,66410797	85,44
63	0.75	0.22	280	78	0.08	42,72104793	0,976059353	0,97089857	0,67216671	80,15
64	0.52	0.29	472	161	0.09	40,16076213	0,975008925	0,96927246	0,66002906	90,68

Tabela 35 – Métricas de qualidade log de eventos $\mathcal{L}_D^2$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	54,4233807	0,976327057	0,97740068	0,80490736	16,53
2	0,5	0,65	1	100	0,05	54,4233807	0,976308777	0,97737819	0,80489033	16,51
3	0,5	0,65	100	1	0,05	54,4233807	0,976305171	0,97737681	0,80484606	16,2
4	0,5	0,65	100	100	0,05	54,4233807	0,976322428	0,97740051	0,80471143	17,69
5	0,5	0,65	500	1	0,05	54,4233807	0,976313497	0,97739574	0,80475568	16,81
6	0,5	0,65	500	100	0,05	54,4233807	0,976311044	0,97737847	0,80493801	16,08
7	0,5	0,35	1	1	0,05	54,4233807	0,976311012	0,97739518	0,80475228	16,57
8	0,5	0,35	1	100	0,05	54,4233807	0,976322428	0,97740051	0,80470802	16,72
9	0,5	0,35	100	1	0,05	54,4497104	0,976343325	0,97742979	0,80483068	16,17
10	0,5	0,35	100	100	0,05	54,4233807	0,976326035	0,97740189	0,80477951	16,67
11	0,5	0,35	500	1	0,05	54,4233807	0,976307438	0,97737709	0,80488011	16,37
12	0,5	0,35	500	100	0,05	54,4233807	0,976298951	0,97736487	0,80488011	16,35
13	0,5	0,85	1	1	0,05	54,4233807	0,976311157	0,97738774	0,80475238	16,48
14	0,5	0,85	1	100	0,05	54,4233807	0,976298652	0,97736487	0,8049313	16,35
15	0,5	0,85	100	1	0,05	54,4233807	0,976299394	0,97737709	0,80475568	16,65
16	0,5	0,85	100	100	0,05	54,4233807	0,976305874	0,97736459	0,8048733	16,31
17	0,5	0,85	500	1	0,05	54,4233807	0,976284767	0,97734649	0,80491427	16,02
18	0,5	0,85	500	100	0,05	54,4233807	0,976323582	0,97738462	0,80495163	16,2
19	0,75	0,65	1	1	0,05	54,4233807	0,976307656	0,97737736	0,80486309	16,71
20	0,75	0,65	1	100	0,05	54,47604	0,976324168	0,97739397	0,80480334	16,61
21	0,75	0,65	100	1	0,05	54,4233807	0,976313642	0,97738829	0,8047694	16,5
22	0,75	0,65	100	100	0,05	54,4233807	0,976301104	0,97738214	0,80475578	16,41
23	0,75	0,65	500	1	0,05	54,4233807	0,976297305	0,97735265	0,80492108	16,21
24	0,75	0,65	500	100	0,05	54,4233807	0,976295788	0,97737571	0,80472504	16,55
25	0,75	0,35	1	1	0,05	54,4233807	0,97628567	0,97734705	0,80495855	16,31
26	0,75	0,35	1	100	0,05	54,4233807	0,976308705	0,97737047	0,80490065	16,19
27	0,75	0,35	100	1	0,05	54,4233807	0,976282064	0,97734567	0,80489043	16,34
28	0,75	0,35	100	100	0,05	54,4233807	0,976293553	0,97735871	0,80484946	16,44
29	0,75	0,35	500	1	0,05	54,4233807	0,97631402	0,9773801	0,80473876	16,77
30	0,75	0,35	500	100	0,05	54,4233807	0,976295788	0,97737571	0,80470462	16,84
31	0,75	0,85	1	1	0,05	54,4233807	0,976317248	0,97738967	0,80483409	16,33
32	0,75	0,85	1	100	0,05	54,4233807	0,976307405	0,97739381	0,8046842	16,57
33	0,75	0,85	100	1	0,05	54,4233807	0,976301104	0,97738214	0,80476259	16,3

34	0,75	0,85	100	100	0,05	54,4233807	0,976287001	0,97736349	0,80476599	16,61
35	0,75	0,85	500	1	0,05	54,4233807	0,976305874	0,97736459	0,80486649	16,43
36	0,75	0,85	500	100	0,05	54,4233807	0,976299394	0,97737709	0,80476589	16,2
37	0,25	0,65	1	1	0,05	54,4233807	0,976313642	0,97738829	0,80477961	16,38
38	0,25	0,65	1	100	0,05	54,4233807	0,976293335	0,97735844	0,80488011	16,08
39	0,25	0,65	100	1	0,05	54,4233807	0,97631119	0,97737102	0,80494152	16,04
40	0,25	0,65	100	100	0,05	54,4233807	0,976298869	0,97736514	0,8049313	16,25
41	0,25	0,65	500	1	0,05	54,4233807	0,976302476	0,97736652	0,80497218	16,07
42	0,25	0,65	500	100	0,05	54,4233807	0,976319976	0,97738324	0,80488011	16,26
43	0,25	0,35	1	1	0,05	54,4233807	0,976316505	0,97738065	0,8047694	16,42
44	0,25	0,35	1	100	0,05	54,4233807	0,976291068	0,97735816	0,80482903	20,57
45	0,25	0,35	100	1	0,05	54,4233807	0,976303223	0,97737102	0,8049279	17,65
46	0,25	0,35	100	100	0,05	54,4233807	0,976297055	0,97736909	0,80473876	19,57
47	0,25	0,35	500	1	0,05	54,4233807	0,97629954	0,97736964	0,80476599	16,66
48	0,25	0,35	500	100	0,05	54,4233807	0,97631119	0,97737102	0,80494493	16,13
49	0,25	0,85	1	1	0,05	54,4233807	0,976311262	0,97737874	0,80491757	16,8
50	0,25	0,85	1	100	0,05	54,5023697	0,976324705	0,9774017	0,80493801	16,36
51	0,25	0,85	100	1	0,05	54,4233807	0,976293553	0,97735871	0,80486309	16,31
52	0,25	0,85	100	100	0,05	54,4233807	0,976321315	0,97738434	0,80489373	16,39
53	0,25	0,85	500	1	0,05	54,4233807	0,976290851	0,97735789	0,80484606	16,58
54	0,25	0,85	500	100	0,05	54,5813586	0,976551185	0,97765829	0,80496196	16,2
55	0.52	0.46	296	72	0.06	54,4233807	0,976296384	0,97736459	0,80489043	21,54
56	0.64	0.60	500	13	0.02	54,4233807	0,976306995	0,97736542	0,80491076	21,55
57	0.46	0.37	231	200	0.07	54,4233807	0,976322428	0,97740051	0,80470802	21,72
58	0.33	0.75	290	37	0.08	54,4233807	0,976294674	0,97735954	0,80490055	21,54
59	0.78	0.56	174	121	0.09	54,4233807	0,976284767	0,97734649	0,8049313	21,45
60	0.67	0.27	282	109	0.05	54,4233807	0,976299991	0,97736597	0,80495855	21,3
61	0.69	0.29	43	82	0.02	54,4233807	0,976293553	0,97735871	0,80486309	21,63
62	0.69	0.24	274	58	0.05	54,4233807	0,97632355	0,97740134	0,80474547	22,27
63	0.75	0.22	280	78	0.08	54,4233807	0,976313642	0,97738829	0,80476599	21,79
64	0.52	0.29	472	161	0.09	54,4233807	0,976309447	0,97738269	0,80475908	22,06

Tabela 36 – Métricas de qualidade log de eventos $\mathcal{L}_D^3$

#	P1	P2	P3	P4	P5	MF1	MF2	MF3	MP	T (s)
1	0,5	0,65	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,39
2	0,5	0,65	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,38
3	0,5	0,65	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,87
4	0,5	0,65	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,57
5	0,5	0,65	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,56
6	0,5	0,65	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,52
7	0,5	0,35	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,47
8	0,5	0,35	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,38
9	0,5	0,35	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,51
10	0,5	0,35	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,33
11	0,5	0,35	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,35
12	0,5	0,35	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,24
13	0,5	0,85	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,54
14	0,5	0,85	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,45
15	0,5	0,85	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,46
16	0,5	0,85	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,34
17	0,5	0,85	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,48
18	0,5	0,85	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,3
19	0,75	0,65	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,27
20	0,75	0,65	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,45
21	0,75	0,65	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,48
22	0,75	0,65	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,37
23	0,75	0,65	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,24
24	0,75	0,65	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,4
25	0,75	0,35	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,46
26	0,75	0,35	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,38
27	0,75	0,35	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,5
28	0,75	0,35	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,47
29	0,75	0,35	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	22,48
30	0,75	0,35	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,6
31	0,75	0,85	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,44
32	0,75	0,85	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,37
33	0,75	0,85	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,23

34	0,75	0,85	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,23
35	0,75	0,85	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,48
36	0,75	0,85	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,37
37	0,25	0,65	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,37
38	0,25	0,65	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,44
39	0,25	0,65	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,58
40	0,25	0,65	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,33
41	0,25	0,65	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,44
42	0,25	0,65	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,6
43	0,25	0,35	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,48
44	0,25	0,35	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,42
45	0,25	0,35	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,33
46	0,25	0,35	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,47
47	0,25	0,35	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,7
48	0,25	0,35	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,49
49	0,25	0,85	1	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	24,25
50	0,25	0,85	1	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,38
51	0,25	0,85	100	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,41
52	0,25	0,85	100	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,24
53	0,25	0,85	500	1	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,41
54	0,25	0,85	500	100	0,05	76,542887	0,92197307	0,9854924	0,85003095	21,61
55	0.52	0.46	296	72	0.06	76,542887	0,92197307	0,9854924	0,85003095	27,91
56	0.64	0.60	500	13	0.02	76,542887	0,92197307	0,9854924	0,85003095	29,44
57	0.46	0.37	231	200	0.07	76,542887	0,92197307	0,9854924	0,85003095	28,68
58	0.33	0.75	290	37	0.08	76,542887	0,92197307	0,9854924	0,85003095	28,89
59	0.78	0.56	174	121	0.09	76,542887	0,92197307	0,9854924	0,85003095	32,41
60	0.67	0.27	282	109	0.05	76,542887	0,92197307	0,9854924	0,85003095	29,49
61	0.69	0.29	43	82	0.02	76,542887	0,92197307	0,9854924	0,85003095	29,12
62	0.69	0.24	274	58	0.05	76,542887	0,92197307	0,9854924	0,85003095	28,98
63	0.75	0.22	280	78	0.08	76,542887	0,92197307	0,9854924	0,85003095	28,51
64	0.52	0.29	472	161	0.09	76,542887	0,92197307	0,9854924	0,85003095	29,41

Para alguns logs de eventos avaliados pode-se observar uma diferença na duração média do cálculo das métricas de qualidade registrado na coluna T(s) das Tabelas 25 a 36. A diferença se justifica pelo fato dos modelos terem sido calculados

em computadores distintos para redução. Sendo as métricas de qualidade para os modelos de configuração entre 1 a 55 rodados em um Intel Core i5-7200U (2,5 GHz) e os modelos de configuração 56 a 64 em um Dual-Core Intel Core i5 (1,8 GHz).