

ELIO RIBEIRO FARIA JUNIOR

COMPARATIVO DE TÉCNICAS DE AGRUPAMENTO
UTILIZADAS EM ABORDAGENS PARA DESCOBERTA
DE MÚLTIPLOS MODELOS DE PROCESSOS

Dissertação apresentada ao Programa de Pós-Graduação em
Informática da Pontifícia Universidade Católica do Paraná
como requisito parcial para obtenção do título de Mestre em
Informática.

Curitiba
2019

ELIO RIBEIRO FARIA JUNIOR

COMPARATIVO DE TÉCNICAS DE AGRUPAMENTO PARA DESCOBERTA DE MODELOS DE PROCESSOS

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. Edson Emílio Scalabrin

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Edilene de Oliveira dos Santos CRB 9 /1636

F224c
2019 Faria Junior, Elio Ribeiro
Comparativo de técnicas de agrupamento para descoberta de modelos de processos / Elio Ribeiro Faria Junior ; orientador, Edson Emílio Scalabrin. -- 2019
88 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná,
Curitiba, 2019
Bibliografia: f.78-88

1. Engenharia de software. 2. Mineração de dados (Computação). 3. Análise por agrupamento. 4. Controle de processo. 5. Máquinas – Indústria.
I. Scalabrin, Edson Emílio. II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática. III. Título.

CDD 20. ed. – 005.10681



Pontifícia Universidade Católica do Paraná
Escola Politécnica
Programa de Pós-Graduação em Informática

DECLARAÇÃO

Declaro para os devidos fins que o aluno **ELIO RIBEIRO FARIA JUNIOR**, defendeu sua dissertação de Mestrado intitulada "**COMPARATIVO DE TÉCNICAS DE AGRUPAMENTO UTILIZADAS EM ABORDAGENS PARA DESCOBERTA DE MÚLTIPLOS MODELOS DE PROCESSOS**", na área de concentração Ciência da Computação, no dia 05 de dezembro de 2019, no qual foi aprovado.

Declaro ainda que foram feitas todas as alterações solicitadas pela Banca Examinadora, cumprindo todas as normas de formatação definidas pelo Programa.

Por ser verdade, firmo a presente declaração.

Curitiba, 06 de novembro de 2019.



Emerson

Prof. Dr. Emerson Cabrera Paraiso
Coordenador do Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná

AGRADECIMENTOS

Ao Centro Universitário de União da Vitória (UNIUV), por fornece todo o apoio necessário.

Ao meu orientador Edson Emílio Scalabrin, que sempre foi muito paciente, conselheiro e acessível.

Ao Cleiton dos Santos Garcia, que me ajudou muito no desenvolvimento da pesquisa.

Aos meus pais Élio e Nelsi por sempre apoiarem.

Por fim, a todos os amigos que direta ou indiretamente contribuíram para a execução deste trabalho.

DEDICATÓRIAS

À memória do meu avô Ulysses Ribeiro Faria (in memoriam), que sempre me apoiou e nunca perdeu a fé nos meus sonhos. A minha Avó (mãe) Rosélia De Lima, por todo carinho e compreensão, essa mulher guerreira sempre é a maior fonte de inspiração para mim. Finalmente a minha esposa Geruza Vieira e ao meu filho Otto Humberto Vieira Ribeiro, meus grandes amores.

RESUMO

A Mineração de Processos é uma área de estudo crescente e promissora focada em entender os processos e ajudar a capturar as descobertas mais significativas sobre a execução real em vez de um modelo de processo idealizado. Os *logs* de eventos extraídos dos sistemas permitem descobrir modelos de processos; avaliar a conformidade; e expandir um modelo existente com indicações sobre desempenho. Entretanto, técnicas tradicionais de Mineração de Processos possuem dificuldades com processos não estruturados e flexíveis, gerando modelos de processos de difícil compreensão. Assim sendo, nosso objetivo é avaliar mecanismos de agrupamento de traços de processos utilizados em abordagens para descoberta de múltiplos modelos de processos, como etapa de pré-processamento à descoberta/geração de modelos de processos. Hipótese: o agrupamento de traços reduz a complexidade de múltiplos modelos de processos descobertos. Neste contexto, foi realizado um estudo comparativo para avaliar: (1) se as técnicas de agrupamento utilizadas em abordagens para descoberta de múltiplos modelos de processos são capazes de separar um *log* de eventos contendo processos diferentes; e (2) se tais técnicas também ajudam na redução da complexidade dos modelos descobertos. Os *logs* de eventos utilizados são de processos semiestruturados e não estruturados. A partir dos resultados obtidos concluiu-se que as técnicas de agrupamento selecionadas para o experimento não são capazes de separar um *log* de eventos contendo processos diferentes. Entretanto, os agrupamentos de traços gerados a partir das abordagens *Sequence Clustering*, *ActiTraC* e *Guilde Tree Miner* permitem obter modelos de processos com menor complexidade. Aqui, a principal contribuição está ligada a hipótese de que técnicas de agrupamentos utilizadas pelas abordagens impactam diretamente no resultado da Mineração de Processos.

Palavras-chaves: *Mineração de Processos, Técnicas de Agrupamento, Não Estruturado, Modelos de Processos.*

ABSTRACT

Process Mining is a growing and promising area of study focused on understanding processes and helping capture the most significant findings about actual execution rather than an idealized process model. The event logs extracted from the systems allow you to discover the process model; checking conformance; and expanding the model exists with indications about performance. However, traditional Process Mining techniques have difficulties with unstructured and flexible processes, generating models of processes that are difficult to understand. Therefore, our objective is to evaluate mechanisms of grouping of traces of processes used in approaches for the discovery of multiple process models, such as preprocessing stage, the discovery / generation of process models. The hypothesis is that the grouping of traits should reduce the complexity of the generated models. In this context, a comparative study was carried out to evaluate: (1) whether clustering techniques, used in approaches to discover multiple process models, are capable of separating a log of events containing different processes; (2) the reduction of the complexity of the models generated. The event logs used are semi-structured and unstructured processes. From the results obtained, it was concluded that the clustering techniques selected for the experiment are not capable of separating a log of events with different processes. However, the clusters generated from the Sequence Clustering, ActiTraC and Guilde Tree Miner approaches lead to models with less complexity than the log file without the use of clustering techniques in terms of complexity metrics. The main contribution is linked to the hypothesis that clustering techniques used by the approaches have a direct impact on the results of Process Mining.

Keywords: *Process mining, clustering techniques, unstructured, Process models*

SUMÁRIO

CAPÍTULO 1 - INTRODUÇÃO	1
1.1 Problema de Pesquisa	4
1.2 Objetivos	5
1.3 Motivação	6
1.4 Processo de Trabalho	6
1.5 Estrutura do documento da dissertação	7
1.6 Considerações sobre o Capítulo	7
CAPÍTULO 2 - MAPEAMENTO SISTEMÁTICO	8
2.1 Método de Pesquisa	8
2.1.1 Definição da Questão da Pesquisa	9
2.1.2 Condução de Pesquisa	12
2.1.3 Publicações e Extração de Dados	14
2.2 Subáreas de conhecimento na mineração de processos	16
2.2.1 Descoberta de processos	19
2.2.2 Conformidade do processo, <i>compliance e auditoria</i>	20
2.2.3 Aprimoramento, extensão e ênfase no desempenho	21
2.2.4 Minerador Organizacional/Social	22
2.2.5 Métodos de Aplicação	22
2.2.6 Arquitetura e Ferramentas	24
2.2.7 Métricas de Qualidade e Similaridade	25
2.2.8 Controle de Variantes do Processo e Agrupamento	26
2.2.9 Técnicas preditivas	28
2.3 Algoritmos de Descoberta de Processos	28
2.4 Aplicações da Mineração de Processos	30
2.4.4 Saúde	32
2.4.5 Tecnologia da Informação e Comunicação	34
2.4.6 Industrial	34
2.4.7 Educação	35
2.4.8 Finanças	35
2.4.9 Outros Domínios de Aplicações	36
2.5 Considerações sobre o Capítulo	36
CAPÍTULO 3 - TÉCNICAS DE AGRUPAMENTO	39
3.1 Abordagens	40
3.2 Resumo das abordagens	47
3.3 Considerações sobre o Capítulo	49
CAPÍTULO 4 - ESTRUTURAÇÃO DA PESQUISA	51
4.1 Método de Pesquisa	51
4.2 Questões de Pesquisa	51
4.3 Estratégia de Pesquisa	52
4.3.1 Seleção das abordagens	52
4.3.2 Construção do ambiente experimental	53
4.3.3 Coleta de dados	59
4.3.4 Avaliação dos resultados	61

4.4	Considerações sobre o Capítulo	61
CAPÍTULO 5 - ANÁLISE E DISCUSSÃO DOS RESULTADOS		62
5.1	Cenário 1: Separação de um arquivo de log contendo três processos diferentes	63
5.2	Cenário 2: Redução da Complexidade dos Modelos	64
5.2.1	Log de eventos TIC	64
5.2.2	Log de eventos Manufatura	67
5.2.3	Log de eventos de Telecomunicações	70
5.2.4	Log de eventos Seguros	71
5.3	Análise dos Resultados	72
5.4	Limitações da Pesquisa	74
5.5	Considerações sobre o Capítulo	74
CAPÍTULO 6 - CONCLUSÕES		75
REFERÊNCIAS BIBLIOGRÁFICAS		78

LISTA DE FIGURAS

Figura 1 - Tipos de Mineração de Processos, adaptado de [van der Aalst 2016] -----	2
Figura 2 - Modelo de Rede de Petri. Fonte: [Song, Günther e van der Aalst 2009]. -----	4
Figura 3 - Modelos de processos obtidos a partir do agrupamento. Fonte: [Song, Günther e van der Aalst 2009]. -----	4
Figura 4 - Processo de mapeamento sistemático. Adaptado de [Petersen et al. 2008]-----	8
Figura 5 - Processo de triagem por critérios. Fonte: o Autor, 2018 -----	14
Figura 6 - Quantidade de publicações sobre Mineração de Processos por ano. Fonte: o Autor, 2018 -----	15
Figura 7 - Periódicos e conferências por país. Fonte: o Autor, 2018 -----	15
Figura 8 - Subáreas de conhecimento da mineração de processos. Fonte: o Autor, 2018 --	17
Figura 9 - Distribuição de artigos por categorias de conhecimento da mineração de processos. Fonte: o Autor, 2018-----	18
Figura 10 - Mapa da distribuição de artigos por categorias de conhecimento. Fonte: o Autor, 2018 -----	18
Figura 11 - Algoritmos de Mineração de Processos. Fonte: o Autor, 2018 -----	29
Figura 12. Distribuição do uso dos algoritmos de mineração de processos. Fonte: o Autor, 2018 -----	30
Figura 13. Número de artigos por domínio de aplicação. Fonte: o Autor, 2018-----	32
Figura 14. Publicações por ano dos tipos de mineração de processos. Fonte: o Autor, 2018 -----	37
Figura 15 - Publicações por ano dos principais domínios de aplicação. Fonte: o Autor, 2018 -----	38
Figura 16. Resultado de um modelo de processo com o agrupamento. [Song et al.2013]---	39
Figura 17. Etapa Sequence Clustering. Fonte: [Rebuge e Ferreira 2012].-----	42
Figura 18. Métodos de ligação. Fonte: [Johnson e Wichern 1998].-----	43
Figura 19 - Etapas do planejamento do experimento. Fonte: o Autor, 2018-----	52
Figura 20 - Parâmetros DWS. Fonte: o Autor, 2019. -----	55
Figura 21 - Parâmetros Trace Clustering. Fonte: o Autor, 2019 -----	56
Figura 22 - Parâmetros Guilde Tree Miner. Fonte: o Autor, 2019. -----	57
Figura 23 - Parâmetros Sequence Clustering. Fonte: o Autor, 2019.-----	58
Figura 24 - Parâmetros ActiTraC. Fonte: o Autor, 2019. -----	58
Figura 25 - Processo da coleta de dados do primeiro cenário. Fonte: o Autor, 2019. -----	60
Figura 26 - Processo da coleta de dados do segundo cenário. Fonte: o Autor, 2019. -----	61
Figura 27. Modelo gerado sem aplicação de técnicas de agrupamento e com ActiTraC: log de eventos TIC. Fonte: o Autor, 2019. -----	66

Figura 28. Alto custo computacional gerado pela abordagem Trace Clustering. Fonte: o Autor, 2019.-----	67
Figura 29. Log de eventos de Manufatura: (a) modelo gerado sem as técnicas de agrupamento. (b) modelos gerados a partir do resultado da aplicação da abordagem ActiTraC para gerar agrupamentos. Fonte: o Autor, 2019.-----	69
Figura 30. Comparativo CNC e CN. Fonte: o Autor, 2019.-----	73

LISTA DE TABELAS

Tabela 1 - Log de Eventos. Fonte: [Song, Günther e van der Aalst 2009]. -----	3
Tabela 2 - Trabalhos relacionados a revisão distemática e mapeamento sistemático. Fonte: o Autor, 2018. -----	11
Tabela 3 - String de Pesquisa. Fonte: o Autor, 2018. -----	12
Tabela 4 - Número de artigos por base de dados digital. Fonte: o Autor, 2018. -----	13
Tabela 5. Domínios de aplicação. Fonte: o Autor, 2018. -----	31
Tabela 6 - Abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento. Fonte: o Autor, 2019 -----	48
Tabela 7 - Abordagens selecionadas para o experimento. Fonte: o Autor, 2019-----	53
Tabela 8 - Log de eventos selecionados para o experimento. Fonte: o Autor, 2019. -----	59
Tabela 9 - Resultados das aplicações das técnicas de agrupamento. Fonte: o Autor, 2019.	63
Tabela 10 - Complexidade de modelos vs. técnicas de agrupamento na descoberta de múltiplos modelos de processos: log de eventos TIC. Fonte: o Autor, 2019.-----	65
Tabela 11 - Resultados das métricas de complexidade dos modelos gerados a partir das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos no log de eventos Manufatura. Fonte: o Autor, 2019.-----	68
Tabela 12 - Resultados das métricas de complexidade dos modelos gerados a partir das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos no log de eventos da área de Telecom. Fonte: o Autor, 2019. -----	70
Tabela 13 - Resultados das métricas de complexidade dos modelos gerados a partir das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos no log de eventos da área de Seguros. Fonte: o Autor, 2019.-----	71
Tabela 14 - Resultado do teste estatístico não paramétrico de Kruskal-Wallis. Fonte: o Autor, 2019. -----	73

LISTA DE ABREVIATURAS E SIGLAS

AP	<i>Affinity Propagation</i>
AWS	<i>Abstraction Workflow Schema</i>
BAM	Monitoramento de Atividades de Negócio
BPA	Análise de Processos de Negócio
BI	Inteligência de Negócios
BP	Processos de Negócio
BPI	Melhoria do Processo de Negócios
BPM	Gerenciamento de Processos de Negócio
BPMS	Gestão de Processos de Negócios
CE	Critério de Exclusão
CI	Critérios de inclusão
CN	Cyclomatic Number
CNC	Coefficient of Connectivity
CPI	Melhoria Contínua do Processo
CPM	Gerenciamento de Desempenho Corporativo
CRM	Customer Relationship Management
EPM	Gerenciamento de Desempenho Empresarial
DWS	Disjunctive Workflow Schema
ERP	Planejamento de Recursos Empresariais
ILP	Programação Lógica Indutiva
KPI	Indicadores-Chave de Desempenho
LPM	<i>Local Process Model</i>
MES	Sistemas de Execução de Manufatura
OLAP	Processamento Analítico <i>Online</i>

PPM	Modelo de Processo Probabilístico
RSL	Revisão Sistemática da Literatura
SCADA	Controle de Supervisão e Aquisição de Dados
TIC	Tecnologia de informação e comunicação
XES	Fluxo de eventos extensível
WFMS	Sistemas de Gerenciamento de <i>Workflow</i>

CAPÍTULO 1 - INTRODUÇÃO

Atualmente existem eventos de muitos processos de negócios registrados em sistemas corporativos, sistemas de automação e controle, sistemas médicos, atividades diárias, dispositivos IoT, redes sociais, entre outros. Esses eventos oferecem muitas possibilidades de geração de conhecimentos para compreender o que está acontecendo de fato. Esta iniciativa coloca em destaque a área de Mineração de Processos que visa descobrir, monitorar e aprimorar processos reais por meio da descoberta de conhecimentos de logs de eventos prontamente disponíveis em muitos sistemas de informação [Van der Aalst 2011]. A Figura 1 mostra as tarefas da Mineração de Processos. O projeto tradicional de um processo de negócio começa a partir de uma abordagem detalhada de mapeamento, envolvendo vários recursos para estabelecer um modelo consenso com a perspectiva mais reconhecida dos participantes. Como alternativa, a Mineração de Processos pressupõe descobrir de forma automática o modelo de processo a partir de logs de eventos e/ou de documentos temporais prontos e disponíveis na maioria dos sistemas de informação.

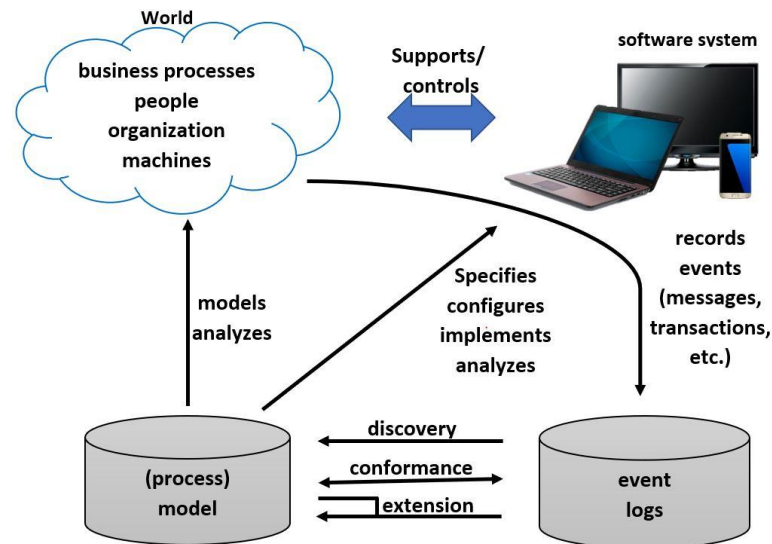


Figura 1 - Tipos de Mineração de Processos, adaptado de [van der Aalst 2016]

O objetivo da Mineração de Processos é extrair um modelo de processo explícito e consistente a partir de *logs* de eventos [Van der Aalst e Weijters 2004]. A Mineração de Processos é definida como um método capaz de gerar uma descrição estruturada ou compreensível de processo a partir de um conjunto de eventos reais de execução devidamente registrada [van der Aalst et al. 2004]. Neste contexto, foram propostas ferramentas para melhorar, aprimorar e auditar os processos de negócios. De acordo com [Westergaard e Wynn 2012], também é possível aplicar o processo de mineração para apoiar processos de conformidade como a lei Sarbanes-Oxley e Basileia II. A Mineração de Processos emergiu como uma ligação entre a análise de processos de negócios (BPA) e monitoramento de atividades de negócios (BAM), aplicando a mineração de dados para detectar automaticamente padrões e modelos reais [Reijers et al. 2016].

De acordo com o manifesto publicado, a Mineração de Processos oferece uma ponte entre mineração de dados, modelagem e análise de processos de negócios [van der Aalst et al. 2012]. Ela não substitui as abordagens de melhorias de processo tradicionais, como a melhoria contínua de processo (CPI), a melhoria de processo de negócios (BPI), *Six Sigma*, gerenciamento de desempenho corporativo (CPM) e outros. Entretanto, a Mineração de Processos é capaz acelerar melhorias em um dado processo. Vários mineradores de processos (especializados na descoberta) são capazes de acelerar o processo de compreensibilidade, modelagem e redesenho [van der Aalst, Weijters, e Maruster 2002], [Weijters e Van der Aalst

2003], [van der Aalst, De Medeiros e Weijters 2005], [Jansen-Vullers, Van der Aalst e Rosemann 2006], [Günther e van der Aalst 2007], [Lamma et al. 2007], [Leemans, Fahland e van der Aalst 2013], [De Smedt, De Weerd e Vanthienen 2014].

Perante este contexto, a Mineração de Processos ainda enfrenta dificuldades para lidar com processos reais que tendem a ser menos estruturados e altamente flexíveis. [Song, Günther e van der Aalst 2009] descreve que o processo de tratamento que ocorre em um hospital é considerado não estruturada, pois existe uma grande diversidade de informações levando a modelos complexos e de baixa compreensão. Técnicas tradicionais de descoberta de processos [van der Aalst et al. 2004], [Weijters e Van der Aalst 2003], [Weijters et al. 2006], [van der Aalst, De Medeiros e Weijters 2005], [Werf et al. 2008], [Lamma et al. 2007], [Leamans et al. 2013] possuem problemas com processos semiestruturados ou não estruturados e geram modelos de difícil compreensão. Em relação a esse problema, os métodos de agrupamento de instâncias de processo podem melhorar os resultados da Mineração de Processos. Eles possibilitam dividir um *log* de eventos existente em subconjuntos menores facilitando assim o processo de descoberta [Weerd et al. 2013].

Com base em um *log* de eventos (Tabela 1) e a utilização do *Alpha Algorithm*, [Song, Günther e van der Aalst 2009] criaram um modelo de Rede de Petri (cf. Figura 2). De acordo com [Meinheim 2018], Redes de Petri são utilizadas para representar modelos de processos descobertos. Deve-se observar que o modelo gerado (cf. Figura 2), não é um resultado ideal considerando as métricas de *fitness* e simplicidade. Entretanto, dividindo o *log* de eventos em vários grupos, pode-se obter modelos mais precisos (cf. Figura 3).

Tabela 1 - Log de Eventos. Fonte: [Song, Günther e van der Aalst 2009].

<i>Case ID</i>	<i>log events</i>
1	(A, John), (B, Mike), (D, Sue), (E, Pete), (F, Mike), (G, Jane), (I, Sue)
2	(A, John), (B, Fred), (C, John), (D, Clare), (E, Robert), (G, Mona), (I, Clare)
3	(A, John), (B, Pete), (D, Sue), (E, Mike), (F, Pete), (G, Jane), (I, Sue)
4	(A, John), (C, John), (B, Fred), (D, Clare), (H, Clare), (I, Clare)
5	(A, John), (C, John), (B, Robert), (D, Clare), (E, Fred), (G, Robert), (I, Clare)
6	(A, John), (B, Mike), (D, Sue), (H, Sue), (I, Sue)

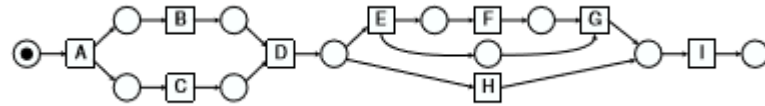


Figura 2 - Modelo de Rede de Petri. Fonte: [Song, Günther e van der Aalst 2009].

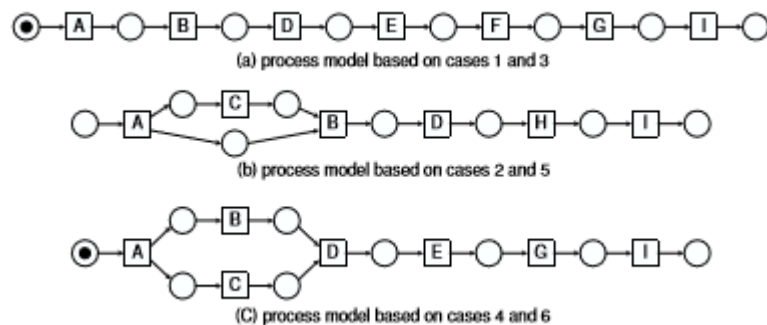


Figura 3 - Modelos de processos obtidos a partir do agrupamento. Fonte: [Song, Günther e van der Aalst 2009].

Neste contexto, este trabalho apresenta um estudo de diferentes técnicas de agrupamento de instância de processos utilizadas em abordagens para descoberta de múltiplos modelos de processos e avaliar como elas impactam na qualidade dos modelos de processos em *logs* de eventos semiestruturados ou não estruturados. Para desenvolver este trabalho foi conduzido um experimento envolvendo diferentes abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento. Foram usadas abordagens para as quais uma implementação de trabalho está disponível (capítulo 3), entretanto para a generalização dos resultados foi utilizado o *Heuristic Miner* [Weijters et al. 2006]. Na sequência, os agrupamentos gerados pelas tais abordagens foram avaliados sob dois cenários: cenário 01, visou a separação de um arquivo de log contendo três processos diferentes; e cenário 02: Baseado nos modelos de processo descobertos foi calculado a redução da complexidade dos modelos minerados.

1.1 Problema de Pesquisa

O avanço da tecnologia e o baixo preço do hardware, ocasionaram um grande volume de informações advinda de vários sistemas de informação e sensores, por exemplo, sistemas

corporativos, sistemas de automação e controle, sistemas médicos, atividades diárias, dispositivos IoT, redes sociais, entre outros. A fim de encontrar padrões úteis nesse grande volume de informações, técnicas de Mineração de Processos foram utilizadas para analisar os *logs* de evento, entretanto, as técnicas tradicionais possuem dificuldades na análise de *logs* de eventos não estruturados e flexíveis [Song et al. 2013], gerando modelos de processos de difícil compreensão.

Uma maneira de superar tais dificuldades e extrair modelos de processos compreensíveis, é a aplicação de técnicas de agrupamento na fase que antecede a mineração de processos, com o objetivo de obter modelos de processos compreensíveis [Rosa et al. 2017].

Em vista que a escolha da técnica de agrupamento impacta diretamente no resultado obtido da mineração de processos, o problema de pesquisa se coloca por meio das seguintes questões: Quais são as técnicas de agrupamento—utilizados nas abordagens para descoberta de múltiplos modelos de processos? As técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos são capazes de separar um log de eventos contendo processos diferentes? As diferentes técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos impactam no resultado da Mineração de Processos?

1.2 Objetivos

O objetivo geral é analisar as técnicas de agrupamento utilizadas em abordagens para descoberta de múltiplos modelos de processos. Tal análise visa estabelecer a técnica de agrupamento mais eficiente utilizada pelas abordagens descoberta de múltiplos modelos de processos. Os objetivos específicos desta pesquisa compreendem:

- Selecionar diferentes técnicas de agrupamento utilizadas em abordagens voltadas para descoberta de múltiplos modelos de processos;
- Construir um experimento para avaliação das técnicas de agrupamento;
- Analisar os resultados quanto a capacidade de separar um arquivo de log de eventos contendo três processos diferentes e à redução da complexidade dos modelos minerados.

1.3 Motivação

As técnicas de mineração de processos possuem, de um lado, uma ótima performance em processos estruturados, de outro lado, uma performance razoável em processos não-estruturados. Entretanto, a mineração de processos é mais interessante em situações que exigem flexibilidade [Song, Günther e van der Aalst 2009], i.e., em processos reais frequentemente não estruturados.

A utilização de técnicas de agrupamento na fase que antecede a mineração de processos deve impactar diretamente na descoberta de modelos de processos, e conseqüentemente, elas devem permitir gerar modelos mais compreensíveis a partir de *logs* de eventos não estruturados, provenientes de processos flexíveis.

Sendo assim, as técnicas de agrupamento utilizadas em abordagens para a descoberta de múltiplos modelos de processos devem permitir gerar modelos de processos com maior qualidade, descrevendo o modelo descoberto com maior compreensibilidade e exatidão.

1.4 Processo de Trabalho

Para organizar o trabalho de pesquisa e atingir os objetivos propostos foi definido um processo inicial com várias fases, que são:

- Fase 1 – Preparação da pesquisa: delimitação da área de estudo, realização de um Mapeamento sistemático sobre Mineração de Processos e uma revisão da literatura mais abrangente sobre abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento.
- Fase 2 – Estruturação da pesquisa: seleção do método e das etapas de pesquisa para execução do trabalho. A experimentação foi o método adotado.
- Fase 3 – Execução da pesquisa: condução da pesquisa realizada conforme o plano experimental.
- Fase 4 – Análise dos Resultados: fase de análise dos resultados obtidos e as conclusões do trabalho.

1.5 Estrutura do documento da dissertação

Este trabalho está estruturado da seguinte maneira:

- O Capítulo 1, aqui apresentado, dá a contextualização sobre o tema deste trabalho, bem como os objetivos gerais e específicos.
- O Capítulo 2 apresenta um mapeamento sistemático sobre Mineração de Processos.
- O Capítulo 3 apresenta uma revisão da literatura sobre abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento.
- O Capítulo 4 apresenta o método de pesquisa utilizado neste trabalho, bem como as etapas e preparação do método.
- O Capítulo 5 apresenta os resultados obtidos e limitações identificadas desta pesquisa.
- O Capítulo 6 apresenta as conclusões e trabalhos futuros.

1.6 Considerações sobre o Capítulo

Este capítulo apresentou as limitações das técnicas de Mineração de Processos em *logs* de eventos não estruturados—gerados frequentemente em processos flexíveis—, onde acabam gerando modelos com alta complexibilidade para a compreensão, e destacou o uso de técnicas de agrupamento em abordagens para descoberta de múltiplos modelos de processos, as quais podem facilitar as análises em processos não estruturados. Por fim, foi apresentado o problema, motivação e objetivos da pesquisa.

CAPÍTULO 2 - MAPEAMENTO SISTEMÁTICO

Neste Capítulo foi desenvolvido um Mapeamento Sistemático. Ele fornece uma ampla visão da área de Mineração de Processos.

2.1 Método de Pesquisa

Este estudo colocou em prática o método de mapeamento sistemático para identificar, organizar e compreender as principais contribuições do estado da arte em relação às técnicas e aplicações de mineração de processo. Os estudos de mapeamento sistemático são projetados para fornecer uma visão ampla de uma área de pesquisa, assim como para estabelecer e procurar evidências sobre um tópico e fornecer medidas quantitativas sobre os trabalhos anteriores [Kitchenham e Charters 2007]. Neste contexto, o método de mapeamento sistemático foi apropriado para procurar áreas de estudos abrangentes. A Figura 4 ilustra o processo de mapeamento sistemático com as etapas essenciais do processo.

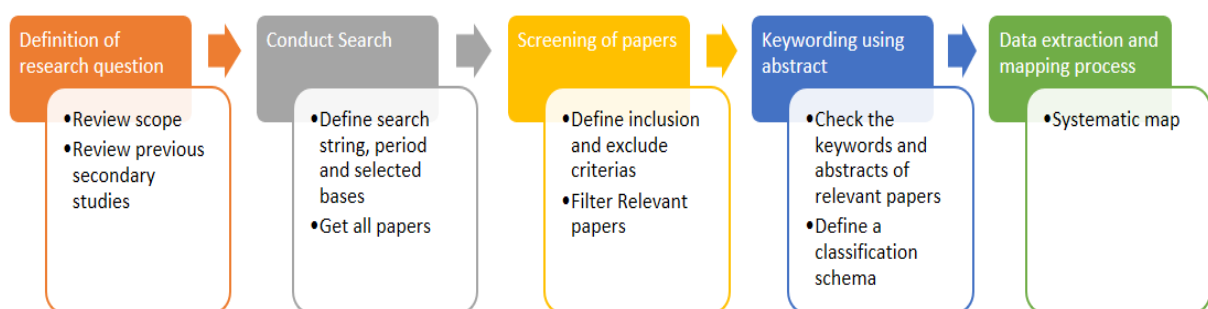


Figura 4 - Processo de mapeamento sistemático. Adaptado de [Petersen et al. 2008]

Segundo Petersen et al. (2008), cada etapa do processo tem um resultado até o mapa sistemático final, conforme descrito a seguir:

- Definição das questões de pesquisa (escopo de pesquisa): são pensadas para fornecer um escopo geral para o estudo;

- Realização da pesquisa: os estudos primários são identificados por meio de *String* de busca nas diversas bases de dados científicas selecionadas;
- Triagem de artigos: são selecionados artigos relevantes, aplicando os critérios de inclusão e exclusão apropriados;
- Palavras-chave: são identificadas e combinadas para desenvolver uma compreensão de alto nível sobre a natureza e contribuição da pesquisa, gerando uma classificação bem-definida;
- Processo de extração e mapeamento de dados: os trabalhos classificados na etapa anterior são extraídos, agregados, visualizados e mapeados com o objetivo de responder às questões de pesquisa.

2.1.1 Definição da Questão da Pesquisa

O objetivo desta etapa de estudo é mapear e compreender as principais contribuições do estado da arte para as técnicas e aplicações de mineração de processos. Dada essa meta foram definidos alguns passos para a realização desta pesquisa, utilizando questões para guiar o estudo do mapeamento sistemático.

RQ1. Como classificar os estudos primários por área de conhecimento?

Esta questão visa identificar as subáreas da mineração de processos e estabelecer um mapeamento categorizado a partir de publicações anteriores. Também é interessante mapear as categorias de trabalhos que estão atraindo atenção e as novas áreas de conhecimentos emergentes.

RQ2. Quais são os algoritmos de mineração de processos propostos?

A intenção dessa questão é destacar as diferentes propostas para essa subárea específica de mineração de processos em termos de descoberta de modelos de processos de negócios a partir de um conjunto de *logs* de eventos. Assim como, mapear os mineradores de processo ao longo do tempo e identificar os autores e a quantidade de artigos aplicados a cada minerador.

RQ3. Onde a mineração de processos está sendo aplicada?

Esta questão destina-se a compreender o domínio de aplicações exploradas anteriores e/ou segmentos da indústria. Pesquisas iniciais referentes a aplicações identificaram muitas publicações em áreas hospitalares ou clínicas. Observou-se também padrões dos clientes de e-commerce e manufatura, entre outras.

Tão importante quanto explorar estudos primários, este mapeamento primeiro realizou uma exploração em bases de dados digitais em relação aos estudos secundários, buscando revisões sistemáticas e mapeamentos sistemáticos. A Tabela 2 resume estudos secundários entre 2003 e 2018.

Duas pesquisas [van der Aalst et al. 2003] [van der Aalst e Weijters 2004] guiaram a área de mineração de processos, compartilhando questões, técnicas e abordagens. Esses estudos reforçaram a colaboração e ao suporte de novas técnicas e ferramentas. Esses trabalhos foram uma etapa inicial para mostrar as muitas lacunas em contribuições futuras, aprimoramentos de algoritmos e aplicações. Estudos secundários de [Peleg 2013] revisaram diretrizes práticas clínicas, onde a mineração de processos foi identificada como uma solução para a manutenção de diretrizes e análise de conformidade. Outro estudo secundário, foi focado em linhas de produtos de software para gerenciamento de processos de negócios e a mineração de processos foi identificada como uma abordagem para otimização de processos [Rocha e Fantinato 2013]. A mineração de processos foi identificada como uma abordagem proeminente para o controle de segurança em uma revisão sistemática sobre segurança [Leitner e Rinderle-Ma 2014]. Outro estudo secundário abordou a mineração de processos com foco no monitoramento de conformidade, que estabeleceu as principais funcionalidades e comparou as abordagens e ferramentas existentes [Ly et al. 2015]. Um estudo de gestão de processos de negócios analisou aspectos de qualidade do mapeamento da evolução, maturidade, impacto e indicou uma atenção crescente da comunidade sobre o processo de mineração, descoberta ou análise [Recker e Mendling 2016]. Finalmente, duas revisões da literatura centraram-se sobre a mineração de processos na saúde, [Rojas et al. 2016] mapearam estudos de caso de acordo com 22 áreas médicas, e [Kurniati et al. 2016] revisou trabalhos em aplicações oncológicas.

Tabela 2 - Trabalhos relacionados a revisão distemática e mapeamento sistemático. Fonte: o Autor, 2018.

Ano	Título	Periódico	Notas
2003	<i>Workflow mining: A survey of issues and approaches</i>	<i>Data & Knowledge Engineering</i>	O documento propõe um formato XML comum para logs de eventos usados para comparar 4 ferramentas de mineração de processos existentes.
2004	<i>Process mining: A research agenda</i>	<i>Computers in Industry</i>	Apresenta uma visão geral, incluindo 11 desafios a serem abordados, tais como ruído, incompletude, loop, etc.
2013	<i>Computer-interpretable clinical guidelines: A methodological review</i>	<i>Biomedical Informatics</i>	A mineração de processos é apenas uma abordagem possível usada para criar diretrizes interpretadas por computador.
2013	<i>The use of software product lines for business process management: A systematic literature review</i>	<i>Information and Software Technology</i>	Entre 25 atributos estabelecidos, apenas um foco nas fases de avaliação de gerenciamento processo de negócio (BPM), com o objetivo de extrair trabalhos primários em otimização de processos usando monitoramento de atividades empresariais e mineração de processos.
2014	<i>Systematic review on security in Process-Aware Information Systems - Constitution, challenges, and future directions</i>	<i>Information and Software Technology</i>	Discute as capacidades para verificar a conformidade, inconsistências de identidade, etapas de salto (detecção de fraudes), análise de causa básica em mineração de processos para melhorar a segurança.
2015	<i>Compliance monitoring in business processes: Functionalities, application, and tool-support</i>	<i>Information Systems</i>	Discute o monitoramento de conformidade, definindo as principais funcionalidades e comparando abordagens.
2016	<i>The State of the Art of Business Process Management Research as Published in the BPM Conference Recommendations for Progressing the Field</i>	<i>Business & Information Systems Engineering</i>	O trabalho está relacionado com métodos de pesquisa, discussão de qualidade, maturidade, índice de citações e progresso na área de gestão de processos de negócios.
2016	<i>Process mining in healthcare: A literature review</i>	<i>Biomedical Informatics</i>	Revisão da literatura sobre o uso da mineração de processos na área da saúde cobrindo 74 artigos.
2016	<i>Process mining in oncology: A literature review</i>	<i>IC on Information Communication</i>	Estudo de mineração de processos em Oncologia cobrindo 37 artigos de 758.

Apesar da existência de mapeamentos e revisões, poucos trabalhos foram diretamente relacionados com técnicas de mineração de processos. Outras lacunas abertas em trabalhos secundários anteriores estão relacionadas ao estabelecimento de uma visão amplamente

horizontal das subáreas de mineração de processos e de diferentes segmentos de aplicativos de negócios.

2.1.2 Condução de Pesquisa

A *String* de pesquisa utilizada para este estudo foi: ("*process mining*" OR "*processes mining*" OR "*workflow mining*" OR "*mining workflow*" OR "*workflows mining*" OR "*mining workflows*"). As fontes de dados escolhidas foram a ACM Digital Library, IEEE Xplore, ScienceDirect e Springer Link. O período de filtragem foi definido a partir de 2002, devido ao algoritmo de descoberta relevante, chamado *Alpha Algorithm* ou α -miner, até 2018. A Tabela 3 apresenta a *String* de pesquisa adaptada a cada biblioteca digital. Na Tabela 4 é possível visualizar a quantidade de artigos extraídos de cada base de dados de acordo com a *String* de pesquisa estabelecida.

Tabela 3 - *String* de Pesquisa. Fonte: o Autor, 2018.

Base de Dados	<i>String de Busca</i>
ACM Digital Library	"query": {content.ftsec:("process mining" "processes mining" "workflow mining" "mining workflow" "workflows mining" "mining workflows")} "filter": {"publicationYear": {"gte":2002}}
IEEE Xplore	("process mining" OR "processes mining" OR "workflow mining" OR "mining workflow" OR "workflows mining" OR "mining workflows") & ranges=2002_2018_Year
ScienceDirect	("process mining" OR "processes mining" OR "workflow mining" OR "mining workflow" OR "workflows mining" OR "mining workflows") AND pub-date > 2001
Springer link	("process+mining"+OR+"processes+mining")+OR+"workflow+mining"+OR+"mining+workflow"+OR+"workflows+mining"+OR+"mining+workflows") & facet-start-year=2002

No processo de seleção foram excluídos artigos com base em resumos e, em alguns casos, na leitura de texto completo. Outros estudos também foram adicionados conforme recomenda a técnica de *bola de neve*. Cada artigo foi revisado apenas por uma única pessoa. Um conjunto de critérios de inclusão e exclusão foi aplicado para identificar todos os trabalhos relevantes para este estudo. Os critérios de inclusão (CI) utilizado foram:

- CI.1: O artigo é disponibilizado eletronicamente e encontrado pela *String* de pesquisa em todos os campos e período;
- CI.2: Pelo menos dois revisores no periódico ou conferência;

- CI.3: Estudos publicados *on-line* no período de 2002 a fevereiro de 2018;
- CI.4: Técnica de *bola de neve* que inclui alguns trabalhos referenciados não indexados em bibliotecas digitais selecionadas.

Tabela 4 - Número de artigos por base de dados digital. Fonte: o Autor, 2018.

Base de Dados	Quantidade de Artigos
<i>ACM Digital Library</i>	470
<i>IEEE Xplore</i>	1357
<i>ScienceDirect</i>	1169
<i>SpringerLink</i>	685
Total	3681

Posteriormente, foi aplicado o critério de exclusão (CE). Esses critérios foram os seguintes:

- CE.1: Não está escrito em inglês;
- CE.2: trabalho duplicado;
- CE.3: Não se trata de mineração de processos, por exemplo: mineração de ferro, mineração de metal, poluição, química ou impacto ambiental;
- CE.4: Mineração de processos referenciada somente na introdução, ou fundamentos, ou parte do estado da arte. Não é uma contribuição ou aplicação de mineração de processos;
- CE.5: Mineração de processos somente como direção adicional;
- CE.6: Apenas relacionado à biografia do autor;
- EC.7: Não é um documento, apenas um guia, índice de conteúdo ou apenas informações de *marketing*;
- EC.8: Contribuição muito curta, menos de três página;
- EC.9: Propostas padrão (ISO / IEC);
- EC.10: Documentos não disponíveis ou removidos da base de pesquisa.

A Figura 5 mostra o total de 3.681 artigos, 31 artigos incluídos pela técnica da bola de neve, e a quantidade de artigos excluídos de acordo com cada critério de exclusão. Após a aplicação de todos os critérios de exclusão, 1.238 artigos foram selecionados.

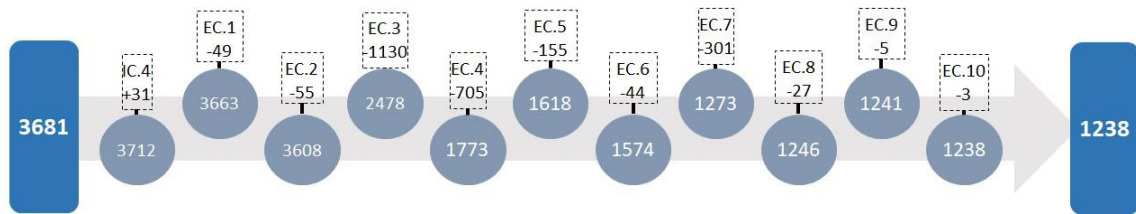


Figura 5 - Processo de triagem por critérios. Fonte: o Autor, 2018

2.1.3 Publicações e Extração de Dados

A Figura 6 fornece uma visão geral do total de trabalhos publicados por ano. A quantidade de publicações sobre mineração de processos cresce ano a ano. Nos últimos anos, podemos observar uma certa instabilidade, mas é provavelmente causada pelo processo de aprovação para alguns artigos. Como este mapeamento sistemático levou dois anos para ser concluído, o primeiro pico de publicações ocorreu em 2014, após carregar o *delta* de artigos, o pico foi movido para 2015. Isso mostra que é uma área de conhecimento crescente com uma instabilidade nos últimos 2 anos causada principalmente pelo período de revisão e edição, e a inclusão dos primeiros meses de 2018.

A extração de dados começa com informações genéricas sobre os países mais ativos em mineração de processos, compostos pela China, Holanda, Itália, Alemanha e EUA (cf. Figura 7). As contribuições desses países foram identificadas com base apenas na filiação do primeiro autor. É notória a grande contribuição do grupo *Process and Data Science* da Holanda. Por muitos anos, Prof.dr.ir. van der Aalst liderou o grupo em *Technische Universiteit Eindhoven* como autor ou coautor de mais de novecentas publicações. Agora, ele lidera o *Rheinisch-Westfälische Technische Hochschule*. Outros grupos de pesquisa de mineração de processos relevantes estão presentes em: *Katholieke Universiteit Leuven* da Bélgica, *Queensland University* da Austrália, *Universität Ulm* da Alemanha, *Università della Calabria* da Itália, *Universität Innsbruck* e *Vienna University of Technology* da Áustria. Pesquisas significativas e contribuições para a área de mineração de processos foram desenvolvidas por meio de parcerias envolvendo mais de um grupo de pesquisa. Como prova, Prof.dr.ir. van der Aalst é coautor em cerca de 29% das publicações de pesquisas cujo primeiro autor é afiliado a entidades de outros países.

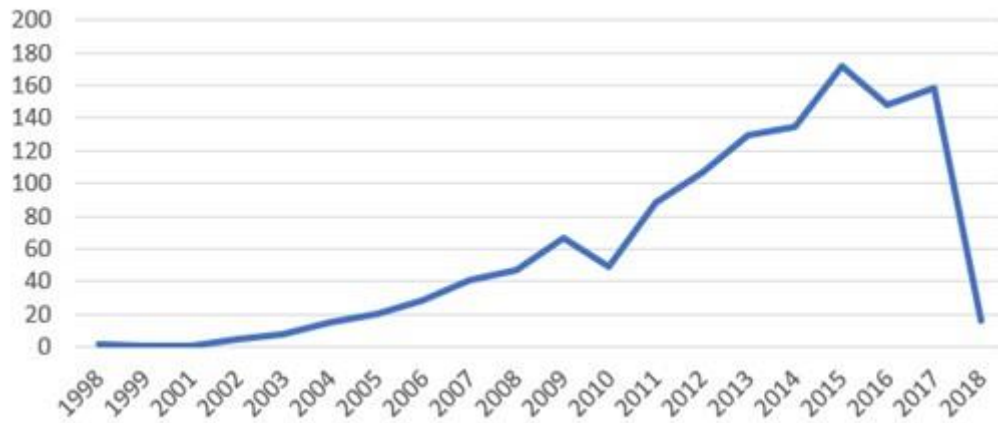


Figura 6 - Quantidade de publicações sobre Mineração de Processos por ano. Fonte: o Autor, 2018

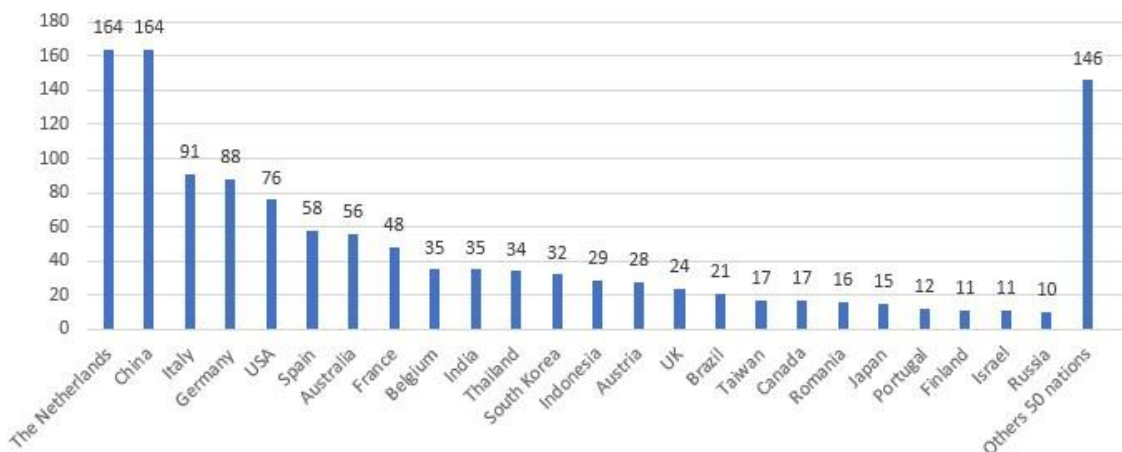


Figura 7 - Periódicos e conferências por país. Fonte: o Autor, 2018

Outras informações relevantes são os periódicos e conferências que tiveram mais contribuições. Identificou-se que os periódicos contribuem com cerca de 611 artigos, os de conferências são mais de 582, seguidos por alguns capítulos de livros. O *Information Systems* é o periódico que possui mais publicações relacionadas a mineração de processos, totalizando 47 artigos, seguido por *Expert Systems with Applications*, com 31 artigos e, *Decision Support Systems*, com 28 artigos.

2.2 Subáreas de conhecimento na mineração de processos

Para responder à pergunta de pesquisa do mapeamento sistemático: "como classificar os estudos primários em áreas de conhecimento?". Todos os estudos selecionados foram analisados e categorizados de acordo com a maior contribuição, uma vez que alguns trabalhos têm um foco secundário, também se identificou uma categorização secundária para os artigos que possuem contribuições relevantes em duas subáreas de conhecimento, aqui também chamadas de categorias. É de salientar que algumas categorias foram propostas de acordo com o conhecimento prévio e as primeiras análises dos estudos secundários. O mapeamento sistemático explorou os três principais tipos de mineração de processos, e observou-se uma categoria relevante relacionada ao minerador organizacional e social. Basicamente, os tipos de mineração de processos são: algoritmos de descoberta de modelos de processos, também chamados de mineradores de processo; conformidade de processo focada na verificação e auditoria de conformidade; e o aprimoramento focado em estender os processos, adicionar ênfase, informações e indicadores-chave de desempenho (KPI); por fim, o minerador organizacional e social identificam a relação entre os executores, funções e unidade.

A Figura 8 ilustra os tipos de mineração de processos identificados com as categorias de conhecimento na parte inferior para oferecer suporte a aplicações de mineração de processos, como: métodos de aplicação, diretrizes para uma estrutura de técnica ou orientação; arquitetura e ferramentas técnicas que combinam várias tecnologias; aquisições de dados de eventos; métricas para melhorar a qualidade do processo; e, ontologias. No lado direito, estão as técnicas avançadas de combinação de novas perspectivas e *insights*, usando técnicas de agrupamento e controle de variante de processo; simulação; alocação de distribuição de recursos; técnicas preditivas; mineração de regra de decisão; detecção de mudança de conceito identificando nova concepção ou realidade.

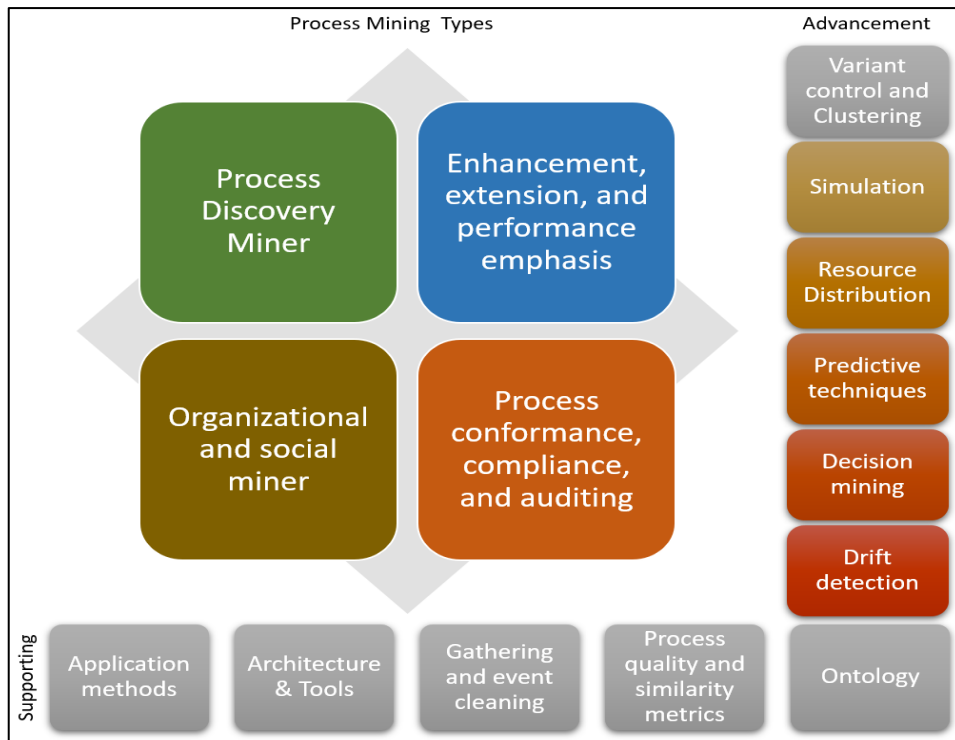


Figura 8 - Subáreas de conhecimento da mineração de processos. Fonte: o Autor, 2018

As Figuras 9 e 10 fornecem as distribuições dos artigos pelas áreas de conhecimento. As nove principais subáreas são responsáveis por 92,1% das contribuições na área de mineração de processos. Em 27,3% (338 artigos), a área de conhecimento mais ativa é a descoberta de processos, destinada principalmente a compreender atividades de controle de fluxo, desde as primeiras publicações em 1998. A segunda área se concentra na conformidade, verificação de conformidade e análise de auditoria, atingindo 16,1% (198 artigos). A terceira categoria mais relevante está relacionada com melhorias e ferramentas de arquitetura em 9,9%, seguido de métodos de aplicação e *frameworks* práticos e conceituais para conduzir e aplicar mineração de processo em 9,6%, e, em quinto, agrupamento, abstração, e processo variante de controle de propostas em 7,2%. Sexto é o aprimoramento do processo, e ênfase de extensão e desempenho em 6,7%, seguido de minerador organizacional e social, e técnicas preditivas, ambos com 5,7%. Por fim, a qualidade do processo e métricas seguem com 4%. Nas próximas subseções, as subáreas mais ativas serão apresentadas.

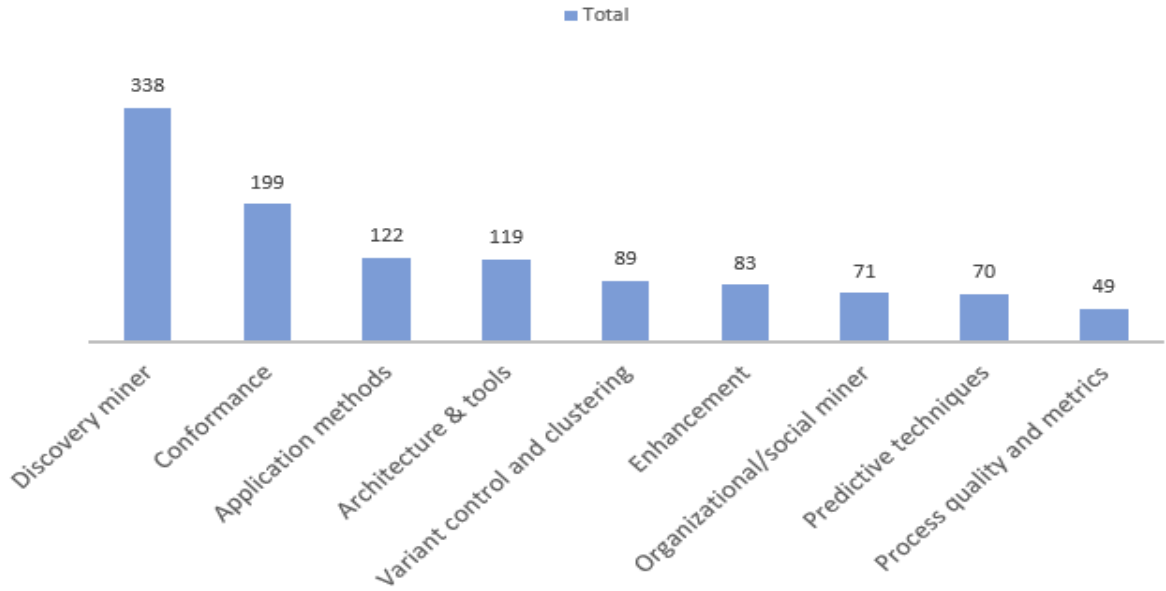


Figura 9 - Distribuição de artigos por categorias de conhecimento da mineração de processos. Fonte: o Autor, 2018

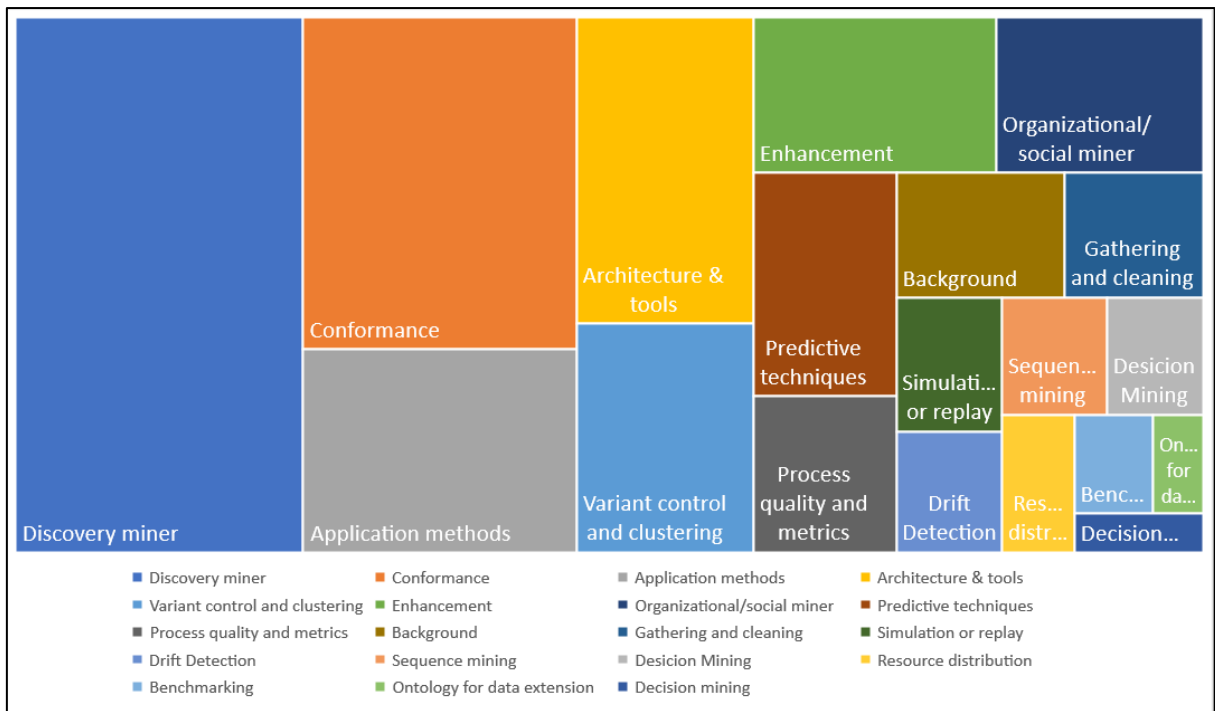


Figura 10 - Mapa da distribuição de artigos por categorias de conhecimento. Fonte: o Autor, 2018

2.2.1 Descoberta de processos

A descoberta de processos é o precursor e a subárea mais explorada da mineração de processos, onde a análise é principalmente dedicada a uma visão *a posteriori* dos *logs* de eventos de execuções de processos de negócios [Weerdt et al. 2012]. O minerador de processo é o termo usado usualmente para a implementação de um *algoritmo* capaz de extrair conhecimentos de *logs* de eventos e criar um modelo de processo. Essa tarefa de descoberta deve entregar um modelo com qualidade, onde se espera que seja compreensível, evitando complexidade desnecessária e, com uma boa precisão, equilibrando o *recall*, a precisão e a generalização [Weerdt et al. 2012].

De acordo com [Medeiros e van der Aalst 2008] a descoberta do processo pode auxiliar os gerentes responder perguntas, como por exemplo: Como os casos estão sendo executados? Qual é o caminho mais frequente para cada modelo de processo?

Diversos algoritmos foram propostos para a descoberta de processos, como o *alpha algorithm* [van der Aalst et al. 2004], *heuristic mining algorithm* [Weijters et al. 2006], *multi-phase mining algorithm* [Van Dongen e van der Aalst 2005], *fuzzy mining algorithm* [Günther e van der Aalst 2007], *genetic mining* [van der Aalst, De Medeiros e Weijters 2005], *region miner* [Carmona, Cortadella e Kishinevsky 2008]. Comparações anteriores relataram que alguns algoritmos envolvem alto custo computacional e várias abordagens são muito difíceis ou inviáveis de aplicar na indústria [Wang et al. 2013].

Os estudos anteriores relatam desafios relacionados à baixa eficiência da descoberta de processos aplicadas em cenários com processos com alto nível de ruído, atividades repetitivas, diferentes tipos e variantes de processos combinados nos mesmos *logs* [Wei e Su 2014]. Este estudo foi aplicado na área de saúde, mas a contribuição pode ser generalizada para outras áreas. Em [Wei e Su 2014] aponta-se para futuros estudos divididos em quatro problemas: análise e identificação das variabilidades de processo; personalização de fluxo com ajustes de acordo com a situação do contexto, disponibilidade de recursos e resultados prováveis; integração com a gestão de processos; e melhorias automatizadas com autoaprendizagem ou recomendações.

Apesar da descoberta de processo ser muito eficaz para capturar o comportamento do controle de um processo, há limitação na descoberta suplementar do modelo de processo

relacionada à identificação de papéis/responsáveis, entrada ou saída de documentos, identificação e generalização dos pontos de decisão [Akman e Demirörs 2009].

2.2.2 Conformidade do processo, *compliance* e auditoria

O monitoramento de processos com base em verificações de conformidade também é um tipo relevante na mineração de processos. Alguns erros são usuais quando se referem à conformidade e *compliance*; estas análises de verificação são semelhantes, mas não são conceitos homogêneos. A principal diferença envolve quando o processo ocorre: a verificação de conformidade é baseada na análise *post mortem* com *logs* de eventos completos (geralmente em um modo *off-line*), enquanto o monitoramento de *compliance* é baseado em *logs* de eventos parciais e em tempo de execução [Ly et al. 2015].

A conformidade pode ser usada para comparar o modelo de referência com a realidade adquirida por meio de *logs* de eventos gerados por sistemas de informação. De acordo com [van der Aalst 2012], a verificação de conformidade pode ser usada para verificar a exatidão dos processos documentados, apontar casos diferentes, e tentar entender o que eles podem ter em comum, para avaliar algoritmos de descoberta de processos e identificar pontos de desvios no modelo de processo. Além disso, para fins de auditoria, a verificação de conformidade pode ser usada para calcular a eficiência de um modelo de processo descoberto e melhorar um modelo novo ou um modelo já existente. A verificação de conformidade é usada em várias ocasiões, tornando-se um dos pilares da mineração de processos.

[Medeiros e van der Aalst 2008] também estabelecem o principal interesse do *c-level* relacionado com a conformidade do processo, por exemplo: Como os casos são complacentes com os modelos de processos implantados? Onde estão os problemas? Quão frequente é a não conformidade? As regras/restrições de um processo estão de fato sendo obedecidas?

A conformidade pode ser aplicada para considerar várias dimensões, como *fitness*, precisão, generalização e simplicidade [van der Aalst 2011b]. Uma revisão sistemática realizada anteriormente por [Ly et al. 2013], identificou 10 funcionalidades relevantes neste tipo de mineração de processos: 1) métricas de tempo; 2) dados do caso; 3) informações de recursos; 4) condições explícitas ou implícitas em atividades não atômicas (associadas a vários eventos que permitam ordens diferentes); 5) dar suporte ao ciclo de vida de uma atividade; 6)

restringir várias instâncias baseadas na métrica de tempo/dados/recursos; 7) detectar e gerenciar violações de conformidade; 8) detectar pro-ativamente e administrar violações; 9) explicar a causa de uma violação; e, 10) quantificar o grau de conformidade. Além disso, em [Ly et al. 2013] realizou-se uma avaliação envolvendo cinco propostas de conformidade (*Mobucon LTL*, *Mobucon EC*, *ECE Rules*, *Supervisory Control Theory*, e *SeaFlows*) com base nessas 10 funcionalidades relevantes.

2.2.3 Aprimoramento, extensão e ênfase no desempenho

Esse tipo de atividade na mineração de processos está focado em ampliar o modelo de processo com informações relevantes. Por exemplo, as aplicações móveis GPS combinam informações *on-line* sobre o tráfego, dando ênfase às ruas congestionadas. A mineração de processos pode executar a mesma função usando o carimbo de data/hora em *logs* de eventos, combinando-os com o modelo de processo para previsões apoiadas em modelos estatísticos ou aprendizagem de máquina.

Os tomadores de decisão nas organizações interessam-se fortemente na contribuição dessa perspectiva de análise de processos, devido ao alto valor agregado em negócios e a percepção facilitada por aprimoramentos de processo. Em [Medeiros e van der Aalst 2008], este tipo de mineração de processos pode dar suporte para responder as seguintes perguntas: Qual é o tempo de processamento médio, mínimo, ou máximo dos casos? Quais caminhos levam mais tempo em média? Quantos casos seguem esses caminhos? Qual é o tempo médio de atendimento para cada tarefa? Quanto tempo foi gasto entre duas tarefas no modelo de processo?

Uma técnica que permite a observação da execução de um processo de negócio é chamada de análise de repetição/*replay*. Esta técnica consiste em uma representação das instâncias de processo dentro de períodos definidos, visualizando a representação sobre o modelo de processo e permitindo a aceleração do tempo para obter *insights* em poucos minutos

sobre o que ocorreu durante um período de semanas. Em [van der Aalst 2012], isto estabelece um significado bem-definido para toda a ênfase visual, tais como:

- O tamanho de uma atividade pode refletir sua frequência ou alguma outra propriedade indicando seu significado (por exemplo, custos ou uso de recursos);
- A cor de uma atividade pode refletir a sua duração ou o seu tempo de serviço;
- A largura de uma conexão/arcos pode refletir a importância de cada dependência causal;
- A coloração das conexões/arcos pode destacar os gargalos existentes; e
- O posicionamento das atividades pode ter um significado bem-definido.

2.2.4 Minerador Organizacional/Social

A pesquisa nesta categoria visa realizar a mineração de estruturas organizacionais, combinando a análise de rede social, o mapeamento de comportamentos de recursos, a colaboração do usuário e a análise de função. A partir da perspectiva de *c-level*, esta categoria estabelece informações organizacionais para apoiar aspectos de identificação e responder perguntas tais como [Medeiros e van der Aalst 2008]: Quantas pessoas estão envolvidas em um caso específico? Qual é a estrutura de comunicação e de dependência entre as pessoas? Quem são as pessoas importantes no fluxo de comunicação? Quem trabalha nas mesmas tarefas?

2.2.5 Métodos de Aplicação

Muitos métodos foram propostos para orientar as aplicações de mineração de processos e combinar técnicas de mineração de processos com outros métodos. Em [van der Aalst, 2011c] foi proposto o modelo *L* life-cycle model*. Esse modelo se desenrola em cinco fases para conduzir uma aplicação de mineração de processos:

- Planejar e justificar a fase—focar orientando-se a dados para explorar e responder a curiosidades e descobrir *insights*; ou, orientado a perguntas focadas em explicar uma situação específica, por que determinado problema ocorre; ou, focada no objetivo de melhorar um KPI, o desempenho em termos de tempo de resposta, explicar um desvio e reduzir custos.

- Extrair eventos relevantes—explorar o domínio de dados do sistema de informação para recuperação de eventos relevantes. Às vezes, isso pode representar uma complexidade de sistema devido à demora, com milhares de registros espalhados entre muitas tabelas.
- Criar um modelo de fluxo de controle e conectá-lo ao *log* de eventos—os algoritmos mineradores de processo podem ser usados para obter um modelo de processo. Quando um modelo de processo compreensível é obtido, é possível iniciar a verificação de conformidade, analisando atividades e desvios do modelo de processo descoberto.
- Criar modelo de processo integrado—aprimoramentos e informações adicionais são integrados ao modelo, promovendo novas perspectivas relacionadas ao tempo, organização, utilização de recursos, comparações de casos e simulação, entre outras possibilidades.
- Suporte operacional—envolve detectar, prever e recomendar. É o nível mais avançado de suporte computacional, por exemplo, a ferramenta de mineração de processos deve ser capaz de gerar alertas em casos de desvios, fornecer aconselhamento sobre gargalos, recomendar configuração de recursos ou realocação, etc.

Com o objetivo de estender o ciclo de vida tradicional dos processos de negócio (BP), [Delgado et al. 2014] propuseram o BPEMM (BP Execution Measurement Model) para suportar a melhoria contínua. Seu trabalho analisa e fornece uma visão da execução de um processo de negócio real. Aqui, o processo de mineração pode oferecer técnicas que são muito úteis. O modelo proposto não está limitado a apenas a obtenção de *insight*, ele também pode ajudar em muitas fases do processo de negócio, por exemplo, na avaliação de desempenho.

Muitas propostas de métodos relevantes focadas em domínios de aplicação específicos, por exemplo, proposto em [Mundbrod, Beuter e Reichert 2015]. Trata-se de um método focado no suporte sistemático de processos de negócios envolvendo colaboração e coordenação mais eficazes entre os colaboradores. Foi validado usando projetos de desenvolvimento para componentes elétricos e eletrônicos. Em [Mans 2013] aplicou-se um método baseado em mineração de processos e simulação de eventos discretos na área saúde para avaliar o impacto das tecnologias digitais. Mais precisamente, em [Mans 2013] avaliou-se os esforços de

desempenho em odontologia para pacientes, dentistas e laboratórios odontológicos de acordo com os cenários de práticas tradicionais e uso de novas tecnologias. O processo de mineração combinado com a simulação de eventos discretos, permite iniciá-lo mais rapidamente em comparação com a abordagem de simulação tradicional, onde os modelos de simulação são criados manualmente.

2.2.6 Arquitetura e Ferramentas

Nesta subseção, o objetivo é verificar os trabalhos publicados para o desenvolvimento de diferentes ferramentas para apoiar e melhorar os processos de negócios. Isso envolve muitas empresas em desenvolvimento, movimentando um alto valor financeiro, grandes parceiros de serviços de consultoria, e dependentes de infraestruturas complexas, robustas e dispendiosas. Dentre essas ferramentas, podemos citar a gestão de processos de negócios (BPMS), análise de processos de negócio (BPA), inteligência de negócios (BI), gerenciamento de desempenho empresarial (EPM) e BAM, entre outros. Desde os anos 90, as ferramentas chamadas sistemas de gerenciamento de workflow (WFMS) têm dado suporte a execução de processo e controle. No entanto, WFMS oferece baixo suporte para identificar instâncias de processos, medir as tarefas executadas ou permitir o gerenciamento de processos de alto nível. BPMS emergiu como uma evolução da WFMS para proporcionar maior qualidade no desempenho de processos, diagnósticos e controle de recursos [Ko 2009]. Ferramentas de BI (Business Intelligence) são úteis para monitorar o desempenho de um dado processo. Essas ferramentas são geralmente centradas em dados, oferecendo pouco ou nenhum suporte a mineração de processos, que é centralizado no processo [Aalst 2016]. Trabalhos anteriores, como o estudo de [Guarda et al. 2013], considerou a integração entre ferramentas de BI e mineração de processos com o objetivo de fornecer flexibilidade, análise quantitativa, explicações de estatística, e para prever a tomada de decisão.

Na última década, muitas ferramentas surgiram para apoiar a mineração de processos. No entanto, a maioria das publicações utilizam o *framework ProM*¹, uma estrutura robusta e código aberto [van Dongen 2005]. O *ProM* oferece arquitetura robusta, e é uma das razões pelas quais a comunidade de pesquisadores adotam esse *framework*. O *ProM* permite flexibilidade

¹ <http://www.promtools.org>

no desenvolvimento de novos algoritmos (*plug-ins*), estendendo e combinando com formatos de entrada e saída padronizados. No entanto, existem outras opções, como o *Aris Process Performance Manager* da Software AG², *Celonis Discovery*³, *Disco* da Fluxicon⁴, *Myivenio* da Cognitive Technology⁵, *Perceptive Process Mining* da Lexmark adquirido por Hyland Software⁶, *Process Gold*⁷, *QPR ProcessAnalyzer*⁸, *SNP Business Process Analysis*⁹, *BAB Framework*, *UpFlux*¹⁰ e outros. Entre este conjunto de ferramentas, observou-se que poucos fornecem recursos para o suporte operacional como um controle de processo ativo ou como uma integração nativa para provedores BPMS.

Outros trabalhos de pesquisa focaram na definição de um formato padrão para questões abstratas relacionadas à heterogeneidade das fontes de dados. Foi proposta uma abordagem para adquirir, trocar e analisar *logs* de eventos, e esse padrão é chamado de fluxo de eventos extensível (XES) [IEEE Standards 2016]. Um esquema XML que descreve a estrutura e os padrões de extensão também está disponível como uma implementação de referência em Java, chamado *OpenXES*. Um processo de certificação também foi desenvolvido para melhorar a portabilidade entre todas as ferramentas de mineração de processos comerciais e acadêmicos.

2.2.7 Métricas de Qualidade e Similaridade

O processo descoberto e os filtros aplicados devem equilibrar as perspectivas de qualidade. As dimensões básicas de qualidade descritas por [Aalst 2016] são:

- *Fitness/Adequação*: capacidade de observar o comportamento do log de eventos no modelo descoberto;
- *Precisão*: qualidade para evitar comportamentos não relacionados ao *log* de eventos usados no processo de descoberta, caso contrário ele é caracterizado como *underfitting*;

² https://www.softwareag.com/corporate/products/aris_alfabet/bpa/aris_ppm/default

³ <https://www.celonis.com>

⁴ <https://fluxicon.com/disco/>

⁵ <https://www.my-invenio.com/>

⁶ <https://www.hyland.com/en/perceptive>

⁷ <https://processgold.com/>

⁸ <https://www.qpr.com/products/qpr-processanalyzer>

⁹ <https://bpm-expo.com/tool/snp-business-process-analysis/>

¹⁰ <https://upflux.net>

- Generalização: o modelo não deve restringir-se a comportamentos muito específicos observados no *log*;
- Simplicidade: qualidade para ser o mais simples possível.

A métrica observada mais dominante é o *fitness* que resulta em um valor entre [0; 1]. 0 para *fitness* fraco e 1 para *fitness* perfeito. Outras três dimensões de qualidade foram exploradas por [Rozinat et al. 2009], que propõe a adequação com foco na ideia da “navalha de *Occam*”. Este trabalho também distingue a adequação estrutural e comportamental, i.e., o foco da adequação estrutural no modelo simples que pode explicar o *log* (maneiras sintáticas de expressar o mesmo comportamento, evitando duplicidade), e a adequação comportamental tenta ser muito genérica e permitir demais comportamentos. As semelhanças comportamentais ou estruturais requerem uma abordagem de mapeamento para identificar cada atividade nos modelos. Para perceber isso, outras técnicas geralmente são combinadas para reconhecer uma similaridade sintática e linguística para identificar cada atividade. [Dijkman et al. 2011].

Em uma revisão sistemática recente, as semelhanças de processos foram organizadas em quatro dimensões [Schoknecht et al. 2017]: linguagem natural; dimensão da estrutura; dimensão de comportamento; e, dimensão de estimativa humana.

2.2.8 Controle de Variantes do Processo e Agrupamento

Como na análise de conformidade, o agrupamento requer uma condição correspondente, como: aplicação lexical ou/e medidas de similaridade semântica em eventos ou rótulos de tarefas; correspondência estrutural com base na topologia dos modelos de processos; e, a correspondência comportamental baseada na verificação da semântica de execução de modelos de processo. Os melhores mineradores de processo descobrem um modelo de processo com alta precisão, mas é comparável a *overfitting*, resultando assim em um processo complexo e não estruturado, por exemplo, um modelo de espaguete.

Com o objetivo de lidar com essa complexidade, uma das primeiras abordagens que propõem a ideia do processo de agrupamento foi [van der Aalst e Günther 2007]. Este trabalho teve como objetivo equilibrar a abstração e agregação, aplicando três etapas: preservar comportamentos altamente significativos, agregar os menos significativos com comportamentos altamente correlacionados, e remover os menos significativos e pouco

correlacionados. No entanto, este trabalho visava extrair uma visão geral de processo não estruturado, sem considerar a visualização de aprimoramento de cada variante do processo relevante. Esta primeira combinação de agrupamento e mineração de processos está disponível na implementação do *fuzzy miner* do *framework ProM*. A mineração de processos pode facilmente acomodar ideias relacionadas a um grupo de variantes de processo, onde verifica-se a variante de um processo específico de acordo com a frequência, o tempo de execução ou outro KPI de processo relevante orientando a melhoria do processo.

Focado na extração de agrupamentos de acordo com variante de processo similar, muitas das técnicas de agrupamento existentes não lidam com a qualidade dos modelos de processo durante o procedimento de agrupamento e isso impacta significativamente em um viés de agrupamento [Weerdt et al. 2013]. A maioria dos algoritmos de descoberta de processo são focados na geração de modelos de adequação ou precisão com simplicidade, no entanto, para comportamentos distintos é mais relevante gerar modelos separados para facilitar a análise, diagnóstico e melhorias [San Pedro e Cortadella 2016]. O algoritmo *ActiTraC* foi proposto por [Weerdt et al. 2013] considerando as medidas na amostragem. É classificado como não supervisionado; foi utilizada uma estratégia de amostragem seletiva baseada principalmente em medições de frequência.

Em [Prodel et al. 2018] foi proposto um novo método para descobrir modelos de processos a partir de *logs* de eventos complexos. As contribuições mais significativas iniciaram com uma representação compacta, hierárquica e flexível de classes de eventos que fornecem agregações significativas e, em seguida, a reprodutibilidade foi melhorada usando um método original baseado em uma estratégia de amostragem de *Monte Carlo*. Essa abordagem permitiu diminuir a complexidade computacional (tamanho do *log* de eventos) e, finalmente, para otimização de modelo de processo, o *tabu search algorithm* foi usado para explorar vizinhos promissores. O método foi testado em uma ampla variedade de *logs* de eventos reais e sintéticos. O método proposto supera tanto as estratégias de criação de processos aleatórios quanto o algoritmo de mineração de processos, *fuzzy miner* implementado no software *DISCOTM*. A proposta de [Prodel et al., 2018] alcançou resultados relevantes combinando algoritmo hierárquico de agregação e otimização para equilibrar a fidelidade do modelo de processo e sua complexidade.

2.2.9 Técnicas preditivas

Muitas abordagens preditivas foram combinadas com a mineração de processos, desde indutores *bayesianos*, árvores de decisão, raciocínio baseado em casos, sistemas de recomendação, redes neurais, entre outras abordagens. Os objetivos da predição são múltiplos, por exemplo, prever a duração de uma instância, identificar atrasos, classificar a causa raiz, prever a próxima atividade e recomendar o melhor caminho, entre outros.

Uma árvore de decisão conhecido como *TILDE* foi usado por [Vasilyev, Ferreira e Iijima 2013] para classificar instâncias de processo em diferentes grupos de acordo com sua duração. Ela é baseada em programação lógica indutiva (ILP), prevendo a duração de uma instância de processo, i.e., a diferença entre os registros de data e hora do último e do primeiro evento registrado para aquela instância, conforme registrado no *log* de eventos.

A predição do tempo de conclusão das instâncias de um processo em execução foi pesquisada por [Polato et al. 2014]. A predição é baseada em duas etapas: prever o caminho do fluxo de controle usando estatística (*Naive bayes*); e por outro lado, todos os dados registrados por cada atividade são coletados e usados para refinar a predição. A técnica foi implementada e avaliada com um estudo de caso real, obtendo resultados significativos de predição.

[Lakshmanan et al. 2015] utilizou algoritmo de classificação de árvore de decisão para determinar a probabilidade de executar da próxima tarefa em uma determinada instância do processo. Usando um modelo de processo probabilístico, a árvore pode ser transformada em uma cadeia de *Markov* estendida para calcular a probabilidade de execução de qualquer tarefa futura em uma instância de processo de execução, oferecendo suporte à execução paralela. O método foi aplicado na validação de seguros de automóveis.

2.3 Algoritmos de Descoberta de Processos

A primeira iniciativa para responder a essa pergunta foi localizar o primeiro trabalho publicado sobre cada algoritmo. A Figura 11 apresenta uma visão geral dos mineradores de processos e lista alguns dos algoritmos mais relevantes, incluindo autor e ano.

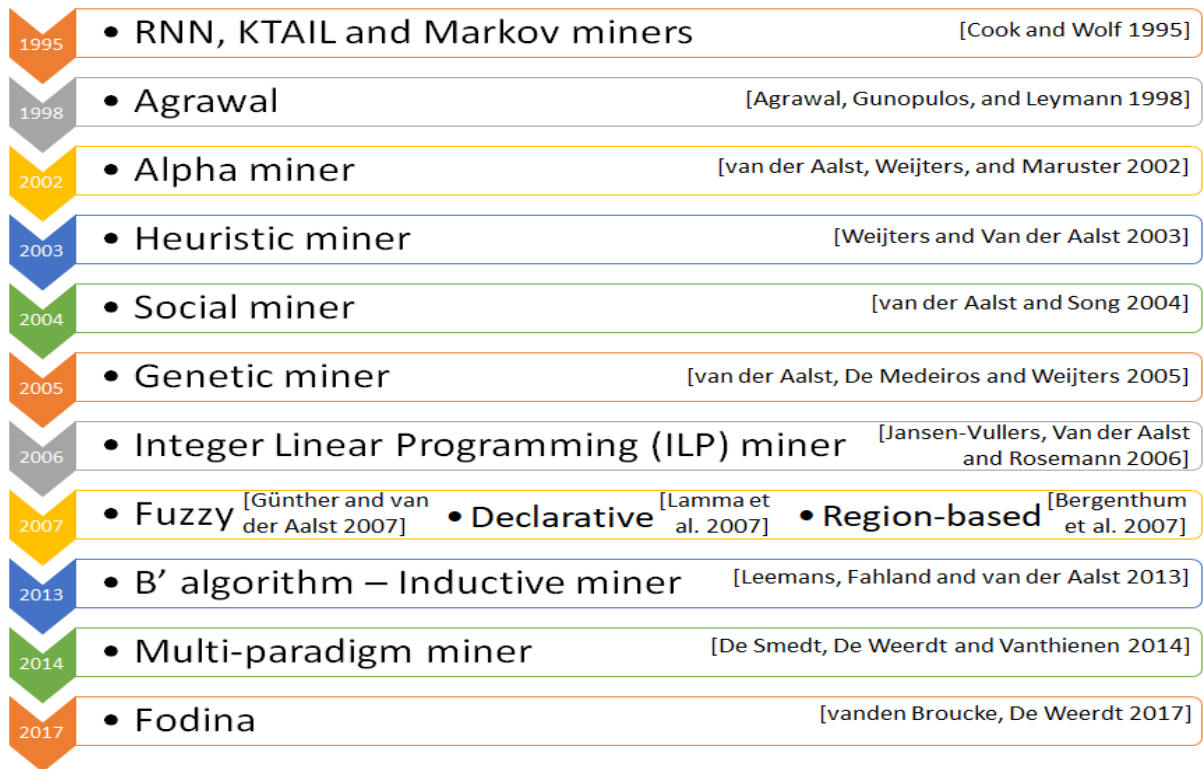


Figura 11 - Algoritmos de Mineração de Processos. Fonte: o Autor, 2018

O gráfico da Figura 12 mostra os algoritmos de mineração de processos mais utilizados. É possível observar que o *Heuristic miner* e o *Fuzzy miner* são os mais utilizados em casos aplicados. Uma possível explicação para esse resultado é a capacidade desses algoritmos lidarem com ruídos e exceções em processos não estruturados [Song et al. 2013], [Günther e van der Aalst 2007].

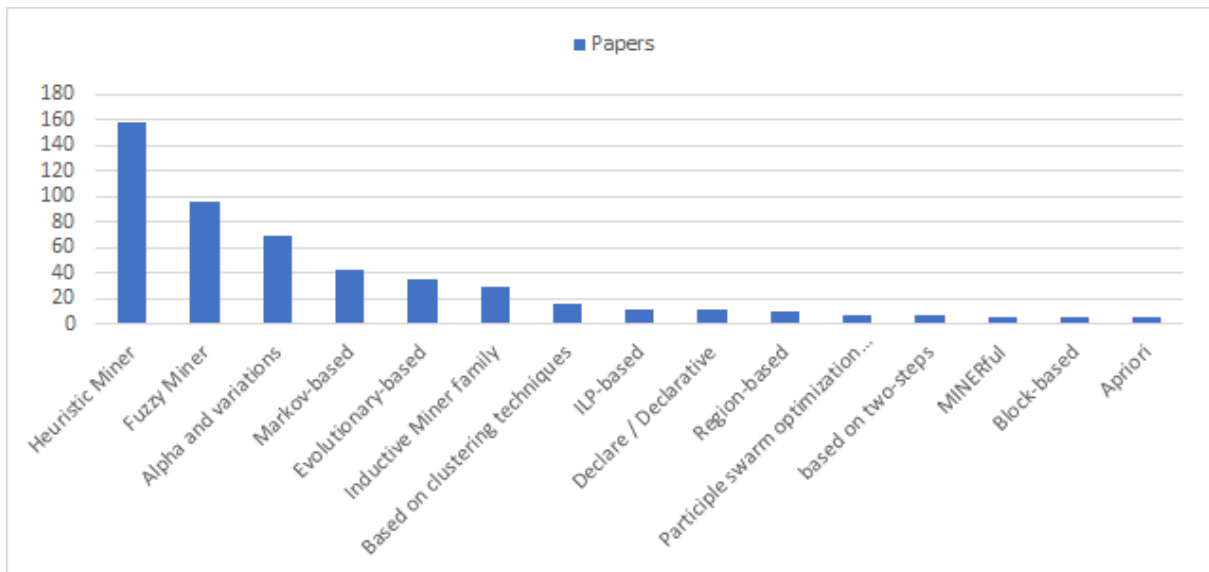


Figura 12. Distribuição do uso dos algoritmos de mineração de processos. Fonte: o Autor, 2018

Entre as ferramentas mais utilizadas, está o *framework ProM* citado em mais de 408 artigos, seguido por *DISCO* da *Fluxicon* em 82 artigos e *Inductive Workflow Learning via Examples (InWolvE)* em 6 artigos. Outras ferramentas também foram aplicadas em algumas pesquisas, mas o número é baixo em trabalhos publicados.

2.4 Aplicações da Mineração de Processos

Esta seção destina-se a responder à pergunta: Onde a mineração de processos está sendo aplicada? Os domínios de aplicação da mineração de processos explorados são bastante amplos. As aplicações mais relevantes estão no setor de saúde, hospitais e caminho clínico. A segunda mais relevante é a tecnologia da informação focada em desenvolvimento de software, manutenção e outros serviços de operação. A manufatura e casos de indústria ocuparam a terceira posição, seguida pela educação e instituições financeiras. A Tabela 5 lista os domínios de aplicativo e o gráfico da Figura 13 apresenta as áreas de aplicação mais ativas de acordo com o número de publicações.

Tabela 5. Domínios de aplicação. Fonte: o Autor, 2018.

Classificação	Descrição
Saúde	O domínio da aplicação que cobre o caminho clínico, tratamento do paciente, ou processos preliminares de um hospital.
TIC	Tecnologia de informação e comunicação, relacionada ao desenvolvimento de software, serviços de TI e empresas de telecomunicações.
Industrial	Aplicação em atividades industriais, realizada por uma fábrica que geralmente recebe material e entrega produtos semiacabados ou acabados. Dentre todos os segmentos industriais o mais significativo é o segmento automotivo.
Educação	Aplicação em educação, <i>e-learning</i> , aplicações científicas e centros de pesquisa com gestão de processos de inovação.
Financeira	Domínio de aplicação relacionado a bancos, seguros, em diferentes processos, tais como pagamentos, investimentos, depósitos, análise de riscos e mitigação, transferências e outras operações financeiras.
Logística	Aplicação relacionada à logística, transporte, armazenamento e gestão de estoques.
Público	Administração pública, governo, serviços municipais, serviços de segurança social e serviços postais.
Segurança	Serviços de segurança em TI ou relacionados com protocolos de processo, tais como restrições de segurança marítima.
<i>Call Center</i>	Assistência técnica e serviços ao cliente para resolver problemas relatados ou responder dúvidas.
Usabilidade	Aplicação para compreender e melhorar a experiência do usuário focada em <i>design</i> , caminho de navegação, <i>websites</i> e melhorias de aplicativos móveis e <i>e-marketing</i> .
Robótica/ <i>Smart</i>	Novos aplicativos usando tecnologias avançadas relacionadas a cenários, como edifícios inteligentes, indústria 4.0 e robótica.
Entretenimento	Aplicações para entretenimento, esportes, jogos, publicidade e mídia.
Utilitário	Este segmento envolve a geração de energia ou a gestão de recursos naturais, tratamento de água, equilíbrio ou entrega.
Vestuário	Aplicações para os produtores de roupas, como tingimento e têxteis.
Consultivo	Serviços de consultoria em assessoria, auditoria e impostos.
Varejo	Aplicações focadas na oferta e venda de produtos.
Biologia	Processos em sistemas biológicos.
Varejo	Lojas físicas ou virtuais.
Farmácia/biologia	Relacionadas com aplicações biomédicas e análises biológicas.
Hotel	Segmento focado em atendimento ao cliente, quartos e hospitalidade.
Agricultura	Agricultura e indústria alimentar.

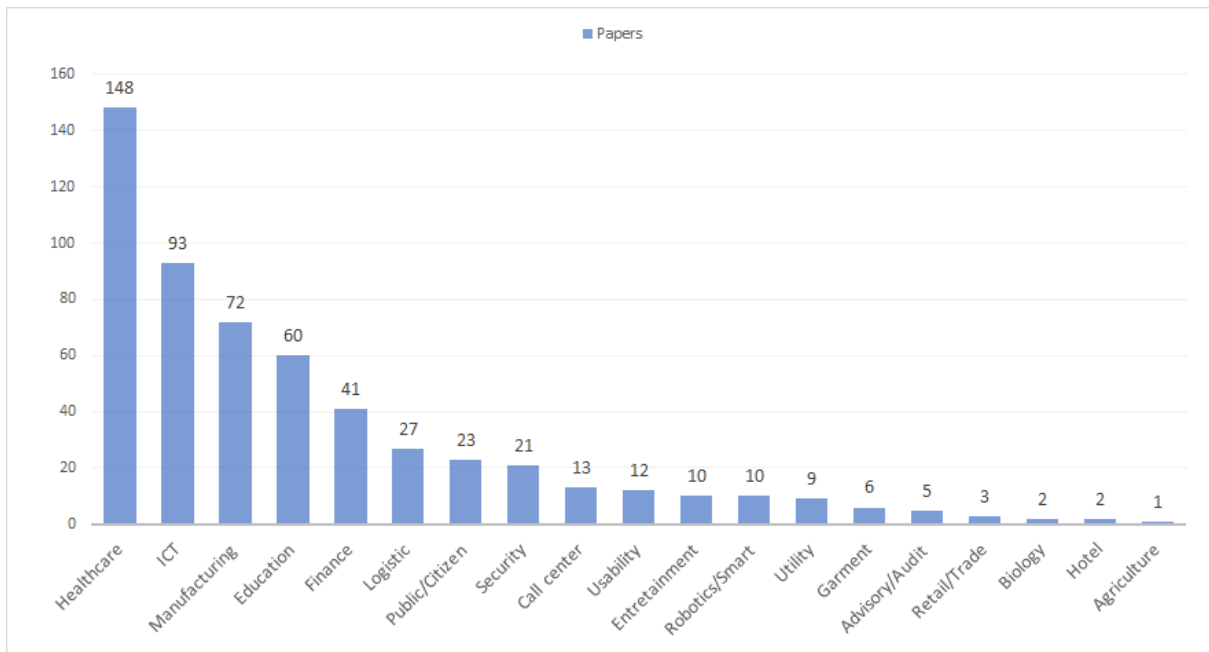


Figura 13. Número de artigos por domínio de aplicação. Fonte: o Autor, 2018

2.4.4 Saúde

Esta seção apresenta aplicações da mineração de processos focadas no segmento de saúde, com base em 147 artigos selecionados, relatando contribuições relevantes e limitações no desenvolvimento de soluções para este domínio de aplicação. Muitos estudos relacionados indicam que os modelos de processo em saúde diferem de outros em relação às suas características, alta variabilidade, complexidade, segurança, privacidade, natureza multidisciplinar de suas atividades, exames inovadores e modernos, tratamentos, ou procedimentos, entre outros ([Rebuge e Ferreira 2012], [Mans, van der Aalst e Vanwersch 2015], [Munoz-Gama e Echizen 2012]). Vale mencionar que a implementação do processo de mineração em saúde é um campo relativamente novo e seu uso pode ajudar na tomada de decisões, redução de custos, identificação de fluxos de pacientes para certas doenças, chances de tratamento, maximização do fluxo com bons resultados, correlações entre os tratamentos administrados, a qualidade dos tratamentos e as complicações, entre outros ([Rebuge e Ferreira 2012], [Rojas et al. 2016]).

Para a melhoria dos cuidados hospitalares, a área de saúde foi equipada com sistemas de informação eficientes para registros clínicos e administrativos de dados do paciente. Estes

dados recolhidos foram explorados para ajudar os gestores na tomada de decisões, com o objetivo de proporcionar serviços mais eficientes e eficazes [Mans, Aalst e Vanwersch 2015] e [Mans et al. 2013]. A mineração de processos tem sido uma das técnicas que se favorece dos dados registrados em sistemas de informação hospitalar [Paster e Helm 2015] e [Huang et al. 2015]. Por meio da aplicação da mineração de processos, as organizações hospitalares podem avaliar como os processos são conduzidos, verificar se certas diretrizes clínicas e protocolos médicos foram realmente seguidos, e obter *insights* sobre gargalos, utilização de recursos, e outros aspectos de processos relacionados ao desempenho [Rovani et al. 2015] e [Cho et al. 2015].

Na área de cuidado da saúde existe um termo bastante consolidado, o caminho clínico. Um percurso clínico pode ser considerado como um processo específico, de acordo com a doença do paciente, que descreve o fluxo de atendimento deste paciente [Huang et al. 2013]. Inúmeros estudos mostram a aplicação da mineração de processos como uma ferramenta útil para verificar as vias clínicas adotadas em configurações hospitalares ([Caron et al. 2014], [Fernandez-Llatas et al. 2015], [Huang et al. 2014], [Bettencourt-Silva et al. 2015], [Baker et al. 2017]).

Outra área da saúde que se beneficiou da aplicação da mineração de processos é a chamada medicina de emergência ([Kim et al. 2013], [Basole et al. 2015], [Garcia et al. 2015], [Rojas et al. 2017], [Alvarez et al. 2018]). Isto é definido como a especialidade médica dedicada ao diagnóstico e tratamento de doenças imprevistas e lesões, incluindo a avaliação inicial do paciente para verificar a complexidade da sua saúde, diagnóstico, tratamento e coordenação dos cuidados entre vários provedores à disposição de qualquer paciente que necessite de cuidados médicos cirúrgicos ou não cirúrgicos [Alvarez et al. 2018].

Há diversas oportunidades para aplicações de mineração de processos em saúde, permitindo que especialistas, por meio de registros de eventos de pacientes, gerenciem o modelo de processo, avaliem a conformidade com diretrizes e protocolos clínicos e identifiquem pontos de melhorias de processos.

2.4.5 Tecnologia da Informação e Comunicação

Foram identificados 83 artigos voltados ao segmento de TI, focados em diferentes objetivos como melhoria de processos, avaliação de processos, gerenciamento ágil de tarefas, verificação de conformidade em operação e gerenciamento de serviços, entre outros.

Em ([Samalikova 2012], [Riera Cruañas 2012], [Rubin et al. 2007]) é apontado o potencial promissor da mineração de processos para apoiar avaliações de processos de software. Estes trabalhos utilizam registros de sistemas de informação para análise e avaliação de processos de software. As oportunidades de análise resultantes são baseadas em considerar as diretrizes CMMI (integração do modelo de maturidade de capacidade) para avaliação. Além desta perspectiva, [Valle et al. 2016] usou mineração de processos para apoiar a avaliação *SCAMPI* (método padrão de avaliação CMMI para a melhoria do processo).

Em [Gupta 2014] foi proposto a aplicação de técnicas de mineração de processos para derivar modelos de processo sobre tempo de execução de softwares, para identificar e remover desvios, para estender os recursos das ferramentas de engenharia de software existentes, para tornar as partes interessadas mais conscientes do processo, para entender o padrão de interação e para melhorar a eficiência de um projeto de software. Em [Gupta et al. 2017] também foi proposto a melhoria do processo de manutenção de software por meio da análise de registros usando a mineração de processos.

2.4.6 Industrial

No segmento industrial, foram selecionados 72 artigos. Devido à crescente presença de sistemas de informação, tais como planejamento de recursos empresariais (ERP), sistemas de execução de manufatura (MES), controle de supervisão e aquisição de dados (SCADA). Vários estudos exploraram o processo mineração usando *log* de eventos nesta área. Um dos primeiros estudos de pesquisa nesta área foi relatado por [Ho and Lau 2006], cujo qual focou em melhorar um processo de produção discreto para descobrir relacionamentos ocultos e fornecer sugestões a partir de uma grande quantidade de eventos. Os métodos propostos ofereceram maior flexibilidade na gestão do processo de produção com capacidade de apoio à decisão e melhoria em tempo real. Em [Rozinat et al. 2009] aplicou-se a mineração de processos sobre registros de

dados de equipamentos industriais. Neste artigo, foram analisados registros de eventos de um fabricante de *chips* e o desafio foi analisar um processo menos estruturado, quando comparado aos processos administrativos.

Outra aplicação recente foi realizada por [Ruschel, Santos e Loures 2017], que integra a mineração de processos com modelos preditivos de rede *bayesiana* para determinar intervalos de manutenção para equipamentos industriais. A ideia é fornecer uma ferramenta para auxiliar os gerentes na tomada de decisões sobre paradas de equipamentos de manutenção, reduzindo o tempo de inatividade e evitando paradas não planejadas.

2.4.7 Educação

Muitas aplicações de mineração de processos visam descobrir, monitorar e aprimorar os processos educacionais, representando aproximadamente 60 artigos. Durante o processo de aprendizado, é possível analisar o conjunto de comportamentos recorrentes que podem ser encontrados nos eventos. Alguns trabalhos se concentram na busca de caminhos de adequados ao aprendizado e de conteúdo com base em um grupo de perfis de usuário, detectando estilos de aprendizado, identificando tendências em estudos de aprendizado *online*, recomendando um caminho de aprendizado para melhores resultados, analisando as relações entre estudantes, apoiando instrutores para aumentar sua conscientização para a colaboração dos alunos, para diagnosticar gargalos, analisar o desempenho, verificar a conformidade e monitorar os alunos, entre outros.

2.4.8 Finanças

Neste domínio de aplicação focou-se no segmento financeiro, com ênfase em bancos, seguradoras, etc. Apesar de todas as oportunidades da mineração de processos em auditoria e da grande importância desse segmento, identificou-se apenas 41 artigos. Os documentos abrangem diferentes aplicações, como análise de risco, tratamento de sinistros, validação de regras contratuais de segurança, detecção gradual de desvios em perfis, auditoria financeira, fraude e análise, aprovação de empréstimos, cheques, cartão de crédito, entre outros.

Em [Lakshmanan et al. 2015] foi proposto um modelo de processo probabilístico usando métodos de *Markov*. Ele foi aplicado em uma companhia de seguros de automóveis com o

objetivo de simular sinistros de seguro que lidam com processos de negócios semiestruturados. A abordagem proposta, usando processos combinados de mineração, resultou em previsões mais precisas do que as abordagens anteriores.

2.4.9 Outros Domínios de Aplicações

Mais de 100 artigos foram publicados em domínios como: serviços públicos municipais, camadas governamentais; serviços de segurança; *call centers* para serviços ao cliente; usabilidade e melhorias da experiência do usuário; entretenimento, esportes e mídia; significativas propostas tecnológicas focadas na robótica, indústria 4.0 e edifícios inteligentes; vestuários; assessoria e auditoria; biologia; hotéis e da agricultura. As aplicações de mineração de processos são diversas e extensíveis a uma ampla gama de negócios.

2.5 Considerações sobre o Capítulo

Neste capítulo um mapeamento sistemático estabeleceu questões de pesquisa formuladas na Seção 2.1.1. No que diz respeito a RQ1: “Como classificar os estudos primários em áreas de conhecimento?” Iniciou-se mapeando os principais tipos de mineração de processos: descoberta—focada em obter um modelo de processo realista baseado em *logs* de eventos; Conformidade—comparação de um modelo de processo (*a priori* ou descoberto) e registro de eventos e vice-versa; Aprimoramento—estender resultados com ênfase em frequências, tempo de trabalho, tempo de espera, gargalos. Além dos principais tipos, sugeriu-se promover a mineração social como um tipo de mineração de processos. A descoberta de processos é a área mais ativa, mas a conformidade, desde 2011, está tomando grande atenção e importância, conforme apresentado na Figura 14.

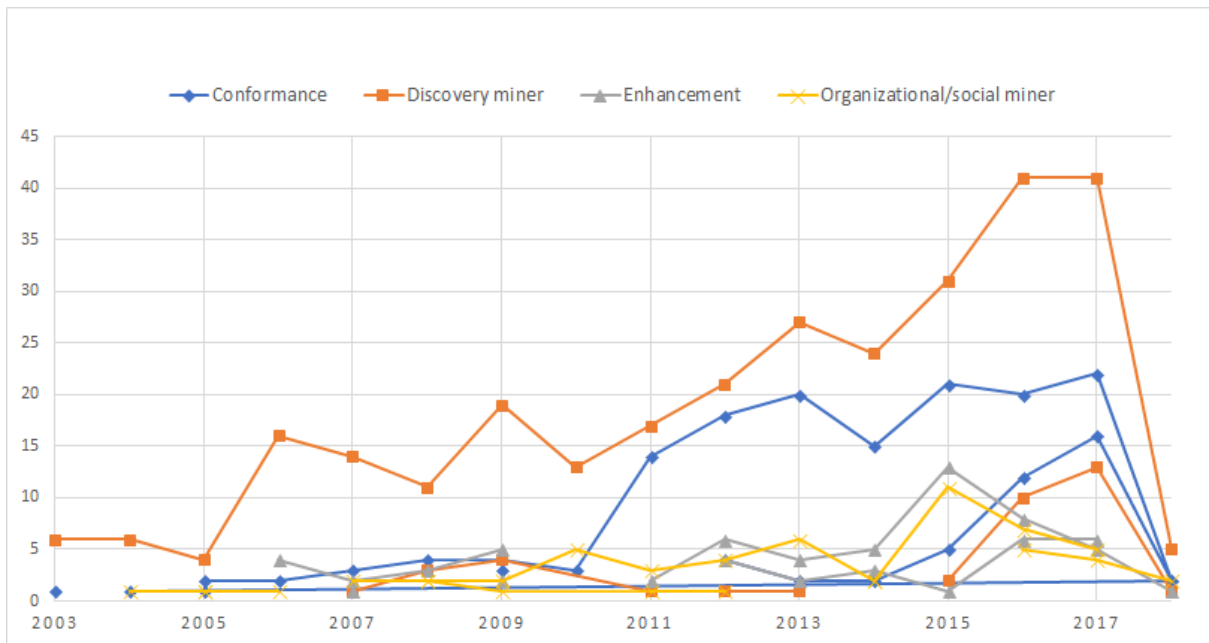


Figura 14. Publicações por ano dos tipos de mineração de processos. Fonte: o Autor, 2018

Foram identificadas cinco categorias: métodos de aplicação, arquitetura e ferramentas, coleta e limpeza de eventos, qualidade de processo e métricas e, ontologia. Como avanços na mineração de processos, mapeou-se categorias como: agrupamento—aplicado no controle de variantes de processos; métodos de simulação—para produzir distribuições, projeções, impacto de risco e incerteza; distribuição de recursos—equilibrando recursos humanos ou máquinas; preditivo—combinando técnicas de aprendizado de máquina; mineração de decisão—mapeamento de regras de entradas no modelo de processo; e, detecção—para identificar um novo conceito ou novos fatos ao longo do tempo.

Na questão RQ2, identificou-se diversos algoritmos. Entretanto, foi possível mapear apenas as principais propostas. Sobre essas últimas, geralmente são sugeridas melhorias, bem como um conjunto de aplicativos que usam esses algoritmos. Por fim, para o RQ3, procurou-se onde a mineração de processos foi aplicada. Identificou-se que a mineração de processos já foi utilizada entre diferentes domínios de aplicações e continua aumentando conforme apresentado na Figura 15.

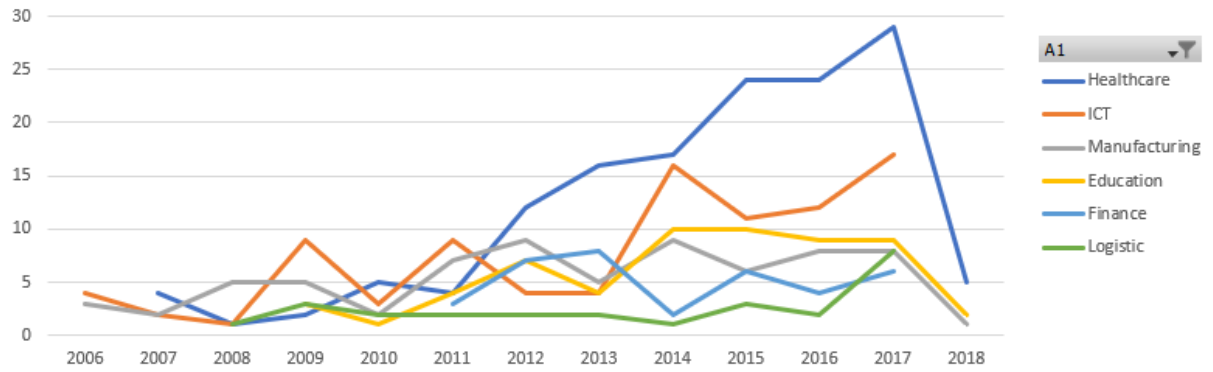


Figura 15 - Publicações por ano dos principais domínios de aplicação. Fonte: o Autor, 2018

O próximo capítulo irá apresentar as abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento.

CAPÍTULO 3 - TÉCNICAS DE AGRUPAMENTO

As técnicas tradicionais de mineração de processos possuem dificuldades em lidar com *logs* de eventos de processos não-estruturados, por exemplo, na área da saúde os *logs* de eventos constituem-se de um grande volume dados, não-estruturados e complexos, gerados a partir de uma grande diversidade de processos e diferentes sequencias de atividades [Song et al. 2013], i.e., embora exista um processo é possível desviar-se dele ou ignorá-lo, gerando modelos processos não-estruturados e flexíveis, o qual também é conhecido na literatura como modelo de processo *espaguete*, onde acaba-se gerando um modelo de alta complexidade de compreensão (cf. Figura 16a).

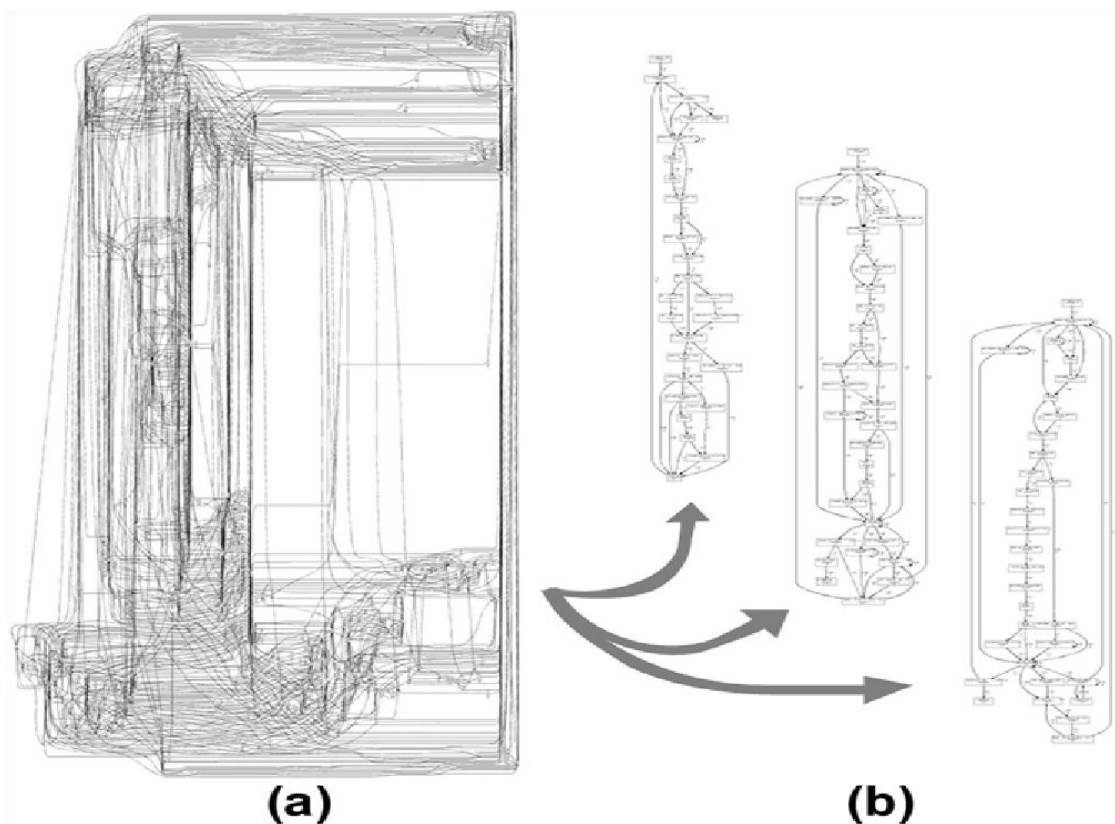


Figura 16. Resultado de um modelo de processo com o agrupamento. [Song et al.2013]

Com o objetivo de gerar modelos de processos mais compreensíveis e exatos (cf. Figura 16b), o agrupamento de instâncias de processos é aplicado como etapa antecessora a mineração de processos, gerando agrupamentos coesos de instâncias de processos e representados, na sequência, por modelos de processos [Bose e van der Aalst 2009].

Diversas técnicas de agrupamento foram desenvolvidas e utilizadas em abordagens para descobrir múltiplos modelos de processos. Todavia, ainda não foi estabelecida qual é a técnica capaz de gerar os modelos de processos mais compreensíveis e exatos a partir de *logs* de eventos de processos não-estruturados e flexíveis, i.e., resolvendo assim uma limitação da mineração de processos.

A seguir são contextualizadas as abordagens para descobrir múltiplos modelos de processos com o suporte de técnicas de agrupamento.

3.1 Abordagens

Focado em logs de agrupamento de traços, cada variante de um processo é separada em um modelo de processo distinto [Greco et al. 2006]. Uma abordagem de espaço vetorial considerou o *log* de eventos para extrair de cada traço e transformá-lo em um conjunto de vetores. A transformação pode ser para valores binários—0 ou 1—para a presença de cada atividade; ou, numérica—usando a frequência/ocorrência de uma atividade. Cada variante de processo resultou em um esquema disjuntivo de fluxo de trabalho (DWS). O conjunto de esquemas disjuntivos de fluxos de trabalho permite a compreensão do processo como um todo. Ele estende o problema de descoberta máxima de processo (MPD) para um conjunto de processos. O agrupamento hierárquico foi aplicado como tarefa antecessora na descoberta do processo, resultando em um esquema disjuntivo de fluxo de trabalho. A abordagem foi baseada no algoritmo *k-means*, operando sobre um espaço vetorial.

Para uma aplicação que mapeou elementos de processo em um processo de desenvolvimento [Santos, Oliveira e Abreu 2015], usou o *plugin* DWS ProM para separar processos complexos em grupos mais simples com fluxos similares. Realizou-se, como passo seguinte, uma verificação de conformidade considerando uma extensão do BPMN, denominada BPMNt. Essa extensão integra recursos de variação da especificação de meta-modelo da engenharia de processo de software (SPEM) à BPMN. Depois de mapear a variação do

processo, o método requer avaliação de especialista para incluir tal variação no processo padrão para dar suporte à adaptação de outros projetos. Essa combinação de descoberta de variantes de processo e avaliação por especialista acelera a evolução do processo padrão/referência da organização, preservando a qualidade do modelo.

Em outro estudo [Greco, Guzzo e Pontieri 2008], combinou tarefas de agrupamento com técnicas de abstração *ad hoc*, com as quais surgiu uma nova abordagem capaz de construir uma taxonomia de modelos de processos; abordagem implementada ProM—plug-in AWS (*Abstraction Workflow Schema*).

Em [Song, Günther e van der Aalst 2009] foi proposta uma abordagem em que os traços são caracterizados por perfis. Estes últimos são formados por um conjunto de itens relacionados que descrevem o traço de uma perspectiva específica, onde o perfil pode ser representado como um vetor *n-dimensional*. Esses vetores são compostos por *atividade*, *originador*, *atributos de caso*, *transição*, *atributos de evento* e *atributos de desempenho*. Essa abordagem foi usada em registros de processos do hospital AMC¹¹. Todavia, o uso destes vetores limitou-se apenas ao perfil *atividade* para computar diferentes distâncias, e.g., *Euclidean*, *Hamming* e *Jaccard* para calcular distâncias entre *casos*. A estratégia básica é seguinte: cria-se uma classe de casos homogênea com ajuda de um algoritmo de Agrupamento, e.g., *K-means*, *Quality Threshold*, *Agglomerative Hierarchical Clustering* e *Self-Organizing Map*. A melhor combinação entre distâncias foi obtida com abordagem *Euclidean* e *Self-Organizing Map*, que geraram modelos mais simples, quando comparados aos modelos de processos gerados diretamente—i.e., sem a etapa de preparação/agrupamento—pelo algoritmo de mineração heurística no *log* de eventos. Em [Song et al. 2013] também adotou uma abordagem de redução de dimensionalidade para lidar com processos complexos e melhorar o desempenho dos algoritmos. A abordagem foi a seguinte: dado um conjunto de perfis calcula-se um vetor para cada instância de processo e, em seguida, aplica-se a redução de dimensionalidade, e.g., SVD—decomposição de valor singular, PCA—análise de componentes principais. Na próxima etapa, aplica-se uma técnica de agrupamento, e.g., *K-means*, *Agglomerative Hierarchical Clustering* e *Self-Organizing Map*.

Em [Bose e van der Aalst 2015] foi proposta uma nova abordagem de agrupamento de traços, chamada *Guide Tree Miner*¹². Ela permite definir passo a passo um contexto para a

¹¹ É um grande hospital acadêmico em Amsterdã na Holanda.

¹² Implementada na forma de *Plugin* no ProM

aplicação de uma tarefa aprendizagem não supervisionada. A configuração de contexto permite, de um lado, selecionar recursos, viés de agrupamento, medidas de distância/similaridade, entre outras opções, e de outro lado, gerar um agrupamento coerente de instâncias de processos apresentadas na forma de modelos mais simples. A abordagem foi aplicada em registros dados da *Philips Healthcare*.

Outra abordagem é a *Sequence Clustering*, usada principalmente para resolver desafios da bioinformática. A *Sequence Clustering* é formada por uma coleção de métodos, cuja finalidade é dividir o log de eventos em agrupamentos de sequências similares [Ferreira et al. 2007]. O algoritmo *Sequence Clustering* foi baseado em cadeias de *Markov*. Para avaliação da abordagem foram feitos dois experimentos. A conclusão foi que a técnica é eficiente para classificar diferentes comportamentos em um ambiente desconhecido. Entretanto, os dados de entrada para uso da técnica devem ser preparados e pré-processados para alcançar melhores resultados.

Em [Rebuge e Ferreira 2012] foi usada a abordagem *Sequence Clustering* como uma etapa metodológica para o desenvolvimento de uma ferramenta de mineração de processos, chamada *Medtrix Process Mining Studio*¹³. Devido à alta complexidade dos processos de saúde, a abordagem *Sequence Clustering* (cf. Figura 17) desempenhou um papel fundamental na identificação de comportamentos regulares e variantes de processos. Uma extensão desta abordagem foi proposta em [Veiga e Ferreira 2010]¹⁴.

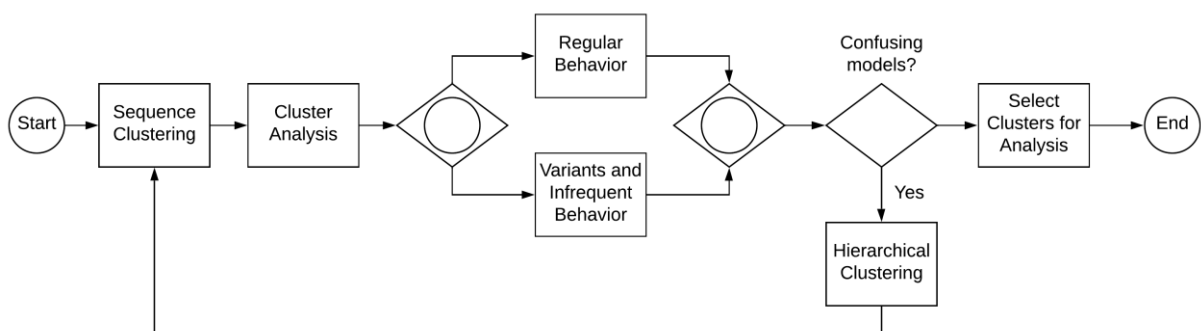


Figura 17. Etapa *Sequence Clustering*. Fonte: [Rebuge e Ferreira 2012].

¹³ Desenvolvida para o Hospital São Sebastião localizado em Portugal.

¹⁴ A primeira implementação foi feita usando SQL Server e Visual Studio. Ela foi migrada na sequência para o *ProM* e nomeada como *Sequence Clustering Mining Plugin*

Com foco na melhoria de processos para a indústria naval, em [Lee et al. 2013] foi proposta a abordagem *Trace Clustering*. Eles combinaram o agrupamento hierárquico com 4 métricas de distância (*Jaccard*, *cosine*, *correlation* e *Hamming*) e 3 métodos de ligação (cf. Figura 18). Os melhores resultados foram alcançados usando o método de ligação média baseado na distância *Jaccard*. Então, para a descoberta do modelo de processos, a abordagem foi combinada com regras de associação, permitindo extrair o fluxo de tarefas frequentes. Em [Park et al. 2014], também na indústria naval, foi proposta uma abordagem que fez o uso do *Trace Clustering*. Aqui, também foi usado a distância de *Jaccard* para avaliar as semelhanças entre traços; para o agrupamento de traços foi usado método *K-means*. A abordagem suporta a análise das diferenças entre trabalho planejado e executado. Além disso, ela fornece diretrizes para a melhoria de desempenho insuficiente em relação aos processos de fabricação.

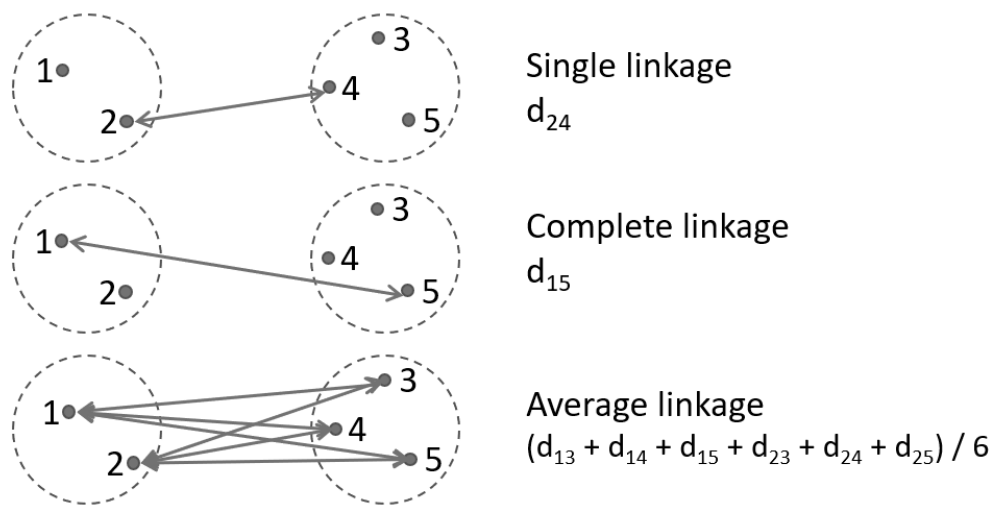


Figura 18. Métodos de ligação. Fonte: [Johnson e Wichern 1998].

Para resolver o problema de modelos de processos complexos gerados a partir do percurso clínico de pacientes com síndrome coronariana aguda¹⁵, [Funkner, Yakovlev e Kovalchuk 2017] propuseram três abordagens para descobrir e agrupar percursos clínicos. A primeira abordagem utilizou o método *k-means* e minimização do desvio de pontos pareados para a descoberta de percursos clínicos. A segunda abordagem utilizou também o método *K-means* e modelos evolutivos para a descoberta de percursos clínicos. A terceira abordagem foi

¹⁵ Admitidos no centro nacional de pesquisa médica Almazov durante 2010-2015 na Rússia.

baseada na descoberta de modelos por um algoritmo genético e agrupamento, que usa os modelos mais adequados considerando a fronteira de Pareto. De acordo com experimentos, a primeira abordagem fornece os melhores resultados.

Em [Chen et al. 2018] foram feitos experimentos com grandes volumes de registros médicos eletrônicos de pacientes com AVC. O propósito era redesenhar e melhorar os percursos clínicos existentes. Desenvolveu-se então um método para extrair processos/atividades típicos de tratamento, e.g., uso de aspirina¹⁶. O método foi dividido em quatro etapas: medir a similaridade de dois registros por meio de um método de similaridade; executar um algoritmo de agrupamento de propagação de afinidade—em inglês, *affinity propagation*; extrair uma descrição do processo de cada agrupamento; avaliar os modelos de processos obtidos. Em [Chen et al. 2018] foram utilizados dados de registros médicos¹⁷ e por meio de uma estrutura que relacionou fenótipos e fluxos de trabalho, foram obtidos agrupamentos considerados aceitáveis pelos especialistas.

Em [Liu, Liu e Ding 2013] foram propostos modelos padrões de comportamento de usuários, com o objetivo de incorporar resultados no processo de agrupamento. Esses modelos padrões de comportamento de usuários são representados como árvore de sufixos probabilísticos. Em [Liu et al. 2018] foram comparados quatro algoritmos de sequência, i.e., *structured-based*, *user-based*, *co-clustering* e um algoritmo de agrupamento que usa apenas a edição de distância. Em dois estudos experimentais conduzidos, mostrou-se que a incorporação de informações relacionadas ao usuário é eficaz na melhoria da precisão de agrupamento.

Com o objetivo de melhorar a compreensão, [De Weerd et al. 2012] propuseram uma técnica de agrupamento de traços para segregar *logs* de eventos em grupos. A abordagem foi dividir os *logs* de eventos em diferentes padrões de grupo. A tarefa de agrupamento aplica otimização gulosa e considera um ajuste do modelo de processo para cada agrupamento. A seleção de traços, para cada agrupamento, foi feita usando a distância *Euclidean*. As tarefas de mineração de textos e geração de árvores de decisão foram combinadas para melhorar a classificação; essa abordagem foi usada quando dois comportamentos distintos são buscados: típico/característico—descoberta de processo padrão—ou atípico—caso com baixa frequência.

¹⁶ Droga antiplaquetária que aparece em quase todos os tratamentos para infarto cerebral.

¹⁷ Do Hospital Memorial *Northwestern*

Em [De Weerd et al. 2013] foi usada a abordagem de agrupamento, denominada *Active Trace*¹⁸, para apoiar a descoberta de processos na presença de comportamentos altamente complexos. As técnicas propostas focam em agrupamentos de traços, permitindo que sejam seguidas por qualquer algoritmo de descoberta de processos. Agrupar logs de eventos similares de acordo com os recursos resulta em um problema fundamental, chamado *bias*, e.g., maximizar as semelhanças dentro do agrupamento e minimizar as semelhanças fora do agrupamento. Deve-se observar que o agrupamento foi baseado na distribuição ótima em um dado K , definido *a priori*. Esta abordagem segue a abordagem *top-down*, usando um algoritmo guloso. Ela foi dividida em três fases: seleção, análise prospectiva e avaliação de caminhos residuais. Primeiro, uma estratégia de amostragem seletiva foi empregada principalmente com base em medições de frequência usando a distância *Euclidean* entre instâncias de processos distintos. Descoberta de processos é aplicada na sequência, então o *fitness* e o tamanho mínimo do agrupamento são verificados. Na segunda fase, continua-se até o máximo predefinido de k -agrupamentos. Finalmente, as instâncias restantes do processo devem ser divididas entre agrupamentos, considerando o *fitness*.

ActiTraC permite executar [De Weerd et al. 2013]:

- amostragem seletiva baseada em frequência, usando tamanho de janela w ;
- pior caso, definindo o alvo w como 1 e o *fitness* para $-\infty$;
- amostragem seletiva baseada na distância, agrupando traços mais similares baseados no alfabeto de repetição máxima de acordo com a distância *Euclidean*.

Em [Accorsi e Stocker 2012] foi proposta uma abordagem para agrupamento de logs com base em informações de tempo, resultando em uma estrutura dinâmica de processo. A abordagem foi dividida em três etapas: cálculo de distância por par de atividades; realização de cortes de agrupamentos, e a descoberta de processos.

Para melhorar a compreensão dos processos não estruturados, [Marchetto e Francescomarino 2011] propuseram o agrupamento de traços para obter os comportamentos relevantes com todos os casos semelhantes e descartando variantes não relevantes. A abordagem inicia com um agrupamento *k-grammar* no conjunto de traços para estabelecer entradas consecutivas e analisar subconjuntos de agrupamentos considerando cada par de traços

¹⁸ Também conhecimento *ActiTraC plugin no ProM*.

baseado na similaridade de cosseno, e depois como um agrupamento *agglomerative hierarchical*, considerando a distância *Euclidean* para medir a distância entre pares de vetores em diferentes agrupamentos e um limite $t = 10\%$.

[San Pedro e Cortadella 2016] também focaram em um novo método para facilitar a compreensão de conjuntos de modelos de processos. A abordagem divide os comportamentos estruturais para obter modelos de processos mais simples, em vez de ignorar o ruído nos *logs* de eventos. O agrupamento de *log* de eventos foi baseado em um sistema de transição rotulado. Como etapa final, uma síntese é obtida estendendo a síntese baseada na região¹⁹. A implementação é baseada no pacote de mineração de processos em *Python*. Os resultados foram comparados com o *ActiTraC* e superam essa abordagem na maioria dos aspectos.

Em [Juneja, Kundra e Sureka 2016] simplificou a análise do modelo de processo considerando o agrupamento de traços usando *k-medoid*, *i.e.*, onde os agrupamentos são representados usando uma mediana dos dados em vez da média, com base em duas métricas diferentes: *Longest Common Subsequence* e *Dynamic Time Warping*. Como próximo passo, a descoberta de modelos de processo foi baseada no *Fuzzy Miner* e implementada na ferramenta *Disco*. A avaliação foi feita em um processo de software usando uma base de dados de rastreamento de problemas e considerando duas medidas: *Fitness* e *McCabe's cyclomatic Complexity*. Portanto, seu método foi eficaz para este estudo de caso.

Em [Tax et al. 2016] foram exploradas três heurísticas para descobrir padrões de comportamento frequentes, chamados de *Local Process Model* (LPM). Foi analisado um subconjunto de atividades com o objetivo de acelerar a descoberta do LPM, mas mantendo a qualidade. Três abordagens diferentes foram comparadas em experimentação sistemática. A primeira abordagem aplicou agrupamento *Markov* para criar conjuntos de projeção, resultando em uma melhoria do tempo de execução, com LPMs descobertos de boa qualidade. A segunda abordagem segue uma heurística baseada na entropia de *log*; exigiu mais tempo de execução para uma velocidade moderada, mas permitiu a descoberta de LPMs de maior qualidade. A terceira proposta é relativa ao ganho de informação, mostrando desempenho instável. Nos experimentos, para alguns conjuntos de dados, a velocidade e a qualidade do LPM são maiores

¹⁹ Este processo foi feito usando o plugin *Prom* para *Petrify*

do que com o método baseado em entropia de *log*, enquanto que para outros conjuntos de dados não há nenhuma aceleração.

Focada na análise de padrões de tratamento para os serviços de saúde do Câncer Gástrico na Colômbia, [Villamil et al. 2017] combinou o agrupamento sequencial com uma análise sequencial baseada em cadeia de *Markov*. O agrupamento foi baseado em uma medida que considerou a distância dos agrupamentos internos. A análise foi relevante para detectar ineficiências na descoberta de processo, considerando agrupamentos e anos de intervalo. A técnica permitiu que especialistas melhorassem o conhecimento do processo e encontrassem problemas de qualidade no tratamento clínico do câncer gástrico.

Um caso aplicado na indústria manufatureira apresenta um processo não estruturado e problemas em indicadores de desempenho de produção [Meincheim et al. 2017]. Uma combinação de técnicas foi usada para entender como o processo ocorre na prática, quantos agrupamentos de traços devem ser identificados como variantes de processos homogêneos e o que causa a ineficiência de produção. Essas técnicas combinaram um agrupamento incremental considerando a distância interna do agrupamento. Para medir a distância, foi aplicada uma similaridade baseada no número de atividades que ocorrem entre dois modelos de processos. Depois, a descoberta do processo foi realizada usando o *Inductive Miner*. Como resultado, identificou-se gargalos causados por decisões errôneas em tempo de execução; os quais serviram para apoiar a melhoria do processo, validando as descobertas com especialistas.

Em um processo educacional, [Ariouat et al. 2016] combinou duas abordagens baseadas em etapas de agrupamento e descoberta de processo. O primeiro passo foi agrupar usando *k-means*, tentando *k* entre 2 e 5, com a medida da distância de *Hamming* entre dois traços, usando o operador lógico *XOR*. Como segundo passo, foi aplicado o *Heuristic Miner*. A abordagem permitiu obter uma boa compreensão dos principais comportamentos dos alunos.

3.2 Resumo das abordagens

A Tabela 6 apresenta um resumo das abordagens para a descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento.

Tabela 6 - Abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento. Fonte: o Autor, 2019

Author Approach	Clustering technique	Distance Heuristic	Process Discovery	Tool
Greco 2006 DWS [Greco et al. 2006] [Santos, Oliveira e Abreu 2015]	Hierarchical clustering results in disjunctive workflow schema then k-means / vector space	Number of combines a maxLevels in the hierarchy of clusters	Maximum Process Discovery	ProM 5.2
Greco 2008 AWS [Greco, Guzzo e Pontieri 2008]	Hierarchical clustering with abstraction techniques	Number of combines a maxLevels in the hierarchy of clusters	Maximum Process Discovery	ProM 4.2
Song 2009 Trace Clustering [Song, Gubther e van der Aalst 2009]	Self-Organizing Map	Euclidean	Heuristic Miner	ProM 5.2
Bose 2010 Guide Tree Miner [Bose e van der Aalst 2010]	Unsupervised k-activities	1) Maximal Pair 2) Maximal..Repeat	Heuristic Miner	ProM 6.8
Ferreira 2007 Sequence Clustering [Ferreira et al. 2007]	Each cluster is represented by corresponding Markov chain	Markov probability	Markov-based	SQL Server Visual Studio
Veiga 2010 Sequence Clustering [Veiga e Ferreira 2010]	Each cluster is represented by corresponding Markov chain	Markov probability	Markov-based	ProM 5.2
Rebuge 2012 Sequence Clustering [Rebuge e Ferreira 2012]	First-order Markov chain extracted from the behavior of those sequences	Distance between two probability distributions that computing the distance between all pairs of clusters of transition matrix	Heuristic Miner	Medtrix
Lee 2013 [Lee et al. 2013]	Hierarchical clustering	Jaccard distance based average linkage method	Association rule to disc. freq.flow	Not available
Park 2014 [Park, Lee e Zhu 2014]	k-means	Similarities for the task and transition vector	Operation sequence	Not available
Funkner 2017 Cluster clinical pathways [Funkner, Yakovlev e Kovalchuk 2017]	k-means	1) Minimize the pairwise squared deviations 2) Template defined by evolutionary technique most with sequences of alignment states 3) GA Templates choosing fittest among Pareto	Not combined	Not available
Chen 2018 [Chen et al. 2018]	Affinity Propagation (AP) clustering to doctor	1) Markov chain to measure the similarity	DOSSC-TTPE	MATLAB

	order set sequence (DOSS)	between pairwise DOSSs; 2) Manhattan distance.		
Liu 2013 [Liu, Liu e Ding 2013] [Liu et al. 2018]	User behavior patterns and sequence co-clustering algorithm	Sequence distance matrix	Not combined	Not available
Chen 2018 [Chen et al. 2018]	Community detection algorithm to infer health condition collections	Cosine similarities	Modified LDA algorithm	ProM
Weerd 2013 ActiTraC [Weerd et al. 2012] [Weerd et al. 2013]	Forward clustering with maximum k finding an optimal distribution	1) Frequency-based - Worst case 2) from 1 to infinity window 3) Distance-based selective using MRA	Heuristic Miner	ProM 6.8
Accorsi 2012 [Accorsi e Stocker 2012]	Cluster cuts using a dynamic k	Activity pair distance matrix	Not combined	Not available
Marchetto 2011 Trace Clustering [Marchetto e Francescomarino 2011]	1)k-grammar clustering technique 2)agglomerative hierarchical clustering	1) Similarity based on cosine similarity 2) Euclidean distance	Not combined	Not available
San Pedro 2016 [San Pedro e Cortadella 2016]	Clustering was based in a labelled transition system (LTS)	Well-Behaved (WB) slices	Petrify	Python
Juneja 2016 [Juneja, Kundra e Sureka 2016]	Adaptation of the k-medoid clustering algorithm	Longest Common Subsequence (LCS) and Dynamic Time Warping (DTW)	Fuzzy Miner	Disco
Tax 2016 Local Process Model (LPM) [Tax et al. 2016]	1) Markov clustering; 2) Entropy; 3) Maximal Relative Information Gain	1) Connectedness of 2 activities by directly follows and directly-precedes ratios; 2) Entropy on a log projection; 3) Maximal Relative Information Gain	Heuristic Miner	Not available
Villamil 2017 [Villamil et al. 2017]	Sequential clustering	Similar internally	Markov-based	ProM 5.2
Meincheim 2017 [Meincheim et al. 2017]	Incremental clustering	Process similarity measure based on the nr activities between 2 cases	Inductive Miner	ProM 6.0
Ariouat 2016 [Ariouat et al. 2016]	k-means trying k between 2 and 5	Hamming distance measure between two traces using the logical XOR operator	Heuristic Miner	ProM

3.3 Considerações sobre o Capítulo

Este capítulo apresentou uma revisão da literatura sobre abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento. Diante disso, com a

Tabela 6 foi possível responder a primeira questão de pesquisa (RQ01), na qual ela identificou diferentes abordagens, contendo na tabela o autor/abordagem, técnica de agrupamento, distância heurística, processo de descoberta e a ferramenta utilizada. Por fim, o próximo capítulo irá selecionar as abordagens que disponibilizam uma implementação para o experimento.

CAPÍTULO 4 - ESTRUTURAÇÃO DA PESQUISA

Este capítulo tem como finalidade descrever o método de pesquisa utilizado neste trabalho. Por isso, inclui-se, neste capítulo, a forma de trabalho e a estratégia, do mesmo modo as etapas utilizadas para organizar o trabalho e alcançar os objetivos previamente definidos.

4.1 Método de Pesquisa

A pesquisa é experimental. As etapas da pesquisa partem da formulação do problema e das hipóteses, seleção das amostras, da revisão do planejamento e a execução do experimento, onde os resultados são submetidos a um teste de hipóteses [Triviños 1987].

4.2 Questões de Pesquisa

O objetivo geral desta pesquisa, lembrado, é comparar diferentes técnicas de agrupamento utilizadas em abordagens para a descoberta de múltiplos modelos de processos, e identificar a técnica mais adequada para segmentar um arquivo de log de eventos contendo três processos diferentes, e também qual segmentação gera à maior redução da complexidade dos modelos minerados. Neste contexto, busca-se responder as seguintes questões de pesquisa:

RQ01: Quais são as técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos?

RQ02: As técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos são capazes de separar um log de eventos contendo processos diferentes?

RQ03: As diferentes técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos impactam no resultado da mineração de processos?

4.3 Estratégia de Pesquisa

Este estudo foi dividido em duas fases: (1) Estudo exploratório para identificar e selecionar abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento; e (2) definição do problema da pesquisa, desenvolvimento do experimento, análise de dados coletas com o intuito de responder as questões de pesquisa. A Figura 19 representa de forma breve a estratégia de pesquisa.

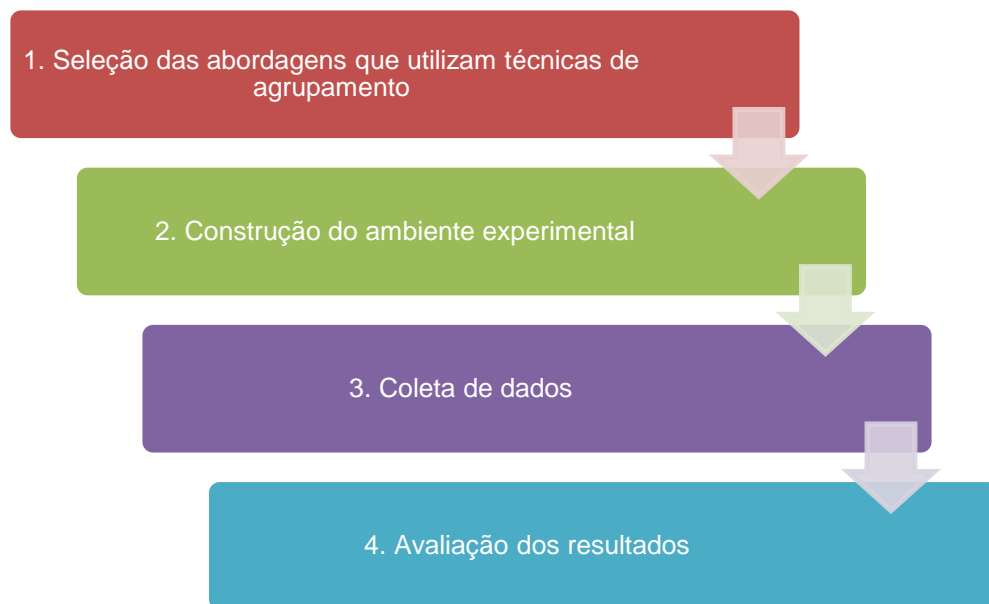


Figura 19 - Etapas do planejamento do experimento. Fonte: o Autor, 2018

4.3.1 Seleção das abordagens

Esta etapa tem como objetivo identificar e selecionar diferentes abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento (cf. Tabela 6).

Apenas as abordagens que disponibilizam uma implementação foram selecionadas. Assim, das 22 abordagens elencadas foram selecionadas para o experimento 5 delas (cf. Tabela 7).

Tabela 7 - Abordagens selecionadas para o experimento. Fonte: o Autor, 2019

Author Approach	Clustering technique	Distance Heuristic	Process Discovery	Tool
Greco 2006 DWS [Greco et al. 2006]	Hierarchical clustering results in disjunctive workflow schema then k-means / vector space	Number of combines a maxLevels in the hierarchy of clusters	Maximum Process Discovery	ProM 5.2
Song 2009 Trace Clustering [Song, Gubther e van der Aalst 2009]	Self-Organizing Map (SOM)	Euclidean	Heuristic Miner	ProM 5.2
Bose 2010 Guide Tree Miner [Bose e van der Aalst 2010]	Unsupervised k-activities	1) Maximal Pair 2) Maximal..Repeat	Heuristic Miner	ProM 6.8
Veiga 2010 Sequence Clustering [Veiga e Ferreira 2010]	Each cluster is represented by corresponding Markov chain	Markov probability	Markov-based	ProM 5.2
Weerdts 2013 ActiTraC [Weerdts et al. 2013]	Forward clustering with maximum k finding an optimal distribution	1) Frequency-based - Worst case 2) from 1 to infinity window 3) Distance-based selective using MRA	Heuristic Miner	ProM 6.8

4.3.2 Construção do ambiente experimental

Esta etapa teve como propósito construir o ambiente experimental para coleta de dados. O experimento foi dividido em dois cenários, o primeiro visou a separação de um arquivo de *log de eventos* em relação a três diferentes processos. E o segundo cenário se concentrou na avaliação da redução da complexidade dos modelos minerados.

As abordagens selecionadas para o experimento fizeram as versões 5.2 e 6.8 da ferramenta ProM, além de diferentes técnicas de agrupamento, distâncias e heurísticas (cf. Tabela 7). Entretanto, para o experimento foi utilizado o *Heuristic Miner* para generalização dos resultados, utilizando-se dos parâmetros padrões do minerador fornecido pelo ProM.

Foram utilizados no experimento os parâmetros padrões das abordagens:

- *DWS* [Greco et al. 2006] o plugin está disponível no ProM 5.2 (cf. Figura 20).
- *Trace Clustering* [Song, Gubther e van der Aalst 2009] o plugin está disponível no ProM 5.2 (cf. Figura 21).
- *Guide Tree Miner* [Bose e van der Aalst 2010] o plugin está disponível no ProM 6.8 (cf. Figura 22).
- *Sequence Clustering* [Veiga e Ferreira 2010] o plugin está disponível no ProM 5.2 (cf. Figura 23).
- *ActiTraC* [Weerd et al. 2013] o plugin está disponível no ProM 6.8 (cf. Figura 24).

DWS mining plugin

Relative-to-best threshold	0.05
Positive observations	10
Dependency threshold	0.9
Length-one-loops threshold	0.9
Length-two-loops threshold	0.9
Long distance threshold	0.9
Dependency divisor	1
AND threshold	0.1

Extra info

Use all-activities-connected-heuristic

Use long distance dependency heuristics

Frequency support - sigma	0.05
Feature relevance threshold - gamma	0.01
(Max.) Number of clusters per split	4
(Max.) Length of features	5
(Max.) Number of splits	2
(Max.) Number of features	2

Help start mining

Figura 20 - Parâmetros DWS. Fonte: o Autor, 2019.

Trace clustering

Profiles configuration

Activity invert active
Compares the occurrences of activities (18 items)
weight:

Activity Char Streams invert active
Compares process instances as streams (0 items)
weight:

Activity Patterns Profile invert active
Compares patterns in process instances (2395 items)
weight:

Activity Pattern Alphabets Profile invert active
Compares patterns with flexible ordering in process instances (666 items)
weight:

Distance metric
Euclidean
Euclidean distance metric

Clustering algorithm
SOM Clustering
SOM Clustering allows the user to specify width and height.

Preprocessing
 TF.IDF
 Preprocessing
SVD dim: 3

start clustering

Figura 21 - Parâmetros Trace Clustering. Fonte: o Autor, 2019

Guide Tree Miner Plugin

Feature Configuration Step

Select Feature Type

- Entire Trace
- Sequence Feature
- Alphabet Feature
- Consider Base Features

Select Sequence Features

- Individual Event
- K-Gram
- Tandem Repeat
- Maximal Repeat
- Super Maximal Repeat
- Near Super Maximal Repeat

Select Feature Count Type

- Nominal
- Numeric

Filter Features

- Frequency
- Instance %
- Min. Alphabet Size
- Max. Alphabet Size

Figura 22 - Parâmetros Guide Tree Miner. Fonte: o Autor, 2019.

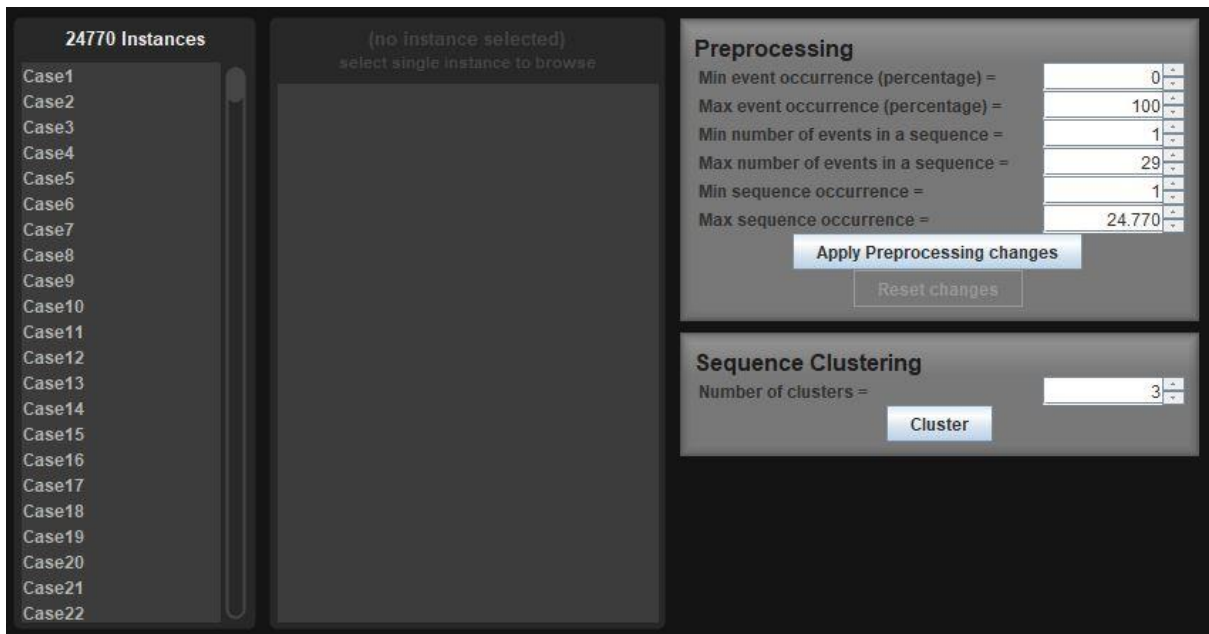


Figura 23 - Parâmetros Sequence Clustering. Fonte: o Autor, 2019.

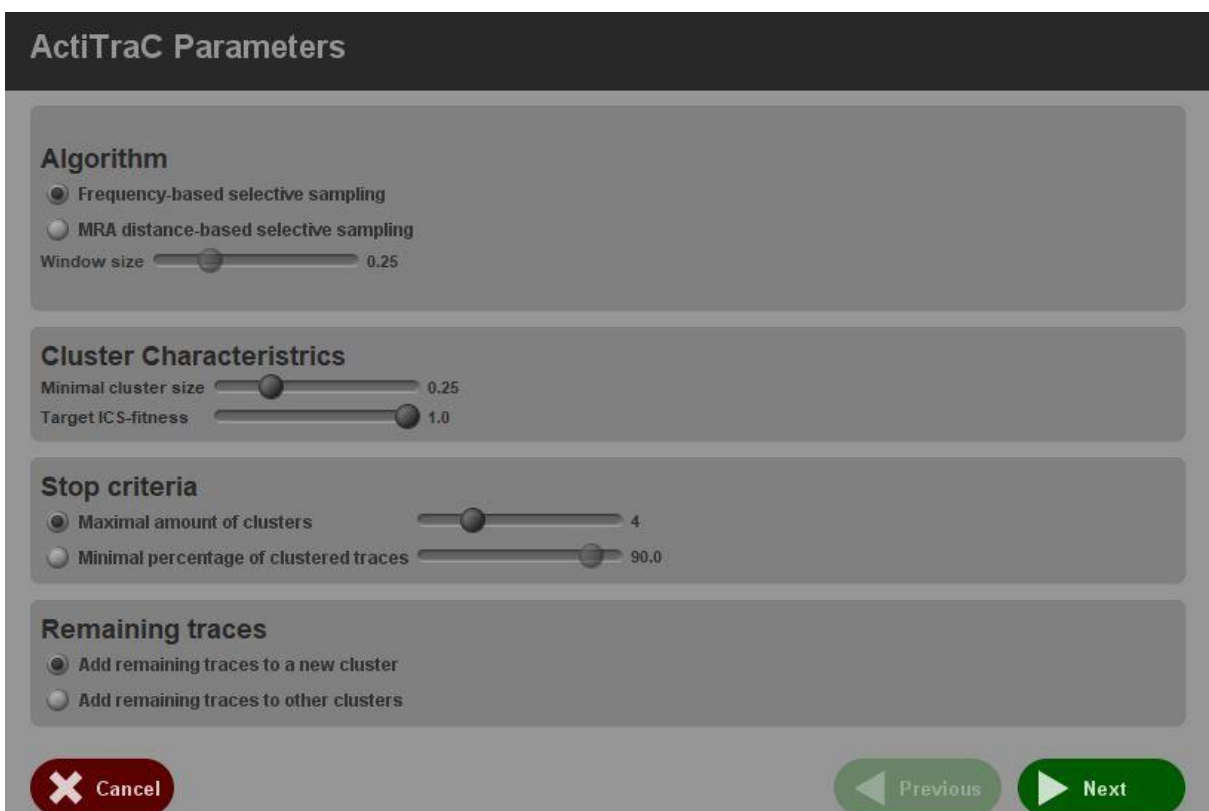


Figura 24 - Parâmetros ActiTraC. Fonte: o Autor, 2019.

4.3.3 Coleta de dados

Esta etapa consistiu na coleta de dados, em que as técnicas de agrupamento utilizadas nas abordagens selecionadas foram aplicadas em 4 *logs* de eventos (cf. Tabela 8). Estes logs de eventos foram utilizados em [Weerd et al. 2013] e são originários de ambientes flexíveis que exibem uma grande variedade de comportamentos de processos.

Tabela 8 - Log de eventos selecionados para o experimento. Fonte: o Autor, 2019.

Segmento	Processo	Casos	Eventos	Fonte
TIC	Processo de helpdesk de um serviço de TIC em uma universidade	24770	124117	[Weerd et al. 2013]
Manufatura	Processo de <i>Customer Relationship Management (CRM)</i> em uma empresa de manufatura	956	11218	[Weerd et al. 2013]
Telecomunicações	Processo de CRM em uma empresa de telecomunicações	17812	83286	[Weerd et al. 2013]
Seguros	Processo de recebimento de documentos em uma companhia de seguros	12391	65653	[Weerd et al. 2013]

O processo de coleta de dados foi dividido em 2 cenários:

- I. O primeiro cenário visou a separação de um arquivo de *log* contendo três processos diferentes. Um processo de *helpdesk* de um serviço de TIC em uma universidade, um processo de CRM em uma empresa de manufatura e um processo de CRM em uma empresa de telecomunicações (cf. Tabela 8), consistindo em 700 casos de cada *log*, os quais foram agregados a um novo arquivo de log contendo 2.100 casos gerais (cf. Figura 25).

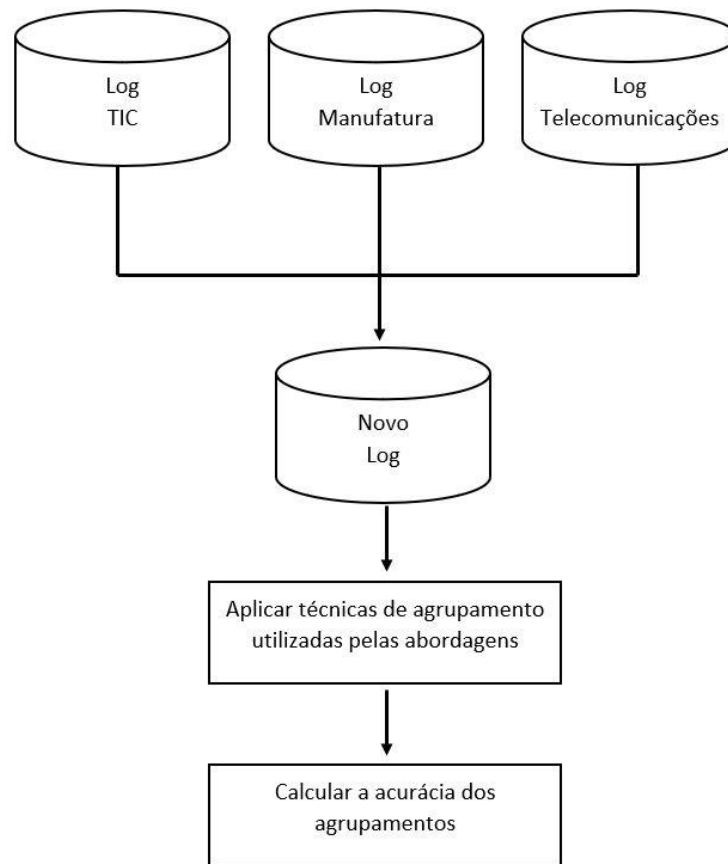


Figura 25 - Processo da coleta de dados do primeiro cenário. Fonte: o Autor, 2019.

- II. Baseado nos modelos de processo descobertos foi calculado a redução da complexidade dos modelos minerados (cf. Figura 26).

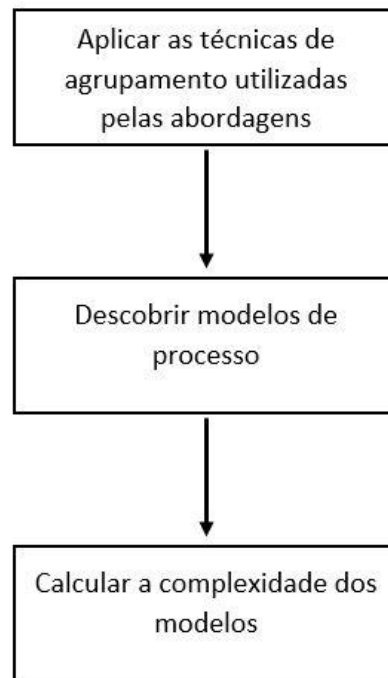


Figura 26 - Processo da coleta de dados do segundo cenário. Fonte: o Autor, 2019.

4.3.4 Avaliação dos resultados

As análises dos dados coletados foram divididas em duas fases. A primeira para verificar se as técnicas de agrupamento utilizado pelas abordagens são capazes de agrupar corretamente o *log* contendo três processos diferentes, e na segunda para calcular à redução da complexidade dos modelos minerados.

4.4 Considerações sobre o Capítulo

Este capítulo apresentou a estruturação da pesquisa e suas etapas para avaliar as técnicas de agrupamento utilizadas nas abordagens para a descoberta de múltiplos modelos de processos.

O método de pesquisa planejado foi experimental, utilizando um estudo em duas etapas: (1) Identificação e seleção das abordagens para descoberta de múltiplos modelos de processos que utilizam técnicas de agrupamento; (2) Realização do experimento em dois cenários, o primeiro cenário visa a utilização das técnicas de agrupamento em um *log* contendo três processos diferentes, e no segundo cenário à redução da complexidade dos modelos minerados.

CAPÍTULO 5 - ANÁLISE E DISCUSSÃO DOS RESULTADOS

Este capítulo apresenta a análise consolidada dos resultados alcançados com base nos experimentos descritos e colocados em prática no capítulo anterior. A partir dos dois cenários descritos anteriormente analisou-se os resultados obtidos das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos, em um log de eventos contendo três processos diferentes. E também analisou-se à redução da complexidade que é frequentemente entendida como o tamanho de um modelo em termos de número de vértices, arestas e a relação entre eles, as métricas de complexidade são representadas por [Melcher 2018]:

- *Coefficient of Connectivity* (CNC): Representa uma relação entre vértices e arestas, e é calculada pela Fórmula F1, onde $|A|$ representa a quantidade de arestas e $|N|$ a quantidade de vértices.

$$CNC(G) := \frac{|A|}{|N|} \quad F1$$

- *Cyclomatic Number* (CN): Fornece uma diferença absoluta e é calculada pela Fórmula F2.

$$CN := |A| - |N| + 1 \quad F2$$

Nas subseções a seguir são apresentados os resultados dos dois cenários descritos no Capítulo 4.

5.1 Cenário 1: Separação de um arquivo de log contendo três processos diferentes

O primeiro cenário visou a utilização de técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos (cf. Tabela 7), em um *log* de eventos contendo três processos diferentes (cf. Figura 25). O número de agrupamentos resultantes para cada abordagem foi definido como três, e todos os outros parâmetros foram definidos como padrão. A Tabela 9 apresenta os resultados obtidos utilizando as técnicas de agrupamento para descoberta de múltiplos modelos de processos. Os valores de cada linha representam as quantidades de casos identificados em cada agrupamento, por exemplo, no primeiro agrupamento da abordagem *Sequence Clustering* foi identificado 486 casos referente ao *log* de eventos TIC, 45 casos referente ao *log* de eventos Manufatura e 310 casos referente ao *log* de eventos Telecomunicações.

Tabela 9 - Resultados das aplicações das técnicas de agrupamento. Fonte: o Autor, 2019.

A		Casos		
		TIC	Manufatura	Telecomunicações
1	<i>Sequence Clustering</i>	486	45	310
2		3	603	182
3		211	52	208
1	<i>ActiTraC</i>	0	405	0
2		536	0	0
3		164	295	700
1	<i>Guilde Tree Miner</i>	0	171	0
2		0	0	700
3		700	529	0
1	<i>DWS</i>	635	593	700
2		0	16	0
3		65	91	0
1	<i>Trace Clustering</i>	285	176	250
2		0	0	7
3		415	524	443

A partir destes resultados, observa-se que as cinco técnicas de agrupamento utilizadas pelas abordagens produzem agrupamentos substancialmente diferentes (cf. Tabela 9). As técnicas *Sequence Clustering* [Veiga e Ferreira 2010], *DWS* [Greco et al. 2006] e *Trace Clustering* [Song, Gubther e van der Aalst 2009] não são capazes de identificar os diferentes processos e separar adequadamente o arquivo de log. Entretanto, os resultados das outras duas

técnicas foram mais relevantes. O primeiro e o segundo agrupamento da abordagem *ActiTraC* [Weerdt et al. 2013] contém respectivamente apenas casos dos processos de manufatura e TIC. No entanto, o terceiro agrupamento contém casos de todos os processos, mas possui todos os casos do processo de telecomunicações. Além disso, o primeiro e o segundo agrupamento resultantes da abordagem *Guide Tree Miner* [Bose e van der Aalst 2010] contém respectivamente apenas casos dos processos de manufatura e telecomunicações, e no terceiro agrupamento possui todos os casos do processo de TIC. Entretanto, diferentemente da abordagem *ActiTraC* [Weerdt et al. 2013], o terceiro agrupamento possui apenas *outliers dos* casos do processo manufatura.

No entanto, nenhuma das abordagens utilizadas foi capaz de gerar três agrupamentos contendo os 700 casos dos processos correspondentes.

5.2 Cenário 2: Redução da Complexidade dos Modelos

O segundo cenário visou mensurar a redução da complexidade dos modelos a partir da aplicação das técnicas de agrupamento utilizadas nas abordagens de descoberta de múltiplos modelos de processos (cf. Tabela 7), diferente do cenário 1 que alterou-se o número de agrupamentos, o cenário 2 utilizou-se os parâmetros *defaults* definidos pela abordagem. Foi utilizado o *Heuristic Miner*²⁰ para generalização dos resultados com os parâmetros *defaults*. Analisou-se à complexidade dos modelos obtidos observando as métricas de complexidade CNC e CN.

Nas subseções a seguir são apresentados os resultados da execução em diferentes bases de dados; considerando os parâmetros *defaults* definidos nas abordagens.

5.2.1 Log de eventos TIC

Esta seção apresentou os resultados do experimento aplicado no log de eventos reais que representam o processo de *helpdesk* de um serviço de TIC em uma universidade. Tal log de eventos é composto por 24.770 casos, 124.217 eventos e 18 atividades. A Tabela 10 apresenta os resultados obtidos das métricas selecionadas para mensurar complexidade dos modelos resultantes.

²⁰ Implementado no PROM.

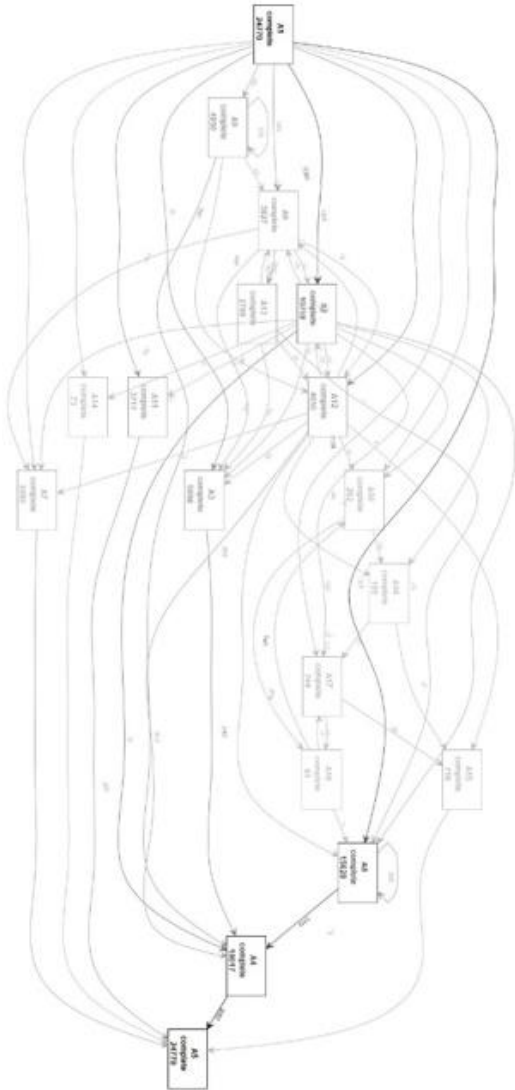
Tabela 10 - Complexidade de modelos vs. técnicas de agrupamento na descoberta de múltiplos modelos de processos: log de eventos TIC. Fonte: o Autor, 2019.

		<i>Sequence Clustering</i>	<i>ActiTraC</i>	<i>Guilde Tree Miner</i>	<i>DWS</i>	<i>Trace Clustering</i>	<i>Somente Heuristic Miner</i>
A	<i>k-cluster</i>	3.00	4.00	4.00	1.00	N.A	1.00
	<i>Mínimo</i>	46.00	19.00	30.00	65.00	N.A	61.00
	<i>Média</i>	49.00	32.75	41.50	65.00	N.A	61.00
	<i>Máximo</i>	53.00	50.00	62.00	65.00	N.A	61.00
	<i>Desvio padrão</i>	2.94	13.55	12.68	N.A	N.A	N.A
N	<i>Mínimo</i>	17.00	10.00	15.00	18.00	N.A	18.00
	<i>Média</i>	17.67	14.50	16.50	18.00	N.A	18.00
	<i>Máximo</i>	18.00	18.00	18.00	18.00	N.A	18.00
	<i>Desvio padrão</i>	0.47	3.57	1.50	N.A	N.A	N.A
CNC	<i>Mínimo</i>	2.56	1.67	2.00	3.61	N.A	3.39
	<i>Média</i>	2.78	2.17	2.47	3.61	N.A	3.39
	<i>Máximo</i>	3.12	2.78	3.44	3.61	N.A	3.39
	<i>Desvio padrão</i>	0.24	0.42	0.57	N.A	N.A	N.A
CN	<i>Mínimo</i>	29.00	9.00	16.00	48.00	N.A	44.00
	<i>Média</i>	32.33	19.25	26.00	48.00	N.A	44.00
	<i>Máximo</i>	37.00	33.00	45.00	48.00	N.A	44.00
	<i>Desvio padrão</i>	3.40	10.16	11.46	N.A	N.A	N.A

|A| = Número de arcos; |N| = Número de vértices; CNC = *Coefficient of Connectivity*;
CN = *Cyclomatic Number*; N.A = Não Avaliado

Os agrupamentos gerados a partir das abordagens *Sequence Clustering*, *ActiTraC* e *Guilde Tree Miner* geram modelos com menor complexidade do que o arquivo de *log* sem a utilização das técnicas de agrupamento em termos das métricas de complexidade mencionadas (cf. Tabela 10). No entanto, existem diferenças substanciais nos agrupamentos produzidos por essas técnicas. A abordagem *Sequence Clustering* produz uma quantidade significativamente maior de vértices e arcos do que as outras abordagens. Todavia, essa abordagem gerou apenas três agrupamentos. As abordagens *ActiTraC* e *Guilde Tree Miner* geraram quatro agrupamentos; a *ActiTraC* foi a abordagem que mais reduziu a complexidade dos modelos descobertos. A Figura 27 apresenta: a) aplicação do *Heuristic Miner* no log de eventos TIC; b) aplicação do *Heuristic Miner* nos agrupamentos gerados pelo *ActiTraC*.

a) Modelo gerado sem a utilização das técnicas de agrupamento



b) Modelos gerados com os agrupamentos da abordagem ActiTraC

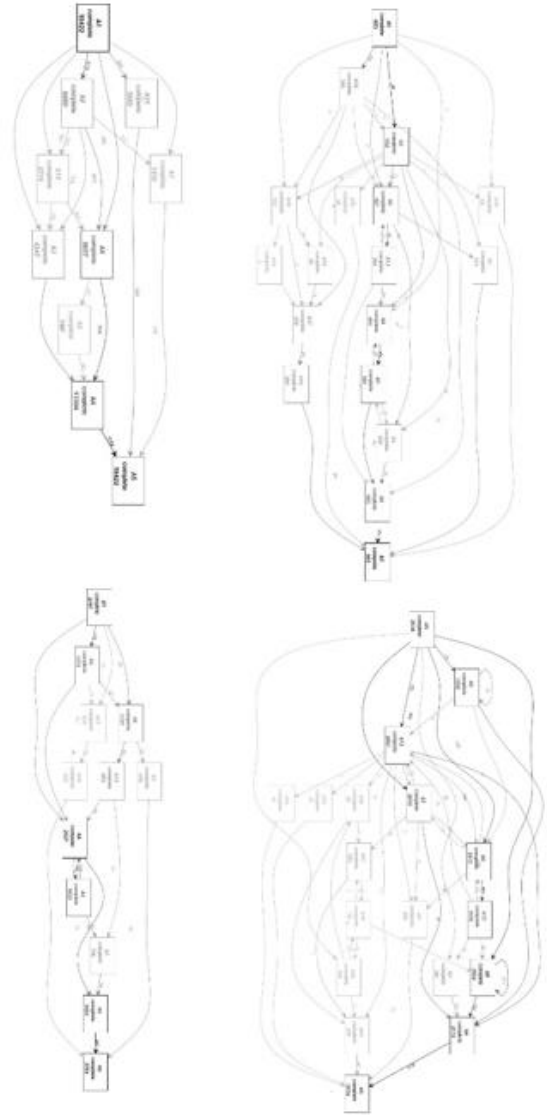


Figura 27. Modelo gerado sem aplicação de técnicas de agrupamento e com ActiTraC: log de eventos TIC.
Fonte: o Autor, 2019.

A abordagem *DWS* com seus parâmetros *defaults* gerou apenas um agrupamento com complexidade maior do que no modelo gerado utilizando o *Heuristic Miner*. Por fim, a abordagem *Trace Clustering* gerou alto custo computacional (cf. Figura 28), i.e., não foi possível utilizar no log de eventos TIC.

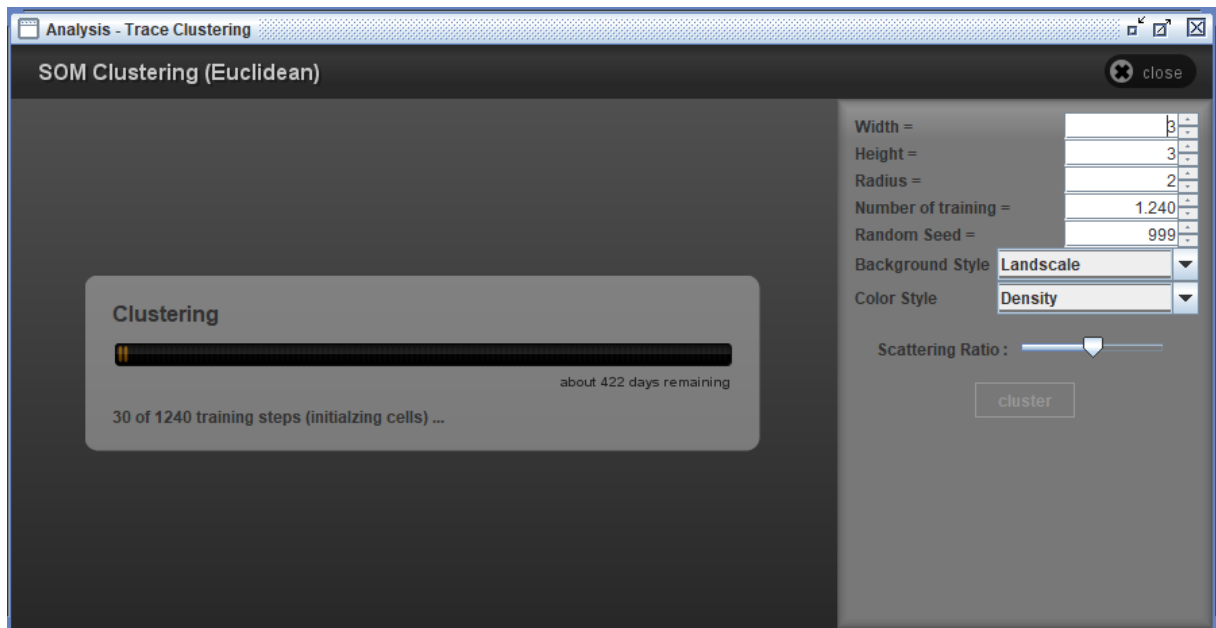


Figura 28. Alto custo computacional gerado pela abordagem Trace Clustering. Fonte: o Autor, 2019.

5.2.2 Log de eventos Manufatura

Esta seção apresenta os resultados obtidos nos experimentos aplicados sobre log de eventos reais de um processo de CRM de uma empresa da área de Manufatura. O log de eventos encerra 956 casos, 11.218 eventos e 22 atividades. A Tabela 11 apresenta os resultados obtidos, em que os valores são medidas resultantes das métricas de complexidade dos modelos.

Tabela 11 - Resultados das métricas de complexidade dos modelos gerados a partir das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos no log de eventos Manufatura. Fonte: o Autor, 2019.

		<i>Sequence Clustering</i>	<i>ActiTraC</i>	<i>Guilde Tree Miner</i>	<i>DWS</i>	<i>Trace Clustering</i>	<i>Somente Heuristic Miner</i>
A	<i>k-cluster</i>	3.00	4.00	4.00	1.00	1.00	1.00
	<i>Mínimo</i>	26.00	16.00	31.00	42.00	42.00	42.00
	<i>Média</i>	36.00	29.00	33.75	42.00	42.00	42.00
	<i>Máximo</i>	45.00	41.00	37.00	42.00	42.00	42.00
	<i>Desvio padrão</i>	7.79	9.03	2.16	N.A	N.A	N.A
N	<i>Mínimo</i>	19.00	13.00	19.00	22.00	22.00	22.00
	<i>Média</i>	20.00	18.50	20.25	22.00	22.00	22.00
	<i>Máximo</i>	22.00	22.00	21.00	22.00	22.00	22.00
	<i>Desvio padrão</i>	1.41	3.35	0.83	N.A	N.A	N.A
CNC	<i>Mínimo</i>	1.37	1.23	1.57	1.91	1.91	1.91
	<i>Média</i>	1.79	1.53	1.68	1.91	1.91	1.91
	<i>Máximo</i>	2.04	1.86	1.85	1.91	1.91	1.91
	<i>Desvio padrão</i>	0.29	0.23	0.11	N.A	N.A	N.A
CN	<i>Mínimo</i>	8.00	4.00	13.00	21.00	21.00	21.00
	<i>Média</i>	17.00	11.50	14.50	21.00	21.00	21.00
	<i>Máximo</i>	24.00	20.00	18.00	21.00	21.00	21.00
	<i>Desvio padrão</i>	6.68	5.85	2.06	N.A	N.A	N.A

|A| = Número de arcos; |N| = Número de vértices; CNC = Coefficient of Connectivity; CN = Cyclomatic Number; N.A = Não Avaliado

Os agrupamentos obtidos com as abordagens *Sequence Clustering*, *ActiTraC*, *Guilde Tree Miner*, *DWS* e *Trace Clustering* permitiram gerar modelos cuja complexidade mante-se igual ou menor aquela observada sobre o conjunto original de logs; i.e., que o de *log* de eventos Manufatura sem a utilização das técnicas de agrupamento em termos das métricas de complexidade (cf. Tabela 11). No entanto, existem diferenças nos agrupamentos produzidos por essas técnicas. A abordagem *Sequence Clustering* produz uma quantidade significativamente maior de vértices e arcos do que a abordagem *ActiTraC*. Entretanto, apesar da *Sequence Clustering* gerar apenas 3 agrupamentos, o resultado foi bem semelhante da abordagem *Guilde Tree Miner*, que gerou 4 agrupamentos. A abordagem *ActiTraC* gerou 4 agrupamentos e foi a abordagem que mais reduziu a complexidade dos modelos minerados. A Figura 29 apresenta aplicação do *Heuristic Miner* sobre : (a) o log de eventos de Manufatura; (b) os agrupamentos gerados pelo *ActiTraC*.

As abordagens *DWS* e *Trace Clustering* com seus parâmetros defaults geraram apenas um agrupamento, e obtiveram uma complexidade igual ao modelo gerado apenas utilizando o *Heuristic Miner*, *i.e.*, sem aplicar a tarefa de agrupamento.

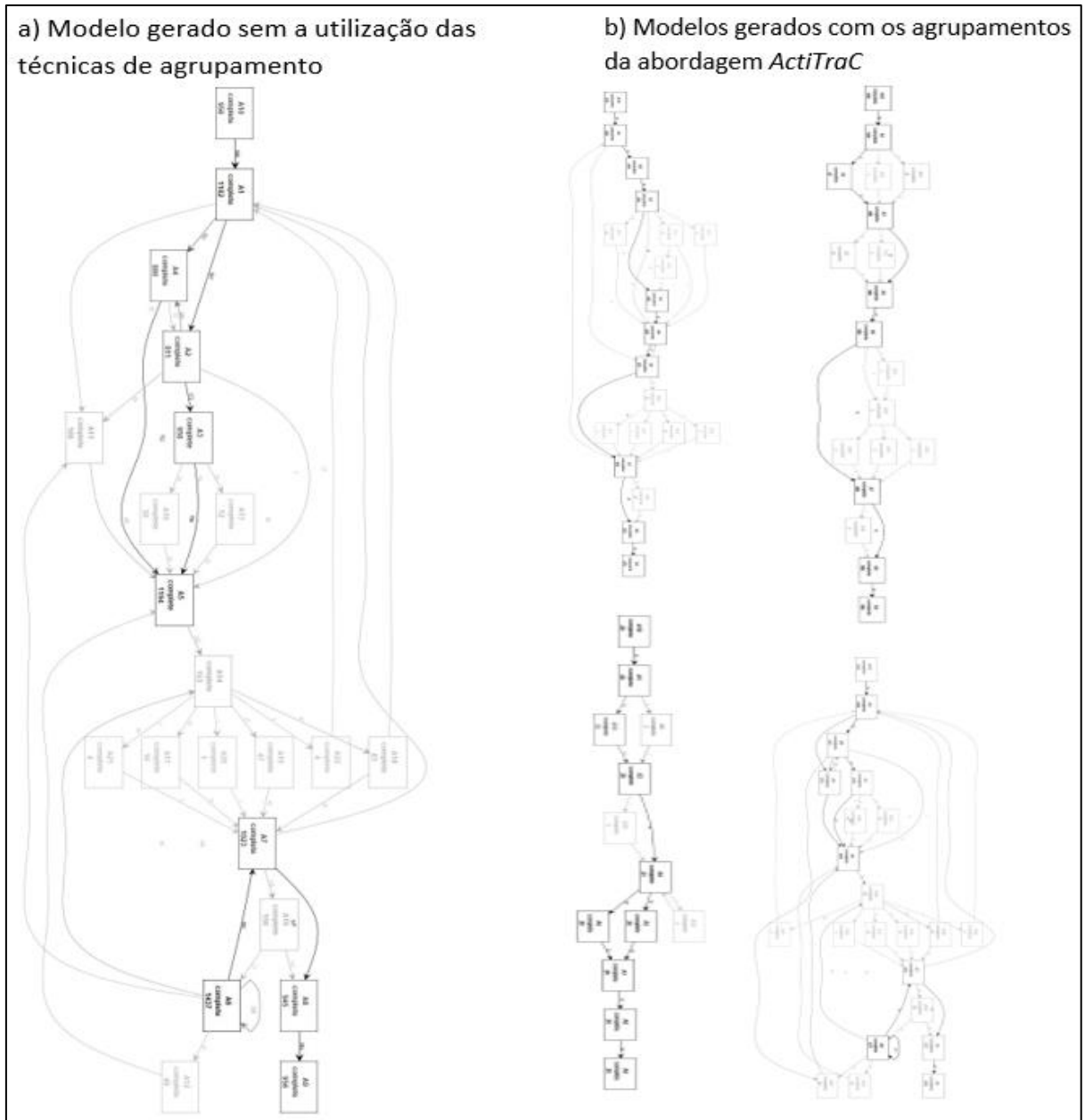


Figura 29. Log de eventos de Manufatura: (a) modelo gerado sem as técnicas de agrupamento. (b) modelos gerados a partir do resultado da aplicação da abordagem *ActiTraC* para gerar agrupamentos. Fonte: o Autor, 2019.

5.2.3 Log de eventos de Telecomunicações

Esta seção apresenta os resultados obtidos dos experimentos aplicados sobre logs de eventos reais de um processo de CRM de uma empresa da área de Telecom. Tal log de eventos encerra 17812 casos, 83.286 eventos e 42 atividades. A Tabela 12 apresenta os resultados obtidos, em que os valores são medidas resultantes das métricas de complexidade dos modelos.

Tabela 12 - Resultados das métricas de complexidade dos modelos gerados a partir das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos no log de eventos da área de Telecom. Fonte: o Autor, 2019.

		<i>Sequence Clustering</i>	<i>ActiTraC</i>	<i>Guilde Tree Miner</i>	<i>DWS</i>	<i>Trace Clustering</i>	<i>Somente Heuristic Miner</i>
	<i>k-cluster</i>	3.00	4.00	4.00	2.00	N.A	1.00
A	<i>Mínimo</i>	90.00	6.00	49.00	80.00	N.A	108.00
	<i>Média</i>	100.33	44.75	69.50	93.00	N.A	108.00
	<i>Máximo</i>	120.00	107.00	96.00	106.00	N.A	108.00
	<i>Desvio padrão</i>	13.91	37.76	18.17	13.00	N.A	N.A
N	<i>Mínimo</i>	38.00	5.00	26.00	36.00	N.A	40.00
	<i>Média</i>	38.00	19.50	32.50	36.50	N.A	40.0
	<i>Máximo</i>	38.00	40.00	38.00	37.00	N.A	40.00
	<i>Desvio padrão</i>	0.00	12.78	5.12	0.50	N.A	N.A
CNC	<i>Mínimo</i>	2.37	1.20	1.88	2.22	N.A	2.70
	<i>Média</i>	2.64	1.96	2.10	2.54	N.A	2.70
	<i>Máximo</i>	3.16	2.67	2.53	2.86	N.A	2.70
	<i>Desvio padrão</i>	0.37	0.53	0.25	0.32	N.A	N.A
CN	<i>Mínimo</i>	53.00	2.00	24.00	45.00	N.A	69.00
	<i>Média</i>	63.33	26.25	38.00	57.50	N.A	69.00
	<i>Máximo</i>	83.00	68.00	59.00	70.00	N.A	69.00
	<i>Desvio padrão</i>	13.91	25.04	13.43	12.50	N.A	N.A

|A| = Número de arcos; |N| = Número de vértices; CNC = *Coefficient of Connectivity*; CN = *Cyclomatic Number*; N.A = Não Avaliado

Os agrupamentos obtidos a partir das abordagens *Sequence Clustering*, *ActiTraC*, *Guilde Tree Miner*, *DWS* permitiram gerar modelos com complexidade menor do que aquela observada no conjunto original de logs de eventos; sem a utilização das técnicas de agrupamento com as métricas de complexidade exibidas na Tabela 12. As abordagens *Sequence Clustering* e *DWS* produziram uma quantidade significativamente maior de vértices e arcos do que as outras abordagens. Entretanto, tais abordagens geraram menor número de agrupamentos

que a demais abordagens. As abordagens *ActiTraC* e *Guilde Tree Miner* geraram quatro agrupamentos; elas foram as abordagens que mais reduziram a complexidade dos modelos minerados. Por fim, dado elevado custo computacional da abordagem *Trace Clustering*, não foi possível avaliá-la para *log* de eventos da área de Telecom.

5.2.4 Log de eventos Seguros

Esta seção apresenta os resultados obtidos dos experimentos aplicados sobre logs de eventos reais de um processo de recebimento de documentos da área de Seguros. Tal log de eventos encerra 12391 casos, 65.653 eventos e 70 atividades. A Tabela 13 apresenta os resultados obtidos, em que os valores são medidas resultantes das métricas de complexidade dos modelos.

Tabela 13 - Resultados das métricas de complexidade dos modelos gerados a partir das técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos no log de eventos da área de Seguros. Fonte: o Autor, 2019.

		<i>Sequence Clustering</i>	<i>ActiTraC</i>	<i>Guilde Tree Miner</i>	<i>DWS</i>	<i>Trace Clustering</i>	<i>Somente Heuristic Miner</i>
	<i>k-cluster</i>	3.00	4.00	4.00	1.00	N.A	1.00
A	<i>Mínimo</i>	150.00	4.00	65.00	218.00	N.A	185.00
	<i>Média</i>	160.67	105.5	107.50	218.00	N.A	185.00
	<i>Máximo</i>	178.00	187.00	172.00	218.00	N.A	185.00
	<i>Desvio padrão</i>	12.36	66.26	40.89	N.A	N.A	N.A
N	<i>Mínimo</i>	68.00	4.00	35.00	70.00	N.A	70.00
	<i>Média</i>	69.00	47.25	52.75	70.00	N.A	70.00
	<i>Máximo</i>	70.00	70.00	70.00	70.00	N.A	70.00
	<i>Desvio padrão</i>	0.81	25.69	13.92	N.A	N.A	N.A
CNC	<i>Mínimo</i>	2.20	1.00	1.81	3.11	N.A	2.64
	<i>Média</i>	2.32	1.91	1.99	3.11	N.A	2.64
	<i>Máximo</i>	2.54	2.67	2.46	3.11	N.A	2.64
	<i>Desvio padrão</i>	0.15	0.60	0.27	N.A	N.A	N.A
CN	<i>Mínimo</i>	83.00	1.00	31.00	149.00	N.A	116.00
	<i>Média</i>	92.67	59.25	55.75	149.00	N.A	116.00
	<i>Máximo</i>	109.00	118.00	103.00	149.00	N.A	116.00
	<i>Desvio padrão</i>	11.61	41.97	28.21	N.A	N.A	N.A

|A| = Número de arcos; |N| = Número de vértices; CNC = *Coefficient of Connectivity*; CN = *Cyclomatic Number*; N.A = Não Avaliado

Os agrupamentos obtidos a partir das abordagens *Sequence Clustering*, *ActiTraC* e *Guilde Tree Miner* permitiram gerar modelos com complexidade menor do que aquela observada no conjunto original de *logs* de eventos: sem a utilização das técnicas de agrupamento com as métricas de complexidade exibidas na Tabela 13. No entanto, existem diferenças substanciais nos agrupamentos produzidos por essas técnicas. A abordagem *Sequence Clustering* produz uma quantidade significativamente maior de vértices e arcos do que as outras abordagens. Todavia, essa abordagem gerou apenas 3 agrupamentos, enquanto as abordagens *ActiTraC* e *Guilde Tree Miner* geraram quatro agrupamentos; a abordagem *ActiTraC* foi que mais reduziu a complexidade dos modelos minerados.

A abordagem *DWS* com seus parâmetros *defaults* gerou apenas um agrupamento, e obteve uma complexidade maior que o modelo obtido com o *Heuristic Miner* sobre o conjunto original de *logs* de eventos. Por fim, dado elevado custo computacional da abordagem *Trace Clustering*, não foi possível avaliá-la para *log* de eventos da área de Seguros.

5.3 Análise dos Resultados

Baseado nos resultados apresentados é possível responder a segunda questão de pesquisa (RQ02), na qual **as técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos não foram capazes de separar um log de eventos contendo processos diferentes**. No entanto, os resultados das abordagens *ActiTraC* e *Guilde Tree Miner* foram promissores, na medida que elas produziram respectivamente uma taxa de acerto de 45 e 41 %.

A terceira questão de pesquisa (RQ03), refere-se se as diferentes técnicas de agrupamento utilizadas nas abordagens para descoberta de múltiplos modelos de processos impactam no resultado da mineração de processos. Conforme o resultado apresentado é possível estabelecer que as técnicas de agrupamento geram redução de complexidade, i.e., os modelos gerados a partir dos agrupamentos são de fáceis interpretação comparado ao modelo que não utilizou (cf. Figura 29). O gráfico da Figura 30 apresenta um comparativo das médias CNC e CN entre a abordagem *ActiTraC*, a qual obteve o melhor desempenho entre as abordagens, e o modelo obtido com o *Heuristic Miner* sobre os conjuntos originais de *logs* de eventos.

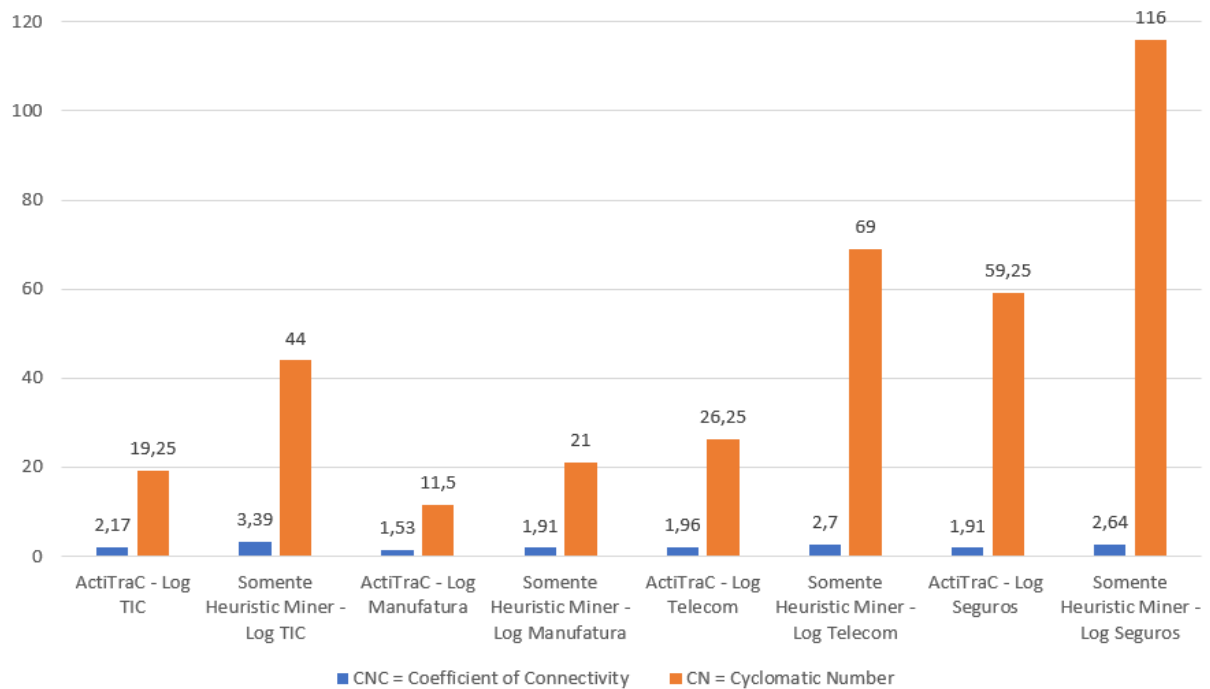


Figura 30. Comparativo CNC e CN. Fonte: o Autor, 2019.

Entretanto, foi utilizado o método estatístico não paramétrico de *Kruskal-Wallis* para verificar a hipótese de que o agrupamento de traços reduz a complexidade de múltiplos modelos de processos descobertos, o nível de significância α foi fixado em 0,05. A partir destes resultados observa-se que não há diferença estatística dos resultados de CNC e CN entre as abordagens e o modelo obtido com o *Heuristic Miner* sobre os conjuntos originais de *logs* de eventos (cf. Tabela 14).

Tabela 14 - Resultado do teste estatístico não paramétrico de *Kruskal-Wallis*. Fonte: o Autor, 2019.

Logs de eventos	p-value	
	CNC	CN
TIC	0.2133	0.2133
Manufatura	0.3728	0.3284
Telecomunicações	0.2318	0.1977
Seguros	0.2233	0.2373

5.4 Limitações da Pesquisa

O estudo realizado apresenta limitações referente a parametrização das abordagens, as quais poderiam ser modificadas em busca de melhores resultados. Em nossos experimentos foram utilizados apenas as parametrizações *defaults* dadas pelas abordagens no Cenário 2. No cenário 1 foram alteradas apenas as quantidades de agrupamentos gerados. Além disso, a técnica de descoberta empregada em binômio com as abordagens de agrupamentos foi apenas a *Heuristic Miner*; outras técnicas podem ser utilizadas, como exemplo: *Genetic Miner* e *Declarative Miner*.

5.5 Considerações sobre o Capítulo

Este capítulo apresentou a análise e discussão dos resultados com o objetivo de responder à questão de pesquisa RQ02 e RQ03. Tal análise foi fundamentada na capacidade, de um lado, das técnicas de agrupamento em separar, a partir de um único conjunto de eventos, os logs dos eventos de três processos diferentes e, de outro lado, das métricas CNC e CN mensurar a complexidade dos modelos minerados. Ao final foi apresentado um resumo e limitações da pesquisa.

CAPÍTULO 6 - CONCLUSÕES

A mineração de processos preenche a lacuna entre a análise tradicional de processos—baseada em modelos—e as técnicas de análise—centradas em dados, como já bem-conhecido no universo da aprendizagem de máquina. A mineração de processos busca confrontar dados de eventos—i.e., comportamentos observados—e modelos de processos—feitos manualmente ou descobertos automaticamente. A técnica é aplicável a qualquer tipo de processo operacional. Neste contexto, pode-se citar algumas áreas de aplicação: hospitalar, manufatura, telecomunicação, logística, auditoria, seguros, etc. Os exemplos observados nestas áreas possuem em comum o fato de que o comportamento dinâmico de cada aplicação precisa estar conexo ao seu modelo de processo. Essa conexão viva é sinônimo de *ciência de dados em ação*.

A mineração de processos é uma subárea de pesquisa que visa extrair representações compactas e abstratas de processos a partir de logs de eventos. Várias abordagens e técnicas foram desenvolvidas para abarcar e generalizar diferentes perspectivas de um processo a partir de logs de eventos. Tais técnicas foram aplicadas com sucesso sobre processos explicitamente projetados, bem-estruturados e rigorosamente aplicados. No entanto e propositalmente, a maioria dos processos na vida real não foram projetados e otimizados, ou nem sequer foram explicitamente definidos, mas eles evoluíram ao longo do tempo. Em tais situações, a aplicação da mineração de processos ganha seu devido destaque, na medida em que não se limita a reencontrar o que já se sabe.

Muitas experiências foram feitas com algoritmos de mineração para experimentar e testar processos da vida real. Observou-se, entretanto, que os algoritmos tendem a funcionar bem em processos estruturados; i.e., eles não fornecem modelos facilmente compreensíveis para processos semiestruturados. As técnicas existentes mais robustas produzem resultados corretos, na medida em que elas garantem que o modelo obtido seja correto no sentido de que a realidade se encaixa em tal modelo. A falta de abstração adequada por parte do modelo resultante é o problema. Em [Günther et al, 2007] o problema foi ilustrado com a seguinte situação: “quando se olha para o mapa de um país onde todas as cidades e bairros são representados por nós idênticos e todas as estradas são retratadas da mesma maneira, o mapa resultante está

correto, mas não é muito adequado”. Portanto, a partir desta ilustração, foi proposta a utilização do conceito de roteiro como metáfora para visualizar os modelos resultantes. Infere-se então que a relevância das atividades e suas relações deve ser destacada. Assim, as atividades e suas relações podem ser agrupadas ou removidas dependendo do seu papel no processo. Em outras palavras, operadores de *drill-up* e *drill-down* podem ser usados sobre o universo de registros de dados gerar abstrações mais adequadas ao que se deseja analisar. Uma forma de melhorar o resultado da mineração de processos, como etapa preliminar, é o agrupamento de instâncias de processos por meio da aplicação medida de similaridade. O agrupamento melhora a acurácia e compreensão dos modelos descobertos, bem como facilita a descoberta de variantes e desvios de processos.

Em geral, a escolha da técnica de agrupamento impacta diretamente na qualidade dos modelos gerados. Isto foi mostrado por meio do estudo ora realizado. A primeira parte do estudo limitou-se a responder a seguinte questão: Quais são até então as técnicas de agrupamento de instâncias similares de processos utilizados para a descoberta de múltiplos modelos de processos? A segunda parte do estudo dedicou-se segregar as técnicas de agrupamento de instâncias similares de processos utilizadas para a descoberta de múltiplos modelos de processos capazes de subdividir um log de eventos geral em subconjuntos disjuntos de eventos, em que cada subconjunto encerrasse os eventos de um processo diferente. Dado o conjunto de técnicas de agrupamento produzido na segunda parte do estudo, a terceira parte do estudo centrou-se em experimentar as técnicas selecionadas na descoberta de múltiplos modelos de processos, sendo que o norte do estudo foi avaliar o impacto de cada técnica de agrupamento *vis-à-vis* a redução da complexidade de cada modelo descoberto.

Para condução de tal estudo foram realizados experimentos, nos quais aplicaram-se diferentes técnicas de agrupamento, utilizadas em binômio com as abordagens de descoberta de múltiplos modelos de processos; foi usada apenas uma técnica de descoberta de modelos de processos, a saber: *Heuristic Miner*. A partir de tais experimentos foram possíveis estabelecer algumas conclusões, em que a primeira conclusão foi que a abordagem *Sequence Clustering* não foi capaz de subdividir um log de eventos geral em subconjuntos disjuntos de eventos, em que cada subconjunto encerrasse os eventos de um processo diferente, e a redução da complexidade dos modelos gerados foi menor comparado as abordagens *ActiTraC* e *Gilde Tree Miner*. A segunda conclusão foi que a abordagem *ActiTraC* gerou os resultados mais

promissores em subdividir um log de eventos geral em subconjuntos disjuntos de eventos (uma taxa de acerto de 45%), em que cada subconjunto encerrasse os eventos de um processo diferente. Ela também foi a abordagem que permitiu obter os melhores resultados quanto a redução da complexidade de cada modelo de processo minerado. A terceira conclusão foi que a abordagem *Guilde Tree Miner* gerou também resultados promissores em subdividir um log de eventos geral em subconjuntos disjuntos de eventos (uma taxa de acerto de 41%), em que cada subconjunto encerrasse os eventos de um processo diferente. Ela permitiu obter um bom desempenho quanto a redução da complexidade de modelo de processo minerado. Finalmente, as abordagens *DWS* e *Trace Clustering* não foram capazes de subdividir um log de eventos geral em subconjuntos disjuntos de eventos, em que cada subconjunto encerrasse os eventos de único processo. A abordagem *DWS* também foi ineficiente quanto a redução da complexidade de cada modelo descoberto. A abordagem *Trace Clustering* impôs alto custo computacional, não sendo possível utilizá-la em 3 logs de eventos dos experimentos. Entretanto, observou-se através do teste estatístico não paramétrico de *Kruskal-Wallis* que não há diferença estatística dos resultados de CNC e CN entre as abordagens e o modelo obtido com o *Heuristic Miner* sobre os conjuntos originais de logs de eventos.

Trabalhos Futuros

Sugere-se:

- implementar outras técnicas de agrupamento de instâncias similares de processos;
- fazer o uso de outros parâmetros de configuração de cada técnica de agrupamento para tentar encontrar melhores resultados;
- avaliar os resultados obtidos usando outras métricas, como por exemplo, *fitness*, generalização e precisão;
- refazer os experimentos usando diferentes mineradores de processos e avaliar a qualidade dos modelos gerados a partir argumentos; e
- generalizar os experimentos na forma de um framework de avaliação de técnicas de agrupamentos de instâncias similares de processos e descoberta de múltiplos processos.

REFERÊNCIAS BIBLIOGRÁFICAS

- A. A. Funkner, A. N. Yakovlev, and S. V. Kovalchuk, “Towards evolutionary discovery of typical clinical pathways in electronic health records,” *Procedia Computer Science*, vol. 119, pp. 234 – 244, 2017, 6th International Young Scientist Conference on Computational Science, YSC 2017, 01-03 November 2017, Kotka, Finland. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050917323918>
- A. J. M. M. Weijters, W. M. P. Van Der Aalst, and A. K. A. De Medeiros, “Process Mining with the Heuristics Miner Algorithm,” *Tech. Univ. Eindhoven, Tech. Rep. WP*, vol. 166, pp. 1–34, 2006.
- A. K. Alves de Medeiros and W. M. P. van der Aalst. 2008. Process Mining towards Semantics. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 35–80. DOI:https://doi.org/10.1007/978-3-540-89784-2_3.
- A. Marchetto and C. D. Francescomarino, “Parameterised trace selection technique for process model recovering,” *IET Software*, vol. 5, no. 6, pp. 563–575, Dec 2011.
- Álvaro Rebugue and Diogo R Ferreira. 2012. Business process analysis in healthcare environments: A methodology based on process mining. *Inf. Syst.* 37, 2 (April 2012), 99–116. DOI:<https://doi.org/10.1016/j.is.2011.01.003>.
- A. Meinheim, C. d. S. Garcia, J. C. Nievola, and E. E. Scalabrin, “Combining process mining with trace clustering: Manufacturing shop floor process - an applied case,” in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, Nov 2017, pp. 498–505.
- Andrea Delgado, Barbara Weber, Francisco Ruiz, Ignacio Garcia-Rodríguez De Guzmán, and Mario Piattini. 2014. An integrated approach based on execution measures for the continuous improvement of business processes realized by services. *Inf. Softw. Technol.* 56, 2 (2014), 134–162. DOI:<https://doi.org/10.1016/j.infsof.2013.08.003>.
- Andreas Schoknecht, Tom Thaler, Peter Fettke, Andreas Oberweis, and Ralf Laue. 2017. Similarity of Business Process Models—A State-of-the-Art Analysis. *ACM Comput. Surv.* 50, 4 (August 2017), 1–33. DOI:<https://doi.org/10.1145/3092694>
- Angelina Prima Kurniati, Owen Johnson, David Hogg, and Geoff Hall. 2016. Process mining in oncology: A literature review. *Proc. 6th Int. Conf. Inf. Commun. Manag. ICICM 2016 i* (2016), 291–297. DOI:<https://doi.org/10.1109/INFOCOMAN.2016.7784260>
- A. Rozinat, I.S.M. de Jong, C.W. Gunther, and W.M.P. van der Aalst. 2009. Process Mining Applied to the Test Process of Wafer Scanners in ASML. *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.* 39, 4 (July 2009), 474–479. DOI:<https://doi.org/10.1109/TSMCC.2009.2014169>.
- Arthur M. Valle, Eduardo A.P. Santos, Eduardo R. Loures. Applying process mining techniques in software process appraisals. *Information and Software Technology*, Volume 87, July 2017, Pages 19-31.

- Baker, K.; Dunwoodie, E.; Jones, R.G.; Newsham, A.; Johnson, O.; Price, C.P.; Wolstenholme, J.; Leal, J.; Mcginley, P.; Twelves, C.; Hall, G., “Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy,” *Int. J. Med. Inform.*, vol. 103, pp. 32–41, 2017.
- B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst. 2005. The ProM Framework: A New Era in Process Mining Tool Support. In *Applications and Theory of Petri Nets 2005*. 444–454. DOI:https://doi.org/10.1007/11494744_25.
- B. F. Van Dongen and W. M. P. van der Aalst. Multi-phase process mining: Aggregating instance graphs into EPCs and Petri nets. 2005
- B. F. Van Dongen, “Real-life event logs - Hospital log.” Eindhoven University of Technology, p. , 2011.
- B. F. Van Dongen, “BPI Challenge 2012.” Eindhoven University of Technology, p. , 2012.
- B. Kitchenham, “Procedures for performing systematic reviews,” KeeleUniversity, Department of Computer Science, Keele, Newcastle ST55BG, UK, Technical Report TR/SE-0401, 2004.
- Basole, R. C.; Braunstein, M. L.; Kumar, V.; Park, H.; Kahng, M.; Chau, D. H.; Tamersoy, A.; Hirsh, D. A.; Serban, N.; Bost, J.; Lesnick, B.; Schissel, B. L.; Thompson, M. Understanding variations in pediatric asthma care processes in the emergency department using visual analytics. *Journal of the American Medical Informatics Association*, v. 22, n. 2, p. 318-323, 2015.
- Burcu Akman and Onur Demirörs. 2009. Applicability of Process Discovery Algorithms for Software Organizations. In *2009 35th Euromicro Conference on Software Engineering and Advanced Applications*, 195–202. DOI:<https://doi.org/10.1109/SEAA.2009.87>.
- Camilo Alvarez, Eric Rojas, Michael Arias, Jorge Munoz-Gama, Marcos Sepúlveda, Valeria Herskovic, Daniel Capurro. Discovering role interaction models in the Emergency Room using Process Mining. *Journal of biomedical informatics*, volume 78, pp 60-77, 2018.
- Caron, F.; Vanthienen, J.; Vanhaecht, K.; Limbergen, E. V.; Weerdt, J. de; Baesens, B. A process mining-based investigation of adverse events in care processes. *Health Information Management Journal*, v. 43, n. 1, p. 16-25, 2014
- Cho, M.; Song, M.; Yoo, S. A systematic methodology for outpatient process analysis based on process mining. *International Journal of Industrial Engineering-Theory Applications and Practice*, v. 22, n. 4, p. 480-493, 2015.
- Christian W. Günther and Wil M. P. van der Aalst. 2007. Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. *Business Process Management*. 5th Intern ed. Springer Berlin Heidelberg. 328–343. DOI:https://doi.org/10.1007/978-3-540-75183-0_24.
- D. Ferreira, M. Zacarias, M. Malheiros, and P. Ferreira, “Approaching process mining with sequence clustering: Experiments and findings,” in *Business Process Management*, G.

- Alonso, P. Dadam, and M. Rosemann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 360–374.
- de Medeiros, A.K.A., Guzzo, A., Greco, G., van der Aalst, W.M.P., Weijters, A.J.M.M., van Dongen, B.F., Saccà, D.: Process Mining Based on Clustering: A Quest for Precision. In: ter Hofstede, A., Benatallah, B., Paik, H.-Y. (eds.) Business Process Management Workshops, LNCS 4928, pp. 17-29. Springer, Berlin (2008)
- Edson Ruschel, Eduardo Alves Portela Santos, and Eduardo de Freitas Rocha Loures. 2017. Mining Shop-Floor Data for Preventive Maintenance Management: Integrating Probabilistic and Predictive Models. *Procedia Manuf.* 11, June (2017), 1127–1134.
- E. Lamma, P. Mello, M. Montali, F. Riguzzi, and S. Storari, “Inducing Declarative Logic-Based Models from Labeled Traces,” *Bus. Process Manag.*, vol. 4714 LNCS, pp. 344–359, 2007.
- Eric Rojas, Jorge Munoz-Gama, Marcos Sepúlveda, and Daniel Capurro. 2016. Process mining in healthcare: A literature review. *J. Biomed. Inform.* 61, (June 2016), 224–236. DOI:<https://doi.org/10.1016/j.jbi.2016.04.007>
- Evgeniy Vasilyev, Diogo R. Ferreira, and Junichi Iijima. 2013. Using Inductive Reasoning to Find the Cause of Process Delays. In 2013 IEEE 15th Conference on Business Informatics, 242–249. DOI:<https://doi.org/10.1109/CBI.2013.41>.
- Fernandez-Llatas, C.; Valdivieso, B.; Traver, V.; Benedi, J. M. Using Process Mining for Automatic Support of Clinical Pathways Design. *Data Mining in Clinical Medicine of the series Methods in Molecular Biology*, v. 1246, p. 79-88, 2015.
- F. Mannhardt, “Sepsis Cases - Event Log.” Eindhoven University of Technology, p. , 2016.
- Garcia, A. O.; Ramirez, Y. E. P.; Armenteros, O. U. L. Process mining in healthcare: analysis and modeling of processes in the emergency área. *IEEE Latin America Transactions*, v. 13, n. 5, p. 1612-1618, 2015.
- Geetika T. Lakshmanan, Davood Shamsi, Yurdaer N. Doganata, Merve Unuvar, and Rania Khalaf. 2015. A markov prediction model for data-driven semi-structured business processes. *Knowl. Inf. Syst.* 42, 1 (January 2015), 97–126. DOI:<https://doi.org/10.1007/s10115-013-0697-8>.
- G. Greco, A. Guzzo, L. Pontieri, and D. Sacca, “Discovering expressive process models by clustering log traces,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 8, pp. 1010–1027, Aug 2006.
- G. Greco, A. Guzzo, and L. Pontieri, “Mining taxonomies of process models,” *Data Knowledge Engineering*, vol. 67, no. 1, pp. 74 – 102, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169023X08000839>
- GIL, A. C. Como elaborar projetos de pesquisa. 4. ed. São Paulo: Atlas, 2002.
- Gupta, M. (2014). Nirikshan: Process mining software repositories to identify inefficiencies, imperfections, and enhance existing process capabilities. Companion Proceedings of the 36th International Conference on Software Engineering. Hyderabad, India (pp. 658-661).

- G. T. S Ho and H. C. W Lau. 2006. Development of an OLAP-Fuzzy Based Process Mining System for Quality Improvement. In *Intelligent Information Processing III*. Springer US, Boston, MA, 243–258. DOI:https://doi.org/10.1007/978-0-387-44641-7_26.
- H. Ariouat, A. H. Cairns, K. Barkaoui, J. Akoka, and N. Khelifa, “A two-step clustering approach for improving educational process model discovery,” in 2016 IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), June 2016, pp. 38–43.
- Huang, Z.; Lu, X.; Duan, H.; Fan, W. Summarizing clinical pathways from event logs. *Journal of Biomedical Informatics*, v. 46, p. 111-127, 2013
- Huang, Z.; Dong, W.; Ji, L.; Gan, C.; Lu, X.; Duan, H. Discovery of Clinical Pathway Patterns from Event Logs Using Probabilistic Topic Models. *Journal of Biomedical Informatics*, v. 47, p. 39-57, 2014.
- Huang, Z.; Dong, W.; Bath, P.; Ji, L.; Duan, H. On mining latent treatment patterns from electronic medical records. *Data Mining and Knowledge Discovery*, v. 29, p. 914-949, 2015.
- IEEE Standards. 2016. P1849/D03, March 2016 Draft Standard for XES -eXtensible Event Stream -for achieving interoperability in event logs and event streams Draft Standard for XES -eXtensible Event Stream -for achieving. Retrieved from <http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html>
- Jan Recker and Jan Mendling. 2016. The state of the art of business process management research as published in the BPM conference: Recommendations for progressing the field. *Bus. Inf. Syst. Eng.* 58, 1 (2016), 55–72. DOI:<https://doi.org/10.1007/s12599-015-0411-3>
- Javier de San Pedro and Jordi Cortadella. 2016. Mining structured petri nets for the visualization of process behavior. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing - SAC '16*, 839–846. DOI:<https://doi.org/10.1145/2851613.2851645>
- J. C. A. M. Buijs, “Receipt phase of an environmental permit application process (‘WABO’), CoSeLoG project.” Eindhoven University of Technology, p. , 2014.
- J. Chen, L. Sun, C. Guo, W. Wei, and Y. Xie, “A data-driven framework of typical treatment process extraction and evaluation,” *Journal of Biomedical Informatics*, vol. 83, pp. 178 – 195, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1532046418301102>
- J. De Weerd, S. K. L. M. vanden Broucke, J. Vanthienen, and B. Baesens, “Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes,” in 2012 IEEE Congress on Evolutionary Computation, June 2012, pp. 1–8.
- J. De Weerd, S. Vanden Broucke, J. Vanthienen, and B. Baesens, “Active trace clustering for improved process discovery,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 12, pp. 2708–2720, 2013.

- J. H. Bettencourt-Silva, J. Clark, C. S. Cooper, R. Mills, V. J. Rayward-Smith, and B. de la Iglesia, "Building Data-Driven Pathways From Routinely Collected Hospital Data: A Case Study on Prostate Cancer," *JMIR Med. Informatics*, vol. 3, no. 3, p. e26, 2015
- J. Carmona, J. Cortadella, and M. Kishinevsky. A region-based algorithm for discovering Petri nets from event logs. In *BPM'08*, pages 358–373, 2008.
- Jianmin Wang, Raymond K. Wong, Jianwei Ding, Qinlong Guo, and Lijie Wen. 2013. Efficient Selection of Process Mining Algorithms. *IEEE Trans. Serv. Comput.* 6, 4 (October 2013), 484–496. DOI:<https://doi.org/10.1109/TSC.2012.20>
- J. M. van der Werf, B. F. van Dongen, C. A. J. Hurkens, K. M. van Hee, and A. Serebrenik, "Process Discovery using Integer Linear Programming," *Appl. Theory Petri Nets*, vol. 5062, no. 3–4, pp. 368–387, 2008.
- Jochen De Weerd, Manu De Backer, Jan Vanthienen, and Bart Baesens. 2012. A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Inf. Syst.* 37, 7 (2012), 654–676. DOI:<https://doi.org/10.1016/j.is.2012.02.004>
- Johannes De Smedt, Jochen De Weerd, and Jan Vanthienen. 2014. Multi-Paradigm Process Mining: Retrieving Better Models by Combining Rules and Sequences. *SSRN Electron. J.* 8841, (2014), 446–453. DOI:<https://doi.org/10.2139/ssrn.2464066>
- Jorge Munoz-Gama and Isao Echizen. 2012. Insuring Sensitive Processes through Process Mining. In *2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing*, 447–454. DOI:<https://doi.org/10.1109/UIC-ATC.2012.83>
- J. Park, D. Lee, and J. Zhu, "An integrated approach for ship block manufacturing process performance evaluation: Case from a korean shipbuilding company," *International Journal of Production Economics*, vol. 156, pp. 214 – 222, 2014. [Online]. Available:<http://www.sciencedirect.com/science/article/pii/S0925527314001959>
- Kim, E.; Kim, S.; Song, M.; Kim, S.; Yoo, D.; Hwang, H.; Yoo, S. Discovery of Outpatient Care Process of a Tertiary University Hospital Using Process Mining. *Healthcare Informatics Research*, v. 19, n. 1, p. 42-49, 2013.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3. *Engineering*, v. 45, n. 4ve, p. 1051.
- Lamma, E., Mello, P., Montali, M., Riguzzi, F. and Storari, S. (2007). Inducing declarative logic-based models from labeled traces. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, v. 4714 LNCS, p. 344–359.
- Lihi Raichelson, Pnina Soffer, and Eric Verbeek. 2017. Merging event logs: Combining granularity levels for process flow analysis. *Inf. Syst.* 71, (November 2017), 211–227. DOI:<https://doi.org/10.1016/j.is.2017.08.010>

- Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil M.P. van der Aalst. 2013. A Framework for the Systematic Comparison and Evaluation of Compliance Monitoring Approaches. In *2013, 17th IEEE International Enterprise Distributed Object Computing Conference*, 7–16. DOI:<https://doi.org/10.1109/EDOC.2013.11>
- Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil M P Van Der Aalst. 2015. Compliance monitoring in business processes: Functionalities, application, and tool-support. *Inf. Syst.* 54, (2015), 209–234. DOI:<https://doi.org/10.1016/j.is.2015.02.007>
- Mans, R., Van Der Aalst, W. M. P., Vanwersch, R. J., Moleman, A. J. Process mining in healthcare: Data challenges when answering frequently posed questions. In: *Process Support and Knowledge Representation in Health Care*. Springer, Berlin, Heidelberg, 2013. p. 140-153.
- Mans, Ronny S.; Van Der Aalst, W. M. P; Vanwersch, Rob JB. *Process mining in healthcare: evaluating and exploiting operational healthcare processes*. Heidelberg: Springer, 2015.
- Maria Leitner and Stefanie Rinderle-Ma. 2014. A systematic review on security in Process-Aware Information Systems - Constitution, challenges, and future directions. *Inf. Softw. Technol.* 56, 3 (2014), 273–293. DOI:<https://doi.org/10.1016/j.infsof.2013.12.004>
- Martin Prodel, Vincent Augusto, Baptiste Jouaneton, Ludovic Lamarsalle, and Xiaolan Xie. 2018. *Optimal Process Mining for Large and Complex Event Logs*. (2018), 1–17.
- M. De Leoni and F. Mannhardt, “Road Traffic Fine Management Process.” Eindhoven University of Technology, p., 2015. APÊNDICES
- M. d. P. Villamil, D. Barrera, N. Velasco, O. Bernal, E. Fajardo, C. Urango, and S. Buitrago, “Strategies for the quality assessment of the health care service providers in the treatment of gastric cancer in colombia,” *BMC Health Services Research*, vol. 17, no. 1, p. 654, Sep 2017. [Online]. Available: <https://doi.org/10.1186/s12913-017-2440-8>
- Meincheim, Alex. *Uma contribuição ao estudo de medidas de similaridade aplicadas no agrupamento incremental de instâncias de processos* Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2018.
- Melcher, J.: *Process Measurement in Business Process Management – Theoretical Framework and Analysis of Several Aspects*. KIT Scientific Publishing, Karlsruhe (2012)
- M. Gupta "Improving software maintenance using process mining and predictive analytics" *IEEE International Conference on Software Maintenance and Evolution (ICSME) 2017*.
- M.H. Jansen-Vullers, W.M.P. van der Aalst, and M. Rosemann. 2006. Mining configurable enterprise information systems. *Data Knowl. Eng.* 56, 3 (March 2006), 195–244. DOI:<https://doi.org/10.1016/j.datak.2005.03.007>
- Mirko Polato, Alessandro Sperduti, Andrea Burattin, and Massimiliano de Leoni. 2014. Data-aware remaining time prediction of business process instances. In *2014 International Joint Conference on Neural Networks (IJCNN)*, 816–823. DOI:<https://doi.org/10.1109/IJCNN.2014.6889360>.

- M. La Rosa, W. M. P. Van Der Aalst, M. Dumas, and F. P. Milani, "Business Process Variability Modeling: A Survey," *ACM Comput. Surv.*, vol. 50, no. 1, p. 2:1--2:45, 2017.
- Mor Peleg. 2013. Computer-interpretable clinical guidelines: A methodological review. *J. Biomed. Inform.* 46, 4 (2013), 744–763. DOI:<https://doi.org/10.1016/j.jbi.2013.06.009>
- M. Song, C. W. Günther, and W. M. P. Van Der Aalst, "Trace clustering in process mining," *Lect. Notes Bus. Inf. Process.*, vol. 17 LNBIP, pp. 109–120, 2009.
- M. Song, H. Yang, S. H. Siadat, and M. Pechenizkiy, "A comparative study of dimensionality reduction techniques to enhance trace clustering performances," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3722–3737, 2013.
- Nicolas Mundbrod, Florian Beuter, and Manfred Reichert. 2015. Supporting Knowledge-Intensive Processes through Integrated Task Lifecycle Support. *Proc. - IEEE Int. Enterp. Distrib. Object Comput. Work. EDOCW 2015–Novem, (2015)*, 19–28. DOI:<https://doi.org/10.1109/EDOC.2015.13>.
- N. Tax, N. Sidorova, W. M. P. van der Aalst, and R. Haakma, "Heuristic approaches for generating local process models through log projections," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI), Dec 2016*, pp. 1–8.
- Paster, F.; Helm, E. From IHE audit trails to XES event logs facilitating process mining. *Digital Healthcare Empowering Europeans: Proceedings of MIE 2015*, v. 210, p. 40-44, 2015.
- Petersen, K., Feldt, R., Mujtaba, S. and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *EASE '08: Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*. <https://www.isl.ce.yildiz.edu.tr/personal/maktas/file/3599/systemticmapping.pdf>.
- R. Accorsi and T. Stocker, "Discovering workflow changes with timebased trace clustering," in *Data-Driven Process Discovery and Analysis*, K. Aberer, E. Damiani, and T. Dillon, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 154–168.
- R. Bose and W. Van Der Aalst, "Context Aware Trace Clustering: Towards Improving Process Mining Results." *Sdm*, pp. 401–412, 2009.
- Remco Dijkman, Marlon Dumas, Boudewijn van Dongen, Reina Käärrik, and Jan Mendling. 2011. Similarity of business process models: Metrics and evaluation. *Inf. Syst.* 36, 2 (April 2011), 498–516. DOI:<https://doi.org/10.1016/j.is.2010.09.006>.
- P. Juneja, D. Kundra, and A. Sureka, "Anvaya: An Algorithm and Case-Study on Improving the Goodness of Software Process Models Generated by Mining Event-Log Data in Issue Tracking Systems," in *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1. IEEE, jun 2016, pp. 53–62. [Online]. Available: <http://ieeexplore.ieee.org/document/7551993/>
- R. M. S. Santos, T. C. Oliveira, and F. B. e Abreu, "Mining software development process variations," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*, ser. SAC '15. New York, NY, USA: ACM, 2015, pp. 1657–1660. [Online]. Available: <http://doi.acm.org/10.1145/2695664.2696046>

- Riera Cruañas. Process Mining Opportunities for CMMI Assessments Universitat Politècnica de Catalunya BarcelonaTech. Master Thesis. 2012. <http://hdl.handle.net/2099.1/14776>
- Roberto Dos Santos Rocha and Marcelo Fantinato. 2013. The use of software product lines for business process management: A systematic literature review. *Inf. Softw. Technol.* 55, 8 (2013), 1355–1373. DOI:<https://doi.org/10.1016/j.infsof.2013.02.007>
- Rojas, E., Sepúlveda, M., Muhoz-Gama, J., Capurro, D., Traver, V., Fernandez-Llatas, C. Question-driven methodology for analyzing emergency room processes using process mining. *Applied Sciences*, 7(3), 302. 2017.
- Ronny S. Mans, Wil M. P. van der Aalst, Rob J Vanwersch. 2015. Process mining in healthcare: evaluating and exploiting operational healthcare processes. Heidelberg: Springer
- Rovani, M.; Maggi, F. M.; Leoni, M. de; Aalst, W. M. P. van der.; Mans, R. S.; Pepino, A. Declarative Process Mining in Healthcare. *Expert Systems with Applications*, v. 42, p. 9236-9251, 2015.
- R. P. J. C. Bose and W. M. P. van der Aalst, “Trace clustering based on conserved patterns: Towards achieving better process models,” in *Business Process Management Workshops*, S. Rinderle-Ma, S. Sadiq, and F. Leymann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 170–181.
- Rubin V., Günther C.W., van der Aalst W.M.P., Kindler E., van Dongen B.F., Schäfer W. (2007) Process Mining Framework for Software Processes. In: Wang Q., Pfahl D., Raffo D.M. (eds) *Software Process Dynamics and Agility*. ICSP 2007. Lecture Notes in Computer Science, vol 4470. Springer, Berlin, Heidelberg.
- Ryan K. L. Ko. 2009. A Computer Scientist’s Introductory Guide to Business Process Management (BPM). *Crossroads XRDS* 15, 4 (2009), 11–18. DOI:<https://doi.org/10.1145/1558897.1558901>
- Samalikova, J. Process mining application in software process assessment. Eindhoven: Technische Universiteit Eindhoven. ((Co-)promot.: prof.dr. R.J. Kusters, prof.dr.ir. P.W.P.J. Grefen & dr.ir. J.J.M. Trienekens). 2012.
- Sander J J Leemans, Dirk Fahland, and Wil M. P. van der Aalst. 2013. Discovering Block-Structured Process Models from Event Logs - A Constructive Approach. . 311–329. DOI:https://doi.org/10.1007/978-3-642-38697-8_17
- S. J. J. Leemans, D. Fahland, and W. M. P. Van Der Aalst, “Discovering block-structured process models from event logs - A constructive approach,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7927 LNCS, pp. 311–329, 2013.
- S. K. Lee, B. Kim, M. Huh, S. Cho, S. Park, and D. Lee, “Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry,” *Expert Systems with Applications*, vol. 40, no. 1, pp. 83–95, 2013.

- T. G. Erdogan and A. Tarhan, "Process mining for healthcare process analytics," Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), pp. 24 543–24 567, January 2017.
- , "Systematic mapping of process mining studies in healthcare," IEEE Access, vol. 6, pp. 24 543–24 567, 2018.
- T. Guarda, M.F. Santos, M.F. Augusto, C. Silva, and F. Pinto. 2013. Process Mining: A framework proposal for Pervasive Business Intelligence. In *Information Systems and Technologies (CISTI), 2013 8th Iberian Conference on*, 1–4.
- TRIVIÑOS, A. N. S. Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987.
- Michael Westergaard, and Moe Wynn. 2012. IEEE Task Force on Process Mining - Process mining manifesto. *Lect. Notes Bus. Inf. Process.* 99 LNBIP, PART 1 (2012), 169–194. DOI:https://doi.org/10.1007/978-3-642-28108-2_19
- van Dongen, B.F., Adriansyah, A.: Process Mining: Fuzzy Clustering and Performance Visualization. In: Rinderle-Ma, S., Sadiq, S., Leymann, F. (eds.) Business Process Management Workshops, vol. 43, pp. 158-169. Springer Berlin Heidelberg (2010)
- Veiga, G.M., Ferreira, D.R.: Understanding Spaghetti Models with Sequence Clustering for ProM. In: Rinderle-Ma, S., Sadiq, S., Leymann, F. (eds.) Business Process Management Workshops, vol. 43, pp. 92-103. Springer Berlin Heidelberg (2010)
- V. Mendes, E. R. F. Junior, C. Garcia, and A. Malucelli, "Kanban and process mining in the task management," in Proceedings of the 17th Brazilian Symposium on Software Quality, ser. SBQS. New York, NY, USA: ACM, 2018, pp. 269–278. [Online]. Available: <http://doi.acm.org/10.1145/3275245.3275286>
- Wil M. P. van der Aalst, W. M. P., Weijters, A. J. M. M. and Maruster, L. (2002). Workflow Mining: Which processes can be rediscovered? In Eindhoven University of Technology
- Wil M.P. van der Aalst, Boudewijn F. Van Dongen, Joachim Herbst, Laura Maruster, Guido Schimm, and A. J.M.M. Weijters. 2003. Workflow mining: A survey of issues and approaches. *Data and Knowledge Engineering*, v. 47, n. 2 (2003), 237–267.
- Wil M. P. van der Aalst and A. J.M.M. Weijters. 2004. Process mining: A research agenda. *Comput. Ind.* 53, 3 (2004), 231–244. DOI:<https://doi.org/10.1016/j.compind.2003.10.001>.
- Wil van der Aalst, W., Weijters, T. and Maruster, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, v. 16, n. 9, p. 1128–1142.
- Wil M. P. van der Aalst, A. K. Alves de Medeiros, and A. J. M. M. Weijters. 2005. Genetic Process Mining. 48–69. DOI:https://doi.org/10.1007/11494744_5.
- Wil M. P. van der Aalst. 2011a. Process Mining. Berlin, Heidelberg: Springer Berlin Heidelberg. v. 53

- Wil M. P. van der Aalst. 2011b. On the Representational Bias in Process Mining. In *2011 IEEE 20th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2–7. DOI:<https://doi.org/10.1109/WETICE.2011.64>
- Wil van der Aalst. 2011c. Process mining: discovering and improving Spaghetti and Lasagna processes. *2011 IEEE Symp. Comput. Intell. Data Min.* c (2011), 1–7. DOI:<https://doi.org/10.1109/CIDM.2011.6129461>.
- Wil M. P. van der Aalst. 2012. Process mining: Overview and opportunities. *ACM Trans. Manage. Inf. Syst.* 3, 2, Article 7 (July 2012), 17 pages. DOI:<https://doi.org/10.1145/2229156.2229157>
- Wil M. P. van der Aalst, Arya Adriansyah, Ana Karla Alves De Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter Van Den Brand, Ronald Brandtjen, Joos Buijs, Andrea Burattin, Josep Carmona, Malu Castellanos, Jan Claes, Jonathan Cook, Nicola Costantini, Francisco Curbera, Ernesto Damiani, Massimiliano De Leoni, Pavlos Delias, Boudewijn F. Van Dongen, Marlon Dumas, Schahram Dustdar, Dirk Fahland, Diogo R. Ferreira, Walid Gaaloul, Frank Van Geffen, Sukriti Goel, Christian Günther, Antonella Guzzo, Paul Harmon, Arthur Ter Hofstede, John Hoogland, Jon Espen Ingvaldsen, Koki Kato, Rudolf Kuhn, Akhil Kumar, Marcello La Rosa, Fabrizio Maggi, Donato Malerba, Ronny S. Mans, Alberto Manuel, Martin McCreesh, Paola Mello, Jan Mendling, Marco Montali, Hamid R. Motahari-Nezhad, Michael Zur Muehlen, Jorge Munoz-Gama, Luigi Pontieri, Joel Ribeiro, Anne Rozinat, Hugo Seguel Pérez, Ricardo Seguel Pérez, Marcos Sepúlveda, Jim Sinur, Pnina Soffer, Minseok Song, Alessandro Sperduti, Giovanni Stilo, Casper Stoel, Keith Swenson, Maurizio Talamo, Wei Tan, Chris Turner, Jan Vanthienen, George Varvaressos, Eric Verbeek, Marc Verdonk, Roberto Vigo, Jianmin Wang, Barbara Weber, Matthias Weidlich, Ton Weijters, Lijie Wen, Michael Westergaard, and Moe Wynn. 2012. IEEE Task Force on Process Mining - Process mining manifesto. *Lect. Notes Bus. Inf. Process.* 99 LNBIP, PART 1 (2012), 169–194. DOI:https://doi.org/10.1007/978-3-642-28108-2_19.
- Wil M. P. van der Aalst. 2016. *Process Mining: Data Science in Action* (second ed.). Springer Berlin Heidelberg, Berlin, Heidelberg. DOI:<https://doi.org/10.1007/978-3-662-49851-4>
- Weijters, a. J. M. M. and van der Aalst, W. M. P. (2003). Rediscovering Workflow Models from Event-Based Data using Little Thumb. *Integrated Computer Aided Engineering*, v. 10, p. 151–162.
- Wei Yang and Qiang Su. 2014. Process mining for clinical pathway: Literature review and future directions. In *2014 11th International Conference on Service Systems and Service Management (ICSSSM)*, 1–5. DOI:<https://doi.org/10.1109/ICSSSM.2014.6943412>
- X. Liu, H. Liu, and C. Ding, “Incorporating user behavior patterns to discover workflow models from event logs,” in *2013 IEEE 20th International Conference on Web Services*, June 2013, pp. 171–178.

- X. Liu, M. Alshangiti, C. Ding, and Q. Yu, “Log sequence clustering for workflow mining in multi-workflow systems,” *Data & Knowledge Engineering*, vol. 117, pp. 1 – 17, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169023X16301355>
- Y. Chen, A. N. Kho, D. Liebovitz, C. Ivory, S. Osmundson, J. Bian, and B. A. Malin, “Learning bundled care opportunities from electronic medical records,” *Journal of Biomedical Informatics*, vol. 77, pp. 1 – 10, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1532046417302605>
- Y. Djenouri, A. Belhadi, and P. Fournier-Viger, “Extracting useful knowledge from event logs: A frequent itemset mining approach,” *Knowledge-Based Systems*, vol. 139, pp. 132–148, 2018.