

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
MESTRADO EM ENGENHARIA DE PRODUÇÃO E SISTEMAS**

**MINERAÇÃO DE PROCESSOS APLICADA NA
GESTÃO DE SUPRIMENTOS: UM ESTUDO DE
CASO COM ERP BRASILEIRO**

Ederson Schwaickardt

2019

MINERAÇÃO DE PROCESSOS APLICADA NA GESTÃO DE SUPRIMENTOS: UM ESTUDO DE CASO COM ERP BRASILEIRO

Ederson Schwaickardt

Dissertação de mestrado apresentada ao Programa de Mestrado em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás como requisito para a obtenção do grau de Mestre em Engenharia de Produção e sistemas.

Orientador: Prof. Dra. Maria José Pereira Dantas

Goiânia – GO

2019

S398m Schwaickardt, Ederson

Mineração de processos aplicada na gestão de suprimentos
: um estudo de caso com ERP brasileiro / Ederson Schwaickardt.--
2019.

98 f.; il.

Texto em português, com resumo em inglês

Dissertação (mestrado) -- Pontifícia Universidade
Católica de Goiás, Escola de Engenharia, Goiânia,
2019

Inclui referências: f. 90-93

1. Mineração de processos. 2. Processos.
3. Compras. I.Dantas, Maria José Pereira. II.Pontifícia
Universidade Católica de Goiás - Programa de Pós-Graduação
em Engenharia de Produção e Sistemas - 2019. III.
Título.

CDU: 005.4(043)

MINERAÇÃO DE PROCESSOS APLICADA NA GESTÃO DE SUPRIMENTOS: UM ESTUDO DE CASO COM ERP BRASILEIRO

EDERSON SCHWAICKARDT

Esta Dissertação julgada adequada para obtenção do título de Mestre em Engenharia de Produção e Sistemas, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica de Goiás em 05/2019.



Prof. Marcos Lajovic Carneiro, Dr. em Engenharia Elétrica.


Coordenador do Programa de Pós-Graduação
em Engenharia de Produção e Sistemas

Banca Examinadora:



Prof. Maria José Pereira Dantas, Dra. em Engenharia Elétrica.

Orientador



Prof. Juliano Lopes de Oliveira, Dr. em Ciência da Computação.



Prof. José Elmo de Menezes, Dr. em Estatística.

Goiânia – Goiás

05/2019

RESUMO

Na realidade dos dias atuais, processos de negócio e dados são ativos muito importantes para as organizações na busca pelo diferencial competitivo. A mineração de processos busca vincular as áreas de gerenciamento de processos e a ciência de dados, possibilitando descobrir os processos reais realizados nos sistemas informatizados, identificar desvios e fraudes nos processos e ainda gerar ideias para melhorar os processos executados nas organizações. Esta pesquisa explora a área de mineração de processos no domínio da gestão de suprimentos, através de uma revisão exploratória da literatura contendo estudos de caso, que foram sumarizados nas perspectivas intraorganizacional e interorganizacional e, posteriormente aplicando em um estudo de caso numa empresa brasileira que utiliza o ERP Protheus. O método de revisão inclui rodadas de pesquisa com as palavras-chaves e o método para aplicar a mineração de processos na empresa estudada foi desenvolvido com base nas melhores práticas dos casos levantados e junção com o método PM² contendo as etapas de planejamento, extração, processamento de dados, mineração e análise, avaliação e melhoria e suporte de processos. As ferramentas utilizadas foram o ProM, com os algoritmos de Mineração Alfa, Mineração Heurística, Mineração *Fuzzy*, Mineração Indutiva e de Avaliação de Conformidade, assim como o software Celonis. Os resultados obtidos no estudo de caso sobre o processo de compras identificaram atividades não levantadas no modelo de referência, assim como casos não conformes e demonstraram como as técnicas de mineração de processos podem auxiliar no dia-a-dia a controlar e melhorar os processos de compras. Tais resultados, em conjunto com a análise dos casos na literatura demonstraram que a mineração de processo é promissora na gestão de processos.

Palavras-chave: Mineração de Processos, Processos de Compra, Gestão de Processos.

ABSTRACT

In today's reality, business processes and data are very important assets for organizations in the search for competitive differential. Process mining seeks to link the areas of process management and data science, making it possible to discover the actual processes carried out in computerized systems, to identify deviations and frauds in the processes and to generate ideas to improve the processes executed in the organizations. This research explores the area of process mining in the field of supply management, through a general review of the literature containing case studies, which were summarized in the intraorganizational and interorganizational perspectives and later applied in a case study in a Brazilian company that uses the ERP Protheus. The review method includes rounds of research with the keywords and the method to apply the process mining in the studied company was developed based on the best practices of the cases raised and joined with the method PM2 containing the steps of planning, extraction, processing data mining and analysis, evaluation and process improvement and support. The tools used were the ProM, with the algorithms of Alpha Mining, Heuristic Mining, Fuzzy Mining, Inductive Mining and of Conformity Assessment, as well as Celonis software. The results obtained in the case study on the procurement process identified activities not raised in the reference model, as well as non-compliant cases and demonstrated how process mining techniques can help in day-to-day control and improve purchasing. Such results, together with case analysis in the literature, have demonstrated that process mining is promising in process management.

Keywords: Process Mining, Procurement Process, Process Management.

Sumário

1	Introdução	10
1.1	Contextualização	10
1.2	Motivação	12
1.3	Objetivos.....	14
1.4	Organização da Dissertação	15
2	Fundamentação Teórica	16
2.1	Processos de Negócio	16
2.2	BPM – <i>Business Process Management</i>	17
2.3	Modelagens de Processos de Negócio	18
2.4	Mineração de Processos	19
2.4.1	Visão Geral.....	20
2.4.2	Perspectivas para Mineração de Processos.....	24
2.4.3	Algoritmos utilizados na Mineração de Processos.....	25
2.4.3.1	Algoritmo Alfa (<i>α algorithm</i>).....	25
2.4.3.2	Mineração Heurística (<i>Heuristics Miner</i>).....	28
2.4.3.3	Mineração Fuzzy (<i>Fuzzy Miner</i>)	30
2.4.3.4	Mineração Genética (<i>Genetic Miner</i>)	32
2.4.3.5	Mineração Indutiva (<i>Inductive Miner</i>).....	35
2.4.3.6	Algoritmos para verificar conformidade	37
2.4.4	Ferramentas de Mineração de Processos	38
2.4.5	Métricas de Análise na Mineração de Processos	38
3	Metodologia	42

3.1	Revisão Exploratória da literatura	42
3.2	Estudo exploratório de caso único e método PM ²	43
4	Revisão Exploratória de literatura: Mineração de Processos na Gestão de Suprimentos	47
4.1.1	Casos intraorganizacionais	49
4.1.2	Casos Interorganizacionais	55
4.1.3	Análise.....	57
5	Estudo de Caso único com método PM ²	62
5.1	Objetivos.....	62
5.2	Planejamento.....	62
5.3	Extração e Processamento dos dados.....	64
5.4	Mineração e Análises	67
5.5	Avaliação	84
5.6	Discussão dos Resultados	85
6	Conclusões	87
	Referências Bibliográficas.....	90
	Anexo 1 – Termo de autorização para o estudo	94

LISTA DE FIGURAS

Figura 1 - Estudo “ <i>Enterprise Application Software Worldwide, 2017 e 2016</i> ” emitido pelo Gartner. Fonte: Totvs (2018).....	13
Figura 2 - Exemplo de ciclo de vida BPM. Fonte: BRAZIL (2013)	17
Figura 3 Mineração de Processos: (a) descoberta, (b) verificação de conformidade e (c) extensão/melhoria. Fonte: Van der Aalst et al. (2012)	22
Figura 4 - Fluxo do processo minerado no prom com <i>inductive visual miner</i>	24
Figura 5 – Exemplo de rede P/T. Fonte: Van der Aalst et al. (2004)	26
Figura 6 - Modelo de processo com escolha não livre. Fonte: Weijters et al. (2006)....	30
Figura 7 - Visão geral da abordagem de mineração genética de processos. Fonte: Van der Aalst (2016)	33
Figura 8 - Tipos de cortes na mineração indutiva. Fonte: Leemans et al. (2013b).....	36
Figura 9 - Avaliação de métricas de qualidade em modelo de processo minerado. fonte: Rozinat et al. (2007).....	39
Figura 10 - Visão geral da metodologia PM ² . Adaptado de Van Eck et al. (2015).	44
Figura 11 - Cadeia de valor de Porter. Fonte: Porter (1986).....	47
Figura 12 - Esquema de partição vertical e horizontal de processos interorganizações. Fonte: (AALST, VAN DER, 2011).....	57
Figura 13 - Estrutura das tabelas de <i>logs</i> no ERP Protheus.....	64
Figura 14 - Tabelas de <i>logs</i> Pivotadas do ERP Protheus.....	65
Figura 15 - Exemplo Tratativa para Criação de Identificadores.....	66
Figura 16 - Estrutura da Tabela de <i>Logs</i>	66
Figura 17 - Painel dos Registros de Eventos.....	68
Figura 18 - Modelo de Processo Minerado com Algoritmo Alpha	69
Figura 19 - Modelo Minerado com Algoritmo Fuzzy (Linhas mais espessas representam os caminhos mais frequentes).....	70

Figura 20 - Modelo Minerado com Algoritmo Heurístico (Todos Eventos).....	71
Figura 21 - Modelo Minerado com Algoritmo Heurístico (10% de Frequência).....	72
Figura 22 - Modelo Minerado com Algoritmo Indutivo (Todos os Eventos)	73
Figura 23 - Modelo Minerado com Algoritmo Indutivo (80%)	74
Figura 24 - Modelo Minerado Filtrando Apenas Casos Completos.....	76
Figura 25 - Modelo Minerado com Celonis (Completo).....	78
Figura 26 - Modelo Minerado com Celonis (80/20).....	79
Figura 27 – Análise da Duração dos Casos.....	79
Figura 28 - Modelo Minerado com Celonis (Log Filtrado).....	80
Figura 29 - Modelo Iniciando no Pedido (Completo)	81
Figura 30 - Modelo Iniciando no Pedido (Log Filtrado)	82
Figura 31 - Emulação no Log Filtrado e Atividade Apenas com Nome da Rotina	83

LISTA DE TABELAS

Tabela 1 - Exemplo de registros de eventos (<i>logs</i>).....	23
Tabela 2 - Algoritmos e aplicações.....	45
Tabela 3 - Taxonomia de Regras de Negócios para Mineração de Processos (Adaptado de Caron et al. (2013a)).....	49
Tabela 4 - Síntese dos casos estudados.....	59
Tabela 5 – Estrutura das Tabelas de Auditoria do ERP Protheus	63
Tabela 6 - Classes de Atividades	67

LISTA DE SIGLAS

ABAP	<i>Advanced Business Application Programming</i> – Programação Avançada de Aplicativos de Negócio
BI	<i>Business Intelligence</i> – Inteligência de Negócio
BPM	<i>Business Process Management</i> – Gestão de Processos de Negócio
BPMN	<i>Business Process Management Notation</i> – Notação de Gestão de Processos de Negócio
BPMS	<i>Business Process Management System</i> – Sistema de Gestão de Processos de Negócio
CBOK	<i>Guide to the Business Process Management Body of Knowledge</i> – Guia para o Gerenciamento de Processos de Negócio – Corpo Comum de Conhecimento
CRM	<i>Customer Relationship Management</i> – Gestão de Relacionamento com cliente
CSV	<i>Comma Separated Values</i> – Valores Separados por Vírgula
DC	<i>Dotted Chart</i> – Gráfico Pontilhado
EDI	<i>Electronic Data Interchange</i> – Troca Eletrônica de Dados
EPC	<i>Event-driven Process Chain</i> – Cadeia de Processos Orientada a Eventos
ERP	<i>Enterprise Resource Planning</i> – Planejamento de Recursos Empresariais
EV	<i>Explore Variants</i> – Explorar Variantes
FLAV	<i>Filter Log by Attribute Values</i> – Filtrar Log por Valores de Atributo
FP	<i>Feature Prediction</i> – Previsão de Recursos
FSM	<i>Final State Machine</i> – Máquina de Estados Finita
IDEF	<i>Integrated Definition Language</i> – Linguagem de Definição Integrada
IvM	<i>Inductive Visual Miner</i> – Minerador Visual Indutivo
LV	<i>Log Visualizer</i> – Visualizador de Log

MRP	<i>Manufacturing Resources Planning</i> – Planejamento de Recursos de Manufatura
MRPII	<i>Manufacturing Resources Planning</i> – Planejamento de Recursos de Manufatura
MXML	<i>Mining Extensible Markup Language</i> – Linguagem de Marcação Extensível de Mineração
OLAP	<i>Online Analytical Processing</i> – Processamento Analítico On-Line
PAS	<i>Process Aware Systems</i> – Sistemas Conscientes do Processo
RFID	<i>Radio Frequency Identification</i> – Identificação por Radiofrequência
RLP/C	<i>Replay a Log for Performance/Conformance Analysis</i> – Repetir um Log para Análise de Desempenho/Conformidade
SCM	<i>Supply Chain Management System</i> – Sistema de Gestão de Cadeia de Suprimentos
SI	Sistema de Informação
UML	<i>Unified Modeling Language</i> – Linguagem de Modelagem Unificada
uSH	<i>Using Simple Heuristics</i> – Usando Heurísticas Simples
WfMC	<i>Workflow Management Coalition</i> – Coalizão de Gerenciamento de Fluxo de Trabalho
XES	<i>Extensible Event Stream</i> – Fluxo de Eventos Extensível

1 Introdução

Este trabalho discute uma pesquisa de dissertação de mestrado realizada na Pontifícia Universidade Católica de Goiás, em breve PUC-GO. Primeiro, o contexto da pesquisa é descrito. Em seguida, o objetivo da pesquisa é proposto. Para atingir este objetivo, várias questões de pesquisa são formuladas. Estas questões de pesquisa são respondidas usando um método de pesquisa.

1.1 Contextualização

Vivemos a era da experiência, na qual o consumidor está cada vez mais exigente, buscando consumir de marcas que tenham mais do que bons produtos e lhes ofereçam agregação de valor e boas experiências. Aspectos de diferenciação estão ligados à agilidade, qualidade e personalização do atendimento, facilidades, economia de tempo e dinheiro. Além da busca pelas melhores experiências, os consumidores também se tornaram mais conscientes de seus direitos e deveres e, assim, procuram selecionar as marcas mais sustentáveis e que contribuam positivamente para a sociedade.

Estudos de caso de centenas de indústrias, revelam que, no novo paradigma internacional de competitividade, as empresas que se sobressaem não são aquelas com os insumos mais baratos ou produção de maior escala, mas sim aquelas com capacidade de melhorar e inovar continuamente Porter; Linde (1995).

Segundo Porter (1986), uma organização pode alcançar a vantagem competitiva de duas maneiras: através do baixo custo relativo ou pela diferenciação. Isto significa que a organização precisa executar suas atividades ao longo da cadeia de valor a um custo menor ou de uma forma única em relação à concorrência.

Na busca por melhores espaços neste ambiente altamente competitivo, as organizações têm investido constantemente em iniciativas para melhorar seus processos, buscando reduzir desperdícios, melhorar seus controles mitigando falhas e riscos e buscando Sistemas de Informação (SI) para automatizar e apoiar seus processos. Sistemas como *Business Process Management System (BPMS)*, *Enterprise*

Resource Planning (ERP), *Customer Relationship Management (CRM)*, *Supply Chain Management System (SCM)* e outros, são exemplos de sistemas que apoiam diversos processos de negócio, gerenciando atividades, integrando áreas e gerindo informações do negócio.

Tais Sistemas de Informação (SI) produzem grandes quantidades de dados, que podem ser tratados e analisados para gerar informações úteis ao negócio, levando a melhores tomadas de decisões e maximização dos resultados. Sistemas, disciplinas e técnicas como *Business Intelligence (BI)*, *Online Analytical Processing (OLAP)*, *Machine Learning*, *Data Mining* e Estatística, apoiam a análise de dados.

Além de manter dados referentes às operações realizadas na empresa, como informações de clientes, fornecedores, produtos, pedidos de venda e outros, grande parte dos Sistemas de Informação (SI) é capaz de registrar históricos de execuções ou eventos (*logs*). Os *logs* geralmente possuem informações sobre qual atividade foi executada, por quem, em qual data e hora, e identificação do caso do processo Van der Aalst et al. (2004). Esses *logs* podem ser utilizados para entender como os processos estão sendo executados na organização e na melhoria e otimização dos processos Gandulfo; Ralha (2016).

Neste contexto, a mineração de processos é uma técnica que extrai conhecimento a partir destes registros de eventos, podendo ser útil para descobrir os reais processos sendo executados na organização, para realizar análises comparativas e, por fim, para melhorar os processos Van der Aalst et al. (2012).

A mineração de processos pode ser vista como uma disciplina que faz o elo entre ciência de dados e ciência de processos. Por um lado, abordagens de ciência de dados, como mineração de dados, estatística e aprendizado de máquina, geralmente não consideram modelos de processos ponta a ponta. Na outra ponta, as abordagens de ciência de processo têm seu cerne no processo em si, porém na maioria das vezes focam na modelagem de processos em vez de aprender com os dados. Este posicionamento único da mineração de processos, que se apoia em duas ciências

primordiais às organizações, torna a mineração de processos em uma ferramenta de grande potencial exploratório de dados e de otimização dos processos Van der Aalst (2016).

Se por um lado a mineração de processos demonstra grande potencial, existem alguns desafios intrínsecos à disciplina. A técnica, em si, apresenta alto nível de complexidade e só pode ser aplicada se um registro de eventos (*log*) estiver disponível. Como mencionado anteriormente, este *log* deve conter informações sobre os casos e os eventos que são executados neles. Embora muitos sistemas produzam *logs* de eventos, a maioria usa seu próprio formato de *log* de eventos, exigindo conhecimento do sistema e grande esforço de preparação dos *logs* para posterior mineração.

1.2 Motivação

A comercialização de produtos está intimamente ligada à aquisição de materiais, uma vez que a produção, distribuição e venda final aos clientes só pode ocorrer de forma ideal se os materiais necessários estejam disponíveis. A complexidade dos processos de suprimentos ao longo da cadeia de valor desencadeou o surgimento de sistemas de Planejamento das Necessidades Materiais (MRP), MRPII e, finalmente, os Sistemas de Planejamento Empresarial (ERP) (ER et al., 2018).

Sistemas ERP são amplamente utilizados para integrar e gerir áreas das organizações, possibilitando automação, armazenamento de informações e otimização das operações. Contudo, por processarem diariamente diversas transações espalhadas por diferentes áreas de negócios, muitos processos se tornam complexos e difíceis de rastrear ou de analisar e entender o processo como um todo.

Neste sentido, a mineração de processos oferece recursos que podem auxiliar na identificação dos reais processos sendo executados no ERP, na avaliação de modelos e normas e, em última instância, na melhoria dos processos para tornar a organização mais competitiva.

Na literatura, existem inúmeros casos de aplicação da mineração de processos em sistemas ERP, em especial o software SAP. A aplicação da mineração de processos em sistemas ERP demonstrou benefícios na identificação de falhas, gargalos, identificação de fraudes e otimização de processos ER et al. (2015, 2018 e ER; Astuti (2014 e Mahendrawathi et al. (2017 e Piessens (2011 e Segers (2007).

Entretanto, não foi identificado na literatura atual, nenhum caso que aplique a mineração de processos no sistema ERP Protheus da Totvs. O Protheus é um sistema ERP amplamente difundido no mercado brasileiro, atendendo organizações de diversos tamanhos nos mais variados setores de mercado.

Segundo estudo “*Enterprise Application Software Worldwide, 2017 e 2016*” emitido pelo Gartner e referenciado pela Totvs em seu relatório de resultados do segundo trimestre de 2018 (ilustrado na **Figura 1**), a empresa detém cerca de cinquenta por cento (50%) do mercado de ERP no Brasil, o que corrobora para a relevância do estudo de caso aplicado neste estudo Totvs (2018).

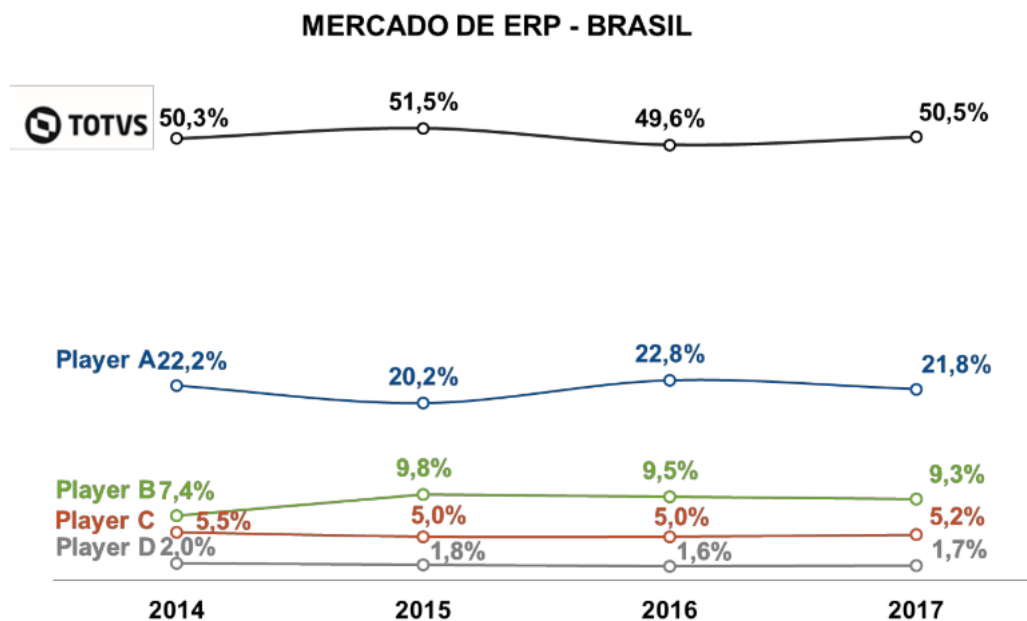


FIGURA 1 - ESTUDO “*ENTERPRISE APPLICATION SOFTWARE WORLDWIDE, 2017 E 2016*” EMITIDO PELO GARTNER. FONTE: Totvs (2018)

Dada a importância do ERP Protheus, principalmente para as pequenas e médias organizações, no mercado brasileiro, a segunda motivação deste trabalho é de aplicar

a mineração de processos em processos de suprimentos de uma empresa brasileira, criando um método para preparação e extração dos *logs* no sistema Protheus da Totvs, que posteriormente poderá ser replicado em outras áreas do negócio.

1.3 Objetivos

Esta pesquisa possui como objetivo principal investigar como a mineração de processos está sendo aplicada nos processos de apoio de gestão de suprimentos.

Para atender a esse objetivo, apresenta dois objetivos específicos:

- (1) Realizar uma revisão exploratória de estudos de casos que aplicam a mineração de processos na área de suprimentos, buscando responder as seguintes questões de pesquisa:
 - Q1: Quais são as características, resultados e tendências dos estudos de casos que aplicam a Mineração de processos na área de gestão de suprimentos?
 - Q2: Como a investigação dos arquivos de *logs* pode permitir a identificação de desvios e fraudes, denominada análise de conformidade, em uma organização?
- (2) Realizar um estudo de caso, afim, de explorar, descrever e avaliar a mineração de processos aplicada na área de gestão de suprimentos em uma empresa que utilize um ERP de relevância nacional, o sistema Protheus da empresa Totvs, aplicando aos dados minerados uma análise de conformidade. Um método de mineração de processos com as melhores práticas é aplicado sobre a estrutura do Protheus, descrevendo os passos envolvidos na preparação dos *logs*, aplicação da mineração de processos, análises conduzidas com métricas, e relatando ferramentas, algoritmos e técnicas utilizadas e, finalmente, os resultados alcançados neste processo investigativo.

1.4 Organização da Dissertação

Esta dissertação está organizada em conforme: no Capítulo 2 é apresentada uma breve revisão de conceitos e métodos relacionados a processos, gestão de processos de negócio, modelagem de processos e mineração de processos; no Capítulo 3 a metodologia de trabalho é apresentada e ilustrada; no Capítulo 4 são descritos e analisados estudos de casos em que a mineração de processos foi aplicada na área de Gestão de Suprimentos; no Capítulo 5 um estudo de caso é apresentado, demonstrando os passos envolvidos e analisando os resultados alcançados; e finalmente, no Capítulo 6 as conclusões e trabalhos futuros são expostos.

2 Fundamentação Teórica

Este capítulo apresenta os principais conceitos relacionados a processos, gerenciamento de processos de negócio, ou BPM (do inglês *Business Process Management*), modelagem de processos e mineração de processos de negócio. Os tipos e perspectivas de mineração de processos serão apresentados, assim como as principais ferramentas, algoritmos e métricas de análise empregadas na mineração de processos.

2.1 Processos de Negócio

Um processo de negócio pode ser definido como um conjunto de tarefas coordenadas que buscam atingir um objetivo de negócio.

No Guia de referência “Guia para o Gerenciamento de Processos de Negócio Corpo Comum de Conhecimento – CBOK”, encontra-se a seguinte definição de processo: “Um processo é uma agregação de atividades e comportamentos executados por humanos ou máquinas para alcançar um ou mais resultados” BRAZIL (2013). Já pela WfMC – Workflow Management Coalition WfMC (1999), um processo é definido como um conjunto de um ou mais procedimentos ou atividades vinculadas que coletivamente alcançam um objetivo do negócio, normalmente dentro do contexto de uma estrutura organizacional com papéis e relacionamentos funcionais.

Em BRAZIL (2013) classifica-se os processos de negócio em três categorias: primários, de suporte e de gerenciamento. Os processos primários estão diretamente ligados à agregação de valor aos clientes. Já os processos de suporte, são aqueles que visam apoiar os processos primários ou outros processos de suporte e gestão. Estes processos de suporte, apesar de não entregarem valor diretamente aos clientes, contribuem indiretamente para uma melhor experiência do cliente e melhoria na vantagem competitiva. Por fim, os processos de gerenciamento são responsáveis por medir, monitorar e controlar as atividades do negócio, buscando garantir que a organização alcance seus objetivos e metas de desempenho.

Os processos constituem parte fundamental do ecossistema de qualquer organização. Eles definem as entradas necessárias, os conjuntos de atividades a serem executadas, ordem das execuções e as saídas que devem ser entregues. Através de estudos e iniciativas sobre os processos é possível aumentar a eficiência, reduzir custos, garantir maior controle reduzindo riscos ao negócio, melhorar o engajamento dos colaboradores, em resumo, conseguir fazer mais com esforço menor e maior qualidade.

2.2 BPM – *Business Process Management*

O Gerenciamento de Processos de Negócios (BPM) é uma disciplina que visa gerenciar a organização integrando as estratégias do negócio com as necessidades dos clientes, por meio de processos ponta a ponta, ou seja, processos que consideram a transversalidade de diversas áreas, cargos e chefias de uma organização. Esta disciplina engloba estratégias, objetivos e metas, estruturas organizacionais, papéis, políticas, métodos e tecnologias para modelar, analisar, desenhar, medir e transformar os processos. O ciclo de vida BPM, exemplo ilustrado na **Figura 2**, implica em comprometimento contínuo da organização e seus colaboradores para com o gerenciamento e melhoria de seus processos BRAZIL (2013).

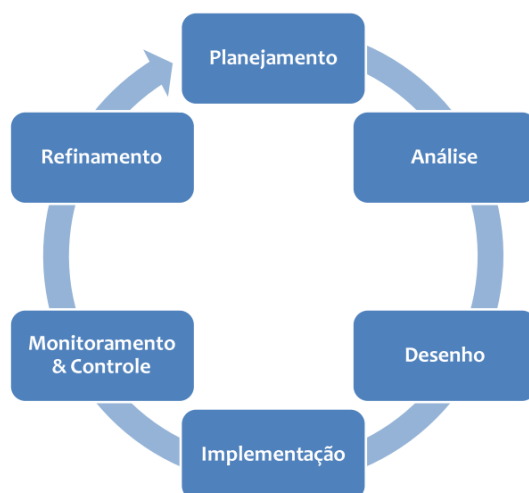


FIGURA 2 - EXEMPLO DE CICLO DE VIDA BPM. FONTE: BRAZIL (2013)

O ciclo exemplificado acima inicia-se na fase de projeto e análise, onde informações sobre o ambiente organizacional são coletadas, possibilitando que os processos de negócio da organização sejam identificados, analisados, validados e representados pela modelagem de processos de negócio na fase de desenho. A notação utilizada para modelar o fluxo de processo deve permitir o entendimento do funcionamento do processo entre as diversas partes interessadas da organização.

Após o processo modelo estar validado passa-se à fase de implementação, na qual o processo pode ser implementado através de um conjunto de políticas, regras e procedimentos. Em seguida, os processos instanciados devem ser monitorados e controlados, permitindo que métricas sejam coletadas e analisadas para possíveis refinamentos na fase subsequente, garantindo assim a melhoria contínua do processo.

Sistemas de informação automatizados podem ser utilizados para desenhar os processos, implementá-los de forma automatizada, controlar as execuções dos casos do processo, gerindo as atividades e seu sequenciamento e regras do negócio, e ainda, para monitorar e medir os processos executados.

2.3 Modelagens de Processos de Negócio

A modelagem de processos de negócio serve para abstrair o funcionamento de processos no mundo real, identificando características intrínsecas ao processo, como recursos, responsáveis, regras e outras. De acordo com IIBA (2015a), um modelo de processos de negócios descreve o fluxo sequencial de trabalho executado em uma ou mais entidades através de atividades e tarefas.

De maneira geral, a modelagem de processos compreende duas grandes atividades: modelagem do estado atual do processo (*As-Is*) e modelagem do estado futuro desejado para o processo (*To-Be*). Ao mapear-se o processo (*As-Is*) é possível identificar problemas, gargalos, falhas e oportunidades de melhorias que conduzirão o modelo futuro desejado (*To-Be*) BRAZIL (2013).

Dentre as principais perspectivas utilizadas na modelagem de processos de negócio pode-se citar Junior; Federson (2017):

- Funcional – representa os elementos do processo.
- Comportamental – representa sequenciamento dos elementos do processo e aspectos como iterações, decisões e outros.
- Organizacional – reflete os perfis e responsabilidades para a realização do processo.
- Informacional – simboliza as informações produzidas ou manipuladas ao longo do processo.

Algumas das notações para modelagem de processos comumente utilizadas no mercado são: *Business Process Management Notation* (BPMN), Rede de Petri, Fluxograma, *Event-driven Process Chain* (EPC), *Unified Modeling Language* (UML), *Integrated Definition Language* (IDEF) e *Value Stream Mapping* Junior; Federson (2017) e Lopes (2017).

A modelagem tradicional de processos é uma atividade nada trivial: ela exige conhecimento da notação utilizada e longas discussões com usuários chaves, especialistas do negócio e gestores envolvidos. Isso torna esta abordagem onerosa, exigindo muito tempo e recursos e sendo muitas vezes inviável economicamente. Alternativamente a esta abordagem manual de modelagem, a mineração de processos tem sido adotada para extrair o modelo de processo, utilizando *logs* de casos de processo gerados por sistemas informatizados Van der Aalst et al. (2004).

Dessa forma, é possível obter modelos *As-Is* que explicitam a situação real dos processos executados na organização. A mineração de processos é estudada em maiores detalhes na seção 2.4 a seguir.

2.4 Mineração de Processos

A mineração de processos pode ser utilizada para extrair conhecimento a partir de dados e registros de eventos geralmente disponíveis em sistemas utilizados nas

organizações. Através dos registros de eventos (*logs*) dos sistemas, que contém informações como processos executados, atividades desempenhadas, responsáveis por cada atividade, momento da execução da atividade e eventos que provocaram determinada atividade, é possível construir o fluxo de trabalho representando os processos reais da organização.

A mineração de processos pode ser útil em pelo menos três contextos. Primeiro na descoberta de como as pessoas e/ou os procedimentos realmente funcionam. Em segundo, os processos minerados podem ser utilizados em análises Delta, por exemplo, comparando um processo atual com algum processo predefinido, modelo de referência, norma/lei, etc. Por fim, mas não menos importante, a mineração de processos busca melhorar os processos, uma vez que pode identificar e eliminar gargalos, prever problemas, descobrir fraudes, desvios e violações de regras, normas e leis, recomendar ajustes e transformar processos Van der Aalst (2011a, 2016 e Van der Aalst et al. (2004, 2012 e Maruster et al. (2002).

Esta seção visa introduzir os principais conceitos da mineração de processos para subsidiar a compreensão teórica e de suas aplicações práticas na área de suprimentos e no estudo de caso.

2.4.1 Visão Geral

Na descoberta e modelagem de processos tradicional, os processos são identificados através de técnicas manuais como: entrevistas, análise de documentos da organização, mapas mentais, listas de partes interessadas, brainstorming, observação de trabalho e outras BRAZIL (2013 e IIBA (2015b).

Estas técnicas manuais parecem ser simples, mas diversos problemas inerentes às áreas de comunicação e engenharia de requisitos e processos tornam a descoberta e modelagem uma tarefa difícil e onerosa, a citar:

- Escopo mal definido que pode introduzir informações desnecessárias gerando confusão, ou falta de informações necessárias.

- Problemas na comunicação entre as partes, gerando entendimento discrepante.
- Omissão de informações.
- Falta de compreensão do processo completo.
- Foco no comportamento do processo, deixando de fora, casos menos significativos como fluxos de exceções.
- Visão tendenciosa do usuário chave sobre o processo.
- Fatores inerentes ao ser humano, como medos, falta de comprometimento e outros.

Neste sentido, a mineração de processos propõe a utilização dos eventos (*logs*) registrados em sistemas informatizados para, alternativamente às técnicas manuais de descoberta e modelagem de processos, extrair conhecimento e construir modelos de processos a partir dos dados reais registrados no sistema Van der Aalst (2016).

A mineração de processos (*Process Mining*) se posiciona na intersecção das áreas Inteligência Computacional (*Computational Intelligence*), Mineração de Dados (*Data Mining*) e Modelagem e Análise de Processos. Seu propósito é identificar, comparar e melhorar os processos de negócio, extraindo conhecimento presente em *logs* (registros de eventos ocorridos) dos diversos sistemas utilizados numa organização Van der Aalst et al. (2012); Van der Aalst (2012); De Leoni et al. (2016); Kalenkova et al. (2015); Van der Aalst et al. (2012); Van der Aalst et al. (2011).

O trabalho de Agrawal et al. (1998) foi pioneiro na mineração de processos, permitindo a construção de gráficos de fluxo de processos através de registros de execução de uma aplicação de fluxo de trabalho. Além disso, eles procuraram lidar com ruídos nos *logs* através de seu algoritmo baseado em grafo direcionado.

Na literatura é comum o posicionamento dos três tipos principais de aplicação da mineração de processos entre os autores do Manifesto de Mineração de Processo, que pode ser resumido na **Figura 3** Van der Aalst et al. (2012).

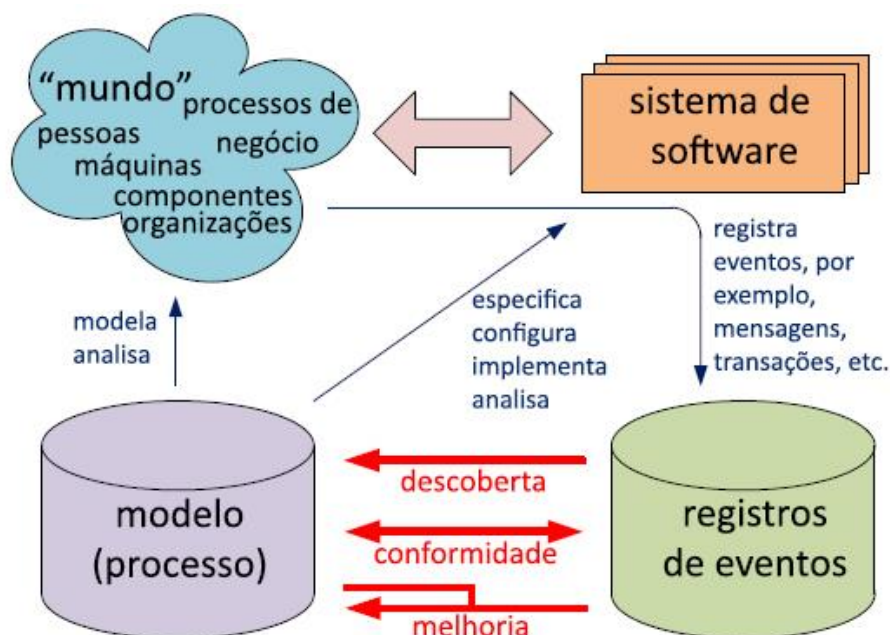


FIGURA 3 MINERAÇÃO DE PROCESSOS: (A) DESCOBERTA, (B) VERIFICAÇÃO DE CONFORMIDADE E (C) EXTENSÃO/MELHORIA. FONTE: Van der Aalst et al. (2012)

Conforme ilustrado na **Figura 3**, a descoberta se concentra na identificação de processos, ou seja, através da entrada de eventos contidos nos registros de eventos e processamento de algoritmos de descoberta de processos, constrói-se o modelo do processo. A verificação de conformidade é usada para checar se os eventos descobertos no *log* coincidem com algum modelo de processo previamente definido. Aqui há uma relação entre a verificação de conformidade e detecção de fraudes. A extensão preocupa-se com a melhoria ou reparo do processo existente.

Na literatura pode-se encontrar diversos algoritmos capazes de identificar processos pela mineração de processos. Destes, o algoritmo Alfa foi um dos primeiros capazes de gerar um modelo de processo, contudo possui problemas para lidar com a construção de rotas complicadas e ruídos. Neste contexto, técnicas mais robustas foram desenvolvidas, por exemplo, Mineração Heurística e Mineração Fuzzy, que podem lidar com *logs* de eventos incompletos, desbalanceados e que possuem ruídos Van der Aalst et al. (2011 e Bezerra; Wainer (2011 e Jans et al. (2011).

A mineração de processos se inicia nos registros de eventos, também referenciados como 'trilhas de auditoria', '*logs* de transações' ou 'história'. Estes registros geralmente

possuem informações sobre quais atividades foram executadas, por quem, em qual data e hora, e identificação do caso do processo Jans et al. (2011).

A **Tabela 1** apresenta um exemplo de registros de eventos com informações sobre dois casos em um processo de aquisições. As linhas representam as várias entradas no registro de eventos, enquanto as colunas representam as diferentes propriedades (atributos) do processo. O caso precisa de um identificador único e terá um conjunto específico de eventos executados. Cada evento individual pode ter vários atributos; mostrados aqui são nome da atividade, data e hora, responsável ou executor, tipo de material e valor.

TABELA 1 - EXEMPLO DE REGISTROS DE EVENTOS (LOGS)

Caso	Atividade	Data e Hora	Responsável	Material	Valor
101060	Criar Solicitação de Compra (A)	08/10/2008 00:00	Pedro	Eletrônico	73000
101060	Criar Pedido de Compra (B)	08/10/2008 13:04	Maria	Eletrônico	73000
101060	Enviar Pedido de Compra (C)	08/10/2008 13:04	Maria	Eletrônico	73000
101060	Receber Confirmação do Pedido (D)	09/10/2008 09:11	Maria	Eletrônico	73000
101060	Registrar Entrada de Mercadorias (E)	20/11/2008 14:03	José	Eletrônico	73000
101064	Criar Pedido de Compra (B)	08/10/2008 14:29	Pedro	Papelaria	21500
101064	Enviar Pedido de Compra (C)	08/10/2008 14:29	Maria	Papelaria	21500
101064	Receber Confirmação do Pedido (D)	08/10/2008 14:52	Maria	Papelaria	21500
101064	Fornecedor fatura (F)	03/11/2008 00:00	André	Papelaria	21500
101064	Registrar recibo de fatura (G)	03/11/2008 17:31	Ana	Papelaria	21500
101064	Registrar Entrada de Mercadorias (E)	04/11/2008 11:47	José	Papelaria	21500

Note que no caso “101060” as atividades “Criar Solicitação de Compra” -> “Criar Pedido de Compra” -> “Enviar Pedido de Compra” -> “Receber Confirmação do Pedido” -> “Registrar Entrada de Mercadorias”, já o caso “101064” executou as atividades “Criar Pedido de Compra” -> “Enviar Pedido de Compra” -> “Receber Confirmação do Pedido” -> “Fornecedor Fatura” -> “Registrar recibo de fatura” -> “Registrar Entrada de Mercadorias”. O fluxo ilustrado na **Figura 4** representa o processo da **Tabela 1**, de acordo com os 2 casos extraídos do registro de eventos.

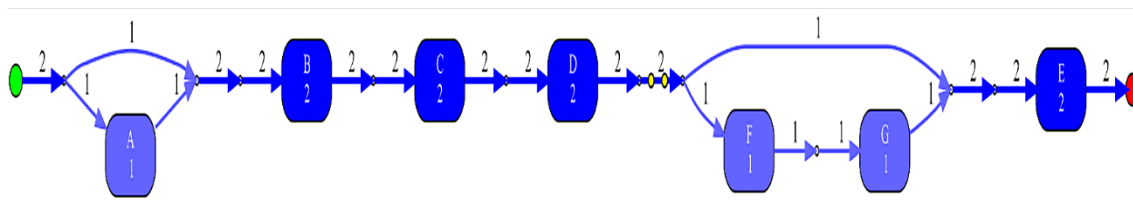


FIGURA 4 - FLUXO DO PROCESSO MINERADO NO PROM COM *INDUCTIVE VISUAL MINER*

Definição 1.0 (Caminho, Registro de evento). Seja $A \in U_A$ um conjunto de atividades em algum universo de atividades U_A . Um caminho $\sigma \in A^*$ é uma sequência de atividades. Um registro de eventos (*log*) é um multiconjunto de caminhos, ou seja, $L \in B(A^*)$ Van der Aalst et al. (2004 e Munoz-Gama (2014). Desta forma, os registros de eventos da **Tabela 1**, são representados como [(A, B, C, D, E), (B, C, D, F, G, E)], contendo informações sobre 11 eventos e 2 casos, onde cada caso seguiu um diferente caminho.

2.4.2 Perspectivas para Mineração de Processos

Os *logs* possuem três perspectivas principais de análise. Uma perspectiva de processo que busca responder como o caso ocorreu, ou seja, qual a ordem das atividades; outra perspectiva organizacional que foca nos atores, identificando quem executou cada atividade e provendo informações dos relacionamentos entre executores e atividades, e por fim, a perspectiva do caso, que atua sobre outras informações pertinentes àquele caso executada Van der Aalst et al. (2007).

A perspectiva de processo foca no fluxo do processo, ou seja, no sequenciamento das atividades. O objetivo da mineração, nesta perspectiva, é descobrir uma caracterização de todos os caminhos possíveis no registro de eventos do processo. Esta perspectiva pode ser estendida incluindo-se outras informações, como duração e frequência dos eventos (*timestamp*), o que permite identificar gargalos e atrasos Lopes (2017).

A perspectiva organizacional se concentra em explicitar os responsáveis pelas atividades de cada caso do processo, estruturando e classificando as pessoas em

termos de funções e unidades organizacionais, ou ainda, mostrando as relações entre os executores das atividades.

A perspectiva de caso foca nas propriedades de cada caso. Os casos podem ser caracterizados pelo fluxo percorrido no próprio caso em análise, ou ainda por informações presentes no mesmo. Por exemplo, nos casos presentes na **Tabela 1**, pode ser interessante saber o material ou tipo de material adquirido, o fornecedor ou valores das aquisições.

2.4.3 Algoritmos utilizados na Mineração de Processos

Inúmeros estudos abordam o desenvolvimento de técnicas e algoritmos de mineração de processos para modelar processos de negócios. O algoritmo Alfa foi um dos primeiros algoritmos capazes de descobrir processos através de registros de eventos, no entanto, possui deficiências para tratar ruídos, comportamentos incompletos e caminhos complexos Van der Aalst et al. (2004 e ER et al. (2018 e De Medeiros et al. (2005). Em Van der Aalst; De Medeiros; et al. (2005 e De Medeiros et al. (2005, 2007) implementaram o algoritmo Genético em torno do processo de seleção natural de Darwin. Outros algoritmos comumente utilizados incluem Minerador *Fuzzy* Van der Aalst (2007), Minerador Heurístico Weijters et al. (2006) e Mineração Indutiva (*Inductive Miner*) Van der Aalst (2016 e Leemans et al. (2013a, 2013b).

A seguir nesta subseção destaca-se algoritmos comumente utilizados para minerar processos de negócio.

2.4.3.1 Algoritmo Alfa (α algorithm)

Com base nos registros de eventos, o algoritmo Alfa é capaz de inferir a relação entre as atividades, considerando que duas atividades estão conectadas desde que sua causalidade possa ser identificada. Por exemplo, se uma atividade A é seguida por B,

mas B nunca é seguida por A, então pode-se inferir que há uma dependência causal entre A e B.

Para expressar o modelo de processo minerado, o algoritmo Alfa utiliza *Workflow Net* (WF-Net), um tipo especial de redes de Petri Van der Aalst et al. (2004). Uma rede de Petri, do tipo rede de Lugares/Transições, ou simplesmente P/T-rede, é uma tupla (P, T, F) , onde:

1. P é um conjunto finito de posições.
2. T é um conjunto finito de transições tal que $P \cap T = \emptyset$.
3. $F \subseteq (P \times T) \cup (T \times P)$, é um conjunto de arcos, chamado de relação de fluxo.

A **Figura 5** ilustra uma rede P/T composta por oito posições e sete transições. Transição A tem uma entrada e uma saída, a transição *And-split* tem uma posição de entrada e duas posições de saída, e a transição *And-join* tem duas posições de entrada e uma posição de saída. Os pontos de entrada de A e saída de D, representam o início e fim do processo, respectivamente.

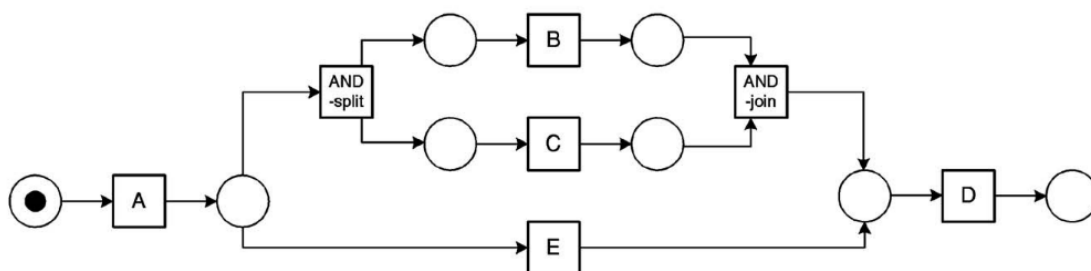


FIGURA 5 – EXEMPLO DE REDE P/T. FONTE: Van der Aalst et al. (2004)

Os passos do algoritmo definidos em Van der Aalst et al. (2004) e sumarizados em Lopes (2017) são descritos a seguir:

1. Cria-se o conjunto T_w de todas as atividades presentes no registro de eventos W .

Definição 1: $T_w = \{t \in T \mid \exists_{\sigma \in W} t \in \sigma\}$.

2. Cria-se o conjunto T_i de todas as atividades que instanciam um processo σ presentes no conjunto de registros de eventos W .

Definição 2: $T_i = \{t \in T \mid \exists \sigma \in W^t = \text{primeiro}(\sigma)\}$.

3. Cria-se o conjunto T_o de todas as atividades que finalizam um caso de processo σ presente no conjunto de registros de eventos W .

Definição 3: $T_o = \{t \in T \mid \exists \sigma \in W^t = \text{último}(\sigma)\}$.

4. Cria-se um conjunto de tuplas (A, B) , buscando encontrar o conjunto $p(a, b)$ que representa as posições da rede, sendo A as transições presentes em T_w que precedem $p(a, b)$, e B as transições presentes em T_w que sucedem $p(a, b)$, levando em consideração que $a \in A$ e $b \in B$, e que a relação de dependência $a \rightarrow_w b$ deve ser verdadeira. Em todas as combinações possíveis entre os elementos presentes em A , representados por a_i e a_j , a relação $a_i \#_w a_j$ é verdadeira, e para todas as possíveis combinações entre os elementos presentes em B , representados por b_i e b_j , a relação $b_i \#_w b_j$ é verdadeira.

Definição 4:

$$X_W = \{(A, B) \mid A \subseteq T_w \wedge B \subseteq T_w \wedge \forall a \in A \forall b \in B a \rightarrow_w b \wedge \forall a_1, a_2 \in A a_1 \#_w a_2 \wedge \forall b_1, b_2 \in B b_1 \#_w b_2\}.$$

5. Elimina-se do conjunto de tuplas (A, B) todas as tuplas (A', B') que representam o subconjunto de elementos de tuplas contidas em (A, B) , ficando apenas as tuplas que contemplam maior quantidade de elementos.

Definição 5: $Y_W = \{(A, B) \in X_W \mid \forall (A', B') \in X_W A \subseteq A' \wedge B \subseteq B' \implies (A, B) = (A', B')\}$.

6. Constrói-se o conjunto de posições da rede P_w , a partir das posições restantes $p(A, B)$ adicionadas às posições $p(\text{início}, T_i)$ e $p(T_o, \text{fim})$, que definem as posições de início e fim, respectivamente.

Definição 6: $P_w = \{p(A, B) \mid (A, B) \in Y_W\} \cup \{i_w, o_w\}$.

7. Constrói-se o conjunto de transições da rede F_w , presentes nos conjuntos A , B , T_i e T_o .

Definição 7:

$$F_w = \{ (a, p_{(A,B)} \mid (A,B) \in Y_w \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_w \wedge b \in B \} \cup \{ i_w, t \mid t \in T_1 \} \cup \{ t, o_w \mid t \in T_o \}.$$

8. O resultado do algoritmo Alfa, dado o conjunto de registros de eventos W , é a rede de Petri definida por (P_w, T_w, F_w) .

2.4.3.2 Mineração Heurística (*Heuristics Miner*)

Mineração Heurística (*Heuristics Miner*) é um algoritmo prático de mineração aplicável que pode lidar com ruído e ser usado para expressar o comportamento total contido nos registros de eventos, ou apenas no comportamento principal (ou seja, nem todos os detalhes e exceções). Para isso, o algoritmo leva em consideração a frequência das relações entre cada par de atividades, para determinar a probabilidade de duas atividades possuírem uma dependência causal, construindo, assim, um grafo de dependência. Apenas as informações de identificador do caso, data e hora e atividade são levados em consideração. Estas informações de data e hora das atividades são utilizadas para determinar a ordenação das atividades Weijters et al. (2006).

As seguintes notações são definidas em Weijters et al. (2006) para analisar as relações entre as atividades. Seja T um conjunto de atividades, i é o identificador da atividade e ti é o número da atividade i no registro de eventos. $\sigma \in T^*$, é um caminho de evento, ou seja, uma sequência arbitrária de identificadores de atividades. $W \subseteq T^*$, é um registro de eventos, por exemplo, um conjunto de rastreamento de eventos. Seja $a, b \in T$:

1. $a > wb$ se existir um caminho $\sigma = t_1 t_2 t_3 \dots t_n$, e $i \in \{1, \dots, n-1\}$, $\sigma \in W$, e $t_i = a$, e $t_{i+1} = b$. Descreve qual atividade aparece na sequência de outra atividade.

2. $a \rightarrow b$, se $a > wb$ e $b \not> wa$. Calcula a relação de dependência direta.
3. $a \# wb$ se $a \not> wb$ e $b > wa$. Indica transições que nunca se seguem diretamente, isto é, não há relação de dependência direta e paralelismo é improvável.
4. $a || b$ se $a > wb$ e $b > wa$. Captura o paralelismo, ou seja, quando duas atividades seguem uma a outra em qualquer ordem.
5. $a \gg wb$ se existe um rastro $\sigma = t_1 t_2 t_3 \dots t_n$ e $i \in \{1, \dots, n-2\}$, tal que $\sigma \in W$ e $t_i = a$ e $t_{i+1} = b$ e $t_{i+2} = a$. Descreve atividades que são executadas várias vezes, neste caso loop de comprimento um.
6. $a \ggg wb$ se existe um rastro $\sigma = t_1 t_2 t_3 \dots t_n$ e $i < j$ e $i, j \in \{1, \dots, n\}$ tal que $\sigma \in W$ e $t_i = a$ e $t_j = b$. Descreve atividades que são executadas várias vezes, neste caso loop de comprimento dois.

Os passos do algoritmo definidos em Weijters et al. (2006) são apresentados a seguir:

1. Minerar o gráfico de dependência: para construir o grafo de dependência, leva-se em consideração a frequência das relações entre cada par de atividades. A probabilidade de existir uma dependência causal entre as atividades a e b , em um $\log W$, é dada por $a \Rightarrow wb$, e definida pela equação: $a \Rightarrow wb = \frac{|a > wb| - |b > wa|}{|a > wb| + |b > wa| + 1}$. O valor de $a \Rightarrow wb$ está sempre entre -1 e 1. Valores próximos a +1 indicam uma forte dependência positiva entre a e b , ou seja, é muito provável que a seja causa de b , mas a dificilmente segue b . Do contrário, valores próximos a -1 significam forte dependência negativa.
2. Identificação de atividades não observáveis e *gateways AND* e *XOR*: Dado um grafo de dependência no qual a atividade A tem relação de dependência com B e C . Se o padrão de conexão for de um *gateway AND*, o padrão BC deverá aparecer no *log*. Por outro lado, se o *gateway* for do tipo *XOR*, o padrão BC não deve aparecer no *log*. Desta forma, define-se: caso $a \Rightarrow b \wedge c$, seja alto,

então as atividades estão conectadas por um *gateway AND*, do contrário estarão conectadas por um *gateway XOR*. A seguinte fórmula é definida para expressar esta ideia: Seja W um registro de evento de T , $a, b, c \in T$, e b e c são dependentes em relação a a , então:

$$a \Rightarrow w b^{\wedge} c = \left(\frac{|b>wc|+|c>wb|}{|a>wb|+|a>wc|+1} \right).$$

3. Minerando dependências em longas distâncias: em alguns modelos de processos a escolha entre duas atividades não pode ser determinada por algum nó do processo, mas pode depender de escolhas realizadas em outras partes do processo. A **Figura 6**, por exemplo, ilustra um caso em que após a execução da atividade D, há uma escolha entre as atividades E e F, porém esta escolha é determinada pela escolha anterior entre B e C. A heurística de dependência de longas distâncias busca lidar com este problema, levando em consideração a relação $a \ggg wb$, que indica a existência da sequência $a...b$, com qualquer número de atividades entre a e b .

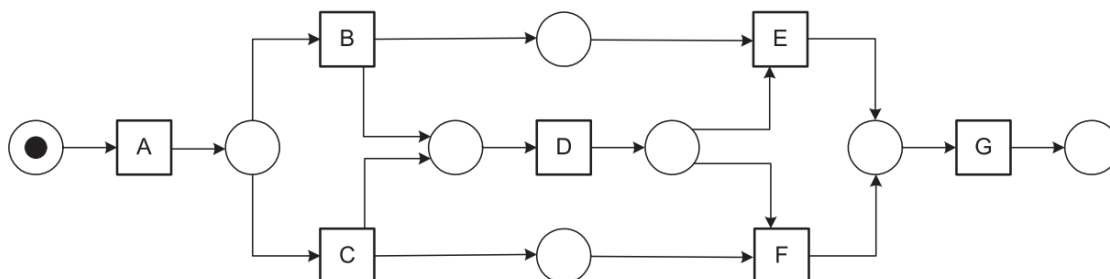


FIGURA 6 - MODELO DE PROCESSO COM ESCOLHA NÃO LIVRE. FONTE: Weijters et al. (2006)

2.4.3.3 Mineração Fuzzy (*Fuzzy Miner*)

A mineração *fuzzy* (*Fuzzy Miner*) é baseada em uma abordagem que utiliza métodos da cartografia temática para fornecer uma visão dinâmica do processo. Na cartografia temática, o conceito principal é a generalização, para abstrair a partir de certos detalhes,

que podem mudar ao longo do tempo. Atividades do processo podem ser relacionadas com localizações numa topologia, por exemplo cidades, e relações de precedência com ligações de tráfego entre elas, por exemplo, vias férreas ou auto estradas Van der Aalst (2007).

No minerador *fuzzy*, duas métricas são identificadas: significância e correlação. A primeira métrica calcula a frequência das ocorrências de eventos e sua ordem: quanto mais frequente é observada uma relação de precedência, mais significativa ela é. A segunda métrica, correlação, mede a proximidade dos eventos, o quão relacionados dois eventos estão seguindo um ao outro. Isso pode ser medido pelos dados que eles compartilham ou pela similaridade de nomes de atividades Van der Aalst (2007).

Para gerar um modelo de processo, o minerador simplifica o processo com base na significância e correlação dos eventos, conforme:

- Comportamento altamente significativo é preservado, ou seja, contidos no modelo simplificado.
- Comportamento altamente correlacionado, mas menos significativo, os eventos são agregados em agrupamentos (*clusters*) e escondidos nos agrupamentos com o modelo simplificado.
- Comportamento pouco correlacionado e menos significante é abstraído, ou seja, removido do modelo simplificado.

Com base nessas medidas, as visualizações de modelos baseados nos registros de eventos podem ser geradas dinamicamente. Isso permite que o analista amplie e reduza determinados aspectos do modelo de processo. Assim, é possível filtrar um certo percentual, por exemplo, 90% dos traços para fazer um modelo. A suposição aqui é que o minerador *Fuzzy* descartaria 10% dos traços que continham ruído, ou comportamento pouco frequente, mas correto, do processo subjacente.

Nesta abordagem pode-se reduzir e focar o comportamento exibido, utilizando os conceitos de agregação e abstração. No modelo simplificado, pode-se dar ênfase, destacando o comportamento mais significativo presente nos registros de eventos.

2.4.3.4 Mineração Genética (*Genetic Miner*)

A mineração genética é um algoritmo que avalia as várias soluções candidatas perante uma função de cobertura (*fitness*), determinando se cada solução está consistente com os registros de eventos. Cada uma das soluções é representada por uma matriz causal, na qual especifica-se o mapa de entrada e saída de dependência de cada atividade. As soluções candidatas são geradas pelo cruzamento genético (seleção e mutação), gerando assim um espaço de busca que contém um conjunto com todas as possíveis combinações de atividades que estão presentes nos registros de eventos Van der Aalst; De Medeiros; et al. (2005 e De Medeiros et al. (2005, 2007).

De acordo com Van der Aalst (2016), as abordagens evolutivas utilizam um procedimento iterativo para imitar o processo de evolução natural. Estas abordagens não são determinísticas e dependem da randomização para encontrar novas alternativas. A **Figura 7**, ilustra uma visão geral desta abordagem, composta por quatro etapas principais, que estão presentes em qualquer algoritmo genético: (a) inicialização, (b) seleção, (c) reprodução e (d) término.

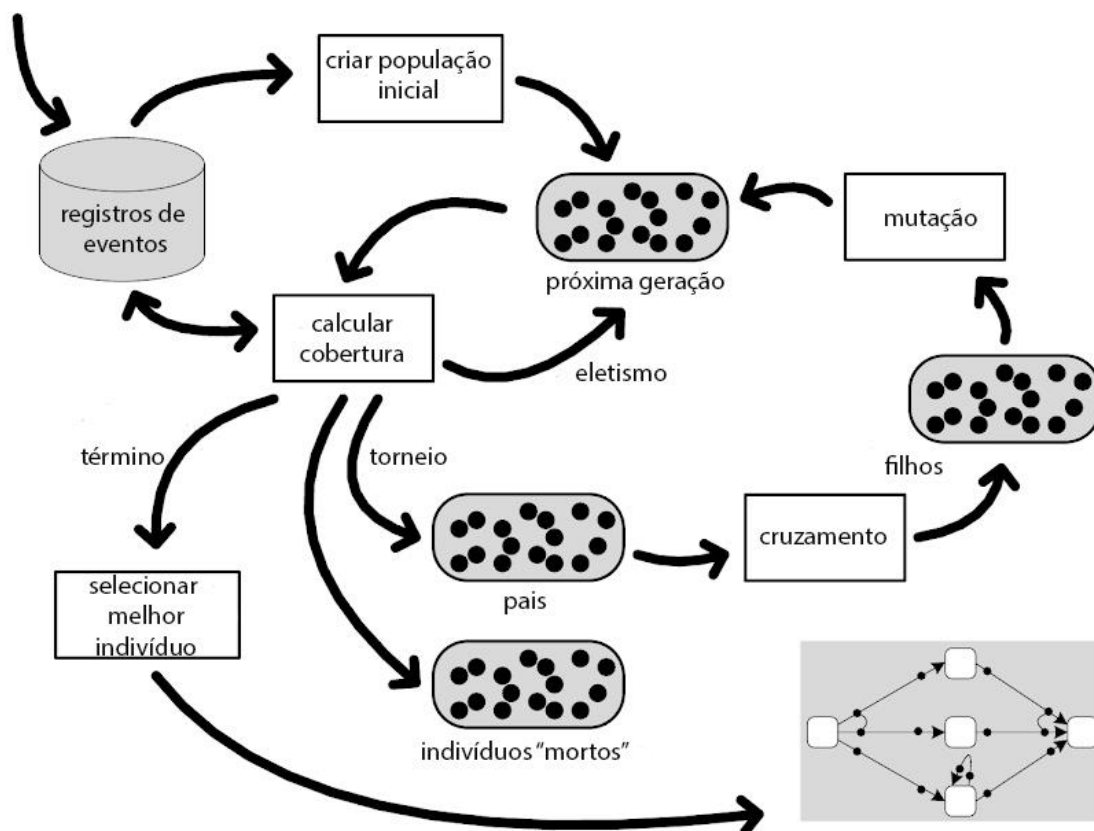


FIGURA 7 - VISÃO GERAL DA ABORDAGEM DE MINERAÇÃO GENÉTICA DE PROCESSOS.
 FONTE: Van der Aalst (2016)

O algoritmo genético de mineração de processos utiliza elitismo, cruzamento e mutação para construir os elementos da população da próxima geração genética. O elitismo seleciona um percentual da população atual mais apto e leva para a próxima geração. O cruzamento gera novos indivíduos com base na população atual de indivíduos mais aptos (pais), recombinando o material genético dos indivíduos mais aptos na esperança que desta combinação dos pais gere um novo elemento ainda mais apto. Por sua vez, a mutação altera algum detalhe menor (gene) de um elemento da população, esperando inserir um novo material útil na população Van der Aalst; De Medeiros; et al. (2005).

Conforme definido em Van der Aalst; De Medeiros; et al. (2005 e De Medeiros et al. (2005), o algoritmo genético para diante das seguintes situações:

1. É encontrado um indivíduo cuja aptidão ou cobertura (*fitness*) é 1.
2. Calcula n gerações, sendo que n é o número máximo de gerações permitidas.

3. O cálculo de indivíduo mais apto não mudou por $n/2$ gerações seguidas.

Caso nenhuma das condições a cima seja válida, o algoritmo genético criará então uma população nova, conforme:

1. Copia-se “taxa de elitismo x tamanho da população” dos melhores indivíduos da população atual para a próxima geração.
2. Enquanto houver indivíduos para serem criados, faça:
 - a. Utilize a seleção de torneio para selecionar pai1.
 - b. Utilize a seleção de torneio para selecionar pai2.
 - c. Selecione um número aleatório r entre 0 (inclusivo) e 1 (exclusivo).
 - d. Se r é menor que a taxa de cruzamento: faça cruzamento entre pai1 e pai2, que gerará filho1 e filho2.
 - e. Se não, filho1 é igual a pai1 e filho2 é igual a pai2.
 - f. Elimina filhos1 e filhos2 (somente necessário quando a taxa de mutação for diferente de zero).
 - g. Copie os filhos1 e filhos2 para a nova população.
3. Retorne a nova população.

Em Van der Aalst (2016) ressalta-se alguns pontos sobre a mineração genética, a citar:

1. A técnica é flexível e robusta, sendo capaz de lidar com ruídos e registros de eventos incompletos.
2. A abordagem pode ser adaptada e ampliada, alterando a função de adequação da população, é possível dar preferência a construções particulares.
3. Para modelos e registros de eventos grandes não é muito eficiente, podendo levar muito tempo para se encontrar um modelo com aptidão aceitável.
4. É interessante combinar heurísticas com a mineração genética, possibilitando que a mineração genética melhore um modelo de processo obtido pela mineração heurística. Tal abordagem pode economizar tempo computacional

e gerar modelos que nunca seriam descobertos pelos algoritmos convencionais que procuram apenas por dependências locais.

2.4.3.5 Mineração Indutiva (*Inductive Miner*)

A mineração indutiva é altamente extensível e permite muitas variantes da abordagem básica, possuindo técnicas capazes de lidar com comportamentos pouco frequente e grandes modelos de registros de eventos. Esta técnica indutiva é atualmente uma das principais abordagens de descoberta de processos, dado a sua flexibilidade, garantias formais e escalabilidade Van der Aalst (2016).

Dada uma relação direta de sequenciamento entre duas atividades $G(L)$, expressada no grafo de um registro de eventos L . As atividades presentes em L , representam os nós do grafo de sequência. Uma aresta (a, b) está presente em $G(L)$, somente se existir em L um caminho $\langle \dots, a, b, \dots \rangle$. Um nó de $G(L)$ é um ponto de início se a atividade está em $Inicio(L)$, definindo-se $Inicio(G(L)) = Inicio(L)$. Da mesma forma, um nó final está em $Fim(L)$, e $Fim(G(L))$ Leemans et al. (2013b).

A ideia do algoritmo de mineração indutiva é encontrar em $G(L)$ estruturas que indiquem o operador “dominante” que ordena o comportamento do processo. Para tal, aplica-se uma abordagem de dividir para conquistar, na qual particiona-se as atividades, seleciona-se a construção de processo mais importante, divide-se o registro de eventos e recorre-se até que um caso base seja encontrado Leemans et al. (2013a). As possíveis divisões ou cortes aplicadas pela mineração indutiva são apresentadas na

Figura 8.

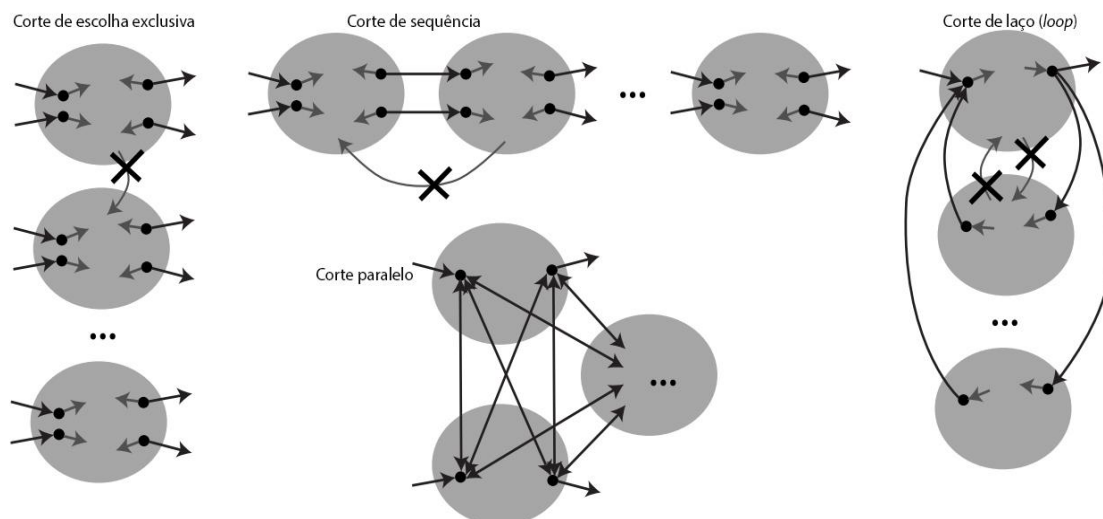


FIGURA 8 - TIPOS DE CORTES NA MINERAÇÃO INDUTIVA. FONTE: Leemans et al. (2013b)

Os passos do algoritmo de mineração indutiva são sumarizados em Van der Aalst (2016) e a sequência de cortes funciona conforme:

1. Havendo um corte de escolha exclusiva não trivial, então aplica-se um corte de escolha exclusiva máximo, dividindo o registro de eventos em registros de eventos menores.
2. Não havendo o corte acima, mas se houver um corte de sequência não trivial, aplica-se um corte de sequência máximo, dividindo o registro de eventos em registros menores.
3. Não havendo os cortes acima, mas se houver um corte paralelo não trivial, aplica-se um corte paralelo máximo, dividindo o registro de eventos em registros menores.
4. Não havendo os cortes acima, mas se houver um corte de laço, aplica-se um corte de laço máximo, dividindo o registro de eventos em registros menores.
5. Após dividir o registro de eventos em sub registros de eventos menores, repete-se o procedimento até que um caso base (sub registro de evento com apenas uma atividade) seja encontrado.

É importante notar que o algoritmo de mineração indutiva sempre produz um modelo de processo sólido, ou seja, livre de impasses (*deadlocks*) e outras anomalias. A métrica

de cobertura (*fitness*) é garantida e devido os modelos serem estruturados em blocos e não haver atividades duplicadas, os modelos tendem a ser simples e com uma boa generalização. Por outro lado, a divisão em cascata pode criar modelos muito generalizados (*underfitting*) Van der Aalst (2016).

2.4.3.6 Algoritmos para verificar conformidade

Em Rozinat; Van der Aalst (2008) o algoritmo *Conformance Checker* é proposto, visando detectar inconsistências entre o modelo de processo minerado e um registro de eventos e quantificar as métricas de cobertura (*fitness*) e precisão. Através da métrica de cobertura, a verificação de conformidade analisa qual extensão dos registros de eventos pode ser associada aos caminhos especificados no modelo do processo minerado. Já pela precisão, a verificação de conformidade analisa o grau de certeza em que o modelo de processo minerado descreve o comportamento observado nos registros de eventos.

O *Conformance Checker* reproduz um registro de eventos dentro de um modelo de rede de Petri de forma não bloqueante, ou seja, cobrindo todos os possíveis caminhos (sem filtros). Durante a reprodução do registro é realizada coleta de informações de diagnóstico que podem ser acessadas posteriormente e o algoritmo se encarrega de tratar de tarefas ocultas que podem permitir que a transição seja reproduzida em seguida, e também é capaz de lidar com tarefas duplicadas Van der Aalst et al. (2010).

Em Van der Aalst; De Beer; et al. (2005) o algoritmo *LTL-Checker (Linear Temporal Logic)* é proposto, no qual é possível especificar e verificar uma variedade de instruções lógicas mantidas nos registros de eventos. O algoritmo leva em conta aspectos temporais e de ordenação das atividades e pode verificar especificações, como: (após a atividade X, a atividade Y deve ser executada), (após a atividade X, eventualmente a atividade Y deve ser executada), sempre (especificando invariantes) e até (até a atividade X, a atividade Y não deve ser executada).

Para detecção de fraudes, restrições de outros atributos também podem ser especificados, e não apenas restrições em atividades. Desta forma, por exemplo, o princípio de “quatro olhos” pode ser verificado. Este princípio estabelece que, embora autorizado a execução de duas atividades, uma pessoa não pode executar ambas as atividades para o mesmo caso do processo Van der Aalst; De Beer; et al. (2005 e Stoop (2012).

2.4.4 Ferramentas de Mineração de Processos

Diversas soluções de mineração de processos estão disponíveis, sejam elas acadêmicas, como o ProM e RapidProM, ou comerciais como o Disco e Celonis. No lado de soluções acadêmicas, o ProM foi selecionado por ser uma solução de código aberto, possuindo mais de 1500 extensões (*plug-ins*) em sua versão atual 6.8, sendo amplamente utilizado em pesquisas e aplicações práticas nesta área Van der Aalst (2016 e Tiwari et al. (2008). Pelo lado de soluções comerciais, o Celonis foi escolhido por ser atualmente a solução líder de mercado, de acordo com o *Gartner* em seu relatório “Guia de Mercado para Mineração de Processos” Kerremans (2018).

Adicionalmente será utilizada a ferramenta XESame Buijs (2010), que atualmente está incorporada no ProM. O objetivo desta ferramenta é mapear as informações nos registros de eventos e exportar no formato XES. Este formato foi desenvolvido pela “*IEEE Task Force on Process Mining*” para servir como um padrão de registros de dados de eventos e atualmente pode ser importado pela maioria das soluções de mineração de processos Buijs (2010).

2.4.5 Métricas de Análise na Mineração de Processos

É importante levar em consideração quesitos de qualidade sobre os modelos de processos minerados, possibilitando comparar o desempenho dos algoritmos utilizados na mineração, ou ainda avaliar os resultados obtidos.

A seguir, são introduzidas algumas métricas de avaliação sugeridas em Rozinat et al. (2007), sendo: cobertura, precisão, simplicidade e generalização.

2.4.5.1 Cobertura (*Fitness*)

A cobertura é medida pela fração de um *log* que pode ser repetido no modelo minerado. Um modelo de processo minerado que tenha um bom índice de cobertura representa o comportamento presente no registro de eventos. Um índice de 100% de cobertura significa que todos os caminhos são representados pelo modelo. O modelo (c) ilustrado na **Figura 9** representa um modelo com índice de cobertura ruim, pois não permite a execução de outras seqüências de atividades.

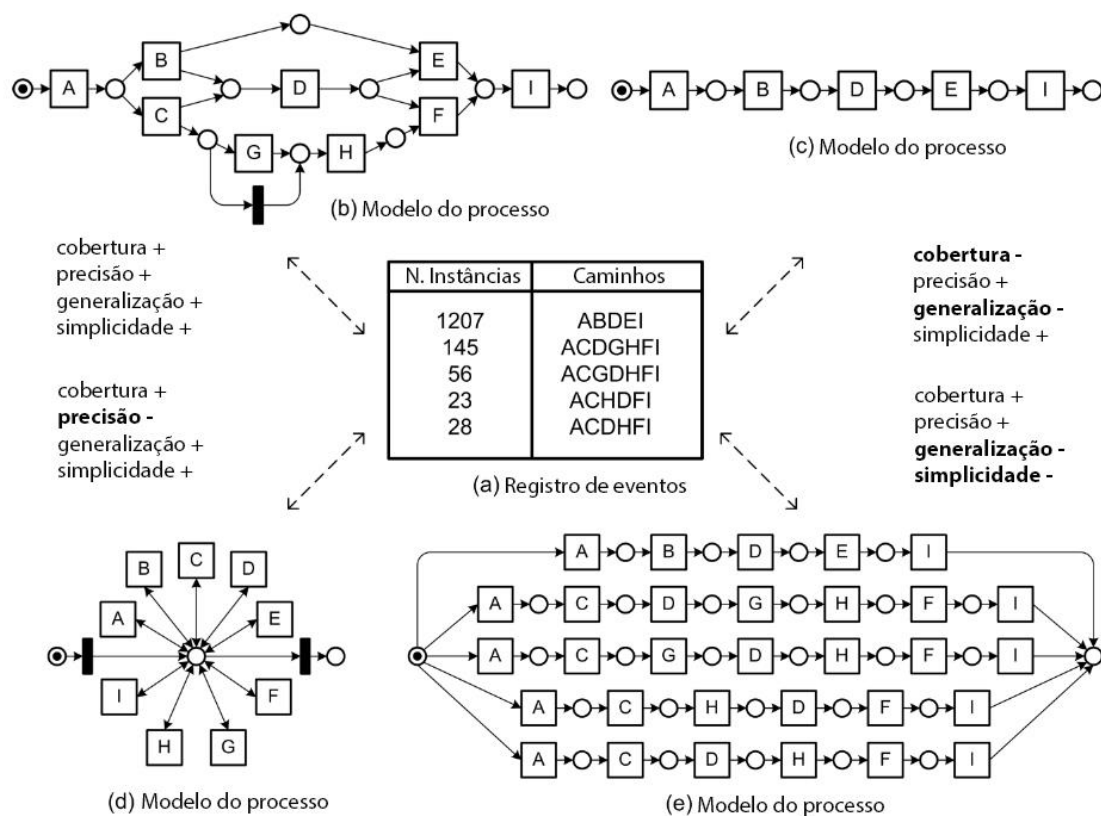


FIGURA 9 - AVALIAÇÃO DE MÉTRICAS DE QUALIDADE EM MODELO DE PROCESSO MINERADO.
FONTE: Rozinat et al. (2007)

Existem diferentes formas de pontuar o nível de cobertura do modelo minerado. Pode ser definido ao nível do caso, por exemplo, a fração de caminhos no registro de eventos que podem ser repetidos completamente. Também pode ser definido ao nível dos eventos, por exemplo, a fração de eventos presentes nos registros de eventos que são realmente possíveis de acordo com o modelo Rozinat et al. (2007).

De acordo com Van der Aalst (2016) é interessante buscar o modelo 80/20, no qual o modelo minerado representa 80% dos casos presentes no registro de eventos (cobertura), e por outro lado mantém a simplicidade ao desconsiderar 20% dos casos menos relevantes ou ruídos do registro de eventos.

2.4.5.2 Precisão (*Behavioral Appropriateness*)

A precisão mede o sub ajuste, ou seja, se o modelo generalizado permite comportamento muito diferente daquele visto no *log*. Um modelo é preciso caso não permita comportamentos muito diferentes daqueles presentes nos registros de eventos. Desta forma, o “modelo de flor” ilustrado na **Figura 9 (d)** não têm precisão, pois permite a execução de atividades de A até I em qualquer ordem.

2.4.5.3 Generalização (*Generalization*)

Por outro lado, a generalização está relacionada ao sobre ajuste. Um modelo sobre ajustado não generaliza suficientemente, ou seja, o modelo é muito específico e orientado demasiadamente a exemplos no *log* de eventos. Na **Figura 9**, o modelo (e) somente permite a execução das cinco sequências exatas do modelo presentes no registro de eventos.

2.4.5.4 Simplicidade (*Simplicity*)

A simplicidade indica que o modelo mais simples capaz de explicar o comportamento do processo é o melhor modelo. Em Van der Aalst (2016) cita-se o princípio “Navalha

de Occam” de William de Occam, que diz: “não se deve aumentar, além do necessário, o número de entidades necessárias para explicar qualquer coisa”.

Nesta medida, a estrutura sintática e semântica do processo de negócio é avaliada. Existem várias formas (sintáticas) de se expressar o mesmo modelo. Desta forma, a métrica de simplicidade avalia se o modelo de processo minerado é o mais simples possível, sem perder a cobertura e precisão. Na **Figura 9 (e)**, por exemplo, a cobertura e precisão do modelo são bons, porém a estrutura do modelo possui muitas tarefas duplicadas, tornando o modelo difícil de compreender.

2.4.5.5 Considerações

Modelos extremos como o “modelo de flor” (modelo d na **Figura 9**), no qual qualquer sequência é possível, e o “modelo de enumeração” (modelo e na **Figura 9**), no qual somente é possível as sequências do *log*, mostram a necessidade de duas dimensões adicionais, precisão e generalização.

Um modelo que não permite muitos comportamentos é muito preciso (*overfit*). Por outro lado, um modelo que é muito generalizado, como o “modelo de flor”, é considerado “*underfitting*”.

Para alcançar qualidade, o modelo minerado deve atingir um equilíbrio entre precisão e generalização (*overfit x underfitting*).

3 Metodologia

Neste capítulo será apresentado o método utilizado para atingir os objetivos desta pesquisa. Primeiramente uma revisão exploratória de literatura é realizada para obter informações sobre a Mineração de Processos, seus conceitos e principais aspectos que possam contribuir para processos da área de Gestão de Suprimentos, e quais benefícios podem ser alcançados com o uso da Mineração de Processos. Em seguida, com base nas revisões realizadas, será realizado um estudo de caso exploratório apresentando o método de mineração de processos aplicado na área de suprimentos de uma empresa brasileira que utiliza o ERP Protheus da Totvs, incluindo os passos envolvidos no planejamento, seleção, extração, organização dos registros de eventos e aplicação dos algoritmos de mineração de processos, assim como as análises conduzidas sobre os modelos, dificuldades encontradas, além dos algoritmos e ferramentas utilizadas.

3.1 Revisão Exploratória da literatura

Esta revisão possui dois objetivos principais. O primeiro é identificar na literatura casos onde a Mineração de Processos foi aplicada na área de Gestão de Suprimentos e gerar uma caracterização dos estudos de caso existentes, descrevendo aspectos importantes, como os métodos e técnicas aplicadas, principais ferramentas, tipo de dados e outros. E o outro objetivo é identificar os estudos que aplicaram a análise de conformidade, identificando desvios/fraudes.

Nenhum dos objetivos anteriores foi explorado em detalhes no passado, fazendo assim, com que esta revisão contribua para a compreensão do contexto geral da área e para promoção da aplicação futura das técnicas de mineração de processos na área de Gestão de Suprimentos. A revisão é apresentada no capítulo 4.

Para responder as perguntas definidas nos objetivos da seção **1.3**, as palavras chaves pesquisadas foram: “process mining”, “workflow mining”, “supply chain”, “purchase process”, “procurement process” e seus similares em língua portuguesa. As

pesquisas se concentraram nos mecanismos de buscas (Google Scholar, Science Direct, Portal de Periódicos da CAPES e site www.processmining.org/publications/start).

3.2 Estudo exploratório de caso único e método PM²

Com o objetivo de aplicar a mineração de processos em uma organização brasileira que utiliza o ERP Protheus, um estudo de caso único foi desenvolvido. Este método de pesquisa científica, de abordagem qualitativa, é justificado por ser um estudo piloto, no qual, busca-se testar as perguntas que norteiam o projeto, hipóteses, e principalmente os instrumentos e procedimentos realizados no projeto. Ao final do estudo exploratório, perguntas podem ser modificadas, retiradas ou acrescentadas, os instrumentos de pesquisa podem ser refinadas, ou hipóteses podem ser reformuladas, com base no que funcionou ou deixou de funcionar Yin (2001).

Segundo Cauchick Miguel et al. (2018), o estudo de caso aplica uma análise profunda de um ou mais objetos, com o uso de diversos instrumentos de coleta de dados e interação entre o pesquisador e o objeto de pesquisa.

Em conjunto com o método de pesquisa, buscando descobrir e analisar os processos de compras e pagamentos, foi escolhido um método de projeto de referência, o PM², que descreve os passos a serem realizados em um projeto de mineração de processos. O método escolhido foi o proposto em Van Eck et al. (2015), por ser um processo iterativo, pelo qual pode-se partir de um questionamento mais abstrato e, através de ciclos de mineração e análise, descer a níveis mais específicos até que os objetivos do projeto sejam alcançados.

Os passos da metodologia PM² são ilustrados na **Figura 10** e definidos em seis etapas: planejamento, extração, processamento de dados, mineração e análise, avaliação e melhoria e suporte de processos. Cada etapa do projeto é relacionada a diferentes objetos de entrada e saída, conforme:

- Objetos relacionados ao objetivo: questões de pesquisa; descobertas de desempenho; descobertas de conformidade e ideias de melhoria.

- Objetos de dados: sistemas de informação; dados de eventos; *logs* de eventos.
- Modelos: modelos de processos; modelos analíticos.

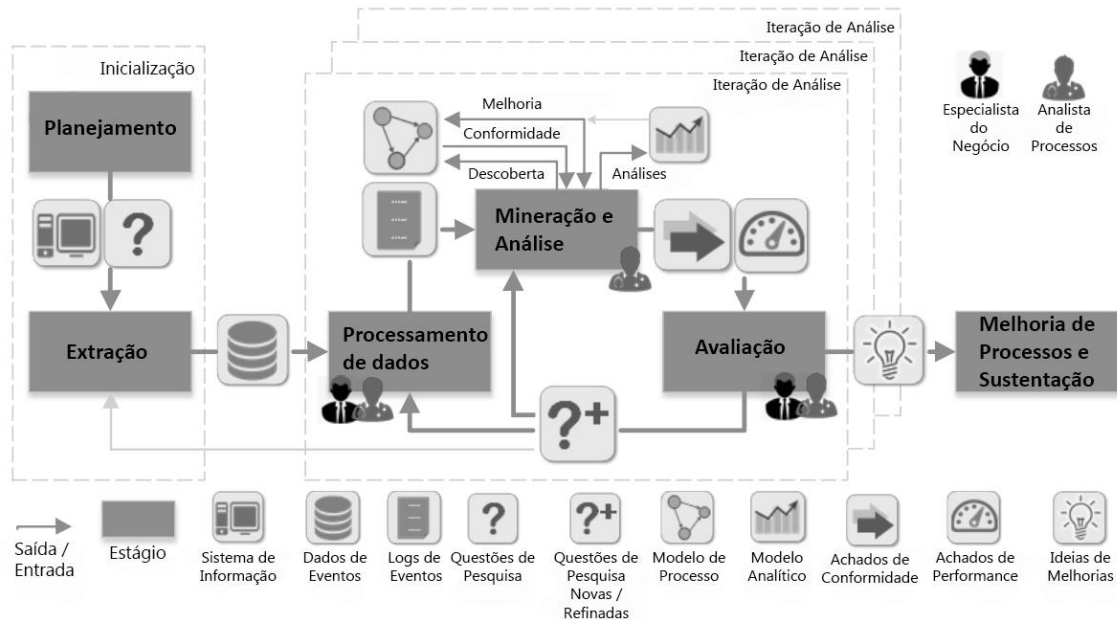


FIGURA 10 - VISÃO GERAL DA METODOLOGIA PM². ADAPTADO DE Van Eck et al. (2015).

Um projeto de mineração de processos seguindo a metodologia PM² começa nas etapas de planejamento e extração, nas quais as questões de pesquisa são definidas e os dados de eventos são extraídos. Neste ponto, em Segers (2007) ressalta-se uma importante decisão que deve ser tomada quanto a cardinalidade da relação entre as tabelas envolvidas na extração. Esta cardinalidade nos registros de eventos pode ser divergente (1:n) ou convergente (n:1), o que dependerá da escolha do caso do processo. Por exemplo, se num processo de suprimentos, o Pedido de Compra for escolhido como o caso do processo, a cardinalidade com a atividade de Entrada de Mercadorias será de divergência, uma vez que várias entregas de mercadorias poderão ser realizadas para atender o pedido. Por outro lado, se o Pagamento for escolhido como caso do processo, a cardinalidade tende a ser de convergência, dado que o pagamento é feito para várias mercadorias.

Na etapa de extração e preparação dos registros de eventos, são criadas consultas nas tabelas do ERP Protheus e, devido ao sistema possuir uma estrutura própria de geração de *logs* que não seguem os padrões MXML ou XES, o resultado das consultas é mapeado na ferramenta XESame para geração dos registros de eventos no padrão XES que poderá ser importado no ProM.

Na etapa de processamento de dados, as atividades de criação dos modelos (visões de processo), agregação de eventos, enriquecimento dos registros de eventos com informações importantes às análises do processo e filtragem dos registros, são executadas com o objetivo de criar o modelo prescrito do processo e conectá-lo ao modelo que será gerado na mineração.

Após, na etapa de mineração e análise, as técnicas/algoritmos de mineração de processos são aplicadas sobre os registros de eventos, a fim de responder as questões de pesquisa e obter informações sobre o processo, identificar possíveis problemas de performance e avaliar conformidade. Para esta etapa, mediante a experiência obtida na análise da literatura e estudos de caso de aplicação na área de suprimentos, foram selecionados os algoritmos descritos na **Tabela 2**.

TABELA 2 - ALGORITMOS E APLICAÇÕES

Algoritmo	Indicações
Mineração Alpha (α Miner)	O algoritmo pode ser utilizado em processos simples e para se gerar uma ideia inicial do modelo do processo. Por outro, lado não consegue lidar com processos complexos, ruídos, tarefas repetidas e <i>loops</i> .
Mineração Heurística (Heuristic Miner)	- Pode lidar com ruído, processos que contenham <i>loops</i> , mas não consegue lidar com tarefas repetidas. - Indicado por ser muito prático, podendo ser usado para expressar o comportamento total contido nos registros de eventos, ou apenas no comportamento principal (ou seja, nem todos os detalhes e exceções).
Mineração Fuzzy (Fuzzy Miner)	- Consegue lidar com ruídos, processos com <i>loops</i> e tarefas repetidas. - É uma técnica robusta indicada para processos mais complexos, pois permite abstrair detalhes do fluxo. - Permite reduzir e focar o comportamento exibido, utilizando os conceitos de agregação e abstração. No modelo simplificado, pode-se dar ênfase, destacando o comportamento mais significativo presente nos registros de eventos.

Mineração Genética (<i>Genetic Miner</i>)	- Comparado aos demais algoritmos, a mineração genética apresenta a melhor precisão do modelo extraído, especialmente para a mineração de processos que apresentam comportamento mais complexo. - Muitas vezes necessita de alto tempo de processamento para chegar ao resultado, além de utilizar apenas registros históricos de <i>log</i> .
Mineração Indutiva (<i>Inductive Miner</i>)	- Altamente extensível e permite muitas variantes da abordagem básica, possuindo técnicas capazes de lidar com comportamentos pouco frequente e grandes modelos e registros de eventos. - É atualmente uma das principais abordagens de descoberta de processos, dado a sua flexibilidade, garantias formais e escalabilidade.
<i>Conformance Checker</i>	- Utilizada para analisar a conformidade entre o modelo prescrito e o modelo minerado.
<i>LTL Checker</i>	- É possível especificar e verificar uma variedade de instruções lógicas mantidas nos registros de eventos.

Em seguida no estágio de avaliação, analisa-se os resultados obtidos na mineração, relacionando-os às ideias de melhoria do processo de negócio. Nesta etapa deve-se diagnosticar, verificar e validar os resultados.

No último estágio de melhoria e sustentação/suporte do processo, busca-se implementar as melhorias levantadas e dar apoio operacional.

No projeto pode-se executar diversas iterações de análises, de forma que cada iteração execute: processamento de dados, mineração e análise e avaliação. Se na iteração, as descobertas encontradas sobre a questão de pesquisa específica forem satisfatórias, pode-se utilizar as descobertas para melhoria e suporte do processo em questão. E caso, novos questionamentos surjam, inicia-se uma nova iteração, a fim de responder a nova pergunta.

Para validar a qualidade do modelo de processo minerado, um teste de conformidade (*Conformance Checker*) entre o registro de evento e o modelo gerado será realizado, quantificando o valor de cobertura (*fitness*) entre 0 e 1. Além da avaliação de cobertura, para verificar a simplicidade e equilíbrio do modelo entre precisão e generalização, será quantificada a quantidade de nós e arcos existentes no modelo gerado, observando se há repetição de nós no modelo.

4 Revisão Exploratória de literatura: Mineração de Processos na Gestão de Suprimentos

Ao longo da cadeia de valor definida em Porter (1986) as atividades podem ser agrupadas em atividades primárias, aquelas envolvidas na produção dos produtos ou serviços, ou em atividades de apoio, estas suportam as atividades primárias provendo insumos ou infraestrutura. Atividades ligadas à gestão de suprimentos, área que se estende por toda a cadeia de valor, suportando todas as atividades, conforme ilustrado na **Figura 11**, devem ser otimizadas para melhorar a competitividade da organização.



FIGURA 11 - CADEIA DE VALOR DE PORTER. FONTE: Porter (1986)

Para Lau et al. (2009), o alto grau de incerteza nas demandas dos clientes e a complexidade de gestão das atividades da cadeia de suprimentos, tem estimulado o uso do gerenciamento da cadeia de suprimentos para melhorar o nível de satisfação dos clientes.

Na gestão de suprimentos o conjunto de processos e atividades podem englobar: criação de planos de materiais; desenvolvimento de requisitos e padrões de qualidade para aquisições; desenvolvimento de estratégias de suprimentos; seleção de fornecedores e desenvolvimento/manutenção de contratos; ordens de materiais e

serviços (requisição, aprovação de requisição, cotações, criação de ordem de compra, aprovação de ordem de compra, receitação de produtos, pagamentos); gestão de fornecedores; e outras.

Tal capilaridade de processos e atividades executadas na área de suprimentos possuem ainda relacionamentos com outras áreas de negócio, por exemplo, gestão de armazenagem e produção, ou ainda, inter-relacionamento com outros negócios, tornando o sistema de gestão de suprimentos complexo e de difícil entendimento holístico pelas partes interessadas.

Segundo Cooper et al. (1997) o escopo da cadeia de suprimentos pode ser definido de acordo com o número de integrantes, atividades e funções envolvidos ao longo da cadeia.

Neste sentido, a mineração de processos pode ser empregada para identificar os reais processos sendo executados nas organizações, comparar com modelos da organização ou modelos de maturidade, ou ainda identificar desvios e fraudes, e em última instância melhorar tais processos na busca pelo diferencial competitivo.

Em Van der Aalst (2011b) divide-se a aplicação da mineração de processos entre as perspectivas intraorganizacional "*intra-organization*" e interorganizacional "*inter-organization*". A primeira respeita os limites da organização individual, já a segunda extrapola tais limites. Apesar de muitos projetos de mineração de processos serem conduzidos sobre uma organização singular, é possível ampliar o escopo da aplicação para processos interorganizacionais, mediante dois pressupostos básicos: deve haver colaboração entre as organizações para lidar com casos comuns de processos e compartilhamento de conhecimento e infraestruturas. Com base nesta divisão de perspectivas, apresenta-se a seguir os estudos de caso que aplicam a mineração de processos no domínio da gestão de suprimentos.

4.1.1 Casos intraorganizacionais

No trabalho intitulado “Verificação abrangente de conformidade baseada em regras com mineração de processos” propõe-se uma abordagem avançada de verificação de conformidade baseada em regras e com a aplicação da mineração de processos Caron et al. (2013a). Nesta abordagem, utiliza-se uma taxonomia bidimensional, na qual os padrões de regras são classificados de acordo com sua perspectiva de mineração de processos. A **Tabela 3** fornece uma visão geral da taxonomia bidimensional e das subcategorias de cada dimensão.

TABELA 3 - TAXONOMIA DE REGRAS DE NEGÓCIOS PARA MINERAÇÃO DE PROCESSOS (ADAPTADO DE CARON ET AL. (2013A)).

	Regras de Cardinalidade	Regras de Coexistência	Regras dinâmicas baseadas em dados	Regras de tempo	Regras de Propriedade Estática
Perspectiva Funcional do Processo	Cardinalidade da Atividade	- Coexistência de Atividades - Coexistência de Eventos de Atividades	Existência orientada por dados	Existência de atividade orientada pelo tempo	Coexistência de Artefato de Eventos
Perspectiva do Fluxo de Controle do Processo		- Atividades não sobrepostas - Ordem das Atividades	Condição de atividade orientada por dados	Regras de Fluxo de Controle Orientadas pelo Tempo	Tempo Absoluto
Perspectiva Organizacional do Processo	Cardinalidade do Originador	- Segregação de Funções - Ligação de Funções - Regras de Envolvimento Temporal	- Autorização Exógena - Regras de Atributos do Originador	Regras de Obrigações Temporais	- Autorização Estática - Atributo Obrigatório do Originador - Regras de Delegação
Perspectiva dos Dados	Cardinalidade dos dados do evento	Coexistência de Dados de Eventos	- Regra derivada de dados de eventos - Regra de Comparação de Dados de Eventos	Integridade Dinâmica	- Regra do valor de dados do evento - Regra de Formato de Dados do Evento

Dimensão de Regras de Negócio

Posteriormente para validar a abordagem de verificação de conformidade baseada em regras de negócio com a aplicação da mineração de processos, os pesquisadores realizaram um estudo de caso aplicando esta abordagem a um processo de compras para pagamento. A pesquisa abordou dois problemas essenciais: o primeiro, busca certificar que contas fraudulentas não sejam registradas e pagas; o segundo, procura garantir responsabilidades adequadas aos funcionários nos sistemas de informação baseado em seus respectivos perfis. A abordagem mostrou vantagens sobre as abordagens tradicionais de auditoria e verificação de conformidade, como: capacidade de verificar e analisar toda a população de casos; alta eficácia de detecção de problemas; além de fornecer evidências suficientes, a abordagem é capaz de garantir independência do auditor Caron et al. (2013b).

Em Jans et al. (2011) utiliza-se a mineração de processos aplicada no contexto de transações fraudulentas, sendo o processo de aquisição foco do estudo de caso em uma organização que utiliza SAP como seu sistema ERP (*Enterprise Resource Planning*). Neste estudo de caso são aplicadas análises nas três perspectivas: de processo, organizacional e de caso; e as atividades do processo de aquisição identificadas como: 'Criação de Ordem de Compra (OC)', 'Alteração de item da OC', 'Assinatura da OC', 'Liberação da OC', 'Recebimento de itens da OC', 'Recebimento de cobrança dos itens da OC' e 'Pagamento'. As técnicas utilizadas no estudo de caso foram a Mineração *Fuzzy*, Análise de Sequência de Desempenho, Máquina de Estados Finita (*Final State Machine – FSM*) e LTL Checker, aplicadas no software ProM, com exceção da FSM aplicada no Petrify. Chega-se à conclusão de que a Mineração de Processos permite a auditoria, fornecendo teoria e algoritmos que verificam a conformidade dos processos, além de fornecer ferramentas para auxiliar auditores na detecção de fraudes ou outros fluxos de trabalhos em estágios muito anteriores aos métodos tradicionais de auditoria.

Na tese defendida por Stoop (2012) aplica-se a mineração de processos no processo de compras. Tal processo está focado na aquisição de materiais com fornecedores e pode ter muitas variações, dependendo das políticas internas e controles. Similar a Jans et al. (2011), também neste estudo de caso busca-se identificar possíveis fraudes nos processos de suprimentos, todavia, identificou-se um aspecto importante a ser tratado na fase de seleção dos *logs*: o momento em que os dados são extraídos pode afetar os resultados da mineração de processos, isto porque durante o processo de extração podem ocorrer mudanças, causando interpretações erradas. Tal situação pode ser contornada filtrando-se os casos situados próximos ao período de corte (data de extração).

O trabalho de Stoop (2012) faz uma ressalva importante quanto ao uso de técnicas para tratar o problema de ruído, como exemplo, Mineração Heurística. Quando o propósito é detectar desvios ou fraudes, os ruídos geralmente são os alvos da investigação. Assim, no contexto de detecção de fraudes, é importante manter todos os caminhos identificados.

Ainda em Stoop (2012) foram utilizadas diversas ferramentas sobre o caso de estudo: ProM, Disco, ProM Heuristic Miner plugin, ProM Fuzzy Miner plugin, ProM Performance Sequence Analysis plugin, ProM Organizational Miner plugin, Role Activity Matrix, ProM LTL Checker plugin.

O trabalho apresentado em Bezerra; Wainer (2011), apesar de não aplicar a Mineração de Processos na área de suprimentos, mas em *logs* de processos gerados sinteticamente, traz contribuições importantes na detecção de fraudes em sistemas baseados em processos (*process aware systems – PAS*) e para os objetivos desta pesquisa. Os autores afirmam que, independente da semântica, classificar um evento raro ou infrequente como uma anomalia ou fraude pode ser uma abordagem ingênua, isto porque outros indicadores, além da frequência no *log*, precisam ser considerados.

Um algoritmo para detecção de fraudes em *logs* de sistemas baseados em processos – PAS, foi proposto em Bezerra; Wainer (2011). Neste são abordados quatro diferentes

métricas: compatibilidade (*fitness*), adequação estrutural, adequação comportamental e tamanho. Os três primeiros foram apresentados no trabalho de (ROZINAT; AALST, VAN DER, 2005) *apud* Bezerra; Wainer (2011), sendo que o primeiro indica quão compatível é um modelo e o *log*, já o segundo e terceiro representam números entre zero e um, onde zero significa que não há compatibilidade entre o modelo e o *log*, enquanto um significa que existe conformidade total entre o modelo e o *log*. Além dos três primeiros, Bezerra; Wainer (2011) sugerem uma nova métrica “tamanho”, que representa a contagem de estágios (*places*), transições e arestas numa Rede de Petri.

Nos resultados alcançados por Bezerra; Wainer (2011) evidenciou-se que os indicadores de adequação estrutural e comportamental não são apropriados para detectar caminhos anômalos. Por outro lado, os indicadores de compatibilidade (*fitness*) e tamanho tiveram performance próxima a 100% na detecção de traços anômalos (positivos).

Para prover uma orientação prática na aplicação da metodologia PM², Van Eck et al. (2015) aplicaram a metodologia a um projeto na IBM, com objetivo de responder questões relacionadas à performance do processo de compras. Tal processo começa com a criação da requisição de compra, após o pedido é enviado ao fornecedor e o processo termina quando todos os itens pedidos são entregues.

No projeto, Van Eck et al. (2015) extraem todos os eventos relacionados às compras do sistema SAP e aplicam a mineração de processos no software ProM com diversas extensões, conforme:

- A criação de *logs* de eventos para diferentes visões ocorreu com o conjunto de ferramentas XESame ou importando arquivo CSV com CSVi no ProM. Tais visões foram criadas com ABAP do SAP e SQL no banco de dados.
- Filtragens foram realizadas com diversas extensões, exemplo, “*Filter Log by Attribute Values (FLAV)*” e “*using Simple Heuristics (uSH)*”.
- Para realizar agregações de subprocessos foi criado um programa (*script*) personalizado.

- Na análise de processos com estatística básica para inspecionar casos individuais e eventos foi utilizado “*Log Visualizer (LV)*”.
- Para analisar os eventos e casos do log sob a perspectiva de tempo, possibilitando revelar padrões de tempo, desvio de conceito e processamentos em lote os autores utilizaram “*Dotted Chart (DC)*”.
- Na descoberta de processos dos diversos algoritmos disponíveis os autores optaram pelo “*Inductive visual Miner (IvM)*”, por ser rápido e capaz de produzir modelos estruturados que podem ser melhorados ou analisados com outras extensões no ProM.
- A extensão “*Replay a Log for Performance/Conformance Analysis (RLP/C)*” foi utilizada para aprimorar o modelo com informações de tempo ou qualidade.
- Para enriquecer os *logs* de eventos com informações adicionais, como duração e outras, utilizou-se a extensão “*Feature Prediction (FP)*”.
- A extensão “*Explore Variants (EV)*” foi utilizada na fase de mineração e análise para examinar duas variações de ordens emergenciais canceladas.

Através do projeto na IBM, usando a metodologia PM², a equipe partiu de uma questão de pesquisa abrangente “Como o processo de compras se parece?”, e foi capaz, ao longo das iterações, de refinar a questão de pesquisa para obter novas ideias valiosas e proposições concretas de melhoria dos processos.

Outro caso de aplicação na perspectiva intraorganizacional é o trabalho apresentado por Mahendrawathi et al. (2017). Eles realizaram um estudo em uma empresa de agrotóxicos, aplicando a mineração de processos para investigar a pós implementação do sistema ERP. O foco da análise foi o processo de aquisição suportado pelo sistema SAP. O estudo de caso foi guiado por perguntas, como: Como a aquisição é suportada pelo sistema SAP? Qual é o caminho mais comum e incomum seguido no processo de aquisição?

Após a extração e preparação dos *logs*, o modelo de fluxo de controle é criado com o sistema Disco, um software de mineração de processos. Em seguida, análises foram conduzidas para comparar os caminhos normais versus os caminhos "incomuns" no processo de aquisição. Os resultados do processo de mineração mostraram que algumas atividades são descontroladas, com grande variação nos tempos de execução. Em alguns casos, as atividades são realizadas rapidamente, mas os procedimentos não são seguidos. Finalmente, os gargalos foram descobertos, indicando oportunidades para melhorar a eficiência operacional.

Diversos casos apresentados anteriormente aplicam a mineração de processos com objetivo de verificação de conformidade e identificação de fraudes. Entretanto, para se tornar uma ferramenta de auditoria interna e externa, é preciso que a mineração de processos seja colocada à prova no contexto da auditoria contemporânea. Neste cenário, o trabalho apresentado em Jans et al. (2013) busca comprovar que, muito obstatante a mineração de processos exija esforços e aquisição de habilidades para ser aplicada, ela pode substituir práticas de auditoria existentes reduzindo a carga de trabalho e aumentando a eficiência da auditoria, ainda, mantendo o trabalho em conformidade com os padrões de auditoria. Como modelo padrão de auditoria é utilizado o *framework* Modelo de Risco de Auditoria – *Audit Risk Model*.

Jans, Alles & Vasarhelyi (2013) identificam as formas de agregação de valor que a mineração de processos pode trazer às práticas de auditoria, a citar:

- A mineração de processos analisa toda a população de dados e não apenas uma amostra.
- Pode-se auditar metadados e não apenas dados inseridos pelo auditado.
- A mineração de processos permite ao auditor formas mais eficazes de desenvolver o modelo de risco de auditoria, conduzindo análises orientadas aos processos e procedimentos.
- A mineração de processos permite análises que não são possíveis com as ferramentas de auditoria atuais, como descobrir as reais formas como os

processos são executados no dia a dia e identificar relações entre os indivíduos que executam as tarefas.

Por fim, para ilustrar os benefícios da mineração de processos aplicada em auditorias, Jans, Alles & Vasarhelyi (2013), recorrem a um estudo de caso usando dados fornecidos por uma empresa multinacional de serviços financeiros na Europa. O processo foco de análise foi o de aquisições, que é suportado pelo sistema SAP da empresa.

4.1.2 Casos Interorganizacionais

Em Maruster et al. (2002) investiga-se processos de gestão de suprimentos ao longo da cadeia que pode conter diversos atores (manufatura, atacadista, varejo, e outros), sob os quais, de modo geral, não há uma visão holística de todo o processo. Os autores apresentam uma visão geral de métodos para descoberta do processo entre e dentre as diversas organizações da cadeia de suprimentos. Em especial, sugere-se uma abordagem prática de métodos heurísticos de aprendizagem contendo três métricas: métrica de causalidade e métricas locais e globais para representar as relações de sucessão, para desenvolver um modelo de regressão logística. Tal método foi capaz de descobrir quase todas as conexões diretas, mesmo os *logs* contendo problemas de paralelismo, desbalanceamento e ruído.

Ainda em Maruster et al. (2002), dispõe-se que o método deve ser constituído de duas etapas, a primeira responsável pela habilitação dos sistemas de cada parte envolvida na cadeia de suprimentos para disposição de registro de todas as tarefas e/ou atividades executadas no processo, sendo requisito primordial a definição de um número comum para identificação do processo entre as partes. A segunda etapa consiste na aplicação das técnicas de mineração de processos para identificação de todo o processo da cadeia de suprimentos e análises pelas partes envolvidas sobre o processo minerado para propor eventuais mudanças.

Rastrear um processo de sistema no qual os eventos não são registrados como um único caso é muito difícil, e um agravante pode ser encontrado em processos de gerenciamento de suprimentos, onde existem diferentes configurações de números de identificação e vários estágios de agregação, o que impede o rastreamento de itens individuais (produtos, ativos ou serviços) como um caso de processo Gerke et al. (2009). Os autores investigaram o padrão EPCglobal para processamento de eventos de rádio frequência (RFID), criando e aplicando um algoritmo capaz de lidar com os desafios dos processos de Gestão de Suprimentos e ainda apresentaram os resultados com um protótipo baseado no Modelo de Referência de Operações de Cadeia de Suprimentos (SCOR) para avaliação. As ferramentas utilizadas na pesquisa foram o “*Supply Chain Editor*” e o ProM, neste último utilizou-se de algoritmos de mineração de processos presentes no sistema, em especial a Mineração Heurística.

No trabalho apresentado por Van der Aalst (2011b) a mineração de processos interorganizacional é particionada em horizontal e vertical. A partição vertical refere-se à distribuição de casos entre diversas organizações com o mesmo processo atual para todas estas, de forma que haverá diferentes casos ou variantes do mesmo processo. Por outro lado, a partição horizontal funciona como um corte do mesmo processo em diversos pedaços, de tal maneira que diferentes organizações executam partes desse processo através da colaboração. Este esquema de partição vertical e horizontal é ilustrado na **Figura 12**.

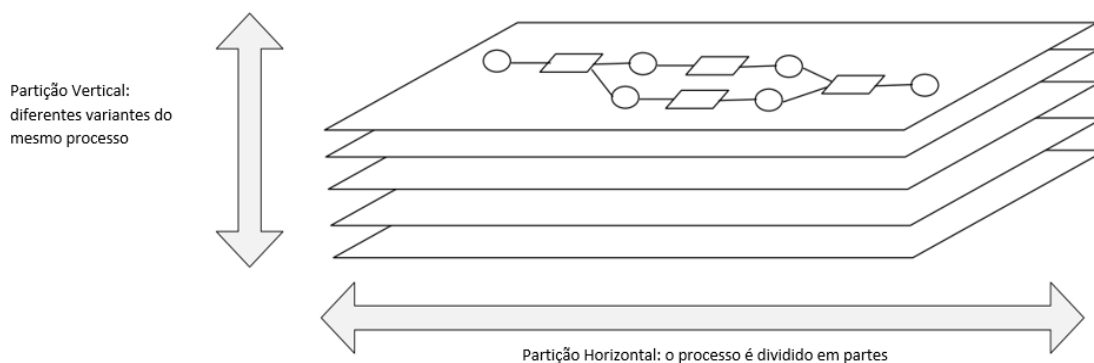


FIGURA 12 - ESQUEMA DE PARTIÇÃO VERTICAL E HORIZONTAL DE PROCESSOS INTERORGANIZAÇÕES.

FONTE: (AALST, VAN DER, 2011)

Em Engel et al. (2016) é apresentado o *framework* EDImine, que tem o objetivo de minerar processos sobre sistemas de colaboração entre negócios suportados por troca de dados eletrônicos (EDI). Tal estudo é classificado na partição horizontal de processos interorganizações, visto que a troca de dados e informações geralmente é conduzida entre diferentes organizações a fim de colaborar sobre um mesmo processo, como exemplo, pode-se citar uma indústria e distribuidora trocando informações sobre um EDI logístico de vendas.

Uma questão importante em padrões de troca de dados – EDI, como EDIFACT, ANSI X.12, e outros, é que estes impõem mais restrições nas mensagens trocadas Van der Aalst (2011b e Engel et al. (2016). Os campos da mensagem possuem significado predefinido, as mensagens podem ser síncronas ou assíncronas e, em algumas situações, uma organização pode observar todas as mensagens trocadas, mas sem poder procurar dentro das mesmas. Dessa forma, a mineração de processos precisa ser realizada com o desafio de ver apenas parte do processo geral.

4.1.3 Análise

A quantidade de estudos de casos em que se aplica a Mineração de Processos na área de Gestão de Suprimentos, apesar de restrita, nos permite responder a primeira pergunta de nosso método de pesquisa, se há casos de aplicação da mineração de processos na área de gestão de suprimentos.

É importante notar que as pesquisas para este estudo foram realizadas de acordo com critérios especificados na metodologia, com palavras-chave, bem como um processo de revisão do resumo e conteúdo dos artigos selecionados que aplicam a mineração de processos na área de suprimentos. Existe a possibilidade de publicações terem sido omitidas, visto que, os resultados das buscas estão sujeitos às limitações das fontes automatizadas de pesquisa.

Por um lado, a existência de autores que analisaram vários conjuntos de dados na área de suprimentos, permite identificar campos onde a mineração de processos provou a sua eficácia em identificar os processos, desvios ou fraudes e oportunidades de melhorias nestes processos, tanto na perspectiva intraorganizacional quanto interorganizações. Por outro lado, identificar campos na área de suprimentos em que nenhum estudo de caso foi ainda relatado, pode encorajar novas pesquisas a explorar o uso da mineração de processos nestes campos no futuro.

Foi possível também responder às questões 2, 3 e 4 que guiam esta revisão, através da análise dos estudos de caso apresentados na Seção 4.1 e Seção 4.2, sintetizados na **Tabela 4** e descritos a seguir.

- 1) Classificação do tipo de aplicação da mineração de processos, podendo ser Descoberta de Processos, Conformidade dos processos descobertos com algum modelo de referência e Melhoria de processos (extensão). Verifica-se que a maioria dos trabalhos aplica a mineração de processos para descobrir e avaliar a conformidade do processo descoberto com um modelo de referência.
- 2) Se a mineração de processos foi aplicada na área de suprimentos. Apenas um artigo selecionado, o de Bezerra; Wainer (2011) não foi aplicado na área, mas foi selecionado por contribuir na área de conformidade e identificação de fraudes.
- 3) Outra diversificação importante da aplicação de mineração de processos na área de suprimentos se dá entre as perspectivas intraorganizacional e interorganizacional. Pode-se verificar que a aplicação da mineração de processos para Melhoria de processos nos casos identificados, se dá principalmente na perspectiva interorganizacional. Isto pode ser explicado devido ao fato de a colaboração entre organizações em uma cadeia de suprimentos ter como foco principal a otimização de tal cadeia.
- 4) É interessante notar que o software para mineração de processos mais utilizado nos casos foi o ProM. ProM é um framework para algoritmos de mineração de processos gratuito que pode ser utilizado por usuários e desenvolvedores.

Todavia existem outros softwares comerciais que realizam a mineração de processos, como Celonis e Disco.

- 5) As técnicas aplicadas à mineração de processos na área de suprimentos de maior frequência nos casos identificados foram de algoritmos personalizados, mineração *fuzzy* e mineração heurística. Contudo não se identificou, em alguns casos, um detalhamento dos algoritmos empregados na mineração de processos.

TABELA 4 - SÍNTESE DOS CASOS ESTUDADOS

Trabalho	Tipo de mineração	Suprimentos?	Perspectiva	Algoritmos	Software	Desvios/Fraudes?
Maruster et al. (2002)	Descoberta e melhoria de processos	Sim	Interorganizacional	Não especificado	Não especificado	Não
(GERKE <i>et al.</i> , 2009)	Descoberta e conformidade	Sim	Interorganizacional	Heuristic Mining	ProM	Não
(BEZERRA; WAINER, 2011)	Conformidade	Não	Intraorganizacional	Anomaly detection algorithm based on α -algorithm	Não especificado	Sim
(JANS <i>et al.</i> , 2011)	Descoberta e conformidade	Sim	Intraorganizacional	Fuzzy Mining, Performance Sequence Analysis, Final State Machine – FSM e LTL Checker	SAP, ProM, Petriify	Sim
(ENGEL <i>et al.</i> , 2016)	Descoberta, conformidade e melhoria de processos	Sim	Interorganizacional	EDImine Framework	ProM	Não

Trabalho	Tipo de mineração	Suprimentos?	Perspectiva	Algoritmos	Software	Desvios/Fraudes?
(STOOP, 2012)	Descoberta e conformidade	Sim	Intraorganizacional	Performance Sequence Analysis, Conformance Analysis, Dotted Chart, Social Mining, Matrix of activities and functions, Heuristic Mining, Fuzzy Mining, Organizational Mining, LTL Checker	ProM, Disco	Sim
Caron et al. (2013b, 2013a)	Descoberta e conformidade	Sim	Intraorganizacional	Heuristic Miner, LTL Checker	ProM	Sim
Jans et al. (2013)	Descoberta e conformidade	Sim	Intraorganizacional	Não especificado	SAP, ProM, ProMimport	Sim
Van Eck et al. (2015)	Descoberta e melhoria de processos	Sim	Intraorganizacional	XESame, CSVi, Filter Log by Attribute Values, using Simple Heuristics, personalized script, Log Visualizer, Dotted Chart, Inductive visual Miner, Replay a Log for Performance/Conformance, Feature Prediction, Explore Variants	SAP, ProM	Não
Mahendrawathi et al. (2017)	Descoberta, conformidade e melhoria de processos	Sim	Intraorganizacional	Fuzzy Miner	Disco	Não

Destes casos estudados, Caron et al. (2013b, 2013a e Van Eck et al. (2015 e Jans et al. (2011 e Mahendrawathi et al. (2017 e Stoop (2012), focam seus estudos nos processos internos de suprimentos e abordam as três principais aplicações da mineração de processos: descoberta, análise de conformidade e melhoria de processos.

Já em Engel et al. (2016 e Gerke et al. (2009 e Maruster et al. (2002), os estudos extrapolam as fronteiras de uma organização para atingir todos os elos de uma cadeia de suprimentos, inserindo novos problemas e maiores complexidades, por outro lado, com a aplicação da mineração de processos é possível identificar os processos interorganizacionais, resultando em uma visão holística da cadeia de suprimentos e possibilitando análises para sua otimização. Tal foco na otimização da cadeia de suprimentos caracteriza o estudo de caso nas três principais áreas da mineração de processos: descoberta, conformidade e extensão.

Esta revisão de estudos de caso foi publicada em Schwaickardt; Dantas (2018).

5 Estudo de Caso único com método PM²

Dentro do universo de organizações que utilizam o sistema ERP Protheus, o subconjunto que habilitam a solução de *log* do sistema é pequeno, isto porque as tabelas de *log* tendem a consumir muitos recursos dos servidores. Somando-se à dificuldade em encontrar uma organização que tenha ativo o *log* do sistema, ainda há o fato de muitas terem receio em abrir seus dados para pesquisa.

Sobre este subconjunto, buscou-se aproximação com organizações por indicações de consultores da Totvs para verificar a disponibilidade em participar do estudo. Dentre as empresas questionadas sobre o interesse em participar do estudo, a Fattu, se disponibilizou em apoiar a pesquisa, conforme termo de aceite disponível no **Anexo 1**.

A Fattu é uma empresa de grande porte, distribuidora de suprimentos para comunicação visual fundada em 1996. A empresa possui centros de distribuição nas cidades do Rio de Janeiro, Campinas, Ribeirão Preto, Goiânia e Espírito Santo, com uma capacidade de armazenamento de mais de 6.000 m².

5.1 Objetivos

O objetivo principal do estudo de caso é entender como o processo de compras e pagamento está sendo realizado no ERP Protheus dentro da organização, validando as regras definidas pela organização, que de maneira geral, definem que um pagamento deve ter previamente uma nota fiscal e da mesma forma a nota fiscal deve ter seus respectivos pedidos de compra registrados.

Seguindo o método PM² adotado nesta pesquisa, a seguir detalha-se o caminho percorrido, as análises realizadas e os resultados alcançados.

5.2 Planejamento

Nesta fase de planejamento busca-se responder às questões: “Como o processo de compras e pagamentos se parece na organização?”, “Todos os pagamentos, referentes

a suprimentos, estão percorrendo o caminho Pedido de Compra, Entrada de Mercadoria (NF) e por fim Pagamento?”.

No ERP Protheus é possível habilitar a geração dos *logs* pelo “*Embedded Audit Trail*”. Nesta solução pode-se definir quais tabelas do sistema serão alvo de auditoria, definindo todos campos ou selecionando-se campos principais das tabelas e quais operações serão auditadas (inclusão, alteração e exclusão). Na **Tabela 5**, a seguir é apresentada a estrutura das tabelas de auditoria que são criadas para cada tabela auditada.

TABELA 5 – ESTRUTURA DAS TABELAS DE AUDITORIA DO ERP PROTHEUS

Campo	Descrição
FIELD	Nome do campo
COLD	Conteúdo antes da atualização
CNEW	Conteúdo depois da atualização
TYPE	Tipo de dado
RECNO	Número de registro
USER	Usuário no Protheus
UDB	Usuário do banco de dados
HOSTNAM	IP do Servidor do Protheus
PROGRAM	Nome da rotina principal a partir do menu do Protheus
OPERATI	Operação no banco de dados (inclusão, alteração, exclusão)
DTIME	Campo composto de data e hora
UNQ	Chave única do registro
OK	indica se o registro do log é confiável ou foi adulterado

Na Fattu a configuração foi definida para gravar todas as tabelas do sistema em que houver movimentações, todos os campos da tabela e todas as operações (inclusão, alteração e exclusão). Devido às tabelas possuírem muitas colunas que não são utilizadas, as consultas foram otimizadas selecionando-se apenas as informações mais importantes de cada tabela e serão apresentadas na seção seguinte de Extração e Processamento dos dados.

5.3 Extração e Processamento dos dados

Após a seleção das colunas pertinentes de cada tabela, o modelo das tabelas é o mesmo, a única diferença é que o campo “FIELD” da estrutura de tabela de *logs* passa a ter apenas informações pertinentes, ou seja, não contém informações nulas ou irrelevantes. A estrutura das tabelas de Pedido de Compra (SC7), Itens de Notas de Entrada (SD1) e Contas a Pagar (SE2), estão ilustradas na **Figura 13**. Neste ponto destaca-se que seria impossível criar um identificador único de caso com esta estrutura, visto que a coluna “FIELD” armazena em suas linhas informações de todos os campos da tabela em questão.

Tabela	Colunas
SC7	TABLE_ID, FIELD, COLD, CNEW, RECNO, USER_NAME, USERID, UDB, PROGRAM, OPERATI, DTIME
SD1	TABLE_ID, FIELD, COLD, CNEW, RECNO, USER_NAME, USERID, UDB, PROGRAM, OPERATI, DTIME
SE2	TABLE_ID, FIELD, COLD, CNEW, RECNO, USER_NAME, USERID, UDB, PROGRAM, OPERATI, DTIME

FIGURA 13 - ESTRUTURA DAS TABELAS DE LOGS NO ERP PROTHEUS.

Para solucionar o problema mencionado, possibilitando gerar um relacionamento entre estas tabelas e criação de um identificador de caso único, é necessário fazer um pivô (operação para converter valores de uma coluna em novas colunas na tabela) nas colunas “FIELD” e “CNEW” de forma a gerar novas colunas com os campos selecionados em “FIELD” e preencher estas colunas com as informações de “CNEW”. A nova estrutura é ilustrada na **Figura 14**.

A cardinalidade dos relacionamentos entre as tabelas é de muitos para muitos (N:N), uma vez que um pedido de compra pode-se dar entrada através de mais de uma nota, uma nota pode conter mais de um pedido de compra, uma nota pode gerar múltiplos títulos (parcelas) e uma ou mais notas podem ser liquidadas em múltiplos títulos.

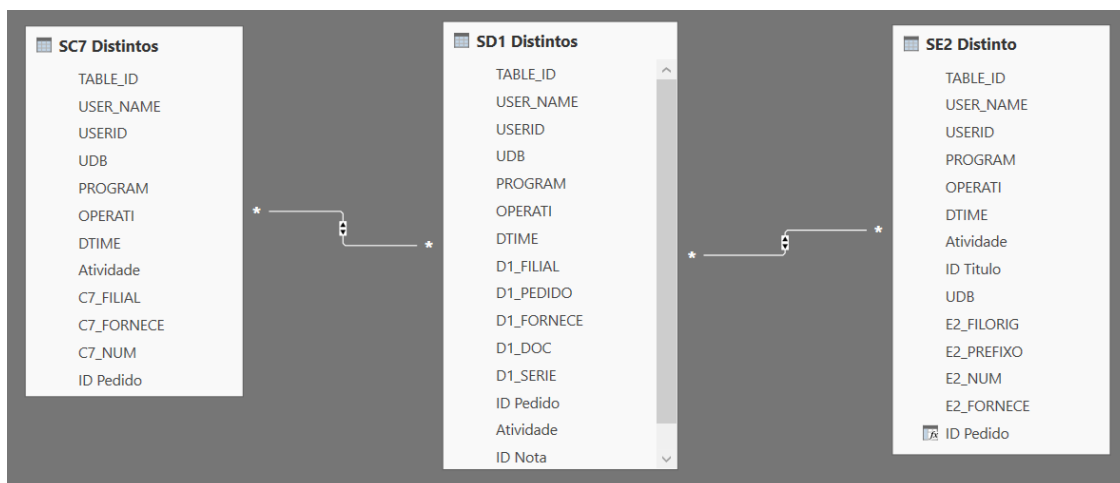


FIGURA 14 - TABELAS DE LOGS PIVOTADAS DO ERP PROTHEUS

Neste ponto, identificou-se que o funcionamento da criação dos registros de eventos pelo sistema gera uma linha de inclusão, atualização e exclusão para cada campo da tabela, separando-os pelos minutos e segundos entre o preenchimento de cada campo. Este grau de especificação no *log* não é interessante para a mineração e será agregado para um nível acima, ou seja, hora e minuto.

Outra descoberta importante sobre os tipos de operações (inclusão, alteração e exclusão) foi de que na tabela de Pedidos de Compra há apenas operações do tipo Inclusão e Exclusão, e cinquenta e seis (56) registros de *log* estão sem informação do número do Pedido de Compra, estes foram desconsiderados.

Já na tabela Notas de Entrada nas operações de Inclusão o sistema não registra no *log* informações de campos, todos são gerados com informações vazias, caracterizando apenas um “*relock*”, função do sistema para reservar um registro no banco de dados. Posteriormente a essas transações de Inclusão na tabela Notas de Entrada, o sistema gera no *log* operações de Alteração (*Update*), preenchendo informações nos campos. Contudo as operações de Alteração em diversos casos não preenchem as informações necessárias em um mesmo registro (linha), separando-os por tempos distintos de gravação, conforme exemplificado na **Figura 15**.

D1_FILIAL	D1_DOC	D1_SERIE	D1_FORNECE	D1_PEDIDO	Pedido	RECNO	ID Nota	ID Pedido
01	000007952	4	001098		608109	158345	010000079524001098	01608109001098
				608109	608109	158345	010000079524001098	01608109001098

FIGURA 15 - EXEMPLO TRATATIVA PARA CRIAÇÃO DE IDENTIFICADORES

Para tratar estes problemas um identificador “ID” foi criado, verificando-se primeiramente o identificador do Pedido de Compra e, quando este não houver, buscando-se o identificador da Nota de Entrada. Este processo para criação do identificador também é realizado na tabela de Contas a Pagar, ligando-a à tabela de Notas de Entrada para buscar o identificador da compra e, quando não encontrado, seleciona-se o identificador do título financeiro.

A tabela estruturada dos registros de eventos contém o identificador gerado, o código da classe da atividade (composto pelo código da rotina e tipo de operação), data e hora do registro, nome do executor da atividade e tabela de origem. Esta tabela é gerada pela união dos registros de eventos das três tabelas apresentadas no modelo da **Figura 14** e ilustrada na **Figura 16** a seguir.

ID	Atividade	Time	Executor	Origem
000105288GPEINPS	FINA750-X	12/06/2018 15:32:00	patricia.oliveira	SE2
01000000001001527	FINA050-X	27/06/2018 09:52:00	bianca.araujo	SE2
01000000001001527	FINA050-X	27/06/2018 09:53:00	bianca.araujo	SE2
01000000001001527	FINA050-X	05/07/2018 13:53:00	bianca.araujo	SE2
01000000001001527	FINA050-X	22/08/2018 14:32:00	bianca.araujo	SE2
01000000001001527	FINA050-X	14/09/2018 11:21:00	bianca.araujo	SE2
01000000001001527	FINA050-X	10/10/2018 17:09:00	bianca.araujo	SE2
01000000001002169	MATA103-I	11/01/2019 13:11:00	eliana.ferreira	SE2
01000000001002169	MATA103-U	11/01/2019 13:11:00	eliana.ferreira	SE2
01000000001002169	MATA103-X	11/01/2019 13:10:48	eliana.ferreira	SD1

FIGURA 16 - ESTRUTURA DA TABELA DE LOGS

As atividades presentes nos registros de eventos são descritas na **Tabela 6**. Das diferentes classes de atividades identificadas, as classes GPEM660 e GPEM670 serão

desconsideradas, por se tratarem de títulos financeiros advindos do módulo de folha de pagamento, visto que o foco dessa pesquisa se dá nos processos de suprimentos para pagamento. As atividades de intervenção direta nas bases de dados MPSDU e “vazio” também são desconsideradas, porém estas foram apresentadas à empresa foco deste estudo para averiguação.

TABELA 6 - CLASSES DE ATIVIDADES

Código	Descrição
FINA050	Contas a Pagar
FINA750	Funções Contas a Pagar
GPEM660	Geração Títulos Folha
GPEM670	Integração Títulos Folha
MATA116	Nota de Conhecimento de Frete
MATA103	Documento de Entrada
MATA121	Pedido de Compra
MATA140	Pré-Documento de Entrada
MATA410	Geração Automática de Pedido de Venda
MATA521	Exclusão Documento de Saída
MATA953	Apuração ICMS
MPSDU	Ferramenta do sistema para intervenção manual na base de dados
“vazio”	Operação direta na base de dados

Esta etapa de extração e processamento é uma das mais dispendiosas do projeto, uma vez que cada sistema de informação pode trabalhar de uma forma diferente e ter uma estrutura de *log* própria. A qualidade da estrutura do *log* impacta diretamente na viabilidade do projeto e na qualidade do modelo de processo minerado e será discutida nas próximas etapas.

5.4 Mineração e Análises

No início da mineração, o primeiro passo executado foi de converter o arquivo de *log* de “.csv” para XES, formato padrão para mineração de processos. Ao processar a conversão no ProM, o painel exibido na **Figura 17** é retornado, no qual identifica-se um processo com 15.841 casos, 261.002 eventos, sendo estes de 60 classes de eventos e

o primeiro evento registrado em 28/05/2018 e último em 05/02/2019 (data da última coleta/extração nas tabelas de *logs*).

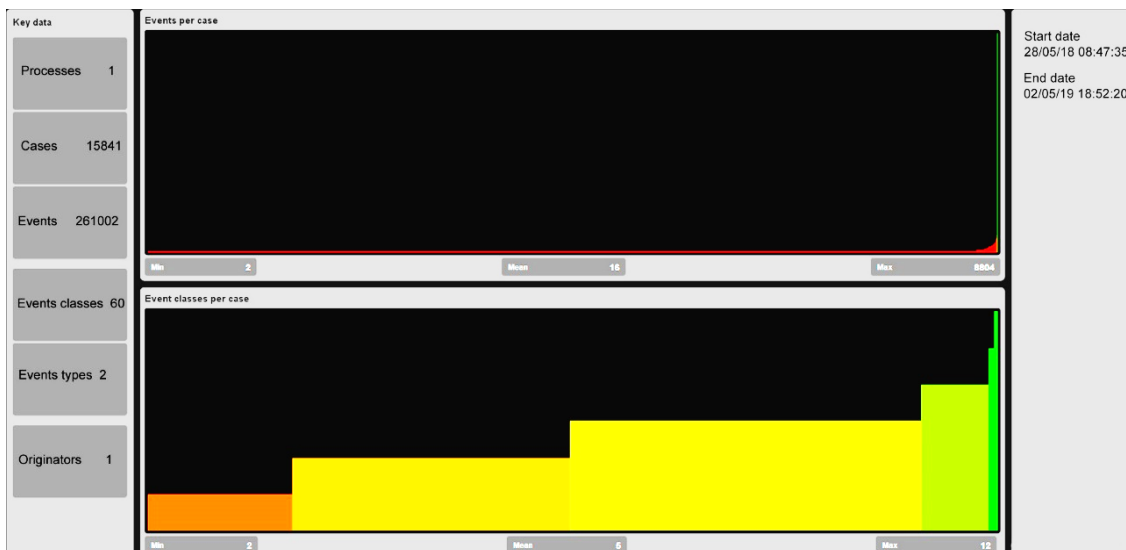


FIGURA 17 - PAINEL DOS REGISTROS DE EVENTOS

Posteriormente os algoritmos de mineração para identificação do modelo do processo foram aplicados, iniciando-se pelo algoritmo Alpha que não é capaz de lidar com *loops* e tarefas repetíveis. Dessa forma, o modelo minerado em rede de Petri apresenta retornos à mesma atividade, não sendo assim o melhor modelo para descrever o processo em análise.

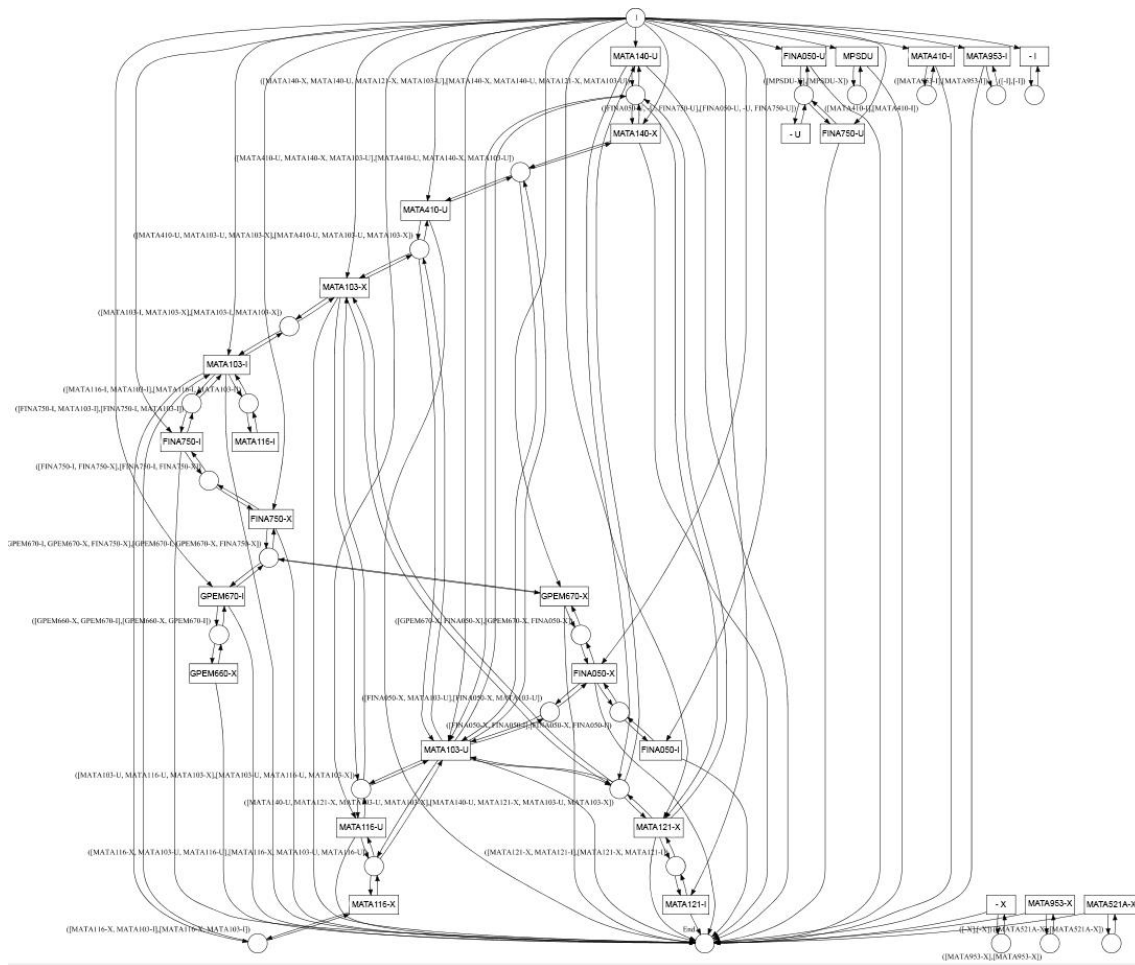


FIGURA 18 - MODELO DE PROCESSO MINERADO COM ALGORITMO ALPHA

Em seguida aplicou-se o algoritmo de mineração *Fuzzy*, no qual foram testados diferentes níveis de cortes para ruídos e eventos poucos frequentes, porém, mesmo considerando-se todos os eventos, o modelo gerado não identifica claramente o início e fim do processo. Por outro lado, o algoritmo identificou modelos à parte do processo principal, como os dos eventos de intervenção direta nas bases de dados. O modelo *Fuzzy* minerado é apresentado na **Figura 19**, na qual as atividades são representadas em forma de retângulos e as conexões pelas linhas, sendo que, as linhas mais espessas representam os caminhos mais frequentes.

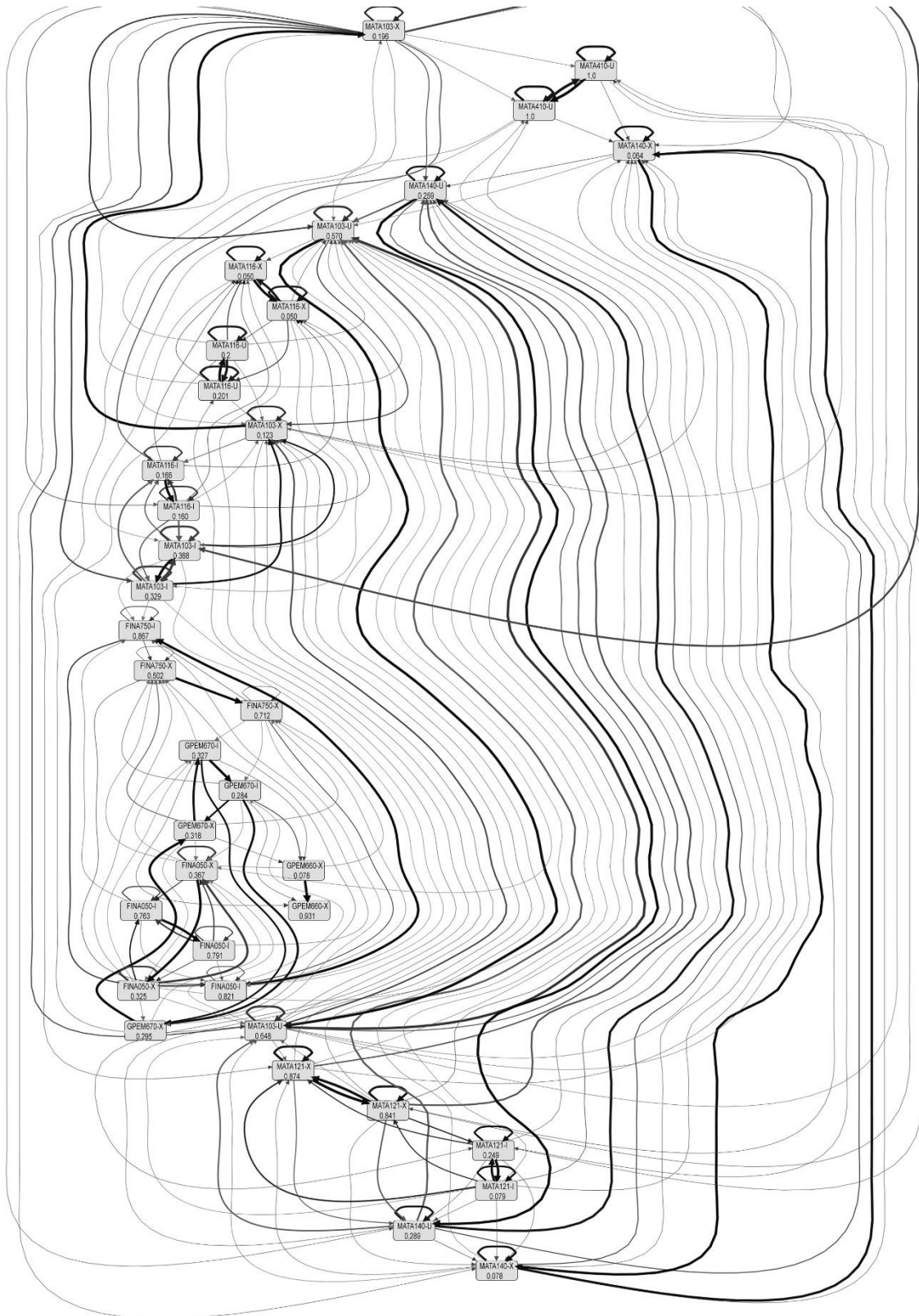


FIGURA 19 - MODELO MINERADO COM ALGORITMO FUZZY (LINHAS MAIS ESPESSAS REPRESENTAM OS CAMINHOS MAIS FREQUENTES)

O próximo algoritmo aplicado foi o de mineração heurística, na qual configurou-se níveis de frequência 0, 10, 30, 50, 80 e 100 por cento. Na configuração de frequência 0,

ilustrado na **Figura 20** todos os eventos contidos nos registros de eventos (*logs*) são considerados e nas demais configurações os eventos são filtrados de acordo com o percentual de frequência de ocorrência nos casos. No modelo gerado os retângulos representam as atividades e os números próximos às linhas representam a frequência do caminho.

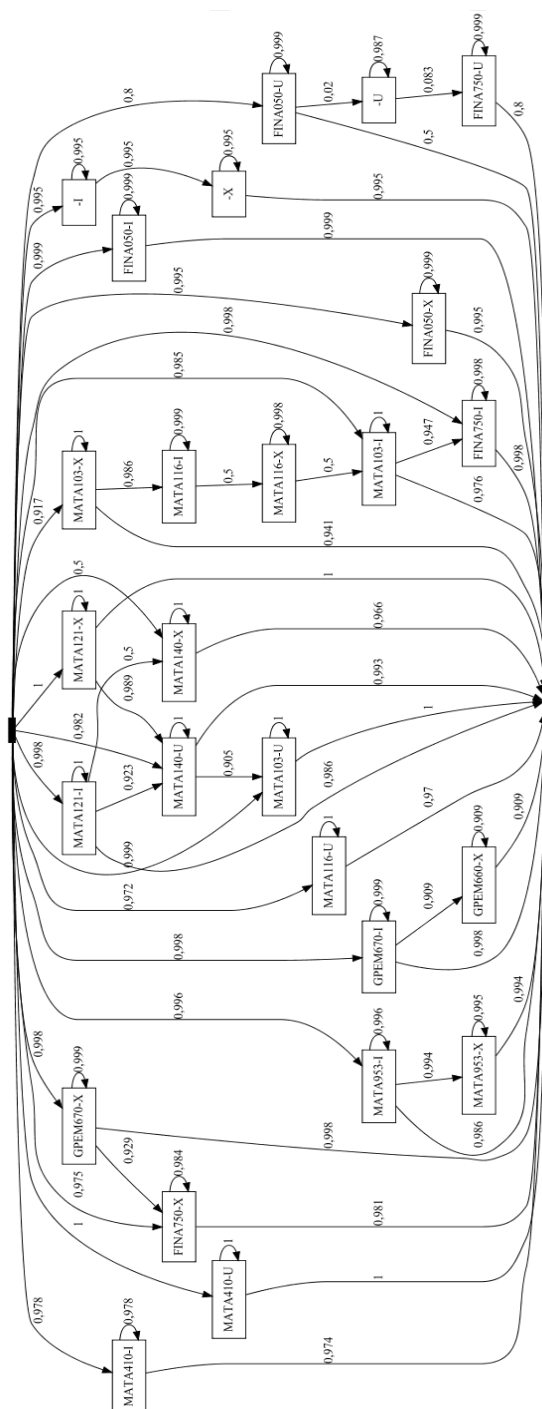


FIGURA 20 - MODELO MINERADO COM ALGORITMO HEURÍSTICO (TODOS EVENTOS)

Destas configurações com filtro de frequência, a configuração de 10% gerou um bom modelo contemplando os principais eventos e desconsiderando os ruídos e eventos que não fazem parte do processo de compras de suprimentos, como os eventos de pagamentos de títulos da folha de pagamento e as movimentações realizadas diretamente na base de dados. O modelo com frequência de 10% é exibido na **Figura 21**.

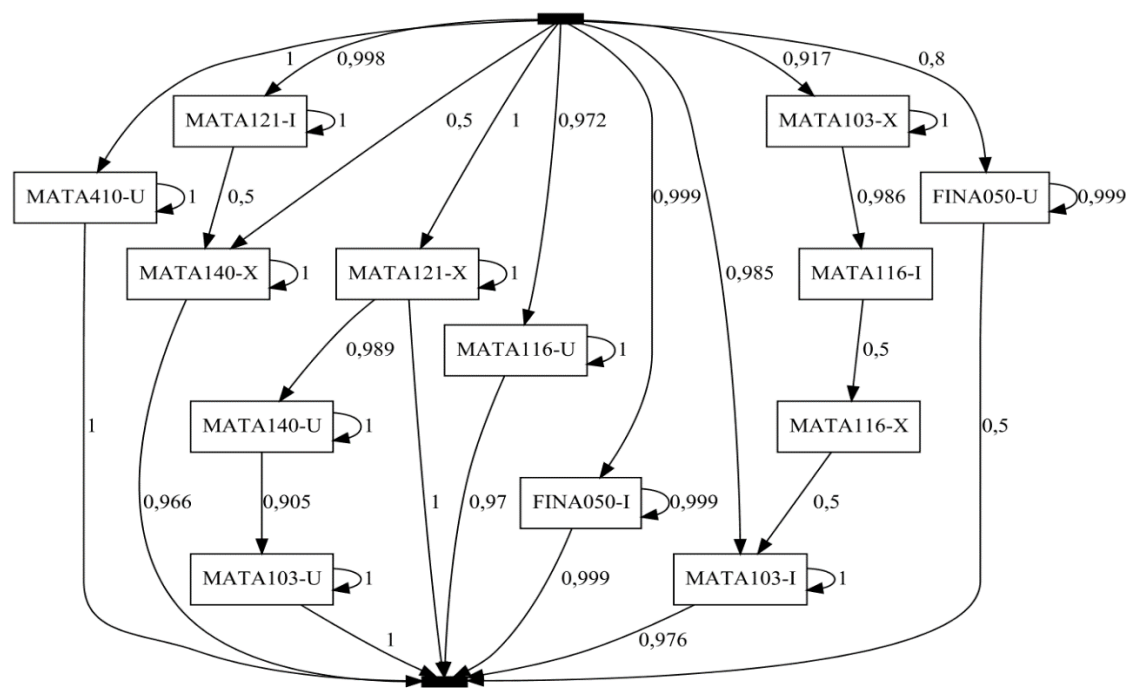


FIGURA 21 - MODELO MINERADO COM ALGORITMO HEURÍSTICO (10% DE FREQUÊNCIA)

O último algoritmo para descoberta do modelo de processo aplicado foi o de mineração indutiva, que se mostrou muito flexível e capaz de lidar com os registros de eventos pouco frequentes. Enquanto na mineração *Fuzzy* os registros de eventos pouco frequentes geraram modelos de fluxo a parte, na mineração indutiva estes foram contemplados no modelo. O modelo que contempla todos os 15.841 casos contidos nos registros de eventos extraídos é exibido na **Figura 22**.

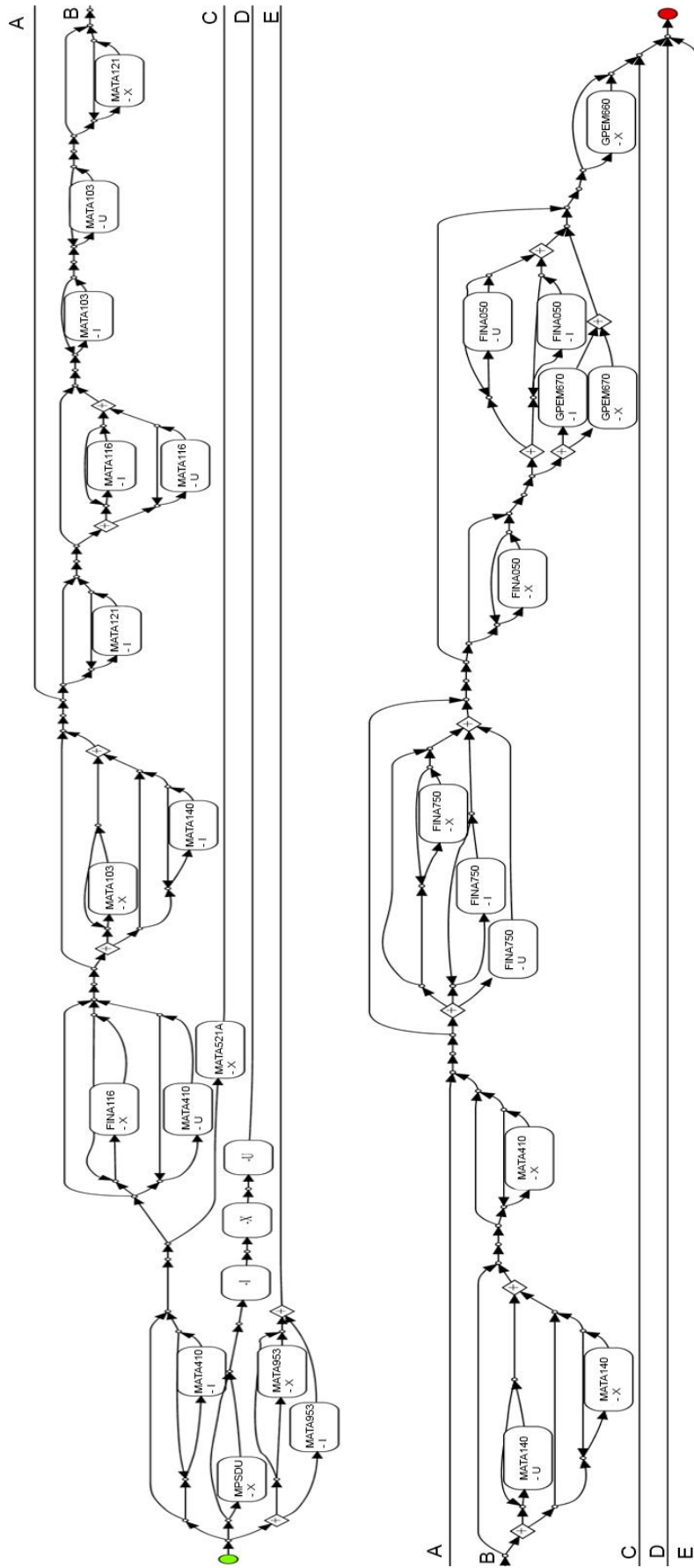


FIGURA 22 - MODELO MINERADO COM ALGORITMO INDUTIVO (TODOS OS EVENTOS)

Ao aplicar a configuração de Pareto, filtrando-se 80% das atividades, os eventos pouco frequentes ou ruídos nos logs foram desconsiderados, desta forma os eventos de pagamento da folha de pagamento e as movimentações diretas nas bases de dados são omitidas no modelo, resultado coerente com o modelo de processo de pagamentos esperado. O modelo minerado nesta configuração é apresentado na **Figura 23**.

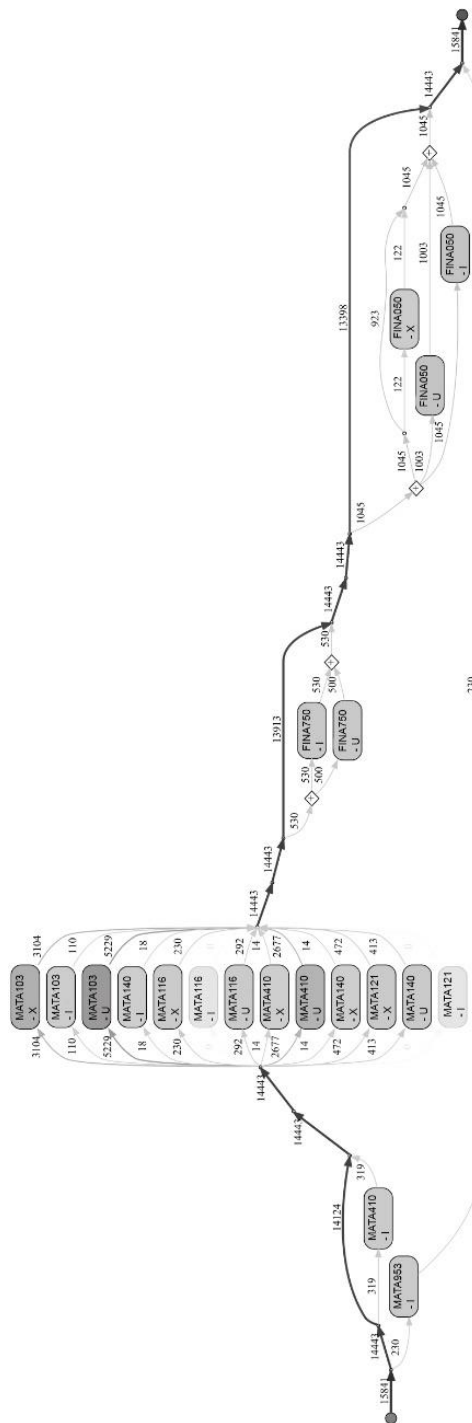


FIGURA 23 - MODELO MINERADO COM ALGORITMO INDUTIVO (80%)

Assim como na mineração heurística, o algoritmo de mineração indutiva foi capaz de filtrar os ruídos e eventos pouco frequentes, como os eventos da folha de pagamento que geram títulos financeiros, as movimentações diretas nas bases de dados e a rotina de apuração de ICMS.

As quatro métricas de qualidade indicadas na literatura foram avaliadas, conforme:

1) Cobertura: para avaliar cobertura no *log* contendo todos os registros de eventos, após identificar o modelo de processo com o algoritmo de mineração indutiva, aplicou-se no ProM o *plugin* “*Replay a Log on Petri Net for Conformance Analysis*” com objetivo de medir a cobertura. Uma cobertura igual a 1.0 indica que todos os casos contidos no *log* são representados pelo modelo minerado, por outro lado, valores menores que 1.0 indicam que alguns casos não se enquadram no modelo. Os indicadores de cobertura retornados pelo algoritmo foram de 0,37 para cobertura do caso (*Trace Fitness*), 0,98 para cobertura modelo (*Move-model Fitness*) e 0,38 para cobertura do *log* (*Move-log Fitness*).

2) Precisão: na análise de precisão do modelo, aplicou-se no ProM *plugin* “*Check Precision based on Align-ETConformance*” observou-se 0,77 no valor do indicador “Precisão”, 0,83 no indicador “Precisão Invertida” e 0,80 no indicador “Precisão Balanceada”. Os indicadores próximos a 1.0 indicam um bom nível de precisão, ademais o modelo não permite caminhos muito diferentes dos registros de eventos, o que poderia ser verificado caso os modelos fossem parecidos com o “modelo de flor” ilustrado na **Figura 9 (d)**.

3) Generalização: ao aplicar o *plugin* “*Measure Precision/Generalization*” sobre o *replay* de avaliação de conformidade, obteve-se o valor de 0,99 para o indicador de generalização. O indicador demonstra um bom nível de generalização, o que se pode verificar também na análise visual dos modelos que não se parecem com o “modelo de enumeração” ilustrado na **Figura 9 (e)**.

4) Simplicidade: o modelo é simples, o que se pode verificar por não haver objetos

mais do que o necessário para ilustrar o modelo e não há repetição das classes de eventos ao longo do modelo.

Os registros de eventos contemplados no período possuem casos incompletos, ou seja, casos em que apenas parte das atividades foram realizadas, que podem ser casos que ainda estão em curso ou casos que realmente não executam todas as atividades previstas no modelo descrito pela empresa. Para analisar apenas os casos completos, gerou-se um *log* que filtra apenas os casos em que o identificador (ID) do caso percorre desde o Pedido de Compra até o pagamento do título financeiro. Estes casos são ilustrados no modelo da **Figura 24**.

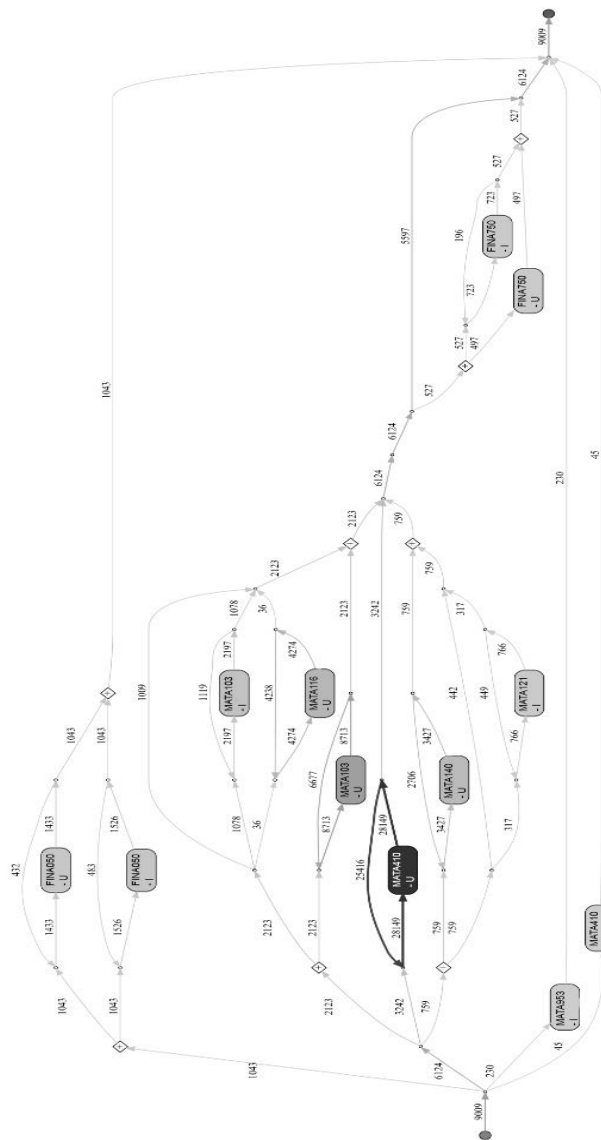


FIGURA 24 - MODELO MINERADO FILTRANDO APENAS CASOS COMPLETOS

Até este ponto, a mineração foi realizada com o software ProM, solução acadêmica de código aberto e que possui grande flexibilidade de escolha de algoritmos. Contudo, aplicar a mineração de processos no ProM pode ser um desafio a usuários que não tenham um profundo conhecimento na área. Para avaliar a mineração em uma solução mais amigável, selecionou-se o software Celonis, solução comercial, mas que permite o uso acadêmico através de uma licença acadêmica.

Importando o arquivo “.csv” no Celonis, primeiro configura-se os tipos de dados de cada coluna e em seguida informa-se as colunas de identificação do caso, atividade, tempo e executor, respectivamente. Assim que criado o modelo, o software aplica seu algoritmo interno automaticamente e retorna um conjunto de análises sobre os casos do processo e o modelo do processo minerado. O modelo minerado com base no *log* completo, contendo todos os casos e atividades é ilustrado na **Figura 25**. Ademais aplicando-se a configuração de 80/20 o modelo minerado é exibido na **Figura 26**, no qual pode-se verificar que as atividades infrequentes e que não fazem parte do processo de compras de suprimentos, como as rotinas da folha de pagamento foram desconsideradas.

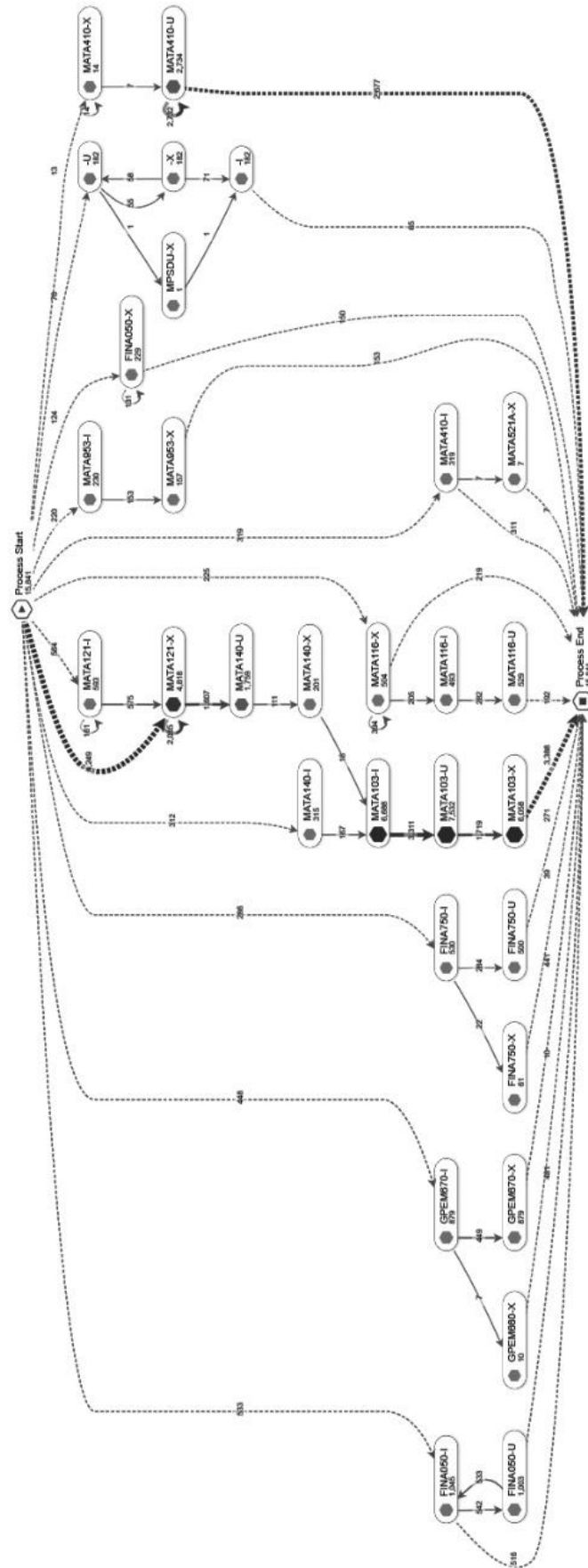


FIGURA 25 - MODELO MINERADO COM CELONIS (COMPLETO)

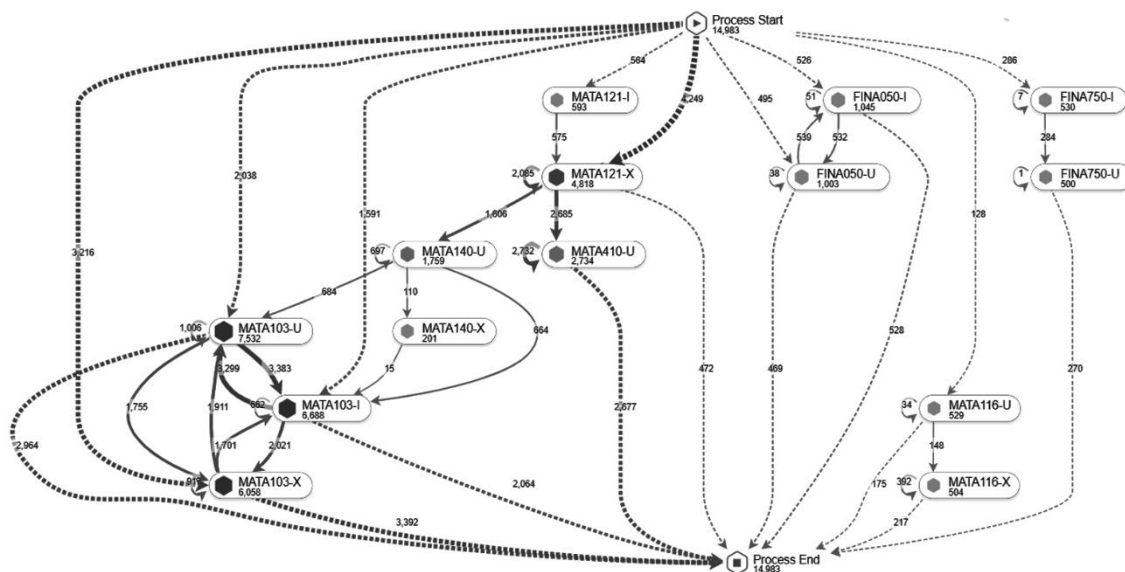


FIGURA 26 - MODELO MINERADO COM CELONIS (80/20)

Dos casos contidos no modelo minerado completo, em 78% dos casos, o processo é finalizado em até 2 dias, mas há uma quantidade significativa de casos, 621 que duraram entre 18 a 254 dias. Essa avaliação da duração dos casos faz parte da perspectiva de análise do processo e é ilustrada na **Figura 27**.

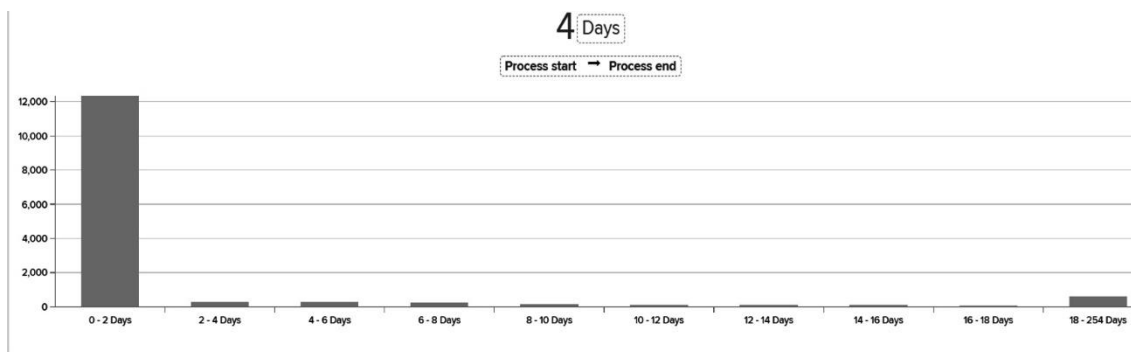


FIGURA 27 – ANÁLISE DA DURAÇÃO DOS CASOS

Em todos os modelos minerados, pode-se verificar que o processo se inicia bifurcando em mais de uma atividade, podendo iniciar pela inclusão do pedido de compra, documento de entrada, título financeiro, ou até por atividades de exclusão de registros. Para criar um modelo mais simples, filtrou-se do conjunto de casos todas as atividades de exclusão, atividades da folha de pagamento e de intervenções manuais nas bases de dados. O modelo contendo todas as possíveis atividades e conexões deste *log* filtrado foi minerado no Celonis e é exibido na **Figura 28**.

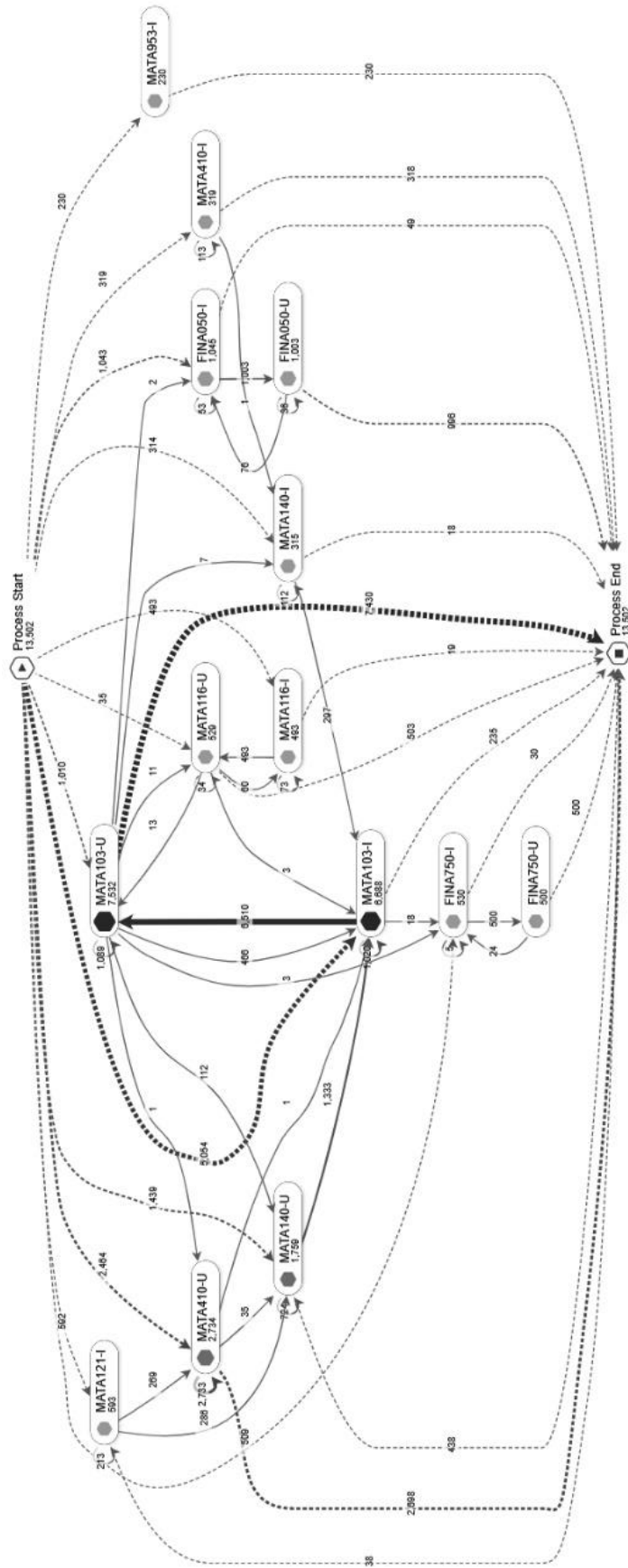


FIGURA 28 - MODELO MINERADO COM CELONIS (LOG FILTRADO)

No modelo completo e no modelo com *log* filtrado aplicou-se um novo filtro no fluxo do processo para retornar apenas os casos focados na compra e que se iniciaram pelo pedido de compra. Os modelos resultantes são ilustrados na **Figura 29** e **Figura 30**, nos quais é possível verificar que muitos pedidos após passarem pela atividade de pré-documento de entrada o processo se encerrou, indicando provável processo que ainda está em andamento.

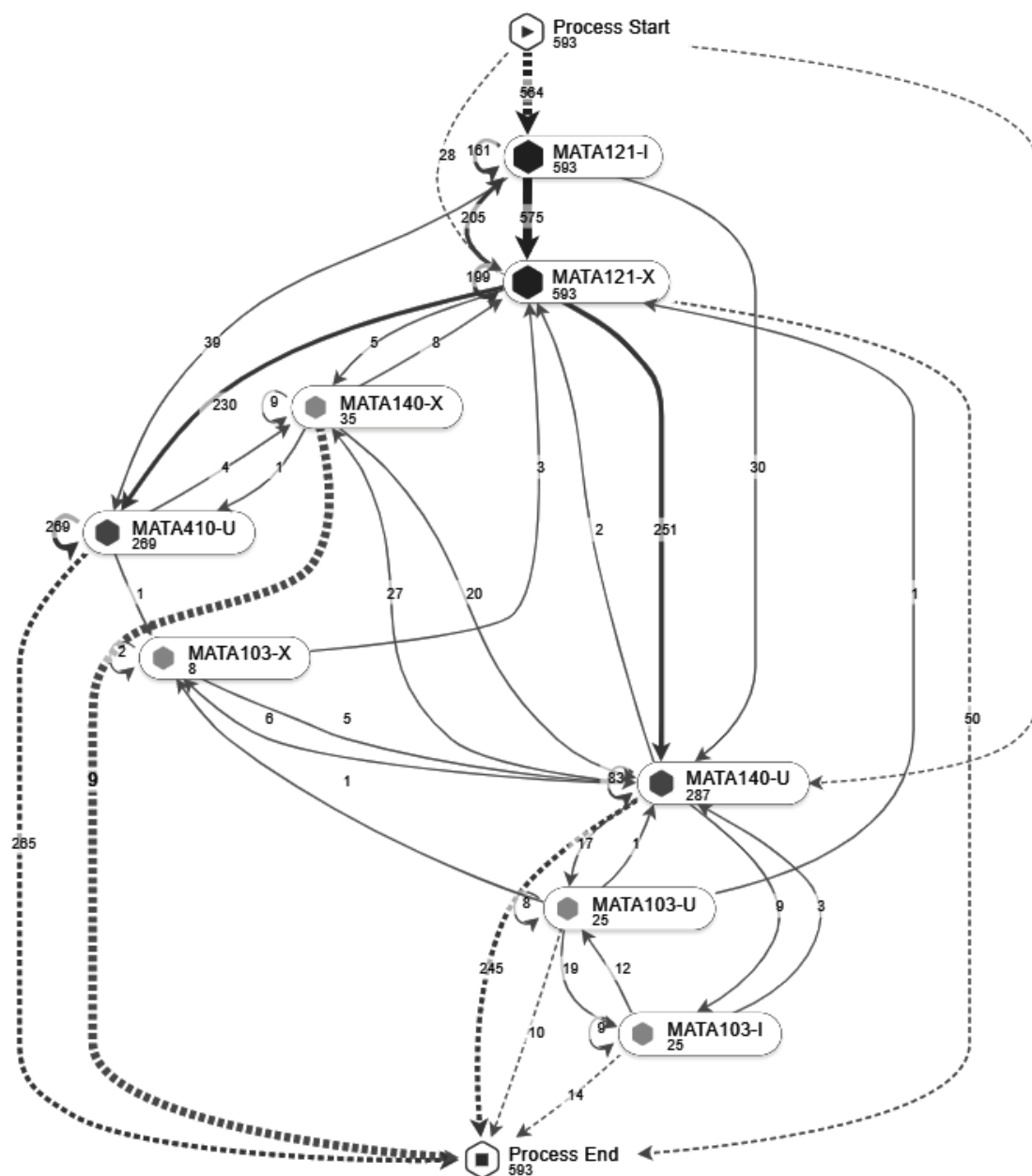


FIGURA 29 - MODELO INICIANDO NO PEDIDO (COMPLETO)

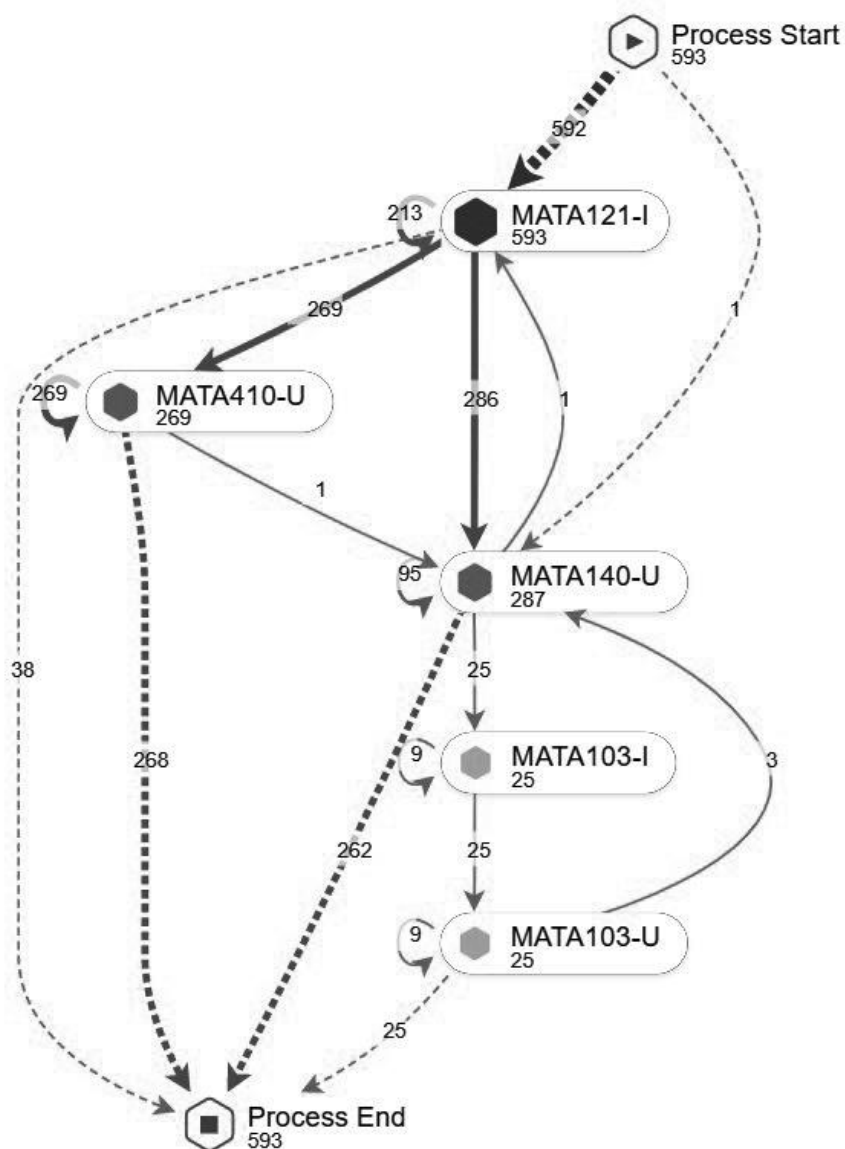


FIGURA 30 - MODELO INICIANDO NO PEDIDO (LOG FILTRADO)

Durante as iterações anteriores, a estrutura de granularidade do *log*, na qual a atividade estava identificada pelo código da rotina mais a operação realizada (inclusão, alteração ou exclusão), não permitiram a emulação dos casos em ambos os softwares de mineração, ProM e Celonis. Contudo, em uma nova iteração, considerou-se os casos no *log* que possuem um identificador completo e a atividade composta apenas pelo nome da rotina, omitindo-se assim os tipos de operações. O *log* transformado foi aplicado no Celonis resultando no modelo ilustrado na **Figura 31**.

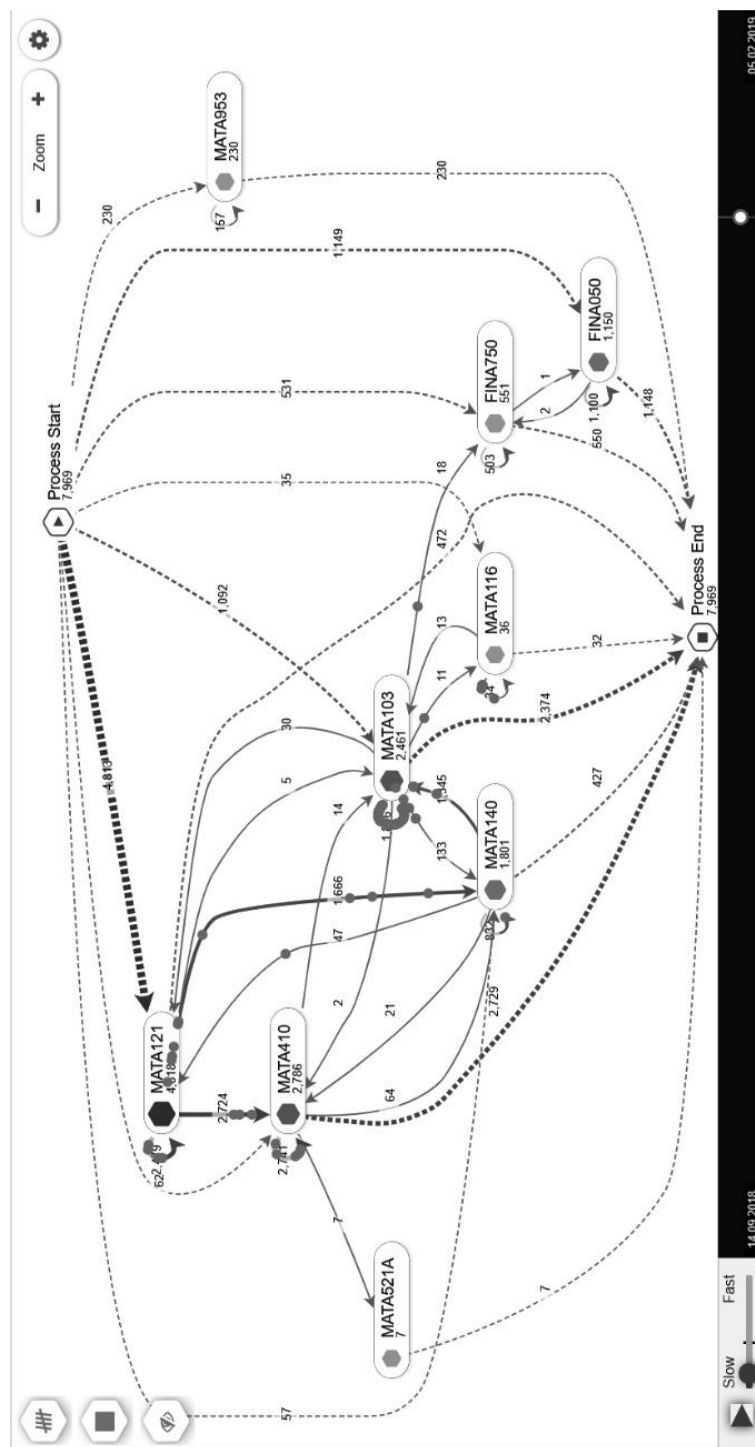


FIGURA 31 - EMULAÇÃO NO LOG FILTRADO E ATIVIDADE APENAS COM NOME DA ROTINA

Através da emulação é possível analisar como os casos se comportam ao longo do tempo, evidenciando no modelo acima uma concentração em torno das atividades de Pedido de Compra e Documento de Entrada. Esta funcionalidade é extremamente útil no monitoramento e priorização de esforços de melhoria sobre as atividades que representam gargalo no processo.

5.5 Avaliação

Apesar do processo levantado na empresa parecer simples, definido pela sequência de atividades de Pedido de Compra, Entrada da Nota Fiscal e Pagamento, na realidade exibida nos modelos de processos minerados, outras atividades são realizadas, permitindo outros caminhos no processo.

As atividades de Pré-Documento de Entrada, Geração Automática de Pedido de Venda e Nota de Conhecimento de Frete, não foram detalhadas no processo levantado previamente com a empresa, mas fazem parte do processo. Uma vez identificadas no modelo minerado, a empresa foi questionada sobre estas e confirmou que realmente fazem parte do processo, em especial, destaca-se a atividade “Geração Automática de Pedido de Venda” que gerou o questionamento “Qual a relação de um Pedido de Venda com o processo de compras?”. A empresa explicou que em processos de compras para distribuição às filiais do grupo, o sistema foi customizado para gerar automaticamente Pedido de Venda na unidade que está distribuindo os materiais.

As outras atividades citadas, também fazem parte do processo, mas foram omitidas no levantamento do processo, evidenciando um problema comum nas técnicas tradicionais de levantamento de processos como entrevistas, questionários, observações e outras.

Focando no modelo de processo minerado apenas as atividades relacionadas à compra em si (Pedido de Compra, Pré-Documento Entrada, Documento de Entrada, Nota de Conhecimento de Frete e Geração Automática de Pedido de Venda), pode-se identificar alguns padrões que podem ser detalhados na perspectiva de caso, sendo:

- 1) Pedidos de Compra que ainda não foram atendidos com entrada de Documento de Entrada. Para estes casos, é interessante manter monitoramento e definir

prazos limites que uma vez alcançados podem disparar alertas e direcionar uma análise no caso.

- 2) Documentos de Entrada diretos sem Pedido de Compra prévio, evidenciando que a diretriz no processo não é seguida no dia a dia.

Na etapa de pagamento, por sua vez, identificou-se que há pagamentos não originados pelo Documento de Entrada da compra.

5.6 Discussão dos Resultados

Se, por um lado, os modelos minerados não alcançaram índices das métricas de qualidade iguais a 1.0 (100%), que se justificam pela própria qualidade do *log* gerado pelo sistema estudado, por outro lado, observa-se nos modelos minerados o balanceamento entre precisão e generalização, o que é indicado na literatura, níveis de cobertura do modelo de 0,98 e a simplicidade (pode ser verificada ao não se observar no modelo repetições de atividades).

Foi possível, após realizar diversas transformações na estrutura do *log*, alcançar uma estrutura com perspectiva de caso, com a identificação das atividades e data e hora de execução destas. Isso permitiu aplicar os algoritmos de descoberta, gerando os modelos ilustrados nas seções anteriores deste capítulo.

Estes modelos trouxeram à vista atividades que não foram descritas no levantamento preliminar do processo e a constatação de que a maioria dos casos não segue o fluxo definido de precedência das atividades: pedido de compra; documento de entrada e finalmente pagamento.

Outra descoberta importante está no tempo de conclusão dos processos, que apresenta uma média geral de 4 dias, mas há casos observados com duração muito acima da média, chegando a durar até 254 dias.

A realização de iterações sobre o método PM² adotado no caso estudado se mostrou uma abordagem promissora, na qual partiu-se de uma pergunta mais abrangente e conseqüentemente um *log* completo, e em seguida a cada ciclo, restrições foram aplicadas buscando descobrir e entender aspectos do processo, levando a criação de *logs* filtrados para cada situação estudada.

6 Conclusões

Esta pesquisa se propôs a explorar a mineração de processos e suas aplicações na área de suprimentos, fazendo uma revisão de casos reportados na literatura e desenvolvendo um estudo de caso na empresa Fattu que utiliza o sistema ERP Protheus. O trabalho apresenta contribuições nesta temática, uma vez que inclui estudos de casos previamente realizados na área de suprimentos, ilustra através de um estudo de caso um projeto de mineração de processo sobre o ERP Protheus e destaca aspectos que devem ser considerados ao se implementar projetos de mineração de processos. Aplicações nesta área buscam resolver problemas nada triviais, mas com resultados promissores, motivando novos estudos e aplicações.

O método aplicado no estudo de caso na empresa Fattu sobre o ERP Protheus, visa explorar o processo de compras e pagamentos, identificando os passos envolvidos, em especial as extrações, transformações e cargas que devem ser realizadas para preparar os registros de eventos, a fim de minerar os modelos de processos.

No estudo de caso pode-se colocar à prova as técnicas de mineração de processos em um sistema amplamente utilizado no Brasil e que até a conclusão desta pesquisa, não havia sido alvo de estudo e enfrentar desafios da mineração de processos, em especial:

- 1) Lidar com registros de eventos complexos e características diversas, originadas pela estrutura dos registros de eventos do ERP e a cardinalidade de muitos para muitos (N:N) entre as tabelas de transações.
- 2) Encontrar, mesclar e limpar os registros de eventos, adequando ao formato necessário para aplicação da mineração de processos.
- 3) Buscar o balanceamento entre os critérios de qualidade como cobertura, precisão, generalização e simplicidade. Por não haver outros casos aplicando sobre o ERP alvo do estudo de caso, não há valores das métricas para comparação dos resultados alcançados nesta pesquisa.

- 4) Melhorar o entendimento e usabilidade da mineração de processos para usuários não especialistas, através da aplicação de tecnologia amigável e de fácil utilização como o software Celonis aplicado no estudo de caso.

As transformações aplicadas ao *log*, método de projeto, algoritmos, técnicas e ferramentas utilizadas neste estudo de caso podem ser generalizadas para aplicação em outras organizações que utilizem o ERP Protheus, inclusive em outras áreas da organização, mediante revisão das perguntas, seleção das tabelas a auditar e cardinalidade dos relacionamentos que farão parte do modelo.

Verificou-se que a etapa mais trabalhosa e crítica do projeto é a de preparação dos *logs* para a estrutura adequada contendo a perspectiva de caso, ou seja, a criação de um identificador de caso que permeie as diversas atividades realizadas sobre cada caso. Neste ponto, outra dificuldade está relacionada à cardinalidade de relacionamento entre as tabelas adotada, que foi de muitos para muitos (N:N). Os *logs* mantidos pelo ERP Protheus possuem uma estrutura muito peculiar e de difícil adaptação para a estrutura ideal, reduzindo a qualidade dos *logs*, e assim, dos modelos minerados.

Por outro lado, pode-se identificar o processo executado no dia-a-dia da empresa estudada, duração dos casos, descobrir atividades omitidas no levantamento do processo, emular os casos para monitorar o comportamento dos casos ao longo do tempo e concluir que a maioria dos casos não segue a regra de precedência no processo de compras.

Trabalhos futuros nesta linha de pesquisa podem envolver análises sobre outras regras de negócio, como avaliação de conflitos de interesse, ou seja, verificar se há atividades sendo realizadas por pessoas sem as devidas permissões. Outra sugestão é avaliar processos de outras áreas da organização, novos algoritmos e ferramentas, além do desenvolvimento de uma ferramenta específica para geração de *logs* sobre o ERP Protheus estruturada sobre a perspectiva de caso, que facilite e dê maior agilidade ao projeto.

Agradecimentos

Este estudo contou com o apoio da Fundação de Amparo à Pesquisa do Estado de Goiás - FAPEG, contrato número: 2017 10216 7000 588.

Referências Bibliográficas

VAN DER AALST, W. M. P. Fuzzy mining - adaptive process simplification based on multi-perspective metrics. Proceedings of the 5th International Conference on Business Process Management (BPM 2007). v. 4714, p.24–28, 2007. Brisbane, Australia. Disponível em: <<http://link.springer.com/10.1007/978-3-540-75183-0>>.

VAN DER AALST, W. M. P. Using process mining to bridge the gap between BI and BPM. **Computer**, v. 44, n. 12, p. 77–80, 2011a.

VAN DER AALST, W. M. P. Intra- and inter-organizational process mining: Discovering processes within and between organizations. **Lecture Notes in Business Information Processing**, v. 92 LNBIP, p. 1–11, 2011b.

VAN DER AALST, W. M. P. Process mining. **Communications of the ACM**, v. 55, n. 8, p. 76–83, 2012.

VAN DER AALST, W. M. P. **Process Mining - Data Science in Action**. 2 edition ed. Springer, Berlin, Heidelberg, 2016.

VAN DER AALST, W. M. P.; ADRIANSYAH, A.; DE MEDEIROS, A. K. A.; et al. Process Mining Manifesto. **Lecture Notes in Business Information Processing**. v. 99 LNBIP, p.169–194, 2012. Disponível em: <http://link.springer.com/10.1007/978-3-642-28108-2_19>.

VAN DER AALST, W. M. P.; DE BEER, H. T.; VAN DONGEN, B. F. Process mining and verification of properties: An approach based on temporal logic. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 3760 LNCS, p. 130–147, 2005. Disponível em: <http://link.springer.com/chapter/10.1007/11575771_11%5Cnhttp://www.scopus.com/inward/record.url?eid=2-s2.0-33646698687&partnerID=40&md5=91c6969ab75adec763c2908480dc3c62>.

VAN DER AALST, W. M. P.; DUMAS, M.; OUYANG, C.; ROZINAT, A.; H.M.W., V. Choreography Conformance Checking : An Approach based on BPEL and Petri Nets. **Babel**, p. 1–71, 2010. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.79.4871&rep=rep1&type=pdf>>.

VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A.; WEIJTERS, A. J. M. M. Genetic Process Mining. , , n. i, p. 48–69, 2005. Disponível em: <http://link.springer.com/10.1007/11494744_5>.

VAN DER AALST, W. M. P.; REIJERS, H. A.; WEIJTERS, A. J. M. M.; et al. Business process mining: An industrial application. **Information Systems**, v. 32, n. 5, p. 713–732, 2007.

VAN DER AALST, W. M. P.; SCHONENBERG, M. H.; SONG, M. Time prediction based on process mining. **Information Systems**, v. 36, n. 2, p. 450–475, 2011. Elsevier. Disponível em: <<http://dx.doi.org/10.1016/j.is.2010.09.001>>.

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

AGRAWAL, R.; GUNOPULOS, D.; LEYMANN, F. Mining Process Models from Workflow Logs. **Sixth International Conference on Extending Database Technology**, p. 469–483, 1998.

BEZERRA, F.; WAINER, J. Fraud detection in process aware systems. **International Journal of Business Process Integration and Management**, v. 5, n. 2, p. 121–129, 2011. Disponível em: <<http://www.inderscience.com/link.php?id=40204>>.

BRAZIL, A. O. B. P. M. P. **Guia para o Gerenciamento de Processos de Negócio Corpo Comum de Conhecimento**. 1 Edição ed. Association of Business Process Management Professionals, 2013.

BUIJS, J. C. A. M. **Mapping Data Sources to XES in a Generic Way**Chelsea, 2010. Eindhoven University of Technology. Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Mapping+Data+Sources+to+XES+in+a+Generic+Way#0%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Mapping+Data+Sources+to+XES+in+a+Generic+Way%231>>.

CARON, F.; VANTHIENEN, J.; BAESENS, B. Comprehensive rule-based compliance checking and risk management with process mining. **Decision Support Systems**, v. 54, n. 3, p. 1357–1369, 2013a. Elsevier B.V. Disponível em: <<http://dx.doi.org/10.1016/j.dss.2012.12.012>>.

CARON, F.; VANTHIENEN, J.; BAESENS, B. Business rule patterns and their application to process analytics. **Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC**, p. 13–20, 2013b.

CAUCHICK MIGUEL, P. A.; FLEURY, A.; PEREIRA MELLO, C. H.; et al. **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. 3 Edição ed. Rio de Janeiro: Elsevier, 2018.

COOPER, M. C.; LAMBERT, D. M.; PAGH, J. D. Supply Chain Management: More Than a New Name for Logistics. **The International Journal of Logistics Management**, v. 8, n. 1, p. 1–14, 1997.

VAN ECK, M. L.; LU, X.; LEEMANS, S. J. J.; VAN DER AALST, W. M. P. PM2: A process mining project methodology. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 9097, p. 297–313, 2015.

ENGEL, R.; KRATHU, W.; ZAPLETAL, M.; et al. Analyzing inter-organizational business processes: Process mining and business performance analysis using electronic data interchange messages. **Information Systems and e-Business Management**, v. 14, n. 3, p. 577–612, 2016. Springer Berlin Heidelberg.

ER, M.; ARSAD, N.; ASTUTI, H. M.; KUSUMAWARDANI, R. P.; UTAMI, R. A. Analysis of production planning in a global manufacturing company with process mining. **Journal of Enterprise Information Management**, v. 31, n. 2, p. 317–337, 2018. Disponível em: <<http://www.emeraldinsight.com/doi/10.1108/09574090910954864>>.

ER, M.; ASTUTI, H. M. A Case Study on Process Mining Implementation in Modelling Supply Chain Business Process: a Lesson Learnt. 6th International Conference on Operations and Supply Chain Management, Bali, 2014. p.808–819, 2014.

ER, M.; ASTUTI, H. M.; PRAMITASARI, D. Modeling and Analysis of Incoming Raw Materials Business Process: A Process Mining Approach. **International Journal of Computer and Communication Engineering**, v. 4, n. 3, p. 196–203, 2015. Disponível em: <<http://www.ijcce.org/index.php?m=content&c=index&a=show&catid=51&id=473>>.

GANDULFO, P. I.; RALHA, C. G. **Método de Mineração de Processos para Auxílio à Tomada de Decisão: um Estudo de Caso no controle de Férias**, 2016. Universidade de Brasília.

GERKE, K.; CLAUS, A.; MENDLING, J. Process mining of RFID-based supply chains.

2009 IEEE Conference on Commerce and Enterprise Computing, CEC 2009, p. 285–292, 2009.

IIBA. **A Guide to the Business Analysis Body of Knowledge (BABOK® Guide) – Version 3.0**. International Institute of Business Analysis, 2015a.

IIBA. **A Guide to the Business Analysis Body of Knowledge (BABOK® Guide) – Version 3.0**. 2015b.

JANS, M.; ALLES, M.; VASARHELYI, M. The case for process mining in auditing: Sources of value added and areas of application. **International Journal of Accounting Information Systems**, v. 14, n. 1, p. 1–20, 2013. Elsevier Inc. Disponível em: <<http://dx.doi.org/10.1016/j.accinf.2012.06.015>>.

JANS, M.; VAN DER WERF, J. M.; LYBAERT, N.; VANHOOF, K. A business process mining application for internal transaction fraud mitigation. **Expert Systems with Applications**, v. 38, n. 10, p. 13351–13359, 2011. Elsevier Ltd. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2011.04.159>>.

JUNIOR, G. T.; FEDERSON, F. M. **Modelagem de Sistemas de Informação para Mineração de Processos : Características e Propriedades das Linguagens**, 2017. Universidade Federal de Goiás.

KALENKOVA, A. A.; VAN DER AALST, W. M. P.; LOMAZOVA, I. A.; RUBIN, V. A. Process mining using BPMN: relating event logs and process models. **Software and Systems Modeling**, 2015. Springer Berlin Heidelberg.

KERREMANS, M. **Market Guide for Process Mining**. 2018.

LAU, H. C. W.; HO, G. T. S.; ZHAO, Y.; CHUNG, N. S. H. Development of a process mining system for supporting knowledge discovery in a supply chain network. **International Journal of Production Economics**, v. 122, n. 1, p. 176–187, 2009. Elsevier. Disponível em: <<http://dx.doi.org/10.1016/j.ijpe.2009.05.014>>.

LEEMANS, S. J. J.; FAHLAND, D.; VAN DER AALST, W. M. P. Discovering block-structured process models from event logs. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 7927 LNCS, p. 311–329, 2013a.

LEEMANS, S. J. J.; FAHLAND, D.; VAN DER AALST, W. M. P. Discovering block-structured process models from event logs - A constructive approach. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 7927 LNCS, p. 311–329, 2013b.

DE LEONI, M.; VAN DER AALST, W. M. P.; DEES, M. A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. **Information Systems**, v. 56, p. 235–257, 2016. Elsevier. Disponível em: <<http://dx.doi.org/10.1016/j.is.2015.07.003>>.

LOPES, N. C. DE S. **MODELO DE GESTÃO POR PROCESSOS BASEADO EM MINERAÇÃO**, 2017. Universidade Federal do Rio de Janeiro.

MAHENDRAWATHI, E. R.; ZAYIN, S. O.; PAMUNGKAS, F. J. ERP Post Implementation Review with Process Mining: A Case of Procurement Process. **Procedia Computer Science**, v. 124, p. 216–223, 2017. Elsevier B.V. Disponível em: <<https://doi.org/10.1016/j.procs.2017.12.149>> .

MARUSTER, L.; WORTMANN, J. C. H.; WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P. Discovering Distributed Processes in Supply Chains. **Proc. of the International Conference on Advanced Production Management Systems**, p. 119–128, 2002.

DE MEDEIROS, A. K. A.; VAN DER AALST, W. M. P.; WEIJTERS, A. J. M. M. **Genetic Process Mining**, 2005. Disponível em: <http://link.springer.com/10.1007/11494744_5>.

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

MUNOZ-GAMA, J. **Conformance checking and diagnosis in process mining**, 2014. Universitat Politècnica de Catalunya - BarcelonaTech. Disponível em: <<http://www.tdx.cesca.cat/handle/10803/284964>>.

PIESSENS, D. A. M. **Event Log Extraction from SAP ECC 6.0** Event Log Extraction from SAP ECC 6.0, 2011. Eindhoven University of Technology.

PORTER, M. E. Changing patterns of international competition. **California Management Review**, v. 28, n. 2, p. 13–14, 1986. Disponível em: <http://books.google.com/books?hl=en&lr=&id=DJjiOUGUg2kC&oi=fnd&pg=PA49&dq=Changing+Patterns+of+International+Competition&ots=h4Ugiy059w&sig=TnsDf-EdF_Mn0BQQ3BvqUQNYyiM%5Cnhttp://books.google.com/books?hl=en&lr=&id=DJjiOUGUg2kC&oi=fnd&pg=PA49&dq=Changing+p>.

PORTER, M. E.; LINDE, C. VAN DER. Toward a New Conception of the Environment-Competitiveness Relationship. **Journal of Economic Perspectives**, v. 9, n. 4, p. 97–118, 1995.

ROZINAT, A.; VAN DER AALST, W. M. P. Conformance checking of processes based on monitoring real behavior. **Information Systems**, v. 33, n. 1, p. 64–95, 2008.

ROZINAT, A.; DE MEDEIROS, A. K. A.; GÜNTHER, C. W.; WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P. Towards an Evaluation Framework for Process Mining Algorithms. , p. 1–20, 2007. Disponível em: <<http://is.tm.tue.nl/staff/aweijters/WP224.pdf>>.

SCHWAICKARDT, E.; DANTAS, M. J. P. Process Mining Applied in Supply Management Processes. **European Journal of Scientific Research**, v. 151, n. 2, p. 160–171, 2018.

SEGRS, I. E. A. **Investigating the application of process mining for auditing purposes**, 2007. Eindhoven University of Technology.

STOOP, J. J. **Process Mining and Fraud Detection**, 2012. Twente University.

TIWARI, A.; TURNER, C. J.; MAJEED, B. A review of business process mining: state-of-the-art and future trends. **Business Process Management Journal**, v. 14, n. 1, p. 5–22, 2008. Disponível em: <<http://www.emeraldinsight.com/doi/10.1108/14637150810849373>>.

TOTVS. Divulgação de Resultados 2T2018. Disponível em: <<http://ri.totvs.com/ptb/3242/634877.pdf>>.

WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A. Process Mining with the Heuristics Miner Algorithm. **Technische Universiteit Eindhoven, Tech. Rep. WP**, v. 166, n. January, p. 1–34, 2006. Disponível em: <http://cms.ieis.tue.nl/Beta/Files/WorkingPapers/Beta_wp166.pdf>.

WFMC. Workflow Management Coalition Terminology & Glossary. **ISA Transactions**, v. 39, n. 2, p. 153–167, 1999.

YIN, R. K. **Estudo de Caso: Planejamento e Métodos**. 2 Edição ed. Porto Alegre: Bookman, 2001.

Anexo 1 – Termo de autorização para o estudo

Goiânia, 28 de agosto de 2018.

Prezado Senhor,


Apresentamos o acadêmico Ederson Schwaickardt, do curso de Mestrado em Engenharia de Produção e Sistemas PROPE/MEPROS, da PUC Goiás, matrícula 20171210400050, para desenvolver seu trabalho dissertativo “Mineração de Processos Aplicada na Gestão de Suprimentos: um Estudo de caso com ERP Brasileiro”, sob a orientação da Professora Dra. Maria José Pereira Dantas.

Esta pesquisa têm o objetivo de desenvolver um estudo de caso aplicando a mineração de processos na área de gestão de suprimentos em um ERP de relevância nacional, o sistema Protheus da empresa Totvs. Um método de mineração de processos que contemple as melhores práticas será aplicado sobre a estrutura do Protheus, descrevendo os passos envolvidos na preparação dos *logs*, aplicação da mineração de processos, análises conduzidas, e relatando ferramentas, algoritmos e técnicas utilizadas, e finalmente os resultados alcançados neste processo investigativo.

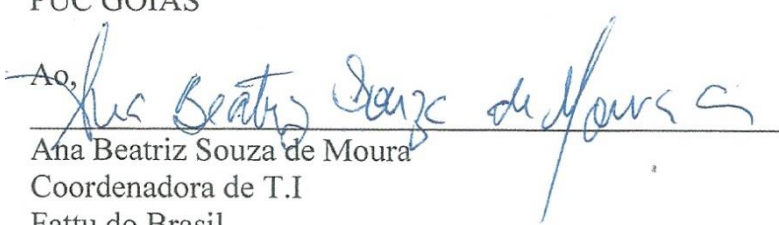
Desta forma, solicitamos a autorização para que o referido aluno possa pesquisar os dados necessários nos registros de *logs* do sistema ERP. Ressaltamos, que os dados pesquisados têm finalidade acadêmica e serão utilizados somente no trabalho dissertativos e na produção de artigos científicos. Assumimos o compromisso que os dados somente serão utilizados para fins acadêmicos.

Sem mais para o momento desde já agradecemos a atenção dispensada.

Atenciosamente,



Professor Dr. Marcos Lajovic Carneiro
Coordenador – PROPE/MEPROS
PUC GOIÁS

Ao, 

Ana Beatriz Souza de Moura
Coordenadora de T.I
Fattu do Brasil