

**GILBERTO ALECES DOS SANTOS**

**MÉTODO- $\beta$ : UM MÉTODO ITERATIVO PARA A  
REDUÇÃO DO *LEAD-TIME* DE MODELOS DE  
PROCESSOS DE NEGÓCIOS**

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

**CURITIBA**

**2020**

**GILBERTO ALECES DOS SANTOS**

**MÉTODO- $\beta$ : UM MÉTODO ITERATIVO PARA A  
REDUÇÃO DO *LEAD-TIME* DE MODELOS DE  
PROCESOS DE NEGÓCIOS**

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

Área de Concentração: *Ciência da  
Computação*

Orientador: Prof. Dr. Edson Emílio Scalabrin

**CURITIBA**

**2020**

Dedico esse trabalho aos meus filhos, Mateus e Maria, pela inspiração de cada dia e a minha esposa Neuza, pelo apoio e incentivo de sempre.

## **Agradecimentos**

Agradeço primeiramente ao nosso grandioso Deus, por me proporcionar o dom da vida e pela oportunidade de estar realizando este trabalho.

Aos meus pais Lauro e Adelair por todo o esforço que fizeram para que eu pudesse ter uma educação base e por todo apoio e incentivo.

A minha esposa Neuza por todo carinho, compreensão e incentivo e por ser essa mulher guerreira, sempre preocupada em dar o melhor de si em função da nossa família.

Aos meus filhos Mateus e Maria por serem esses anjos enviados por Deus para dar inspiração e força.

Ao meu orientador Edson Emílio Scalabrin que nunca mediu esforços para me guiar em todo o desenvolvimento deste trabalho.

Ao meu amigo Alex Meinheim responsável pelo início desta trajetória e por todas suas contribuições e direcionamento ao longo dela.

À UpFlux que me proporcionou suporte base para o desenvolvimento deste projeto, e à seus responsáveis por todo apoio.

À PUCPR, ao Programa de Pós-Graduação em Informática (PPGIa) e a Fundação CAPES pela oportunidade e suporte oferecido.

Por fim, agradeço a todas as pessoas que direta ou indiretamente contribuíram para a execução deste trabalho.

## Resumo

A mineração de processo permite realizar a descoberta automatizada de modelos de processos a partir de logs de eventos. Esses modelos fornecem informações que permitem realizar vários tipos de análise baseada em modelos, dentre as quais destacam-se: verificação de conformidade, análise de desempenho e identificação de gargalos nos processos negócios. A identificação de gargalos é altamente relevante para os gestores de processos sugerirem melhorias. Aqui, a melhoria que nos interessa é a redução do *lead-time*. Uma dupla hipótese foi feita. A primeira é que a redução do *lead-time* passa pela identificação e remoção de gargalos em processos de negócio. A segunda é que a intervenção focada no caminho crítico permite reduzir o *lead-time* provocando pouca alteração no processo corrente. A partir da execução da tarefa de descoberta de modelo de processo pode-se obter informações básicas do processo de negócio, tais como: duração mínima, média e máxima de cada atividade. E tais informações permitem aplicar a técnica PERT/CPM para calcular a média ponderada do tempo de execução esperado de cada atividade e o caminho crítico. Este último dá a direção para identificar a atividade relevante *vis-à-vis* o foco de interesse de análise; i.e., o gargalo no caminho crítico do processo de negócio. Os experimentos realizados mostraram que o método desenvolvido permite reduzir o *lead-time* de um processo e indicar o grau de comprometimento de recursos em cada iteração do método, em termos de similaridade entre modelo otimizado e modelo original.

**Palavras-Chave:** Mineração de processo; Caminho crítico; *Lead-time*; Otimização.

## Abstract

Process mining allows for automated discovery of process models from event logs. These models provide information that allows you to perform various types of analysis based on models, among which stand out: compliance verification, performance analysis and identification of bottlenecks in business processes. Bottleneck identification is highly relevant for process managers to suggest improvements. Here, the improvement that interests us is the reduction of delivery time. A double hypothesis has been made. The first is that the reduction in lead time involves the identification and removal of bottlenecks in business processes. The second is that the intervention focused on the critical path reduces delivery time, causing little change in the current process. Process mining allows, through the task of discovering the process model, to obtain basic statistical information of a business process, such as: minimum, average and maximum duration of each activity. The PERT/CPM technique allows to calculate the weighted average of the expected execution time for each activity and the critical path. The latter guides the identification of the relevant activity in relation to the focus of the analysis's interest; that is, the bottleneck in the critical path of a business process. The experiments show that the method allows to reduce the lead time of a process and to indicate the degree of commitment of resources in each iteration of the method, in terms of similarity between the optimized model and the original model.

**Keywords:** Process mining; Quality metrics; Lead-time; Optimization.

## Lista de Figuras

Figura 1 – Estrutura básica de um processo de negócio.....	18
Figura 2 – Componentes da mineração de processo.....	19
Figura 3 - Mineração de processos: descoberta, verificação de conformidade, extensão. Fonte [1].....	21
Figura 4 - Modelo de processo descoberto com o <i>Alpha Miner</i> . Fonte: [6].....	22
Figura 5. Modelo de processo descoberto pela ferramenta UpFlux.....	24
Figura 6. Representação da rede AoN com uma ligação tecnológica entre as atividades i e j. Fonte [7] .....	35
Figura 7. Rede de projeto AoN. Fonte [7].....	36
Figura 8 - Uma rede de exemplo AoN da Tabela 4. Fonte [7] .....	37
Figura 9. Distribuição <i>beta</i> utilizada para expressar a variabilidade da duração da atividade em PERT. Fonte [7].....	42
Figura 10. <i>Lead-time</i> físico e financeiro. Fonte [16].....	43
Figura 11 - Modelo de processo e caminho crítico em destaque.....	51
Figura 12 - Estrutura do método.....	58
Figura 13 – Redução do <i>lead-time</i> e similaridade entre modelos – log de eventos artificiais. 64	
Figura 14 – Redução do <i>lead-time</i> e similaridade entre modelos – casos de sepsis. ....	66
Figura 15 – Redução do <i>lead-time</i> e similaridade entre modelos – faturamento hospitalar. ...	68
Figura 16 – Redução do <i>lead-time</i> e similaridade entre modelos – log de produção.....	70
Figura 17 – Redução do <i>lead-time</i> e similaridade entre modelos – todas as bases. ....	71
Figura 18 – Desvio padrão da otimização e similaridade – estabilidade.....	72

## Lista de Tabelas

Tabela 1 -Fragmento de alguns logs de eventos: cada linha corresponde a um evento. Fonte [1].....	20
Tabela 2 - Abordagens de otimização de modelos de processos de negócio. Fonte [5] .....	29
Tabela 3. Abordagens de otimização para modelos de processos de negócios. Fonte [5] .....	33
Tabela 4. Lista de atividades com seus antecessores. Fonte [7].....	35
Tabela 5. Um exemplo de projeto fictício com 12 atividades não fictícias. Fonte [7].....	37
Tabela 6 - Enumeração de todos os caminhos possíveis do projeto da Tabela 4: Fonte [7] ...	38
Tabela 7. Folgas das atividades de projeto de exemplo da Figura 6. Fonte [7] .....	40
Tabela 8. Três estimativas de tempo para as atividades de projeto. Fonte [7] .....	41
Tabela 9. Log de eventos. ....	48
Tabela 10. Tempos das atividades do modelo de processo descoberto. ....	50
Tabela 11. Lista de atividades e seus tempos de execução médio. ....	50
Tabela 12. Lead-time de cada caminho no modelo de processo .....	50
Tabela 13. Notações para os cálculos de otimização.....	51
Tabela 14. Cálculos das atividades C e E.....	52
Tabela 15. Cálculos das atividades F e H.....	52
Tabela 16. Etapas do experimento.....	57
Tabela 17. Base de dados. ....	60
Tabela 18. Resultados da otimização do <i>lead-time</i> – log de eventos artificiais. ....	63
Tabela 19. Resultados da otimização do <i>lead-time</i> – casos de sepse. ....	65
Tabela 20. Resultados da otimização do <i>lead-time</i> – casos de faturamento médico-hospitalar. .....	67
Tabela 21. Resultados da otimização do <i>lead-time</i> – casos de produção. ....	69

## Lista de Abreviaturas e Siglas

PM	<i>Process mining</i>
PERT	<i>Program Evaluation and Review Technique</i>
CPM	<i>Critical Path Method</i>
CSV	<i>Comma-separated values</i>
XES	<i>X-Genics EManager Skins</i>
MXML	<i>Macromedia eXtensible Markup Language</i>

# Sumário

1.1	Motivação .....	14
1.2	Objetivo.....	14
1.3	Hipóteses.....	15
1.4	Contribuição.....	15
1.5	Escopo.....	15
1.6	Organização do Trabalho .....	16
2.1	Processo .....	17
2.1.1	Processo de Negócio .....	17
2.2	Mineração de Processos .....	19
2.2.1	Log de Eventos .....	20
2.2.2	Tipos de Mineração de Processos .....	20
2.2.3	Algoritmos de Descoberta de Processos .....	22
2.2.4	Ferramentas.....	23
2.3	Medida de Similaridade .....	24
2.3.1	Medidas baseadas na correspondência entre nós e arestas .....	25
2.3.2	Medidas baseadas na distância de edição de grafos.....	26
2.3.3	Medidas de dependências causais entre atividades.....	27
2.4	Otimização de Processos de Negócios.....	28
2.5	Caminho Crítico .....	33
2.5.1	Cálculo do caminho crítico .....	34
2.5.2	Avaliação de programação e técnica de revisão .....	40

2.6	Lead-time .....	43
2.7	Trabalhos Relacionados .....	43
2.7.1	Estimativa de tempo.....	43
2.7.2	Simulação e melhoria de processo .....	44
2.8	Considerações .....	45
3.1	Método- $\beta$ .....	47
3.2	Logs de Eventos .....	48
3.3	Descoberta de modelo de processos .....	49
3.4	Identificar e calcular caminho crítico e gargalos .....	50
3.5	Executar a otimização .....	53
3.6	Considerações do Capítulo .....	54
4.1	Método de Pesquisa .....	55
4.2	Hipótese de Pesquisa.....	56
4.3	Estratégia de Pesquisa.....	56
4.3.1	Medidas de similaridade e cálculo de eficiência de redução do <i>lead-time</i> .....	57
4.3.2	Identificação e seleção das técnicas para calcular o caminho crítico e gargalos .....	57
4.3.3	Desenvolvimento do método e construção do ambiente experimental.....	58
4.3.4	Coleta de dados .....	60
5.1	B1: Log artificial.....	62
5.2	B2: Casos de <i>Sepses</i> .....	65
5.3	B3: Faturamento Hospitalar .....	67
5.4	B4: Log de Produção .....	68

5.5	Discussão dos Resultados .....	71
-----	--------------------------------	----

# 1 Introdução

A mineração de processo permite descobrir automaticamente modelos de processos de negócio a partir de *logs* de eventos. Esses modelos fornecem informações que permitem realizar vários tipos de análise baseada em modelo [1]. A mineração de processo congrega basicamente três tarefas significativas, em *prol* do gestor de processos de negócio, na direção da análise da realidade de um processo de negócio com esforços reduzidos. As tarefas são: i) descobrir modelos de processos reais sem conhecimento *a priori* ou com esforços reduzidos; ii) verificar a conformidade do modelo descoberto *vis-à-vis* ao modelo de processo teórico ou de referência; iii) estender modelo de processo com informações de desempenho, gargalos, recursos e custos [2]. Esta última tarefa, apesar de fornecer informações sobre gargalos deixa uma lacuna não preenchida por um método claro e computacional que facilite à análise do gestor de processos de negócio face ao caminho crítico, gargalos reais de um processo de negócio, assim como, geração de alternativas com baixo esforço que reduzam gargalos e consequentemente o *lead-time*.

A nossa primeira aproximação para gerar uma extensão de processo com menos gargalos e menor *lead-time* apoia-se em um método clássico, em que dado um modelo de processo—rede de atividades—tal aproximação consiste, primeiramente, calcular o caminho crítico  $P$  e, em segundo lugar, determinar a atividade crítica  $A$  sobre o caminho crítico  $P$ , aumentar a capacidade da atividade  $A$ , e consequentemente, recalculando o *lead-time* do processo. Neste contexto, a proposta encerra a definição de um método computacional e interativo, denominado Método- $\beta$ , capaz de utilizar informações geradas pelas tarefas de mineração de processos para extrair, a partir de estatísticas básicas sobre logs de eventos, o caminho crítico do processo e a média ponderada do tempo de execução de cada atividade; a técnica do caminho crítico foi inspirada em PERT/CPM [7]. Em seguida, o próximo passo do Método- $\beta$ , consiste em identificar a atividade com gargalo, levando em consideração o tempo de execução médio ponderado de cada atividade do caminho crítico, o tempo médio de espera e a frequência absoluta.

## 1.1 Motivação

Os modelos descobertos pela mineração de processo trazem conhecimentos que permitem gerar diagnóstico de um processo de negócio [2], o que auxilia na tomada de decisão. Porém, muitas vezes o gestor de processos não dispõe de ferramental/recurso para identificar os possíveis gargalos de um processo de negócio e avaliar o quanto eles podem impactar no *lead-time*; deve-se salientar que, mesmo depois de identificá-los, o custo em termos de esforços e sobrecarga cognitiva para derivar modelos alternativos para propor uma melhoria pode ser significativo.

O desafio está em conceber e colocar em prática um método computacional eficiente que identifique de forma automática o caminho crítico em um processo de negócio e seus possíveis gargalos e aplique estratégias eficientes—orientada a caminho crítico—para buscar reduzir o *lead-time* em modelos de processos de negócios, previamente descobertos usando técnicas de mineração de processo; tal desafio encerra nossa motivação para a realização deste trabalho de pesquisa. Deve-se salientar que a redução de *lead-time* pode levar a redução de custos.

## 1.2 Objetivo

Dada a motivação e o desafio supra exposto, o objetivo principal da pesquisa está em conceber e avaliar um método computacional que permita gerar alternativas sobre um modelo de processo de negócio visando reduzir o seu *lead-time*, considerando que o modelo de processo foi descoberto de forma automática por meio da mineração de processos. Para a consecução desse propósito foram definidos três objetivos específicos, que são:

- Realizar um estudo da literatura para identificar métodos computacionais que automatizem a identificação e proposição de alternativas operativas para reduzir atrasos ou *lead-time*.
- Definir um método computacional para identificar e propor alternativas eficientes de redução do *lead-time* de processo de negócio.
- Construir um ambiente para experimentar e avaliar o desempenho do método proposto *vis-à-vis* a diferentes cenários de teste—definidos por diferentes logs de eventos de processos de negócios.

Partindo da premissa que a redução do *lead-time* passa pela identificação, análise e remoção de gargalos, a análise de desempenho será medida em função da redução efetiva do *lead-time* de um processo de negócio. Assume-se que um bom desempenho é quando a redução do *lead-time* do processo alvo é alcançada **com o menor número de modificações no modelo de processo original**. O aumento de recurso para execução de uma atividade é um exemplo de modificação.

### 1.3 Hipóteses

As hipóteses desta pesquisa são:

- H1. A extração de informações de logs de eventos de processos de negócios semiestruturados, por meio da tarefa de descoberta de modelos de instâncias de processos, é suficiente para alimentar um método computacional focado na redução do *lead-time*, com baixo esforço e sobrecarga cognitiva.
- H2. As informações geradas pela técnica de PERT/CPM—média ponderada de cada atividade e caminho crítico—permitem gerar uma estratégia eficiente para identificar atividades significativas/impactantes em um processo de negócio.

### 1.4 Contribuição

Concepção e desenvolvimento de um método computacional capaz de identificar gargalos e sugerir alternativas para reduzir atrasos—ou *lead-time*—em modelos de processos de negócios descobertos por meio de técnicas de mineração de processos.

### 1.5 Escopo

O escopo deste trabalho se limita a: identificar o caminho crítico e os gargalos em modelos de processos de negócio; usar apenas as informações extraídas dos logs de eventos e das saídas das tarefas de mineração de processos; analisar a eficiência do método proposto para reduzir *lead-time* apenas em função do número de movimentações de ajustes no modelo de processo original—assume-se que quanto menor o número de movimentações, melhor; e gerar modelos de processos com as mudanças sugeridas pelo método proposto **sem considerar os custos dos recursos** envolvidos em cada atividade.

## **1.6 Organização do Trabalho**

Este trabalho está organizado em sete capítulos. O Capítulo 2 aborda a fundamentação teórica. O Capítulo 3 apresenta os trabalhos relacionados. O Capítulo 4 descreve o método concebido e desenvolvido para auxiliar o analista de processo de negócio quanto a redução de *lead-time*. O Capítulo 5 apresenta o método de pesquisa empregado, detalhando as etapas: experimentos; coleta de dados; avaliação do método/resultados. O Capítulo 6 apresenta a discussão dos resultados obtidos por meio dos experimentos realizados. O Capítulo 7 apresenta as conclusões do trabalho realizado e trabalhos futuros.

## 2 Fundamentação Teórica

Neste capítulo são apresentadas as principais definições e conceitos que fundamentam o esforço de pesquisa. São abordados conceitos referentes a processos, modelos de processos, mineração de processos e seus tipos, log de eventos, métodos de descoberta de modelo processo, bem como as técnicas e as ferramentas utilizadas para o desenvolvimento do projeto.

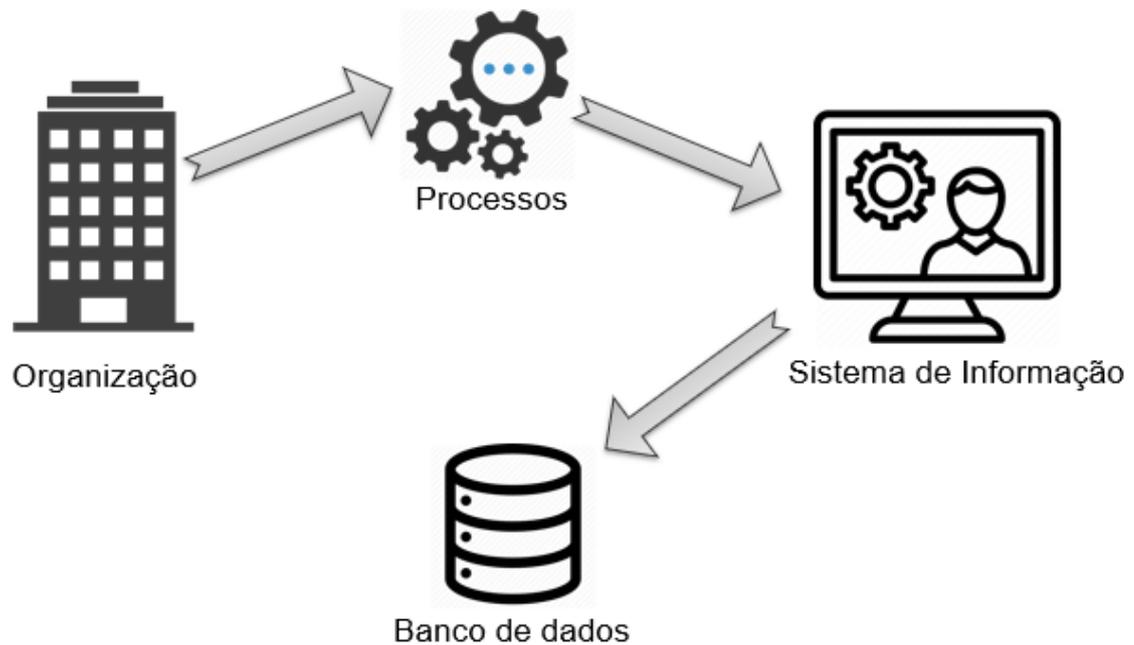
### 2.1 Processo

Um processo pode ser visto como uma ordenação específica de atividades de trabalho definidas no tempo e no espaço, contendo um ponto de início e um ponto de fim, assim como entradas e saídas claramente identificadas [28]. Todo processo pode ser representado por um modelo.

#### 2.1.1 Processo de Negócio

Em Havey [28], propõe-se uma definição simples de processo de negócio, em que um processo é definido como um conjunto de “regras específicas para passo a passo para a resolução de um problema de negócio. ” Desde a década de 1990, quando surgiram as primeiras definições de processos tentava-se focar em processos negócios em direções específicas. No entanto, em quase todas as referências nesta área, duas definições específicas de processos de negócios são recorrentemente citadas. A primeira definição foi dada por Hammer e Champy [23]. Eles afirmam que “um processo de negócio é uma coleção de atividades que toma um ou mais tipos de insumos e cria uma saída com valor para o cliente”. A segunda definição é Davenport [14], em que ele afirma que “um processo de negócio é definido como a cadeia de atividades cujo objetivo final é a produção de um resultado específico para um determinado cliente ou mercado”.

Um processo é descrito por um modelo de processo. E os modelos de processos são amplamente utilizados como uma fonte de informações e procedimentos para auxiliar no gerenciamento da complexidade organizacional [6]. Diversas notações são utilizadas para representar um modelo de processo, como Redes de Petri, BPMN, YAML, EPC [18].



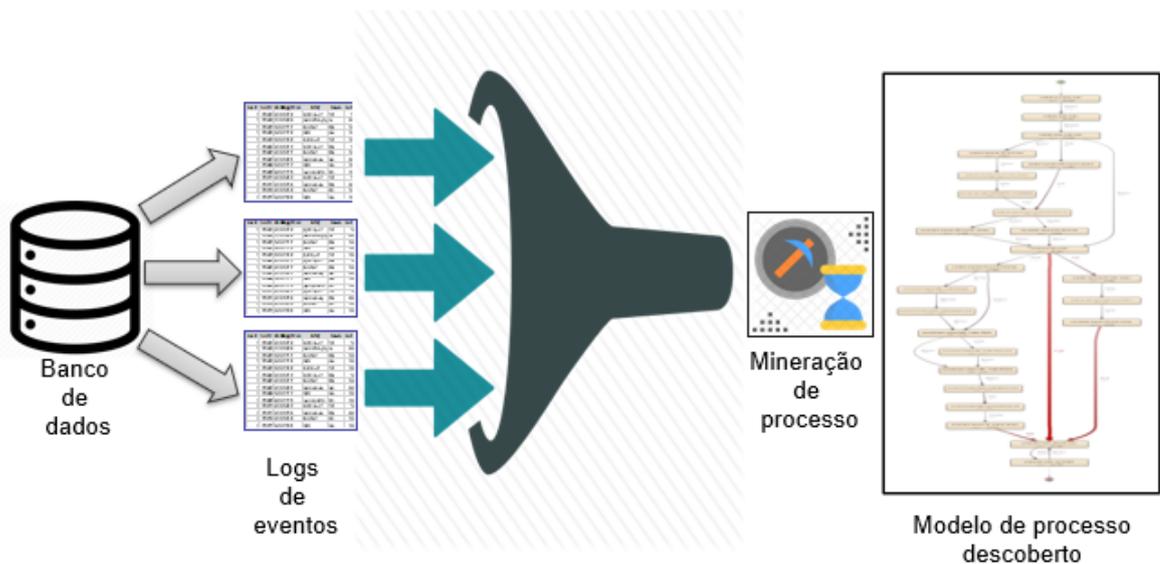
**Figura 1 – Estrutura básica de um processo de negócio.**

Diferentes níveis de graus de liberdade podem ser atribuídos aos executores das atividades de um processo de negócio, dando origem a processos reais que podem ser bem-estruturados, semiestruturados ou não-estruturados. Tais processos em sua maioria são regidos se utilizando de um sistema de informação responsável por armazenar cada traço de eventos que ocorre em um determinado processo (cf. Figura 1). Os traços dos processos semiestruturados são os alvos que despertam maiores interesses *vis-à-vis* a aplicação da mineração de processos, pois eles representam potenciais significativos para a descoberta de conhecimentos relevantes em negócio. Deve-se notar que a mineração de processos pouco contribui quando toma como entrada traços de processos bem-estruturados, pois o potencial de descobrir conhecimentos relevantes é reduzido, por outro lado, o esforço de preparação de logs de eventos de um processo não-estruturado pode inviabilizar a tarefa, em particular, no curto prazo.

## 2.2 Mineração de Processos

A mineração de processo é uma disciplina emergente que fornece um conjunto de ferramentas baseadas em fatos para apoiar a melhoria de processos [6]. Além disso, ela também permite monitorar a execução de um processo de negócio com base em dados de eventos reais e propor métodos e ferramentas para fornecer agilidade a tarefa de diagnóstico [6], reduzindo a lacuna entre praticado e projetado—modelo conceitual. A mineração de processos pode ser utilizada para cumprir as seguintes tarefas [6]:

- 1 descobrir modelos de processos a partir de dados reais sem conhecimento *a priori*;
- 2 verificar a conformidade de um modelo de processo teórico *vis-à-vis* a um modelo descoberto;
- 3 estender o modelo de processo de negócio com informações de desempenho, gargalos, recursos e custos visando fazer emergir novas perspectivas de análises.



**Figura 2 – Componentes da mineração de processo.**

A Figura 2 apresenta componentes básicos que compõem a estrutura da mineração de processos. Com base em logs de eventos armazenados em uma base de dados e por meio de algoritmos específicos, realiza-se a tarefa de descoberta de modelo de processo.

### 2.2.1 Log de Eventos

Um *log* de eventos é um registo histórico do fluxo de informações e das interações entre colaboradores e recursos para executar um conjunto de atividades no âmbito de um processo de produção [4]. Há uma multiplicidade de ferramentas que geram e mantêm estes registos. Entretanto, faz-se necessário manter tais registos em sistemas de informação seguindo uma estrutura padronizada para facilitar a análise de um processo de negócio, pois torna a tarefa de análise bem mais efetiva no curto prazo e com sobrecarga cognitiva reduzida.

Para executar a tarefa de descoberta de modelos da mineração de processos, cada evento registo em log deve ter uma estrutura canônica mínima, a saber: identificação de caso, atividade executada e alguma ordem de execução, podendo ainda, em algumas situações, também existir o registo do início e fim da atividade [2].

**Tabela 1 -Fragmento de alguns logs de eventos: cada linha corresponde a um evento.  
Fonte [1]**

Case Id	Event id	Properties				
		Timestamp	Activity	Resource	Cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	...
...	...	...	...	...	...	...

A Tabela 1 apresenta as informações comuns presentes em um *log* de eventos.

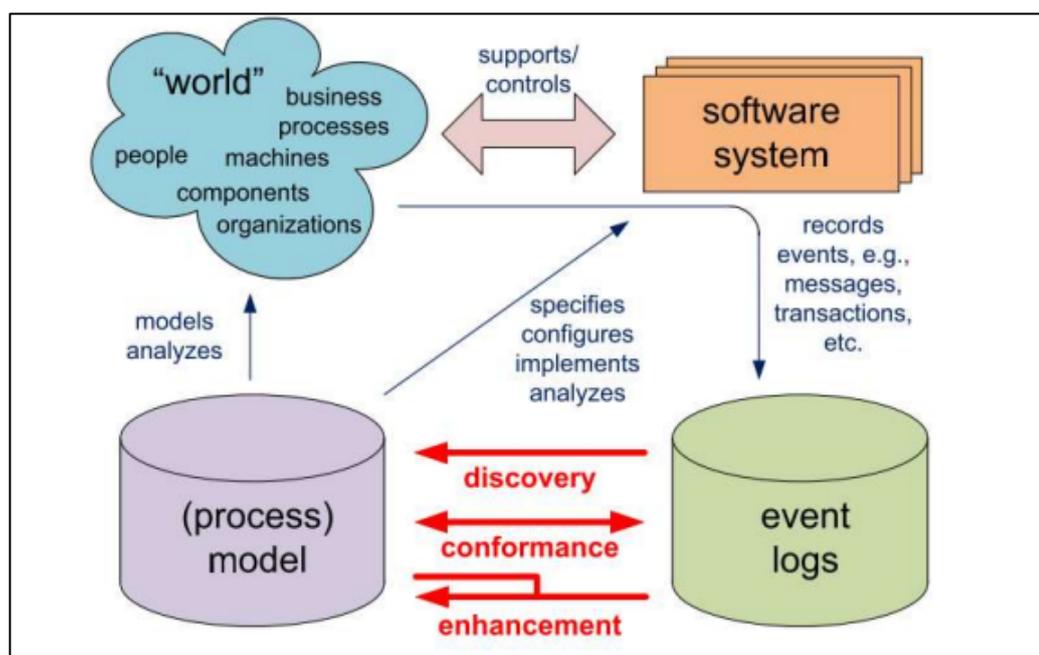
### 2.2.2 Tipos de Mineração de Processos

Existem três tipos de mineração de processo, sendo eles [1]:

- 1 Descoberta: toma-se como ponto de partida um log de eventos e deriva-se desse log um modelo de processo. Tais eventos podem ser de diversos tipos ou ter diferentes origens, por exemplo, as transações de um negócio registradas em um sistema de informação, bem como os históricos de utilização de recursos, jornada de vida de pacientes, ciclo/jornada de vida de equipamentos.

- 2 Verificação de conformidade: toma-se como ponto de partida um modelo de processo de negócio e os eventos e/ou traços de eventos deste mesmo processo e os compara. O objetivo é a verificação de possíveis desvios entre o processo teórico—de referência e o processo que ocorreu de fato no mundo real.
- 3 Extensão: toma-se como ponto de partida um modelo de processo descoberto e o enriquece com as informações dos eventos observados. A extensão visa compreender e melhorar um modelo de processo de negócio e não apenas verificar sua conformidade.

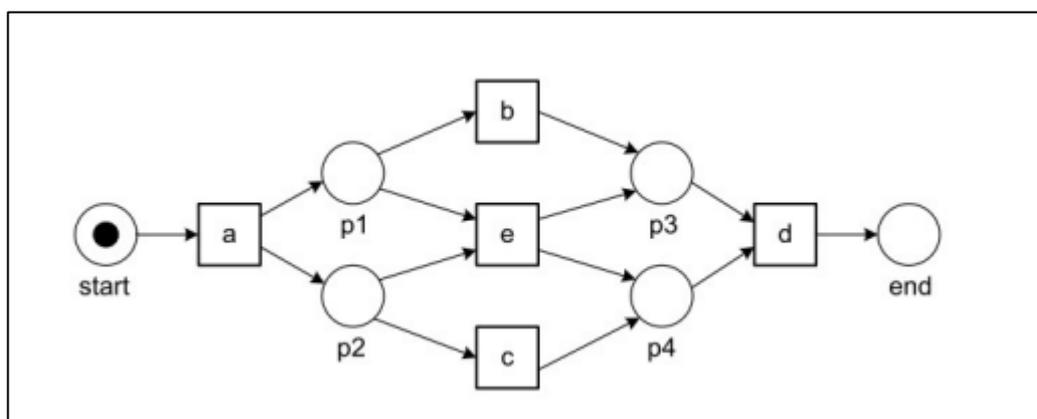
A Figura 3 ilustra esses tipos de mineração de processos de negócios, colocando em sinergia: sistema de informação, log de eventos, modelo de processo, controle e interação com o ambiente de negócio.



**Figura 3 - Mineração de processos: descoberta, verificação de conformidade, extensão. Fonte [1]**

A tarefa de descoberta é responsável por produzir um modelo de processo sem utilizar qualquer tipo de informação prévia do processo [6]. Ao longo dos anos diversas abordagens foram propostas com essa finalidade, sendo o *Alpha Miner* a abordagem precursora. O

resultado obtido da aplicação do *Alpha Miner* é um modelo de processo representado por uma Rede de Petri (cf. Figura 4).



**Figura 4 - Modelo de processo descoberto com o *Alpha Miner*. Fonte: [6].**

A análise de conformidade baseada em modelos visa comparar o modelo de processo de referência com o modelo descoberto a partir de um log de eventos. Essa iniciativa pode ser utilizada para verificar se o processo real, registrado em log, está em conformidade com o modelo previamente estabelecido e *vice-versa*. Desta forma, pode-se detectar, localizar e explicar desvios em um processo, bem como mensurar a importância e gravidade dos mesmos [6].

A extensão é o terceiro tipo de mineração de processos. A extensão visa melhorar e enriquecer um modelo de processo a partir de informações provenientes do processo em operação, que podem ser extraídas de logs de eventos. Um exemplo de enriquecimento é a combinação de um modelo de processo com dados de desempenho, obtidos a partir da data de início e término de cada atividade [7].

### 2.2.3 Algoritmos de Descoberta de Processos

Desde o início da mineração de processos vários algoritmos foram desenvolvidos, como por exemplo: *Alpha Algorithm* [17], *Heuristic Miner* [19] [21], *Fuzzy Miner* [22], *Genetic Miner* [25], *Integer Linear Programming Miner* [29], Mineração Declarativa [25], e *Inductive Miner* [31].

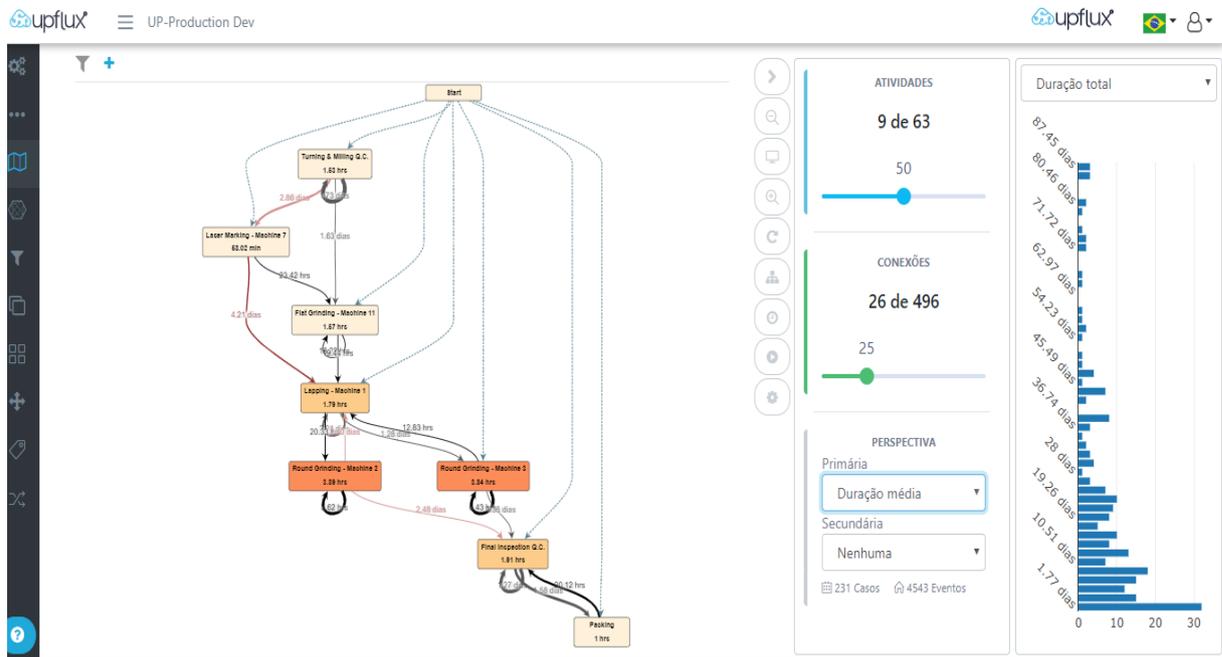
Dentre o conjunto de abordagens, o *Fuzzy Miner* [22] se destaca em aplicações do mundo real, em particular, tomando como entrada traços de eventos de processos semiestruturados. Ele permite analisar o modelo descoberto dando ênfase as variantes mais relevantes, filtrando-as por meio de medidas de significância. Quando um processo semiestruturado é analisado, filtros com base nestas medidas são úteis para aumentar a compreensão do processo de negócio. Esta abordagem se baseia em quatro pilares:

- 1 Agregação: o número de elementos exibidos é limitado;
- 2 Abstração: as informações insignificantes no contexto escolhido são omitidas;
- 3 Ênfase: as informações mais significativas são destacadas, i.e., os caminhos com maior tempo de espera ou frequência; e
- 4 Customização: diferentes visões podem ser geradas de acordo com as perspectivas de análise. A visualização de processo de negócio é dada por um diagrama de transições, sem representar paralelismo, junções e divisões no processo de negócio.

#### 2.2.4 Ferramentas

Nos últimos anos diversas ferramentas de mineração de processos foram desenvolvidas para os mais variados tipos de aplicação. A ferramenta de mineração de processo ProM [33], se destaca no meio acadêmico devido as diversas técnicas de mineração que ela implementa [6]. Essa ferramenta suporta *logs* de eventos nos formatos XES, MXML e CSV, que em conjunto com as técnicas: *Alpha Algorithm* [17], *Heuristic Miner* [19] [21], *Fuzzy Miner* [22], *Genetic Miner* [25], *Integer Linear Programming Miner* [29], *Mineração Declarativa* [25] e *Inductive Miner* [31] descobrem modelos e extensões de modelos de processos de negócios.

Outra ferramenta de mineração de processos é a UpFlux [34]. Ela implementa uma versão paralela do algoritmo *Fuzzy Miner* [22] para descobrir modelos processos. O paralelismo permite operar de forma eficiente com grandes *logs* de eventos. Ela tem suporte de integração via API. O formato do *log* de eventos para a tarefa de descoberta de modelos é CSV. A Figura 5 mostra um modelo de processo de negócio descoberto pela ferramenta UpFlux.



**Figura 5. Modelo de processo descoberto pela ferramenta UpFlux.**

A Figura 5 mostra um modelo de processo de negócio enriquecido. As cores mais intensas dão ênfase as atividades mais significativas em termos de frequência. A direita da mesma figura tem-se um gráfico de barras com a duração total das instâncias do processo.

### 2.3 Medida de Similaridade

Em geral, a similaridade é dada por um conjunto de pressupostos, e não diretamente por uma fórmula. Esses pressupostos são divididos em três proposições [35]: (a) a similaridade entre A e B está relacionado com a sua semelhança, de tal forma que, quanto mais elementos em comum, mais similares eles são; (b) a similaridade entre A e B está relacionada com as diferenças entre eles, em que quanto mais diferenças eles têm, menos similares eles são; e (c) a máxima similaridade entre A e B é atingida quando ambos são idênticos.

As medidas de similaridade de modelos de processos que podem ser segmentadas em 4 diferentes grupos [38]: (a) medidas baseadas na correspondência entre nós e arestas; (b) medidas baseadas na distância de edição de grafos; (c) medidas baseadas na análise de dependências causais entre atividades; e (d) medidas baseadas na comparação de conjunto de traços ou *logs* de eventos.

### 2.3.1 Medidas baseadas na correspondência entre nós e arestas

A ideia principal das medidas de similaridades baseadas na correspondência entre nós e aresta, está em comparar os rótulos das atividades presentes nos nós para determinar o grau de similaridade entre dois modelos de processos. Elas são consideradas simples e de baixo custo computacional. Por exemplo, em [53] os autores apresentam uma abordagem simples para calcular a similaridade de modelos de processos, em que se considera o número de rótulos de atividades equivalentes entre dois modelos de processos. Basicamente, o grau de similaridade é calculado em duas etapas. Na primeira etapa são identificados os rótulos comuns das atividades entre processos e, na segunda etapa, é calculada a similaridade dos rótulos (cf. E1).

$$sim(M_0, M_1) = 2 \times \frac{\sum A_0 \cap A_1}{A_0 + A_1} \quad (E1)$$

Onde,

- $A_0$  representa as atividades pertencentes ao conjunto de nós do modelo de processos  $M_0$ ; e
- $A_1$  representa as atividades pertencentes ao conjunto de nós do modelo de processos  $M_1$ .

Em [54] os autores definem que a equivalência entre modelos, se dá pelo grau de similaridade entre dois modelos de processos a partir de uma função de correlação de rótulos de atividades. A função de correlação que mede a similaridade entre rótulos individuais serve para calcular a equivalência de atividades. O cálculo do grau de similaridade entre dois modelos de processos (cf. E2)

$$sim(M_0, M_1) = \frac{2 \sum_{x \in A_0} corr(x, map(x))}{\sum N_0 + \sum E_0 \sum N_1 + \sum E_1} \quad (E2)$$

Onde,

- $N_0$  representa o conjunto de nós do modelo de processos  $M_0$ ;
- $E_0$  representa o conjunto de arestas do modelo de processos  $M_0$ ;
- $A_0$  representa as atividades pertencentes ao conjunto de nós do modelo de processos  $M_0$ ;

- $N_I$  representa o conjunto de nós do modelo de processos de  $M_I$ ; e
- $E_I$  representa o conjunto de arestas do modelo de processos de  $M_I$ .

### 2.3.2 Medidas baseadas na distância de edição de grafos

As medidas apresentadas em [49] e [50] visam medir a similaridade entre modelos de processos considerando a estrutura dos modelos comparados, i.e., o quanto os nós e as transições são equivalentes entre os mesmos processos.

Em [52], os autores utilizam regras que visam remover todas as estruturas de um grafo de processo que estão corretas. Já, em [51], as regras de [52] são utilizadas para reduzir um modelo  $M_o$  para algo que seja equivalente ou esteja contido em  $M_1$ . Em [49], o grau de similaridade entre dois modelos considera apenas o número de arestas comuns entre dois modelos de processos (cf. E3)

$$sim(M_0, M_1) = \frac{\Sigma(E_1^{red} \cap E_0)}{\Sigma E_1^{red}} \quad (E3)$$

Onde,

- $E_0$  é conjunto de arestas do modelo de processo  $M_0$ ; e
- $E_1^{red}$  é o conjunto reduzido de arestas do modelo de processos  $M_1$ .

Em [50], os autores propuseram uma abordagem para fazer a junção de dois modelos de processos similares, visando consolidar e eliminar as redundâncias nos modelos de processos comparados, em que primeiramente, calcula-se a similaridade entre dois modelos de processos para posteriormente sugerir a junção dos mesmos modelos (cf. E4)

$$sim(M_0, M_1) = \mathbf{1.0} - \frac{wskipn \times fskipn + wskipe \times fskipe + wsubn \times fsubn}{wskipn + wskipe + wsubn} \quad (E4)$$

Onde,

- $wskipn$  representa os pesos dos nós inseridos e removidos;
- $wskipe$  representa os pesos das arestas inseridas e removidas; e

- $w_{subn}$  representa os pesos atribuídos aos nós substituídos.

As frações dos nós inseridos/removidos, das arestas inseridas/removidas e da média da distância dos nós substituídos são dadas por  $f_{skipn}$ ,  $f_{skipe}$  e  $f_{subn}$ .

### 2.3.3 Medidas de dependências causais entre atividades

Medidas de similaridade de dependências causais entre atividades visam aferir a similaridade entre dois modelos de processos de acordo com as relações entre atividades e transições. A principal diferença dessas medidas em relação as que consideram somente a estrutura dos modelos, está relacionada ao fato que em alguma medida as relações adjacentes, implícitas e diretas entre os nós são consideradas [54].

Em [56], os autores representam as dependências causais entre atividades usando dois vetores: a) **vetor de atividades**: ele dá as probabilidades de execução de cada atividade de  $M_o$  pertencente a  $M_1$  onde  $A_o \cup A_1$ ; b) **vetor de transições**: ele é o resultado da multiplicação das probabilidades de execução pela distância entre duas atividades, em que a distância entre duas atividades é proporcional ao número de transições entre as mesmas.

Os valores atribuídos as transições entre atividades são definidos pelo inverso da distância entre transições diretas de duas atividades, de modo que o peso para transições diretas é 1 e o peso de uma transição implícita é um valor entre [0, 1].

A similaridade de dois modelos de processos, com base nestes vetores, é dada pelo coeficiente de cosseno. Este último permite obter um valor alto para dois vetores que possuem elementos em comum com valores também altos. As medidas de similaridade para os vetores de atividades e transições são definidas da seguinte forma (cf. E5 e E6).

$$sim_{act}(P_x, P_y) = \frac{a_x \cdot a_y}{|a_x||a_y|} = \quad (E5)$$

$$sim_{trans}(P_x, P_y) = \frac{t_x \cdot t_y}{|t_x||t_y|} = \frac{\sum t_{i,x}t_{i,y}}{\sqrt{\sum t_{i,x}^2 \sum t_{i,y}^2}} \quad (E6)$$

Portanto, a similaridade total de dois modelos de processos é obtida por meio da soma dos pesos dos coeficientes de cosseno dos vetores das atividades e das transições (cf. E7).

$$sim(P_x, P_y) = \alpha sim_{act}(P_x, P_y) + (1 - \alpha) sim_{trans}(P_x, P_y) \quad (E7)$$

Esta seção apresentou algumas formas de como realizar o cálculo de similaridade entre dois modelos de processo, já a seção seguinte aborda a otimização voltada para processos de negócio, que assim como a similaridade é um item essencial no desenvolvimento deste trabalho.

## 2.4 Otimização de Processos de Negócios

A otimização de processos de negócios é a melhoria contínua e automatizada de processos de negócios que usa medidas quantitativas de desempenho pré-especificadas [5]. Em [3], argumenta-se que a otimização de processos de negócios é crucial quando as organizações necessitam realizar mudanças nas prioridades de seus negócios, bem como nos casos de negócios com características variadas (e.g., cliente com diferentes preferências), com isso as especificações dos processos de negócio devem permitir flexibilidades, configuradas e reconfiguradas adequadamente. Dentro deste contexto, dois assuntos devem ser considerados. Primeiro, os processos devem ser continuamente melhorados. Segundo os processos devem ser flexíveis o suficiente para lidar com ambiente dinâmico [3].

Em [5], os autores apresentaram uma revisão e uma classificação das principais técnicas de modelagem de processos de negócio relacionadas a capacidade de análise e otimização, visando identificar quais tipos de modelos de processos de negócio são adequados para analisar e otimizar. A Tabela 2 resume as principais abordagens relacionadas a otimização de processo.

**Tabela 2 - Abordagens de otimização de modelos de processos de negócio. Fonte [5]**

Referência	Resumo
Sadiq & Orłowska [42]	A abordagem de Sadiq e Orłowska se apoia na técnica de redução de grafos para detectar conflitos estruturais em um modelo de processo. A abordagem básica consiste em remover todas as estruturas do grafo de fluxo de trabalho que estão definitivamente corretas.
van der Aalst et al. [45]	O trabalho de van der Aalst e seus colegas apresenta uma técnica de análise alternativa para a verificação de grafos de fluxo de trabalho, cf. introduzido em [42]. Essa nova técnica tem melhor desempenho que a técnica apresentada em [42], quando o grafo de fluxo de trabalho não é denso. Aqui, a abordagem se apoia na técnica clássica de rede de Petri para representar fluxo de trabalho.
Lin et al. [40]	A abordagem de Lin e seus colegas é semelhante à descrita por van der Aalst e seus colegas em [45]. Ela define um conjunto completo e mínimo de regras, bem como um novo algoritmo para identificar conflitos estruturais em modelos de processo. Porém, a técnica de redução de grafos não foi relacionada com medida de desempenho quantificável, embora tenha base algorítmica.
Soliman [43]	A abordagem desenvolvida por Soliman resultou em uma ferramenta realista para estimar os custos dos processos de mapeamento e determinar o nível econômico de mapeamento. Essa abordagem é baseada em um equilíbrio entre a utilidade e o custo da coleta de informações por meio do mapeamento de processos. A principal vantagem dessa abordagem é a redução geral do custo de reengenharia de processos de negócio. Portanto, essa abordagem pode ser chamada de mapeamento do processo de menor custo. Isso ocorre porque esse método pode ser usado para determinar o nível ideal de mapeamento de processo e o menor custo de mapeamento de processo.
Hofacker & Vetschera [30]	A abordagem proposta por Hofacker e Vetschera está relacionada às questões de otimização. Ela fornece suporte analítico para a otimização multicritério do projeto de um processo de negócio. A otimização multicritério é bastante atraente, já que processos de negócio geralmente

	<p>encerram critérios conflitantes. O suporte analítico para a otimização do projeto de um processo de negócio se fundamenta em três técnicas: programação matemática, método <i>branch and bound</i> e algoritmo genético.</p>
<p>Tiwari et al. [44] Vergidis et al. [46]</p>	<p>A abordagem proposta por Tiwari e seus colegas apresenta uma estrutura para aplicação da otimização multiobjetivo aos processos de negócios. Os esforços foram na aplicação de uma série de algoritmos evolutivos de otimização de multiobjetivo na tentativa de gerar diversos projetos de processos de negócios otimizados. A estrutura de otimização proposta inclui uma representação quantitativa do processo de negócio envolvendo duas matrizes, uma para capturar os detalhes do <i>design</i> do processo e outra para calcular e avaliar os atributos do processo. A proposta inclui um algoritmo que verifica a viabilidade de cada solução candidata (i.e., <i>design</i> do processo). De forma análoga, Vergidis e seus colegas propõem uma abordagem de otimização multiobjetivo automatizada focada na melhoria de processos de negócio. A estrutura proposta usa um modelo de processo de negócio genérico formalmente definido. E parte multiobjetivo é expressa em termos de custo e duração do processo. O modelo de processo de negócio é representado, e incorporado a uma plataforma de otimização, na qual dispõe-se de vários algoritmos de otimização multiobjetivo.</p>
<p>Gutjahr et al. [24]</p>	<p>A abordagem de Gutjahr e colegas apresenta um processo de otimização estocástica para determinar um conjunto de ações eficientes que fixam/travam atividades selecionadas em uma rede de atividades estocásticas. Esta abordagem modifica o método <i>branch-and-bound</i> estocástico convencional, usando uma heurística de pesquisa local. Tal modificação economiza tempo computacional e, ao fazer isso, fornece um método apropriado para resolver várias aplicações relacionadas a otimização estocástica combinatória. Um estudo computacional comparativo mostra que esta abordagem não apenas supera as técnicas convencionais, mas também melhora o método <i>branch-and-bound</i> estocástico convencional.</p>

Jaeger et al. [32]	A abordagem de Jaeger e seus colegas fornece uma estrutura de otimização fundamentada na avaliação de desempenho, fazendo intervenções nos recursos e processos para melhorar o desempenho de um sistema. A proposta se divide em 4 partes: a) um modelo de sistema executável baseado em workflows, gatilhos e recursos de execução; b) um ambiente de execução que mede o desempenho do sistema; c) operadores de melhoria que podem modificar o modelo do sistema; e d) um mecanismo automático de condução de melhorias para orientar as modificações do modelo do sistema <i>vis-à-vis</i> uma meta de desempenho.
Han [26]	A abordagem de Han encerra uma estrutura algorítmica para apoiar a confecção do projeto do processo de negócio usando modelos de decisão. O objetivo desta metodologia é reduzir o custo total de implementação de decisões criando um modelo quantitativo e usando quatro padrões de mudança de projeto: a) simplificação de um processo por meio de automações simples; b) sequenciamento linear; c) re-sequenciamento envolvendo paralelização de processos; e d) integração radical de processos implementados por algoritmos.
Zhou & Chen [49]	A abordagem de Zhou e Chen está voltada principalmente as questões de como atribuir recursos de acordo com as informações históricas dos recursos empregados nas execuções das atividades, buscando garantir o alto desempenho de um processo de negócio. A qualidade da atribuição de recursos é descrita pelo grau médio de aproximação à atribuição ideal de recursos e pelo grau de não-conformidade da atribuição de recursos com base no conceito de entropia da informação, que pode ser considerada a qualidade preditiva do processo de negócio. Os modelos de otimização combinados à avaliação multiobjetivo de processos de negócio foram desenvolvidos para processos de negócio <i>intra-empresa</i> e <i>inter-empresas</i> . A técnica evolucionista <i>algoritmo genético</i> foi utilizada para resolver o problema de otimização multiobjetivo.

Zhou & Chen [47]	A proposta de Zhou e Chen centrou-se seus esforços principalmente nas restrições de tempo, custos e recursos de um modelo de processo de negócio, visando otimizá-lo, utilizando também a técnica evolucionista <i>algoritmo genético</i> para minimizar o custo do processo.
Zhou & Chen [48]	Aqui, Zhou e Chen centraram-se seus esforços no desenvolvimento de uma metodologia para sistematizar o projeto para otimização de um processo de negócio a partir de perspectivas estratégicas, táticas e operacionais, usando métodos estruturados e quantitativos que suportam o <i>design</i> .
Rummel et al. [41]	Rummel e seus colegas propuseram um modelo para reduzir o tempo de ciclo de vida de um processo de negócio, consolidando atividades—atribuindo várias atividades a um mesmo ator—eliminando assim a coordenação e o atraso de entrega entre diferentes atividades que ocorreram quando atribuídos a diferentes atores. Como essa abordagem é focada na atividade (ou tarefa), ela ignora o atraso de interatividade que pode contribuir significativamente para o tempo total do ciclo de vida do processo.
Dewan et al. [15]	A abordagem de Dewan e seus colegas visa otimizar as tarefas de modo a reduzir o tempo total do ciclo de um processo. Não há metodologia sistemática para determinar o reprocessamento ideal de tarefas intensivas em informação. Um modelo matemático para redesenhar redes de processos complexos é apresentado, mas uma limitação do texto publicado é que ele se refere a processos de negócio com apenas fluxos de informação. Sua principal contribuição é a reestruturação efetiva do processo de negócio e redução do tempo total das tarefas, usando a redução ou a eliminação de atraso de entrega como resultado de uma reunião unificada.

A Tabela 3 classifica as abordagens apresentadas na Tabela 2, destacando modelo de processo de negócio, tipo de modelagem e tipo de otimização de processo.

**Tabela 3. Abordagens de otimização para modelos de processos de negócios. Fonte [5]**

<b>Modelo</b>	<b>Tipo de Modelagem</b>	<b>Tipo de otimização</b>	<b>Autor(es)</b>
Rede de Petri (e fluxo de trabalho)	Modelo diagramático Matemática / modelo formal	Técnica de redução de grafos	(Sadliq and Orłowska,2000) (van der Aalst et at.,200) (Lin et al., 2002)
Modelo matemático	Modelo matemático formal	Abordagem algorítmica	(Han, 2003) (Gutjahr et al., 2000) (Jaeger et al., 1995) (hofacker and Vetshera, 2001) (Soliman, 1998) (Tiwari et al., 2006) (Vergidis et al.,2006) (Volkner and Weners,200) (Zhou and Chen, 2003a) (Zhou and Chen, 2002) (Zhou and Chen, 2003b)
		Consolidação de atividades/tarefas	(Dewan et al., 1998) (Rummel et al., 2005)

A otimização e melhoria de um processo requer inicialmente a localização, identificação e caracterização de gargalos do processo de negócio, assim como do caminho do processo de negócio com maior tempo de duração. Essa situação pode ser endereçada com métodos e técnicas desenvolvidos na área de gerenciamento de projetos, dentre estes destacam-se: o método CPM (*Critical Path Method*), e a técnica PERT (*Program Evaluation and Review Technique*). Deve-se notar que além da localização, identificação e caracterização do caminho crítico, PERT/CPM contribui também minimizar a duração de um projeto, minimizar os custos de um projeto, maximizar as receitas de um projeto, entre outros [7].

## 2.5 Caminho Crítico

O caminho crítico em um esquema de fluxo de trabalho pode ser definido como: o caminho com a duração mais longa, partindo da atividade inicial até a atividade final. Ele pode ser

utilizado como informação básica para ajudar a resolver problemas envolvendo fluxo de trabalho, por exemplo, gerenciamento de tempo [8].

O método CPM é uma técnica clássica de modelagem de projetos<sup>1</sup> datada dos 50 do século XX. Um ano depois foi feito o primeiro teste da proposta de modelam. Ela foi aplicada na construção de uma fábrica de produtos químicos. Nesta mesma época, o escritório de projetos especiais da Marinha dos Estado Unidos criou a técnica PERT. Esta última visava gerenciar o programa de mísseis POLARIS. Devido às semelhanças entre CPM e PERT, elas são frequentemente chamadas de técnica PERT/CPM [7].

Dado a importância da PERT/CPM na abordagem do projeto de pesquisa ora em desenvolvimento, a próxima subseção descreve em detalhes o cálculo do caminho crítico.

### **2.5.1 Cálculo do caminho crítico**

O cálculo do caminho crítico envolve conceitos ligados a: rede de projeto, atividade no nó (AoN), tempo mínimo, tempo máximo de atraso, tempo mais cedo (ESS) e tempo mais tarde (LSS).

#### ***Rede de Projeto***

Uma rede de projeto parte de uma sequência lógica que envolve as várias atividades durante a execução de um projeto. Os vínculos entre tais atividades são chamados de relações de precedência tecnológica. A incorporação dessas ligações tecnológicas entre qualquer pares de atividades é o primeiro passo para a construção da rede de projeto [7]. Uma rede de projeto pode ser vista como um grafo  $G(N, A)$ , onde  $N$  denota o conjunto de nós e  $A$  denota o conjunto de arcos.

Para efeito prático, o conjunto de atividades de um projeto e suas relações de precedência tecnológica podem ser exibidas como uma rede usando dois formatos: atividade-no-nó (AoN) e atividade-em-arco (AoA). Para este trabalho será detalhado somente o formato (AoN), para mais detalhes do formato (AoA), ver [7].

---

<sup>1</sup> Proposta por: M. R. Walker da E.I. Du Pont de Nemours e Co. e J. E. Kelly da Remington Rand, por volta de 1957

### **Atividade no nó (AoN)**

Uma rede AoN exibe as atividades dos nós e as relações de precedência dos arcos. A maioria das ferramentas computacionais dependem do formato de AoN. A Figura 6 apresenta a estrutura da rede AoN por meio da ligação tecnológica entre as atividades  $i$  e  $j$ .

As três etapas a seguir mostram como construir uma rede de AoN:

1. desenhar um nó para cada atividade da rede de projeto;
2. desenhar um arco para cada relação de precedência entre duas atividades  $i$  e  $j$ ; e
3. adicionar uma atividade fictícia no início e uma atividade fictícia no final, assegurando um único início e um único fim para a rede de projeto.

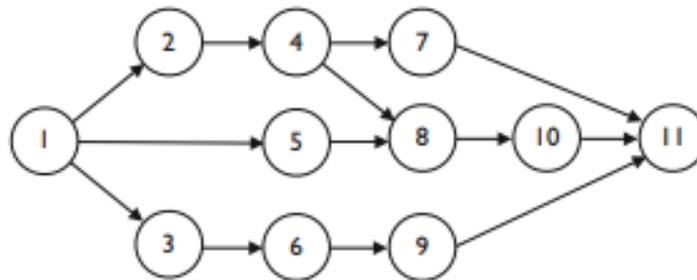


**Figura 6. Representação da rede AoN com uma ligação tecnológica entre as atividades  $i$  e  $j$ . Fonte [7]**

A rede de projeto AoN, da Tabela 4, exibe uma lista de atividades de projeto com seus antecessores. Deve-se notar a inserção de dois nós fictícios—1 e 11—para assegurar todas as relações de precedência.

**Tabela 4. Lista de atividades com seus antecessores. Fonte [7]**

Atividade	Antecessores
<b>1</b>	-
2	1
3	1
4	2
5	2
6	3
7	4
8	4, 5
9	6
10	8
<b>11</b>	7, 9, 10



**Figura 7. Rede de projeto AoN. Fonte [7]**

A Figura 7 mostra uma rede de projeto AoN, onde o conjunto  $N$  de nós representa atividades e o conjunto  $A$  de arcos representa as restrições de precedência. As atividades são numeradas. O nó  $1$  e o nó  $11$  são usados para denotar início e fim; ambos são atividades fictícias. A atividade fictícia de início  $1$  é antecessora para todas as outras atividades da rede e é usada para denotar o início do projeto. De forma semelhante, a atividade fictícia fim  $11$  denota o término do projeto e é sucessor para todas as outras atividades do projeto [7].

### **Mínimo/máximo**

Uma relação de início-término, com um intervalo de tempo mínimo  $n$ , pode ser representada da seguinte forma (cf. E8).

$$FS_{ij}^{min} = n \quad (E8)$$

Onde, a atividade  $j$  só pode iniciar  $n$  ou mais períodos após ao término da atividade  $i$ .

Uma relação de término-início, com um tempo de atraso máximo de  $n$ , pode ser representada da seguinte maneira (cf. E9):

$$FS_{ij}^{max} = n \quad (E9)$$

Onde, a atividade  $j$  só pode iniciar  $n$  ou menos períodos após ao término da atividade  $i$ .

A Tabela 5 exemplifica um projeto fictício com 12 atividades não fictícias (e uma atividade fictícia de início (1) e uma de final (14)). A Figura 6 exhibe a rede de projeto AoN correspondente ao exemplo da Tabela 4, onde o valor ao lado de cada nó da rede de projeto denota a duração da atividade.

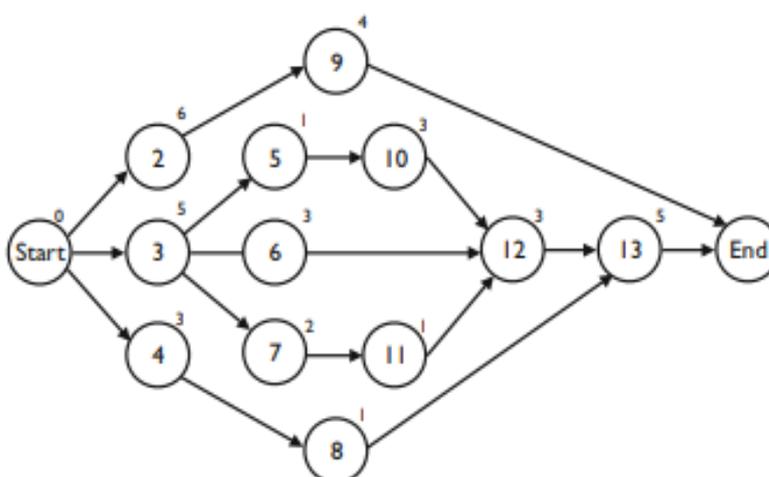
Um caminho em uma rede de projeto pode ser definido como uma sequência de atividades conectadas, partindo do nó de início até o nó final do projeto. Todas as atividades—e conseqüentemente, todos os caminhos—devem ser concluídas para finalizar o

projeto. A Tabela 6 enumera todos os caminhos possíveis da rede de projeto exibido na Figura 8 com sua correspondente duração total.

O primeiro intervalo de tempo para a possível conclusão do projeto é igual ao caminho mais longo na rede de projeto. Esse caminho é conhecido como o caminho crítico. Ele determina a duração geral do projeto. Para que o projeto não atrase é importante manter as atividades do caminho crítico no planejamento, na medida que atrasos em qualquer uma dessas atividades resultam em uma violação de toda a duração do projeto [7].

**Tabela 5. Um exemplo de projeto fictício com 12 atividades não fictícias. Fonte [7]**

Atividade	Antecessores	Duração(dias)
1	-	0
2	1	6
3	1	5
4	1	3
5	3	1
6	3	3
7	3	2
8	4	1
9	2	4
10	5	3
11	7	1
12	6, 10, 11	3
13	8, 12	5
14	9, 13	0



**Figura 8 - Uma rede de exemplo AoN da Tabela 4. Fonte [7]**

**Tabela 6 - Enumeração de todos os caminhos possíveis do projeto da Tabela 4: Fonte [7]**

Caminho	Duração
1 → 2 → 9 →	10
1 → 3 → 5 → 10 → 12 → 13 → 14	17
1 → 3 → 6 → 12 → 13 → 14	16
1 → 3 → 7 → 11 → 12 → 13 → 14	16
1 → 4 → 8 → 13 → 14	9

Na fase de construção de um projeto pode ser tentador gerar todos os caminhos possíveis para determinar o caminho mais longo do referido projeto. Mas infelizmente, devido à grande quantidade de caminhos que uma rede de projeto pode produzir, o tempo de CPU necessário para gerá-los torna tal iniciativa inaplicável. Geralmente, as ferramentas computacionais dependem de uma abordagem em três etapas para detectar o caminho crítico de uma rede [7]:

1. calcular o tempo mais cedo;
2. calcular o tempo mais tarde; e
3. calcular a folga para cada atividade.

### ***Tempo mais cedo (ESS)***

O primeiro início  $es_i$  de cada atividade  $i$  pode ser determinado usando cálculos futuros na rede do projeto. O início mais tardio de uma atividade é *maior ou igual* que o *primeiro término* de todas as suas atividades predecessoras. O *primeiro término*  $ef_i$  de uma atividade  $i$  é definido como sua *primeira hora de início* acrescida da sua duração estimada. Os primeiros tempos de início podem ser calculados usando os seguintes cálculos de encaminhamento, começando a partir do nó fictício de início 1 (cf. E10 e E11):

$$es_1 = 0 \quad (E10)$$

$$es_j = \max(es_i + d_i | i \in P_j) \quad (E11)$$

os primeiros tempos de término são dados (cf. E12):

$$ef_i = es_i + d_i \quad (E12)$$

Pode-se verificar que as primeiras horas de início das atividades do projeto da Tabela 8 são dadas (cf. E13):

$$\begin{aligned} es_1 = 0, es_2 = 0, es_3 = 0, es_4 = 0, es_5 = 5, \\ es_6 = 5, es_7 = 5, es_8 = 3, es_9 = 6, es_{10} = 6, \\ es_{11} = 7, es_{12} = 9, es_{13} = 12, es_{14} = 17 \end{aligned} \quad (E13)$$

Portanto a duração total mínima do projeto é 17 unidades de tempo.

### **Tempo mais tarde (LSS)**

O *último término*  $lf_n$  de cada atividade  $i$  pode ser calculado de forma análoga, usando cálculos para trás, a partir da duração do projeto no nó final fictício do projeto. O *último término*  $lf_i$  de uma atividade é *menor ou igual* que o *início mais cedo* de todas as suas atividades sucessoras. O *início mais cedo*  $ls_i$  de uma atividade  $i$  é definido como seu *último tempo de término* menos a sua duração estimada. Os últimos tempos de término podem ser calculados usando os seguintes cálculos, começando com o nó final fictício  $n$  (cf. E14 e E15):

$$lf_n = 8_n \quad (E14)$$

$$ls_i = \min (lf_j - d_j | j \in S_i) \quad (E15)$$

as últimas horas de início são dadas (cf. E16).

$$ls_i = lf_i - d_i \quad (E16)$$

Dado o prazo final do projeto de 17 unidades de tempo, calculado como o início da atividade fictícia final na etapa anterior, as últimas horas de início de cada atividade são dadas (cf. E18):

$$\begin{aligned} ls_1 = 0, ls_2 = 7, ls_3 = 0, ls_4 = 8, ls_5 = 5, \\ ls_6 = 6, ls_7 = 6, ls_8 = 11, ls_9 = 13, ls_{10} = 6, \\ ls_{11} = 8, ls_{12} = 9, ls_{13} = 12 \text{ e } ls_{14} = 17 \end{aligned} \quad (E18)$$

### **Folga de atividade**

A quantidade de tempo de folga associada a cada atividade é usada para denotar o tempo livre de cada atividade dentro do *ESS* e do *LSS*. Ele denota a quantidade de tempo que cada atividade pode ser adiada sem violar a duração final do projeto. A folga (ou flutuação) da atividade  $i$  pode ser calculada (cf. E19):

$$ls_i - es_i = lf_i - ef_i \quad (E19)$$

As atividades com folga zero não podem ser atrasadas sem afetar a duração final do projeto e são chamadas de atividades críticas. Portanto, o caminho crítico consiste no caminho formado pelas atividades críticas, por exemplo, dado pelas atividades 1, 3, 5, 10, 12, 13 e 14 da Tabela 7.

**Tabela 7. Folgas das atividades de projeto de exemplo da Figura 6. Fonte [7]**

Atividade	Folga
1	0
2	7
3	0
4	8
5	0
6	1
7	1
8	8
9	7
10	0
11	1
12	0
13	0
14	0

Para concluir um projeto no tempo esperado, o cálculo de *caminho crítico* contribui para que o gestor se concentre nas atividades críticas às quais requeem mais atenção e recursos.

### **2.5.2 Avaliação de programação e técnica de revisão**

A abordagem CPM supõe-se que as durações das atividades não são alteradas ou que as mudanças são muito pequenas e insignificantes. Entretanto, na abordagem PERT pressupõe

que pode haver alterações nas durações das atividades. Assim, o PERT é uma solução estatística para atividades com durações incertas. Ele permite estimar a probabilidade de conclusão de cada atividade [7]. Face tal incerteza sobre os tempos de cada atividade, a técnica PERT fornece uma técnica particular para estimar a duração de cada atividade. A técnica permite obter três estimativas de duração para cada atividade individual [7]:

- Estimativa de *tempo otimista*: o menor tempo possível em que a atividade poderá ser concluída, onde tudo deve ser perfeito.
- Estimativa de *tempo pessimista*: o tempo mais longo possível em que a atividade poderá ser concluída—assume-se aqui o cenário de pior caso.
- Estimativa de *tempo realista*: o tempo mais provável em que a atividade poderá ser concluída em circunstâncias normais.

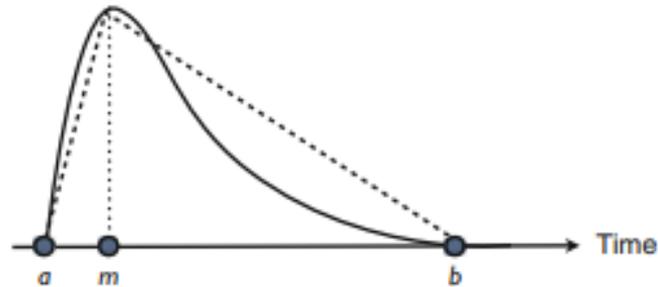
A Tabela 8 exhibe as três estimativas de tempos às atividades de projeto, tomando como exemplo a rede da Figura 8. A estimativa de *tempo otimista* é denotada por  $a$ , a estimativa de *tempo pessimista* é indicada por  $b$  e  $m$  é usada para se referir à estimativa de *tempo realista*.

**Tabela 8. Três estimativas de tempo para as atividades de projeto. Fonte [7]**

Atividade	Otimista $a$	Realista $m$	Pessimista $b$	$t$	$\sigma$
1	0,0	0,0	0,0	0	0,00
2	2,0	5,0	13,0	6	1,67
3	3,0	3,5	14,0	5	2,00
4	1,0	3,0	5,0	3	1,67
5	1,0	1,0	1,0	1	2,00
6	1,0	2,0	9,0	3	1,33
7	1,0	2,0	3,0	2	0,33
8	1,0	1,0	1,0	1	0,00
9	1,0	4,0	7,0	4	1,00
10	1,0	3,0	5,0	3	0,67
11	0,5	1,0	1,5	1	0,17
12	1,0	2,5	7,0	3	1,00
13	1,0	4,0	13,0	5	2,00
14	0,0	0,0	0,0	0	0,00

Em PERT, cada duração da atividade é uma variável aleatória entre dois valores extremos—i.e.  $a$  e  $b$ —e segue uma distribuição de probabilidade *beta*. A Figura 9 mostra uma típica função de distribuição *beta* e sua aproximação triangular. Note que a diferença entre  $m$  e  $b$  é

frequente, mas não necessariamente maior que a diferença entre  $a$  e  $m$  para expressar assimetria positiva.



**Figura 9. Distribuição *beta* utilizada para expressar a variabilidade da duração da atividade em PERT. Fonte [7]**

O tempo esperado  $t$  de uma distribuição *beta* pode ser aproximado pela média ponderada, da seguinte maneira (cf. E20):

$$t = \frac{a + 4m + b}{6} \quad (\text{E20})$$

Onde,  $a$  é o tempo otimista,  $m$  o tempo realista e  $b$  o tempo pessimista.

O desvio padrão da duração de uma atividade, que serve como medida de risco, pode ser calculado usando um intervalo de três desvios padrão (cf. E21):

$$\sigma = \frac{b - a}{6} \quad (\text{E21})$$

Este cálculo é de fato inspirado no canônico intervalo de três *sigma* para a distribuição normal. Um intervalo de três *sigma* para a distribuição normal é o intervalo entre a média menos três vezes o desvio padrão e a média mais três vezes o desvio padrão, pois exatamente 99,73% das observações se encontram nesse intervalo. Com base neste contexto, o PERT supõe que quase nenhuma das durações reais das atividades estará de fora desse intervalo [7].

## 2.6 Lead-time

O trabalho desenvolvido em [16], destaca a importância do *lead-time* para um processo de negócio. Com um exemplo prático da área de manufatura, define-se o *lead-time físico* como sendo o momento entre a chegada da matéria prima ao armazém do fabricante até a entrega ao cliente, além do *lead-time financeiro* que fica entre o pagamento do fornecedor e o recebimento do cliente (cf. Figura 10).

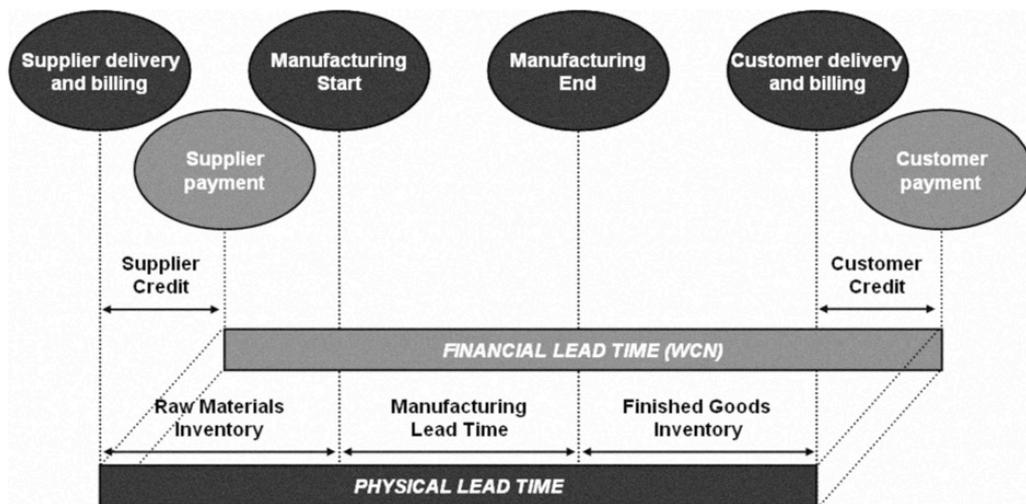


Figura 10. *Lead-time* físico e financeiro. Fonte [16]

Os autores também relatam que a maioria das empresas tem seus esforços direcionados para o controle de custos, porém eles defendem que a redução do *lead-time* pode ser muito mais lucrativa para a saúde dos negócios.

## 2.7 Trabalhos Relacionados

Essa seção destaca os principais trabalhos relacionados a mineração de processos que aborda questões relacionadas a identificação de gargalos, otimização e melhoria de processo.

### 2.7.1 Estimativa de tempo

Em [8] foi proposto um método para estimar o tempo de ciclo de um processo de negócio com base em listas de trabalhos individuais e suas relações envolvendo recursos e atividades. A

abordagem proposta foi construída com base no modelo de filas e na teoria da distribuição conjunta de variáveis aleatórias.

Em [9] assumiu-se que uma previsão precisa do desempenho do processo de negócio durante sua fase de concepção pode facilitar a avaliação do processo existente e a geração de alternativas. Assim, foi proposto um método de aproximação para estimar o tempo de ciclo de um processo de negócio. Utilizou-se o tempo de ciclo para definir o período entre o recebimento de um pedido de um cliente e a conclusão do pedido, pois processos executados por um BPMS são dinâmicos, e informações completas são raramente conhecidas antes de serem executadas. Logo, se o tempo do ciclo puder ser previsto no tempo de *design* do processo, ele poderá facilitar a avaliação e a racionalização do processo existente, bem como o esboço de um novo processo. O método apresentado usou um modelo de fila para estimar o tempo de ciclo de um processo de negócio.

Em [12] foi proposto um método para estimar a duração das tarefas, que geralmente são incertas e não-deterministas. A metodologia introduz estimadores *fuzzy* não assintóticos para definir durações de tarefas e proporcionar boas estimativas de tempos de execução, que não dependam apenas da duração média e da variação dos dados de amostra.

Em [13] foram usadas informações geradas por meio da tarefa de descoberta de modelos de processos para prever o tempo de conclusão das instâncias em execução. Foi desenvolvida uma abordagem configurável que permitiu construir um modelo de processo com informações de tempo aprendidas a partir de instâncias anteriores; usou-se o que foi aprendido para prever o tempo de conclusão. Essa abordagem foi implementada no ProM [33].

### **2.7.2 Simulação e melhoria de processo**

Em [10] foi abordada uma metodologia para o redesenho de processo. A metodologia *combinou* técnicas de simulação, ferramentas de mineração de dados e processos. Técnicas específicas de mineração de processos—e.g., *alpha miner*, *heuristic miner*, *genetic algorithm*—, permitiram descobrir modelos de processos reais e verificar a conformidade *vis-à-vis* as regras de negócios. Técnicas de mineração de dados, como árvores de decisão e análise de *cluster*, foram úteis para determinar as variáveis que influenciavam nos tempos de ciclo do processo e para determinar as chances de executar um processo dentro de um

determinado limite de tempo. A partir do modelo de processo descoberto foi criado um modelo de simulação. A simulação permitiu testar diferentes alternativas de *redesign* de processo, visando avaliar melhorias antes de serem colocadas em prática.

Em [11], de forma muito semelhante a abordagem apresentado em [10], utilizou-se de dados de execuções de processos para a construção de modelos de simulação. Tal abordagem visou extrair conhecimentos de arquivos de *logs* por meio de técnicas de mineração de processos para complementar às fontes de informação tradicionais, tais como, documentação do processo, entrevistas e observações.

## **2.8 Considerações**

Os trabalhos relacionados apresentam soluções que utilizam basicamente a mineração de processos para estimar o tempo de conclusão de um projeto e realizar simulações de cenários alternativos no *redesign* de um processo. Porém, não foi encontrado nada que utilize especificamente a mineração de processos como base para encontrar o caminho crítico e em função deste tentar reduzir o *lead-time* do processo de negócio em questão. É nesta lacuna que esse trabalho de pesquisa busca estabelecer sua contribuição para a otimização de processos de negócio.

### 3 MÉTODO- $\beta$ : REDUÇÃO DO LEAD-TIME

A proposta visa definir e avaliar um método computacional eficiente para gerar alternativas e avaliá-las na direção de reduzir o *lead-time* de um processo de negócio, apoiando-se fortemente nas informações e modelos gerados pela mineração de processos, assim como nos conceitos e nas técnicas canônicas PERT/CPM.

A relevância da mineração de processos é notória, na medida que ela permite realizar de forma automática a descoberta de modelos de processos a partir de logs de eventos. Esses modelos fornecem informações que permitem realizar vários tipos de análises baseadas em modelos: verificação de conformidade, avaliação de desempenho e identificação de gargalos nos processos de negócio. A identificação de gargalos é altamente relevante para o gestor de processos de negócio e particularmente quando feita com suporte computacional para analisá-los e sugerir melhorias, o que lhe permite atuar no curto prazo.

A abordagem proposta de redução de *lead-time* deve provocar pouca alteração no modelo de processo analisado. Entende-se como pequena alteração, aquela cujo o ajuste no modelo de processo corrente porta apenas sobre uma única atividade. Dada essa restrição, o método proposto para a redução de *lead-time* de processos de negócio passa pela identificação e remoção de gargalos. A eficiência do método está na intervenção focada no caminho crítico, permitindo assim reduzir *lead-time* provocando pouca alteração no processo corrente.

Naturalmente, o método proposto baseia-se fortemente na mineração de processos. Mais precisamente, o método usa a mineração para descobrir modelos de instâncias de processos e obter informações estatísticas básicas de um processo negócio, como por exemplo: duração mínima, média e máxima de cada atividade. O método também usa a técnica PERT/CPM para calcular a duração média ponderada de cada atividade e caminho crítico de uma rede de atividades ou modelo de processo. O caminho crítico dá a direção para identificar de forma efetiva a atividade relevante *vis-à-vis* o foco de interesse da análise; i.e., a atividade com gargalo no caminho crítico do processo de negócio em análise. A identificação do gargalo de uma determinada atividade leva em consideração o seu tempo médio ponderado

e a frequência absoluta, assim como o tempo médio de espera para realizar a execução das instâncias de um processo de negócio.

### 3.1 Método- $\beta$

O seu funcionamento básico do método parte da indicação de um log de eventos inicial  $E_0$ , uma taxa de redução desejada  $R$  para o *lead-time* do modelo de processo de negócio  $M_0$  e um número máximo de iteração  $N$ . As etapas do método são as seguintes:

1. descobrir modelo de processo de negócio  $M_0$  a partir do log de eventos  $E_0$ ;
2. calcular capacidade de execução de cada atividade do modelo de processo  $M_0$ ; assume-se como capacidade o tempo médio ponderado de cada atividade (seguindo modelo PERT/CPM);
3. calcular caminho crítico  $P_0$  do modelo  $M_0$ ;
4. calcular *lead-time*  $LT_0$  do modelo  $M_0$  com  $P_0$ ;
5.  $r = 0$ ;  $i = 1$ ;  $S_0 = 100\%$ ;  $RP = \{<M_0, S_0, r>\}$  #  $S_0$  é similaridade de  $M_0$  com  $M_0$  e  $r$  é melhoria;
6. enquanto ( $r < R$ ) e ( $i < N$ ) faça:
  - encontrar no modelo  $M_{i-1}$  a atividade  $A$  com maior gargalo;
  - duplicar capacidade da atividade  $A$ , gerando  $A_1$  e  $A_2$ , onde  $2|A| = |A_1| + |A_2|$ ;
  - substituir, no *log* de eventos  $E_{i-1}$ , atividade  $A$  pelas atividades  $A_1$  e  $A_2$ , gerando um novo *log* de eventos  $E_i$ ;
  - descobrir modelo de processo  $M_i$  a partir do *log* de eventos  $E_i$ ;
  - calcular similaridade  $S_i$  a partir de  $M_0$  e  $M_i$ ;
  - calcular caminho crítico  $P_i$  do modelo  $M_i$ ;
  - calcular *lead-time*  $LT_i$  do modelo  $M_i$  com  $P_i$ ;
  - $r = (LT_0 - LT_i) / LT_0$ ;
  - $i = i + 1$ ;
  - $RP = RP = \{<M_i, S_i, r>\}$ ;
7. retornar  $RP$ .

A seguir será mostrado o passo-a-passo do método e as ferramentas de mineração processos que foram utilizadas. O resultado em termos de melhoria produzido pelo método está no cálculo da diferença entre  $M_0$  e  $M_i$  para verificar a eficiência da redução do *lead-time*.

As subseções a seguir descrevem cada etapa da construção e avaliação do Método- $\beta$ .

### 3.2 Logs de Eventos

Os dados de entrada para mineração de processos são *logs* de eventos. Tais logs podem ser extraídos de um sistema de informação prontamente disponível ou gerados artificialmente, conforme a necessidade de análise.

**Tabela 9. Log de eventos.**

<b>Caso</b>	<b>Atividade</b>	<b>Data Inicial</b>	<b>Data Final</b>	<b>Recurso</b>	<b>Custo</b>
1	A	11/04/2020 00:00	11/04/2020 00:10	Jose	52,00
1	D	11/04/2020 00:10	11/04/2020 00:20	Rafael	52,00
2	B	11/04/2020 00:00	11/04/2020 00:05	Gilberto	82,00
3	C	11/04/2020 00:00	11/04/2020 00:10	Eduardo	98,00
3	E	11/04/2020 00:10	11/04/2020 00:15	Denise	53,00
3	F	11/04/2020 00:15	11/04/2020 00:25	Marcos	35,00
3	H	11/04/2020 00:25	11/04/2020 00:30	Sheila	42,00
4	C	11/04/2020 00:10	11/04/2020 00:30	Eduardo	98,00
4	E	11/04/2020 00:30	11/04/2020 00:40	Denise	53,00
4	F	11/04/2020 00:40	11/04/2020 01:00	Marcos	35,00
4	H	11/04/2020 01:00	11/04/2020 01:05	Sheila	42,00
5	C	11/04/2020 00:30	11/04/2020 01:00	Eduardo	98,00
5	E	11/04/2020 01:00	11/04/2020 01:15	Denise	53,00
5	F	11/04/2020 01:15	11/04/2020 01:45	Marcos	35,00
5	H	11/04/2020 01:45	11/04/2020 01:50	Sheila	42,00
6	C	11/04/2020 01:00	11/04/2020 01:20	Eduardo	98,00
6	E	11/04/2020 01:20	11/04/2020 01:30	Denise	53,00
6	G	11/04/2020 01:30	11/04/2020 01:40	Pedro	35,00
6	H	11/04/2020 01:40	11/04/2020 01:45	Sheila	42,00

A próxima definição consiste na abordagem que será usada para a descoberta do modelo de processo, utilizando as técnicas de mineração de processos.

### 3.3 Descoberta de modelo de processos

Há várias abordagens algorítmicas para descobrir modelos de instâncias de processos. Optou-se aqui pela abordagem *Fuzzy Miner* [22], na sua versão algorítmica-paralela implementada na plataforma *UpFlux* [34]. O paralelismo permite operar de forma eficiente—em termos de tempo de resposta—com grandes logs de eventos. O formato de entrada para o log de eventos é CSV.

A *UpFlux* permite além de descobrir um modelo de processo de negócio a partir de um log de eventos, ela também gera informações estatísticas como: tempo médio, mínimo, máximo e mediana para cada atividade, assim como a frequência média e mediana de cada instância de processo de negócio. Deve-se notar que essas informações são persistidas juntamente como o modelo descoberto e podem ser usadas na próxima fase do Método- $\beta$ —a otimização de *lead-time*. Essa próxima etapa será dedicada a geração de informações complementares ao modelo de processo descoberto, com vista a localizar, identificar e caracterizar a *atividade mais crítica*, i.e., a atividade com o maior gargalo e situada no caminho crítico do processo de negócio.

As três estimativas de tempo mínimo, máximo e médio de cada atividade da Tabela 10 são obtidas a partir da execução da tarefa de descoberta de modelo de instâncias processos. Com base nas informações do modelo descoberto, pode-se aplicar a técnica de PERT/CPM para calcular o tempo esperado para cada atividade, dado pelo tempo médio ponderado de cada atividade (cf. E22):

$$wAveT = \frac{maxT + 4.aveT + minT}{6} \quad (E22)$$

**Tabela 10. Tempos das atividades do modelo de processo descoberto.**

Atividade	Tempo			
	Mínimo ( <i>minT</i> )	Máximo ( <i>maxT</i> )	Médio ( <i>aveT</i> )	Médio Ponderado ( <i>wAveT</i> )
A	10	10	10	10
B	5	5	5	5
C	10	30	20	20
D	10	10	10	10
E	5	15	10	10
F	10	30	20	20
G	10	10	10	10
H	5	5	5	5

A próxima etapa consiste em calcular e identificar o caminho crítico do processo de negócio.

### 3.4 Identificar e calcular caminho crítico e gargalos

A Tabela 10 contém oito atividades com seus respectivos tempos de execução. A Tabela 11 mostra todos os caminhos do processo de negócio, identificando o maior tempo de execução. Deve-se observar que há um tempo médio de espera ( $T_e$ ) de 77.5 para iniciar atividade C.

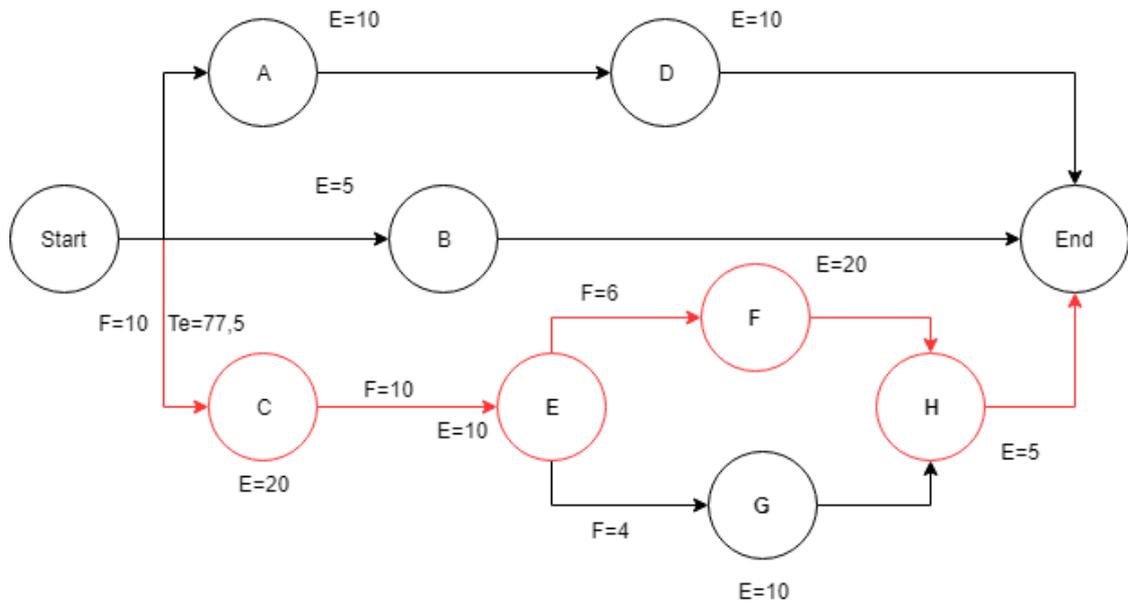
**Tabela 11. Lista de atividades e seus tempos de execução médio.**

Activity	Start	A	B	C	D	E	F	G	H	End
Time	0	10	5	20	10	10	20	10	5	0

**Tabela 12. Lead-time de cada caminho no modelo de processo**

Caminho	Duração média – atividade/transição
Start→A→D→End	$0 + 0 + 10 + 0 + 10 + 0 + 0 = 20$
Start→B→End	$0 + 0 + 5 + 0 + 0 = 5$
Start→C→E→F→H→End	$0 + 77,5 + 20 + 0 + 10 + 0 + 20 + 0 + 5 + 0 = 132,5$
Start→C→E→G→H→End	$0 + 77,5 + 20 + 0 + 10 + 0 + 10 + 0 + 5 + 0 = 122,5$

A Figura 12 mostra o modelo de processo de negócio descoberto com o seu caminho crítico identificado. A letra  $E$ , ao lado de cada atividade, representa o tempo de execução da atividade,  $F$  representa a frequência absoluta das instâncias do modelo processo de negócio por onde passa o caminho crítico.  $Te$  é o tempo de espera.



**Figura 11 - Modelo de processo e caminho crítico em destaque.**

A Tabela 13 resume algumas fórmulas e as notações usadas para ilustrar os cálculos do Método- $\beta$ .

**Tabela 13. Notações para os cálculos de otimização**

Notação	Descrição
$T_s$	Tempo de saída da atividade anterior.
$A_{fa}$	Tempo de fim da instância anterior.
$A_i$	Tempo de início da atividade: se $A_{fa} \geq T_s$ então $A_{fa}$ senão $T_s$
$E$	Tempo de execução da atividade.
$A_f$	Tempo de fim da atividade: $A_i + E$
$F$	Folga da atividade: $T_s - A_{fa}$
$T_e$	Tempo de espera: $A_i - T_s$
$I$	Instância de processo.

Deve-se notar que cada linha das próximas tabelas (14 e 15) correspondem as valores de uma instância de processo para uma dada atividade. Os valores da coluna  $A_i$  representam os tempos para iniciar a execução de cada instância de processo  $I_i$ .

**Tabela 14. Cálculos das atividades C e E.**

<b>I</b>	<b>Atividade C</b>							<b>Atividade E</b>						
	$T_s$	$A_i$	$E$	$A_f$	$A_{fa}$	$F$	$T_e$	$T_s$	$A_i$	$E$	$A_f$	$A_{fa}$	$F$	$T_e$
1	0	0	20	20	0	0	0	20	20	10	30	0	20	0
2	0	20	20	40	20	0	20	40	40	10	50	30	10	0
3	0	40	20	60	40	0	40	60	60	10	70	50	10	0
4	0	60	20	80	60	0	60	80	80	10	90	70	10	0
5	10	80	20	100	80	0	70	100	100	10	110	90	10	0
6	10	100	20	120	100	0	90	120	120	10	130	110	10	0
7	15	120	20	140	120	0	105	140	140	10	150	130	10	0
8	20	140	20	160	140	0	120	160	160	10	170	150	10	0
9	30	160	20	180	160	0	130	180	180	10	190	170	10	0
10	40	180	20	200	180	0	140	200	200	10	210	190	10	0

**Tabela 15. Cálculos das atividades F e H.**

<b>I</b>	<b>Atividade F</b>							<b>Atividade H</b>						
	$T_s$	$A_i$	$E$	$A_f$	$A_{fa}$	$F$	$T_e$	$T_s$	$A_i$	$E$	$A_f$	$A_{fa}$	$F$	$T_e$
1	30	30	20	50	0	30	0	20	20	5	25	0	20	0
2	50	50	20	70	50	0	0	40	40	5	45	25	15	0
3	70	70	20	90	70	0	0	60	60	5	65	45	15	0
4	90	90	20	110	90	0	0	80	80	5	85	65	15	0
5	110	110	20	130	110	0	0	100	100	5	105	85	15	0
6	130	130	20	150	130	0	0	120	120	5	125	105	15	0

Para identificar a atividade que deve ter sua capacidade de execução incrementada, soma-se os tempos de espera de cada atividade (cf. E23).

$$A_{cr} = \max_{x \in \Pi} \sum_{t \in T_e^x} t \quad (E23)$$

Onde  $\Pi = \{C, E, F, H\}$  é o conjunto de atividades do caminho crítico e  $A_{cr}$  é a atividade com o maior tempo de espera. Data a identificação de  $A_{cr}$  executa-se a primeira aproximação de otimização, observando a seguinte condição: caso a frequência absoluta do caminho crítico seja menor que dois, a próxima fase do Método- $\beta$ , não é realizada.

### 3.5 Executar a otimização

Para reduzir o *lead-time* de um processo dentro da abordagem proposta, a primeira aproximação consiste em duplicar a capacidade da atividade identificada como crítica  $A_{cr}$  e distribuir as instâncias dos processos de forma proporcional às atividades frutos da duplicação. A ilustração desta operação consiste em substituir a atividade crítica  $A_{cr}$  corrente por duas atividades com a mesma capacidade da atividade substituída (cf. E24).

$$|A_1| + |A_2| = 2|A_{cr}| \quad (E24)$$

Onde  $|x|$  representa a capacidade de execução da atividade  $x$ .

Esta operação de duplicação requer que o *log* de eventos  $E_0$  seja atualizado, gerando um novo *log* de eventos  $E_{i=1}$ . Tal atualização consiste em redistribuir proporcionalmente as instâncias da atividade  $A_{cr}$  para as atividades  $A_1$  e  $A_2$ .

Deve-se notar que o índice zero foi reservado para indicar o *log* de eventos original ( $E_0$ ) e o seu modelo de processo correspondente ( $M_0$ ). O próximo passo consiste em descobrir o novo modelo de processos  $M_i$  a partir do novo *log* de eventos  $E_i$ . Relembrando que o método é iterativo e uma das condições é  $i \in [1, N]$ , onde  $N$  é o número máximo de iterações.

$$M_{i \in N} = \text{discoveryProcess}(E_{i \in N}) \quad (F1)$$

Dado o novo modelo de processo  $M_i$ , avalia-se a capacidade de execução de  $M_i$ , considerando o *log* de eventos  $E_i$ , i.e., avalia-se a capacidade de execução em paralelo das atividades  $A_1$  e  $A_2$  em substituição a atividade crítica  $A_{cr}$ .

O próximo passo consiste em calcular o caminho crítico do modelo de processo  $M_i$  e seu *lead-time*  $LT_i$ . O interesse aqui é obter a razão da diferença entre *lead-time* do processo  $M_0$  e *lead-time* do processo  $M_i$ .  $r_i = \frac{(LT_0 - LT_i)}{LT_0}$  (E25)

Onde  $LT_0$  é o *lead-time* do processo original.

Além da obtenção da redução do *lead-time*, o nosso interesse porta também sobre o quanto o novo modelo  $M_i$  é similar ao modelo  $M_0$ . Espera-se obter um modelo de processo  $M_i$  com a maior similar possível com  $M_0$  e com a maior redução de *lead-time*.

$$S_{i \in N} = \text{similarity}(M_0, M_{i \in N}) \quad (F2)$$

Para realizar o cálculo da similaridade foi utilizado o cálculo de medidas baseadas na correspondência entre nós e arestas, i.e., identifica-se os rótulos das atividades comuns entre os modelos de processos [17]:

$$\text{sim}(M_0, M_1) = 2 \times \frac{\sum A_0 \cap A_1}{A_0 + A_1} \quad (E26)$$

Onde  $A_0$  e  $A_1$  são respectivamente os rótulos das atividades dos processos  $M_0$  e  $M_1$ .

Neste ponto, encerra-se a primeira iteração do método. A condição de continuação é: se  $i \in [1, N]$  e  $r_i \in [R, 1]$ , então *continua* senão *encerra*. Relembramos que  $R$  é o valor de redução desejado sobre o *lead-time*  $LT_0$ .

### 3.6 Considerações do Capítulo

Esta seção apresentou o comportamento básico do *Método-β*, focado na redução do *lead-time* de um processo de negócio, buscando também provocar a menor perturbação possível no modelo do processo original. A menor perturbação é altamente relevante para evitar custos elevados associados as grandes mudanças no ambiente de execução do referido processo.

A próxima seção apresenta como foi realizada a pesquisa, detalhando a fases de construção dos experimentos e as bases de dados que foram utilizadas.

## 4 Estruturação da Pesquisa

Esta seção apresenta os procedimentos metodológicos que foram realizados para colocar em execução o desenvolvimento desta pesquisa. Nesta linha, descrevemos de forma breve o método de pesquisa, a estratégia de pesquisa com etapas dos experimentos, a construção do ambiente de experimentação, a coleta de dados, bem como a avaliação dos resultados obtidos.

### 4.1 Método de Pesquisa

Foram realizados experimentos com objetivo de analisar e avaliar a funcionalidade do método proposto. Para [27] o principal objetivo de um experimento é coletar dados suficientes de uma população e todos de um mesmo ambiente para obter um resultado estatisticamente significativo sobre o atributo estudado em comparação com outro atributo. Os tópicos a seguir mostram as fases do experimento:

<b>Tópico</b>	<b>Descrição</b>
Definição	Definição do problema e do objetivo de pesquisa.
Planejamento	Definição do projeto do experimento, considerando aspectos de validade do mesmo.
Execução	Coleta de dados provenientes do experimento.
Análise	Análise de dados experimentais coletados.
Apresentação dos Resultados	Discussão e apresentação dos resultados.

Definida as fases do método escolhido, a seguir são apresentadas as questões de pesquisa.

## 4.2 Hipótese de Pesquisa

Relembramos que o objetivo desta pesquisa é desenvolver um método capaz de analisar um modelo de processo de negócio e propor alternativas para reduzir o seu *lead-time*, com base na premissa de que essa redução passa pela identificação, análise e remoção de gargalos. O modelo de processo de negócio é obtido a partir da execução da tarefa de descoberta de modelos de instâncias de processo. De forma mais precisa, o método proposto visa otimizar um modelo de processo de negócio com base na redução do *lead-time*, tomando como ponto de partida o caminho crítico do referido processo de negócio alvo. Neste contexto, relembramos também as nossas hipóteses de pesquisa, que são:

- H1. A extração de informações de *log* de eventos de um processo de negócio semiestruturado, por meio das tarefas de mineração de processos, é suficiente alimentar um método computacional para reduzir o *lead-time* de um processo de negócio com baixo esforço e baixa sobrecarga cognitiva.
- H2. As informações geradas pela técnica PERT/CPM—média ponderada de cada atividade e caminho crítico—permitem gerar uma estratégia eficiente de identificação de atividades significativas em um processo de negócio.

Na direção das hipóteses acima postuladas, a pergunta é a seguinte: como explorar informações geradas em tornos de um modelo de processo—descoberto automaticamente a partir de *logs* de eventos—e pelas técnicas PERT/CPM para compor um método computacional para reduzir o *lead-time* de um processo de negócio, com baixo esforço e baixa sobrecarga cognitiva?

## 4.3 Estratégia de Pesquisa

Esta seção apresenta as etapas que foram desenvolvidas para alcançar o objetivo estabelecido. As subseções a seguir detalham de cada etapa da Tabela 16, onde estão contidos os assuntos relacionados ao: i) estudo que visa selecionar técnicas para auxiliar na identificação do caminho crítico e dos gargalos em um modelo de processo de negócio descoberto automaticamente; e ii) experimento que visa responder à questão de pesquisa por meio do planejamento, do desenvolvimento do Método- $\beta$ , da coleta e da análise dos dados.

Tabela 16. Etapas do experimento

<b>Etapa</b>	<b>Descrição</b>
<b>Etapa 1</b>	Selecionar técnicas para calcular o caminho crítico de um modelo de processo de negócio, assim como para localizar e identificar gargalos.
<b>Etapa 2</b>	Desenvolver o Método- $\beta$ e construir o ambiente experimental.
<b>Etapa 3</b>	Coletar dados do experimento.
<b>Etapa 4</b>	Avaliar os resultados em termos de desempenho e simplicidade—redução do <i>lead-time</i> alvo com o menor número de alterações no modelo original.

#### 4.3.1 Medidas de similaridade e cálculo de eficiência de redução do *lead-time*

O cálculo de similaridade entre o modelo descoberto original e do modelo modificado é feito por meio da confrontação de grafos de dependência, usada para verificar a conformidade entre dois modelos de processos. Em seguida, aplica-se a fórmula para calcular o coeficiente de cosseno e assume-se o resultado desta operação como o grau de similaridade dos modelos de processos em questão. Para isto, cada grafo de dependência é dividido em vetores, um com as atividades e outro com as transições. Essa estratégia permite medir a similaridade entre modelos, utilizando para isso, os vetores gerados e os pesos atribuídos (cf. Seção 2.3).

O cálculo de eficiência de redução do *lead-time* é feito pela comparação entre o número de atividades do modelo original e número de atividades do modelo modificado.

#### 4.3.2 Identificação e seleção das técnicas para calcular o caminho crítico e gargalos

Esta etapa visa identificar as técnicas utilizadas para descobrir o caminho crítico de um modelo de processo de negócio, bem como os gargalos existentes neste caminho, além de descobrir o tempo esperado para cada atividade.

A revisão da literatura nos leva as técnicas PERT/CPM. A CPM é usada como elemento de base para localizar e identificar o caminho mais longo em rede de projeto, levando em consideração os tempos de cada atividade. Essa mesma abordagem clássica nos permite, com alguns ajustes, obter a partir de um modelo de processo representado na forma

de grafo—com nós e arestas—informações relevantes sobre as atividades críticas; cada nó é uma atividade. A PERT pode ser usada para calcular o tempo esperado de cada atividade com base nas estimativas de tempo *otimista*, *realista* e *pessimista*.

### 4.3.3 Desenvolvimento do método e construção do ambiente experimental

O esquema da Figura 12 apresenta as tarefas para execução passo-a-passo de cada experimento. Três grandes atividades são indicadas. A primeira consiste em descobrir e armazenar modelos de processos. A segunda consiste em *otimizar* cada modelo de processo descoberto. A terceira consiste em comparar o modelo de processo descoberto e o modelo de processo *otimizado*. As grandes atividades também representam os desafios de desenvolvimento para compor o ambiente de experimentação. Salientamos que a localização e identificação de gargalos são altamente relevantes para o método proposto e representam o ponto de concentração de esforços.

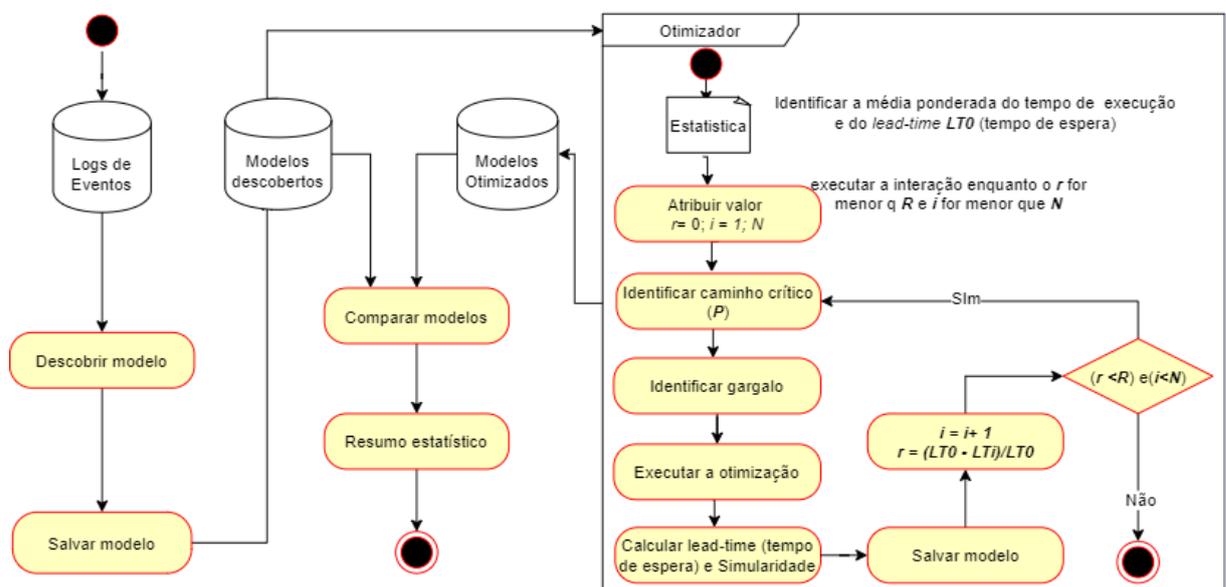


Figura 12 - Estrutura do método.

A Figura 12 mostra esquematicamente o encadeamento dos elementos envolvidos na proposta. Inicialmente, (1) a mineração de processos que permite, por meio da tarefa de descoberta de modelo de instâncias de processo, obter informações estatísticas básicas de um processo negócio como: duração mínima, média e máxima de cada atividade. (2) A comparação do modelo de instâncias de processo descoberto original/alvo e do modelo de

instâncias de processo *otimizado* é feita por meio de medidas de distâncias de grafos, utilizando o coeficiente de cosseno para gerar um grau de similaridade, normalmente usada como uma das formas de verificação de conformidade entre modelos de processos. (3) A técnica PERT/CPM permite calcular a média ponderada de duração de cada atividade e o caminho crítico. Este último dá a direção para identificar de forma efetiva a(s) atividade(s) relevante(s) *vis-à-vis* o foco de interesse de análise; i.e., atividades com gargalos no caminho crítico do processo de negócio. A quantificação da severidade do gargalo de uma dada atividade leva em consideração a soma do tempo de espera, sempre observando a seguinte condição: caso a frequência absoluta do caminho crítico seja menor que dois, a próxima fase do método, não é realizada.

Conforme mostrado na Figura 12, a parte relativa à *otimização* do Método- $\beta$  implementa um comportamento iterativo. O método parte da indicação de um *log* de eventos inicial  $E_0$ , uma taxa de redução desejada  $R$  para o *lead-time* do modelo de processo de negócio  $M_0$  e um número máximo de iteração  $N$ . As etapas do método são as seguintes:

1. descobrir modelo de processo de negócio  $M_0$  a partir do *log* de eventos  $E_0$ ;
2. calcular capacidade de execução de cada atividade do modelo de processo  $M_0$ —assume-se como capacidade o tempo médio ponderado de cada atividade (seguindo modelo PERT/CPM);
3. calcular caminho crítico  $P_0$  do modelo  $M_0$ ;
4. calcular *lead-time*  $LT_0$  do modelo  $M_0$  com  $P_0$ ;
5.  $r = 0$ ;  $i = 1$ ;  $S_0 = 100\%$ ,  $RP = \{<M_0, S_0, r>\}$  #  $S_0$  similaridade em de  $M_0$  com  $M_0$ ,  $r =$  redução do  $LT_0$ ;
6. enquanto ( $r < R$ ) e ( $i < N$ ) faça:
  - encontrar no modelo  $M_{i-1}$  a atividade  $A$  com maior gargalo;
  - duplicar capacidade da atividade  $A$ , gerando  $A_1$  e  $A_2$ , onde  $2|A| = |A_1| + |A_2|$ ;
  - substituir, no *log* de eventos  $E_{i-1}$ , atividade  $A$  pelas atividades  $A_1$  e  $A_2$ , gerando um novo *log* de eventos  $E_i$ ;
  - descobrir modelo de processo  $M_i$  a partir do *log* de eventos  $E_i$ ;
  - calcular similaridade  $S_i$  a partir de  $M_0$  e  $M_i$ ;
  - calcular caminho crítico  $P_i$  do modelo  $M_i$ ;
  - calcular *lead-time*  $LT_i$  do modelo  $M_i$  com  $P_i$ ;
  - $r = (LT_0 - LT_i) / LT_0$ ;
  - $i = i + 1$ ;
  - $RP = RP = \{<M_i, S_i, r>\}$ ;
7. retornar  $RP$ .

A avaliação do resultado do método está no cálculo da diferença entre  $M_0$  de cada modelo  $M_i$  para verificar a eficiência da redução do *lead-time*.

#### 4.3.4 Coleta de dados

Nesta etapa foi realizada a coleta de dados. A fonte de dados é a saída da execução da tarefa de descoberta de modelos de processos aplicada em 4 bases de dados (cf. Tabela 17). Exceto a primeira base de dados da referida tabela, que foi criada artificialmente, todas as demais bases de dados contêm conjuntos/*logs* de eventos de processos reais, que foram utilizados em vários estudos na área de mineração de processos.

**Tabela 17. Base de dados.**

Segmento	Processo	N. Casos	N. Atividades	N. Eventos	Ano de Criação
<b>Artificial</b>	B1: Log Artificial [Autor]	10	6	60	2020
<b>Saúde</b>	B2: Casos de Sepses [36]	1.000	16	8.577	2016
<b>Saúde</b>	B3: Faturamento Hospitalar [37]	100.000	19	451.359	2016
<b>Industria</b>	B4: Log de Produção [39]	225	56	4.543	2014

Para realizar a descoberta do modelo de processo a partir de cada base de dados ou *log* de eventos constante na tabela anterior, foi utilizada a ferramenta de mineração de processos UpFlux [34], que implementa uma versão paralela do algoritmo *Fuzzy Miner* [22]. O resultado da aplicação da implementação do *Fuzzy Miner*, sobre tais logs de eventos, é um conjunto de modelos de processos  $\Pi$ , estruturados na forma de grafos de dependências causais, com informações sobre as atividades, suas sequencias, e um resumo estatístico.

Dado o conjunto de modelos de processos descobertos  $\Pi$ , o próximo passo foi a aplicação do Método- $\beta$ . Os dados de interesse são:

- a) o número de interações que o método/algoritmo foi executado para reduzir o *lead-time* do processo em questão, tomando como entrada uma taxa (em %) de redução alvo denotada no método por  $R$ ;
- b) o número de atividades do caminho crítico, duração do caminho crítico, atividade crítica—ou com gargalo—; e
- c) a similaridade entre o modelo original e os otimizados.

Definida a estruturação da pesquisa e realizada todas as etapas necessárias para a implementação e validação do Método- $\beta$ , o Capítulo seguinte apresenta os resultados obtidos pela aplicação do método.

## 5 Avaliação dos Resultados

Esta seção consolida os resultados obtidos a partir da aplicação do Método- $\beta$  em quatro bases de dados (cf. Tabela 17), sendo uma base de dados com *log* de eventos artificiais e três de processos reais. Para cada base de dados analisou-se o percentual de otimização do *lead-time* e à similaridade do modelo em otimização para cada iteração, de um total de 20 interações. A análise também levou em consideração o número de atividades adicionais que foram criadas para os novos modelos *otimizados*, comparando-as ao percentual de redução do *lead-time*.

As subseções a seguir apresentaram os detalhes dos experimentos realizado tomando como ponto de partida cada base de dados ou *log* de eventos, a fim de mostrar a eficácia do método. A primeira subseção apresenta uma base de dados artificial, criada manualmente, com o objetivo de validar e mostrar passo a passo a execução do método. As demais seções apresentam os resultados de forma sucinta.

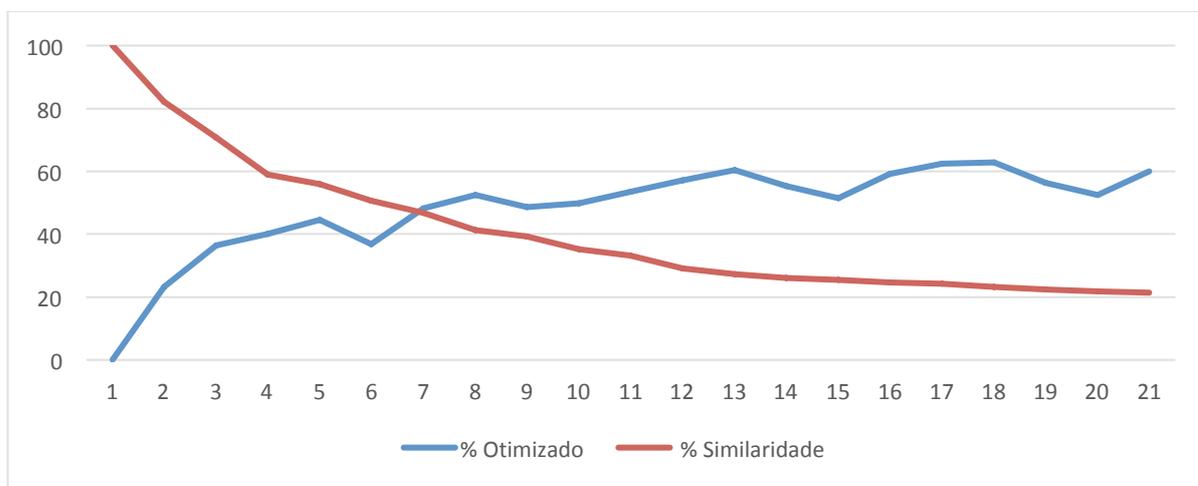
### 5.1 B1: Log artificial

Esta seção apresenta os resultados do experimento tomando como ponto de partida uma base de dados artificial. Tal base possui dez casos e sessenta eventos. Os eventos representam um processo de negócio simples composto por apenas seis atividades (*Production-Order*, C, E, F, H, *Expedition*). A Tabela 18 resume os valores de redução de *lead-time* (*em minutos*), otimização (em %) e similaridade (em %).

**Tabela 18. Resultados da otimização do *lead-time* – log de eventos artificiais.**

Iteração	<i>Lead-time</i> (em min.)	% Otimizado	% Similaridade
0	7.661,00	0,00	100
1	5.875,50	23,31	82,14
2	4.865,17	36,49	70,82
3	4.595,17	40,02	58,91
4	4.253,79	44,47	55,93
5	4.833,79	36,90	50,60
6	3.963,79	48,26	46,70
7	3.643,79	52,44	41,31
8	3.936,29	48,62	39,22
9	3.838,79	49,89	35,29
10	3.558,79	53,55	33,12
11	3.288,79	57,07	29,16
12	3.031,29	60,43	27,30
13	3.416,26	55,41	26,05
14	3.714,26	51,52	25,37
15	3.122,26	59,24	24,75
16	2.870,26	62,53	24,16
17	2.840,26	62,93	23,13
18	3.347,96	56,30	22,32
19	3.647,54	52,39	21,85
20	3.057,13	60,09	21,41

Os valores da Tabela 18 mostram uma oscilação para cima e para baixo. Isso ocorre devido a duplicação da capacidade de execução da atividade crítica e em algumas vezes transfere o problema de gargalo para a atividade subsequente. Todavia, pode-se perceber já com poucas iterações certa estabilidade, bem como um ponto de cruzamento entre a redução de *lead-time* e o grau de similaridade entre o modelo de processo original e o modelo de processo melhorado (cf. Figura 13).



**Figura 13 – Redução do *lead-time* e similaridade entre modelos – log de eventos artificiais.**

Uma utilização importante da análise de redução do *lead-time* está diretamente ligada ao custo do tempo de entrega de um produto ou serviço. A nossa proposta é simples e o esforço para gerar as alternativas é mínimo, considerando que ele é automatizado. E a aplicação do seu resultado é potencialmente viável na medida que as melhorias são focadas nas atividades críticas presentes no caminho crítico do processo de negócio. Hipoteticamente, se uma melhoria de 20% no *lead-time* do processo analisado é suficiente, a alteração no conjunto de atividades é pequena, considerando que ação de melhoria foi a duplicação da capacidade de apenas uma atividade; além disso, a perturbação no ambiente pode ser também considerada mínima ou local.

Ao analisar esses dois indicadores (cf. Figura 13), pode-se inferir que à medida que o percentual de otimização aumenta a similaridade diminui. Isso é esperado, pois a cada iteração do método a capacidade de execução da atividade crítica é duplicada no modelo otimizado  $M_i$ , e assim em comparação com o modelo original  $M_0$ , quanto maior o percentual de otimização menor é sua similaridade. A elevação da capacidade de execução da atividade crítica é inserida no modelo pela duplicação da atividade, mantendo o rótulo da atividade crítica/alvo e adicionando um novo rótulo para atividade criada. Desta forma, pode-se calcular a similaridade com base nos rótulos das atividades.

O experimento mostrou que o método proposto permitiu otimizar modelo do processo em questão em até 62.93% com 23.13% de similaridade em 20 iterações. Por outro

lado, deve-se também frisar que já na primeira iteração foi possível otimizar o modelo em 23.31% com uma similaridade de 82.14%. Pode-se assumir que o ponto importante de tomada de decisão, considerando os valores de otimização do modelo e dissimilaridade, se encontra entre as iterações 6 e 7, quando as curvas se cruzam. Deve-se notar que não foi levado em conta o custo de elevação da capacidade de execução da atividade crítica. Em outras palavras, assume-se que: como atividade é crítica *vis-à-vis* ao *lead-time*, o custo é menos relevante. Além disso, o custo deve ser minimizado dado que o método propõe ajustes localizados.

## 5.2 B2: Casos de Sepses

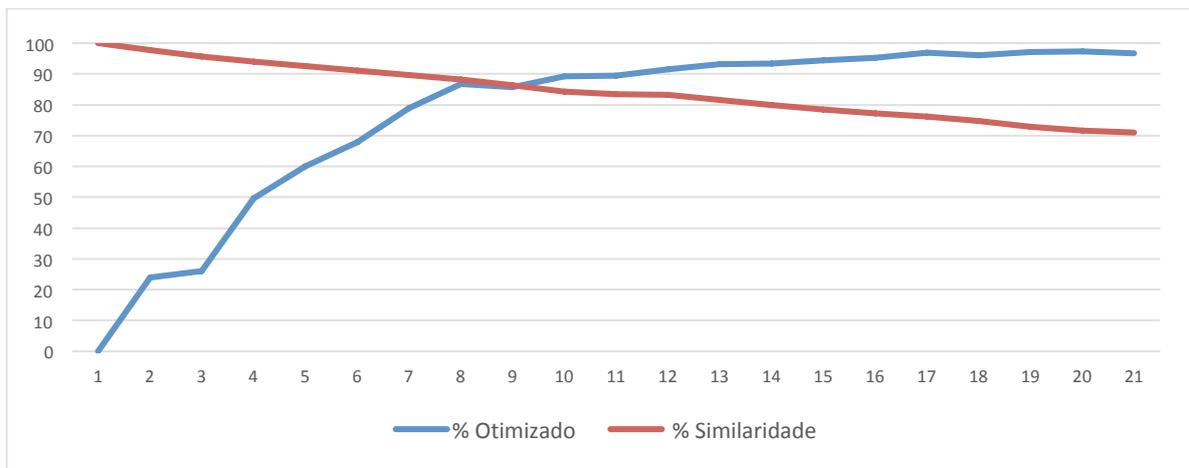
Esta seção apresenta os resultados de experimento tomando como ponto de partida uma base de dados da área da saúde contendo logs de eventos de casos de sepse. Essa base possui cerca de 1.000 casos, 15.000 eventos e 16 atividades distintas. A Tabela 19 mostra os resultados obtidos em cada iteração de Método- $\beta$ .

**Tabela 19. Resultados da otimização do *lead-time* – casos de sepse.**

Iteração	<i>Lead-time</i> (em min.)	% Otimizado	% Similaridade
0	162.975,97	0,00	100
1	124.063,61	23,88	97,70
2	120.698,74	25,94	95,72
3	82.152,56	49,59	93,92
4	64.986,69	60,12	92,58
5	52.220,95	67,96	91,14
6	34.560,50	78,79	89,66
7	21.431,24	86,85	88,19
8	23.251,43	85,73	86,45
9	17.493,12	89,27	84,30
10	17.229,22	89,43	83,40
11	13.680,86	91,61	83,19
12	11.199,73	93,13	81,52
13	10.919,98	93,30	79,94

14	9.116,41	94,41	78,58
15	7.676,61	95,29	77,23
16	5.138,66	96,85	76,14
17	6.250,21	96,16	74,66
18	4.782,27	97,07	72,94
19	4.168,71	97,44	71,67
20	5.302,85	96,75	70,92

Com base nos dados da Tabela 19, a Figura 14 mostra que a evolução do processo de melhoria/otimização, nesta base de dados, é similar no tocante ao cruzamento das linhas de otimização e similaridade, elas se cruzaram próxima da 9ª iteração.



**Figura 14 – Redução do lead-time e similaridade entre modelos – casos de sepse.**

O experimento mostrou que é possível otimizar o modelo de processo original para os casos de sepse em até 97,97% com 70,92% de similaridade em 20 iterações. Trata-se de uma taxa de otimização considerável, levando-se em conta o percentual de similaridade é de 70,92%. Deve-se também destacar que já na primeira iteração, com apenas uma modificação no modelo, foi possível uma otimização de 24,28% e mantendo uma similaridade de 97,70% com o modelo de processo original. As curvas da Figura 12 mostram que o melhor compromisso entre modelo otimizado e dissimilaridade se encontra entre as iterações 8 e 9.

### 5.3 B3: Faturamento Hospitalar

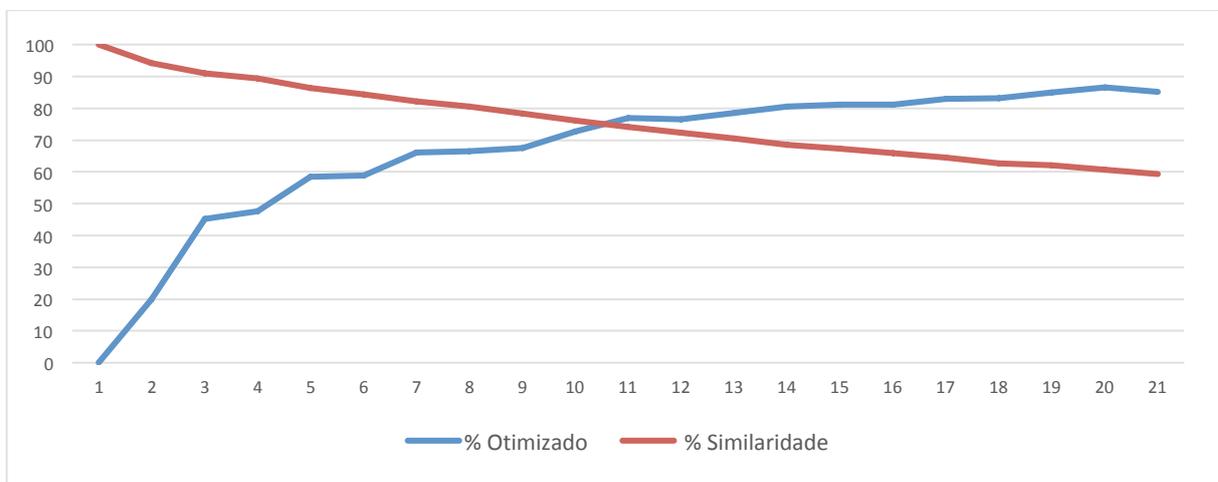
Esta seção apresenta os resultados de um experimento tomando como ponto de partida uma base de dados da área da saúde, em que os *logs* de eventos são relativos a faturamento de serviços médico-hospitalar. A base de dados possui cerca de 100.000 casos, 451.359 eventos e 19 atividades. A Tabela 20 mostra os resultados obtidos em cada iteração de Método- $\beta$ .

**Tabela 20. Resultados da otimização do *lead-time* – casos de faturamento médico-hospitalar.**

Iteração	<i>Lead-time</i> (em min.)	% Otimizado	% Similaridade
0	1.484.799,00	0,00	100,00
1	1.189.557,81	19,88	94,20
2	812.333,96	45,29	90,94
3	777.125,83	47,66	89,41
4	615.637,96	58,54	86,41
5	609.507,24	58,95	84,40
6	502.171,17	66,18	82,13
7	497.894,75	66,47	80,61
8	483.783,59	67,42	78,27
9	404.075,42	72,79	76,19
10	341.985,51	76,97	74,07
11	348.543,20	76,53	72,27
12	319.824,25	78,46	70,50
13	288.056,34	80,60	68,54
14	280.120,19	81,13	67,36
15	280.120,19	81,13	65,95
16	251.732,33	83,05	64,48
17	249.729,77	83,18	62,78
18	222.713,86	85,00	62,09
19	198.947,87	86,60	60,63
20	219.827,98	85,19	59,33

A Figura 15 apresenta graficamente os resultados detalhados na Tabela 19.

Com base nos dados da Tabela 20, a Figura 15 mostra que a evolução do processo de melhoria/otimização, nesta base de dados, é similar ao que aconteceu no experimento anterior (cf. Figura 14), no tocante ao cruzamento das linhas de otimização e similaridade, elas se cruzaram na 10ª iteração.



**Figura 15 – Redução do lead-time e similaridade entre modelos – faturamento hospitalar.**

O experimento mostrou que é possível otimizar o modelo de processo original para os casos de faturamento médico-hospitalar em até 86,60% com 60,63% de similaridade em 20 iterações. Deve-se também frisar que já na primeira iteração, com apenas uma única modificação no modelo original ou de base, foi possível otimizar 19,88% e mantendo uma similaridade de 94,20%. E as curvas da Figura 15 mostram que o melhor compromisso entre modelo otimizado e dissimilaridade se encontra entre as iterações 10 e 11.

#### 5.4 B4: Log de Produção

Esta seção apresenta os resultados de um experimento tomando como ponto de partida uma base de dados da área industrial, contendo *logs* de eventos de uma linha de produção. Essa

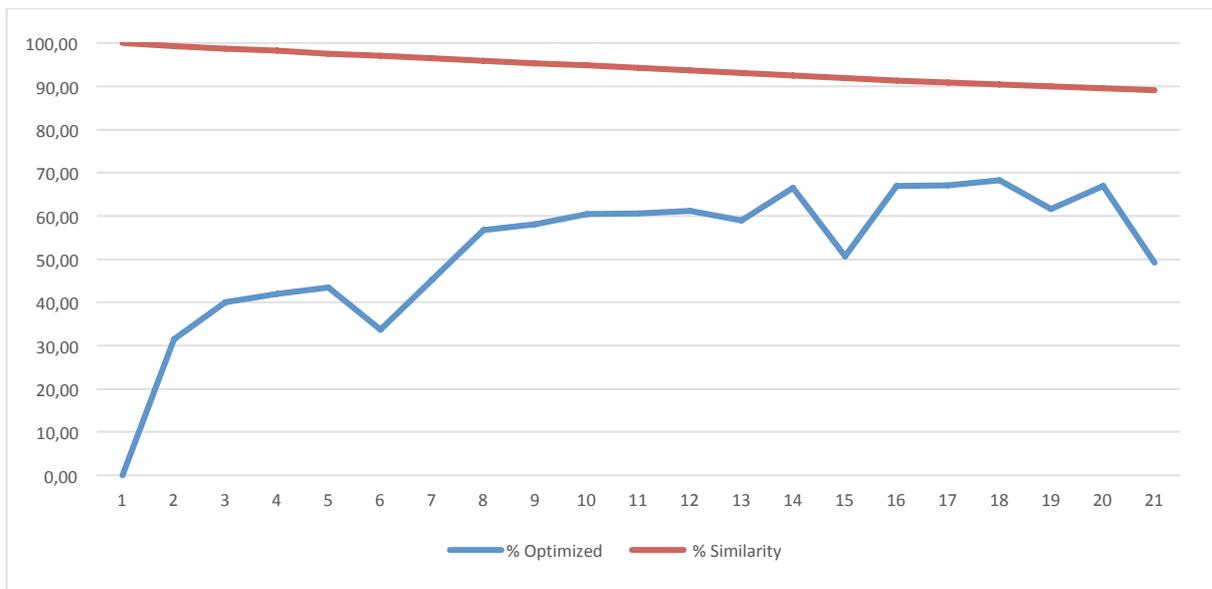
base de dados possui cerca de 225 casos, 4.543 eventos e 56 atividades. A Tabela 21 mostra os resultados obtidos em cada iteração de Método- $\beta$ .

**Tabela 21. Resultados da otimização do lead-time – casos de produção.**

Iteração	Lead-time (em min.)	% Otimizado	% Similaridade
0	97.370,44	0,00	100
1	66.716,67	31,48	99,34
2	58.426,03	40,00	98,66
3	56.438,61	42,04	98,19
4	55.027,64	43,49	97,54
5	64.529,02	33,73	97,07
6	53.327,70	45,23	96,46
7	42.088,84	56,77	95,89
8	40.837,11	58,06	95,31
9	38.536,64	60,42	94,86
10	38.336,67	60,63	94,23
11	37.761,01	61,22	93,64
12	39.907,79	59,01	93,09
13	32.661,46	66,46	92,52
14	47.965,79	50,74	91,91
15	32.176,34	66,95	91,34
16	32.032,14	67,10	90,91
17	30.969,03	68,19	90,39
18	37.375,91	61,61	89,91
19	32.241,04	66,89	89,59
20	49.495,47	49,17	89,07

Com base nos dados da Tabela 21, a Figura 16 apresenta graficamente o comportamento do algoritmo e sua evolução.

Com base nos dados da Tabela 21, a Figura 16 mostra que a evolução do processo de melhoria/otimização, nesta base de dados, é similar ao que aconteceu no experimento anterior (cf. Figura 15), no tocante a otimização, dado que os melhores valores aparecem na 10ª iteração. Deve-se, no entanto, observar que para o número de 20 iterações as linhas de otimização e similaridade não se cruzam.



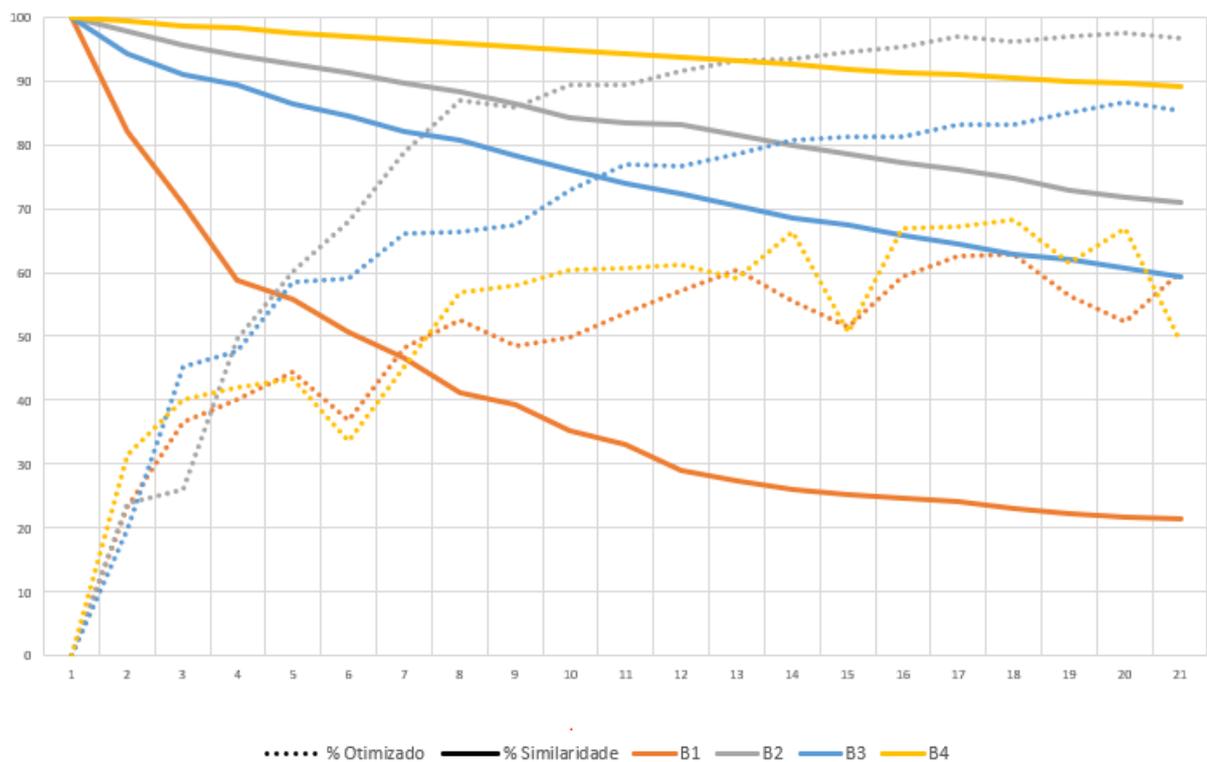
**Figura 16 – Redução do lead-time e similaridade entre modelos – log de produção.**

O experimento mostrou que é possível otimizar o modelo de processo original em até 68,19% com 90,39% de similaridade. Deve-se também frisar que já na primeira iteração, com apenas uma modificação no modelo, foi possível obter uma taxa otimização de 31,48% e mantendo uma similaridade de 99,34%. A Figura 16 mostra que ao realizar as 20 iterações, ou seja, mesmo depois de alterar o modelo por 20 vezes, a sua similaridade quase não foi alterada ficando próximo aos 90%. Obviamente, deve-se ponderar tal valor elevando de similaridade, pois ele se manteve alta em função do número de atividades, que é significativamente superior aos demais casos analisados. E as curvas da Figura 16 mostram que o melhor compromisso entre modelo otimizado e dissimilaridade se encontra entre as iterações 8 e 10. Embora, a mais expressiva taxa de otimização tenha ocorrido na iteração 17, com o valor de 68,19%, percebe-se que existe certa estabilidade em termos valores de otimização a partir da 9ª ou 10ª iteração.

## 5.5 Discussão dos Resultados

Este capítulo centrou-se na realização dos experimentos com quatro bases de dados distintas, em que foram apresentadas as análises e discussões detalhadas dos resultados com o objetivo de responder as questões de pesquisa elaboradas.

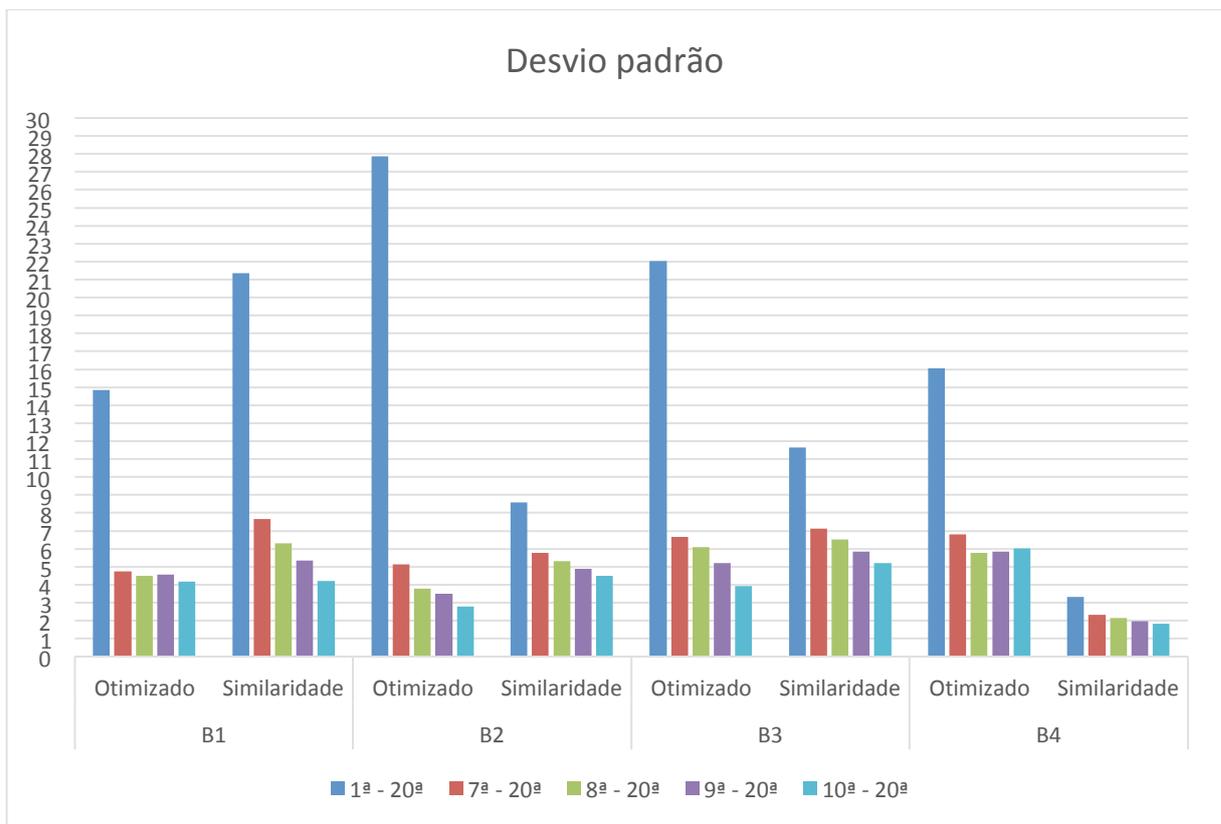
O gráfico da Figura 17 apresenta os resultados de todas as bases de dados, em que é possível verificar uma tendência, como já comentado em cada experimento individualmente, que a medida que avança na otimização do *lead-time* do processo  $M_{t+1}$  a similaridade sua diminui com relação a sua original  $M_0$ .



**Figura 17 – Redução do lead-time e similaridade entre modelos – todas as bases.**

Analisando o gráfico da Figura 17, também é possível observar, com relação a otimização vs. similaridade, que as curvas para as bases de dados B1, B2 e B3 tiveram um comportamento similar, visto que o cruzamento das linhas ocorreu entre a 7<sup>a</sup> e 11<sup>a</sup> iteração e apenas com o experimento com a base de dados B4 que não ocorreu tal cruzamento; ele não ocorreu devido ao alto número de atividades presente neste processo.

Com o objetivo de verificar o comportamento padrão de execução do método e sua estabilidade, o gráfico da Figura 18 apresenta para cada experimento—associado a cada base de dados—o desvio padrão da otimização e similaridade para todas as iterações—da 1<sup>a</sup> até 20<sup>a</sup>—e também os valores obtidos entre intervalos de iterações, a saber: [7<sup>a</sup>, 20<sup>a</sup>], [8<sup>a</sup>, 20<sup>a</sup>], [9<sup>a</sup>, 20<sup>a</sup>] e [10<sup>a</sup>, 20<sup>a</sup>].



**Figura 18 – Desvio padrão da otimização e similaridade – estabilidade.**

Ao analisar os resultados da Figura 18, pode-se assumir que a partir da 7<sup>a</sup> iteração ocorre uma certa estabilidade na otimização vs. similaridade, visto que o desvio padrão, com exceção da similaridade do B4, se encontra em  $5 \pm 2$ .

Com esses resultados foi possível observar um potencial significativo do método *vis-à-vis* na redução do *lead-time* em modelos de processos, mantendo controlado o número de ajustes no modelo original. I.e., levando sempre em consideração o menor número de alterações na estrutura do processo original, realizado com base no caminho crítico e na identificação da atividade com maior gargalo.

## 6 Conclusões

A identificação automática de gargalos em processos de negócio é altamente relevante para os gestores sugerir melhorias efetivas. No contexto deste projeto, o foco da melhoria está centrado na redução de *lead-time* de processos de negócio. E a estratégia principal de melhoria tomou como ponto de partida o caminho crítico e sua eficácia permitiu reduzir o *lead-time* de processo de negócio, provocando poucas perturbações localmente no processo analisado.

A tarefa de descoberta de modelo de processo permitiu gerar informações estatísticas básicas, como por exemplo, duração mínima, média e máxima de cada atividade. Tais informações viabilizaram a aplicação da técnica PERT/CPM para calcular a média ponderada do tempo de execução esperado de cada atividade e o caminho crítico. Este último deu a direção para identificar a atividade relevante *vis-à-vis* o foco de interesse de análise. O resultado foi uma abordagem simples que permite recalcular parcialmente a rede de atividades e obter rapidamente informações de melhoria em termos de *lead-time*.

Ao analisar os resultados de cada experimento, pode-se concluir que o método proposto neste trabalho teve um desempenho significativo quanto à otimização do *lead-time*, i.e., com foco na redução do *lead-time*. Com poucas alterações—aumento da capacidade de execução da atividade crítica—nos modelos de processo foi possível observar uma otimização do *lead-time* considerável em todos os experimentos, sem comprometer exageradamente o modelo de processo original.

### ***Trabalhos futuros***

A duplicação da capacidade de execução da atividade crítica de um processo de negócio revelou-se bastante satisfatória. Contudo, quando uma certa atividade—de um certo processo—é executada por um recurso escasso e/ou especializado, normalmente, a demanda é por um aumento da execução inferior a 100% da capacidade atual. Assim, o principal trabalho futuro, pretende avaliar outras estratégias, além da duplicação da capacidade de execução de uma atividade, permitindo aumentar a capacidade de execução de uma atividade de forma

fracionada. Um segundo trabalho futuro encerra o cálculo do fluxo máximo sobre o caminho crítico a fim de identificar previamente se o aumento da capacidade da atividade crítica não irá gerar gargalos nas atividades subsequentes.

## Referências

- [1] Van der Aalst, W. (2011). *Process Mining – discovery, conformance and enhancement of business processes*. Springer.
- [2] Karla A. de Medeiros, Ana & Weijters, A & Aalst, Wil M. P.. (2005). *Genetic Process Mining: A Basic Approach and Its Challenges..* 203-215.
- [3] Vergidis, Kostas & Tiwari, Ashutosh & Majeed, Basim. (2008). *Business Process Analysis and Optimization: Beyond Reengineering. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on.* 38. 69 - 82. 10.1109/TSMCC.2007.905812.
- [4] Dias, A. A. S. (2015). *Modelo de monitorização da execução de processos aplicado ao sector da Saúde. Dissertação em Engenharia e Gestão Industrial.*
- [5] M. Aghdasi and S. E. Malihi, "Rule based business process optimization," 2010 IEEE International Conference on Industrial Engineering and Engineering Management, Macao, 2010, pp. 305-309.
- [6] W. van der Aalst, *Process Mining*, vol. 5. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016.
- [7] Vanhoucke, Mario. (2012). *Project management with dynamic scheduling: Baseline scheduling, risk analysis and project control.* 10.1007/978-3-642-25175-7.
- [8] Yi Xie, Chen-Fu Chien, and Ren-Zhong Tang. 2013. A method for estimating the cycle time of business processes with many-to-many relationships among the resources and activities based on individual worklists. *Comput. Ind. Eng.* 65, 2 (June 2013), 194-206.
- [9] Ha, B. H., Reijers, H. A., Bae, J., & Bae, H. (2006). An approximate analysis of expected cycle time in business process execution. In J. Eder, & S. Dustdar (Eds.), *Proceedings of the BPM workshops 2006* (pp. 65-74).

- [10] Aguirre S., Parra C., Alvarado J. (2013) Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach. In: Cudre-Mauroux P., Ceravolo P., Gašević D. (eds) Data-Driven Process Discovery and Analysis. SIMPDA 2012.
- [11] N. Martin, B. Benoît and A. Caris, "Event log knowledge as a complementary simulation model construction input," 2014 4th International Conference On Simulation And Modeling Methodologies, Technologies And Applications (SIMULTECH), Vienna, Austria, 2014, pp. 456-462.
- [12] L. C. Tsironis, D. S. Sfiris and B. K. Papadopoulos, "Fuzzy Performance Evaluation of Workflow Stochastic Petri Nets by Means of Block Reduction," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 40, no. 2, pp. 352-362, March 2010.
- [13] Aalst, Wil M. P. & Schonenberg, Helen & Song, Minseok. (2011). Time prediction based on process mining. *Inf. Syst.*. 36. 450-475. 10.1016/j.is.2010.09.001.
- [14] T. H. Davenport, *Process Innovation: Reengineering Work Through Information Technology*. Boston, MA: Harvard Business School Press, 1993.
- [15] R. Dewan, A. Seidmann, and Z. Walter, "Workflow optimization through task redesign in business information processes," in Proc. Hawaii Int. Conf. Syst. Sci., Jan. 1998, vol. 1, pp. 240–252.
- [16] J. Gramdi, "Elaborating actual lead time with both management and execution data," 2009 International Conference on Computers & Industrial Engineering, Troyes, 2009, pp. 674-677, doi: 10.1109/ICCIE.2009.5223807.
- [17] W. M. P. Van Der Aalst, A. J. M. M. Weijters, and L. Maruster, "Workflow Mining: Which Processes can be Rediscovered," BETA Work. Pap. Ser. WP 74, p. 25, 2002.
- [18] T. Basten and W. M. P. Van Der Aalst, "Inheritance of behavior," *J. Log. Algebr. Program.*, vol. 47, no. 2, pp. 47–145, 2001.
- [19] A. J. M. M. Weijters and W. M. P. van der Aalst, "Rediscovering Workflow Models from Event-Based Data using Little Thumb," *Integr. Comput. Eng.*, vol. 10, pp. 151–162, 2003.

- [20] M V, Manoj & Thomas, Likewin & Basava, Annappa. (2014). Efficient process mining through critical path network analysis. Souvenir of the 2014 IEEE International Advance Computing Conference, IACC 2014. 10.1109/IAdCC.2014.6779377.
- [21] Habibi, F., Birgani, O., Koppelaar, H & Radenović, S. (2018). Using fuzzy logic to improve the project time and cost estimation based on Project Evaluation and Review Technique (PERT). *Journal of Project Management*, 3(4), 183-196.
- [22] C. W. Günther and W. M. P. van der Aalst, "Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics," *Bus. Process Manag. - Lect. Notes Comput. Sci.*, vol. 4714, pp. 328–343, 2007.
- [23] M. Hammer and J. Champy, *Reengineering the Corporation: A Manifesto for Business Revolution*. London, U.K.: Brealey, 1993.
- [24] W. J. Gutjahr, C. Strauss, and E. Wagner, "A stochastic branch-and-bound approach to activity crashing in project management," *INFORMS J. Comput.*, vol. 12, no. 2, pp. 125–135, 2000.
- [25] E. Lamma, P. Mello, M. Montali, F. Riguzzi, and S. Storari, "Inducing declarative logic-based models from labeled traces," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4714 LNCS, pp. 344–359, 2007.
- [26] H.-S. Han, "Business process change design from decision model perspective," *Omega, Int. J. Manage. Sci.*, vol. 9, no. 2, pp. 21–45, 2003.
- [27] M. V. Zelkowitz and D. R. Wallace, "Experimental models for validating technology," *Computer (Long. Beach. Calif.)*, vol. 31, no. 5, pp. 23–31, 1998.
- [28] M. Havey, *Essential Business Process Modelling*. Sebastopol, CA: O'Reilly, 2005.
- [29] J. M. E. van der Werf, B. F. van Dongen, C. A. Hurkens, and A. Serebrenik, "Process Discovery using Integer Linear Programming," *Fundam. Informaticae*, vol. 94, no. 3–4, pp. 387–412, 2009.
- [30] I. Hofacker and R. Vetschera, "Algorithmical approaches to business process design," *Comput. Oper. Res.*, vol. 28, pp. 1253–1275, 2001.

- [31] S. J. J. Leemans, D. Fahland, and W. M. P. Van Der Aalst, “Discovering block-structured process models from event logs - A constructive approach,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7927 LNCS, pp. 311–329, 2013.
- [32] T. Jaeger, A. Prakash, and M. Ishikawa, “A framework for automatic improvement of workflows to meet performance goals,” in *Proc. 6th Int. Conf. Tools Artif. Intell.*, Nov. 1994, pp. 640–646.
- [33] <http://promtools.org/doku.php>
- [34] <https://app.upflux.net/>
- [35] D. Lin, “An Information-Theoretic Definition of Similarity,” *Proc. ICML*, pp. 296–304, 1998.
- [36] <https://data.4tu.nl/repository/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460>.
- [37] <https://data.4tu.nl/repository/uuid:76c46b83-c930-4798-a1c9-be94dfeb741>.
- [38] M. Becker and R. Laue, “A comparative survey of business process similarity measures,” *Comput. Ind.*, vol. 63, no. 2, pp. 148–167, 2012.
- [39] <https://data.4tu.nl/repository/uuid:68726926-5ac5-4fab-b873-ee76ea412399>.
- [40] H. Lin, Z. Zhap, H. Li, and Z. Chen, “A novel graph reduction algorithm to identify structural conflicts,” in *Proc. 35th Hawaii Int. Conf. Syst. Sci.*, 2002, vol. 9, p. 289.
- [41] J. L. Rummel, Z. Walter, R. Dewan, and A. Seidmann, “Activity consolidation to improve responsiveness,” *Eur. J. Oper. Res.*, vol. 161, pp. 683–703, 2005.
- [42] W. Sadiq and M. Orłowska, “Analyzing process models using graph reduction techniques,” *Inf. Syst.*, vol. 25, no. 2, pp. 117–134, 2000.
- [43] F. Soliman, “Optimum level of process mapping and least cost business process re-engineering,” *Int. J. Oper. Prod. Manage.*, vol. 18, no. 9/10, pp. 810–816, 1998.
- [44] A. Tiwari, K. Vergidis, and B. Majeed, “Evolutionary multi-objective optimisation of business processes,” in *Proc. IEEE Congr. Evol. Comput.*, Jul. 2006, pp. 3091–3097.

- [45] W. M. P. van der Aalst, A. Hirsenschall, and H. M. W. Verbeek, “An alternative way to analyse workflow graphs,” in *Lecture Notes in Computer Science*, Springer-Verlag, 2002, vol. 2348, p. 535.
- [46] K. Vergidis, A. Tiwari, and B. Majeed, “Business process improvement using multi-objective optimisation,” *BT Technol. J.*, vol. 24, no. 2, pp. 229–235, 2006.
- [47] Y. Zhou and Y. Chen, “Business process assignment optimisation,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2002, vol. 3, pp. 540–545.
- [48] Y. Zhou and Y. Chen, “The methodology for business process optimised design,” in *Proc. Ind. Electron. Conf. (IECON)*, 2003, vol. 2, pp. 1819–1824.
- [49] Y. Zhou and Y. Chen, “Project-oriented business process performance optimization,” in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2003, vol. 5, pp. 4079–4084.
- [50] R. Lu and S. Sadiq, “On the Discovery of Preferred Work Practice Through Business Process Variants,” in *Conceptual Modeling - ER 2007: 26th International Conference on Conceptual Modeling*, Auckland, New Zealand, November 5-9, 2007. Proceedings, C. Parent, K.-D. Schewe, V. C. Storey, and B. Thalheim, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 165–180.
- [51] M. La Rosa, M. Dumas, R. Uba, and R. Dijkman, “Merging Business Process Models,” *Move to Meaningful Internet Syst. (OTM 2010)*, pp. 96–113, 2010.
- [52] R. Lu and S. Sadiq, “Business Process Management,” *Bus. Process Manag.* 2006, vol. 4102, no. 1, pp. 426–431, 2006.
- [53] W. Sadiq and M. E. Orłowska, “Analyzing process models using graph reduction techniques,” *Inf. Syst.*, vol. 25, no. 2, pp. 117–134, 2000.
- [54] R. Akkiraju and A. Ivan, “Discovering business process similarities: An empirical study with SAP best practice business processes,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6470 LNCS, pp. 515–526, 2010.
- [55] R. Dijkman, M. Dumas, B. Van Dongen, R. Krik, and J. Mendling, “Similarity of business process models: Metrics and evaluation,” *Inf. Syst.*, vol. 36, no. 2, pp. 498–516, 2011.

- [56] J.-Y. J. Y. Jung, J. Bae, and L. Liu, “Hierarchical business process clustering,” Proc. - 2008 IEEE Int. Conf. Serv. Comput. SCC 2008, vol. 2, no. 12, pp. 613–616, 2008.