



*Seleção de características de modelos
de processos usando técnicas de
Inteligência de Enxames*

Renato Barbosa Cirne

Orientador: Fernando Buarque de Lima Neto

University of Pernambuco – UPE
School of Engineering – POLI
Computational Intelligent Research Group – CIRG



At CIRG-UPE Nature insights nurture our creativity

Agenda

- Introdução, Motivações e Objetivos
- Fundamentação Teórica
 - Gestão de Processos, Riscos e *Compliance*
 - *Process Mining* (PM)
 - *Ant Colony Optimization* - ACO
 - Seleção de Características.
- Contribuições
- Resultados
- Conclusões

INTRODUÇÃO, MOTIVAÇÕES E OBJETIVOS

Recife-PE, Brasil, 09/09/2020

Seleção de características de modelos de processos usando técnicas de Inteligência de Enxames

Motivações

- Em média, **5% de receitas perdidas** com atividades fraudulentas (ACFE, 2018);
- **49%** das organizações já foram vítimas de fraudes ou crimes econômicos; (PWC, 2018)
- As inovações e as iniciativas de otimização devem se concentrar nos **processos de alto impacto (20%)** ABPMP (2016);
- **70%** das empresas com Inteligência Artificial - IA até 2030 (McKinsey & Company, 2018);
- **IA pouco aplicada** na área de **compliance** (Pinheiro et al., 2020);

Motivações

- *Process Mining* - PM **relevante** no gerenciamento de processos, em especial, na **geração e análise de modelos** (van der Aalst, 2016);
- PM têm problemas com o grande volume e/ou a complexidade de logs de eventos (van der Aalst (2014));
- Comunidade científica atenta às técnicas de Inteligência de Enxames (IE) para **Seleção de Características – SC** (Xue et al., 2016)

Objetivo Geral

Desenvolver um método de seleção de características na geração de modelos de processos de negócio usando técnicas de Inteligência de Enxames, **Ant Colony Optimization (ACO)**.

Objetivos Específicos

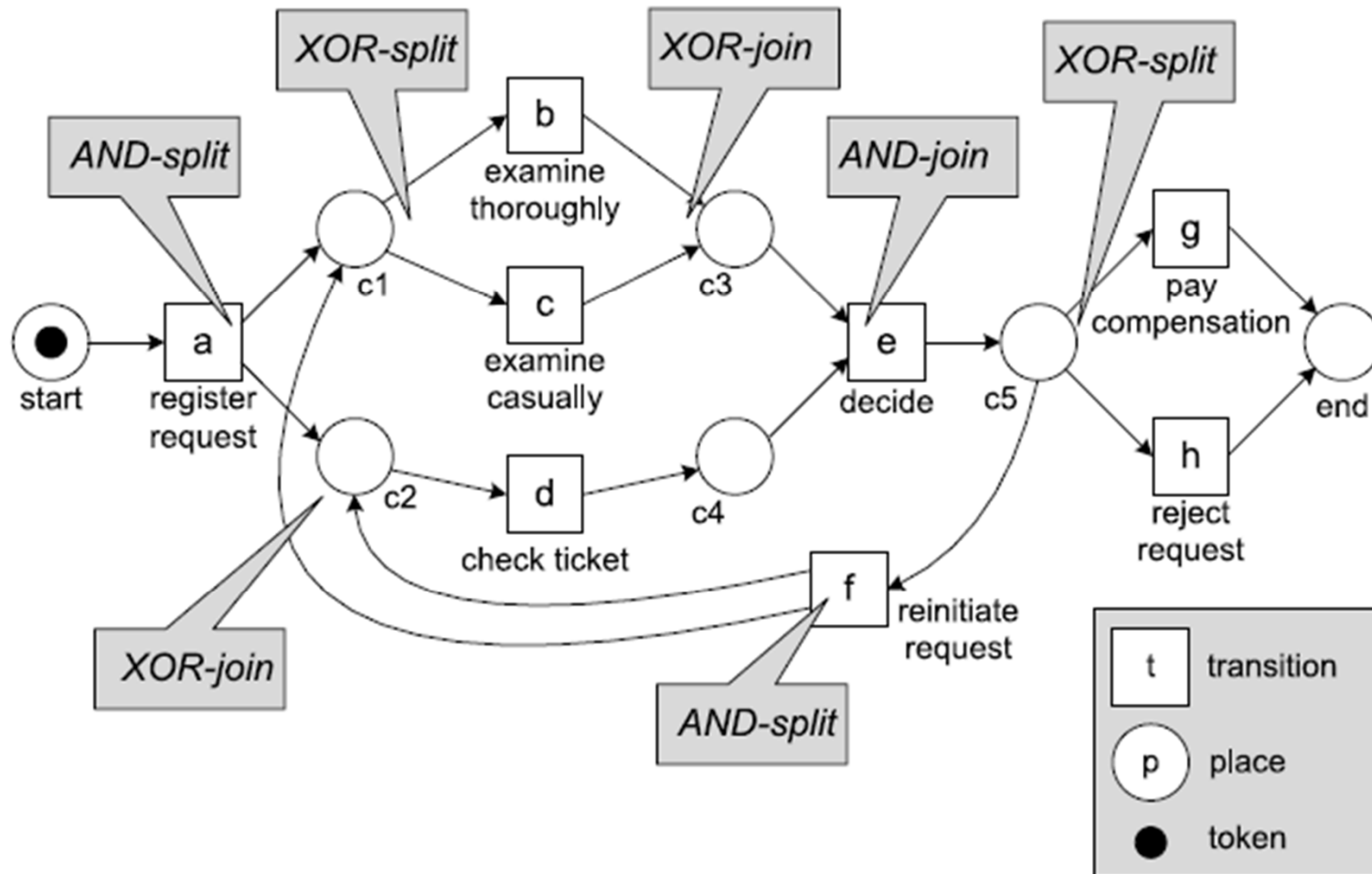
- **Identificar** técnicas de SC mais adequadas ao problema;
- Descobrir **oportunidades** de pesquisa;
- **Desenvolver** método de SC com ACO na descoberta de modelo de processos;
- **Avaliar** os **resultados** quanto à **adequabilidade, simplicidade e precisão** dos modelos descobertos.

FUNDAMENTAÇÃO TEÓRICA

Recife-PE, Brasil, 09/09/2020

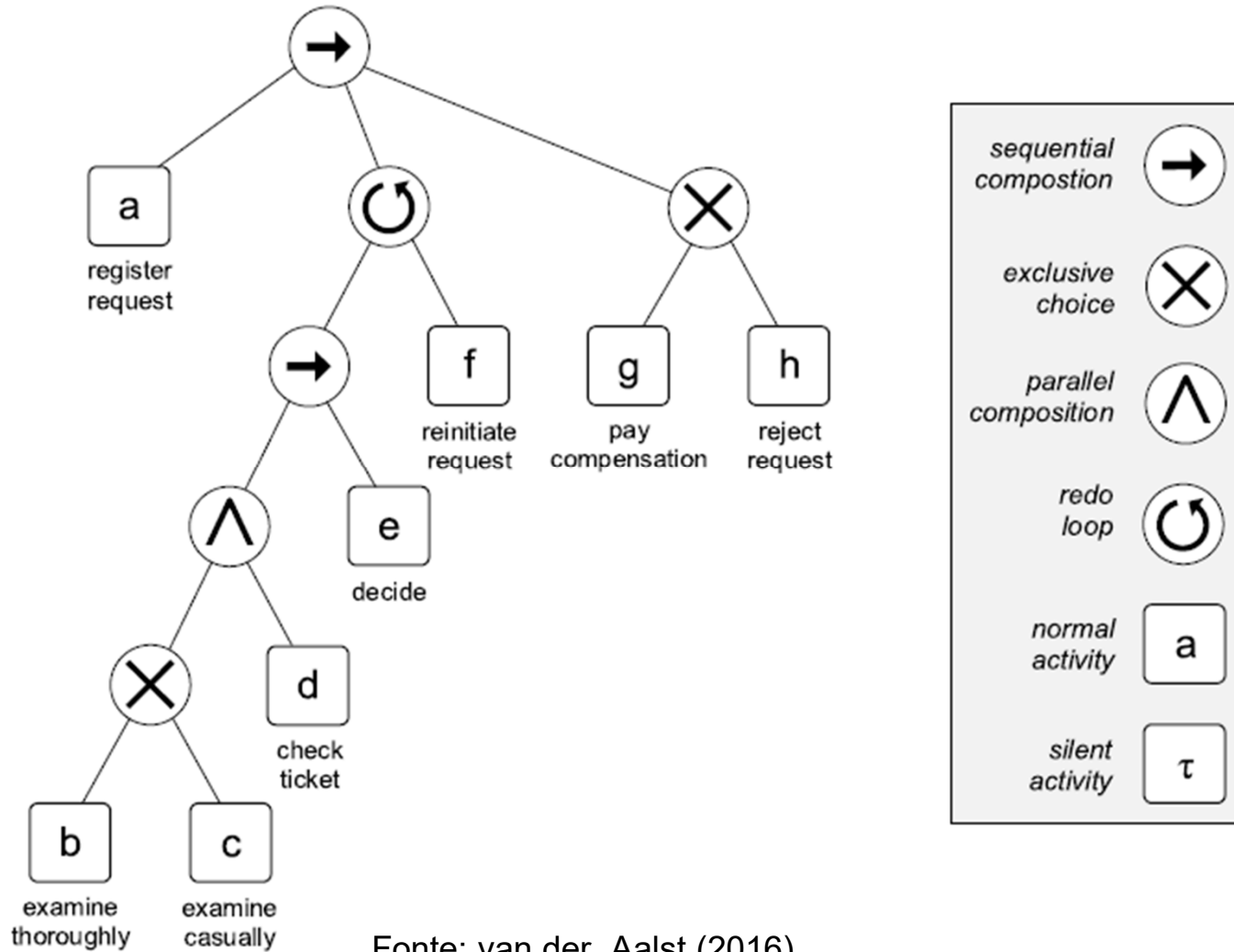
Seleção de características de modelos de processos usando técnicas de Inteligência de Enxames

Gestão de Processos



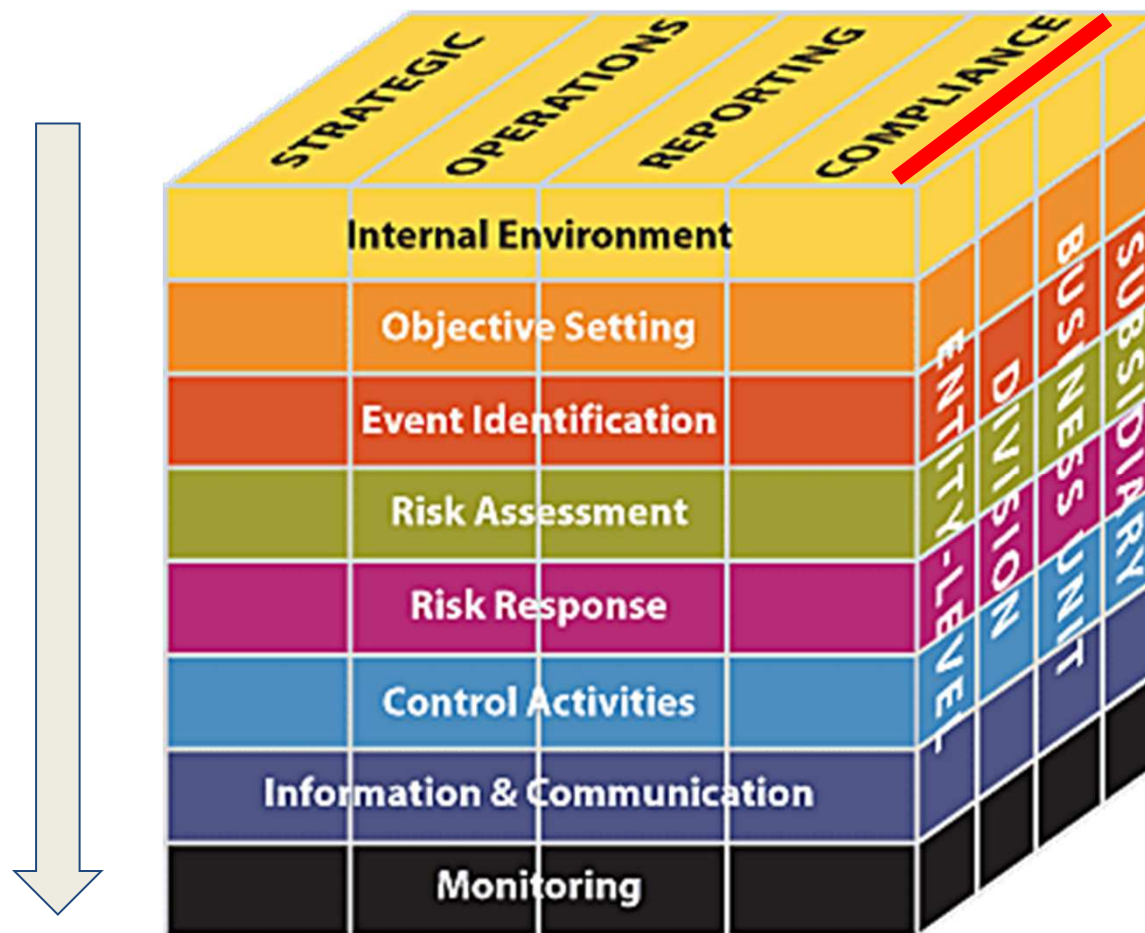
Fonte: van der, Aalst (2016)

Gestão de Processos



Fonte: van der, Aalst (2016)

Gestão de Riscos e *Compliance*



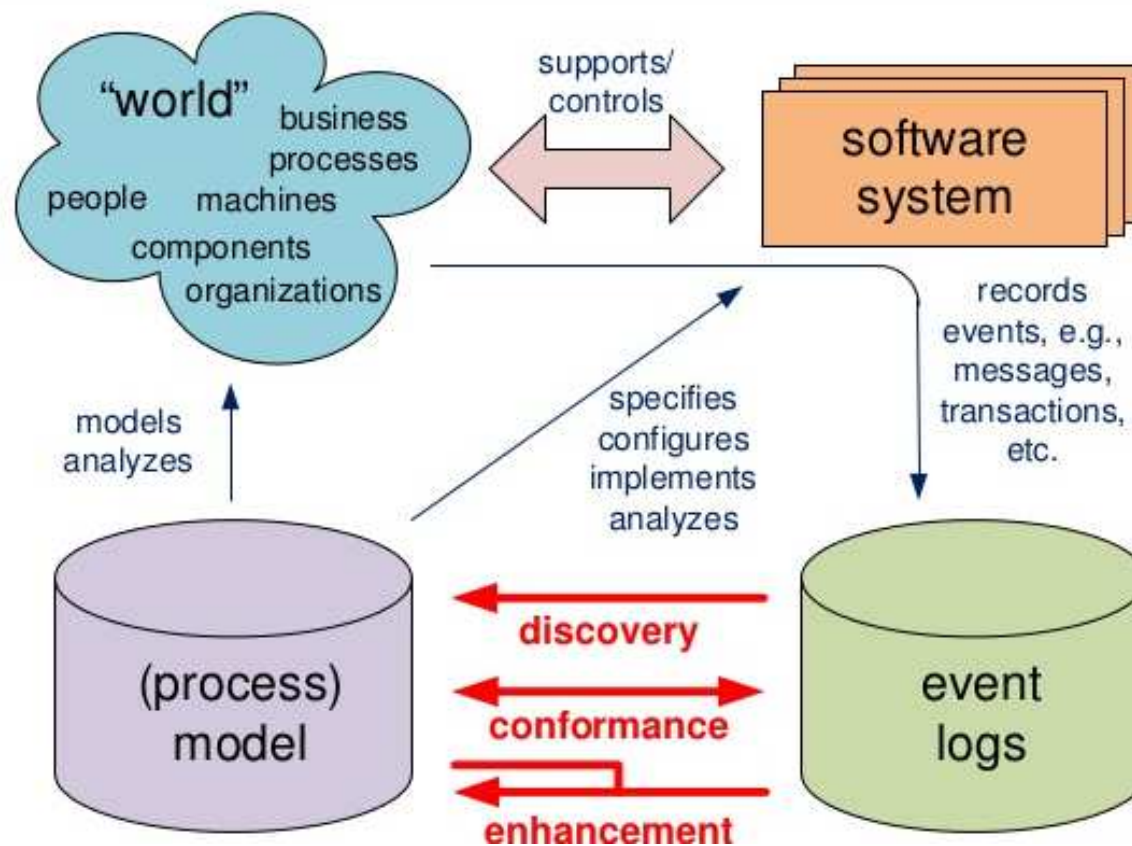
Fonte: COSO.2007.

Inteligência Artificial e *Compliance*



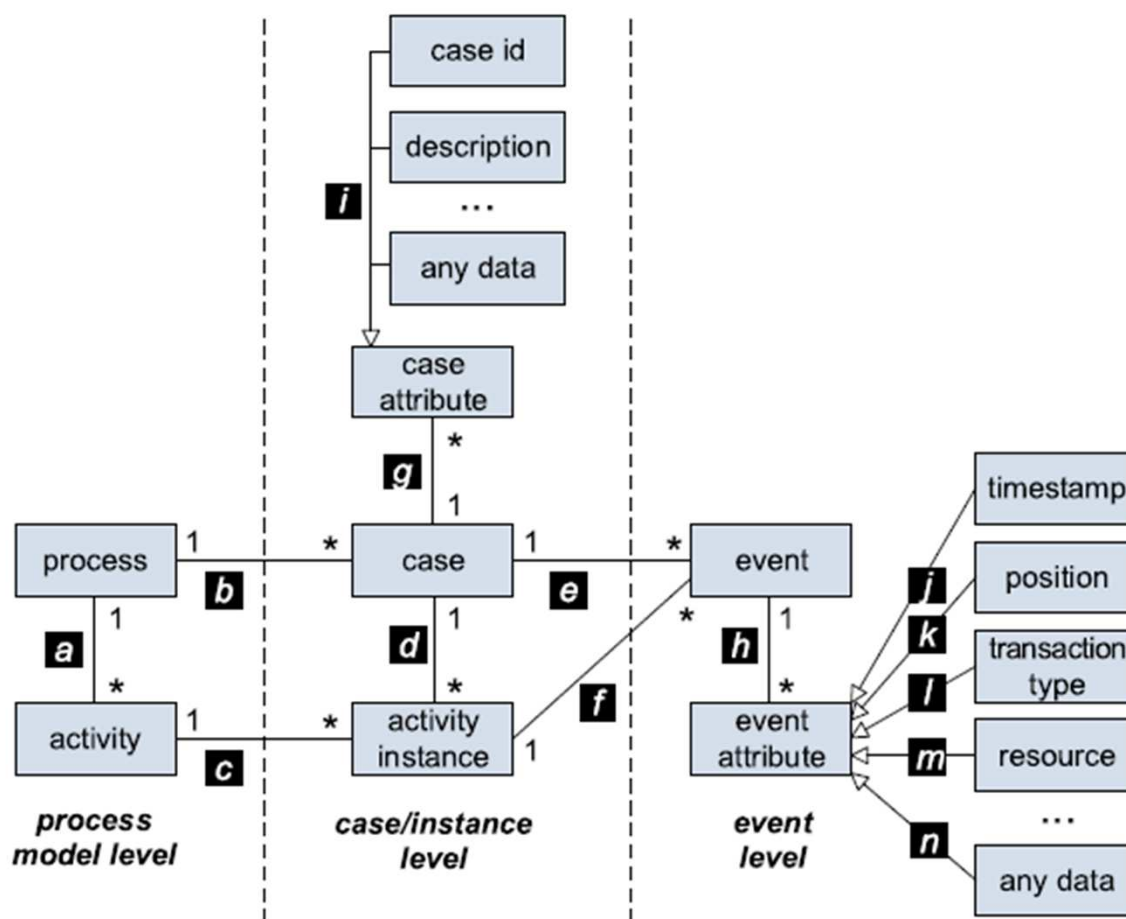
Fonte: Pinheiro et al., 2020

Process Mining - PM



Fonte: van der, Aalst (2012)

Process Mining - PM



Fonte: van der, Aalst (2016)

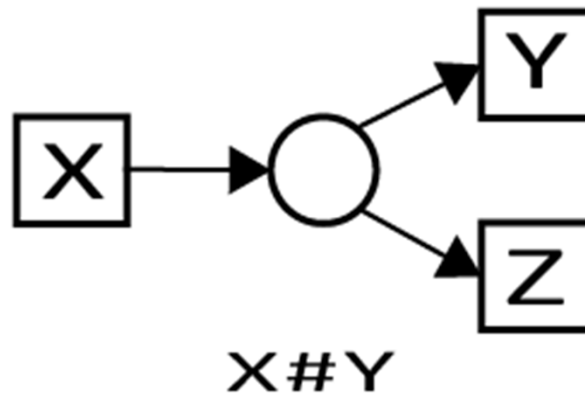
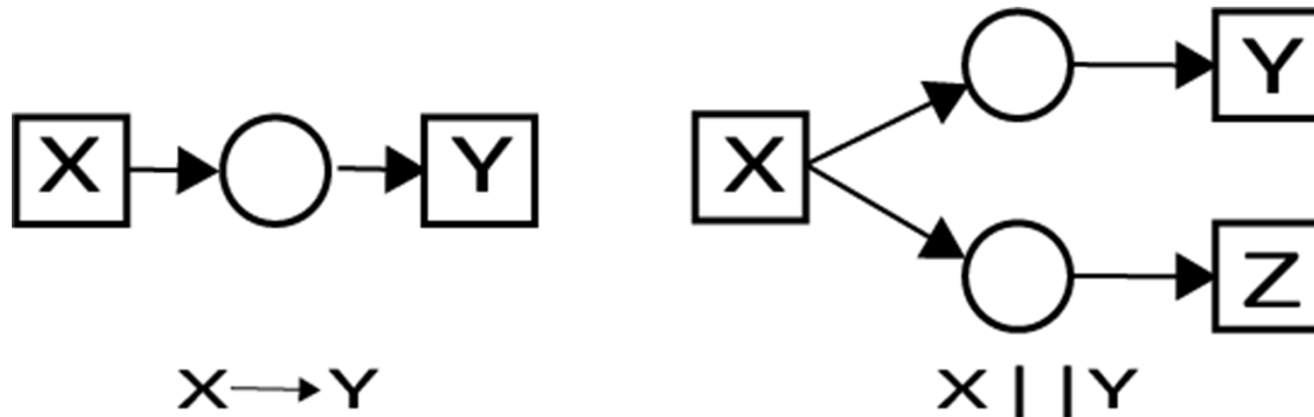
Process Discovery

Extração de um log de eventos e produção de um modelo de processo, elucidando o comportamento registrado no log (van der Aalst, 2016).



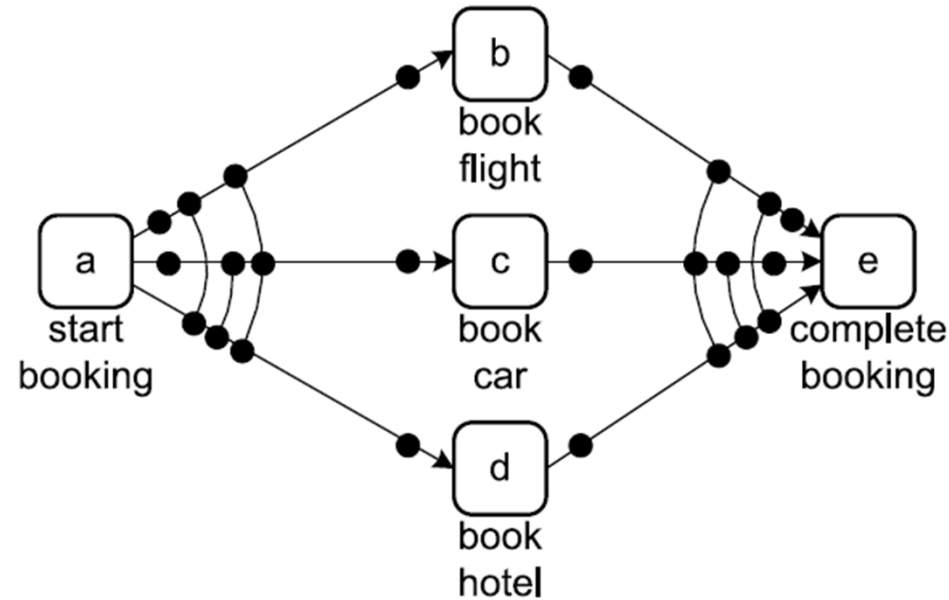
Fonte: Adaptado de Cook e Wolf ,1998.

α -Algorithm



Fonte: van der, Aalst (2011)

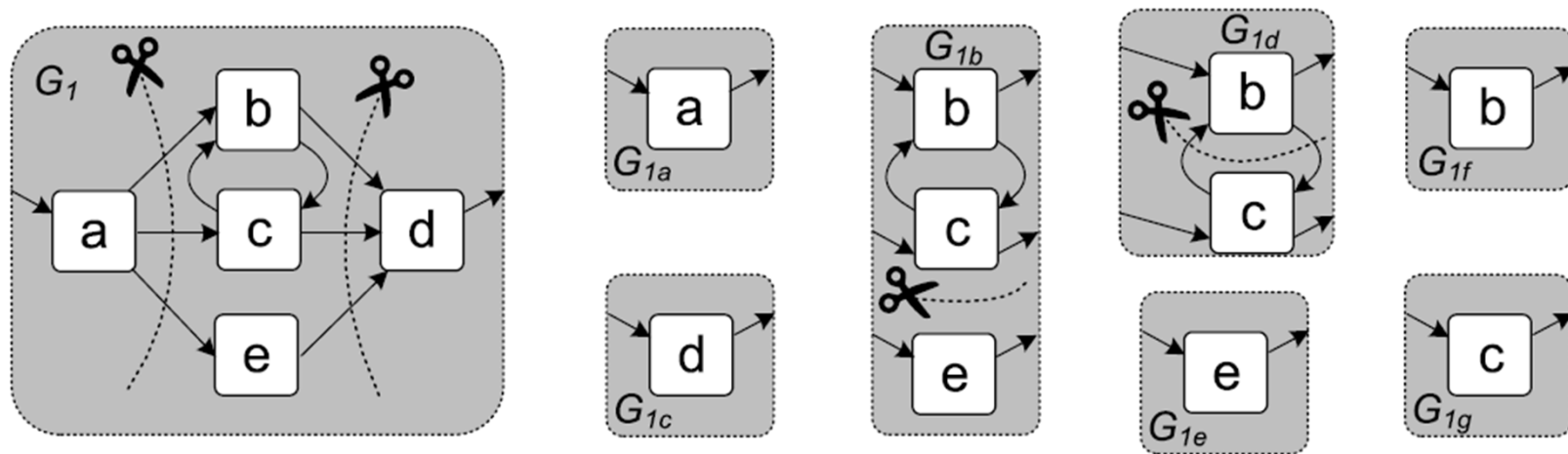
Heuristic Miner



$$|a \Rightarrow_L b| = \begin{cases} \frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{se } a \neq b \\ \frac{|a >_L a|}{|a >_L a|} & \text{se } a = b \end{cases}$$

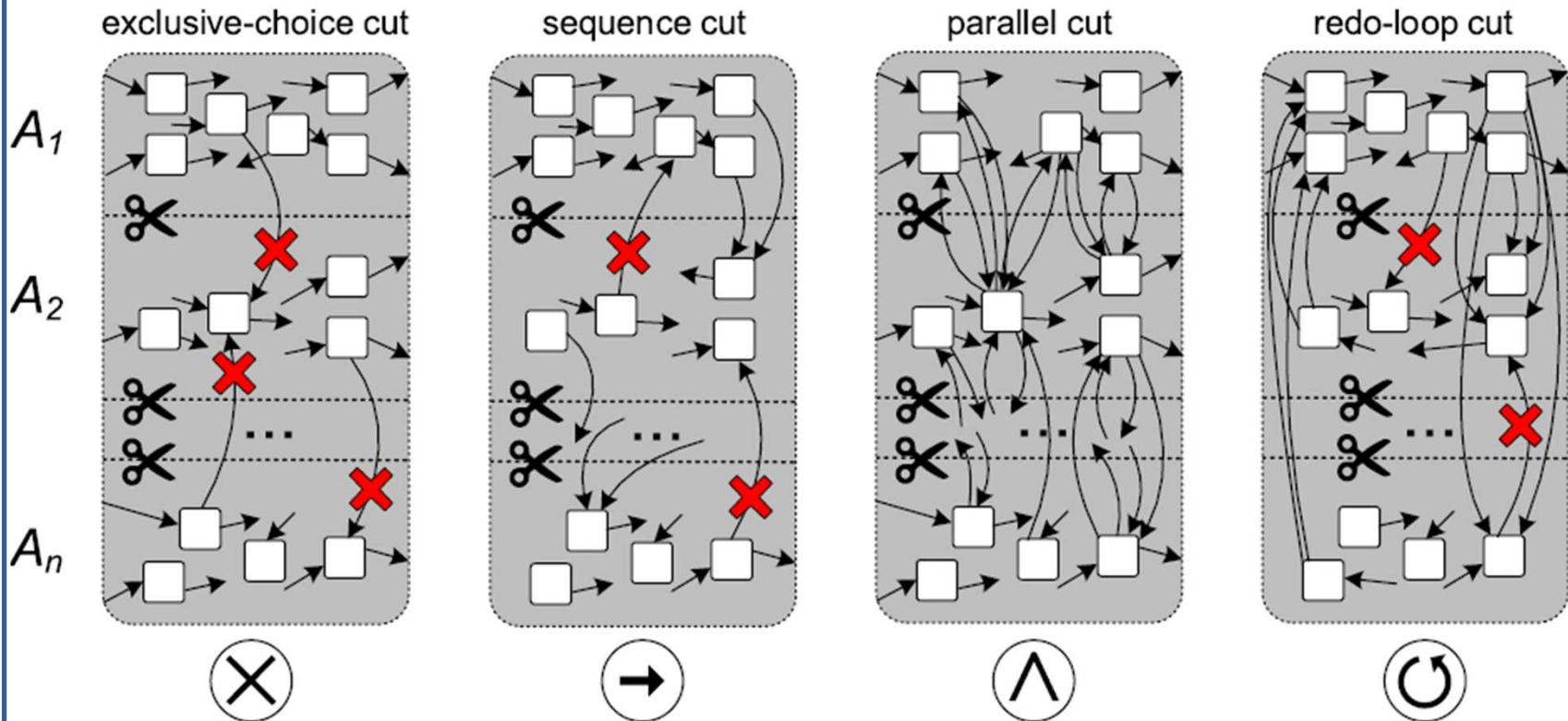
Fonte: van der, Aalst (2016)

Inductive Miner



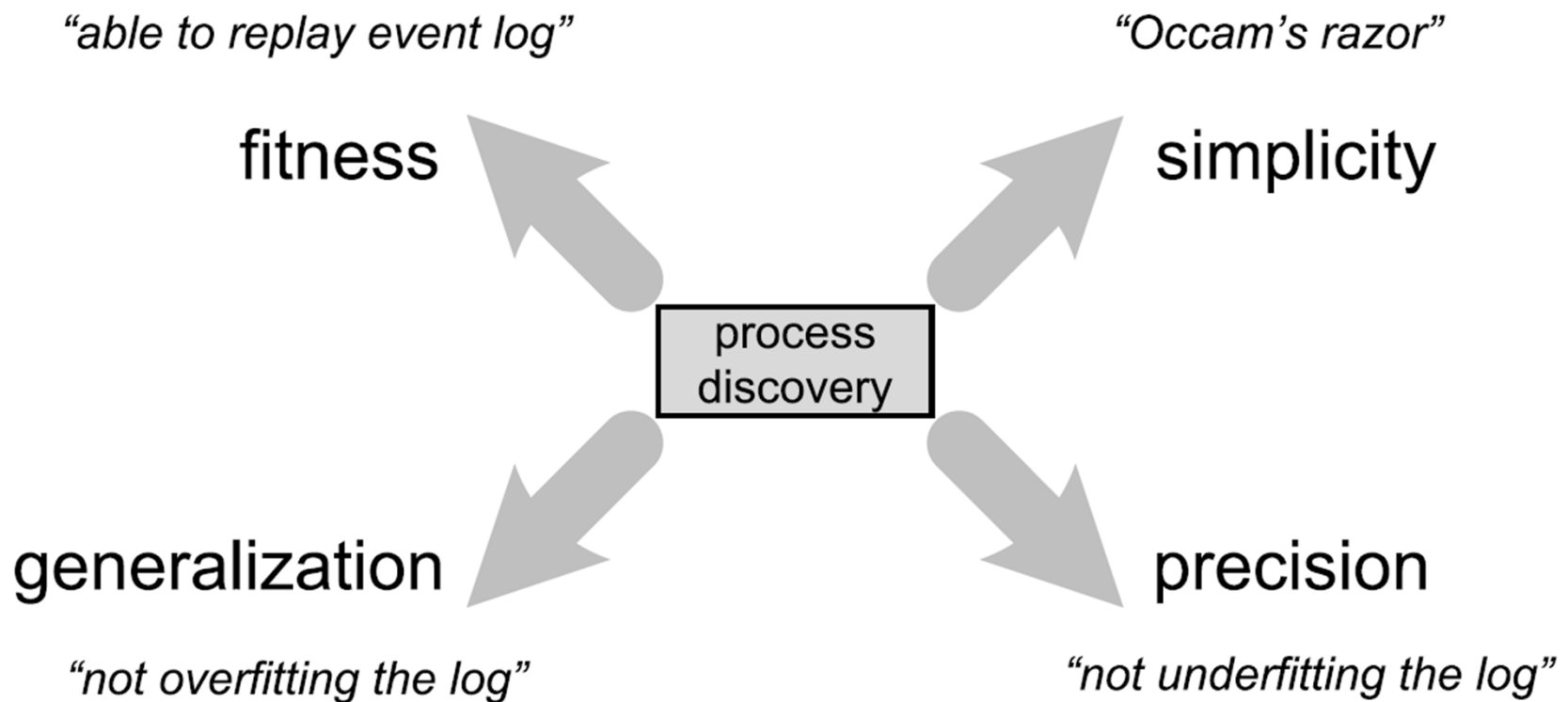
Fonte: van der, Aalst (2016)

Inductive Miner



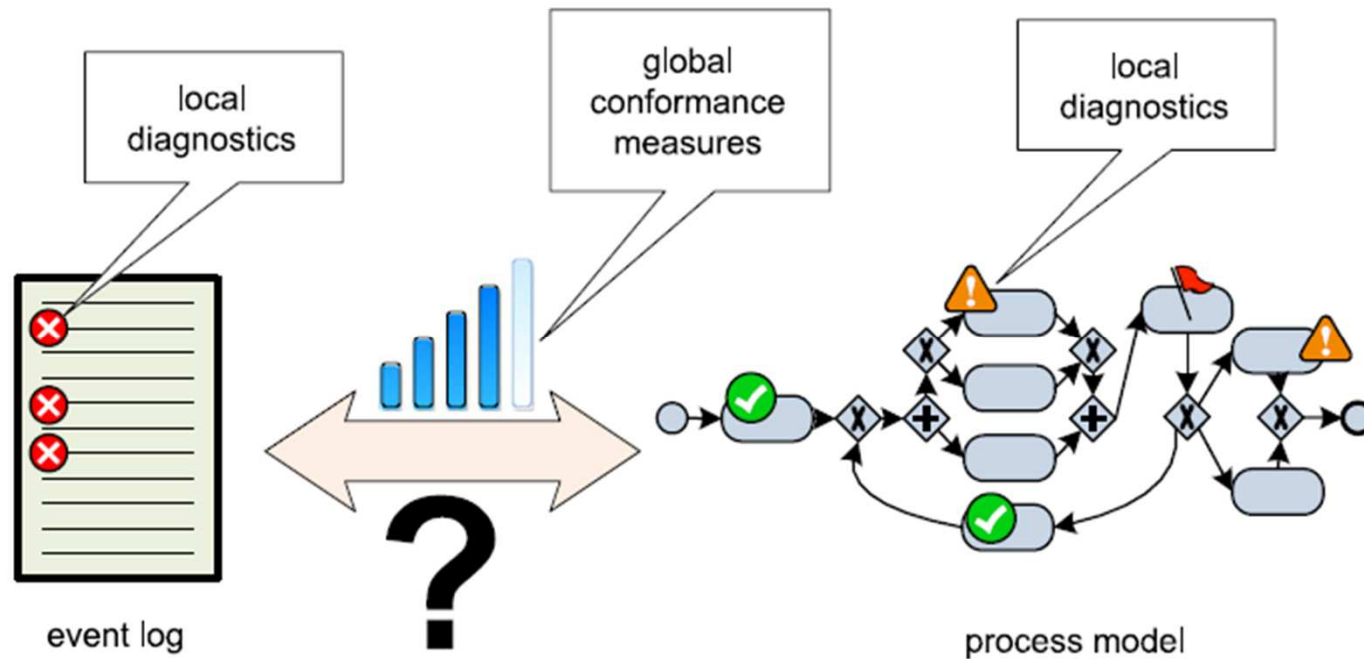
Fonte: van der, Aalst (2016)

As Quatro Dimensões de Qualidade



Conformance Checking

Busca verificar o quão bem um log de eventos e um modelo de rede do sistema se encaixam.



Fonte: van der, Aalst (2016)

Token Based Replay (TBR)

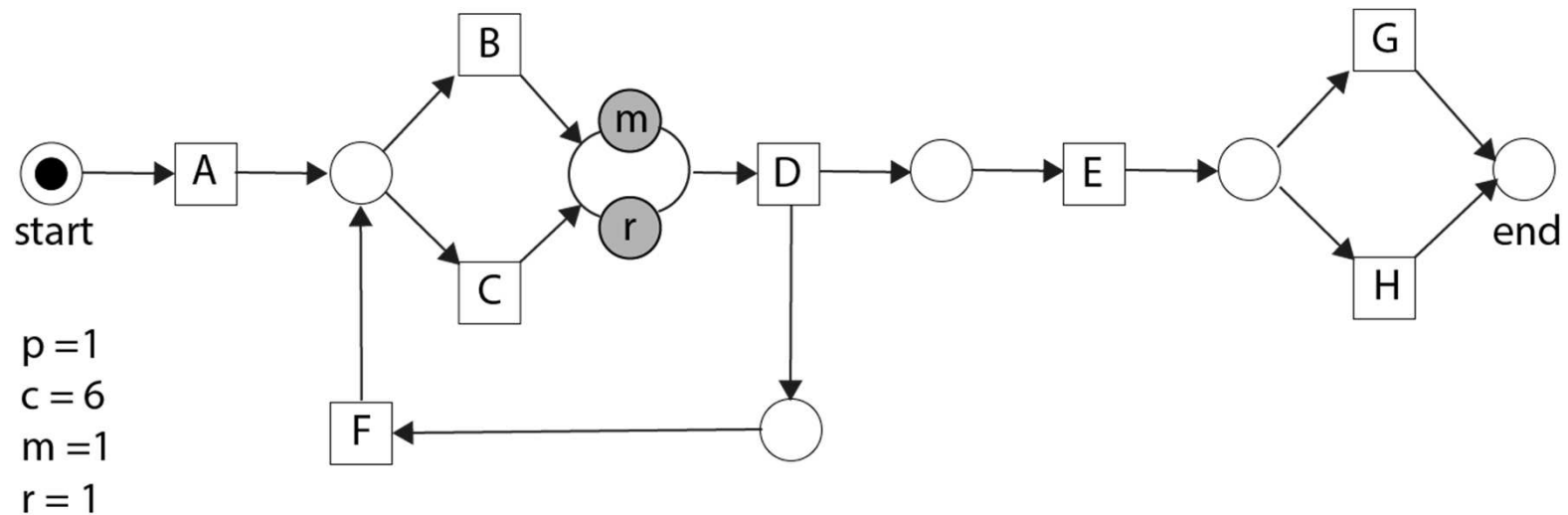
Permite de alguma forma descobrir a fração do cenário que está em conformidade com o modelo e a fração que não está (van der Aalst, 2014).

$$TBR = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in S} S(\sigma) \cdot m_{\sigma}}{\sum_{\sigma \in S} S(\sigma) \cdot c_{\sigma}} \right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in S} S(\sigma) \cdot r_{\sigma}}{\sum_{\sigma \in S} S(\sigma) \cdot p_{\sigma}} \right)$$

Fonte: Rozinat (2010)

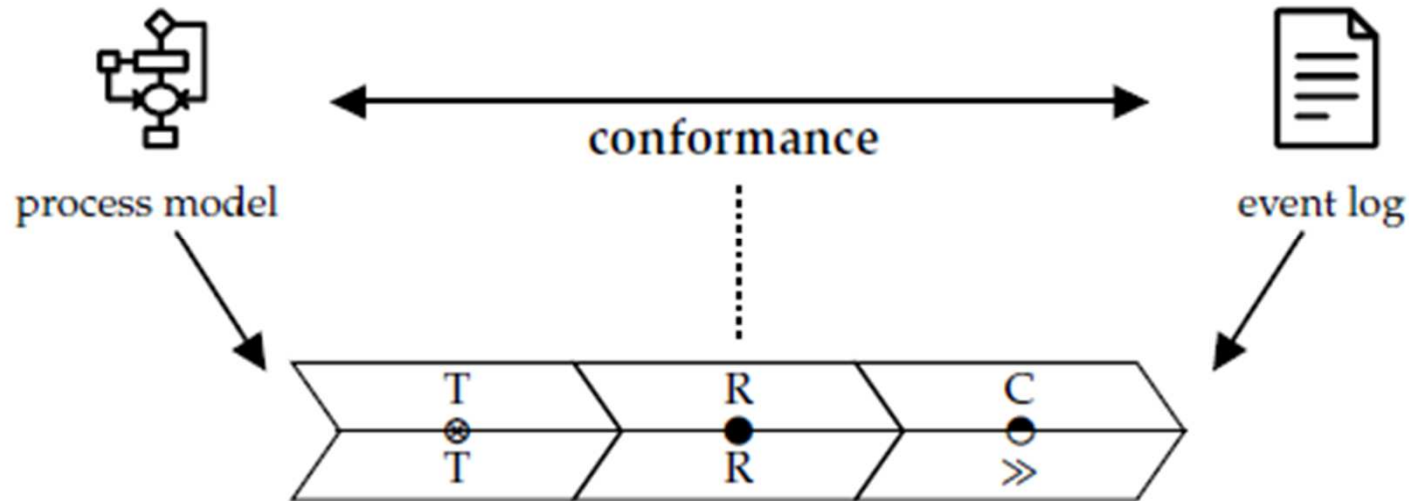
Token Based Replay (TBR)

<A, D, C, E, H>



$$\text{Fitness} = \frac{1}{2} \left(1 - \frac{1}{6}\right) + \frac{1}{2} \left(1 - \frac{1}{6}\right)$$

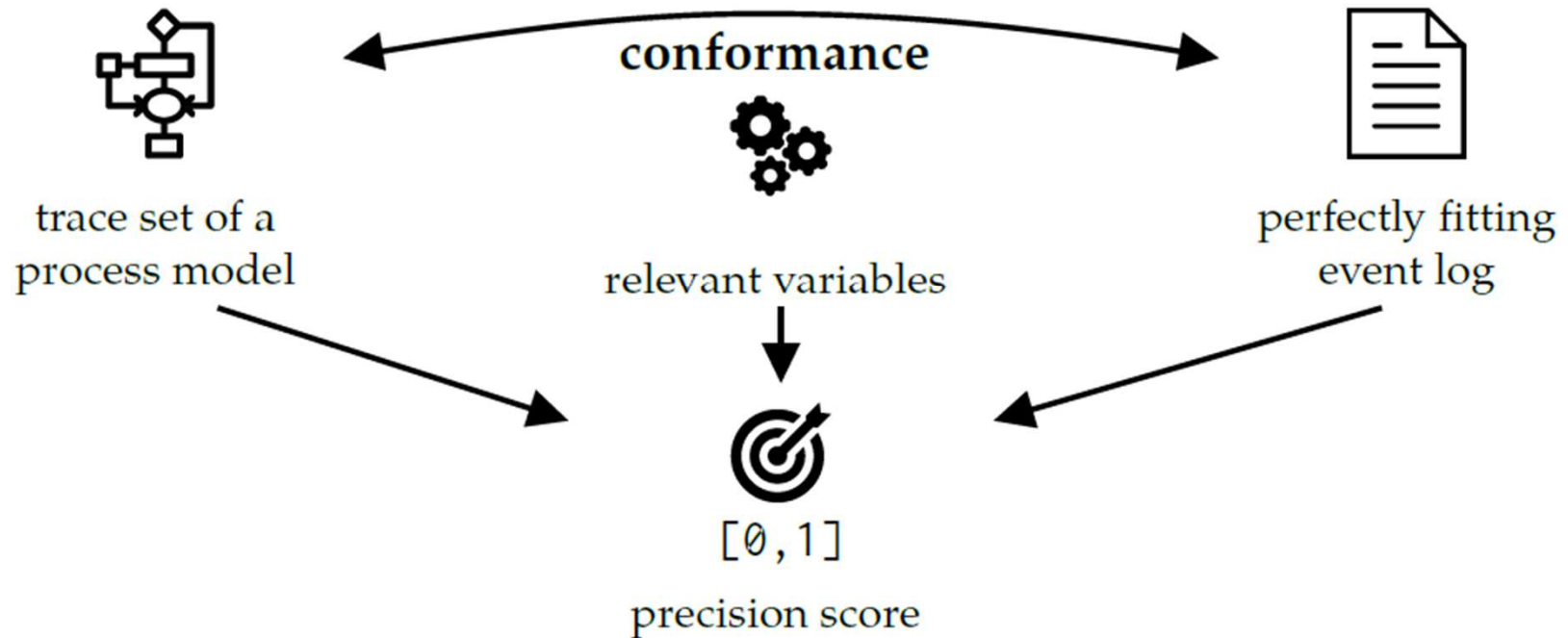
Alignment



$$Fitness - alignment(E, PN) = 1 - \frac{\sum_{\sigma \in E} E(\sigma) \cdot \lambda_{otm}(\sigma, PN)}{\sum_{\sigma \in E} E(\sigma) \cdot \lambda_{prior}(\sigma, PN)}$$

Fonte: Mannhardt (2018)

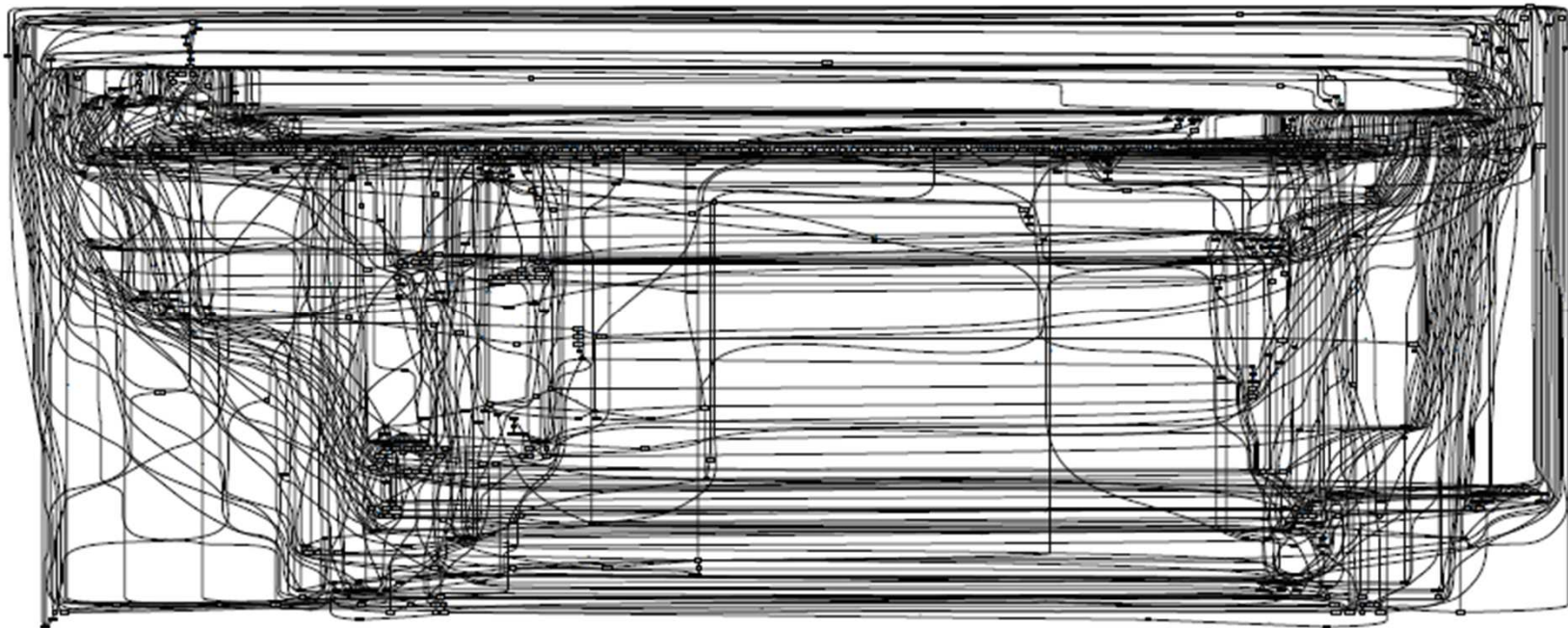
Precision



$$\text{Multi - perspective Precision}(LTS, L, V_{PR}) = \frac{\sum_{e \in E} |obs_{LTS,L}(e, V_{PR})|}{\sum_{e \in E} |pos_{LTS,L}(e, V_{PR})|}$$

Fonte: Mannhardt (2018)

Simplicity



$$Simplicity = 1 - \frac{n^{\circ} \text{ de atividades duplicadas} + n^{\circ} \text{ de atividades faltantes}}{n^{\circ} \text{ de nós em uma árvore} + n^{\circ} \text{ de classes em um log de eventos}}$$

Fonte: Buijs et al., 2012

Ant Colony Optimization – ACO

Pseudo-Algoritmo 1 - ACO

Input: $G(C, L), f, \Omega$

Output: $G(C, L^*)$

procedure ACOMetaheuristic

ScheduleActivities

ConstructAntsSolutions

UpdatePheromones

DaemonActions % optional

end-ScheduleActivities

end-procedure

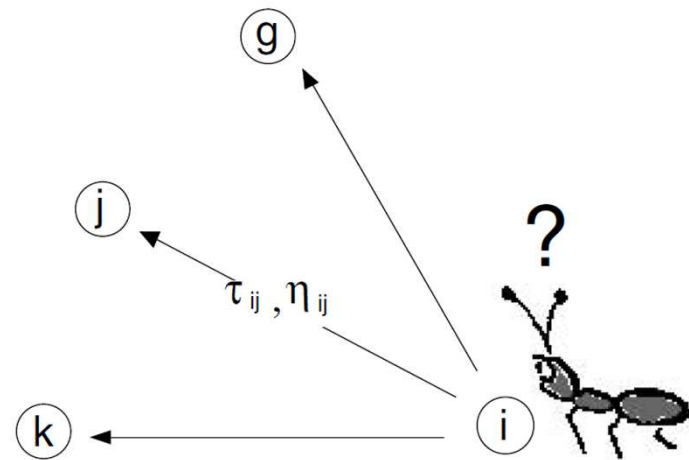
Fonte: Dorigo e Stützle (2004)



Ant Colony Optimization – ACO

$$P_T(c_{h+1} = j | x_h) = \begin{cases} \frac{F_{ij}(\tau_{ij})}{\sum_{(i,l) \in N_i^k} F_{il}(\tau_{il})}, & \text{se } (i,j) \in AS \\ 0, & \text{do contrário} \end{cases}$$

$$F_{ij}(z) = z^\alpha \eta_{ij}^\beta$$



Fonte: Dorigo e Stützle (2004)

Inteligência de Enxames em *Process Mining*

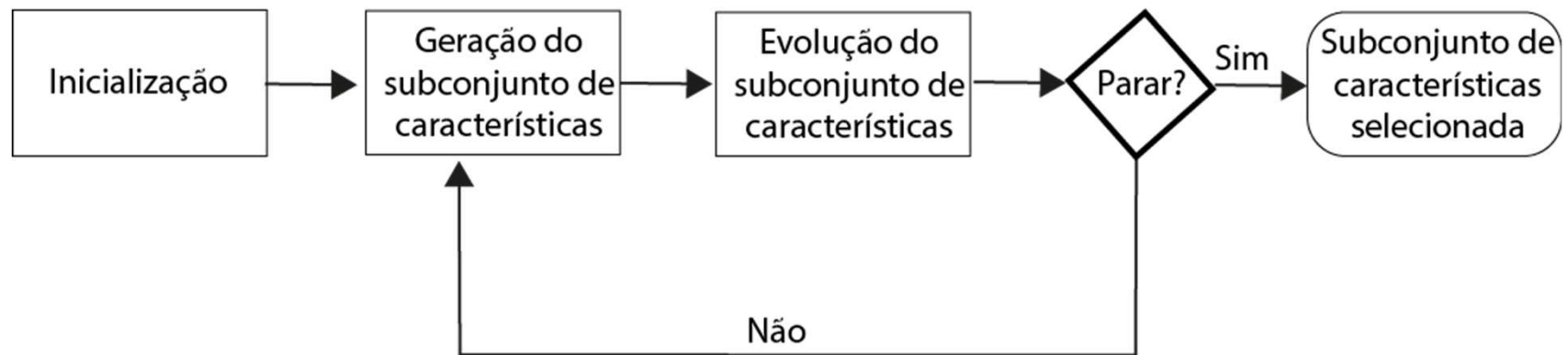
Hybrid Particle Swarm Optimization method for process mining (Chifu et al., 2012)

- *Process Discovery*;
- Otimização da *Matrix Casual*.

Process Discovery Using Ant Colony Optimization
(Chinces e Salomie, 2013)

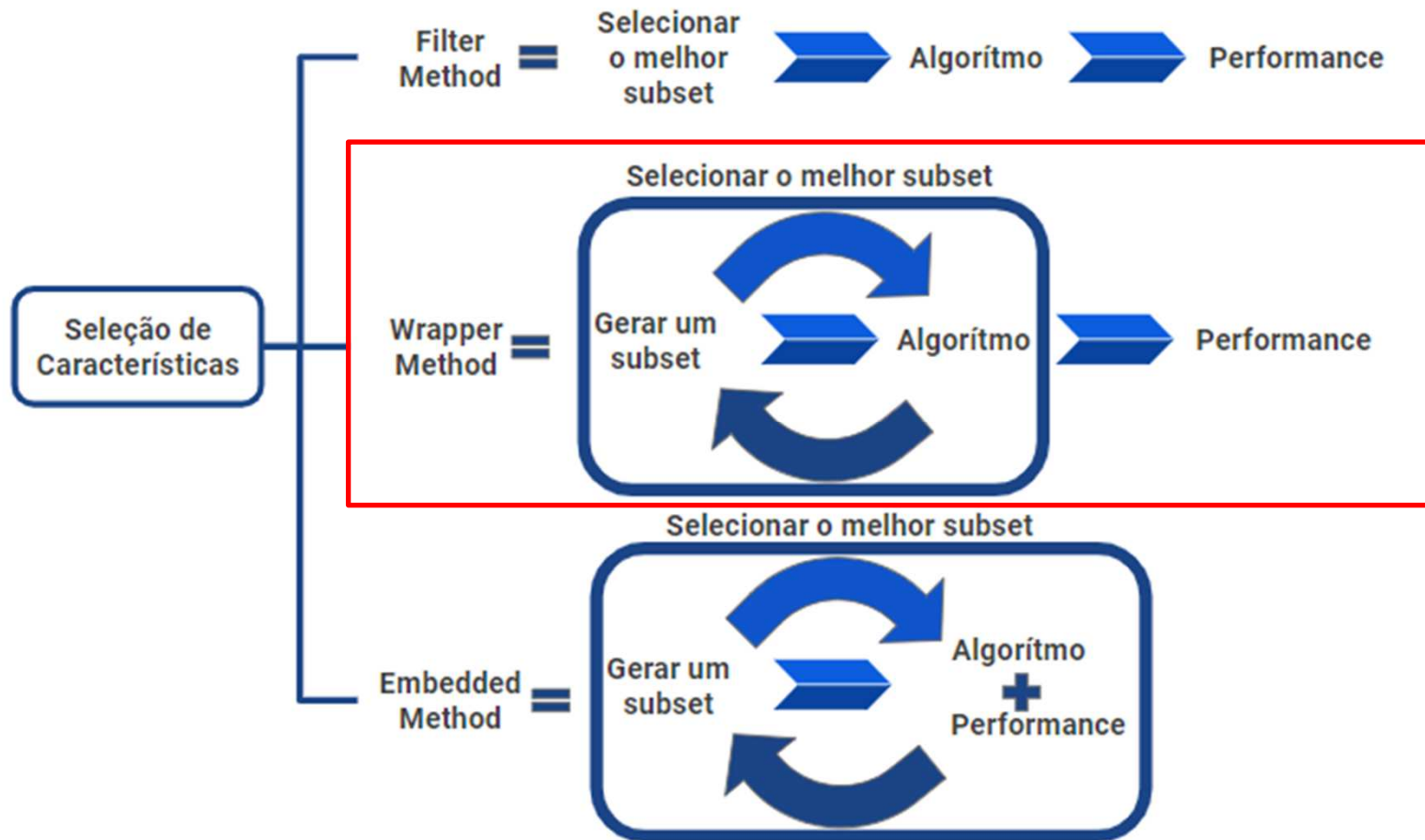
- *Process Discovery*;
- Uso de C-Nets;
- Identificação das transições (2-gram) e não consideram os efeitos do ruído e da incompletude.

Seleção de Características



Fonte: adaptado de Dash e Liu (1997)

Seleção de Características



Fonte: Adaptado de Julie Hamon,, 2013.

Seleção de Características em *Process Mining*

- Divisão iterativa dos traços em *clusters* até que o log estivesse particionado em *clusters* ótimos (de Medeiros et al.,2008);
- Agrupamento hierárquico em que cada **traço como um ponto de um espaço de características** devidamente identificado (Grecco et al.,2006);
- **Seleção de característica estrutural** extraída de propriedades de sequências de atividades derivadas de log de eventos (2-gram) (Hinka et al, 2018);
- **Análise** de vários métodos de seleção de subconjuntos de traços, analisando seus respectivos desempenhos (Fani Sani et al.,2019)
- Abordagem de agrupamento focada nas medidas de *conformance checking* de cada traço, com objetivo de gerar modelos de processos com **ganhos na dimensão *fitness*** (Cirne et al.,2020)

CONTRIBUIÇÕES

Recife-PE, Brasil, 09/09/2020

Seleção de características de modelos de processos usando técnicas de Inteligência de Enxames

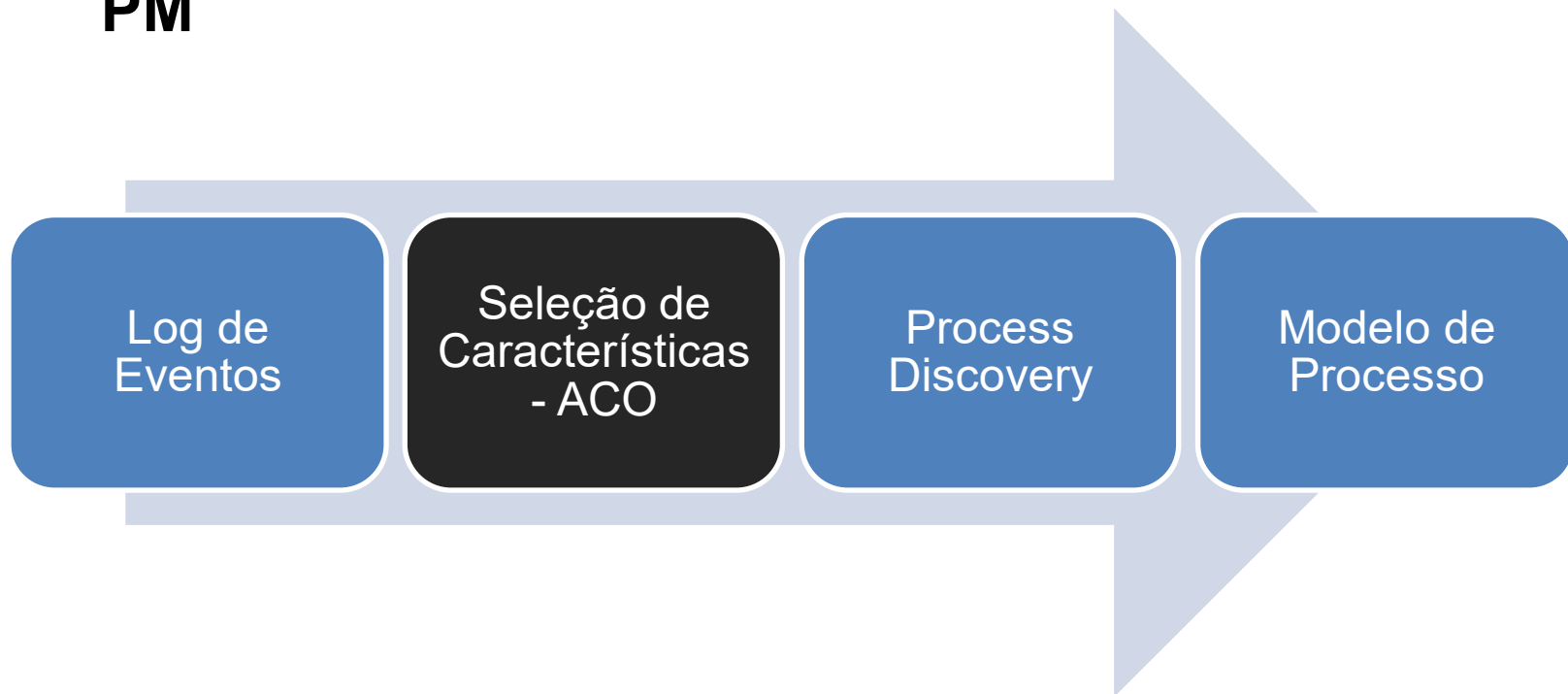
Oportunidades

Utilização de seleção de características com Inteligência de Enxames em processos de negócio

- Não há pesquisa anterior;
- Problema NP-completo ;
- Variáveis estão inter-relacionadas;
- Sequenciamento e dependências (representação em grafos);
- Bons resultados com ACO na seleção de características;
- Experiências anteriores com adaptações no uso de IE em *Process Mining*.

Oportunidades

- **Utilização de seleção de características com IE em PM**



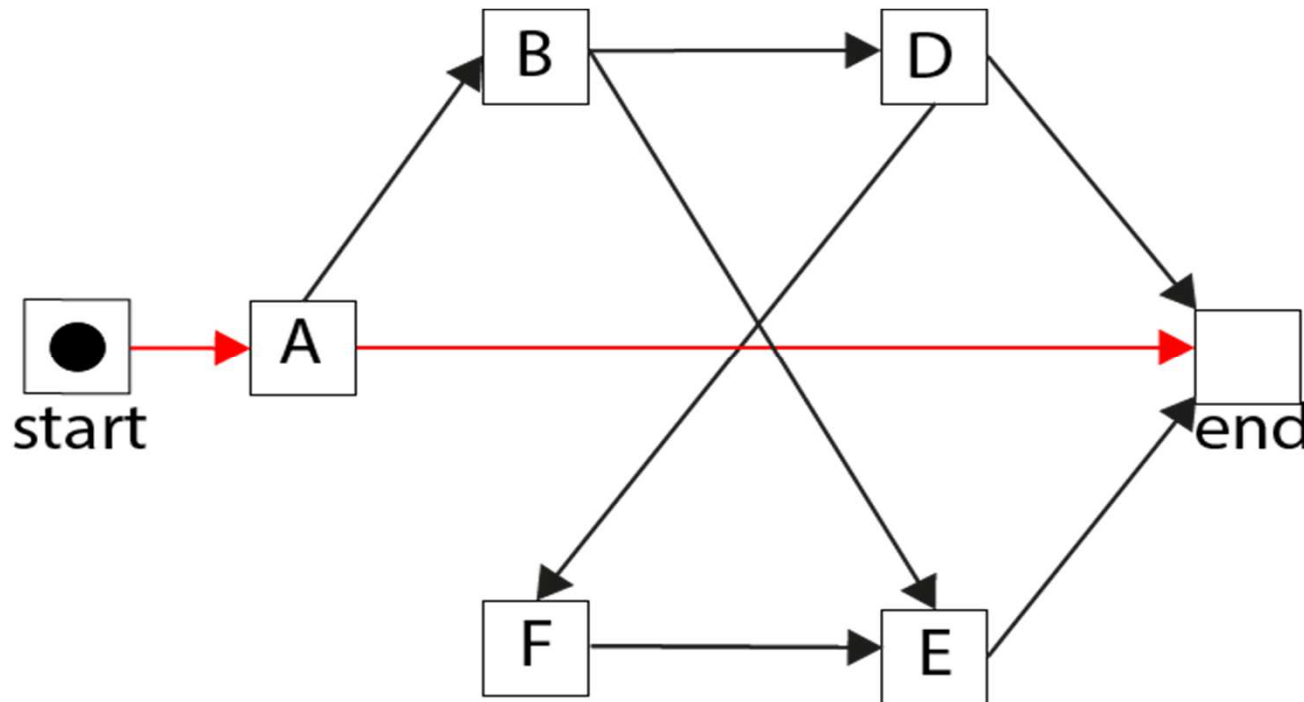
Oportunidades

Seleção de Característica - ACO com Feromônio Repelente

- Aplicações focadas na identificação da rota mais curta
- Adaptações a problemas de maximização
- Quatro dimensões de avaliação da qualidade de um modelo de descoberto
- Estudos considerando a existência de fenômenos naturais com ação repelente
- Existência do feromônio de “Entrada Proibida” nas formigas-faraó.

Oportunidades

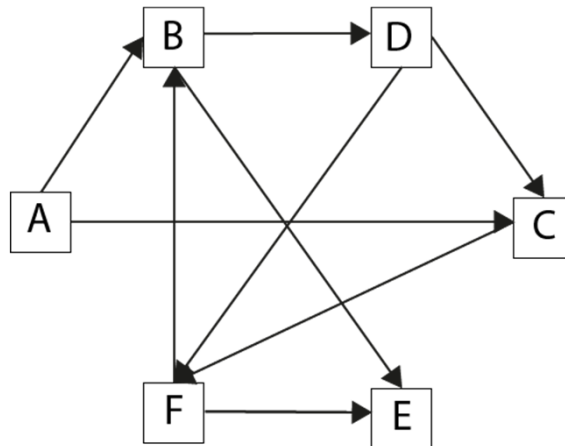
Seleção de Característica - ACO com Feromônio Repelente



Proposta de Método

Modelagem Espaço de Busca

Função 1 (Geração *C-Net*) Seja L um log de eventos. Sendo σ um caso de L , i.e., $\sigma \in L$. Sendo a e b eventos de σ , i.e., a e $b \in \sigma$. $\forall \sigma \in L \quad \forall a, b \in \sigma$ se $a \Rightarrow_L b \Rightarrow (a, b) + AS$, i.e., a relação entre a e b gera uma transição, sendo b o nó.



$\langle A, B, D \rangle^{21}$

$\langle A, C, F, E \rangle^3$

$\langle A, B, E \rangle^{12}$

$\langle A, B, D, C, F, E \rangle^{22}$

$\langle A, B, D, F, E \rangle^{14}$

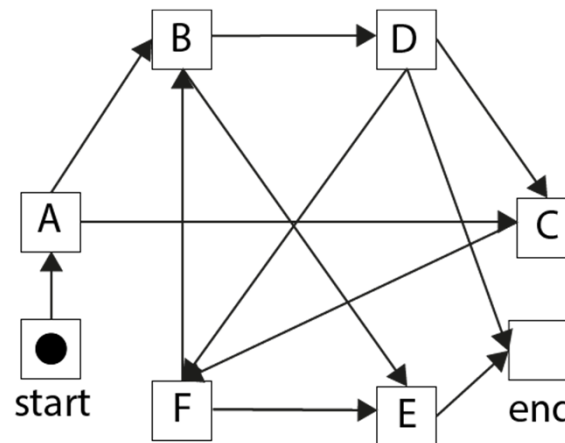
$\langle A, B, D, F, B, E \rangle^{11}$

Proposta de Método

Modelagem Espaço de Busca

Função 2 (Adição de início/fim artificiais) Seja W um log de eventos de T . Então W^+ é um (artificial) início e fim adicionais sobre T^+ com:

1. $T^+ = T \cup \{início, fim\}$,
2. $W^+ = \{início, \sigma, fim \mid \sigma \in W\}$

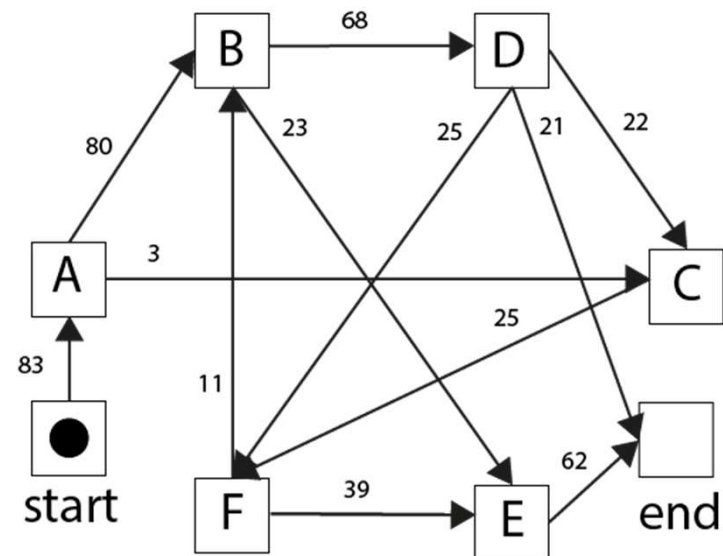


Proposta de Método

Modelagem Espaço de Busca

Função 3 (Cálculo da Medida de Dependência) Seja L um log de evento de A e $a, b \in A$. Assim como $|a >_L b|$ é o número de vezes que a é diretamente seguido por b em L . Então, $|a \Rightarrow_L b|$ o valor da medida de dependência entre a e b , sendo $a \neq b$ é dado por:

$$|a \Rightarrow_L b| = |a >_L b|$$



Proposta de Método

Seleção de Características

Função 4 (Restrições Ω) Seja S^* um conjunto não-vazio de seleções de características ótimas, com $S^* \subseteq S$. As restrições Ω condicionarão S^* a um conjunto de seleções de características viáveis. A busca de uma dada formiga $m \in M$, sendo M uma colônia de formigas, é restrita às seguintes condições $\Omega.n$:

- $\Omega.1 - \forall_{m \in M}$, sendo σ o traço de seleção de características, $hd_1(\sigma) = início$;
- $\Omega.2 - \forall_{m_1, m_2 \in M}$, se $m_1 \neq m_2 \Rightarrow \sigma_1 \neq \sigma_2$;
- $\Omega.3 - \forall_{m \in M}$, sendo σ o traço de seleção de características, $tlk(\sigma) = fim$

Proposta de Método

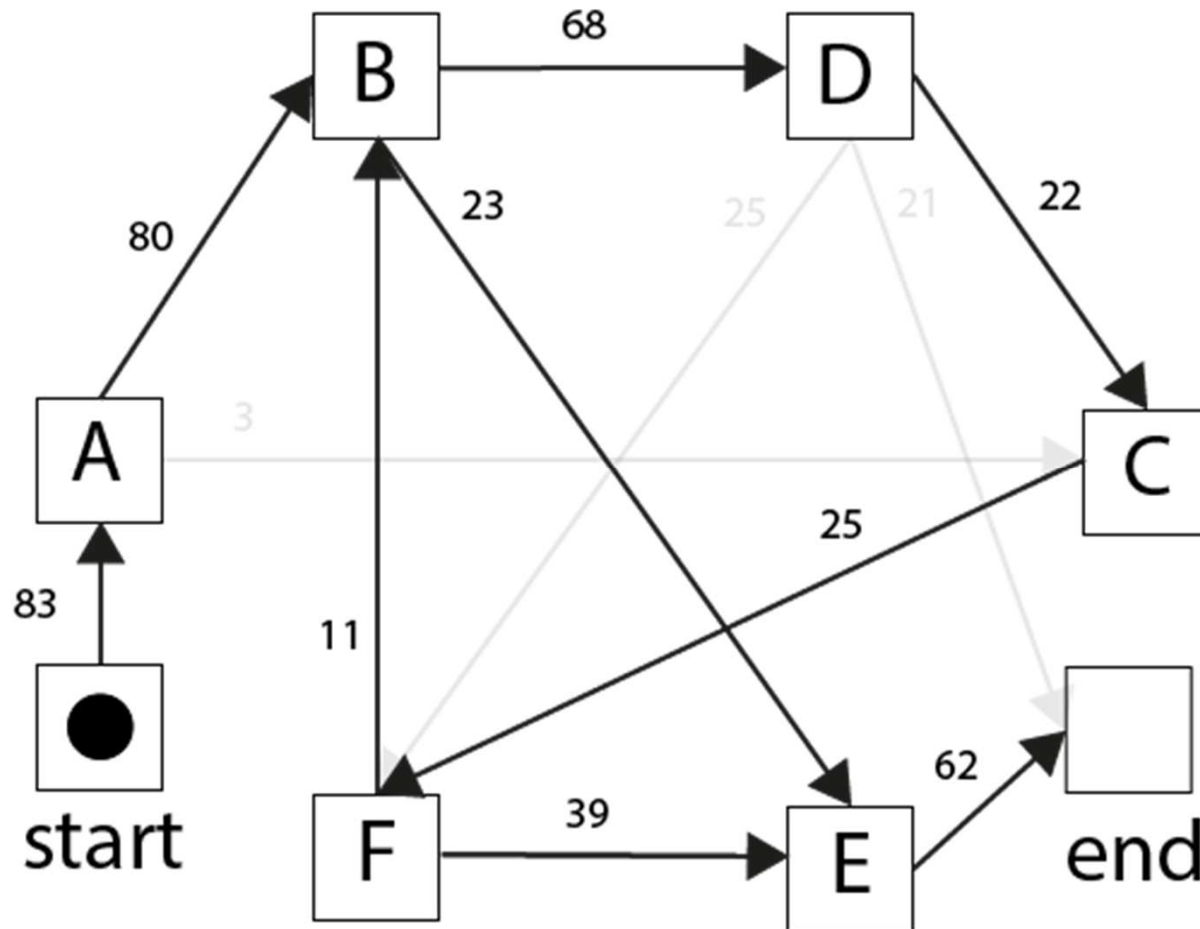
Seleção de Características

Função 5 (Probabilidade de seleção do nó P) A função $P(c_{h+1} = j | \sigma_h)$ é a função de probabilidade de uma formiga em uma dada atividade i e na etapa h , escolher uma atividade j , da seguinte forma:

$$P(c_{h+1} = j | \sigma_h) = \begin{cases} \frac{(1 + z_{ij})^\alpha |i \Rightarrow_L j|^\beta}{\sum_{k \in A} (1 + z_{ik})^\alpha |i \Rightarrow_L k|^\beta}, & \text{se } (i, j) \in AS \\ 0, & \text{do contrário} \end{cases}$$

Proposta de Método

Seleção de Características



Proposta de Método

Seleção de Características

Função 6 (*F-Fitness*) Seja $s \in S^*$ uma seleção de características ótima de uma formiga $m \in M$, sendo σ o traço gerado. $f(s)$ é a função fitness dada por:

$$f(s) = \frac{\sum_{i,j \in \sigma} |i \Rightarrow_L j|}{|\sigma|}$$

onde, i e j são atividades sequenciadas de σ , i.e., sendo $\sigma = \langle t_1, t_2, t_3, \dots, t_k \rangle$ e $i \in \{1, \dots, k-1\} \Rightarrow t_x = i$ e $t_{x+1} = j$.

Proposta de Método

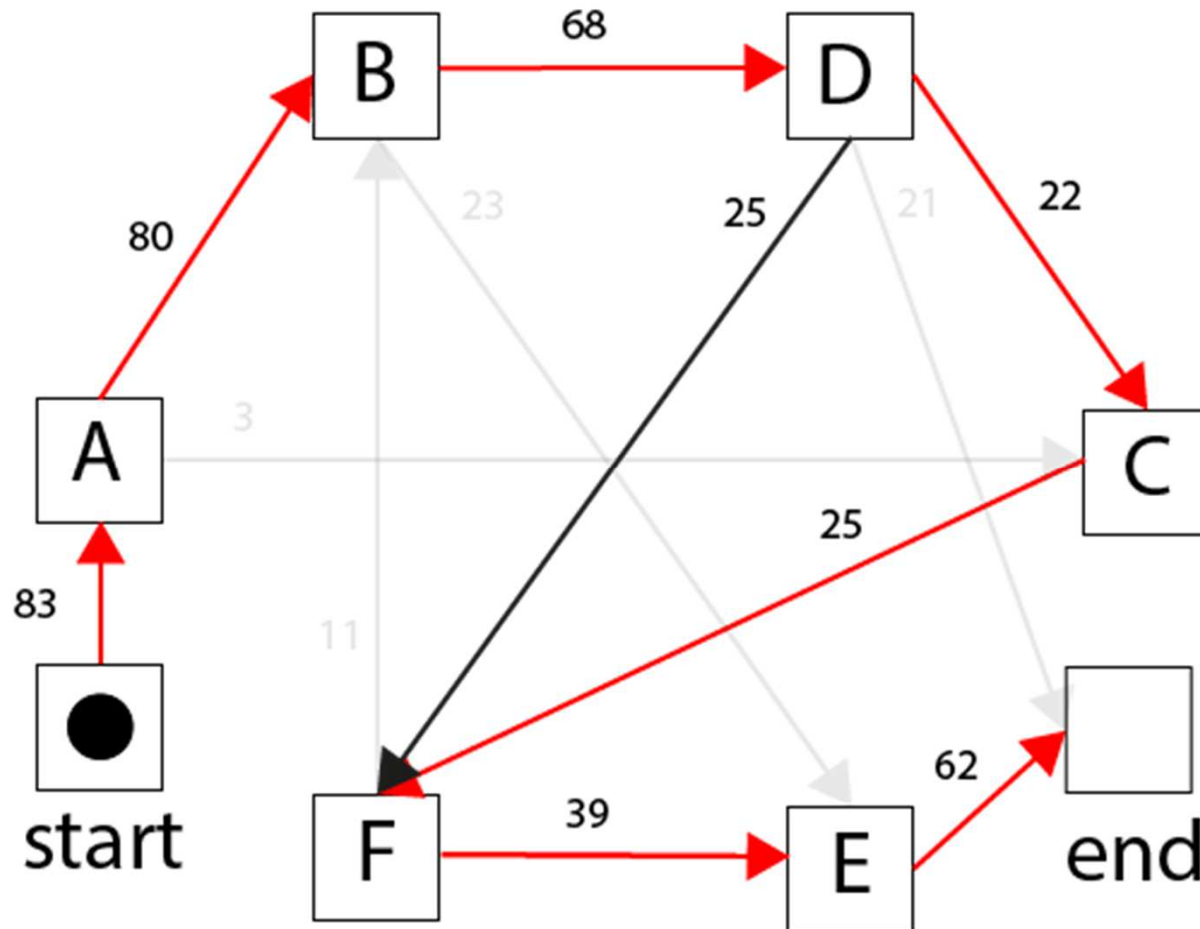
Seleção de Características

Função 7 (Atualização Feromônio) Seja $U(z_{ij}, \rho)$ a função de atualização do feromônio na transição (i,j) . Seja $f(s^{bs}) < f(s^y)$ a melhor solução de iteração (na iteração atual y).

$$U(z_{ij}, \rho) = \begin{cases} z_{ij} = z_{ij} \cdot (1 - \rho) + \frac{q}{f(s^{bs})}, & \text{se } (i,j) \in s^{bs}; \\ z_{ij} = z_{ij} \cdot (1 - \rho), & \text{se } (i,j) \notin s^{bs}; \end{cases}$$

Proposta de Método

Seleção de Características



Proposta de Método

Algoritmo *Feature Selection - Ant Process Mining (FS-APM)*

1. configurar parâmetros ($\mu, \omega, \tau_0, \rho, \alpha, \beta, q$)

μ – número de formigas;

ω – número de iterações ;

τ_0 – feromônio inicial;

ρ – taxa evaporação;

α – parâmetro de efeito do feromônio;

β – parâmetro de efeito visibilidade;

q – parâmetro de depósito de feromônio.

2. gerar Espaço de Busca

2.1. gerar *C-Net*; (**Função 1**)

2.2. adicionar início/fim artificiais; (**Função 2**)

2.3. calcular medidas de dependência. (**Função 3**)

Proposta de Método

3. selecionar características

3.1. for i in ω :

3.1.1. for x in μ :

$h \leftarrow 1$

$x_h \leftarrow \langle \text{início} \rangle$ (**Função 4 - $\Omega.1$**)

3.1.1.1. while ($c_h \neq fim$): (**Função 4 - $\Omega.3$**)

$j \leftarrow \text{SelecionarPróximoNó}(x_h, P(c_{h+1} = j | x_h))$ (**Função 5**)

$x_h \leftarrow x_h \otimes j$

end-while

Proposta de Método

3.1.1.1. **while** ($c_h \neq fim$): (**Função 4 - $\Omega.3$**)

$j \leftarrow \text{SelecionarPróximoNó}(x_h, P(c_{h+1} = j | x_h))$ (**Função 5**)

$x_h \leftarrow x_h \otimes j$

end-while

3.1.1.2. **if** $x_h \in S$ (**Função 4 - $\Omega.3$**):

then 3.1.1.1

else return x_h

end-if

3.1.1.3. **end-for**

3.1.2. **calcular** *F-Fitness* (**Função 5**)

3.1.3. **atualizar feromônio** *U* (**Função 6**)

3.1.4. **end-for**

Proposta de Método

- ***Fitness*** \Rightarrow Problema Estático de Seleção de Característica – ACO;
- ***Precision*** \Rightarrow Geração da *C-Net* e Feromônio Repelente ($\alpha < 0$ em *P*)
- ***Generalization*** \Rightarrow O componente $\sum_{i,j \in \sigma} |i \Rightarrow_L j|$ da função *F-Fitness* e a restrição $\Omega.2$.
- ***Simplicity*** \Rightarrow O componente $|\sigma|$ da função *F-Fitness* e a restrição $\Omega.2$;

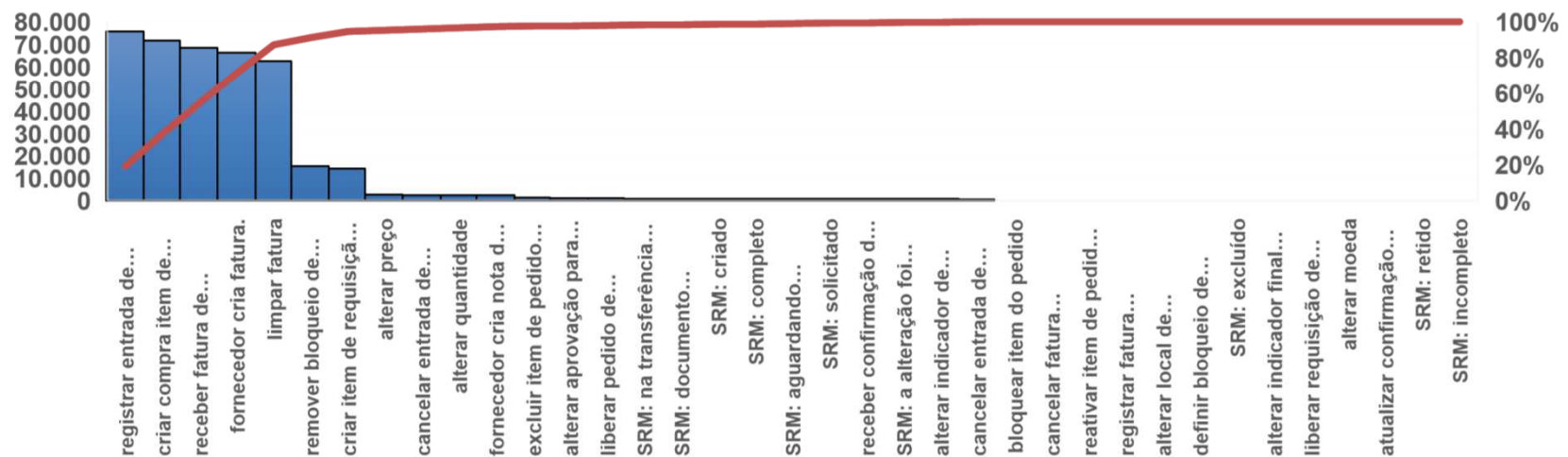
RESULTADOS

Recife-PE, Brasil, 09/09/2020

Seleção de características de modelos de processos usando técnicas de Inteligência de Enxames

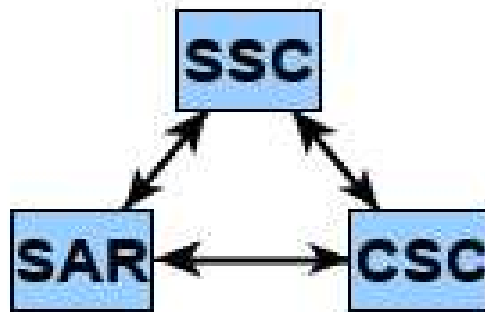
Análise Descritiva

Histograma de Atividades por Quantidade de Casos

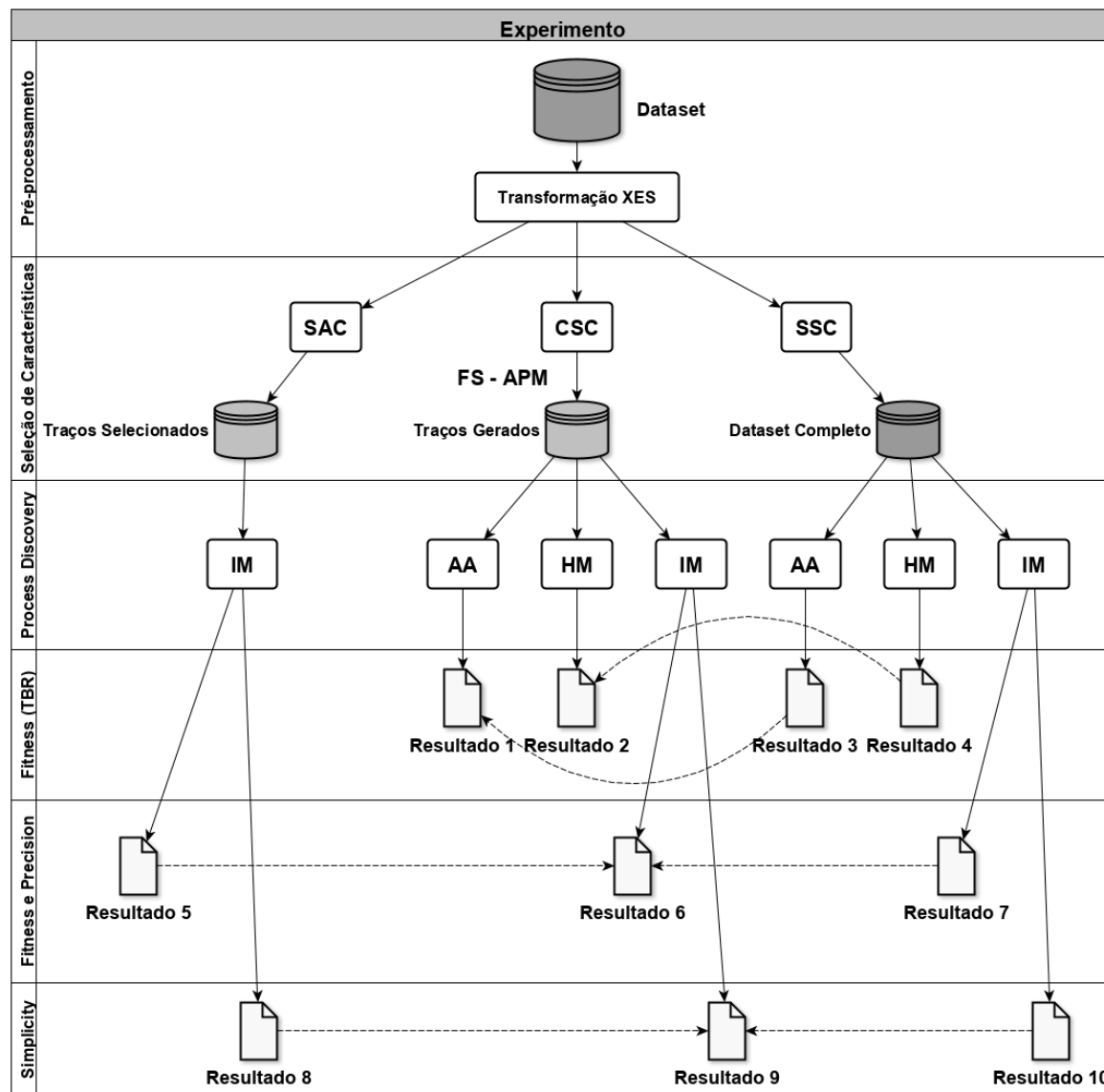


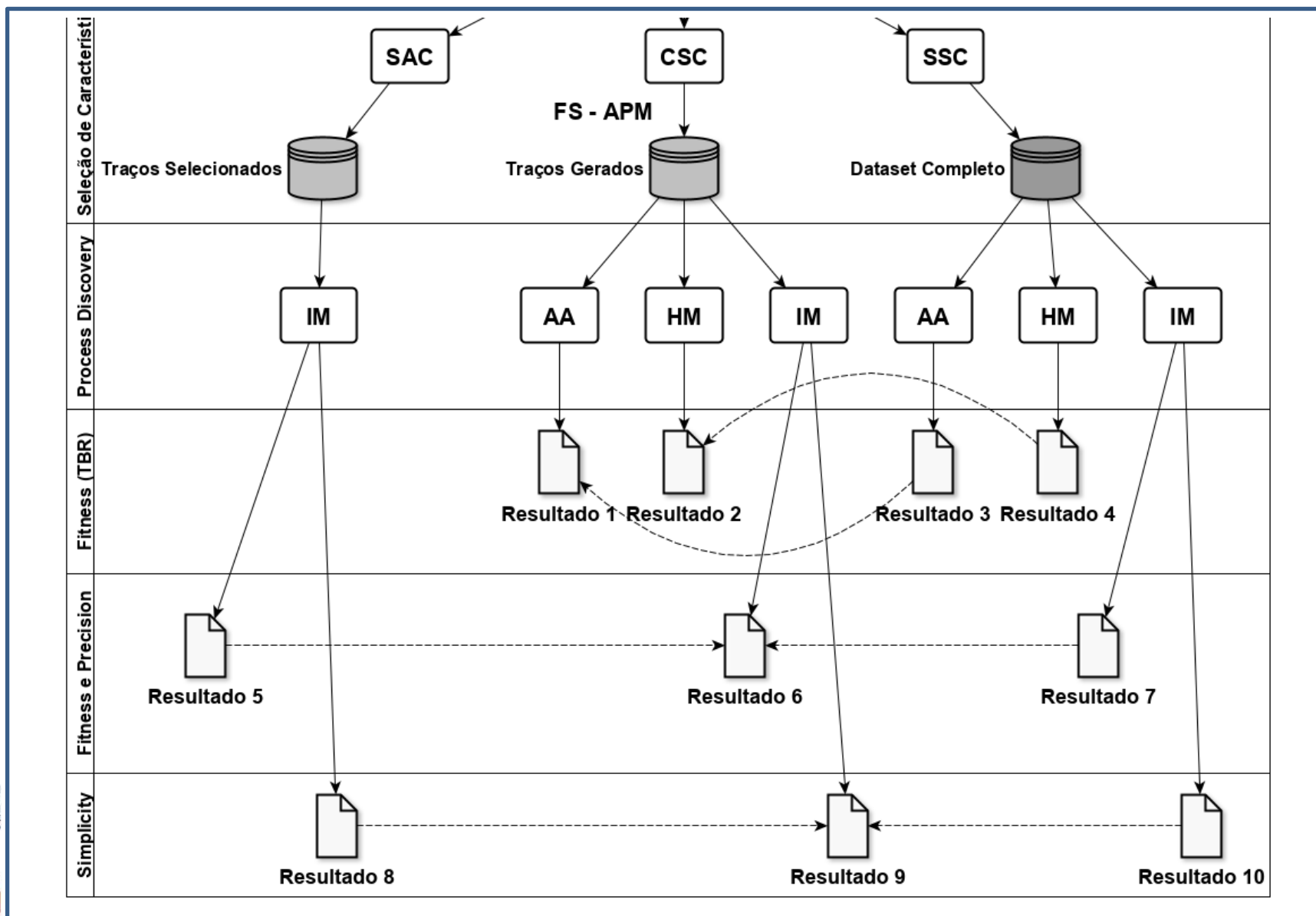
- 71.681 casos
- 395.017 eventos
- 558 vendedores
- 37 tipos de atividades

Experimento



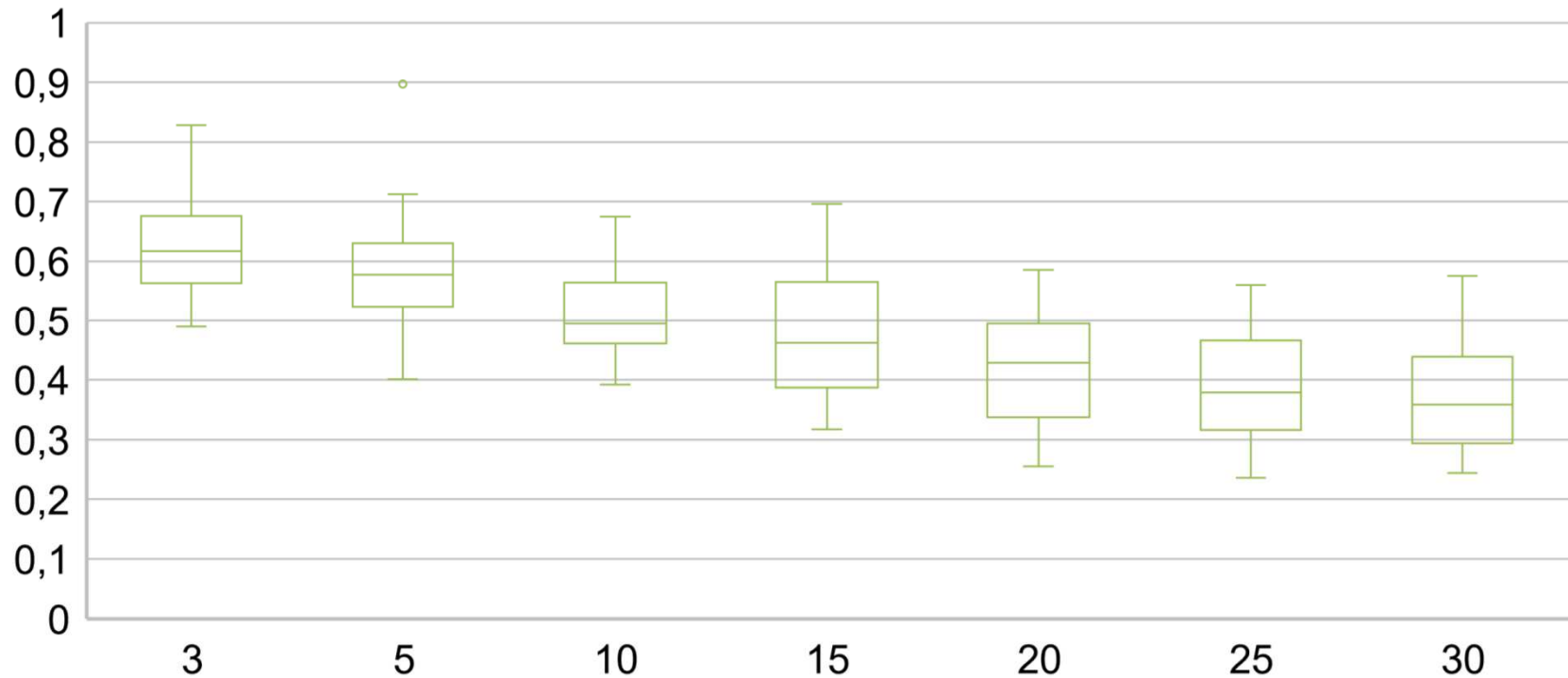
$$Qualidade = \frac{3}{\left(\frac{1}{Fitness} + \frac{1}{Precision} + \frac{1}{Simplicity} \right)}$$



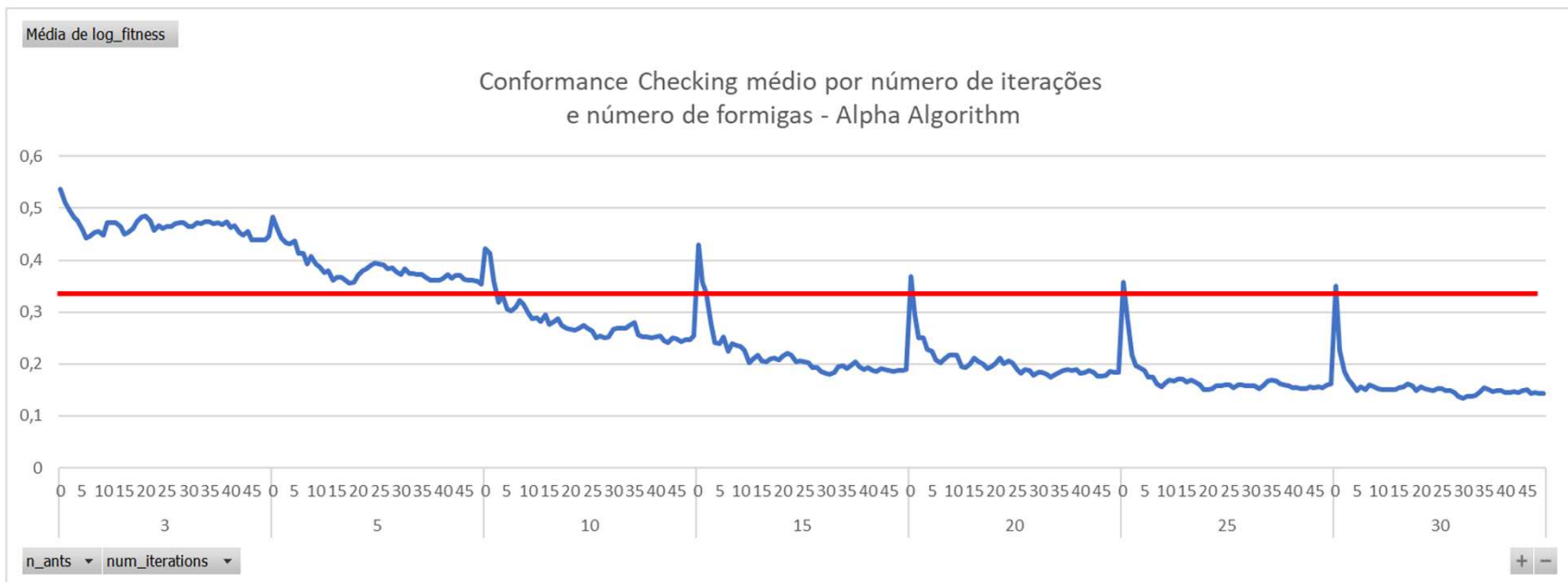


Experimento Preliminar

FITNESS POR N° DE FORMIGAS - AA

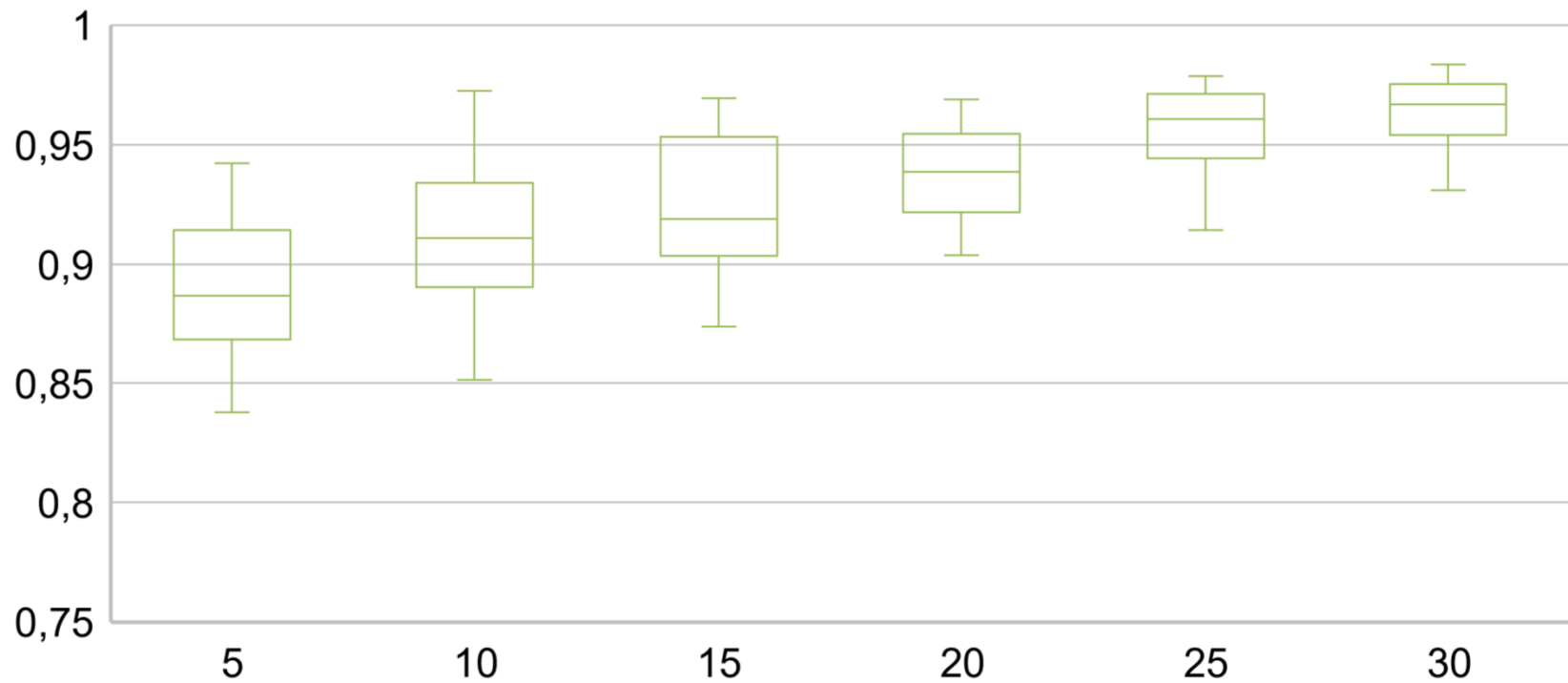


Experimento Preliminar

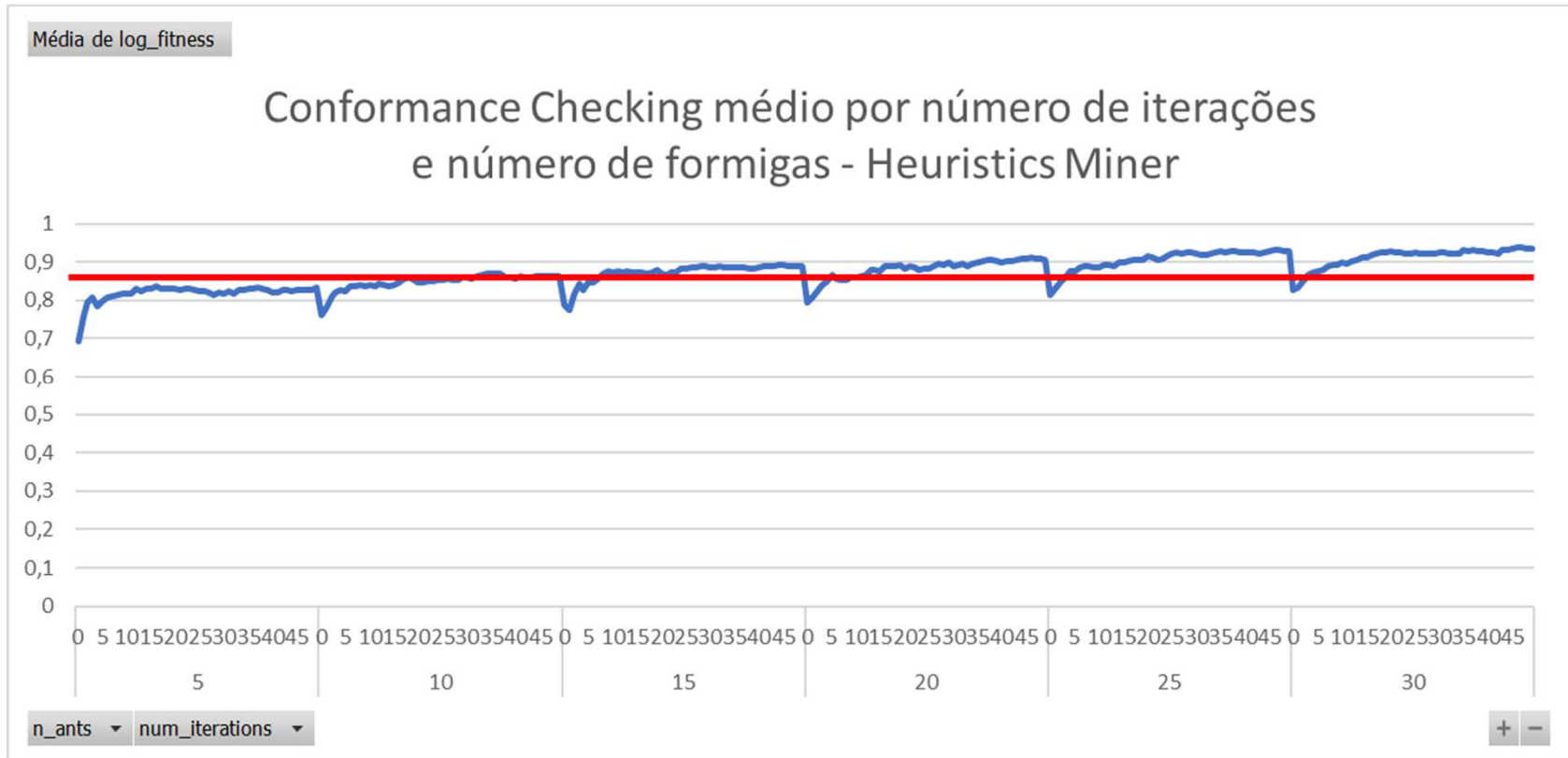


Experimento Preliminar

FITNESS POR N° DE FORMIGAS - HM

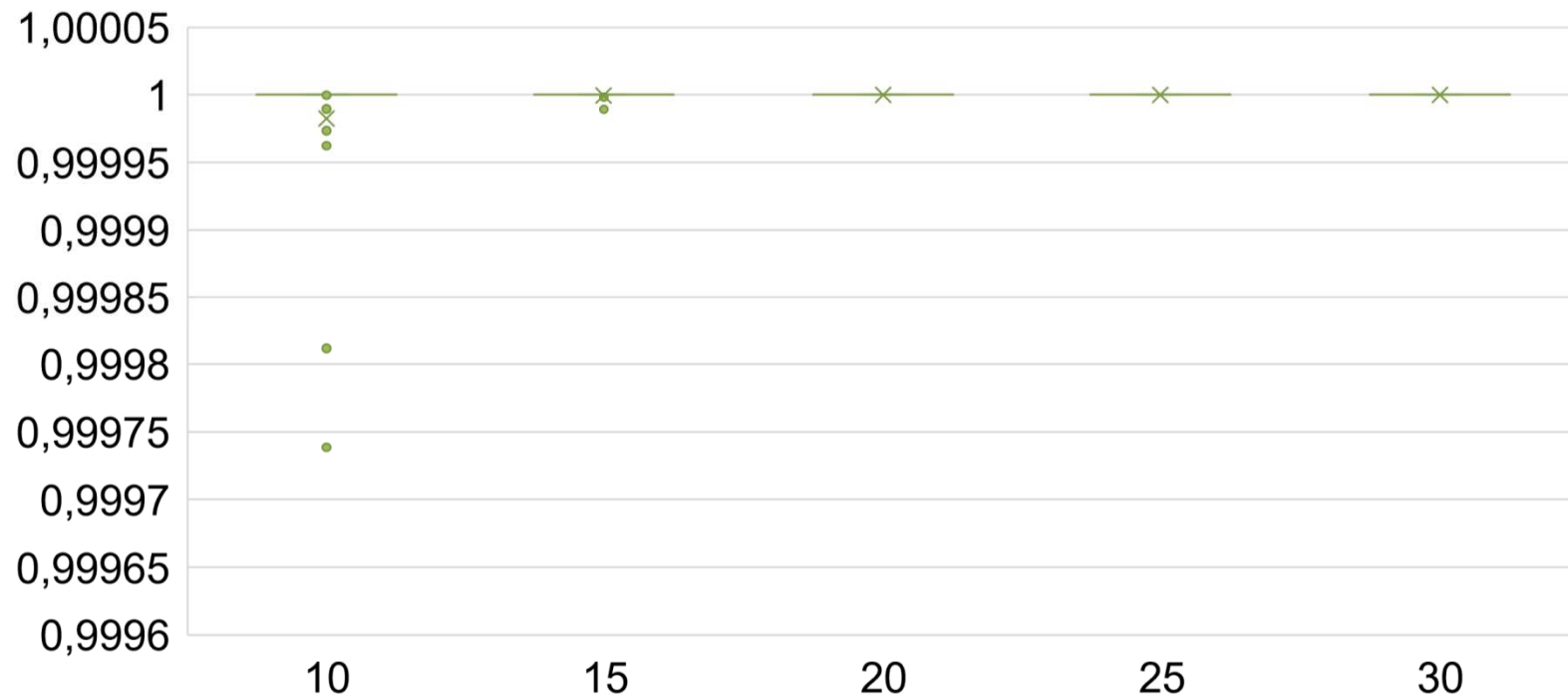


Experimento Preliminar



Experimento Preliminar

FITNESS POR N° DE FORMIGAS - IM



Experimento Preliminar



Experimento Final

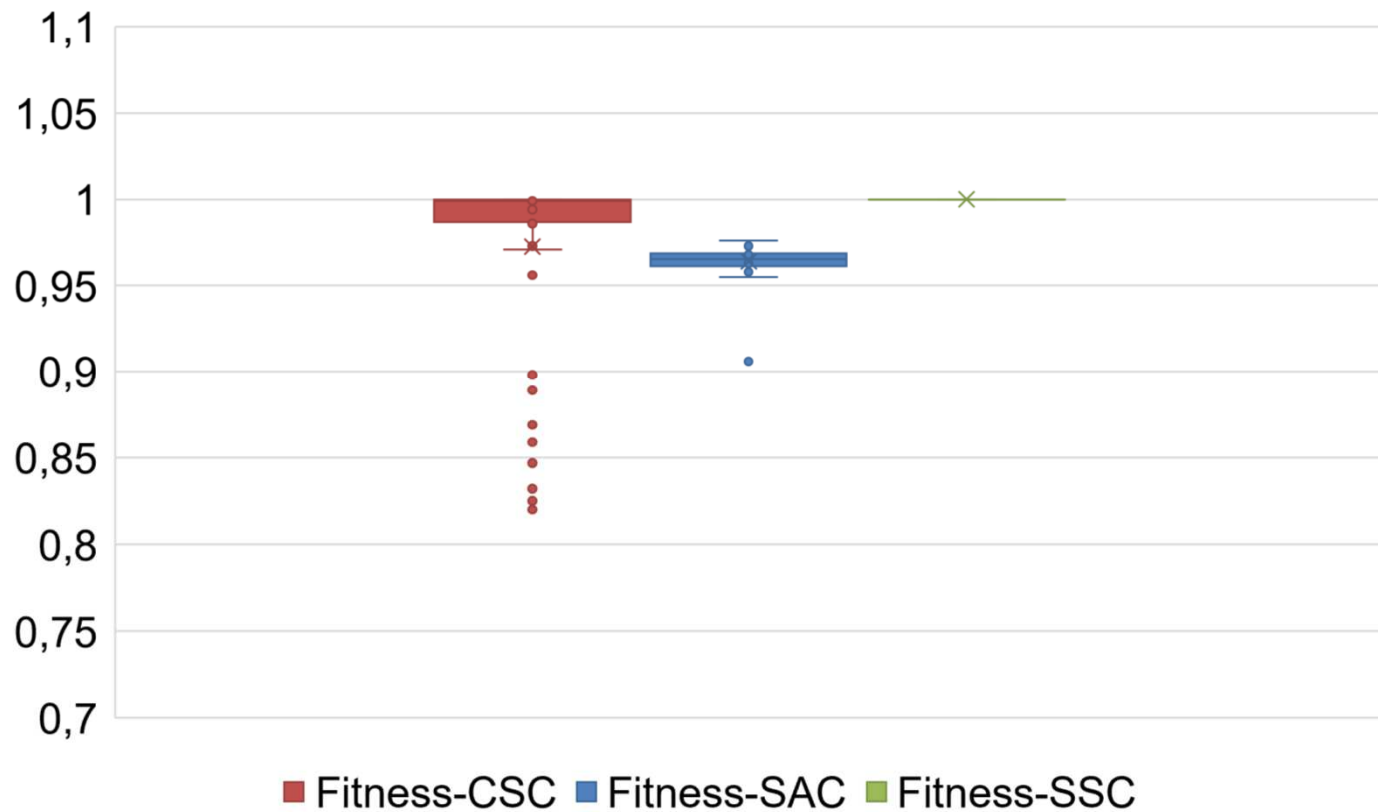
Inductive Miner - IM

Qualidade-SSC	Fitness-SSC	Precision-SSC	Simplicity-SSC
0,6033	1,0000	0,3430	0,9459

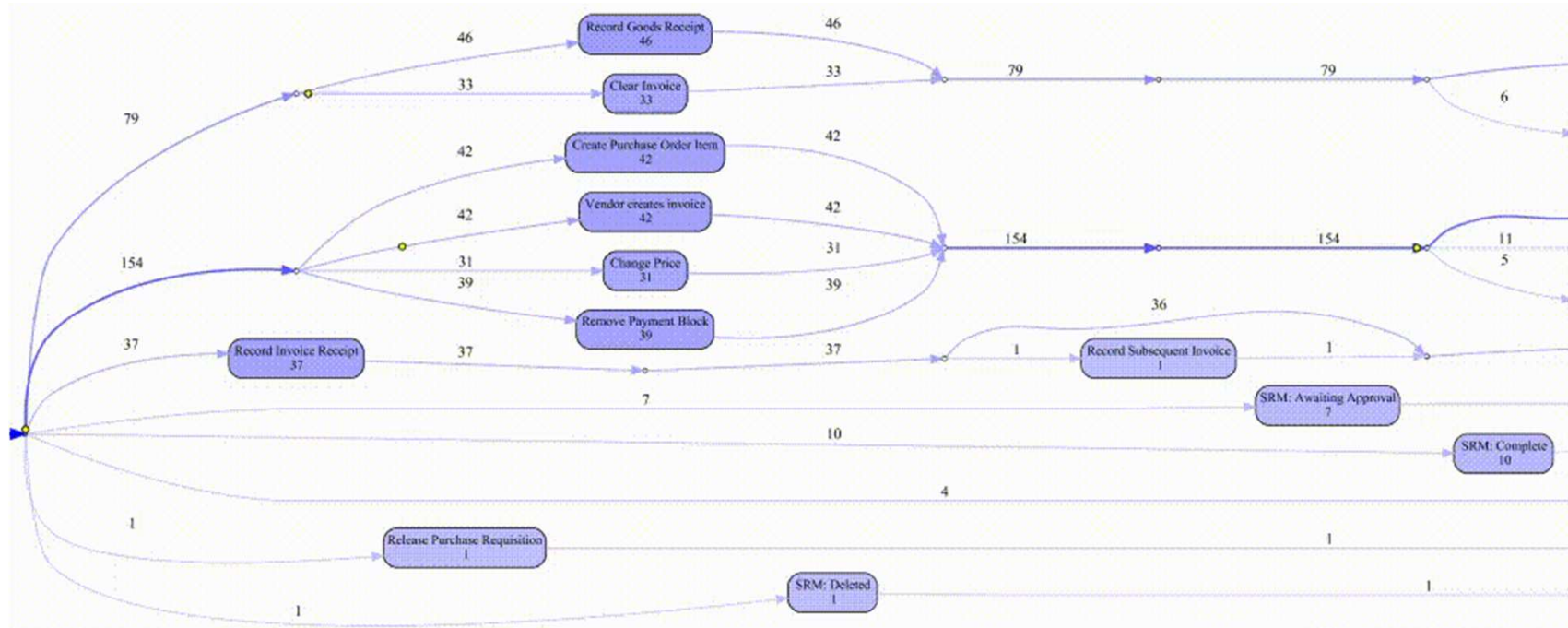
Parâmetros	Valores
ω	50,00
τ_0	0,10
ρ	0,10
α	- 1,00
β	1,00
q	100,00

Experimento Final

Dimensão *Fitness* (*Alignment*)



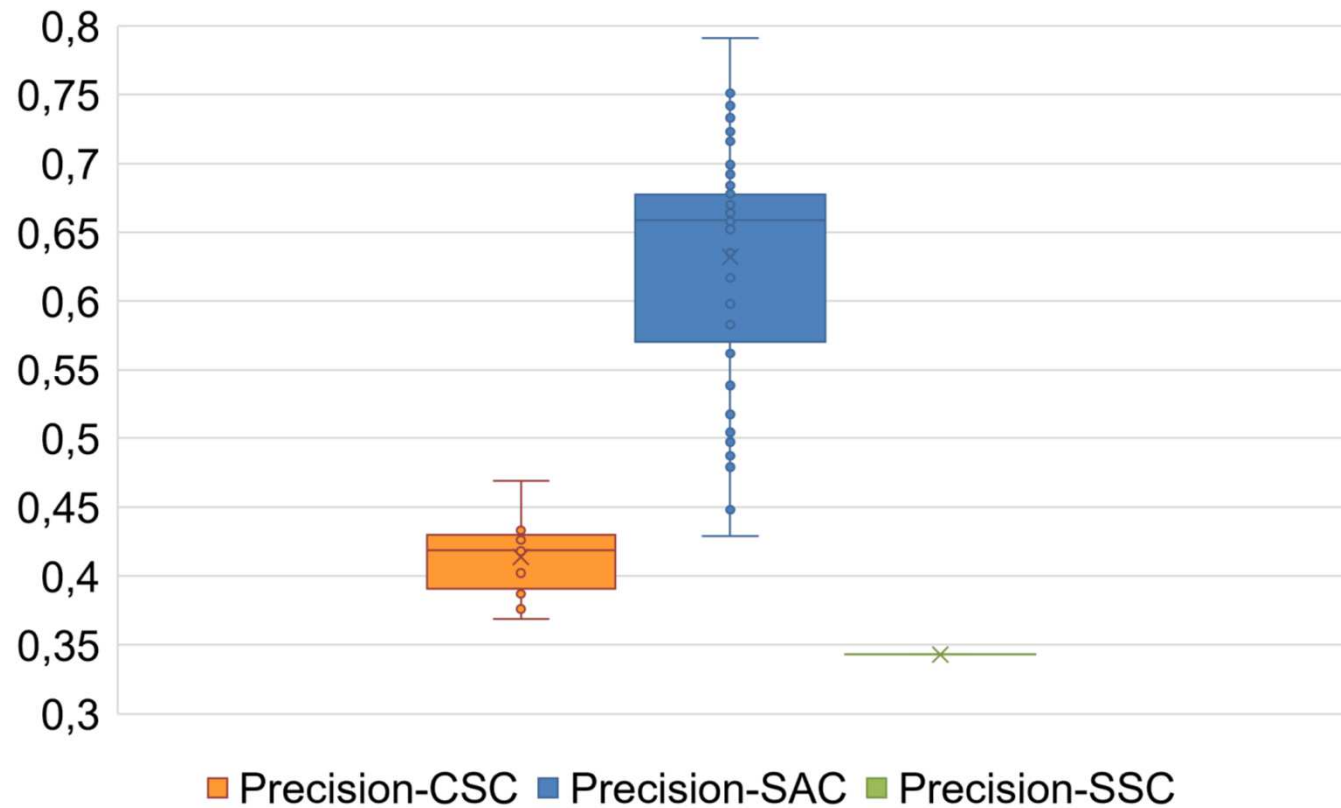
Experimento Final



Título: Modelo descoberto CSC – Exp. 48.

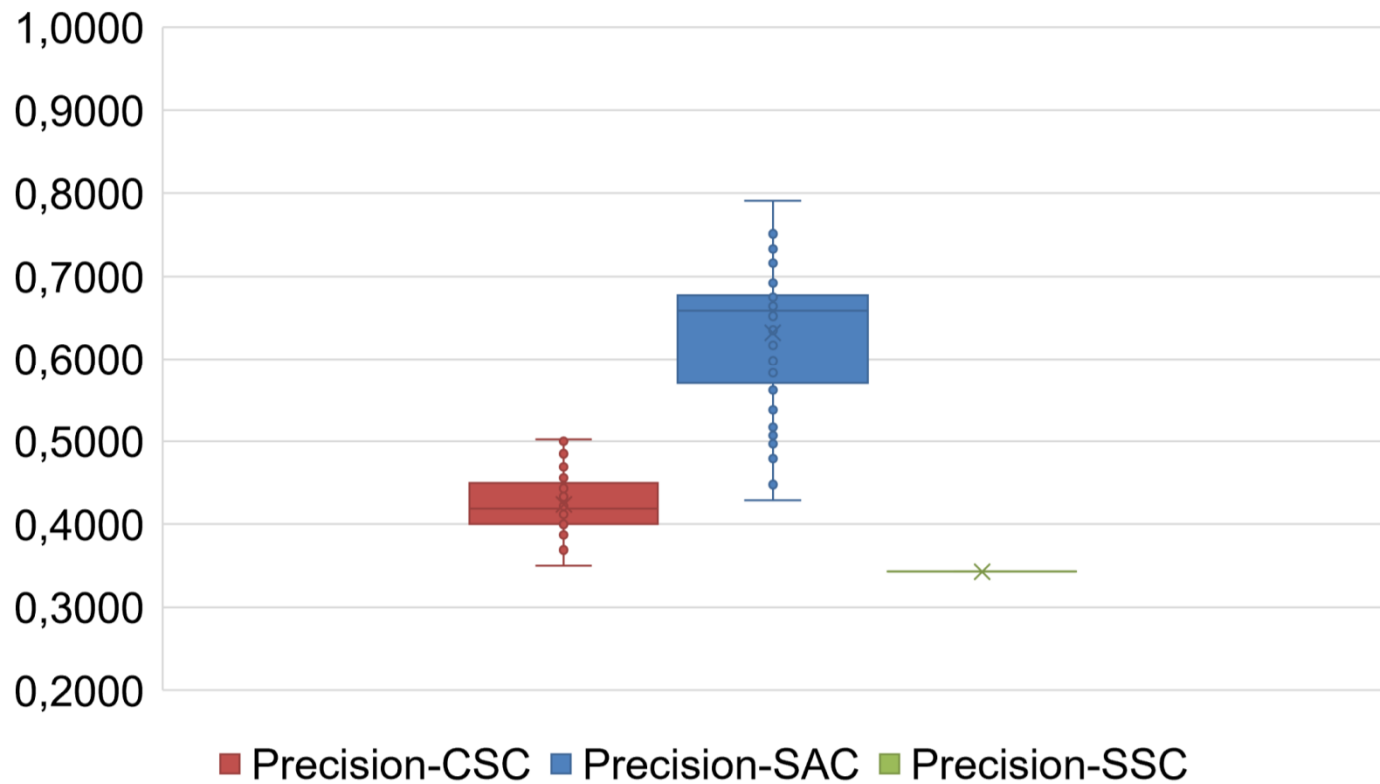
Experimento Final

Dimensão *Precision*



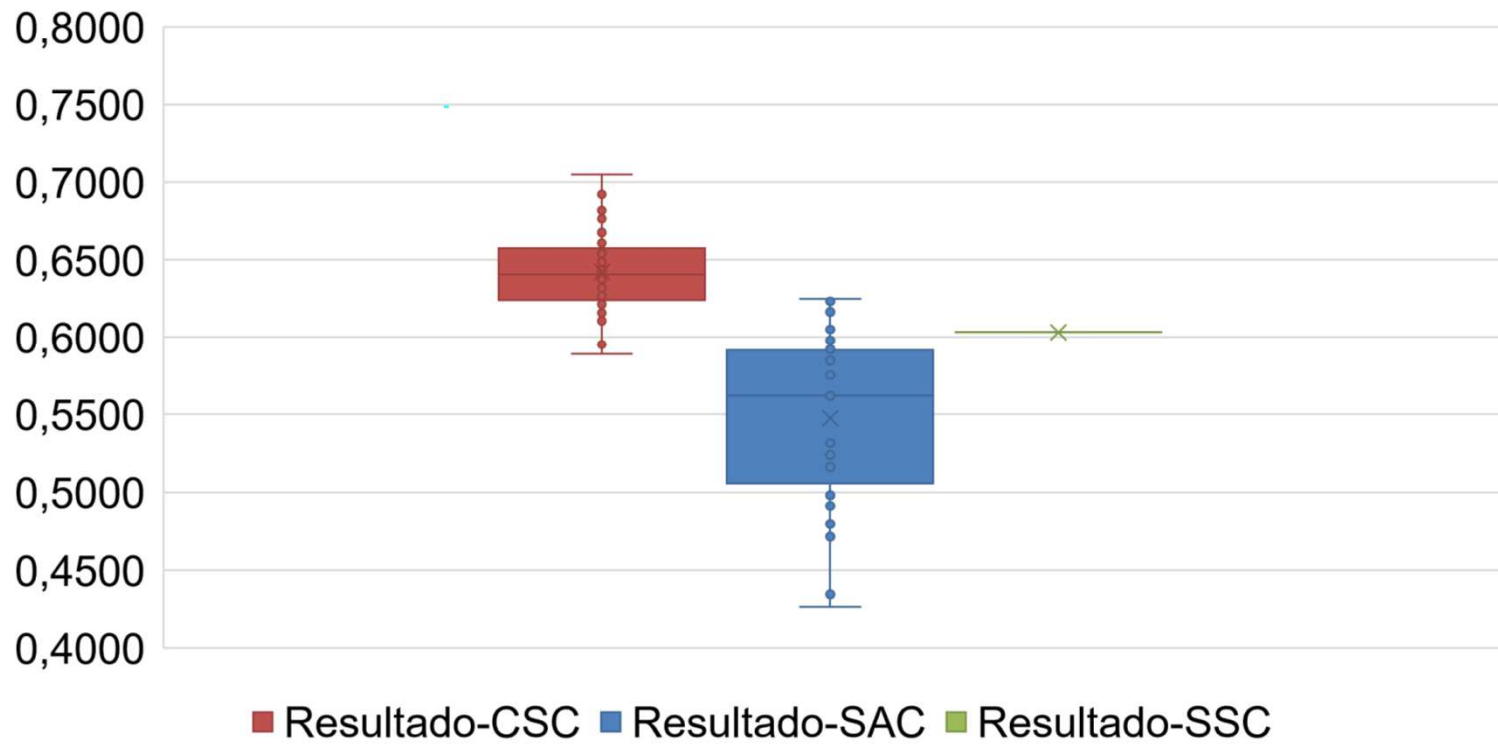
Experimento Final

Dimensão *Simplicity*



Experimento Final

Qualidade



CONCLUSÕES

Recife-PE, Brasil, 09/09/2020

Seleção de características de modelos de processos usando técnicas de Inteligência de Enxames

Conclusões

- **Demonstrada a adequabilidade** das técnicas de seleção de características à área de *Process Mining*;
- **Identificadas oportunidades** de pesquisa na área;
- **Desenvolvido um novo método** adaptado de seleção de características;
- O método FS-APM produziu **níveis de Qualidade** dos modelos descobertos **superiores ao inicialmente aferidos**;
- **Simplificada a etapa de geração de modelos** de processos com a redução de entradas;

Conclusões

- O conjunto final médio de seleção de características representou apenas **0,25%** da quantidade total de **eventos** (395.017);
- O resultado foi produzido com apenas 30 traços ótimos, ou seja, **0,042% dos casos** do log de eventos;
- É possível tornar **o modelo escalável**, reduzindo o tempo de processamento na aplicação de outras técnicas;
- Poderá produzir impacto na eficácia das organizações com a **redução dos custos** processuais e o **aperfeiçoamento das técnicas de *Process Mining***.

Contribuições

- 1) **Utilização efetiva de Inteligência de Enxames** para seleção de características com foco na descoberta de processos;
- 2) **Seleção de características** a partir da geração de **traços ótimos**;
- 3) **Geração de espaço de busca** para metaheurísticas;
- 4) **Incorporação das quatro dimensões *fitness*, *precision*, *simplicity* e *generalization*** como objetivos múltiplos;
- 5) **Feromônio repelente** para garantir **diversidade** de busca da colônia;
- 6) **Resultados** considerando **três dimensões (*fitness*, *precision* e *simplicity*)**.

Trabalhos Futuros

- 1) **Uso e aperfeiçoamento** de Inteligência de Enxames para log de eventos reais para a produção de melhores resultados em *Process Mining*;
- 2) **Estudo** na variação dos parâmetros do FS-APM;
- 3) **Avaliação da eficiência produzida** considerando o custo computacional do método proposto;
- 4) **Execução do FS-APM multiobjetivo**;
- 5) **Incorporação** de feromônios atrativos e repulsivos conjuntamente com diferentes medidas de decaimento;
- 6) **Estudo das medidas *Precision*** na aferição da qualidade considerando o impacto causado pela **filtragem de classes de atividades**.



OBRIGADO

*Seleção de características de modelos de processos
usando técnicas de Inteligência de Enxames*

Renato Barbosa Cirne

Orientador: Fernando Buarque de Lima Neto

University of Pernambuco – UPE

School of Engineering – POLI

Computational Intelligent Research Group – CIRG



At CIRG-UPE Nature insights nurture our creativity