Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Heuristic Process-Mining

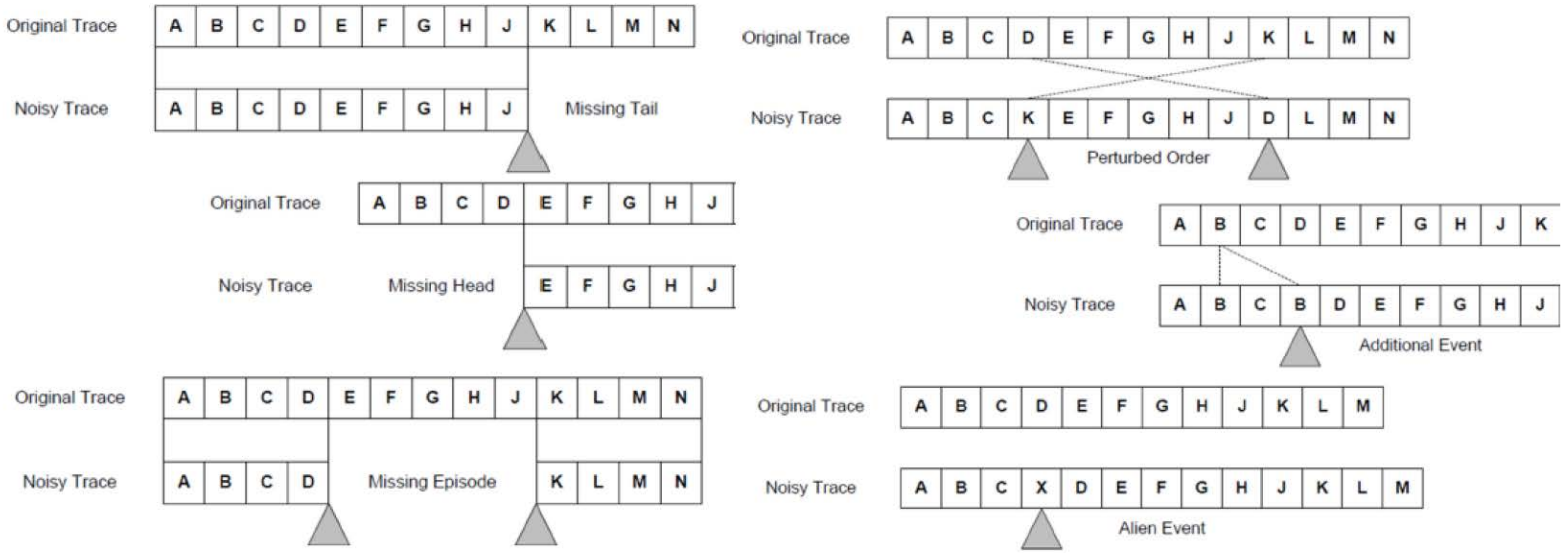
Jani Podlesny

basic problem - Alpha Algorithmus

α(+/++) – Algorithms are not robust against noise

- (!) Any direct successor relation is processed
- Process model is created, which also reflects the problem exactly!

basic problem - Noise



Quelle: Prof. Dr. Weske - POIS II Vorlesung

basic problem – noisy behavoir

„It is impossible to make reliable statements about noisy behavoir given the small set of observation“

Prof. Dr.ir. W.M.P. (Wil) van der Aalst

Example with Alpha Algorithm

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

Example with Alpha Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

a > e	a > b	a > c	a > b	a > c	a > d	a > d	a > d
	b > c	c > b	b > e	c > e	d > e	d > d	d > d
	c > e	b > e				d > e	d > e
		c > e					

Example with Alpha Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

a > e	a > b	a > c	a > b	a > c	a > d	a > d	a > d
	b > c	c > b	b > e	c > e	d > e	d > d	d > d
	c > e	b > e				d > e	d > e
		c > e					

Example with Alpha Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

1. a > b
   a > c
   a > d
   a > e
   b > c
   b > e
   c > b
   c > e
   d > d
   d > e

Example with Alpha Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

1.
a > b
a > c
a > d
a > e
b > c
b > e
c > b
c > e
d > d
d > e

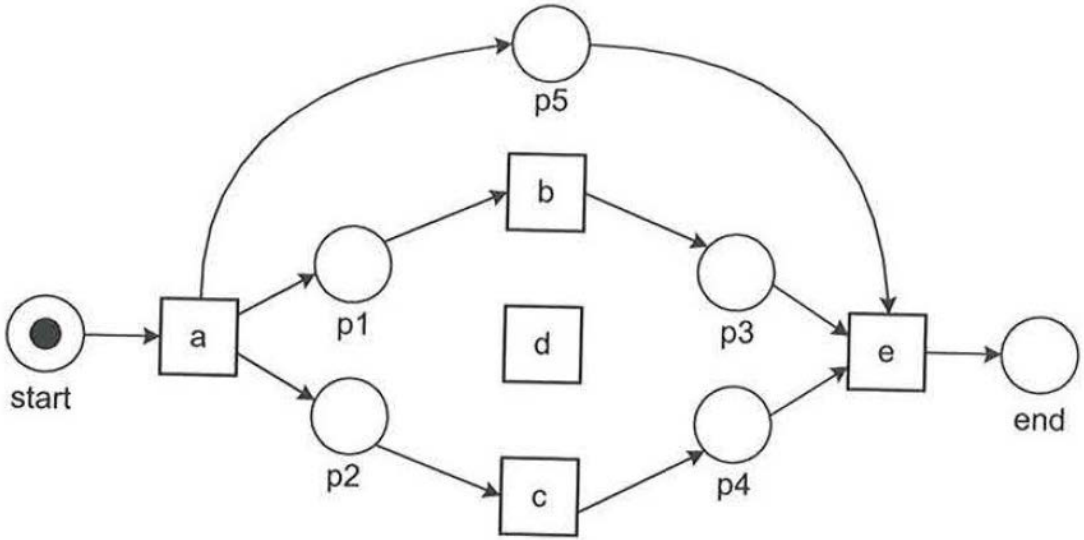
2.
a -> b b || c
a -> c
a -> d
a -> e
b -> e
c -> e
d -> e

3.
{(a),(d,e)}
{(a),(c)}
{(a),(b)}
{(b),(e)}
{(c),(e)}
{(d,a),(e)}

4.
$T_i = \{a\}$
$T_o = \{e\}$

Example with Alpha Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

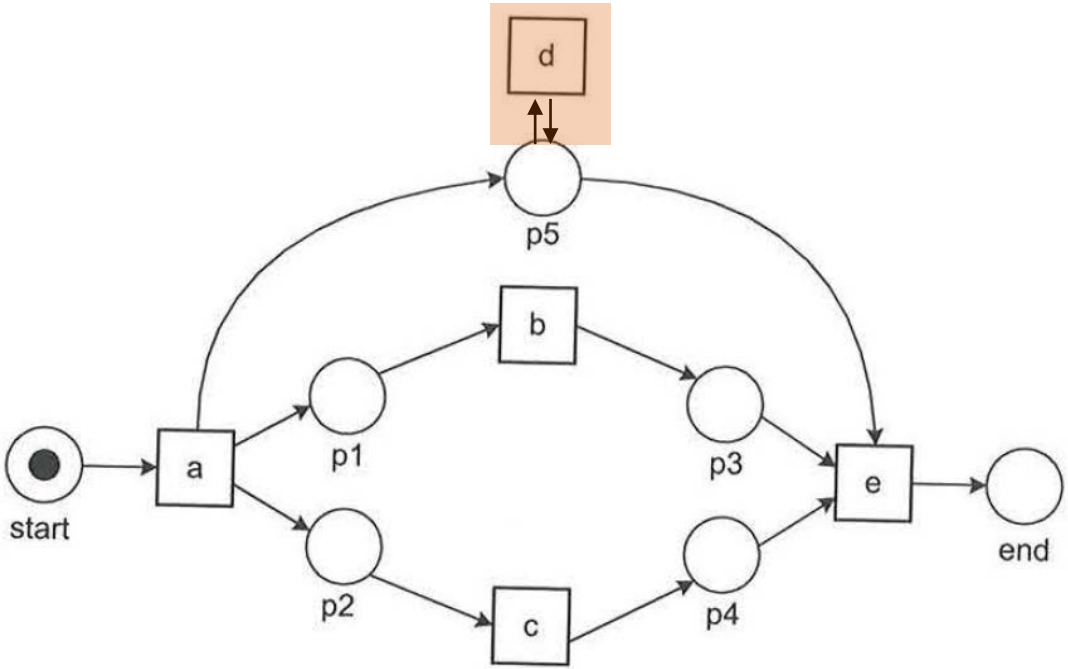


3. {(a),(d,e)}
{(a),(c)}
{(a),(b)}
{(b),(e)}
{(c),(e)}
{(d,a),(e)}

4. $T_i = \{a\}$
$T_o = \{e\}$

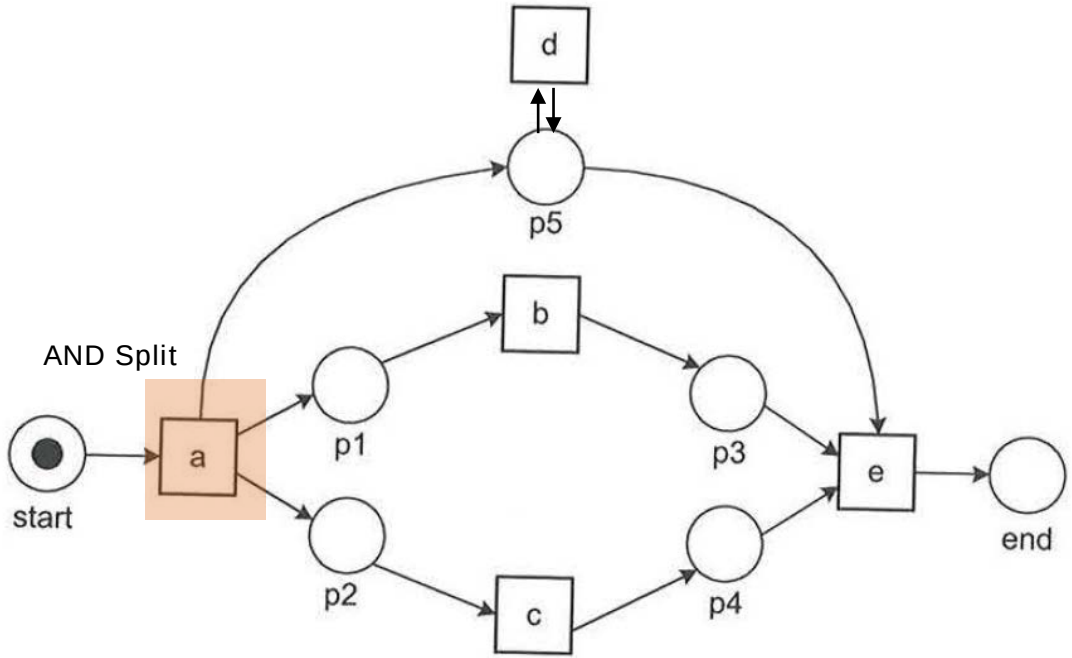
Example with Alpha(+) Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]



Example with Alpha(+) Algorithm – „directly follows“

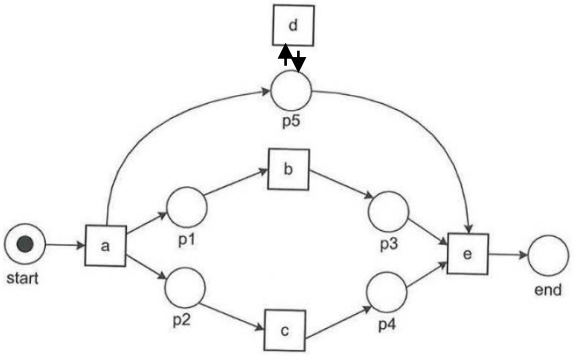
L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]



Problems with Alpha & Alpha(+) Algorithm

- Has a representational bias, which does not allow for skipping activities
  (e.g. jumping from a to e)

- Does not consider frequencies



Heuristic Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

$$|d >_L d| = 4$$

Heuristic Algorithm – „directly follows“

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

$|d >_L d| = 4$

$\|>_L\|$	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0

Heuristic Algorithm – „directly follows“

Definition 6.1 (Dependency measure) Let L be an event log[1] over $\mathscr{A}$ and $a, b \in \mathscr{A}$. $|a >_L b|$ is the number of times a is directly followed by b in L, i.e.,

$$|a >_L b| = \sum_{\sigma \in L} L(\sigma) \times \left|\left\{1 \leq i < |\sigma| \mid \sigma(i) = a \wedge \sigma(i+1) = b\right\}\right|$$

Source: „Process Mining“, Will M.P. van der Aalst, Springer Verlag

Heuristic Algorithm – dependency relation

If $|a \Rightarrow_L b|$

$$f(x) = \frac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1}$$

dependency between a and b

-1 alias strong negative

1 alias strong positive

Heuristic Algorithm – dependency relation (exception)

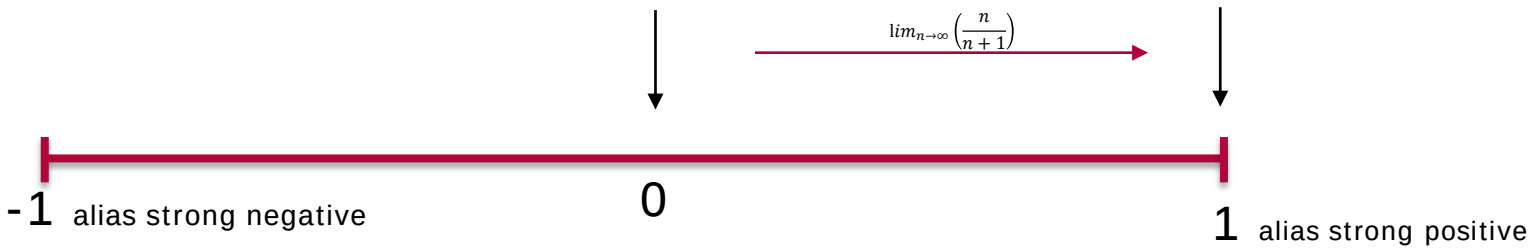
If $|a \Rightarrow_L b|$ where a = b

$$f(x) = \frac{|a>_L a| - |a>_L a|}{|a>_L a| + |a>_L a| + 1} = 0$$

Heuristic Algorithm – dependency relation (exception)

If $|a \Rightarrow_L b|$ where a = b

$$f(x) = \frac{|a>_L a|}{|a>_L a|+1} = 0 \qquad \Leftrightarrow \lim_{n\to\infty}\left(\frac{n}{n+1}\right) = 1$$



Heuristic Algorithm – dependency relation

Function for calculating dependency relation between a and b

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\[2ex] \dfrac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases}$$

Heuristic Algorithm – dependency relation

$$L = [\ (a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1]$$

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\[2ex] \dfrac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases}$$

$\lvert\Rightarrow_L\rvert$	a	b	c	d	e
a	$\frac{0}{0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{13-0}{13+0+1} = 0.93$	$\frac{5-0}{5+0+1} = 0.83$
b	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0}{0+1} = 0$	$\frac{10-10}{10+10+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
c	$\frac{0-11}{0+11+1} = -0.92$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
d	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0-0}{0+0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{4}{4+1} = 0.80$	$\frac{13-0}{13+0+1} = 0.93$
e	$\frac{0-5}{0+5+1} = -0.83$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0}{0+1} = 0$

Heuristic Algorithm – dependency relation

$$L = [\ (a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1]$$

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\[2ex] \dfrac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases}$$

$\lvert\Rightarrow_L\rvert$	a	b	c	d	e
a	$\frac{0}{0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{13-0}{13+0+1} = 0.93$	$\frac{5-0}{5+0+1} = 0.83$
b	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0}{0+1} = 0$	$\frac{10-10}{10+10+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
c	$\frac{0-11}{0+11+1} = -0.92$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
d	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0-0}{0+0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{4}{4+1} = 0.80$	$\frac{13-0}{13+0+1} = 0.93$
e	$\frac{0-5}{0+5+1} = -0.83$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0}{0+1} = 0$

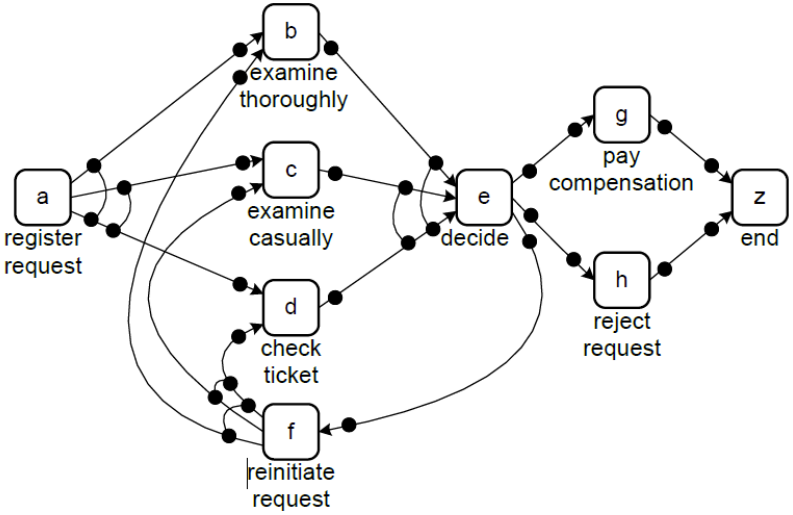
Heuristic Algorithm – dependency relation

$$L = [\ (a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1]$$

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\[2ex] \dfrac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases}$$

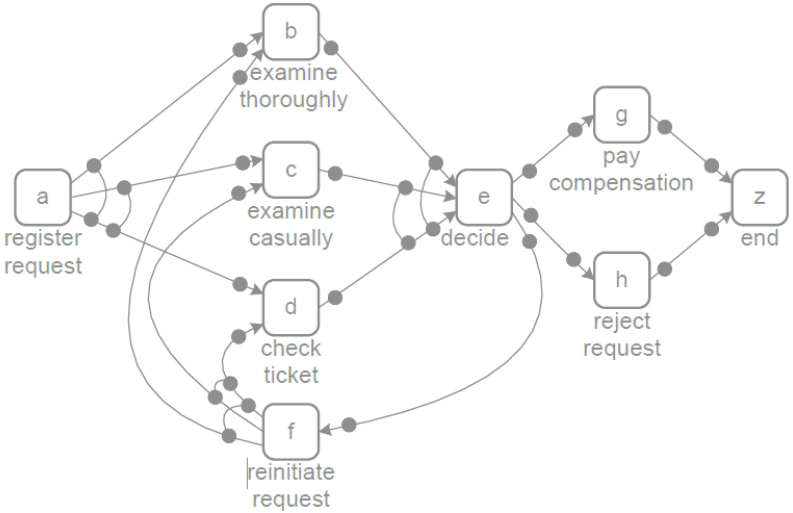
$\lvert\Rightarrow_L\rvert$	a	b	c	d	e
a	$\frac{0}{0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{13-0}{13+0+1} = 0.93$	$\frac{5-0}{5+0+1} = 0.83$
b	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0}{0+1} = 0$	$\frac{10-10}{10+10+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
c	$\frac{0-11}{0+11+1} = -0.92$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
d	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0-0}{0+0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{4}{4+1} = 0.80$	$\frac{13-0}{13+0+1} = 0.93$
e	$\frac{0-5}{0+5+1} = -0.83$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0}{0+1} = 0$

Causal Net

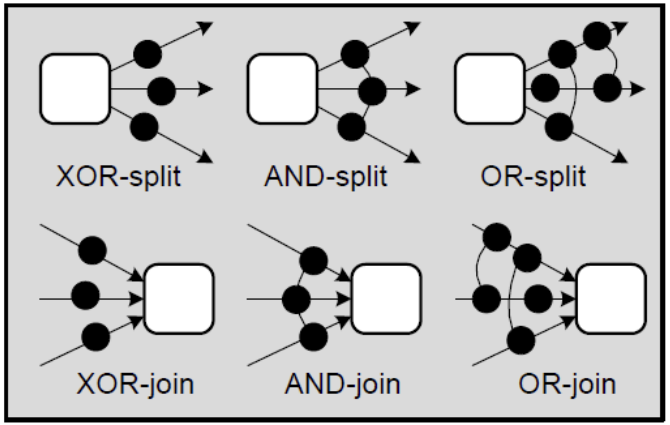


Quelle: „Process Mining“, Will M.P. van der Aalst, Springer Verlag

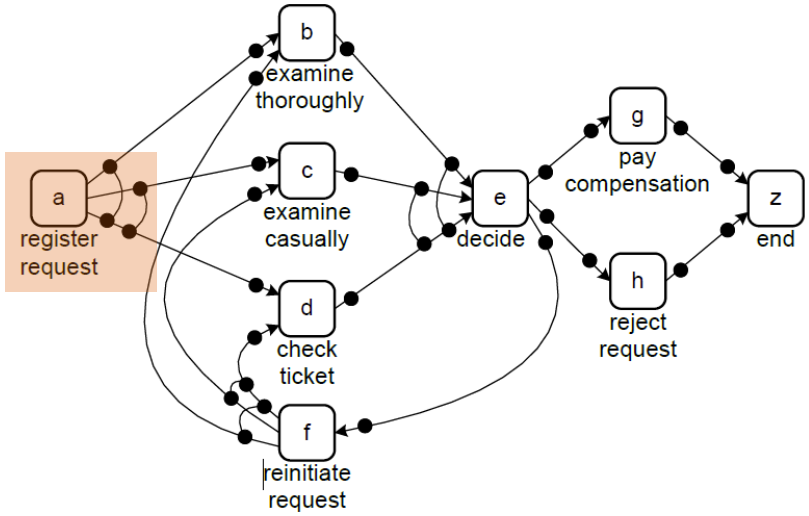
Causal Net - Gateways



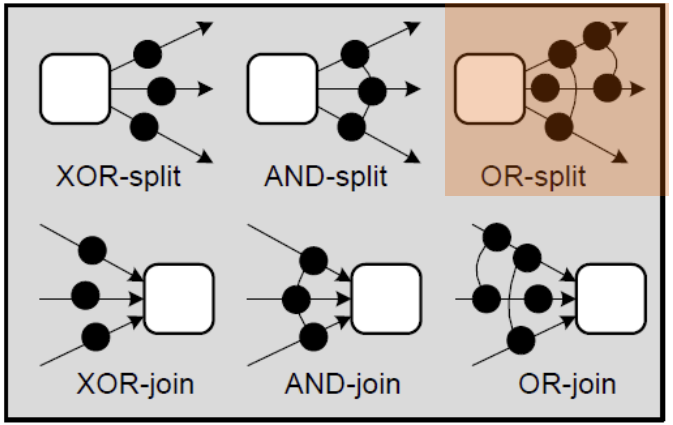
Quelle: „Process Mining“, Will M.P. van der Aalst, Springer Verlag



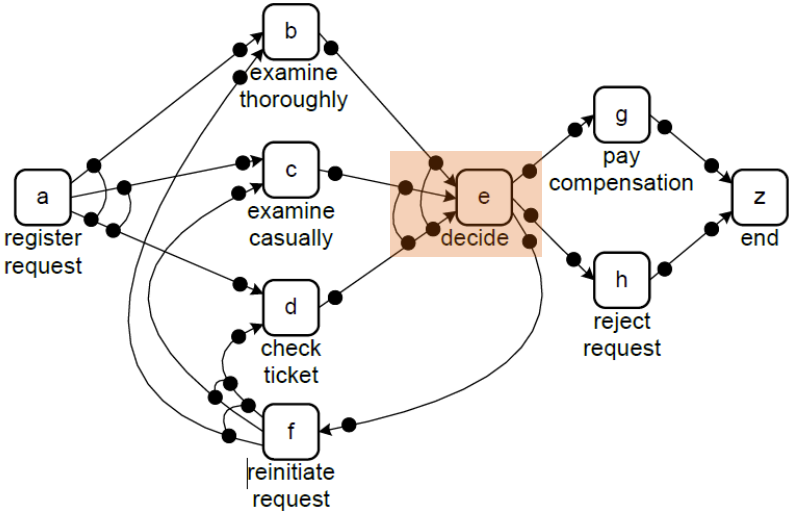
Causal Net - Gateways



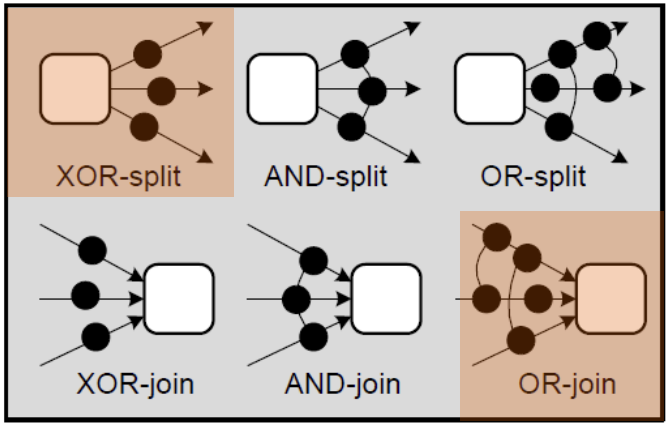
Quelle: „Process Mining“, Will M.P. van der Aalst, Springer Verlag



Causal Net - Gateways



Quelle: „Process Mining“, Will M.P. van der Aalst, Springer Verlag



Causal Net – formal Definition

Definition 2.8 (Causal net) A *Causal net* (C-net) is a tuple $C = (A, a_i, a_o, D, I, O)$ where:

- $A \subseteq \mathscr{A}$ is a finite set of *activities*
- $a_i \in A$ is the *start activity*
- $a_o \in A$ is the *end activity*
- $D \subseteq A \times A$ is the *dependency relation*
- $AS = \{X \subseteq \mathscr{P}(A) \mid X = \{\emptyset\} \vee \emptyset \notin X\}^2$
- $I \in A \rightarrow AS$ defines the set of possible *input bindings* per activity
- $O \in A \rightarrow AS$ defines the set of possible *output bindings* per activity

such that

- $D = \{(a_1, a_2) \in A \times A \mid a_1 \in \bigcup_{as \in I(a_2)} as\}$
- $D = \{(a_1, a_2) \in A \times A \mid a_2 \in \bigcup_{as \in O(a_1)} as\}$
- $\{a_i\} = \{a \in A \mid I(a) = \{\emptyset\}\}$
- $\{a_o\} = \{a \in A \mid O(a) = \{\emptyset\}\}$
- All activities in the graph (A, D) are on a path from a_i to a_o

Source: „Process Mining“, Will M.P. van der Aalst, Springer Verlag

Heuristic Algorithm – dependency relation

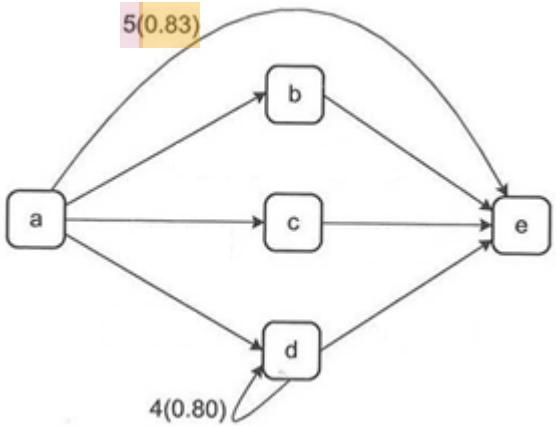
$$L = [\ (a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1]$$

$$|a \Rightarrow_L b| = \begin{cases} \dfrac{|a >_L b| - |b >_L a|}{|a >_L b| + |b >_L a| + 1} & \text{if } a \neq b \\[2ex] \dfrac{|a >_L a|}{|a >_L a| + 1} & \text{if } a = b \end{cases}$$

$\lvert\Rightarrow_L\rvert$	a	b	c	d	e
a	$\frac{0}{0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{11-0}{11+0+1} = 0.92$	$\frac{13-0}{13+0+1} = 0.93$	$\frac{5-0}{5+0+1} = 0.83$
b	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0}{0+1} = 0$	$\frac{10-10}{10+10+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
c	$\frac{0-11}{0+11+1} = -0.92$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{11-0}{11+0+1} = 0.92$
d	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0-0}{0+0+1} = 0$	$\frac{0-0}{0+0+1} = 0$	$\frac{4}{4+1} = 0.80$	$\frac{13-0}{13+0+1} = 0.93$
e	$\frac{0-5}{0+5+1} = -0.83$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-11}{0+11+1} = -0.92$	$\frac{0-13}{0+13+1} = -0.93$	$\frac{0}{0+1} = 0$

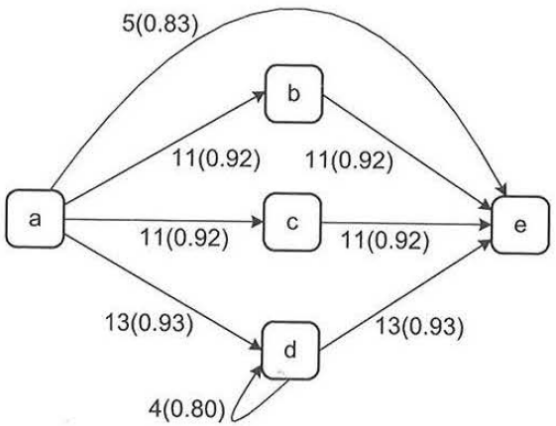
Heuristic Algorithm – Dependency Graph

L = [$(a,e)^{5}, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^{1}, (a,c,e)^{1}, (a,d,e)^{10}, (a,d,d,e)^{2}, (a,d,d,d,e)^{1}$]



Heuristic Algorithm – Dependency Graph

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]



$|\Rightarrow_L|$

$\|>_L\|$	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0

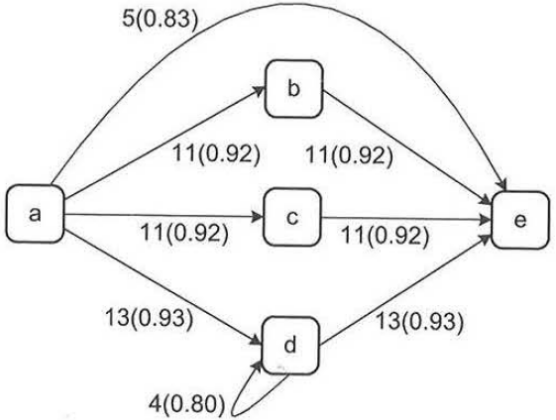
$|>_L|$

$\|\Rightarrow_L\|$	a	b	c	d	e
a	$\frac{0}{0+1}=0$	$\frac{11-0}{11+0+1}=0.92$	$\frac{11-0}{11+0+1}=0.92$	$\frac{13-0}{13+0+1}=0.93$	$\frac{5-0}{5+0+1}=0.83$
b	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0}{0+1}=0$	$\frac{10-10}{10+10+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
c	$\frac{0-11}{0+11+1}=-0.92$	$\frac{10-10}{10+10+1}=0$	$\frac{0}{0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{11-0}{11+0+1}=0.92$
d	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0-0}{0+0+1}=0$	$\frac{0-0}{0+0+1}=0$	$\frac{4}{4+1}=0.80$	$\frac{13-0}{13+0+1}=0.93$
e	$\frac{0-5}{0+5+1}=-0.83$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-11}{0+11+1}=-0.92$	$\frac{0-13}{0+13+1}=-0.93$	$\frac{0}{0+1}=0$

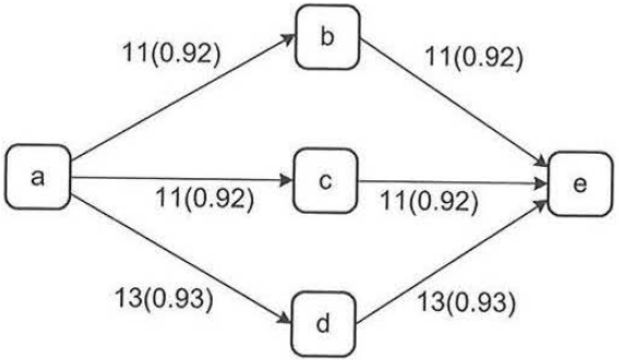
Heuristic Algorithm – Dependency Graph

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

Threshold:
$|\Rightarrow_L| = 2$
$|>_L| = 0.7$



Threshold:
$|\Rightarrow_L| = 5$
$|>_L| = 0.9$



Heuristic Algorithm – Splits & Joins

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

Replay for treshold $\begin{cases} |\Rightarrow_L| = 2 \\ |>_L| = 0.7 \end{cases}$

$(a,e)^5$ ⇔ a is 5 times directly followed by e

$(a,b,c,e)^{10}, (a,c,b,e)^{10}$ ⇔ a is 20 times followed by both b <u>and</u> c

$(a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$ ⇔ a is 13 times followed by d

Heuristic Algorithm – Splits & Joins

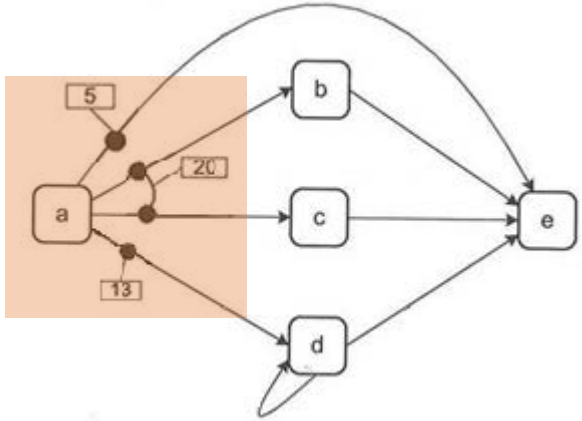
L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]

Replay for treshold $\begin{cases} |\Rightarrow_L| = 2 \\ |>_L| = 0.7 \end{cases}$

$(a,e)^5$ ⇔ a is 5 times directly followed by e

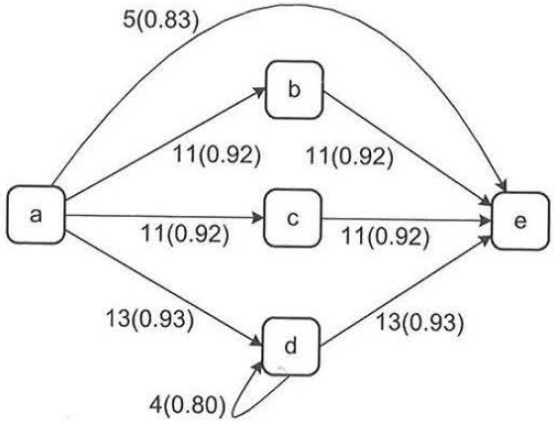
$(a,b,c,e)^{10}, (a,c,b,e)^{10}$ ⇔ a is 20 times followed by both b <u>and</u> c

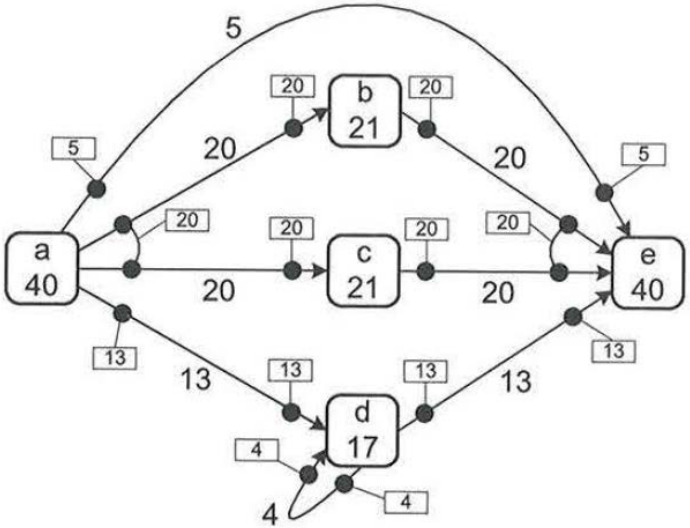
$(a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$ ⇔ a is 13 times followed by d



Heuristic Algorithm – Splits & Joins

$$L = [\ (a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1]$$

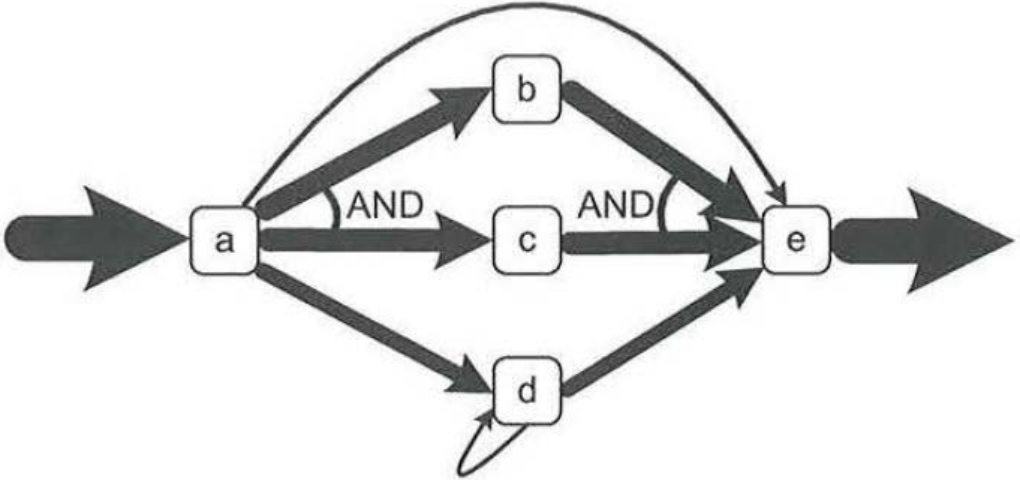




Source: „Process Mining“, Will M.P. van der Aalst, Springer Verlag

Heuristic Algorithm – Splits & Joins

L = [$(a,e)^5, (a,b,c,e)^{10}, (a,c,b,e)^{10}, (a,b,e)^1, (a,c,e)^1, (a,d,e)^{10}, (a,d,d,e)^2, (a,d,d,d,e)^1$]



Source: „Process Mining“, Will M.P. van der Aalst, Springer Verlag

Resumeé - Heuristic Mining

Representational Bias

(provided by causal nets)

&

usage of frequencies

→ makes Heuristic Mining more robust than most other approaches

Heuristic Mining - Issues

L = [$(a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}$]

Noise?!

Heuristic Mining - issues

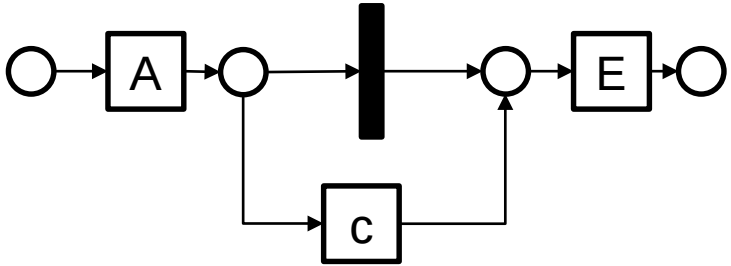
L = [$(a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}$]

- threshold apply to the entire net
- final safety only brings expert knowledge
- Should we set threshold to 3 ?

Heuristic Mining - issues

L = [$(a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}$]

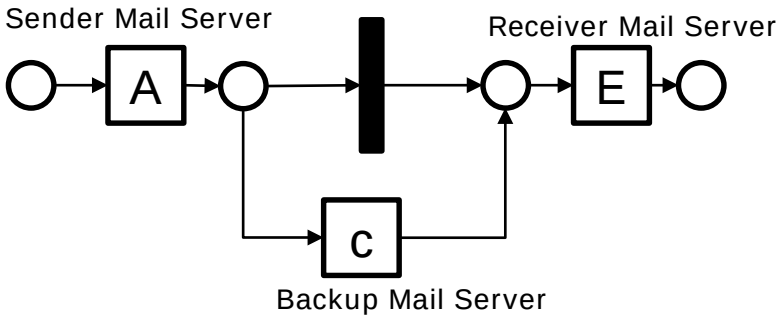
- threshold apply to the entire net
- final safety only brings expert knowledge
- should we set threshold to 3 ?



Heuristic Mining - issues

$L = [\ (a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}]$

- threshold apply to the entire net
- final safety only brings expert knowledge
- Should we set threshold to 3 ?



Pref	Hostname	IP Address
10	mail2.hpi.uni-potsdam.de	141.89.225.122
10	mail2.hpi.uni-potsdam.de	2001:638:807:204::8d59:e17a
20	mail3.hpi.uni-potsdam.de	141.89.225.123
20	mail3.hpi.uni-potsdam.de	2001:638:807:204::8d59:e17b

Heuristic Mining – my own attempt

1. Preprocessing:
   - analyse log entrie‘s name and description
   - compare keywords with matching algorithm (Collaborative Filtering)
   - if comparison was successful memorize correlating model

2. Postprocessing:
   - compare generated model with correlating model from preprocessing
   - if necessary adapt generated model

Heuristic Mining – my own attempt - example

$L = [\ (a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}]$

CaseID	Timestamp	Name
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
3	000006	Send Mail (a)
3	000007	NSA Bounce (f)
3	000008	CIA Bounce (g)
3	000009	Response Backup Mail Server (c)
3	000010	Receive Mail (e)

9999x

3x

$L = [\ (a,e), (a,c,e)]$

CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Heuristic Mining – my own attempt - example

$L = [\ (a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}]$

CaseID	Timestamp	Name
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
3	000006	Send Mail (a)
3	000007	NSA Bounce (f)
3	000008	CIA Bounce (g)
3	000009	Response Backup Mail Server (c)
3	000010	Receive Mail (e)

9999x

3x

$L = [\ (a,e), (a,c,e)]$

CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Heuristic Mining – my own attempt - example

$L = [\ (a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}]$

CaseID	Timestamp	Name	
1	000001	Send Mail (a)	9999x
1	000002	Receive Mail (e)	
2	000003	Send Mail (a)	
2	000004	Response Backup Mail Server (c)	
2	000005	Receive Mail (e)	
3	000006	Send Mail (a)	3x
3	000007	NSA Bounce (f)	
3	000008	CIA Bounce (g)	
3	000009	Response Backup Mail Server (c)	
3	000010	Receive Mail (e)	

$L = [\ (a,e), (a,c,e)]$

CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Heuristic Mining – my own attempt - example

$L = [\ (a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}]$

CaseID	Timestamp	Name
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
3	000006	Send Mail (a)
3	000007	NSA Bounce (f)
3	000008	CIA Bounce (g)
3	000009	Response Backup Mail Server (c)
3	000010	Receive Mail (e)

9999x

3x

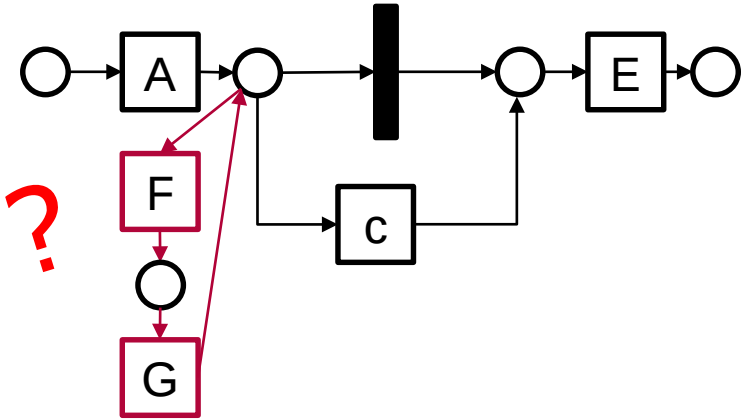
$L = [\ (a,e), (a,c,e)]$

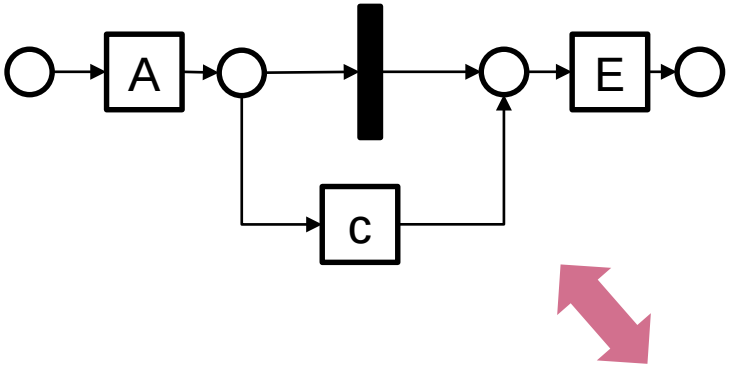
CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Heuristic Mining – my own attempt - example

L = [$(a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}$]

L = [$(a,e), (a,c,e)$]

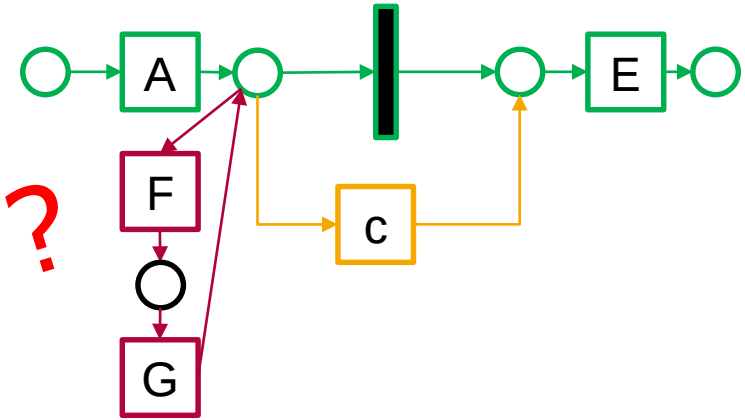




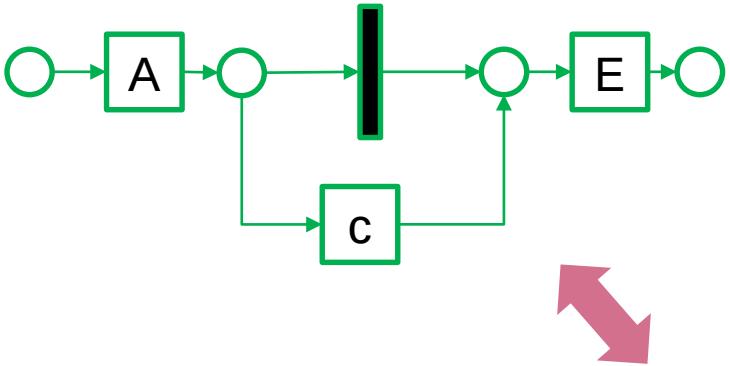
CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Heuristic Mining – my own attempt - example

L = [$(a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}$]



L = [$(a,e), (a,c,e)$]

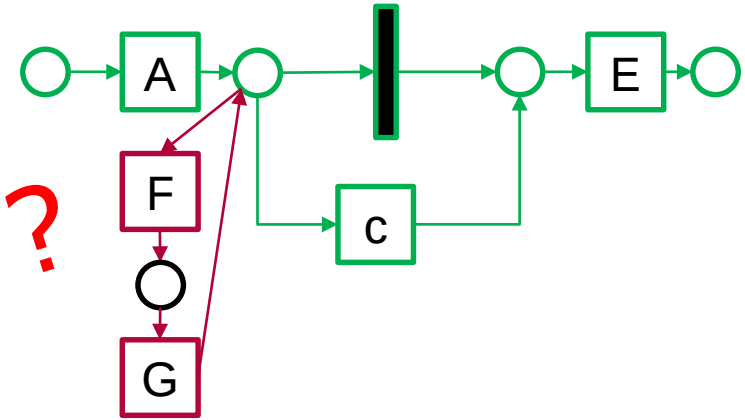


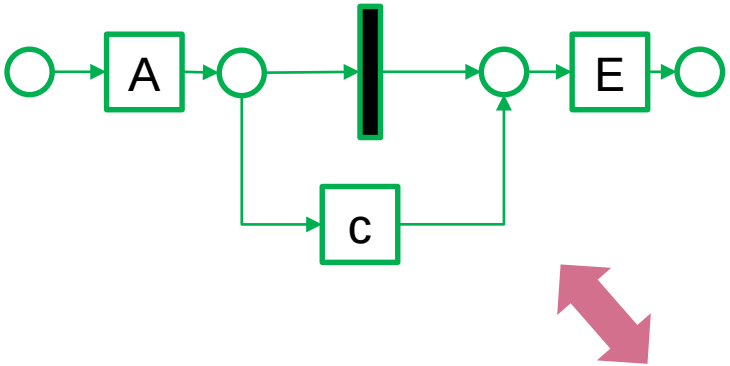
CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Heuristic Mining – my own attempt - example

L = [$(a,e)^{9999}, (a,c,e)^{1}, (a,f,g,c,e)^{3}$]

L = [$(a,e), (a,c,e)$]





CaseID	Timestamp	Name
...	...	...
1	000001	Send Mail (a)
1	000002	Receive Mail (e)
2	000003	Send Mail (a)
2	000004	Response Backup Mail Server (c)
2	000005	Receive Mail (e)
...	...	...

Conclusion

underlying problem:

- α-Algorithmus is not resistant against noise

attempt:

- Order relation presents frequency of occurence of relationships within event log

realisation:

- Causal Nets

Sources

„Process Mining“, Will M.P. van der Aalst, Springer Verlag

„Genetic Mining und Heuristic Mining”, Jannik Arndt, Universität Oldenburg

„Process Mining with the Heuristics Miner Algorithm“, Weijters, van der Aalst, Eindhoven University of Technology

„Zur Eignung des Process Mining für die Überprüfung von Sicherheitsanforderungen in prozessgestützten Informationssystemen“, Meike Ullrich, Universität Freiburg

„Heuristics Miner 6.0: Users Guide“, Weijters, Riberiro, Eindhoven University of Technology