

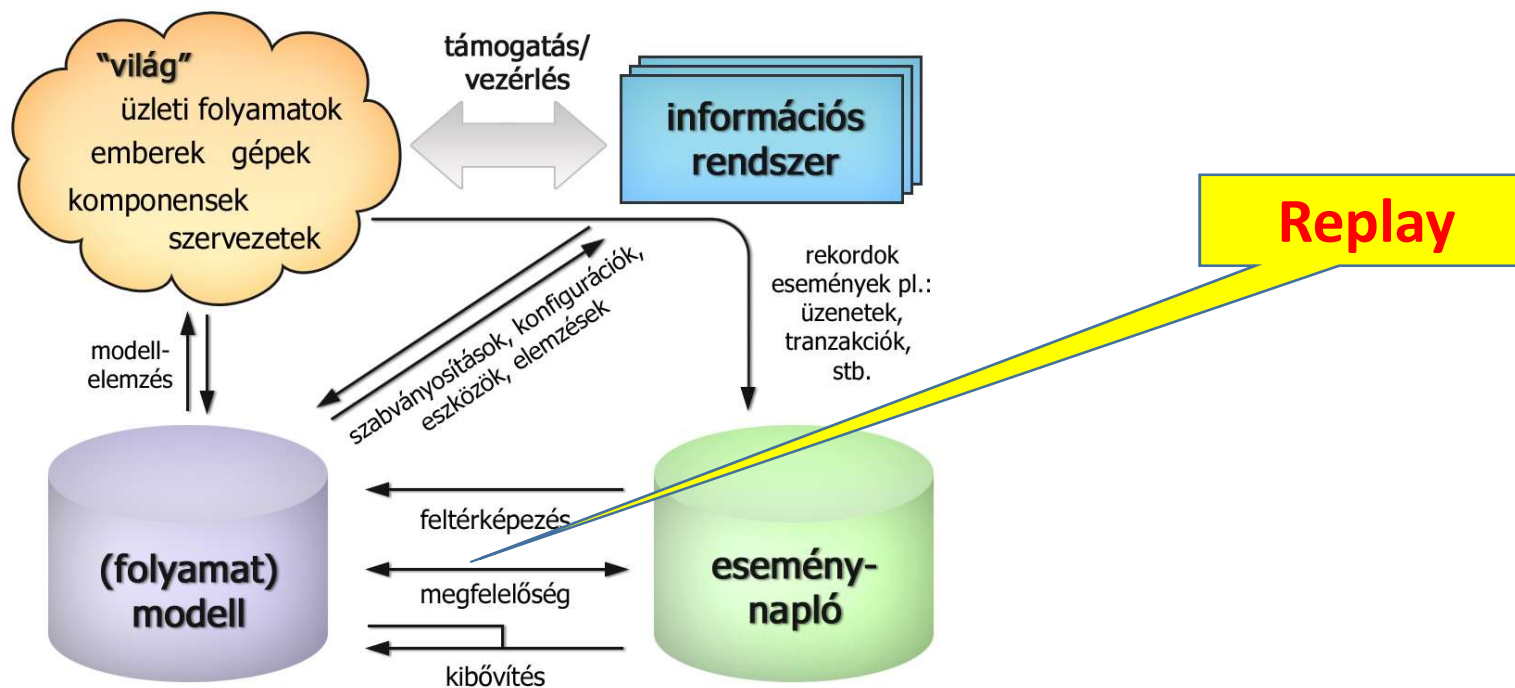
Folyamatbányászat

Konformancia elemzés – megfelelőség elemzés II.

Token alapú és illesztés alapú
megfelelőség ellenőrzés

A konformancia ellenőrzés módszerei

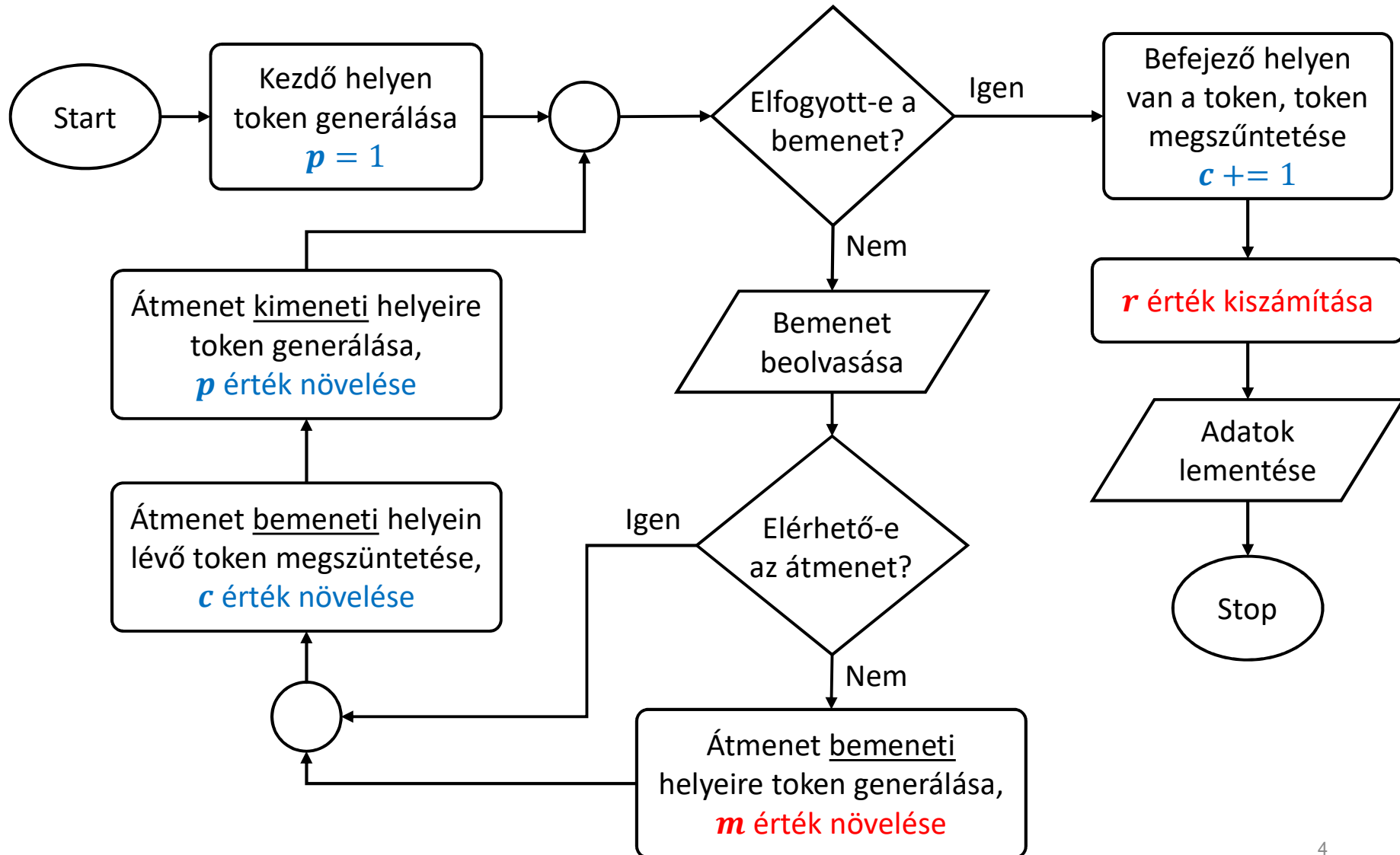
1. Konformancia ellenőrzés **ok-okozati lábnyomokkal**,
2. **Konformancia ellenőrzés token alapú visszajátszással**,
3. **Illesztés alapú** megfelelés ellenőrzés.



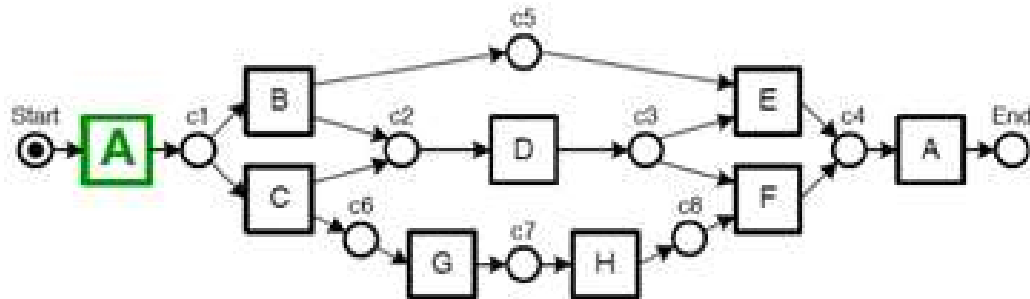
Konformancia token alapú egyezőség-vizsgálata

- Ehhez szükség van a **folymat modelljére**, pl. Petri háló formájában
- A **napló fájlok** rengeteg információt tartalmaznak, melyek sok esetben nem feltétlen szükségesek a vizsgálat elvégzéséhez. Ezért az adatoknak előzetes átalakításon kell átesniük, mely folyamán csak **azon esemény tulajdonságokat tartjuk meg, melyek számunkra szükségesek.**
- Bevezetünk négy különböző mérőszámot:
 1. ***p*** - létrehozott token (produced token),
 2. ***c*** - elhasznált token (consumed token),
 3. ***m*** - hiányzó token (missing token),
 4. ***r*** - megmaradt token (remaining token).

Konformancia token alapú egyezőség-vizsgálatának folyamatábrája

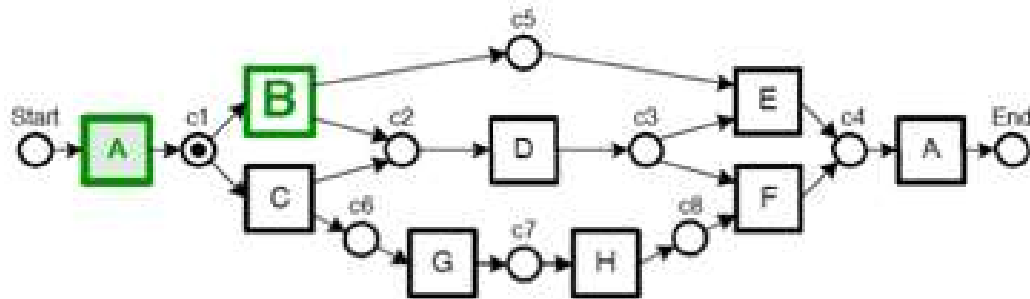


Fitnesz érték figyelése az események végig követésével egy trace-ben



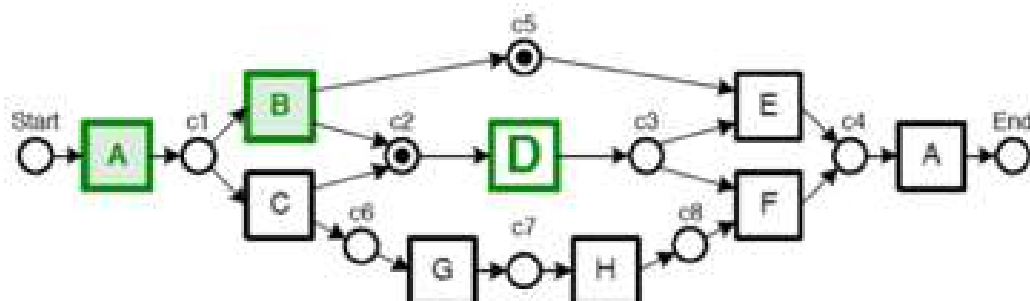
$m = 0$
 $r = 0$
 $c = 0$
 $p = 1$

No. of Instances	Log Traces
1207	→ A BDEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 1$
 $p = 2$

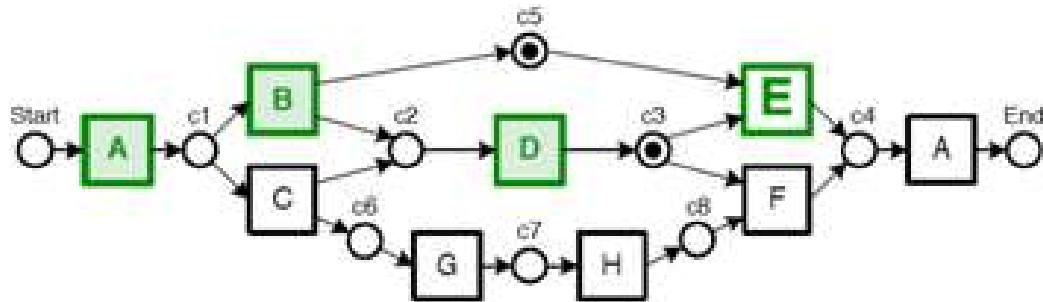
No. of Instances	Log Traces
1207	→ A B DEA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 2$
 $p = 4$

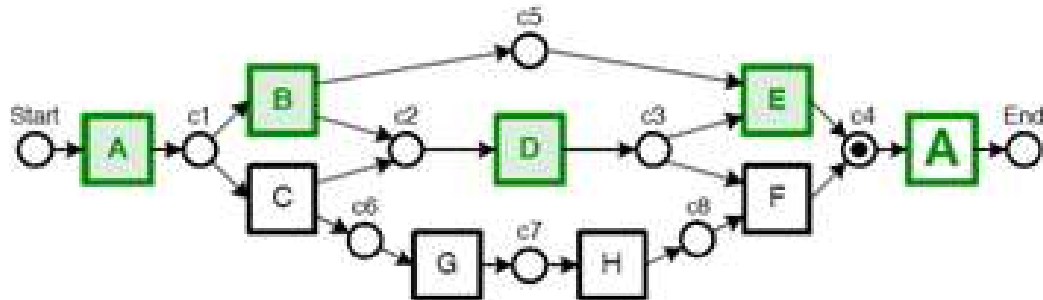
No. of Instances	Log Traces
1207	→ A B D EA
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Ha nincs probléma: $m = 0 \wedge r = 0$



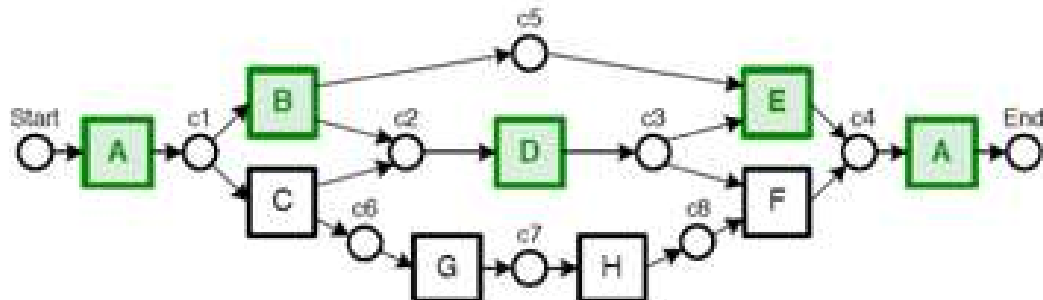
$m = 0$
 $r = 0$
 $c = 3$
 $p = 5$

No. of Instances	Log Traces
1207	→ ABDE A
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 5$
 $p = 6$

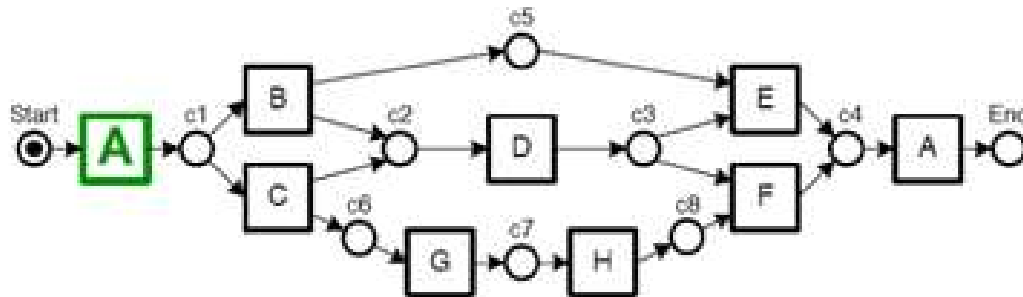
No. of Instances	Log Traces
1207	→ ABDE A
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 7$
 $p = 7$

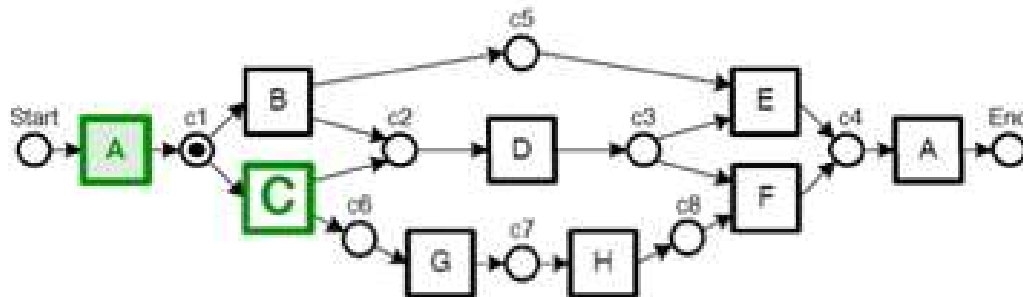
No. of Instances	Log Traces
1207	→ ABDE A
145	ACDGHFA
56	ACGDHFA
23	ACHDFA
28	ACDHFA

Ha probléma van: $m > 0$ v $r > 0$



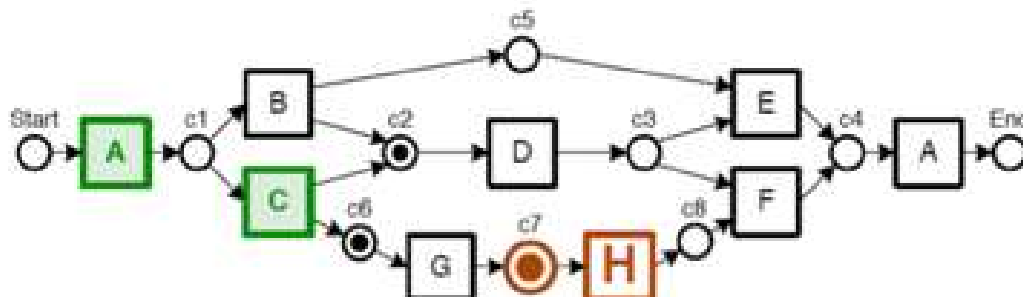
$m = 0$
 $r = 0$
 $c = 0$
 $p = 1$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 0$
 $r = 0$
 $c = 1$
 $p = 2$

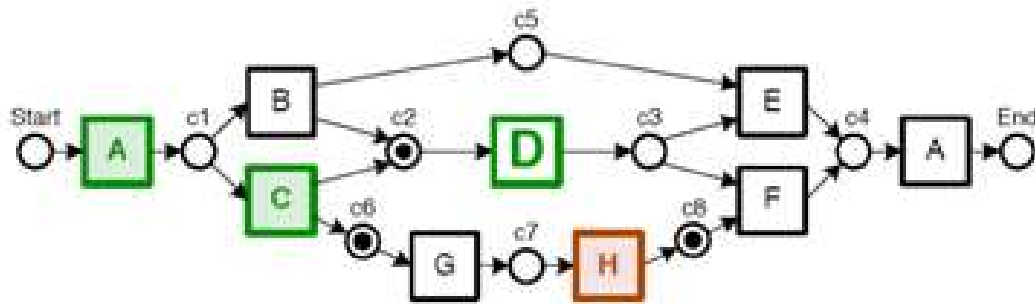
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 2$
 $p = 4$

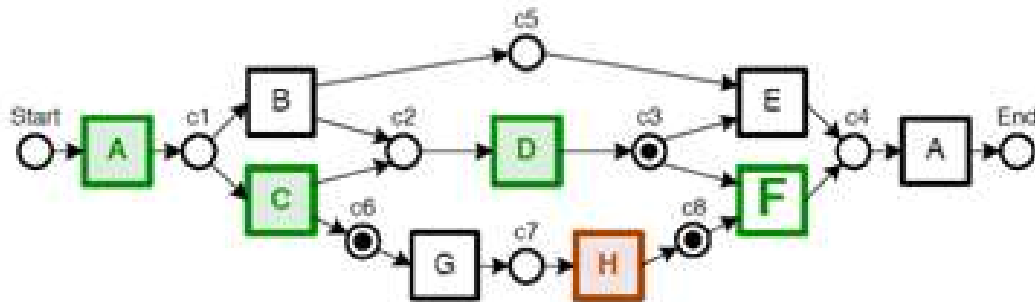
No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA

Folytatás



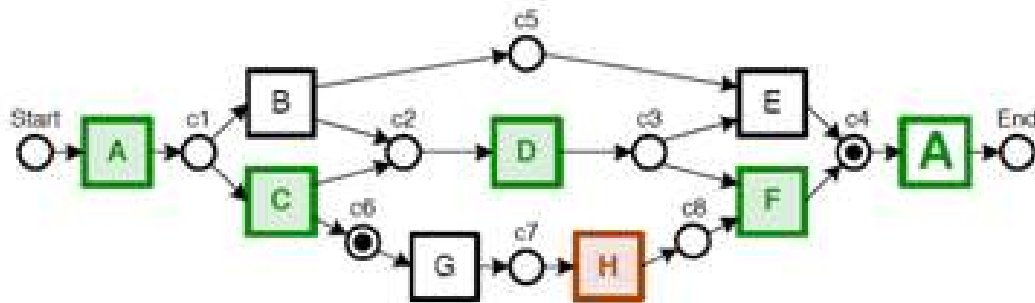
$m = 1$
 $r = 0$
 $c = 3$
 $p = 5$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 4$
 $p = 6$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA



$m = 1$
 $r = 0$
 $c = 6$
 $p = 7$

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA

Fitness érték számítása

Token alapú egyezőség-vizsgálat esetén

Egy trace (t) **fitness értéke:**

$$f(M, t) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right)$$

hiányzó token(ek) megmaradt token(ek)

elhasznált token(ek) létrehozott token(ek)

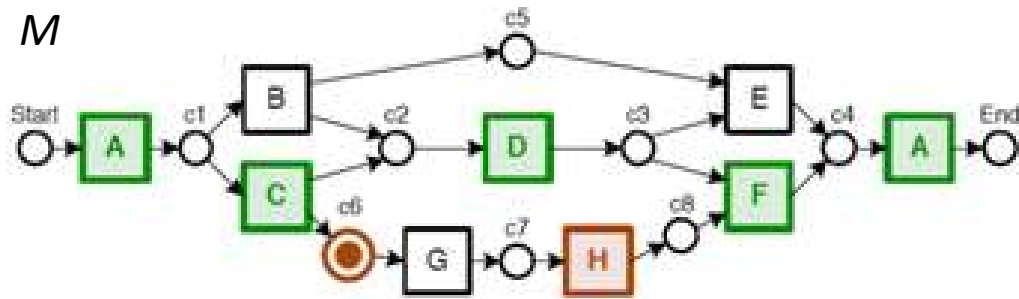
Egy log (L) **fitness értéke:**

$$f(M, L) = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i p_i} \right)$$

különböző tracek száma azonos tracek száma

Fitness érték számítása

Token alapú egyezőség-vizsgálat esetén



$m = 1$
 $r = 1$
 $c = 8$
 $p = 8$

L

No. of Instances	Log Traces
1207	ABDEA
145	ACDGHFA
56	ACGDHFA
23	→ ACHDFA
28	ACDHFA

A fenti trace **fitness értéke**:

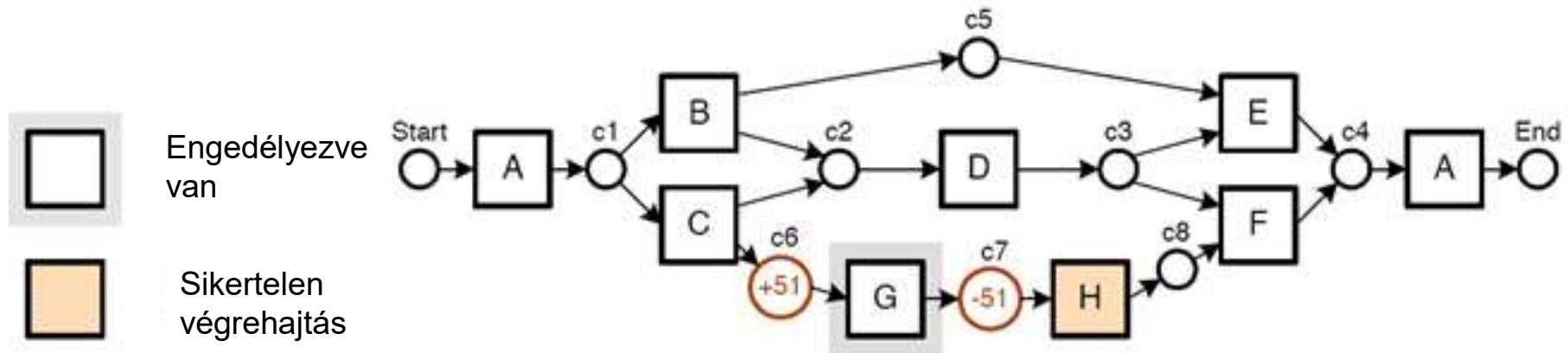
$$f(M, t) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right) = \frac{1}{2} \left(1 - \frac{1}{8} \right) + \frac{1}{2} \left(1 - \frac{1}{8} \right) = \frac{7}{8} = \mathbf{0,875}$$

A fenti **log fitness értéke**:

$$f(M, L) = \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i r_i}{\sum_{i=1}^k n_i p_i} \right) = \dots =$$

$$= \frac{1}{2} \left(1 - \frac{51}{10666} \right) + \frac{1}{2} \left(1 - \frac{51}{10666} \right) \approx \mathbf{0,995}$$

Hogyan használhatjuk a módszert diagnosztikához



A **H** művelet valamilyen oknál fogva nem került végrehajtásra. Ki kell deríteni, hogy mi volt az oka!



A probléma lehetséges forrásai:

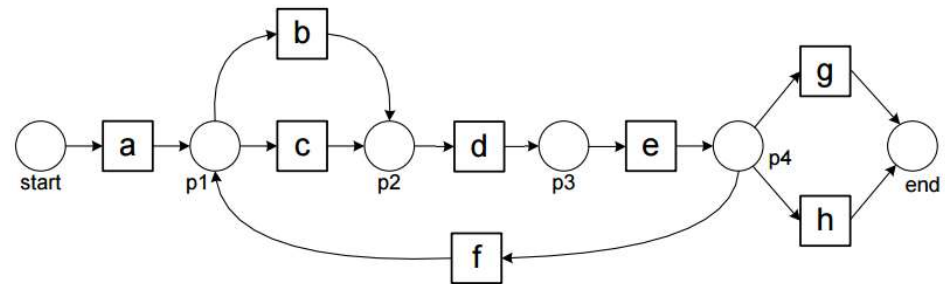
- **végrehajtási probléma** → **G** nem lett végrehajtva
- **rögzítési probléma** → **G** végre lett hajtva, csak nem került rögzítésre a logban
- **modellezési probléma** → **G** végrehajtása opcionális vagy (az adott modellben nem megjelenő) feltételekhez kötött

Másik példa

- A napló fájl átalakítása után csak azon **esemény tulajdonságokat** tartjuk meg melyek számunkra szükségesek.
- Ezután **előfordulási gyakoriság** szerint összerendezzük őket.
- A vizsgálat elvégzéséhez **szükséges adatok** maradnak meg: az *azonosító*, a *gyakoriság* és az *útvonal*

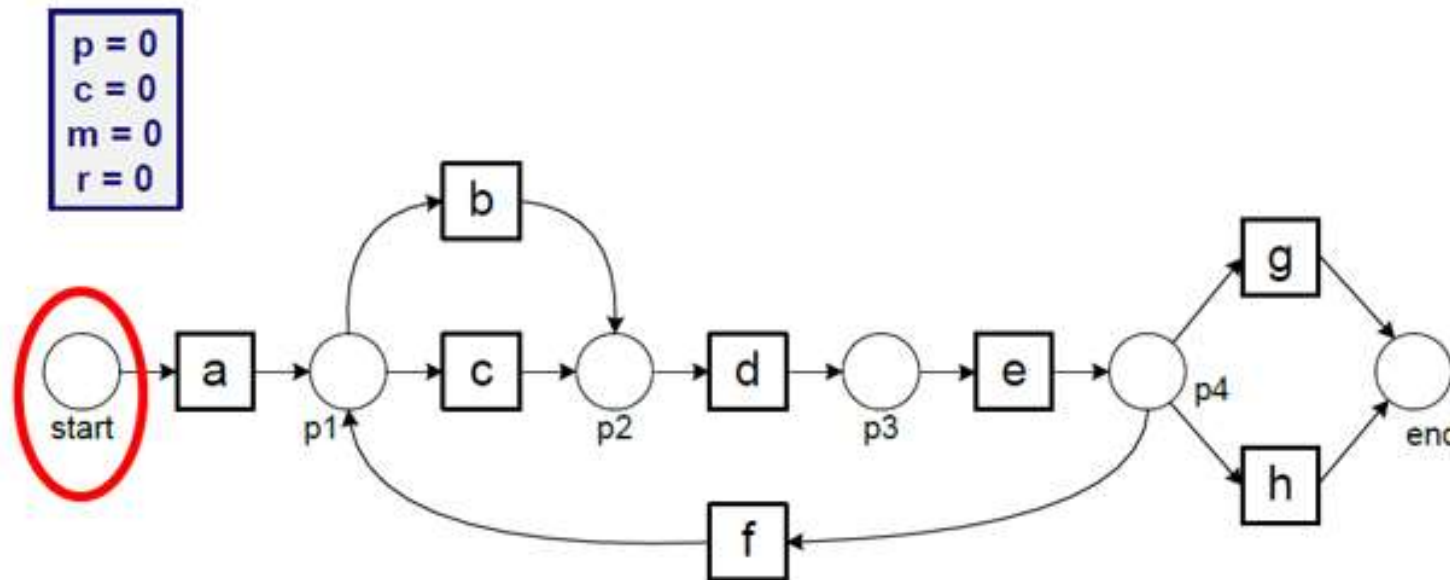
455	σ_1	$\langle a, c, d, e, h \rangle$
191	σ_2	$\langle a, b, d, e, g \rangle$
177	σ_3	$\langle a, d, c, e, h \rangle$
144	σ_4	$\langle a, b, d, e, h \rangle$
111	σ_5	$\langle a, c, d, e, g \rangle$
82	σ_6	$\langle a, d, c, e, g \rangle$
56	σ_7	$\langle a, d, b, e, h \rangle$
47	σ_8	$\langle a, c, d, e, f, d, b, e, h \rangle$
38	σ_9	$\langle a, d, b, e, g \rangle$
33	σ_{10}	$\langle a, c, d, e, f, b, d, e, h \rangle$
14	σ_{11}	$\langle a, c, d, e, f, b, d, e, g \rangle$
11	σ_{12}	$\langle a, c, d, e, f, d, b, e, g \rangle$
9	σ_{13}	$\langle a, d, c, e, f, c, d, e, h \rangle$

Elemezzük a megjelölt eseménysorozatot a rendelkezésre álló modell alapján!



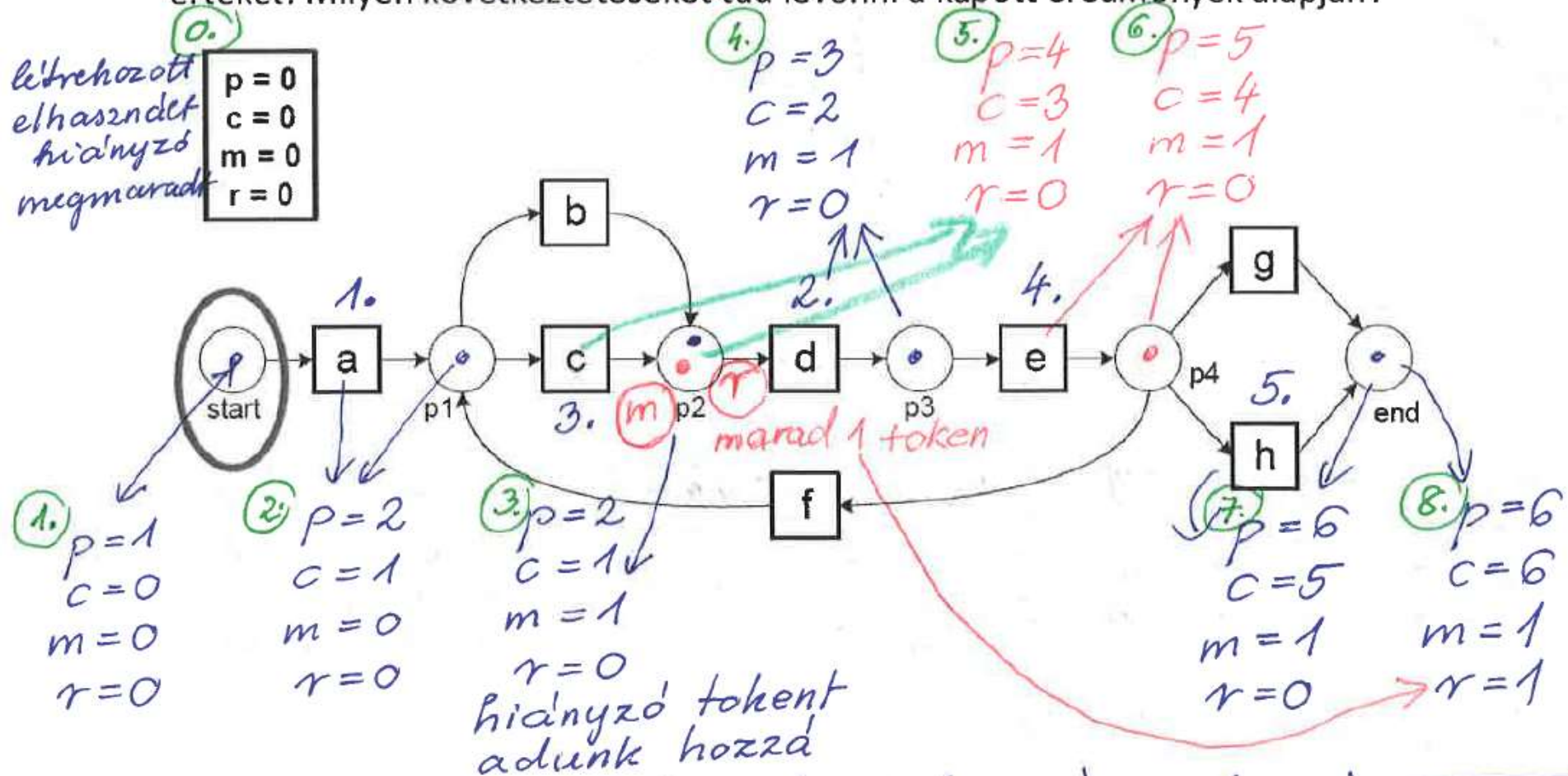
Számolja ki az alábbi trace és modell alapján a megfelelőséget meghatározó **fitness értéket** token alapú egyezőség-vizsgálat használatával! Milyen következtetéseket tud levonni a kapott eredmények alapján?

$$t = \langle a, d, c, e, h \rangle$$



$$f(M, t) = \frac{1}{2} \left(1 - \frac{m}{c} \right) - \frac{1}{2} \left(1 - \frac{r}{p} \right)$$

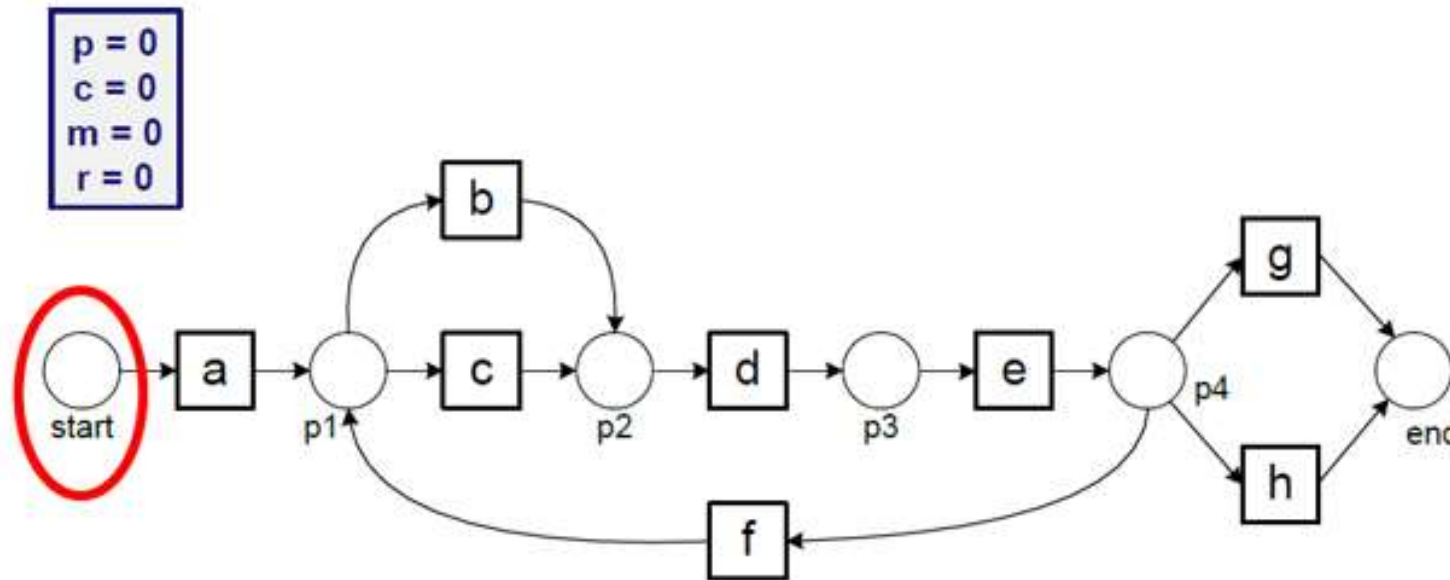
Számolja ki a $t = \langle a, d, c, e, h \rangle$ trace alapján a megfelelőséget meghatározó fitness értéket! Milyen következtetéseket tud levonni a kapott eredmények alapján?



$$f = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right) = \frac{1}{2} \left(1 - \frac{1}{6} \right) + \frac{1}{2} \left(1 - \frac{1}{6} \right) = \boxed{0,833}$$

Számolja ki az alábbi trace és modell alapján a megfelelőséget meghatározó **fitness értéket** token alapú egyezőség-vizsgálat használatával! Milyen következtetéseket tud levonni a kapott eredmények alapján?

$$t = \langle a, d, c, e, f, c, d, e, h \rangle$$



$$f(M, t) = \frac{1}{2} \left(1 - \frac{m}{c} \right) - \frac{1}{2} \left(1 - \frac{r}{p} \right)$$

t = <a, d, c, e, f, c, d, e, h>
1. 2. 3. 4. 5. 6. 7. 8. 9.

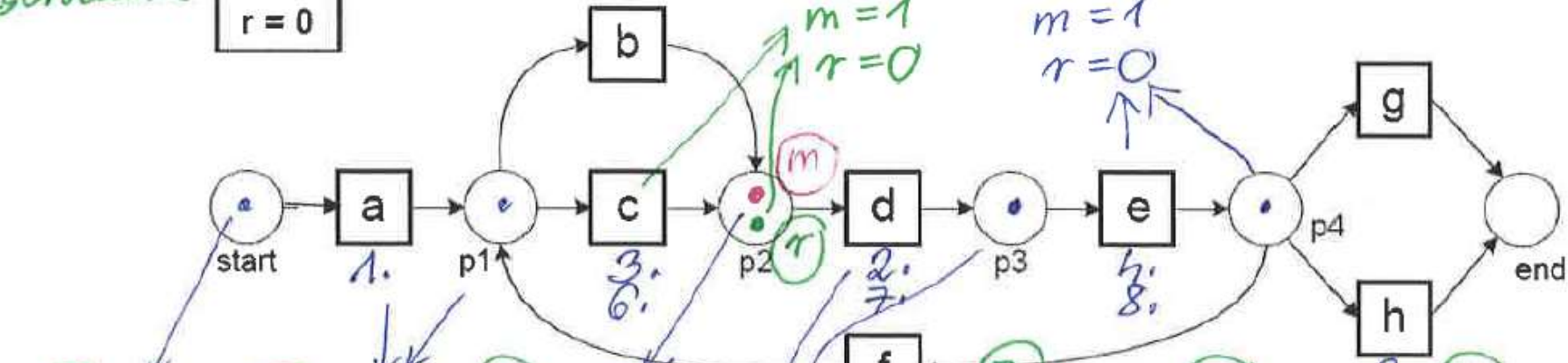
0.
lépés sorozatma

p = 0
c = 0
m = 0
r = 0

5. 3.
p = 4
c = 3
m = 1
r = 0

6. 4.
p = 5
c = 4
m = 1
r = 0

- művelet sorozatma



1. p = 1
c = 0
m = 0
r = 0

2. p = 2
c = 1
m = 0
r = 0

3. p = 2
c = 1
m = 1
r = 0

4. p = 3
c = 2
m = 1
r = 0

7. p = 6
c = 5
m = 1
r = 0

8. p = 7
c = 6
m = 1
r = 0

9. p = 8
c = 7
m = 1
r = 0

10. p = 9
c = 8
m = 1
r = 0

11. p = 10
c = 9
m = 1
r = 0

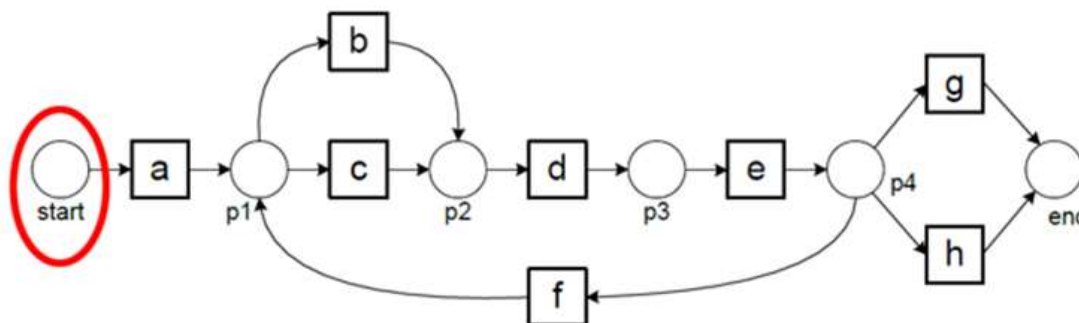
12. p = 10
c = 10
m = 1
r = 1

$$P = \frac{1}{2} \left(1 - \frac{m}{c}\right) + \frac{1}{2} \left(1 - \frac{r}{p}\right) = \frac{1}{2} \left(1 - \frac{1}{10}\right) + \frac{1}{2} \left(1 - \frac{1}{10}\right) = \underline{\underline{0.9}}$$

Számolja ki az alábbi log és modell alapján a megfelelőséget meghatározó **fitnessz értéket** token alapú egyezőség-vizsgálat használatával!

Milyen következtetéseket tud levonni a kapott eredmények alapján?

$$L = [\langle a, d, c, e, h \rangle^{177}, \langle a, d, c, e, f, c, d, e, h \rangle^9]$$



$$k = 2$$

$$t_1 = \langle a, d, c, e, h \rangle \quad n_1 = 177 \quad m_1 = 1 \quad c_1 = 6 \quad r_1 = 1 \quad p_1 = 6$$

$$t_2 = \langle a, d, c, e, f, c, d, e, h \rangle \quad n_2 = 9 \quad m_2 = 1 \quad c_2 = 10 \quad r_2 = 1 \quad p_2 = 10$$

$$\begin{aligned} f(M, L) &= \frac{1}{2} \left(1 - \frac{\sum_{i=1}^k n_i m_i}{\sum_{i=1}^k n_i c_i} \right) = \frac{1}{2} \left(1 - \frac{177 \cdot 1 + 9 \cdot 1}{177 \cdot 6 + 9 \cdot 10} \right) + \frac{1}{2} \left(1 - \frac{177 \cdot 1 + 9 \cdot 1}{177 \cdot 6 + 9 \cdot 10} \right) = \\ &= 1 - \frac{177 + 9}{177 \cdot 6 + 9 \cdot 10} = \mathbf{0,8385} \end{aligned}$$

Az elemzést elvégző alkalmazás felhasználói felülete, eredmények

The screenshot displays a software interface with several panels. At the top left, a 'File' panel shows the file path: 'D:\My Backups_20161219\va_tantargyak 2010-to\Process mining\konformancia szakdoga\Programkod\Teszt\Eseménynapló\teszt_események.bt'. Below this, a table lists event data:

455	q1	<a,c,d,e,h>
191	q2	<a,b,d,e,g>
177	q3	<a,d,c,e,h>
144	q4	<a,b,d,e,h>

To the right, a 'Modell file Name' panel shows the path: 'D:\My Backups_20161219\va_tantargyak 2010-to\Process mining\konformancia szakdoga\Programkod\Teszt\Modell\teszt_modell.pnml'. Below this, a list of places and transitions is shown:

- Place id: 1 Place name: start
- Place id: 2 Place name: p1
- Place id: 3 Place name: p2
- Place id: 4 Place name: p3
- Place id: 5 Place name: p4
- Place id: 6 Place name: end
- Transition id: 7 Transition name: a
- Transition id: 8 Transition name: b
- Transition id: 9 Transition name: c
- Transition id: 10 Transition name: d
- Transition id: 11 Transition name: e
- Transition id: 12 Transition name: g
- Transition id: 13 Transition name: h
- Transition id: 14 Transition name: f

On the far right, there is a blue button labeled 'Analysis'. Below the main panels, a scrollable area displays analysis results for two events:

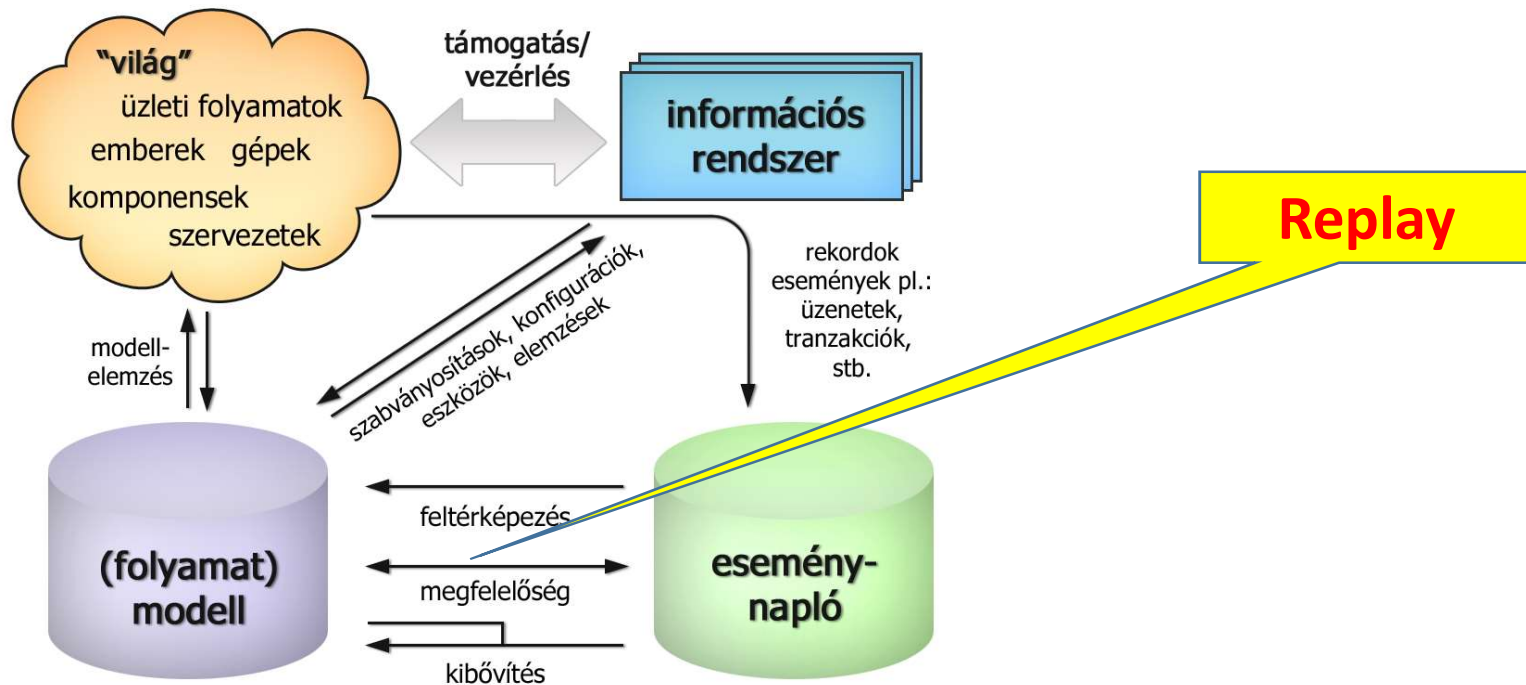
.....
Frekvency: 191
Event_name: q2
Produce: 6
Consumed: 6
Missing: 0
Remain: 0
Fitness: 1,000
.....
Frekvency: 177
Event_name: q3
Produce: 6
Consumed: 6
Missing: 1
Remain: 1
Missing_place_id: 3 Name: p2
Remain_place_id: 3 Name: p2
Fitness: 0,833
.....

Eredmények

- q1, q2, q4: Tökéletes illeszkedés.
 - Egyezőség mértéke: 1
 - Létrehozott token-ek száma: 6
 - Megszüntetett token-ek száma: 6
 - Hiányzó token-ek száma: 0
 - Megmaradt token-ek száma: 0
 - Egyéb hiba: Nincs
- q3: Hiányos illeszkedés.
 - Egyezőség mértéke: 0.833
 - Létrehozott token-ek száma: 6
 - Megszüntetett token-ek száma: 6
 - Hiányzó token-ek száma: 1
 - Megmaradt token-ek száma: 1
 - Egyéb hiba:
 - Hiányzó token hiba: Hely neve p2
 - Megmaradt token hiba: Hely neve p2
- A eseménynaplóban lévő összes eseményre kivetített egyezőség mértéke (teljes fitness): 0.97.

A konformancia ellenőrzés módszerei

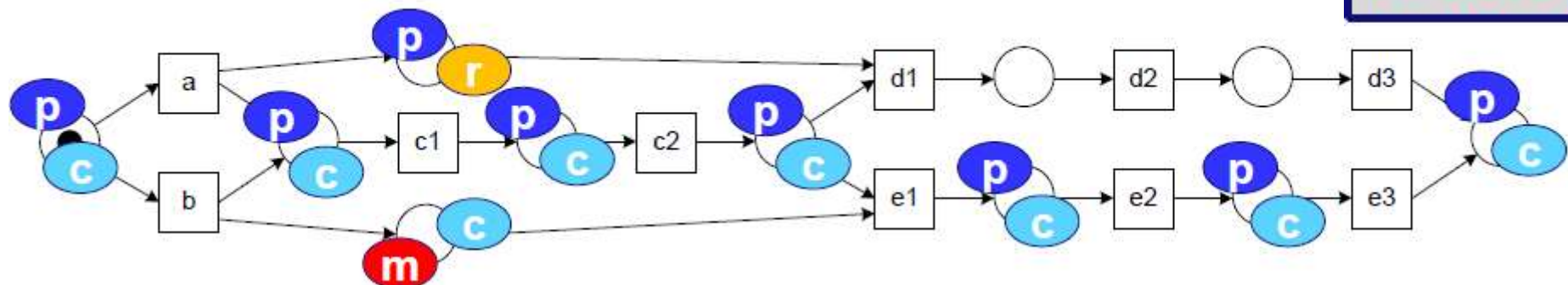
1. Konformancia ellenőrzés **ok-okozati lábnyomokkal**,
2. Konformancia ellenőrzés **token alapú visszajátszással**,
3. **Illesztés alapú** megfelelőség ellenőrzés. ←



A token alapú egyezőség módszer korlátai

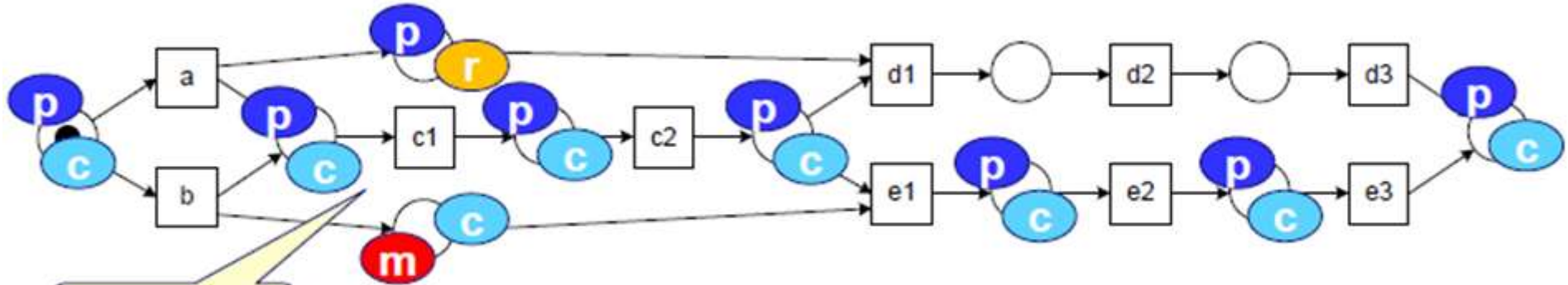
- A módszer feltételezi a láthatóságot és az egyediséget az átmenetek azonosítójának,
- A megfelelőségi értékek bizonyos esetekben túl optimisták,
- A helyi döntés nem mindig helyes,
- A replay technika a modellen keresztül nem feltétlenül ad helyes útvonalat, pedig ez lényeges a megfelelőség vizsgálathoz, a teljesítmény elemzéshez és más diagnosztikákhoz,
- **Szeretnénk látni a „legzártabb útvonalat”, pl. $\langle b, c1, c2, e1, e2, e3 \rangle$**

$\langle a, c1, c2, e1, e2, e3 \rangle$



Illesztés alapú megfelelés ellenőrzés

$\langle a, c1, c2, e1, e2, e3 \rangle$



???

Log mozgás:
csak a logban mozog

Szinkron mozgás:
mindkettőben mozog

A felső sor megfelel egy **trace**-nek a **logban**

Az alsó sor megfelel egy **útvonal**nak a **modellben**

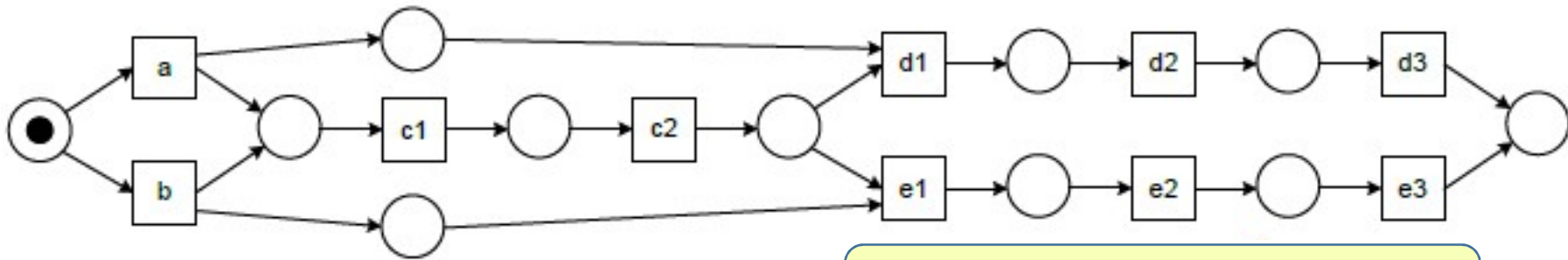
a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

Modell mozgás:
csak a modellben mozog

» - **nem mozog**

Trace a logban: $\langle a, c1, c2, e1, e2, e3 \rangle$

Összehasonlítjuk a **log**beli lefutást a **modell**beli lehetséges útvonalakkal.



Optimális, nincs másik illesztés, amelynek alacsonyabb a költsége

a	»	c1	c2	e1	e2	e3
»	b	c1	c2	e1	e2	e3

2

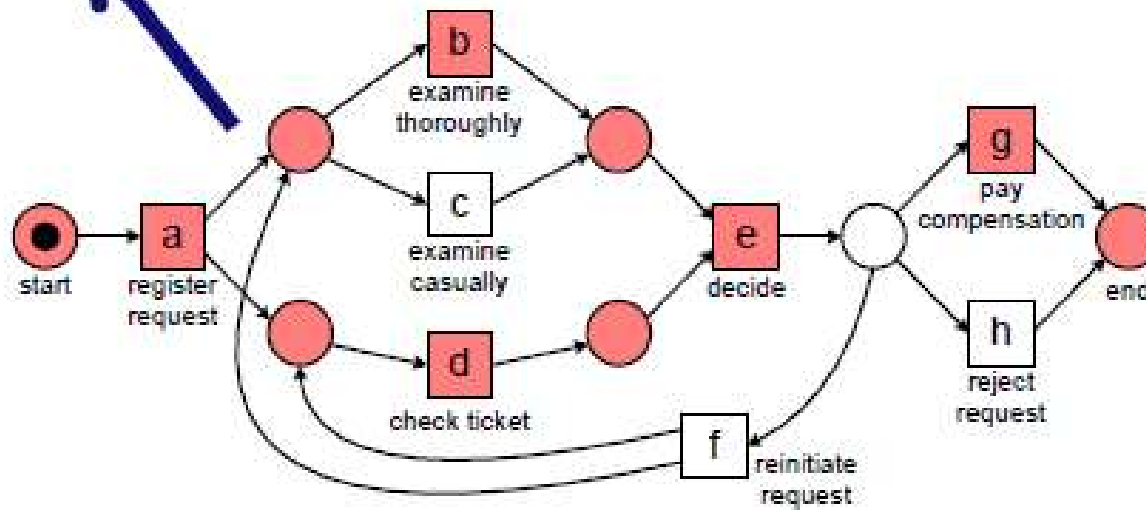
a	c1	c2	»	»	»	e1	e2	e3
a	c1	c2	d1	d2	d3	»	»	»

6

Másik példa: Illesztés vizsgálata



← **<a,b,d,e,g>**



#	trace
455	acdeh
191	abdeg
177	acdeh
144	abdeh
111	acdeg
82	adbeg
56	adbeg
47	acdefdbeg
38	adbeg
33	acdefdbeg
14	acdefdbeg
11	acdefdbeg
9	acdefdbeg
6	acdefdbeg
5	acdefdbeg
3	acdefdbefdbeg
2	acdefdbeg
2	acdefdbefdbeg
1	acdefdbefdbeg
1	adbefdbefdbeg
1	acdefdbefdbefdbeg
1391	

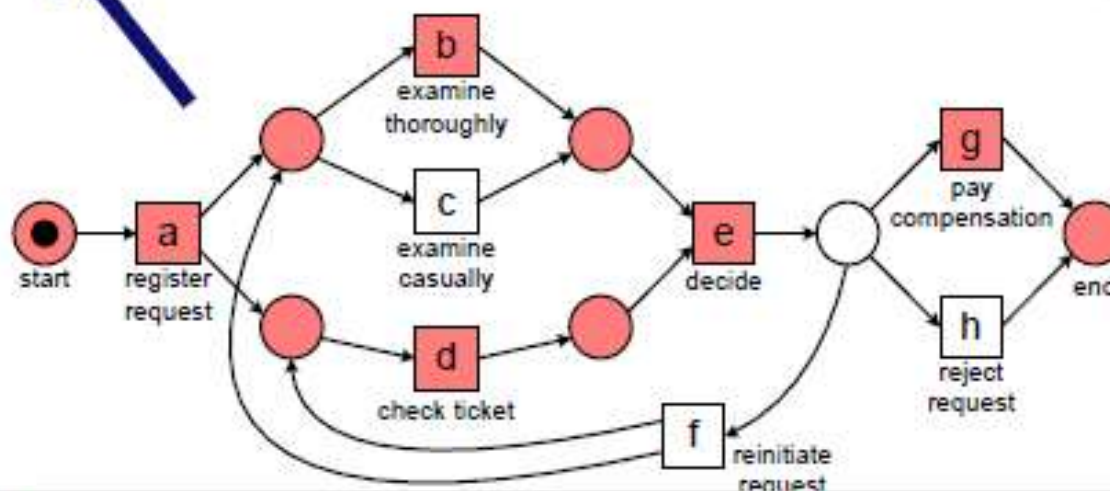
Illesztés vizsgálata

Problémák:

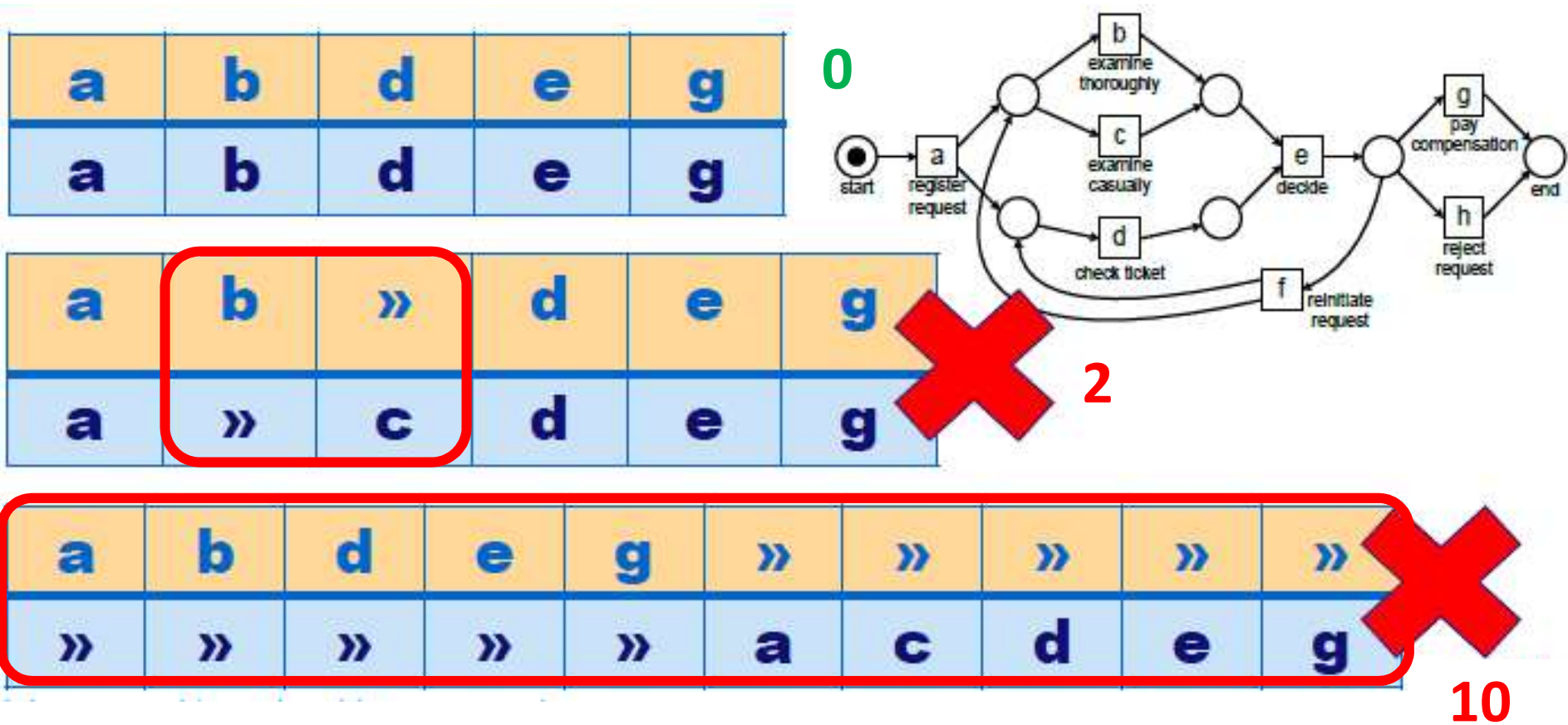
- **b** vagy **c** hiányzik a lefutás során
- **g** és **h** közül csak az egyik hajtódhat végre

a	»	d	e	g	h
a	b	d	e	g	»

← **⟨a,d,e,g,h⟩**



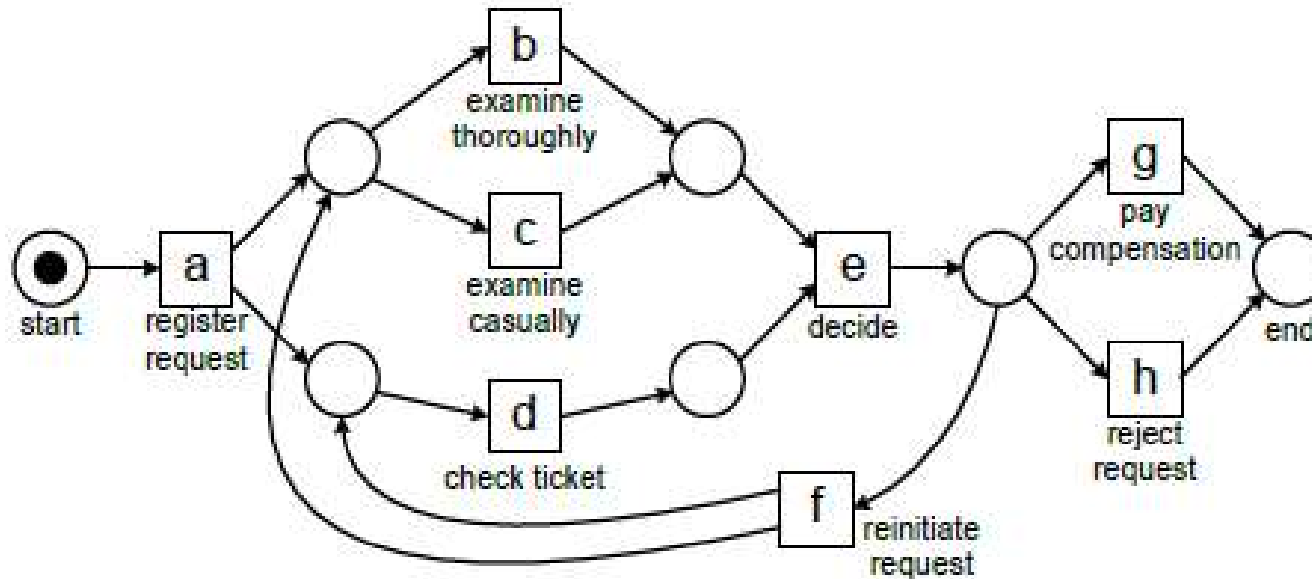
Optimális illesztés $\langle a, b, d, e, g \rangle$ -re



Egy másik optimális illesztés $\langle a, b, e, f, d, e, g \rangle$ -re

a	b	»	e	f	d	»	e	g	2
a	b	d	e	f	d	b	e	g	

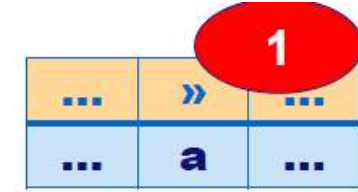
A ciklust tekintjük, így a **d** és a **b** esemény hiányzik a **logban**!



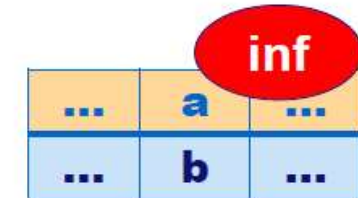
Az optimális illesztés **költségfüggvény** függő!

Az alap költségfüggvény a „>>” jel alapján számol

- A **log mozgás** és **modell mozgás** költsége **1**.



- A **szinkron mozgás** költsége **0** (ha a művelet megegyezik).



azt mondjuk, hogy a költség végtelen

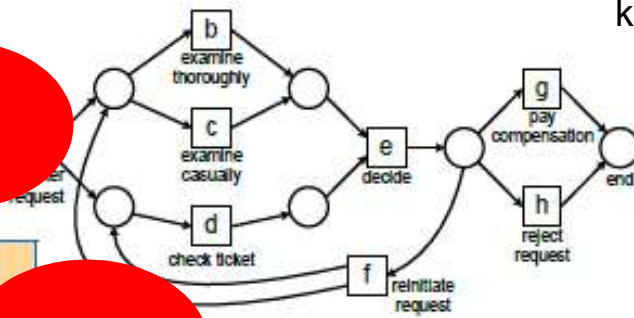
a	b	d	e	g
a	b	d	e	g



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



a	b	d	e	g	>>	>>	>>	>>	f
>>	>>	>>	>>	>>	a	c	d	e	g



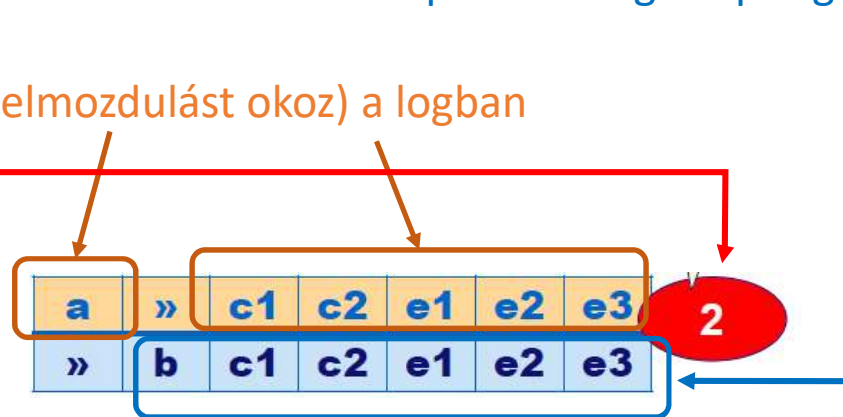
Fitness érték számítása

Illesztés alapú egyezőség-vizsgálat esetén

Egy trace (t) **fitness értéke:**

$$f(M, t) = 1 - \frac{K}{M + N}$$

- N : a legrövidebb útvonal eseményeinek száma a kezdő állapottól a vég állapotig a modellben
- M : az összes esemény száma (amely elmozdulást okoz) a logban
- K : az optimális illesztés költsége



Egy log (L) **fitness értéke:**

$$f(M, L) = 1 - \frac{\sum_{i=1}^k n_i K_i}{\sum_{i=1}^k n_i (M_i + N_i)}$$

azonos tracek száma

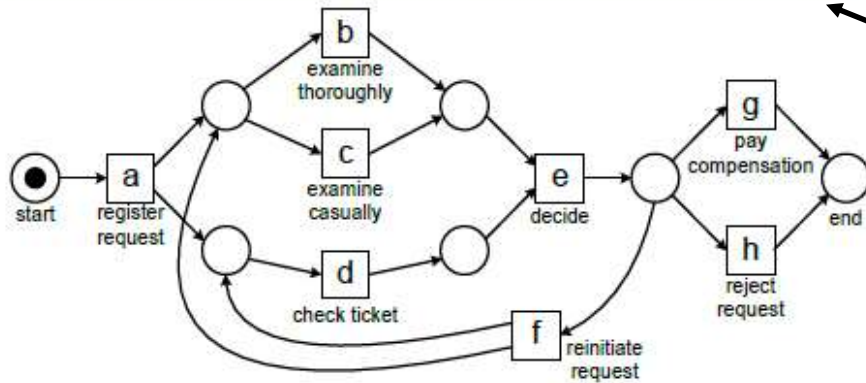
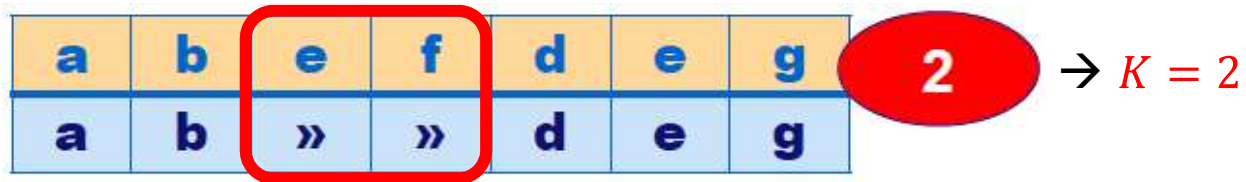
különböző tracek száma

Fitness érték számítása

Illesztés alapú egyezőség-vizsgálat esetén

Példa egy trace **fitness értékének** kiszámítására:

$$t = \langle a, b, e, f, d, e, g \rangle \rightarrow M = 7$$

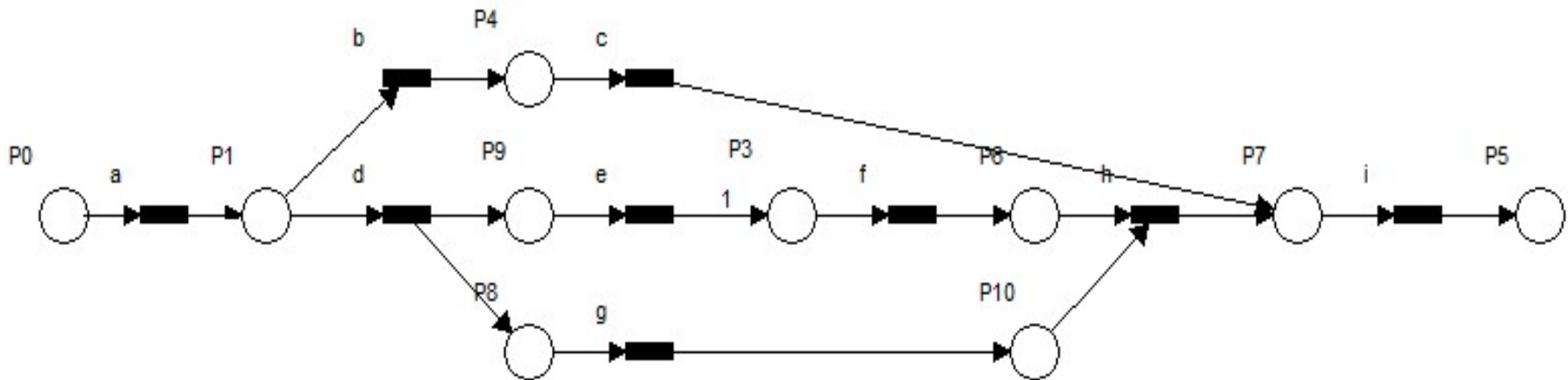


legrövidebb útvonal a modellben → N = 5

$$f(M, t) = 1 - \frac{K}{M + N} = 1 - \frac{2}{7 + 5} = 0,833$$

Számolja ki az **egyezőség mértékét** az illesztés alapú módszer segítségével az alábbi log és modell alapján az egyes tracekre és a teljes logra is!

$$L = [\langle a, b, c, h, i \rangle^5, \langle a, b, e, f, g, h, i \rangle^4, \langle a, d, e, g, f, h, i \rangle]$$

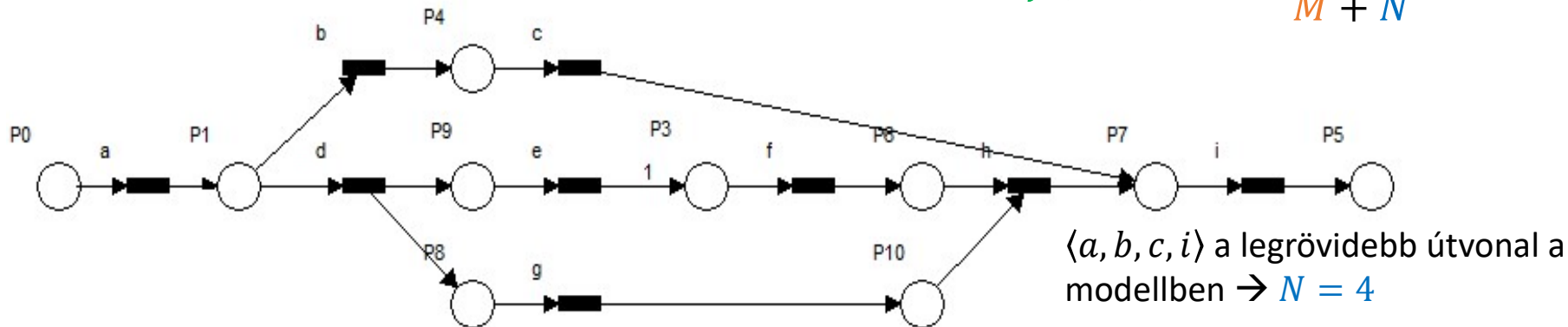


$$f(M, t) = 1 - \frac{K}{M + N}$$

$$f(M, L) = 1 - \frac{\sum_{i=1}^k n_i K_i}{\sum_{i=1}^k n_i (M_i + N_i)}$$

Számolja ki az **egyezőség mértékét** az illesztés alapú módszer segítségével az alábbi log és modell alapján!

$$L = [\langle a, b, c, h, i \rangle^5, \langle a, b, e, f, g, h, i \rangle^4, \langle a, d, e, g, f, h, i \rangle] \quad f(M, t) = 1 - \frac{K}{M + N}$$



t_1	a	b	c	h	i
M	a	b	c	>>	i

$$K = 1 \quad M = 5$$

$$f(M, t_1) = 1 - \frac{1}{5 + 4} = 1 - \frac{1}{9} = \frac{8}{9}$$

t_2	a	b	>>	e	f	g	h	i
M	a	>>	d	e	f	g	h	i

$$K = 2 \quad M = 7$$

$$f(M, t_2) = 1 - \frac{2}{7 + 4} = 1 - \frac{2}{11} = \frac{9}{11}$$

t_3	a	d	e	g	f	h	i
M	a	d	e	g	f	h	i

$$K = 0 \quad M = 7$$

$$f(M, t_3) = 1 - \frac{0}{7 + 4} = 1 - \frac{0}{11} = 1$$

$$f(M, L) = 1 - \frac{\sum_{i=1}^k n_i K_i}{\sum_{i=1}^k n_i (M_i + N_i)} = 1 - \frac{5 * 1 + 4 * 2 + 1 * 0}{5 * 9 + 4 * 11 + 1 * 11} = 1 - \frac{13}{100} = \mathbf{0,87}$$

Összehasonlíthatjuk az eseményeket nagyon finom szinten, ha a műveleteken kívül más adatok is rendelkezésre állnak az eseményhez kapcsolódóan

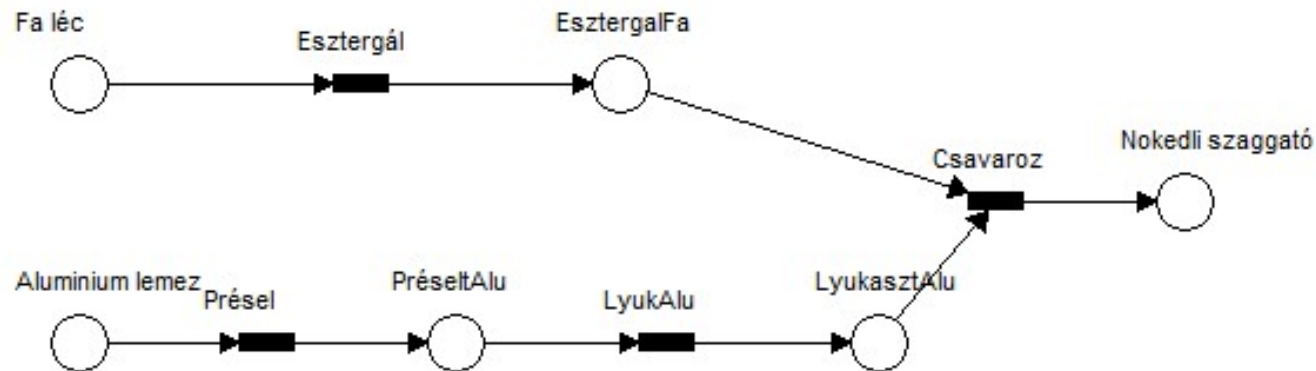
- Bármilyen költséget számolhatunk
- Számolhatunk kockázatot

...	Levél küldés (János, március. 8., 50e Ft)	...
...	Email küldés (Patrícia, március. 22., 20e Ft)	...

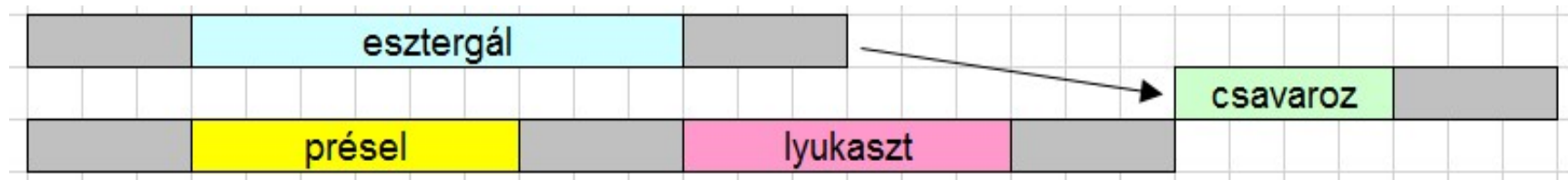
- Pl. megállapíthatjuk, hogy rossz személy hajtotta végre a tevékenységet, nem emailt küldött, hanem postai levelet, rossz összeget közölt az ügyféllel
- stb.

A konformancia elemzés alternatív megközelítései

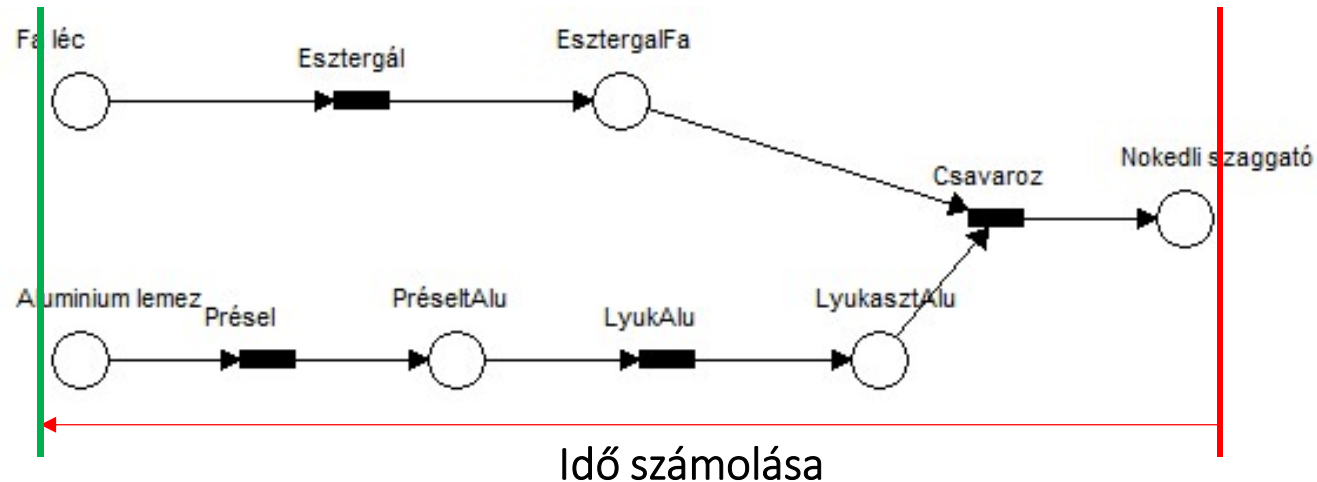
- Tekintsük az alábbi **gyártási folyamatot**:



- A folyamat lényege, hogy mindkét részfolyamatnak el kell készülnie mielőtt összecsavarozásra kerülne a termék.
- Másfajta szemléltetése a folyamatnak: A **Gantt-diagramon** a szürke részek az egyes műveletek közötti szállítást jelölik, ez a Petri-hálóba nem jelenik meg.



Petri háló kiértékelése

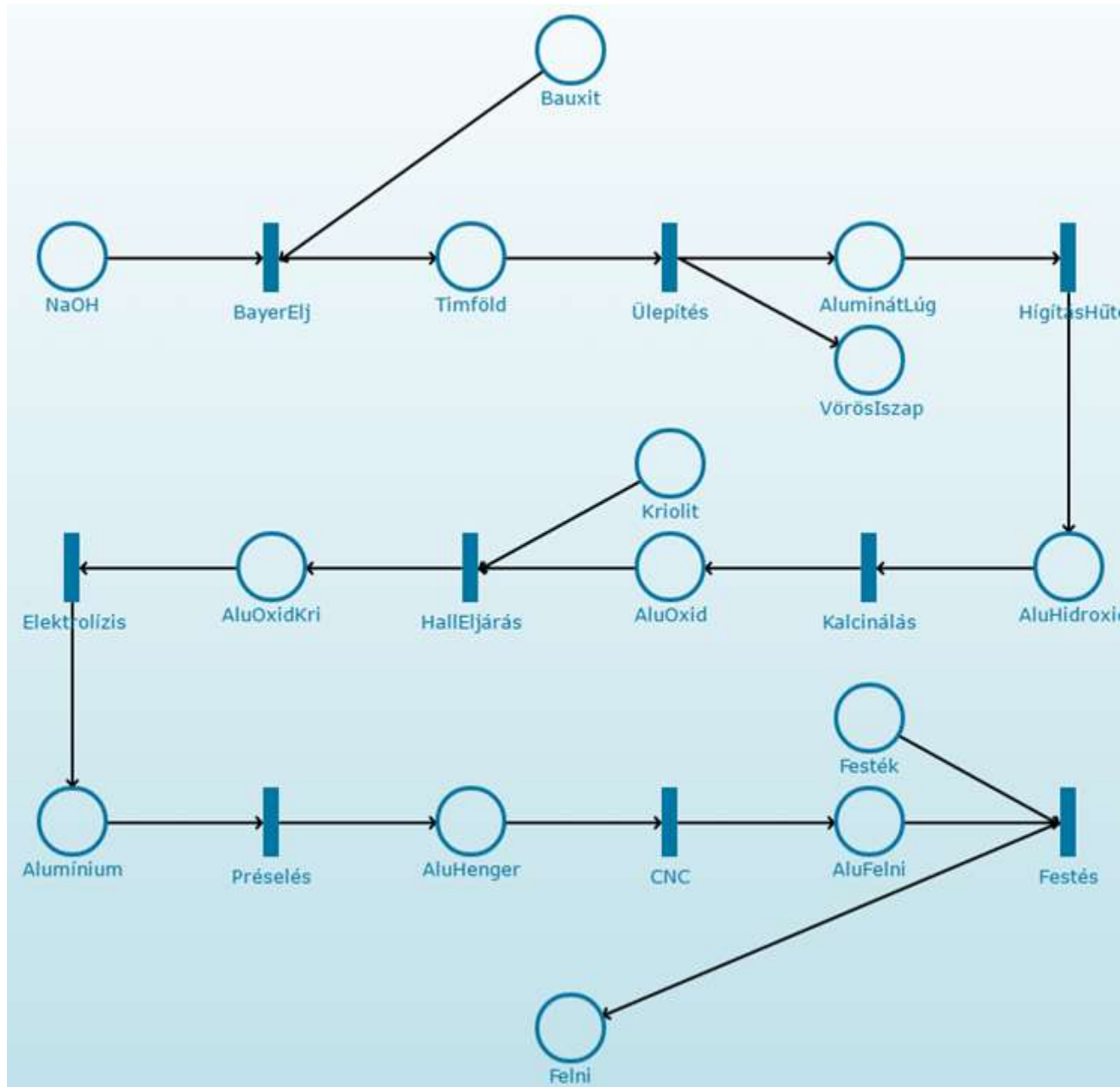


- Az algoritmus a készterméktől indulva számolja ki, hogy az egyes félkész munkadaraboknak a kezdéstől számítva **mennyi idő alatt** kell készen lenniük.
- A célállapotból kiindulva **rekurzívan** kiszámoljuk az addig eltelt időt.
- Amennyiben egy átmenetbe több helyről jön él (pl.: csavaroz), akkor a bejövő élekből a **legtöbb időt igénylőt választjuk**, mert nem lehet elindítani a műveletet, amíg a legtöbb időt igénylő előfeltétel nem készült el.

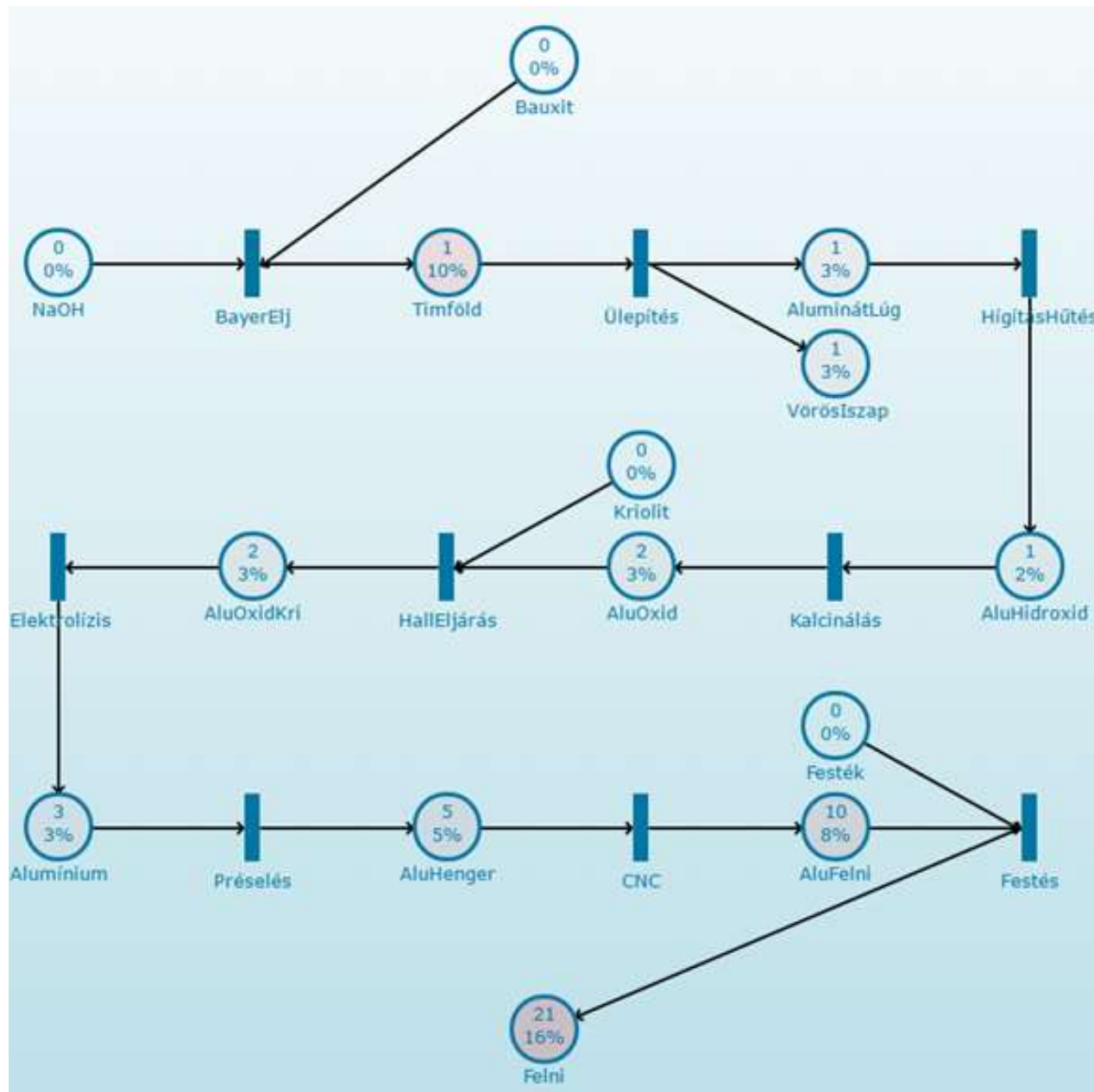
Eltérések keresése

- **Időbeli eltéréseket** keresünk jelen állapotban (vizsgálható lehet még a költség vagy az erőforrások).
- Az eltérések kereséséhez egy olyan **log fájlt** kell feldolgozni, amely **idő adatokat** is tartalmaz. A fájl az egyes **műveletek kezdeti és befejezési idejét** tartalmazza.
- Ezeket az időket megfeleltetjük a **hálóban** az **átmenet előtti és az azt követő hely idejével**, majd **letranszformáljuk** úgy, hogy a folyamat nullával kezdődjön.
- **Például**, ha a log fájlban az esztergálás 2021-01-13 11:39:32-től 2021-01-13 11:39:56-ig (**24 másodpercig**) tart, akkor a transzformálás után a *Fa* hely 0 időértéket, az *EsztergaltFa* hely pedig **24 értéket** kap.
- Ha ezt az egész **háló minden helyére** elvégeztük, akkor hasonló adatszerkezetet kapunk mint a Petri-háló feldolgozásakor kaptunk. Ezen két adatszerkezetet összehasonlítva megkapjuk, hogy **a folyamat egyes állomásai időben hogy viszonyulnak a modellhez**.
- A rendszer ezeket **az eltéréseket színekkel segítségével jelölheti a Petri-hálón**, ezzel vizualizálva, hogy hol vannak komolyabb időbeli eltérések, illetve hol sikerült lefaragni a modell szerinti időből.

Alumínium felni készítés modellje



Az elemzés eredménye



- Az egyes helyeken megjelennek az **időbeli eltérések értékei és százalékos arányuk**.
- Bizonyos erősségű **vörös háttérszínnel** vannak a **helyek** a Petri-hálóban jelölve.
- **PI.** meg tudjuk nézni, hogy az adott műveletet **ki végezte** abban az időben és meg lehet vizsgálni, hogy **mi a lassulás oka**.

Többperspektívás és online megfelelés ellenőrzés

- A megfelelés ellenőrzési módszerek túlnyomó többsége **offline** és csak a **control-flow perspektívára** (azaz csak az események sorrendjére) fókuszál.
- Az **online megfelelés ellenőrzés** alkalmazásával az eltéréseket hamarabb lehet detektálni, így a helyreállítási intézkedéseket akár a folyamat lefutásának befejeződése előtt végre lehet hajtani.

Bemenetei: **eseményfolyam** és folyamat modell

- Mivel nem csak az események sorrendjében lehet eltérés, ezért szükség van **többperspektívás megfelelés ellenőrzési** megoldásokra is, amelyek más perspektívákat (pl.: adat vagy erőforrás) is figyelembe vesz.

Bemenetei: log és **többperspektívát figyelembe vevő folyamat modell** (pl. adat Petri-háló)

- Még nincs kiforrott egyszerre **többperspektívás** és **online megfelelés ellenőrzési** algoritmus

Bemenetei: **eseményfolyam** és **többperspektívát figyelembe vevő folyamat modell**

Eseményfolyam

- csak **control-flow** perspektíva

eset azonosító × tevékenység azonosító

..., (1352, *a*), (1351, *b*), (1351, *c*), ...



- **control-flow** perspektíva + más perspektívák

eset azonosító × tevékenység azonosító × az esemény attribútumainak értékei

..., (1352, *a*, (Bas, 20/10/26 10:45)), (1351, *b*, (Daniel, 20/10/26 11:11)), ...

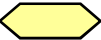
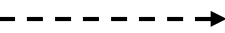






Naplófájlba rögzíthető:

Eset	Tevékenység	Erőforrás	Időbélyeg
...
13152	create account (<i>a</i>)	Bas	20/10/26 10:45
13151	submit order (<i>b</i>)	Daniel	20/10/26 11:11
13151	request quote (<i>c</i>)	Daniel	20/10/26 11:39
...

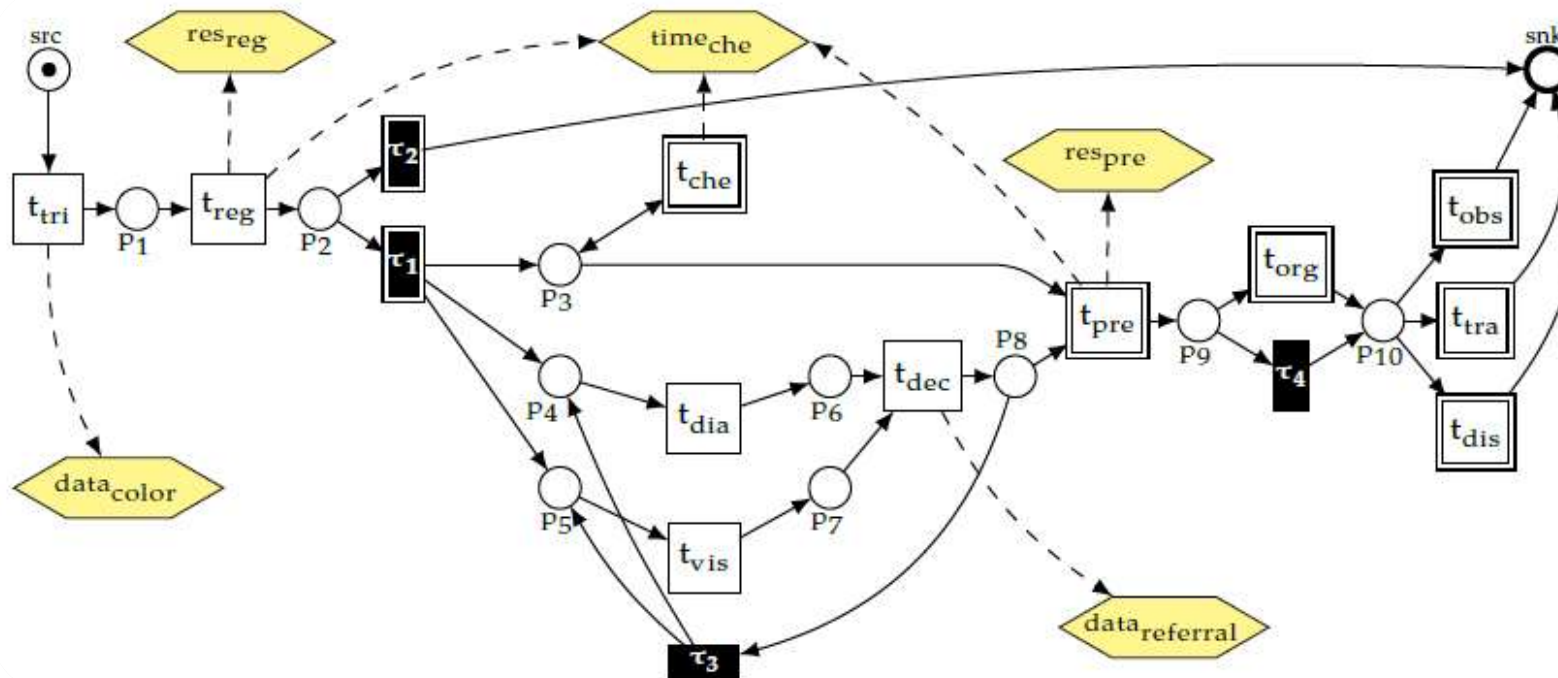
Adat Petri-háló (Data Petri Net, DPN)

Fontos fogalmak:

- változó (v), prime változó (v') 
- írási művelet 
- átmenet adatfüggő őrrrel   
- láthatatlan átmenet 

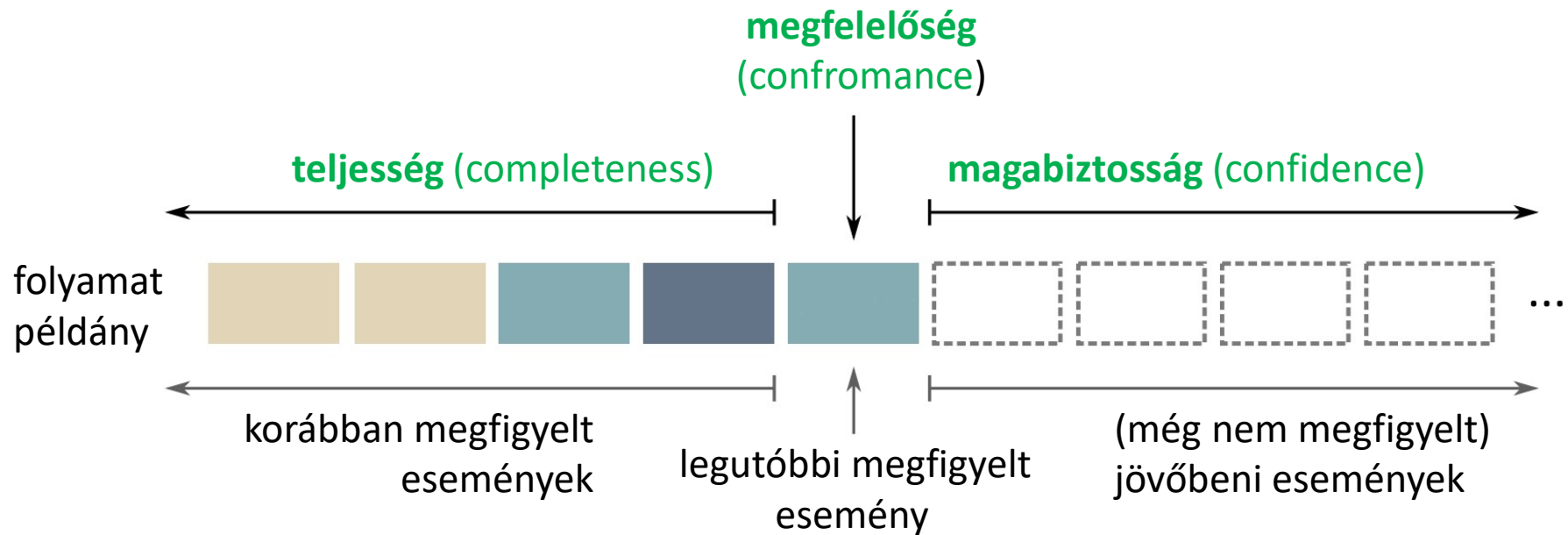
Transition	Guard expression
τ_1	$data_{color} \neq White$
τ_2	$data_{color} = White$
t_{che}	$time'_{che} \leq (time_{che} + 1h)$
t_{pre}	$time'_{che} \leq (time_{che} + 1h) \wedge res_{reg} = res'_{pre}$
t_{org}	$data_{referral} = Tertiary$
t_{obs}	$data_{referral} \neq Home$
t_{tra}	$data_{referral} \neq Home$
t_{dis}	$data_{referral} = Home$

örkifejezések



Online megfelelés ellenőrzés

Fő kihívás: a lefutás a vizsgálatkor még nem befejezett, az eleje és/vagy a vége hiányozhat



3 megfeleléségi mutató:

- **megfelelés:** a megfigyelt jó viselkedések aránya az összes viselkedéshez képest
- **teljesség:** azt jelzi, hogy az adott folyamat végrehajtás a kezdettől lett-e megfigyelve
- **magabiztosság:** azt jelzi, hogy az adott folyamat végrehajtás milyen közel van a befejezéshez

Többperspektívás online megfelelés ellenőrzés

Többperspektívás illesztés alapú módszert használ:

- Lehetséges mozgások:
 - **helyes szinkron mozgás** → a modell változó értékei megegyeznek az esemény attribútum értékeivel
 - **helytelen szinkron mozgás** → a modell változó értékei nem egyeznek meg az esemény attribútum értékeivel
 - **log mozgás**
 - **modell mozgás**
- Két költségfüggvényt használ:
 - tevékenység-illesztés költségfüggvénye
 - változó-illesztés költségfüggvénye → az egyes változó/attribútum értékeknek különböző súlyt lehet adni (pl. az súlyosabb probléma lehet, ha nem megfelelő személy végezte el a tevékenységet, mint az, hogy az esemény hosszabb ideig tartott az elvártnál)

Kihívások:

- a lefutások a vizsgálatkor még nem befejezettek
- az többperspektívás illesztés alapú módszer használta számítás igényes lehet → cache használata

Többperspektívás online megfelelés ellenőrzés

Gyártási folyamat példa

Több lefutásra is a következő kimenetet kapjuk:

	művelet	t_{s10}	t_{s20}	...	t_{s70}	t_{s80}
log rész	attribútumok	-	-	...	-	qib = 1 tray_row = 0 tray_column = 0
	átmenet	t_{s10}	t_{s20}	...	t_{s70}	t_{s80}
folyamat-modell rész	írt változók	-	-	...	-	qib = 1 tray_row = 1 tray_column = 1



Jelentése: A késztermékek egy része nem kerül tálcára a 8. állomás végén.

Kérdés: A késztermékek eldobásra kerülnek? Vagy csak hibásan érzékeli és/vagy rögzíti a rendszer? Ez helyszíni vizsgálatot igényel!

Megfelelőség ellenőrzés a ProM-ben

Online megfelelés ellenőrzés

- Nincs elérhető plugin!
(A megvalósított verzió pm4py folyamatbányászati platformban érhető el.)

Több perspektívás megfelelés ellenőrzés

- Kiegyensúlyozott több perspektívás megfelelés ellenőrzési módszer
(Balanced Multi-perspective Conformance Checking, rövidítve: BMCC)
- A BMCC megvalósított verziója “**Conformance checking of DPN**” néven érhető el ProM 6.9-es pluginként, a “**DataAwareReplayer**” nevű csomag részeként.

ProM 5.2 – néhány plugin itt működik igazán jól
Betöltjük azt a fájlt, amelyet vizsgálni szeretnénk (.mxml)

The screenshot displays the ProM 5.2 software interface. The window title is "ProM [5.2]". The menu bar includes "File", "Mining", "Analysis", "Conversion", "Exports", "Window", and "Help". The toolbar contains icons for file operations, search, and analysis. The main window displays the file "log_garazskapu_hibamentes_es_hibas.mxml (2)".

The dashboard shows the following data:

Category	Value
Processes	1
Cases	3
Events	26
Event classes	12
Event types	2
Originators	2

The "Events per case" chart shows a distribution with a minimum of 8, a mean of 8, and a maximum of 9. The "Event classes per case" chart shows a similar distribution with a minimum of 8, a mean of 8, and a maximum of 9.

The "Log info" section provides the following details:

- Source: MXML generator
- Source program: MXML generator
- Start date: 2010-02-10 13:46:25
- End date: 2010-02-10 15:50:35
- Description: Created by manual, using an existing Petri net

A button labeled "start analyzing this log" is located at the bottom right of the dashboard.

Használhatjuk az **Analysis** menü **Trace Comparison** lehetőségét

Össze tudjuk hasonlítani a traceket műveleti szinten.

A különbségek **pirossal** kiemelésre kerülnek.

ProM [5.2]

File Mining Analysis Conversion Exports Window Help

Analysis - Trace Comparison

Trace Diff Analysis

1 9 events	2 9 events
auto_beall #1 normal @Auto 10.02.2010 13:46:25.000	auto_beall #1 normal @Auto 10.02.2010 14:46:02.000
automata_gombnyomasra_var #2 normal @Automata 10.02.2010 13:47:00.000	automata_gombnyomasra_var #2 normal @Automata 10.02.2010 14:47:00.000
gombnyomas #3 normal @Auto 10.02.2010 13:47:10.000	gombnyomas #3 normal @Auto 10.02.2010 14:47:10.000
jegykiadas #4 normal @Automata 10.02.2010 13:47:15.000	jegykiadas_rosszul #4 hibas @Automata 10.02.2010 14:47:15.000
jegyelvetel #5 normal @Auto 10.02.2010 13:47:20.000	jegyelvetel #5 normal @Auto 10.02.2010 14:47:20.000
sorompo_fel #6 normal @Automata 10.02.2010 13:47:30.000	sorompo_fel #6 normal @Automata 10.02.2010 14:47:30.000
behajtas #7 normal @Auto 10.02.2010 13:47:55.000	behajtas #7 normal @Auto 10.02.2010 14:47:55.000
sorompo_le #8 normal @Automata 10.02.2010 13:48:25.000	sorompo_le #8 normal @Automata 10.02.2010 14:48:25.000
parkolas #9 normal @Auto 10.02.2010 13:50:35.000	parkolas #9 normal @Auto 10.02.2010 14:50:35.000

Jump to Difference: Previous Next Go Back

Egy másik össze-
hasonlítás:
A hiányzó
műveletek **zölddel**
vannak kiemelve.

ProM [5.2]

File Mining Analysis Conversion Exports Window Help

Analysis - Trace Comparison (2)

Trace Diff Analysis

1 9 events		3 8 events
auto_beall #1 normal @Auto 10.02.2010 13:46:25.000		auto_beall #1 normal @Auto 10.02.2010 15:46:02.000
automata_gombnyomasra_var #2 normal @Automata 10.02.2010 13:47:00.000		automata_gombnyomasra_var #2 normal @Auto 10.02.2010 15:47:10.000
gombnyomas #3 normal @Auto 10.02.2010 13:47:10.000	↔	jegy_mar_kiadva #3 hibas @Automata 10.02.2010 15:47:15.000
jegykiadas #4 normal @Automata 10.02.2010 13:47:15.000	↔	jegyvetel_rossz #4 hibas @Auto 10.02.2010 15:47:20.000
jegyvetel #5 normal @Auto 10.02.2010 13:47:20.000	←	
sorompo_fel #6 normal @Automata 10.02.2010 13:47:30.000		sorompo_fel #5 normal @Automata 10.02.2010 15:47:30.000
behajtas #7 normal @Auto 10.02.2010 13:47:55.000		behajtas #6 normal @Auto 10.02.2010 15:47:55.000
sorompo_le #8 normal @Automata 10.02.2010 13:48:25.000		sorompo_le #7 normal @Automata 10.02.2010 15:48:25.000
parkolas #9 normal @Auto 10.02.2010 13:50:35.000		parkolas #8 normal @Auto 10.02.2010 15:50:35.000

Jump to Difference: [Previous](#) [Next](#) [Go Back](#)

Ahhoz, hogy használhassuk a **Footprint Similarity** plugint, először be töltjük azt a két .mxml állományt, amelyeket össze akarunk hasonlítani.

A betöltés után számos információt azonnal megtudhatunk a log tartalmáról.

The screenshot shows the ProM 5.2 software interface with two windows open. The top window, titled 'log_garazskapu_hibamentes.mxml', displays a dashboard with the following key data:

Category	Value
Processes	1
Cases	1
Events	9

The 'Events per case' chart shows a single bar with a value of 9. Summary statistics are: Min 9, Mean 9, Max 9.

The bottom window, titled 'log_garazskapu_hibamentes_es_hibas.mxml (2)', displays a dashboard with the following key data:

Category	Value
Processes	1
Cases	3
Events	26
Event classes	12
Event types	2
Originators	2

The 'Events per case' chart shows a distribution with a mean of 8. Summary statistics are: Min 8, Mean 8, Max 9.

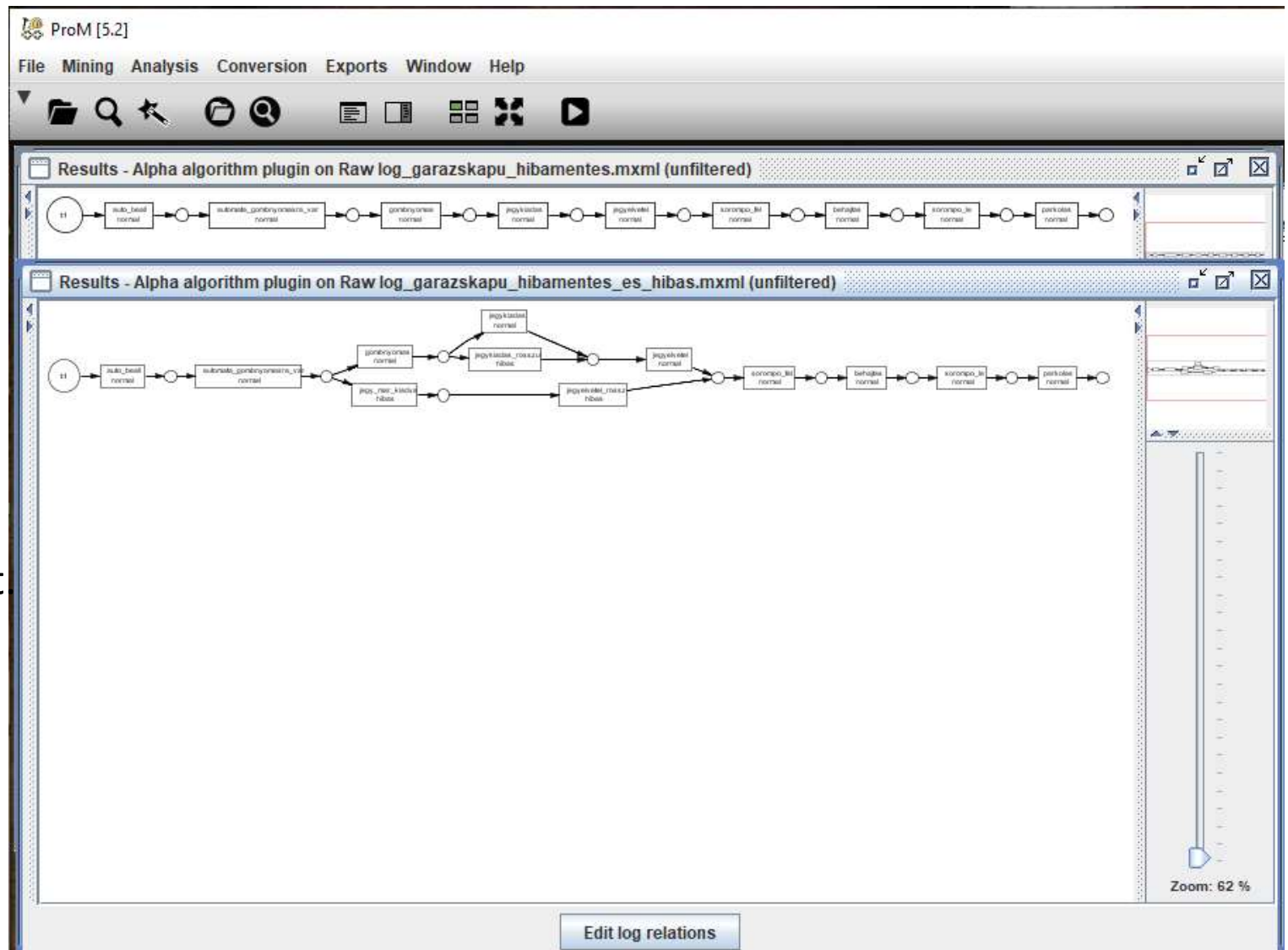
The 'Event classes per case' chart shows a distribution with a mean of 8. Summary statistics are: Min 8, Mean 8, Max 9.

Log info for both windows:

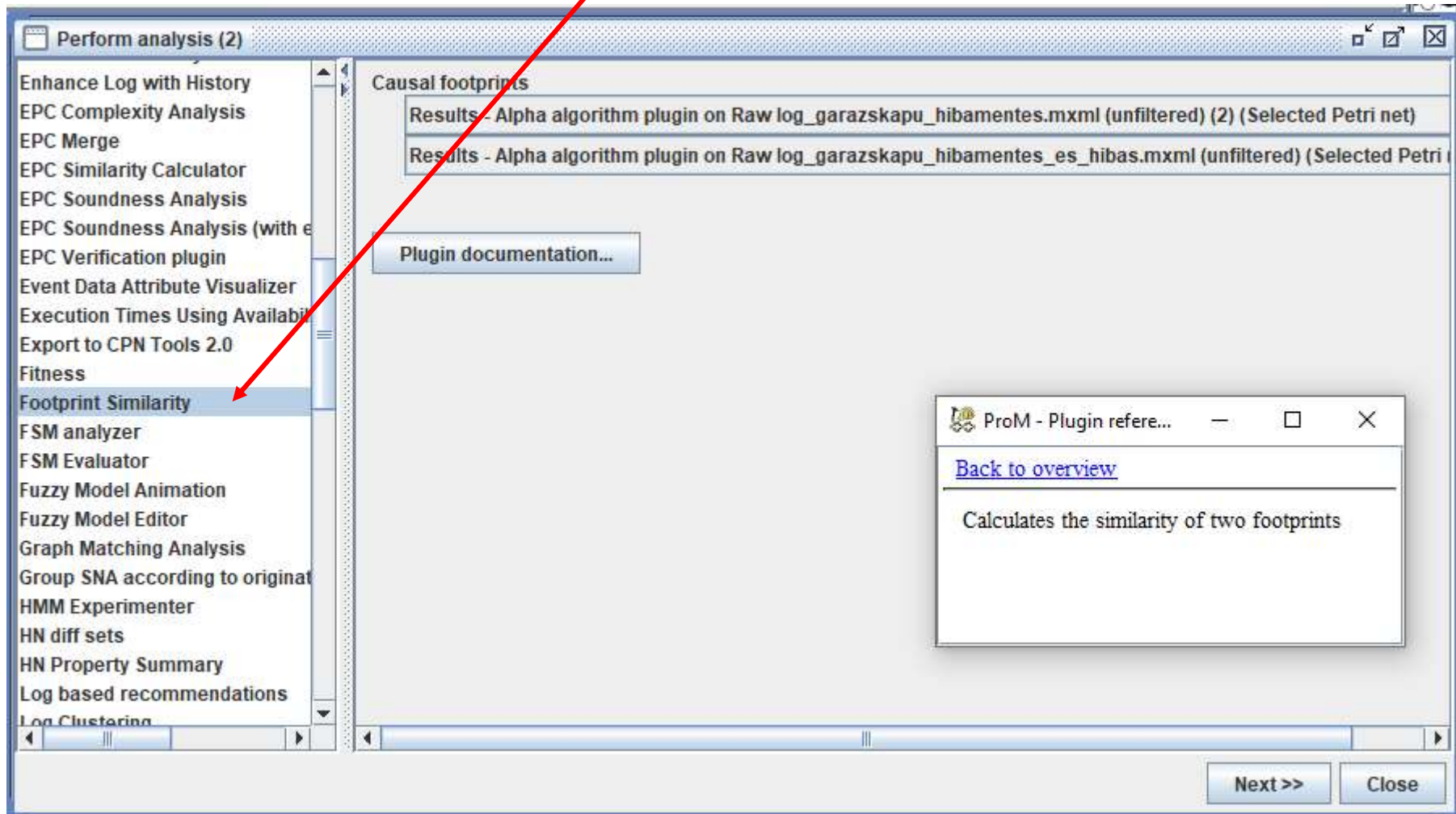
- Source: MXML generator
- Source program: MXML generator
- Start date: 2010-02-10 13:46:25
- End date: 2010-02-10 13:50:35 (top) / 2010-02-10 15:50:35 (bottom)
- Description: Created by manual, using an existing Petri net

A 'start analyzing this log' button is visible at the bottom right of the second window.

Ahhoz, hogy használhassuk a **Footprint Similarity** plugint, másodszor mindkét .mxml állományon lefuttatjuk az Alpha algoritmust. Ezt elérhetjük a **Mining** menü keresztül **Alpha algorithm plugin** néven.



A Footprint Similarity algoritmust az **Analysis** menü keresztül a **More Analysis** csoportban érhetjük el



Eredmény a két modell összehasonlítása után

The screenshot displays the 'Footprint Similarity' application window. On the left, 'modell 1' is shown with a simple flow diagram. On the right, 'modell 2' is shown with a more complex flow diagram. The central panel lists elements from both models for comparison:

Element from left model	Element from right model
auto_beall (normal)	auto_beall (normal)
automata_gombnyomasra_var (normal)	automata_gombnyomasra_v...
behajtas (normal)	behajtas (normal)
gombnyomas (normal)	gombnyomas (normal)
jegyelvetel (normal)	jegyelvetel (normal)
jegykiadas (normal)	jegykiadas (normal)
parkolas (normal)	parkolas (normal)
sorompo_fel (normal)	sorompo_fel (normal)
sorompo_le (normal)	sorompo_le (normal)

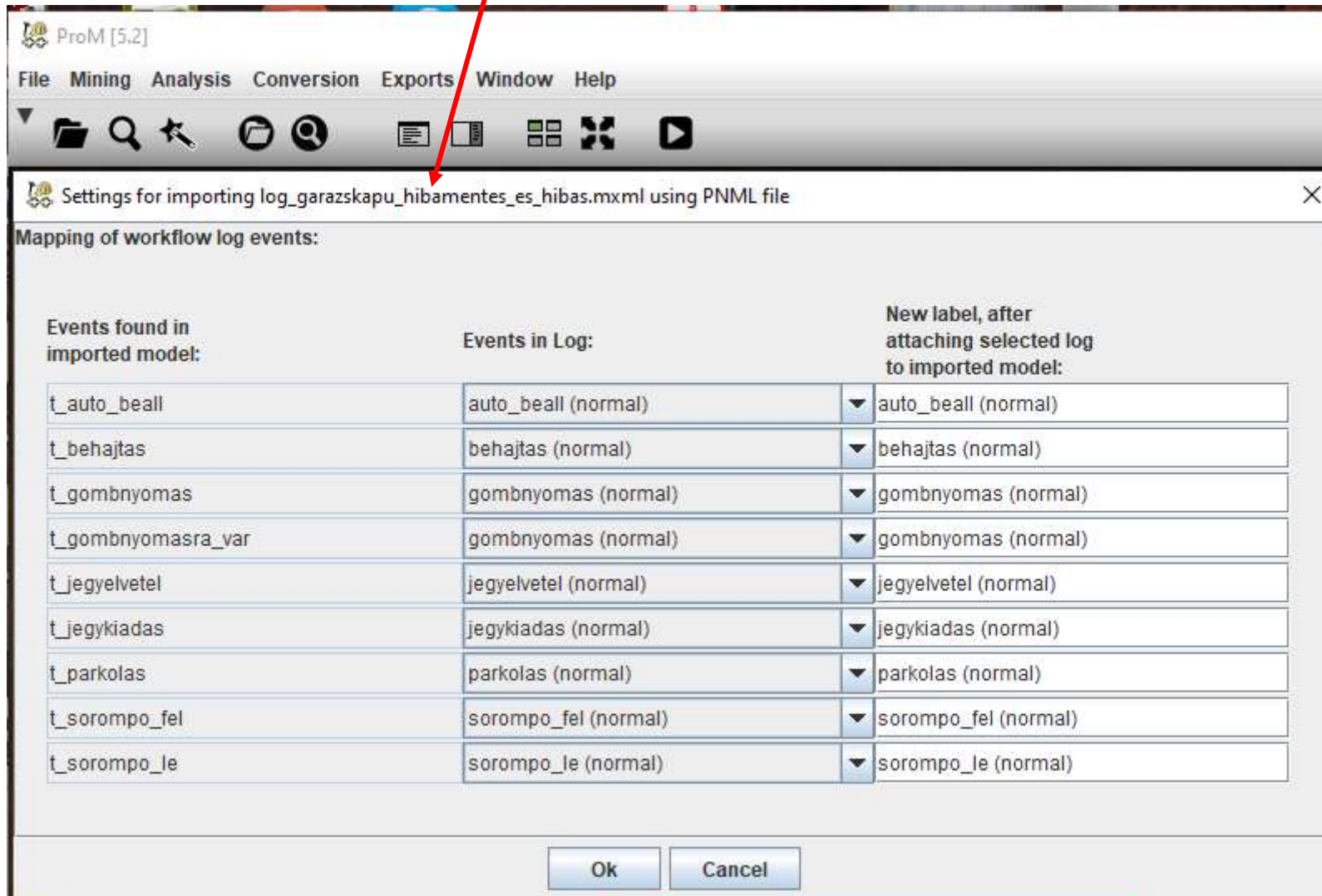
Below the list is a red-bordered button labeled 'Calculate Similarity'. To its right, the similarity result '58,81 %' is displayed in a green circle. A red text annotation 'Kattintás a gombra!' points to the button. The interface also features zoom controls at the bottom, set to 50%.

Token alapú egyezőség vizsgálata:

Betöltöttük a .mxml fájl



Betöltöttük a hibamentes modellt a hibás loggal
Ehhez használjuk a **File menü Open PNML file**
parancsát



Nem egyezett a modell és a log, megváltoztatjuk:

Settings for importing log_garazskapu_hibamentes_es_hibas.mxml using PNML file

Mapping of workflow log events:

Events found in imported model:	Events in Log:	New label, after attaching selected log to imported model:
t_auto_beall	auto_beall (normal)	auto_beall (normal)
t_behajtas	behajtas (normal)	behajtas (normal)
t_gombnyomas	gombnyomas (normal)	gombnyomas (normal)
t_gombnyomasra_var	gombnyomas (normal)	gombnyomas (normal)
t_jegyvetel	jegyvetel (normal)	jegyvetel (normal)
t_jegykiadas	jegykiadas (normal)	jegykiadas (normal)
t_parkolas	parkolas (normal)	parkolas (normal)
t_sorompo_fel	sorompo_fel (normal)	sorompo_fel (normal)
t_sorompo_le	sorompo_le (normal)	sorompo_le (normal)

Ok Cancel

Help

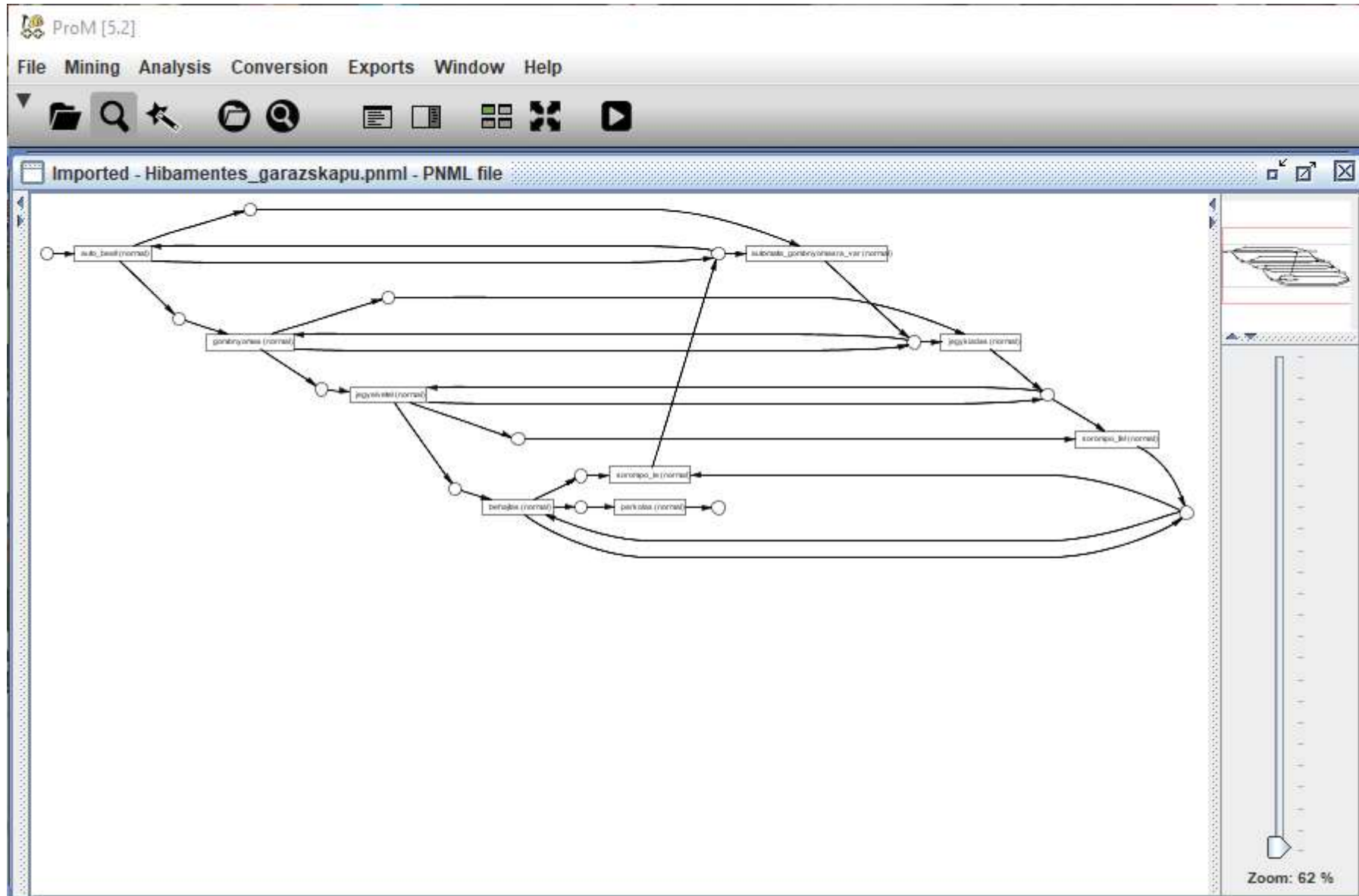
Settings for importing log_garazskapu_hibamentes_es_hibas.mxml using PNML file

Mapping of workflow log events:

Events found in imported model:	Events in Log:	New label, after attaching selected log to imported model:
t_auto_beall	auto_beall (normal)	auto_beall (normal)
t_behajtas	behajtas (normal)	behajtas (normal)
t_gombnyomas	gombnyomas (normal)	gombnyomas (normal)
t_gombnyomasra_var	automata_gombnyomasra_var (normal)	automata_gombnyomasra_var (normal)
t_jegyvetel	jegyvetel (normal)	jegyvetel (normal)
t_jegykiadas	jegykiadas (normal)	jegykiadas (normal)
t_parkolas	parkolas (normal)	parkolas (normal)
t_sorompo_fel	sorompo_fel (normal)	sorompo_fel (normal)
t_sorompo_le	sorompo_le (normal)	sorompo_le (normal)

Ok Cancel

Modell betöltése



Beállíthatjuk, hogy mit szeretnénk kiszámolni, látni:

ProM [5.2]

File Mining Analysis Conversion Exports Window Help

Analysis - Conformance Checker

The Conformance Checker has automatically determined the maximum search depth needed to transparently fire invisible tasks during the replay of your model (if any). In the case of computability problems, one might want to decrease the search depth to get a response (setting it to 0 will result in not searching at all). However, this is likely to yield pessimistic measurements.

Restrict search depth for invisible tasks Maximum depth:

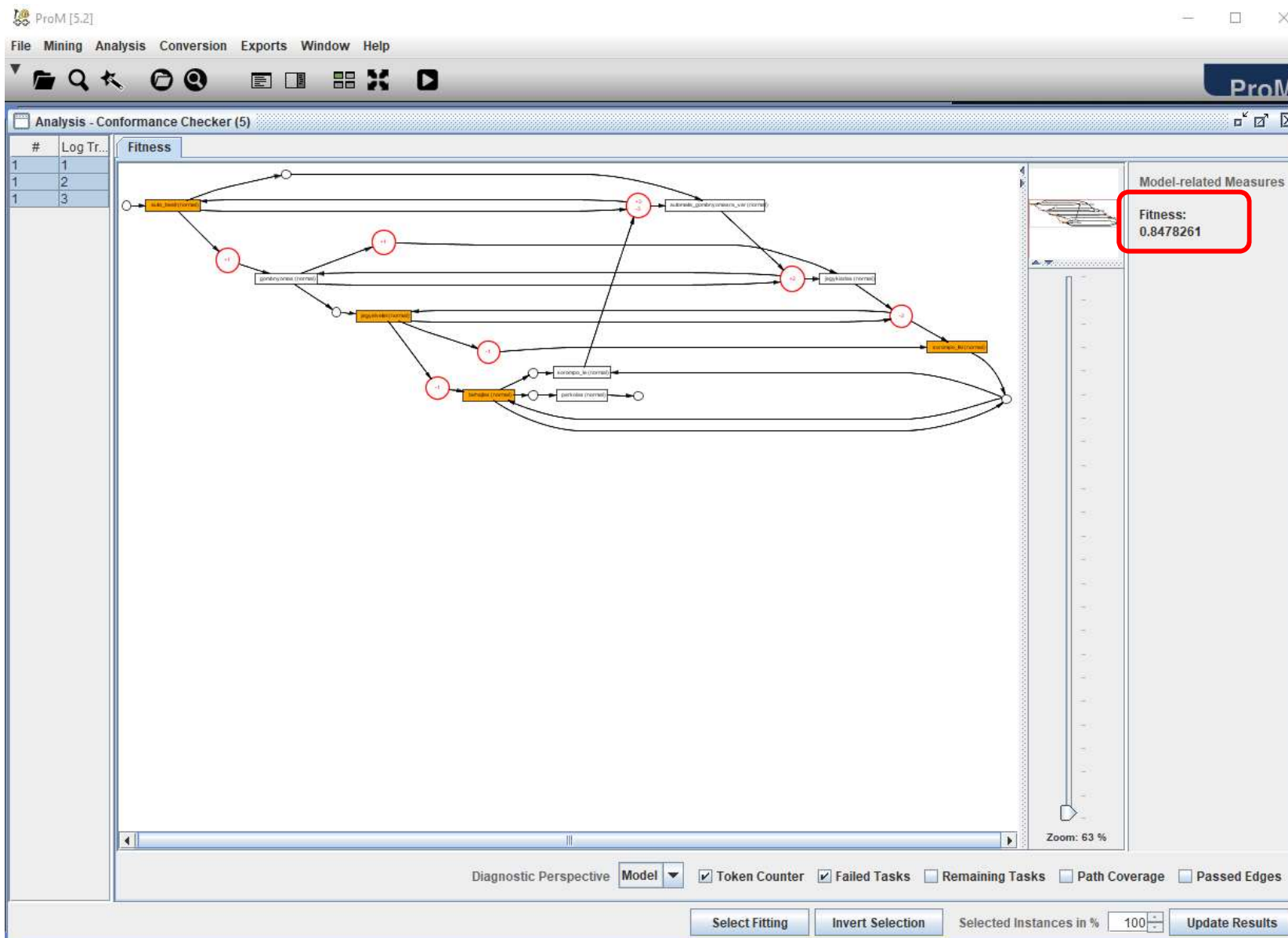
Choose best shortest sequence of invisible tasks

Furthermore, you can choose which kind of analysis you would like to perform. The computation process may speed up if you deselect the categories (fitness, precision, structure), or specific metrics, in which you are not interested.

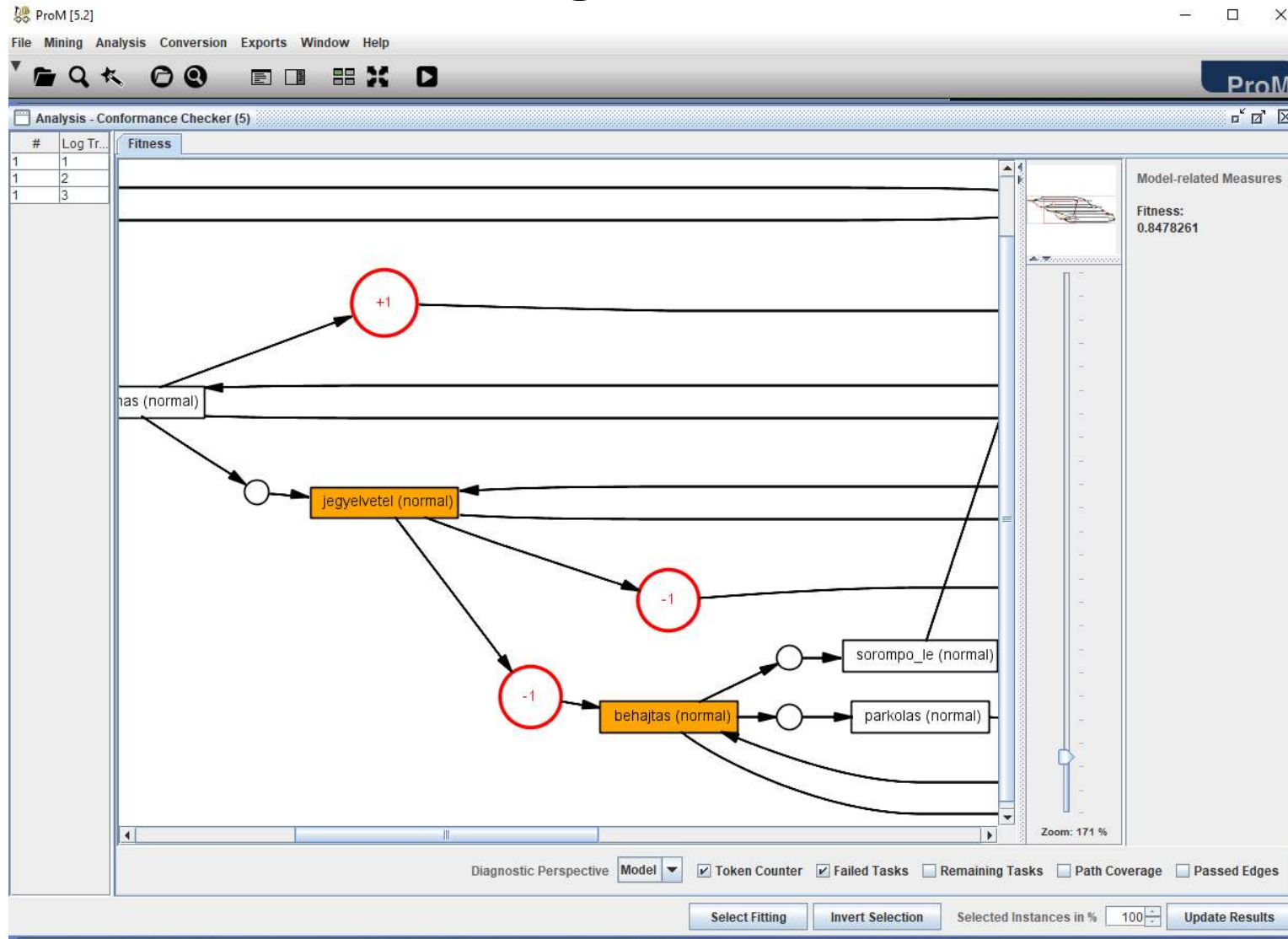
- Fitness** Fitness evaluates whether the observed process *complies with* the control flow specified by the process. One way to investigate the fitness is to replay the log in the Petri net. The log replay is carried out in a non-blocking way, i.e., if there are tokens missing to fire the transition in question they are created artificially and replay proceeds. While doing so, diagnostic data is collected and can be accessed afterwards.
 - f** The token-based **fitness** metric f relates the amount of missing tokens during log replay with the amount of consumed ones and the amount of remaining tokens with the produced ones. If the log could be replayed correctly, that is, there were no tokens missing nor remaining, it evaluates to 1.
 - pSE** The **successful execution** metric p_{SE} determines the fraction of successfully executed process instances (taking the number of occurrences per trace into account).
 - pPC** The **proper completion** metric p_{PC} determines the fraction of properly completed process instances (taking the number of occurrences per trace into account).
- Precision** Precision, or Behavioral Appropriateness, evaluates *how precisely* the model describes the observed process.
 - saB** The **simple behavioral appropriateness** metric sa_B is based on the mean number of enabled transitions during log replay (the greater the value the less behavior is allowed by the process model and the more precisely the behavior observed in the log is captured). Note that this metric should only be used as a comparative means for models without alternative duplicate tasks. Note further that in order to determine the mean number of enabled tasks in the presence of invisible tasks requires to build the state space from the current marking after each replay step. Since this may greatly decrease the performance of the computational process, you might want to switch this feature off.
 - aaB** The **advanced behavioral appropriateness** metric aa_B is based on successorship relations among activities with respect the event relations observed in the log (the greater the value the more precisely the behavior observed in the log is captured).
- Structure** Structural Appropriateness evaluates whether the model describes the observed process in a *structurally suitable* way.
 - saS** The **simple structural appropriateness** metric sa_S is a simple metric based on the graph size of the model (the greater the value the more compact is the model). Note that this metric should only be used as a comparative means for models allowing for the same amount of behavior.
 - aaS** The **advanced structural appropriateness** metric aa_S is based on the detection of redundant invisible tasks (simply superfluous) and alternative duplicate tasks (list alternative behavior rather than expressing it in a meaningful way).

Help... Start Analysis

Az összehasonlítás eredménye:



Megtekinthetjük, hogy hol volt hiányzó token, hol keletkezett megmaradt token:



Egy másik példa, ami betölthető a ProM 6.9-be: Lfull.mxml betöltése

The screenshot shows the ProM 6.9 software interface. The title bar reads 'ProM UITopia'. The main window has a 'Workspace' header with an 'import...' button. On the left, there are navigation buttons: 'All', 'Favorites', 'Imported', and 'Selection'. The central workspace shows a resource named 'L-conf XLog' with a star icon and a play button. A red arrow points from the text above to this resource. A tooltip on the right displays the log content, which is a list of cases and their associated letters.

sort by ABC

L-conf
XLog

L-conf
XLog
just created
imported

Show
 Show

Rename resource

```
455x Case1 a c d e h
191x Case2 a b d e g
177x Case3 a d c e h
144x Case4 a b d e h
111x Case5 a c d e g
82x Case6 a d c e g
56x Case7 a d b e h
47x Case8 a c d e f d b e h
38x Case9 a d b e g
33x Case10 a c d e f b d e h
14x Case11 a c d e f b d e g
11x Case12 a c d e f d b e g
9x Case13 a d c e f c d e h
8x Case14 a d c e f d b e h
5x Case15 a d c e f b d e g
3x Case16 a c d e f b d e f d b e g
2x Case17 a d c e f d b e g
2x Case18 a d c e f b d e f b d e g
1x Case19 a d c e f d b e f b d e h
1x Case20 a d b e f b d e f d b e g
1x Case21 a d c e f d b e f c d e f d b e g
```

Ez a log tartalma

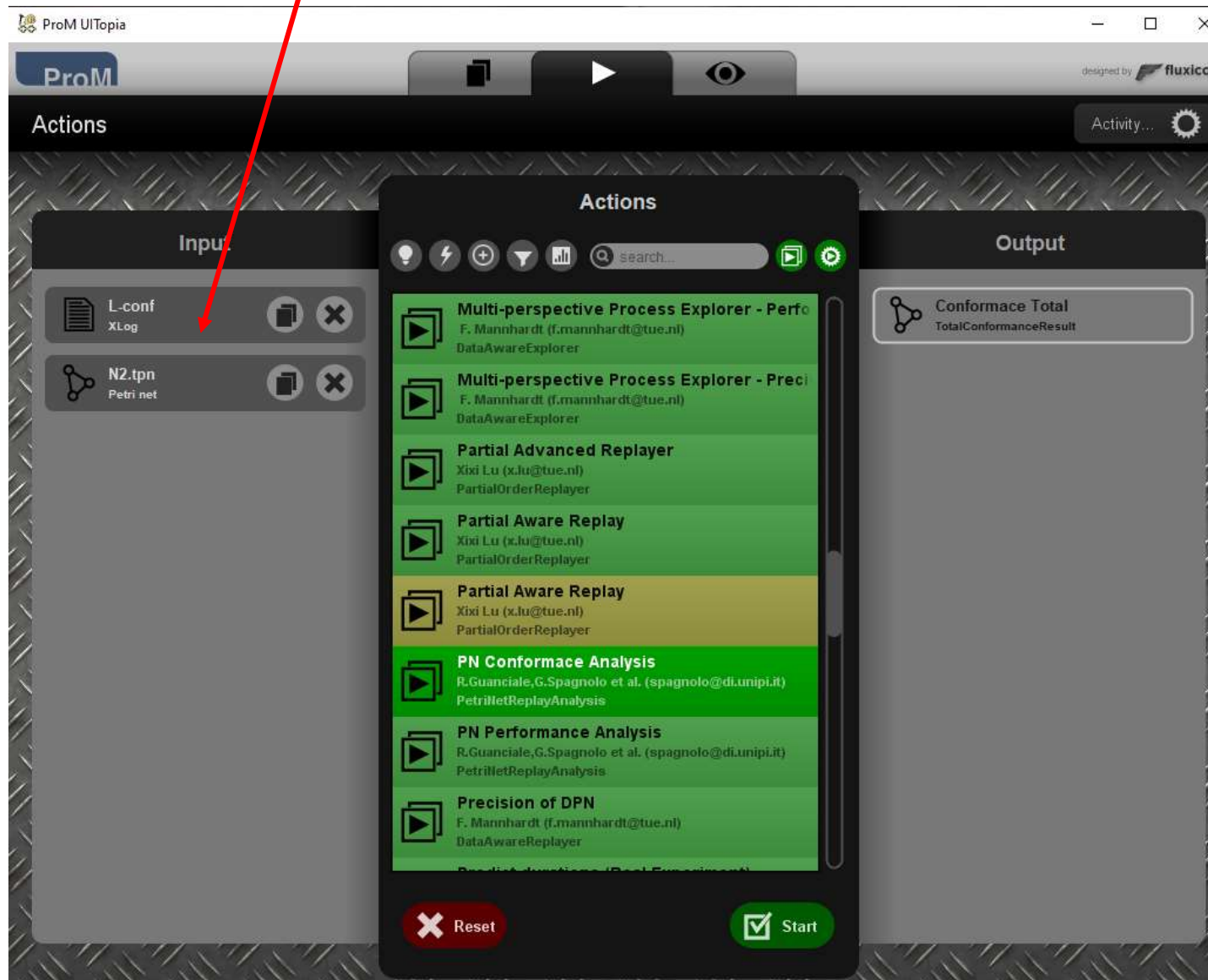
N1.pnml betöltése

The screenshot shows the ProM UI interface. The top bar includes the ProM logo, a workspace title, and an 'import...' button. The main workspace area contains a file list on the left and a detailed view of the selected file on the right. The file list shows two items: 'N1.tpn' (Petri net) and 'L-conf' (XLog). The 'L-conf' file is selected and highlighted in green. The detailed view on the right shows the 'L-conf' XLog file, which was imported. Below the file name are several icons: a star, an eye, a play button, and a close button. There are also buttons for 'Show parents', 'Show children', 'Rename resource', and 'Export to disk'. A red arrow points from the title 'N1.pnml betöltése' to the 'N1.tpn' file in the list.

```
place "start" init 1;
place "p1";
place "p2";
place "p3";
place "p4";
place "end";
trans "a"~"a\ncomplete" in "start" out "p1";
trans "b"~"b\ncomplete" in "p1" out "p2" ;
trans "c"~"c\ncomplete" in "p1" out "p2" ;
trans "d"~"d\ncomplete" in "p2" out "p3" ;
trans "e"~"e\ncomplete" in "p3" out "p4" ;
trans "f"~"f\ncomplete" in "p4" out "p1" ;
trans "g"~"g\ncomplete" in "p4" out "end" ;
trans "h"~"h\ncomplete" in "p4" out "end" ;
```

Ez a modell felépítése

Ha mind a két fájl betöltésre került, akkor a **PN Conformance Analysis** plugint használhatjuk



Eredmények:
f=0,87
d 443 trace-nél

```
place "start" init 1;  
place "p1";  
place "p2";  
place "p3";  
place "p4";  
place "end";  
trans "a~"a\ncomplete" in "start" out "p1";  
trans "b~"b\ncomplete" in "p1" out "p2" ;  
trans "c~"c\ncomplete" in "p1" out "p2" ;  
trans "d~"d\ncomplete" in "p2" out "p3" ;  
trans "e~"e\ncomplete" in "p3" out "p4" ;  
trans "f~"f\ncomplete" in "p4" out "p1" ;  
trans "g~"g\ncomplete" in "p4" out "end" ;  
trans "h~"h\ncomplete" in "p4" out "end" ;
```

