Mediating Between Modeled and Observed Behavior: The Quest for the "Right" Process

prof.dr.ir. Wil van der Aalst

PAIS lab

Лаборатория ПОИС

HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

Outline

Introduction to Process Mining (short)

The 4+ dimensions of conformance

Importance of alignments to relate observed and modeled behavior

Representational bias

Mediating between a reference model and observed behavior (model repair)

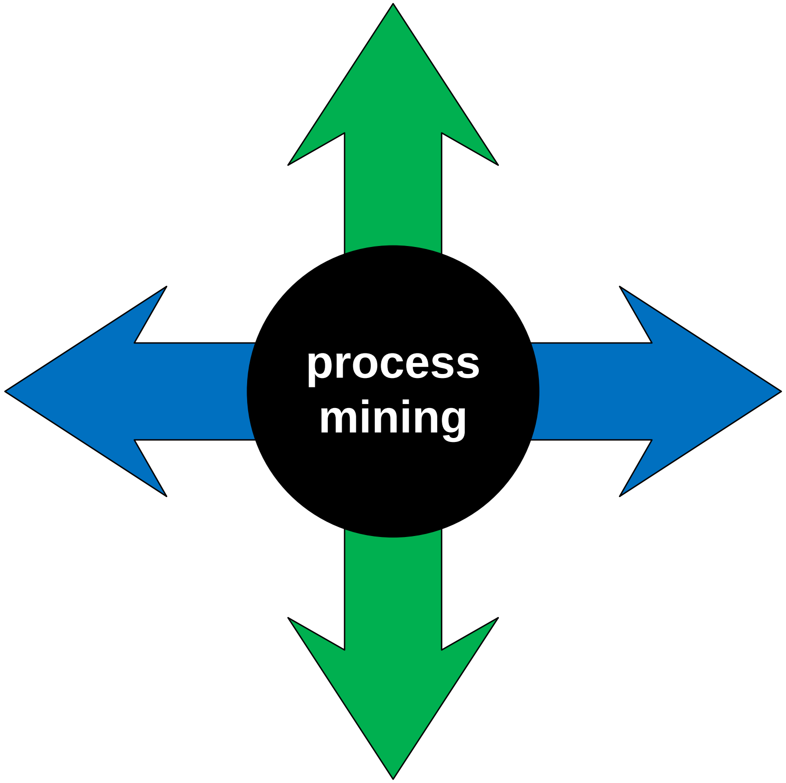
Discovering configurable process models

Decomposing process mining problems to deal with Big Data

Positioning Process Mining

process model analysis

(simulation, verification, etc.)



process mining

performance-oriented questions, problems and solutions

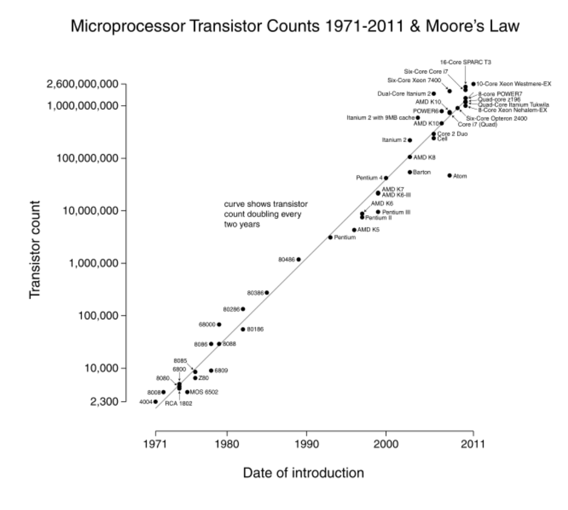
compliance-oriented questions, problems and solutions

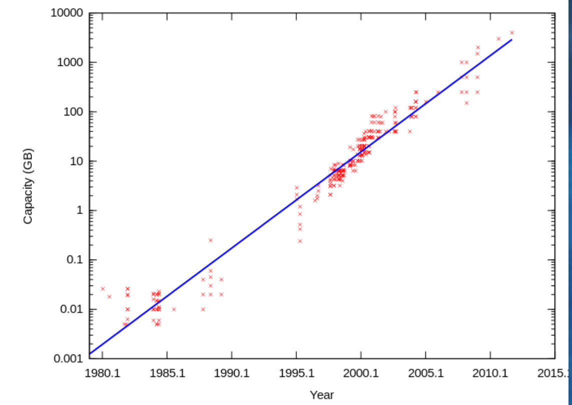
data-oriented analysis

(data mining, machine learning, business intelligence)

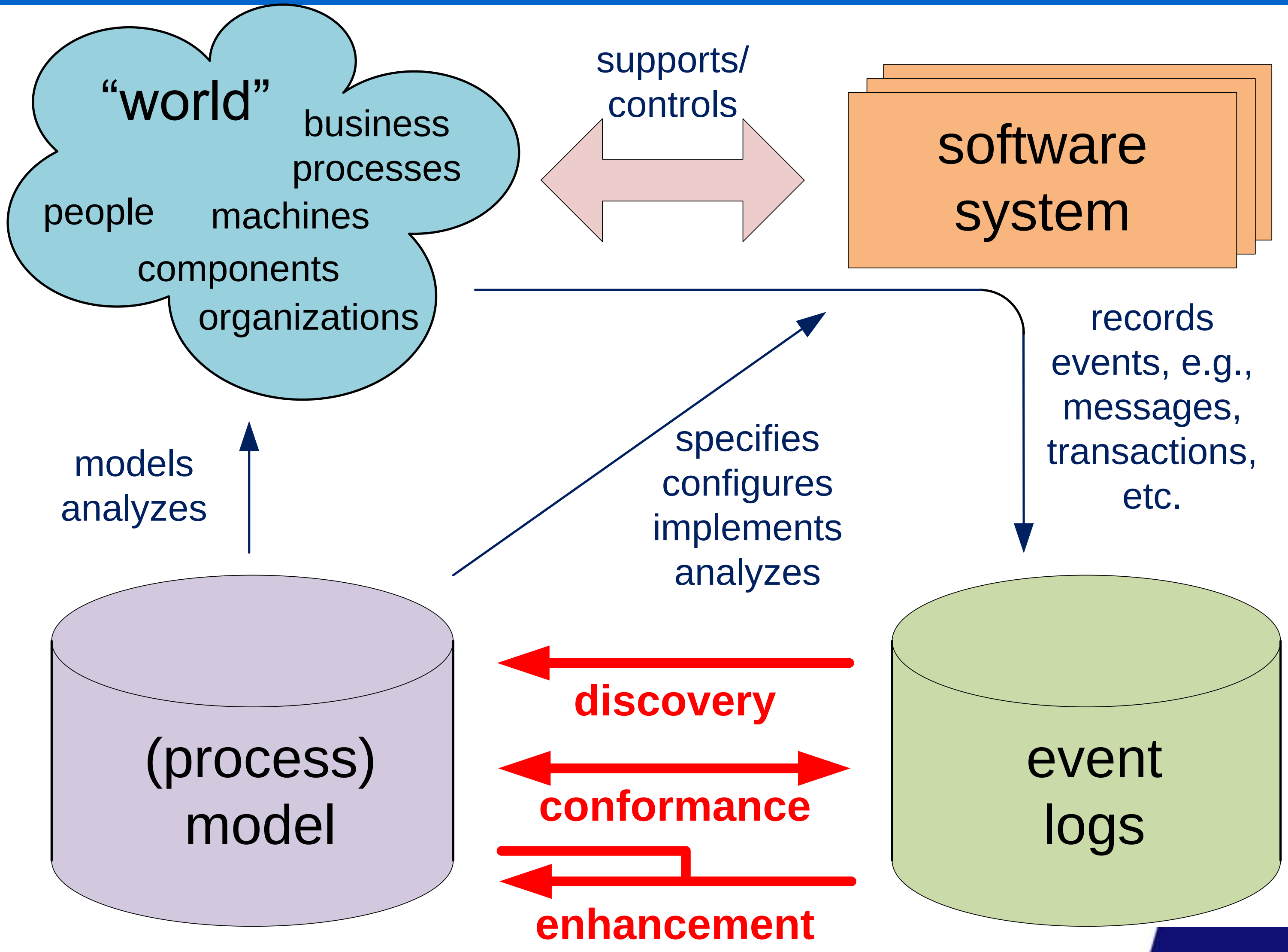
Moore's Law

2013

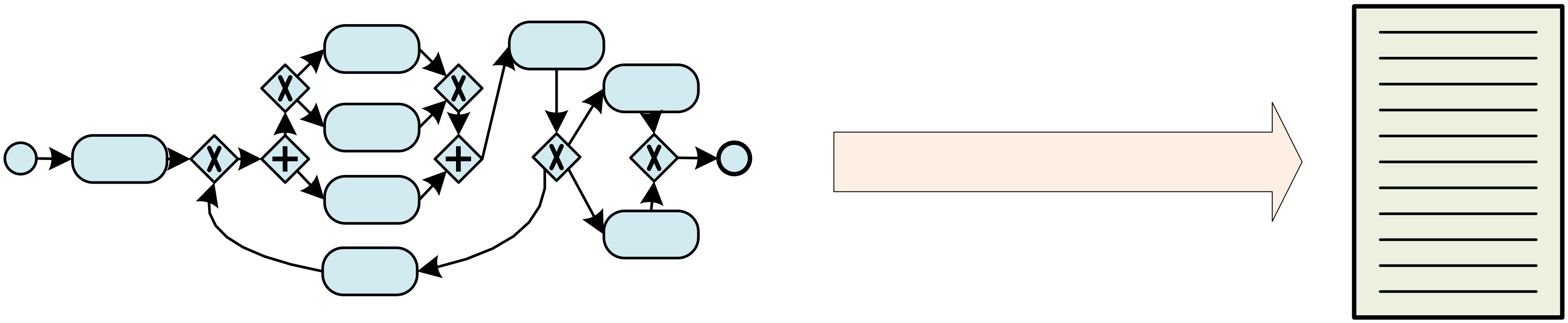




Three Types of Process Mining



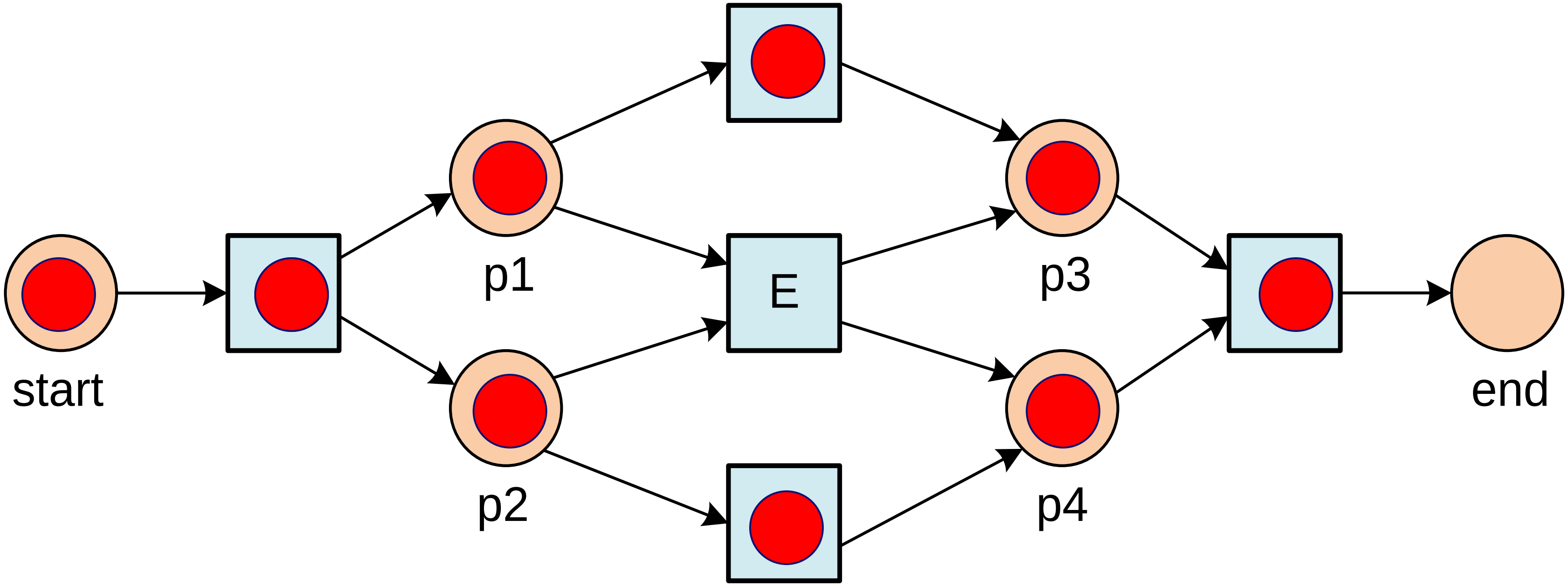
Play-Out



process model

event log

Play-Out (Classical use of models)



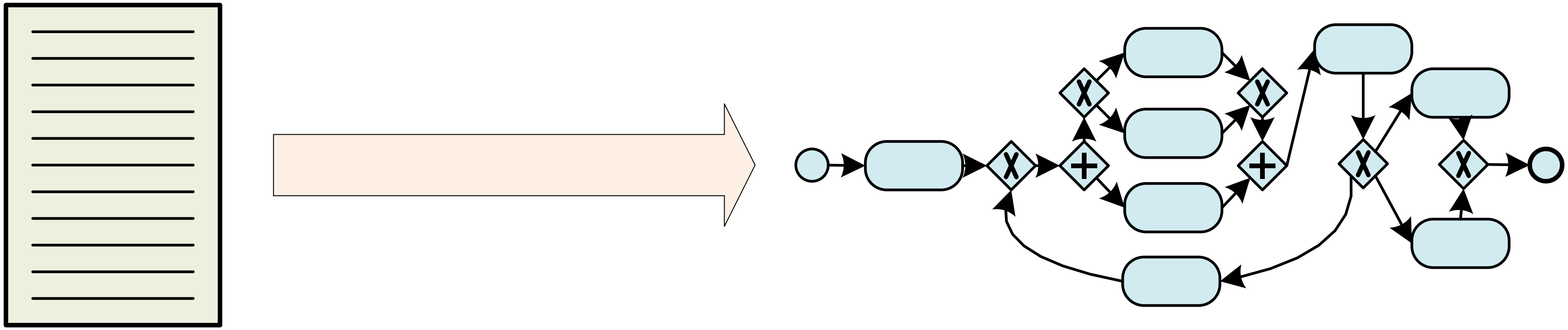
A B C D A E D A E D

A B C D A C B D

A C B D

A E D A C B D

Play-In



event log

process model

Play-In

A B C D

A E D

A E D

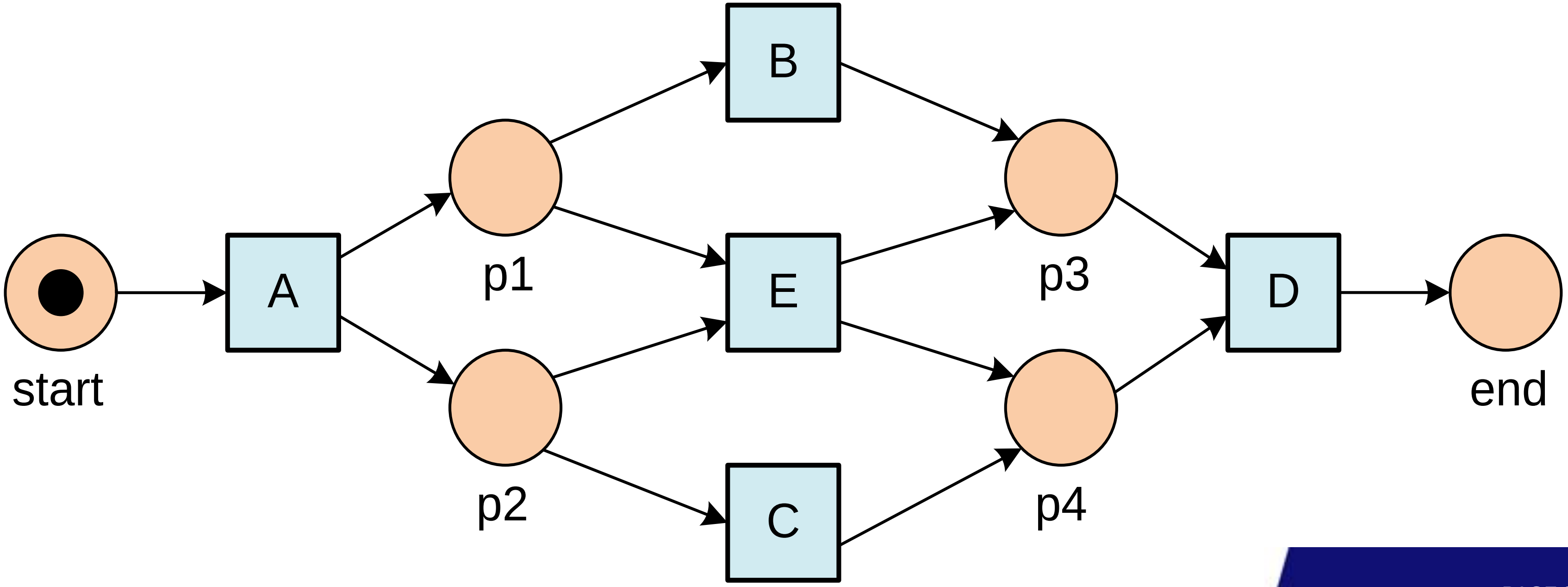
A B C D

A C B D

A C B D

A E D

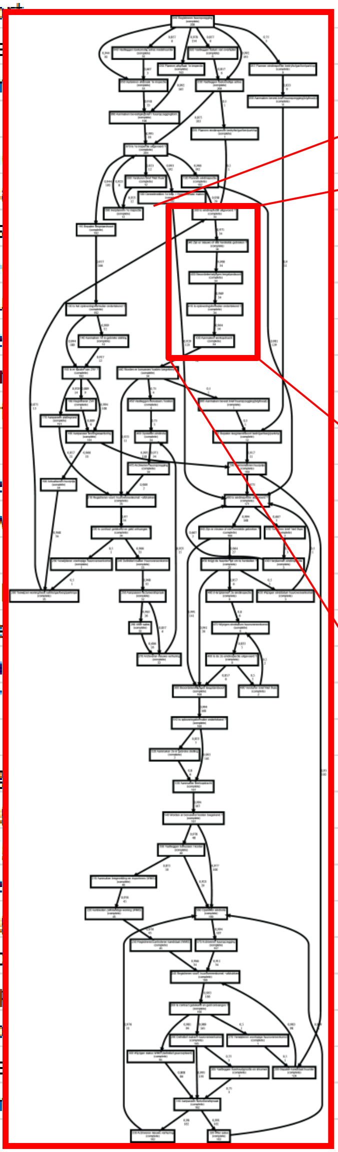
A C B D

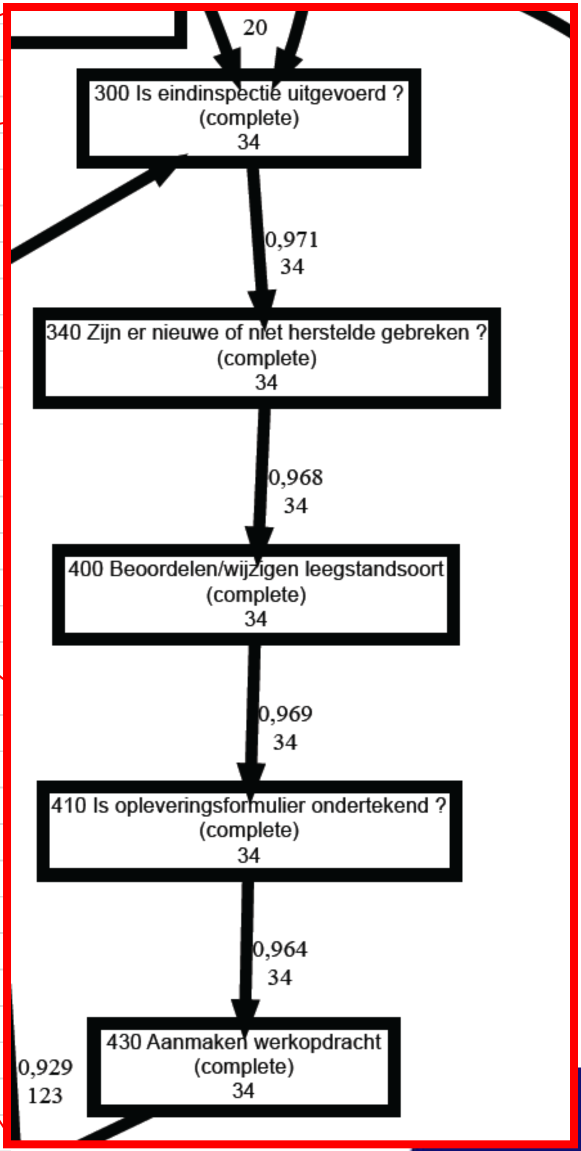


Example Process Discovery

(Vestia, Dutch housing agency, 208 cases, 5987 events)

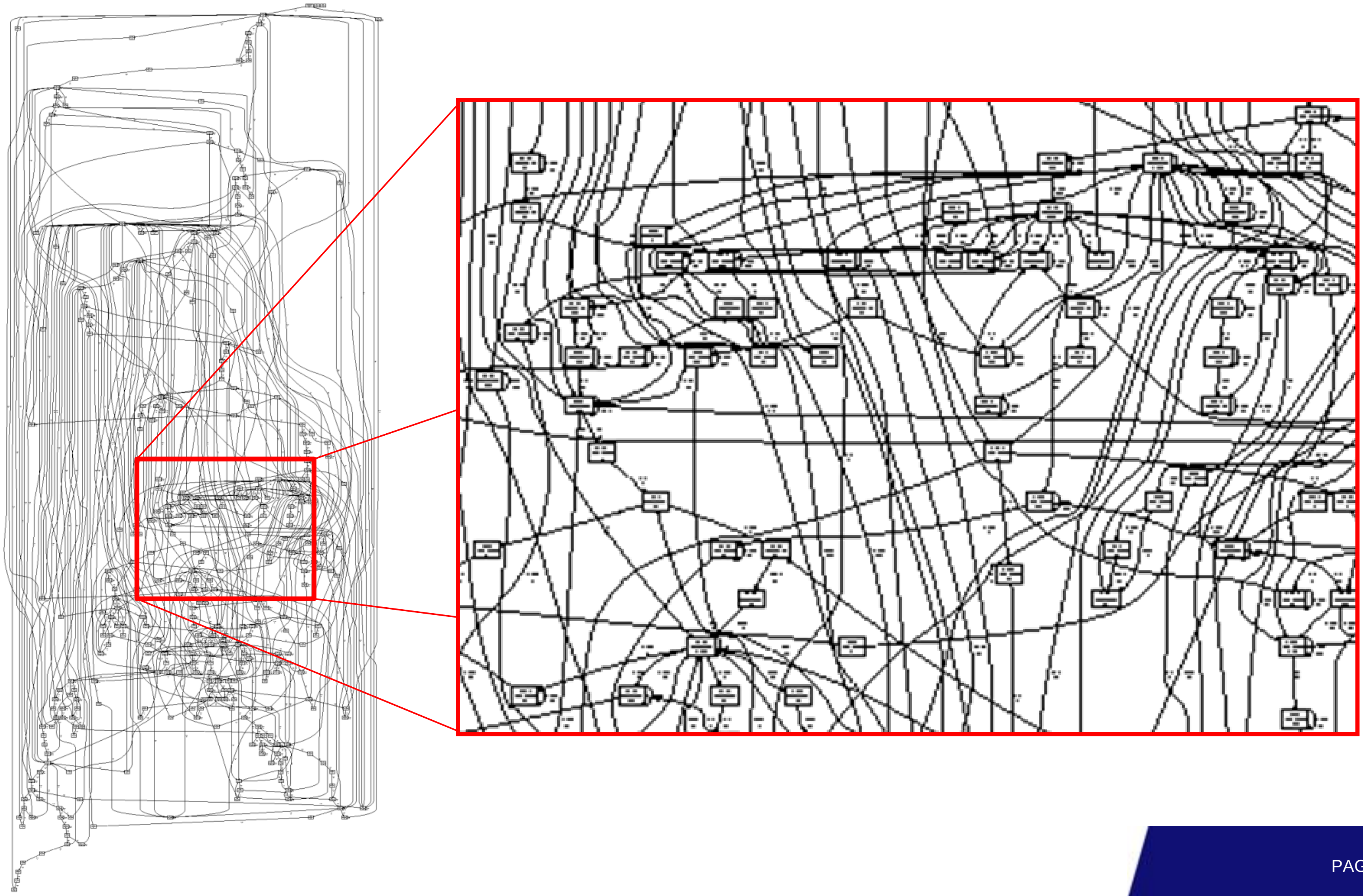
117315	110 Bepalen leegstandsoo	16.05.2007 14:06:23
117315	120 Plannen eindinspectie	16.05.2007 14:36:01
117315	130 Is het opleveringsform	23.05.2007 09:41:40
117315	150 Is er sprake van ZAV ?	23.05.2007 09:41:51
117315	170 Aanpassen plattegron	23.05.2007 11:57:18
117315	180 Aanpassen woningwa	23.05.2007 09:42:37
117315	190 Actualiseren huurprijs	23.05.2007 09:48:23
117315	200 Toewijzen woning/be	23.05.2007 09:48:29
117315	210 Registreren voorl. huu	10.09.2007 16:24:36
117315	220 Is contract getekend e	11.09.2007 14:56:18
117315	240 Definitief maken Huur	31.03.2008 16:17:12
117315	250 Aanpassen factureera	09.09.2008 15:39:59
117315	260 After sales	09.09.2008 16:51:24
117315	270 Archiveren nieuwe ve	10.09.2008 07:52:08
117315	300 Is eindinspectie uitge	07.06.2007 14:47:04
117315	340 Zijn er nieuwe of niet	07.06.2007 14:47:06
117315	400 Beoordelen/wijzigen	07.06.2007 14:51:16
117315	410 Is opleveringsformulie	07.06.2007 14:51:26
117315	430 Aanmaken werkopdra	11.06.2007 09:21:39
117315	440 Worden er bonussen/	11.06.2007 09:21:49
117315	460 Opstellen eindnota	08.08.2007 16:18:26
117315	470 Archiveren huuropzeg	09.08.2007 14:42:23
119763	010 Registreren huuropze	09.05.2007 11:19:14
119763	030 Vastleggen toekomsti	09.05.2007 12:25:01
119763	050 Inplannen afspraak 1e	09.05.2007 11:59:52
119763	060 Aanmaken bevestigin	09.05.2007 12:31:57
119763	070 Is 1e inspectie uitgevo	16.05.2007 13:04:26
119763	100 Gereedmelden 1e insp	16.05.2007 13:43:39
119763	110 Bepalen leegstandsoo	16.05.2007 13:43:28
119763	120 Plannen eindinspectie	16.05.2007 13:42:58
119763	130 Is het opleveringsform	16.05.2007 13:34:49
119763	150 Is er sprake van ZAV ?	16.05.2007 13:34:56




20
300 Is eindinspectie uitgevoerd ?
(complete)
34
0,971
34
340 Zijn er nieuwe of niet herstelde gebreken ?
(complete)
34
0,968
34
400 Beoordelen/wijzigen leegstandsoort
(complete)
34
0,969
34
410 Is opleveringsformulier ondertekend ?
(complete)
34
0,964
34
0,929
123
430 Aanmaken werkopdracht
(complete)
34

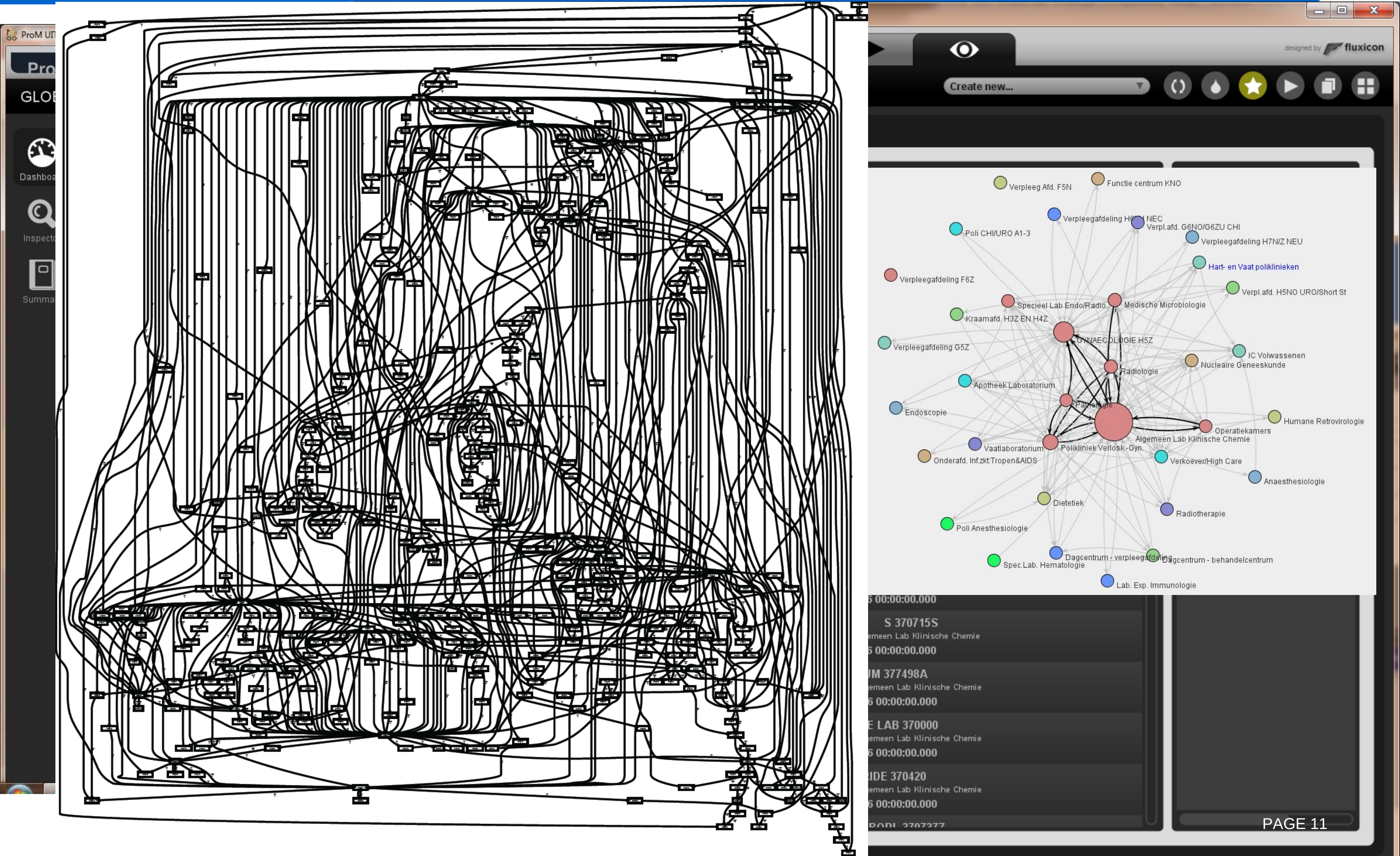
Example Process Discovery

(ASML, test process lithography systems, 154966 events)



Example Process Discovery

(AMC, 627 gynecological oncology patients, 24331 events)



6 00:00:00.000

S 370715S
emeen Lab Klinische Chemie
6 00:00:00.000

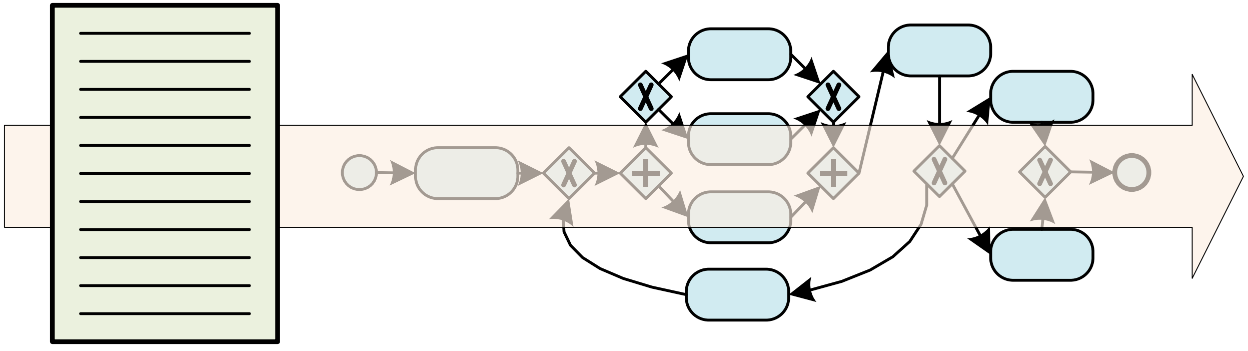
JM 377498A
emeen Lab Klinische Chemie
6 00:00:00.000

E LAB 370000
emeen Lab Klinische Chemie
6 00:00:00.000

RIDE 370420
emeen Lab Klinische Chemie
6 00:00:00.000

RODI 3707277

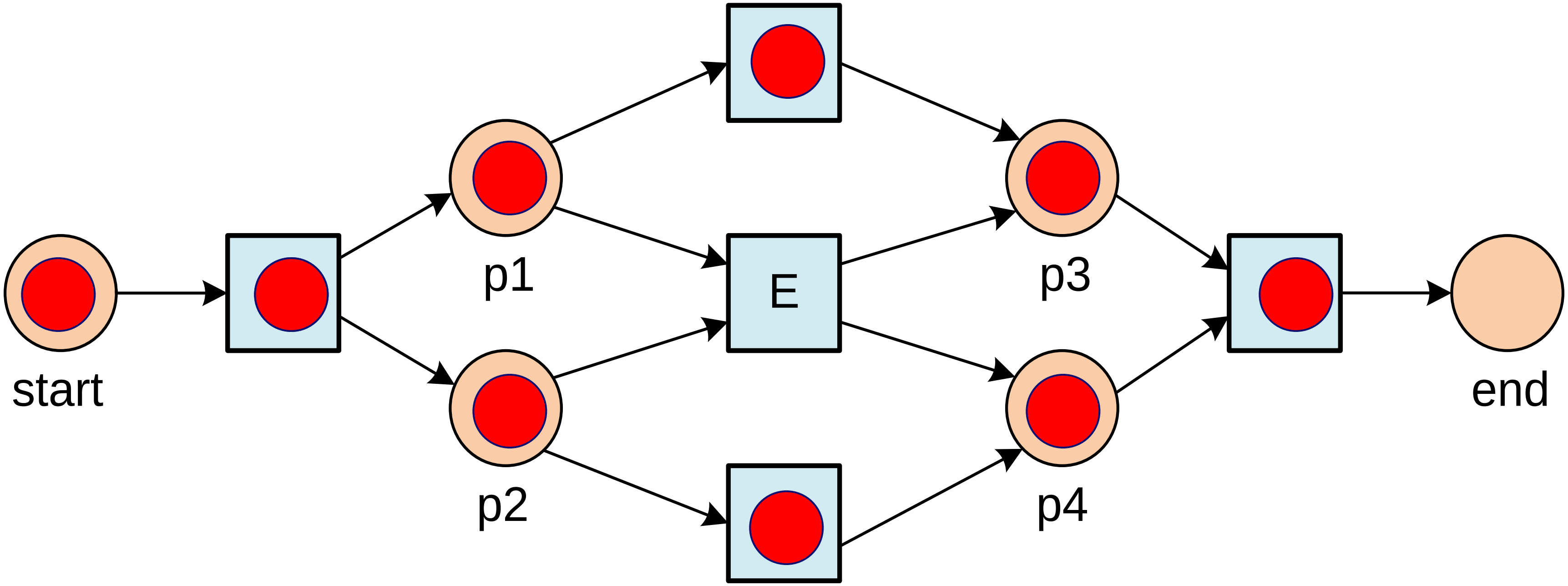
Replay



- extended model showing times, frequencies, etc.
- diagnostics
- predictions
- recommendations

Replay

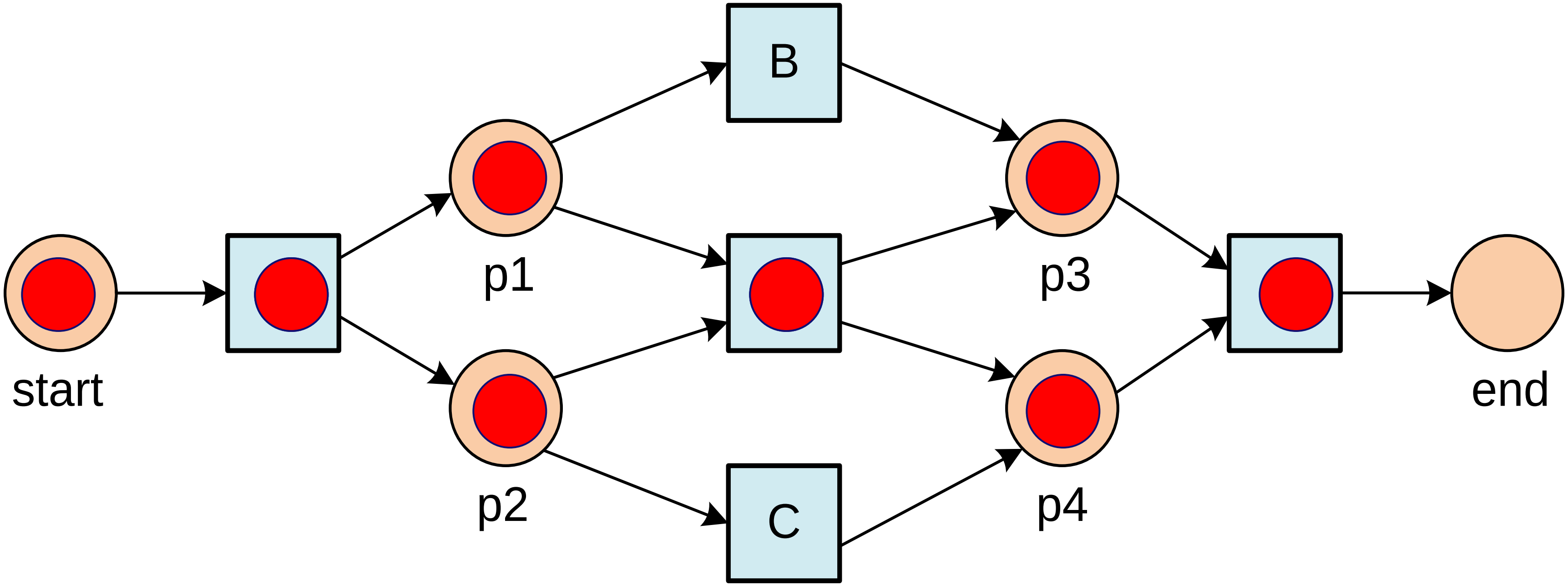
A B C D



Replay

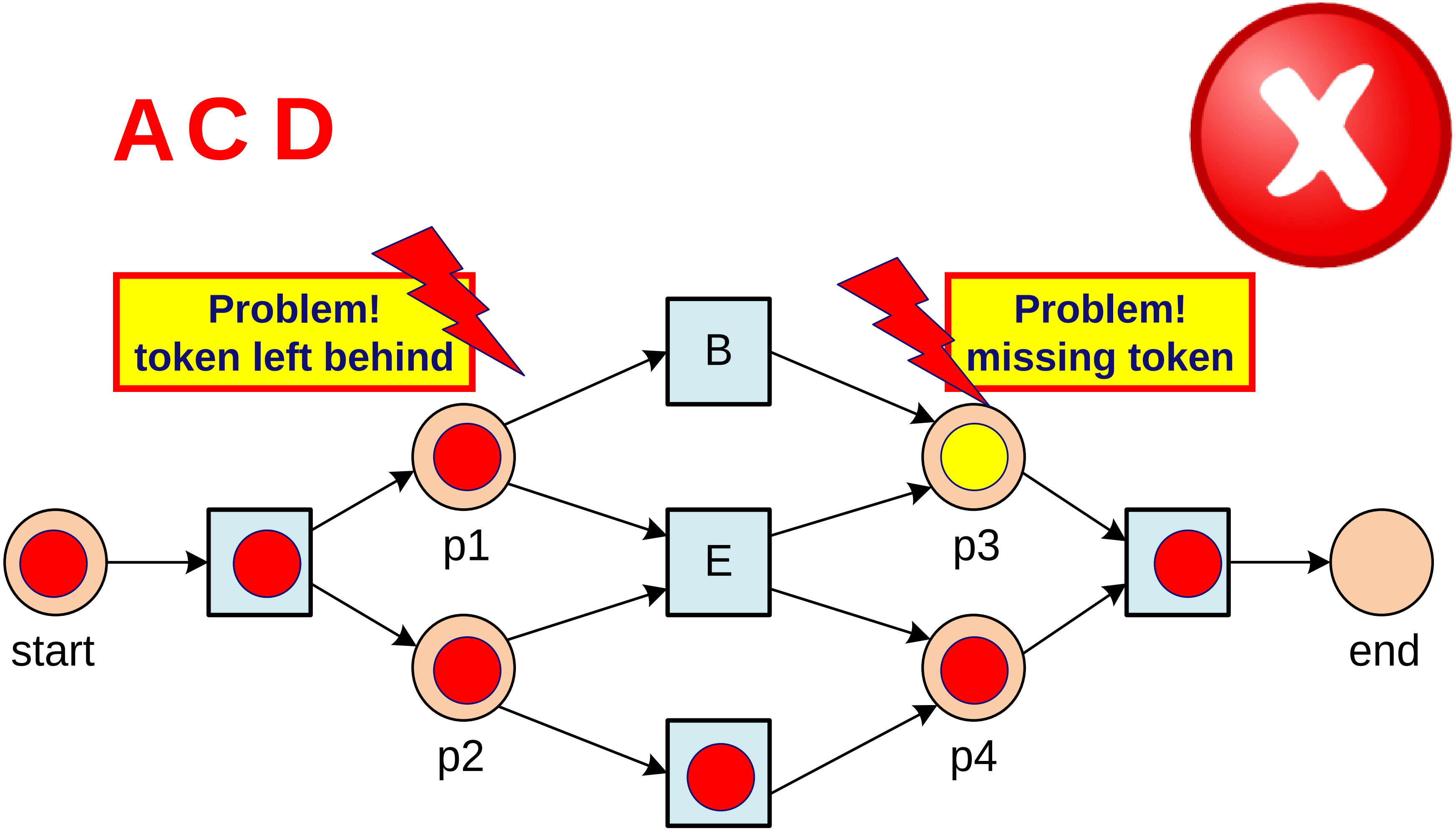
A E D

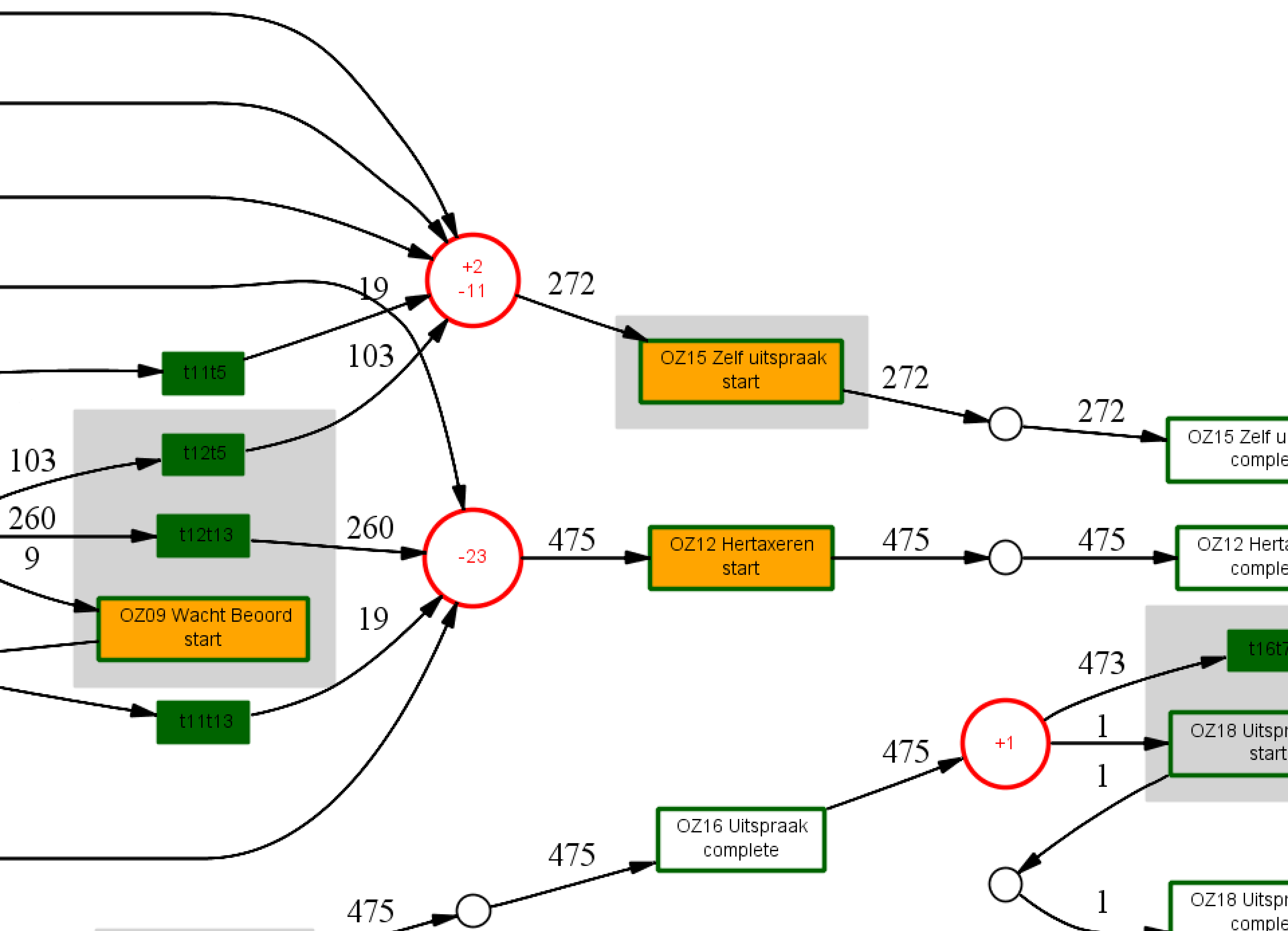
✓



Replay can detect problems

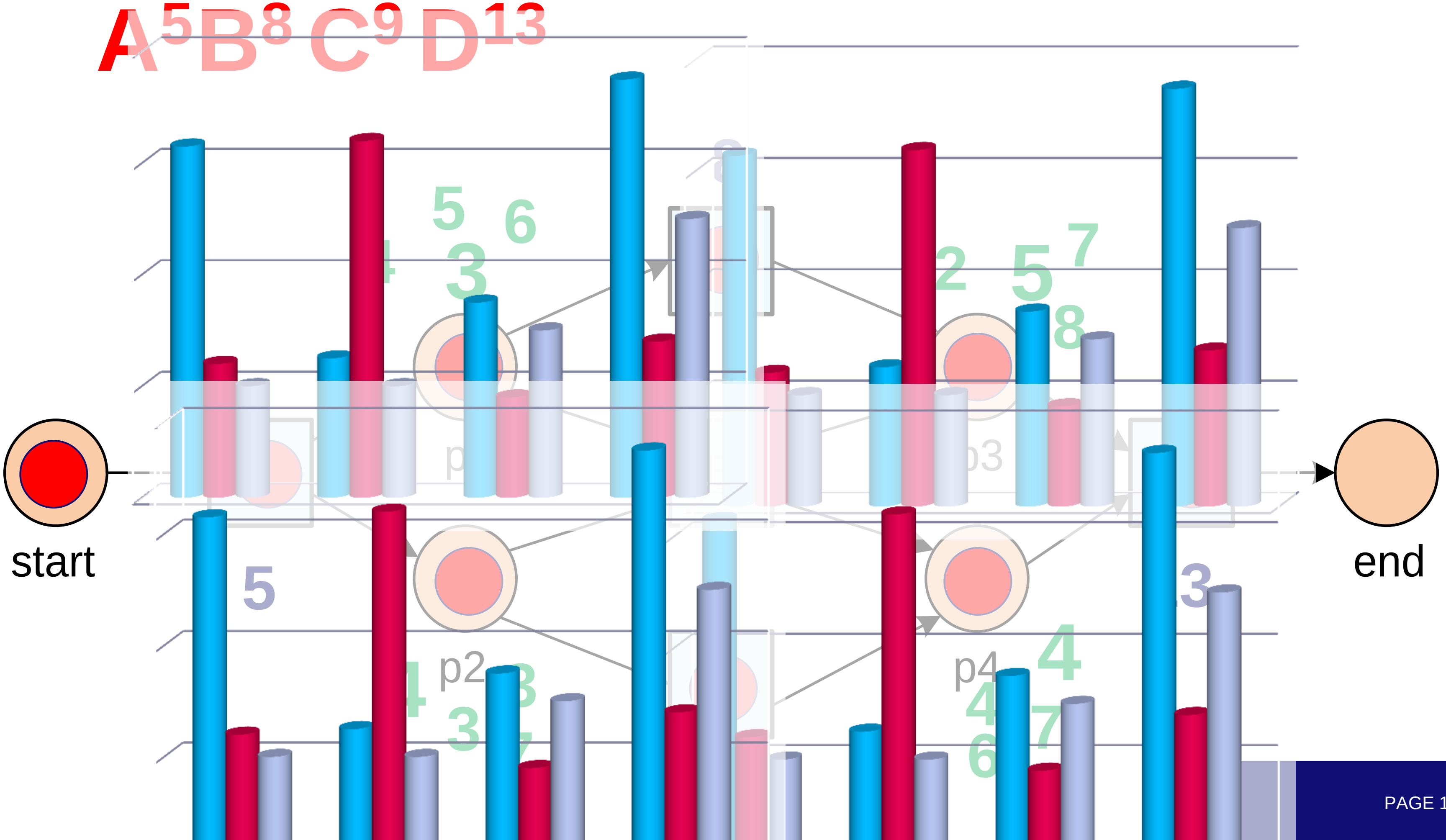
A C D




+2
-11
272
19
103
OZ15 Zelf uitspraak start
272
272
OZ15 Zelf u
comple
t11t5
103
t12t5
260
t12t13
9
260
OZ09 Wacht Beoord start
19
t11t13
-23
475
OZ12 Hertaxeren start
475
475
OZ12 Herta
comple
473
t16t7
+1
1
OZ18 Uitspr
start
1
475
OZ16 Uitspraak complete
475
475
1
OZ18 Uitspr
comple

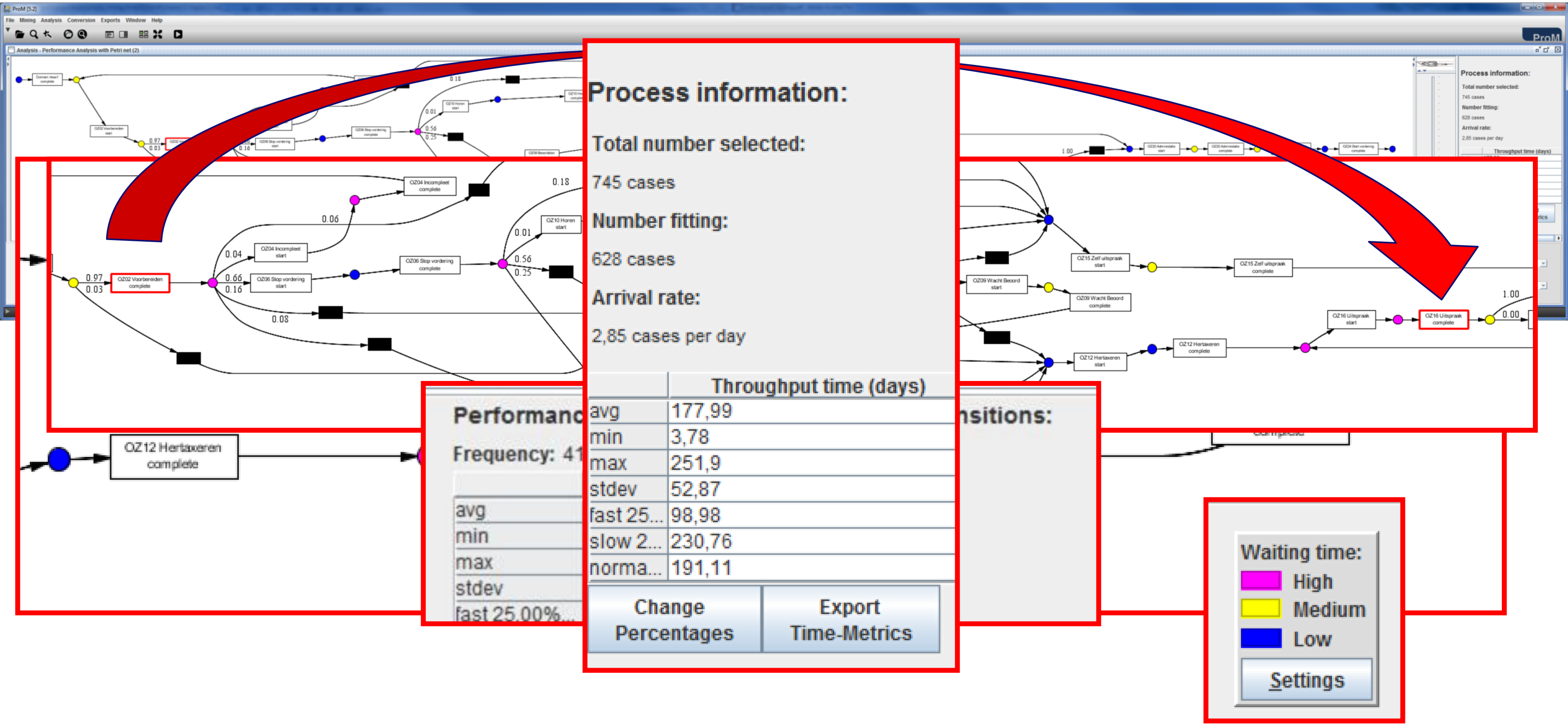
Replay can extract timing information

$A^5 B^8 C^9 D^{13}$



Performance Analysis Using Replay

(WOZ objections Dutch municipality, 745 objections, 9583 event, f= 0.988)



Process information:

Total number selected:

745 cases

Number fitting:

628 cases

Arrival rate:

2,85 cases per day

	Throughput time (days)
avg	177,99
min	3,78
max	251,9
stdev	52,87
fast 25...	98,98
slow 2...	230,76
norma...	191,11

Change Percentages | Export Time-Metrics

Waiting time:

🟪 High

🟨 Medium

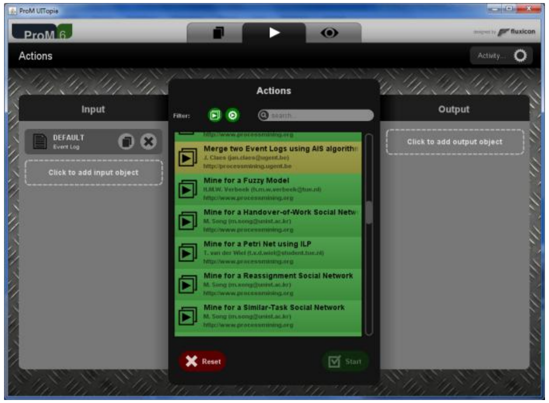
🟦 Low

Settings

Hundreds of plug-ins available covering the whole process mining spectrum

ProM
process mining workbench

open-source (L-GPL)



DEFAULT
Event Log

Click to add input object

Click to add output object

Merge two Event Logs using AIS algorithm
J. Claes (jan.claes@ugent.be)
http://processmining.ugent.be

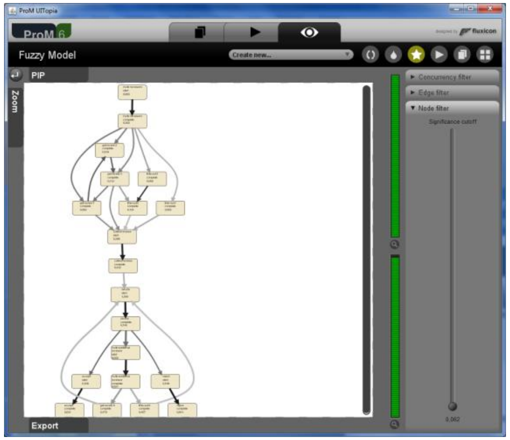
Mine for a Fuzzy Model
H.M.W. Verbeek (h.m.w.verbeek@tue.nl)
http://www.processmining.org

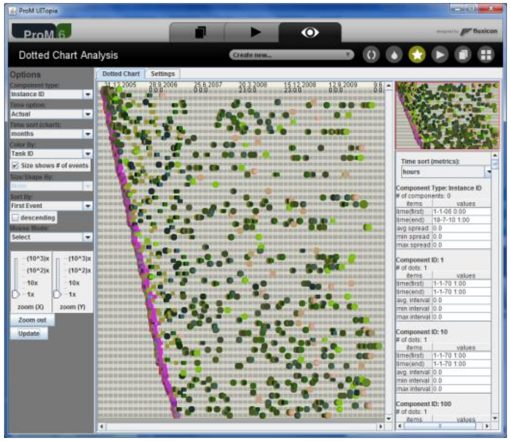
Mine for a Handover-of-Work Social Network
M. Song (m.song@unist.ac.kr)
http://www.processmining.org

Mine for a Petri Net using ILP
T. van der Wiel (t.v.d.wiel@student.tue.nl)
http://www.processmining.org

Mine for a Reassignment Social Network
M. Song (m.song@unist.ac.kr)
http://www.processmining.org

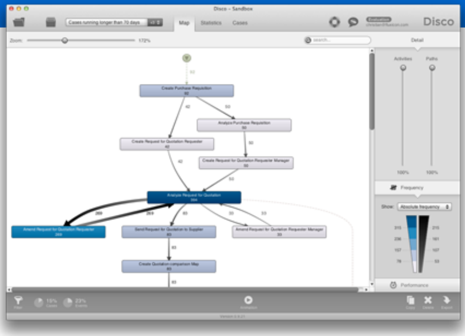
Mine for a Similar-Task Social Network
M. Song (m.song@unist.ac.kr)
http://www.processmining.org

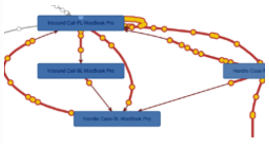




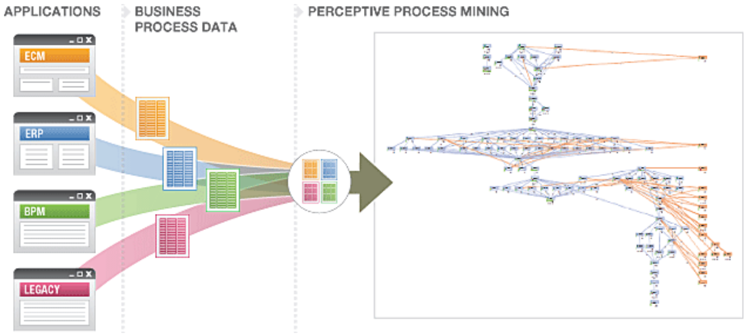
Download from: www.processmining.org

Commercial Alternatives



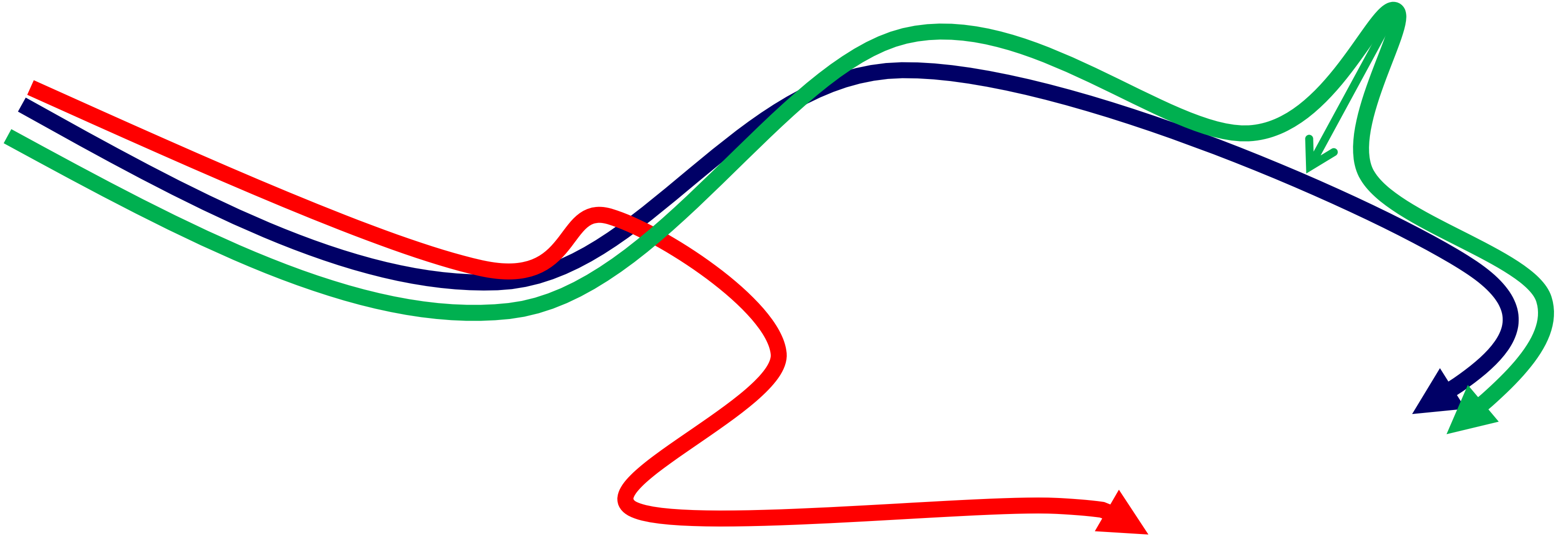


- Disco (Fluxicon)
- Perceptive Process Mining (before Futura Reflect and BPM|one)
- ARIS Process Performance Manager
- QPR ProcessAnalyzer
- Interstage Process Discovery (Fujitsu)
- Discovery Analyst (StereoLOGIC)
- XMAnalyzer (XMPro)
- …



Three Key Observations

#1 Alignments are essential!



- **conformance checking to diagnose deviations**
- **squeezing reality into the model to do model-based analysis**

a	c	$\gg$	d	$\gg$	f	$\gg$
a $t1$	c $t4$	b $t3$	d $t5$	τ $t7$	$\gg$	h $t10$

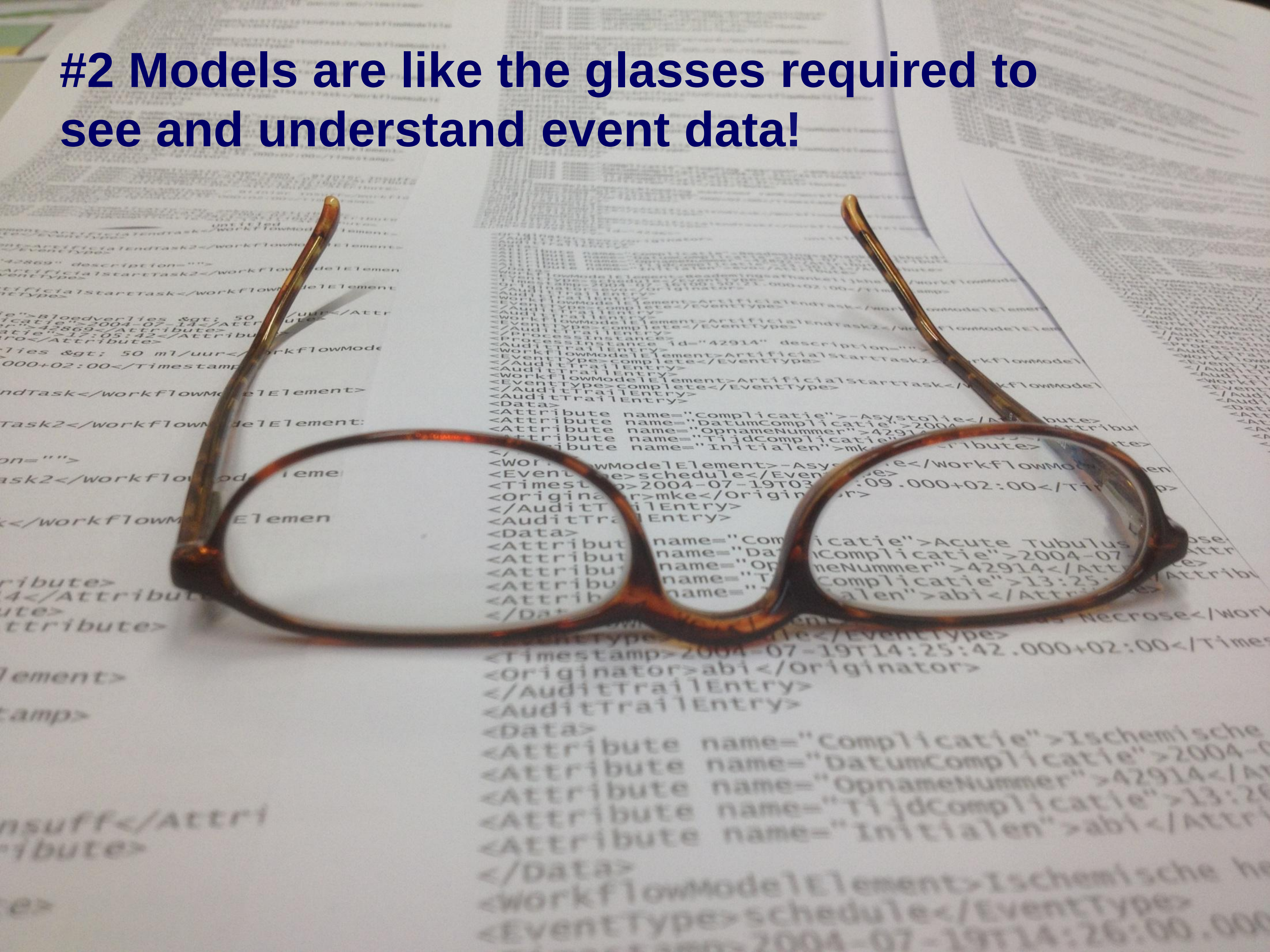
move on log

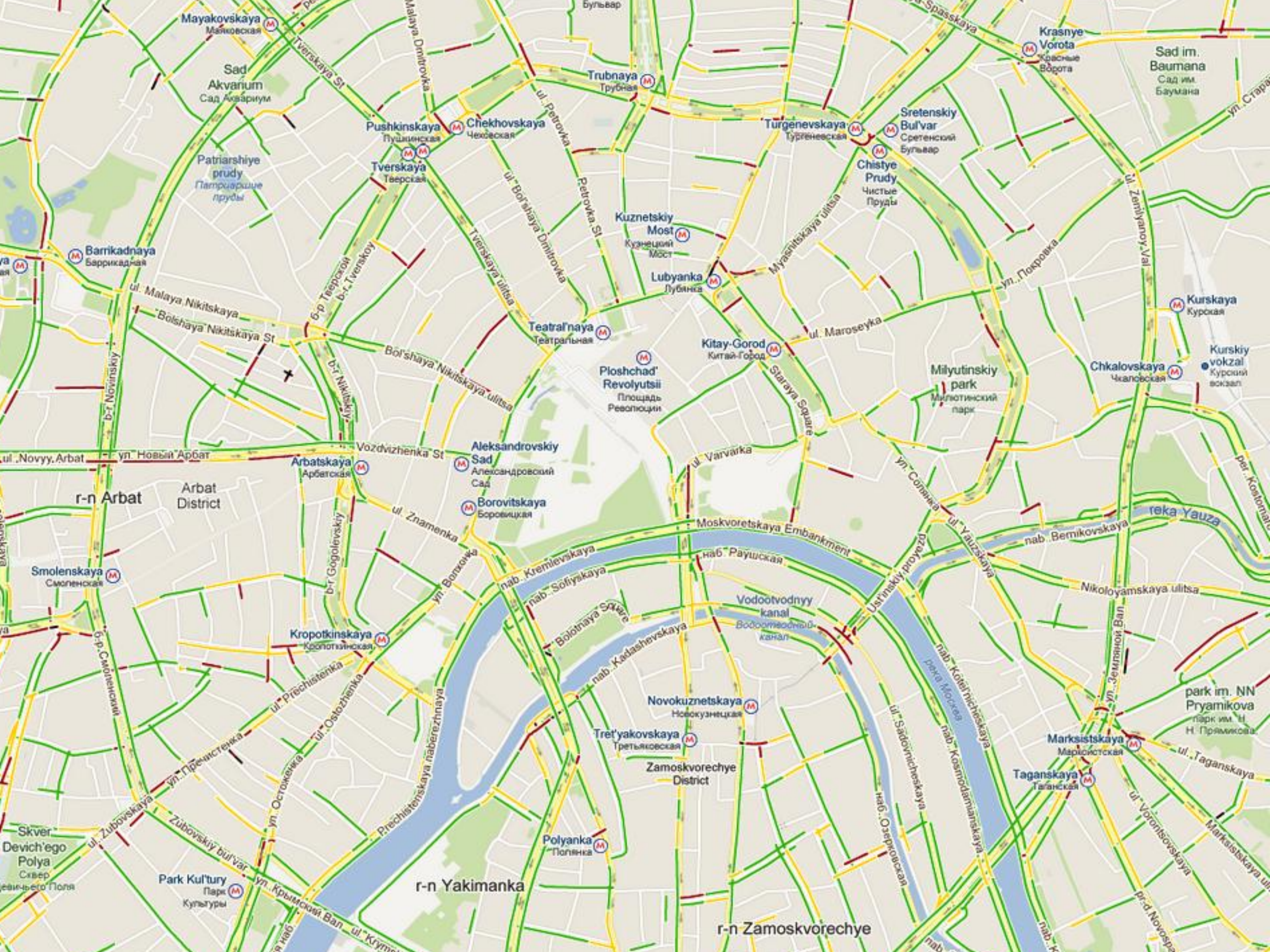
move on model

move on model

move on model (harmless)

#2 Models are like the glasses required to see and understand event data!


<Timestamp>2004-07-19T14:25:42.000+02:00</Times
<Originator>abi</Originator>
</AuditTrailEntry>
<AuditTrailEntry>
<Data>
<Attribute name="Complicatie">Ischemische
<Attribute name="DatumComplicatie">2004-0
<Attribute name="OpnameNummer">42914</A
<Attribute name="TijdComplicatie">13:2
<Attribute name="Initialen">abi</Attr
</Data>
<WorkflowModelElement>Ischemische he
<EventType>schedule</EventType>



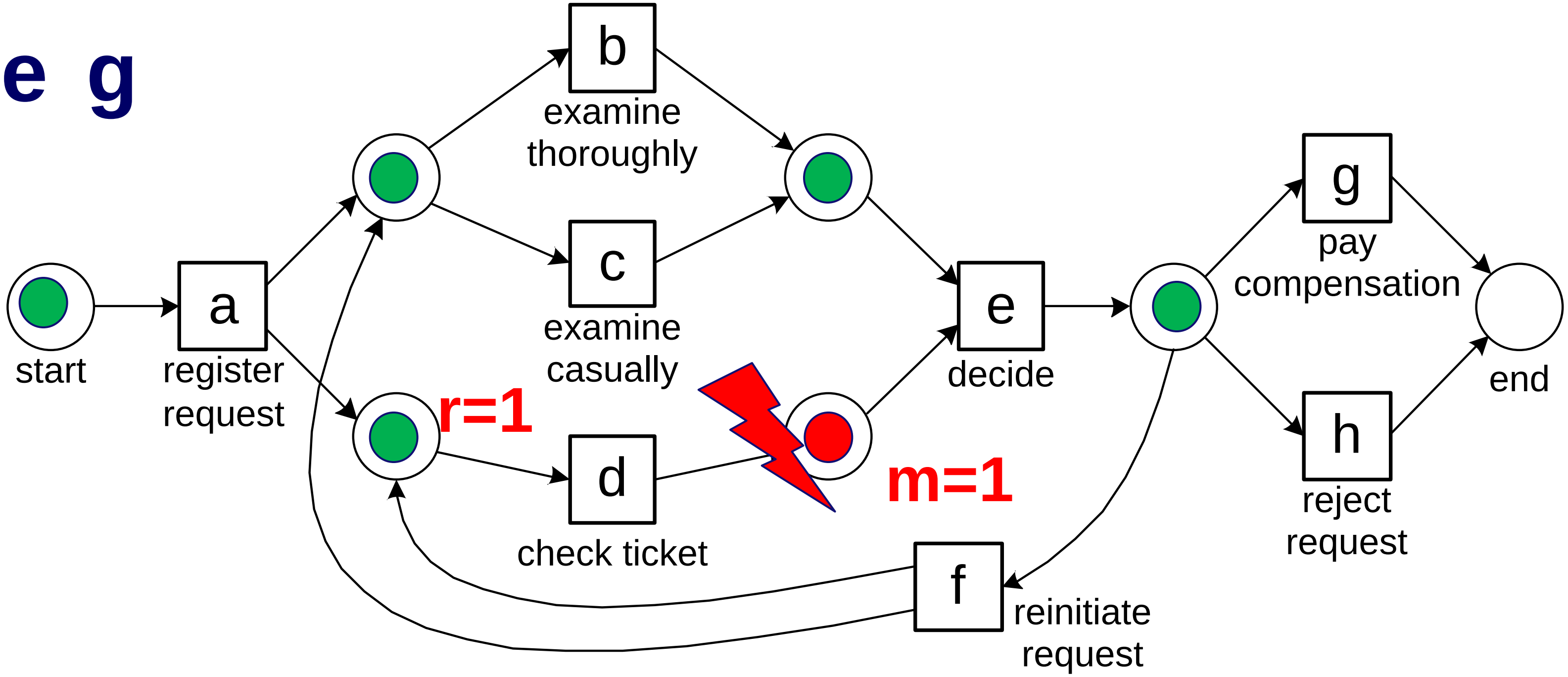
Бульвар
Mayakovskaya
Маяковская
Malaya Dmitrovka
Tverskaya St
Sad Akvarium
Сад Аквариум
Trubnaya
Трубная
ul. Petrovka
Krasnye Vorota
Красные Ворота
Sad im. Baumana
Сад им. Баумана
ul. Staraya
Pushkinskaya
Пушкинская
Chekhovskaya
Чеховская
Tverskaya
Тверская
Turgenevskaya
Тургеневская
Sretenskiy Bul'var
Сретенский Бульвар
Chistye Prudy
Чистые Пруды
Patriarshiye prudy
Патриаршие пруды
ul. Bol'shaya Dmitrovka
Petrovka St
Tverskaya ulitsa
Kuznetskiy Most
Кузнецкий Мост
Myasnitskaya ulitsa
ul. Zemlyanoy Val
Barrikadnaya
Баррикадная
б-р Тверской
b-r Tverskoy
Lubyanka
Лубянка
ul. Pokrovka
Kurskaya
Курская
ul. Malaya Nikitskaya
Bolshaya Nikitskaya St
Teatral'naya
Театральная
Kitay-Gorod
Китай-Город
ul. Maroseyka
Kurskiy vokzal
Курский вокзал
Chkalovskaya
Чкаловская
Bol'shaya Nikitskaya ulitsa
b-r Nikitskiy
Ploshchad' Revolyutsii
Площадь Революции
Staraya Square
Milyutinskiy park
Милютинский парк
b-r Novinskiy
ul. Novyy Arbat
ул. Новый Арбат
Vozdvizhenka St
Arbatskaya
Арбатская
Aleksandrovskiy Sad
Александровский Сад
ul. Varvarka
r-n Arbat
Arbat District
Borovitskaya
Боровицкая
ul. Solyanka
per. Kostomarovsky
ul. Znamenka
b-r Gogolevskiy
Moskvoretskaya Embankment
reka Yauza
ul. Yauzskaya
nab. Bernikovskaya
ul. Volkhonka
nab. Kremlevskaya
наб. Раушская
Ust'inskiy proyezd
Smolenskaya
Смоленская
nab. Sofiyskaya
Nikoloyamskaya ulitsa
Vodootvodnyy kanal
Водоотводный канал
Bolotnaya Square
Kropotkinskaya
Кропоткинская
nab. Kadashevskaya
б-р Смоленский
ul. Prechistenka
ul. Ostozhenka
Prechistenskaya naberezhnaya
nab. Kotel'nicheskaya
река Москва
ul. Sadovnicheskaya
nab. Kosmodamianskaya
ul. Zemlyanoy Val
Novokuznetskaya
Новокузнецкая
park im. NN Pryamikova
парк им. Н. Н. Прямикова
Tret'yakovskaya
Третьяковская
ул. Пречистенка
Marksistskaya
Марксистская
ul. Taganskaya
Zamoskvorechye District
Taganskaya
Таганская
ул. Остоженка
наб. Озерковская
ul. Vorontsovskaya
Marksistskaya
Skver Devich'ego Polya
Сквер Девичьего Поля
ul. Zubovskaya
Zubovskiy bul'var
Polyanka
Полянка
Park Kul'tury
Парк Культуры
ul. Krymskiy Val
ul. Krymskiy
r-n Yakimanka
r-n Zamoskvorechye
pr-d Novospassky
nab.

Alignments

Joint work with Arya Adriansyah, Boudewijn van Dongen, Elham Ramezani, Dirk Fahland, Massimiliano de Leoni, et al.

Replaying trace “abeg”

a b e g



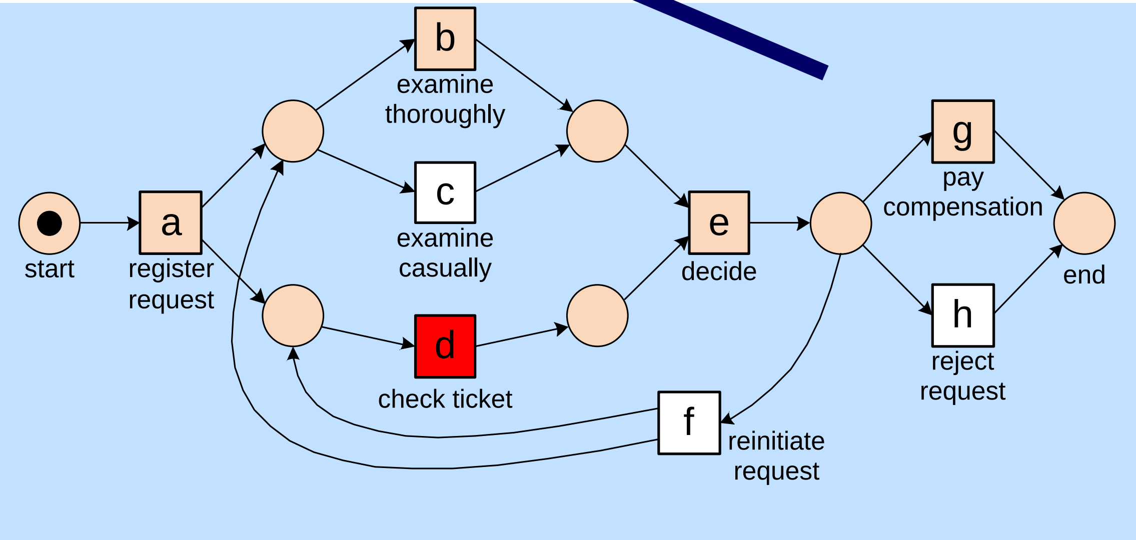
$$fitness(\sigma, N) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.83333$$

From "playing the token game" to optimal alignments …

observed trace: "abeg"

a	b	»	e	g
a	b	d	e	g

move in model only

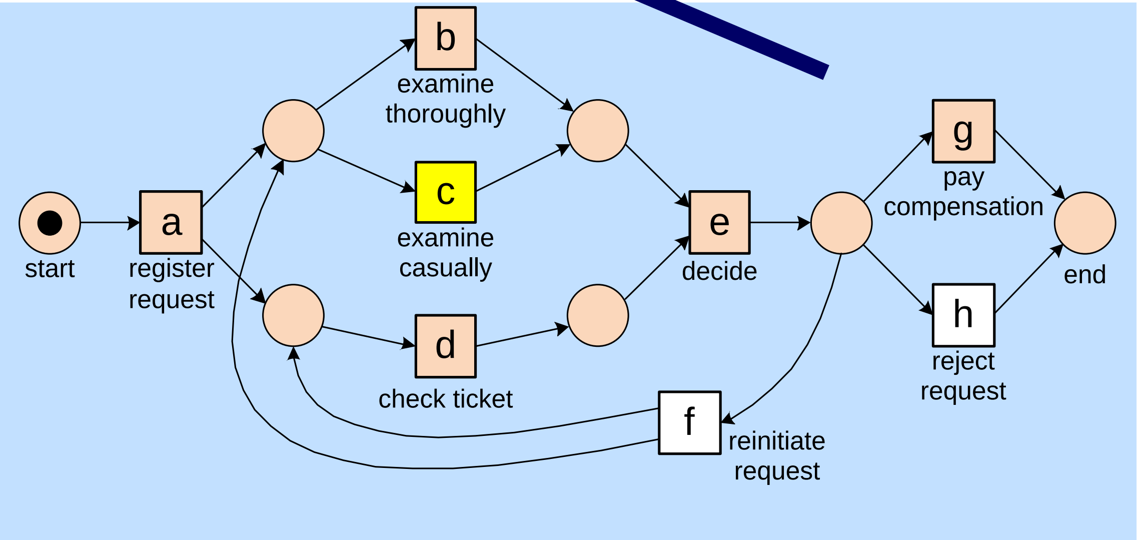


Another alignment

observed trace: “abcdeg”

a	b	c	d	e	g
a	b	»	d	e	g

move in log only



Moves have costs

...	a	...
...	»	...

...	»	...
...	a	...

...	a	...
...	a	...

...	a	...
...	b	...

- Standard cost function:
  - $c(x,\text{»}) = 1$
  - $c(\text{»},y) = 1$
  - $c(x,y) = 0$, if $x=y$
  - $c(x,y) = \infty$, if $x \neq y$

Any cost structure is possible

...	send-letter(John,2 weeks, $400)	...
...	send-email(Sue,3 weeks,$500)	...

- Similar activities (more similarity implies lower costs).
- Resource conformance (done by someone that does not have the specified role).
- Data conformance (path is not possible for this customer).
- Time conformance (missed the legal deadline)

Using Alignments

- An optimal alignment has the lowest possible costs.
- If multiple alignments are optimal, pick one or use all.
- Like an "oracle" revealing paths in the model.
- These paths can be used for further analysis!

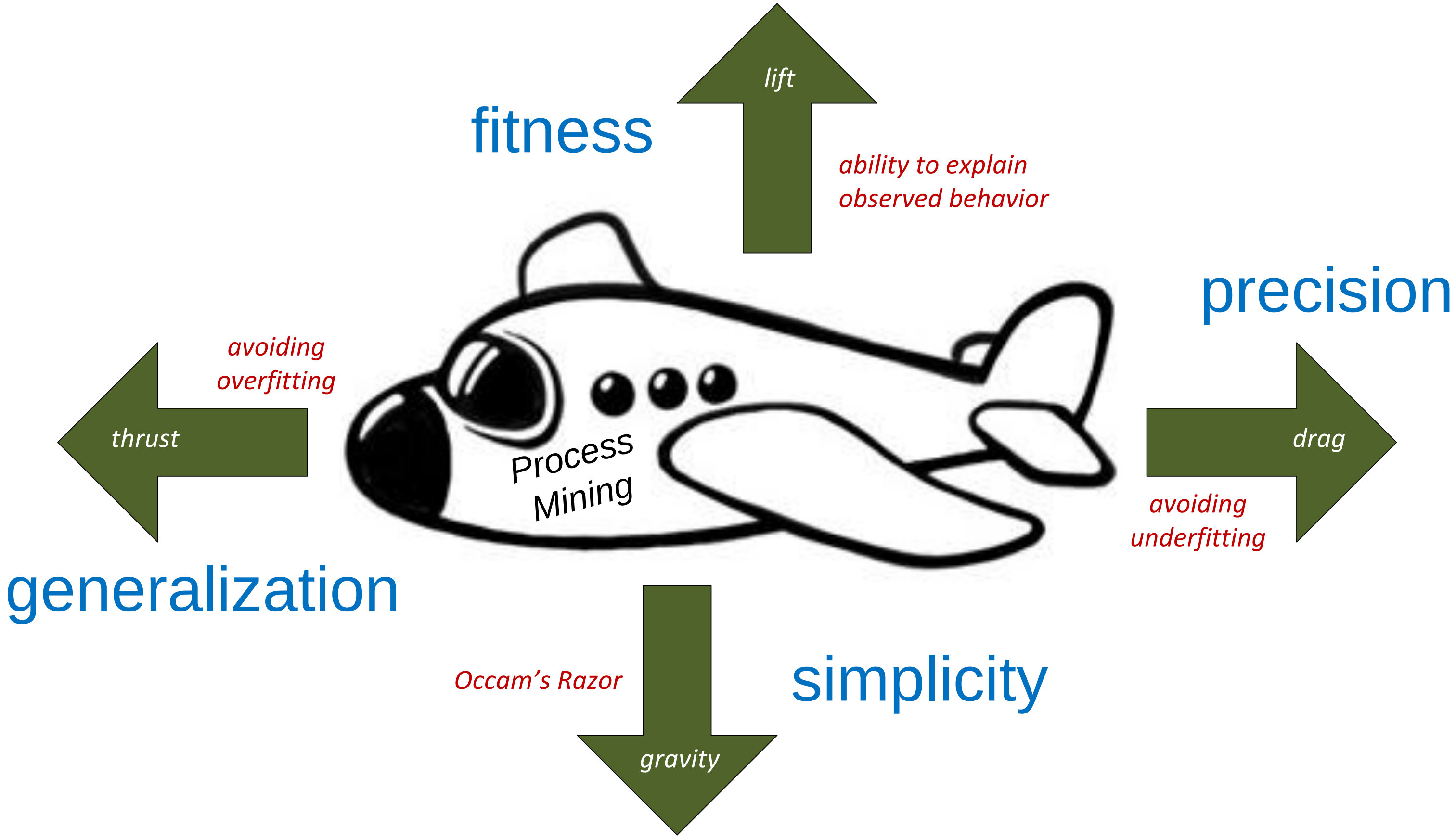
- Can be used for quantifying various types conformance (possibly using a different cost function).

Some pointers

- Wil M. P. van der Aalst, Arya Adriansyah, Boudewijn F. van Dongen: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery 2(2): 182-192 (2012)
- Arya Adriansyah, Jorge Munoz-Gama, Josep Carmona, Boudewijn F. van Dongen, Wil M. P. van der Aalst: Alignment Based Precision Checking. Business Process Management Workshops 2012: 137-149
- Arya Adriansyah, Boudewijn F. van Dongen, Wil M. P. van der Aalst: Conformance Checking Using Cost-Based Fitness Analysis. EDOC 2011: 55-64
- Massimiliano de Leoni, Wil M. P. van der Aalst, Boudewijn F. van Dongen: Data- and Resource-Aware Conformance Checking of Business Processes. BIS 2012: 48-59
- Joos C. A. M. Buijs, Boudewijn F. van Dongen, Wil M. P. van der Aalst: On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. OTM Conferences (1) 2012: 305-322
- Elham Ramezani, Dirk Fahland, Wil M. P. van der Aalst: Where Did I Misbehave? Diagnostic Information in Compliance Checking. BPM 2012: 262-278
- A. Adriansyah, B.F. van Dongen, W.M.P. van der Aalst. Memory-Efficient Alignment of Observed and Modeled Behavior. BPM Center Report BPM-13-03, BPMcenter.org, 2013

Conformance: Taking a Step Back

Conventional Conformance Notions



fitness

ability to explain observed behavior

precision

avoiding underfitting

generalization

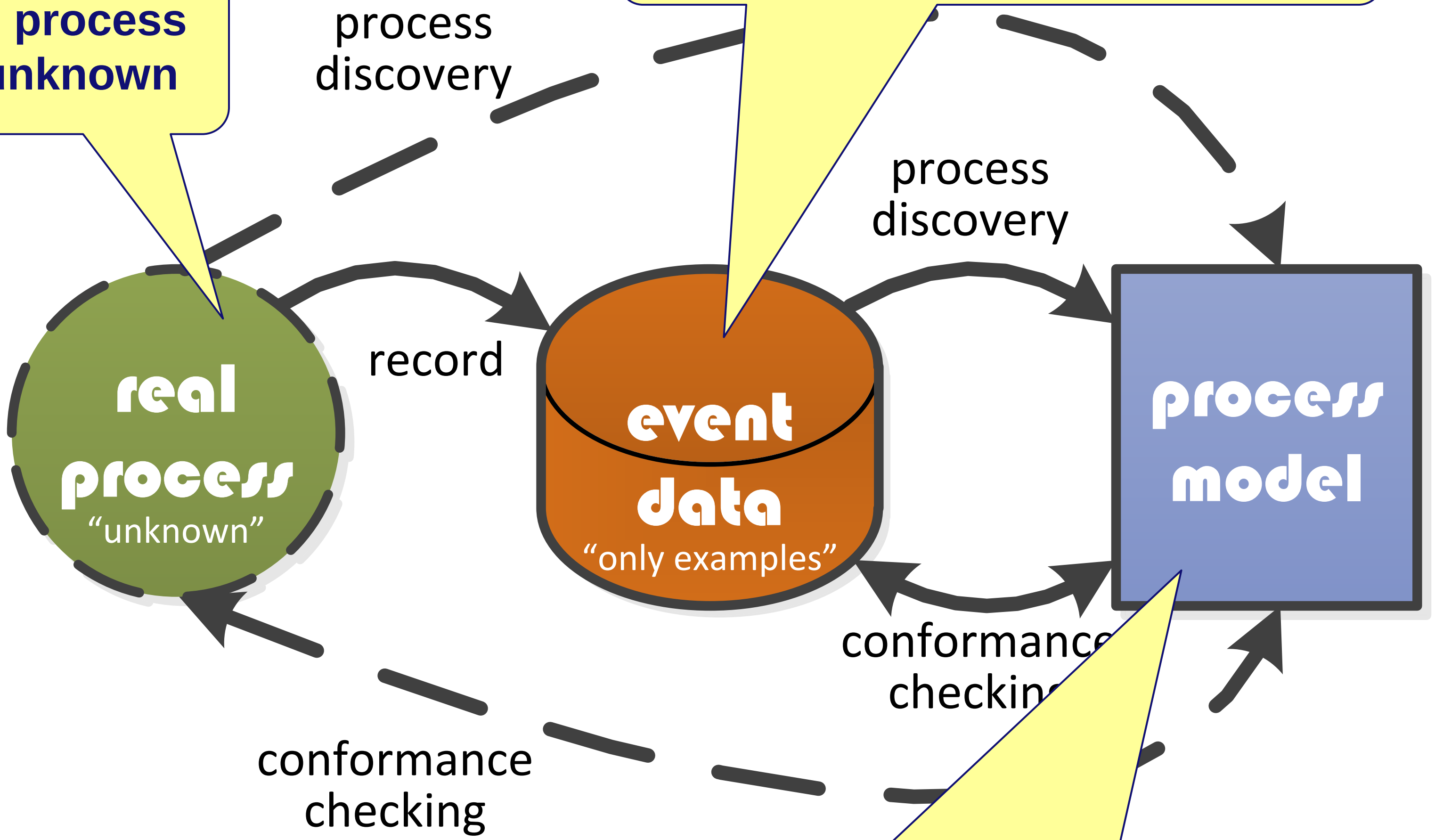
avoiding overfitting

simplicity

Occam's Razor

Leaving out one of these dimensions during discovery will lead to degenerate cases!

Problem

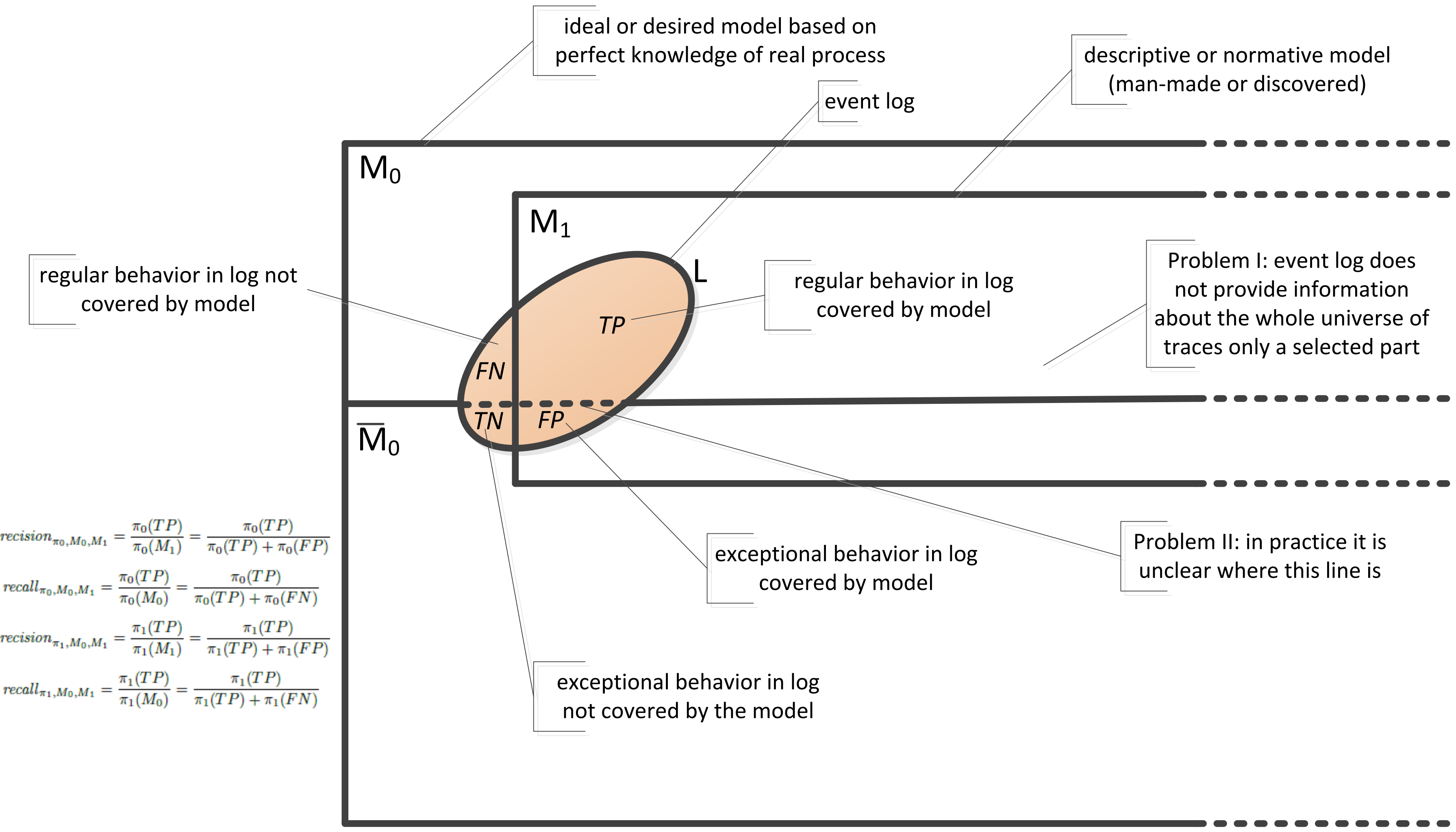


real process
is unknown

event logs covers only a fraction
of all possible behavior

model needs to provide an abstraction:
Murphy's Law of Process Mining

Traditional Notions such as Precision and Recall do NOT Apply


ideal or desired model based on perfect knowledge of real process
event log
descriptive or normative model (man-made or discovered)
M_0
M_1
L
regular behavior in log not covered by model
TP
regular behavior in log covered by model
Problem I: event log does not provide information about the whole universe of traces only a selected part
FN
TN
FP
$\overline{M}_0$
exceptional behavior in log covered by model
Problem II: in practice it is unclear where this line is
exceptional behavior in log not covered by the model

$$precision_{\pi_0,M_0,M_1} = \frac{\pi_0(TP)}{\pi_0(M_1)} = \frac{\pi_0(TP)}{\pi_0(TP)+\pi_0(FP)}$$

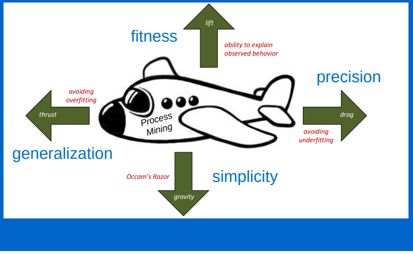
$$recall_{\pi_0,M_0,M_1} = \frac{\pi_0(TP)}{\pi_0(M_0)} = \frac{\pi_0(TP)}{\pi_0(TP)+\pi_0(FN)}$$

$$precision_{\pi_1,M_0,M_1} = \frac{\pi_1(TP)}{\pi_1(M_1)} = \frac{\pi_1(TP)}{\pi_1(TP)+\pi_1(FP)}$$

$$recall_{\pi_1,M_0,M_1} = \frac{\pi_1(TP)}{\pi_1(M_0)} = \frac{\pi_1(TP)}{\pi_1(TP)+\pi_1(FN)}$$

$$precision_{\pi_1,M_L,M_1} = \frac{\pi_1(TP_L)}{\pi_1(M_1)} = \pi_1(TP_L) = \pi_1(M_L) \qquad recall_{\pi_L,M_L,M_1} = \frac{\pi_L(TP_L)}{\pi_L(M_L)} = \pi_L(TP_L) = \pi_L(M_1)$$

Operationalizing the Four Conformance dimensions



- **Fitness (fraction of observed behavior possible according to the model).**
  - **Measure at the case or event level.**
  - **How to continue after a deviation (where to put the "blame"), cf. duplicate and silent activities.**
- **Precision (avoiding underfitting; fraction of allowed behavior never observed).**
  - **Log only contains examples.**
  - **Metrics are e.g. based on escaping edges.**
- **Generalization (avoiding overfitting; probability that the next unseen case will not fit).**
  - **Reasoning about unseen behavior, strongly related to log completeness.**
- **Simplicity (Occam's Razor: the simplest of two or more competing theories is preferable)**
  - **Easy to operationalize (e.g., number of nodes or arcs).**
  - **Often subjective (valuation of AND/XOR/OR-split/joins).**

Some pointers

- Wil M. P. van der Aalst: Mediating Between Modeled and Observed
- Behavior: The Quest for the “Right” Process. Seventh IEEE International Conference on Research Challenges in Information Science (RCIS 2013), (2013)
- Wil M. P. van der Aalst, Arya Adriansyah, Boudewijn F. van Dongen: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery 2(2): 182-192 (2012)
- Joos C. A. M. Buijs, Boudewijn F. van Dongen, Wil M. P. van der Aalst: On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. OTM Conferences (1) 2012: 305-322
- Wil M. P. van der Aalst: Process Mining - Discovery, Conformance and Enhancement of Business Processes. Springer 2011, isbn 978-3-642-19344-6, pp. I-XVI, 1-352

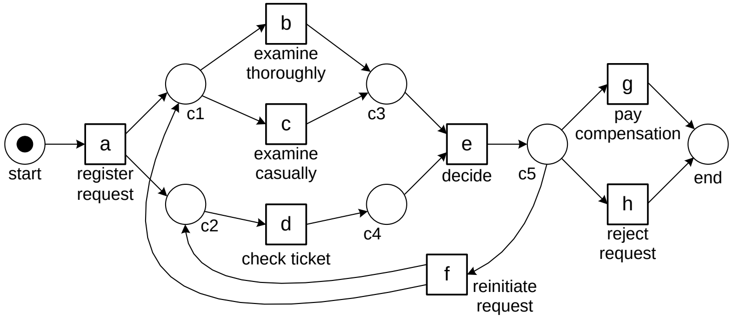
Representational Bias in Process Mining

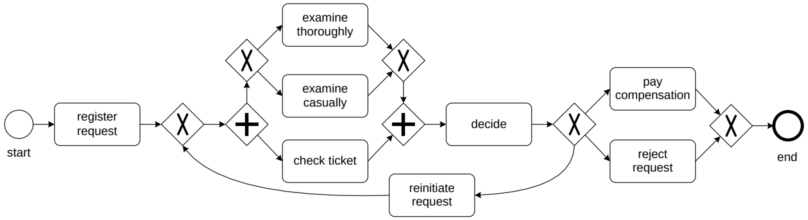
(not about visualization)

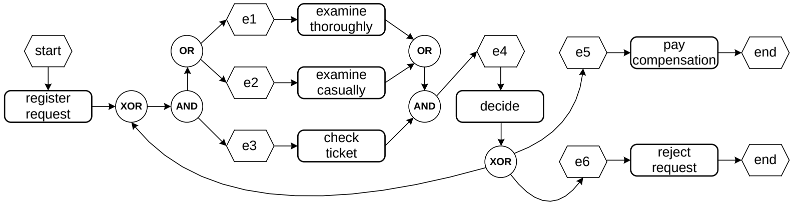
Joint work with Joos Buijs, Sander Leemans, and Boudewijn van Dongen.

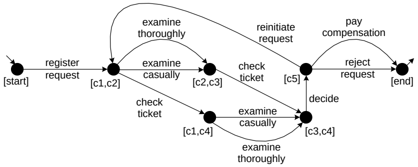
Typical Representational Bias

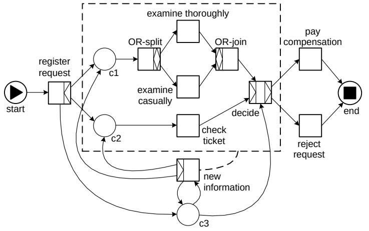
- (Labeled) Petri Nets, WF-nets, etc.
- Subsets of
  - BPMN diagrams,
  - UML Activity Diagrams,
  - Event-Driven Process Chains (EPCs),
  - YAWL,
  - etc.
- Transition Systems
- (Hidden) Markov Models
- …



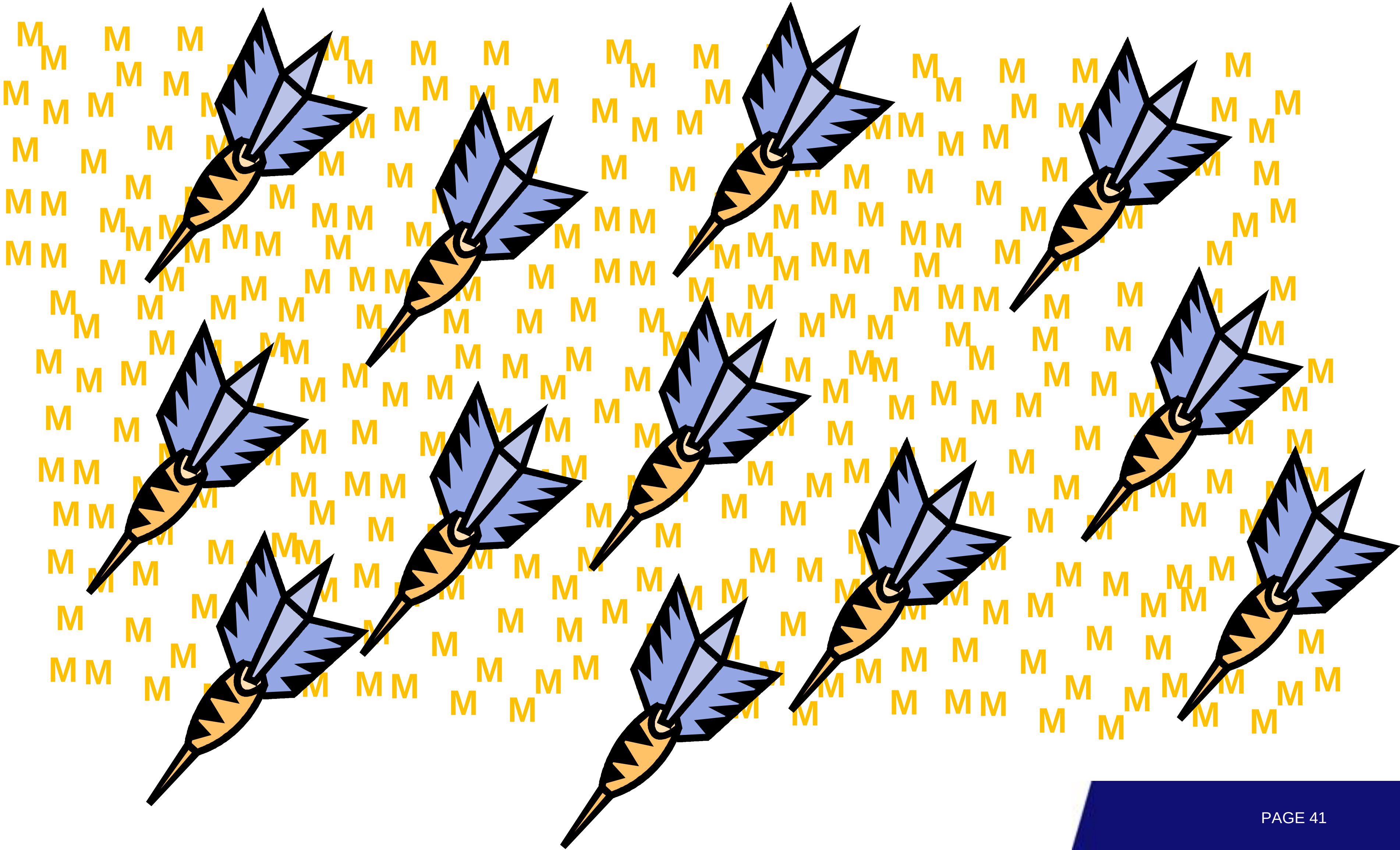




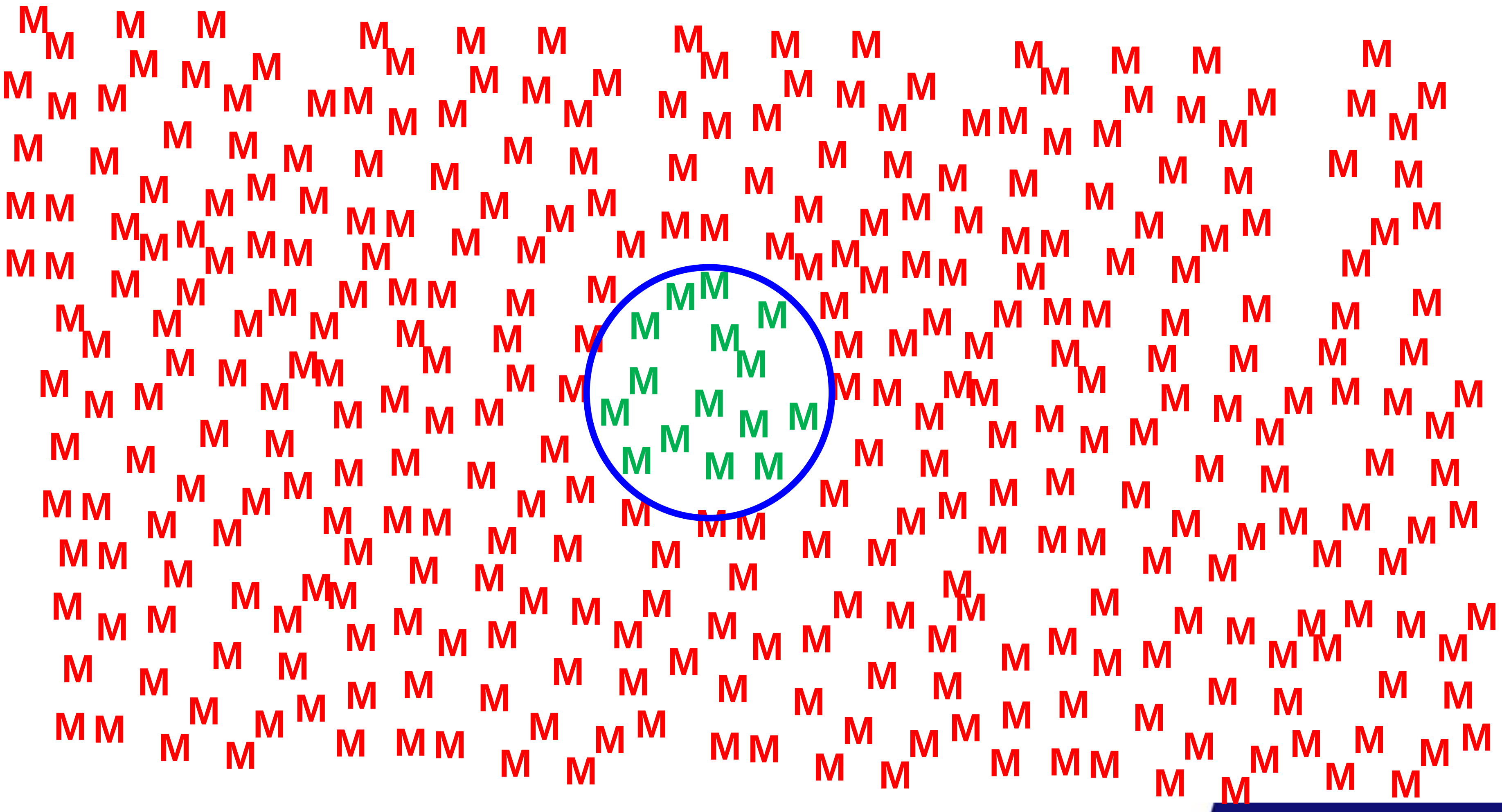




Huge Search Space When Discovering a Petri Net, BPMN model, and the like …

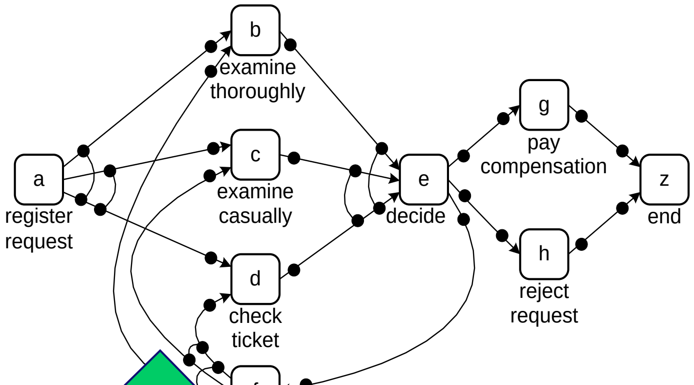


… with just a few interesting candidates

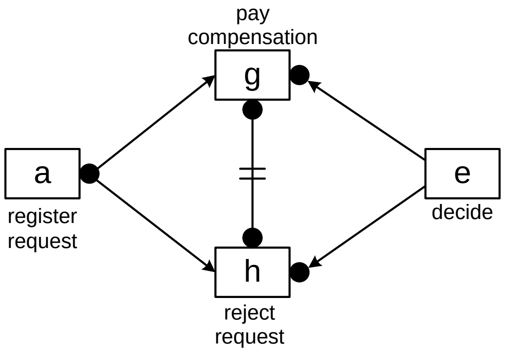


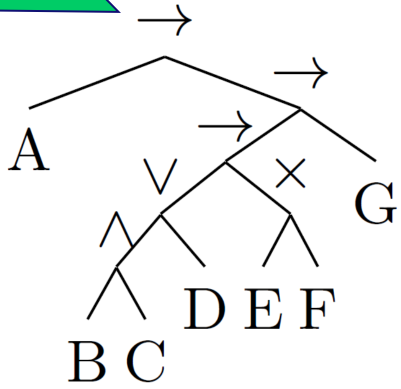
Alternative Representational Bias

1. C-nets (XOR/AND/OR-split/join graphs; more likely to be sound due to declarative semantics).
2. Declare models (constraint based, grounded in LTL; anything is possible unless forbidden)
3. Process Trees (similar to subsets of various process algebras; sound by structure)

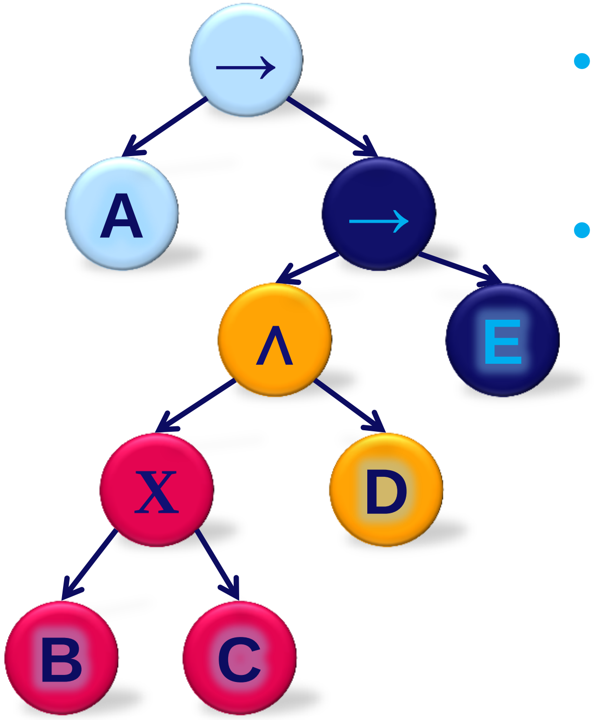


today's focus

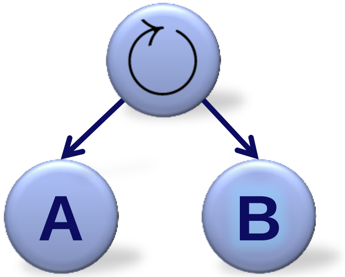




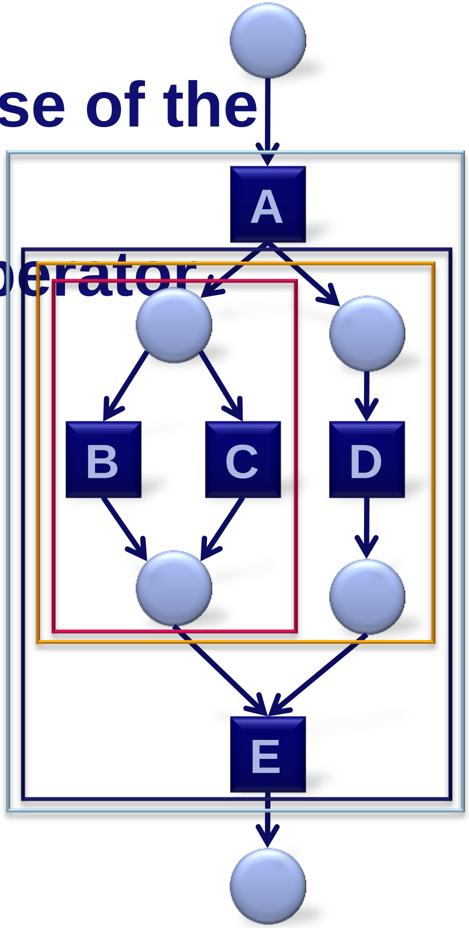
Another Representational Bias: Process Trees



- Always sound because of the block structure
- Also Loop and OR operator

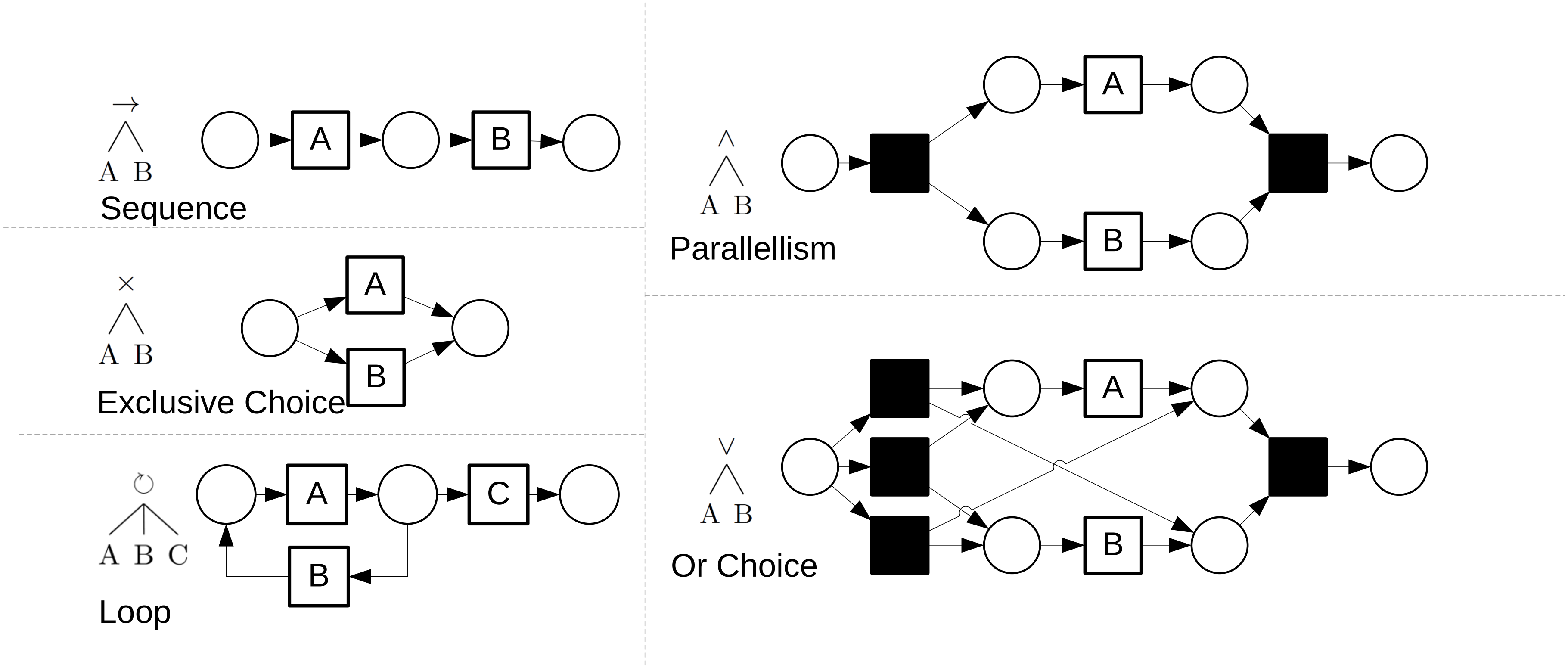


A (BA)*

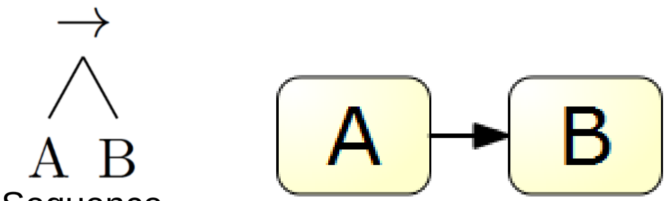


Petri Net Semantics

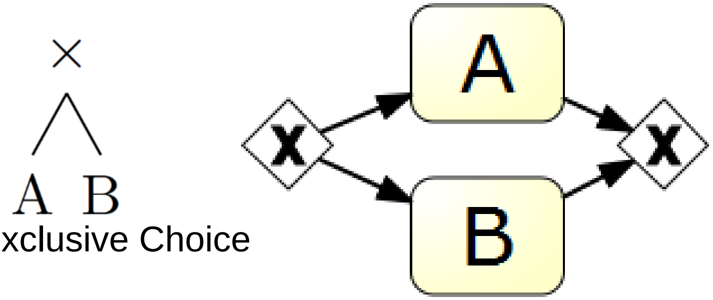
(used for comparison and conformance checking only)



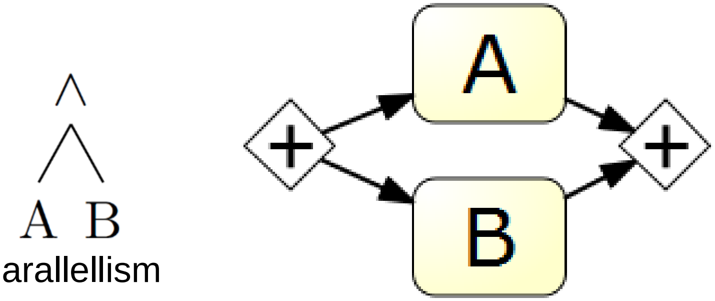
… and BPMN.



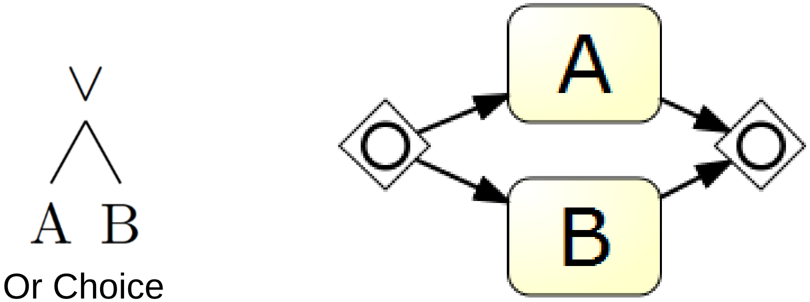
Sequence



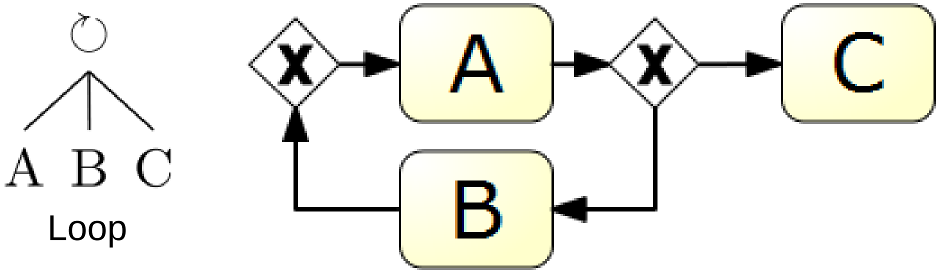
Exclusive Choice



Parallellism



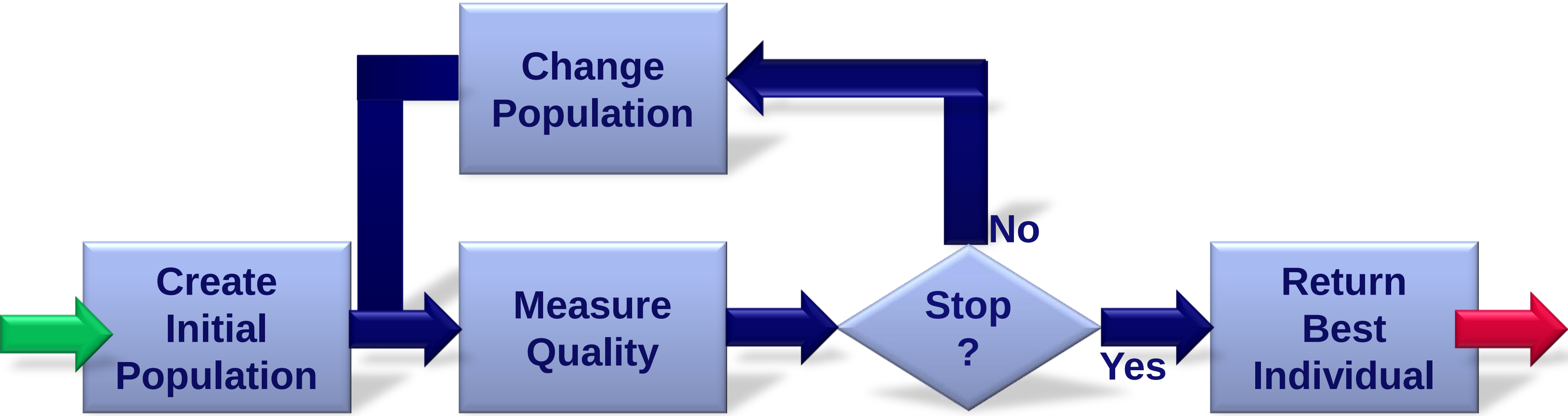
Or Choice



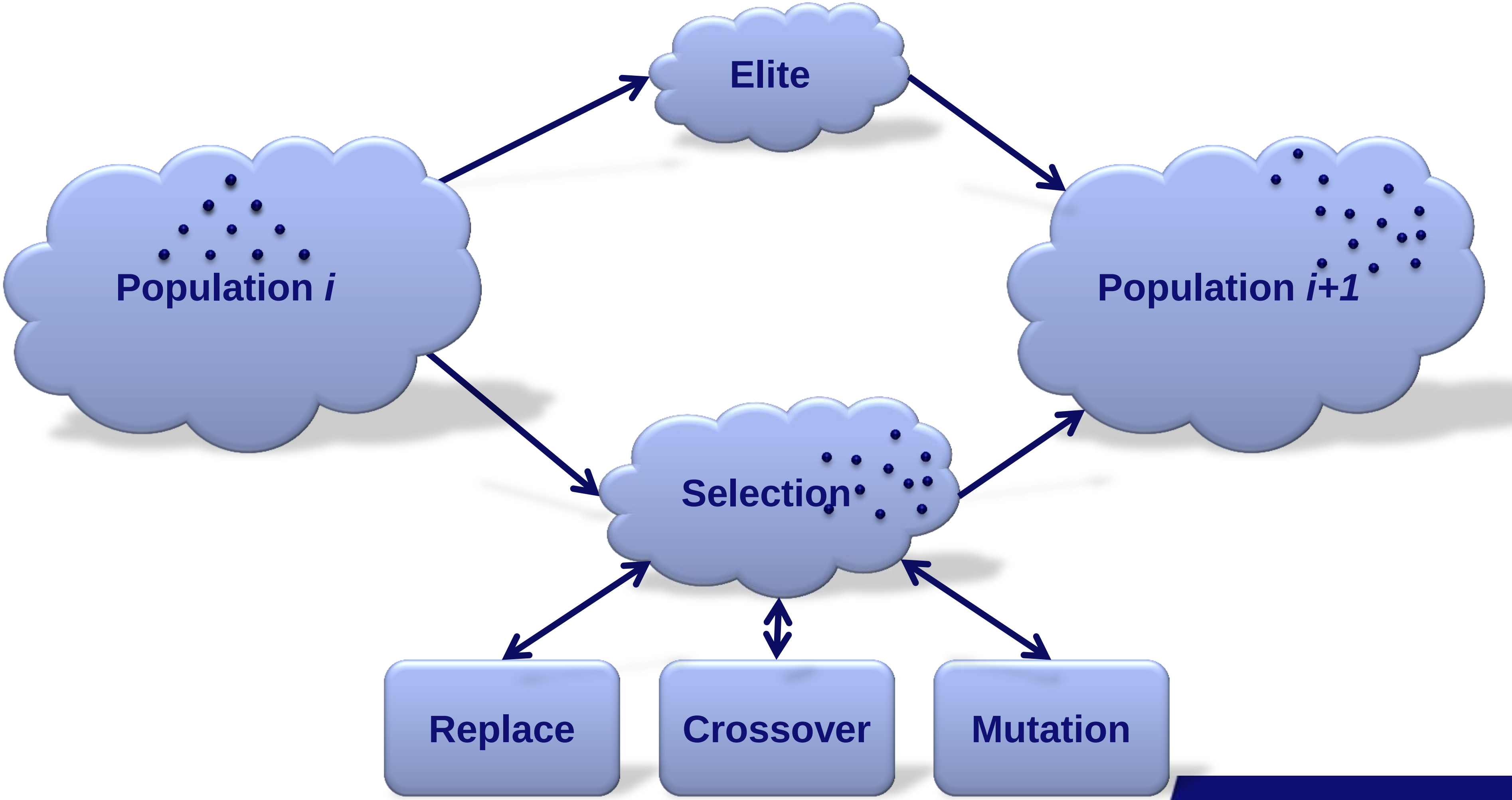
Loop

A Discovery Algorithm Using Process Trees: Evolutionary Tree Miner (ETM)

- **Process trees** as representation (= limit search space to "good" models).
- **Genetic** approach (= very flexible)
- Fitness function uses all **four criteria** (= seamlessly balance the different "forces")

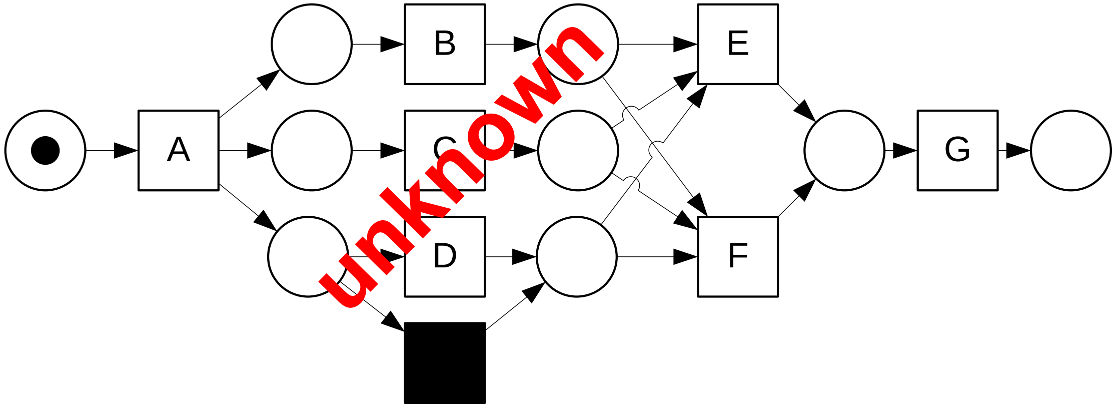


Population Change



Example

Trace	#
A B C D E G	6
A B C D F G	38
A B D C E G	12
A B D C F G	26
A B C F G	8
A C B E G	1
A D B C F G	1
A D B C E G	1
A D C B F G	4
A C D B F G	2
A C B F G	1

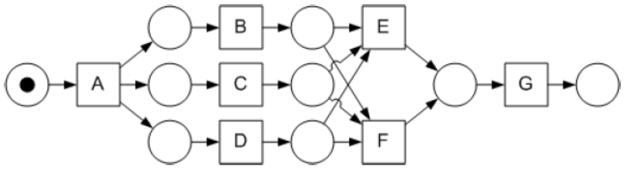


A = send e-mail, B = check credit,
C = calculate capacity, D = check system,
E = accept, F = reject, G = send e-mail

Conventional Algorithms (1/3)

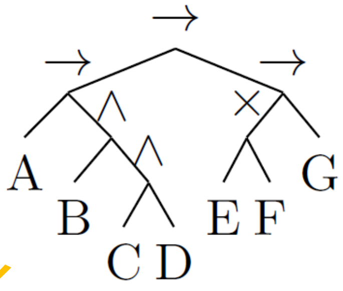
("best effort" mapping to process trees to allow for comparison)

alpha miner



sound

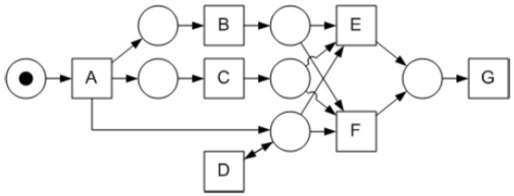
… lucky



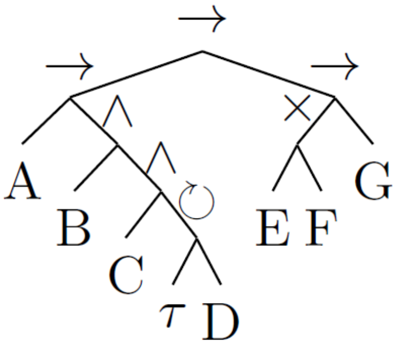
f: 0,992	p: 0,995
s: 1,000	g: 0,889

low fitness

ILP miner



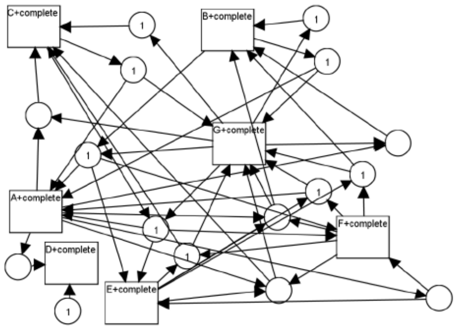
sound



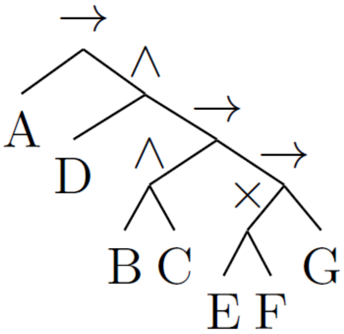
f: 1,000	p: 0,784
s: 0,933	g: 0,830

low precision

language-based region miner



sound

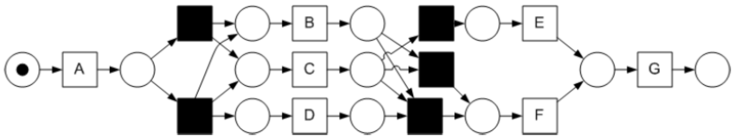


f: 0,992	p: 0,957
s: 1,000	g: 0,889

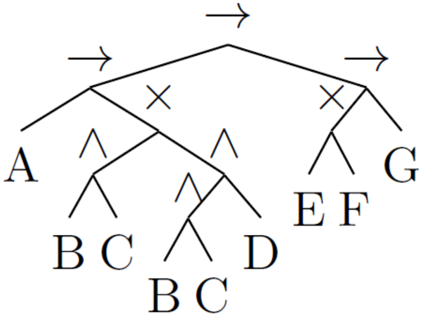
low fitness

Conventional Algorithms (2/3)

heuristic miner

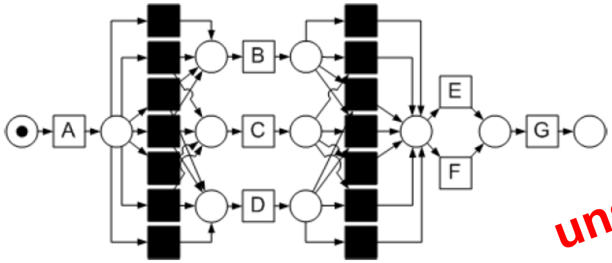


unsound

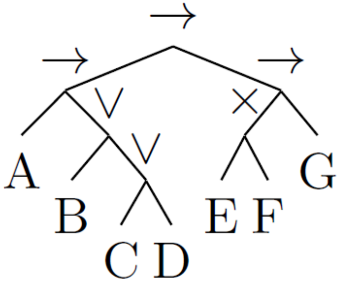


f: 1,000	p: 0,986
s: 0,875	g: 0,852

multi-phase miner



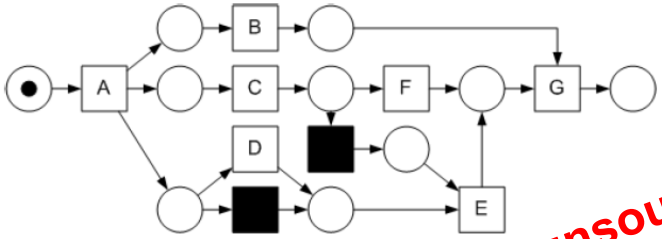
unsound (relaxed sound)



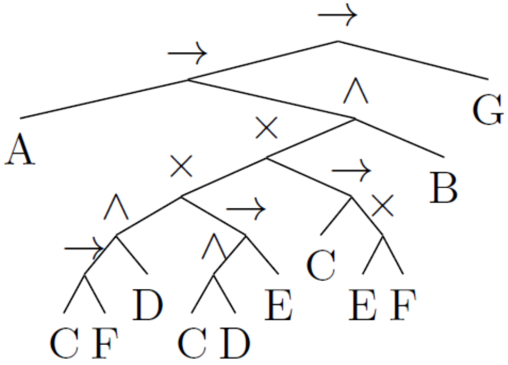
f: 1,000	p: 0,830
s: 1,000	g: 0,889

Conventional Algorithms (3/3)

genetic miner

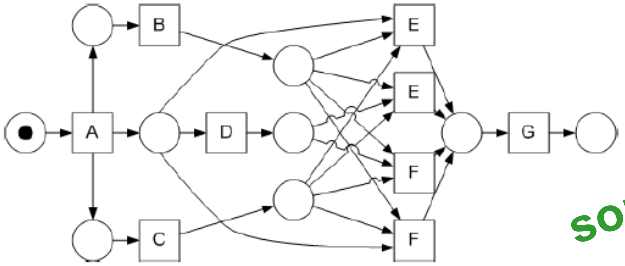


unsound

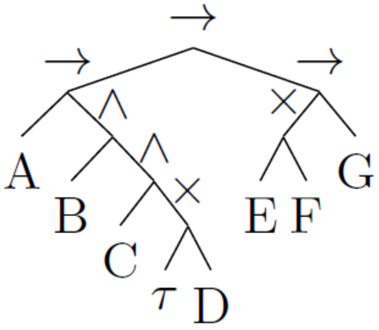


f: 1,000	p: 0,922
s: 0,737	g: 0,790

state-based region miner

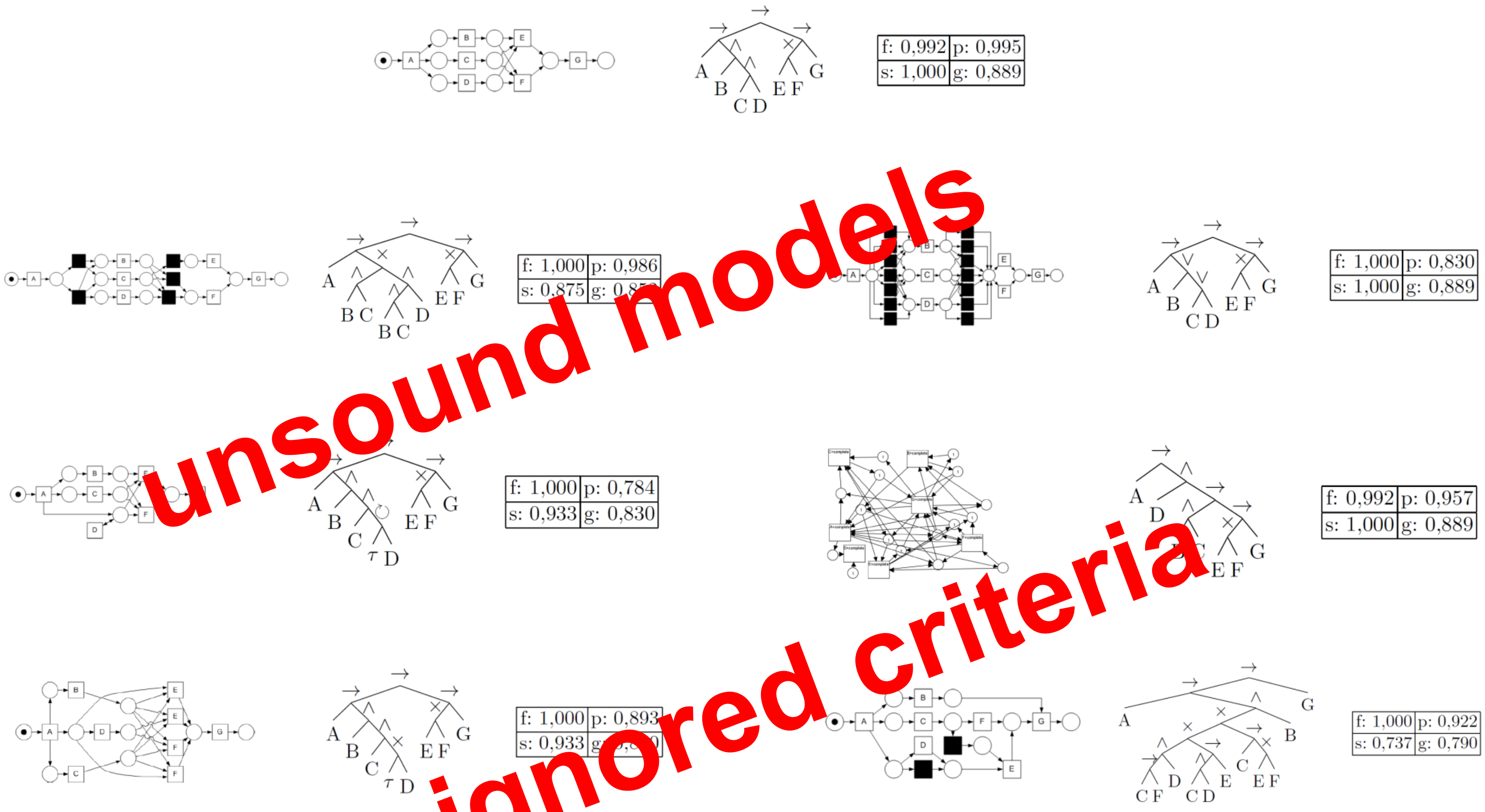


sound



f: 1,000	p: 0,893
s: 0,933	g: 0,830

Often unsound result and no mechanism to seamlessly balance the four criteria



f: 0,992	p: 0,995
s: 1,000	g: 0,889

f: 1,000	p: 0,986
s: 0,875	g: 0,85[illegible]

f: 1,000	p: 0,830
s: 1,000	g: 0,889

f: 1,000	p: 0,784
s: 0,933	g: 0,830

f: 0,992	p: 0,957
s: 1,000	g: 0,889

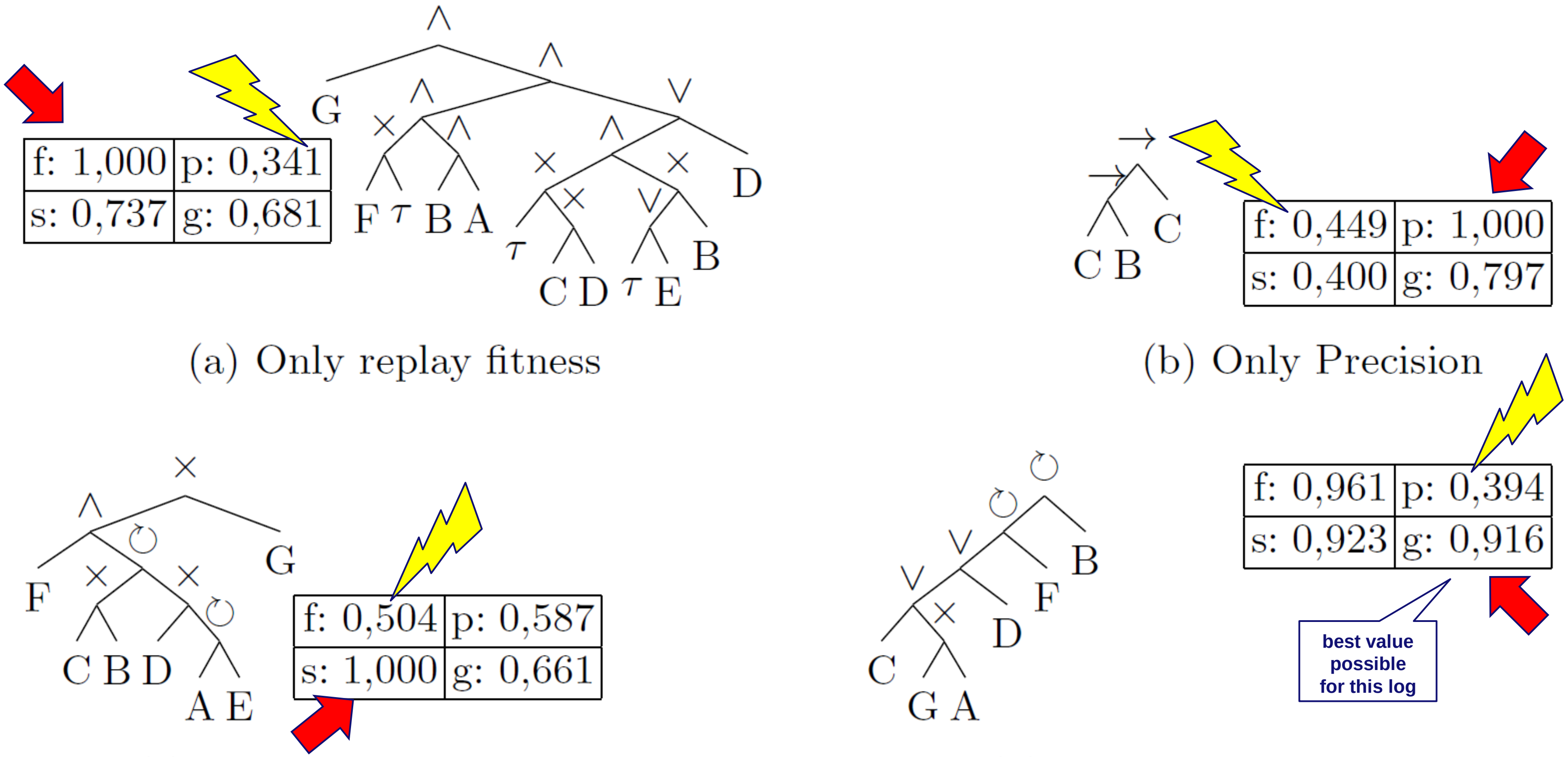
f: 1,000	p: 0,893
s: 0,933	g: [illegible]

f: 1,000	p: 0,922
s: 0,737	g: 0,790

unsound models

ignored criteria

Genetic Mining (ETM) While Considering Only One Criterion



f: 1,000	p: 0,341
s: 0,737	g: 0,681

(a) Only replay fitness

f: 0,449	p: 1,000
s: 0,400	g: 0,797

(b) Only Precision

f: 0,504	p: 0,587
s: 1,000	g: 0,661

(c) Only Simplicity

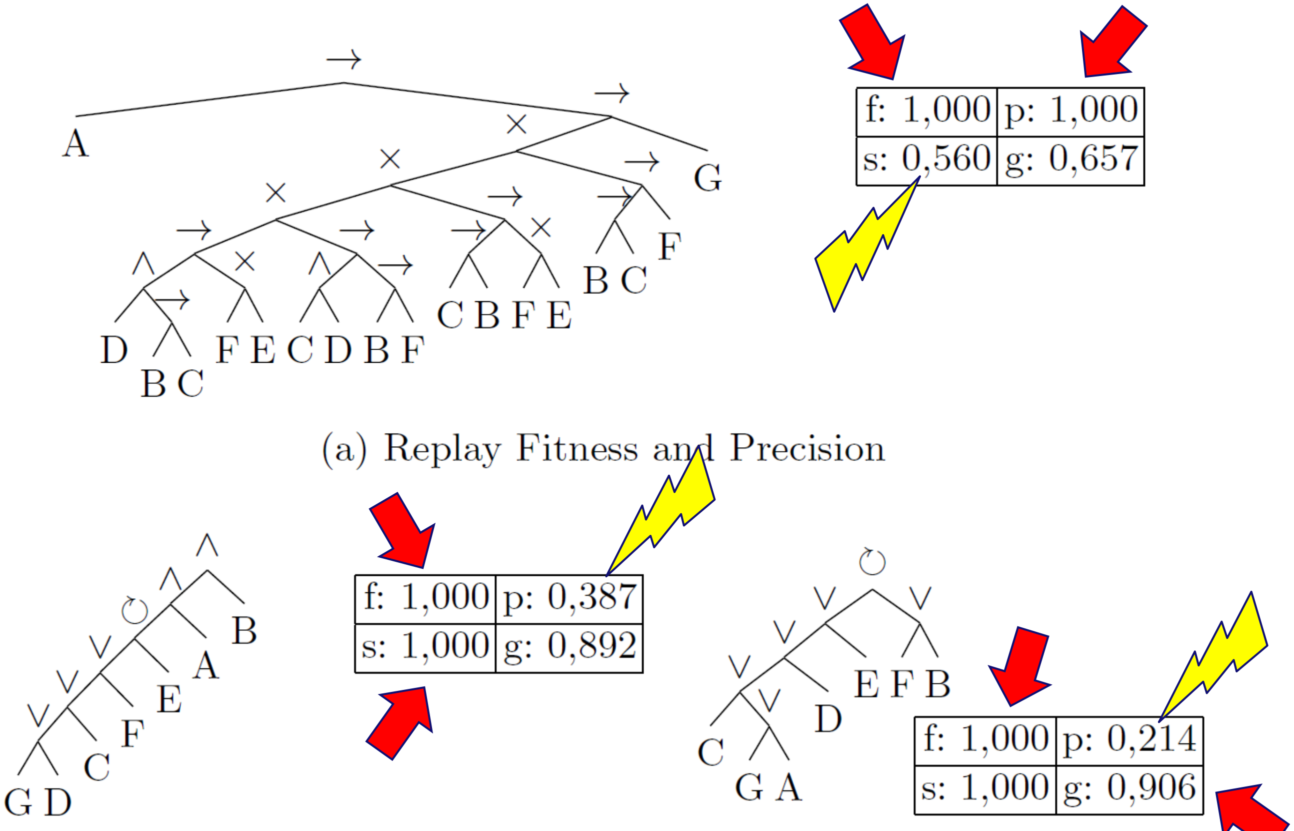
f: 0,961	p: 0,394
s: 0,923	g: 0,916

best value possible for this log

(d) Only Generalization

ETM with weight zero to three out of four perspectives.

Considering Replay Fitness and One Other Criterion



f: 1,000	p: 1,000
s: 0,560	g: 0,657

(a) Replay Fitness and Precision

f: 1,000	p: 0,387
s: 1,000	g: 0,892

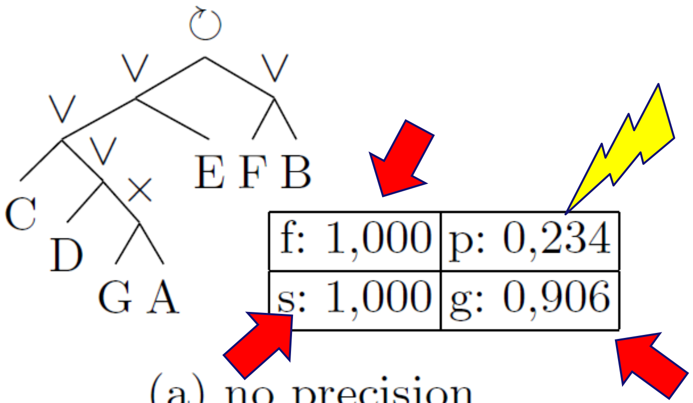
(b) Replay Fitness and Simplicity

f: 1,000	p: 0,214
s: 1,000	g: 0,906

(c) Replay Fitness and Generalization

ETM with weight zero to two out of four perspectives.

Considering 3 of 4 Criteria



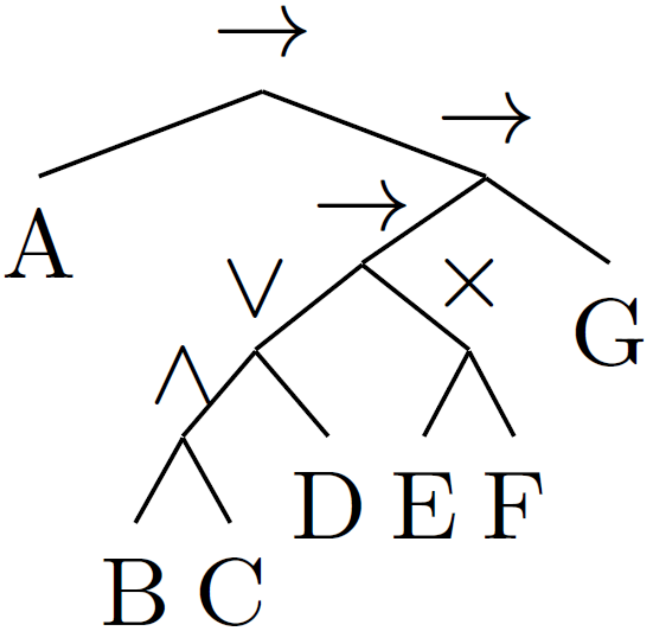
f: 1,000	p: 0,234
s: 1,000	g: 0,906

(a) no precision

replay fitness needs to have a larger weight

Considering All Four Criteria with Emphasis on Fitness

f: 1,000	p: 0,923
s: 1,000	g: 0,889



fitness has weight 10

Initial Model Versus Discovered Model

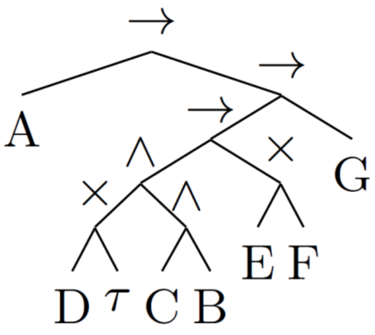
Better than existing algorithms (but patience is needed)!

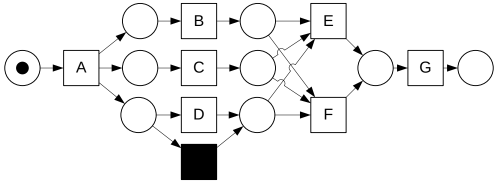
Trace	#
A B C D E G	6
A B C D F G	38
A B D C E G	12
A B D C F G	26
A B C F G	8
A C B E G	1
A D B C F G	1
A D B C E G	1
A D C B F G	4
A C D B F G	2
A C B F G	1

Discovered model outperforms initial model with respect too all criteria!

simulated

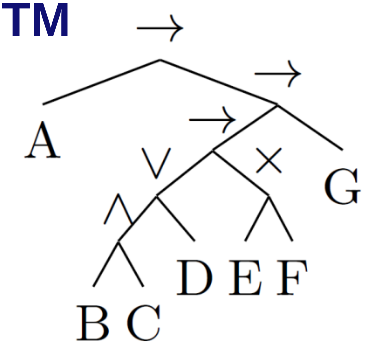
f: 1,000	p: 0,893
s: 0,933	g: 0,830





discovered by ETM

f: 1,000	p: 0,923
s: 1,000	g: 0,889



1) Carefully choose your representational bias during discovery: Unrelated to presentation/visualization!

2) Consider all conformance dimensions (replay fitness, precision, generalization, and simplicity)!

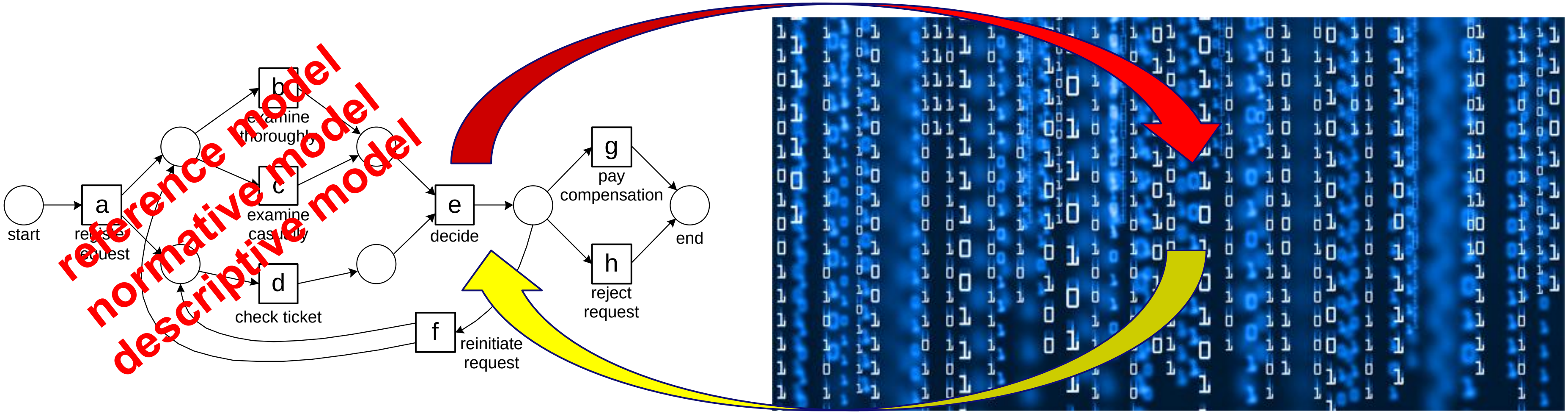
Some pointers

- Wil M. P. van der Aalst, Arya Adriansyah, Boudewijn F. van Dongen: Causal Nets: A Modeling Language Tailored towards Process Discovery. CONCUR 2011: 28-42
- Joos C. A. M. Buijs, Boudewijn F. van Dongen, Wil M. P. van der Aalst: On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. OTM Conferences (1) 2012: 305-322
- Joos C. A. M. Buijs, Boudewijn F. van Dongen, Wil M. P. van der Aalst: A genetic algorithm for discovering process trees. IEEE Congress on Evolutionary Computation 2012: 1-8
- S.J.J. Leemans, D. Fahland, W.M.P. van der Aalst. Discovering Block-Structured Process Models From Event Logs – A Constructive Approach. BPM Center Report BPM-13-06, BPMcenter.org, 2013

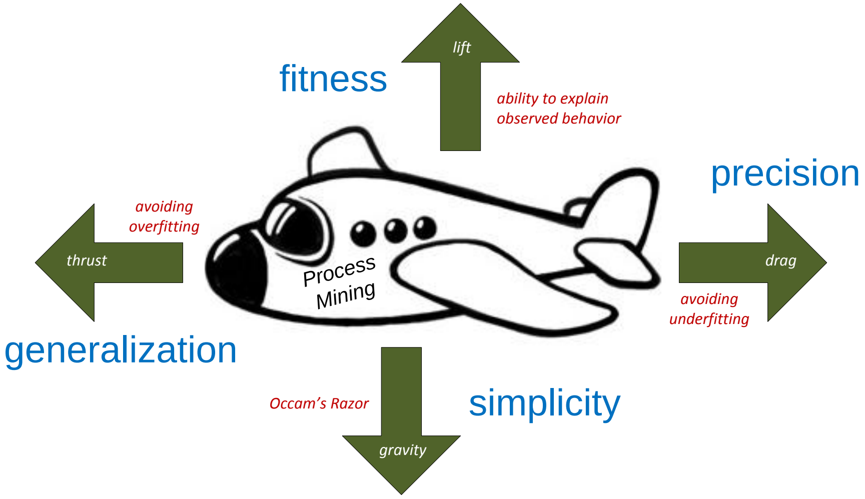
Mediating Between a Reference Model and Real Observed Behavior

Joint work with Joos Buijs, Boudewijn van Dongen, and Dirk Fahland.

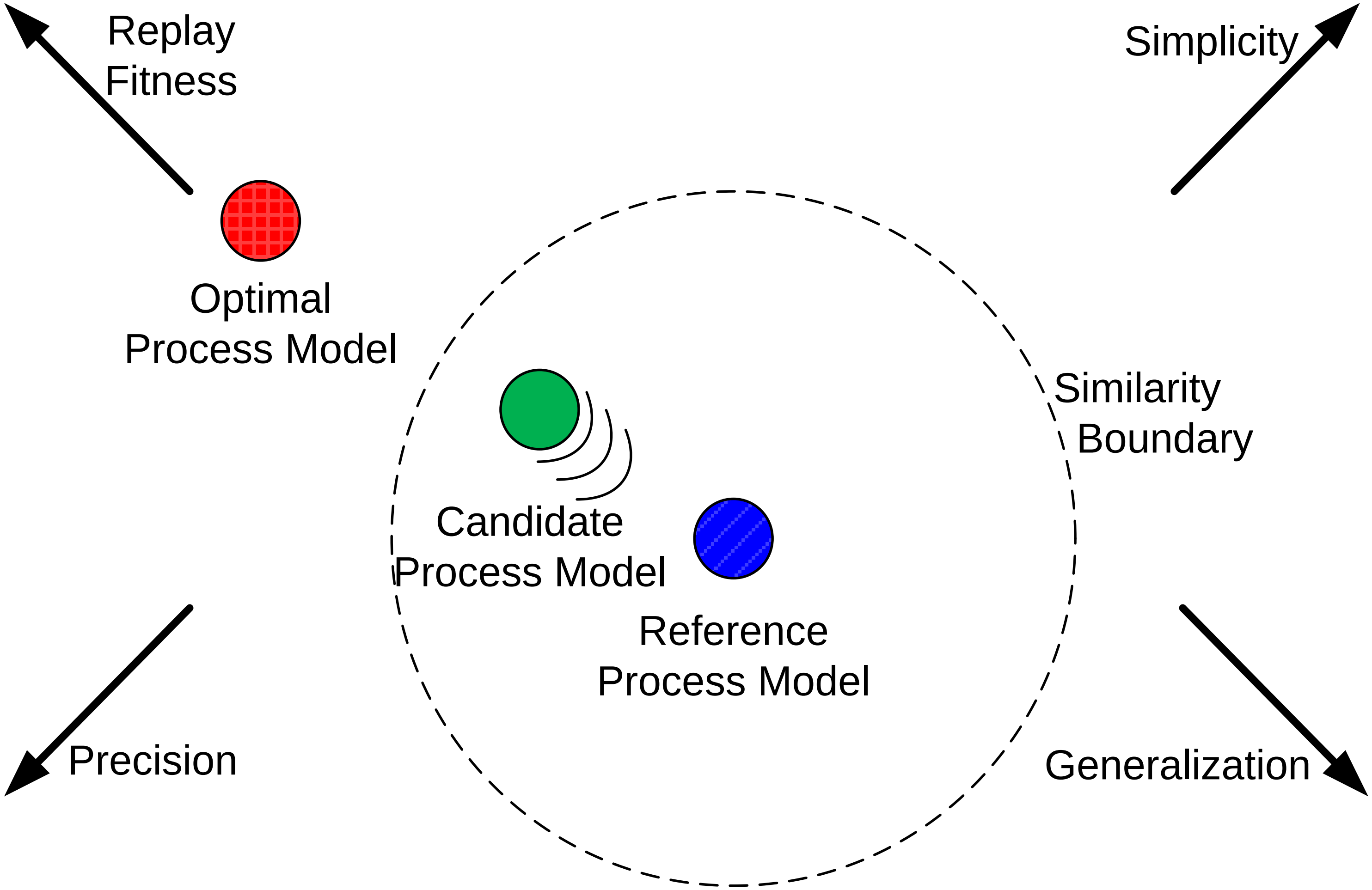
Compromise Based on Two Main Forces



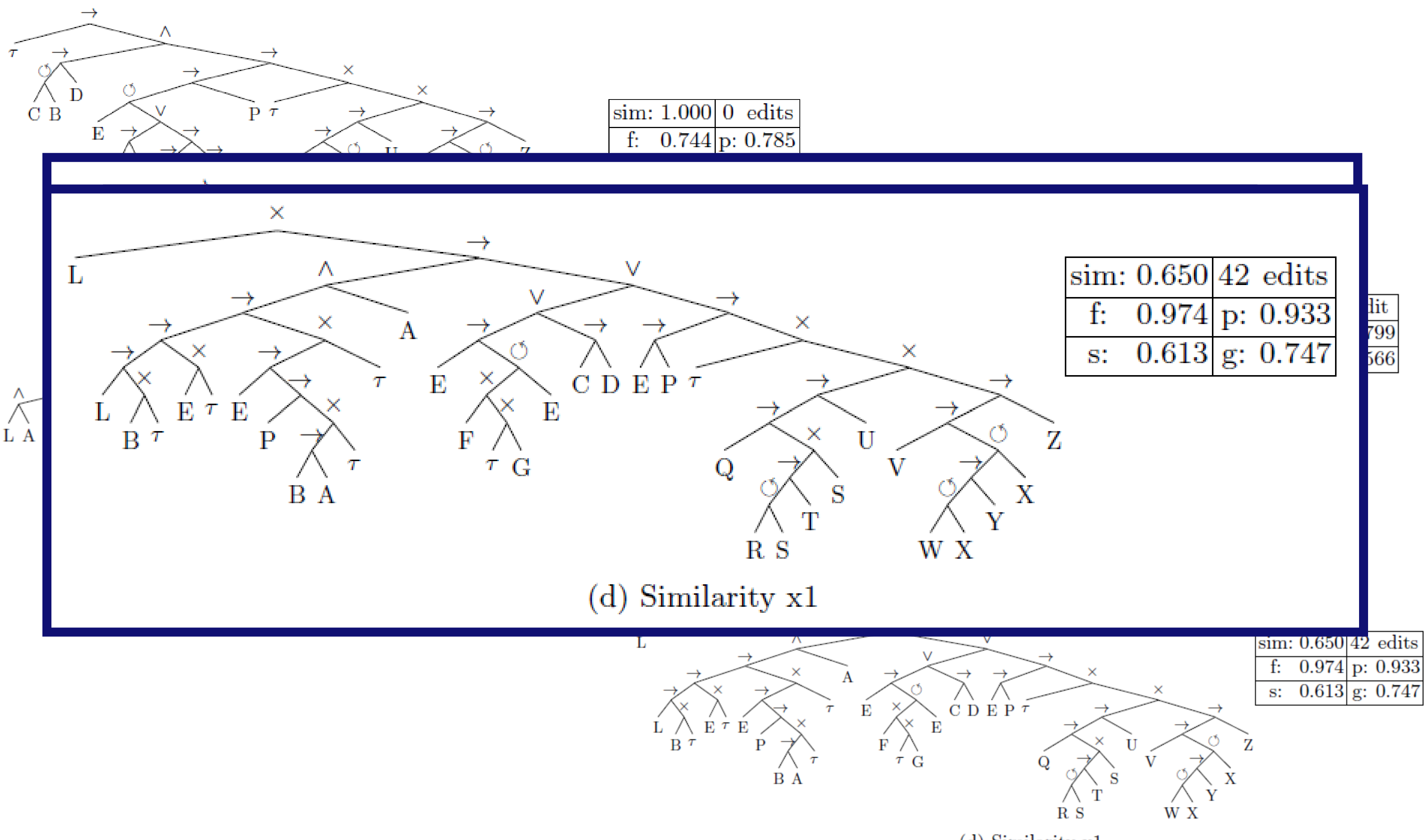
Various techniques to compare graphs, e.g., edit distance notions (add, remove, replace).



Force Between Reference Model and Candidate Model Can be Viewed as a 5th Conformance Notion



Example From CoSeLoG Project



sim: 1.000	0 edits
f: 0.744	p: 0.785

sim: 0.650	42 edits
f: 0.974	p: 0.933
s: 0.613	g: 0.747

(d) Similarity x1

sim: 0.650	42 edits
f: 0.974	p: 0.933
s: 0.613	g: 0.747

(d) Similarity x1

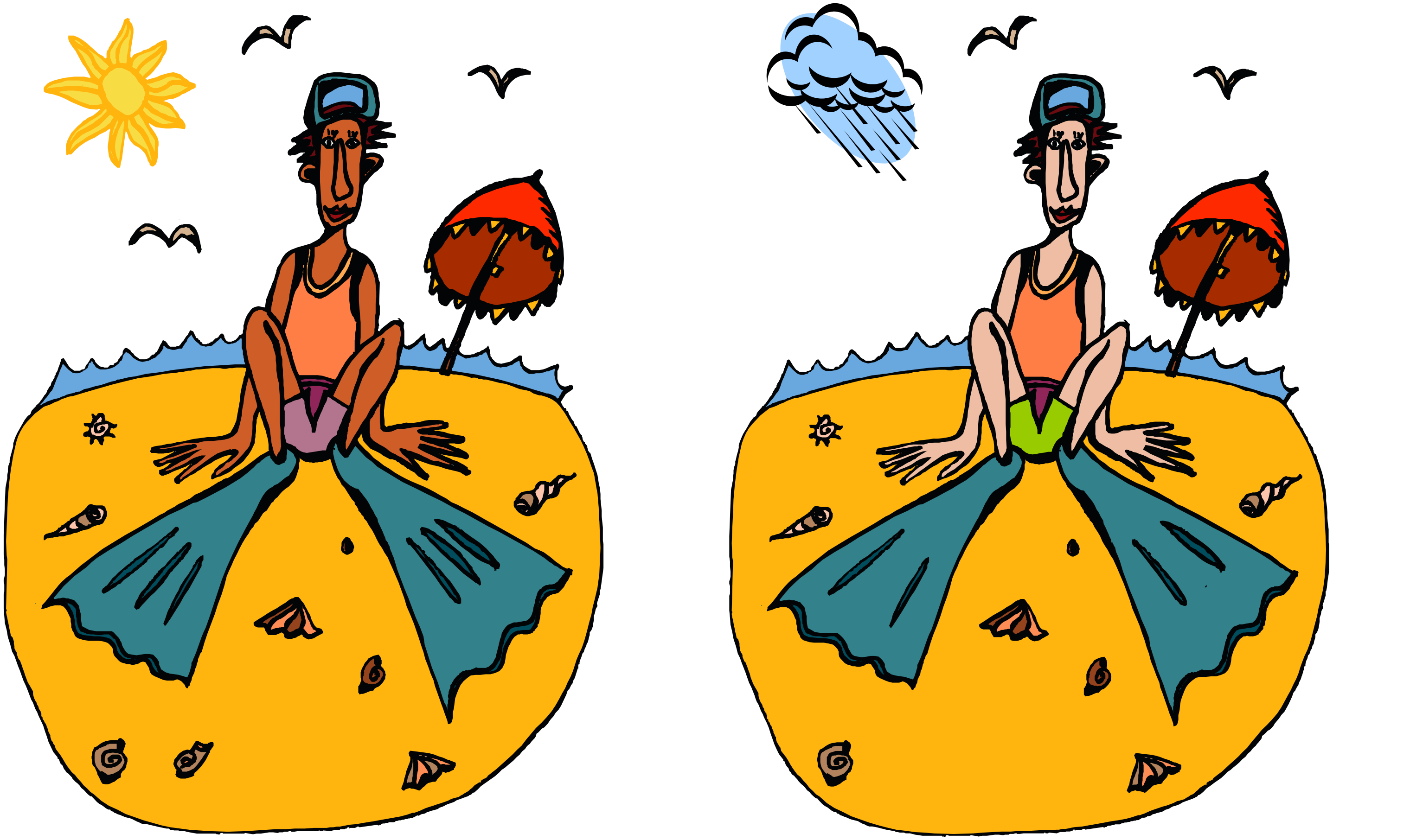
Some pointers

- J.C.A.M. Buijs ,M. La Rosa, H.A. Reijers, B.F. van Dongen, and W.M.P. van der Aalst: Improving Business Process Models using Observed Behavior, SIMPDA 2012 post-proceedings, Lecture Notes in Business Information Processing, 2013.
- Joos C. A. M. Buijs, Boudewijn F. van Dongen, Wil M. P. van der Aalst: On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. OTM Conferences (1) 2012: 305-322
- Dirk Fahland, Wil M. P. van der Aalst: Repairing Process Models to Reflect Reality. BPM 2012: 229-245.
- Dirk Fahland, Wil M. P. van der Aalst: Simplifying discovered process models in a controlled manner. Inf. Syst. 38(4): 585-605 (2013).

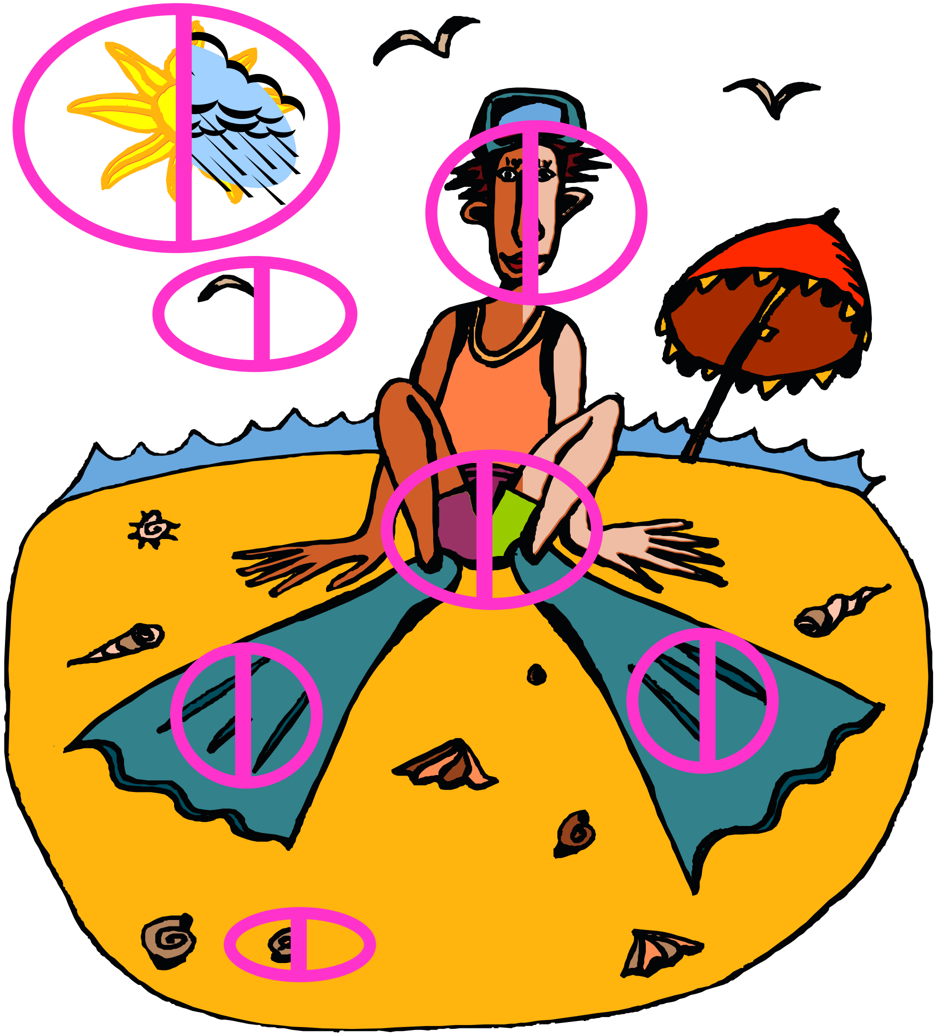
Mining Configurable Process Models

Joint work with Joos Buijs, Boudewijn van Dongen, and Florian Gottschalk.

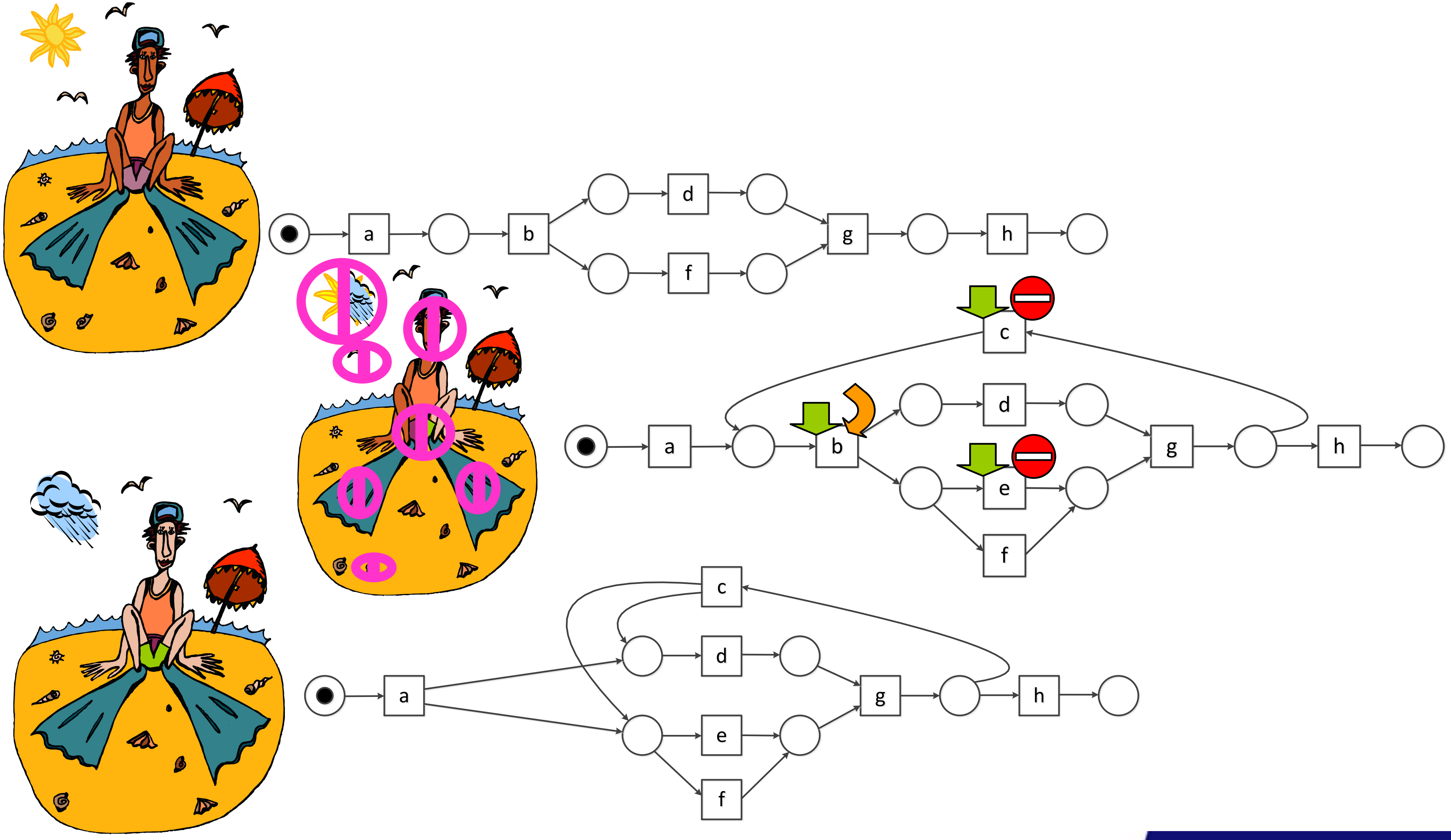
Two variants of the same process …



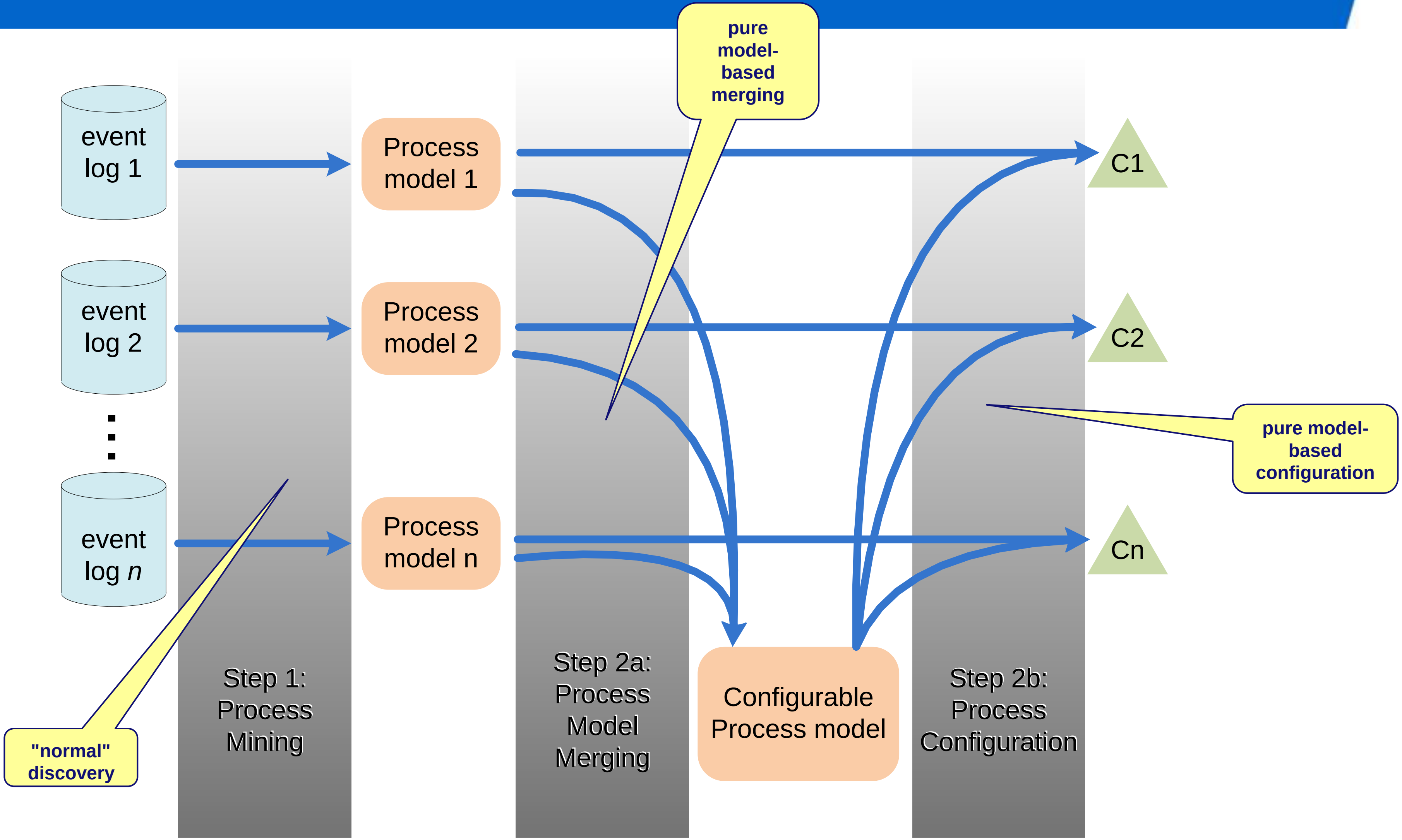
Configurable Process Model



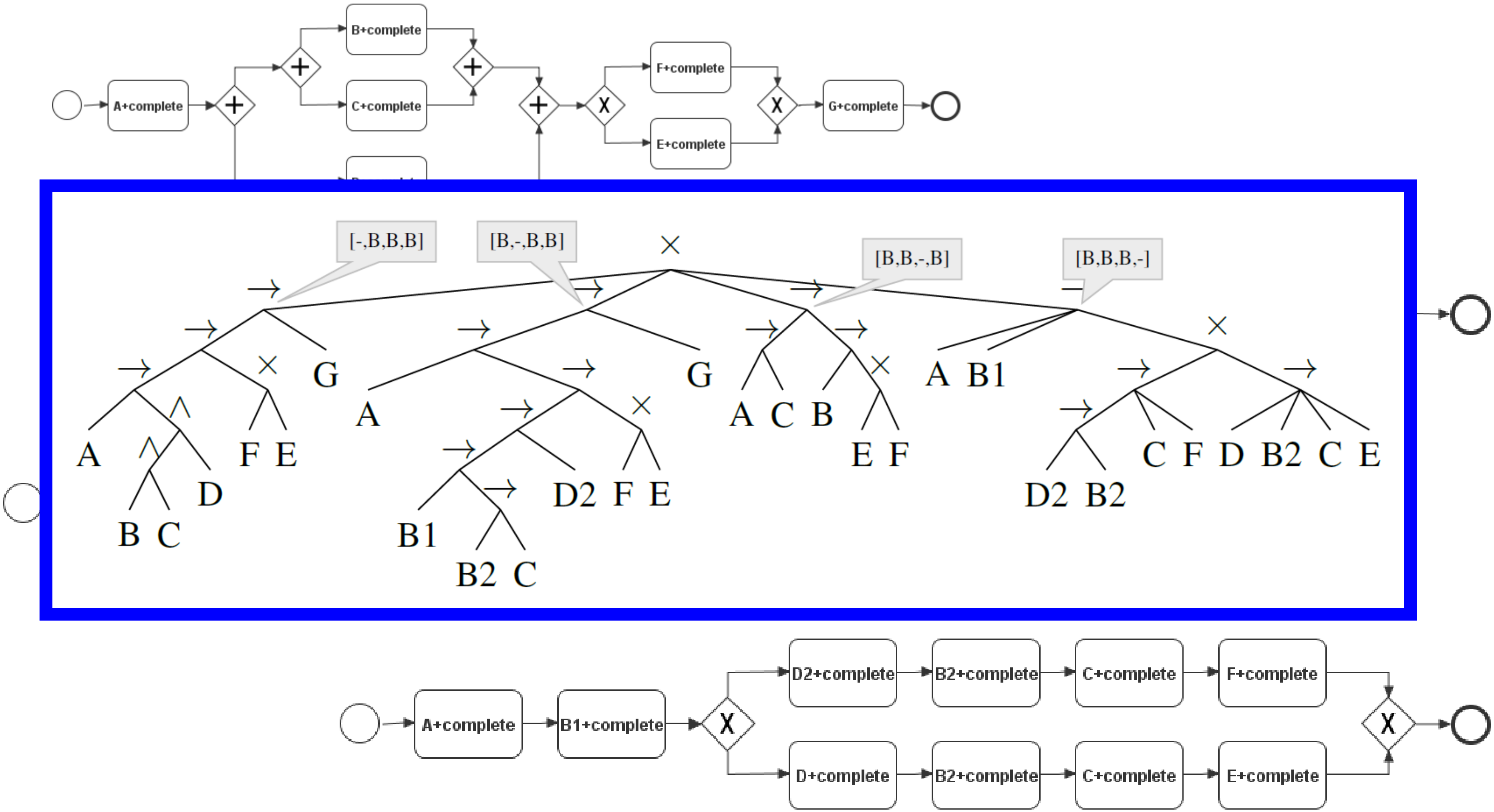
Variants of the same process



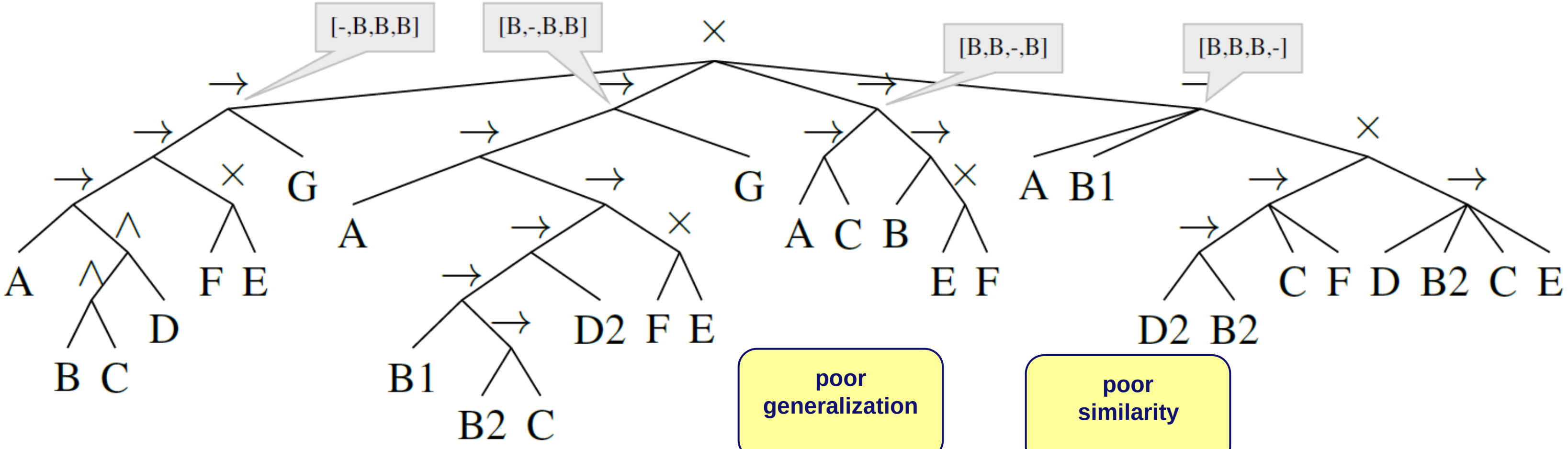
Approach 1: Merge Separately Mined Models



Results Approach 1 for running example



Results Approach 1 for running example



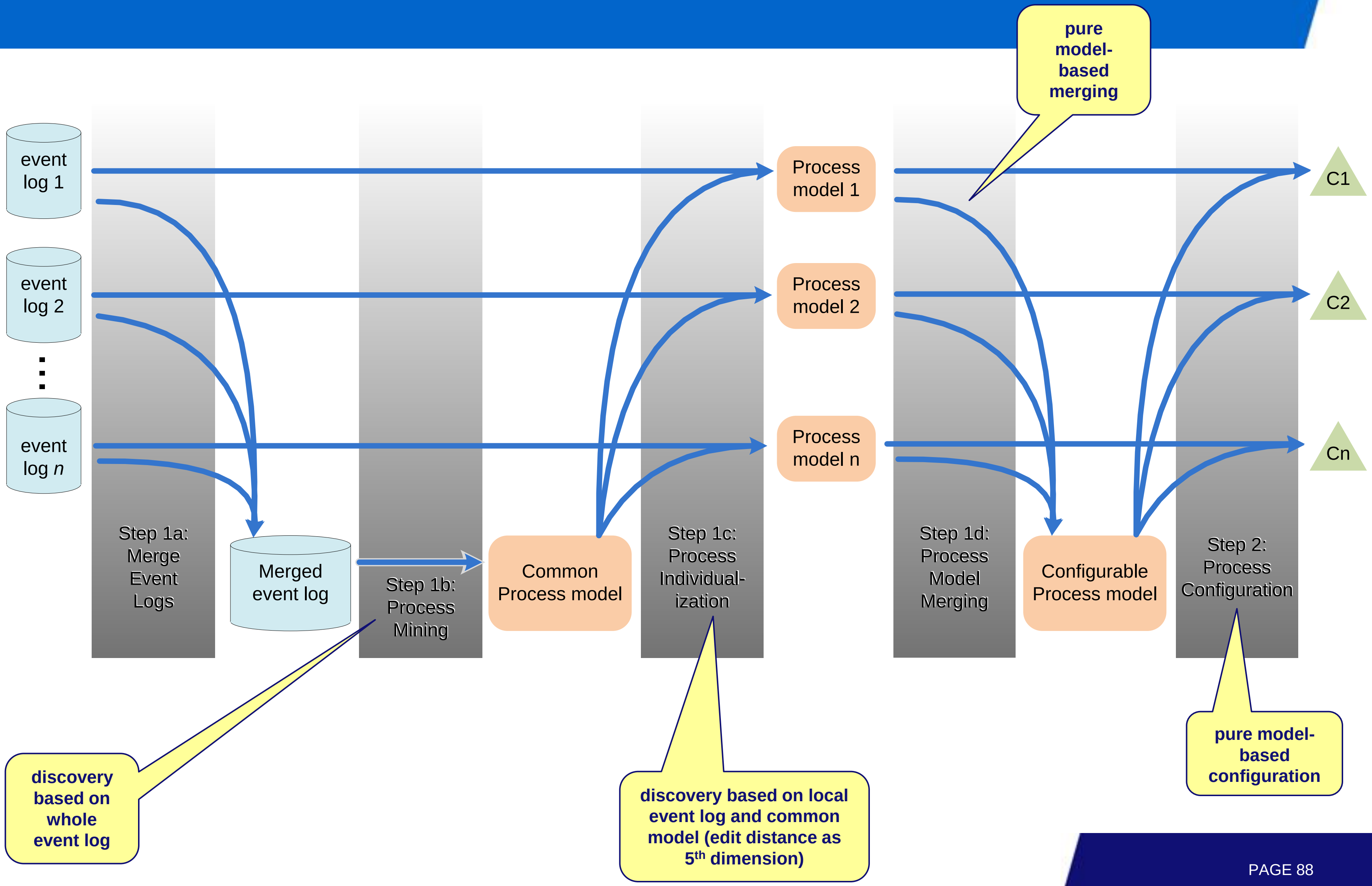
poor generalization

poor similarity

	Overall	Fitness	Precision	Simplicity	Generalization
Combined	0.989	0.999	0.999	0.981	0.220
Variant 0	0.986	0.995	0.995	0.981	0.235
Variant 1	0.989	1.000	1.000	0.981	0.263
Variant 2	0.989	1.000	1.000	0.981	0.174
Variant 3	0.989	1.000	1.000	0.981	0.264

Size	#C.P.	Similarity
53	4	-
14	3	0.418
16	3	0.464
10	3	0.317
16	3	0.464

Approach 2



event log 1
event log 2
event log *n*
Step 1a: Merge Event Logs
Merged event log
Step 1b: Process Mining
Common Process model
Step 1c: Process Individual-ization
Process model 1
Process model 2
Process model n
Step 1d: Process Model Merging
Configurable Process model
Step 2: Process Configuration
C1
C2
Cn

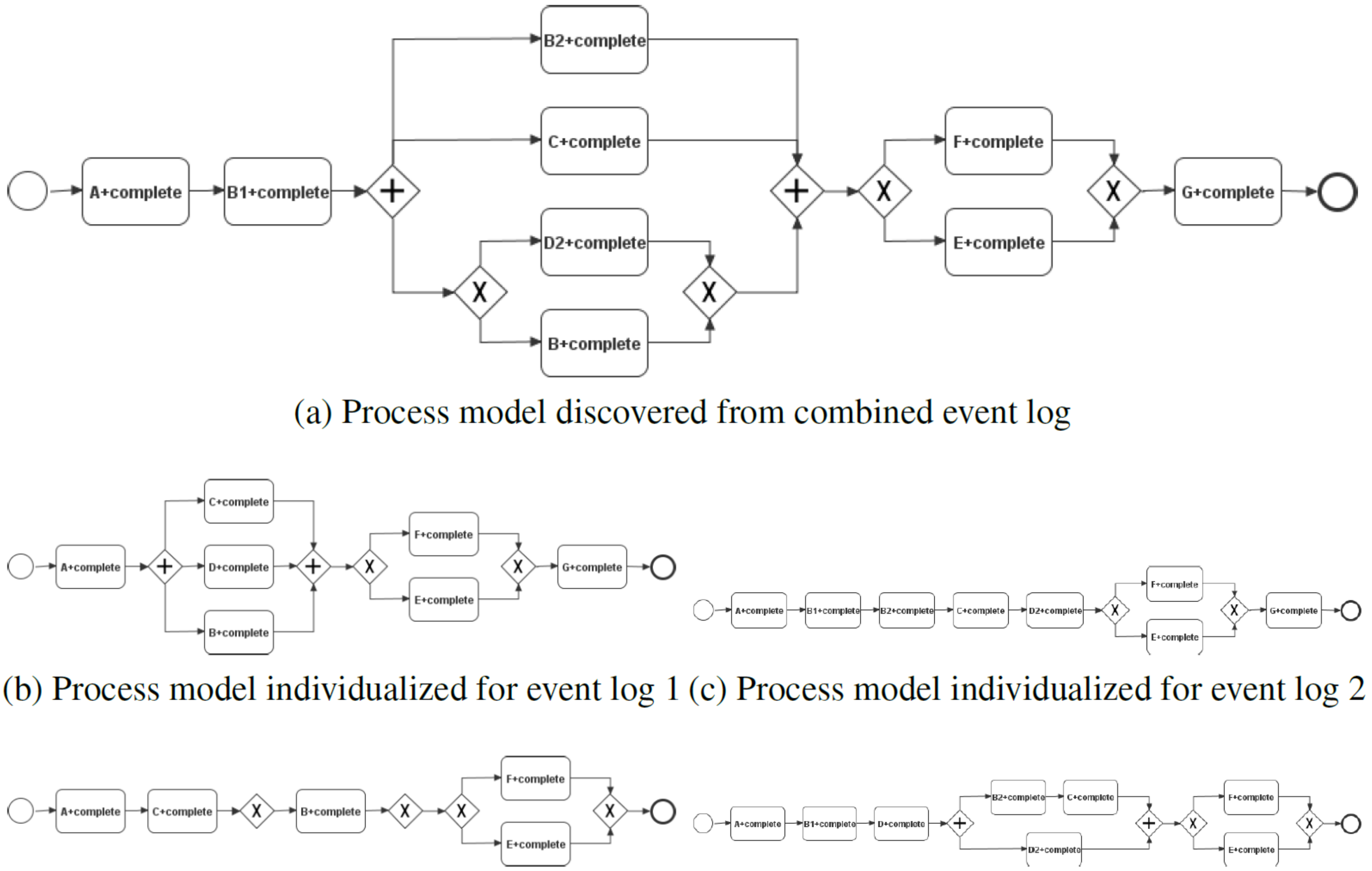
pure model-based merging

discovery based on whole event log

discovery based on local event log and common model (edit distance as 5th dimension)

pure model-based configuration

Results Approach 2 for running example

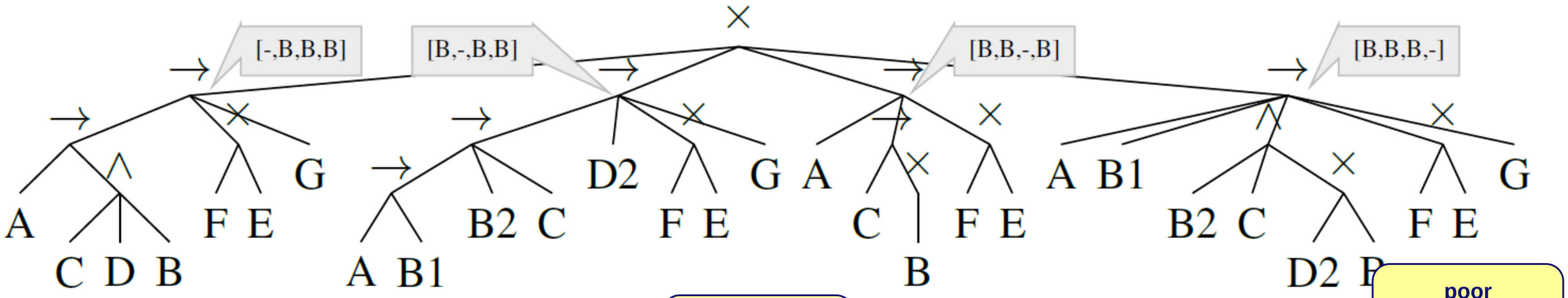


(a) Process model discovered from combined event log

(b) Process model individualized for event log 1 (c) Process model individualized for event log 2

(d) Process model individualized for event log 3 (e) Process model individualized for event log 4

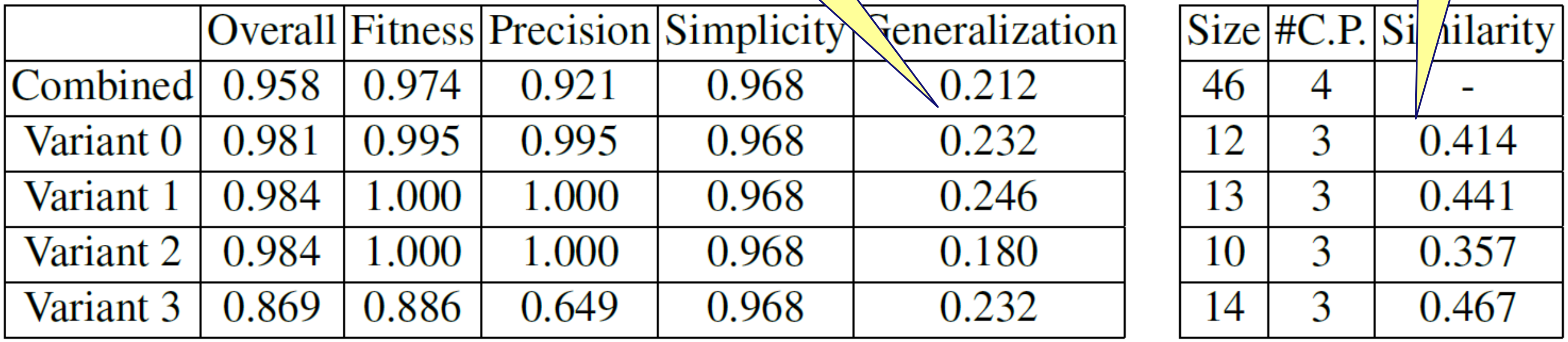
Results Approach 2 for running example



(f) Configurable process model o er merging models (b) throug

poor generalization

poor similarity

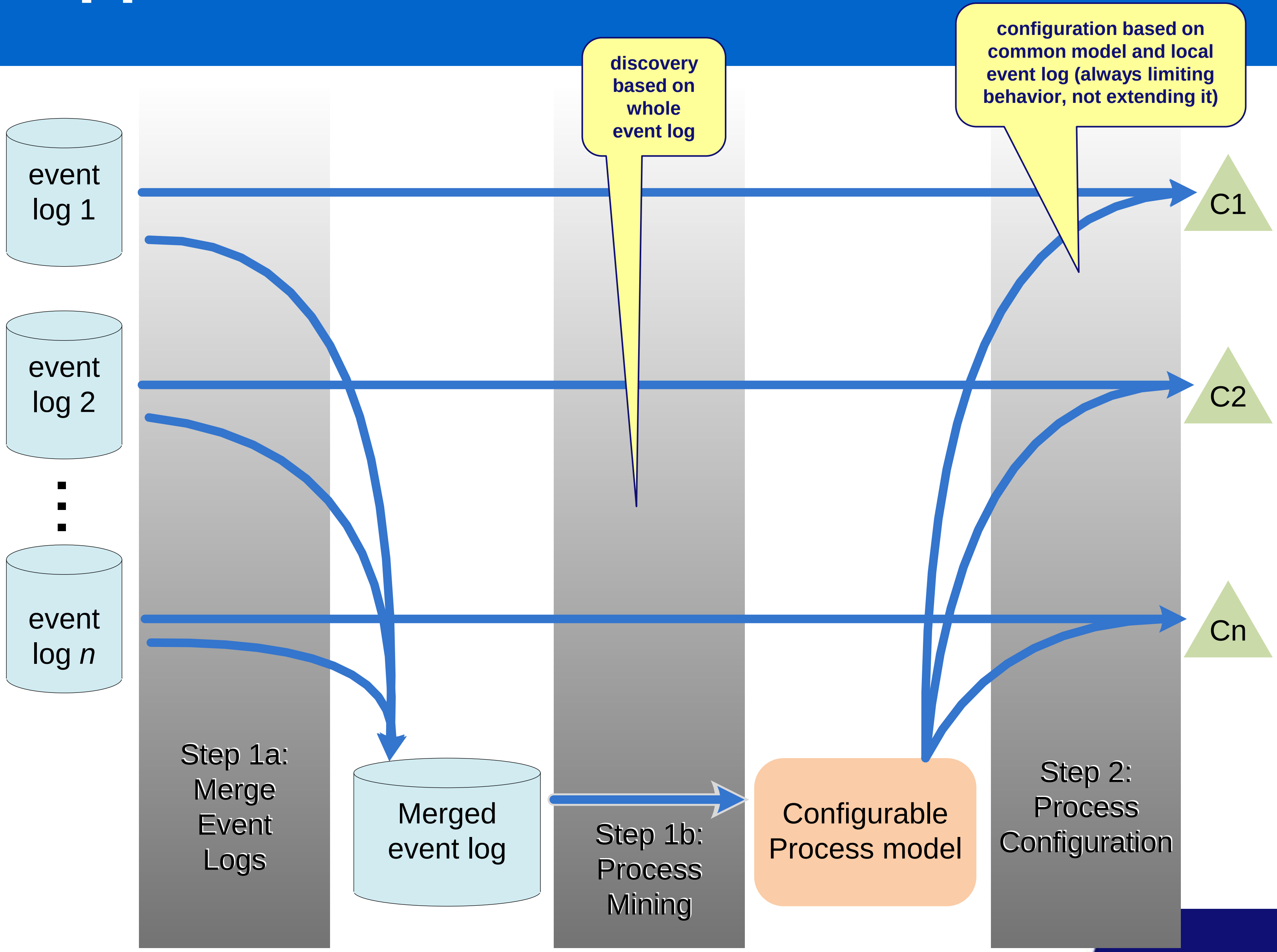


	Overall	Fitness	Precision	Simplicity	Generalization
Combined	0.958	0.974	0.921	0.968	0.212
Variant 0	0.981	0.995	0.995	0.968	0.232
Variant 1	0.984	1.000	1.000	0.968	0.246
Variant 2	0.984	1.000	1.000	0.968	0.180
Variant 3	0.869	0.886	0.649	0.968	0.232

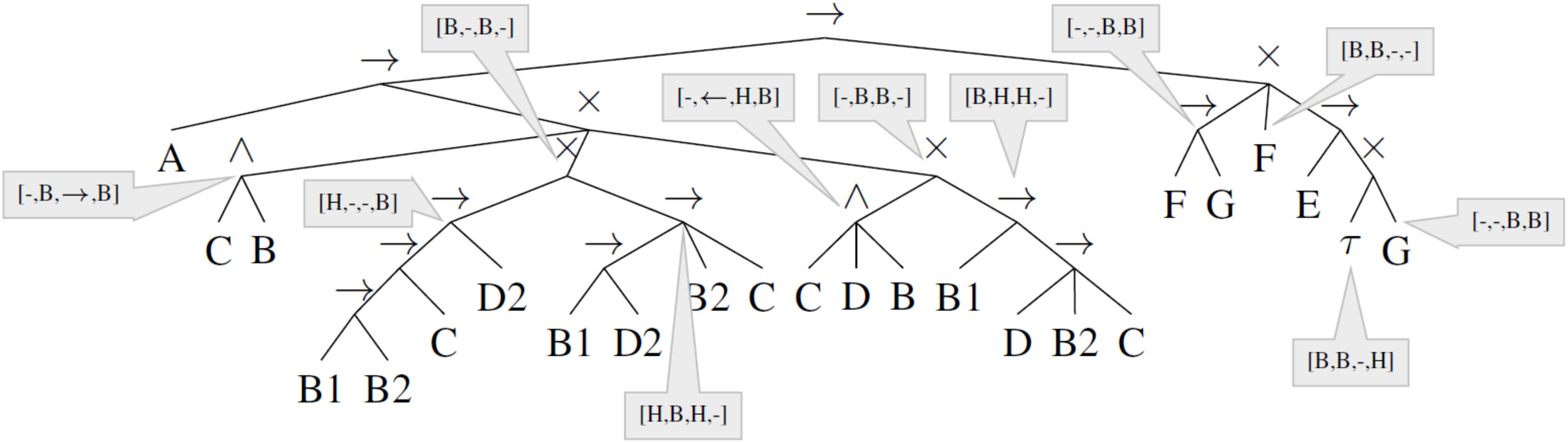
Size	#C.P.	Similarity
46	4	-
12	3	0.414
13	3	0.441
10	3	0.357
14	3	0.467

(g) Quality statistics of the configurable process model of (f)

Approach 3


event log 1
event log 2
event log n
Step 1a: Merge Event Logs
Merged event log
Step 1b: Process Mining
discovery based on whole event log
Configurable Process model
Step 2: Process Configuration
configuration based on common model and local event log (always limiting behavior, not extending it)
C1
C2
Cn

Results Approach 3 for running example



(a) Configurable process model discovered when first discovering the process model and then the configurations

better generalization

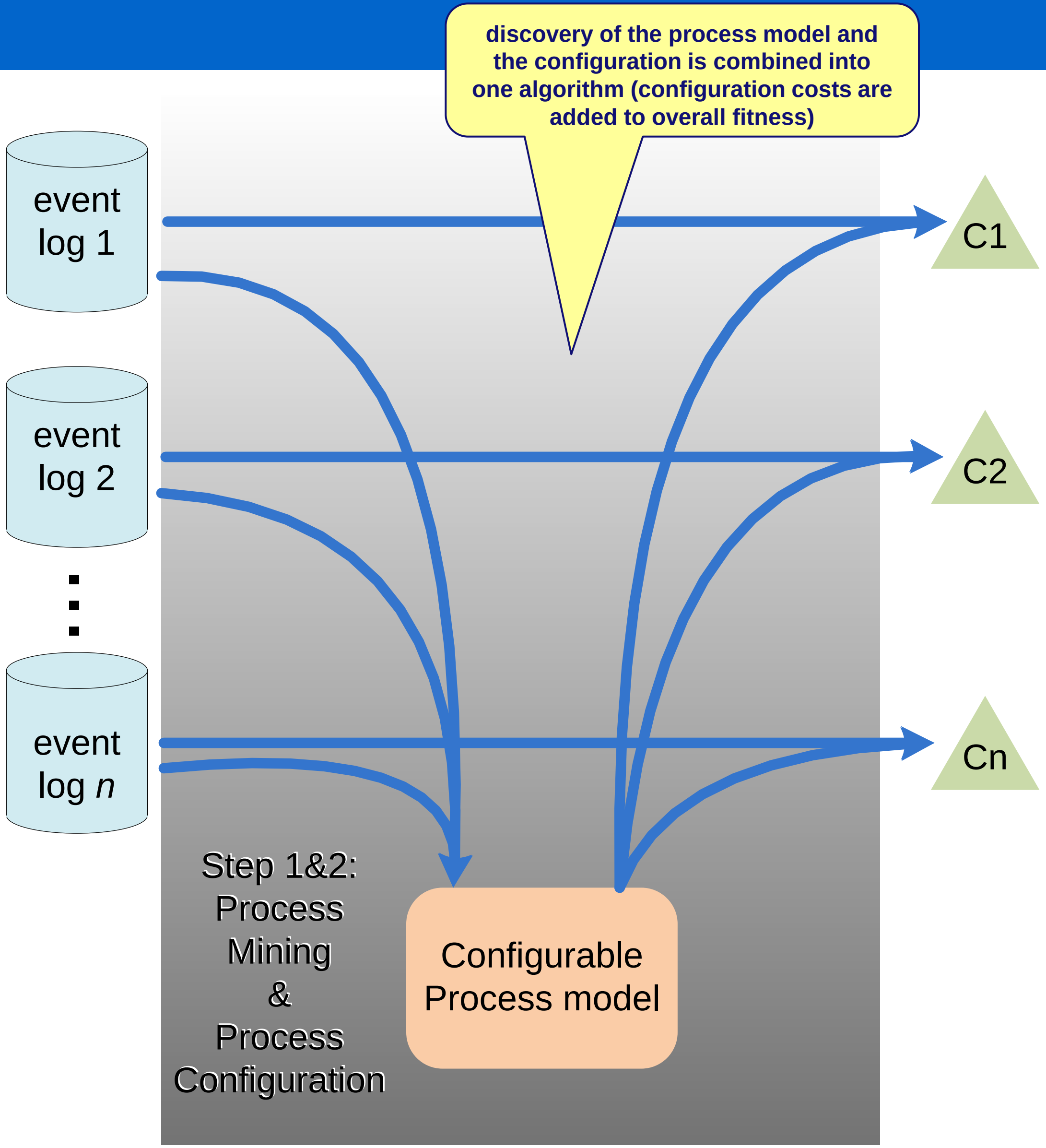
better similarity

	Overall	Fitness	Precision	Simplicity	Generalization
Combined	0.988	1.000	0.981	0.986	0.374
Variant 0	0.990	1.000	0.990	0.986	0.400
Variant 1	0.992	1.000	1.000	0.986	0.408
Variant 2	0.992	1.000	1.000	0.986	0.285
Variant 3	0.977	1.000	0.922	0.986	0.496

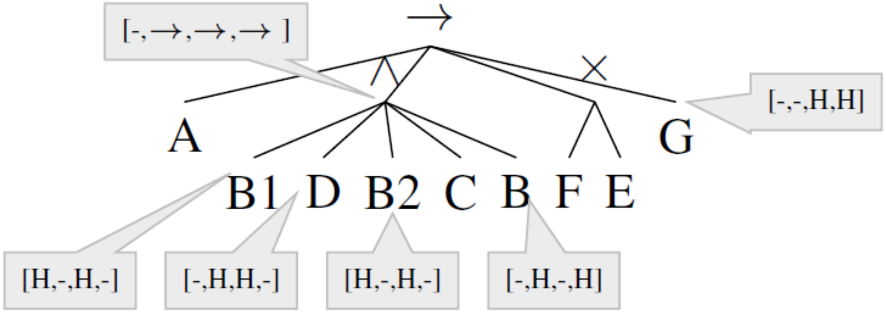
Size	#C.P.	Similarity
42	11	-
20	6	0.645
20	7	0.645
13	8	0.473
24	6	0.727

(b) Quality statistics of the configurable process model of (a)

Approach 4


discovery of the process model and the configuration is combined into one algorithm (configuration costs are added to overall fitness)
event log 1
event log 2
event log *n*
C1
C2
Cn
Step 1&2: Process Mining & Process Configuration
Configurable Process model

Results Approach 4 for running example



(a) Configurable odel discov using the integrate y approach

small drop in replay fitness

better simplicity

even better generalization

even better similarity

	Overall	Fitness	Precision	Simplicity	Generalization
Combined	0.983	0.962	0.999	1.000	0.684
Variant 0	0.996	0.995	0.994	1.000	0.738
Variant 1	0.957	0.894	1.000	1.000	0.723
Variant 2	0.998	1.000	1.000	1.000	0.614
Variant 3	0.961	0.905	1.000	1.000	0.741

Size	#C.P.	Similarity
12	6	-
10	2	0.909
10	3	0.909
8	5	0.800
10	3	0.909

(b) Quality statistics of the configurable process model of (a)

Genetic algorithms are versatile and can consider different forces at the same time

… but often not fast enough

Some pointers

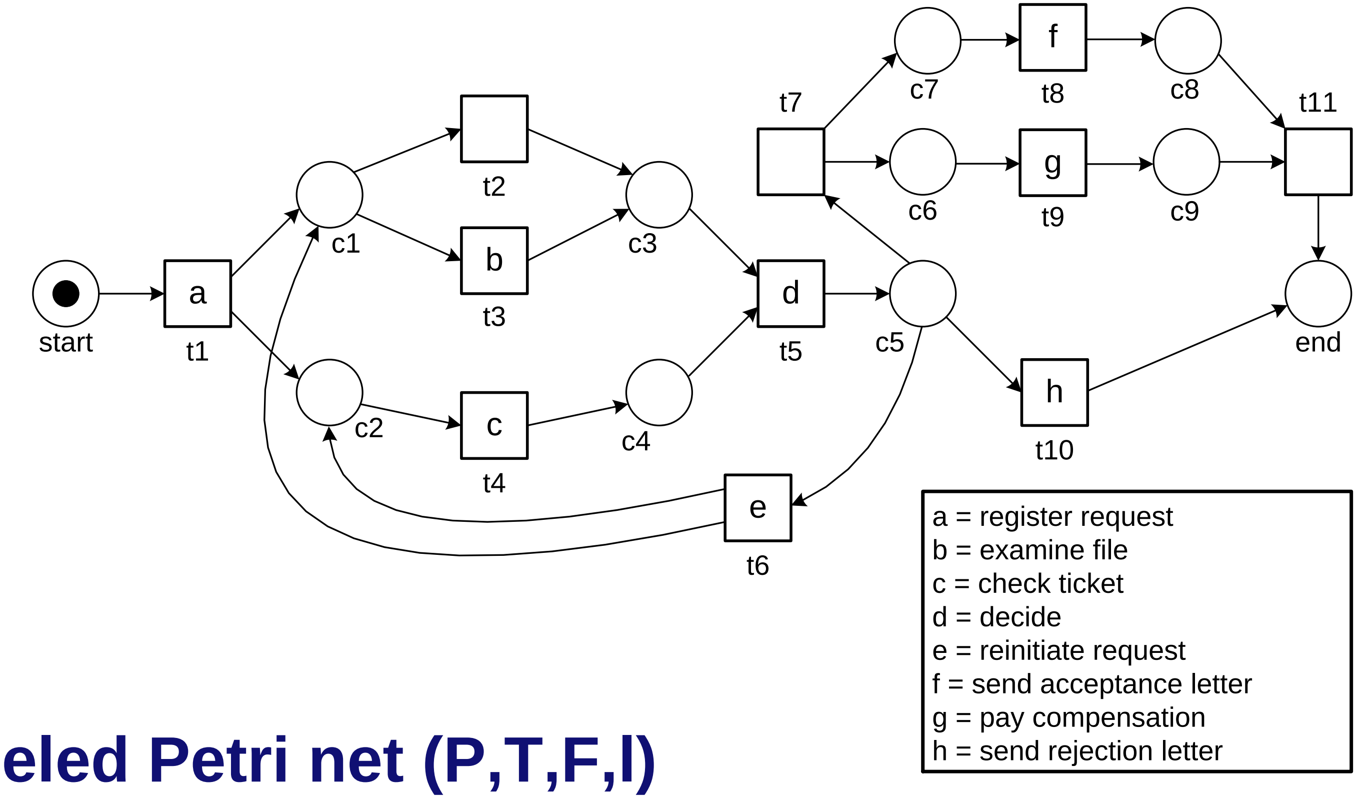
- J.C.A.M. Buijs, B.F. van Dongen, and W.M.P. van der Aalst: Mining Configurable Process Models from Collections of Event Logs, under review, 2013.
- Florian Gottschalk, Teun A. C. Wagemakers, Monique H. Jansen-Vullers, Wil M. P. van der Aalst, Marcello La Rosa: Configurable Process Models: Experiences from a Municipality Case Study. CAiSE 2009: 486-500
- Florian Gottschalk, Wil M. P. van der Aalst, Monique H. Jansen-Vullers: Merging Event-Driven Process Chains. OTM Conferences (1) 2008: 418-426
- Wil M. P. van der Aalst: Business Process Configuration in the Cloud: How to Support and Analyze Multi-tenant Processes? ECOWS 2011: 3-10

Decomposing Process Mining Problems

Joint work with Eric Verbeek, Jorge Munoz , and Josep Carmona.

Big Data: Opportunities and Challenges

System Net



a = register request
b = examine file
c = check ticket
d = decide
e = reinitiate request
f = send acceptance letter
g = pay compensation
h = send rejection letter

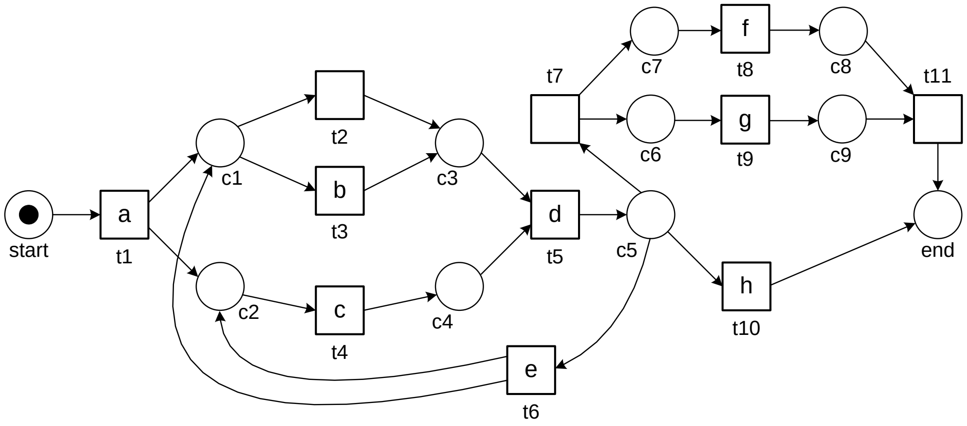
- **Labeled Petri net (P,T,F,l)**
- **Silent transitions and visible transitions (unique or not),**
- **One initial marking M_{init}, one final marking M_{final}**

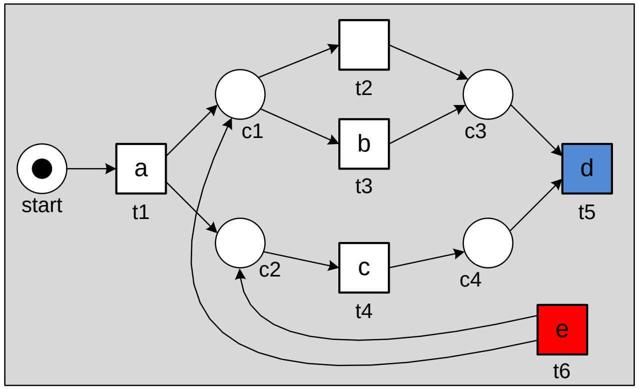
Traces and Logs

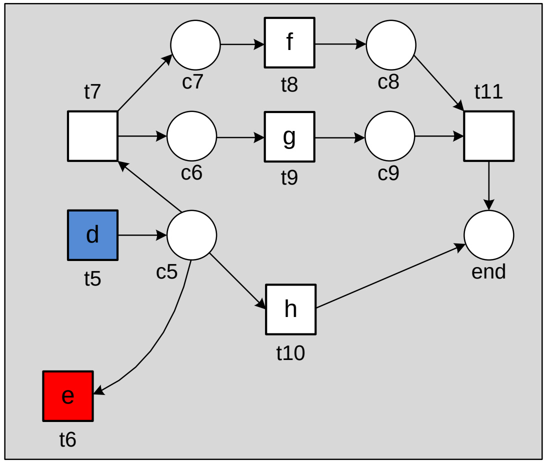
- A system net SN has a set of possible visible traces ϕ(SN) starting in M_{init} and ending M_{final} only showing the visible steps.
- An event log L is a multiset of traces.
- Two main process mining problems:

1. Conformance checking: Given L and SN, evaluate the "conformance" (e.g., fitness, precision, generalization, etc.) of L and ϕ(SN)
2. Process discovery: Given L, create SN such that the conformance of L and ϕ(SN) is "as good as possible"

Valid Decomposition

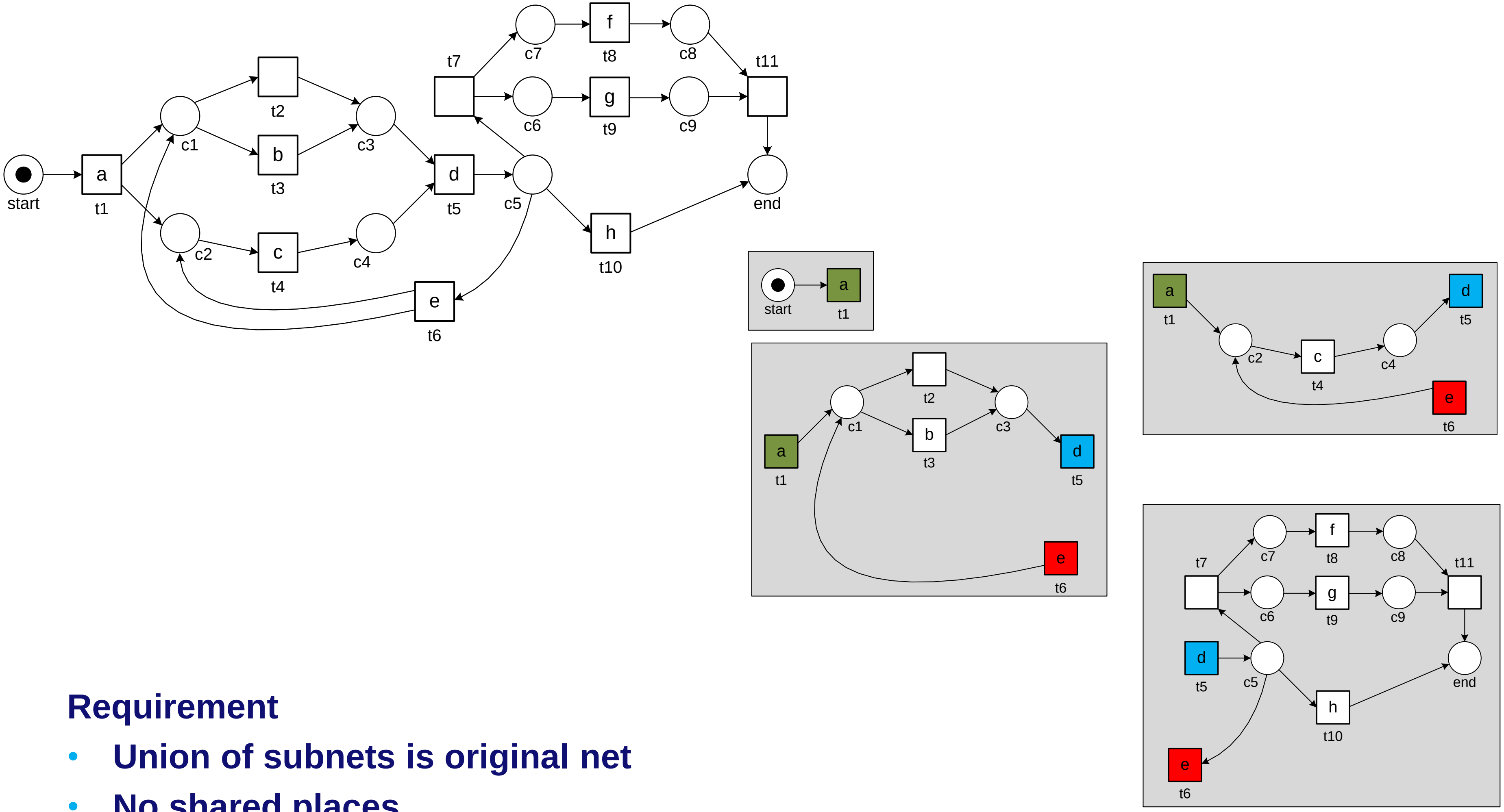






- **Union of subnets is original net**
- **No shared places**
- **Shared transitions are visible and have unique label**

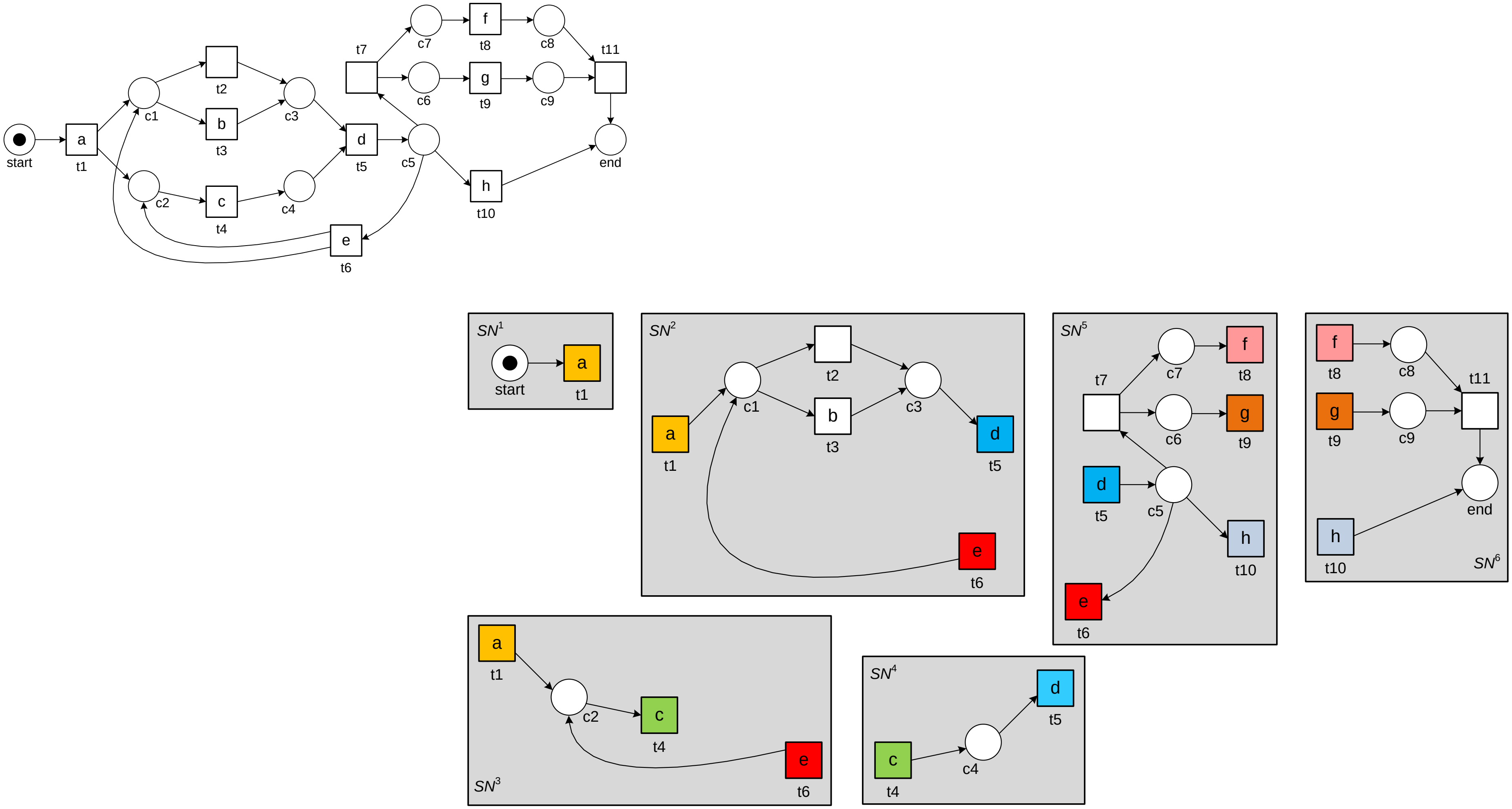
Another Decomposition



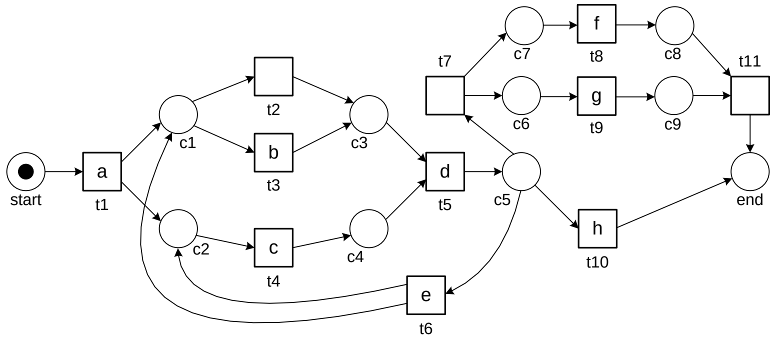
Requirement

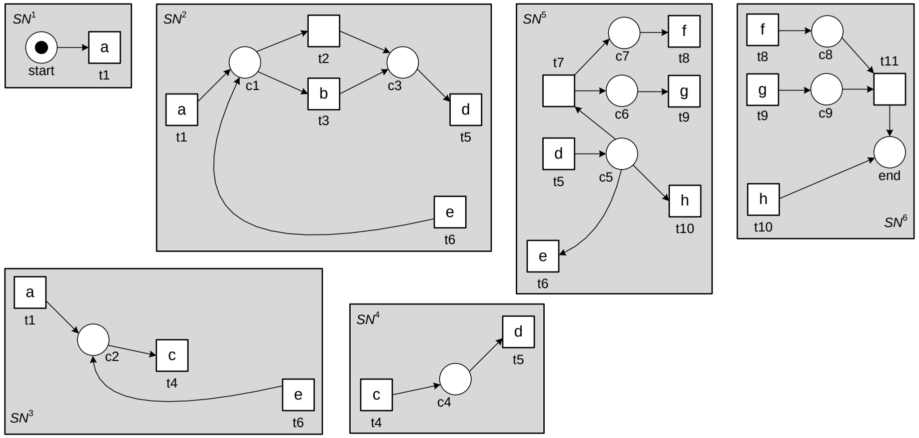
- Union of subnets is original net
- No shared places
- Shared transitions are visible and unique

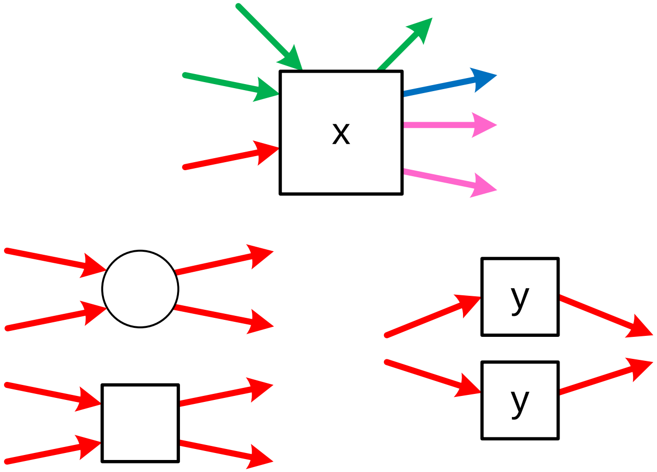
Maximal Valid Decomposition



Maximal Decomposition

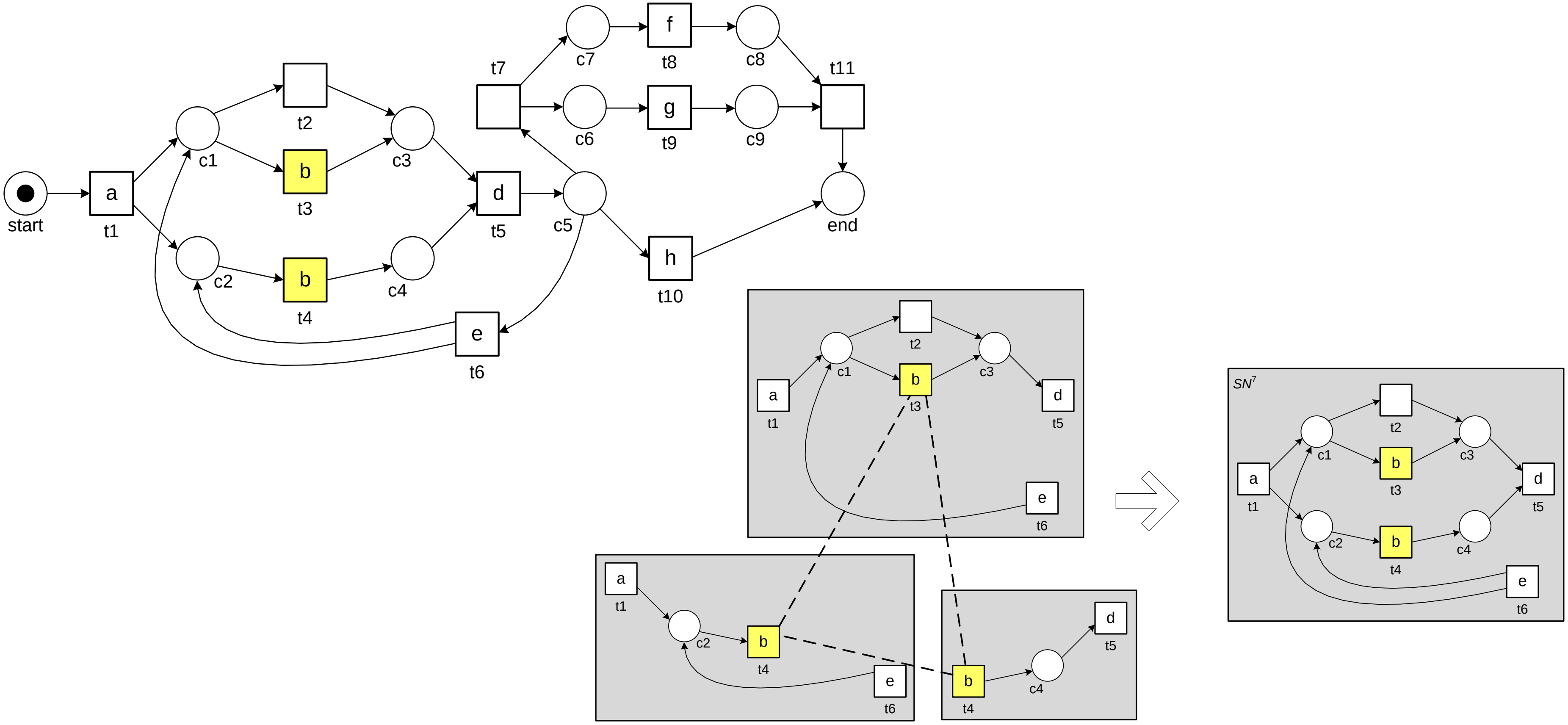






- Construction: group arcs iteratively
- Maximal decomposition is unique
- There is always a valid decomposition

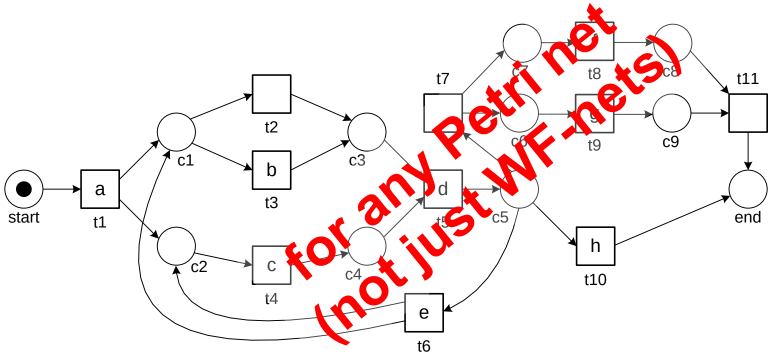
Non-unique visible labels



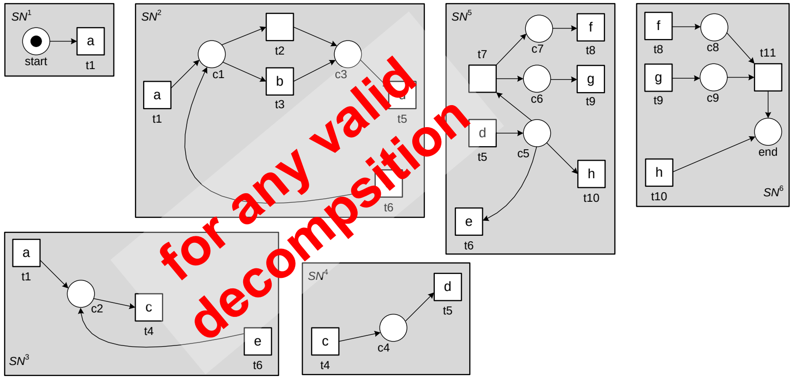
- **Union of subnets is original net**
- **No shared places**
- **Shared transitions are visible and unique**

Conformance checking can be decomposed !!!

- Let L be an event log, SN a system net, and D={SN^1,SN^2, … SN^n} a valid decomposition
- L^i is the sublog of SN^i (L projected onto visible transitions of SN^i)



for any Petri net (not just WF-nets)

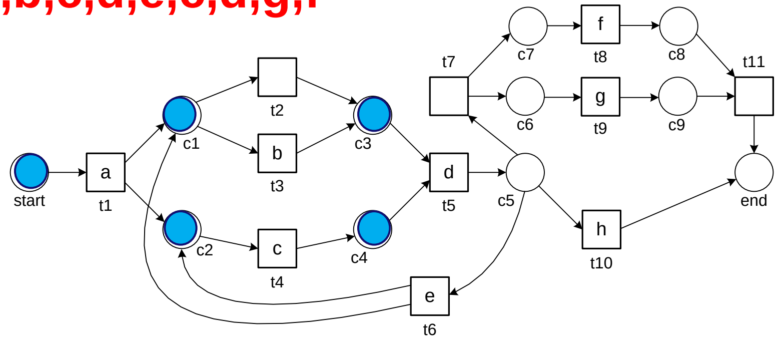


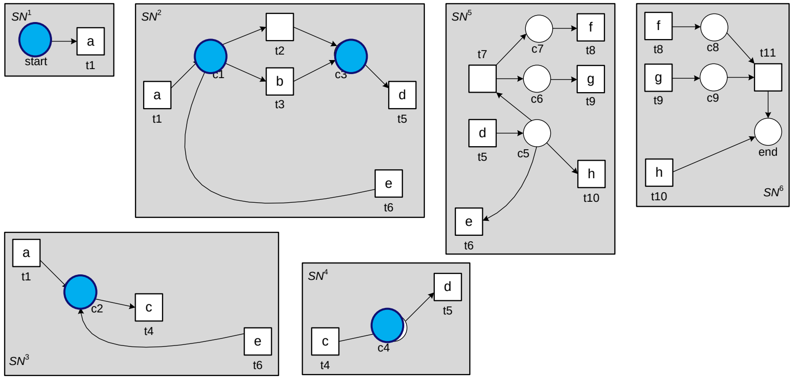
for any valid decomposition

L is perfectly fitting SN
if and only if
each projected log is L^i is perfectly fitting SN^i

Example of alignment for observed trace a,b,c,d,e,c,d,g,f

a,b,c,d,e,c,d,g,f





$$\gamma_3 = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ \hline a & b & c & d & e & c & \gg & d & \gg & g & f & \gg \\ \hline a & b & c & d & e & c & \tau & d & \tau & g & f & \tau \\ t1 & t3 & t4 & t5 & t6 & t4 & t2 & t5 & t7 & t9 & t8 & t11 \end{array}$$

Etc.

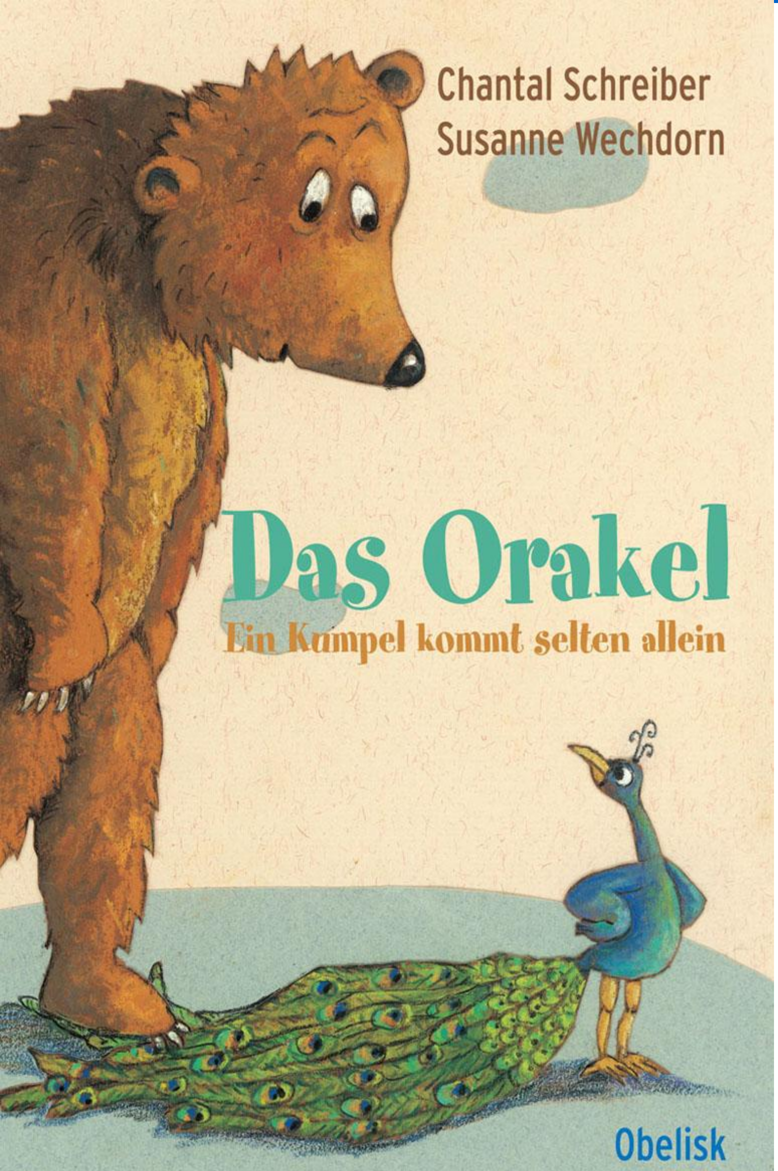
$$\gamma_3^1 = \begin{array}{|c|} 1 \\ \hline a \\ \hline a \\ t1 \end{array} \qquad \gamma_3^2 = \begin{array}{|c|c|c|c|c|c|} 1 & 2 & 4 & 5 & 7 & 8 \\ \hline a & b & d & e & \gg & d \\ \hline a & b & d & e & \tau & d \\ t1 & t3 & t5 & t6 & t2 & t5 \end{array} \qquad \gamma_3^3 = \begin{array}{|c|c|c|c|} 1 & 3 & 5 & 6 \\ \hline a & c & e & c \\ \hline a & c & e & c \\ t1 & t4 & t6 & t4 \end{array}$$

$$\gamma_3^4 = \begin{array}{|c|c|c|c|} 3 & 4 & 6 & 8 \\ \hline c & d & c & d \\ \hline c & d & c & d \\ t4 & t5 & t4 & t5 \end{array} \qquad \gamma_3^5 = \begin{array}{|c|c|c|c|c|c|} 4 & 5 & 8 & 9 & 10 & 11 \\ \hline d & e & d & \gg & g & f \\ \hline d & e & d & \tau & g & f \\ t5 & t6 & t5 & t7 & t9 & t8 \end{array} \qquad \gamma_3^6 = \begin{array}{|c|c|c|} 10 & 11 & 12 \\ \hline g & f & \gg \\ \hline g & f & \tau \\ t9 & t8 & t11 \end{array}$$

So What?
Quantifying Conformance

- **Exact result:** Fraction of cases perfectly fitting SN **equals** the fraction of cases for which each projected log is L^i is perfectly fitting SN^i.
- **Bound:** For fitness at the event level (costs based alignment) it is possible to compute an **optimistic** value.

Discovery can also be distributed!



1. **Assume the set of activities is split in overlapping sets.**
2. **Split log L in sublogs L^i based on these sets.**
3. **Discover a model SN^i per sublog.**
4. **Merge the models SN^i into SN and return result.**

All the earlier guarantees hold, e.g., L is perfectly fitting SN if and only if each projected log is L^i is perfectly fitting SN^i

Example

$$L = [\langle a,b,c,d,h\rangle^{30}, \langle a,c,b,d,e,b,c,d,j,g,f,k\rangle^{28}, \langle a,c,b,d,e,c,b,d,h\rangle^{28}, \langle a,c,b,d,e,c,b,d,j,f,g,k\rangle^{27}, \langle a,b,c,d,e,c,i,d,j,g,f,k\rangle^{26}, \langle a,b,c,d,e,i,c,d,h\rangle^{26}, \langle a,i,c,d,e,c,b,d,j,g,f,k\rangle^{25}, \langle a,i,c,d,j,f,g,k\rangle^{24}, \langle a,c,b,d,e,b,c,d,h\rangle^{24}, \langle a,b,c,d,e,b,c,d,j,g,f,k\rangle^{24}, \langle a,c,b,d,e,c,i,d,h\rangle^{22}, \langle a,b,c,d,e,c,i,d,h\rangle^{22}, \langle a,c,i,d,j,f,g,k\rangle^{21}, \langle a,c,i,d,j,g,f,k\rangle^{20}, \langle a,c,i,d,e,b,c,d,h\rangle^{18}, \langle a,c,i,d,e,c,i,d,h\rangle^{17}, \langle a,i,c,d,h\rangle^{17}, \langle a,c,b,d,j,g,f,k\rangle^{17}, \langle a,c,b,d,h\rangle^{15}, \langle a,i,c,d,e,i,c,d,j,f,g,k\rangle^{14}, \langle a,c,i,d,e,c,b,d,h\rangle^{14}, \langle a,c,i,d,e,c,b,d,j,f,g,k\rangle^{12}, \langle a,b,c,d,e,i,c,d,j,f,g,k\rangle^{12}, \langle a,b,c,d,e,c,b,d,h\rangle^{11}, \langle a,c,b,d,j,f,g,k\rangle^{10}, \langle a,i,c,d,j,g,f,k\rangle^{9}, \langle a,i,c,d,e,b,c,d,h\rangle^{9}, \langle a,i,c,d,e,c,b,d,h\rangle^{9}, \langle a,i,c,d,e,c,b,d,j,f,g,k\rangle^{8}, \langle a,c,b,d,e,b,c,d,j,f,g,k\rangle^{8}, \langle a,b,c,d,e,b,c,d,h\rangle^{7}, \langle a,c,b,d,e,i,c,d,j,g,f,k\rangle^{7}, \langle a,i,c,d,e,c,i,d,j,g,f,k\rangle^{6}, \langle a,c,i,d,h\rangle^{6}, \langle a,c,b,d,e,i,c,d,h\rangle^{6}, \langle a,i,c,d,e,b,c,d,j,g,f,k\rangle^{6}, \langle a,b,c,d,j,f,g,k\rangle^{5}, \langle a,i,c,d,e,i,c,d,h\rangle^{5}, \langle a,i,c,d,e,c,i,d,h\rangle^{4}, \langle a,c,i,d,e,i,c,d,h\rangle^{4}, \langle a,b,c,d,e,c,i,d,j,f,g,k\rangle^{4}, \langle a,b,c,d,e,c,b,d,j,g,f,k\rangle^{4}, \langle a,b,c,d,j,g,f,k\rangle^{3}, \langle a,c,i,d,e,b,c,d,j,f,g,k\rangle^{3}, \langle a,b,c,d,e,b,c,d,j,f,g,k\rangle^{3}, \langle a,i,c,d,e,b,c,d,j,f,g,k\rangle^{3}, \langle a,b,c,d,e,c,b,d,j,f,g,k\rangle^{3}, \langle a,c,i,d,e,i,c,d,j,f,g,k\rangle^{2}, \langle a,c,i,d,e,i,c,d,e,b,c,d,e,b,c,d,j,g,f,k\rangle^{2}, \langle a,b,c,d,e,i,c,d,j,g,f,k\rangle^{2}, \langle a,c,b,d,e,c,i,d,j,g,f,k\rangle^{2}, \langle a,c,i,d,e,c,i,d,j,f,g,k\rangle^{2}, \langle a,c,b,d,e,b,c,d,e,c,b,d,j,g,f,k\rangle^{2}, \langle a,c,i,d,e,c,b,d,j,g,f,k\rangle^{2}, \langle a,c,b,d,e,i,c,d,j,f,g,k\rangle^{2}, \langle a,c,i,d,e,i,c,d,e,i,c,d,j,g,f,k\rangle^{1}, \langle a,c,b,d,e,c,i,d,j,f,g,k\rangle^{1}, \langle a,i,c,d,e,c,i,d,e,c,i,d,e,i,c,d,e,i,c,d,j,g,f,k\rangle^{1}, \langle a,i,c,d,e,c,i,d,j,f,g,k\rangle^{1}, \langle a,c,i,d,e,c,i,d,j,g,f,k\rangle^{1}]$$

$$A^1 = \{a, b, d, e, i\}$$

$$A^2 = \{a, c, d, e\}$$

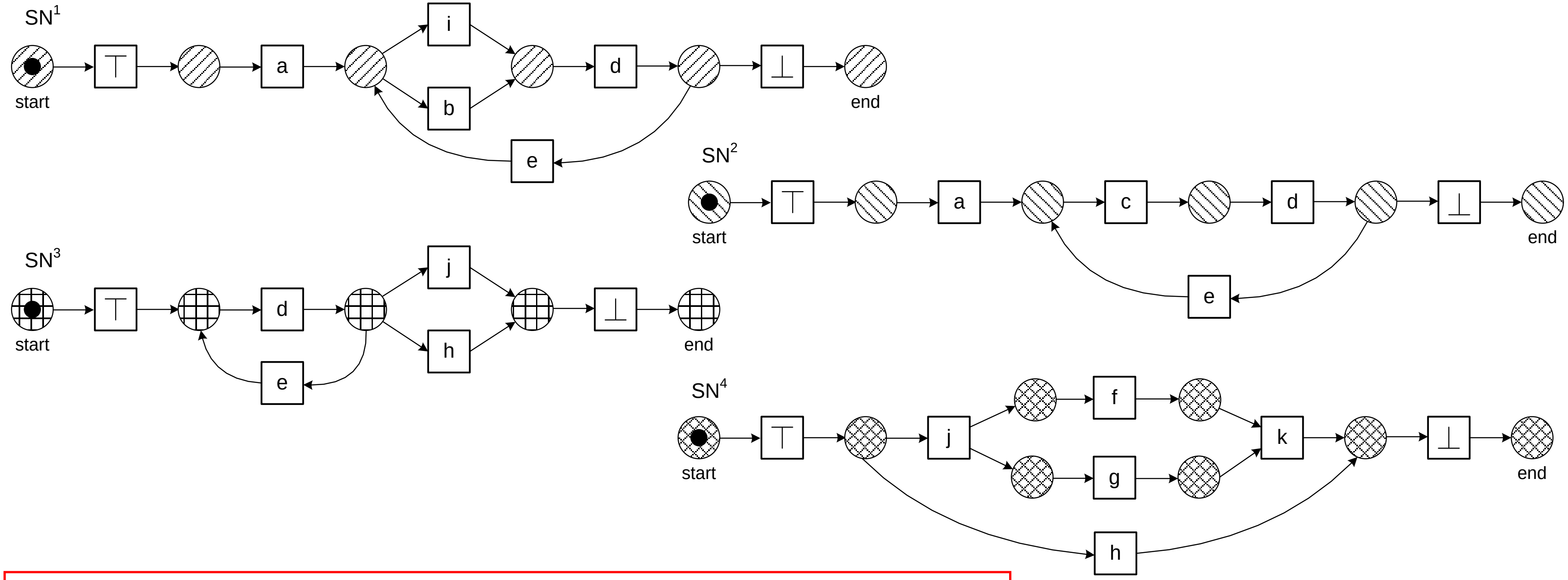
$$A^3 = \{d, e, h, j\}$$

$$A^4 = \{f, g, h, j, k\}$$

- **Let's use the Alpha algorithm**
- **Alpha algorithm is not very suitable for discovering transition bordered subnets.**
- **By adding start and end activities we get (in this case) perfectly fitting subnets.**

$$L' = [\langle\top\rangle \cdot \sigma \cdot \langle\bot\rangle \mid \sigma \in L_o] = [\langle\top,a,b,c,d,h,\bot\rangle^{30}, \langle\top,a,c,b,d,e,b,c,d,j,g,f,k,\bot\rangle^{28}, \langle\top,a,c,b,d,e,c,b,d,h,\bot\rangle^{28}, \langle\top,a,c,b,d,e,c,b,d,j,f,g,k,\bot\rangle^{27}, \langle\top,a,b,c,d,e,c,i,d,j,g,f,k,\bot\rangle^{26}, \ldots].$$

Discover model per sublog (Alpha algorithm)



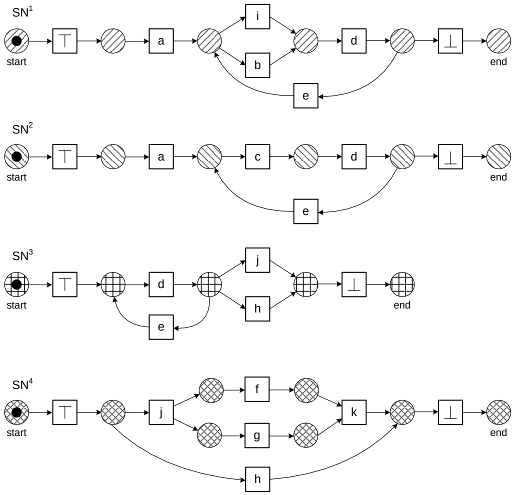
$$L^1 = [\langle \top, a, b, d, \bot \rangle^{30}, \langle \top, a, b, d, e, b, d, \bot \rangle^{28}, \langle \top, a, b, d, e, b, d, \bot \rangle^{28}, \langle \top, a, b, d, e, b, d, \bot \rangle^{27}, \langle \top, a, b, d, e, i, d, \bot \rangle^{26}, \ldots]$$

$$L^2 = [\langle \top, a, c, d, \bot \rangle^{30}, \langle \top, a, c, d, e, c, d, \bot \rangle^{28}, \langle \top, a, c, d, e, c, d, \bot \rangle^{28}, \langle \top, a, c, d, e, c, d, \bot \rangle^{27}, \langle \top, a, c, d, e, c, d, \bot \rangle^{26}, \ldots]$$

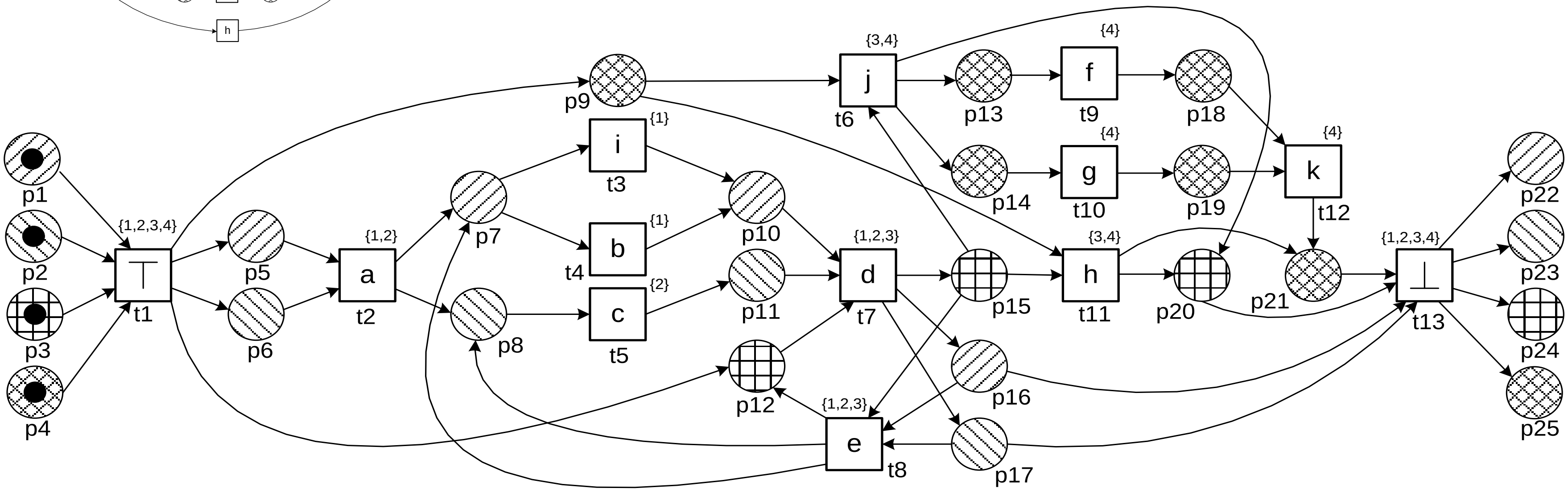
$$L_3 = [\langle \top, d, h, \bot \rangle^{30}, \langle \top, d, e, d, j, \bot \rangle^{28}, \langle \top, d, e, d, h, \bot \rangle^{28}, \langle \top, d, e, d, j, \bot \rangle^{27}, \langle \top, d, e, d, j, \bot \rangle^{26}, \ldots]$$

$$L_4 = [\langle \top, h, \bot \rangle^{30}, \langle \top, j, g, f, k, \bot \rangle^{28}, \langle \top, h, \bot \rangle^{28}, \langle \top, j, f, g, k, \bot \rangle^{27}, \langle \top, j, g, f, k, \bot \rangle^{26}, \ldots]$$

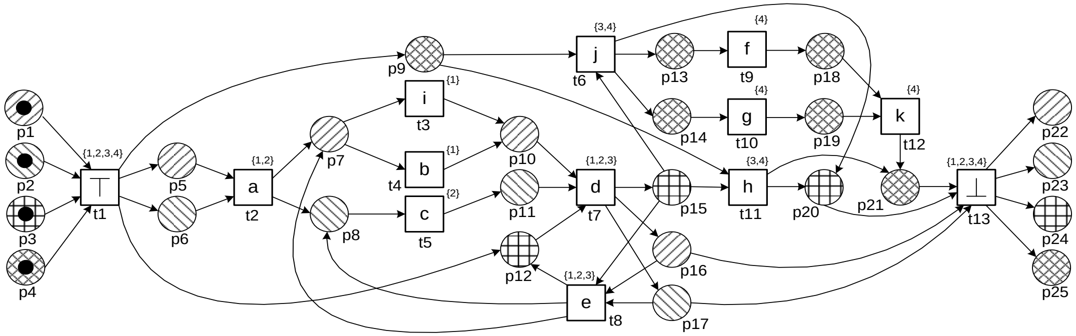
Merge models

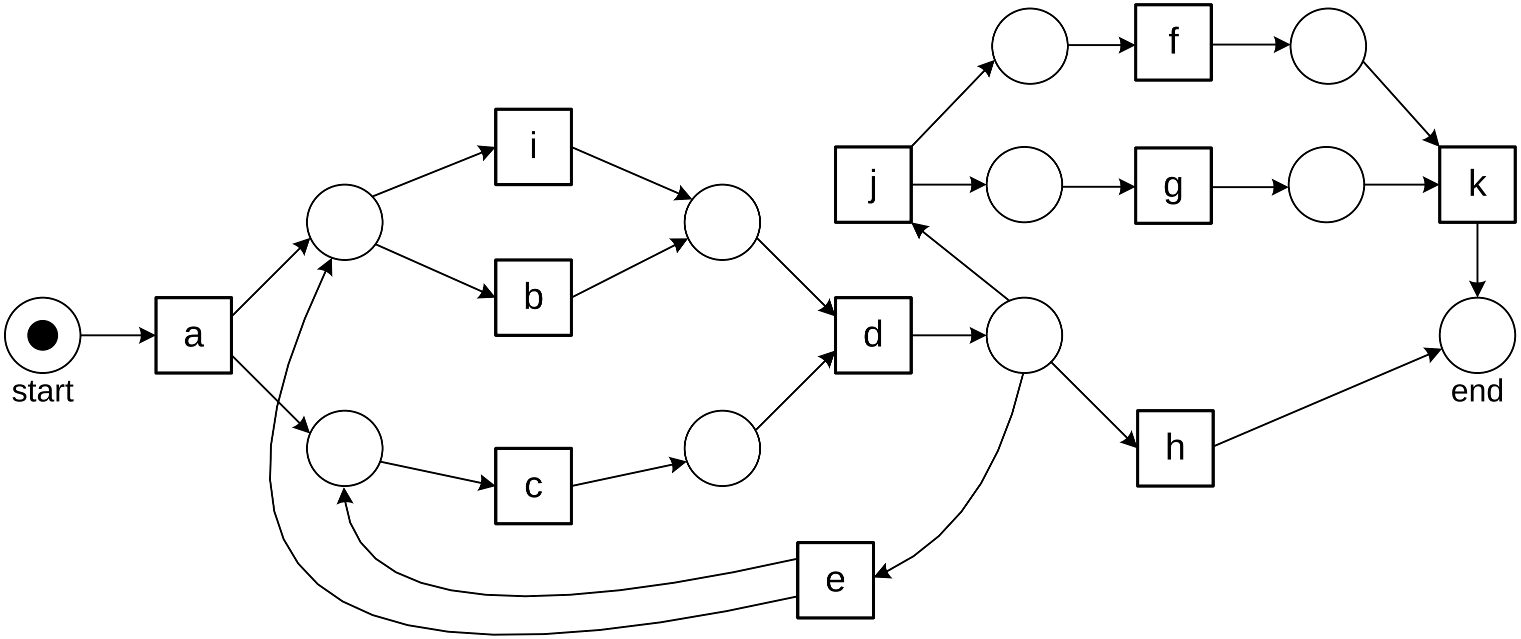


L is perfectly fitting SN because each projected log is L^i is perfectly fitting SN^i



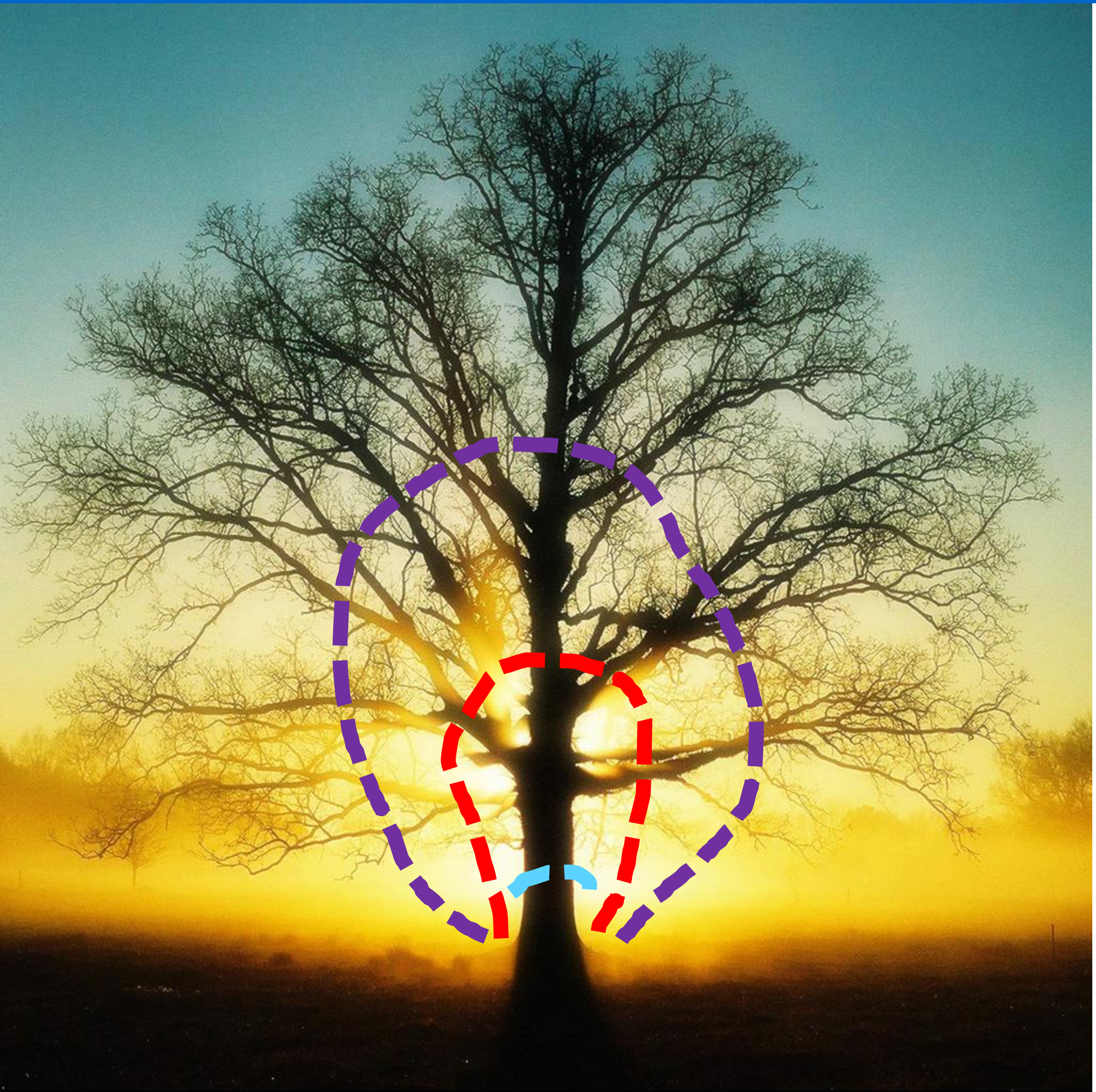
Simplify by removing redundant places and initialization





Log is perfectly fitting this model !

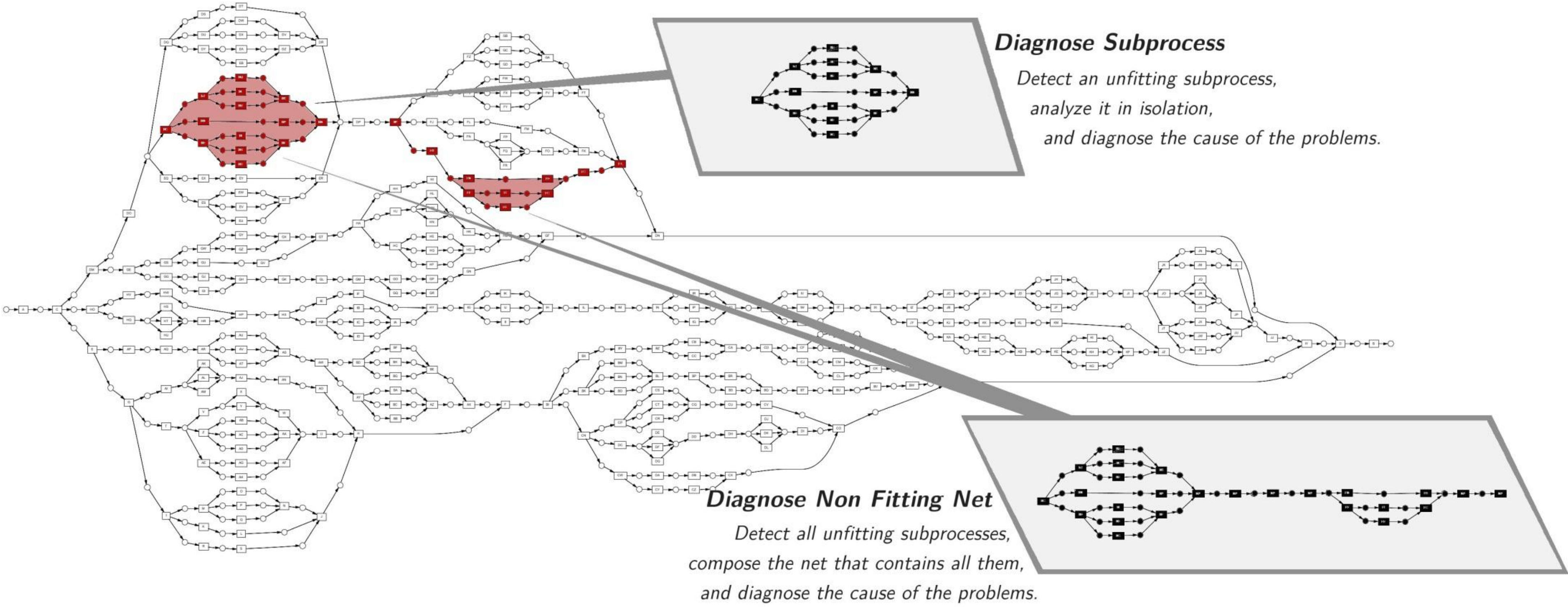
SESEs, Passages, and general result all define trees on process graph/event log



- Few large process mining tasks or many smaller process mining tasks.
- Choice of level is driven by computation time and desired diagnostics!
- All implemented in ProM.

ProM
process mining workbench

Diagnosis



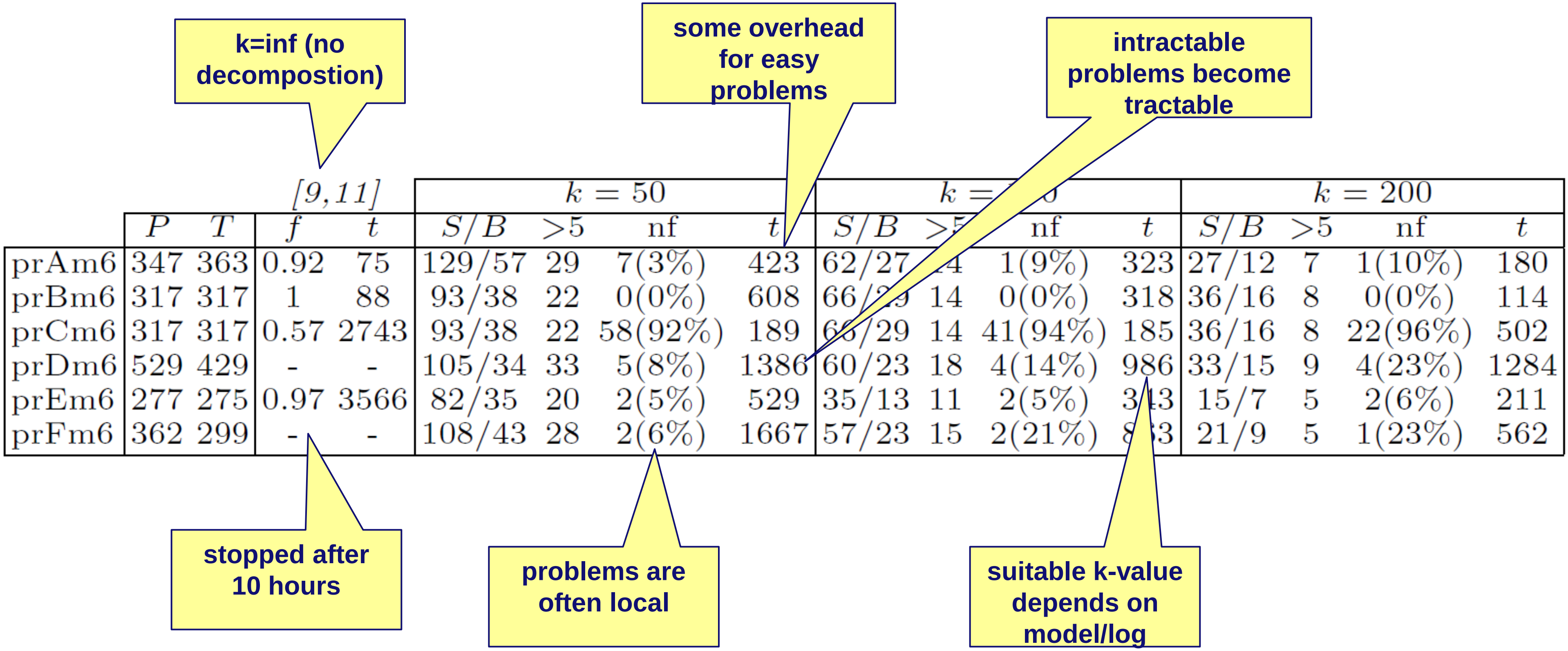
Diagnose Subprocess

Detect an unfitting subprocess,
analyze it in isolation,
and diagnose the cause of the problems.

Diagnose Non Fitting Net

Detect all unfitting subprocesses,
compose the net that contains all them,
and diagnose the cause of the problems.

Example: Conformance Checking based on SESEs



k=inf (no decompostion)

some overhead for easy problems

intractable problems become tractable

			[9,11]		k = 50				k = 100				k = 200			
	P	T	f	t	S/B	>5	nf	t	S/B	>5	nf	t	S/B	>5	nf	t
prAm6	347	363	0.92	75	129/57	29	7(3%)	423	62/27	14	1(9%)	323	27/12	7	1(10%)	180
prBm6	317	317	1	88	93/38	22	0(0%)	608	66/29	14	0(0%)	318	36/16	8	0(0%)	114
prCm6	317	317	0.57	2743	93/38	22	58(92%)	189	66/29	14	41(94%)	185	36/16	8	22(96%)	502
prDm6	529	429	-	-	105/34	33	5(8%)	1386	60/23	18	4(14%)	986	33/15	9	4(23%)	1284
prEm6	277	275	0.97	3566	82/35	20	2(5%)	529	35/13	11	2(5%)	343	15/7	5	2(6%)	211
prFm6	362	299	-	-	108/43	28	2(6%)	1667	57/23	15	2(21%)	863	21/9	5	1(23%)	562

stopped after 10 hours

problems are often local

suitable k-value depends on model/log

k = maximal number of arcs in one SESE, P = # places, T = # transitions, f = fitness, t = time in seconds, S = # transition bounded SESEs, B = # bridges, >5 = # more than 5 arcs, nf = # non-fitting parts.

Some pointers

- Wil M.P. van der Aalst. Decomposing Petri Nets for Process Mining: A Generic Approach. BPM Center Report BPM-12-20, BPMcenter.org, 2012
- Wil M. P. van der Aalst: Distributed Process Discovery and Conformance Checking. FASE 2012: 1-25
- Wil M. P. van der Aalst: Decomposing Process Mining Problems Using Passages. Petri Nets 2012: 72-91
- H. M. W. (Eric) Verbeek, Wil M. P. van der Aalst: An Experimental Evaluation of Passage-Based Process Discovery. Business Process Management Workshops 2012: 205-210
- Wil M.P. van der Aalst and Eric Verbeek. Process Discovery and Conformance Checking Using Passages. BPM Center Report BPM-12-21, BPMcenter.org, 2012
- J. Munoz-Gama, J. Carmona, W. van der Aalst: Hierarchical Conformance Checking of Process Models Based on Event Logs. In: Applications and Theory of Petri Nets, 2013

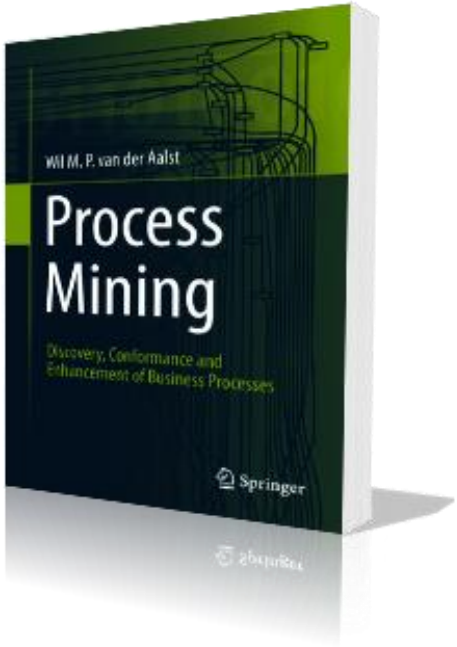
Conclusion

Conclusion (1/2)

- Alignments are essential for relating observed and modeled behavior!
- Conformance has (at least) four dimensions!
- Representational bias is important (and should not be confused with visualization)!
- New questions are emerging:
  - mediating between a reference model and observed behavior
  - discovering configurable process models
- Decomposing process mining problems to deal with Big Data.

Conclusion (2/2)

Still many challenging and highly relevant open problems in process mining!

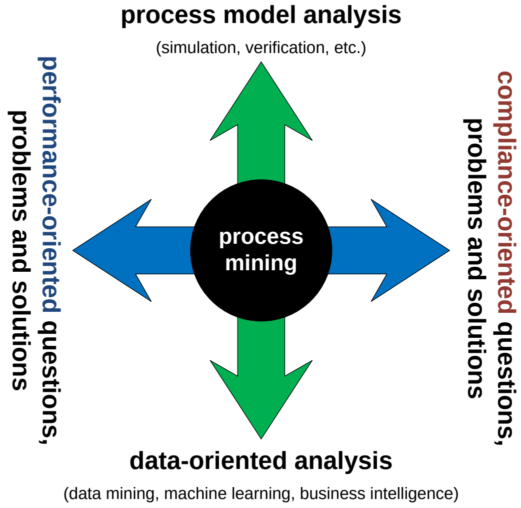


Wil M. P. van der Aalst

Process Mining

Discovery, Conformance and Enhancement of Business Processes

Springer



process model analysis

(simulation, verification, etc.)

performance-oriented questions, problems and solutions

process mining

compliance-oriented questions, problems and solutions

data-oriented analysis

(data mining, machine learning, business intelligence)

Harvard Business Review

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

ProM

process mining workbench

processmining.org