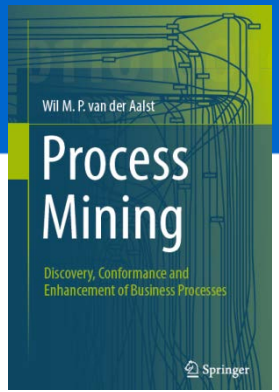
Process Mining: Data Science in Action

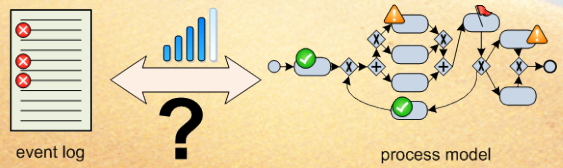
Conformance Checking Using Causal Footprints

prof.dr.ir. Wil van der Aalst
www.processmining.org



Technische Universiteit Eindhoven University of Technology

Where innovation starts



causal footprints

Footprint of L_1

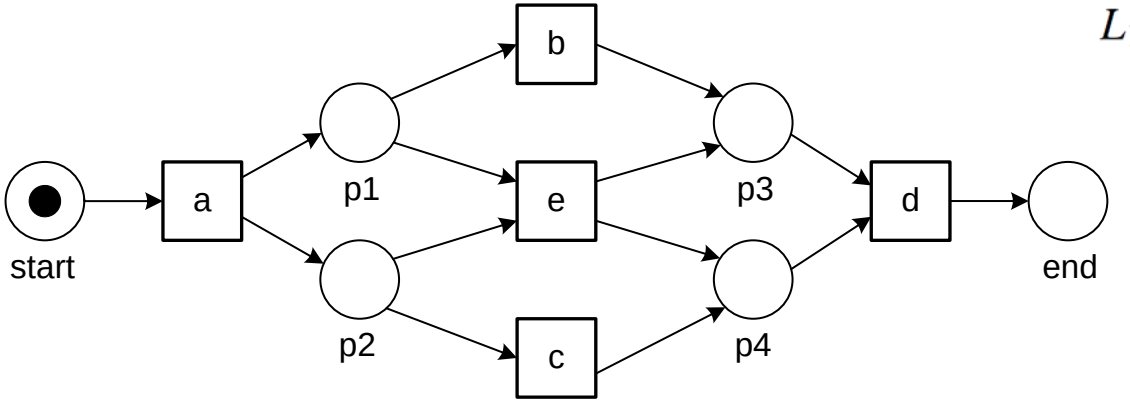
$$L_1 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^2, \langle a,e,d\rangle]$$

	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\#_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

- Direct succession: **x>y iff for some case x is directly followed by y.**
- Causality: **x→y iff x>y and not y>x.**
- Parallel: **x||y iff x>y and y>x**
- Choice: **x#y iff not x>y and not y>x.**

TU/e

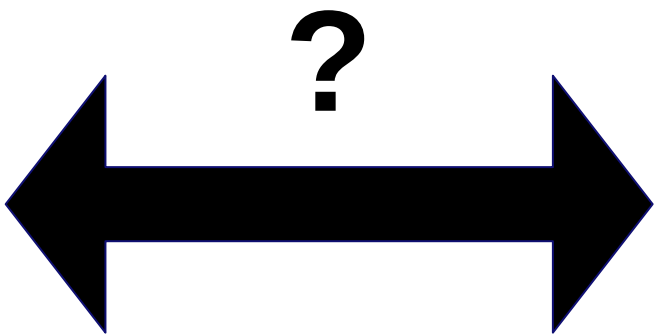
Discovered model has the same footprint

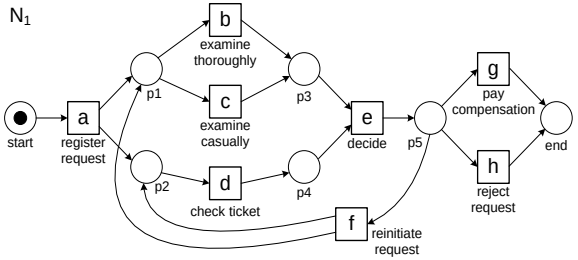


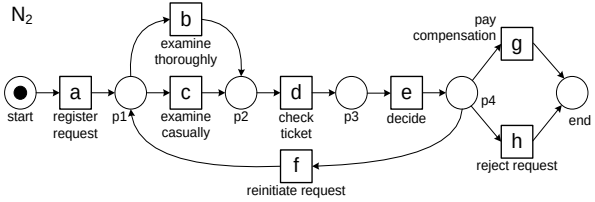
$$L_1 = [\langle a,b,c,d\rangle^3, \langle a,c,b,d\rangle^2, \langle a,e,d\rangle]$$

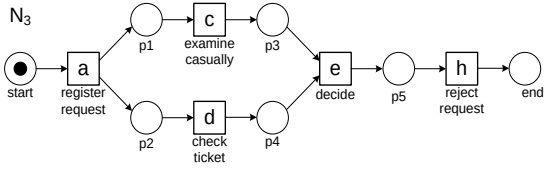
	a	b	c	d	e
a	$\#_{L_1}$	$\rightarrow_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$
b	$\leftarrow_{L_1}$	$\#_{L_1}$	$\|\|_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
c	$\leftarrow_{L_1}$	$\|\|_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$
d	$\#_{L_1}$	$\leftarrow_{L_1}$	$\leftarrow_{L_1}$	$\#_{L_1}$	$\leftarrow_{L_1}$
e	$\leftarrow_{L_1}$	$\#_{L_1}$	$\#_{L_1}$	$\rightarrow_{L_1}$	$\#_{L_1}$

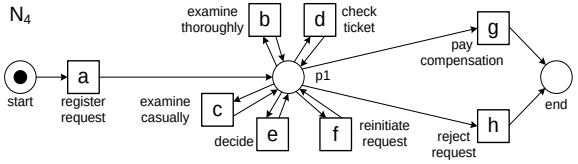
#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	





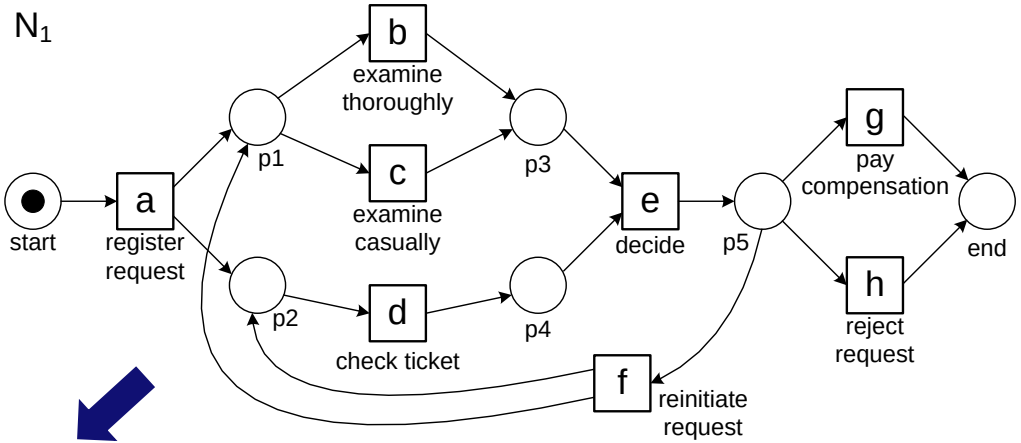






#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

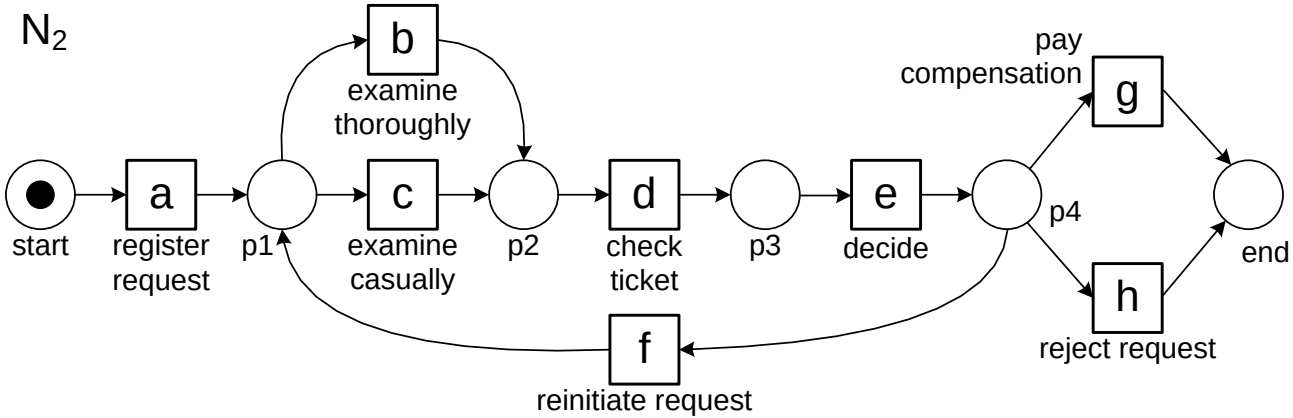
L_{full} and N_1



	a	b	c	d	e	f	g	h
a	#	→						#
b	←	#						#
c	←	#						#
d	←	‖	‖	#	→	←	#	#
e	#	←	←	←	#	→	→	→
f	#	→	→	→	←	#	#	#
g								#
h	#	#	#	#	←	#	#	#

footprint-based conformance = 1
(1 = perfect match)
(0 = worst match possible)

footprints of log and model coincide



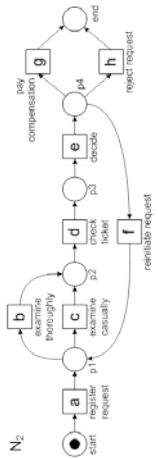
	a	b	c	d	e	f	g	h
a	#	$\rightarrow$	$\rightarrow$	#	#	#	#	#
b	$\leftarrow$	#	#	$\rightarrow$	#	$\leftarrow$	#	#
c	$\leftarrow$	#	#	$\rightarrow$	#	$\leftarrow$	#	#
d	#	$\leftarrow$	$\leftarrow$	#	$\rightarrow$	#	#	#
e	#	#	#	$\leftarrow$	#	$\rightarrow$	$\rightarrow$	$\rightarrow$
f	#	$\rightarrow$	$\rightarrow$	#	$\leftarrow$	#	#	#
g	#	#	#	#	$\leftarrow$	#	#	#
h	#	#	#	#	$\leftarrow$	#	#	#

L_{full}

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

	a	b	c	d	e	f	g	h
a	#	→	→	→	#	#	#	#
b	←	#	#	‖	→	←	#	#
c	←	#	#	‖	→	←	#	#
d	←	‖	‖	#	→	←	#	#
e	#	←	←	←	#	→	→	→
f	#	→	→	→	←	#	#	#
g	#	#	#	#	←	#	#	#
h	#	#	#	#	←	#	#	#

N_2



	a	b	c	d	e	f	g	h
a	#	→	→	#	#	#	#	#
b	←	#	#	→	#	←	#	#
c	←	#	#	→	#	←	#	#
d	#	←	←	#	→	#	#	#
e	#	#	#	←	#	→	→	→
f	#	→	→	#	←	#	#	#
g	#	#	#	#	←	#	#	#
h	#	#	#	#	←	#	#	#

Quantifying the differences

	a	b	c	d	e	f	g	h
a				$\rightarrow: \#$				
b				$\parallel : \rightarrow$	$\rightarrow: \#$			
c				$\parallel : \rightarrow$	$\rightarrow: \#$			
d	$\leftarrow: \#$	$\parallel : \leftarrow$	$\parallel : \leftarrow$			$\leftarrow: \#$		
e		$\leftarrow: \#$	$\leftarrow: \#$					
f				$\rightarrow: \#$				
g								
h								

$$1 - \frac{12}{64} = 0.8125$$

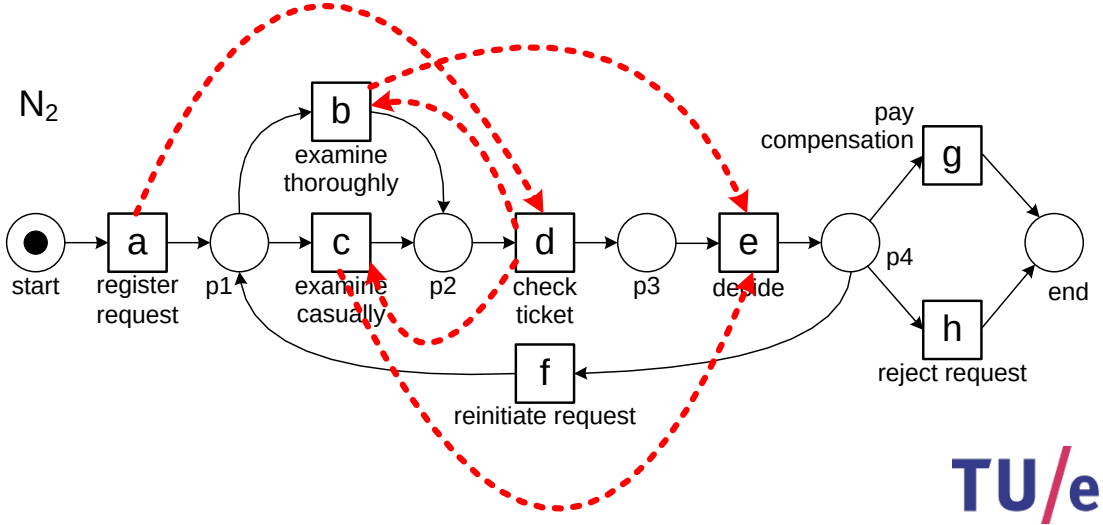
(x:y where x is in log and y in N_2)

footprint-based conformance

TU/e

Diagnostics

	a	b	c	d	e	f	g	h
a				$\rightarrow: \#$				
b				$\parallel : \rightarrow$	$\rightarrow: \#$			
c				$\parallel : \rightarrow$	$\rightarrow: \#$			
d	$\leftarrow: \#$	$\parallel : \leftarrow$	$\parallel : \leftarrow$			$\leftarrow: \#$		
e		$\leftarrow: \#$	$\leftarrow: \#$					
f				$\rightarrow: \#$				
g								
h								

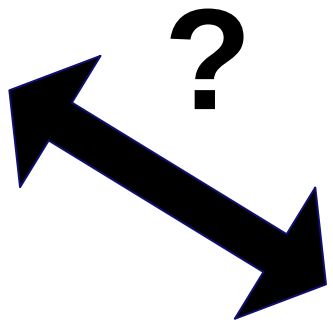


Question

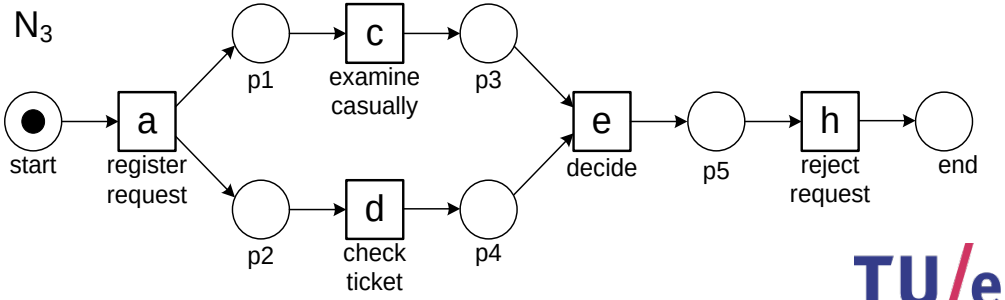
Estimate footprint-based conformance

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

	a	b	c	d	e	f	g	h
a	#	→	→	→	#	#	#	#
b	←	#	#	‖	→	←	#	#
c	←	#	#	‖	→	←	#	#
d	←	‖	‖	#	→	←	#	#
e	#	←	←	←	#	→	→	→
f	#	→	→	→	←	#	#	#
g	#	#	#	#	←	#	#	#
h	#	#	#	#	←	#	#	#



Estimate the fraction of matching cells in footprint matrices



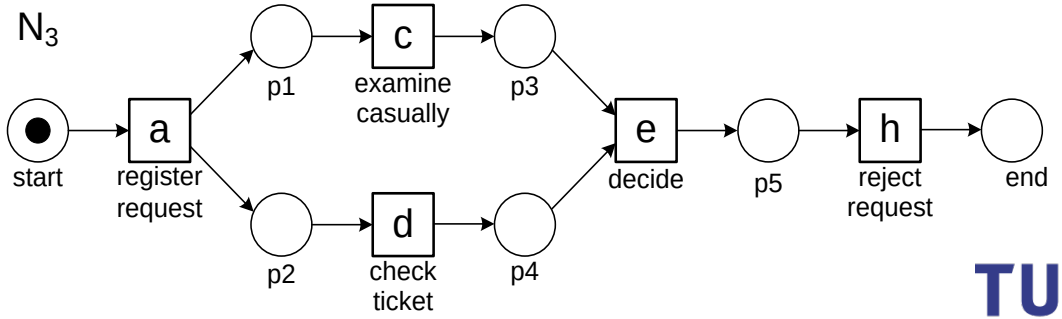
TU/e

Answer

	a	b	c	d	e	f	g	h
a	#	#	→	→	#	#	#	#
b	#	#	#	#	#	#	#	#
c	←	#	#	‖	→	#	#	#
d	←	#	‖	#	→	#	#	#
e	#	#	←	←	#	#	#	→
f	#	#	#	#	#	#	#	#
g	#	#	#	#	#	#	#	#
h	#	#	#	#	←	#	#	#

$$1 - \frac{16}{64} = 0.75$$

footprint-based conformance



TU/e

Question

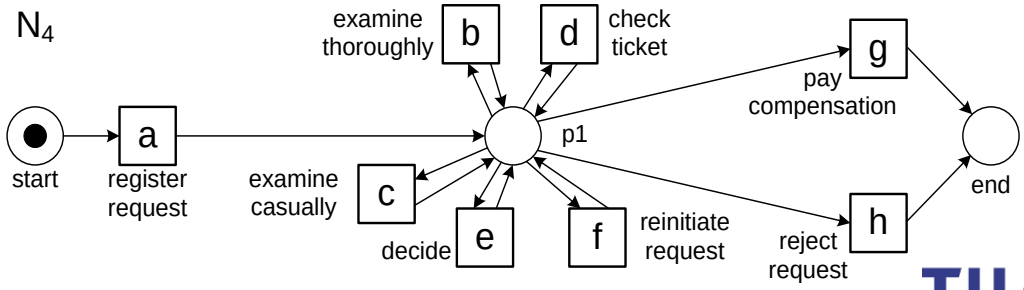
Estimate footprint-based conformance

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	

	a	b	c	d	e	f	g	h
a	#	→	→	→	#	#	#	#
b	←	#	#	‖	→	←	#	#
c	←	#	#	‖	→	←	#	#
d	←	‖	‖	#	→	←	#	#
e	#	←	←	←	#	→	→	→
f	#	→	→	→	←	#	#	#
g	#	#	#	#	←	#	#	#
h	#	#	#	#	←	#	#	#

Estimate the fraction of matching cells in footprint matrices

?



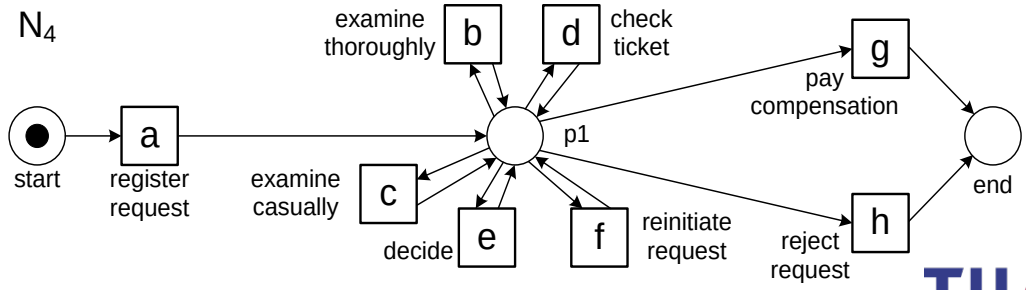
TU/e

Answer

	a	b	c	d	e	f	g	h
a	#	→	→	→	#	#	#	#
b	←	#	#	‖	#	#	#	#
c	←	#	#	‖	#	#	#	#
d	←	‖	‖	#	#	#	#	#
e	#	#	#	#	#	#	→	→
f	#	#	#	#	#	#	#	#
g	#	#	#	#	←	#	#	#
h	#	#	#	#	←	#	#	#

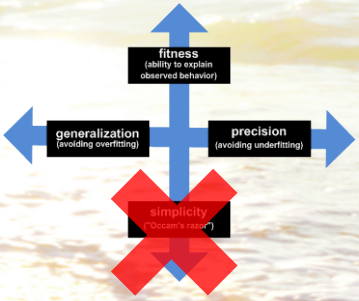
$$1 - \frac{45}{64} = 0.296875$$

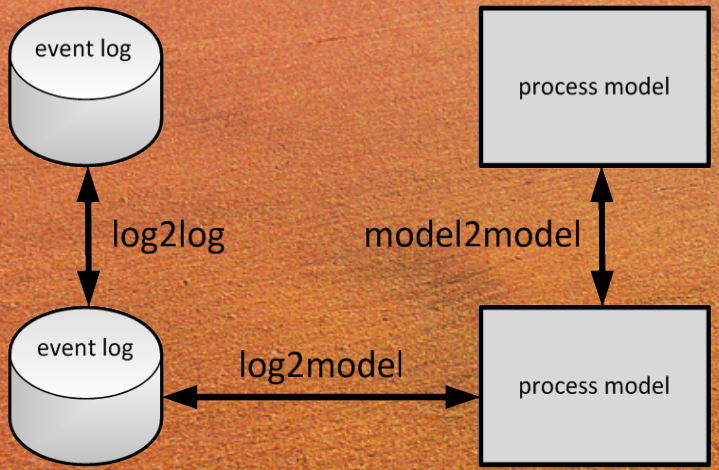
footprint-based conformance

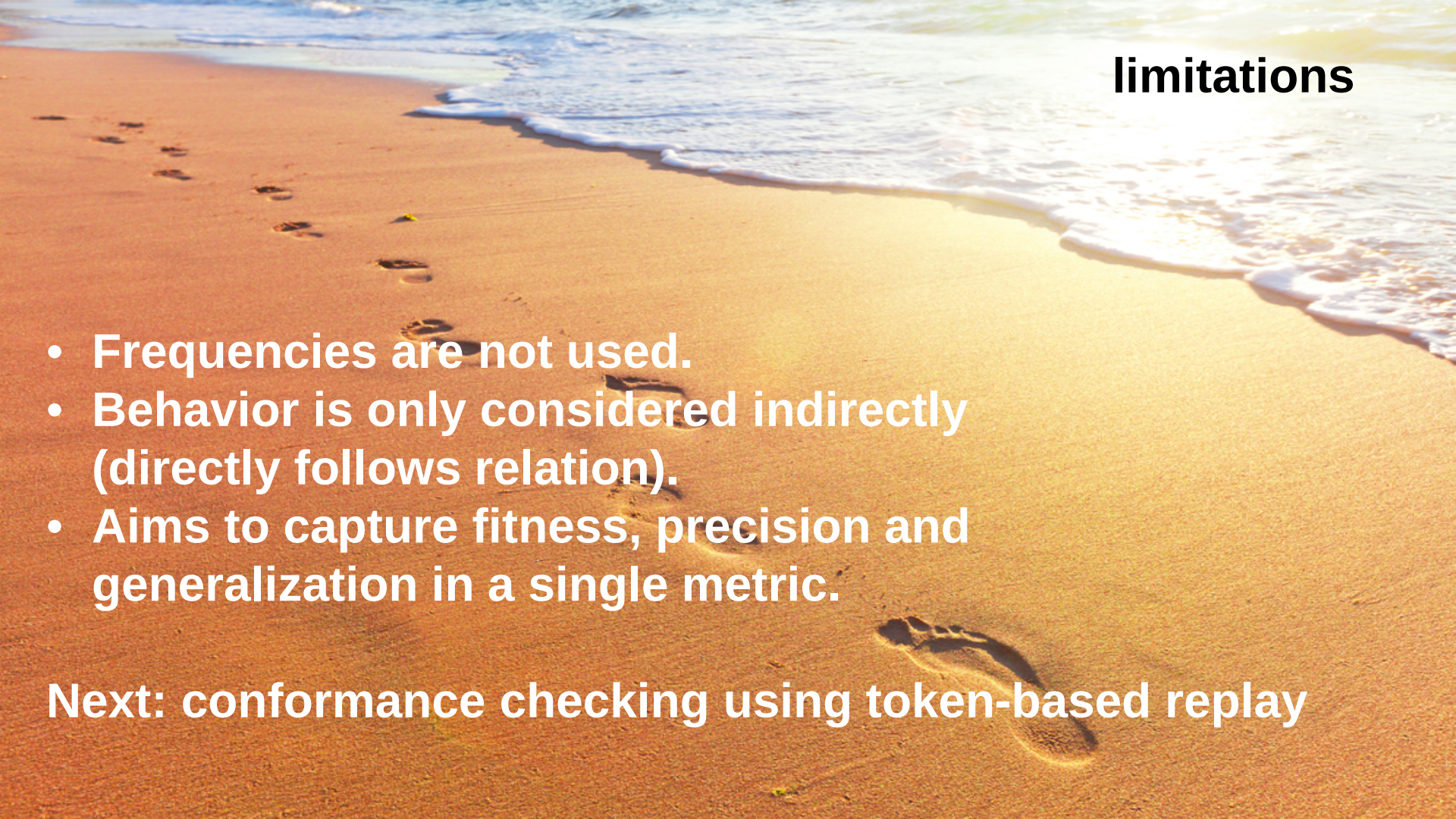


TU/e

flexible


fitness
(ability to explain observed behavior)
generalization
(avoiding overfitting)
precision
(avoiding underfitting)
simplicity
("Occam's razor")


event log
process model
log2log
model2model
event log
log2model
process model



limitations

- Frequencies are not used.
- Behavior is only considered indirectly (directly follows relation).
- Aims to capture fitness, precision and generalization in a single metric.

Next: conformance checking using token-based replay

Part I: Preliminaries

- Chapter 1: Introduction
- Chapter 2: Process Modeling and Analysis
- Chapter 3: Data Mining

Part II: From Event Logs to Process Models

- Chapter 4: Getting the Data
- Chapter 5: Process Discovery: An Introduction
- Chapter 6: Advanced Process Discovery Techniques

Part III: Beyond Process Discovery

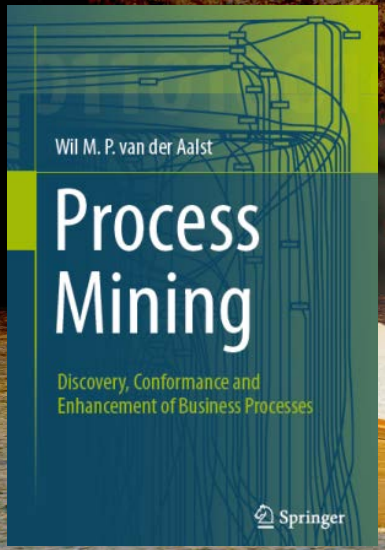
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

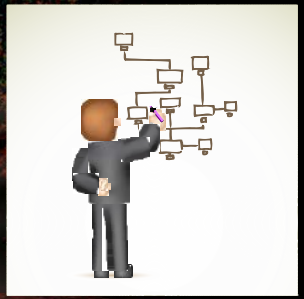
Part IV: Putting Process Mining to Work

- Chapter 10: Tool Support
- Chapter 11: Analyzing "Lasagna Processes"
- Chapter 12: Analyzing "Spaghetti Processes"

Part V: Reflection

- Chapter 13: Cartography and Navigation
- Chapter 14: Epilogue





TU/e