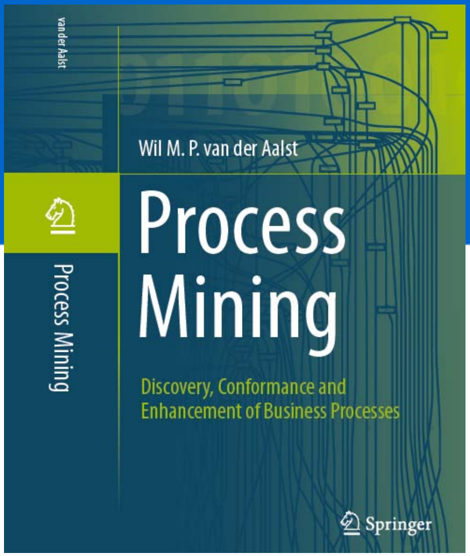
Chapter 8
Mining Additional Perspectives

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



TU/e Technische Universiteit
Eindhoven
University of Technology

Where innovation starts

Overview

Chapter 1
Introduction

Part I: Preliminaries

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"

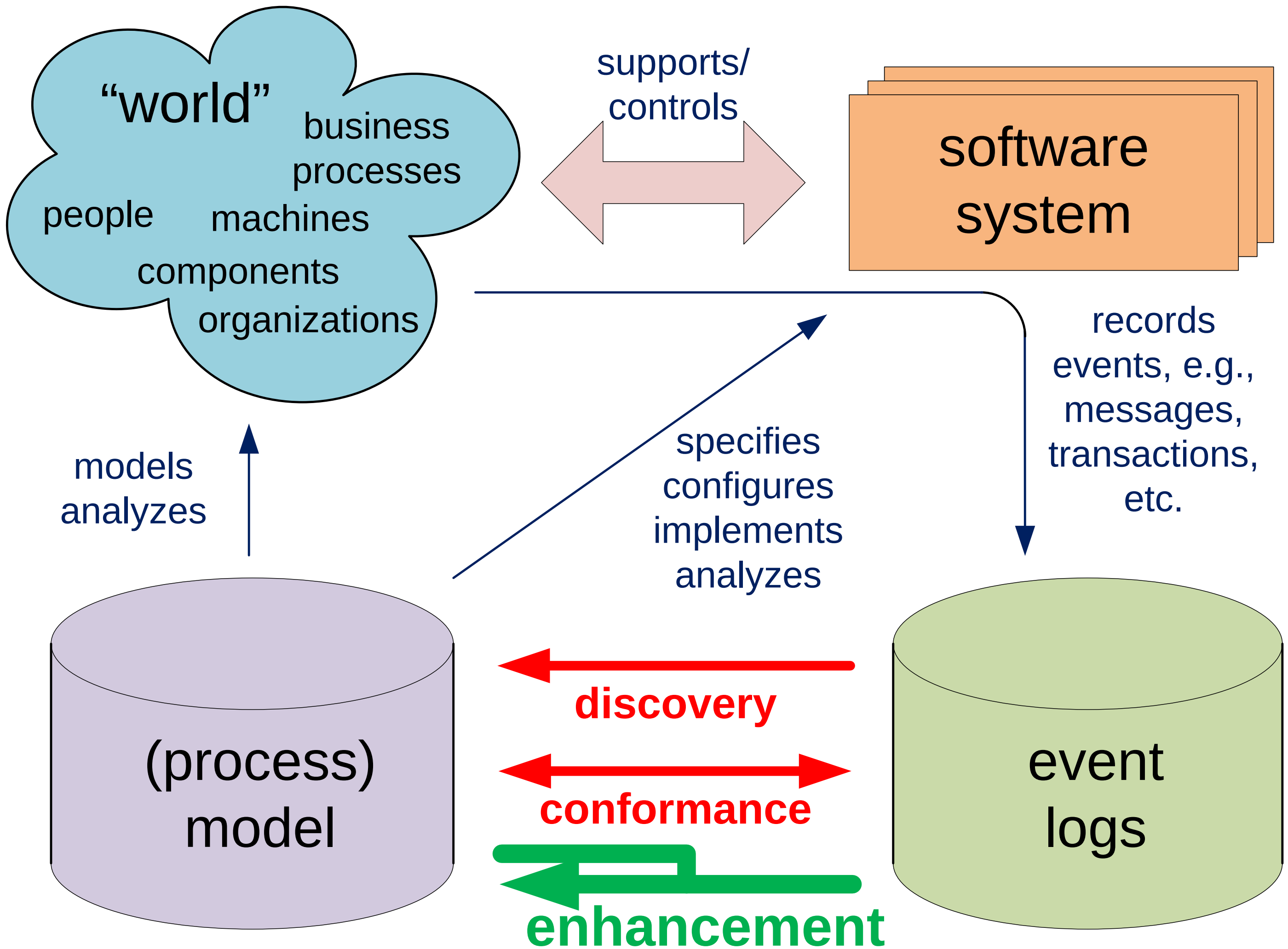
Part V: Reflection

Chapter 13
Cartography and Navigation

Chapter 14
Epilogue

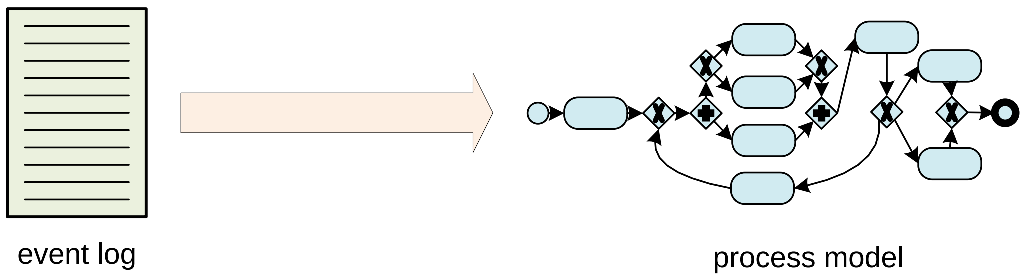
Mining additional perspectives

(one type of enhancement, cf. repair in context of conformance checking)

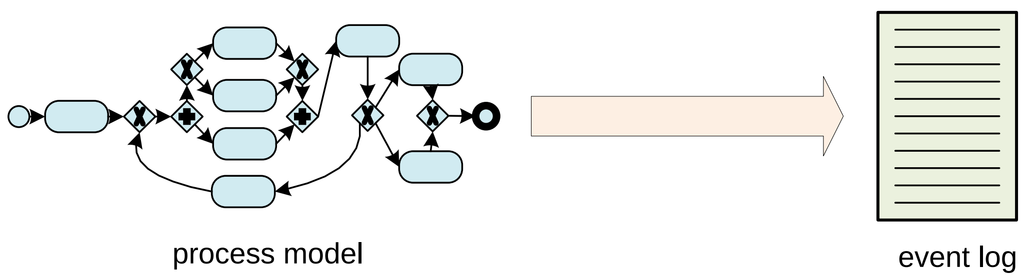


Replay: Connecting events to model elements is essential for process mining

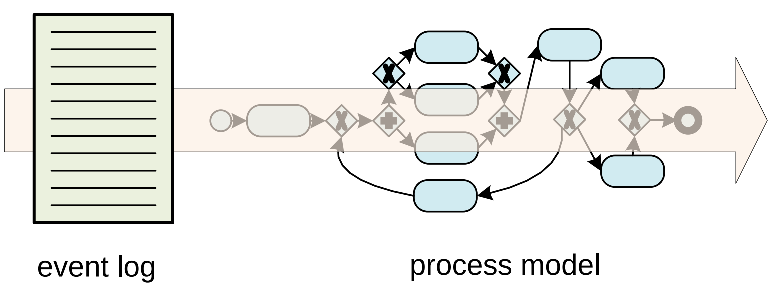
Play-In



Play-Out



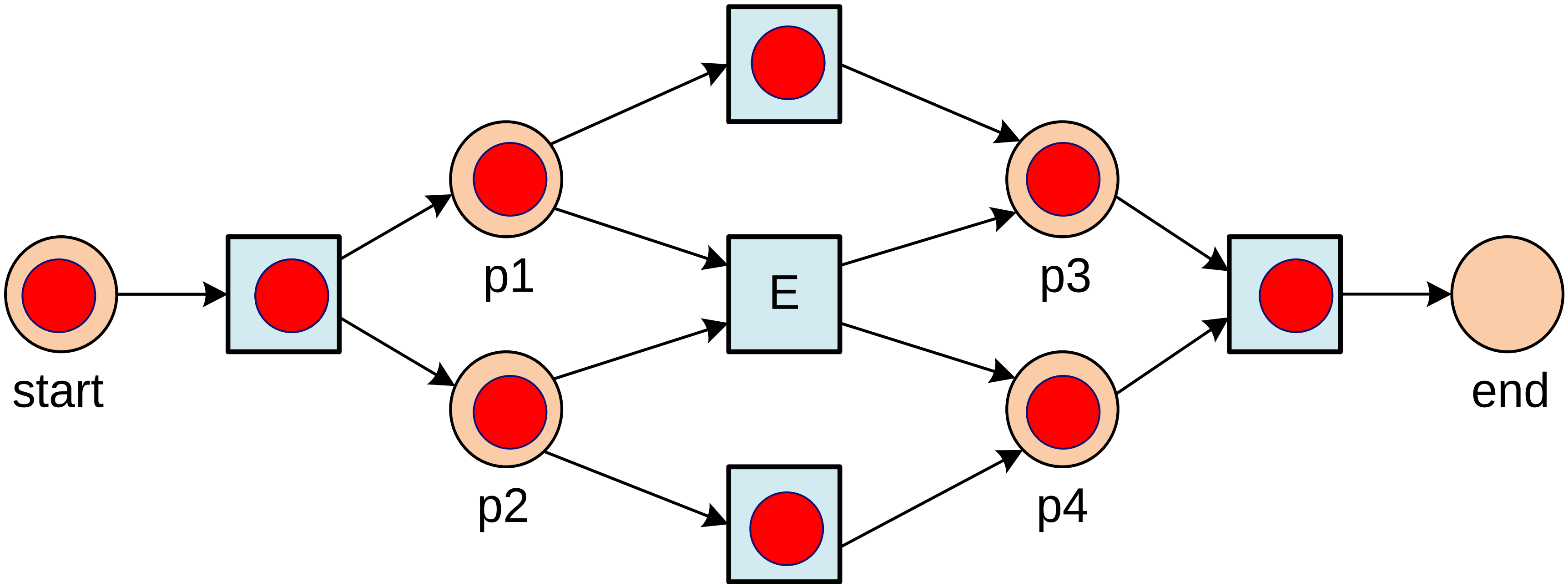
Replay



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

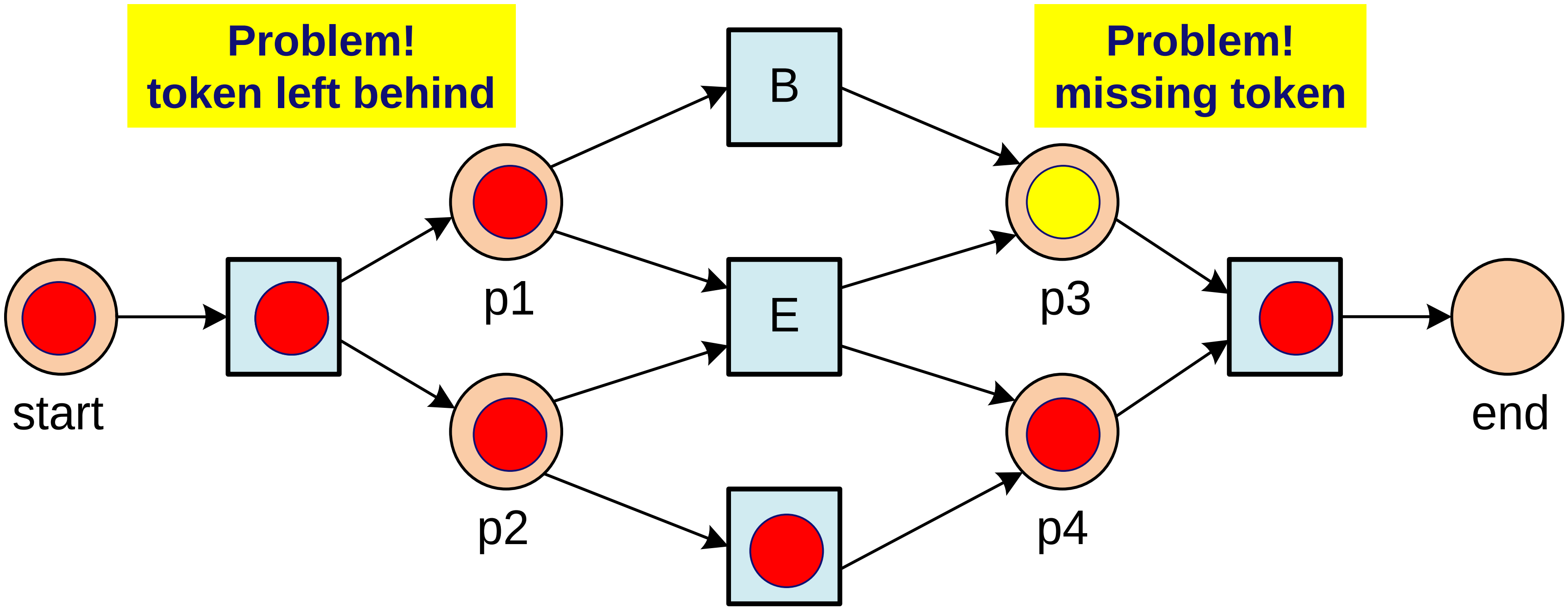
Remember: Replay!

A B C D



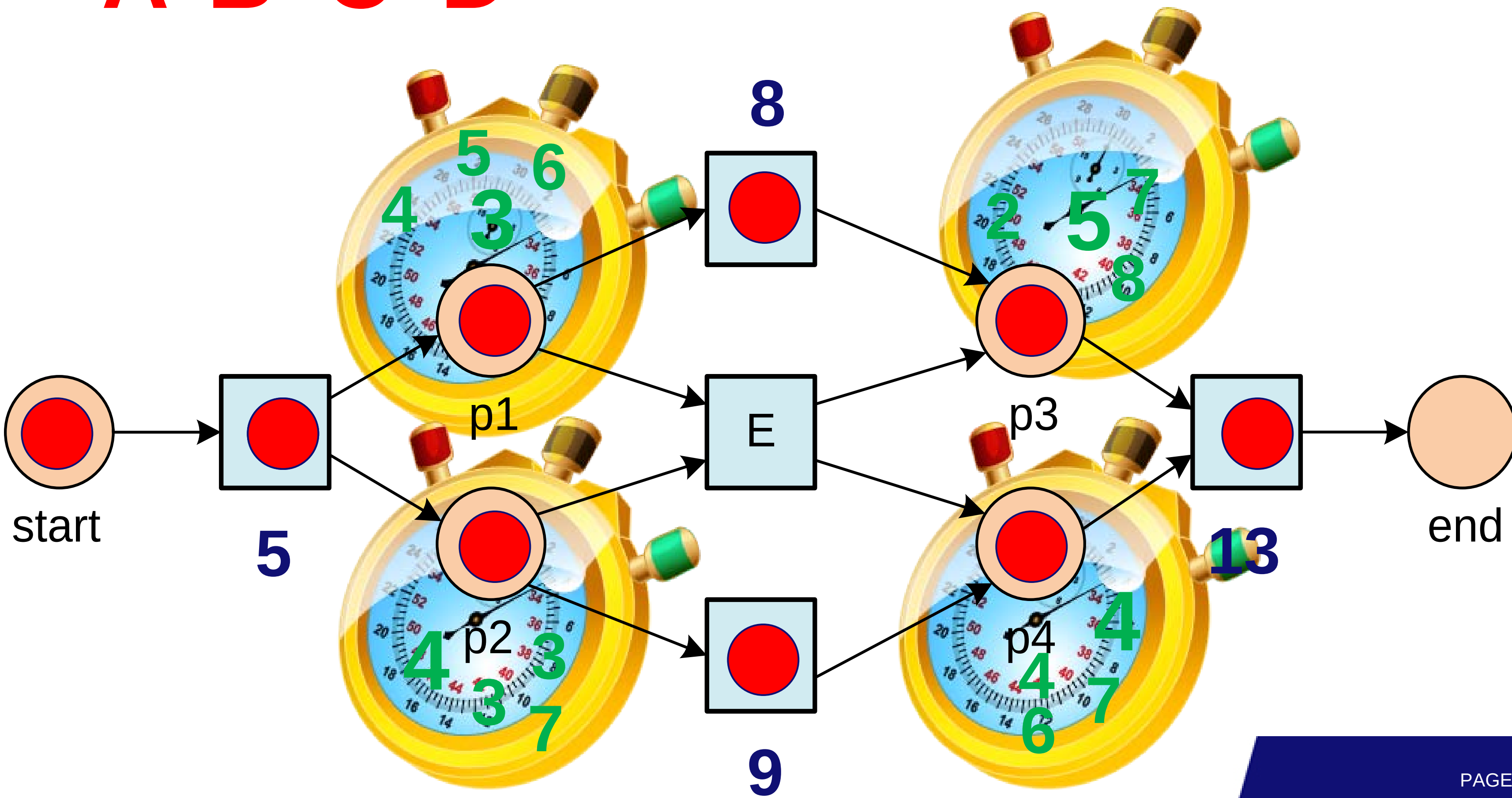
Replay can detect problems

A C D



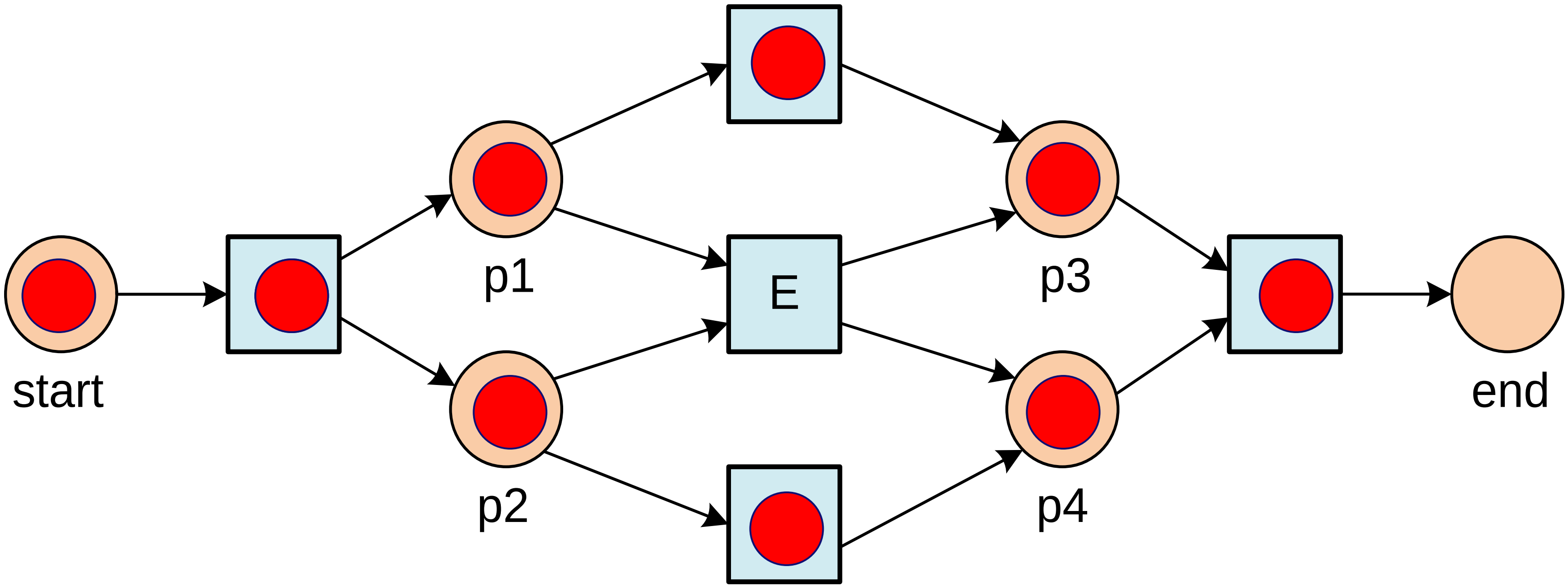
Replay can extract timing information

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



Decision mining: “Red” cases

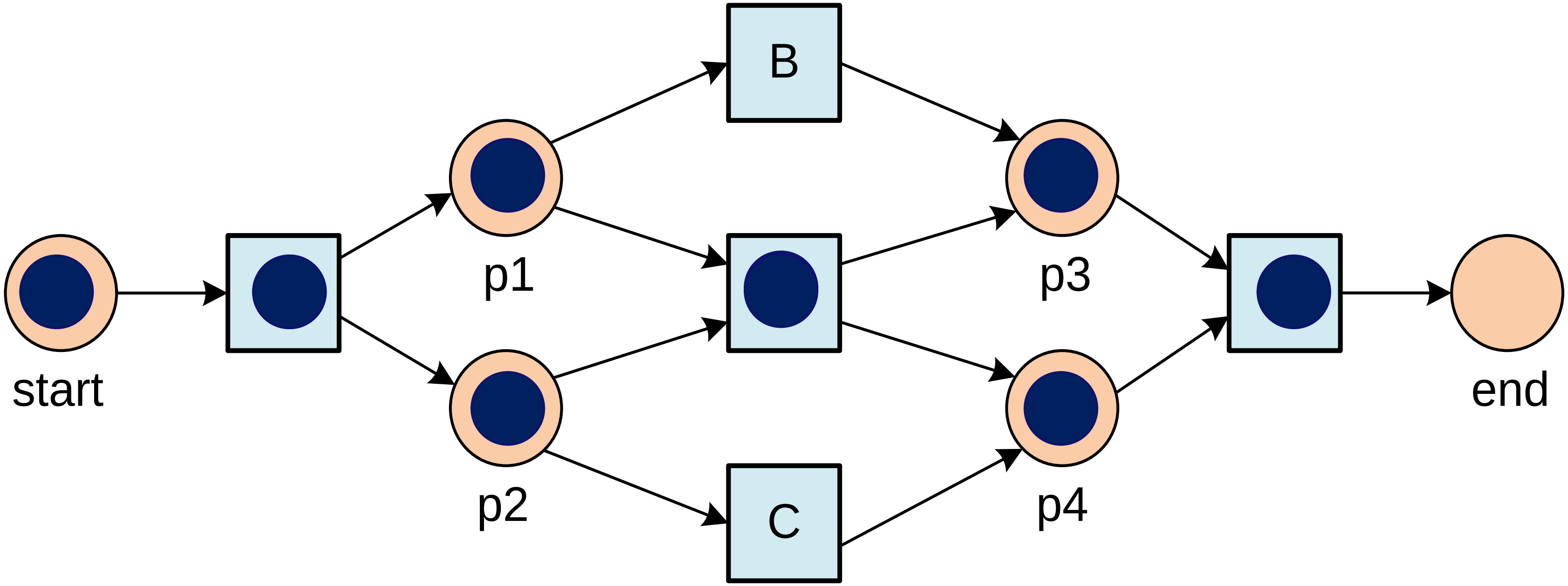
A B C D



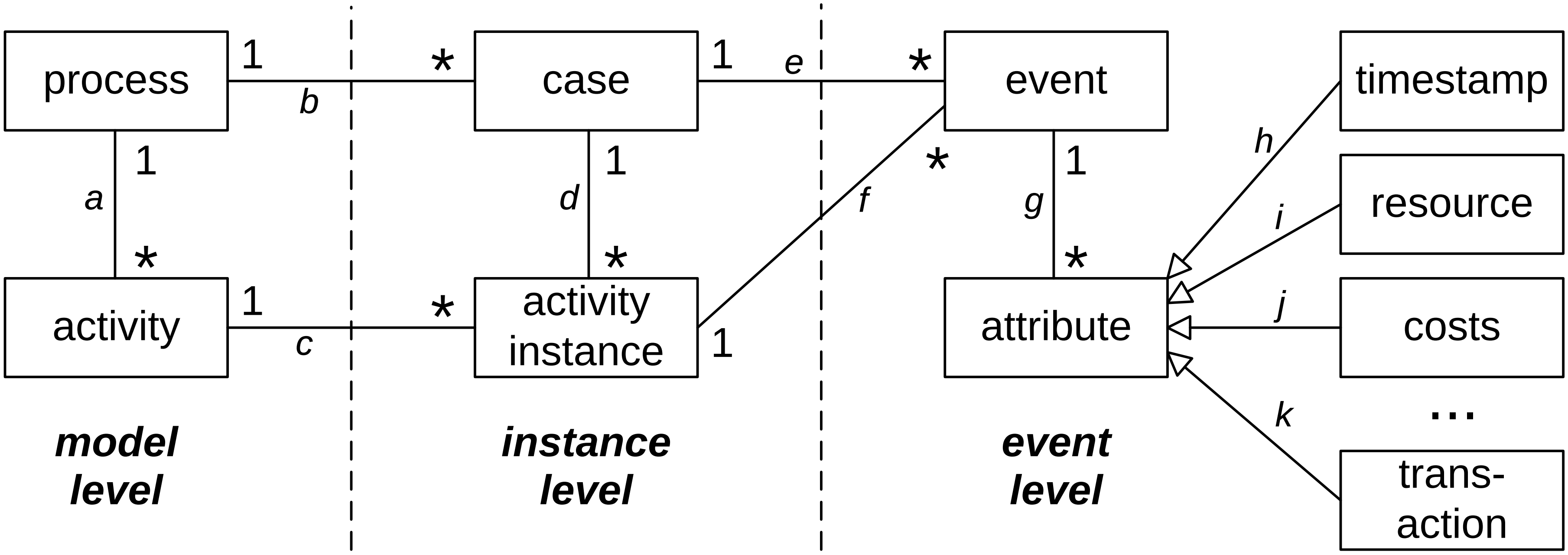
Decision mining: “Blue” cases

A E D

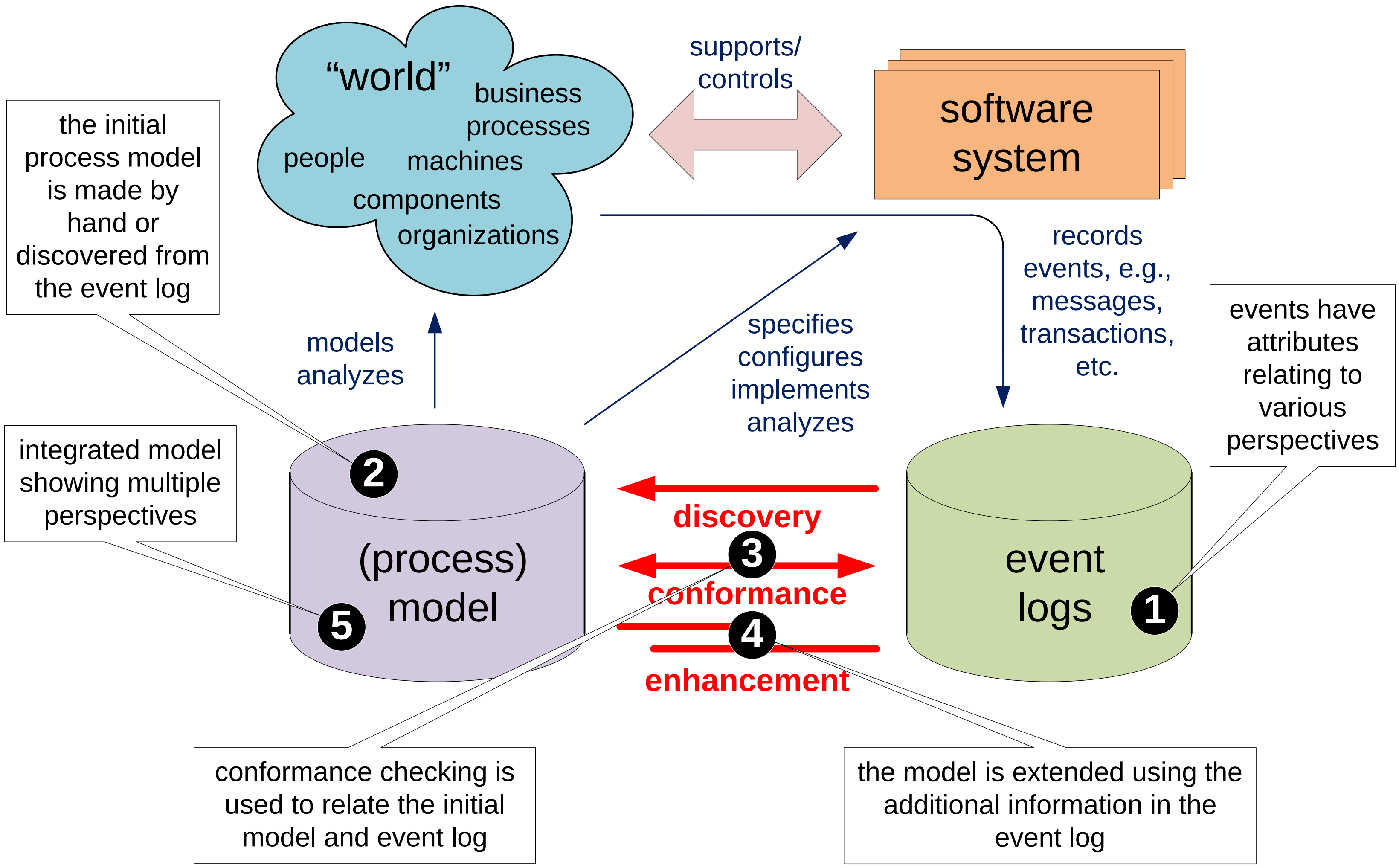
If red then B+C;
If blue then E;



Starting point: connected event log and model



Process



"world"
business processes
people
machines
components
organizations
supports/ controls
software system
records events, e.g., messages, transactions, etc.
models analyzes
specifies configures implements analyzes
the initial process model is made by hand or discovered from the event log
integrated model showing multiple perspectives
events have attributes relating to various perspectives
(process) model
event logs
discovery
conformance
enhancement
1
2
3
4
5
conformance checking is used to relate the initial model and event log
the model is extended using the additional information in the event log

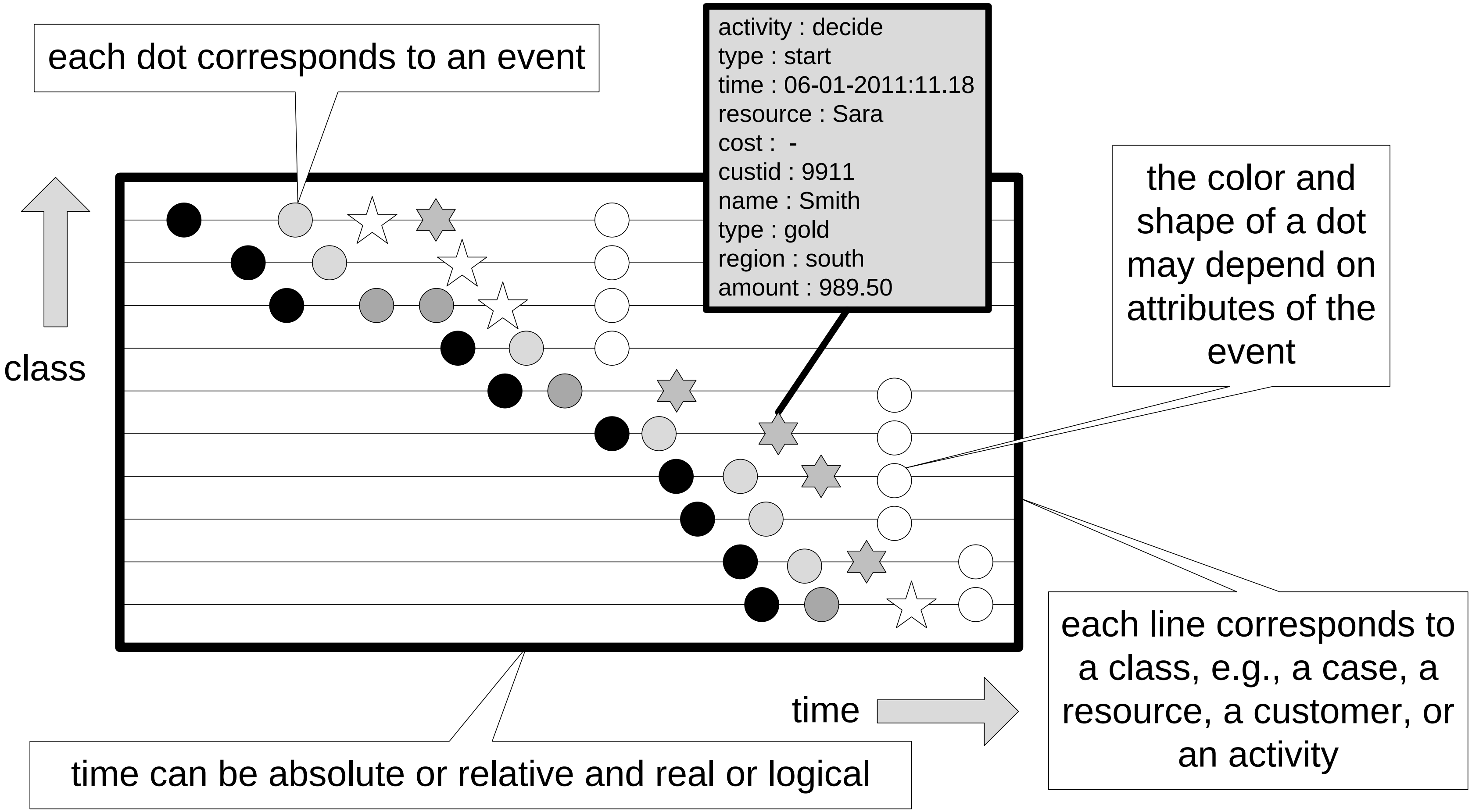
Attributes in event logs

case id	event id	properties				
		time	activity	trans	resource	cost
1	35654423	30-12-2010:11.02	register request	start	Pete	
	35654424	30-12-2010:11.08	register request	complete	Pete	50
	35654425	31-12-2010:10.06	examine thoroughly	start	Sue	
	35654427	31-12-2010:10.08	check ticket	start	Mike	
	35654428	31-12-2010:10.12	examine thoroughly	complete	Sue	400
	35654429	31-12-2010:10.20	check ticket	complete	Mike	100
	35654430	06-01-2011:11.18	decide	start	Sara	
	35654431	06-01-2011:11.22	decide	complete	Sara	200
	35654432	07-01-2011:14.24	reject request	start	Pete	
	35654433	07-01-2011:14.32	reject request	complete	Pete	200
2	35654483	30-12-2010:11.32	register request	start	Mike	
	35654484	30-12-2010:11.40	register request	complete	Mike	50
	35654485	30-12-2010:12.12	check ticket	start	Mike	
	35654486	30-12-2010:12.24	check ticket	complete	Mike	100
	35654487	30-12-2010:14.16	examine casually	start	Pete	
	35654488	30-12-2010:14.22	examine casually	complete	Pete	400
	35654489	05-01-2011:11.22	decide	start	Sara	
	35654490	05-01-2011:11.29	decide	complete	Sara	200
	35654491	08-01-2011:12.05	pay compensation	start	Ellen	
	35654492	08-01-2011:12.15	pay compensation	complete	Ellen	200
...	...	...	...	...	...	...

Cases may also have attributes

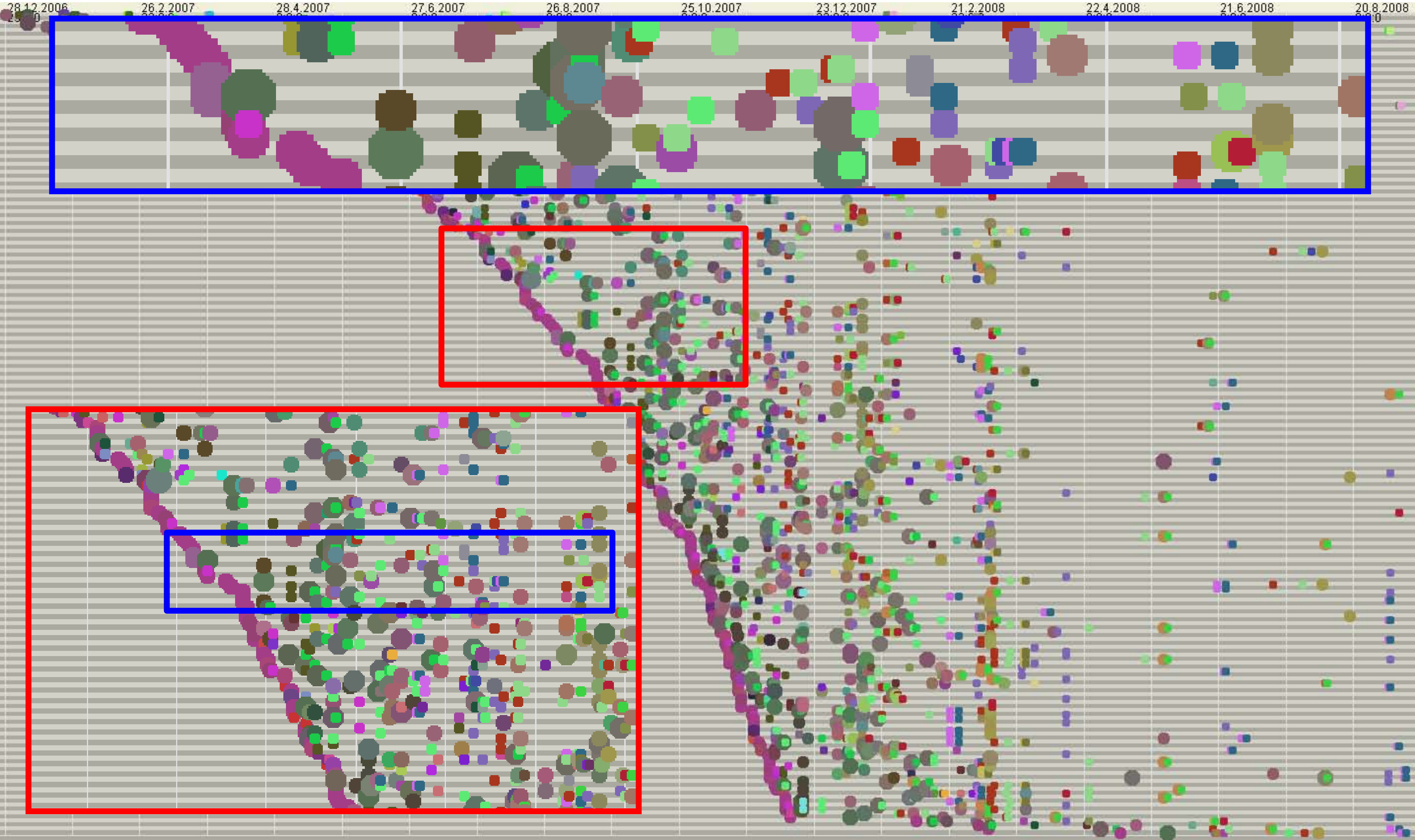
case id	custid	name	type	region	amount
1	9911	Smith	gold	south	989.50
2	9915	Jones	silver	west	546.00
3	9912	Anderson	silver	north	763.20
4	9904	Thompson	silver	west	911.70
5	9911	Smith	gold	south	812.10
6	9944	Baker	silver	east	788.00
7	9944	Baker	silver	east	792.80
8	9911	Smith	gold	south	544.70
...	...	...	...	...	...

Helicopter view: Dotted charts



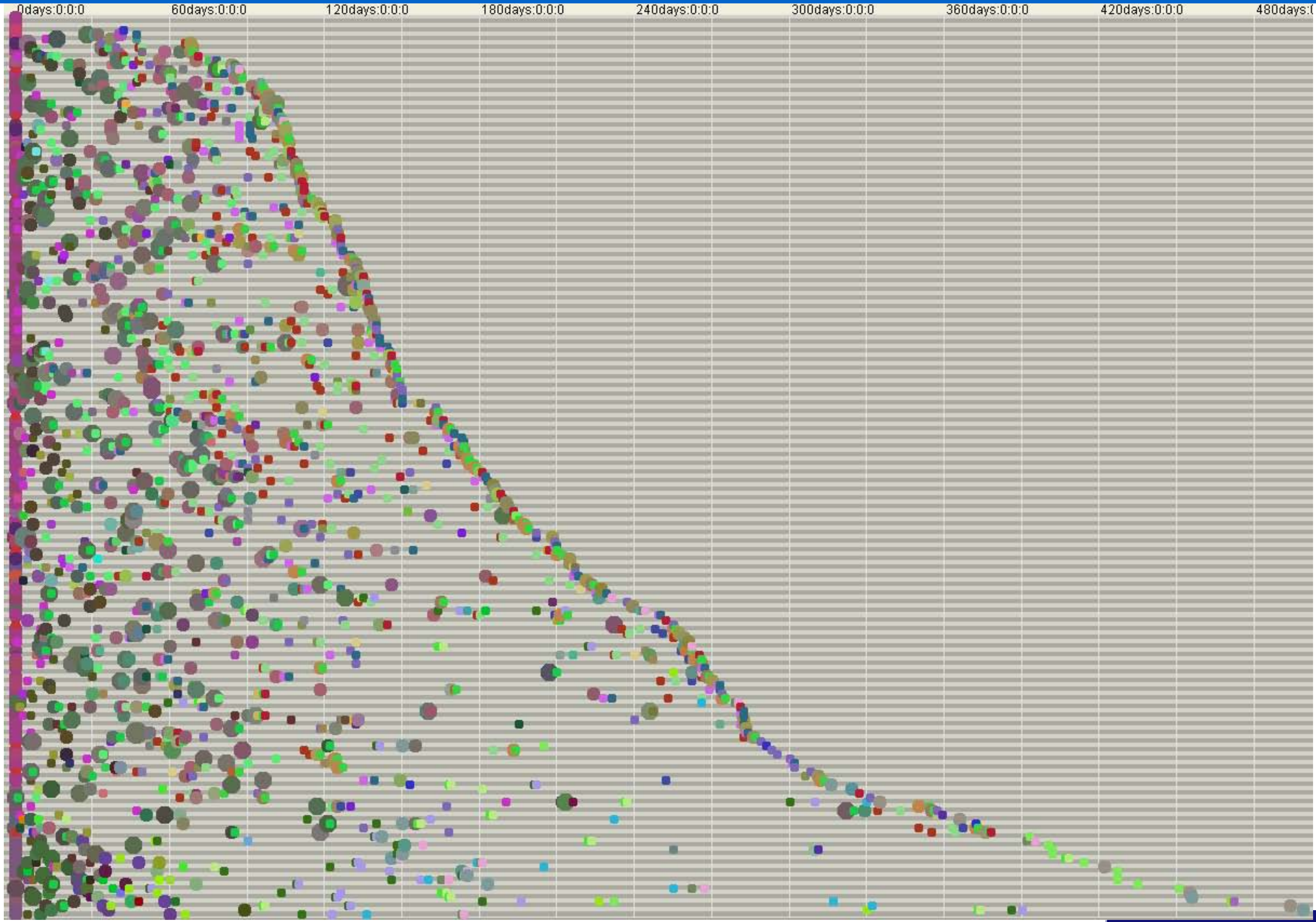
Dotted chart for a process of a housing agency using absolute time

Zooming in



28.12.2006
26.2.2007
28.4.2007
27.6.2007
26.8.2007
25.10.2007
23.12.2007
21.2.2008
22.4.2008
21.6.2008
20.8.2008

Same log, relative time



Organizational mining

case id	trace
1	$\langle a^{Pete}, b^{Sue}, d^{Mike}, e^{Sara}, h^{Pete} \rangle$
2	$\langle a^{Mike}, d^{Mike}, c^{Pete}, e^{Sara}, g^{Ellen} \rangle$
3	$\langle a^{Pete}, c^{Mike}, d^{Ellen}, e^{Sara}, f^{Sara}, b^{Sean}, d^{Pete}, e^{Sara}, g^{Ellen} \rangle$
4	$\langle a^{Pete}, d^{Mike}, b^{Sean}, e^{Sara}, h^{Ellen} \rangle$
5	$\langle a^{Ellen}, c^{Mike}, d^{Pete}, e^{Sara}, f^{Sara}, d^{Ellen}, c^{Mike}, e^{Sara}, f^{Sara}, b^{Sue}, d^{Pete}, e^{Sara}, h^{Mike} \rangle$
6	$\langle a^{Mike}, c^{Ellen}, d^{Mike}, e^{Sara}, g^{Mike} \rangle$
...	...

(a = *register request*, b = *examine thoroughly*, c = *examine casually*, d = *check ticket*, e = *decide*, f = *reinitiate request*, g = *pay compensation*, and h = *reject request*)

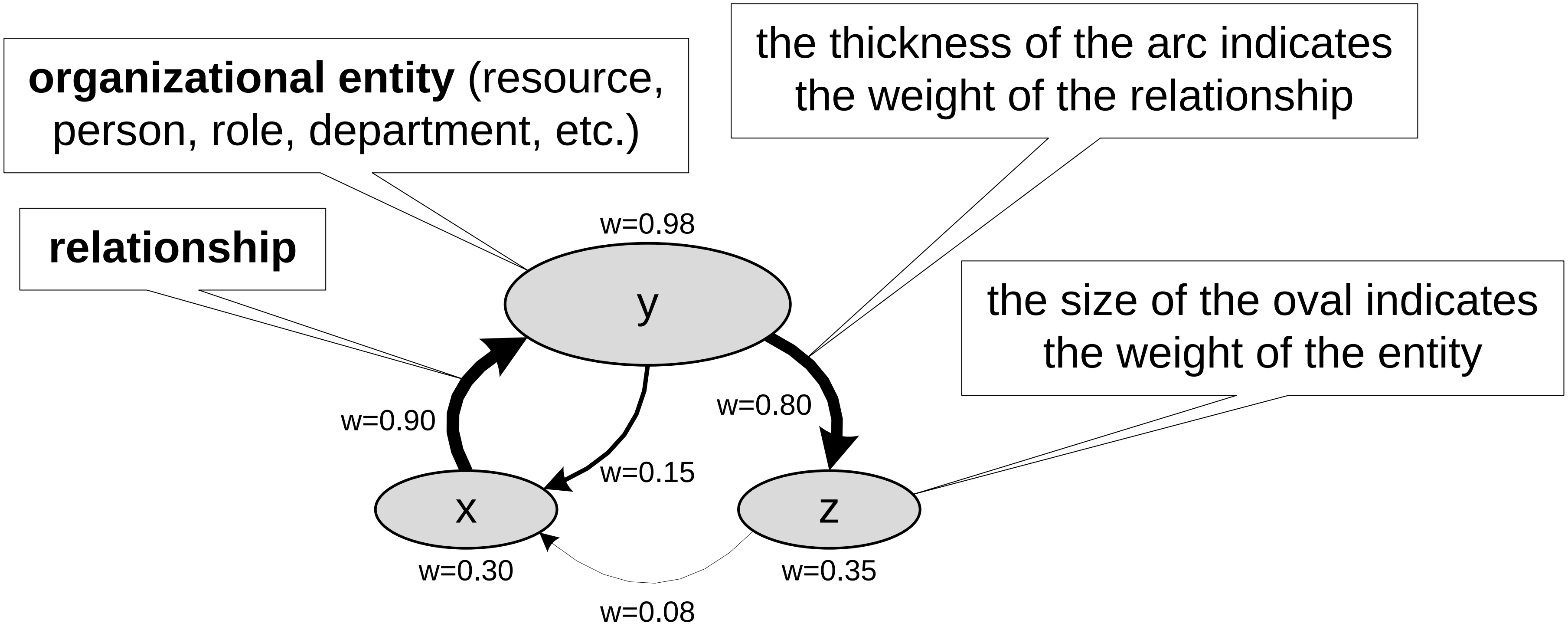
Resource-activity matrix

mean number of times a resource performs an activity per case

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Pete	0.3	0	0.345	0.69	0	0	0.135	0.165
Mike	0.5	0	0.575	1.15	0	0	0.225	0.275
Ellen	0.2	0	0.23	0.46	0	0	0.09	0.11
Sue	0	0.46	0	0	0	0	0	0
Sean	0	0.69	0	0	0	0	0	0
Sara	0	0	0	0	2.3	1.3	0	0

Activity a is executed exactly once for each case (take the sum of the first column). Pete, Mike, and Ellen are the only ones executing this activity. In 30% of the cases, a is executed by Pete, 50% is executed by Pete, and 20% is executed by Ellen. Activities e and f are always executed by Sara. Activity e is executed, on average, 2.3 times per case. Etc.

Social network analysis



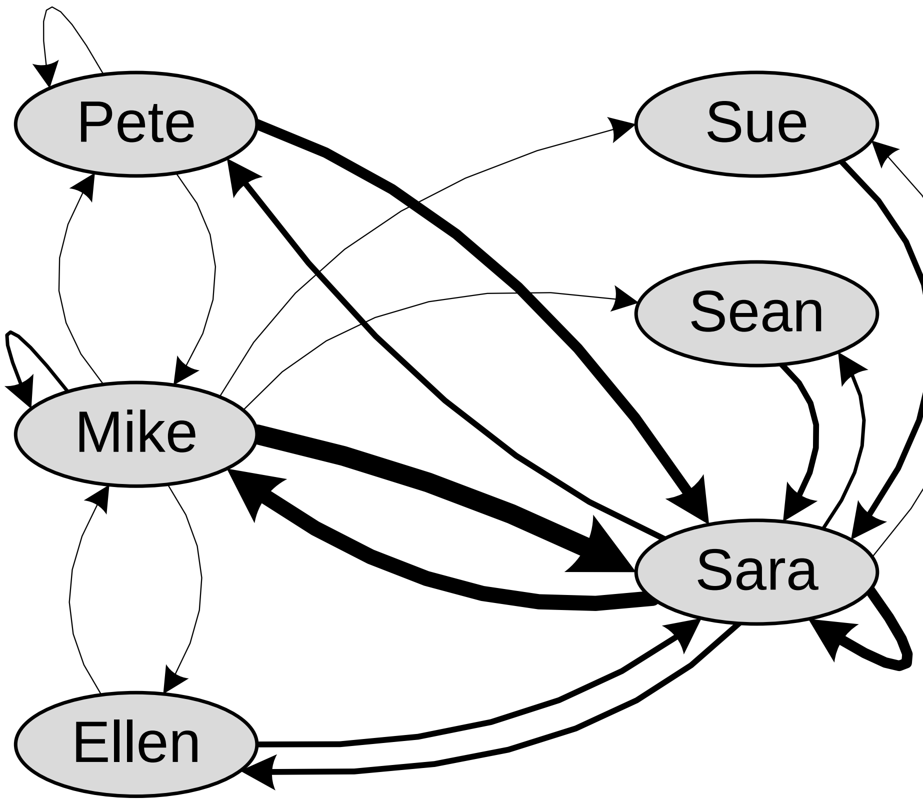
Handover of work matrix

	Pete	Mike	Ellen	Sue	Sean	Sara
Pete	0.135	0.225	0.09	0.06	0.09	1.035
Mike	0.225	0.375	0.15	0.1	0.15	1.725
Ellen	0.09	0.15	0.06	0.04	0.06	0.69
Sue	0	0	0	0	0	0.46
Sean	0	0	0	0	0	0.69
Sara	0.885	1.475	0.59	0.26	0.39	1.3

Count the number of times work is handed over from one resource to another (on average per case).

The causal dependencies in the process model are used to count handovers in the event log.

Social network based on handover of work (threshold of 0.1)

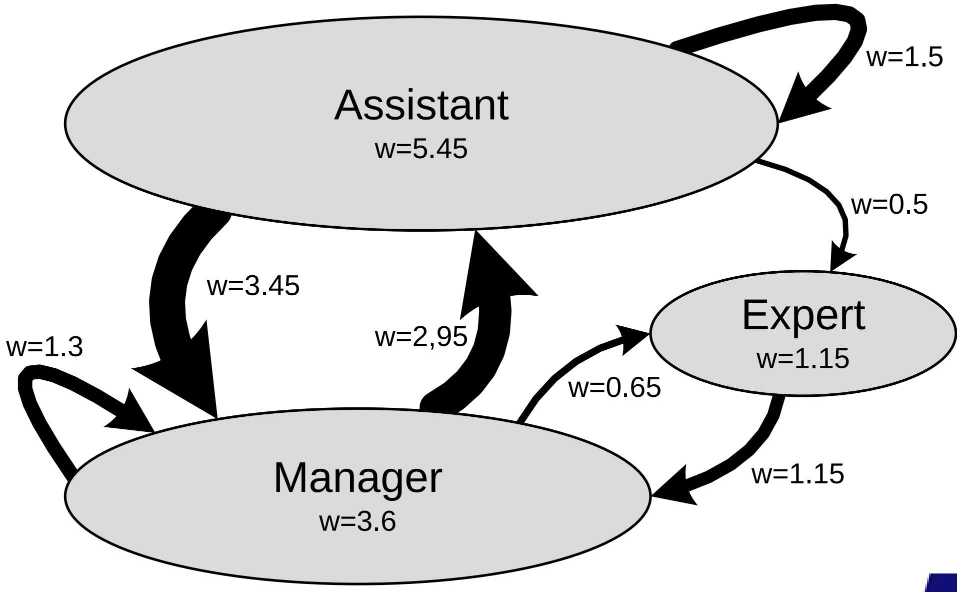


	Pete	Mike	Ellen	Sue	Sean	Sara
Pete	0.135	0.225	0.09	0.06	0.09	1.035
Mike	0.225	0.375	0.15	0.1	0.15	1.725
Ellen	0.09	0.15	0.06	0.04	0.06	0.69
Sue	0	0	0	0	0	0.46
Sean	0	0	0	0	0	0.69
Sara	0.885	1.475	0.59	0.26	0.39	1.3

In this figure only the thickness of the arcs is based on frequencies.

Handover of work at role level

	Assistant	Expert	Manager
Assistant	1.5	0.5	3.45
Expert	0	0	1.15
Manager	2.95	0.65	1.3

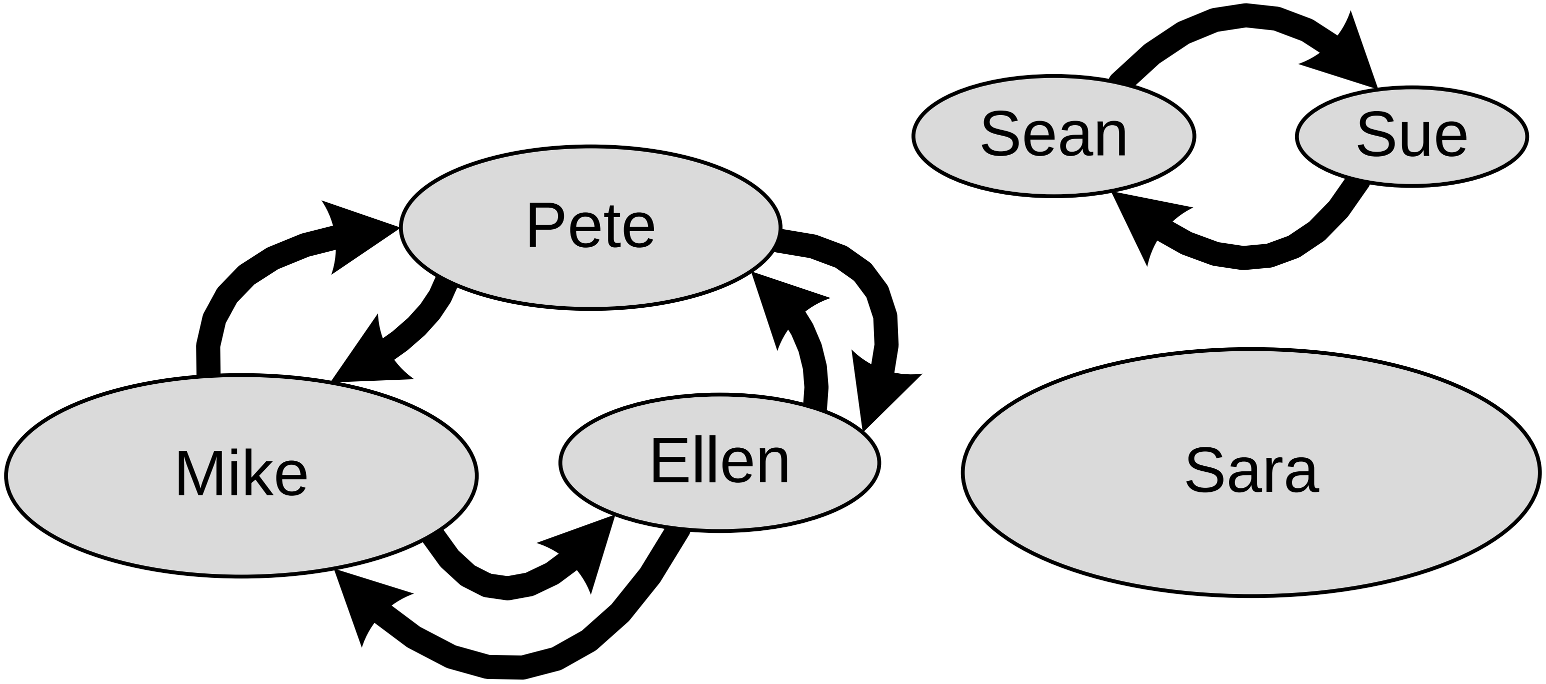


In this figure also the size of each node is based on frequencies.

Profile

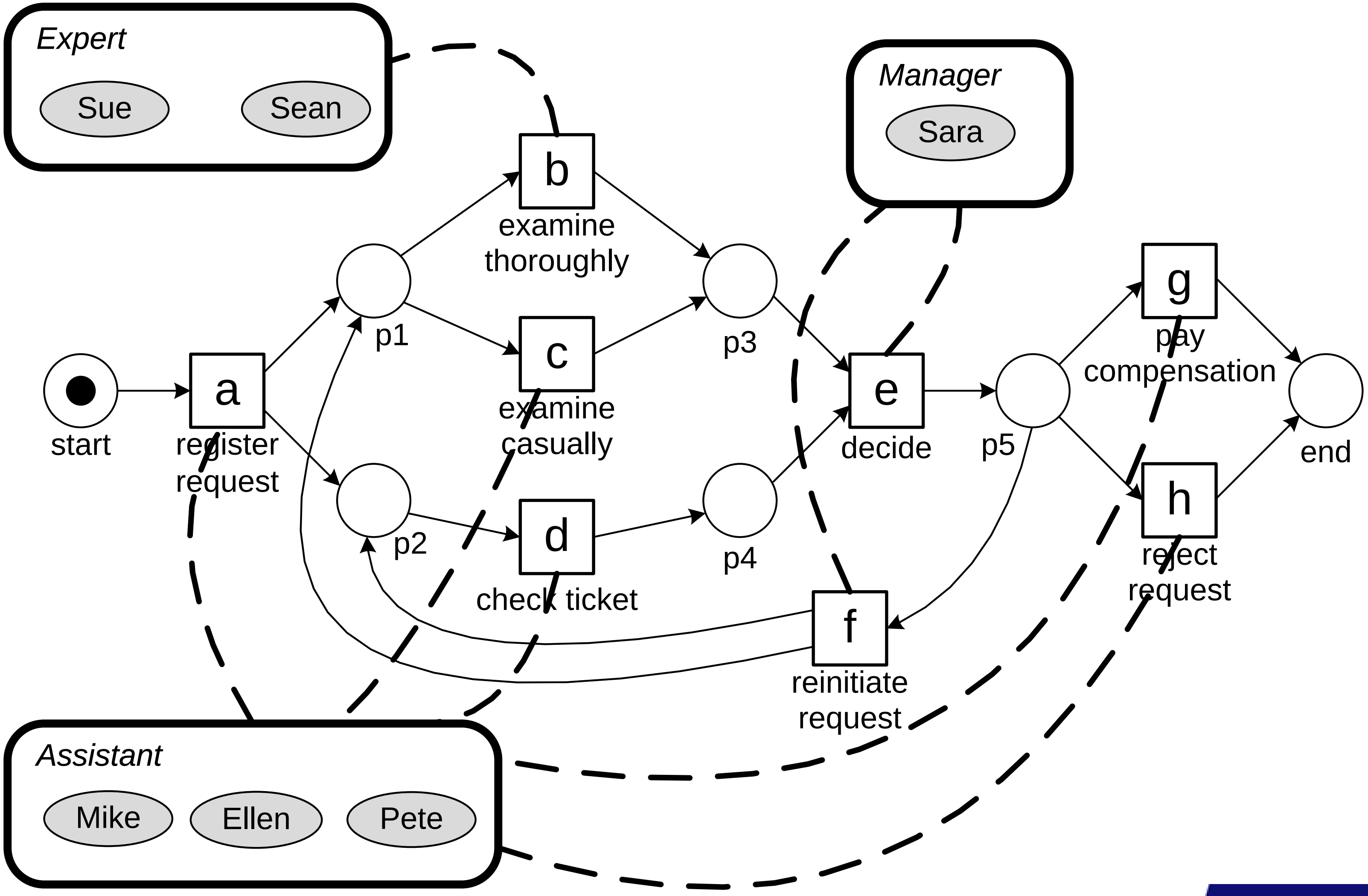
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Pete	0.3	0	0.345	0.69	0	0	0.135	0.165
Mike	0.5	0	0.575	1.15	0	0	0.225	0.275
Ellen	0.2	0	0.23	0.46	0	0	0.09	0.11
Sue	0	0.46	0	0	0	0	0	0
Sean	0	0.69	0	0	0	0	0	0
Sara	0	0	0	0	2.3	1.3	0	0

Social network based on similarity of profiles

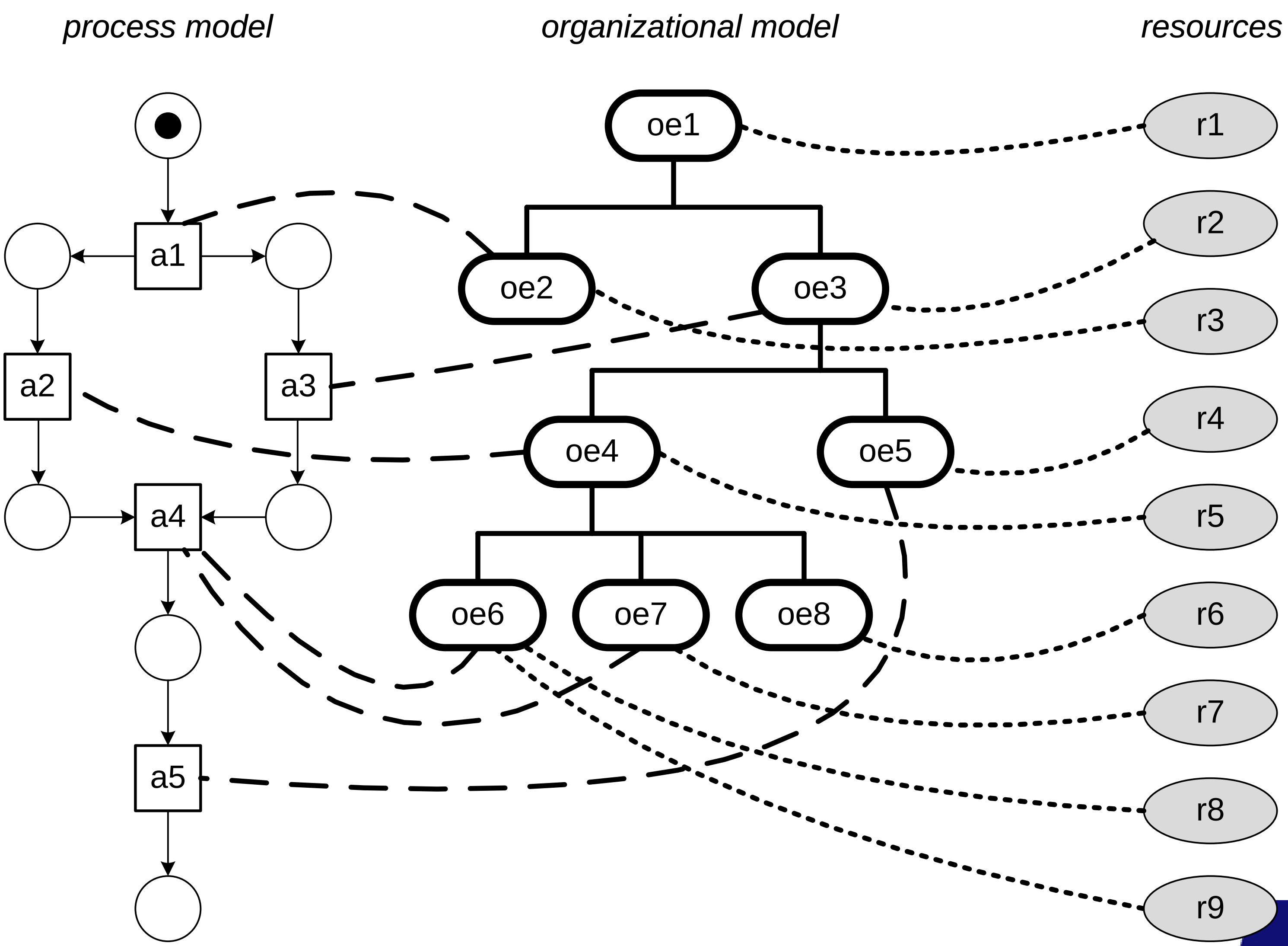


Resources that execute similar collections of activities are related. Sara is the only resource executing e and f . Therefore, she is not connected to other resources. Self-loops are suppressed as they contain no information (self-similarity)

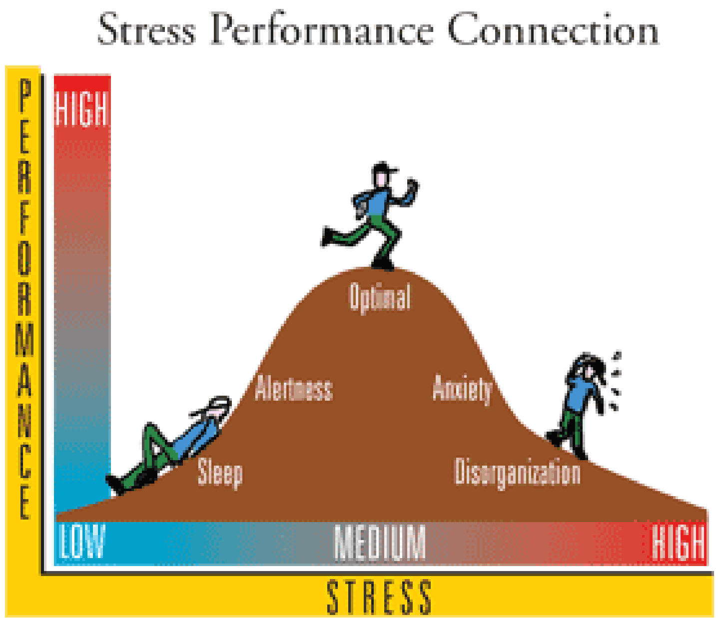
Discovering organizational structures



Another example



Analyzing resource behavior, e.g., Yerkes-Dodson law of arousal

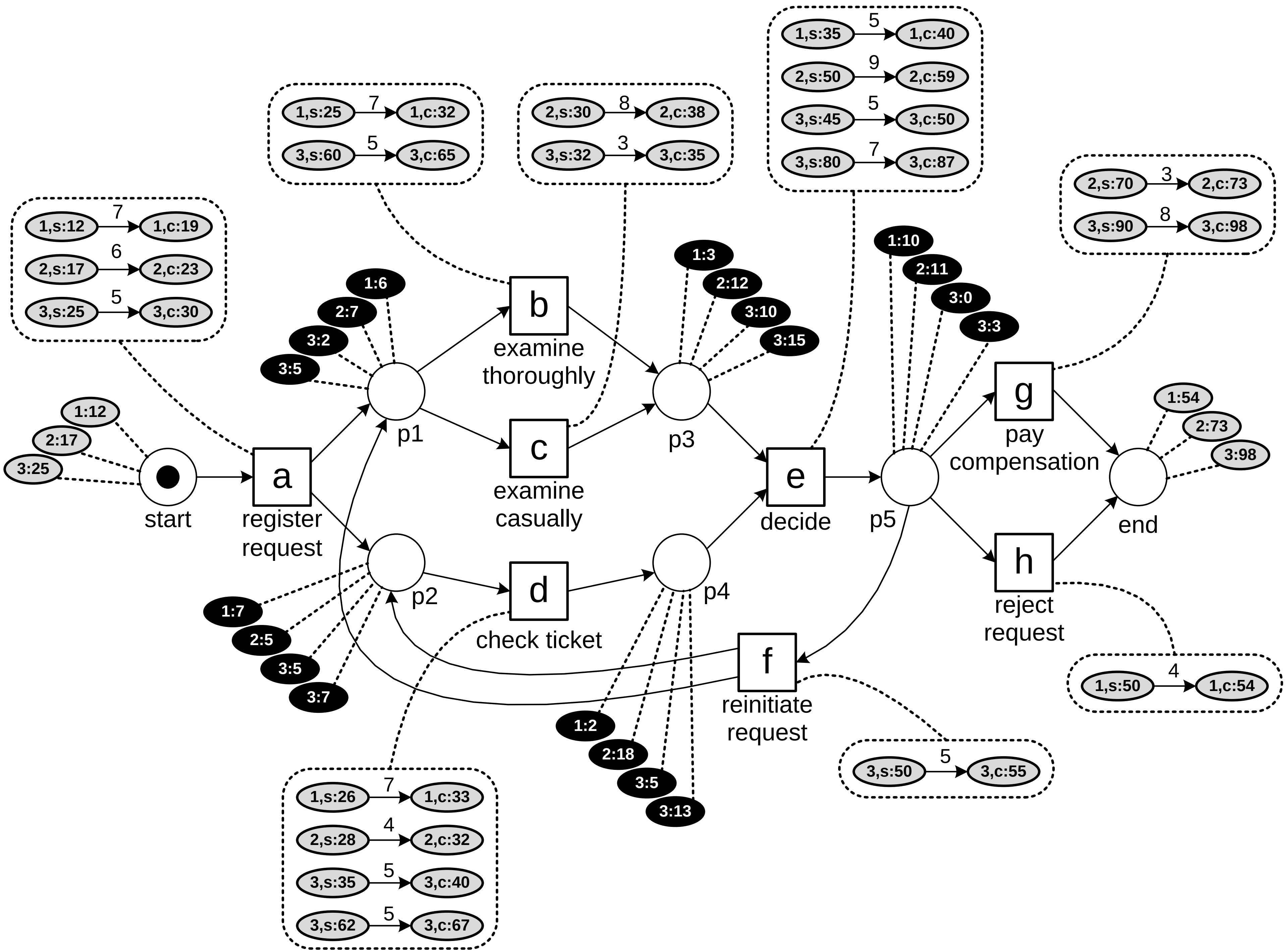


case id	event id	properties				
		time	activity	trans	resource	cost
	35654423	30-12-2010:11.02	register request	start	Pete	
	35654424	30-12-2010:11.08	register request	complete	Pete	50
	35654425	31-12-2010:10.06	examine thoroughly	start	Sue	
	35654427	31-12-2010:10.08	check ticket	start	Mike	
1	35654428	31-12-2010:10.12	examine thoroughly	complete	Sue	400
	35654429	31-12-2010:10.20	check ticket	complete	Mike	100
	35654430	06-01-2011:11.18	decide	start	Sara	
	35654431	06-01-2011:11.22	decide	complete	Sara	200
	35654432	07-01-2011:14.24	reject request	start	Pete	
	35654433	07-01-2011:14.32	reject request	complete	Pete	200
	35654483	30-12-2010:11.32	register request	start	Mike	

Learning time and probabilities

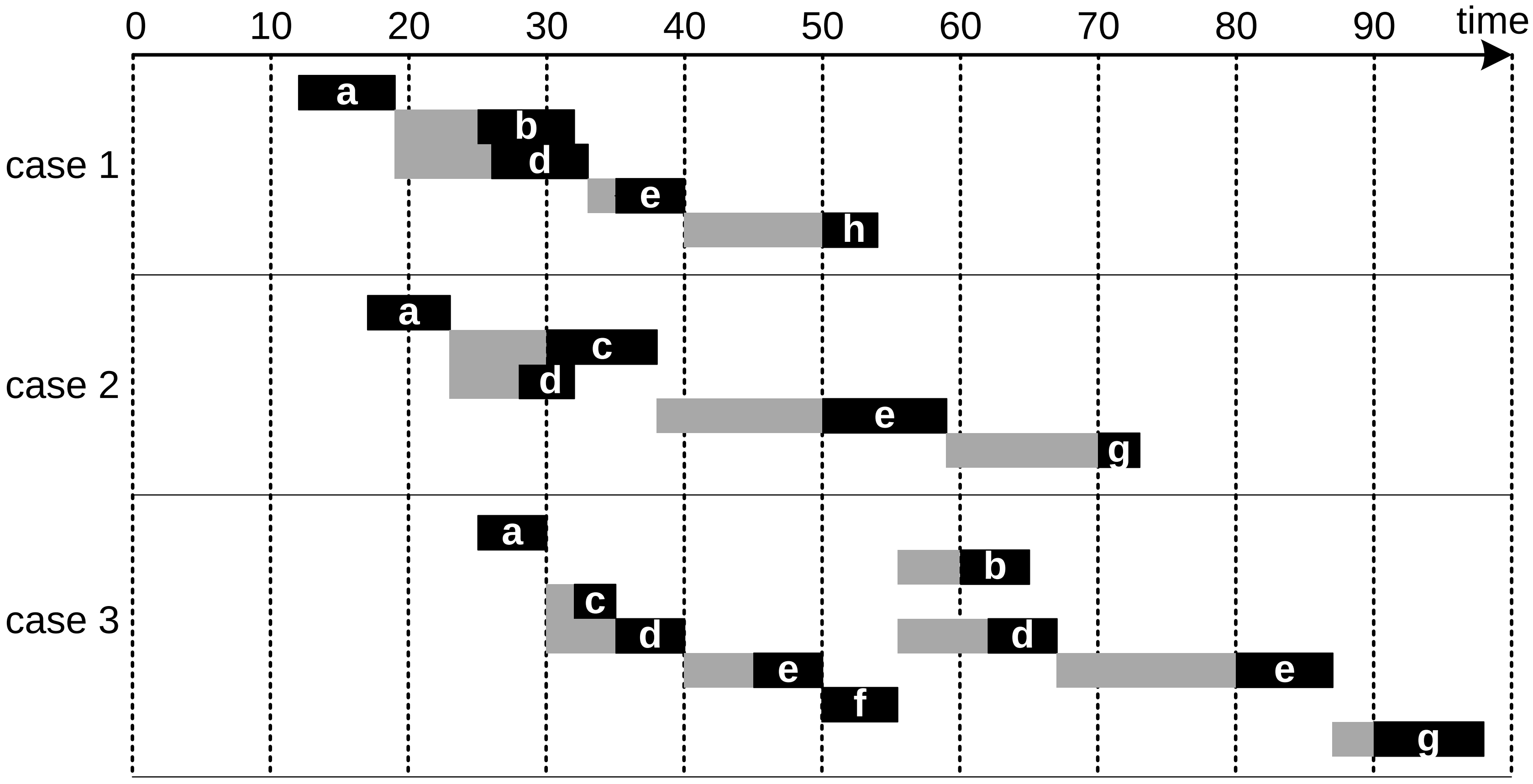
case id	trace
1	$\langle a^{12}_{start}, a^{19}_{complete}, b^{25}_{start}, d^{26}_{start}, b^{32}_{complete}, d^{33}_{complete}, e^{35}_{start}, e^{40}_{complete}, h^{50}_{start}, h^{54}_{complete}\rangle$
2	$\langle a^{17}_{start}, a^{23}_{complete}, d^{28}_{start}, c^{30}_{start}, d^{32}_{complete}, c^{38}_{complete}, e^{50}_{start}, e^{59}_{complete}, g^{70}_{start}, g^{73}_{complete}\rangle$
3	$\langle a^{25}_{start}, a^{30}_{complete}, c^{32}_{start}, c^{35}_{complete}, d^{35}_{start}, d^{40}_{complete}, e^{45}_{start}, e^{50}_{complete}, f^{50}_{start}, f^{55}_{complete}, b^{60}_{start}, d^{62}_{start}, b^{65}_{complete}, d^{67}_{complete}, e^{80}_{start}, e^{87}_{complete}, g^{90}_{start}, g^{98}_{complete}\rangle$
...	...

- **Replay, as before, but now considering timestamps.**
- **Let us replay the first three cases in the event log:**
  - **case 1 starts at time 12 and ends at time 54,**
  - **case 2 starts at time 17 and ends at time 73,**
  - **case 3 starts at time 25 and ends at time 98.**

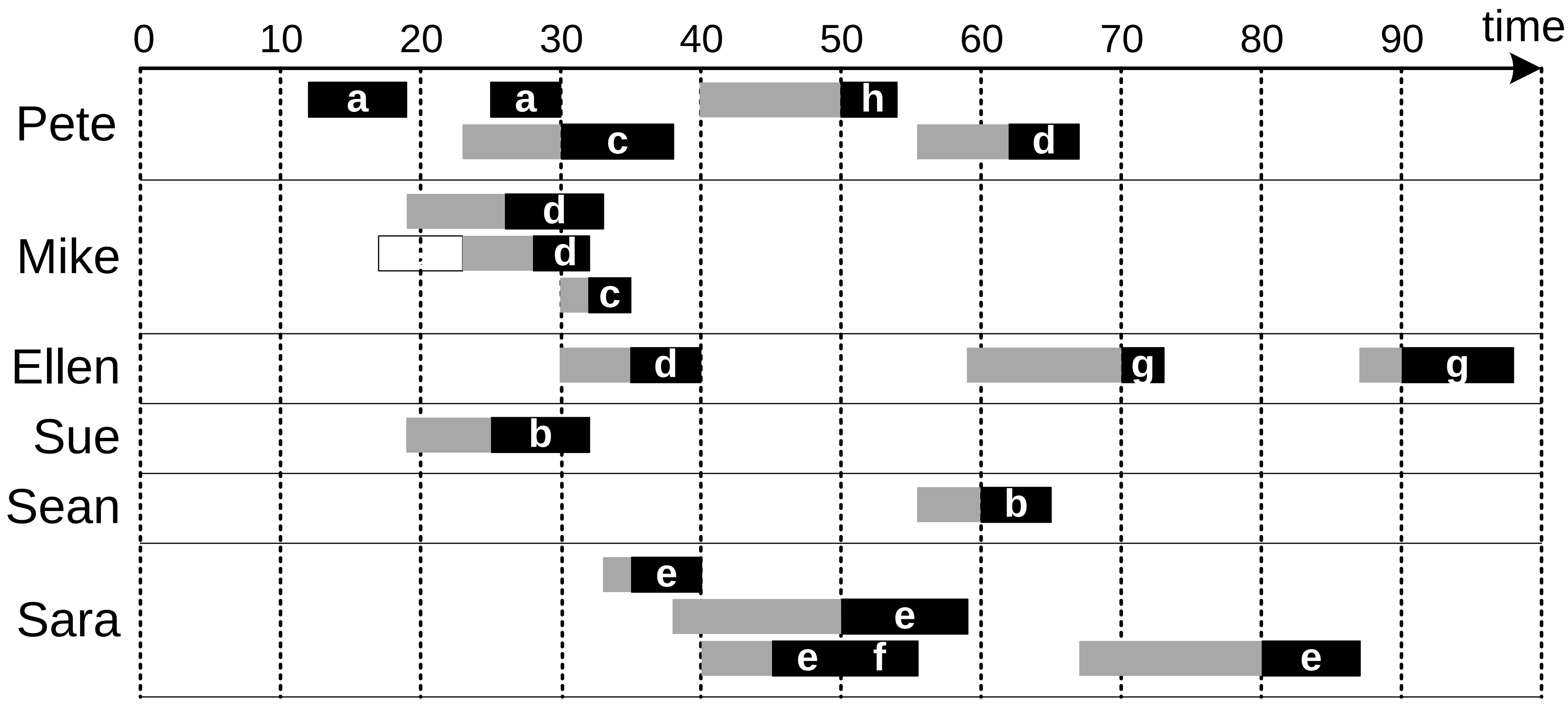


start
1:12
2:17
3:25
a
register request
1,s:12 7 1,c:19
2,s:17 6 2,c:23
3,s:25 5 3,c:30
p1
1:6
2:7
3:2
3:5
p2
1:7
2:5
3:5
3:7
b
examine thoroughly
1,s:25 7 1,c:32
3,s:60 5 3,c:65
c
examine casually
2,s:30 8 2,c:38
3,s:32 3 3,c:35
d
check ticket
1,s:26 7 1,c:33
2,s:28 4 2,c:32
3,s:35 5 3,c:40
3,s:62 5 3,c:67
p3
1:3
2:12
3:10
3:15
p4
1:2
2:18
3:5
3:13
e
decide
1,s:35 5 1,c:40
2,s:50 9 2,c:59
3,s:45 5 3,c:50
3,s:80 7 3,c:87
p5
1:10
2:11
3:0
3:3
f
reinitiate request
3,s:50 5 3,c:55
g
pay compensation
2,s:70 3 2,c:73
3,s:90 8 3,c:98
h
reject request
1,s:50 4 1,c:54
end
1:54
2:73
3:98

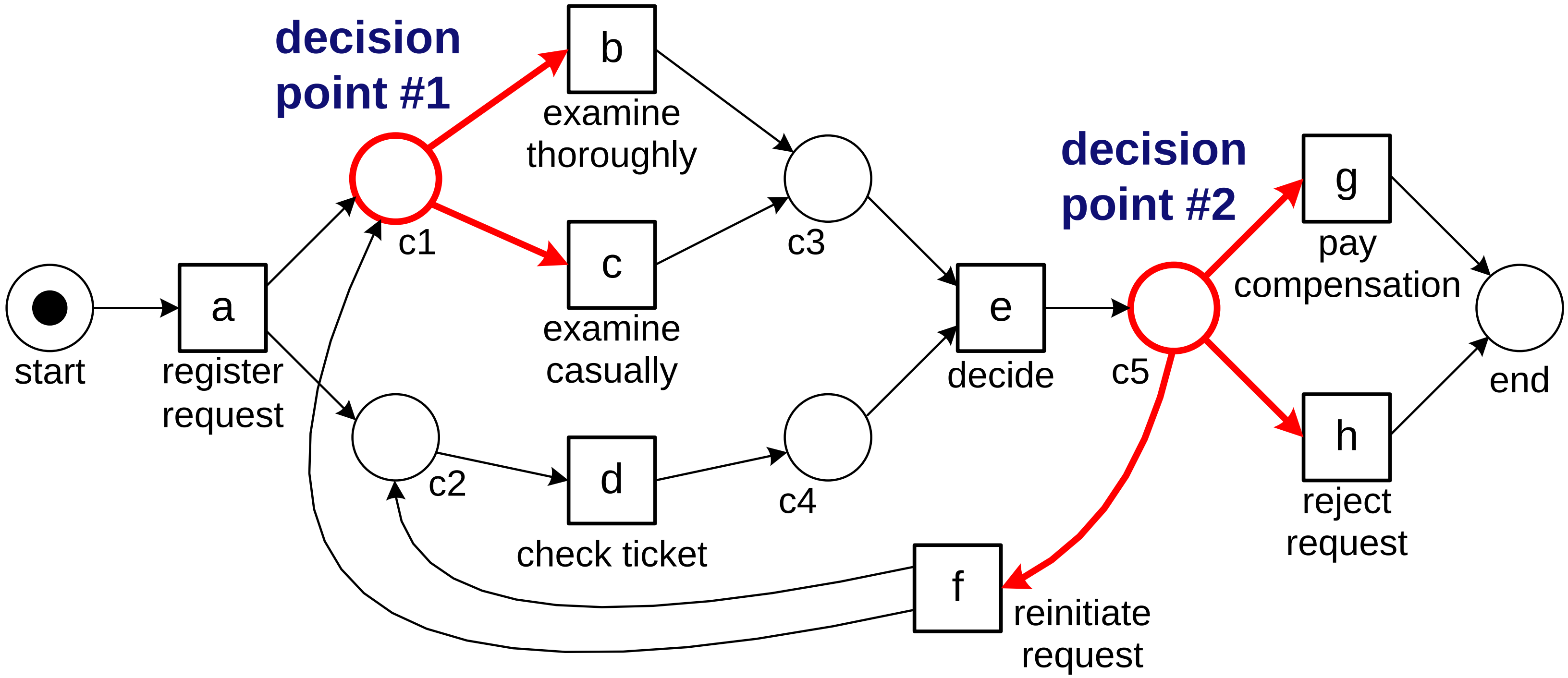
Another view on the timed replay of the first three cases



Timed replay projected onto resources

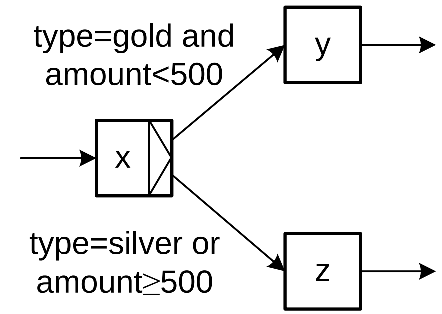


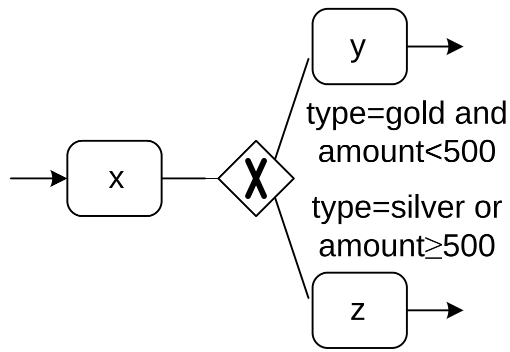
Decision mining

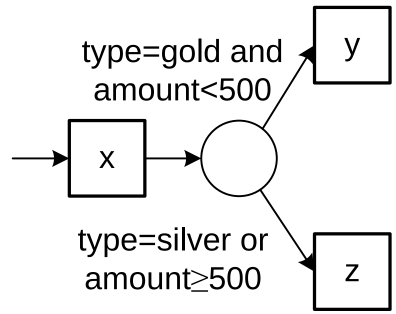


Example: XOR-split

type	region	amount	activity
gold	south	987.30	z
silver	north	178.70	z
gold	south	211.50	y
silver	west	587.70	z
silver	east	224.70	z
silver	south	278.50	z
gold	north	488.50	y
silver	west	443.20	z
silver	south	673.70	z
gold	west	413.50	y
silver	south	687.70	z
gold	south	987.30	z
silver	north	378.80	z
gold	south	314.50	y
silver	north	537.70	z
silver	west	158.70	z
gold	east	344.50	y
...	...	...	...





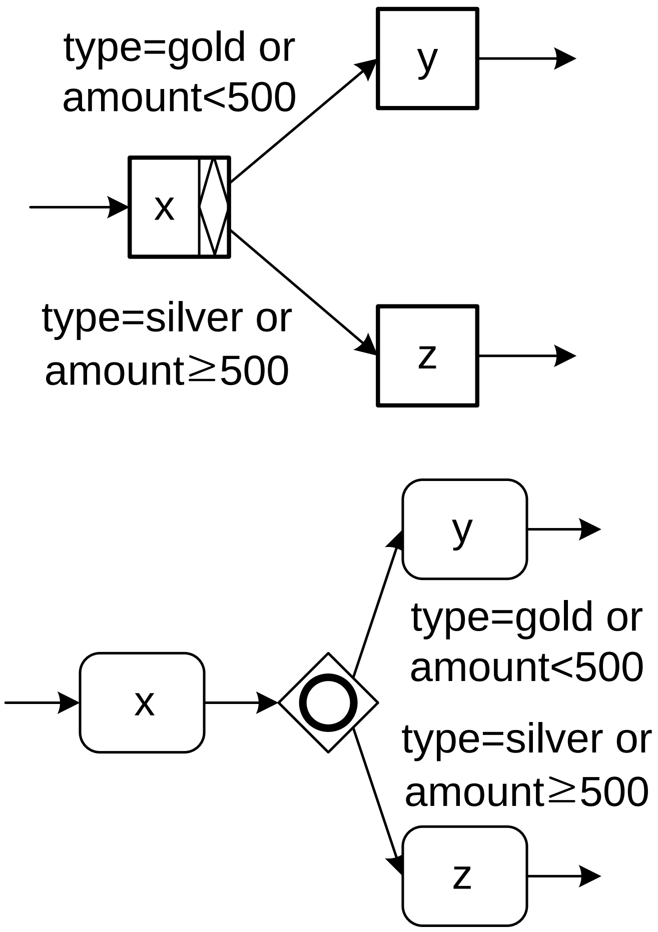


What are the "features" (predictor variables) influencing the decision?

A classification technique like decision tree learning can be used to find such rules: :explain response variable (dependent variable) in terms of predictor variables (independent variables).

Example: OR-split

type	region	amount	activity
gold	south	987.30	y and z
silver	north	178.70	y and z
gold	south	211.50	just y
silver	west	587.70	just z
silver	east	224.70	y and z
silver	south	278.50	y and z
gold	north	488.50	just y
silver	west	443.20	y and z
silver	south	673.70	just z
...	...	...	...



Classification in process mining

- **The application of classification techniques like decision tree learning is not limited to decision mining based on event/case data only.**
- **Additional predictor variables may be used:**
  - **behavioral information (count number of loops)**
  - **performance information (processing times)**
  - **contextual information (weather, queues, etc.)**
- **Alternative response variables can be analyzed:**
  - **uncover reasons for non-conformance (split instances in two groups)**
  - **uncover reasons for delays**

Bringing it all together