CS 4072 - Topics in CS
Process Mining

Lecture # 26

May 31, 2022

Spring 2022

FAST - NUCES, CFD Campus

Dr. Rabia Maqsood

rabia.maqsood@nu.edu.pk

Today's Topics

- Advanced topics in Process Mining
- Process Mining and Data Mining
- Mid 2 solution

Advanced topics in PM

Mining Additional Perspectives

So far, we have focused on control-flow perspective

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
...	...	...	...	...	...	...

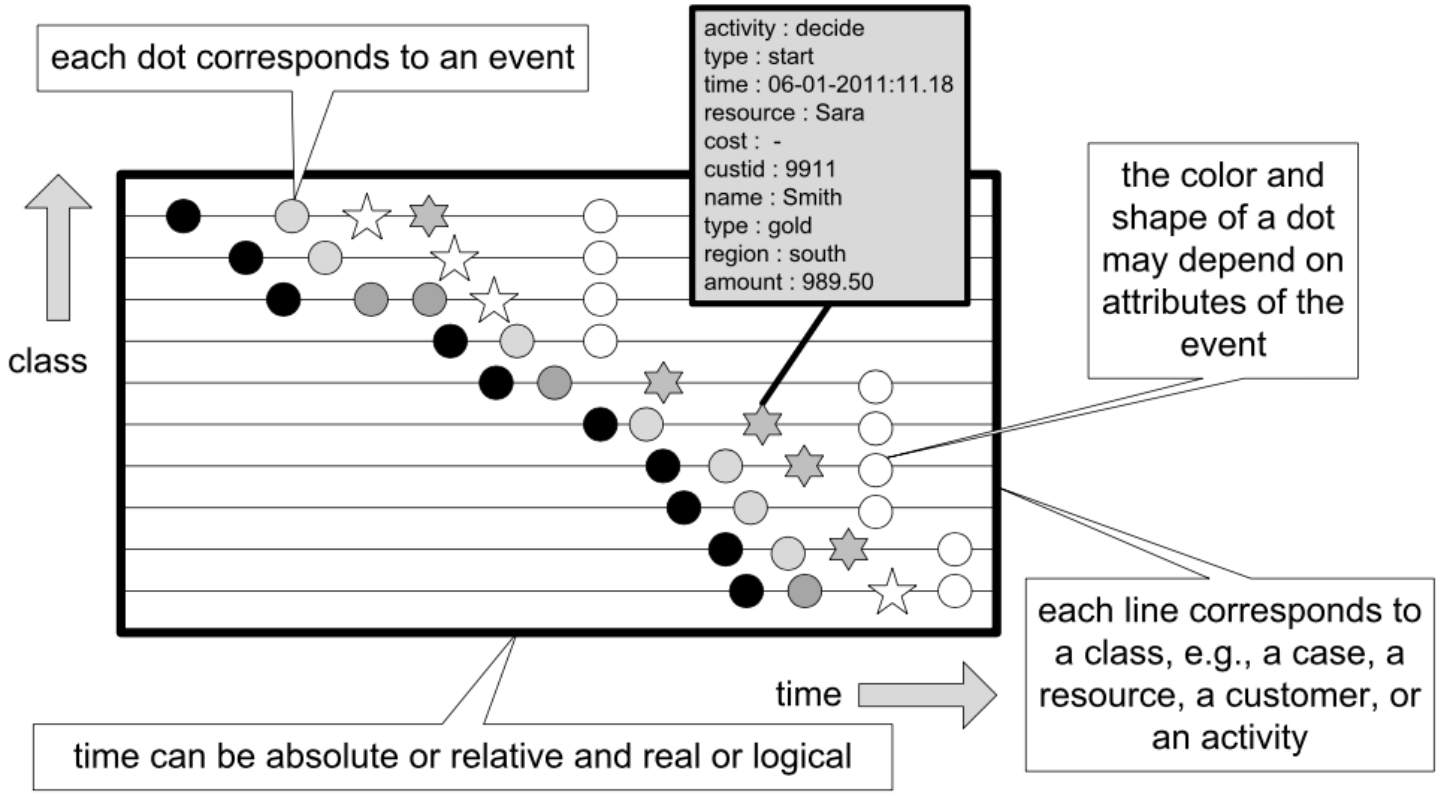
Useful information can be extracted from other data attributes in the event log

Attributes of cases

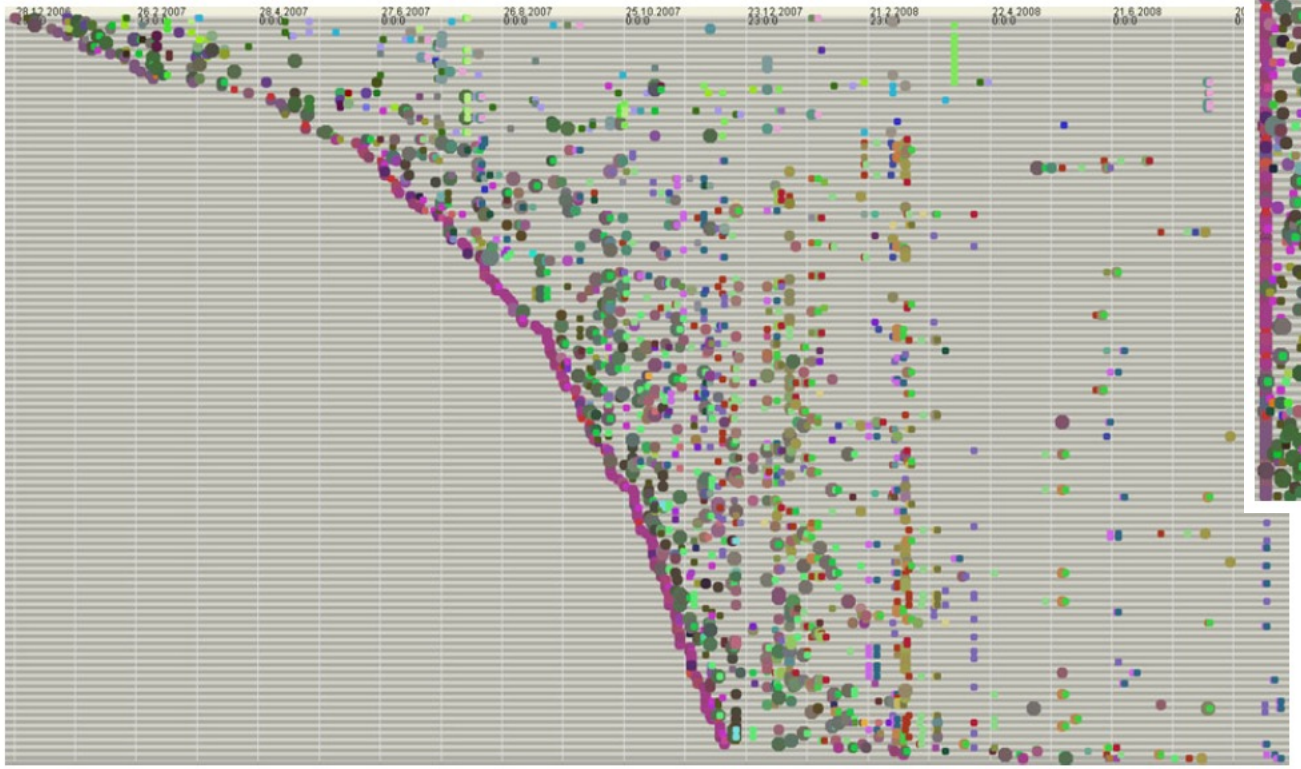
Case id	Custid	Name	Type	Region	Amount
1	9911	Smith	gold	south	989.50
2	9915	Jones	silver	west	546.00
3	9912	Anderson	silver	north	763.20
4	9904	Thompson	silver	west	911.70
5	9911	Smith	gold	south	812.10
6	9944	Baker	silver	east	788.00
7	9944	Baker	silver	east	792.80
8	9911	Smith	gold	south	544.70
...	...	...	...	...	...

Case attributes refers to a complete case instead of an event

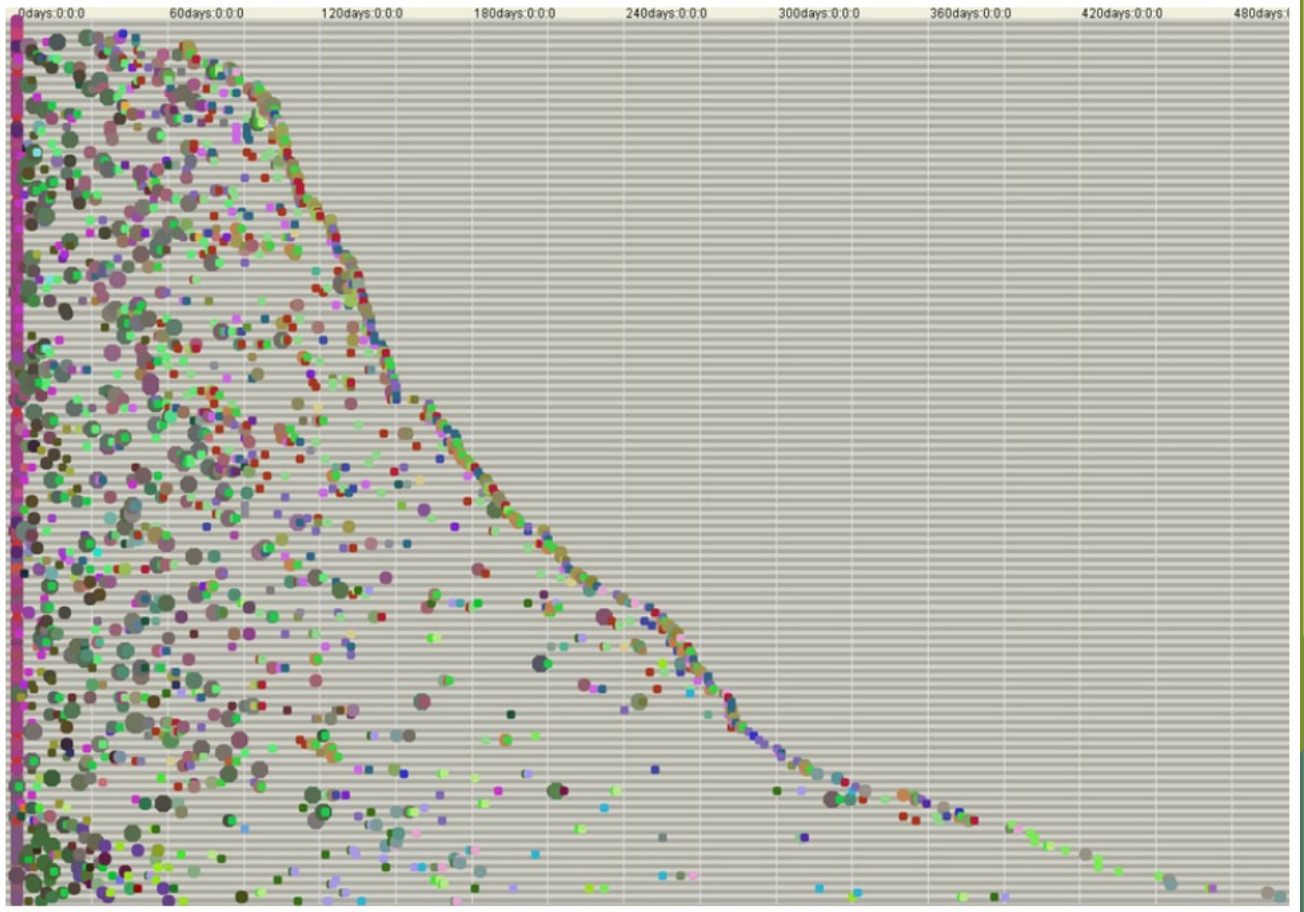
Dotted Chart
Helicopter View



Dotted Chart
Helicopter View



Absolute time



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	$\langle a^{Mike}, c^{Ellen}, d^{Mike}, e^{Sara}, g^{Mike} \rangle$
...	...

Input event log, highlighting “resource” with each event

Table 9.4 Resource-activity matrix showing the mean number of times a person performed 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

Social Network Analysis

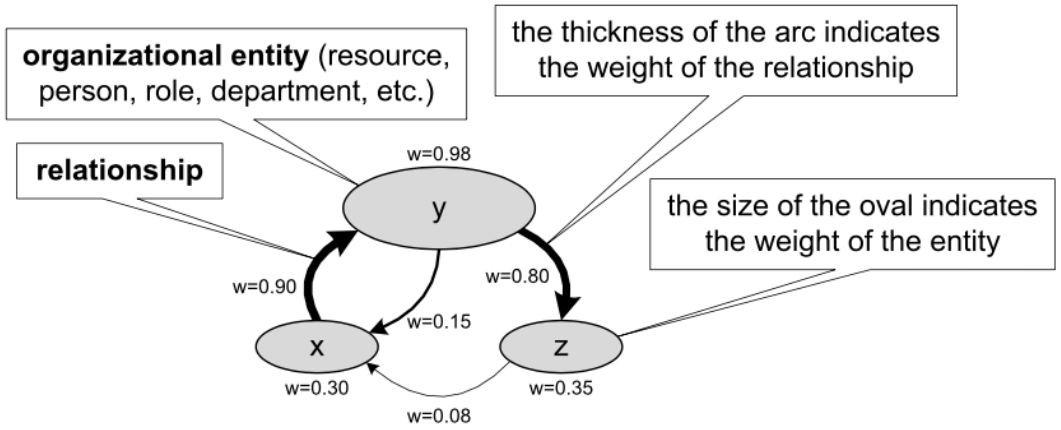


Fig. 9.5 A social network consists of nodes representing organizational entities and arcs representing relationships. Both nodes and arcs can have weights indicated by "$w = \cdots$" and the size of the shape

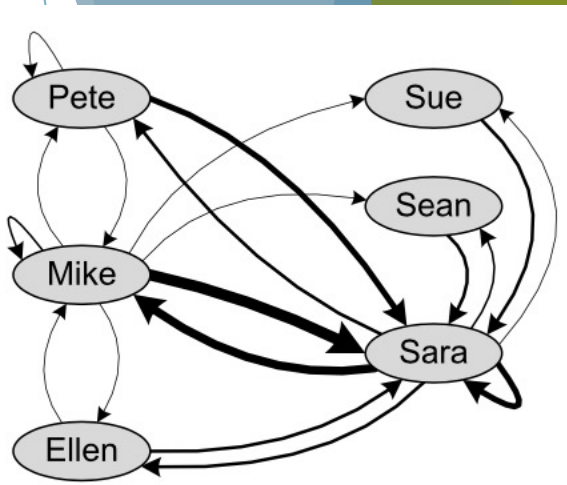


Fig. 9.6 Social network based on handover of work at the level of individual resources using a threshold of 0.1. The thickness of the arcs is based on the frequency of handovers from one person to another

Table 9.5 Handover of work matrix showing the mean number of handovers from one person to another per case

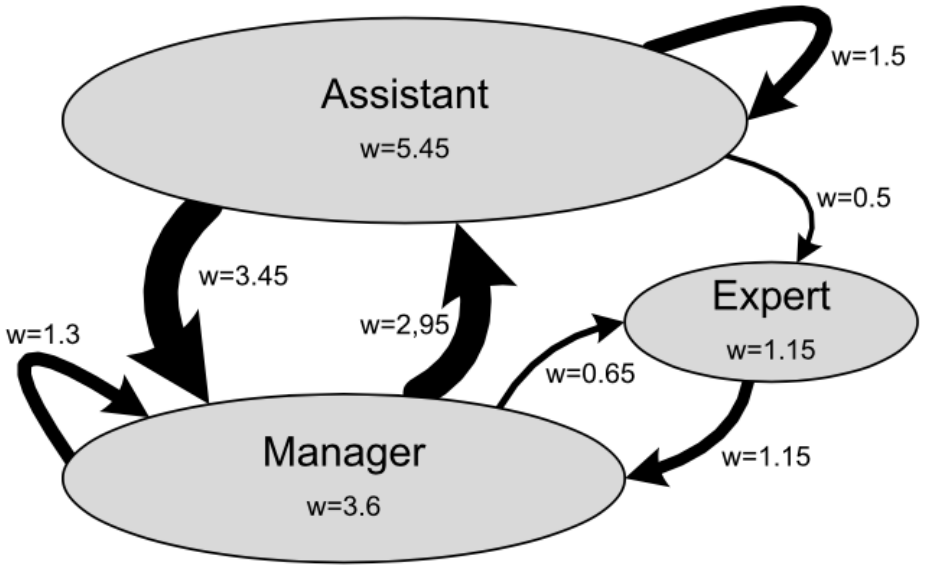
	Pete	Mike	Ellen	Sue	Sean	Sara
Pete	0.135	0.225	0.09	0.06	0.09	1.035
Mike	0.225	0.375	0.15	0.1	0.15	1.725
Ellen	0.09	0.15	0.06	0.04	0.06	0.69
Sue	0	0	0	0	0	0.46
Sean	0	0	0	0	0	0.69
Sara	0.885	1.475	0.59	0.26	0.39	1.3

Social Network Analysis

Table 9.6 Handover of work matrix at the role level

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

Fig. 9.7 Social network based on handover of work at the level of roles. The weights of nodes are based on the number of times a resource having the role performs an activity. The weights of the arcs are based of the average number of times a handover takes place from one role to another per case



Discovering Organizational Structures

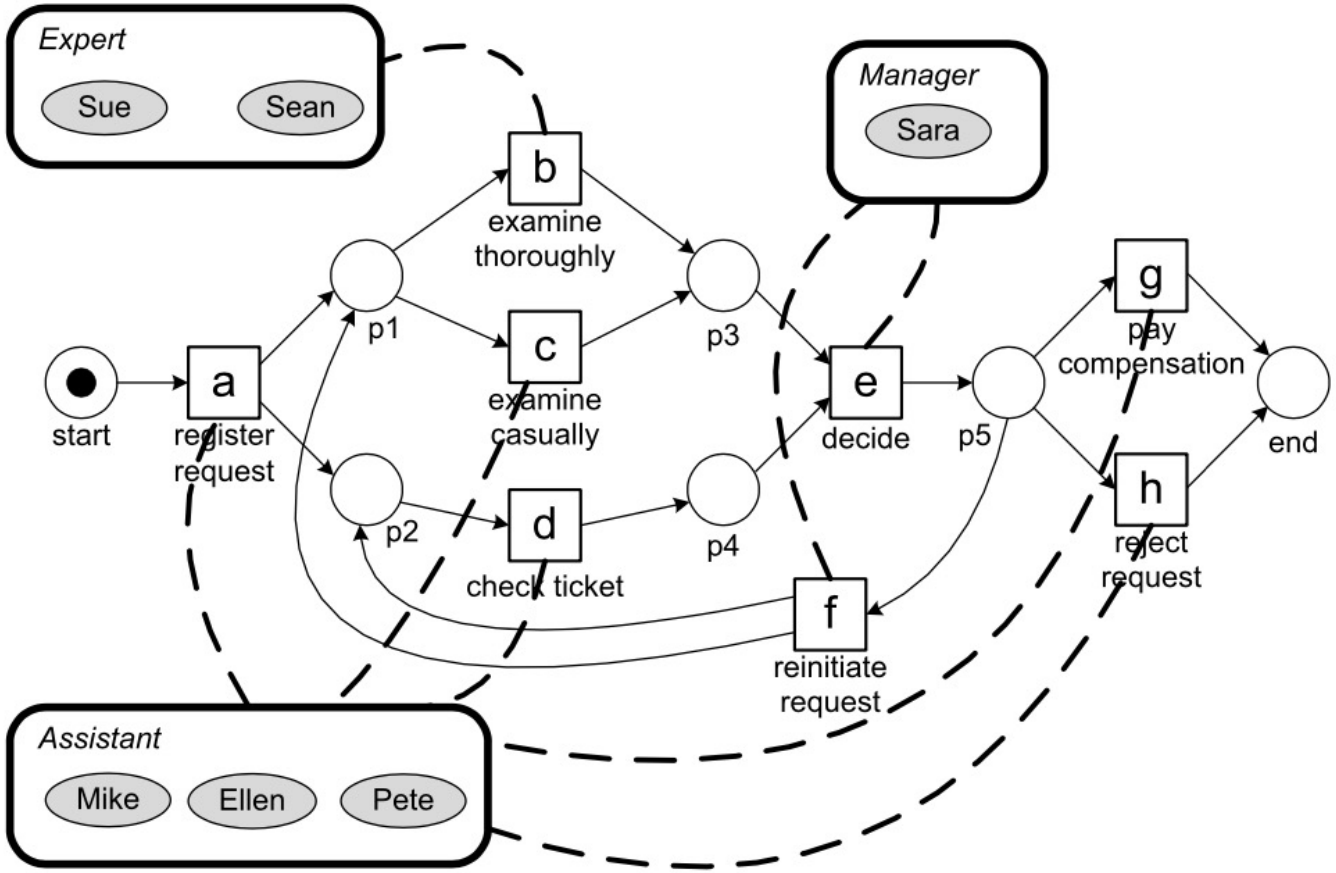


Fig. 9.9 Organizational model discovered based on the event log

Discovering Organizational Structures

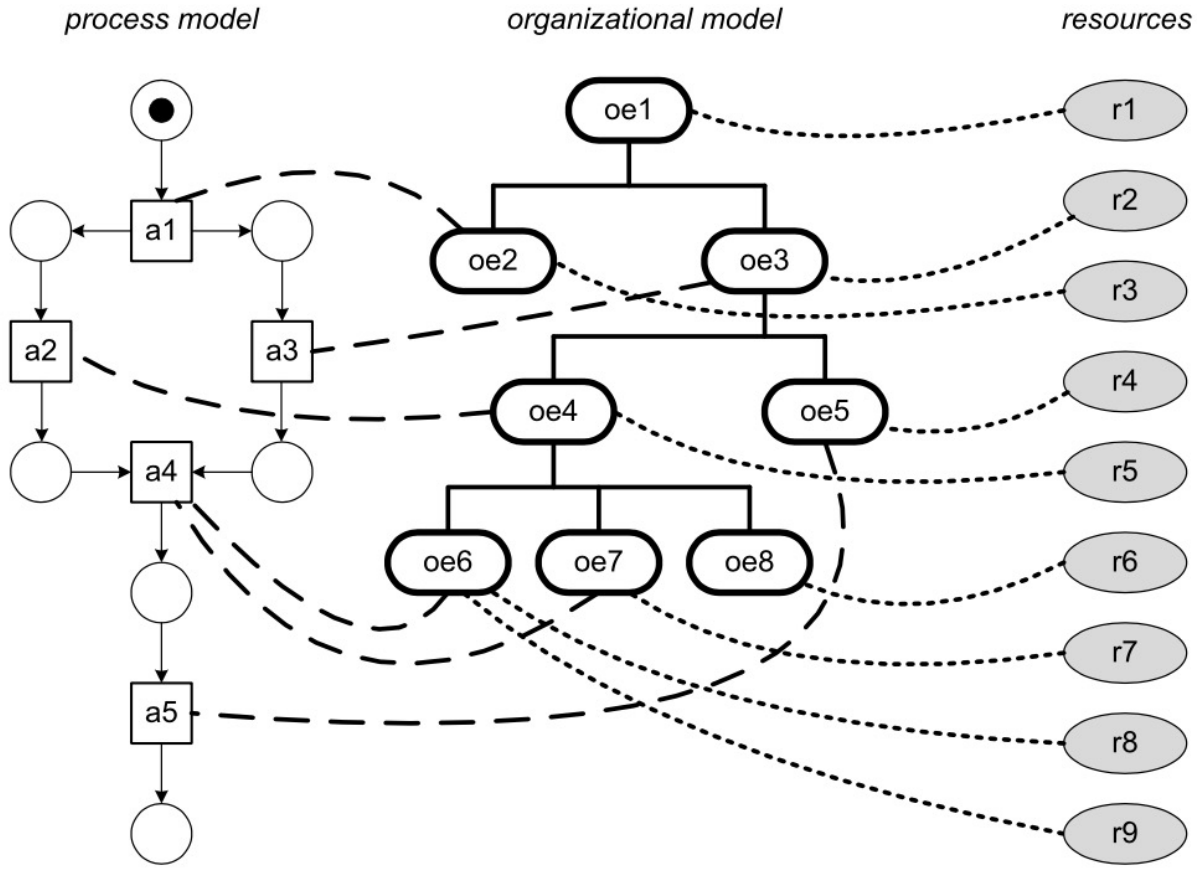


Fig. 9.10 The organizational entities discovered connect activities in the process model to sets resources

Analyzing Resource Behavior

Table 9.7 Compact representation of the event log highlighting timestamps; artificial timestamps are used to simplify the presentation of the time-based replay approach

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	$\langle a^{25}_{start}, a^{30}_{complete}, c^{32}_{start}, c^{35}_{complete}, d^{35}_{start}, d^{40}_{complete}, e^{45}_{start}, e^{50}_{complete}, f^{50}_{start}, f^{55}_{complete}, b^{60}_{start}, d^{62}_{start}, b^{65}_{complete}, d^{67}_{complete}, e^{80}_{start}, e^{87}_{complete}, g^{90}_{start}, g^{98}_{complete}\rangle$
...	...

And even more…

- Model repairing
- Model enhancement
- Decision mining

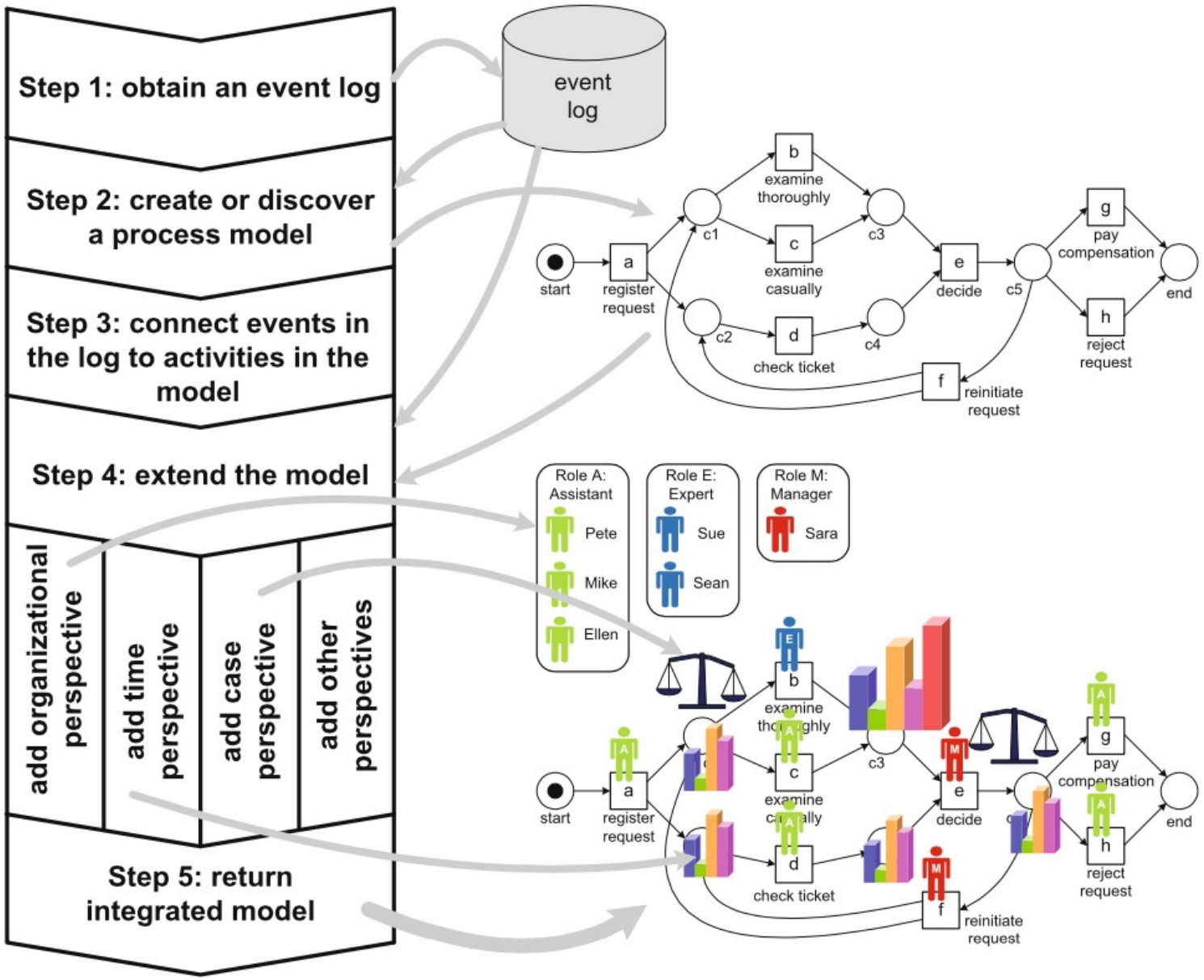


Fig. 9.16 Approach to come to a fully integrated model covering the organizational, time, and case perspectives

PM and DM: overview

Process Mining and Data Mining

- Process mining builds on two pillars:
  - Process modeling and analysis
  - Data mining
- Some process mining techniques build on classical data mining techniques,
  - E.g., discovery and enhancement approaches focusing on data and resources.
- Ideas originating from the data mining field can be used for the evaluation of process mining results.

Process Mining and Data Mining

- Existing data mining techniques are of little use for control-flow discovery, conformance checking, and other process mining tasks.

- Data mining can be defined as:

  "the analysis of (often large) data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner"

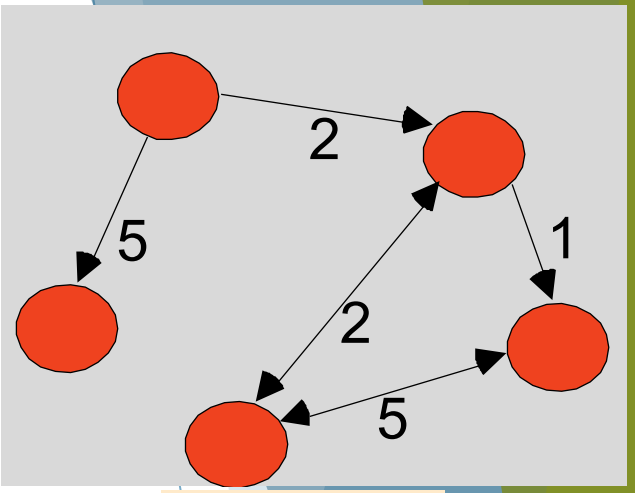
Data Formats in Data Mining

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Record Data

	team	coach	play	ball	score	game	win	lost	timeout	season
Document 1	3	0	5	0	2	6	0	2	0	2
Document 2	0	7	0	2	1	0	0	3	0	0
Document 3	0	1	0	0	1	2	2	0	3	0

Document Data



Graph Data

Projection of x Load	Projection of y load	Distance	Load	Thickness
10.23	5.27	15.22	2.7	1.2
12.65	6.25	16.22	2.2	1.1

Data Matrix

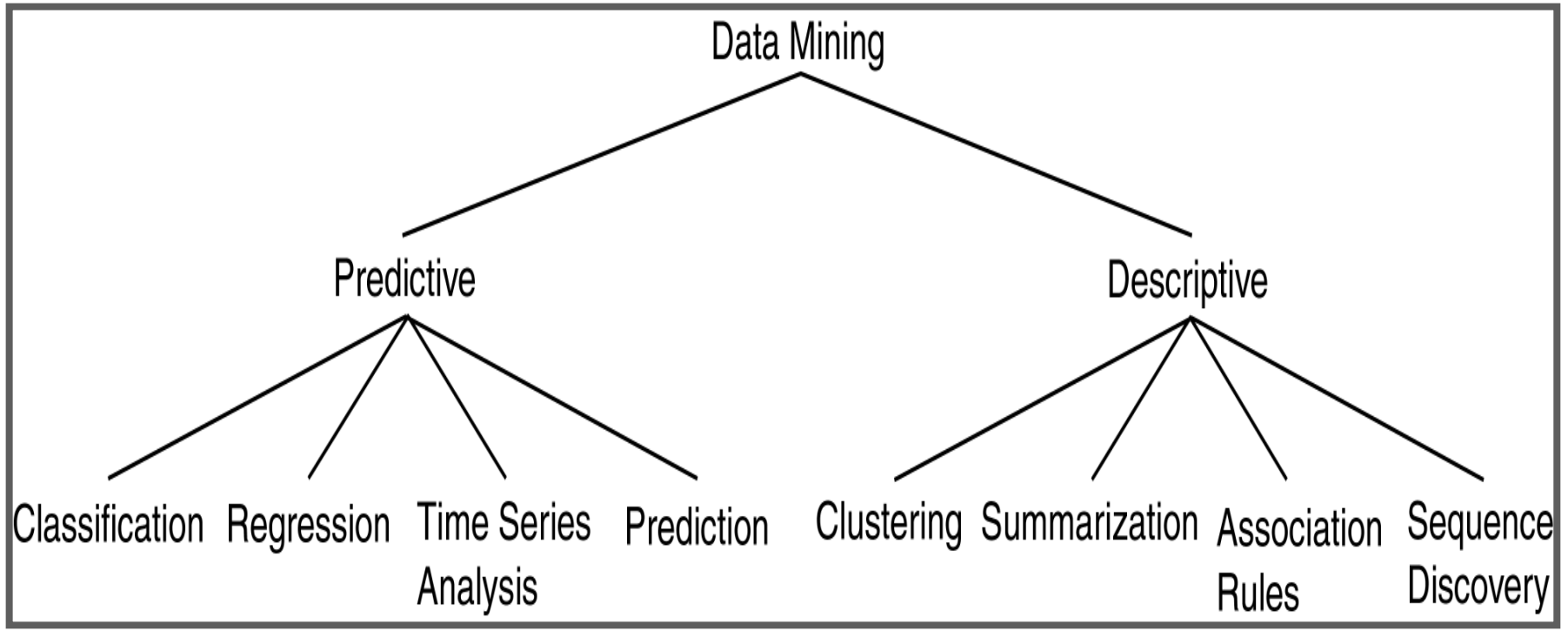
TID	Items
1	Bread, Coke, Milk
2	Beer, Bread
3	Beer, Coke, Diaper, Milk
4	Beer, Bread, Diaper, Milk
5	Coke, Diaper, Milk

Transaction Data

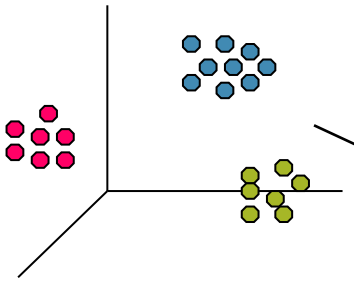
```
GGTTCCGCCTTCAGCCCCGCGCC
CGCAGGGCCCGCCCCGCGCCGTC
GAGAAGGGCCCGCCTGGCGGGCG
GGGGGAGGCGGGGCCGCCCGAGC
CCAACCGAGTCCGACCAGGTGCC
CCCTCTGCTCGGCCTAGACCTGA
GCTCATTAGGCGGCAGCGGACAG
GCCAAGTAGAACACGCGAAGCGC
TGGGCTGCCTGCTGCGACCAGGG
```

Ordered Data

Data Mining Tasks



Data Mining Tasks

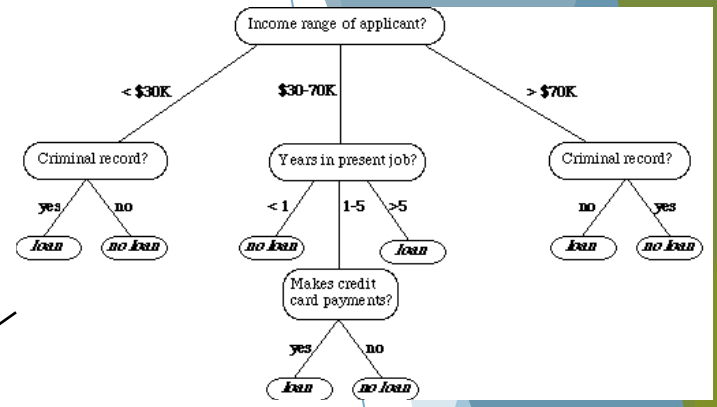


Clustering

Data

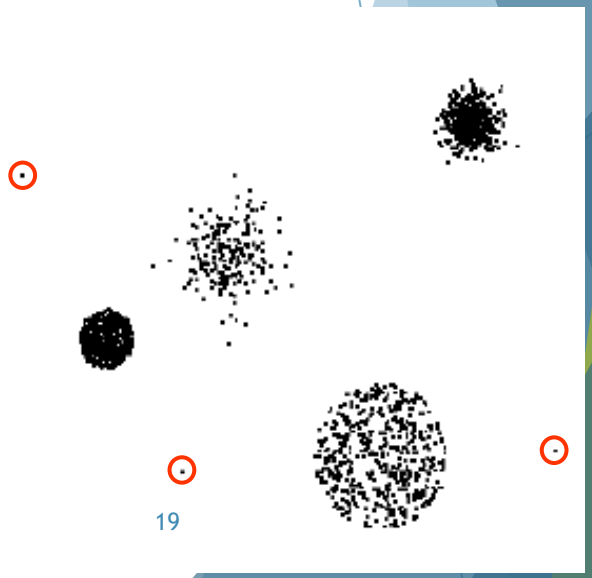
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes
11	No	Married	60K	No
12	Yes	Divorced	220K	No
13	No	Single	85K	Yes
14	No	Married	75K	No
15	No	Single	90K	Yes

Predictive Modeling



Association Rules

Anomaly Detection



Thank you