Mining Additional Perspectives

Andrea Polini

Process Mining
MSc in Computer Science (LM-18)
University of Camerino

Summary

Summary

Idea

What we want to do?

Event logs may contain a wealth of information relating to other perspectives such as the organizational perspective, the case perspective, and the time perspective

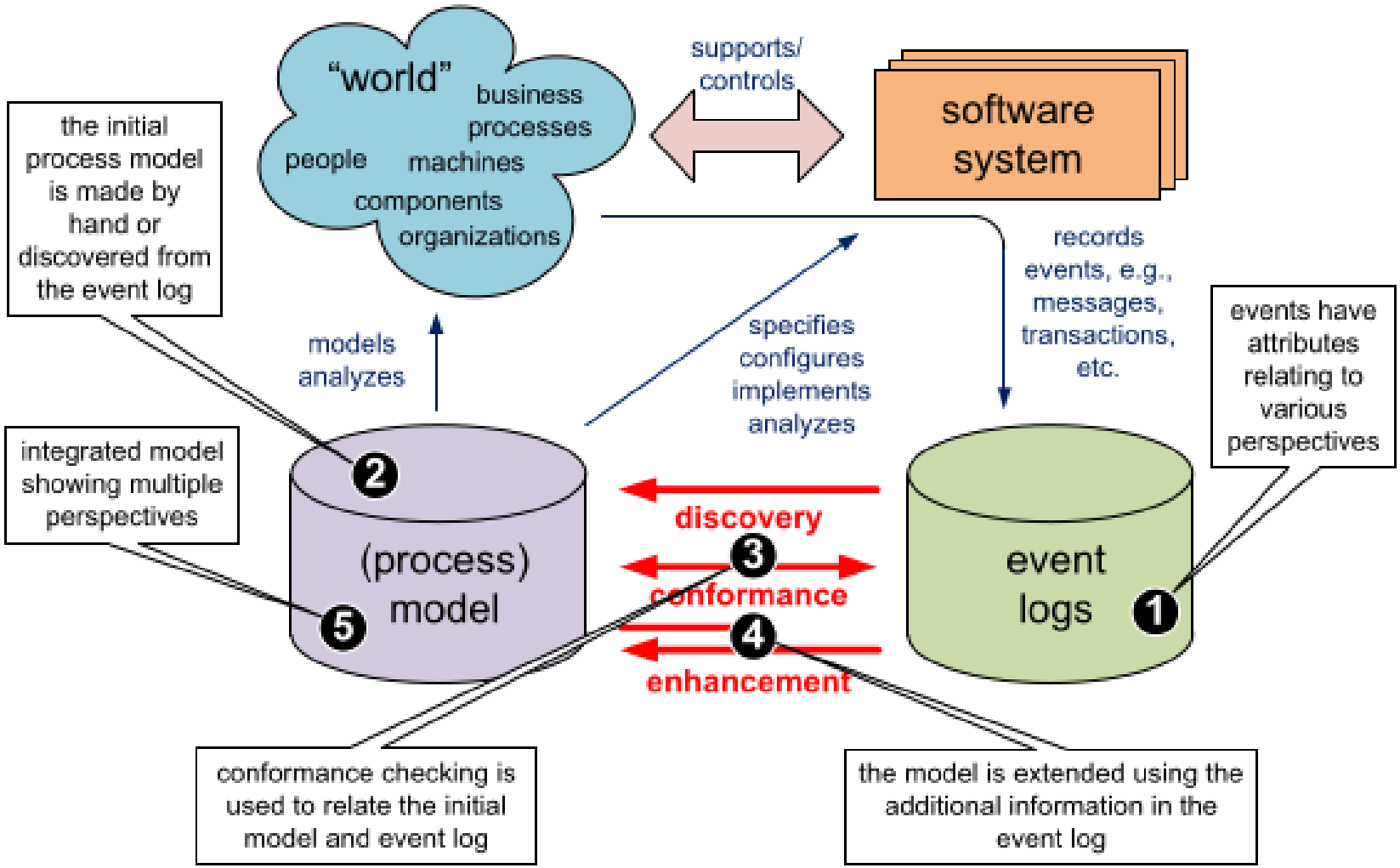
- Organizational mining can be used to get insight into typical work patterns, organizational structures, and social networks
- Timestamps and frequencies of activities can be used to identify bottlenecks and diagnose other performance related problems
- Case data can be used to better understand decision-making and analyze differences among cases

Idea

What we want to do?

Event logs may contain a wealth of information relating to other perspectives such as the organizational perspective, the case perspective, and the time perspective

- Organizational mining can be used to get insight into typical work patterns, organizational structures, and social networks
- Timestamps and frequencies of activities can be used to identify bottlenecks and diagnose other performance related problems
- Case data can be used to better understand decision-making and analyze differences among cases


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

Extended Log example

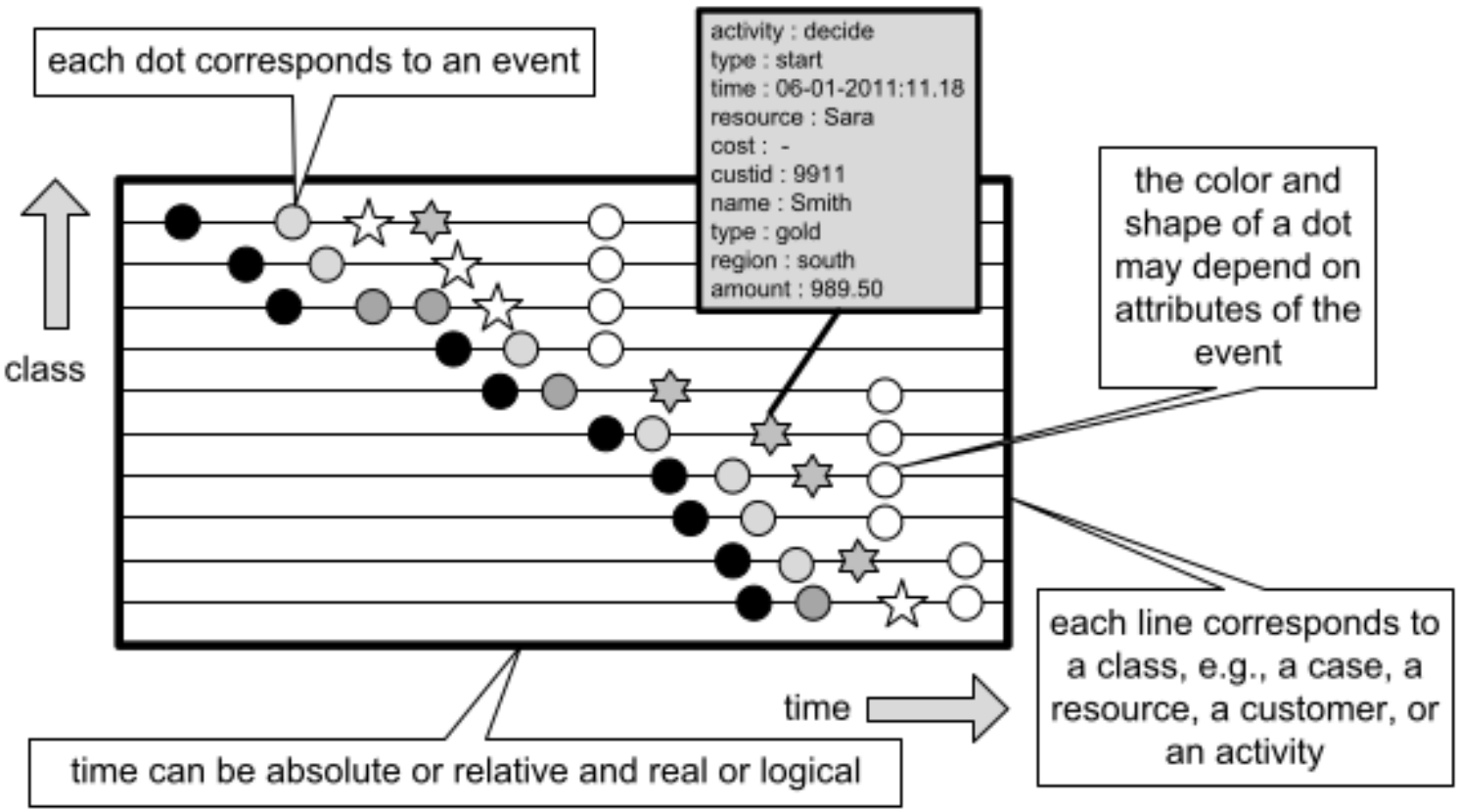
Case id	Event id	Properties				
		Time	Activity	Trans	Resource	Cost
1	35654423	30-12-2010:11.02	register request	start	Pete	
	35654424	30-12-2010:11.08	register request	complete	Pete	50
	35654425	31-12-2010:10.06	examine thoroughly	start	Sue	
	35654427	31-12-2010:10.08	check ticket	start	Mike	
	35654428	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

Dotted charts

A dotted chart provides an “helicopter view” over a log where each event is represented by a dot

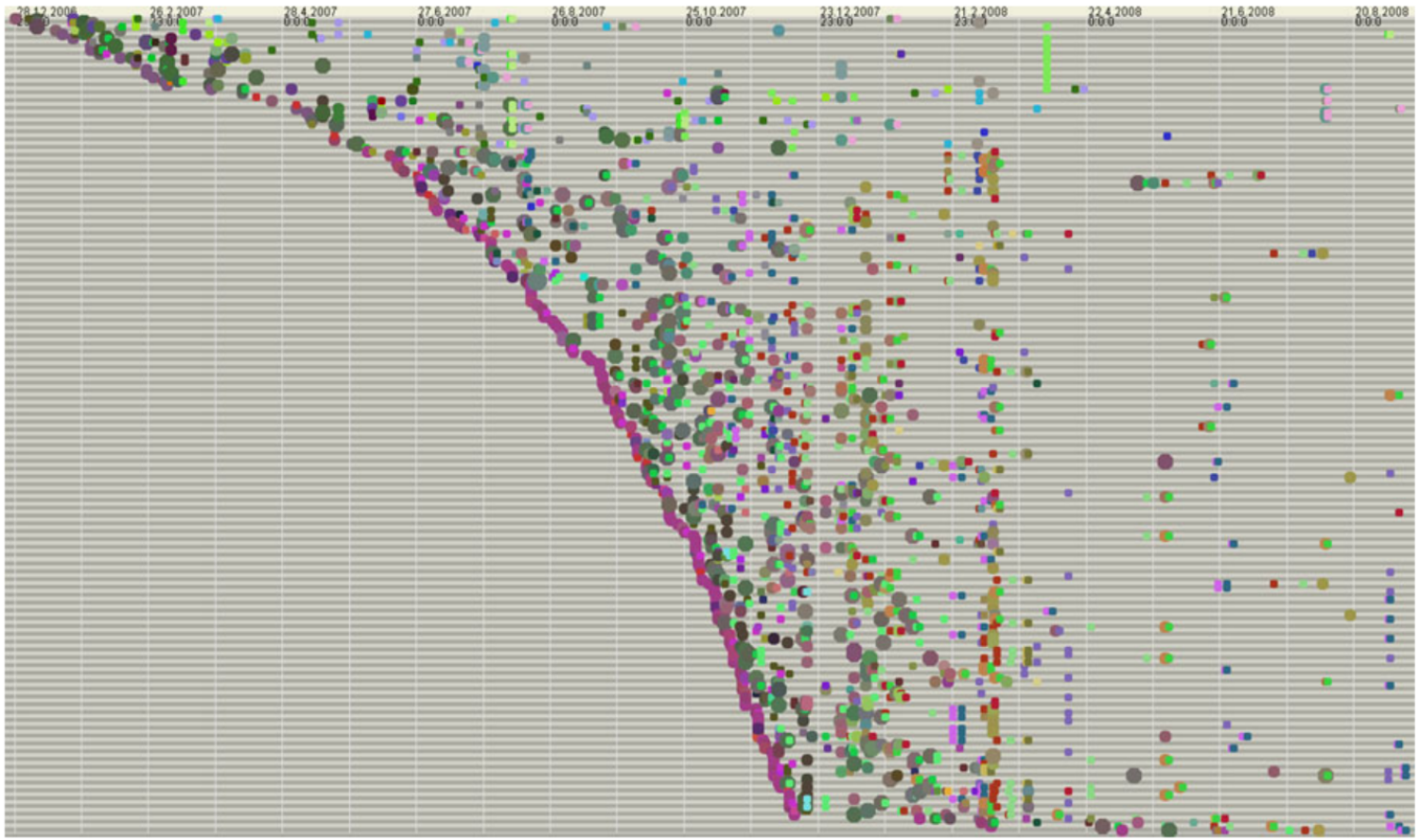
- Horizontal axis represent the time (absolute or relative - real or logical)
- Vertical axis related to a classifier (e.g. resource, case)

Dotted charts



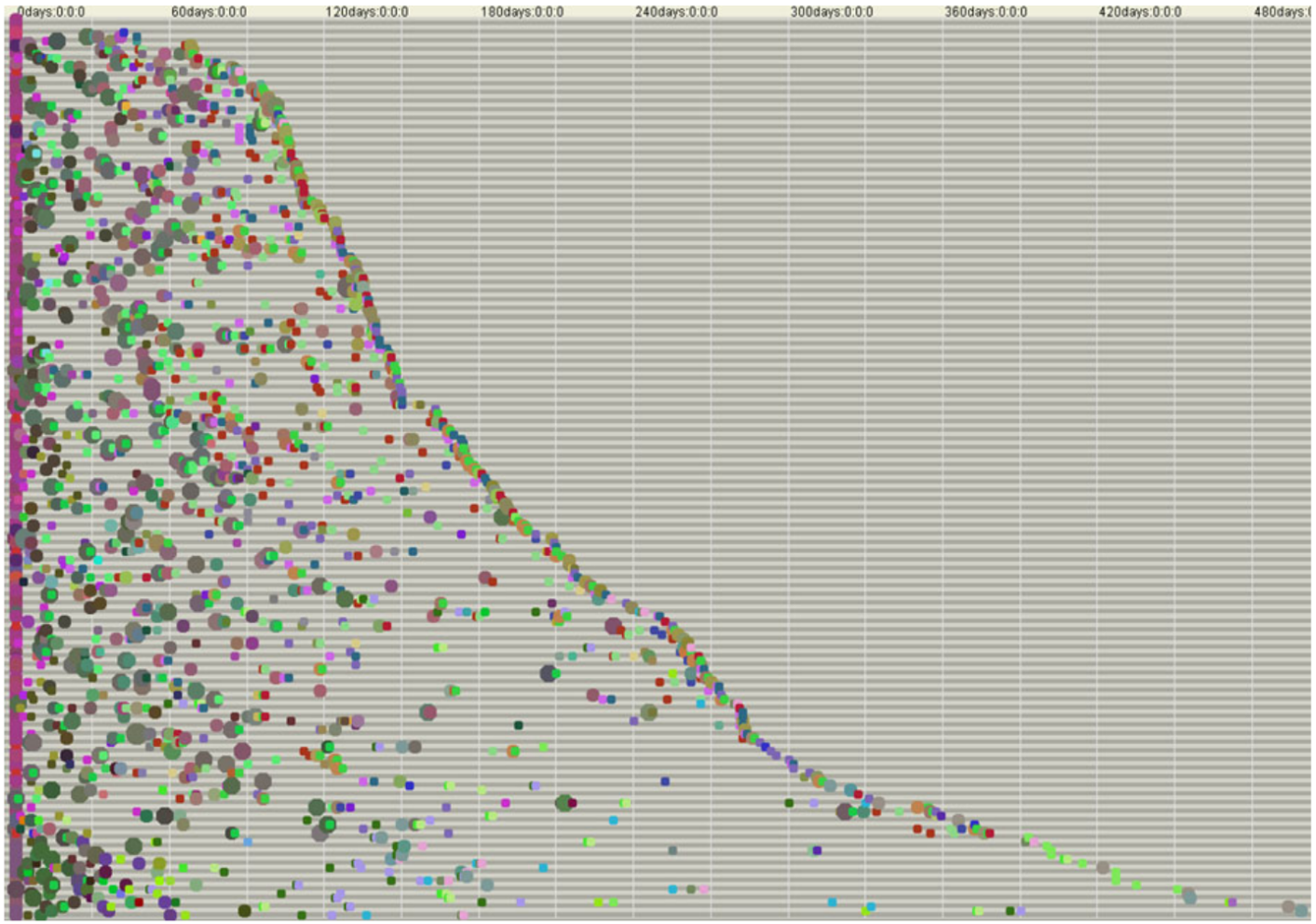
Dotted charts example

Events related to a housing agency related process (x-time/y-case)



Dotted charts example

Events related to a housing agency related process (x-time/y-case)



Organizational mining

Starting point for organizational mining is typically the $\#_{resource}(e)$ attribute present in most event logs

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

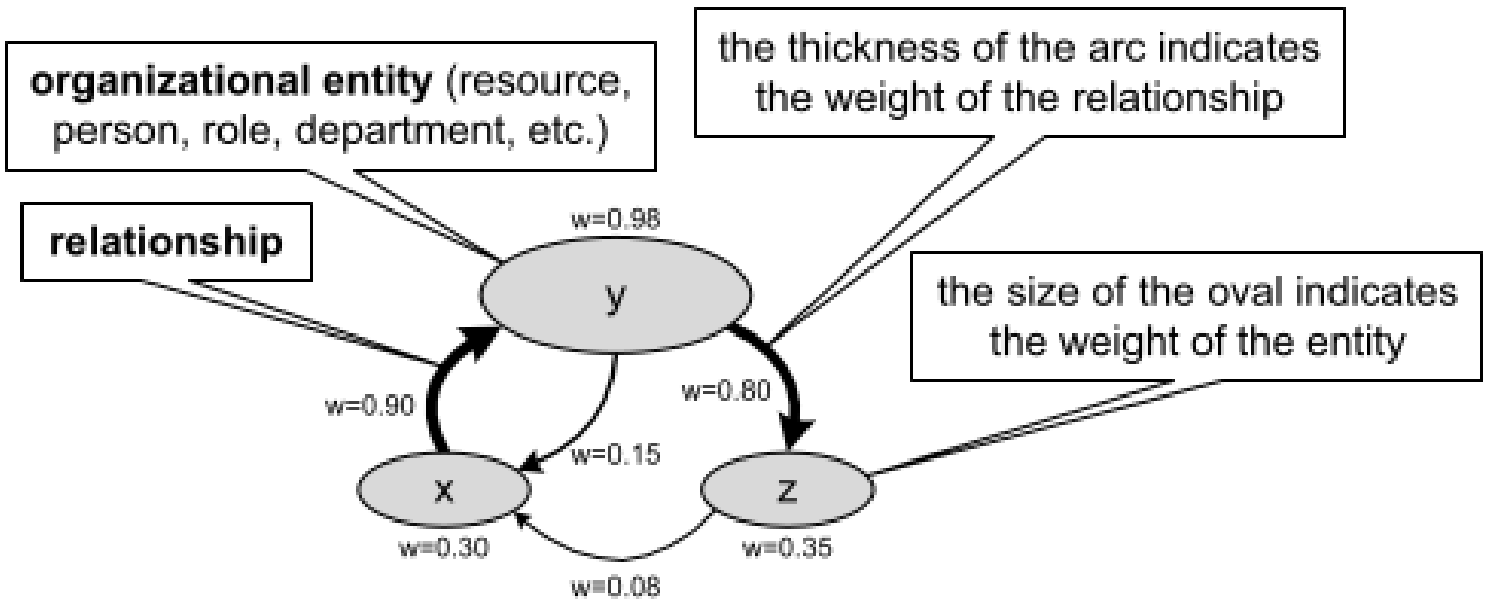
	a	b	c	d	e	f	g	h
Pete	0.3	0	0.345	0.69	0	0	0.135	0.165
Mike	0.5	0	0.575	1.15	0	0	0.225	0.275
Ellen	0.2	0	0.23	0.46	0	0	0.09	0.11
Sue	0	0.46	0	0	0	0	0	0
Sean	0	0.69	0	0	0	0	0	0
Sara	0	0	0	0	2.3	1.3	0	0

a = register request, b = examine thoroughly, c = examine casually, d = check ticket, e = decide,

f = reinitiate request, g = pay compensation, and h = reject request

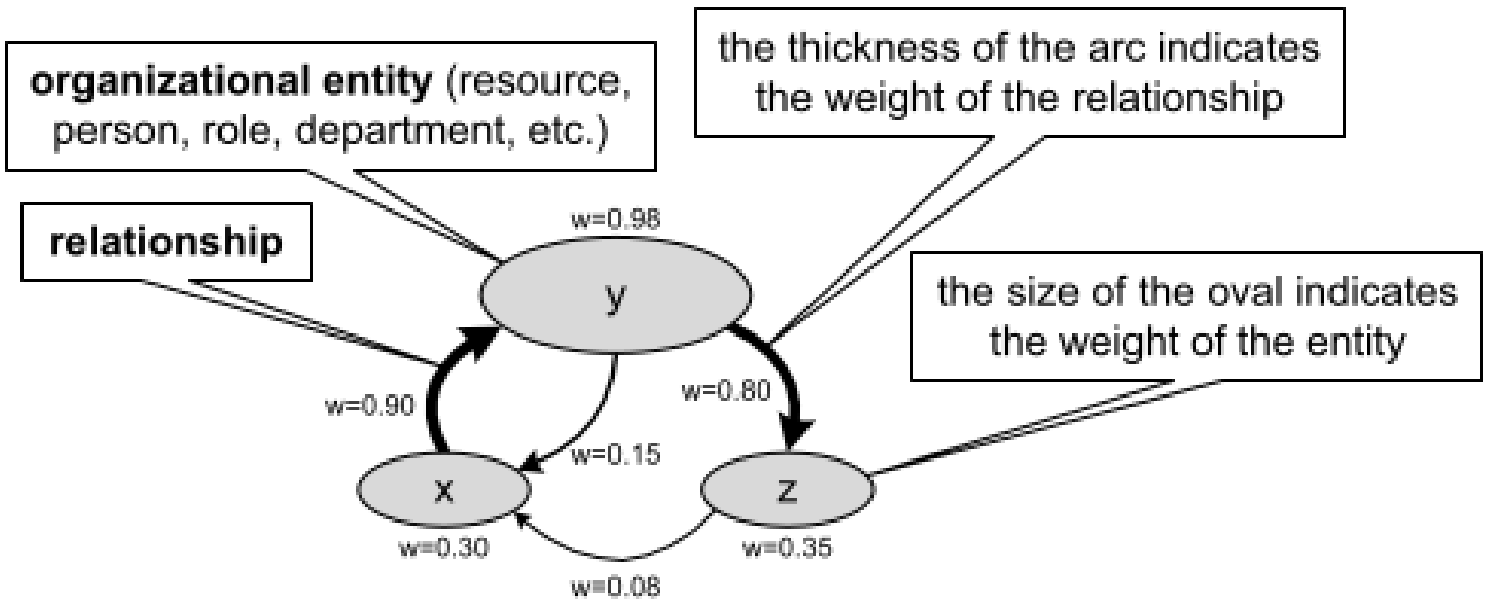
Social Network Analysis

Data coming from the log can be used to reconstruct “social relation” among the resources



Social Network Analysis

Data coming from the log can be used to reconstruct “social relation” among the resources



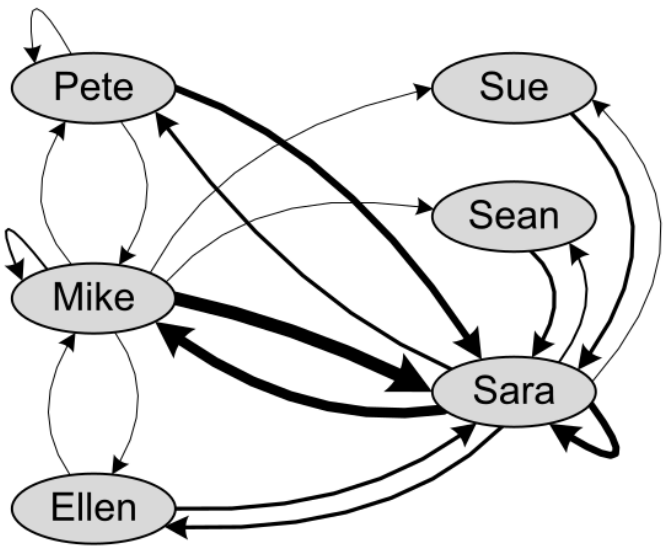
A wide variety of metrics have been defined to analyze social networks and to characterize the role of individual nodes in such a diagram.
E.g. Bavelas-Leavitt index of centrality (based on geodesic paths):

$$BL(i) = \frac{\Sigma_{j,k} D_{j,k}}{\Sigma_{j,k} (D_{j,i} + D_{i,k})}$$

Social network example

The matrix reports the handover of work showing the mean number of handovers from one person to another per case, the figures represent the corresponding social network using a threshold of 0,1:

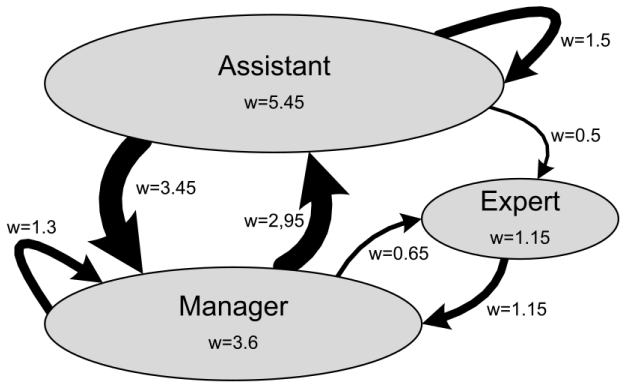
	Pete	Mike	Ellen	Sue	Sean	Sara
Pete	0.135	0.225	0.09	0.06	0.09	1.035
Mike	0.225	0.375	0.15	0.1	0.15	1.725
Ellen	0.09	0.15	0.06	0.04	0.06	0.69
Sue	0	0	0	0	0	0.46
Sean	0	0	0	0	0	0.69
Sara	0.885	1.475	0.59	0.26	0.39	1.3



Social networks additional insights

Resources can be clustered to consider roles in the organizational model

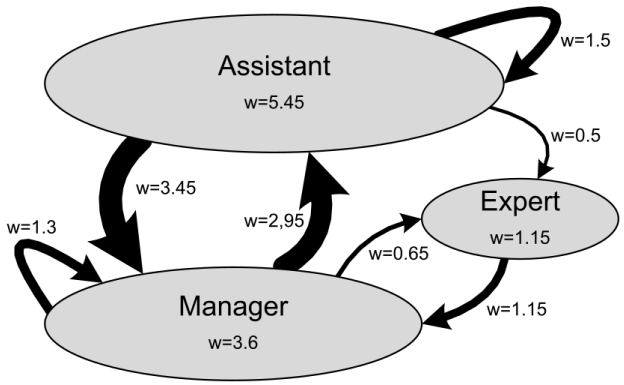
	Assistant	Expert	Manager
Assistant	1.5	0.5	3.45
Expert	0	0	1.15
Manager	2.95	0.65	1.3



Social networks additional insights

Resources can be clustered to consider roles in the organizational model

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

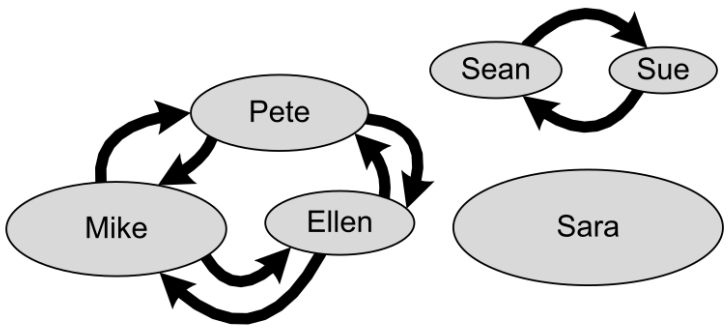


Network can be built on relations defined on similarities of profiles:

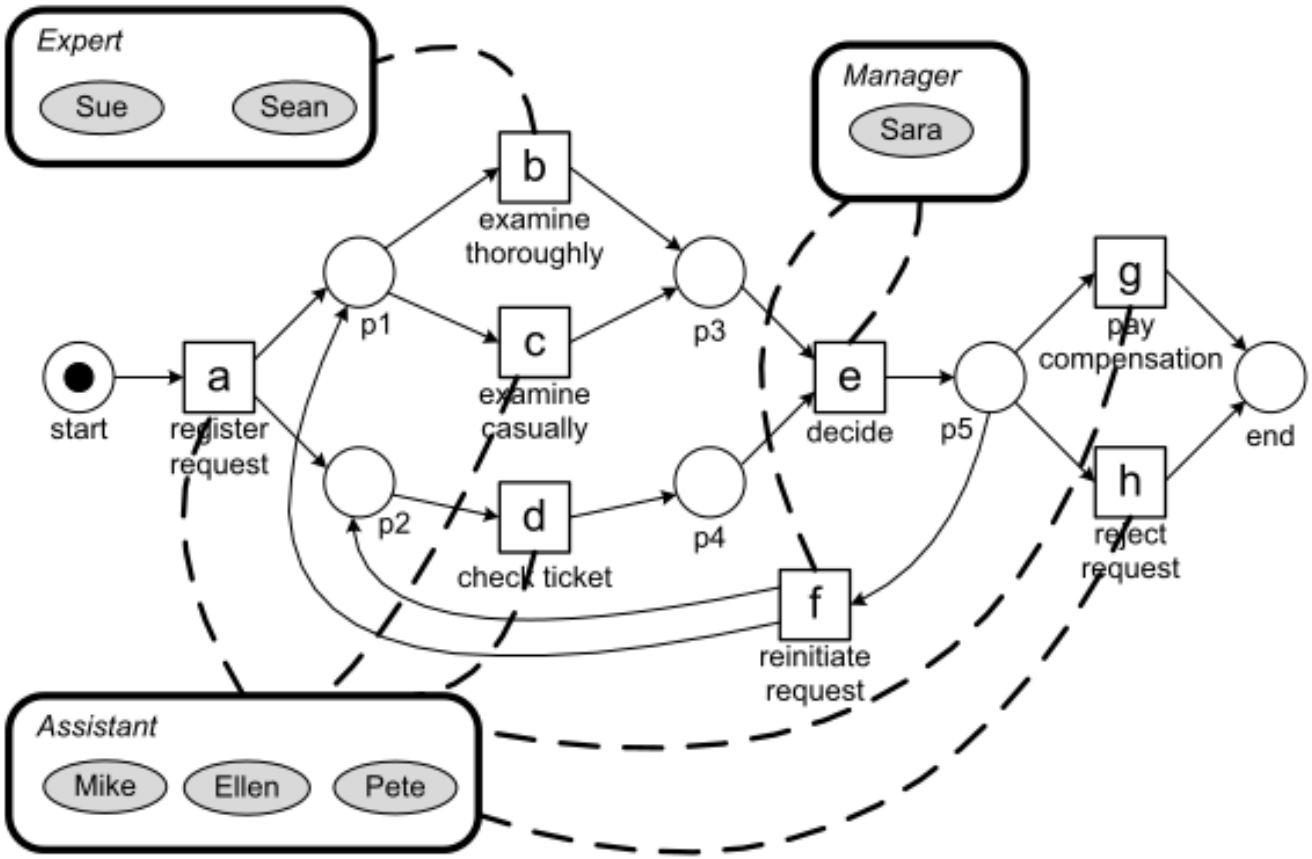
- different metrics can be adopted (e.g. Minkowski distance, Hamming distance) Pearson's correlation coefficient:

$$r_{x,y} = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}}$$

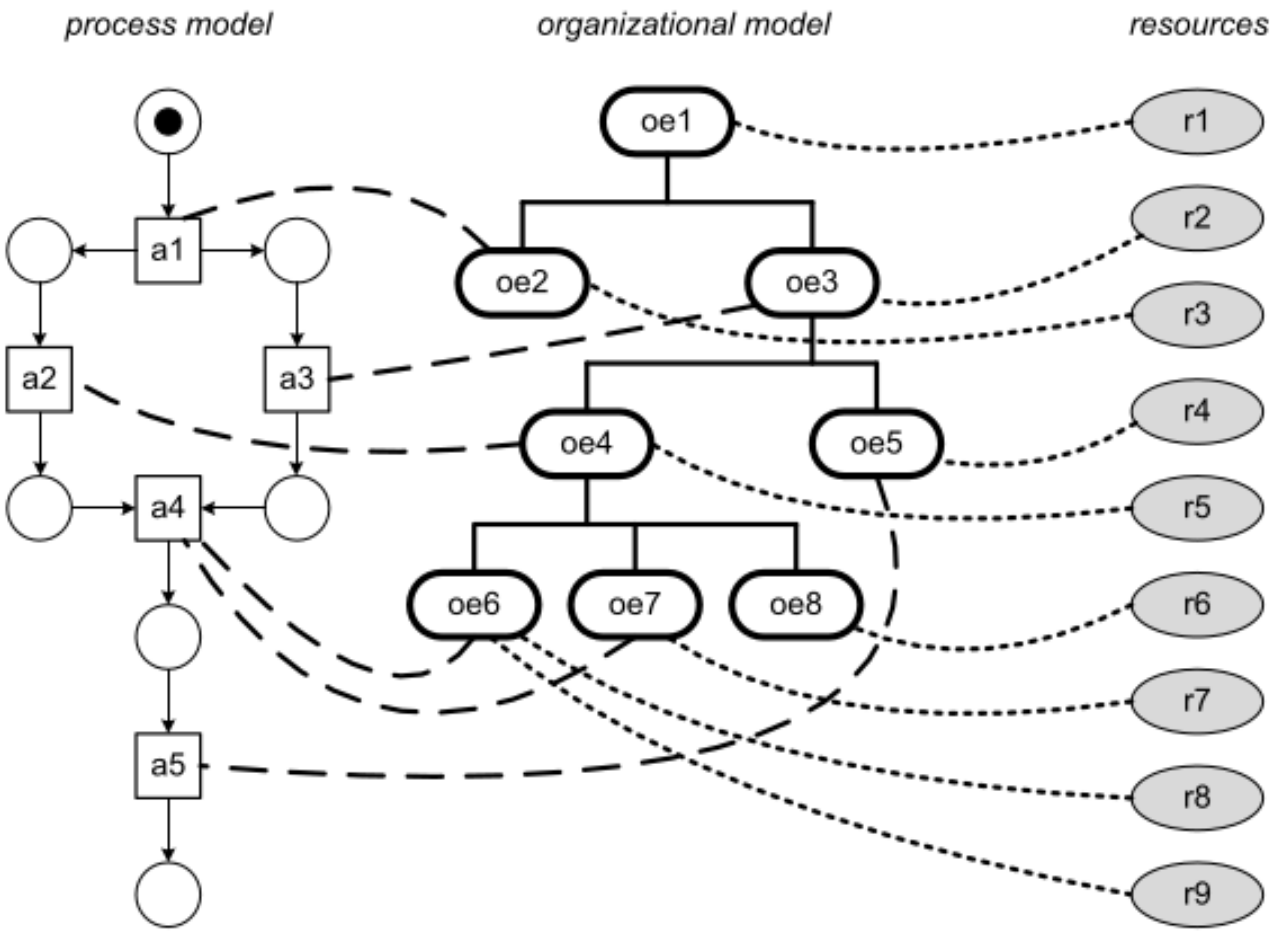
- clustering techniques can be used



Disovered organizational model



Organization and activities



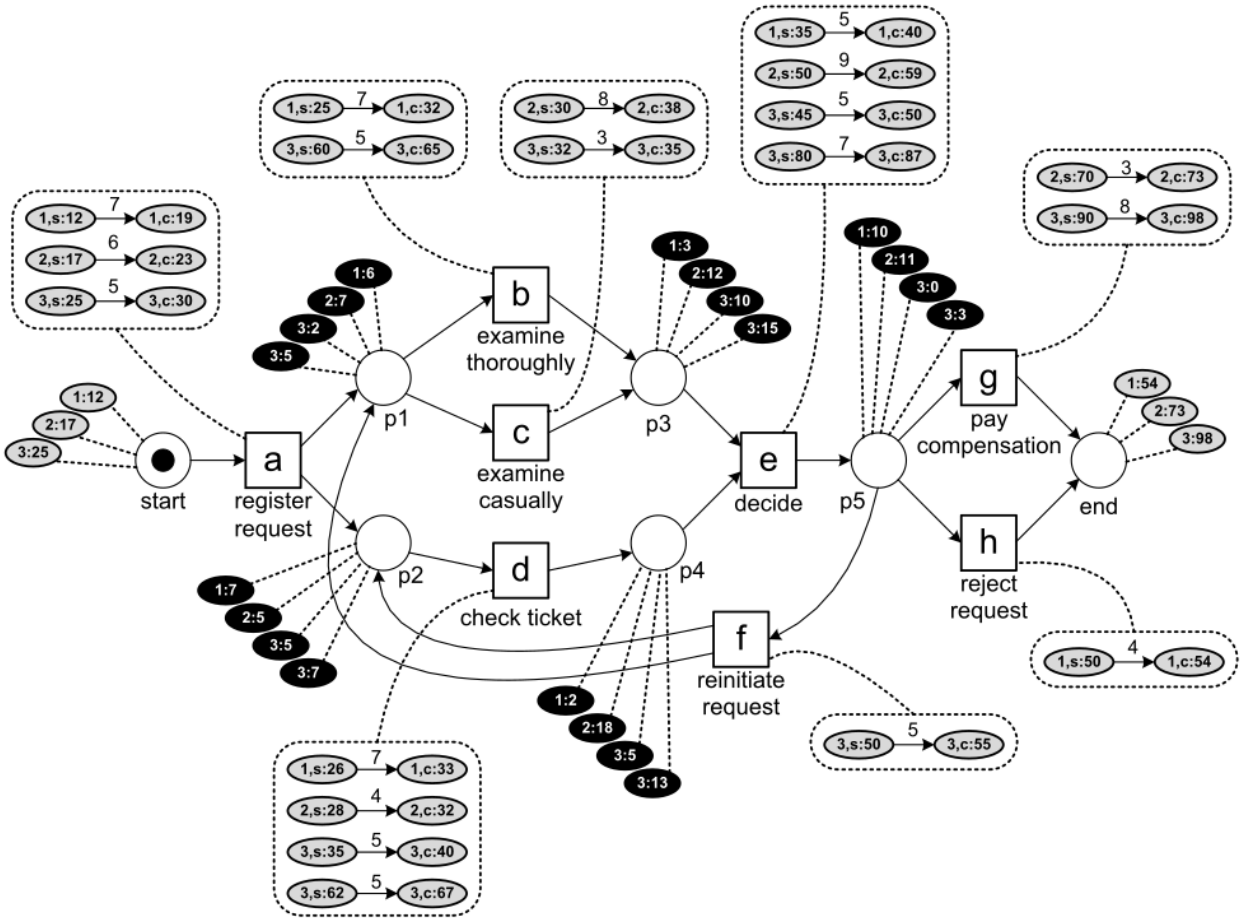
Privacy and anonymization

Event logs may contain sensitive or private data

- workers rights
- patients rights
- ...

Substituting identifiers with unrelated strings could not solve the problem

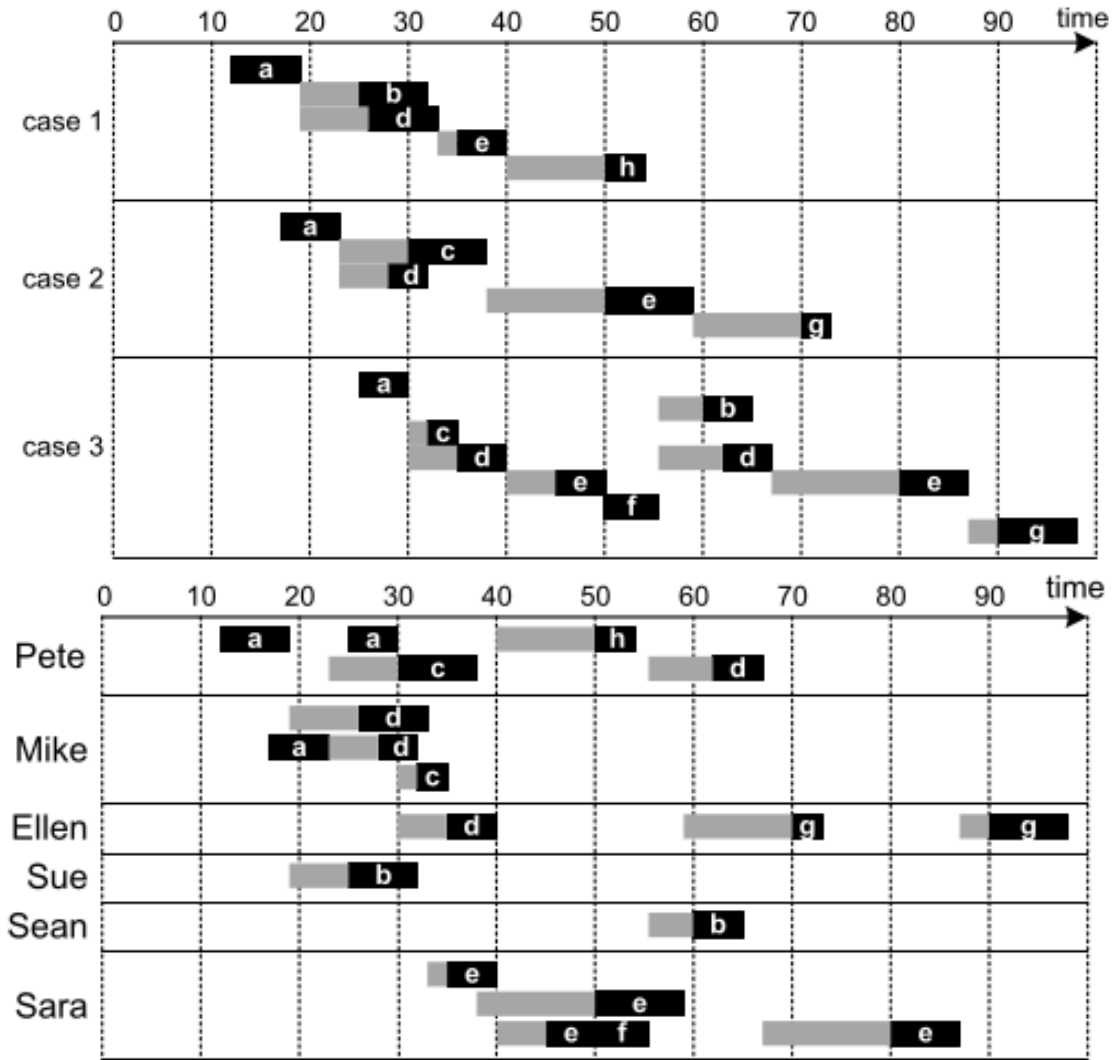
Time and probabilities



Intresting analysis

- Visualization of waiting and service times
- Bottleneck detection and analysis
- Flow time and SLA analysis
- Analysis of frequencies and utilization

Timeline related visualization



Decision Mining

Decision mining aims to find rules explaining choices in terms of characteristics of the case

- a classification technique like decision tree learning can be used to find rules (results have to be carefully evaluated)
- In case relevant data (predictor variables) are not included in the log it is still possible to infer probabilities for the decision

Bringing it all together