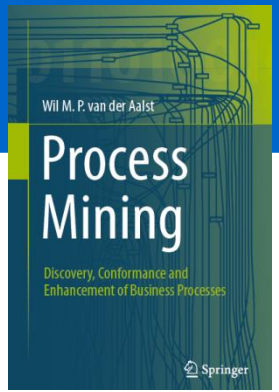
Process Mining: Data Science in Action

Organizational Mining

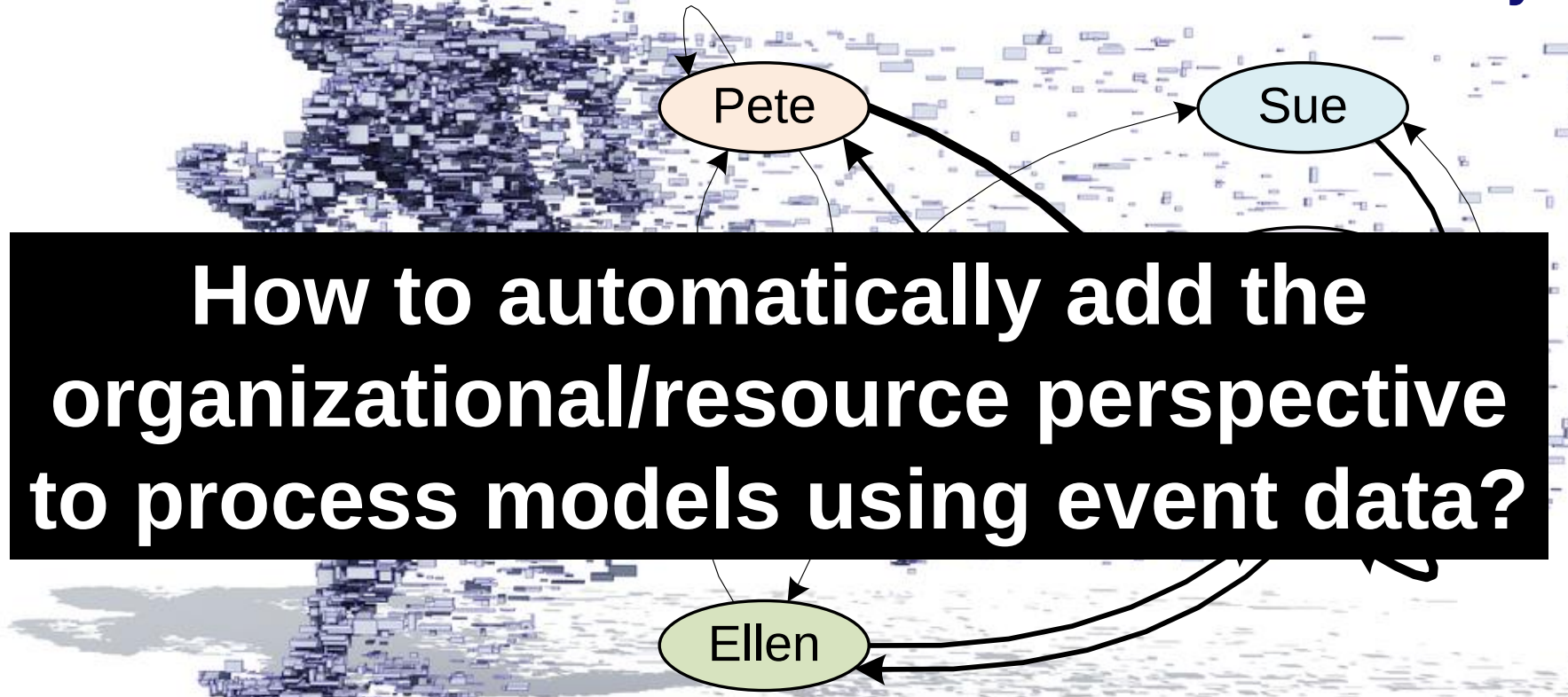


prof.dr.ir. Wil van der Aalst
www.processmining.org

Technische Universiteit Eindhoven University of Technology

Where innovation starts

Previous lecture:
Social network analysis



How to automatically add the organizational/resource perspective to process models using event data?

Resource-activity matrix

(see previous lecture)

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	$\langle a^{Pete}, c^{Mike}, d^{Ellen}, e^{Sara}, f^{Sara}, b^{Sean}, d^{Pe}$
4	
5	
6	
..	

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

TU/e

Question: Which resources are similar?

	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

Suppose that you need to make three groups with similar resources. What would these groups be?

TU/e

Answer

	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

{ Pete, Mike, Ellen } { Sue, Sean } { Sara }

TU/e

Distance based on resource-activity matrix

Standard notions of "distance" can be used e.g., Euclidian distance, Manhattan distance, Minkowski distance, and Pearson's correlation coefficient.

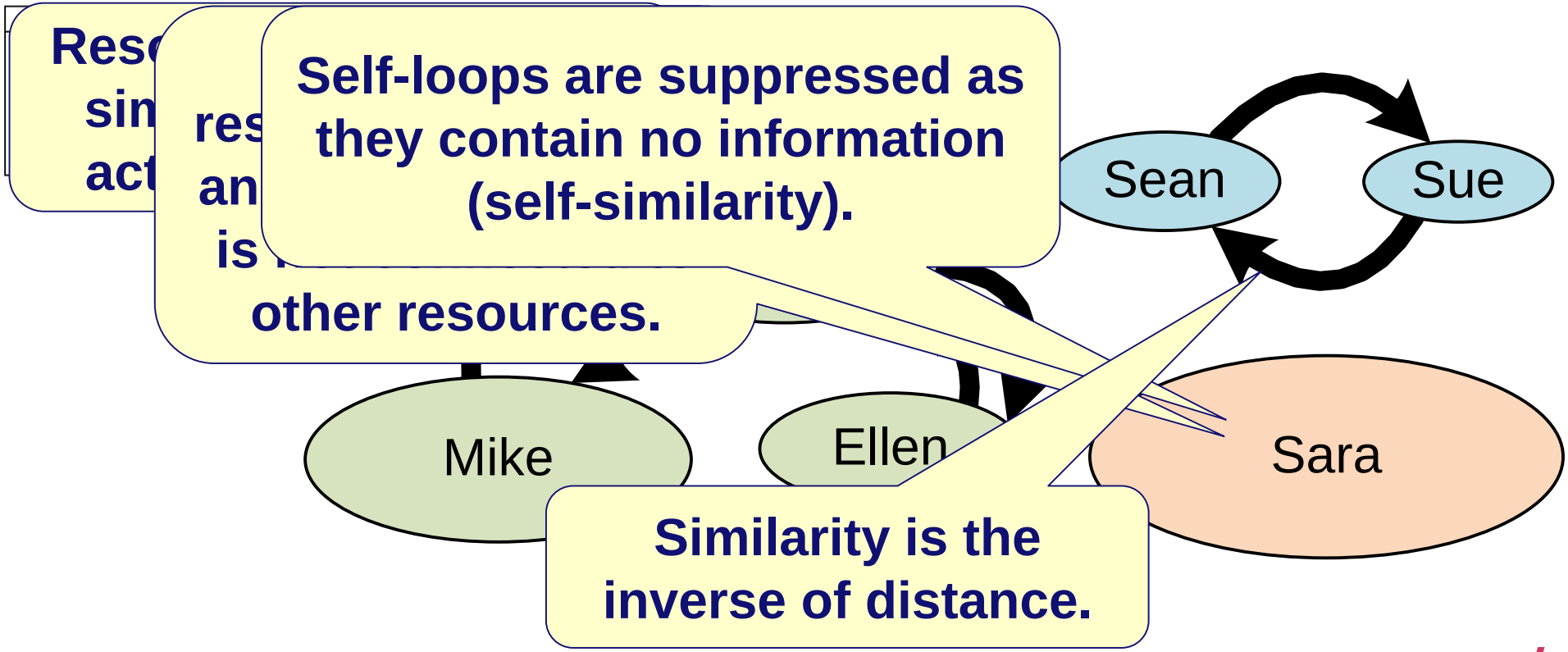
P_{Pete} = (0.3 , 0, 0.345 , 0.69 , 0 , 0 , 0.135 , 0.165)

P_{Mike} = (0.5 , 0 , 0.575, 1.15 , 0 , 0 , 0.225, 0.275)

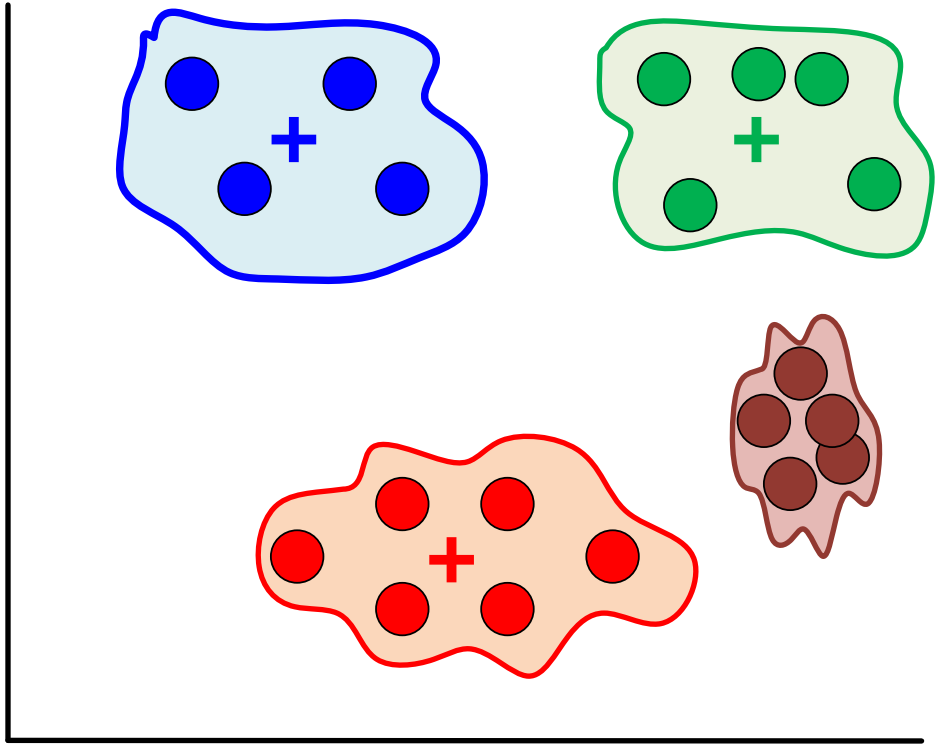
P_{Sara} = (0 , 0 , 0 , 0 , 2.3 , 1.3 , 0 , 0)

TU/e

Social network based on similarity of profiles



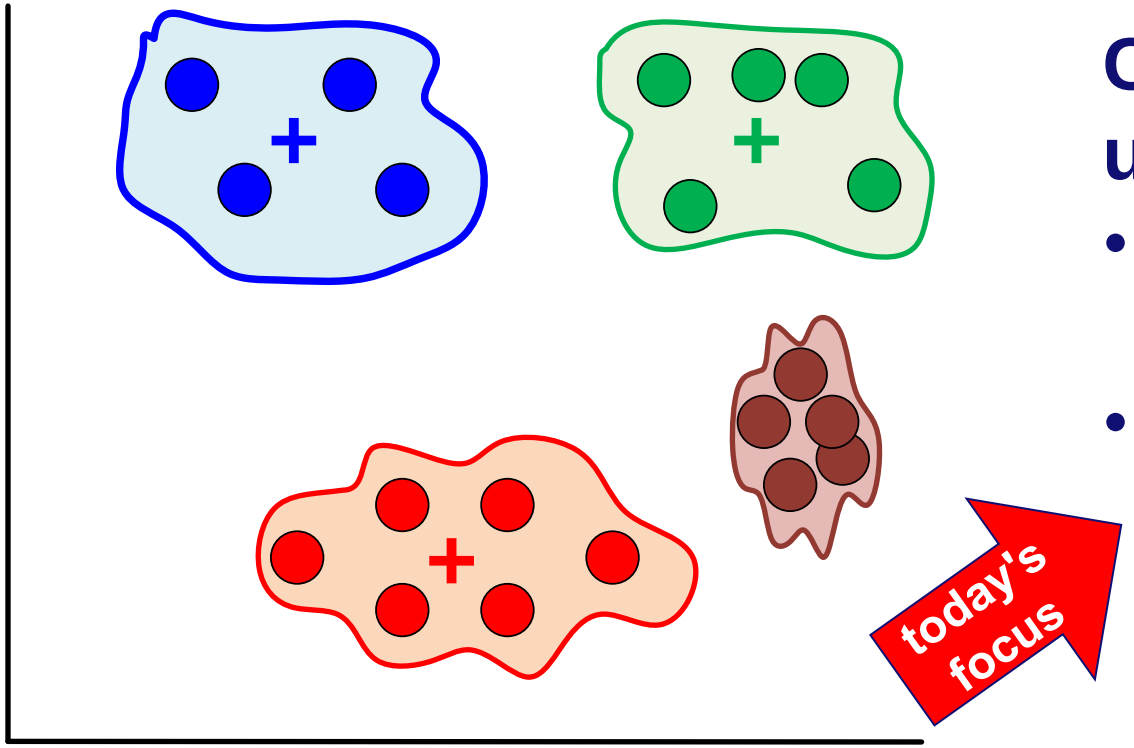
Related to clustering
(basic data mining technique discussed before)



- *k*-means clustering
- agglomerative hierarchical clustering

TU/e

Clustering: cases and resources



Clustering may be used for:

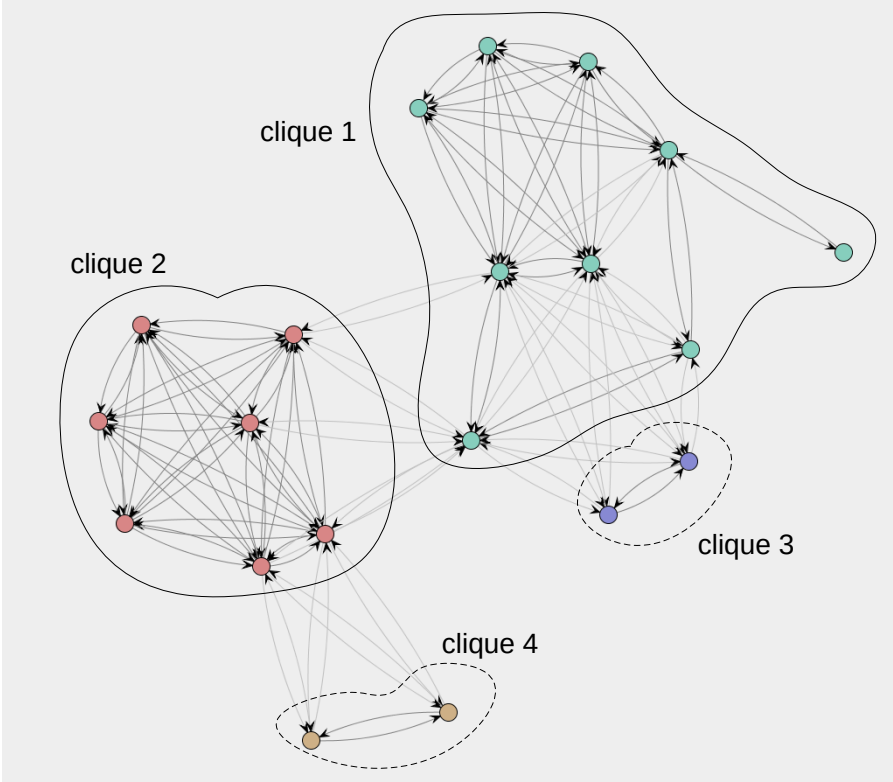
- **grouping cases (process variants)**
- **grouping resources (identifying roles)**

TU/e

Real-life example: Roles found by ProM

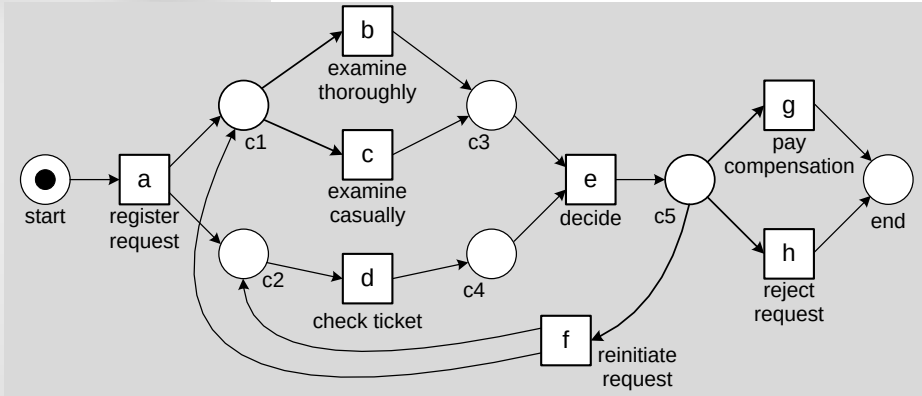
user	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}
user 1	0	0	51	0	0	0	0	0	0	0	0	0	0
user 2	1	2	0	0	2	0	0	0	0	38	0	69	0
user 3	0	9	0	0	0	0	0	0	0	0	0	0	0
user 4	2	0	0	0	0	0	0	0	0	0	0	0	0
user 5	117	0	4	0	3	0	0	0	0	1	0	20	6
user 6	172	6	14	0	7	3	0	0	1	2	0	48	53
user 7	1	41	8	14	275	8	8	865	55	180	0	128	5
user 8	2	868	7	6	105	0	0	79	266	441	0	844	3
user 9	90	0	2	0	1	2	0	0	1	2	0	27	28
user 10	0	0	0	899	0	0	0	0	0	0	0	0	1019
user 11	336	1	3	1	4	2	0	0	0	1	0	18	23
user 12	1	645	13	21	419	3	0	3	217	281	1	334	9
user 13	0	1	0	0	0	0	0	0	0	0	0	0	0
user 14	0	0	0	0	0	0	0	0	0	1	0	0	0
user 15	0	0	0	0	0	0	0	2	2	0	0	2	0
user 16	1	3	3	2	1	0	0	1	2	3	1	0	0
user 17	0	4	0	0	0	0	0	0	0	0	0	0	0
user 18	9	0	0	0	0	0	0	0	0	0	0	0	0
user 19	13	1	0	0	1	0	0	0	0	0	0	4	0
user 20	0	0	0	21	0	0	0	0	0	0	0	0	258

(not normalized per case)

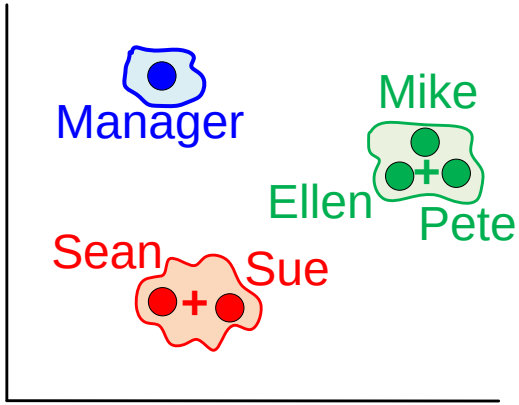


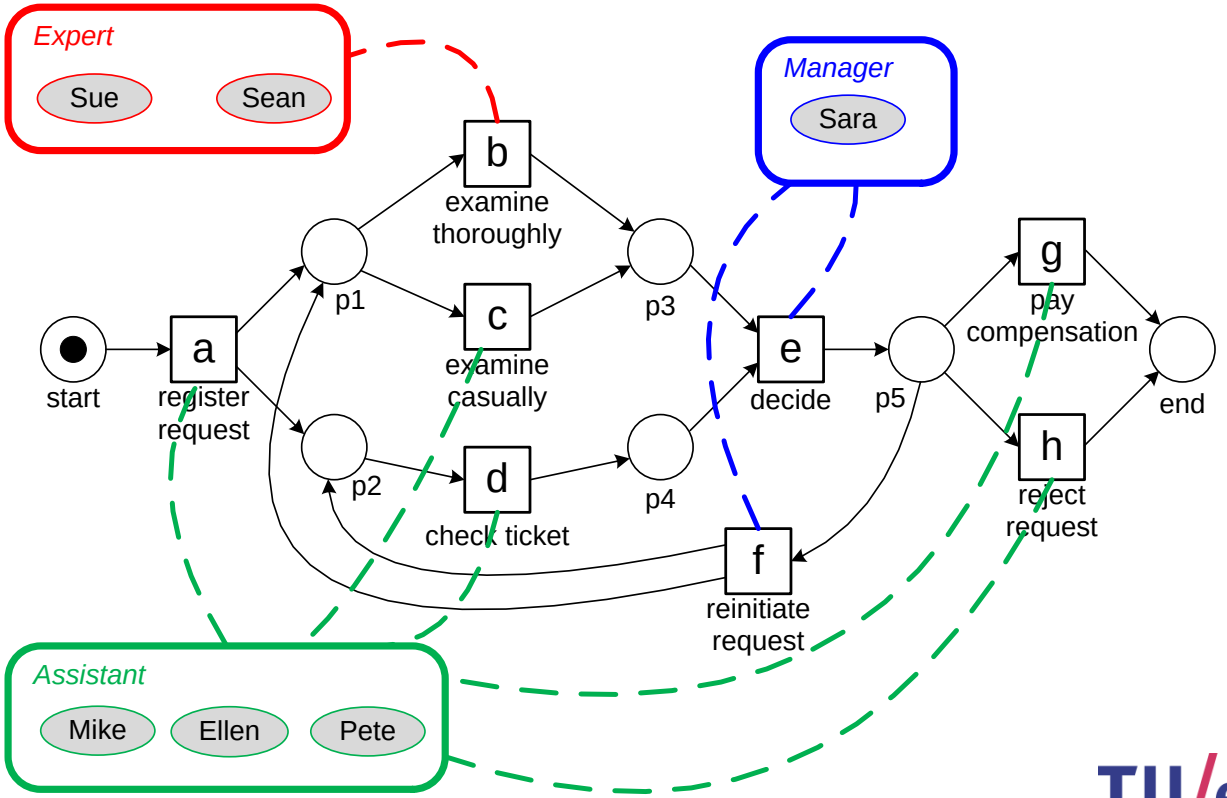
TU/e

assign activities to roles, groups, or other organizational entities



Extending process models with the organizational perspective



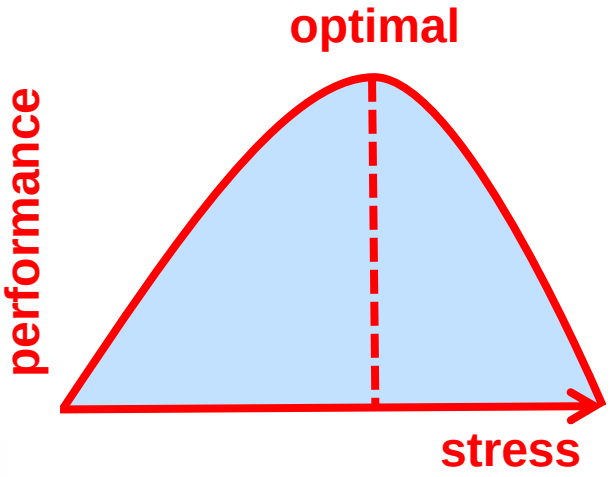


TU/e

Learning more about resources

- When are resources available?
  - part-time, vacation, illness, …
  - shared among different processes
- Which resources collaborate well?
- Which resources perform well on specific activities?
- …

Yerkes-Dodson law of arousal



All-inclusive process models are needed …

Processes, people, roles, and other organizational entities are intertwined and cannot be viewed in isolation.

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

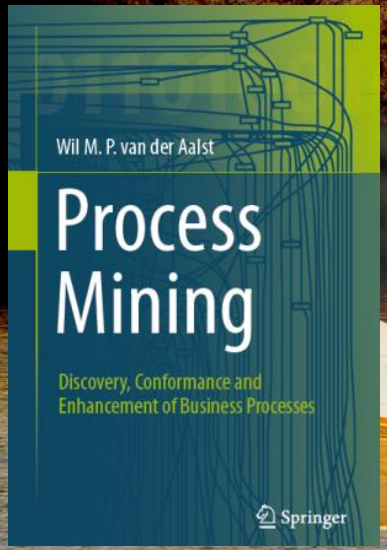
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

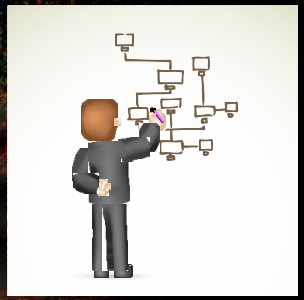
Part IV: Putting Process Mining to W...

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

Part V: Reflection

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





TU/e