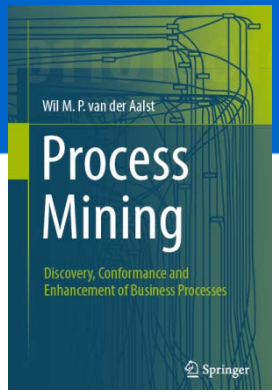
Process Mining: Data Science in Action

Mining Social Networks

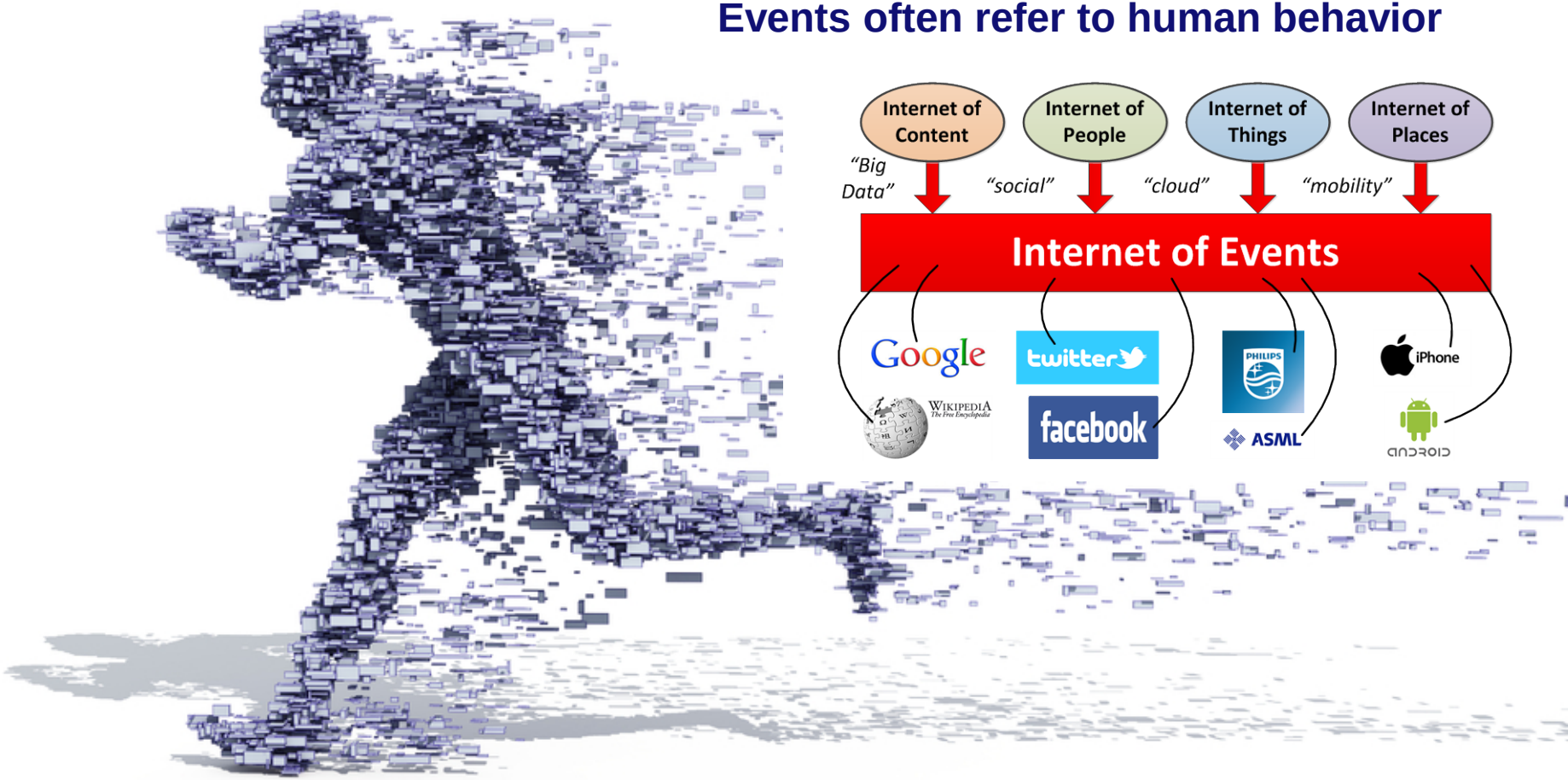


prof.dr.ir. Wil van der Aalst
www.processmining.org

TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

Events often refer to human behavior



Events having a resource attribute

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*)

TU/e

Resource-activity matrix

(mean number of times a resource performs an activity per case)

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

Resource-activity matrix

(mean number of times a resource performs an activity per case)

In 30% of the cases, a is executed by Pete, 50% is executed by Mike, and 20% is executed by Ellen.

Activities e and f

matrix provides basic

who is doing what".

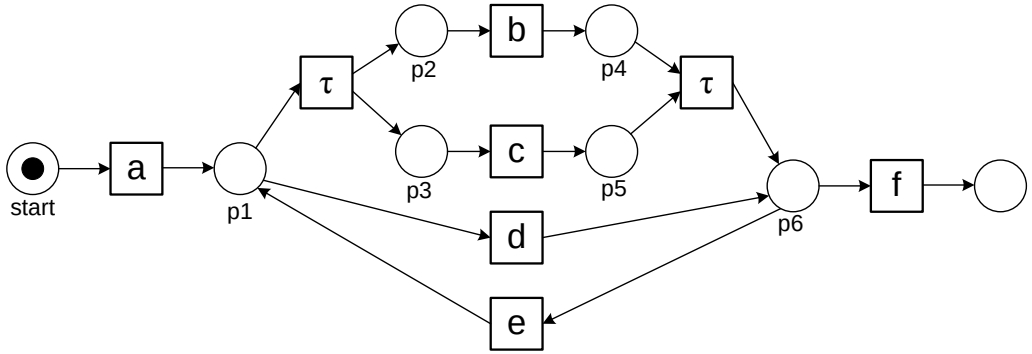
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
Pete	0.3	0	0.345	.69	0	0	0.135	0.165
Mike	0.5	0	0.575	15	0	0	0.225	0.275
Ellen	0.2	0	0.23	0.	0	0	0.09	0.11
Sue	0	0.46	0	0	0	0	0	0
Sean	0	0.69	0	0	0	0	0	0
Sara	0	0	0	0	2.3	1.3	0	0

Question: Create resource activity matrix

case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete

Question: Create resource activity matrix

Process context:



may take some time …

case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol

TU/e

Answer: Resource activity matrix

mean number of times a resource performs an activity per case

	a	b	c	d	e	f
Pete	1.00	0.00	0.00	0.00	0.00	0.00
Mary	0.00	0.67	0.00	0.67	0.00	0.00
Sue	0.00	0.00	0.67	0.00	0.00	0.00
Kirsten	0.00	0.00	0.00	0.00	0.33	0.00
Carol	0.00	0.00	0.00	0.00	0.00	1.00

case id	activity	type	time	resource
1	a	start	10	Pete
1	a	complete	12	Pete
1	c	start	15	Sue
2	a	start	16	Pete
2	a	complete	17	Pete
1	c	complete	18	Sue
3	a	start	20	Pete
2	b	start	22	Mary
2	b	complete	25	Mary
3	a	complete	28	Pete
1	b	start	30	Mary
1	b	complete	34	Mary
3	d	start	35	Mary
3	d	complete	37	Mary
2	c	start	40	Sue
1	f	start	42	Carol
2	c	complete	45	Sue
1	f	complete	46	Carol
2	e	start	50	Kirsten
3	f	start	51	Carol
2	e	complete	52	Kirsten
2	d	start	53	Mary
3	f	complete	55	Carol
2	d	complete	56	Mary
2	f	start	57	Carol
2	f	complete	60	Carol

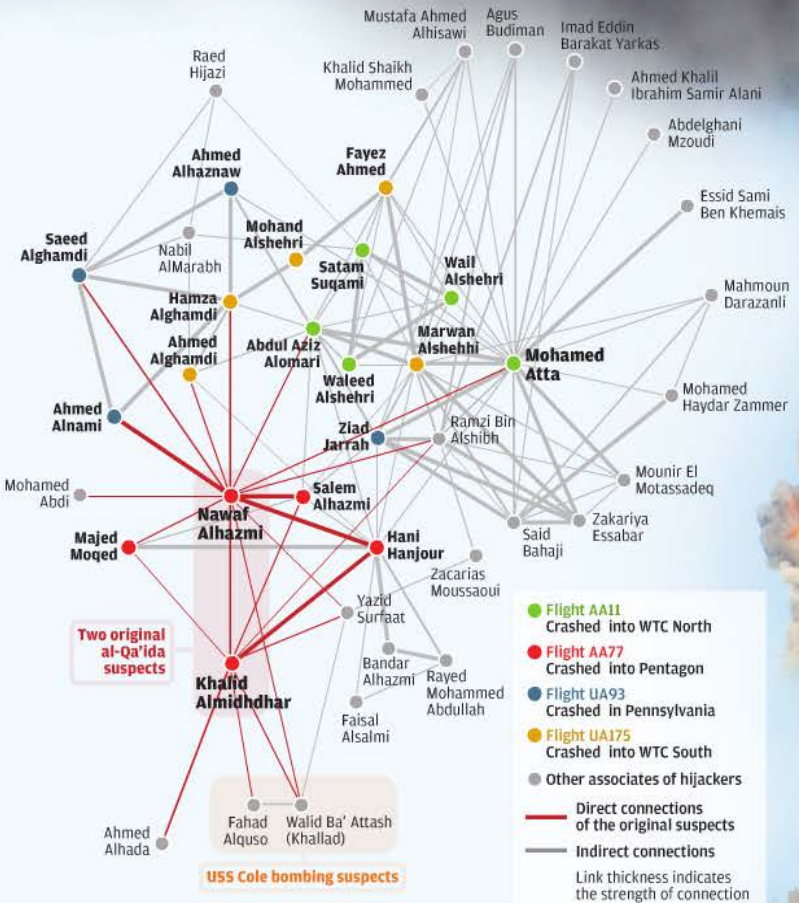
social networks

How two names and a sheaf of newspaper cuttings revealed the 9/11 team

This social network of the 19 hijackers behind the 9/11 attacks in the United States, and their associates, was drawn up at the end of 2001. Valdis Krebs, a commercial consultant in network analysis, started with newspaper reports of the two original terrorist suspects, Nawaf Alhazmi and Khalid Almihdhar. He then plotted the position of the other hijackers and associates. His analysis highlighted the central role played by Mohamed Atta. It also shows the close associations between the "Hamburg cell" that Atta set up, as well as the close links with the two original suspects – critical information that may have helped to avert an attack had it been known.

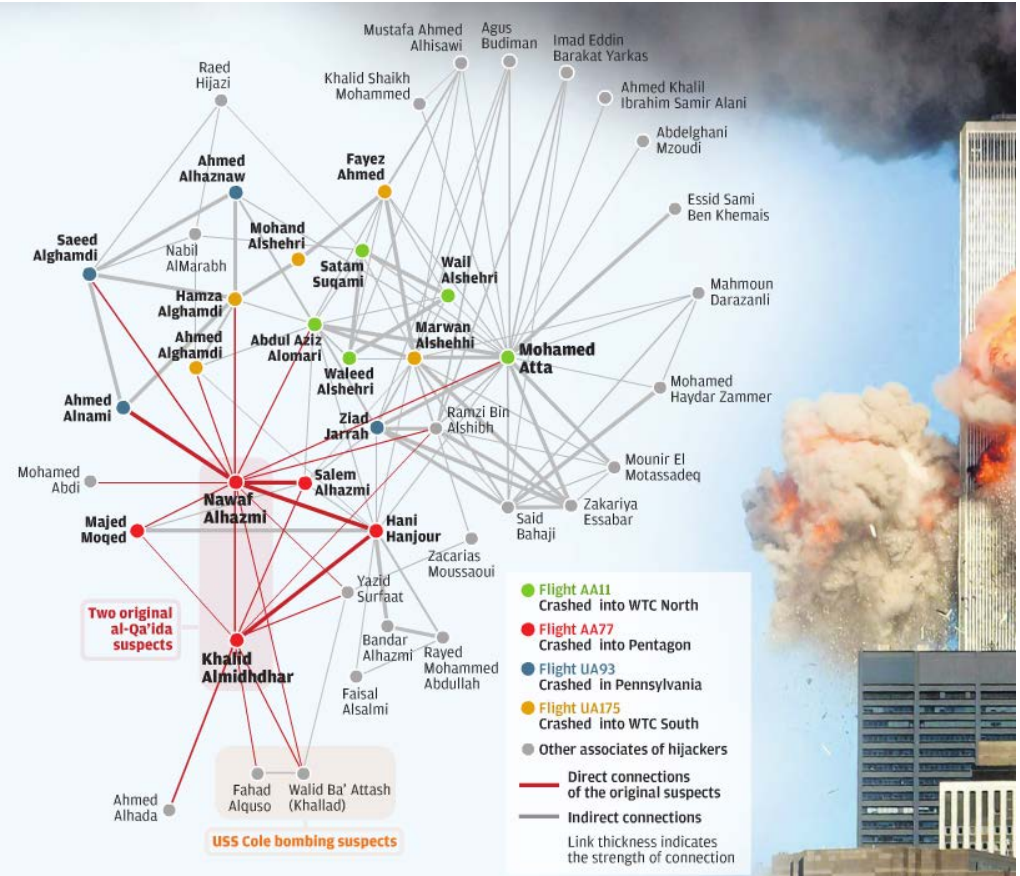
Mohamed Atta | Nawaf Alhazmi | Khalid Almidhdhar

Emergency services attend the scene after Flight AA77 crashes into the Pentagon



The second plane, Flight UA175, crashes into the South Tower of the World Trade Centre

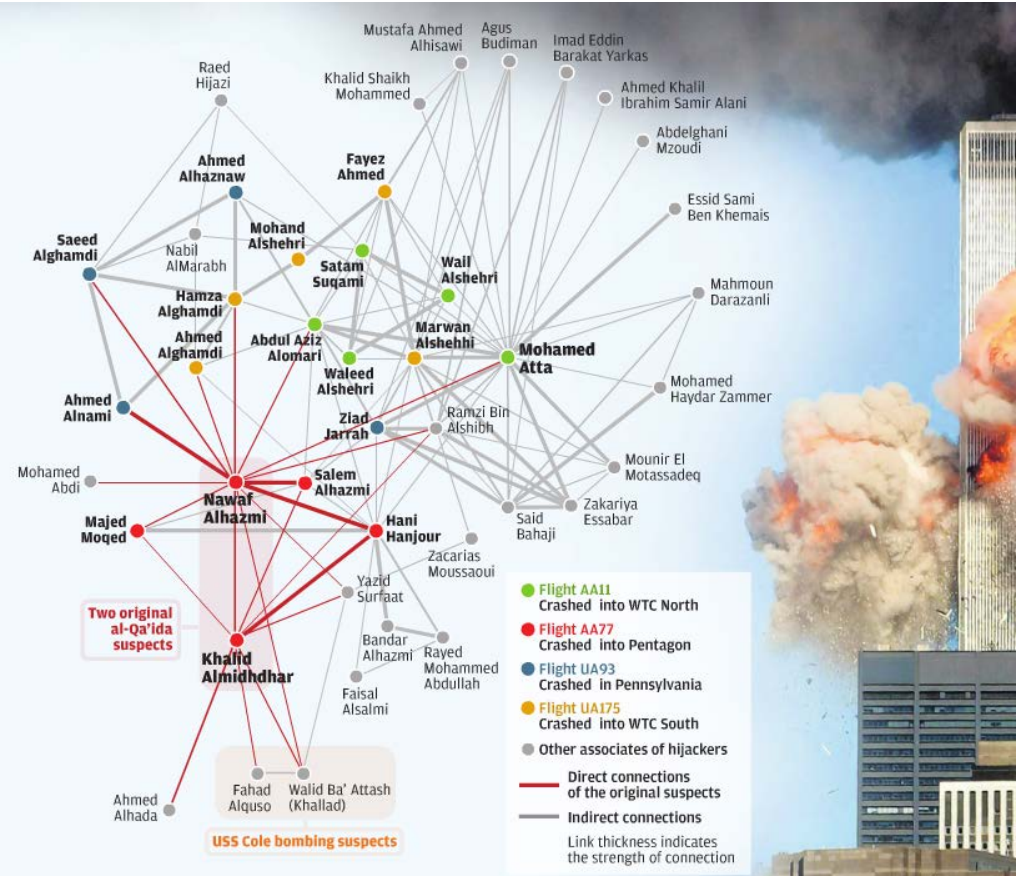
Social network analysis



- **Sociometry: present data on interpersonal relationships in graph or matrix form.**
- **Jacob Levy Moreno used such techniques in the 1930s to better assign students to residential cottages.**

TU/e

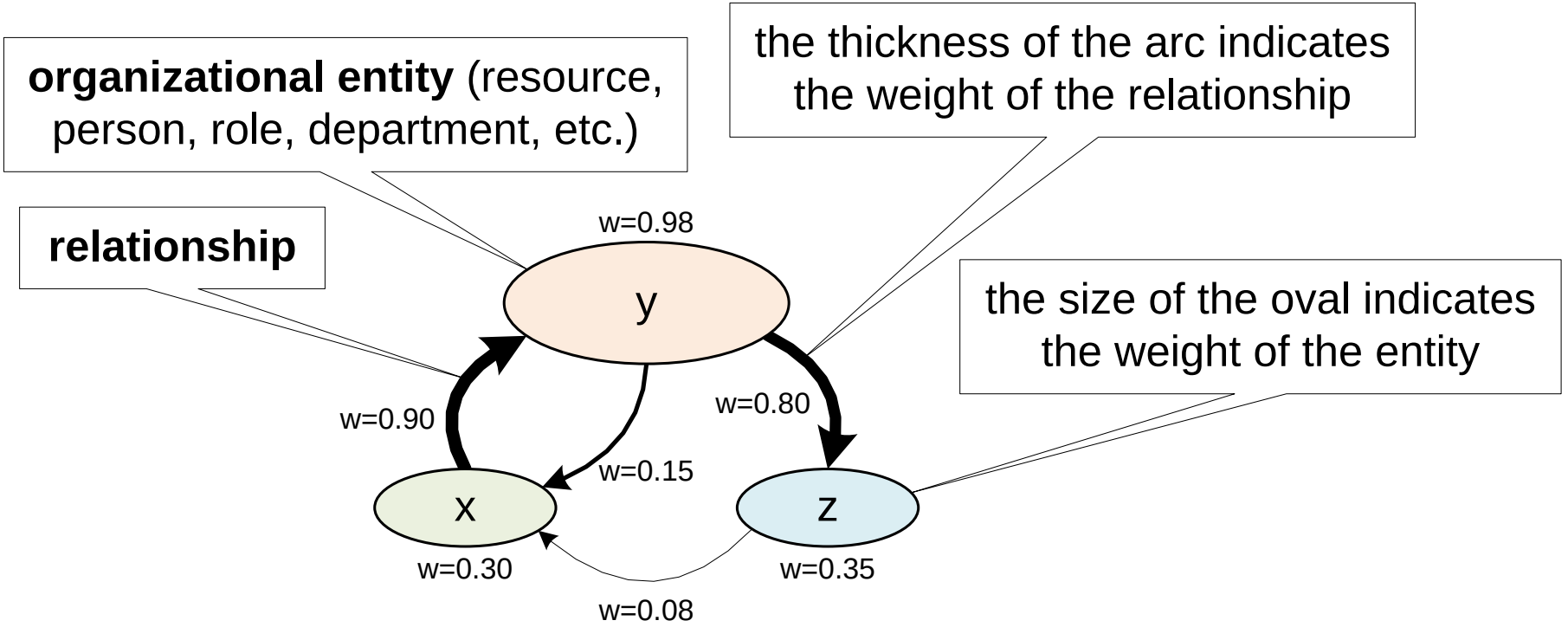
Social network analysis



- Arcs: weights or (inverted) distance.
- Metrics to denote importance:
  - centrality,
  - closeness,
  - betweenness.
  - …
- Identification of cliques.

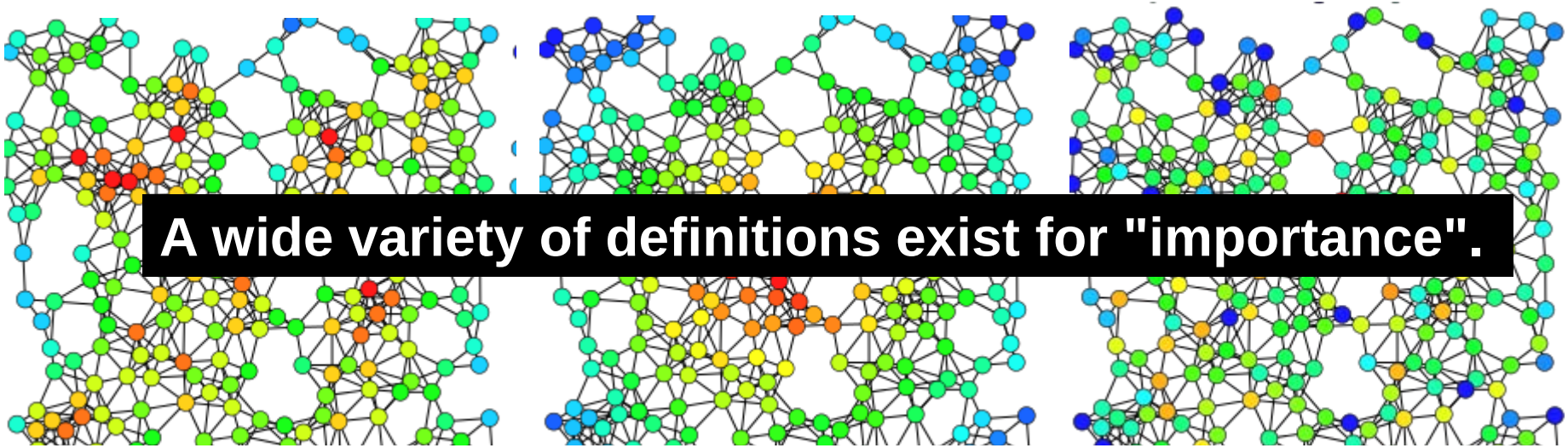
TU/e

Social network



Importance of nodes in a social network

(figures by Claudio Rocchini, CC BY-SA 3.0)



A wide variety of definitions exist for "importance".

degree centrality: number of connections a particular node has

closeness centrality: 1 divided by the sum of all shortest paths to a particular node

betweenness centrality: fraction of shortest paths between any two nodes passing a particular node

TU/e

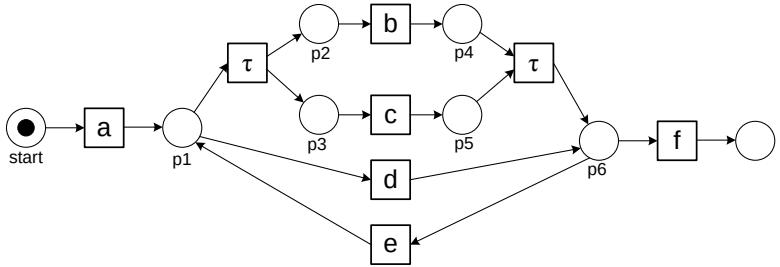
Handover of work matrix

On average Pete hands over work to himself 0.135 times per case.

On average Sara hands over work to Mike 1.475 times per case.

	Pe	Mike	Elle	Sue	Sean	Sara
Pete	0.135	0.225	0.	0.06	0.09	1.035
Mike	0.225	0.375		0.1	0.15	1.725
Ellen	0.09	0.15	.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

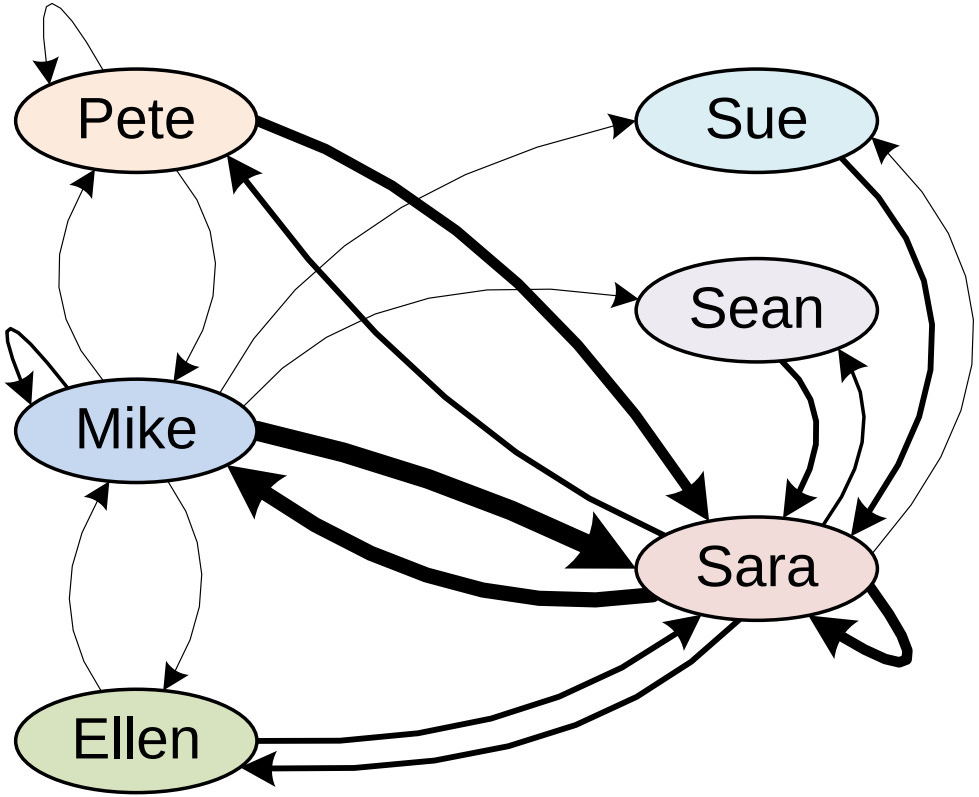
Process model defines "causality"



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

	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 based on handover of work (threshold of 0.1)

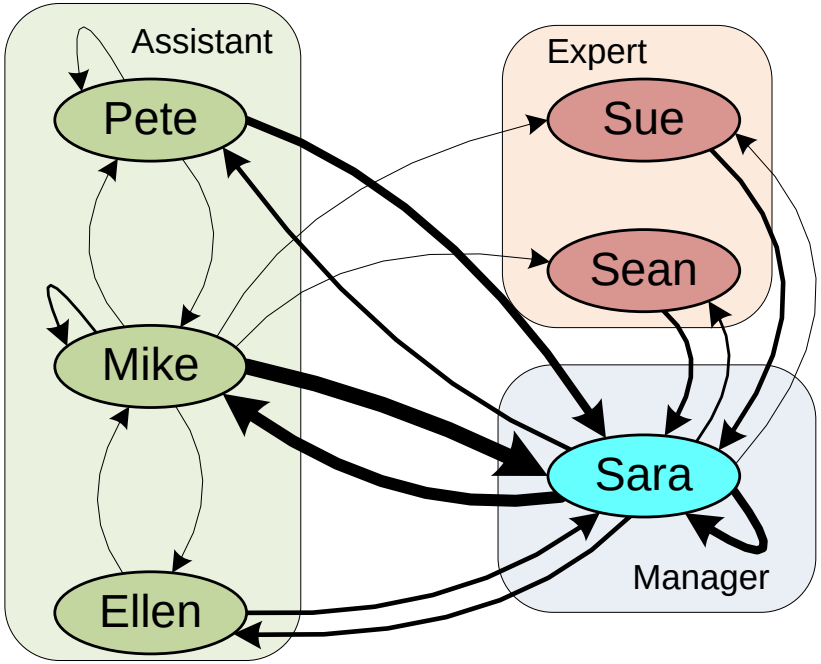


In this figure only the thickness of the arcs is based on frequencies. All nodes have the same size.

	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

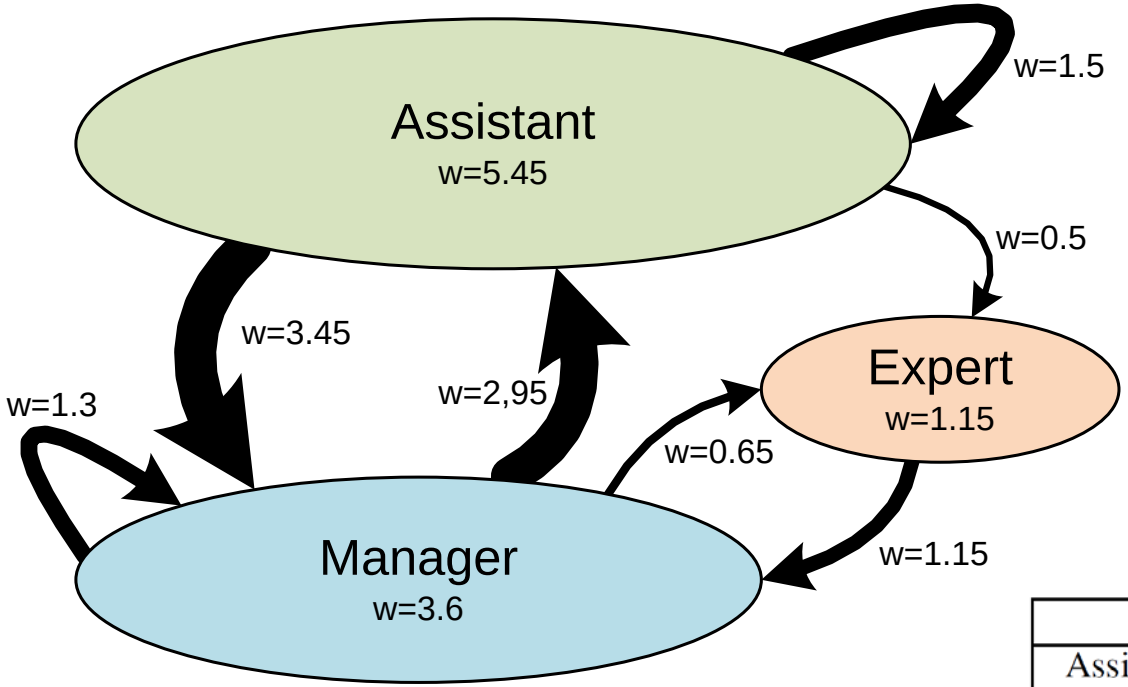
Sometimes we have explicit role or group information

- Information system can provide such information (like an address book or directory).
- It may also be recorded with the event itself.
- Let's assume three roles: Assistant, Expert, and Manager.



TU/e

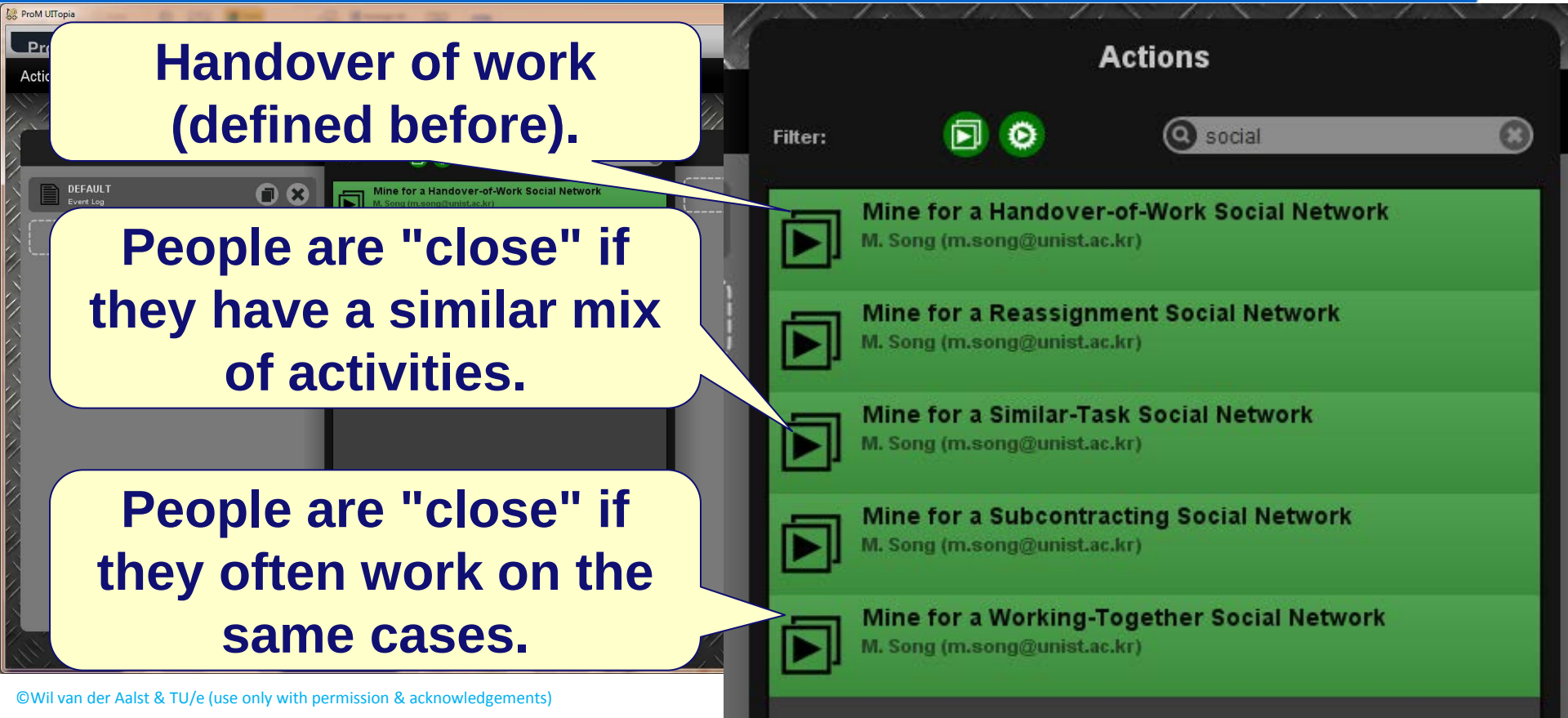
Handover of work at role level



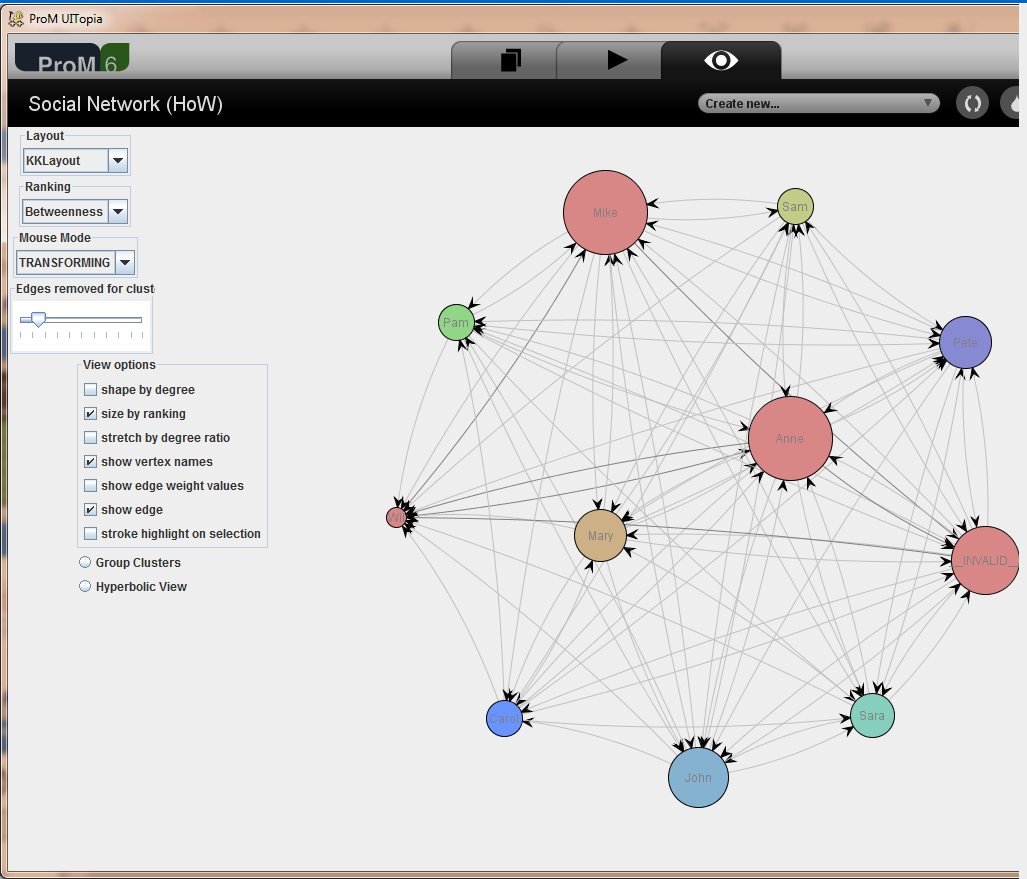
In this figure also the size of each node is based on frequencies.

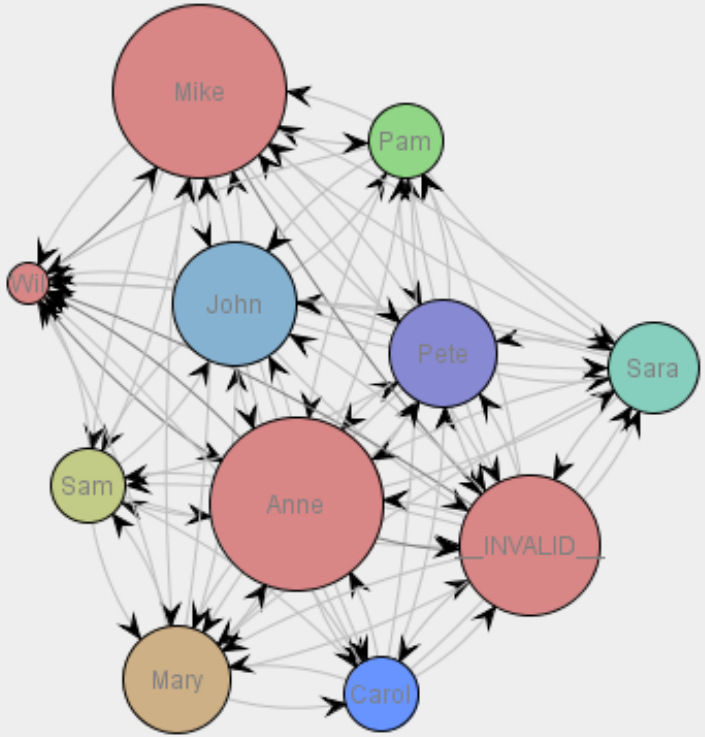
	Assistant	Expert	Manager
Assistant	1.5	0.5	3.45
Expert	0	0	1.15
Manager	2.95	0.65	1.3

Social network miner in ProM

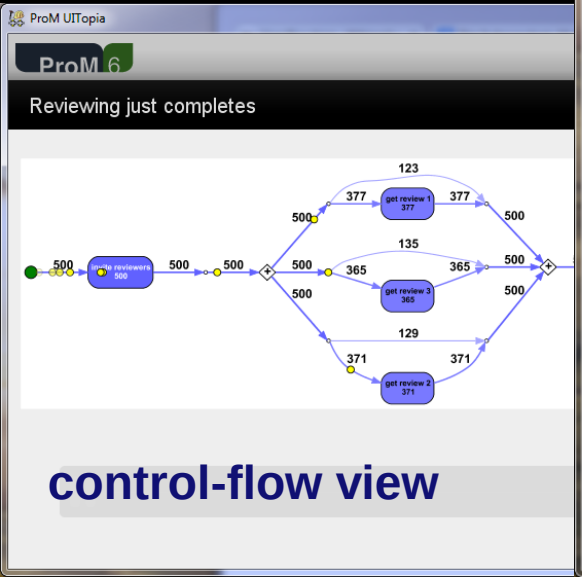


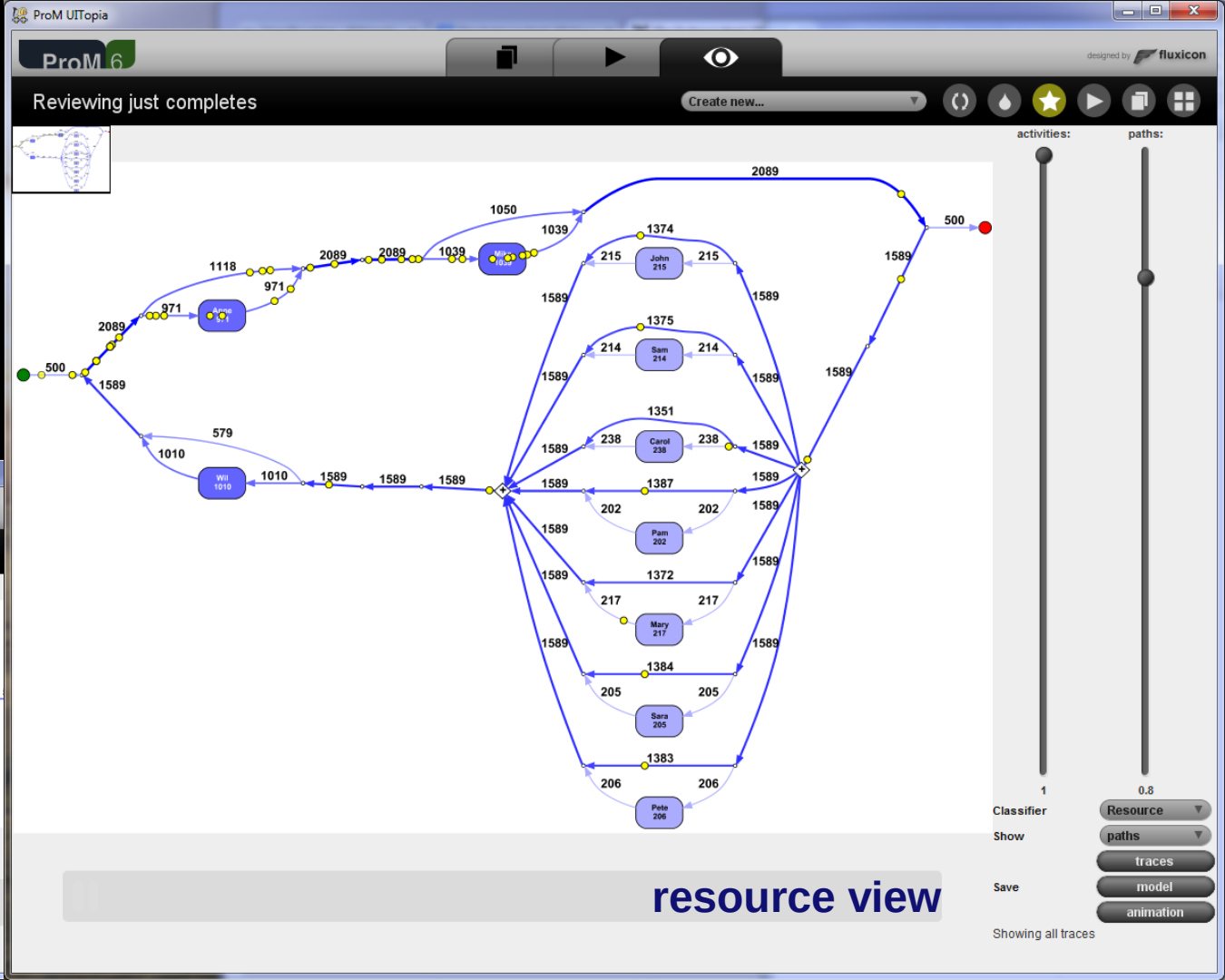
Social network based on hand-over of work





Changing the classifier from activity name to resource name.





- **Resource-activity matrix (Who is doing what?).**
- **Handover of work matrix (How is work passed on?)**
- **Used to create social networks (one of many possibilities).**
- **Social network can be analyzed in many ways.**

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

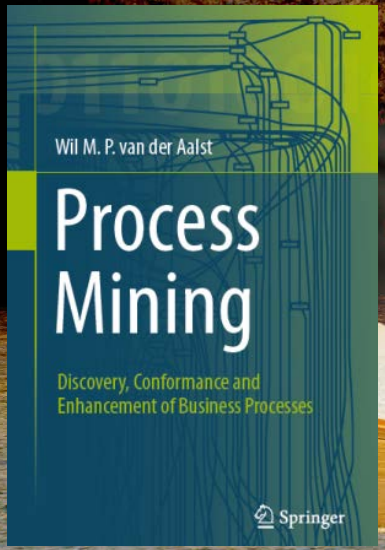
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

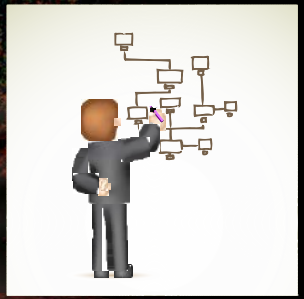
Part IV: Putting Process Mining to W...

- 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