An introduction to **Process Mining**

Diogo R. Ferreira

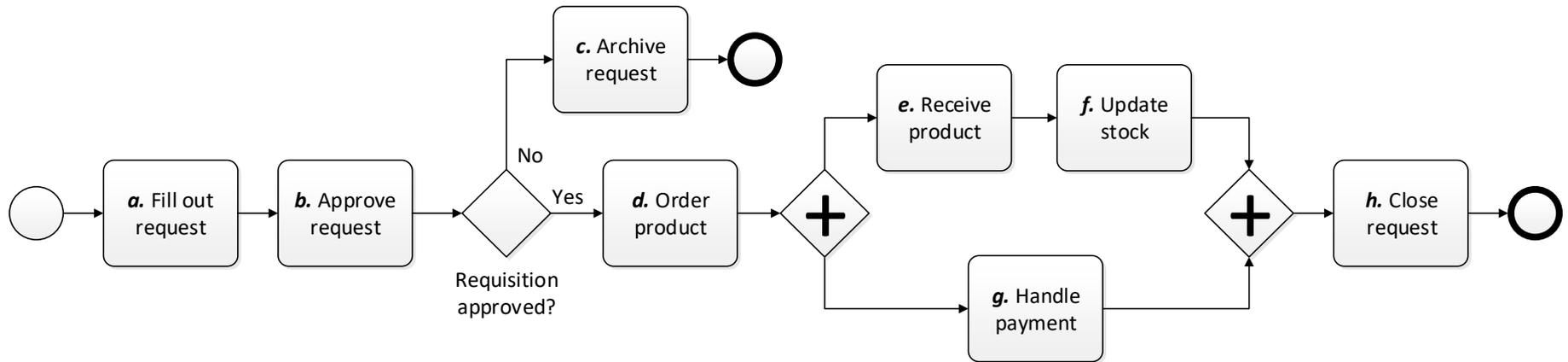diogo.ferreira@tecnico.ulisboa.pt

Introduction

- Example of a business process
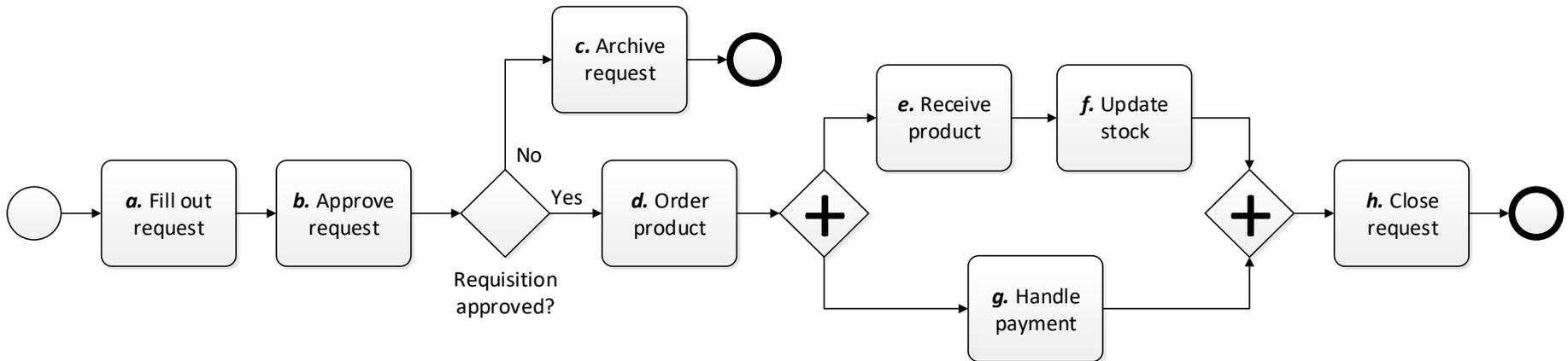
- sequence of activities
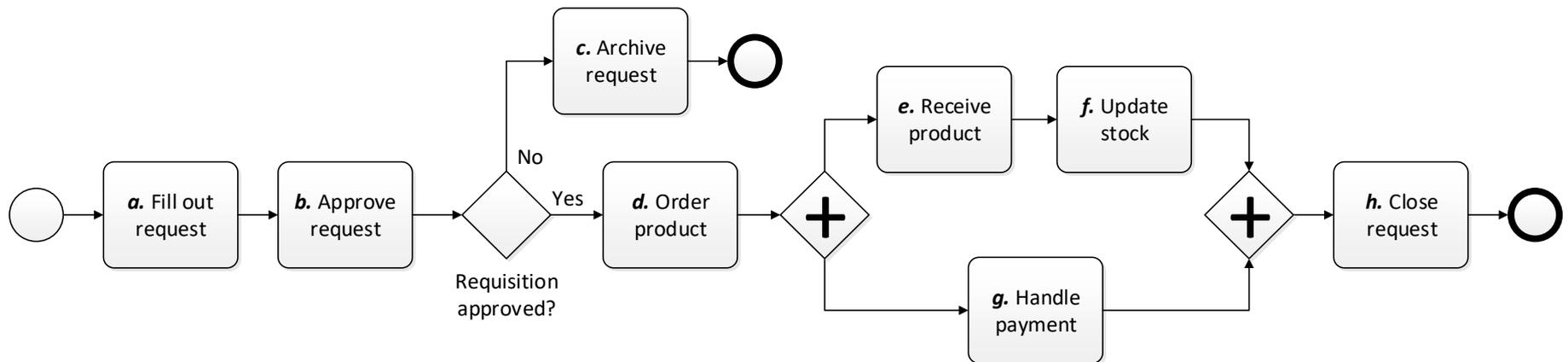- branching and parallel behavior
- process model vs. process instances

Introduction
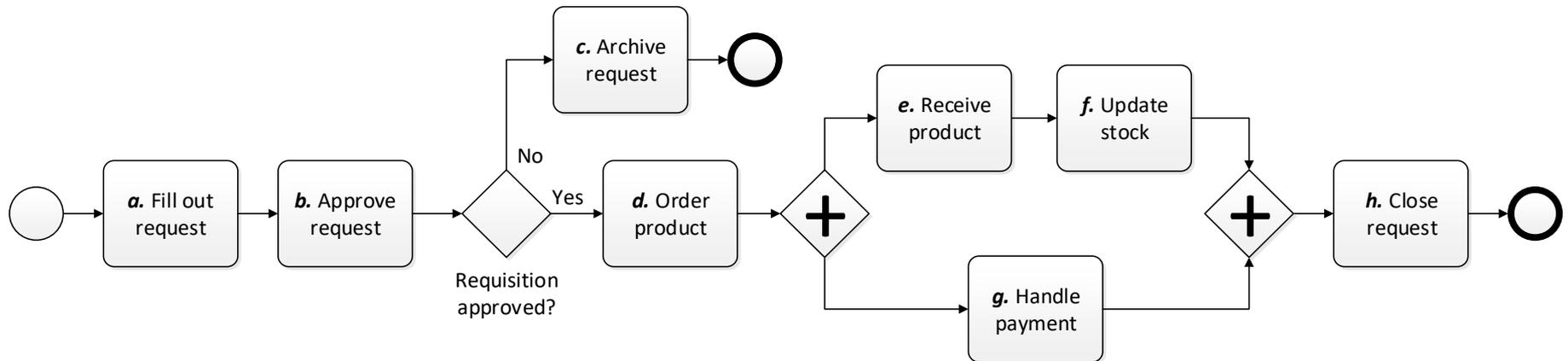
Introduction

a b c

a b d e f g h

a b d e g f h
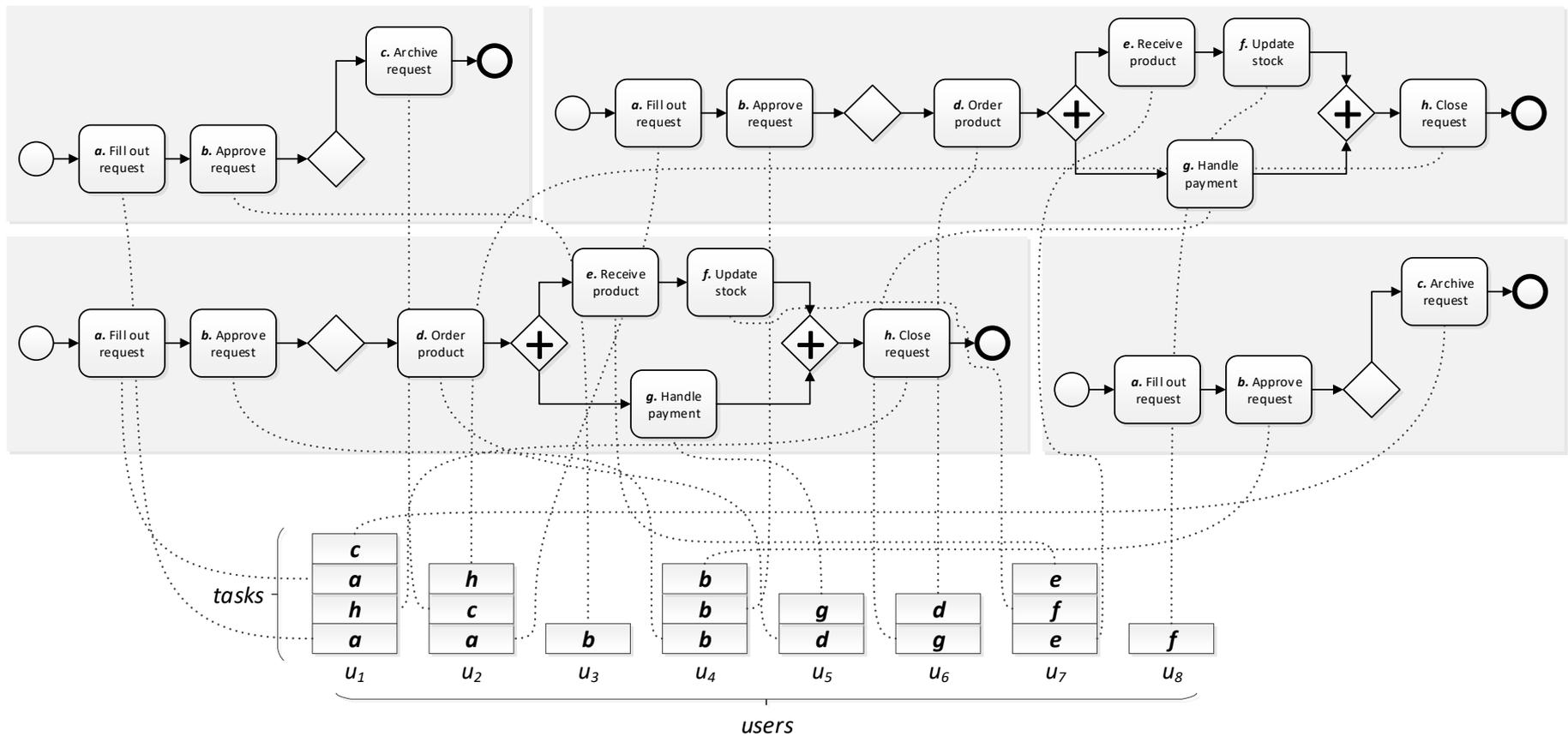
a b d g e f h

Introduction

- Possible questions
  - what is the most frequent path?
  - how many requests (%) get approved?
  - how much time does the process (or each activity) take?
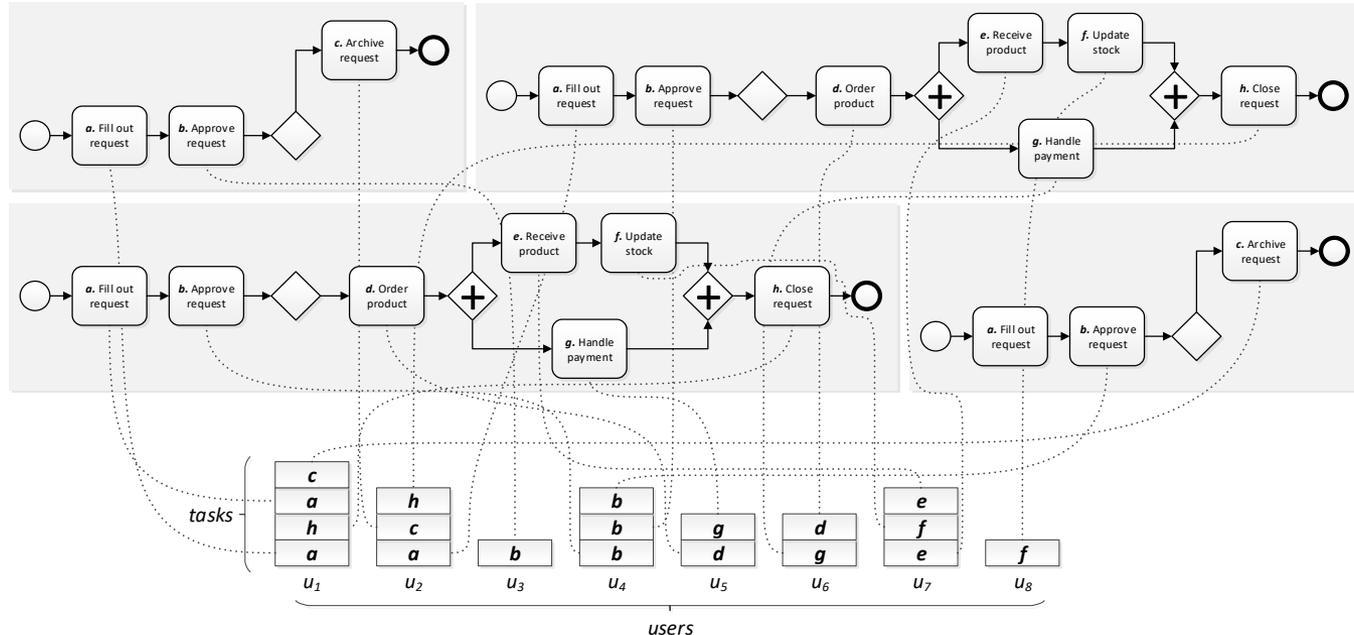 - does every instance comply with the model?

Introduction

- Task allocation

Introduction

- More possible questions
  - what is the distribution of work, i.e. who does what?
  - how do users interact and collaborate with each other?
  - what is the performance of each user in each activity?
  - are certain resources being overloaded?

Event logs

- Example of an event log

case id	task	user	timestamp
1	a	u_1	2016-04-09 17:36:47
1	b	u_3	2016-04-11 09:11:13
1	d	u_6	2016-04-12 10:00:12
1	e	u_7	2016-04-12 18:21:32
1	f	u_8	2016-04-13 13:27:41
2	a	u_2	2016-04-14 08:56:09
2	b	u_3	2016-04-14 09:36:02
2	d	u_5	2016-04-15 10:16:40
1	g	u_6	2016-04-18 19:14:14
2	g	u_6	2016-04-19 15:39:15
1	h	u_2	2016-04-19 16:48:16
2	e	u_7	2016-04-20 14:39:45
2	f	u_8	2016-04-22 09:16:16
3	a	u_2	2016-04-25 08:39:24
2	h	u_1	2016-04-26 12:19:46
3	b	u_4	2016-04-29 10:56:14
3	c	u_1	2016-04-30 15:41:22

Event logs

- Process mining perspectives
  - control-flow perspective (*task* column)
  - organizational perspective (*user* column)
  - performance perspective (*timestamp* column)

case id	task	user	timestamp
1	a	u_1	2016-04-09 17:36:47
1	b	u_3	2016-04-11 09:11:13
1	d	u_6	2016-04-12 10:00:12
1	e	u_7	2016-04-12 18:21:32
1	f	u_8	2016-04-13 13:27:41
2	a	u_2	2016-04-14 08:56:09
2	b	u_3	2016-04-14 09:36:02
2	d	u_5	2016-04-15 10:16:40
1	g	u_6	2016-04-18 19:14:14
2	g	u_6	2016-04-19 15:39:15
1	h	u_2	2016-04-19 16:48:16
2	e	u_7	2016-04-20 14:39:45
2	f	u_8	2016-04-22 09:16:16
3	a	u_2	2016-04-25 08:39:24
2	h	u_1	2016-04-26 12:19:46
3	b	u_4	2016-04-29 10:56:14
3	c	u_1	2016-04-30 15:41:22

Control-flow perspective

- Task transitions within each case id

case id	task	user	timestamp
1	a	u_1	2016-04-09 17:36:47
1	b	u_3	2016-04-11 09:11:13
1	d	u_6	2016-04-12 10:00:12
1	e	u_7	2016-04-12 18:21:32
1	f	u_8	2016-04-13 13:27:41
2	a	u_2	2016-04-14 08:56:09
2	b	u_3	2016-04-14 09:36:02
2	d	u_5	2016-04-15 10:16:40
1	g	u_6	2016-04-18 19:14:14
2	g	u_6	2016-04-19 15:39:15
1	h	u_2	2016-04-19 16:48:16
2	e	u_7	2016-04-20 14:39:45
2	f	u_8	2016-04-22 09:16:16
3	a	u_2	2016-04-25 08:39:24
2	h	u_1	2016-04-26 12:19:46
3	b	u_4	2016-04-29 10:56:14
3	c	u_1	2016-04-30 15:41:22

Control-flow perspective

- Transition matrix

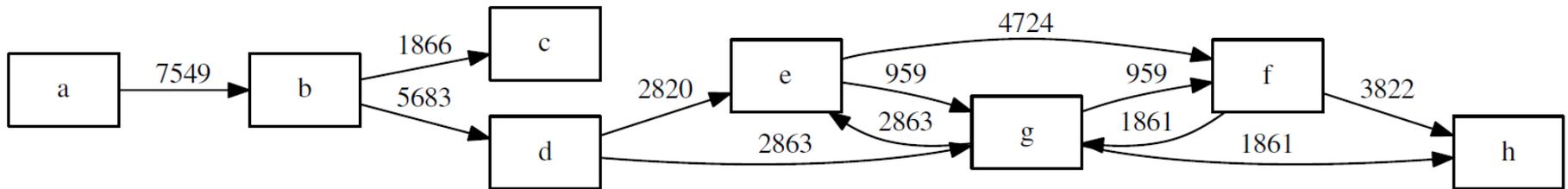|	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*
a		3						
b			1	2				
c								
d					1		1	
e						2		
f							1	1
g					1			1
h								

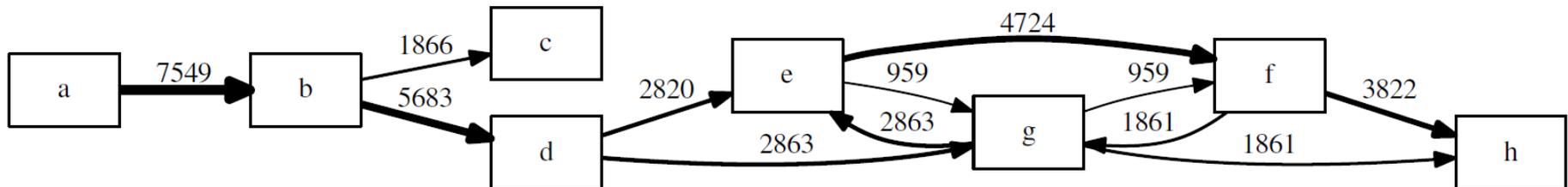# Control-flow perspective
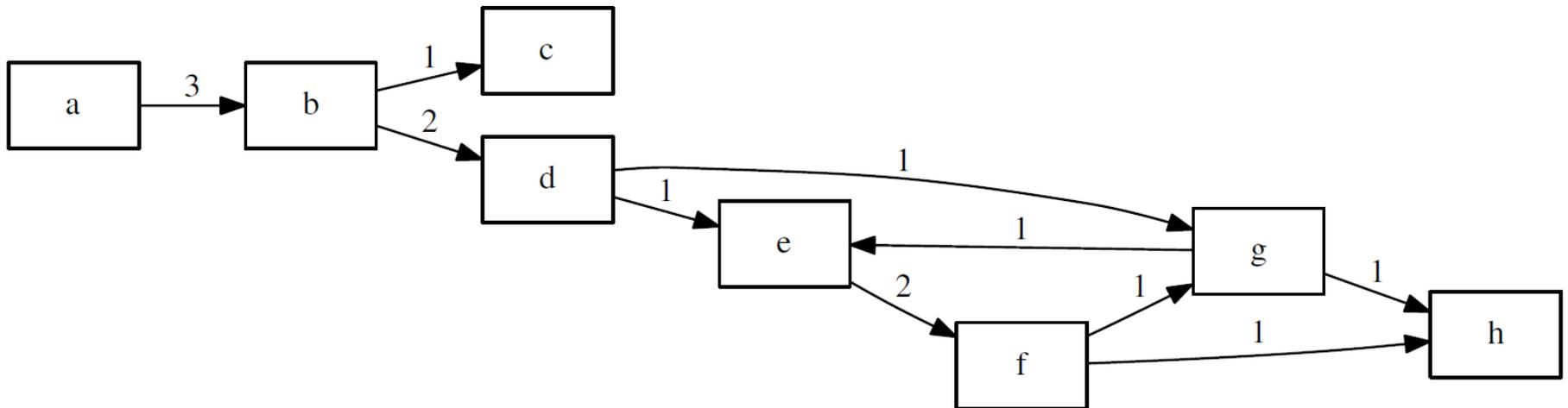
- Transition graph

Control-flow perspective
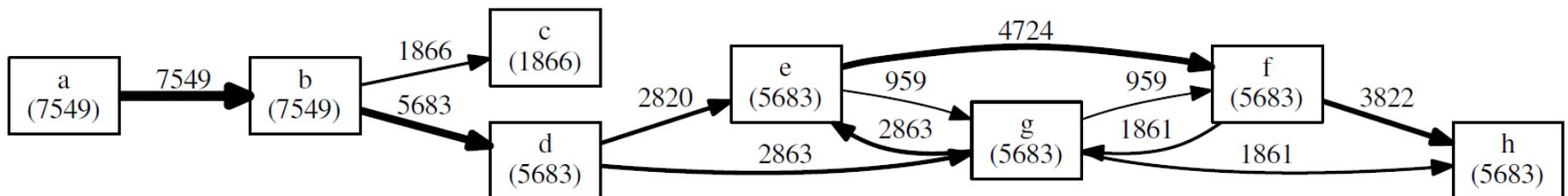
- Transition graph for a larger event log

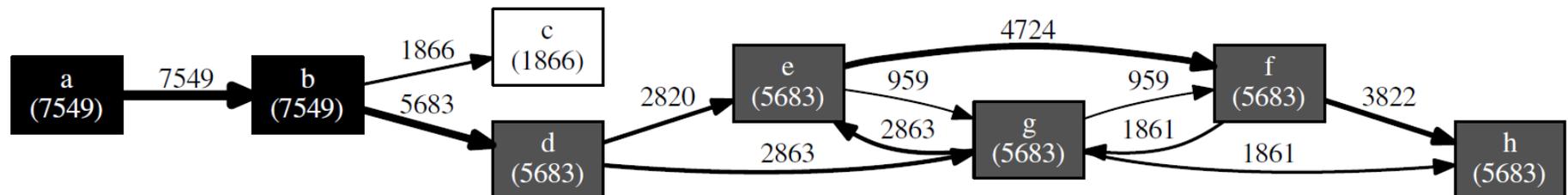- percentage of approvals: 5683 / (1866 + 5683) ≈ 75%

Control-flow perspective

- Improving the visualization
  - edge thickness

  - activity counts

  - node coloring

Organizational perspective

- Handover of work

case id	task	user	timestamp
1	a	u_1	2016-04-09 17:36:47
1	b	u_3	2016-04-11 09:11:13
1	d	u_6	2016-04-12 10:00:12
1	e	u_7	2016-04-12 18:21:32
1	f	u_8	2016-04-13 13:27:41
2	a	u_2	2016-04-14 08:56:09
2	b	u_3	2016-04-14 09:36:02
2	d	u_5	2016-04-15 10:16:40
1	g	u_6	2016-04-18 19:14:14
2	g	u_6	2016-04-19 15:39:15
1	h	u_2	2016-04-19 16:48:16
2	e	u_7	2016-04-20 14:39:45
2	f	u_8	2016-04-22 09:16:16
3	a	u_2	2016-04-25 08:39:24
2	h	u_1	2016-04-26 12:19:46
3	b	u_4	2016-04-29 10:56:14
3	c	u_1	2016-04-30 15:41:22

Organizational perspective

- Transition matrix

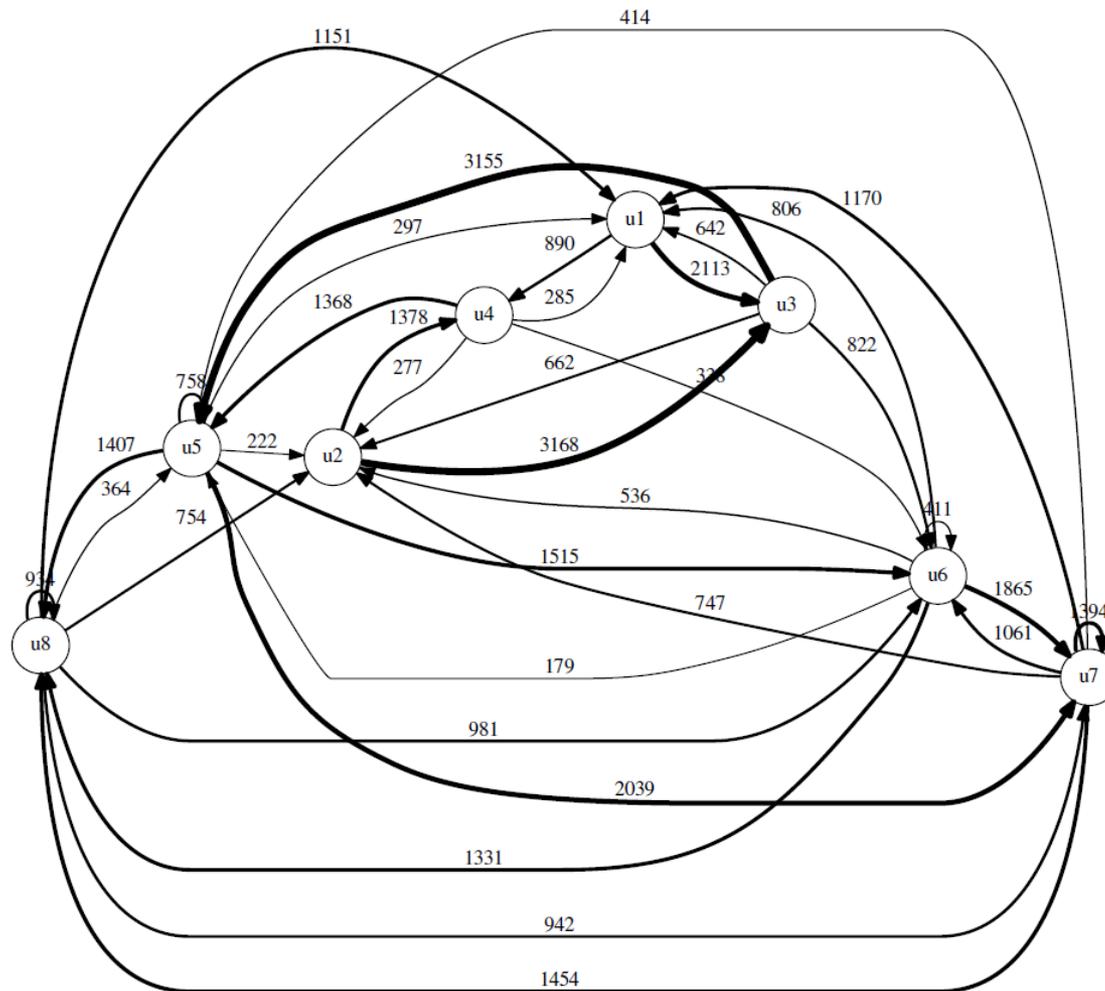|	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
u_1			1					
u_2			1	1				
u_3					1	1		
u_4	1							
u_5						1		
u_6		1					2	
u_7								2
u_8	1					1		

Organizational perspective
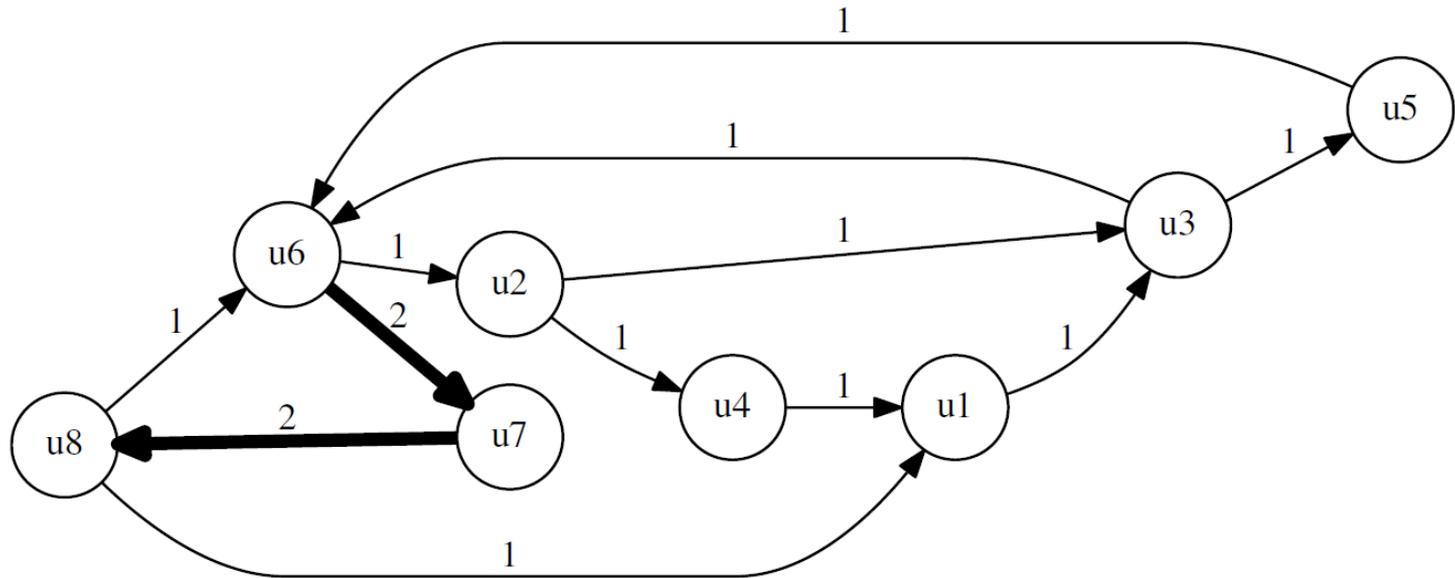
- Transition graph

Organizational perspective

- Transition graph for a larger event log

Organizational perspective

- Working together

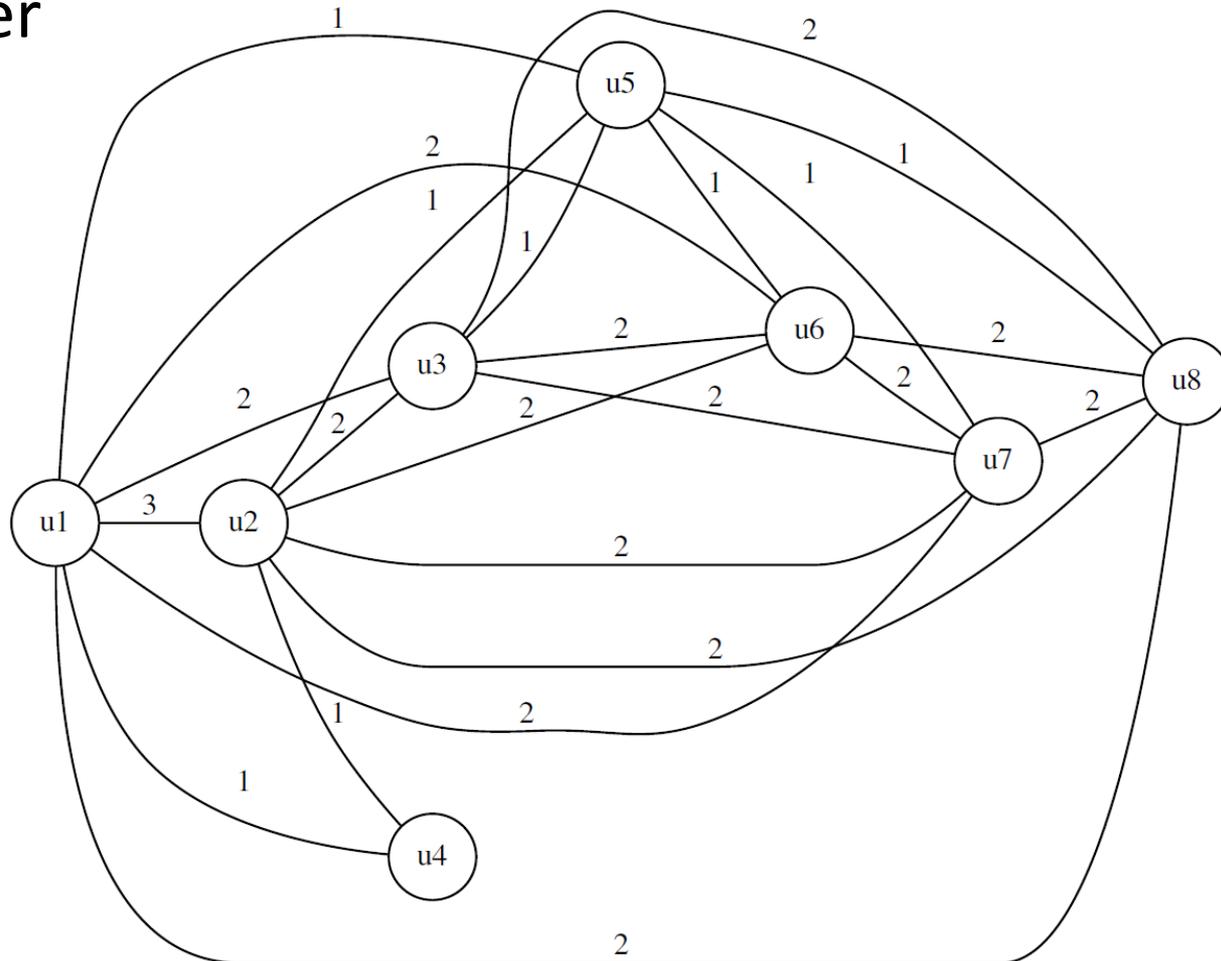| case id	task	user	timestamp
1	a	u_1	2016-04-09 17:36:47
1	b	u_3	2016-04-11 09:11:13
1	d	u_6	2016-04-12 10:00:12
1	e	u_7	2016-04-12 18:21:32
1	f	u_8	2016-04-13 13:27:41
2	a	u_2	2016-04-14 08:56:09
2	b	u_3	2016-04-14 09:36:02
2	d	u_5	2016-04-15 10:16:40
1	g	u_6	2016-04-18 19:14:14
2	g	u_6	2016-04-19 15:39:15
1	h	u_2	2016-04-19 16:48:16
2	e	u_7	2016-04-20 14:39:45
2	f	u_8	2016-04-22 09:16:16
3	a	u_2	2016-04-25 08:39:24
2	h	u_1	2016-04-26 12:19:46
3	b	u_4	2016-04-29 10:56:14
3	c	u_1	2016-04-30 15:41:22

Organizational perspective

- Number of cases where each pair of users have worked together

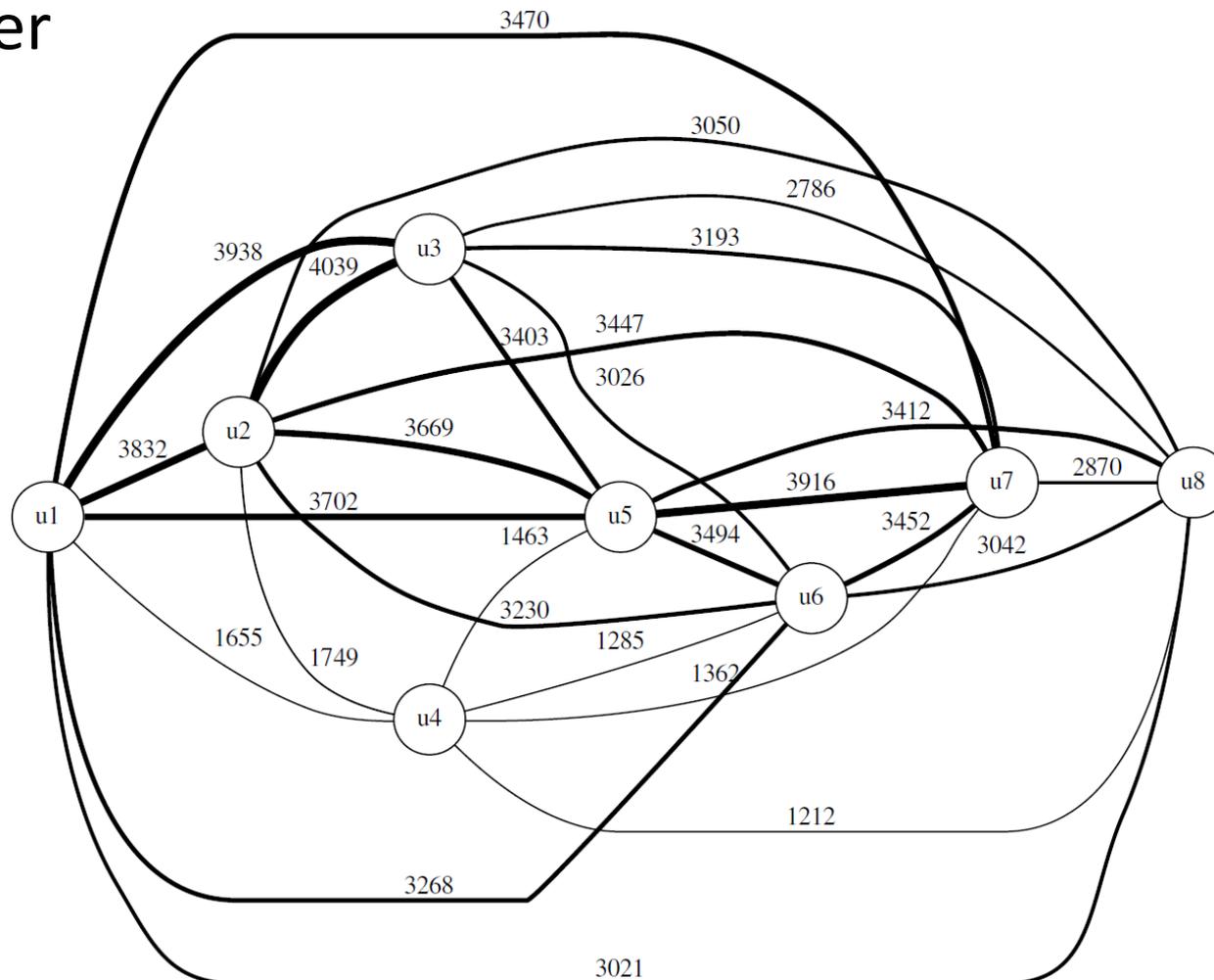|	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8
u_1		3	2	1	1	2	2	2
u_2	3		2	1	1	2	2	2
u_3	2	2			1	2	2	2
u_4	1	1						
u_5	1	1	1			1	1	1
u_6	2	2	2		1		2	2
u_7	2	2	2		1	2		2
u_8	2	2	2		1	2	2	

Organizational perspective

- Number of cases where each pair of users have worked together

- Number of cases where each pair of users have worked together

Organizational perspective
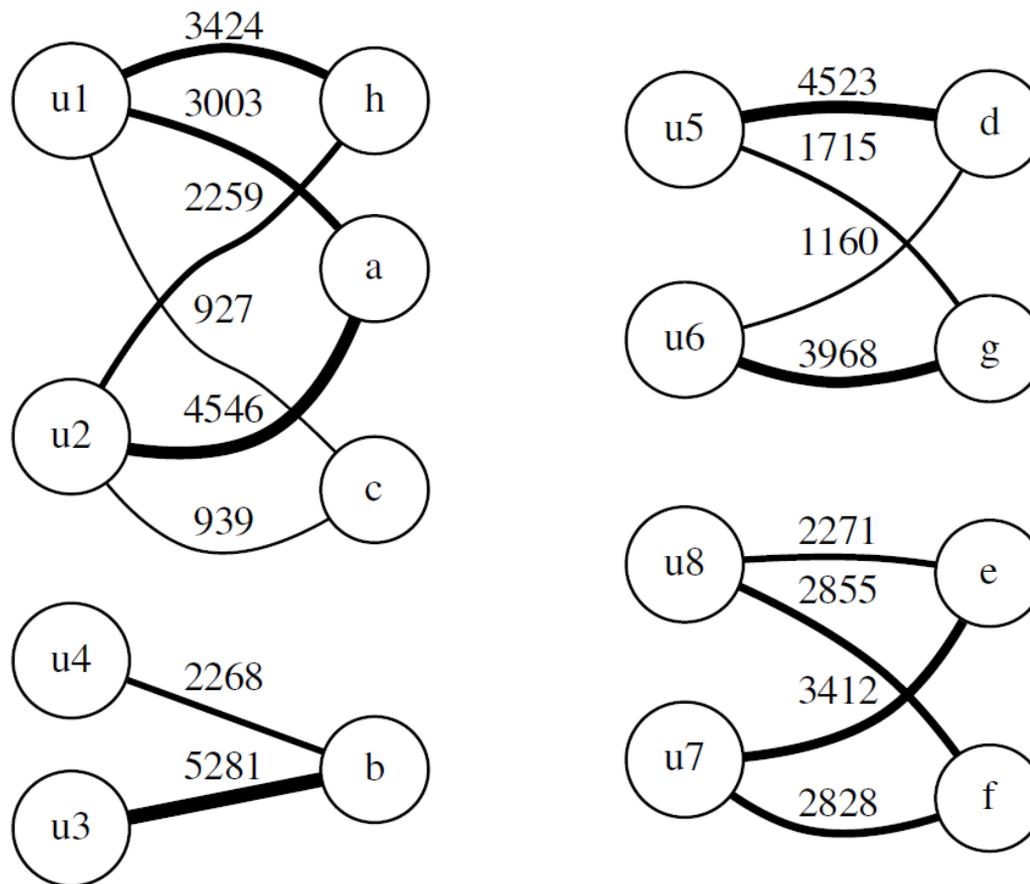
- Distribution of work

case id	task	user	timestamp
1	a	u_1	2016-04-09 17:36:47
1	b	u_3	2016-04-11 09:11:13
1	d	u_6	2016-04-12 10:00:12
1	e	u_7	2016-04-12 18:21:32
1	f	u_8	2016-04-13 13:27:41
2	a	u_2	2016-04-14 08:56:09
2	b	u_3	2016-04-14 09:36:02
2	d	u_5	2016-04-15 10:16:40
1	g	u_6	2016-04-18 19:14:14
2	g	u_6	2016-04-19 15:39:15
1	h	u_2	2016-04-19 16:48:16
2	e	u_7	2016-04-20 14:39:45
2	f	u_8	2016-04-22 09:16:16
3	a	u_2	2016-04-25 08:39:24
2	h	u_1	2016-04-26 12:19:46
3	b	u_4	2016-04-29 10:56:14
3	c	u_1	2016-04-30 15:41:22

Organizational perspective

- Distribution of work

Performance perspective

- Timestamp difference between events

| case id	task	timestamp
1	a	2016-04-09 17:36:47
1	b	2016-04-11 09:11:13
1	d	2016-04-12 10:00:12
1	e	2016-04-12 18:21:32
1	f	2016-04-13 13:27:41
2	a	2016-04-14 08:56:09
2	b	2016-04-14 09:36:02
2	d	2016-04-15 10:16:40
1	g	2016-04-18 19:14:14
2	g	2016-04-19 15:39:15
1	h	2016-04-19 16:48:16
2	e	2016-04-20 14:39:45
2	f	2016-04-22 09:16:16
3	a	2016-04-25 08:39:24
2	h	2016-04-26 12:19:46
3	b	2016-04-29 10:56:14
3	c	2016-04-30 15:41:22

Performance perspective
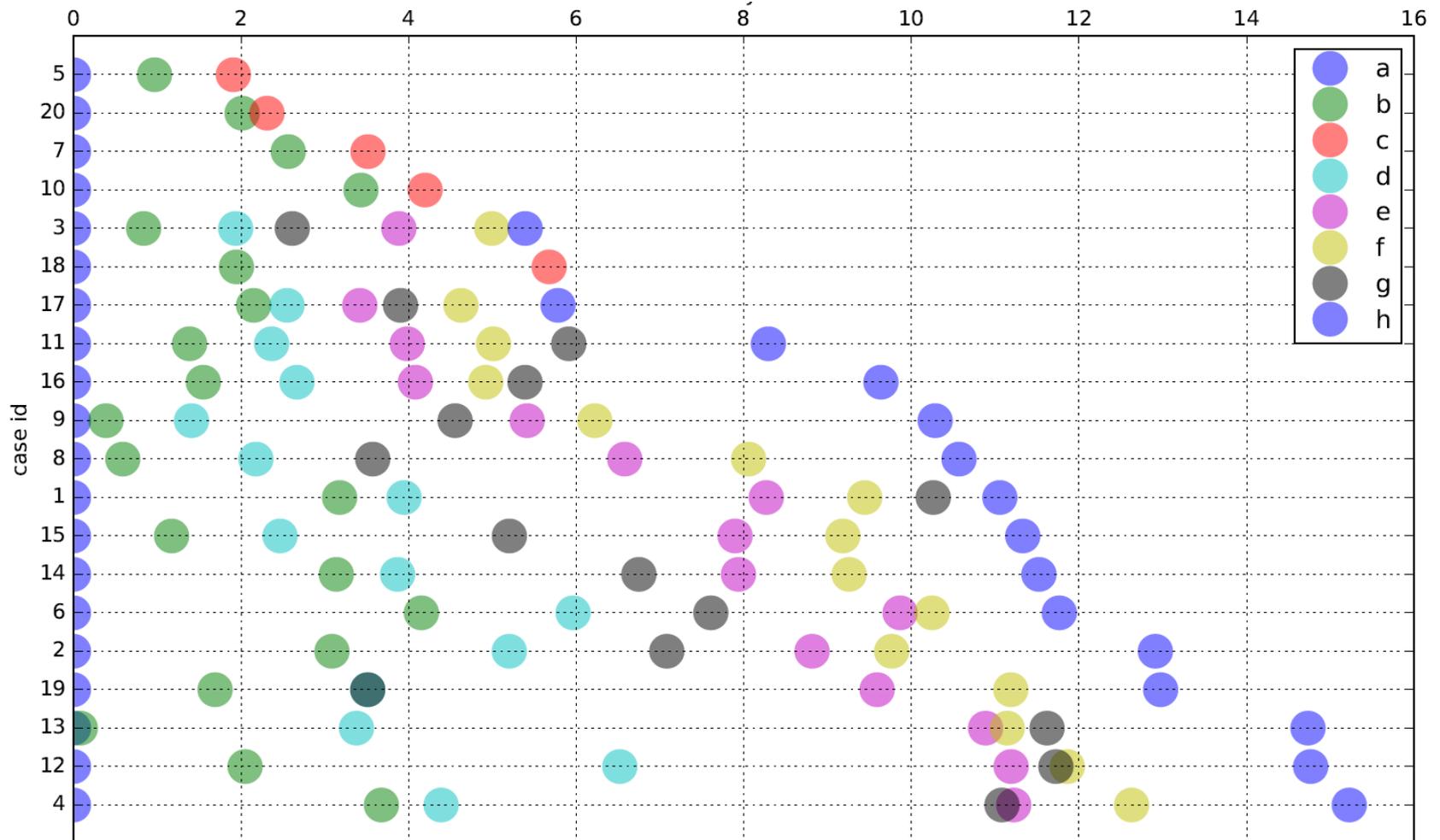
- Timestamp difference between events
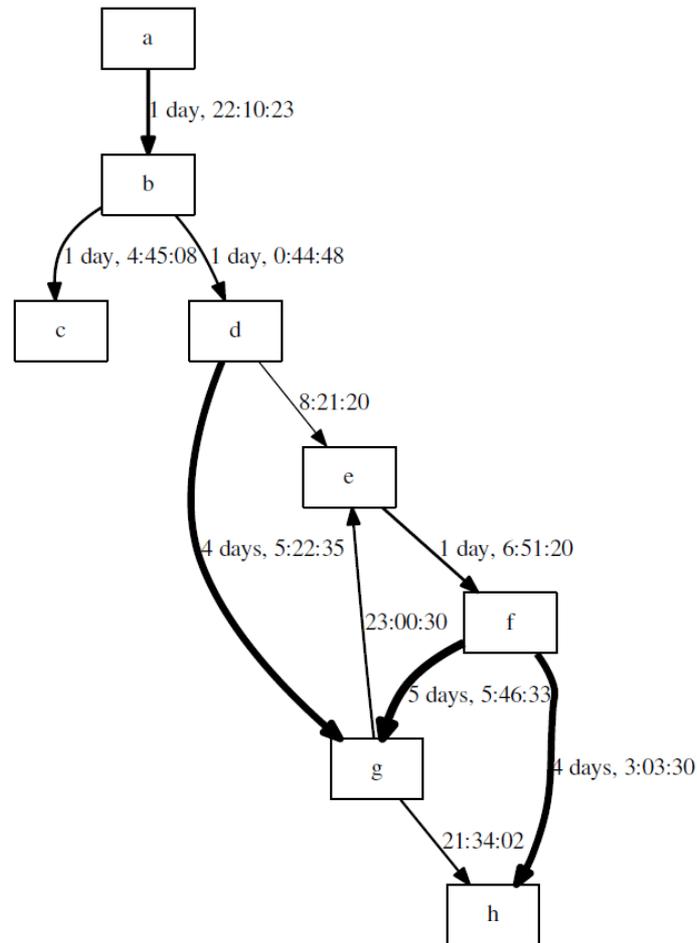
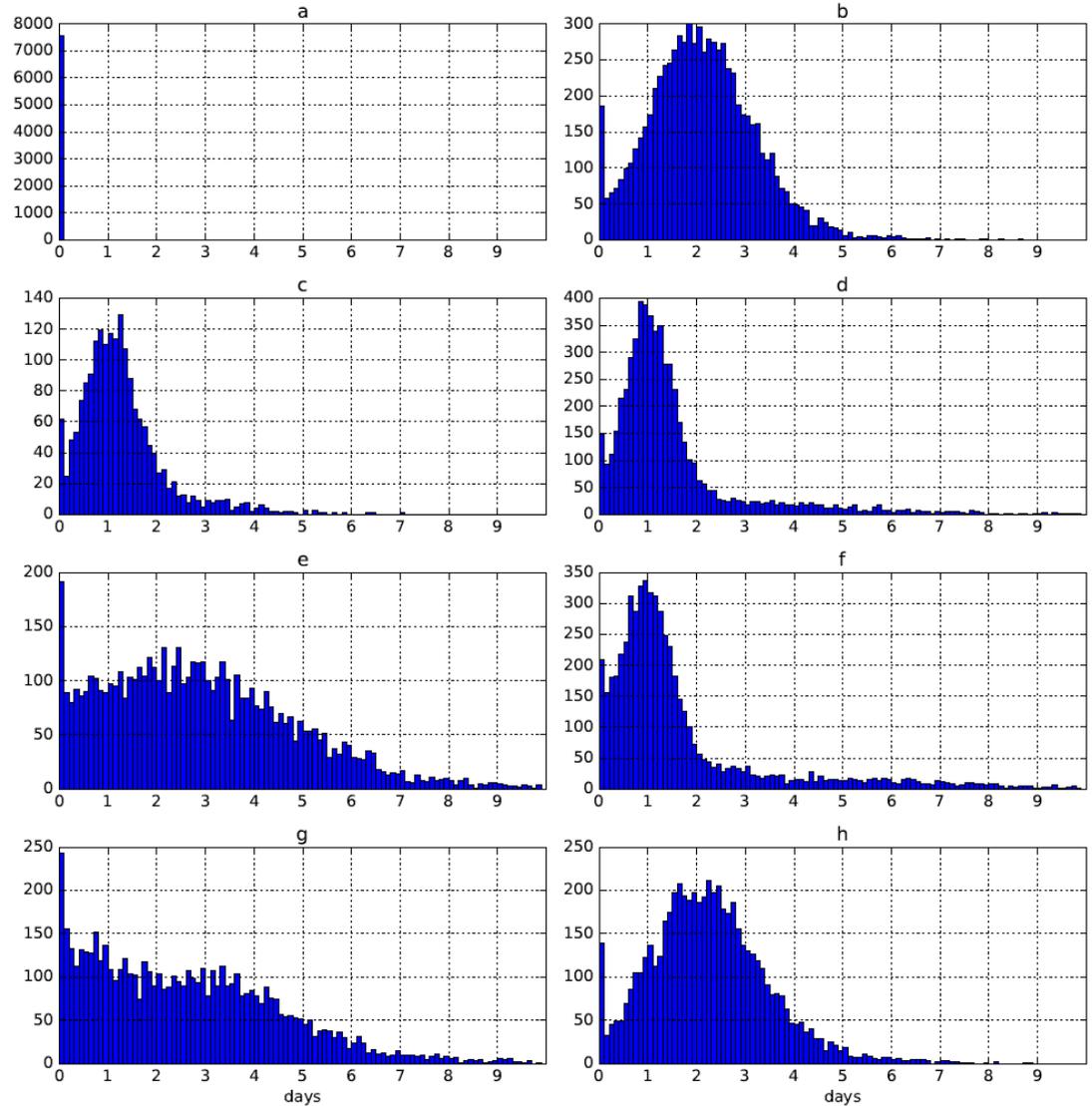# Performance perspective

- Timeline of events
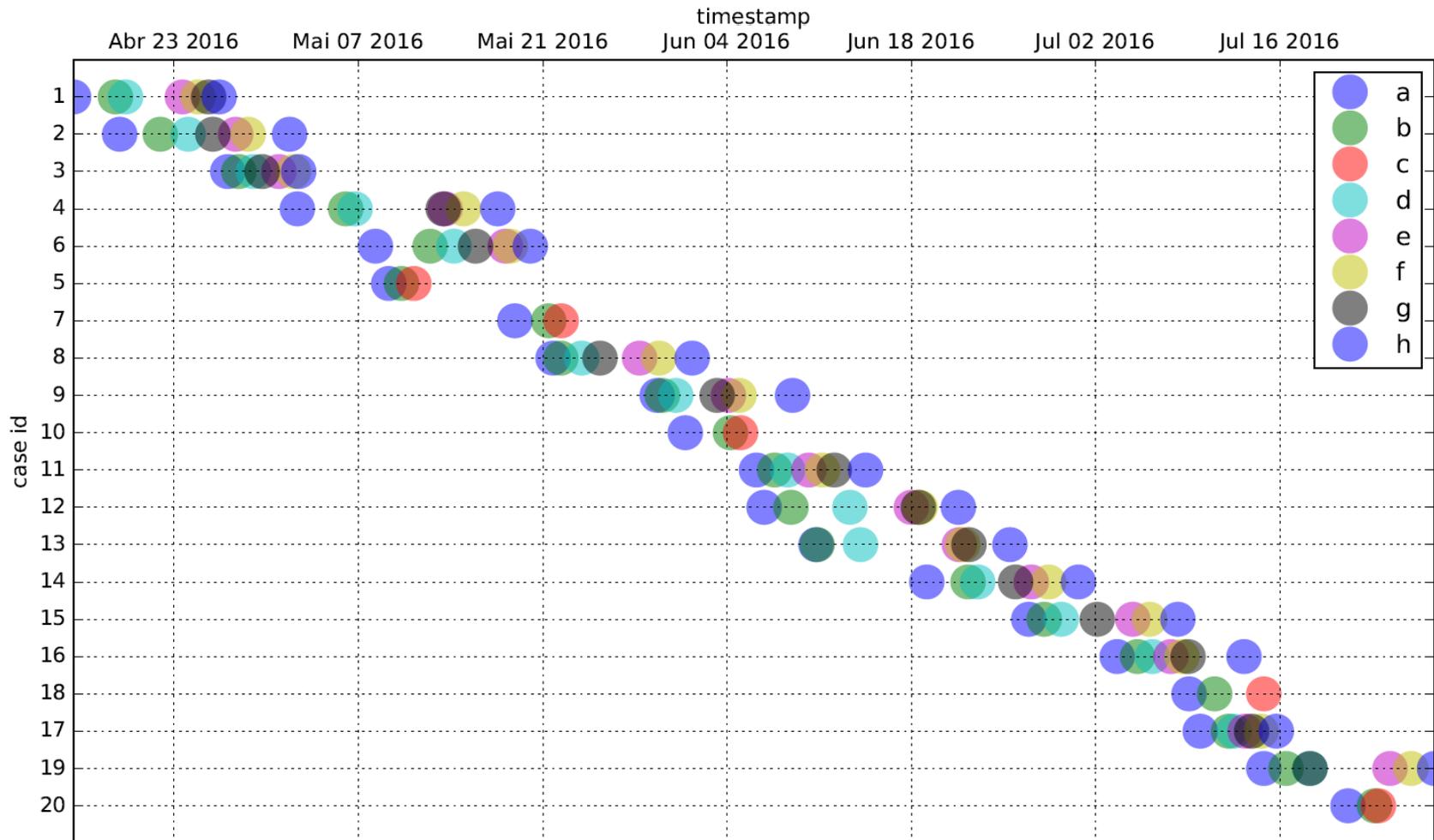
Performance perspective

- Relative time

Performance perspective

- Activity duration

Conclusion

- Where to go from here
  - process mining books
    - *A Primer on Process Mining: Practical Skills with Python and Graphviz*, D. R. Ferreira (Springer, 2017)
    - *Process Mining: Data Science in Action*, W. van der Aalst (Springer, 2016)
  - process mining courses
    - *Process Mining: Data Science in Action* https://www.coursera.org/learn/process-mining
    - *Introduction to Process Mining with ProM* https://www.futurelearn.com/courses/process-mining
  - process mining website
    - http://www.processmining.org
  - process mining tools: ProM, Disco, etc.