Практическое применение методов извлечения и анализа моделей процессов: принципы и примеры

Александр Давидович Брейман,
научный сотрудник МЛПОИС,
к.т.н., доцент

Process mining framework

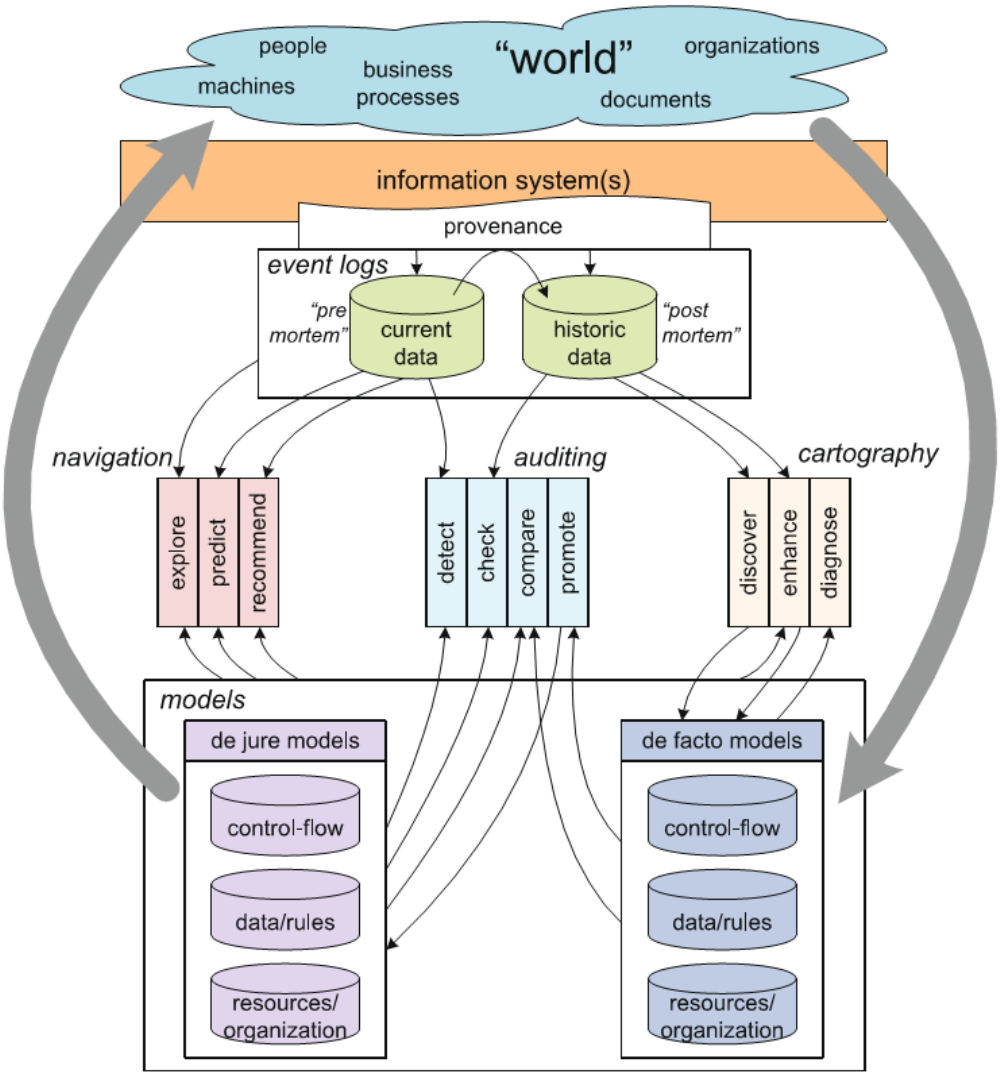


fig. 11.5 from W.M.P. van der Aalst, Process Mining , Springer, 2011.

L* life-cycle model

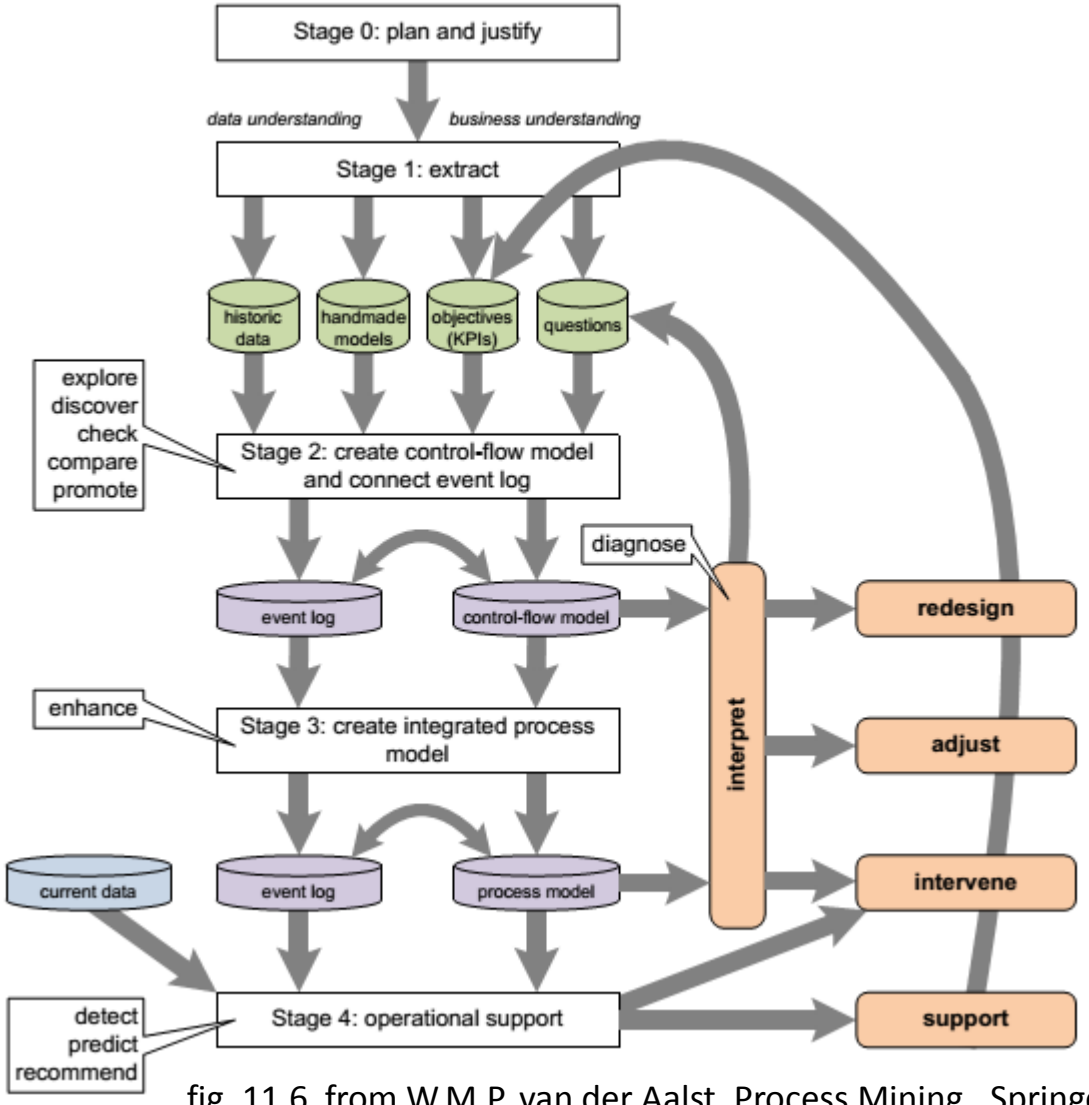


fig. 11.6 from W.M.P. van der Aalst, Process Mining , Springer, 2011.

Types of process mining projects

- Data-driven (curiosity driven)
  - Powered by availability of event data
- Question-driven
  - “Why do cases handled by X takes longer than cases handled by Y?”
- Goal-driven
  - To improve a process with respect to KPIs

Functional areas

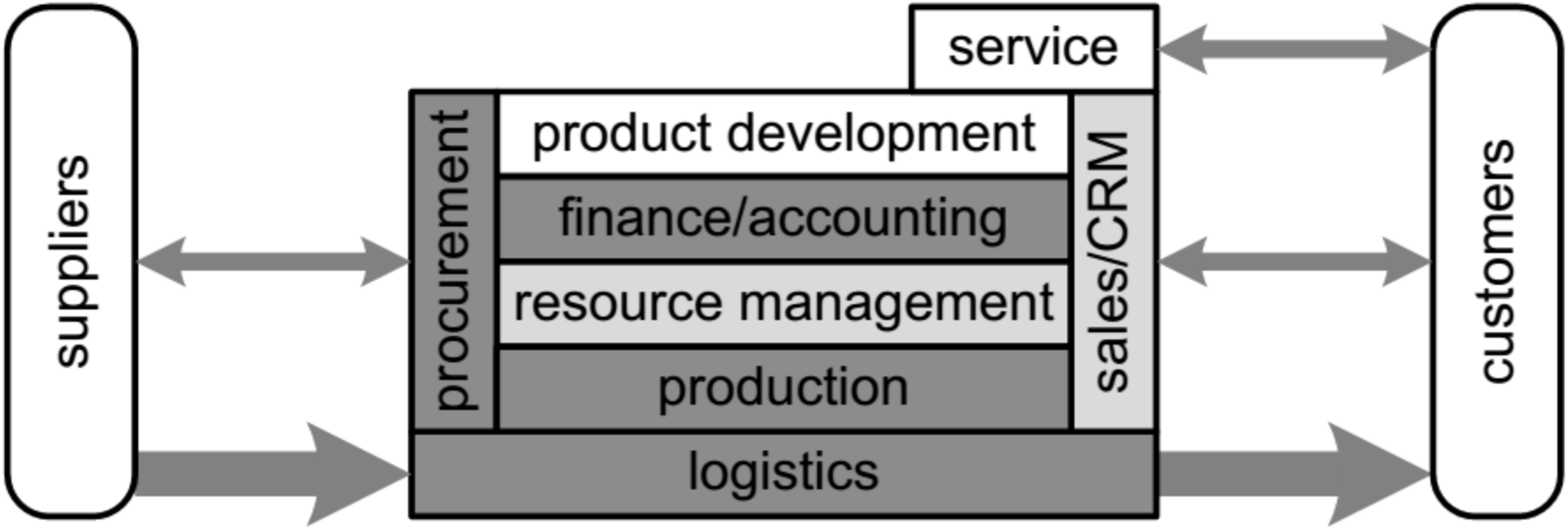
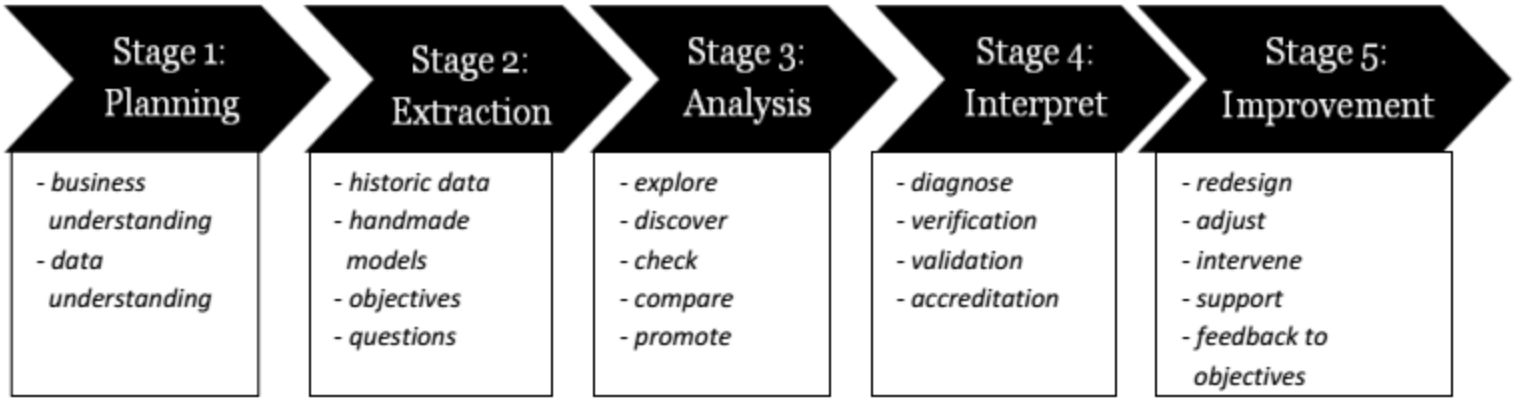


fig. 11.7 from W.M.P. van der Aalst, Process Mining , Springer, 2011.

Case 1: Simple/complex and fast/slow insurance claims

- Understanding process behaviors in a large insurance company in Australia : a case study.
- Authors: Suriadi Suriadi, Moe T. Wynn, Chun Ouyang, Arthur H.M. ter Hofstede, Nienke van Dijk.
- http://eprints.qut.edu.au/55502/

Case 1: Simplified L* life-cycle



Case 1: Stakeholder's main question

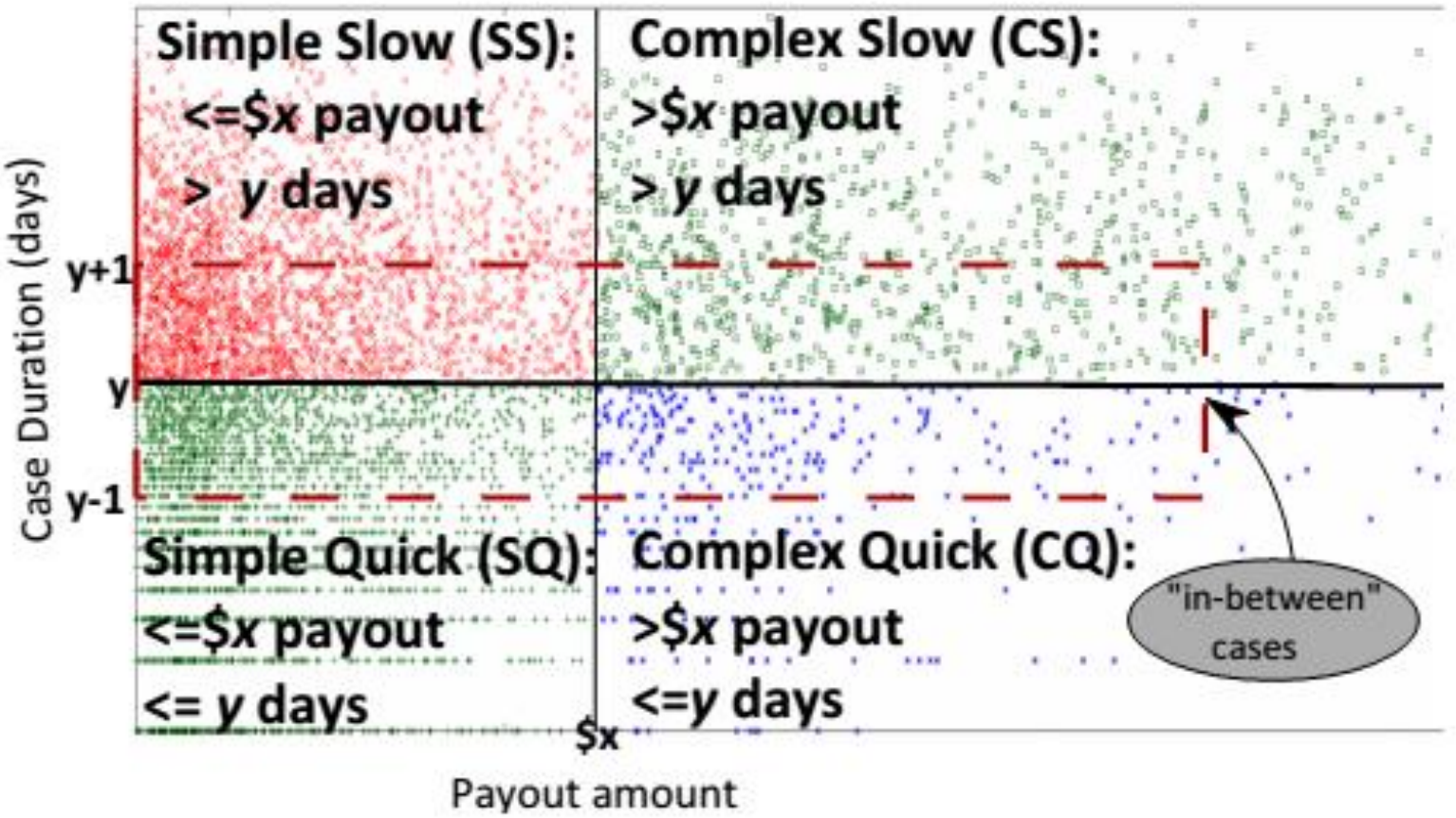
- Why did the processing of certain 'simple' claims takes an unexpectedly long time to complete?

- 'Simple' claim is a claim whose net payout value is less then $x dollars and should be completed no later than y-number of days.

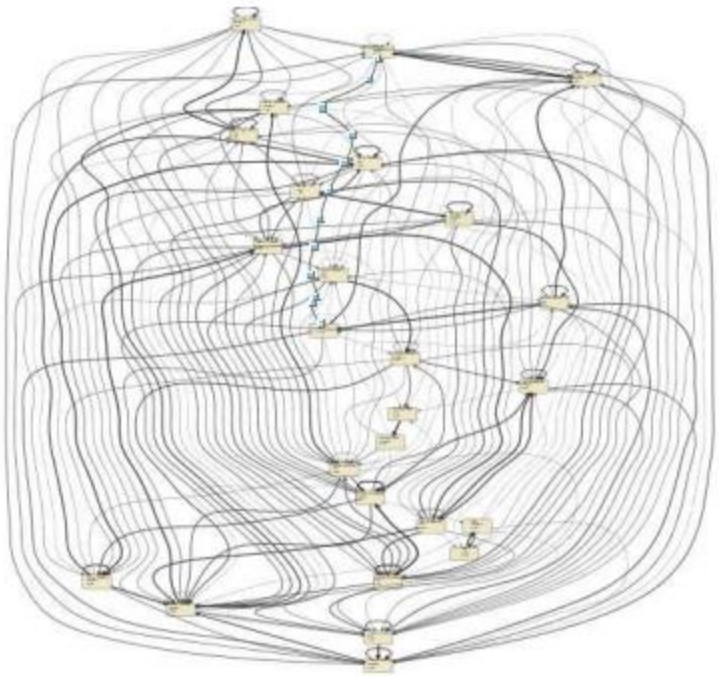
Case 1: Process mining questions

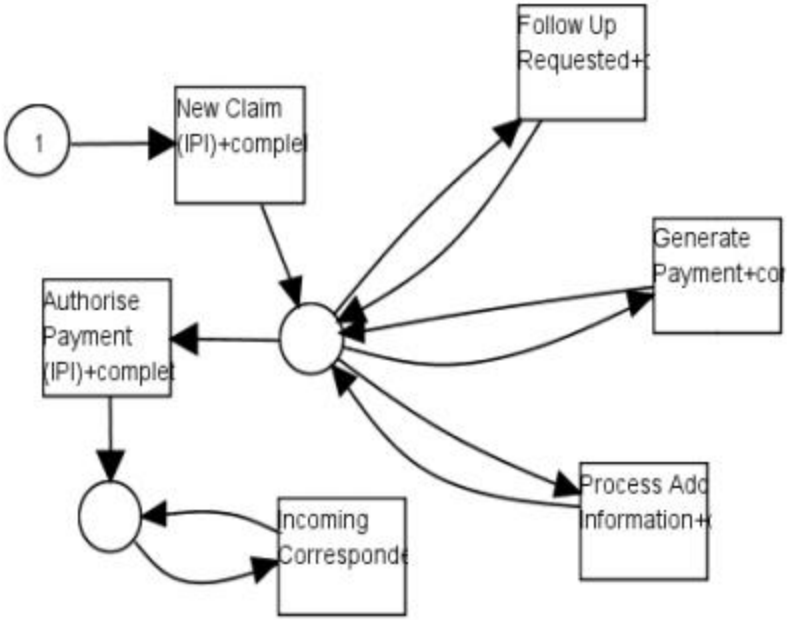
- Q1: What is the performance distribution of simple and non-simple (=complex) claims
- Q2: What do the process models look like for simple and complex claims? What are frequently-taken paths in the processing of these claims?
- Q3: What are the key differences in the way claims were processed between those strictly simple clams (completed on-time) and those supposedly simple claims (completed longer that y-number of days)?

Case 1: Log split by payout and duration

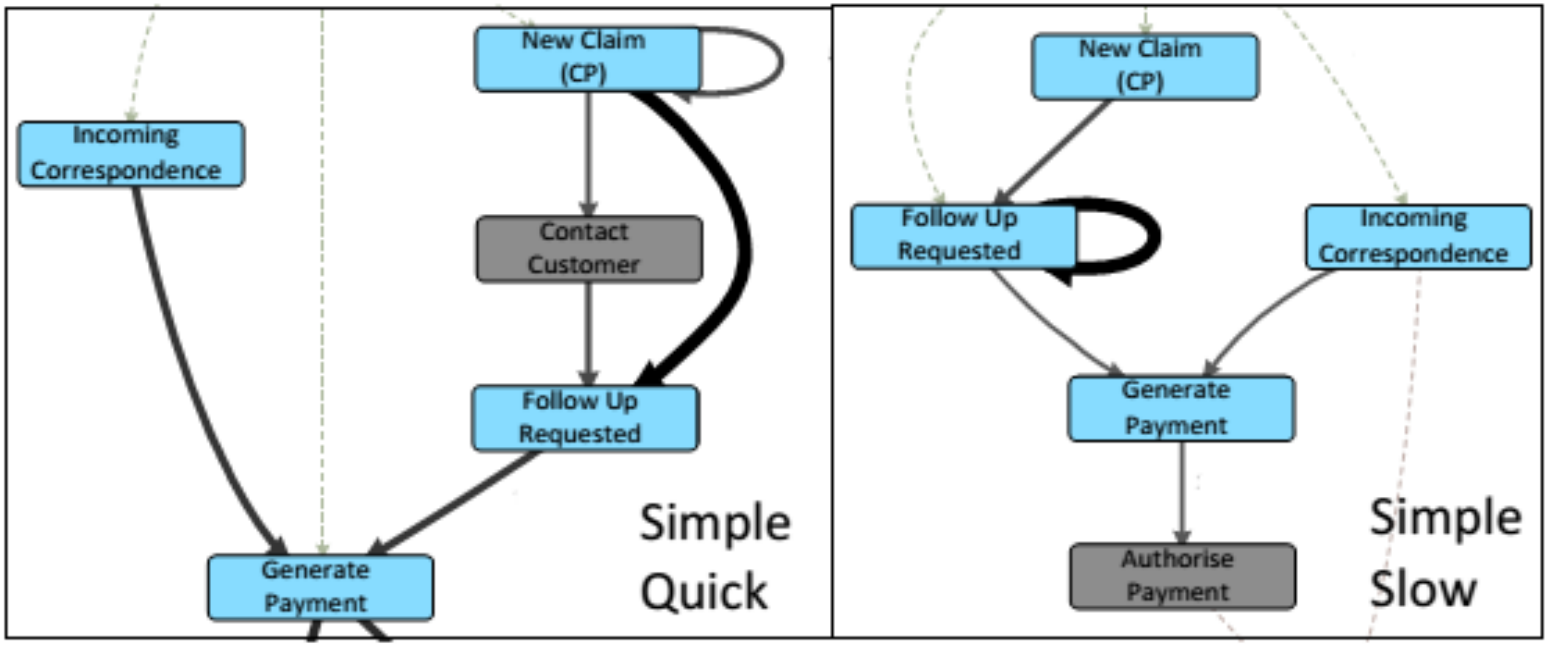


Case 1: Fuzzy Miner Spaghetti-like and ILP Miner Flower-like Models





Case 1: Disco-discovered Dominant Paths for Simple Quick and Simple Slow



Case 1: Frequencies difference of Simple Quick and Simple Slow

Activity	Simple Quick		Simple Slow	
	actFreq	actDist	actFreq	actDist
Follow Up Requested	1.86	74.4%	5.79	92.3%
Incoming Correspondence	1.75	81.6%	4.27	90.1%
Contact Customer	0.66	46.8%	1.29	63.3%
Contact Assessor	0.11	4.9%	1.36	21.5%
Conduct File Review	2.03	89.8%	6.11	96.9%

actFreq – average per-case execution frequency of an activity-X
(total occurrence of activity-X/number of cases)
actDist – distribution of an activity-X over all cases
(total number of cases with activity-X/number of cases)

Case 1: Improvements

- Findings from Q1 were useful to the stakeholders as they were able to quickly understand the performance distributions of claims
- High number of SS cases was surprising to stakeholders
- Insights from Q2 not only validated their suspicions w.r.t. non-standard manner in which claims were processed, but also highlighted need for process standardization
- Insights from Q3 have precisely identified areas for improvement

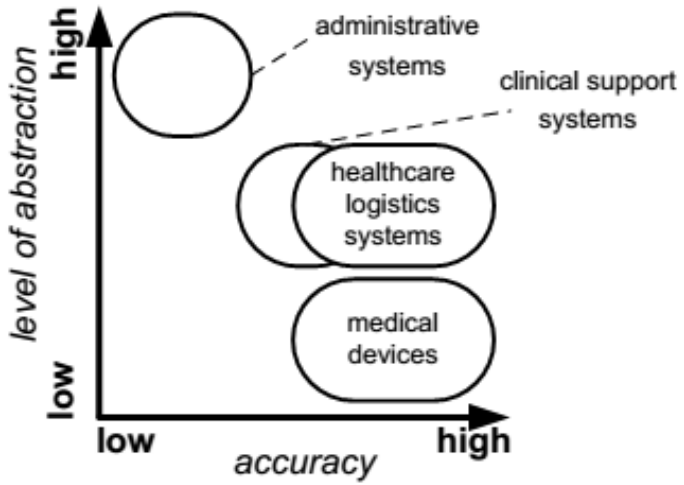
Case 2: Hospital Information Systems

- Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions
- Authors: R.S. Mans, W.M.P. van der Aalst, R.J.B. Vanwersch, A.J. Moleman. (TU/e, Academic Medical Center, University of Amsterdam)
- http://www.tue.nl/en/publication/ep/p/d/ep-uid/280589/

Case 2: Hospital Information Systems Frequently Posed Questions

- Q1: What are the most followed paths and what exceptional paths are followed?
- Q2: Are there differences in care paths followed by different patient groups?
- Q3: Do we comply with external and internal guidelines?
- Q4: Where are the bottlenecks in the process?

Case 2: Hospital Information Systems Data Spectrum



a) Visualization of the spectrum

	Level of abstraction	Accuracy	Granularity	Directness	Correctness
Administrative systems	High	Low	Low	Low	Average
Clinical support systems	Average	Average	Average	Low	High
Healthcare logistics systems	Average	High/Average	Average/Low	High	High
Medical devices	Low	High/Average	High	High	Average

b) For each type of system, it is indicated which value is given for the 'level of abstraction' and 'accuracy' dimension. Also, for the 'accuracy' dimension it is indicated which value is given for each subdimension.

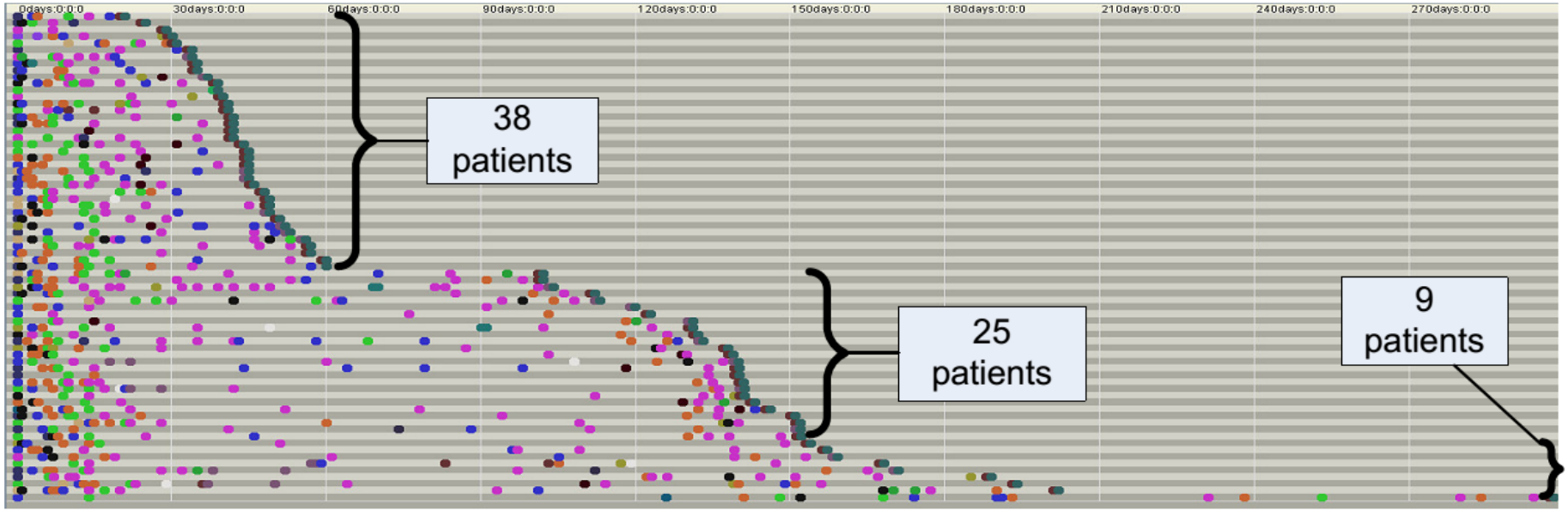
- For solving Q1..Q4 typically data from administrative systems are used
- Data from other systems may be required

Case 2: Gastro-enterology department, Surgery for colorectal cancer

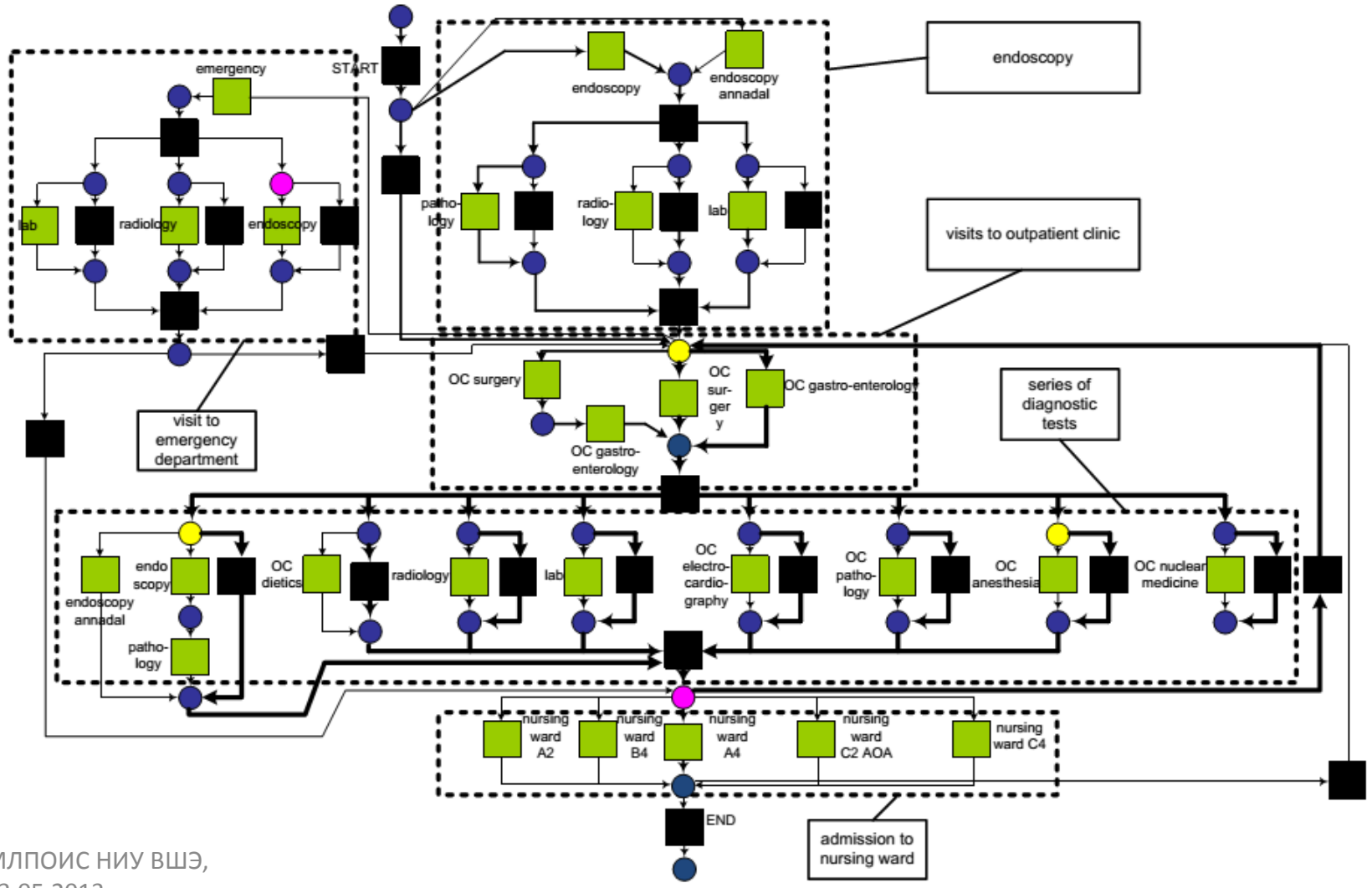
	patient identifier	day of birth	sex	address	requesting relation	executing doctor	requesting depart-ment	executing department	description department	operation	description operation
	A	B	C	D	E	F	G	H	I	J	K
1	Patiënt	Geboorte datum	Ge slach	BAS: plaats	Aanvragende Relatie	Uitvoerend Arts	Aanvr. verpl. OE	Uitvoerende OE	Beschrijving OE	Ver richtir	Beschrijving verrichting
2	999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	678700	m.c.v. mean corpusculair v
3	'999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	370407D	hemoglobine foto-elektrisch
4	'999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	370712B	leukocyten tellen elektr.hem
5	'999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	370715A	trombocyten tellen-electron
6	'999999	7-11-1950	Man	Maastricht	Mans	LKC algemeer	PINT	LCHE	LAB - KLIN.CHEMIE A.	370423	alkalische fosfatasebloed
7	'999999	7-11-1950	Man	Maastricht	Mans	LKC algemeer	PINT	LCHE	LAB - KLIN.CHEMIE A.	370442	natriumbl

	movement identifier	trajectory identifier	external status	trajectory code	start date operation	start year of trajectory	diag-nosis	description diagnosis	start date trajectory	number of operations	department identifier
	L	M	N	O	P	Q	R	S	T	U	V
1	Beweging	Traject Nummer	Status extern	Traject Code	Begindatum Verrichtin	Begin jaar	Diagnose	Diagnose omschrijving	Begindatum Traject	Aant. Verricht	OE_ Verrichtin
2	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_678700
3	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_370407D
4	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_370712B
5	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_370715A
6	1000144779/14	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LCHE_370423
7	1000144779/14	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LCHE_370442

Case 2: Patient process before surgery in dotted chart



Case 2: Patient process before surgery in Petri net



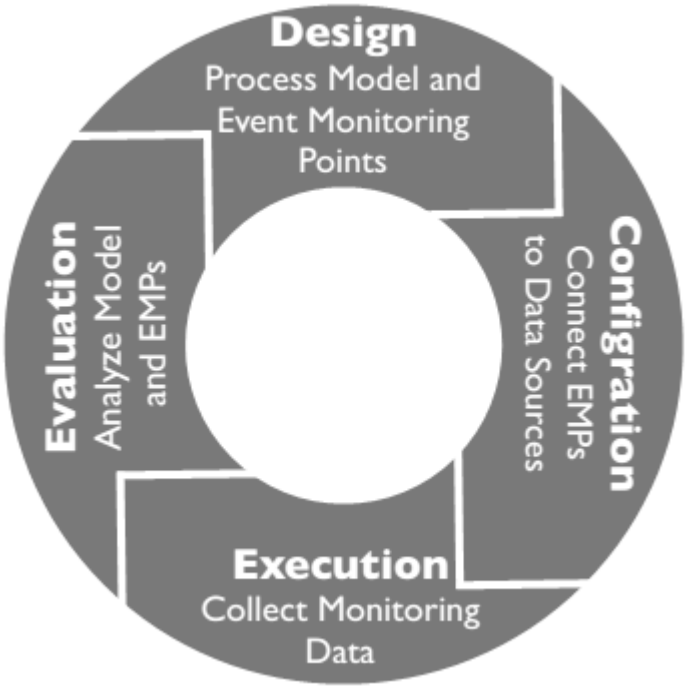
Case 2: Findings from process mining

- Three groups of patients can be distinguished
  - Did not receive radiotherapy before surgery (38 patients, 21..60 days)
  - Received radiotherapy before surgery (25 patients, 101..154 days)
  - Complex cases with individual treatment (6 patients, >154 days)
- Performance information in Petri Net:
  - Blue places: <5 days
  - Yellow places: 5..10 days
  - Pink places: >10 days
- Waiting time before the patient can be admitted (avg. 12,87 days, std.dev. 5.68 days)
- Waiting time before the patient has a contact with doctor (gastro-ent.: 4.25 days / 15.32 hours, surgery: 6.63 days / 16.61 hours)

Case 3: Conformance Checking in the Setting of Sparse Process Executing Information

- Embedding Conformance Checking in a Process Intelligence System in Hospital Environments?
- Authors: Kathrin Kirchner, Nico Herzberg, Andreas Rogge-Solti, Mathias Weske.(University Hospital of Jena, Hasso Plattner Institute at the Univ. of Potsdam)
- http://bpt.hpi.uni-potsdam.de/pub/Public/AndreasRoggeSolti/Conformance_Checking_BPI_Healthcare_ProHealth_12.pdf

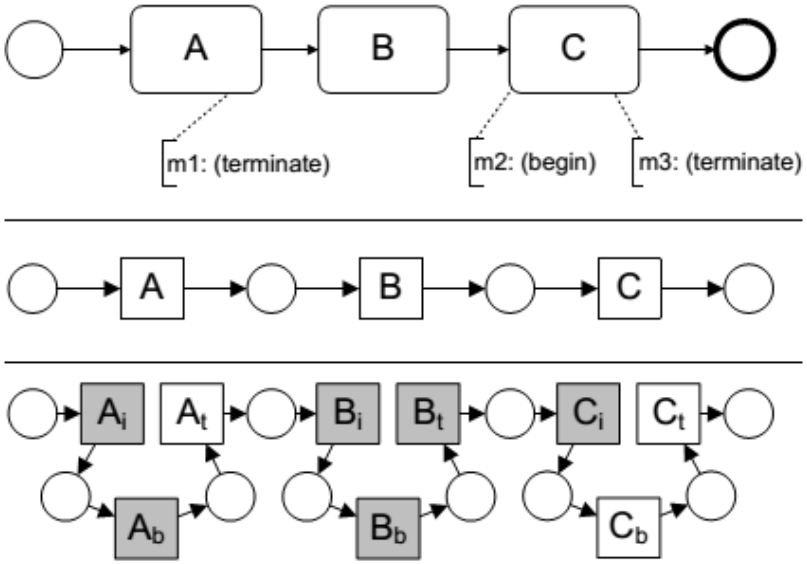
Case 3:Cyclic approach



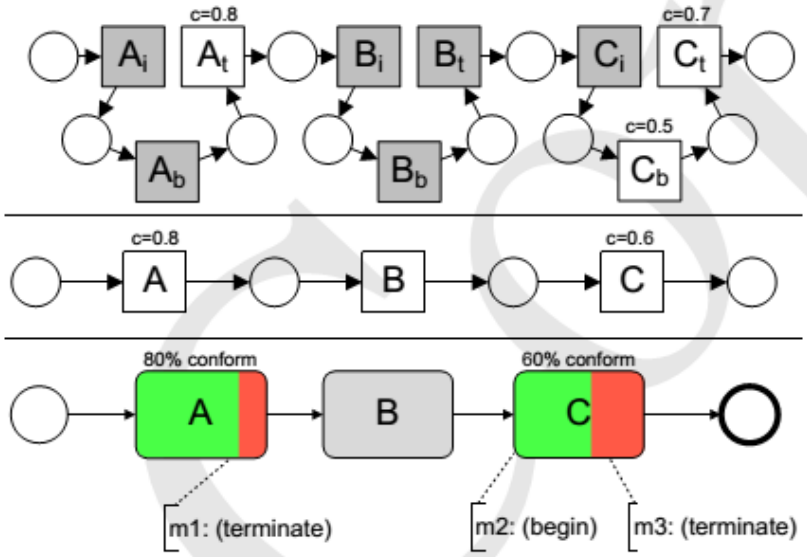
Case 3: Event monitoring points (EMP)

- EMPs have to be defined by domain experts that can describe when a certain task started or ended
  - When does the patient, having one or several liver diseases, arrive at surgery department for the first time?
  - When does the evaluation for the liver transplantation start?
  - When does the patient arrive at the operational room? When does she leave?
- EMPs have to be bound to implementation, e.g.
  - Web service call
  - Reading a certain cell in a spreadsheet
  - Executing SQL query

Case 3: Dealing with incomplete event data



(a) Translation of an annotated operational business process in BPMN into a Petri net. The annotations describe monitoring points attached to the model at certain state transitions. First step uses translations as in [9] and the second is a postprocessing step replacing each activity with a simplistic lifecycle of *init*, *begin*, and *terminate*. The observable transitions of the resulting Petri net are white, the unobservable are marked gray.

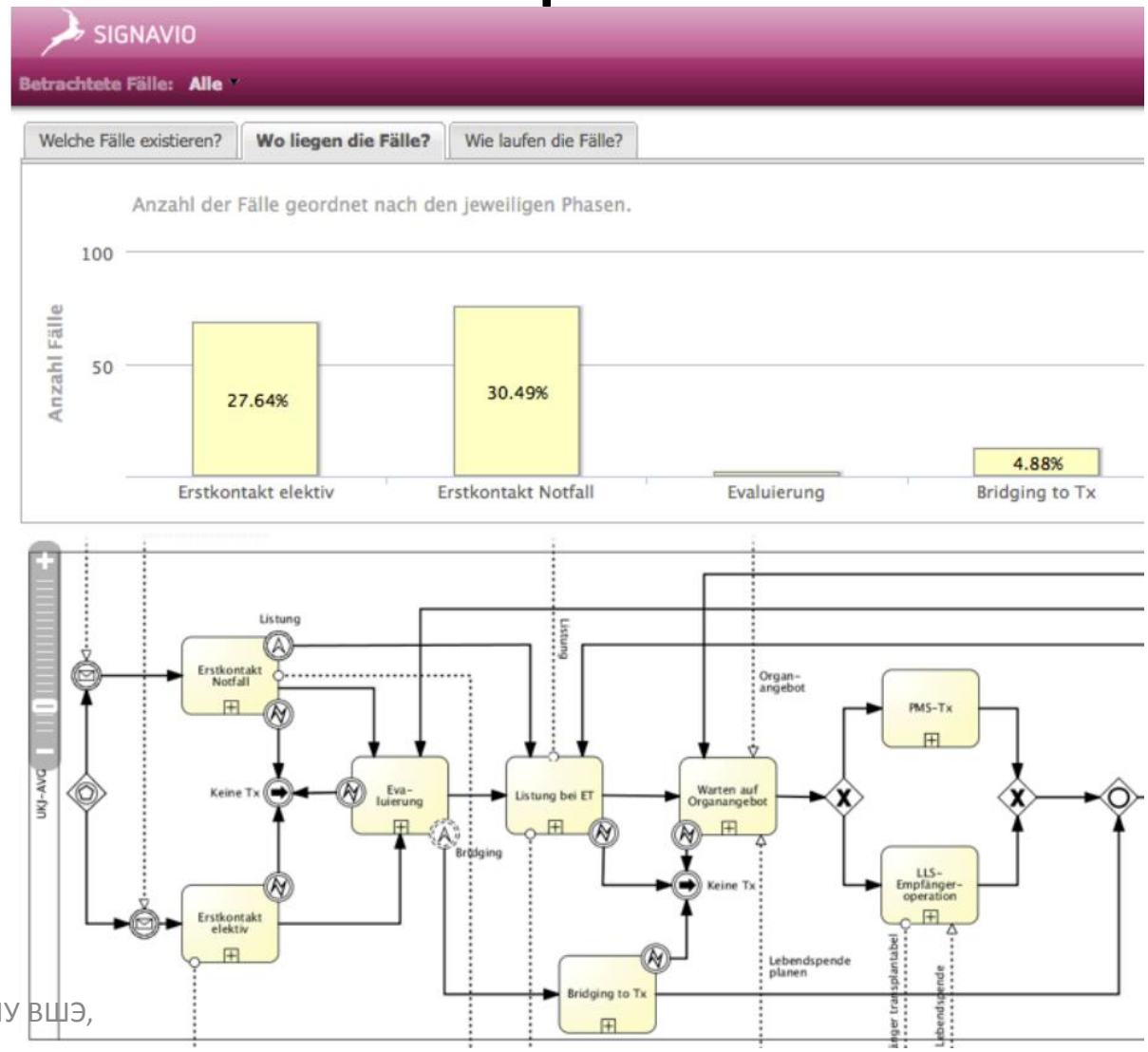


(b) With the conformance results for each monitoring point for a state transition, we calculate the error rate of the activity by the weighted sum of the error rates of each participating EMP. Further, we can propagate this information to the operational process model in BPMN. In this view, we mark unmonitored activities gray for the user to distinguish between elements that are recorded and those that only exist in the model.

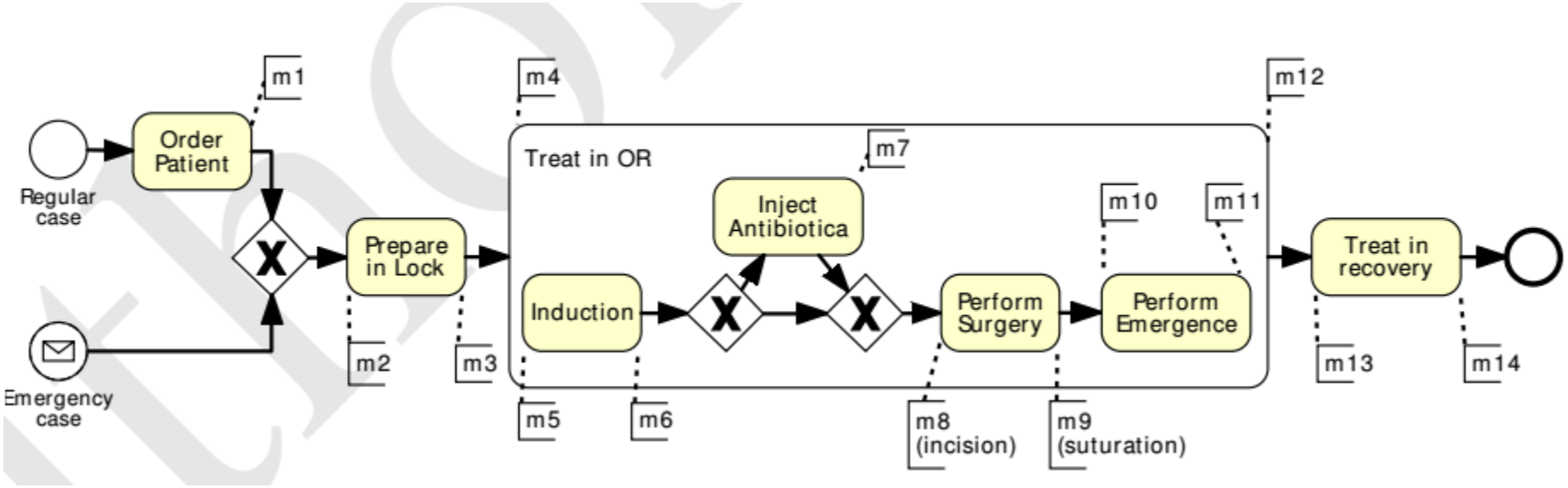
Case 3: Questions to be answered by monitoring system

- How did all of my cases perform?
- Which activities are performed for running instances at the moment?
- How does the process execution perform?
- Which cases exist?
- Where are the cases currently?

Case 3: Process model for liver transplantation



Case 3: Process model of surgery



Case 3: Cost-based alignment of model and log

EMP	event name	fitting	inserted	other pos.	all	failing	fitness
m_1	Patient ordered	552	0	0	552	0	1.000
m_2	Arrival in Lock	1006	280	9	1295	289	0.777
m_3	Departure of Lock	862	297	136	1295	433	0.666
m_4	Arrival in OR	1270	11	14	1295	25	0.981
m_5	Start of induction	1261	34	0	1295	34	0.974
m_6	End of induction	1253	42	0	1295	42	0.968
m_7	Antibiotics prophylaxis	183	0	112	295	112	0.620
m_8	Incision	1253	42	0	1295	42	0.968
m_9	Suturation	1239	55	1	1295	56	0.957
m_{10}	Start of emergence	1129	88	78	1295	166	0.872
m_{11}	End of emergence	1167	106	22	1295	128	0.901
m_{12}	Departure of OR	1219	62	14	1295	76	0.941
m_{13}	Arrival in recovery	1223	32	40	1295	72	0.944
m_{14}	Departure of recovery	1271	24	0	1295	24	0.981

Case 4: Process mining in huge, diverse and complex event logs

- Monitoring Deployed Application Usage with Process Mining.
- Authors: C.W. Gunther, A. Rozinat, W.M.P. van der Aalst, Kenny van Uden. (TU/e, Phillips Healthcare)
- http://bpmcenter.org/wp-content/uploads/reports/2008/BPM-08-11.pdf

Case 4: Setting

- >1400 Philips Healthcare's X-Ray machines distributed all over the world
- Fine grained event logs remotely transferred to Phillips Healthcare, millions of events
- Event types: single command (e.g. move a table, capture an image), boundaries of higher level abstractions (e.g. specific blood vessel analysis)

Case 4: Managing Diversity

- Diversity: several cases may be structured completely different, single cases differ significantly from one another
- Cause: different modes of X-ray machine
- Answer: Trace clustering (agglomerative hierarchical clustering) produce clusters = profiles, each measuring a number of features for each case

Case 4: Managing Complexity

- Complexity: high level of detail, richness in potential behavior, overwhelming number of artifacts (e.g. events in a log) combined with a large number of ordering relations
- Answer: Fuzzy mining
  - Metrics: significance (relative importance of artifact) and correlation (how closely events following one another are)
  - Highly significant events: preserve
  - Less significant, highly correlated: aggregate
  - Less significant, lowly correlated: remove

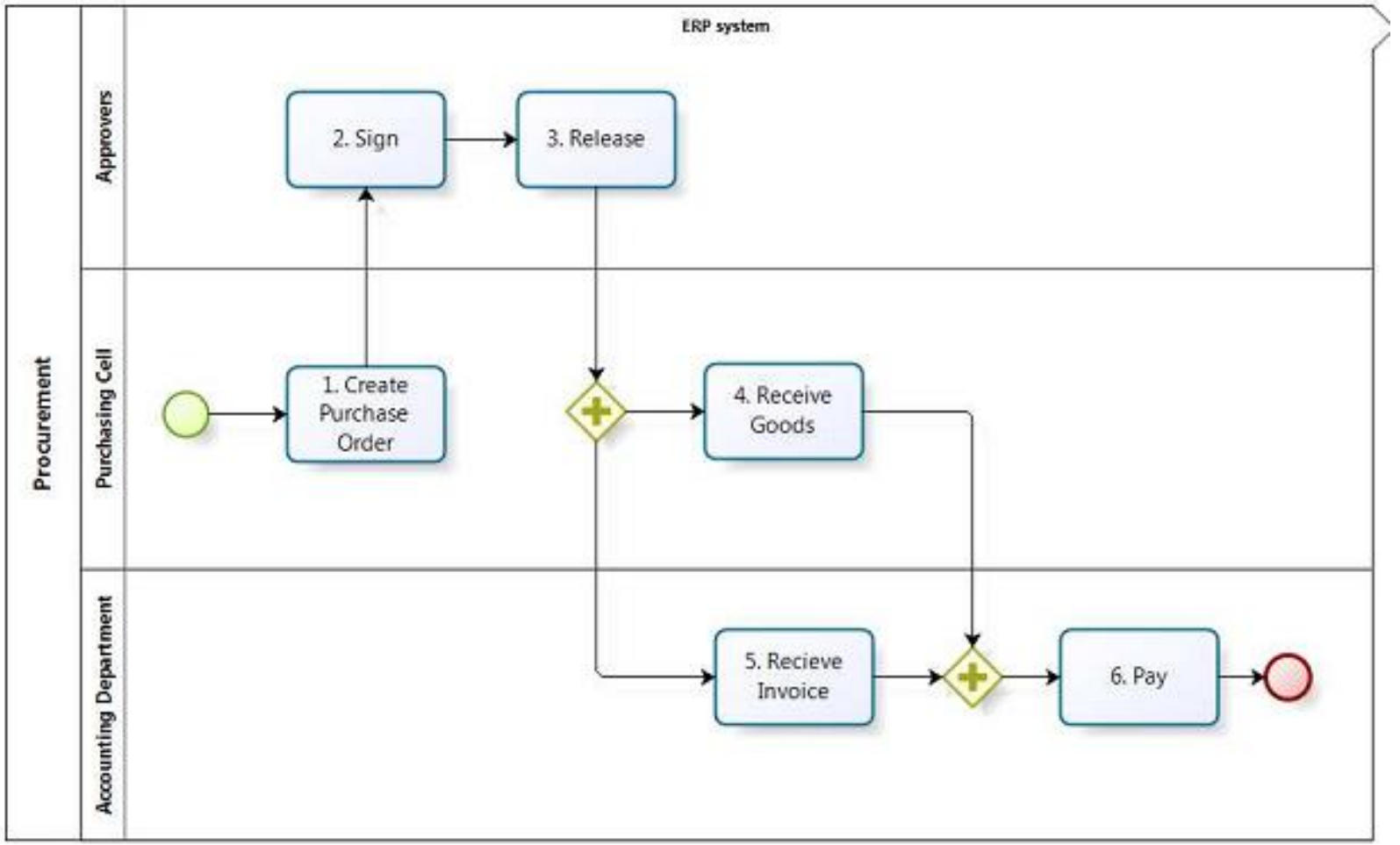
Case 4: Example data

- 41 system in US and Canada, 1801 logged days
- Most frequent (59/438) user profile: PatientPreparaton, ProcedureLeftCoronary, PocedureAbdomen, Patient Completion
- Surprise: majority of fluoroscopy runs (70%) consists of StartFluoroscopy and StopFluoroscopy with no actions in between (testing patterns doesn't reflect real usage)
- Confirm: diversity, no single event flow

Case 5: Procurement process rules conformance checking

- Process Mining of Event Logs in Internal Auditing: a Case Study.
- Authors: Mieke Jans, Michael Alles, Miklos Vasarhelyi. (Hasselt University, Rutgers Business School)
- http://www.systemsthinking.nl/Paper%20Jans%20Alles%20Vasarhelyi.pdf

Case 5: Procurement process



Case 5: Most frequent activity patterns

		Pattern Frequency		Cumulative total	throughput time (days)			
Pattern	**Sequence**	**#**	**%**	**%**	**avg**	**min**	**max**	**st.dev**
1	Create PO → Sign → Release → GR → IR → Pay	11,608	44.3%	44.3%	27.78	1	334	20.05
2	Create PO → Change Line → Sign → Release → GR → IR → Pay	6,955	26.6%	70.9%	32.33	2	343	57.72
3	Create PO → Change Line → Release → IR → Pay	2,488	9.5%	80.4%	75.63	3	344	38.99
4	Create PO → Release → IR → Pay	640	2.4%	82.8%	16.8	3	338	26.38
5	Create PO → Change Line → Sign → Release → IR → Pay	491	1.9%	84.7%	50.85	6	237	24.07
6	Create PO → Change Line → Sign → Release → IR → GR → Pay	393	1.5%	86.2%	56.36	9	295	40.16

Case 5: Process discovered using Fuzzy miner

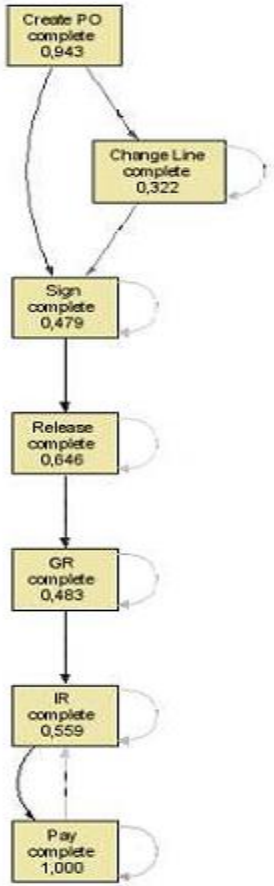


Figure 2. Output of Fuzzy Miner with default settings to uncover the core process in the event log.

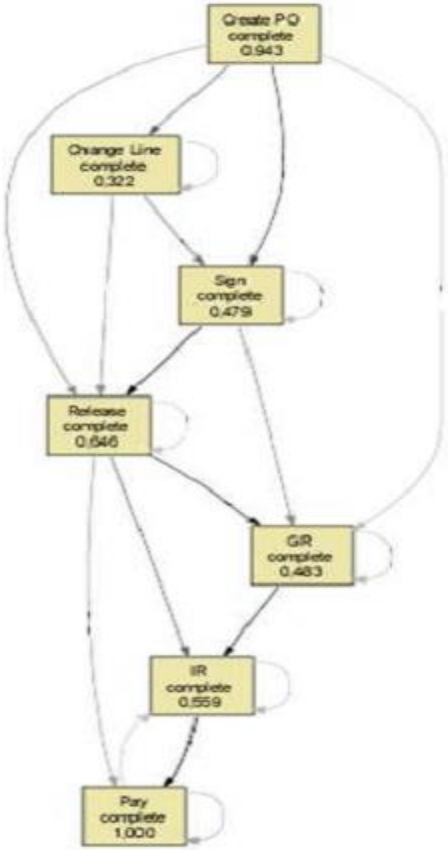
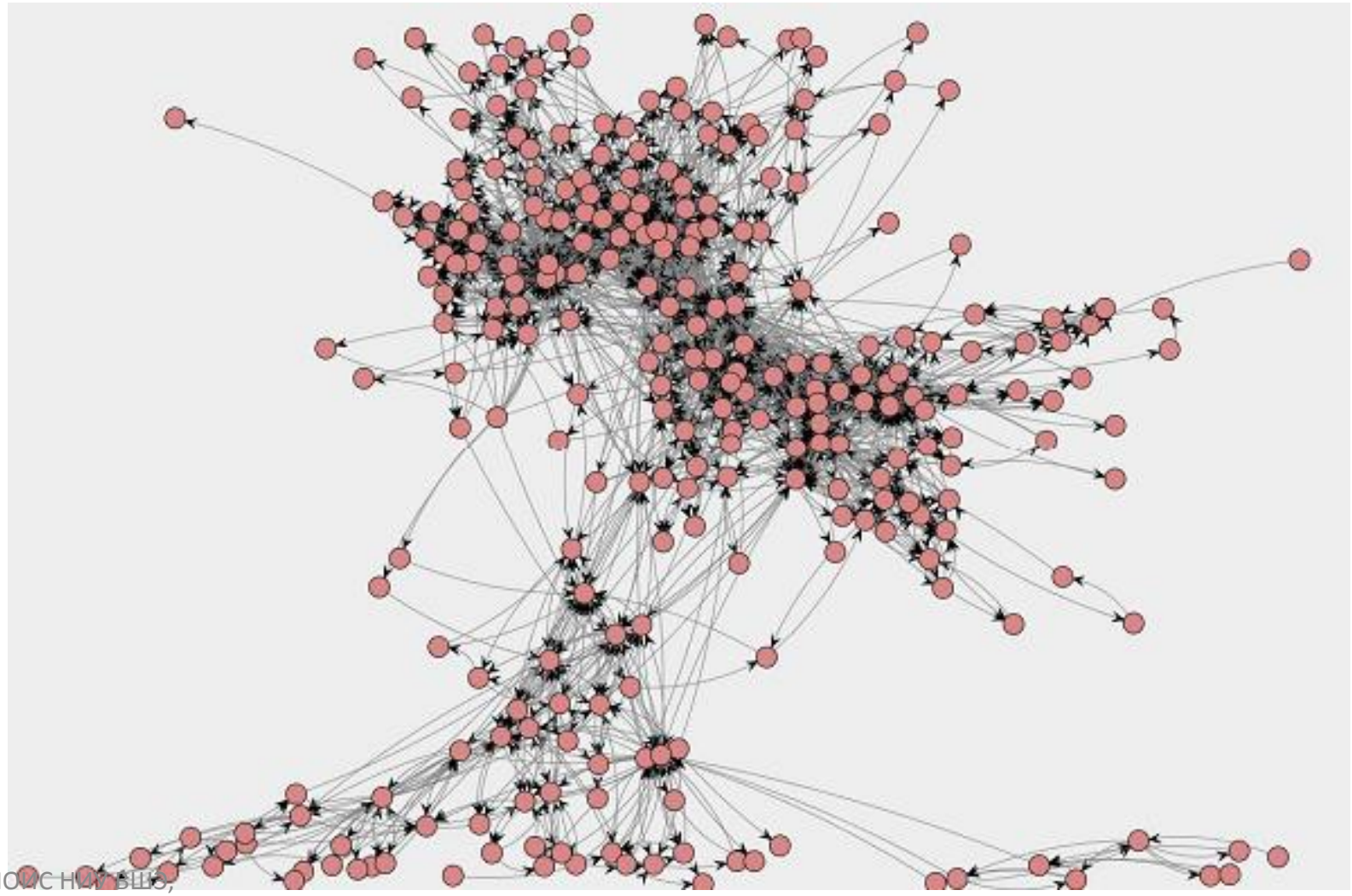


Figure 3. Output of Fuzzy Miner with lower threshold settings to reveal less-frequent flows

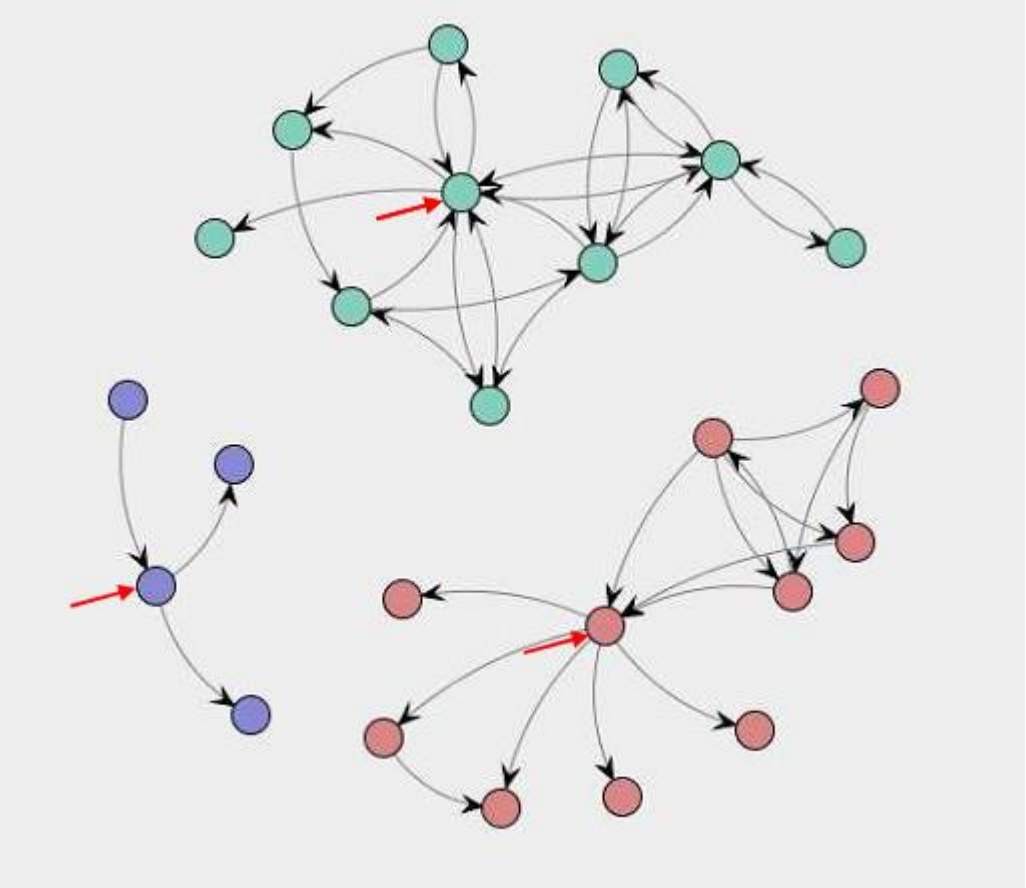
Case 5: Flows for further investigations

	Extra flows	Occurrences	Result
1	Create PO → GR	0	OK
2	Create PO → Release	739	Verification required for omitting Sign
3	Change Line → Release	2.790	Verification required for omitting Sign
4	Sign →GR	11	Further investigation → OK
5	Release → IR	4.973	Verification required on GR indicator
6	Release → Pay	244	Verification required on GR indicator and IR
7	Pay → IR	227	Verification required on IR

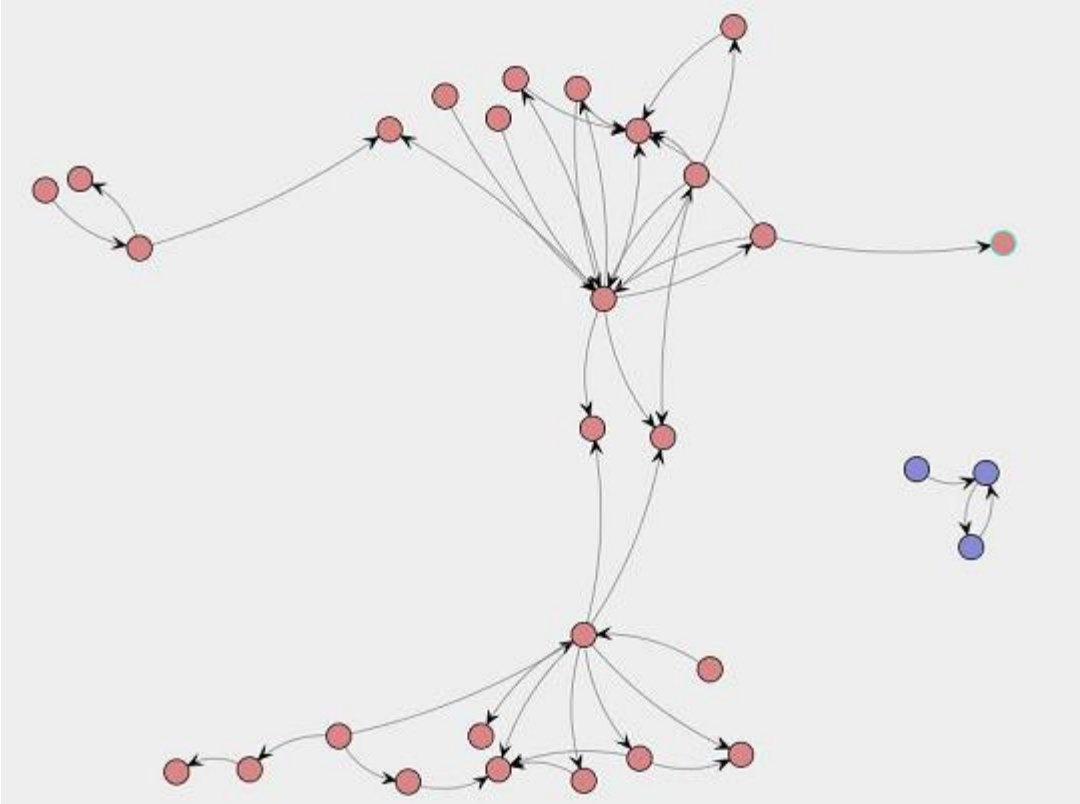
Case 5: Social network of all originators (272 individuals)



Case 5: 175 cases by three individuals violating controls on Release and GoodsReceipt



Case 5: 742 cases without Sign



Case 5: Rule violations unnoted by internal auditors and found by process mining

- Three PO's which passed without any Sign or Release
- 175 violations of segregation of duty principle
- 265 payment without matching invoice
- 3 PO's which did not show Good Receipt entry, although Good Receipt indicator was flagged
- 742 cases which did not show Sign activity

Thank you!
Questions?