

Dr. Pavlos Delias, EMATTECH, Greece
pdelias@teiemt.gr, @PavlosDelias



Process Analytics – Make it Work!





Office of Naval Research



US Naval Academy



Dec.	1	Midn't.	31 52	50 51		N. N. W.	Light breeze.	Nimbus and Cum. Stratus.	10	S. W.	30.10
		8 A. M.	31 59	49 48		S. W.	Moderate.	Nimbus and Cirro Cum. Stratus.	10	S'd. & W'd.	30.12
		4 P. M.	32 00	49.13		S. S. E.	Light breezes.	Nimbus.	10	Southward.	30.12
"	2	Midn't.	32 00	48 50		S. S. E.	Light breezes.	Nimbus.	10	Southward.	30.12
		8 A. M.	32 24	48 26	N. 53° W. 0.5	S. E.	Moderate.	Cum. Stratus.	8	S'd. & E'd.	30.22
		4 P. M.	32 25 20	47 50	Westward. 0.5	South.	Moderate.	Cir. Cum. Strat.	5	Southward.	30.18
"	3	Midn't.	32 25	47 1		S. by E.	Moderate.	Cum. Stratus.	6	Southward.	30.24

6	8	"	Thick rainy sea.	
7	7	"		
8	7	"		11
9	8	"	M.C. Dennis in Bay	
10	8	"		
11	8	"	Inconstant rain, with heavy squalls.	
12	8	"		P.S. No observations.




Course & Dist. *N 70 E 171* Lat { obs. *10. 10* } Lon { acc. *16. 10* } Lat { obs. *14. 10* } Lon { acc. *14. 10* } { mde } { chr. *19. 25* } { pm. *10* } { am. *11* }

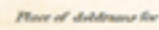

Depart. *60 E* of Acc. *v* of Acc. *v* Bar. *Ther. 68°*

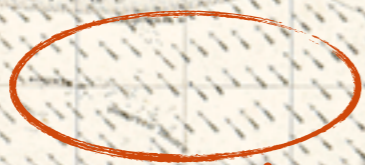
1	8	4	SE breeze fresh	Sunday 18 th April 1830.
2	9	"	heavy rain & squalls	
3	9	"		First had fresh breeze @ NE drawing to the North & NW in the middle, with inconstant heavy rain, & hard squalls: latterly moderate breeze @ NW & cloudy weather.
4	9	"		
5	9	"	North breeze fresh	
6	9	4	light rain & squalls	
7	10	4	Strong breeze with rain	
8	10	4	gale & constant rain	
9	10	"	Heavy rain	
10	10	4	Squally	
11	10	4	Hard rain	
12	10	"	Heavy rain	
1	10	"		All Perform'd Divine Service in the Stairage, the unsettled weather preventing it on Deck. At Noon muster'd the Ship's Comp ^y .
2	10	4	North heavy squalls & rain	
3	10	4		
4	10	"	SW	
5	9	4		
6	10	"		
7	9	4	Fine	
8	13	"	Out-rigged & made all sail.	Sick List 9 Seamen.
9	7	4		
10	7	4		Water expended this week Gallons 1061

EXPLANATION

-  North East Trades
-  South East Trades
-  South East & South West Monsoons
-  North East & North West Monsoons
-  Prevailing Westerly Winds

-  Route & average passage in days
-  Fair Winds
-  Good Winds

-  Floor of atmosphere for March line
-  Sept



Winds



Ideal Routes



©PHOTO Rienk Mebius 2007





©PHOTO Rienk Mebius 2007

Event (Big) Data exist...

	A	B	C
1	CaseID	Event	TimeStamp
2	1	Arrival	1/6/12 21:50
3	1	Assortment	1/6/12 21:55
4	1	Diagnosis	1/6/12 21:57
5	1	Blood_Test	1/6/12 22:05
6	1	Blood_Test	1/6/12 22:45
7	1	Biochem_Test	1/6/12 22:05
8	1	Biochem_Test	1/6/12 23:20
9	1	Entrance_to_ER_Room	1/6/12 21:55
10	1	Exit_from_ER_Room	1/6/12 23:35
11	1	ER_Exit	1/6/12 23:35
12	2	Arrival	1/6/12 22:30
13	2	Assortment	1/6/12 22:35
14	2	Diagnosis	1/6/12 22:36
15	2	Blood_Test	1/6/12 22:40
16	2	Blood_Test	1/6/12 23:20
17	2	Biochem_Test	1/6/12 22:40
18	2	Biochem_Test	1/6/12 23:55
19	2	Additional_Test	1/6/12 22:50
20	2	Additional_Test	1/6/12 23:50
21	2	Entrance_to_ER_Room	1/6/12 22:35
22	2	Exit_from_ER_Room	2/6/12 0:09
23	2	Prescription	2/6/12 0:05
24	2	ER_Exit	2/6/12 0:10
25	3	Arrival	1/6/12 23:05
26	3	Assortment	1/6/12 23:10
27	3	Diagnosis	1/6/12 23:12
28	3	Blood_Test	1/6/12 23:15
29	3	Blood_Test	1/6/12 23:55
30	3	Biochem_Test	1/6/12 23:15
31	3	Biochem_Test	2/6/12 0:35
32	3	Additional_Test	1/6/12 23:25
33	3	Additional_Test	2/6/12 0:20
34	3	Entrance_to_ER_Room	1/6/12 23:10
35	3	Exit_from_ER_Room	2/6/12 0:17

...in healthcare

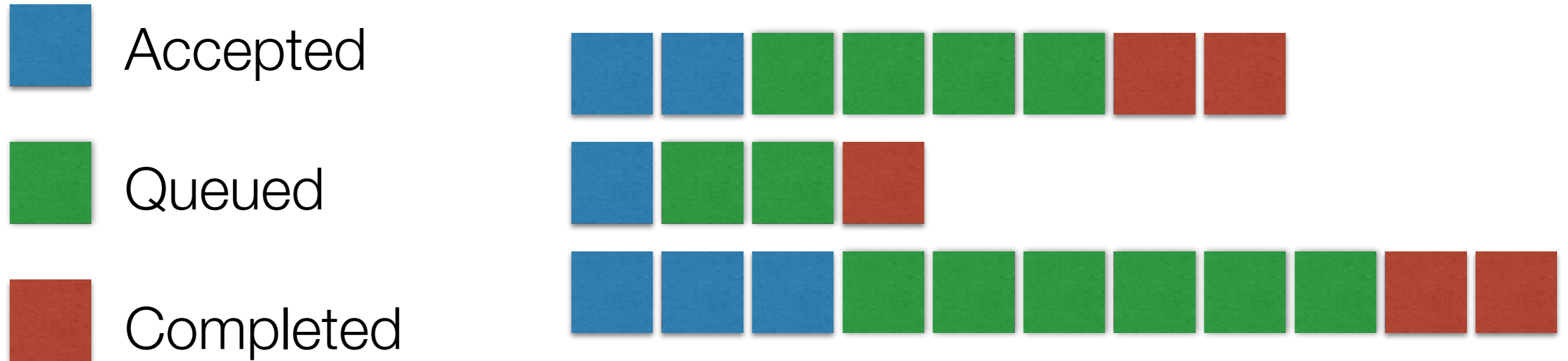
Event (Big) Data exist...

	A	B	C
1	CaseID	Event	TimeStamp
2	1	Arrival	1/6/12 21:50
3	1	Assortment	1/6/12 21:55
4	1	Diagnosis	1/6/12 21:57
5	1	Blood_Test	1/6/12 22:05
6	1	Blood_Test	1/6/12 22:45
7	1	Biochem_Test	1/6/12 22:05
8	1	Biochem_Test	1/6/12 23:20
9	1	Entrance_to_ER_Room	1/6/12 21:55
10	1	Exit_from_ER_Room	1/6/12 23:35
11	1	ER_Exit	1/6/12 23:35
12	2	Arrival	1/6/12 22:30
13	2	Assortment	1/6/12 22:35
14	2	Diagnosis	1/6/12 22:36
15	2	Blood_Test	1/6/12 22:40
16	2	Blood_Test	1/6/12 23:20
17	2	Biochem_Test	1/6/12 22:40
18	2	Biochem_Test	1/6/12 23:55
19	2	Additional_Test	1/6/12 22:50
20	2	Additional_Test	1/6/12 23:50
21	2	Entrance_to_ER_Room	1/6/12 22:35
22	2	Exit_from_ER_Room	2/6/12 0:09
23	2	Prescription	2/6/12 0:05
24	2	ER_Exit	2/6/12 0:10
25	3	Arrival	1/6/12 23:05
26	3	Assortment	1/6/12 23:10
27	3	Diagnosis	1/6/12 23:12
28	3	Blood_Test	1/6/12 23:15
29	3	Blood_Test	1/6/12 23:55
30	3	Biochem_Test	1/6/12 23:15
31	3	Biochem_Test	2/6/12 0:35
32	3	Additional_Test	1/6/12 23:25
33	3	Additional_Test	2/6/12 0:20
34	3	Entrance_to_ER_Room	1/6/12 23:10
35	3	Exit_from_ER_Room	2/6/12 0:17



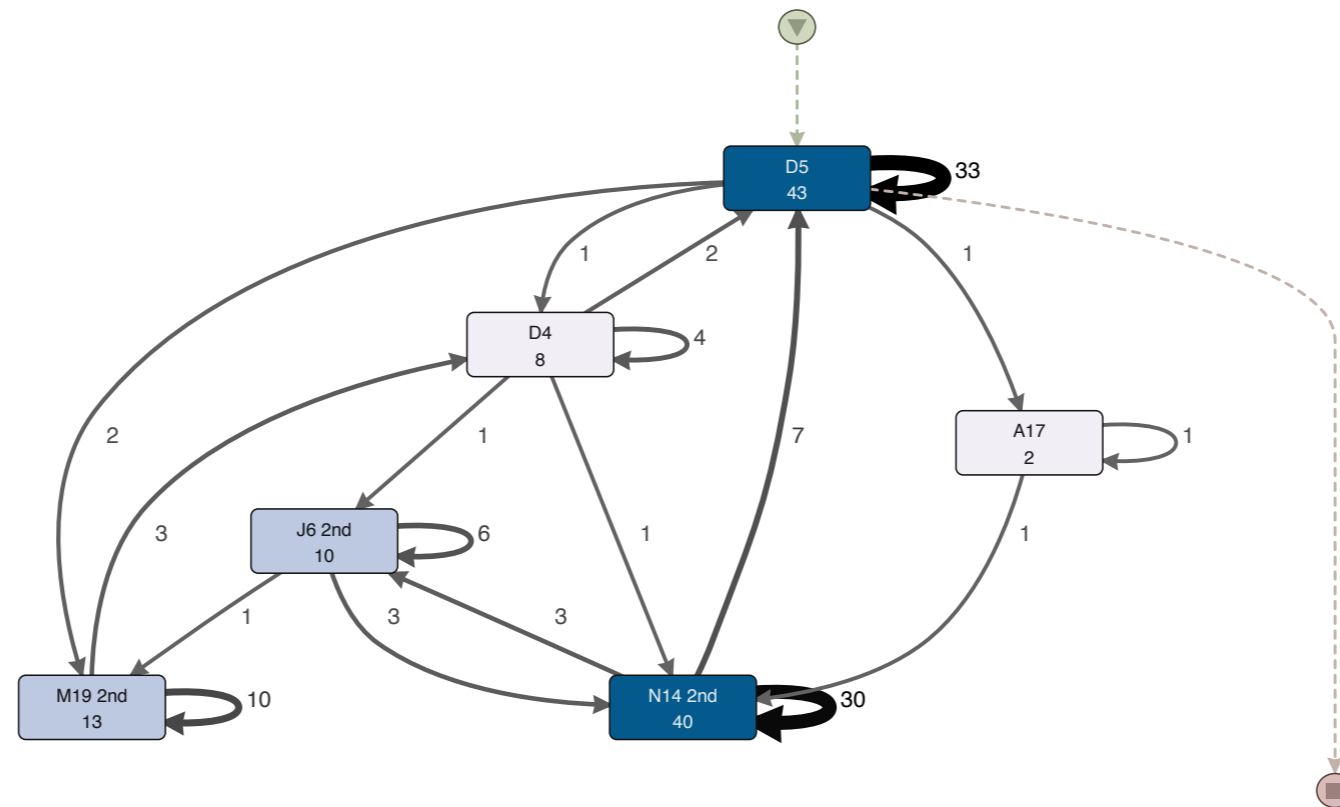
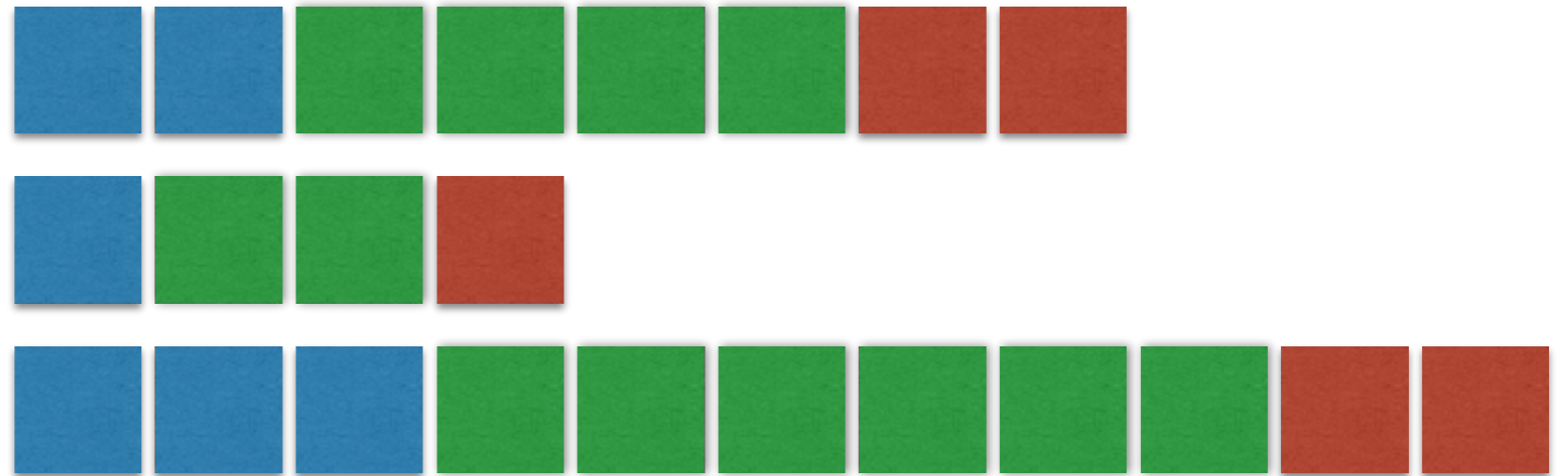
...in healthcare

Event (Big) Data exist...



...in CRM systems

Event (Big) Data exist...



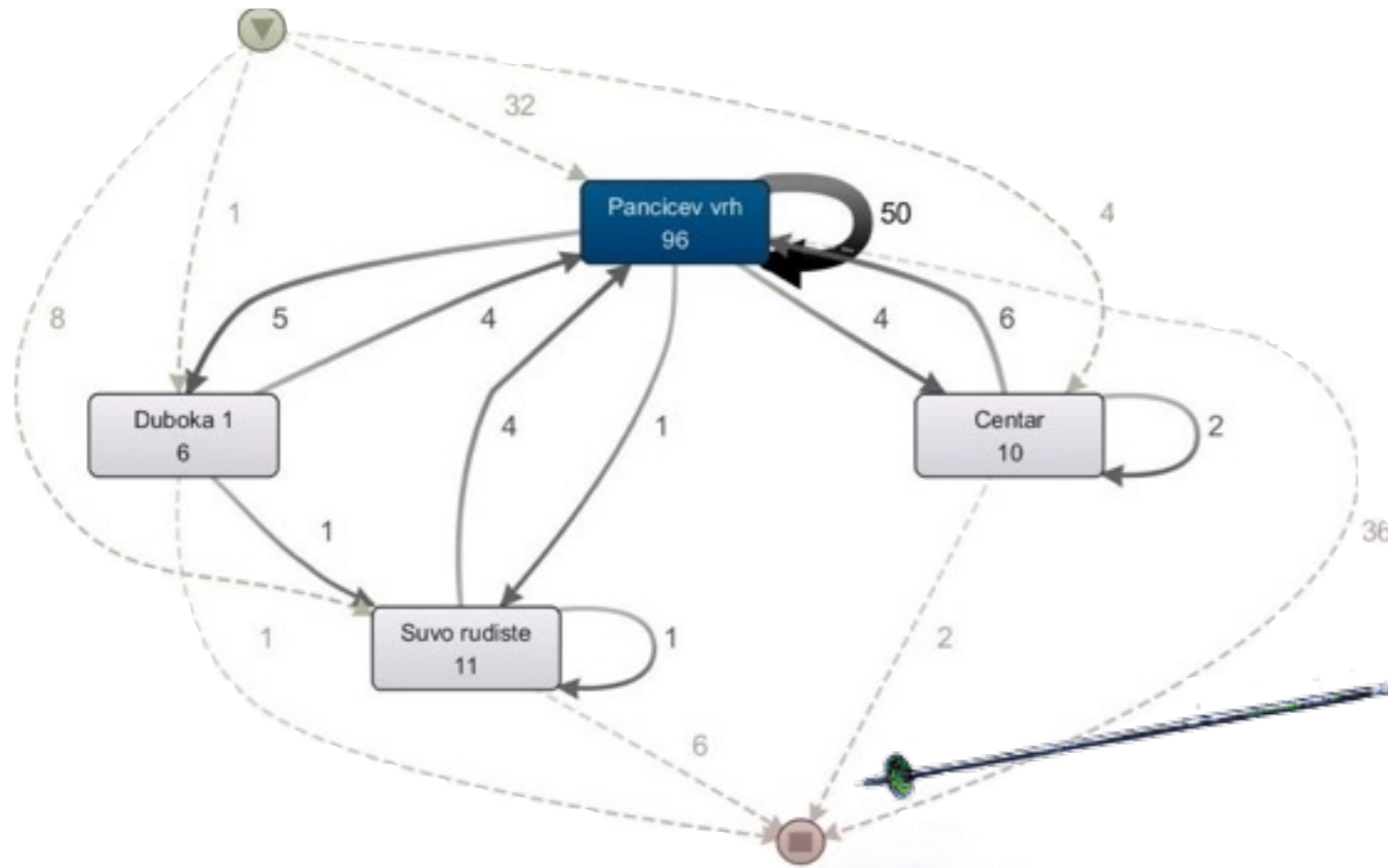
...in CRM systems

Event (Big) Data exist...

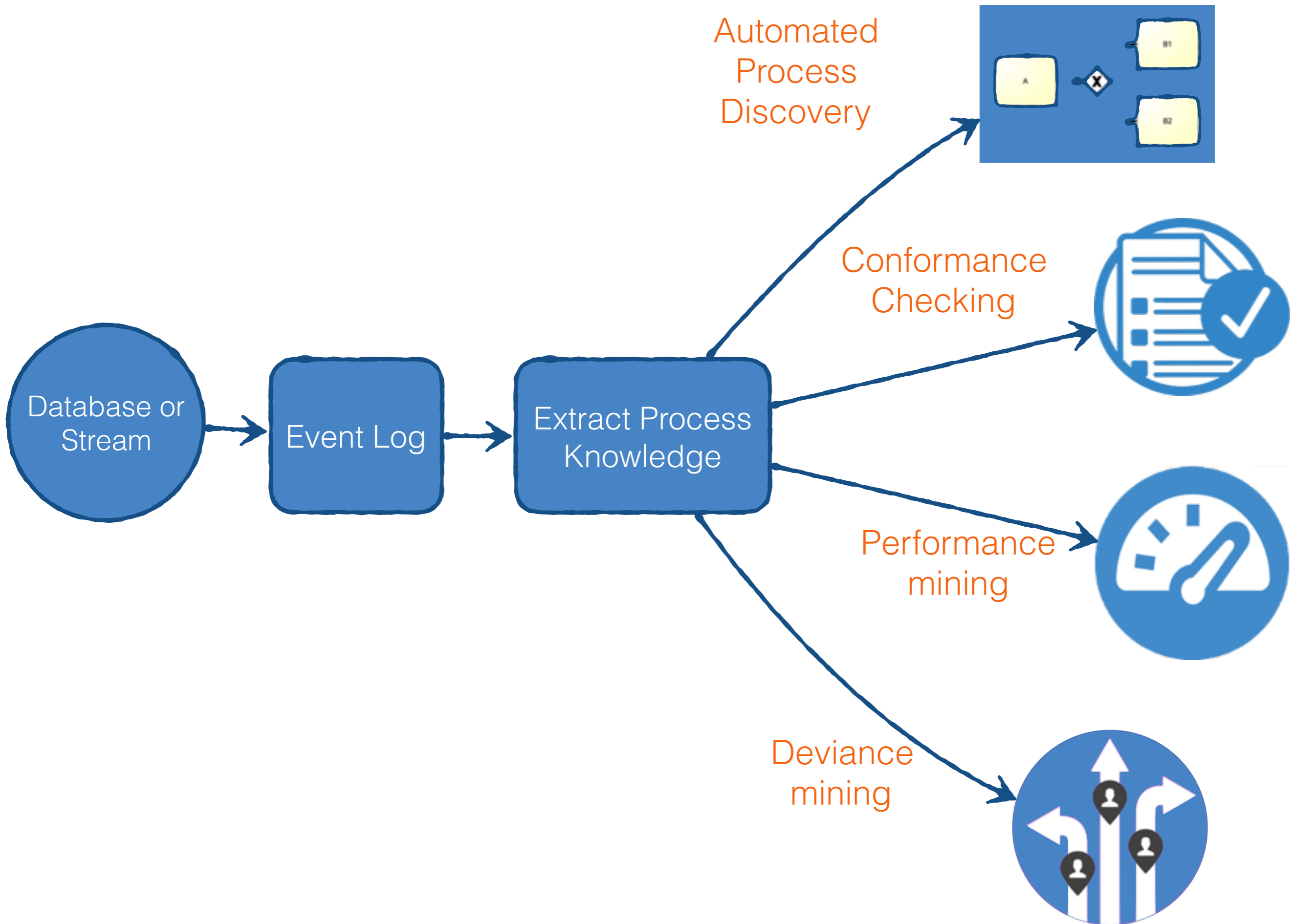


...in RFID logs

Event (Big) Data exist...



...in RFID logs



Value Proposition

- Understand your processes as they are. Not as you imagine them
- Back your hypotheses with evidence. Not only intuitions and beliefs
- Quantify the impact of redesign options. Before and after

Event Log

Mandatory Elements



Case ID



Activity



Timestamp

Optional Elements



Resource

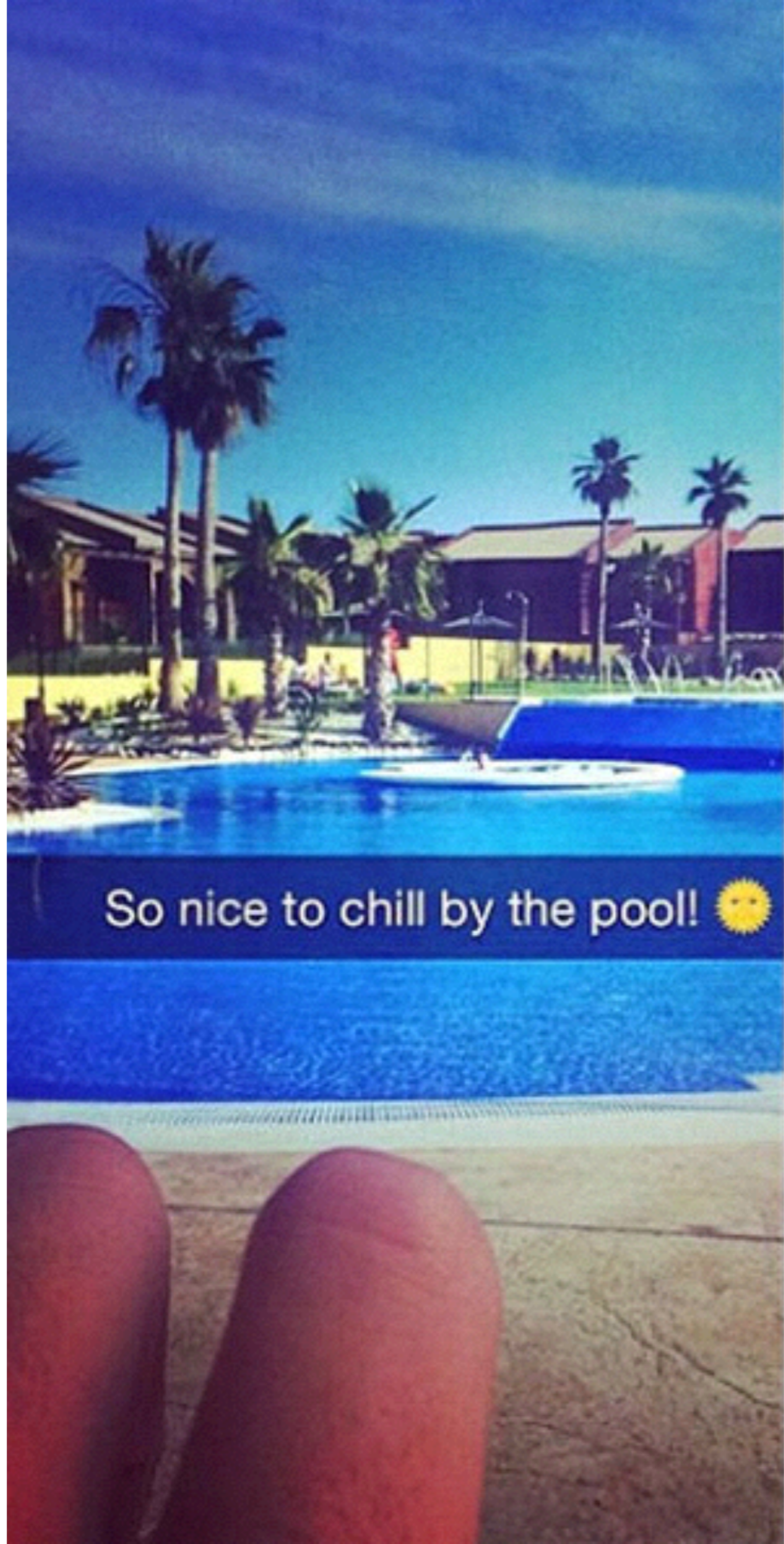
...



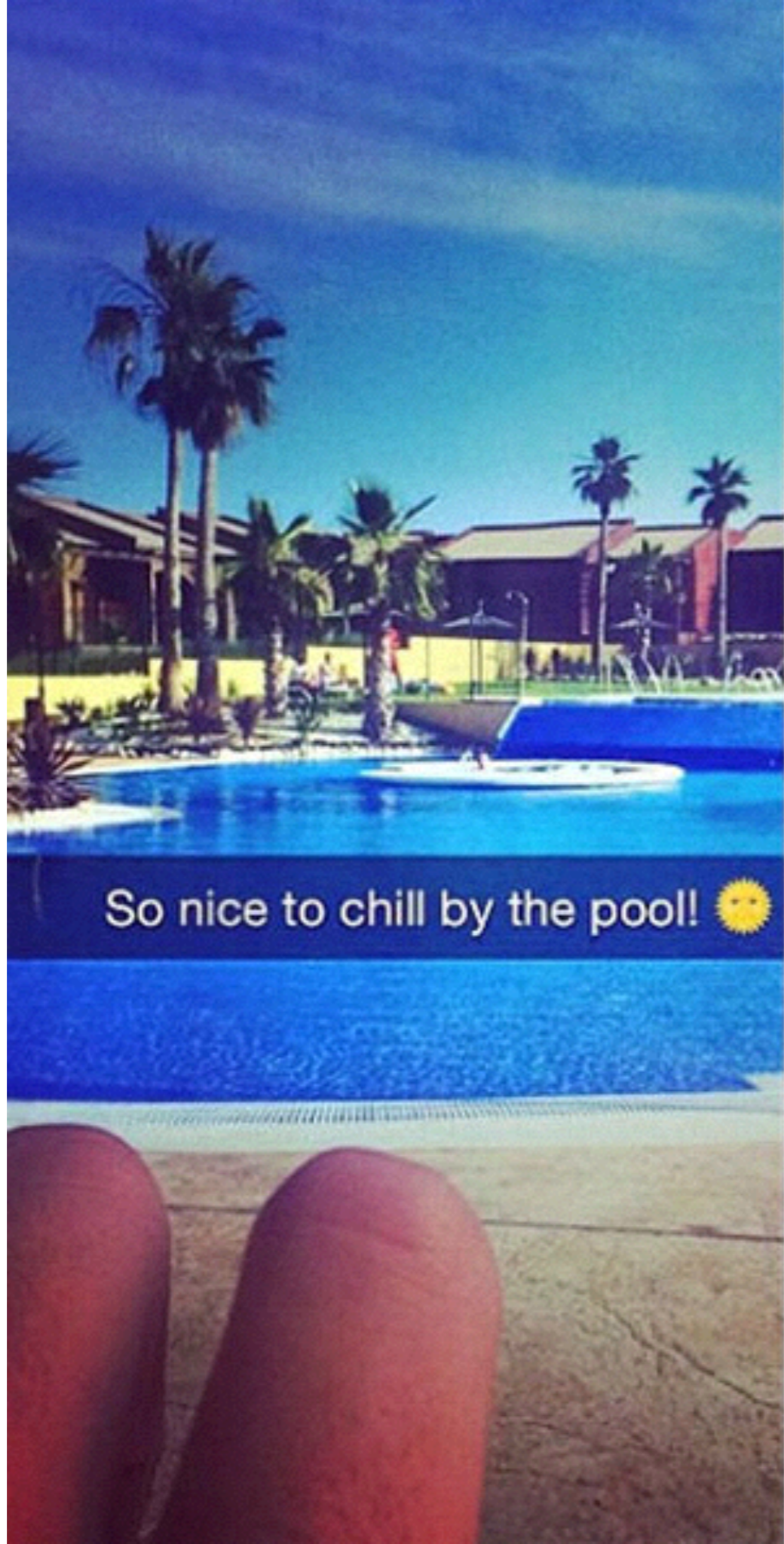
Attributes

Challenges in Event Log Extracting

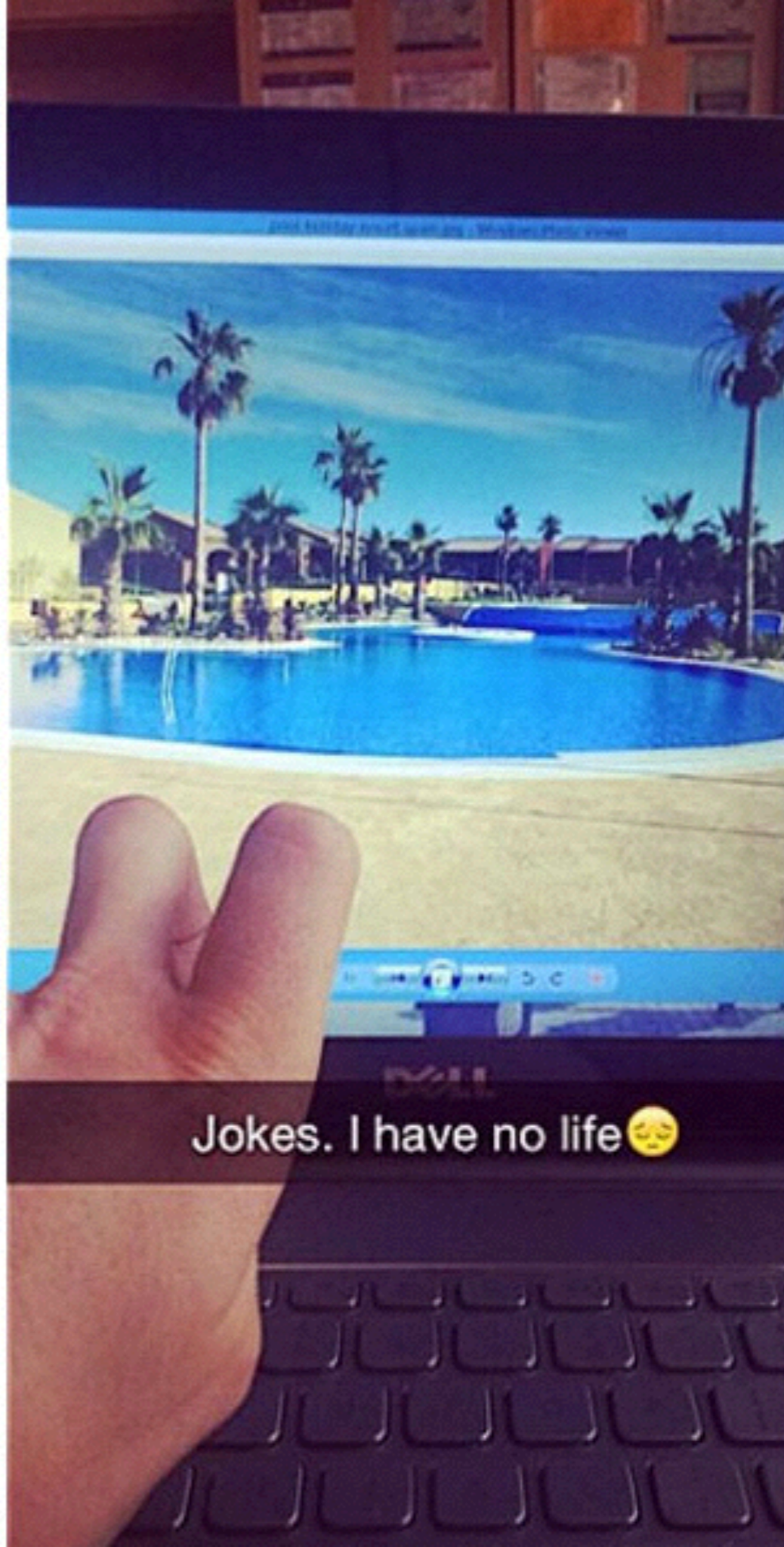
- **Correlation** (case ID is not explicitly recorded)
- **Timestamps** (reliability, different sources)
- **Snapshots** (duration / completeness of the log)
- **Scoping** (Which tables of the DB do we need?)
- **Granularity**



So nice to chill by the pool! ☀️



So nice to chill by the pool! ☀️



Jokes. I have no life 😞

Discovery Basics: The a- algorithm

- Ordering Relations $>$, \rightarrow , \parallel , $\#$
- **Direct succession:** $x > y$ iff for some case x is directly followed by y
- **Causality:** $x \rightarrow y$ iff $x > y$ and not $y > x$
- **Parallel:** $x \parallel y$ iff $x > y$ and $y > x$
- **Unrelated:** $x \# y$ iff not $x > y$ and not $y > x$

$A > B$

$A > C$

$B > C$

$B > D$

$C > B$

$C > D$

$E > F$

ABCD

ACBD

EF

$B \parallel C$

$C \parallel B$

$A \rightarrow B$

$A \rightarrow C$

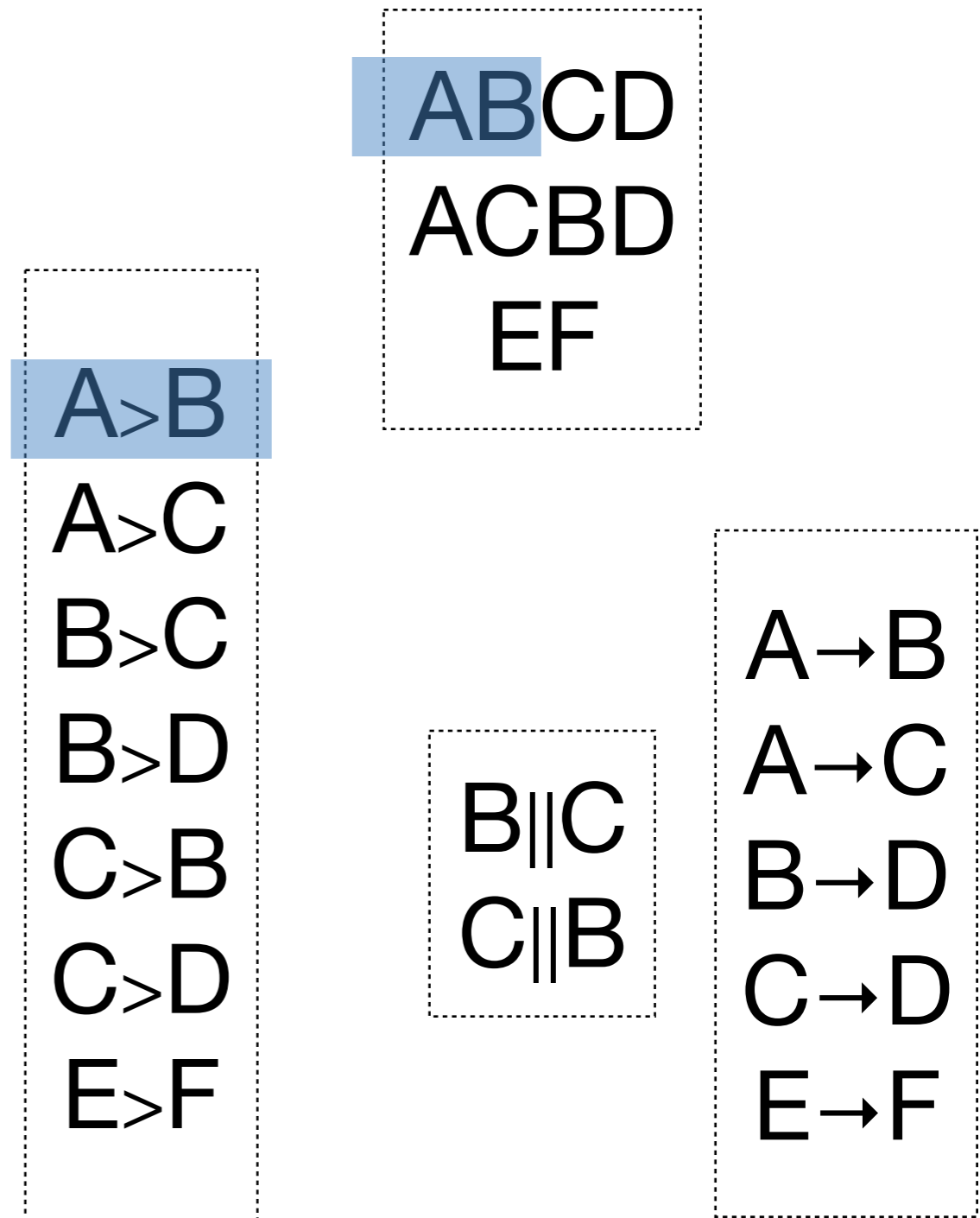
$B \rightarrow D$

$C \rightarrow D$

$E \rightarrow F$

Discovery Basics: The a- algorithm

- Ordering Relations $>$, \rightarrow , \parallel , $\#$
- **Direct succession:** $x > y$ iff for some case x is directly followed by y
- **Causality:** $x \rightarrow y$ iff $x > y$ and not $y > x$
- **Parallel:** $x \parallel y$ iff $x > y$ and $y > x$
- **Unrelated:** $x \# y$ iff not $x > y$ and not $y > x$

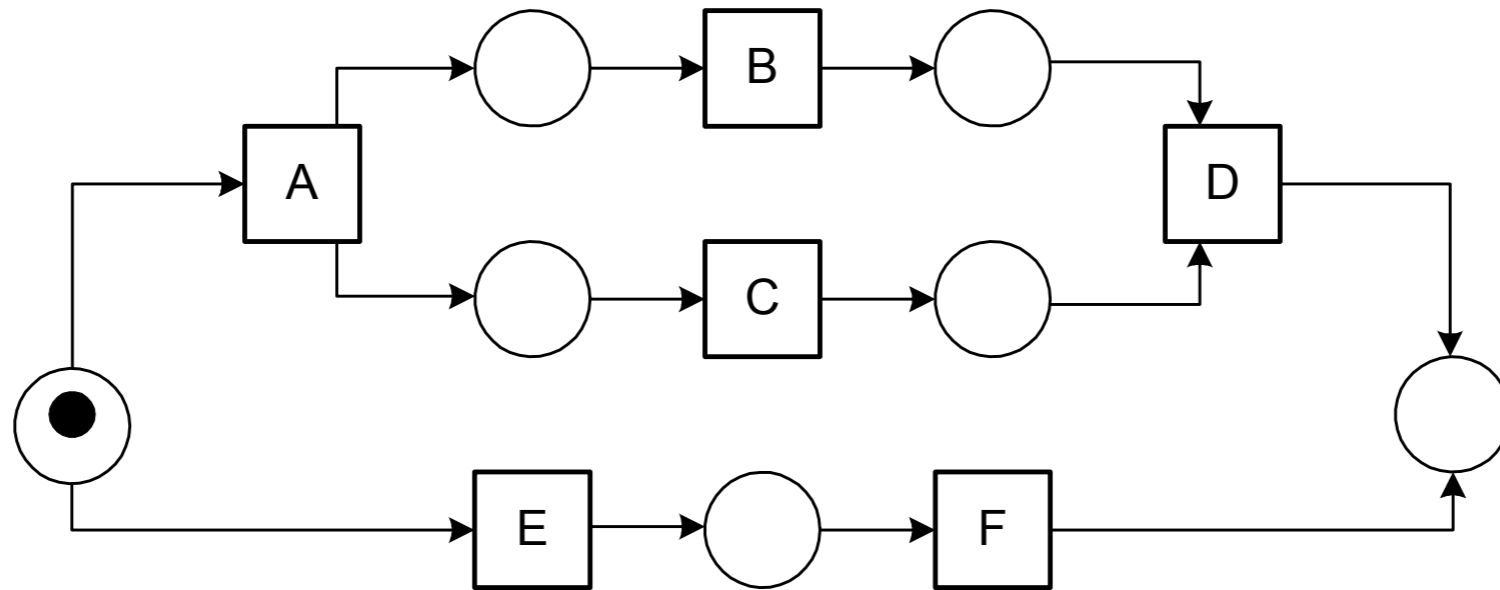


a-algorithm continued...

ABCD
ACBD
EF

B||C
C||B

A → B
A → C
B → D
C → D
E → F

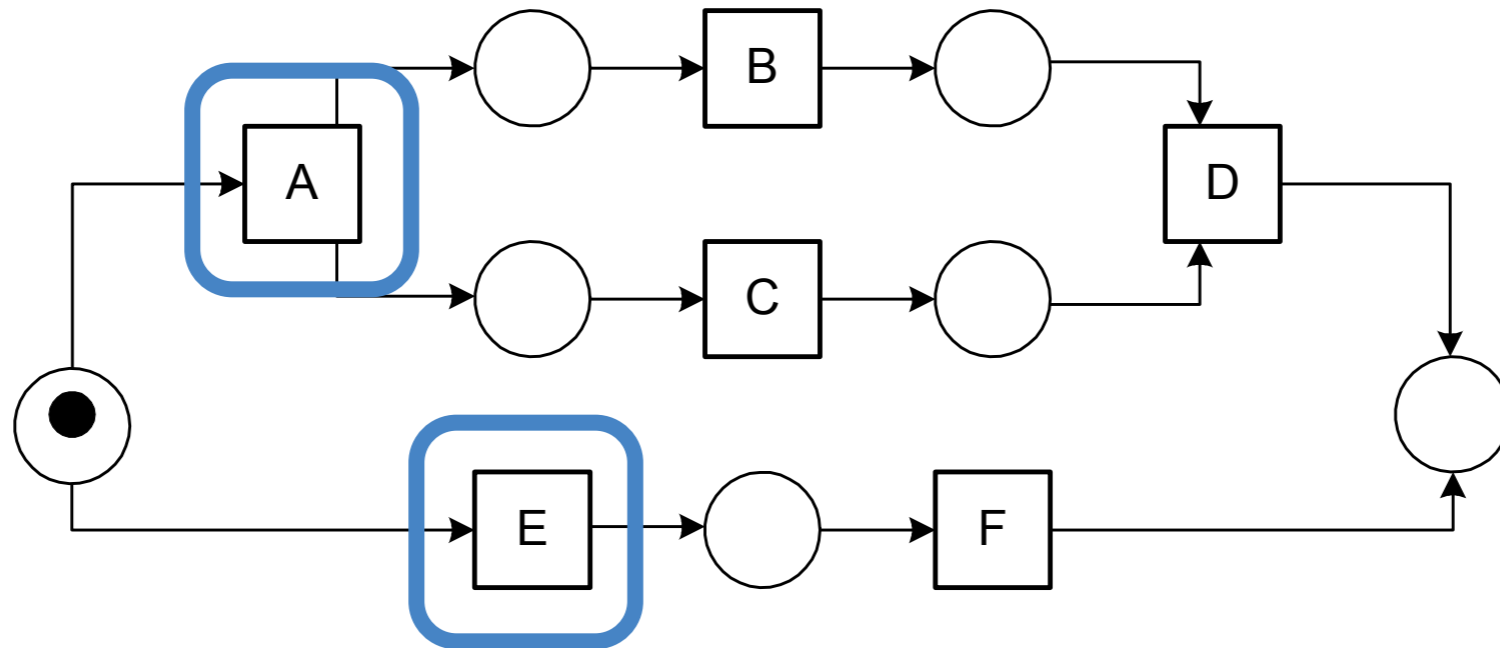


a-algorithm continued...

A B C D
A C B D
E F

B||C
C||B

A → B
A → C
B → D
C → D
E → F

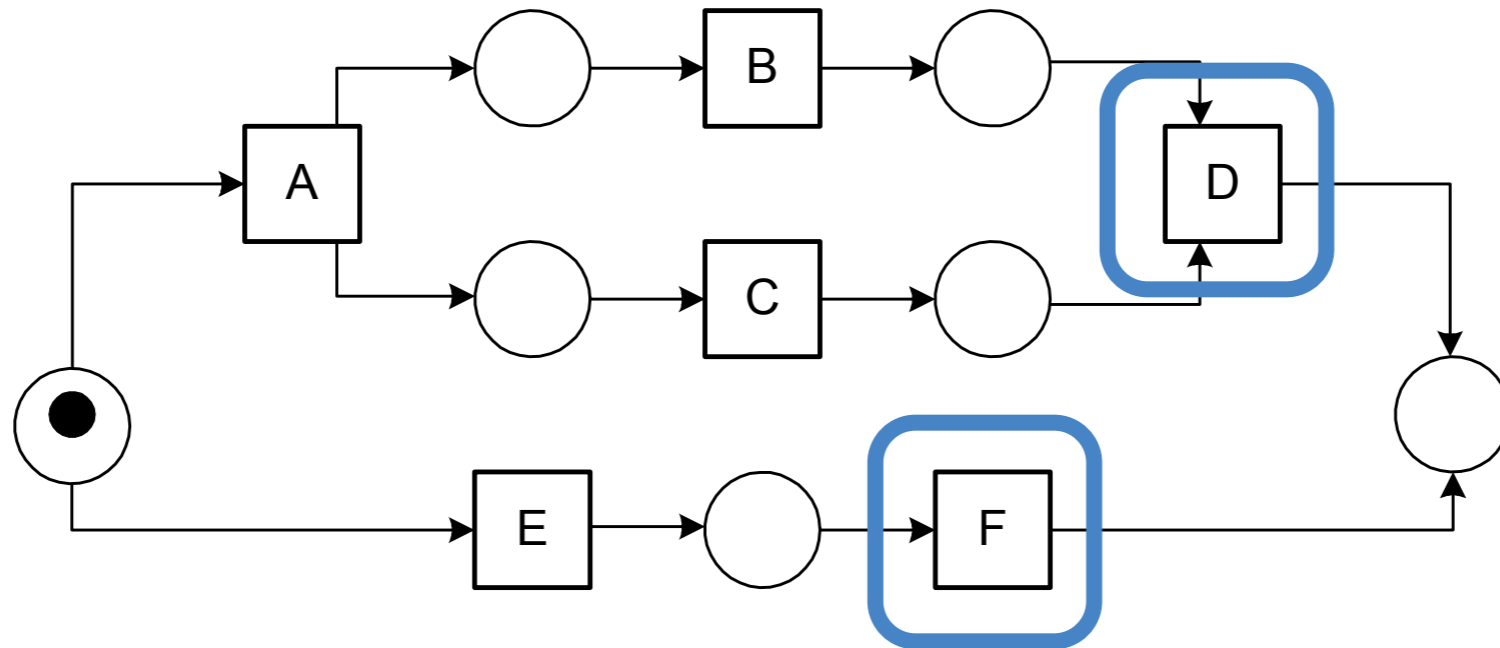


a-algorithm continued...

ABCD
ACBD
EF

B||C
C||B

A → B
A → C
B → D
C → D
E → F

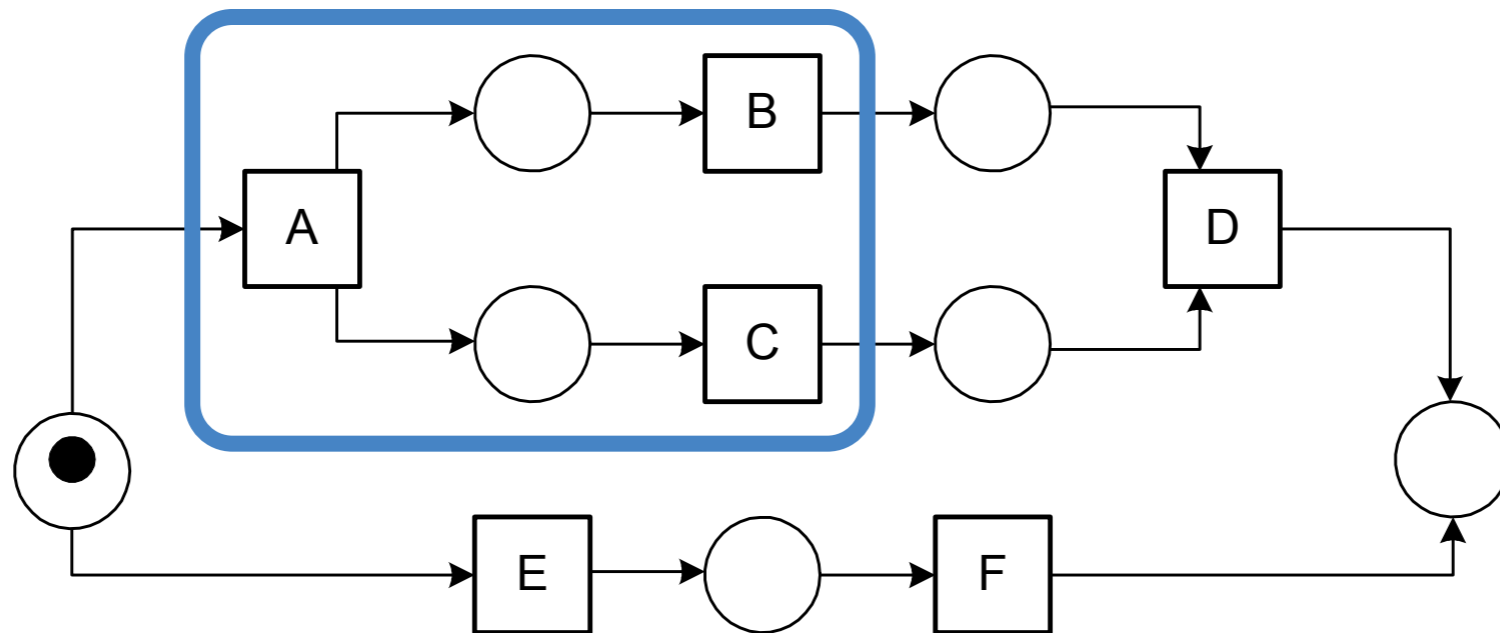


a-algorithm continued...

ABCD
ACBD
EF

B||C
C||B

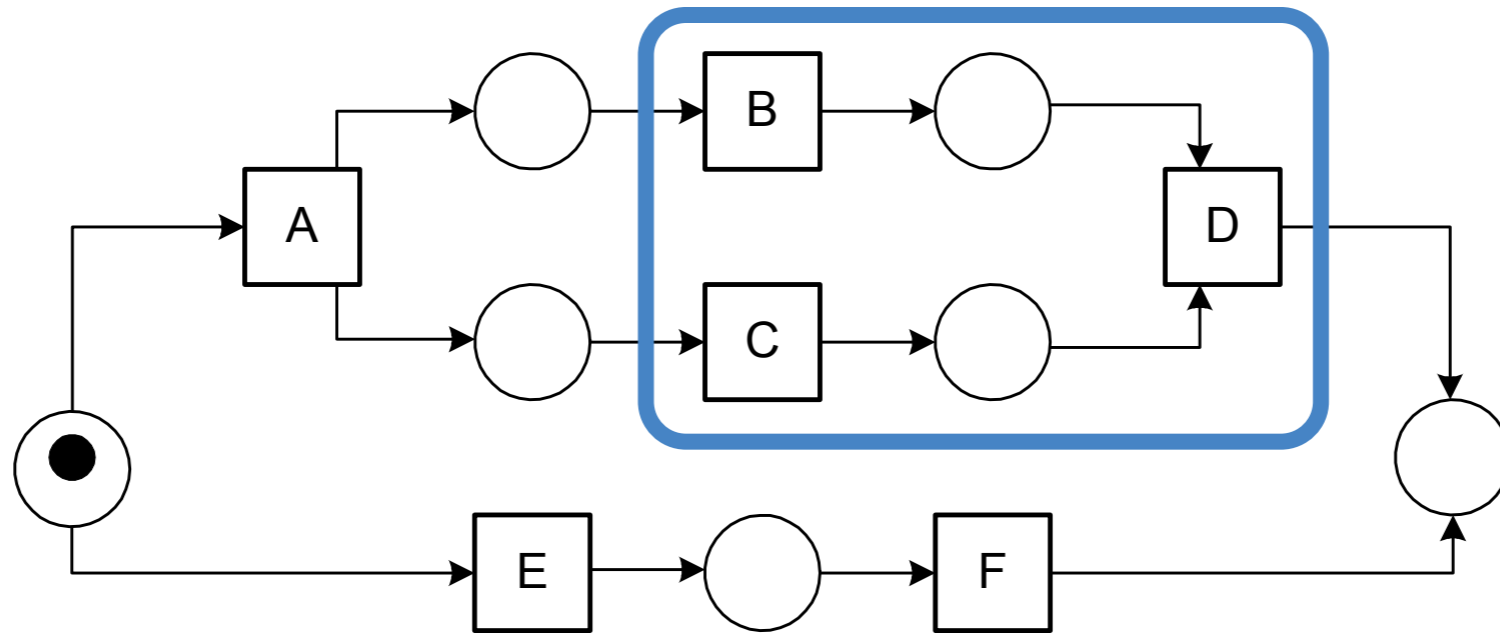
A → B
A → C
B → D
C → D
E → F



a-algorithm continued...

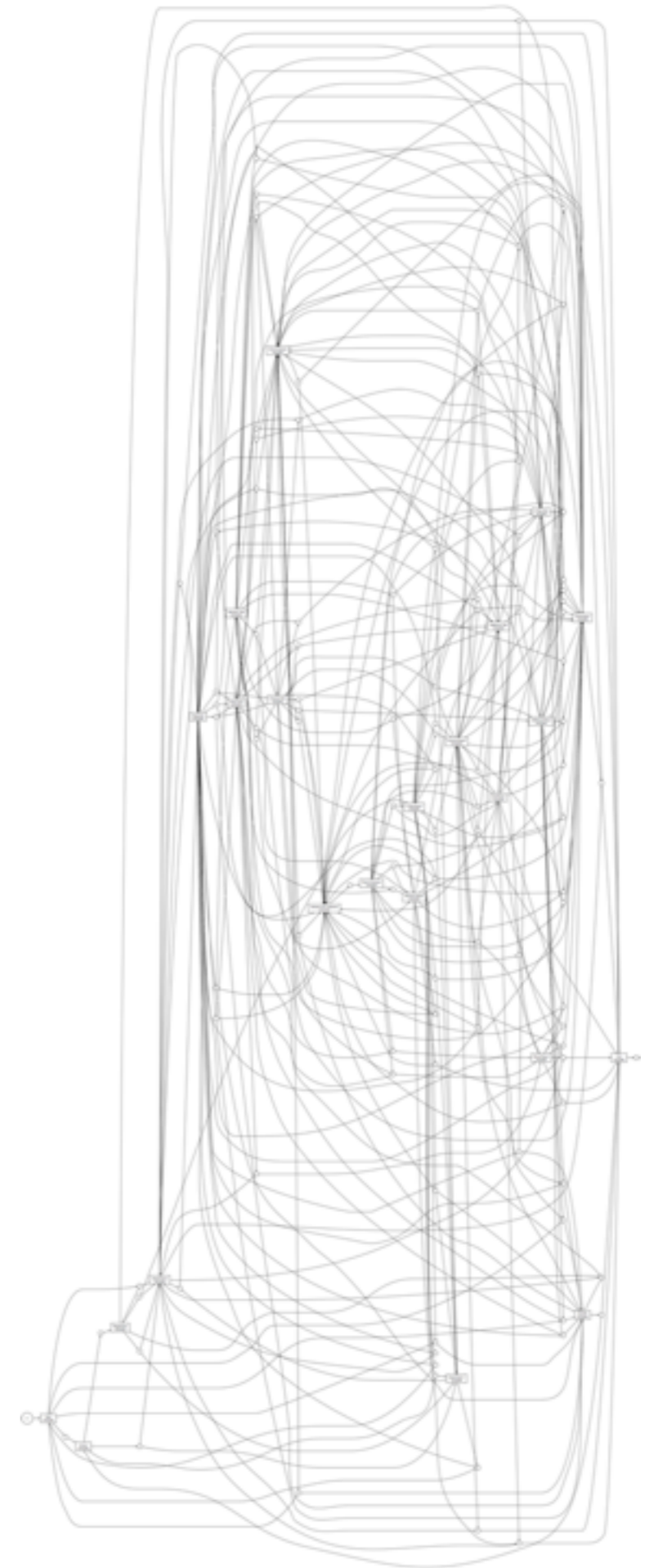
ABCD
ACBD
EF

B C	A → B
C B	A → C
	B → D
	C → D
	E → F

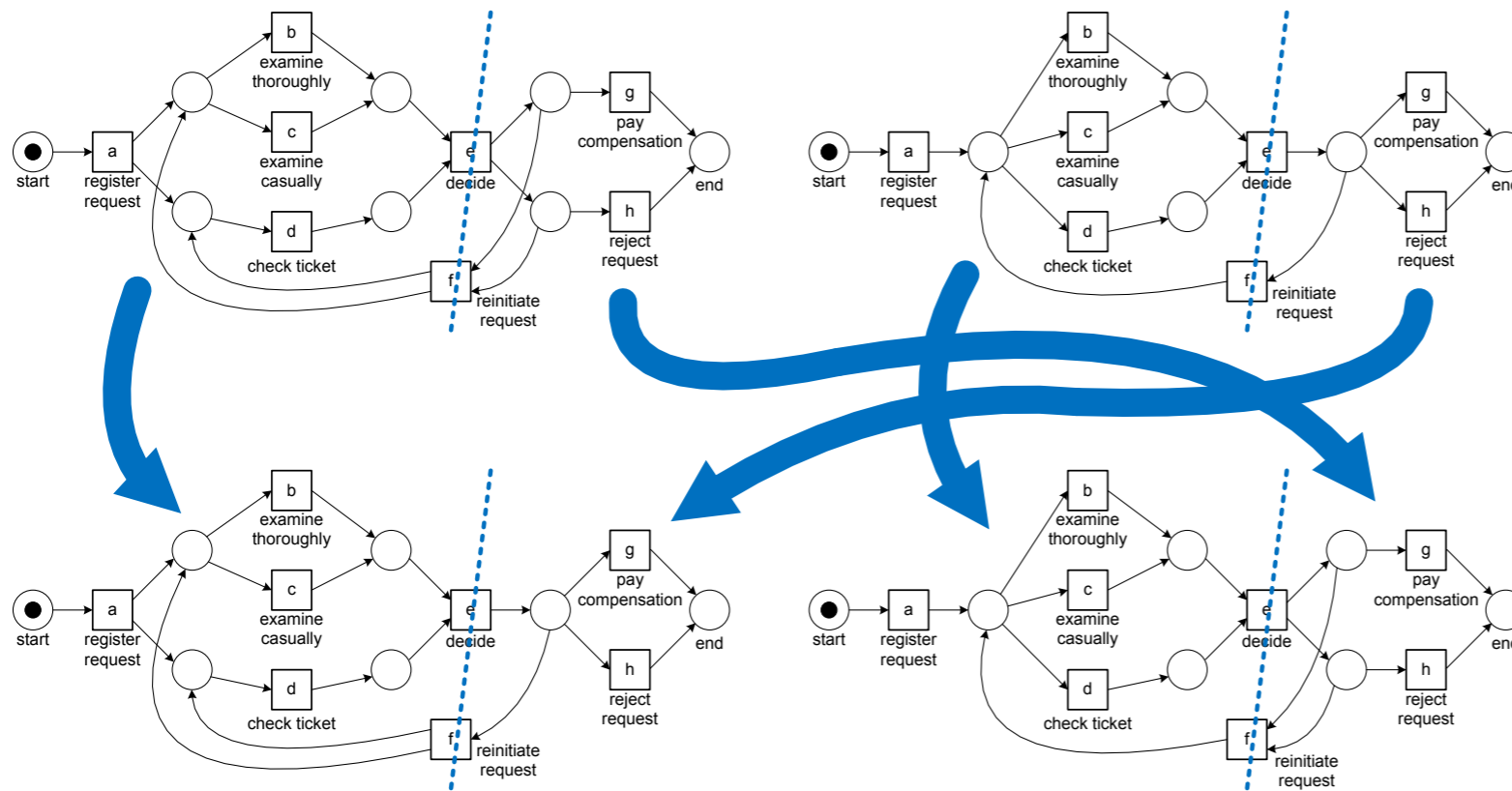


But...

- Spaghetti (needlessly complex)
- Limitations about
 - Short-loops
 - Non-local dependencies
- Non robust
- Big computational time

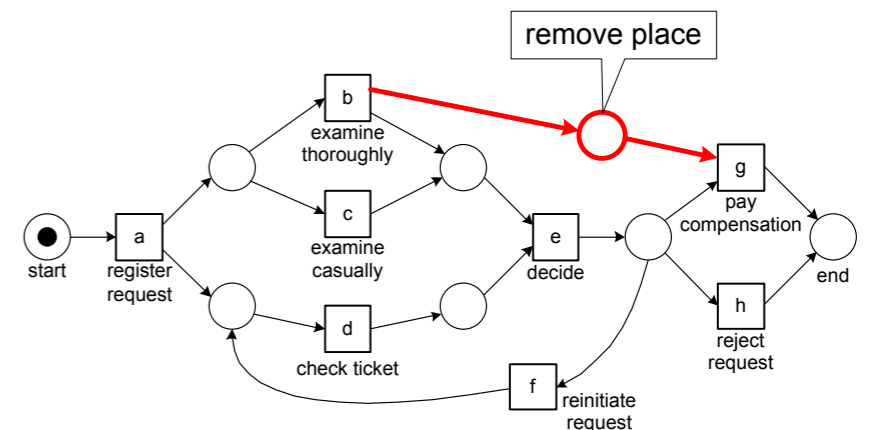


A global approach: Genetics



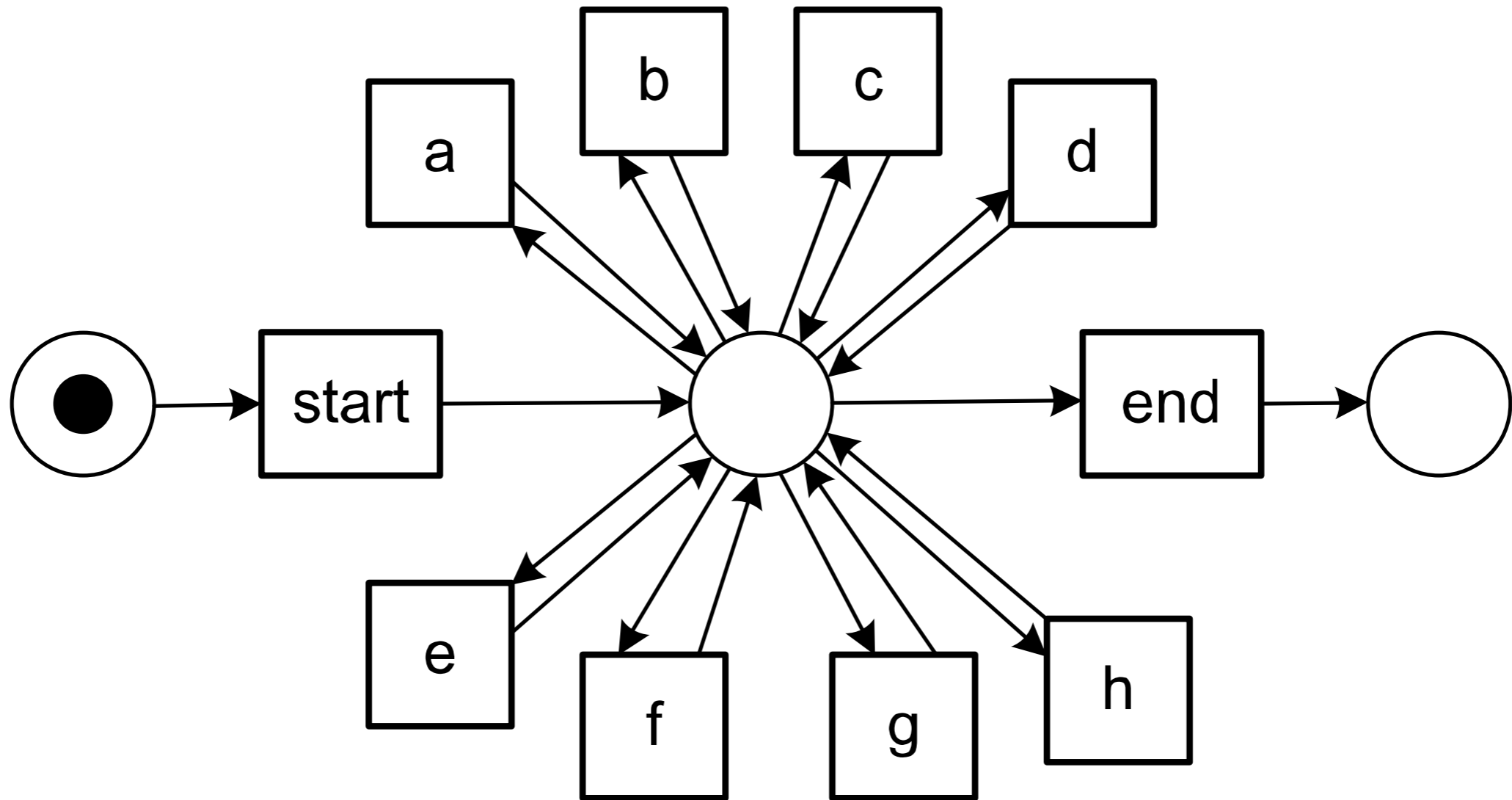
Crossover

Mutation



Is fitness enough?

Is fitness enough?



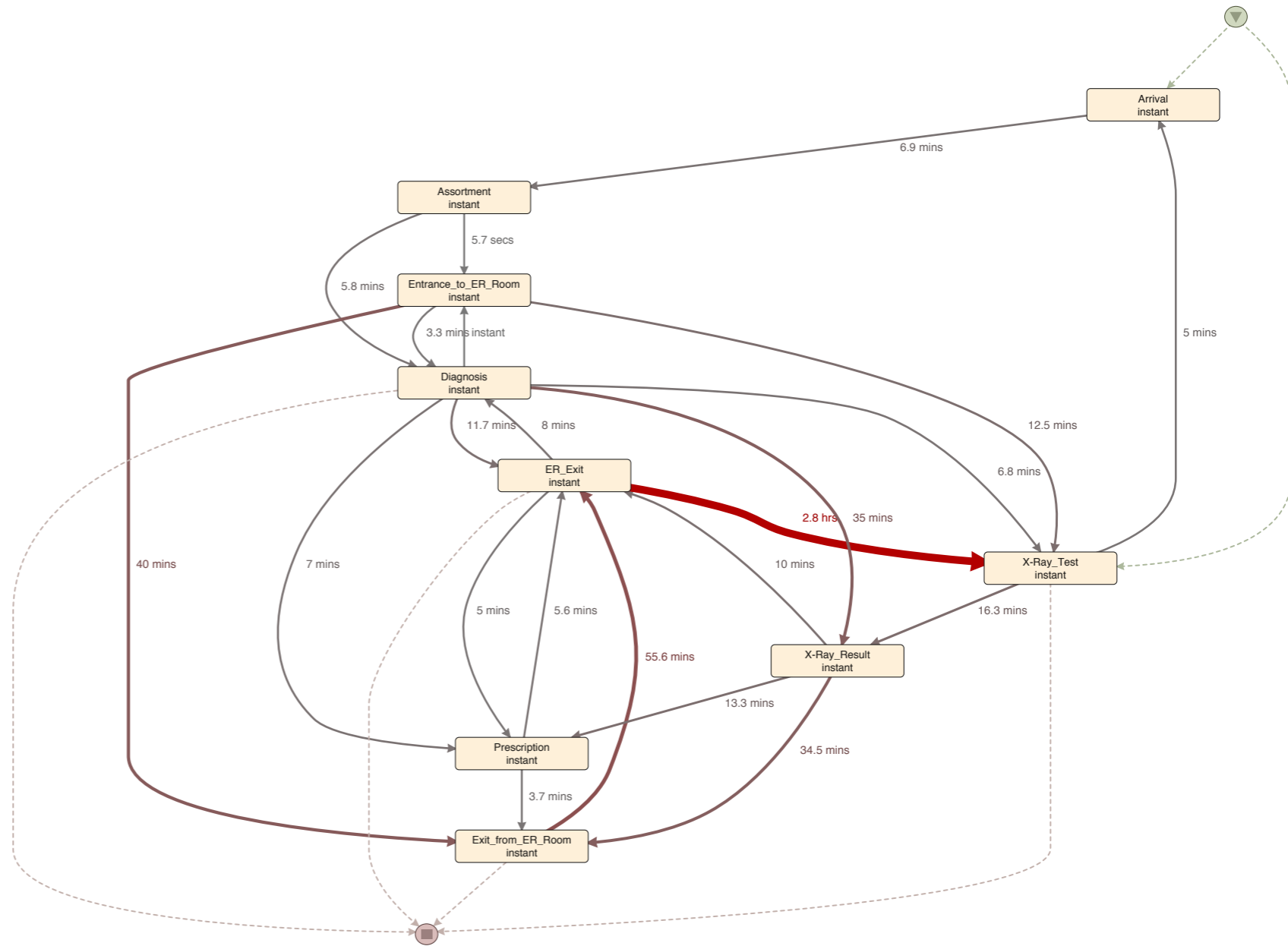
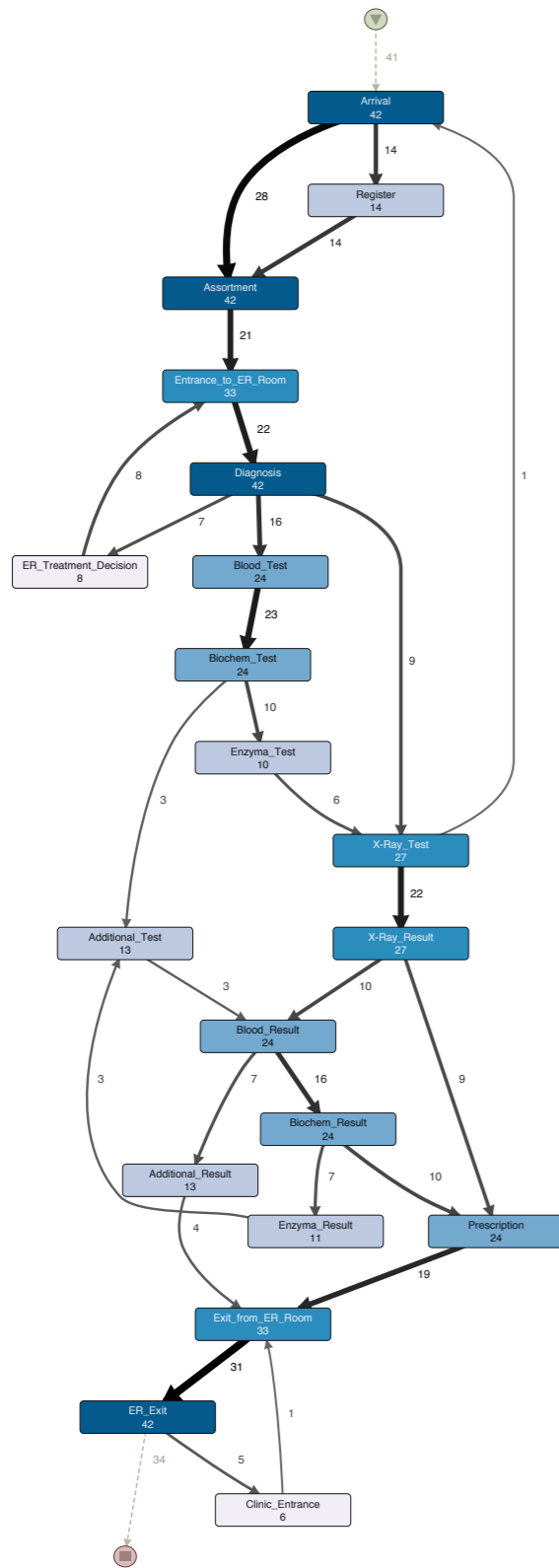
4 model quality criteria

- **Fitness** (be able to replay the observed behavior)
- **Precision** (do not allow too much additional behavior)
- **Generalization** (avoid overfitting)
- **Simplicity** (do not increase, beyond what is necessary, the number of entities required to explain the behavior)

Discovery Algorithms

- Heuristic Mining
- Genetic Mining
- Stochastic Task Graphs
- α ++ algorithm
- ILP mining
- LTL mining
- Fuzzy mining
- Hidden Markov models
- State-based regions
- ETM genetic algorithm
- Multi-phase mining
- ...

Real good results...



Challenges

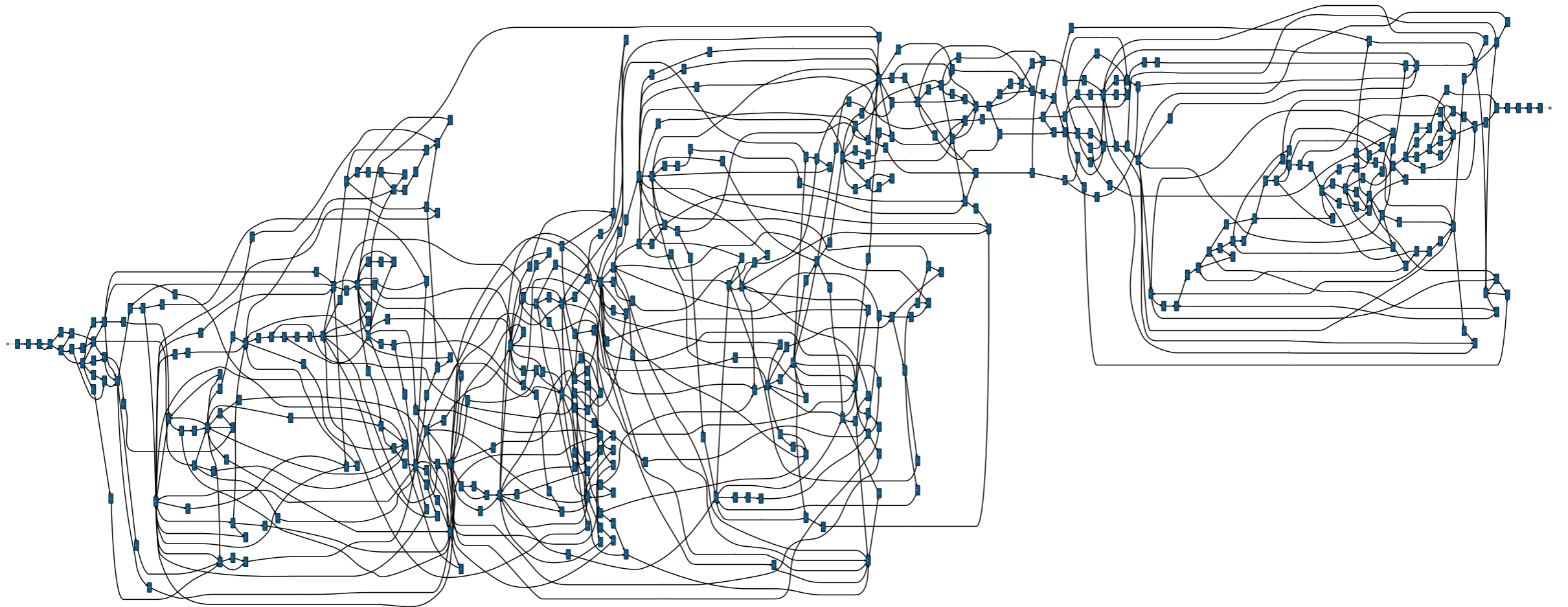
- The number of cases in the event log
- The average number of events per case
- The number of unique activities
- The number of distinct traces
- Multiple Perspectives / Context Awareness

Challenges

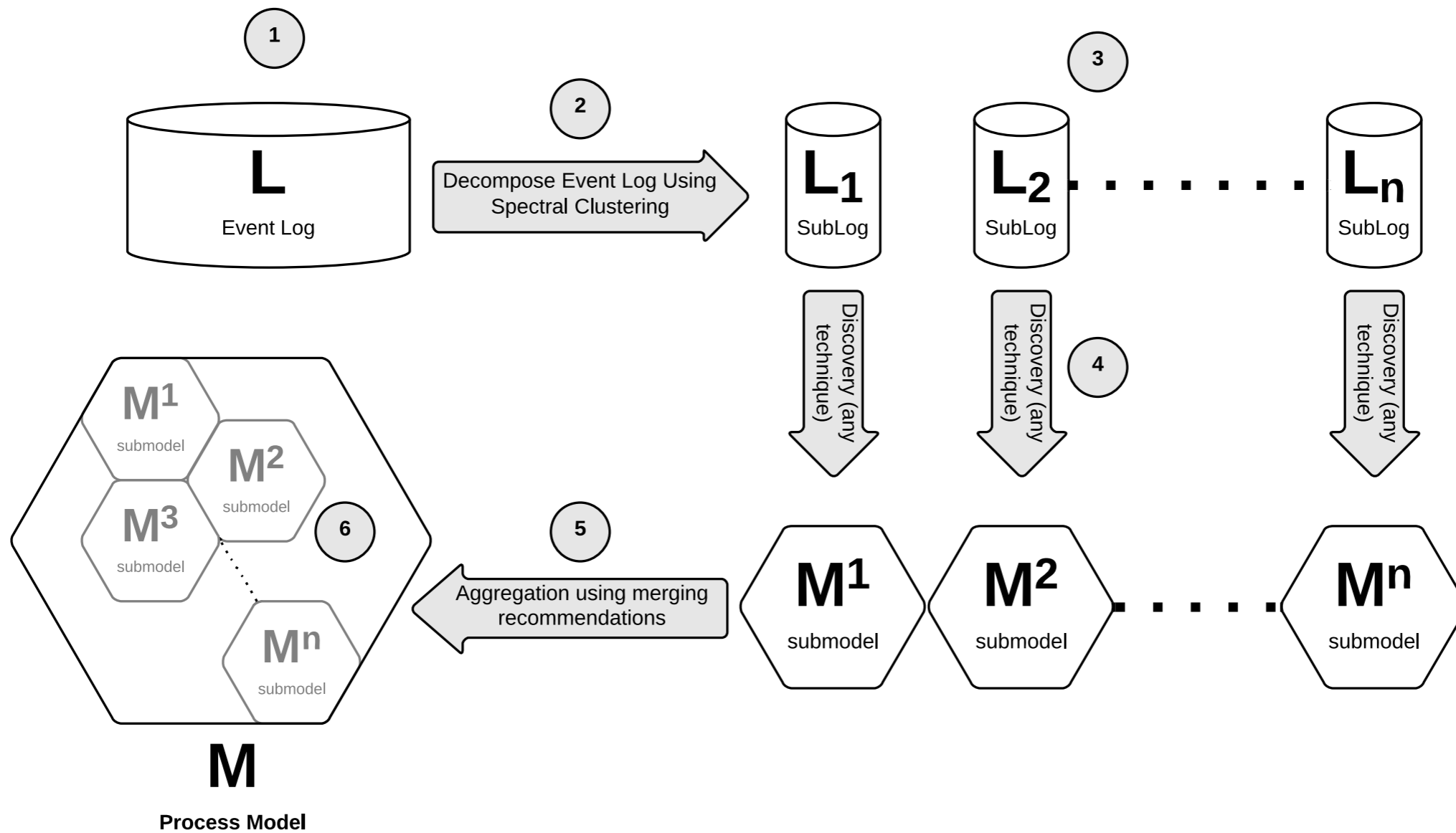
- The number of cases in the event log
- The average number of events per case
- The number of unique activities
- The number of distinct traces
- Multiple Perspectives / Context Awareness

The number of unique activities

- A matter of feasibility, performance & comprehension



Divide & Conquer



Graph partitioning to Divide & Conquer

- A connectivity metric

$$c_{ij} = \frac{\text{number of traces where } i \& j \text{ are directly connected}}{\text{total number of traces}}$$

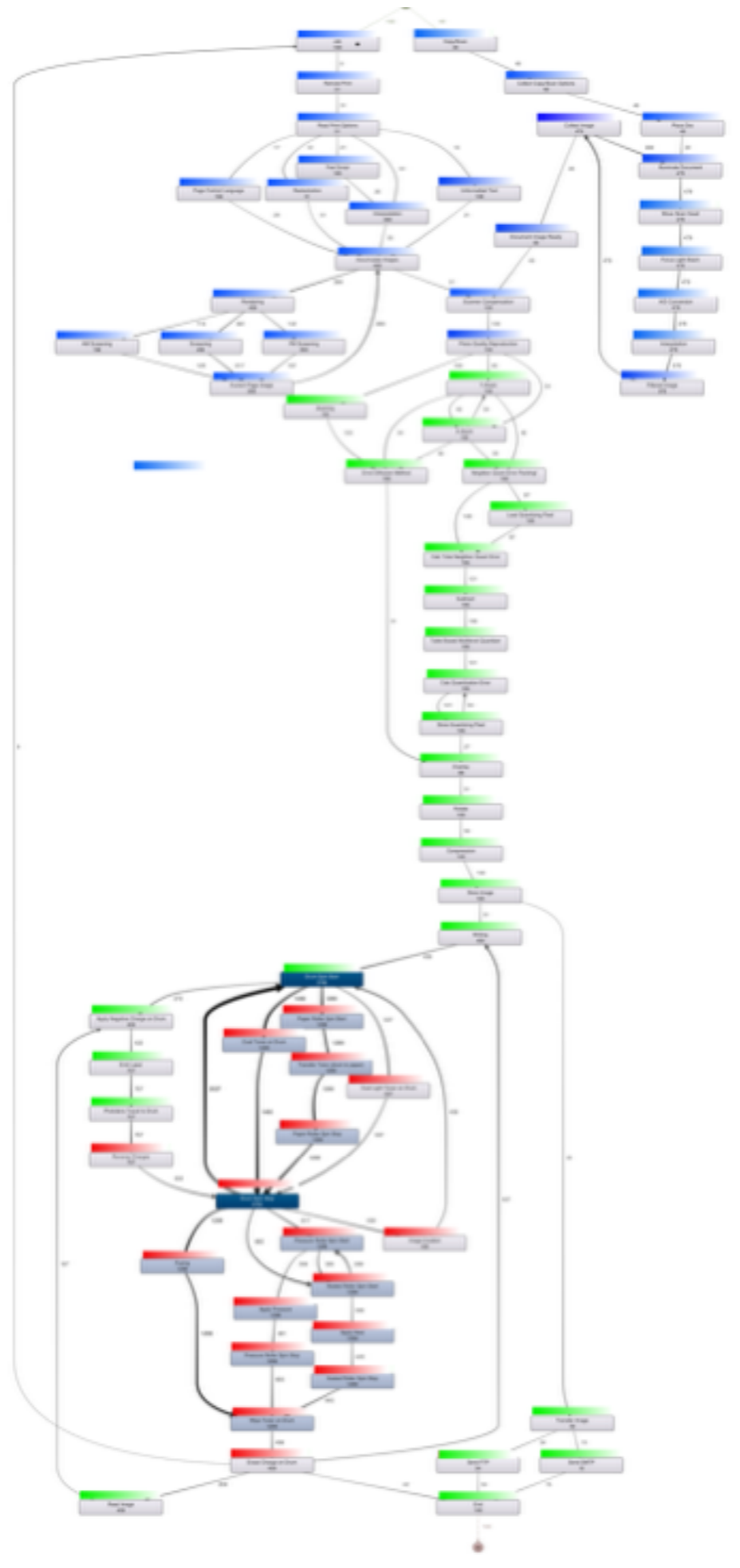
- Two optimization criteria
(Informativeness, Robustness)

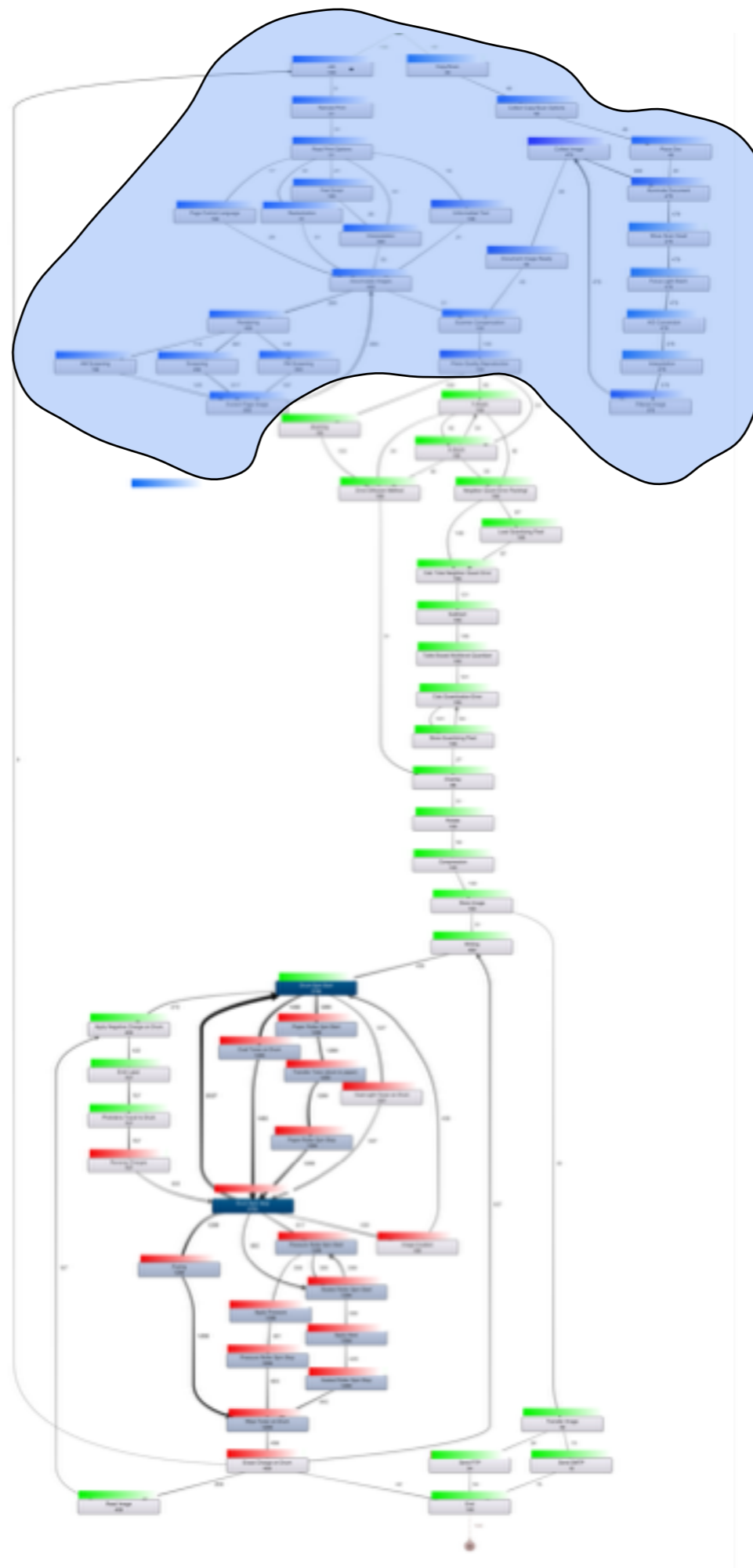
$$I_n = \frac{\sum_{i \in S_n, j \in S_n} c_{ij}}{\sum_{i \in S_n, j \in A} c_{ij}}$$

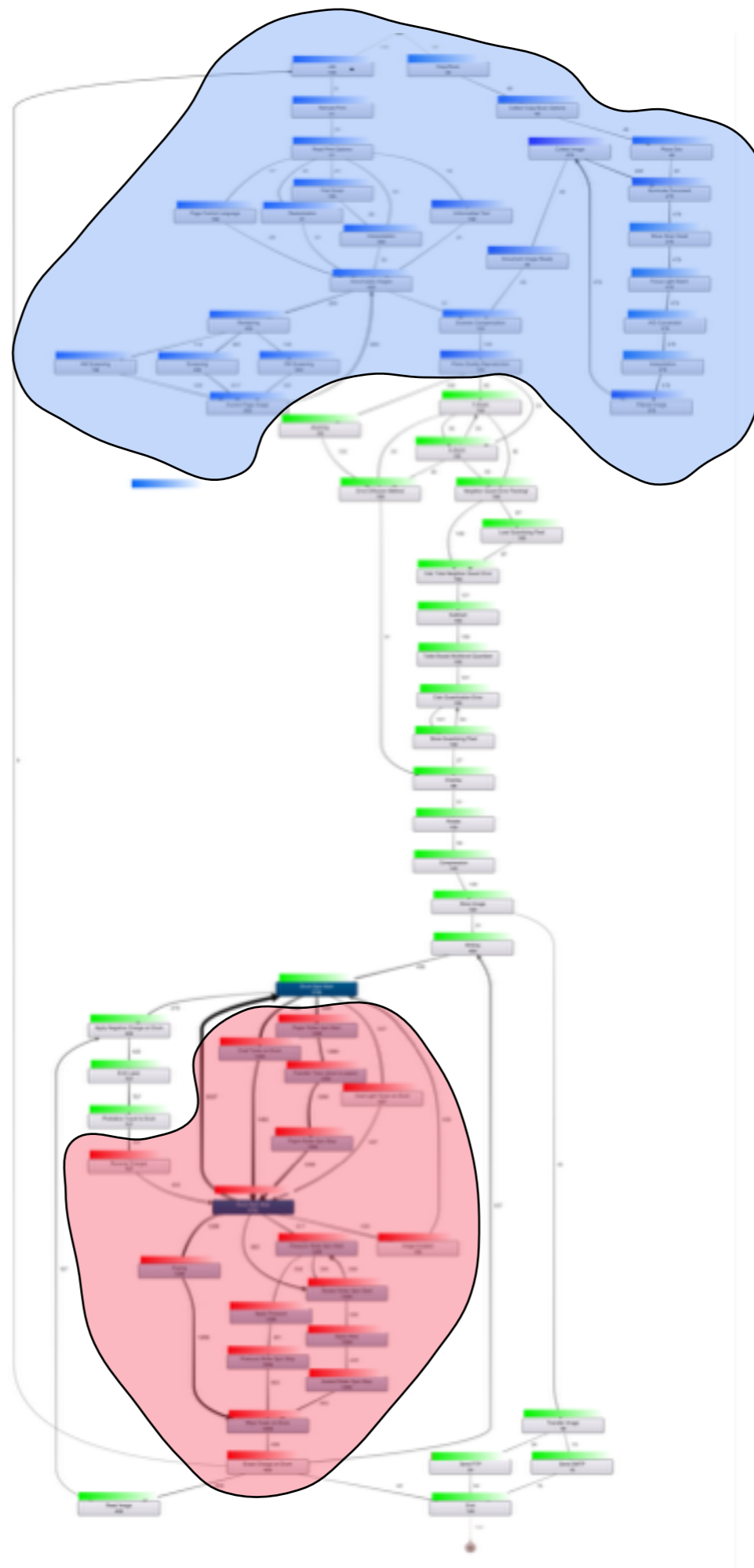
$$R_n = \frac{\sum_{i \in S_n, j \notin S_n} c_{ij}}{\sum_{i \in S_n, j \in A} c_{ij}}$$

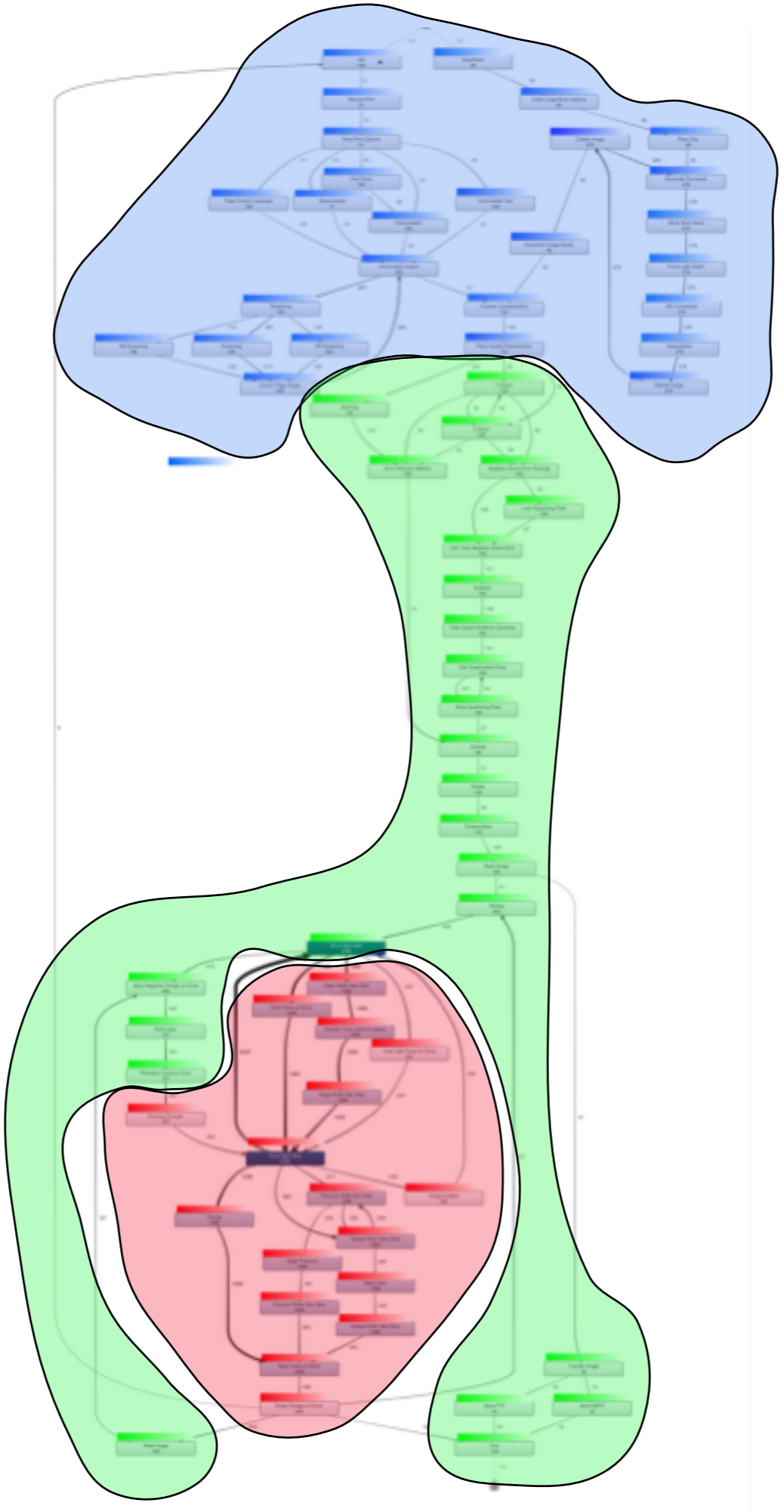
- Graph partitioning

$$\min R = N - \sum_{i=1}^N \lambda_i$$

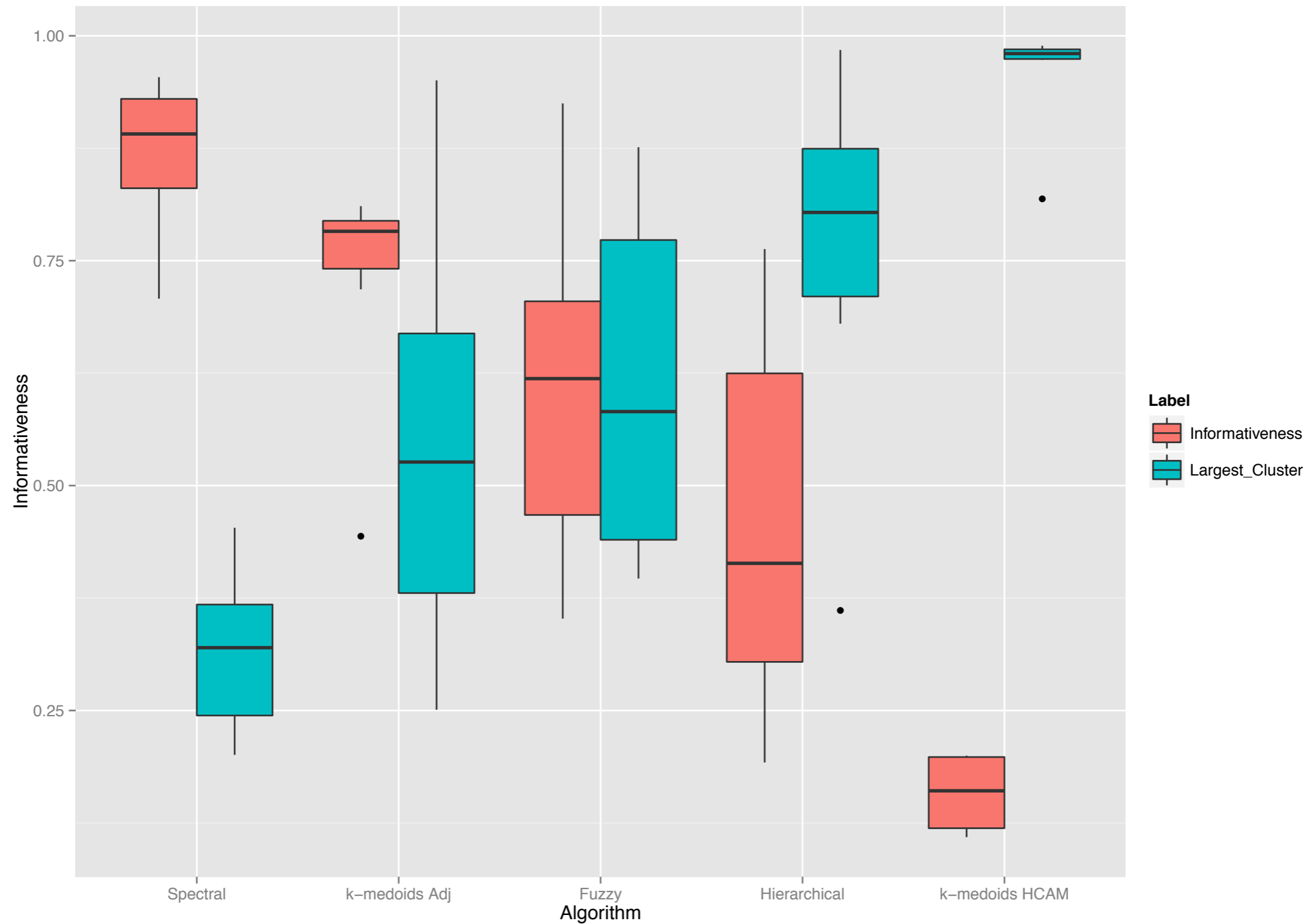








Encouraging results...



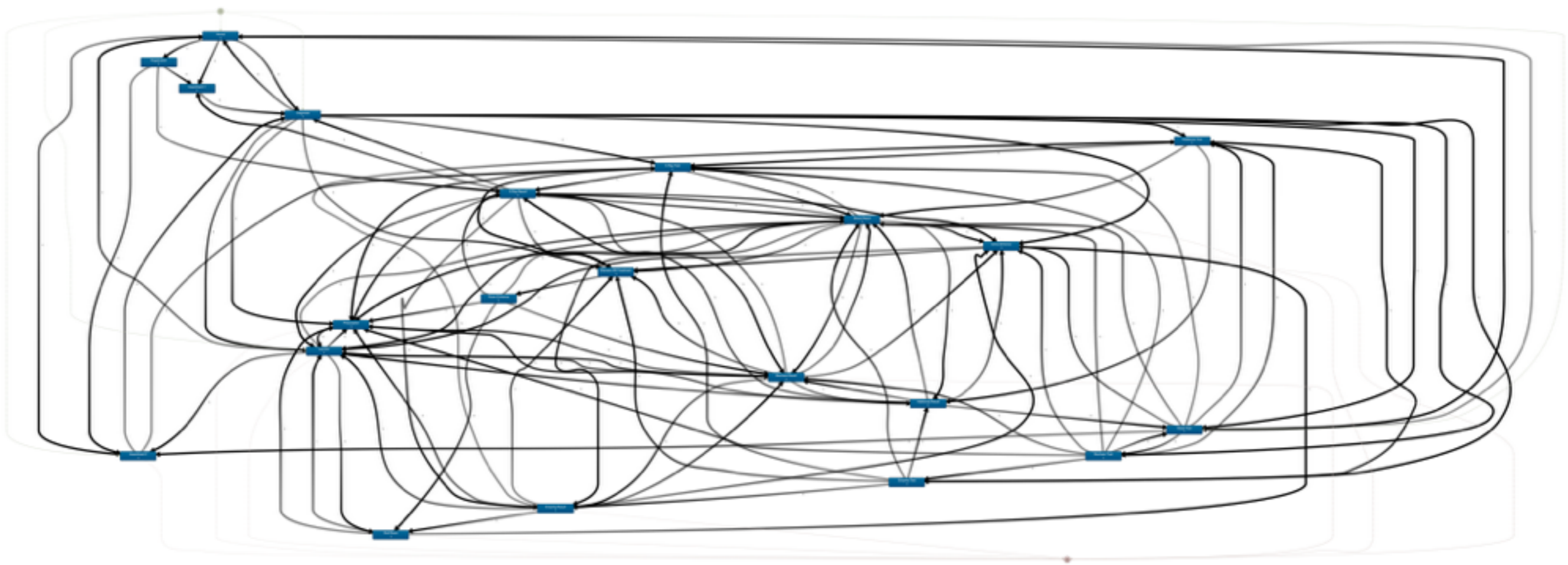
Challenges

- The number of cases in the event log
- The average number of events per case
- The number of unique activities
- The number of distinct traces
- Multiple Perspectives / Context Awareness

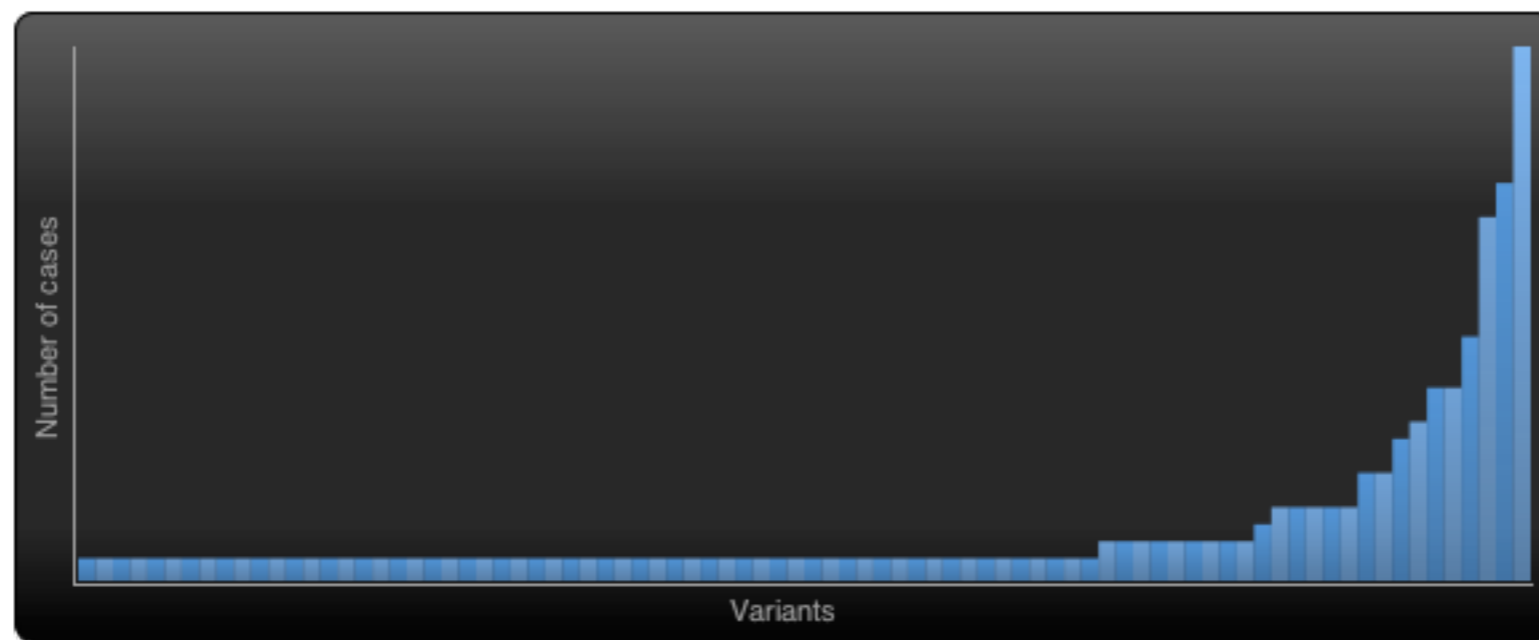
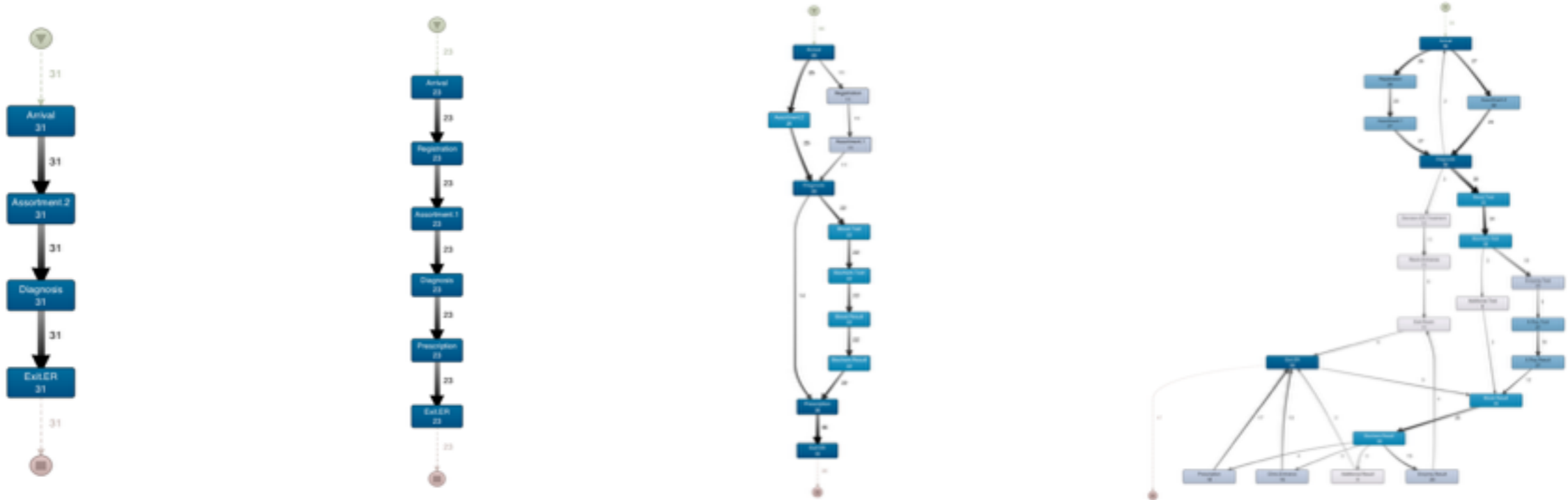
Challenges

- The number of cases in the event log
- The average number of events per case
- The number of unique activities
- The number of distinct traces
- Multiple Perspectives / Context Awareness

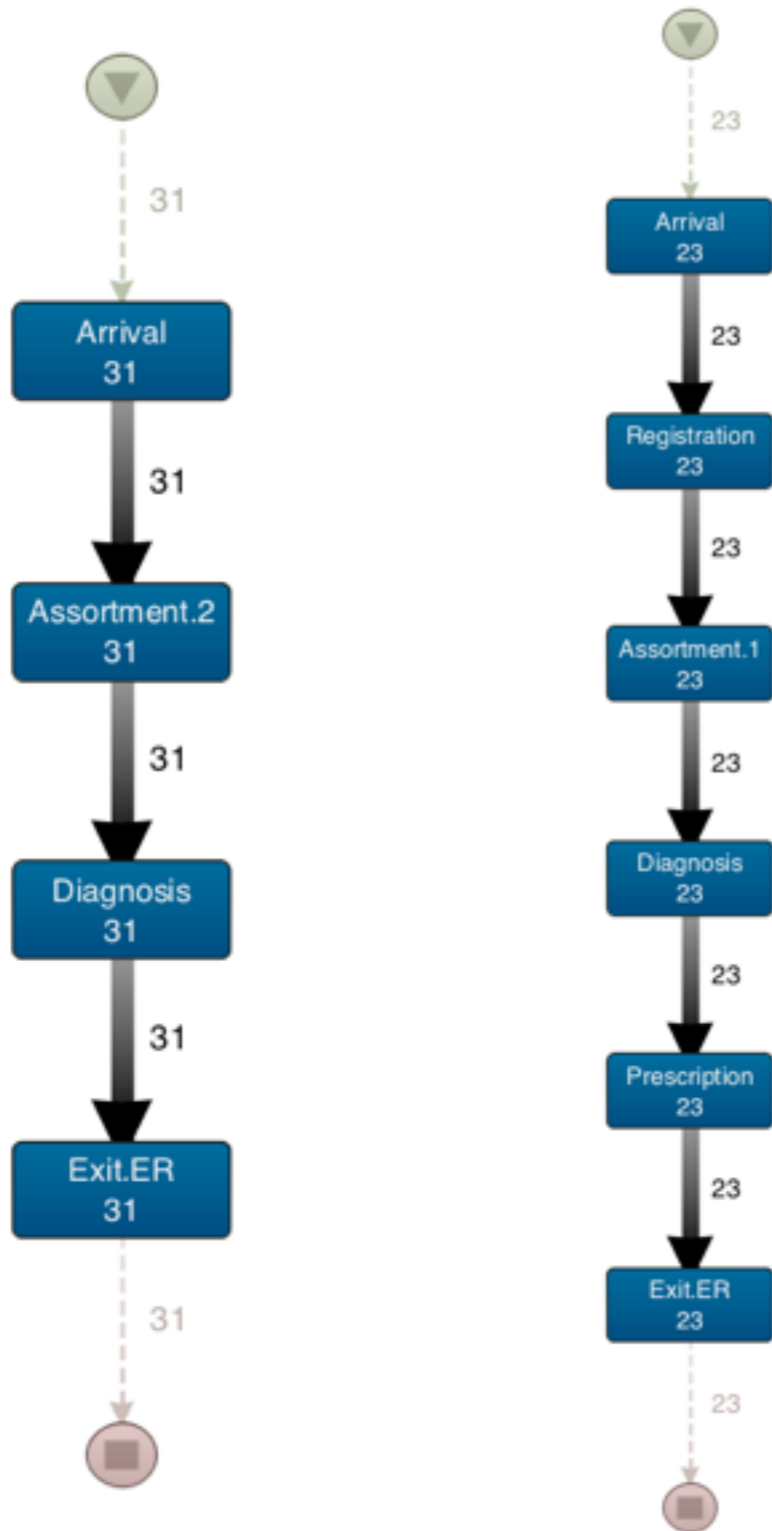
The number of distinct traces(variants)



Large Variability in Flow



Are Paths Similar?



$$sim_{activities}(T_i, T_j) = \frac{a(i) \cdot a(j)}{|a(i)| |a(j)|} = \frac{\sum_k a_k(i) \times a_k(j)}{\sqrt{\sum_k a_k(i)^2 \times \sum_k a_k(j)^2}}$$

$$sim_{transitions}(T_i, T_j) = \frac{t(i) \cdot t(j)}{|t(i)| |t(j)|} = \frac{\sum_k t_k(i) \times t_k(j)}{\sqrt{\sum_k t_k(i)^2 \times \sum_k t_k(j)^2}}$$

Spectral Clustering

- Degree Matrix D
- Laplacian matrix L
(unnormalized)
- Largest eigenvectors
matrix U
- Kmeans on U

$$d_i = \sum_j s_{ij}$$

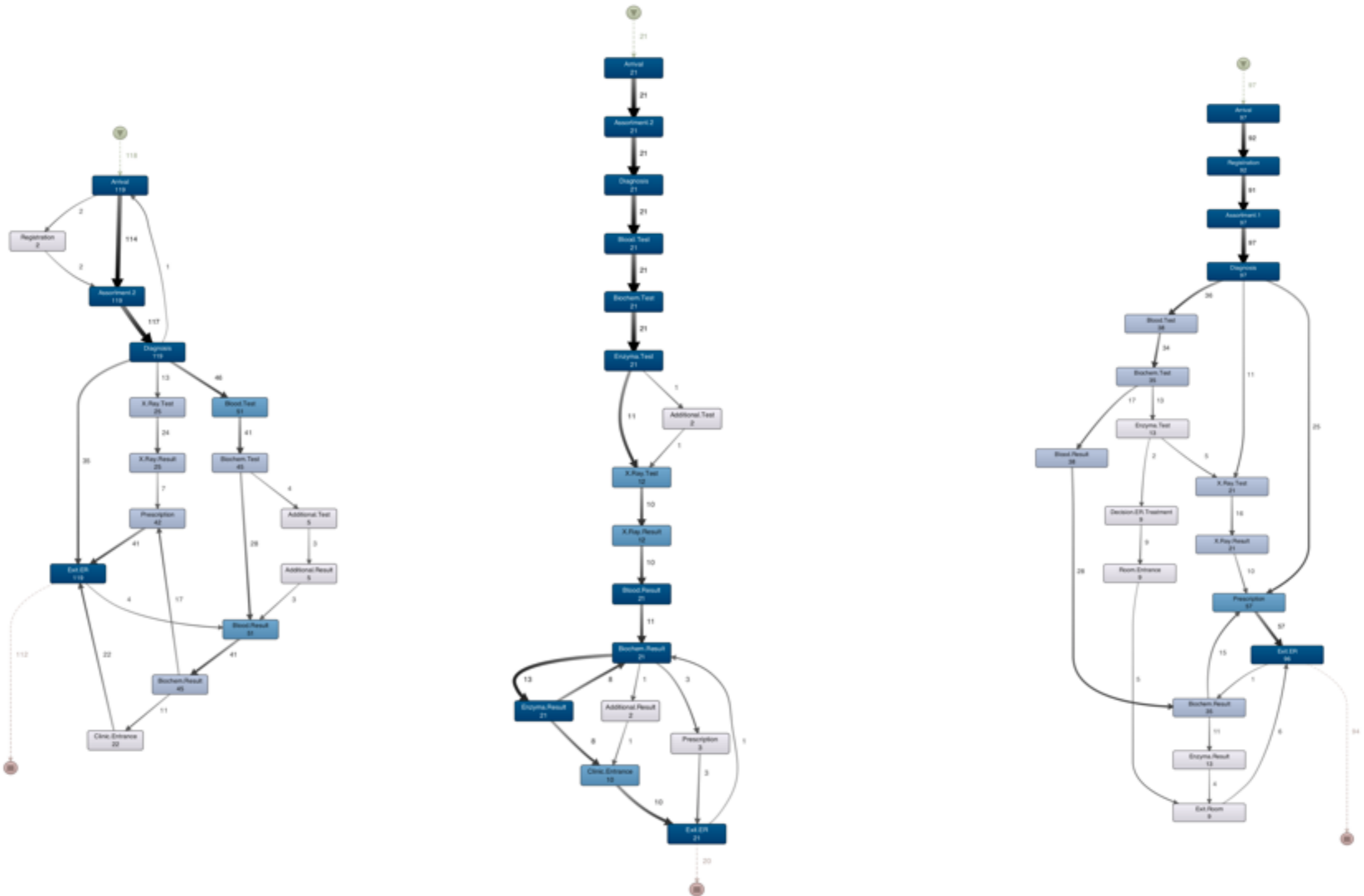
$$D = \text{diag}\{d_1, \dots, d_n\}$$

$$L = D - S$$

$$Lu = \lambda Du$$

$$U \in \mathbb{R}^{n \times k}$$

Now, it Makes More Sense



Why is this Important? (Hospital Case study)

- In retrospect, we were able to correlate the patients' clusters with their triage.
- Communicate the parameters of operations management to doctors
- Provide interesting insights (e.g. economic crisis and night visitors)

Challenges

- The number of cases in the event log
- The average number of events per case
- The number of unique activities
- The number of distinct traces
- Multiple Perspectives / Context Awareness

Challenges

- The number of cases in the event log
- The average number of events per case
- The number of unique activities
- The number of distinct traces
- Multiple Perspectives / Context Awareness

The Social Perspective - Volvo Case

“...restoring a customer’s normal service operation as quickly as possible when incidents arise ensuring that the best possible levels of service quality and availability are maintained.”

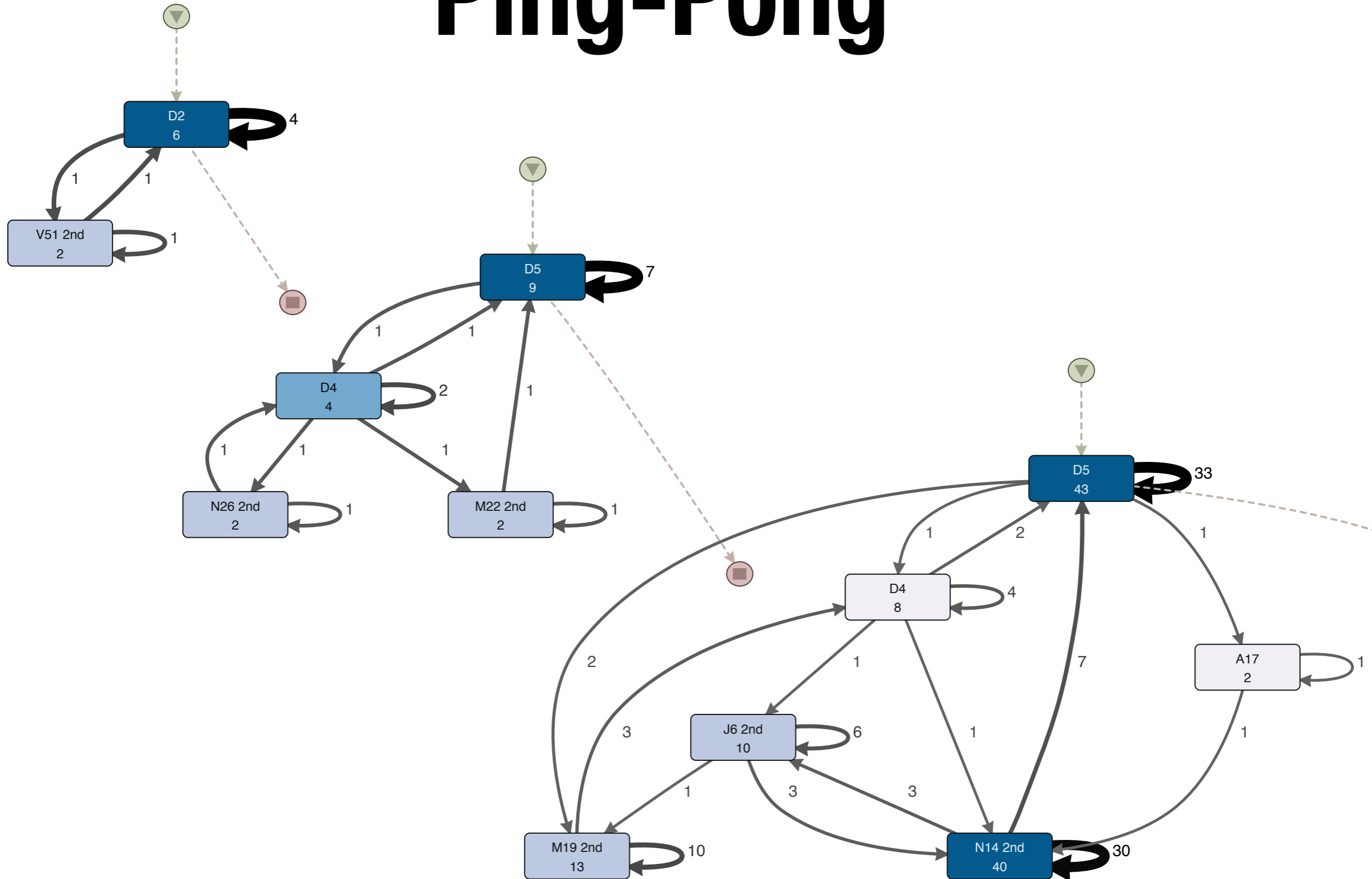


Social Patterns

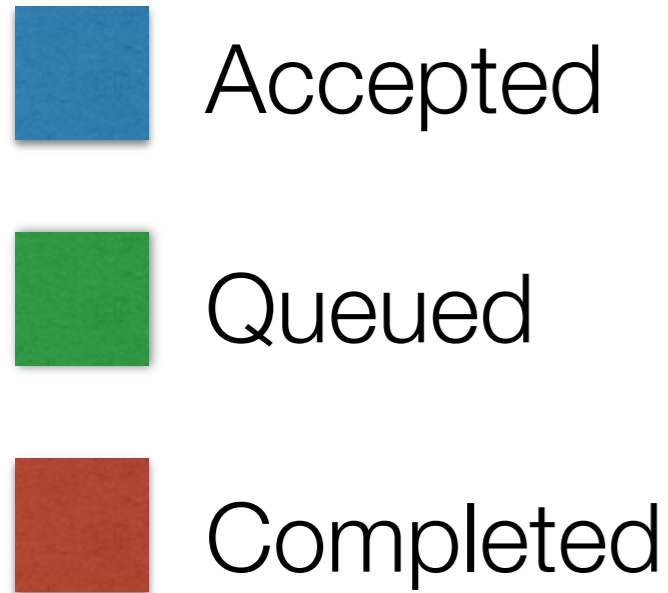
- **Push to Front** (1st line can resolve the service request alone)
- **Ping-Pong** (teams send the same case to each other again and again)



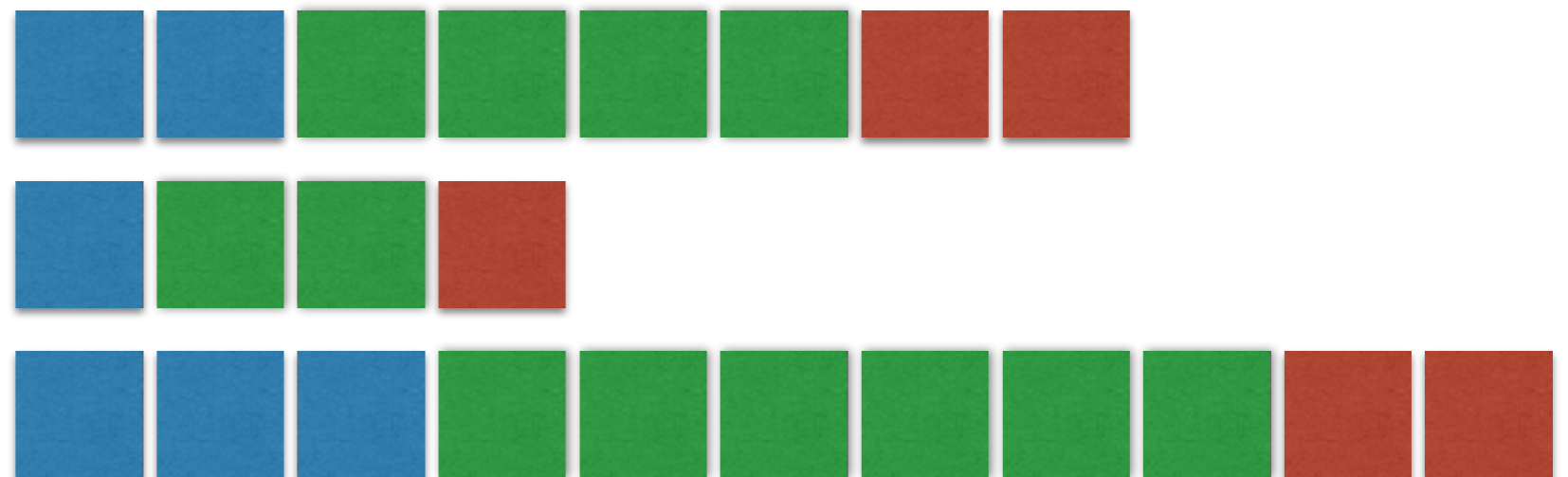
Ping-Pong

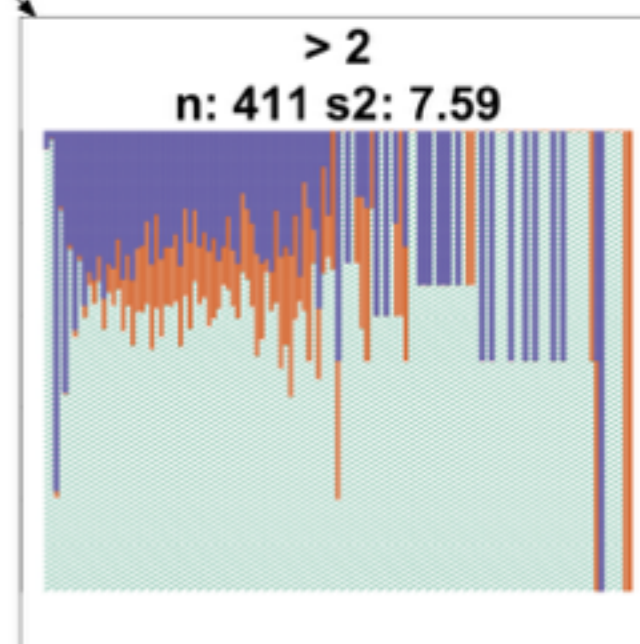
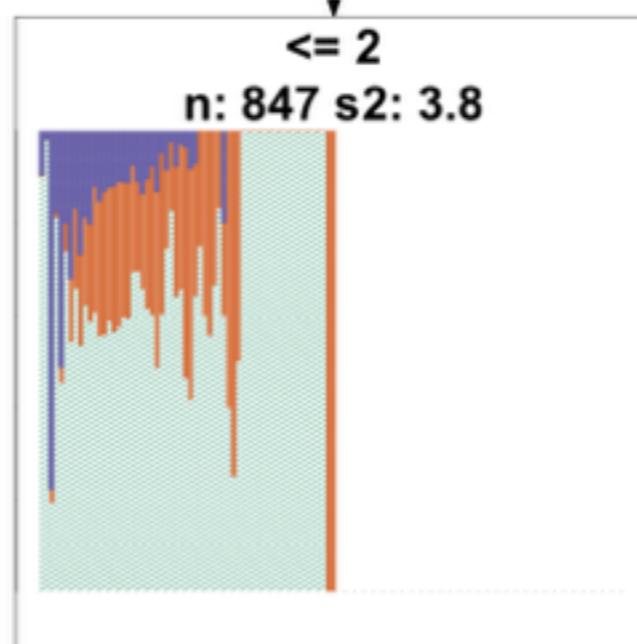
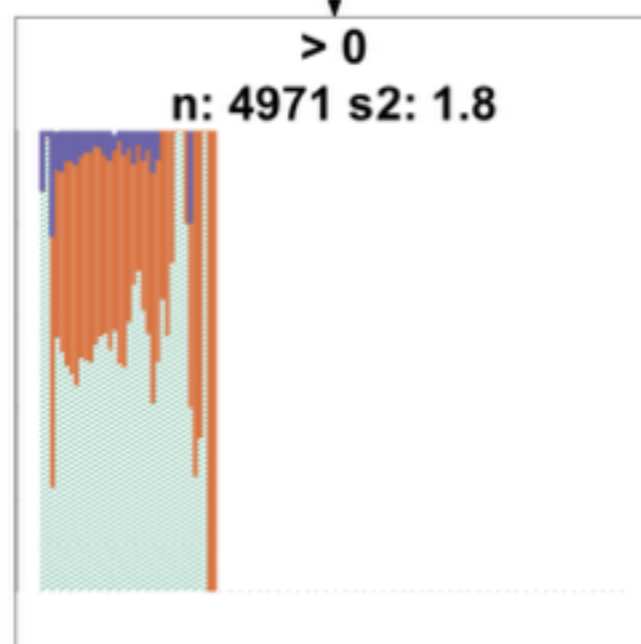
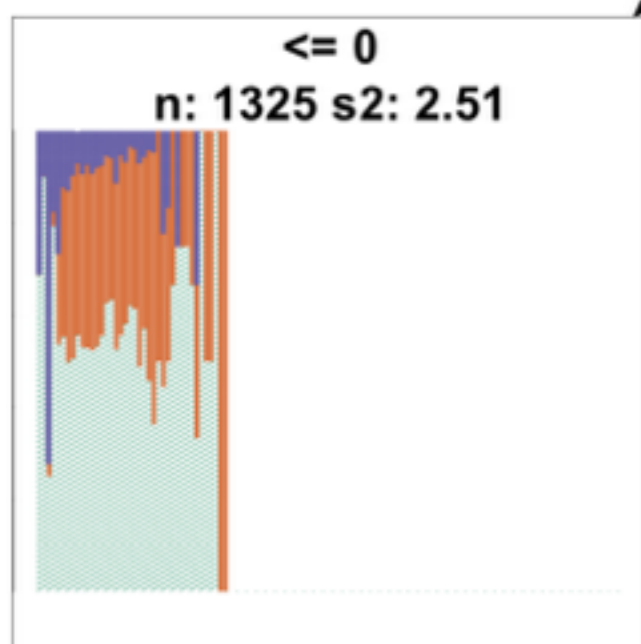
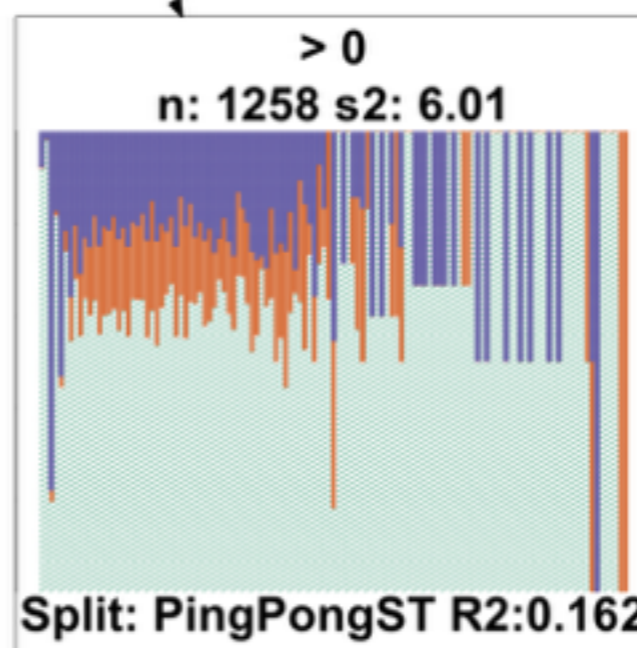
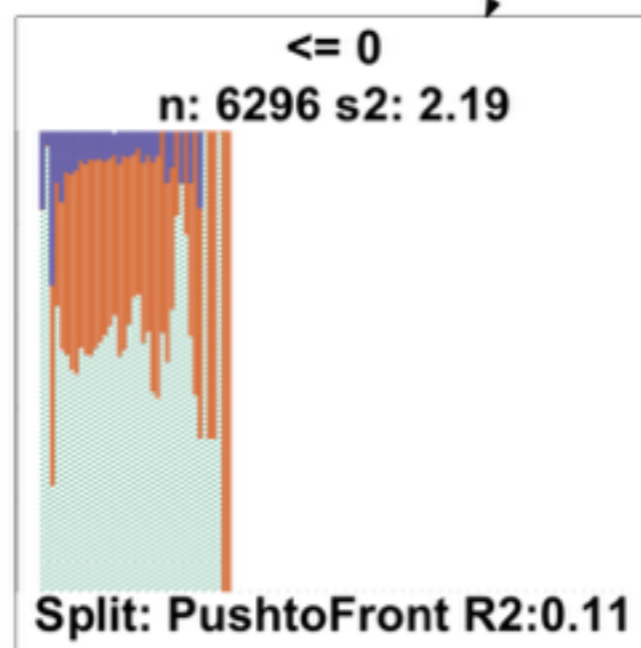
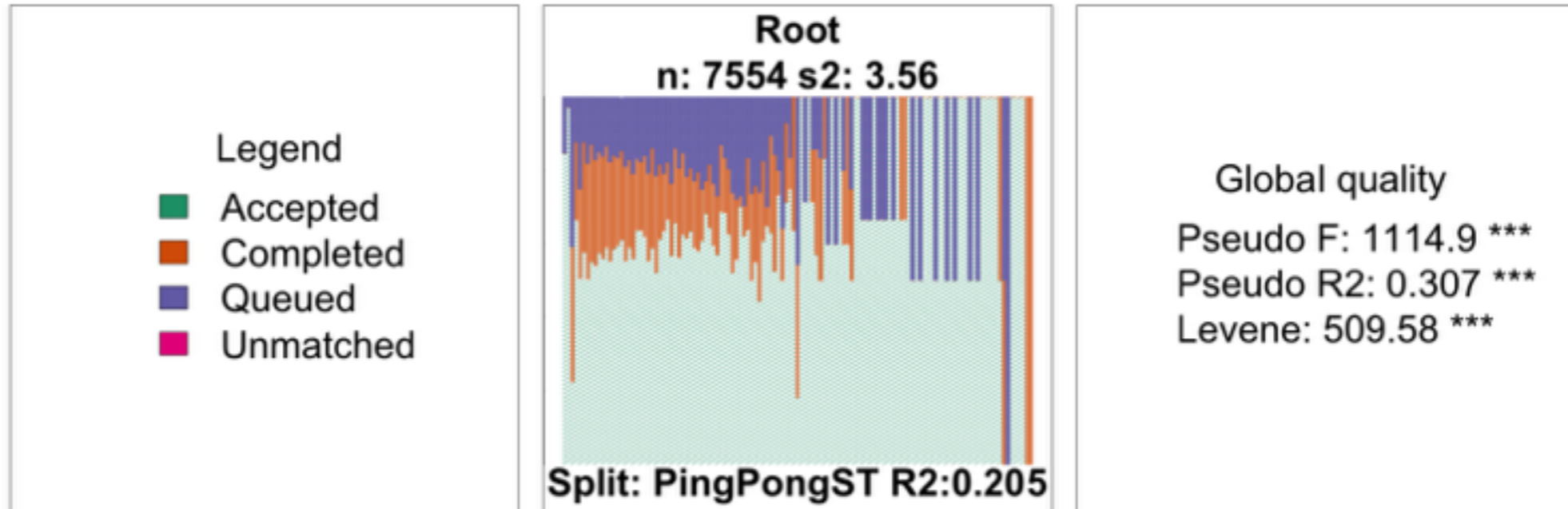


Volvo Case Revisited - Lifecycle of an incident



Cases Visualisation





Open Issues (Discovery)

- Computationally efficient process discovery algorithms
- Evaluating process discovery algorithms
- Model Quality Criteria
- Balancing multiple criteria in Discovery & Conformance
- Supervised / Semi-supervised process discovery
- Mining Resource behavior - Social Networks

Open Issues (Other)

- Auditing - Diagnosis Analytics
- Visual Analytics for Process Mining
- Comparing / Merging Models
- Decomposing process mining problems
- Prediction - Recommendation for operational support (On-line PM)
- Concept Drift and context-aware process mining
- Trace Alignment (similar to sequence alignment in biology)

THANK YOU!

Dr. Pavlos Delias, EMATTECH, Greece
pdelias@teiemt.gr, [@PavlosDelias](https://twitter.com/PavlosDelias)
