University of Pisa

MSc in Computer Engineering

Process-driven Information Systems

LECTURE 20

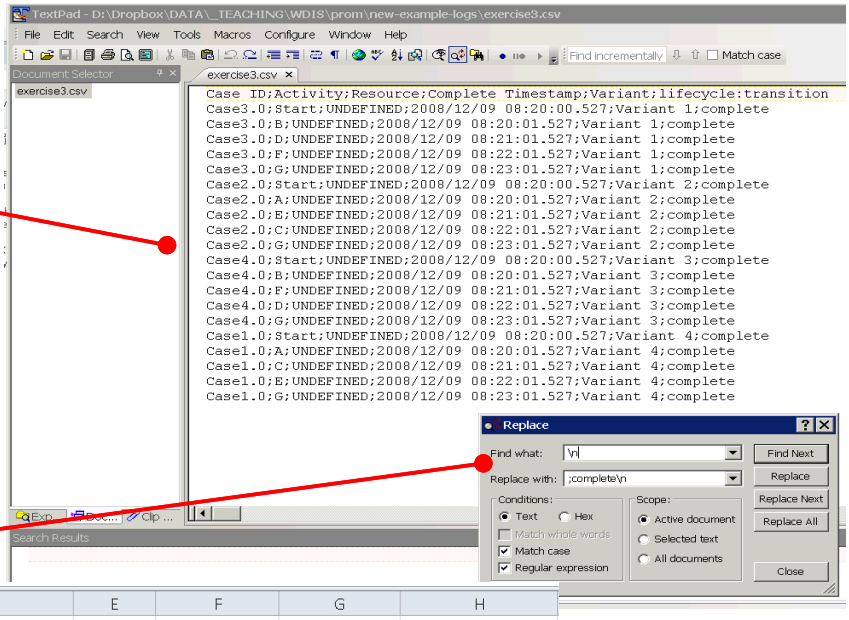
http://www.iet.unipi.it/m.cimino/wdis/

Mario G.C.A. Cimino

Department of Information Engineering

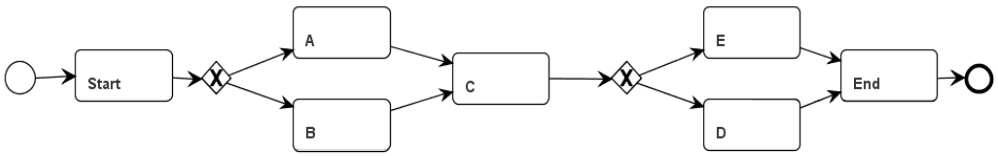
Introduction to Process Mining tools

- To manually edit the log, convert it from XES to CSV. Then edit it with Textpad or Excel and then use again Disco to convert it to XES.
- With Textpad regular expressions can be used



A	B	C	D	E	F	G	H
Case ID	Activity	Resource	Complete Timestamp	Variant	(case) creator	(case) variant	lifecycle:transition
Case2.0	Start	Start	2008/12/09 08:20:01.527	Variant 1	Fluxicon Disco	Variant 1	complete
Case2.0	B	UNDEFINED	2008/12/09 08:20:01.527	Variant 1	Fluxicon Disco	Variant 1	complete
Case2.0	C	UNDEFINED	2008/12/09 08:21:01.527	Variant 1	Fluxicon Disco	Variant 1	complete
Case2.0	E	UNDEFINED	2008/12/09 08:22:01.527	Variant 1	Fluxicon Disco	Variant 1	complete
Case2.0	End	End	2008/12/09 08:22:01.527	Variant 1	Fluxicon Disco	Variant 1	complete
Case1.0	Start	Start	2008/12/09 08:20:01.527	Variant 2	Fluxicon Disco	Variant 2	complete
Case1.0	A	UNDEFINED	2008/12/09 08:20:01.527	Variant 2	Fluxicon Disco	Variant 2	complete
Case1.0	C	UNDEFINED	2008/12/09 08:21:01.527	Variant 2	Fluxicon Disco	Variant 2	complete
Case1.0	D	UNDEFINED	2008/12/09 08:22:01.527	Variant 2	Fluxicon Disco	Variant 2	complete
Case1.0	End	End	2008/12/09 08:22:01.527	Variant 2	Fluxicon Disco	Variant 2	complete

- Exercise 2



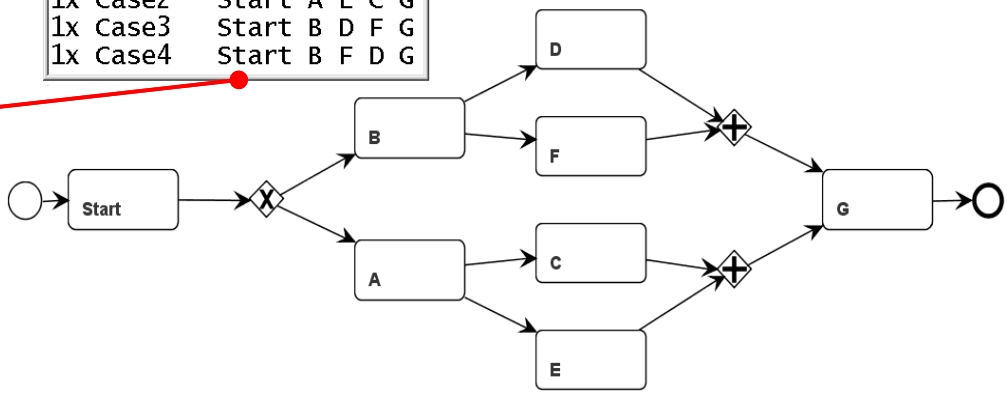
1x Case1	Start A C D End
1x Case2	Start B C E End

BPMN Analysis (Using causal net miner)

- Exercise 3 (the Start event must be added)

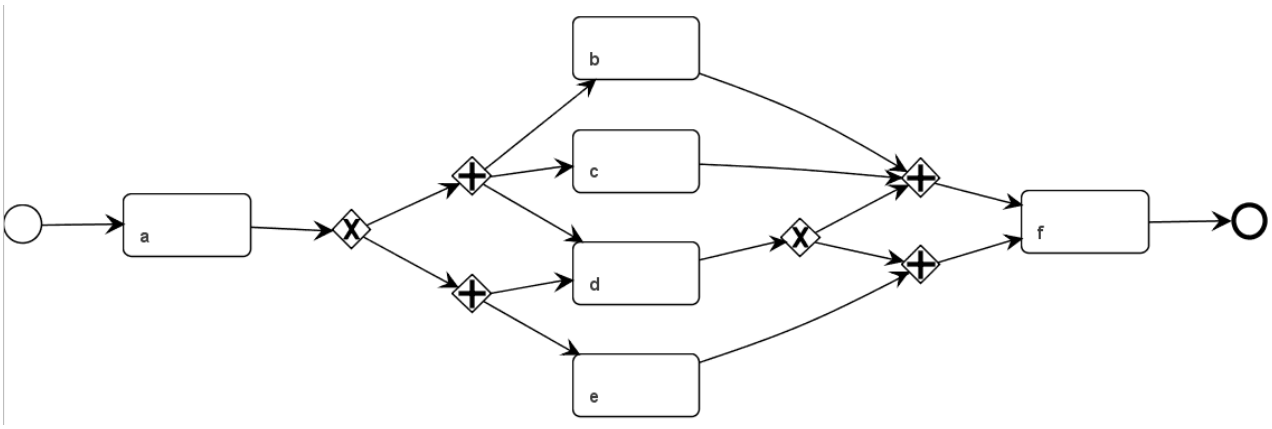
1x Case1	Start A C E G
1x Case2	Start A E C G
1x Case3	Start B D F G
1x Case4	Start B F D G

BPMN Analysis (Using causal net miner)



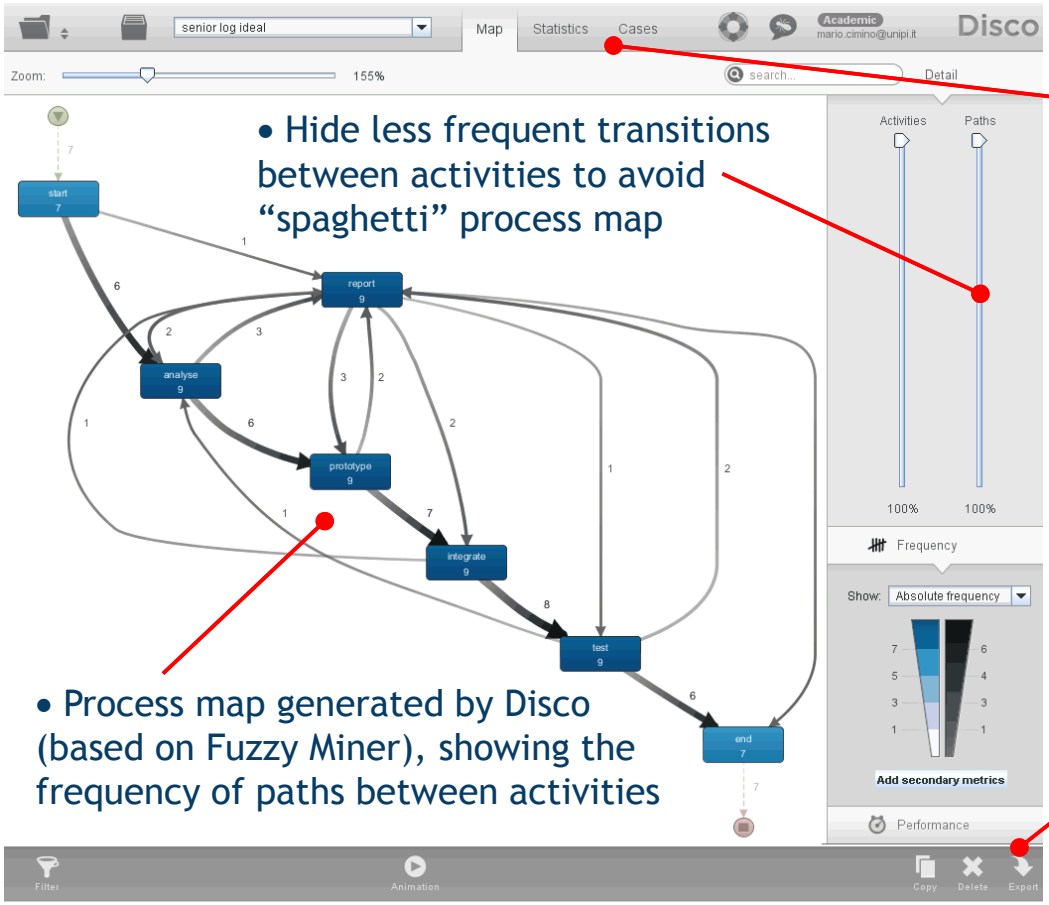
- Exercise 4

1x Case1	a b c d f
1x Case2	a c b d f
1x Case3	a b d c f
1x Case4	a c d b f
1x Case5	a d e f
1x Case6	a e d f



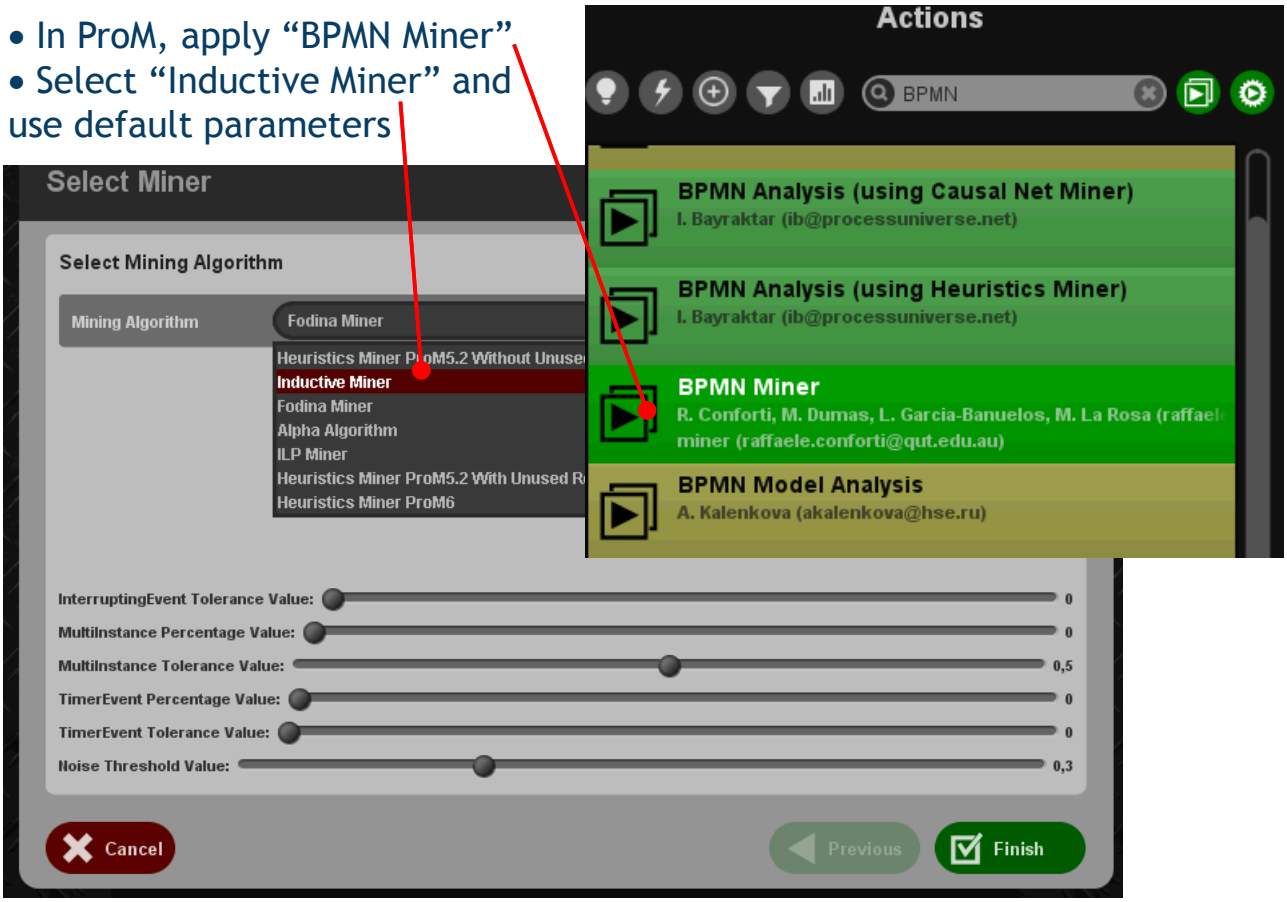
- SENIOR log (Software ENgineering for Input/Output pRoblems)
- It contains different cases of problem solving. Each case involves a worker improving the input/output (i/o) of a Java-based software application
- Synthetic log: cases without violations; the base pattern, iterated one or many times is $(a \rightarrow p \rightarrow i \rightarrow t) \parallel r$

Case ID	Timestamp	Activity
0	05/02/2015 00.00	**start**
0	05/02/2015 01.00	report
0	05/02/2015 02.00	analyse
0	05/02/2015 03.00	prototype
0	05/02/2015 04.00	integrate
0	05/02/2015 05.00	test
0	05/02/2015 06.00	end
1	05/02/2015 00.00	**start**
1	05/02/2015 01.00	analyse
1	05/02/2015 02.00	report
1	05/02/2015 03.00	prototype
1	05/02/2015 04.00	integrate
1	05/02/2015 05.00	test
1	05/02/2015 06.00	end
2	05/02/2015 00.00	**start**
2	05/02/2015 01.00	analyse
2	05/02/2015 02.00	prototype
2	05/02/2015 03.00	report
2	05/02/2015 04.00	integrate
2	05/02/2015 05.00	test
2	05/02/2015 06.00	end
3	05/02/2015 00.00	**start**
3	05/02/2015 01.00	analyse
3	05/02/2015 02.00	prototype
3	05/02/2015 03.00	integrate
3	05/02/2015 04.00	report
3	05/02/2015 05.00	test
3	05/02/2015 06.00	end
4	05/02/2015 00.00	**start**
4	05/02/2015 01.00	analyse
4	05/02/2015 02.00	prototype
4	05/02/2015 03.00	integrate
4	05/02/2015 04.00	test
4	05/02/2015 05.00	report
4	05/02/2015 06.00	end
5	05/02/2015 00.00	**start**
5	05/02/2015 01.00	analyse
5	05/02/2015 02.00	prototype
5	05/02/2015 03.00	integrate
5	05/02/2015 04.00	test
5	05/02/2015 05.00	report
5	05/02/2015 07.00	analyse
5	05/02/2015 08.00	report
5	05/02/2015 09.00	prototype
5	05/02/2015 10.00	integrate
5	05/02/2015 11.00	test
5	05/02/2015 12.00	end
6	05/02/2015 00.00	**start**
6	05/02/2015 01.00	analyse
6	05/02/2015 02.00	report
6	05/02/2015 03.00	prototype
6	05/02/2015 04.00	integrate
6	05/02/2015 05.00	test
6	05/02/2015 06.00	analyse
6	05/02/2015 07.00	prototype
6	05/02/2015 08.00	report
6	05/02/2015 09.00	integrate
6	05/02/2015 10.00	test
6	05/02/2015 11.00	end

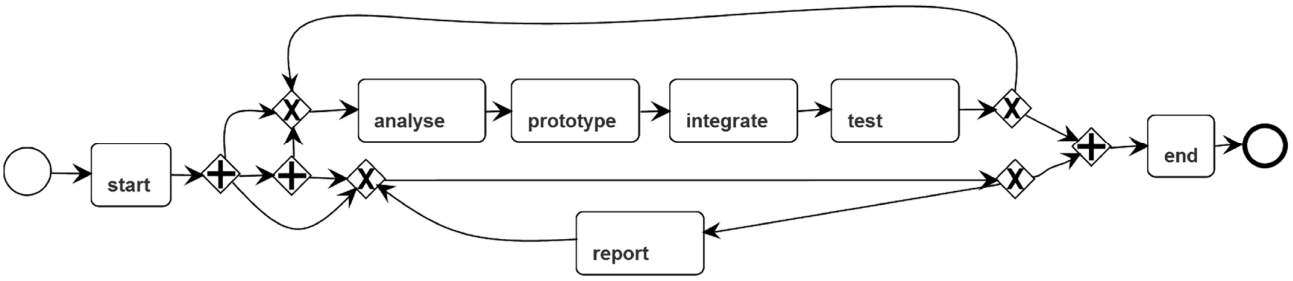


- Hide less frequent transitions between activities to avoid "spaghetti" process map
- See statistics and Cases
- Process map generated by Disco (based on Fuzzy Miner), showing the frequency of paths between activities
- Export data in a number of formats

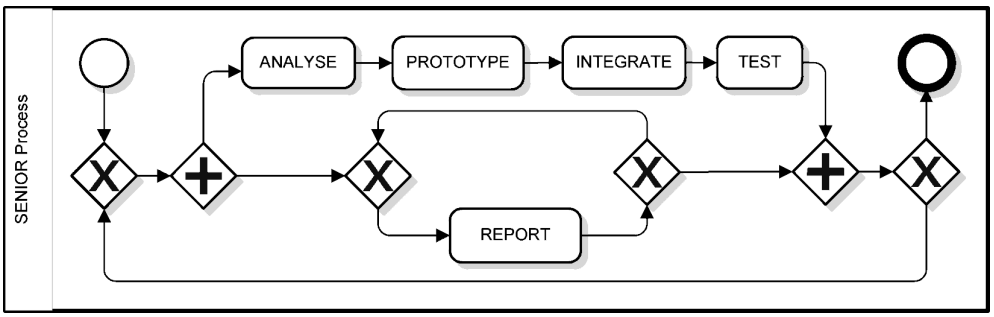
- In ProM, apply “BPMN Miner”
- Select “Inductive Miner” and use default parameters



- Model generated by the inductive miner algorithm



- Since there is no violation in the event log, the generated model is very similar to the normative process:



- In brief: the **Inductive Miner** aims to discover block-structured process models fitting the behavior represented in event log. IM partitions the activities, select the most important process constructs, splits the log and recurses until a base case is encountered.

- A *process tree* is the hierarchical representation of a block-structured workflow net. The leaves of the tree are activities, representing transitions. The nodes of the tree, *operators*, describe how their children are combined: exclusive choice ($\times$), sequential composition ($\rightarrow$), parallel composition ($\wedge$), and loop ($\circlearrowleft$).

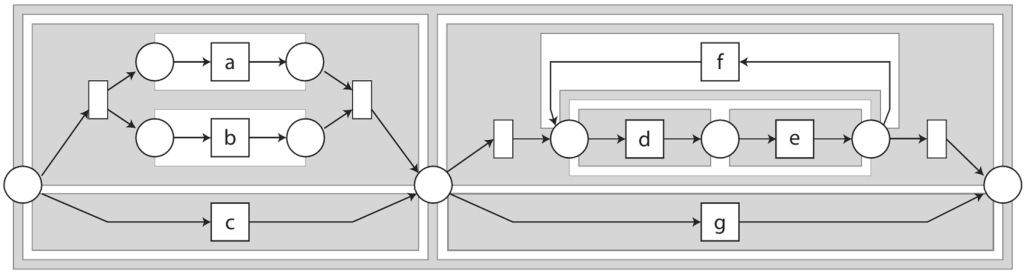
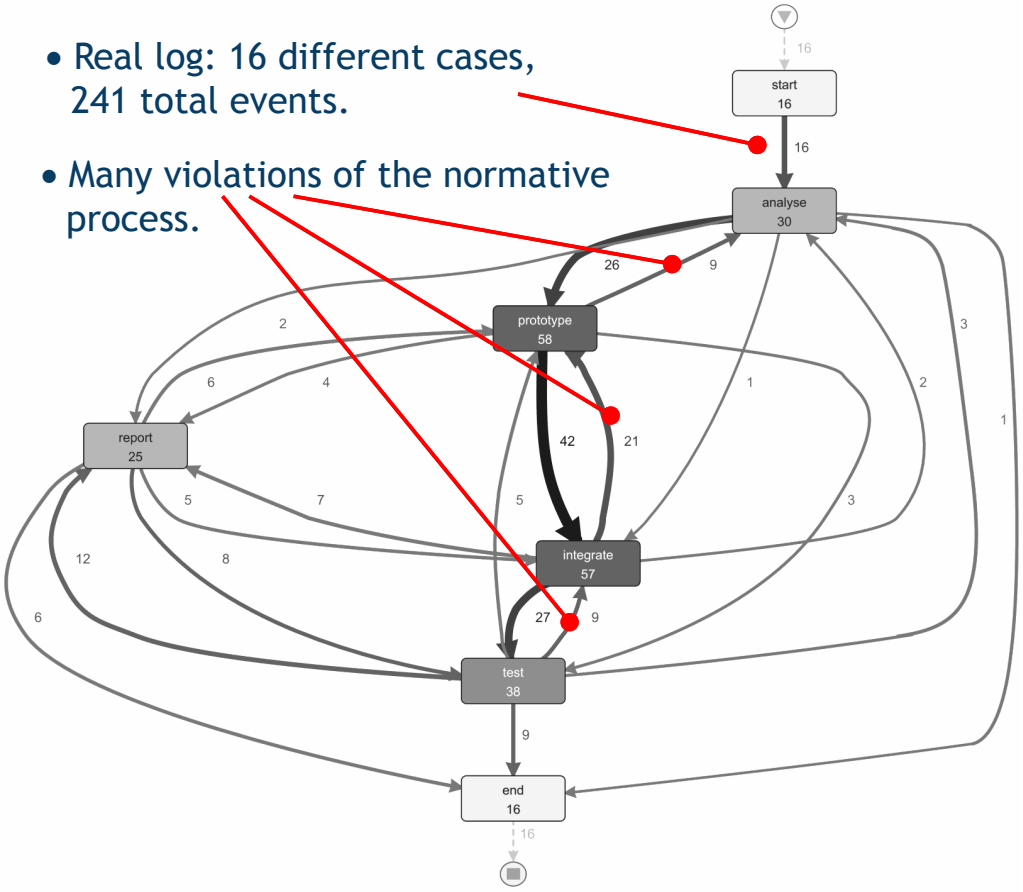
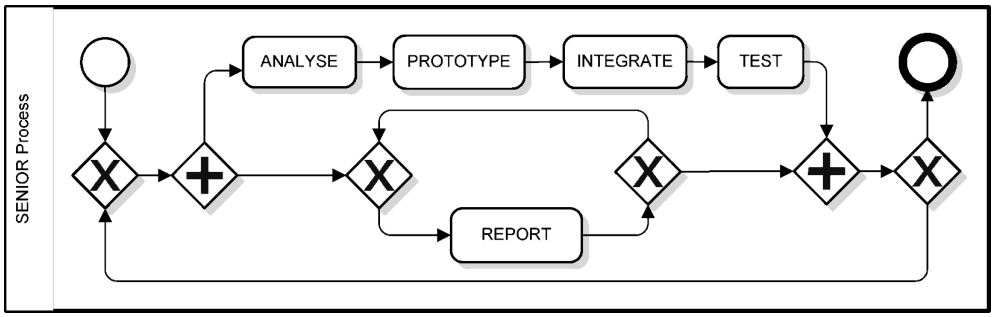


Figure: A block-structured workflow net M_E; filled regions denote the block-structure; process tree $\rightarrow(\times(\wedge(a,b),c),\times(\circlearrowleft(\rightarrow(d,e),f),g))$ corresponds to this net.

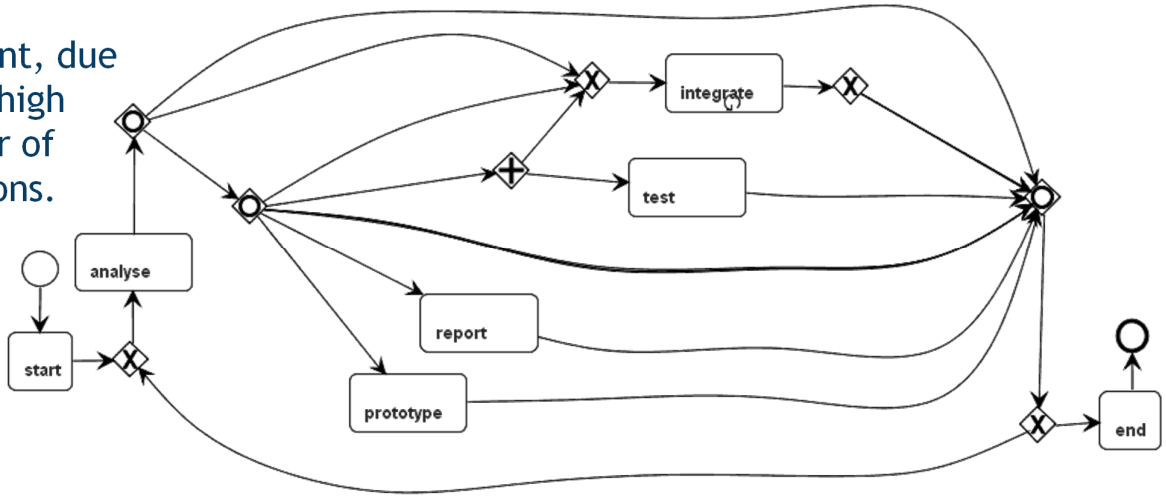
- Real log: 16 different cases, 241 total events.
- Many violations of the normative process.



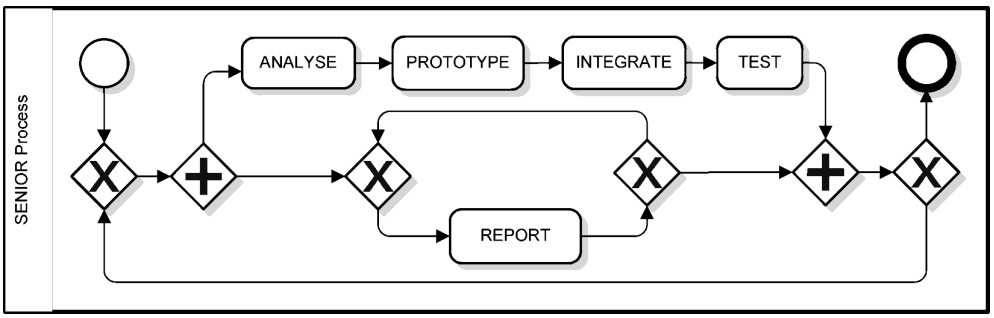
- Normative process:



- The "spaghetti" model generated by the inductive miner algorithm is very different, due to the high number of violations.



- Conformance checking: to check the real log against the normative process



- Steps on ProM:

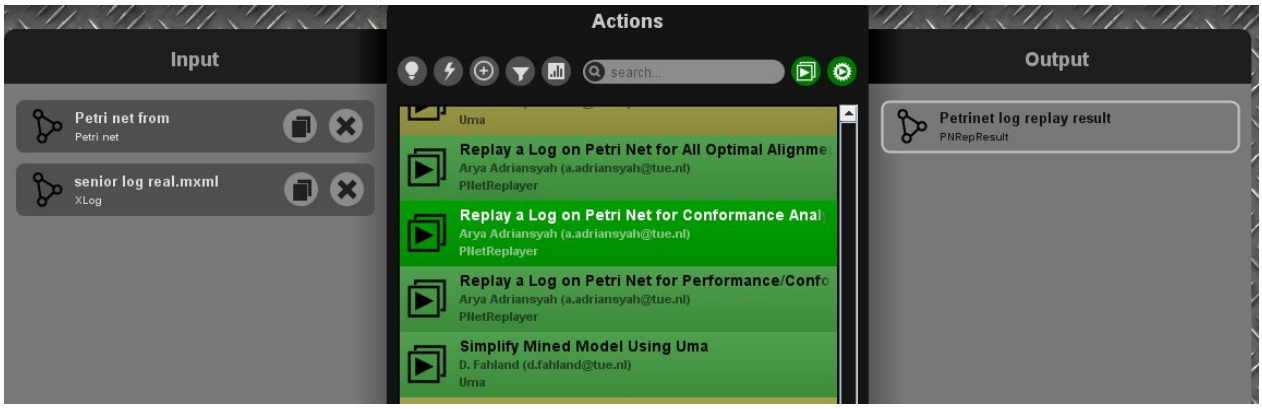
1: Import both the normative model (senior bpmn.xml) and the real log (senior log real.mxml)

2: In the action tab, click on "Select BPMN Diagram"; the normative process model appears.

3: Select the BPMN Diagram in the workspace tab, and click the action button; then select "Convert BPMN Diagram to Petri net (control-flow)"; a Petri net appears: you do not need to inspect it.

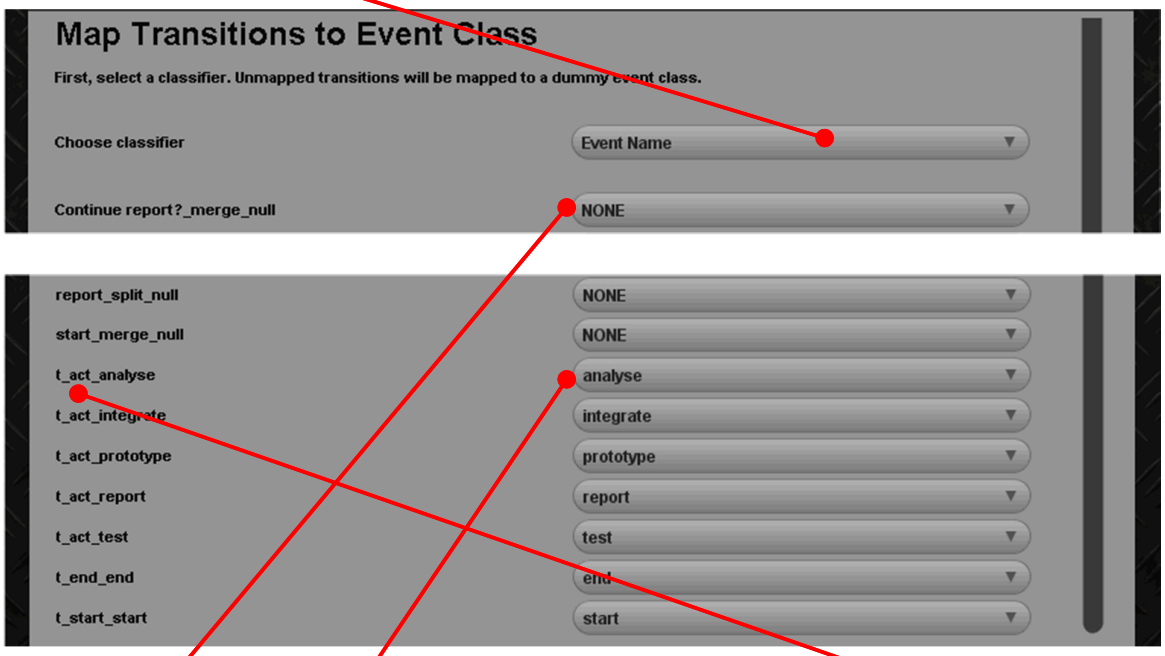
4: In the workspace tab, select the Petri net, and click the action button; then click to add input object and select "senior log real.mxml"

5: In the action tab, select "Replay Log on Petri Net for Conformance Analysis"



6: Answer YES to the question "No final marking is foud on this model. Do you want to create one?"; select "p_end_end" as a candidate final marking; Click on "Add place".

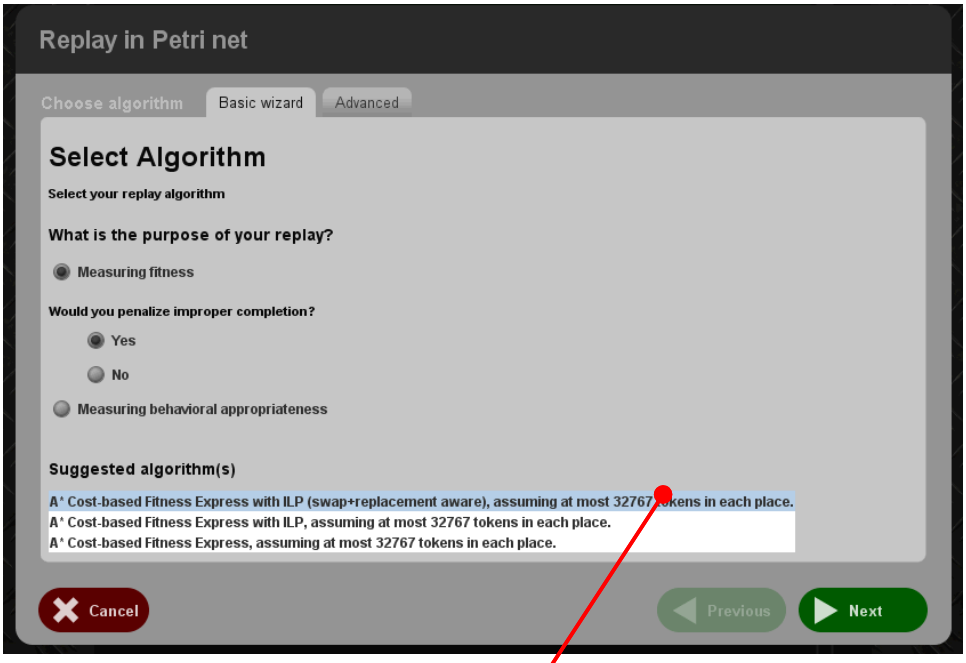
7: Select "Event Name" as a classifier



8: Select NONE for all transitions, except for those with "t_act","t_end", "t_start" as a name prefix, whose mapping must be accurately checked

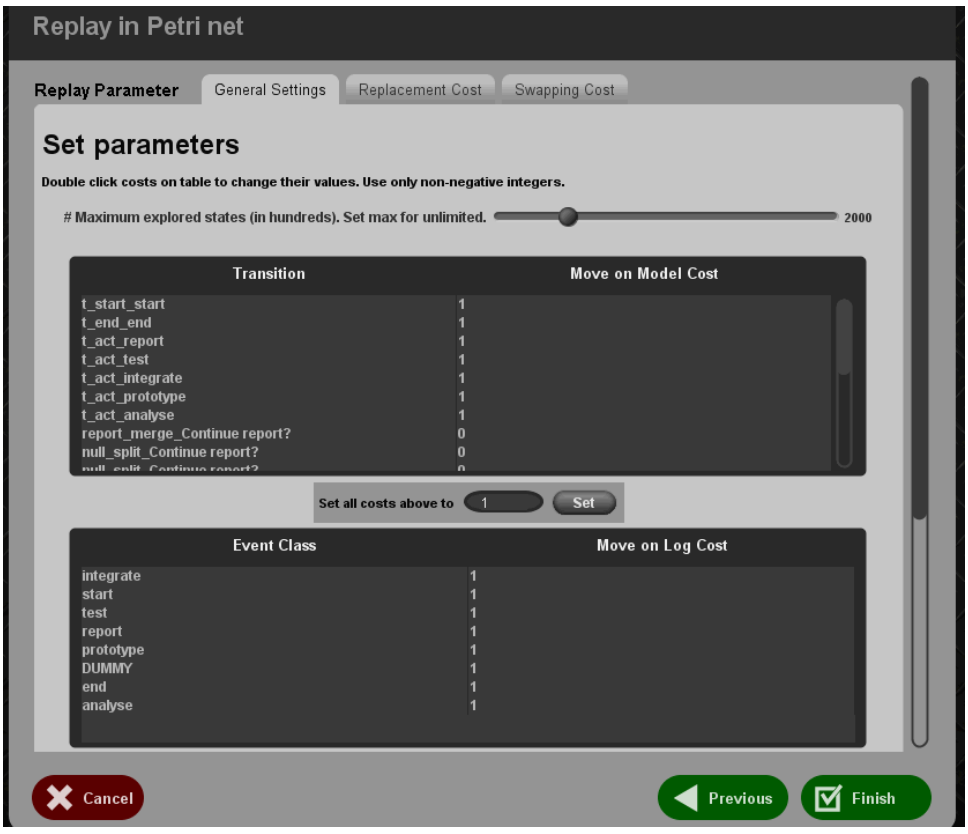
9: Select “YES, set them to invisible” in the dialog windows on the visibility of unmapped transitions

10: Wait the processing for abut a half a mninute

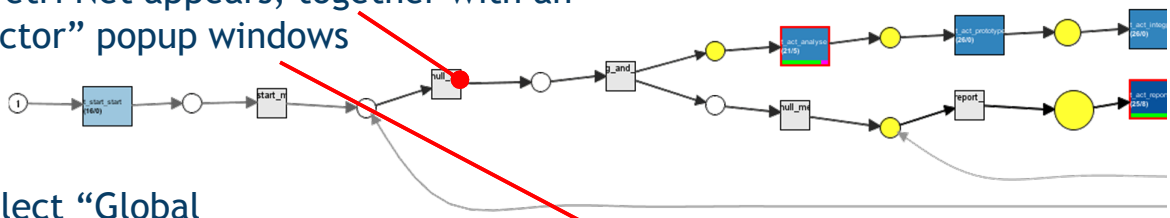


11: Leave the default algorithm for measuring fitness

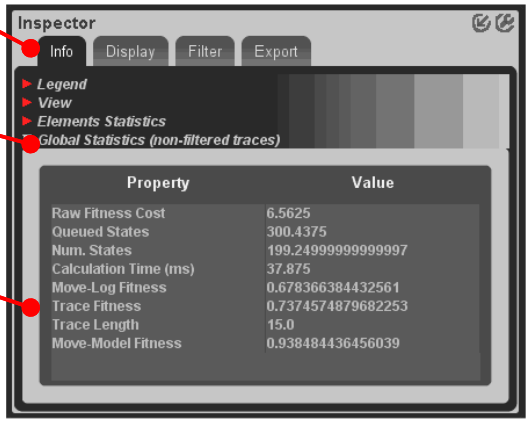
12: Leave the default parameters, and click on “Finish”. Wait some seconds



13: A Petri Net appears, together with an "Inspector" popup windows



14: Select "Global Statistics" to see a table with the most important average fitness properties.



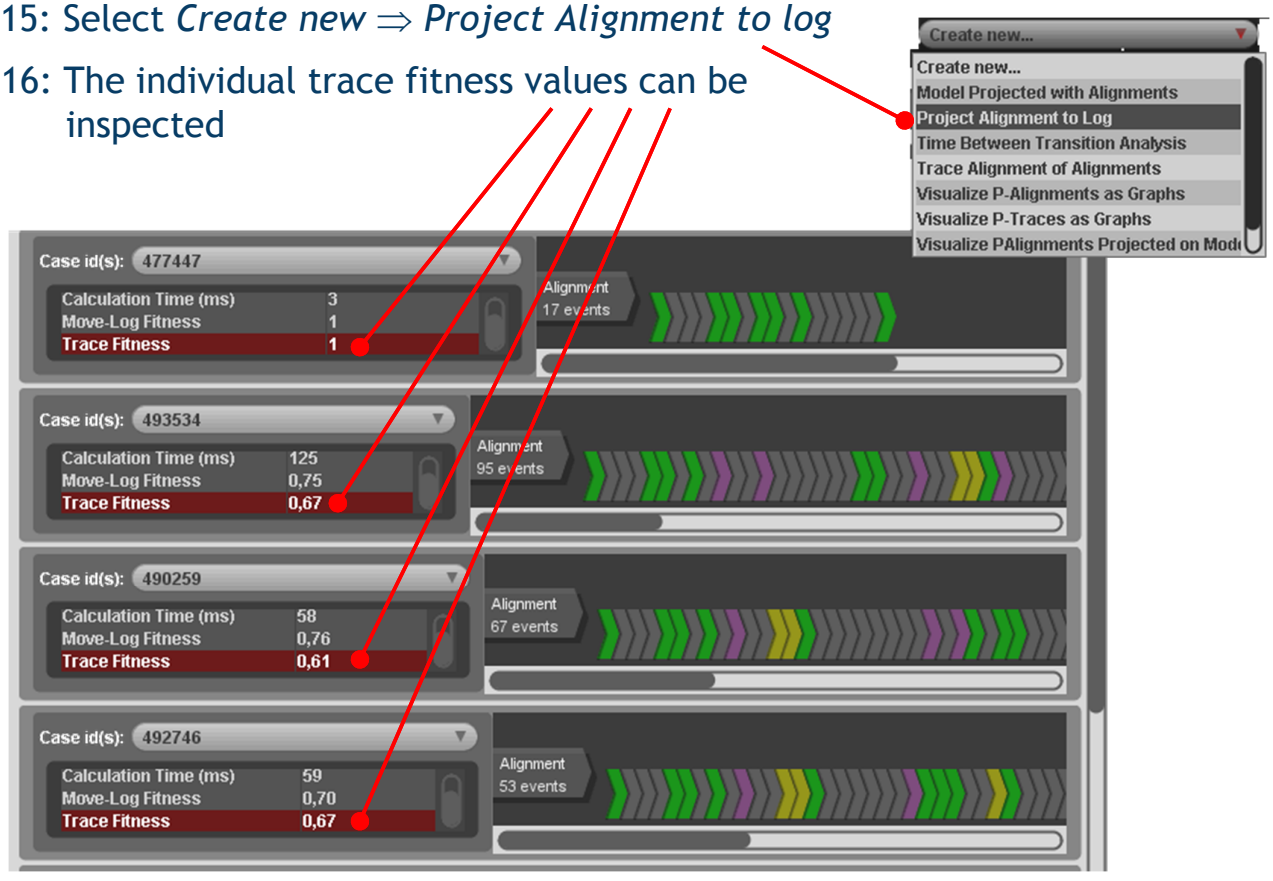
- In particular, the average *Trace Fitness* is shown: 0.737457...

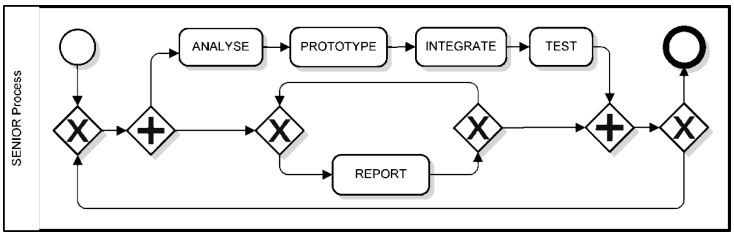
- The Trace-Fitness value represents the fitness value of the Petri Net with the log, and indicates how well the event log can be replayed in the discovered Petri Net.

- A fitness value of 1 means that the log can be successfully replayed, whereas a value of 0 means that this is completely not the case.

15: Select *Create new* ⇒ *Project Alignment to log*

16: The individual trace fitness values can be inspected





← The normative process

LEGEND
- Synchronous move (move log+model)
- Unobservable move (move model only)
- Skipped event class (move model only)
- Inserted event class (move log only)
- Replaced violation (move log+model)
- Swapped violation (move log+model)

- Case 491381: Fitness=1

491381, 11/12/2014 22.11, start
491381, 11/12/2014 22.12, analyse
491381, 15/12/2014 17.05, prototype
491381, 06/01/2015 12.07, integrate
491381, 22/01/2015 11.22, test
491381, 24/01/2015 09.58, report
491381, 24/01/2015 16.47, end

- Case 477089: Fitness = 0.92 (report event skipped)

477089, 16/12/2014 21.59, start
477089, 16/12/2014 22.00, analyse
477089, 26/12/2014 14.49, prototype
477089, 08/01/2015 21.22, integrate
477089, 11/01/2015 15.34, test
477089, 11/01/2015 18.27, end