Model Quality, Alpha Miner & Process Exploration

Benedikt Knopp

BPI-Instruction 4

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Today‘s Agenda

Model Quality Assessment

Alpha Miner: Step-by-Step Example

Alpha Miner: Many Examples

Process Exploration (ProM, Celonis)

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Recap: Model Quality Assessment

Fitness (Recall):
To what extend can the behavior recorded in the log be replayed by the model?

Precision:
To what extend is the model behavior present in the log?

Generalization:
How likely is it that an unseen trace can be replayed by the model?

Hard / Ambiguous answers.
Requires knowledge or assumptions about the process apart from what is seen in the log.

Simplicity:
How simple and readable is the model?

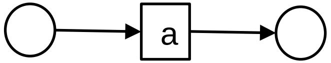
E.g. number of nodes, no label duplications…
Let's keep the discussion informal here

Recap: Model Quality Assessment

Generalization:
How likely is it that an unseen trace can be replayed by the model?

Log-based assessment: Gain confidence from observing log frequencies.

Non-Log based assessment: Make assumptions about the process apart from what is seen in the log.



$L_1 = [\langle a \rangle^1]$
$L_2 = [\langle a \rangle^{100}]$

For this model, generalization w.r.t L_2 is higher than w.r.t L_1 because we have more evidence that future traces look the same.

See next slides…

Chair of Process and Data Science

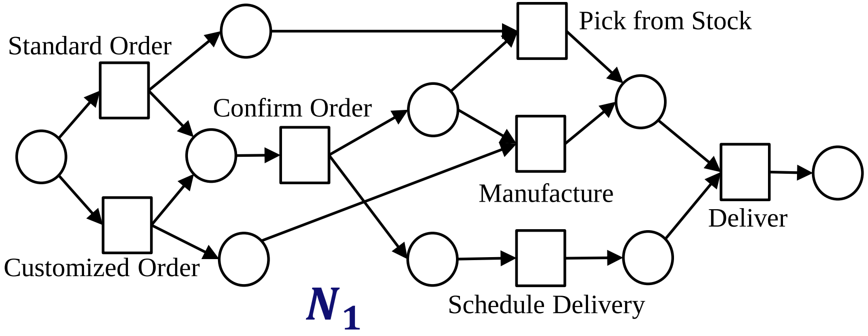
RWTH AACHEN UNIVERSITY

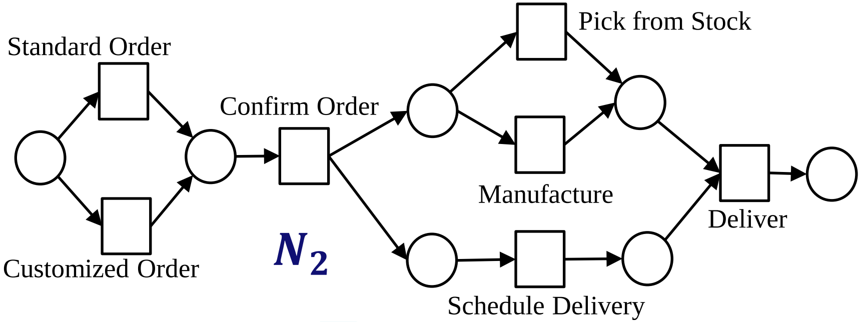
Exercise 1a: Model Quality Assessment

Recall the „carpenter“ process from last instruction (Exercise 6). Given the following event log L describing observed behavior:

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver\rangle, \\ \langle Customized, Confirm, Manufacture, Schedule, Deliver\rangle]$$

argue which of the following two Petri nets N_1, N_2 is better with respect to precision, fitness, generalization and simplicity.



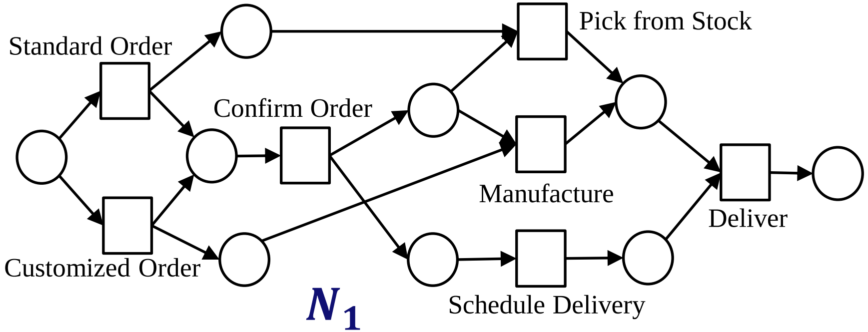


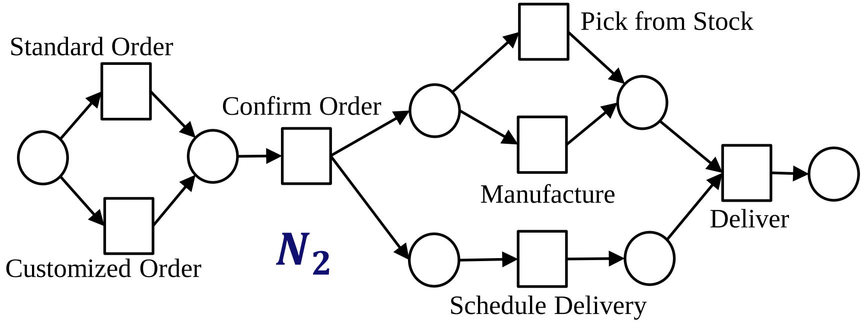
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 1a: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver \rangle, \langle Customized, Confirm, Manufacture, Schedule, Deliver \rangle]$$





Fitness (Recall):
To what extend can the behavior recorded in the log be replayed by the model?

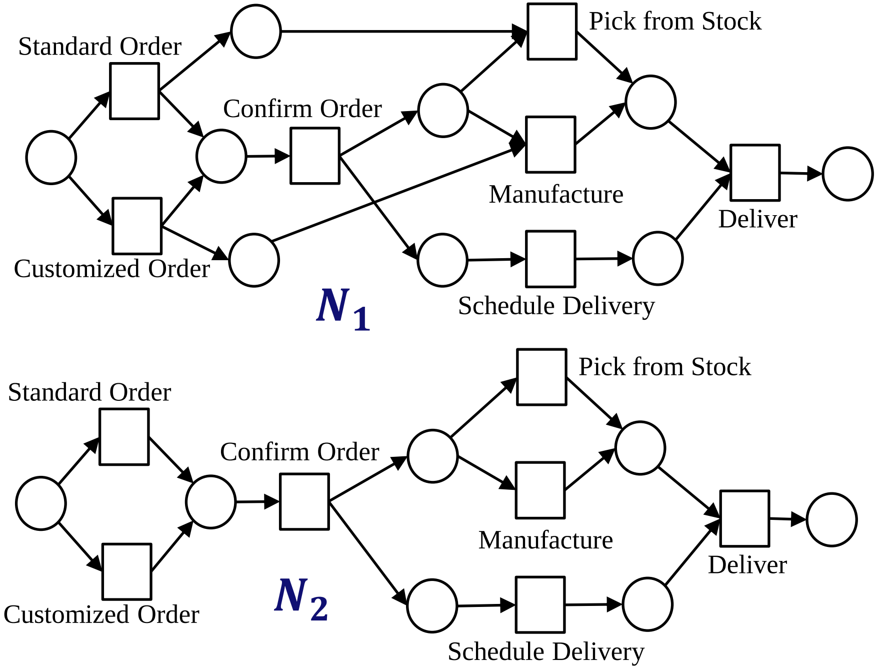
Both N_1 and N_2 can replay L completely.
N_1 is as fitting as N_2.

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 1a: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver\rangle, \langle Customized, Confirm, Manufacture, Schedule, Deliver\rangle]$$



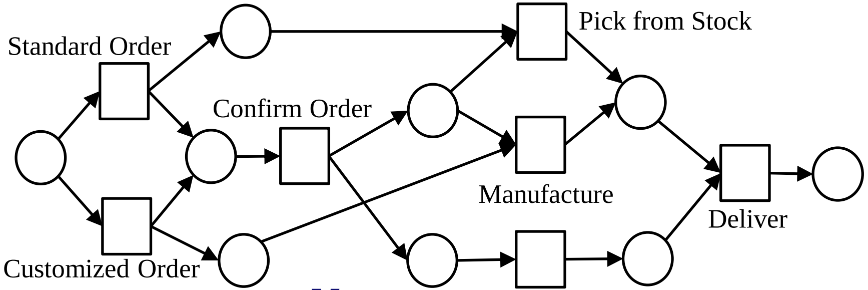
Precision:
To what extend is the model behavior present in the log?

N_2 allows for more further traces than N_1.
N_1 is more precise than N_2.

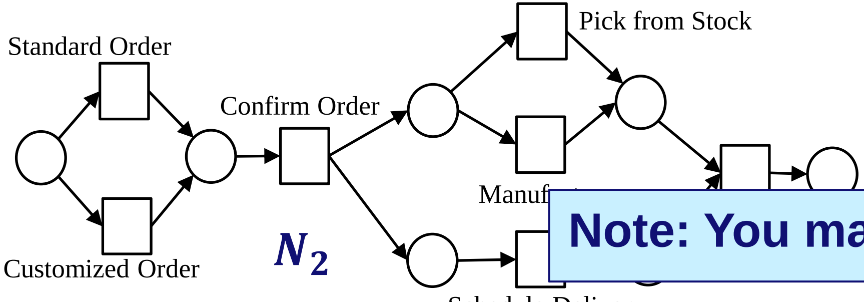
Exercise 1a: Mod ssment (Sol.)

"…the order is picked from stock in case of a catalog order, and manufactured in case of a customized order…"

$L = [\langle Standard, Confirm, Sch$
$\langle Customized, Confirm, Manuf \quad ule, Deliver\rangle]$



N_1



N_2

Generalization:
How likely is it that an unseen trace can be replayed by the model?

Assuming that the process description faithfully describes reality, N_1 generalizes as well as N_2.

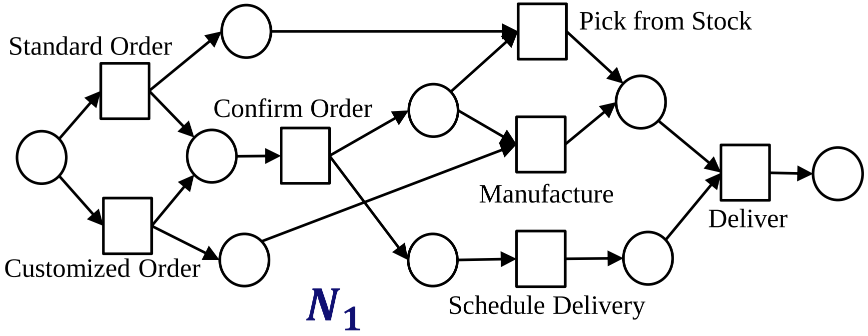
Note: You may also argue differently.

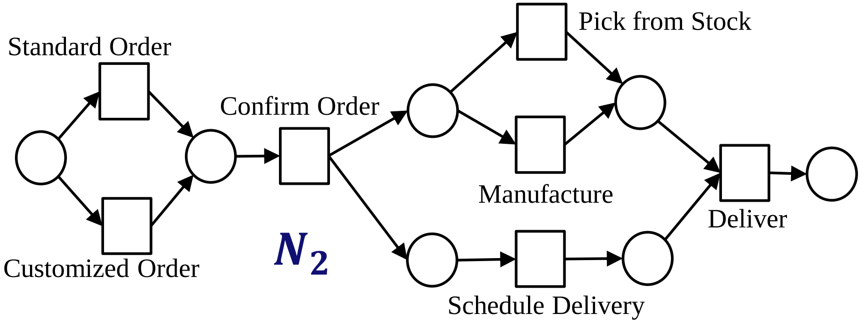
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 1a: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver \rangle, \langle Customized, Confirm, Manufacture, Schedule, Deliver \rangle]$$





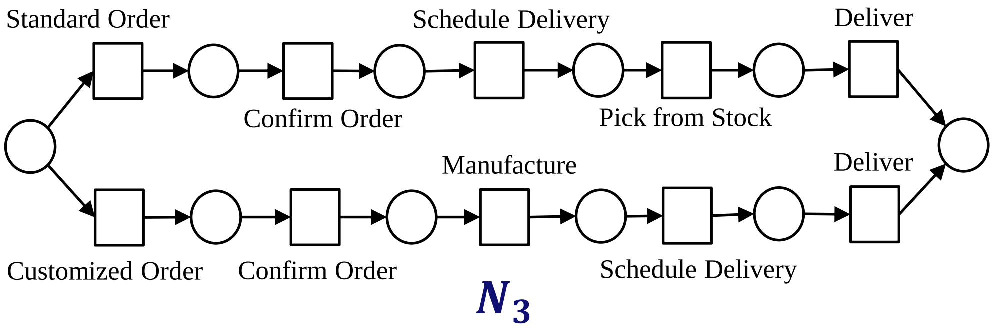
Simplicity:

How simple and readable is the model?

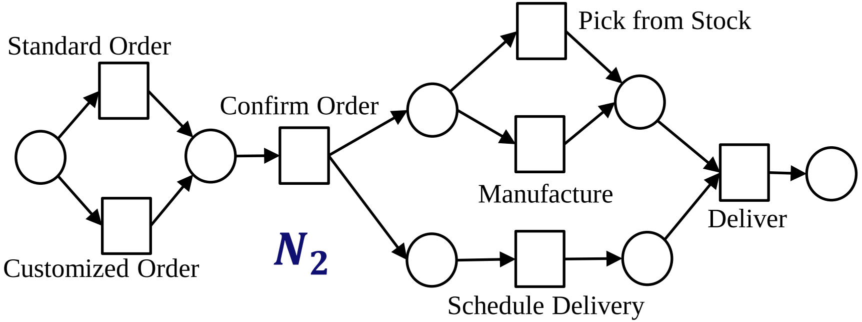
N_2 has two places less.

N_2 is slightly simpler than N_1.

Exercise 1b: Model Quality Assessment (Sol.)

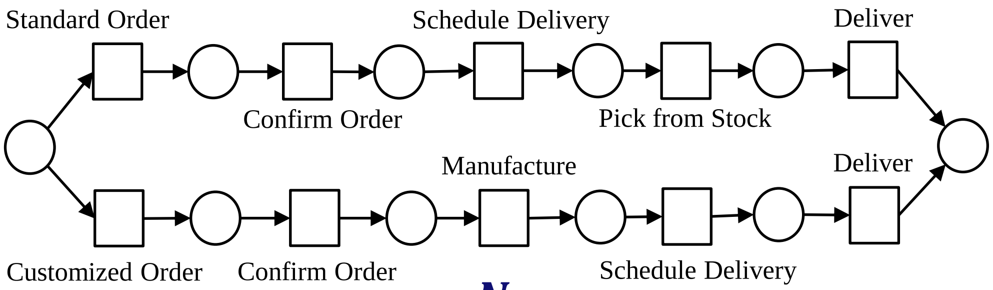


VS.

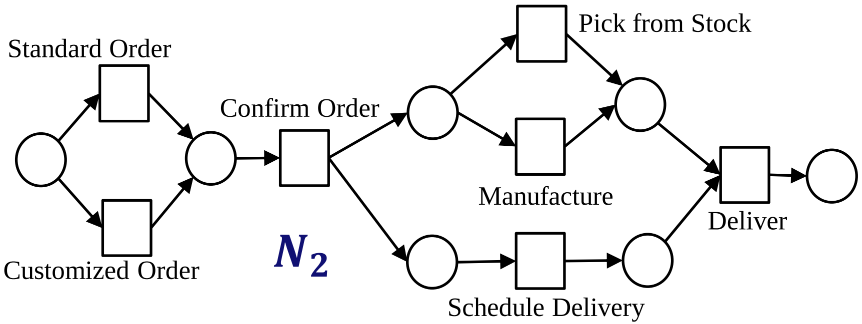


Exercise 1b: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver\rangle,$$
$$\langle Customized, Confirm, Manufacture, Schedule, Deliver\rangle]$$



N_3



N_2

Fitness (Recall):
To what extend can the behavior recorded in the log be replayed by the model?

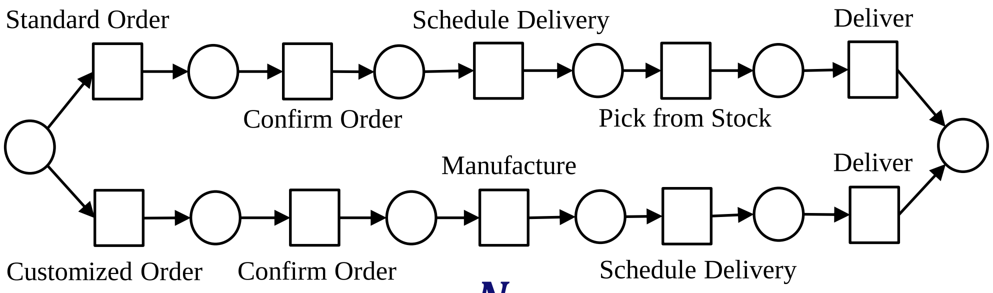
Both N_3 and N_2 can replay L completely.
N_3 is as fitting as N_2.

Chair of Process and Data Science

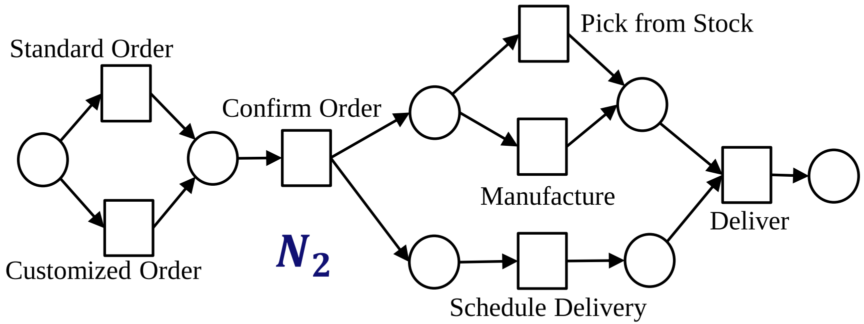
RWTH AACHEN UNIVERSITY

Exercise 1b: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver\rangle, \langle Customized, Confirm, Manufacture, Schedule, Deliver\rangle]$$



N_3



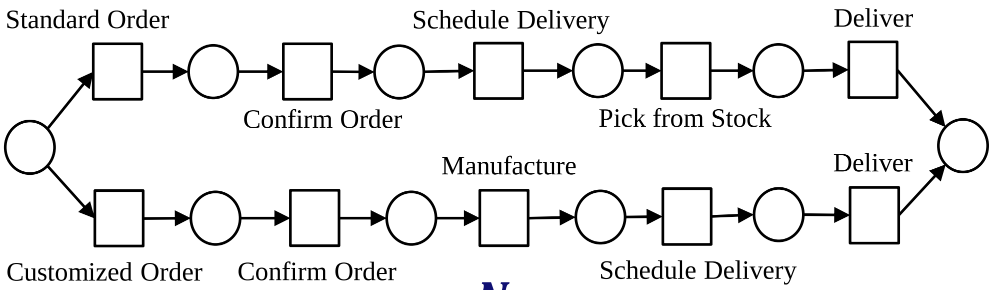
N_2

Precision:
To what extend is the model behavior present in the log?

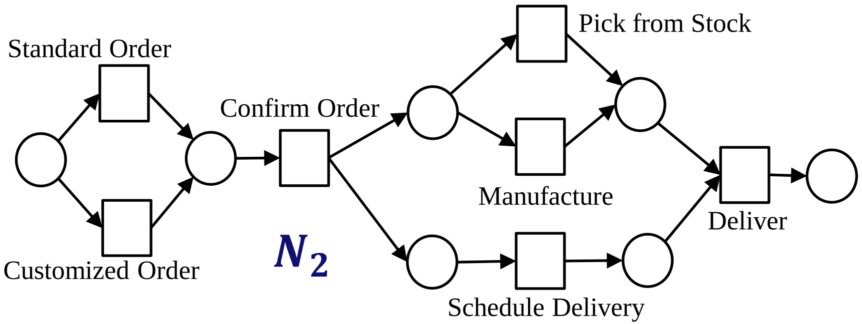
Only N_2 allows for traces that are not in L.
N_3 is more precise than N_2.

Exercise 1b: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver \rangle, \langle Customized, Confirm, Manufacture, Schedule, Deliver \rangle]$$



N_3



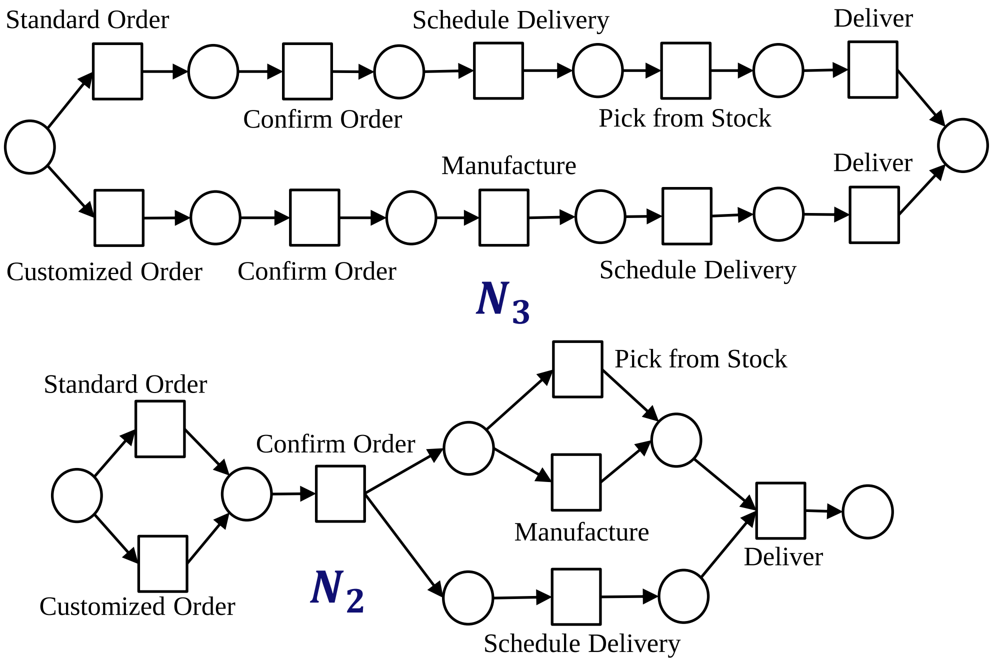
N_2

Generalization:
How likely is it that an unseen trace can be replayed by the model?

N_3 cannot replay e.g. the first trace in L with *Pick* and *Schedule* being swapped.
N_2 generalizes better than N_3.

Exercise 1b: Model Quality Assessment (Sol.)

$$L = [\langle Standard, Confirm, Schedule, Pick, Deliver \rangle, \langle Customized, Confirm, Manufacture, Schedule, Deliver \rangle]$$



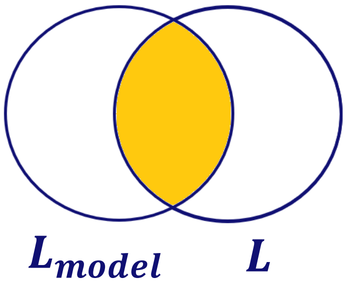
Simplicity:
How simple and readable is the model?

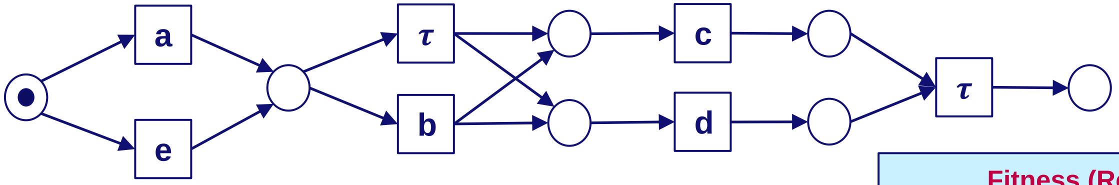
N_3 has duplicated transition labels and more nodes.
N_2 is simpler than N_3.

Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

Exercise 2a: Model Quality Metrics

Consider the following model and the following event log. Argue about fitness and precision of the model using the complete model and log traces as model/log behavior.





$$L = [\langle a, c\rangle^2, \langle e, d, c\rangle^5, \langle e, b, c, d\rangle^5]$$

Fitness (Recall):
To what extend can the behavior recorded in the log be replayed by the model?

Precision:
To what extend is the model behavior present in the log?

Fitness: $\frac{|L_{model} \cap L|}{|L|}$

With respect to trace frequencies

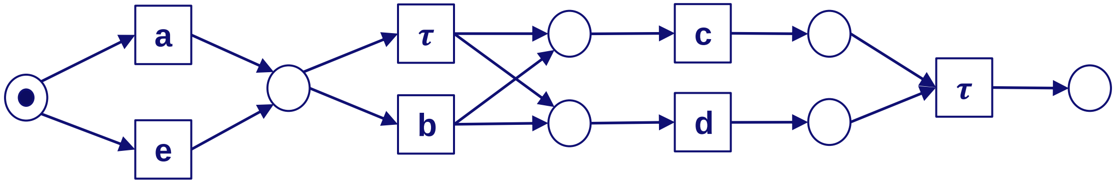
Precision: $\frac{|L_{model} \cap L|}{|L_{model}|}$

Without respect to trace frequencies

Chair of Process and Data Science

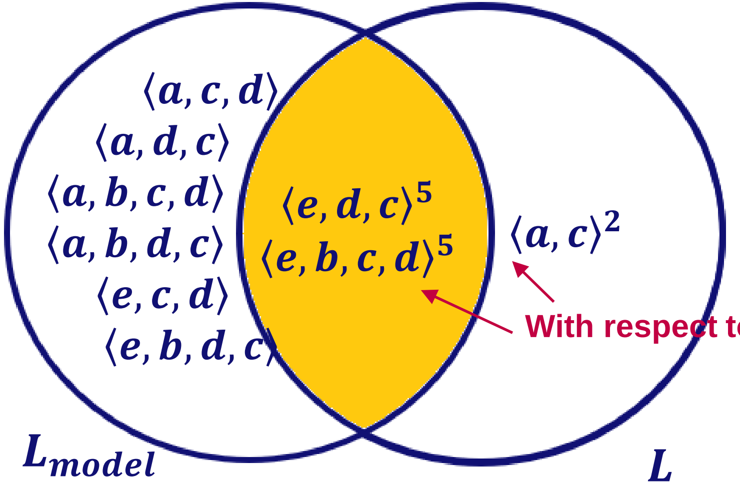
RWTH AACHEN UNIVERSITY

Exercise 2a: Model Quality Metrics (Sol.)



$$L_{model} = [\langle a,c,d\rangle, \langle a,d,c\rangle, \langle a,b,c,d\rangle, \langle a,b,d,c\rangle, \langle e,c,d\rangle, \langle e,d,c\rangle, \langle e,b,c,d\rangle, \langle e,b,d,c\rangle]$$

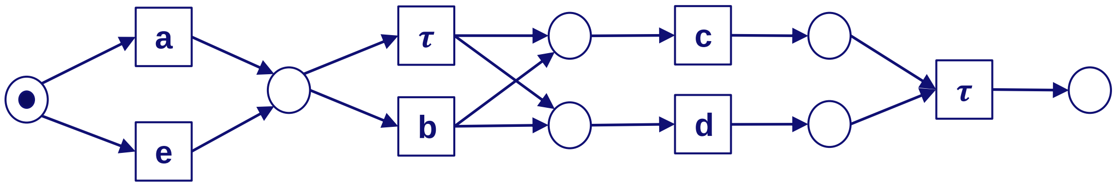
$$L = [\langle a,c\rangle^2, \langle e,d,c\rangle^5, \langle e,b,c,d\rangle^5]$$



With respect to trace frequencies

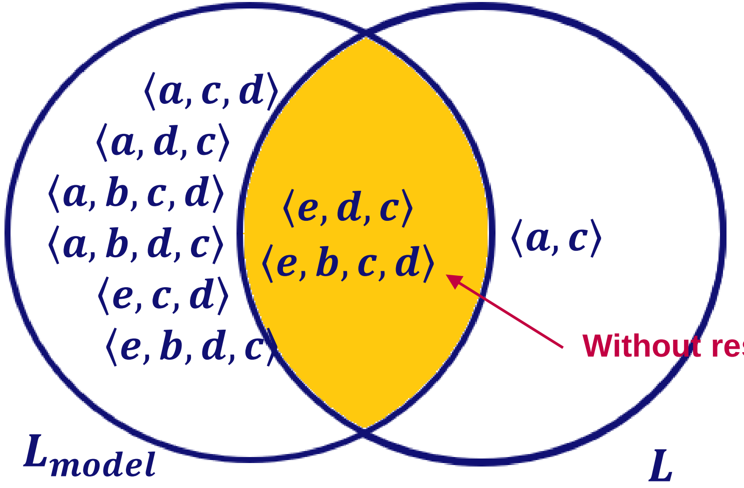
Fitness: $\frac{|L_{model} \cap L|}{|L|} = \frac{10}{12}$

Exercise 2a: Model Quality Metrics (Sol.)



$$L_{model} = [\langle a,c,d\rangle, \langle a,d,c\rangle, \langle a,b,c,d\rangle, \langle a,b,d,c\rangle, \langle e,c,d\rangle, \langle e,d,c\rangle, \langle e,b,c,d\rangle, \langle e,b,d,c\rangle]$$

$$L = [\langle a,c\rangle^2, \langle e,d,c\rangle^5, \langle e,b,c,d\rangle^5]$$

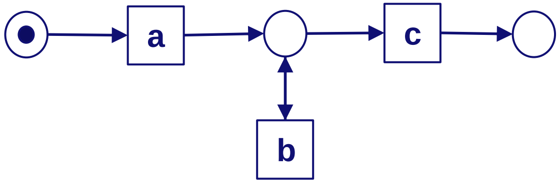


Without respect to trace frequencies

Precision: $\frac{|L_{model} \cap L|}{|L_{model}|} = \frac{2}{8}$

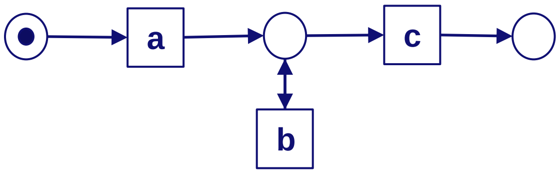
Exercise 2b: Model Quality Metrics

Consider the following model and the following event log. When reasoning about model quality as before, what is the problem here?



$$L = [\langle a, c\rangle^5, \langle a, b, c\rangle^3, \langle a, b, b, c\rangle^3]$$

Exercise 2b: Model Quality Metrics (Sol.)



$$L = [\langle a,c\rangle^5, \langle a,b,c\rangle^3, \langle a,b,b,c\rangle^3]$$

$$L_{model} = [\langle a,c\rangle, \langle a,b,c\rangle, \langle a,b,b,c\rangle, \langle a,b,b,b,c\rangle, \langle a,b,b,b,b,c\rangle, \ldots]$$

There are infinitely many accepted traces.

Precision: $\dfrac{|L_{model} \cap L|}{|L_{model}|} = \dfrac{3}{\infty} = 0$

This might not be what we want.
We need more advanced evaluation metrics.

Chair of Process and Data Science

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma\}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I\} \cup \{ (t, o_L) \mid t \in T_O\}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma \}$,

Identifying activities

2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma\}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a,p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)},b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L,t) \mid t \in T_I\} \cup \{ (t,o_L) \mid t \in T_O\}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Set of start activities

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma\}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$, ← **Set of end activities**
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I\} \cup \{ (t, o_L) \mid t \in T_O\}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma \}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

(A, B) pairs

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma \}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Deleting non maximal pairs

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma \}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$, ← **Places from pairs**
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

Recap – Alpha Miner

Let L be an event log over T. $\alpha(L)$ is defined as follows.

1. $T_L = \{ t \in T \mid \exists_{\sigma \in L}\ t \in \sigma \}$,
2. $T_I = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{first}(\sigma) \}$,
3. $T_O = \{ t \in T \mid \exists_{\sigma \in L}\ t = \mathit{last}(\sigma) \}$,
4. $X_L = \{ (A,B) \mid A \subseteq T_L \land A \neq \emptyset \land B \subseteq T_L \land B \neq \emptyset \land \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \land \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \land \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$,
5. $Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \land B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a,p_{(A,B)}) \mid (A,B) \in Y_L \land a \in A \} \cup \{ (p_{(A,B)},b) \mid (A,B) \in Y_L \land b \in B \} \cup \{ (i_L,t) \mid t \in T_I \} \cup \{ (t,o_L) \mid t \in T_O \}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Add arcs

Alpha Miner - Example

Consider the following event log L:

$$L = [\langle a, b, c\rangle, \langle a, c\rangle]$$

1. **Give the directly-follows matrix and the footprint matrix of L.**
2. **Derive the set of pairs of transition sets X and identify the maximal pairs Y.**
3. **Write down the resulting Petri net $\alpha(L)$.**
4. **Is the discovered model a sound workflow net?**
5. **Does the discovered process model represent both traces in the event log? If not, explain why.**

Alpha Miner - Example

Give the directly-follows matrix and the footprint matrix of L.

$$L = [\langle a, b, c\rangle, \langle a, c\rangle]$$

Directly-follows relations

	a	b	c
a		>	>
b			>
c			

Footprint Matrix

	a	b	c
a	#	$\rightarrow$	$\rightarrow$
b	$\leftarrow$	#	$\rightarrow$
c	$\leftarrow$	$\leftarrow$	#

Causality: **x $\rightarrow$ y iff x > y and not y > x.**
Parallelism: **x || y iff x > y and y > x**
Independence: **x # y iff not x > y and not y > x.**

Alpha Miner - Example

Derive the set of pairs of maximal transition sets Y.

We look at all pairs of non-empty activity sets

$$X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge$$
$$\forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \ \wedge\ \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \ \wedge\ \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$$

Each activity in A should be causal $\rightarrow_L$ to each in B

All activities in each A and B should be independent

Then Y is the maximal pairs in X.

Suggestion:...

Chair of Process and Data Science

Alpha Miner - Example

Derive the set of pairs of maximal transition sets Y.

Suggestion: Write down pairs of sets A, B. Start with what you find in the footprint matrix. Mind that here already, a#a, b#b, c#c needs to hold.

A	B
$\{a\}$	$\{b\}$
$\{a\}$	$\{c\}$
$\{b\}$	$\{c\}$

Alpha Miner - Example

Derive the set of pairs of maximal transition sets Y.

Iteratively, combine rows by joining A with A and B with B, if applicable.

Footprint Matrix

	a	b	c
a	#	→	→
b	←	#	→
c	←	←	#

	A	B
1	$\{a\}$	$\{b\}$
2	$\{a\}$	$\{c\}$
3	$\{b\}$	$\{c\}$
1+2	~~$\{a\}$~~	~~$\{b, c\}$~~
2+3	~~$\{a, b\}$~~	~~$\{c\}$~~
1+3	~~$\{a, b\}$~~	~~$\{b, c\}$~~

Cannot do that: b#c does not hold.

Cannot do that: a#b does not hold.

Cannot do that: a#b and b#c and b→b do not hold.

Chair of Process and Data Science

Alpha Miner - Example

Derive the set of pairs of maximal transition sets Y.

We are done: This is X.

A	B
$\{a\}$	$\{b\}$
$\{a\}$	$\{c\}$
$\{b\}$	$\{c\}$

All pairs A, B are maximal. Therefore, $Y = X$ in this case.

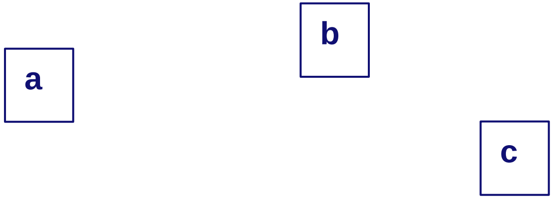
Alpha Miner - Example

L= [<a, b, c>, <a, c>]

1. T_L = { a, b, c},
2. T_I = { a },
3. T_O = { c },

Footprint

	a	b	c
a	#	→	→
b	←	#	→
c	←	←	#



Chair of Process and Data Science

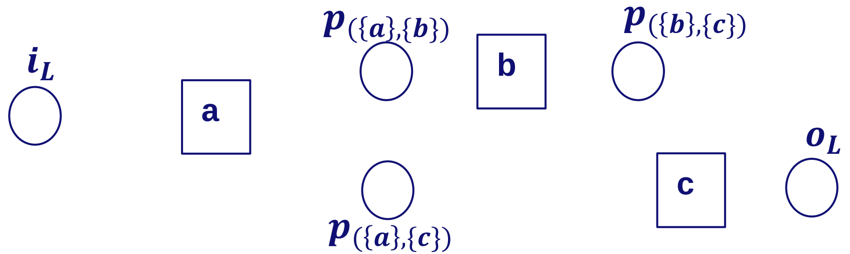
Alpha Miner - Example

L= [<a, b, c>, <a, c>]

1. T_L = { a, b, c},
2. T_I = { a },
3. T_O = { c },
4. X_L = { ({a}, {b}), ({b}, {c}), ({a}, {c}) }
5. Y_L = { ({a}, {b}), ({b}, {c}), ({a}, {c}) }
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,

Footprint

	a	b	c
a	#	→	→
b	←	#	→
c	←	←	#



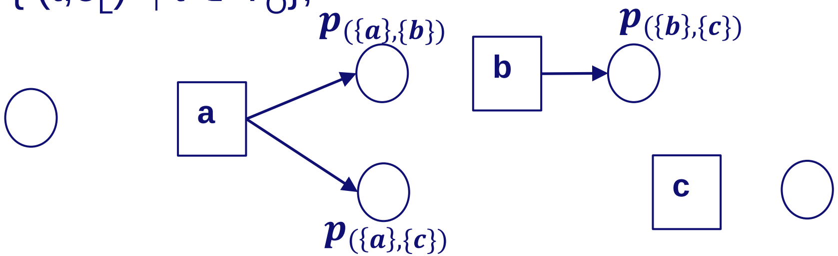
Alpha Miner - Example

L= [<a, b, c>, <a, c>]

1. T_L = { a, b, c},
2. T_I = { a },
3. T_O = { c },
4. X_L = { ({a}, {b}), ({b}, {c}), ({a}, {c}) }
5. Y_L = { ({a}, {b}), ({b}, {c}), ({a}, {c}) }
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \boxed{\{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \}} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I\} \cup \{ (t, o_L) \mid t \in T_O\}$,

Footprint

	a	b	c
a	#	→	→
b	←	#	→
c	←	←	#



Chair of Process and Data Science

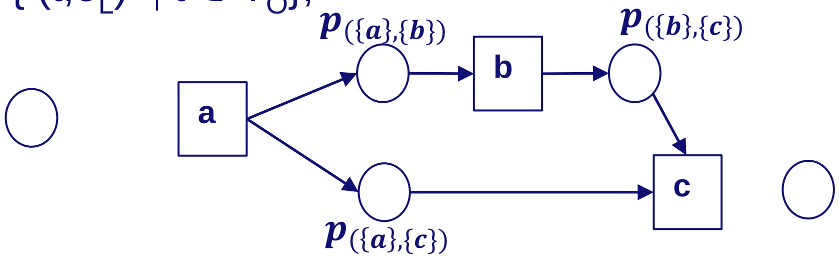
Alpha Miner - Example

L= [<a, b, c>, <a, c>]

1. $T_L = \{ a, b, c\}$,
2. $T_I = \{ a \}$,
3. $T_O = \{ c \}$,
4. $X_L = \{ (\{a\}, \{b\}), (\{b\}, \{c\}), (\{a\}, \{c\}) \}$
5. $Y_L = \{ (\{a\}, \{b\}), (\{b\}, \{c\}), (\{a\}, \{c\}) \}$
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \boxed{\{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \}} \cup \{ (i_L, t) \mid t \in T_I\} \cup \{ (t, o_L) \mid t \in T_O\}$,

Footprint

	a	b	c
a	#	→	→
b	←	#	→
c	←	←	#



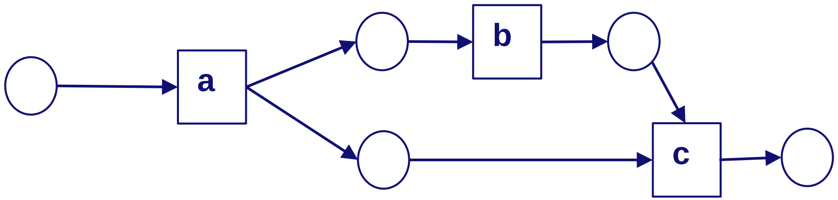
Alpha Miner - Example

L= [<a, b, c>, <a, c>]

1. $T_L = \{ a, b, c \}$,
2. $T_I = \{ a \}$,
3. $T_O = \{ c \}$,
4. $X_L = \{ (\{a\}, \{b\}), (\{b\}, \{c\}), (\{a\}, \{c\}) \}$
5. $Y_L = \{ (\{a\}, \{b\}), (\{b\}, \{c\}), (\{a\}, \{c\}) \}$
6. $P_L = \{ p_{(A,B)} \mid (A,B) \in Y_L \} \cup \{i_L, o_L\}$,
7. $F_L = \{ (a, p_{(A,B)}) \mid (A,B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A,B)}, b) \mid (A,B) \in Y_L \wedge b \in B \} \cup \boxed{\{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \}}$,

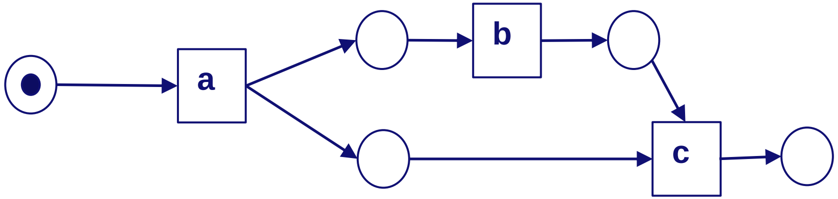
Footprint

	a	b	c
a	#	→	→
b	←	#	→
c	←	←	#



Chair of Process and Data Science

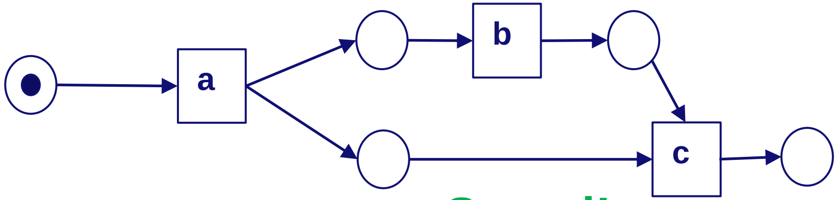
Exercise 3: Alpha Miner Example



$$L = [\langle a, b, c \rangle, \langle a, c \rangle]$$

1. Is the discovered model a sound workflow net?
2. Does the discovered process model represent both traces in the event log? If not, explain why.

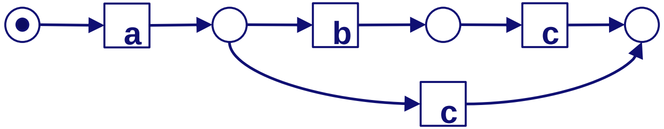
Exercise 3: Alpha Miner Example (Solution)



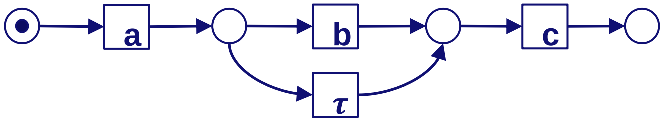
Sound!

$$L = [\langle a, b, c\rangle, \langle a, c\rangle]$$

The model can only replay $\langle a, b, c\rangle$.
A fitting model would look as follows:



or



Silent Transition

The Alpha algorithm can by definition never discover such models with duplicated labels or silent transitions (representational bias)

Chair of Process and Data Science

Exercise 4: Alpha Miner – Practice, Practice

Consider the event logs given below. For each event log, think about what a Petri net might look like that describes the behavior suitably. Apply the Alpha algorithm, validate your result with ProM, and compare the result with your conceived net.
If the net is a sound workflow net, argue about precision and fitness (and generalization).

a) $L_1 = [\langle$*order, pay*$\rangle$, $\langle$*order, reminder, pay*$\rangle$, $\langle$*order, reminder, reminder, pay*$\rangle]$

b) $L_2 = [\langle$*a,b,c,d,e,f,b,d,c,e,g*$\rangle$, $\langle$*a,b,d,c,e,g*$\rangle$, $\langle$*a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g*$\rangle]$

c) $L_3 = [\langle$*awake, eat, study, sleep*$\rangle$, $\langle$*awake, study, eat, sleep*$\rangle]$

d) $L_4 = [\langle$*awake, eat, study, eat, study, sleep*$\rangle]$

Let $L_5 = [\langle$*awake, eat, study, eat, study, sleep*$\rangle$, $\langle$*awake, study, eat, study, eat, sleep*$\rangle]$. What Petri net would describe this process well? Compare the directly-follows matrices of L_3, L_4 and L_5. What is $\alpha(L_5)$?

Chair of Process and Data Science

Exercise 4: Alpha Miner – Practice, Practice

Consider the event logs given below. For each event log, think about what a Petri net might look like that describes the behavior suitably. Apply the Alpha algorithm, validate your result with ProM, and compare the result with your conceived net.
If the net is a sound workflow net, argue about precision and fitness (and generalization).

Note: This should simply trigger some thoughts – There is no comparison between the conceived model and the actual outcome on the later slides.

a) $L_1 = [\langle$*order, pay*$\rangle$, … *minder, pay*$\rangle]$

b) $L_2 = [\langle$*a,b,c,d,e,f,b,* … *d,c,e,g*$\rangle]$

c) $L_3 = [\langle$*awake, eat,* …

d) $L_4 = [\langle$*awake, eat, study, eat, study, sleep*$\rangle]$

Let $L_5 = [\langle$*awake, eat, study, eat, study, sleep*$\rangle$, $\langle$*awake, study, eat, study, eat, sleep*$\rangle]$.
What Petri net would describe this process well? Compare the directly-follows matrices of L_3, L_4 and L_5. What is $\alpha(\mathrm{L}_5)$?

Chair of Process and Data Science

Exercise 4: Alpha Miner – Practice, Practice

Consider the event logs given below. For each event log, think about what a Petri net might look like that describes the behavior suitably. Apply the Alpha algorithm, validate your result with ProM, and compare the result with your conceived net.
If the net is a sound workflow net, argue about precision and fitness (and generalization).

a) $L_1 = [\langle$*order, pay*$\rangle$, $\langle$*order, reminder, pay*$\rangle$, $\langle$*order, reminder, reminder, pay*$\rangle]$

b) $L_2 = [\langle$*a,b,c,d,e,f,b,d,c,e,g*$\rangle$, ... $\langle$*a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g*$\rangle]$

c) $L_3 = [\langle$*awake, eat, study* ...

d) $L_4 = [\langle$*awake, eat, study* ...

Let $L_5 = [\langle$*awake, eat, study* ...
What Petri net would descr...
matrices of L_3, L_4 and L_5. W... $\alpha(L_5)$?

Use the alpha_L[xxx] files. If you want to practice with more logs, you can just modify any of these files (with a standard Texteditor) so that it features the log you want.
Take care to have a correct sorting (timestamps) and that each trace has a unique case id!

Chair of Process and Data Science

Exercise 4: Alpha Miner – Practice, Practice

Consider the event logs given below. For each event log, think about what a Petri net might look like that describes the behavior suitably. Apply the Alpha algorithm, validate your result with ProM, and compare the result with your conceived net.
If the net is a sound workflow net, argue about precision and fitness (and generalization).

Be careful here!

a) $L_1 = [\langle order, pay \rangle, \langle order, reminder, pay \rangle, \langle order, reminder, reminder, pay \rangle]$

b) $L_2 = [\langle a,b,c,d,e,f,b,d,c,e,g \rangle, \langle a,b,d,c,e,g \rangle, \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g \rangle]$

c) $L_3 = [\langle awake, eat, study, sleep \rangle, \langle awake, study, eat, sleep \rangle]$

d) $L_4 = [\langle awake, eat, study, eat, study, sleep \rangle]$

Hint: c and d are concurrent

Let $L_5 = [\langle awake, eat, study, eat, study, sleep \rangle, \langle awake, study, eat, study, eat, sleep \rangle]$. What Petri net would describe this process well? Compare the directly-follows matrices of L_3, L_4 and L_5. What is $\alpha(L_5)$?

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4: Alpha Miner – Practice, Practice

Consider the event logs given below. For each event log, think about what a Petri net might look like that describes the behavior suitably. Apply the Alpha algorithm, validate your result with ProM, and compare the result with your conceived net.
If the net is a sound workflow net, argue about precision and fitness (and generalization).

a) $L_1 = [\langle order, pay \rangle, \langle order, reminder, pay \rangle, \langle order, reminder, reminder, pay \rangle]$

b) $L_2 = [\langle a,b,c,d,e,f,b,d,c,e,g \rangle, \langle a,b,d,c,e,g \rangle, \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g \rangle]$

c) $L_3 = [\langle awake, eat, study, sleep \rangle, \langle awake, study, eat, sleep \rangle]$

d) $L_4 = [\langle awake, eat, study, eat, study, sleep \rangle]$

Note:
Loop of length 1
Loop of length 2
Loop of length 3

Let $L_5 = [\langle awake, eat, study, eat, study, sleep \rangle, \langle awake, study, eat, study, eat, sleep \rangle]$.
What Petri net would describe this process well? Compare the directly-follows matrices of L_3, L_4 **and** L_5**. What is** $\alpha(L_5)$?

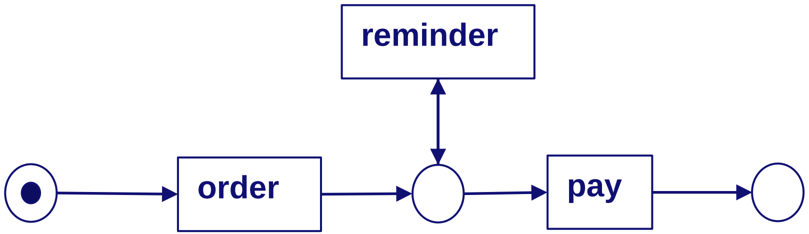
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4a: Alpha Miner – Practice (Solution)

$L_1 = [\langle \textit{order, pay} \rangle, \langle \textit{order, reminder, pay} \rangle, \langle \textit{order, reminder, reminder, pay} \rangle]$

***Think about what a Petri net might look like that describes the behavior suitably*.**
This is a suggestion how a suitable model may look like:



Note: This should simply trigger some thoughts – There is no comparison between this and the actual outcome on the later slides.

Exercise 4a: Alpha Miner – Practice (Solution)

$L_1 = [\langle \textit{order, pay} \rangle, \langle \textit{order, reminder, pay} \rangle, \langle \textit{order, reminder, reminder, pay} \rangle]$

Directly-follows relations

	o	r	p
o		>	>
r		>	>
p			

Footprint Matrix

	o	r	p
o	#	→	→
r	←	||	→
p	←	←	#

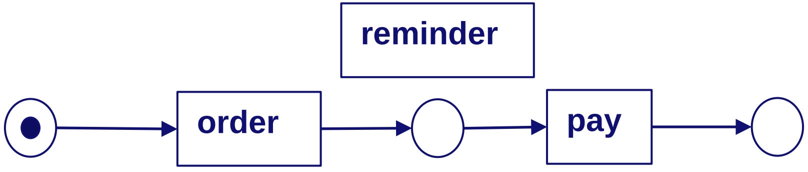
A	B
$\{o\}$	$\{p\}$
$\cancel{\{o\}}$	$\cancel{\{r\}}$
$\cancel{\{r\}}$	$\cancel{\{p\}}$

Cannot do that: r#r does not hold.

Cannot do that: r#r does not hold.

$X = \{(\{o\},\{p\})\}$

$Y = \{(\{o\},\{p\})\}$



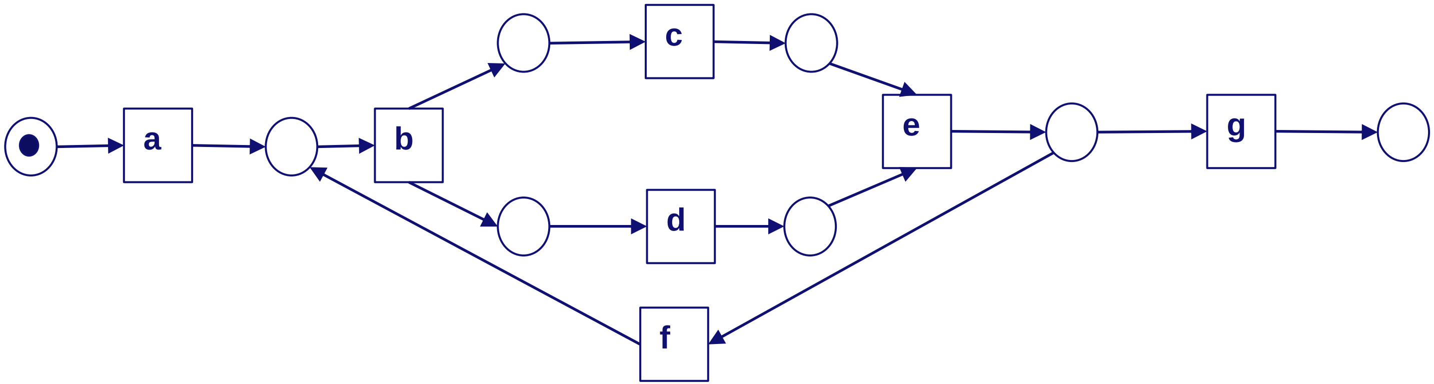
$\alpha(L_1)$

not a workflow net!

Exercise 4b: Alpha Miner – Practice (Solution)

$L_2 = [\langle a,b,c,d,e,f,b,d,c,e,g \rangle, \langle a,b,d,c,e,g \rangle, \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g \rangle]$

Suggested Model:



Note: This should simply trigger some thoughts – There is no comparison between this and the actual outcome on the later slides.

Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Practice (Solution)

This one a bit more detailed again.

$$L_2 = [\langle a,b,c,d,e,f,b,d,c,e,g\rangle, \langle a,b,d,c,e,g\rangle, \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g\rangle]$$

Directly-follows relations

	a	b	c	d	e	f	g
a		>					
b			>	>			
c				>	>		
d			>		>		
e						>	>
f		>					
g							

Footprint Matrix (omit # for brevity)

	a	b	c	d	e	f	g
a		→					
b	←		→	→		←	
c		←		||	→		
d		←	||		→		
e			←	←		→	→
f		→			←		
g					←		

Exercise 4b: Alpha Miner – Practice (Solution)

Start again by writing down all pairs of singleton sets that meet the definition of X.

Footprint Matrix

	a	b	c	d	e	f	g
a		$\rightarrow$					
b	$\leftarrow$		$\rightarrow$	$\rightarrow$		$\leftarrow$	
c		$\leftarrow$		$\|$	$\rightarrow$		
d		$\leftarrow$	$\|$		$\rightarrow$		
e			$\leftarrow$	$\leftarrow$		$\rightarrow$	$\rightarrow$
f		$\rightarrow$			$\leftarrow$		
g							

	A	B
1	$\{a\}$	$\{b\}$
2	$\{b\}$	$\{c\}$
3	$\{b\}$	$\{d\}$
4	$\{c\}$	$\{e\}$
5	$\{d\}$	$\{e\}$
6	$\{e\}$	$\{f\}$
7	$\{e\}$	$\{g\}$
8	$\{f\}$	$\{b\}$

$$X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\, a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\, a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\, b_1 \#_L b_2 \}$$

Exercise 4b: Alpha Miner –

Check for all possible combinations of rows whether they satisfy the definition of X.

Footprint Matrix

	a	b	c	d	e	f	g
a		→					
b	←		→	→		←	
c		←		||	→		
d		←	||		→		
e			←	←		→	→
f		→			←		
g					←		

	A	B
1	$\{a\}$	$\{b\}$
2	$\{b\}$	$\{c\}$
3	$\{b\}$	$\{d\}$
4	$\{c\}$	$\{e\}$
5	$\{d\}$	$\{e\}$
6	$\{e\}$	$\{f\}$
7	$\{e\}$	$\{g\}$
8	$\{f\}$	$\{b\}$

merge step

	A	B
1+2	$\{a, b\}$	$\{b, c\}$
1+3	$\{a, b\}$	$\{b, d\}$
1+4	$\{a, c\}$	$\{b, e\}$
1+5	$\{a, d\}$	$\{b, e\}$
1+6	$\{a, e\}$	$\{b, f\}$
1+7	$\{a, e\}$	$\{b, g\}$
1+8	$\{a, f\}$	$\{b\}$

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner –

Check for all possible combinations of rows whether they satisfy the definition of X (continued).

Footprint Matrix

	a	b	c	d	e	f	g
a		→					
b	←		→	→		←	
c		←		‖	→		
d		←	‖		→		
e			←	←		→	→
f		→			←		
g							

	A	B
1	$\{a\}$	$\{b\}$
2	$\{b\}$	$\{c\}$
3	$\{b\}$	$\{d\}$
4	$\{c\}$	$\{e\}$
5	$\{d\}$	$\{e\}$
6	$\{e\}$	$\{f\}$
7	$\{e\}$	$\{g\}$
8	$\{f\}$	$\{b\}$

merge step

	A	B
~~1+2~~	~~$\{a, b\}$~~	~~$\{b, c\}$~~
~~1+3~~	~~$\{a, b\}$~~	~~$\{b, d\}$~~
~~1+4~~	~~$\{a, c\}$~~	~~$\{b, e\}$~~
~~1+5~~	~~$\{a, d\}$~~	~~$\{b, e\}$~~
~~1+6~~	~~$\{a, e\}$~~	~~$\{b, f\}$~~
~~1+7~~	~~$\{a, e\}$~~	~~$\{b, g\}$~~
1+8	$\{a, f\}$	$\{b\}$

$$X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\, a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\, a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\, b_1 \#_L b_2 \}$$

Exercise 4b: Alpha Miner – Check for all possible combinations of rows whether they satisfy the definition of X (continued).

Footprint Matrix

	a	b	c	d	e	f	g
a		$\rightarrow$					
b	$\leftarrow$		$\rightarrow$	$\rightarrow$		$\leftarrow$	
c		$\leftarrow$		$\parallel$	$\rightarrow$		
d		$\leftarrow$	$\parallel$		$\rightarrow$		
e			$\leftarrow$	$\leftarrow$		$\rightarrow$	$\rightarrow$
f		$\rightarrow$			$\leftarrow$		
g					$\leftarrow$		

	A	B
1	$\{a\}$	$\{b\}$
2	$\{b\}$	$\{c\}$
3	$\{b\}$	$\{d\}$
4	$\{c\}$	$\{e\}$
5	$\{d\}$	$\{e\}$
6	$\{e\}$	$\{f\}$
7	$\{e\}$	$\{g\}$
8	$\{f\}$	$\{b\}$

merge step

	A	B
2+3	$\{b\}$	$\{c,d\}$
2+4	$\{b,c\}$	$\{c,e\}$
2+5	$\{b,d\}$	$\{c,e\}$
2+6	$\{b,e\}$	$\{c,f\}$
2+7	$\{b,e\}$	$\{c,g\}$
2+8	$\{b,f\}$	$\{c,b\}$

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Check for all possible combinations of rows whether they satisfy the definition of X (continued).

Footprint Matrix

	a	b	c	d	e	f	g
a		→					
b	←		→	→		←	
c		←		‖	→		
d		←	‖		→		
e			←	←		→	→
f		→			←		
g							

	A	B
1	$\{a\}$	$\{b\}$
2	$\{b\}$	$\{c\}$
3	$\{b\}$	$\{d\}$
4	$\{c\}$	$\{e\}$
5	$\{d\}$	$\{e\}$
6	$\{e\}$	$\{f\}$
7	$\{e\}$	$\{g\}$
8	$\{f\}$	$\{b\}$

merge step

	A	B
~~2+3~~	~~$\{b\}$~~	~~$\{c, d\}$~~
~~2+4~~	~~$\{b, c\}$~~	~~$\{c, e\}$~~
~~2+5~~	~~$\{b, d\}$~~	~~$\{c, e\}$~~
~~2+6~~	~~$\{b, e\}$~~	~~$\{c, f\}$~~
~~2+7~~	~~$\{b, e\}$~~	~~$\{c, g\}$~~
~~2+8~~	~~$\{b, f\}$~~	~~$\{c, b\}$~~

$$X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\; a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\; a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\; b_1 \#_L b_2 \}$$

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		$\rightarrow$					
e			$\leftarrow$	$\leftarrow$		$\rightarrow$	$\rightarrow$
f		$\rightarrow$			$\leftarrow$		
g					$\leftarrow$		

Fast forward: This is what remains after the first merge step (after checking all possible combinations).

A	B
$\{a\}$	$\{b\}$
$\{b\}$	$\{c\}$
$\{b\}$	$\{d\}$
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
$\{e\}$	$\{f\}$
$\{e\}$	$\{g\}$
$\{f\}$	$\{b\}$

1st merge step

A	B
$\{a, f\}$	$\{b\}$
$\{e\}$	$\{f, g\}$

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		→					
c		←		∥	→		
d		←	∥		→		
e			←	←		→	→
f		→			←		
g					←		

Next, check if you can merge any of the new pairs together.

A	B
	$\{b\}$
	$\{c\}$
$\{b\}$	$\{d\}$
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
$\{e\}$	$\{f\}$
$\{e\}$	$\{g\}$
$\{f\}$	$\{b\}$

1st merge step

A	B
$\{a, f\}$	$\{b\}$
$\{e\}$	$\{f, g\}$

2nd merge step

A	B
$\{a, f, e\}$	$\{b, f, g\}$

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		→					
b							
c		←		||	→		
d		←	||		→		
e			←	←		→	→
f		→			←		
g							

You can't! You're done.

A	B
$\}$	$\{b\}$
$\{b\}$	$\{c\}$
$\{b\}$	$\{d\}$
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
$\{e\}$	$\{f\}$
$\{e\}$	$\{g\}$
$\{f\}$	$\{b\}$

1st merge step

2nd merge step

A	B
~~$\{a, f, e\}$~~	~~$\{b, f, g\}$~~

$X_L = \{ (A,B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge$
$\forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a1,a2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b1,b2 \in B}\ b_1 \#_L b_2 \}$

Chair of Process and Data Science

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		$\rightarrow$					
b	$\leftarrow$		$\rightarrow$	$\rightarrow$		$\leftarrow$	
c		$\leftarrow$		$\parallel$	$\rightarrow$		
d		$\leftarrow$	$\parallel$		$\rightarrow$		
e			$\leftarrow$	$\leftarrow$		$\rightarrow$	$\rightarrow$
f		$\rightarrow$			$\leftarrow$		
g					$\leftarrow$		

A	B
$\{a\}$	$\{b\}$
$\{b\}$	$\{c\}$
$\{b\}$	$\{d\}$
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
$\{e\}$	$\{f\}$
$\{e\}$	$\{g\}$
$\{f\}$	$\{b\}$
$\{a, f\}$	$\{b\}$
$\{e\}$	$\{f, g\}$

This is X.

Chair of Process and Data Science

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		$\rightarrow$					
e			$\leftarrow$	$\leftarrow$		$\rightarrow$	$\rightarrow$
f		$\rightarrow$			$\leftarrow$		

To obtain Y, eliminate all pairs A’, B’ where there is another A, B such that A’ is subsumed A and B’ is subsumed by B.

$Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}\ A \subseteq A' \wedge B \subseteq B' \Rightarrow (A,B) = (A',B') \}$,

A	B
$\{a\}$	$\{b\}$
$\{b\}$	$\{c\}$
$\{b\}$	$\{d\}$
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
$\{e\}$	$\{f\}$
$\{e\}$	$\{g\}$
$\{f\}$	$\{b\}$
$\{a,f\}$	$\{b\}$
$\{e\}$	$\{f,g\}$

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		→					
e			←	←		→	→
f		→			←		

To obtain Y, eliminate all pairs A', B' where there is another A, B such that A' is subsumed A and B' is subsumed by B.

$Y_L = \{ (A,B) \in X_L \mid \forall_{(A',B') \in X_L}$
$A \subseteq A' \wedge B \subseteq B' \Rightarrow$
$(A,B) = (A',B') \}$,

A	B
~~$\{a\}$~~ $\subseteq \{a,f\}$	~~$\{b\}$~~ $\subseteq \{b\}$
$\{b\}$	$\{c\}$
[illegible]	[illegible]
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
~~$\{e\}$~~	~~$\{f\}$~~
~~$\{e\}$~~	~~$\{g\}$~~
~~$\{f\}$~~	~~$\{b\}$~~
$\{a,f\}$	$\{b\}$
$\{e\}$	$\{f,g\}$

Example: subsumed by second last row

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Practice (Solution)

Footprint Matrix

	a	b	c	d	e	f	g
a		$\rightarrow$					
b	$\leftarrow$		$\rightarrow$	$\rightarrow$		$\leftarrow$	
c		$\leftarrow$		$\parallel$	$\rightarrow$		
d		$\leftarrow$	$\parallel$		$\rightarrow$		
e			$\leftarrow$	$\leftarrow$		$\rightarrow$	$\rightarrow$
f		$\rightarrow$			$\leftarrow$		
g					$\leftarrow$		

A	B
~~$\{a\}$~~	~~$\{b\}$~~
$\{b\}$	$\{c\}$
$\{b\}$	$\{d\}$
$\{c\}$	$\{e\}$
$\{d\}$	$\{e\}$
~~$\{e\}$~~	~~$\{f\}$~~
~~$\{e\}$~~	~~$\{g\}$~~
~~$\{f\}$~~	~~$\{b\}$~~
$\{a, f\}$	$\{b\}$
$\{e\}$	$\{f, g\}$

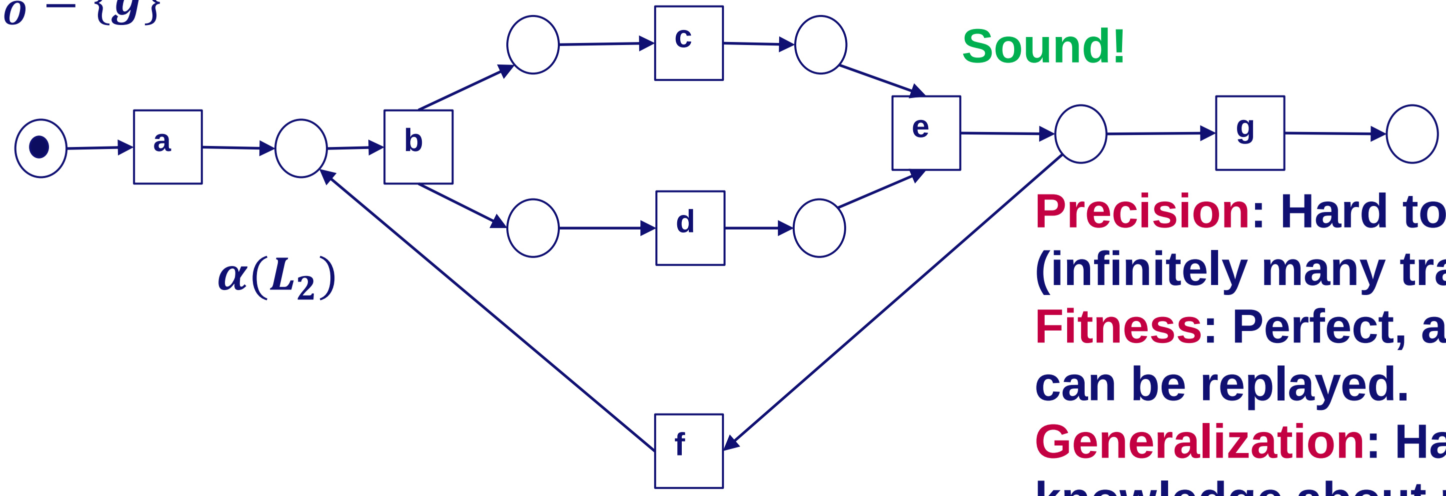
X = every pair listed
Y = every pair listed that is not crossed out

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4b: Alpha Miner – Practice (Solution)

$Y = \{(\{b\},\{c\}),(\{b\},\{d\}),(\{c\},\{e\}),(\{d\},\{e\}),(\{a,f\},\{b\}),(\{e\},\{f,g\})\}$
$T_I = \{a\}$
$T_O = \{g\}$



$\alpha(L_2)$

Sound!

Precision: Hard to reason about (infinitely many traces possible).
Fitness: Perfect, all log traces can be replayed.
Generalization: Hard to argue (no knowledge about real process).

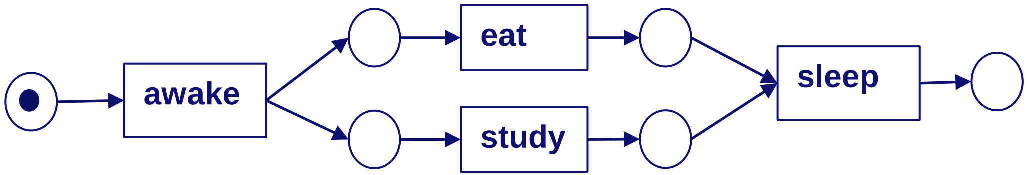
$L_2 = [\langle a,b,c,d,e,f,b,d,c,e,g\rangle, \langle a,b,d,c,e,g\rangle, \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g\rangle]$

Chair of Process and Data Science
RWTH AACHEN UNIVERSITY

Exercise 4c: Alpha Miner – Practice (Solution)

$L_3 = [\langle \textit{awake, eat, study, sleep} \rangle, \langle \textit{awake, study, eat, sleep} \rangle]$

Suggested Model:



Exercise 4c: Alpha Miner – Practice (Solution)

$L_3 = [\langle \textit{awake, eat, study, sleep} \rangle, \langle \textit{awake, study, eat, sleep} \rangle]$

a: awake, e: eat,
s: study, z: sleep

Directly-follows relations

	a	e	s	z
a		>	>	
e			>	>
s		>		>
z				

Footprint Matrix (omit # for brevity)

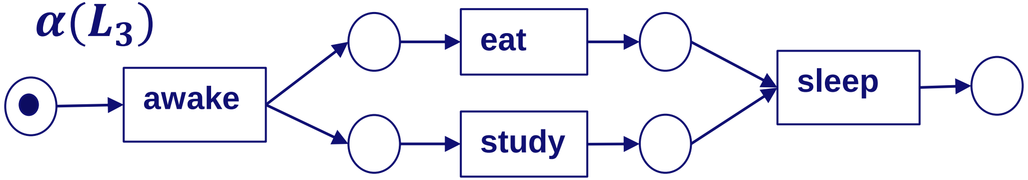
	a	e	s	z
a		→	→	
e	←		‖	→
s	←	‖		→
z		←	←	

A	B
$\{a\}$	$\{e\}$
$\{a\}$	$\{s\}$
$\{e\}$	$\{z\}$
$\{s\}$	$\{z\}$

No merge possible.

$X = Y =$ **every pair listed**

$\alpha(L_3)$



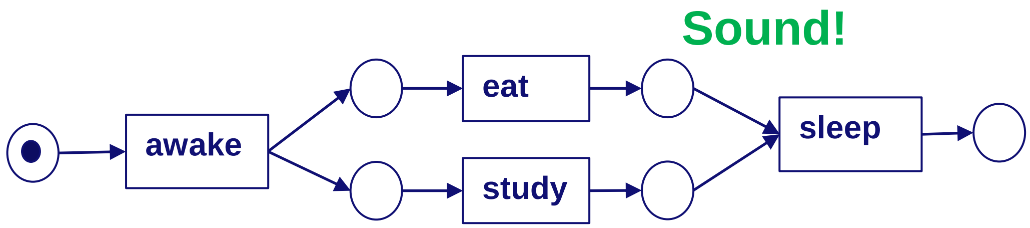
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4c: Alpha Miner – Practice (Solution)

a: awake, e: eat,
s: study, z: sleep

$$Y = \{(\{a\},\{e\}),(\{a\},\{s\}),(\{e\},\{z\}),(\{s\},\{z\})\}$$
$$T_I = \{a\}$$
$$T_O = \{z\}$$



Precision: Perfect, all model traces are in the log.
Fitness: Perfect, all log traces can be replayed.
Generalization: hard to argue (no knowledge about real process).

$L_4 = [\langle$***awake, eat, study, sleep***$\rangle$, $\langle$***awake, study, eat, sleep***$\rangle]$

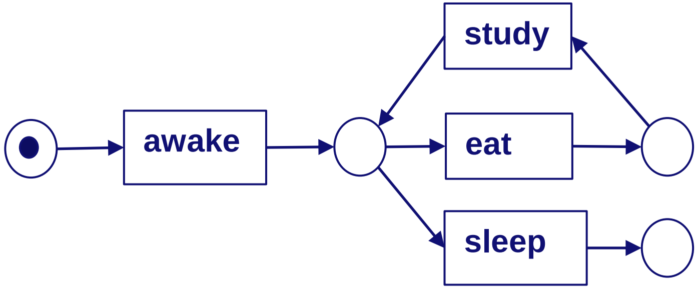
Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4d: Alpha Miner – Practice (Solution)

$L_4 = [\langle \textit{awake, eat, study, eat, study, sleep} \rangle]$

Suggested Model:



Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4d: Alpha Miner – Practice (Solution)

$L_4 = [\langle \textit{awake, eat, study, eat, study, sleep} \rangle]$

a: awake, e: eat, s: study, z: sleep

Directly-follows relations

	a	e	s	z
a		>		
e			>	
s		>		>
z				

Footprint Matrix (omit # for brevity)

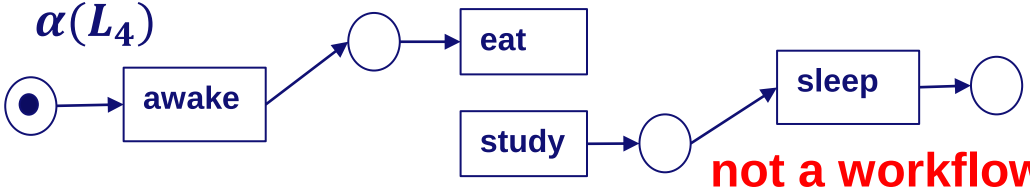
	a	e	s	z
a		$\rightarrow$		
e	$\leftarrow$		$\parallel$	
s		$\parallel$		$\rightarrow$
z			$\leftarrow$	

A	B
$\{a\}$	$\{e\}$
$\{s\}$	$\{z\}$

No merge possible.

$X = Y =$ **every pair listed**

$\alpha(L_4)$

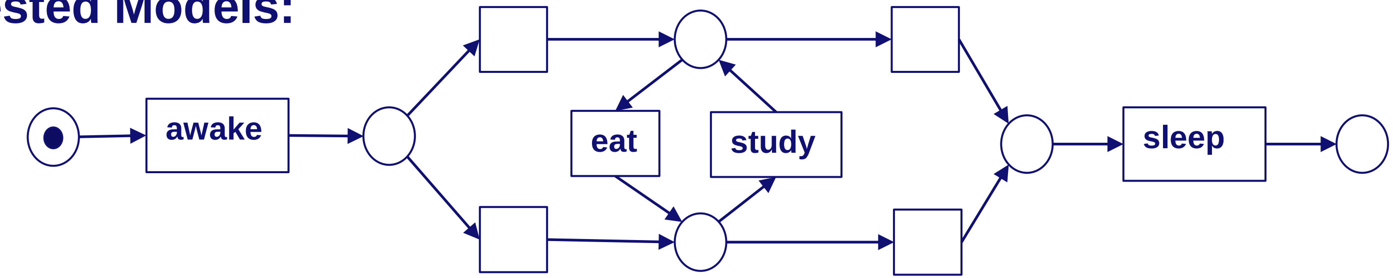


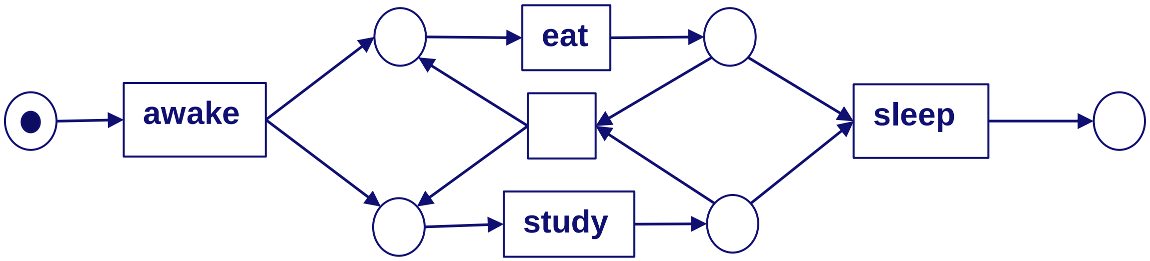
not a workflow net!

Exercise 4e: Alpha Miner – Practice (Solution)

$L_5 = [\langle awake, eat, study, eat, study, sleep\rangle, \langle awake, study, eat, study, eat, sleep\rangle]$

Suggested Models:





Exercise 4e: Alpha Miner – Practice (Solution)

$$L_5 = [\langle \textit{awake, eat, study, eat, study, sleep} \rangle, \langle \textit{awake, study, eat, study, eat, sleep} \rangle]$$

Directly-follows relations

	a	e	s	z
a		>	>	
e			>	>
s		>		>
z				

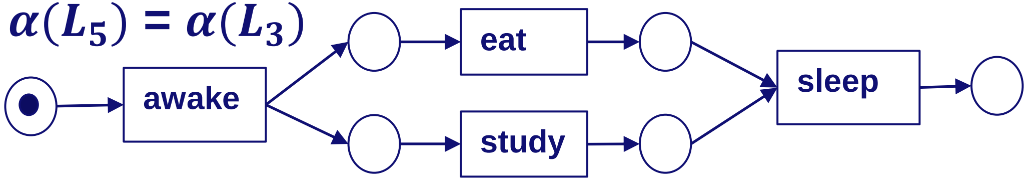
Directly-follows relations L_3

	a	e	s	z
a		>	>	
e			>	>
s		>		>
z				

Directly-follows relations L_4

	a	e	s	z
a		>		
e			>	
s		>		>
z				

$\alpha(L_5) = \alpha(L_3)$



Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 4e: Alpha Miner – Practice (Solution)

$L_5 = [\langle awake, eat, study, eat, study, sleep\rangle, \langle awake, study, eat, study, eat, sleep\rangle]$

Directly-follows relations

	a	e	s	z
a		>	>	
e			>	>
s		>		
z				

Directly-follows relations L_3

	a	e	s	z
e			>	>

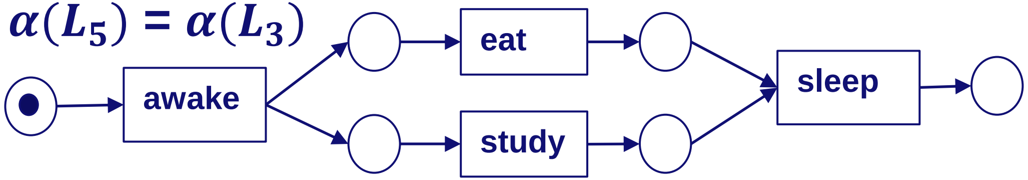
Directly-follows relations L_4

	a	e	s	z
e			>	

$L_3 = [\langle awake, eat, study, sleep\rangle, \langle awake, study, eat, sleep\rangle]$
$L_4 = [\langle awake, eat, study, eat, study, sleep\rangle]$

One can argue that L_5 "looks more like" L_4 than L_3, but the directly-follows relations tell a different story.
Conclusion: We need different discovery techniques.

$\alpha(L_5) = \alpha(L_3)$



Precision: bad.
Fitness: bad.
Generalization: bad.

Exercise 5 – Process Exploration

Hands On ProM and Celonis

Load the provided *delivery_management* event log into both ProM and Celonis to:

	Suggestions / Hints	
	ProM Plugins	**Celonis Sheets**
1. Give an overview of what days are busy (ProM) and the average throughput times of cases (Celonis).	Dotted Chart	Process Overview, (New Sheet with Line Chart)
2. Show the two most frequent variants and their DFGs (both tools).	(Explore Event Log), Inductive Visual Miner	New Sheet with Variant Explorer
3. Discover a model. *Advanced*: Discover a model only for cases that **do not** contain “failed delivery“ (both tools).	(any) Miner, Filter Event Log	Conformance

Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

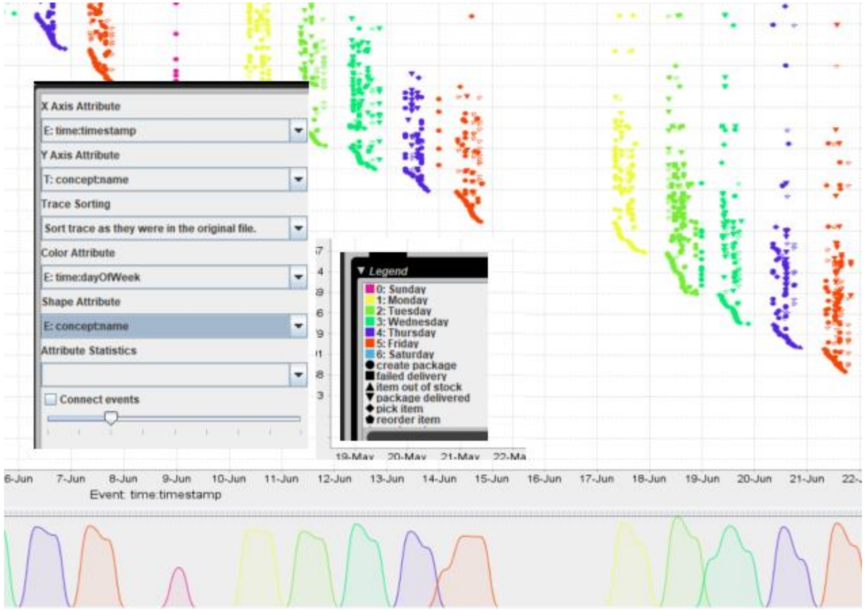
Exercise 5 – Process Exploration (Solution)

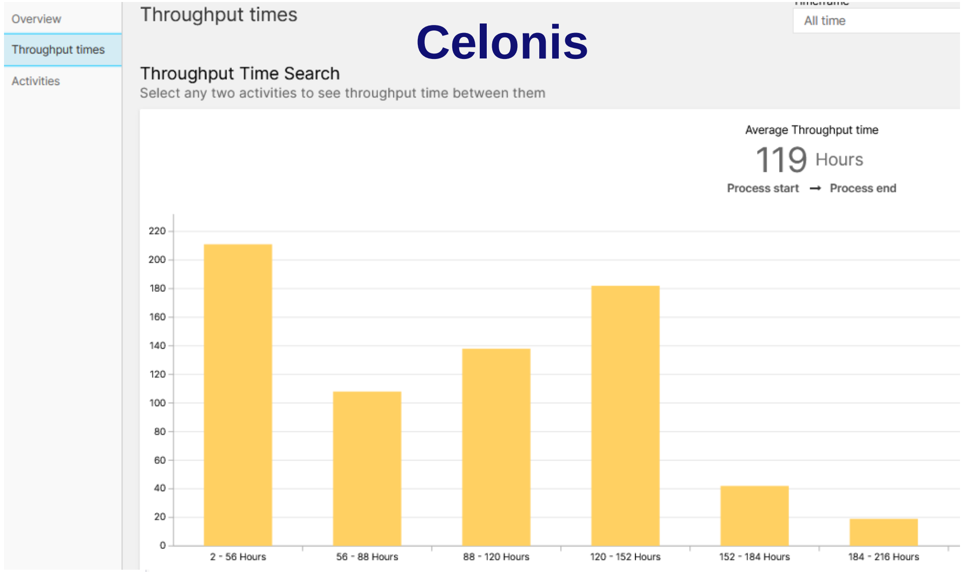
Hands On ProM and Celonis

1. Give an overview of what days are busy…

…and the average throughput times of cases.

ProM





Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Exercise 5 – Process Exploration (Solution)

Hands On ProM and Celonis

2. Show the two most frequent variants and their DFGs.

ProM

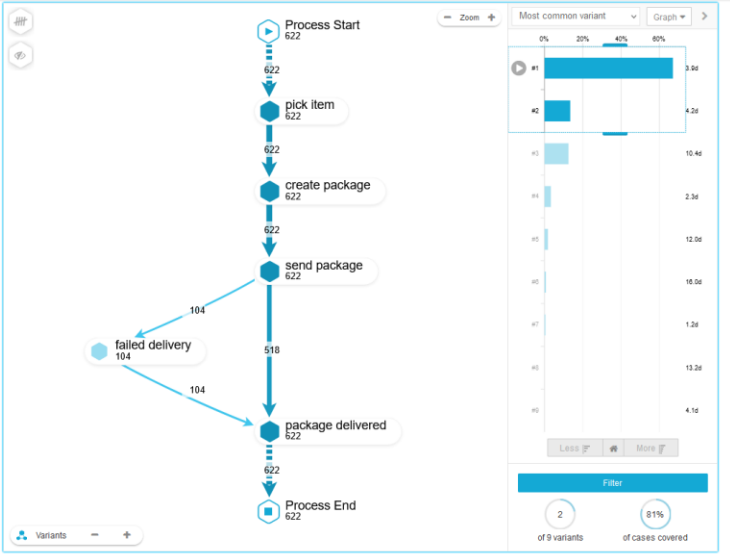
DFG of the filtered log containing only the top 2 variants (correct solution)

DFG of the original log with adapted path threshold (incorrect solution)

Note that the numbers are not the same.

Celonis

Variant Explorer



Chair of Process and Data Science

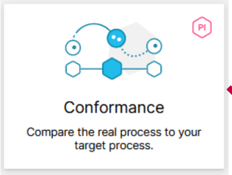
RWTH AACHEN UNIVERSITY

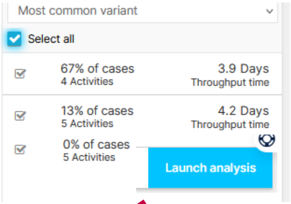
Exercise 5 – Process Exploration (Solution)

Hands On ProM and Celonis

3. Discover a model.

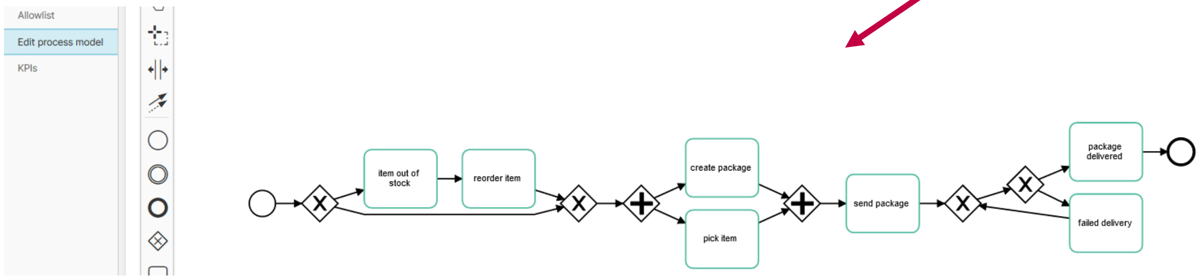
Celonis





ProM

Choose your favorite ☺



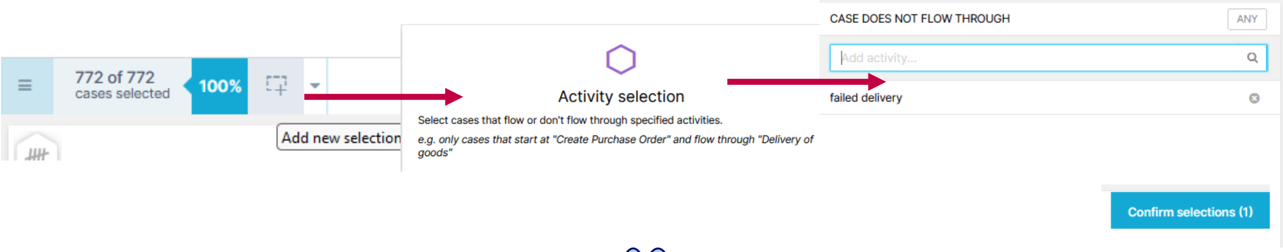
Exercise 5 – Process Exploration (Solution)

Hands On ProM and Celonis

3. *Advanced*: Discover a model only for cases that do not contain “failed delivery“.

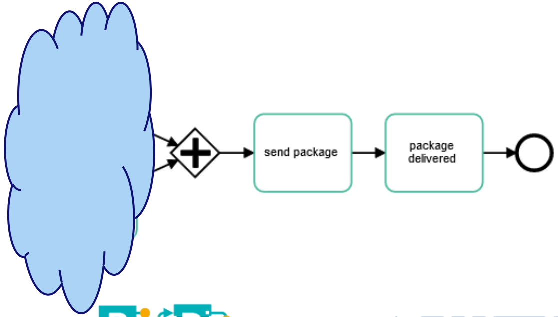
Celonis

ProM



Perform the filtering as shown on the next slide and mine a model.

Repeat the steps shown on the previous slide



Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

ProM: Advanced Filtering
Example