CS 4072 – Topics in CS
Process Mining

Lecture # 11

March 29, 2022

Spring 2022

FAST – NUCES, CFD Campus

Dr. Rabia Maqsood

rabia.maqsood@nu.edu.pk

Today's Topics

- Alpha algorithm: limitations

Alpha Algorithm

- (α-algorithm): Let L be an event log over T $\subseteq \mathcal{A}$. α(L) is defined as follows:

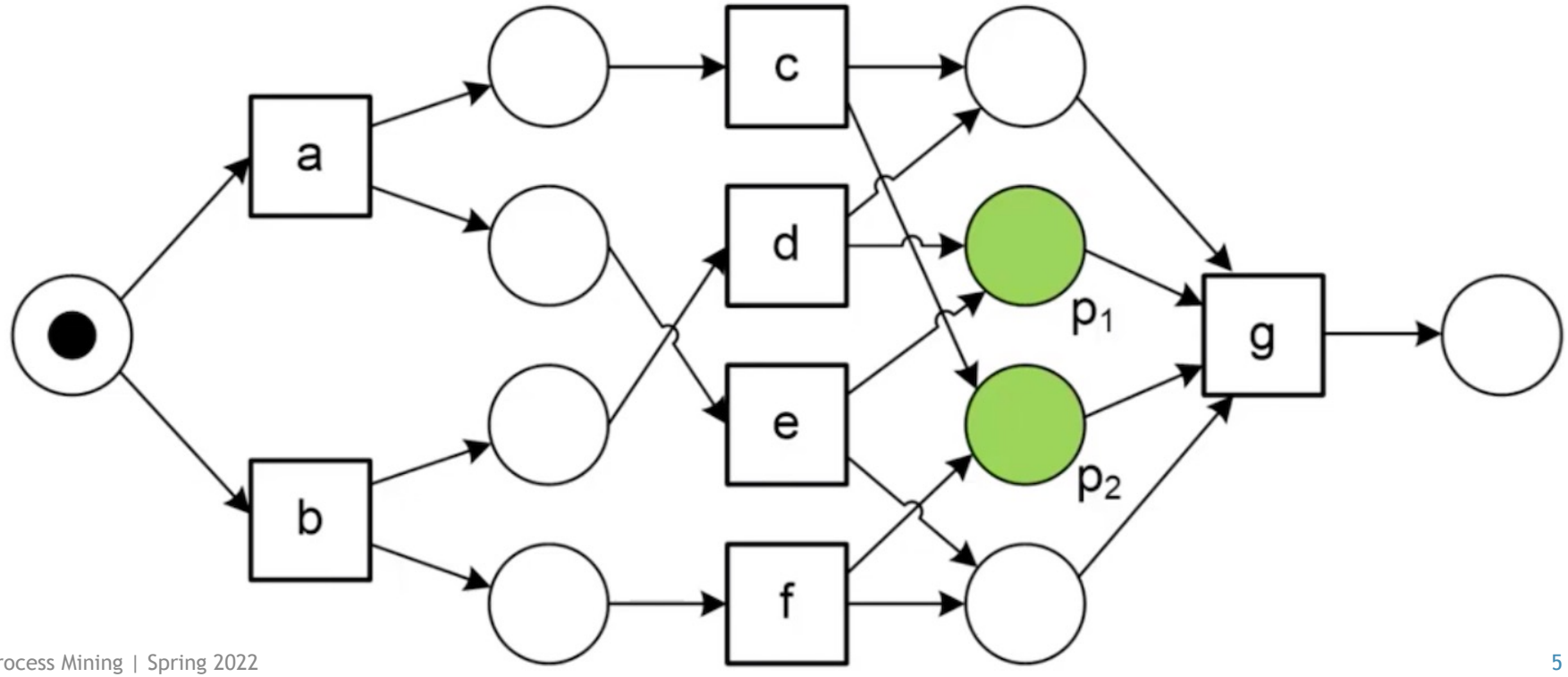
1. $T_L = \{t \in T \mid \exists_{\sigma \in L}\ t \in \sigma\}$,
2. $T_I = \{t \in T \mid \exists_{\sigma \in L}\ t = first(\sigma)\}$,
3. $T_O = \{t \in T \mid \exists_{\sigma \in L}\ t = last(\sigma)\}$,
4. $X_L = \{(A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B}\ a \rightarrow_L b \wedge \forall_{a_1, a_2 \in A}\ a_1 \#_L a_2 \wedge \forall_{b_1, b_2 \in B}\ b_1 \#_L b_2\}$,
5. $Y_L = \{(A, B) \in X_L \mid \forall_{(A', B') \in X_L} A \subseteq A' \wedge B \subseteq B' \Longrightarrow (A, B) = (A', B')\}$,
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. $\alpha(L) = (P_L, T_L, F_L)$.

Limitations of the α-algorithm

- There are many different WF-nets that have the same possible behavior, i.e., two models can be structurally different but trace equivalent.

Limitations of the α-algorithm
Implicit places

$L_6 = [\langle a,c,e,g\rangle^2, \langle a,e,c,g\rangle^3, \langle b,d,f,g\rangle^2, \langle b,f,d,g\rangle^4]$



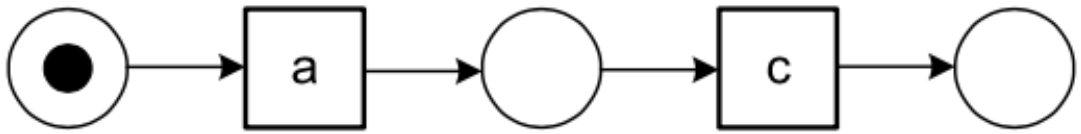
p1 and p2 are implicit places

And can be removed without affecting the set of possible firing sequences.

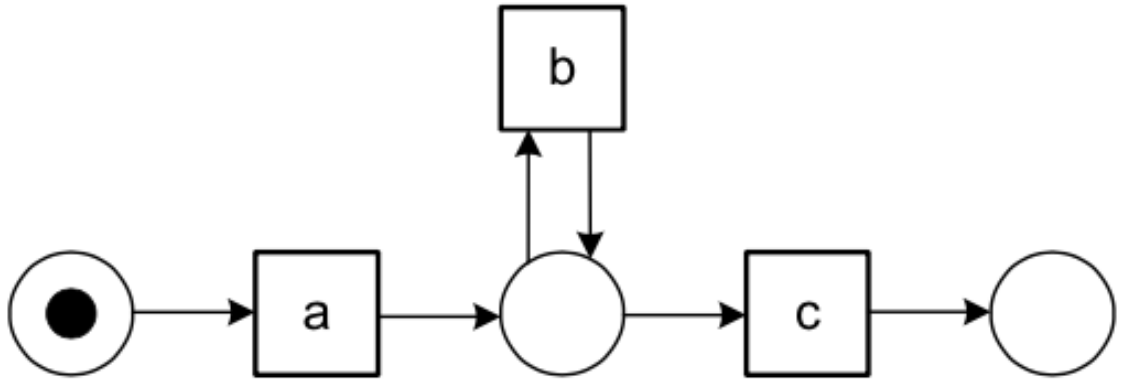
Limitations of the α-algorithm
Loops of length 1

$L_7 = [\langle a,c\rangle^2, \langle a,b,c\rangle^3, \langle a,b,b,c\rangle^2, \langle a,b,b,b,b,c\rangle^1]$

a>b	a→b	b||b	a#a
a>c	a→c		c#c
b>b	b→c		...
b>c			



Incorrect WF-net for L_7



Desired model

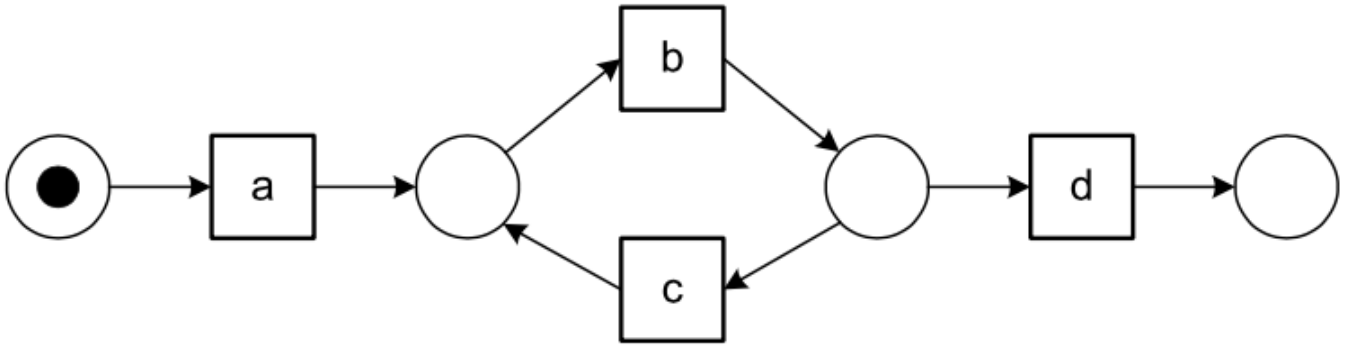
Limitations of the α-algorithm
Loops of length 2

$L_8 = [\langle a,b,d\rangle^3, \langle a,b,c,b,d\rangle^2, \langle a,b,c,b,c,b,d\rangle]$

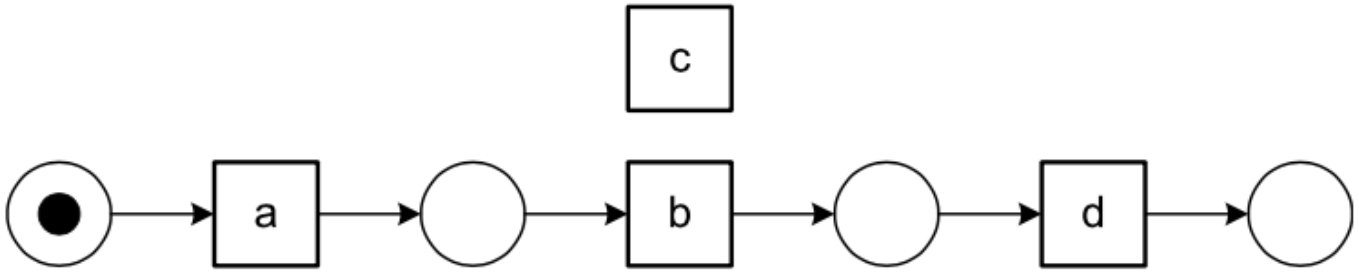
a>b
b>c
b>d
c>b

a→b
b→d

b||c



Desired model

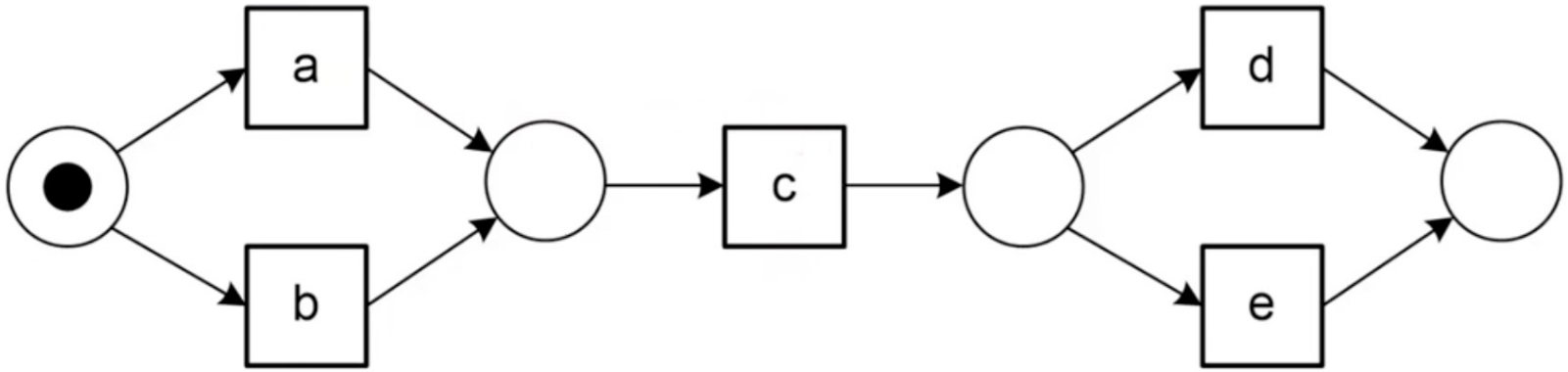


Incorrect WF-net for L_8

Self study:
Alpha algorithm can deal with loops of length>2

Limitations of the α-algorithm
Non-local dependencies

$L_9 = [\langle a,c,d \rangle^{45}, \langle b,c,e \rangle^{42}]$

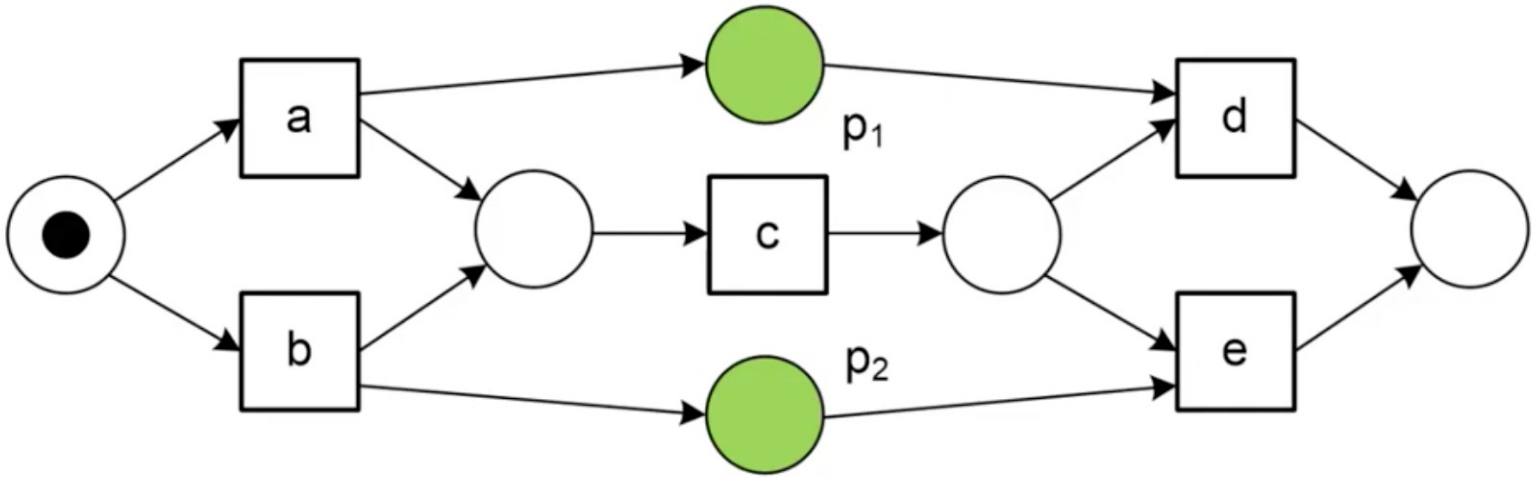


Discovered WF-net for L_9

Any issue?

Limitations of the α-algorithm
Non-local dependencies

$L_9 = [\langle a,c,d\rangle^{45}, \langle b,c,e\rangle^{42}]$



Discovered WF-net for L_9

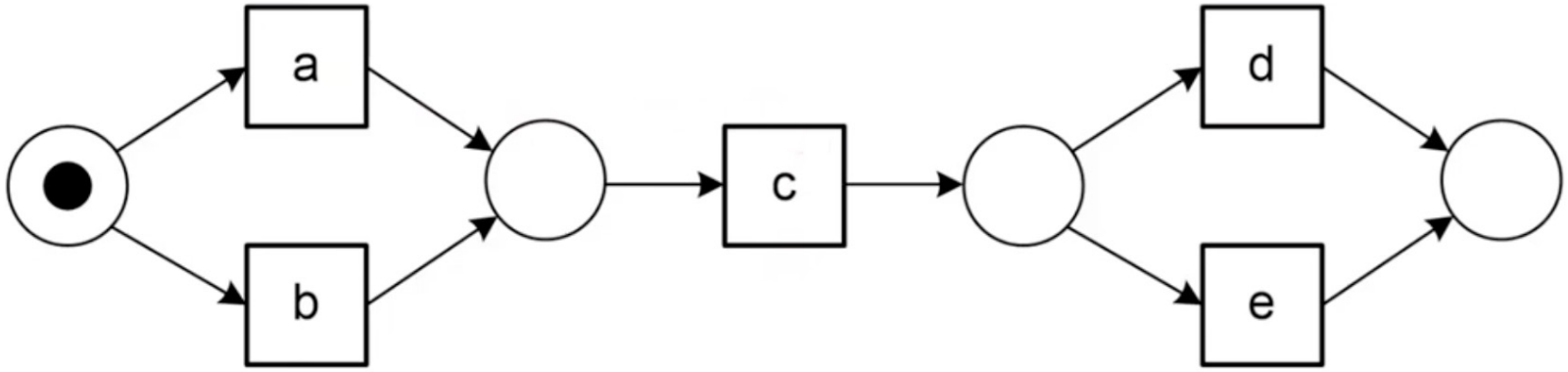
Places p1 and p2 are not discovered

Non-local dependencies result from non-free choice process constructs

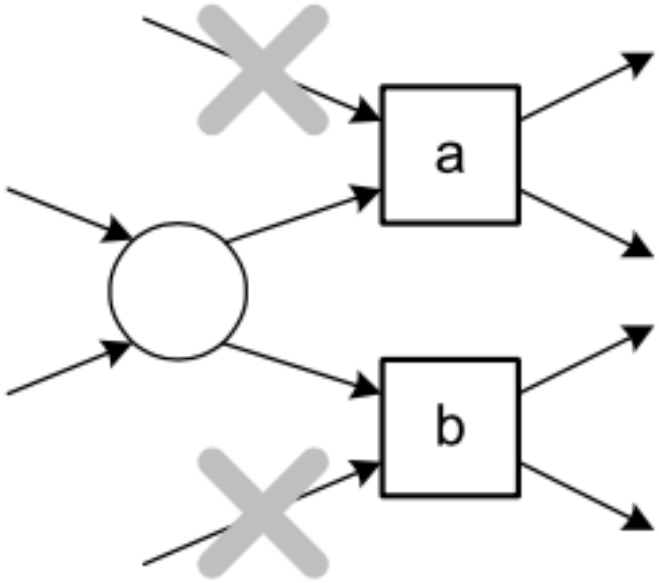
Limitations of the α-algorithm
Two event logs, same discovered model

$L_4 = [\langle a, c, d\rangle^{45}, \langle b, c, d\rangle^{42}, \langle a, c, e\rangle^{38}, \langle b, c, e\rangle^{22}]$

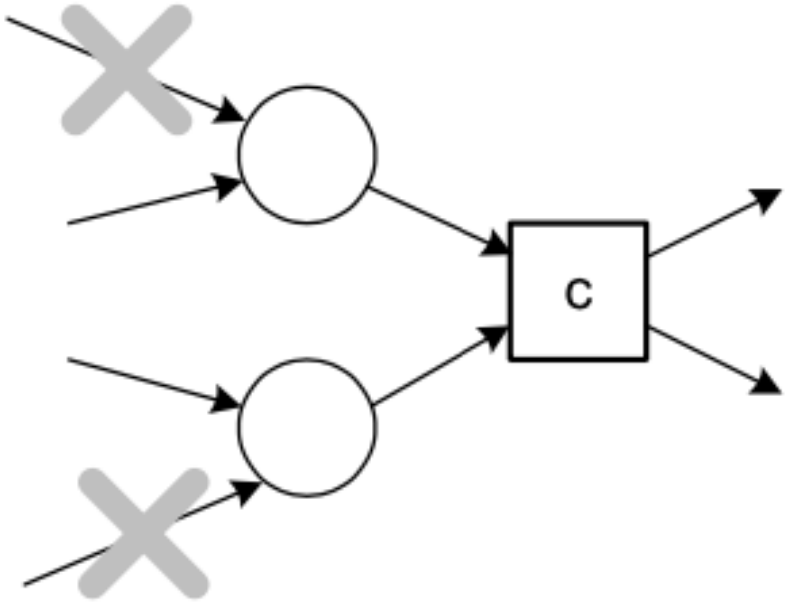
$L_9 = [\langle a,c,d\rangle^{45},\langle b,c,e\rangle^{42}]$



Limitations of the α-algorithm
Difficult constructs



Non-free choice split

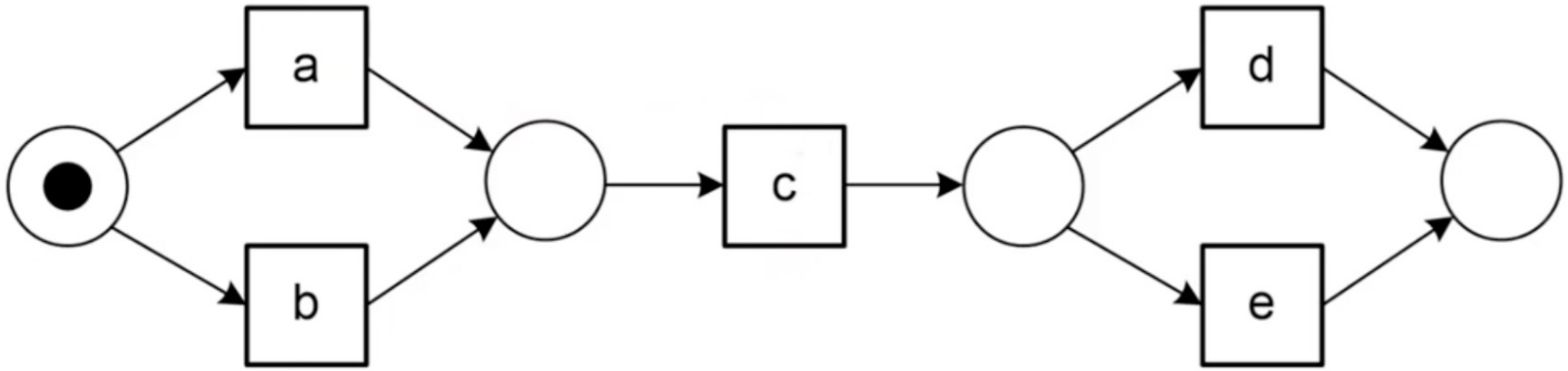


Non-free choice synchronization

Limitations of the α-algorithm
Duplicate transitions

$L = [\langle a, c, d\rangle^{45}, \langle b, c, e\rangle^{42}, \langle a, c, e\rangle^{20}]$

We want to generate a sound WF-net that can produce the observed behavior and nothing more.



Model generated by the Alpha algorithm

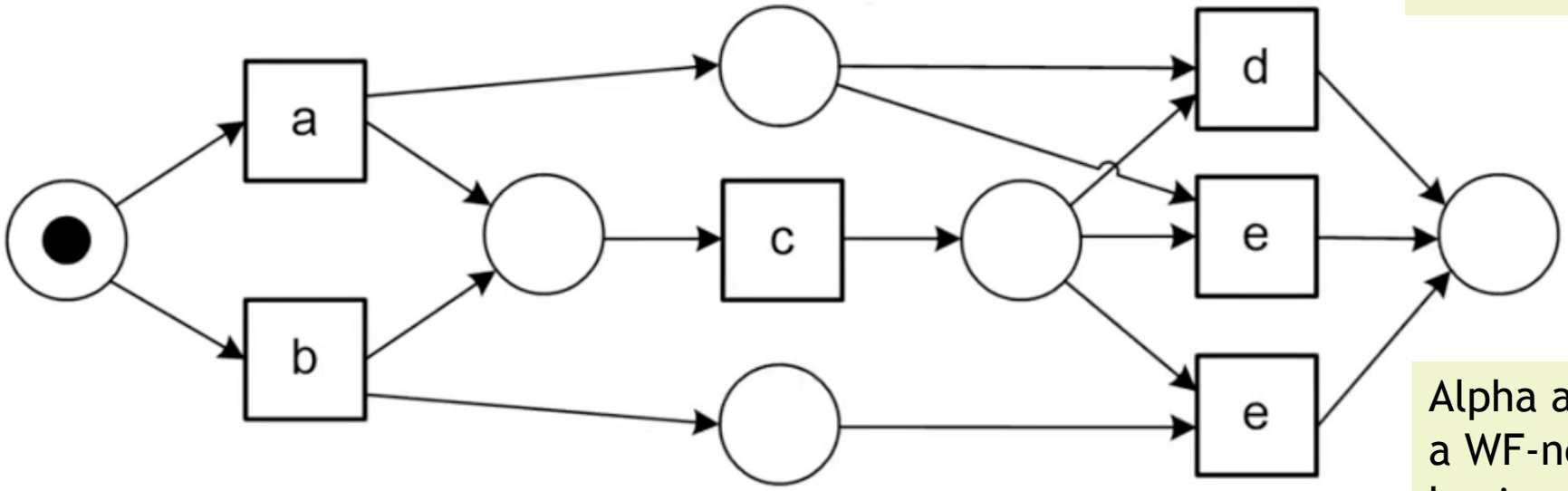
Also allows traces that start with b and ends with d!

Limitations of the α-algorithm
Duplicate transitions

$L = [\langle a, c, d\rangle^{45}, \langle b, c, e\rangle^{42}, \langle a, c, e\rangle^{20}]$

We want to generate a sound WF-net that can produce the observed behavior and nothing more.

Note the duplicate **e** transition



Desired model

Alpha algorithm cannot create a WF-net with two transitions having same label.

Limitations of the α-algorithm
Representational Bias

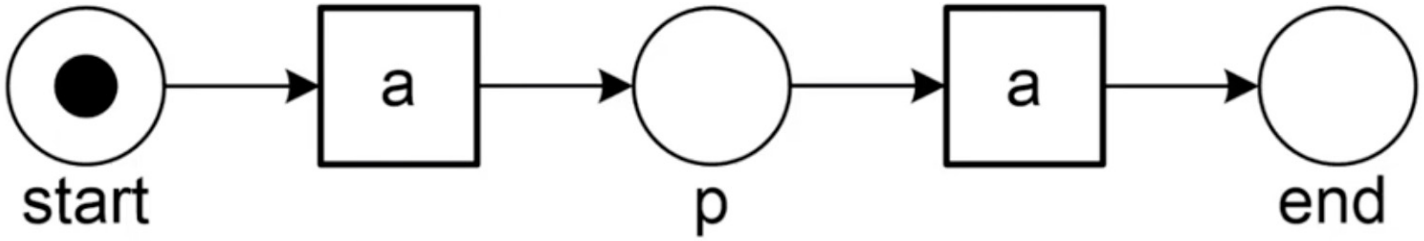
- Any discovery technique requires such a *representational bias*.

- For alpha algorithm, we assumed that the process to be discovered is a sound WF-net
  - Where each transition bears a unique and visible label.

- It is not possible to have two transitions with the same label (i.e., l(t1) = l(t2) implies t1 = t2)
- Or transitions whose occurrences remain invisible (i.e., it is not possible to have a so-called silent transition, so for all transitions t , l(t) ≠ τ)

Limitations of the α-algorithm
Representational Bias

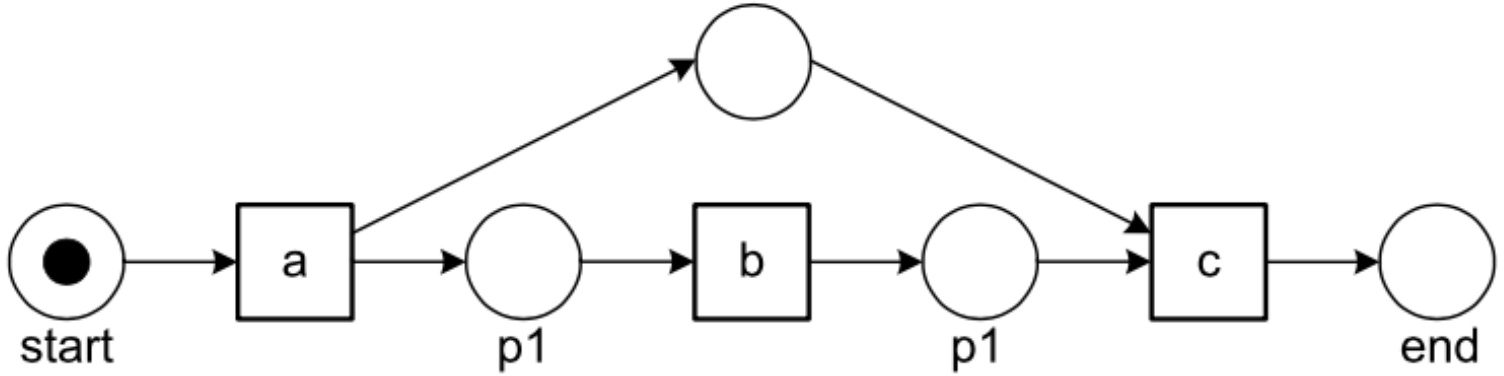
$L_{10} = [\langle a, a \rangle^{55}]$



There is no WF-net with unique visible labels that exhibits this behavior.

Another example

$L_{11} = [\langle a, b, c\rangle^{20}, \langle a, c\rangle^{30}]$

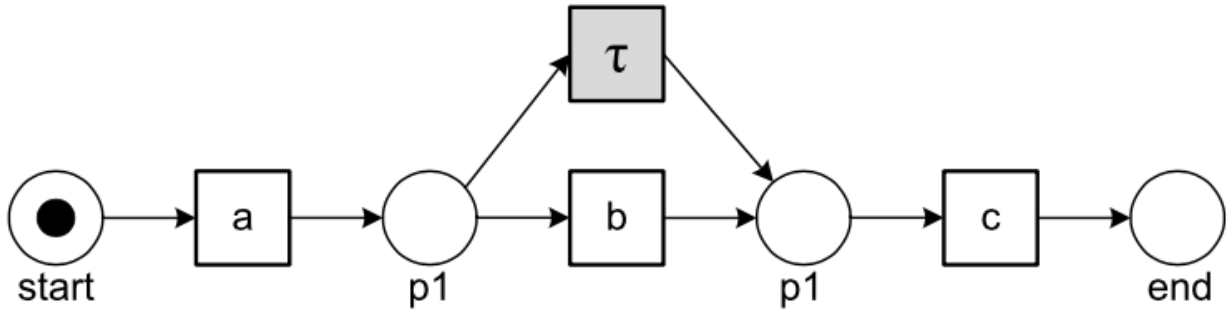


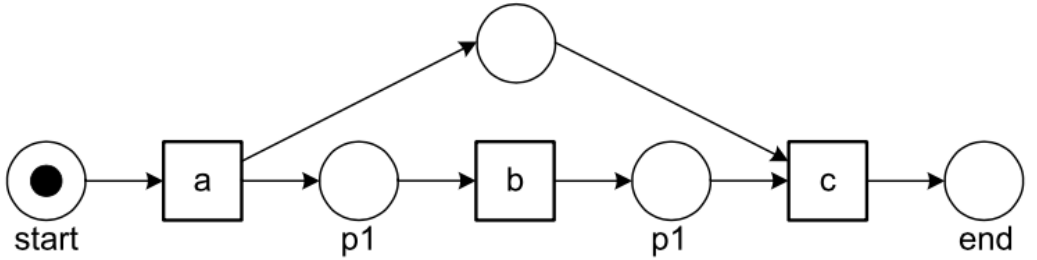
Any issue?

Cannot reproduce <a,c>

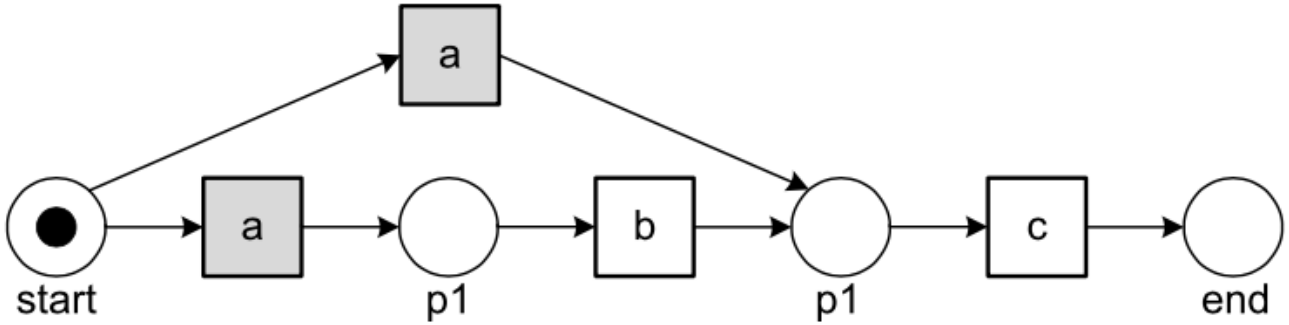
Another example

$L_{11} = [\langle a, b, c\rangle^{20}, \langle a, c\rangle^{30}]$





Cannot reproduce <a,c>



There is no WF-net with unique visible labels that exhibits this behavior.

Limitations of the α-algorithm
Traces' frequencies are not considered

- Resultant issues:
  - Noise
  - Incompleteness

Reading Material

- Chapter 6: Aalst