1 Process discovery Exercises

Exercise 1.1 The following ten event logs were extracted from an enterprise information system.

1. $L_1 = [5 \cdot \langle a,b,c,d\rangle,\ 8 \cdot \langle a,c,b,d\rangle,\ 9 \cdot \langle a,e,d\rangle]$
2. $L_2 = [\langle a,b,c,d,e,f,b,d,c,e,g\rangle,\ \langle a,b,d,c,e,g\rangle,\ \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g\rangle]$
3. $L_3 = [45 \cdot \langle a,c,d\rangle,\ 42 \cdot \langle b,c,d\rangle,\ 38 \cdot \langle a,c,e\rangle,\ 22 \cdot \langle b,c,e\rangle]$
4. $L_4 = [2 \cdot \langle a,b,e,f\rangle,$
   $3 \cdot \langle a,b,e,c,d,b,f\rangle,\ 2 \cdot \langle a,b,c,e,d,b,f\rangle,\ 4 \cdot \langle a,b,c,d,e,b,f\rangle,\ 3 \cdot \langle a,e,b,c,d,b,f\rangle]$
5. $L_5 = [\langle a,d,e,f,h\rangle, \langle a,e,d,f,h\rangle, \langle g,h\rangle, \langle a,b,c,d,f,h\rangle, \langle a,c,b,d,f,h\rangle,$
   $\langle a,b,d,c,f,h\rangle, \langle a,c,d,b,f,h\rangle]$
6. $L_6 = [\langle a,b,c,d,e,f,b,d,c,e,g\rangle, \langle a,b,d,c,e,g\rangle, \langle a,b,c,d,e,f,b,c,d,e,f,b,d,c,e,g\rangle]$
7. $L_7 = [\langle a,b,c,d,f\rangle, \langle a,c,b,d,f\rangle, \langle a,b,d,c,f\rangle, \langle a,c,d,b,f\rangle, \langle a,d,e,f\rangle, \langle a,e,d,f\rangle]$
8. $L_8 = [\langle a,b,c,d,e,f,b,d,c,e,g\rangle, \langle a,b,c,d,e,g\rangle, \langle h,e,g\rangle]$
9. $L_9 = [\langle a,d,f,h\rangle, \langle a,c,e,g,c,e,h\rangle, \langle b,f,g,d,f,h\rangle]$
10. $L_{10} = [\langle a,b,d,e,g,i,j\rangle, \langle a,b,e,d,h,i,j\rangle, \langle a,b,f,d,g,i,j\rangle, \langle a,b,f,g,d,i,j\rangle,$
    $\langle a,b,g,d,f,i,j\rangle, \langle a,b,d,f,h,i,c,b,e,h,d,i,j\rangle, \langle a,b,g,f,d,i,c,b,h,f,d,i,j\rangle,$
    $\langle a,b,h,e,d,i,c,b,d,g,e,i,j\rangle]$
11. $L_{11} = [\langle a,c,d,f\rangle, \langle a,c,d,f\rangle, \langle b,e,f\rangle, \langle a,d,c,f\rangle, \langle b,e,f\rangle, \langle a,d,c,f\rangle, \langle b,e,f\rangle]$
12. $L_{12} = [\langle a,c,d,f\rangle, \langle a,c,d,g\rangle, \langle b,e,f\rangle, \langle a,d,c,g\rangle, \langle b,e,g\rangle, \langle a,d,c,f\rangle, \langle b,e,f\rangle]$
13. $L_{13} = [\langle f,b,g,a,d\rangle, \langle f,c,a,h,i,e\rangle, \langle f,c,h,a,i,e\rangle, \langle f,b,a,g,d\rangle, \langle f,c,h,i,a,e\rangle]$
14. $L_{14} = [\langle d,a,g,f,b,c\rangle, \langle d,a,h,f,b,i,a,g,f,b,c\rangle, \langle e,a,f,h,b,i,a,f,h,b,c\rangle,$
    $\langle e,a,f,g,b,c\rangle, \langle e,a,h,f,b,c\rangle]$

- Derive the $\rightarrow_L$ relation.
- Use the eight steps of the α-algorithm to construct the corresponding Petri net and draw the Petri net (delivering all of the intermediate results is not necessary, only the resulting Petri net is required).
- If possible, give a trace possible according to the discovered model but not (yet) observed in the log.
- Try the Heuristics and Inductive Miner (WARNING: no solution provided for this, but use the ProM tool introduced in the lab to see the solutions.).

Solutions

Solution 1.1

1.
   - $\rightarrow_{L_1} = \{(a,b),(a,c),(a,e),(b,d),(c,d),(e,d)\}$
   - See Figure 1.
   - There is no such trace.

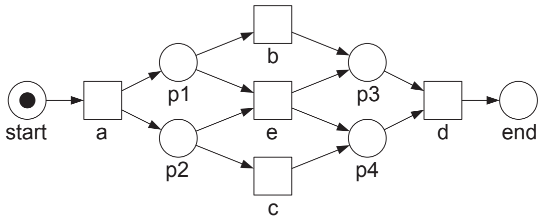


Figure 1: The process model discovered from L_1.

2.
   - $\rightarrow_{L_2} = \{(a,b),(b,c),(b,d),(c,e),(d,e),(e,f),(e,g),(f,b)\}$
   - See Figure 2.
   - For example, $\langle a,b,d,c,e,f,b,c,d,e,g\rangle$.

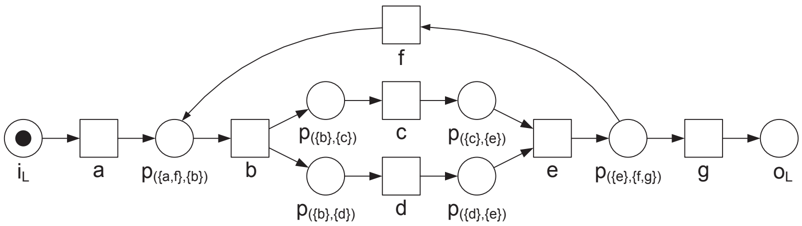


Figure 2: The process model discovered from L_2.

3.
   - $\rightarrow_{L_3} = \{(a,c),(b,c),(c,d),(c,e)\}$
   - See Figure 3.
   - There is no such trace.

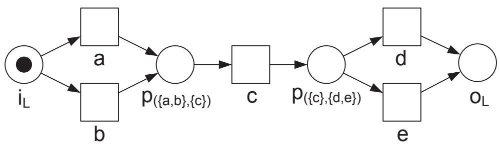


Figure 3: The process model discovered from L_3.

4.
   - $\rightarrow_{L_4} = \{(a,b),(a,e),(b,c),(b,f),(c,d),(d,b),(e,f)\}$
   - See Figure 4.
   - For example, $\langle a,e,b,f\rangle$.

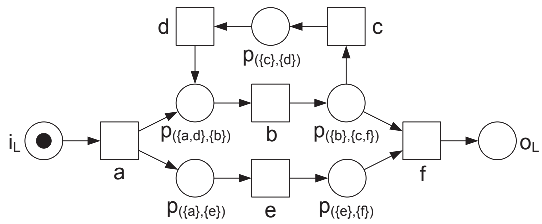


Figure 4: The process model discovered from L_4.

5. - $\rightarrow_{L_5} = \{(a,b),(a,c),(a,d),(a,e),\ (b,f),(c,f),(d,f),(e,f),(f,h),(g,h)\}$
   - See Figure 5.
   - $\langle a,d,b,c,f,h\rangle$ and $\langle a,d,c,b,f,h\rangle$.

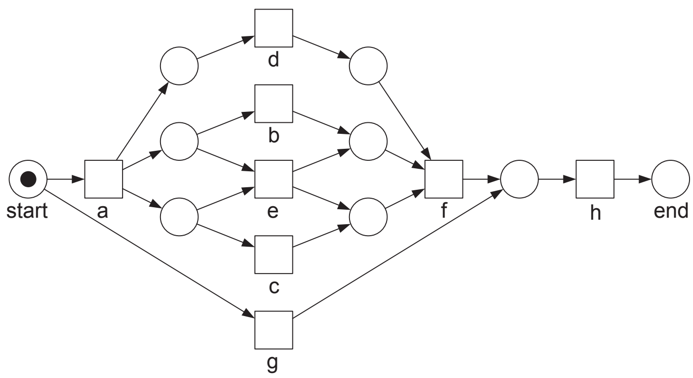


Figure 5: The process model discovered from L_5.

6. - $\rightarrow_{L_6} = \{(a,b),(b,c),(b,d),(c,e),(d,e),(e,f),(e,g),(f,b)\}$
   - See Figure 6.
   - For example, $\langle a,b,c,d,e,g\rangle$.

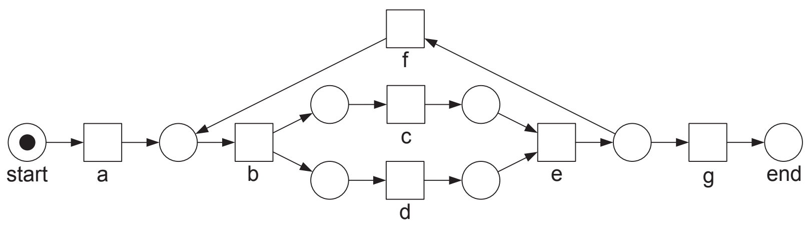


Figure 6: The process model discovered from L_6.

7. - $\rightarrow_{L_7} = \{(a,b),(a,c),(a,d),(a,e),(b,f),(c,f),(d,f),(e,f)\}$
   - See Figure 7.
   - For example, $\langle a,d,b,c,f\rangle$ and $\langle a,d,c,b,f\rangle$.

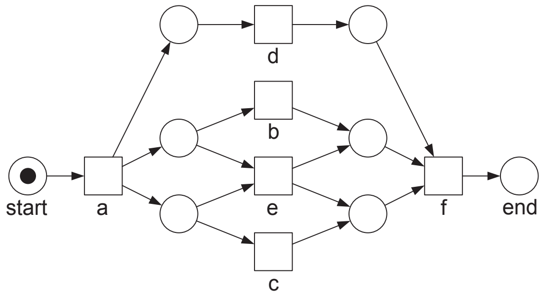


Figure 7: The process model discovered from L_7.

8. - $\rightarrow_{L_8} = \{(a,b),(b,c),(b,d),(c,e),(d,e),(e,f),(e,g),(f,b),(h,e)\}$
   - See Figure 8.
   - For example, $\langle h,e,f,b,c,d,e,g\rangle$.

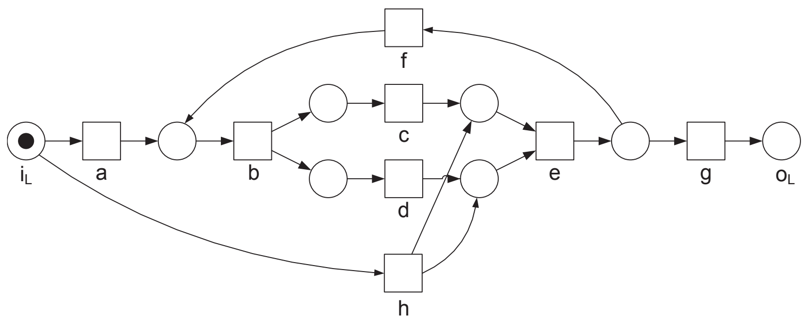


Figure 8: The process model discovered from L_8.

9. - $\rightarrow_{L_9} = \{(a,c),(a,d),(b,f),(c,e),(d,f),(e,g),(e,h),(f,g),(f,h),(g,c),(g,d)\}$
   - See Figure 9.
   - For example, $\langle b,f,g,c,e,h\rangle$.

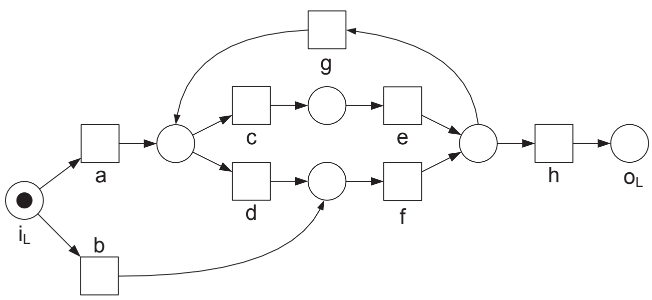


Figure 9: The process model discovered from L_9.

10. - $\rightarrow_{L_{10}} = \{(a,b),(b,d),(b,e),(b,f),(b,g),(b,h),(c,b),(d,i),(e,i),(f,i)$
    $,(h,i),(g,i),(i,c),(i,j)\}$
    - See Figure 10.
    - For example, $\langle a,b,d,h,e,i,j\rangle$.

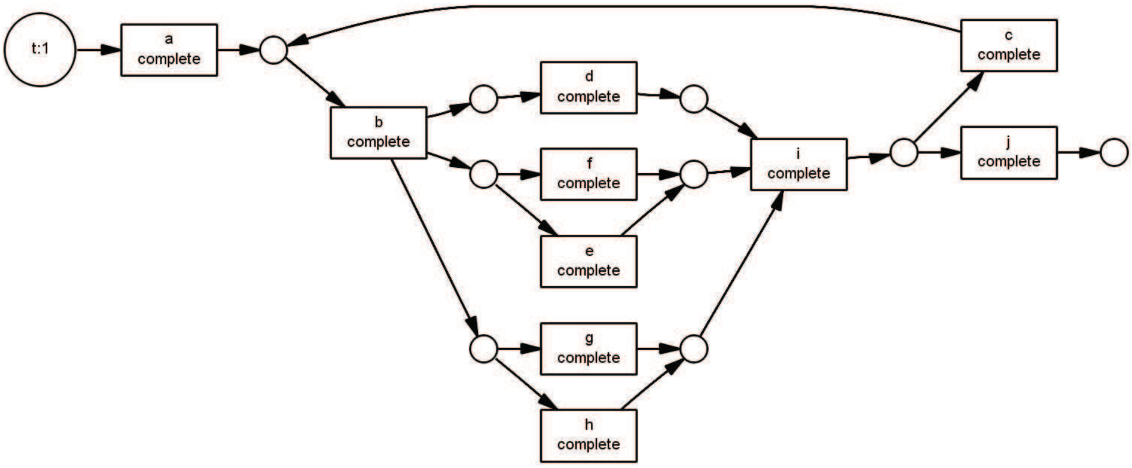


Figure 10: The process model discovered from L_{10}.

11.
- $\rightarrow_{L_{11}} = \{(a,c),(a,d),(b,e),(c,f),(d,f),(e,f)\}$
- See Figure 11.
- There is no such trace.

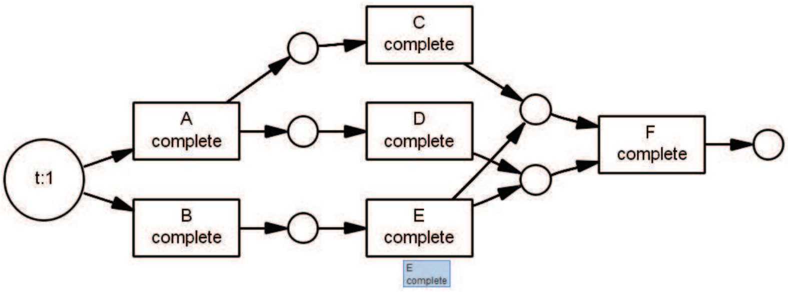


Figure 11: The process model discovered from L_{11}.

12.
- $\rightarrow_{L_{12}} = \{(a,c),(a,d),(b,e),(c,f),(c,g),(d,f),(d,g),(e,f),(e,g)\}$
- See Figure 12.
- There is no such trace.

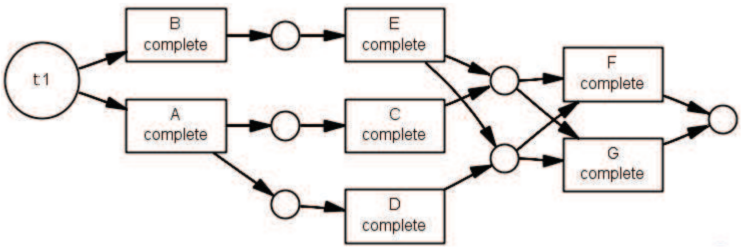


Figure 12: The process model discovered from L_{12}.

13.
- $\rightarrow_{L_{13}} = \{(f,b),(f,c),(b,g),(b,a),(c,a),(c,h),(g,d),(a,d),(h,i),(a,e),(i,e)\}$
- See Figure 13.

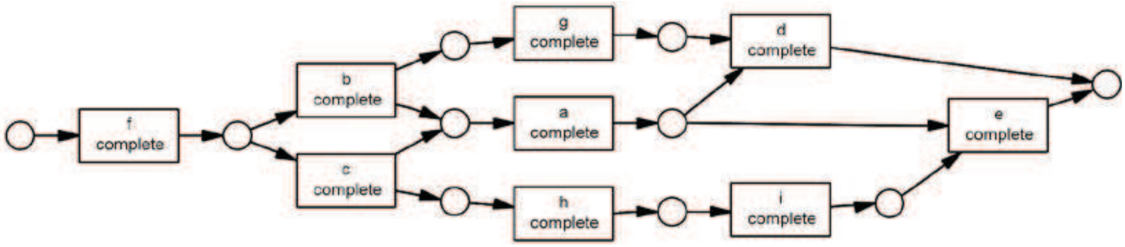


Figure 13: The process model discovered from L_{13}.

- There is no such trace.

14. - $\rightarrow_{L_{14}} = \{(d,a),(e,a),(a,g),(a,h),(a,f),(g,b),(h,b),(f,b),(b,i),(b,c),(i,a)\}$
    - See Figure 14.
    - For example $\langle d,a,g,f,b,i,a,g,f,b,c\rangle$.

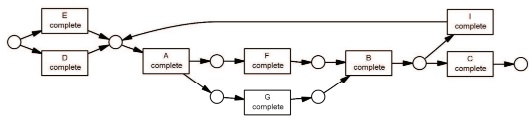
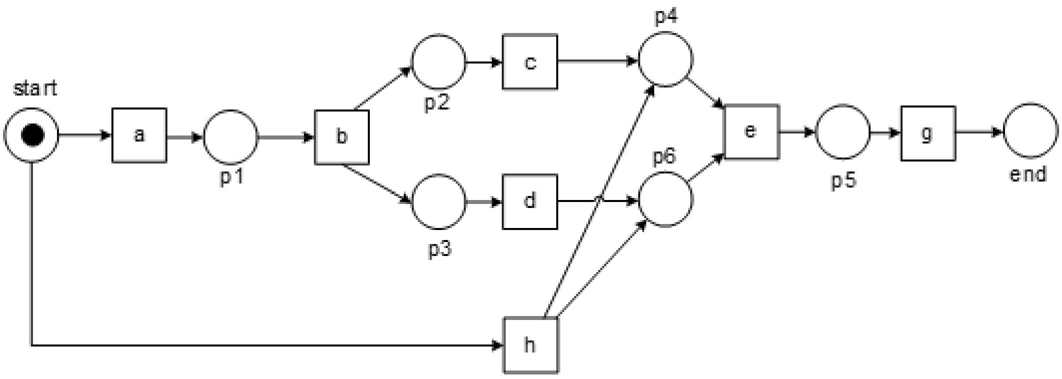


Figure 14: The process model discovered from L_{14}.

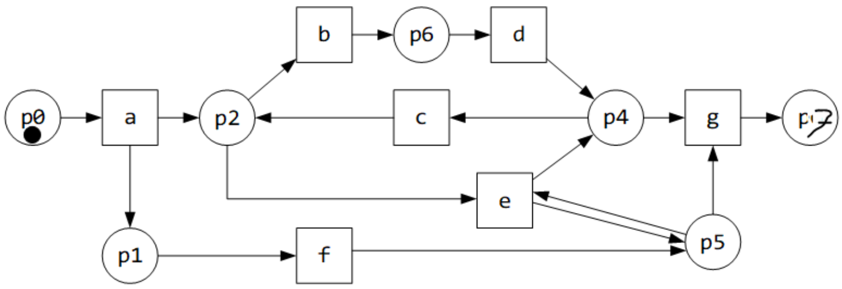
2 Conformance Checking

Exercise 2.1 Consider the following model:



Moreover, consider the following event log: $L = [\langle a, g\rangle^3, \langle h, e, g\rangle^3]$. Compute fitness for the given event log and the given model using *token-based fitness*.

Exercise 2.2 Consider event log $L = [\langle < a, b, e, d, g\rangle^{10}, \langle a, b, f, d, g\rangle^4]$ and the following Petri net:



1. Compute the fitness for the given log on the given model using token-based fitness.
2. Illustrate an example of a potential trace where the number of tokens that is consumed is not equal to the number of tokens that are produced.
3. For the following trace, select the appropriate optimal alignment (a), (b) or (c). Assume the default cost function assigning cost 1 to each deviation. $\langle a, c, e, f, e, g\rangle$

(a)

log trace	a	$\gg$	$\gg$	c	e	f	e	g
execution sequence	a	b	d	c	$\gg$	f	e	g

(b)

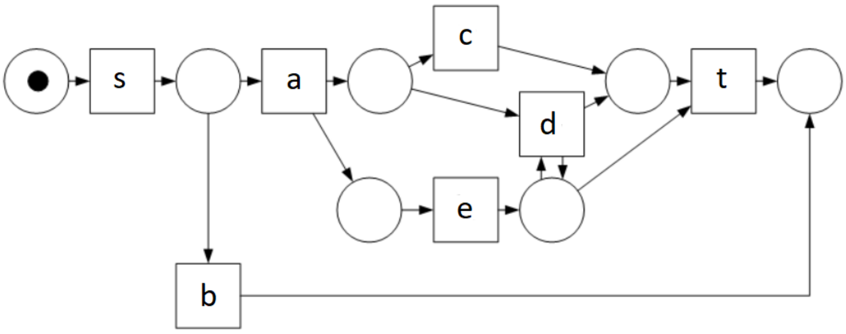
log trace	a	c	e	f	e	g
execution sequence	a	$\gg$	$\gg$	f	e	g

(c)

log trace	a	$\gg$	$\gg$	e	f	e	g
execution sequence	a	b	d	e	f	e	g

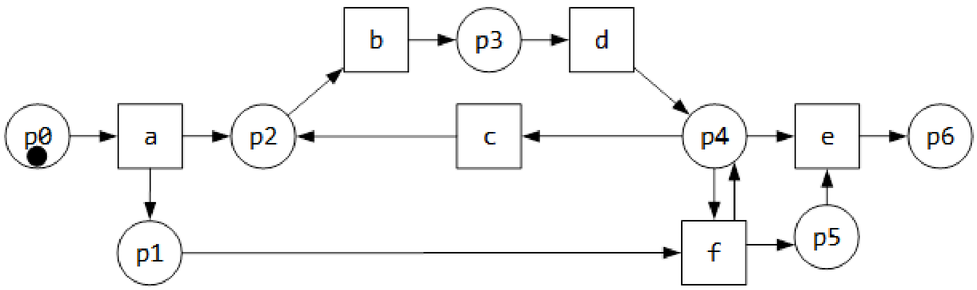
4. For the following trace, draw the synchronous product and use it to find the appropriate optimal alignment. Assume the default cost function assigning cost 1 to each deviation. $\langle a, b, d, c, e, g\rangle$

Exercise 2.3 Consider event log $L = [\langle s, a, c, e, t\rangle^5, \langle a, e, d, t\rangle^3, \langle s, a, e, t\rangle^4, \langle s, b\rangle^{10}]$ and the following Petri net:



1. Compute fitness of the event log with respect to the model using token-based fitness.
2. Give an optimal alignment for the trace $\langle a, b, d, c, d, e, t\rangle$, and, its corresponding fitness. Use the standard cost function, i.e. log/model moves have cost 1, synchronous moves have value 0.

Exercise 2.4 Consider event log $L = [\langle a, b, e, d, c\rangle^{10}, \langle a, b, d, f, e\rangle^5]$ and the following Petri net:



1. Compute the fitness for the given log on the given model using token-based fitness.
2. Illustrate an example of a potential trace where the number of tokens that is consumed is not equal to the number of tokens that are produced.
3. For the following trace, select the appropriate optimal alignment (a), (b) or (c). Assume the default cost function assigning cost 1 to each deviation. $\langle a, c, e, f, e, g\rangle$

(a)

log trace	a	≫	≫	c	e	f	e	g
execution sequence	a	b	d	≫	≫	f	e	≫

(b)

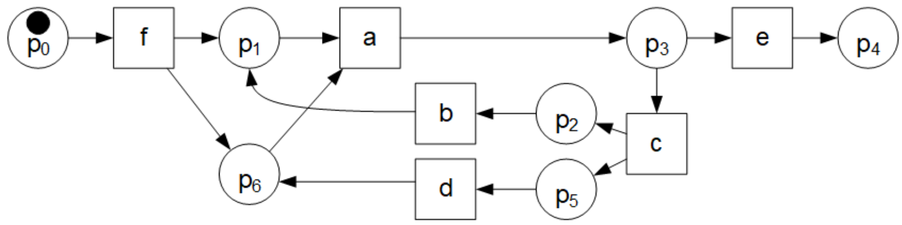
log trace	a	c	e	f	e	g
execution sequence	a	c	≫	f	e	≫

(c)

log trace	a	≫	≫	c	e	≫	≫	f	e	g
execution sequence	a	b	d	c	≫	b	d	f	e	≫

4. For the following trace, provide the synchronous product and use it to compute an optimal alignment. Assume the default cost function assigning cost 1 to each deviation. $\langle a, b, f, c, d, e\rangle$

Exercise 2.5 Consider event log $L = [\langle f, a, b, d, a\rangle^{10}, \langle f, a, c, b, e\rangle^5]$ and the following Petri net:



1. Compute the fitness for the given log on the given model using token-based fitness.

2. Illustrate an example of a potential trace where the number of tokens that is consumed is not equal to the number of tokens that are produced.

3. For the following trace, select the appropriate optimal alignment (a), (b) or (c). Assume the default cost function assigning cost 1 to each deviation. $\langle f, a, c, a, e\rangle$

(a)

log trace	f	a	c	≫	≫	a	e
execution sequence	f	a	c	b	d	a	e

(b)

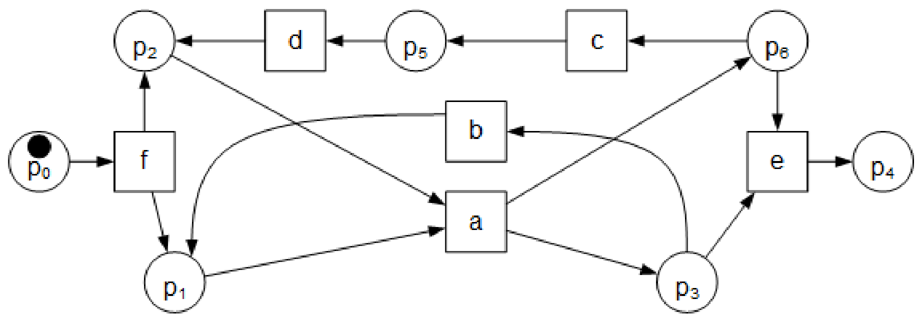
log trace	f	a	c	≫	a	e
execution sequence	f	a	c	d	a	e

(c)

log trace	f	a	c	a	e
execution sequence	f	≫	≫	≫	e

4. For the following trace, compute the synchronous product and use it to compute an optimal alignment. Assume the default cost function assigning cost 1 to each deviation. $\langle a, d, b, e, g\rangle$

Exercise 2.6 Consider event log $L = [\langle f, a, b, c, d, a, e\rangle^{10}, \langle f, a, c, b, a\rangle^{5}]$ and the following Petri net:



1. Compute the fitness for the given log on the given model using token-based fitness.

2. Illustrate an example of a potential trace where the number of tokens that is consumed is not equal to the number of tokens that are produced.

3. For the following trace, select the appropriate optimal alignment (a), (b) or (c). Assume the default cost function assigning cost 1 to each deviation. $\langle f, a, c, e\rangle$

(a)

log trace	f	a	c	≫	≫	≫	e
execution sequence	f	a	c	b	d	a	e

(b)

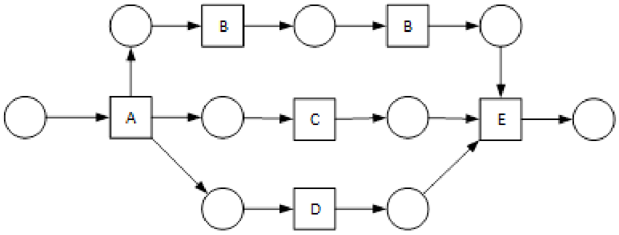
log trace	f	a	c	≫	e
execution sequence	f	a	c	d	e

(c)

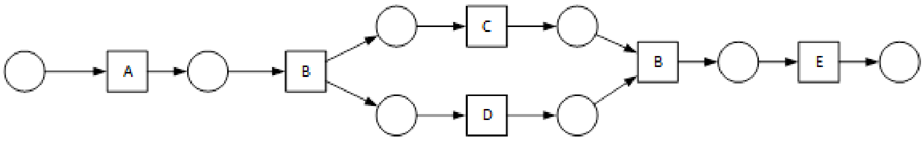
log trace	f	a	c	e
execution sequence	f	a	≫	e

4. For the following trace, provide the synchronous product and use it to compute an optimal alignment. Assume the default cost function assigning cost 1 to each deviation. $\langle a, c, d, a, b, a, e\rangle$

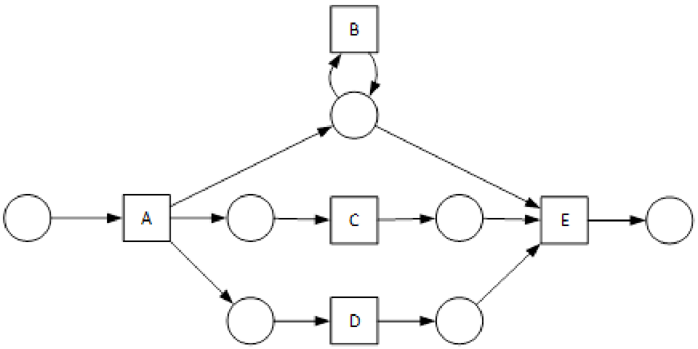
Exercise 2.7 Consider event log $L = [\langle a, b, c, d, b, e\rangle^{200}, \langle a, b, d, c, b, e\rangle^{100}, \langle a, d, b, c, e\rangle^{2}, \langle a, d, b, b, c, e\rangle^{2}]$ and the following Petri nets:



(A)



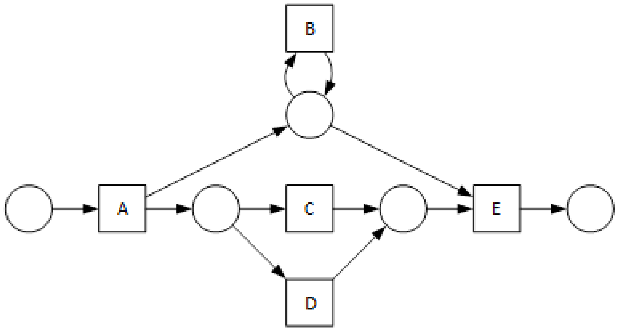
(B)



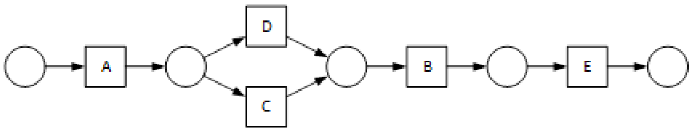
(C)

Rank these models from most fitting to least fitting with respect to log L.

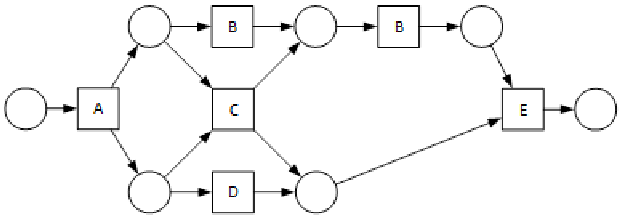
Exercise 2.8 Consider event log $L = [\langle a, b, d, b, e\rangle^{200}, \langle a, c, b, e\rangle^{100}, \langle a, b, b, d, e\rangle^{3}, \langle a, b, c, e\rangle^{4}]$ and the following Petri nets:



(A)



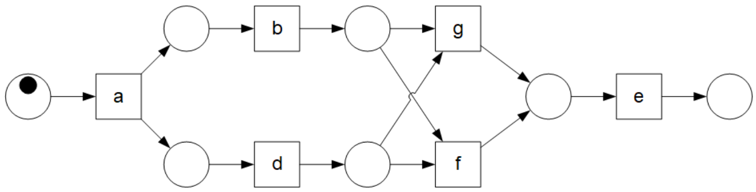
(B)



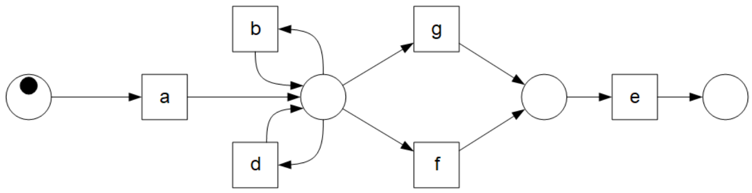
(C)

Rank these models from most fitting to least fitting with respect to log L.

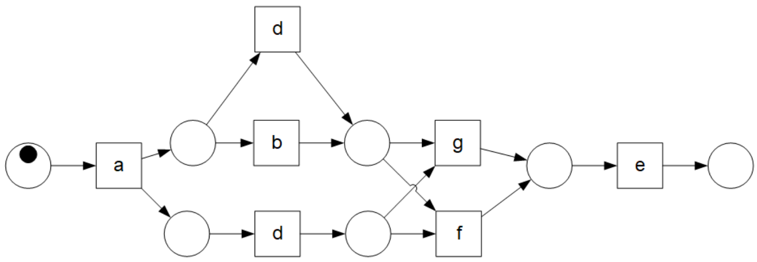
Exercise 2.9 Consider event log $L = [\langle a,b,d,f,e\rangle^{200}, \langle a,d,b,g,e\rangle^{50}, \langle a,d,d,g,e\rangle^{5}, \langle a,g,e\rangle^{2}]$ and the following Petri nets:



(A)



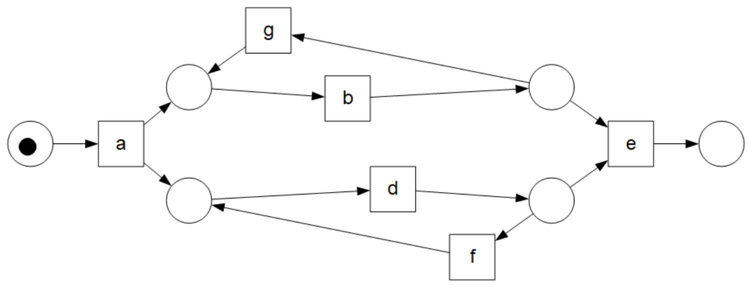
(B)



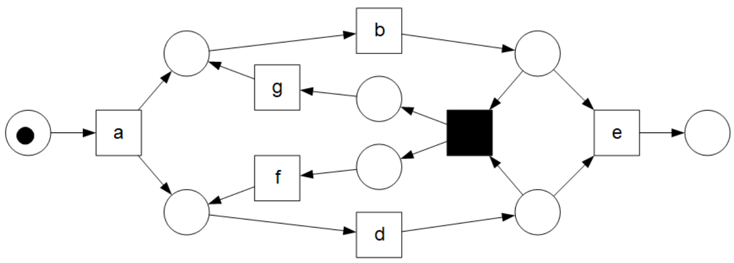
(C)

Rank these models from most fitting to least fitting with respect to log L.

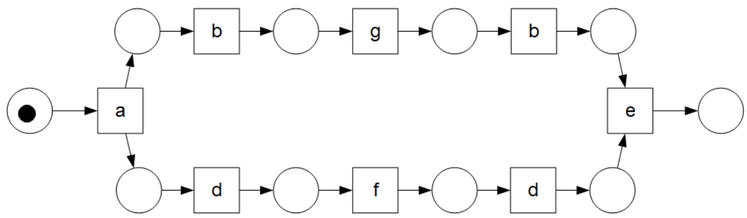
Exercise 2.10 Consider event log $L = [\langle a, b, d, f, e\rangle^{200}, \langle a, d, b, g, e\rangle^{50}, \langle a, d, d, g, e\rangle^{5}, \langle a, g, e\rangle^{2}]$ and the following Petri nets (note that one model contains an 'invisible' transition, i.e. a transition which does not correspond to an activity logged in the event log):



(A)



(B)



(C)

Rank these models from most fitting to least fitting with respect to log L.

Solutions

Solution 2.1

The table for token-based fitness is depicted below:

Token based fitness:

Trace	Freq	P	R	C	M	P (all)	R (all)	C (all)	M (all)
$\langle a, g\rangle$	3	3	$1^{(p_1)}$	3	$1^{(p_5)}$	9	3	9	3
$\langle h, e, g\rangle$	3	4	0	5	0	50	0	50	0
total						**59**	**3**	**59**	**3**

fitness $= \frac{1}{2}(1 - 3/59) + \frac{1}{2}(1 - 3/59) = 0.9492$

Solution 2.2

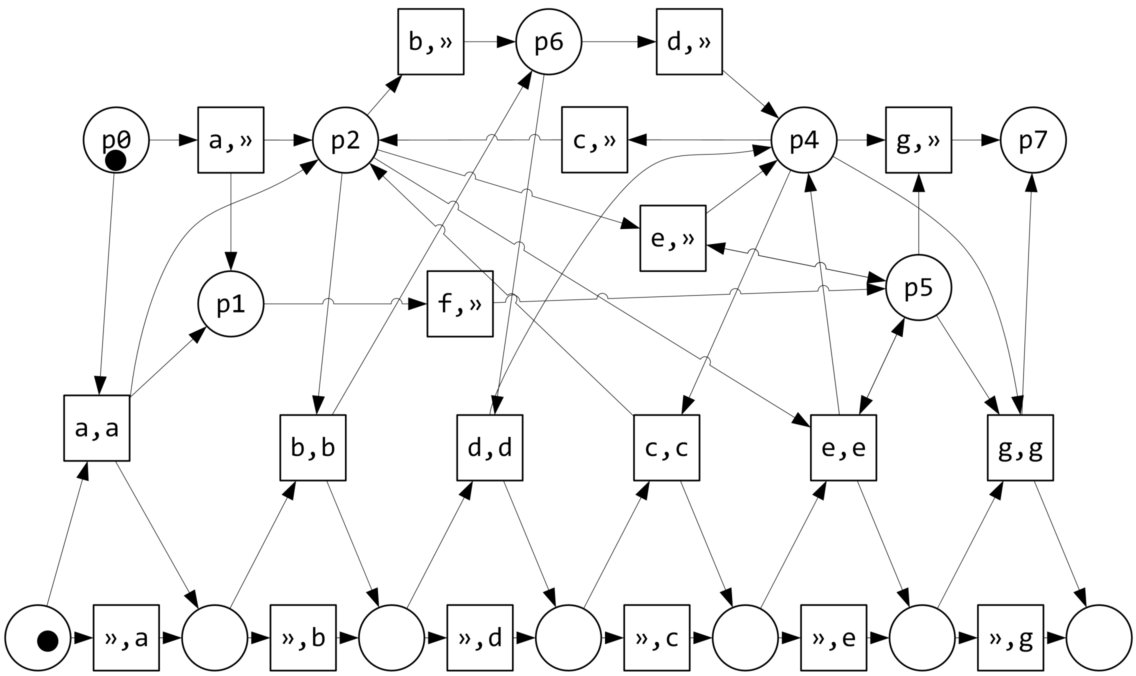
1. The table for token-based fitness is depicted below:

Token based fitness:

Trace	Freq	P	R	C	M	P (all)	R (all)	C (all)	M (all)
$\langle a, b, e, d, g\rangle$	10	8	$2^{(p_1, p_4)}$	8	$2^{(p_2, p_5)}$	80	20	80	20
$\langle a, b, f, d, g\rangle$	4	7	0	7	0	28	0	28	0
total						**108**	**20**	**108**	**20**

fitness $= \frac{1}{2}(1 - 20/108) + \frac{1}{2}(1 - 20/108) = \frac{88}{108} = 0.8148$

2. An example of such a trace is $\langle a\rangle$.
3. b is the correct alignment.
4. Below is the synchronous product



An optimal alignment with cost 1 is:

log trace	a	b	d	c	$\gg$	e	g
execution sequence	a	b	d	c	f	e	g

Solution 2.3

1.

Token based fitness:

Trace	**Freq**	**P**	**R**	**C**	**M**	**P (all)**	**R (all)**	**C (all)**	**M (all)**
s, a, c, e, t	5	7	0	7	0	35	0	35	0
a, e, d, t	3	7	1	7	1	21	3	21	3
s, a, e, t	4	6	1	6	1	24	4	24	4
s, b	10	3	0	3	0	30	0	30	0
total						**110**	**7**	**110**	**7**

fitness $= \frac{1}{2}(1 - 7/110) + \frac{1}{2}(1 - 7/110) = 0.9364$

2.

log trace	$\gg$	a	b	d	c	d	e	t
execution sequence	s	a	$\gg$	$\gg$	c	$\gg$	e	t

Cost of alignment: 4
Cost of "best of worst" alignment: $2 + 7 = 9$
Fitness $= 1 - (\frac{4}{9}) = 5/9 = 0.55$

Solution 2.4

1. The table for token-based fitness is depicted below:

Token based fitness:

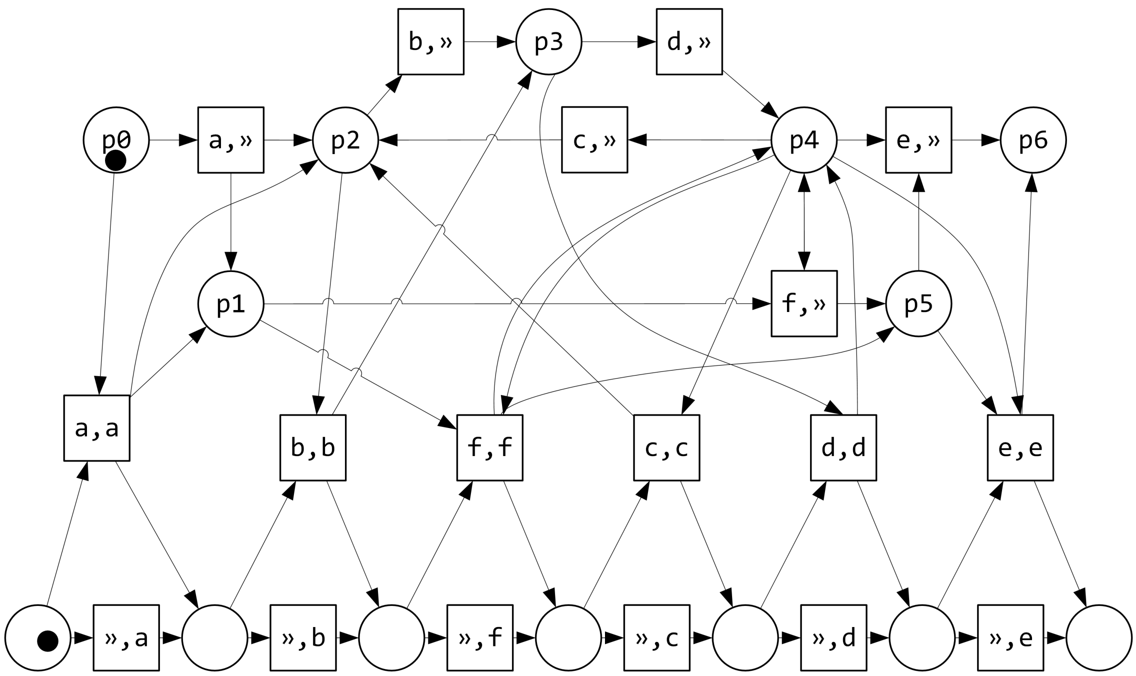
		per trace				per trace class			
Trace	**Freq**	**P**	**C**	**R**	**M**	**P**	**C**	**R**	**M**
$\langle a, b, e, d, c \rangle$	10	1,2,1,1,1,1	1,1,2,1,1,1	p_1, p_2	p_4, p_5	70	70	20	20
$\langle a, b, d, f, e \rangle$	5	1,2,1,1,2,1	1,1,1,2,2,1	0	0	40	40	0	0
total						**110**	**110**	**20**	**20**

fitness $= \frac{1}{2}(1 - \frac{M}{C}) + \frac{1}{2}(1 - \frac{R}{P}) = \frac{1}{2}(1 - 20/110) + \frac{1}{2}(1 - 20/110) = \frac{90}{110} = 0.8182$

2. An example of such a trace is $\langle a \rangle$.

3. a is the correct alignment.

4. The synchronous product looks as follows:



And an optimal alignment with cost 2 is:

log trace	a	b	≫	f	c	≫	d	e
execution sequence	a	b	d	f	c	b	d	e

Solution 2.5

1. The table for token-based fitness is depicted below:

   Token based fitness:

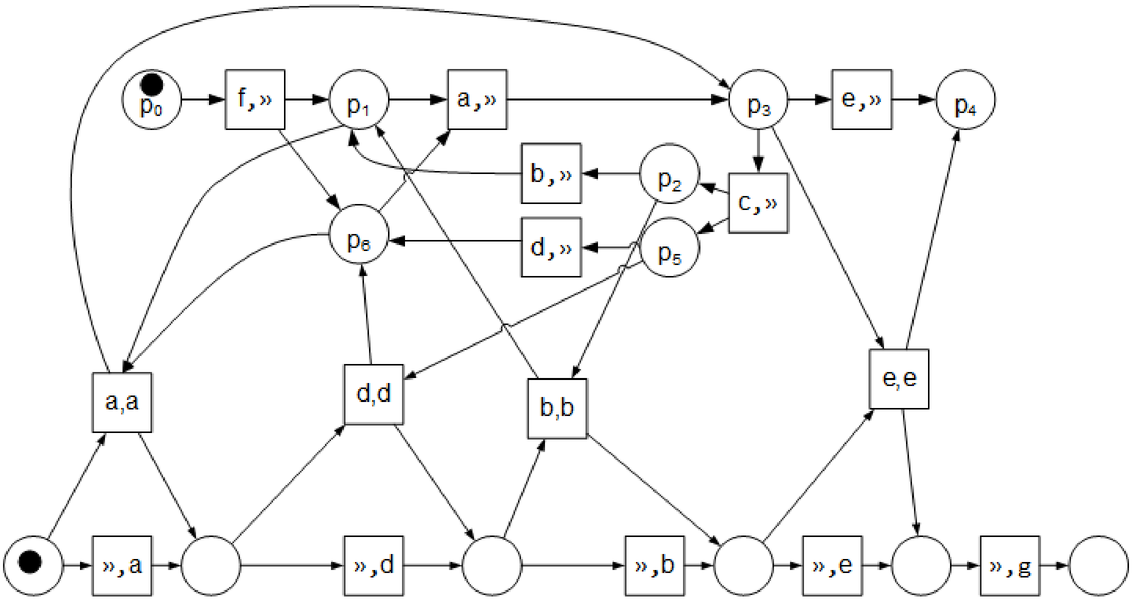
		per trace				per trace class			
Trace	**Freq**	**P**	**C**	**R**	**M**	**P**	**C**	**R**	**M**
$\langle f, a, b, d, a\rangle$	10	7	8	2	3	70	80	20	30
$\langle f, a, c, b, e\rangle$	5	8	7	2	1	40	35	10	5
total						**110**	**115**	**30**	**35**

   fitness $= \frac{1}{2}(1 - \frac{M}{C}) + \frac{1}{2}(1 - \frac{R}{P}) = \frac{1}{2}(1 - 35/115) + \frac{1}{2}(1 - 30/110) = \frac{180}{253} = 0.7115$

2. An example of such a trace is $\langle f\rangle$.

3. a is the correct alignment.

4. The synchronous product is depicted below (note the extra transition "g" which does not appear in the model)



An optimal alignment, with cost 4 is:

log trace	$\gg$	a	$\gg$	d	b	$\gg$	e	g
execution sequence	f	a	c	d	b	a	e	$\gg$

Or, equally optimal:

log trace	$\gg$	a	d	b	e	g
execution sequence	f	a	$\gg$	$\gg$	e	$\gg$

Solution 2.6

1. The table for token-based fitness is depicted below:

   Token based fitness:

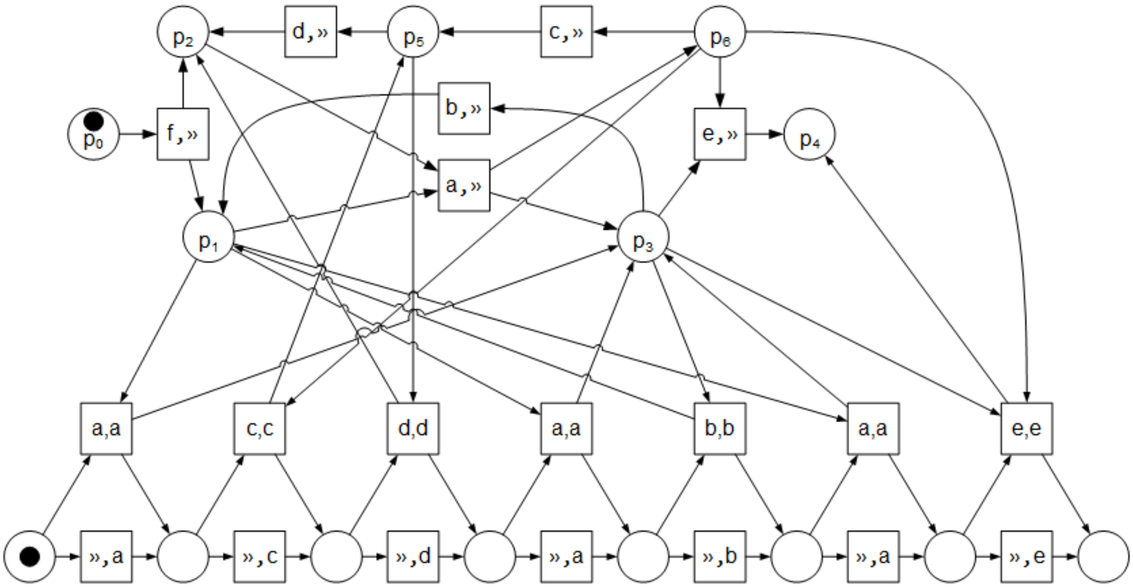
		per trace				per trace class			
Trace	**Freq**	**P**	**C**	**R**	**M**	**P**	**C**	**R**	**M**
$\langle f, a, b, c, d, a, e \rangle$	10	11	11	0	0	110	110	0	0
$\langle f, a, c, b, a \rangle$	5	9	8	3	2	45	40	15	10
total						**155**	**150**	**15**	**10**

fitness $= \frac{1}{2}(1 - \frac{M}{C}) + \frac{1}{2}(1 - \frac{R}{P}) = \frac{1}{2}(1 - 10/150) + \frac{1}{2}(1 - 15/155) = \frac{427}{465} = 0.9183$

2. An example of such a trace is $\langle f \rangle$.

3. c is the correct alignment.

4. The synchronous product is depicted below:



An optimal alignment with cost 2 is shown below:

log trace	$\gg$	a	c	d	a	b	a	e
execution sequence	f	a	c	d	$\gg$	b	a	e

Solution 2.7
From most fitting to least fitting, the models are C, A, B.

Solution 2.8
From most fitting to least fitting, the models are A, C, B.

Solution 2.9
From most fitting to least fitting, the models are B, C, A.

Solution 2.10
From most fitting to least fitting, the models are A, B, C.