

Advanced Process Mining

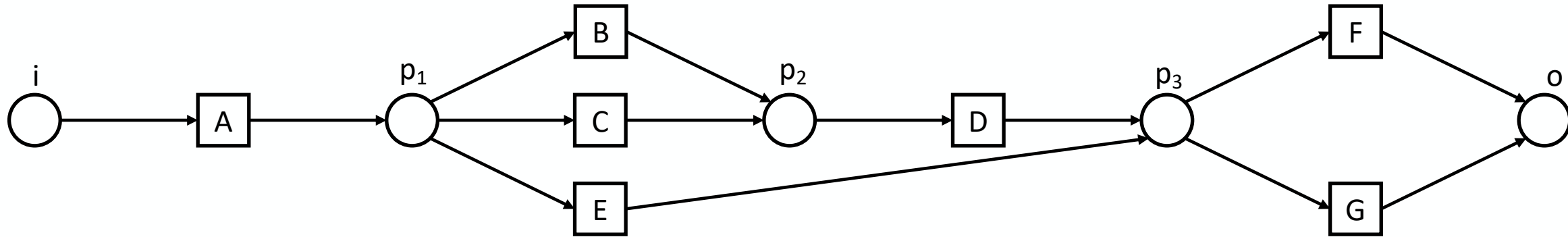
Sommer term 2020

Exercise sheet 6

Revision

Four dimensions of quality

Exercise 1a

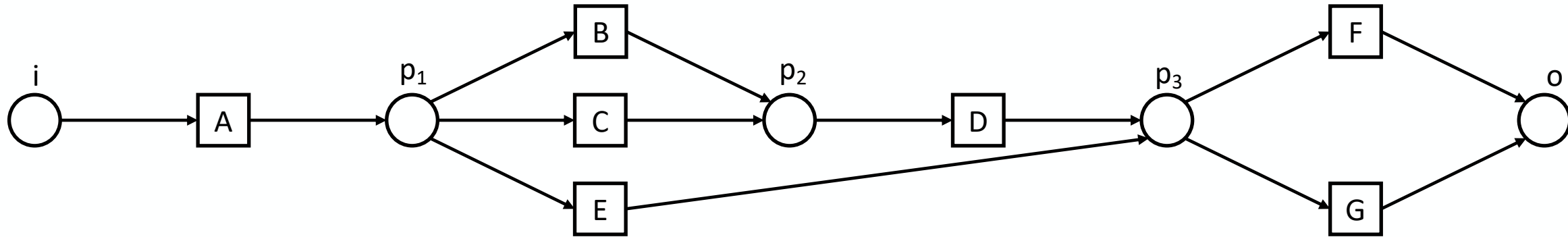


Given the process model above generate an event log with the least possible number of entries that will result in:

Precision = 1

Four dimensions of quality

Exercise 1a



Given the process model above generate an event log with the least possible number of entries that will result in:

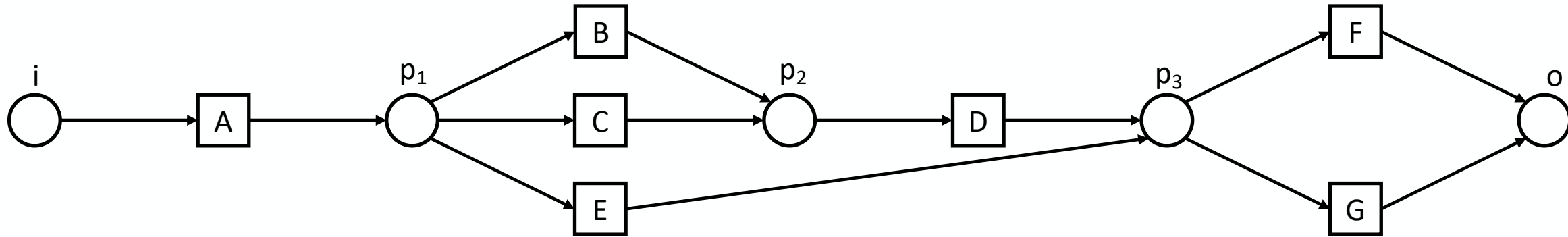
Precision = 1

For a perfect precision the event log has to contain every possible trace reproducible by the model.

#	Trace
n	ABDF
m	ABDG
o	ACDF
p	ACDG
q	AEF
r	AEG

Four dimensions of quality

Exercise 1a

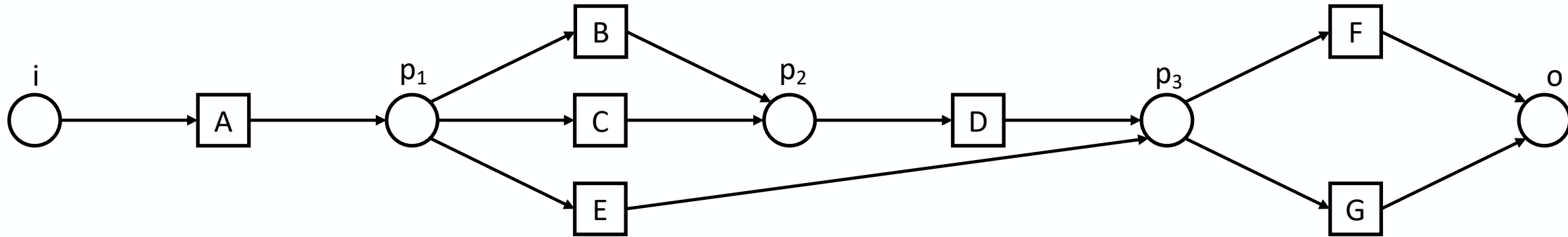


Given the process model above generate an event log with the least possible number of entries that will result in:

Fitness = 1

Four dimensions of quality

Exercise 1a



Given the process model above generate an event log with the least possible number of entries that will result in:

Fitness = 1

#	Trace
n	AEG

For a perfect fitness the event log has to contain only one single entry, that can be reproduced exactly by the model.

Four dimensions of quality

Exercise 1b

Calculate the F1-Score for the given values:

Fitness = 0.80

Precision = 0.95

Four dimensions of quality

Exercise 1b

Calculate the F1-Score for the given values:

Fitness = 0.80

Precision = 0.95

$$\text{F1 Score} = \frac{2 \times \text{fitness} \times \text{precision}}{\text{fitness} + \text{precision}} = \frac{2 \times 0.8 \times 0.95}{0.8 + 0.95} = 0.8686$$

Four dimensions of quality

Exercise 1c

Name two measurements to reduce the complexity of real-life logs.

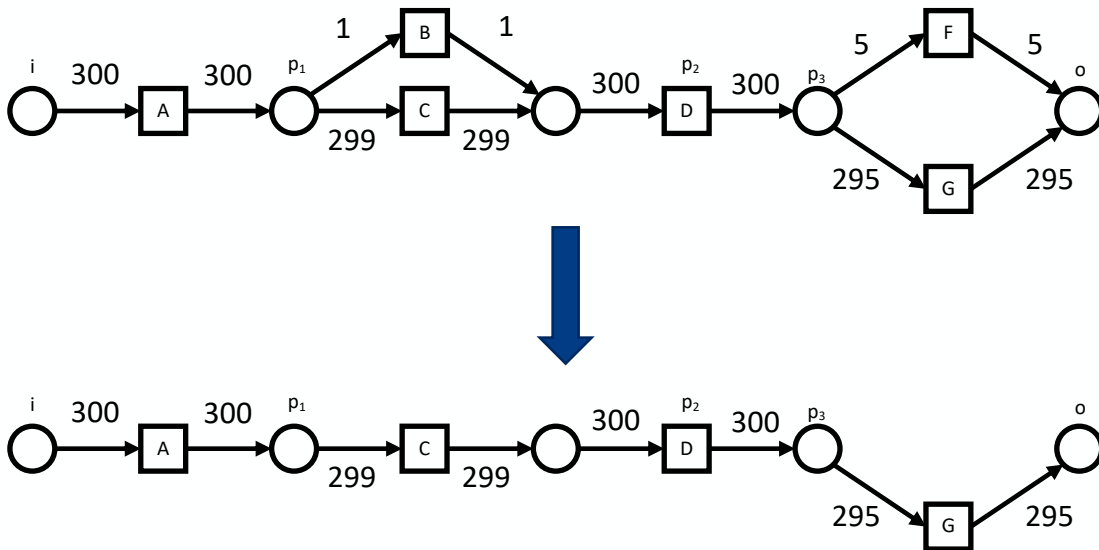
Four dimensions of quality

Exercise 1c

Name two measurements to reduce the complexity of real-life logs.

Abstraction

Only most frequent arcs/activities



Filtering

Remove traces that fulfil a certain condition

#	Trace
200	ABDF
190	ABDG
100	ACDF
10	DG



#	Trace
200	ABDF
190	ABDG
100	ACDF
100	ACDF

Name one possible problem with Directly-Follows graphs.

Activities that have a flexible ordering (e.g., due to concurrency) lead to Spaghetti-like DFGs with loops even when activities are executed at most once.

DFGs can be simplified using frequency-based thresholds. However, this may lead to all kinds of interpretation problems due to "invisible gaps" in the model.

Performance information mapped onto DFGs can be misleading, e.g., the average time reported between two activities is conditional (only the situations where they directly follow each other are considered)

The precision of a Directly-Follows graph is very high.
True or False?

False

A DFG allows for more behaviour than was recorded in the log.

**The effort to find the optimal alignment is independent from the trace length and model size.
True or False?**

False

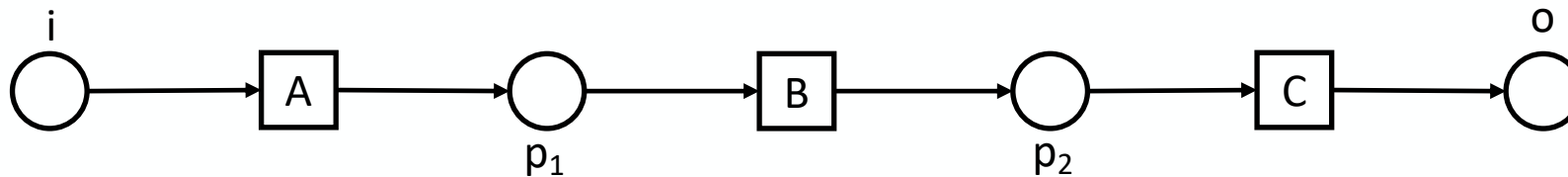
The size of the search space is determined by the length of the trace and the size of the model.
The longer a trace and the more complicated a process model, the more combinations of alignments have to be considered.

There are always multiple optimal alignments for a given trace and model if the trace is not exactly representable by the model.

True or False?

False

Counterexample:



σ_1	\gg	B	\gg
N	A	B	C

Name two reasons why conformance checking can be useful.

Auditing can be a reason. It helps the auditor to detect fraud or malpractice.

Improving of the process. Risks and inefficiencies can be detected with conformance checking.