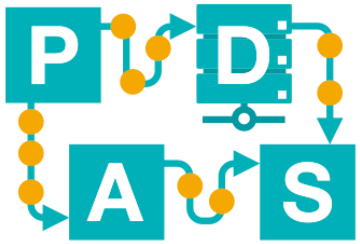
Process Mining - Instruction

Lecture 14

IDS-L14-I



Chair of Process and Data Science

RWTH AACHEN UNIVERSITY

Process Discovery

Process Discovery

In the lecture you have seen the key concepts of Process Discovery:

- **Process data often comes in form of events (event logs), recorded with a timestamp, case ID, and activity of the event**
- **Doing Process Discovery means creating a model representing the behavior in the event data**
- **Many formalisms for models (Petri nets, process trees, others)**

Chair of Process and Data Science

Process Discovery

How do people create process models without Process Mining?

By hand!

Process models are often designed by hand. Usually, these models are drawn by experts of a certain process and reflect how thing *should* go in reality (normative process model).

Process Discovery

Process Discovery is hard even when a human and a computer work together. Two (of many) reasons:

- **As it is also the case in Machine Learning, you only get to observe a part of reality. Your model has to balance between fitting both the data you have and also unseen process instances.**

Process Discovery

Process Discovery is hard even when a human and a computer work together. Two (of many) reasons:

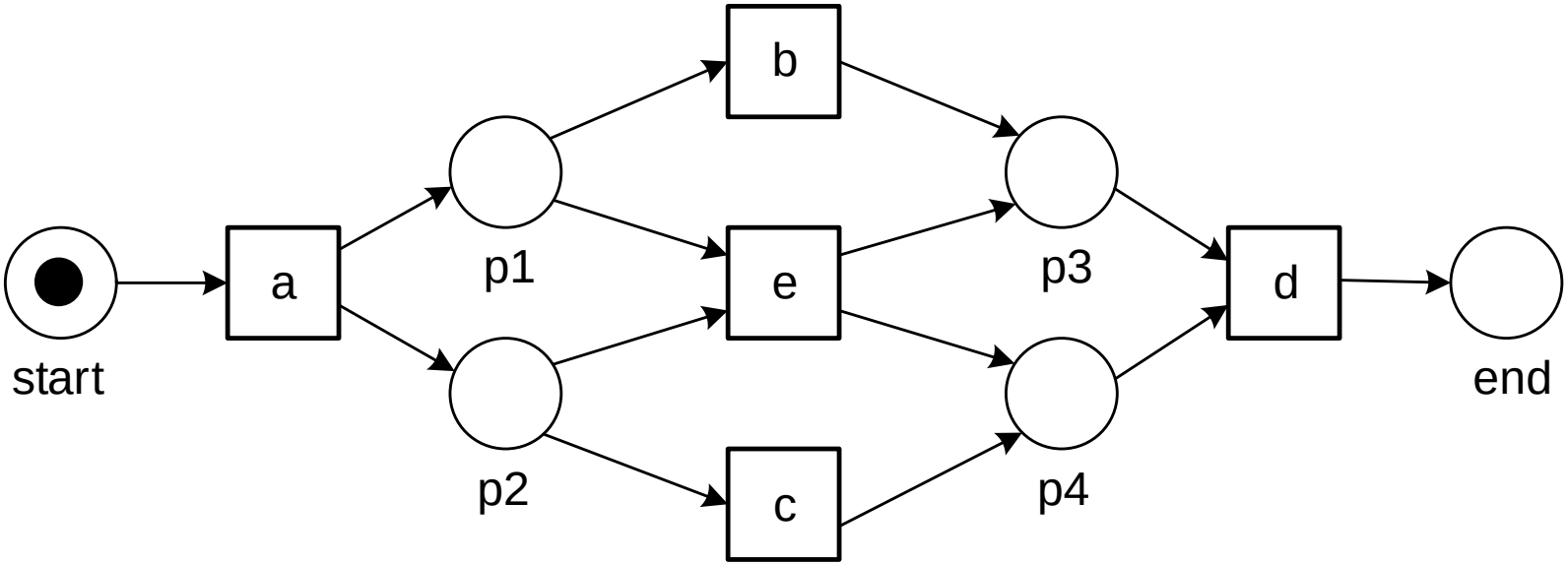
- **A problem more specific to Process Mining: you only get to observe positive behavior. Event logs cannot contain negative examples, and you do not have examples of what *cannot happen*.**

Process Models: Petri Nets

Recall from the lecture:

- **Petri nets are composed by directed arcs, places, and transitions.**
- **Places contain tokens. A specific configuration of tokens in a net is called a marking.**
- **Transition can fire, consuming a token from the input places, putting a token in the output places, and "producing" and event.**
- **Initial and final markings define the start and end of a trace.**

“Playing out”

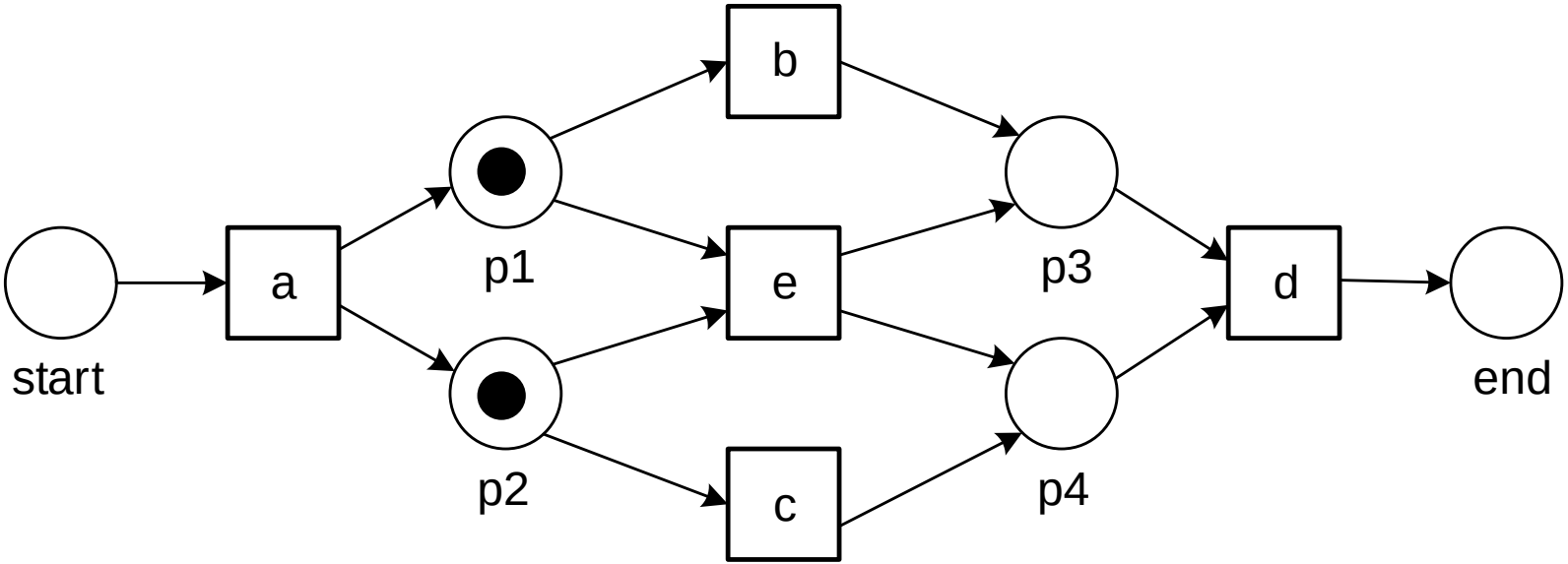


Initial marking

<>

Chair of Process and Data Science

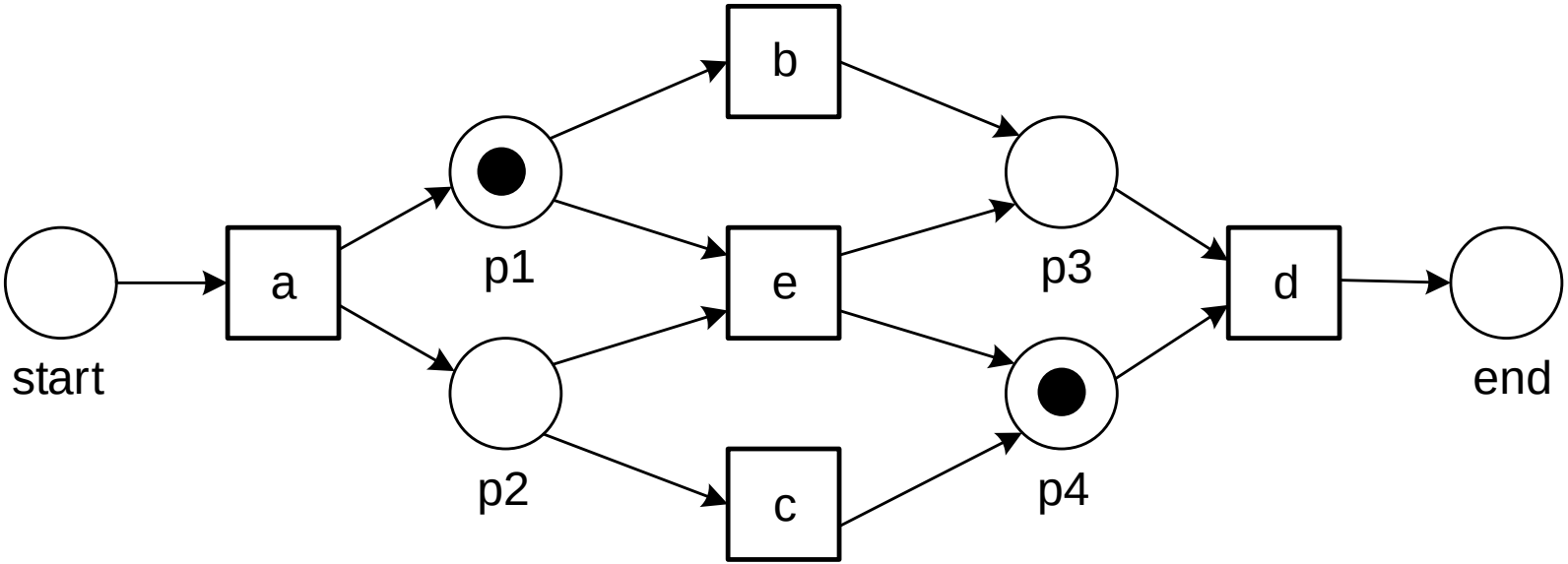
“Playing out”



<a>

Chair of Process and Data Science

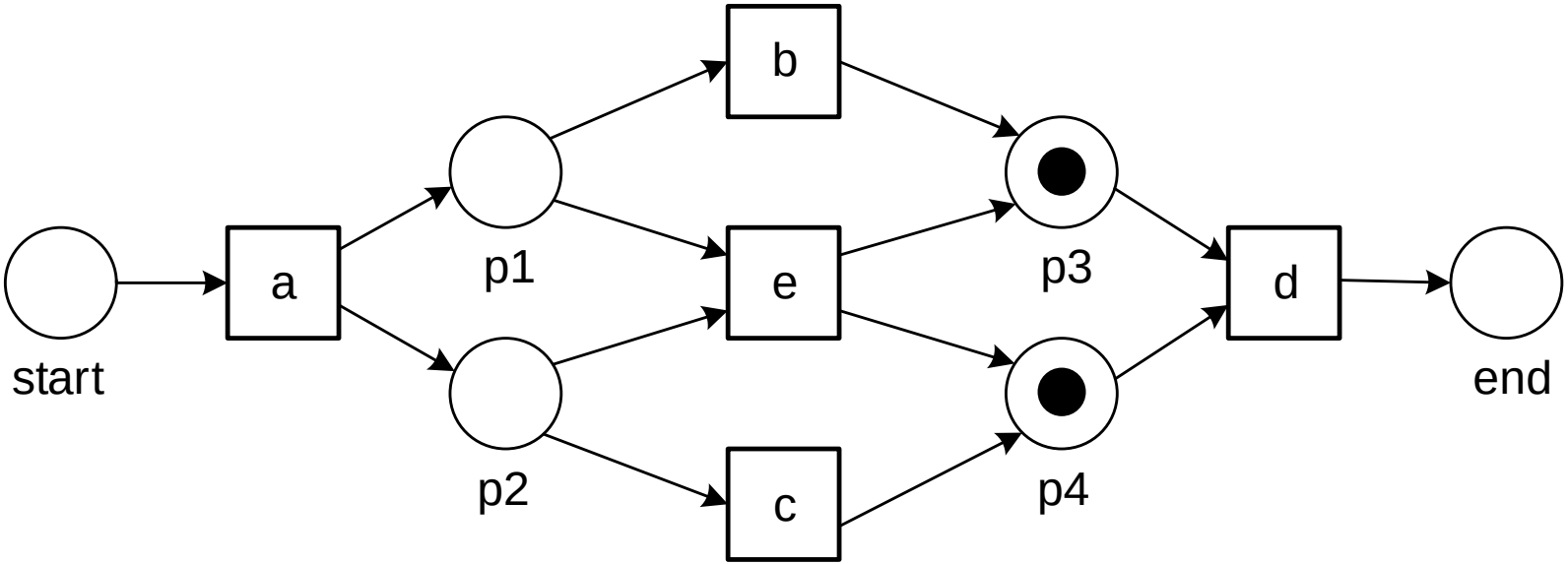
“Playing out”



<a,c>

Chair of Process and Data Science

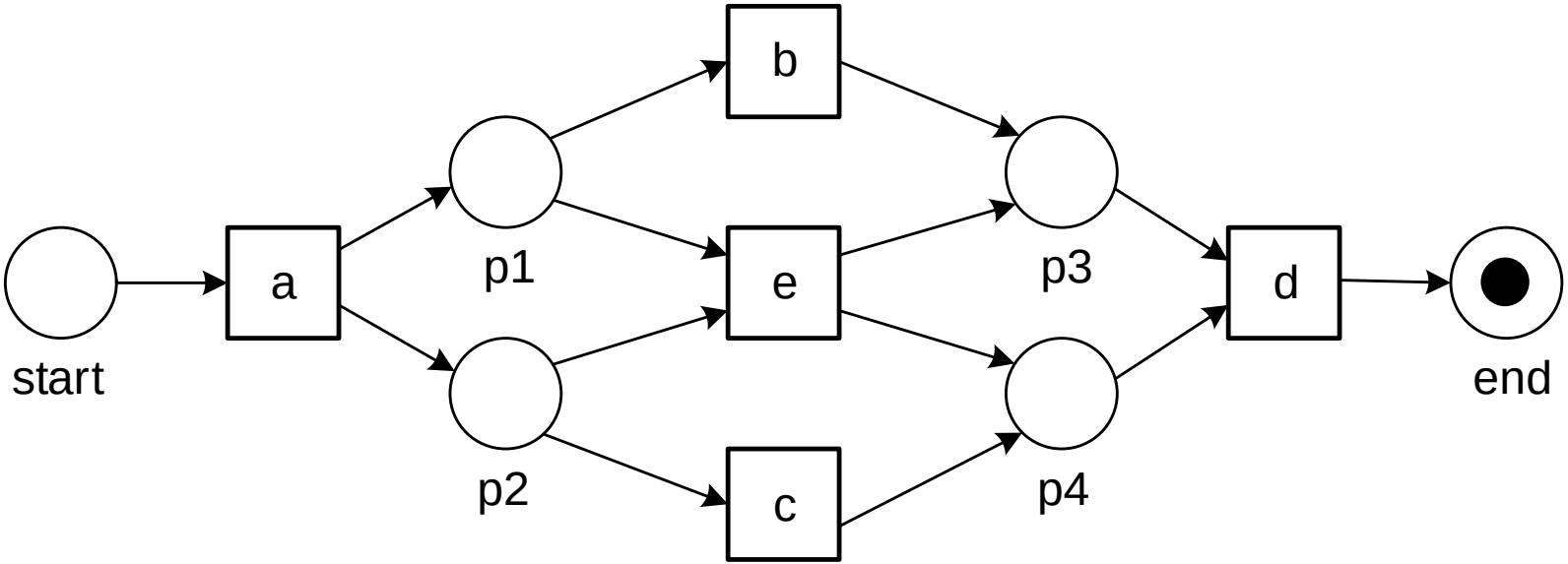
“Playing out”



<a,c,b>

Chair of Process and Data Science

“Playing out”



<a,c,b,d>

Final marking

Process Discovery

Given some traces, can you come up with a Petri net that can replay all of them?

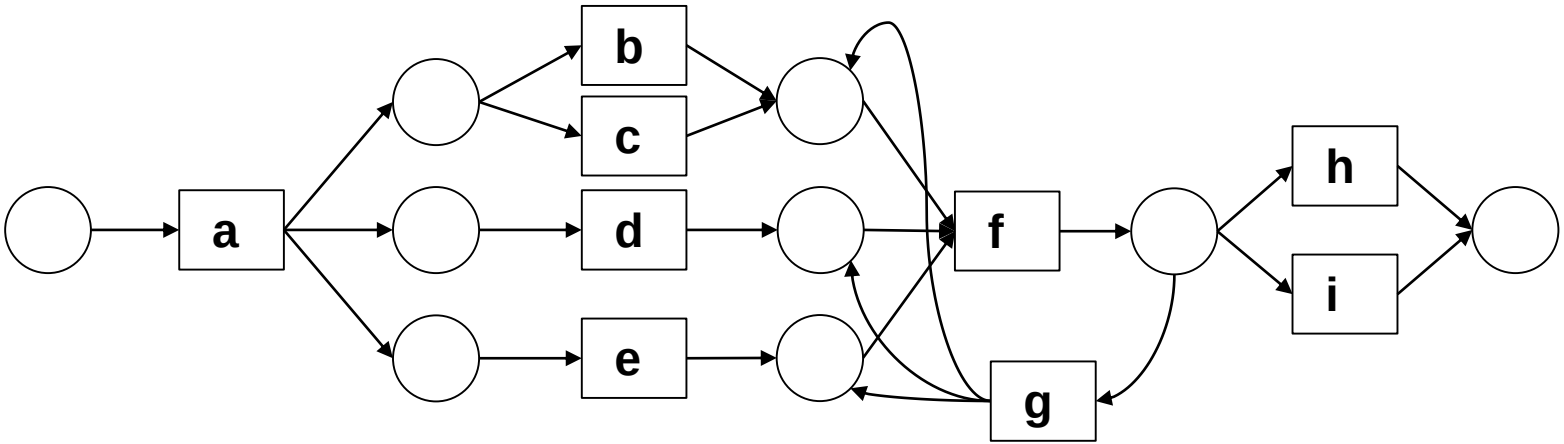
<a,b,d,e,f,h>

<a,e,c,d,f,g,f,i>

<a,d,e,b,f,g,f,g,f,h>

Process Discovery: solution

<a,b,d,e,f,h> <a,e,c,d,f,g,f,i> <a,d,e,b,f,g,f,g,f,h>



Chair of Process and Data Science