Dirk **Fahland**

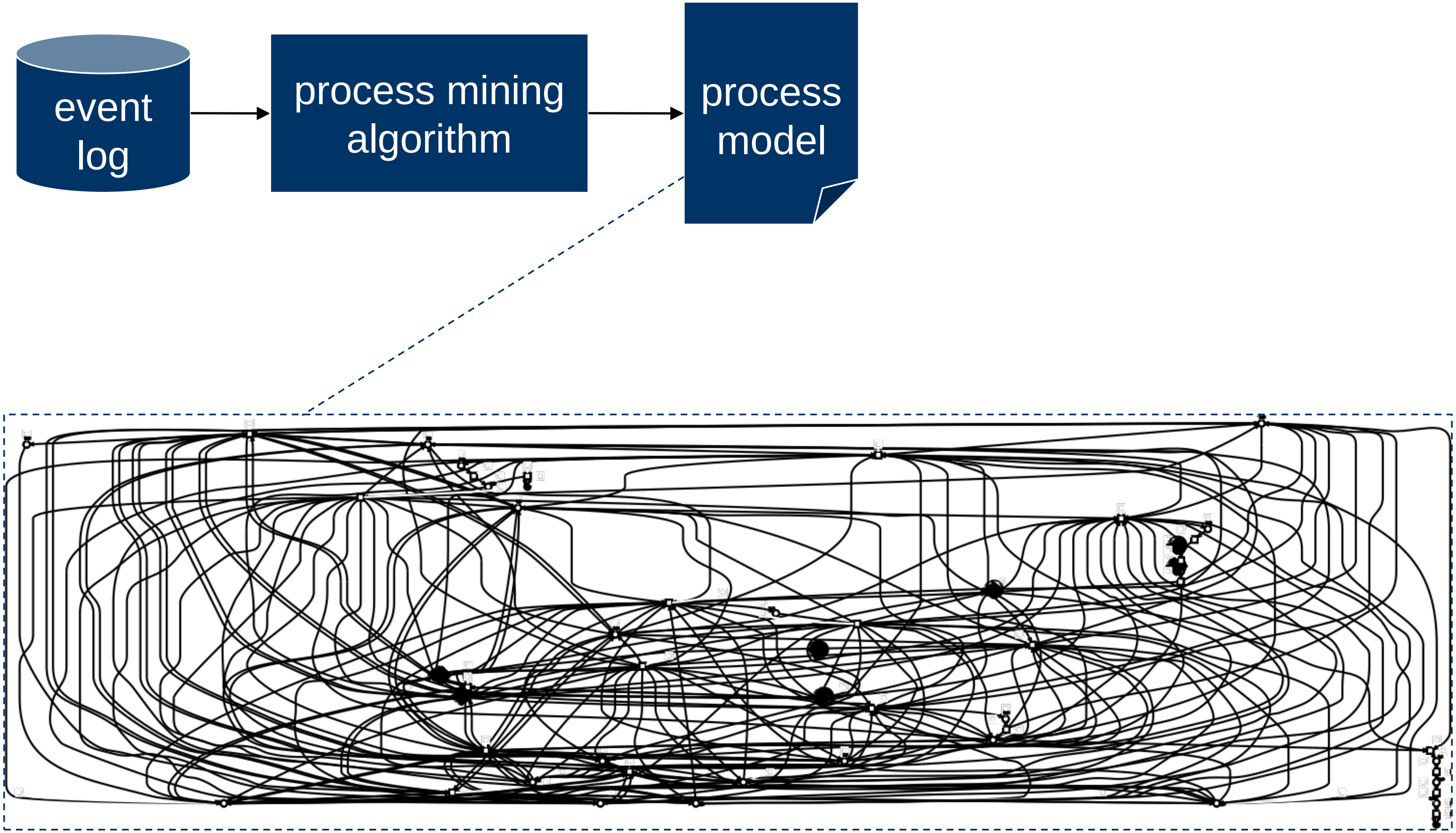
Wil M.P. **van der Aalst**

Simplifying Mined Process Models

Technische Universiteit **Eindhoven** University of Technology

Where innovation starts

Process Mining, Currently

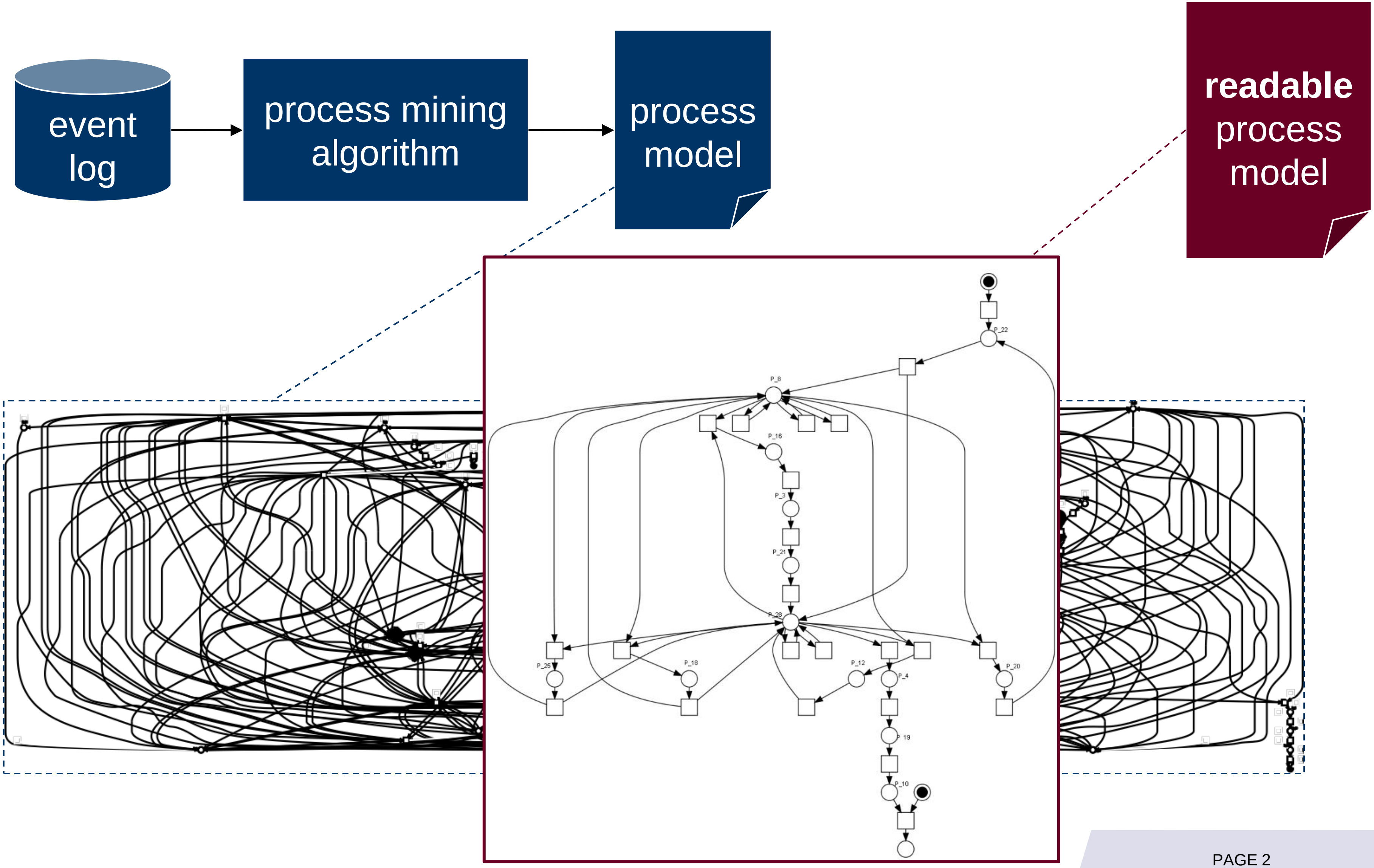


event log

process mining algorithm

process model

Process Mining, Currently



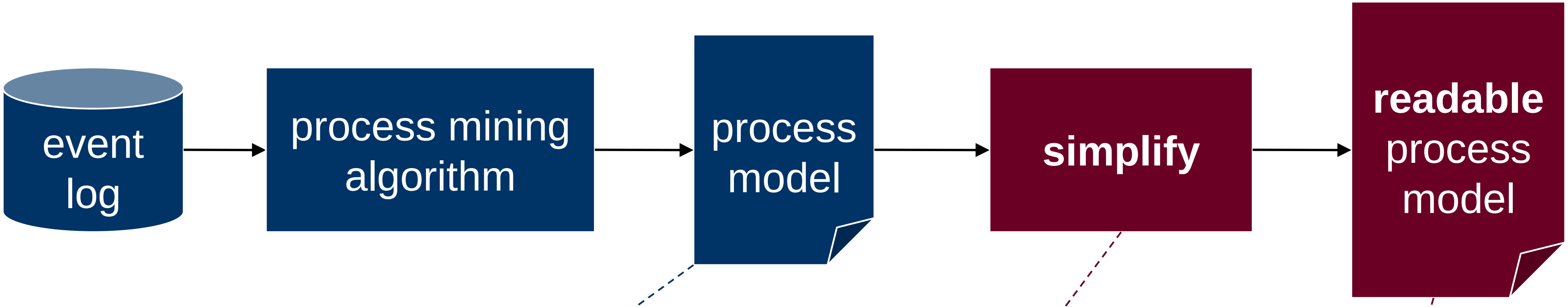
event log

process mining algorithm

process model

readable process model

Post-Process the Model

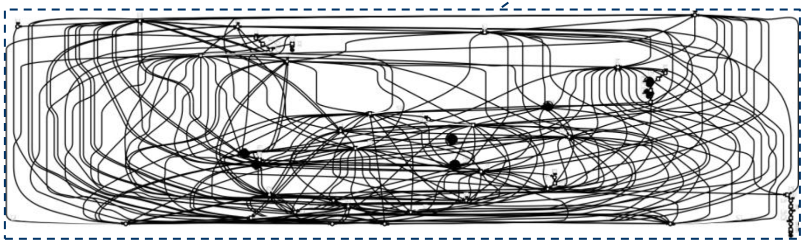

event log

process mining algorithm

process model

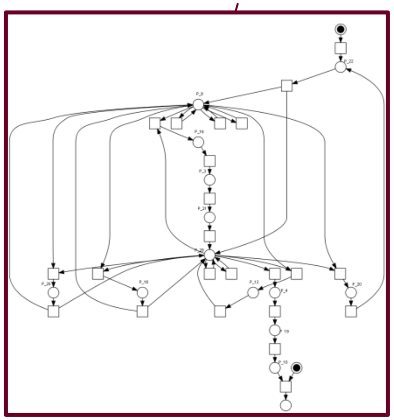
simplify

readable process model



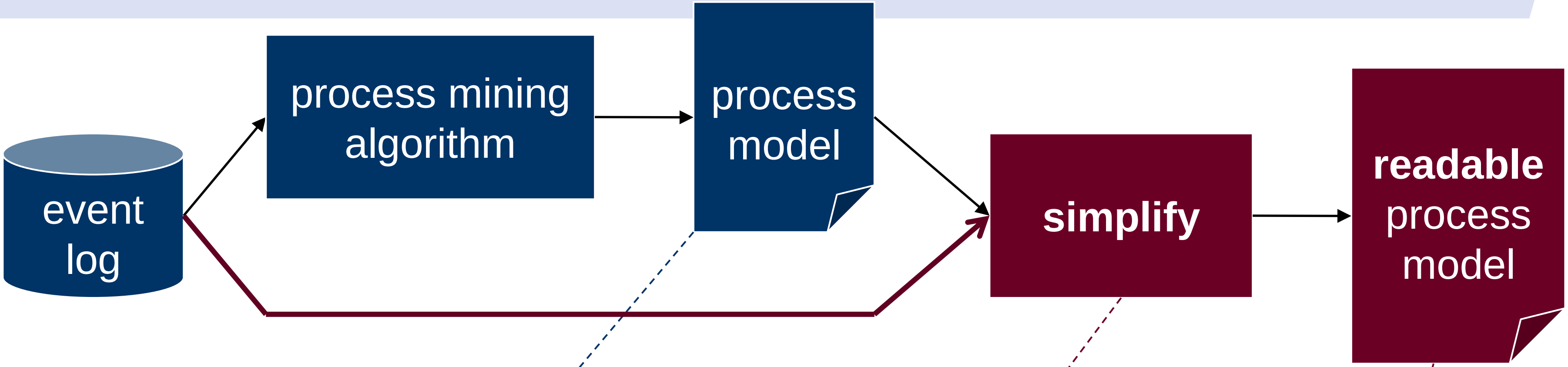
can replay the entire log

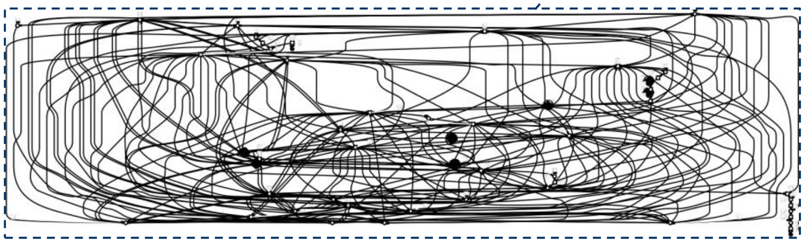
introduce:
post-processing operations on the mined model



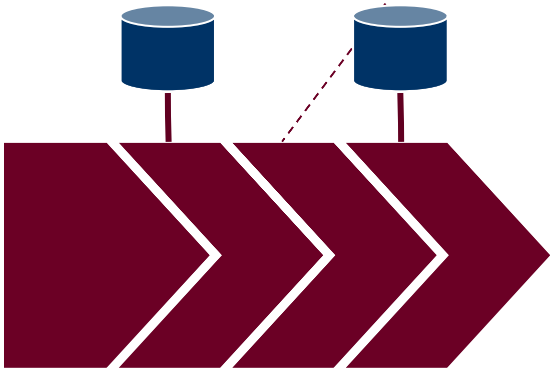
can replay the entire log

…Based on Original Event Log

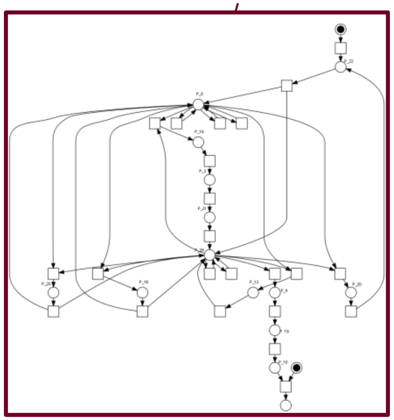




can replay the entire log

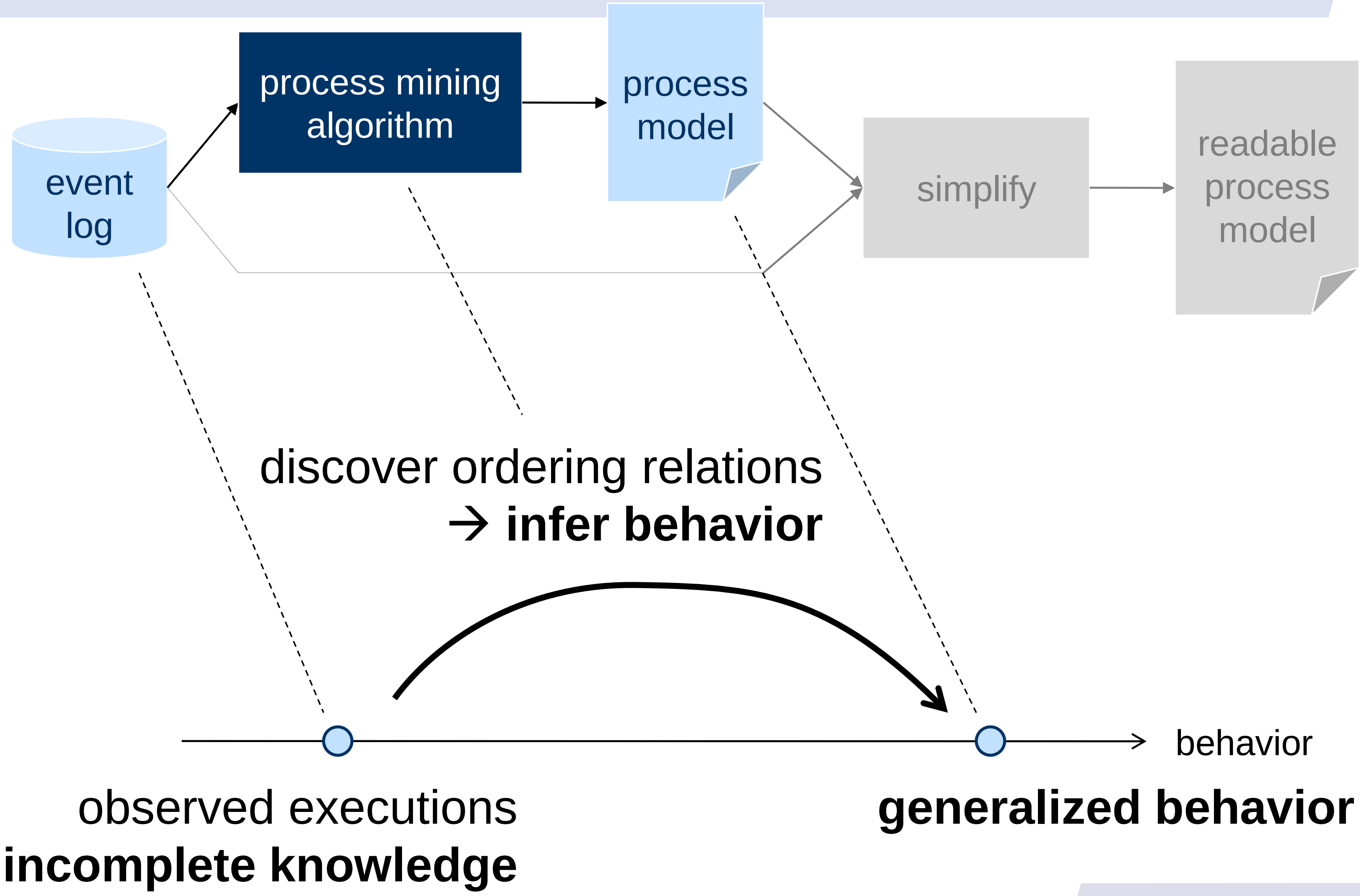


introduce: **post-processing operations on the mined model**



can replay the entire log

Analysis



event log

process mining algorithm

process model

simplify

readable process model

discover ordering relations

→ **infer behavior**

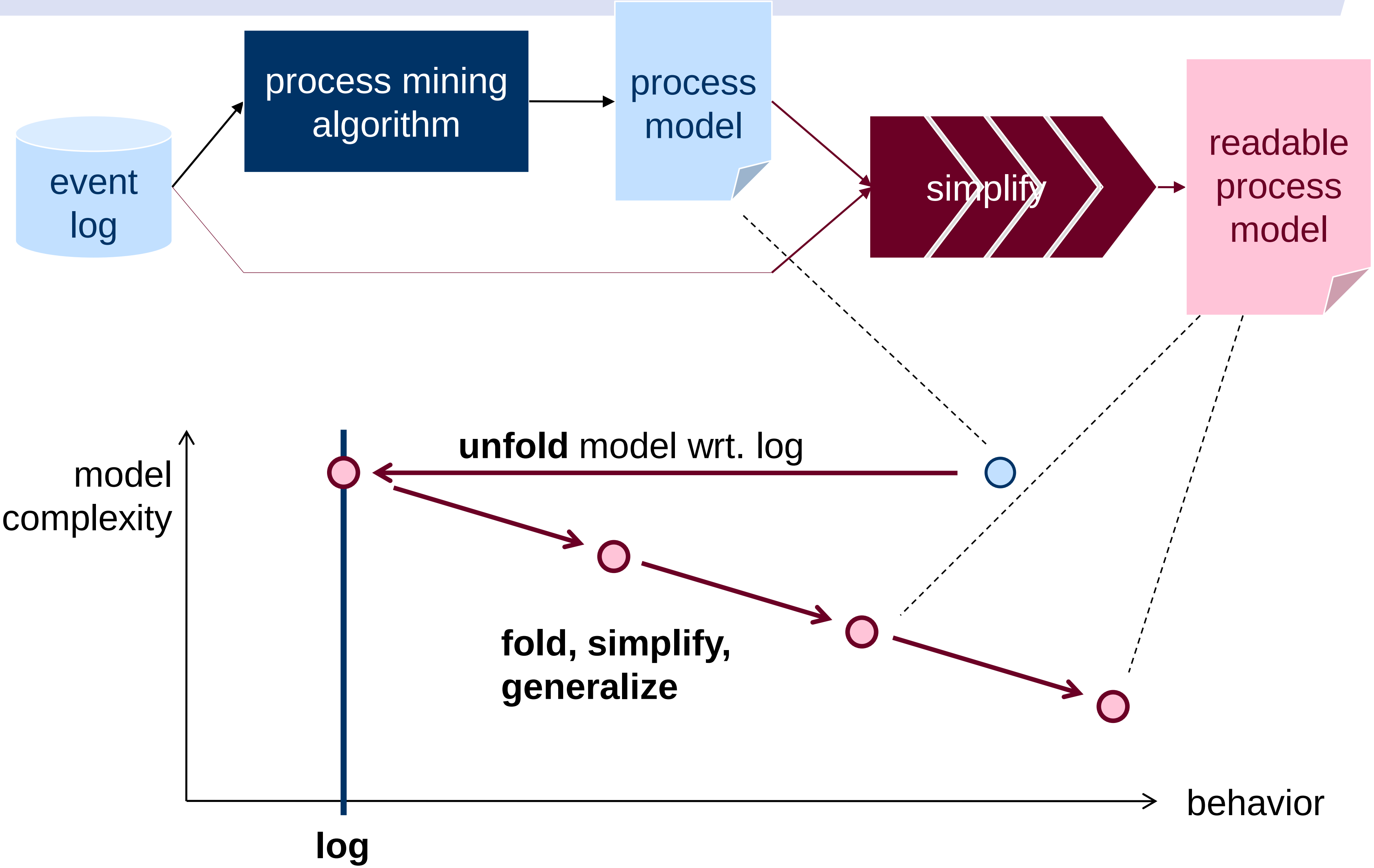
behavior

observed executions

incomplete knowledge

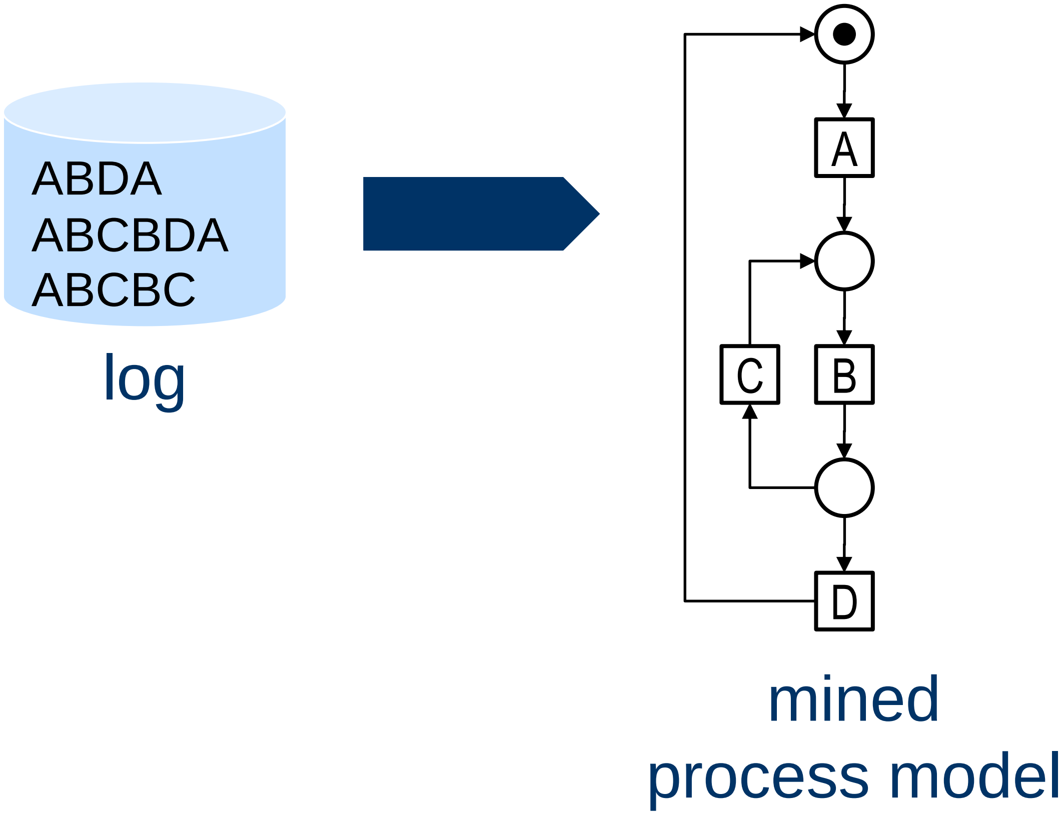
generalized behavior

Idea: Re-Adjust Generalization



Unfold a Spaghetti-Model

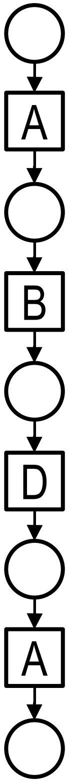
Unfold Model wrt. a Log


ABDA
ABCBDA
ABCBC

log

mined
process model

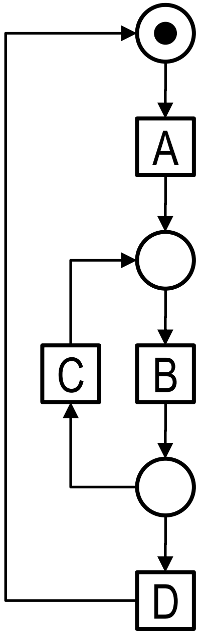
Unfold Model wrt. a Log



unfold

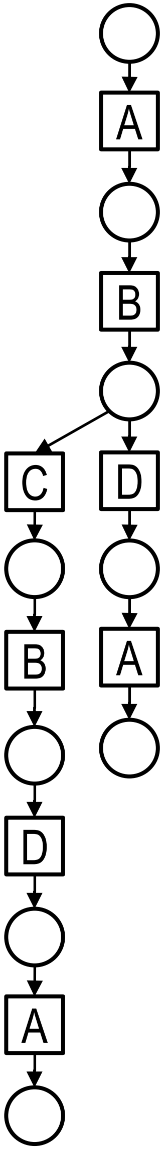
ABDA
ABCBDA
ABCBC

log



mined process model

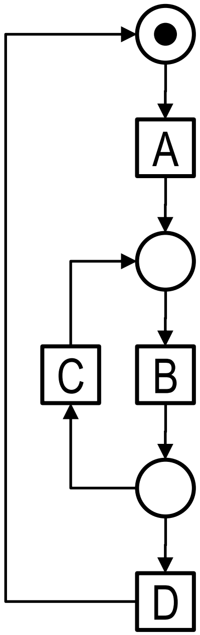
Unfold Model wrt. a Log



unfold

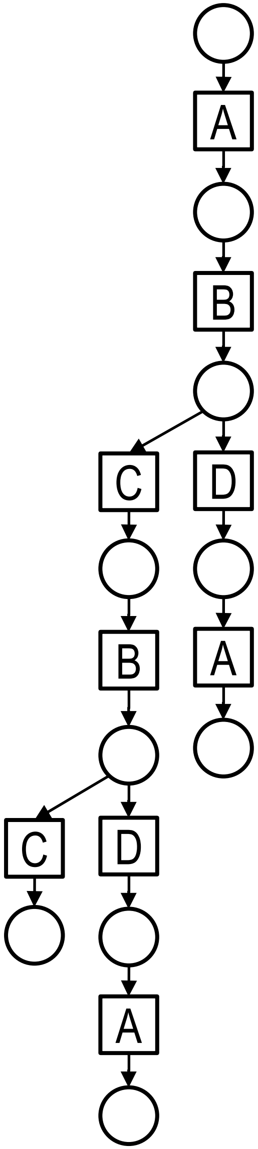
ABDA
ABCBDA
ABCBC

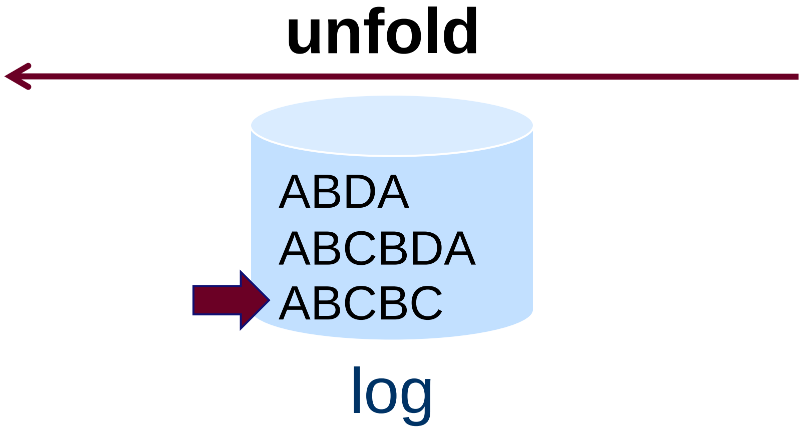
log



mined
process model

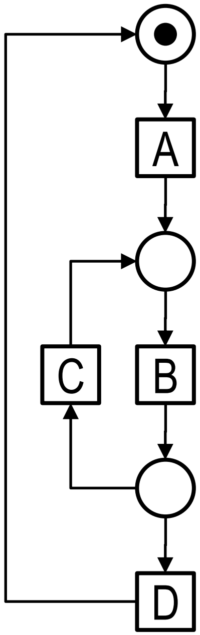
Unfold Model wrt. a Log




unfold

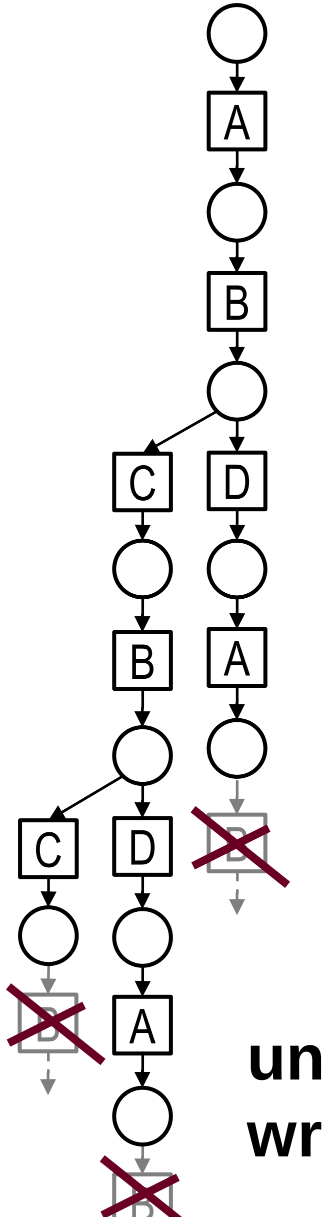
ABDA
ABCBDA
ABCBC

log

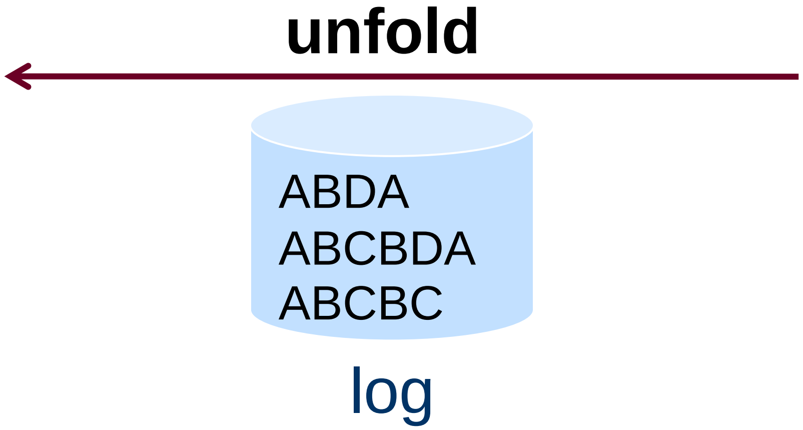


mined
process model

Unfold Model wrt. a Log

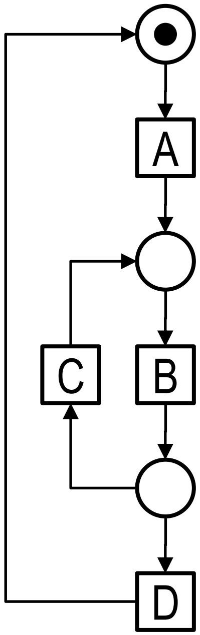


unfolding
wrt. the log


unfold

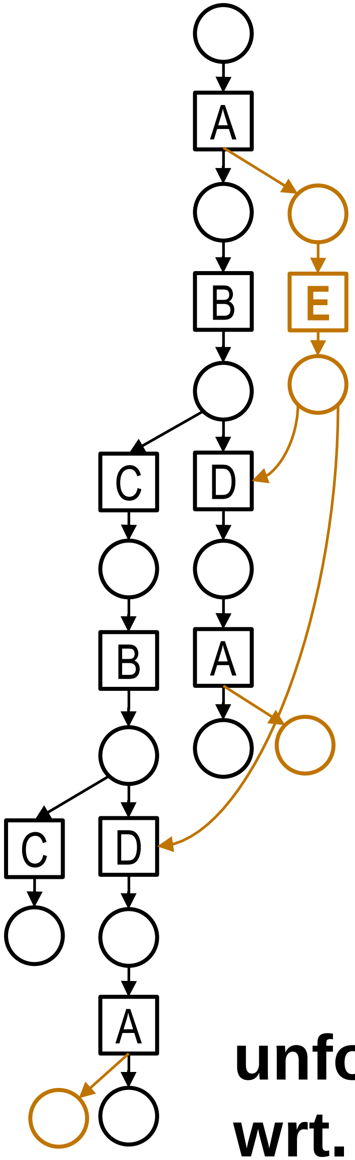
ABDA
ABCBDA
ABCBC

log



mined
process model

Represents Concurrency

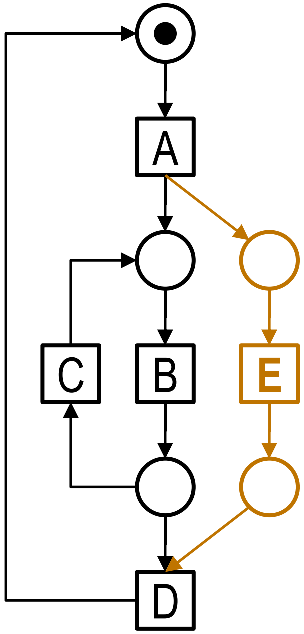


unfolding
wrt. the log

unfold

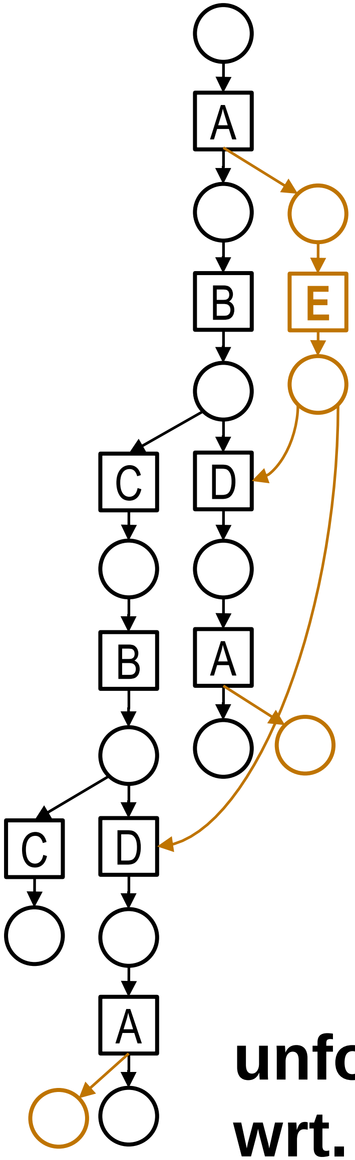
AEBDA
ABECBDA
ABCBC

log

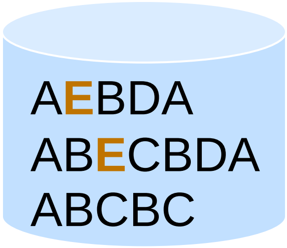


mined
process model

Represents Concurrency



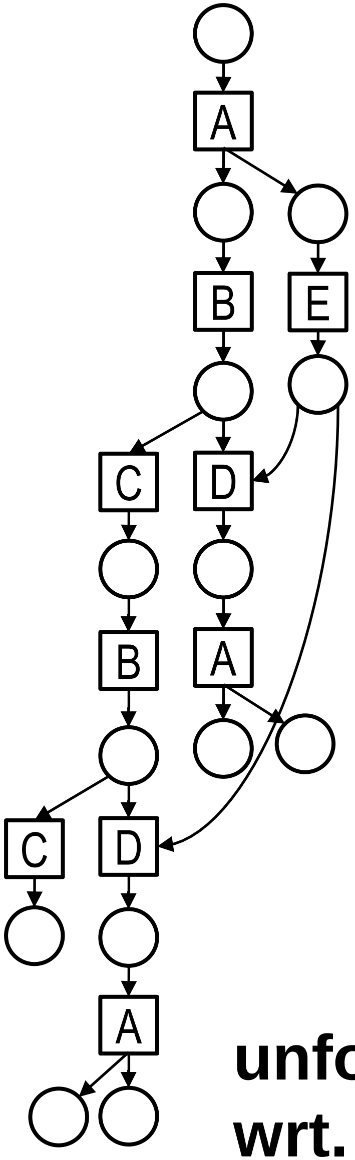
unfolding wrt. the log


AEBDA
ABECBDA
ABCBC

log

- **is a process model**
- contains only behavior in the log
- is acyclic
- represents concurrency explicitly
- labeled
  (several tasks with same label)

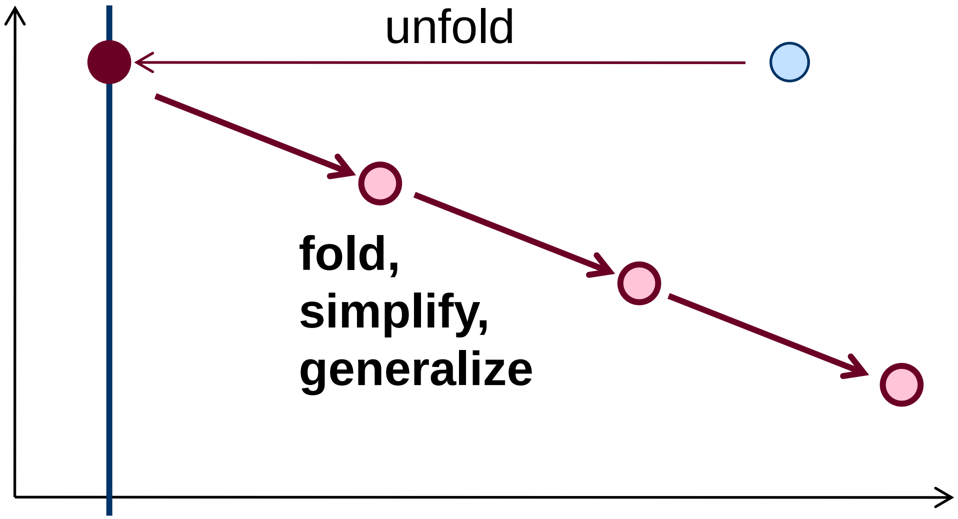
Represents Concurrency



unfolding wrt. the log

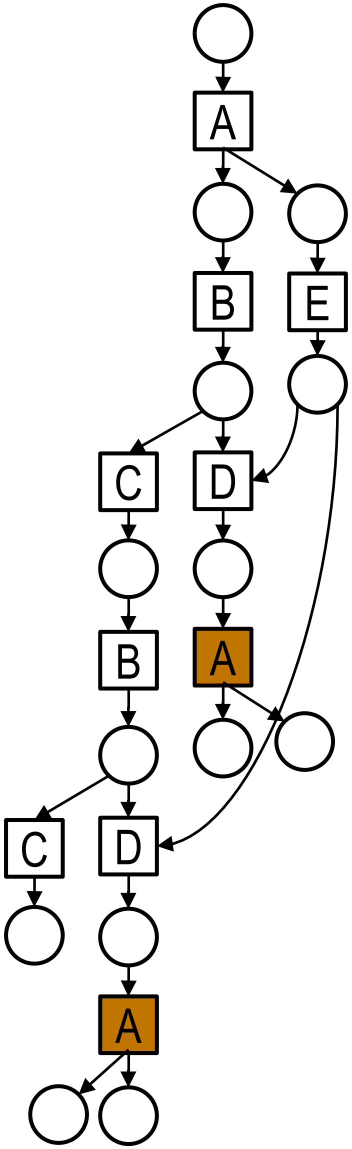
AEBDA
ABECBDA
ABCBC

log



fold, simplify, generalize

Fold an unfolded model

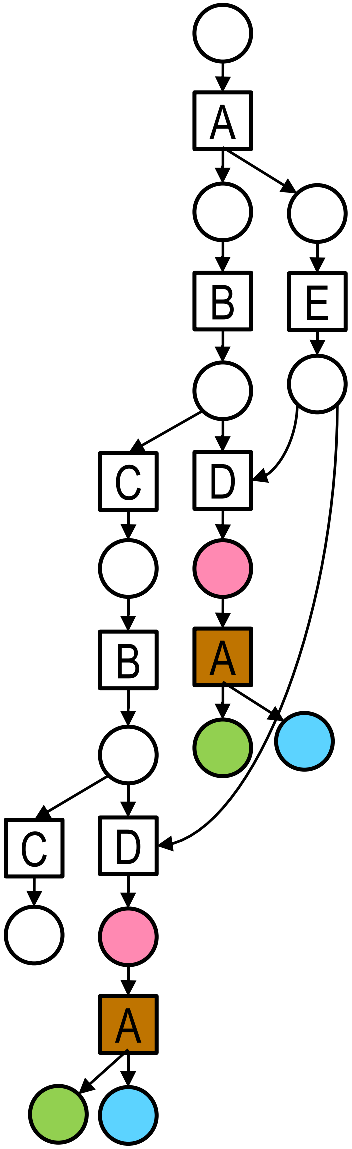


merge **equivalent** nodes

necessary condition on equivalent **transitions**

- same label

Fold an unfolded model

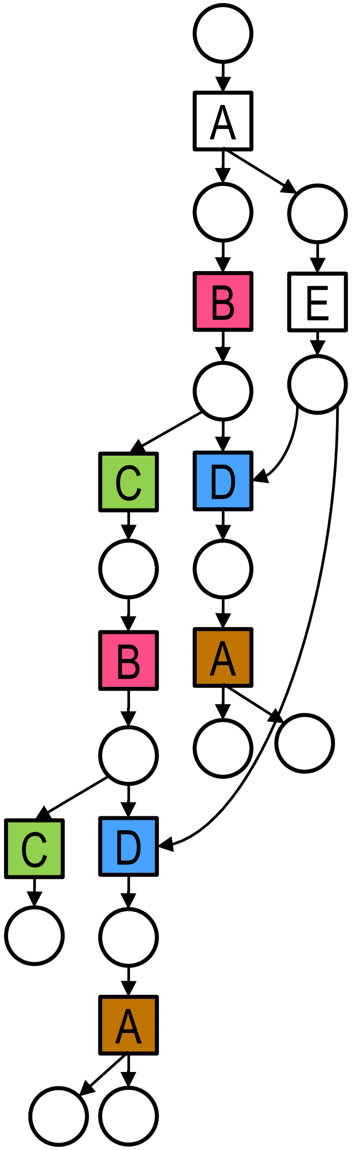


merge **equivalent** nodes

necessary condition on equivalent **transitions**

- same label
- equivalent pre-/post-places

Fold an unfolded model



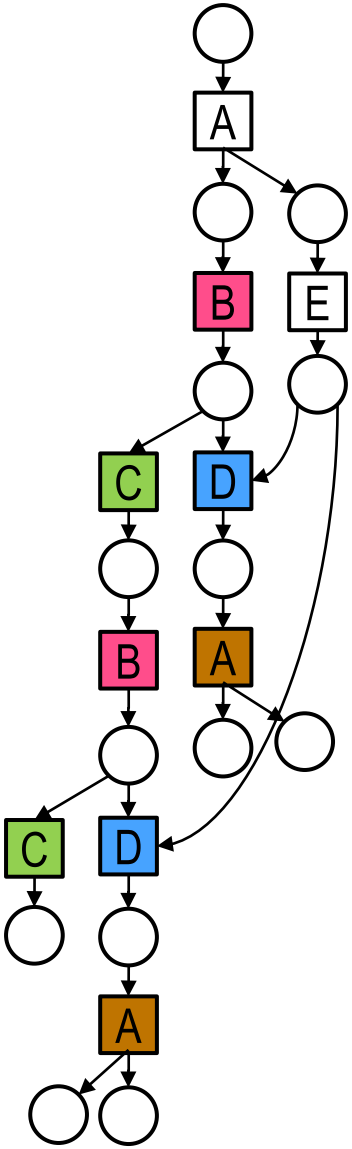
merge **equivalent** nodes

necessary condition on equivalent **transitions**

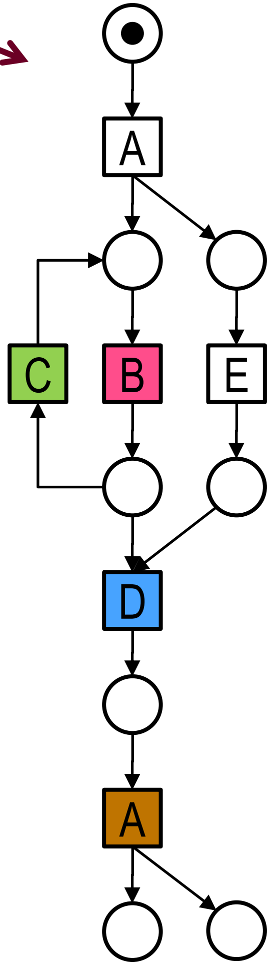
- same label
- equivalent pre-/post-places

various equivalences possible (see paper for some)

Fold an unfolded model

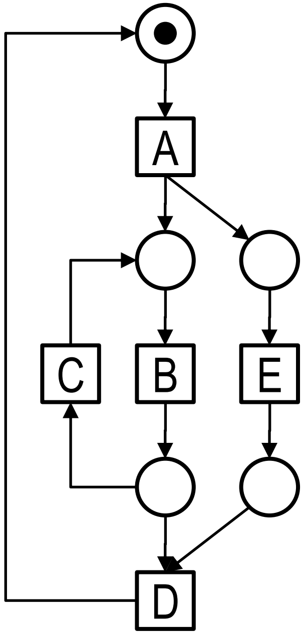


merge **equivalent** nodes

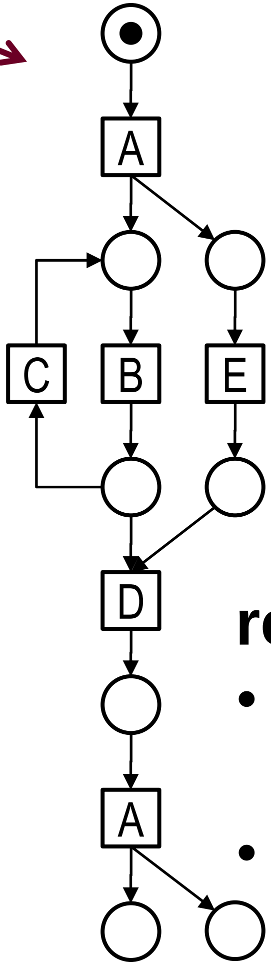


Unfolding and Refolding

unfold



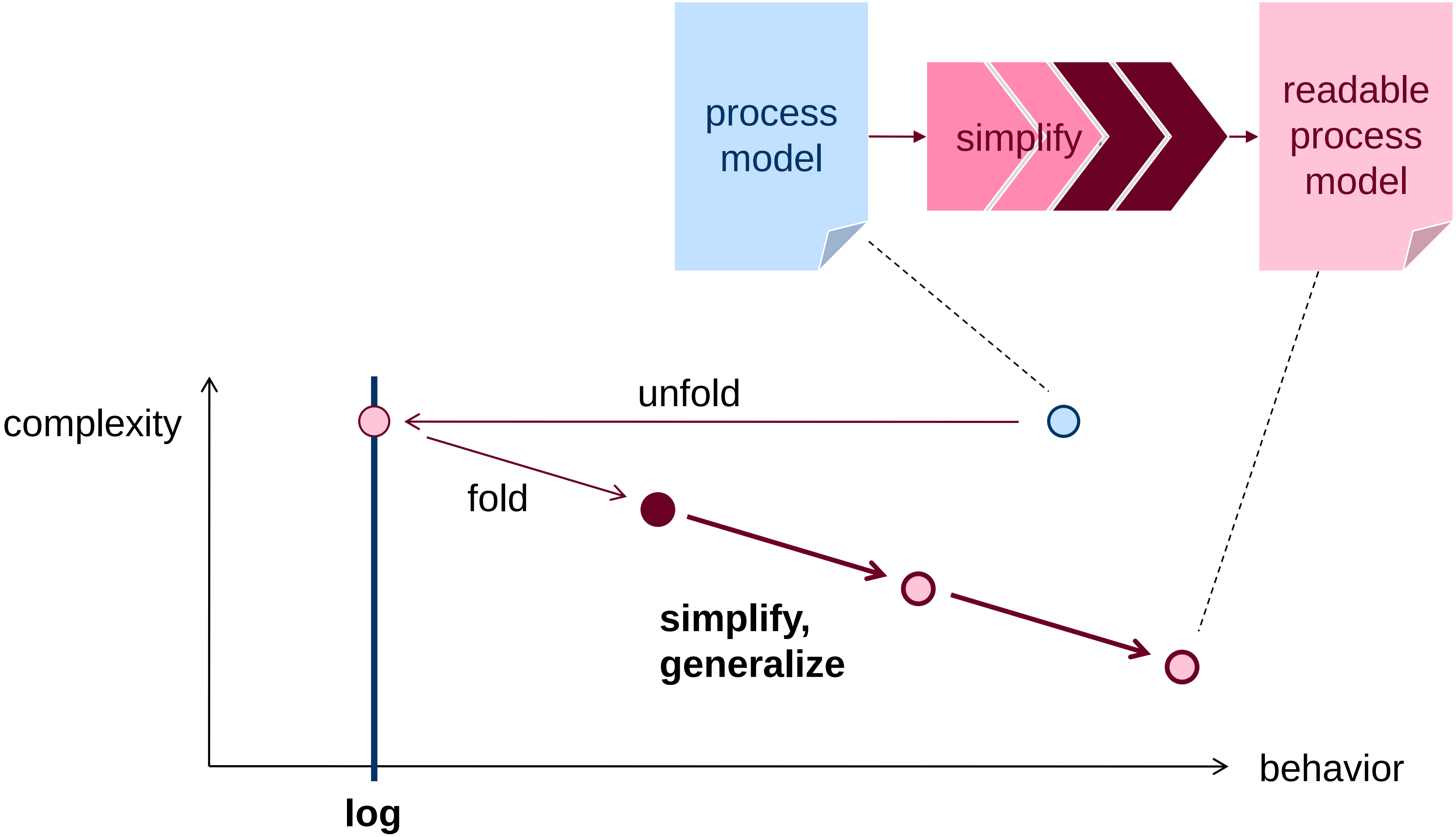
fold



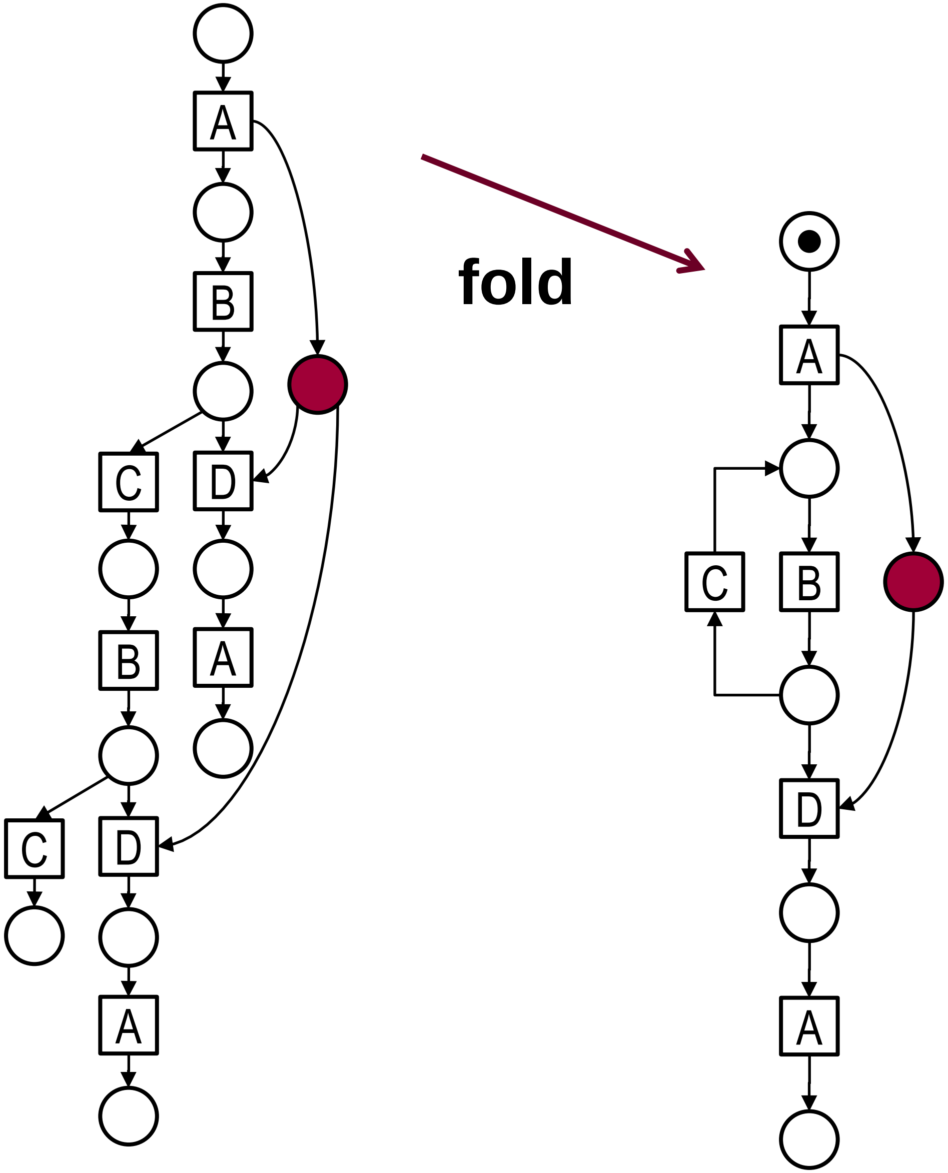
refolded vs. original model

- less behavior
  (replays the log and more)
- simpler structure

Next: Simplifying and Generalizing



Implied Places



fold

implied place

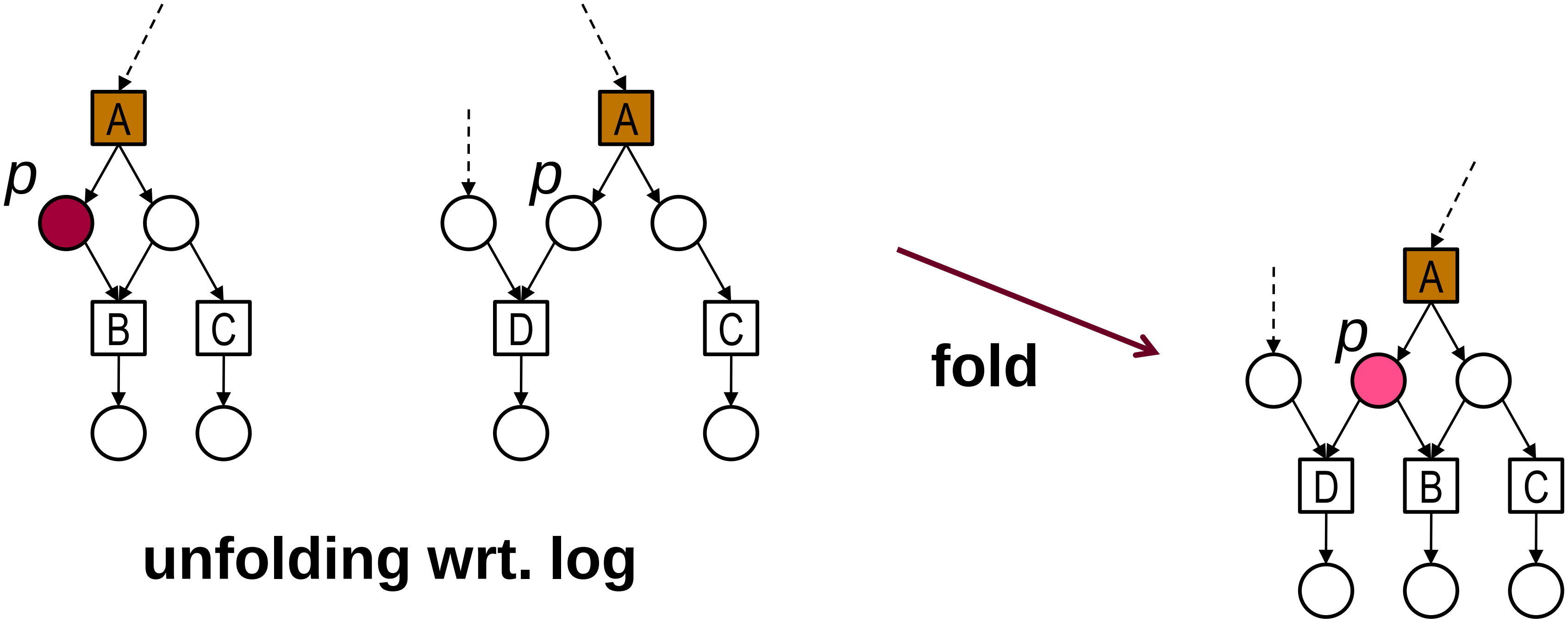
- does not restrict transitions

remove from folded model

- simpler model
- same behavior

various techniques to find implied places

Special: Implied Places and Folding

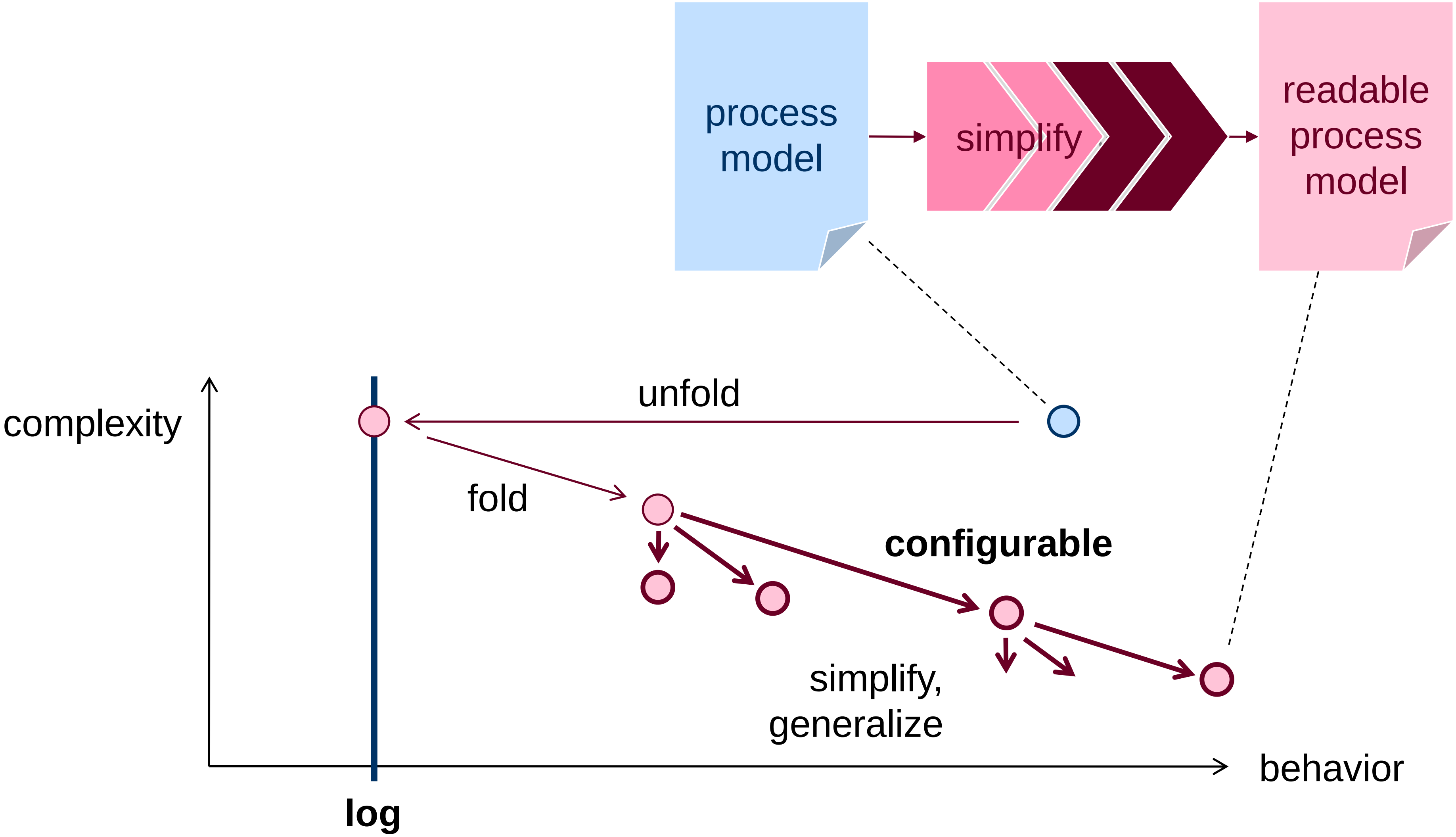


unfolding wrt. log

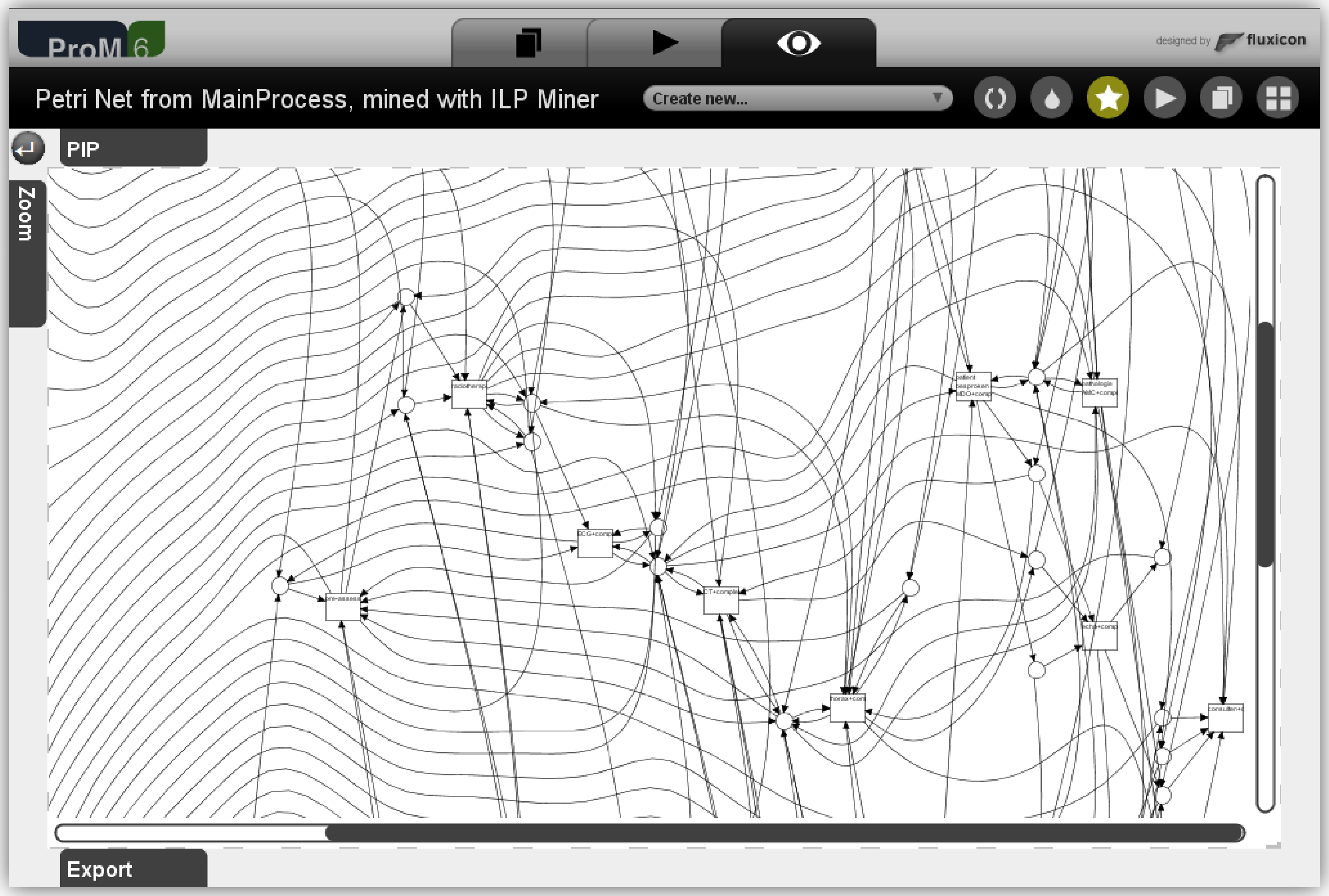
fold

- folding may merge implied and non-implied places
- remove *p*: simpler model,
  more behavior (generalization)
- **let user decide**

Configurable Simplification

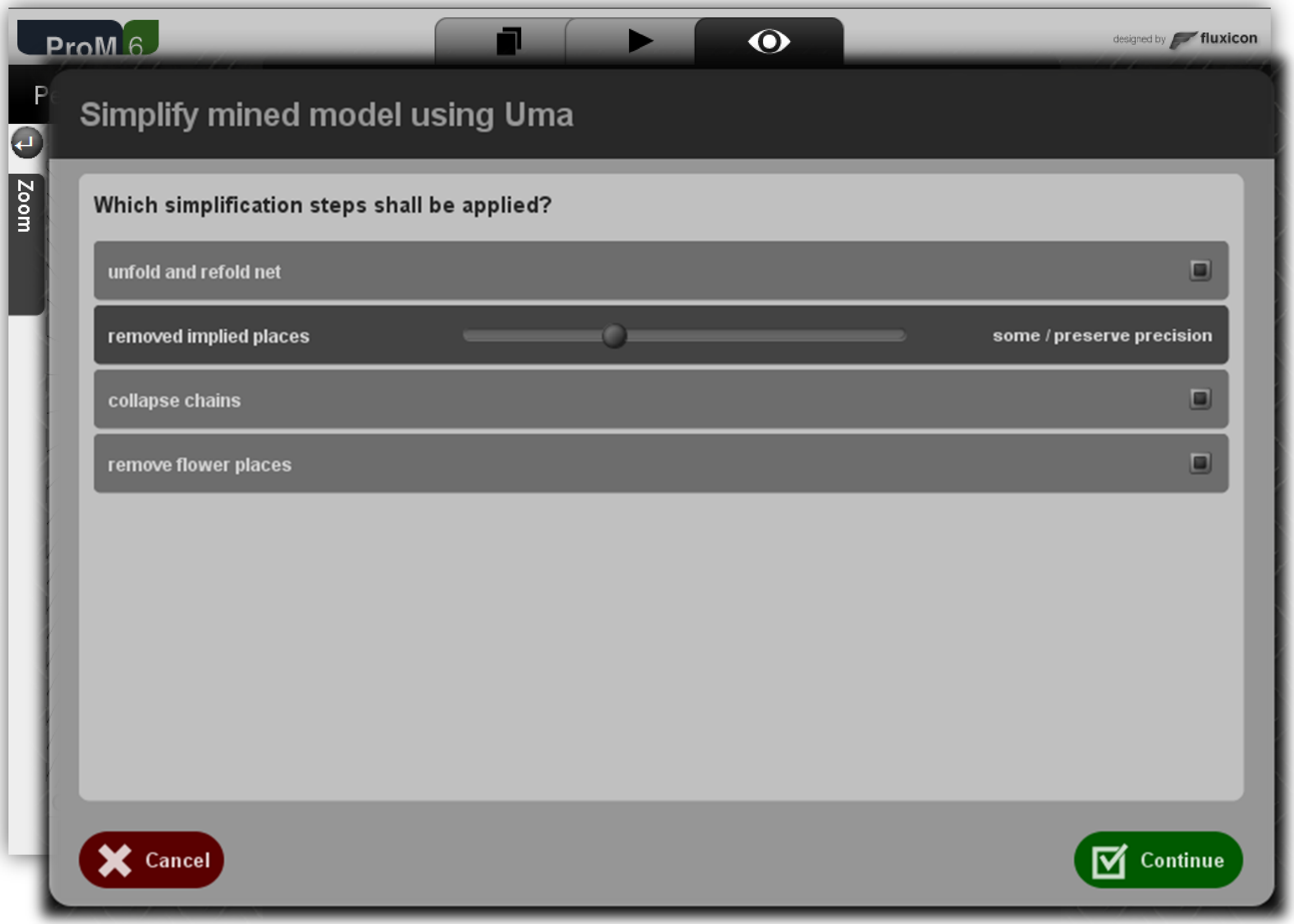


ProM6 / Uma > www.processmining.org



Petri Net from MainProcess, mined with ILP Miner

ProM6 / Uma > www.processmining.org

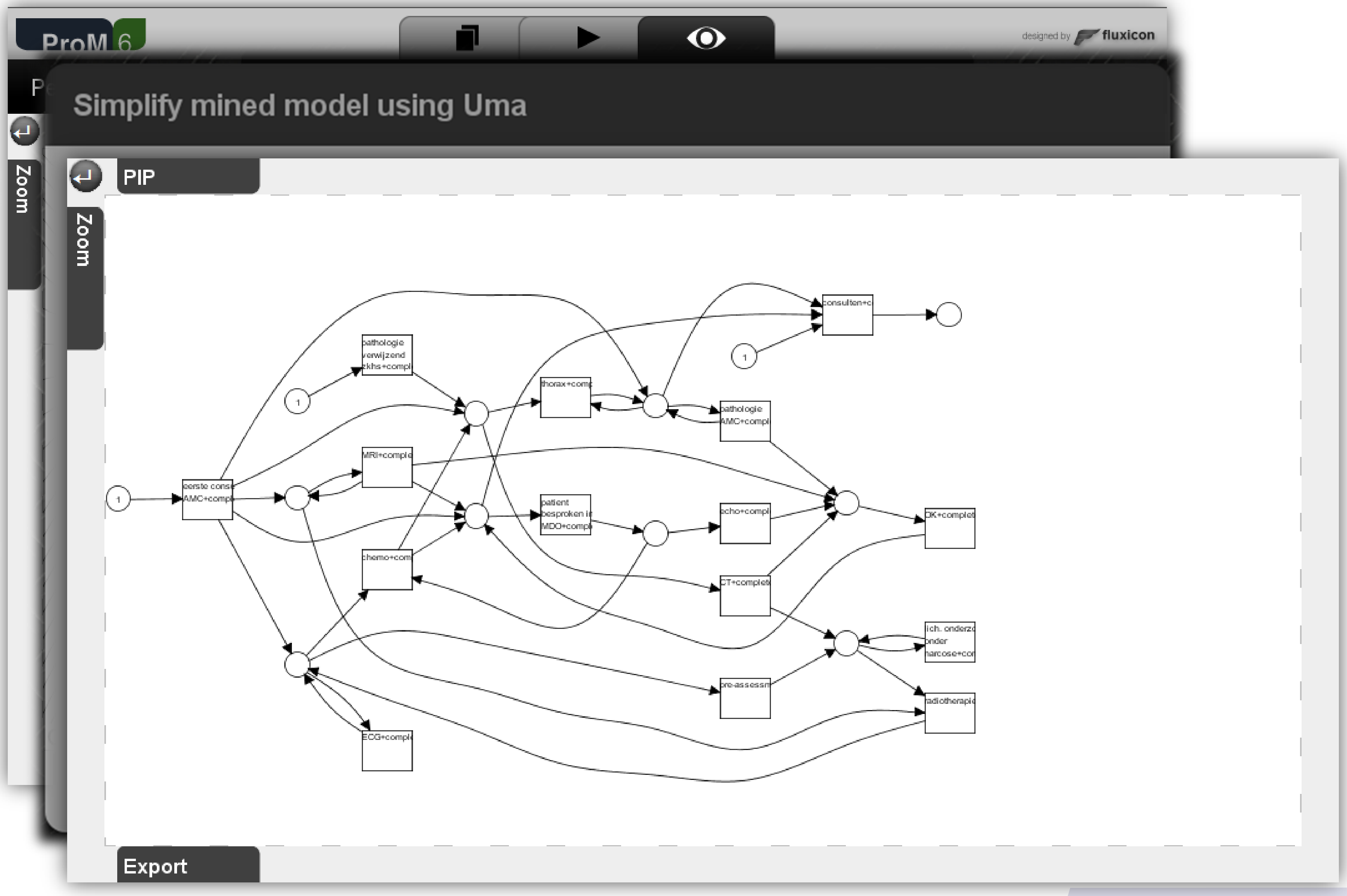


Simplify mined model using Uma

Which simplification steps shall be applied?

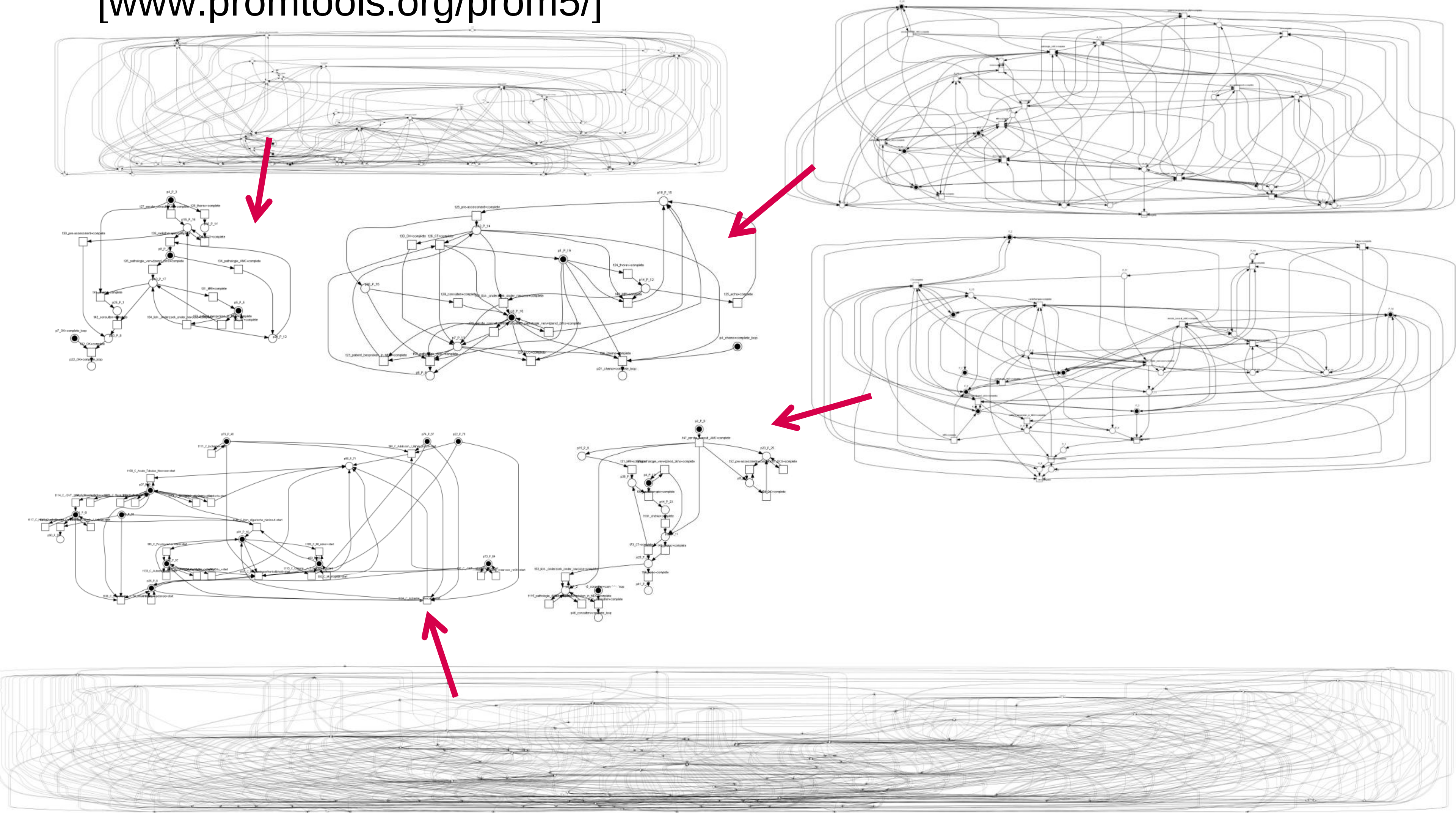
- unfold and refold net □
- removed implied places — some / preserve precision
- collapse chains □
- remove flower places □

ProM6 / Uma > www.processmining.org



Experimental Results

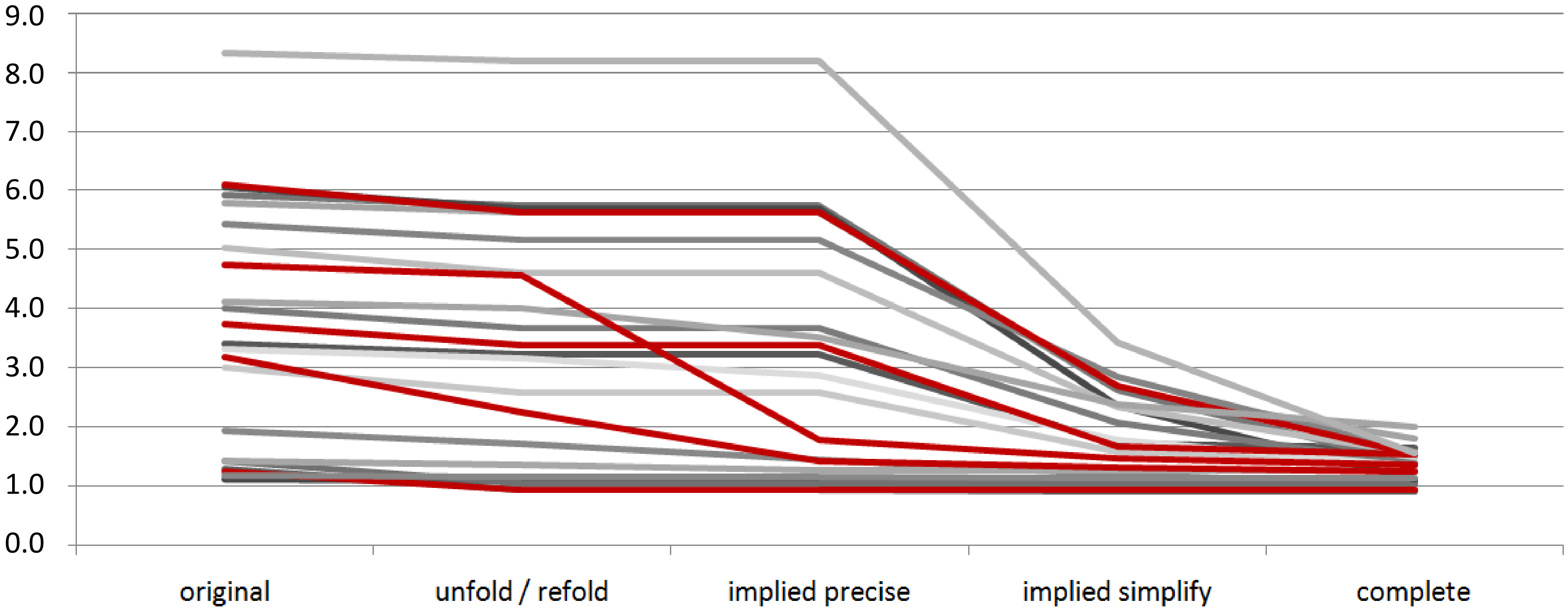
- 15 benchmark logs, 6 industrial logs
  [www.promtools.org/prom5/]



Experimental Results

- 15 benchmark logs, 6 industrial logs
  [www.promtools.org/prom5/]

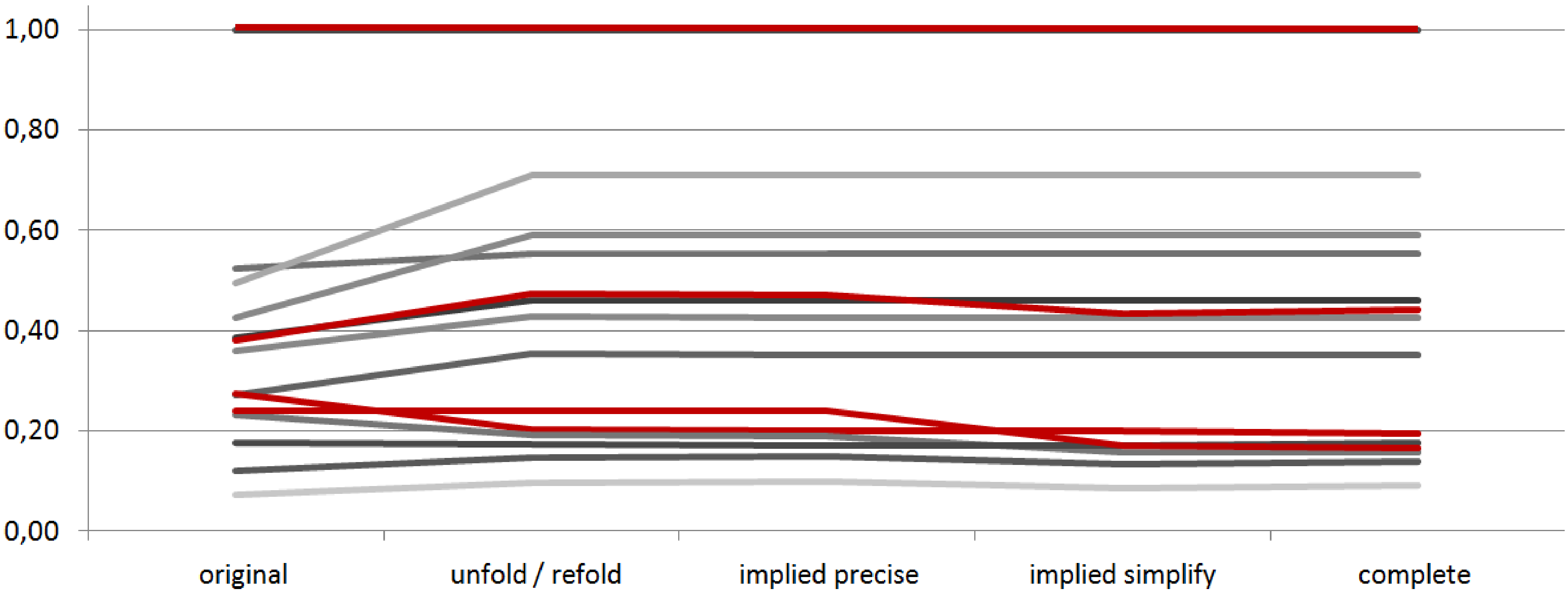
model complexity = #arcs / #nodes



time < 1 sec for industrial logs

Experimental Results

- **precision:** traces allowed by model and not in log
- 1.0 = only log behavior allowed



- rises/falls **within limits** (can be controlled)

from
spaghetti

to
lasagna?

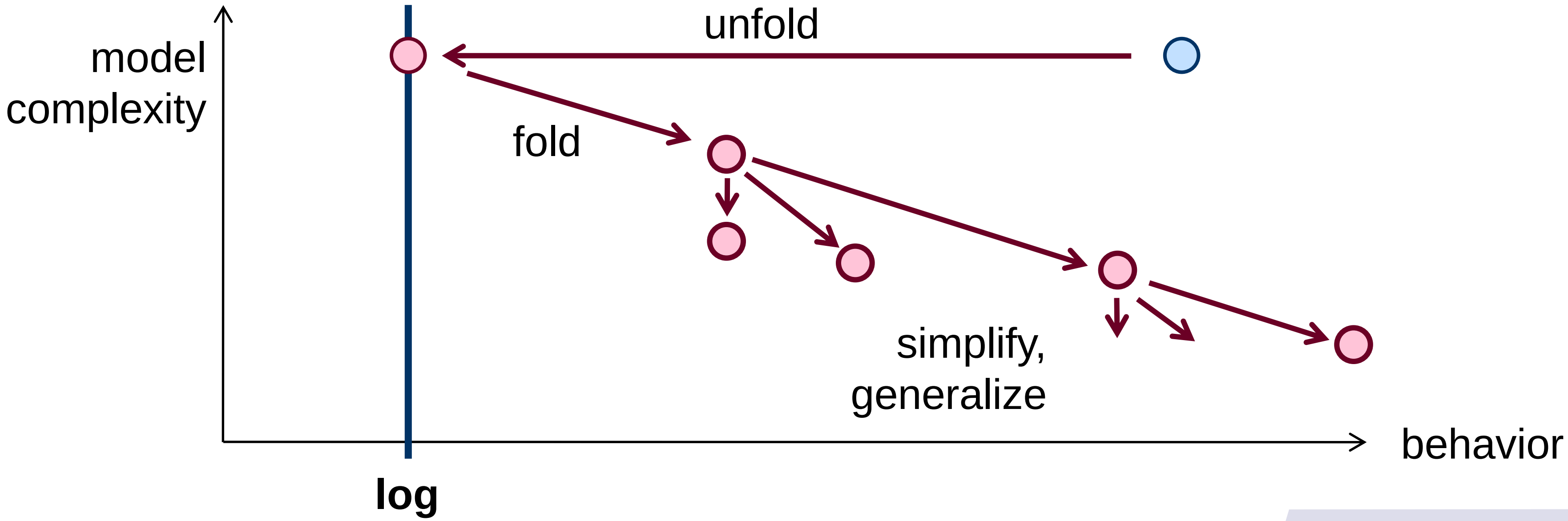
Genuss aus Bayern
Mit bestem Hartweizen und frischen Eiern aus kontrollierter Bodenhaltung
1898
Spaghetti
Bermbacher
Guten
Feine Frischei-Nudeln

from **spaghetti**

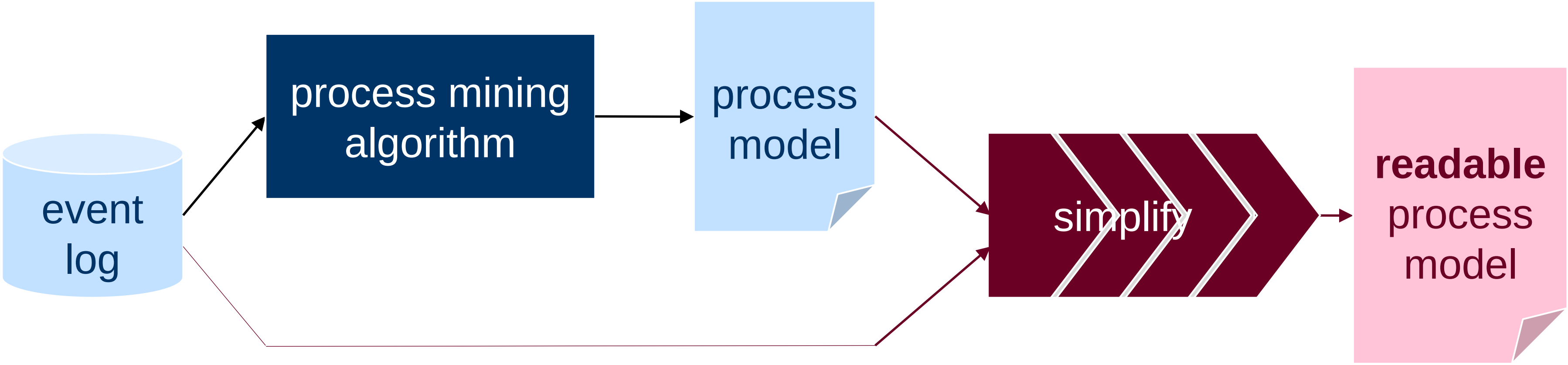
to **less complex spaghetti**

Lessons Learned

- techniques to navigate the **model/behavior space**
- use **model and log together**
- use **model unfoldings**
- break a rule and see what happens



And next?



- **process views**
most simple model covering 80% of the log

- **improve mining algorithms?**
we showed: there is room for improvement

Dirk **Fahland**

about.me/**dirk.fahland**

Simplifying Mined Process Models

Technische Universiteit **Eindhoven** University of Technology

Where innovation starts