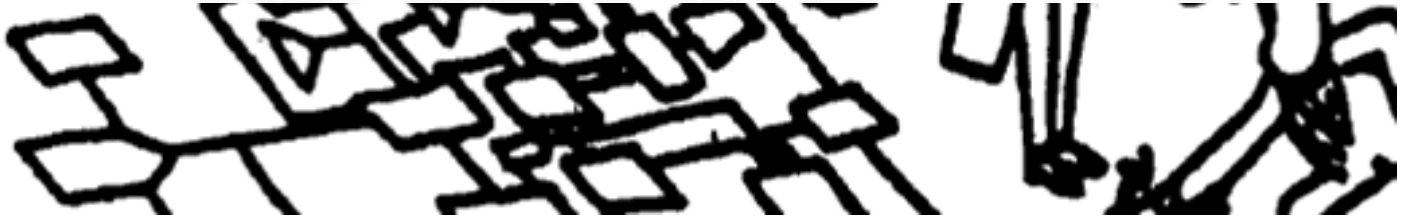
KU LEUVEN

Bidimensional Process Discovery for Mining BPMN Models

DeMiMoP 2014, ~~Haifa~~ Eindhoven

- Jochen De Weerdt, KU Leuven (@jochendw)
- Seppe vanden Broucke, KU Leuven (@macuyiko) (presenter)
- Filip Caron, KU Leuven

Jochen.deweerdt@kuleuven.be



Outline

- Introduction
- Rationale and Relevance
- Bidimensional Process Discovery
- Conclusions

Outline

- ✓ **Introduction**
- Rationale and Relevance
- Bidimensional Process Discovery
- Conclusions

Introduction

- Process mining: knowledge discovery from event logs
- Process discovery: discovery of some representational model for control-flow, social aspect, performance, etc.
- Many techniques proposed to mine: Petri nets, Heuristic Nets, Causal nets, EPCs, Declarative models

Where is the BPMN miner?

Introduction (2)

BPMN miner would be beneficial:

- Intuitive
- Known by many practitioners
- Easy to use and extend
- Comparison with as-is models

Outline

- ✓ Introduction
- ✓ **Rationale and Relevance**
- Bidimensional Process Discovery
- Conclusions

Relevance

Relevance for practitioners:

- BPMN as the de facto standard
- Widely adopted
- Used for documenting, improving, simulating and implementing business processes

Availability of BPMN miner could aid in:

- Automated process identification
- Facilitating the process re-engineering cycle
- Improved communication of process mining results

Relevance (2)

Relevance for education:

- Courses and text books often give BPMN a great deal of attention
- Later on, process mining is brought up, but requires introduction of new modeling standards (mostly Petri nets and Causal nets)
- Tools to mine BPMN is likely to lower effort required by educators to incorporate process mining in their units

Relevance (3)

Relevance for research:

- Process mining key contributions have mostly been technical in nature
- Opportunity for research on ease of use, user acceptance, etc. of process mining
- Process discovery technique with BPMN would lower barriers to conducting such studies

Comparison

Modeling Notation	Ease of Interpretation	Suitability Rep. Bias Proc. Disc.	Popularity (Modeling)	Popularity (Mining)
Petri net	●●○○○	●●○○○	●○○○○	●●●●○
Heuristic net	●●●○○	●●●●○	○○○○○	●●●●●
Fuzzy model	●●●●○	●●●●○	○○○○○	●●●●●
Causal net	●○○○○	●●●●●	○○○○○	●●○○○
EPC	●●●●○	●○○○○	●●●●○	●●○○○
BPMN	●●●●●	●●○○○	●●●●●	○○○○○

Outline

- ✓ Introduction
- ✓ Rationale and Relevance
- ✓ **Bidimensional Process Discovery**
- ▪ Conclusions

Preliminaries

- Event log L: multiset of traces
- Trace $\sigma \in L$: finite sequence of events

- O_L: set of originators
- T_L: set of activities

- $a: L \to T_L$: function returning actitvity for given event
- $o: L \to O_L$: function returning originator for given event

$$\sigma = \langle start^{alice}, register^{bob}, \dots \rangle$$

Control-flow Discovery

- BPMN construct assumed to be known:
  - Flow objects
  - Connecting objects
  - Artifacts
  - Swim lanes

- Our approach (currently) considers:
  - Gateways (XOR, AND)
  - Tasks
  - Sequence flow
  - Start and end events
  - Swim lanes

- Other works illustrate that only small subset of BPMN is used in most real-life environments

Control-flow Discovery (2)

- Control-flow discovery similar as Heuristics Miner algorithm:
  - Dependency information is derived to construct dependency graph
    $$D = \{(a,b) \mid a \in T_L \wedge b \in T_L \wedge \exists \sigma \in L: [\exists \sigma_i \in \sigma: [a(\sigma_i) = a \wedge a(\sigma_{i+1}) = b]]\}$$
  - Next, split and join information is derived. E.g. $\{(a,b),(a,c),(a,d)\} \subset D$: investigate whether *b, c, d,* occur in parallel, independently or a mixture of both. This is done by iterating over all traces and investigating succession and precedence relations
    $I: T_L \to P(P(T_L))$ and $O: T_L \to P(P(T_L))$
  - BPMN model constructed: add start and end events, construct BPMN graph with XOR and AND gateways, tasks and sequence flows
  - Simplification step: removal of all gateways which only contain single entry/exit and merging all AND gateways with the same outgoing activities and a single incoming activity (with additional XOR to connect incoming activities)

Filtering and Abstraction

- Dealing with variability and noise:
  - Standard Heuristics Miner dependency thresholds
  - Split/join threshold implement to select split/join patterns
  - For event logs with many low-frequent activities: filter/merging possibility

Bidimensional Discovery

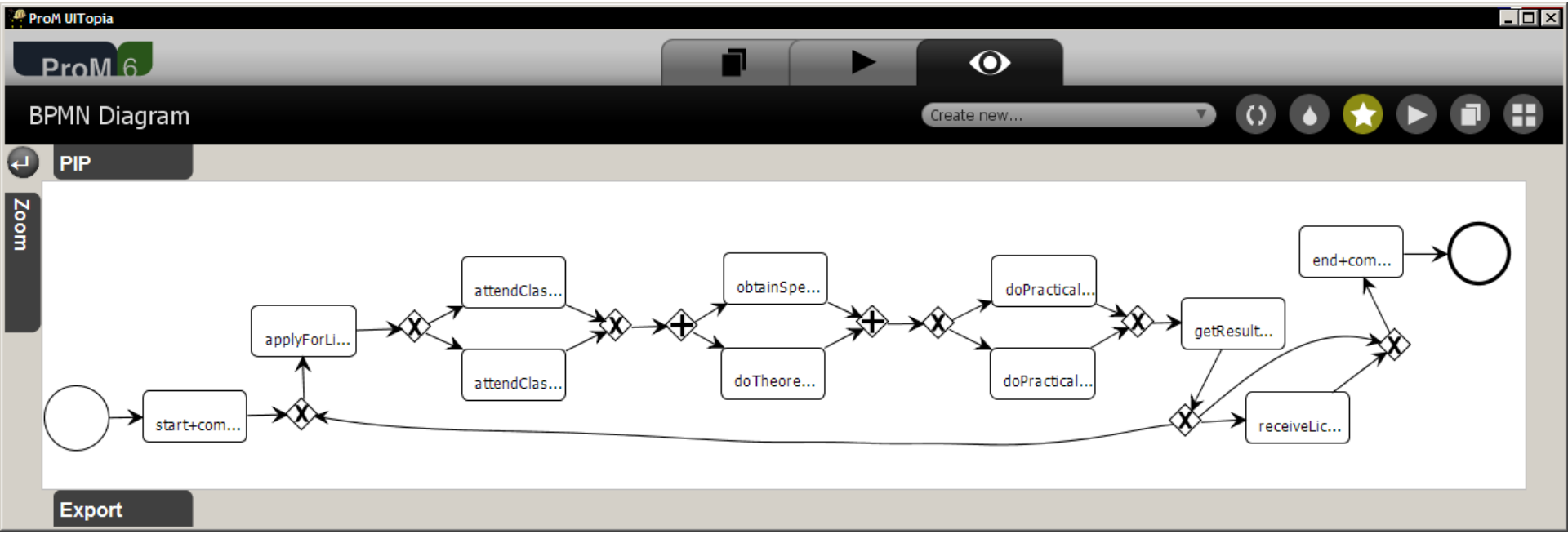
- Combine control-flow with other perspective: originator information
  - Using the swimlane construct of BPMN
  - Each swimlane represents a “worker pool”, or “role”
  - For each activity $a \in T_L$ a swimlane pool $S_i = \{a\}$ is constructed. Next, swimlanes S_i and S_j are merged so long iff $\exists o \in O_L, a_i \in S_i, a_j \in S_j, \sigma_i \in L, \sigma_j \in L$: $[a(\sigma_i) = a_i \wedge a(\sigma_j) = a_j \wedge o(\sigma_i) = o(\sigma_j) = o]$
  - Leads to set of merged swimlanes such that each swimlane represents a role (group of originators) responsible for a set of activities

Conformance Analysis and Exporting

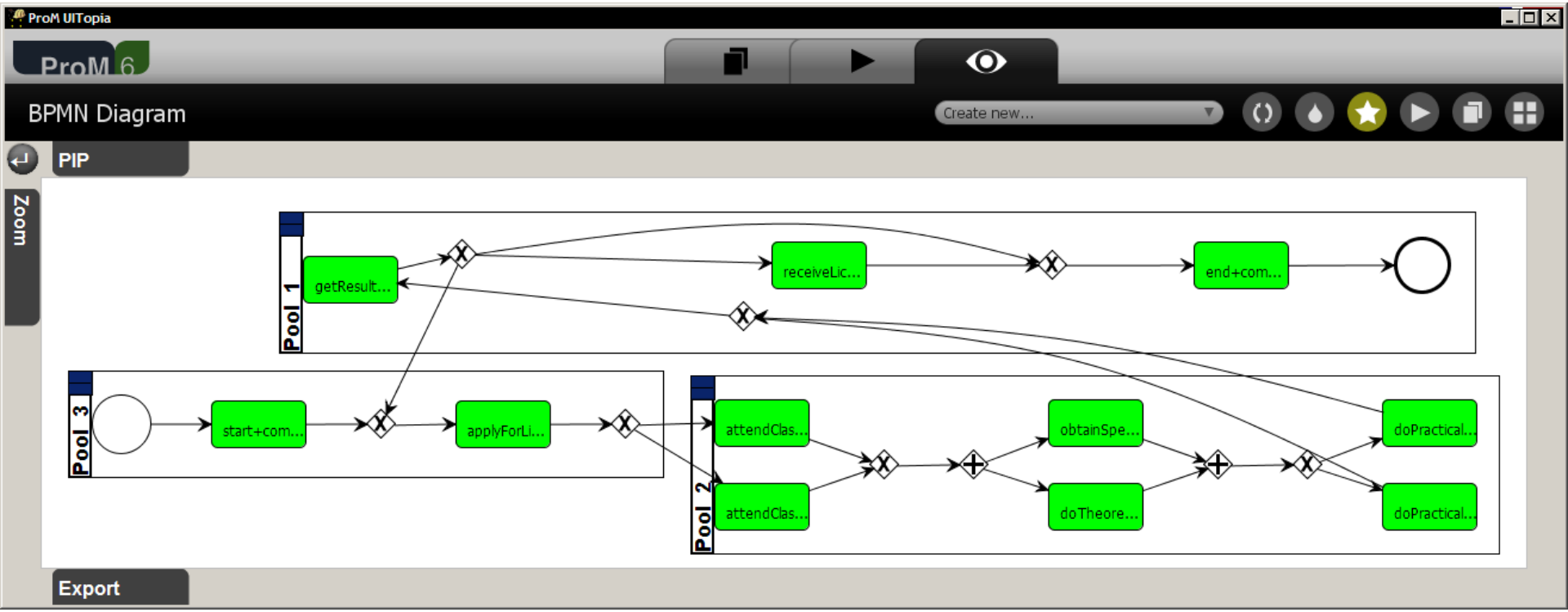
- Add ability to replay an event log over a discovered BPMN model (using token-based execution semantics) to derive fitness metric
  - Highlight conformance issues on activity-level

- Exporting to XPLD
  - Use in Aris, Bizagi, Signavio, Activity

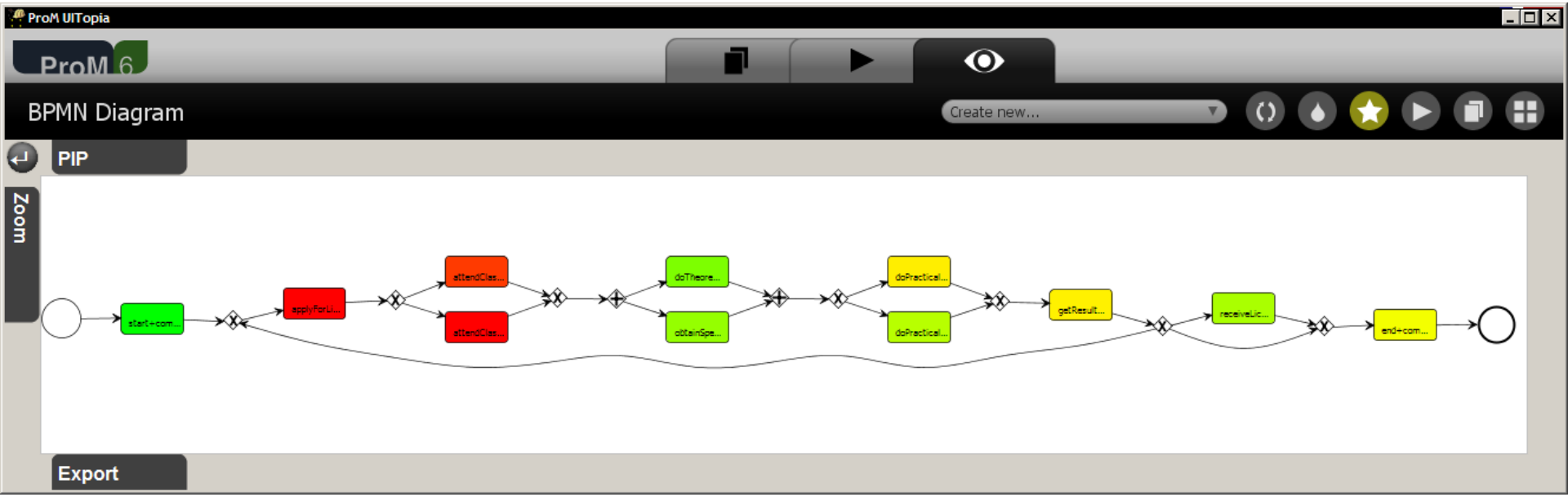
Illustrations



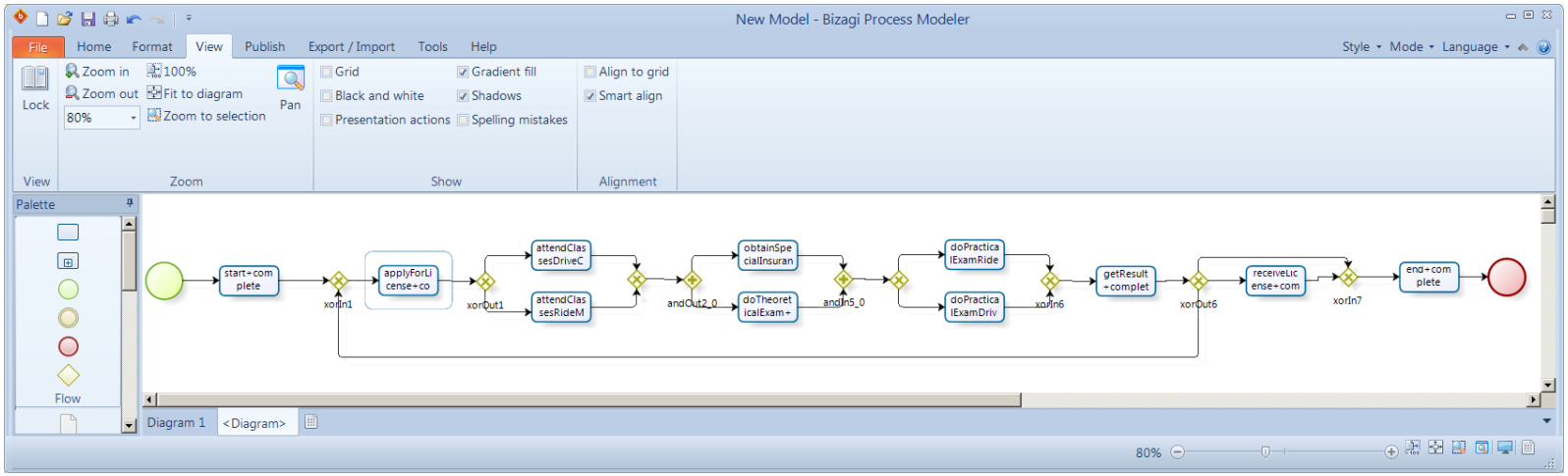
Illustrations (2)

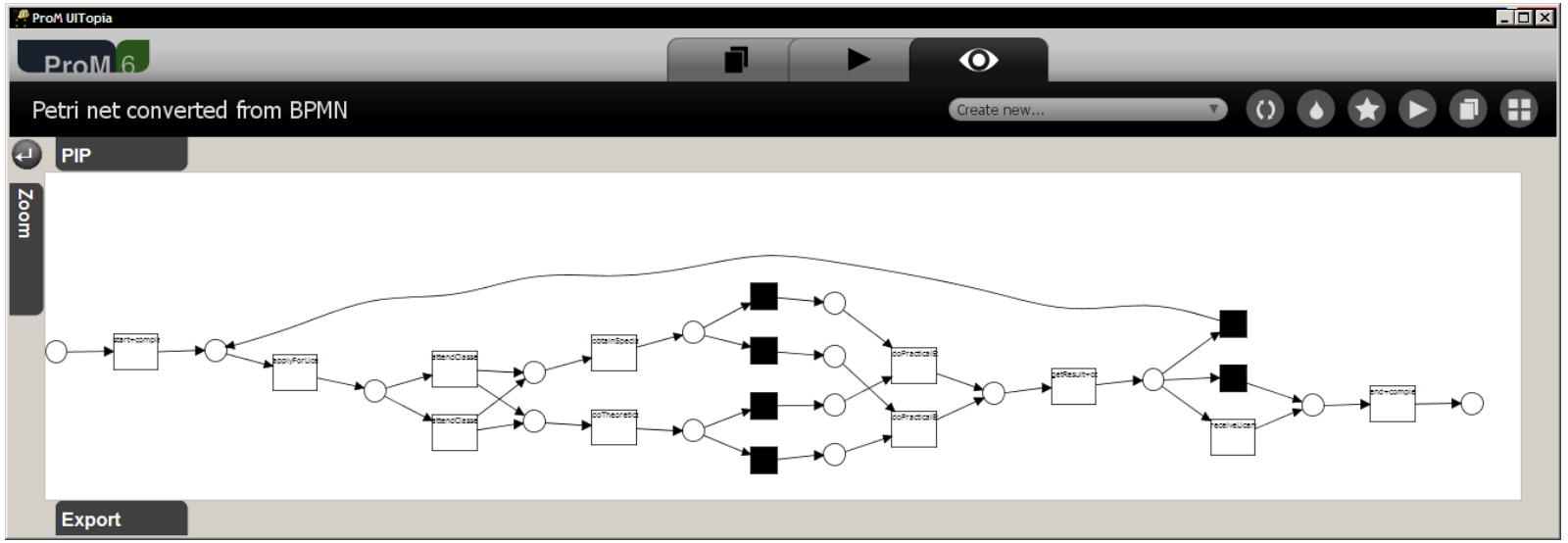


Illustrations (3)



Illustrations (4)





Outline

- ✓ Introduction
- ✓ Rationale and Relevance
- ✓ Bidimensional Process Discovery
- ✓ **Conclusions**

Conclusions

Summary:

- BPMN miner presented
- Combines control-flow with social/originator data
- Conformance checking, exporting, filtering functionality
- Hope to lower adoption barrier of process mining

Future work:

- Discover exceptions
- Discover sub-processes
- Discover data-flows
- Discover decision gateways (additional data attributes in event log required)
- Discover hierarchical process structures

Questions and Answers