



# Declarative Process Mining: Reducing Discovered Models Complexity by Pre-Processing Event Logs

**Pedro H. Piccoli Richetti**, Fernanda Araujo Baião, Flávia Maria Santoro

Department of Applied Informatics

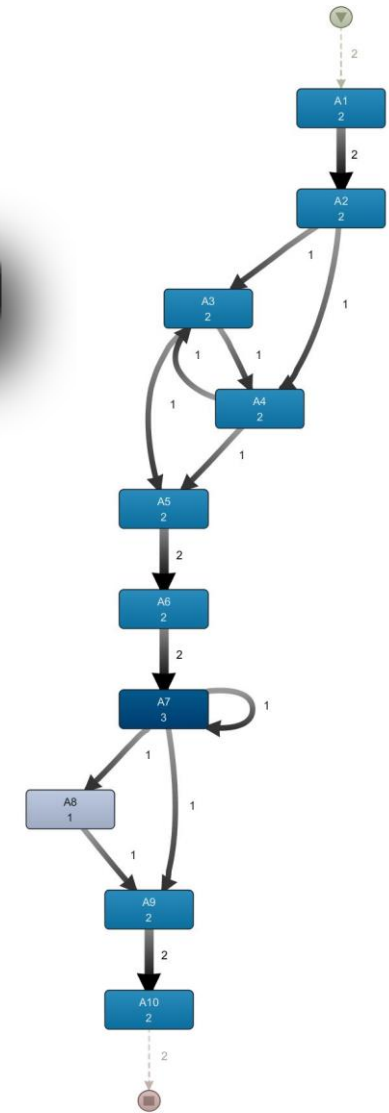
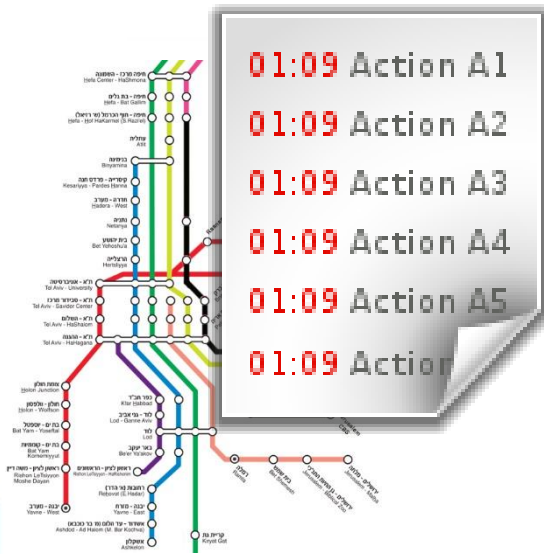
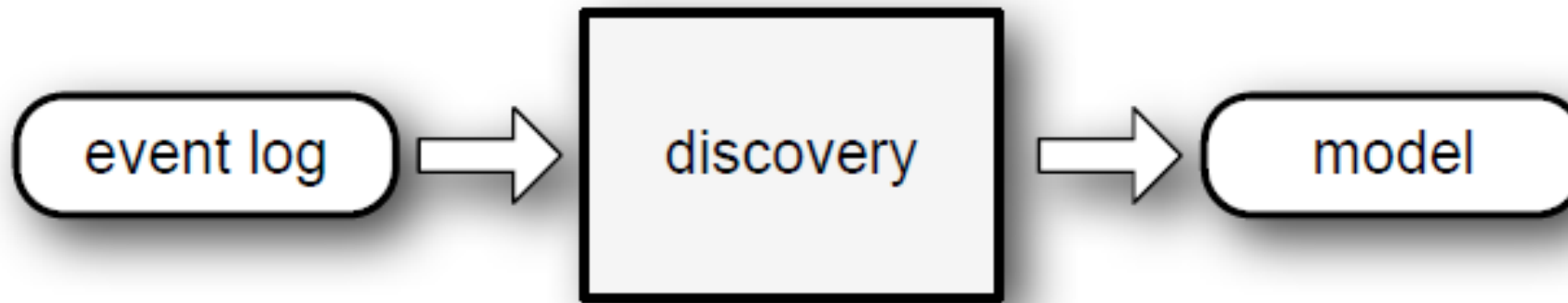
UNIRIO - Federal University of the State of Rio de Janeiro

Rio de Janeiro, Brazil

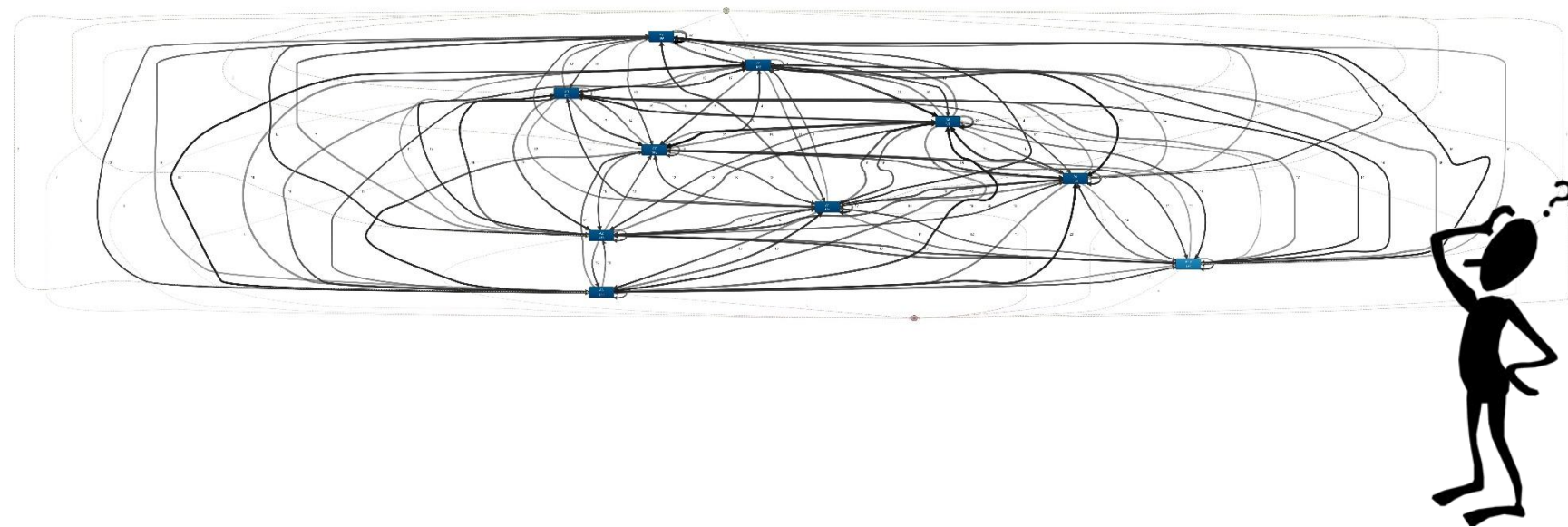
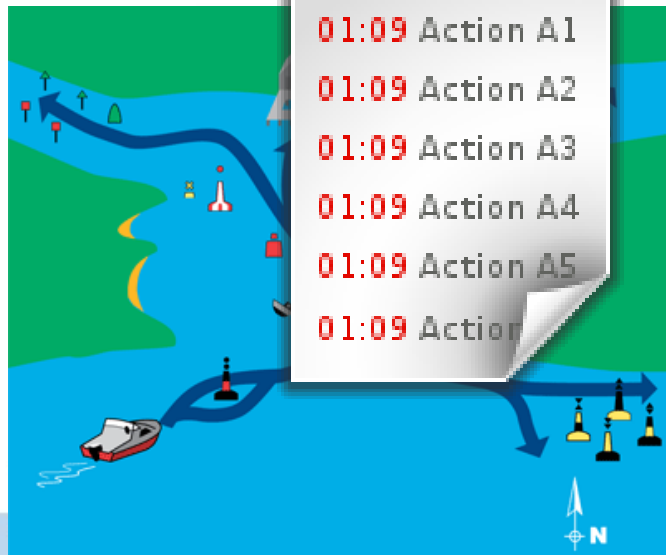
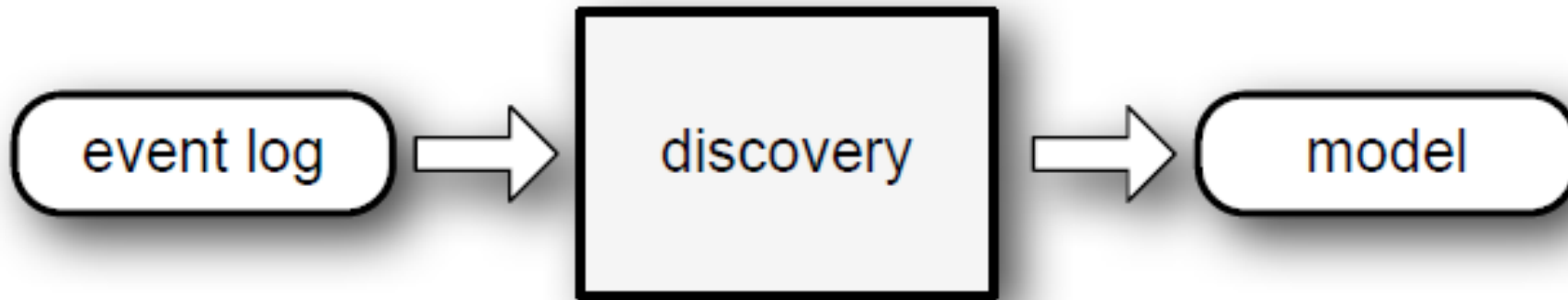
# Imperative x Declarative Paradigms



# Imperative Process Model Mining

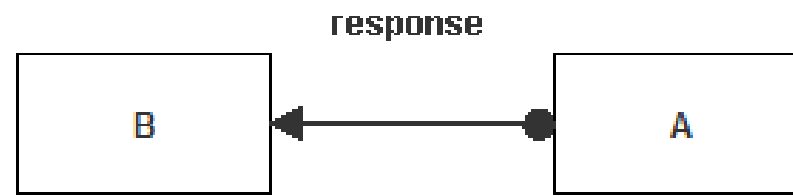


# Imperative Process Model Mining

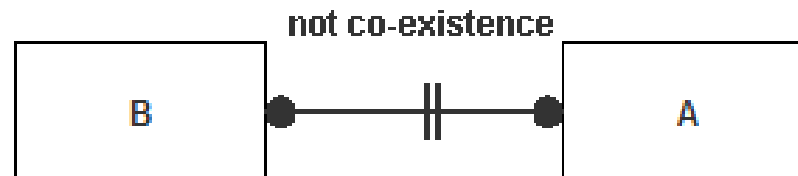


# Declarative Process Modeling

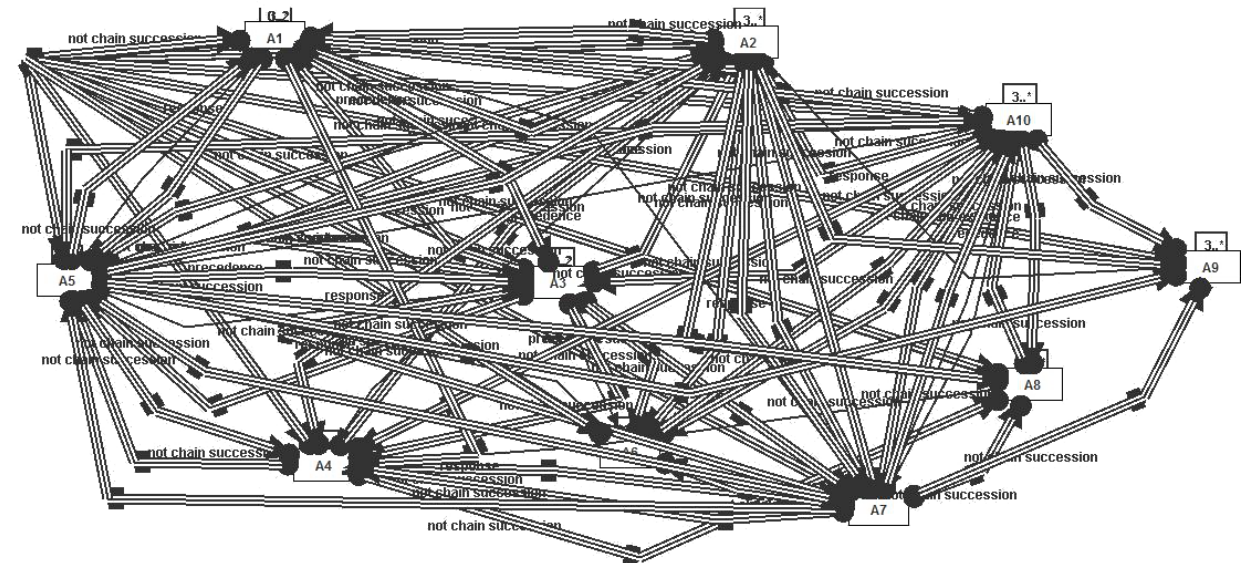
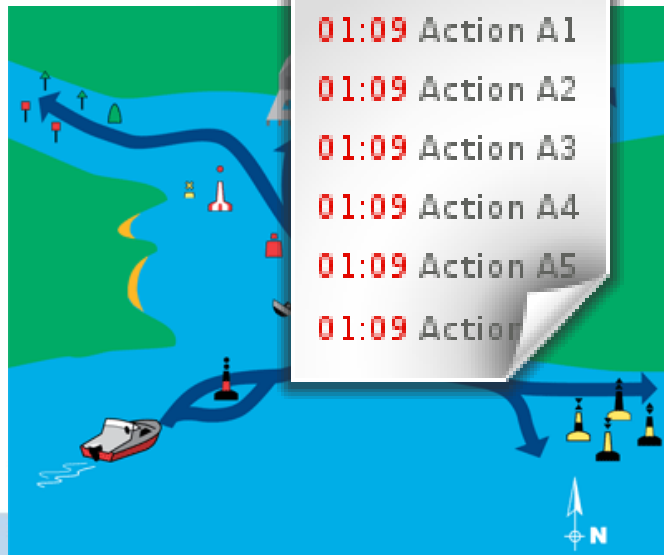
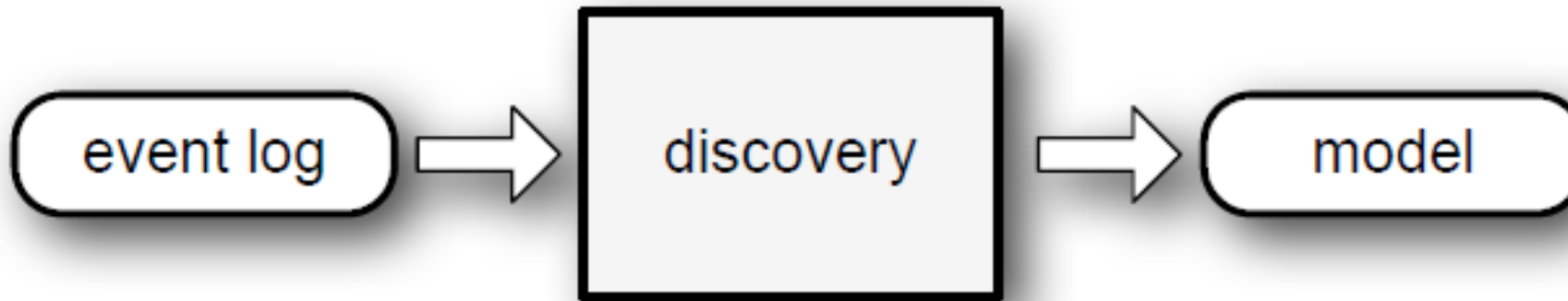
- Declare language (van der Aalst et al., 2009);
- Whenever activity "A" is executed, activity "B" has to be eventually executed afterwards.



- Only one of the two tasks "A" or "B" can be executed, but not both.



# Declarative Process Model Mining





# The problem of incomprehensibility of discovered declarative process models

- Declarative process mining techniques may produce models with a **high quantity of constraints, which may be incomprehensible for humans.** (Bose et al., 2013)
- The **combination of constraints** in a declarative process model might generate new hidden dependencies, which are **complex and difficult** to be identified by humans (Haisjackl et al., 2013).
- The **increasing number of restrictions negatively impacts** on the model quality. (Reijers et al., 2013)







# The problem of incomprehensibility of discovered declarative process models

- Declarative process mining techniques may produce models with a **high quality** **for** human

How to address this problem?

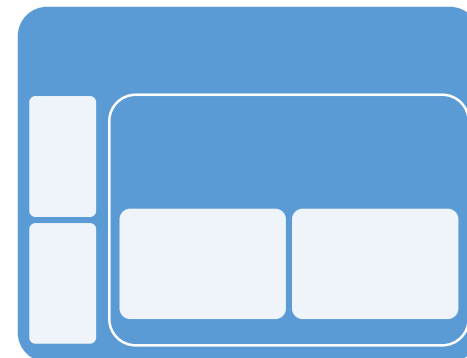
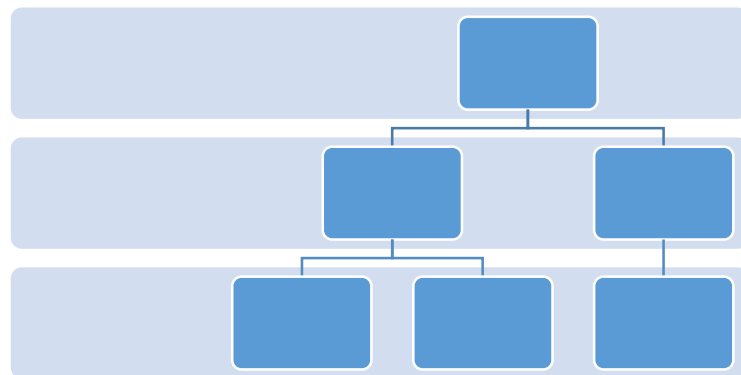
- The **combination of constraints** in a declarative process model might generate new hidden dependencies, which are **complex and difficult** to be identified by humans (Haisjackl et al., 2013).
- The **increasing number of restrictions negatively impacts** on the model quality. (Reijers et al., 2013)





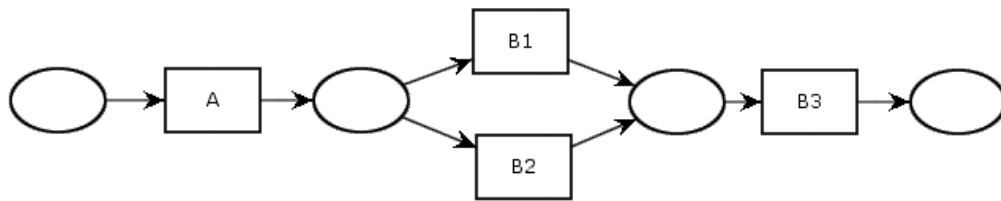
# Hierarchies on Business Process Models

- “**Abstraction** is seen as an effective approach to represent readable **models**, showing aggregated activities and hiding irrelevant details.” (Smirnov et al., 2011)
- “**Hierarchies** may be used to perform aggregation, thus **reducing the mental effort** to understand a model.” (Zugal et al., 2013)



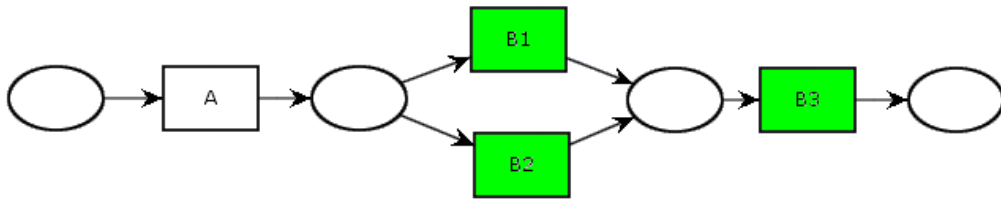
# Hierarchies on Business Process Models

- On imperative models, every process fragment ranging from a single entry and a single exit (SESE) can be grouped as a complex activity. (Weber et al., 2011)



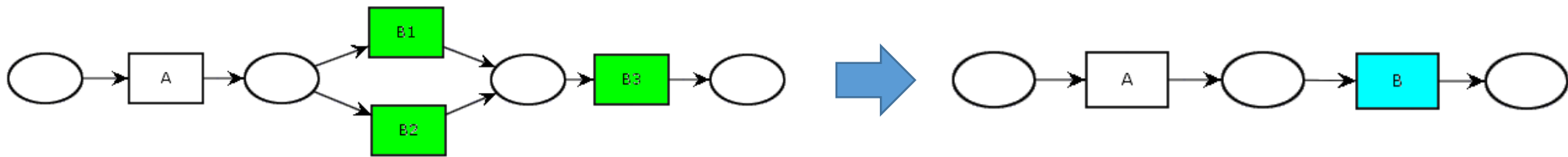
# Hierarchies on Business Process Models

- On imperative models, every process fragment ranging from a single entry and a single exit (SESE) can be grouped as a complex activity. (Weber et al., 2011)



# Hierarchies on Business Process Models

- On imperative models, every process fragment ranging from a single entry and a single exit (SESE) can be grouped as a complex activity. (Weber et al., 2011)



# Hierarchies on Business Process Models

- On imperative models, every process fragment ranging from a single entry and a single exit (SESE) can be grouped as a complex activity. (Weber et al., 2011)
- On **declarative models this structure is not informative enough**, because the activities' sequence is not rigid.
- The structural grouping of activities is inadequate and, for declarative models, **it should consider a common objective of the grouped activities**. (Zugal et al., 2013).

# Related Work

Approach	Authors
The search for sequential patterns on event logs and their replacement by abstract activities.	Li et al. (2011)
A user-guided discovery of declarative process models and a collection of post processing techniques to simplify and repair discovered declarative models.	Maggi et al. (2011), (2013)
The discovery of hierarchical process models using ProM, by preprocessing an event log, based on pattern abstractions relative to sequences in event log traces.	Bose et al. (2012)
The construction of abstraction layers in process models by matching events and activities.	Baier et al. (2013)

# Related Work

Approach	Authors
The search for sequential patterns on event logs and their	Li et al. (2011)
None of these approaches addresses abstraction techniques on automatically <u>mined declarative process models</u> in order to reduce their complexity.	
The discovery of hierarchical process models using ProM, by preprocessing an event log, based on pattern abstractions relative to sequences in event log traces.	Bose et al. (2012)
The construction of abstraction layers in process models by matching events and activities.	Baier et al. (2013)



# Objective

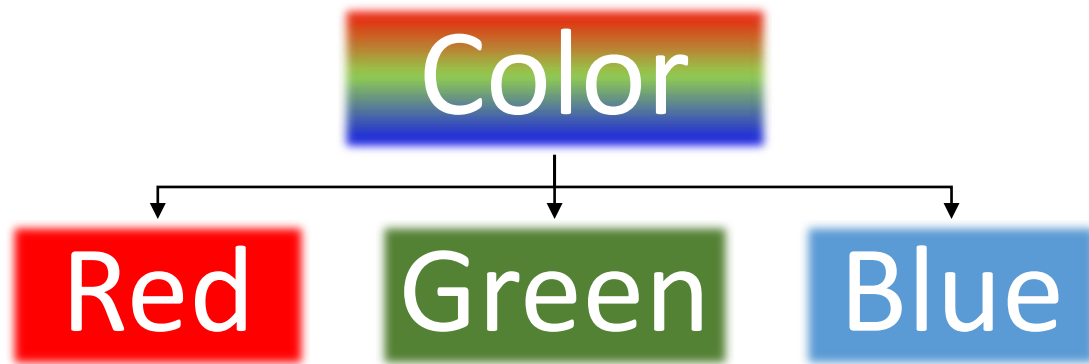
- Mining hierarchical Declare models using a linguistic hierarchy of activities.
- The idea is to group activities with common semantics instead of using process structure to create groups.

# A Method to Abstract Activities through Semantic Relations

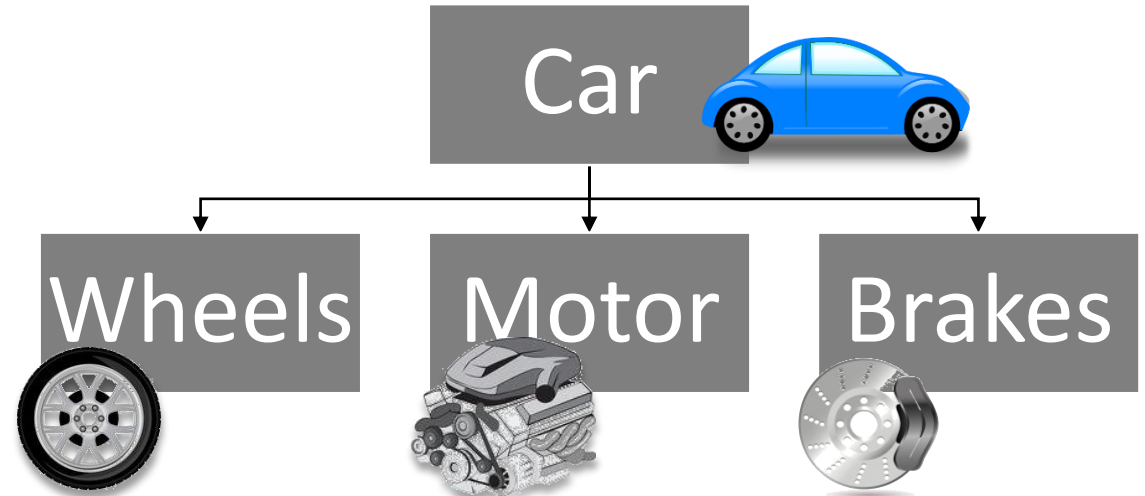
- Inspired by the semantic approach of Leopold et al. (2013) to name imperative process models and fragments, our approach applies [Natural Language Processing](#) to identify common objectives between activity labels, and then abstracts these activities into hierarchies.

# A Method to Abstract Activities through Semantic Relations

## Hypernymy



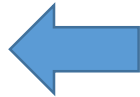
## Holonymy



# A Method to Abstract Activities through Semantic Relations



1	<prepare teaching sequence, decide on teaching method, give lessons>
2	<decide on teaching method, prepare a lesson in detail, give lessons>
3	<prepare a lesson in detail, give lessons>



- decide on teaching method
- prepare teaching sequence
- prepare a lesson in detail
- give lessons

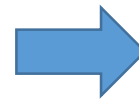
Process name: “How to prepare oneself and materials for teaching pupils”  
(Haisjackl et al. 2013)

# A Method to Abstract Activities through Semantic Relations



- Decide/V teaching/N method/N
- Prepare/V teaching/N sequence/N
- Prepare/V lesson/N detail/N
- Give/V lessons/N

For each Noun and Verb, we look for its hypernyms and holonyms.



teaching#n,0  
doctrine#n#1  
education#n#4  
activity#n#1  
act#n#2  
belief#n#1  
education#n#4  
activity#n#1  
profession#n#2  
doctrine#n#1

# A Method to Abstract Activities through Semantic Relations

$p_a = [\text{Prepare/V lesson/N detail/N}], [\text{Give/V lessons/N}]$

Pairs of activity labels are generated

prepare#v,0

sound#v

make#v

learn#v

change#v

educate#v

teach#v

cause#v

initiate#v

inform#v

give#v,0

move#v

release#v

change#v

submit#v

use#v

make#v

submit#v

use#v

inform#v

prepare#v,0

give#v,0

Best match: **make#v#39**

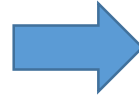
# A Method to Abstract Activities through Semantic Relations

$p_a = [\text{Prepare/V lesson/N detail/N}], [\text{Give/V lessons/N}]$

prepare#v

give#v

Best match: **make#v#39**



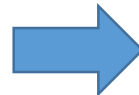
prepare#v#2, make#v#39 = 1

give#v#13, make#v#39 = 0.407

lesson#n

lesson#n

Best match: **lesson#n#4**



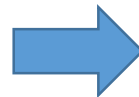
lesson#n#4, lesson#n#4 = 1

lesson#n#4, lesson#n#4 = 1

detail#n

lesson#n

Best match: **none**



detail#n, none = 0

lesson#n, none = 0



# A Method to Abstract Activities through Semantic Relations

$p_a = [\text{Prepare/V lesson/N detail/N}], [\text{Give/V lessons/N}]$

Lin's Similarity  
Metric

prepare#v

prepare#v#2, make#v#39 = 1

= 0.407

The similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are.

Lin, Dekang. "An information-theoretic definition of similarity." *ICML*. Vol. 98. 1998.

#4 = 1

lesson#n

lesson#n#4, lesson#n#4 = 1

Best

This similarity definition has good correlation with human judgments.

detail#n

detail#n, none = 0

lesson#n

lesson#n, none = 0

Best match: none

# A Method to Abstract Activities through Semantic Relations

$p_a = [\text{Prepare/V lesson/N detail/N}], [\text{Give/V lessons/N}]$

prepare#v

give#v

Best m

$$\text{average semantic relatedness value} = \frac{\sum[Lin]}{\text{No. of concepts}}$$

prepare#v#2, make#v#39 = 1

give#v#13, make#v#39 = 0.407

lesson#n

lesson#n

Best match

[prepare a lesson in detail, give lessons] = 0.567

lesson#n#4, lesson#n#4 = 1

#n#4 = 1

detail#n

lesson#n

Best match: none

detail#n, none = 0

lesson#n, none = 0

# A Method to Abstract Activities through Semantic Relations

[decide on teaching method; prepare teaching sequence; 0.421]

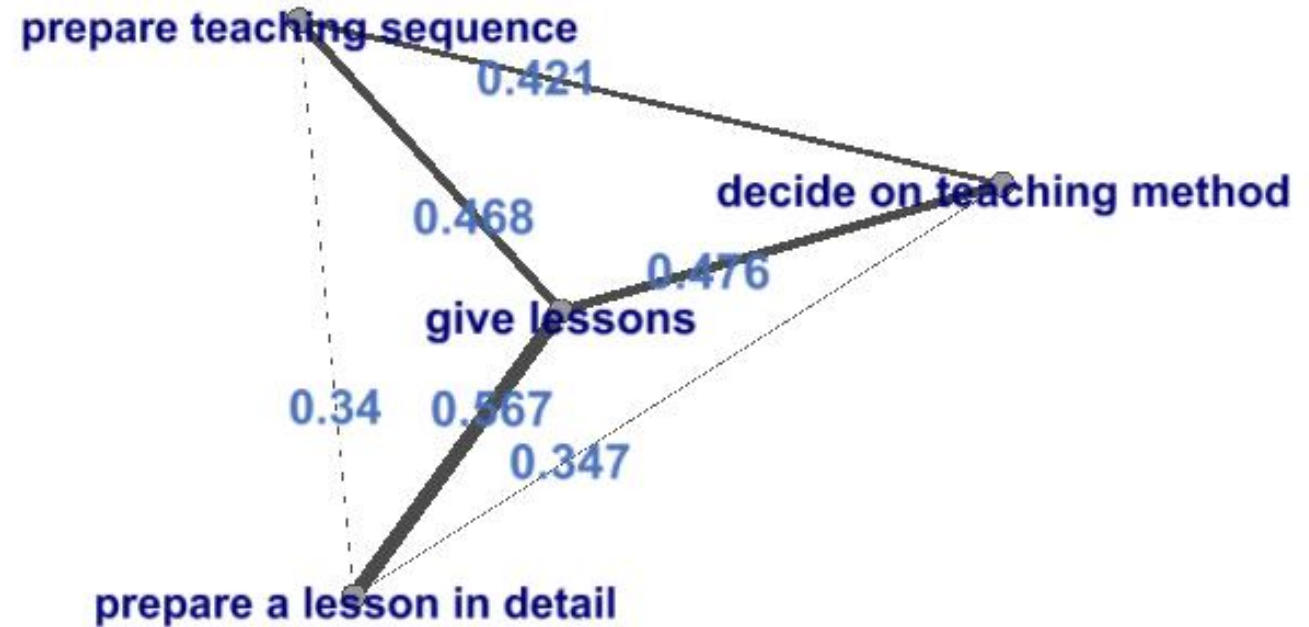
[decide on teaching method; prepare a lesson in detail; 0.347]

[decide on teaching method; give lessons; 0.476]

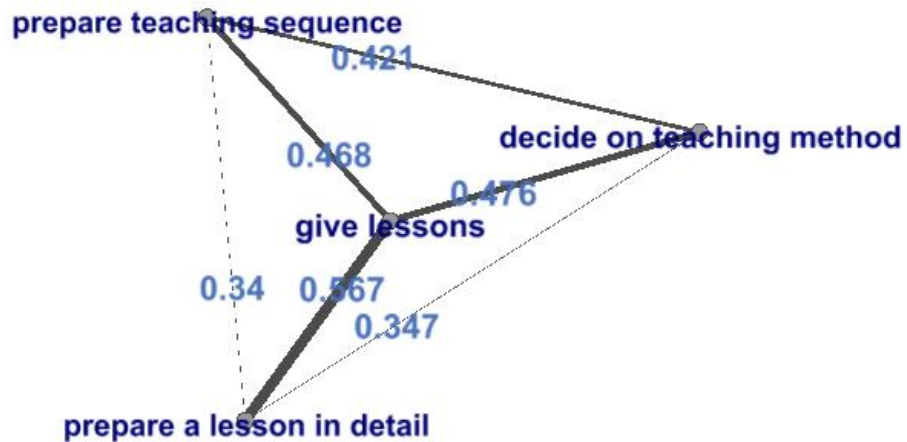
[prepare teaching sequence; prepare a lesson in detail; 0.340]

[prepare teaching sequence; give lessons; 0.468]

[prepare a lesson in detail; give lessons; 0.567]



# A Method to Abstract Activities through Semantic Relations



0->decide on teaching method  
1->prepare teaching sequence  
2->prepare a lesson in detail  
3->give lessons

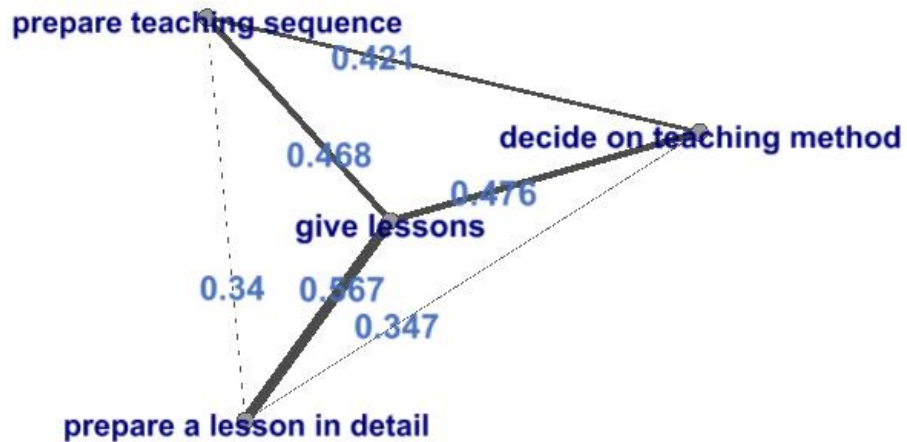
Threshold: 0

Group

[0, 1]  
[0, 2]  
[1, 2]  
[0, 1, 2]  
[0, 3]  
[1, 3]  
[0, 1, 3]  
[2, 3]  
[0, 2, 3]  
[1, 2, 3]  
[0, 1, 2, 3]

Analysis of all possible fully connected subgraphs

# A Method to Abstract Activities through Semantic Relations



0->decide on teaching method  
1->prepare teaching sequence  
2->prepare a lesson in detail  
3->give lessons

Threshold: 0

Group	EdgeSum
[0, 1]	0.421
[0, 2]	0.347
[1, 2]	0.340
[0, 1, 2]	1.108
[0, 3]	0.476
[1, 3]	0.468
[0, 1, 3]	1.365
[2, 3]	0.567
[0, 2, 3]	1.391
[1, 2, 3]	1.376
[0, 1, 2, 3]	2.621

Analysis of all possible fully connected subgraphs

# A Method to Abstract Activities through Semantic Relations

prepare teaching sequence

decide on teaching method

give lessons

0.567

prepare a lesson in detail

Only edges with weight above 0.5 are considered (User selection).

0->prepare a lesson in detail

1->give lessons

Threshold: 0.5

Group

EdgeSum

[0, 1]

0.567

Selected Group:

[prepare a lesson in detail, give lessons]

# Preprocessing the event log

Selected Group:  
[prepare a lesson in detail, give lessons]

Suggested complex activity:  
[prepare and give lessons]

1	<prepare teaching sequence, decide on teaching method, <a href="#">give lessons</a> >
2	<decide on teaching method, <a href="#">prepare a lesson in detail</a> , <a href="#">give lessons</a> >
3	< <a href="#">prepare a lesson in detail</a> , <a href="#">give lessons</a> >

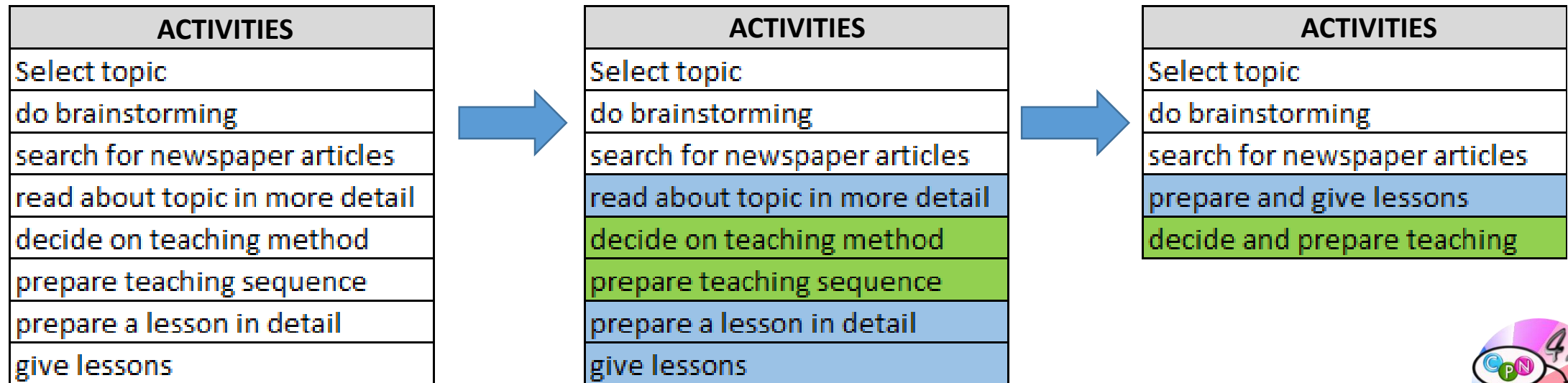


1	<prepare teaching sequence, decide on teaching method, <a href="#">prepare and give lessons</a> >
2	<decide on teaching method, <a href="#">prepare and give lessons</a> , <a href="#">prepare and give lessons</a> >
3	< <a href="#">prepare and give lessons</a> >

“Spaghetteness” of process models can be reduced by first mining common constructs or functionalities, abstract them and then discovering process models on the abstracted log. (Bose, 2009)



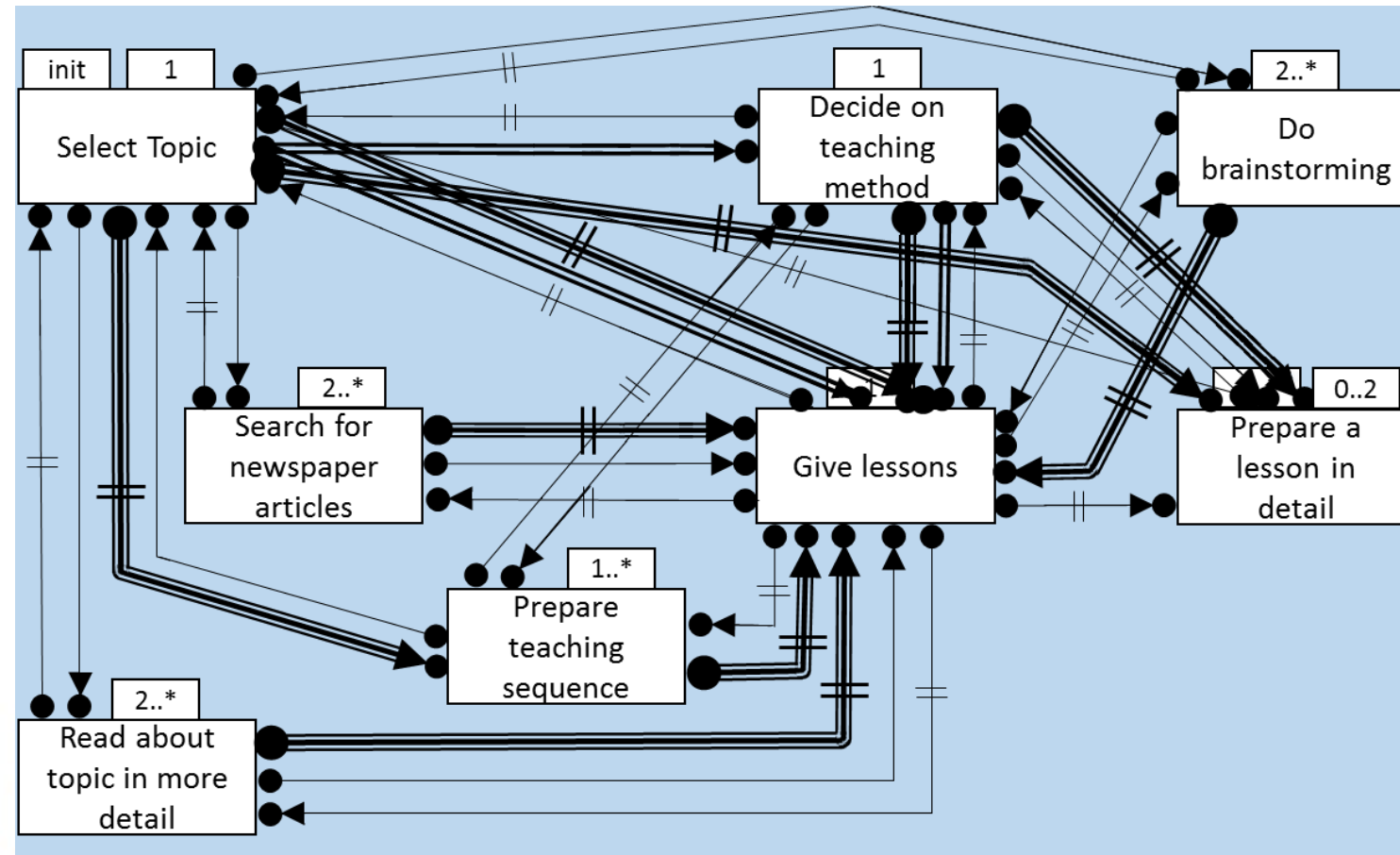
# Case Study – Abstraction Method Execution



Process name: **“How to prepare oneself and materials for teaching pupils”**

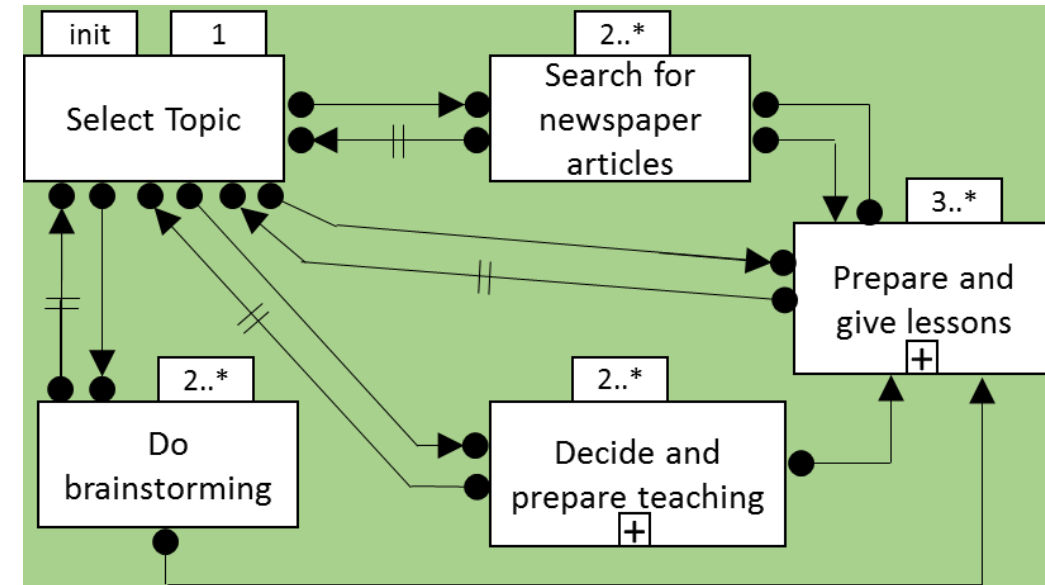
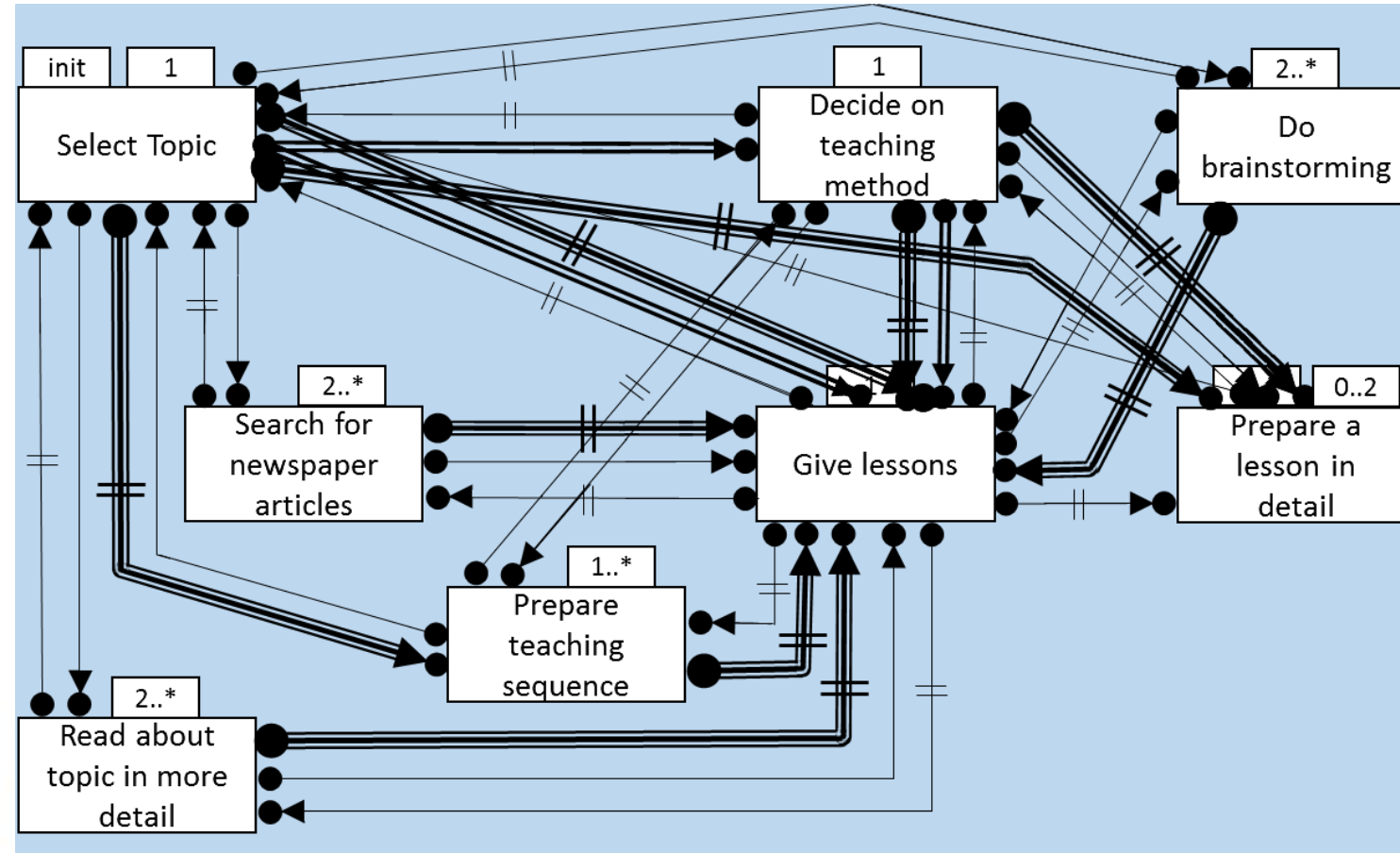
Haisjackl, C., et al.: Making Sense of Declarative Process Models: Common Strategies and Typical Pitfalls. In: BPMDS 2013 and EMMSAD 2013. LNBIP, vol. 147, pp. 2–17. Springer, Heidelberg (2013)

# Case Study – Flat Model Discovery



Declare Model from the original event log.  
(Flat model)

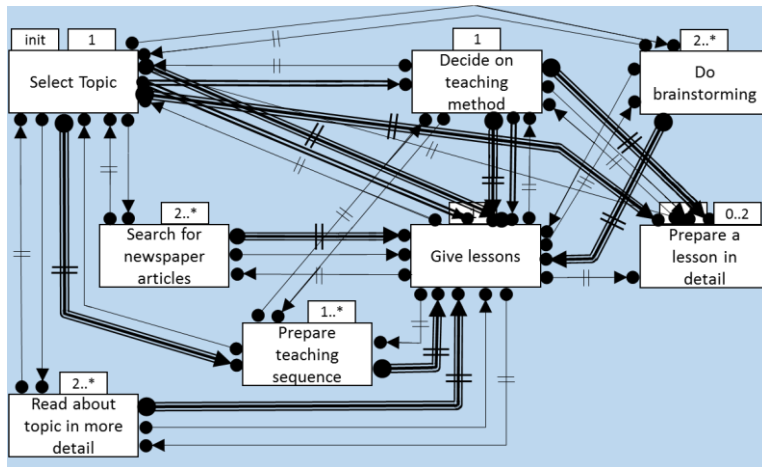
# Case Study – Hierarchical Model Discovery



Declare Model from the event log with complex activities.  
(Hierarchical model)

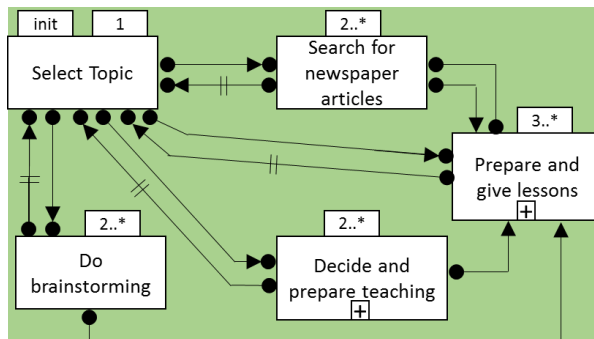
# Case Study – Complexity Reduction Evaluation

## Flat



	Flat	Hierarchical
<b>No. of Activities</b>	8	5
<b>No. of Constraints</b>	45	18
<b>No. of Different Constraints</b>	9	8
<b>No. of Complex Activities</b>	0	2
<b>Constraint/Activity Ratio</b>	5.63	3.60

## Hierarchical



La Rosa, et al.: Managing Process Model Complexity Via Abstract Syntax Modifications. IEEE Transactions Industrial Informatics 7, 614–629 (2011)

# Contributions

- A novel problem formulation.
- A technical solution addressing the problem:
  - 2 Algorithms to suggest complex activities, in a user guided fashion.

# Limitations

- The method is dependent on the labeling quality.
- Declare is not stuttering invariant and the exact ordering on some events matter.
- There exists information loss, caused by the preprocessing method.

# Future Work

- Address the previous limitations.
- Discussion about quality dimensions measures: fitness, precision, simplicity and generalizability.
- Analysis of real life event logs.
- Point out where the method performs better, regarding the labeling styles and the losses caused by the aggregation.

תודה  
Dankie Gracias  
Спасибо شكراً  
Merci Takk  
Köszönjük Terima kasih  
Grazie Dziękujemy Děkojame  
Ďakujeme Vielen Dank Paldies  
Kiitos Täname teid 谢谢  
**Thank You** Tak  
感謝您 Obrigado Teşekkür Ederiz  
감사합니다  
Σας ευχαριστούμε ඔබට  
Bedankt Děkujeme vám  
ありがとうございます  
Tack



✉ [pedro.richetti@uniriotec.br](mailto:pedro.richetti@uniriotec.br)



Universidade Federal do Estado do Rio de Janeiro

