

Seminar Machine Learning Applications in Process Mining

**Assessment of  
*Discovering Causal Factors*  
*Explaining Business Process Performance*  
*Variation***

Tom-Hendrik Hülsmann

RWTH Aachen University  
tom.huelsmann@rwth-aachen.de

# Agenda

- Introduction
- Method
- Example Application
- Conclusion

# Introduction

# Discovering Causal Factors Explaining Business Process Performance Variation <sup>[1]</sup>

- Bart F.A Hompes et al.
- 2017 CAiSE

## Discovering Causal Factors Explaining Business Process Performance Variation

Bart F.A. Hompes<sup>1,2(✉)</sup>, Abderrahmane Maaradji<sup>3</sup>, Marcello La Rosa<sup>3</sup>,  
Marlon Dumas<sup>4</sup>, Joos C.A.M. Buijs<sup>1</sup>, and Wil M.P. van der Aalst<sup>1</sup>

<sup>1</sup> Eindhoven University of Technology, Eindhoven, The Netherlands  
{b.f.a.hompes,j.c.a.m.buijs,w.m.p.v.d.aalst}@tue.nl

<sup>2</sup> Philips Research, Eindhoven, The Netherlands

<sup>3</sup> Queensland University of Technology, Brisbane, Australia  
{abderrahmane.maaradji,m.larosa}@qut.edu.au

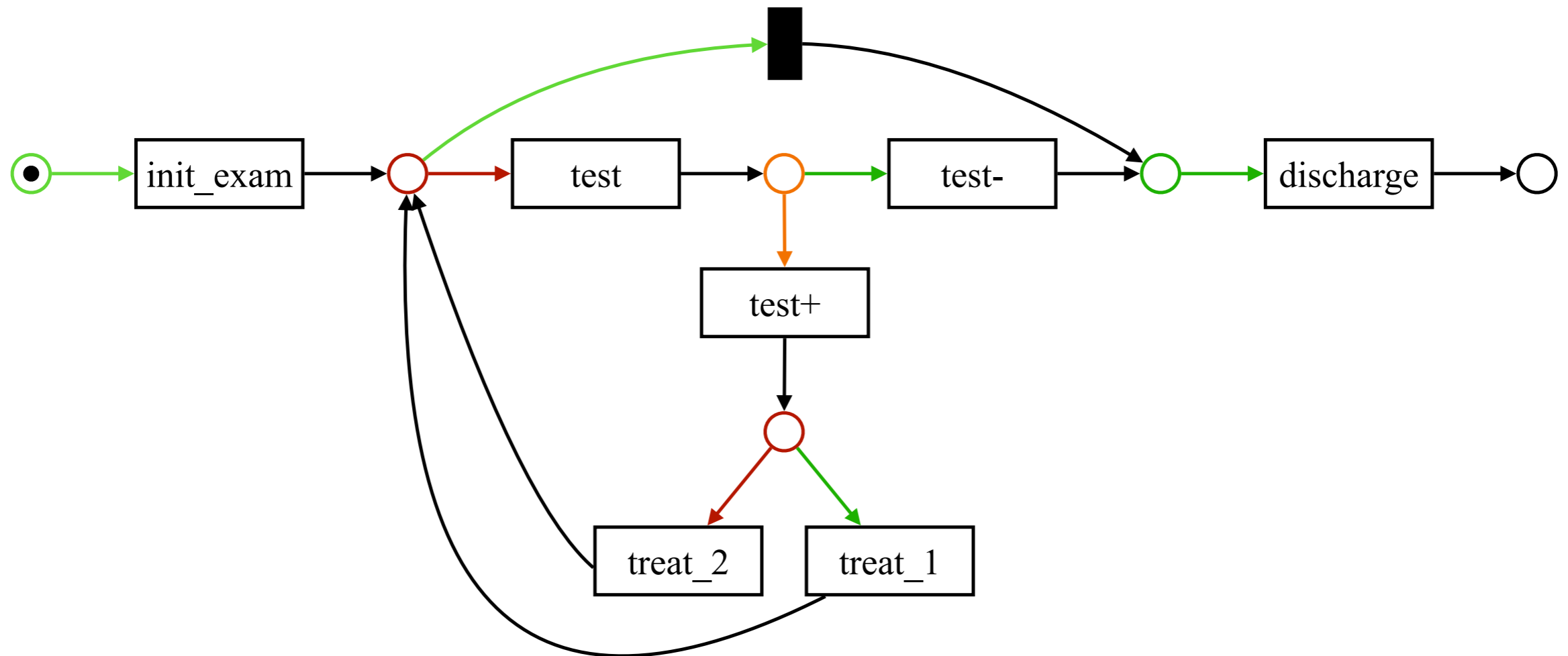
<sup>4</sup> University of Tartu, Tartu, Estonia  
marlon.dumas@ut.ee

**Abstract.** Business process performance may be affected by a range of factors, such as the volume and characteristics of ongoing cases or the performance and availability of individual resources. Event logs collected by modern information systems provide a wealth of data about the execution of business processes. However, extracting root causes for performance issues from these event logs is a major challenge. Processes may

# Business Process Performance

- Measures a qualitative aspect of the process
- Most often time based
  - Case Duration
  - Activity Duration
  - Waiting time
- Discovery of bottlenecks

# Business Process Performance



# Business Process Performance

- Improve processes based on findings
- Decrease execution times
- Decrease costs
- Increase Customer Satisfaction

# Causal Factors

- Not *where* is the problem but *what* causes the problem
- More detailed than bottleneck analysis
- Allows statements like:
  - *"The involvement of Resource A has a causal effect on the performance of the Test activity"*



# Causal Factors

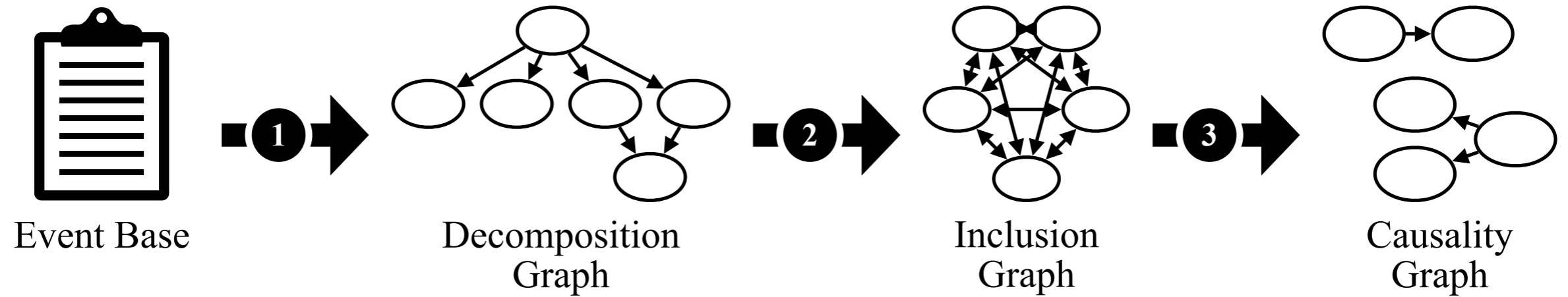
- Novel approach to process performance analysis
- Not a big research focus compared to
  - Performance prediction
  - Replay based bottleneck detection

# Related Work

- Performance studies applied to real processes
  - Software Development [2], Container Shipping [3], Hierarchical Process Models [4]
- Replay based on alignments
  - van der Aalst et al. (2012) [5]
- Prediction of running cases
  - van der Aalst et al. (2011) [6]
  - Making prediction results more transparent by Verenich Ily et al. (2017) [7]

# Method

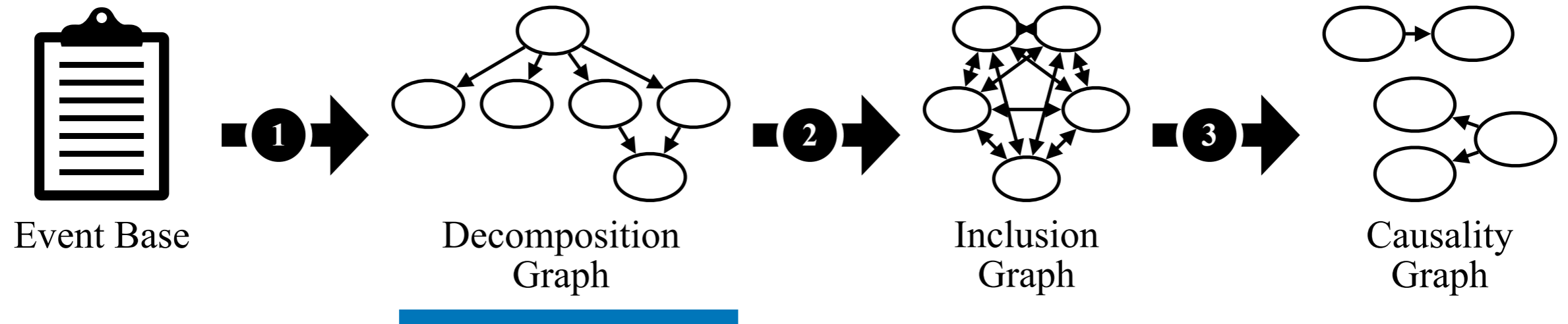
# Method



# Event Base

- Based on the event log
- Different formal representation
- Set of properties  $P$ 
  - e.g.  $P = \{activity, resource, \dots\}$
- Set of event identifiers  $E$ 
  - e.g.  $E = \{(test, 12A, 2020-07-24\ 12:00), \dots\}$
- Family of functions  $\pi_P: E \rightarrow V_P$ 
  - e.g.  $\pi_{resource}(test, 12A, 2020-07-24\ 12:00) = "Peter"$

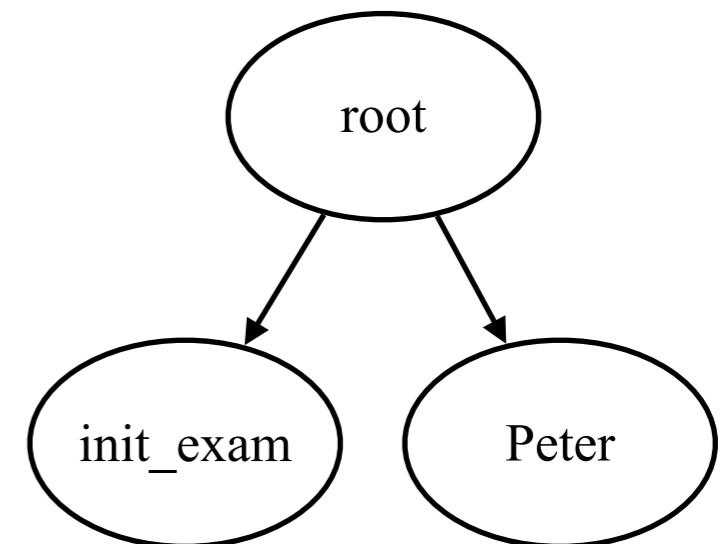
# Method



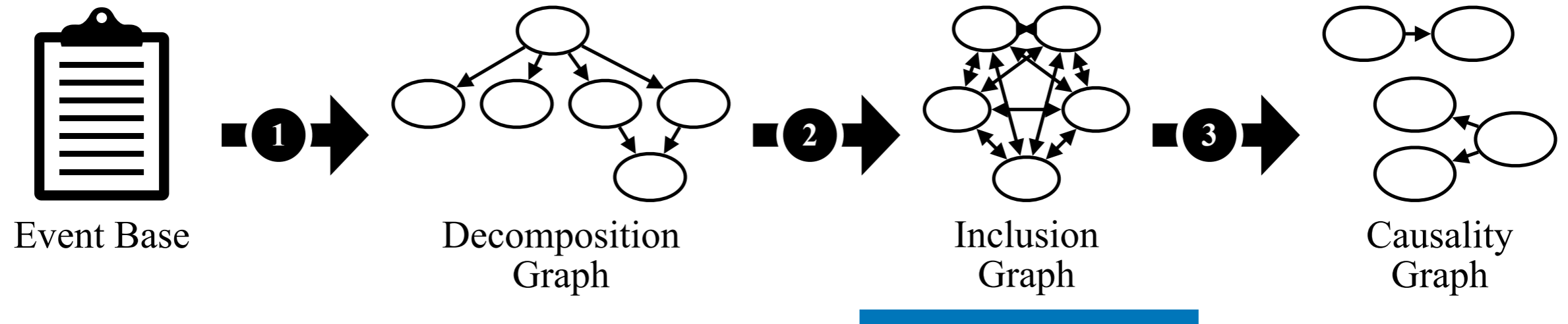
# Decomposition Graph

1

- Partitions events based on common properties
- Hierarchical structure from top to bottom
- Events in deeper nodes are more similar



# Method

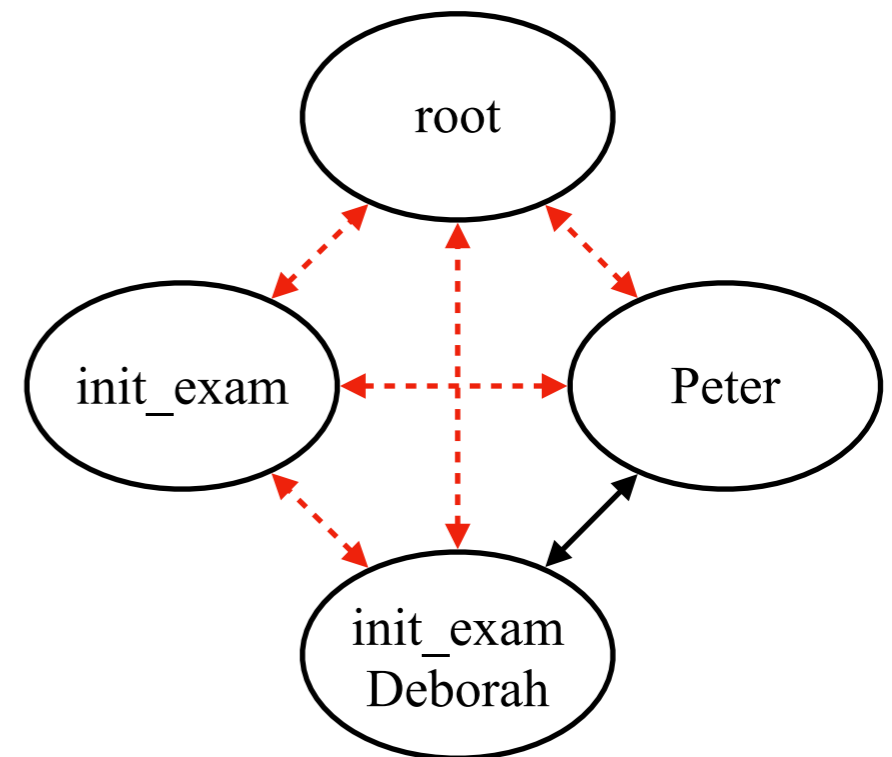




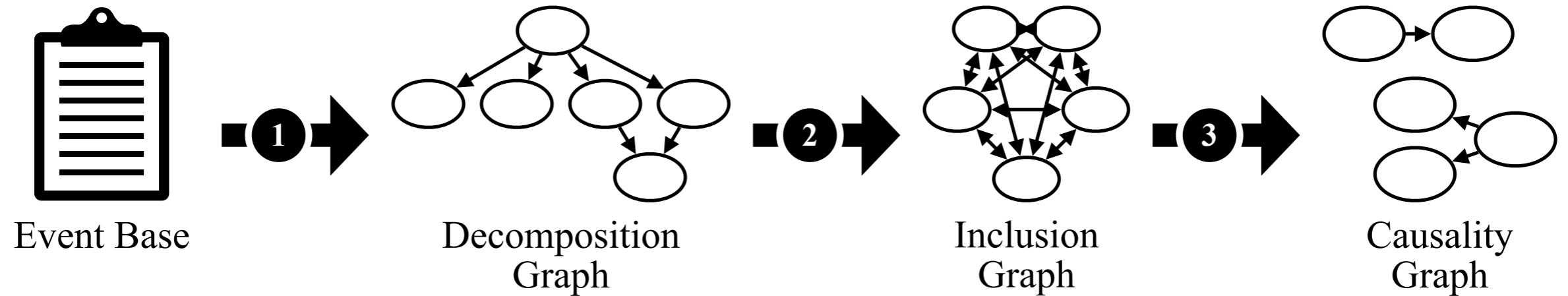
# Inclusion Graph

2

- Based on Decomposition graph
- Start with fully connected graph
- Prune based on Decomposition graph
- Contains candidate causal relations

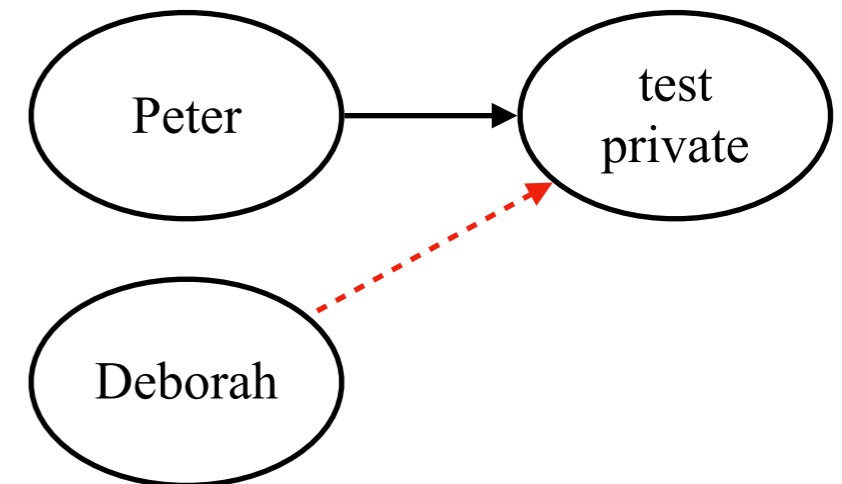


# Method



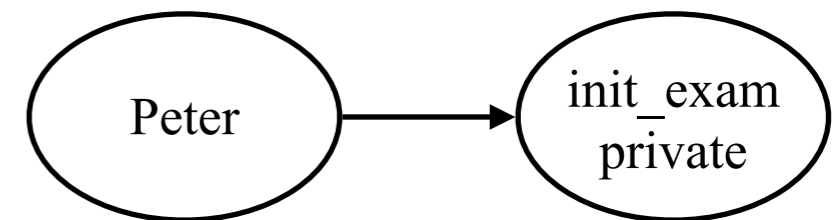
# Causality Test

- **Granger Causality Test** [8] is performed for each edge of the Inclusion graph
- Statistical test to find causal relations in time series data
- Edges with negative test are discarded



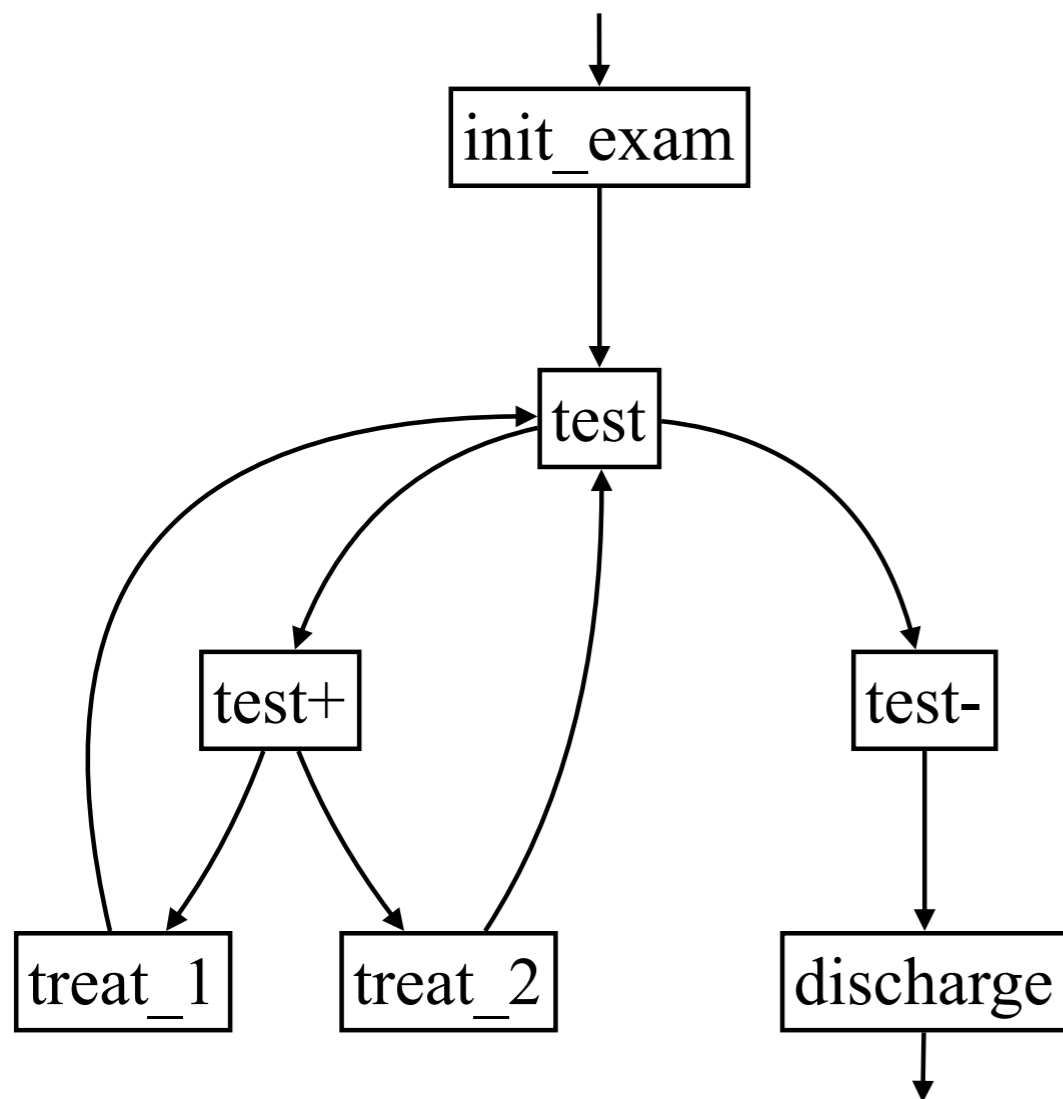
# Causality Graph

- Connected nodes represent causal relations
- *"The performance of events performed by resource **Peter** causes the performance of **init\_exam** events related to **private** insurance patients"*



# Example Application

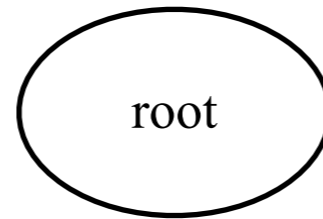
# Example Process



patient_id	activity	timestamp	resource	insurance
1	init_exam	13:12	Peter	State
2	test-	13:24	lab	Private
1	test	13:25	Susan	State
3	tes-	13:33	Deborah	Private
...	...	...	...	...

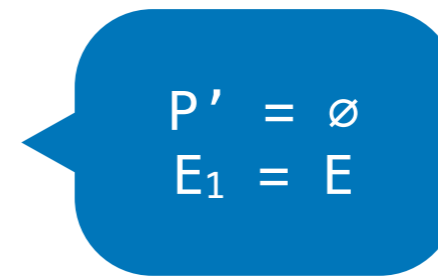
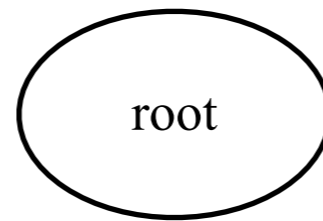
# Decomposition Graph

1



# Decomposition Graph

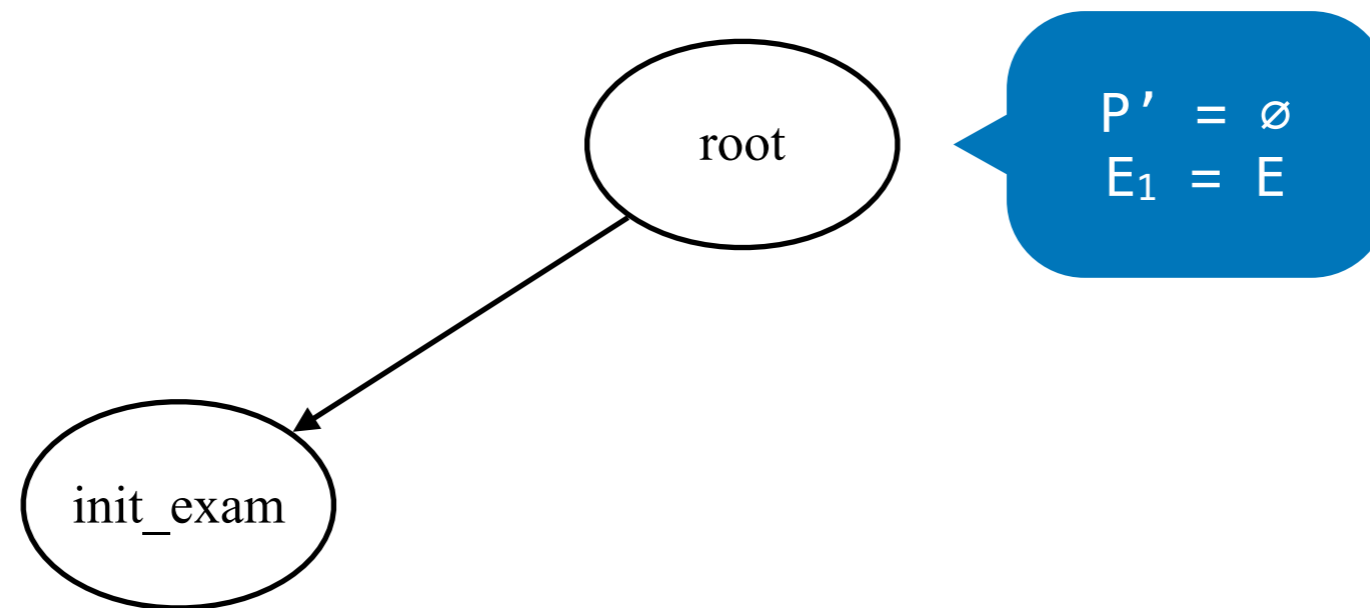
1





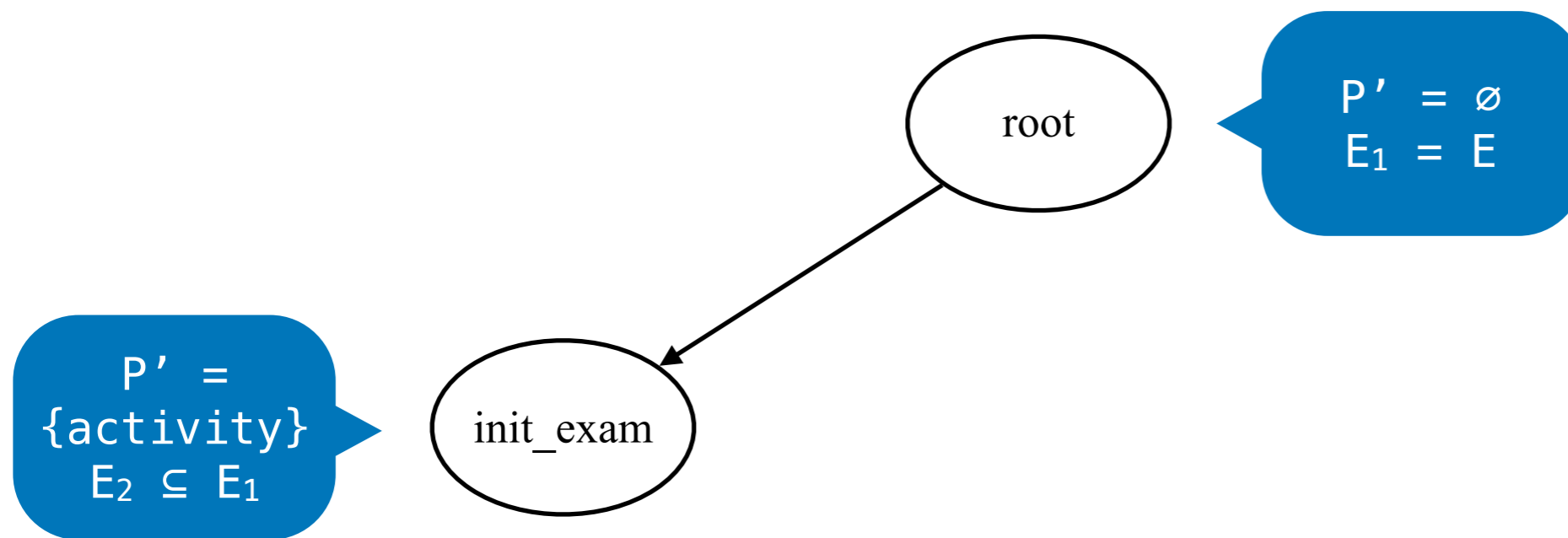
# Decomposition Graph

1



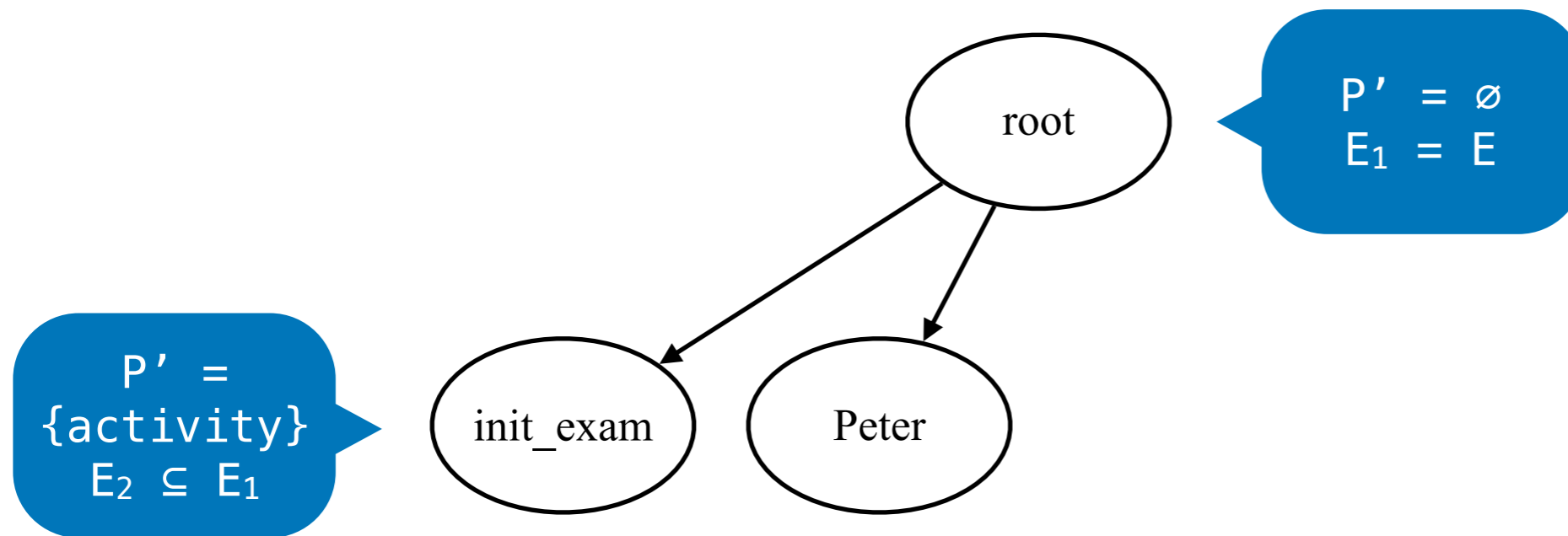
# Decomposition Graph

1



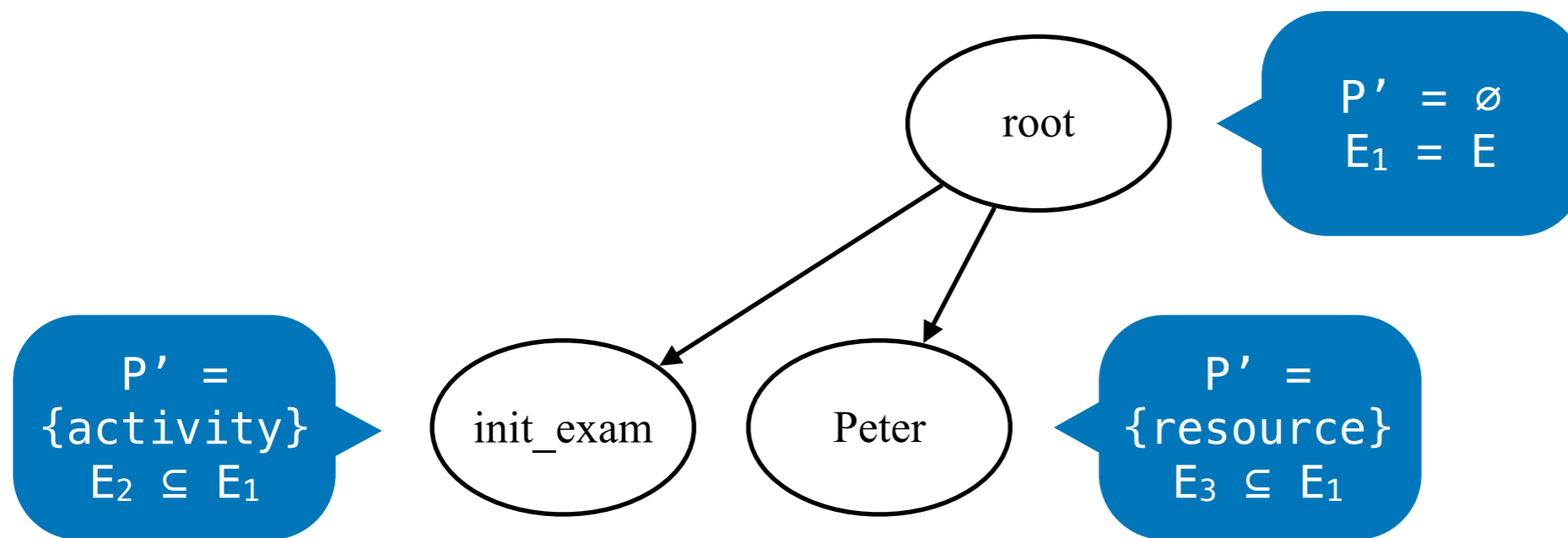
# Decomposition Graph

1



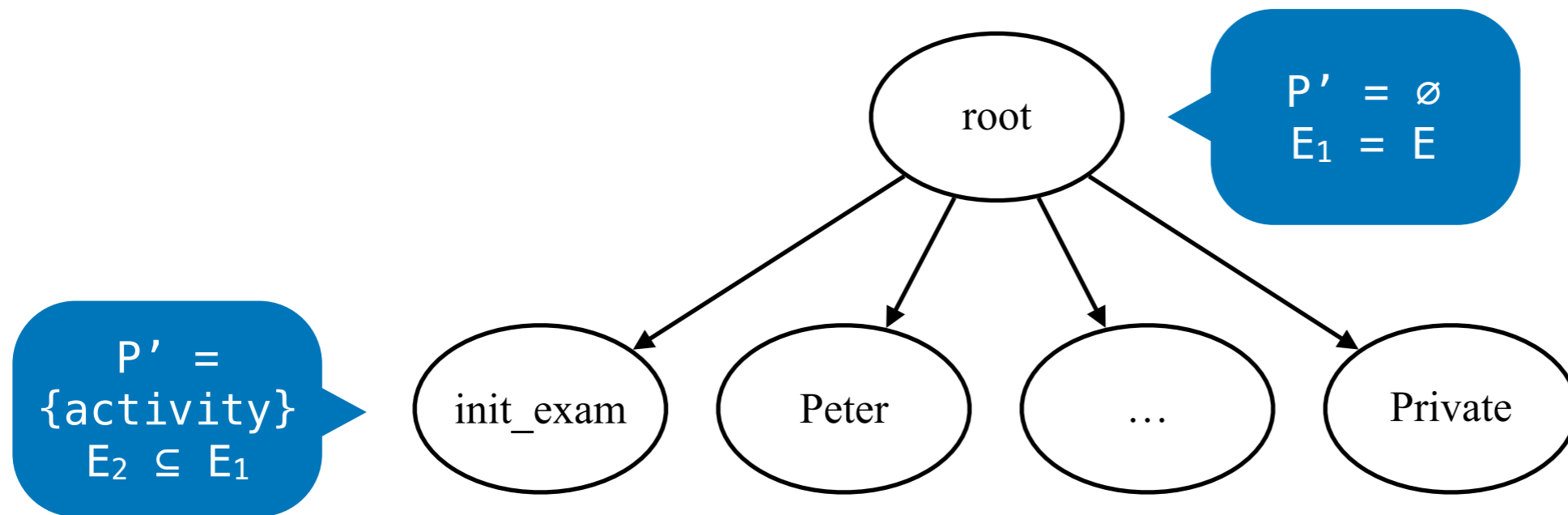
# Decomposition Graph

1



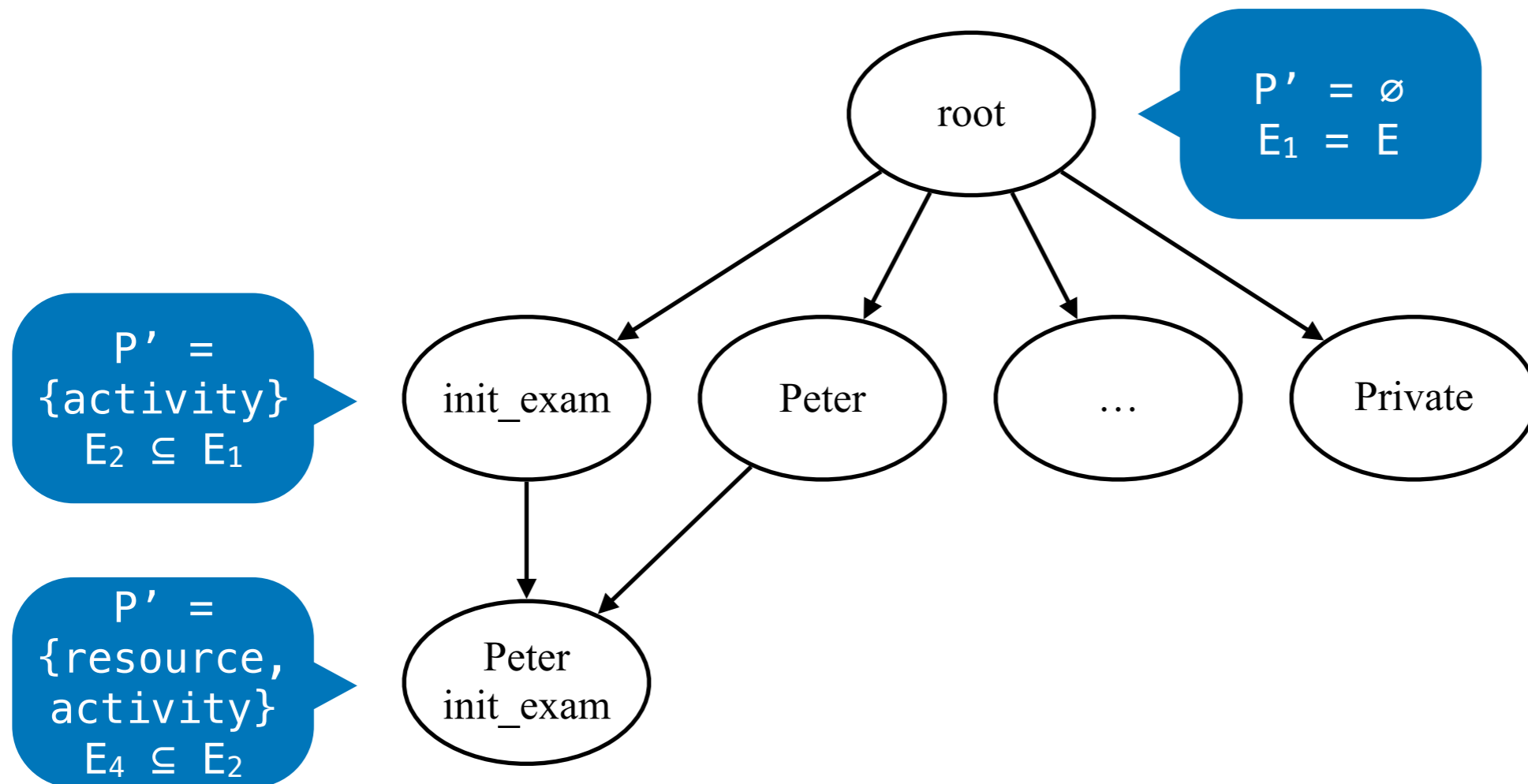
# Decomposition Graph

1



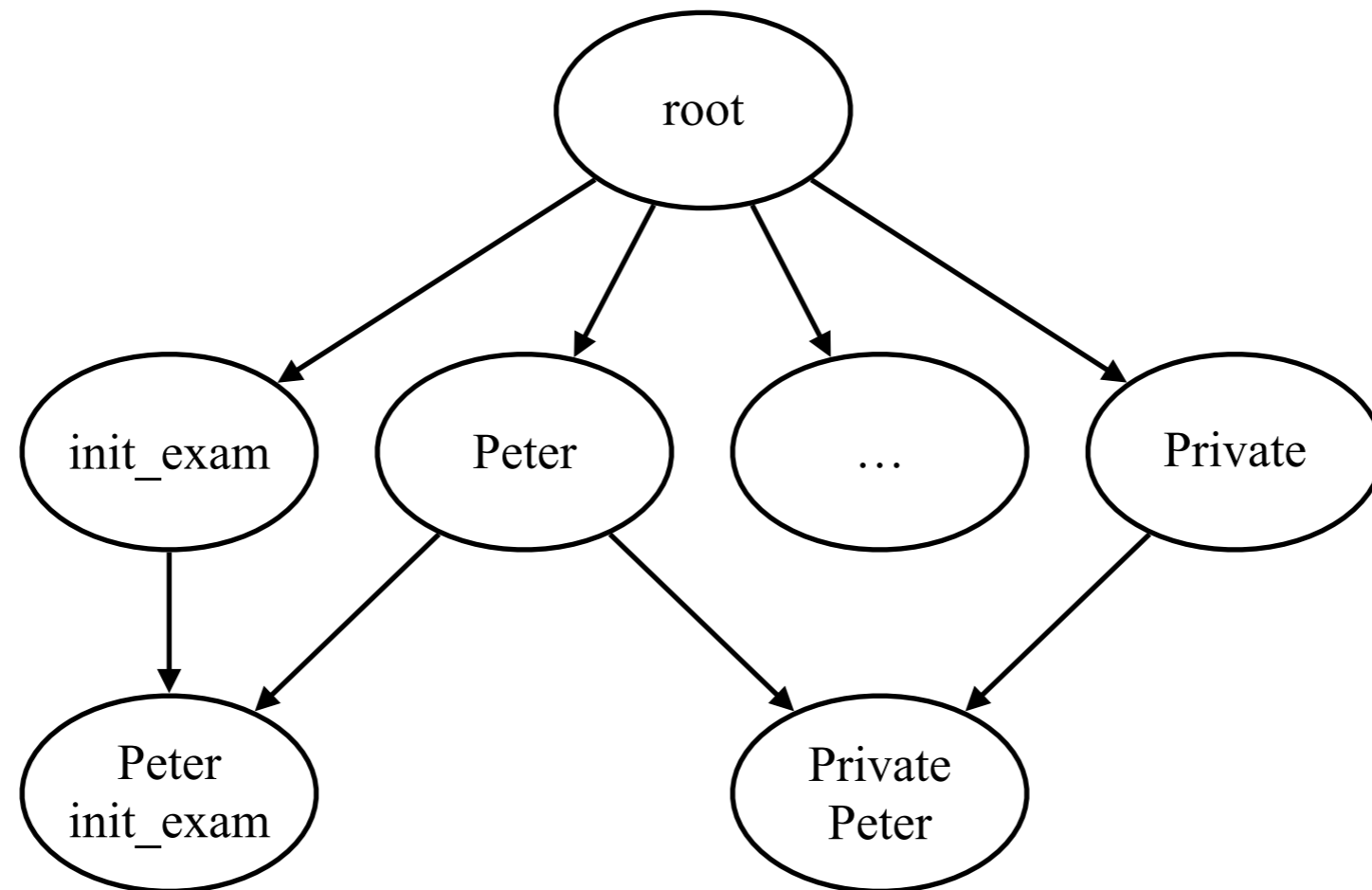
# Decomposition Graph

1



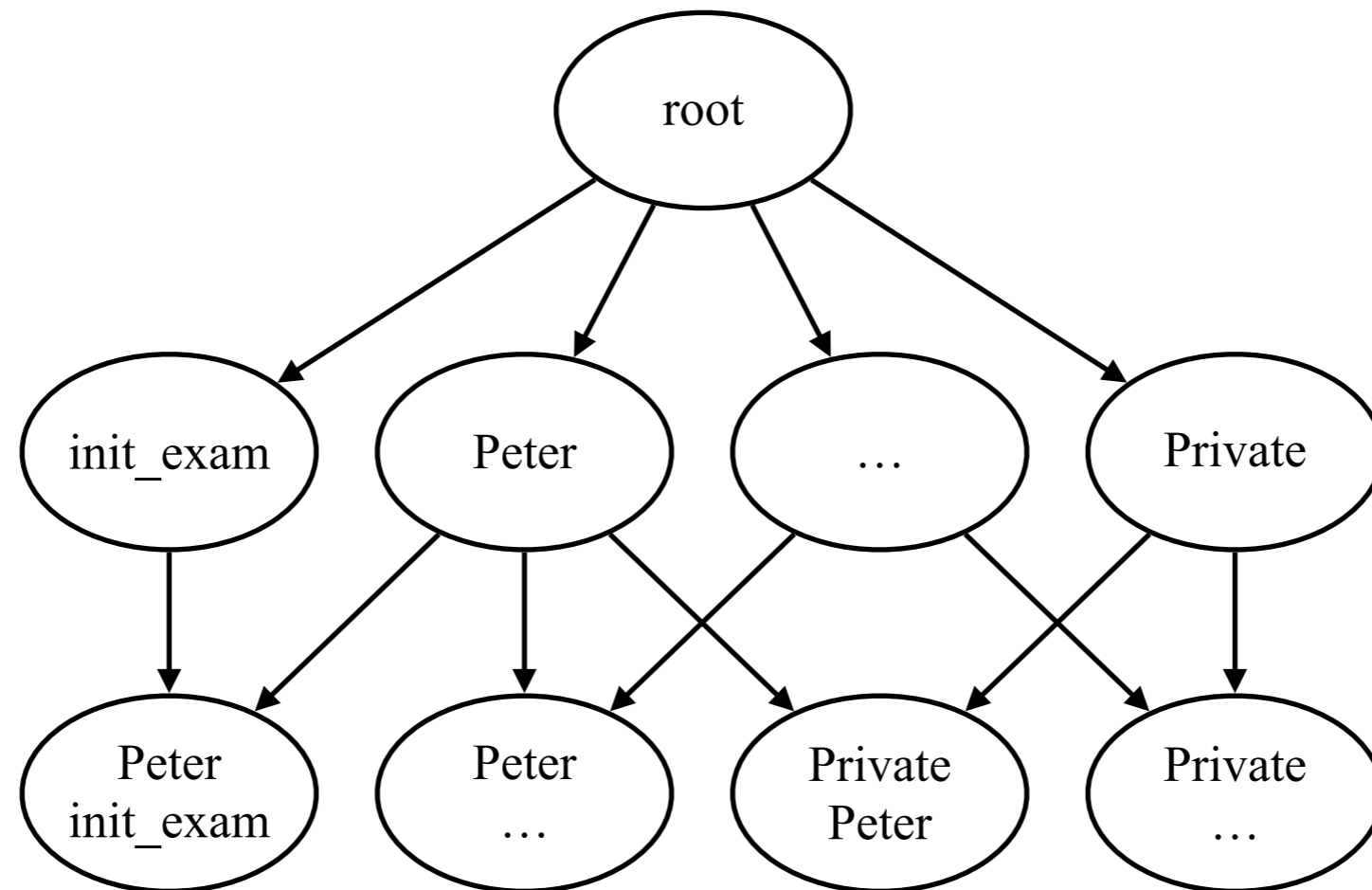
# Decomposition Graph

1



# Decomposition Graph

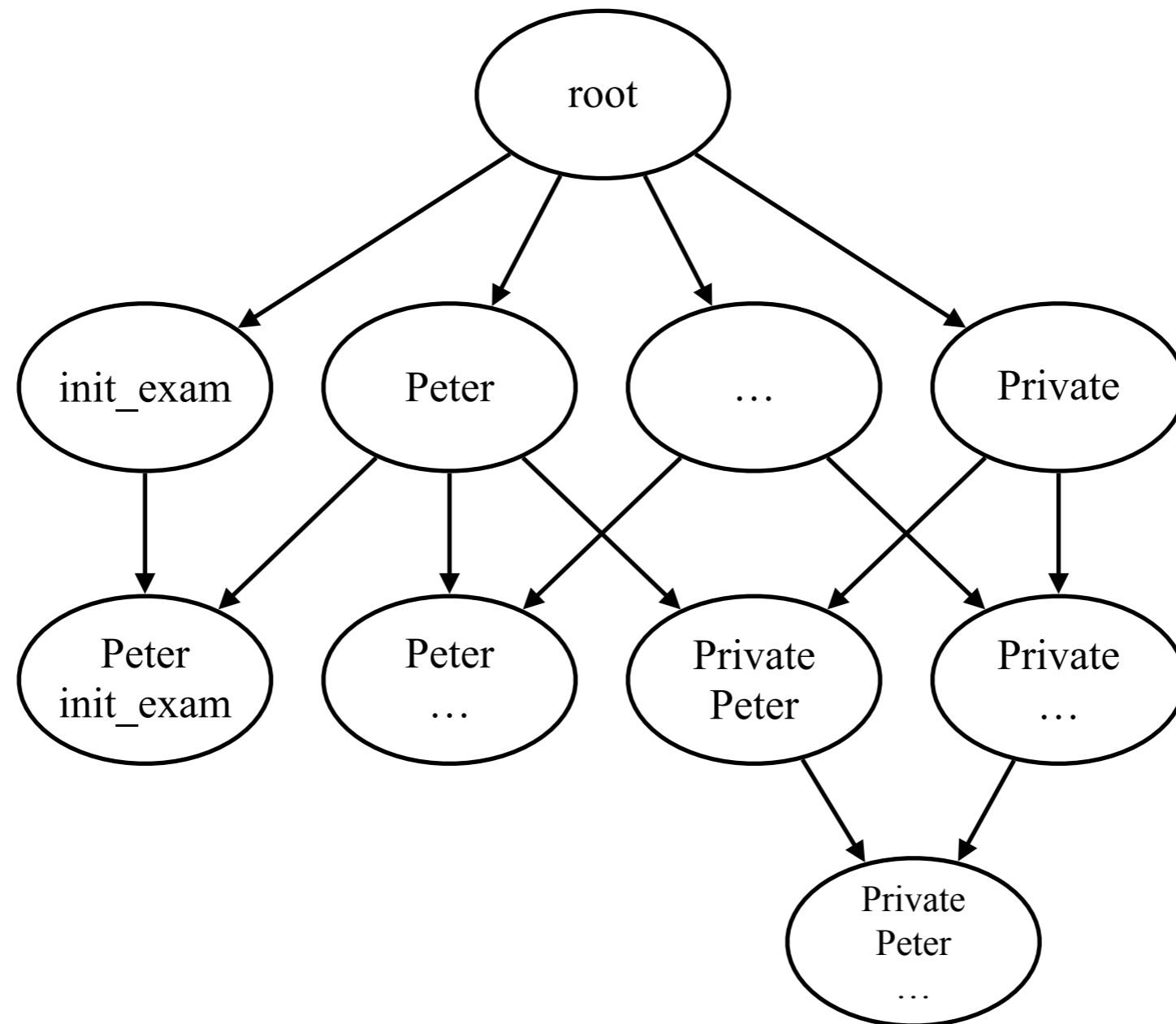
1



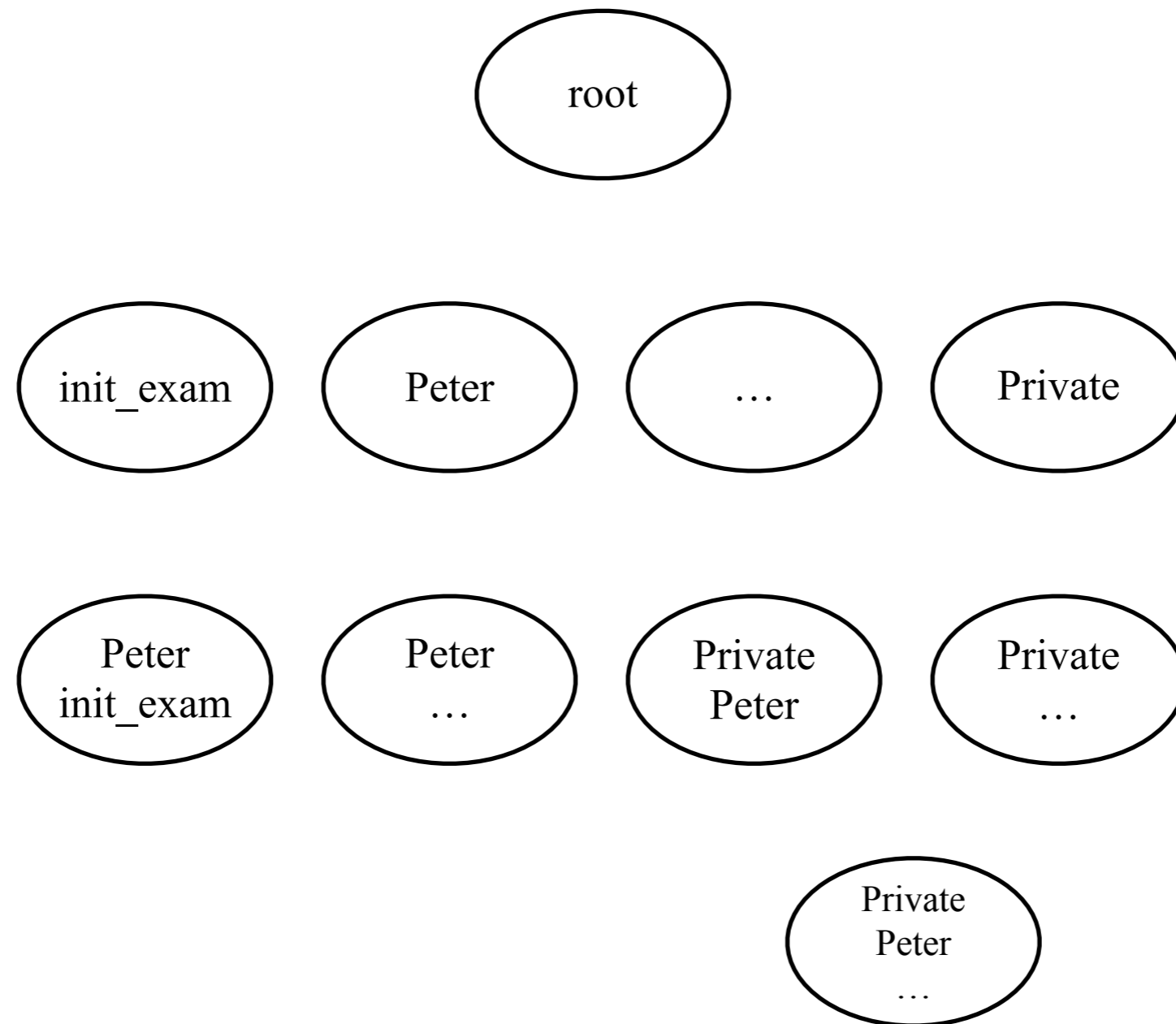


# Decomposition Graph

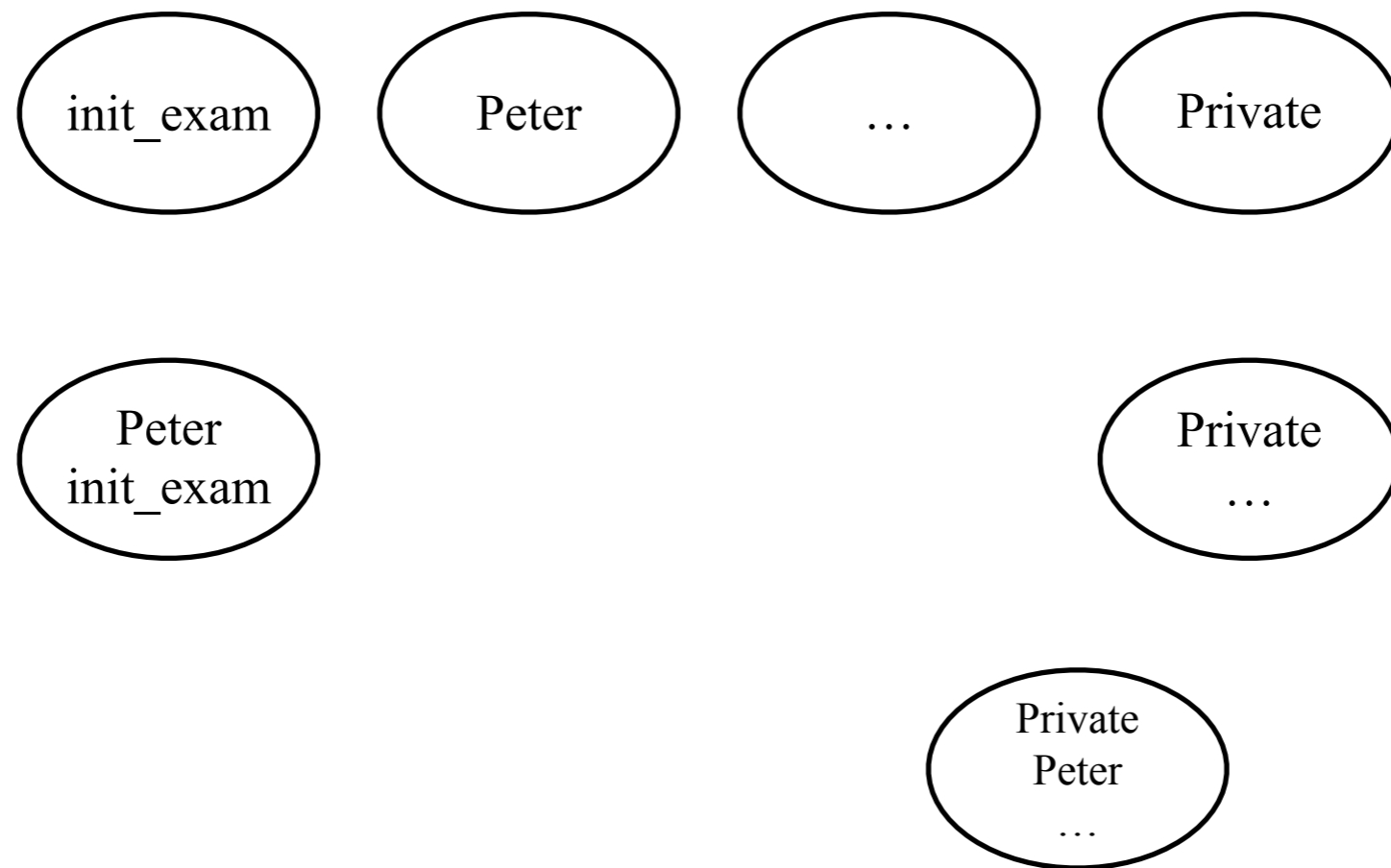
1



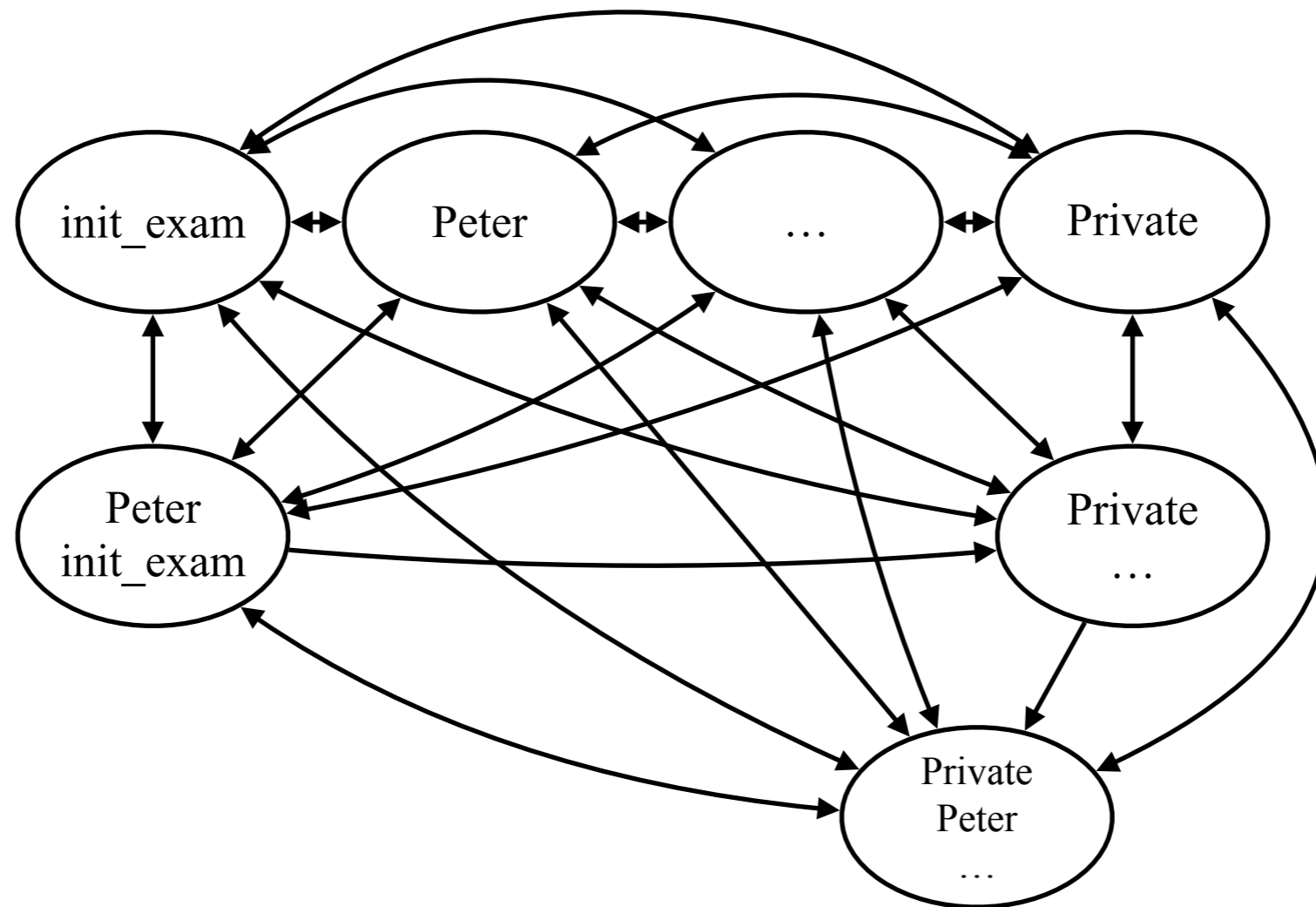
# Inclusion Graph



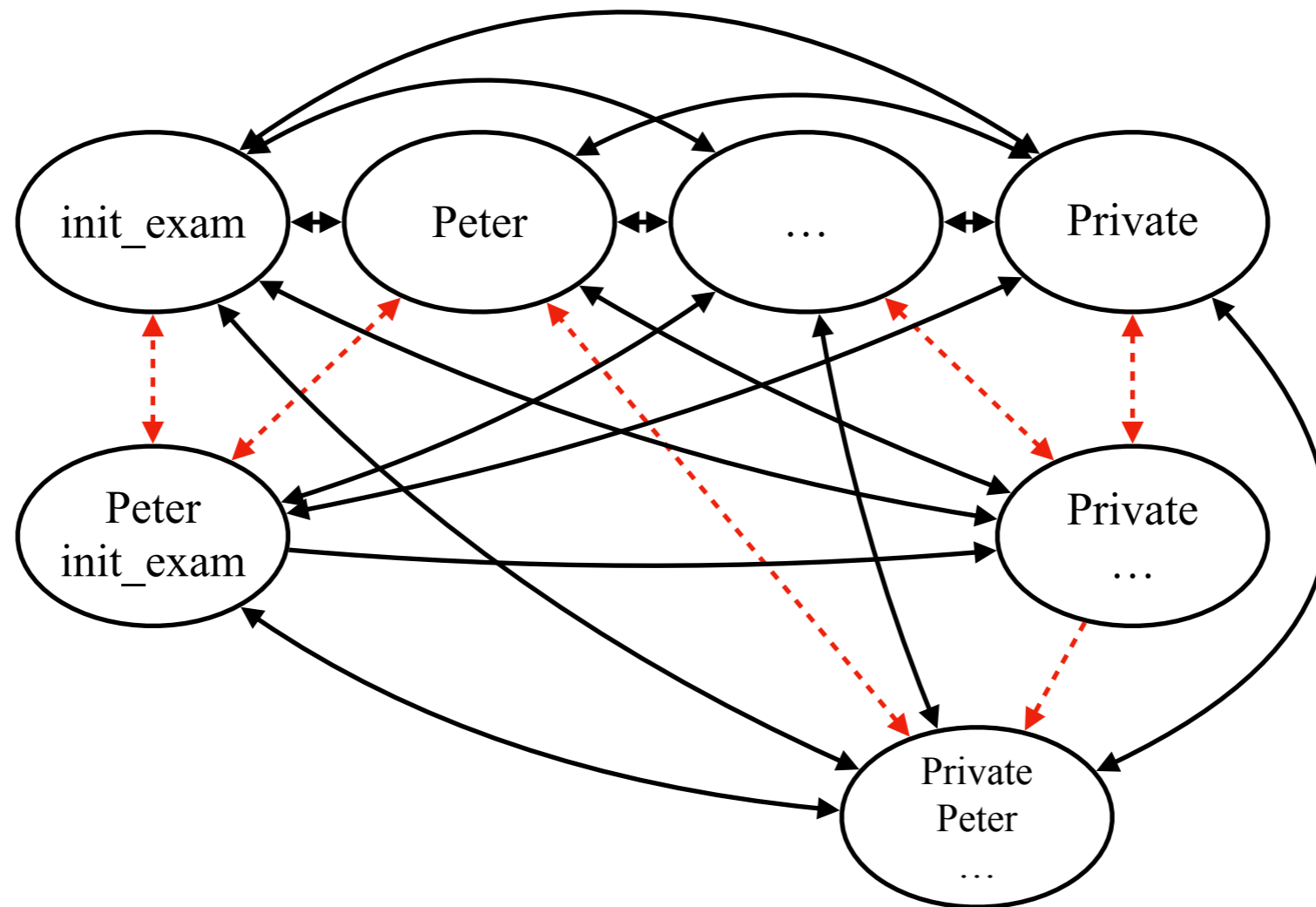
# Inclusion Graph



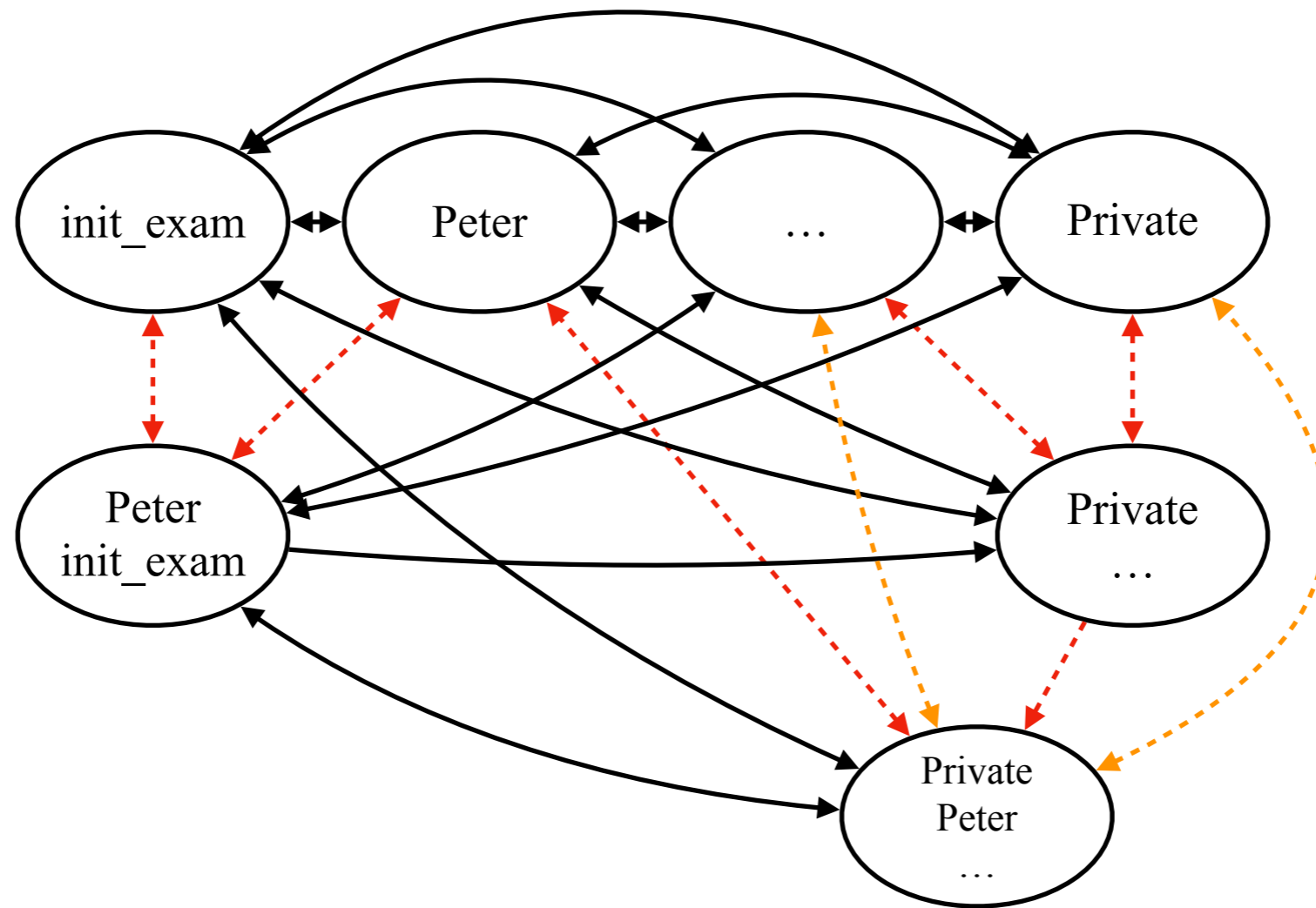
# Inclusion Graph



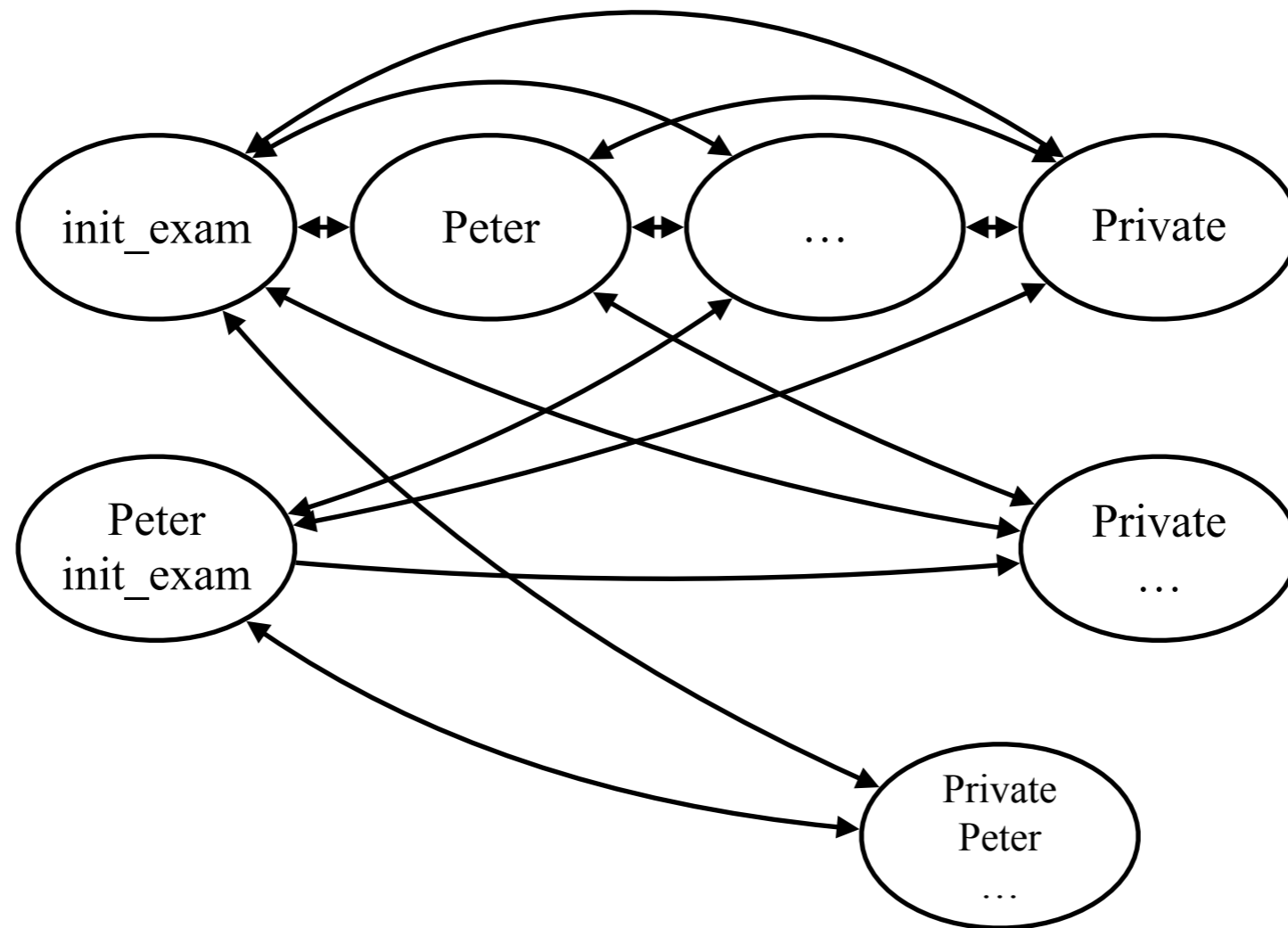
# Inclusion Graph



# Inclusion Graph

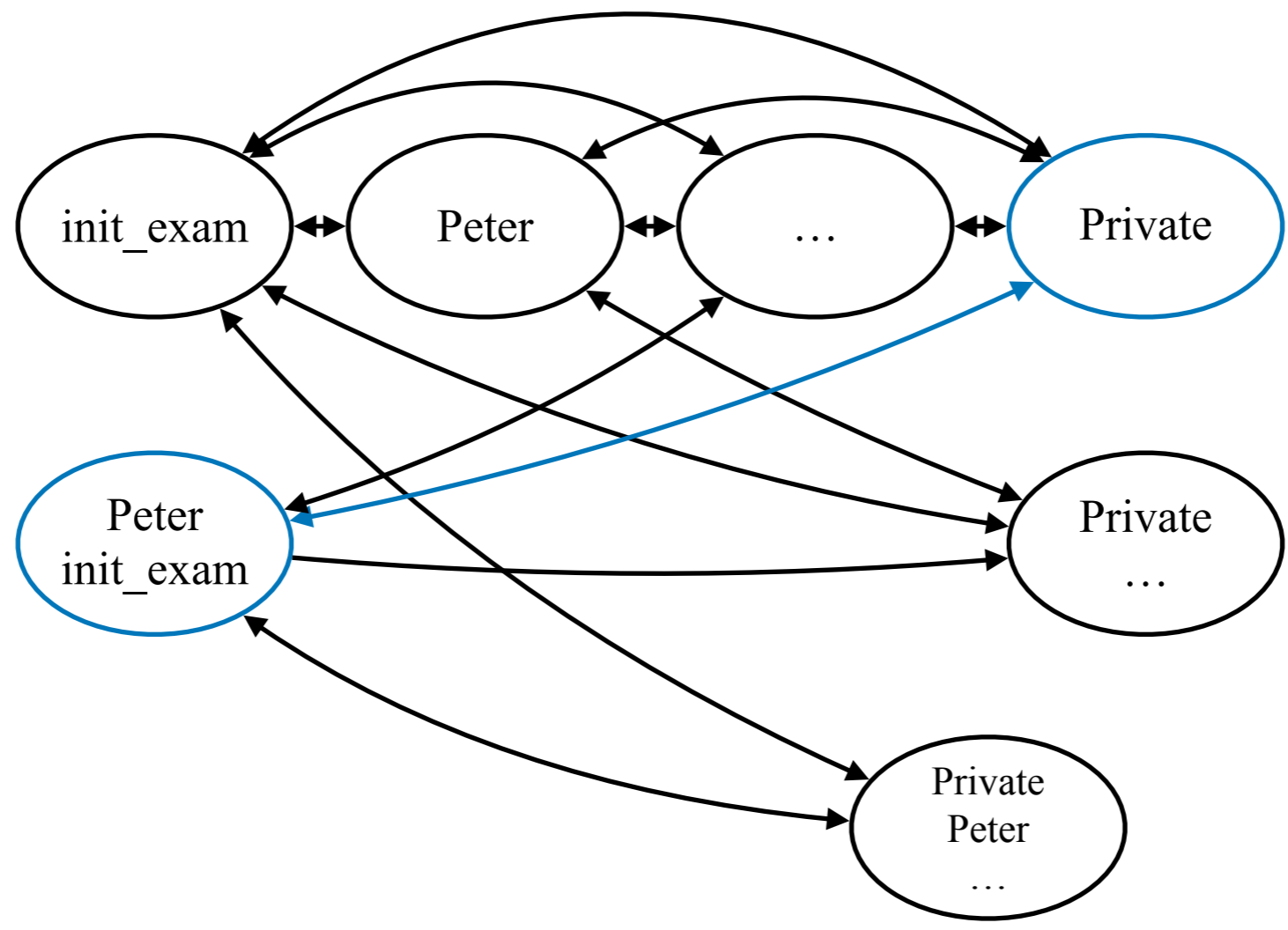


# Inclusion Graph



**in this case 38%  
reduction of edges**

# Causality Test

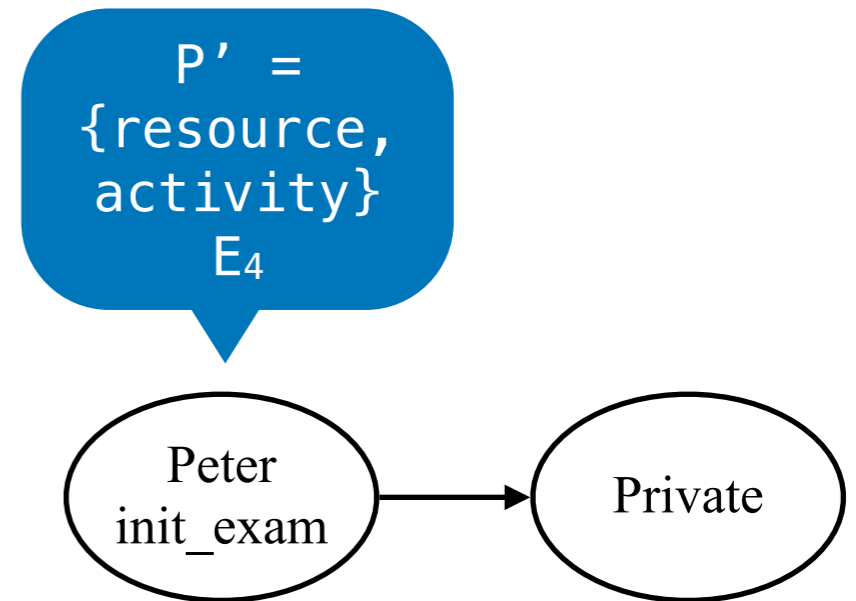




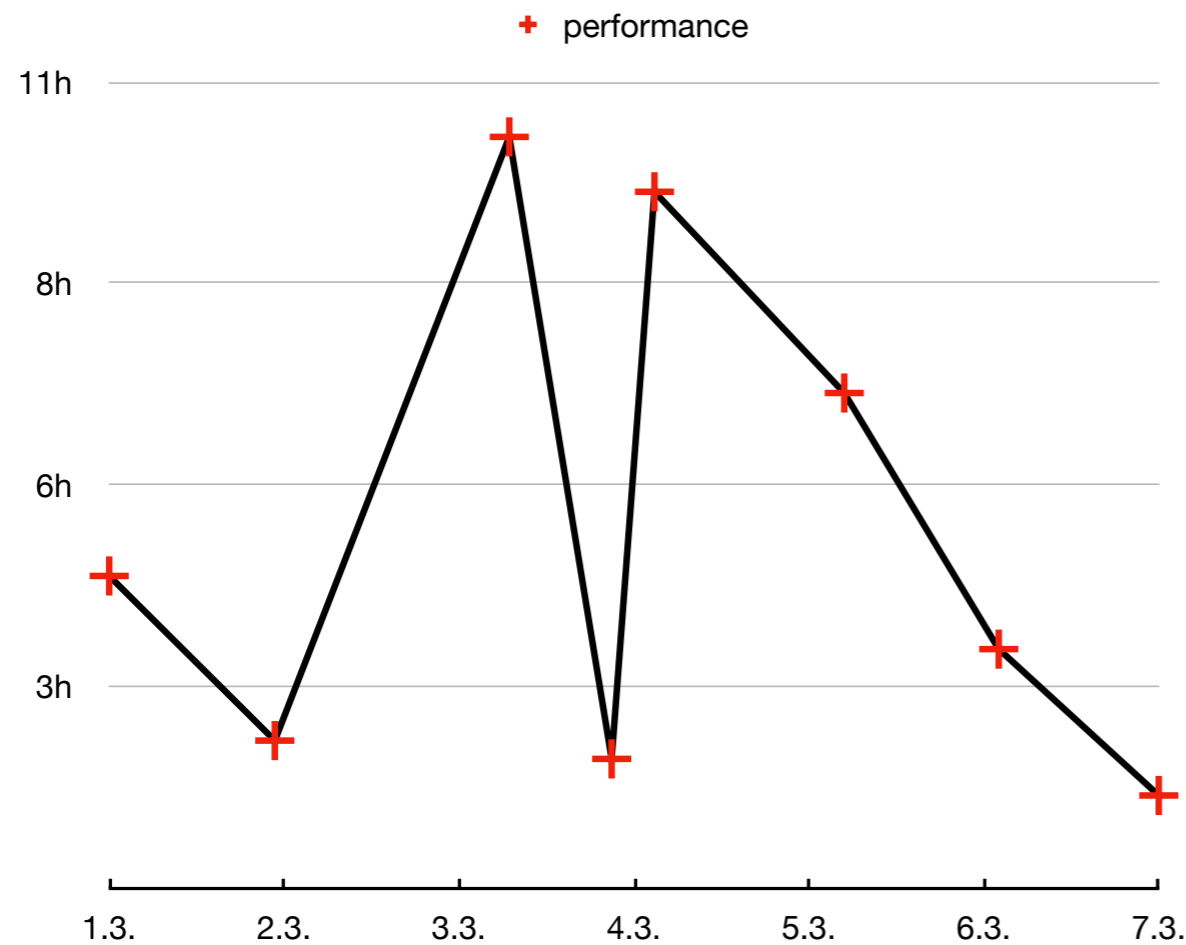
# Time Series

E<sub>4</sub>

patient_id	activity	timestamp	resource	performance
5	init_exam	20-03-01 18:00	Peter	5
3	init_exam	20-03-02 14:30	Peter	2.6
7	init_exam	20-03-03 11:15	Peter	9
2	init_exam	20-03-04 17:45	Peter	2.5
...	...	...	...	...



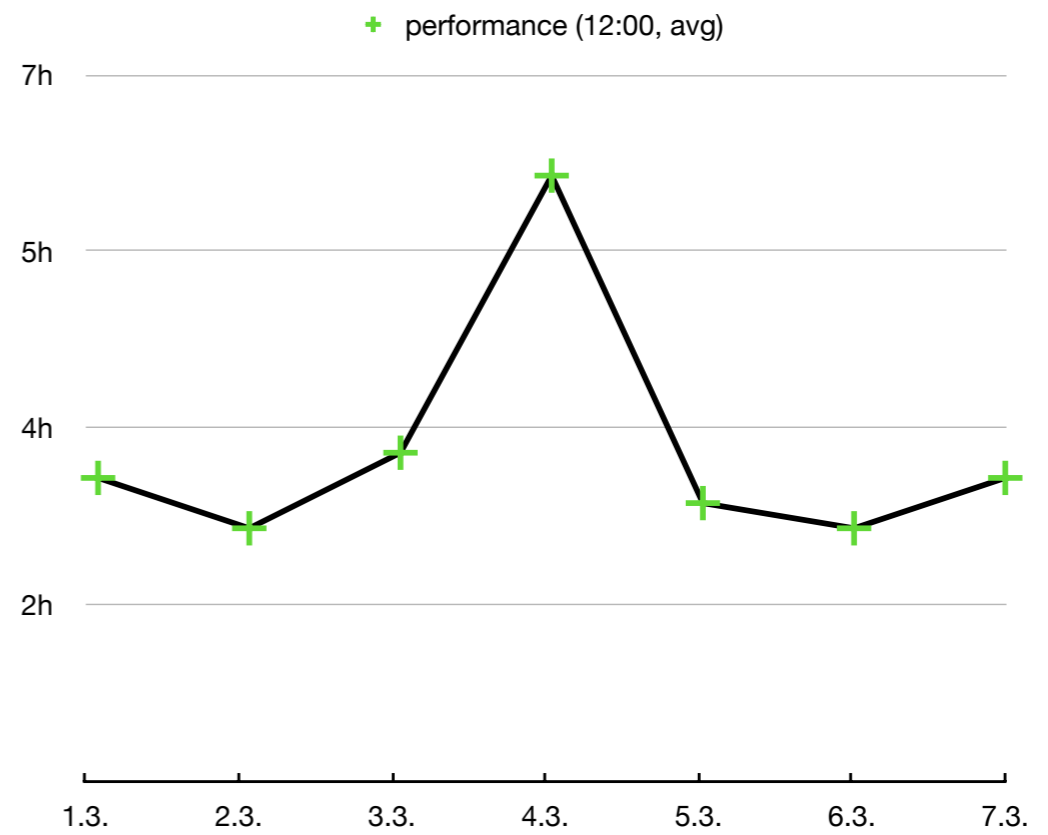
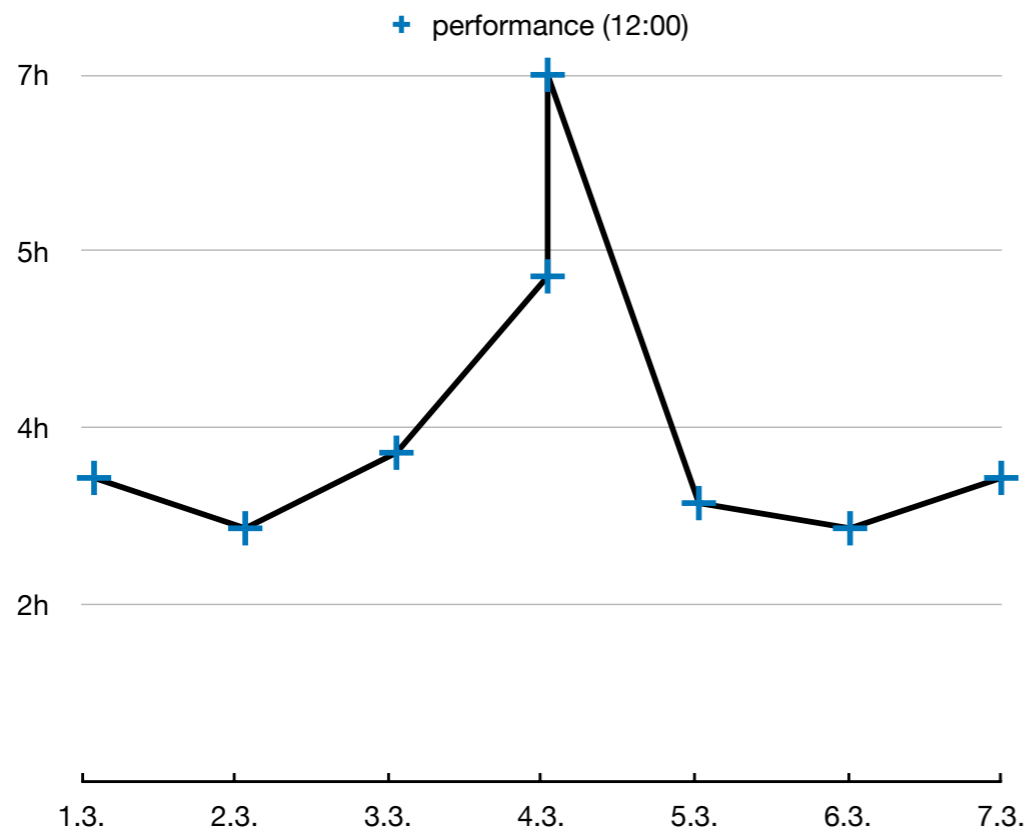
# Time Series



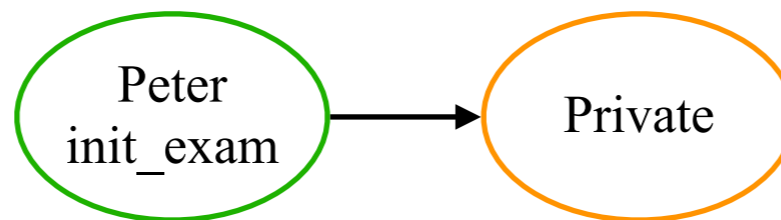
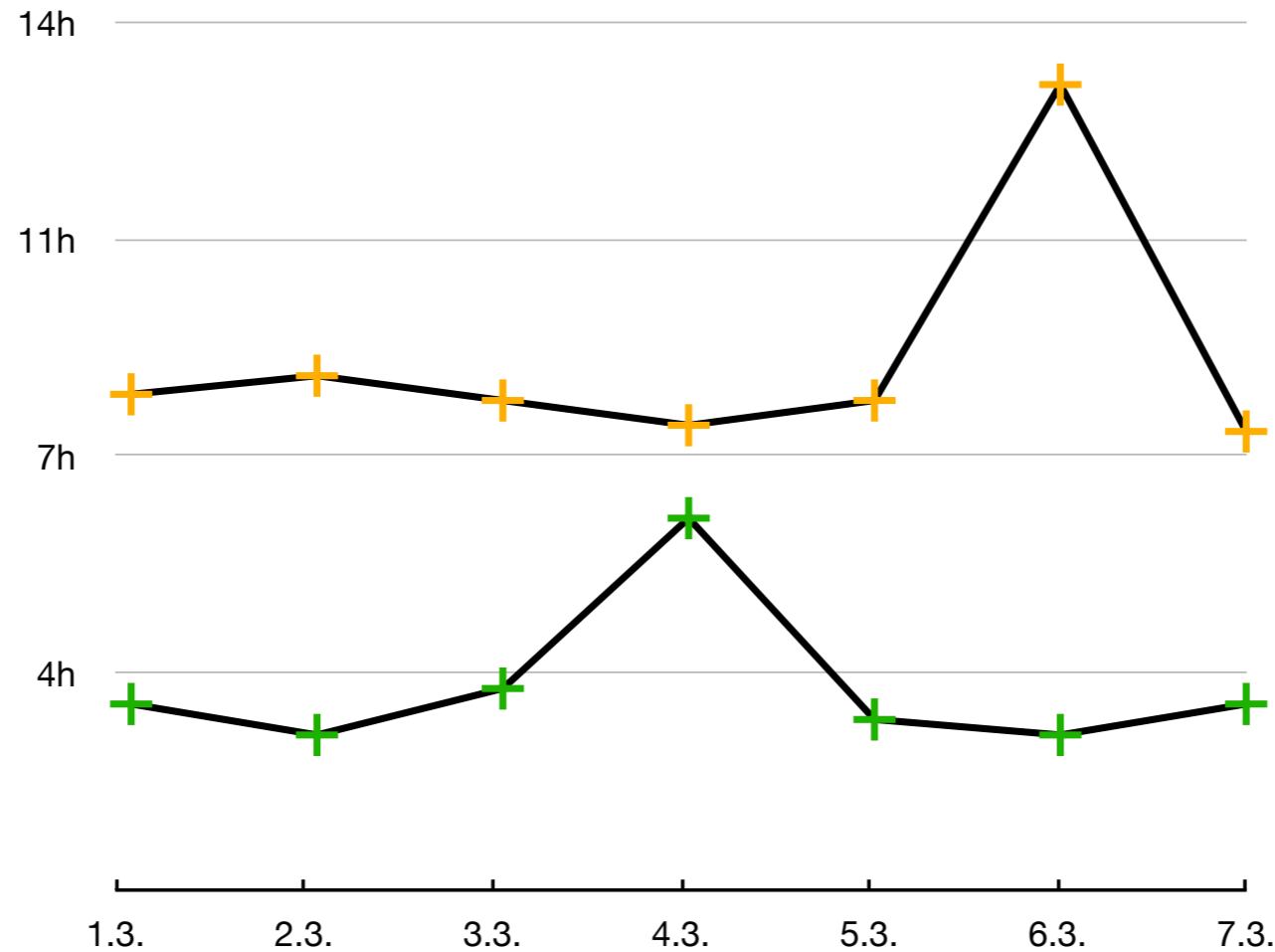
E4

patient_id	activity	timestamp	resource	performance
5	init_exam	20-03-01 18:00	Peter	5
3	init_exam	20-03-02 14:30	Peter	2.6
7	init_exam	20-03-03 11:15	Peter	9
2	init_exam	20-03-04 17:45	Peter	2.5
...	...	...	...	...

# Time Series

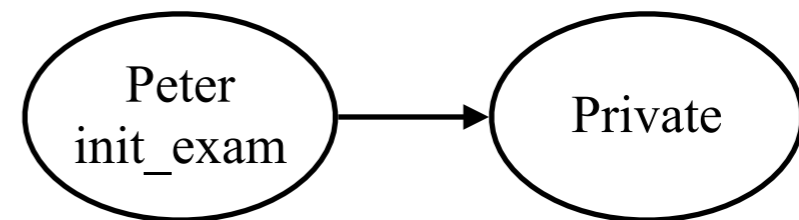


# Causality Test



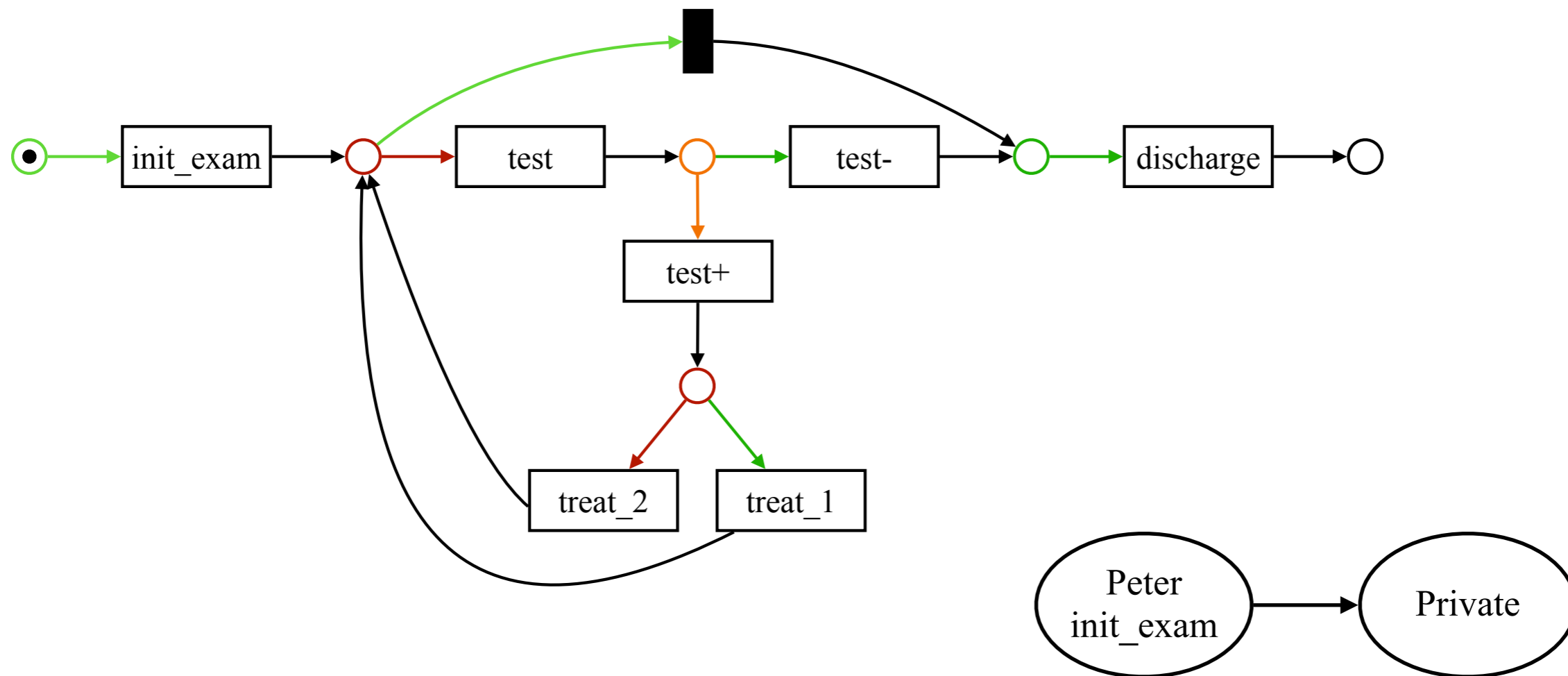
# Causality Graph

- *"The performance of initial exams performed by Peter causes the performance of events related to privately insured patients."*
- Type of relation is not known
  - Poor  $\Rightarrow$  Poor
  - Good  $\Rightarrow$  Good



# Conclusion

# Results



*"There is a performance problem with the test activity."*

*"When Peter performs the initial exam, the performance of cases of private patients decreases."*

# Conclusion

- Higher amount of information compared to traditional approaches
- Results are more actionable
- Less expert knowledge required



# Conclusion

- Not all kinds of causal relations can be found
- Granger Causality Test is controversial [9]
- Further pruning of the Inclusion graph could improve runtime
- Statistical approach to Causal Factors, others are possible

# Questions

# References

**[1]** Bart F.A. Hompes, Abderrahmane Maaradji, Marcello La Rosa, Marlon Dumas, Joos C.A.M. Buijs, Wil M.P. van der Aalst: Discovering Causal Factors Explaining Business Process Performance Variation. International Conference on Advanced Information Systems Engineering, 177-192 (2017)

**[2]** A. H. M. Shani, R. Sarno, K. R. Sungkono and C. S. Wahyuni: Time Performance Evaluation of Agile Software Development. 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 202-207 (2019)

**[3]** R. Rahardianto, R. Sarno and G. Intani Budiawati: Performance Time Evaluation of Domestic Container Terminal Using Process Mining and PERT. 2018 International Seminar on Application for Technology of Information and Communication, pp. 469-475 (2018)

**[4]** A. N. Rofiif, M. M. Wildan, K. R. Sungkono, R. Sarno and C. S. Wahyuni: Performance Analysis of Hierarchical Process Model. 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 171-177 (2019)

...

# References

- [5] van der Aalst, Wil and Adriansyah, Arya and van Dongen, Boudewijn: Replay- ing history on process models for conformance checking and performance analysis. WIREs Data Mining and Knowledge Discovery Volume 2, pp. 182-192 (2012)
- [6] van der Aalst and M.H. Schonenberg and M.Song: Time prediction based on process mining. Information Systems Volume 36, pp. 450-475 (2011)
- [7] Verenich, Ilya and Nguyen, Hoang and La Rosa, Marcello and Dumas, Marlon: White-Box Prediction of Process Performance Indicators via Flow Analysis. Pro- ceedings of the 2017 International Conference on Software and System Process, pp. 85–94 (2017)
- [8] Granger, C. W. J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, pp. 424 - 438 (1969)
- [9] Hamilton, James D: Time Series Analysis. Princeton University Press, pp 306-308 (1994)

