



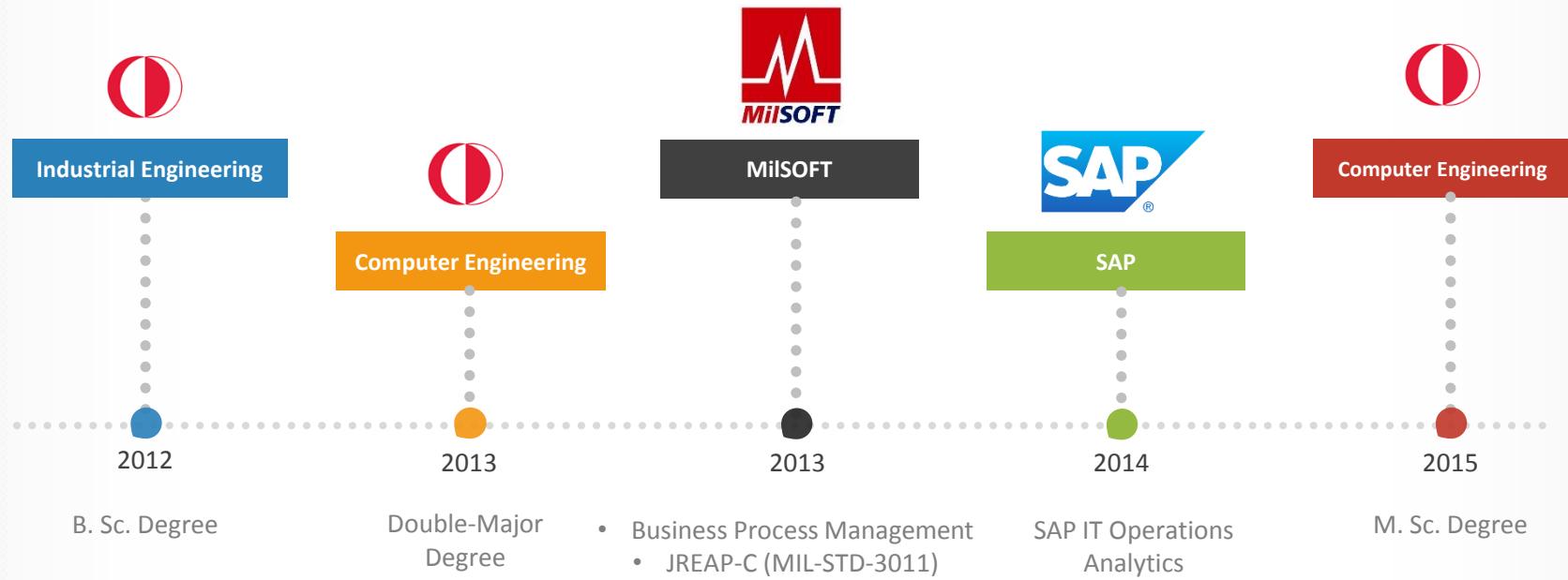
Recommendation Generation for Performance Improvement by using Cross-Organizational Process Mining

Onur Yılmaz

Supervisor: Assoc. Prof. Pınar Karagöz

September 1, 2015

Onur Yılmaz



Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

Introduction

Process Mining



Relatively young and developing research area and main idea is to

- discover,
- monitor and;
- improve processes by extracting information from event logs. [1]



Events recorded and available



Competitive business life

Introduction

Cross- Organizational Process Mining



Cloud computing
and shared
infrastructures

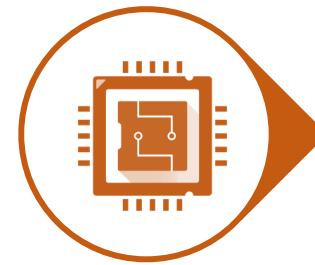


Event logs of
multiple
organizations [2]



Analyze the big
picture

Work together to execute
the same process



Learn from
each other

Execute the same task
on shared infrastructure [3]

Introduction

Focus of this Study



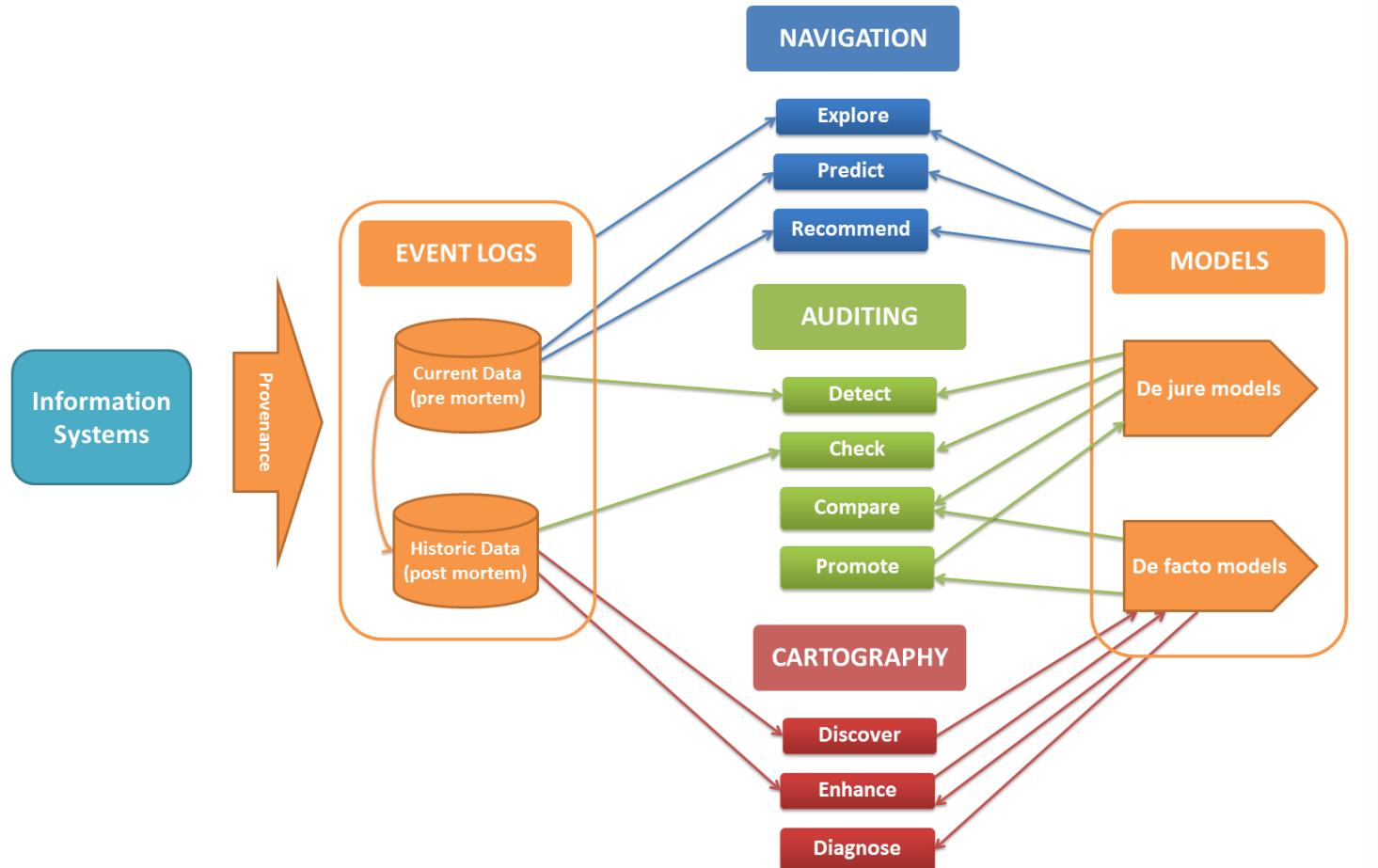
- A hybrid approach
 - Using different process mining subfields to create a new point of view
- Cross-organizational process mining
 - Processes are executed on several organizations,
 - Unsupervised learning using performances of organizations

Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

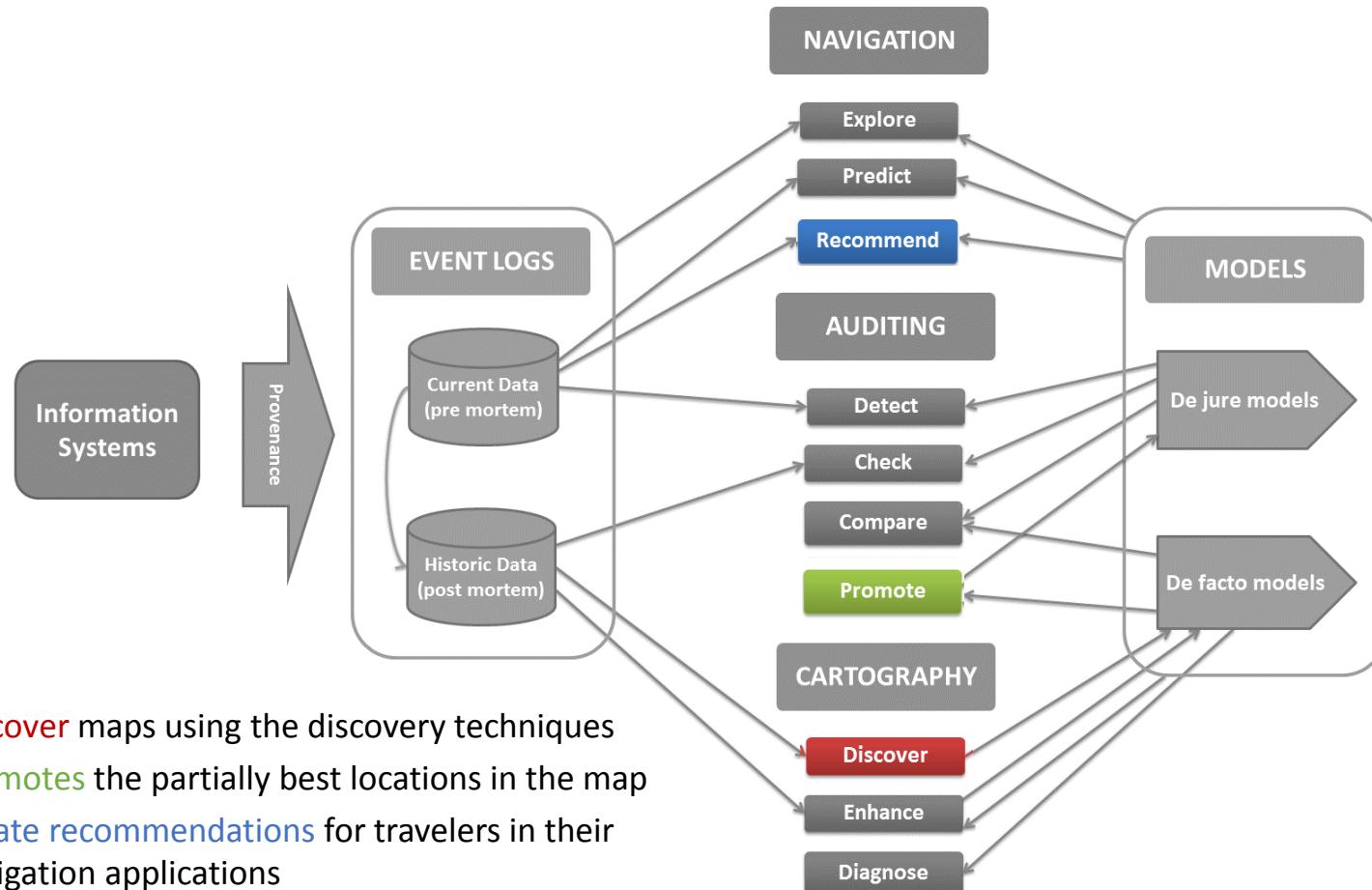
Related Work

State of the Art in Process Mining



Related Work

State of the Art in Process Mining



- **Discover** maps using the discovery techniques
- **Promotes** the partially best locations in the map
- **Create recommendations** for travelers in their navigation applications

Related Work

Process Discovery in Process Mining



- Various different process mining algorithms for solving different challenges in the process mining area
 - Alpha Algorithm [5] [6]
 - Inductive Approach [7] [8]
 - Hierarchical Clustering [9]
 - Genetic Approach [10] [11]
 - Heuristic Approach [12]

Robust, repeatable
and mature set of
approaches

Related Work

Cross-organizational Process Mining

- Analyze processes of organizations in an objective manner
- Matching behaviors and process models of organizations [2]
- Configurable process models for organizations [13]
- Intra and inter-organizational process mining:
 - Collaboration
 - Exploiting commonality



Learning opportunity of organizations from each other

Related Work

Process Similarity in Process Mining



- Emerging attention in business processes [2]
- Repository of process models of similar business operations [14]
- Methods in the literature
 - Similarity Metrics [15]
 - Alignment Matrix [16]
 - Delta Analysis [17] [18]
 - Mismatch Patterns [14]

Point the pattern differences between organizations

Related Work

Contributions of This Study

- Cross-organizational process mining approach for process performance improvement
- Generic, noise-capable process mining method for mining process models of different organizations
- Clustering of organizations based on their performance indicators
 - Unlike the clustering methods based on process structures in the literature [9]



Related Work

Contributions of This Study



- Mismatch analysis for spotting differences between processes of organizations
 - Formulation and implementation of patterns and analyzers
- Recommendation generation to show how organizations can learn from other organizations which perform better
- Open-source, extensible and configurable set of plugins in ProM framework



Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

Background

Event Log

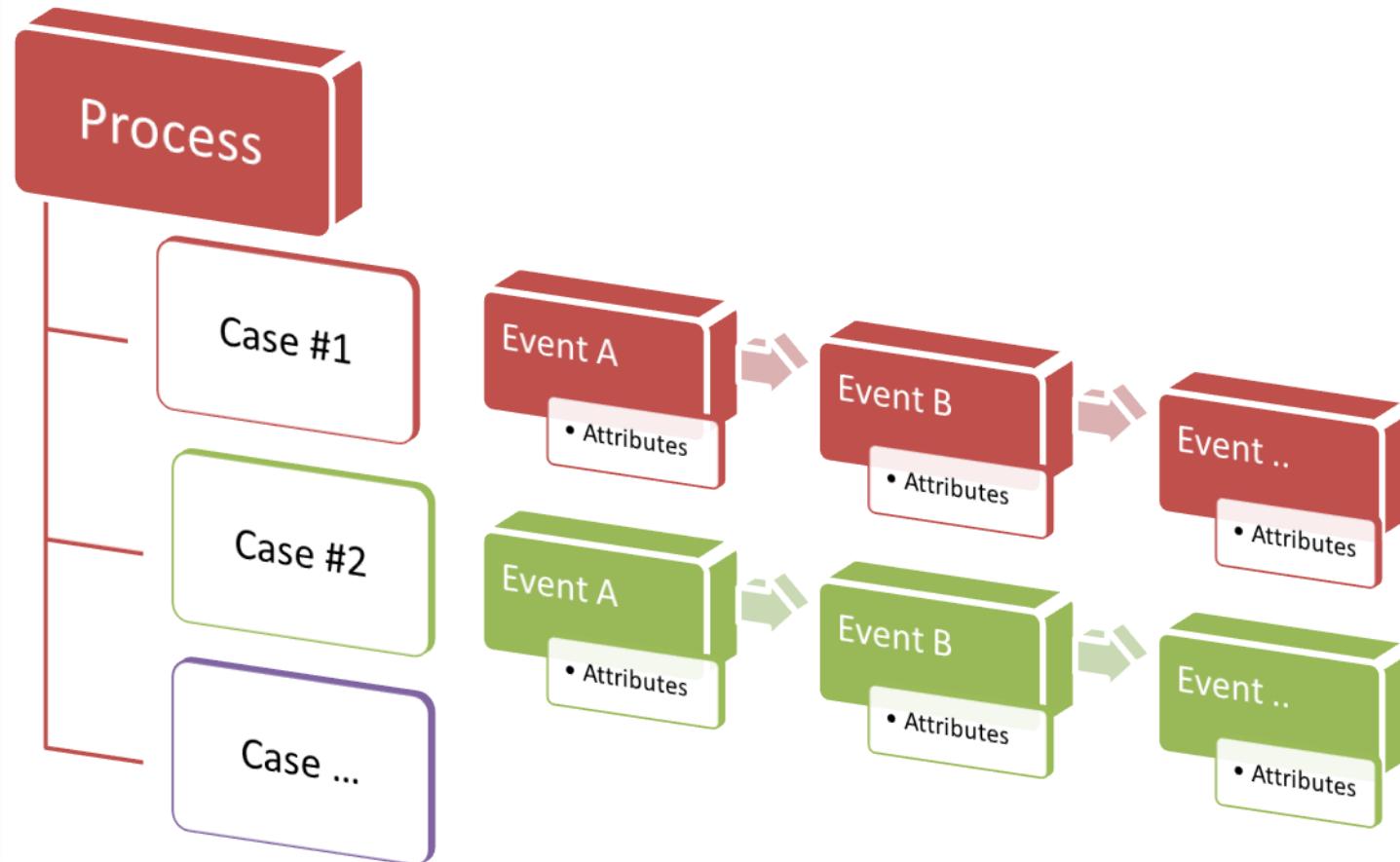
- Outputs of the software systems like Enterprise Resource Planning (ERP) or Business Process Management (BPM)



Event Log				
Attributes				
	Event	Date	Time	Transition
Case #1	Register Application	16.04.2013	14:37:27	Complete
	Check Credit	16.04.2013	14:41:19	Complete
	Check System	16.04.2013	14:47:35	Complete
	Calculate Capacity	16.04.2013	14:50:21	Complete
	Accept	16.04.2013	14:53:22	Complete
	Send decision e-mail	16.04.2013	14:55:11	Complete
Case #2	Register Application	16.04.2013	16:28:19	Complete
	Check Credit	16.04.2013	16:36:22	Complete

Background

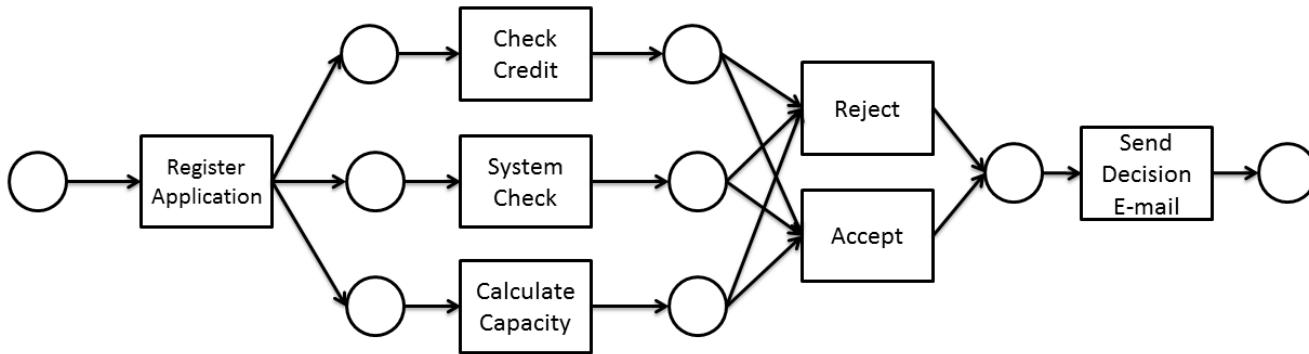
Event Log



Background

Process Modeling

- Workflow Net
 - Petri net with a start node, end node and connectedness

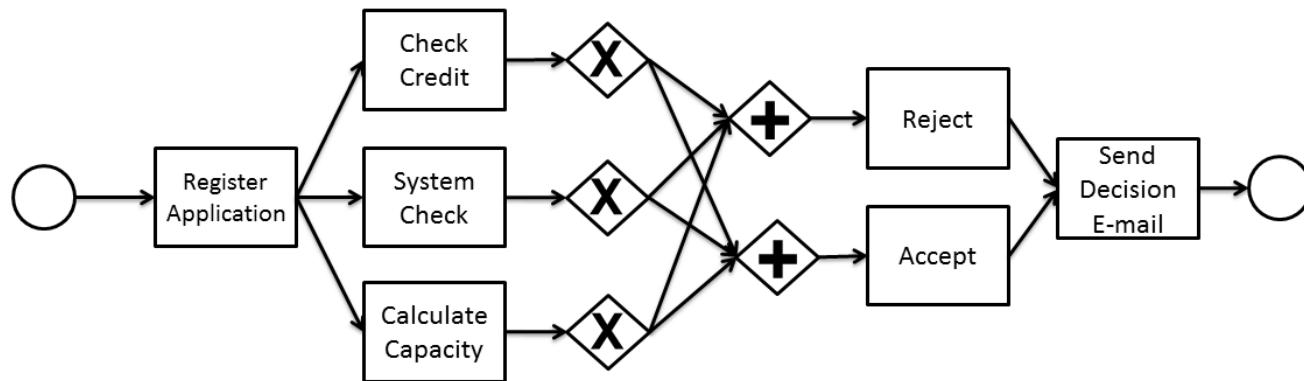


- Mathematical background

Background

Process Modeling

- Business Process Modelling Notation (BPMN)
 - Standardized and easy to understand by stakeholders



- Business oriented

Background

Process Discovery



- One of the most challenging tasks is to construct a process model based on the behavior in the event logs
- Inductive Process Mining is used since it is **simple, highly applicable** and **configurable**
 - Block-structured Workflow Nets
 - Rediscoverability [19]

Background

Process Discovery

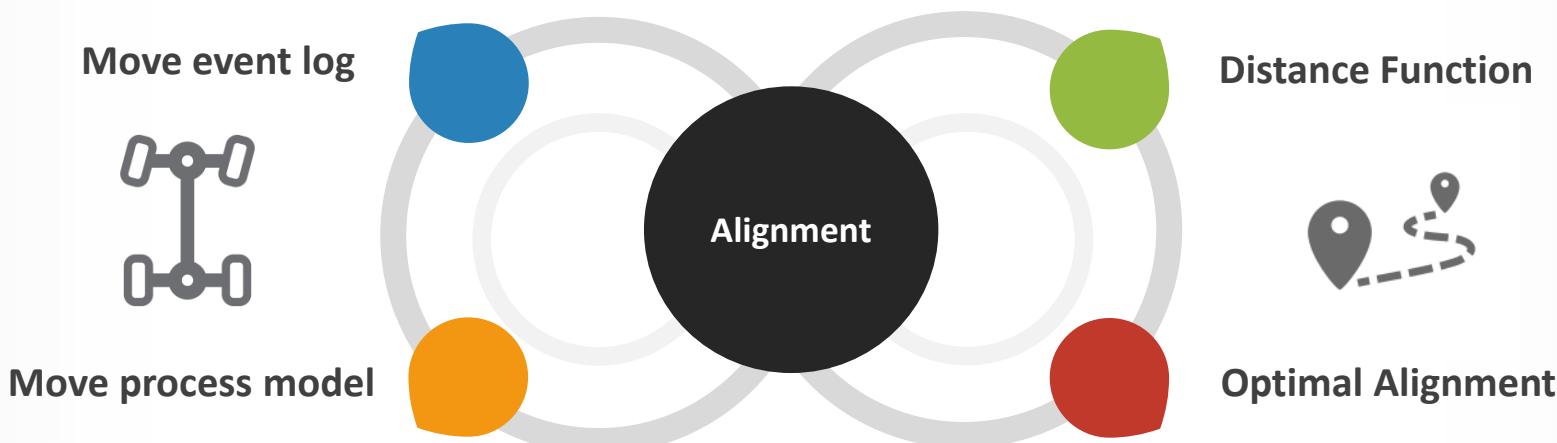


- **Inductive Miner Infrequent (IMi)** is proposed as an extension to Inductive Miner to handle noise in the event logs:
 1. **Activity Sets:** Split the activities in log to disjoint sets.
 2. **Sublogs:** Split the log by using activity sets.
 3. **Recursive Mining:** Mine sublogs with these steps until a sublog contains only single activity. [20]
- Filter infrequent behavior at every step by a user defined threshold

Background

Process Performance Analysis

- Discover relationships between event logs and process models for conformance and performance analysis [21]



Background

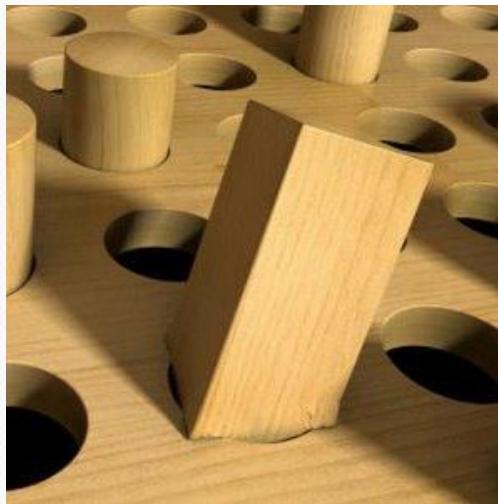
Clustering

- k-means algorithm (Lloyd's algorithm) and its variation based on random initialization:
 - k-means++ by Arthur and Vassilvitskii [22]
- Clustering of performance analysis results of organizations
- Implementation of k-means++ in WEKA [23]



Background

Mismatch Patterns in Process Models



- Patterns for frequent mismatches between the similar process models by Dijkman [14]



Authorization



Activity

- Skipped Activity
- Refined Activity



Control Flow

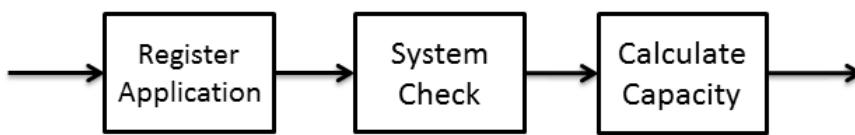
- Activities at Different Moments in Processes
- Different Conditions for Occurrence
- Different Dependencies
- Additional Dependencies

Background

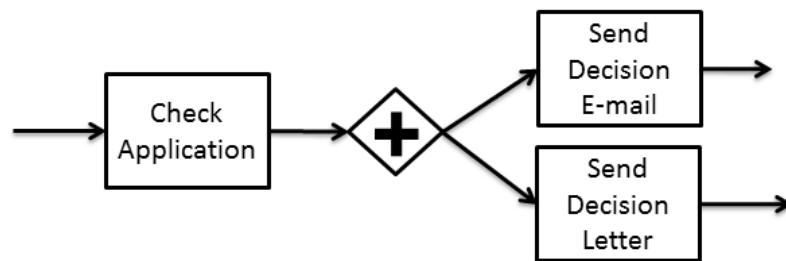
Mismatch Patterns in Process Models

Activity Mismatch Patterns [14]

- Skipped Activity



- Refined Activity



Background

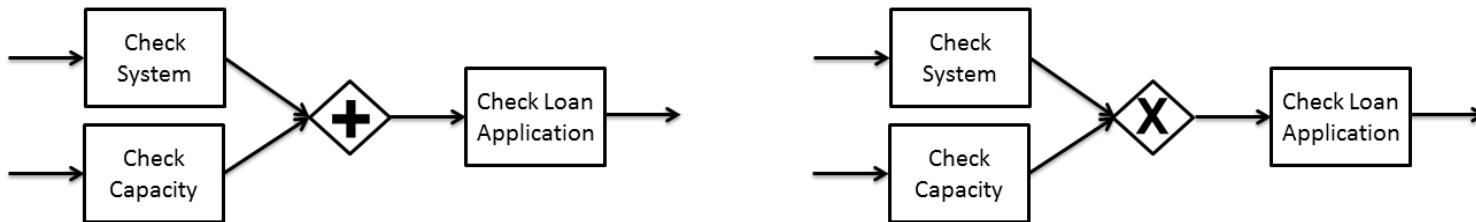
Mismatch Patterns in Process Models

Control Flow Mismatch Patterns [14]

- Activities at Different Moments in Processes



- Different Conditions for Occurrence

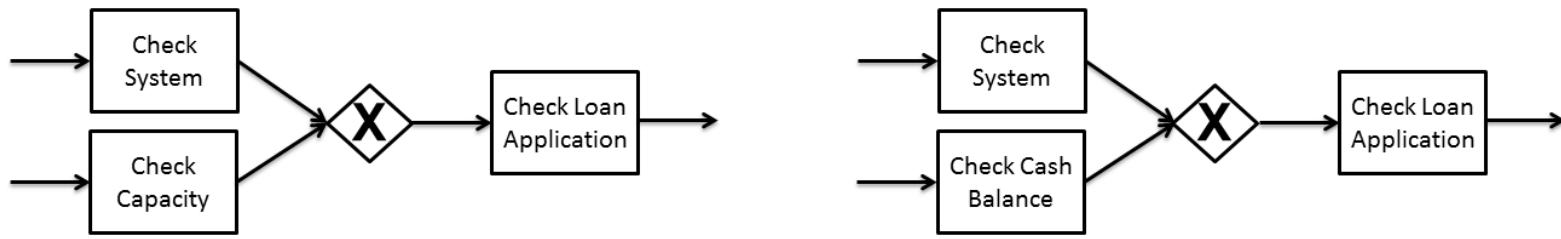


Background

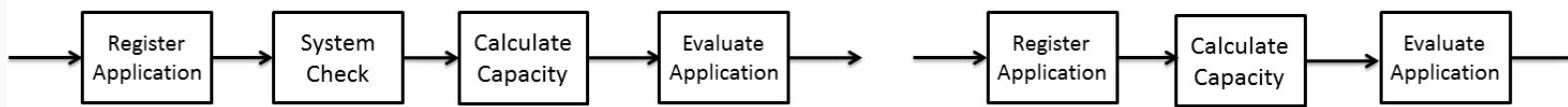
Mismatch Patterns in Process Models

Control Flow Mismatch Patterns [14]

- Different Dependencies



- Additional Dependencies



Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

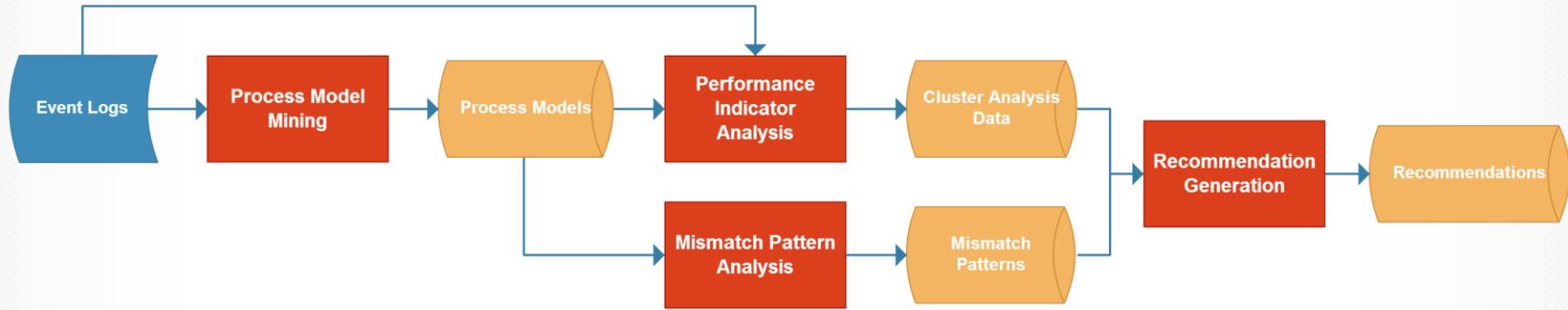
Methodology

Approach Overview



Methodology

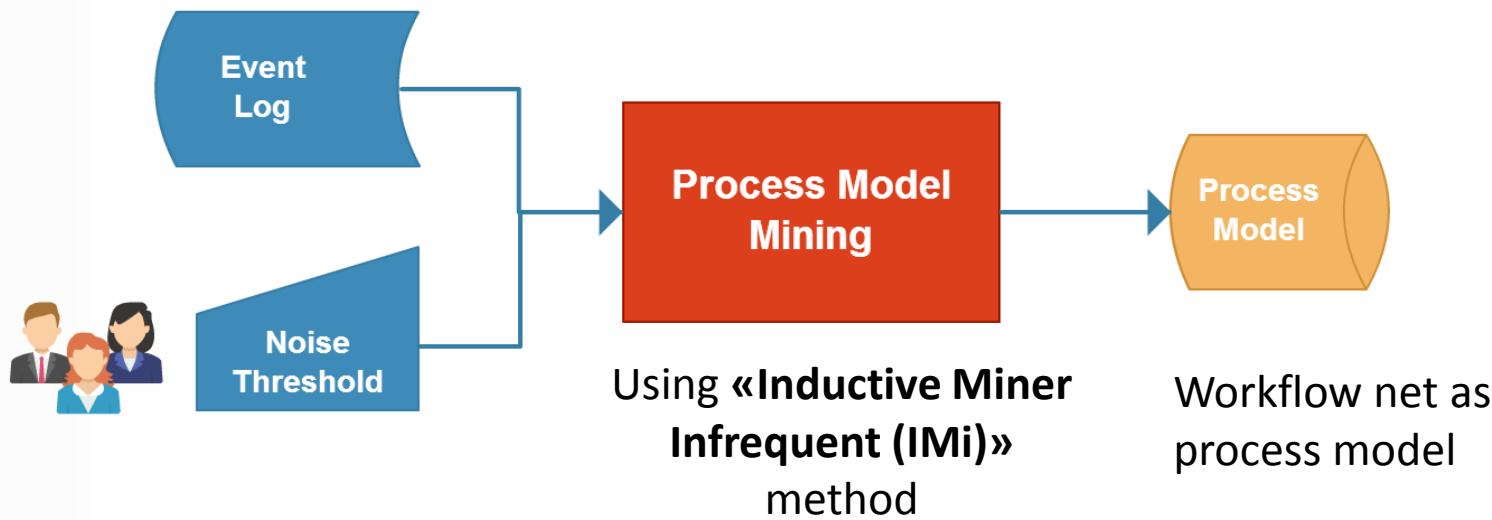
Approach Overview



Methodology

Process Model Mining

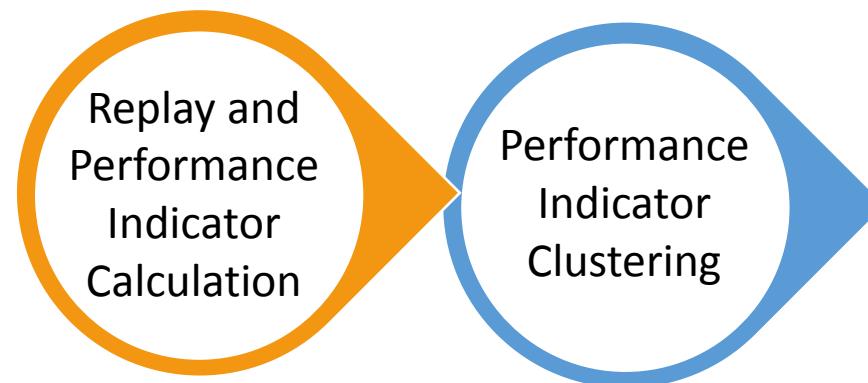
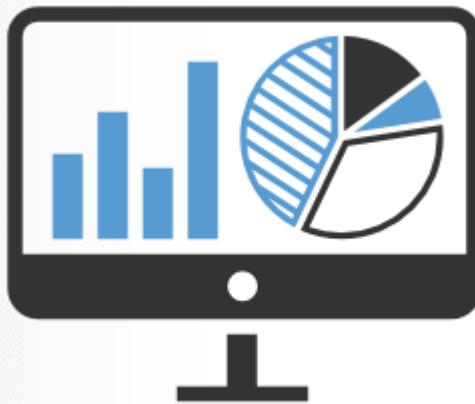
Applied for each organization:



Methodology

Performance Indicator Analysis

- Two steps:



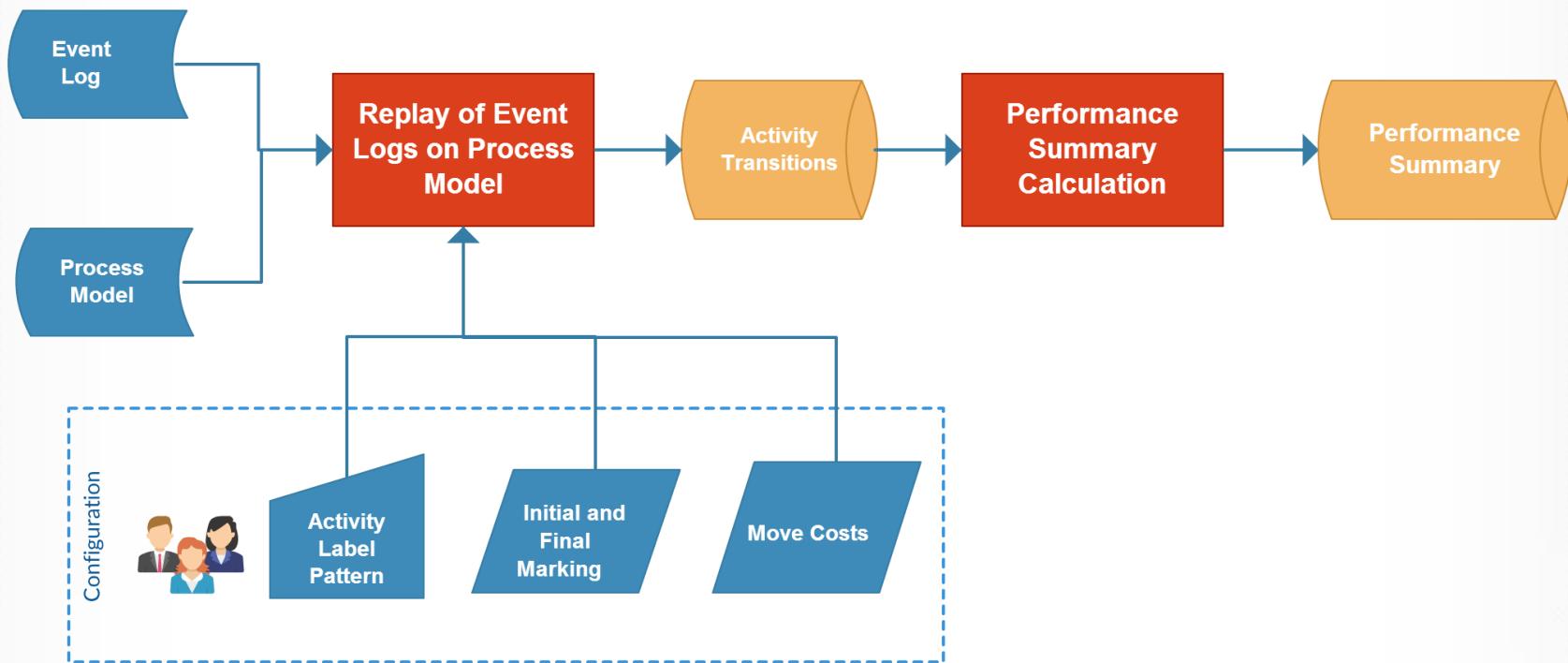
- Performance Indicators:
 - Average Time Between Activities
 - Standard Deviation of Time Between Activities



Methodology

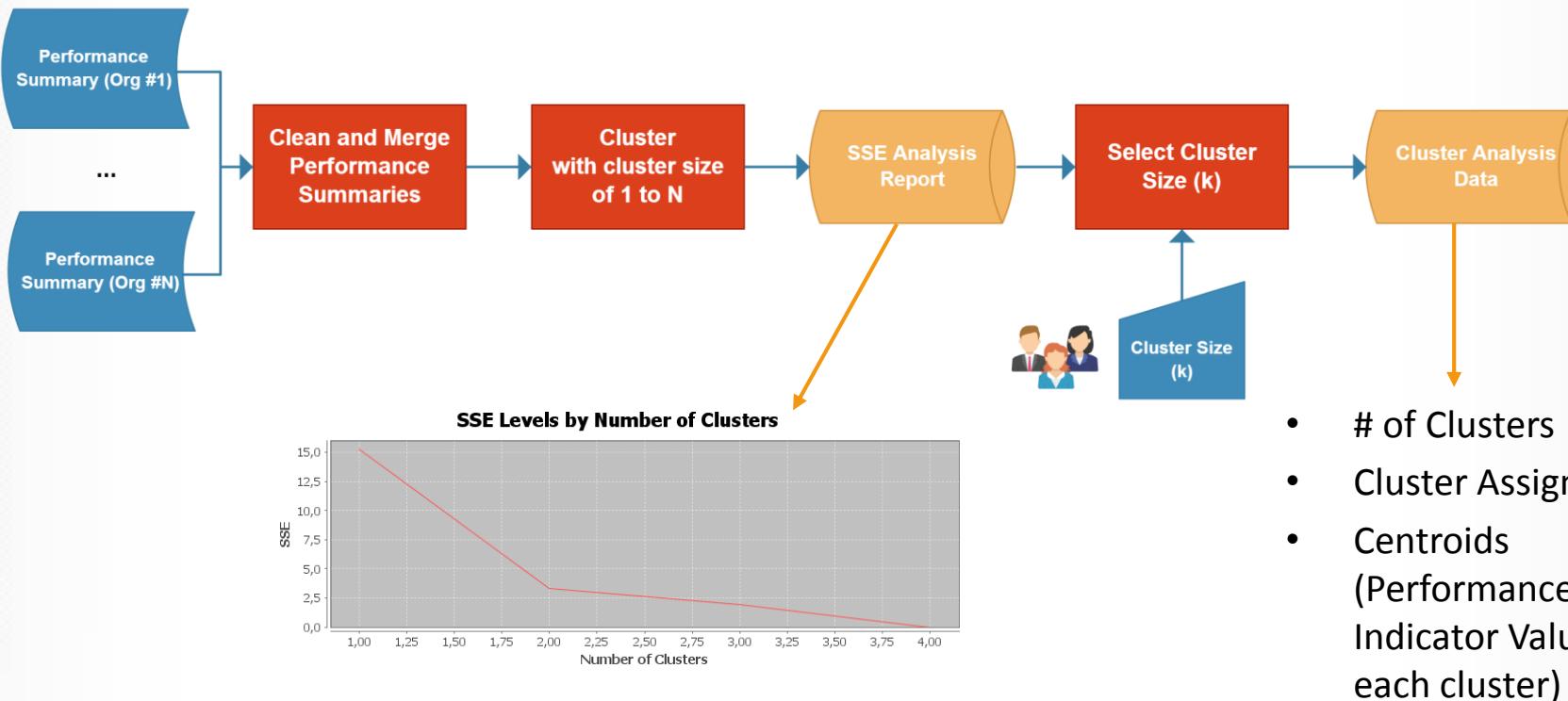
Performance Indicator Analysis - Replay and Performance Indicator Calculation

Applied for each organization:



Methodology

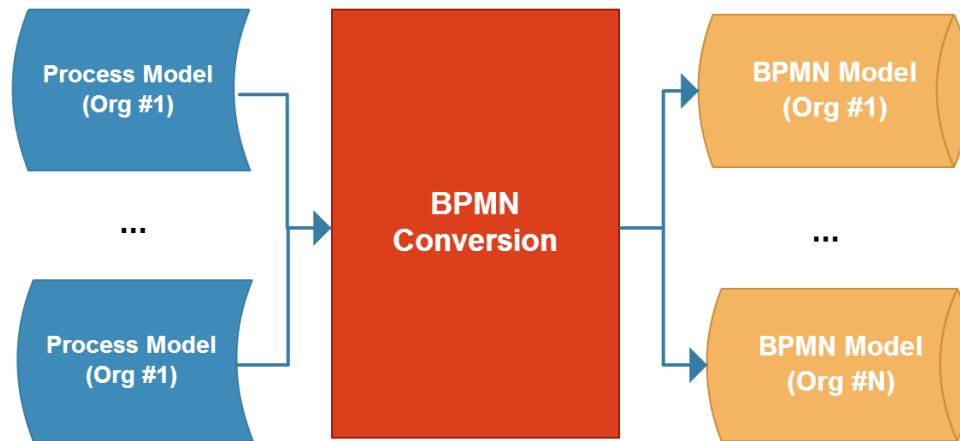
Performance Indicator Analysis – Performance Indicator Clustering



Methodology

Mismatch Pattern Analysis

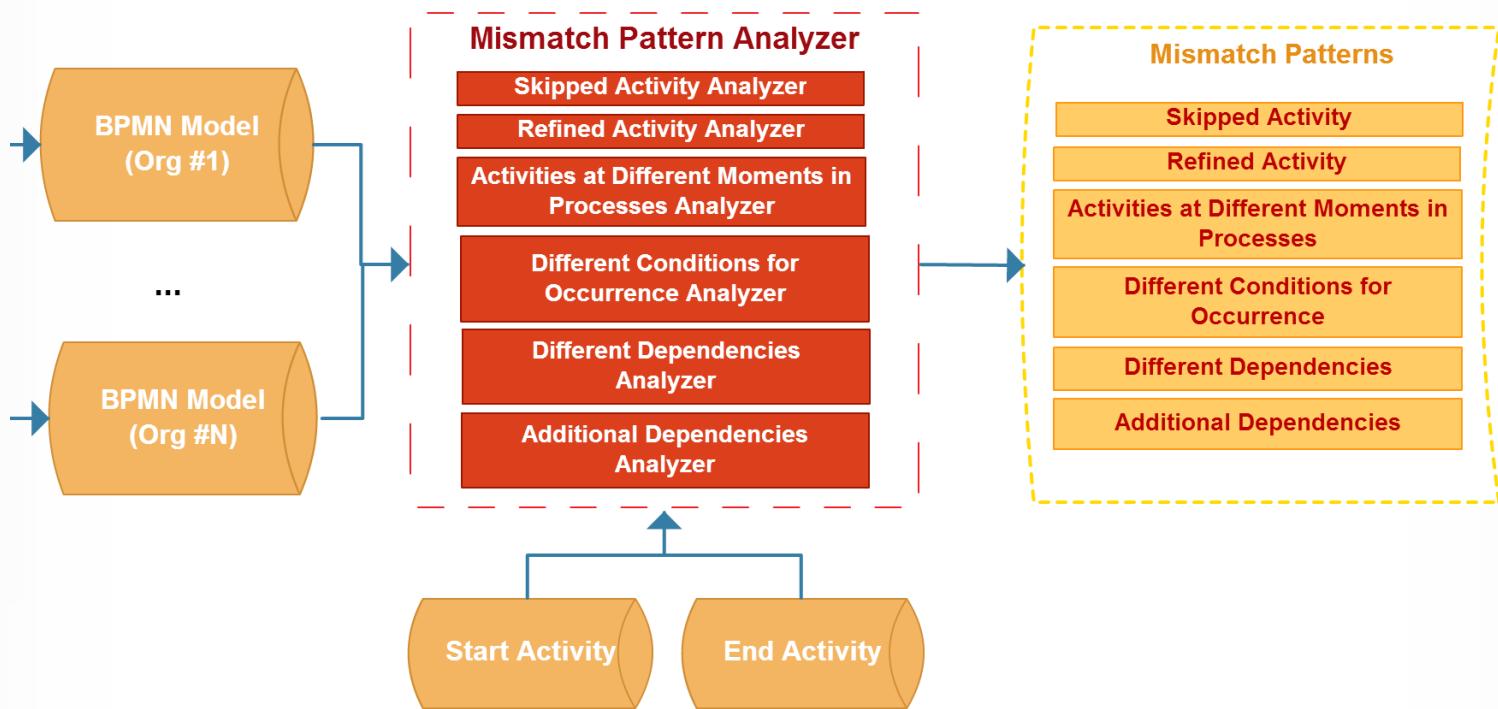
- Spot the differences between process models of different organizations as mismatch patterns
- BPMN used since notation is more appropriate to formulate mismatch patterns



Methodology

Mismatch Pattern Analysis

- Mismatch patterns and analyzers are developed



Methodology

Mismatch Pattern Analysis

- Mismatch patterns and analyzers formulated

input: O_1 first organization, O_2 second organization, A_{start} starting activity, A_{end} ending activity

output: MismatchPatterns a set of mismatch patterns

```
MismatchPatterns <-- {}
MismatchPatterns <-- SkippedActivityAnalyzer( $O_1$ ,  $A_{start}$ ,  $A_{end}$ )
MismatchPatterns <-- RefinedActivityAnalyzer( $O_1$ ,  $O_2$ ,  $A_{start}$ ,  $A_{end}$ )
MismatchPatterns <-- DifferentMomentsAnalyzer( $O_1$ ,  $O_2$ ,  $A_{start}$ ,  $A_{end}$ )
MismatchPatterns <-- DifferentConditionsAnalyzer( $O_1$ ,  $O_2$ ,  $A_{start}$ ,  $A_{end}$ )
MismatchPatterns <-- DifferentDependencyAnalyzer( $O_1$ ,  $O_2$ ,  $A_{start}$ ,  $A_{end}$ )
MismatchPatterns <-- AdditionalDependencyAnalyzer( $O_1$ ,  $O_2$ ,  $A_{start}$ ,  $A_{end}$ )
return MismatchPatterns
```

Methodology

Recommendation Generation

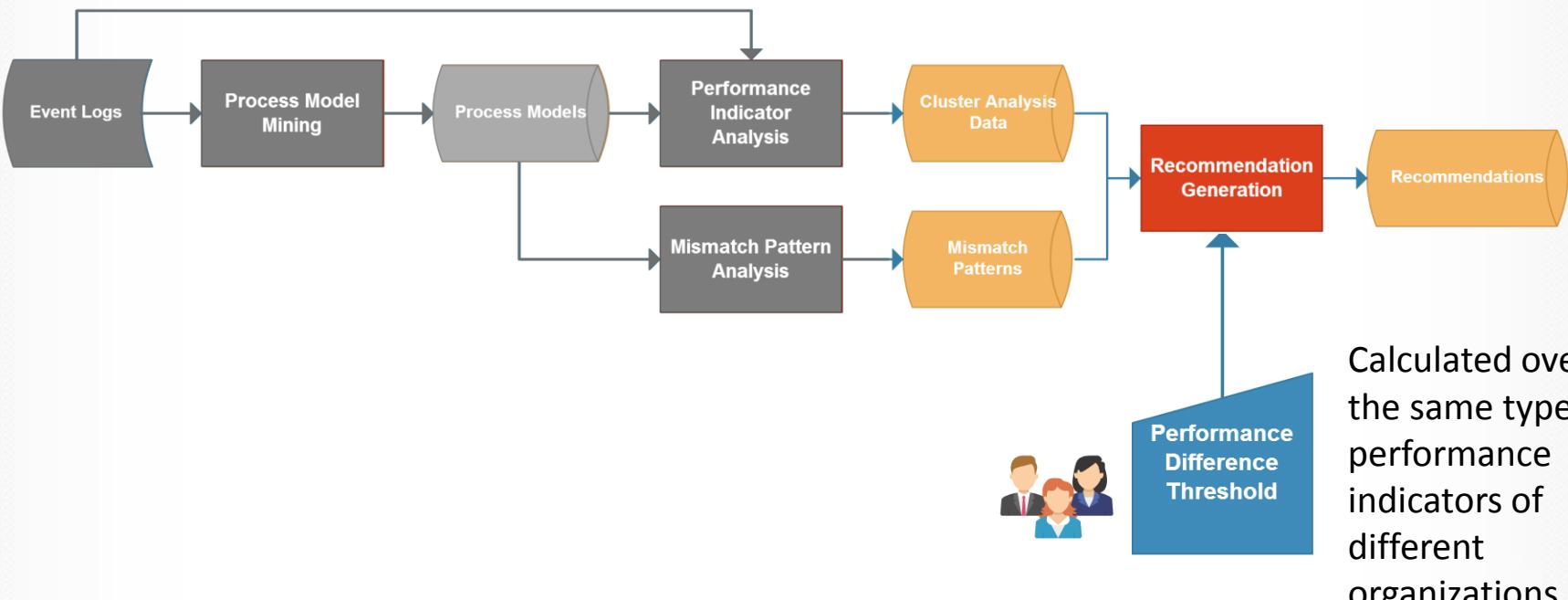


- Providing a set of mismatch patterns for each organization to enhance their processes.
 - Mismatch patterns between organizations, which are **performing better** in terms of their performance indicator values.

Recommendation = (Organization, A_{start}, A_{end}, Mismatch Patterns)

Methodology

Recommendation Generation



Methodology

Recommendation Generation

RecommendationGeneration

input: **O** organization, **C** Cluster Analysis Data, **P** performance difference threshold
output: **Recommendations** a set of recommendations

```
Recommendations ← {}
i ← Cluster of organization O
for each centroid for cluster i
    Get other cluster j with the centroids of Astart and Aend; and
    value difference larger than P
    for each organization O' in the cluster j
        MismatchPatterns ← Mismatch Pattern Analysis(O, O', Astart, Aend)
        Recommendations ← Recommendation(O, Astart, Aend, MismatchPatterns)
```

Methodology

ProM Implementation



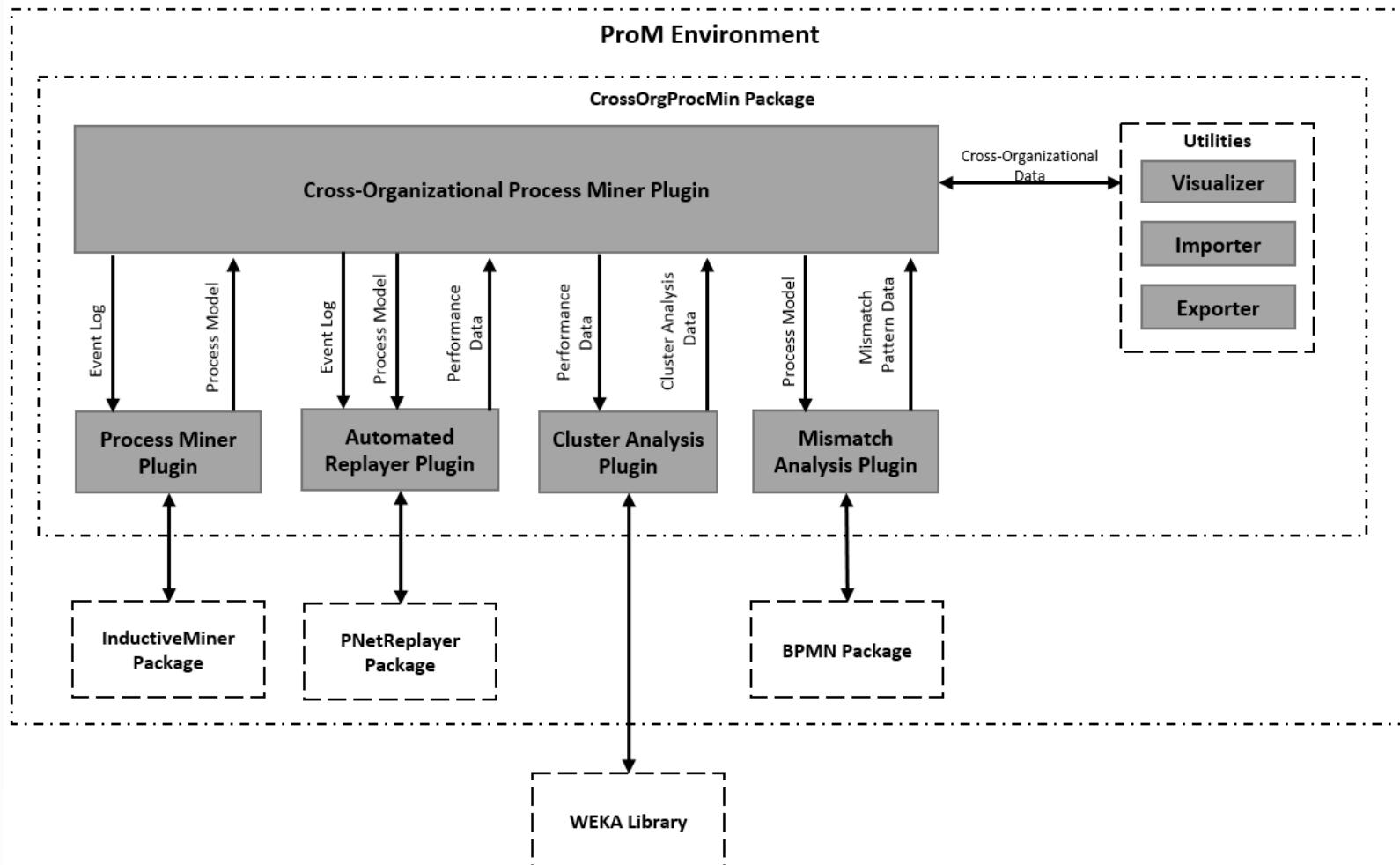
- ProM, extensible framework that supports a wide variety of process mining techniques in form of plugins [24]
- Widely accepted in industry and academia with an active community
- Developed set of plugins are packaged with the name of **«CrossOrgProcMin»** and published on Github.



[onuryilmaz/cross-orgc-proc-min](https://github.com/onuryilmaz/cross-orgc-proc-min)

Methodology

ProM Implementation

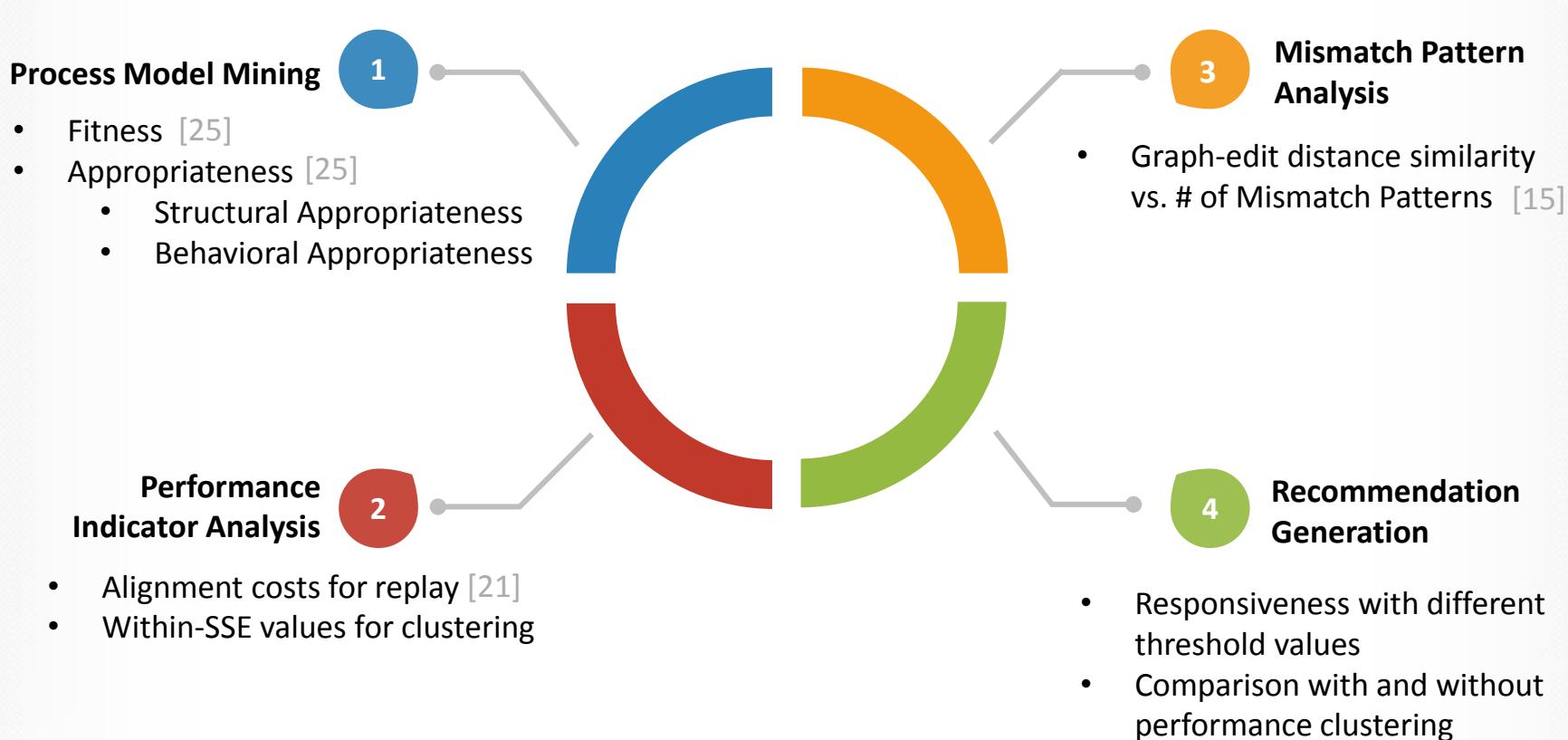


Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

Results & Discussions

Evaluation Metrics



Results & Discussions

Dataset Selection



- **Loan Application Process** [26]
 - Synthetically generated
 - 4 variants of a simple loan application in a financial institute
- **Environmental Permit Application Process** [27]
 - Real-life event log from "Configurable Services for Local Governments (CoSeLoG)" project [13]
 - «Environmental Permit Application Process» of 5 municipalities in Netherlands

Results & Discussions

Loan Application Process

	Cases	Events	Percentage
Variant #1	100	590	24 %
Variant #2	70	420	17 %
Variant #3	200	800	33 %
Variant #4	105	630	26 %

Cases Events
475 2440

- These variants are used as organizational logs

Results & Discussions

Loan Application Process – Process Model Mining

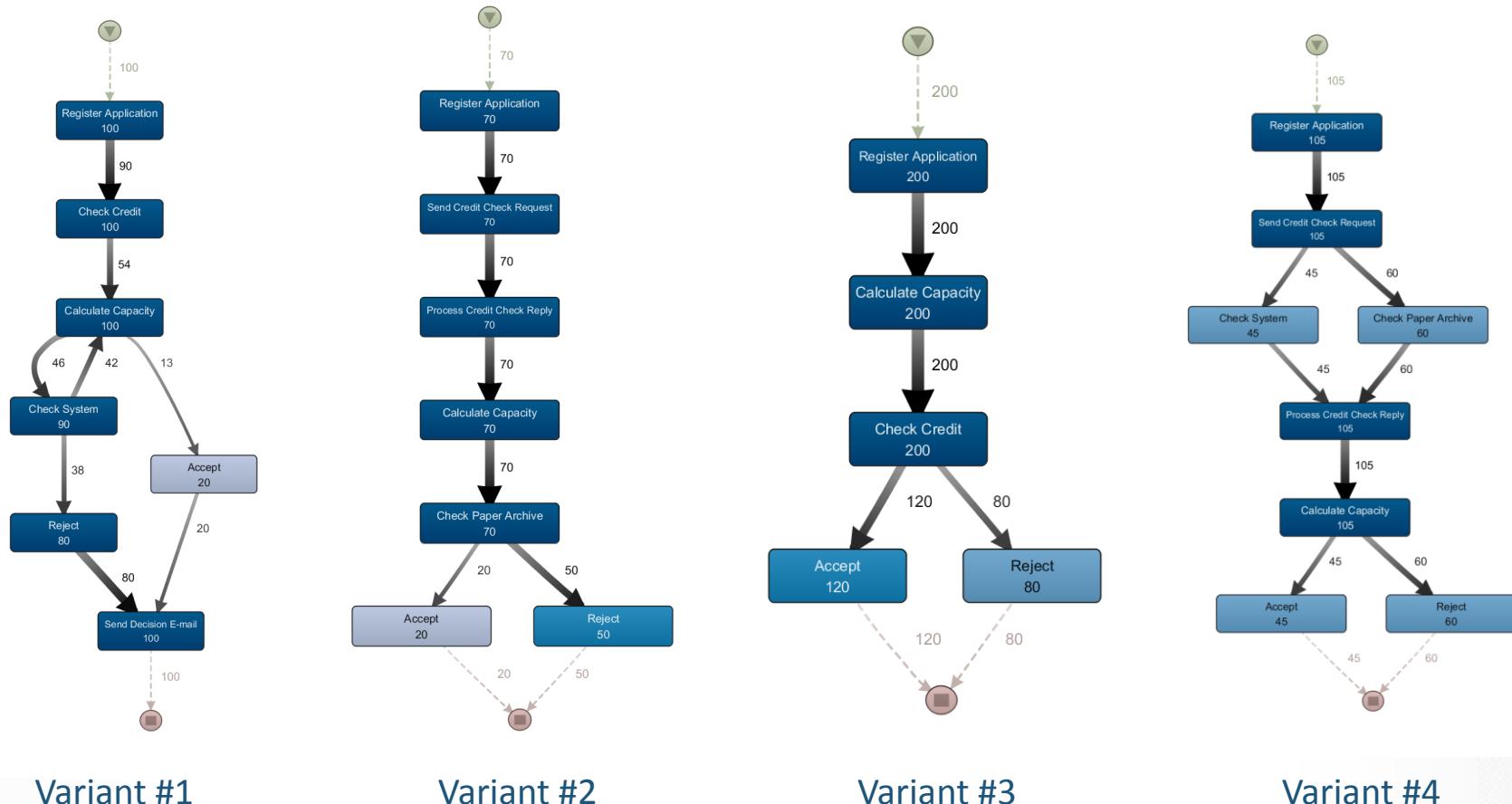
- With «0» noise threshold

	Fitness	Structural Appropriateness	Behavioral Appropriateness	Average Appropriateness
Variant #1	100 %	70 %	98.5 %	84.2 %
Variant #2	100 %	100 %	100 %	100 %
Variant #3	100 %	100 %	100 %	100 %
Variant #4	100 %	100 %	98.2 %	99.1 %
Average	100 %	92.5 %	99.7 %	96.6 %

- Each process model is fitting to event log and «appropriate»

Results & Discussions

Loan Application Process – Process Model Mining



Variant #1

Variant #2

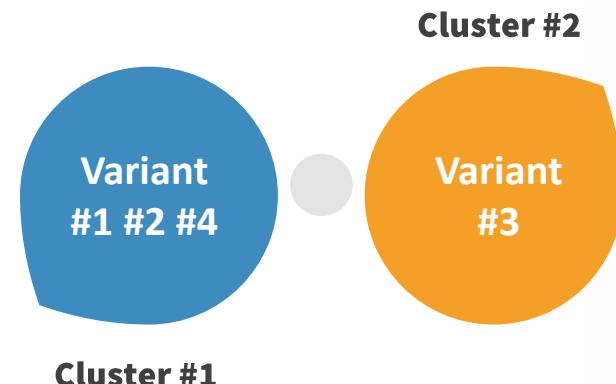
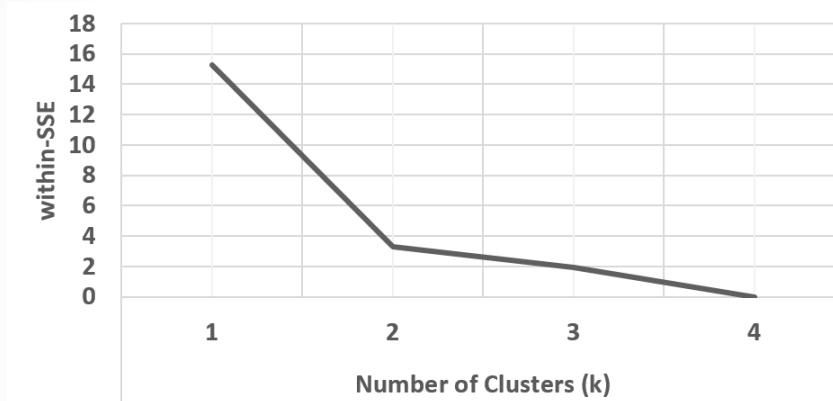
Variant #3

Variant #4

Results & Discussions

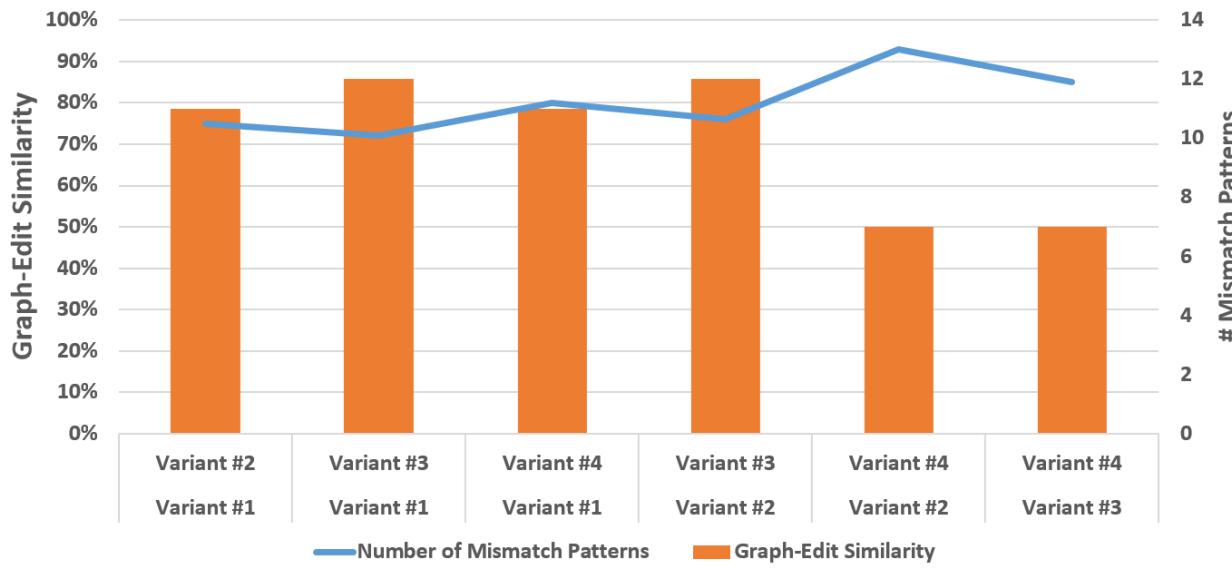
Loan Application Process – Performance Indicator Analysis

- **Performance Indicator Calculation:**
 - Replay costs are all zero since 100 % fitness is achieved
- **Clustering:**



Results & Discussions

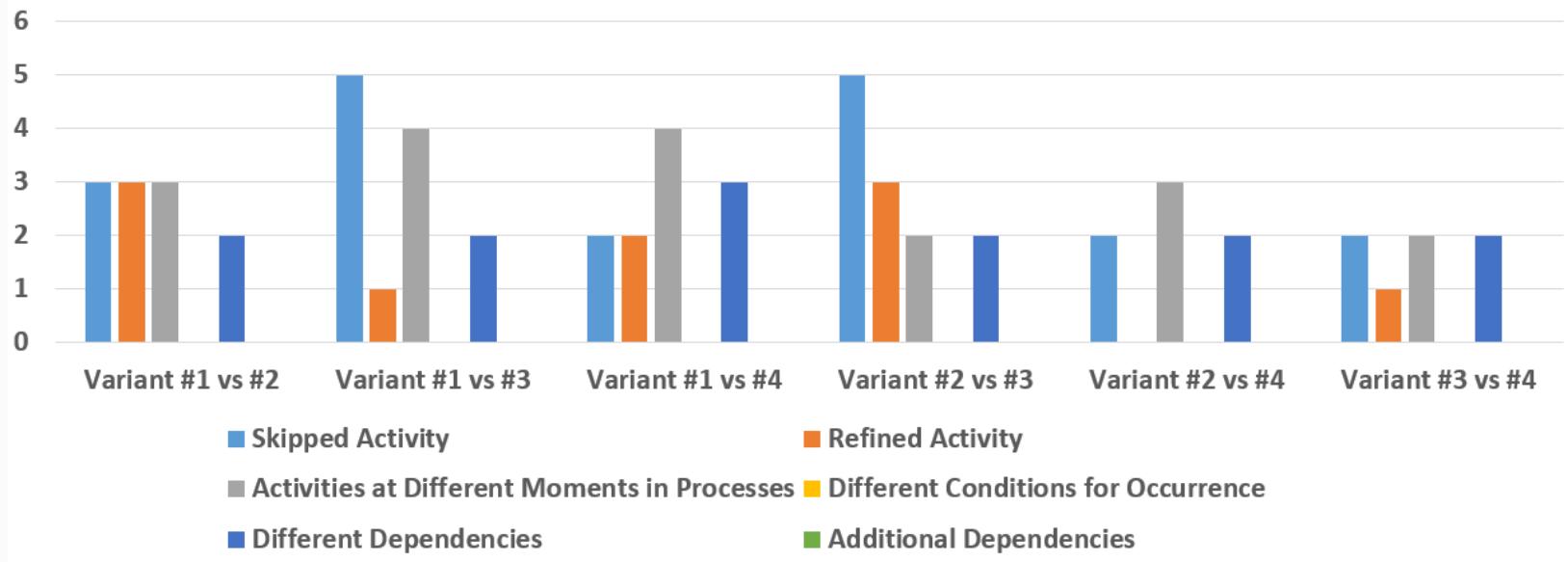
Loan Application Process – Mismatch Pattern Analysis



Correlation between graph-edit similarity and number of mismatch patterns

Results & Discussions

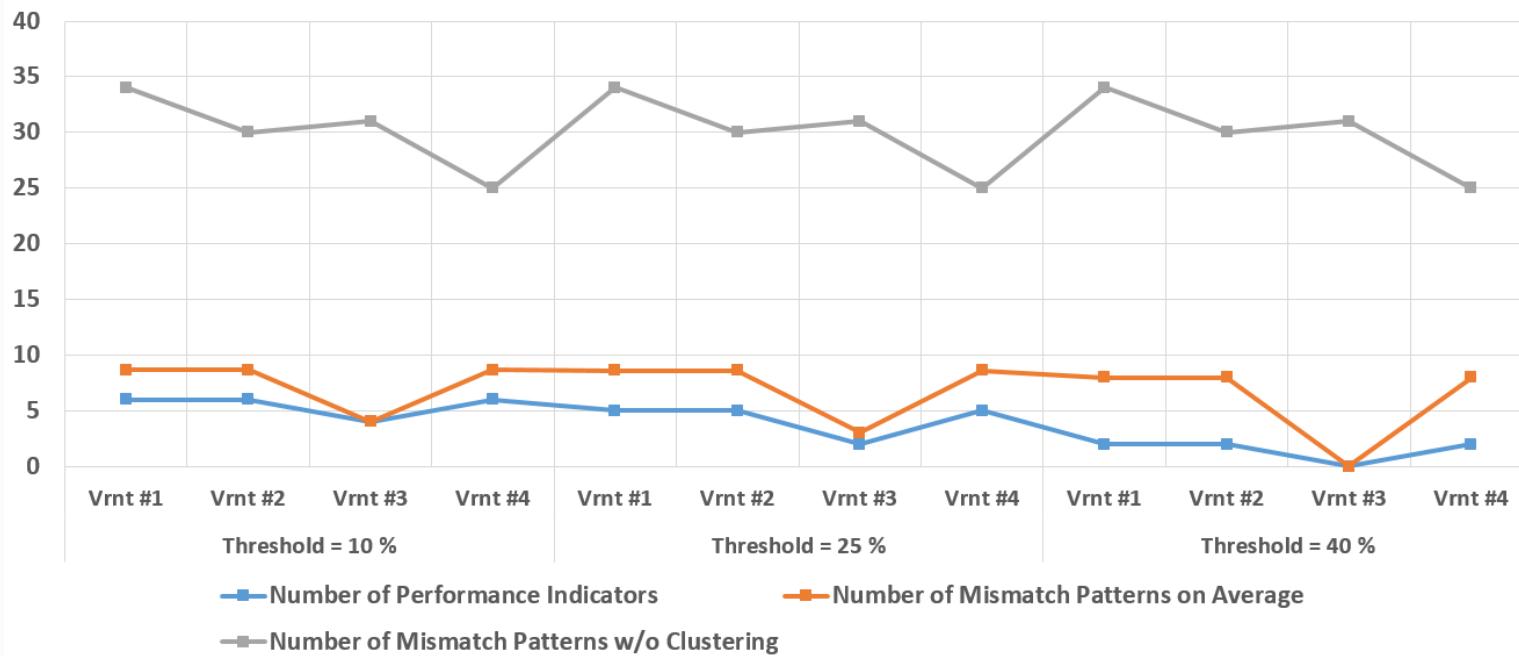
Loan Application Process – Mismatch Pattern Analysis



- "Skipped Activity" and "Activities at Different Moments" mostly
- No "Different Conditions for Occurrence" or "Additional Dependencies"

Results & Discussions

Loan Application Process – Recommendation Generation



- Responsiveness and degree of helping the user to focus on the performance improvement

Results & Discussions

Environmental Permit Application Process

	Cases	Events	Percentage
Municipality #1	54	131	6.1 %
Municipality #2	302	586	27.3 %
Municipality #3	37	73	3.4 %
Municipality #4	340	507	23.7 %
Municipality #5	481	845	39.4 %

Cases Events
1214 **2142**

- Preprocessing is undertaken on the raw dataset [28]
- These municipalities are used as organizational logs

Results & Discussions

Environmental Permit Application Process – Process Model Mining

- With 10 % noise threshold

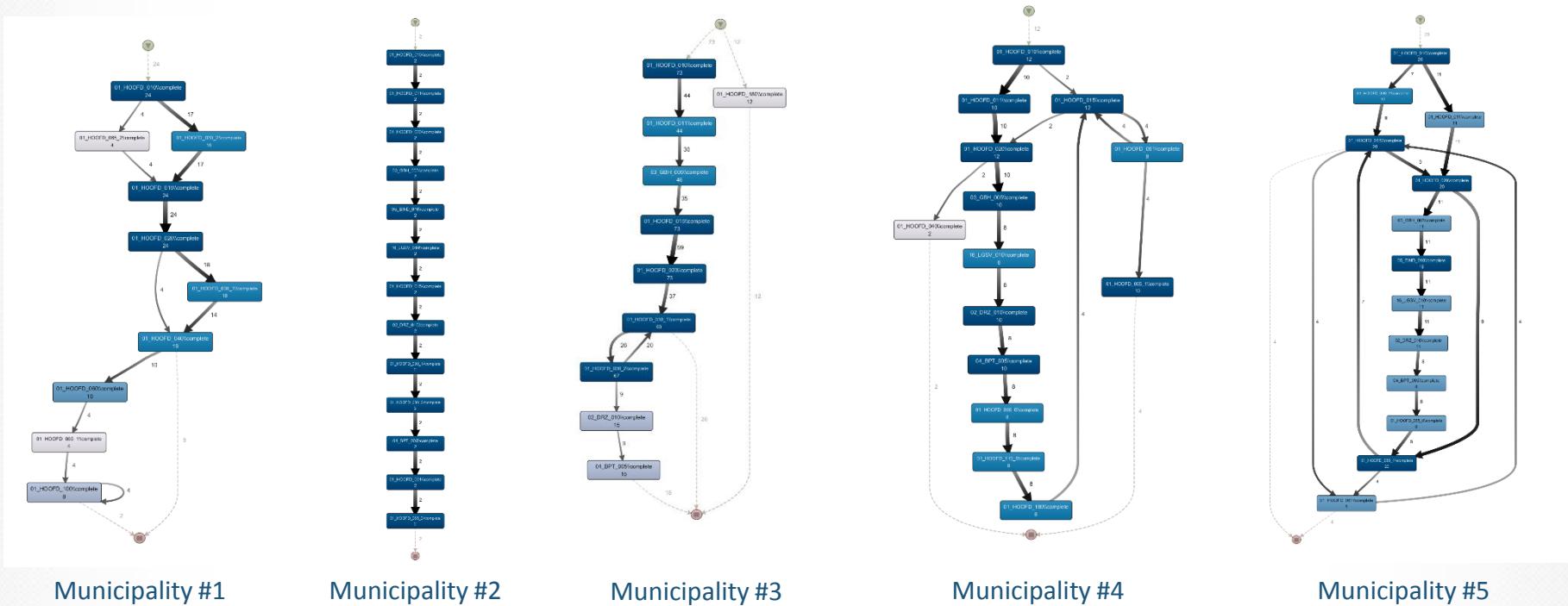
	Fitness	Structural Appropriateness	Behavioral Appropriateness	Average Appropriateness
Municipality #1	86 %	97.5 %	54.4 %	76 %
Municipality #2	100 %	100 %	100 %	100 %
Municipality #3	92.3 %	71.1 %	67.2 %	69.1 %
Municipality #4	96.8 %	65.7 %	64 %	64.9 %
Municipality #5	94.5 %	58.8 %	39.7 %	49.3 %
Average	93.9 %	78.6 %	65.1 %	71.9 %

- High fitness values except Municipality #4 and #5

Results & Discussions

Environmental Permit Application Process – Process Model Mining

- 10 to 20 times simplified process models



Results & Discussions

Environmental Permit Application Process – Performance Indicator Analysis

- **Performance Indicator Calculation:**

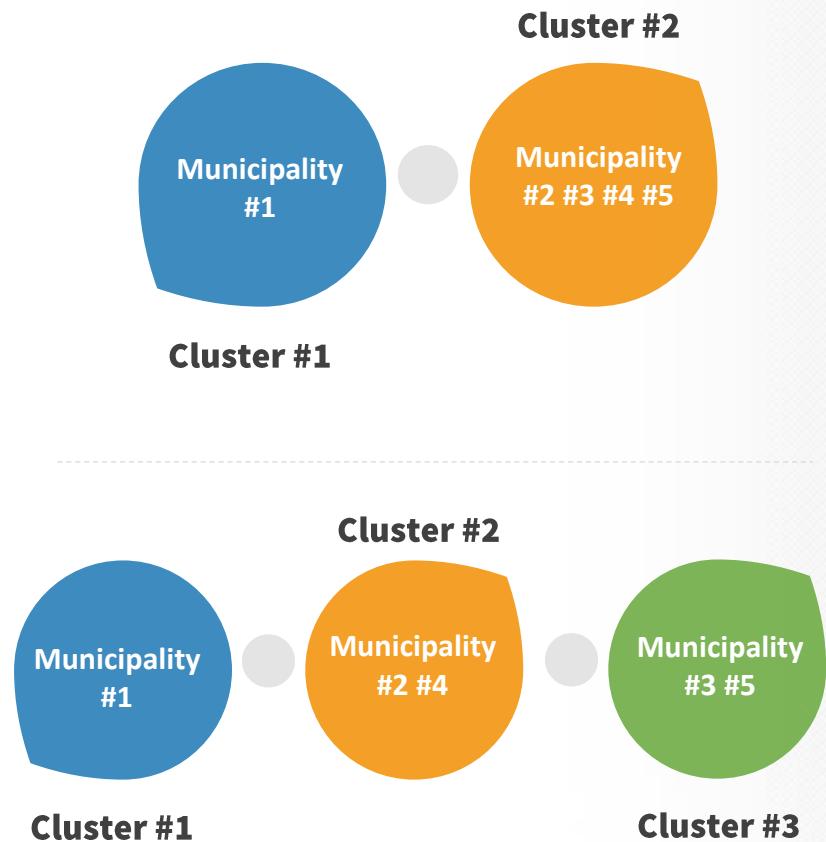
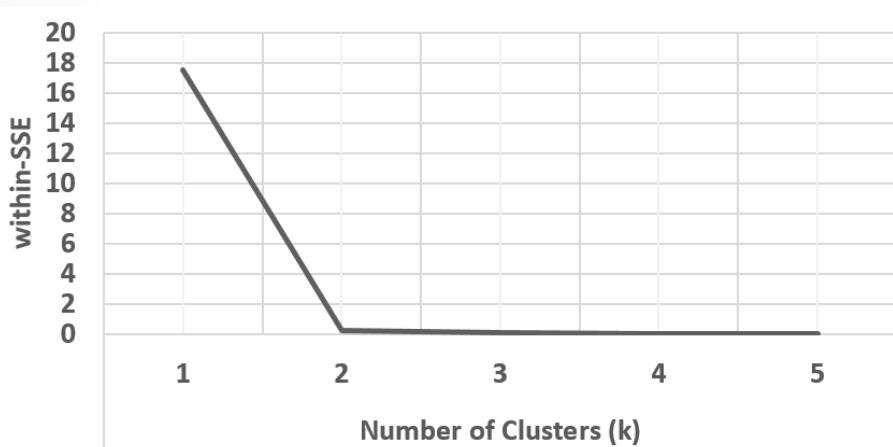
- As appropriateness and fitness decrease; alignment costs increase for the municipalities
- Performance indicators calculated over replay are acceptable

	Fitness	Average Appropriateness	Alignment Cost
Municipality #1	86 %	76 %	173.2
Municipality #2	100 %	100 %	0
Municipality #3	92.3 %	69.1 %	323.3
Municipality #4	96.8 %	64.9 %	9.1
Municipality #5	94.5 %	49.3 %	35.8

Results & Discussions

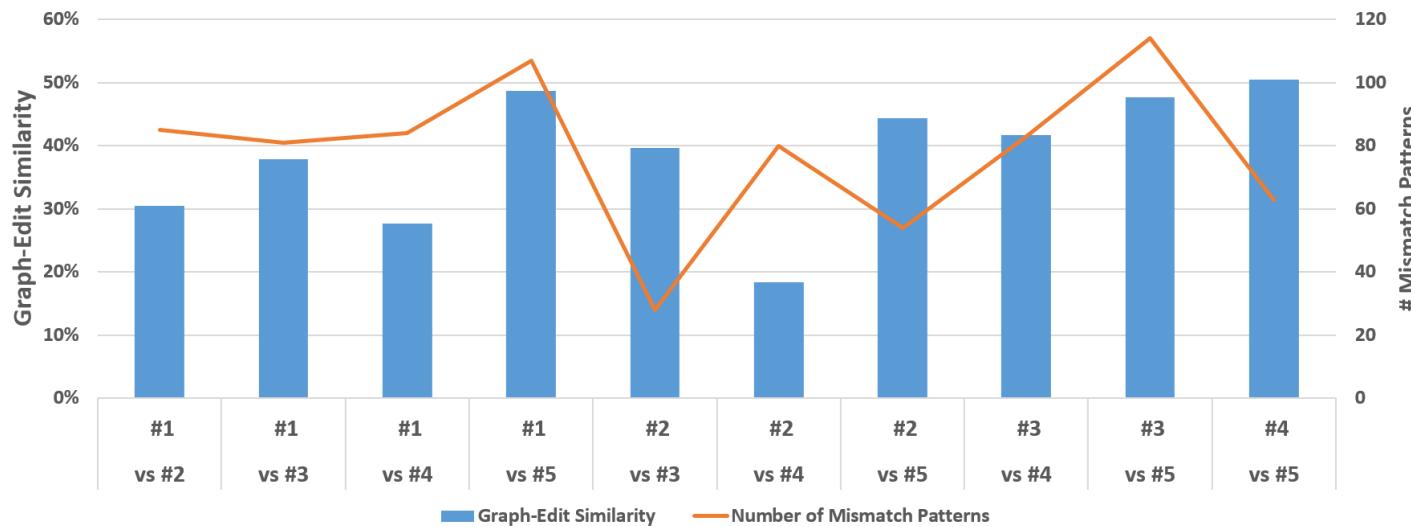
Environmental Permit Application Process – Performance Indicator Analysis

- **Clustering:**



Results & Discussions

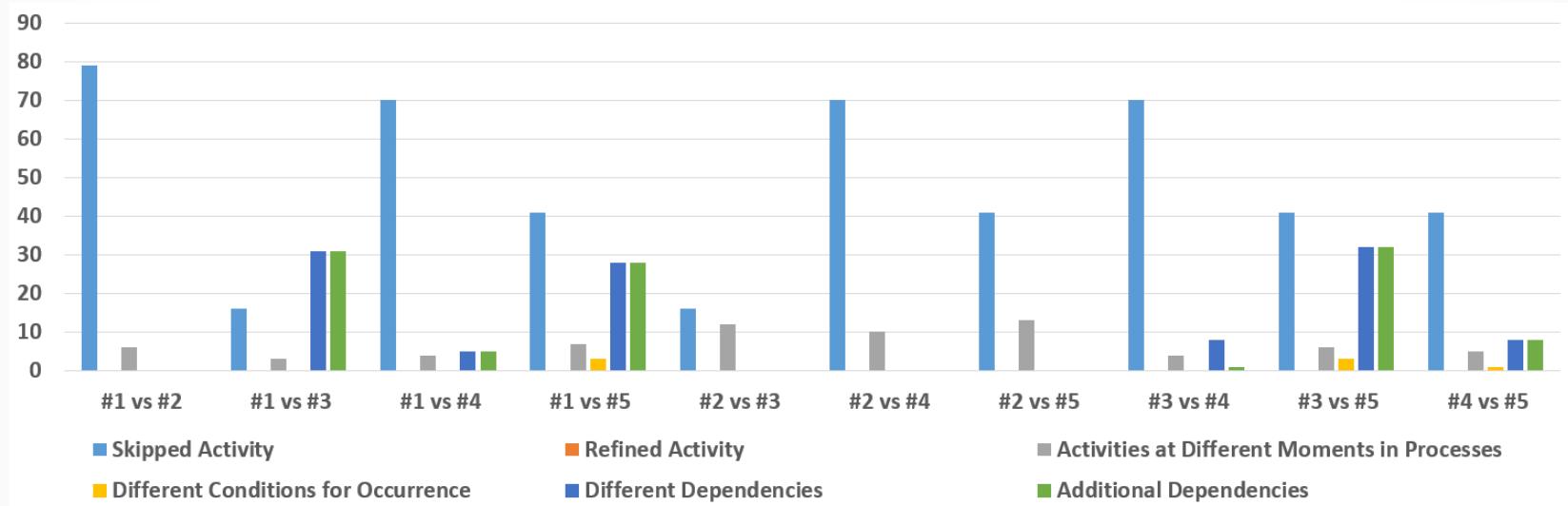
Environmental Permit Application Process – Mismatch Pattern Analysis



- Correlation between graph-edit similarity and number of mismatch patterns except Municipality #4 and #5

Results & Discussions

Environmental Permit Application Process – Mismatch Pattern Analysis

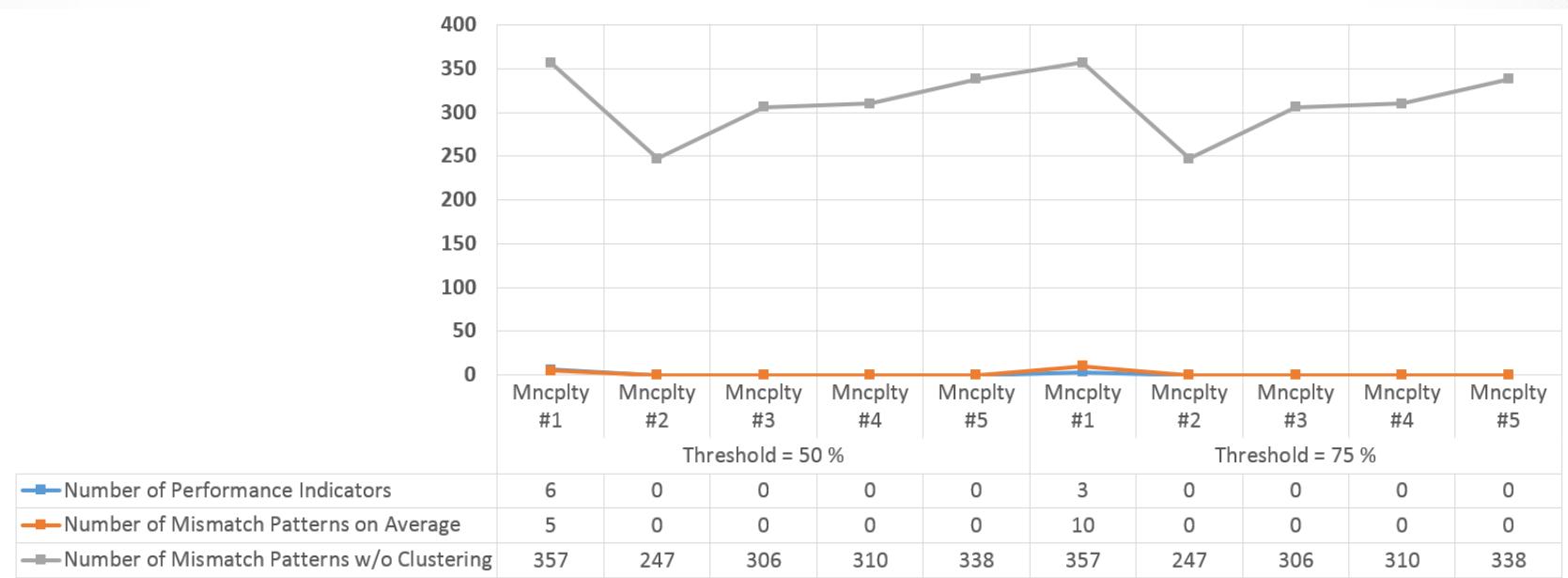


- "Skipped Activity" mostly
- "Refined Activity" is eliminated since codes are used instead of activity names

Results & Discussions

Environmental Permit Application Process – Recommendation Generation

(with 2 clusters)

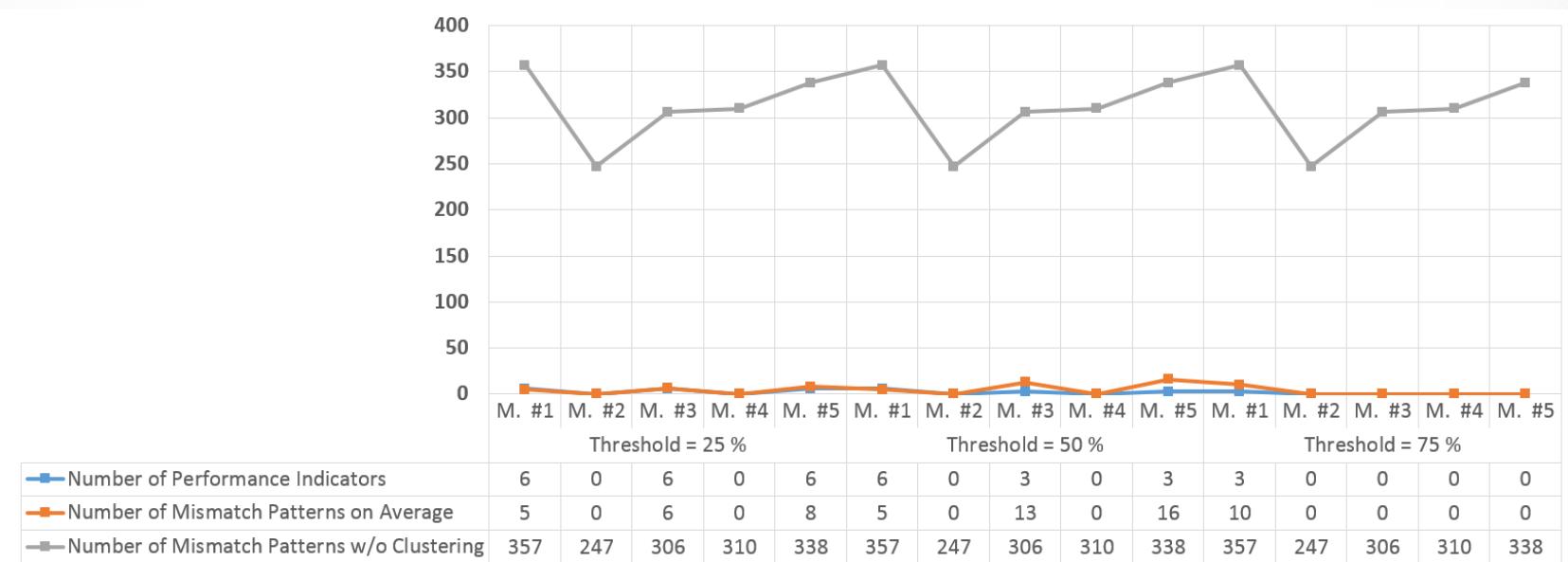


- Number of mismatch patterns to check significantly decreases with performance clustering

Results & Discussions

Environmental Permit Application Process – Recommendation Generation

(with 3 clusters)



- Learning opportunities increases as number of cluster increases
- Number of mismatch patterns to check significantly decreases with performance clustering

Results & Discussions

Discussions



- Process mining stage:
 - Mines the process models with high fitness and high appropriateness from event logs
- Performance indicator calculation stage:
 - With the high quality process models, calculated performance indicators are acceptable
- Clustering stage:
 - Organizations can be clustered based on their performance indicators

Results & Discussions

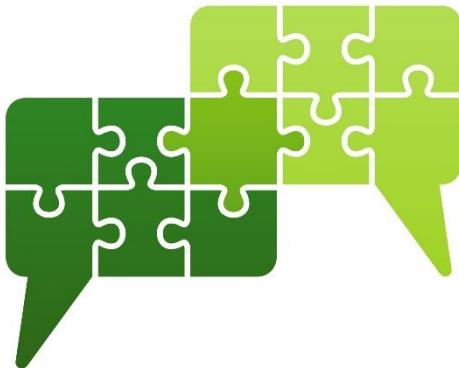
Discussions



- Mismatch analysis stage:
 - It finds differences in accordance with similarity metrics
 - Not all mismatch patterns are discovered
 - Information value of mismatch patterns are not equal
- Recommendation generation stage:
 - Performance clustering helps to focus on differences
 - 3 times more in Loan Application Process
 - 100 times more in Environmental Permit Application Process

Results & Discussions

Discussions



- Business value of generated recommendations:
 - Results may be
 - important or
 - infeasible and irrelevant for business environment
 - Some insights about results can be provided but business environment and case related assessment is also required.

Results & Discussions

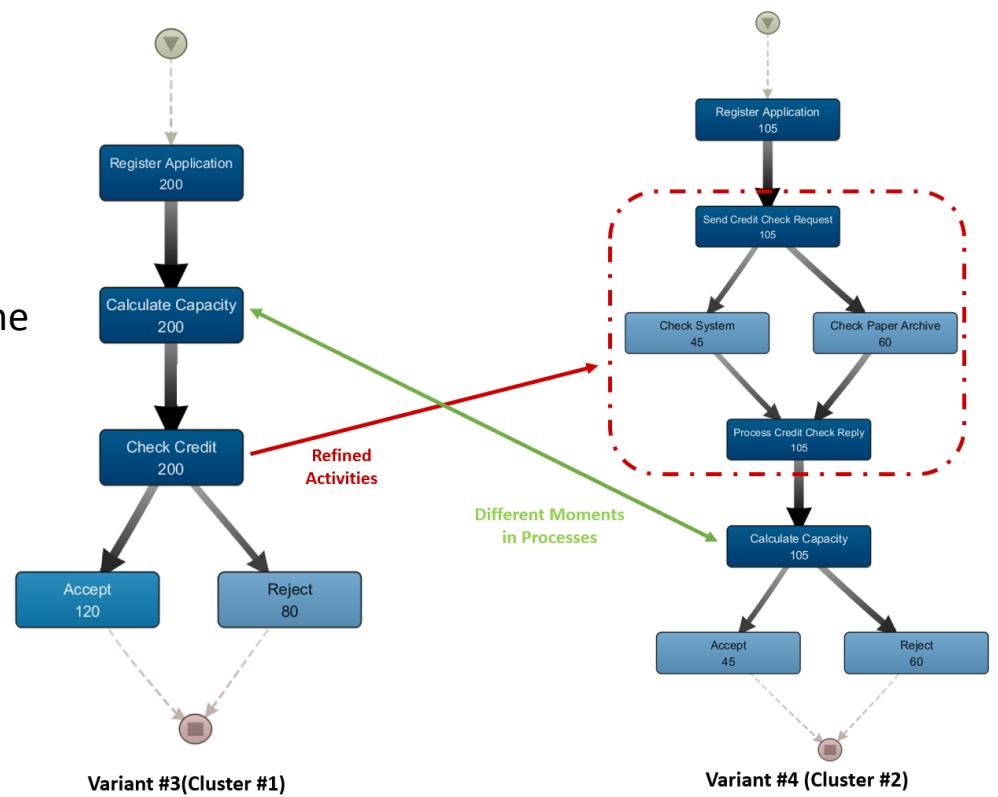
Discussions

In Loan Application Process,
Variant #3 performs worse

- 27 % on average time and
- 12 % on standard deviation time

between activities

Calculate Capacity» → Accept



Results & Discussions

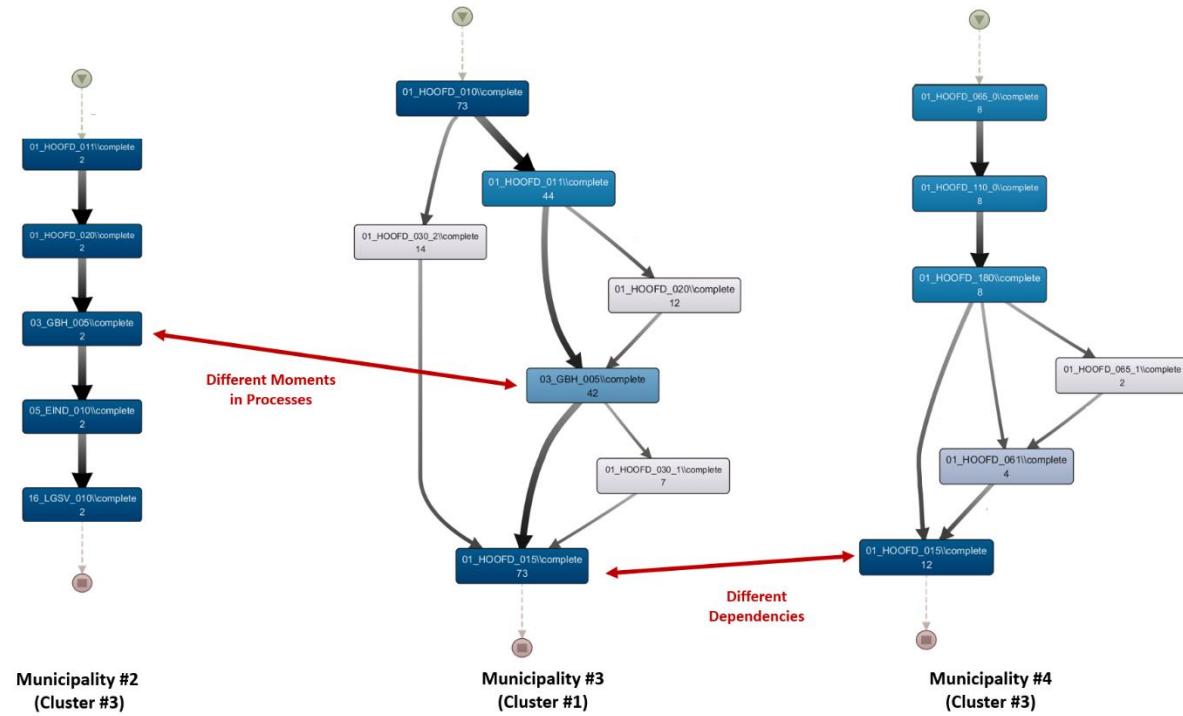
Discussions

In Environmental Permit Application Process,
Cluster #3 performs better

- 40 % on average time and
- 53% on standard deviation time

between activities

**01_HOOFD_010 →
01_HOOFD_015**



Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

Conclusion



- Cross-organizational process mining is applied
 - Unsupervised learning with predictor variables as performances of organizations
 - In an environment where processes are executed on several organizations
- Results show that it is possible to use cross-organizational process mining and mismatch patterns for performance improvement recommendations

Conclusion

- A four-stage solution is presented and their performances are explained



Conclusion

Future Work



- Process mining stage:
 - Different techniques can be used to mine complex process models
- Performance indicator stage:
 - New indicators based on business environment
- Mismatch pattern analysis:
 - New patterns can be introduced

Conclusion

Future Work

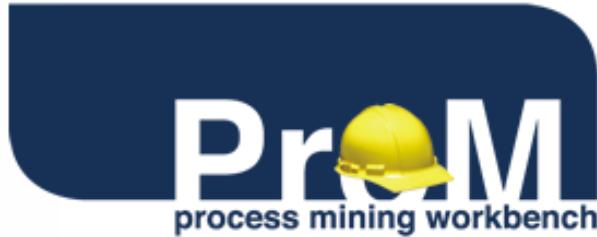


- Recommendation generation:
 - Domain or BPM expertise to assess the quality of recommendations
- ProM implementation:
 - Visually selecting the interested area of process models

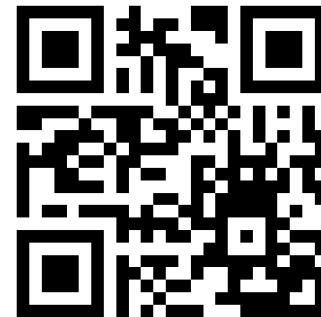
Agenda

- 
- 1 Introduction
 - 2 Related Work
 - 3 Background
 - 4 Methodology
 - 5 Results & Discussions
 - 6 Conclusion & Future Work
 - 7 Demonstration

Demonstration



Also available on YouTube
<https://youtu.be/T92UrRfl3r0>



References

- [1] W. M. P. van der Aalst, A. Adriansyah, A.K.A. de Medeiros, and et al. Process mining manifesto. In Business process management workshops, pages 169–194. Springer, 2012.
- [2] J. C. Buijs, B. F. van Dongen, and W. M. P. van der Aalst. Towards crossorganizational process mining in collections of process models and their executions. In Business Process Management Workshops, pages 2–13. Springer, 2012.
- [3] W. M. P. van der Aalst. Intra-and inter-organizational process mining: Discovering processes within and between organizations. In The Practice of Enterprise Modeling, pages 1–11. Springer, 2011.
- [4] W. M. P. van der Aalst. Process mining: discovery, conformance and enhancement of business processes. Springer Science & Business Media, 2011.
- [5] A. K. A. de Medeiros, B. F. van Dongen,W. M. P. van der Aalst, and A. J. M. M. Weijters. Process mining: Extending the -algorithm to mine short loops. 2004.
- [6] W. M. P. van der Aalst, T.Weijters, and L. Maruster. Workflow mining: Discovering process models from event logs. Knowledge and Data Engineering, IEEE Transactions on, 16(9):1128–1142, 2004.
- [7] J. Herbst. Dealing with concurrency in workflow induction. In European Concurrent Engineering Conference. SCS Europe. Citeseer, 2000.

References

- [8] J. Herbst and D. Karagiannis. Integrating machine learning and workflow management to support acquisition and adaptation of workflow models. In Database and Expert Systems Applications, 1998. Proceedings. Ninth International Workshop on, pages 745–752. IEEE, 1998.
- [9] G. Greco, A. Guzzo, and L. Pontieri. Mining hierarchies of models: From abstract views to concrete specifications. In Business Process Management, pages 32–47. Springer, 2005.
- [10] W. M. P. van der Aalst, A.K.A. de Medeiros, and A. J. M. M. Weijters. Genetic process mining. In Applications and theory of Petri nets 2005, pages 48–69. Springer, 2005.
- [11] E. Esgin, P. Senkul, and C. Cimenbicer. A hybrid approach for process mining: using from-to chart arranged by genetic algorithms. In Hybrid Artificial Intelligence Systems, pages 178–186. Springer, 2010.
- [12] E. Esgin and R. Senkul. A hybrid approach to process mining: Finding immediate successors of a process by using from-to chart. In Machine Learning and Applications, 2009. ICMLA'09. International Conference on, pages 664–668. IEEE, 2009.
- [13] W. M. P. van der Aalst. Business process configuration in the cloud: how to support and analyze multi-tenant processes? In Web Services (ECOWS), 2011 Ninth IEEE European Conference on, pages 3–10. IEEE, 2011.

References

- [14] R. Dijkman. Mismatch Patterns in Similar Business Processes. Beta, Research School for Operations Management and Logistics, 2007.
- [15] R. Dijkman, M. Dumas, B. van Dongen, R. Käärik, and J. Mendling. Similarity of business process models: Metrics and evaluation. *Information Systems*, 36(2):498–516, 2011.
- [16] J. C. Buijs and H. A. Reijers. Comparing business process variants using models and event logs. In *Enterprise, Business-Process and Information Systems Modeling*, pages 154–168. Springer, 2014.
- [17] E. Esgin and P. Karagoz. Sequence alignment adaptation for process diagnostics and delta analysis. In *Hybrid Artificial Intelligent Systems*, pages 191–201. Springer, 2013.
- [18] E. Esgin and P. Senkul. Delta analysis: a hybrid quantitative approach for measuring discrepancies between business process models. In *Hybrid Artificial Intelligent Systems*, pages 296–304. Springer, 2011.
- [19] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst. Discovering blockstructured process models from event logs-a constructive approach. In *Application and Theory of Petri Nets and Concurrency*, pages 311–329. Springer, 2013.
- [20] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst. Discovering block-structured process models from event logs containing infrequent behaviour. In *Business Process Management Workshops*, pages 66–78. Springer, 2014.

References

- [21] W. M. P. van der Aalst, A. Adriansyah, and B. van Dongen. Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2):182–192, 2012.
- [22] D. Arthur and S. Vassilvitskii. k-means++: the advantages of carefull seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 1027–1035, 2007.
- [23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. SIGKDD Explor. Newsl., 11(1):10–18, November 2009.
- [24] H. M. W. Verbeek, J. C Buijs, B. F. van Dongen, and W. M. P. van der Aalst. Prom 6: The process mining toolkit. Proc. of BPM Demonstration Track, 615:34–39, 2010.
- [25] A. Rozinat and W. M. P. van der Aalst. Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1):64–95, 2008.
- [26] J. C. Buijs. Loan application example, 2013.
- [27] J.C. Buijs. Environmental permit application process ('wabo'), coselog project, 2014.
- [28] J. C. Buijs. Flexible Evolutionary Algorithms for Mining Structured Process Models. PhD thesis, PhD thesis. Eindhoven, The Netherlands: Technische Universiteit Eindhoven, 2014 (cit. on p. 179), 2014.

Questions & Comments



Thank you for your
attention!



Recommendation Generation for Performance Improvement by using Cross-Organizational Process Mining

Onur Yılmaz

Supervisor: Assoc. Prof. Pınar Karagöz

September 1, 2015