



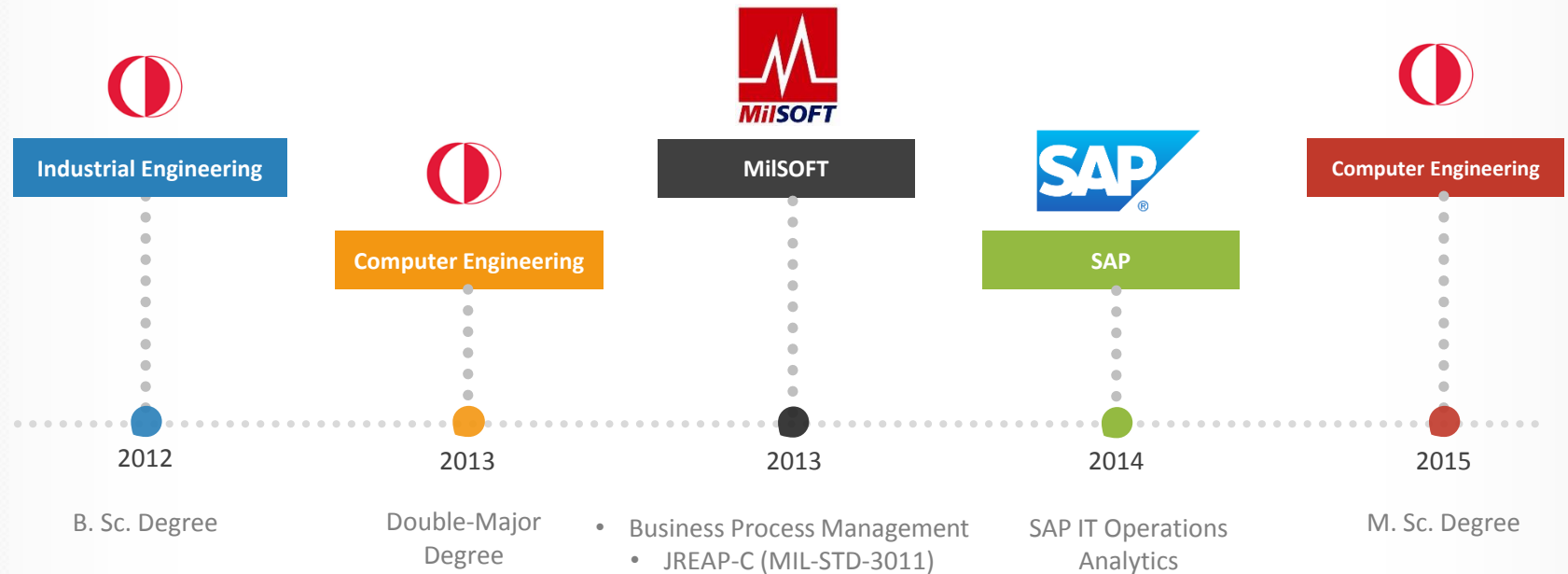
Recommendation Generation for Performance Improvement by using Cross-Organizational Process Mining

Onur Yılmaz

Supervisor: Assoc. Prof. Pınar Karagöz

September 1, 2015

Onur Yilmaz



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.



Events recorded and available



Competitive business life

Introduction

Cross- Organizational Process Mining



Cloud computing
and shared
infrastructures

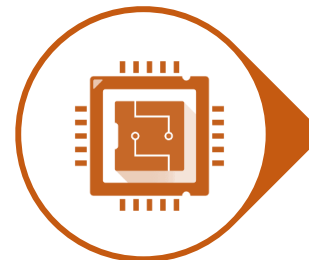


Event logs of
multiple
organizations



**Analyze the big
picture**

Work together to execute
the same process



**Learn from
each other**

Execute the same task
on shared infrastructure

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

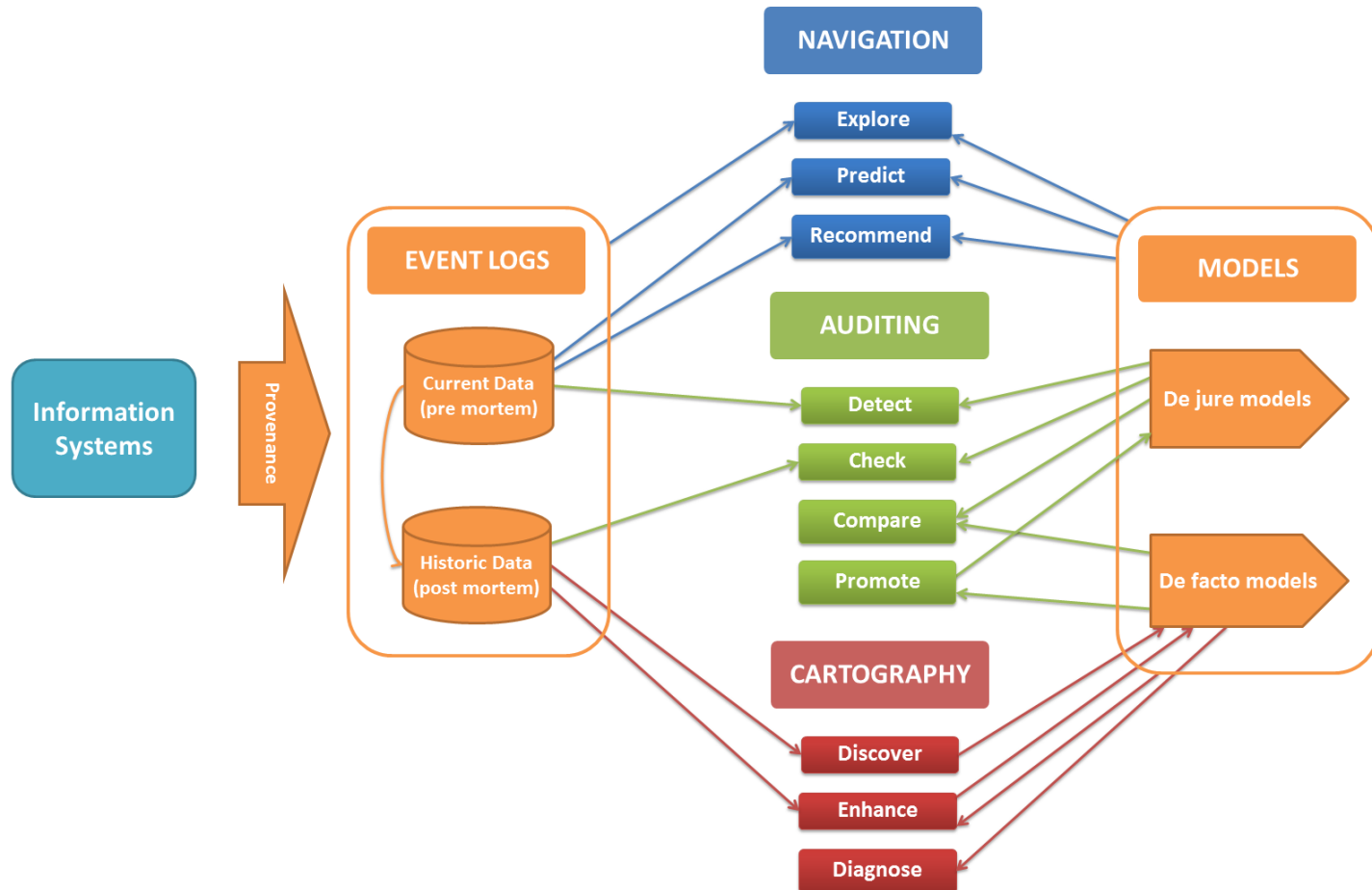
Conclusion & Future Work

7

Demonstration

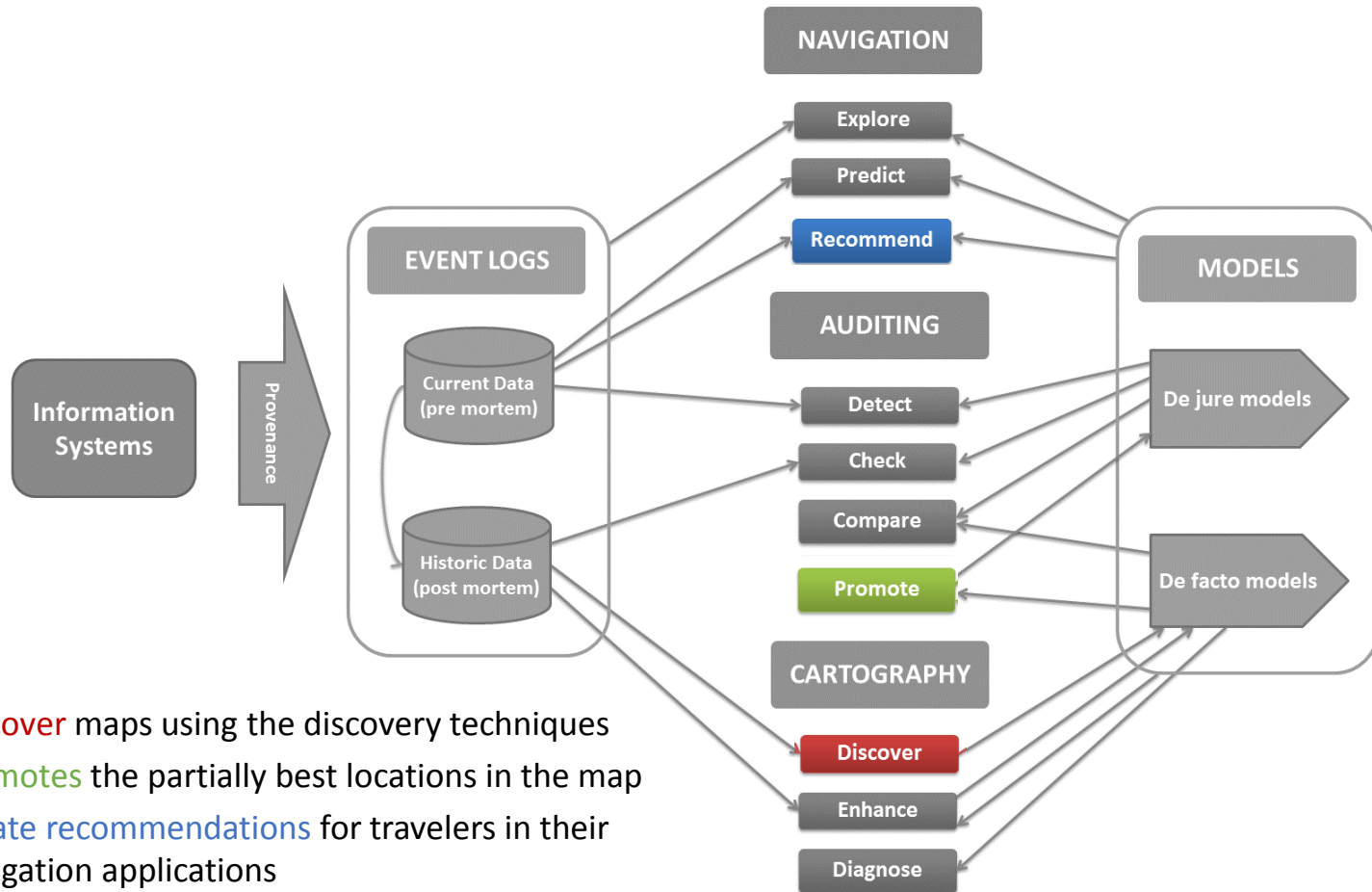
Related Work

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
 - Inductive Approach
 - Hierarchical Clustering
 - Genetic Approach
 - Heuristic Approach

**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
- Configurable process models for organizations
- 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
- Repository of process models of similar business operations
- Methods in the literature
 - Similarity Metrics
 - Alignment Matrix
 - Delta Analysis
 - Mismatch Patterns

**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

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

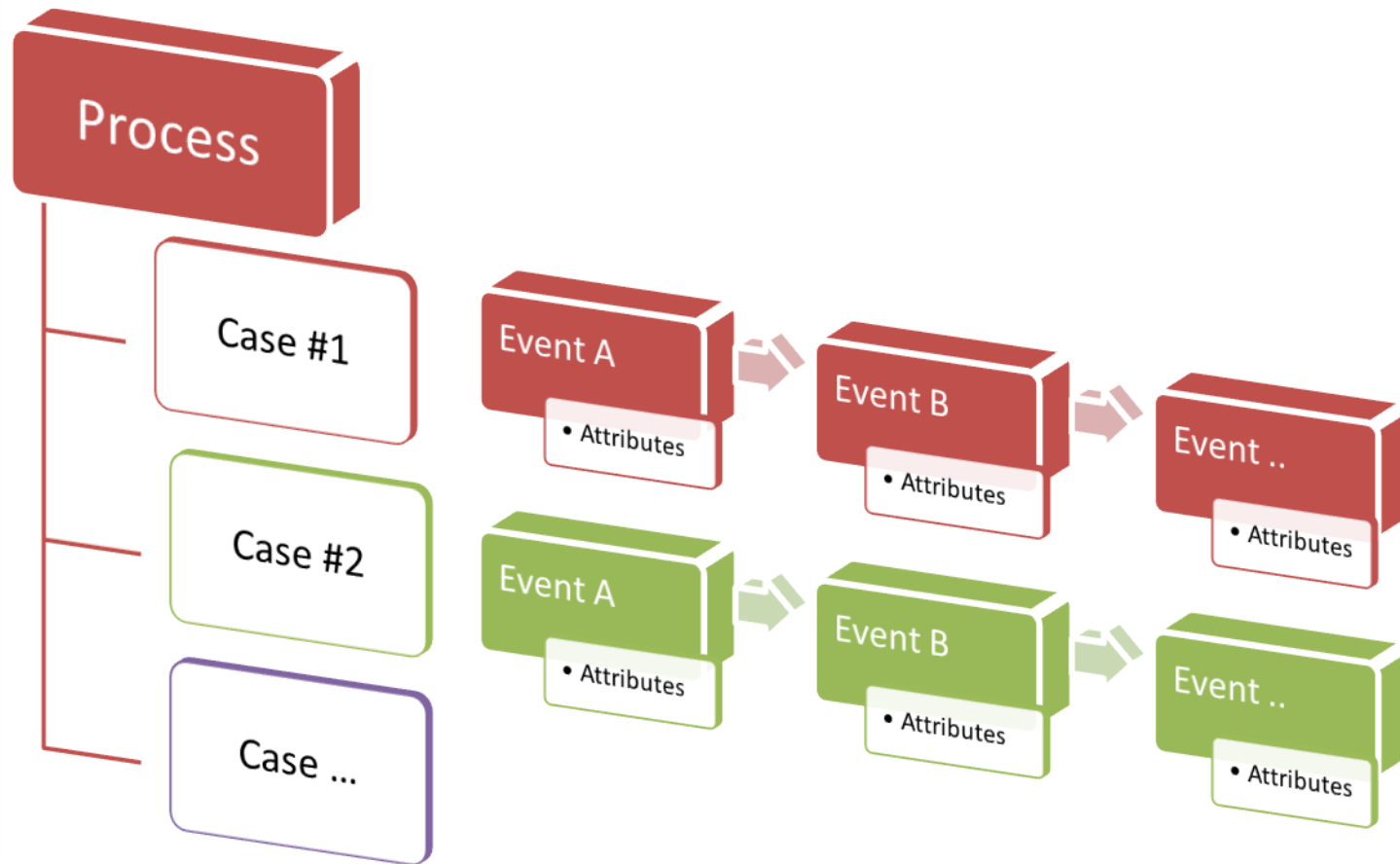
- Outputs of the software systems like Enterprise Resource Planning (ERP) or Business Process Management (BPM)



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

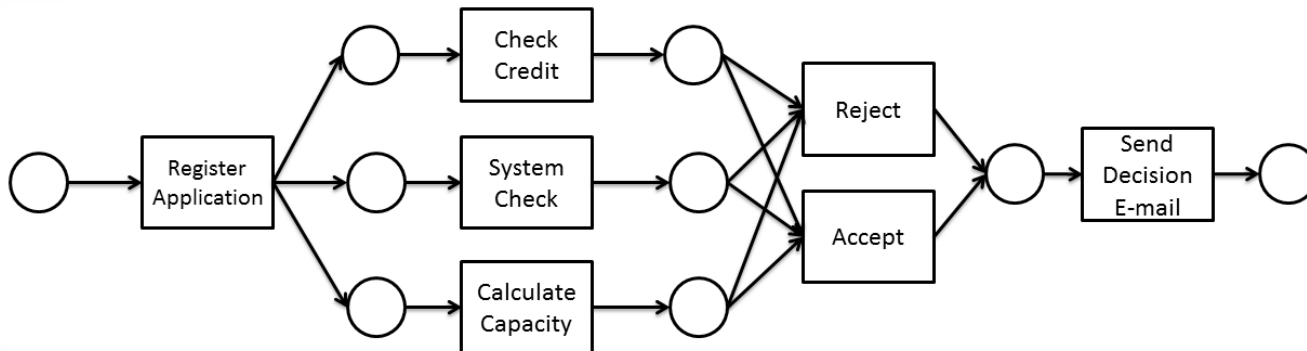
Event Log



Background

Process Modeling

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

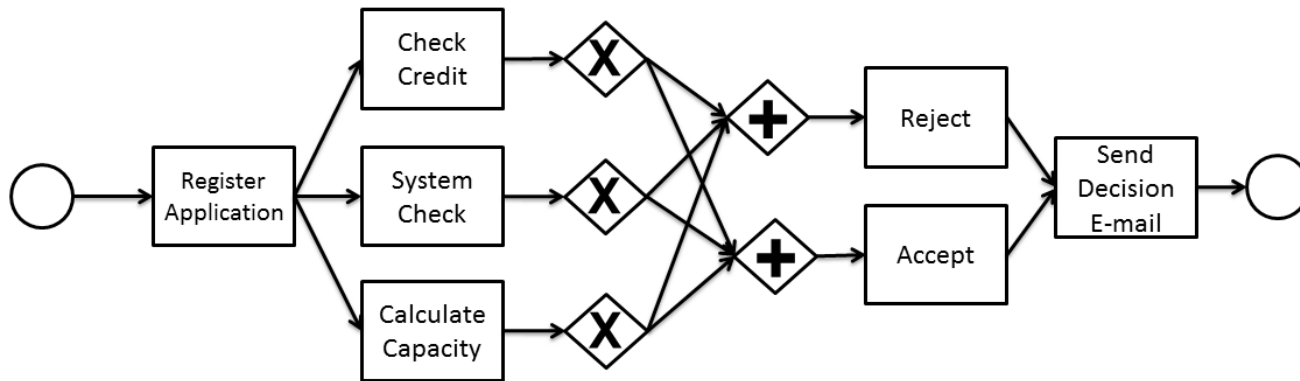


- Mathematical background

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

Background

Process Discovery



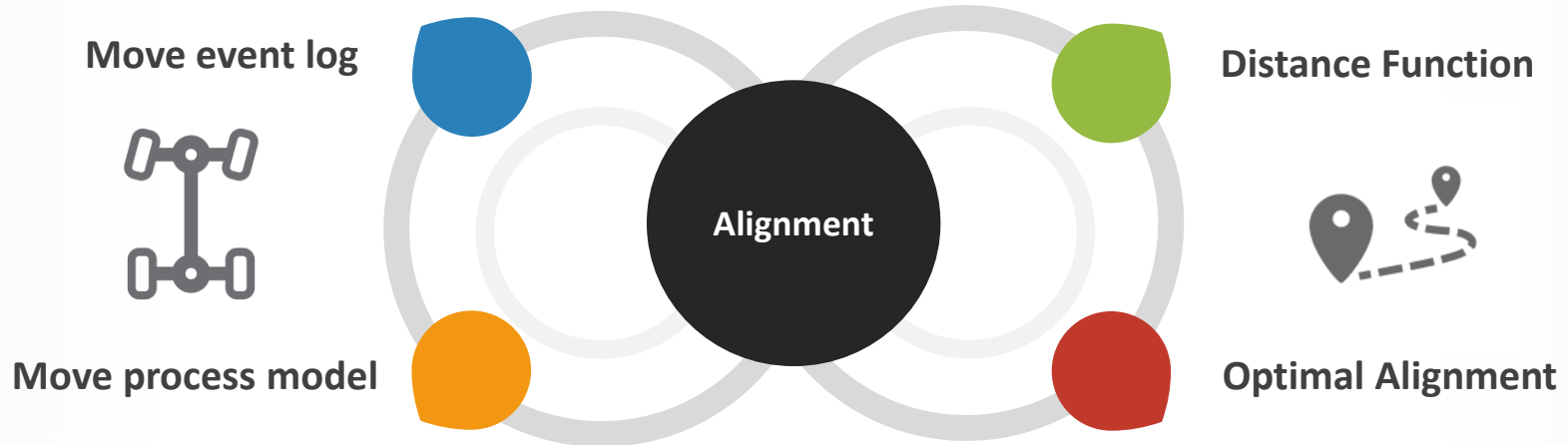
- **Inductive Miner Infrequent (IMi)** is proposed as an extension to Inductive Miner to handle noise in the event logs:
 1. **Activity Sets:** Split the activities in log to disjoint sets.
 2. **Sublogs:** Split the log by using activity sets.
 3. **Recursive Mining:** Mine sublogs with these steps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

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



Background

Clustering

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



Background

Mismatch Patterns in Process Models



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



Authorization



Activity

- Skipped Activity
- Refined Activity



Control Flow

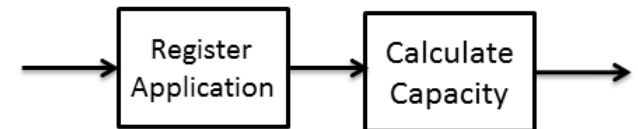
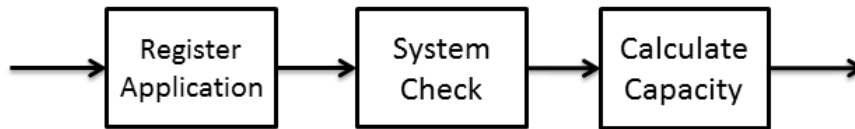
- Activities at Different Moments in Processes
- Different Conditions for Occurrence
- Different Dependencies
- Additional Dependencies

Background

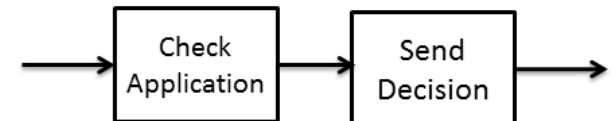
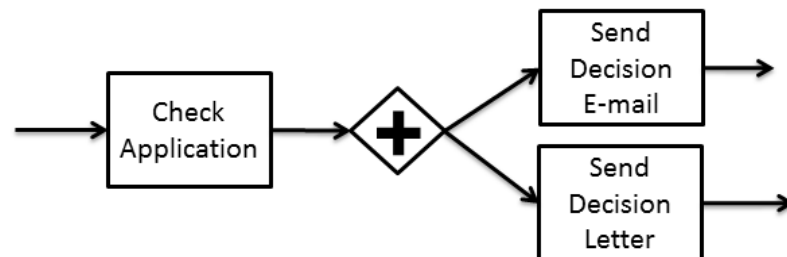
Mismatch Patterns in Process Models

Activity Mismatch Patterns

- Skipped Activity



- Refined Activity



Background

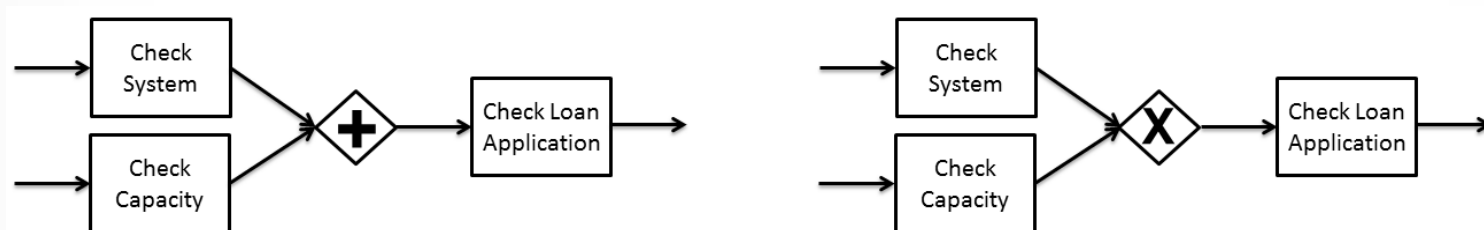
Mismatch Patterns in Process Models

Control Flow Mismatch Patterns

- Activities at Different Moments in Processes



- Different Conditions for Occurrence

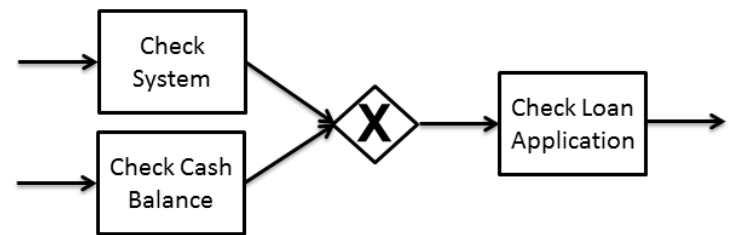
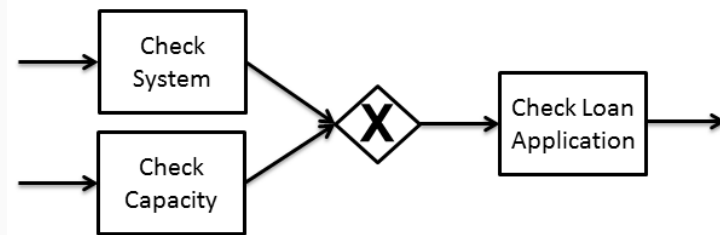


Background

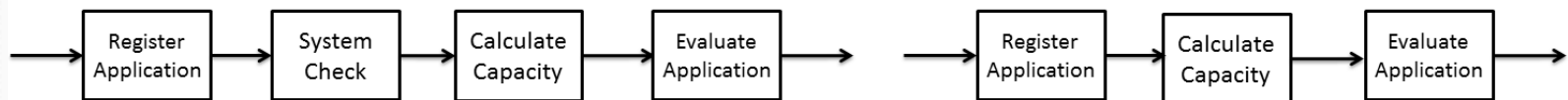
Mismatch Patterns in Process Models

Control Flow Mismatch Patterns

- Different Dependencies



- Additional Dependencies



Agenda

1

Introduction

2

Related Work

3

Background

4

Methodology

5

Results & Discussions

6

Conclusion & Future Work

7

Demonstration

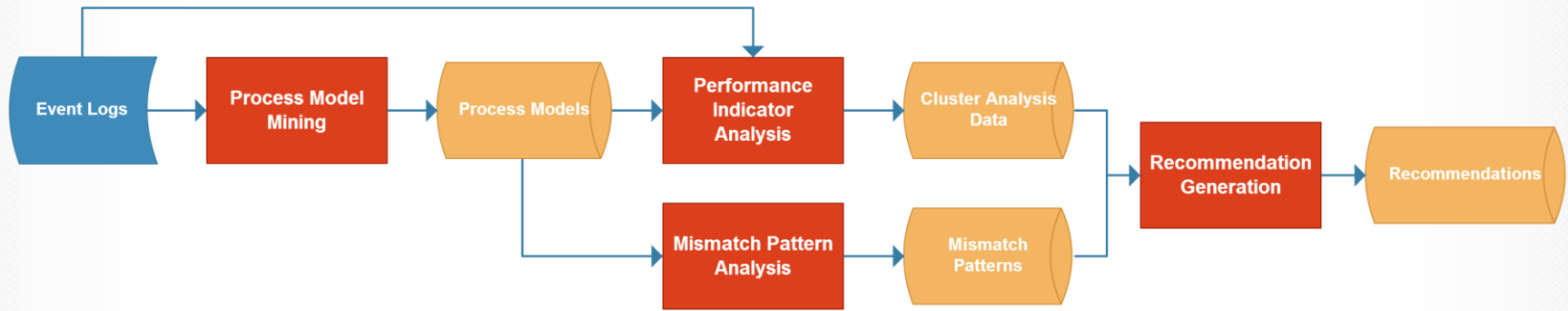
Methodology

Approach Overview



Methodology

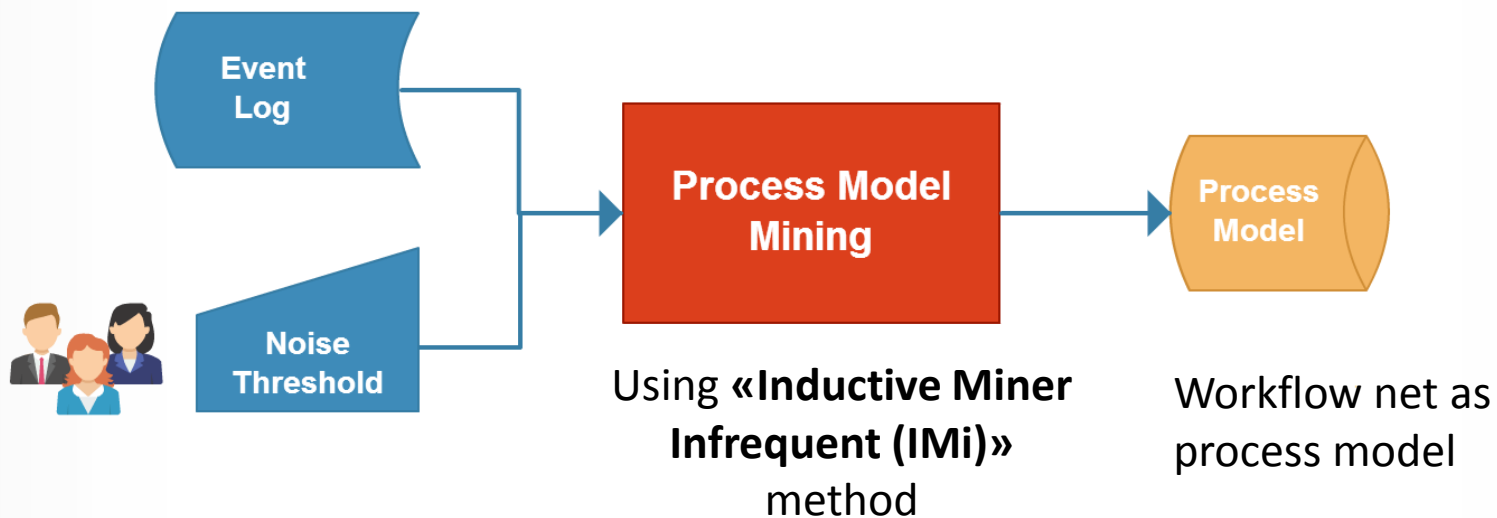
Approach Overview



Methodology

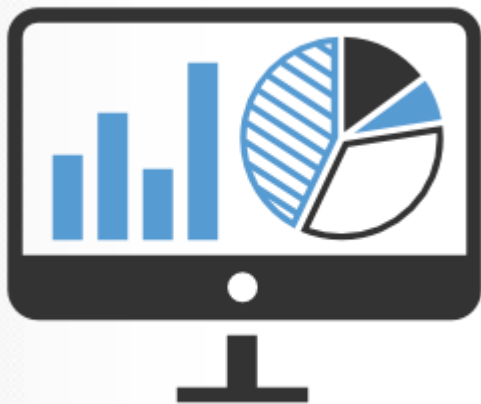
Process Model Mining

Applied for each organization:

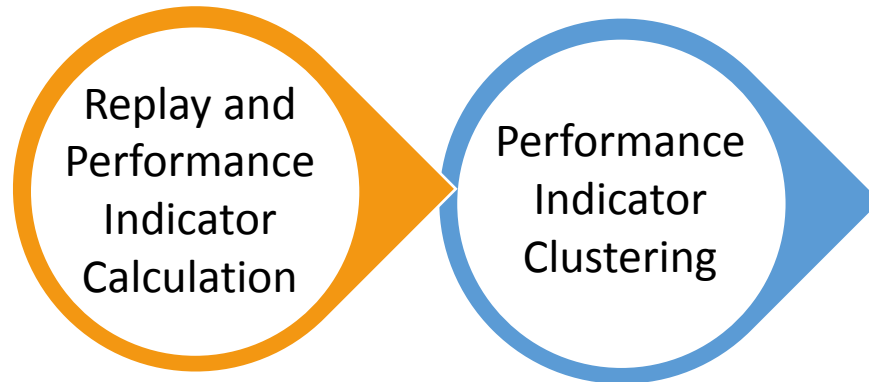


Methodology

Performance Indicator Analysis



- Two steps:



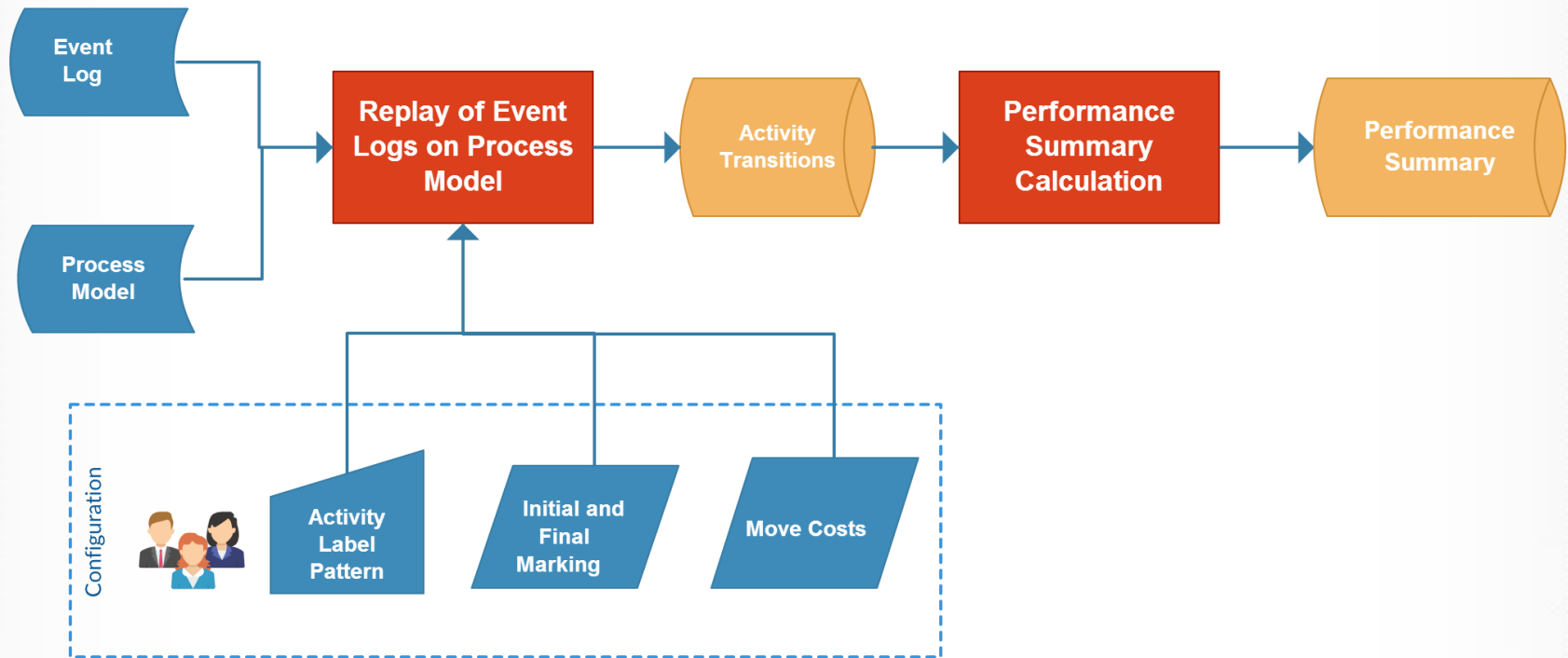
- Performance Indicators:
 - Average Time Between Activities
 - Standard Deviation of Time Between Activities



Methodology

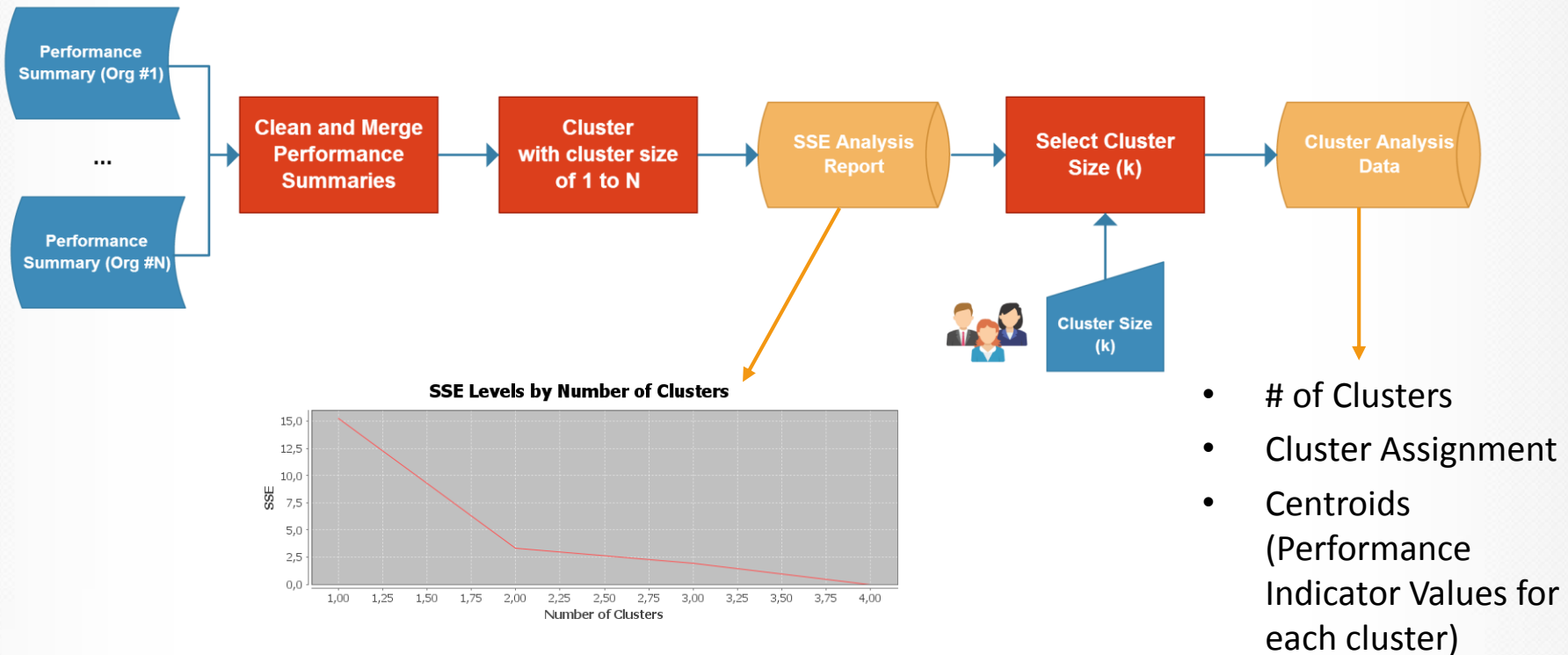
Performance Indicator Analysis - Replay and Performance Indicator Calculation

Applied for each organization:



Methodology

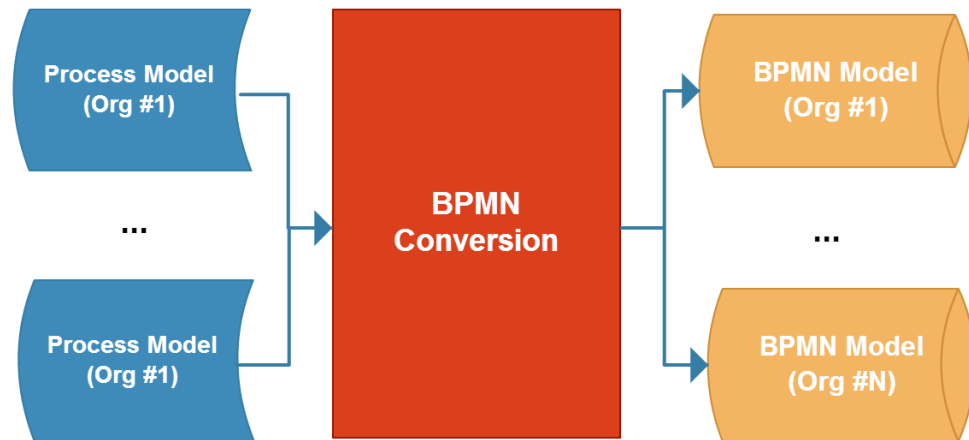
Performance Indicator Analysis – Performance Indicator Clustering



Methodology

Mismatch Pattern Analysis

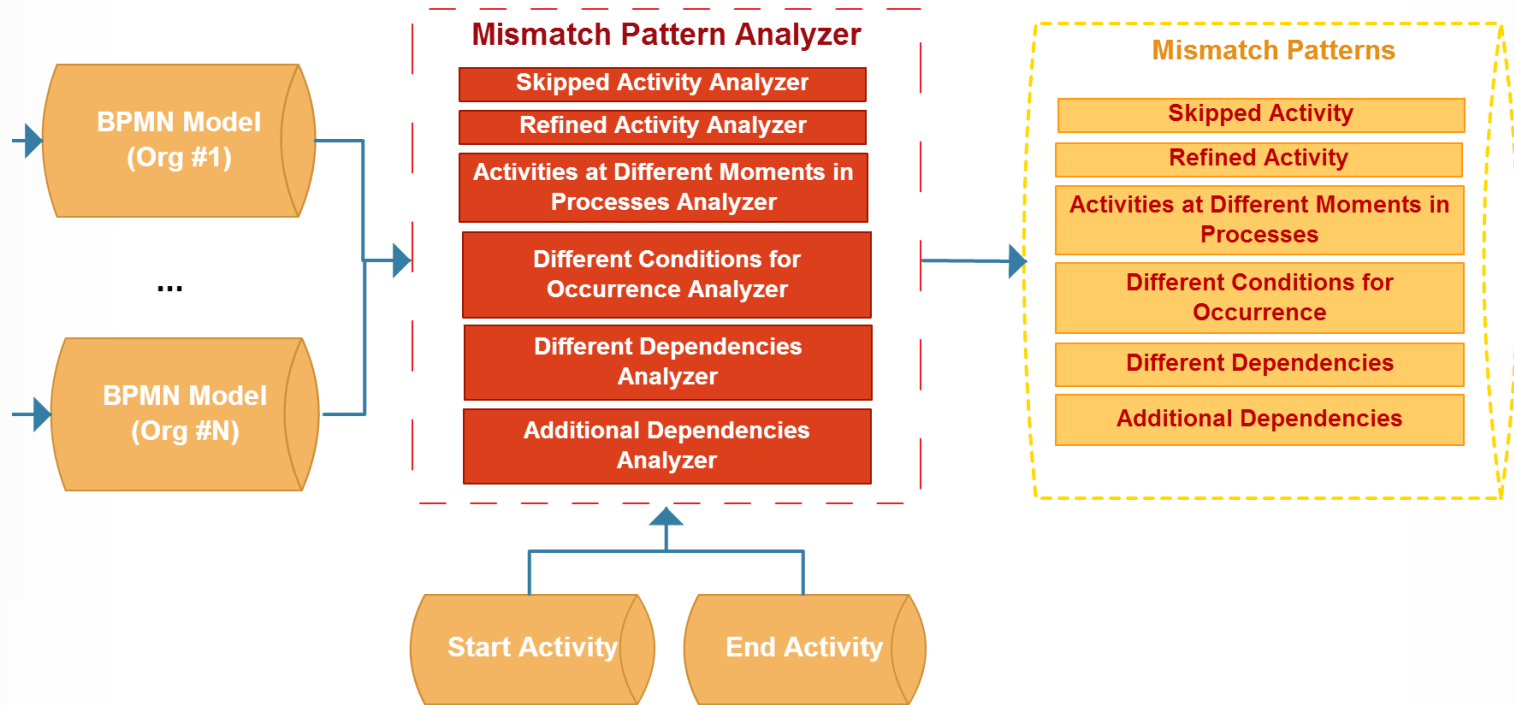
- Spot the differences between process models of different organizations as mismatch patterns
- BPMN used since notation is more appropriate to formulate mismatch patterns



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

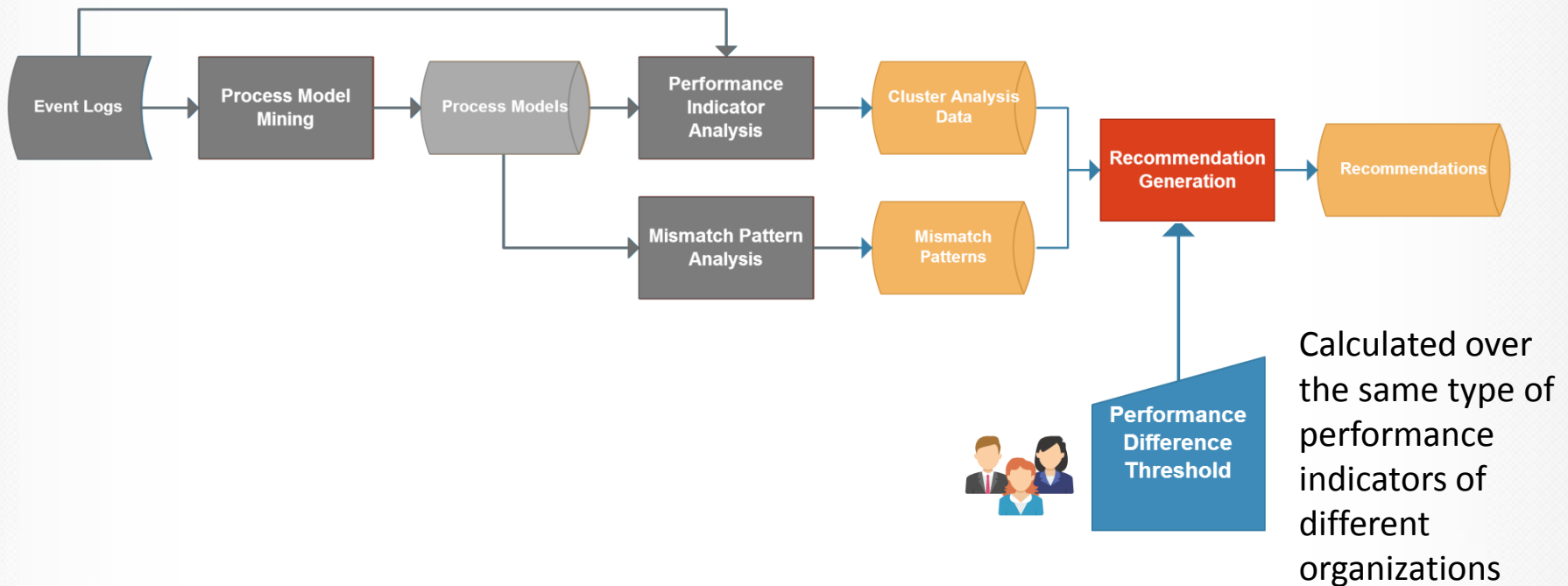


- Providing a set of mismatch patterns for each organization to enhance their processes.
 - Mismatch patterns between organizations, which are **performing better** in terms of their performance indicator values.

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 \leftarrow {}

i \leftarrow Cluster of organization **O**

for each centroid for cluster **i**

Get other cluster **j** with the centroids of A_{start} and A_{end} ; and value difference larger than **P**

for each organization **O'** in the cluster **j**

MismatchPatterns \leftarrow Mismatch Pattern Analysis(**O**, **O'**, A_{start} , A_{end})

Recommendations \leftarrow Recommendation(**O**, A_{start} , A_{end} , **MismatchPatterns**)

Methodology

ProM Implementation



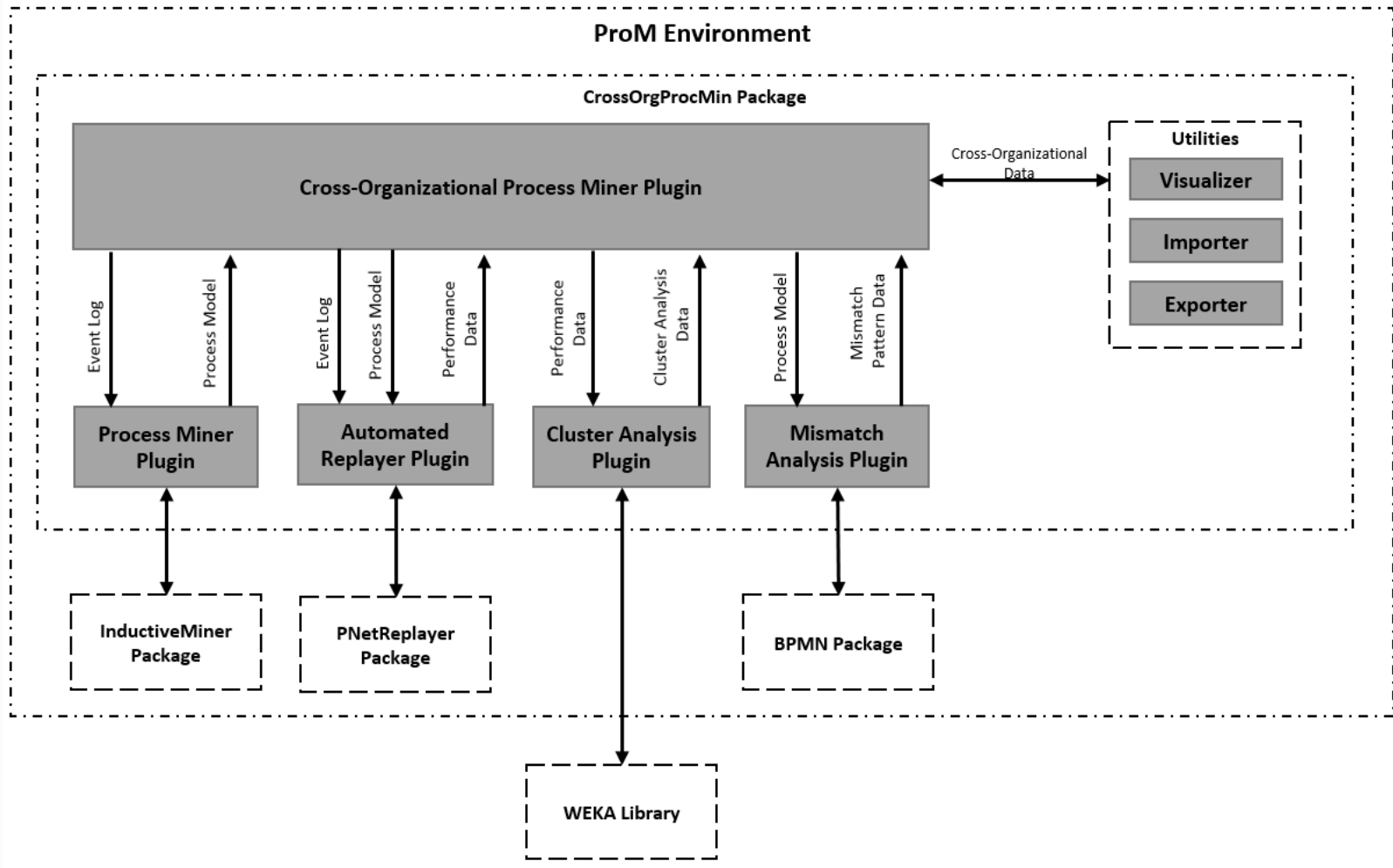
- ProM, extensible framework that supports a wide variety of process mining techniques in form of plugins
- Widely accepted in industry and academia with an active community
- Developed set of plugins are packaged with the name of «**CrossOrgProcMin**» and published on Github.



onuryilmaz/cross-orc-proc-min

Methodology

ProM Implementation



Agenda

1

Introduction

2

Related Work

3

Background

4

Methodology

5

Results & Discussions

6

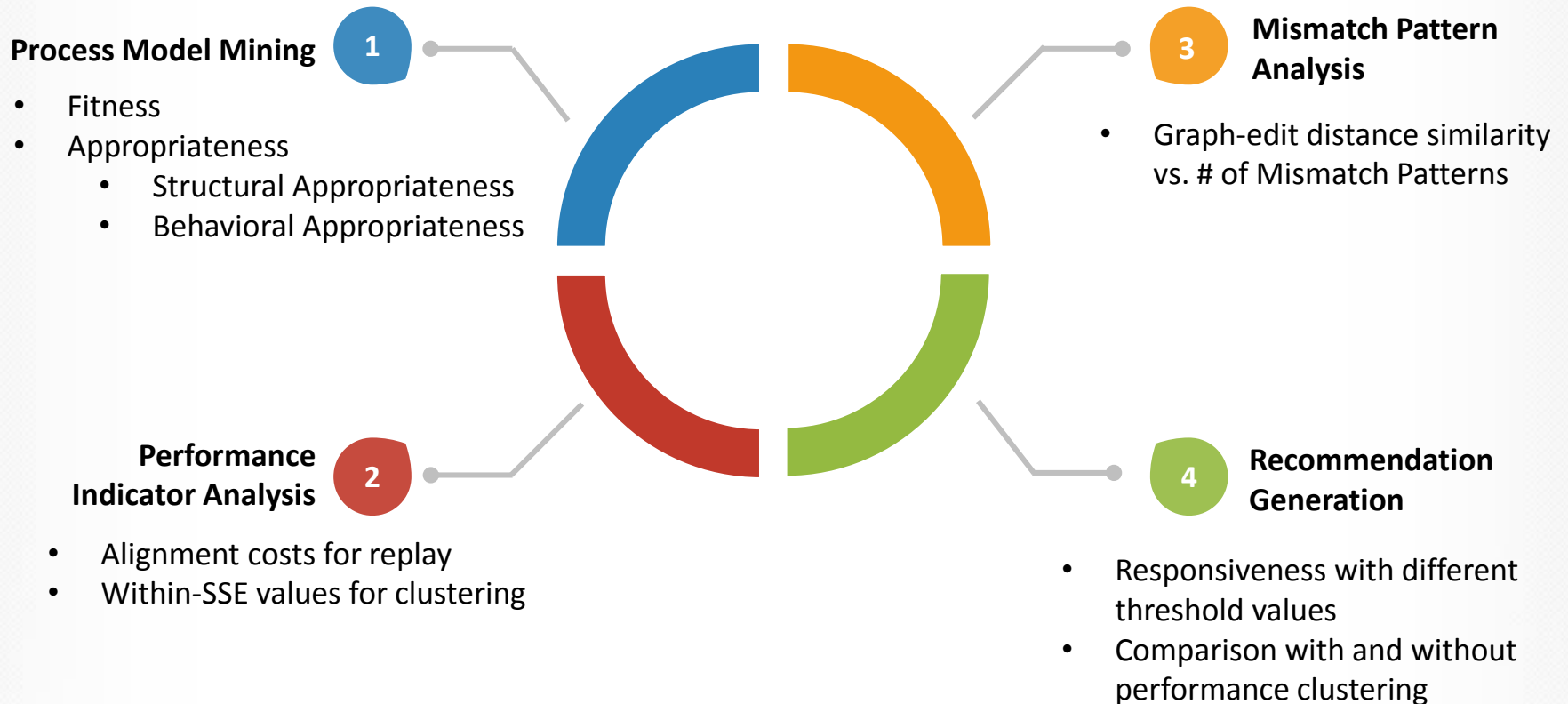
Conclusion & Future Work

7

Demonstration

Results & Discussions

Evaluation Metrics



Results & Discussions

Dataset Selection



- **Loan Application Process**
 - Synthetically generated
 - 4 variants of a simple loan application in a financial institute
- **Environmental Permit Application Process**
 - Real-life event log from "Configurable Services for Local Governments (CoSeLoG)" project
 - «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
475

Events
2440

- These variants are used as organizational logs

Results & Discussions

Loan Application Process – Process Model Mining

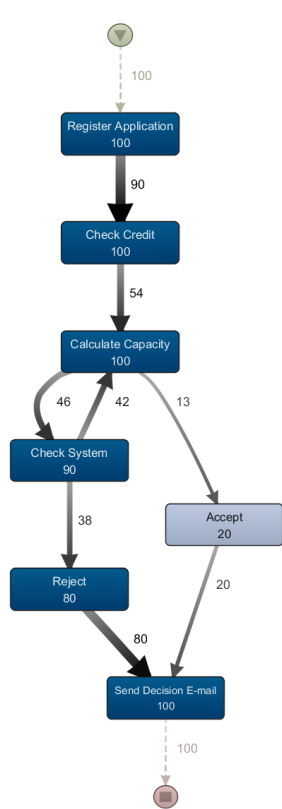
- With «0» noise threshold

	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 %

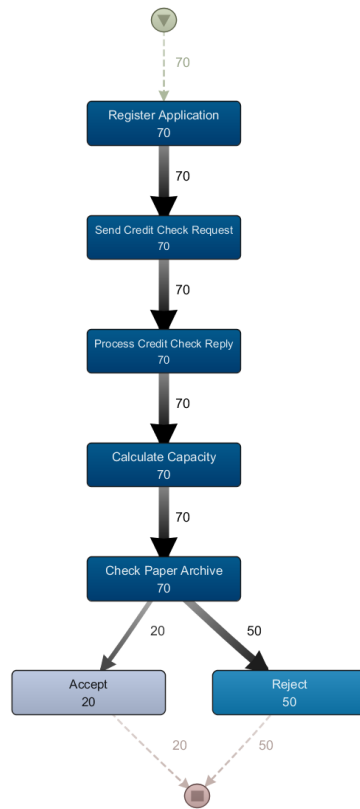
- Each process model is fitting to event log and «appropriate»

Results & Discussions

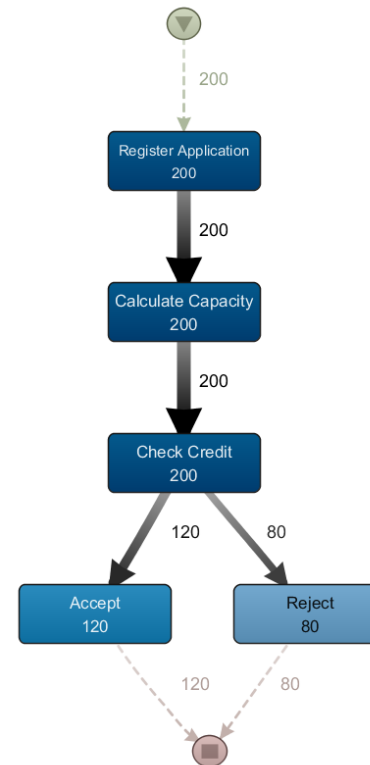
Loan Application Process – Process Model Mining



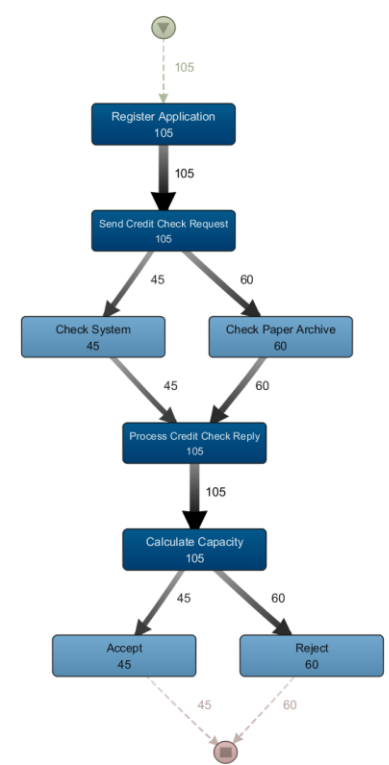
Variant #1



Variant #2



Variant #3

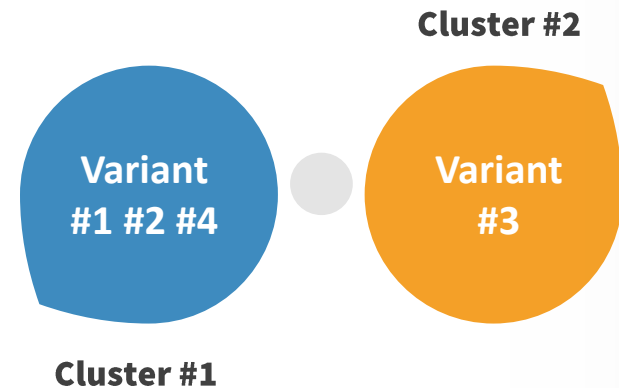
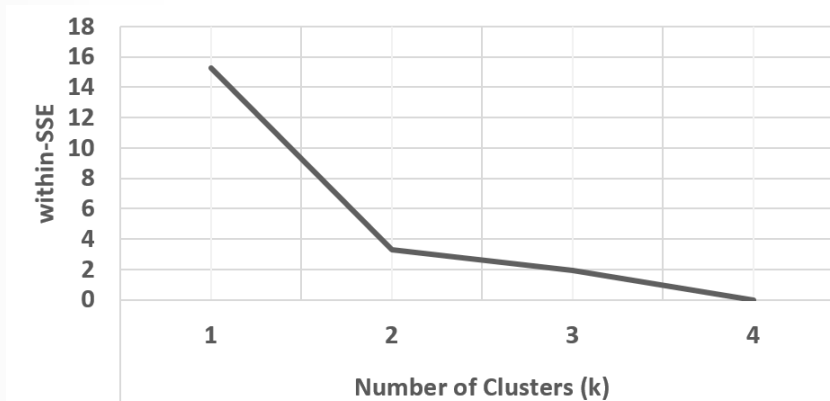


Variant #4

Results & Discussions

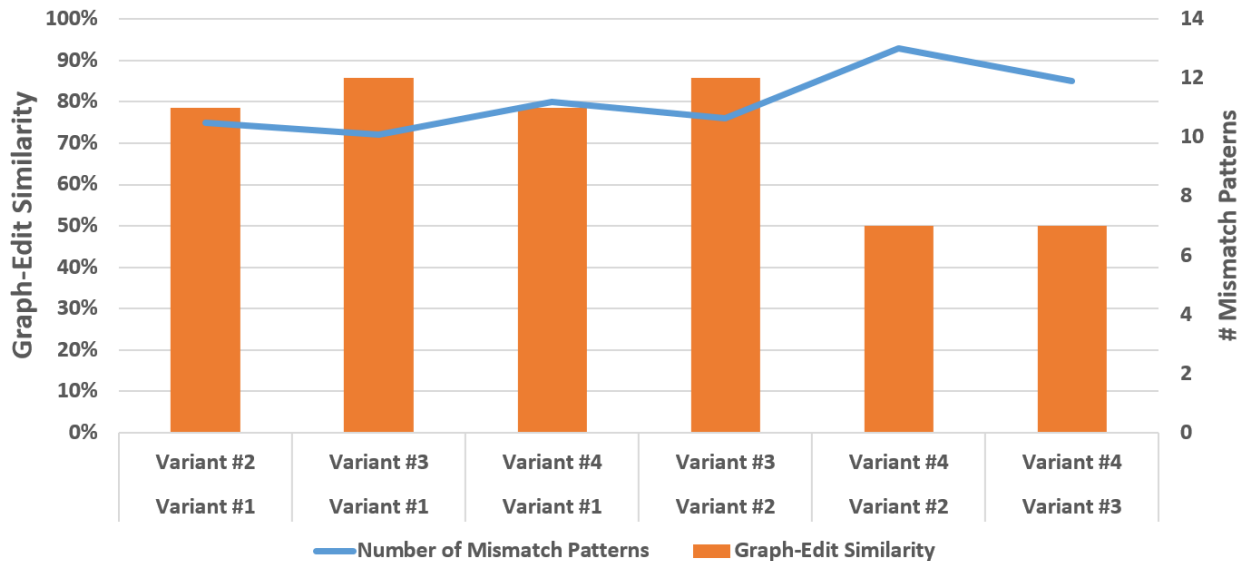
Loan Application Process – Performance Indicator Analysis

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



Results & Discussions

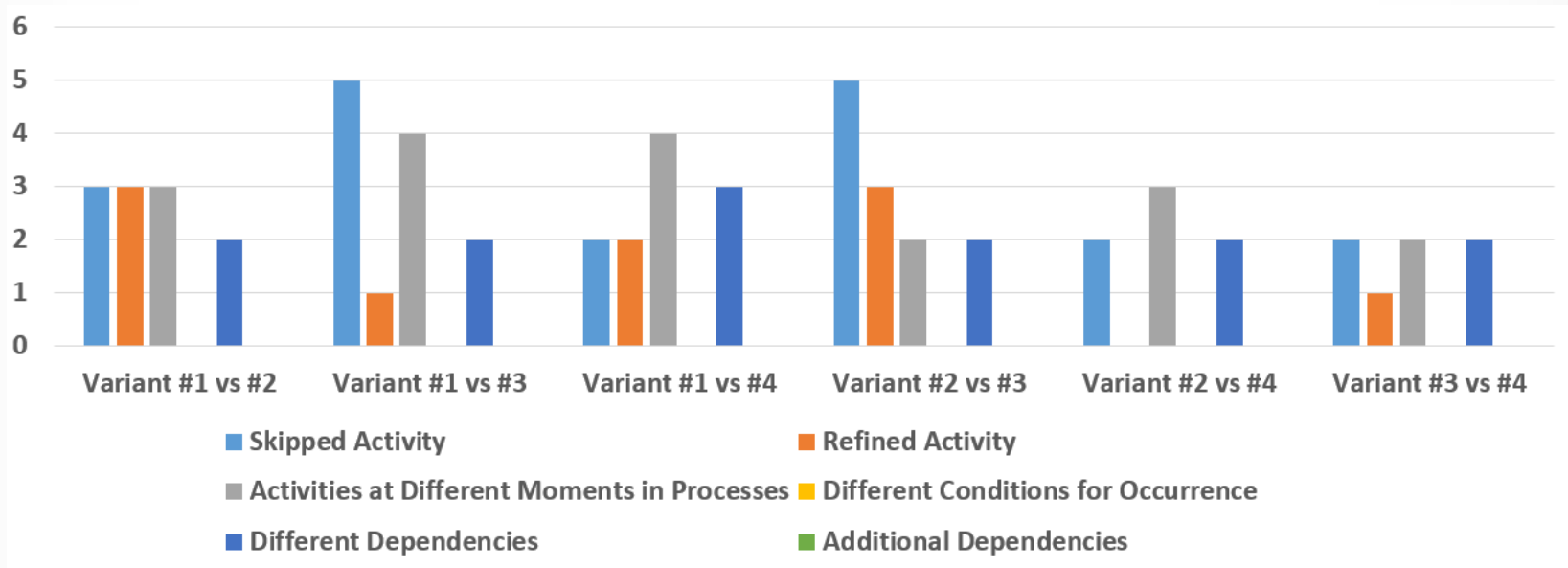
Loan Application Process – Mismatch Pattern Analysis



Correlation between graph-edit similarity and number of mismatch patterns

Results & Discussions

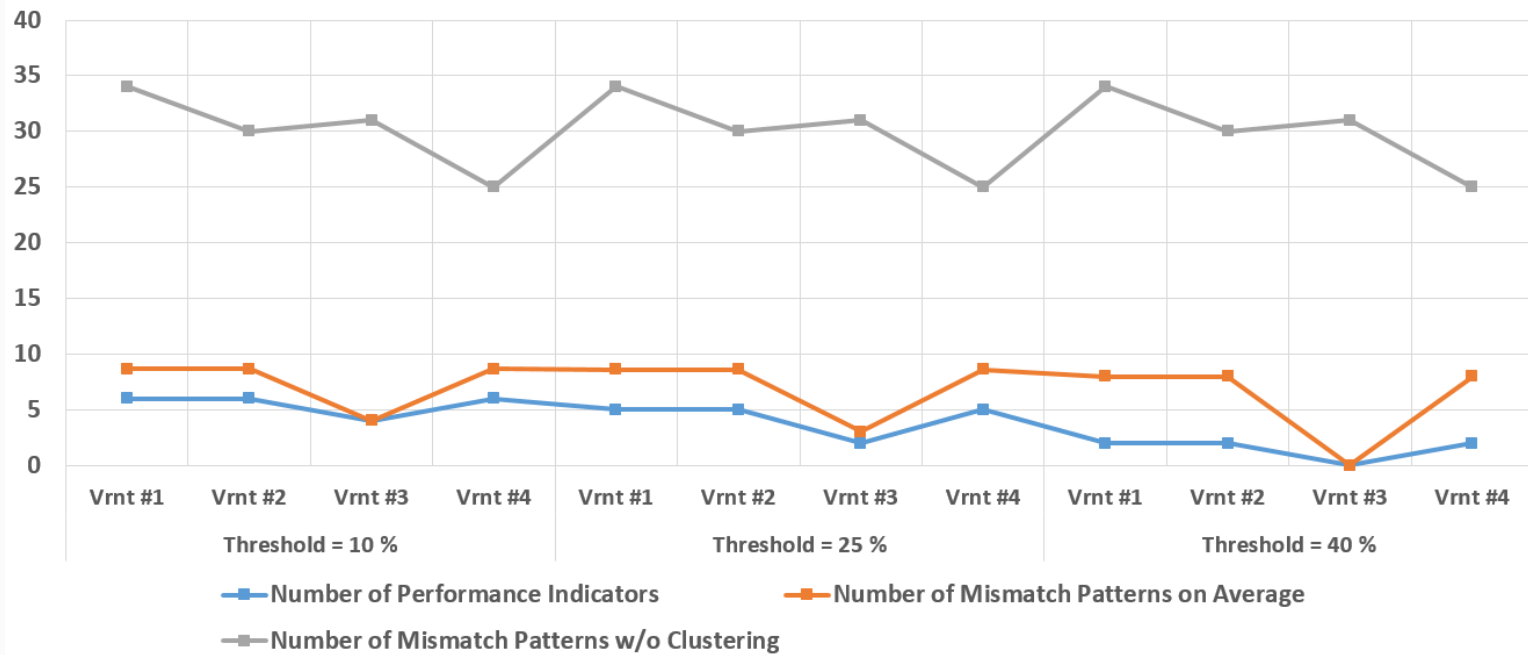
Loan Application Process – Mismatch Pattern Analysis



- "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 1214 Events 2142

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

Results & Discussions

Environmental Permit Application Process – Process Model Mining

- With 10 % noise threshold

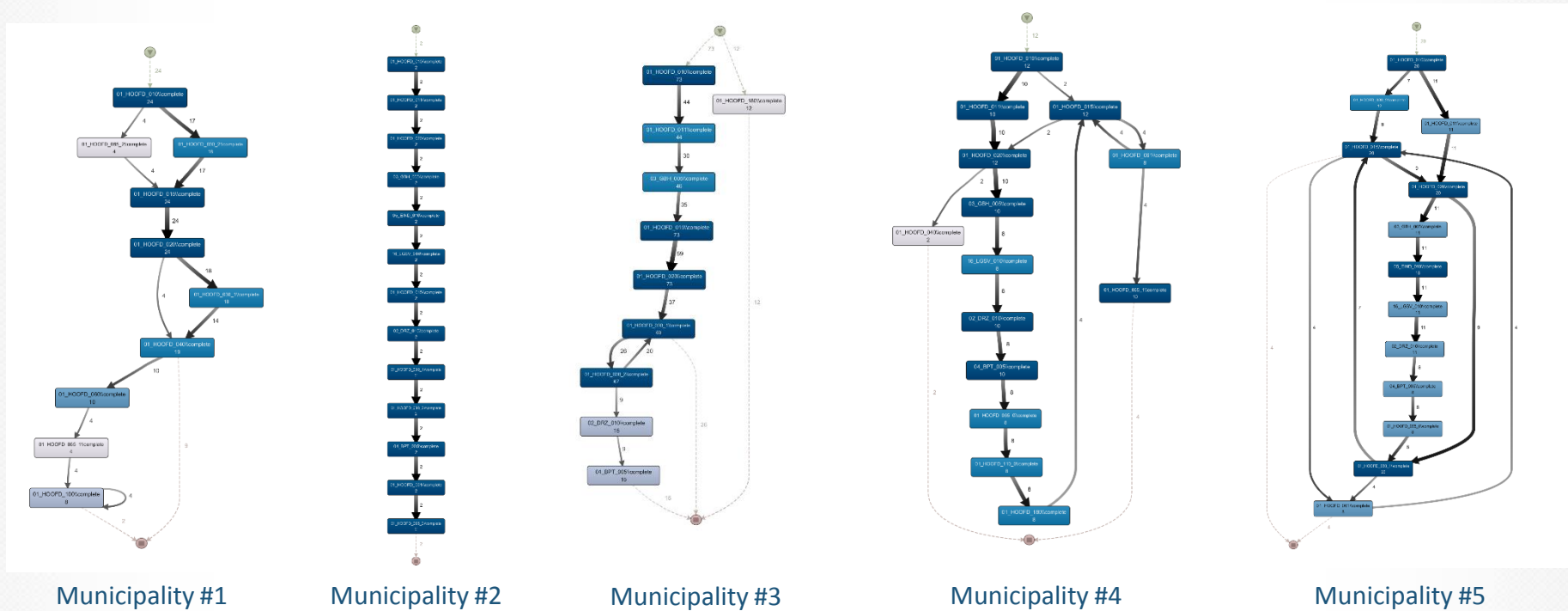
	Fitness	Structural Appropriateness	Behavioral Appropriateness	Average Appropriateness
Municipality #1	86 %	97.5 %	54.4 %	76 %
Municipality #2	100 %	100 %	100 %	100 %
Municipality #3	92.3 %	71.1 %	67.2 %	69.1 %
Municipality #4	96.8 %	65.7 %	64 %	64.9 %
Municipality #5	94.5 %	58.8 %	39.7 %	49.3 %
Average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

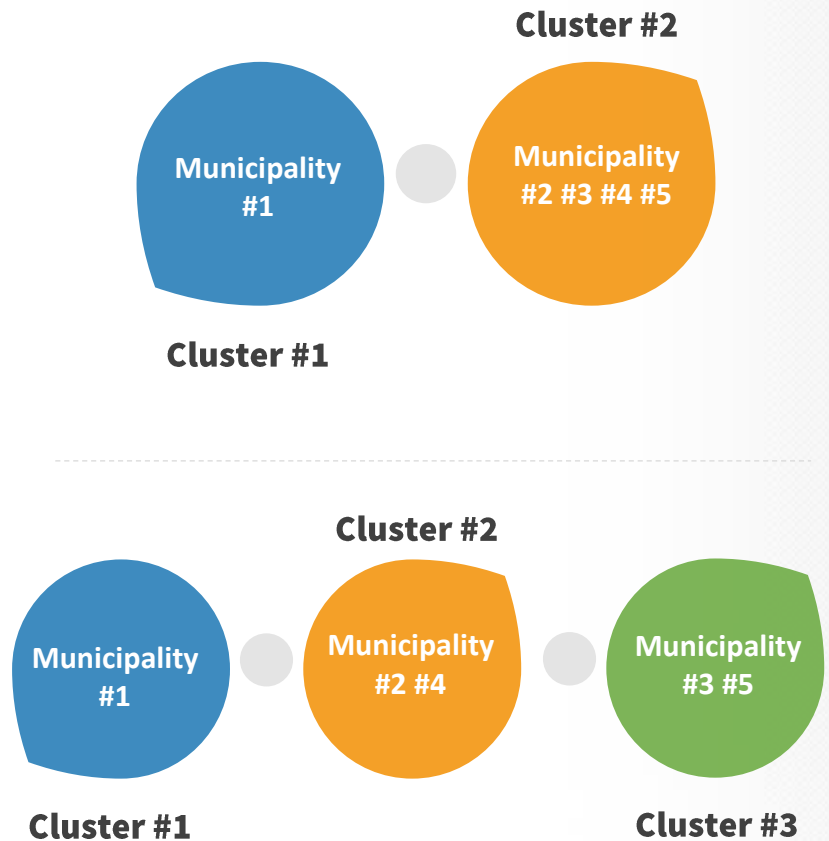
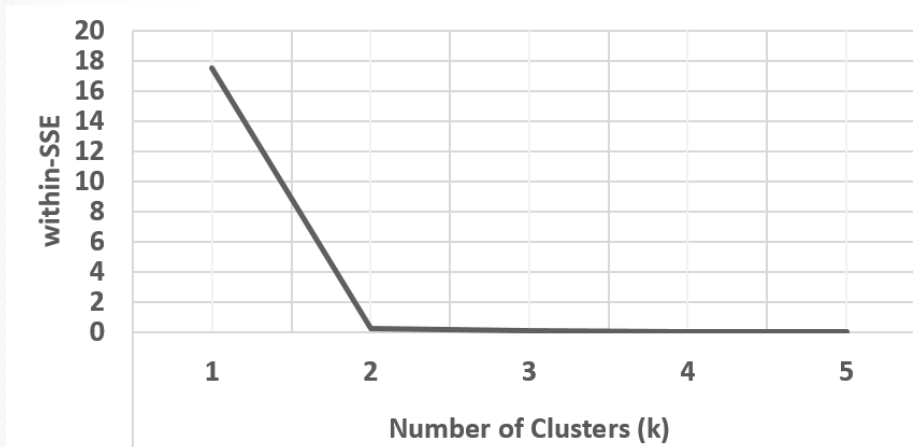
- **Performance Indicator Calculation:**
 - As appropriateness and fitness decrease; alignment costs increase for the municipalities
 - Performance indicators calculated over replay are acceptable

	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

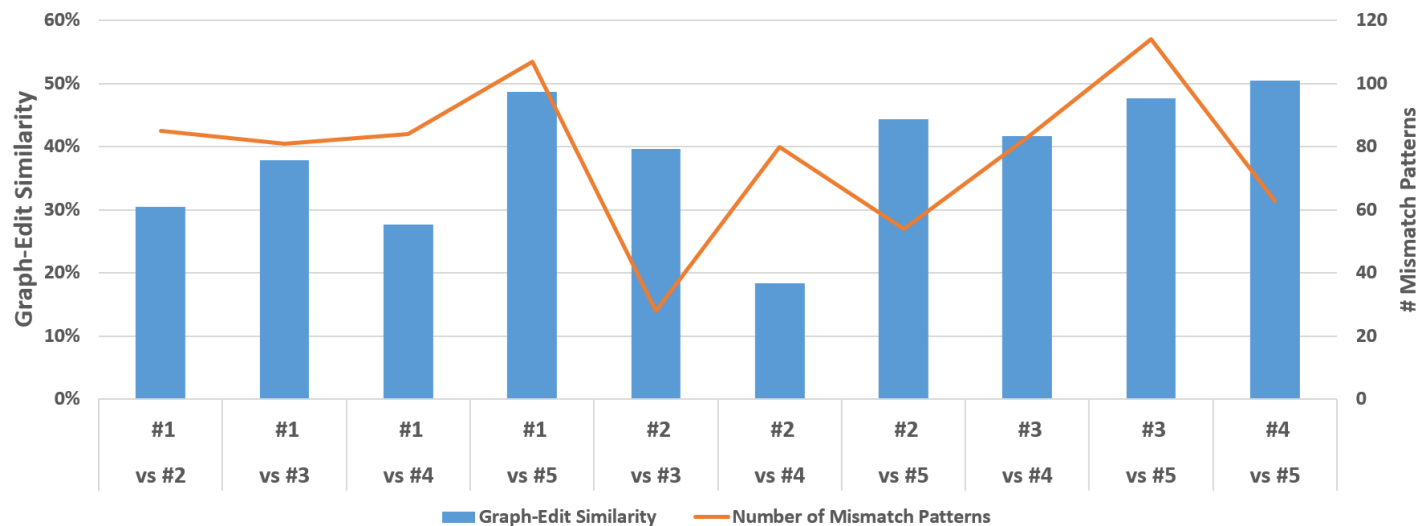
Environmental Permit Application Process – Performance Indicator Analysis

- Clustering:**



Results & Discussions

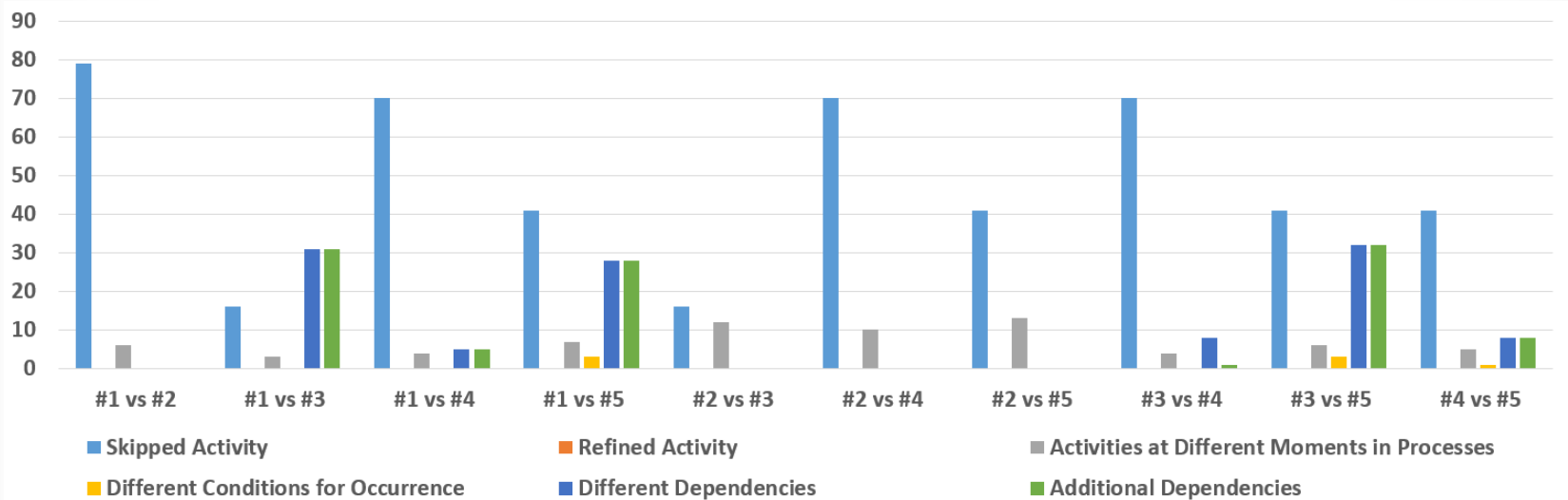
Environmental Permit Application Process – Mismatch Pattern Analysis



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

Results & Discussions

Environmental Permit Application Process – Mismatch Pattern Analysis

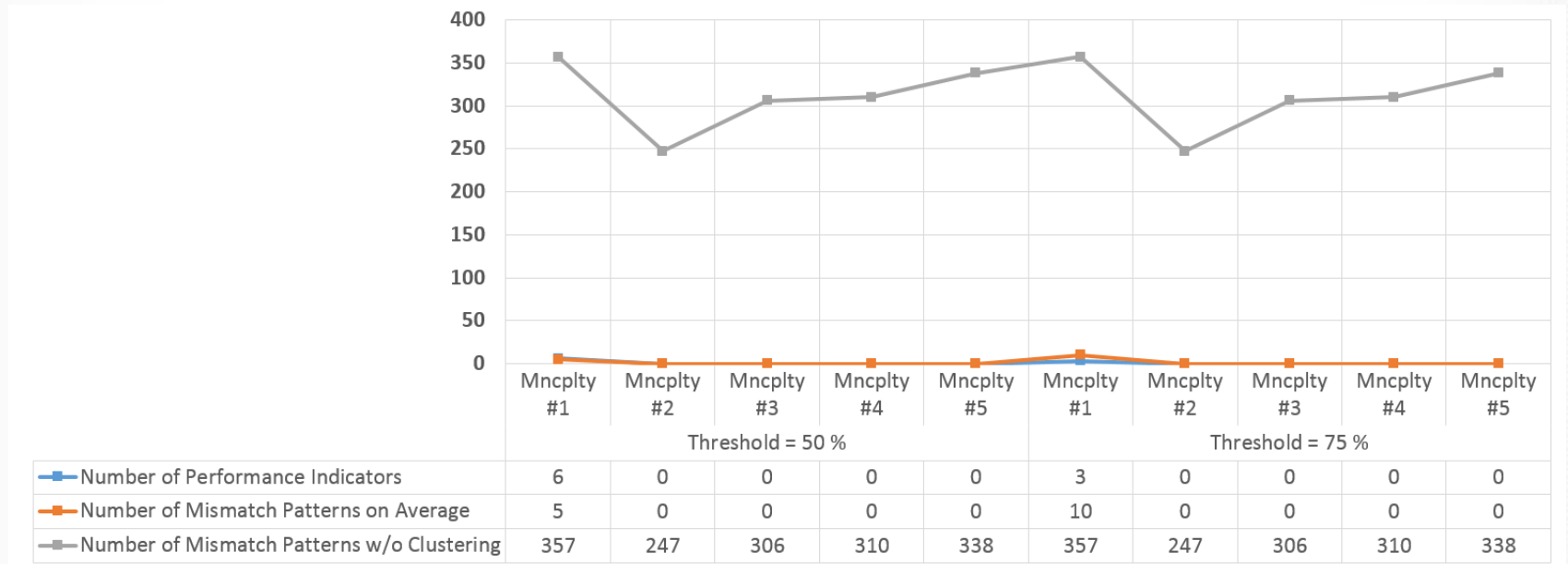


- "Skipped Activity" mostly
- "Refined Activity" is eliminated since codes are used instead of activity names

Results & Discussions

Environmental Permit Application Process – Recommendation Generation

(with 2 clusters)

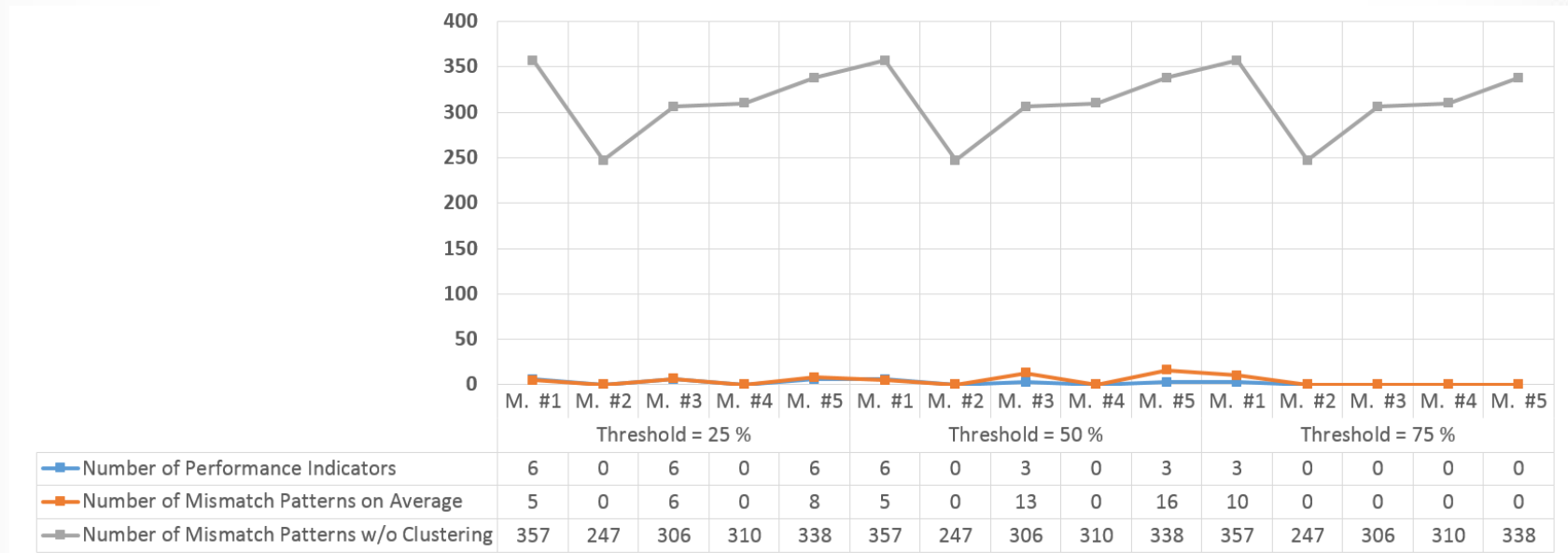


- Number of mismatch patterns to check significantly decreases with performance clustering

Results & Discussions

Environmental Permit Application Process – Recommendation Generation

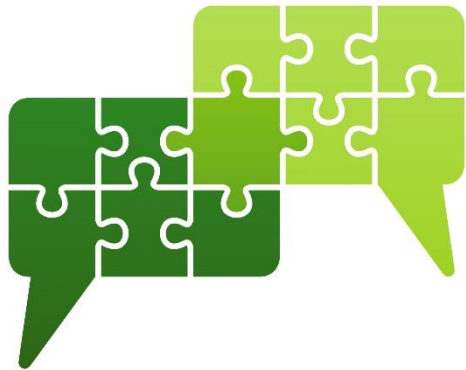
(with 3 clusters)



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

Results & Discussions

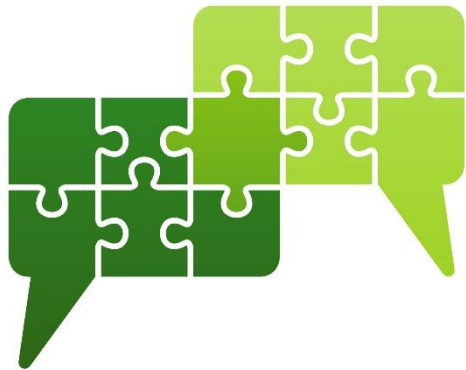
Discussions



- 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

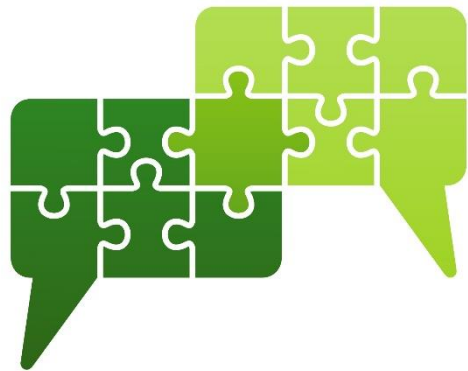
Discussions



- 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

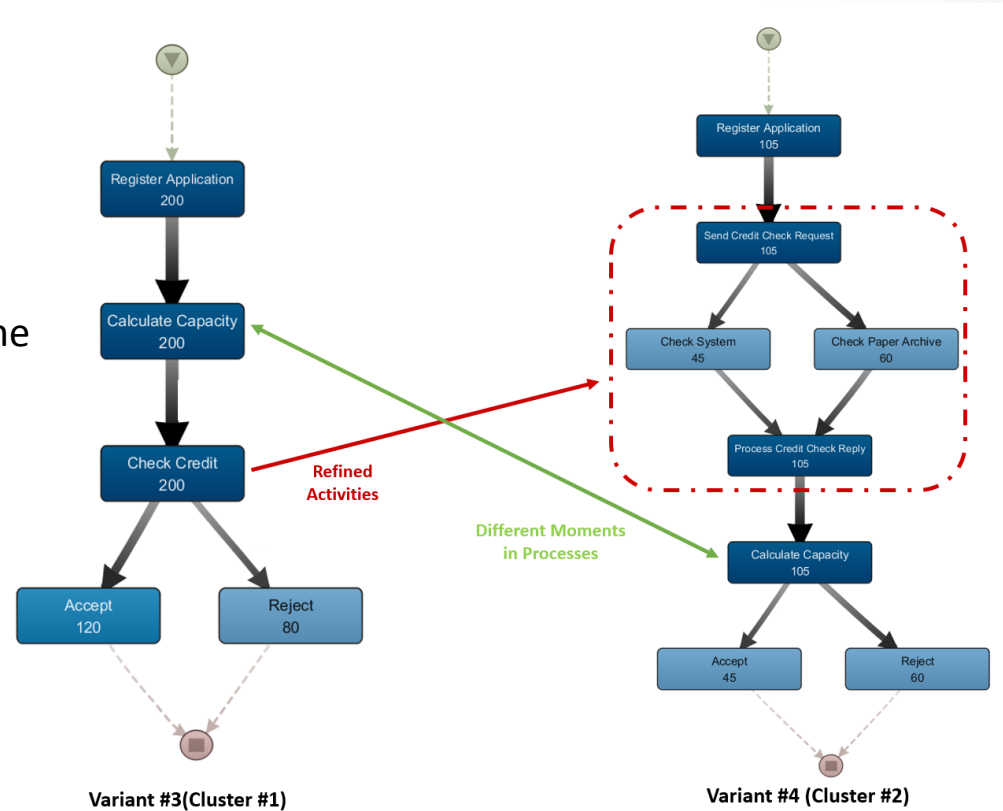
Discussions

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

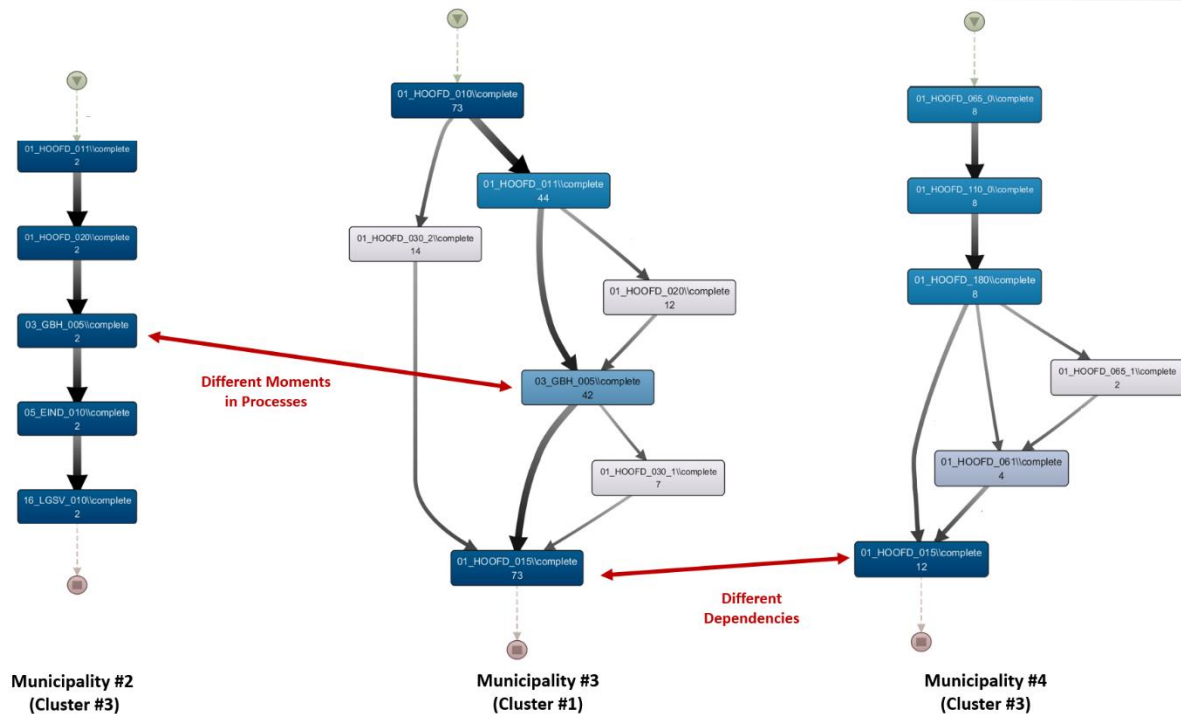
Discussions

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

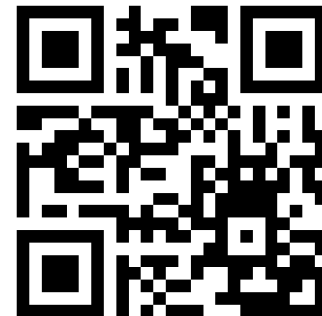
7

Demonstration

Demonstration



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



References

TODO

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