

Assessment of BPMN Models Complexity for Redesign using Cluster Analysis

Chrysoula Fotoglou

Supervisor: Dr Kostas Vergidis

MSc in Applied Informatics, University of Macedonia

29 October 2019

Presentation Context

Basic concepts

Aim & Objectives

Methodology

1st method – Cluster Analysis

2nd method – Weighted Sum

Examples from Literature

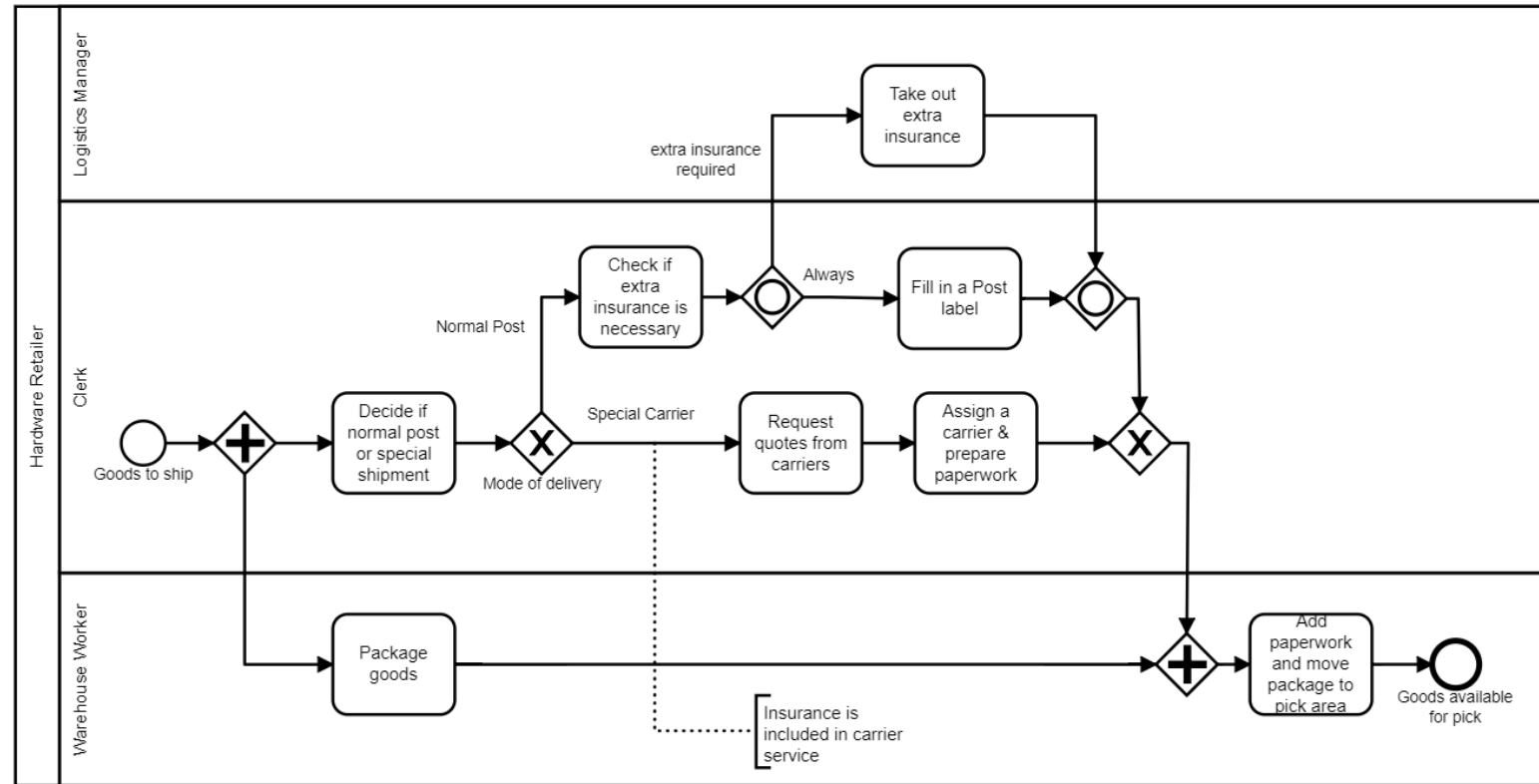
Conclusions

Business Process Model & Notation (BPMN)

- Popular Modeling standard for Business Processes
- Able to represent complex business concepts

Basic Elements of BPMN:

- Activities
- Gateways (XOR, OR, AND)
- Sequence flows
- Events



Business Process Complexity

“The degree to which a system or component has a design or implementation that is difficult to analyze, understand or explain”¹

- Complexity Measurement → Metrics
- High Complexity → Unsuccessful Business Process Redesign (BPR)
- BPR Dimensions : Cost, Time, Quality, Flexibility
- Complexity Reduction → Normalization

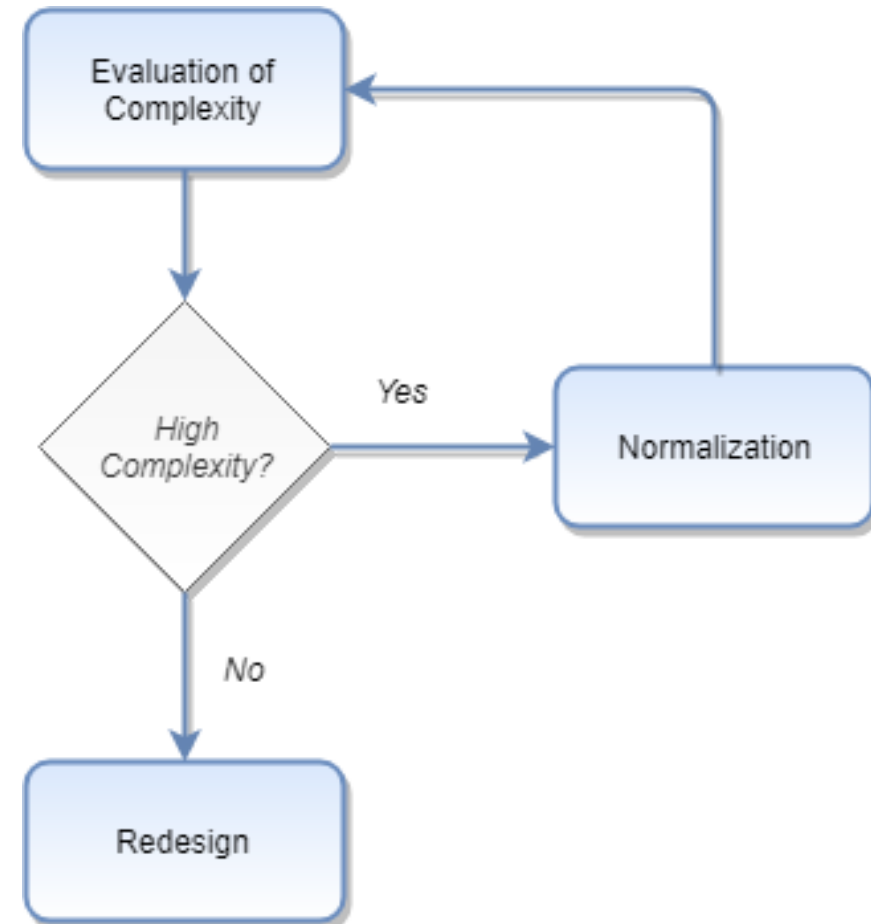
¹J. Cardoso, “Control-flow complexity measurement of processes and Weyuker’s properties,” in 6th International Enformatika Conference, 2005, vol. 8, pp. 213–218.

Aim & Objectives

Main Purpose: Complexity assessment of BPMN models in relation to Redesign

Objectives:

1. Selection of suitable metrics from literature to measure complexity in relation to Redesign
2. Adaptation of metrics to BPMN models
3. Development of methods that provide efficient complexity assessment
4. Definition of reference values that facilitate the identification of highly complex models



Selection of Metrics

Selection criteria:

- Covering all relevant complexity aspects (size, control flow, structure)
- Empirically validated
- Easily adapted to BPMN
- Simple and straightforward calculation

Complexity Type	Metric
<i>Activity</i>	NOA
	NOAJS
<i>Control flow</i>	CFC
<i>Structural</i>	CNC
	Density

NOA: Number of Activities

NOAJS: Number of Activities, Joins and Splits

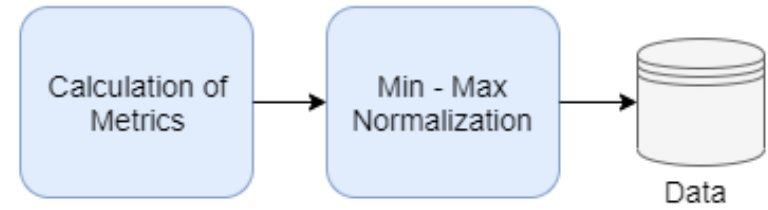
$$CFC = \left(\sum_{i \in (XOR\text{-splits of } P)} CFC_{XOR\text{-split } i} \right) + \left(\sum_{j \in (OR\text{-splits of } P)} CFC_{OR\text{-split } j} \right) + \left(\sum_{k \in (AND\text{-splits of } P)} CFC_{AND\text{-split } k} \right)$$

$$CNC = \frac{Arcs}{Nodes}$$

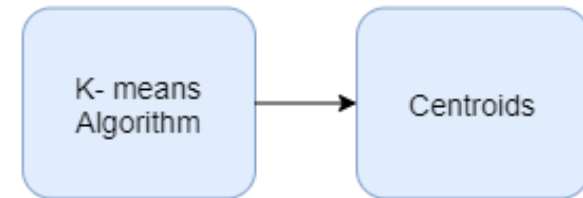
$$Density = \frac{Arcs}{Nodes * (Nodes - 1)}$$

Methodology

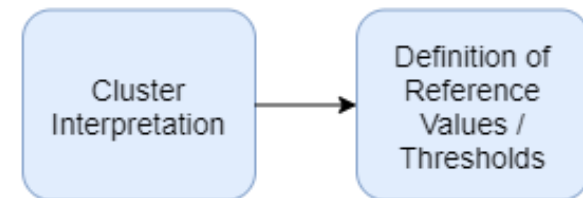
Data Preprocessing Phase



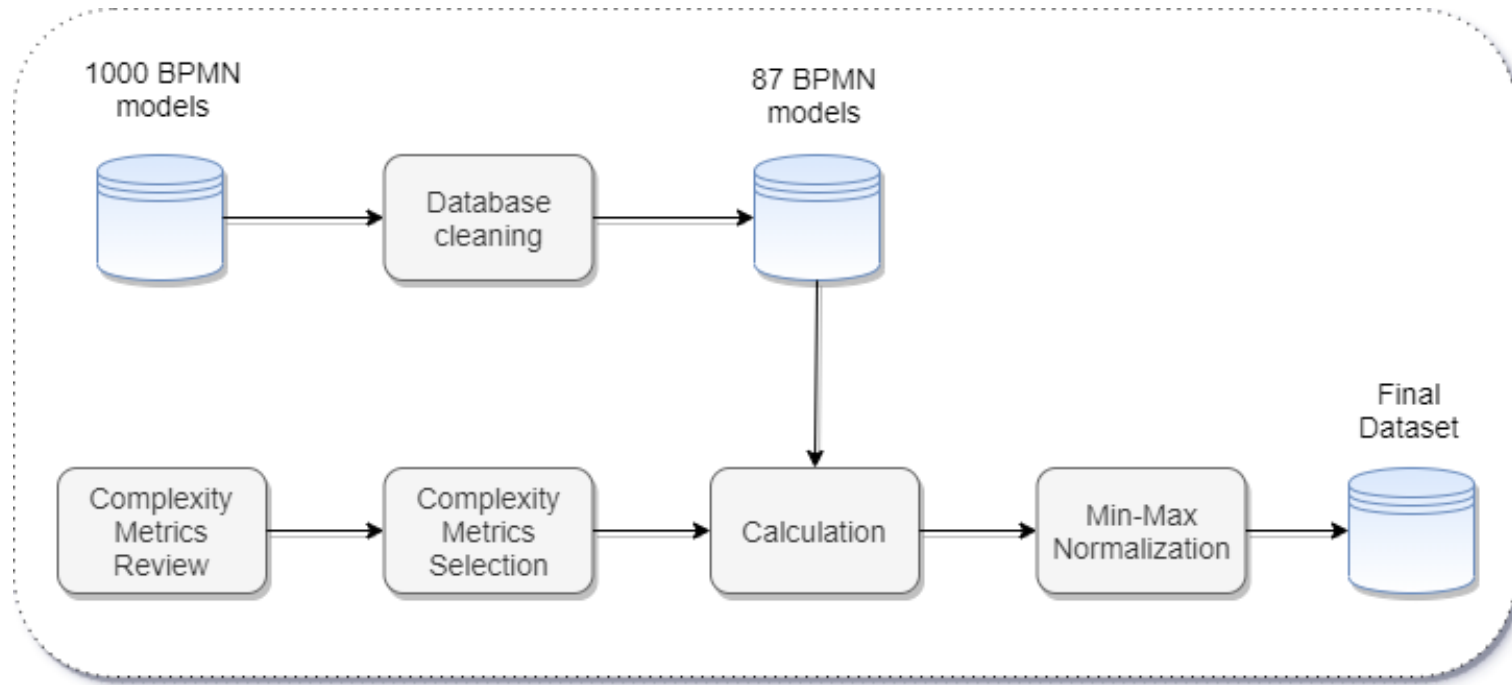
Clustering Phase



Analysis Phase



Data Preprocessing



Clustering

➤ K-means Algorithm

- Data mining technique
- Most popular for Clustering
- Centroid-based partitioning

➤ Parameters

- Num of clusters (K) → 3
- Initialization method → Random
- Similarity measure → Euclidean distance

Steps:

1. Specify the number of clusters (K)
2. Randomly select K points as initial cluster centroids
3. Assign the data points to the closest centroid
4. Recompute the centroids of the clusters
5. Repeat the process from step 3
6. Stop when the clusters do not change

Analysis of the output

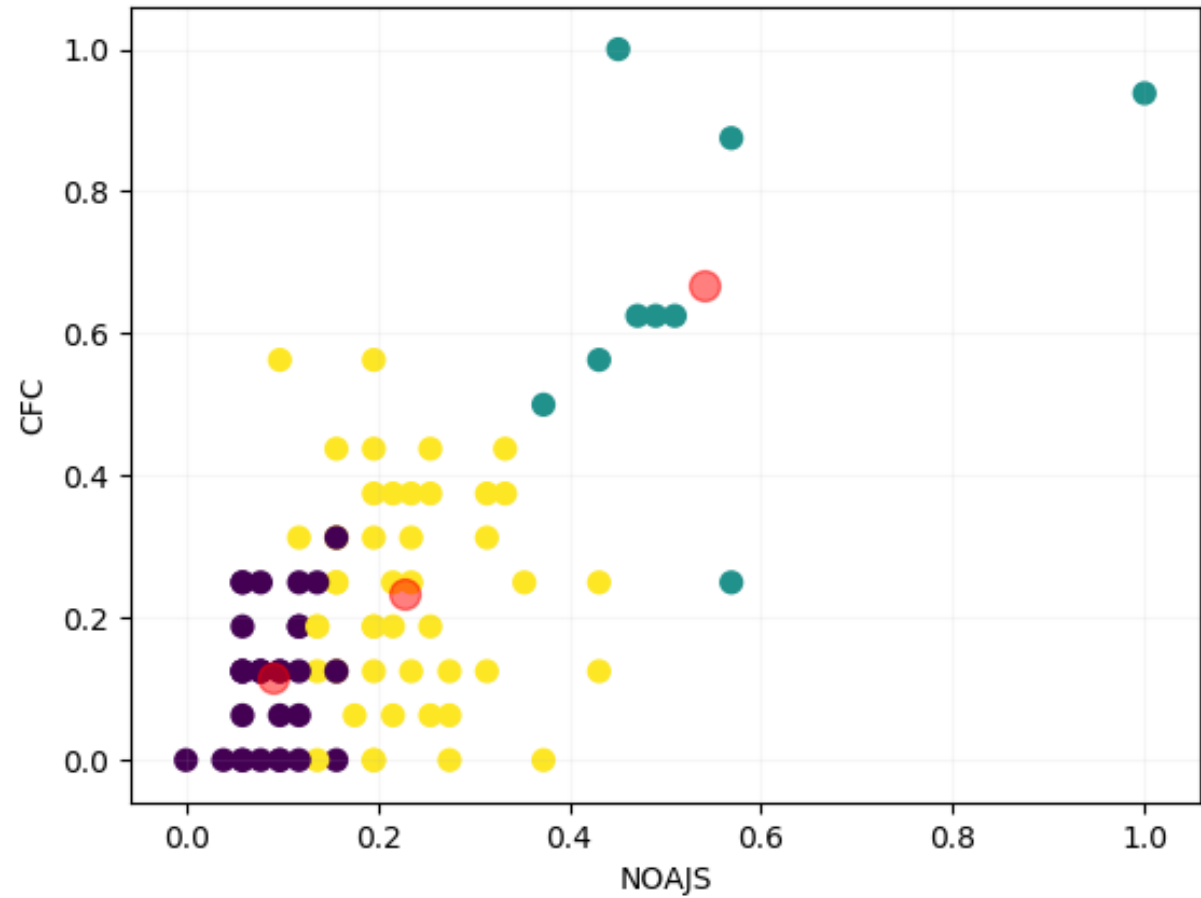
- Centroids as representatives of the Clusters
- Low values → Low Complexity
- Moderate values → Moderate Complexity
- High values → High Complexity

1st Approach – Initial Clustering

K-means on 5 Features

	NOA	NOAJS	CFC	D	CNC
Centroid 1	0.091	0.091	0.115	0.430	0.393
Centroid 2	0.243	0.228	0.233	0.181	0.397
Centroid 3	0.529	0.540	0.667	0.073	0.559

- NOA Excluded
- Density Excluded

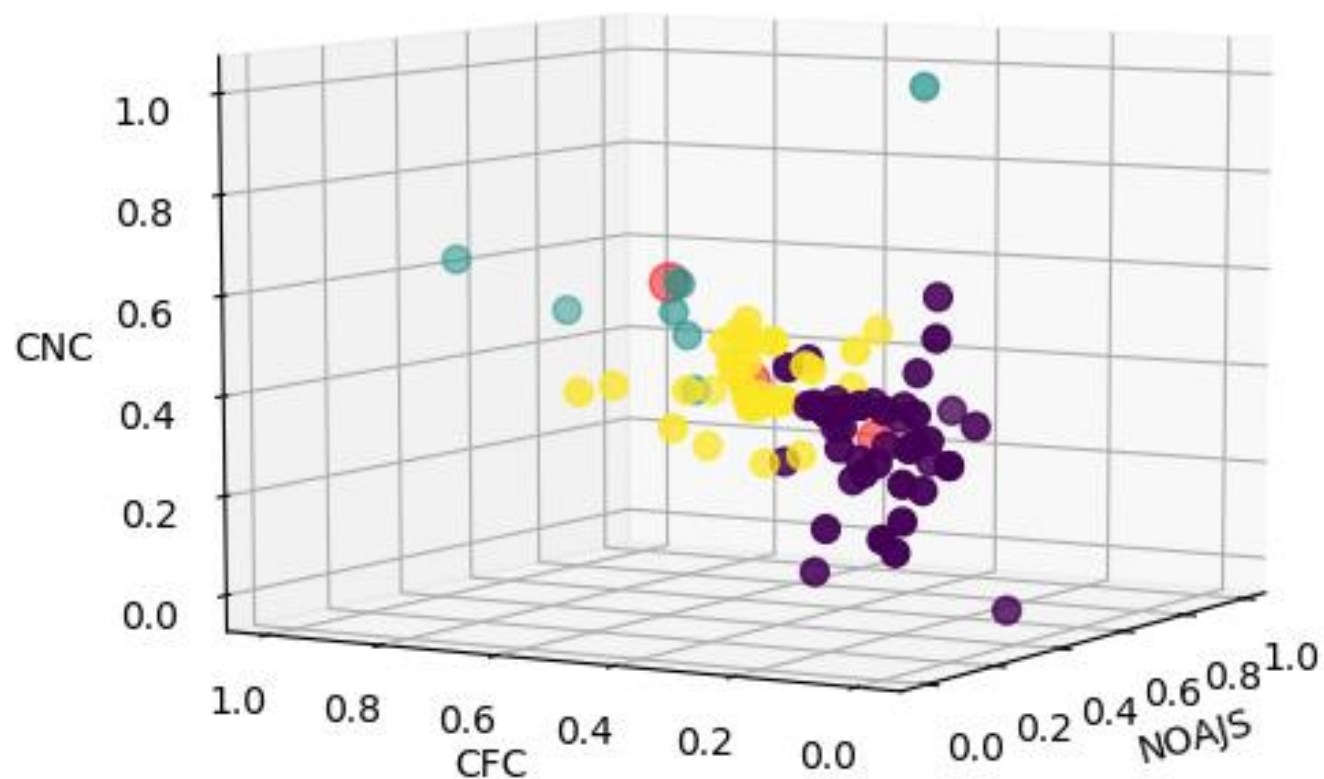


1st Approach – Updated Clustering

K-means on NOAJS, CFC, CNC

	NOAJS	CFC	CNC
Centroid 1	0.149	0.093	0.363
Centroid 2	0.213	0.339	0.447
Centroid 3	0.561	0.688	0.580

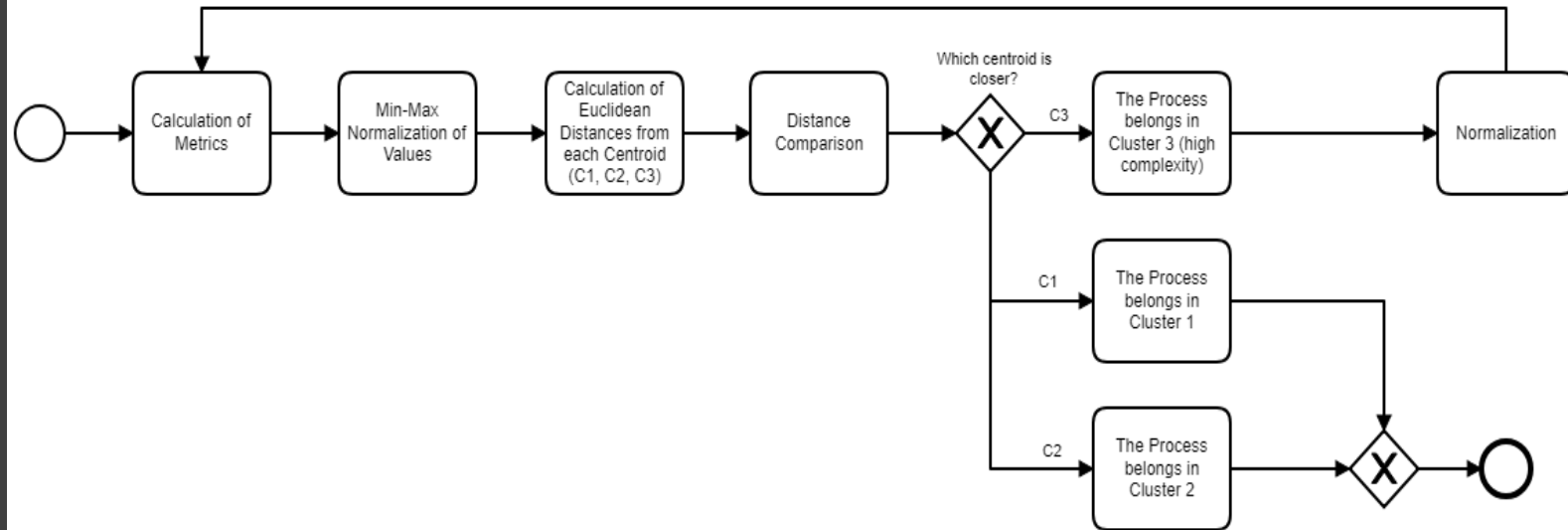
	Complexity	Color
Cluster 1	Low	Purple
Cluster 2	Moderate	Yellow
Cluster 3	High	Green



Assessment using Centroids

Steps

1. Calculation of metrics
2. Min-Max normalization of the metric values
3. Calculation of Euclidean distance from each centroid (C1, C2, C3)
4. Comparison of the distances to decide which centroid is closer
5. Assign model to the most similar cluster
6. Model assigned to highly complex cluster → Normalization



2nd Approach – Weighted Sum

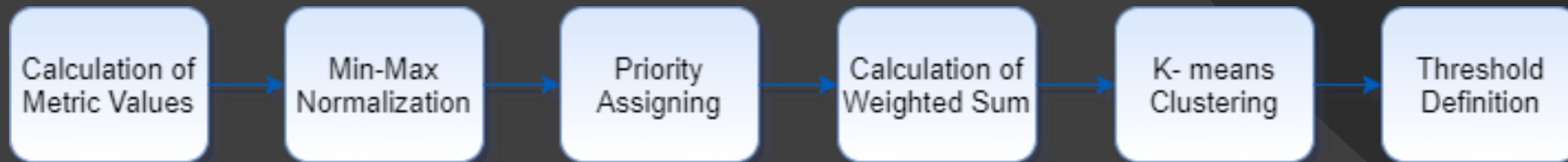
Let $p_1, p_2 \dots p_n$ be the priorities assigned to n metrics and $w_1, w_2 \dots w_n$ their weight

The weight of each metric is calculated by the formula:

$$w_i = \frac{p_i}{\sum_{i=1}^n p_i}$$

If $v_1, v_2 \dots v_n$ are the values of each of the n metrics, then:

$$WS = \sum_{i=1}^n w_i v_i$$

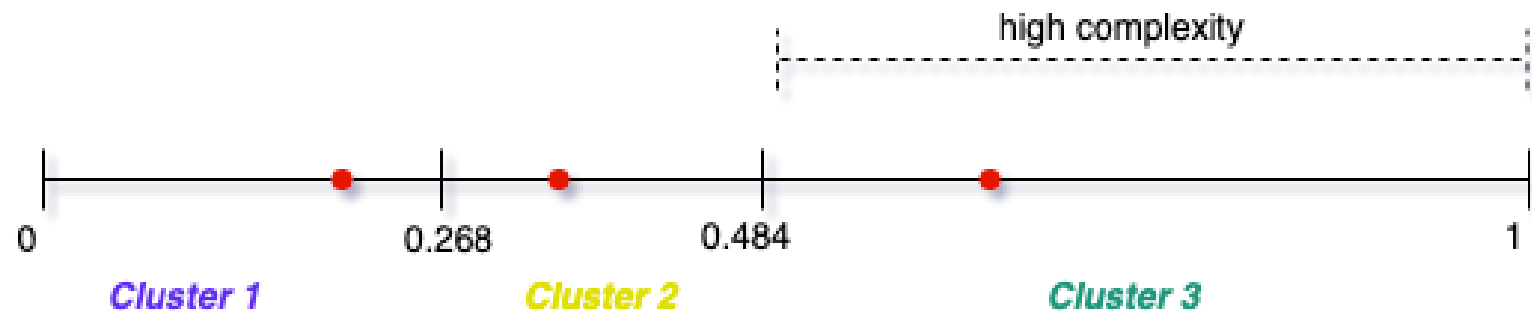


2nd Approach Weighted Sum – Scenario 1

Equal priorities for all metrics

- One-dimensional Illustration
↓
- Mean of Intervals between 2 adjacent centroids act as thresholds
- Easier threshold extraction
- Similar results as previous method

	Weighted Sum
Centroid 1	0.197
Centroid 2	0.339
Centroid 3	0.630

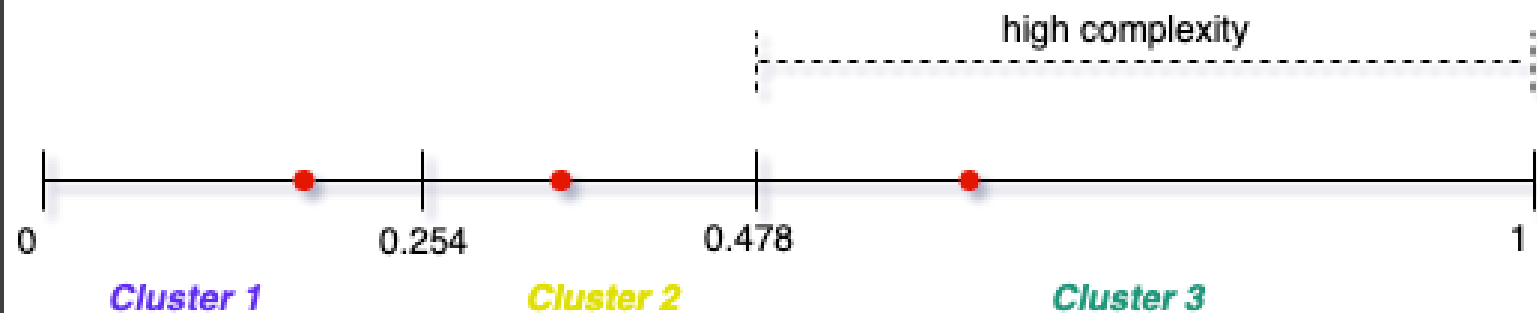


2nd Approach Weighted Sum – Scenario 2

Different priorities for each metric

- High CFC Priority
- Low NOAJIS Priority
- Moderate CNC Priority

	Weighted Sum
Centroid 1	0.1681
Centroid 2	0.3408
Centroid 3	0.6154



Example Processes

	Complexity Level		
Process	<u>Cluster Analysis</u>	<u>Weighted Sum – Scenario 1</u>	<u>Weighted Sum – Scenario 2</u>
Evaluate Quote Process	Moderate	Moderate	High
Property Valuation	Moderate	Moderate	Moderate
Healthcare Scenario	Moderate	Moderate	Moderate
Medical Assessment	High	High	High
Loan Application	Low	Low	Moderate
Bank Account Opening	High	High	High
Baking Workflow Process	Moderate	Moderate	Moderate
Emergency Ward of a Hospital	High	High	High
Auction	High	High	High
Admission Process	Moderate	Moderate	Moderate

Example Processes

	Complexity Level		
Process	<u>Cluster Analysis</u>	<u>Weighted Sum – Scenario 1</u>	<u>Weighted Sum – Scenario 2</u>
Evaluate Quote Process	Moderate	Moderate	High
Property Valuation	Moderate	Moderate	Moderate
Healthcare Scenario	Moderate	Moderate	Moderate
Medical Assessment	High	High	High
Loan Application	Low	Low	Moderate
Bank Account Opening	High	High	High
Baking Workflow Process	Moderate	Moderate	Moderate
Emergency Ward of a Hospital	High	High	High
Auction	High	High	High
Admission Process	Moderate	Moderate	Moderate

Conclusions

➤ Research Contribution

- Methods that combine metrics from literature to assess BPMN models complexity for Redesign
- Cluster analysis → Centroids as reference values
- Weighted Sum → Definition of exact thresholds
 - Assignment of priorities to metrics
 - Highlighting the impact of each metric on complexity and identify problematic areas

➤ Limitations

- Limited number of metrics
- Pre-defined number of complexity categories/clusters
- Dataset dependent techniques

Thank you!

Questions?