UNIVERSITÄT BAYREUTH

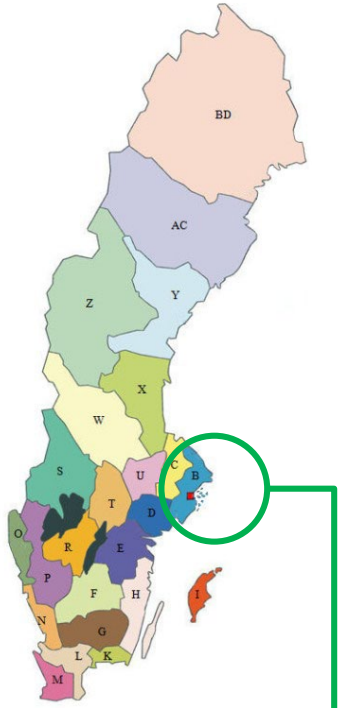
Beyond Singular:

Dealing with Multi-Case Notions in Process Mining

Amin Jalali, PhD.,
Associate Professor
Stockholm University
aj@dsv.su.se

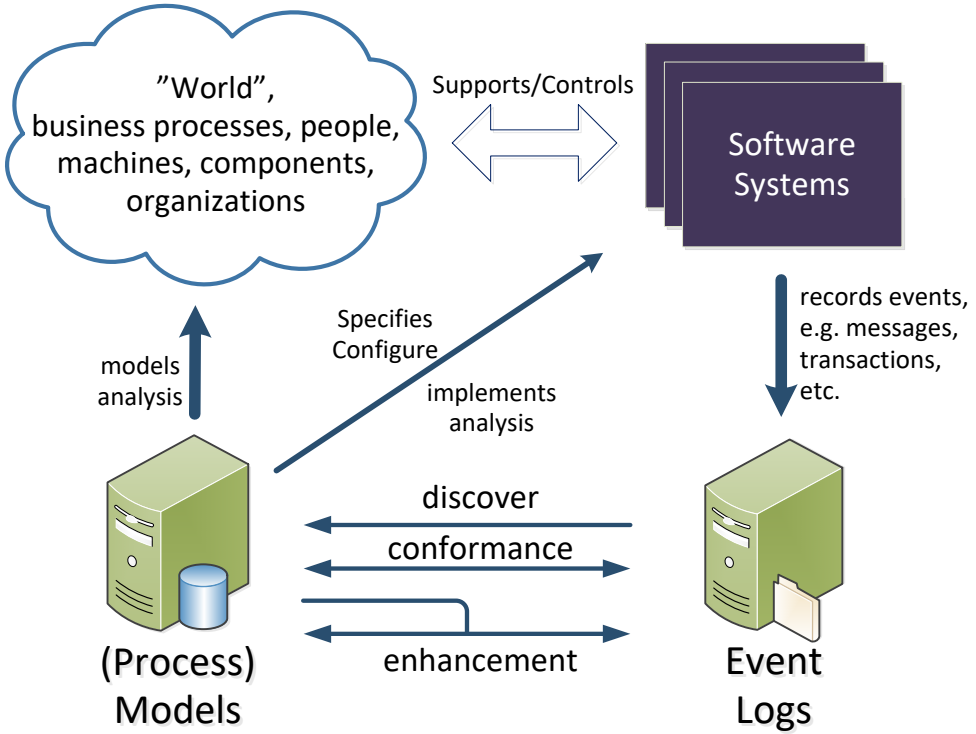
https://www.linkedin.com/in/aminjalali/
https://twitter.com/amin_jalali

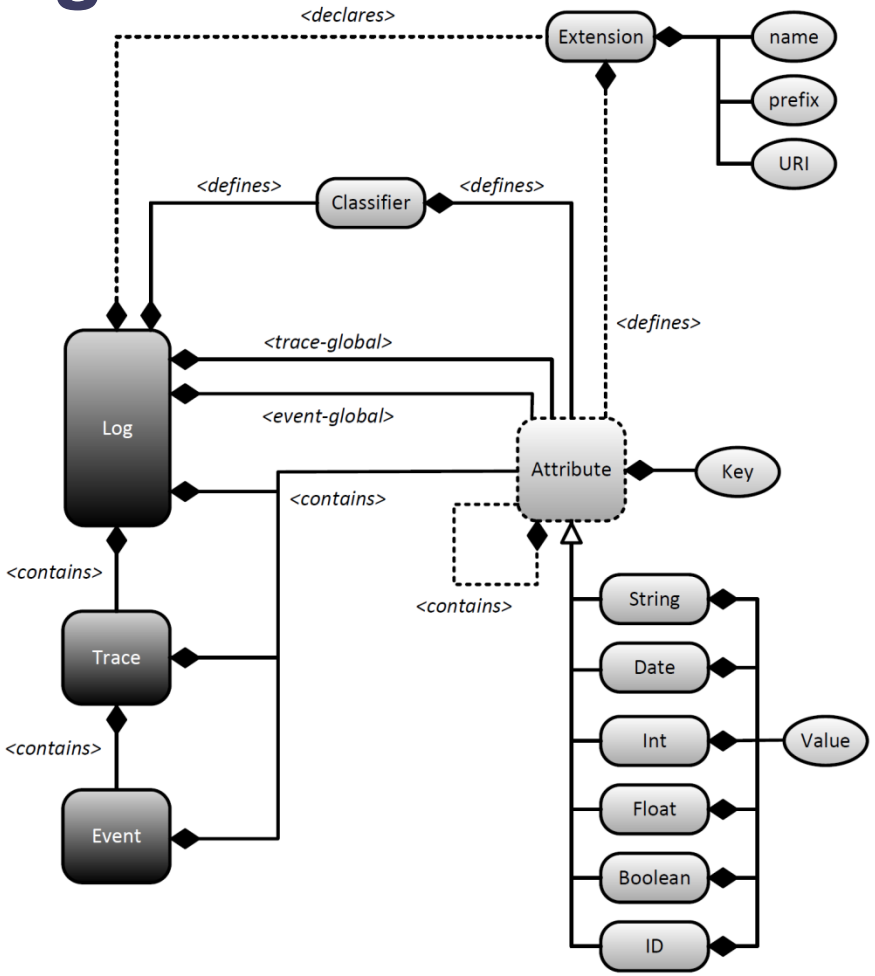
About us



Established 1878

Process Mining





XES
Extensible Event Stream

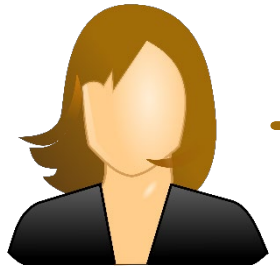
How to export a log file?

We need a case identifier!
Based on which, we flat the log!

Order
o1
o2

Item
i1
i2
i3

OrderItem	
o1	i1
o2	i2
o2	i3



o1, create_order, 2023-01-10T08:15:00, Amin
i1, pick_item, 2023-01-10T08:20:00, Amin
i1, confirm_item, 2023-01-10T08:22:00, Amin
o1, confirm_order, 2023-01-10T08:50:00, Amin
o2, create_order, 2023-01-11T08:15:00, Agnes
i2, pick_item, 2023-01-11T08:20:00, Agnes
i2, confirm_item, 2023-01-11T08:22:00, Agnes
i3, pick_item, 2023-01-11T08:25:00, Agnes
i3, confirm_item, 2023-01-11T08:27:00, Agnes
o2, confirm_order, 2023-01-11T08:29:00, Agnes

Selecting Item as the identifier

Convergence (one event may be related to different cases)

Wrong Analysis!

confirm_order = ~~3~~ 2

i1, create_order, 2023-01-10T08:15:00, Amin
i1, pick_item, 2023-01-10T08:20:00, Amin
i1, confirm_item, 2023-01-10T08:22:00, Amin
i1, **confirm_order**, 2023-01-10T08:50:00, Amin
i2, create_order, 2023-01-11T08:15:00, Agnes
i2, pick_item, 2023-01-11T08:20:00, Agnes
i2, confirm_item, 2023-01-11T08:22:00, Agnes
i2, **confirm_order**, 2023-01-11T08:29:00, Agnes
i3, create_order, 2023-01-11T08:15:00, Agnes
i3, pick_item, 2023-01-11T08:25:00, Agnes
i3, confirm_item, 2023-01-11T08:27:00, Agnes
i3, **confirm_order**, 2023-01-11T08:29:00, Agnes

Map them to related ones!

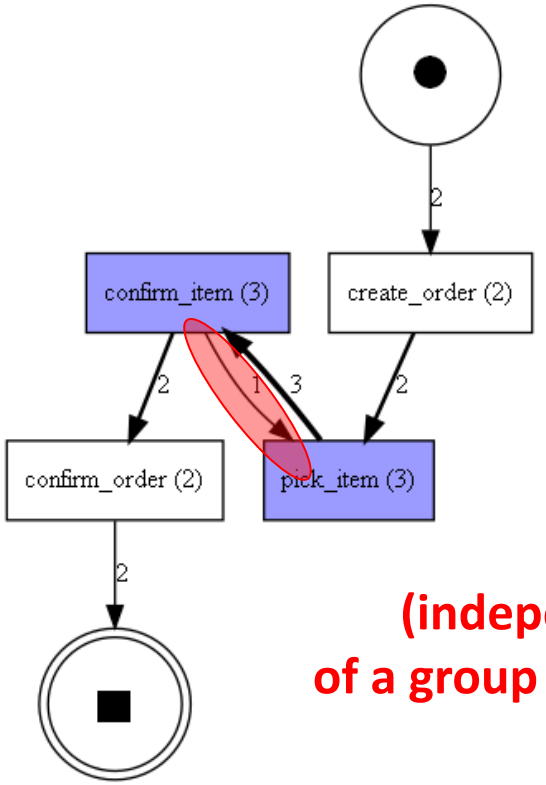
o1, create_order, 2023-01-10T08:15:00, Amin
i1, pick_item, 2023-01-10T08:20:00, Amin
i1, confirm_item, 2023-01-10T08:22:00, Amin
o1, confirm_order, 2023-01-10T08:50:00, Amin
o2, create_order, 2023-01-11T08:15:00, Agnes
i2, pick_item, 2023-01-11T08:20:00, Agnes
i2, confirm_item, 2023-01-11T08:22:00, Agnes
i3, pick_item, 2023-01-11T08:25:00, Agnes
i3, confirm_item, 2023-01-11T08:27:00, Agnes
o2, confirm_order, 2023-01-11T08:29:00, Agnes

Ignore other Events!

Not capturing the whole process!

i1, pick_item, 2023-01-10T08:20:00, Amin
i1, confirm_item, 2023-01-10T08:22:00, Amin
i2, pick_item, 2023-01-11T08:20:00, Agnes
i2, confirm_item, 2023-01-11T08:22:00, Agnes
i3, pick_item, 2023-01-11T08:25:00, Agnes
i3, confirm_item, 2023-01-11T08:27:00, Agnes

Selecting Order as the identifier



divergence
(independent, repeated executions of a group of activities within a single case)

Not capturing the whole process!

o1, create_order, 2023-01-10T08:15:00, Amin
o1, confirm_order, 2023-01-10T08:50:00, Amin
o2, create_order, 2023-01-11T08:15:00, Agnes
o2, confirm_order, 2023-01-11T08:29:00, Agnes

Ignore other Events!

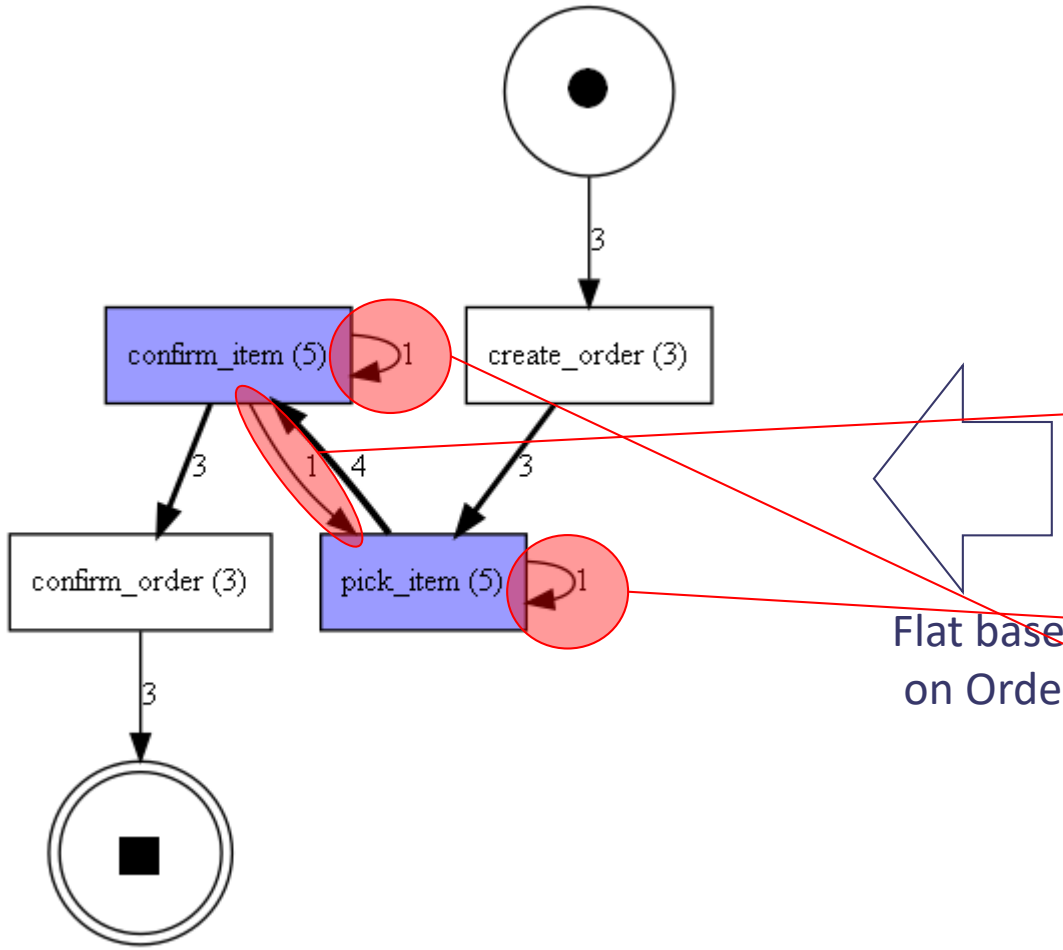
o1, create_order, 2023-01-10T08:15:00, Amin
o1, **pick_item**, 2023-01-10T08:20:00, Amin
o1, **confirm_item**, 2023-01-10T08:22:00, Amin
o1, confirm_order, 2023-01-10T08:50:00, Amin
o2, create_order, 2023-01-11T08:15:00, Agnes
o2, **pick_item**, 2023-01-11T08:20:00, Agnes
o2, **confirm_item**, 2023-01-11T08:22:00, Agnes
o2, **pick_item**, 2023-01-11T08:25:00, Agnes
o2, **confirm_item**, 2023-01-11T08:27:00, Agnes
o2, confirm_order, 2023-01-11T08:29:00, Agnes

Map them to related ones!

o1, create_order, 2023-01-10T08:15:00, Amin
i1, pick_item, 2023-01-10T08:20:00, Amin
i1, confirm_item, 2023-01-10T08:22:00, Amin
o1, confirm_order, 2023-01-10T08:50:00, Amin
o2, create_order, 2023-01-11T08:15:00, Agnes
i2, pick_item, 2023-01-11T08:20:00, Agnes
i2, confirm_item, 2023-01-11T08:22:00, Agnes
i3, pick_item, 2023-01-11T08:25:00, Agnes
i3, confirm_item, 2023-01-11T08:27:00, Agnes
o2, confirm_order, 2023-01-11T08:29:00, Agnes

Selecting Order as the identifier

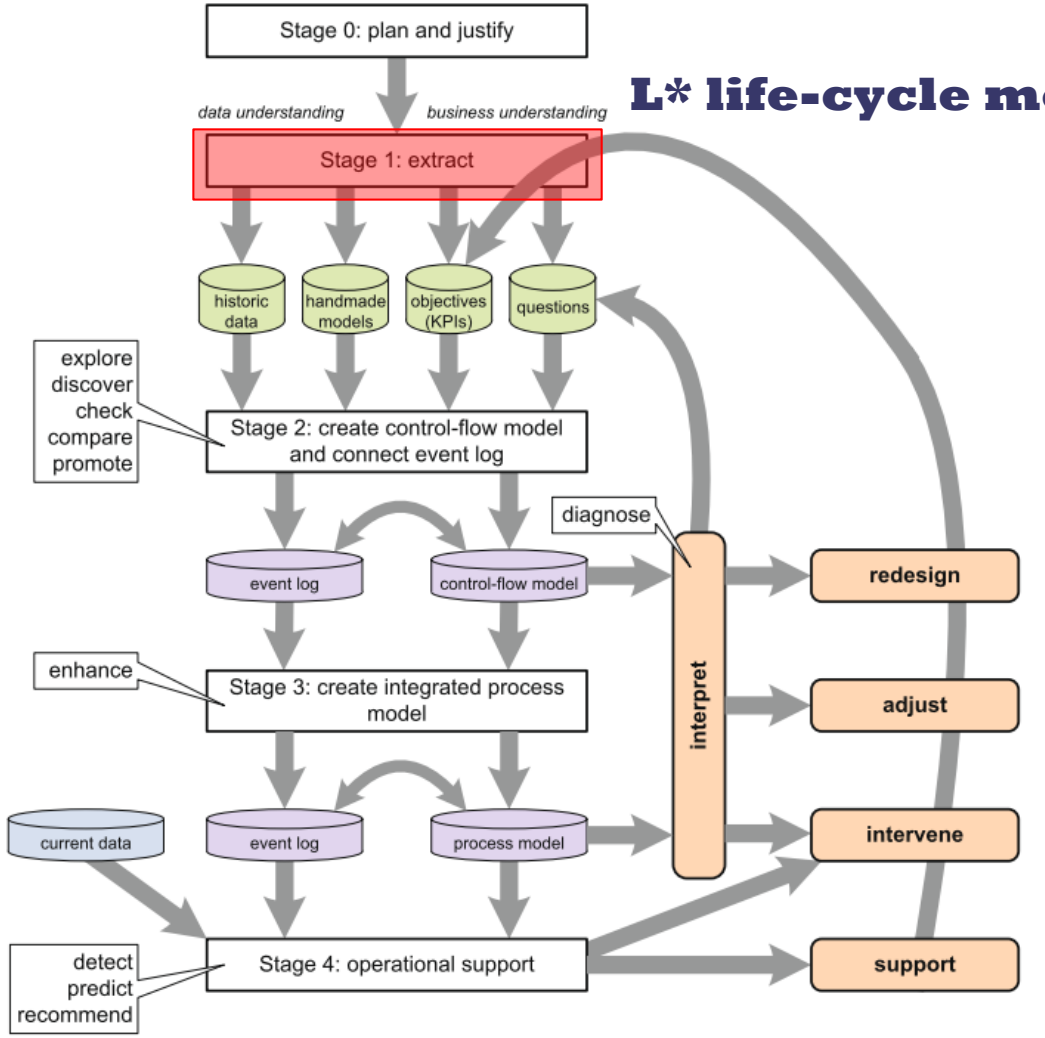
Open in Colab



Flat based on Order

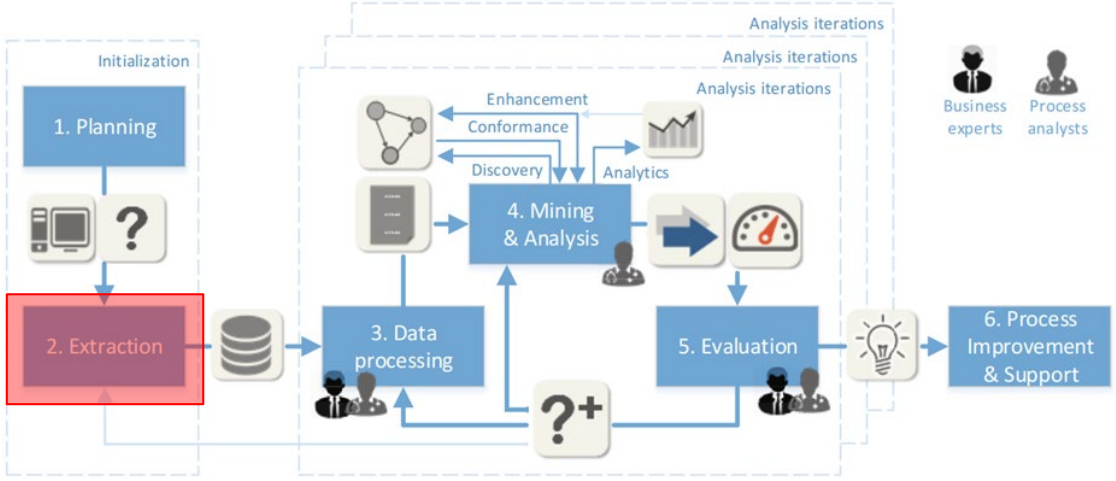
o1, create_order, 2023-01-10T08:15:00, Amin
i1, pick_item, 2023-01-10T08:20:00, Amin
i1, confirm_item, 2023-01-10T08:22:00, Amin
o1, confirm_order, 2023-01-10T08:50:00, Amin
o2, create_order, 2023-01-11T08:15:00, Agnes
i2, pick_item, 2023-01-11T08:20:00, Agnes
i2, confirm_item, 2023-01-11T08:22:00, Agnes
i3, pick_item, 2023-01-11T08:25:00, Agnes
i3, confirm_item, 2023-01-11T08:27:00, Agnes
o2, confirm_order, 2023-01-11T08:29:00, Agnes
o2, create_order, 2023-01-11T08:15:00, Linda
i2, pick_item, 2023-01-11T08:20:00, Linda
i3, pick_item, 2023-01-11T08:21:00, Linda
i2, confirm_item, 2023-01-11T08:22:00, Linda
i3, confirm_item, 2023-01-11T08:23:00, Linda
o2, confirm_order, 2023-01-11T08:29:00, Linda

One analysis may need several Log Extraction

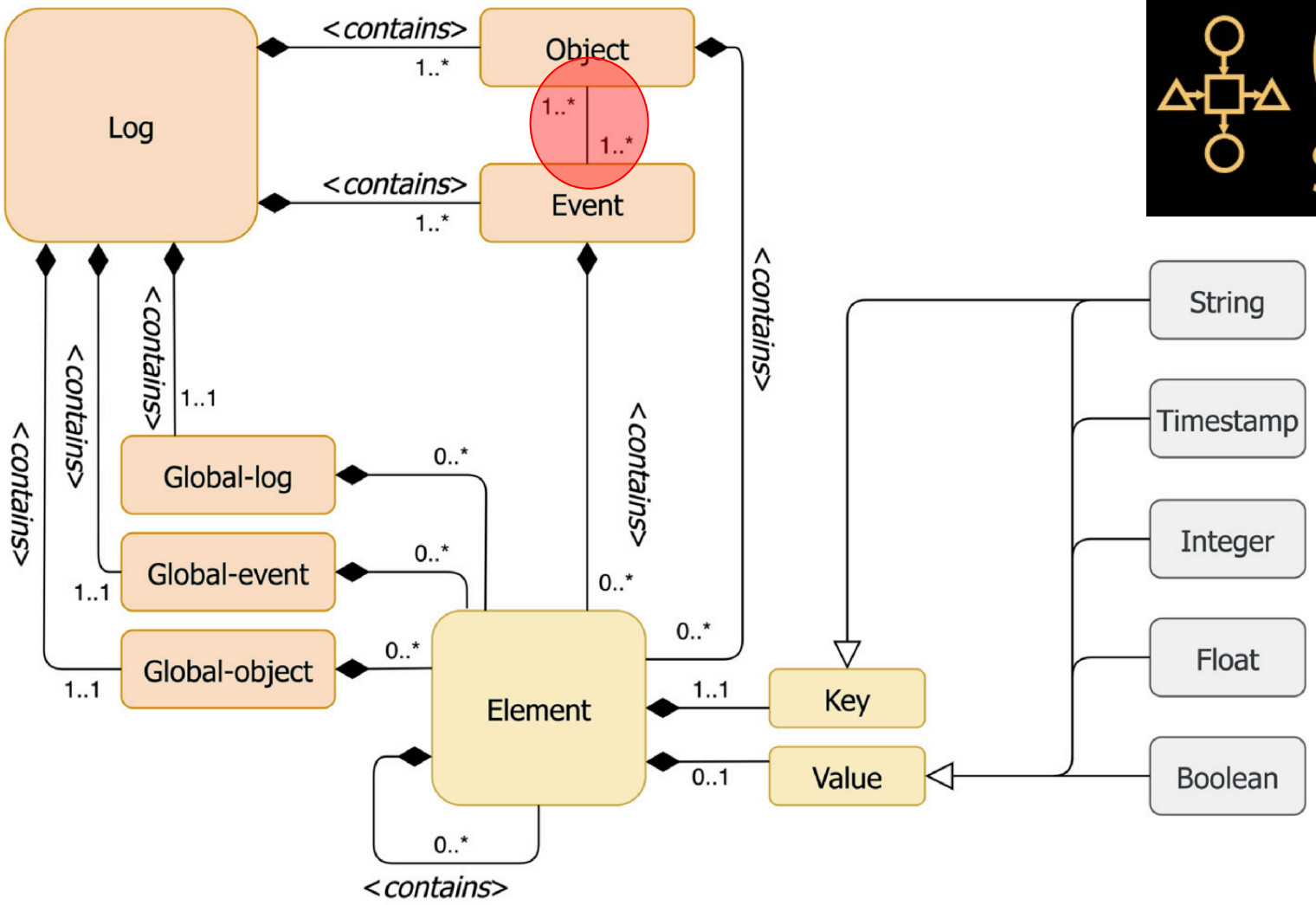


L* life-cycle model

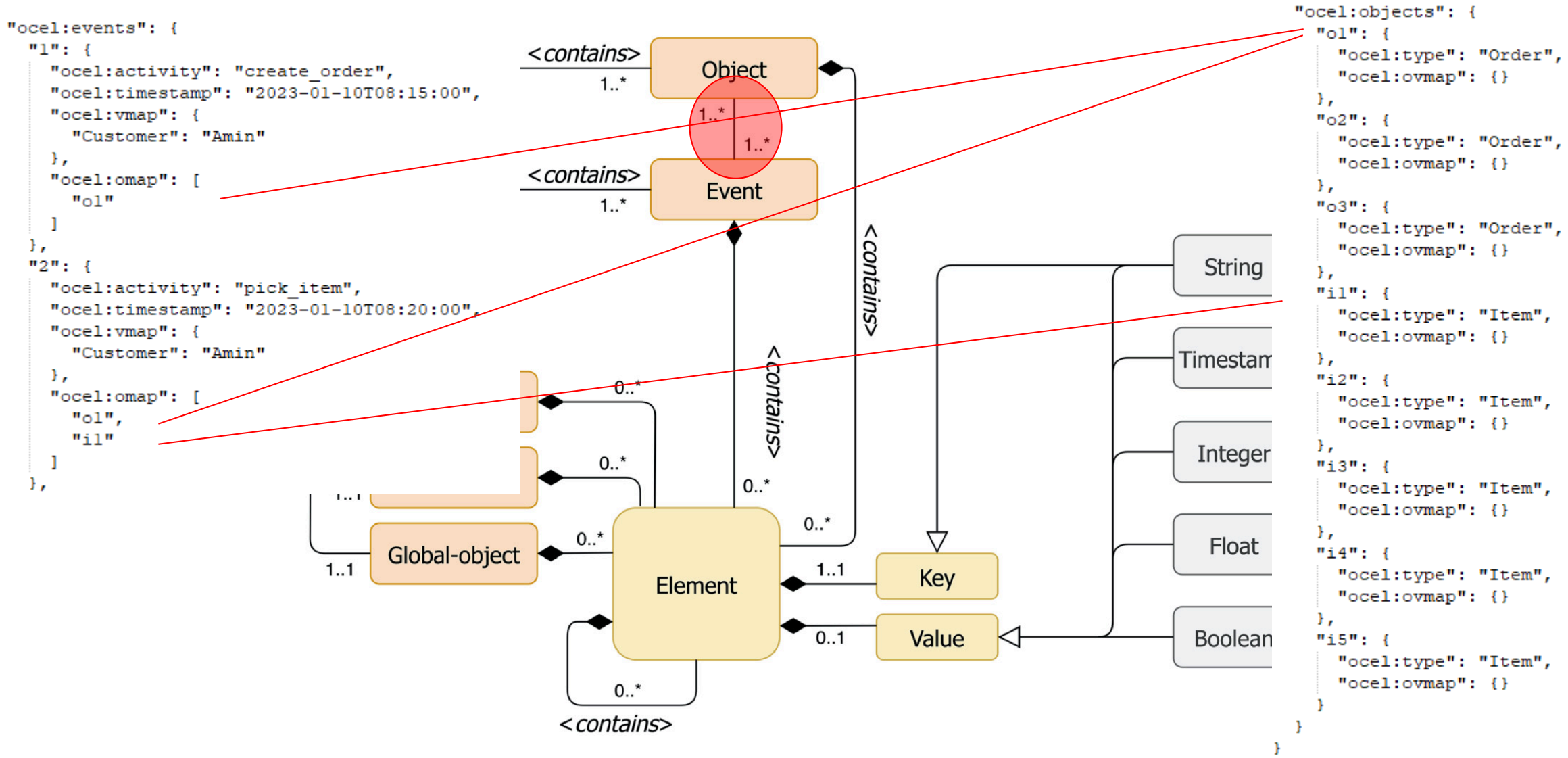
PM2 methodology



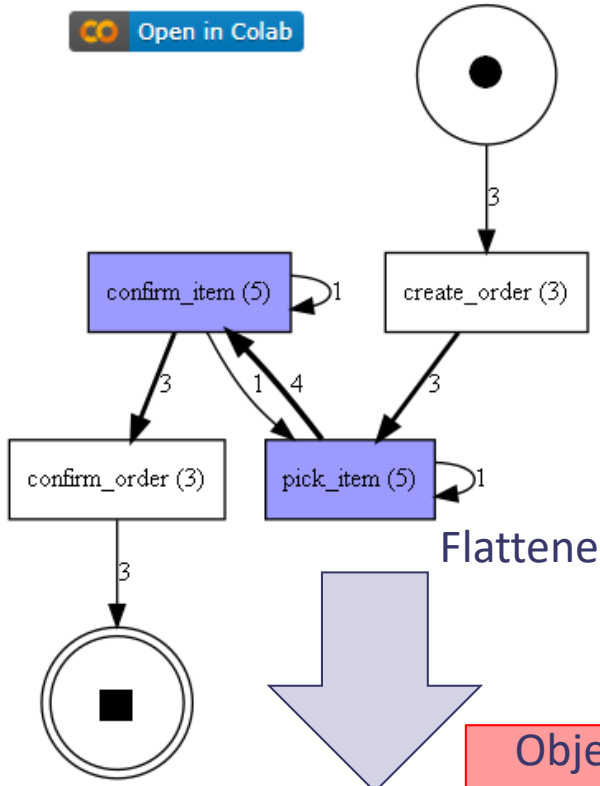
OCEL: A Standard for Object-Centric Event Logs

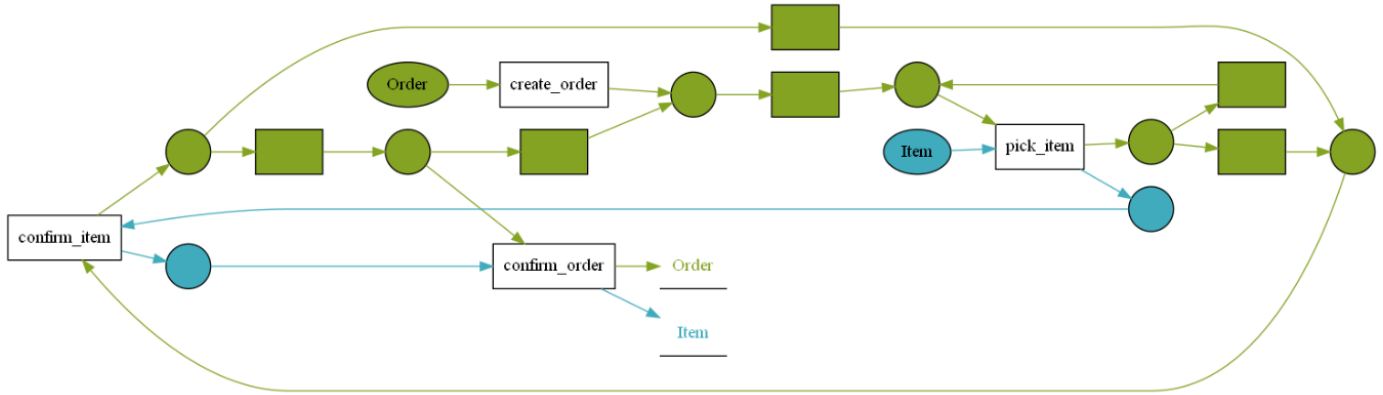


How our log will look like?



DFG vs DFM (OC-DFG) to OC-Petri nets





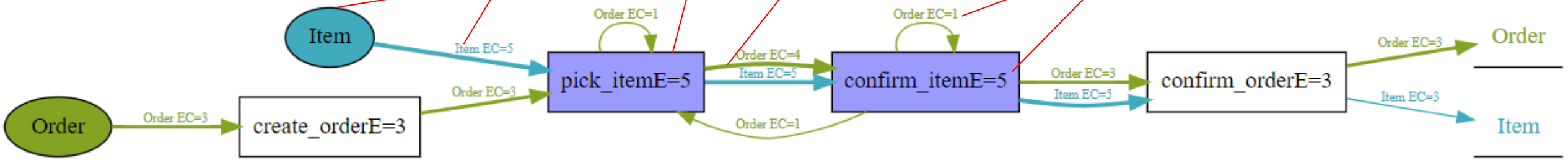
Flattened by Order

Flow

Task

Object Type

Frequency



Single case vs Multi-case notions

Feature	Traditional Process Mining	Object-Centric Process Mining
Tools Availability	+	-
Simplicity	+	-
Level of maturity	+	-
Convergence	-	+
Divergence	-	+
log extraction	-	+

Do we need to select one of them?
Can they go hand in hand?

How to make a balance?

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 November 2022, accepted 1 December 2022, date of publication 2 December 2022,
date of current version 7 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3226573

Object Type Clustering Using Markov Directly-Follow Multigraph in Object-Centric Process Mining

https://ieeexplore.ieee.org/document/9969591/

AMIN JALALI
Department of Computer and Systems Sciences, Stockholm University, 164 07 Kista, Sweden
e-mail: aj@dsv.su.se

A sample DFM

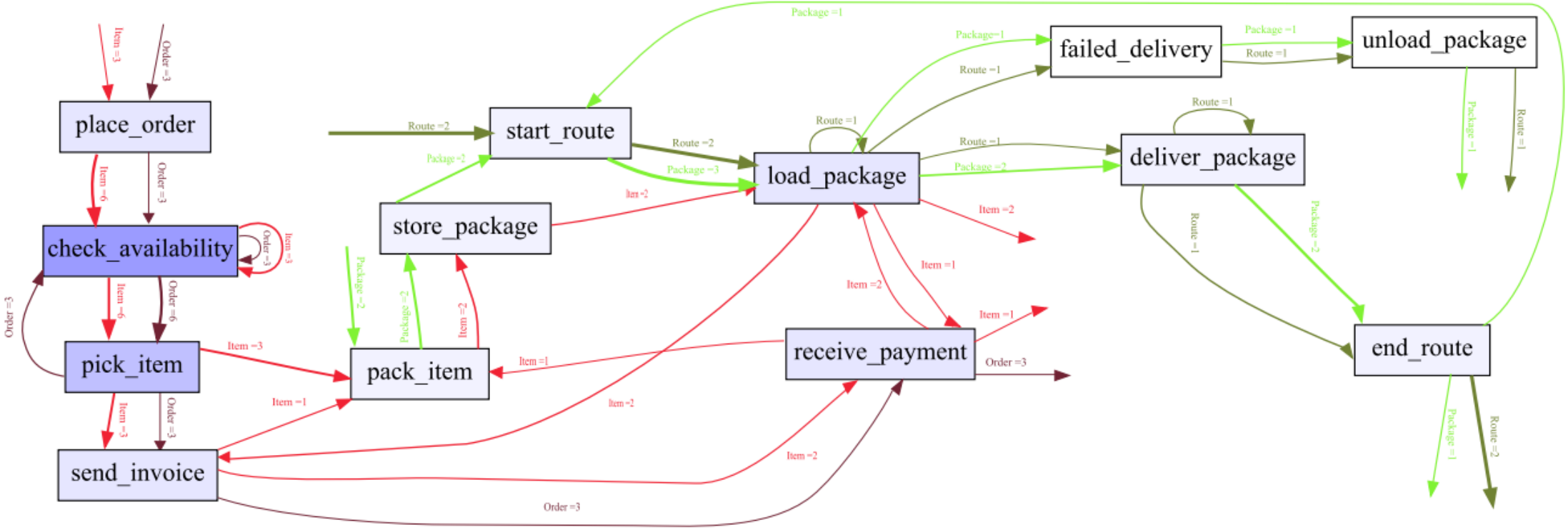


FIGURE 1. A Directly-Follows Multigraph (DFM), discovered from 39 events, indicates how process models incorporating all case notions can become complex.

Marov DFM

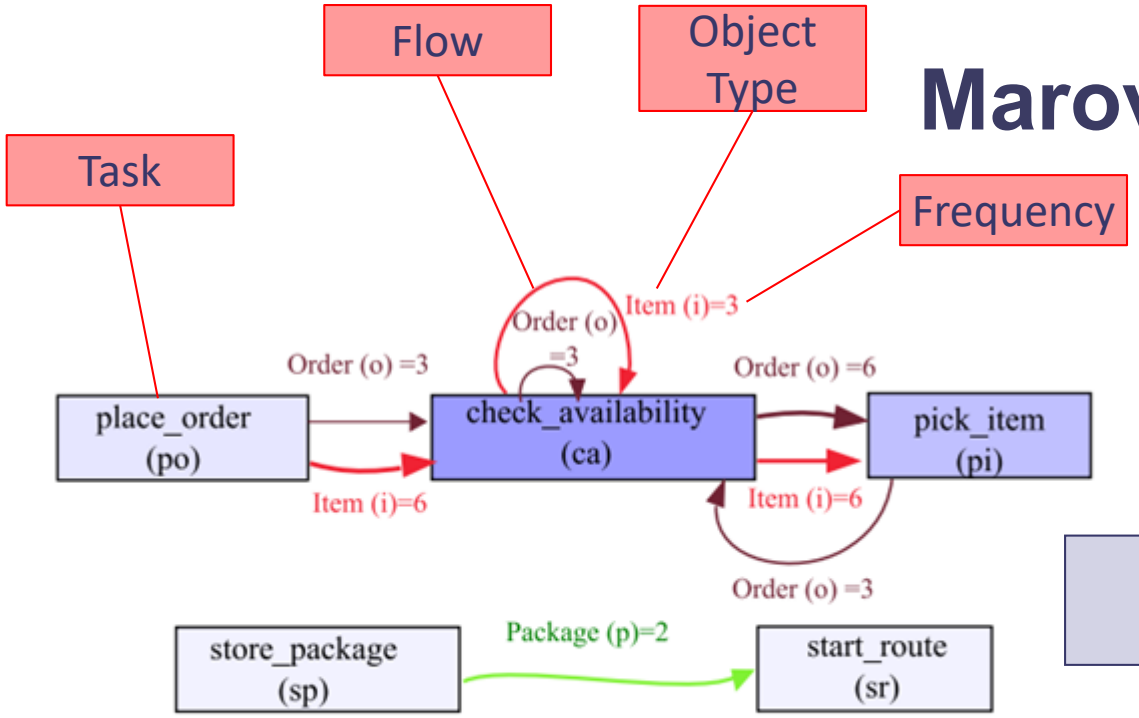


FIGURE 2. **A simple DFM taken from FIGURE 1 for explaining the approach.**

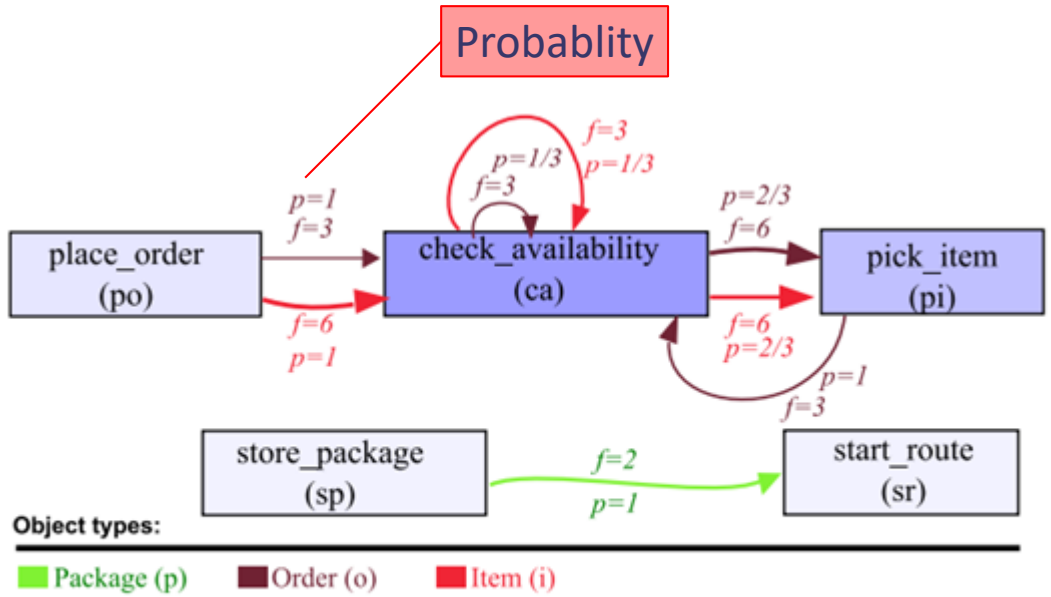


FIGURE 3. **A Markov DFM of the DFM presented in FIGURE 2.**

Object Type

$$p\left((t,\theta,t')\right) \leftarrow \frac{f\left((t,\theta,t')\right)}{\sum_{\forall t'' \in t^{\{\theta\}}_{\bullet}} f\left((t,\theta,t'')\right)} \qquad (1)$$

Source task

Target task

Markov Adjacency matrix for each object type

$$p\big((t,\theta,t')\big) \leftarrow \frac{f\big((t,\theta,t')\big)}{\sum_{\forall t'' \in t^{\{\theta\}}_{\bullet}} f\big((t,\theta,t'')\big)} \qquad (1)$$

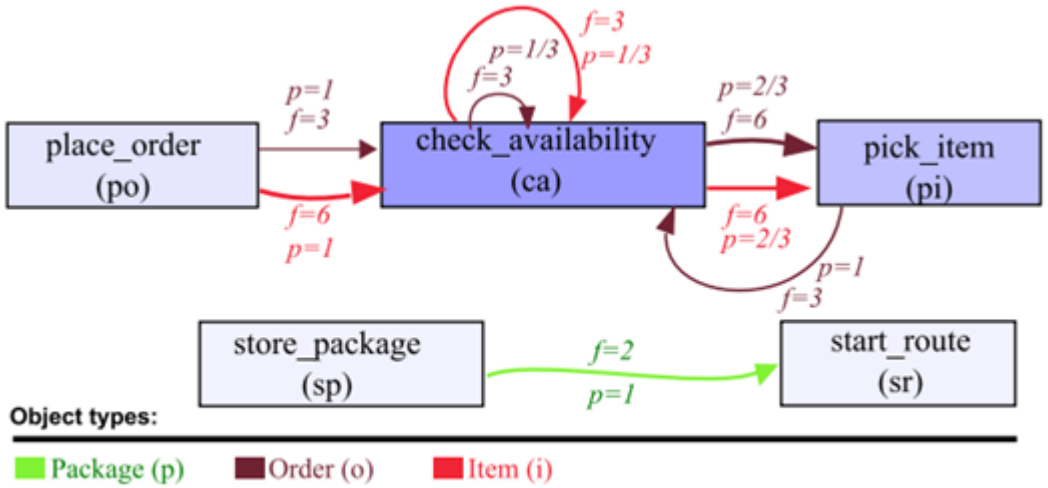


FIGURE 3. A Markov DFM of the DFM presented in FIGURE 2.

	ca	pi	po	sp	sr
ca	**1/3**	**2/3**	0	0	0
pi	0	0	0	0	0
po	**1**	0	0	0	0
sp	0	0	0	0	0
sr	0	0	0	0	0

(a) Probability of relations for Item

	ca	pi	po	sp	sr
ca	**1/3**	**2/3**	0	0	0
pi	**1**	0	0	0	0
po	**1**	0	0	0	0
sp	0	0	0	0	0
sr	0	0	0	0	0

(b) Probability of relations for Order

	ca	pi	po	sp	sr
ca	0	0	0	0	0
pi	0	0	0	0	0
po	0	0	0	0	0
sp	0	0	0	0	**1**
sr	0	0	0	0	0

(c) Probability of relations for Package

Calculating similarity

$$sim(\theta_1, \theta_2) \leftarrow \frac{\sum_{\forall t,t' \in T} \left(p(t, \theta_1, t') * p(t, \theta_2, t')\right)}{\sum_{\forall t_1,t_2 \in T} \left(\frac{p(t_1, \theta_1, t_2)^2 + p(t_1, \theta_2, t_2)^2}{2}\right)} \quad (2)$$

TABLE 2. Calculated Similarity Matrix that shows the similarity of the process for object type pairs.

	o	i	p
o	1.0	0.76	0.0
i	0.76	1.0	0.0
p	0.0	0.0	1.0

	ca	pi	po	sp	sr
ca	**1/3**	**2/3**	0	0	0
pi	0	0	0	0	0
po	**1**	0	0	0	0
sp	0	0	0	0	0
sr	0	0	0	0	0

(a) Probability of relations for Item

	ca	pi	po	sp	sr
ca	**1/3**	**2/3**	0	0	0
pi	**1**	0	0	0	0
po	**1**	0	0	0	0
sp	0	0	0	0	0
sr	0	0	0	0	0

(b) Probability of relations for Order

	ca	pi	po	sp	sr
ca	0	0	0	0	0
pi	0	0	0	0	0
po	0	0	0	0	0
sp	0	0	0	0	**1**
sr	0	0	0	0	0

(c) Probability of relations for Package

Identifying clusters by a threshold

Algorithm 1: discoverClusters

Data: $\big((OT, T, R, f), p, sim\big), threshold$
Result: *clusters*
begin
  $clusters \leftarrow \{\}$;
  foreach $\theta_1, \theta_2 \in OT$ **do**
    if $sim(\theta_1, \theta_2) >= threshold$ **then**
      $X \leftarrow$
      $\bigcup_{C \in clusters} \{C | \{\theta_1\} \subseteq C \ \vee \ \{\theta_2\} \subseteq C\}$;
      $clusters \leftarrow$
      $clusters \backslash X \ \cup \ \{\{\bigcup_{C \in X} C \ \cup \ \{\theta_1, \theta_2\}\}\}$;
  return *clusters*;

TABLE 3. Filtered similarity matrix and Identified clusters for the running example by setting different thresholds.

	o	i	p
o	1.0	0.76	0.0
i	0.76	1.0	0.0
p	0.0	0.0	1.0

(a) 1 cluster when threshold=0, i.e., $\{\{i, o, p\}\}$

	o	i	p
o	1.0	0.76	
i	0.76	1.0	
p			1.0

(b) 2 clusters when threshold=0.01, i.e., $\{\{i, o\}, \{p\}\}$

	o	i	p
o	1.0		
i		1.0	
p			1.0

(c) 3 clusters when threshold=0.77, i.e., $\{\{i\}, \{o\}, \{p\}\}$

Threshold Tuning

Algorithm 2: tuneClusters

Data: $(M, threshold, res)$ such that M is a
Result: res such that
begin
  if $res = \{\}$ **then**
    $res \leftarrow \{(0, discoverClusters(M, 0))\}$;
    $res \leftarrow res \cup \{(1, discoverClusters(M, 1))\}$;
    return $tuneClusters(M, 0.5, res)$;
  else
    if $(threshold, \_) \in res$ **then**
      return res;
    else
      $CT \leftarrow discoverClusters(M, threshold)$;
      $res \leftarrow res \cup \{(threshold, CT)\}$;
      $u \leftarrow min\{i \mid \forall_{(i,-)\in res}\ i > threshold\}$;
      $l \leftarrow max\{i \mid \forall_{(i,-)\in res}\ i < threshold\}$;
      if $|\{C|\forall_{(t,C)\in res}\ t = u\}| \neq |CT|$ **then**
        $t \leftarrow round((threshold + u)/2, 2)$;
        $res \leftarrow$
        $res \cup \{(t, discoverClusters(M, t))\}$;
      if $|\{C|\forall_{(t,C)\in res}\ t = l\}| \neq |CT|$ **then**
        $t \leftarrow round((threshold + l)/2, 2)$;
        $res \leftarrow$
        $res \cup \{(t, discoverClusters(M, t))\}$;
      return res;

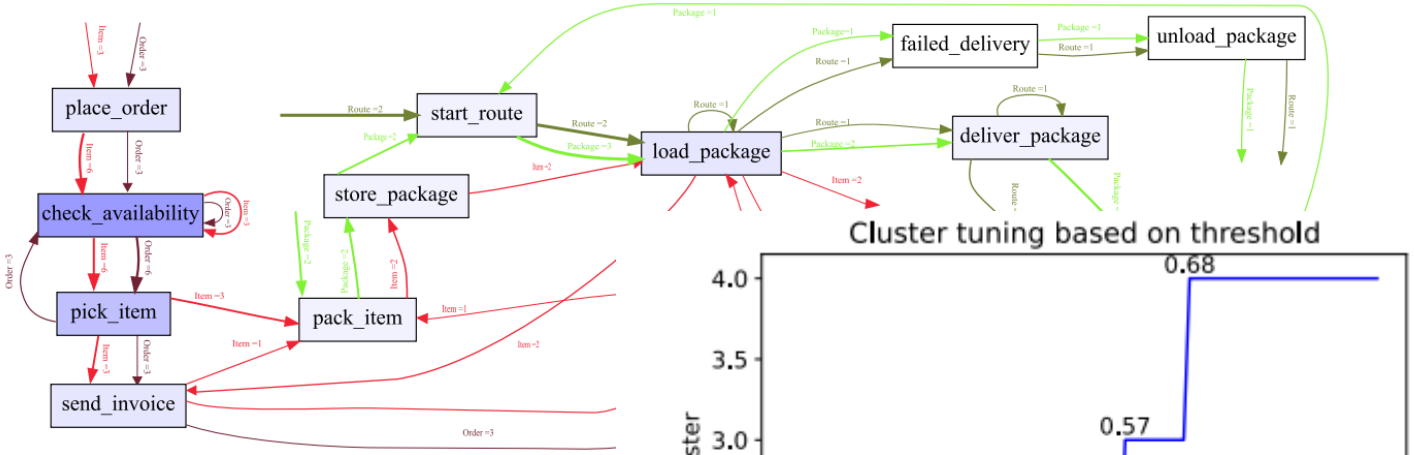


FIGURE 1. A Directly-Follows Multigraph (DFM), discovered from 39 e
complex.

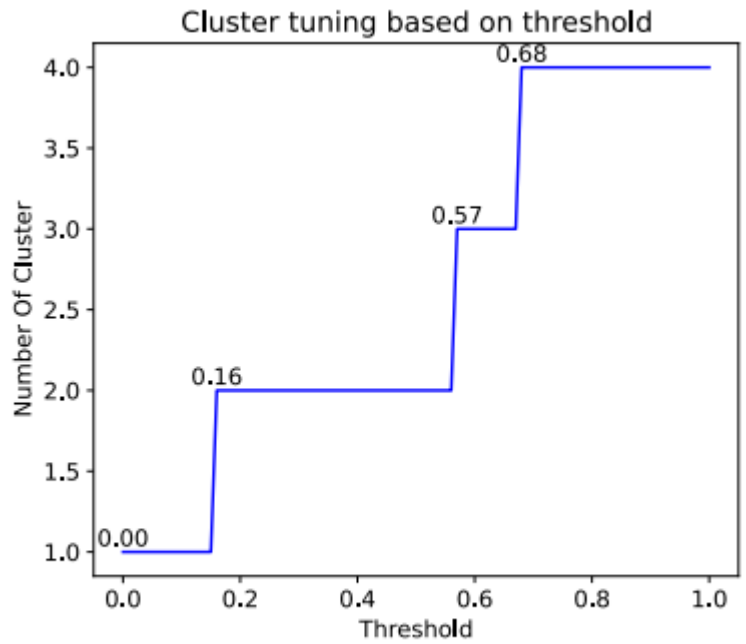
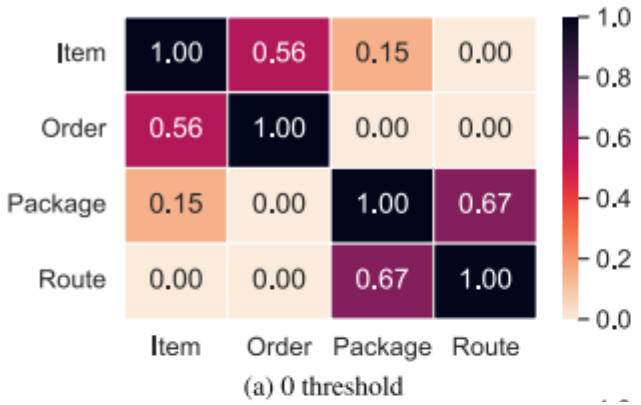


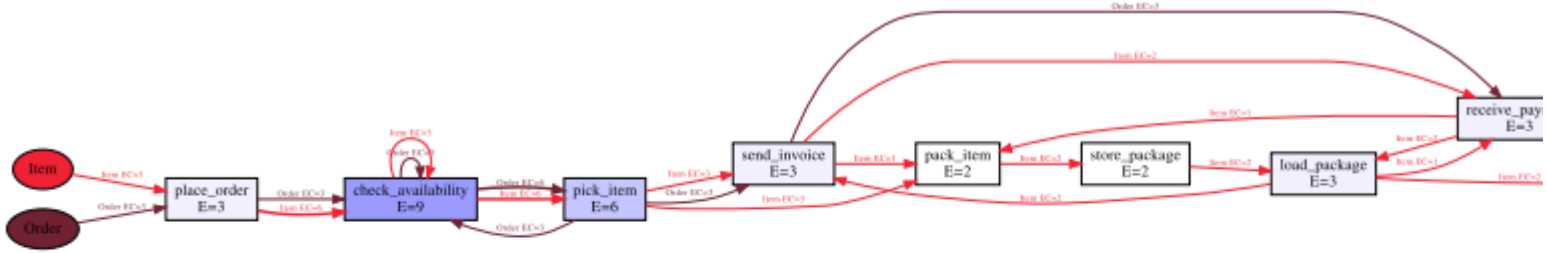
FIGURE 4. The cluster tuning result for DFM in FIGURE 1.

How to flat based on similar object types

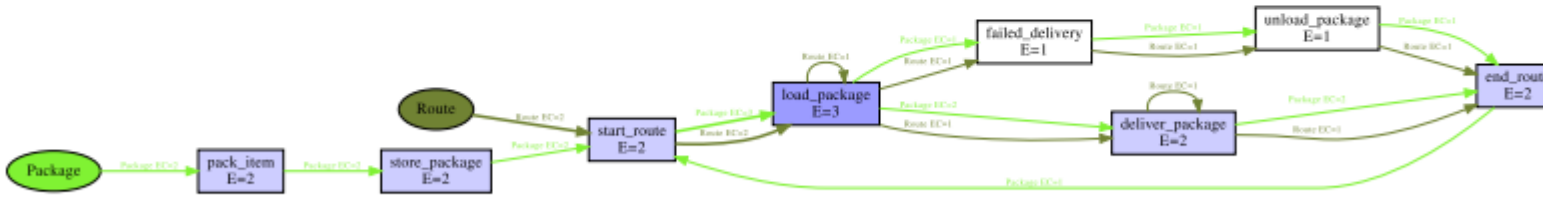


(a) 0 threshold

(b) 0.16 threshold



(a) DFM for the cluster that include Order and Item object types



(b) DFM for the cluster that include Package and Route object types

FIGURE 6. **Discovered DFMs based on two identified clusters by a similarity threshold of 0.16. The figure is made intentionally small just to show supporting the separation of similar object types.**

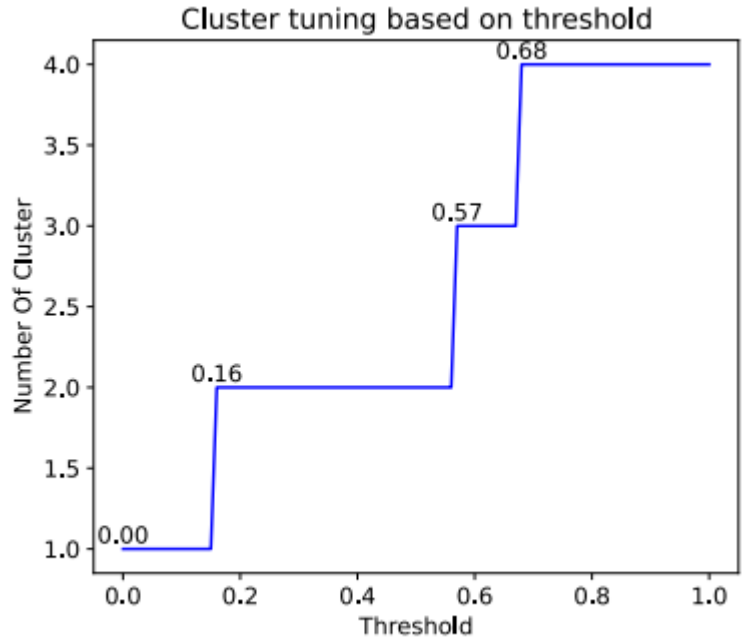


FIGURE 4. **The cluster tuning result for DFM in FIGURE 1.**

Let’s try it together!

https://github.com/jalaliamin/ResearchCode/blob/main/Invited_Lectures/2023Bayreuth/Exercise.ipynb

The End