Similarity Resonance For Improving Process Model Matching Accuracy

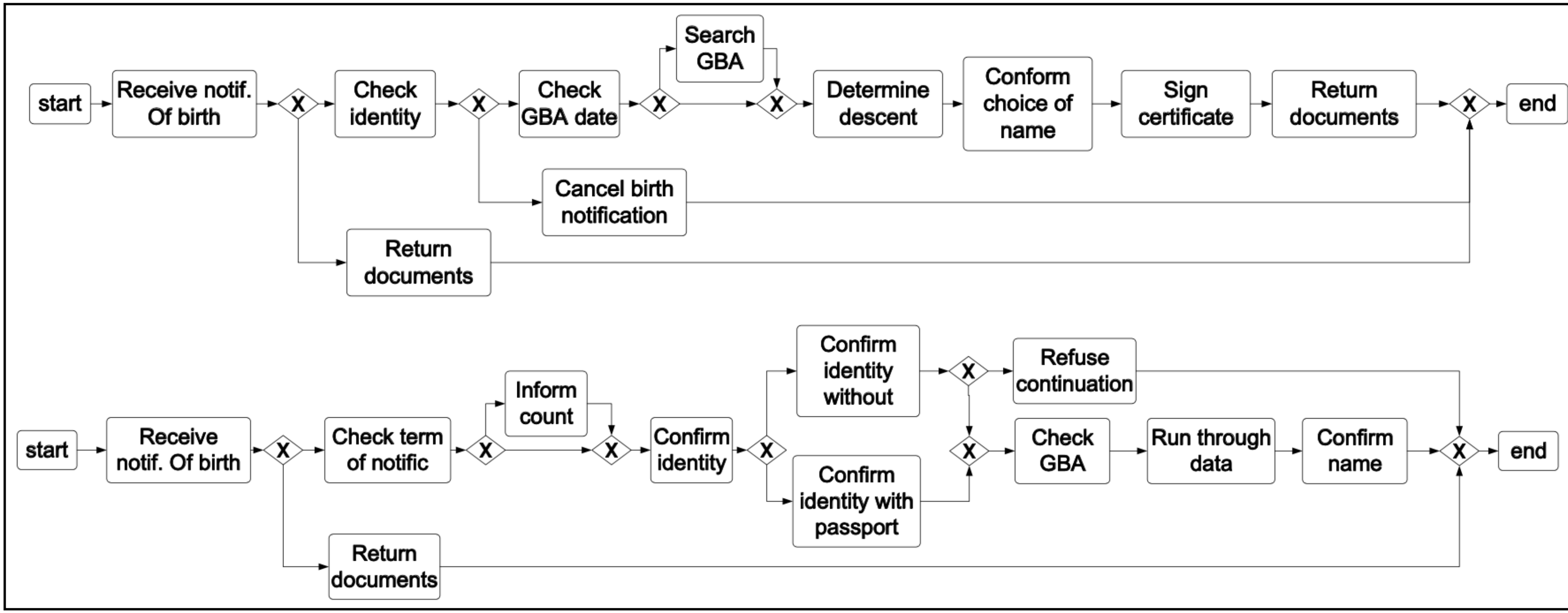
Presented By:

Mukander Kumar
M.Sc. Data Science

Research Background

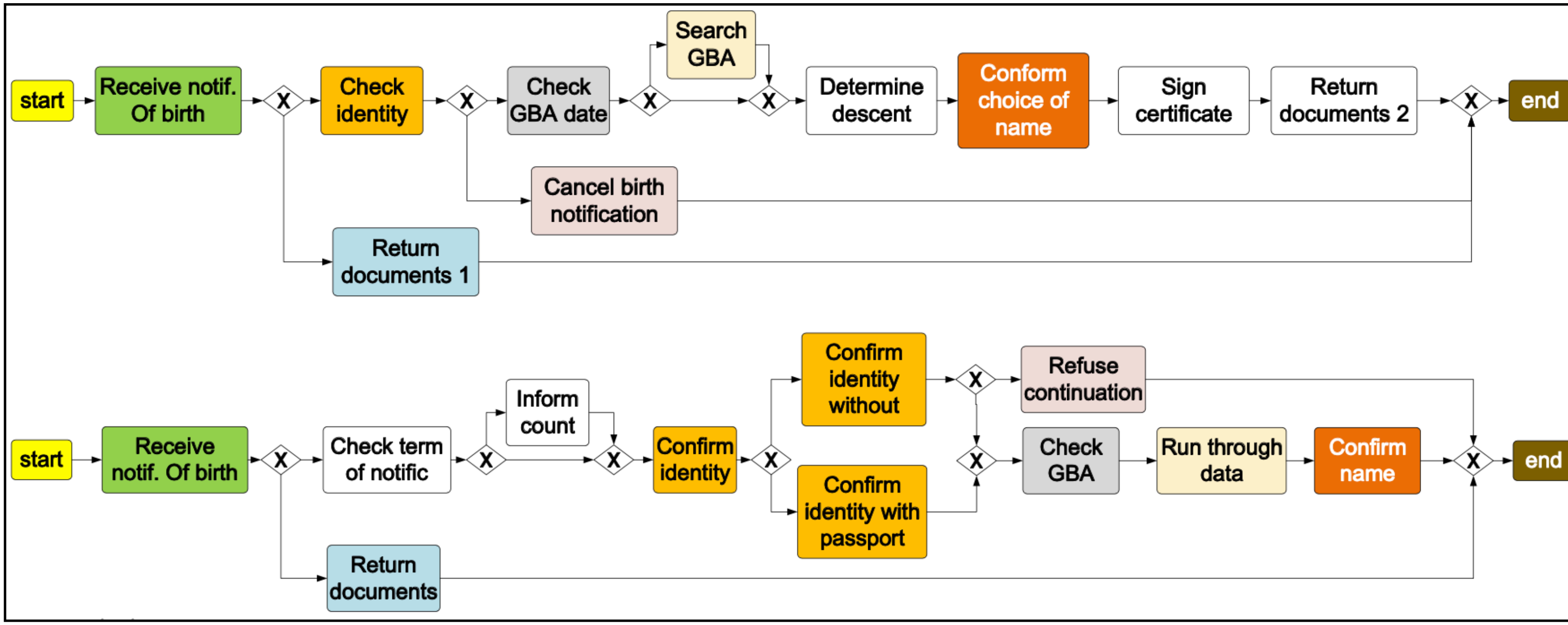
- Process Model Matching:
  - Given 2 process models, creates a mapping between process activities based on a similarity notion.
  - Many applications such as:
    - Querying large process model repositories (e.g. retrieving processes similar to a given one)
    - Refactoring process model repositories (e.g. merging similar processes to reduce redundancy, creating a version management system of process models)

Example - Birth Registration Process



Example - Birth Registration Process

- Matching According to Gold Standard



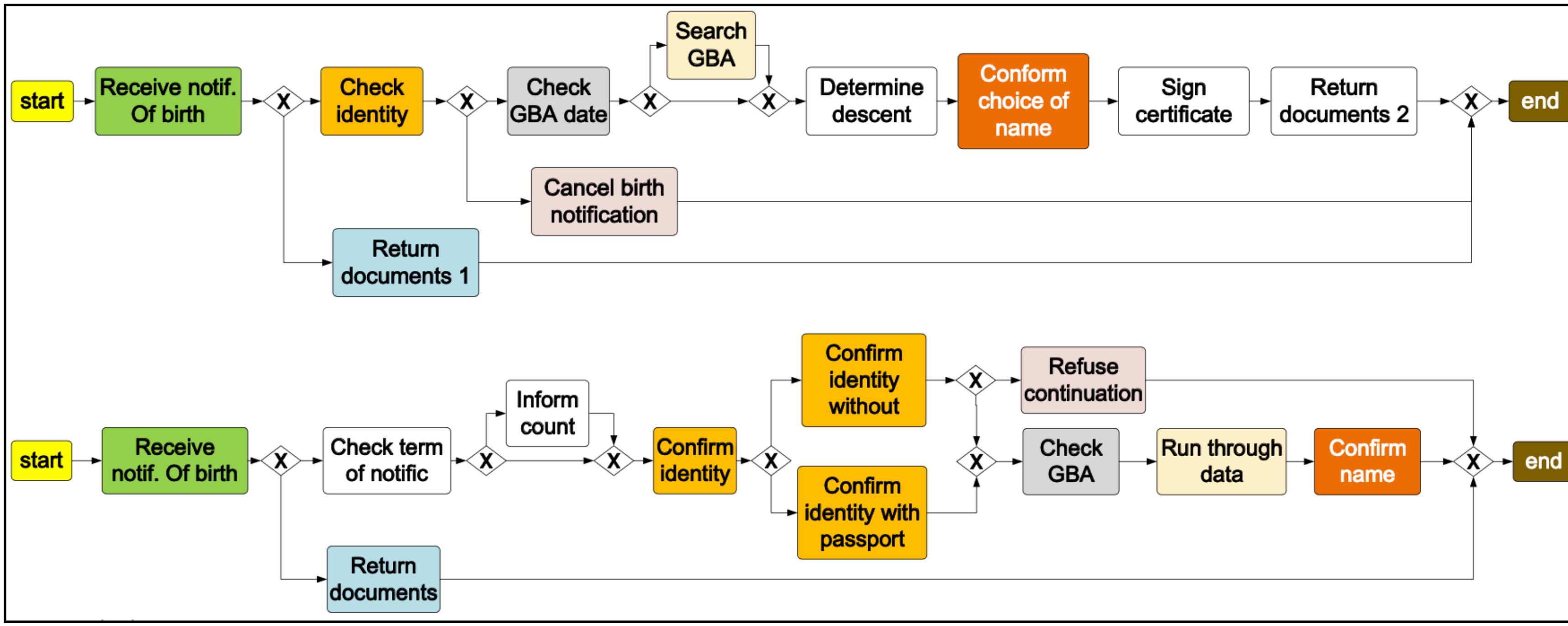
Process Model Matching - 2 Steps

1. Compute pairwise similarity between process activities.

2. Derive an optimal mapping based on the computed similarity.

Example - Birth Registration Process

- Matching According to Gold Standard



1) Compute pairwise similarity between process activities.

Process 1	Process 2	Sim
Start	Start	1
Start	Receive notif. Of birth	0.04
…	…	
Receive notif of birth	Receive notif of birth	1
…	…	
Return documents 1	Return documents	1
Return documents 2	Return documents	1
…	…	
Search GBA	Run through data	0.14
Cancel birth notification	Refuse continuation	0.13
Cancel birth notification	Receive notif of birth	0.35

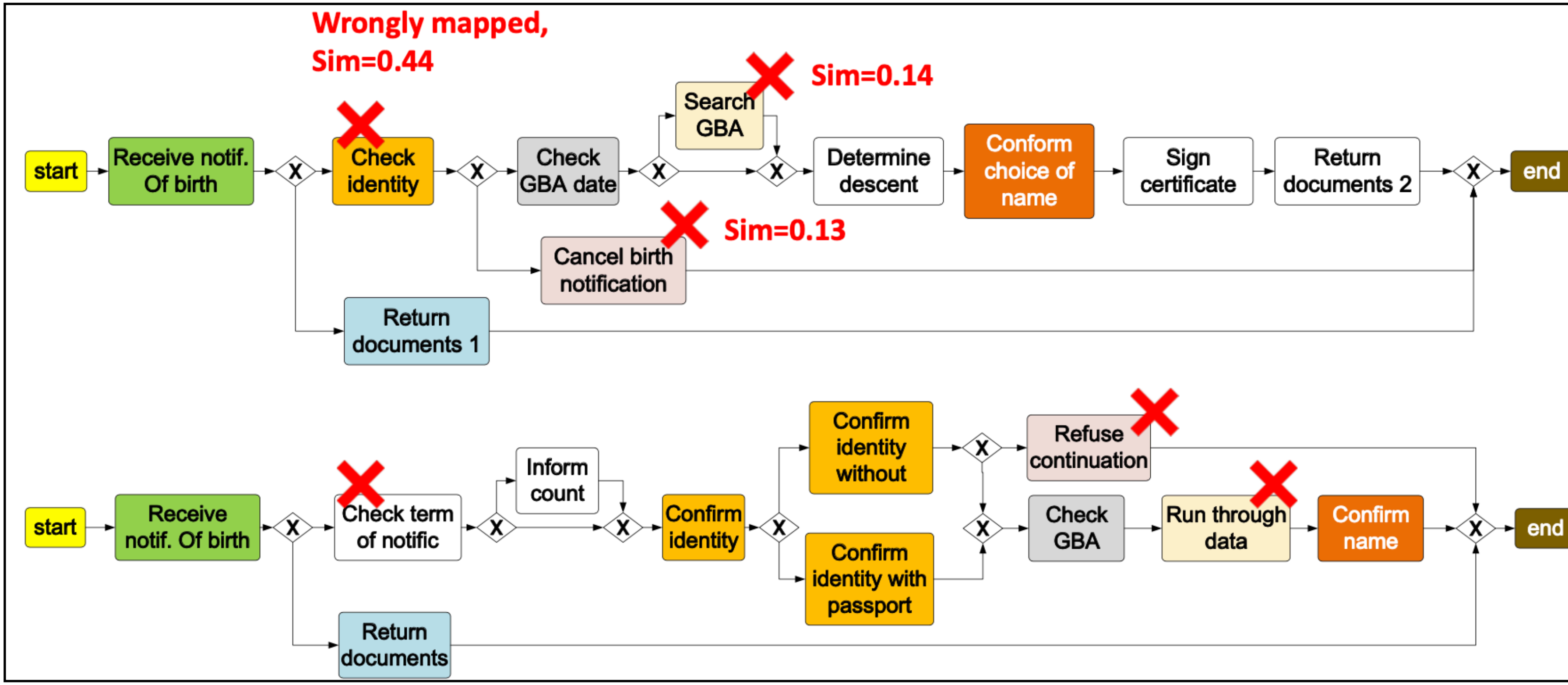
- Computed based on activities labels
  - Syntactic(e.g.Levenshtein
  - distance)
  - Linguistic(Using WordNet)

1) Derive an optimal mapping based on the computed similarity.

Process 1	Process 2	Sim
Start	**Start**	**1**
Start	Receive notif. Of birth	0.04
…	…	
Receive notif of birth	**Receive notif of birth**	**1**
…	…	
Return documents 1	**Return documents**	**1**
Return documents 2	Return documents	1
…	…	
Search GBA	Run through data	0.14
Cancel birth notification	Refuse continuation	0.13
Cancel birth notification	Receive notif of birth	0.35

- Using a matching algorithm with a similarity threshold:
- Graph edit distance, threshold=0.3
- Greedy algorithms:
  - all pairs with a similarity above threshold (1:n, 1:1 mapping)

Example - Birth Registration Process



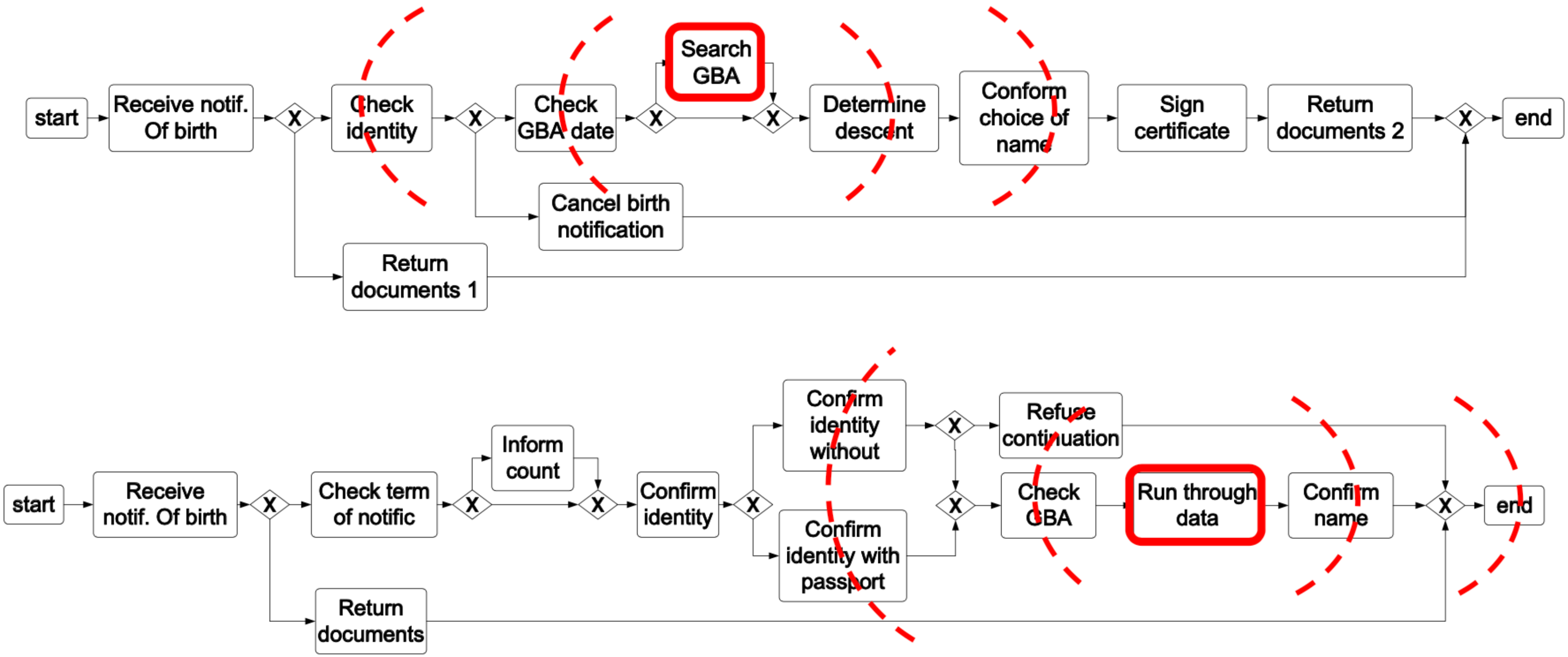
Research Problem

- If two activities are classified as highly (respectively slightly) similar in the **first** step, they get a high(respectively low) chance to be mapped in the **second** step

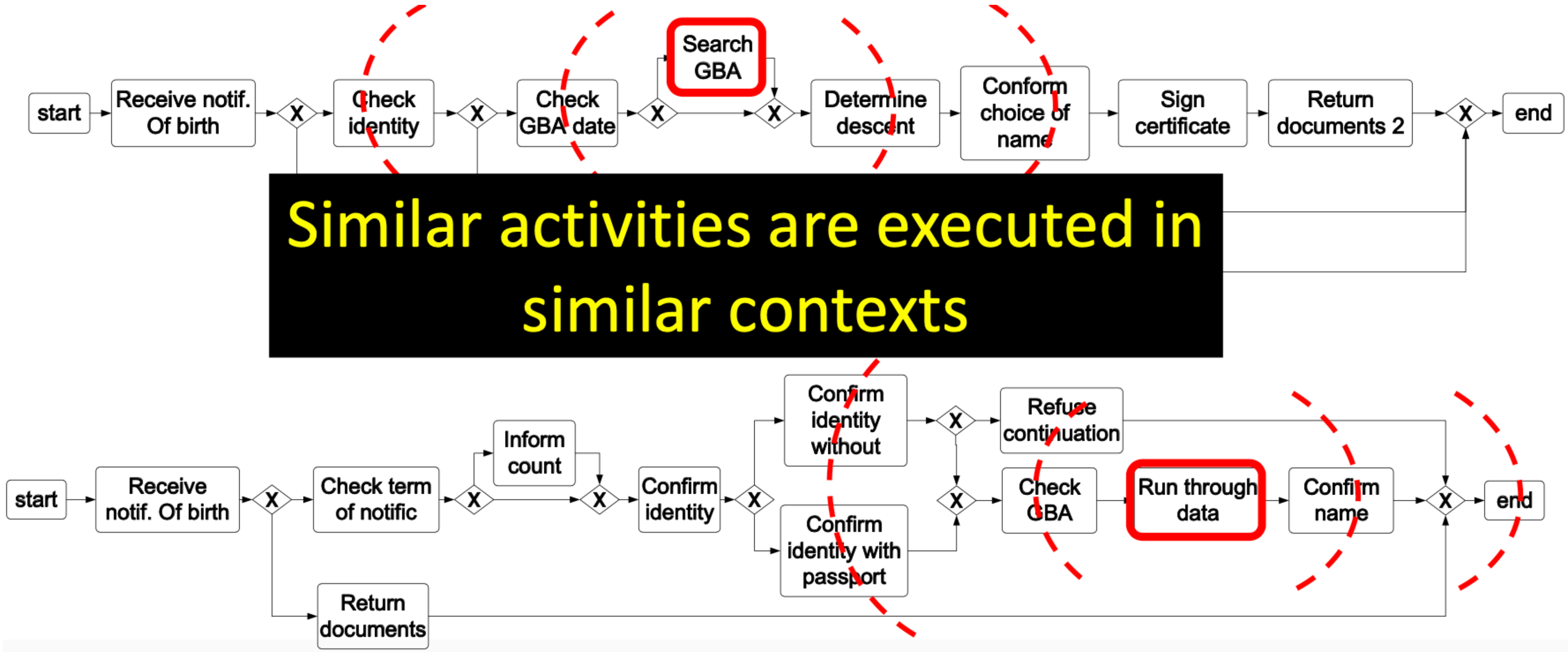
- Improving the accuracy of activities’ similarity, improves the accuracy of process model matching.

- **How to compute an accurate similarity between activities?**

Observations that motivates our work



Observations that motivates our work



Similar activities are executed in similar contexts

Our Approach: Similarity Resonance

- Similarity between 2 activities is computed based on the similarity between neighboring activities
- Similarity between neighboring activities is computed based on the similarity between their neighbors
- Iterative system in which similarity values are computed and updated until convergence —> Global contextual similarity.

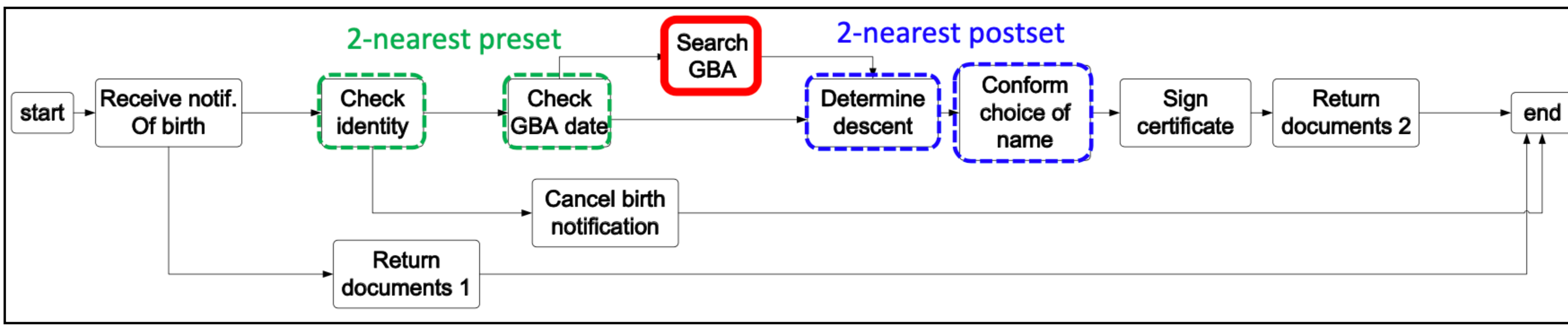
Approach Overview

- 1) Derive k-nearest neighbors
  - Local context, contributes directly to the similarity of 2 activities
- 2) Construct k-resonance graph
  - How similarity propagates between contexts
- 3) Compute similarity resonance
  - How similarity is computed and how it converges

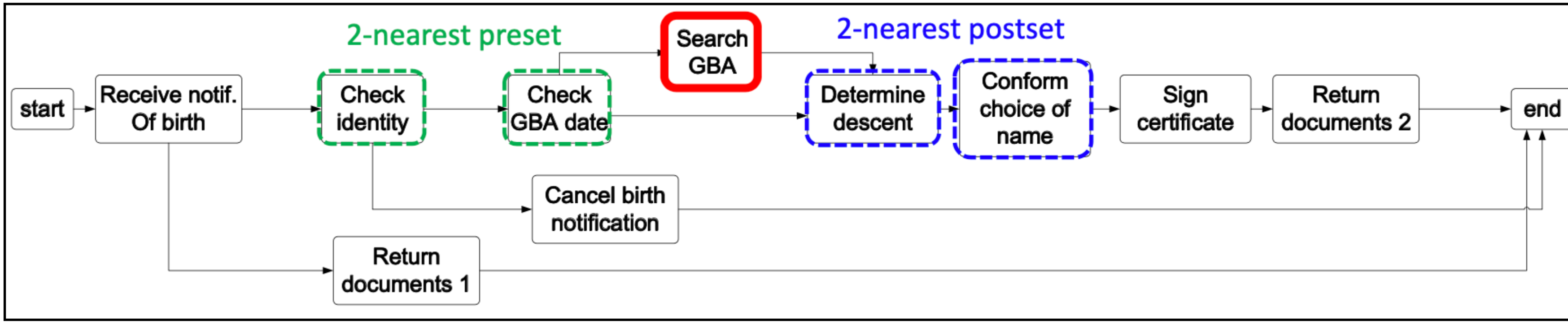
1) k-nearest neighbors of an activity “a"

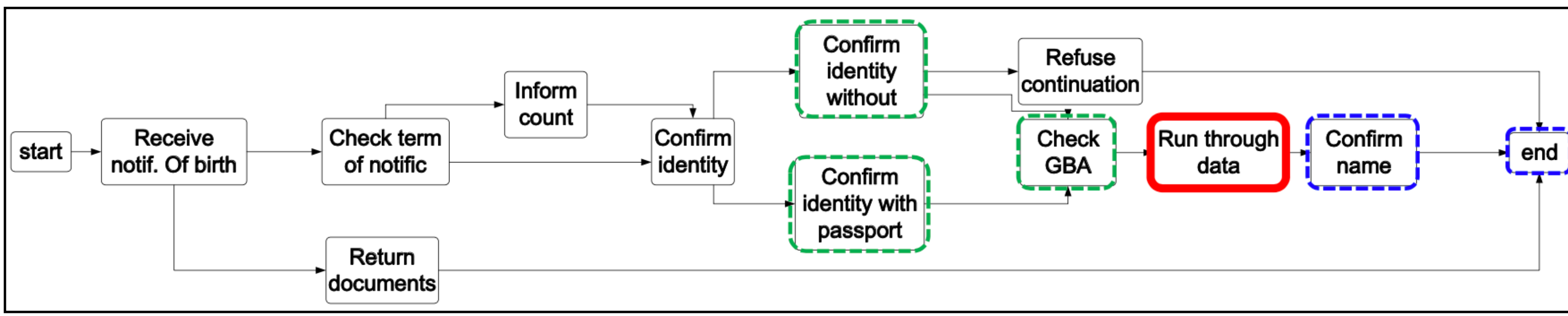
- a’ is **k-nearest preset** if
  - exists a path from a’ to a
  - length of the shortest path from a’ to a is less or equal to k

- E.g. 2-nearest neighbours



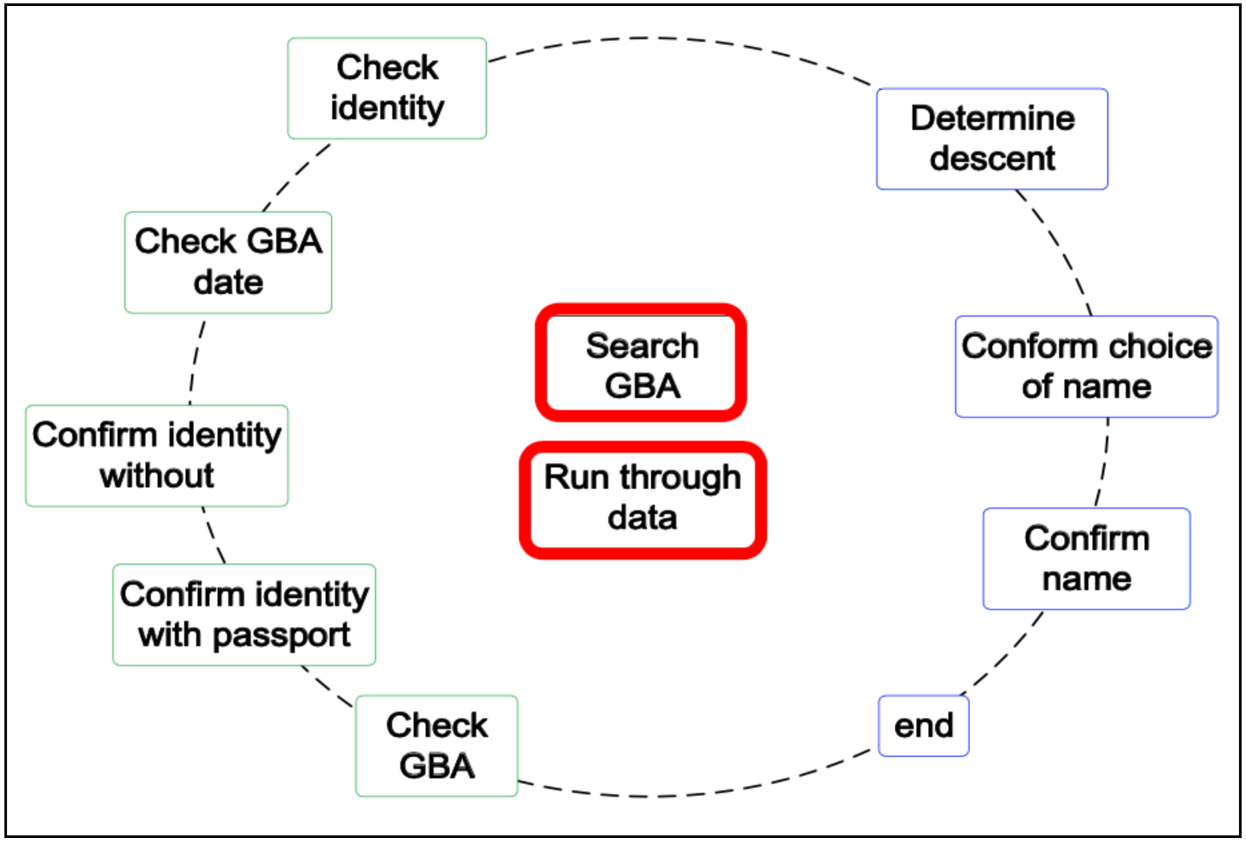
1) k-nearest neighbors of an activity “a"



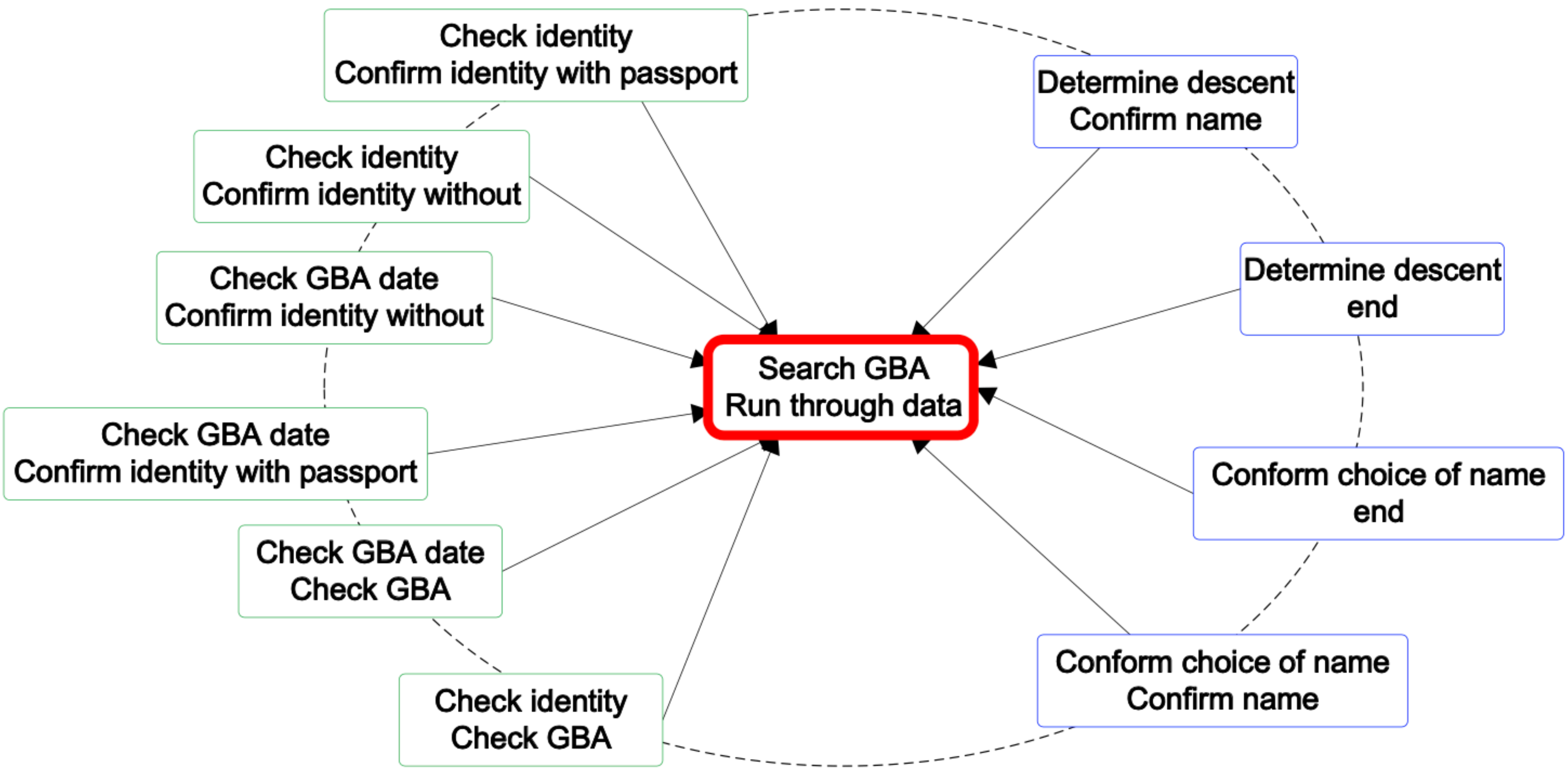


1) k-nearest neighbors of an activity “a"

- k-nearest neighbors of two activities contribute directly to their similarity



2) k-resonance graph (weighted directed graph)



2) k-resonance graph (weighted directed graph)

- How to define the similarity propagation factor(weight of edges)??
  - Less Likely similar activities will receive less amount of similarity
  - More Likely similar activities will receive more amount of similarity

- How to identify likely similar activities?

$a \in M_1$, $a' \in M_2$, Sim_L is a label similarity measure

$$P_1(a,a') = \frac{Sim_L(a,a')}{\sum_{a'_2 \in P_2} Sim_L(a,a'_2)}$$

$$P_2(a',a) = \frac{Sim_L(a,a')}{\sum_{a'_1 \in P_1} Sim_L(a',a'_1)}$$

$$P(a,a') = \max(P_1, P_2)$$

Likely similar if $P(a,a') \geq t_P$

t_P user defined threshold

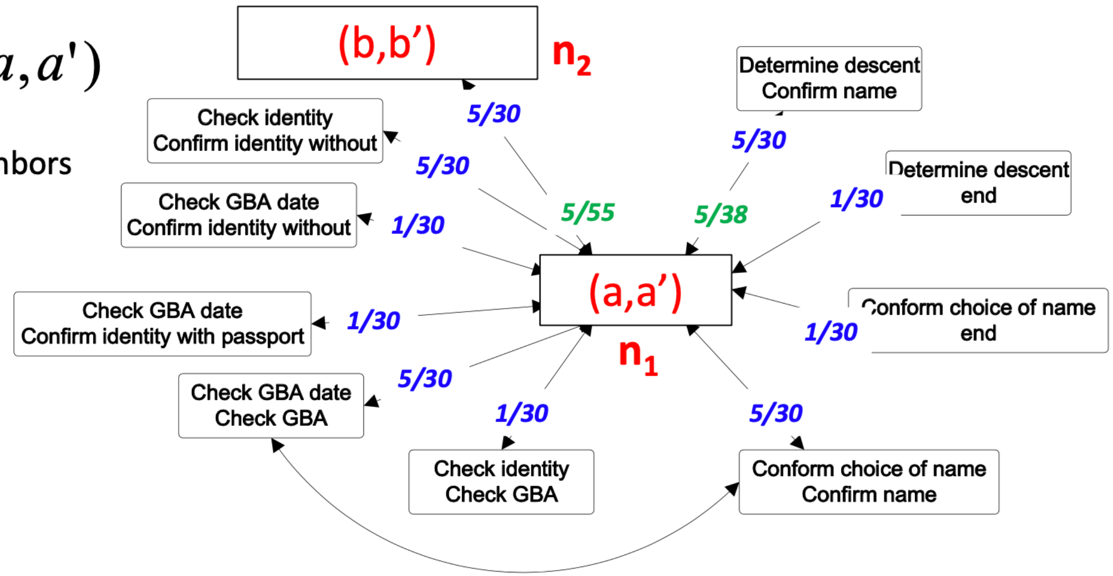
3) Compute Similarity Resonance

At iteration i+1

$$Sim_{i+1}(a,a') = \boxed{\alpha\left(\sum_{n_2 \in \bullet n_1} W((n_2,n_1))Sim_i(b,b')\right)} + \boxed{(1-\alpha)Sim_L(a,a')}$$

$$Sim_{i+1}(a,a') = Sim_L(a,a')$$

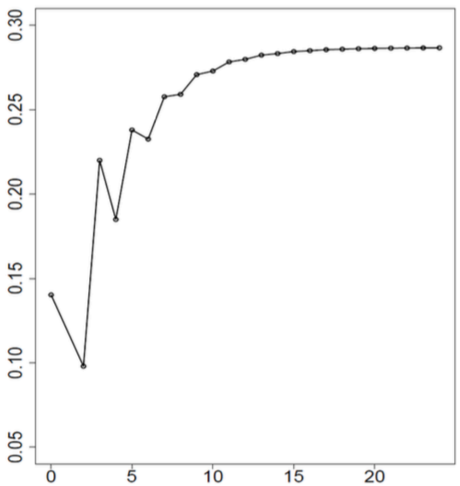
If node (a,a') doesn't have neighbors

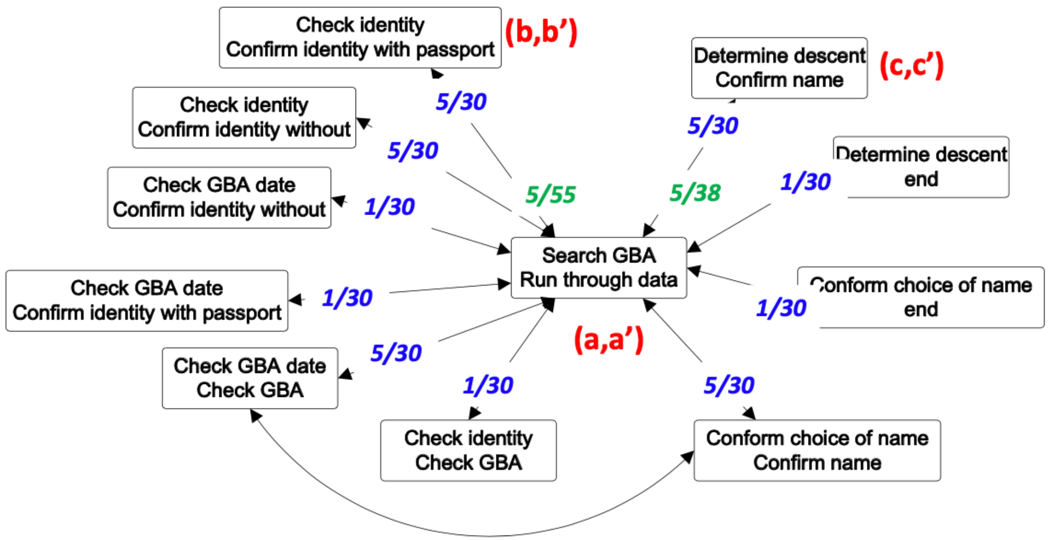


3) Compute Similarity Resonance

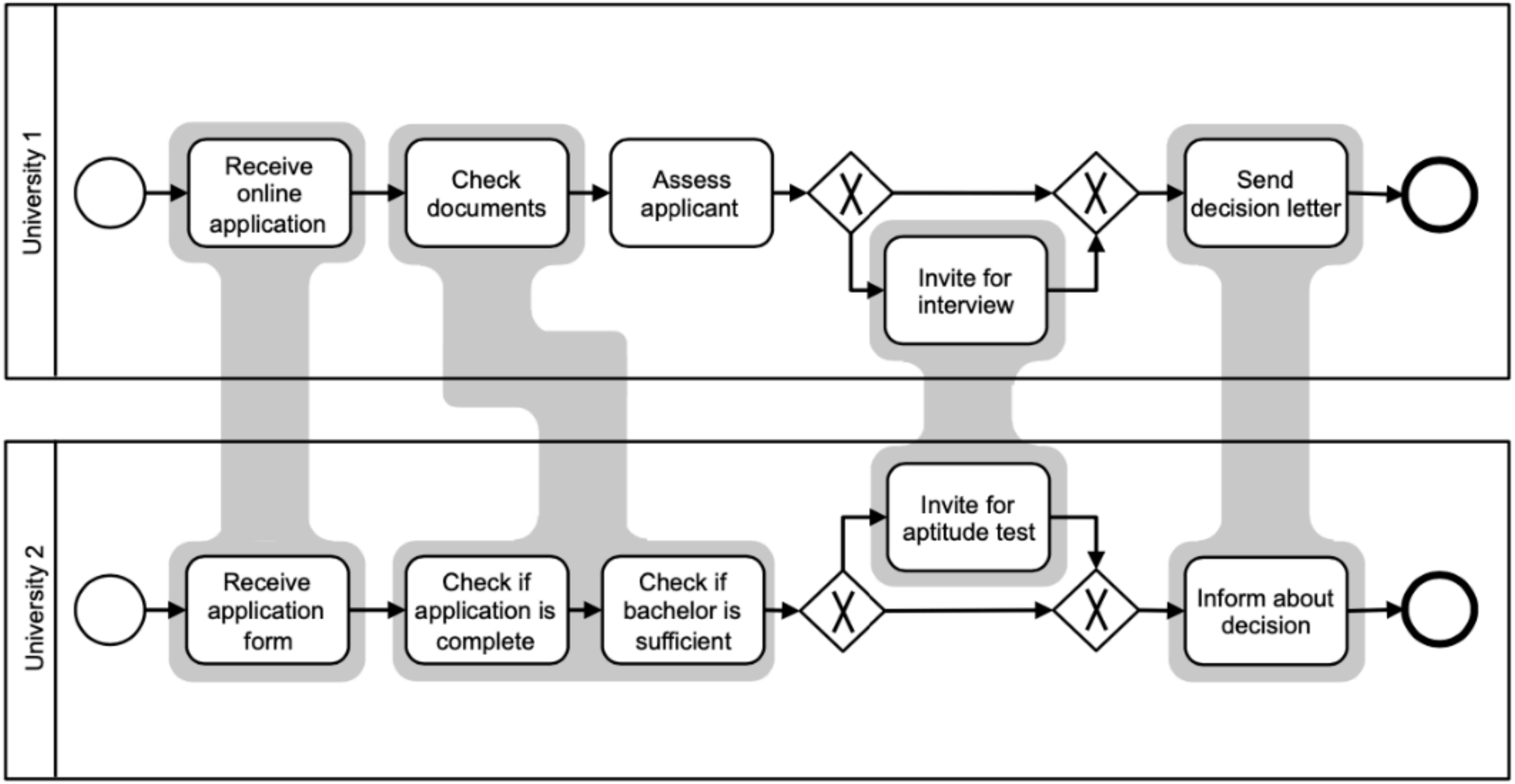
At iteration i=0 $Sim_0 = Sim_L$

At iteration i=1 $Sim_1(a,a') = 0.9 \times \left(\frac{5}{55} Sim_0(b,b') + \frac{5}{38} Sim_0(c,c') + \ldots \right) + 0.1 \times Sim_L(a,a')$





Application



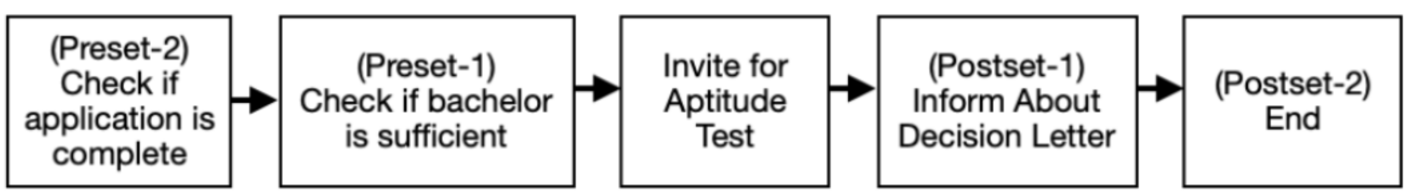
Our Approach

- Derive k-nearest neighbors for the activity pair
- Construct k-resonance graph and compute similarity resonance

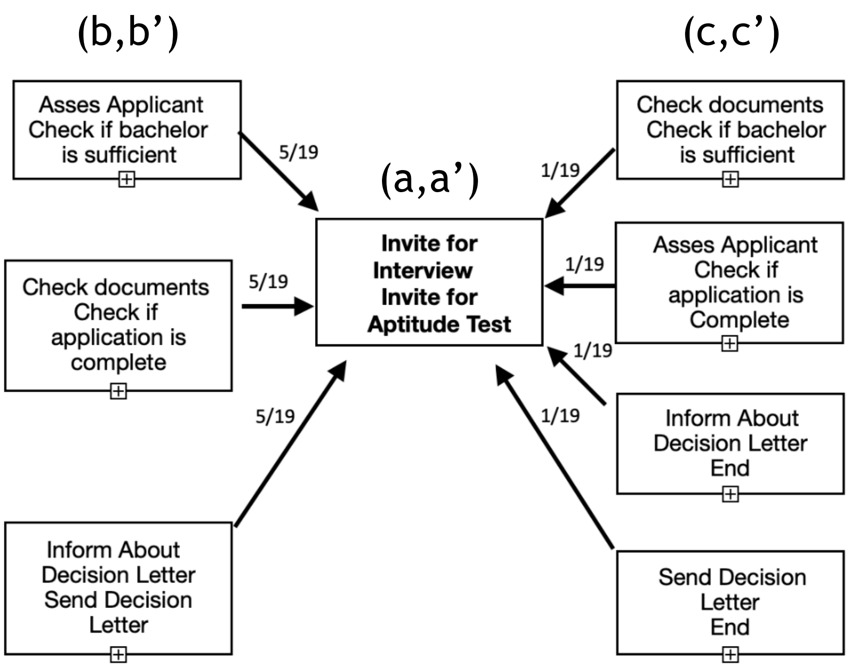
k-nearest neighbors of an activity

- 2-Nearest Neighbors of Invite for Interview

- 2-Nearest Neighbors of Invite for Aptitude Test



k-resonance graph and Similarity Resonance



- At Iteration i=0

  $Sim_0 = Sim_L$

- At Iteration i=1

$Sim_1(a,a') = 0.9 \times (5/15\, Sim0(b,b') + 5/38\, Sim0(c,c') + \ldots) + 0.1 \times Sim_L(a,a')$

Results

- Similarity Score Sim_L and Sim_R

$id : (n_1, n_2) \in N_1 \times N_2$	Sim_L	Sim_R
1: (Receive notif. of birth, Receive notif. of birth)	1.0	0.93
2: (Return documents 1, Return documents)	1.0	0.86
-: (Return documents 2, Return documents)	1.0	0.21
3: (Check identity, Confirm identity)	0.52	0.47
3: (Check identity, Confirm identity without)	0.52	0.47
3: (Check identity, Confirm identity with passport)	0.52	0.45
4: (Check GBA date, Check GBA)	0.82	0.7
5: (Search GBA, Run through data)	0.14	0.29
6: (Cancel birth notification, Refuse continuation)	0.13	0.31
-: (Cancel birth notification, Receive notif. of birth)	0.35	0.07
7: (Conform choice of name, Confirm name)	0.42	0.36

Results of Application

- Similarity Score Sim_L and Sim_R

$id:(n_1,n_2) \in N_1 \times N_2$	Sim_L	Sim_R
1: Receive Online Application, Receive Application Form	0.14	0.4
2: Check Documents,Check if application is complete	0.12	0.47
3: Check Documents,Check if bachelors is sufficient	0.32	0.3
4: Invite for Interview,Invite for aptitide test	0.34	0.24
5: Send decision letter,Inform about decision	0.2	0.31
6: Asses Applicant,Check if bachelor is sufficient	0.1	0.41
7:Asses Applicant,Check if application is complete	0.01	0.45

Evaluation

- 3 freely available process model datasets:
  - birth registration (petri net) - 9 process models, 36 process pairs to match
  - university admission (bpmn) - 9 process models, 36 process pairs to match
  - asset management (EPC) – 36 pairs of process models retrieved from SAP
- 2 experiments:
  - Accuracy of process matching using standard label similarity vs using resonance similarity
  - Comparison with best results from PMMC’16 and OAEI’16 challenges

Evaluation

- Label Similarity:
  - bag-of-words technique combined with maximum between Levenshtein distance and Lin metric as a similarity measure
- Matching algorithms:
  - graph edit distance with default values
  - 2 variants of a greedy algorithm (1:n and 1:1 mapping)

1) Label vs Resonance Similarity

	SAP			Uni			Birth		
Compared approaches	P_R/P (std.)	R_R/R (std.)	F_R/F (std.)	P_R/P (std.)	R_R/R (std.)	F_R/F (std.)	P_R/P (std.)	R_R/R (std.)	F_R/F (std.)
Resonance/Label	**.82/.75** (.27/.31)	**.90/.78** (.16/.24)	**.83/.73** (.22/.26)	**.70/.64** (.40/.44)	**.70/.56** (.31/.35)	**.60/.53** (.36/.39)	**.58/.49** (.27/.27)	**.57/.53** (.18/.19)	**.54/.46** (.19/.18)

2) Comparison with best results from PMMC'15 and OAEI'16

	SAP			Uni			Birth		
Compared approaches	P_R/P (std.)	R_R/R (std.)	F_R/F (std.)	P_R/P (std.)	R_R/R (std.)	F_R/F (std.)	P_R/P (std.)	R_R/R (std.)	F_R/F (std.)
Resonance/OAEI'16	-	-	-	**.77/.72** (.28/-)	**.78/.69** (.26/-)	**.75/.70** (.25/-)	-	-	-
Resonance/PMMC'15	**.82/.89** (.27/.31)	**.90/.52** (.16/.42)	**.83/.49** (.22/.43)	**.70/.64** (.40/.34)	**.70/.62** (.31/.31)	**.60/.60** (.36/.30)	**.58/.68** (.27/.18)	**.57/.47** (.18/.24)	**.54/.54** (.19/.22)

We used the same label similarity technique used by the best algorithm in OAEI'16

Conclusion

- Introduced a new activity similarity based on a global notion of contextual similarity
- Showed that global contextual similarity improves process model matching accuracy
- ….and outperforms local contextual similarity

FUTURE WORK

- Incorporate resource/data/time and behavioral dimensions into the similarity computation.

- Varying the k from k=2 for better accuracy based on contexts and behavioural aspects.

- Investigate the problem of prediction in process mining
  - This work's idea has been widely applied in social network and web prediction

THANKYOU