

Fast Comparative Analysis of Merge Tree Using Locality Sensitive Hashing

Weiran Lyu, Raghavendra Sridharamurthy, Jeff M. Phillips, Bei Wang



Problem Statement

Motivation

- Measuring similarity between objects is fundamental in data analysis.
- The need for meaningful and efficient similarity measures is especially important in studying time-varying scientific data and ensembles.
- LSH remains unexplored in topological data analysis and visualization.

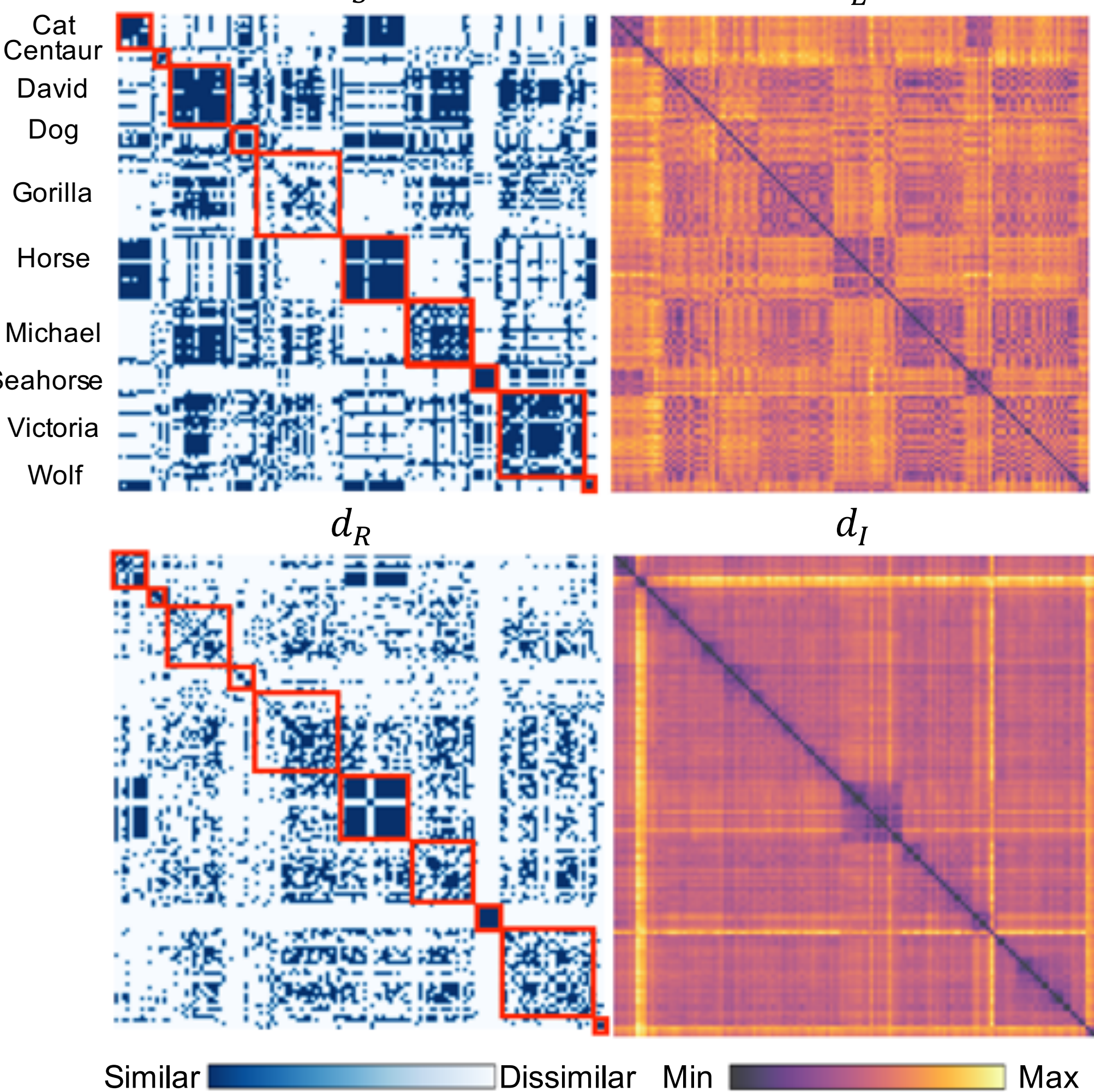
Challenges

- LSH framework for hierarchical data that can be extended to merge tree
- Labeling strategy that preserves topology
- LSH Parameter tuning: K , b , r
- Similarity matrix design

Main Contributions

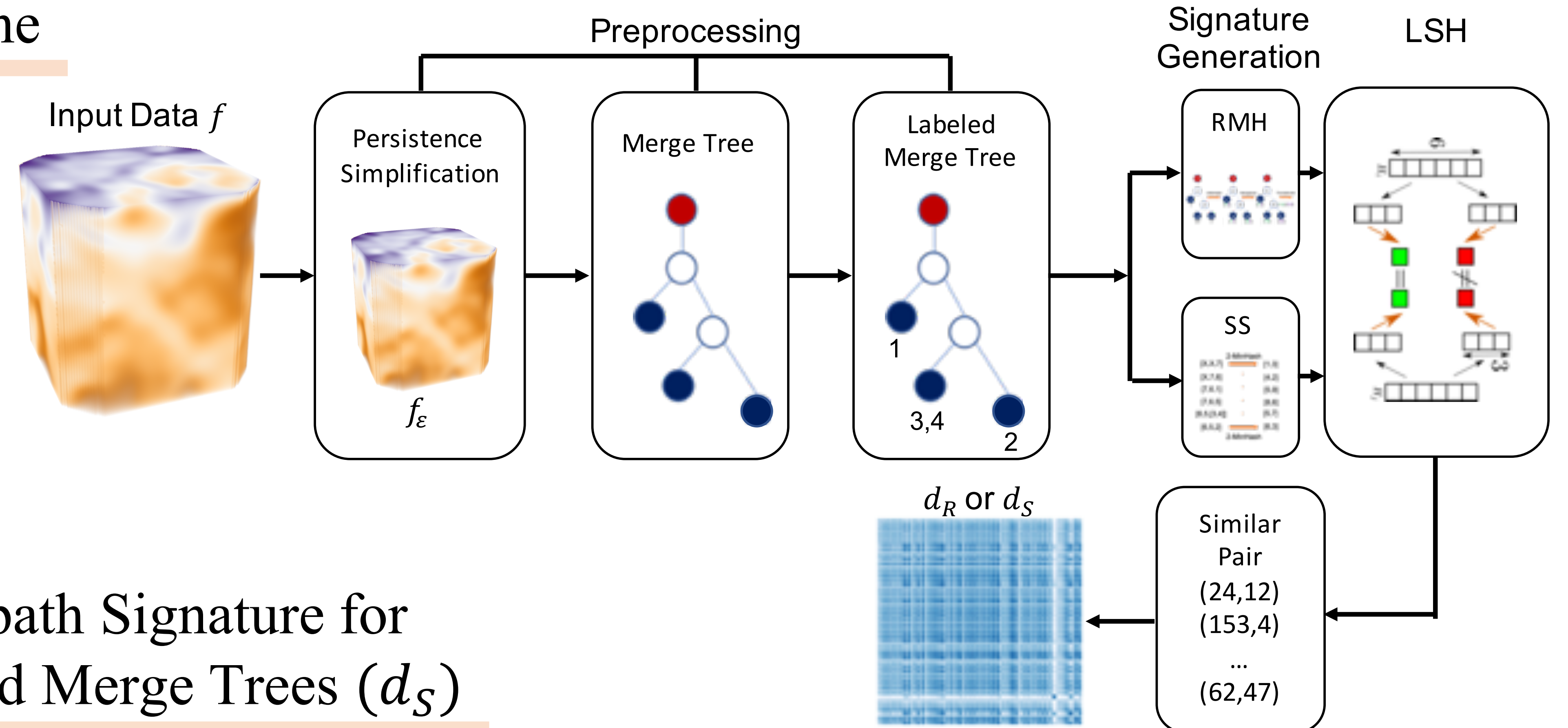
- Two new similarity measures for labeled merge trees that can be computed via LSH, using new extensions to Recursive MinHash and subpath signature, respectively.
- Our similarity measures are extremely efficient to compute and closely resemble the results of merge tree edit distance (d_E) or geometric interleaving distance (d_I)
- Efficiency with up to a 1000x speed-up

Results

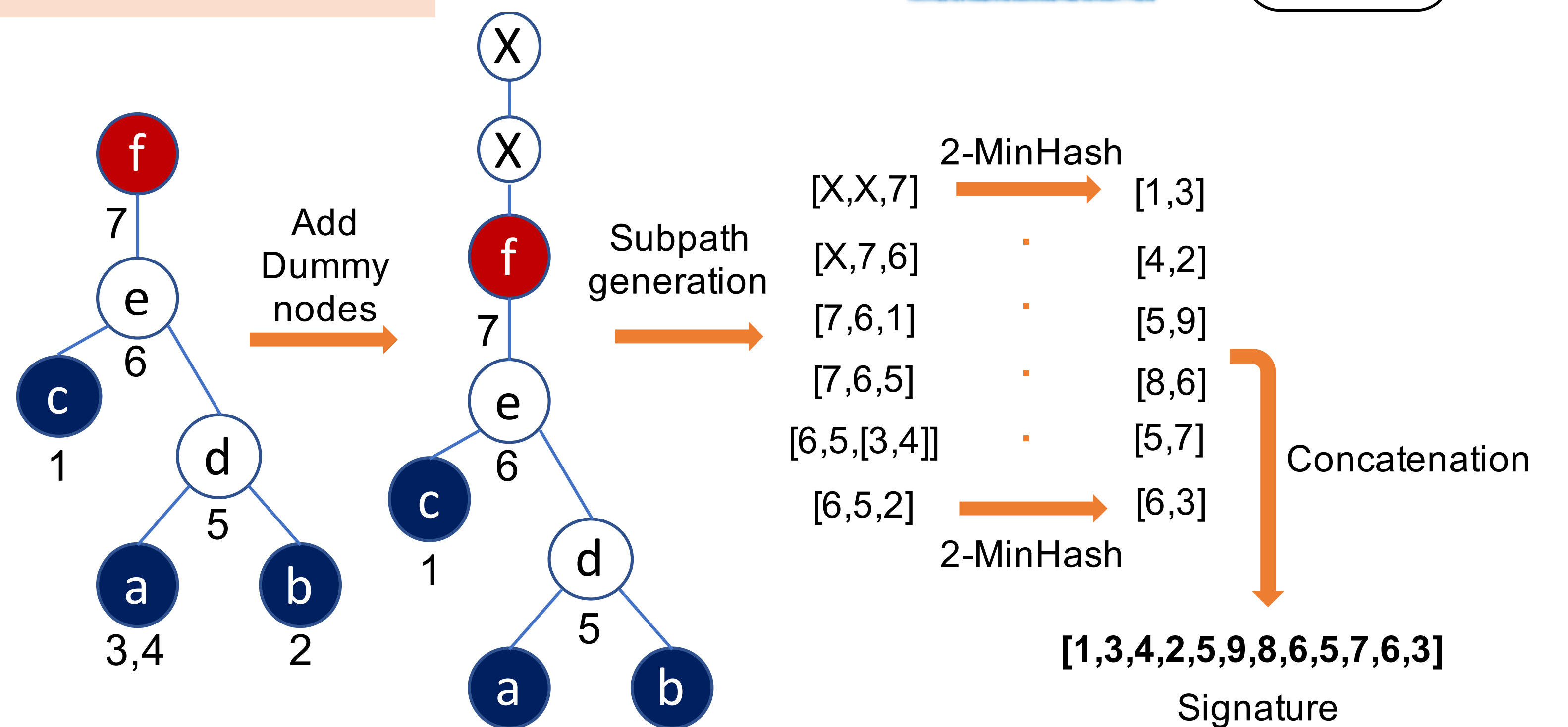


LSH for Comparing Merge Trees

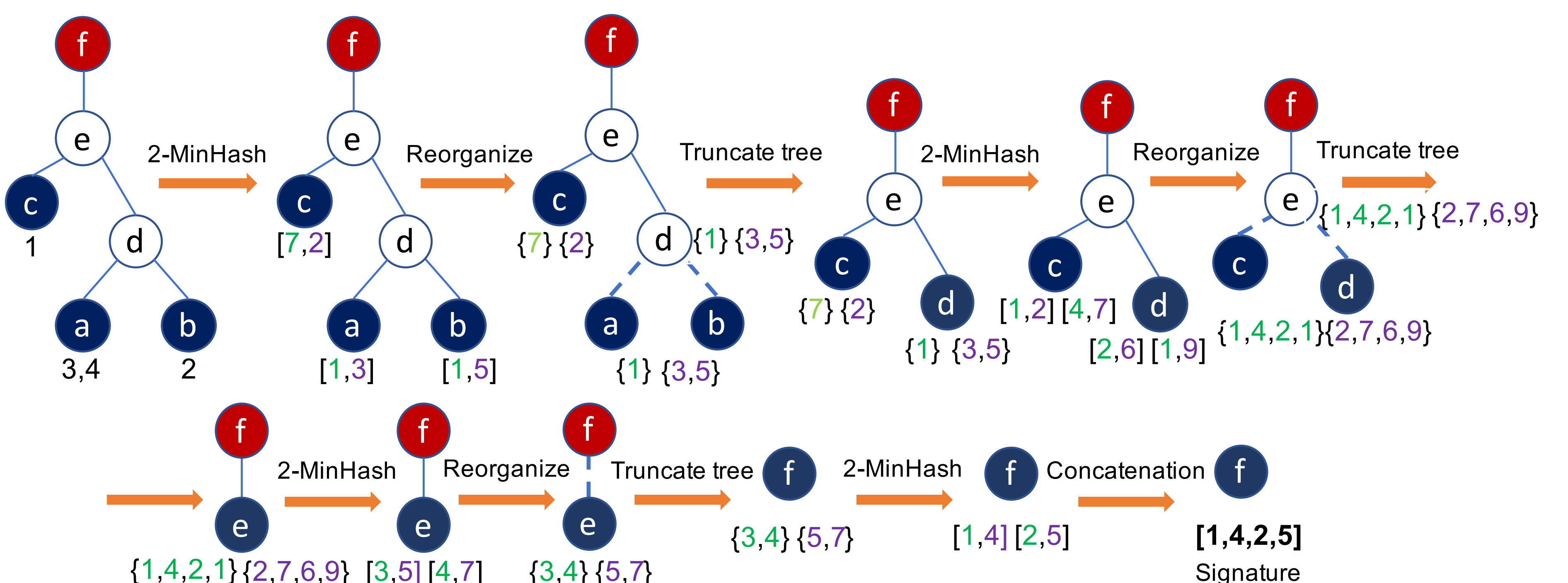
Pipeline



Subpath Signature for Labeled Merge Trees (d_S)



Recursive MinHash for Labeled Merge Trees (d_R)



Shape Matching

- Data: TOSCA shape dataset contains various non-rigid shapes in different poses
- Experiments: Aim to classify shapes correctly using labeled merge trees, irrespective of their poses
- Key result: d_S and d_R classifies most shapes into correct categories

Dataset	Method	K = 20		K = 40		K = 60		K = 80	
		LSH	Total	LSH	Total	LSH	Total	LSH	Total
Heated Flow (2000)	d_S	1.50	2660.7	1.72	2661.0	2.03	2661.3	2.31	2661.5 (9x)(80x)
	d_R	76.90	2736.1	254.00	2913.2	663.20	3322.4	1357	4016 (6x)(5x)
	d_I				23686				
	d_E				211688				
Viscous Finger (5746)	d_S	45.10	13825	55.18	13835	65.45	13845	72.18	13845 (33x)(1000x)
	d_R	374	14154	1156	14936	2843	16624	5636	19416 (23x)(800x)
	d_I				455388*				
	d_E				15965520*				

