



HAL
open science

A New Similarity Measure for Temporal Semantic Sequences

Hiba Merakchi, Thomas Devogele, Veronika Peralta, Cyril de Runz

► **To cite this version:**

Hiba Merakchi, Thomas Devogele, Veronika Peralta, Cyril de Runz. A New Similarity Measure for Temporal Semantic Sequences. Symposium MaDICS, May 2025, Toulouse, France. <hal-05349031>

HAL Id: hal-05349031

<https://hal.science/hal-05349031v1>

Submitted on 5 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



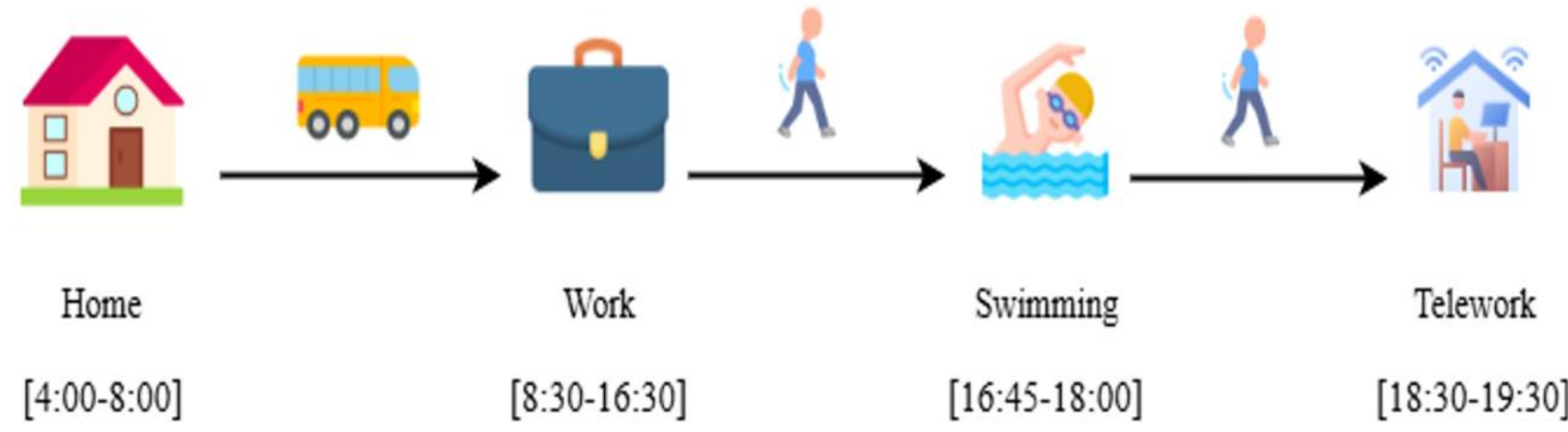
HAL Authorization

A New Similarity Measure for Temporal Semantic Sequences

Context

Semantic Sequences

- Chronologically ordered elements.
- Semantic meaning.
- Potentially a duration.



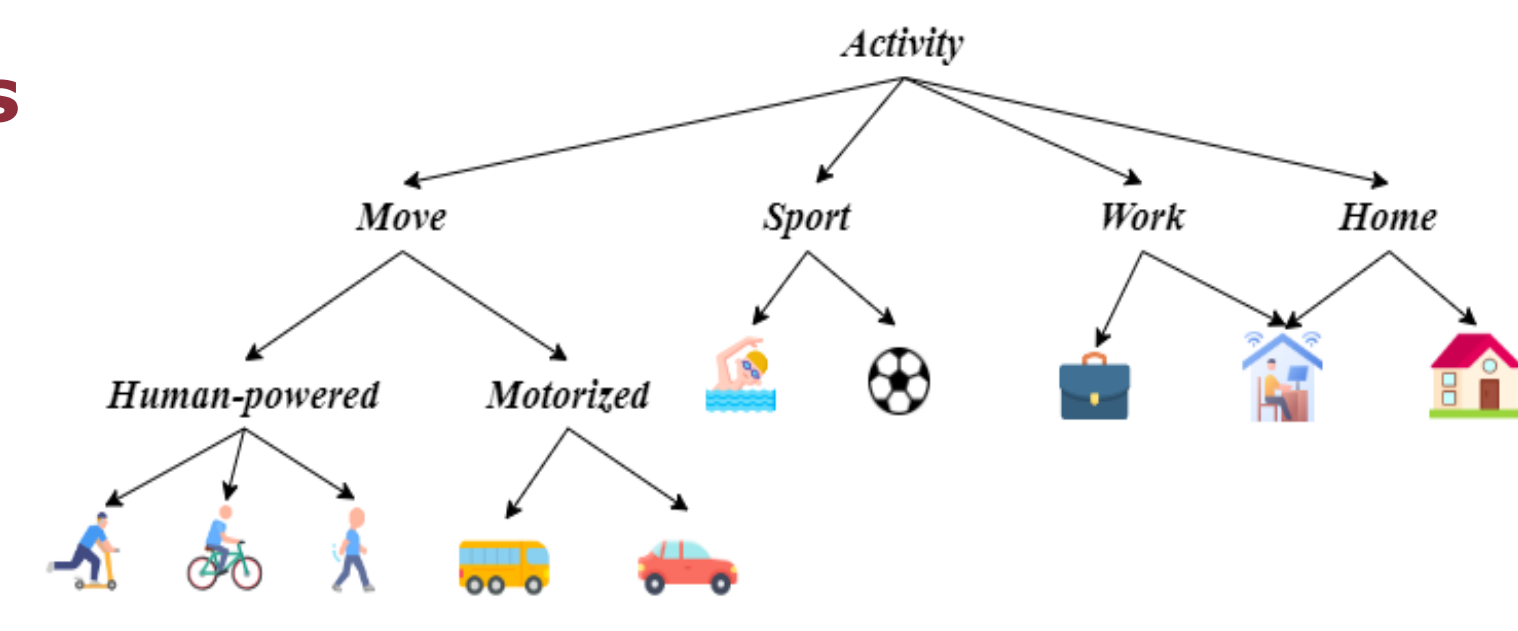
Variety of Semantic Sequences

- Mobility trajectories.
- Life paths.
- Patient records.
- Various activity flows :
 - Production chain steps.
 - E-learning exercises.
 - Database queries.
 - Songs in a playlist.

Comparing Semantic Sequences

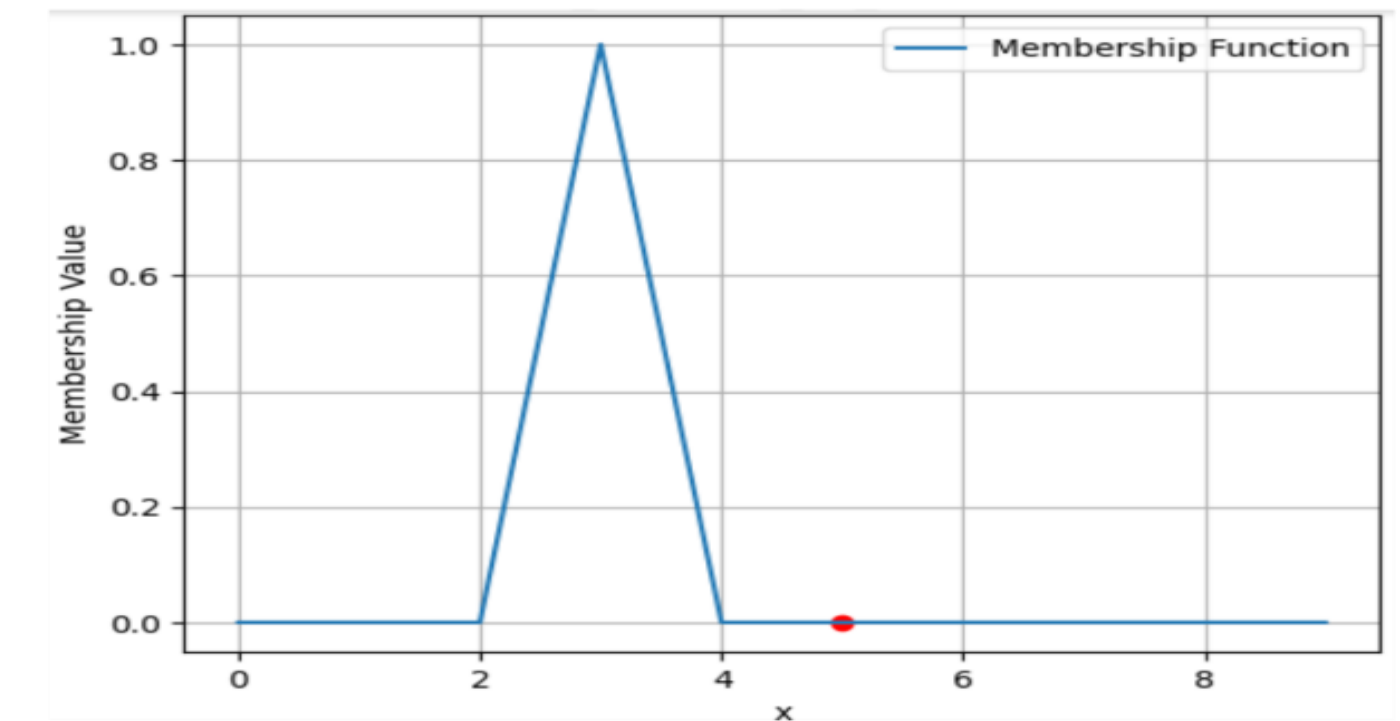
Semantic Similarity

- Ontologies.
- Sequences with close elements are semantically closer.



Temporal Similarity

- Fuzzy Logic.
- Sequences with smaller time shift are temporally closer.



Relative Fuzzy Temporal Hamming

1. Time-Length Comparability

- Fuzzy index that reflects the comparability of two sequences in terms of their time length.
- Meaningful and consistent comparison.
- Similarity value based on the difference in their time lengths.

$$I_c(S_a, S_b) = \begin{cases} 0, & \text{if } |l_t(S_a) - l_t(S_b)| > l, \\ 1 - \frac{|l_t(S_a) - l_t(S_b)|}{l}, & \text{otherwise.} \end{cases}$$

2. Duration Normalization

- Adjusting the element durations of each sequence to relative values.
- Sum of normalized durations equals 1.

$$\hat{d}_i = \frac{d_i}{\sum_{j=1}^n d_j}$$

3. Application of FTH

- Time-weighted cost function that ensures that longer elements incur a higher editing cost.

$$FTH_{\Delta}(\hat{S}_a, \hat{S}_b) = \max\{FTH_{\Delta}(\hat{S}_a \rightarrow \hat{S}_b), FTH_{\Delta}(\hat{S}_b \rightarrow \hat{S}_a)\}$$

4. From FTH Δ to Sim Δ

- We obtain a distance measure that we transform into a similarity.
- FTH Δ is proven to be bounded by the sequence time length (in this case 1).

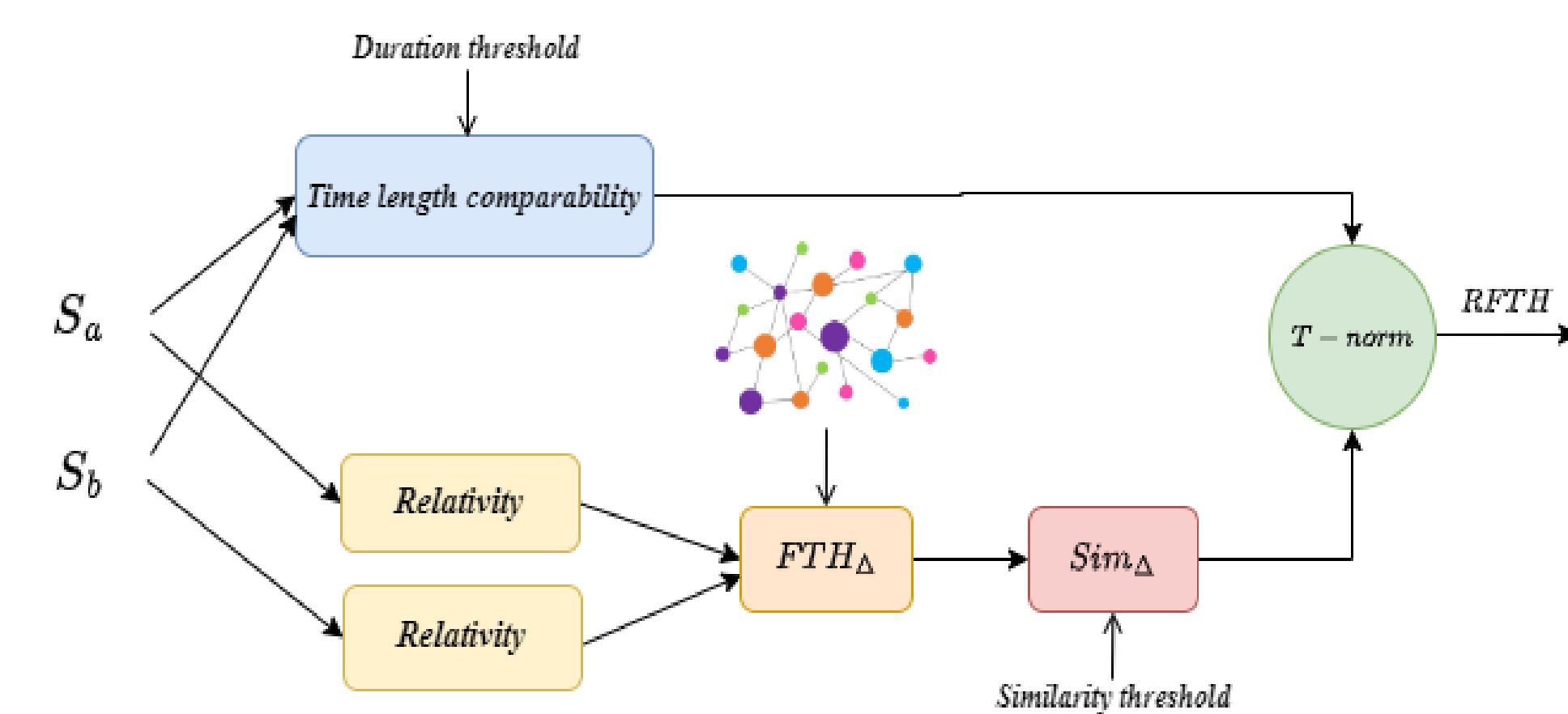
$$sim_{\Delta}(\hat{S}_a, \hat{S}_b) = 1 - FTH_{\Delta}(\hat{S}_a, \hat{S}_b)$$

5. Similarity Agregation

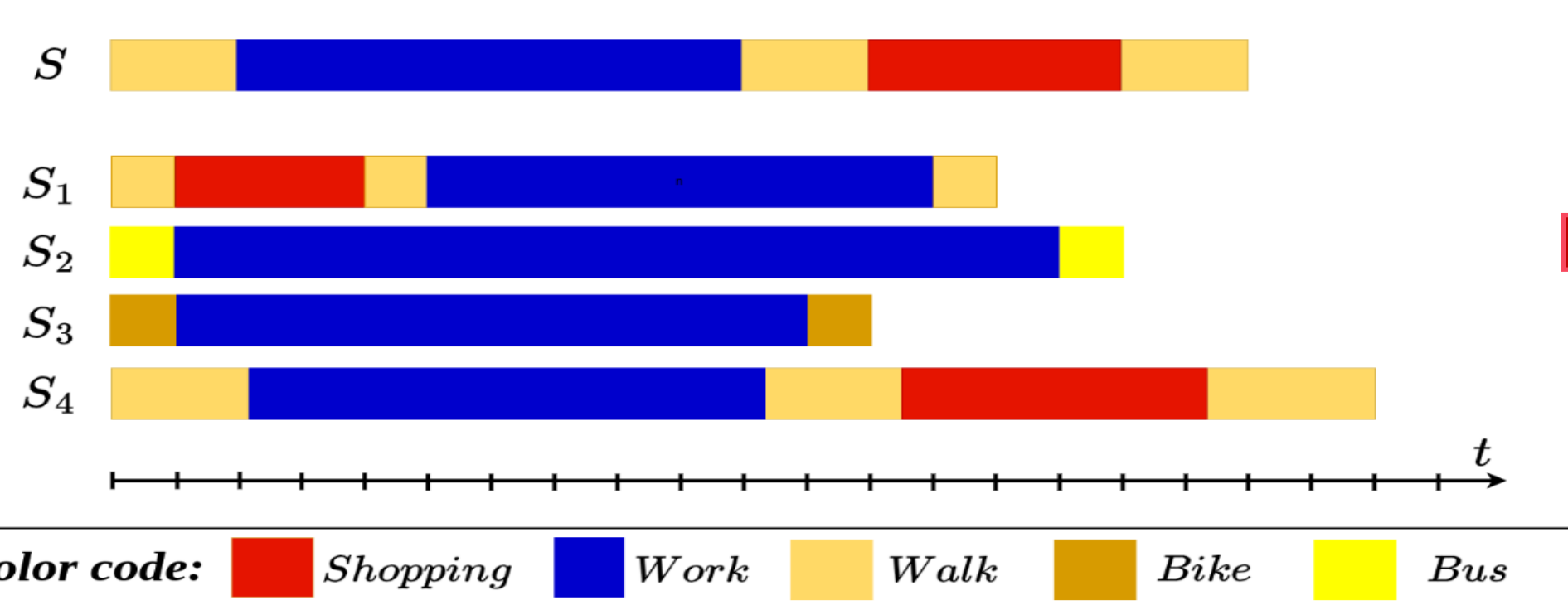
- Unified Metric using a **T-norm**.
- Mathematically sound and proportional fusion of both measures.

$$RFTH(S_a, S_b) = T(I_c(S_a, S_b), sim_{\Delta}(\hat{S}_a, \hat{S}_b))$$

Pipeline for Relative FTH



Example

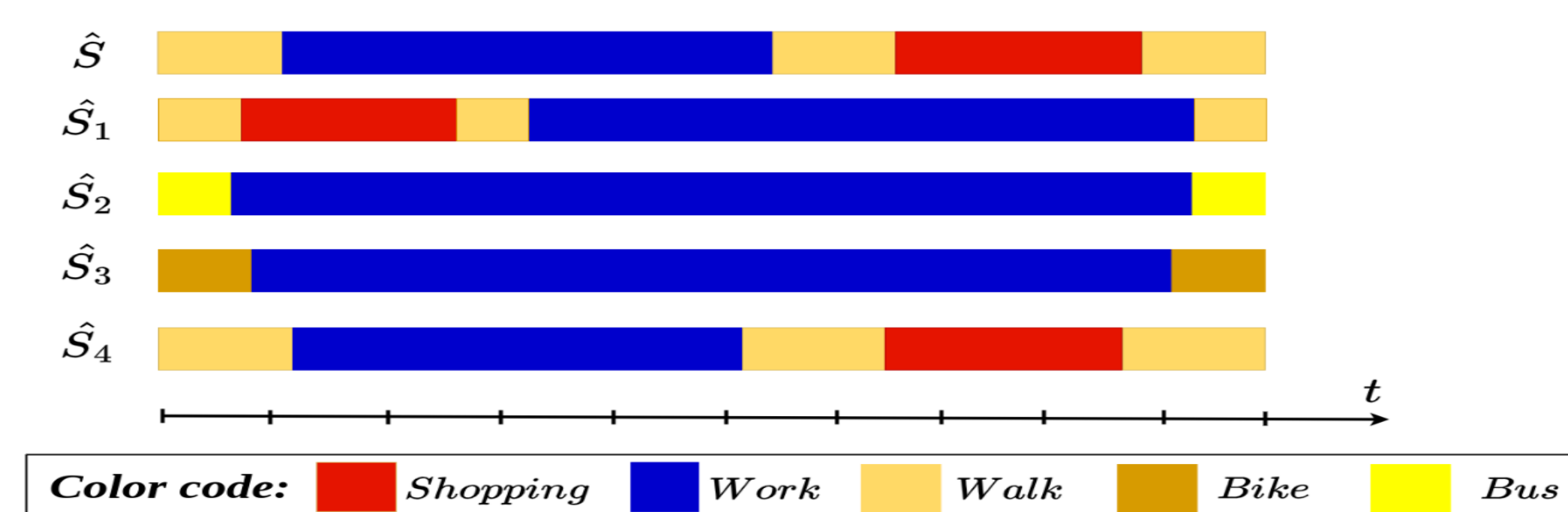


1. Time-Length Comparability

Sequence	S1	S2	S3	S4
Ic	0.5	0.75	0	0.75

Time-Length Comparability Index between sequence S and each of the sequences S1, S2, S3, and S4.

2. Duration Normalization



Representation of the normalized versions of sequences S, S1, S2, S3, and S4. Each time slot represents 0.1

5. Similarity Agregation

4. From FTH Δ to Sim Δ

3. Application of FTH

Sequence	S1	S2	S3	S4
Ic	0.5	0.75	0	0.75
FTH Δ	0.42	0.51	0.44	0.04
Sim Δ	0.58	0.49	0.56	0.96
Zadeh: Min(Ic, Sim Δ)	0.5	0.49	0	0.75
Probabilistic T-norm: Ic x Sim Δ	0.29	0.36	0	0.72
Łukasiewicz: max(0, Ic + Sim Δ - 1)	0.08	0.24	0	0.71

Sequence	S1	S2	S3	S4
Sim Δ	0.58	0.49	0.56	0.96

SIM Δ between sequences S1, S2, S3, and S4.

Sequence	S1	S2	S3	S4
FTH Δ	0.42	0.51	0.44	0.04

FTH Δ distance between sequences S1, S2, S3, and S4.

Future Work

- Comparison of sequences with missing data or irregular event patterns.
- Query Language capable of formulating complex queries on temporal semantic sequences.