

Data Analytics and Big Spatio-Temporal Data

PhD Candidate:

Luca Colomba

1. Introduction

The high diffusion of IoT and connected devices enables the acquisition of great amount of data, often characterized with spatial and temporal information, e.g., bike and car sharing systems. Consequently, the adoption and development of data mining and knowledge extraction algorithms can provide useful insights to users and characterize spatio-temporal events.

2. Goals

The goal of this thesis is to develop distributed and scalable data mining algorithms to analyse spatio-temporal data. More specifically, given a database of spatio-temporal events, the goal is to mine frequent patterns of spatio-temporal events and characterize them, without any absolute reference in space and time (**spatio-temporal invariance property**). An example is shown in Fig. 1.

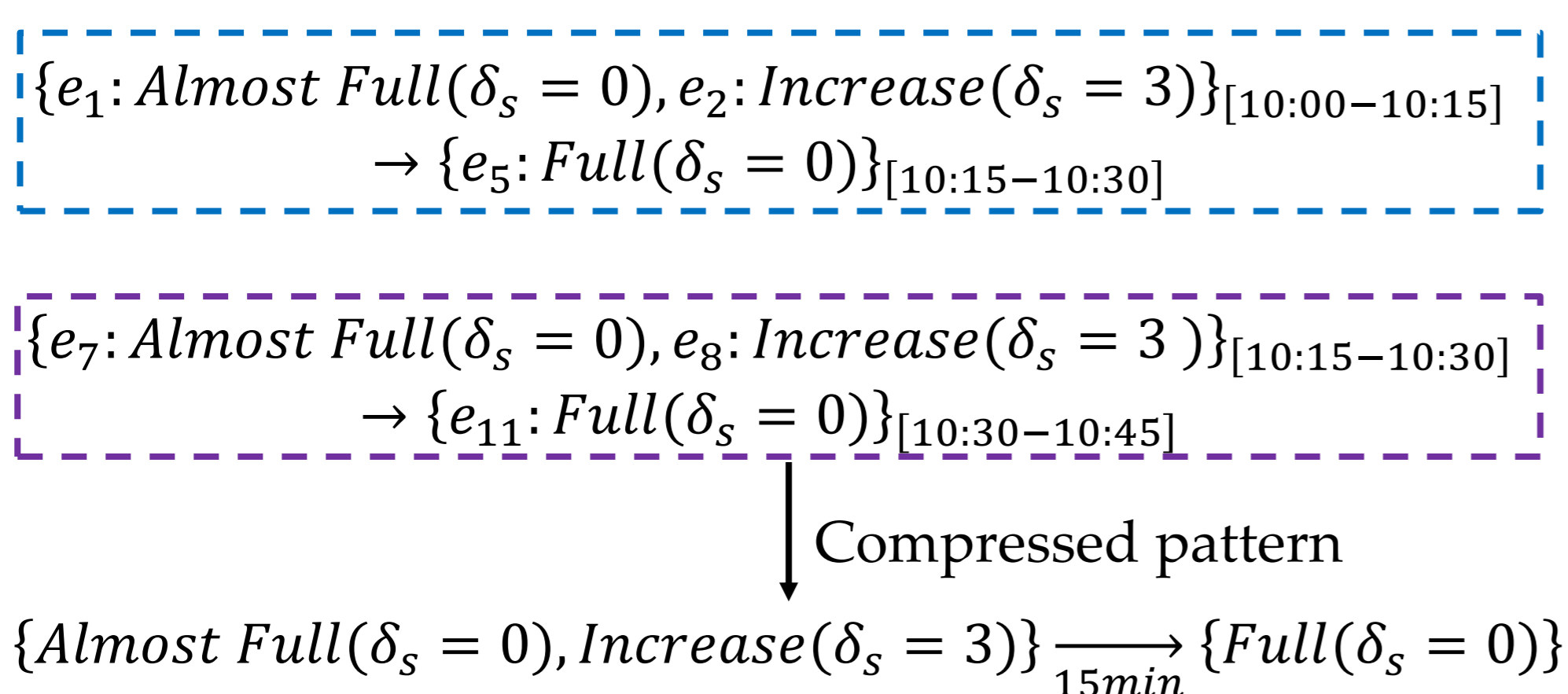
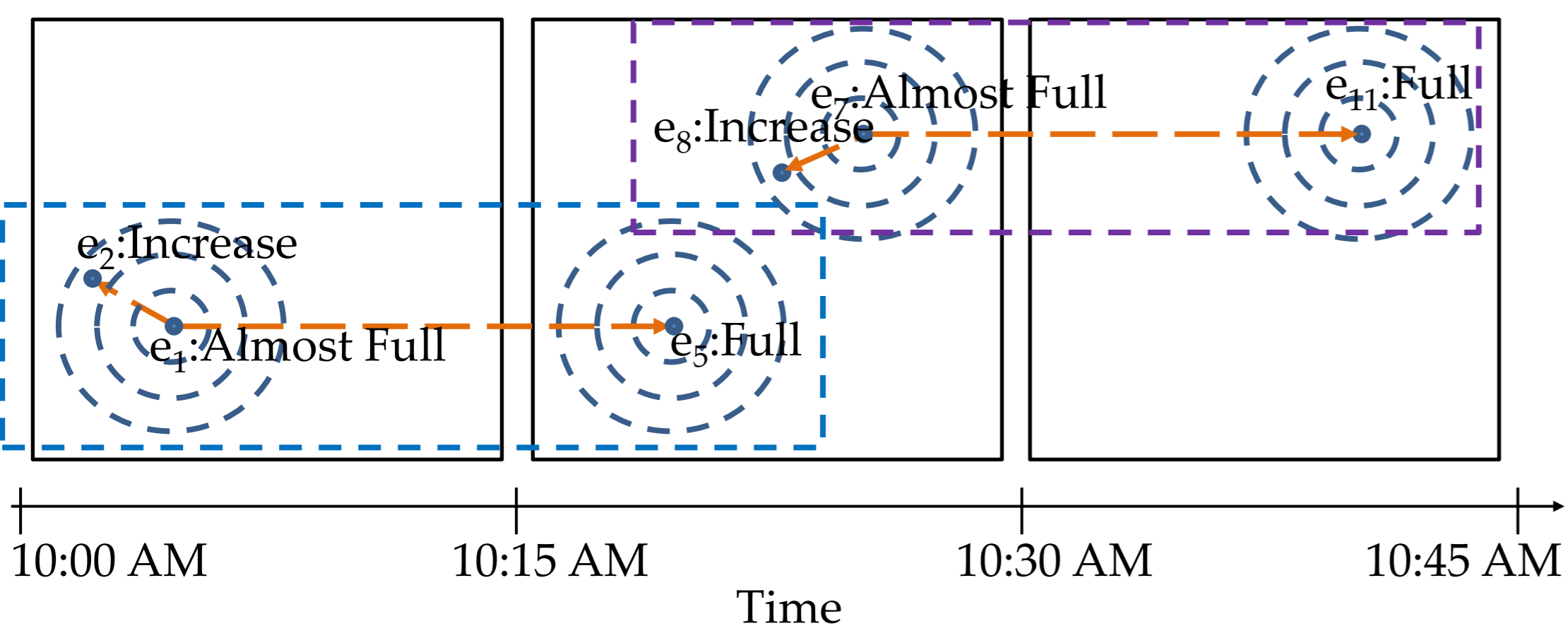
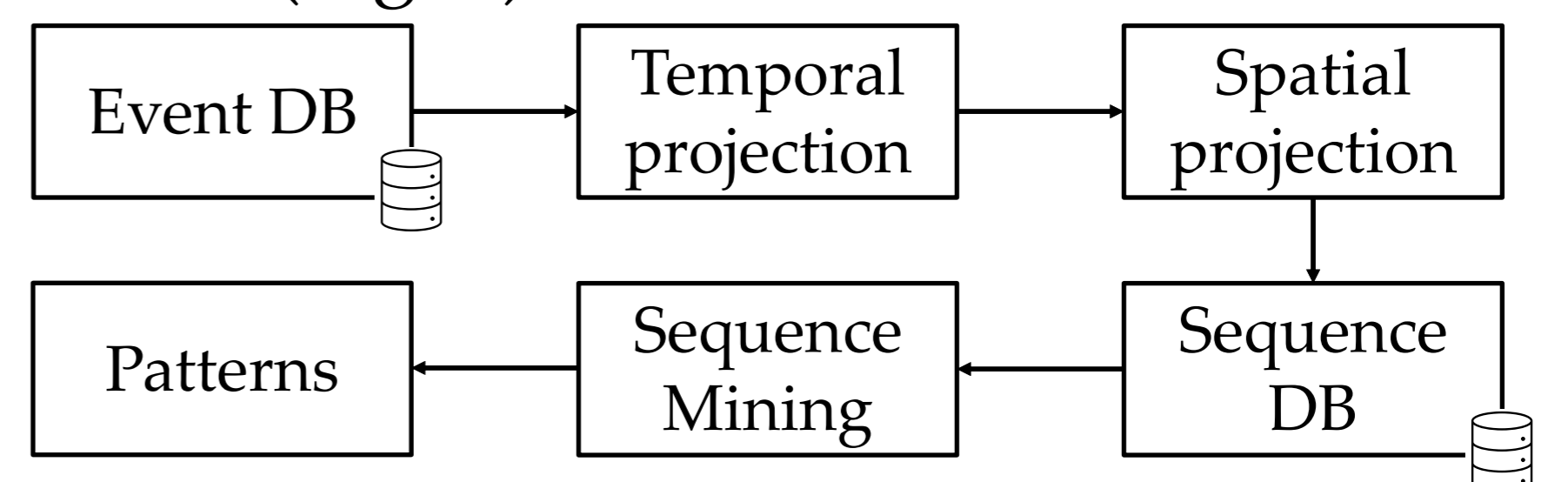


Fig. 1

4. Method

In this context, we propose Spatio-Temporally Invariant Pattern Miner (STInv Miner). The algorithm analyses a database of events, reprojecting them on temporal and spatial dimensions to satisfy the spatio-temporal invariance property. Consequently, a sequence database is generated, on which a distributed sequence mining algorithm is executed (Fig. 2).



5. Experimental results

The extracted patterns (Table 1) show how STInv miner is able to mine non-static patterns. Moreover, the proposed patterns demonstrated effective compression ratio (CR) compared to the patterns mined when removing the spatial invariance property.

| | Conf. 1 | Conf. 2 |
|--|---------|---------|
| Total # of STInvs | 45.3M | 11.0M |
| Mean # triplets per sequence | 5.90 | 5.77 |
| Mean discrete spatial distance | 2.40 | 8.31 |
| Mean discrete temporal distance | 2.13 | 2.17 |
| #sequences with at least one $\delta_s > 0$ (%) | 99.95 | 99.89 |
| #sequences with at least one $\delta_t > 0$ (%) | 99.99 | 99.99 |
| #sequences with at least one $\delta_s > 0$ and $\delta_t > 0$ (%) | 99.29 | 98.36 |

Table 1

| | Config. 1 | | | Config. 2 | | |
|---------------|-----------|-------|-------|-----------|-------|------|
| | #STInvs | #Abs | CR | #STInvs | #Abs | CR |
| San Francisco | 45.3M | 2350M | 51.90 | 6.24M | 28.2M | 4.53 |
| Palo Alto | 111K | 143K | 1.29 | 344K | 592K | 1.72 |

Table 2

6. References

[1] Colomba, L., Cagliero, L., Garza, P. (2022). Mining SpatioTemporally Invariant Patterns, SIGSPATIAL2022 Conference