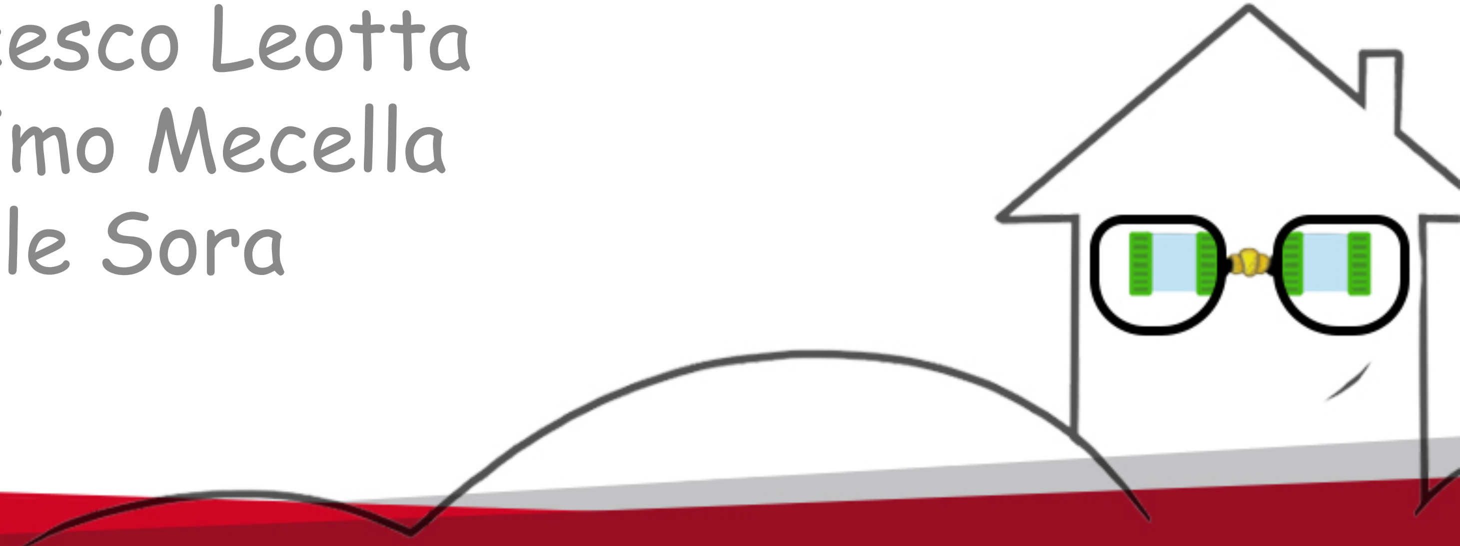
SAPIENZA
UNIVERSITÀ DI ROMA

An habit is a process: a BPM-based approach for smart spaces

Francesco Leotta
Massimo Mecella
Daniele Sora

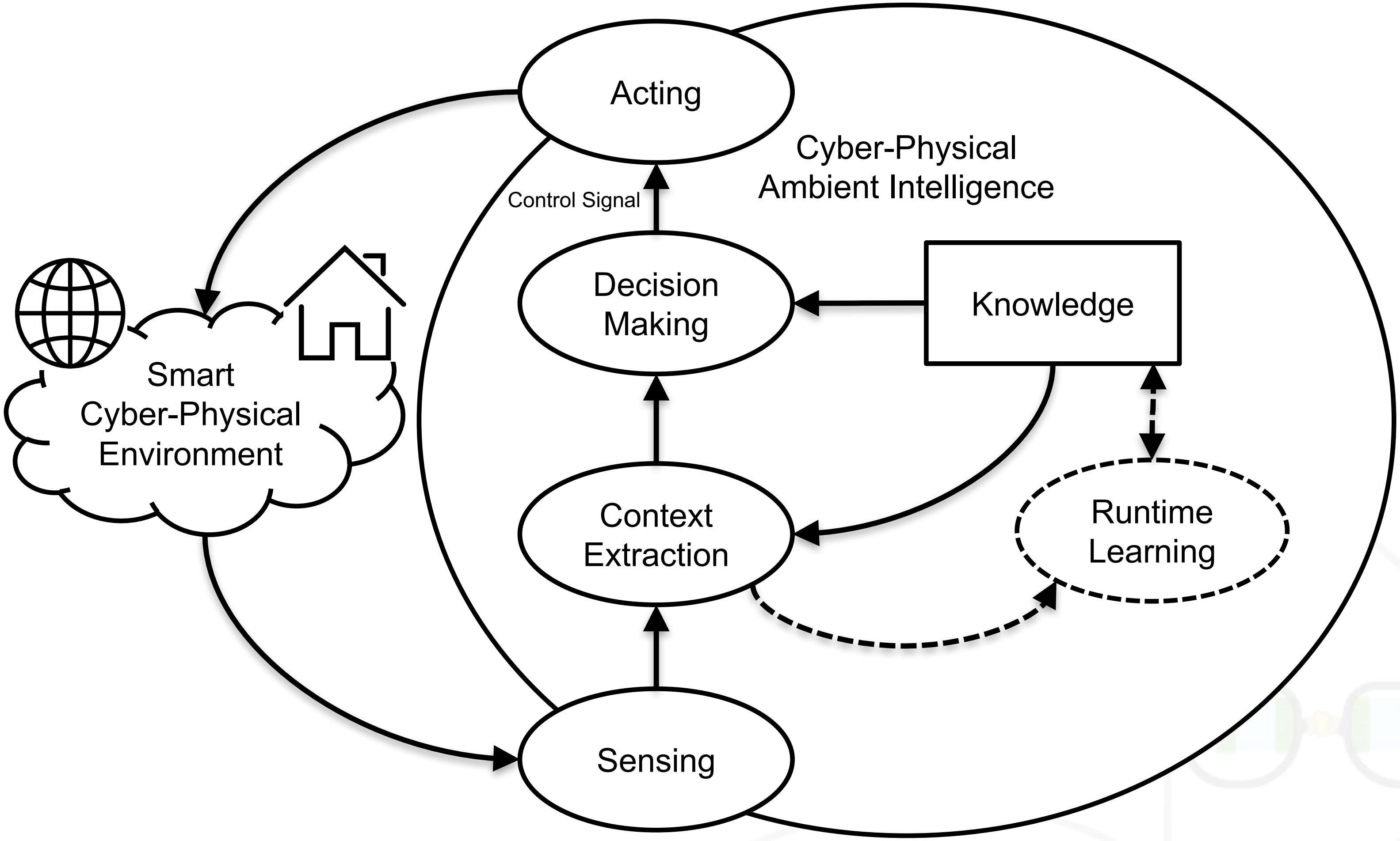


Smart Spaces and AmI

"A Smart Space is an environment centered on its human users in which a set of embedded networked artefacts, both hardware and software, collectively realize the paradigm of ambient intelligence (AmI)"

- Strongly related to **internet of things (IoT)**
  - Large availability of small and powerful embedded devices
  - Interaction through both physical actions and digital identities
- E.g., smart houses and offices
- Realize the paradigm of **Ambient Intelligence (AmI)**
  - different research areas (e.g., AI, HCI, social sciences)
- Immediate impacts on society
  - Daily life and work experience
  - Energy saving
  - Safety and security

The AmI Loop


Acting
Cyber-Physical
Ambient Intelligence
Control Signal
Decision
Making
Knowledge
Smart
Cyber-Physical
Environment
Context
Extraction
Runtime
Learning
Sensing

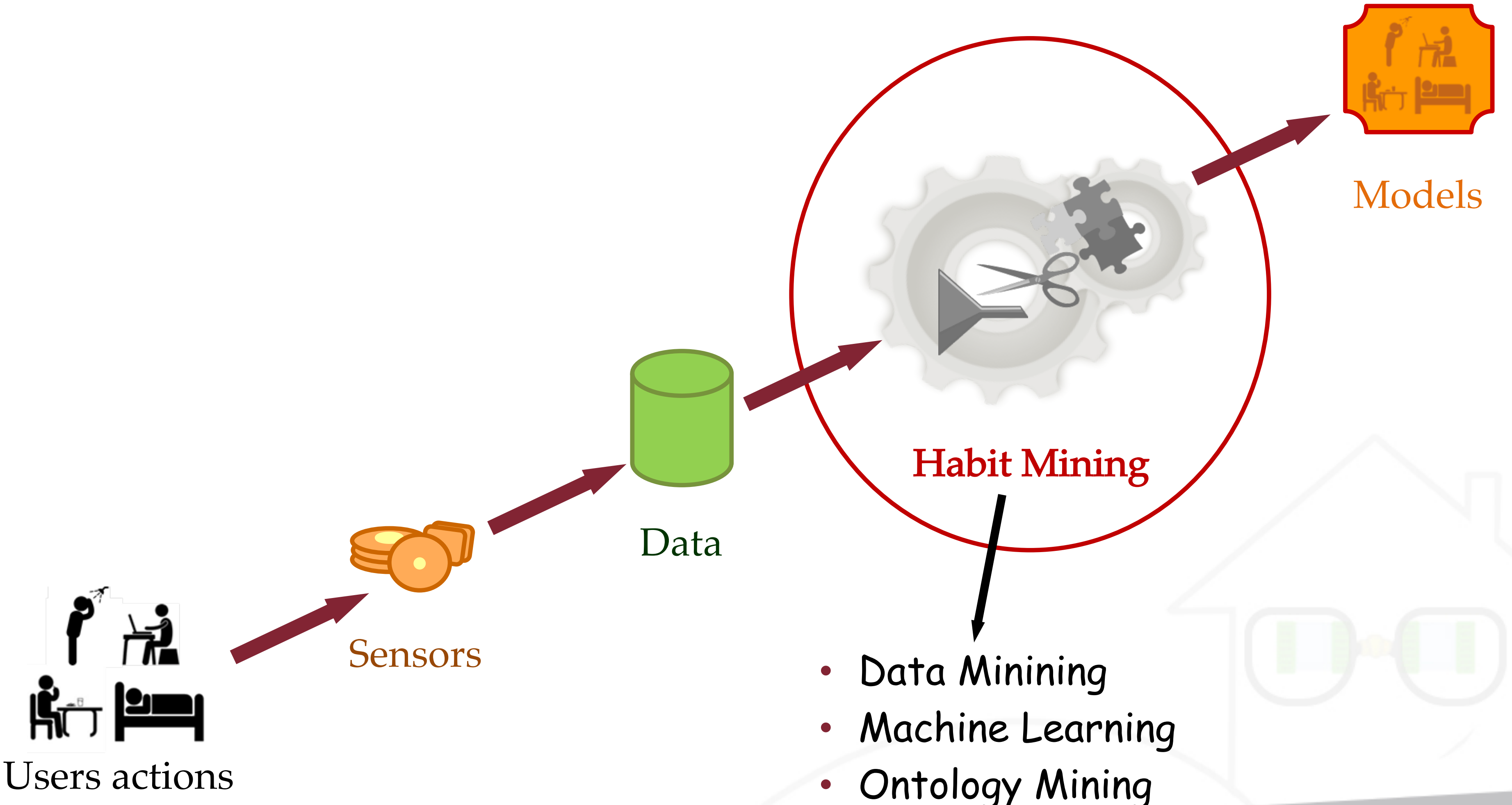
The Role of Knowledge

- Knowledge plays a central role in AmI systems
- It takes the form of a set of models describing:
  - **Human routines (habits)**
  - Environment/device dynamics
  - User preferences
- Specification-Based
  - Represent **hand-made** expert knowledge using logic formalisms
  - Reasoning engines to infer conclusions and to make decisions
  - **Impractical ☹ but human-readable ☺**
- Learning-Based
  - Represented by using mathematical and statistical formalisms (e.g., HMM)
  - **Automatically extracted ☺ but difficult to be revised ☹**

Mining-based Approaches and Process Mining

- Apply unsupervised or semi-supervised learning approaches to obtain human readable models
  - The best of two worlds!!!
  - E.g., Pattern-mining approaches (e.g., CASAS project [Cook2013])
- Approaches that model human habits as workflows have been proposed [Aztiria2010]
  - **Methods applied in controlled conditions!!!**
- **Is it possible to apply process mining to smart spaces?**
  - **Growing availability and maturity of process mining techniques ☺**
  - **Different challenges must be addressed in a real setting ☹**

The Context


Users actions
Sensors
Data
Habit Mining
Models
• Data Minining
• Machine Learning
• Ontology Mining

Current Approaches (1)

Knowledge Driven Approaches:

- ECA Rules:
  - Active Database
  - Event-Condition-Action (ECA) rules
    ("ON event IF condition THEN action")
- Ontologies:
  - Hand Made
  - Models describe relations between sensors

Current Approaches(2)

Data Driven approaches:

- Goals:
  - pattern mining
  - iterative compression of the sensor log
  - Enactment or recognition

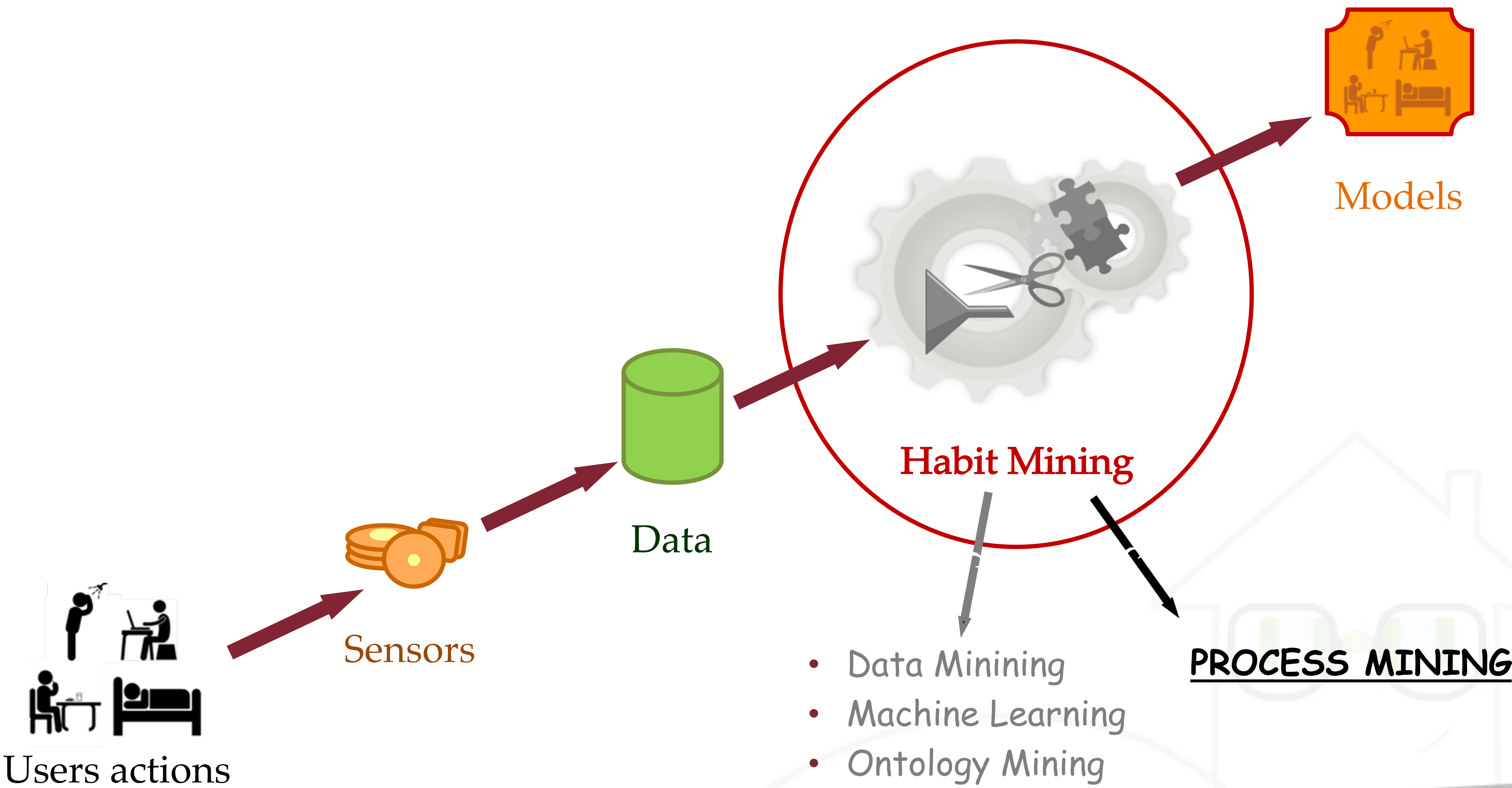
Often used machine learning algorithms (HMM, SVN), low readability, big amount of data

Current Approaches(3)

Hybrid Approaches:

- Knowledge-driven + data-driven approach
  - Clustering algorithms for activities
  - Learning algorithm for models extraction from actions clusters
  - Ontology «seed», enriched using learning

The Objective



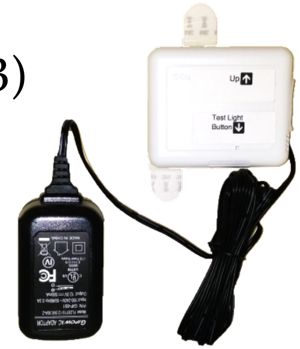
Habit = Human Process

- If we consider habit as human process, then Process Mining techniques can be applied.
- Models produced by Process Mining are in general designed for being analyzed by humans → more readable
- Which one? Fuzzy miner, designed for low structured processes (see better later!)

Smart Home in a Box

(A)	1	Server
(B)	3	Relays
(C)	24	Infrared motion sensors
(D)	2	Temperature sensors
(E)	1	Magnetic door sensor
(F)	40	Adhesive strips
(G)	34	Batteries (9V, AA)

(A)


(B)


(C)


(D)


(E)


(F)

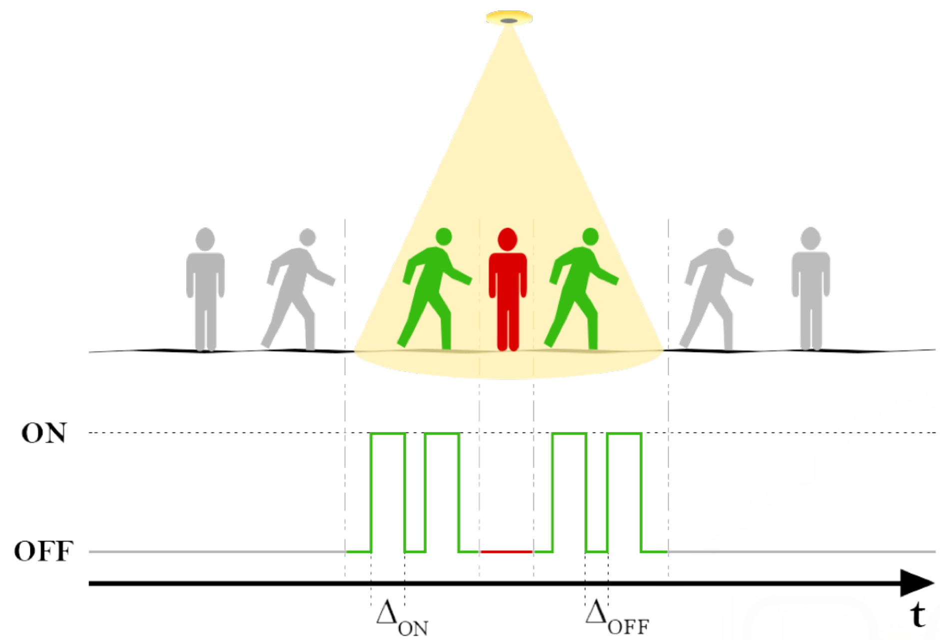

(G)


(Washington State University)

Casas Dataset

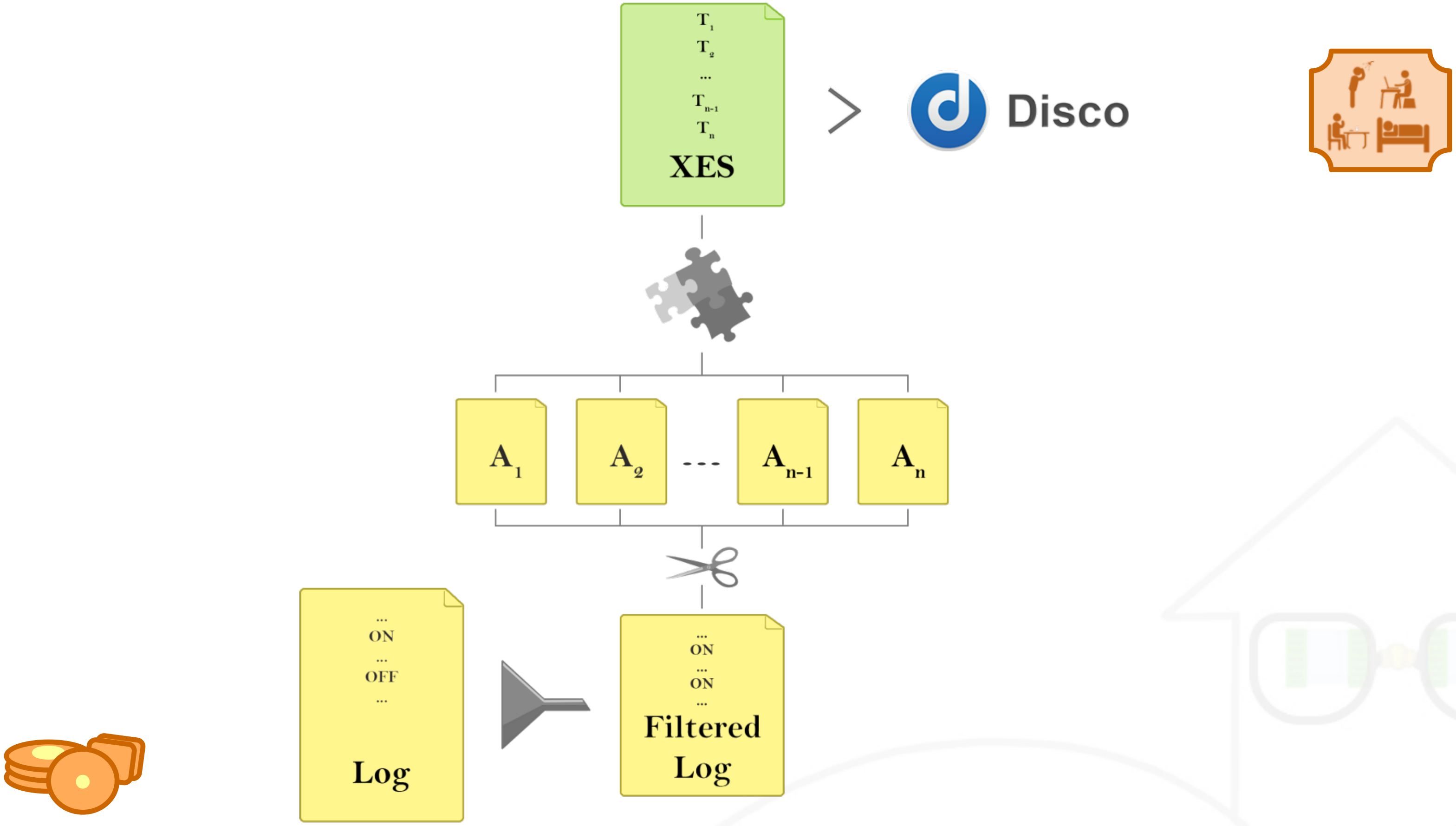
Infrared motion sensors

- Area
- Position

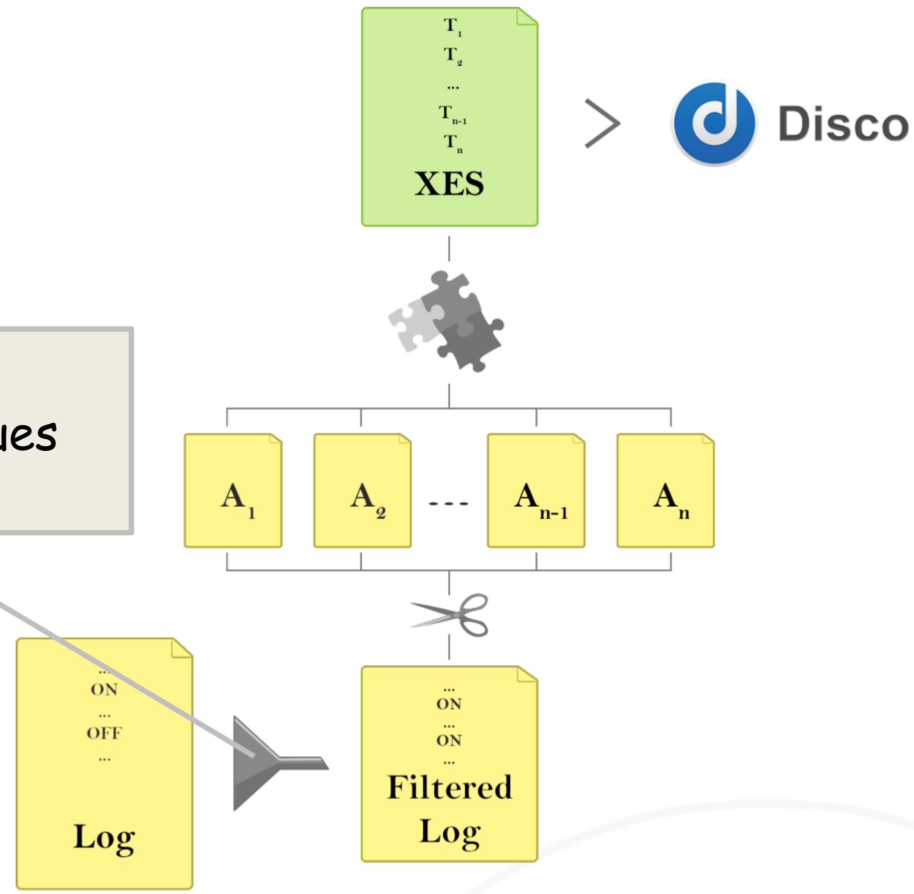


yyyy-MM-dd HH:mm:ss.SSSSSS SensorName VALUE

The Technique of UIC 2016



The Technique of UIC 2016

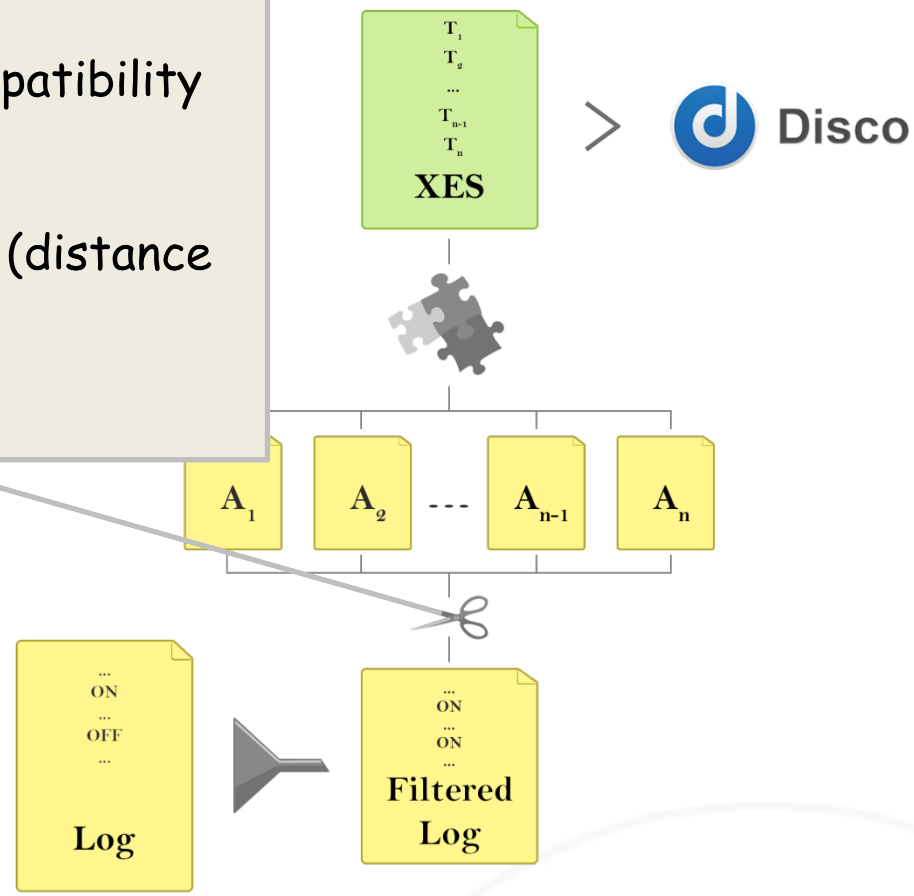


Only ON values

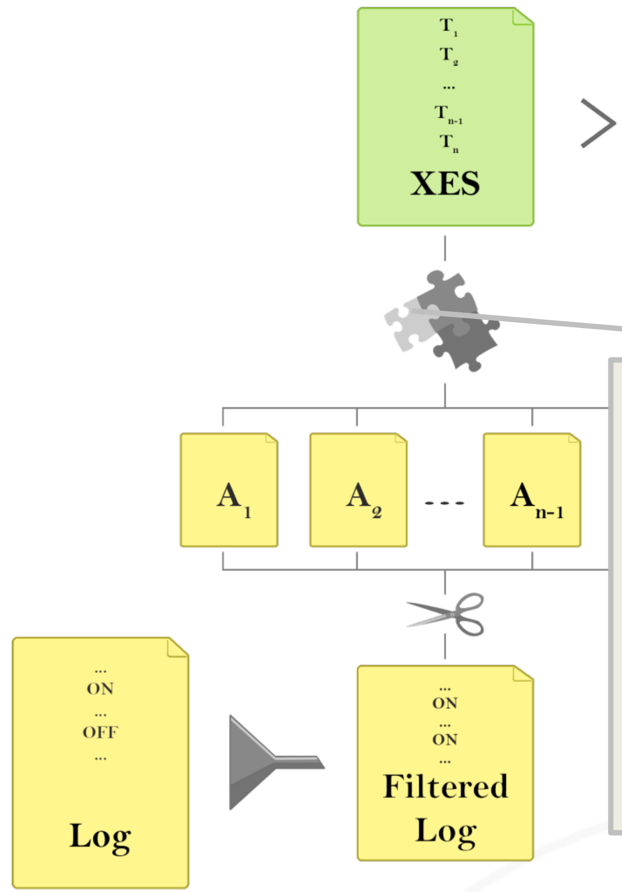
SAPIENZA UNIVERSITÀ DI ROMA

The Technique of UIC 2016

- Topological compatibility matrix
- Velocity vector (distance matrix)



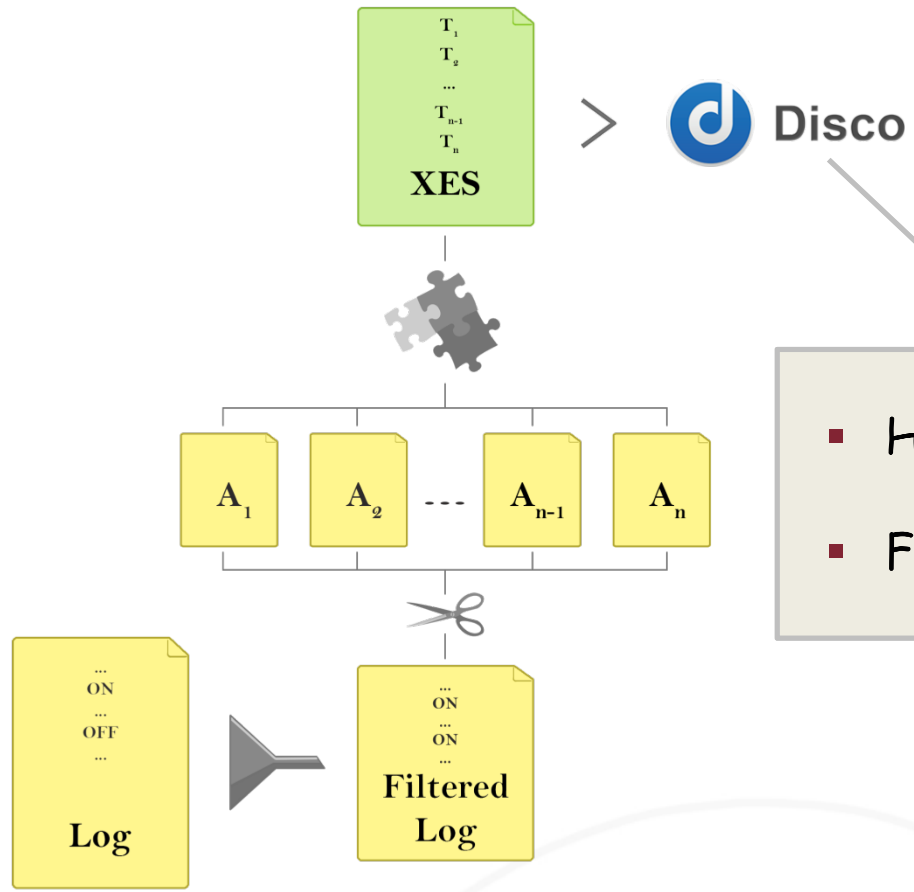
The Technique of UIC 2016



- traces of events

```
<string key="concept:name" value="name"/>
<date key="time:timestamp"
   value="2015-11-30T18:25:49.308+01:00"/>
<string key="Sensor" value="string"/>
<classifier name="Activity" keys="Sensor"/>
```

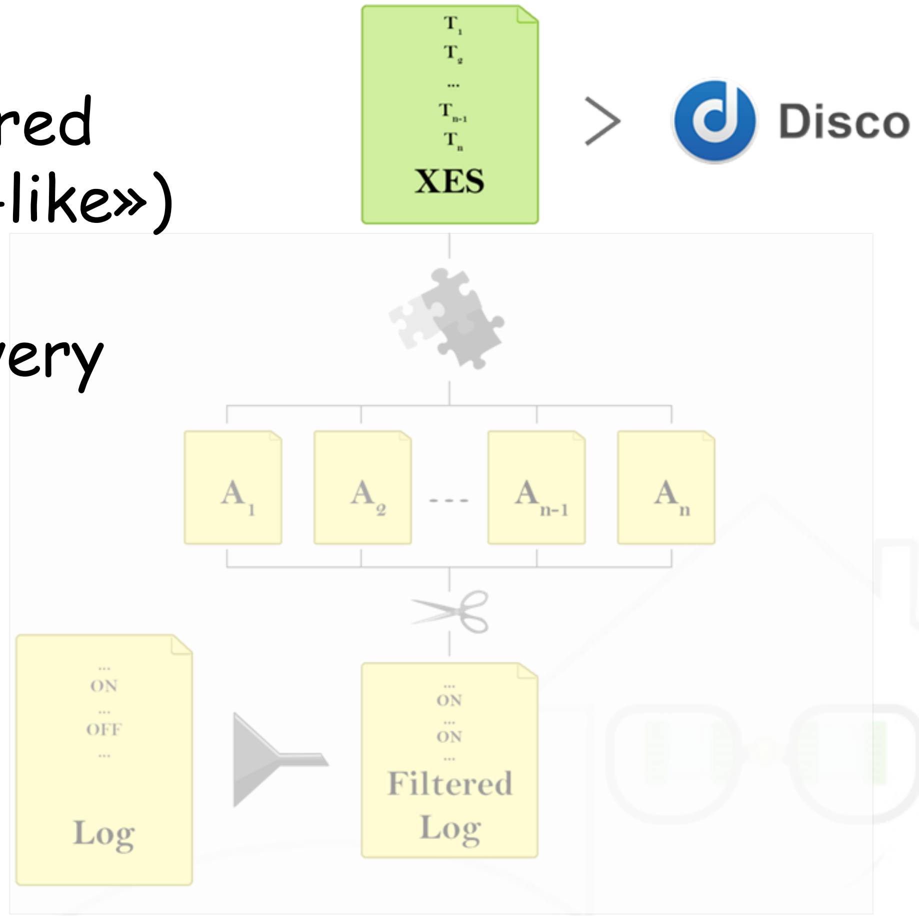
The Technique of UIC 2016



- Habit mining
- Fuzzy miner

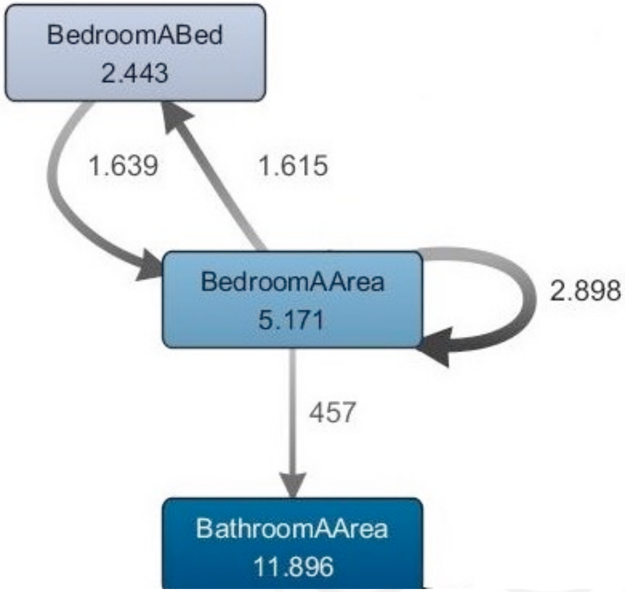
Fuzzy Mining (1)

- Well suited for unstructured process (alias «Spaghetti-like»)
- Automated process discovery
- «Road Map of the processes»: importance of connections graphically underlined

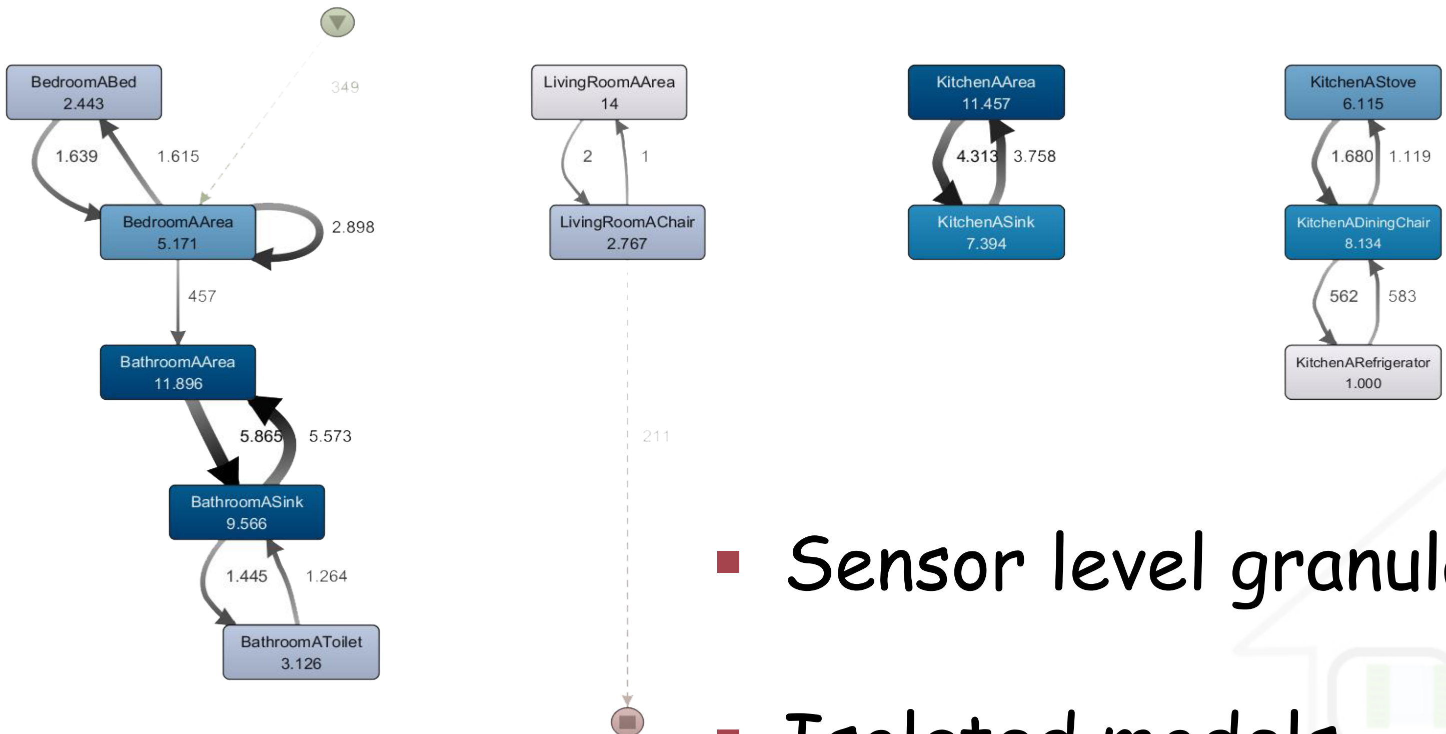


Fuzzy Mining (2)

- Fuzzy Model is composed by:
  - Nodes representing activities (sensors)
  - Edges representing connections
  - Metrics for dynamically filtering edges/nodes (significance/correlation)
  - Importance of each element is graphically showed



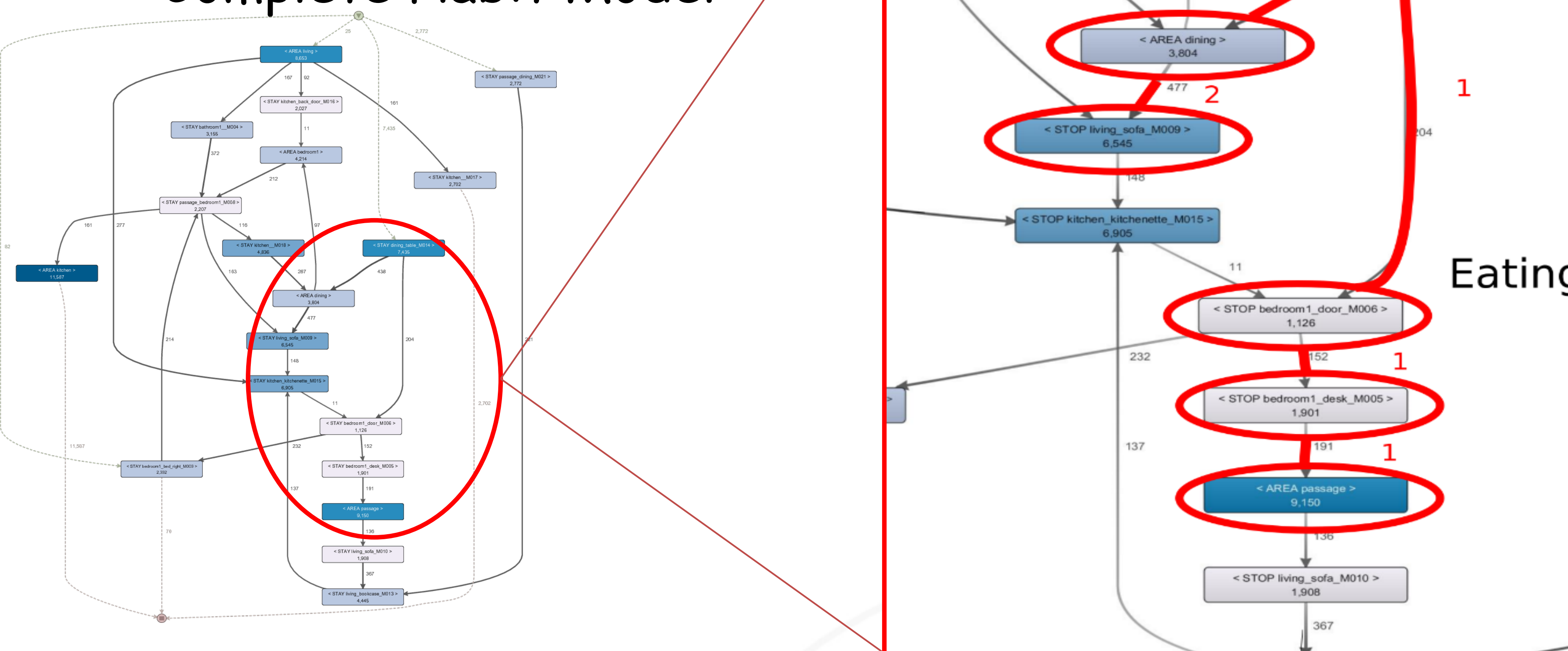
UIC 2016



- Sensor level granularity
- Isolated models

From UIC 2016 to UIC 2017

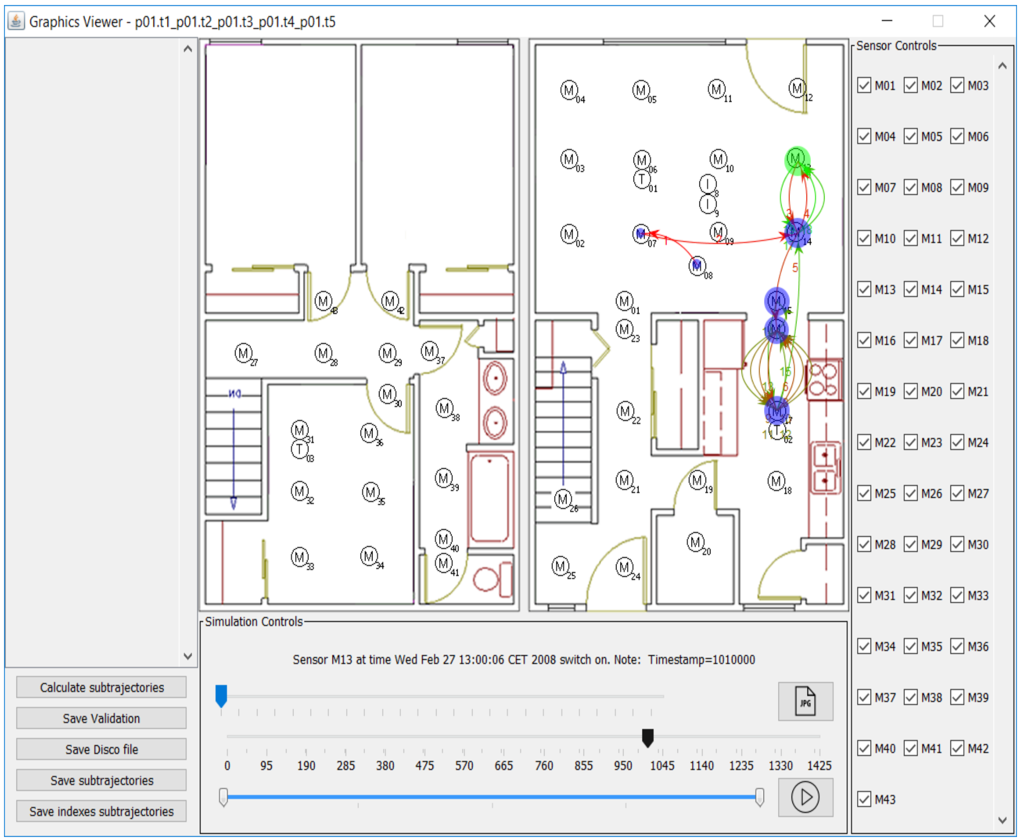
- Trajectory analysis
- Higher Granularity
- Complete Habit Model



Definitions

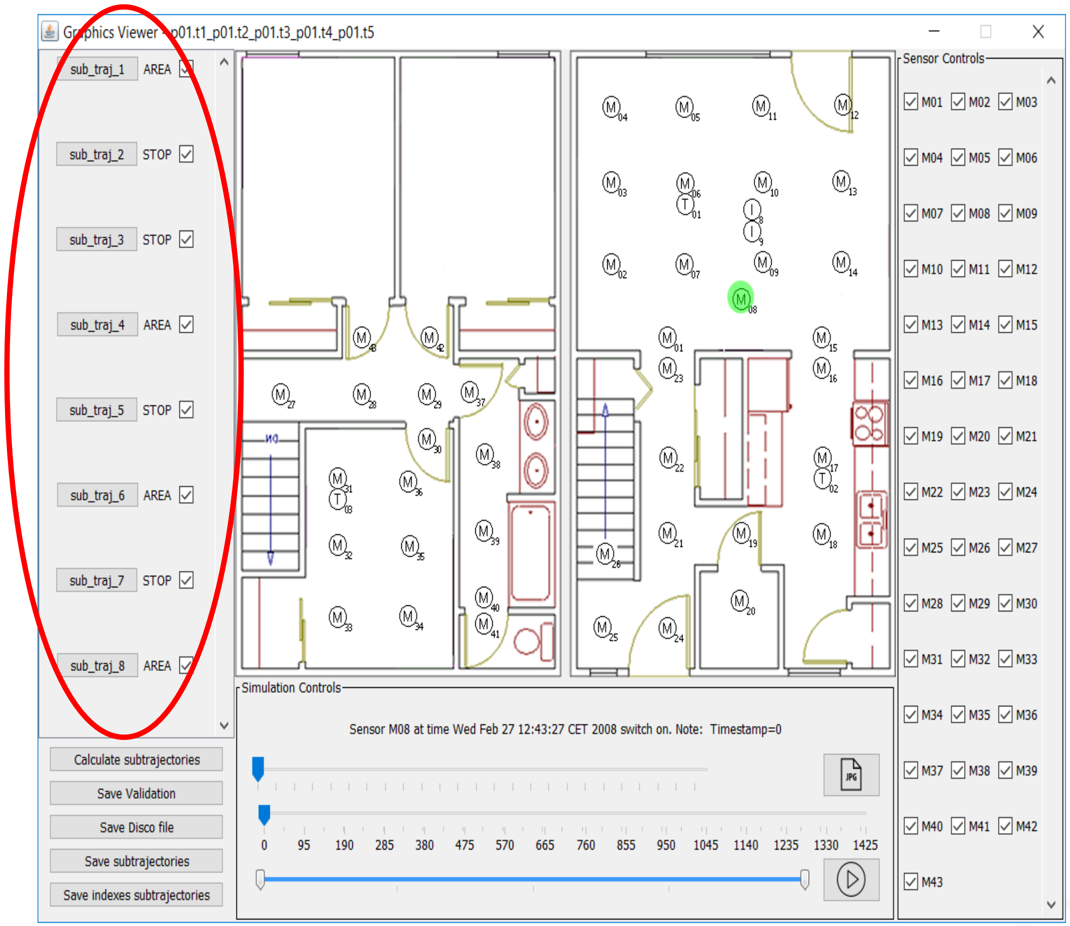
- Action: atomic interaction with the environment
- Activity: a sequence of actions (just one in extreme cases) or sensor measurements/events with a final goal. In some cases an action can be an activity itself
- Habit: a sequence or interleaving of activities that happen in specific contextual conditions

Trajectory Analysis Tool (1)



- Time Interval Selection
- Evolution Simulation
- Sensor Selection

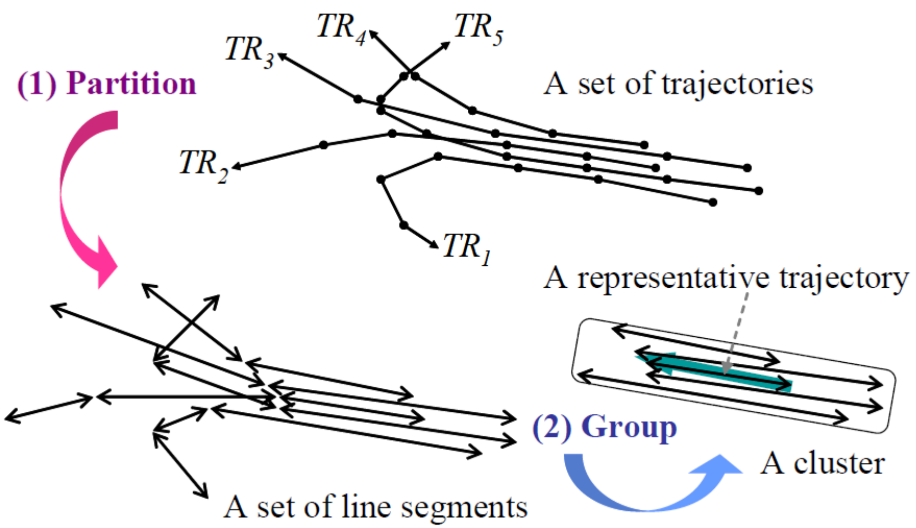
Trajectory Analysis Tool (2)



- Pattern Miner: extraction of pattern. They can be visualized on the map.
- Traclus Algorithm

TRACLUS algorithm*

- Trajectory clustering algorithm, devised for describing hurricanes' trajectories.
- Two phases:
  - Trajectory partitioning
  - Density-based line-segment clustering



*Lee, Han, Whang «Trajectory clustering: a partition-and-group framework» in Proceedings of 2007 ACM SIGMOD international conference on Management of data, 2007

Subtrajectories classification

Sub-trajectories → log segmentation in actions

Trajectory partitioning algorithm output is categorized in 3 classes:

STAY

AREA

MOVEMENT

Actions classification Indices

$\boldsymbol{I_m(\delta)}$ is the index related to a quick and heterogeneous movement

$$I_m(\delta) = \frac{number\ of\ distinct\ sensors}{total\ number\ of\ sensors}$$

$\boldsymbol{I_a(\delta)}$ is the index related to a movement in a given area

$$I_a(\delta) = Gini\ coefficient$$

$\boldsymbol{I_s(\delta)}$ is the index related to a static activities

$$I_s(\delta) = \frac{attivation\ time\ of\ the\ most\ triggered\ sensor}{total\ subtrajectory\ duration}$$

Actions classification

Classification Index:

$$I_{tot}(\delta) = w_m I_m(\delta) + w_a I_a(\delta) + w_s I_s(\delta)$$

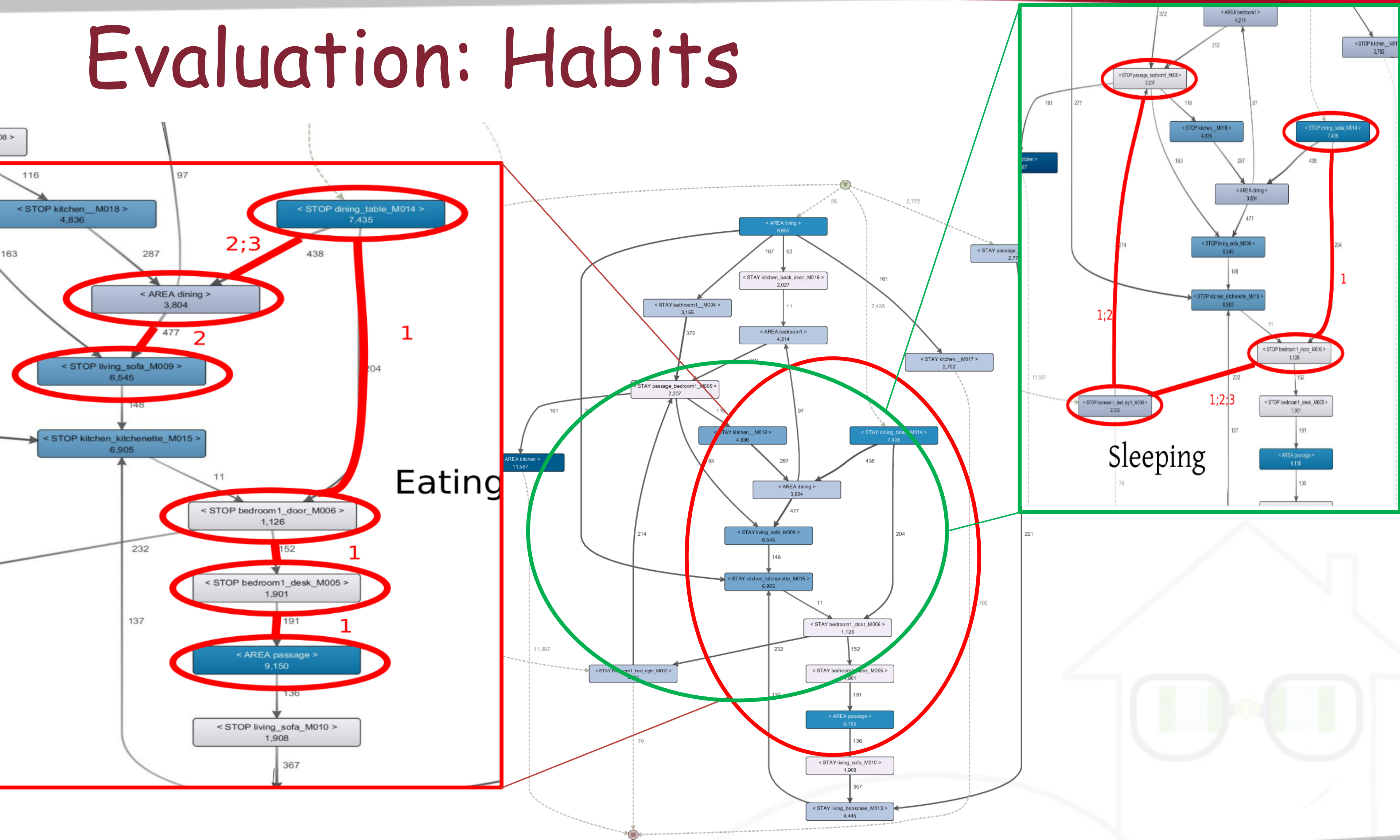
With:

$$w_m + w_a + w_s = 1$$

Subtrajectory classification:

$$f(\delta) = \begin{cases} STAY, & 0 \le I_{tot}(\delta) < T_a \\ AREA, & T_a \le I_{tot}(\delta) < T_m \\ MOVEMENT, & T_m \le I_{tot}(\delta) \le 1 \end{cases}$$

Evaluation: Habits



Evaluation: Similarity Metric

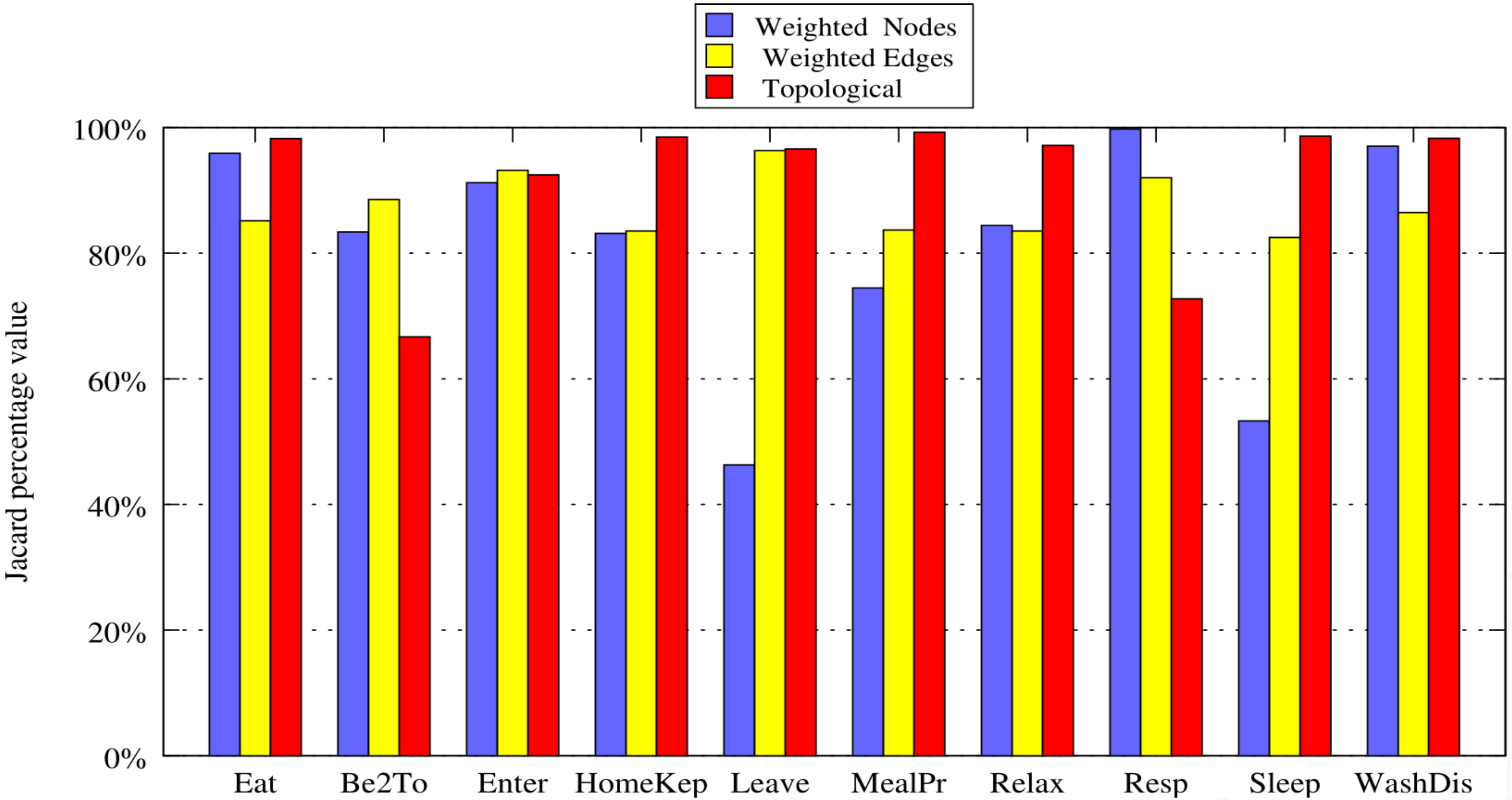
Weighted Jaccard similarity

$$J(G_1, G_2) = \frac{\sum_i \sum_j \min\left(G_1(n_i,n_j),G_2(n_i,n_j)\right)+\sum_i \min(G_1(n_i),G_2(n_i))}{\sum_i \sum_j \max\left(G_1(n_i,n_j),G_2(n_i,n_j)\right)+\sum_i \max(G_1(n_i),G_2(n_i))}$$

with

- G_k a graph
- $G_k(n_i, n_j)$ = weight arc from n_i to n_j
- $G_k(n_i)$ = weight node n_i

Evaluation


Weighted Nodes
Weighted Edges
Topological
Jacard percentage value
100%
80%
60%
40%
20%
0%
Eat
Be2To
Enter
HomeKep
Leave
MealPr
Relax
Resp
Sleep
WashDis

Conclusions…

- Process discovery technique to mine human behavior
- Elaborated unsupervised models
- High confidence of the models

…and future works

- User evaluation for readability
- Exploitation of models at runtime
- Anticipation of users decisions