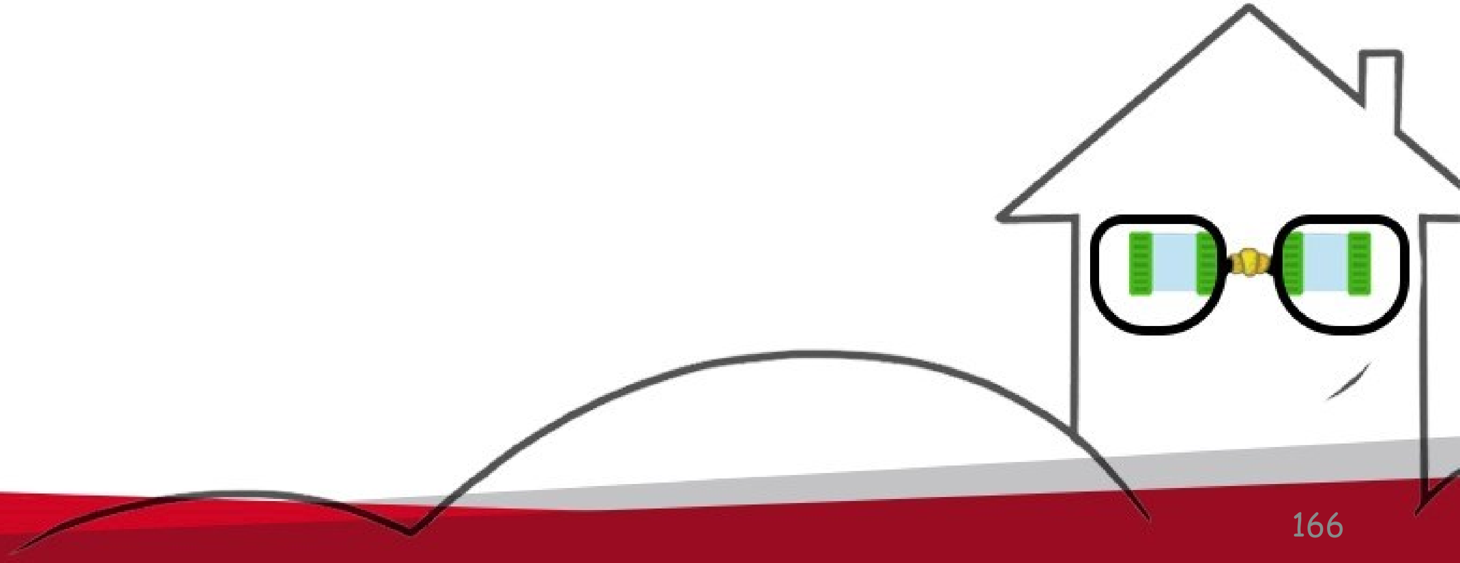
SAPIENZA
UNIVERSITÀ DI ROMA

Hands-On Visual Process Maps

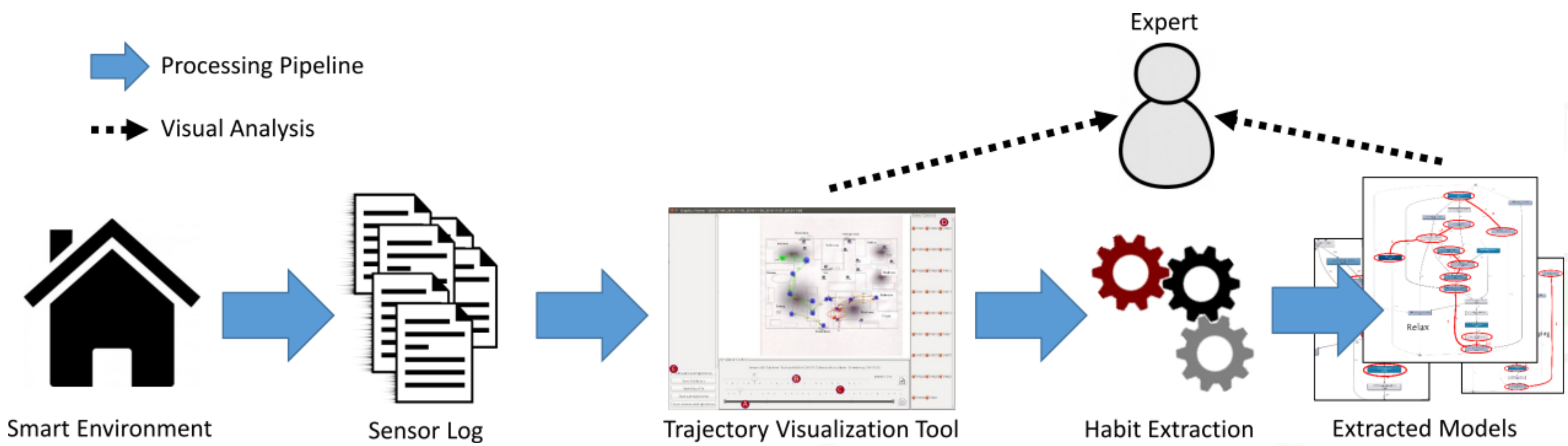


Visual Process Maps (1)

- Our proposed pipeline for
  - learning (cf. step 1 slide 19)
  - visual analysis (cf. step 5 slide 19) of sensor logs
- Based on process mining techniques
- Approach developed in:
  - Leotta F. Mecella M., Sora D., Spinelli G. "Pipelining user trajectory analysis and visual process maps for habit mining." 2017 IEEE Ubiquitous Intelligence & Computing, 2017.
  - Leotta F., Mecella M., Sora D. "Visual analysis of sensor logs in smart spaces: Activities vs. situations." 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2018.
  - Leotta F., Mecella M., Sora D. "Visual Process Maps: A Visualization Tool for Discovering Habits in Smart Homes." Journal of Ambient Intelligence and Humanized Computing, 2019.

Visual Process Maps (2)

- The pipeline consists of several steps
- Human expert is involved in visually analyzing the log and the extracted models
- Pipeline is intended for analysis



Useful Material

- Access the USB pen drive distributed to attendees during the tutorial

OR

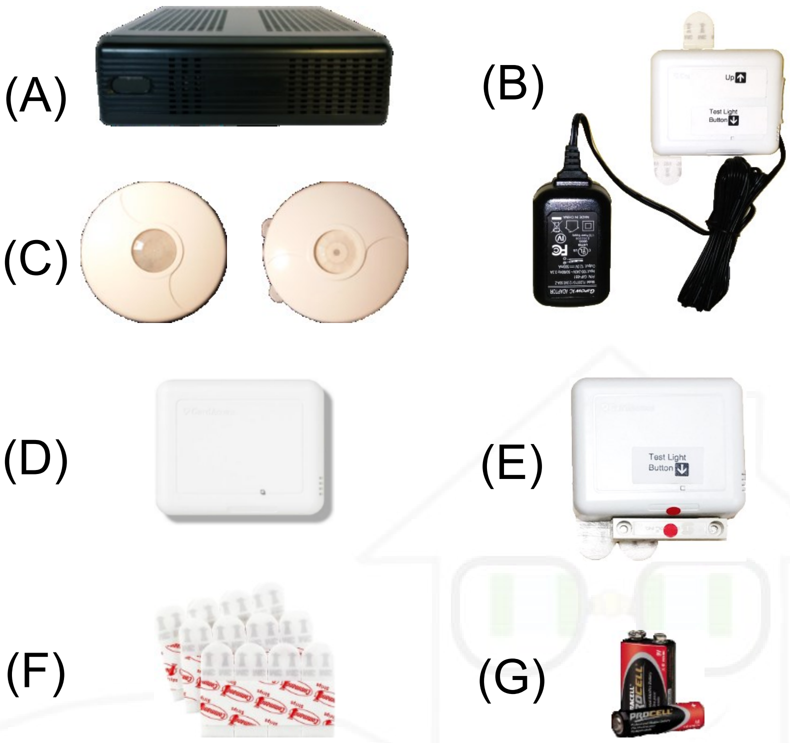
- Download the tool and data at
  - https://www.dropbox.com/s/17k5igmbzsnx651/VPMv1.0.zip?dl=0

Software/Hardware Requirements

- Any operating system
- Java Runtime Environment 8 installed
- Python 3 installed (optional)
- At least 1GB of free space on disk
- At least 4GB of RAM

The Dataset (1/3)

- Aruba Dataset from CASAS project http://casas.wsu.edu/datasets/
  - Dataset covering life of a woman for two years
  - Labeled dataset
    - Meal_Preparation (1606)
    - Relax (2910)
    - Eating (257)
    - Work (171)
    - Sleeping (401)
    - Wash_Dishes (65)
    - Bed_to_Toilet (157)
    - Enter_Home (431)
    - Leave_Home (431)



The Dataset (2/3)

- PIR sensors MXXX
- Door closure sensors DXXX
- Temperature sensors TXXX
- Sensor row format <date time sensor value [label]>
  - The label denotes whether an activity starts or ends
  - Interleaved activities

The Dataset (3/3)

- **The aruba folder in the tutorial kit contains:**
  - The data file containing the dataset
  - The aruba.jpg file containing a map of the aruba experiment
  - The README file describing the dataset
  - The aruba_sensor_map.csv containing rows in the format <Sensor X Y floor Room Object Note> where:
    - Sensor is the name of a sensor inside the dataset
    - X Y floor and Room represent the location with respect to the aruba.jpg file (X and Y are pixels)
    - Object is the name of the physical object in correspondence of the sensor

Segmenting the Dataset

- We want to split into process cases the original dataset

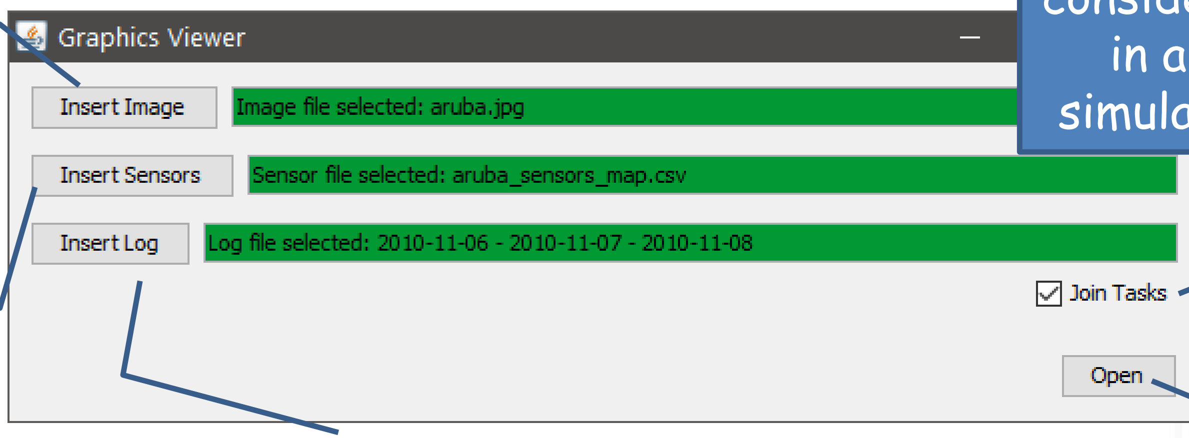
- Instructions:
  - Copy the data file in the segmentation folder
  - Run segmentation.py (you need Python3)
    - To run it on Python2 you only need to change print() to print (no round brackets) inside segmentation.py
- Two folders are generated in the segmentation folder:
  - **Output_date**: here we have a file for each day (the habit here is the daily routine)
  - **Output_task**: here we have a file for each task (here we consider the activities separately)
  - A different file for each repetition of the habit/activity (process case)

Playing the Log (1)

- Execute GraphicViewerIntegrated-1.0-SNAPSHOT-shaded.jar
  - Trajectory Analysis Tool
  - One of our original contributions

1) Load the map of the house

2) Load sensor positions

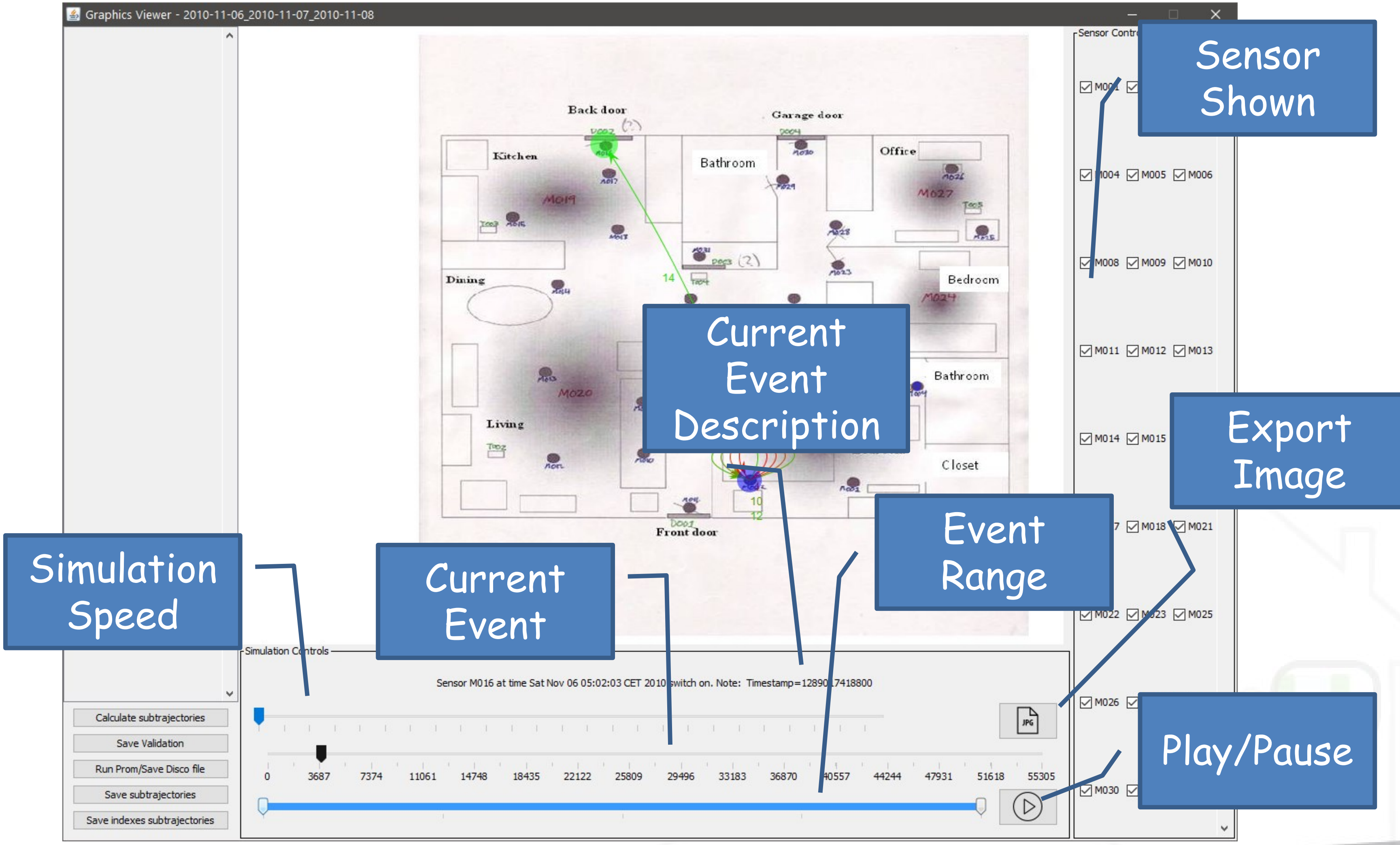


3) Load multiple files (cases) from the same habit/activity

4) If unchecked, consider every case in a different simulation window.

5) Load simulation window

Playing the Log (2)



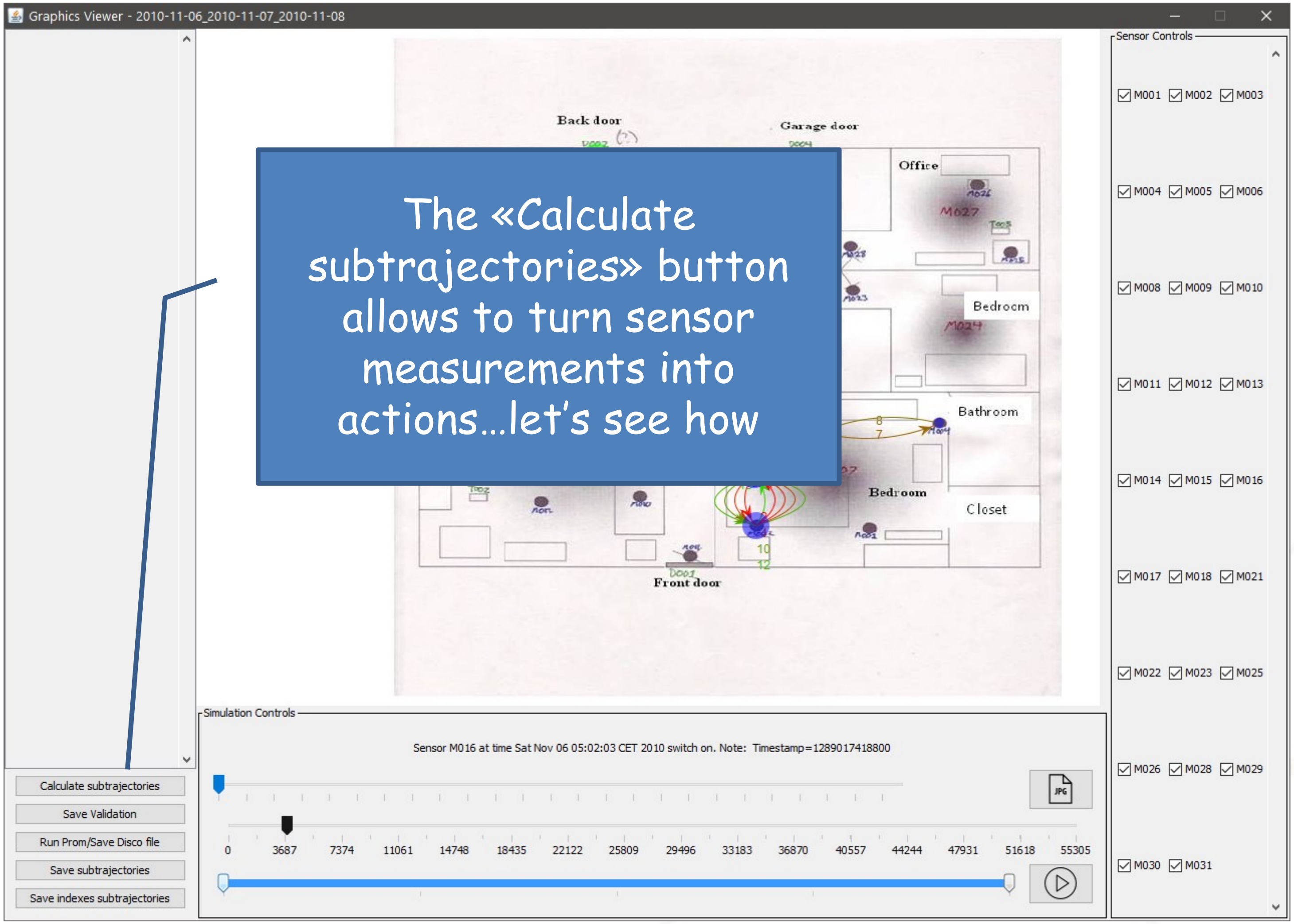
SAPIENZA UNIVERSITÀ DI ROMA

Playing the Log (3)

While playing, the path of the user is shown:

- Path color denotes the age (from green – more recent – to blue – older) of the measurement
- A number is associated to each edge (the bigger the newer)
- The size of the dot reflects the time spent under the sensor

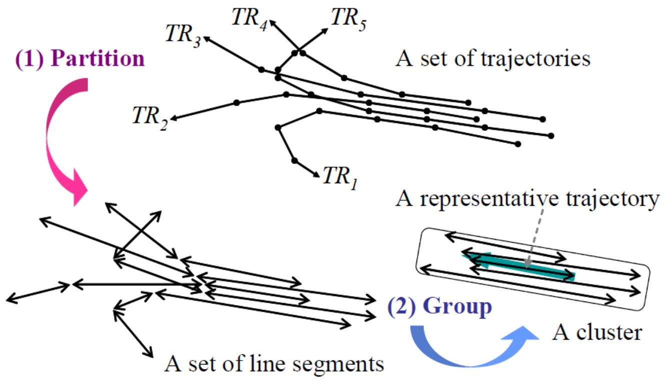
Computing Subtrajectories



The «Calculate subtrajectories» button allows to turn sensor measurements into actions…let's see how

Bridging the Gap between Sensor Logs and Event Logs

- TRACLUS [Lee2007]: Trajectory clustering algorithm
  - Two phases:
    - Trajectory partitioning
    - Density-based line-segment clustering



- We can now classify each trajectory as a specific movement action: STAY, AREA, MOVEMENT

Bridging the Gap between Sensor Logs and Event Logs

Given a trajectory δ returned by TRACLUS

$\boldsymbol{I_m(\delta)}$ reflects how many sensors are involved in the trajectory

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

$\boldsymbol{I_a(\delta)}$ reflects how trajectory time is distributed among sensors (Gini coefficient)

$\boldsymbol{I_s(\delta)}$ reflects how much time is spent under a single sensor

$$I_s(\delta) = \frac{time\ spent\ under\ the\ most\ frequent\ sensor}{total\ time\ of\ trajectory}$$

Bridging the Gap between Sensor Logs and Event Logs

Classification Index:

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

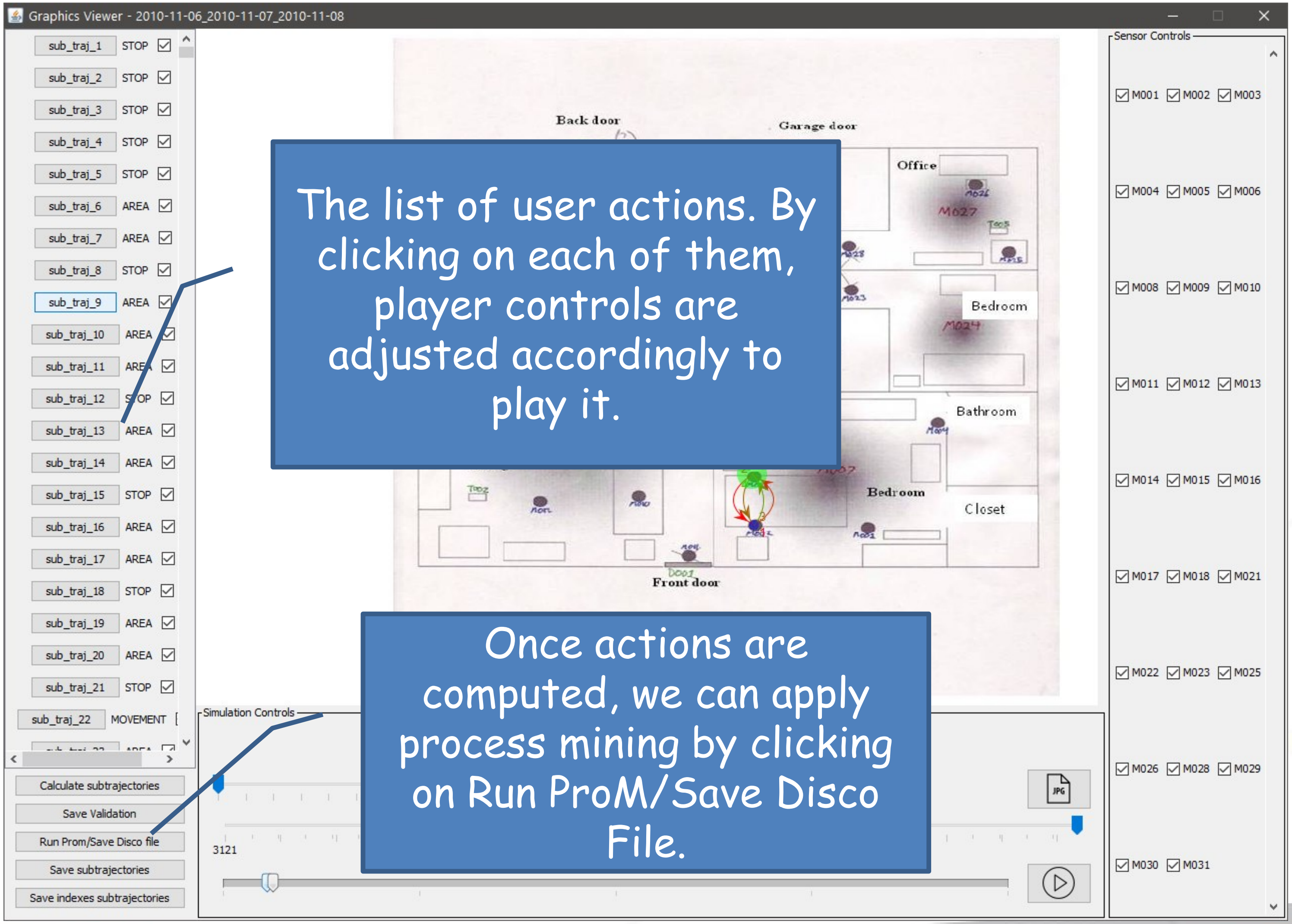
With:

$$w_m + w_a + w_s = 1$$

Subtrajectory classification:

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

Computing Subtrajectories

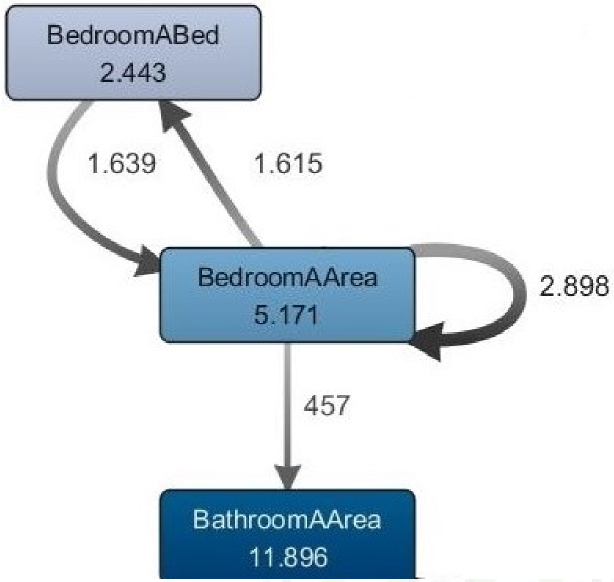


The list of user actions. By clicking on each of them, player controls are adjusted accordingly to play it.

Once actions are computed, we can apply process mining by clicking on Run ProM/Save Disco File.

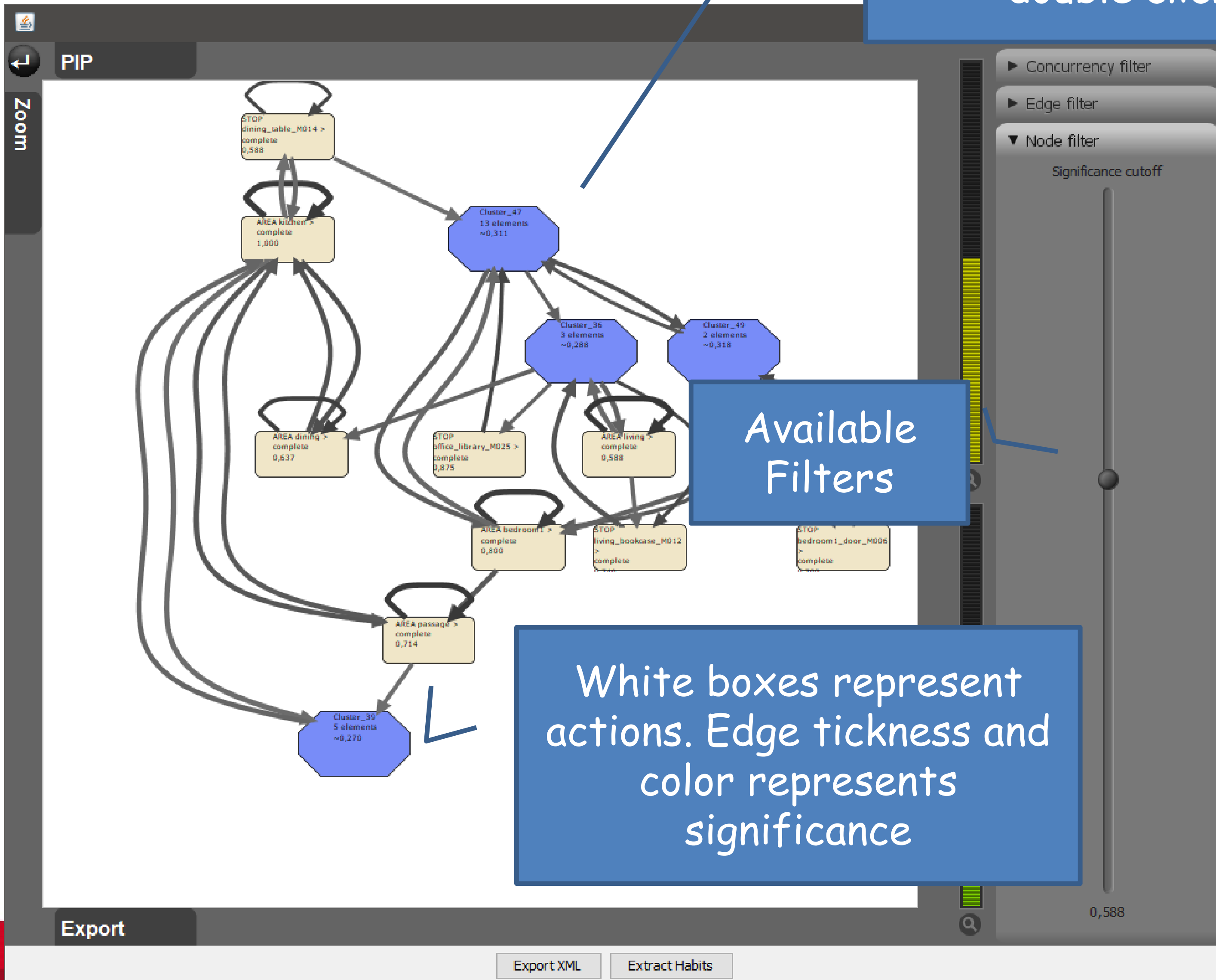
Discovering Human Habits

- Once the sensor log is turned into a (movement) action log, we can apply fuzzy mining:
  - Automated process discovery by using ProM http://www.promtools.org/doku.php
  - ProM fuzzy mining classes imported
  - The functionality saves the event log in a csv format compliant to ProM
    - Explicitly set .csv extension
  - Nodes representing actions
    - **In our case STAY or AREA actions**
    - ***MOVEMENT* actions ignored**



Fuzzy Mining Outcome

Violet nodes represents aggregation of events. You can navigate them by double clicking

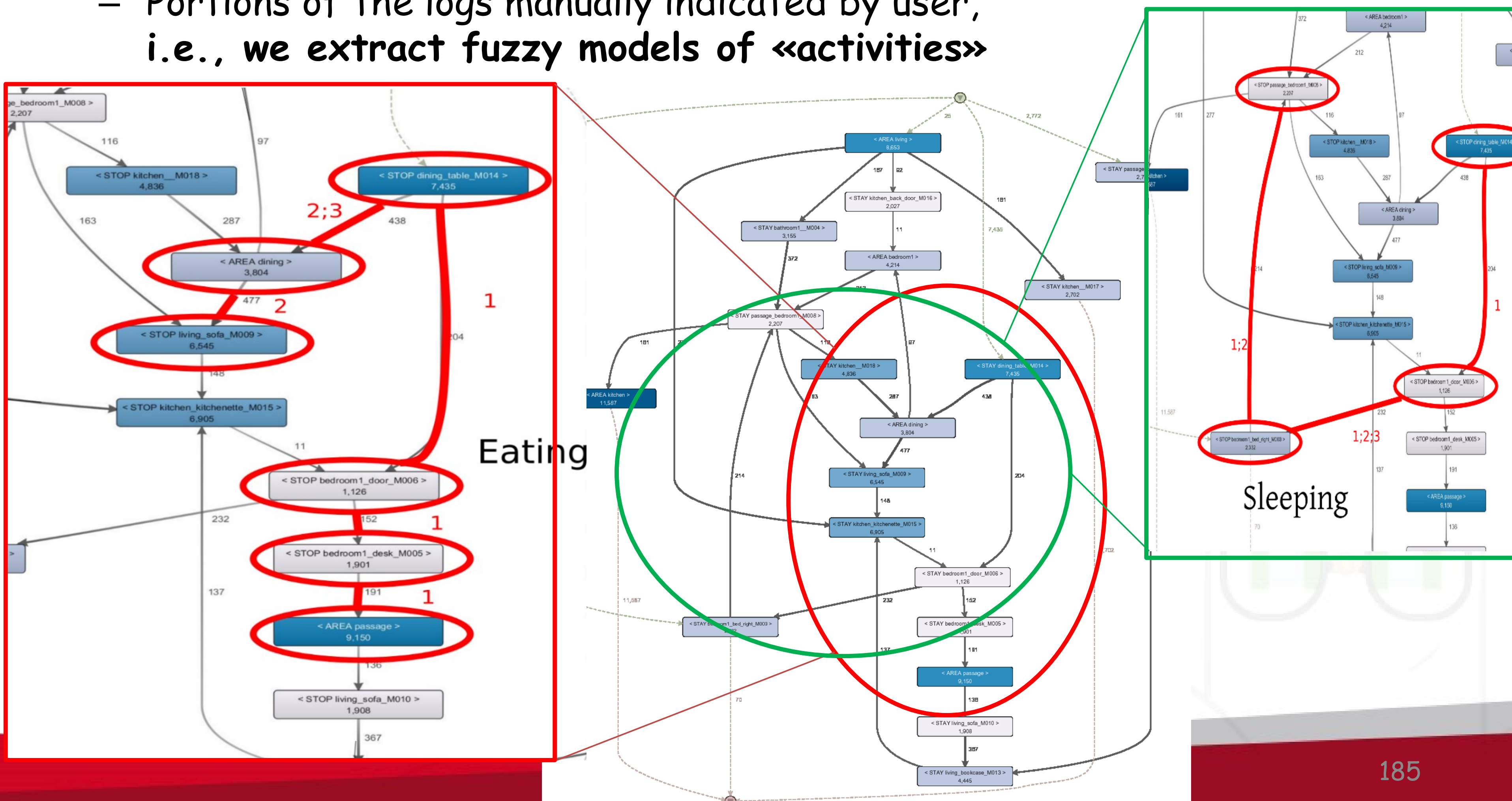


Available Filters

White boxes represent actions. Edge tickness and color represents significance

Discovering Human Habits

- We initially segment traces splitting on:
  - Entire days, **i.e., we extract fuzzy models of the «daily habit»**
  - Portions of the logs manually indicated by user, **i.e., we extract fuzzy models of «activities»**



Concluding Remarks on VPM

- Validation performed:
  - By comparing the discovered daily habit model wrt. models obtained from the labeled activities of CASAS
  - By comparing VPM against other visual representation tool, namely SITUVIS
  - Full details in:
    - Leotta F., Mecella M., Sora D. "Visual Process Maps: A Visualization Tool for Discovering Habits in Smart Homes." Journal of Ambient Intelligence and Humanized Computing, 2019.

Current Work: Multiple Users

- Presence of multiple users requires an additional level of label
- It is possible to recognize single user traces by using tracking?
  - Experiments on SVM based tracking
- Future experiments on deep learning

Current Work: Decision Mining

- Usually mined process models do not contain contextual information used to make decisions
- Integration of ECA rules in process models
- Imperative models instead of fuzzy mining

Current Work: Applications

- Previous work only focused on visual inspections of human habits
- How to include other process mining techniques for applications:
  - Conformance Checking → Anomaly detection
  - Predictive analysis → Smart Space automation

What to take home?

- An analysis of the state of the art demonstrating that further research is needed in:
  - Representing activities/habits
  - Very few has been done for decision making beyond simple reactive behaviors
    - Cf. step 5 of slide 19
- Playing with real software is helpful to understand issues and proposed solution
  - Repeatability is crucial for the community
  - Cf. VLDB/SIGMOD repeatability initiative
  - Standard benchmark datasets and tools for building them
    - Different levels of complexity based on available sensor types

What to take home?

- Process Mining techniques can be adopted in the context of smart spaces
- VPM is a first attempt toward this direction
  - The BPM community is already moving in this direction
    - More generally towards IoT

Thanks for Your Attention

- If you have further questions please send us an email at:
  - mecella@diag.uniroma1.it
  - leotta@diag.uniroma1.it