



SAPIENZA
UNIVERSITÀ DI ROMA

Applying Process Mining to Smart Spaces: Challenges



What is a Process in Smart Spaces

- A process is defined as a set of tasks with precedence relations
- **Activity and habits can be considered the equivalent of processes**
 - **Activity:** a sequence of actions (one in the extreme case) or sensor measurements/events with a final goal
 - Activities can be collaborative
 - **Habit:** a set of interleaving of activities that happen in specific contextual conditions
 - E.g., what a user does each morning between 08:00 and 10:00am
 - E.g., what a user does between very specific actions (e.g., leaving the bed and leaving the house)
- **Tasks of activity and habits are actions**

Dealing with Granularity

- Clear gap between the granularity of sensor logs and the traces used for process mining [Baier2013]
- No one-to-one correspondence between sensor measurements and performed actions (tasks)
 - A single user action may trigger many sensor measurements
 - A single sensor measurement may be related to several actions
- **Required approach:**
 1. **Aggregate sensor measurements to recognize actions**
 2. **Apply process mining**
- The kind of available sensors strongly influences the granularity and confidence of recognized actions

Log Segmentation (1/2)

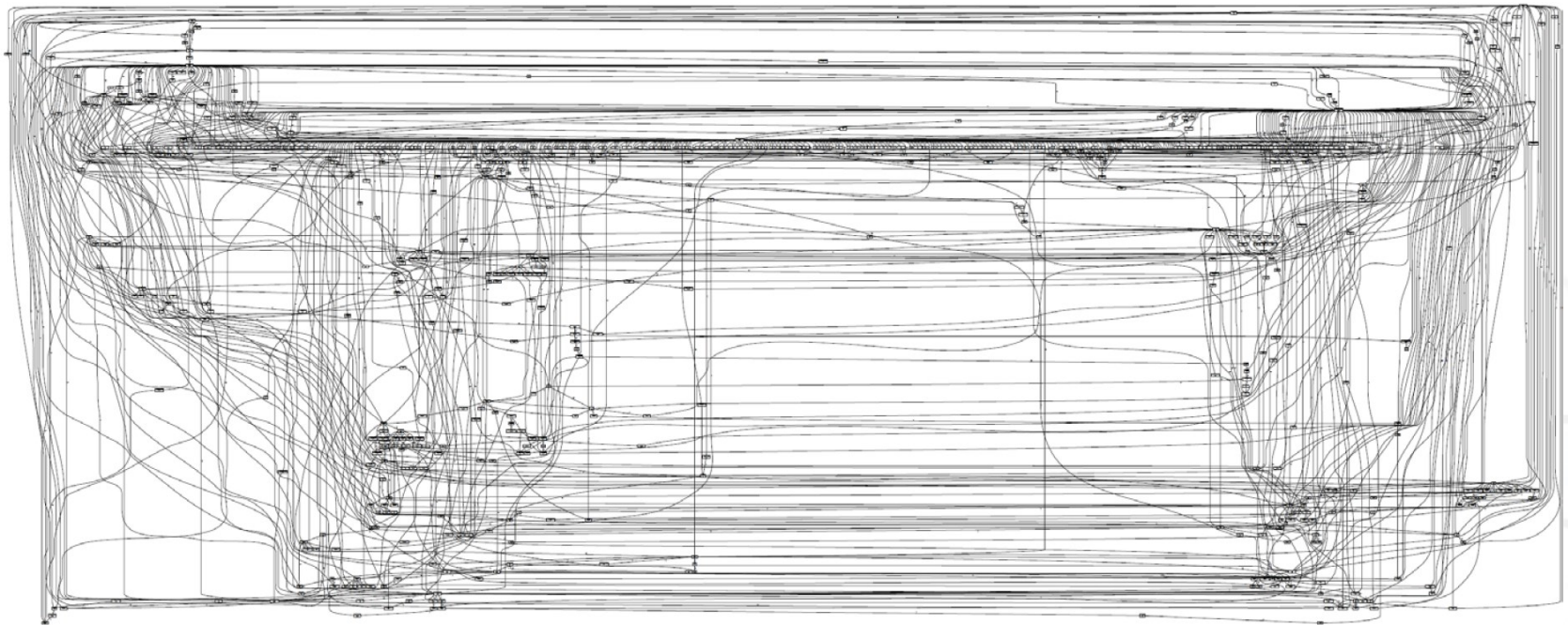
- A common prerequisite of process mining techniques is to have an event log explicitly segmented into cases (process instances)
 - Case “start” and case “end” events
 - For each event, which case it belongs to
 - Relatively easy to instrument a process in an industrial or business environments
- This assumption is usually not met by sensor logs, as labeling is generally an expensive task to be performed by humans
 - *Especially difficult to associate actions (derived from sensor measurements) to activities and habits in the interleaved case and in presence of multiple users*

Log Segmentation (2/2)

- How do we define habits and activities?
 - Manually defined?
 - Automatically learned and adapted?
 - **Active learning?**
- What about multiple users?
 - Usually sensor logs do not contain any information about which user(s) caused a certain sensor to trigger or to provide a specific measurement
 - The employment of body-area sensors and tags is usually perceived as invasive by the user and do not solve all the issues
 - **Mining habits in a multi-user scenario is significantly harder**
 - e.g., even though multiple users can be identified by the spatial distance between PIRs triggering close in time, when trajectories intersect
 - tracking techniques or reasoning must be employed to keep following users

Which Formalism? (1/2)

- Question: Does a human habit resemble a “spaghetti” process?
 - Approaches to deal with unstructured processes do exist as both imperative and declarative modeling formalisms
 - Human processes in smart spaces are very similar to “artful” processes (e.g., treating patients in hospitals)



Which Formalism? (2/2)

- Declarative modeling formalisms
 - Usually based on temporal logics (e.g., DECLARE [Pesic2007])
 - Already applied to smart spaces for reasoning [Magherini2013]
 - The notion of time is qualitative and not quantitative
 - Time is a first-class property of a measurement
 - Attempts to support a quantitative notion of time [Westergaard2012]
 - **Are typical constraints enough?**
- Fuzzy mining [Günther2007]
 - Borrows concepts from the world of maps and cartography
 - Zoom in and out on a process model highlighting the importance of certain tasks and connection between tasks
 - **More suitable for offline analysis than for online monitoring**

Westergaard, M., Maggi, F.M.: Looking into the future. In OTM 2012

Magherini, T., Fantechi, A., Nugent, C.D., Vicario, E.: Using temporal logic and model checking in automated recognition of human activities for ambient-assisted living. IEEE Trans. Hum. Mach. Syst. 2013

Pesic, M., Schonenberg, H., van der Aalst, W.M.P.: Declare: full support for loosely structured processes. In EDOC 2007

Günther, C.W., van der Aalst, W.M.P.: Fuzzy mining - adaptive process simplification based on multi-perspective metrics. In BPM 2007

Maps and Cartography (1/2)



- Road map of Italy
- Abstract from small cities and roads
- Big cities aggregate local roads
- Usage of color and size

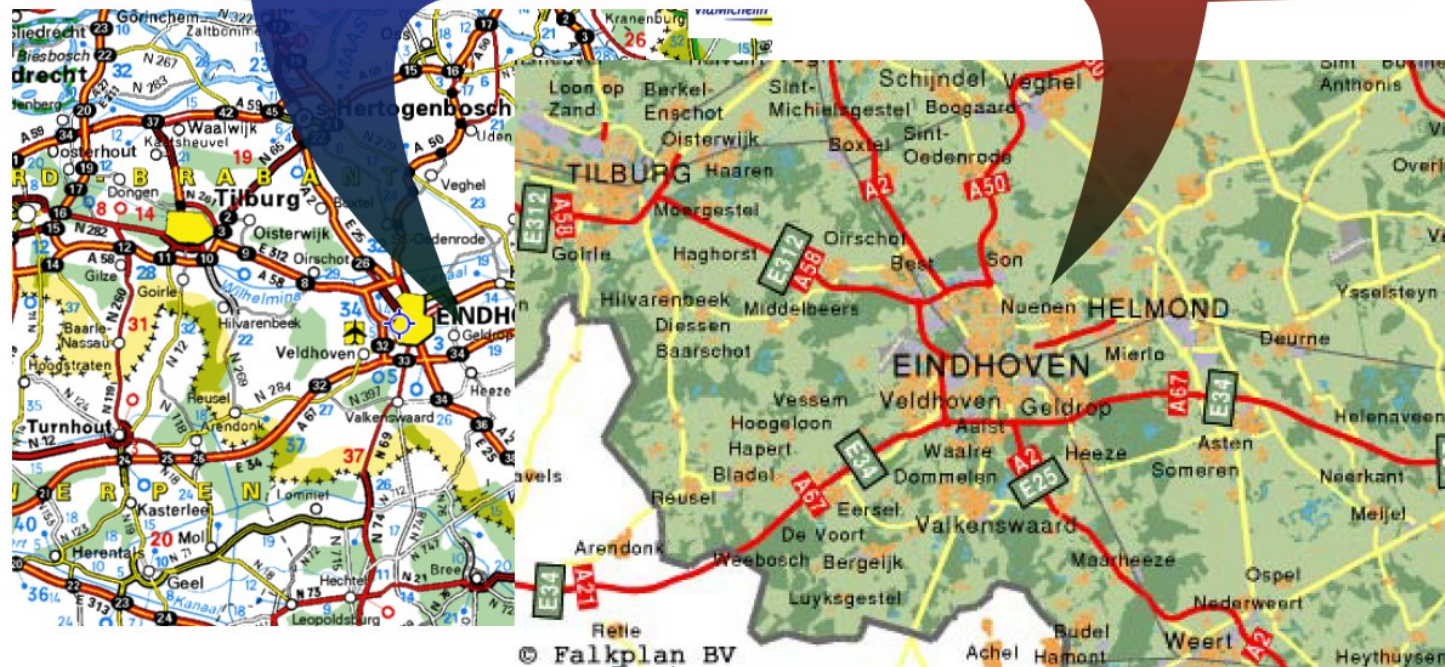
Maps and Cartography (2/2)

Aggregation

Clustering of coherent, less significant structures

Abstraction

Removing isolated, less significant structures



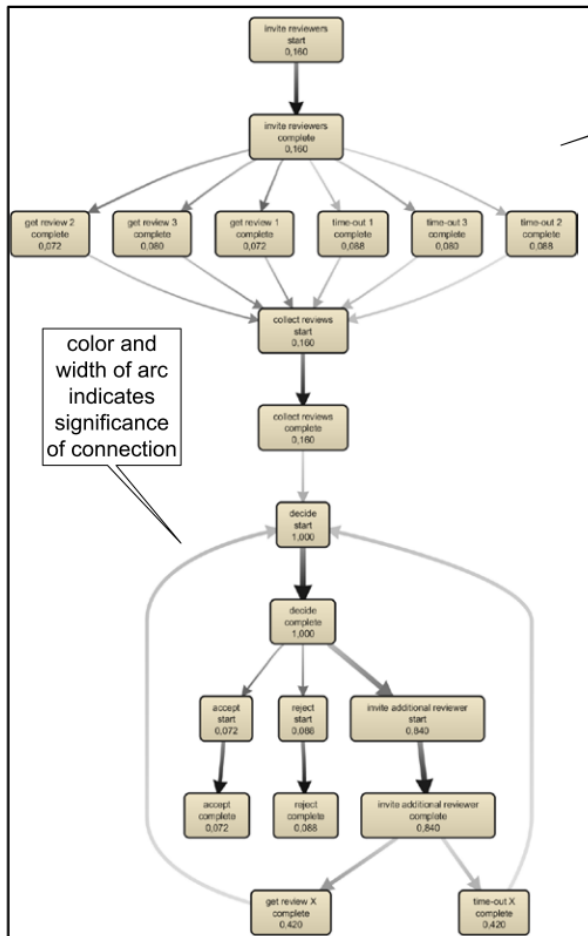
Fuzzy Mining (1/2)

- Fuzzy Mining takes as input different cases of a process and compute a graph $G = \langle V, E \rangle$ where V is a set of nodes and E is a set of directed arcs
 - An arc between two nodes $v1$ and $v2$ is present if $v1$ precedes (even not immediately) $v2$
- Significance Metrics
 - Measure the importance of an event (unary metric) or a precedence relation (binary metric)
 - Frequency in the log
- Correlation Metrics
 - Binary metrics showing how closely related two events are
 - Distance in log is taken into account
 - Deep comparison (e.g., names of the events)

Fuzzy Mining (2/2)

- Aggregated metrics are obtained by combining significance and correlation metrics
- What kind of metrics are considered can be tuned
- During the analysis, filtering based on thresholds is employed to filter out and to aggregate events and arcs
- Fuzzy mining is supported by commercial tools (e.g., Disco by Fluxicon) and open source tools (e.g., ProM)

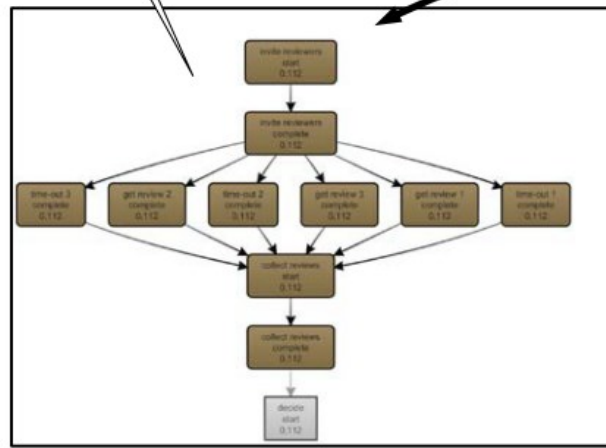
Fuzzy Mining: Applying filtering



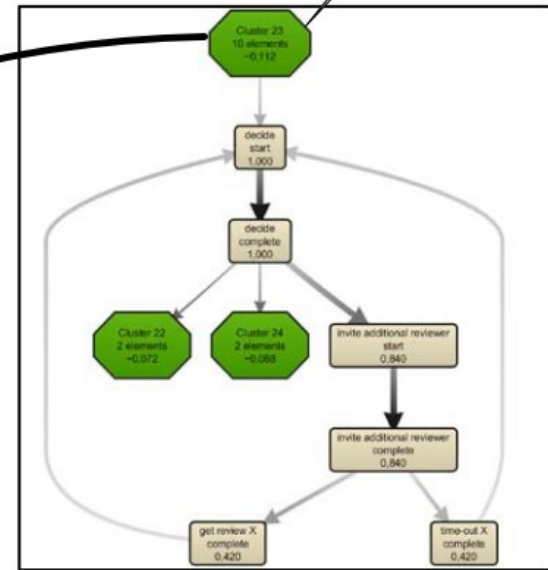
fuzzy model showing all activities

color and width of arc indicates significance of connection

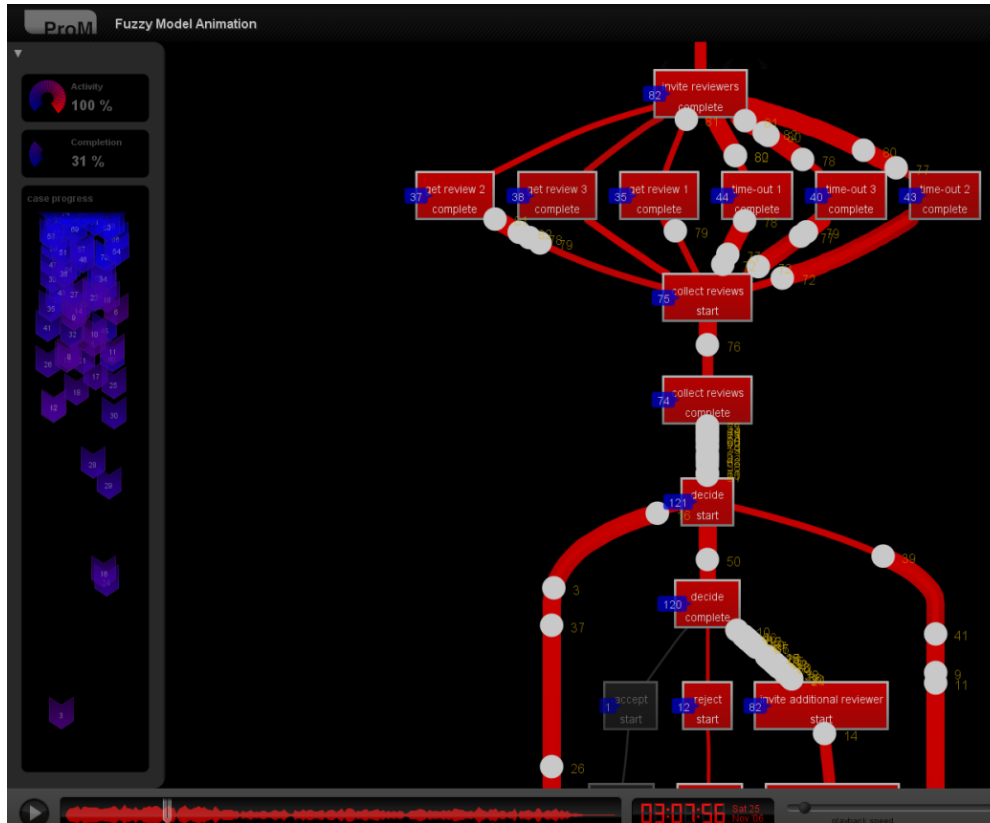
inner structure of aggregated node



aggregated node containing 10 activities



Fuzzy Mining: Replaying



- Fuzzy Mining main intent is analysis
 - No enactment
- The availability of players allows to replay logs on the models