



*Big Data Analytics:*
lançando luz dos genes ao cosmos

# Multiple Aspect Trajectories Modeling and Integration

**Ronaldo dos Santos Mello**

UFSC/INE/PPGCC/GBD
*r.mello@ufsc.br*

Programa de Pós-Graduação em Ciência da Computação UFSC

GBD INE-UFSC

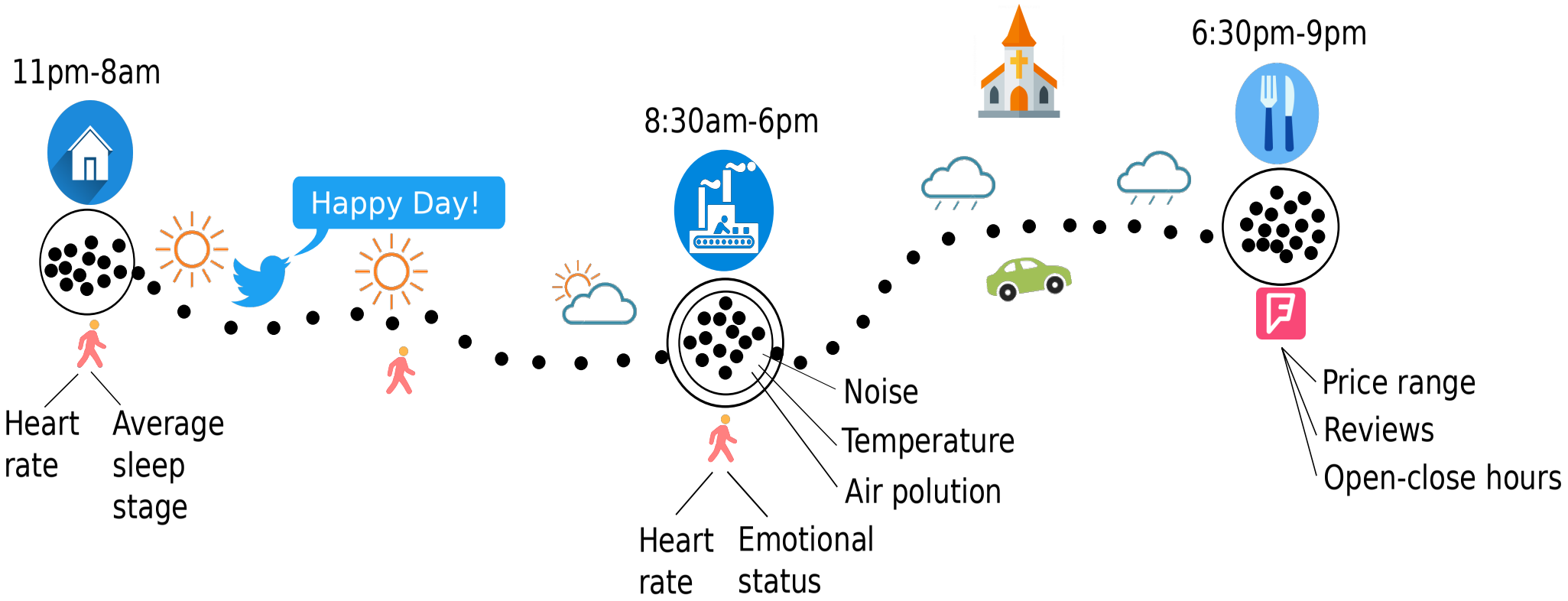
# Ronaldo dos Santos Mello

- Professor/researcher at Informatics and Statistics Department of UFSC (INE/UFSC, Santa Catarina state, Brazil)
- Teaching/research area: Databases
- Current research interests
  - data modeling for complex data
  - data integration for complex data
  - integrity constraints for complex data
  - Big Data management issues
    - NoSQL, NewSQL, Data Lakes, ...

# Multiple Aspect Trajectory (MAT)

- A trajectory (or any part of it) that may be enriched with different aspects
- Aspect
  - a semantic context associated to an aspect type with specific properties
  - examples
    - *Sheraton* (a *hotel* aspect type with *stars*, *facilities*, …)
    - *Pisa Tower* (a *POI* aspect type with *location*, *founded*, …)
    - *Happy* (an *emotional* aspect type with *intensity*, *emoticon*, …)

# MAT Example


11pm-8am
Happy Day!
8:30am-6pm
6:30pm-9pm
Heart rate
Average sleep stage
Noise
Temperature
Air polution
Heart rate
Emotional status
Price range
Reviews
Open-close hours

# Research Activities related to MAT

- MAT modeling
  - complete DB design of MATs
  - modeling of relevant data for MAT analytics



- MAT integration
  - generation of a representative MAT for a set of similar MATs

# Research Activities related to MAT

- MAT modeling
  - complete DB design of MATs
  - modeling of relevant data for MAT analytics

Transactions in GIS WILEY

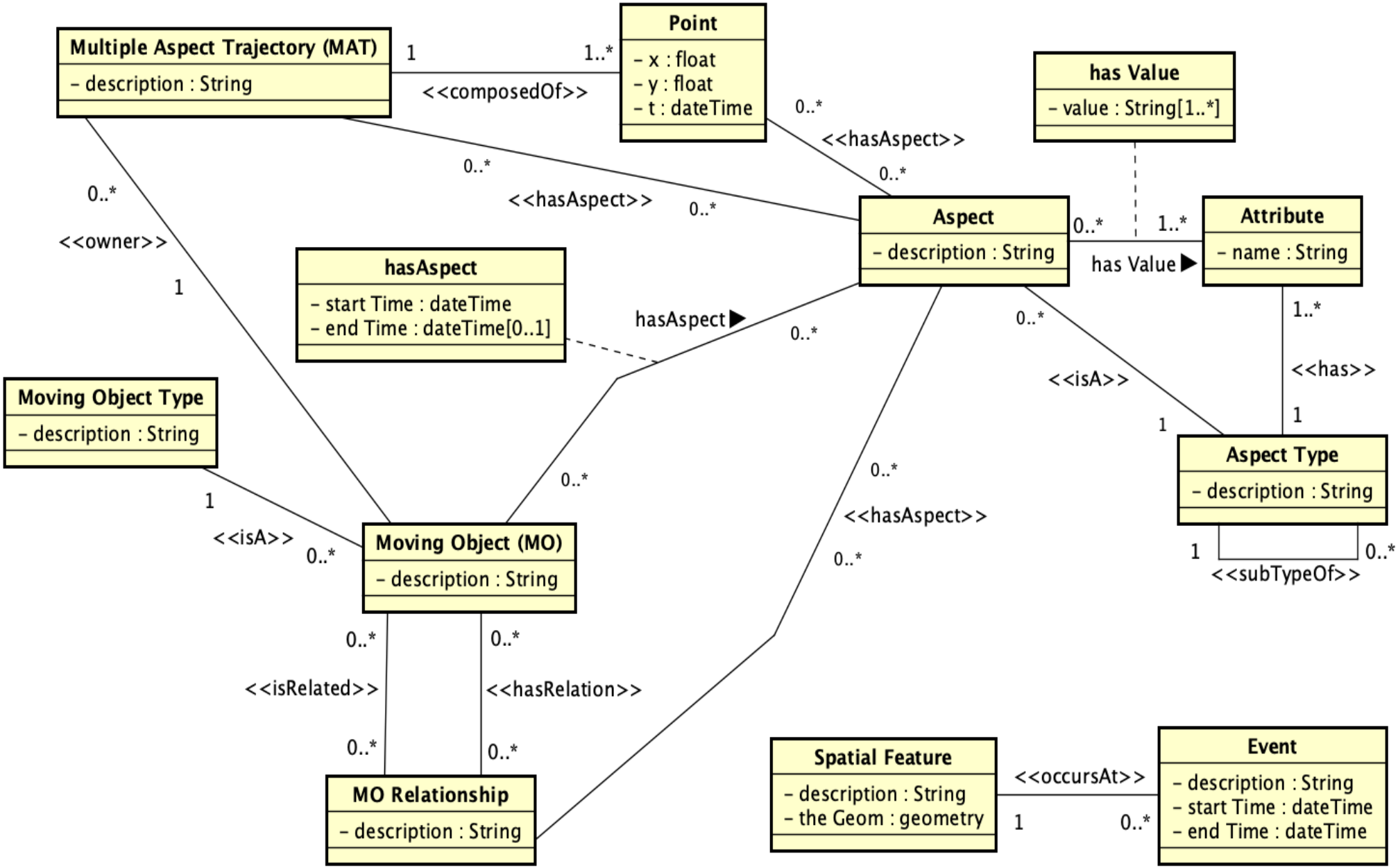
acm SAC2020

- MAT integration
  - generation of a representative MAT for a set of similar MATs

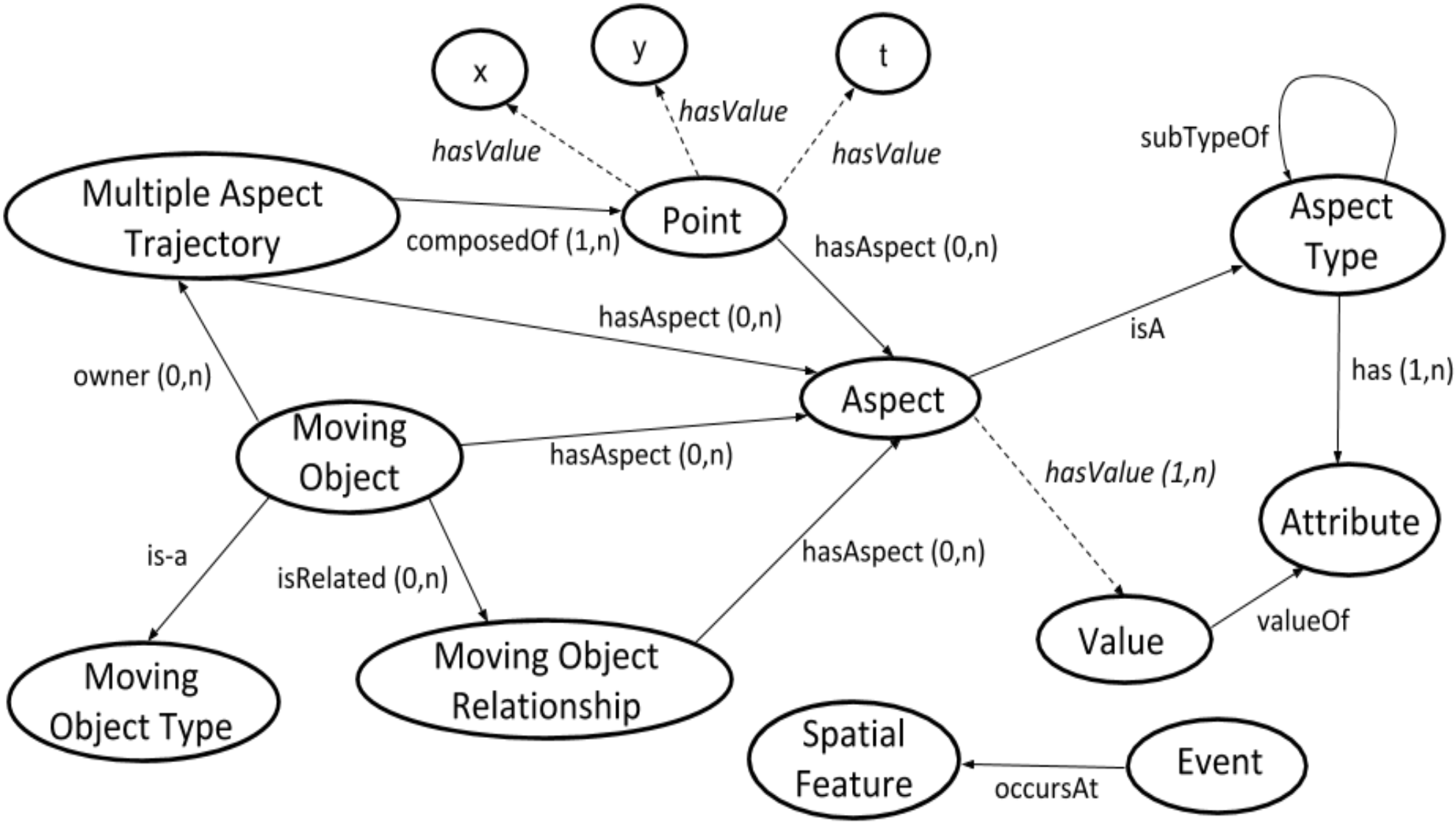
# Complete DB Design of MAT

- Goal
  - a reference data model for any application that intends to represent, manipulate and store MATs
    - a simple but expressive data model
  - we consider all DB design phases
    - conceptual, logical and physical design
- Related Work
  - focus on one or few trajectory semantic features
  - no detailed DB design

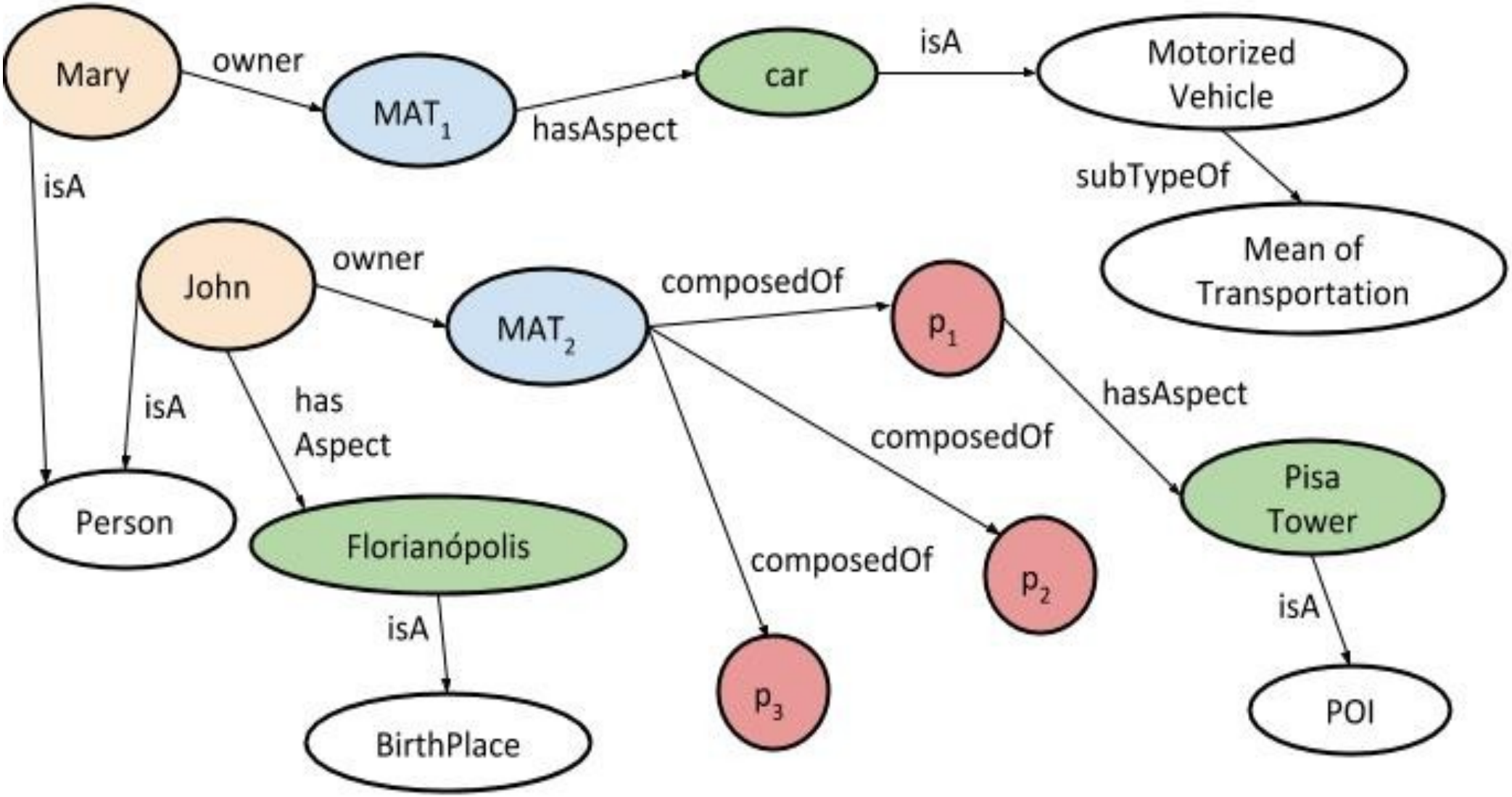
# MAT Conceptual Model



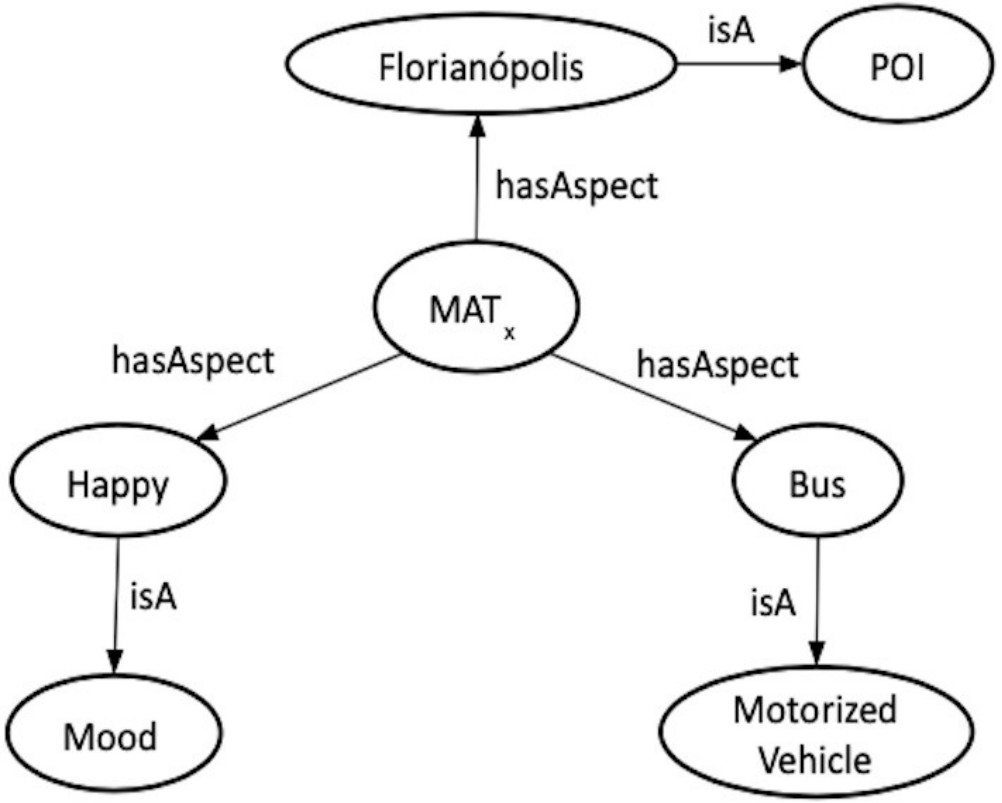
# MAT Logical Model (Graph)



# Example of MAT Instance



# MAT Storage (NoSQL DBs)

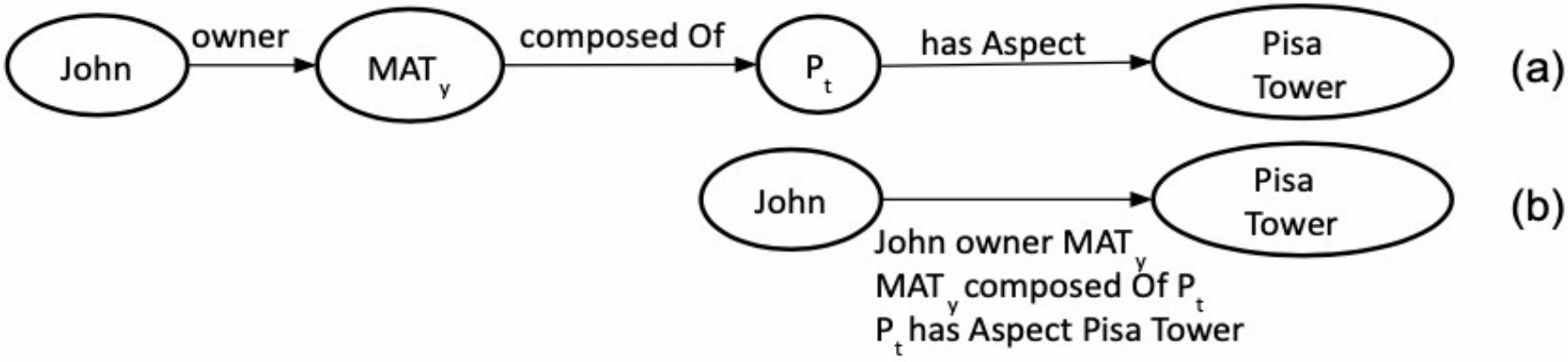



(a)

```
{
  "subject": "Mat x",
  "hasAspect": [
      {
        "aspect": "Bus",
        "aspectType": "Motorized Vehicle"
      },
      {
        "aspect": "Florianópolis",
        "aspectType: "POI"
      },
      {
        "aspect": "Happy",
        "aspectType": "Mood"
      },
                  ]
}
```

(b)

Document DB (MongoDB)




Graph DB (Neo4j)

# MAT Modeling Experimental Evaluation

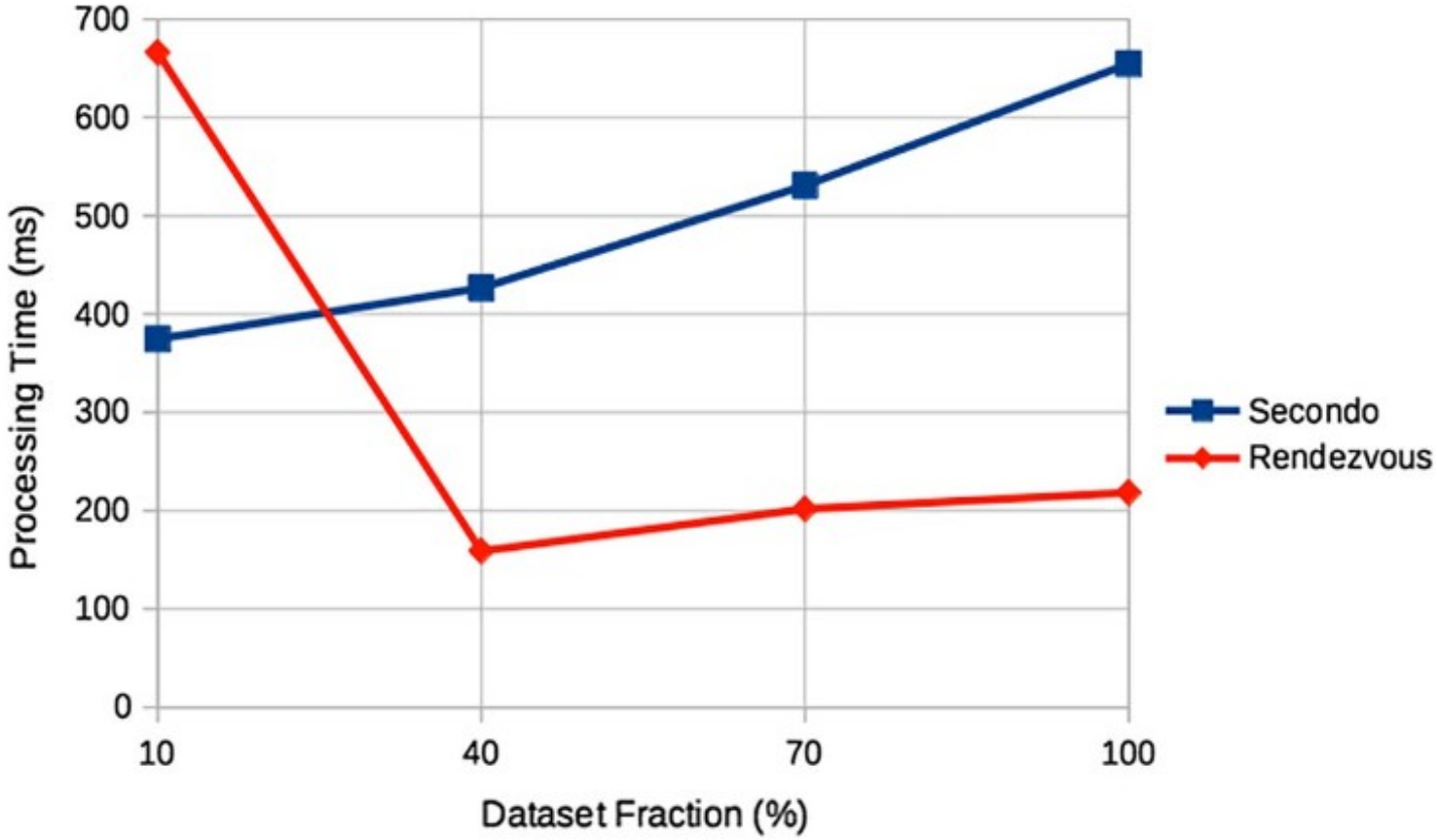



**FIGURE 9** Comparison of query performance between SECONDO and Rendezvous over a larger number of nodes

# Research Activities related to MAT

- MAT modeling
  - complete DB design of MATs
  - modeling of relevant data for MAT analytics



- MAT integration
  - generation of a representative MAT for a set of similar MATs

# MAT Modeling for Analytics (*MAT Analytics*)

- Motivation
  - several valuable analysis may involve trajectory data + aspect data
  - an application scenario

> "*Suppose that trajectories of people that go to Copacabana restaurants (a POI aspect) by car (a mean of transportation aspect) usually post bad reviews (a sentiment aspect). From this analysis, restaurant owners could take several actions to minimize such a kind of reviews, like an upgrade of their menu items or an improvement of their parking and security facilities.*"

# MAT Analytics Data Modeling

- Open Issue
  - no work related to modeling and storage of patterns involving moving objects, trajectories and aspects at the same time
- Challenges
  - flexibility to model heterogeneous correlations among MAT features
    - based on the result of different data mining techniques
  - efficient storage and querying
    - databases for Big Data (NoSQL, NewSQL, ...) could be good solutions
- Main contribution
  - no need to rerun data mining processes over large datasets in order to access discovered patterns because these patters will be stored and available for querying

# Staying at ISTI/CNR-Pisa (2018)

- A *SoBigData* Consortium visiting research grant


- Goal
  - to analyze trajectory datasets to get insights about MAT analytics modeling
- Considered data repository
  - *GPS Tracks Tuscany by Volunteers*
    - a group of datasets with 8392 trajectories of 129 people
    - one of the few repositories that holds aspects related to trajectories

# Analyzed Datasets

- Two embedded *GPS Tracks Tuscany by Volunteers* datasets were analyzed
  - Diaries
    - Aspects: *distance, average speed, day of the week (weekend/week day), duration, period of the movement (start hour-end hour) and trajectory purpose (goal)*
  - MP
    - Aspects: *trajectory purpose (goal), mean of transportation and duration*
- Both datasets are relational DBs where some tables have trajectory data enriched with some aspects

# Methodology

1) To identify the relational tables that maintain aspects related to trajectory data
2) To define and execute grouping queries (`SELECT … GROUP BY`) by varying the combination and amount of considered aspect attributes
3) To manually analyze the resulting groups in order to discover patterns

# Some Analysis Results

- MP Dataset

a) goal = ‘Leisure (sport, excursion, ...)’ OR goal = ‘Carburator Fixing’ ⇒ mean of transportation = ‘Motorcycle’ OR mean of transportation = ‘Automobile’ (100% of confidence)

b) goal = ‘Shopping (supermarket, ...) ⇒ mean of transportation = ‘Automobile’ (80% of confidence)

c) goal = ‘Pick up or drop out someone’ ⇒ mean of transportation = ‘Automobile’ (100% of confidence)

- Diaries Dataset

a) goal = "Work" OR goal = "Restaurant" OR goal = "Service" OR goal = "Study" OR goal = "Return Home" ⇒ NOT(day_of_the_week = "Sunday") (87% of confidence)

b) goal = "Shopping" ⇒ day_period = "12-18" (70% of confidence)

c) goal = "Restaurant" ⇒ day_period = "6-12" OR day_period = "12-18" (~ 80% of confidence)

d) goal = "Supermarket" ⇒ day_period = "6-12" OR day_period = "12-18" (99% of confidence)
(people do not go to supermarkets at night!) 

# Analytical Dependency (AD)

- The analysis of the discovered patterns reveals that
  - dependencies may be complex, *i.e.*, *determinant* and *determined* parts may hold a set of predicates connected by logical operators
  - despite the datasets have only aspects related to *whole trajectories*, it is possible that some dependencies rule only part of a trajectory or even the behaviour of the moving object
- We call *Analytical Dependency (AD)* this kind of dependency
- AD definition

"*a set of constraints over some aspects (determinant aspects) that usually determines a set of constraints over other aspects (determined aspects) in the context of a moving object, a trajectory as a whole, a trajectory point or a moving object relationship*"

# AD Preliminary Formal Definition

```
MO | MAT | POINT | MOR

{determinant aspect constraint}

⇒

{determined aspect constraint}
```

- *MO, MAT, POINT or MOR*: domain entities of the MAT data model
- *determinant/determined aspect constraint*: complex conditions connected by AND or OR, or enclosed by NOT

# Examples of ADs

## 1) People that visit Pisa Tower (POI) usually do a post in a social network:

```
POINT | composedOf.owner.is-a[description = 'Person'] AND
hasAspect[description = ’Pisa Tower’].is-a[description = ‘POI’]
=>
hasAspect[description = 'Post']
```

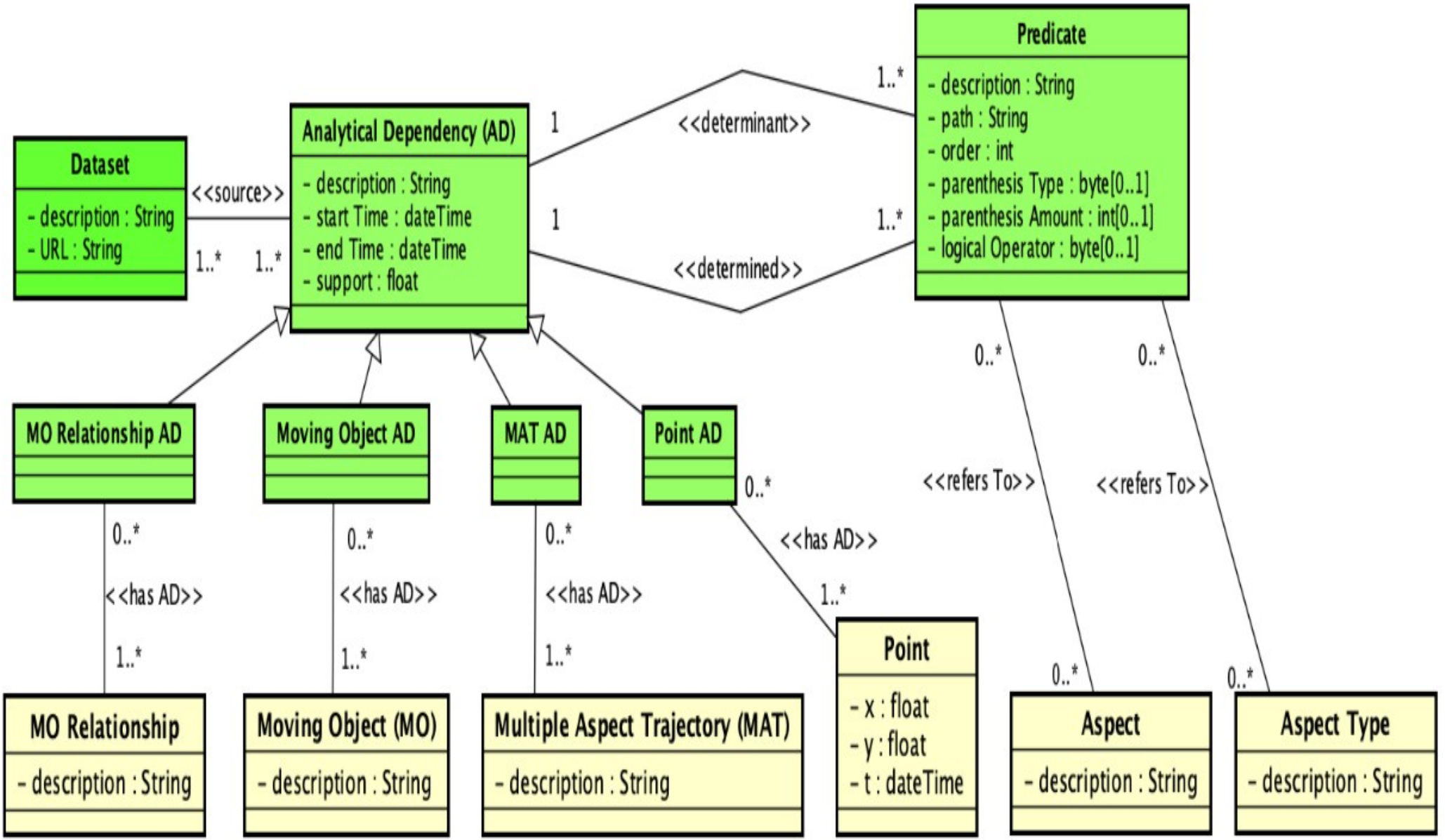
## 2) Trajectories of persons that visit Pisa and also Florence (POIs) moves by car or by train:

```
MAT | owner.is-a[description = 'Person'] AND
composedOf.hasAspect[description = 'Pisa'].is-a[description = ‘POI’]
AND
composedOf.hasAspect[description = 'Florence'].is-a[description = ‘POI’]
=>
hasAspect[description = 'car'] OR hasAspect[description = 'train']
```

# AD Inspiration

- *Association Rule (AR)*: a common representation for discovered knowledge
- *Functional Dependency (DF)*: a concept used in relational database normalization
  - both of them basically allow the definition of value dependencies between data items
- Different from AR and DF, an AD
  - also defines the domain entity (the focus of the AD)
  - allows the definition of complex conditions

# MAT Analytics Conceptual Model



# AD Power Expression

- An AD is able to represent several classes of data integrity constraints
  - boolean expressions
  - conditional rules
  - it may traverse all MASTER conceptual model entities and relationship (broad data range)
  - it may constraint the state of any attribute of an entity or relationship of the MASTER conceptual model

# Research Activities related to MAT

- MAT modeling
  - complete DB design of MATs
  - modeling of relevant data for MAT analytics



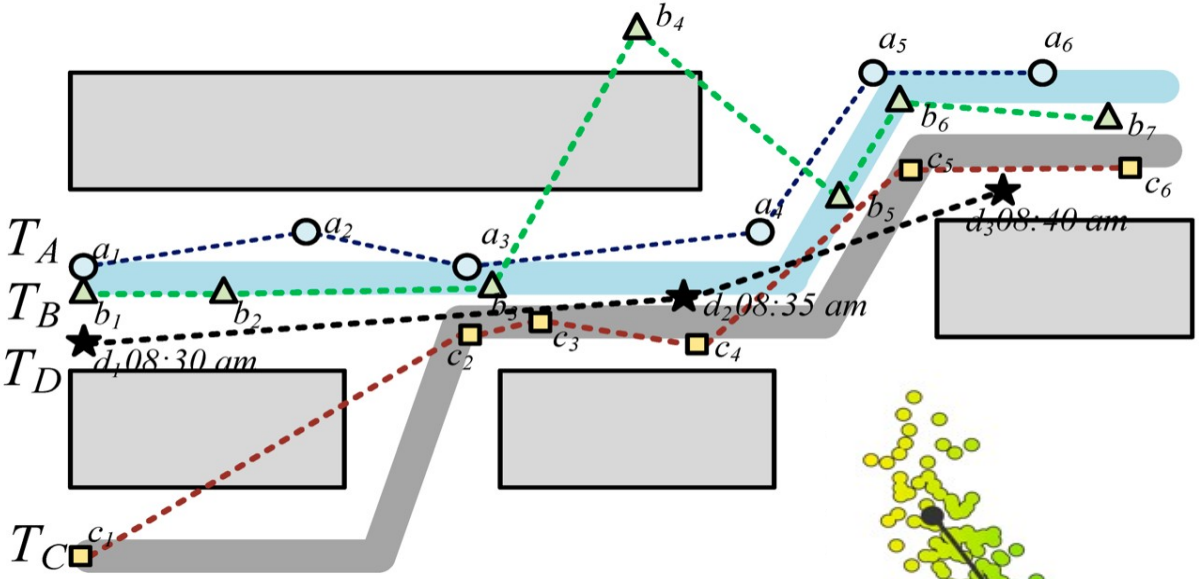
- MAT integration
  - generation of a representative MAT for a set of similar MATs

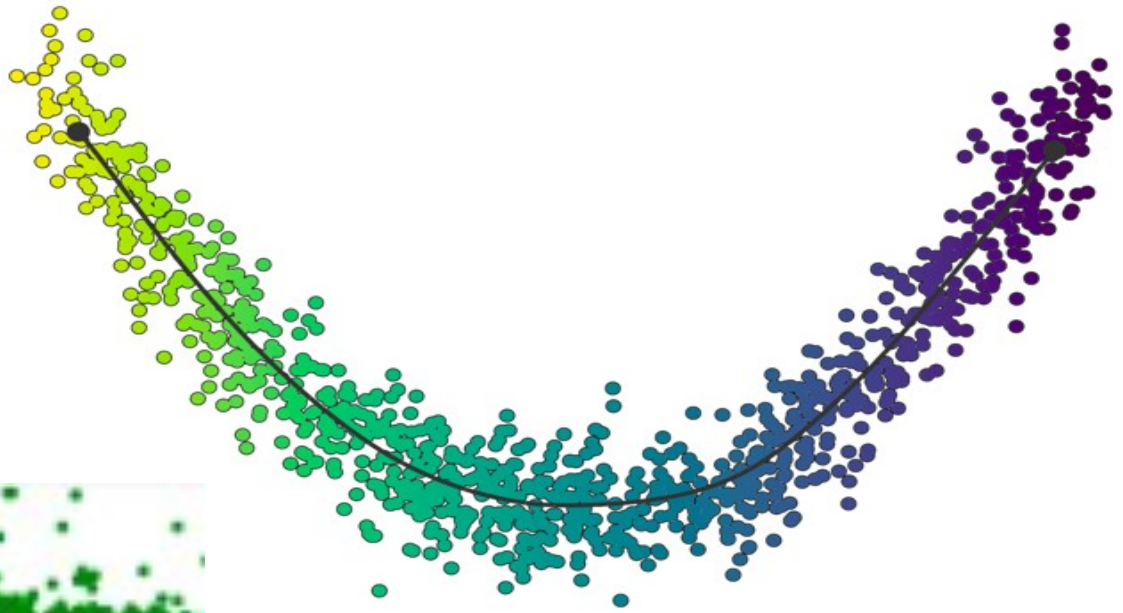
# MAT Integration

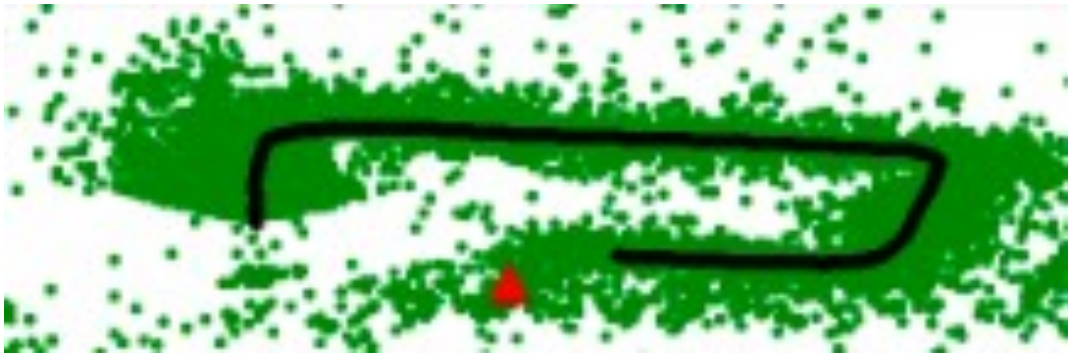
- Motivation
  - during a data analysis task, a data scientist or decision maker may get confused and lost when visualizing a large amout of trajectories
  - an application scenario

"*Suppose the Florianopolis mayor wants to analyze the behaviour of bikers in the island in order to improve their moving conditions. But there are so many bikers in the island and so many bikers` trajectories that it is impossible to get conclusions about their behaviours!! So, it would be more productive if the mayor could restrict this huge amount of trajectories to a more limited set of trajectories that could represent the most relevant trajectory behaviours of the bikers.*"

# MAT Integration – the Problem



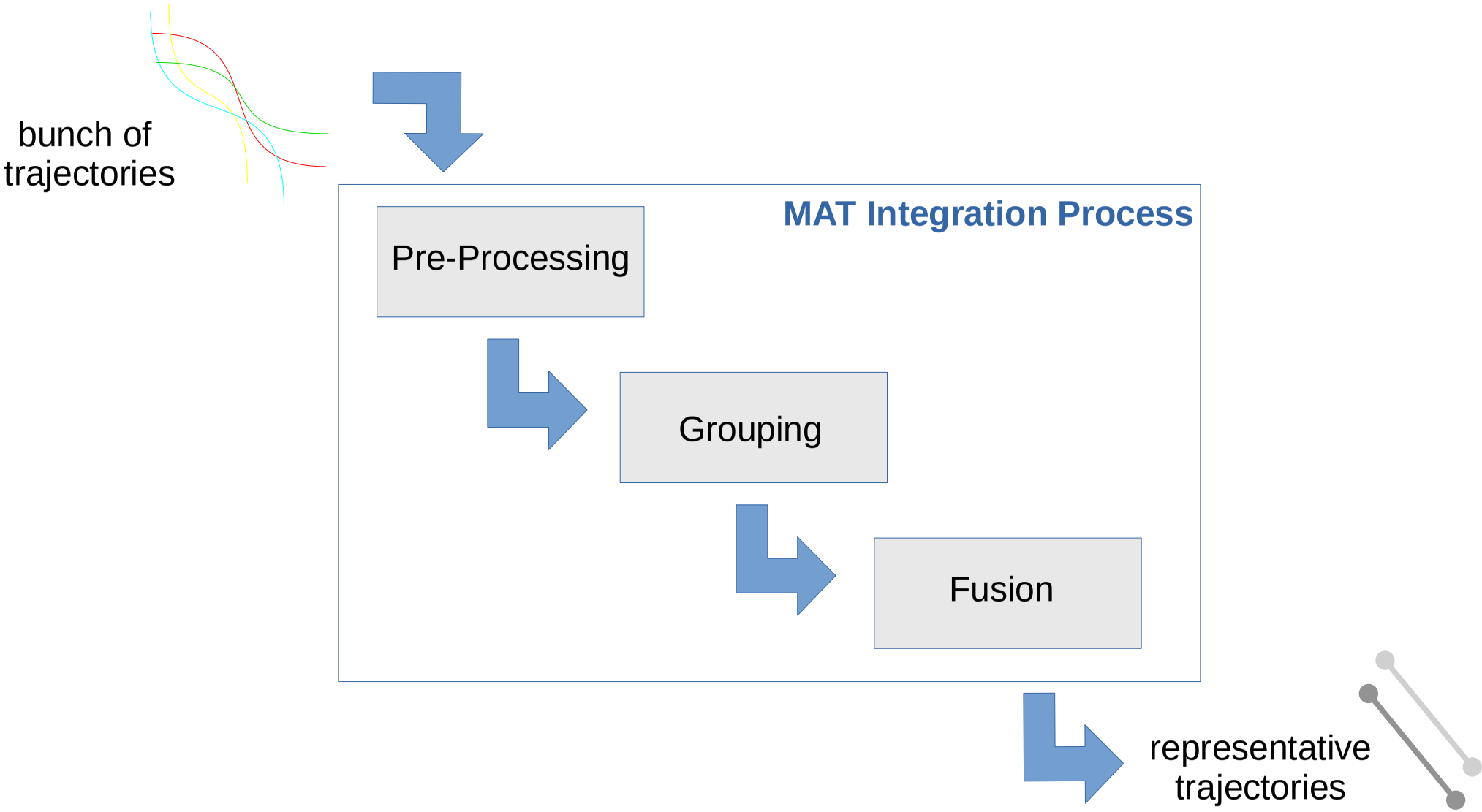





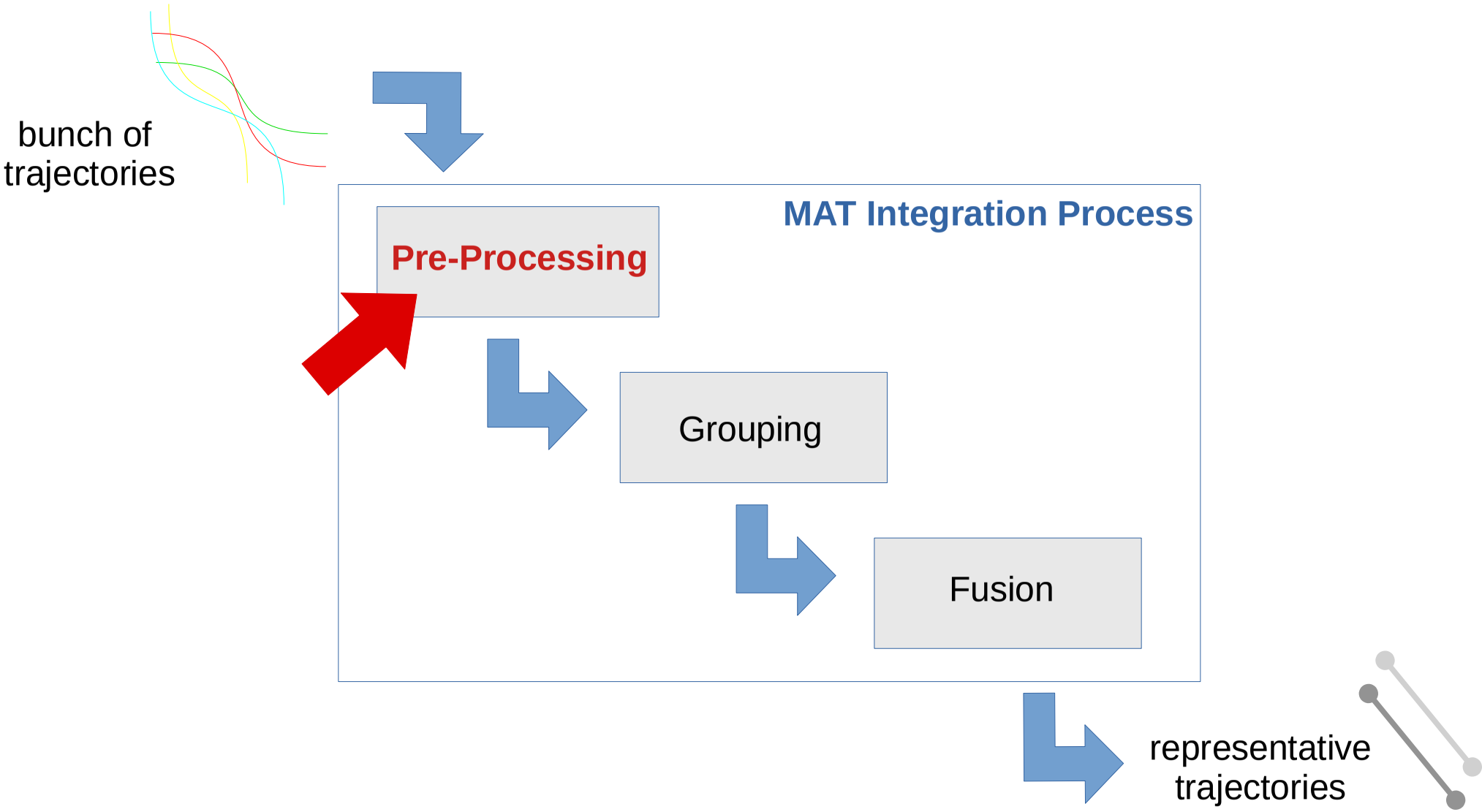
# MAT Integration – Related Work

- Trajectory integration approaches are based mainly on space and time dimensions
  - semantics (aspects) is not taken into account
- Provides a single (fixed) integrated view of trajectories
  - no flexibility to define different views of integrated trajectories according to the user needs
    - space, time, aspects or any combination of these 3 dimensions

# MAT Integration – A Process

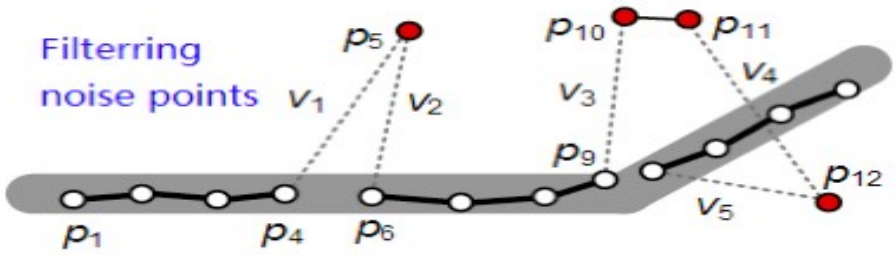


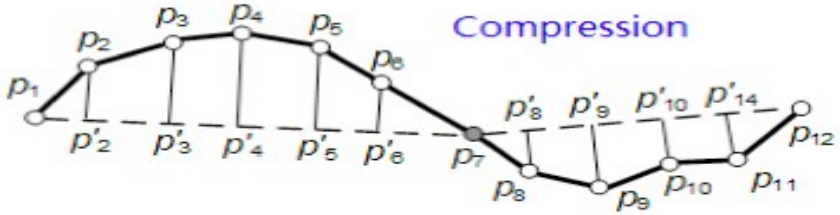
# MAT Integration – A Process

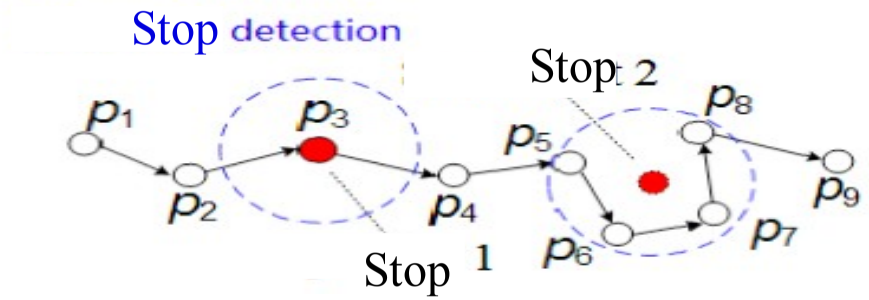


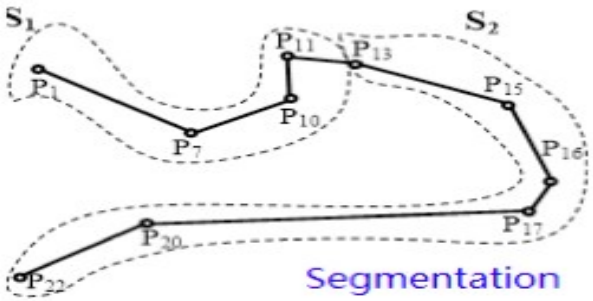
# Pre-Processing Phase

- Goal
  - To reduce the amount of MAT details to be considered for integration purposes
  - It is necessary when we have *Big trajectory Data*
- An optional phase
- Several techniques may be considered/adapted

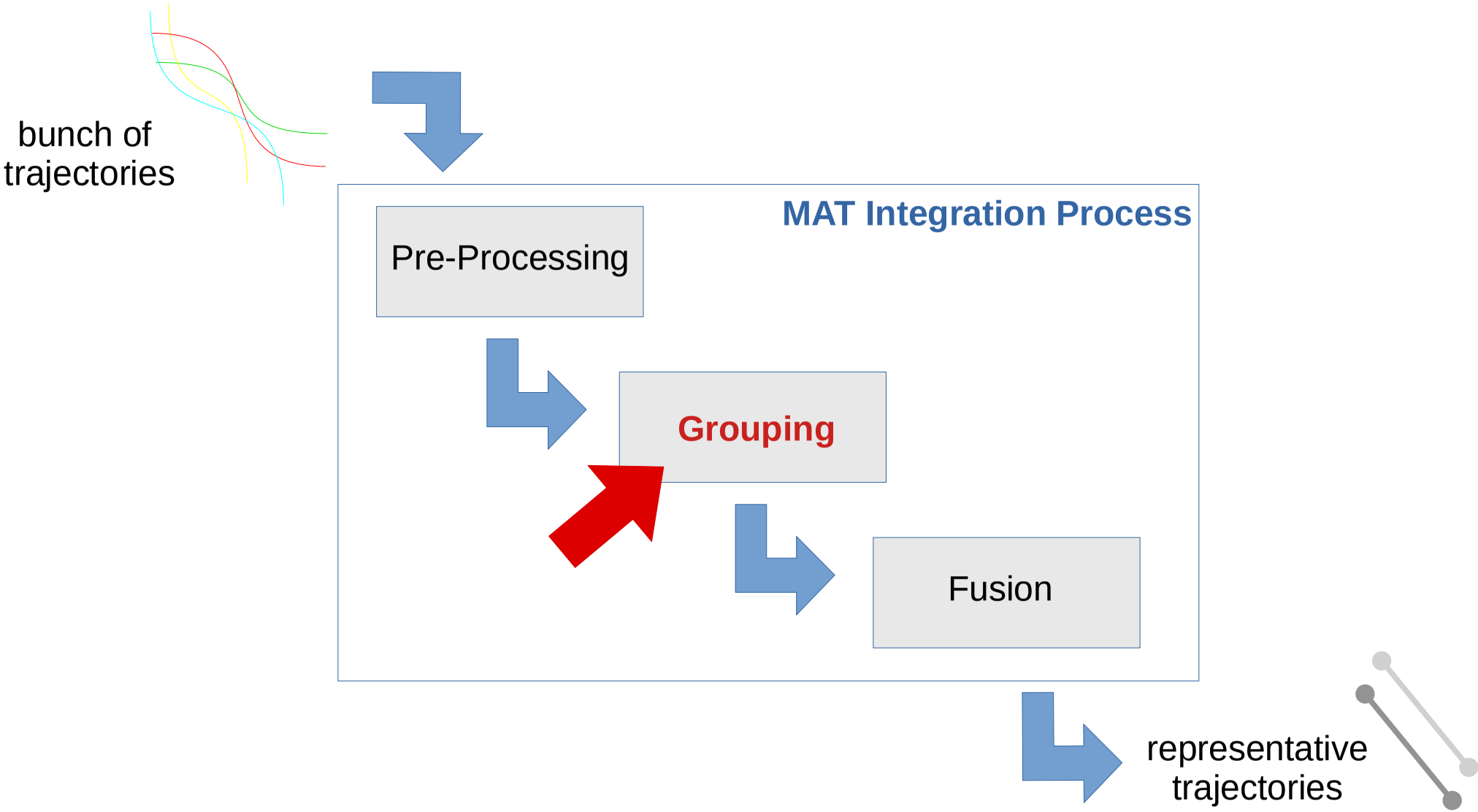


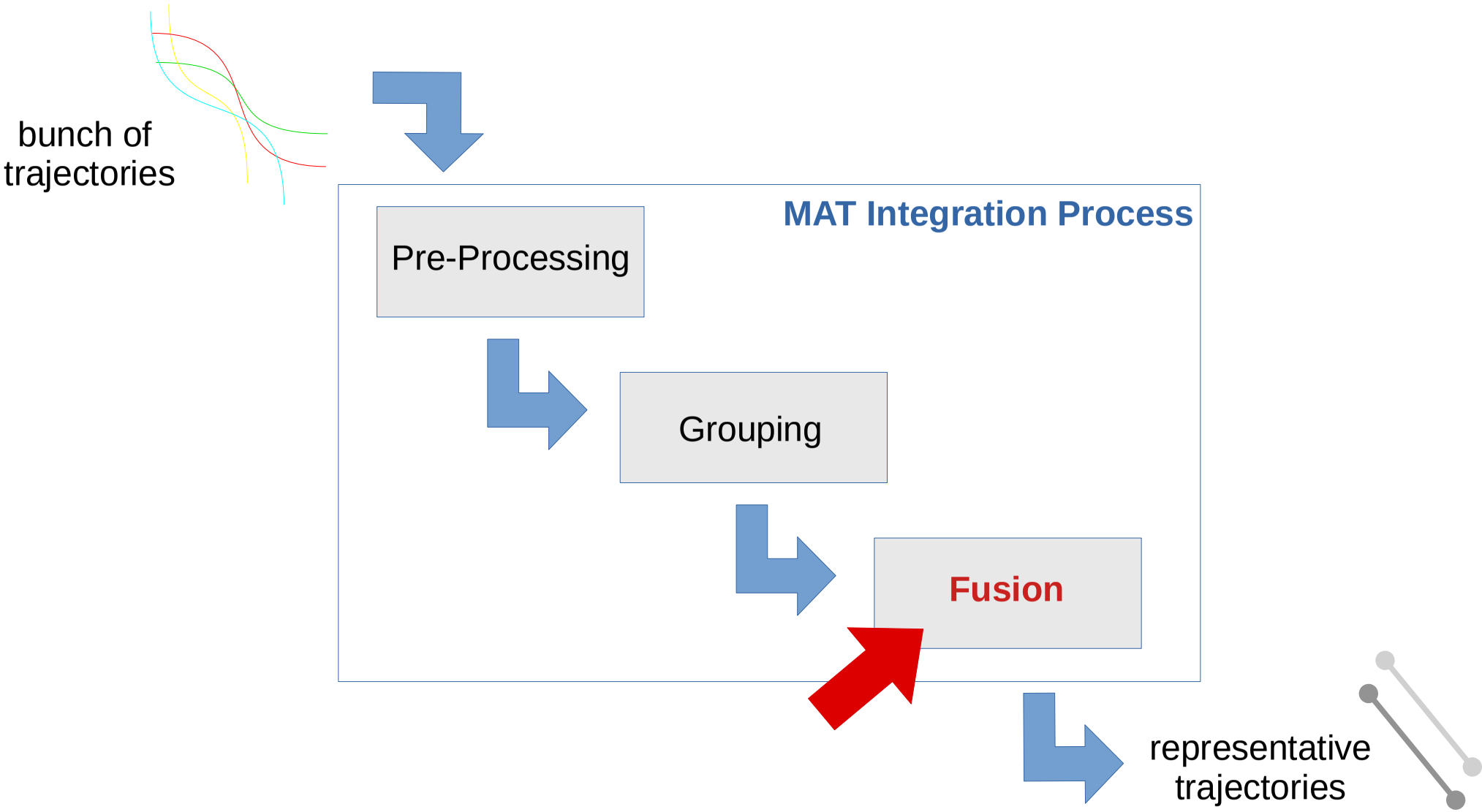
# MAT Integration – A Process



# Grouping Phase

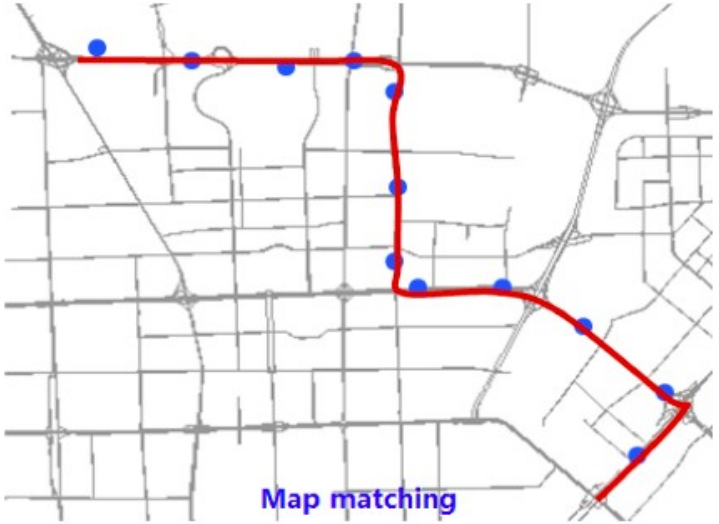
- Goal
  - To choose (and constraint) the relevant dimensions (space, time, aspects) to be considered for MAT integration
- Several possibilities
  - Focus on time
    - bike trajectories in the evening
  - Focus on space
    - bike trajectories from downtown to UFSC
    - bike trajectories in the Florianopolis continental area
  - Focus on aspects
    - bike trajectories during rainy days
  - Focus on space + aspects
    - bike trajectories during rainy days at the Florianopolis continental area
  - …
- Trajectory similarity approaches may be considered/adapted

# MAT Integration – A Process



# Fusion Phase

- Goal
  - To generate a representative MAT for each MAT grouping
- Challenges
  - To solve MAT representation conflicts
    - different number of points
    - different number of aspects related to the trajectories, points or moving objects
    - conflicting aspects (*rainy vs sunny*) in the same group
    - ...
  - To solve constraints related to space and time that must be considered for generating the representative MAT
    - example: street map matching





Big Data Analytics:
lançando luz dos genes ao cosmos

# Multiple Aspect Trajectories Modeling and Integration

**Ronaldo dos Santos Mello**

UFSC/INE/PPGCC/GBD
*r.mello@ufsc.br*