

# Modularity-Based Clustering for Network-Constrained Trajectories

Mohamed K. El Mahrsi, Télécom ParisTech – Département Informatique et Réseaux  
 Fabrice Rossi, Équipe SAMM EA 4543, Université Paris 1 Panthéon-Sorbonne

## Context and Motivations

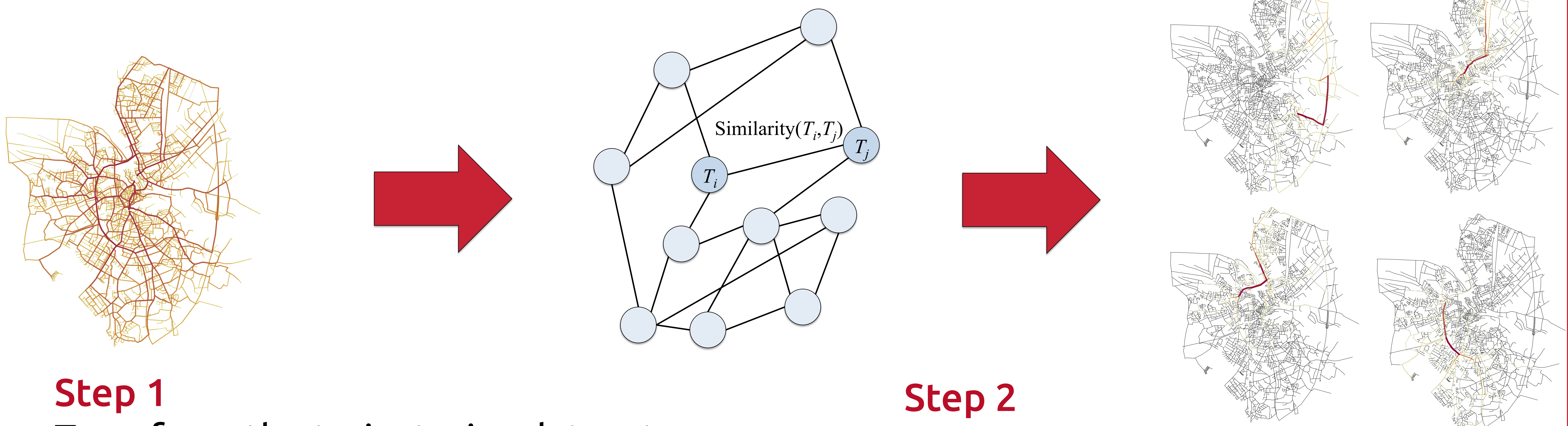
- Recent explosion in location-aware devices (GPS, smartphones, etc.).
- Eventually, trajectories of vehicles travelling along the road network can be collected at a large scale.
- An opportunity to have a better understanding of the traffic flow and dynamics.
- Useful knowledge for many applications (traffic monitoring, road planning, carpooling, etc.).

## Problem and Data Model

- Our aim is to regroup trajectories into clusters that moved along the same part of the road network.
- The road network is an oriented graph  $G=(V, E)$  where:
  - Nodes are road intersections
  - Edges are road segments
- A trajectory  $T$  is a sequence of travelled segments (eventually timestamped with the date of entry on the segment)

$$T = \{(t_1, e_1), \dots, (t_i, e_i), \dots, (t_n, e_n)\}$$

## Approach



### Step 1

Transform the trajectories dataset  $\mathcal{T}$  into a similarity graph

- Use a cosine similarity with spatial weighting
- Weights indicate relevance of road segments

$$\text{Similarity}(T_i, T_j) = \frac{\sum_{e \in E} \omega_{e, T_i} \cdot \omega_{e, T_j}}{\sqrt{\sum_{e \in E} \omega_{e, T_i}^2} \cdot \sqrt{\sum_{e \in E} \omega_{e, T_j}^2}}$$

$$\omega_{e, T} = \frac{\text{length}(e)}{\sum_{e' \in T} \text{length}(e')} \cdot \log \frac{|\mathcal{T}|}{|\{T' : e \in T'\}|}$$

### Step 2

Apply modularity optimization clustering to discover a hierarchy of meaningful clusters

$$Q = \frac{1}{2m} \sum_{k=1}^K \sum_{i, j \in C_k} \left( \omega_{ij} - \frac{d_i d_j}{2m} \right)$$

## Conclusions

- Tested on synthetic datasets.
- Parameter-free, unlike existing approaches which are very sensitive to threshold values.
- Multi-level hierarchy of nested clusters, suitable for exploring large trajectory datasets on various levels of detail.

Presented at the European Symposium in Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012)

Contact: khalil.mahrsi@telecom.paristech.fr