Data-driven Mobility Analytics

Danyang SUN, Ph.D. Candidate, ENPC
Lab recherche environnement VINCI | PARISTECH (ENPC)
Nov 30th, 2021

“ Good information helps make right decisions ”

Massive digital traces → Enhance mobility diagnosis

Mobility challenges

- Traffic congestion
- Energy concern
- Emissions and pollutions

Need insights for solutions

- Understand the needs
- Diagnose the problems
- Capture the trends

Massive Data

- Wide-coverage
- Up-to-date
- Rich info for obtainning knowledge of human movements

Modern Mobility Data

What is modern? (vs traditional surveys)

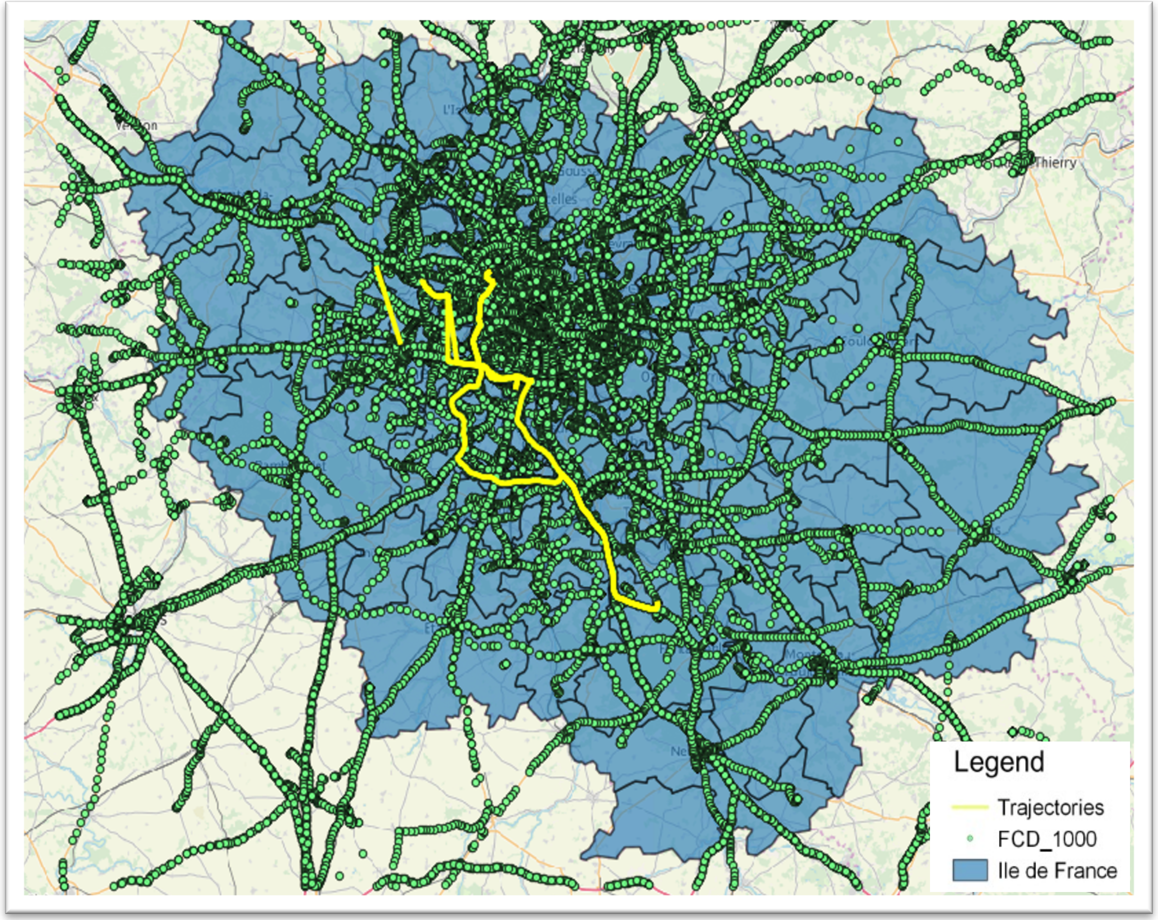
- By sensing technologies
- Mobility localization Info generated automatically

Digital trajectory data (as typical forms)

- GPS receivers (mobile phones, cars, etc.)
- Public transit records (ticket validations)
- Geo-tagged tweets etc.

Floating Car Data

- GPS traces of moving vehicles
- Per record per 30s~60s (sourced from Coyote)
- 3~5% penetration rate (160,000 observed vehicles in Île-de-France)



FCD examples over Île-de-France

lab recherche environnement
VINCI | PARISTECH

Research Topic

A Ph.D. thesis (2019.01-2021.12 @LVMT, ENPC, supervised by Fabien LEURENT and Xiaoyan XIE)

Objective

- Mobility pattern diagnosis by Trajectory data mining

Issues

Two levels of issues

Individual-Centered Mobility Patterns

- Vehicle usage type (trips)
  - User differentiation
- **Mobility regularities (stays)**
  - Anchor places (e.g. homes and work places)
- Travel time estimation (prediction)

Place-Based Mobility Patterns

- Functional zones (intra-)
  - Land occupations (e.g commercial vs residential)
- **Spatial organizations (inter-)**
  - Employment cores and their catchment areas
- Spatial interactions (inter-)
  - Origin – Destination trip flow

Case study:

Explore home-to-work spatial relations of Île-de-France

- Commuting situations
- Centers and catchment areas

Case Study:
"Explore home-to-work spatial relations of the IDF region"

Analytical steps

Individual analysis

Phase 1:
Trajectory processing → detect home and workplaces

Geographical analysis

Phase 2:
Spatial distribution analysis → identify employment centers

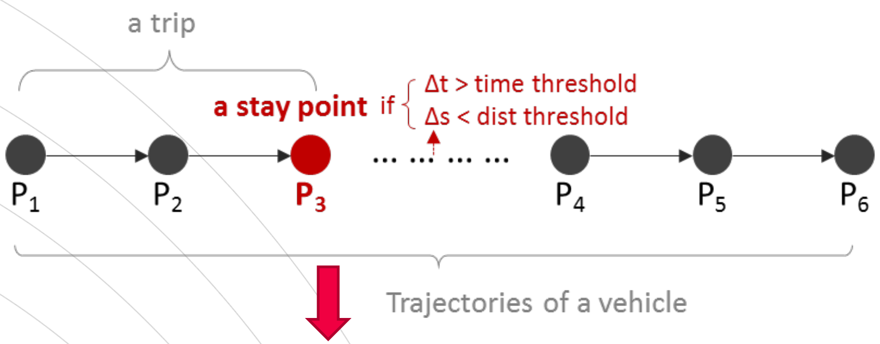
Phase 3:
Spatial relation analysis → discover bonded spatial communities

Phase 1: Home and workplace detection (anonymized)

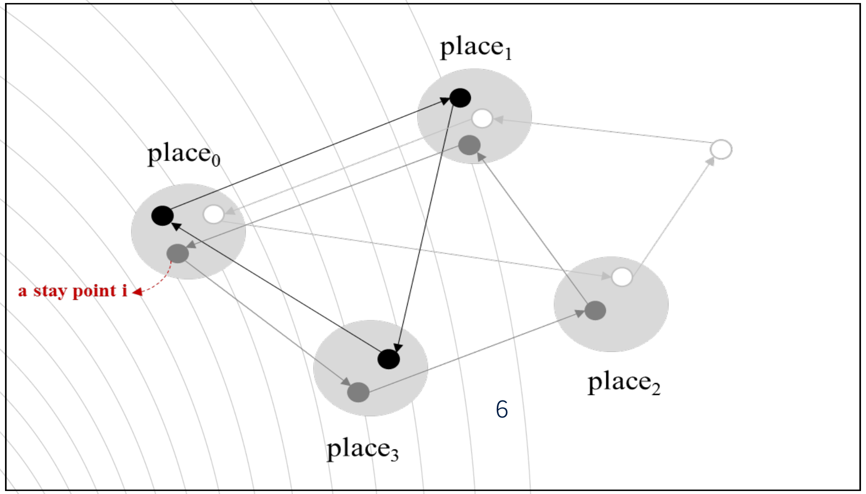
Aim: Recognize the home and workplaces

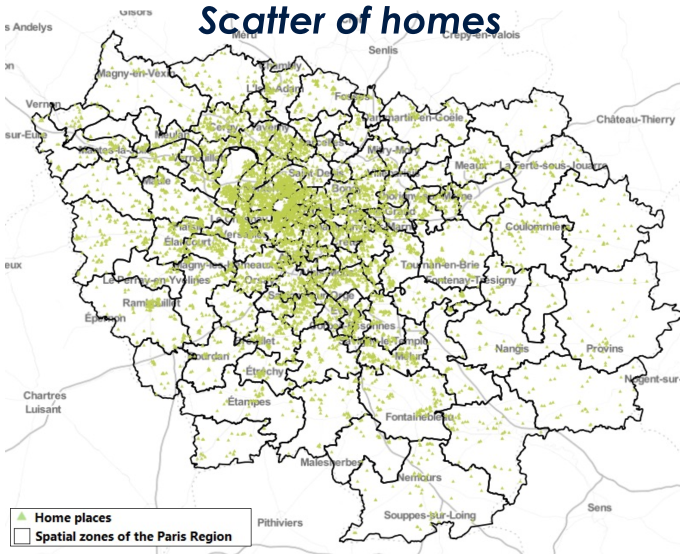
Approach: featuring mining + unsupervised learning -> Learn regularities

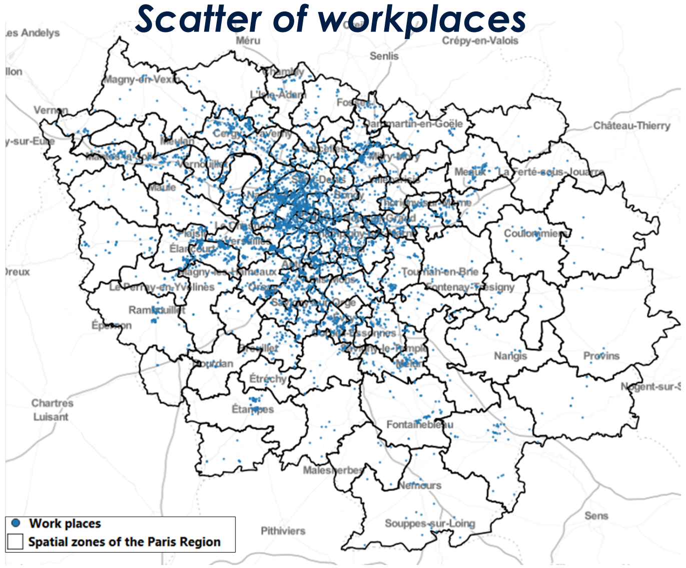
1) Trajectory → **stay points** (by sequencing algorithm)



2) Stay points → **functional places** (by DBSCAN)







3) Functional places → **homes, workplaces** (by Gaussian Mixture Model + Kmeans)

According to :

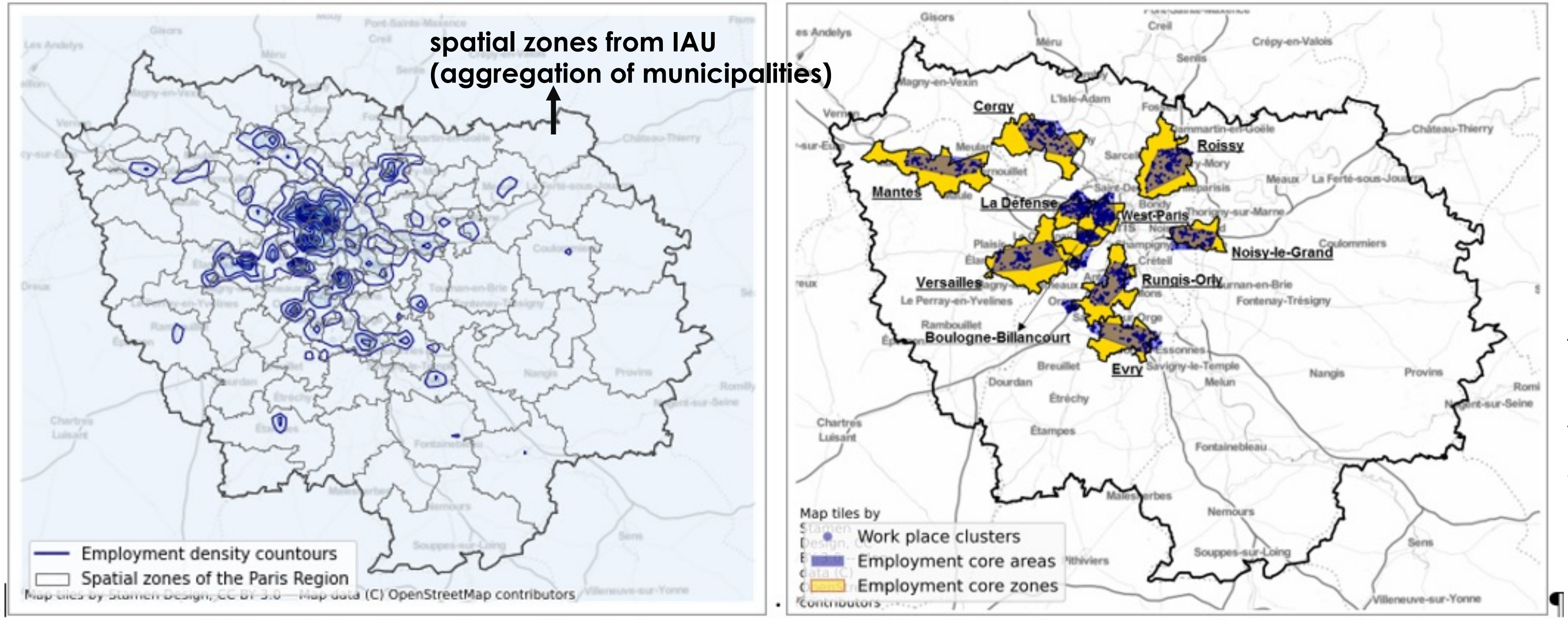
- Visit duration and frequentation
- E.g.: homes: more likely with patterns of visits at the end of day, on almost every day

Phase 2: Spatial distribution analysis

Aim: Identify employment core areas

Approach: density modeling + spatial clustering → delineate core areas based on workplaces

1) Kernel Density Estimation-> density contours **b) HDBSCAN (spatial clustering) -> core clusters -> core zones**



(a) Spatial density contours of work places (b) Employment core areas and core zones

Results:

10 employment core zones

- La-Défense
- Boulogne-Billancourt
- Versailles
- Noisy-le-grand
- Etc.

Account for 43% of total jobs

Notes:

Various-density clustering

- Finer-grained in dense areas and coarser in sparse areas

Results based on auto-mobilists only

lab recherche environnement
VINCI | PARISTECH

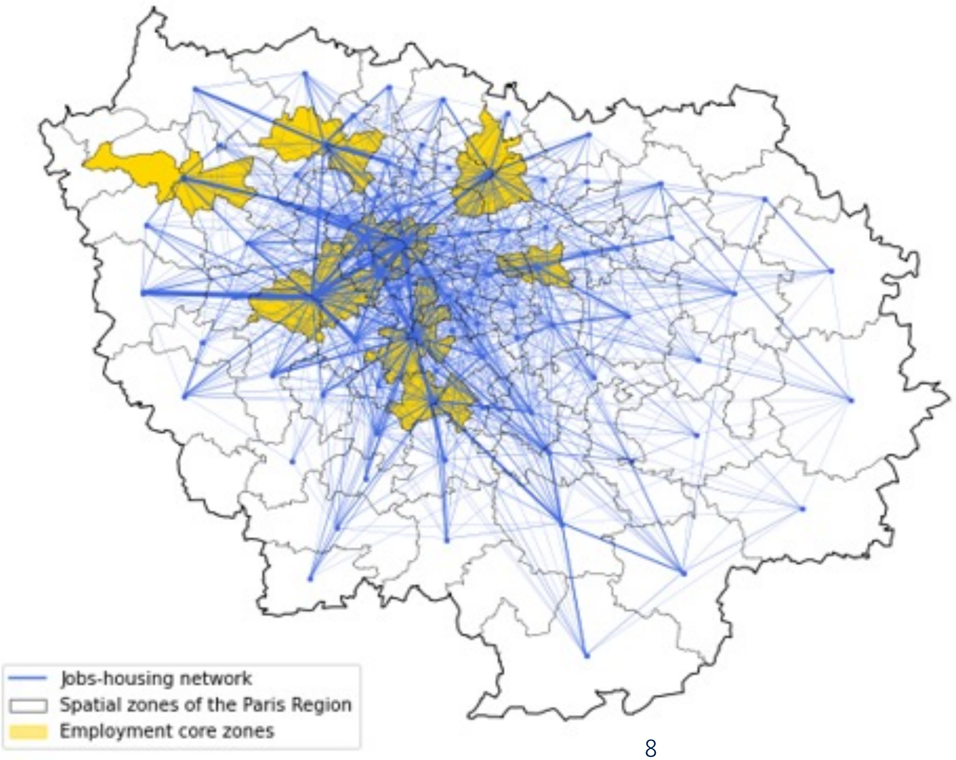
Phase 3: Spatial relation analysis (1)

Aim: Identify spatial communities

Approach: home-to-work flow + graph partition → detect bonded areas

1) Build jobs-housing graph network

- Based on home-to-work flow



2) Graph partition → find communities

- Method: find sub-networks with denser intra-connections and sparser inter-connections
- Output: communities, which are then used to interpret bonded territorial regions

e.g. Community detection by finding inner-dense sub-networks

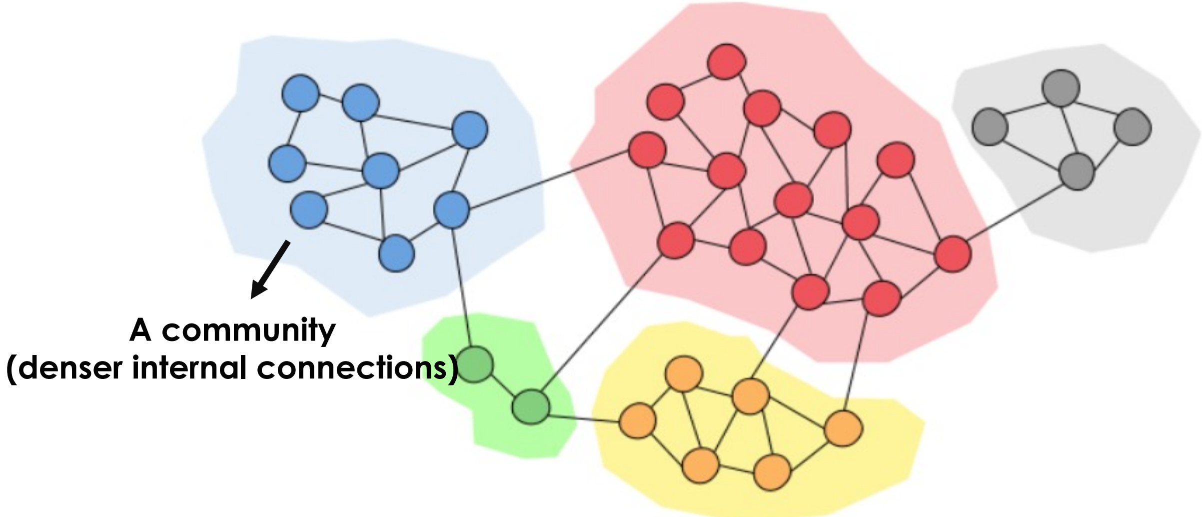
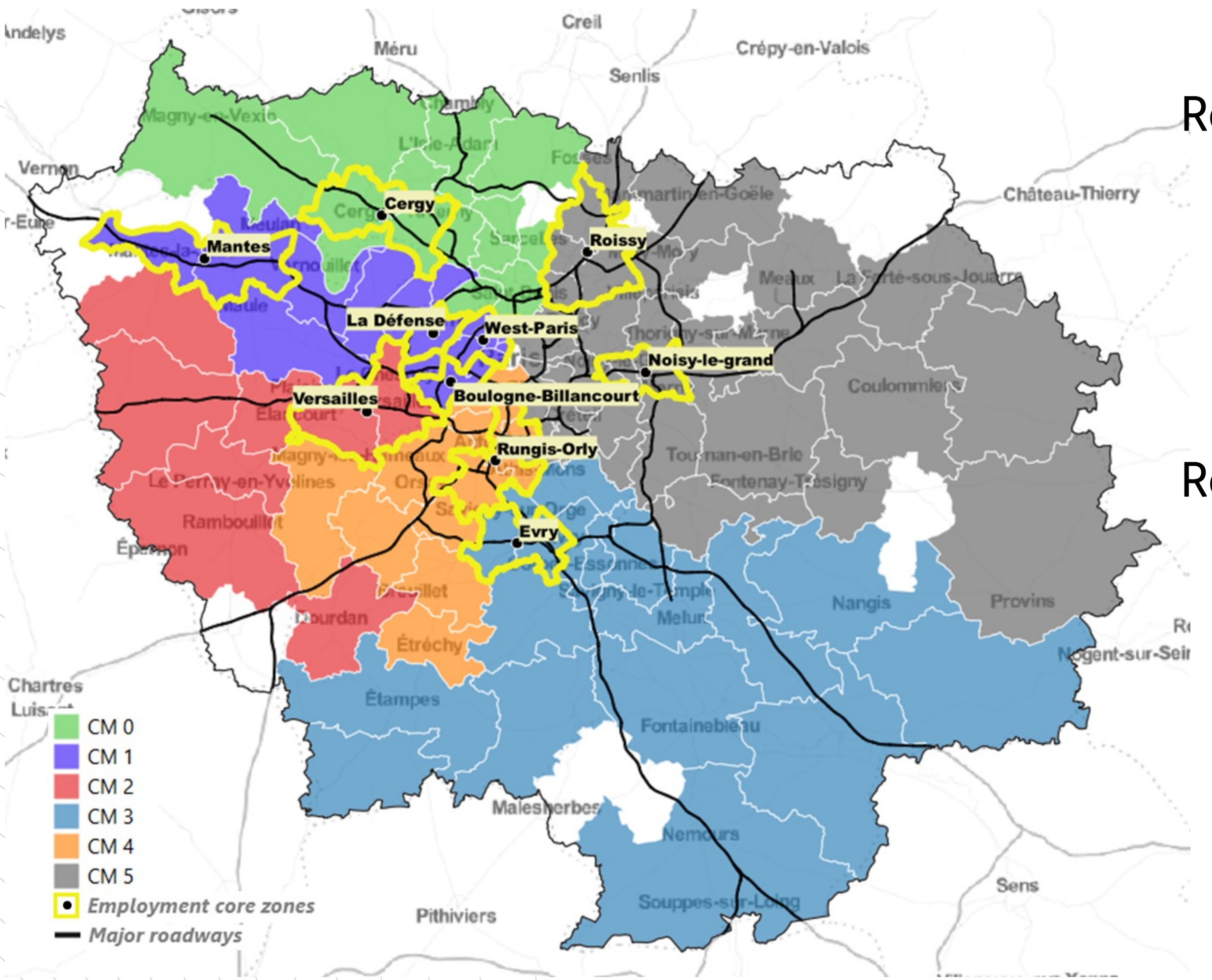


Figure source: by Thamindu Dilshan Jayawickrama @ towardsdatascience.com

Phase 3: Spatial relation analysis (2)

Results of detected spatial communities and their sub-centers



Results:

- 6 spatial communities
- Consistent with the layout of major highways

  (since results based on auto-mobilists from FCD)

Remarks

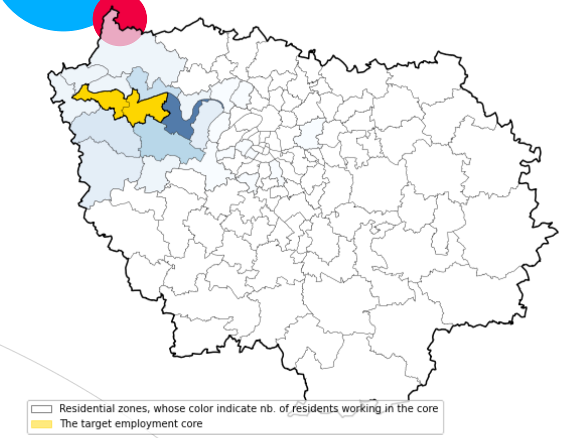
- Communities are mainly formed by adjacent zones

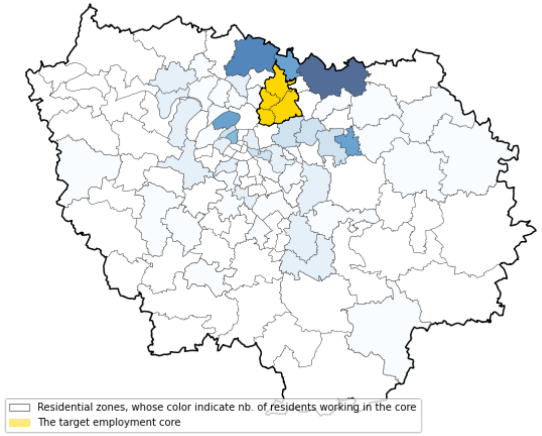
  → show the nature of spatial cohesion
- Employment cores are embedded in the communities

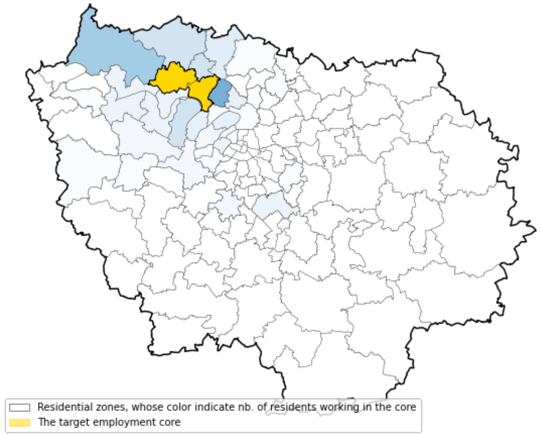
  → interpret the “catchment” property

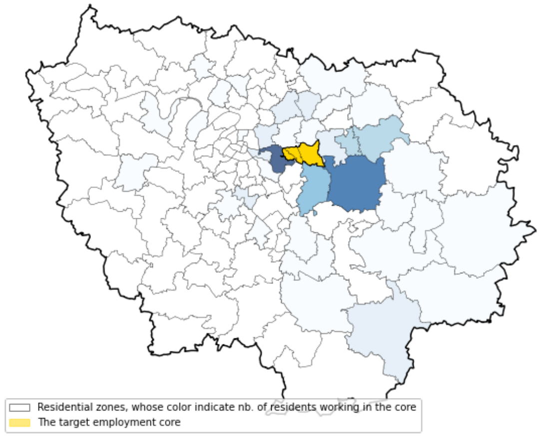
lab recherche environnement VINCI PARISTECH

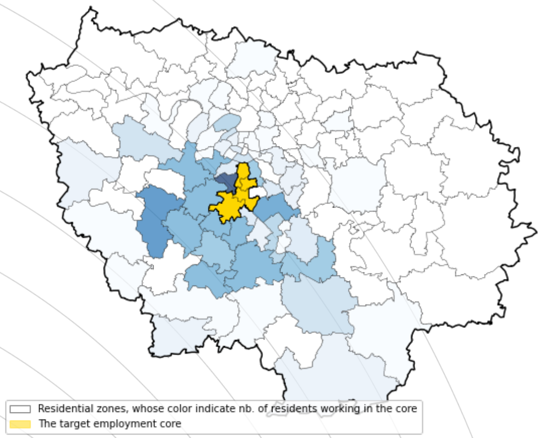
Core-periphery patterns: the "radiation" of cores

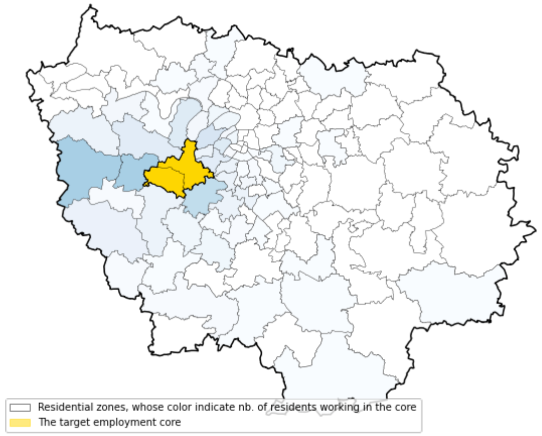


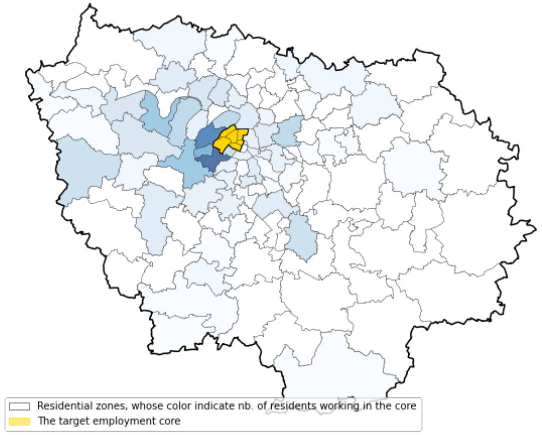


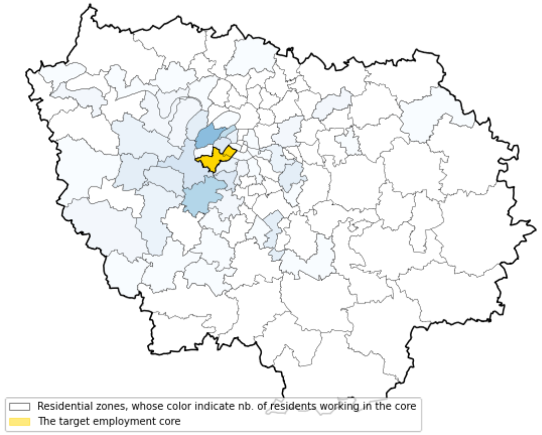




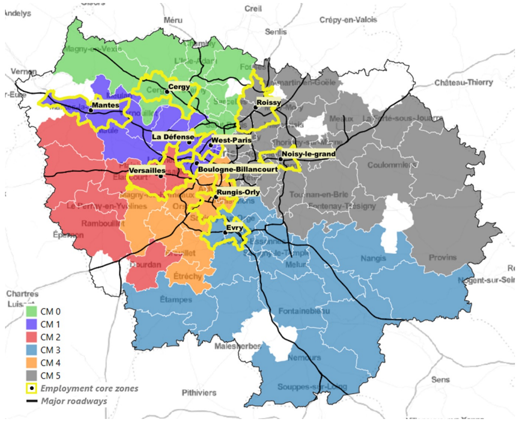


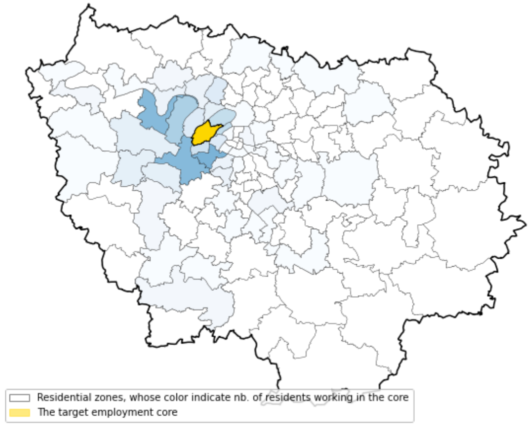


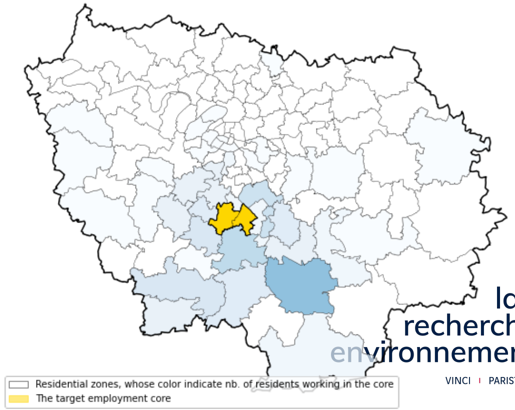




- The "blueness" indicates the number residents working in the core
- Confirm that the communities were not randomly aggregated







lab recherche environnement
VINCI PARISTECH

Contribution

- Data-science driven method for automated mobility analytics
- Insights for facilitating network planning and land-use evaluation

Future expansion

- Currently only applied on car mode
- Replicable and scalable to multimodal analysis
  - by using a database involving multimodal trips

    (e.g. mobile phone data, fused datasets etc.)

lab
recherche
environnement
VINCI | PARISTECH

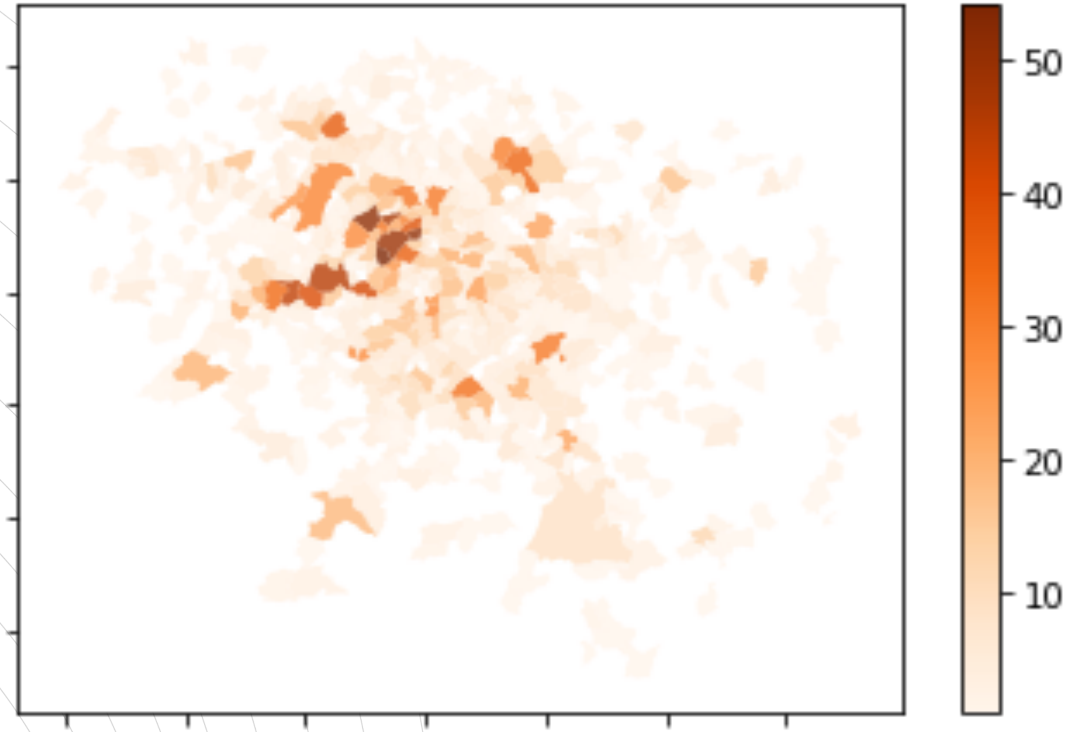
Des questions ?

Danyang SUN
danyang.sun@enpc.fr
lab-recherche-environnement.org

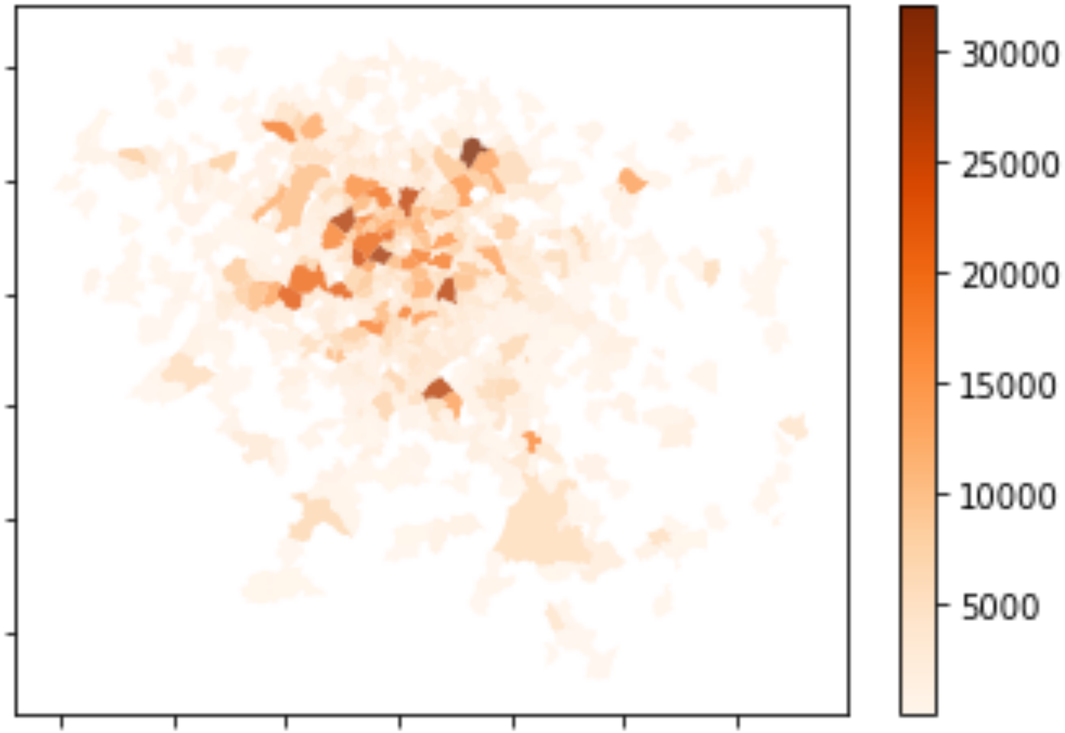
Appendix (1)

Result evaluation

- Comparing with census data (Insee, 2017)
  - ➢ Consistent → the dense areas of jobs in space can be well discovered based on FCD



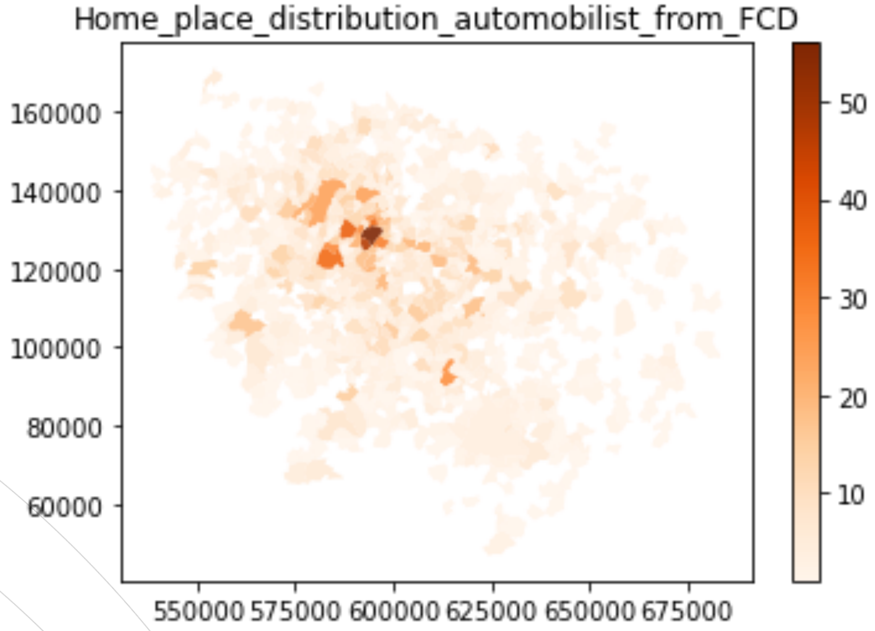
Job numbers by municipalities from FCD (2019)



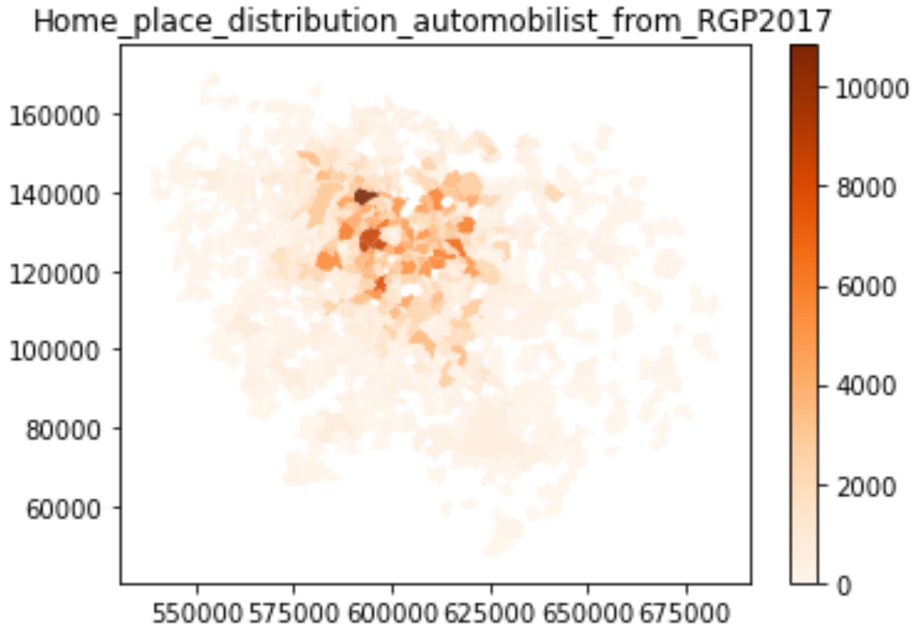
Job numbers by municipalities from census data (2017)

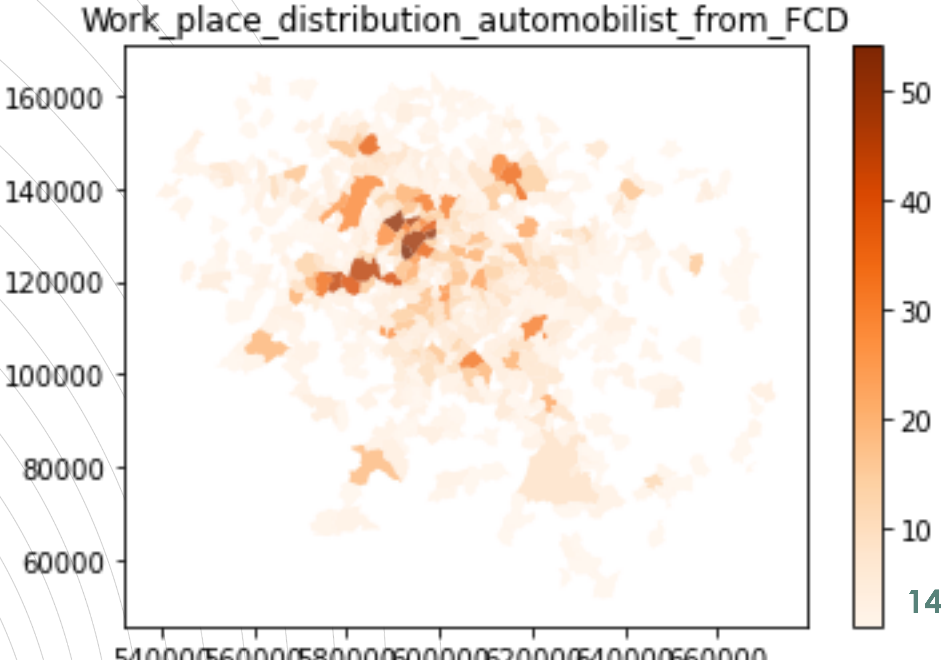
lab recherche environnement VINCI PARISTECH

Appendix (2) - Data representativeness

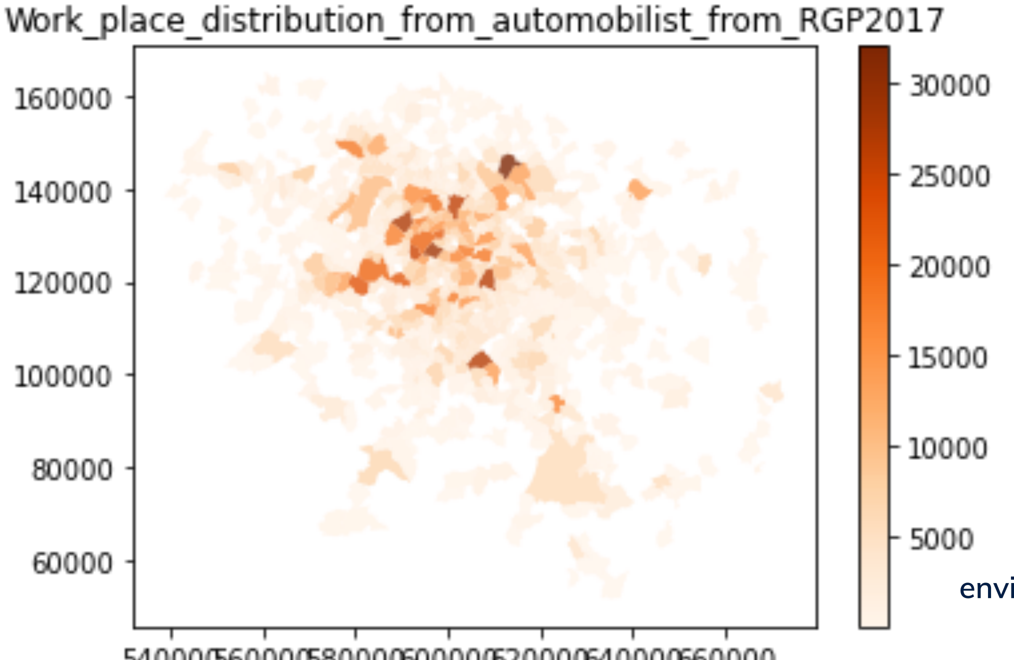


VS





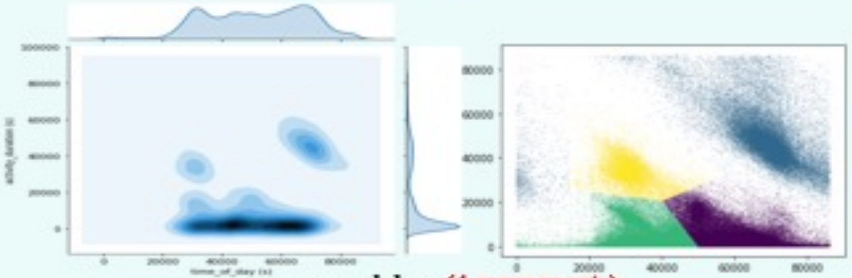
VS



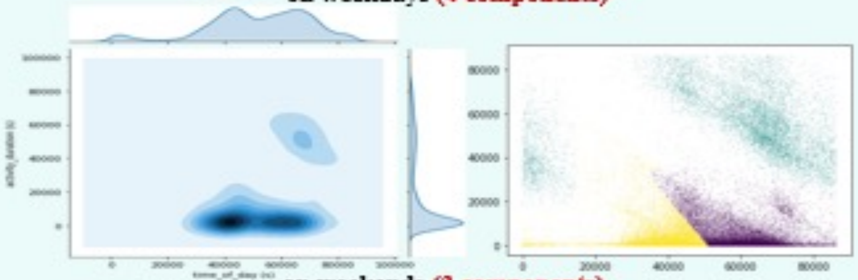
Appendix (3) - Two-level place type identification

Activity Type Identification

- **Features**: 1) time of day and 2) activity duration
- **Approach**: Gaussian Mixture Modeling
- **Results**:
  - Weekdays: 4 activity types
  - Weekends: 3 activity types



on weekdays (4 components)



on weekends (3 components)

Evaluation

Compare with Rule-Based Identification

- Rule-based identification are widely adopted on identifying home/work places
- A common lack of ground truth for such data
- Showed a highly consistent matching

Significant-Place Type Identification

- **Features**: Activity profiles of each place (conditional relative frequencies over each activity type)
- **Approach**: K-means clustering
- **Results**: 7 clusters -> 5 types by characterization
  - Home (c0 & c5): *significant late-day long activities*
  - Workplaces (c2 & c6): *significant early-day activities*
  - Secondary places etc.

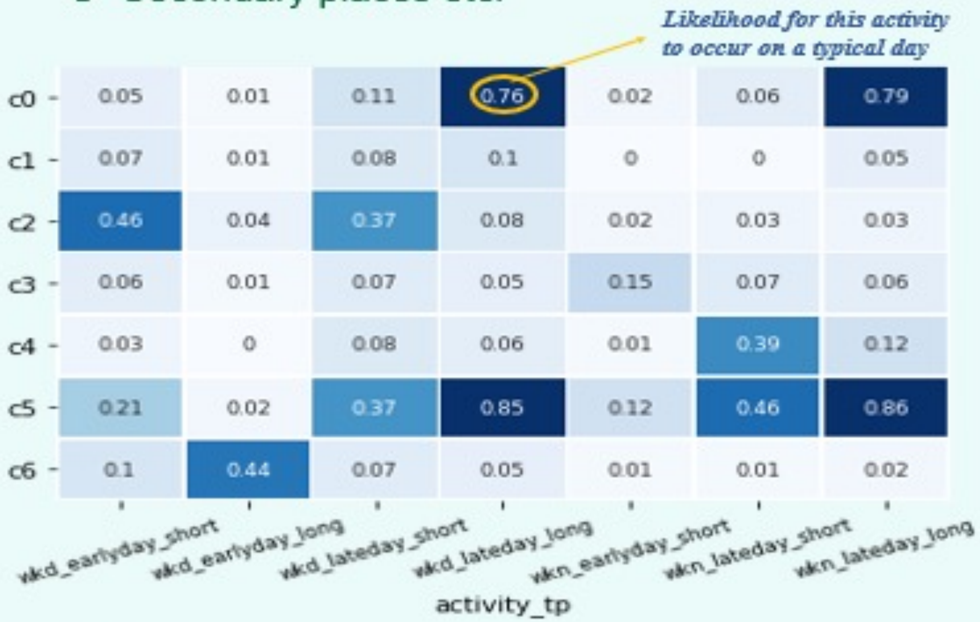


Table. Results of Evaluation

By our method	By rule-based identification (considered as base reference) Home	Work	Secondary	Consistency ratio
Home	5601	107	116	96.2%
Work	26	1254	108	90.3%
Secondary	188	2063	18282	89.0%