National Research Council of Italy

Institute of Electronics Computer and Telecommunication Engineering

NoF 2017
London, UK
November 22, 2017

Hidden structures in mobile network traffic

Marco Fiore

http://perso.citi.insa-lyon.fr/mfiore/
marco.fiore@ieiit.cnr.it
@marc0_fi0re

1
Context
Mobile networking meets data analytics

Context

- **High expectations for 5G (& beyond) networks**
  - Accommodate ***7-fold mobile traffic growth*** by 2021 [1]
  - ***Orders-of-magnitude performance upgrade*** over LTE [2]
    - **1,000 times** capacity per unit area, **100 times** connected devices

A. increasing capacity

new spectrum, waveforms, MIMO, multi-RAT, denser deployment, D2D

B. better managing capacity

dynamic resource (re)configuration paradigms: CR, C-RAN, MEC, SDN, NFV, network slicing

- **5G will feature [2] cognitive network management [3]**

"*A **cognitive network** has a cognitive process that can perceive current network conditions, and then plan, decide and act on those conditions. The network can learn from these adaptations and use them to make future decisions*"

self-organizing networking ***anticipatory networking*** ***machine learning*** ***data analytics***

talk focus

[1] Cisco VNI Forecast, Global Mobile Data Traffic Forecast Update 2016–2021
[2] EC H2020 5G Infrastructure PPP. Pre-structuring Model, version 2.0. 2014
[3] R.W. Thomas, L.A. DaSilva, A.B. MacKenzie. IEEE DySPAN 2005

Context

- **Classification of mobile traffic demands**

  “*identifying network-wide profiles of mobile traffic*”

- **Two orthogonal perspectives**
  - mobile traffic is a ***spatiotemporal*** phenomenon

1. *spatial classification*

*at which **locations** does mobile traffic follow comparable time dynamics?*

talk topic

2. *temporal classification*

*during which **time periods** does mobile traffic show similar geographical distributions?*

unveil ***locality*** of traffic fluctuations

expose ***long-timescale dynamics***

  - ***required inputs*** for cognitive networking, via network-wide resource orchestration in, e.g., C-RAN, MEC [4,5]

[4] H. Assem, T. Sandra Buda, L. Xu, H2020 5G-PPP CogNet, Deliverable D2.1, 2015
[5] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, W. Xiang, IEEE Network, 30(1), 2016

2

Spatial classification

A geography of mobile network demands

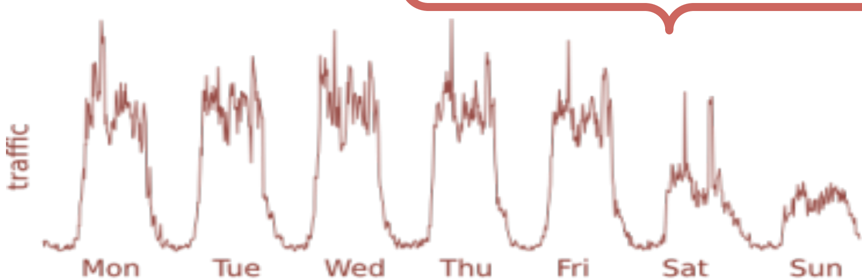
Methodology

*At which **locations** (in a target geographical area) does mobile traffic follow similar dynamics? How do such dynamics look like? And what induces them?*

1. Mobile traffic **signature** at one location

- metric → ***aggregate*** *mobile traffic volume*
- temporal support → *one **week*** [6,7]
- (filtering) → *none*
- normalization → ***standard score***

median *hourly volume*



2. Pairwise **signature distance** measure

$$p_{ab} = \frac{\sum\limits_{t\in\mathcal{T}} (\mathbf{s_a}(t) - \mu_{\hat{a}}) \cdot (\mathbf{s_b}(t) - \mu_{\hat{b}})}{\sqrt{\sum\limits_{t\in\mathcal{T}} (\mathbf{s_a}(t) - \mu_{\hat{a}})^2} \cdot \sqrt{\sum\limits_{t\in\mathcal{T}} (\mathbf{s_b}(t) - \mu_{\hat{b}})^2}}$$

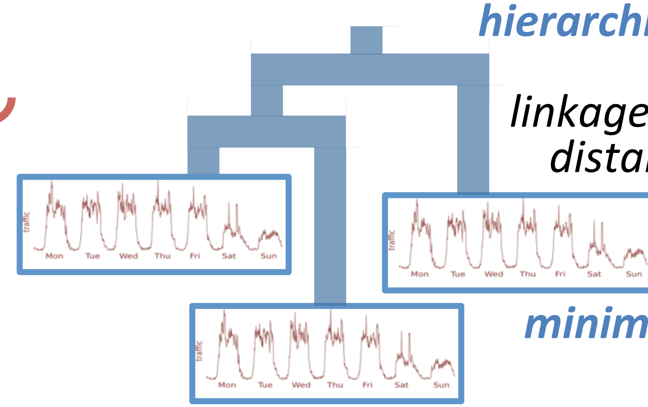
correlation-based *distance*

3. Signature **clustering algorithm**

agglomerative ***hierarchical*** *clustering*

linkage with average distance – ***UPGMA***

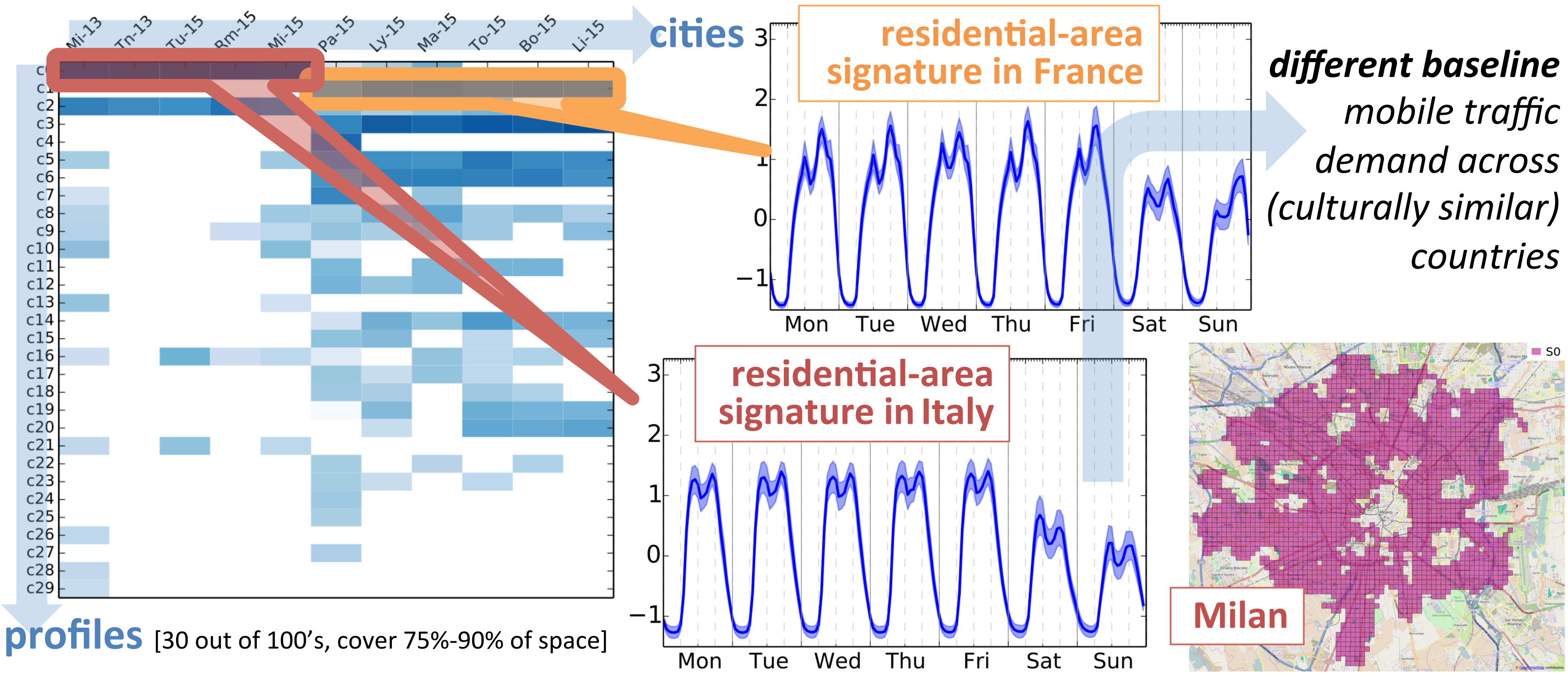
minimum skewness *stopping rule*



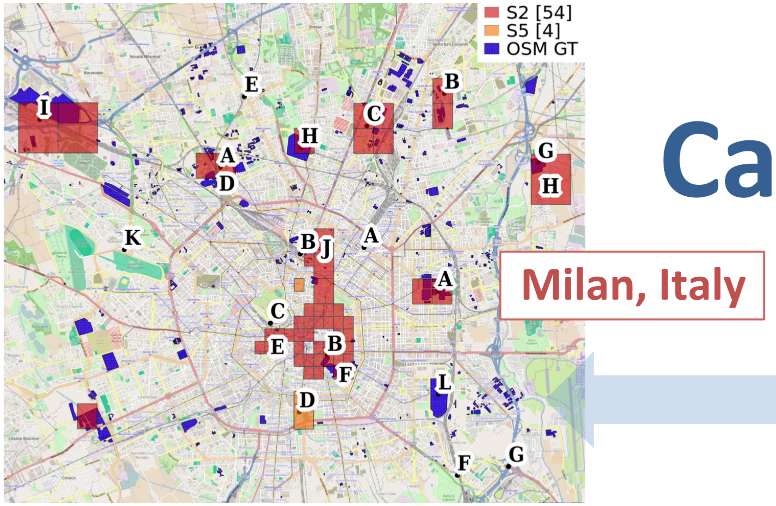
[6] R. Keralapura et al., ACM MobiCom 2010; [7] M.Z. Shafiq et al., ACM SIGMETRICS 2011

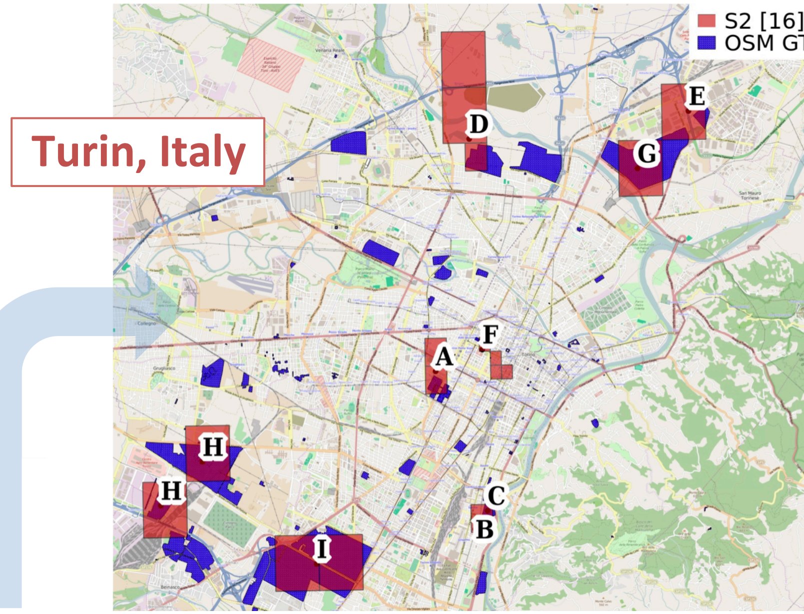
Case study

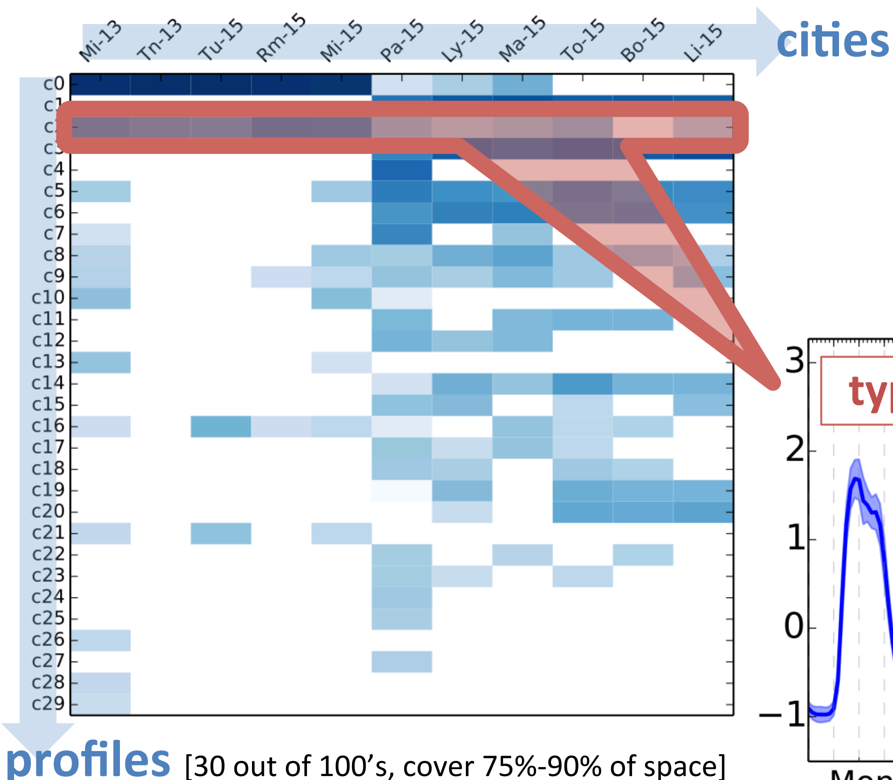
- Real-world mobile network traffic datasets
  - **Orange** 2014-15 [6 main cities in France, 4 months, antenna cells]
  - **TIM** BDC 2013-15 [4 main cities in Italy, 2 months, grid]



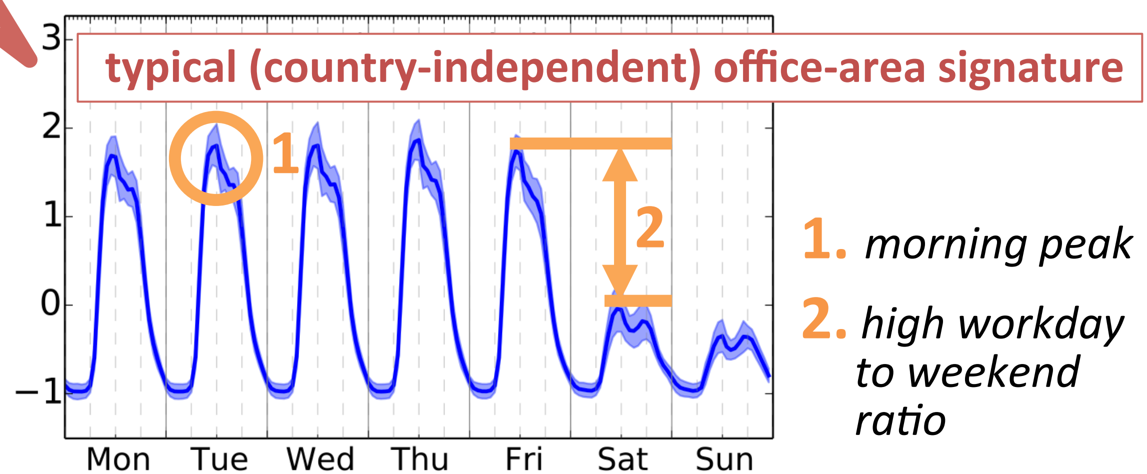
Case study





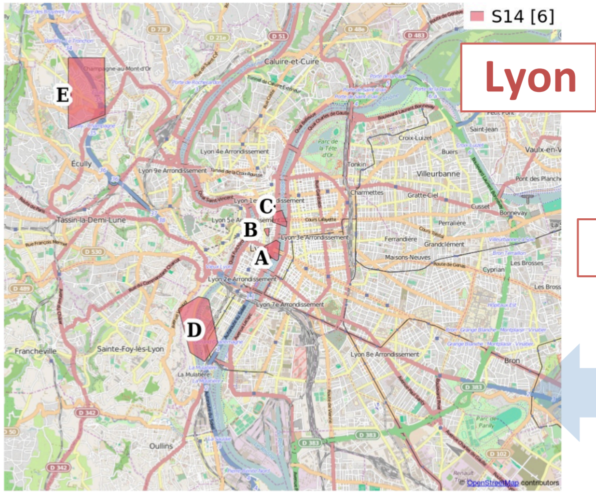


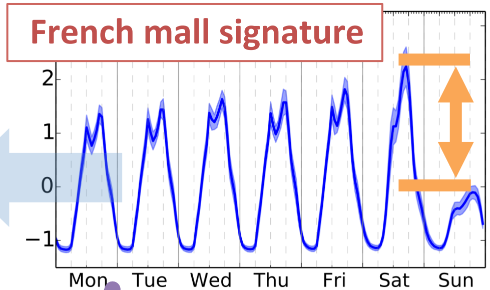
[30 out of 100's, cover 75%-90% of space]

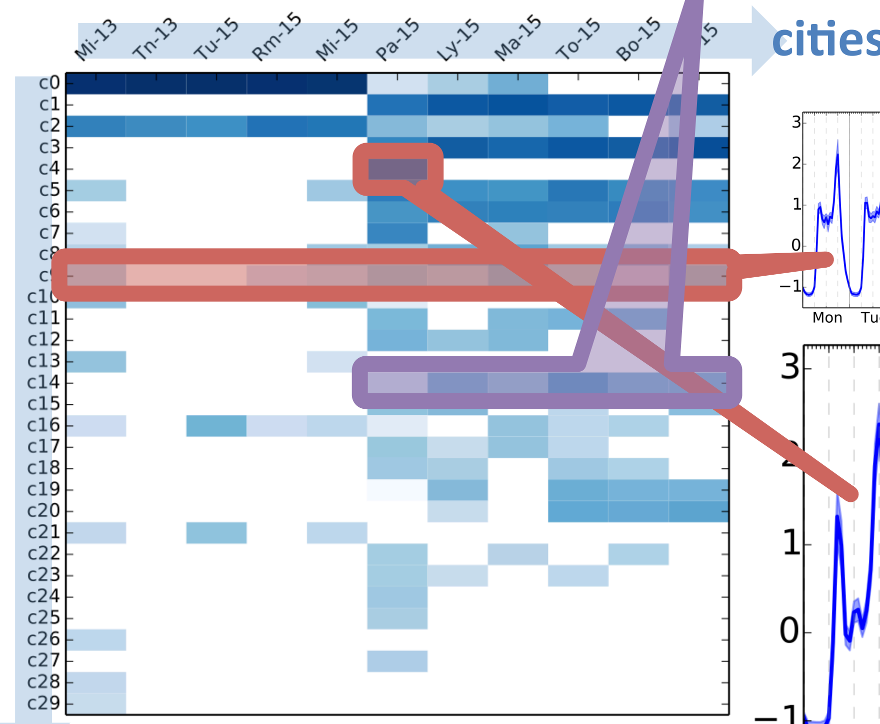


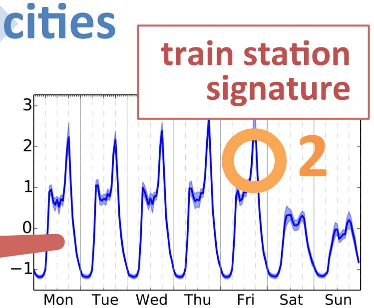
1. *morning peak*
2. *high workday to weekend ratio*

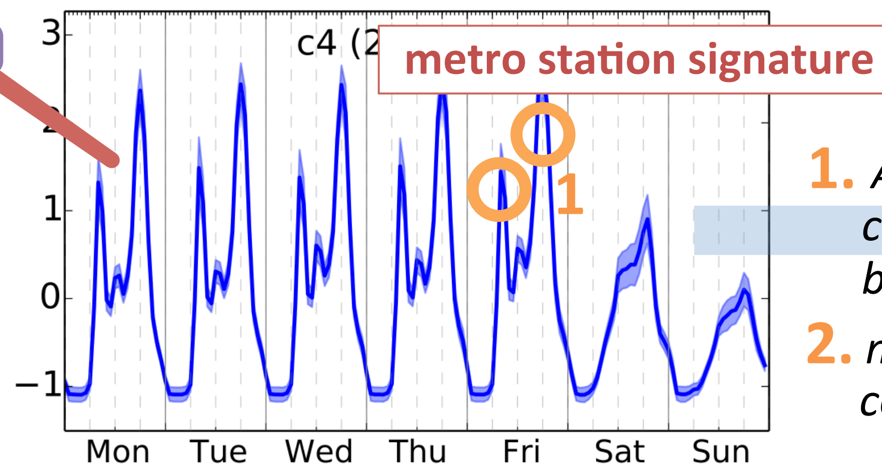
Case study

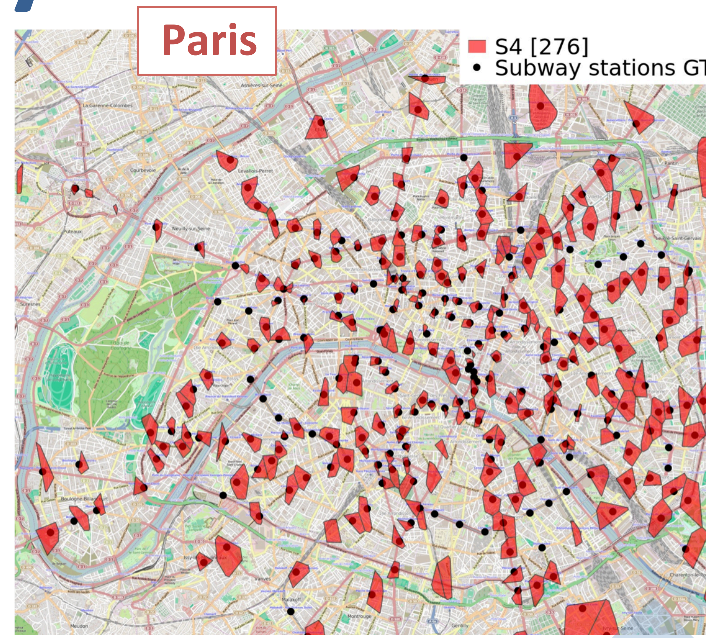












profiles [30 out of 100's, cover 75%-90% of space]

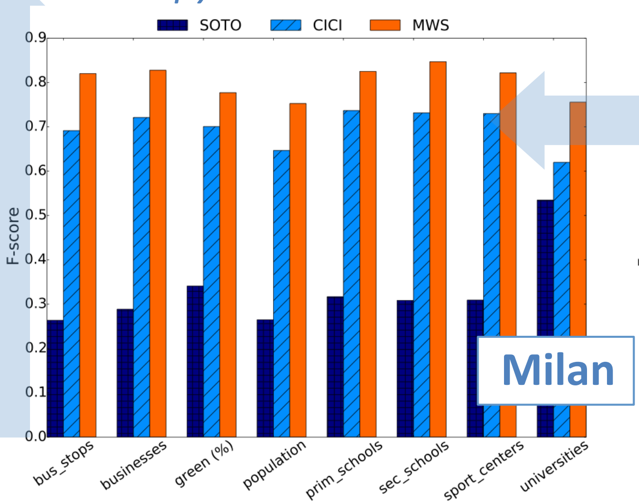
1. *AM-PM commuting behavior*
2. *mostly PM commuting*

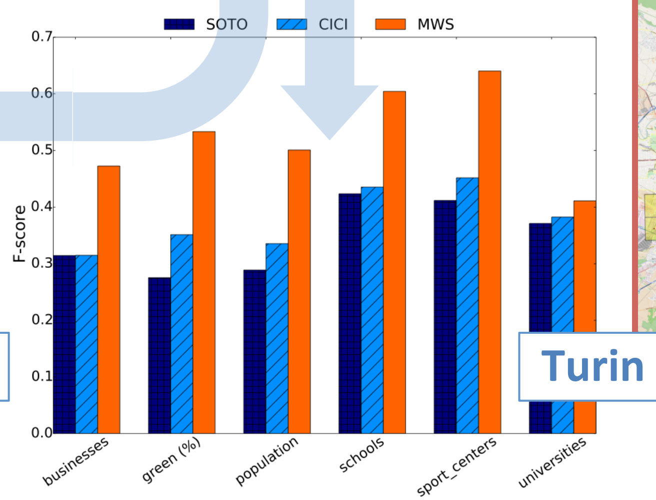
Land use detection

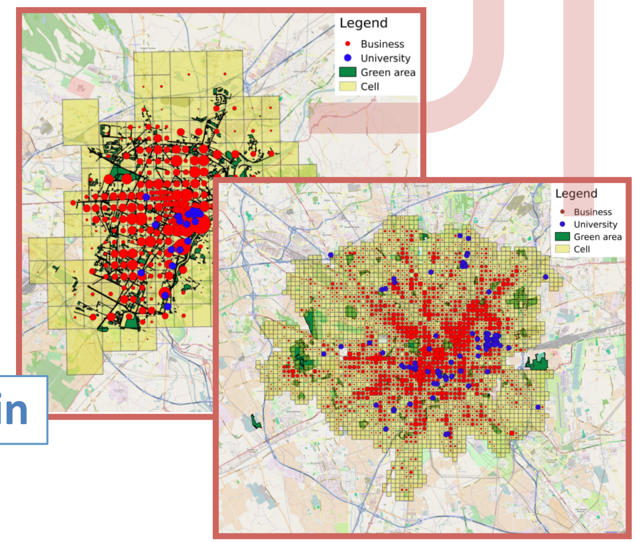
land use ***cartographies*** *provided by municipalities*

- Dual use of the methodology in **geoinformatics**
  - Complement traditional land use mapmaking
    - *census data, surveys, satellite imagery, points-of-interest*
  - Several strategies[8-11] ***evaluated*** against ***ground truth***

f-score coverage and entropy







[8] V. Soto et al., ACM HotPlanet 2011; [9] B. Cici et al., ACM MobiHoc 2015
[10] S. Grauwin et al., Geotechnologies and the Environment 2015
[11] A. Furno et al., IEEE Transactions on Mobile Computing 2017

3

An alternative approach

Spatial classification with EFA

Methodology

- **Exploratory Factor Analysis (EFA)**

inter-related by

Kx1 vector of ***common factors***, i.e., a small number (K<<N) of complex relationships among variables

Nx1 vector of ***variables***, i.e., phenomena of interest

$$\boxed{\mathbf{X}} = \boxed{\Lambda}\boxed{\mathbf{F}} + \boxed{\mathbf{U}}$$

Nx1 vector of unknown ***unique factors*** that are specific to a single variable

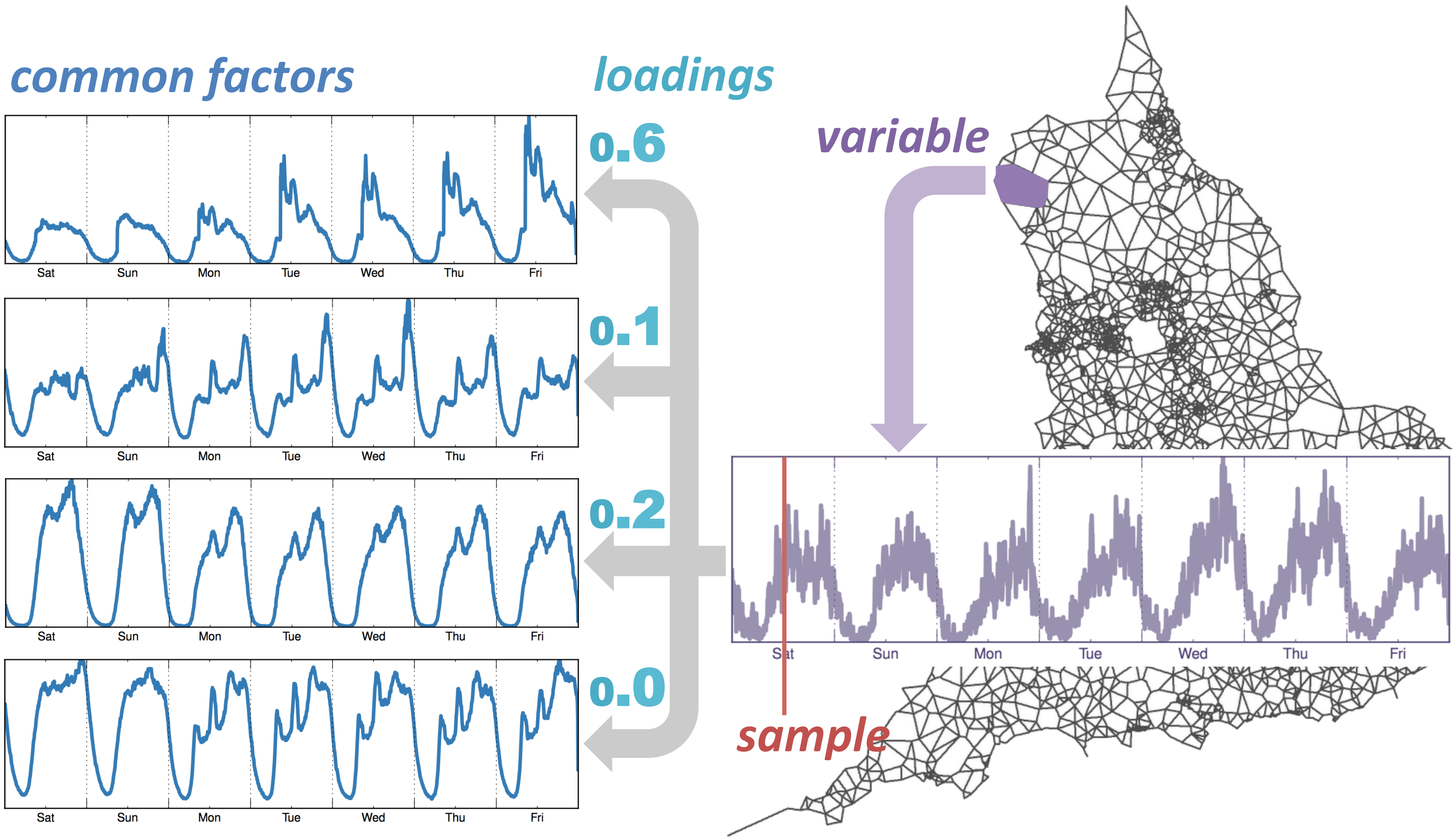
NxK matrix of unknown ***loadings*** that describe how important is each factor to every variable

interpreted through

- **EFA solution**
  - by analyzing variable observations from a set of samples, ***EFA identifies common/unique factors, and loadings*** [6]

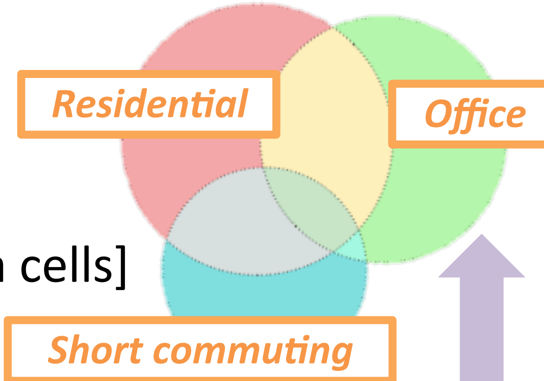
[6] S.A. Mulaik, Foundations of Factor Analysis, CRC Press, 2009

Methodology

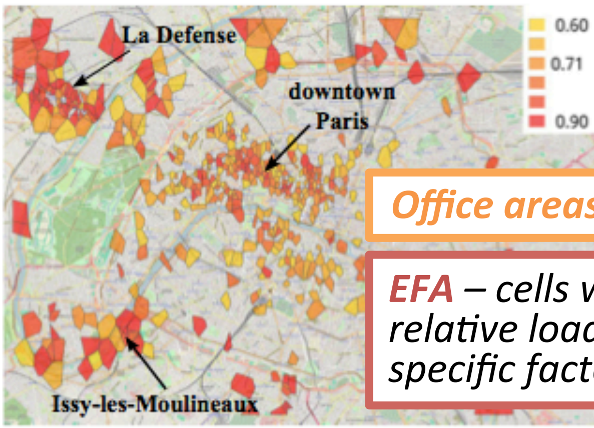


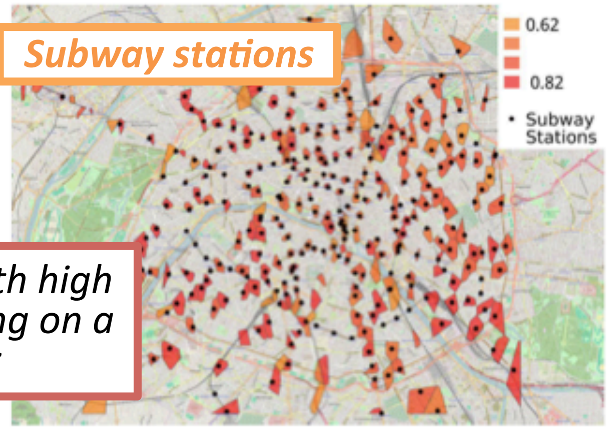
Case study

mixed land use detection

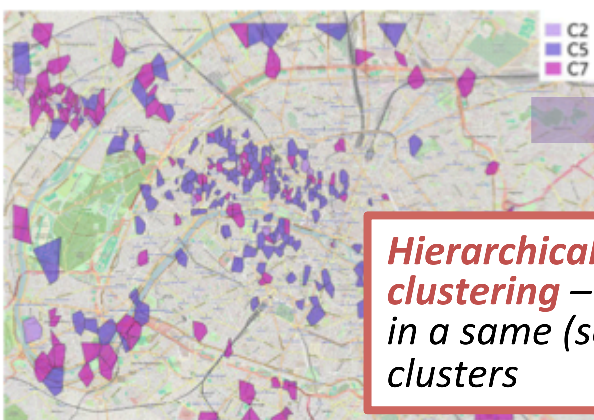


- **Orange** 2014-15 [Paris, France, 4 months, antenna cells]
  - *14 EFA factors* are identified





EFA – cells with high relative loading on a specific factor



Hierarchical clustering – cells in a same (set of) clusters

- 14 factors versus *hundreds of clusters*
  - multiple signature clusters just capture *different intensities of a same phenomenon*
  - many clusters are *unique factors*
  - traffic demands are in fact a *mixture* of actual common factors

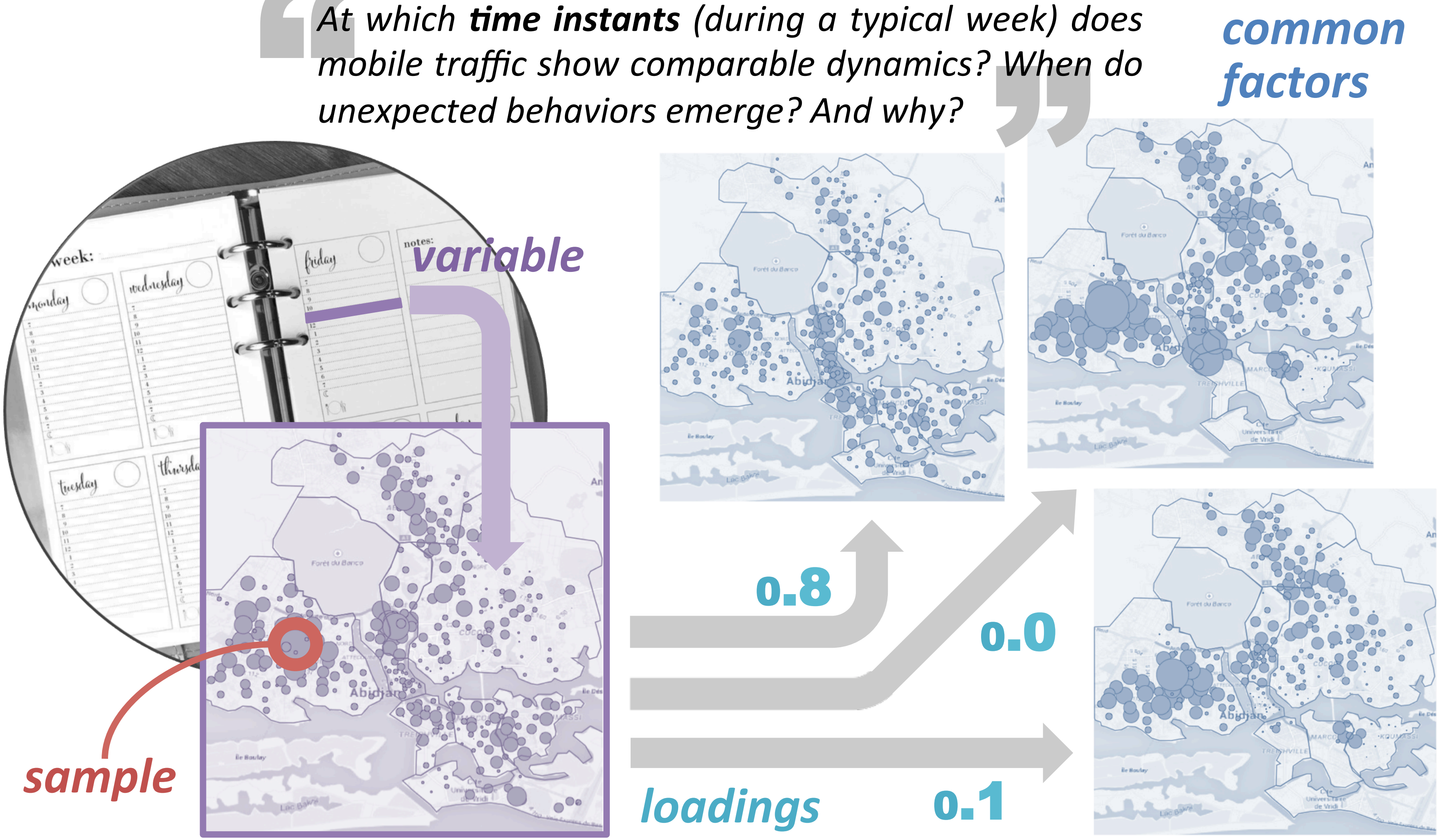
3

Temporal classification

Circadian rhythms in mobile network activity

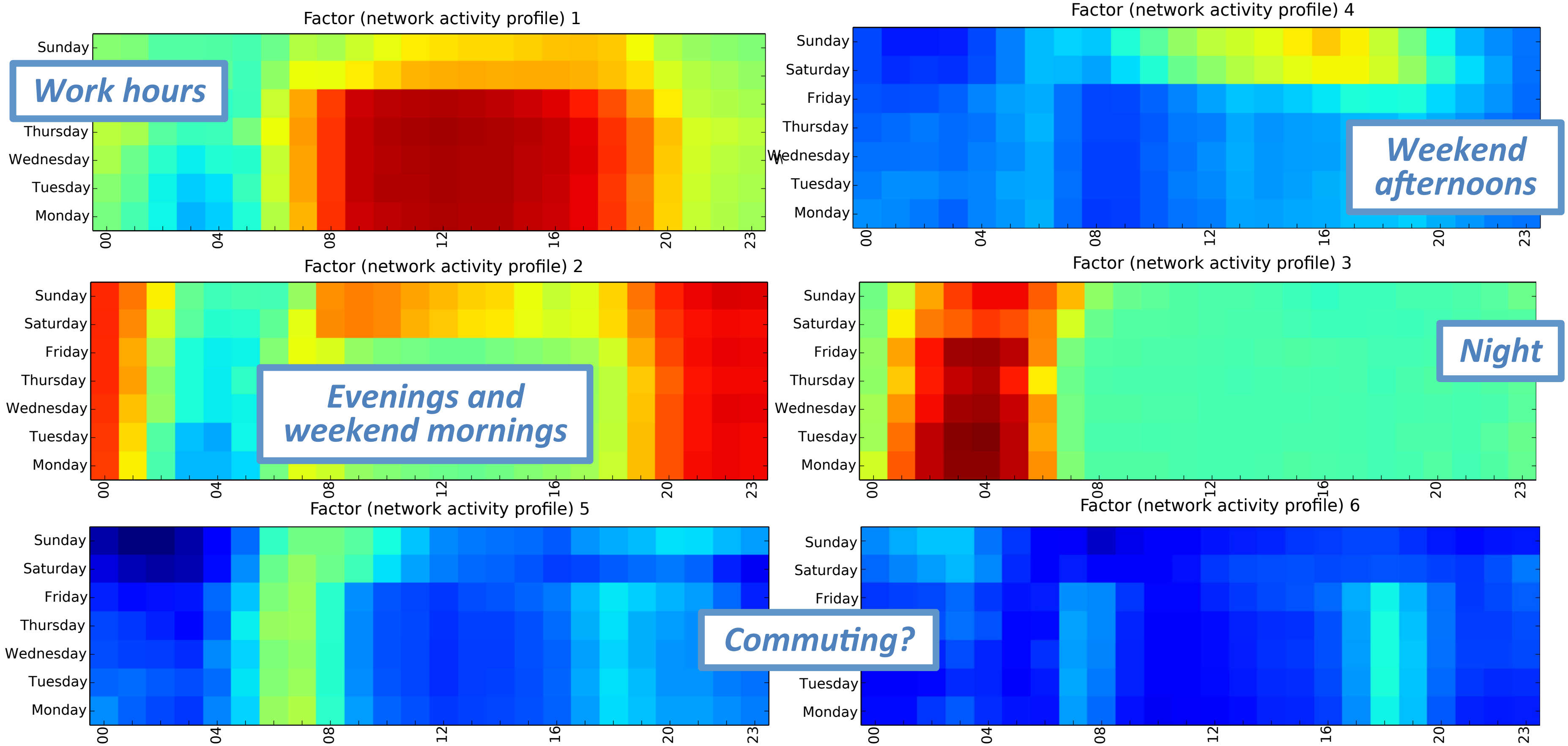
Methodology

*At which **time instants** (during a typical week) does mobile traffic show comparable dynamics? When do unexpected behaviors emerge? And why?*



Case study

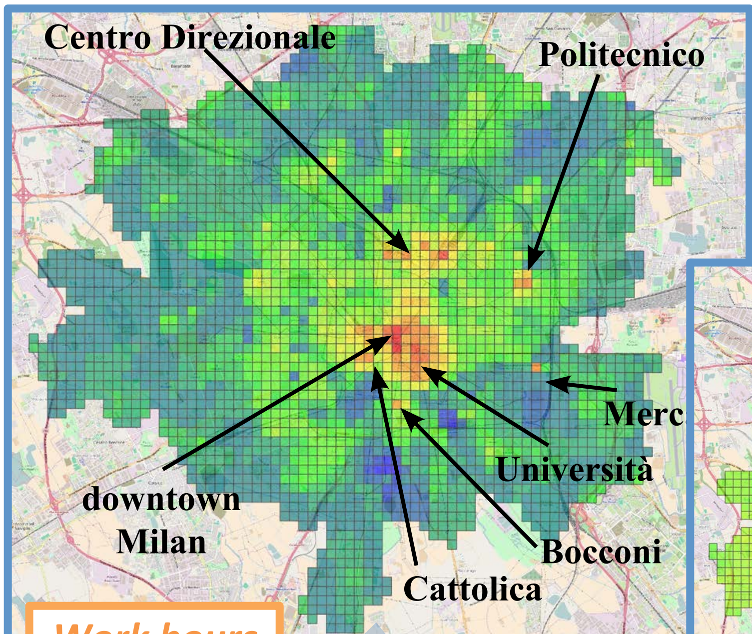
- **TIM** BDC 2013 [Milan, Italy, 2 months, grid]
  - *6 EFA factors* are identified



Case studies

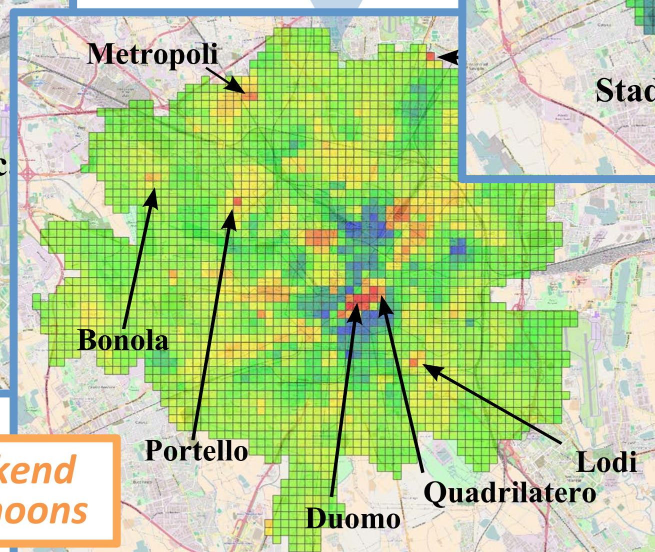
- EFA returns **scores**

  1. scores show ***in which areas*** each factor best characterizes the demand



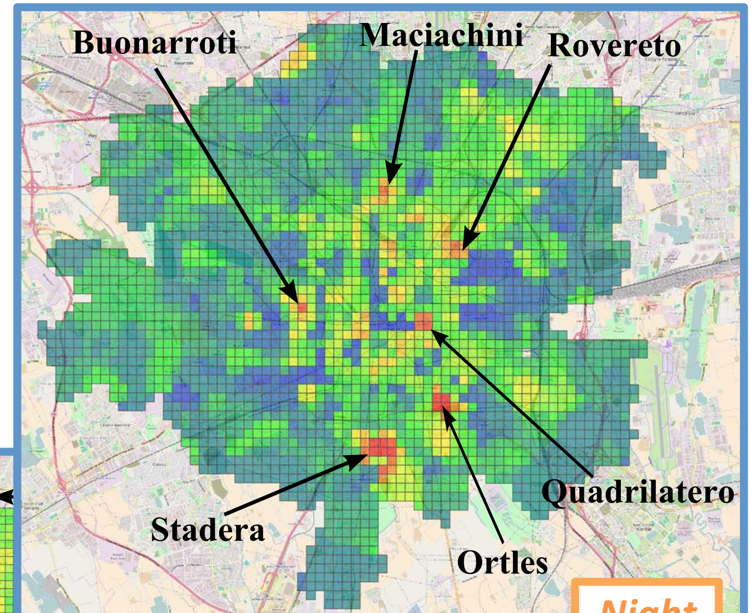
Work hours

downtown, universities, commercial areas



touristic, shopping, and entertainment areas

Weekend afternoons

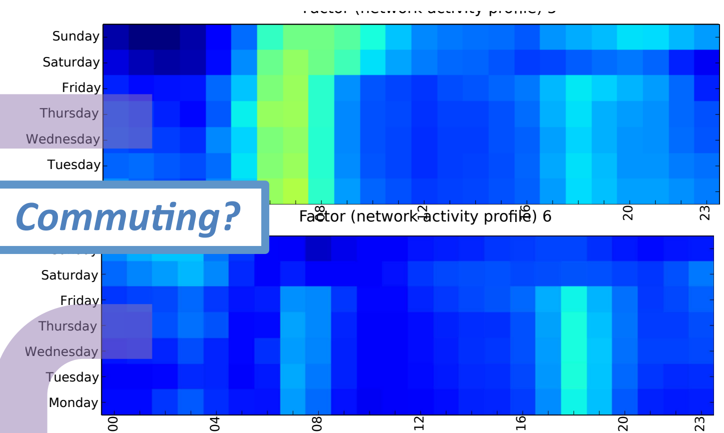


Night

ill-famed neighborhoods

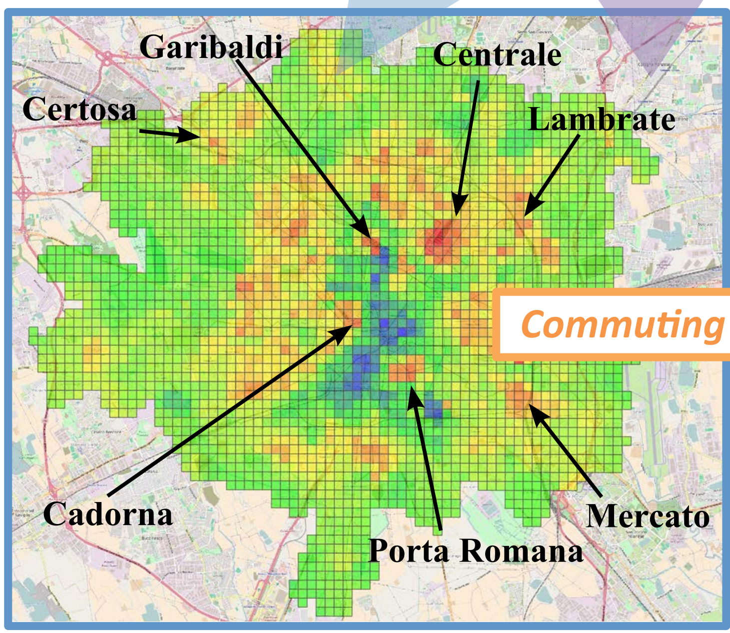
Case studies

- EFA returns **scores**
  - **2.** scores help ***disambiguating*** among apparently similar temporal profiles

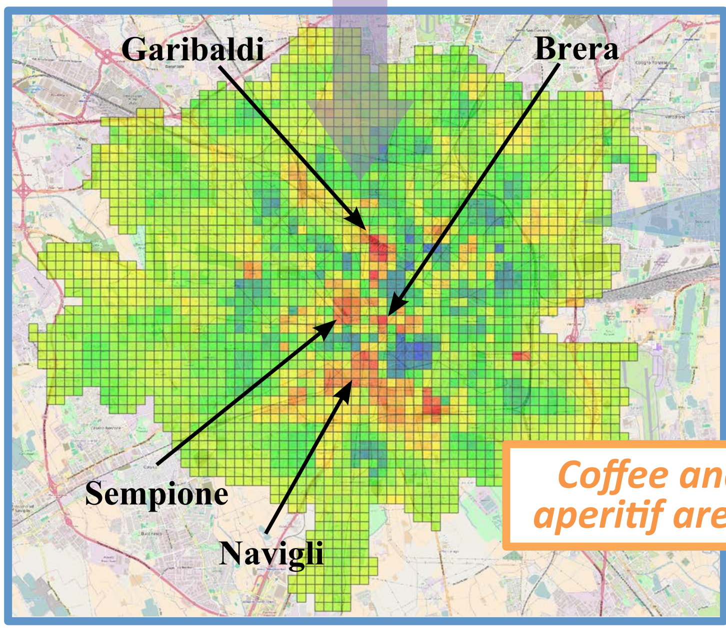


Commuting?

actual commuting areas, train stations, suburbs, beltway



Commuting



Coffee and aperitif areas

popular places where people relax before and after work

4

Outlook

Conclusions and perspectives

Outlook

- **Contributions**
  - Tools to **profile/understand typical** mobile network usages
    - over **space** and **time**
- **Perspectives**
  - Plenty of space for research
    - **Data** level
      - ***per-mobile service*** data collected through deep packet inspection (DPI) by dedicated probes at GGSN/PGW
      - ***fine-grained*** end terminal positioning data from control collected at RNC/eNodeB
      - ***merged*** fine-grained per-mobile service data from the sources above
    - **Algorithmic** level
      - ***alternative / more effective*** approaches to spatiotemporal classification
      - ***on-line*** operation on streaming data
      - ***privacy-by-design*** mobile traffic classification

Outlook

- **Applications**
  - A first step towards analytics for ***cognitive mobile networks***
    - **centralized orchestration mechanisms** (e.g., in 5GPPP pre-structuring model) need to be fed with knowledge of mobile traffic dynamics
    - macroscopic spatiotemporal profiles outline a (limited) number of **network-wide configurations** for long-timescale control operations[9]
      - C-RAN planning and optimization; spectrum assignment; centralized RAT selection; load balancing or traffic engineering in the CN; network resource de- and re-allocation; base station switch on/off; dynamic pricing
    - need for integration with new network technologies/paradigms
      - C-RAN resource allocation; cloudlet deployment; network slicing
  - Results are also relevant to ***other disciplines***
    - Geoinformatics, sociology, demographics, urban planning, etc.

[9] T. Chen et al., IEEE Comm. Mag., 2015

?

Thanks!

http://perso.citi.insa-lyon.fr/mfiore/

marco.fiore@ieiit.cnr.it

@marc0_fi0re

References

- **Temporal classification**
  - D. Naboulsi, R. Stanica, M. Fiore, *"Classifying Call Profiles in Large-scale Mobile Traffic Datasets"*, ***IEEE INFOCOM***, Toronto, Canada, 2014
  - A. Furno, D. Naboulsi, R. Stanica, M. Fiore, *"Mobile Demand Profiling for Cellular Cognitive Networking,"* ***IEEE Transactions on Mobile Computing***, 16(3), 2017
- **Spatial classification**
  - A. Furno, R. Stanica, M. Fiore, *"Comparative Evaluation of Urban Fabric Detection Techniques Based on Mobile Traffic Data,"* ***ACM/IEEE ASONAM***, Paris, France, 2015
  - A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, Z. Smoreda, *"A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas"*, ***IEEE Transactions on Mobile Computing***, 16(10), 2017
- **Spatiotemporal classification**
  - A. Furno, M. Fiore, R. Stanica, *"Joint Spatial and Temporal Classification of Mobile Traffic Demands"*, ***IEEE INFOCOM***, Atlanta, GA, USA, April 2017