

Diritto alla privacy e diritto alla conoscenza: un dilemma per la società dell'informazione

Dino Pedreschi

Pisa KDD Lab www-kdd.isti.cnr.it

University of Pisa and ISTI-CNR, Italy

Lezioni In Piazza – Non pagheremo la vostra crisi!

Pisa, Logge di Banchi, 12 Novembre 2008



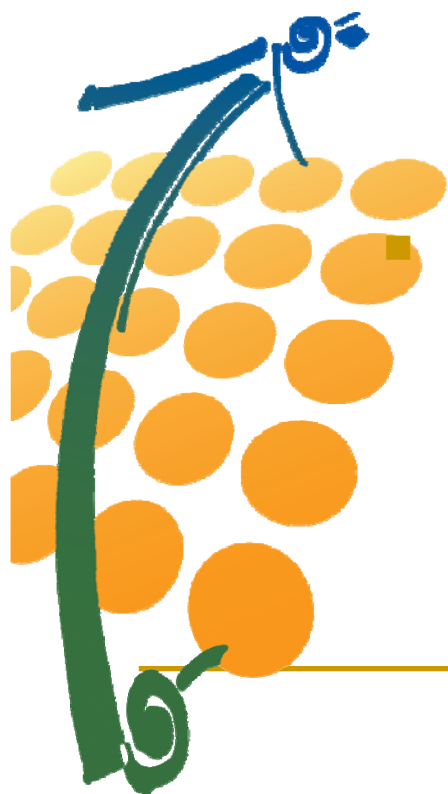
Le reti wireless e i dati di mobilità

- Le reti wireless sono la nervatura del territorio
- Oltre a fornire servizi, raccolgono tracce molto informative sulle attività umane
- Le infrastrutture UbiComp porteranno questo fenomeno al liite
- Miniaturizzazione, wearability, pervasività → tracce di crescente
 - accuratezza
 - ricchezza semantica



Quali dati di mobilità?

- Dati di ubicazione da cellulari, posizione di antenne nella rete GSM/UMTS.
- Dati di ubicazione da dispositivi GPS – forse un giorno Galileo?
 - L'attuale generazione di smartphones Nokia, iPhone, HTC, ... ha un ricevitore GPS a bordo e può trasmettere traiettorie GPS tramite SMS/MMS
- Dati di ubicazione da ...
 - peer-to-peer mobile networks
 - intelligent transportation environments – VANET
 - ad hoc sensor networks, RFIDs (radio-frequency ids)



Mobility, Data Mining and Privacy

- Verso una **archeologia del presente**?
- Uno scenario denso di opportunità e rischi:
 - Il **mobility data mining** può produrre conoscenza utile;
 - ma la **privacy** di ciascuno è a rischio.
- Una nuova area di ricerca multi-disciplinare sta emergendo, con un alto potenziale sociale ed economico
 - F. Giannotti and D. Pedreschi (Eds.)
Mobility, Data Mining and Privacy. Springer, 2008.





Un progetto paradigmatico: **GeoPKDD**

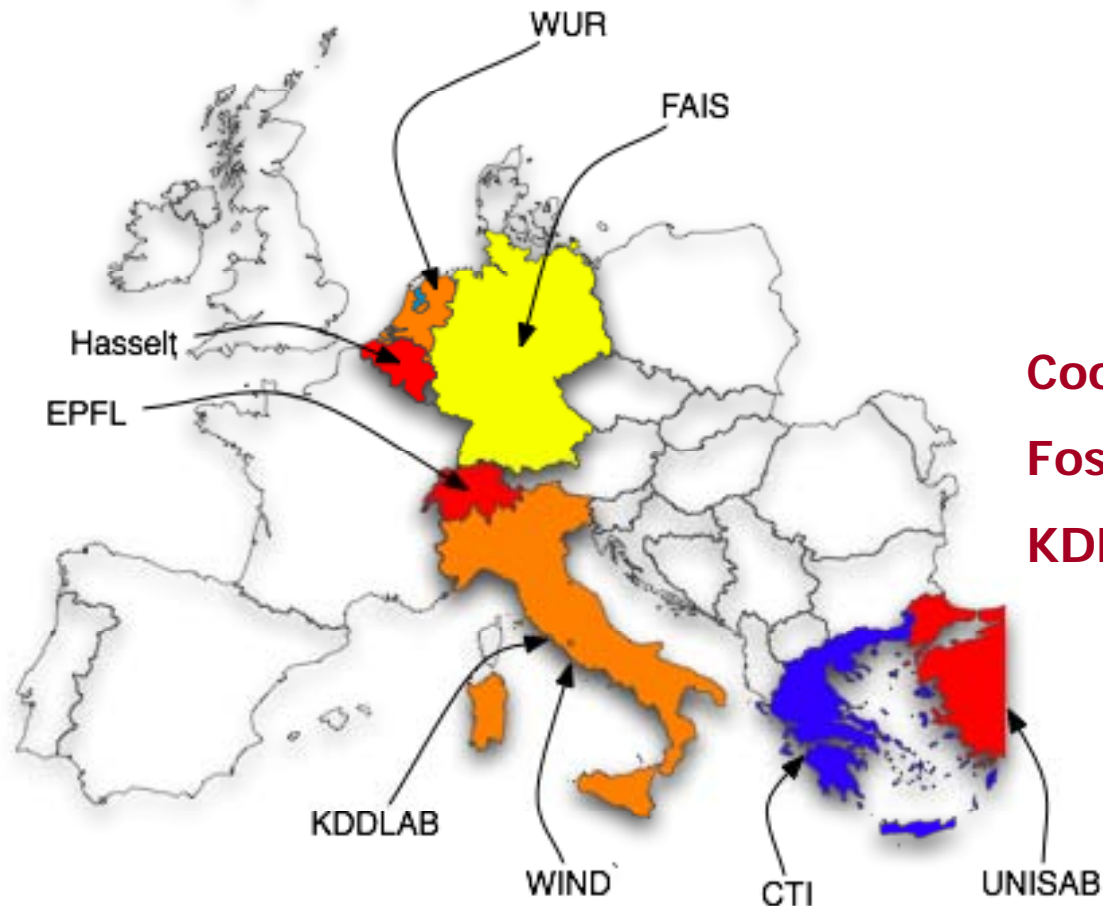
<http://www.geopkdd.eu>

A European FP6 project

Geographic Privacy-aware

Knowledge Discovery and Delivery

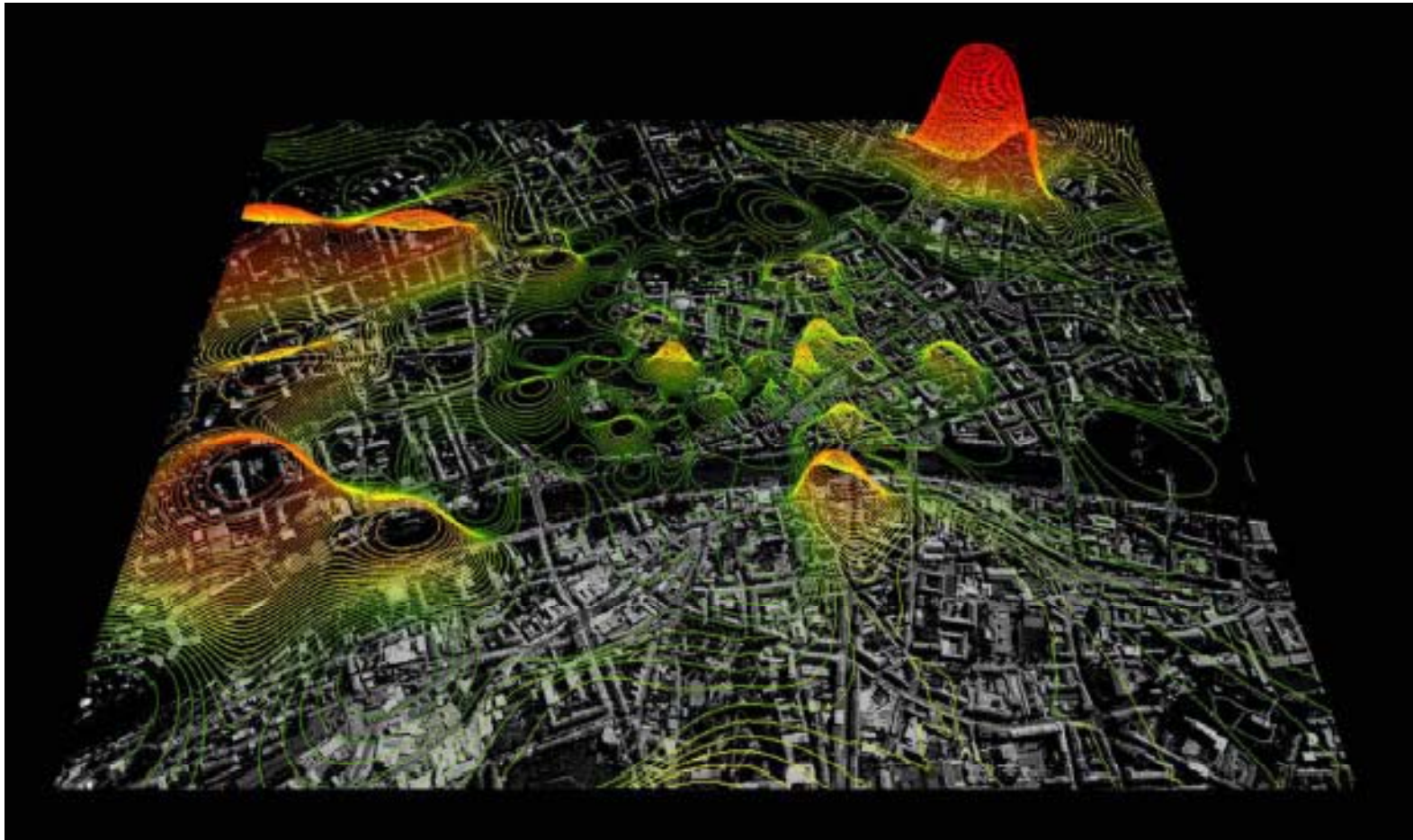




Coordinatore:
Fosca GIANNOTTI
KDD-LAB Pisa, ISTI-CNR



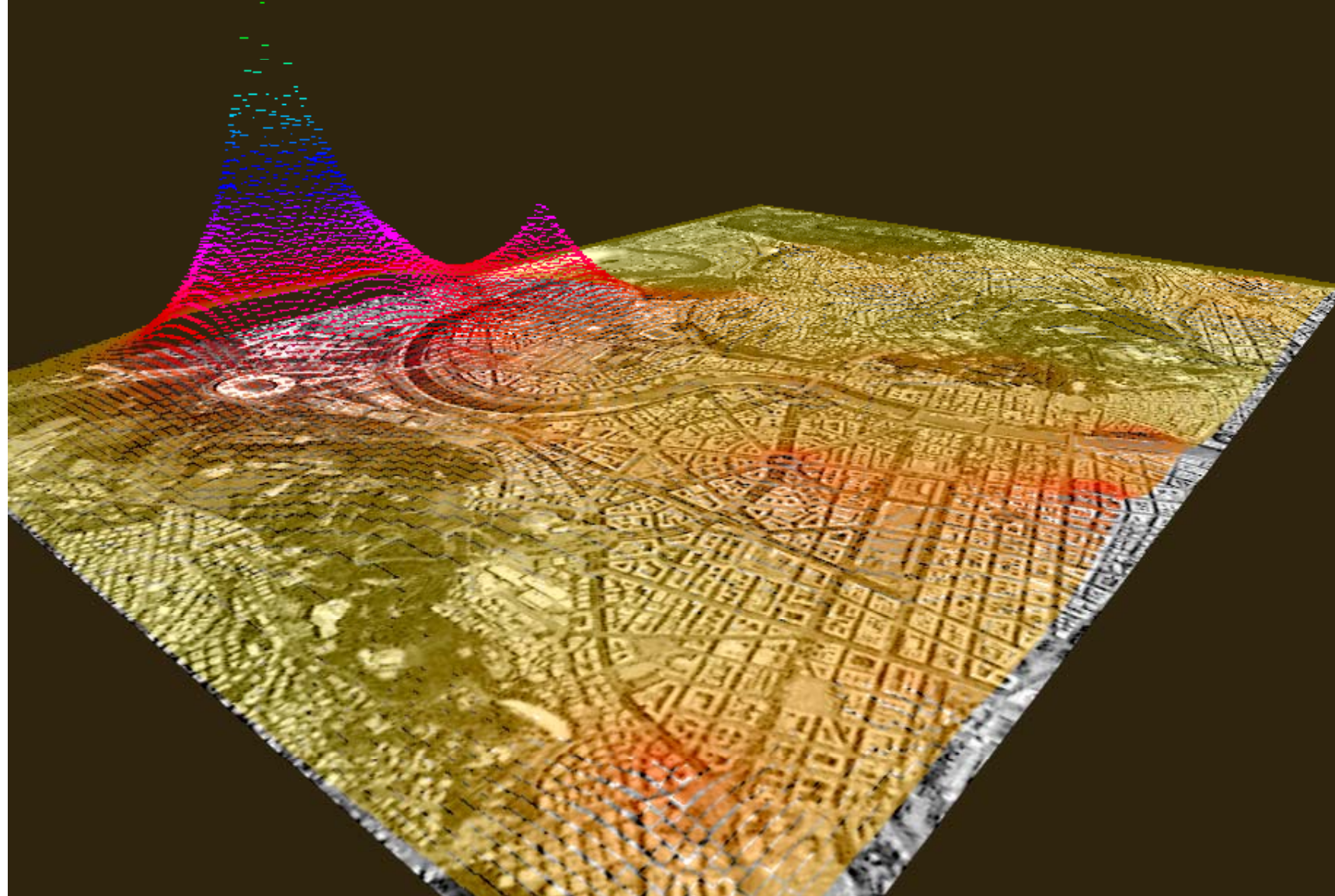
Real-time density estimation in urban areas



The senseable project: <http://senseable.mit.edu/grazrealtime/>

Madonna Concert
Cellphone activity in Stadio Olimpico Rome
2006-08-06

At Rome's Olympic Stadium
Located about three kilometres from the Vatican
During the song Live to Tell...
Madonna appeared against a mirrored cross



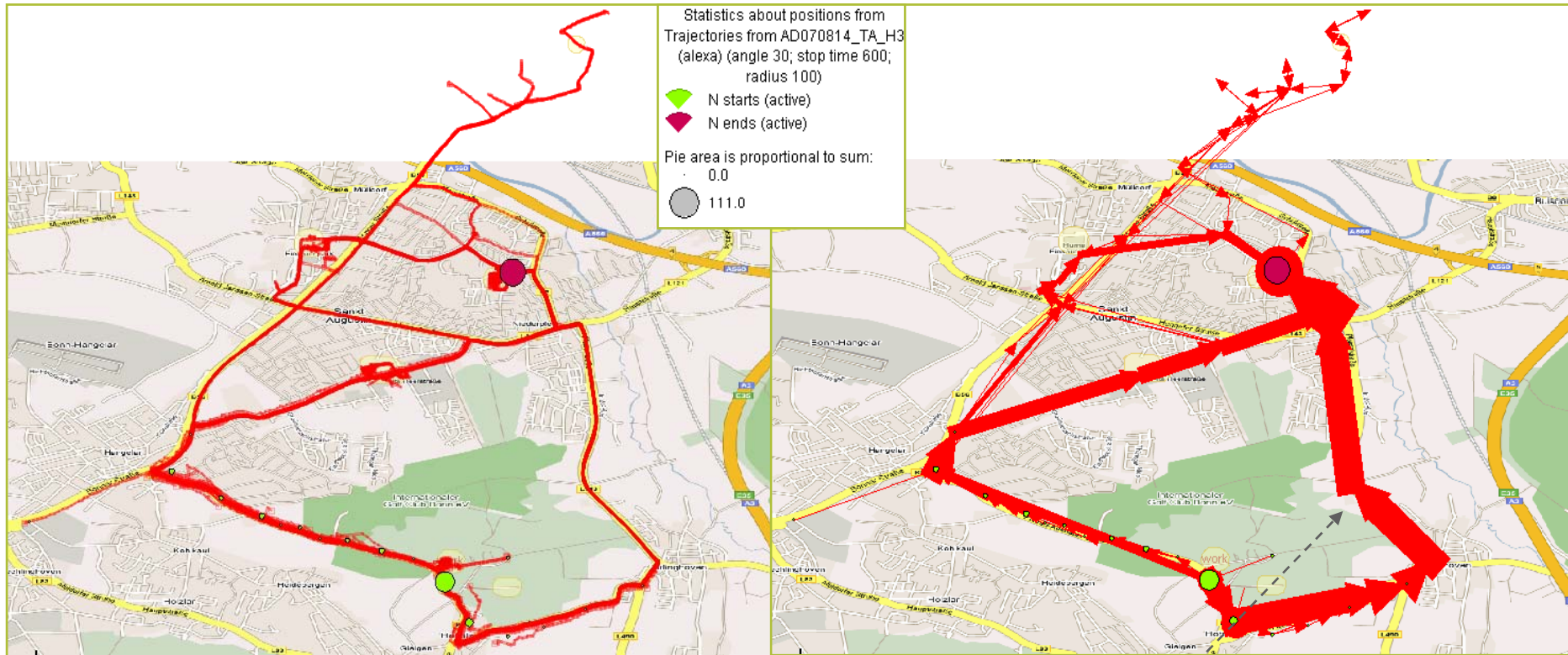
Più ambiziosamente: pattern di mobilità



Dai dati di mobilità ai pattern di mobilità



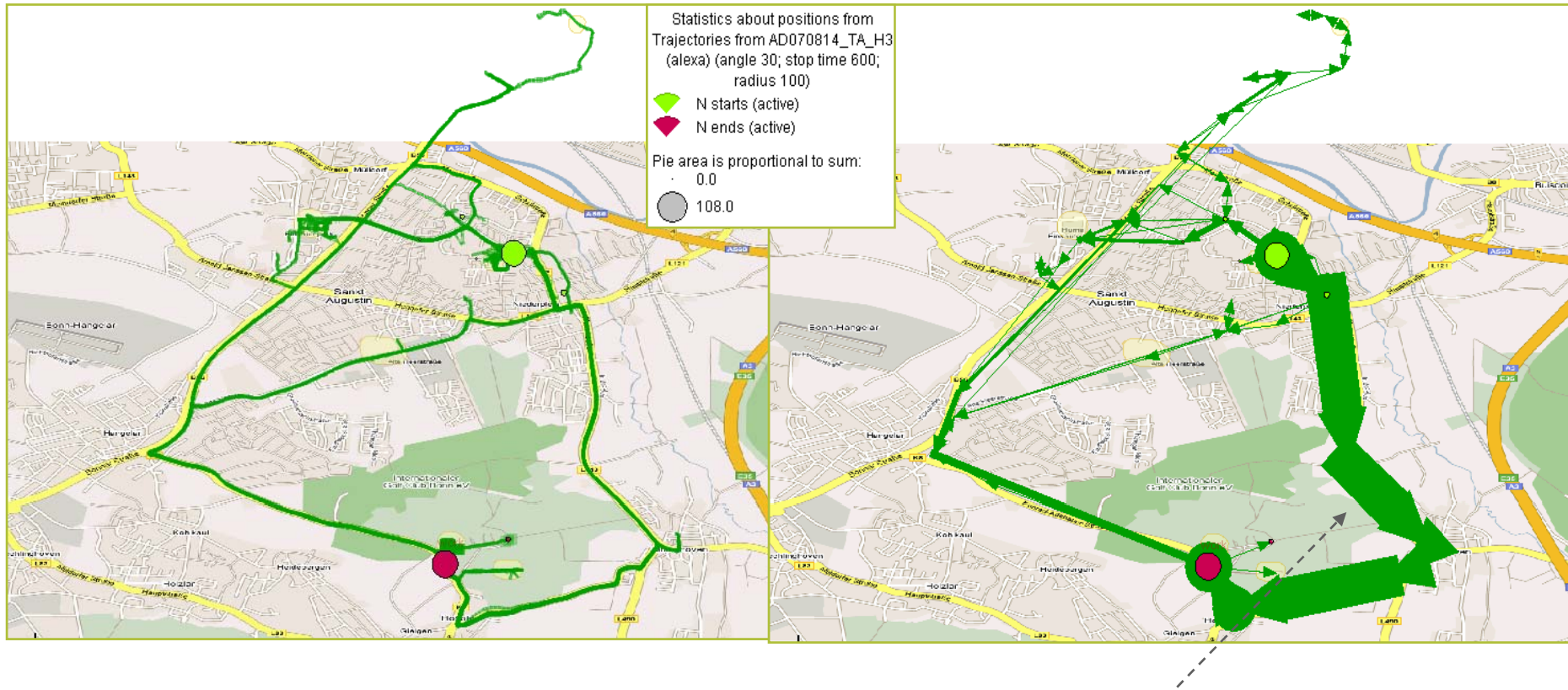
Cluster 1: casa → lavoro



Observation: the eastern route is chosen more often



Cluster 2: lavoro → casa

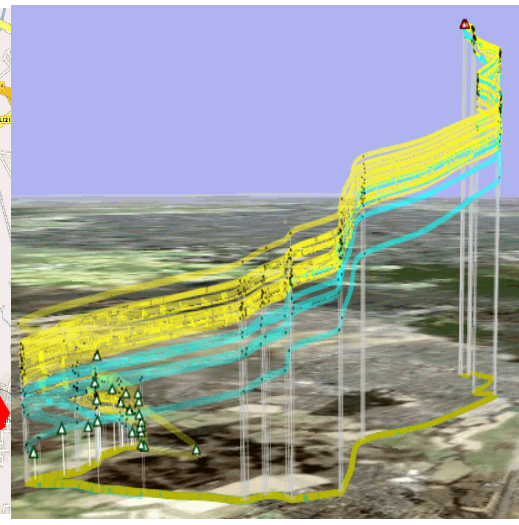
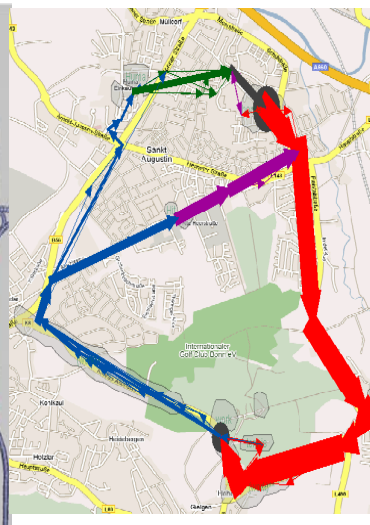


Observation: the eastern route is chosen much more often

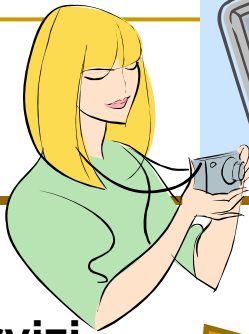
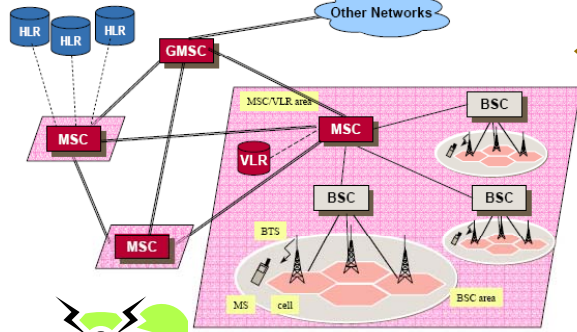


Dai dati di mobilità ai pattern di mobilità

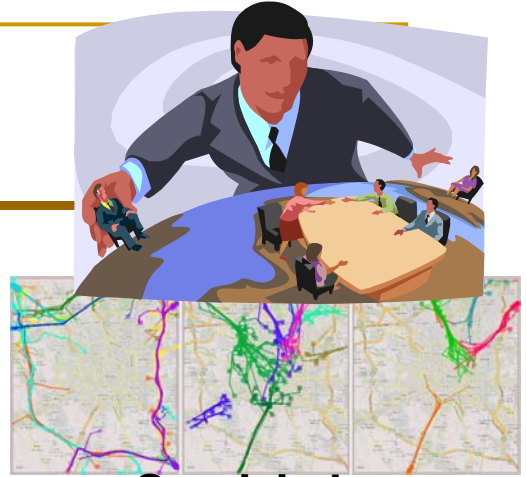
```
name|date|y|x  
Prinzessin|08.20.1998|52.118|12.087  
Prinzessin|08.23.1998|51.019|15.309  
Prinzessin|08.26.1998|47.723|22.786  
Prinzessin|08.29.1998|43.040|27.119  
Prinzessin|08.31.1998|38.715|32.165  
Prinzessin|09.01.1998|37.195|35.255  
Prinzessin|09.03.1998|32.979|36.021  
Prinzessin|09.05.1998|28.513|33.437  
Prinzessin|09.06.1998|23.961|32.937  
Prinzessin|09.07.1998|19.418|33.446  
Prinzessin|09.12.1998|15.823|34.094  
Prinzessin|10.11.1998|14.685|32.848  
Prinzessin|11.03.1998|11.510|32.591  
Prinzessin|11.24.1998|13.888|35.667  
Prinzessin|12.08.1998|12.562|34.777  
Prinzessin|12.10.1998|9.124|35.644  
...
```



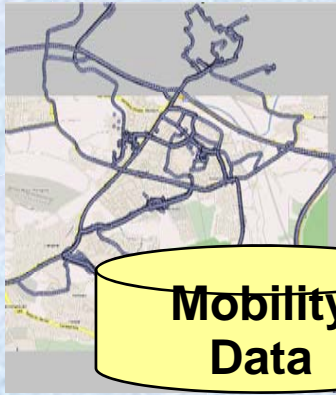
GSM network, WSN, GPS



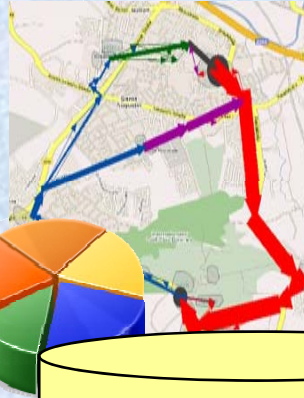
Servizi all'end user



Servizi al Mobility manager



Mobility Data



Mobility Patterns

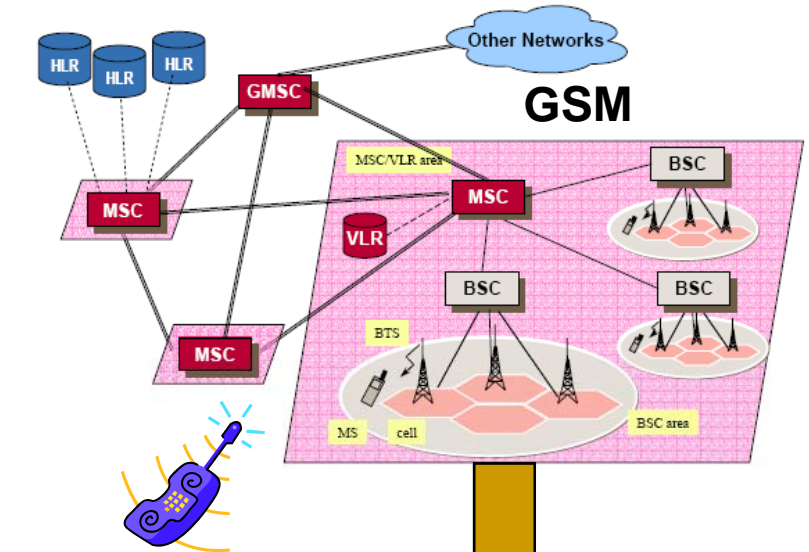


```

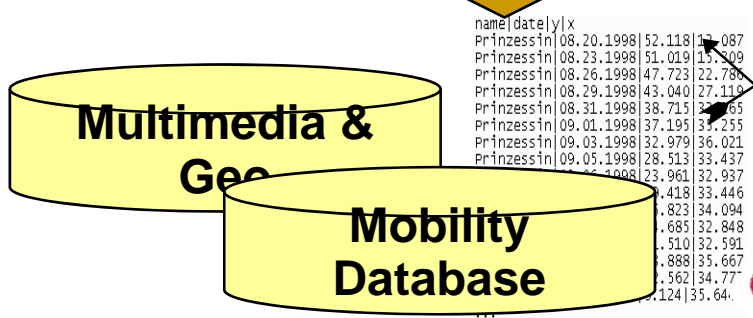
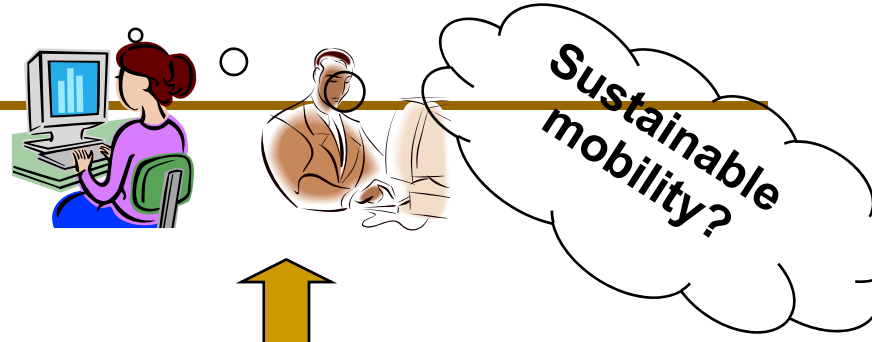
name|date|y|x
Prinzessin|08.20.1998|52.118|12.087
Prinzessin|08.23.1998|51.019|15.309
Prinzessin|08.26.1998|47.723|22.786
Prinzessin|08.29.1998|43.040|27.119
Prinzessin|08.31.1998|38.715|32.165
Prinzessin|09.01.1998|37.195|35.255
Prinzessin|09.03.1998|33.070|26.021
Prinzessin|09.05.1998|31.945|21.896
Prinzessin|09.07.1998|30.820|17.771
Prinzessin|09.09.1998|29.695|13.646
Prinzessin|09.11.1998|28.570|9.521
Prinzessin|12.08.1998|12.562|34.777
Prinzessin|12.10.1998|9.124|35.644
...
    
```

Raw data

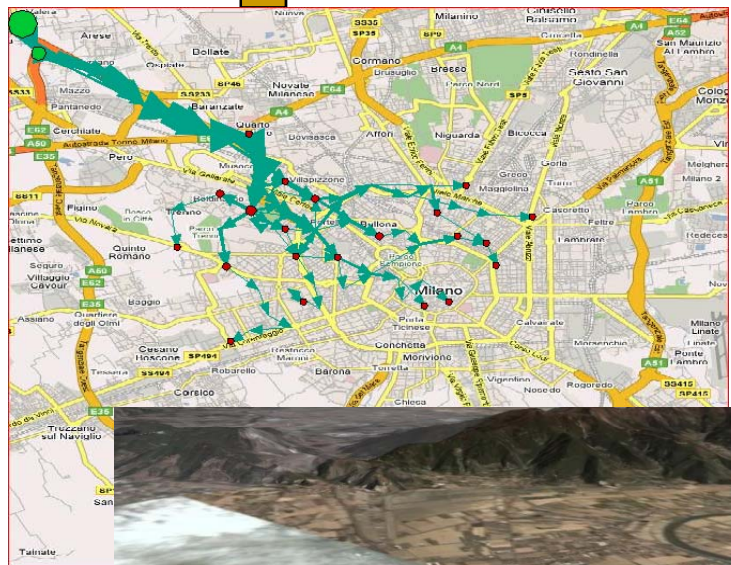
Privacy and anonymity protection



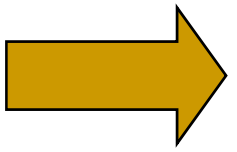
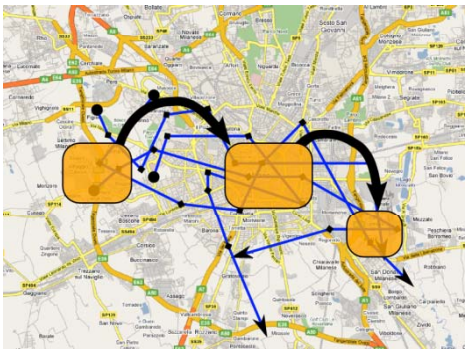
Mobility management



name	date	ly	x
Prinzessin	08.20.1998	52.118	12.087
Prinzessin	08.23.1998	51.019	15.509
Prinzessin	08.26.1998	47.723	22.786
Prinzessin	08.29.1998	43.040	27.119
Prinzessin	08.31.1998	38.715	32.665
Prinzessin	09.01.1998	37.195	33.255
Prinzessin	09.03.1998	32.979	36.021
Prinzessin	09.05.1998	28.513	33.437
		23.961	32.937
		20.418	33.446
		18.823	34.094
		16.685	32.848
		15.110	32.591
		13.888	35.667
		12.562	34.777
		11.124	35.641



Mobility models

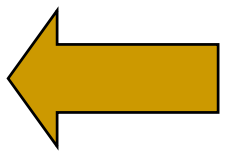
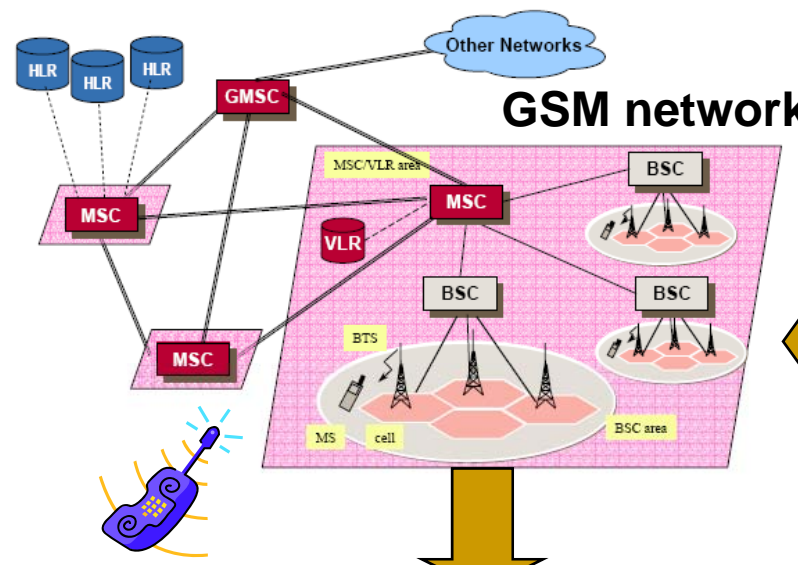


End user



Where should I go next?

GSM network



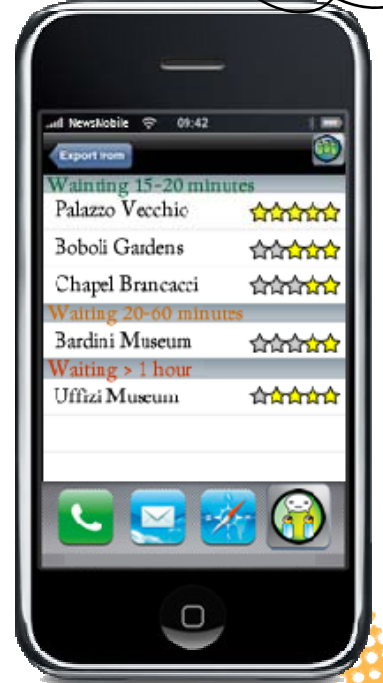
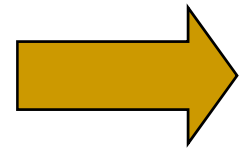
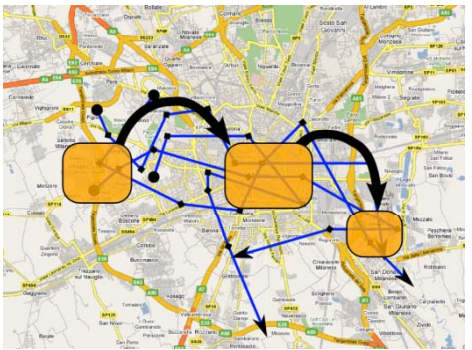
Multimedia & Geo

Mobility Database

name	date	ly	x
Prinzessin	08.20.1998	52.118	12.087
Prinzessin	08.23.1998	51.019	15.809
Prinzessin	08.26.1998	47.723	22.788
Prinzessin	08.29.1998	43.040	27.119
Prinzessin	08.31.1998	38.715	32.665
Prinzessin	09.01.1998	37.195	33.255
Prinzessin	09.03.1998	32.979	36.021
Prinzessin	09.05.1998	28.513	33.437
Prinzessin	09.08.1998	23.961	32.937
Prinzessin	09.11.1998	19.418	33.446
Prinzessin	09.14.1998	14.873	34.094
Prinzessin	09.17.1998	10.323	34.848
Prinzessin	09.20.1998	5.778	35.602
Prinzessin	09.23.1998	1.233	36.356
Prinzessin	09.26.1998	-3.312	37.110
Prinzessin	09.29.1998	-8.867	37.864
Prinzessin	10.02.1998	-14.422	38.618
Prinzessin	10.05.1998	-19.977	39.372
Prinzessin	10.08.1998	-25.532	40.126
Prinzessin	10.11.1998	-31.087	40.880
Prinzessin	10.14.1998	-36.642	41.634
Prinzessin	10.17.1998	-42.197	42.388
Prinzessin	10.20.1998	-47.752	43.142
Prinzessin	10.23.1998	-53.307	43.896
Prinzessin	10.26.1998	-58.862	44.650
Prinzessin	10.29.1998	-64.417	45.404
Prinzessin	11.01.1998	-70.000	46.158
Prinzessin	11.04.1998	-75.583	46.912
Prinzessin	11.07.1998	-81.166	47.666
Prinzessin	11.10.1998	-86.749	48.420
Prinzessin	11.13.1998	-92.332	49.174
Prinzessin	11.16.1998	-97.915	49.928
Prinzessin	11.19.1998	-103.498	50.682
Prinzessin	11.22.1998	-109.081	51.436
Prinzessin	11.25.1998	-114.664	52.190
Prinzessin	11.28.1998	-120.247	52.944
Prinzessin	12.01.1998	-125.830	53.698
Prinzessin	12.04.1998	-131.413	54.452
Prinzessin	12.07.1998	-137.000	55.206
Prinzessin	12.10.1998	-142.583	55.960
Prinzessin	12.13.1998	-148.166	56.714
Prinzessin	12.16.1998	-153.749	57.468
Prinzessin	12.19.1998	-159.332	58.222
Prinzessin	12.22.1998	-164.915	58.976
Prinzessin	12.25.1998	-170.498	59.730
Prinzessin	12.28.1998	-176.081	60.484
Prinzessin	12.31.1998	-181.664	61.238



Mobility models



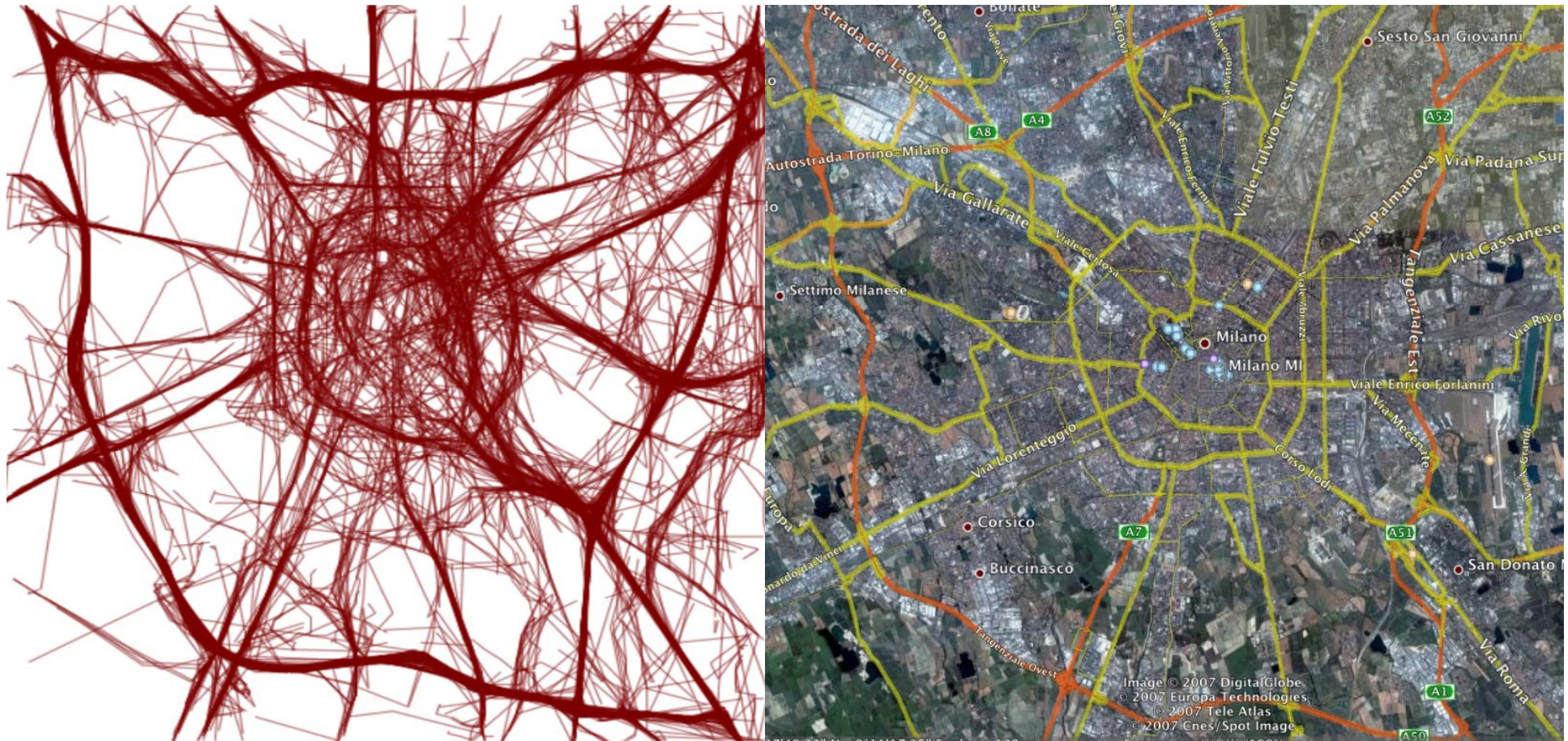
Analisi di Mobilità a Milano

- WIND Telecomunicazioni spa, in collaborazione con
 - Comune di Milano, Agenzia di Mobilità,
 - Infoblu e OctoTelematics (ricevitori GPS nell' auto con speciali contratti di assicurazione)
- Obiettivi
 - Estendere il concetto di analisi del traffico cittadino:
 - Costruire modelli di spostamento (ad es matrice origine-destinazione) più aggiornati e affidabili
 - Analizzare le correlazioni tra traffico e inquinamento, collezione di dati prima e dopo l'introduzione di ECOPASS



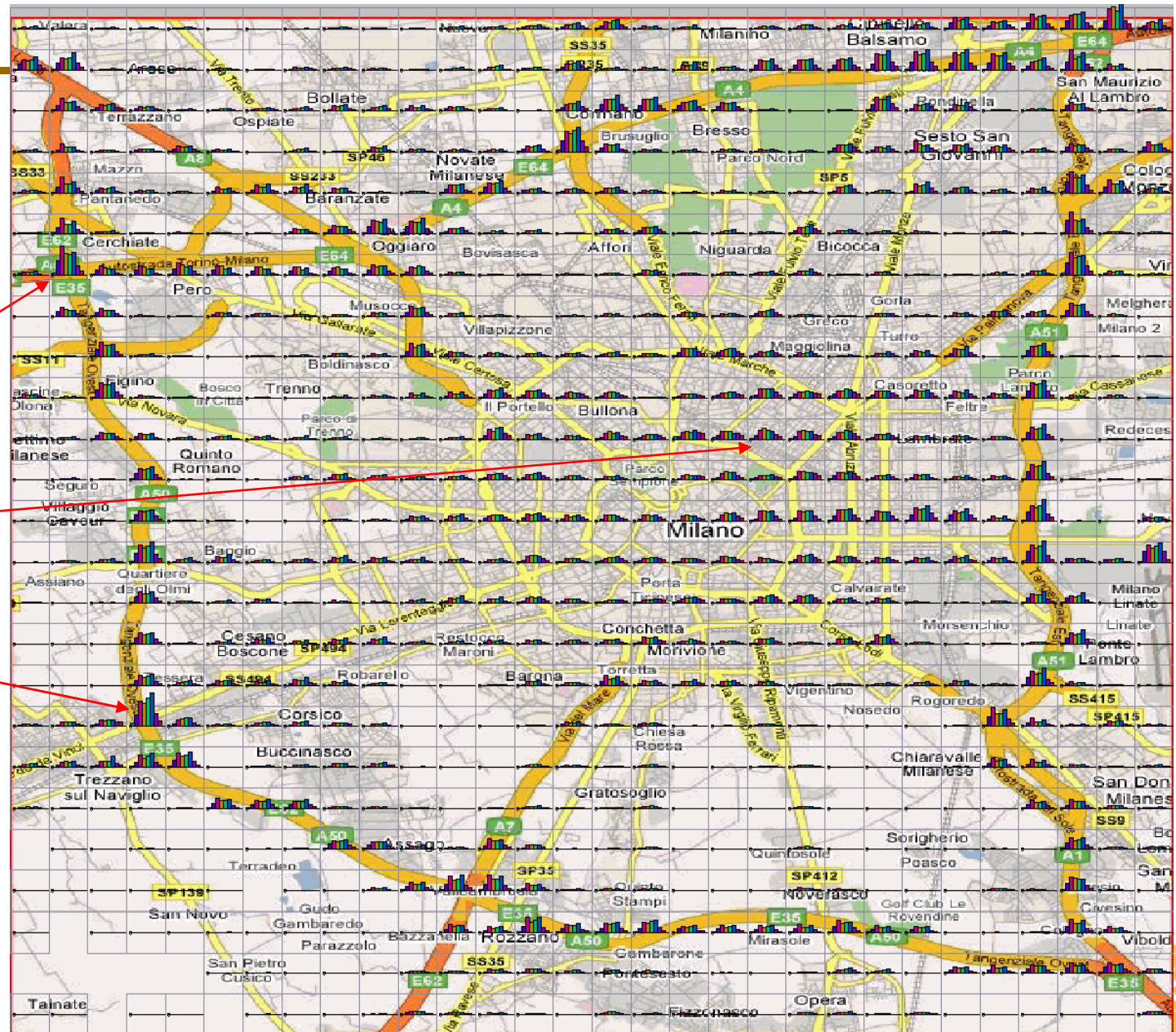
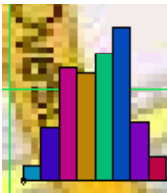
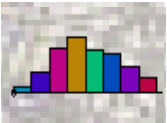
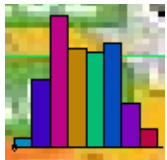
Analisi di Mobilità a Milano

- ❑ Fonti dei dati: Octotelematics
- ❑ Una settimana, 1-Apr-07 al 7-Apr-07
- ❑ 2.075M punti, 17.000 veicoli, 200.000 traiettorie



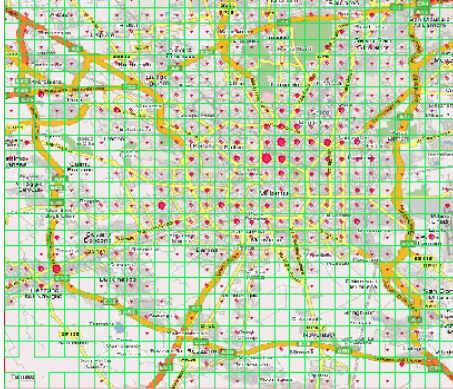
Traffic density patterns (spatio-temporal aggregation)

- count, parameter 0=0..2
- count, parameter 0=3..5
- count, parameter 0=6..8
- count, parameter 0=9..11
- count, parameter 0=12..14
- count, parameter 0=15..17
- count, parameter 0=18..20
- count, parameter 0=21..23

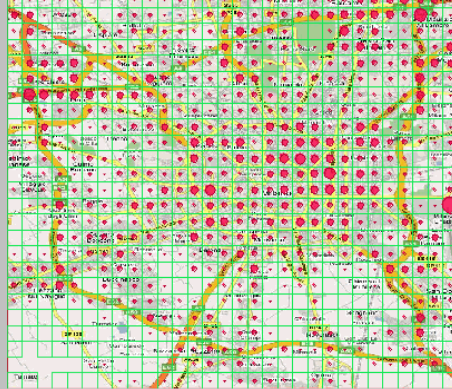


Low-speed movement (counts, 3h intervals)

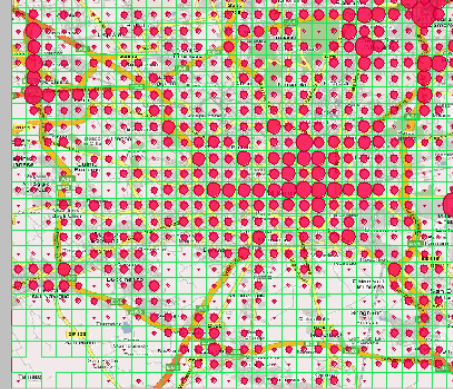
count, parameter 0=0, parameter 1=0..2



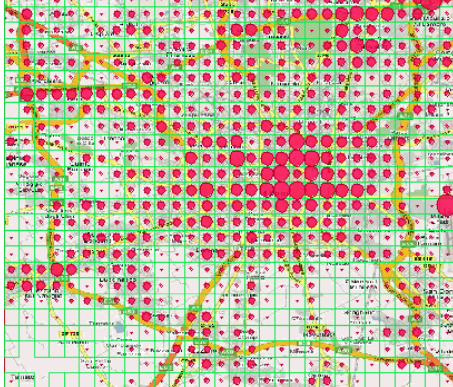
count, parameter 0=0, parameter 1=3..5



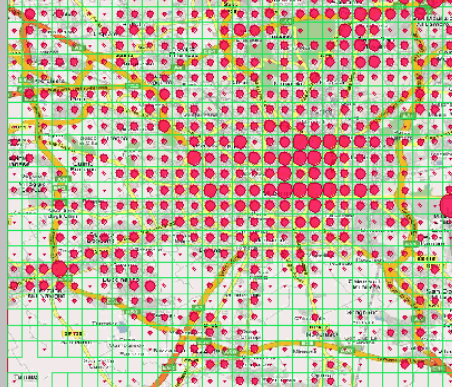
count, parameter 0=0, parameter 1=6..8



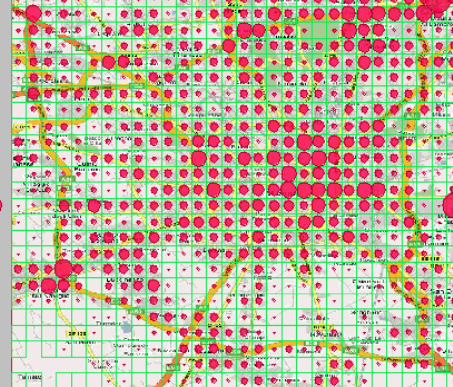
count, parameter 0=0, parameter 1=9..11



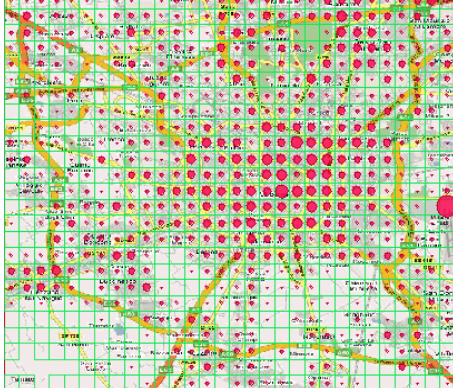
count, parameter 0=0, parameter 1=12..14



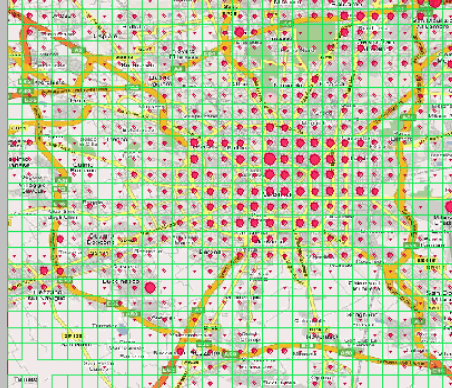
count, parameter 0=0, parameter 1=15..17



count, parameter 0=0, parameter 1=18..20



count, parameter 0=0, parameter 1=21..23



Trajectory Patterns

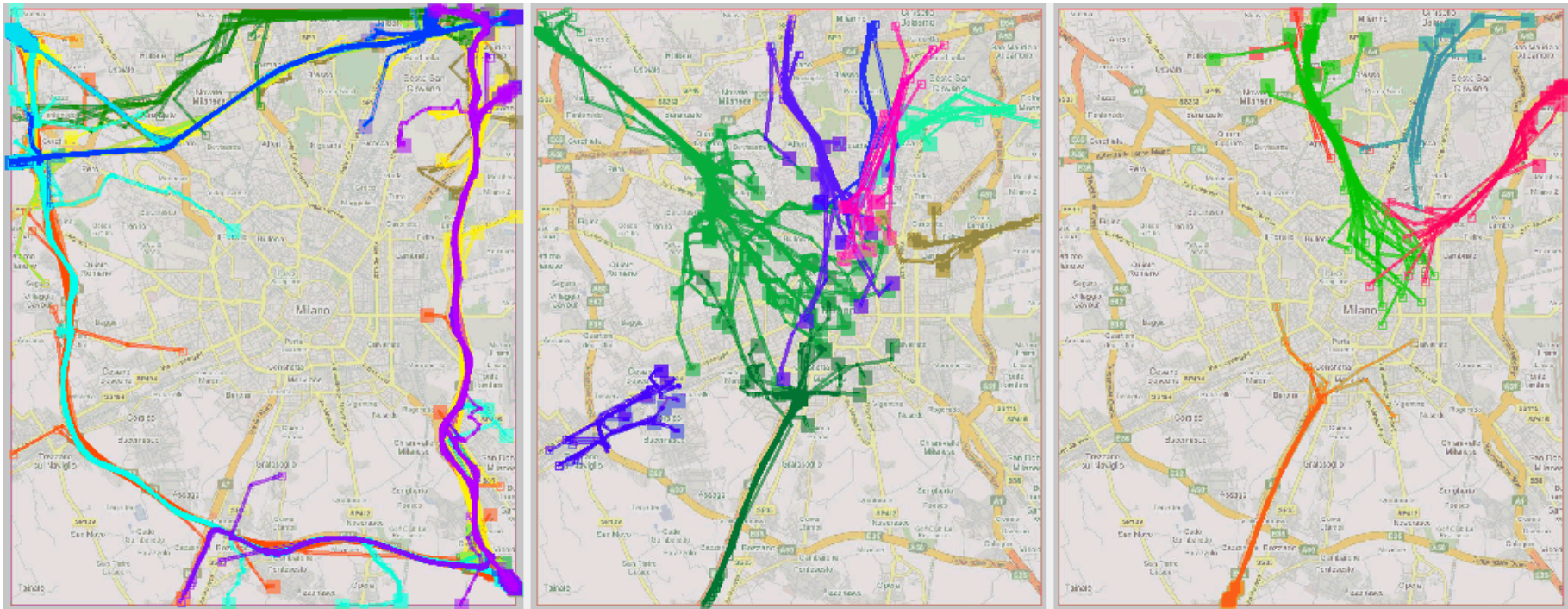


Working days

Sunday



Clustering trajectories on “route similarity”



Left: peripheral routes; middle: inward routes; right: outward routes.

- Rinzivillo, Pedreschi, Nanni, Giannotti, Andrienko, Andrienko
Visually-driven analysis of movement data by progressive clustering. J. of Information Visualization, 2008



February 8, 2008 5:56 PM PST

Nokia turns people into traffic sensors

Posted by [Erica Ogg](#)

[8 comments](#)

UNION CITY, Calif.--On a cool, overcast morning in the parking lot of a Lowe's hardware store, 100 UC Berkeley students lined up in rows ready to jump into a bevy of idling vehicles.

With media and VIPs from companies like Nokia, Navteq, General Motors, BMW, and CalTrans looking on, wave after wave of students left the parking lot to drive a 10-mile stretch of the nearby 880 freeway as part of a large-scale experiment to test how cell phones can monitor and predict traffic.

The test, conducted all day Friday, was put on by the California Center for Innovative Transportation (CCIT) as a joint project between Nokia, CalTrans, and Berkeley's Department of Civil and Environmental Engineering.

Each student car was issued a Nokia N95 phone with GPS and special traffic-monitoring software developed by Nokia's Palo Alto, Calif.-based research lab--plus a Bluetooth headset. As the students drove the freeway, the phone sent data about each car's speed and position back to the company's research facility. The data is compiled and used to predict traffic patterns and help drivers get where they need to be quickly. Nokia hopes that one day the system could be a significantly cheaper way to track traffic than the permanent sensors installed in roadways or next to them because it uses equipment most people already own: cell phones.



Video: Using cell phones to track traffic

Alex Bayen, a professor of civil and environmental engineering and lead researcher on the project for Berkeley, called the experiment "a glimpse into the future of traffic information

[Ad Feedback](#)

Dice discussions
The Career Hub for Tech Insiders™

Q: Should I bother trying to gain mainframe experience?

Get all kinds of perspectives at DICE DISCUSSIONS

About News Blog

Recent posts on technology, trends, and more.

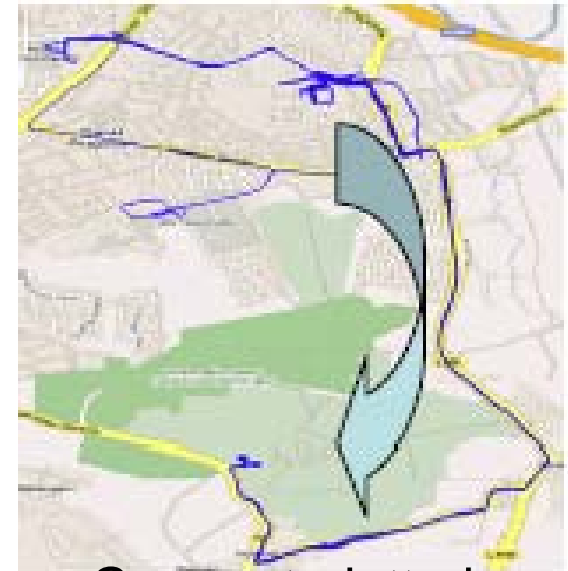
[Subscribe to this blog](#)
[Click this link to view this blog as XML.](#)

Add this feed to your online news reader

- [Add to Google](#)
- [Add to my Yahoo](#)
- [Add to MSN](#)
- [Add to Bloglines](#)
- [Add to Newsgator](#)

Dalle opportunità ai rischi

- I dati di localizzazione spaziale e temporale consentono di inferire informazioni sensibili
 - Possono rivelare abitudini, costumi sociali, orientamenti religiosi e sessuali degli individui
- Allo stesso tempo, consentono di identificare gli individui
 - Esempio: una traiettoria che parte da casa e termina al luogo di lavoro, in orari che si ripetono durante la settimana
 - Cancellare l'identificatore della traiettoria (nome, cognome, ecc.) non è sufficiente !!



Gruppo traiettorie
Casa → Ufficio



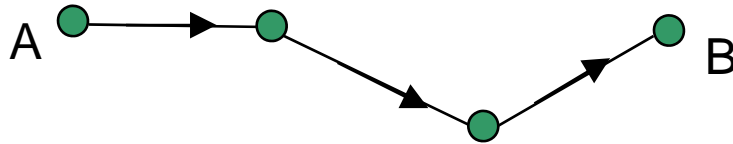
Anonimato dei dati personali

- Rendere anonimi i dati personali è difficile
- Spesso è possibile risalire all'identità degli interessati a partire da dati de-identificati.
- Molti famosi esempi di re-identificazione:
 - I dati clinici del Governatore del Massachusetts (Sweeney's experiment, 2001)
 - La crisi America On Line dell'agosto 2006 : re-identificazione di utenti attraverso i search logs
- Due pericoli:
 - **Molte osservazioni** sullo stesso soggetto "anonimo"
 - **Collegamento fra dati** in database separati

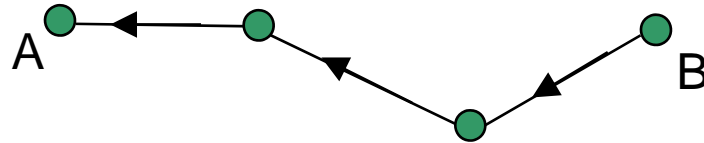


Spatio-temporal linkage nei dati di mobilità

Id:
34567



[tutti i giorni feriali dalle
7:45 alle 8:15]



[tutti i giorni feriali dalle
17:45 alle 18:15]

- Solo una persona vive in A e lavora in B.
- Id:34567 = Prof. Smith
- ... poi dalle altre tracce di Id:34567 si scopre che il sabato sera ...



Privacy-preserving spatio-temporal data mining

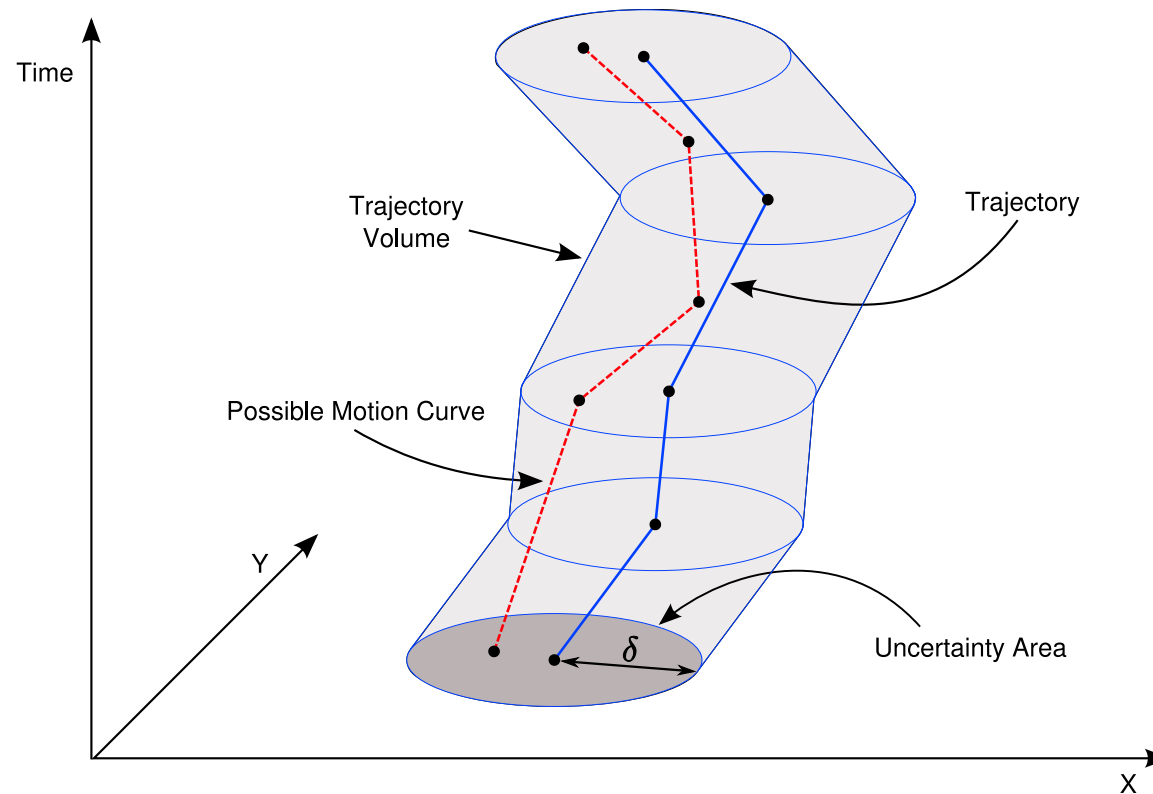
Trajectory anonymization



Tecniche per l'anonimato dei dati

■ Anonimizzazione

- Modificare i dati in modo che ogni traiettoria sia indistinguibile da almeno altre K
- ... minimizzando la distorsione introdotta nei dati



Concludendo ...



Mentre i dati di mobilità ci sommergeranno ...

- ... il mobility data mining sta emergendo come un nuovo tema scottante
- L'Ubiquitous Computing creerà **sistemi informativi pervasivi**, attraverso flussi di dati semanticamente arricchiti (in un contesto decentralizzato)
- La sfida è agli inizi!
- Attenzione: senza una alleanza fra etica, tecnologia e legge non si conquisterà la fiducia dei cittadini!



Fosca Giannotti
Dino Pedreschi (Eds.)

Giannotti
Pedreschi (Eds.)



Mobility, Data Mining
and Privacy

Mobility, Data Mining and Privacy

Geographic Knowledge Discovery

Giannotti · Pedreschi (Eds.)

Mobility, Data Mining and Privacy

The technologies of mobile communications and ubiquitous computing permeate our society, and wireless networks sense the movement of people and vehicles, generating large volumes of mobility data. This is a source of great opportunities and risks: on one side, mining this data can produce useful knowledge, supporting sustainable mobility and intelligent transportation systems; on the other side, individual privacy is at risk, as the mobility data contain sensitive personal information. A new multidisciplinary research area is emerging at the crossroads of mobility, data mining, and privacy.

This book assesses this research frontier from a computer science perspective, investigating the various scientific and technological issues, open problems, and roadmap. The editors manage a research project called GeoPDD (Geographic Privacy-Aware Knowledge Discovery and Delivery), funded by the EU Commission and involving 40 researchers from 7 countries, and this book tightly integrates and relates their findings in 13 chapters covering all related subjects, including the concepts of movement data and knowledge discovery from movement data; privacy-aware geographic knowledge discovery; wireless network and next-generation mobile technologies; trajectory data models, systems and warehouses; privacy and security aspects of technologies and related regulations; querying, mining and reasoning on spatiotemporal data; and local analytics methods for movement data.

This book will benefit researchers and practitioners in the related areas of computer science, geography, social science, statistics, law, telecommunication and transportation engineering.

ISBN 978-3-540-75176-2



springer.com

 Springer

Pisa, Italy - December 15th - 19th, 2008

ICDM 2008

IEEE International Conference on Data Mining



<http://icdm08.isti.cnr.it/>

Sponsors



Organizations



Conference Co-chairs

Franco Turini (KDD Lab, Univ. Pisa, Italy)
Carlo Zaniolo (UCLA, USA)
Naren Ramakrishnan (Virginia Tech, USA)

Program Committee Chairs

Fosca Giannotti (KDD Lab, ISTI-CNR, Italy)
Dimitrios Gunopoulos (UC Riverside, USA)

Steering Committee Chair

Xindong Wu (Univ. Vermont, USA)

Tutorials Chairs

Dino Pedreschi (KDD Lab, Univ. Pisa, Italy)
Arno Siebes (Utrecht Univ., The Netherlands)

Workshops Chairs

Francesco Bonchi (Yahoo! Research, Spain)
Bettina Berendt (Humbolt Univ. Berlin, Germany)

Award Chair

Katharina Morik (Univ. Dortmund, Germany)

Panels Chair

Jean-Francois Boulicaut (INSA Lyon, France)

Exhibit and Demo Chairs

Haijun Wang (IBM T. J. Watson Research Center, USA)
Michail Vlachos (IBM T. J. Watson Research Center, USA)

Important Dates

- July 7, 2008 Deadline for paper submission
- September 15, 2008 Notification to authors
- October 7, 2008 Deadline for camera-ready copies
- December 15 - 19, 2008 Conference

ICDM 2008

The IEEE International Conference on Data Mining

December 15-19, 2008
Pisa, Italy
<http://icdm08.isti.cnr.it>

