

Big Data Analytics & Social Mining for Science and Society

Dino Pedreschi



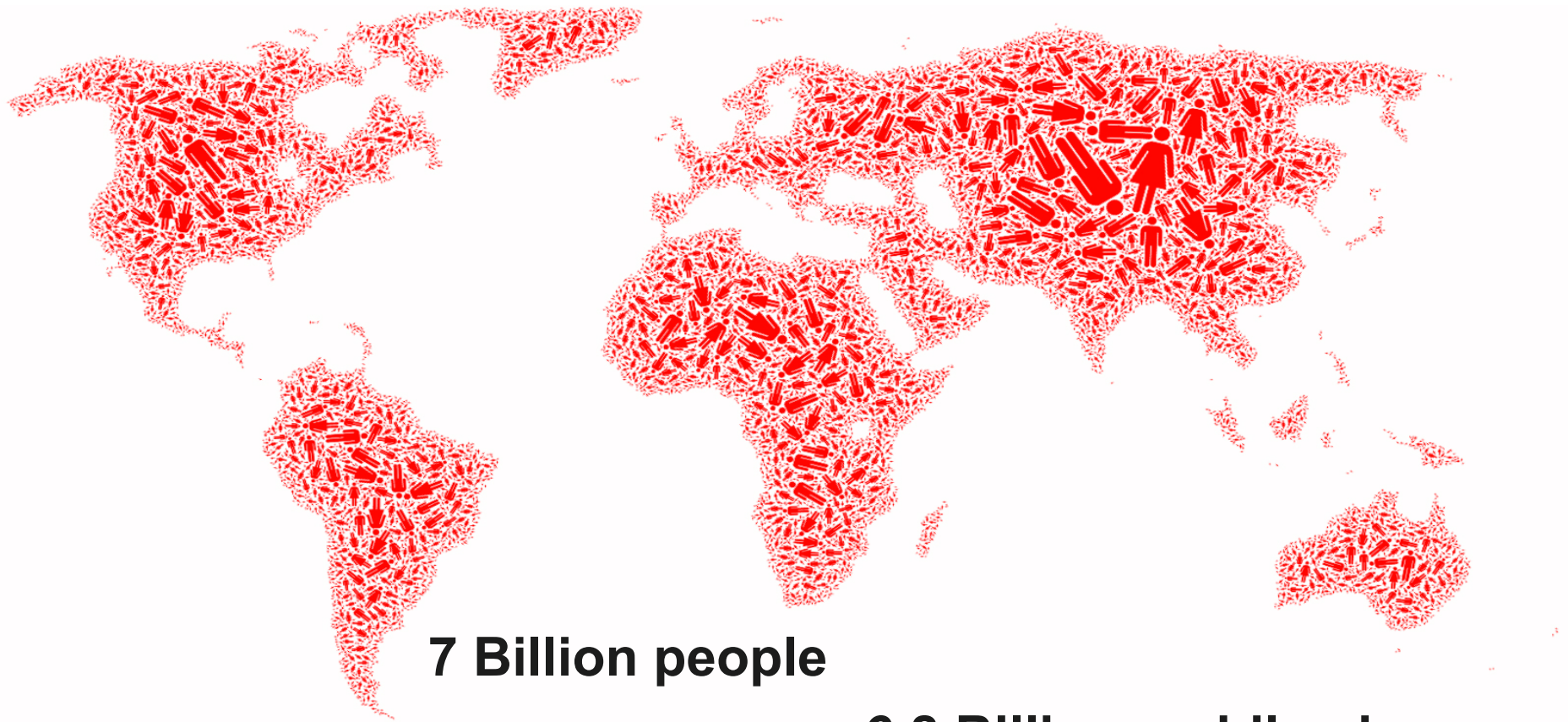
Università di Pisa & ISTI-CNR



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







7 Billion people

6.8 Billion mobile phones



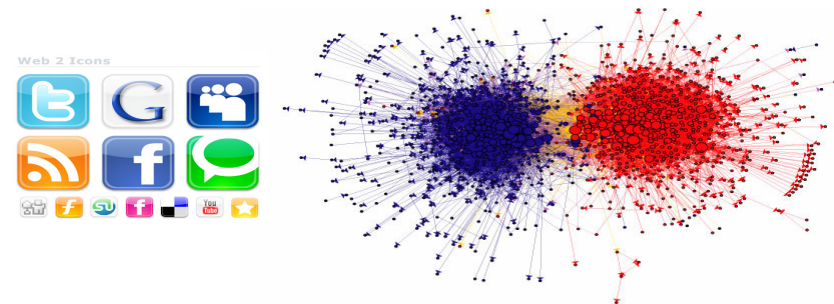
SIAMO TUTTI POLLICINI DIGITALI

Big data proxies of social life

SHOPPING PATTERNS & LIFESTYLE



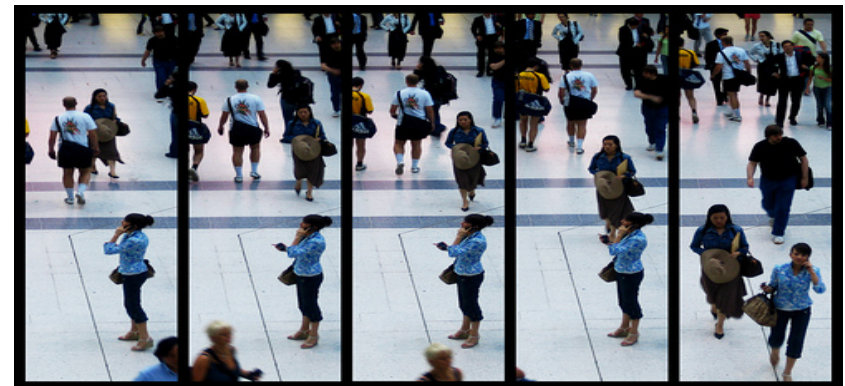
RELATIONSHIPS & SOCIAL TIES

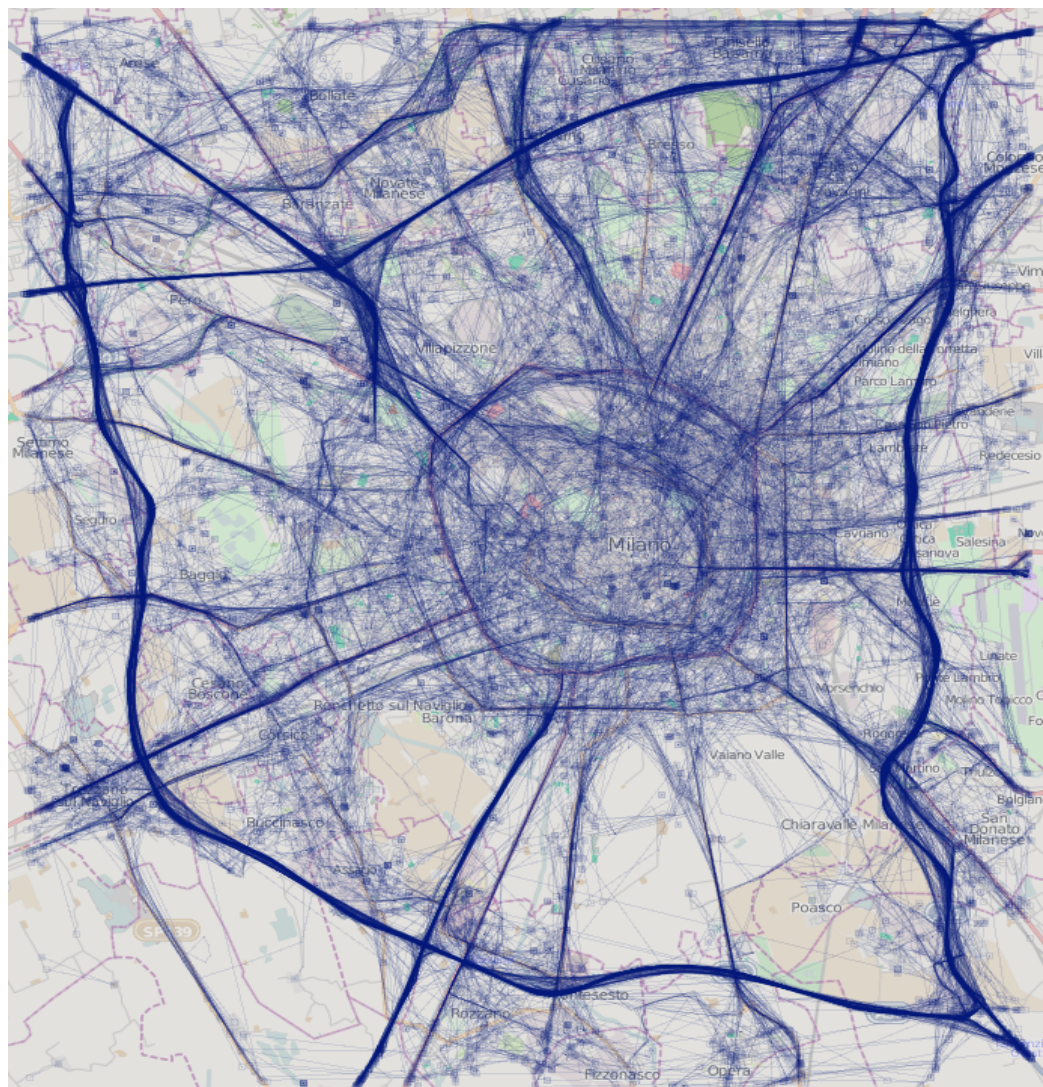


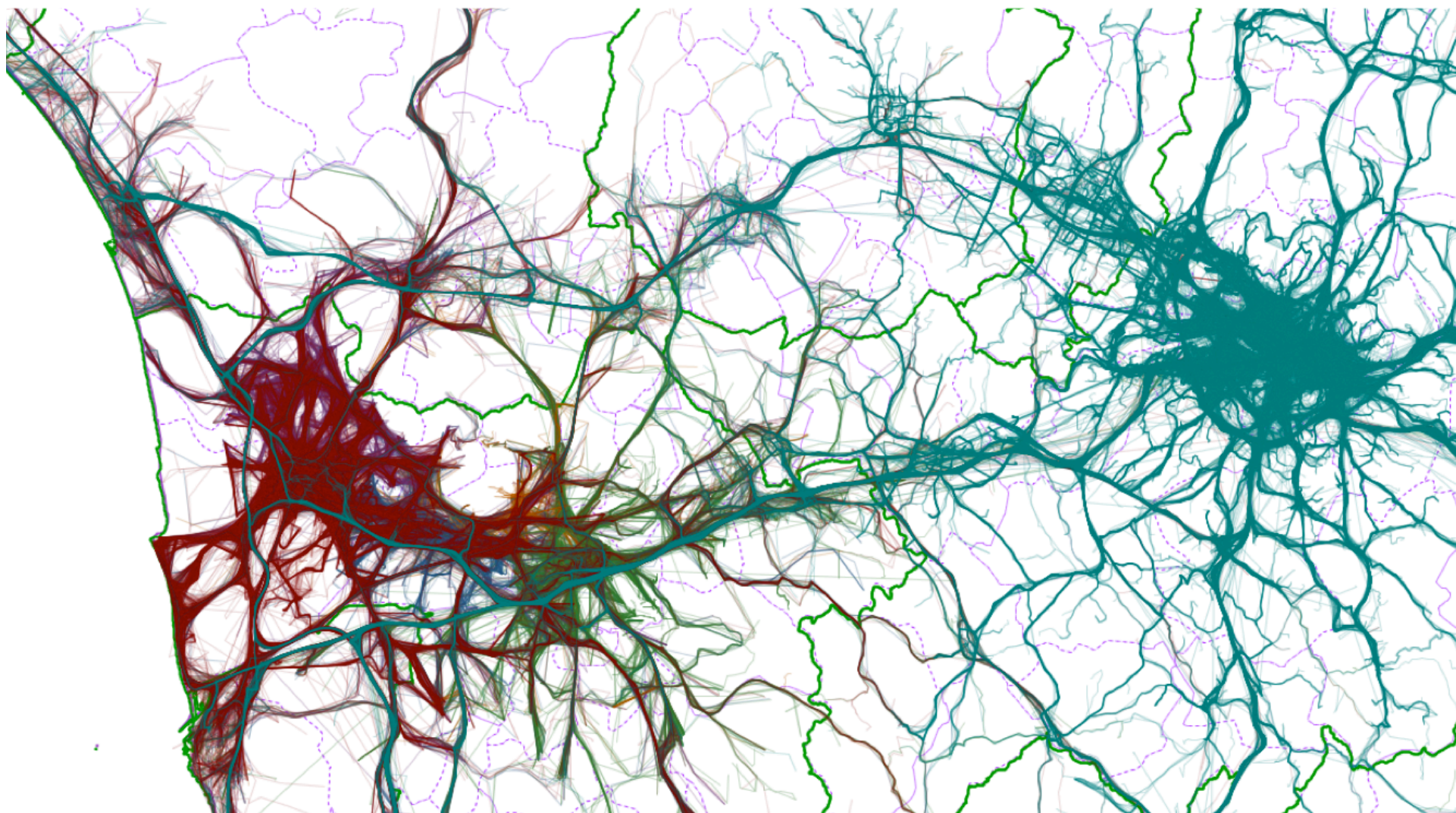
DESIRES, OPINIONS, SENTIMENTS

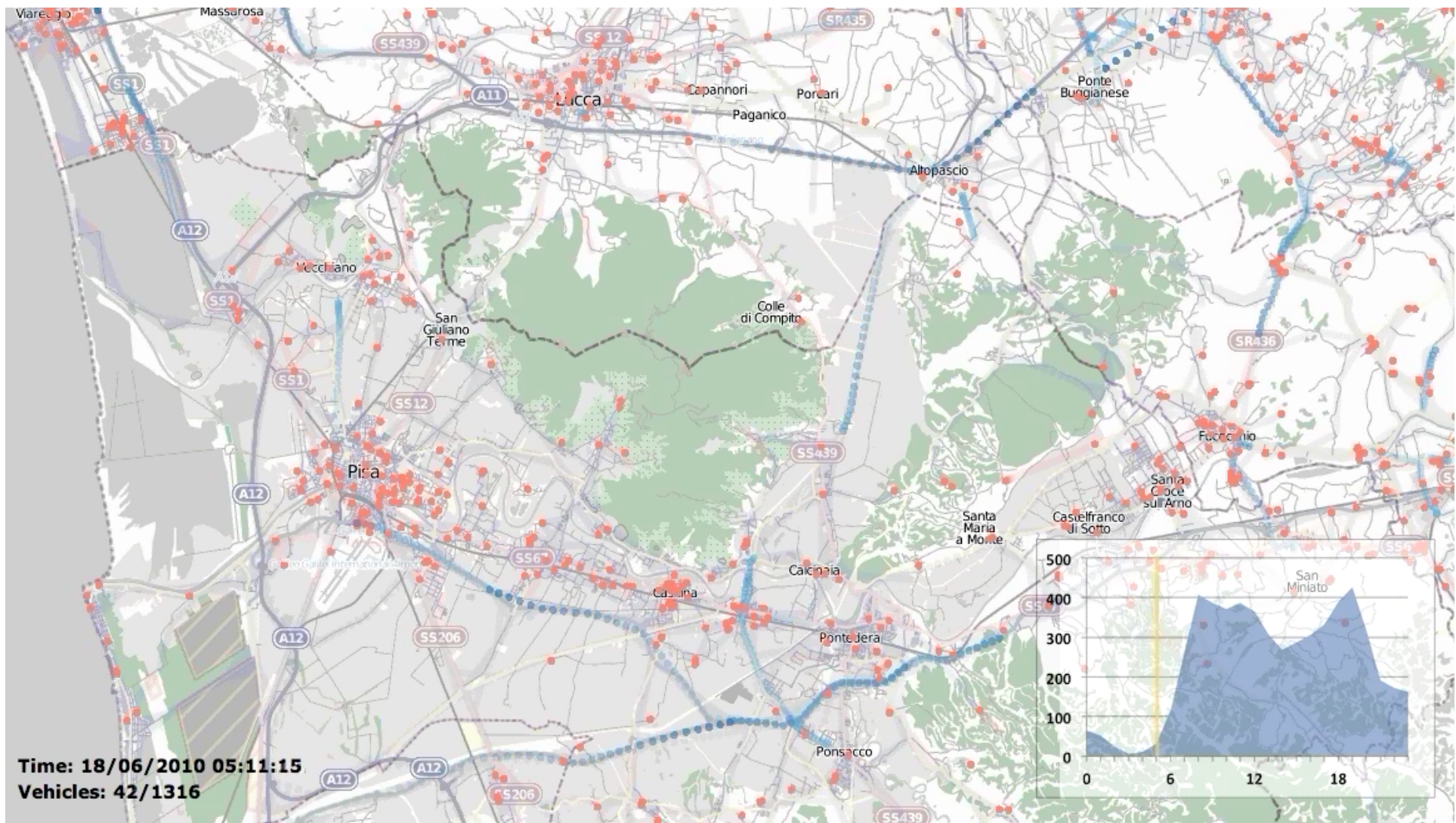


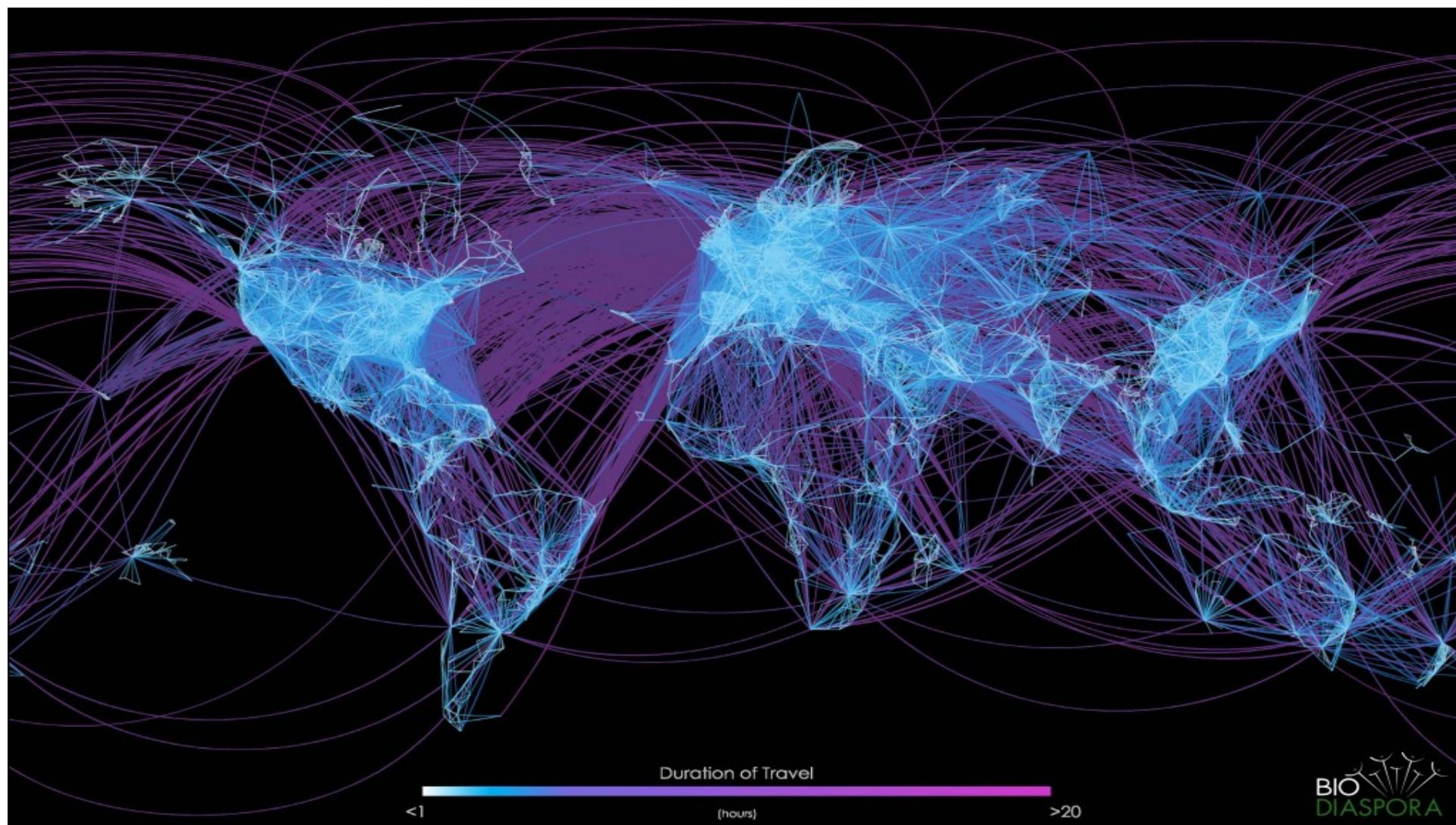
MOVEMENTS



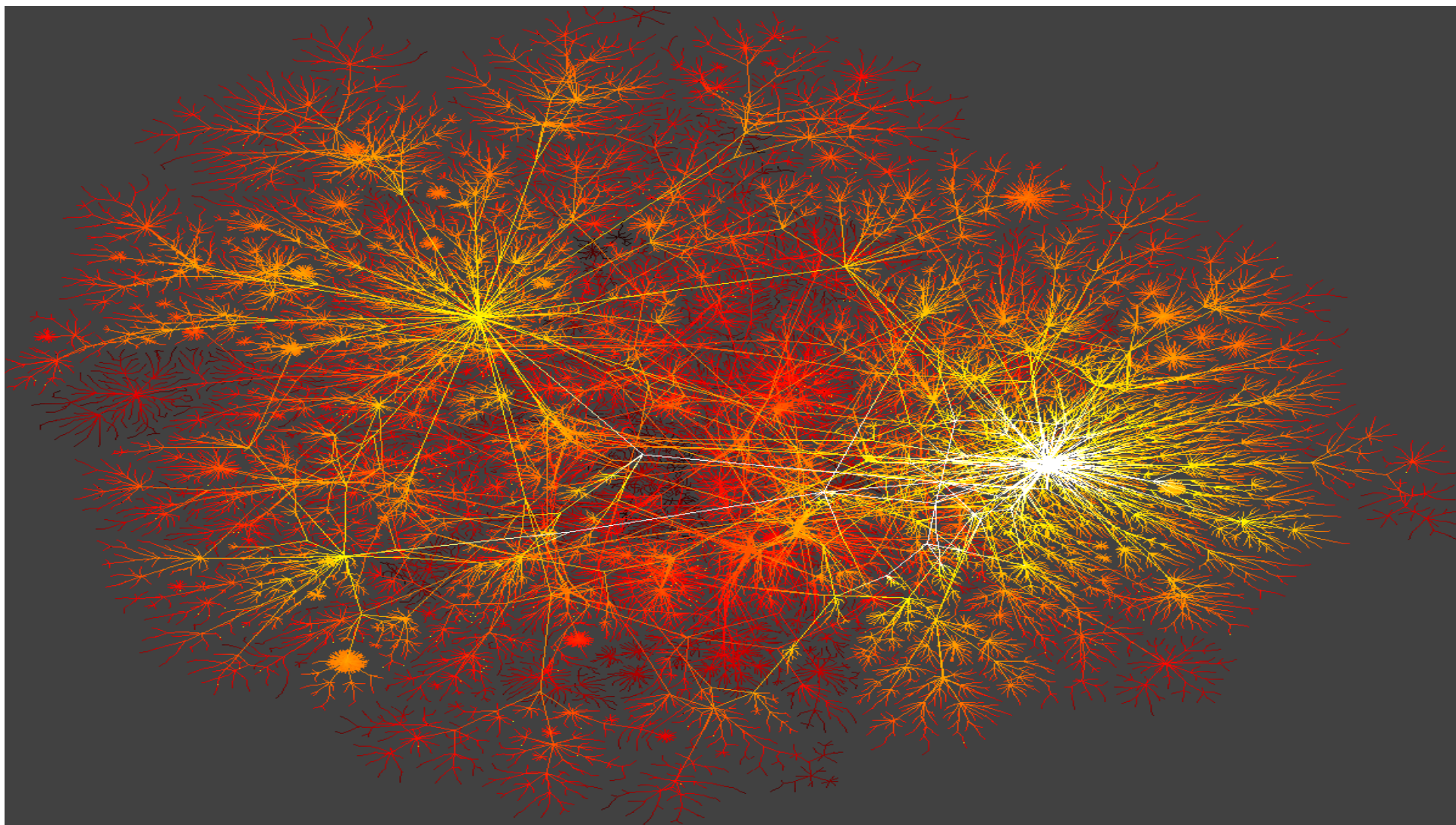


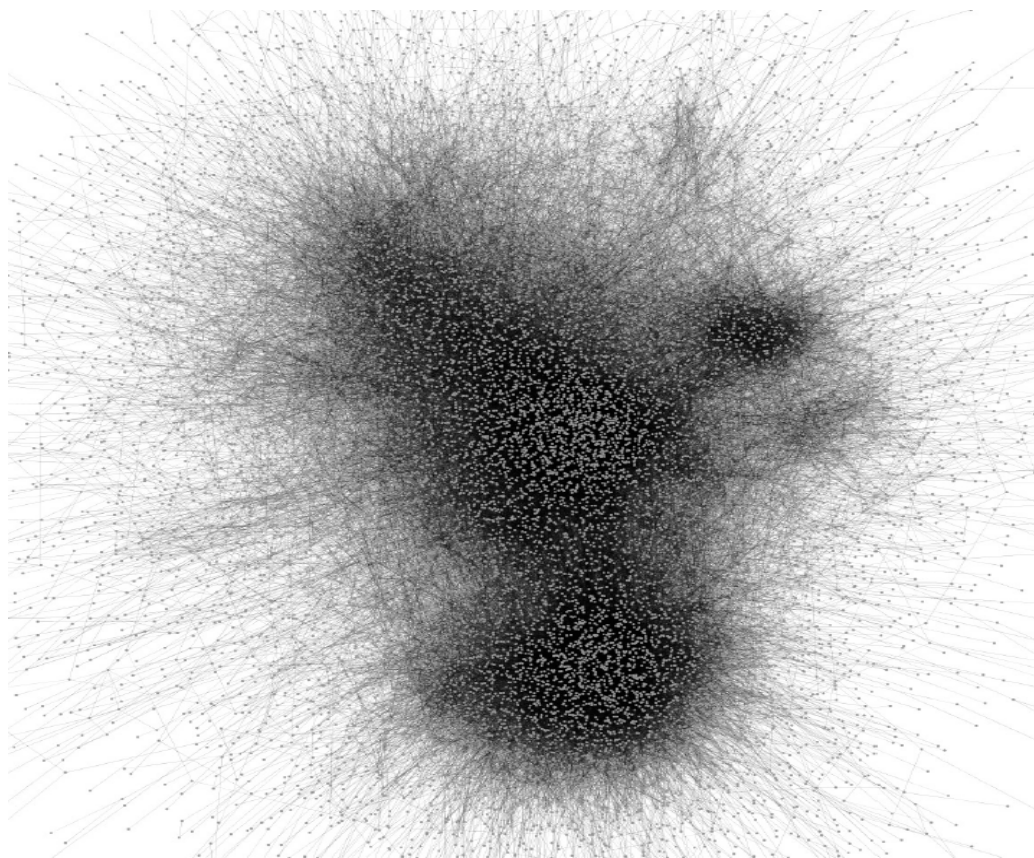


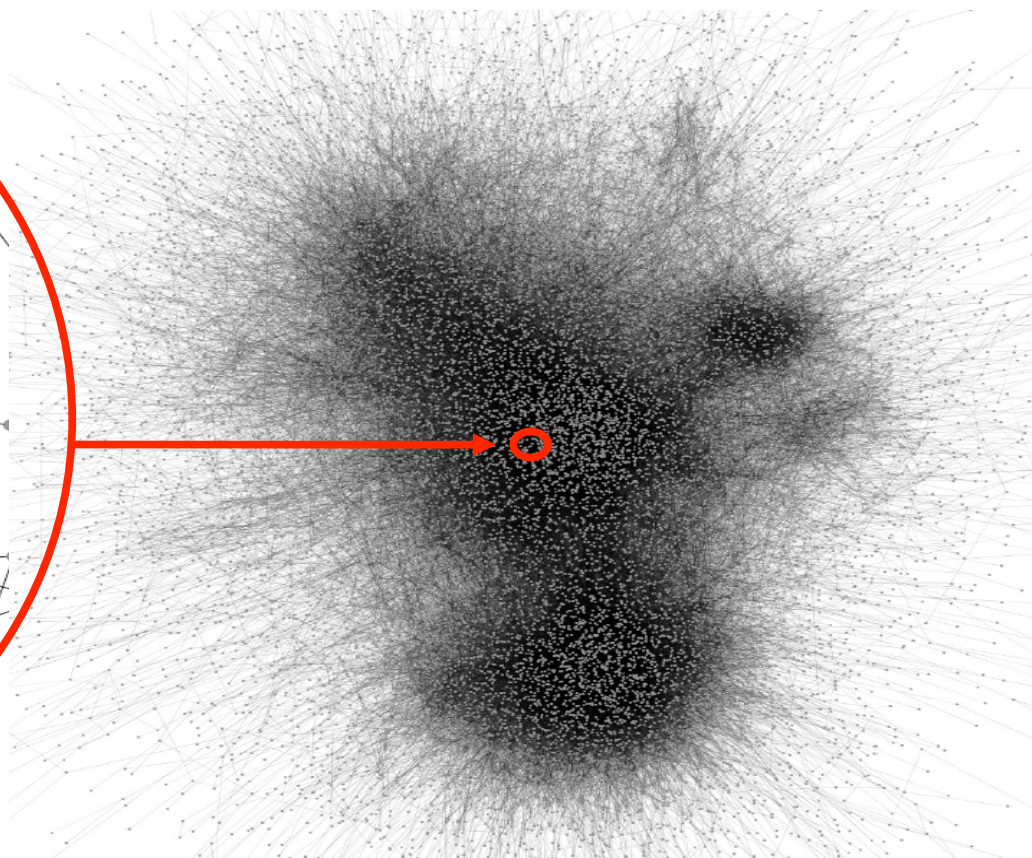
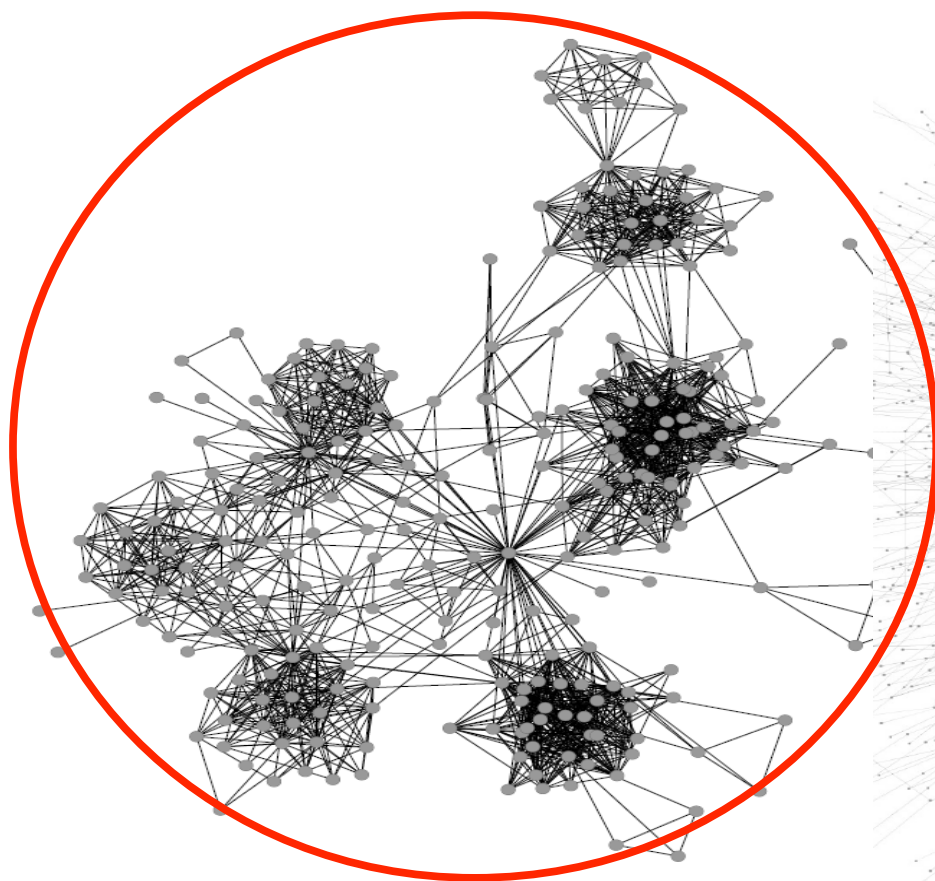








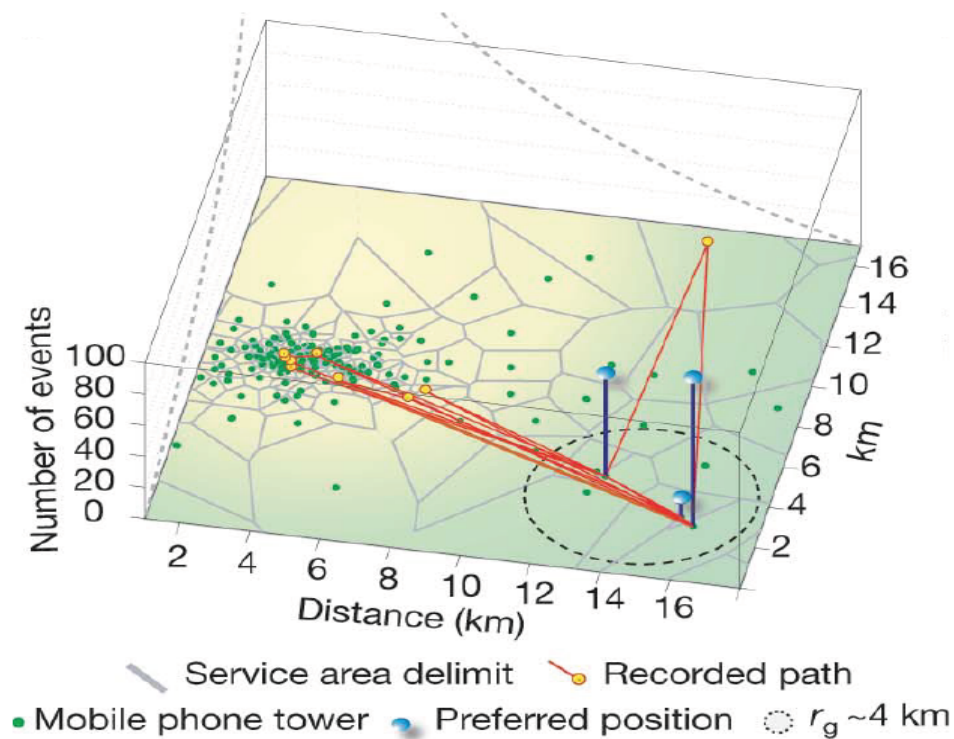




SOCIAL MINING

MAKING SENSE OF BIG DATA

FOCUS ON COUNTRY-WIDE CDR DATA



when
you
call

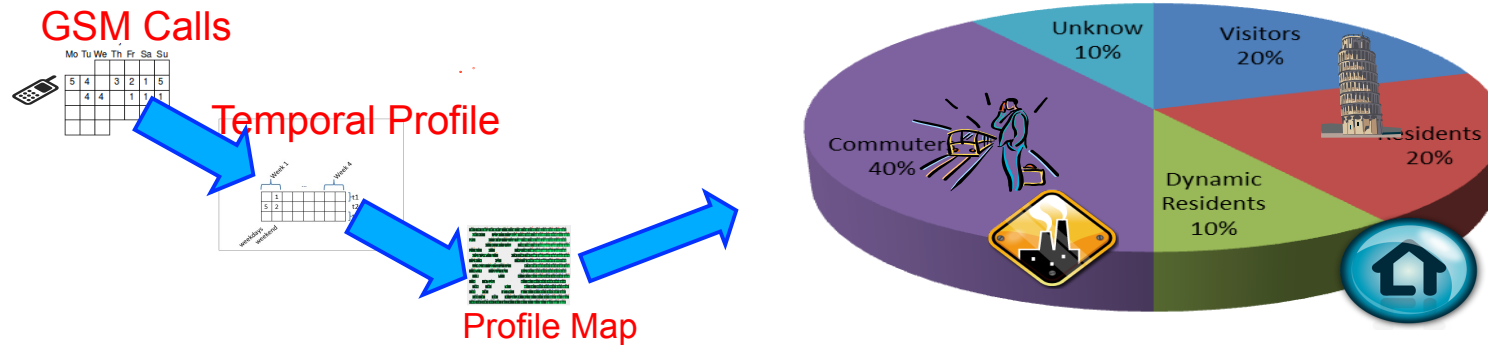


where
you
call



who
you
call

A SOCIOMETER BASED ON MOBILE PHONE DATA FOR REAL TIME DEMOGRAPHICS



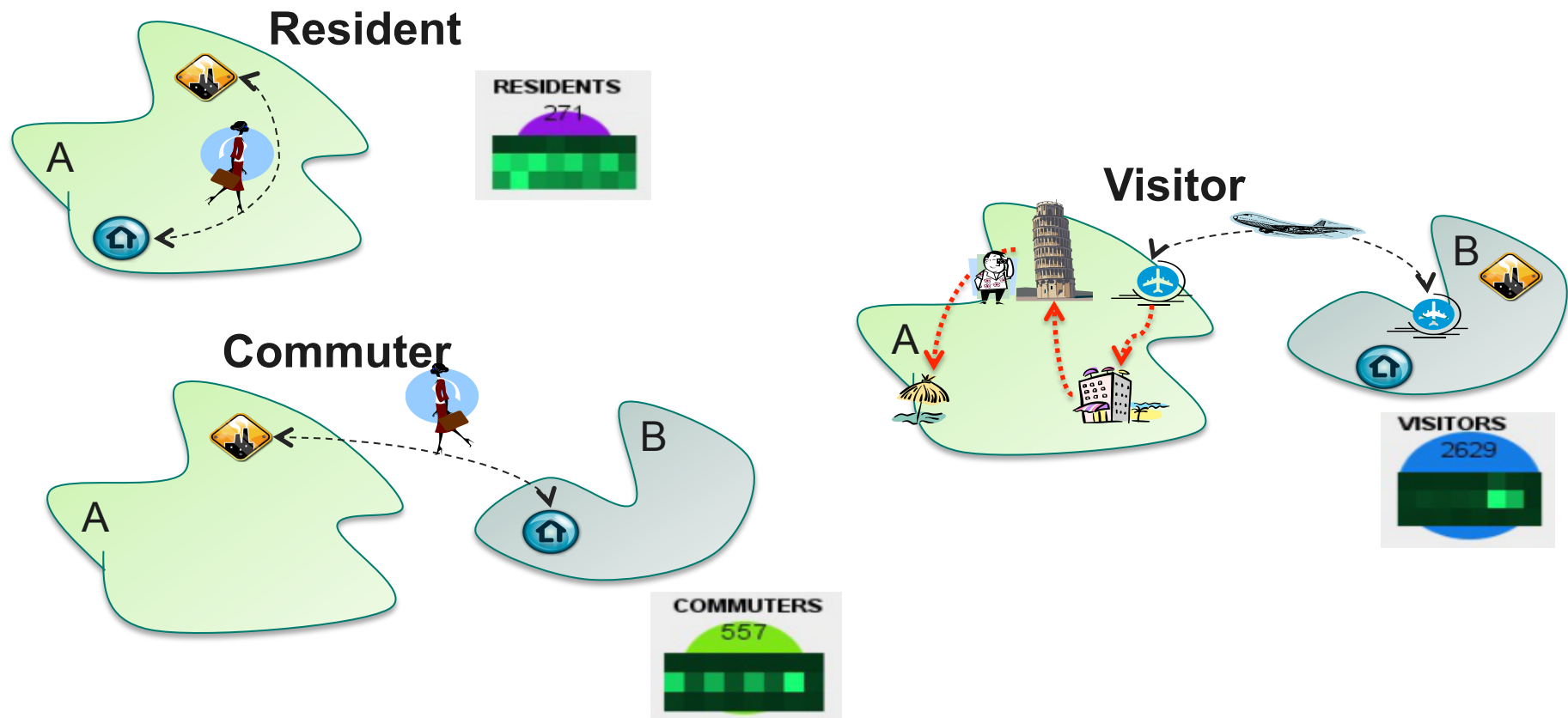
ISTITUTO DI SCIENZA E TECNOLOGIE
DELL'INFORMAZIONE "A. FAEDO"



UNIVERSITÀ DI PISA

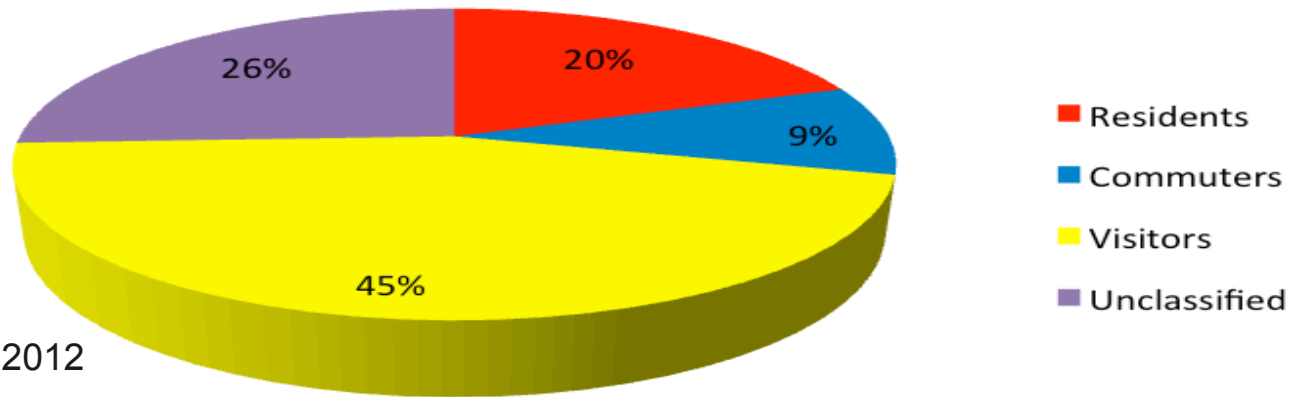


CALLING PROFILES



URBAN SOCIOMETER: THE CITY USER METER

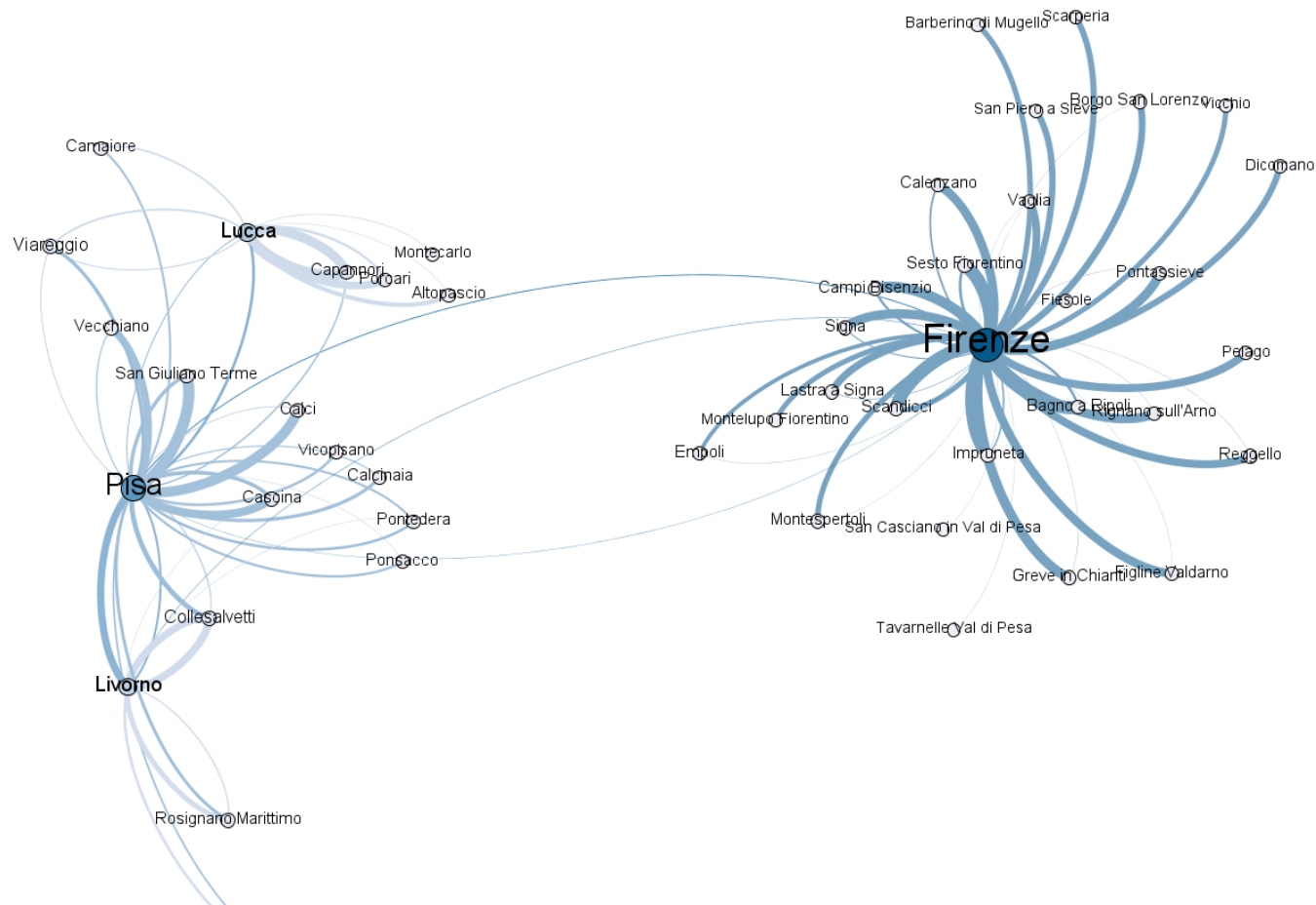
Classification outcome



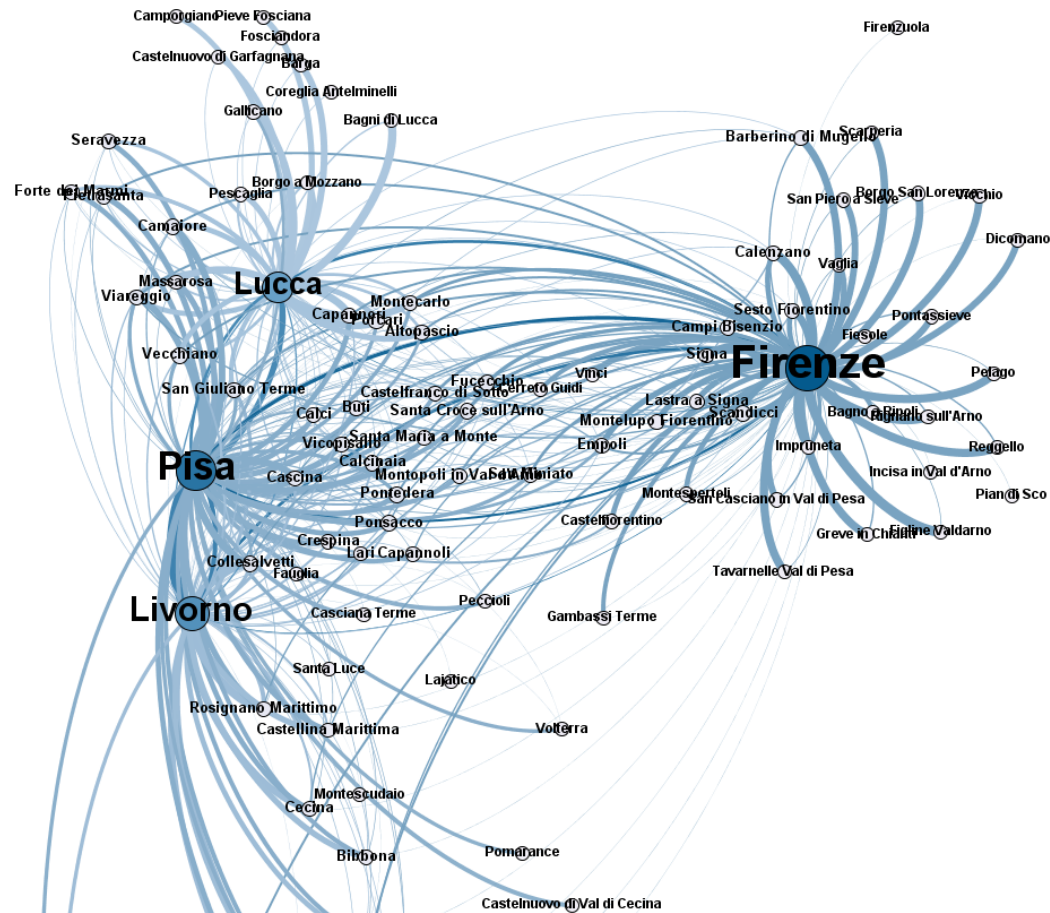
Pisa, January 2012

Analysis of GSM calls data for understanding user mobility behavior
B Furletti, L Gabrielli, C Renso, S Rinzivillo
Big Data, 2013 IEEE International Conference on, 550-555

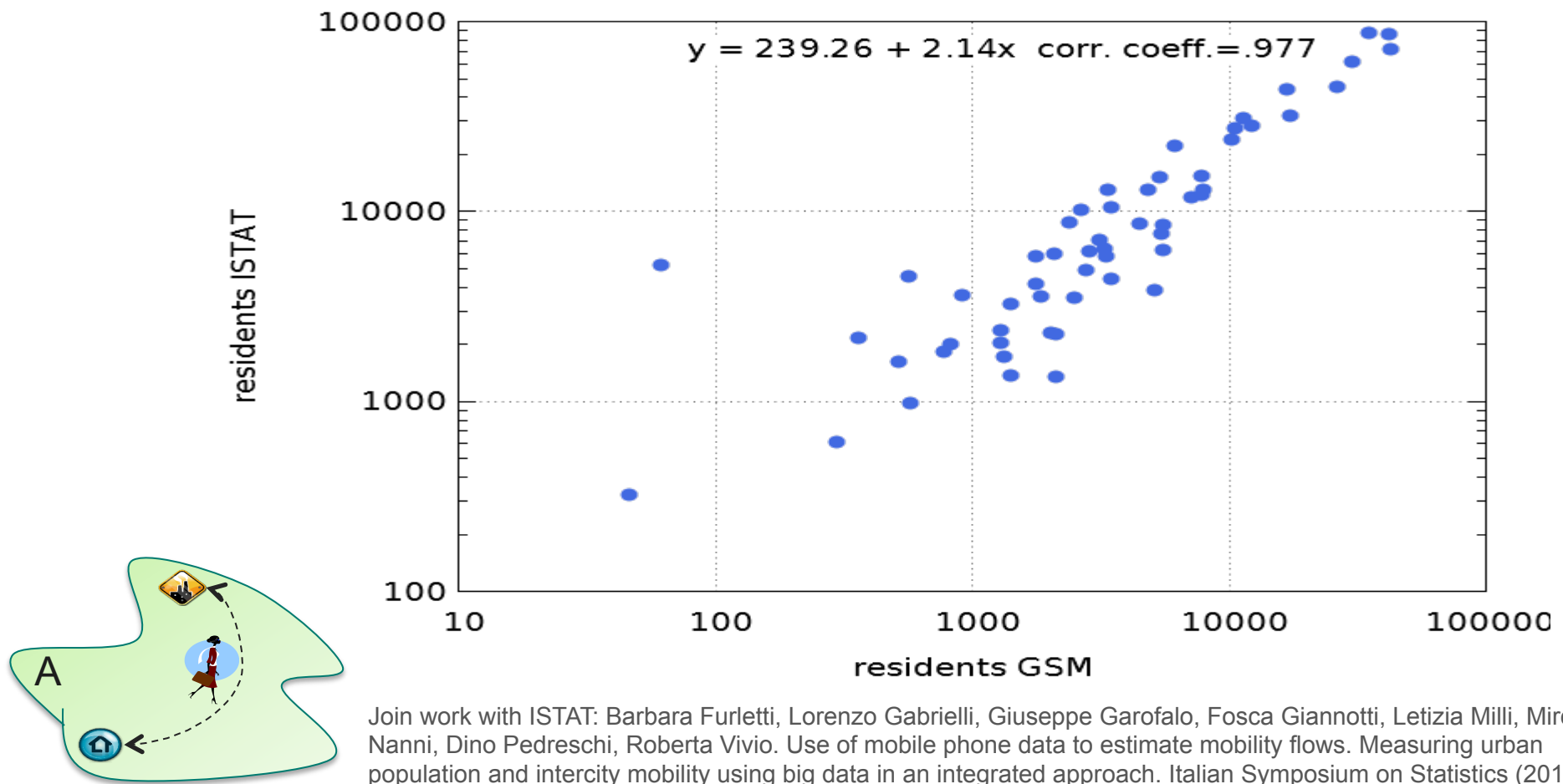
COMMUTER IN-OUT FLOWS



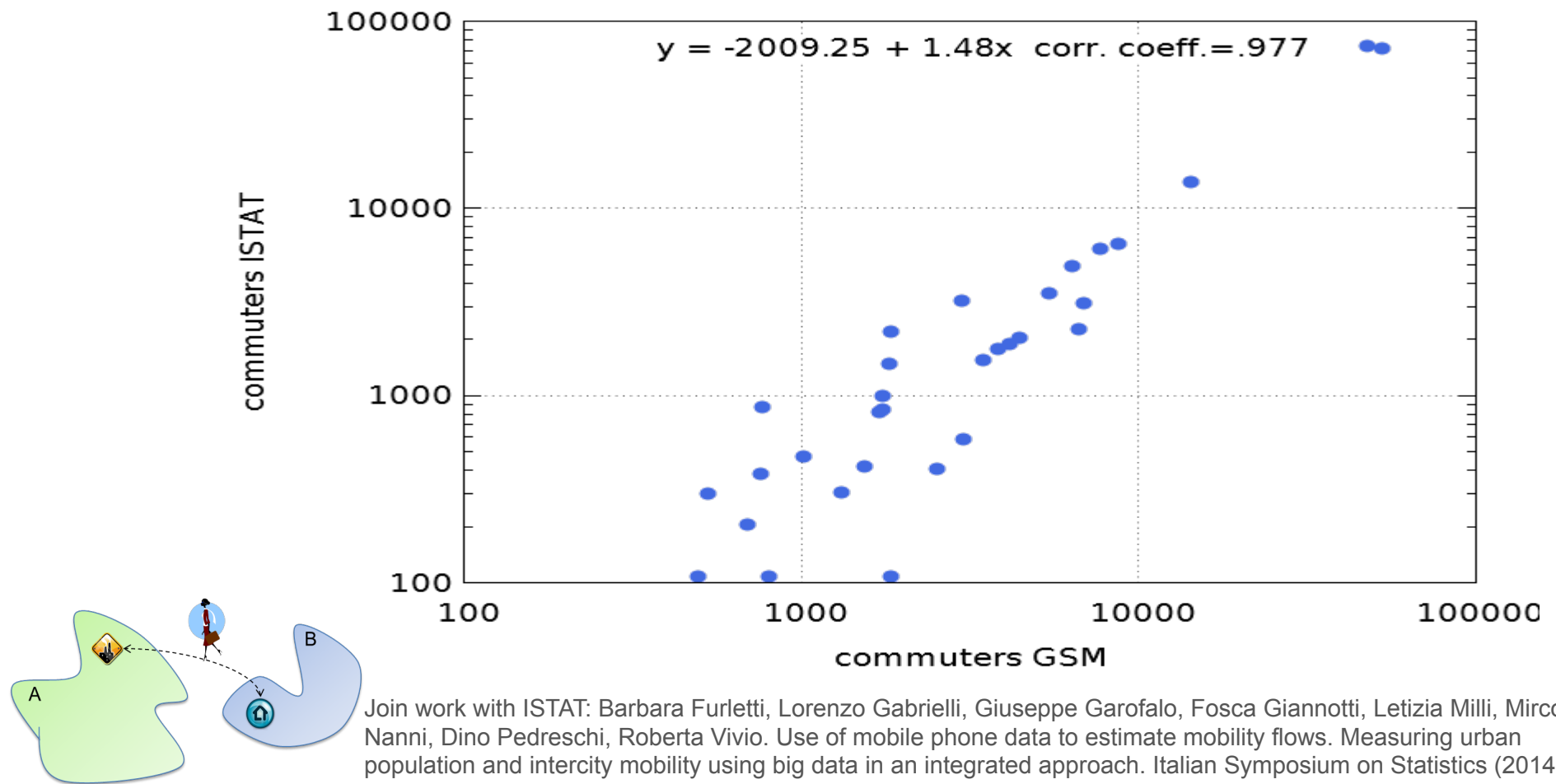
VISITOR IN-OUT FLOWS



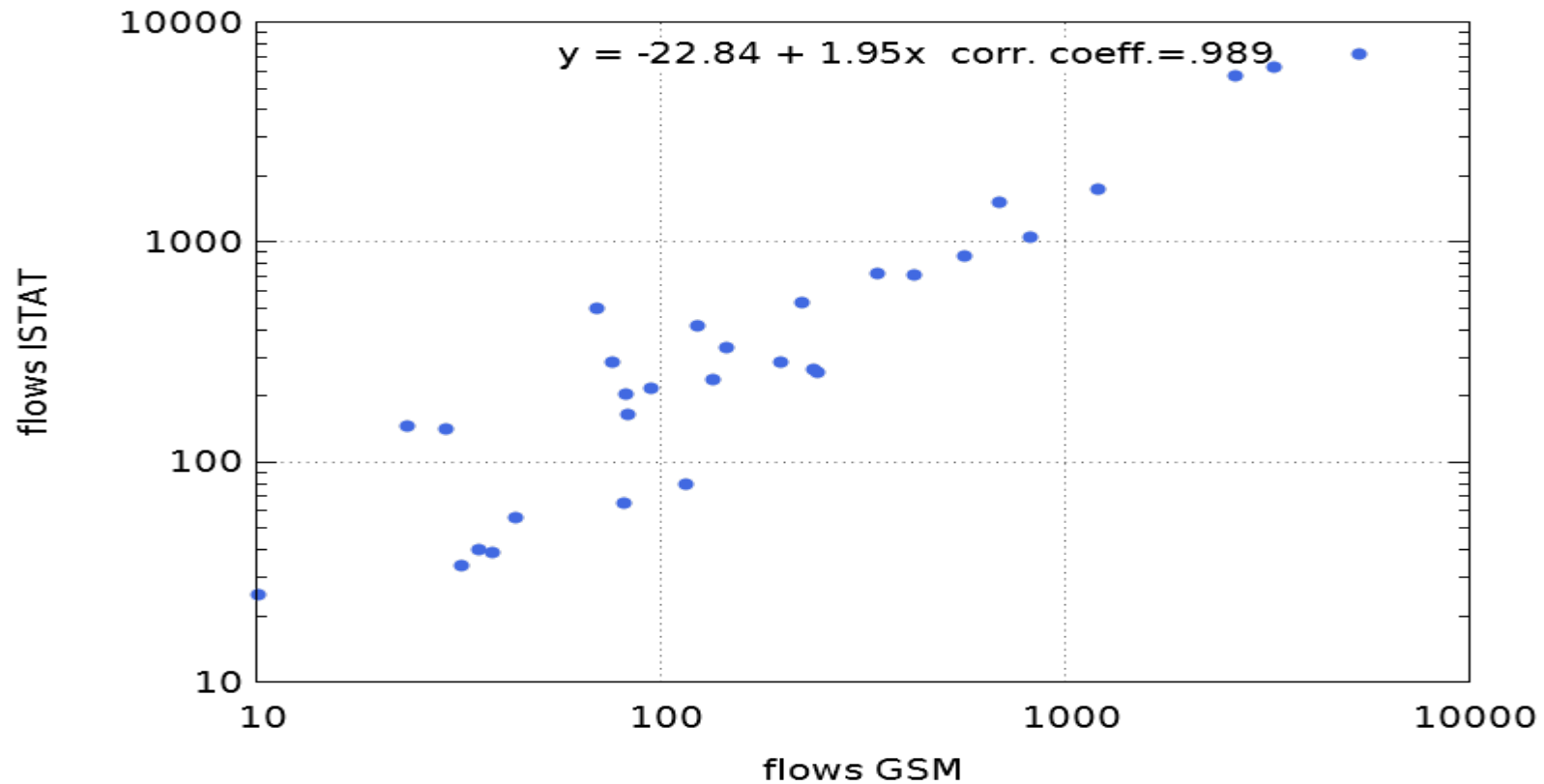
RESIDENTS – VALIDATION WITH ADMINISTRATIVE DATA



COMMUTERS - VALIDATION WITH ADMINISTRATIVE DATA



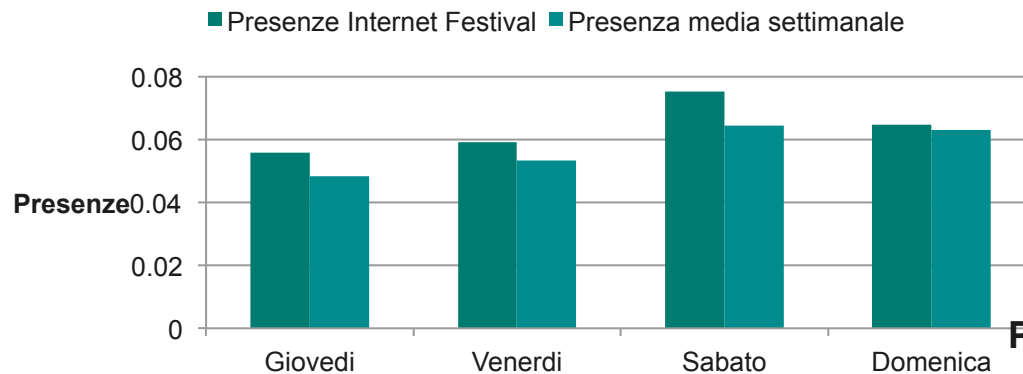
INTER-CITY FLOWS – VALIDATION WITH ISTAT DATA



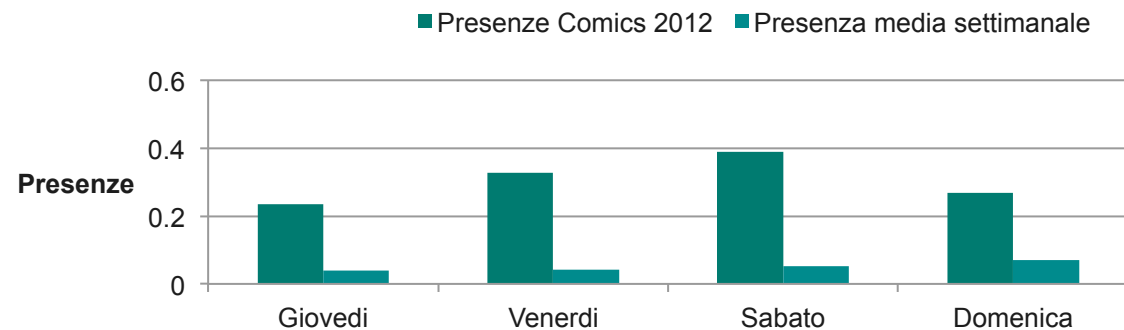
Join work with ISTAT: Barbara Furletti, Lorenzo Gabrielli, Giuseppe Garofalo, Fosca Giannotti, Letizia Milli, Mirco Nanni, Dino Pedreschi, Roberta Vivio. Use of mobile phone data to estimate mobility flows. Measuring urban population and intercity mobility using big data in an integrated approach. Italian Symposium on Statistics (2014).

MEASURING EXCEPTIONAL EVENTS

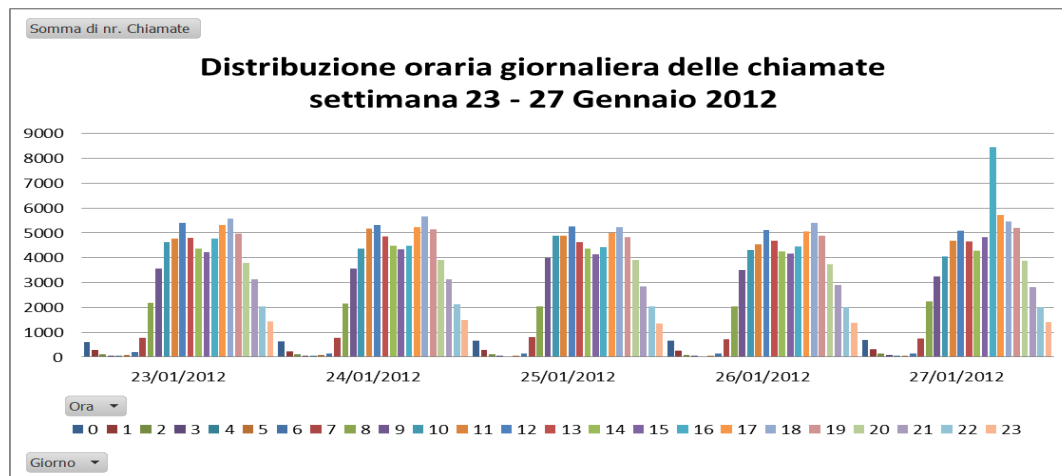
Presence of Visitors GSM - Pisa - Historical Center



Presence of Visitors GSM - Lucca Comics 2012



EXTRAORDINARY EVENTS



LA NAZIONE

Firenze / Arezzo / Empoli / Grosseto / La Spezia / Livorno / Lucca / Massa Carrara / Montecatini / Pisa / Pontedera / Piombino

HOME SPORT MOTORI MAGAZINE LIFESTYLE SPETTACOLO TECNOLOGIA BLOG MULTIMEDIA

Home Toscana Cinema

Homepage > Toscana > Terremoto: trema tutta la Toscana Massa, crolla il tetto in chiesa Miracolosamente illeso il

Terremoto: trema tutta la Toscana Massa, crolla il tetto in chiesa Miracolosamente illeso due fedeli

Evacuato l'Ateneo di Pisa, telefoni in tilt **Terremoto a Massa: buco nel tetto in chiesa** Commenti

L'epicentro del sisma di **magnitudo 5.4** (scala Richter) sull'Appennino toscano-emiliano. La scossa è stata avvertita soprattutto in Lunigiana e sulla costa nord della regione

Facebook 72 Twitter 1

Email Stampa



Sommario Terremoto con epicentro nel parmense, scosse avvertite in tutta la Toscana

ARTICOLI CORRELATI

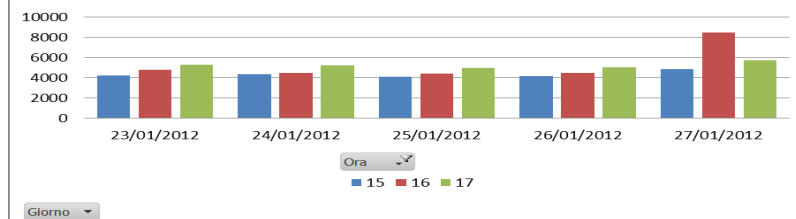
Il terremoto a Empoli

Hai sentito il terremoto? Il questionario INGV

L'allarme dei geologi: "Il 60% delle case è a rischio sismico"

Toscana, 27 gennaio 2012 - Paura in Toscana per la scossa di terremoto di magnitudo 5.4 della scala Richter, avvertita in tutto il Nord Italia intorno alle 15 e 53, con epicentro nella zona dell'Appennino toscano-emiliano. Dicono i costruttori

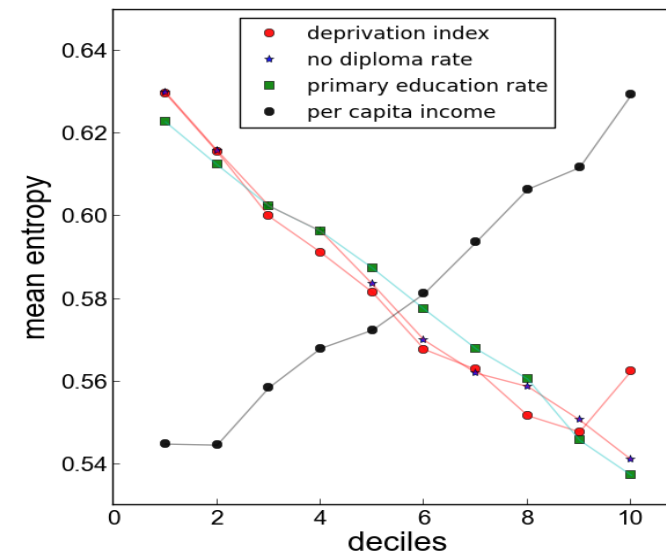
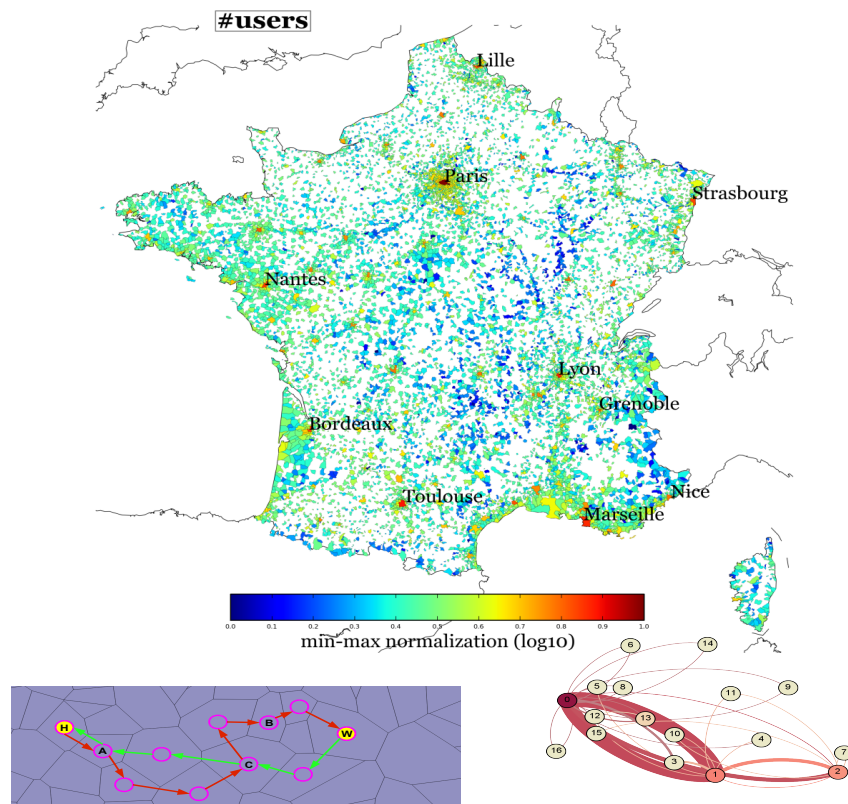
settimana 23 - 27 Gennaio 2012



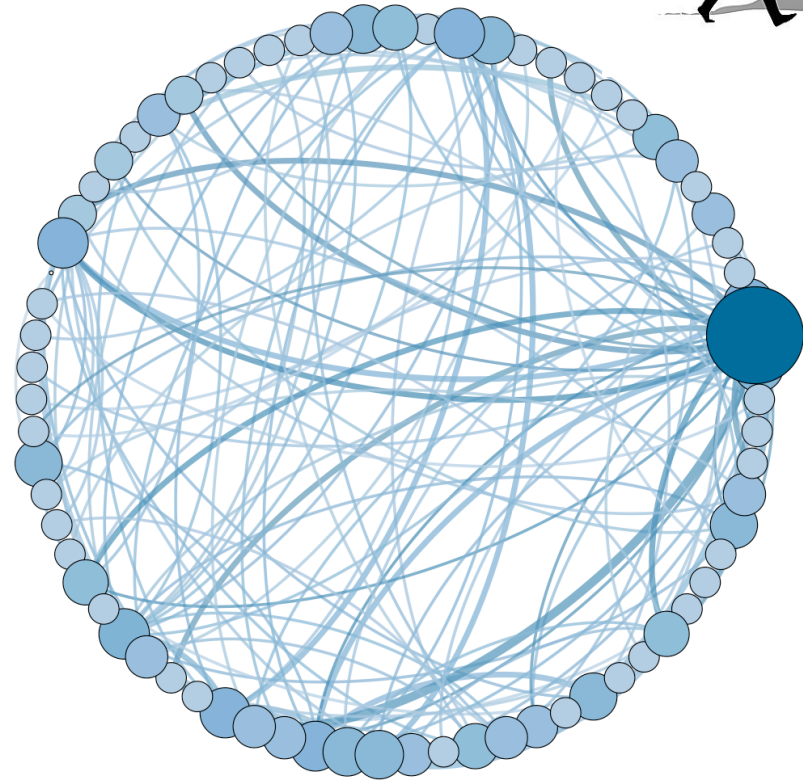
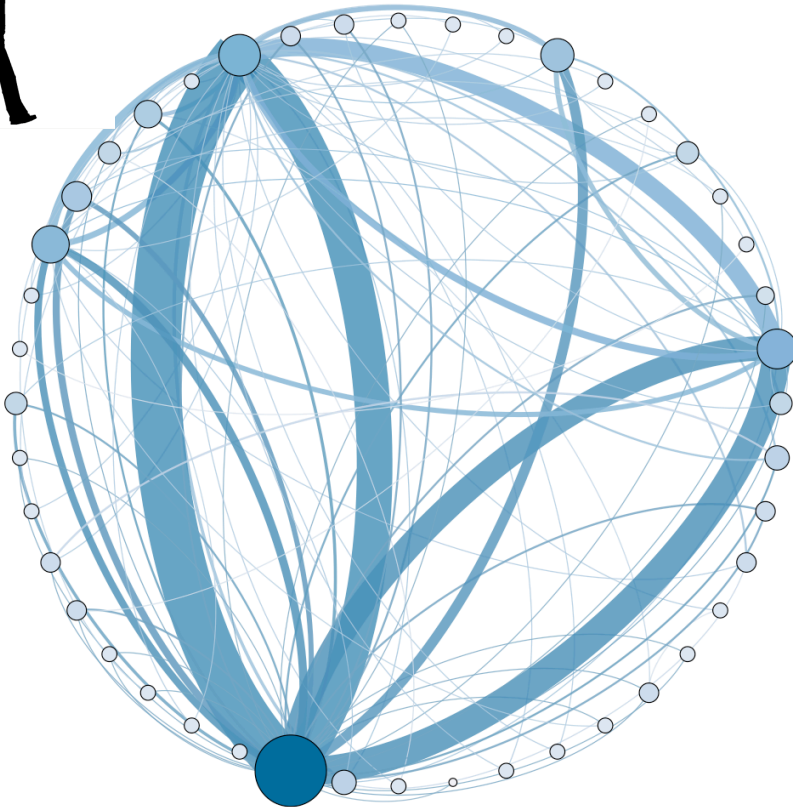
BIG DATA FOR DEVELOPING COUNTRIES

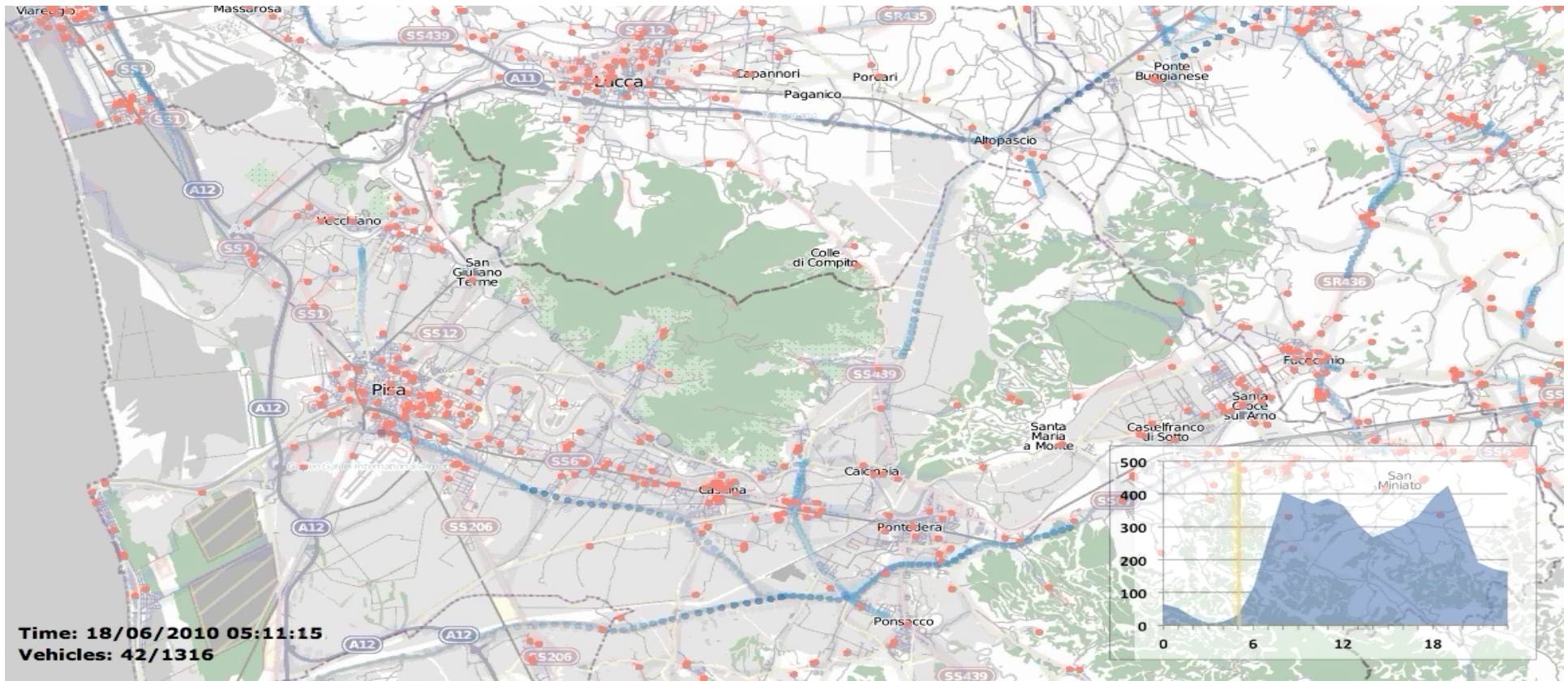


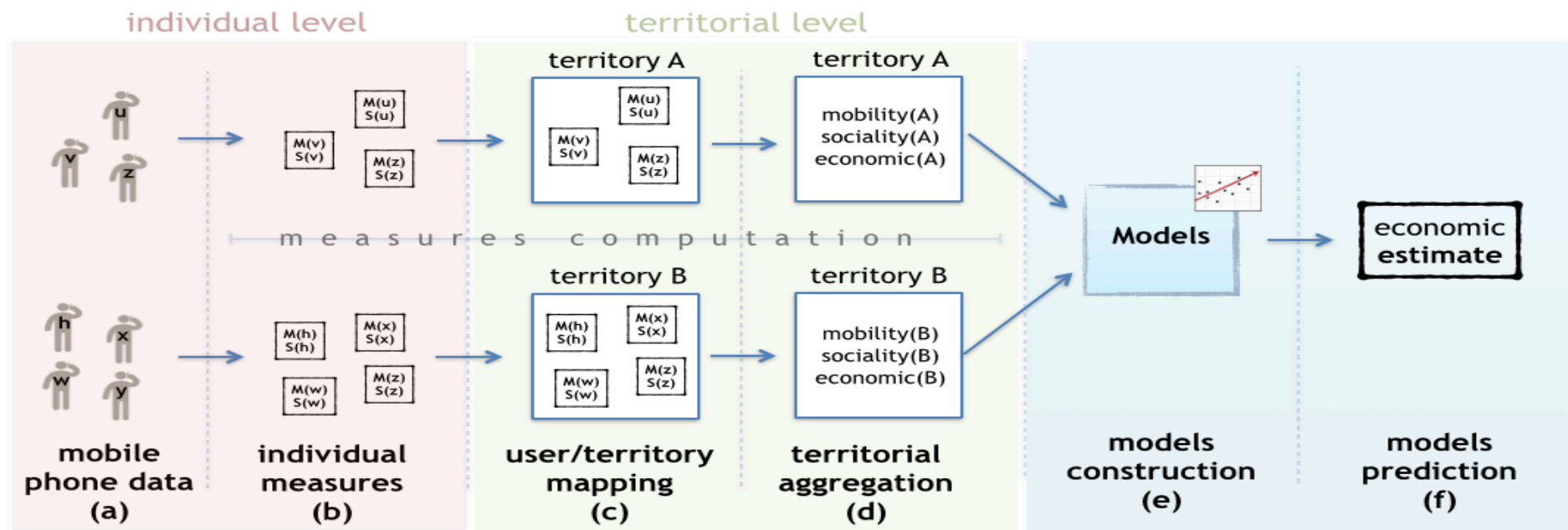
BIG DATA: DIVERSITY AND ECONOMIC DEVELOPMENT

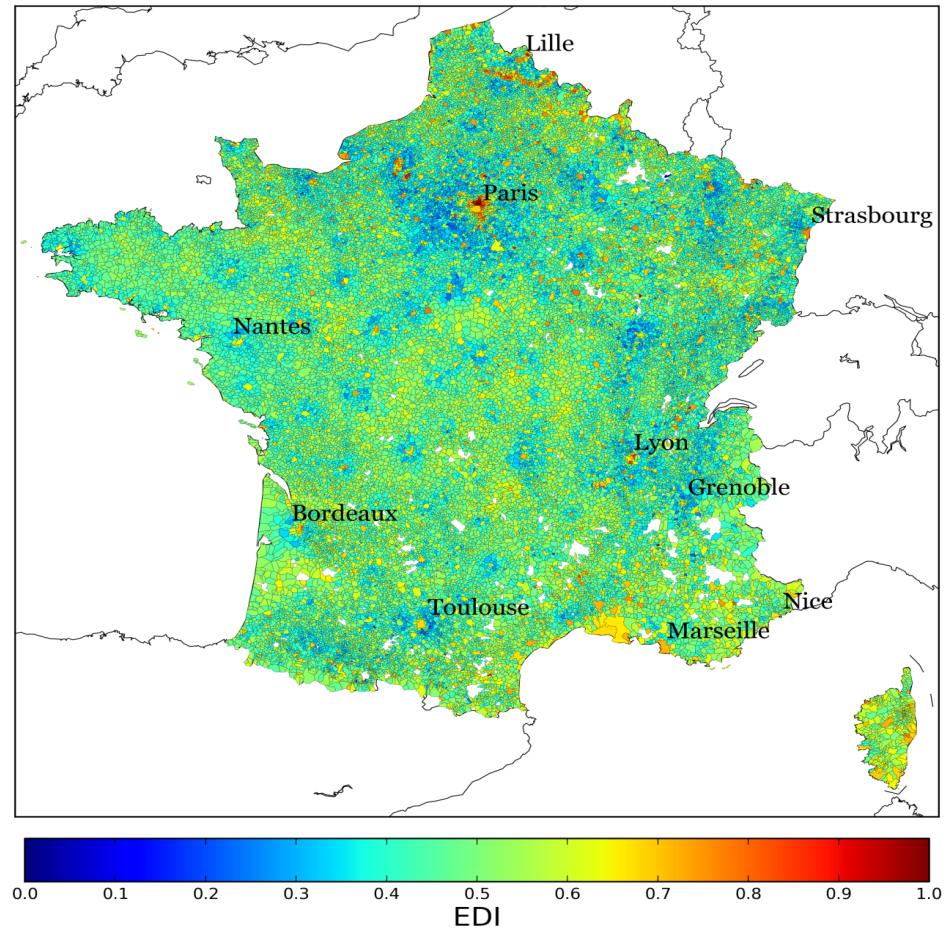


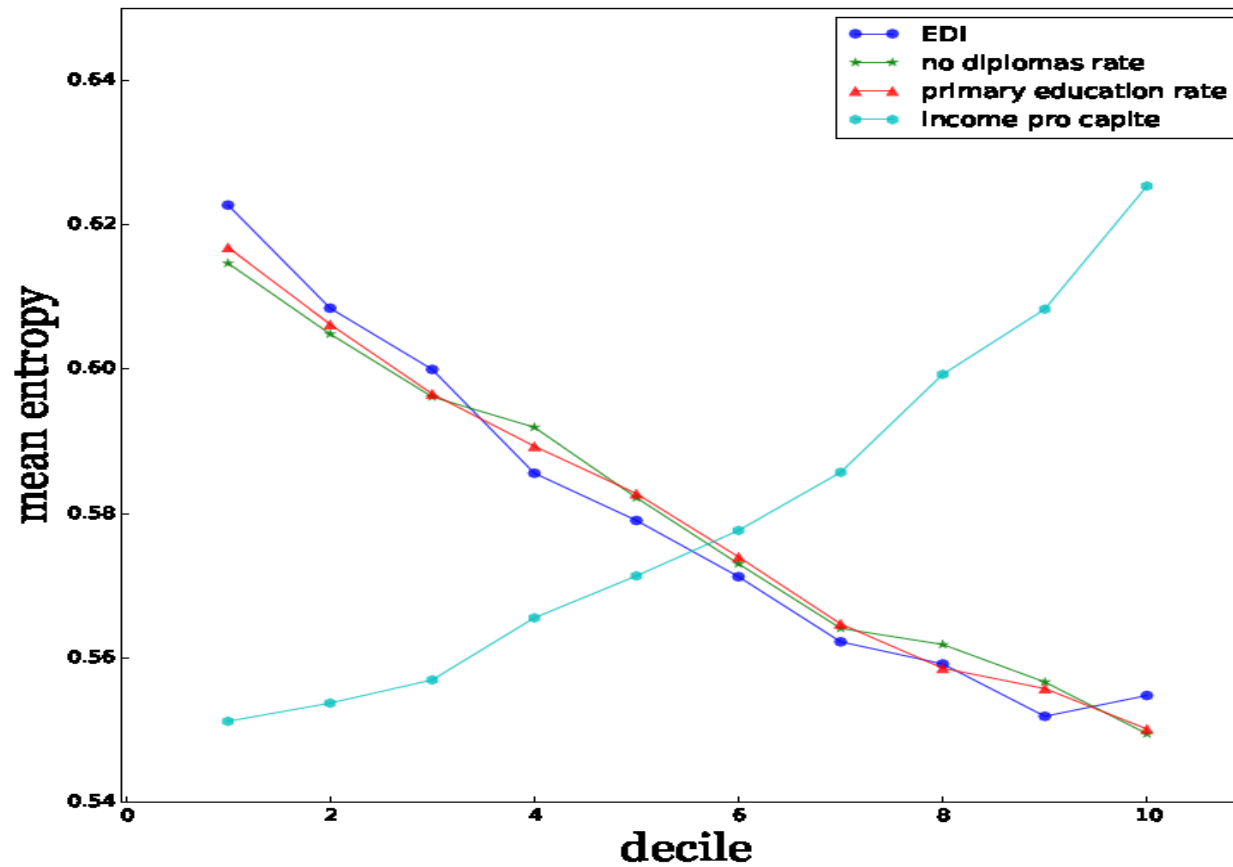
DIVERSITY OF INDIVIDUAL MOBILITY NETWORKS





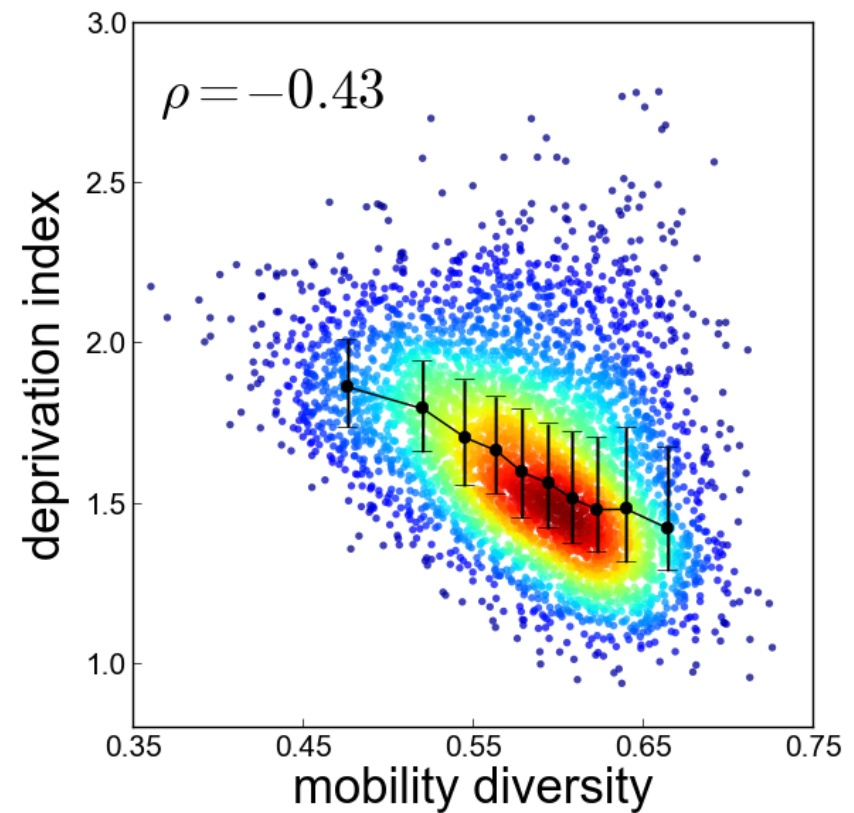
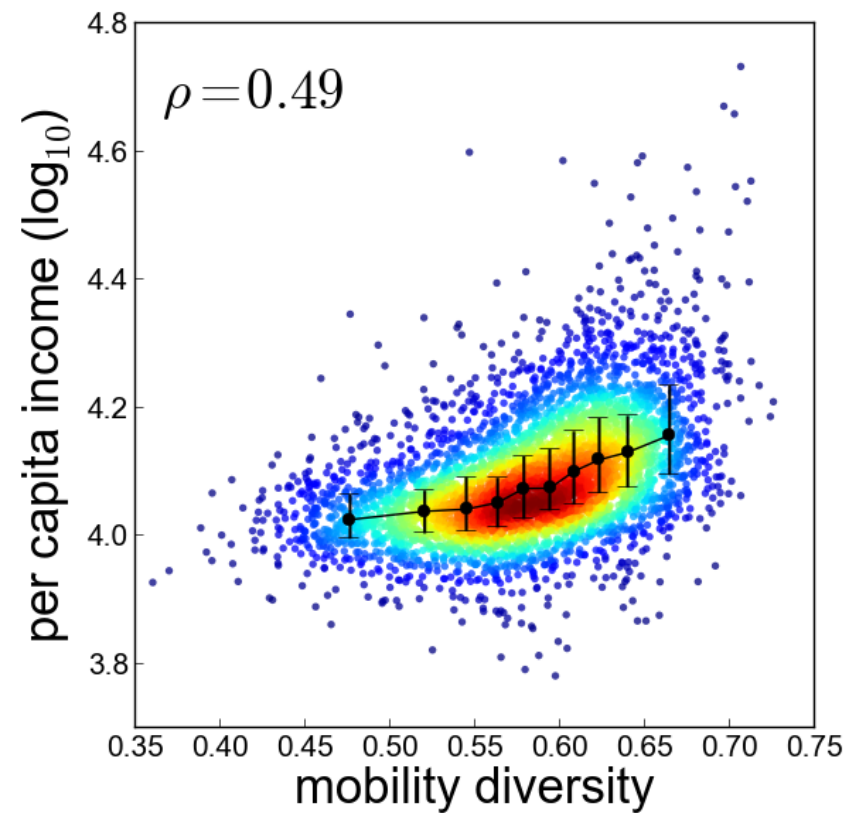




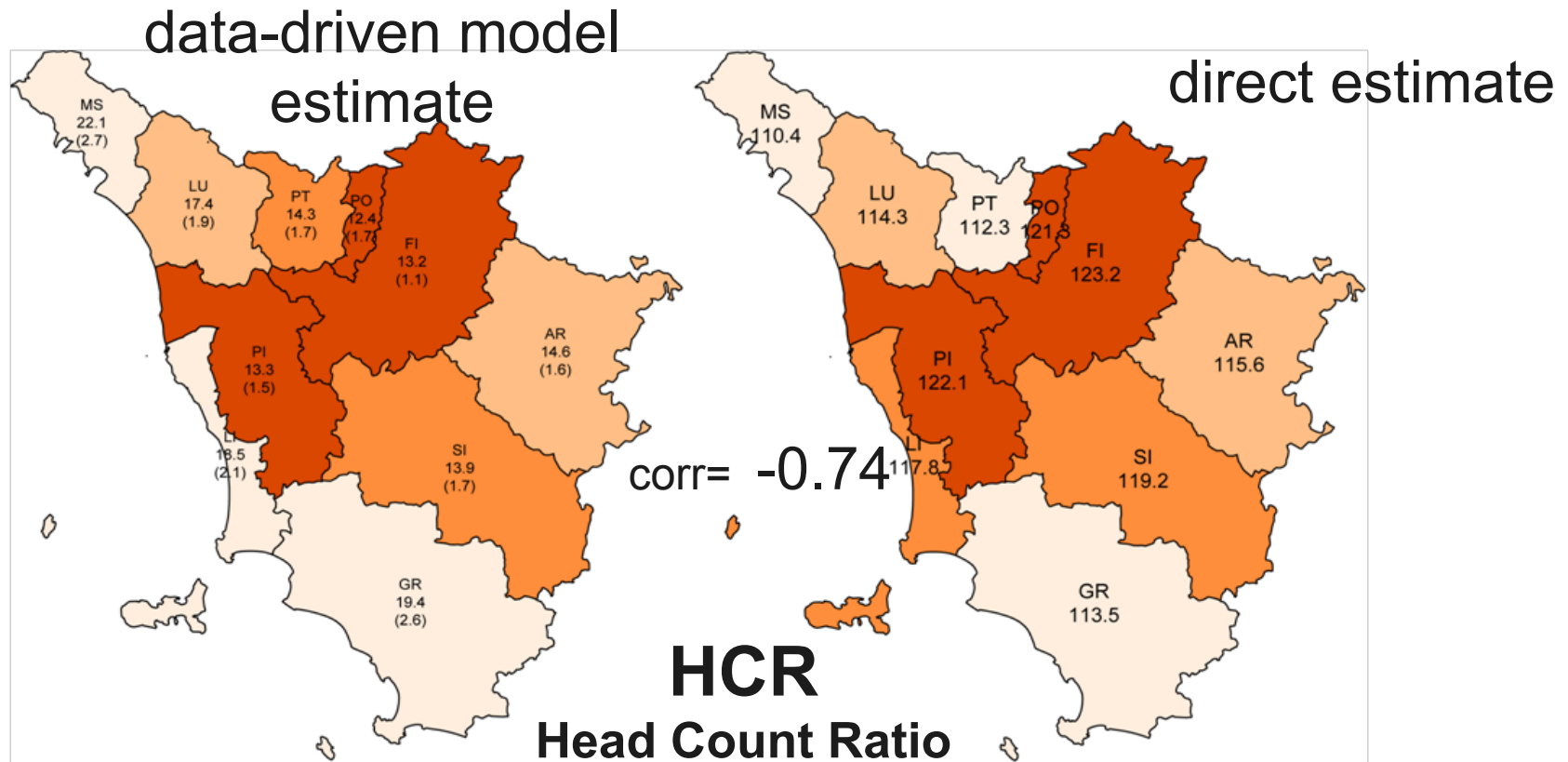


Dino Pedreschi. Big Data, Social Mining, Diversity, and Wellbeing. Italian Symposium on Statistics (2014).
Luca Pappalardo, Maarten Vanhoof, Zbigniew Smoreda, Dino Pedreschi, Fosca Giannotti. Human Mobility, Social Networks and Economic Development. Submitted (2015).

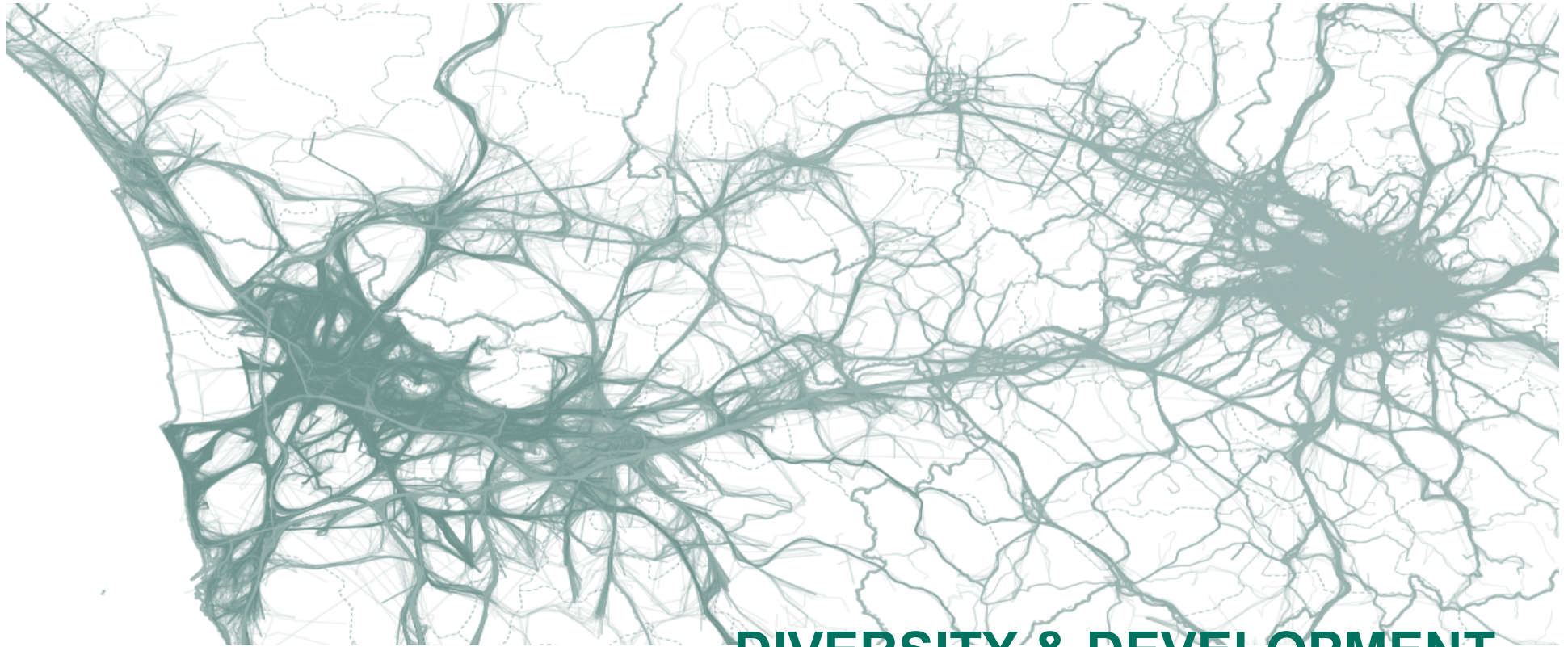
DIVERSITY AND DEVELOPMENT



...ALSO IN TUSCANY WITH GPS DATA



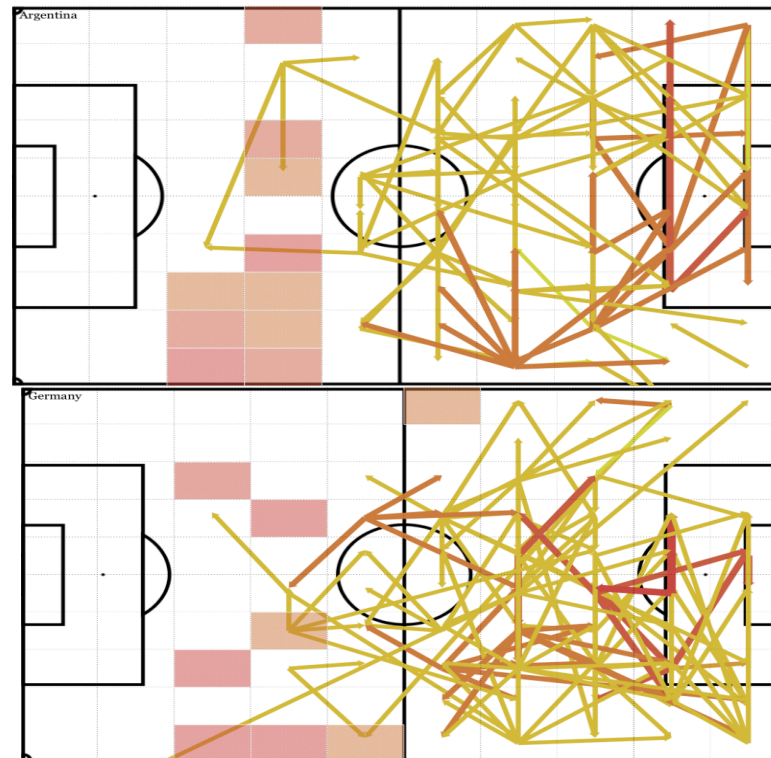
Giusti, Marchetti, Pratesi, Salvati, Pedreschi, Giannotti, Rinzivillo, Pappalardo, Gabrielli. **Small area model based estimators using Big Data Sources**. Journal of Official Statistics, vol. 31(2) 2015.



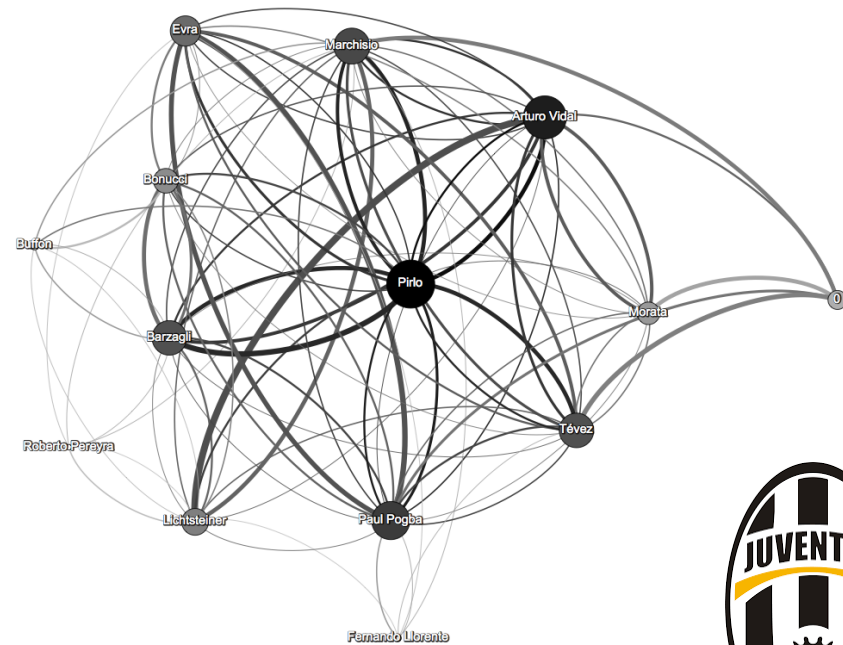
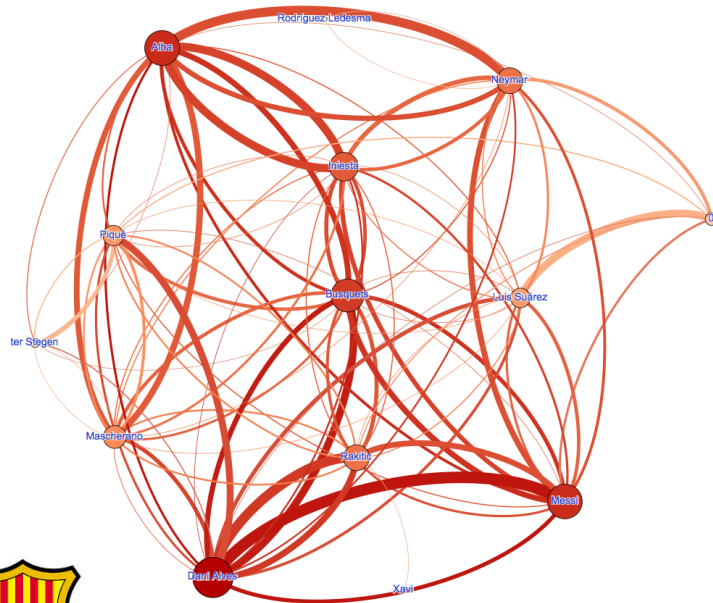
DIVERSITY & DEVELOPMENT

THE PATTERNS OF SUCCESS IN SPORTS

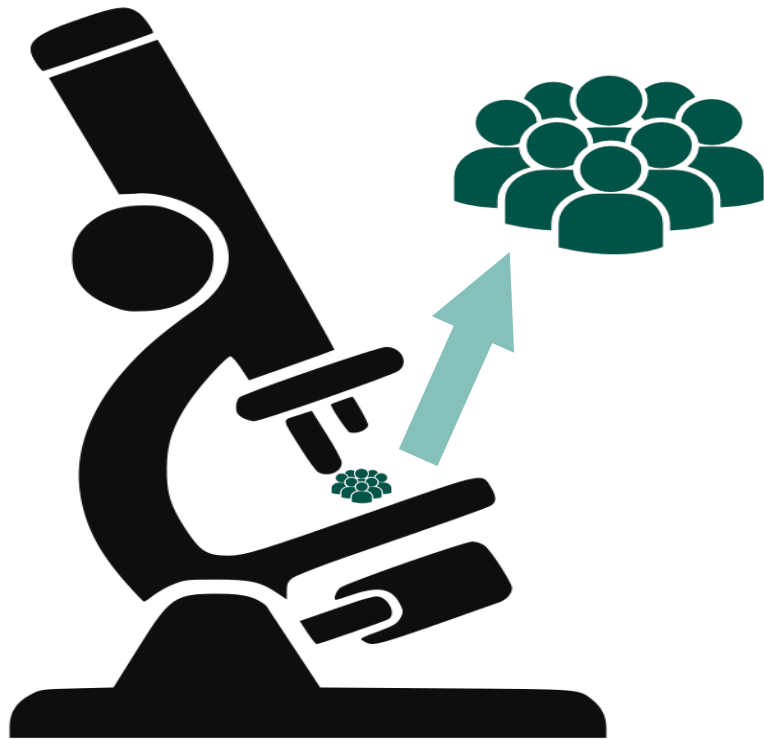
“Football is a simple game: 22 men chase a ball for 90 minutes and at the end, the Germans always win” -- Gary Lineker (after Italy 1990 Final)



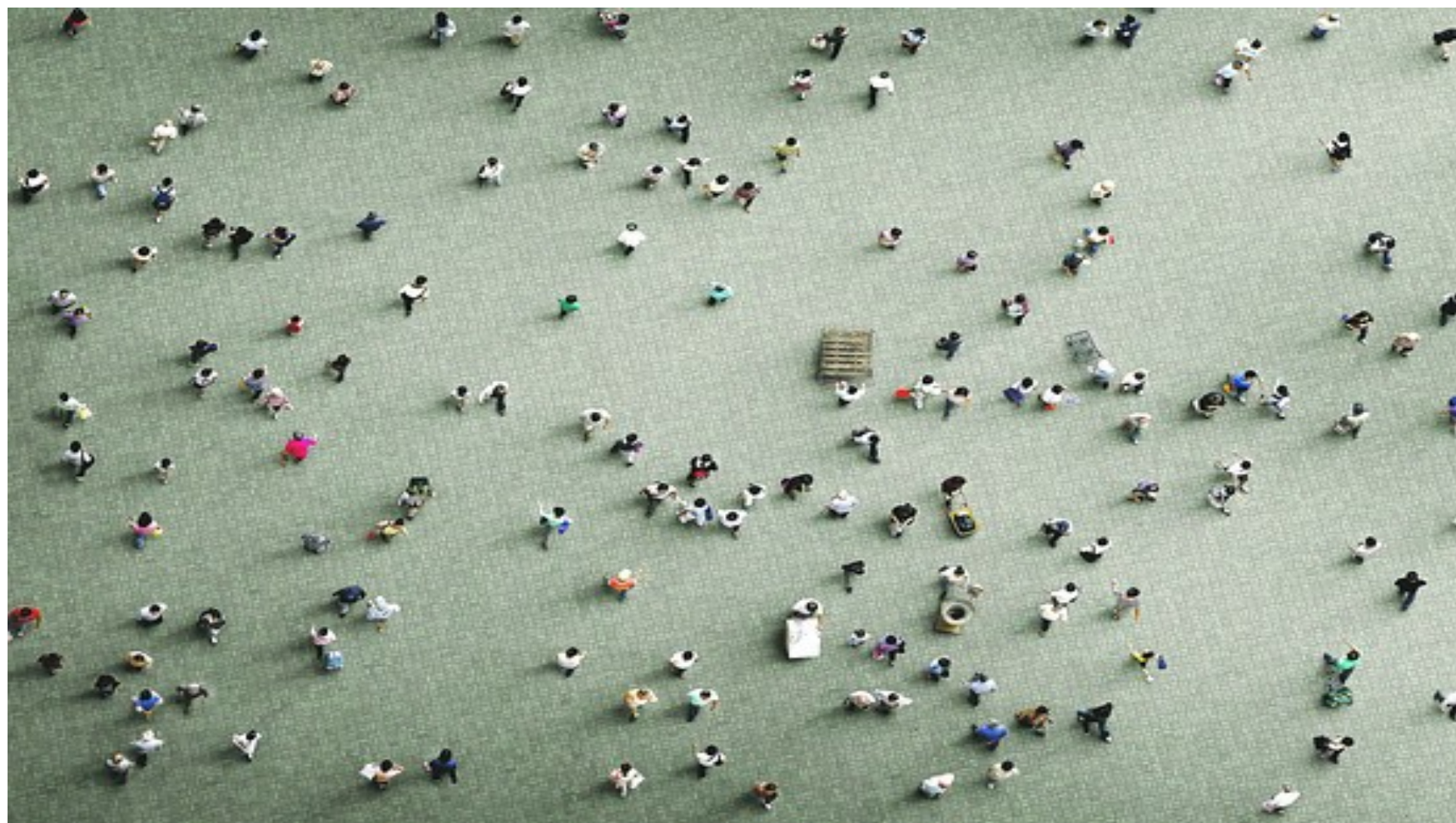
BIG DATA, DIVERSITY AND SPORT SUCCESS



Big Data Analytics & Social Mining



A TOOL TO
MEASURE,
UNDERSTAND,
AND POSSIBLY PREDICT
HUMAN BEHAVIOR



WHO'S GOT THE BENEFITS OF BIG DATA SO FAR?

A few **latifundists of data**

GAFA



- Profiling for behavioral advertising and target marketing

NSA



- Profiling for discovering potential threats to homeland security
- Mass surveillance

A world map where the landmasses are filled with a dense pattern of small red icons. These icons include stylized human figures, arrows, and other symbols that suggest movement, communication, and data. The map is centered on the Atlantic Ocean, with North and South America on the left, Europe and Africa in the center, and Asia and Australia on the right. The overall color scheme is a light gray background with red icons and black text.

A NEW DEAL ON PERSONAL DATA

DEMOCRATIZING BIG DATA

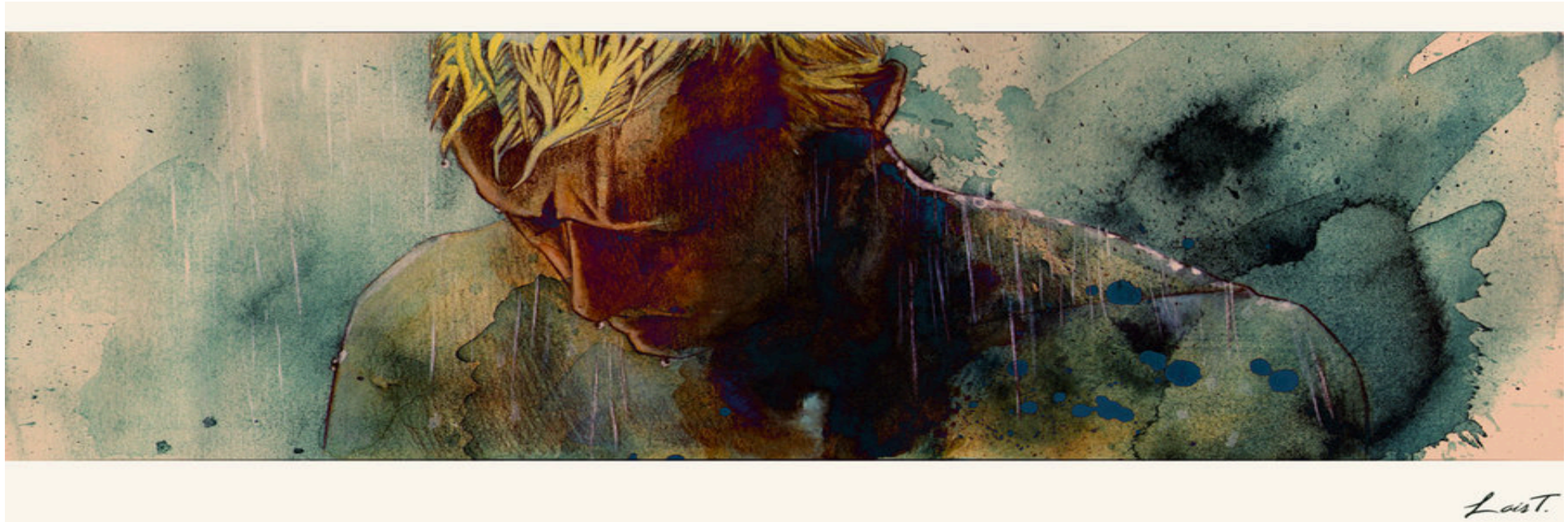
A USER-CENTRIC ECOSYSTEM FOR BIG DATA

Engage people in the creation and use of big data and knowledge, by empowering individuals with **self-knowledge**

Incentivize individual to **participate**, to align own self-interest with broader societal goals

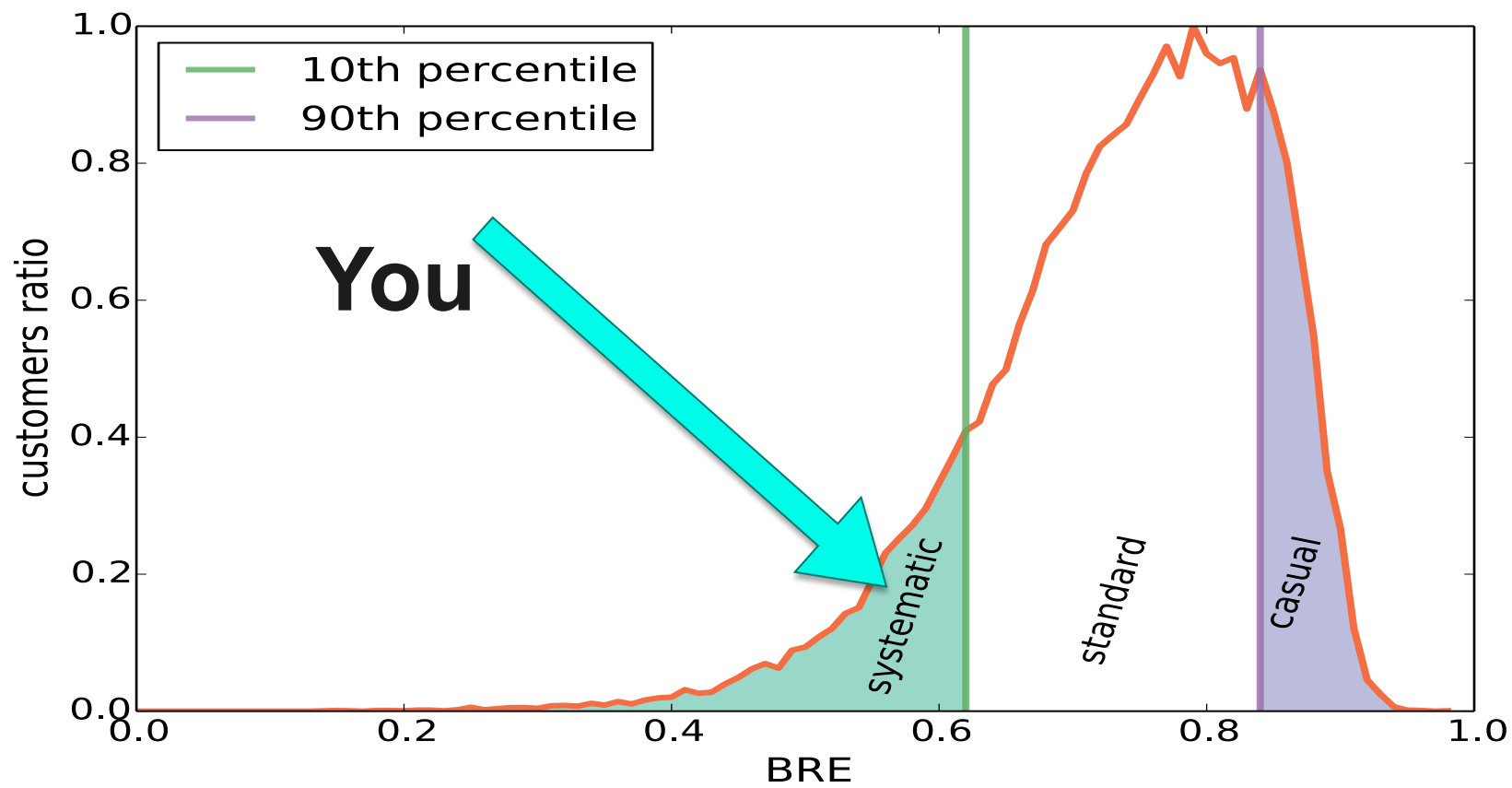
Based on **transparency, trust & privacy, peer-to-peer networks**

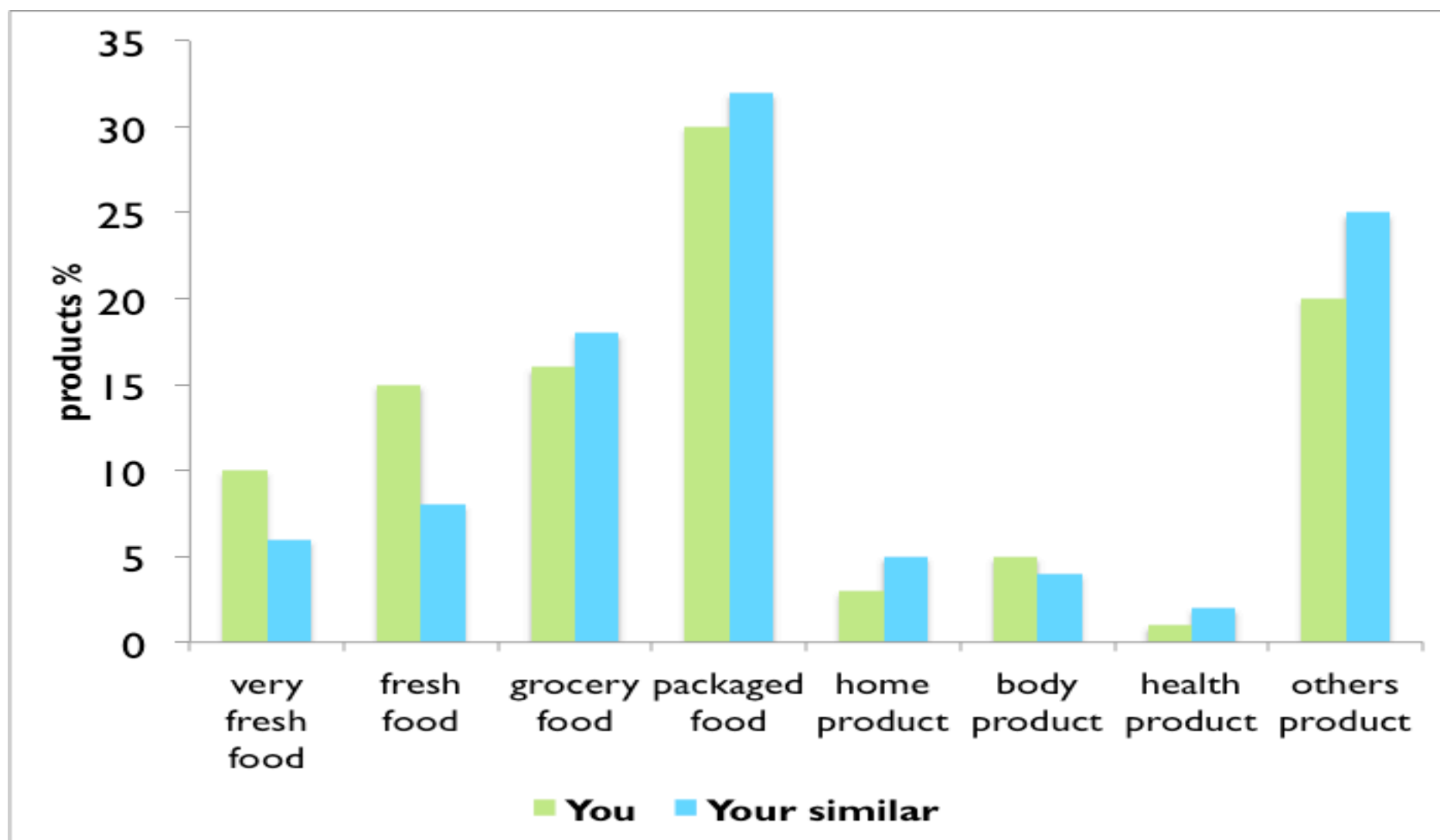




“I've... seen things you people wouldn't believe...
Attack ships on fire off the shoulder of Orion.
I watched c-beams glitter in the dark near the Tannhäuser Gate.
All those... moments...
will be lost in time, like...
tears... in... rain.”



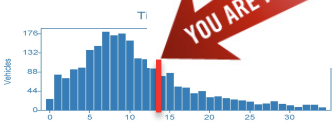
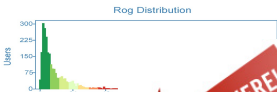
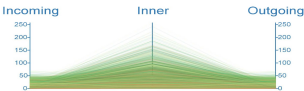
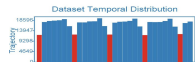




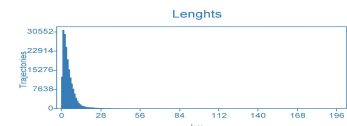
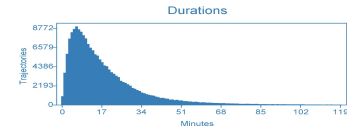
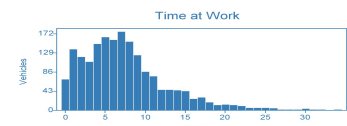
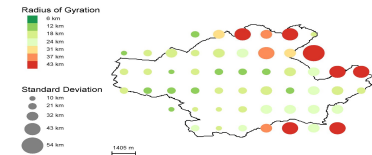
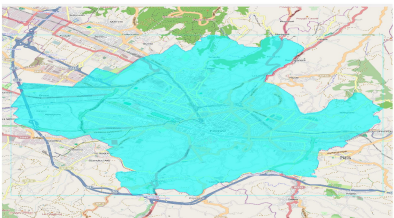
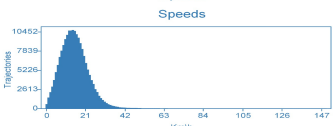
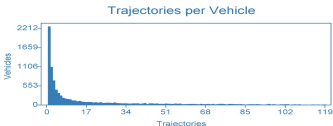
HOW ARE YOU COMPARED TO YOUR CITY'S MOBILITY ATLAS?

Firenze

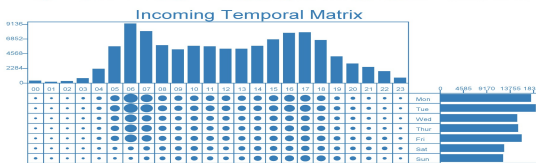
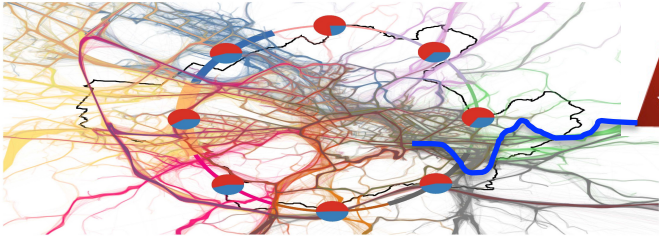
Surface area: 106 km²
Coordinates: 43,78 11,24
Vehicles: 32.752
From: 2011-05-01 To: 2011-05-31



Inner Traffic (145.034 Trajectories)



Incoming Traffic (105.881 Trajectories)

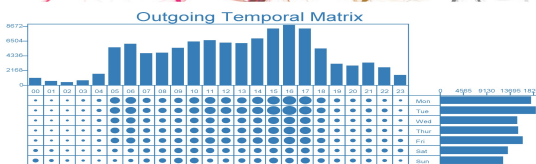
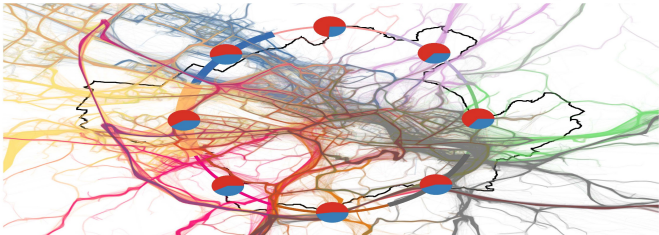


	Traj	Perc
SUD 16%	Lastra a Signa	2.342 99%
	Impruneta	3.853 87%
	San Casciano L.	1.836 75%
	Figline Vald'.	1.150 81%
	Greve in Chian.	868 36%
	Tavarnelle Val.	744 93%
	Bagno a Ripoi	7.314 93%
EST 19%	Fiesole	3.970 95%
	Pontassieve	2.797 93%
	Greve in Chian.	1.516 63%
	Rignano sul A.	774 92%

Regular VS Occasional



Outgoing Traffic (106.418 Trajectories)



	City	Traj	Perc
NORD 12%	Sesto Fiorenti.	8.356	60%
	Calenzano	1.235	36%
	Voglia	845	87%
	Campi Bisenzio	487	7%
	Borgo San Lore.	456	54%
OVEST 52%	Scandicci	13.439	98%
	Prato	6.168	95%
	Campi Bisenzio	5.846	92%
	Sesto Fiorenti.	5.521	39%
	Lastra a Signa	2.423	98%
SUD 14%	Impruneta	3.905	85%
	San Casciano I.	1.801	72%
	Figline Valdai.	1.155	77%
	Tavarnelle Val.	742	92%
	Greve in Chian.	739	29%
EST 20%	Bagno a Ripoi	7.701	96%
	Fiesole	3.982	94%
	Pontassieve	2.806	98%
	Greve in Chian.	1.670	67%
	Rignano sul A.	818	96%

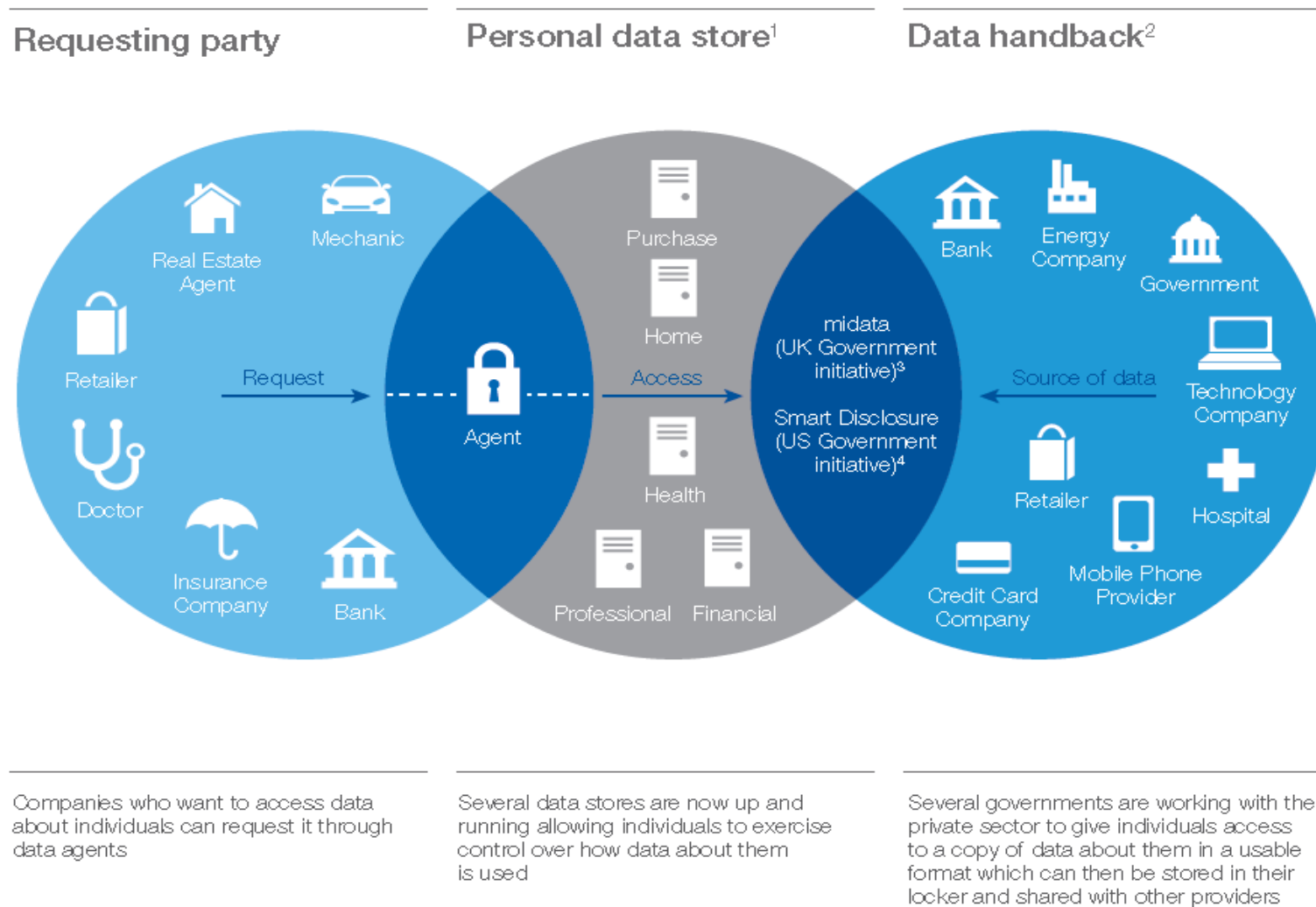
Regular VS Occasional



KNOW THYSELF



Empower everyone with tools to integrate own digital breadcrumbs into meaningful knowledge, to boost **self-awareness** of own behavioral patterns: health, mobility, consumptions, ...



A world map rendered in a digital, network-like style. The continents are outlined by a complex web of glowing blue lines and nodes, suggesting a global network or data flow. The background is dark blue with faint, swirling patterns. The overall aesthetic is high-tech and digital.

An ecosystem of personal data stores

**A DIGITAL NERVOUS SYSTEM TO UNDERSTAND
AND MANAGE OUR COMPLEX WORLD**

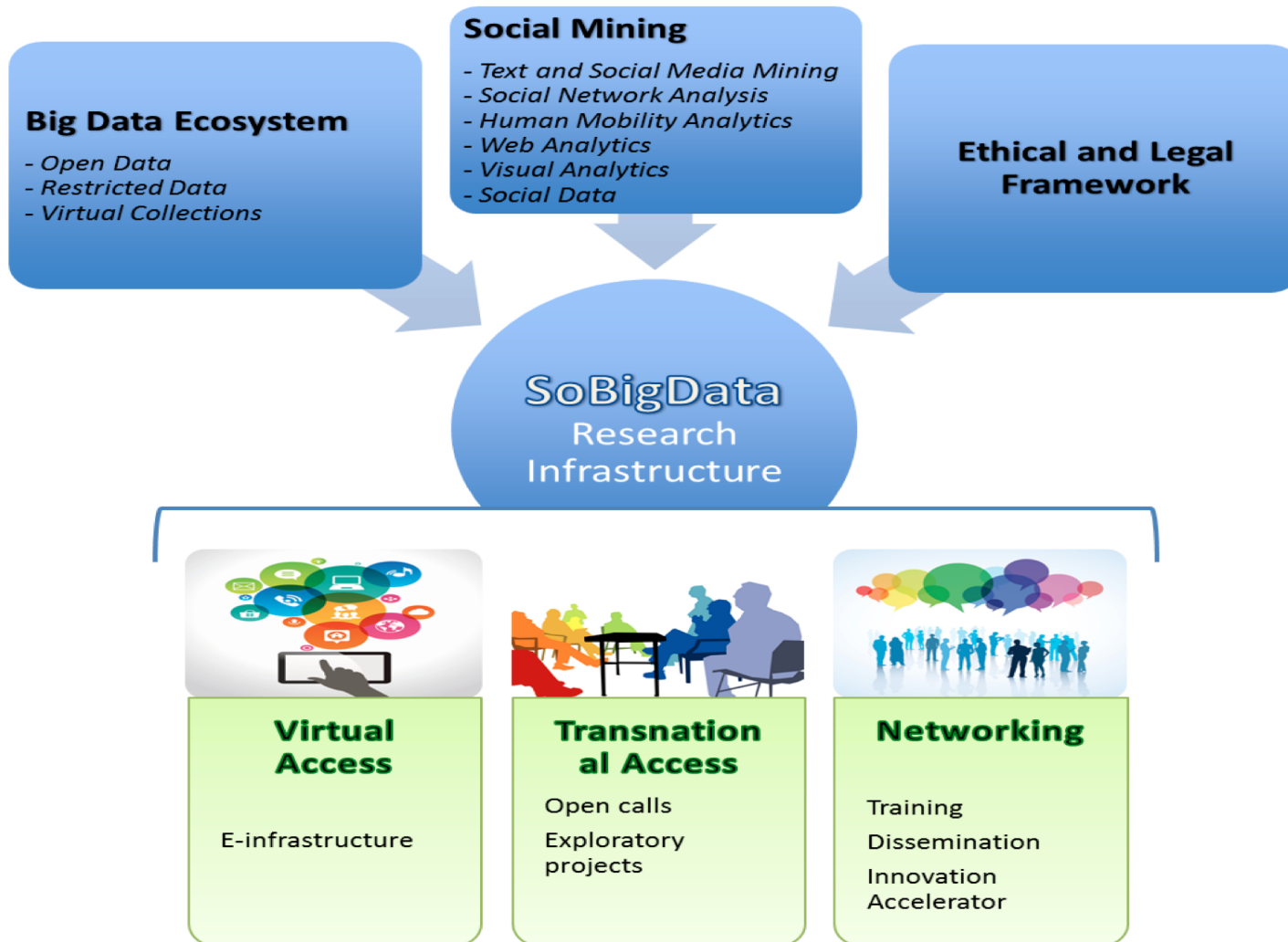


Research Infrastructure

WWW.SOBIGDATA.EU

- 1 - CNR Consiglio Nazionale delle Ricerche Italy
- 2 - USFD University of Sheffield UK
- 3 - UNIPI Università di Pisa Italy
- 4 - FRH Fraunhofer IAIS and IGD Germany
- 5 - UT Tartu Ulikool Estonia
- 6 - IMT Scuola IMT Lucca Italy
- 7 - LUH Leibniz Universitaet Hannover Germany
- 8 - KCL King's College London UK
- 9 - SNS Scuola Normale Superiore di Pisa Italy
- 10 - AALTO Aalto University Finland
- 11 - ETHZ ETH Zurich Switzerland
- 12 - TUDelft Technische Universiteit Delft Netherlands

Starts: September 2015





SoBigData.eu – the CERN for Big Data Analytics

KEY PUBLICATIONS

F Giannotti, M Nanni, F Pinelli, D Pedreschi. Trajectory pattern mining. ACM SIGKDD 2007

F Giannotti, D Pedreschi. Mobility, data mining and privacy: Geographic knowledge discovery. Springer, 2008

A Monreale, F Pinelli, R Trasarti, F Giannotti. WhereNext: a location predictor on trajectory pattern mining. ACM SIGKDD 2009

S Rinzivillo, D Pedreschi, M Nanni, F Giannotti, N Andrienko, G Andrienko. Visually driven analysis of movement data by progressive clustering. Information Visualization 7 (3-4), 225-239. 2008

D Wang, D Pedreschi, C Song, F Giannotti, AL Barabasi. Human mobility, social ties, and link prediction. ACM SIGKDD 2011

F Giannotti, M Nanni, D Pedreschi, F Pinelli, C Renso, S Rinzivillo, R Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. The VLDB 20(5) 2011

R Trasarti, F Pinelli, M Nanni, F Giannotti. Mining mobility user profiles for car pooling. ACM SIGKDD 2011

KEY PUBLICATIONS

M Coscia, G Rossetti, F Giannotti, D Pedreschi. Demon: a local-first discovery method for overlapping communities. ACM SIGKDD 2012

S Rinzivillo, S Mainardi, F Pezzoni, M Coscia, D Pedreschi, F Giannotti. Discovering the geographical borders of human mobility. KI-Künstliche Intelligenz 26 (3) 2012

D Pennacchioli, M Coscia, S Rinzivillo, D Pedreschi, F Giannotti. Explaining the Product Range Effect in Purchase Data. IEEE BIGDATA 2013

B Furletti, L Gabrielli, C Renso, S Rinzivillo. Analysis of GSM Calls Data for Understanding User Mobility Behavior. IEEE BIGDATA 2013

L Milli, A Monreale, G Rossetti, D Pedreschi, F Giannotti, F Sebastiani. Quantification trees. IEEE ICDM 2013

Giusti, Marchetti, Pratesi, Salvati, Pedreschi, Giannotti, Rinzivillo, Pappalardo, Gabrielli. Small area model based estimators using Big Data Sources. Journal of Official Statistics, 31(2) 2015.

Furletti, Gabrielli, Garofalo, Giannotti, Milli, Nanni, Pedreschi, Vivio. Use of mobile phone data to estimate mobility flows. Measuring urban population and intercity mobility using big data in an integrated approach. Italian Symposium on Statistics, 2014.

Luca Pappalardo, Maarten Vanhoof, Zbigniew Smoreda, Dino Pedreschi, Fosca Giannotti. Human Mobility, Social Networks and Economic Development. Submitted (2015).

VISION PAPERS

F Giannotti, D Pedreschi, A Pentland, P Lukowicz, D Kossmann, J Crowley, D Helbing. **A planetary nervous system for social mining and collective awareness.** The European Physical Journal Special Topics 214 (1), 49-75, 2012

J van den Hoven, D Helbing, D Pedreschi, J Domingo-Ferrer, F Giannotti . **FuturICT—The road towards ethical ICT.** The European Physical Journal Special Topics 214 (1), 153-181, 2012

M Batty, KW Axhausen, F Giannotti, A Pozdnoukhov, A Bazzani, M Wachowicz. **Smart cities of the future.** The European Physical Journal Special Topics 214 (1), 481-518, 2012

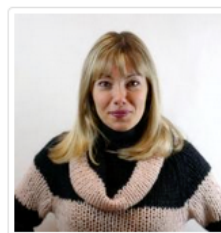
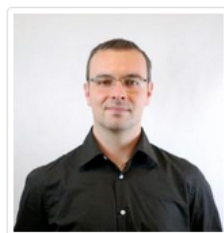
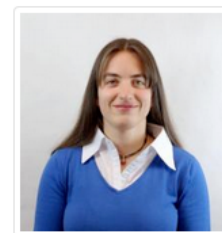
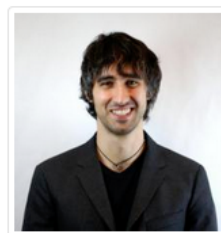
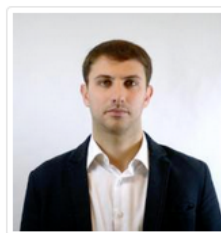
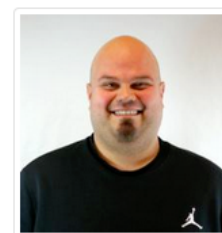
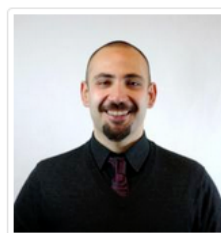
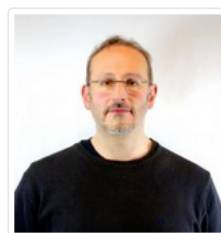
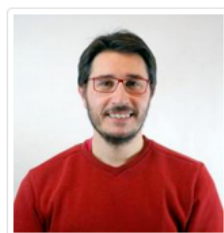
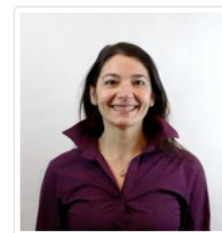
THIS IS THE WORK OF MANY PEOPLE FOR A LONG TIME

Fosca Giannotti, co-lead of KDD LAB

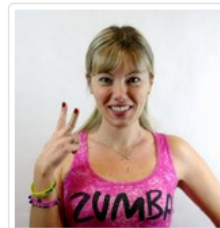
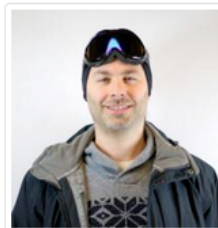
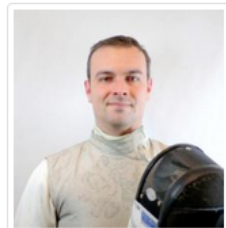
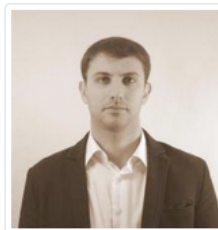
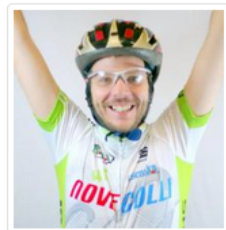
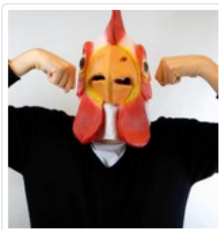
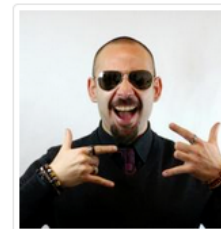
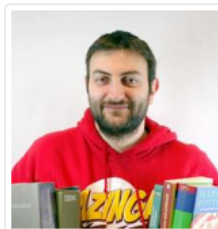
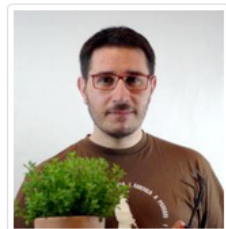
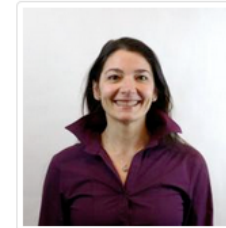
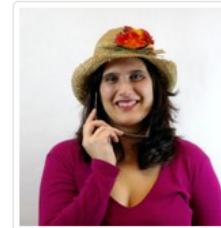
Salvo Rinzivillo, Mirco Nanni, Roberto Trasarti, Salvatore Ruggieri, Chiara Renso,
Anna Monreale, Franco Turini

all the fantastic folks at KDD LAB Pisa

many international collaborators



**Knowledge Discovery
& Data Mining Lab**
<http://kdd.isti.cnr.it>



**Knowledge Discovery
& Data Mining Lab**
<http://kdd.isti.cnr.it>