Social Mining &

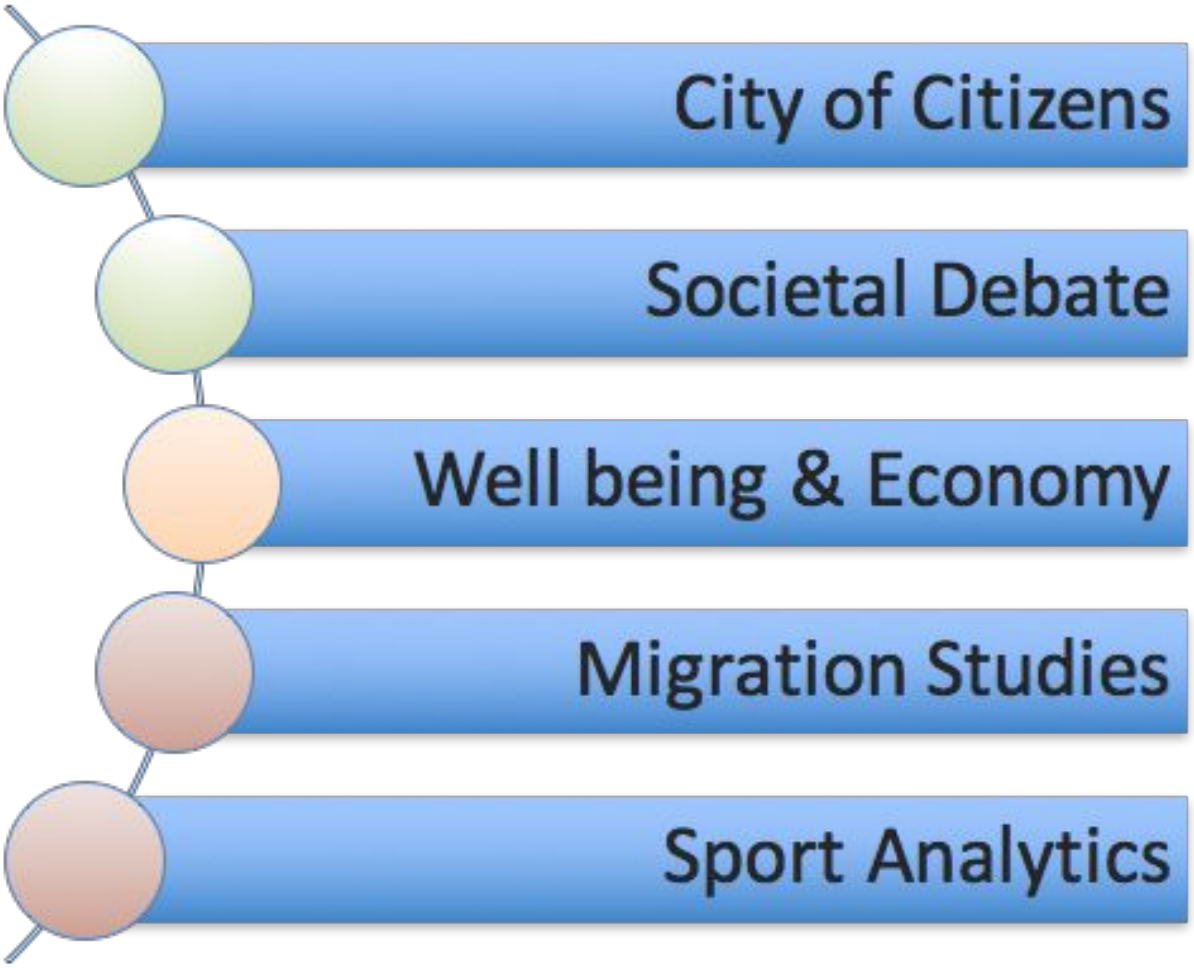
Big Data Ecosystem

H2020 - www.sobigdata.eu

September 2015- August 2019

@SoBigData (https://twitter.com/SoBigData)
https://www.facebook.com/SoBigData

Big Data as proxy of Human behavior



Smart City, Mobility Data Analysis, Transportation system, Multi-modal planning, Car-Pooling, etc.

Social Media analysis, newspaper articles, social debate, text analysis, sentiment analysis, etc.

Economic indicators, health, diffusion model, cost of life, crisis prediction, etc.

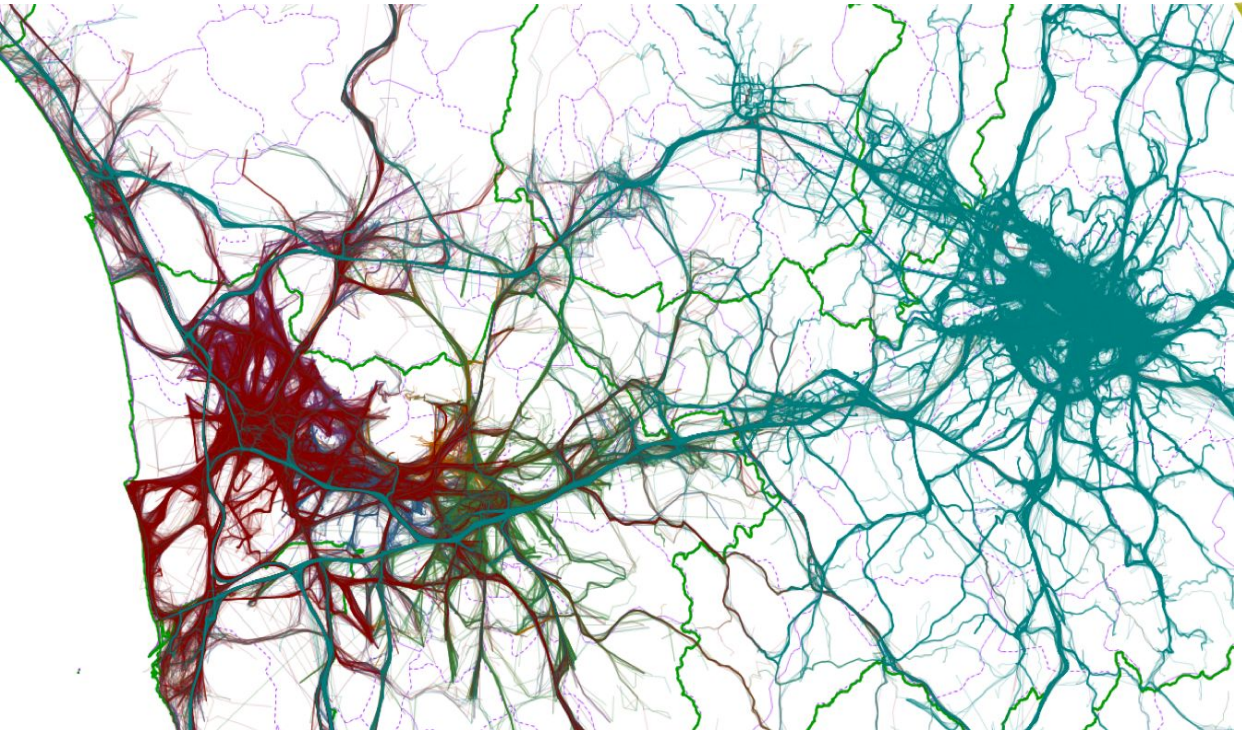
Migration flows, sentiment analysis, complex network analysis, etc.

Success Indicators, Training models, Tactics impact, etc.

What defines the human division of territory?

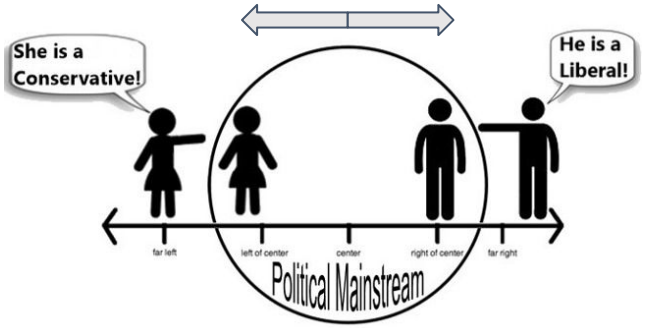
cities are placed in particular areas for a number of good reasons: communication routes, natural resources, migration flows. But once cities are located in a given spot, **who decides where one city ends and another begins?**

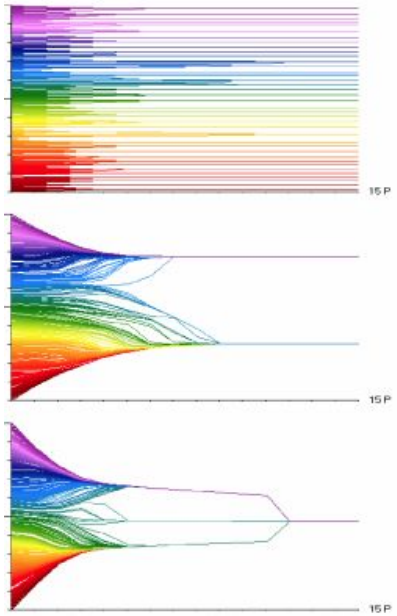
Network analysis can be useful in this context, because it can provide an objective way to divide the territory according to a particular theory.

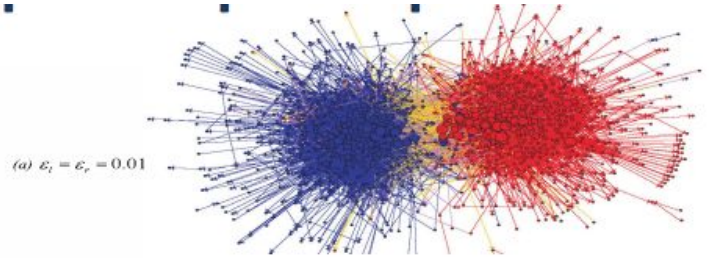


Societal Debate

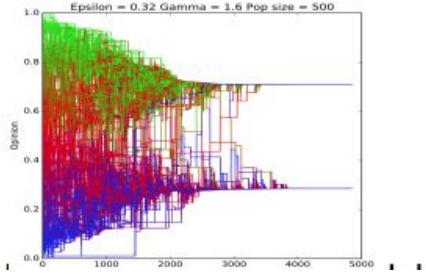
What is the effect of Topics/Posts Recommendation systems in Social Networks?
Algorithmic bias amplifies opinion polarization of the users showing them only a specific (their) view of the reality.







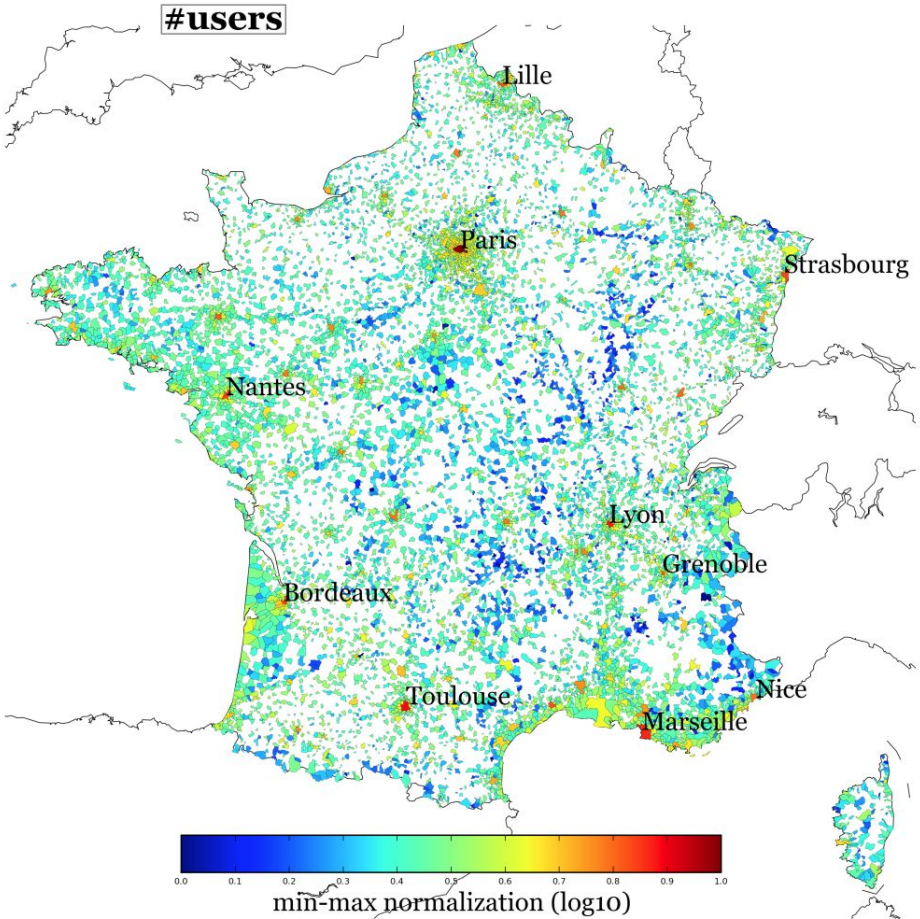
$$p_i(j) = \frac{d_{ij}^{-\gamma}}{\sum_{k \neq i} d_{ik}^{-\gamma}}$$

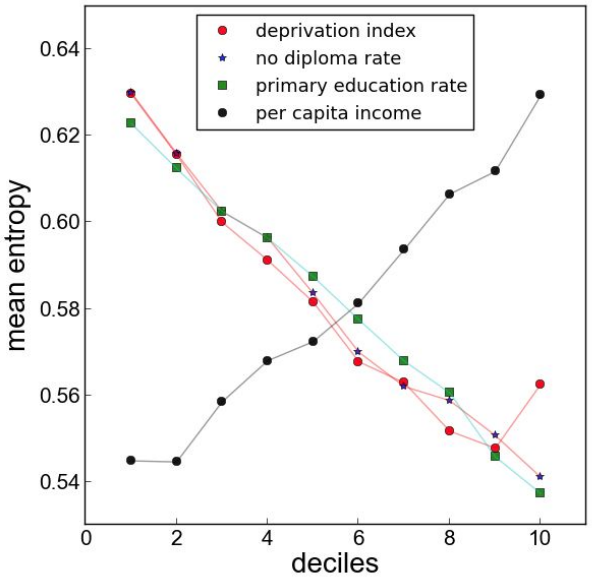


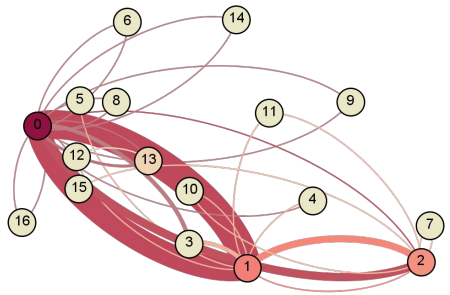
WellBeign & Economy

Can we estimate Country well-being using new Big Data sources?

We studied human behavior through the lens of phone data records by means of new statistical indicators that quantify and possibly “nowcast” the well-being and the socio-economic development of a territory.

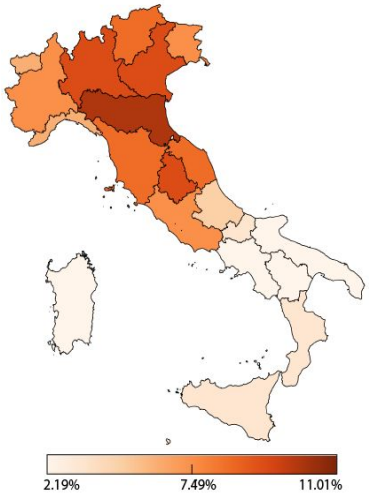




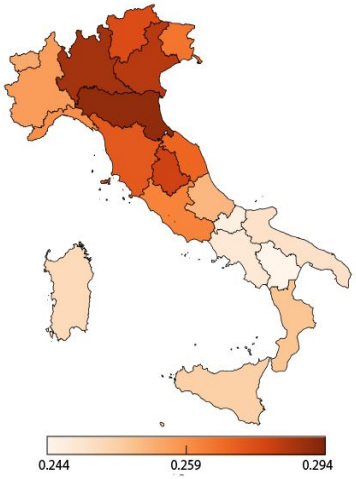


How to estimate the Migration stock in a country?

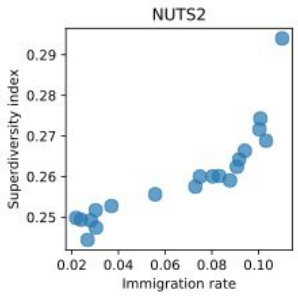
Definition of Superdiversity: distance between the emotional content of words in the standard language and on Twitter. It is strongly correlated with the immigrant presences.

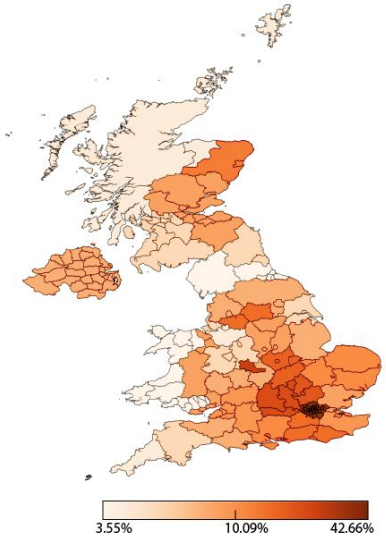


Immigration Rate

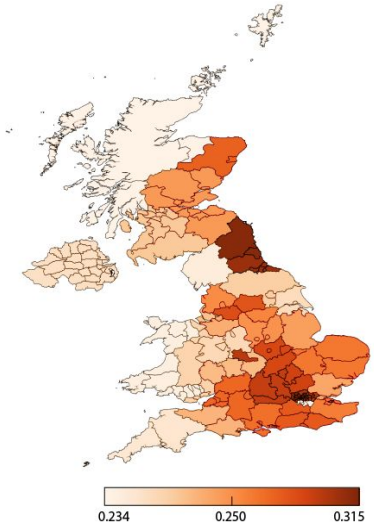


Superdiversity

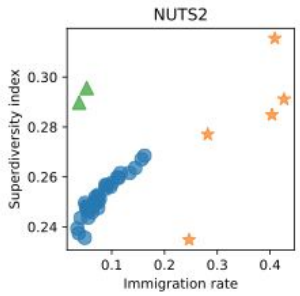




Immigration Rate

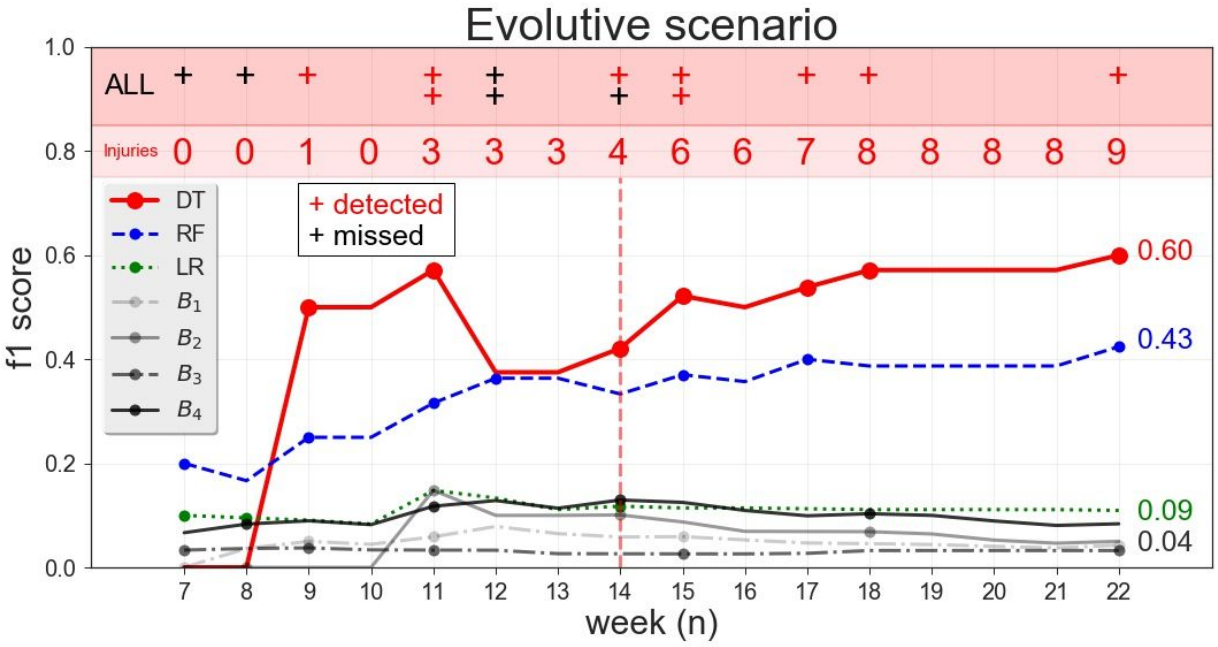


Superdiversity

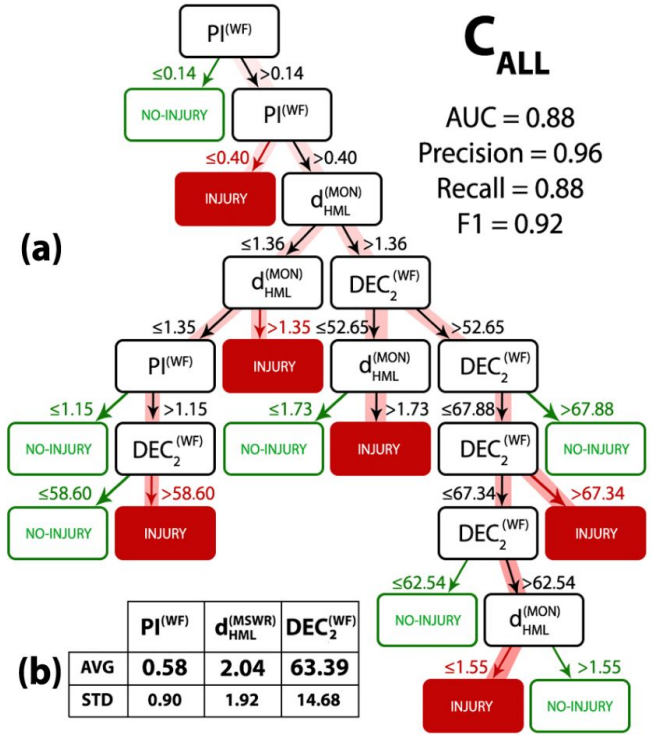


Can we predict players injuries?

Tracking (GPS) the players activities we collected data describing the training workload of players. Using this training set and a multidimensional machine learning approach, it is possible to generate rules and predict injuries.



(a)



(b)

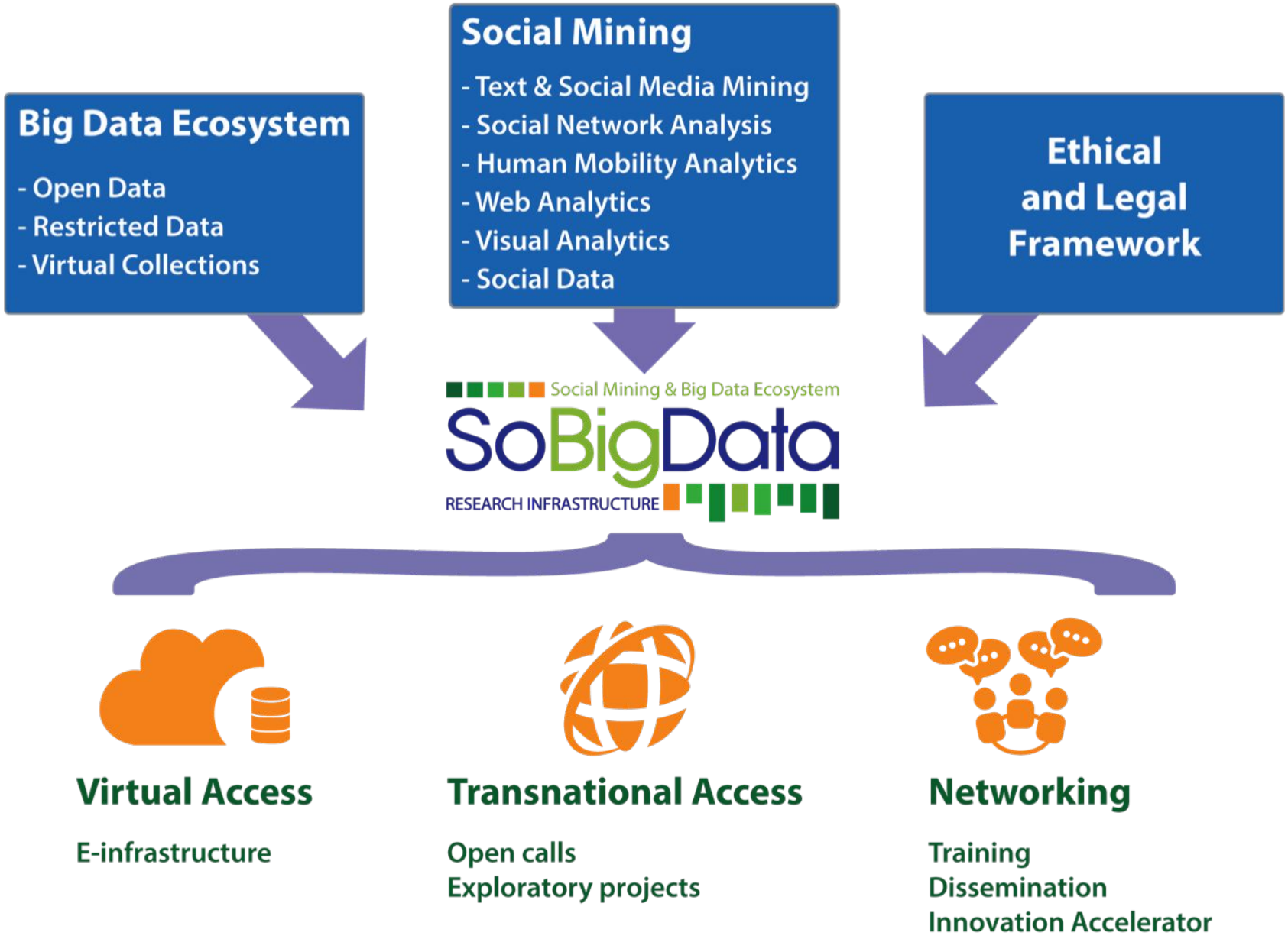
	$PI^{(WF)}$	$d^{(MSWR)}_{HML}$	$DEC_2^{(WF)}$
AVG	0.58	2.04	63.39
STD	0.90	1.92	14.68

(c)

Injury detection rules

	$PI^{(WF)}$	$d^{(MSWR)}_{HML}$	$DEC_2^{(WF)}$	freq %	acc %
RULE 1	(0.14, 0.40]			58	100
RULE 2	>0.40	>1.73	>52.65	16	66
RULE 3	>1.15	≤1.35	>58.60	11	50
RULE 4	>0.40	>1.36	(67.34, 67.88]	5	100
RULE 5	>0.40	(1.36, 1.55]	(62.54, 67.34]	5	100
RULE 6	>0.40	(1.35, 1.36]		5	100

The Project


Social Mining
- Text & Social Media Mining
- Social Network Analysis
- Human Mobility Analytics
- Web Analytics
- Visual Analytics
- Social Data
Big Data Ecosystem
- Open Data
- Restricted Data
- Virtual Collections
Ethical and Legal Framework
Social Mining & Big Data Ecosystem
SoBigData
RESEARCH INFRASTRUCTURE
Virtual Access
E-infrastructure
Transnational Access
Open calls
Exploratory projects
Networking
Training
Dissemination
Innovation Accelerator