

Towards a Digital Time Machine fueled by Big Data and Social Mining

Dino Pedreschi

KDD LAB Knowledge Discovery and Data Mining Lab
Universit'a di Pisa e Istituto di Scienza e Tecnologie dell'Informazione del CNR
`dino.pedreschi@unipi.it`

A sliding window has been opened by the ICTs, and now we live a measurable global society. “Big Data”, i.e., the digital breadcrumbs of human activities, sensed as a by-product of the technologies that we use, let us observe the individual and collective behavior of people at an unprecedented detail, scale and speed. All dimensions of our social life have big data “proxies”:

- our desires, opinions and sentiments leave their traces in the social media we participate in, in the query logs of the search engines we use, in the tweets we send and receive
- our relationships and social ties leave their traces in the network of our phone or email contacts, in the friendship links of our favorite social network
- our shopping patterns and lifestyles leave their traces in the transaction records of our purchases;
- our movements leave their traces in the records of our mobile phone calls, in the GPS tracks of our onboard navigation system.

Big Data at a societal scale provide a powerful microscope, which together with Social Mining the ability of discovering knowledge from these data can help us understand, nowcast and forecast many complex and hidden socio-economic phenomena, from the diffusion of viruses, information, innovation and crises to the unequal distribution of resources and opportunities. Scientific research is being revolutionized by this new wave, and policy making is next in line, because big data and social mining are providing novel means for measuring and monitoring well-being in our societies more realistically, beyond the GDP, more precisely, continuously, everywhere. A fair use of big data and the development of policies and standards for empowering people whose digital traces are recorded in the data, has the potential to yield us a conceptual framework for exploring the past, the present and the near future(s) by making sense of the digital traces a Digital Time Machine. In the first place, the digital time machine is meant to empower individual citizens, which can explore the past and the present to gain better knowledge of self and own position in society, and explore plausible futures to reason on the consequences of decision making. But the digital time machine also works for communities, institutions, and businesses. It will be based on a fair use of big data and modeling, rooted on ethical values such as trust, privacy, transparency, and public good. This new wave has profound repercussions on many ethically sensitive issues, including privacy and data protection (Who can access my data?), data ownership and exploitation (Who owns

my data? For what purposes?), transparency (Who does what with my data?), self-awareness and self-empowerment (Can I access the knowledge hidden in my data?), monopoly of knowledge (How to counterbalance the power of web mega-corporations?). In our globally interconnected world, we cannot afford to miss the opportunity disclosed by big data, but within a socio-techno-legal framework that allows for knowledge to become a safe, public good. My seminar discusses the novel questions that big data and social mining allow to raise and answer, how a new paradigm for scientific exploration, statistics and policy making is emerging, and the major scientific, technological and societal barriers to be overcome to realize this vision. I will focus on two concrete projects with telco providers and official statistics bureau in Italy and France aimed at measuring, quantifying and possibly predicting key demographic and socio-economic indicators based on nation-wide mobile phone data: the population of different categories of city users (residents, commuters, visitors) in urban spaces, the inter-city mobility, the level of well-being and economic development of geographical units at various scales.

References

1. 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) p. 49-75. Springer (2012). Available open access at <http://link.springer.com/article/10.1140/epjst/e2012-01688-9>
2. M. Nanni, F. Giannotti, C. Thanos, A. Rauber (editors). *Big Data Analytics: towards a European research agenda*. ERCIM White Paper on Big Data Analytics (2014). Available online at <http://www.ercim.eu/news/387-ercim-white-paper-on-big-data-analytics>
3. Fosca Giannotti, Dino Pedreschi. *Mobility, data mining and privacy*. Springer (2008)
4. Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, Albert-Laszlo Barabasi. Human mobility, social ties, and link prediction. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge discovery and data mining*, p. 1100-1108. ACM Press (2011)
5. S Rinzivillo, S Mainardi, F Pezzoni, M Coscia, D Pedreschi, F Giannotti. Discovering the geographical borders of human mobility. *KI-Kunstliche Intelligenz* 26 (3), 253-260 (2012)
6. D Pedreschi, S Ruggieri, F Turini. The Discovery of Discrimination. In *Discrimination and Privacy in the Information Society*, 91-108. Springer (2012)
7. M Atzori, F Bonchi, F Giannotti, D Pedreschi. Anonymity preserving pattern discovery. *The VLDB Journal* 17 (4), 703-727 (2007)
8. 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 Journal* 20(5), 695-719 (2011)
9. 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 inter-city mobility using big data in an integrated approach*. Italian Symposium on Statistics (2014).

10. A Monreale, S Rinzivillo, F Pratesi, F Giannotti, D Pedreschi. Privacy-by-design in big data analytics and social mining. *EPJ Data Science* 3 (1), 1-26, 2014
11. Dino Pedreschi. Big Data, Social Mining, Diversity, and Wellbeing. *Italian Symposium on Statistics* (2014).
12. Luca Pappalardo, Maarten Vanhoof, Zbigniew Smoreda, Dino Pedreschi, Fosca Giannotti. Human Mobility, Social Networks and Economic Development. Submitted (2014).