Information society: how to use mathematical models.

Francesco De Pretis
Link Campus University, Rome - Italy
f.depretis@unilink.it

Abstract

Information society is constantly fueling data of whatever sort, from social network to geo-localized records: their magnitude is posing new questions over the way such data have to be analyzed and interpreted. In this paper, we focus on a new mathematical approach arising from the conceptual core of statistical mechanics, which has recently proved well in offering valuable insights for large data analysis. After having examined this new mathematical perspective, we provide a case-study centered on data regarding immigrant integration in Italy.

Information society and rising of big data

Back to 2009, in a mile-stone journalistic report which opened the big data era, the Economist titled "Data, data everywhere" (The Economist, 2009). Today, living in the information society means accessing enormous quantities of data, as it never occurred in the past. Just to mention an example, the world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s (Hilbert & Lopez, 2011): as of 2012, every day 2.5 exabytes (2.5×10^{18}) of data were created (IBM, 2013). Accessing big data, however, is only the starting point: nowadays, the big stake comes when the analysis sets in. For this reason - as this article will point out - we need new paradigms, we need new mathematical approaches of analysis.

Modelling data of the information society: the statistical mechanics perspective

Albeit we live in the information society, the inability of giving credible forecasts over phenomena that we observe every day is something which has been perceived even from the most conservative establishment: as reported by the British newspaper The Daily Telegraph (Pierce, 2008), in occasion of the opening of the academic year at London School of Economics on November 2008, Her Majesty Elizabeth II abruptly asked the economists "why did nobody notice it ?", referring to the ongoing financial crisis.

The question whether none could really predict the so-called "credit-crunch" started in 2007 with the default of Lehman Brothers (at that time, the fourth-largest investment bank in the US) and – to mention another example – the reason why it has been so difficult to handle immigration flows in Europe since the beginning of the '90s (Willekens, 1994; Bijak & Wiśniowski, 2009) put in doubt the actual modelling of economic and social phenomena. Juxtaposing immigration and economic phenomena could sound unusual but, on the other hand, it turns out that they have more things in common than one might initially expect: such phenomena are processes characterized by occurring among a very large number of people, more resembling the particles of gas rather than a spare group of players; for this reason, their analogy with thermodynamic systems has been advocated not only by physicists (Georgescu-Roegen, 1971).

However, despite these properties, such phenomena have been for a long time modeled and studied by social scientists through a classical physics approach (Jevons, 1871), designing social processes as an idealized system of perfectly rational, optimizing institutions and individuals, who, by trading in markets, bring the economy and society to a balanced, efficient equilibrium; a "postcard-scenery" which has been completely blasted by the turmoil of the ongoing financial crisis.

The disappointment related to the low predictability of such models has found endorsers even in the most unexpected people, such as real economic agents, seen as "insiders" in the world of finance. This is the case – for example – of George Soros, who claims to have been for a long time aware of the low consistency of the models used to give explanation to social phenomena; in 2009 such beliefs pushed Soros to contribute with a 50 million US dollars pledge to the foundation of the Institute for New Economic Thinking (INET), a New York City-based nonprofit think tank whose purpose is to support academic research and teaching in economics "outside the dominant paradigms of efficient markets and rational expectations" (Weber, 2011).

The ideas beneath INET and the new vision of social science the institute embodies are well exemplified by the initial part of a speech Soros gave at the

INET Conference at King's College on April 2010: "Economic theory has modeled itself on theoretical physics. It has sought to establish timelessly valid laws that govern economic behavior and can be used reversibly both to explain and to predict events. But instead of seeking laws capable of being falsified through testing, economics has increasingly turned itself into an axiomatic discipline consisting of assumptions and mathematical deductions – similar to Euclidean geometry. Rational expectations theory and the efficient market hypothesis are products of this approach. Unfortunately they proved to be unsound. To be useful, the axioms must resemble reality. Euclid's axioms meet that condition; rational expectations theory does not. It postulates that there is a correct view of the future to which the views of all the participants tend to converge. But the correct view is correct only if it is universally adopted by all the participants – an unlikely prospect. Indeed, if it is unrealistic to expect all participants to subscribe to the theory of rational expectations, it is irrational for any participant to adopt it. Anyhow, rational expectations theory was pretty conclusively falsified by the crash of 2008 which caught most participants and most regulators unawares. The crash of 2008 also falsified the Efficient Market Hypothesis because it was generated by internal developments within the financial markets, not by external shocks, as the hypothesis postulates." (Soros, 2010)

As Soros correctly outlines, the information society needs new paradigms and new models, able to embrace the complex processes that take place into the real world between human beings. Such complexity cannot be captured any more by a classical physics approach: the point of view underlying this article is that this new way, this novel approach to the study of such phenomena necessarily starts with statistical mechanics, a field between mathematics and physics used for "statistical probabilistic predictions about systems which either contain elements which are too small to see or too numerous to keep track of; or usually both" (Susskind, 2013).

Which reasons could drive us to endorse such shift, this passage from the classical approach to the statistical one? In order to reply to such question, it is worth to recall in this place which are the main differences between classical and statistical mechanics, observing the methodologies such fields implement to give reason to the phenomena under their own lens.

As summarized in (Susskind, 2013), classical mechanics mainly concerns with the concept of perfect predictability: the rough idea is that while witnessing the phenomena in a "closed system", (i.e. separated and not interacting with anything else outside such system), one can make predictions with a maximum precision or at a given level of precision (or predictability), just knowing two things, i.e. the initial conditions and the laws of evolution of the system – a vision that dates backs already the beginning of the nineteenth century, as popularized by Pier Simon Laplace (Laplace, 1814).

The point is that complete predictability can turn out to be totally useless. This assumption has not to be viewed as a criticism since the basic laws of classical mechanics have shown to be very powerful in their predictability: the point – again – is that in many cases such laws turns to be very useless to actually analyzing what it is really going on. For example, having a list of the positions and velocities of every particle in a certain room would not be very useful to us, since the list would be too long and would not give explicit account of other macroscopic properties of major interest, such temperature, pressure and so on. To shift such example to the set of economic and social phenomena, it could be not such worth (or useful) to cast predictions over the amount of each single bank account of a given population, to know the exact price with which each barrel of crude oil is exchanged world-wide at a given time or with whom and when each mixed marriage among people coming from two populations takes place. Surely instead, it would be of foremost interest to know with a good predictability the gross domestic product of a nation, the international price of crude-oil or the way mixed marriages and other integration phenomena occur in a given nation, in a given time. In other words, it could sound more interesting to know not what the single does but what comes out when many singles interact together.

This is exactly the case of statistical mechanics, a theory which provides tools that are applicable when the initial conditions of a system are not known with infinite precision (the so-called "starting point"), the laws of evolution of the system are neither completely known or when the system is not closed (for example, if it is interacting with other elements outside). Therefore, to say that in (Susskind, 2013) words: "when ideal predictability is not possible or it is not practicable way, you resort to probability and statistical mechanics".

However, the renounce to classical mechanics total predictability does not mean that statistical mechanics is not itself a sound field which can offer precise and viable answers. It is known that when the number of elements in a systems tend to be very large, also probability tend to be a very precise predictor (when – of course – the law of large numbers is applicable), so that statistical mechanics itself can be highly predictable but not for everything: even for the more complicated properties of a given system, like fluctuations and large deviations, statistical mechanics is always a feasible tool not to forecast predictions over the exact time when such unusual events might happen but to give precise predictions of the probability associated to such unusual events. From another point of view, statistical mechanics cannot predict information on every element of a given system but can afford precise and reliable information about collective properties arising from the interaction of a large number of elements, providing a framework for relating the microscopic properties of individual elements to the macroscopic bulk properties of a given system or population.

A case-study from Italy

As example of how a statistical mechanics model could infer unknown information from a large dataset, we offer here a case-study regarding immigrant integration phenomena in Italy.

In (De Pretis, 2014), the aim is to cast light over the way immigrant integration phenomena occur, assuming a quantitative point of view based on the exploitation of large empirical datasets. Starting from seminal works on national (Barra, Contucci, Sandell & Vernia, 2014) and regional immigration data (De Pretis & Vernia, 2014), we share the simple observation that very little is known about the mechanisms that bring about integration. For example, elementary questions concerning how integration responds to an increase in immigration density or to what extent the intensity of interaction modifies the level of integration still beg coherent empirical and theoretical answers. It is therefore manifest that the missing of knowledge over the mechanisms internally governing such phenomena undermines the effectiveness of formulating social policies concretely able to promote integration. (De Pretis, 2014) thus proposes new perspectives according the theoretical modelling based on statistical mechanics methods, whose growing importance in social sciences (Galam & Moscovici, 1991; Durlauf, 1996; Durlauf, 1999; Brock & Durlauf, 2001; Castellano, Fortunato & Loreto, 2009; Montanari & Saberi, 2010; De Pretis, 2012) has been extensively underlined in the previous paragraph.

According to this perspective, the main result of (De Pretis, 2014) is the identification and theoretical interpretation of two empirical laws (a linear one and a square root one) which connect two quantifiers of integration (i.e. the percentage of mixed marriages and the percentage of mixed newborns) to the density of immigrants related to the geographical areas where the previous quantifiers have been measured. This result has been obtained analyzing three large datasets containing over 10^6 information regarding marriages and births registered in Italy during an eleven years span. It is important to stress that both Italian datasets have been characterized by showing the two distinct patterns only when partitioning data according to the size of their municipalities' population.

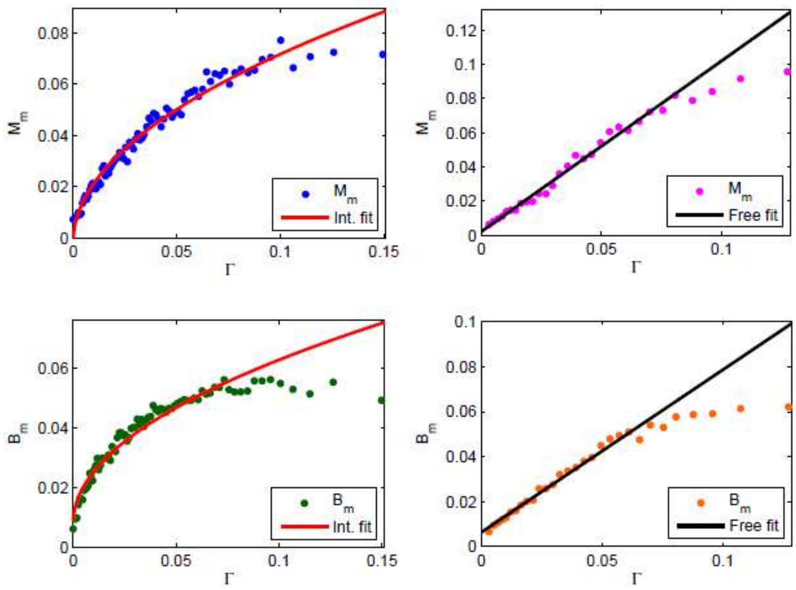

Figure 1. *Mixed marriages and newborns in Italy*. Dots are average quantities versus Γ where $\Gamma = \gamma(1-\gamma)$ and γ is the immigrant density. Left upper panel: quantifier M_m^s (blue dots), fraction of mixed marriages occurred in municipalities with less than 10 000 inhabitants, fitted by $a\sqrt{\Gamma}+b$ (red curve). Right upper panel: quantifier M_m^b (pink dots), fraction of mixed marriages occurred in municipalities with more than 10 000 inhabitants, fitted by $a\Gamma+b$ (black curve). Similar analyses conducted in lower panels for quantifier B_m (fraction of mixed newborns) both for small-sized and large-sized municipalities.

Looking at figure 1, Italian data suggest - according to statistical mechanics model presented in (Barra, Contucci, Sandell & Vernia, 2014) - that in small-sized municipalities (whose population does not exceed 10.000 inhabitants) imitative phenomena mainly take place while independent choices seem to be the most common patterns in large-sized municipalities. This result turns out to be even more impressive if compared with subtle considerations made by classical sociology authors, who understood a crisp difference in the emergence of social actions occurring in small-sized or large-sized municipalities already more than one hundred years ago (Durkheim, 1897).

In conclusion, the introduction of new mathematical models based on statistical mechanics is paving the way to more affordable analyses of data coming from the information society.

References

Barra, A., Contucci, P., Sandell, R. & Vernia, C. (2014). An analysis of a large dataset on immigrant integration in Spain. The Statistical Mechanics perspective on Social Action. Scientific Reports, Nature, 4 : 4174.

Bijak, J. & Wiśniowski, A. (2009). Forecasting of immigration flows until 2025 for selected European countries using expert information. IDEA Working Papers n° 7.

Brock W. & Durlauf S.N. (2001). Discrete choices with social interactions, Rev. Econ. St. 68.

Castellano C., Fortunato S. & Loreto V. (2009). Statistical physics of social dynamics, Rev. Mod. Phys. 81, 591-646.

De Pretis, F. (2012, June 27). Nella tempesta globale quali numeri ci salveranno? La Stampa.

De Pretis, F. (2014). A quantitative approach to the modelling of interacting system from empirical data: the statistical mechanics perspective and a case study from social sciences. Modena: University of Modena and Reggio Emilia.

De Pretis, F. & Vernia C. (2014). A Statistical Mechanics Approach to Immigrant Integration in Emilia Romagna (Italy)**.** Published in (P. Contucci, R. Menezes, A. Omicini, J. Poncela-Casasnovas - Complex Network V, Studies in computational intelligence - Springer International Publishing CHE) - n. volume 549 - pp. 57-62

Durkheim, É. (1897). Le Suicide : Étude de sociologie. Paris: Félix Alcan.

Durlauf, S.N. (1996). Statistical mechanics approaches to socioeconomic behavior, Technical Working Paper, 203, Natl. Bur. Econ. Res.

Durlauf, S.N. (1999). How can statistical mechanics contribute to social science?, Proc. Natl. Acad. Sc. USA 96.

Galam, S. & Moscovici, S. (1991). Towards a theory of collective phenomena: Consensus and attitude changes in groups, Europ. J. Social. Psycho. 21, 1, 49-74.

Georgescu-Roegen, N. (1971). The Entropy Law and the Economic Process. Cambridge, Massachusetts: Harvard University Press.

Hilbert, M. & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. Science 332 (6025): 60–65.

IBM (2013). What is big data? — Bringing big data to the enterprise. IBM website.

Jevons, W. S. (1871). The Theory of Political Economy. London: Macmillan & Co.

Laplace, P. S. (1814). Essai philosophique sur les probabilités. Paris: Courcier.

Montanari, A. & Saberi, A. (2010). The spread of innovations in social networks, Proc. Natl. Acad. Sc. USA, N.107, 20196.

Pierce, A. (2008, November 5). The Queen asks why no one saw the credit crunch coming. The Daily Telegraph.

Soros, G. (2010). Anatomy of Crisis – The Living History of the Last 30 years: Economic Theory, Politics and Policy. INET Inaugural 2010 Conference (p. 1-10). King's College Cambridge, England: INET.

Susskind, L. (2013). The Theoretical Minimum Lectures. Stanford, USA: Stanford University Web-Resources.

The Economist (2009). Data, Data Everywhere. Special Report. The Economist website.

Weber, T. (2011, January 26). Davos 2011: Soros warns UK cuts could cause recession. BBC News website.

Willekens, F. (1994). Monitoring international migration flows in Europe. European Journal of Population / Revue européenne de Démographie, 10(1), 1-42.