**Big Data: fashion, need or opportunity? The difference between description and comprehension.**

Day by day we are bombarded with a mass of information which is hard to keep up. Indeed, it is difficult to select those relevant to our needs, hard to remember where we had found or where we archived. The increased capability of computer systems and communication has enabled collection and access to many and many typologies of data and, at the same time, has increased the difficulty of "smart navigation". The main thing is that there is a huge difference between data, information, knowledge and wisdom. In fact, if statistically the increase of the amount of data could mean an increase accuracy in the forecast or precision of a measure, this is true only under certain assumptions on how the system behaves.

*Complicated* does not mean *complex*, *simple* does not mean *linear*. The complication is often synonymous of inconvenience, difficulty, and especially when dealing with calculation or prediction. But it does not mean inability to understand or solve. Unconsciously, it is associated to lack of technical capacity or information. A multiplication between two very large numbers is complicated if done by hand, as well as the budget of a large company. But if it is supported by data, procedures and calculators, this results as simplified.

Complexity is an intrinsic characteristic of a system linking the variables influencing it. Small deviations of a variable, or its measure as a data point, may dramatically impact on the evolution of the system. This means that we are not be able to describe it adequately, sometimes losing information which is very influential. The famous example of complex systems is that on climate change, where the beating of wings of a butterfly on one hemisphere of the planet might unleash a storm on the other hemisphere. So: complicated does not mean complex, complex often means complicated.

Let’s also reflect on the Ptolemaic approach to the movement of the planets around the Sun. In principle, today we could still use that procedure to calculate the position of the planets: with high performing computers and many observations we would be able to predict, with great accuracy, how the launch of a satellite would behave. When Kepler introduced his famous laws, everything became easier! But Kepler has simplified the problem, not linearized: on the contrary, Ptolemy was using the sum of many linear terms, the famous celestial circles, while Kepler introduced ellipses, more complicated from a conceptual point of view. So simple does not mean linear, but often simplicity follows the knowledge of the process that regulates the system: Kepler's laws are associated with Newton's laws of dynamics of masses.

This being said, we have some elements to understand better the difference between having a lot of data available and understanding what is happening: the difference between information and knowledge.

The difference between correlation and cause-effect relationship probably is well known, but often forgotten. Let’s imagine a train always leaving a station immediately after the bell tower ringing: probably it leaves each quarter of an hour, even if the bell does not ring. The correlation does not imply cause-effect. Data and laws that regulate the system are not necessarily closely related, unless there are two fundamental conditions: all the variables affecting the system are measured properly, you know the laws linking them.

Anyway, having many data often helps. None of us has ever seen an apple flying upwards when dropped. This gives us a clue that there is something attracting it, as the Earth. It 'a clue, not an evidence. And here we come to the fashion or utility of Big Data (meant as the collection and analysis of a large amount of information). First of all, we have to cesarly distinguish if we address a large number of data or a large number of variables. In fact, we could have tons of experiments on an dropped apples under different conditions, or tons of experiments on the formation of a cloud: it is not the same thing. It is intuitive to prefer the first case to the second, but it is not always so. Let's understand why.

Statistically, we have said that it is better to have a lot of data to increase the so-called robustness in the system description: higher accuracy, higher accuracy, greater ability to predict the future. Having a lot of data on "many" variables helps to better describe the system. Regarding complexity, increasing the information on more variables facilitates the discovery of unknown behaviors.

Then, the analysis of large amounts of data does not provide understanding, but allows a description with greater detail, and eventually the test of some solutions. This ability to test the system, however, is not always possible (see method of analogues and Kac's lemma): as an example, let’s think about the complexity of the socio-economic-political system at the global level, where experiments on different planets cannot be performed or a reversal in time to repeat the experiment with different conditions, in order to observe what would otherwise be. Therefore, the analysis of the correlations allows to predict the evolution of a system, within the statistical probability and in short timescales, but cannot suggest the laws which regulate it.

Big Data could be a fashion, but also an opportunity, hiding two great aspects: the economic and political use to predict/control social behaviors, the inability to understand the current complex system, asking for clues from a detailed description of the same. To make the next breakthrough step, however, we need someone who will have the intuition to transform the clues into a new paradigm: a new Kepler, Dirac, Einstein...