Science & Technology
FORESIGHT
from society to research

Background Document

Challenges for Complexity Science in the XXI Century

WG COMPLE XXI

National Research Council of Italy

CNR project Foresight

Working Group CompleXXI

Challenges for Complexity science in the XXI century[1]

Background paper

The CNR Foresight project aims to define research strategies for the next future. It is a common feeling in our society, even among non-scientist, that the XXI century will pose formidable challenges to Science. One is how to deal effectively with Complexity in the physical, biological, ecological, and social universe. Stephen Hawking himself declared that this 'will be the century of complexity'. Although the word Complexity is sometimes ambiguous is generally accepted that it refers to the emergence of *unexpected collective properties*, a priori unexpected from microscopic interactions. On general grounds, complex systems are characterized by: many heterogeneous interacting parts; multiple scales; complicated transition laws; unpredicted emergence; sensitive dependence on initial conditions; path-dependent dynamics; networked hierarchical connectivities; interaction of autonomous agents; self organisation; non-equilibrium dynamics etc [SanMiguel 2012].

A large progress has been achieved in the recent decades: concepts like self-organization, emergence, complex networks, pattern formation, criticality, deterministic chaos, synchronization etc were conceptualized and constitute a powerful conceptual background that is currently being exported from physics to other disciplines [Badii 1999].

However, there is indeed a urgent need to reduce the gap between pure and applied science, hoping to manage and control systems with levels of complexity exceeding the capacity of current approaches. A new generation of scientist should emerge, trained to understand how complex systems behave, how to live with them, to control them and to design them well. Examples are countless: in the natural sciences we may just mention genetic networks, bio-molecules, habitats and ecology, weather and climate among many. In technology one has to face complex hierarchical networks in distributed computing, complex materials and process control. In social sciences and economics problem like transport management, risk evaluation epidemic spreading, social networks etc. will all benefit from a deeper knowledge of Complexity Sciences.

Big data? Big theory! The need for understanding

In all domains, complex systems are studied through increasingly large quantities of data, stimulating revolutionary scientific breakthroughs. Those developments will pose new and

1 Tentative title and acronym

fundamental theoretical issues to scientists that are to face such *Data Deluge*. However, data, whether produced by massive simulations or automated experimentation, is not understanding. [Crutchfield2014]. The impressive advances in computing power and data acquisition technologies have led some to claim that theory is somehow obsolete: large-scale simulations can now include all system components and so are complete: there's no need to suffer the inevitable simplifications theory requires. This point of view is questionable: Complexity science needs to go beyond *data literalism* ("data describes Nature") and *computationalism* ("a computer code is the theory of the phenomenon it simulates"). Though [Crutchfield2014],

[...] there is now a new and very real possibility for a novel synthesis of advances in experimental technique, high-performance computing, and theory: automatically building theories from data. That is,to the extent we understand pattern, we can use machines to find emergent organization in our vast data sets. And from there, the machines can build our theories, most likely with guidance from a new generation of theorists.

Altogether this calls for a critical, balanced interplay experiment, computing, and theory will be required where *Big Theory, g*iven its critical role in understanding such complex systems, deserves a place.

Inference: learning from data

The recent success of machine learning as implemented through deep neural networks has been embraced with enthusiasm. Without doubt, the impact of such recent achievements on our societies and lives cannot be underestimated [Mezard 2018]. But in spite of practical successes, it is recognized that the underlying functioning mechanisms are largely not understood. For physicists expert in spin glasses, the energy landscape is unexpectedly smooth. Indeed, the good performance can be explained only in a nonequilibrium statistical physics framework: regions of the optimization landscape exist that are both robust and accessible and that their existence is crucial to achieve good performance on a class of particularly difficult learning problems [Baldassi 2016].

This is by no means only an academic issue: without a more deep understanding we are not guaranteed that any deep network will always perform the task for which it was trained. This poses obvious problem and challenges in a world were sensible decision may be taken based on such inferences.

Limitations in learning from data arise also when considering the analogy between derivation of macrolaws from microscopic in statistical physics and statistical inference from data [Moore2018]. Since data are necessarily noisy and incomplete, there could indeed be a phase transition that hinders the inference process. When the amount of noise in a data set crosses a critical threshold, it can suddenly become impossible to find underlying patterns in it, or even tell if a pattern is really there. This includes finding communities in social and biological networks, or clusters in high-dimensional data, or structure in noisy matrices and tensors. How can we locate these phase transitions, and design algorithms that perform as well as possible? What informational and computational barriers do these transitions create? Once more only a deeper insight of the theory behind will be needed.

Predicting, forecasting, modeling

Complexity unavoidably leads to large fluctuations and managing the resulting uncertainty is a major challenge: to cope with it we need to perform forecasts. Also we need to understand what are the limitation and applicability ranges of our predictions.

Recently, availability of huge data set revived the inductivist approach, which just relies on knowledge of the past, basically by finding *analogues*. The underlying belief is that *big data* will lead to *much better* forecasts. However, this point of view can be in practice useless: the required analog, whose existence is guaranteed in theory, sometimes cannot be expected to be found in practice, even if complete and precise information about the system is available. In fact, the mean recurrence time is *exponentially large* in the system phase-space dimension so that in practice, a recurrence is never observed [Hosni 2018] and this approach may be of little use.

The case of weather forecast is also instructive: in principle, future weather can be predicted by solving the proper partial differential equations with initial conditions given by the present state of the atmosphere. But, they are too accurate: they also describe high-frequency wave motions that are irrelevant for meteorology. A much more useful approach is to construct *effective equations* eliminating irrelevant (fast) variables. Such equations have great computational advantages making the numerical computations satisfactorily efficient. But, even more importantly, capture the essence of the phenomena of interest, which could otherwise be hidden in too detailed a description, as in the case of the complete set of original equations. Thus the effective equations are a form of *clever reductionism:* they are not mere approximations but rather require a subtle mixture of hypotheses, theory and observations [Baldovin 2019].

In the study of modeling and prediction of natural, as well social, phenomena, a crucial aspect is the understanding of the *cause-effect relationships* among different variables; perhaps the most popular example is, in the context of the climatic change, the debated link between temperature and CO_2. Today it is well known that often the first natural approach in terms of correlations is not enough accurate and can produce paradoxical conclusions [Pearl 2000], therefore it is necessary to use more sophisticated approaches. Looking at causality as a *flow of information* among processes one can use mathematical methods based on ideas from dynamical systems and information theory [Schreiber 2000, Palus et al 2007].

Altogether, the above considerations can be summarized in a word by stating that modeling is an art that is necessary to master to cope with Complexity.

Inferring connectivities and dynamics

Emergent dynamical properties are intimately related to the topology of the underlying network of connections among the constituent parts of the system. This concept, which is central in so called *network science* leads to the issue of the inverse problem, namely how to infer the connectivities from data and use them in network model building. To mention a concrete case, one may wish to reconstruct the distribution of connectivity of a neural network from measurement of global neural activity measurements [Adam 2019]. As discussed above, reductionism calls for a suitable modeling, with previous insight of relevant variables: in the above case, one may start from a theoretical model (e.g. in this example the "Heterogenous Mean-Field approximation") and use it as a criteria try to organize the neurons in different classes, depending on their associated degree and current. The resulting connection can be than back-tested against further measurements.

A challenging perspective regards the reconstruction of the dynamics: to what extent one can use sophisticated approaches like for instance *reservoir computing* to perform predictions and attractor reconstruction of chaotic dynamical systems from time series data ? Theoretical framework that describes conditions under which reservoir computing can create an empirical model capable of skillful short-term forecasts and accurate long-term ergodic behavior are being investigated [Lu 2018]. It is presumable that those approaches will be central in the future.

Challenges

Even from a short and partial account like the one presented in the present document, it is easy to realize that the field is so vast that it would be a challenge by itself to outline the main research objectives. We mention here three possible keywords/themes for the future Working Group activity:

- *Predictability*: how to improve it to the limits and formulate new approaches to prediction, forecasting and risk? How to infer causal relations between observables?
- *Modeling*: moving from data to dynamical models; how to develop a more integrated approach to and understand fundamental limitations? How can we construct effective models that can help prediction and understanding?
- *Control and management*: how to use the acquired knowledge to manage a system properties in an open and changing environment?

The above challenges are general and can be formulated in different scientific and technological research context. This also requires finding *common languages* among scientist with diverse background, a task that is by itself difficult and requires close interaction and sharing of knowledge and expertise. This will one of the aims of the future events to be organized within the Foresight project.

Stefano Lepri

Consiglio Nazionale delle Ricerche - Istituto dei Sistemi Complessi

stefano.lepri@isc.cnr.it

References

[Adam 2019] I. Adam,G. Cecchini,D. Fanelli, T. Kreuz, R. Livi, M. di Volo, A Mascaro, E. Conti, A.Scaglione, L. Silvestri, F. S. Pavone, *Inferring network structure and local dynamics from neuronal patterns with quenched disorder*, preprint arXiv:1910.05761 (2019)

[Badii 1999] R. Badii, A. Politi, *Complexity: hierarchical structures and scaling in physics*, Cambridge University Press (1999)

[Baldassi 2016] C. Baldassi, C. Borgs, J. T. Chayes, A. Ingrosso, C. Lucibello, L. Saglietti, R. Zecchina, *Unreasonable effectiveness of learning neural networks: from accessible states and robust ensembles to basic algorithmic schemes*, Proceedings of the National Academy of Sciences 113 (48), E7655-E7662 (2016)

[Baldovin 2018] M. Baldovin, F. Cecconi, M. Cencini, A. Puglisi, A. Vulpiani, *The role of data in model building and prediction: a survey through examples*. Entropy 20, 807 (2018)

[Crutchfied 2014] J.P. Crutchfield, *The dreams of a theory*, *WIREs Comput Stat* , 6:75–7 (2014)

[Hosni2018] H.Hosni, A. Vulpiani, *Forecasting in the light of big data*, Philosophy and tecnology, 31, pages 557–569(2018)

[Lu 2018] Z. Lu, B. R. Hunt, E. Ott, *Attractor reconstruction by machine learning* Chaos: An Interdisciplinary Journal of Nonlinear Science 28 (6), 061104 (2018)

[Moore 2017] C. Moore, *The computer science and phyics of community detection: landscapes, phase transitions and hardness complexity*, preprint arXiv:1702.00467 (2017)

[Mezard 2018] M. Mezard, *Artificial intelligence and its limits*, Europhysics News 49/5&6, (2018).

[Pearl 2000] J. Pearl, *Causality*, Cambridge Univ. Press (2000)

[Schreiber 2000] T. Schreiber, *Measuring information transfer* Phys. Rev. Lett. 85(2), 461 (2000).

[Palus 2007] M. Palus and M. Vejmelka, *Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections*. Phys. Rev. E 75, 056211 (2007)

[SanMiguel 2012] San Miguel et al. *Challenges in complex systems science*, Eur. Phys. J. Special Topics 214, 245–271 (2012)