

Integrating induction and deduction in order to complement experiments and priors for the understanding, forecast, control and foresight of apparently complex processes in internet of Beings.

Diego Liberati

National Research Council of Italy

### Introduction

Biology and biomedicine is lesser and lesser a matter of evidence, letting so called "clinical eyes" implicitly infer causal reasons behind epiphenomena, nor just a matter of axiomatic deduction, being each of us - getting elderly is itself an illness, as already Lucius Annaeus Seneca - Nero's praeceptor - was stating (but the alternative is to die young [my maternal grand father, personal communication, since 1960];) not just a corollary of a theorem, thus a case of a certain illness, but a very special individual with one's own DNA (except for monozygote twins, but then epigenetic is anyway differentiations ever since, being even the position in uterus - even if a mother would psychologically feel identified with all her children since conception to death :), sometimes, and on average more often getting older, malfunctioning in one's own way, thus needing a "LEGO co-player with God" to

help to repair: to that, one should have a design, like a scientist, or at least an engineer, not just try and fail!

This century fashion of inference from data does often understand how things are going on and even how to control them via mimicking our neural approach of learning by examples. One of the main drawbacks, besides time and computation power needed in the learning phase, is that often even deep learning does learn how to do, but is less good than us in then explaining why. Logical Networks (Muselli & Liberati Ieee Trans KDE 2002) do overcome such drawbacks by inferring in the canonical OR of ANDs form of electrical circuits fashion (Muselli & Liberati IEEE Trans CAS I 2000), namely in deduction form, then easily keen to introduce priors just a posteriori modifying inferred rules (think about inferring the need to run at no more than 45.96 Km/h to avoid ticket, it could probably a posteriori set at 50 by enforcing an obvious prior to whom the inference was asymptotically tending).

When data are available not just as single shots but in time course, a Piece Wise Affine identification approach within Hybrid dynamic and logic framework does generalize the above, by optimally cutting even complex non linear hysteretical multivariable processes in pieces of almost linear behavior among almost

stationarity borders (Ferrari Trecate, Muselli, Liberati & Morari, Automatica 2003). Rain attenuation in mobile communication could for instance be easily taken into account in this way (Formentin, Luin, Capsoni, Nebuloni & Liberati, 8th Advanced Satellite Multimedia Systems Conference 2016)

When within each linear piece, even simple approaches like principal components, carefully revisited, could then allow to indentify a hierarchy among the conditioning factors, by considering each salient factor influence on the main discriminating principal component: in this way, even famous discriminations about pathologies (Golub et al Science 1999) could easily be outperformed (Garatti, Bittanti, Liberati & Maffezzoii, Intelligent Data Analysis 2007) even discovering classification errors, or hypothesis could be tested as not falsified by key experiments, as in confirming a new WNT pathway in Leukemia discrimination (Grassi, Liberati et al., Frontiers in Oncology 2019)

A powerful general purpose profolio of tools, thus helping in fostering not just thing-to-thing interacion, but, though things, among human Beings.

A couple of ongoing applications are worth to be mentioned as advanced work in

progress already mature for publication but still in the last review phases before final submission:

### Community analysis in interactomic regulation networks

The powerful though simplistic idea that every gene just codifies for its own protein, then free to interact within the cell, is nowadays over ever since. It would not account for most of the well-known epigenetic properties, modulating gene expression to the different context within, for instance, different organs, in different individuals

Networks of interactomics actors, including genes and codified proteins themselves, are nowadays known, though not yet all completely identified, as responsible for the exhibited complex para-social interaction resulting in the beautiful diversity, within similarity, of key factors of life

The same happened many years ago at a different scale within the central nervous system neuroscience, when the so-called "grandmother" neuron, formerly believed to be responsible for memorizing the beloved, was substituted by the task-recruited neural network including several actors, each of which also in turn still available to contribute to other tasks within (partially) different other networks.

Still the same, at an even bigger scale, is every day's everybody's social multi-interactions experience of each homo oeconomicus of us.

In order to investigate such a kind of networks, a simple but powerful idea, as proved by Google usefulness and consequential success, is to investigate the ranking of each actor relationships to each other, being such actor either an internet page, as in the original ... Page algorithm, or a gene or codified protein in our case.

It is worth noticing, as pointed out in a seminal review by Vidyasagar (2011), that the recent randomized approach introduced by Ishii and Tempo (2010) would be the technological key to drastically reduce, at

the cheap price of a limited loss of precision, the overwhelming computational complexity that would prevent to apply Page ranking to the analysis of every significant -omic network, besides the almost-toy subnetworks already investigated, as for instance in Hannah & Zaki (2014)

A possible complementary approach, proposed by Landi & Piccardi (2014), could be taken to our -omic regulation network. As a public available benchmark, the data (Gavin et al 2006) used in the reported work by Zaki (...) can be used, in order to investigate which features of alternatively proposed approach are possibly useful as a complement to even improve the already powerful Page approach on one side, or eventually able to surrogate it in a less costly way, even taking into account the recalled randomized economy. A probably even better solution will be some improvement on randomized Page Ranking we are willing to make also in memory of our beloved colleague Roberto Tempo (Ideaki Ishii, personal communication, 2019)

A bachelor thesis under our co-supervision with professor Piccardi has been undertaken by Dr Nicoletta, now postdoc in Canada, and a colleague, over a couple of public datasets, as described in Gavin et al. (...), already used also by Hannah & Zaki (2014) to test alternative approaches to Page ranking. They describe proteic interactions in *Saccharomyces cerevisiae*, namely:

PPI-D1, obtained via mass spectrometry, includes 1430 proteins linked by 6531 interactions

PPI-D2 combines 6 different experiments with hybrid techniques, including 3869 proteins linked by 23399 interactions.

The recorded interactions are weighted in order to sharpen information, by pruning the non trustable information via adjusted weights, under the assumption that proteins with the same neighborhood usually should share similar functions. For the two

considered data sets, such correction yields, for PPI-D1, in 990 proteins with 4687 interaction, while for PPI-D2 the network results in 1443 proteins with 6993 interactions.

A few approaches fully described elsewhere have been considered. For the sake of completeness some of their salient features are nevertheless recalled within the next few sentences

Performances of the various approaches are evaluated and compared in the bi-dimensional space of Recall.....and Precision.... taking into account a combination of positive and negative false rejection, namely type A and B statistical errors.

By considering Precision and Recall for several approaches on the used data sets, among the 4 approaches reminded in the following, IOCP generally deserves good results, yielding the subsets of salient communities confirming our priors.

-omic regulation networks can thus be seen as a special case of the more general network community analysis, quite popular nowadays in the interdisciplinary field of complex systems, at the edge, among other disciplines, of physics, control theory, operating research.

4 approaches have mainly be taken into account: two of them, Louvain and LMC, are based on network partition, while the one, IOPC, proposed by Landi & Piccardi and based on a quasi-local search allowing partial superposition of communities, looks more general, usually outperforming the 2 other approaches as for both Recall and Precision. The fourth approach is the Page ranking, possibly randomized, sometimes comparable to IOPC, as for instance on one of the 2 used test networks, while for the other one the loss of Recall with Page rank is dramatic.

Among such 4, themselves considered the best ones among the many other approaches available in literature, the very best, on the pair of benchmarks employed, appears to be IOPC, very strict for false positive, thus even improvable by pruning or merging the smallest found communities.

Other further possible improvements may imply to take into account the time evolution of the interaction graph (like PRISM) in order to also take into account the dynamics of interactions. This would allow to deal not only with the identification of the salient variables, but also with their dynamical interplay, thus leading to the so-called path defining the destiny of the considered -omic interaction

Improvements in either Louvain or LMC, in the direction not forcing anymore partitions, could make their performances comparable with IOPC, while Page ranking, already often not worse than IOPC, is the simplest to be randomized, offering thus advantages in speed, as well as keen to be improved (Ishii personal communication, 2019).

All the offered results are in some sense qualitative (Balestra e Liberati 1993), or at most semi-quantitative, offering figures of merit to judge of their usefulness : they provide the involved actors, but not the (dynamical) weights balancing their interaction

To become fully quantitative, one should resort to ideas like the ones able to mine for instance linear (Garatti et al 2007), binary (Muselli & Liberati 2002), inferential (Bosin et al., 2005) or even hybrid (Ferrari Trecate et al, 2003) quantitative models learned from richer data, weighting, even dynamically, the arcs of the interactomic graphs, and thus able to generalize.

## Lotka & Volterra once more revisited

A second even more mature example - thus even shorterly reported in order not to spoil imminent publication, is relative to a different scale, not anymore gene protein interaction within cells but interindividual and machine to individual social interaction.

The peculiar society we are talking for is the nowadays fashionable so called Industry 4.0, and the investigated relationship is the human-robot interaction: again competition (does God play dice no matter of us?) or cooperation (does Him instead play LEGO WITH us ?), or something in between?

Looking at data across 2010, thus encompassing the big 2008-2009 world crisis, and writing a simple prey(s)-predator(s) model with some degree of freedom [Fantini et al, submitted] one could induce-deduce that in Korea, Japan, USA and Germany robots are for humans like cows: they eat some grass (less qualified jobs) giving in exchange meat (more qualified jobs, sorry vegetarians ;)

Then robots do not seem to destroy occupation, just pushing to a more qualified one, thus implying the need to re-qualify loyal employees, AND hire junger and more qualified ones, but even impaired, more and more able to do sophysticated tasks thanks to new techs, like for instance even Brain Computer Interfacing, and working more relaxed, thanks to for instance stress monitoring via Hearth Rate Variability analysis, AND reinvest the increased gain in new qualified manpower instead of escape with the sudden money like unfortunately in



Italy, where less mature family capitalism is often blind or myopic, accumulating or even dilapidating for family and friends, instead of physiological reinvesting for the sake of society.

## Discussion

Technology will thus be keeping doing its traditional role thus of social elevator, in a less direct, but maybe even more interesting and anyway actual and forced way than nowadays, like University, deeply involved in the re-qualification process.

Also University in fact, even since social elevator within families - parents invested in more cultured children to give them a better future also in terms of income - is becoming social elevator within Countries: our best students including my best pupils, are more and more often getting prestigious jobs elsewhere - meaning London Switzerland, Germany, USA, French: brain drain not desperate emigration; on the other side, bright pupils of less favorite countries - say for instance Serbia where we have just being offered to serve as scientific attaché at Italian Ambassade - do work remotely in their country at a salary in between, in a win-win approach helping them to get more and italian little companies to spend less thanks to the

vision by His Excellence Ambassador Lo Cascio and coworkers

Internet of Beings thus, even more and better than so called internet of things: what really matter is not just machine to machine interaction, but human to human interaction also through machines and Industry 4.0: cooperation among humans, instead of competition, is the secret, like ants, neurons, proteins, genes have already understood ever since, evolutionary living that Gospel we sometimes worship then dumbly disattend in our daily life: let us finally start evolving, by learning from nature to cooperatively play LEGO instead of competitively loose playing dice!

### References

Muselli & Liberati: Binary rule generation via Hamming Clustering, IEEE Transactions Knowledge and Data Engineering, 14(6): 1256-1268, 2002

Muselli & Liberati: Training digital circuits with Hamming Clustering, IEEE Transactions on Circuits And Systems I: Fundamental Theory and Applications 47(4): 513-527, 2000

G Ferrari Trecate, M Muselli, D Liberati & M Morari, A clustering technique for the identification of piecewise affine

systems, *Automatica* 39 (2), 205–217, 2003

Formentin, Luin, Capsoni, Nebuloni & Liberati, 8th Advanced Satellite Multimedia Systems Conference, Mallorca, 2016

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* 15;286(5439):531–7, 1999

S Garatti, S Bittanti, D Liberati & A Maffezzoli, An unsupervised clustering approach for leukemia classification based on DNA micro-arrays data, *Intelligent Data Analysis* 11(2): 175–188, 2007

Susanna Grassi, Sara Palumbo, Veronica Mariotti, Diego Liberati, Francesca Guerrini, Elena Ciabatti, Serena Salehzadeh, Claudia Baratè, Serena Balducci, Federica Ricci, Gabriele Buda, Lorenzo Iovino, Francesco Mazziotta, Francesco Ghio, Giacomo Ercolano, Antonello Di Paolo, Antonela Cecchettini, Chiara Baldini, Letizia Mattii, Silvia Pellegrini, Mario Petrini, Sara Galimberti: the Wnt Pathway is relevant for the BCR-ABL-1 independent resistance in Chronic Myeloid Leukemia, *Frontiers in Oncology* 9:e532-on, 2019

M. Vidyasagar, "Probabilistic methods in cancer biology," *European Journal of Control*, 17(5-6), 483-511, September-December 2011

Ideaki Ishii and Roberto Tempo: Distributed Randomized Algorithms for the

PageRank Computation, IEEE TRANSACTIONS ON AUTOMATIC CONTROL 55(9):1987-2002,

2010

Hanna, E.M., Zaki, N. Detecting protein complexes in protein interaction networks using a ranking algorithm with a refined merging procedure. BMC Bioinformatics 15: 204, 2014

Pietro Landi, Carlo Piccardi: Community analysis in directed networks: In-, out- and pseudo-communities, Physical Review E 89(1): 012814, 2014

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, et al: Proteome survey reveals modularity of the yeast cell machinery. Nature 440: 631-636, 2006

Ideaki Ishii, personal communication, 2019

Fantini P, Maffezzoni P., Daniel L., Liberati D. & Taisch M: submitted, 2019