Abstract Booklet

CHALLENGES IN DATA SCIENCE: A COMPLEX SYSTEMS PERSPECTIVE

13-18 NOVEMBER, 2023 POLITECNICO DI TORINO

DASC 2023

Title: A neuroscience-inspired approach to large-scale communication network monitoring

Author(s): Luc Berthouze
Affiliation(s): University of Sussex (UK)

Abstract: Much of our lives today, both at work and at home, relies on the continuous and reliable provision of ICT infrastructure. Network outages are extremely disruptive and very costly. However, preventing outages (or indeed, identifying their cause), is extremely challenging because of how complex and large today's infrastructures have become. With services, traffic workloads and user needs ever-growing, new approaches are needed. In this talk, I will discuss how we are leveraging and adapting methods typically used to understand another large-scale complex system -- the brain -- to (a) detect unknown, unseen and continuously changing inter-dependencies and (b) exploit this knowledge to develop scalable and data-driven approaches for anomaly detection and root cause analysis in both large-scale networks and microservice architectures.

Keywords: Communication Networks, Functional Connectivity, Root-Cause Analysis

Author Profiles (s): https://scholar.google.com/citations?user=WvYeh5IAAAAJ&hl=en&oi=ao

Title: The transition to synchronization of networked systems

Author: Stefano Boccaletti
Affiliation: CNR - Institute of Complex Systems, Via Madonna del Piano 10, Sesto Fiorentino, Italy

Abstract: I will show that the transition to synchronization of a generic networked dynamical system is a feature that only depends on the topology of the network's connections and can be entirely predicted and completely characterized with the only help of eigenvalues and eigenvectors of the graph's Laplacian matrix. In particular, the transition is made of a well-defined sequence of events, each of which corresponds to either the nucleation of one (or several) cluster(s) of synchronized nodes or to the merging of multiple synchronized clusters into a single one. The network's nodes involved in each of such clusters can be exactly identified, and the value of the coupling strength at which such events are taking place (and therefore, the complete events' sequence) can be rigorously ascertained. I will moreover clarify that the synchronized clusters are formed by those nodes which are indistinguishable at the eyes of any other network's vertex, and as so they receive the same dynamical input from the rest of the network. Therefore, such clusters are more general subsets of nodes than those defined by the graph's symmetry orbits, and at the same time more specific than those described by network's equitable partitions. Finally, I will present large-scale simulations which show how accurate are our predictions in describing the synchronization transition of both synthetic and real-world large size networks, and we even report that the observed sequence of clusters is preserved in heterogeneous networks made of slightly non-identical systems.

Keywords: Synchronization; Complex Networks; Phase transition.

Author Profile: https://scholar.google.com/citations?user=BEC76f4AAAAJ&hl=it&oi=ao

Title: Leveraging Social Contagion to Foster Consensus in Collective Decision-Making

Author(s): Mohammad Savari[1], Nikolaj Horsevad[2], and Roland Bouffanais[1,3]
Affiliation(s): (1) University of Ottawa, Ontario, Canada; (2) Energinet, Eritsø, Denmark; (3) University of Geneva, Switzerland

Abstract: Appropriate social transfer of information among units of a multi-agent system (MAS) is a prerequisite for an effective collective response to changing environmental conditions. From the network perspective, this social information transfer requires understanding the interplay between network topology and agents' dynamics [1]. Specifically, information propagation through the MAS can either take the form of a simple contagion—associated with pairwise interactions—or a complex contagion—involving social influence and reinforcement [2, 3]. The key role played by the network topology in this social information transfer has been acknowledged [1]. In that work, a nontrivial relationship between the pace of external perturbations and the network degree is reported. Subsequently, Horsevad et al. [2] revealed the possibility of complex contagion with a leader-follower consensus model of distributed decision-making lacking thresholds and/or nonlinearities. Prior to that work, complex contagions were limited to decision-making models based on a binary decision variable with a threshold [3]. Reference [2] highlights that other network properties, beyond the degree distribution, influence the social contagion process. The existence of a transition from a simple contagion to a complex one hinges on knowing which network property plays a key role. One serious limitation of these works is the lack of a systematic way of characterizing the type of social contagion for a given collective decision-making protocol. What has been found true for the first-order leader-follower consensus might not hold for other forms of distributed decision-making. For instance, the Kirchhoff index and clustering coefficient may not be the appropriate metrics to decipher which type of social contagion is unfolding. It is worth stressing that these network metrics only incorporate features of the network topology without accounting for the agents' dynamics taking place over this network. Here, we propose a novel approach based on spectral graph theory to address this issue by considering a generalized metric that would embody network topology along with agents' dynamics. Specifically, the spectrum of the graph Laplacian offers valuable information about both network structure and agents' dynamics. The eigenvalues of the Laplacian matrix have been used for community detection and spectral clustering. As a matter of fact, the spectrum of the graph Laplacian can reveal information about both global and local properties of the network, such as the number of connected components, clustering coefficient, and spectral gap, etc. Furthermore, the Kirchhoff index can be expressed as the sum of the inverse of the eigenvalues. Also, different parts of the spectrum can be associated with community structures, motif multiplication, and bipartiteness of the network graph. This approach has the potential to extend our results to any collective decision-making protocol beyond the simple leader-follower consensus.

1. Mateo, D., Horsevad, N., Hassani, V., Chamanbaz, M., & Bouffanais, R. (2019). Optimal network topology for responsive collective behavior. *Science Advances*, 5(4), eaau0999.
2. Horsevad, N., Mateo, D., Kooij, R. E., Barrat, A., & Bouffanais, R. (2022). Transition from simple to complex contagion in collective decision-making. *Nature Communications*, 13(1), 1442.
3. Centola, D., Eguíluz, V. M., & Macy, M. W. (2007). Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications*, 374(1), 449-456.

Keywords: Collective decision-making, Social contagion, Distributed consensus.

Author Profiles (s): R. Bouffanais: Scholar, Linkedin, ORCID

Title: How to open the lid in complex systems - from telecommunication to electrophysiology

Author: Teodor, Buchner (1)
Affiliation(s): Faculty of Physics, Warsaw University of Technology (1); EXATEL S.A. (2)

Abstract: As prominent members of energy preserving (i.e. lazy) mankind, we tend to use/abuse intellectual cliches to project any new knowledge onto our semantic space. This reduces cognitive dissonance and brings us back to safety. I would like to refer to two distinct examples in the area of complex systems.

In bioengineering, we tend to use equivalent circuits, which are treated as representation of many biological, chemical and physical processes. They ease interpretation of any process up to the level of absurd. If any voltage drop can be represented as a resistor, what does it actually explain from the rich physical reality? Here I show our recent results on propagation of ECG as an electromagnetic wave. They show what can you achieve if you move only one step deeper in theory, to consider a complex nature of human tissue, instead of a typical assumption, that a man is made of metal. It is also important to remember words of caution expressed by Kenneth Cole and Alexander Mauro, on how not to abuse permittivity spectra. We get to conclusion, that complex system physics as an important tool to understand the reality, and that interpretation of large data using domain knowledge always has advantage over plain AI methods.

In telecommunication, in turn, we (complex systems physicists) have a tendency to treat internet as a huge, flat, evolutionary network. I argue, that if you ask network engineers, they rather see a network of networks, which are present in many different layers. Some of them, indeed, represent evolutionary networks, where physics of complex systems should be applied in order to find answer of well-posed and important business problems. In telecommunication we definitely know the answer, but we are still looking for good questions and good areas of application, that would be acknowledged by network traffic engineers. It is not about answering our own questions. It is about the huge amounts of data, which should be opened in a meaningful way. I also refer to presence of specific traffic patterns, which are important both for traffic engineering and for cybersecurity.

Keywords: Electrocardiography (1); Impedance spectra (2); Telecommunication network (3).

Author Profiles (s):   https://www.linkedin.com/in/teodor-buchner-163a272
https://scholar.google.com/citations?user=rnwu4gQAAAAJ&hl=en
https://orcid.org/0000-0003-0030-3194
https://www.researchgate.net/profile/Teodor-Buchner

Title: Reconstruction of horizontal wind fields from spaceborne doppler radar data using a continuity equation constraint

Author(s): Cinzia Cambiotti (1), Sérgio da Silva (1), Anna Carbone (1), Alessandro Battaglia (2)
Affiliation(s): DISAT, Politecnico di Torino, Turin, Italy (1); DIATI, Politecnico di Torino, Turin, Italy (2)

Abstract: In this work, we present a data inversion methodology for reconstructing 2D horizontal wind fields, analyzing measured line-of-sight velocity data obtained from a conically scanning doppler radar system envisaged as payload of the ESA Earth Explorer 11 WIVERN (WInd VElocity Radar Nephoscope, https://wivern.polito.it/) [1,2]. We employ the least squares approach while imposing constraints dictated by the continuity equation to obtain physically consistent and interpretable results. The physics associated with the continuity equation is included as a regularization term for the data-inverse problem. Our results highlight that care should be taken to determine an optimal regularization parameter to balance data constraint and data fidelity. The results also reveal the effectiveness of our methodology in satisfactorily reconstructing horizontal wind fields, thereby contributing to the advancement of data science applications in spaceborne radar remote sensing.

[1] Illingworth, A. J., Battaglia, A. et al., 2018: Wivern: A new satellite concept to provide global in-cloud winds, precipitation and cloud properties. Bull. Amer. Met. Soc., DOI: 10.1175/BAMS-D-16-0047.1, 1669-1687.

[2] Battaglia, A., Martire, P., Caubet, E., Phalippou, L., Stesina, F., Kollias, P., Illingworth, A., 2022: Observation error analysis for the WInd VElocity Radar Nephoscope W-band Doppler conically scanning spaceborne radar via end-to-end simulations, Atmos. Meas. Tech., DOI: https://doi.org/10.5194/amt-15-3011-2022

Keywords: Data inversion; Spaceborne radar; Continuity equation.

Author Profiles (s): Cinzia Cambiotti : https://www.polito.it/personale?p=cinzia.cambiotti
Sérgio da Silva : https://scholar.google.com/citations?user=PznkCCsAAAAJ&hl=it
Anna Carbone : https://scholar.google.com/citations?hl=it&user=INBplj8AAAAJ
Alessandro Battaglia : https://scholar.google.com/citations?hl=it&user=Ob6UR_oAAAAJ

Title: From Quantum Picturalism to Interpretrable Quantum AI

Author(s): Bob Coecke
Affiliation(s): Quantinuum & Perimeter Institute

Abstract: In 2020 our Oxford-based Quantinuum team performed Quantum Natural Language Processing (QNLP) on IBM quantum hardware [1,2,3]. Key to having been able to achieve what is conceived as a heavily data-driven task, is the observation that quantum theory and natural language are governed by much of the same compositional structure [4,5,6] – a.k.a. tensor structure. Hence our language model is in a sense quantum-native, and we provide an analogy with simulation of quantum systems in terms of algorithmic speed-up. Meanwhile we have made all our software available open-source, and with support [7]. We will also introduce the notion of compositional intelligence, exploiting the fact that the compositional match between natural language and quantum extends to other domains as well, such as patio-temporal perception [8], we will argue that a new generation of AI can emerge when fully pushing this analogy. The so-called ZX-calculus [9,10] for quantum theory (and linear algebra more generally) has been proven to be complete, so can be conceived as a full-bodied reasoning system that go hand-in-hand with modern machine learning.

1. B. Coecke, G. de Felice, K. Meichanetzidis, and A. Toumi. Foundations for near-term quantum natural language processing, 2020. arXiv:2012.03755. doi:10.48550/arXiv.2012.03755

2. K. Meichanetzidis, A. Toumi, G. de Felice, and B. Coecke. Grammar-aware question-answering on quantum computers. arXiv:2012.03756, 2020. doi:10.48550/arXiv.2012.03756

3. R. Lorenz, A. Pearson, K. Meichanetzidis, D. Kartsalkis, and B. Coecke. Qnlp in practice: Running compositional models of meaning on a quantum computer. arXiv:2102.12846, 2021. doi:10.48550/arXiv.2102.12846

4. S. Clark, B. Coecke, and M. Sadrzadeh. A compositional distributional model of meaning. In Proceedings of the Second Quantum Interaction Symposium (QI-2008), pages 133–140, 2008. url:http://www.cs.ox.ac.uk/people/stephen.clark/papers/qai08.pdf

5. B. Coecke, M. Sadrzadeh, and S. Clark. Mathematical foundations for a compositional distributional model of meaning. In J. van Benthem, M. Moortgat, and W. Buszkowski, editors, A Festschrift for Jim Lambek, volume 36 of Linguistic Analysis, pages 345–384. 2010. arXiv:1003.4394. doi:10.48550/arXiv.1003.4394

6. S. Clark, B. Coecke, E. Grefenstette, S. Pulman, and M. Sadrzadeh. A quantum teleportation inspired algorithm produces sentence meaning from word meaning and grammatical structure. Malaysian Journal of Mathematical Sciences, 8:15–25, 2014. arXiv:1305.0556. doi:10.48550/arXiv.1305.0556

7. D. Kartsaklis, I. Fan, R. Yeung, A. Pearson, R. Lorenz, A. Toumi, G. de Felice, K. Meichanetzidis, S. Clark, and B. Coecke. lambeq: An efficient high-level Python library for quantum NLP. arXiv:2110.04236, 2021. doi:10.48550/arXiv.2110.04236

8. V. Wang-Mascianica and B. Coecke. Talking space: Inference from spa- tial linguistic meanings. Journal of Cognitive Science, 22(3):421–463, 2021. doi:10.48550/arXiv.2109.06554

9. B. Coecke and A. Kissinger. Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning. Cambridge University Press, 2017. doi:10.1017/9781316219317

10. B. Coecke, D. Horsman, A. Kissinger, and Q. Wang. Kindergarden quantum mechanics graduates... or how i learned to stop gluing LEGO together and love the ZX-calculus. Theoretical Computer Science, 897:1–22, 2022. arXiv:2102.10984. doi:10.48550/arXiv.2102.10984

Keywords: Quantum Picturalism, Compositional Intelligence

Author Profiles (s): https://twitter.com/coecke
https://scholar.google.ca/citations?user=fO17CXgAAAAJ&hl=en&oi=ao
https://www.linkedin.com/in/bob-coecke-9389627/

Title: Non-Gaussian models for high-frequency returns

Authors: Federica, De Domenico[1,2]; Giacomo, Livan[1]; Guido, Montagna[1,2]; Oreste, Nicrosini[2].
Affiliations: [1]University of Pavia; [2]Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, Pavia, Italy.

Abstract: It is well known that the probability distribution function of high-frequency returns is characterized by a leptokurtic, heavy-tailed shape. Yet there is no general consensus on which distribution should be adopted to describe them, since different models have demonstrated, to varying extent, an ability to reproduce empirically observed stylized facts. The most important examples that fit empirical data are the Truncated Lévy, the Student's t, the q-Gaussian and the Modified Weibull distribution. In order to provide some clarity, we perform a comparative study of the aforementioned non-Gaussian models. To introduce a coherent framework, each distribution is reported to zero mean, unit variance and evaluated over a bounded range to mimic realistic outliers. Given the above setting, a noticeable consistency emerges as they show a similar leptokurtic shape and a quite good agreement along the tails. We generate large samples of non-Gaussian random numbers as synthetic data representations of financial fluctuations. We then simulate their dynamical evolution and analyze their statistical properties, underlying differences with respect to the normal scenario. In particular, we focus on the convergence rate to the asymptotic distributions and, in order to further investigate the behavior of the outliers, we compare the complementary distribution functions and the kurtosis. As a first example of applications, we study the impact of our results on option pricing. In particular, we implement a comparison between Gaussian and non-Gaussian fluctuations for plain vanilla and path-dependent options for short and long maturities.

Keywords: High-frequency models (1); simulations (2); option pricing (3).

Author Profiles: www.linkedin.com/in/federica22

Title: Opinion dynamics model revealing yet undetected cognitive biases

Author(s): Guillaume Deffuant (1) (2)
Affiliation(s): Université Clermont-Auvergne, Inrae, UR LISC (1); Université Clermont-Auvergne, LAPSCO (2)

Abstract: This talk is about opinion dynamics models suggesting the existence of yet unknown cognitive biases. Indeed, the mathematical analysis establishes that the emergent patterns of the model are due to specific biases in self-evaluation and in evaluation of others. In particular, the positive bias in self-evaluation identified in the model is different from the positive bias identified in the literature in social psychology. We describe a specifically designed experiment that detects this specific bias in human subjects, for the first time. We finally discuss the role of the model in this work, which is revealing phenomena that are almost impossible to imagine without its simulations.

Keywords: Opinion dynamics, moment approximation, positive bias, self-enhancement, experiment.

Author Profiles (s): https://scholar.google.fr/citations?user=xwYTdNwAAAAJ&hl=fr&oi=ao

Title: Acoustic metafluids based on random microstructure networks

Author(s): Aljaž Draškovič-Bračun (1,2) , Tilen Potisk (1,2) , Matej Praprotnik (1,2) and Daniel Svenšek (1,2)
Affiliation(s): (1) Laboratory of Molecular Modeling, National Institute of Chemistry, Ljubljana, Slovenia. (2) Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

Abstract: In metamaterials, the interaction between waves and matter leads to unusual wave-propagating properties of the medium, which can be used for various purposes [1]. In acoustic metamaterials, the parameters that determine sound propagation, i.e., the effective compressibility χ and density ρ of the wave-carrying medium, can become simultaneously negative in certain frequency windows due to resonance effects of sub-wavelength inclusions [1,2,3]. In our work [3] we focus on irregularly shaped inclusions - micro-oscillators – that are allowed to be randomly distributed throughout the host fluid and have irregular modal shapes. We show that the metafluid concept need not necessarily be based on position periodicity or correlation of the suspended micro-oscillators, and in this case not even on ideally designed micro-oscillators. We formulate the detailed operating principle of such a metafluid model, give explicit formulas for its effective dynamic moduli in terms of the modal structure of the micro-oscillators, and discuss basic practical issues of performance optimization in terms of their mass and size. In our model the micro-oscillators consist of point masses connected by harmonic potentials. Further we discuss how the amount and distribution of such connections affects the effectiveness of the micro-oscillators in modification of apparent acoustic parameters of the fluid. It turns out that the so called floppy-modes, which appear at very low frequencies due to under-constrained regions in the system [4], can have a substantial effect on acoustic parameters of the medium. This makes them acoustically accessible and additionally the absence of need for an intricately designed structure brings experimental realizations that much closer.

[1] M. Kadic, T. Bückmann, R. Schittny, and M. Wegener, Metamaterials beyond electromagnetism, Rep. Prog. Phys. 76, 126501 (2013)
[2] T. Brunet, A. Merlin, B. Mascaro, K. Zimny, J. Leng, O. Poncelet, C. Aristégui, and O. Mondain-Monval, Soft 3d acoustic metamaterial with negative index, Nat. Mater. 14, 384 (2015).
[3] A. Draškovič-Bračun, T. Potisk, M. Praprotnik, and D. Svenšek, Suspension of discrete microscopic oscillators as a model of an ultrasonic metafluid, Phys. Rev. B 105, 224317 (2022).
[4]A. J. Rader, B. M. Hespenheide, L. A. Kuhn, and M. F. Thorpe, Protein unfolding: Rigidity lost, Proceedings of the National Academy of Sciences 99, 3540 (2002)
Keywords: Acoustic metamaterial; ultrasound; networks;

Author Profiles(s): Aljaž Draškovič-Bračun https://orcid.org/0000-0002-1250-3583

Title: Complex systems perspective on natural language

Author(s): Stanisław Drożdż
Affiliation(s): Faculty of Computer Science and Telecommunications, Cracow University of Technology and Institute of Nuclear Physics, Polish Academy of Sciences, Cracow, Poland

Abstract : The science of complexity seeks to address the fundamental question of the governing principles that Nature employs when organizing the basic constituents of matter and energy into intricate structures and dynamic patterns that permeate all levels of the Universe's hierarchy. A closely connected phenomenon, namely natural language, has exhibited remarkable abilities in swiftly emerging and being adopted by humans. This linguistic phenomenon effectively mirrors these intricate patterns, as evidenced by its capacity to encode and communicate information pertaining to them and among them. Consequently, it is entirely justified to anticipate that natural language encapsulates the core essence of complexity. Indeed, this assertion holds particularly true in the context of human speech and writing, where the fact that *more is different* becomes strikingly evident. Therefore, it is only fitting for natural language to occupy a central role in the quantitative examination within the realm of complexity science.

Referring to such a perspective, this presentation [1] aims to consolidate the key methodological principles employed within this field and to assess their effectiveness in distinguishing between universal characteristics and language-specific traits within written representations of natural language across major Western languages. It thus delves into the examination of word frequencies in texts across major Western languages, highlighting the significant finding that accounting for punctuation largely restores the scaling behavior that is typically disrupted in the Zipf's law, especially for the most frequently used words, a phenomenon often addressed through the Mandelbrot correction.

Subsequently, the time series analysis techniques is utilised to investigate different forms of long-range correlations within written texts, drawing inspiration from complex systems. These time series are derived by segmenting the text into sentences or fragments between consecutive punctuation marks. Intriguingly, these series exhibit characteristics commonly observed in signals originating from complex systems, including the presence of long-range correlations and the emergence of fractal or even multifractal structures. Furthermore, a noteworthy observation is that the fluctuations in the distances between consecutive punctuation marks appear to universally follow the discrete Weibull distribution, a pattern often encountered in survival analysis.

Lastly, the utilization of complex network methodologies in the realm of linguistic structures, with a focus on word-adjacency networks is explored. These networks capture the relationships between words based on their co-occurrence within texts. The findings from such analyses suggest that the network metrics derived from these structures can serve as effective tools for tasks such as text classification, including authorship attribution. Complex networks have also been applied to a different category of linguistic networks known as word-association networks, which are constructed using data gathered from specific psycholinguistic experiments. Throughout all of the analyses presented in relation to written language, punctuation emerges as a pivotal factor, exerting a profound influence on the quantifiable characteristics of language.

[1] Based on collaboration with Jarosław Kwapień and Tomasz Stanisz

 Keywords: Complex Systems, Natural Language, Correlations

Title: Simulations of water and aqueous solutions under extreme conditions: the important role of molecular dynamics simulations

Author: Paola Gallo
Affiliation; Università degli Studi Roma Tre, Dipartimento di Matematica e Fisica

Abstract: Water is a complex system where computer science plays a crucial role. Computer molecular dynamics simulations in particular are pivotal to study the behavior of water and aqueous solutions under extreme conditions. Especially in the supercooled region that is to-date only partially accessed by experiments. I will show water systems where computer simulations show results in line with experiments and produce further results in regions of the phase diagram not explored so far by experiments. I will show results on thermodynamics, slow dynamics and structure upon cooling of bulk water, and water in solutions with inorganic and organic solutes. The results that I will discuss are of interest also for outer planets science, and in particular for the research about water on Mars, and for cryoprotection techniques.

Keywords: Supercooled water (1); Computer simulations (2); aqueous solutions (3).

Author Profiles: Google Scholar https://scholar.google.it/citations?user=-OD3YEQAAAAJ&hl=it

Title: Statistics for a SAFE AI

Author(s): Paolo Giudici
Affiliation(s): Statistical laboratory, University of Pavia

Abstract: The current widespread use of AI motivates the need to develop advanced statistical methods that can measure its "trustworthiness", in line with the Artificial Intelligence Act recently proposed by the European Commission (European Commission, 2021).

To measure trustworthiness of AI, we propose statistical metrics that consist of a set of four integrated statistical measures of trustworthiness, all based on the extension of the Lorenz Curve (Lorenz, 1905): from the measurement of income concentration to the measurement of the concentration of machine learning predictions. The four statistical metrics can be summarised with the acronym S.A.F.E., which derives from the four considered variables: Sustainability, which refers to the resilience of the AI outputs under anomalous extreme events and/or cyber attacks; Accuracy, which refers to the predictive accuracy of the model outputs; Fairness, which refers to the absence of biases towards population groups, induced by the AI output; Explainability, which refers to the capability of the model output to be understood and oversight by humans, particularly in its driving causes. While the former two requirements are more technical, and "internal" to the AI process, the latter two are more ethical, and "external" to the AI process, involving the stakeholders of an AI system.

We remark that tbs proposed metrics consist of "agnostic" statistical tools, able to postprocess the predictive output of a machine learning model in a general way, independently on the underlying data structure and statistical model.

Keywords: machine learning (1); Lorenz zonoids (2); explainable AI (3).

Author Profiles (s): https://scholar.google.com/citations?user=ogeVB1kAAAAJ&hl=en

Title: The shape of data: Manifolds, Cities and Biodiveristy

Author(s): Thilo Gross (123)
Affiliation(s): Helmholtz Institute for functional marine biodiversity (1); University of Oldenburg, Institute for Chemistry and Biology of the Marine Environement (2); Alfred-Wegener Institute, Helmholtz Centre for Marine and Polar research (3)

Abstract: The aim of this talk is to highlight the need for a geometrical understanding of datasets as an important challenge at the interface of complexity and data science. Consider a dataset that describes a collection of objects. Each object then corresponds to one point in the, typically high-dimensional data-space, spanned by the dimensions recorded in a dataset. For example in a dataset on lakes, each lake could be a datapoint, and the length, width and depth of the lake could be three of the axes that span the data-space. In general the datapoints do not fill the dataspace uniformly. Instead they typically lie on realtively thin, effectively low-dimensional manifolds. Thus for example the lakes in our example reveal the manifold of plausible lakes. All points correspond to lakes that could very well exist in the real world while points that are not on the manifold would likley not be recognied as lakes even if they were encountered in reality. In this talk I am going to present how the geometric exploration of data manifolds can yield interesting insights into datasets. In particular I will consider data on cities and ecosystems to demonstrate that the natural coordinate system implied by the shape of the data manifolds yield coordinates that have an intuitive interpretation.

Keywords: Manifold Learning, Urban Structure, Biodiversity.
Author Profiles (s): exTwitter: @thilogross ; homepage: biond.org ; scholar: eogyTQUAAAAJ ; LinkedIn: thilo-gross-75b9a86 ; youtube: @complexitypapers

Title: How to make clinical predictions when we do not know everything?

Author(s): Haralampos Hatzikirou
Affiliation(s): 1. Department of Mathematics, Khalifa University, Abu Dhabi, United Arab Emirates, 2. Centre for information services and high performance computing (ZIH), TU Dresden, Germany

Abstract: In clinical practice, a plethora of examinations is conducted to assess the state of a certain pathology. These span from blood sample analysis, clinical imaging (e.g. CT, MRI) and biopsy sampling are among the most important diagnostic and prognostic tools. Such medical data correspond to snapshots in time of the patient's state, since current standard of care (SoC) is not based on emergent technologies of real-time measurements, such as liquid biopsies or biosensors. Moreover, clinical data refer to different biological scales since imaging, such as MRI, typically provides an organ level picture of a disease (macroscopic), biopsies represent cellular patterns at a tissue (mesoscopic) level and -omics, FACS or molecular markers allow for sub-cellular insights. Finally, the biophysical mechanisms that regulate phenomena in all these scales are not completely known. Therefore, current clinical care faces the following challenges: (C1) data collection is sparse in time since it relies on patient's clinical presentation, (C2) we lack the knowledge/uncertainty of the mechanisms involved in regulating these data variables across different scales (structural uncertainty), and (C3) medical data are multiscale. Therefore, integrating these data to predict the future of a disease and propose an appropriate treatment is a formidable task. I propose to harness the ability of mechanistic models to integrating the existing biological knowledge and deal with the emerging dynamics. At the same time complete the missing knowledge by using data intensive techniques. Here I will present (i) a Bayesian regression framework of combining models and machine learning to predict tumor growth and (ii) model-driven classification method to assess the graft loss risk in kidney transplantation patients.

Keywords: Mechanism uncertainty, dynamic modelling, machine learning, clinical predictions
Author Profiles (s): https://www.linkedin.com/in/haralampos-hatzikirou-69b7a817, www. hatzikirou.gr

Title: From Big Data to Digital Twins: Promises and Perils

Author(s): Dirk Helbing
Affiliation(s): ETH Zurich, Switzerland

Abstract: Digital technology is reinventing the ways we study, simulate and manage the world. Around the world, Big-Data-driven, AI-controlled forms of cybernetic societies are in the making. What are the promises? What are the perils? And why is science still needed, 15 years after Chris Anderson proclaimed "the end of theory" and that "the data deluge makes the scientific method obsolete". This has to do with the properties of complex, networked systems.

So, was the FuturICT message, with the recent warning of "existential threats" from AI, voiced by "AI godfather".

Geoffrey Hinton and others, finally understood? And what scientific challenges should we now be working on?

Title: Digital Twin-Learning: A Complex Systems Framework

Author(s): Dzintars Jankovskis (1); Iveta Cirule (2); Anna Carbone(3)
Affiliation(s): Liepaja University, Liepaja, Latvia (1); Liepaja University, Liepaja, Latvia; BIORGANIK5 Ltd, Riga, Latvia (2); Politecnico di Torino, Turin 10129, Italy (3)

Abstract: The Digital Twin (DT) term has become extremely popular among various sectors like manufacturing, healthcare, urban and transport infrastructure, agriculture, and other industries, however, there can be a significant difference between the understanding of this term, even in each of the industries. Several authors have explained the main difference between the Digital Model, Shadow, and Twin and we believe there has to be a common framework for how DT can be taught so there is always a clear understanding between those terms. DT term has to be well defined so not only can industry experts understand and use it correctly but there is also a common understanding in the academic field of this term.

1. Wright, L., Davidson, S. How to tell the difference between a model and a digital twin. Adv. Model. and Simul. in Eng. Sci. 7, 13 (2020). https://doi.org/10.1186/s40323-020-00147-4
2. Fuller, Z. Fan, C. Day and C. Barlow, "Digital Twin: Enabling Technologies, Challenges and Open Research," in IEEE Access, vol. 8, pp. 108952-108971, 2020, doi: https://doi.org/10.1109/ACCESS.2020.2998358
3. Johnson, J., Buckingham Shum, S., Willis, A. et al. The FuturICT education accelerator. Eur. Phys. J. Spec. Top. 214, 215–243 (2012). https://doi.org/10.1140/epjst/e2012-01693-0
4. Cirule, I.; Uvarova, I. Open Innovation and Determinants of Technology-Driven Sustainable Value Creation in Incubated Start-Ups. J. Open Innov. Technol. Mark. Complex. 2022, 8, 162. https://doi.org/10.3390/joitmc8030162
5. D. Jankovskis, I. Cirule and Anna Carbone, Digital Twins, and E-Learning: Challenges and Opportunities. 5th International Conference on Higher Education Learning Methodologies and Technologies Online University of Foggia. 2023.

Keywords: Digital Twin, Learning, DT-Learning, Digital Model, Digital Shadow

Author Profiles (s): Dzintars Jankovskis[1(0009-0007-4911-0031)], Iveta Cirule[2[0000-0003-2545-7085],] and Anna Carbone[3[0000-0003-4945-9165]]

Title: Microscopic data analysis for financial market microstructure: a quantitative test of a microscopic econophysics model for the long memory in market-order flows

Author(s): Yuki Sato and Kiyoshi Kanazawa
Affiliation(s): Department of Physics, Kyoto University, Japan

Abstract: In this study, we examine a detailed financial dataset from the Tokyo Stock Exchange (TSE) to validate a microscopic econophysics model. A widely recognized observation in financial market microstructure is the persistent nature of market-order flows over long periods. Specifically, let a buy (sell) market order sign at time t be represented as $\epsilon_t := +1(-1)$. The long-term persistence is marked by a power-law decay in the autocorrelation function of these market-order signs, described as $C_\tau := E[\epsilon_t \epsilon_{t+\tau}] \propto \tau^{-\gamma}$ with $0 < \gamma < 1$. In econophysics, the origin of this phenomenon has been a topic of debate for a long time. The most promising microscopic hypothesis is the order-splitting behaviour at the level of individual traders. In 2005, Lillo, Mike, and Farmer (LMF) proposed a corresponding microscopic model based on this order-splitting hypothesis. Furthermore, they predicted that the macroscopic power-law exponent γ should be given by $\gamma = \alpha - 1$ with the microscopic power-law exponent α that is related to the distribution of the total number of order splittings. However, the LMF prediction has not been validated quantitatively by data analyses for 18 years without appropriate large microscopic datasets. In this talk, we solve this long-lasting problem by analysing a large microscopic dataset in the TSE market. Finally, we find that the LMF prediction holds even at a quantitative level.

1. Y. Sato and K. Kanazawa, Inferring microscopic financial information from the long memory in market-order flow: A quantitative test of the Lillo-Mike-Farmer model, to appear in Phys. Rev. Lett.
2. Y. Sato and K. Kanazawa, Quantitative statistical analysis of order-splitting behavior of individual trading accounts in the Japanese stock market over nine years, to appear in Phys. Rev. Res.

Keywords: econophysics; financial market microstructure; long memory ; Lillo-Mike-Farmer model

Author Profiles (s): Scholar: https://scholar.google.com/citations?user=XE2XapQAAAAJ, ORCID: https://orcid.org/0000-0002-6531-1438, Website: https://kanazawa.scphys.kyoto-u.ac.jp/

Title: On the structure of the world economy

Author(s): Ljupco Kocarev
Affiliation(s): Macedonian Academy of Sciences and Arts, Krste Misirkov 2, Skopje, Macedonia

Abstract: The expansion of global production networks has raised many important questions about the interdependence among countries and how future changes in the world economy are likely to affect the countries' positioning in global value chains. We are approaching the structure and lengths of value chains from a completely different perspective than has been available so far. By assigning a random endogenous variable to a network linkage representing the number of intermediate sales/purchases before absorption (final use or value added), the discrete-time absorbing Markov chains proposed here shed new light on the world input/output networks. The variance of this variable can help assess the risk when shaping the chain length and optimize the level of production. Contrary to what might be expected simply on the basis of comparative advantage, the results reveal that both the input and output chains exhibit the same quasi-stationary product distribution. Put differently, the expected proportion of time spent in a state before absorption is invariant to changes of the network type. Finally, the several global metrics proposed here, including the probability distribution of global value added/final output, provide guidance for policy makers when estimating the resilience of world trading system and forecasting the macroeconomic developments.
Keywords: world economy; global production networks; discrete-time absorbing Markov chain; sensitivity analysis

Title: Phase Diagram and Structure of Supercooled Water in Sodium Perchlorate Solutions

Authors: Paolo La Francesca, Paola, Gallo
Affiliation: Dipartimento di Matematica e Fisica, Università degli studi Roma Tre, Italy

Abstract: Phase transitions constitute a collective change in the behaviour of a system with a high number of particles and are perhaps the most important example of critical phenomena in condensed matter physics. The anomalous behaviour of water can be explained by the existence, in its supercooled region, of a second-order phase transition occurring at a liquid-liquid critical point below which water would exist in two separate phases: high-density liquid (HDL) and low-density liquid (LDL). In this context, since water is often easier to supercool in solutions, it is natural to study the thermodynamic behaviour of aqueous solutions. Aqueous perchlorate solutions are of special interest due to their supposed existence in a liquid form in the Martian subsurface. By analysing the phase diagram of supercooled solutions of sodium perchlorate in water and its structure through numerical simulations of molecular dynamics, it is possible to investigate the collective changes in the behaviour of water that makes it possible for it to remain liquid at temperatures much lower than its freezing point. In particular, it is observed that the LDL phase, where nucleation is much more likely to occur, appears to shrink upon increasing the concentration, while the liquid-liquid phase transition occurs at slightly higher temperatures and significantly lower pressures.

Keywords: Numerical simulations; phase transitions; statistical mechanics; aqueous solutions.

Author Profiles (s): https://orcid.org/0009-0007-4097-4650, https://scholar.google.com/citations?user=-OD3YEQAAAAJ

Title: Complexity science in innovation management

Author: Raffaella Manzini
Affiliation: LIUC Università Cattaneo, Castellanza, Italy

Abstract: Innovation is a dynamic, complex phenomenon. Most scholar and practitioners would agree on that, and an empirical confirmation can be found in the recent dynamic reaction of the innovation eco-system to the COVID pandemic. During the pandemic, a wide set of actors (agents) co-evolved as a complex system in which hospitals, institutions, doctors, large pharmaceutical companies, and small science-technology based ones, universities, and individuals leveraged their resources and capabilities and created new ones, to let new solutions to new problems emerge in very short time. So, studying innovation eco-systems with the lens of complexity science would help a better understanding of these systems and would have dramatic managerial implications.

Despite the above, the use of complexity science and its methodologies in the field of innovation management is still very limited, probably because it would require a strong change in the way we approach research on innovation management. Just to mention a few examples, there would be the need to abandon the traditional distinction among the micro, meso and macro perspecitves of investigation; to change the way rationality, efficiency and redundancy are conceptualized and evaluated; to extend research well beyond "theory testing", dealing with emerging phenomenon.

It is not yet clear whether, and to what extent, these changes are actually taking place.

Keywords: Innovation ecosystems; innovation management; innovation network.

Author Profiles: Raffaella Manzini https://liuc.esploro.exlibrisgroup.com/esploro/profile/raffaella_manzini/
http://linkedin.com/in/raffaella-manzini-41b22982
https://www.researchgate.net/profile/Raffaella-Manzini

Title: Nonlinear climate data analysis: A complex systems perspective

Author(s): Cristina Masoller
Affiliation(s): Departament de Fisica, Universitat Politecnica de Catalunya, Spain

Abstract: The 2021 Nobel Prize in Physics recognized the fundamental role of complex systems in understanding our climate and the origin of climate change, and it has put the interdisciplinary research field of complex systems in the spotlight [1]. In this talk, I will present our work, where we use the Hilbert transform to unveil significant changes in surface air temperature across the globe that have occurred in the last decades [2]. I will also discuss how information measures such as the permutation entropy and the transfer entropy have allowed us to infer bivariate causal interdependencies among climatic indices [3, 4] and to detect changes in spatial vegetation fields [5].

[1] G. Bianconi et al, "Complex systems in the spotlight: next steps after the 2021 Nobel Prize in Physics", J. of Phys: Complexity 4, 010201 (2023).
[2] D. A. Zappala, M. Barreiro, C. Masoller, "Quantifying changes in spatial patterns of surface air temperature dynamics over several decades", Earth Syst. Dynam. 9, 383–391 (2018).
[3] R. Silini, C. Masoller "Fast and effective pseudo transfer entropy for bivariate data-driven causal inference", Sci. Rep. 11, 8423 (2021).
[4] R. Silini, G. Tirabassi, M Barreiro, L. Ferranti, C. Masoller, "Assessing causal dependencies in climatic indices", Climate Dynamics 61, 79–89 (2023).
[5] G. Tirabassi, C. Masoller, "Entropy-based early detection of critical transitions in spatial vegetation fields", PNAS 120, e2215667120 (2023).

Title: Design of artificial intelligence-enhanced competency framework on digital skills: usage Sentinel satellite data as a use-case

Author: Arturs Mietulis
Affiliation: Liepāja Unversity, Latvia

Abstract: Data is constantly produced and consumed, so need exists to understand and analyze relevant data flows [1]. Introduction of modern digital technologies requires improvement of quality of human capital [2]. Utilisation of disruptive digitally enabled technologies requires certain skills, knowledge, and attitudes.

Skills', knowledge, and attitude development is core in professional development of any specialist. In current research, civil protection personnel are considered as a target auditorium for whom a competency framework will be developed. Scoping review did not reveal any digital skills competency frameworks for these specialists, so the research will create added value.

A competency framework facilitates the identification of training needs and guides the design of a professional development program [3]. In current research, identified competencies will not just shape a static framework, but will be used as a basis for creation of an updatable (living) digital competency framework for civil protection personnel. Updatability shall be ensured by means of feeding the manually identified competencies from the framework to an artificial intelligence algorithm, that would be able to conduct further text mining to identify new competencies related to previous ones already included in the framework. New, highly relevant competencies identified by means of text mining will become a part of an automatically updated framework.

Main research question is whether it is *per se* possible to create an automatically updatable digital competency framework for civil protection personnel that would cover all levels of responsibilities, all areas

of activities, and all currently existing digital technologies. To answer the question, a use-case will be developed around digital skills required to utilise Sentinel satellite data. When framework part related to area of competencies relevant to Sentinel programme will be populated with keywords, the automation part of the study may begin. A machine learning model shall be developed to further perform independent text-mining and keywords' identification. Output shall be evaluated using the classification accuracy of the model that was trained on previously collected data. Evaluation shall be based on creation of misclassification table for the practical group (actual group) and the prediction group (classification group) [4].

While it is impossible to cover all sorts of application of digital skills, it has been decided to establish a use case. The use-case will focus on usage of Sentinel data, i.e., the pilot section of prospective competency framework will focus on knowledge, skills, and attitudes required to utilise digital Sentinel data.

A use-case will be developed around digital skills required to utilise Sentinel satellite data. When framework part related to area of competencies relevant to Sentinel programme will be populated with keywords, the automation part of the study may begin. A machine learning model shall be developed to further perform independent text-mining and keywords' identification.
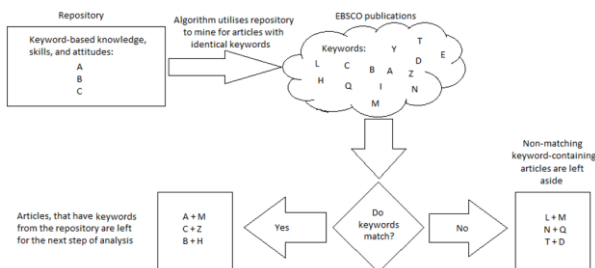

Illustration 1. Outline of processes of pre-selection of scientific articles for further analysis based on absence or presence of specific keywords
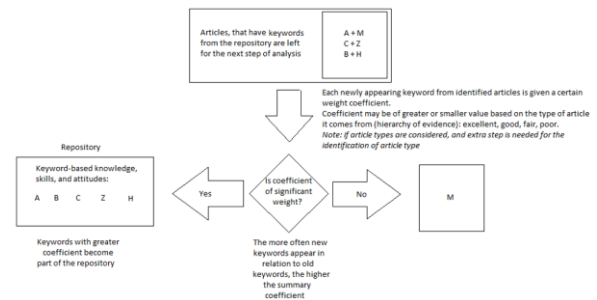

Illustration 2. Development of a threshold-based "filter" to exclude false positive and negative results

1. T.H. Davenport et al., How 'big data' is different, MIT Sloan Manag. Rev., 2012
2. Rodchenko, V., Rekun, G., Fedoryshyna, L., Roshchin, I., & Gazarian, S. (2021). The Effectiveness of Human Capital in the Context of the Digital Transformation of the Economy: The Case of Ukraine. Journal of Eastern European & Central Asian Research, 8(2), 202–213. https://doi.org/10.15549/jeecar.v8i2.686
3. Tetiana Ivanivna Kovalchuk, Oleksandr Yevhenovych Korystin, & Nataliia Petrivna Sviridyuk. (2019). Hybrid threats in the civil security sector in Ukraine. Проблеми Законності, 147, 163–175. https://doi.org/10.21564/2414-990x.147.180550 Competency Framework for Software Engineers, 2010
4. Juyoung Song, Author. Big Data Analysis Using Machine Learning for Social Scientists and Criminologists. Cambridge Scholars Publishing, 2019

Keywords: artificial intelligence, competency framework, digital skills, Sentinel

Author's profile: https://orcid.org/0009-0005-4453-7592

Title: Adventures with random walks: from counting birds to searching for global maxima on fitness landscapes

Author(s): Alexandre V. Morozov
Affiliation(s): Department of Physics and Astronomy (1) and Center for Quantitative Biology (2), Rutgers University, Piscataway, NJ 08854, USA

Abstract: Large-scale networks represent a broad spectrum of systems in nature, science, technology, and human societies. Computer networks such as the World Wide Web and the Internet, social networks such as Twitter/X and Facebook, and online knowledge-sharing platforms such as Wikipedia have become part and parcel of our lives. These networks tend to be very large and time-dependent, making predictions of their properties a challenging task. I will describe a novel methodology, based on random walks, for the inference of various properties of complex networks. I will show that our formalism yields reliable estimates of global network properties, such as the network size, after only a small fraction of its nodes has been explored. I will also introduce a novel algorithm for partitioning network nodes into non-overlapping communities - a key step in revealing network modularity and hierarchical organization. I will apply this algorithm to various benchmarks, including a large-scale map of roads and intersections in the state of Colorado. Next, I will demonstrate how these ideas can be extended to global optimization - the search for a global maximum or minimum on fitness or energy landscapes. Finally, I will describe a Monte Carlo sampling technique which can be used to infer key thermodynamic quantities in systems with discrete states from small, non-equilibrium samples. Thus, random walks can be used to reveal modular organization and global structure of complex networks, find global optima, and infer thermodynamics of physical systems such as spin glasses.

Title: Rejuvenation and memory in spin glass: when simulations meet experiments

Author: Ilaria Paga

Affiliation(s): CNR, Italy

Abstract: Driven by the synergy between the experimental group of Ray Orbach (Texas) and the Janus collaboration, we were able to solve a twenty-year problem concerning the magnetic response of a spin glass system in finite dimension. First, we have learned how to quantitatively extract the spin-glass coherence length (i.e. the size of the glassy domains) from simulations of non-equilibrium spin-glass dynamics. Second, thanks to the advent of the dedicated super-computer Janus II (numerical side) and of the single crystal of CuMn (experimental one), we have fulfilled the time-scale and length-scale gap between experiments and numerical simulations, so we have finally been allowed to compare them. Third, Janus II has provided a crucial understanding of how temperature chaos in non-equilibrium dynamics works. These milestones have allowed us to set up a successful simulation of memory and rejuvenation. In this talk, I will describe our recent success in reproducing in a simulation the spectacular memory and rejuvenation effects of spin glasses, reviewing the crucial ingredients that have brought us to this point.

Title: Data-driven new targets identification for chronic liver diseases

Author: Dimitris Polychronopoulos
Affiliation: Ochre Bio, Ltd, UK

Abstract: To the best of our knowledge, Ochre Bio has generated more human liver deep phenotyping data than anyone else. Deep phenotyping involves studying how perturbation of genes affects the liver, by juxtaposing tissue, blood and clinical phenotypes with functional genomics data to produce a map, or knowledge graph, of the relationships between genes and phenotype. We are building on this data to enable knowledge-graph based gene prioritisation. Running such 'in silico' screens allows us to narrow down the universe of genes for further study in our cellular, tissue, and organ models.

Keywords: biotech R&D; bioinformatics; liver disease; data-driven drug discovery; AI/ML; knowledge graphs.

Author Profiles (s): Scholar , LinkedIn , ORCID

Title: How scaling laws challenge geographical theories of urban systems

Author: Denise Pumain
Affiliation: University Paris 1 Pantheon-Sorbonne, France

Abstract: Since some twenty years the literature on scaling laws is booming, especially regarding their applications to urban systems. Is this a real breakthrough in the understanding of urban dynamics, or just a passing fad for a new scientific fashion? What are the results of this research trend? Can we draw interesting directions from it to deepen theoretical knowledge and to improve urban policies? I will answer these questions by recalling the conditions in which the first work on scaling laws applied to cities emerged, between our Parisian team of geographers and those of the Santa Fe Institute and London Imperial College. I will then show how the conception of these scaling laws evolved as they were confronted with a variety of empirical data and a diversity of systems of cities. By comparing the results obtained with those of other modelling approaches, I will make a provisional assessment of the contribution of this work to the evolution of urban theories. Some basic principles explain the recurrent difficulty of transferring and integrating models between the natural and social sciences. These issues need to be widely discussed in order to avoid too risky interventions in the dynamics of complex systems before new models can be proposed to planners.

Keywords: Systems of cities; rank-size rule; urban dynamics; co-evolution; urban theories

Title: The Development of Personalization method within Digital Product Ecosystems

Author(s): Alberts Pumpurs
Affiliation(s): Riga Technical University, Institute of Information Technology, Latvia

A few words about the research contents and outcomes:

1. Personalization in the Digital Domain: The research begins by highlighting the significance of personalization in the contemporary digital landscape, emphasizing its role in product development.

2. Current Personalization Methodologies: A review of existing methodologies that translate user needs into individual product personalization requirements.

3. Gap Identification in Ecosystem Personalization: The study identifies the limited application of current personalization strategies in the broader context of product ecosystems.

4. Proposed Personalization Methodology: The research proposes an approach tailored to identify areas within product ecosystems that could benefit from customization.

5. Empirical Evaluation: The Apple Headphone Ecosystem is used as a case study, with products from this ecosystem being evaluated using tools like the Kano model, QFD matrix, and TRIZ.

Abstract: Today personalization stands as an integral aspect of product development. This research aims to dive into the aspects of personalization methodologies applied within digital product ecosystems. Currently, there are established methodologies for converting user needs into requirements for individual product personalization, but these methods are not adapted in the context of product ecosystems. As the ecosystem domain has gained interest in academia and industry it is essential to asses the possibility of personalization approaches within an ecosystem framework. The objective is to explain the principles and methodologies in use, amongst the challenges of personalization techniques. As the result of this, the research proposes a personalization methodology tailored to pinpoint areas within product ecosystems ripe for customization. For empirical validation, the proposed personalization methodology will be applied to products within Apple's Headphone Ecosystem. This evaluation will employ a suite of analytical tools, including the Kano model, QFD matrix, TRIZ, and others that may emerge as pertinent. Conclusively, the research aspires to design methodologies that offer insights for ecosystem optimization – determining which components should be incorporated, excluded, or personalized – with the overarching goal of enhancing end-user satisfaction within the product ecosystem.

Keywords: Ecosystems, Personalization, User requirements.

Author Profiles (s): https://www.linkedin.com/in/albertspumpurs/

https://www.researchgate.net/profile/Alberts-Pumpurs

Title: Theoretical analysis of co-occurrence relationship in data characterized by the frequency distribution

Author(s): Koutarou Tamura
Affiliation(s): UB Research, Uzabase, Inc., Japan

Abstract: When we evaluate the characteristics of relationship between elements of data, such as tokens or tags for text data, the relationship is usually characterized by co-occurrence of the elements. The intensity of this relationship can be defined as the number or rate of pairs of two elements that occur in the same unit of data. And, the concept is applied for various fundamental metrics and modelling such as similarity indices, matching, and building a bipartite graph.

The frequency of co-occurrence strongly depends on the frequency of occurrence of the elements themselves. In this study, we theoretically derive the number of pairs of the elements in given frequency, and tried to understand the characteristics of co-occurrence in terms of the frequency of the elements. As a result of this theoretical derivation, we demonstrated that we obtain theoretical values of similarity indices such as the Jaccard index and cosine similarity and their statistical significance can be discussed.

Keywords: Co-occurrence, Similarity index Author Profiles (s):
https://scholar.google.co.jp/citations?user=plxcX0QAAAAJ&hl=ja&authuser=1

Title: Feeding the models with voluminous operational data and make them exploitable by users: the experience from ESA's Copernicus Sentinels and DestinE platform

Author(s): Alessandra Tassa; Jolyon Martin; Ines Sanz Morere.
Affiliation(s): European Space Agency, Belgium

Abstract: Measurements of the Earth are key for understanding the complexity of our planet and support decrypting the climate-resources-population conundrum. Satellites provide an advantageous viewpoint to observe the Earth and many of its phenomena, particularly the impacts of climate change. Space-based Earth Observation data are used by scientists and analysts from all over the world across many different sectors, with long time series feeding complex system models and supporting various types of reanalysis and forecasts in support to decision making in complex environments. Under the the European Commission's Destination Earth initiative, tera-byte-scale space-based data from the operational European Programme Copernicus will be integrated within digital models of the Earth to provide alternative what-if scenarios related to e.g. natural disasters, weather hazards, and climate change paths. The complexity of the data flows is exacerbated by the open nature of the platform which aims to serve a broad variety of users and user scenarios. In this context, the European Space Agency acts as provider of Copernicus Sentinels data on the one side and as the developer of the DestinE platform, the front-end of DestinE system, on the other side. In this lecture, we will describe the complexity of the system and the challenges faced in this first development phase. We will also argue for the future scenarios of operation of the system and the related challenges and opportunities.

Keywords: Earth observations (1); climate (2); digital twins (3).
Author Profiles (s): Jolyon Martin https://www.linkedin.com/in/jolyon-martin-828286a2/
                Inés Sanz Morère https://www.linkedin.com/in/inés-sanz-morère-73a4b450/
Useful links: Copernicus Programme: https://copernicus.eu
        Destination Earth initiative: https://destination-earth.eu/
        Copernicus Sentinels Data Access: https://dataspace.copernicus.eu/

Title: Chat GPT as an assistant to educator

Author(s): Viktorija Tataurova, Jekaterina Martinova
Affiliation(s): Riga Technical University, Latvia

Abstract: In recent years, artificial intelligence has been used as an assistance to humans in various fields, such as medicine, human resources, law. Although students are increasingly applying artificial intelligence as an assistance in their studies, educators are still reluctant to use artificial intelligence in the studying process. For the reason of its increasing application by modern students as the source of ideas for classroom projects, Chat GPT has been selected as the subject of the research. The present research aims at estimating the successful realization of Chat GPT by educators in the preparation of the classroom activities and the evaluation stage. For one month, two university educators will be implementing the advice Chat GPT gives on the preparation of the classroom activities, as well as the suggestions regarding the assignment evaluation. The evaluation is carried out by artificial intelligence on the following parameters – language, grammar, creativity, and compliance with the task requirements. The educators are expected to reflect on their experience with the use of Chat GPT as an assistance to a teacher, listing its benefits and drawbacks, and comparing the preparation and assessment time with and without the help of Chat GPT.

Keywords: Artificial Intelligence (1); Chat GPT (2); evaluation (3).

Author Profiles (s): ORCID: 0000-0002-2943-6970 (1); ORCID: 0000-0002-6934-9556 (2)

Title: Architecture Design and Challenges of Big Data Ecosystems

Author(s): Bedir Tekinerdogan
Affiliation(s): Wageningen University & Research, The Netherlands

Abstract: A big data ecosystem encompasses various tools and processes to handle, process, and analyze voluminous and complex data sets. It integrates multiple platforms and components to extract valuable insights from data to support the decision-making process. The task of developing and managing these big data ecosystems is notably non-trivial, given the heterogeneity and dynamic nature intrinsic to these systems. In this talk, our attention will center on the architecture modeling and design pertinent to big data ecosystems. We will delve into two industrial case studies, focusing on big data ecosystems applied in the realms of precision agriculture and healthcare systems. Throughout, key challenges and impediments related to the design and management of these expansive big data ecosystems will be explored and discussed in depth.

Keywords: Big Data; Digital Ecosystem; System of Systems; Architecture Design

Author Profiles (s): https://nl.linkedin.com/in/bedir;

https://www.researchgate.net/profile/Bedir-Tekinerdogan;

https://orcid.org/0000-0002-8538-7261

Title: The Targeted Energy Transfer Nonlinear Model Analyzed Through Machine Learning Methods

Author(s): Giorgos P. Tsironis
Affiliation(s): Department of Physics, University of Crete and FORTH, Greece

Abstract: The Targeted Energy Transfer (TET) mechanism involves resonant transfer of energy in a non-resonant but nonlinear system [1]. The original idea was motivated by ultrafast electron transfer in chlorophyl molecules and was formulated in the context of the Discrete Nonlinear Schrödinger (DNLS) equation. The transfer is perfect when a specific constrain is fulfilled that connects nonlinearity with energy disparity in the context of a two state model. While it applies to classical models it can be readily extended to quantum systems as well [2]. In both classical and quantum cases it is very important to find the constrain that enables efficient transfer. While this is possible in simple dimer units, it becomes a very challenging task in more extended systems. In order to bypass this difficulty we may apply techniques from Machine Learning (ML). Specifically, by selecting appropriate loss function and minimizing it in the system parameter space one may obtain a direct formulation of the appropriate constrain that leads to resonant transfer. In the classical oscillator case we can find easily the TET transfer condition by employing this method in the dimer case that is known analytically [3]. Furthermore, we can use the method in more complex geometries such as the a trimer that involves a nonlinear dimer unit that is separated by a linear state. While this trimer model is not analytically tractable the application of the ML approach gives readily the resonant transfer parameter landscape. A similar approach can be applied to the quantized version of the DNLS model [4]. Extension of the classical loss function to the quantum case and subsequent minimization through learning and back-propagation results in the precise analytical TET result for the quantum system. The method is then applied to the fully quantum trimer case where the intermediate stat is in general nonlinear. Assuming that two of the oscillators fulfil the quantum TET condition leads to optimization of transfer through the third state. This transfer enables movement of arbitrary number of bosons in unison from the donor to the acceptor state. The successful application of ML techniques in this model is now being extended to the cases where in addition to the electronic degrees of freedom we also have vibrational degrees as well [5] or even Parity-Time (PT) symmetry [6]. We show analytically and numerically that TET works quite efficiently also in these more complex cases. We focus on the temperature dependence of the phenomenon and apply our ML technique in order to find parameter regimes for optimal transfer. In conclusion the TET model is quite general and through the application of ML methods one may uncover its full applicability.

[1] G. Kopidakis, S. Aubry, and G. P. Tsironis, Phys. Rev. Lett. 87, 165501 (2001).
[2] P. Maniadis, G. Kopidakis, and S. Aubry, Physica D: Nonlinear Phenomena 188, 153 (2004).
[3] G. D. Barmparis and G. P. Tsironis, JOSA B, 38, C120 (2021)
[4] J. Andronis, G. Arapantonis, G. D. Barmparis and G. P. Tsironis, Phys. Rev. E 107, 065301 (2023)
[5] N. Almazova, S. Aubry and G. P. Tsironis, in preparation
[6] A. Tselikis, H. Tselikis, G. D. Barmparis, K. Makris and G. P. Tsironis, in preparation.

Keywords: Targeted energy transfer (1); Nonlinear resonance (2); Machine Learning(3).

Author Profiles (s): Scholar: https://scholar.google.gr/citations?user=L7vdsyMAAAAJ&hl=en
ORCID: https://orcid.org/0000-0003-4793-0224
Linkedin : https://www.linkedin.com/in/giorgos-tsironis-47048a3b/?originalSubdomain=gr

Title: Exploring Preferential Attachment in Evolving Cryptocurrency Transaction Networks

Author(s): Dániel Kondor 1, Nikola Bulatovic 2, József Stéger 2, István Csabai 2 and Gábor Vattay 2
Affiliation(s): Complexity Science Hub Vienna, Vienna, Austria (1); Department of Physics of Complex Systems, Eötvös Loránd University, Budapest, Hungary (2);

Abstract: Cryptocurrency transaction networks, exemplified by Bitcoin and Ethereum, represent intricate real-world complex systems with detailed temporal evolution. These networks have attracted extensive scrutiny from the network science, economic, and cryptographic research communities. In our previous analysis of early Bitcoin transactions, we elucidated the phenomenon of preferential attachment, where wealthier entities tend to accumulate more resources—an observation that has spurred ongoing interest. In this study, we revisit this phenomenon in the context of the drastically expanded Bitcoin network, which has grown nearly a hundredfold since our initial investigation. Moreover, we undertake a comparative analysis with Ethereum, the second-largest cryptocurrency, to discern commonalities and distinctions in their evolutionary patterns. Our research substantiates that preferential attachment continues to be a pivotal force shaping the trajectories of both Bitcoin and Ethereum transaction networks. To facilitate comprehensive analysis, we provide access to the latest versions of both transaction networks and an efficient software implementation. This tool empowers researchers to delve into the intricate linking statistics vital for understanding preferential attachment in networks comprising several hundred million edges. Our findings contribute valuable insights to understanding evolving cryptocurrency networks, shedding light on the enduring influence of preferential attachment in shaping their structures. This research advances our theoretical understanding and provides practical tools for further exploration in the realm of complex network dynamics.

Keywords: Cryptocurrency, Complex Networks, Preferential Attachment, Bitcoin, Ethereum, Network Evolution, Transaction Networks, Comparative Analysis.

Author Profiles (s): https://orcid.org/0000-0003-3720-7462

https://orcid.org/0000-0003-2836-1855

https://orcid.org/0000-0001-9232-9898

https://orcid.org/0000-0002-0919-9429

Title: Human - Environment Interaction and Well-being Monitoring: Systematic Review

Author(s): Ieva Vitolina[1], Atis Kapenieks[1], Janis Kapenieks Sen[1], Aleksandrs Gorbunovs[1], Janis Kaminskis[1], Una Krutova[1], Janis Kapenieks Jun[1], Guna Jakobsone-Snepste[1], Jean-Denis Mathias[2]

Affiliation(s): [1] Riga technical University, Latvia [2]National Research Institute for Agriculture, Food and Environment (INRAE), France

Abstract: Nowadays, the well-being of people, the planet, and the environment is a topic of increasing global attention. The 2030 Agenda for Sustainable Development is aimed at improving the quality of people's lives, reducing the degradation of the planet, and ensuring the sustainability of the environment. The aim of the study is to conduct a literature review on the concept of human - environment well-being monitoring and to highlight criteria impacting well-being, the data sources and models used for characterizing well-being processes. The focus of the well-being research is from the aspect of interaction between people and the environment. Research questions: RQ1: What is the role of human-nature interactions on well-being? And how is it taken into account in the literature? RQ2: What is the influence of space/rural dimension to well-being monitoring? According to predefined keywords 73 articles were determined to be relevant for the performance of a systematic review report. Although the systematic review is still in process, for the moment it is possible to draw some conclusions: 1) The dominant approach is anthropocentric approach (ecosystem, nature is a tool or service for human well-being). A holistic approach (human as ecosystem part) or an ecocentric approach does not dominate as a research goal in the documents. 2) The concept of well-being is interpreted through a combination of different criteria and indicators depending on the context and the goals of the research authors. Mainly criteria and indicators are grouped to characterize: economic, social, human and environmental aspects. 3) Well-being is mainly determined at the national or regional level, not measured for small territories (local scale). The planned further works are as follows: (1) To continue to perform systematic review related to well-being concept, to synthesize the results. (2) To focus on local scale research used geolocalized well-being factors. (3) To draw conclusions about approaches and dynamic data modelling useful for Latvia. (4) Identify new approaches based on technologies for well-being data collection and analysis such as satellite imaging analysis.

Keywords: well-being (1); Human - Environment Interaction (2); models (3).

Author Profiles (s): Ieva Vitolina: https://orcid.org/0000-0002-7089-1326

Atis Kapenieks: https://orcid.org/0000-0002-3082-5922

Janis Kapenieks Sen: https://orcid.org/0000-0001-6244-2355

Aleksandrs Gorbunovs: https://orcid.org/0000-0003-3770-7168

Janis Kaminskis: https://orcid.org/0000-0001-6345-8084

Una Krutova: https://www.linkedin.com/in/krutova-una-32842895/

Janis Kapenieks Jun: https://www.linkedin.com/in/janis-kapenieks-5436017/

Guna Jakobsone-Snepste: https://www.linkedin.com/in/mba-guna-jakobsone-snepste-1008b043/

Jean-Denis Mathias: https://orcid.org/0000-0002-6172-9079

Title: Unraveling the Dance of Bacteria: exploring bacteria mobility through simulation in confined settings

Author(s): Dario Javier Zamora, Roberto Artuso
Affiliation(s): Universita' degli Studi dell'Insubria, Italy

Abstract: Understanding the dynamics of bacteria mobility within complex systems is essential for elucidating their ecological impact, disease transmission, and biotechnological applications. In this work, we delve into the world of bacteria mobility through the lens of complex systems. Our research focuses on employing random walks and run-and-tumble models as tools to describe and predict the movement patterns of bacteria in heterogeneous environments. By using the capabilities of data science, we use experimental data to validate these models, providing insights into the intricate mechanisms governing bacterial motility. Through simulations, we explore the behavior of bacteria in confined settings. Our findings shed light to diverse areas, spanning from applied questions, such as open doors for the development of innovative strategies in biotechnology, to fundamental mathematical ones, such as the generalized mean chord theorem. The mean chord theorem states that the mean path length of ballistic trajectories randomly crossing a domain only depend on the ratio between the region area and its perimeter. This theorem has been recently generalized for non-convex domains and extended to the case of Brownian motion to find many applications in various fields including biological locomotion.

Keywords: Random Walk (1); Run and Tumble (2); Bacteria mobility (3).
Author Profiles (s): Scholar: https://scholar.google.com/citations?user=BC3zaLAAAAAJ&hl=it&oi=ao
OrcID: 0000-0001-5672-6212

Title: Computationally efficient stochastic space-time prediction method for regular and scattered data in arbitrary dimension

Author(s): Milan Žukovič (1); Dionissios Hristopulos (2)
Affiliation(s): Institute of Physics, Pavol Jozef Šafárik University in Košice, Slovakia (1); Department of Electrical and Computer Engineering, Technical University of Crete, Greece (2)

Abstract: The work introduces a physically inspired machine learning method for spatial/temporal regression called the modified planar rotator method (MPRS). MPRS is a non-parametric model which incorporates spatial or temporal correlations via short-range, distance-dependent ``interactions'' without assuming a specific form for the underlying probability distribution. Predictions are obtained by means of a fully autonomous learning algorithm which employs equilibrium conditional Monte Carlo simulations. MPRS is able to handle scattered data and arbitrary spatial dimensions. Its prediction performance is tested on various synthetic and real-word data in one, two and three dimensions and the results demonstrate that MPRS is competitive with standard interpolation methods, such as ordinary kriging and inverse distance weighting, even without parameter tuning. It is particularly effective for rough and non-Gaussian data, such as daily precipitation time series. MPRS shows superior computational efficiency and scalability for large samples. Massive data sets involving millions of nodes can be processed in a few seconds on a standard personal computer.

Keywords: space-time prediction (1); autonomous algorithm (2); non-Gaussian model (3).
Author Profiles (s): Milan Žukovič - ORCID: https://orcid.org/0000-0001-6241-299X
ResearcherID: http://www.researcherid.com/rid/H-1600-2016
Scopus Author ID: http://www.scopus.com/inward/authorDetails.url?authorID=55960223500&partnerID=MN8TOARS
Dionissios Hristopulos - Linkedin: https://www.linkedin.com/in/dthchania/
Google profile: https://scholar.google.gr/citations?hl=en&pli=1&user=UosHluAAAAAJ
Personal web page: http://www.geostatistics.tuc.gr/index.php?id=4908