UNIVERSITÀ DEGLI STUDI DI MILANO
Selezione pubblica per n. 1 posto di Ricercatore a tempo determinato ai sensi dell'art. 24, comma 3, lettera a) della Legge 240/2010 per il settore concorsuale 01/B1 - INFORMATICA, settore scientifico-disciplinare INF/01 - INFORMATICA presso il Dipartimento di INFORMATICA "GIOVANNI DEGLI ANTONI", (avviso bando pubblicato sulla G.U. n. 7 del 25/01/2019) Codice concorso 3979

JIANYI LIN
CURRICULUM VITAE

Informazioni Personali

Cognome	LIN
Nome	JIANYI
Data di Nascita	24/01/1981

Research Positions

Sep 2016 – ongoing	**Assistant Professor** at the Department of Mathematics, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
Apr 2013 – Sep 2016	**Post-doc Fellow** at the Department of Computer Science, University of Milan, Milan, Italy Research activity title: A computational approach to the automatic recognition of facial expressions Supervisor: dr. Raffaella Lanzarotti
Mar 2012 – Feb 2013	**Post-doc Fellow** at the Department of Computer Science, University of Milan, Milan, Italy Research activity title: Development of an automatic method for feed-back control in industrial applications Supervisor: prof. Paola Campadelli
Jun 2011 – Oct 2011	**Research Assistant** at Airett Onlus & Department of Human Physiology, University of Milan, Milan, Italy Research activity title: Multi-factorial analysis of gait and movement in the Rett syndrome Supervisor: prof. Ioannis U. Isaias
Nov 2007	**Temporary Research Fellowship** at the Department of Electronics and Information, Politecnico di Milano, Milan, Italy Research activity title: Development a cell-based model for the simulation of forest fires in spatially extended forests and statistical analysis of burnt areas series Supervisor: prof. Sergio Rinaldi

Education

Mar 2012	**Doctor of Philosophy in Mathematics and Statistics for Computational Science**, University of Milan, Milan, Italy Thesis: "Exact algorithms for size constrained clustering" — Supervisor: prof. Alberto Bertoni
Oct 2007	**Master of Science in Computing Systems Engineering**, Politecnico di Milano, Milan, Italy Thesis: "A cell-based spatial model for the analysis of forest fires" — Supervisor: prof. Sergio Rinaldi

Oct 2004	**Bachelor of Science in Computer Science**, University of Milan, Milan, Italy Thesis: "Pattern statistics in multi-component stochastic models" — Supervisor: prof. Massimiliano Goldwurm

Professional Qualification

Feb 2009	**Ingegnere dell'Informazione Senior** Qualification obtained upon completion of Italian national examination (comparable to US Professional Engineer)

Awards and Scholarships

Mar 2009	Nomination in the List of Best Graduates in 2007-2008, Politecnico di Milano, Milan
Oct 2008	2008 Best Thesis Award with monetary prize for studies on modelling and/or computer science applied to environmental science, CIRITA (Interdepartmental Centre for Research in Environmental Informatics), Milan
Feb 2008	Scholarship for the International PhD School on Randomized Algorithms, Ennio De Giorgi Mathematics Research Centre, S.N.S., Pisa
2008 – 2010	3 year Fellowship for the PhD Programme from Italian Ministry of Education, University and Research
2002 – 2007	5 year Scholarship for undergraduate and graduate studies from ISU (Italian national Institute for the Right to University Education)

Visiting Scholar

May-Aug 2018 (3 months)	Birkbeck Institute for Data Analytics, Birkbeck College, **University of London**, UK Visiting researcher with focus on the study of structural properties and extension of centrality measures in complex networks. Invited by: Alessandro Provetti
Jun-Jul 2017 (2 months)	Center for Applied Statistics in Business and Economics, Department of Statistical Sciences, **Università Cattolica del Sacro Cuore**, Italy Visiting researcher with focus on topics of computational methods for statistical data analysis with applications to information and social sciences. Invited by: Emiliano Sironi
Jun 2013 (1 week)	Bioinformatics Group, Institute of Molecular Life Sciences, **University of Zurich**, Switzerland Short-term visiting with the aim of investigating suitable statistical corrections of off-target effects in siRNA screen data. Invited by: Andrea Franceschini

Teaching Activities

Lectures

Sep 2016–ongoing	**Instructor** for the courses: Probability & Statistics for Engineers; Random Variables and Stochastic Processes; Differential Equations, at the College of Engineering, Khalifa University $\simeq 450$ hours, taught in English
Oct 2014–Jan 2016	**Teaching Assistant** of the course Optimization, MSc in Economics and Finance, University of Milan 20 hours, taught in English
Oct–Dec 2015	**Teaching Assistant** of the course Geometry, BSc in Civil Engineering, Politecnico di Milano 24 hours
Mar–Jun 2015	**Teaching Assistant** of the course Probability and Statistics, BSc in Computer Science, University of Milan 16 hours

Apr–Jun 2014	**Temporary Lecturer** of the course Curves and Surfaces (Geometrical-Differential Analysis), BSc in Industrial Design, Politecnico di Milano 60 hours
Feb 2012–Sep 2016	**Examiner** of the course Signal Processing, BSc in Computer Science for New Media Communications, University of Milan $\simeq$ 40 hours
Oct 2011–Sep 2012	**Contract Professor** (professore a contratto) of the course Basic Computer Science, BSc in Biotechnology, University of Milan 40 hours
Oct 2009–Jan 2012	**Teaching Assistant** of the course Computer Science (Programming), BSc in Physics, University of Milan 120 hours
Co-supervisor[1] **of** Master's Theses	
ongoing	Matteo Maffetti, "Graph signals and sparse representation for emotion recognition" Supervisor: G. Grossi
Apr 2018	Gabriele Marancina, "ECG signal compression with sparse and adapted dictionary" Supervisor: G. Grossi
Jun 2017	Guido Rocco, "On the number of occurrences of a letter in strings generated at random from a regular language" Supervisor: M. Goldwurm
Sep 2015	Claudio Panelli, "Algorithms for 2-clustering in the Euclidean plane" Supervisor: M. Goldwurm
Jun 2015	Alessandro Stellino, "Algorithms for the clustering problem in the plane" Supervisor: M. Goldwurm
Dec 2014	Linda Pini, "Size-constrained clustering in the plane with Manhattan norm" Supervisor: M. Goldwurm
Bachelor's Theses	
Oct 2016	Lorenzo Milone, "Dynamic programming algorithms and applications to clustering problems" Supervisor: M. Goldwurm
Oct 2014	Antonio Lombardo, "Analysis of pathologies in voice signals" Supervisor: G. Grossi
Feb 2013	Andrea De Lentinis, "Algorithms for the constrained clustering problem" Supervisor M. Goldwurm
Feb 2012	Alessandro Stellino, "Heuristics for k-clustering with constraints" Supervisor A. Bertoni
Oct 2011	Luca Zanetti, "Dynamic data structure for convex hulls in the plane" Supervisor A. Bertoni
Supervisor of	Senior Design Project: "Construction and Validation of an Automated Laboratory Scale Plant Phenotyping Robotic Platform"
Department service	Dr. Lin has taken part in thesis examination boards, course syllabus design committee and degree program presentations

Journal and Conference Service

Editorial board member	International Journal of Computer Science and Application (DEStech) Canadian Journal of Biomedical Research (MaplesPub)
Journal referee	Data Science and Engineering (Springer) Journal of Imaging Science and Technology (IS&T Society) Journal of Optical Society of America A (OSA) Computer and Electrical Engineering (Elsevier) Simulation Modelling Practice and Theory (Elsevier)

[1] Thesis titles are translated from Italian

	Journal of Computational Methods in Sciences and Engineering (IOS Press) Journal of Environmental Informatics (ISEIS) Mathematical Problems in Engineering (Hindawi) Transactions on Internet Technology (ACM) Quality and Quantity (Springer) Biomedical Signal Processing and Control (Elsevier) IEEE Access (IEEE)
Technical program committee member	2nd International Conference on Data Science and Business Analytics (ICDSBA 2018) 4th International Conference on Intelligent Computing and Signal Processing (ICSP 2019) 4th International Conference on Networks, Communications, Wireless and Mobile Computing (NCWMC 2019) 1st IEEE SERVICES Workshop on Cyber Security and Resilience in the Internet of Things (within IEEE Services 2019)
Organization chair	International Workshop on Machine Learning for Wireless Communications (MLWCOM 2019: workshop to be held in conjunction with ISCC 2019)
Local organization member	40th International Symposium on Mathematical Foundations of Computer Science (MFCS 2015)
Conference reviewer	2016 International Conference on Image Analysis and Recognition (ICIAR 2016) 40th International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 2014) 21st Italian Workshop on Neural Networks (WIRN 2011) 11th International Conference on Artificial Immune Systems (ICARIS 2012) 12th World Congress on Intelligent Control and Automation (WCICA 2016) 7th International Conference on Complex Networks and Their Applications (COMPLEX NETWORKS 2018) 12thInternational workshop on Graph-Based Representation in Pattern Recognition (GbR 2019)

Affiliation

Associations	European Association for Theoretical Computer Science (EATCS)
Laboratories	At the Department of Computer Science, University of Milan: - PHuSeLab (Perceptual Computing and Human Sensing Lab)
Groups	Quantum Computing Research Group, Khalifa University of Science and Technology

Oral presentations

- Invited Speaker for Research Seminar of the Department of Statistics, UAE University, Al Ain, United Arab Emirates, 25 Oct 2018
  Presentation title: Pattern statistics in bicomponent rational stochastic models
- 20th International Conference on Descriptional Complexity of Formal Systems (DCFS 2018), Halifax, Canada, 27 Jul 2018
  Presentation title: A local limit property for pattern statistics in bicomponent stochastic models
- Research seminar at Birkbeck Institute for Data Analysis, University of London, United Kingdom, 16 Jul 2018
  Presentation title: Size-constrained clustering problems: computational complexity and algorithms
- AMS Department Research Seminar - Khalifa University, Abu Dhabi, United Arab Emirates, 6 Feb 2017
  Presentation title: Computational complexity of geometric clustering problems with size constraints
- 11th International Conference on Algorithmic Aspects in Information and Management (AAIM 2016),

Bergamo, Italy, 18 Jul 2016
Presentation title: On the Complexity of Clustering with Relaxed Size Constraints

- 15th Italian Conference on Theoretical Computer Science (ICTCS 2014), Perugia, Italy, 19 Sep 2014
  Presentation title: Size-constrained 2-Clustering in the Plane with Manhattan Distance
- 13th Italian Conference on Theoretical Computer Science (ICTCS 2012), Varese, Italy, 19 Sep 2012
  Presentation title: Size constrained clustering problems in fixed dimension

Other attendances at conferences and schools

Aug 2015	International Symposium on Mathematical Foundations of Computer Science, Milan
May 2015	Workshop on Clustering within the PRIN 2010-2011 Project, Naples,
Jun 2009	ECS10 The 10th European Congress of Stereology and Image Analysis, Milan
Jun 2009	Trends in Bifurcation Analysis: Methods and Applications, Milan
Mar 2008	BiSS Bertinoro International Spring School, Bertinoro
Feb 2008	BiCi-SNS International PhD School on Randomized Algorithms, Pisa

Project Participation

2019	**Co-Investigator member in research project with funding** Title: "Developing a new integrated abdominal Electrography system for noninvasively detecting fetal brain signals" Fund: CIRA, Khalifa University, United Arab Emirates – Project ID: CIRA-2019-023 Principal Investigator: Ahsan Khandoker Duration: 36 months Description: We investigate a new method for the noninvasive measurement of human fetal brain development (35 weeks gestation till one month after birth) by abdominal electrical signal, comparing the signals with fetal magnetic resonance images and fetal Electrocardiogram from fetal brain & heart respectively, and creating new technology called brain evoked electrography, BEEG.
2018	**Co-Investigator member in research project proposal under peer-review** Title: "Scouting for binding sites of potential drug candidates on the structures of oncogenic Ras superfamily proteins" Fund: Terry Fox International Research Grant 2018, Terry Fox Research Institute, Canada Principal Investigator: Suryani Lukman Description: Our overall objective is to discover novel allosteric sites on the 3D structures of oncogenic proteins belonging to the Ras superfamily (particularly Rab and Rho families), and chemical compounds that can bind to them to regulate/restore the normal functions.
2013–2016	**Participation as Post-doc fellow in research project activities** Title: "Facial expressions and interpretation of emotions: a computational approach integrating between image processing and physiological signals based on shape analysis and Bayesian networks" Fund: FIRB 2012—Italian Ministry of Education - Project ID: RBFR12VHR7 Description: The main contributions to the project consist, firstly, in the development of face image recognition systems based on sparse decomposition that are able to robustly classify under several uncontrolled conditions, such as variations in illumination, occlusion, expression, pose. Secondly, we proposed a novel sparse decomposition iterative scheme based on Lipschitzian mappings for application to ECG signal compression and dictionary learning integrating orthogonal Procrustes shape analysis.

2012–2015 **Participation as Post-doc fellow in research project activities**
Title: "Automata and formal languages: mathematical and applicative aspects"
Fund: PRIN 2010-2011—Italian Ministry of Education - Project ID: H41J12000190001
Description: In the context of project's research Line A4. "Generation, enumeration and compression of combinatorial structures", we studied the computational complexity of clustering problems subjected to cluster size constraints. For the polynomial-time solvable problems we designed efficient clustering algorithms, especially in low dimension, using results on convex hulls and k-sets, and adapting a technique from algebraic geometry called CAD cylindrical algebraic decomposition.

2011–2012 **Participation as Post-doc fellow in industrial research project activities**
Title: "Development of an automatic method for feed-back control in industrial context"
Funded by: Luxottica S.p.A. (eyewear industry company)
Description: The tasks consisted in the capability analysis and control for the process of insertion of metal core wire-hinge into (cellulose) acetate temple, and in the evaluation/extension of matching methods for the automatic left-right lens pairing on limited mechanical queues based on morphological and colorimetric measures.

Description of Research Topics

The primary interests of my research activity focus on modern computational aspects of semi-supervised learning problems and algorithms, from both theoretical and applicative viewpoint. I have given contributions mainly in the wide-sense *machine learning* areas of **constrained data clustering**, **sparse biosignal processing** and **bioinformatics**, as well as in the areas of **stochastic models for pattern statistics** and **applied dynamical modelling**, as outlined hereafter.

Constrained Clustering: Complexity and Algorithms

Problem statement In the research topic of geometric clustering a central and long-standing problem is the so-called *Minimum Sum-of-Squares Clustering* (MSSC) problem, that consists in finding a k-partition $\{A_1, ..., A_k\}$ of a given set of n points in d-dimensional space, that **globally** minimises the weight $W(A_1, ..., A_k) := \sum_{i=1}^{k} \sum_{x \in A_i} ||x - C_{A_i}||_2^2$ where C_{A_i} is the ℓ_2-norm based centroid of the cluster A_i. This NP-hard problem admits the well-known heuristic k-Means, that however can require exponential time. Incorporating background information into the problem in form of constraints can improve clustering techniques, and often represents workload/data balancing requirements or physical constraints. Recent works in literature have been focusing on problems with **cluster size constraints**, i.e. clustering problems where the sought k-partition is required to satisfy also some cardinality conditions on each cluster, such as $|A_i| = m_i$ (SCC variant) or $|A_i| \in \mathcal{M}$ (RSC variant), where m_i or $\mathcal{M}$ is given. Introducing size constraints can avoid the issues of unbalanced clusters and outliers. However, these problems are more challenging than MSSC due to the additional constraints [J8, M1], and hence are mostly analysed fixing the dimension d and the number of clusters k.
Linear case Initially, such problems were studied in the 1-dimensional case obtaining a computational complexity characterisation for SCC also with different ℓ_p-norms (other than ℓ_2) and for the centroid localisation problem [J13]. The former is proven NP-hard, while the latter is polynomial-time reducible from the long-standing open problem SQRT-Sum, which is in the Counting Hierarchy, introduced by K. Wagner in 1986, lying between PP and PSPACE. Existence of a polynomial time algorithm for SCC would notably place SQRT-Sum in P. Instead, 1-dimensional RSC with Manhattan and Euclidean norms are proven to be solvable in $O(n(k|\mathcal{M}| + n))$ time and $O(n^2)$ space by a dynamic programming technique [C6].
Planar case I then investigated the 2-dimensional extensions [C10, M1], showing that planar SCC and RSC [C6] are strongly NP-hard and do not admit FPTAS, while the case SCC with $k = 2$ in the Euclidean plane can be solved in subquadratic time [C8] within the frame of computational geometry by means of a bound on Erdös k-sets in the plane and the dynamic convex hulls. Instead, endowing the plane with the ℓ_1-norm leads to an **optimal solution** computable in subcubic time [C9] exploiting a separation property by suitable curves, that are generalisations of Voronoi lines to ℓ_1-norm.

Recent advances [J2, C6] show that RSC with Manhattan norm is also NP-hard, and this result implies that even the wide-spread k-**Median variant** of MSSC is NP-hard.

Higher dimensional spaces Furthermore, using real algebraic geometry arguments more general results were obtained on the 2-clustering SCC problem in any fixed d-dimensional space endowed with ℓ_p-norm, for integer $p \geq 1$. In particular, I proved that this problem is solvable in time and space polynomial w.r.t. the instance size and the parameter p by appropriately enumerating the 2-clusterings, which can be bijectively associated to the irreducible algebraic varieties $Z_i = \{(\mu, \lambda, \gamma) \in \mathbb{R}^{2d+1} : \|x_i - \mu\|_p^p - \|x_i - \lambda\|_p^p - \gamma = 0\}$, $i = 1, ..., n$, constructed from the n given points of X; such class of cluster separating hypersurfaces can be polynomial-time enumerated with an adapted Cylindrical Algebraic Decomposition, yielding optimal solutions [M1].

Sparse Biosignal Processing

Sparsity problem model Based on Occam's Razor principle, the sparse representation model is concerned with finding a solution of an underdetermined linear system having the minimal number of non-null components. Precisely, denoting with $\|x\|_0 = \#\text{supp}(x) = \#\{i : x_i \neq 0\}$ the ℓ_0-norm of a vector x, the sparsest linear representation problem consists in $\operatorname{argmin}_{x \in \mathbb{R}^m} \|x\|_0$ subject to $Ax = b$ for a given dictionary matrix $A \in \mathbb{R}^{n \times m}$ and measurement vector $b \in \mathbb{R}^n$. The fact that this problem is combinatorial in nature and was proven to be NP-hard raised the relevant question of devising efficient techniques for finding good approximate solutions, leading often to heuristics and ℓ_1-norm **convex relaxations** in place of ℓ_0.

Theoretical development In the PHuSeLab research group we introduced a novel sparsity recovery technique called Limaps, Lipschitzian Mappings for Sparsity [J3], which is a simple yet efficient scheme obtained by iterating the self-map $G_\lambda : \text{Aff}_{A,b} \to \text{Aff}_{A,b}$ over the affine space $\text{Aff}_{A,b}$ of (feasible) solutions for the above-stated problem, with control parameter $\lambda > 0$, and defined as $G_\lambda(x) = x - Px \odot e^{-\lambda|x|}$. This transformation is composed by the two phases of a Lipschitzian map and the orthogonal projection P onto the null-space $\mathcal{N}(A)$.

First, the repeated application of the map G_λ with a proper parameter scheduling $\{\lambda_i\}_{i=0}^{\infty}$ in Limaps [J3] yields a sequence converging to a solution vector $\bar{x}$, that satisfies necessary conditions for minimality of a smooth ℓ_0-norm relaxation. As the number of iterations increases the smooth relaxation gets closer and closer to the NP-hard combinatorial problem, thus justifying the validity of the iterative scheme. It is also shown that, under loose positive-definiteness conditions, such a point $\bar{x}$ is asymptotically stable in the Lyapunov sense [J3], which entails **robustness** of the algorithm **to trajectory perturbations**. Second, after introducing ℓ_1 relaxations of the problem, it is demonstrated that, under regularity conditions on Lagrange multipliers, the solution thus obtained is asymptotically well approximated with the limit point $\bar{x}$ of Limaps scheme. Moreover, the algorithm also turns out to be **robust to noise** on the measurement b, in the sense that if initial data $\tilde{b} = b + \varepsilon$ is taken, then the noisy solution $\tilde{x}$ differs from noiseless version $\bar{x}$ slightly w.r.t. the noise magnitude $\|\varepsilon\|$.

Application to Image Processing Basing on the common idea of applying sparse representation to capture the essential information, the Limaps algorithm was applied to automatic recognition of face image identity (AFR) [J11, J7]. This general problem strongly depends on many conditions, such as variations in illumination, occlusion, expression, pose, background clutter. The research efforts addressed robust techniques for fully automatic recognition in such uncontrolled conditions with small training sample size, i.e. SSS condition [J1, C1], or large-scale datasets [C7]. Despite these quite difficult working hypotheses, **high precision recognition** can be obtained combining dimensionality reduction by linear discriminant analysis and sparse representation [J11], outperforming other sparsity based classifiers such as SRC, CRC, Lasso and WRSC. When images are transformed into more structured spaces with multiple features and image condition corrections, the AFR system is more robust in predicting the identity. Moreover, a decision process with cascade stages in the system is able to provide a useful reliability degree of the face identity prediction at each stage [J7].

Application to Signal Compression In recent years many techniques have been proposed for ECG signal compression, mostly using wavelet decomposition methods. Within the frame of sparse representation models we developed a compression technique based on the Limaps algorithm that works on dictionaries constructed over natural basis of the ECG signal [J10]. Indeed, after segmentation and period normalisation of the given signal, every segment can be expressed as a vector y and encoded by a solution x of the sparse approximation problem where an error tolerance $\|Ax - y\| \leq \varepsilon$ is accepted in the reconstruction Ax. The dictionary A used for sparse coding is pre-defined through initial fitting. Combining the arithmetic encoding after Limaps yields a sparsity-based algorithm that effectively exploits the typical pattern occurrences in heart beats, leading to very good compression [J10]. An improvement can be obtained by Tikhonov regularisation of the dictionary's

pseudo-inverse in the sparse coding step. Introducing this novelty together with a backup procedure for few hardly sparsifiable segments (typical of beats with heart disease) leads to a mixed scheme with **high compression rate** and with reconstruction quality guarantee in terms of both PRDN and clinically validated measure WWPRD, that performs considerably better than well-established algorithms, such as SPIHT, ARLE, TRE [J9]. The results were also validated on long-term ECG signals, thus leading to potential implementation into portable devices for at-home monitoring.

Learning dictionaries for sparsity Recently, the methods of dictionary design for sparsity based on learning by examples have been demonstrated to outperform those using structured dictionaries, such as generated wavelets, shearlet or other transforms. In this current context I developed R-SVD [J4], an algorithm for **learning dictionaries in sparse representation** models, inspired by the orthogonal Procrustes shape analysis. This substantially consists in applying generalised rotations to groups of dictionary atoms in order to optimise the Frobenius error measure of the sparse representation of the training samples. While maintaining the classical alternating scheme of MOD and K-SVD that consists in repeatedly optimising first the sparse encoding and then the dictionary, the R-SVD algorithm is able to find an optimal dictionary, after few iterations, with **sensibly higher SNR** in extensive experiments with both synthetic and real-world data [J4].

Fractal image compression In fractal image coding based on Generalised Fractal Transforms, one seeks to approximate a target image or signal by the fixed point of a contractive fractal transform operator. Such fixed point can be calculated to arbitrary precision by minimising an IFSM-derived functional that accounts for collage distance (induced by contractive operator), negentropy (negative entropy) and sparsity. We relax the sparsity with the ℓ_1-norm, thus convexifying the functional, which is then minimised by the Sequential Quadratic Programming method [C4], since it is effective on strongly non-linear problems. The reformulated multi-objective scheme augmented with total-variation [C4] leads to highly precise approximation of the target image.

Bioinformatics

My recent research activity in bioinformatics aims at developing tools for the analysis of biological networks characterised by high volume and variety of data, which also poses relevant scalability issues.

Functional enrichment detection One of the problem in analysing protein interaction datasets is the detection of enrichment of functional systems in a network, which is a non-trivial task and requires an explicit null model, owing to the non-uniform distribution of the connectivity degrees of proteins. To this aim, we used the Random Graph with Given Degree Sequence (RGGDS), a model that preserves the degree distribution in a given subgraph. More precisely, given a set V of vertices, a subset $U = \{v_1, v_2, ..., v_k\} \subset V$ and a sequence $\{d_1, d_2, ..., d_k\} \subset \mathbb{N}$ of degrees for vertices of U, a RGGDS is an undirected graph $G =< V, E >$ picked (uniformly) at random from those with an induced subgraph $G(U)$ having degrees $\deg(v_i) = d_i$, for $i = 1, ..., k$. We found that the random variable X_U counting the inner edges of the induced subgraph $G(U)$ is well approximated by the Poisson distribution. This allows for simplification of the enrichment analysis, that abstractly consists in detecting subgraphs with a significantly deviated number of inner edges, and hence allows to **avoid computationally demanding Monte Carlo methods**. We experimented the prediction based on X_U on various PPI networks with very good resulting estimate quality. This improvement was integrated into the well-known STRING system [A3] and the work was published in a highly cited article [J12].

Phylogenetic profiling Comparison of phylogenetic distributions can be a successful approach for predicting functional associations between non-homologous genes that encode proteins from the same complex or pathway. In this sense, we designed a phylogenetic profiling algorithm, called SVD-Phy [J6], based on truncated SVD decomposition of the normalised matrix of alignment bit scores. In particular, this addresses the problem of uninformative profiles giving rise to false positive predictions. Extensive tests on SIMAP database validated against the KEGG pathway demonstrated that the algorithm has substantially improved performance over other phylogenetic profiling methods.

Semi-supervised node ranking Several semi-supervised learning tasks on biomolecular networks can be formalised as a problem of node ranking with a given partial labelling of the nodes. We devised RANKS, RAnking of Nodes with Kernelized Score functions [J5], a flexible algorithmic scheme implemented in R for solving node ranking and classification problems. It uses a semi-supervised learning strategy which relies on similarity between nodes, which is achieved by label propagation, random walk, or kernel functions, one of which is random walk kernel. In particular, the latter type is computed by **not explicitly forming the Gram matrix**, but rather by reducing it to an iterative formula with sequential matrix-vector products. This model of computation is also being implemented in the vertex-centric graph framework GraphChi [C5, A2], whose workspace lies

on secondary memory, thus allowing also for evaluation of large-scale networks. Moreover, RANKS has been successfully applied to gene function prediction, gene disease prioritisation and drug repositioning problems.

Applied Dynamical Models

Neurocomputing models The development of intracellular and multi-unit extracellular recordings allowed researchers to characterize the neuronal receptive fields in primary sensory cortices of several sensory systems. Most works on sensory processing focused on evoked activity, while functional role and dynamics of spontaneous activity (SA) were not well understood yet. We investigated the SA from a dynamical perspective, successfully modelling its firing activity patterns through intermittent chaos generated from logistic map [C11], and studying the predictability of higher order loose synchronies [A4] which shows presence of diverse tasks merged in the shared functional state of SA.
Ecological modelling The analysis of various forest fires has pointed out that burned areas can often be described by different power-law distributions for small, medium and large fires and that a scaling law for the time intervals between consecutive fires is fulfilled. Theoretical arguments explaining such laws have not been fully satisfactory because important physical and/or biological factors controlling forest fires were not taken into account. To overcome this, we developed a two-layer spatially extended forest model [J14] which is essentially deterministic and encapsulates the main characteristics of vegetation growth, and fire ignition and propagation. This model, consisting substantially in a cellular automaton with logistic population ODE cells modified to Lotka-Volterra type, is able to extensively support the various empirical laws.

Bibliometry

Total citations	Scopus: 2272, Google Scholar: 2971
H-index	Scopus: 7, Google Scholar: 8
ORCID iD	https://orcid.org/0000-0002-3299-448X

Software contributed

On Bioconductor	SVD-Phy: R language implementation of the phylogenetic profiling algorithm which uses truncated SVD for predicting functional associations between non-homologous genes. https://omictools.com/svd-phy-tool
On CRAN	RANKS: R language implementation of the node ranking tool for biological networks with kernelised score functions. https://cran.r-project.org/web/packages/RANKS
On Web	R-SVD: Matlab package containing the implementation of R-SVD, an algorithm for dictionary learning in sparsity models based on the orthogonal Procrustes shape analysis. LiMapS & k-LiMapS: Matlab implementation of fast iterative methods for finding sparse solutions to underdetermined linear systems, based on a fixed-point iteration scheme which combines nonconvex Lipschitzian-type mappings with canonical orthogonal projectors. http://phuselab.di.unimi.it/resources.php#software

Computer Skills

Scientific computing:	solid skills in Matlab, Mathematica, R, Python
Other skills:	C/C++, Java, PHP, LaTeX, Linux, Windows

Languages

Italian, Chinese:	Mother tongue
English:	Fluent
German:	Zertifikat Deutsch als Fremdsprache (self-assessment: B2)

Publications

Journals

[J1] V. Cuculo, A. D'Amelio, G. Grossi, R. Lanzarotti, and J. Lin. Robust single-sample face recognition by sparsity-driven sub-dictionary learning using deep features. *Sensors*, 19(1):146, 2019.

[J2] M. Goldwurm, J. Lin, and F. Saccà. On the complexity of clustering with relaxed size constraints in fixed dimension. *Theoretical Computer Science*, 717:37–46, 2018.

[J3] A. Adamo, G. Grossi, R. Lanzarotti, and J. Lin. Sparse Decomposition by Iterating Lipschitzian-type Mappings. *Theoretical Computer Science*, 664:12–28, 2017.

[J4] G. Grossi, R. Lanzarotti, and J. Lin. Orthogonal Procrustes analysis for dictionary learning in sparse representation. *PLoS One*, 12(1):e0169663(1–16), 2017.

[J5] G. Valentini, G. Armano, M. Frasca, J. Lin, M. Mesiti, and M. Re. RANKS: a flexible tool for node label ranking and classification in biological networks. *Bioinformatics*, 32(18):2872–2874, 2016.

[J6] A. Franceschini, J. Lin, C. von Mering, and L. J. Jensen. SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles. *Bioinformatics*, 32(7):1085–1087, 2016.

[J7] G. Grossi, R. Lanzarotti, and J. Lin. Robust face recognition providing the identity and its reliability degree combining sparse representation and multiple features. *International Journal of Pattern Recognition and Artificial Intelligence*, 30(10):1656007–(1–18), 2016.

[J8] J. Lin, A. Bertoni, and M. Goldwurm. Exact algorithms for size constrained 2-clustering in the plane. *Theoretical Computer Science*, 629:80–95, 2016.

[J9] G. Grossi, R. Lanzarotti, and J. Lin. High-rate compression of ECG signals by an accuracy-driven sparsity model relying on natural basis. *Digital Signal Processing*, 45:96–106, 2015.

[J10] A. Adamo, G. Grossi, R. Lanzarotti, and J. Lin. ECG compression retaining the best natural basis k-coefficients via sparse decomposition. *Biomedical Signal Processing and Control*, 15:11–17, 2015.

[J11] A. Adamo, G. Grossi, R. Lanzarotti, and J. Lin. Robust face recognition using sparse representation in LDA space. *Machine Vision and Applications*, 26(6):837–847, 2015.

[J12] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, and L. J. Jensen. String v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(D1):D808–D815, 2013.

[J13] A. Bertoni, M. Goldwurm, J. Lin, and F. Saccà. Size Constrained Distance Clustering: Separation Properties and Some Complexity Results. *Fundamenta Informaticae*, 115(1):125–139, 2012.

[J14] J. Lin and S. Rinaldi. A derivation of the statistical characteristics of forest fires. *Ecological Modelling*, 220(7):898–903, 2009.

Conference proceedings

[C1] M. Bodini, A. D'Amelio, G. Grossi, R. Lanzarotti, and J. Lin. Single Sample Face Recognition by Sparse Recovery of Deep-learned LDA Features. In *Proc. 19th Int. Conf. on Advanced Concepts for Intelligent Vision Systems (ACIVS 2018)*, volume 11182 of *Lecture Notes in Computer Science*, pages 297–308, 2018.

[C2] M. Goldwurm, J. Lin, and M. Vignati. A local limit property for pattern statistics in bicomponent stochastic models. In *Proc. 20th Int. Conf. Descriptional Complexity of Formal Systems (DCFS 2018)*, volume 10952 of *Lecture Notes in Computer Science*, pages 114–125, 2018.

[C3] A. D'Amelio, V. Cuculo, G. Grossi, R. Lanzarotti, and J. Lin. A note on modelling a somatic motor space for affective facial expressions. In *New Trends in Image Analysis and Processing – ICIAP 2017*, volume 10590 of *Lecture Notes in Computer Science*, pages 181–188. Springer, 2017.

[C4] H. Kunze, D. La Torre, and J. Lin. IFSM fractal image compression with entropy and sparsity constraints: A sequential quadratic programming approach. In *ICNPAA 2016 World Congress*, volume 1798 of *American Institute of Physics Conference Series*, pages 020090–1(7), 2017.

[C5] J. Lin, M. Mesiti, M. Re, and G. Valentini. Within network learning on big graphs using secondary memory-based random walk kernels. In *Complex Networks & Their Applications V*, volume 693 of *Studies in Computational Intelligence*, pages 235–245. 2016.

[C6] M. Goldwurm, J. Lin, and F. Saccà. On the Complexity of Clustering with Relaxed Size Constraints. In *Algorithmic Aspects in Information and Management - 11th International Conference (AAIM 2016) Proceedings*, volume 9778 of *Lecture Notes in Computer Science*, pages 26–38, 2016.

[C7] G. Grossi, R. Lanzarotti, and J. Lin. A selection module for large-scale face recognition systems. In *Image Analysis and Processing - ICIAP 2015, Part II*, volume 9280 of *Lecture Notes in Computer Science*, pages 529–539, 2015.

[C8] A. Bertoni, M. Goldwurm, and J. Lin. Exact algorithms for 2-clustering with size constraints in the euclidean plane. In *SOFSEM 2015: Theory and Practice of Computer Science*, volume 8939 of *Lecture Notes in Computer Science*, pages 128–139. 2015.

[C9] A. Bertoni, M. Goldwurm, J. Lin, and L. Pini. Size-constrained 2-Clustering in the Plane with Manhattan Distance. In *Proc. 15th Italian Conference on Theoretical Computer Science (ICTCS 2014)*, volume 1231 of *CEUR Workshop Proceedings*, pages 33–44, 2014.

[C10] J. Lin. Size constrained clustering problems in fixed dimension. In *Proc. 13th Italian Conference on Theoretical Computer Science (ICTCS 2012)*, pages 155–158, Varese, Italy, 2012.

[C11] R. Storchi, A. Zippo, G. Caramenti, J. Lin, and M. Valente. Modeling neuronal ensemble firing activity through intermittent chaos. In *Proc. IEEE 5th International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2010)*, pages 1593–1598, 2010.

Conference abstracts/posters and conferences without proceedings

[A1] S. Lukman, A. P. Kumar, S. Azzam, K. Al Adem, M. N. Nguyen, C. Verma, J. Lin, and S. M. Lee. Integrating artificial intelligence (AI) methods in structural-dynamics based drug discovery workflow: case studies on amylin, protein tyrosine phosphatase 1B and Ras superfamily proteins. 21st-Century Drug Discovery and Development for Global Health (S3) Conference – Keystone Symposia. Berlin, Germany, 2018.

[A2] M. Re, M. Mesiti, M. Frasca, J. Lin, and G. Valentini. Analysis of bio-molecular networks through semi-supervised graph-based learning methods. Third Italian Workshop on Machine Learning and Data Mining - XIII AI*IA Symposium on Artificial Intelligence, Pisa, Italy, 2014.

[A3] A. Franceschini, A. Roth, D. Szklarczyk, M. Simonovic, M. Kuhn, P. Minguez, J. Lin, P. Bork, L. J. Jensen, and C. von Mering. String 9.1 — a resource for protein interaction networks. In: SIB Days, Swiss Institue of Bioinformatics, 2014.

[A4] A. Zippo, R. Storchi, J. Lin, G. Caramenti, M. Valente, and G. E. Biella. Higher order synchrony predictability in somatosensory cortex during spontaneous activity. In *Frontiers in Computational Neuroscience*, Conference Abstract: Bernstein Conference on Computational Neuroscience, 2010. ISSN 1662-5188.

Monographs

[M1] J. Lin. *Exact algorithms for size constrained clustering*. Ledizioni Publishing, Italy, 2013. ISBN 8867050656.

[M2] J. Lin. *Pattern statistics in multicomponent stochastic models: local results in bicomponent models*. Lambert Academic Publishing, Saarbrucken, 2012. ISBN 9783659119835.

Working papers

[U1] M. Goldwurm, J. Lin, and M. Vignati. Saddle point method in the analysis of pattern statistics for words of regular languages. In preparation, 2019.

[U2] D. Conte, G. Grossi, R. Lanzarotti, J. Lin, and A. Petrini. Parallel-Efficient Independent Task Partitioning by MCMC Random Graph Coloring. In preparation, 2019.

[U3] J. Lin, I. Pozzana, Y. Prifti, and A. Provetti. A polynomial extension to eigenvector centrality. In preparation, 2019.

[U4] M. Re, M. Mesiti, J. Lin, and G. Valentini. Notes on local strategies for node label ranking algorithms. In preparation, 2019.

Data 24 febbraio 2019 Luogo Abu Dhabi