BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Luis M. Rocha

eRA COMMONS USER NAME (credential, e.g., agency login): lmrocha

POSITION TITLE: George J. Klir Professor of Systems Science

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Instituto Superior Técnico, Portugal	B.S.	1985-1988	Mechanical Engineering
Instituto Superior Técnico, Portugal	M.S.	1988-1990	Systems Engineering
State University of New York at Binghamton	Ph.D.	1992-1997	Systems Science
Los Alamos National Laboratory	Post Doc	1997-1998	Complex Systems

A. Personal Statement

My research, leadership, and training are on complex networks & systems, computational biology, systems biology, and computational intelligence. Thus, I am very well suited to tackle two key project components, network science and machine learning methods, to integrate and analyze multilevel, heterogeneous biomedical data ranging from multiomics datasets, to pharmacology, electronic health records and social media. My lab has considerable experience in mining literature, social media, and electronic health records in the biomedical and public health domains (§C1-2.) I also have an established record in complex systems and network science, with a focus on multivariate dynamics in biomedical problems (§C3-5). I have strong ongoing collaborations with the relevant personnel on this renewal application. Drs. Borner and Miller were also PIs on the first period of this project (R01LM012832-01), which was very fruitful in joint publications and computational products, and we have been co-PIs in other projects during the last decade. I have also collaborated with the new PI, Dr. Valencia, for many years, from early BioCreative text mining challenges to recent papers on drug interactions and comorbidities using electronic health records. I am committed to interdisciplinary research and graduate training, having created and directed various interdisciplinary PhD programs, including serving as the PI and Director of a recent NSF-NRT interdisciplinary PhD training program in complex networks & systems (NSF-1735095). It included multi-level complexity in health as one of its main areas of research (Dr. Borner was a co-PI of this training grant as well). I have trained and graduated 13 PhD students and 6 postdoctoral fellows in these topics, and currently train 6 PhD students and 2 postdoctoral fellows. I am co-Director of the Consortium for Social and Biomedical Complexity between Indiana and Binghamton Universities. Four recent publications most related to the current project:

a. R.B. Correia, A. Barrat, L. M. Rocha [2023]. "Contact networks have small metric backbones that maintain community structure and are primary transmission subgraphs". *PLoS Computational Biology*. **19**(2): e1010854. DOI:10.1371/journal.pcbi.1010854. PMC9949650.

b. A. Gates, Correia, R.B., X. Wang, L.M. Rocha [2021]. "The effective graph reveals redundancy, canalization, and control pathways in biochemical regulation and signaling." *Proceedings of the National Academy of Sciences*. **118** (12) e2022598118. PMC8000424.

c. Simas, T., R.B Correia and L.M. Rocha [2021]. "The distance backbone of complex networks". *Journal of Complex Networks*. **9** (6), cnab021. DOI: 10.1093/comnet/cnab021. PMC10859673.

d. I.B. Wood, R.B. Correia, W.R. Miller, and L.M. Rocha [2022]. "Small Cohort of Epilepsy Patients Showed Increased Activity on Facebook before Sudden Unexpected Death*". Epilepsy & Behavior*. **128**: 108580. DOI:10.1016/j.yebeh.2022.108580. PMC10582639.

B. Positions, Scientific Appointments, and Honors

Positions and Employment:
1990-1991   <u>Graduate Research Assistant</u>, Laboratório Nacional de Engenharia Civil, Lisbon, Portugal
1995-1997   <u>Adjunct Professor</u>, State University of New York, Binghamton, Department of Systems Science
1997-1999   <u>Postdoctoral Fellow</u>, Los Alamos National Laboratory
1998-2002   <u>Team Leader</u> of the Complex Systems Modeling Team, Los Alamos National Laboratory
1999-2004   <u>Technical Staff Member</u>, Los Alamos National Laboratory
2002-2012   <u>Director</u>, Computational Biology Collaboratorium, and <u>co-director</u> of Ph.D Program in
               Computational Biology, Instituto Gulbenkian de Ciencia, Portugal
2002-   <u>Principal Investigator</u>, Instituto Gulbenkian de Ciencia, Portugal
2004-2013   <u>Associate Professor</u>, Indiana University, Luddy School of Informatics, Computing & Engineering.
2008-2020   <u>Director</u>, Indiana University, Complex Networks & Systems track, Informatics PhD Program,
2013-2021   <u>Professor</u>, Indiana University, Luddy School of Informatics, Computing & Engineering.
2016-2018   <u>Visiting Professor</u>, Neuroscience & Clinical Sciences, Champalimaud Foundation, Portugal
2016-2018   <u>Visiting Professor</u>, Center for Theoretical Physics, University of Aix-Marseille, France
2017-2023   <u>Director</u>, NSF-NRT Interdisciplinary Training Program in Complex Networks and Systems.
2019-2021   <u>Visiting Professor</u>, NOVA School of Business and Economics, Portugal
2019-   <u>Director</u>, Consortium for Social and Biomedical Complexity, Binghamton & Indiana University
2021-   <u>George J. Klir Professor of Systems Science</u>, Binghamton University, SUNY

Other Experience and Professional Service (selected):
2006   Chair, Organizing Committee, Artificial Life X
2011-   Editorial Board, PLoS ONE
2011   Member, NSF, Networks, Synthetic Biology and Evolution review panel
2014-   Editorial Board, Frontiers in Robotics and AI | Computational Intelligence
2014   Site Visit Committee of the *Biocomputational Evolution in Action CONsortium* (BEACON)
               NSF Science and Technology Center at Michigan State University
2017-   Editorial Board, BMC Bioinformatics
2017-   Editorial Board, Complexity
2018 -2021   Advisory Council, *Indiana University Network Science Institute* (IUNI)
2018   External Review Committee, *Center for the Study of Complex Systems*, University of Michigan
2018 -2021   Council (elected member), Complex Systems Society (CCS)
2019   General Chair, Complex Networks 2019
2020-2024   Program Chair, Complex Networks 2020, 2021, 2022, 2023, 2024
2020   Reviewer, Biomedical Computing and Health Informatics (BCHI) study section at the CSR NIH
2021,2023   Reviewer, Clinical Informatics and Digital Health (CIDH) study section at the CSR NIH
2021-   Editorial Board, Journal of Computational Science
2022-   Editorial Board, Frontiers in Physics - Complex Systems section
2023   Reviewer, Clinical Data Management and Analysis (CDMA) and study section at the CSR NIH
2023-   Section Editor, *PLoS Complex Systems*, Complex Systems in Nature section
2024-   External Advisory Board, Environmental Health Institute (Instituto de Saúde Ambiental),
               University of Lisbon Medical School (ISAMB-FMUL), Portugal

I regularly serve on the program committees of conferences in my field and referee articles for peer-review journals such as: *Adaptive Behavior, Adv. Complex Syst., Artificial Intelligence, Artificial Life, Biosystems, BMC Bioinformatics, BMC Syst. Biology, Cognitive Science, Complex Systems, Data Mining & Knowledge Discovery, J. Biomedical Informatics, Genome Biology, IEEE/ACM Trans. Comp. Biology & Bioinformatics, J. of Theor. Biology, Molec. Biosystems, Nature Communications, Network Science, Nucleic Acids Research, PLOS Computational Biology, PLOS ONE, PNAS, Phys. Rev. E, Science Advances, Scientific Reports*.

Honors and Awards (selected):
2015   Trustees Award for Teaching Excellence Award 2015, Indiana University
2016   Top 30 paper (PMC4720984) in translational bioinformatics. Russ Altman 2016 year in review.
2016-   Fulbright Scholar, United States Department of State Bureau of Educational and Cultural Affairs

Media Coverage (selected):
Coverage of our *PLoS Computational Biology* paper on backbones for epidemic modeling. February-March,

2023: The Print, Technology Networks, Reporter Health, MedicalXpress, EurekaAlert, laSexta, Expresso, etc.

Coverage of our *Epilepsy & Behavior* paper, March 1, 2022, *National Geographic, CNN Portugal, Observador, RTP,Jornal Online, etc.*

March 16th, 2021. Coverage of our PNAS paper and cover on "Control pathways in biochemical regulation and signaling": Publico, Sabado, Jornal Economico, etc.

Medical News, Globo News, Futurity, Publico, Medical X Press, Diario de Noticias, SIC TV, etc. Wide Coverage of our work on "drug interactions and gender and age biases." August, 2019.

Time, Newsweek, The Independent, The Times, Mother Jones, The Atalantic, etc. Viral Coverage of our paper "Human Sexual Cycles are Driven by Culture and Match Collective Moods." December21-30, 2017.

Reddit top trending topic: "After studying 129 countries, academics discovered that the majority of people become more interested in sex and pornography around religious festivals."

Nature **522**, 395 (25 June 2015): "Computer fact-checker and news reader grab attention online."

Computational Fact-Checking, June 2015: various news pieces at *Wired, ACM Tech news, Deutschlandfunk, GCN, Pacific Standard, Phys.org, Scientific Computing, GizModo, Big Think*, etc.

EuroScientist, January 28, 2015: "When Real Science Falls Short In Hollywood."

Nautilus, USA, April 16, 2014: "Scientists Create Cybernetic Links Between People—by DJing"

RDP-Antena 2, Portugal: Three radio interviews devoted to Alan Turing and AI: Dec. 2012 - Jan2013

Público, Portugal: "O cérebro é uma máquina de Turing?", December 27, 2012.

La RepubBlica, Italy: "L'Italia, un paese felice e lo scrive su Facebook", December 23, 2011.

The Daily Telegraph, U.K.: "Can we make software that comes to life?", August 5th, 2008.

New York Times. "On Facebook, Scholars Link Up With Data." December 17, 2007.

The Washington Post: "Web May Hold the Key to Achieving Artificial Intelligence" in Sep. 6, 2002;

Recent Invited Lectures (selected):

Social Media Mining for Health 2023 (#SMM4H). Workshop at AMIA 2023 Annual Symposium, New Orleans, 11/11/2023. Keynote Speaker: "Network science to characterize health vulnerabilities and biases."

Multiscale & Integrative Complex Networks (MIX-NEXT), Satellite Symposium at NetSci 2023, Vienna, Austria, 7/10/2023. Keynote Speaker: "Redundancy in the Structure and Dynamics of Complex Networks."

Network Science for The Sustainable Development Goals, Satellite Symposium NetSci 2023, Vienna, Austria, 7/11/2023. Keynote Speaker: "Network science to characterize health vulnerabilities and biases."

Boolean networks and related modeling frameworks Symposium at the ECMTB 2022, Heidelberg, Germany, 9/19-23/2022. Invited Speaker: "The effective graph: nonlinear canalization in biochemical systems."

Northwestern Institute on Complex Systems Seminar Series, Northwestern University, 5/11/2022 Colloquium Speaker: "Redundancy in the Structure and Dynamics of Complex Systems."

Gender Bias in Artificial Intelligence and Health: Building a Future for Equality, Barcelona Supercomputing Center, Spain, April 14, 2021. Keynote Speaker: "Sex and Gender perspective in Personalized Medicine."

NERCCS 2021: 4thNortheast Reg. Conf. on Complex Systems, Online, March 31-April 2. Keynote Speaker.

5th European Conference on Translational Bioinformatics: Biomedical Big Data Supporting Precision Medicine, Madrid, November 30th-December 1st, 2020. Keynote Speaker.

Universidad Nacional del Sur, Bahia Blanca, Argentina, November 20-22, 2018 Lecture Series Speaker: "Interdisciplinarity, Data Science and Complex Systems".

Network Medicine: Personalized Medicine in the Era of Big Data. NetSci 2018 Satellite. Paris, France. June 11 Keynote Speaker: "The multi-level complexity of human health: integrating cohort-specific data".

2nd Week of Complexity Sciences, Universidad Nacional Autonoma de México, Mexico City, January 31st 2018 Keynote Speaker: "Towards understanding the multi-level complexity of human health: from drug-interaction to human reproduction cycles".

C. Contributions to Science

1. **Computational Social Science for Public Health Monitoring**. My group is one of the first to use social media data to study collective social behavior in biomedical problems. For instance, we were the first to use *Instagram* for pharmacovigilance of drug interactions, adverse reactions, and behavior pathology, focusing on depression and epilepsy, and *Facebook* to study sudden death in epilepsy. The work demonstrates that social media provides a very promising source of large-scale data for monitoring and understanding public health in ways that have not been hitherto possible. Indeed, given the large number of users, social media data allows us to identify under-reported, population-level pathology. My group and collaborators have used other sources of large-scale data to tackle problems of biomedical and public-

health relevance, such as: *electronic health records* for studying gender and age biases and comorbidity in medical care, *Twitter* and *Google Trend*s data to study human reproductive behavior on a global scale, and *Wikipedia* to automatically establish the veracity of online statements. Another very important contribution is a focus on the design of human-centered, personalized applications and biomedical corpora.

a. A. Min, W.R. Miller, L.M. Rocha, K. Borner, R.B. Correia, and P.C. Shih. [2023]. "Understanding Contexts and Challenges of Information Management for Epilepsy Care" In: *Proc. SIGCHI Conf Human Factors Computing Systems* (CHI '23). DOI:/10.1145/3544548.3580949. PMC10544776.
b. R.B. Correia, L.P. de Araújo, M.M. Mattos, D. Wild and L.M. Rocha [2019]. "City-wide Analysis of Electronic Health Records Reveals Gender and Age Biases in the Administration of Known Drug-Drug Interactions." *NPJ Digital Medicine*. **2**:74. PMC6650500.
c. I. B. Wood, P.L. Varela, J. Bollen, L.M. Rocha, and M.J. Sá [2017] "Human Sexual Behavior is driven by culture and collective moods." *Scientific reports* **7** (1): 17973. PMC5740080.
a. G.L.Ciampaglia, P.Shiralkar, L.M.Rocha, J. Bollen, F. Menczer, A. Flammini [2015]. "Computational fact checking from knowledge networks."*PloS ONE*. **10**(6): e0128193. PMC4471100.

2. **Multilevel Complex Systems Approach to Biomedical Literature Mining**. My group has been involved in this field from its very start, having participated successfully in the first four BioCreAtIvE (Critical Assessment for Information Extraction) between 2004 and 2012. Much of biomedical research relies on the inference of correlations and interactions from data at multiple levels of the biological organization from the molecular to the social. To understand the biochemical, functional, and behavioral roles of genes and proteins in organisms, there is a need to integrate and study multilevel associations reported in the literature and multiomics databases, as well as exposome data via electronic health records, social media, and digital phenotyping. My contributions to this goal have been the development of novel methods based on network science and complex systems. This data-driven approach has enabled the automatic discovery, classification and annotation of protein-protein and drug-drug interactions, biosocial health risks, pharmacokinetic parameters in drug interaction and adverse reaction studies, protein sequence and structure prediction, functional annotation of transcription data, enzyme annotation publications, etc.

a. R.B. Correia, I.B Wood, J. Bollen, L.M. Rocha [2020]. "Mining social media data for biomedical signals and health-related behavior". *Annual Review of Biomedical Data Science, 3:1.* DOI: 10.1146/annurev-biodatasci-030320-040844. PMC7299233.
b. J. Sánchez-Valle, R.B. Correia, M. Camacho-Artacho, R. Lepore, M.M. Mattos, L. M. Rocha, A. Valencia [2023]. "Analysis of electronic health records from three distinct and large populations reveals high prevalence and biases in the co-administration of drugs known to interact." *medRxiv* 2023.02.06.23285566; DOI:10.1101/2023.02.06.23285566.
c. A. Kolchinsky, A. Lourenço, H. Wu, L. Li, and L.M. Rocha. [2015] "Extraction of Pharmacokinetic Evidence of Drug-drug Interactions from the literature." *PLoS ONE* **10**(5): e0122199. PMC4427505.
d. S. Zhang, H. Wu, L. Wang, G. Zhang, L.M. Rocha, H. Shatkay, L. Li [2022]. "Translational drug–interaction corpus". *Database*, **baac031**. DOI:10.1093/database/baac031. PMC9216474.

3. **Redundancy and Control in Complex Networks.** Network science has provided many insights into the organization of complex systems, capturing multilevel interactions as (multilayer) graphs or multivariate dynamical systems. As the field matures, however, there is a need to move from understanding to controlling the organization of complex systems. This is particularly true in systems biology and medicine, where increasingly accurate models of biochemical regulation are being produced. I have contributed to this advancement with two mathematical concepts developed in my group that allow us to remove different forms of redundancy in networks: distance closures and dynamical canalization. The first concept allows us to infer the invariant subgraph that is sufficient to compute all shortest paths in a weighted graph. This has demonstrated that there is massive redundancy in many networks in different domains, whereby most edges in a network are not necessary to compute shortest paths (e.g., 99% of conserved protein interactions in male germ cells). Removing redundant edges facilitates computation and discovery of important regulatory pathways. The second concept is used to remove redundancy from the logical rules of biochemical regulation models in systems biology, revealing that most variables (e.g., chemical species) rely on a small subset of their inputs for regulation (canalization). The removal of this redundancy simplifies multivariate dynamics into scalable graphs that reveal how to control large biochemical models, which are otherwise too large to study analytically.

    a. R.B. Correia, et al [2024]."The conserved transcriptional program of metazoan male germ cells uncovers ancient origins of human infertility". *eLife*. In Press. *bioRxiv* 2022.03.02.482557, DOI:10.1101/2022.03.02.482557.

    b. A. Gates, A. and L.M. Rocha [2016]. "Control of complex networks requires both structure and dynamics". *Scientific Reports*. 6, 24456. PMC4834509.

    c. S. Manicka, M. Marques-Pita, and L.M. Rocha [2022]. Effective connectivity determines the critical dynamics of biochemical networks. *J. R. Soc. Interface*. **19**(186):20210659. PMC8767216.

    d. L.M. Rocha [2022]. "On the feasibility of dynamical analysis of network models of biochemical regulation." *Bioinformatics*. **btac360**, DOI: 10.1093/bioinformatics/btac360. PMC9272802.

4. **Multivariate and Network Dynamics.** Modern social and biomedical analysis relies on large complex multivariate timeseries data. It is thus essential to develop methods to infer time-varying data associations such as pairwise variable interactions for reverse-engineering networks, discovering subsets of variables that mostly interact with one another (modularity). Our contributions to these problems have been in the areas of spectral methods, statistical inference, information theory, and multilayer network analysis. These are being used to uncover interactions and multiscale modularity in various domains, such as gene regulation, transcriptomics, multiomics and phylogenetics in disease, epidemiology, and brain activity.

    a. A. Kolchinsky, M.P. van den Heuvel, A. Griffa, P. Hagmann, L.M. Rocha, O. Sporns and J. Goñi [2014]. "Multi-scale Integration and Predictability in Resting State Brain Activity". *Front. Neuroinformatics. 8:66.* PMC4109611.

    b. D.S.Paños, , F.X. Costa, and L.M. Rocha. [2023]. "Semi-metric topology characterizes epidemic spreading on complex networks". *arXiv*:2311.14817. DOI: 10.48550/arXiv.2311.14817

    c. T. Parmer, L.M. Rocha, F. Radicchi [2022]. "Influence maximization in Boolean Networks." *Nature Communications*. **13**, 3457, DOI:10.1038/s41467-022-31066-0. PMC9203747.

    d. M.E. Wall, A. Rechtesteiner, and L.M. Rocha [2003]. "Singular Value Decomposition and Principal Component Analysis ". In: *Understanding and Using Microarray Analysis Techniques: A Practical Guide*. D. P. Berrar, W. Dubitzky, and M. Granzow (Eds.). Kluwer Academic, pp. 91-109.

5. **Automata Models of Evolutionary Systems**. I have also contributed to theoretical biology by developing computational models of systems whose evolutionary role or adaptive capabilities are not well understood. This includes the first computational model of RNA Editing and an agent-based model of T-Cell cross-regulation. We have also studyied how the dynamics of many existing systems biology models relate to their evolvability. More generally, I have contributed to the study of the interplay between self-organization and natural selection by introducing the concept of selected self-organization and developing bio-inspired computing and agent-based models to study it.

    a. M. Marques-Pita and L.M.Rocha [2013]. "Canalization and control in automata networks: body segmentation in Drosophila melanogaster." *PLOS One*, **8**(3): e55946. PMC3592869.

    b. L.M. Rocha, C. Huang, A. Maguitman, J. Kaur. [2007]."Agent-based Model of Genotype Editing". *Evolutionary Computation*. **15** (3): 253-89. PMID17705779.

    c. L.M. Rocha [2001]. "Evolution with material symbol systems". *Biosystems*. **60**: 95-121. PMID11325506

    a. K.H. Park, F.X. Costa, L.M. Rocha, R. Albert, J.C. Rozum [2023]. "Robustness of biomolecular networks suggests functional modules far from the edge of chaos". *PRX Life*. **1**, 023009. DOI: 10.1103/PRXLife.1.023009. NIHMS1965674.

Complete List of Published Work:
http://www.ncbi.nlm.nih.gov/sites/myncbi/luis.rocha.1/bibliography/49501628/public/
or http://orcid.org/0000-0001-9402-887X