# 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. This expertise is very well suited to tackle two key components of this project: social media mining for sentiment and biomedical signals, and analysis of multivariate time-series data. My lab has considerable experience in literature, social media, and electronic health records mining in the biomedical and public health domains (§C1-2.) Indeed, I received a Fulbright scholarship to work in this domain, and our pioneering work on longitudinal analysis of Instagram and Twitter user timelines to monitor biomedical phenomena is well known. I also have an established record in multivariate time-series analysis and modeling of dynamical processes in biology (§C3-5), with one of the most cited papers on the singular value decomposition and much work on evolutionary algorithms−both central to the project. I am the PI of an NIH-NLM R01 (1R01LM012832-01) that focuses on integrating many sources of data, including social media, to provide a personalized web service for epilepsy management, as well as an NSF-NRT interdisciplinary PhD training program in complex networks & systems, which includes multi-level complexity in health as one of its main areas of research. I have trained and graduated 11 PhD students and 3 postdoctoral fellows in these topics, and currently train 10 PhD students and two postdoctoral fellows. I am co-Director of the Consortium for Social and Biomedical Complexity with co-PI Bollen, with whom I have a longstanding collaboration. Four recent publications most related to current project are:

a. Parmer, T., L. M. Rocha, F. Radicchi [2022]. "Influence maximization in Boolean Networks." *Nature Communications*. In press.
b. 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.
c. 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.
d. 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.

## B. Positions, Scientific Appointments, and Honors

### Positions and Employment:
1990-1991    Graduate Research Assistant, Laboratório Nacional de Engenharia Civil, Lisbon, Portugal
1995-1997    Adjunct Professor, State University of New York, Binghamton, Department of Systems Science
1997-1999    Postdoctoral Fellow, Los Alamos National Laboratory
1998-2002    Team Leader of the Complex Systems Modeling Team, Los Alamos National Laboratory
1999-2004    Technical Staff Member, Los Alamos National Laboratory
2002-2012    Director, Computational Biology Collaboratorium, and co-director of Ph.D Program in
          Computational Biology, Instituto Gulbenkian de Ciencia, Portugal
2004-2013    Associate Professor, Indiana University, Luddy School of Informatics, Computing & Engineering.
2013-2021    Professor, Indiana University, Luddy School of Informatics, Computing & Engineering.
2016-2018    Visiting Professor, Neuroscience & Clinical Sciences, Champalimaud Foundation, Portugal
2016-2018    Visiting Professor, Center for Theoretical Physics, University of Aix-Marseille, France
2017-    Director, NSF-NRT Interdisciplinary Training Program in Complex Networks and Systems.
2019-2021    Visiting Professor, NOVA School of Business and Economics, Portugal
2019-    Director, Center for Social and Biomedical Complexity, Indiana University
2020-2021    Visiting Professor, Binghamton University, SUNY
2021-    George J. Klir Professor of Systems Science, 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 -    Advisory Council, *Indiana University Network Science Institute* (IUNI)
2018    External Review Committee, *Center for the Study of Complex Systems*, University of Michigan
2018 -    Council (elected member), Complex Systems Society (CCS)
2019    General Chair, Complex Networks 2019
2020    Program Chair, Complex Networks 2020
2020    Reviewer, Biomedical Computing and Health Informatics (BCHI) study section at the CSR NIH.
2021    Reviewer, Clinical Informatics and Digital Health (CIDH) study section at the CSR NIH.
2021    Program Chair, Complex Networks 2021
2021-    Editorial Board, Journal of Computational Science

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 ONE, PNAS, Phys. Rev. E, Science Advances, Scientific Reports,* etc.

### 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):
*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, 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):
*Gender Bias In Artificial Intelligence And Health: Building a Future for Equality, Barcelona Supercomputing Center, Spain, April 14, 2021. Keynote Speaker.*
*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: from biomedicine to collective intelligence"
*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".
*Controlling Complex Systems. NetSci2017 Satellite. Indianapolis,* Indiana. June 19-20, 2017
   Invited Speaker: "Canalization in the dynamics of complex networks drives dynamics, criticality and control".
*Emerging Activity. Relating Things. Humboldt-Universität zu Berlin,* February 17, 2017
   Keynote Speaker: "Material Turing Machines in Life and Collective Intelligence".
*Control and Observability of Network Dynamics, Mathematical Biosciences Institute workshop,* April 11-15, 2016, Columbus, OH. Invited Speaker: "Redundancy and control in complex networks."

## C. Contributions to Science

1. **Computational Social Science for Public Health Monitoring**. My group has been one of the first to use social media data to study collective social behavior in biomedical problems. For instance, my group was the first to use *Instagram* to build public health monitoring and surveillance tools for discovering drug interactions, adverse reactions, and behavior pathology, focusing on depression and epilepsy. This recent work demonstrates that the universe of social media provides a very promising source of large-scale data that can help with 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; *Wikipedia* to automatically establish the veracity of online statements; *Instagram* to predict violence and social unrest situations, etc. Very important for this area is the development of biomedical corpora and dictionaries to mine social media and the literature.

   a. A. Min, W. Miller, L.M. Rocha, K.Börner, R.B. Correia, and P.C. Shih. [2021]. "Just In Time: Challenges and Opportunities of First Aid Care Information Sharing for Supporting Epileptic Seizure Response". *Proc. of the ACM on Human-Computer Interaction*. **5**, CSCW1, Article 113, pp. 1-24.

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. **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 the research presently conducted in the biomedical domain relies on the inference of correlations and interactions from data at multiple levels of the biological organization: from the molecular to the social. Because we ultimately want to increase our knowledge of the biochemical, functional and behavioral roles of genes and proteins in organisms, there is a clear need to integrate the associations and interactions among biological entities that have been reported and accumulated in the literature, electronic health records, and experimental databases. My contributions to the Biomedical literature mining have been the development of novel methods based on network science or bio-inspired computing. This data-driven approach has enabled the automatic discovery, classification and annotation of protein-protein and drug-drug interactions, 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. NIHMSID:1596369.

b. R.B. Correia, L. Li, L.M. Rocha [2016]. "Monitoring potential drug interactions and reactions via network analysis of Instagram user timeliness". *Pac. Symp. Biocomp*. **21**:492-503. PMC4720984.

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. A. Abi-Haidar, J. Kaur, A. Maguitman, P. Radivojac, A. Retchsteiner, K. Verspoor, Z. Wang, and L.M. Rocha [2008]."Uncovering protein-protein interactions in abstracts and text using linear models and word proximity networks ". *Genome Biology*. **9**(Suppl 2):S1. PMC2559982.

3. **Redundancy and Control in Complex Networks.** Network science has provided many insights into the organization of complex systems. The success of this approach is its ability to capture the organization of multivariate interactions as networks or graphs without explicit dynamical rules for node variables. As the field matures, however, there is a need to move from understanding to controlling complex systems. This is particularly true in systems biology and medicine, where increasingly accurate models of biochemical regulation have been produced. I have contributed to this goal with two mathematical concepts developed in my group which allow us to remove different forms of redundancy in networks: 1) distance closures, and 2) canalization via schema re-description. 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. 90% of edges in some brain networks). Removing redundant edges can facilitate computation and discovery of important pathways in many applications. 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 smaller subset of their inputs to be regulated (canalization). The removal of this redundancy simplifies and indeed enables the characterization of control in large biochemical and neural network models, which are otherwise too large to study analytically

a. R.B. Correia, A. Barrat, L. M. Rocha [2022]. "The metric backbone preserves community structure and is a primary transmission subgraph in contact networks". *bioRxiv* 2022.02.02.478784, doi: 10.1101/2022.02.02.478784.

b. Manicka, S., M. Marques-Pita, and L.M. Rocha [2022]. Effective connectivity determines the critical dynamics of biochemical networks. *J. R. Soc. Interface*. **19**(186):20210659. PMID: 35042384

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

d.  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.

4.  **Multivariate Time-Series Analysis and Network Inference.** In the age of data-science, it is essential to develop methods to infer time-varying data associations such as pairwise variable interactions and subsets of variables that mostly interact with one another (modularity). My lab's contribution to the problem of inference on networks and multivariate dynamics has been in the area of spectral methods, statistical inference, and information theory, which has been used to uncover interactions and multiscale modularity in various domains, such as gener regulation, transcriptomics and brain activity time-series data.

    a.  R.B. Correia, J.M. Almeida, Ma.J. Wyrwoll, I.Julca, D. Sobral, C.S. Misra, L.G. Guilgur, H.-C. Schuppe, N. Silva, P. Prudêncio, A. Nóvoa, A. S. Leocádio, J. Bom, M. Mallo, S. Kliesch, M. Mutwil, L.M. Rocha, F.Tüttelmann, J. D. Becker, P. Navarro-Costa [2022]. "The conserved transcriptional program of metazoan male germ cells uncovers ancient origins of human infertility". *bioRxiv* 2022.03.02.482557, doi: 10.1101/2022.03.02.482557.
    b.  A. Kolchinsky, A. Gates and L.M. Rocha. [2015] "Modularity and the spread of perturbations in complex dynamical systems." *Phys. Rev. E Rapid Communications*. **92**, 060801(R).
    c.  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.
    d.  Wall, M.E., 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.  **Agent-Based 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. For instance, I provided the first computational model that demonstrates the evolutionary potential of RNA Editing and a model that proves the ability of T-Cell cross-regulation to classify self from non-self in the presence of populations of changing pathogens. I am also developing bio-inspired methods for spam detection and text classification from that model. 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 evolutionary algorithms and agent-based models to study it.

    a.  A. Abi-Haidar and L.M. Rocha [2011]. "Collective Classification of Textual Documents by Guided Self-Organization in T-Cell Cross-Regulation Dynamics". *Evolutionary Intelligence*. **4**(2):69-80.
    b.  L.M. Rocha, C. Huang, A. Maguitman, J. Kaur. [2007]."Agent-based Model of Genotype Editing". *Evolutionary Computation*. **15** (3): 253-89.
    a.  L.M. Rocha And W. Hordijk [2005]. "Material Representations: From the Genetic Code to the Evolution of Cellular Automata". *Artificial Life*. **11** (1-2), pp. 189 - 214.
    b.  L.M. Rocha [2001]. "Evolution with material symbol systems". *Biosystems*. **60**: 95-121.

**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**

**D. Research Support**

**Relevant Ongoing Research Support**

NSF – 1735095                                                 Rocha (PI)      2017-2022
NRT: *Interdisciplinary Training in Complex Networks and Systems.* Large training grant whereby trainees
    enroll in a dual-major PhD program in Complex Networks and Systems and an empirical domain such as
    biology, neuroscience, economics, cognitive science, sociology, etc. We expect to involve students from
    this training grant in proposed research.


PR-IUB:                                                       Rocha (co-PI) 2019-2021
Prepared for Environmental Change (PfEC) Grand Challenge initiative, Indiana University
*A River Runs Through It: Restoring Biodiversity and Empowering Resilient Community-Environment
    Interactions in Urban Waterways.*

Fundação para a Ciência e Tecnologia, Portugal, DSAIPA/DS/0102/2019-2022.
*Factors to Promote Healthy Dialogue and Behaviours in Online School Communities*
    Co-PI,  Instituto Gulbenkian de Ciencia , €222K (PI Manuel Marques-Pita)

Fundação para a Ciência e Tecnologia, Portugal - DSAIPA/AI/0087/2018  (core team) 2018-2021
*Identification and Forecasting Hospital Emergency Demand*

Fundação para a Ciência e Tecnologia, Portugal, PTDC/MEC-AND/30221/2017 Rocha (core) 2018-2021
*The sperm cell core genetic program: combined clinical and research approach to the diagnosis of male
    infertility*

## **Relevant Completed Research Support**

NIH - R01LM011945-01                              Rocha, Li, Shatkay (MPI)      2014-2019
BLR: Evidence-based Drug-Interaction Discovery: In-Vivo, In-Vitro and Clinical
*Literature and Text mining to uncover distinct types of experimental evidence of drug interactions.*


NSF - 1754207, Division Of Environmental Biology (DEB)          Martinez (PI), Rocha (co-PI) 2018-2019
*OPUS: Synthesizing Ecology with Ecological Networks*

Indiana Clinical Translational Sciences Institute (ICTSI).              Miller, Rocha (MPI)     2016-2018
*Sudden Unexpected Death in Epilepsy: Identifying Risk Factors with Social Media Mining.*


Persistent Systems, Inc.                                      Rocha (PI)      2014-2017
*Large-Scale Text and Social Data Analytics for Health.* Allowed us to develop our pipeline for extraction of DDI
    from Instagram and Twitter.

Intelligence Advanced Research Projects Activity (IARPA),
Open Source Indicators                                       Rocha (IU co-PI)  2012-2014
*EMBERS: Early Model-Based Event Recognition using Surrogates*, with Virginia Tech. Allowed us to develop
    our Twitter analysis pipeline for mood and biomedical dictionaries.

Fundação Luso-Americana para o Desenvolvimento (Portugal) and
National Science Foundation (USA)                            Rocha (co-PI)     2012-2014
*Network Mining for Gene Regulation and Biochemical Signaling*