

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: Professor of Informatics and Computing

Professor of Cognitive 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 systems, computational biology, systems biology, and computational intelligence. This expertise is very well suited to tackle two key components of this project: network analysis and machine learning methods to extract and link personalized biomedical data ranging from molecular pharmacology to electronic health records and social media. My lab has considerable experience in literature, social media, and electronic health records mining in the biomedical and public health domains (sections C1 and C2.) Indeed, recently, I received a Fulbright scholarship to work in this domain, a recent paper of ours (PMC4720984) was deemed one of the most important papers of the year in translational bioinformatics by Russ Altman's 2016 translational science year in review. I am also the PI of a new 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, and a member of the advisory council of the Indiana University Network Science Institute (IUNI). My lab has been involved in many biomedical projects using these techniques at DOE/LANL (1998-2004), Indiana University, and the Instituto Gulbenkian de Ciencia; I have also trained and graduated 9 PhD students and 3 postdoctoral fellows in these topics, and currently train 7 PhD students. I am the chair of the complex networks & systems PhD program at Indiana University and co-director of a PhD program in Computational Biology at the Instituto Gulbenkian de Ciencia. Four recent publications most related to current project are:

- a. R.B. Correia, L.P. de Araújo, M.M. Mattos, D. Wild and L.M. Rocha [2018]. "City-wide Analysis of Electronic Health Records Reveals Gender and Age Biases in the Administration of Known Drug-Drug Interactions." *arXiv:1803.03571*.
- b. 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. 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.
- b. T. Simas and L.M. Rocha [2015]. "Distance Closures on Complex Networks." *Network Science* **3**(2):227-268.

B. Positions 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- 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, School of Informatics, Computing & Engineering.
2013- Professor, Indiana University, School of Informatics, Computing & Engineering.
2016-2018 Visiting Professor, Neuroscience & Clinical Sciences, Champalimaud Foundation, Portugal
2016- Visiting Professor, Center for Theoretical Physics, University of Aix-Marseille, France
2017- Director, NSF-NRT Interdisciplinary Training Program in Complex Networks and Systems, Indiana University

Other Experience and Professional Service (selected):

2006 Chair, Organizing Committee, Artificial Life X
2010 Member, NSF, CCF Core programs, Science and Technology Center Site review
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 Program Chair, Complex Networks 2018
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, Complex Systems Society (CCS)

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*, *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):

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." December 21-30, 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 - Jan 2013
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.
BBC Digital Planet, U.K.: World Service interview about immune-inspired spam-detection system, 2008.

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):

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".

Jardins de Verão, Fundação Gulbenkian, Lisbon, Portugal, July 15, 2017

Keynote Speaker: "From social networks to biological behaviours".

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."

Inference on Networks: New Models and New Data, Santa Fe Institute workshop December 14-18, 2015, Santa Fe, NM. Invited Speaker: "Redundancy and control in complex networks."

Translational Bioinformatics Conference (TCB 2015), November 7-9th, Tokyo, Japan. Highlight Speaker: "Extraction of pharmacokinetic evidence of drug Interactions in literature and social media."

C. Contribution 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 online media to tackle problems of public-health relevance, such as using *Twitter* and *Google Trends* 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. 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.
- 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. 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.
- d. H. Wu, S. Karnik, A. Subhadarshini, Z. Wang, S. Philips, X. Han, C. Chiang, L. Liu, M. Boustani, L.M. Rocha, S.K. Quinney, D.A. Flockhart and L. Li [2013]. "An Integrated Pharmacokinetics Ontology and Corpus for Text Mining". *BMC Bioinformatics.* **14**:35. PMC3571923.

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. 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.
- b. A. Lourenço, M. Conover, A. Nematzadeh, F. Pan, A. Wong, H. Shatkay, and L.M. Rocha [2011]. "A Linear Classifier Based on Entity Recognition Tools and a Statistical Approach to Method Extraction in the Protein-Protein Interaction Literature". *BMC Bioinformatics*, 12 (Suppl 8):S12. PMC3269935.
- c. 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.
- d. K. Verspoor, J. Cohn, C. Joslyn, S. Mniszewski, A. Rechtsteiner, L.M. Rocha, T. Simas [2005]. "Protein Annotation as Term Categorization in the Gene Ontology using Word Proximity Networks". *BMC Bioinformatics*. **6**(Suppl 1): S20. PMC1869013.

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. Correia, R.B., A. Gates, X. Wang, L.M. Rocha [2018]. "CANA: A python package for quantifying control and canalization in Boolean Networks." *Frontiers in Physiology: Systems Biology*. **9**: 1046. DOI: 10.3389/fphys.2018.01046.
- b. A. Gates, A. and L.M. Rocha [2016]. "Control of complex networks requires both structure and dynamics". *Scientific Reports*. **6**, 24456. PMC4834509.
- c. T. Simas and L.M.Rocha [2015]. "Distance Closures on Complex Networks." *Network Science* **3**(2):227-268.
- 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 transcriptomics and brain activity time-series data.

- a. 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).
- b. 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.
- c. J. Challacombe, A. Rechtsteiner, L.M. Rocha, E.P. Brown, T. Shenk, M. Altherr, T. Brettin [2004]. "Evaluation of the host transcriptional response to human cytomegalovirus infection" *Physiol. Genomics*. 10.1152/physiolgenomics.00155.2003.
- d. Wall, M.E., A. Rechtsteiner, 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 computational 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

Ongoing Research Support

NIH - R01LM012832-01	Rocha, Borner, Miller(MPI)	2018-2022
<i>myAURA: Personalized Web Service for Epilepsy Management</i>		
Development of a personalized health library for epilepsy patients. Experts in informatics and epilepsy self-management come together with engaged stakeholders to develop novel data- and network science methods, which can sift through vast amounts of information from medical research, social media, electronic health records, mobile apps, and websites, and extract information best-suited to an individual patient's needs.		
NSF – 1735095	Rocha (PI)	2017-2022
<i>NRT: Interdisciplinary Training in Complex Networks and Systems.</i> 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.		
NIH - R01LM011945-01	Rocha, Li, Shatkay (MPI)	2014-2019
BLR: Evidence-based Drug-Interaction Discovery: In-Vivo, In-Vitro and Clinical <i>Literature and Text mining to uncover distinct types of experimental evidence of drug interactions.</i> In this project we have been developing dictionaries, corpora and classification pipelines for extraction of evidence of DDI at different scales from the molecular, to the individual and population levels.		
Indiana Clinical Translational Sciences Institute (ICTSI).	Miller, Rocha (MPI)	2016-2018

Sudden Unexpected Death in Epilepsy: Identifying Risk Factors with Social Media Mining. Small pilot Study with co-PI Wendy Miller to use Facebook to identify risk factors in SUDEP. It allowed us to build a Facebook App so that families of deceased patients could donate their Facebook timelines.

Completed Research Support

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