+1 647-673-6248
ramanuja.simha@gmail.com
http://ramanujasimha.com/
ramanujasimha

Ramanuja Simha

Research Interests

Machine Learning, Probabilistic Modeling, Bayesian Networks, Deep Learning, Computational Biology, Data Mining, Graph Algorithms, Social Network Analysis

Education

2011–2016 **Doctor of Philosophy (Ph.D.), Computer Science**, *University of Delaware*, USA, *3.97/4.00*.
Advised by Dr. Hagit Shatkay.
Research in *machine learning* and *computational biology*. Dissertation on "*A Probabilistic Framework For Protein Multi-Location Prediction, And Its Applicability To Multi-Label Classification*."

2008–2011 **Master of Science (M.S.), Computer Science**, *University of South Florida*, USA, *4.0/4.0*.
Research in *graph algorithms* and *data mining*. Thesis on "*Mining Associations Using Directed Hypergraphs*."

2001–2005 **Bachelor of Engineering (B.E.), Information Science**, *National Institute of Engineering*, IN.
Thesis on "*High Performance Messaging Layers for Cluster Architecture*."

Publications (Google Scholar: *scholar.google.com/citations?user=4nOtToMAAAAJ*)

In Preparation/Submission

19. **Theory of Reality.** *Ramanuja Simha.* Manuscript submitted to arXiv, 2017.
18. **Improved Multi-label Classification Performance Using A Generative Mixture Model.** *Ramanuja Simha, and Hagit Shatkay.*
Manuscript for submission to Journal of Machine Learning Research (JMLR), 2017.
17. **A Comprehensive Database of Eukaryotic Multi-localized Proteins.** *Ramanuja Simha, Melody Lugo, Dean Arnold, Tongming Li, Sanjeev Patra, and Hagit Shatkay.*
Manuscript for submission to Nucleic Acids Research Database Issue, 2017.

Peer-Reviewed Journal Papers

16. **Protein (Multi-)Location Prediction: Utilizing Interdependencies via a Generative Model.** *Ramanuja Simha, Sebastian Briesemeister, Oliver Kohlbacher, and Hagit Shatkay.*
Bioinformatics 31(12), 2015. **Impact factor: 5.766.**
15. **Protein (Multi-)Location Prediction: Using Location Inter-Dependencies in a Probabilistic Framework.** *Ramanuja Simha and Hagit Shatkay.*
Algorithms for Molecular Biology 9(1), 2014. **Impact factor: 1.439.**
14. **Determining the Subcellular Location of New Proteins from Microscope Images Using Local Features.** *Luis Coelho, Joshua Kangas, Armaghan Naik, Elvira Osuna-Highley, Estelle Glory-Afshar, Margaret Fuhrman, Ramanuja Simha, Peter Berget, Jonathan Jarvik, and Robert Murphy.*
Bioinformatics 29(18), 2013. **Impact factor: 5.766.**
13. **Identifying High Betweenness Centrality Nodes in Large Social Networks.** *Nicolas Kourtellis, Tharaka Alahakoon, Ramanuja Simha, Adriana Iamnitchi, and Rahul Tripathi.*
Social Network Analysis and Mining 3(4), 2013.

Peer-Reviewed Conference / Workshop Papers, Podium Presentations, Posters, and Abstracts

12. **Improved Multi-Label Classification Using Label Inter-dependencies Via A Generative Mixture Model.** *Ramanuja Simha and Hagit Shatkay.*
European Conference on Artificial Intelligence (ECAI), August 2016. **Acceptance rate: ∼27%.**

11. **Protein (Multi-)Location Prediction: Utilizing Interdependencies via a Generative Model.** *Ramanuja Simha, Sebastian Briesemeister, Oliver Kohlbacher, and Hagit Shatkay.*
International Conference on Intelligent Systems for Molecular Biology (ISMB), July 2015. **Acceptance rate: ∼17%.** (Also appears as [18] above.)

10. **Protein (Multi-)Location Prediction: Using Bayesian Networks for Location Interdependencies, and a Mixture Model.** *Ramanuja Simha and Hagit Shatkay.*
International Conference and Exhibition of the Society for Laboratory Automation and Screening (SLAS), February 2015.

9. **Protein (Multi-)Location Prediction: Using Location Inter-Dependencies in a Probabilistic Framework.** *Ramanuja Simha and Hagit Shatkay.*
International Workshop on Algorithms for Bioinformatics (WABI), September 2013.

8. **Mining Associations Using Directed Hypergraphs.** *Ramanuja Simha, Rahul Tripathi, and Mayur Thakur.*
International Workshop on Graph Data Management: Techniques and Applications (GDM) at the International Conference on Data Engineering (ICDE), April 2012.

7. **Branded with a Scarlet C: Cheaters in a Gaming Social Network.** *Jeremy Blackburn, Ramanuja Simha, Nicolas Kourtellis, Xiang Zuo, Matei Ripeanu, John Skvoretz, and Adriana Iamnitchi.*
International World Wide Web Conference (WWW), April 2012. **Acceptance rate: ∼12%.**

6. **Cheaters in a Gaming Metanetwork.** *Jeremy Blackburn, Ramanuja Simha, Clayton Long, Xiang Zuo, John Skvoretz, and Adriana Iamnitchi.*
HPDC / SIGMETRICS Student Posters, June 2011. **Best Student Poster Award**.
SIGMETRICS Performance Evaluation Review 39(3), 2011.

5. **K-Path Centrality: A New Centrality Measure in Social Networks.** *Tharaka Alahakoon, Rahul Tripathi, Nicolas Kourtellis, Ramanuja Simha, and Adriana Iamnitchi.*
Workshop on Social Network Systems (SNS), April 2011.

Theses

4. **A Probabilistic Framework For Protein Multi-Location Prediction, And Its Applicability To Multi-Label Classification.** *Ramanuja Simha.*
PhD Dissertation, University of Delaware, 2016.

3. **Mining Associations Using Directed Hypergraphs.** *Ramanuja Simha.*
Masters Thesis, University of South Florida, 2011.

Other Papers

2. **Influence Maximization in Distribution Networks.** *Ramanuja Simha, Rahul Tripathi, and Balaji Padmanabhan.* Manuscript.

1. **Identifying Minimal Sources in Networks.** *Ramanuja Simha, Rahul Tripathi, and Balaji Padmanabhan.* Manuscript.

Awards and Honors

Aug 2016 European Association for Artificial Intelligence (EurAI, formerly European Coordinating Committee for Artificial Intelligence - ECCAI) Travel Award 2016

2013-14, 14-15 University of Delaware Graduate Fellowship

Feb 2015 Tony B. Academic Travel Award, Society for Laboratory Automation and Screening (SLAS) 2015

Jul 2012 Machine Learning Summer School 2012 Scholarship

Apr 2012 NSF IEEE International Conference on Data Engineering (ICDE) 2012 Scholarship

Jun 2011 High Performance Distributed Computing (HPDC) 2011 Student Travel Grant

Jun 2011 High Performance Distributed Computing (HPDC) / ACM Special Interest Group on Performance Evaluation (SIGMETRICS) 2011 Best Student Poster Award

Jul 2010 Petascale Programing Environments and Tools

Work Experience

2016–Present **University of Toronto and Univ. Health Network**, *Postdoctoral Research Fellow*, Toronto, ON.
Developing machine learning and deep learning methods to predict for dimensionality reduction. Developing statistical tools to derive insights from gene, protein expression data and clinical data.

2011–2016 **University of Delaware**, *Graduate Fellow / Research Assistant*, Newark, DE.

- Developed a probabilistic generative model for assigning possibly multiple locations to proteins. Devised an *Expectation Maximization* based algorithm to learn model parameters.
- Developed a prediction method that incorporates location inter-dependencies using Bayesian network classifiers. Learned structure using greedy hill climbing. Implemented Gibbs sampling and variable elimination for inference.

Summer 2015 **GE Global Research**, *R&D Intern*, Niskayuna, NY.
Employing network-based techniques, developed a software implementation to stratify cancer subtypes based on genomics and transcriptomics information.

Summer 2014 **LinkedIn**, *Applied Data Mining Intern*, Mountain View, CA.
Developed machine-learning techniques to rank start-up companies using probabilistic models, neural networks, and regression. Aggregated data from Crunchbase, AngelList, and LinkedIn HDFS database.

2014–Present **Kaggle**, *www.kaggle.com/users/16351/rsimha*.
Africa Soil Property Prediction Challenge: **Finished in the top ∼7% (87/1233)**; Developed a multiple variable prediction method based on a collection of support vector regression classifiers. *Allstate Purchase Prediction Challenge*: **Finished in the top ∼17% (271/1568)**; Constructed a multi-label classifier using random forests, support vector machines, and Bayesian networks.

Summer 2011 **Carnegie Mellon University**, *Visiting PhD Student*, Pittsburgh, PA.
Developed machine-learning approaches for assigning locations to proteins using image-based features.

Summer 2010 **National Center for Atmospheric Research**, *SIParCS Intern*, Boulder, CO.
Developed a data-mining model to predict performance of MPI-OpenMP software.

2008–2011 **University of South Florida**, *Research Assistant*, Tampa, FL.

- Developed MapReduce algorithms to compute network characteristics such as degree, diameter, triad census, clustering coefficient for a gaming network containing *12M nodes* and *88M edges*.
- Implemented a directed hypergraph model for time-series (e.g. *S&P 500 dataset*) analysis by utilizing attribute-level association rules.
- Utilizing a hash table, implemented an efficient streaming algorithm for computing PageRank.

2005–2008 **Tesco HSC**, *Senior Software Engineer*, Bengaluru, KA & Welwyn Garden City, HRT.
Developed and tested integration applications in the retail domain.

Other Professional Training

Summer 2012 **UC Santa Cruz**, *Machine Learning Summer School Student*, Santa Cruz, CA.
Learned approximate inference methods for graphical models, boosting algorithms, and machine learning techniques for information retrieval applications.

Teaching Experience

Spring 2011 **Analysis of Algorithms**, *Teaching Assistant*, University of South Florida, Tampa, FL.
Held office hours for undergraduate-level course on algorithm design, assisted students with data structures/programming, graded homework and exams.

Fall 2010, Fall 2009 **Intro. to Theory of Algorithms**, *Teaching Assistant*, University of South Florida, Tampa, FL.
Held office hours for graduate-level course on algo. design and complexity analysis, taught classes in instructor's absence, graded homework and exams.

Spring 2010 **Operating Systems**, *Teaching Assistant*, University of South Florida, Tampa, FL.
Held office hours for graduate-level course on design and implementation of operating systems, taught classes in instructor's absence, graded paper reviews, assisted in the design of homework & exams, and their grading.

Fall 2008 **Automata Theory**, *Teaching Assistant*, University of South Florida, Tampa, FL.
Held office hours for undergraduate-level course on formal languages, taught classes in instructor's absence, held weekly problem sessions, graded homework and exams.

Skills

Programming Python, C++, MapReduce, SQL, R, Apache Pig

Software Theano, Keras, numpy, scikit-learn, SWIG, scipy, matplotlib, NLTK, Weka, BLAST, mrjob, Amazon Elastic MapReduce, Amazon EC2, LaTeX

Machine Learning Multi-label Classification, Bayesian Networks, Deep Learning, Autoencoding, Expectation Maximization (EM), Nonnegative Matrix Factorization (NMF), Gibbs Sampling

Data Mining Graph Algorithms, Parallel Programming, PageRank, Association Mining, Social Network Analysis

OS UNIX, Linux, Microsoft Windows

Professional Activities

PC Member Mid-Atlantic Student Colloquium on Speech, Language and Learning 2012.

Reviewer Nature - Scientific Reports; Proteins: Structure, Function, and Bioinformatics; BMC Bioinformatics; Journal of Computational Biology; International Conference on Research in Computational Molecular Biology (RECOMB) 2016; Intl. Conference for High Performance Computing, N/w, Storage and Analysis (SC11) 2011; International Conference on Information Systems (ICIS) 2009

Member The International Society for Computational Biology (ISCB)