

Habil Zare

Email: zare@txstate.edu
Homepage: http://cs.txstate.edu/~h_z14/
Phone: (+1) 206 612-7830
Canadian citizen, and US green card holder

Department of Computer Science
Texas State University
Comal Building, 307B
San Marcos, TX 78666
USA

RESEARCH INTERESTS

Bioinformatics, Computational Biology, Cancer Research, Genetics, Epigenomics, Data Mining and Statistical Machine Learning, Computational Complexity, Graph Theory.

CURRENT POSITION

Assistant Professor in Department of Computer Science, Texas State University, USA, Sept. 2014-current.

Responsibilities: Research, teaching, and academic service.

Principal Investigator, established Oncinfo Lab, which is focused on research in bioinformatics and computational biology, Sept. 2014.

EDUCATION

Ph.D. in Computer Science, University of British Columbia, Canada, January 2010¹-December 2011.

Dissertation: Automatic Analysis of Flow Cytometry Data and its Application in Lymphoma Diagnosis.
Using sophisticated machine learning techniques, developed novel bioinformatics tools to improve data analysis, and discovered a criterion useful for lymphoma diagnosis.

Supervisors: Arvind Gupta and Ryan Brinkman.

Committee members: Andrew Weng, Gabor Tardosh, and Valentine Kabanets.

M.Sc. in Computer Science, Sharif University of Technology, Iran, 2002-2004.

Dissertation: Markov Chain (Random Walk) on Graphs.
Performed a comprehensive study on different definitions for Laplacian of a graph and proposed a unifying viewpoint to spectral graph theory from various aspects of mathematics, such as algebra, geometry, stochastic processes, and mathematical analysis. Application to randomized algorithms were discussed.

B.Sc. in Mathematics, Sharif University of Technology, Iran, 1998-2002.

POSTDOCTORAL

Senior Fellow in Noble Laboratory, Department of Genome Sciences, The University of Washington, USA, January 2012-2014.

¹transferred from Simon Fraser University with my senior supervisor.

PREVIOUS RESEARCH

Department of Genome Sciences, School of Medicine, The University of Washington, January 2012-June 2014.

Developing and applying advanced machine learning techniques to analyze genomic and epigenomic data in Noble Laboratory:

- Studying the relation between histone modification and gene expression; Modeled the epigenomic, sequential data by a dynamic Bayesian network to predict de novo genes and their tissue-specific expression. Also, designed a deep learner to improve state of the art methods for analyzing biological sequence data.
- Discovering clonal structure of breast cancer using high throughput sequencing; Designed and implemented Clomial, an algorithm that models next generation sequencing data from several samples of a single tumor by binomial distributions. The model can be used to identify the clonal structure of tumor, such as number of clones, their genotype, and the proportion of each clone in the tumor.

British Columbia Cancer Agency, Terry Fox Laboratory, April 2009-December 2011.

Developing tools and techniques for analyzing flow cytometry data in Brinkman Laboratory:

- Designed and implemented a novel data reduction scheme to reduce size of flow cytometry data without significant loss in biological information. My approach resulted in enhancing spectral clustering technique such that it could be efficiently applied on flow cytometry data for the first time.
- Designed and implemented a novel feature scoring scheme based on LASSO, useful for feature selection.
- Applied supervised machine learning techniques to build a classifier for differential diagnosis between subtypes of lymphoma in 860 patients.
- Applied my methodology to data of 110 patients diagnosed in BC Cancer Agency and discovered three novel flow cytometry features that discriminate between Mantel Cell Lymphoma (MCL) and Small Lymphocytic Lymphoma (SLL) with high accuracy. The novel features were shown to be clinically useful.

PUBLICATIONS

Published Journal Articles

- **Zare H.**, Shooshtari, P., Gupta, A., Brinkman R.R., Data reduction for spectral clustering to analyze high throughput flow cytometry data. *BMC Bioinformatics* 2010, **11**:403-413. (“**Highly accessed**” according to the journal website.) Impact factor: 3.0, Citations: 43.
Contribution: Spectral clustering cannot be applied directly to large size data because it has time complexity $O(n^3)$. I designed and implemented the novel "faithful sampling" algorithm (95%), the methods for computing similarity matrix (95%), estimating the number of cluster (80%), and post-processing steps (70%). I also performed the experiments (90%) and wrote the manuscript(70%).
- **Zare H.**, Bashashati A., Kridel R., Aghaeepour N., Haffari G., Connors J., Gupta A., Gascoyne R., Brinkman R.R., Weng A., Automated analysis of multidimensional flow cytometry data improves diagnostic accuracy between mantle cell lymphoma and small lymphocytic lymphoma, *American Journal of Clinical Pathology* 2012 **137**:75-85. Impact factor: 2.9, Citations: 13.
Contribution: The novel features improve the accuracy of diagnosis from 70% to 98%. I designed the pipeline (80%), implemented the algorithms (95%), analyzed data (70%), and discovered novel phenotypes (95%).
- Bashashati A., Nathalie A. J., Hadjkhodabakhshi A., Whiteside M., **Zare H.**, Scott D.W., Lo K., Gottardo R., Brinkman F., Connors J., Slack G., Randy D.G., Weng A., Brinkman R.R., B-cells with high side scatter parameter by flow cytometry correlate with inferior survival in diffuse large B cell lymphoma, *American Journal of Clinical Pathology* 2012, **137**:805-814. Impact factor: 2.9, Citations: 3.
Contribution: I contributed in data analysis (5%).

- Aghaeepour N., Chattopadhyay P.K., Ganesan A., O'Neill K., **Zare H.**, Jalali A., Hoos H.H., Roederer M., Brinkman R.R., Early immunologic correlates of HIV protection can be identified from computational analysis of complex multivariate T-cell flow cytometry assays, *Bioinformatics* 2012, **28**:1009-1016. Impact factor: 5.3, Citations: 26.
Contribution: I contributed in data analysis (15%), preparing figures, presenting results and writing the manuscripts (30%), and biological interpretation (5%).
- **Zare H.**, Haffari G., Weng A., Gascoyne R., Gupta A., Brinkman R.R., Scoring relevancy of features based on combinatorial analysis of Lasso with application to lymphoma diagnosis, *BMC Genomics* 2013, **14**:S14. Impact factor: 4.4, Citations: 4.
Contribution: designed a novel feature scoring scheme useful for feature selection (90%), implemented the algorithm (95%), developed a mathematical framework to explain the performance of the method (70%), theoretically proved its efficiency (80%), performed experiments (95%) and wrote the manuscript (70%).
- Aghaeepour N., et al, **Zare H.**,... Critical assessment of automated flow cytometry data analysis techniques, *Nature Methods* 2013, **10**:228-238. Impact factor: 23.6, Citations: 54.
Contribution: This is a comparison between different computational approaches for analyzing flow cytometry data. I contributed in data analysis; running SamSPECTRAL on data (15%), preparing figures (5%), presenting results (5%) and writing the manuscripts (5%).
- **Zare H.**, Wang J., Hu A., Weber K., Smith J., Nickerson D., Witten D., Blau A.C., and Noble W.S., Inferring clonal composition from multiple sections of a breast cancer, *PLOS Computational Biology* 2014, **10**(7). Impact factor: 4.9.
Contribution: I proposed a scheme to model Next Gen Sequencing data of a tumor with binomial distributions, and developed an algorithm to infer the parameters using EM (80%), derived the formulas (90%), ran experiments (95%), prepared figures (60%), and wrote the manuscripts (40%).

Journal Articles In Preparation

- **Zare, H.**, Hoffman M., Bilmes J., Weng Z., Noble W.S., A deep learning strategy to link transcription to combinatorial patterns of histone modifications in the human genome.
Contribution: I developed the deep Dynamic Bayesian network (50%), implemented the model using gmtk (95%), ran experiments (95%), prepared figures (90%), and wrote the manuscripts (50%).

Invited Talks

- “Computational approaches for gene network analysis with applications in studying AML and other cancers”, The University of Texas Health Science Center at San Antonio, TX, USA, March 2015.
- “SamSPECTRAL: Efficient spectral clustering on flow cytometry data”, FlowCAP summit, American National Institutes of Health Campus (NIH), M.D., USA, September 2010.
- “Automatic analysis of flow cytometry data”, British Columbia CLL Day, BC, Canada, February 2010.

Presentations

- “Inferring clonal composition from multiple sections of a breast cancer”, Seattle Genetic Instability and Cancer Symposium, Fred Hutchinson Cancer Research Center, Seattle, April 2014.
- “How to infer clonal composition of a tumor from multiple tissue samples?” Post-Doctoral Presentation, University of Washington, April 2014.
- “Scoring relevancy of features based on combinatorial analysis of Lasso with application to lymphoma” Asia Pacific Bioinformatics Conference, Vancouver, January 2013.
- “Automated analysis of multidimensional flow cytometry data improves diagnostic accuracy between mantle cell lymphoma and small lymphocytic lymphoma”, ImmunoVancouver meeting, University of British Columbia, Vancouver, June 2011.

- “From graph theory to cancer diagnosis”, Bioinformatics Reading Group, University of British Columbia, Vancouver, June 2011.
- “Discrimination between Mantle Cell Lymphoma and Small Lymphocytic Lymphoma can be improved by automated analysis of multidimensional flow cytometry data”, Work in Progress (WIP) seminars, BC Cancer Research Center, Vancouver, February 2011.
- “Is it possible to apply spectral clustering method to analyze high throughput biological data?”, Honors Seminar, Computer Science Dept., University of British Columbia, October 2010.
- “Lymphoma Diagnosis Based on Automated Analysis of Flow Cytometry Data”, Cancer Bioinformatics Workshop, Cambridge Cancer Research, September 2010.
- “Automatic analysis of flow cytometry data for lymphoma diagnosis”, BC Cancer Research Center, Trainee Seminars, May 2010.
- “Automated identification of cell populations in flow cytometry data”, Work in Progress (WIP) seminars, BC Cancer Research Center, Vancouver, November 2009.

PATENTS

- Blau A. C., Noble W., Wang J., Witten D., **Zare H.**, “Inferring clonal composition from multiple sections of a cancer”, US Provisional Patent Application 61/955,118, March 18, 2014.

EDITORIAL BOARD

- Journal of Biological Research & Development.

PEER REVIEWING

- National Science Foundation.
- Bioinformatics.
- Intelligent Systems for Molecular Biology (ISMB)
- Neural Information Processing Systems (NIPS).
- PLOS ONE.
- Cytometry: Part A.
- OMICS: A Journal of Integrative Biology.
- Pacific Symposium on Biocomputing.
- Uncertainty in Artificial Intelligence (UAI) conference.
- Cancer Informatics.

SCIENTIFIC SOFTWARE

- SamSPECTRAL (approved by Bioconductor)
Developed and maintain this R package, an implementation of my enhancement to spectral clustering method that is capable of clustering flow cytometry data with 100K events in minutes.
- FeaLect (approved by CRAN)
Developed and maintain this R package that implements my novel scoring scheme useful for feature selection.

- Clomial (approved by Bioconductor)

Developed and maintain this R package which fits binomial distributions to counts obtained from Next Gen Sequencing data of multiple samples of the same tumor. The trained parameters can be interpreted to infer the clonal structure of the tumor.

Programming Skills and Bioinformatics Tools

Expert: R Statistical Programming Language, Maple, Mathematica, Python, Prolog, \LaTeX , BWA, SAMtools, etc.

Familiar: Matlab, C, Java, Perl, HTML, GTAK, UCSS Genome Browser, IGV, bedtools, SVN, etc.

TEACHING

Texas State University, Department of Computer Science

Instructor, Fall 2014-current.

Formal Languages (two times), Foundations of Computer Science II.

Simon Fraser University, School of Computer Science

Sessional Instructor, Discrete Mathematics 1, Spring 2008.

Teaching Assistant for 11 courses (Fall 2004-Spring 2009):

Introduction to Computing Science and Programming (three times), Computability and Complexity, Computational Complexity, Data Structures and Algorithms, Discrete Mathematics, Cryptography, Software Engineering, Introduction to Web, Social Implications of a Computerized Society.

Sharif University of Technology, Mathematical Science Department

Teaching Assistant for 5 courses (Fall 2001-Spring 2004):

Mathematical Analysis 1, 2, and 3, Elements of Mathematics, Calculus 1.

Mentoring

- Mentoring Amir Foroushani, postdoc in Oncinfo Lab, 2014-current.
- Supervising Rupesh Agrahari, graduate student in Oncinfo Lab, 2014-current.
- Mentoring a rotating student in the Noble Lab at University of Washington, leading to a publication in PLOS Computational Biology, 2013-2014.

PROFESSIONAL SERVICE

- Member of Graduate Committee, Computer Science Department, Texas State University, 2014-current.
- Coordinator of Sarenakh Mathematical Seminar Series, Sharif University of Technology, 2002-2003.
- Secretary of Student Mathematical Association at Sharif University of Technology, 2002-2003.

HONORS and AWARDS

- **1st Rank** in Iranian National Entrance Exam for Master Program in Mathematics among 3000 students, 2002.
- Member of Graduate Faculty, Texas State University, 2014-current.
- National Science Foundation, Directorate for Computer and Information Science and Engineering Travel Award, \$800, 2015.
- International Society for Advances of Cytometry Travel Award, \$650, 2012.
- International Society for Advances of Cytometry Travel Award, \$650, 2011.
- University of British Columbia Faculty of Science Graduate Award, \$2,800, 2011.
- University of British Columbia Travel Award, \$500, 2011.
- University of British Columbia Faculty of Science Graduate Award, \$2,800, 2010.
- University of British Columbia Faculty of Science Graduate Award, \$2,800, 2009.
- Simon Fraser University Graduate Fellowship Award, \$3,000, summer 2008.
- Simon Fraser University Graduate Fellowship Award, \$3,000, summer 2007.
- Simon Fraser University Graduate Fellowship Award, \$6,000, summer 2006.
- Institute for Research in Fundamental Sciences Fellowship Award, \$300, 2000-2002.

REFERENCES

Dr. William Noble

(my postdoc supervisor)
Professor of Genome Sciences
The University of Washington
Seattle, WA, USA.
Phone: (+1) 206.221.4973
Email: noble@gs.washington.edu

Dr. Ryan Brinkman

Associate Professor
Medical Genetics
The University of British Columbia
Vancouver, BC, Canada.
Phone: (+1) 604.675.8132
Email: rbrinkman@bccrc.ca

Dr. Andrew P. Weng, MD, PhD

Senior Scientist
Pathology and Laboratory Medicine
British Columbia Cancer Agency
Vancouver, BC, Canada.
Phone: (+1) 604.675.8136
Email: aweng@bccrc.ca

Dr. Anthony Blau, MD

Professor of Medicine/Hematology
The University of Washington
850 Republican Street, Room N535
Seattle, WA, USA.
Phone: (+1) 206.685.6873
Email: tblau@u.washington.edu

Dr. Jeff Bilmes

Professor
Department of Electrical Engineering
The University of Washington
Seattle, WA, USA.
Phone: (+1) 206.221.5236
Email: bilmes@ee.washington.edu

Dr. Amir Daneshgar

(my MSc supervisor)
Professor of Mathematical Sciences
Sharif University of Technology
Tehran, Iran.
Phone: (+98) 21.6616.5610
Email: daneshgar@sharif.ac.ir

Dr. Arvind Gupta

(my PhD supervisor)
Professor of Computing Science
The University of British Columbia
Vancouver, BC, Canada
Phone: (+1) 604.822.9189
Email: arvind@mitacs.ca

Dr. Zhiping Weng

Professor and Director
Program in Bioinformatics &
Interactive Biology
The University of Massachusetts
368 Plantation Street, Worcester, MA.
Phone: (+1) 508.856.8866
Email: Zhiping.Weng@umassmed.edu

Dr. Gabor Tardos

Professor
School of Computer Science
Simon Fraser University
Burnaby, BC, Canada.
Phone: (+1) 778.7826912
Email: tardos@cs.sfu.ca