

CHIRAG LAKHANI

217 Harvard St., Apt 3L ♦ Cambridge, MA 02139

(919) · 749 · 5810 ♦ chirag.lakhani@gmail.com

EDUCATION

North Carolina State University

December 2010

Ph.D., Mathematics

Thesis: The GIT Compactification of Quintic Threefolds

Advisor: Amassa Fauntleroy

Designated emphasis on new algebraic geometry approaches to topological string theory; close collaboration with the algebraic geometry group at NCSU and the string theory group at Duke. Particular interest in toric varieties, mirror symmetry, and geometric invariant theory.

North Carolina State University

December 2003

B.S. Mathematics

Minor: Physics

Magna Cum Laude

EXPERIENCE

Harvard Medical School - Department of Biomedical Informatics

March 2015 - Present

Postdoctoral Research Fellow

Boston, MA

- Created “Exposome Data Warehouse,” a database of aggregated publicly available environmental data (socioeconomic factors, pollution, and climate) that links to electronic health records (EHR).
- Nearing completion on a project in where we repurpose insurance claims data to estimate the contribution of genetics and the environment to disease. We utilize family structure in the insurance claims data to understand the role of genetics and are utilizing the “Exposome Data Warehouse” to determine the role of the environment for hundreds of traits.

HelloWallet, Inc

April 2014 - March 2015

Data Scientist

Washington, DC

- Led initiative to prototype and productionize a machine learning based transaction classification system in the HelloWallet app. Work ranged from data cleaning, prototyping algorithms in sci-kit learn to developing a technical approach for productionizing app in Java based app stack..
- Led data engineering initiative to migrate analytics database from MySQL environment to a Hadoop/Spark based environment.

Zaloni, Inc

February 2013 - April 2014

Data Scientist

Research Triangle Park, NC

- Led new data science initiative at Zaloni, implementing large-scale machine learning solutions to meet clients needs (such as fraud detection or scalable clustering algorithms for customer segmentation).
- Liaison between clients and development teams in the US and India to create data science products of value.
- Developed “Introduction to Data Science” training course which focused on applying Mahout to data science problems in the real world; taught at several Fortune 500 companies.

Duke University and SAMSI

August 2012 - February 2013

Independent Researcher

Research Triangle Park, NC

- Focused on developing scalable solutions to traditional nonlinear dimensionality reduction algorithms. Collaborated with Duke Prof. Mauro Maggioni and his group on leveraging geometric multi-resolution analysis to compute manifold learning algorithms efficiently.
- At SAMSI primarily focused on working group led by Prof. Ilse Ipsen (NC State), Research Scientist Michael Mahoney (Stanford), and Prof. Petros Drineas (RPI), on applying randomized linear algebra techniques to Kernel PCA algorithms for more efficient, scalable computing.

North Carolina State University

Postdoctoral Research Associate

August 2011 - December 2011

Raleigh, NC

- Advisor: Hamid Krim.
- Collaborated with Prof. Hamid Krim and his group on applying computational topology to problems in data mining. Primarily interested in the use of persistent cohomology for nonlinear dimensionality reduction. Also considered the use of Morse theory for detecting community structures in complex networks.

High Point University

Instructor of Mathematics

August 2010 - May 2011

High Point, NC

- Taught various math courses in the department such as calculus I, calculus II, and mathematics for elementary school teachers.

IN PREPARATION

C. Lakhani, S. Bhavé, B. Tierney, R. Acosta, A. Deonarine, I. Kohane, and C. Patel, “[Exposome DW: A Unified Data Warehouse of Geotemporal Demographic, Economic, and Environmental Information to Enhance Clinical Research](#)”, *In Preparation* (2017)

PUBLICATIONS

C. Lakhani, B. Tierney, A. Manrai, I. Kohane, J. Yiang, P.M. Visscher, and C. Patel, “[Repurposing Large Health Insurance Claims Data to Estimate Genetic and Environmental Contributions in 561 Diseases](#)”, *Submitted for Review* (2017)

C. Lakhani, “[The GIT Compactification of Quintic Threefolds](#)”, *arXiv* **arXiv:1010.3803** (2010)

DATA PRODUCTS

C. Lakhani, B. Tierney, and C. Patel, “[Repurposing Large Health Insurance Claims Data to Estimate Genetic and Environmental Contributions in 561 Diseases](#)”, <https://chiragjp.shinyapps.io/twinsshiny/>

TRAINING AT HARVARD

Harvard Courses:

CS 281: Advanced Machine Learning	Fall 2015
EPI 249: Molecular Biology for Epidemiologists	Fall 2016
BST 227: Introduction to Statistical Genetics	Fall 2016
EPI 511: Advanced Population and Medical Genetics	Spring 2017

Summer Institute in Statistical Genetics:

Association Mapping: GWAS and Sequencing Data	July 2017
Mixed Models in Quantitative Genetics	July 2017
Advanced Quantitative Genetics	July 2017
Statistical and Quantitative Genetics of Disease	July 2017

ACADEMIC TALKS

“Building a Search Engine to Find Environmental and Phenotypic Factors Associated with Health and Disease” NSF Big Data Spokes Meeting - Washington, DC, March 2017.

“Building an Exposome API” Environmental Statistics Seminar - Harvard School Of Public Health, February 2016.

“Systematic and Large-Scale Investigation of Twin and Sibling Concordance of 1723 Traits in a Nationally Representative Health Claims Cohort” American Society of Human Genetics Platform Talk - Baltimore, MD, October 2015.

“Approximation of Kernel Matrices” SAMSI Working Group - SAMSI, November 2012.

“Tensor Decompositions in Signal Processing” Tensor Seminar - North Carolina State University, October 2011.

“Topological Data Analysis and Zig-Zag Persistence” Geometric Methods in Signal Processing - North Carolina State University, August 2011.

“Secondary Fan for Toric Varieties” Toric Varieties Working Seminar - Duke University, November 2009.

“Schemes and Varieties” Visiting Student Seminar - Tata Institute of Fundamental Research, July 2005.

STUDENT MENTORING

Project Mentor, Summer 2016. Harvard Summer Institute of Biomedical Informatics. Supervised two undergraduate students to develop collect and ingest NOAA and EPA data into the “Exposome Data Warehouse” .

COMPUTATIONAL SKILLS

Programming Languages	C/C++, Python, Java
Hadoop Tools	Map-Reduce Programming, Hive, Pig, HBase, Spark
Databases	Hive, MySQL, PostgreSQL, HAWQ, Impala, SparkSQL
Large Scale Machine Learning	Mahout, Graphlab, MADLib, MLlib
Machine Learning Libraries	Various R libraries, sci-kit learn, MATLAB, PyMC, STAN
Web Technologies	HTML, CSS, PHP, Javascript, JQuery

ONLINE PROFILES

Github	https://github.com/cmlakhan
Linkedin	https://www.linkedin.com/in/chiraglakhani
Website	http://cmlakhan.github.io/

HONORS AND AWARDS

Cloud Computing Credits Award (\$100,000) (Microsoft Azure)	May 2017
AMS Travel Grant - MRC (Computational and Applied Toplogy)	June 2011
NSF Travel Grant - MEGA 2011 Conference, Stockholm, Sweden	May 2011
NSF Travel Grant - Aspects of Moduli Conference, Pisa, Italy	June 2008
Department Undergraduate Aware for Research	December 2003
PAMS College Undergraduate Award for Service	December 2003