BIJAN MAZAHERI

github.com/honeybijan bijanmazaheri.com bmazaher@broadinstitue.org
Cambridge, MA (781)-985-0881

I am a computer scientist and mathematician interested in the synthesis and transportability of knowledge from and between multiple data sources. My work spans causality, mixture models, decision fusion, and distribution shift. I am currently at the Broad working to develop new tools to solve unique data science challenges in the biological and health sciences.

EDUCATION

California Institute of Technology - Pasadena, CA *Oct 2017 - Aug 2023*
Ph.D. Candidate
Department of Computing and Mathematical Sciences, GPA: 3.9/4.0
Awarded NSF Graduate Research Fellowship and Amazon AI4Science Research Fellowship

Cambridge University (Emmanuel College) - Cambridge, UK *Oct 2016 - Jun 2017*
Mathematics Part 1B
Supported by a Herchel Smith Fellowship
Additional classes in Computer Science and Mathematics Part II

Williams College - Williamstown, MA *Sep 2012 - Jun 2016*
Bachelor of Arts
Physics and Computer Science, GPA: 3.92/4.00
Highest Honors (Physics), Phi Beta Kappa, Sigma Xi, Magna Cum Laude
Thesis Title: RNA Macrostates and Macrokinetics

WORK EXPERIENCE

Eric and Wendy Schmidt Center at the Broad Institute - Cambridge, MA *Oct 2023 - Now*
Postdoctoral Associate
Working to apply insights from my Ph.D. to biological and hospital data.

Amazon Research Causality Lab - Tübingen, Germany *Oct 2022 - Feb 2023*
Applied Scientist Intern (L5)
Worked with Dr. Michaela Hardt, Dr. Atalanti Mastakouri, and Dr. Dominik Janzing
Lead-authored a paper that was accepted to UAI 2023 and gained experience with Amazon's code review process.

BioDiscovery - El Segundo, CA *Jun 2017 - Sep 2017*
Intern
I developed methods for clustering cancers based on their genomes and implemented it within the company stack. My work has now been integrated into BioDiscovery's software and presented at a conference.

IBM T.J. Watson Research Center - Yorktown Heights, NY *Jun 2016 - Sep 2016*
Intern
Worked with Dr. Victor Kravets (mentor) and Dr. Andrew Sullivan (manager).

Projects included non-greedy and map-reduce algorithms for factoring sum of products representations. The goal of this project was to find more efficient mappings of circuits onto 2-dimensional chips.

TEACHING EXPERIENCE

Markov Chain Monte Carlo *Spring 2022*
Head TA for new class on MCMC in theoretical computer science. Developed solutions and grading rubrics for problem sets.

Physics and Mathematics *Sep 2013-Jun 2016*
TAed for undergraduate classes in Electricity and Magnetism, Classical Mechanics, Mathematical Methods for Scientists, Premed Physics, Discrete Mathematics.

AWARDS AND GRANTS

National Science Foundation Graduate Research Fellowship *Awarded Spring 2019*
Awarded in 2019 for a proposal to research confounding influence in causal networks.
Amazon AI4Science Research Fellowship *Awarded Spring 2022*
Funding for research with the potential to aid scientific discovery. **Herchel Smith Fellowship** *Awarded Spring 2016*
Funding for 1-3 years of study at Cambridge University.

PUBLICATIONS

Bijan Mazaheri, Atalanti Mastakouri, Dominik Janzing, and Michaela Hardt. Causal Information Splitting: Engineering Proxy Features for Robustness to Distribution Shifts. In *The 39th Conference on Uncertainty in Artificial Intelligence*, 2023.

Spencer Gordon, ***Bijan Mazaheri**, Yuval Rabani, and Leonard J Schulman. Causal Inference Despite Limited Global Confounding via Mixture Models. In *2nd Conference on Causal Learning and Reasoning*, 2023.

Siddharth Jain, **Bijan Mazaheri**, Netanel Raviv, and Jehoshua Bruck. Glioblastoma signature in the DNA of blood-derived cells. *PLOS ONE* 16(9): e0256831. 2021.

Bijan Mazaheri, Siddharth Jain, and Jehoshua Bruck. Expert Graphs: Synthesizing New Expertise via Collaboration. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 2447–2452, 2021.

Spencer Gordon, ***Bijan Mazaheri**, Yuval Rabani, and Leonard Schulman. Source Identification for Mixtures of Product Distributions. In *Conference on Learning Theory*, pages 2193–2216. PMLR, 2021.

Bijan Mazaheri, Siddharth Jain, and Jehoshua Bruck. Robust Correction of Sampling Bias using Cumulative Distribution Functions. *Advances in Neural Information Processing Systems*, volume 33, pages 3546–3556. Curran Associates, Inc., 2020.

* = Authorship order is alphabetical.

PREPRINTS

Bijan Mazaheri, Spencer Gordon, Yuval Rabani, and Leonard J Schulman. Causal Discovery under Latent Class Confounding. *arXiv:2311.07454*, 2023.

Bijan Mazaheri, Siddharth Jain, Matthew Cook, and Jehoshua Bruck. Distribution Re-weighting and Voting Paradoxes. *arXiv:2311.06840*, 2023.

Spencer Gordon, Eric Jahn, ***Bijan Mazaheri**, Yuval Rabani, and Leonard J Schulman. Identification of Mixtures of Discrete Product Distributions in Near-Optimal Sample and Time Complexity. *arXiv:2309.13993*, 2023.

Spencer Gordon, ***Bijan Mazaheri**, Leonard J Schulman, and Yuval Rabani. The sparse Hausdorff moment problem, with application to topic models. *arXiv:2007.08101*, 2020.

Siddharth Jain, **Bijan Mazaheri**, Netanel Raviv, and Jehoshua Bruck. Cancer Classification from Healthy DNA using Machine Learning. *BioRxiv*, page 517839, 2019.

Siddharth Jain, **Bijan Mazaheri**, Netanel Raviv, and Jehoshua Bruck. Short Tandem Repeats Information in TCGA is Statistically Biased by Amplification. *BioRxiv*, page 518878, 2019.

* = Authorship order is alphabetical.

PATENTS

Siddharth Jain, **Bijan Mazaheri**, Netanel Raviv, and Jehoshua Bruck. Mutation profile and related labeled genomic components, methods and systems. 2019.

WORKSHOPS

Simon's Institute for Theory of Computing: Causality *Spring 2022*
4 week workshop on Causal inference methods.

INVITED TALKS

Simon's Institute for Theory of Computing *May 2023*
Title: "Causal Discovery under Limited Global Confounding"

PROJECTS

LACCTiC *Sep 2021 - present*
I maintain a website for collegiate cross country with 10,000 regular users based on an original algorithm for ranking performances on varying terrain. The backend runs on Python and Django and the frontend uses React, and the database is hosted on AWS. I have helped advise 16 student projects using this data.

MENTORSHIP

Caltech Cross Country Team *Sep 2018 - Sep 2022*
Assistant Coach
Mentoring and supporting undergraduate students at Caltech.
Data Science Projects *Sep 2019 - present*
I have supported 16 projects with undergraduate students using data on my website, and have advised some of these students in applying to graduate school.

SERVICE

Reviewer for AIStats, CLeaR.
Sports statistics outreach for Cross Country and correspondent at D3 Glory Days.
Caltech Community Associate - volunteer position for building community in Caltech housing.