Michael Bradshaw

✉ michaelscottbradshaw@gmail.com · in msbradshaw · ⚹ MSBradshaw · ▢ msbradshaw.github.io

Education

University of Colorado Boulder
Boulder, CO
PhD Computer Science
2019 - 2024
- **Advisor**: Ryan Layer
- **Dissertation Title**: Software and algorithms to improve outcomes for patients with rare-and-undiagnosed genetic-disease

Brigham Young University
Provo, UT
BS Bioinformatics
2019
- **Minors**: Computer Science & Molecular Biology

Experience

University of Colorado Boulder
Boulder, CO
PhD Student
2020-2024
- Demonstrated knowledge-graph embeddings link prediction models are biased against certain classes of genes and diseases. Proposed a new theory on the threefold origin and cyclical nature of these biases. Showed biases can be reduced by training on carefully selected and curated datasets.
- Using an ensemble of network clustering algorithms and XGBoost classifier, created a tool to predict groups of genes and phenotypes in a knowledge graph with latent connections. Applied this tool to undiagnosed disease patients and generated novel hypotheses for 49 cases. For this project, I integrated data from the human phenotype ontology (HPO), STRING protein database, OMIM, Orphanet, Comparative Toxicogenomics Database, Mouse Genome Informatics.
- Collaborated closely with clinicians and wet lab scientists to create a copy number visualization and curation tool that met their needs. The tool, SeeNV, uses a snakemake pipeline and a variety of cli tools to generate and visualize statistics about CNVs from whole exome data. Created statics to control for the many technical biases of whole exome data, variation with library preparation, individuals, across populations and between batches. SeeNV is has been used in Children's Hospital of Colorado (CHCO) clinical workflow since 2022.
- Developed a high throughput Nextflow pipeline for read alignment, variant calling, and quality control of whole exome sequencing data used by the Precision Diagnostics team at CHCO.

Department Lead Teaching Assistant
2020-2024
- Managed the department's pool of ∼90 teaching assistants
- Organized and led monthly in-service training, and taught a yearly six-week-long workshop on teaching skills

Teaching Assistant
2019-2020
- Taught a weekly 1.5-hour recitation to a group of 50 students, designed homework assignments, projects, and exams, held office hours, graded, all centered on teaching C++
- Receive the Outstanding TA Award

Max Plank Institute for Chemical Ecology
Jena, Germany
Intern Bioinformatician
2018 - 2019
- Performed quantitative trait loci (QTL) mapping to understand the genetic control of microbiome recruitment of 600 *Nicotiana attenuata* root samples

Publications & Graphics
Provo, UT
Associate Lead Wed Developer
11/2016 – 12/2018
- Worked with my team of 7 web developers to build and maintain six websites using a tech stack of PHP, JavaScript, jQuery, VUE, Docker and Github
- Revamped our team's git processes for continuous testing and integration of new features

Awards

Best Poster   **2022**
- International Society of Molecular Biology (ISMB) Conference, Bio-ontologies

Lead Graduate Student Fellowship   **2021-2022**
- Center for Teaching and Learning, CU Boulder
- Awarded to one PhD student from each department to receive discipline-specific teacher training

Department Service Award   **2021**
- Computer Science Department, CU Boulder

Lead Graduate Student Fellowship   **2020-2021**
- Center for Teaching and Learning, CU Boulder
- Awarded to one PhD student from each department to receive discipline-specific teacher training

Mentor's Choice Award   **2020**
- Copenhagen Bioinformatics Hackathon
- Of 20+ competing groups, the organizers and mentors of the competition chose my project on identifying fake research articles made by paper mills as the best

Most Voted Project Award   **2020**
- Copenhagen Bioinformatics Hackathon
- Of 20+ competing groups, participants of the hackathon voted my project the best

Outstanding Teaching Assistant Award   **2020**
- Computer Science Department, University of Colorado Boulder
- Given to 2 out of 90+ TA in the Department based on the student reviews

Publications

1. **Bradshaw M.S.**, Layer M. Understanding and addressing study, topological and model biases in biological knowledge-graph embedding link prediction tasks. (In progress).

2. **Bradshaw M.S.**, Firman T., Gaskell A., And Layer M. SeeNV: a tool for visualizing and assessing copy number variations in whole exome data. (In progress).

3. **Bradshaw M.S.**, Gibbs C.P., Martin S., Firman T., Gaskell A., Fosdick B.K., and Layer R.M. Hypothesis generation for rare and undiagnosed diseases through clustering and classifying time-versioned biological ontologies (Under review / preprint on biorxiv).

4. **Bradshaw, M. S.** & Payne, S. H. Detecting fabrication in large-scale molecular omics data. *PLoS One* 16, e0260395 (2021).

5. Smallegan M.J., Shehata S., Spradlin S.F., Swearingen A., Wheeler G., Das A., Corbet G., Nebenfuehr B., Ahrens D, Tauber D., Lennon S., Choi K., Huynh T., Wieser T., Schneider K., **Bradshaw M.**, Basken J., Lai M., Read T., and Hynes-Grace M., and Timmons D., Demasi J., Rinn J.L. Genome-wide binding analysis of 195 DNA binding proteins reveals "reservoir" promoters and human-specific SVA-repeat family regulation. *PLoS One* 16, e0237055 (2021).

6. Layer, R., Fosdick, B., **Bradshaw, M.**, Larremore, D. & Doherty, P. Case study: Using Facebook data to monitor adherence to stay-at-home orders in Colorado and Utah. *medRxiv* (2020).

7. **Bradshaw, M.** et al. Characterizing the ribosomal tandem repeat and its utility as a DNA barcode in lichen-forming fungi. *BMC Evol. Biol.* 20, 2 (2020).

8. Sumsion, G. R. and **Bradshaw, M.S.** et al. Diverse approaches to predicting drug-induced liver injury using gene-expression profiles. *Biol. Direct* 15, 1 (2020)

9. Sumsion, G. R., **Bradshaw, M. S.**, Hill, K. T., Pinto, L. D. G. & Piccolo, S. R. Remote sensing tree classification with a multilayer perception. PeerJ (2019).

Teaching

Guest Lecturer   October 2022 & October 2023
- Taught a week-long series of workshop-style lessons to a class of 20-30 graduate students in a software engineering for scientists course. Covered core techniques for data science like loading, formatting, cleaning, analyzing, and plotting data.

Lead Graduate Student Fellowship Workshop   2021 & 2022
- Organized and ran workshops on "Building Community in the Classroom" and "Benefits of Self-Reflection in the Classroom" for groups of educators.

Teaching Assistant - Introduction to Computer Science   2019-2020
- Planned and taught weekly 1.5-hour recitation to sections of 50 undergraduates
- Held office hours to help guide students to answers to their questions for 17 hours a week
- Substitute taught a week of lectures for a course of 200+ students

Personal Projects

Pulse - bioRxiv word trends   April 2020 - 2022
- Created a web-app for plotting the frequency of search terms in scientific literature overtime. The Github repository and front end can be found here. The back end (no longer active due to storage costs) was an AWS Relational Database Service (RDS) loaded with all abstracts scraped from bioRxiv and medRxiv. The Database was accessible by the front end via a rest API (AWS API Gateway) that called an AWS Lambda function to perform the actual queries and process the data.

Home Weather Monitoring System   April 2020 - 2021
- A network of Raspberry-pis constantly monitoring temperature and humidity of the various levels of a townhouse. All nodes reported to one main server that displayed the rolling temperature stats on the living room TV.

Fake Article Detector   2020
- Trained an ensemble of machine learning classifiers to identify research articles produced by paper mills based on their abstract and tern-frequency-inverse-document-frequency (TF-IDF). This identified 38 previously unknown fake articles on PubMed. Find the slide deck of the project results here

Technical Skills

Languages: Python, R, Bash, C++, JavaScript, HTML, CSS
Developer Tools: git, Github Actions, AWS RDS, AWS Lambda, AWS API Gateway
Technologies/Frameworks: Nextflow, Snakemake, Networkx, Pandas, Matplotlib, tidy-verse, Plotly, Selenium, Scikit-learn, TensorFlow, Slurm, PyKeen, Shiny
Bioinformatics Tools: GATK, Savvy CNV, BWA, bedtools, vcftools, Picard, Mutec, quiime2