ALUN BLENNETH STOKES

McMaster University, Hamilton, ON
stokea1@mcmaster.ca
+1 (647)-287-2418

RESEARCH INTERESTS

My interests lie in **mathematics and computing**, particularly **number theory** and **symbolic algebra**. Currently, I study **dessins d'enfants** and (often) the computation of **Belyi maps** from permutation representations and passports. For this, I write software products (here, in all of **Julia, Python, and CUDA**) for **high-performance, parallel, and distributed computing**. Perhaps what most people are interested in is my experience across a broad range of **machine learning (ML)** methods, particularly with **modern ML** techniques – including **natural-language processing (NLP), computer vision (CV), some more standard neural networks** (including adversarial and generative networks), and **graph neural networks (GNNs)**. What I bring to the table beyond many data scientists is the rigorous mathematical training in ideas such as **σ-algebras, Borel spaces, measure theory, topology**, and the like – all of which give me the fundamental understanding to work more meaningfully with these models as a data scientist.

EDUCATION

Master of Science (Pure Mathematics) September 2021 - April 2023
McMaster University
Supervisor: *Dr Cameron Franc*

Bachelor of Integrated Science (Mathematics & Statistics) September 2017 - June 2021
McMaster University Honours: *summa cum laude* (10.8/12 cGPA)
Supervisor: *Dr Cameron Franc* *The Search for Self-Contained Numbers*

PUBLICATIONS

[1] **Stokes, A.** Hum, W., Zaslavsky, J. **STEM Fellowship Journal.** **6**(1): 1-5. Available at A Minimal-Input Multilayer Perceptron for Predicting Drug-Drug Interactions.

[2] † **Stokes, A.** Automatically Solving Square-Piece Jigsaw Puzzles using Convolutional Neural Networks with Gradient Boosted Decision Trees. **The Undergraduate Journal.** (12th edition). Accessible at: Automatically Solving Square-Piece Jigsaw Puzzles.

[3] † * **Stokes, A.** The search for self-contained numbers: k-special 3-smooth representations and the Collatz conjecture. **MacSphere**, 2021, [Online]. Available at: The search for self-contained numbers.

*Entries marked with † have **not** been peer-reviewed.* *Entries marked with * are theses.*

EMPLOYMENT

Graduate Research and Teaching Assistant (Current) September 2021 - April 2023
McMaster University *Dr Cameron Franc, various*

- Continuing my theoretical work on dessins d'enfants for my research.
- Working as a teaching assistant in at least 2 courses per semester at both the graduate and undergraduate level.

Research Assistant (Number Theory and Symbolic ML) May 2021 - August 2021
McMaster University *Dr Cameron Franc*

- Investigated machine learning strategies to discriminate non-congruence finite-index subgroups of the modular group and compute Belyi maps corresponding to dessins d'enfants.
- Designed symbolic evolutionary learning framework to allow number theoretic and algebraic problems to be exactly (rather than probabilistically) taught subgroup property identification.

Data Scientist (NLP and the CPI) June 2020 - August 2020
Statistics Canada *Consumer Prices Division (Serge Goussev)*

- Employed numerous NLP methods for hierarchical data structure mapping to aid in calculating the consumer price index.
- Included data manipulation and cleaning before use, and exploratory data techniques to determine appropriate methods.
- Learned to quickly write several literature reviews on current state-of-the-art methods and technical reports on the results of my work.

Research Assistant (Quasi-Hyperbolicity and GNNs) May 2020 - July 2020
McMaster University *Drs George Dragomir and Andy Nicas*

- Built on recent work to investigate how quasi-hyperbolicity could be exploited to reduce roughness and distortion in graph embeddings.
- Using GNNs, I achieved unprecedented (and previously unseen) accuracy at the task using novel node features to encode global structure at the node level.

Research Assistant (CNNs for Biomedical Applications) May 2019 - May 2020
McMaster University *Dr Ned Nedialkov*

- Developed novel convolutional neural networks to segment photoacoustic cancerous breast tissue images.
- Used sophisticated techniques to mitigate the unique style of photoacoustic noise not present in other medical imaging.
- Developed data pipeline and infrastructure with an automated experiment tracking, ranking, monitoring, and batching software to train 100s of models simultaneously for aggressive hyperparameter optimization.
- Networks used for intrasurgical device to assess tumour boundary *during* operations without a radiologist, which may likely reduce reoccurrence rate.

FUNDING, GRANTS, AND AWARDS

Ontario Graduate Scholarship May 2022 - April 2023
$ 15,000 **Competitive**

NSERC USRA May 2021 - August 2021
$ 8,120 **Competitive**

Oriel College (Oxford University) General Funding †*Declined*
£10,000 **Non-competitive**

Dean's Honour List September 2017 - April 2021
Awarded all 4 years of undergraduate degree **Non-competitive**

McMaster Stewart Award May 2020
$ 3,750 **Competitive**

STEM Fellowship Big Data Competition July 2019
$ 3,000 **Competitive**

McMaster President's Award September 2017
$ 2,500 **Non-competitive**

† indicates an award was declined due to not attending the funding institution.

TEACHING ASSISTANTSHIPS

McMaster University January 2022 - April 2022
Graduate Topics in Risk Management (Financial Mathematics) MFM 763
Introductory Number Theory MATH 3H03

- As part MFM 763, I had to learn the CreditMetrics system and other fintech concepts (eg. bond pricing) in order to effectively teach these graduate students (often my seniors) the content.

McMaster University September 2021 - December 2021
Numerical Linear Algebra MATH 3NA3
Linear Algebra I MATH 1B03

McMaster University January 2021 - April 2021
Introduction to Discrete Mathematics *CS 1DM3*

INVITED TALKS AND SEMINARS

Algebra and Algebraic Geometry Seminar November 2021
McMaster University *An Introduction to Belyi Maps*

- Gave a 30-minute presentation on dessins d'enfants, their relevance, and pertinent computational techniques used in my research open to McMaster's math faculty and graduate students.

CANDEV January 2020
Government of Canada *Transformer Embeddings to Identify Course Redundancies*

- Gave a short talk on our use of transfer-learning with a transformer model to cluster courses offered by the Canadian School of Public Service and identify redundancies in course offerings.
- Received several offers to interview given the quality of our work (led to StatCan job!)

Undergraduate Big Data Competition July 2019
STEM Fellowship *Predicting in-vivo Drug Interactions Without Drug Structure*

- A talk given with coauthors on our ML model for predicting *in-vivo* drug-drug interactions using only analytical chemical properties (which was *not* in the literature at the time).
- Used around 1.2 million drug interactions for model training.
- Conference held at York University.

TECHNICAL SKILLS

Languages†	**Python, Julia**, Java, SQL, C/C++, CUDA, MATLAB.
Major Libraries†	**SageMath**, Pytorch, HomotopyContinuation.jl, Tensorflow.
Software & Tools	LaTeX, Git, Zotero, Macuaulay2.
Operating Systems†	**GNU/Linux** (Ubuntu, primarily), MacOS, Windows
Misc.	Cloud-based computing (AWS & GCP), high adaptability – both to technical issues, and methodological ones.

NON-TECHNICAL SKILLS

Working Environments	Work well in both **independent** and team settings
Communication	Strong **technical written skills**, confident public orator
Growth	I am quite good at taking and retaining what I learn from all my various experiences.
Client Interaction	**Fastidious note-taking** and my ability to quickly learn has led to quite good feedback from both **academic supervisors** and **commercial clients** I've worked with.
Teaching & Leadership	My extensive experience TAing and tutoring has given me a good sense of confidence when making well-informed decisions for a group, and **relaying complex information in a digestible way** to others.

†*Listed in order of proficiency* ***Bolding*** *indicates preferentiality*

OTHER PROJECTS

Global Undergraduate Awards September 2021

Dr Ned Nedialkov *Fully Automated Jigsaw Puzzle Solving by Hybrid ML*

- Won first place in North America for a paper on modular CNN with random forest classification technique to solve square-piece jigsaws.
- Reported state-of-the-art piece-wise matching accuracy (around 98% on 16×16 pixel piece pairs and 99% on 32×32).

National Big Data Competition June 2020

Dr Yasaman Amannejad *Medication Recommendation by Matrix Factorization*

- Devised a matrix factorization-based recommender system to predict effective drugs for treating several mental illnesses, given a patient's history with other drugs.
- Performed data-scraping of approximately 50,000 records from Drugs.com.