

MODELING THE MATHEMATICAL SCIENCES COMMUNITY USING MATHEMATICS, STATISTICS, AND

organized by

Ron Buckmire, Katherine Kinnaird, Drew Lewis, and Andres Vindas Melendez

Workshop Summary

Introduction

This workshop sought to catalyze research in the emerging field of “MetaMath.” MetaMath is a subfield of the Science of Science (which applies scientific tools to study the scientific enterprise) that seeks to apply mathematical, statistical, and data science tools to study the mathematical sciences community. The workshop brought together researchers from across the mathematical sciences to collaborate on research problems centered on describing and analyzing the mathematics community, with a particular focus on producing actionable insight that will help the field become more equitable and inclusive. During the week, morning lectures focused on topics such as existing work in MetaMath and Science of Science, as well as lectures on tools that might be applied to these questions such as machine learning techniques and topological data analysis. All of the active working groups had a keen focus on data either finding and curating data or leveraging data to make meaning about the mathematics community.

Working Groups

Many projects were proposed and many participants brought projects that they were currently working on to the workshop. During our workshop, we had approximately eight groups meeting throughout the week. A summary of the work by the groups that reported progress to us by the deadline for this report are given below.

Documenting Elitism at Small Liberal Arts Colleges. Our group was inspired by the work of Buckmire et al. [Buckmire] and Clauset et al. [Clauset] and wanted to see if we can replicate or do a similar analysis with small liberal arts colleges (SLACs) focusing on mathematics faculty. We defined elitism among SLACs as those that are subject to the Endowment Tax, that is, those with endowments exceeding \$500,000 per full-time student. We also identified small for SLAC as institutions that had student populations between 500 and 4000, based on the Carnegie Classification. We narrowed our sample further by considering only institutions with mathematics departments. From this pool, we identified 50 schools that met our criteria and began examining their mathematics department websites to collect information on faculty members’ PhD-granting institutions and academic ranks. The goal of this project is to understand the academic origins of faculty at elite SLACs. Our next goals are to complete our database and do some analysis of the data.

Enumerating Black Math Ph.D.s. During the AIM Workshop we focused on collecting data on Black women mathematicians. Since then, we have nearly doubled the known number of Black women who received PhDs in mathematics. Our current count is 139; we started with 77. To aid in this collection process, we are spreading the word about

this ongoing research project by sending out blurbs to the newsletters of various mathematical organizations during Women’s History Month with a link to a google form (located at <https://tinyurl.com/blackwomenmathphds>) so that individuals can (self) report information about Black women PhDs in the US. We also have identified a graduate student who will be able to support our efforts in data collection and have had conversations with program officers in the Science of Social, Behavioral, and Economic Sciences (SBE) Directorate at NSF about funding a potential proposal related to this project.

Our plan is to use this data collection to do a network analysis on Black mathematicians as a whole—studying movements between institutions, research collaboration and advisor networks, and historical degree patterns. To this end, we also have access to a larger data set of Black Mathematicians in the US that was scraped from the Mathematicians of the African Diaspora pages [MAD] provided by Professor Edray Goins of Pomona College. We are going to modify this dataset utilizing our current collection process for Black women mathematicians. This list currently has 501 total Black mathematicians.

MathData Hub. This working group was organized to make progress on the idea of creating a central hub for data about the mathematics community. The most prominent members are Sarah Bryant, Chief Scientific Advancement Officer at the AMS, and David Bressoud, Emeritus Professor of Mathematics at Macalester College and current PI of the latest NSF award (https://www.nsf.gov/awardsearch/show-award?AWD_ID=2336355DUE-2336355) *to fund the 2025 CBM and Four – Year Institutions.*)

Models of the Mathematics Degree Pipeline. Our group focused on developing models to understand the dynamics of academic pathways within the mathematical sciences community. During the workshop, we engaged in collaborative discussions on how mathematical modeling can be used to capture degree inflows and academic progression within the community. The workshop format strongly encouraged open exchange of ideas, and many of our most productive research directions emerged from informal conversations and extended discussions during the week.

By the end of the workshop, our group had made substantial progress on several different mathematical models of the mathematical degree pipeline from 1970 to 2020 in the United States. The group continued to work synchronously and asynchronously online and coalesced around three research products. One product involves a two-compartment model where the unknown parameters are assumed to vary over time and only flows from each compartment are known [Topaz]. The second product is a data anchored discrete time stock flow model with constant parameters that reconstructs undergraduate and graduate populations from observed degree data and uses vacancy limited hiring to examine postdoctoral accumulation and faculty workforce dynamics [Babasola]. The third product is a pedagogically oriented paper intended to describe how ordinary and delay differential equations can be used to mathematically model the real-world dynamics of the mathematical degree pipeline. The first two products have been submitted for publication ([Babasola],[Topaz]) and the third is still in development.

The highly collaborative and intellectually stimulating atmosphere of the MetaMath workshop played an important role in refining our models, clarifying research questions, and identifying directions for future work.

#disruptJMM.

Nonnegative matrix factorization was introduced by MetaMath workshop participant Deanna Needell as a machine learning technique with a clear algorithm where it is easy to trace how it made its decisions [Kassab1,Kassab2]. We used the algorithm to understand the #DisruptJMM twitter data set from a new direction. The #DisruptJMM data set comes from previous MetaMath work [Roca, Lewis], where Twitter data using the hashtag was obtained to study the push for social justice at the Joint Mathematics Meetings. This data set had already been cleaned, qualitatively coded, and anonymized, so it was an ideal candidate to use while learning about nonnegative matrix factorization. We had very promising results with just using the algorithm directly, without experimenting with rank or seed words.

Our future work will entail seeing if experimenting with rank or seed words finds new or similar relationships as those found in prior work. We are now meeting virtually to write up the results and think about other data candidates to study using NMF and other machine learning techniques.

Bibliography

- [Babasola] Oluwatosin Babasola, Olayemi Adeyemi, Ron Buckmire, Daozhou Gao, Maila Hallare, Olaniyi Iyiola, Deanna Needell, Chad M. Topaz, and Andrés R. Vindas-Meléndez. “A Discrete-Time Model of the Academic Pipeline in Mathematical Sciences with Constrained Hiring in the United States”. <https://arxiv.org/abs/2602.12188>
- [Buckmire] Ron Buckmire, Carrie Diaz Eaton, Joseph E. Hibdon Jr, Jakini Auset Kauba, Drew Lewis, Omayra Ortega, José L. Pabón, Rachel Roca, and Andrés R. Vindas-Meléndez. “Quantifying and Documenting Gender-Based Inequalities in the Mathematical Sciences in the United States.” In *Advances in Data Science: Women in Data Science and Mathematics (WiSDM) 2023*, pp. 335-352. Cham: Springer Nature Switzerland, 2025.
- [Clauset] Clauset, Aaron, Samuel Arbesman, and Daniel B. Larremore. “Systematic inequality and hierarchy in faculty hiring networks.” *Science advances* 1, no. 1 (2015): e1400005.
- [Kassab1] Lara Kassab, Alona Kryshchenko, Hanbaek Lyu, Denali Molitor, Deanna Needell, and Elizaveta Rebrova. On nonnegative matrix and tensor decompositions for covid-19 twitter dynamics. arXiv preprint arXiv:2010.01600, 2020.
- [Kassab2] Lara Kassab, Alona Kryshchenko, Hanbaek Lyu, Denali Molitor, Deanna Needell, Elizaveta Rebrova, and Jiahong Yuan. Sparseness-constrained nonnegative tensor factorization for detecting topics at different time scales. *Frontiers in Applied Mathematics and Statistics*, Volume 10 - 2024, 2024.
- [Lewis] Drew Lewis, Rachel Roca, Stephanie Marshall, Joseph E. Hibdon, and Carrie Diaz Eaton. Key network disruptors: A structural analysis of the #disruptJMM network influencers and actions. December 2025.
- [MAD] Mathematicians of the African Diaspora. <https://www.mathad.com>.
- [Roca] Rachel Roca, Carrie Diaz Eaton, Drew Lewis, Joesph Hibdon, and Stefanie Marshall. #disruptJMM: Online social justice advocacy and community building in mathematics. *The Journal of Humanistic Mathematics*, 13(2):39–63, 2023.
- [Topaz] Chad M. Topaz, Oluwatosin Babasola, Ron Buckmire, Daozhou Gao, Maila Hallare, Olaniyi Iyiola, Deanna Needell, and Andrés R. Vindas-Meléndez. “A dynamical model of the U.S. mathematics graduate degree pipeline”. <https://arxiv.org/abs/2601.08525>