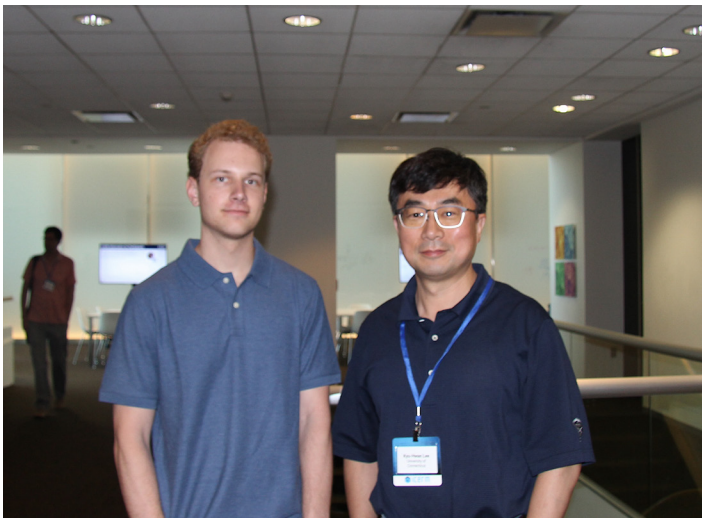
Mathematical Databases and AI: How a Univers project led to a major breakthrough

By Halle Bryant

From July 6-8, ICERM hosted the Hot Topic Workshop Murmurations in Arithmetic, organized by Yang-Hui He (London Institute for Mathematical Sciences and Merton College, Oxford University), Kyu-Hwan Lee (University of Connecticut), and Thomas Oliver (Teesside University). The workshop brought scholars together to discuss the murmurations phenomenon, which was first uncovered in 2022 as an outcome of the organizers' research collaboration applying machine learning to datasets from the L-functions and Modular Forms Database (LMFDB). It was University of Connecticut undergraduate Alexey Pozdnyakov who first observed the phenomenon while exploring a dataset given to him by his professor, Kyu-Hwan Lee.



Pozdnyakov (left) and Lee at ICERM during Murmurations in Arithmetic

After enrolling in Lee's Fall 2021 graduate Abstract Algebra course, Pozdnyakov told his professor he was looking for a research project combining mathematics and computer science. In what turned out to be a career-shifting gesture, Lee started offering Pozdnyakov datasets from the LMFDB to experiment on. In the spring of 2022, Pozdnyakov's experimentation with a dataset of elliptic curves produced an unexpected picture.

By performing logistic regression and principal component analysis on the dataset, he obtained a graph showing a bias that had previously gone undetected. "Pretty soon," he said, "this interesting picture came up and I showed it to [Lee]... He was surprised to see it, and at that point, we knew we had something interesting." The points on the graph behaved like flocks of birds with local interactions impacting the larger structure, inspiring the name "murmurations phenomenon."

Out of that discovery came the paper "Murmurations of elliptic curves," released in April 2022 by He, Lee, Oliver, and Pozdnyakov. The authors study the flocking or gathering of elliptic curves according to their rank. "These oscillations were striking and not expected, so whenever we showed this picture to number theorists they were very surprised because they expected [a] separation, but never these [patterns]," Lee explained.

One of the number theorists they consulted was Andrew Sutherland (MIT), managing editor of the LMFDB and Simons Collaboration Principal Investigator. Pozdnyakov and Lee credited Sutherland with the realization that their results with elliptic curves applied to a wide range of mathematical objects. "He took the computation we did and experimented with it in a bunch of other settings. And what he noticed was that, just like with elliptic curves, you have this bias that seems to oscillate with other curves and other arithmetic objects," said Pozdnyakov.

After consulting with Sutherland, the group set out to study the phenomenon more widely. They turned first to Dirichlet characters, which appealed to them because of their simplicity

Source: Alexey Pozdnyakov. A plot of murmurations where blue, red, and green correspond to elliptic curves of ranks 0, 1, and 2, respectively

compared to elliptic curves. The follow-up paper "Murmurations of Dirichlet characters" was published in July by Lee, Oliver, and Pozdnyakov. All four authors of "Murmurations of Elliptic Curves" are currently working on a paper reporting general observations of murmurations in various settings.

Commenting on how the breakthrough will impact his research and teaching going forward, Lee explained how "AI-guided math research is emerging in number theory, in representation theory, in combinatorics, and also in geometry." Thanks to open-access resources such as the LMFDB, he sees growing potential for students and non-academics to make significant contributions to this work. "[Even if one] doesn't have a strong background in mathematics traditionally or a long education in sophisticated math research, I can invite anyone who has the desire and some commitment and new ideas."

Pozdnyakov's experience is proof of these emerging possibilities. He thanked the developers of the LMFDB for creating such a comprehensive resource. "Before this," he shared, "I'd never taken a class on number theory. But the way I was able to learn about a lot of these things was by looking at the data on LMFDB [since] they have explanations of basically every technical term. So just by going on there, clicking, and reading up on the information they provide, I was able to learn and become involved." He also thanked Lee, Yang-Hui He, and Thomas Oliver for welcoming him as a collaborator.

This fall, Pozdnyakov will enter his senior year at the University of Connecticut where he is majoring in mathematics and computer science with a minor in physics. After visiting ICERM to attend Murmurations in Arithmetic and the subsequent Simons Collaboration Workshop LMFDB, Computation, and Number Theory (LuCaNT), he is excited to continue utilizing machine learning to advance theoretical math. "I'm pretty set on this idea going forward of trying to study math and do math research with the tools of AI," he said. "I want people to be aware of these tools and know how to use them and hopefully I can help pioneer this direction of using AI and machine learning to do mathematics research."

Lee will be back at ICERM for the Fall 2025 Semester Program Combinatorics and Representation Theory, which he is organizing alongside Chris Bowman, Nicolle Gonzalez, Nicolas Libedinsky, Rosa Orellana, Greta Panova, Anne Schilling, Lauren Williams, Jamie Vicary, and Adam Wagner.

The pair thanked the National Science Foundation, the Simons Foundation, ICERM, and the creators of the LMFDB for their ongoing support.

The LMFDB is supported in part by the Simons Collaboration. Below, you can watch an animation showing the growth of the database since its founding in 2011!

Source: Andrew Sutherland