Pranay Mundra pranay.mundra@yale.edu https://mundrapranay.github.io

Programming Skills

Languages: C++, Java, Python, SQL, Go, R, JavaScript, Rust

Technologies: CUDA, AWS, Azure, Pytorch, MongoDB, Boost-Python, React, Faiss, TensorFlow, PostgreSQL

Experience

Yale University New Haven, CT

Graduate Research Assistant - Advisor Quanquan C. Liu *August 2024 - Present*

- Developing provably practical and accurate locally edge differentially private (LEDP) graph algorithms

University of Rochester Medical Center – Office of Research IT Rochester, NY

Research Data Engineer II - Biostatistics & Computational Biology: McCall Research Group *Jan 2024 - August 2024*

- Developing intelligent **storage solutions** for large sequencing data in the MicroRNA project, **optimizing data retrieval** for faster analysis and inference.
- Leading the creation of an **open-source end-to-end software** for microglia image analysis in collaboration with the team, packaging research ideas into accessible and usable software.

Massachusetts Institute of Technology – CSAIL Group Remote

Graduate Summer Researcher - Advisors : Quanquan Liu & Julian Shun *June 2023 - August 2023*

- Implement a benchmark suite for **privacy-preserving locally adjustable graph algorithms** in **parallel and distributed settings** with the Parallel Computing Group. Code available on request.

Paris Lodron Universität Salzburg – Database Research Group Salzburg, Austria

Graduate Summer Researcher - Advisor : Martin Schäler *June 2023 - August 2023*

- Developed an alignment algorithm for **dynamic bipartite graph matching**, to uncover intricate language patterns and semantic similarities in biblical texts across languages and historical epochs.
- Collaborated with a team of linguists and researchers to integrate natural language processing and set similarity search algorithms into the BOSS project to improve search accuracy.

University of Rochester – Computer Science Department Rochester, NY

Graduate Research Assistant - Advisor : Fatemeh Nargesian *July 2021 - May 2023*

- Developed and implemented **KOIOS**, a **novel, exact, efficient, and generic** filter verification system for **top-k** set similarity search using semantic overlap, achieving **5.5x ehanced performance** over current state of the art methods. Published in **IEEE ICDE 2023**.
- Developed an algorithm **fair coreset selection** with **400x speedup** over existing ML-based techniques. Evaluated on MNIST, FashionMNIST, CIFAR10, obtaining **70**% accuracy with only **24**% data.
- Worked on developing **Quok**, an innovative system for **approximate query answering** over Open Knowledge. Addressed challenges related to diverse, noisy, and incomplete data. Published in **HILDA 2023**.

California Institute of Technology – Anima AI Lab Pasadena, CA

Undergraduate ML Researcher - Advisors : Forough Arabshahi & Animashree Anandkumar *June 2020 - April 2021*

- Developed a **novel recursive neural network architecture**, Tree Stack Memory Units (Tree-SMU), to enable **compositional generalization** in the domain of mathematical reasoning.
- Evaluated the generalization of Tree-SMU on four different compositionality tests. We showed that Tree-SMU consistently outperforms the compositional generalization of powerful baselines such as transformers, tree transformers and Tree-LSTMs. (**arXiv Preprint**)

University of Washington – Database Group Seattle, WA

Undergraduate Research Assistant - Advisors: Brandon Haynes, Batya Kenig & Dan Suciu *April 2019 - December 2020*

- Developed high-performance Python API for **LightDB**, accelerating query speed and enabling access to various video data, including VR and AR videos.
- Optimized Python API mapping to low-level constructs, **reducing device transfer time** and delivering faster query execution for users.

• Implemented **boost-python** framework, expanding query expression usage in LightDB and **increasing user adoption**.
• Created **Maimon**, a pioneering system for discovering approximate **MultiValued Dependencies** and acyclic schemas, employing information theory principles. Optimized Maimon to **minimize file scans** by leveraging information theory for MVD pruning and entropy calculations, comparing performance with in-memory and MySQL databases. Published in **ACM SIGMOD 2020**.

Education

Yale University — New Haven, CT
Ph.D. Computer Science — *August 2024 - Present*

University of Rochester — Rochester, NY
M.Sc. Computer Science (3.8/4.0) — Full Scholarship & Research Assistantship — *August 2021 - December 2023*

University of Washington — Seattle, WA
B.Sc. Mathematics (3.0/4.0) — *September 2017 - June 2021*

Publications

Approximate Query Answering over Open Data: Mengqi Zhang, **Pranay Mundra**, Chukwubuikem Chikweze, Fatemeh Nargesian, Gerhard Weikum. (HILDA 2023)[Paper]
KOIOS : Top-k Semantic Overlap Set Search: **Pranay Mundra**, Jianhao Zhang, Fatemeh Nargesian, and Nikolaus Augsten. (IEEE ICDE 2023)[Paper]
Mining approximate acyclic schemes from relations: Batya Kenig, **Pranay Mundra**, Guna Prasaad, Babak Salimi, and Dan Suciu. (ACM SIGMOD 2020)[Paper]
Compositional Generalization with Tree Stack Memory Units: Forough Arabshahi, Zhichu Lu, **Pranay Mundra**, Sameer Singh, Animashree Anandkumar. (arXiv Preprint)[Paper]

Relevant Courses

Computer Science: Computer Programming I, II; Introduction to Database Systems; Database Systems Internals; Data Structures & Algorithms; Linux Fundamentals; Introduction to Artificial Intelligence; Advanced Algorithms; Analytical Methods in Computer Science; Machine Learning, Parallel & Distributed Systems; Computer Networks; Data Mining; Computational Complexity; End to End Deep Learning; Collaborative Programming & Software Design; Introduction to Cryptography.
Mathematics: Honors Calculus I, II, III; Real Analysis I, II; Linear Analysis; Probability I, II; Differential Equations, Linear Algebra, Numerical Analysis I, II; Modern Algebra I, II; Combinatorial Theory I, II.

Teaching Experience

University of Rochester — Rochester, NY
Department of Computer Science : Graduate Teaching Assistant

- CSC 261/461 - **Database Systems** (Spring 2023)
- CSC 263/463 - **Data Management Systems** (Spring 2022)
- CSC 244/444 - **Knowledge Representation in AI**, (Fall 2022)

University of Washington — Seattle, WA
Paul G. Allen School of Computer Science & Engineering : Undergraduate Teaching Assistant

- CSE 444 - **Database Systems Internals**, (Winter 2021)
- CSE 414/344 - **Introduction to Database Systems**, (Fall 2020, Winter 2020, & Spring 2019)

Projects

AquaDB & SimpleDB: Implemented a multi-user transactional database server written in Go and Java respectively.
Gene Regulatory System: Optimized the code to leverage GPU parallelism using the CUDA framework for the following paper: McMurray et al. Gene network modeling via TopNet reveals functional dependencies between diverse tumor-critical mediator genes.
Husky Map Server: Created a google map for the University of Washington campus, which shows the shortest path between two locations.
Flight Booking Application: Implemented a flight booking service with user management, transaction support, itinerary search & reservations.
Spotify Song Explorer: Web Application that allows visualization of different audio features for Top 50 songs, fetched using the Spotify API.