

SHANGDI YU

MIT CSAIL
32 Vassar Street, 32-G728
Cambridge, MA 02139

shangdiy@mit.edu
<https://yushangdi.github.io>

EDUCATION

Massachusetts Institute of Technology, Cambridge, MA Expected 2024
Ph.D. in Electrical Engineering and Computer Science
Dissertation title: Parallel Algorithms, Optimizations, and Benchmarks for Spatial and Graph Clustering
Advisor: Julian Shun
Cornell University, Ithaca, NY 2016 – 2019
B.Sc. in Computer Science; B.Sc. in Operations Research and Engineering
Summa Cum Laude

RESEARCH INTERESTS

Parallel algorithms, parallel computing, performance engineering, software engineering, clustering algorithms, graph algorithms, dynamic algorithms

SELECTED RESEARCH PROJECTS

Parallel algorithms for metric data. Metric datasets contain points where each point represents a numeric vector. I developed the following scalable parallel clustering algorithms for low- and high-dimensional metric data: HDBSCAN [8], HAC [7], DPC [3], and DBHT [1]. I also developed parallel computational geometry libraries, such as ParGeo [6][13] and GeoGraph [14], and algorithms, such as finding the closest pair of points [9].

Parallel algorithms for graphs. A graph contains a set of nodes and a set of edges that connect the nodes and is used to model relationships between objects. I have developed parallel algorithms for the k -core decomposition problem [4][5] and k -clique counting [10].

AWARDS & HONORS

| | |
|---|------|
| Best Student Presentation Award at ACDA [3] | 2023 |
| Best Paper Award at SPAA [4] | 2022 |
| Two Sigma Diversity PhD Fellowship Finalist | 2022 |
| Siebel Scholar Fellowship, Class of 2022 | 2021 |
| MIT EECS Department, Edwin S. Webster Graduate Fellowship | 2019 |
| Cornell University, Computer Science Prize for Academic Excellence and Leadership | 2019 |
| Cornell University, Operations Research Prize for Academic Excellence | 2019 |

PUBLICATIONS

CONFERENCE PAPERS

- [1] **Shangdi Yu**, Julian Shun, “[Parallel Filtered Graphs for Hierarchical Clustering](#)”, Proceedings of the IEEE International Conference on Data Engineering (ICDE), pp. 1967-1980. 2023
- [2] **Shangdi Yu**, Horace He, “[Transcending Runtime-Memory Tradeoffs in Checkpointing by being Fusion Aware](#)”, Proceedings of Machine Learning and Systems 5 (MLSys). 2023
- [3] Michael Huang*, **Shangdi Yu***, and Julian Shun “[Faster Parallel Exact Density Peaks Clustering](#)”, Proceedings of the SIAM Conference on Applied and Computational Discrete Algorithms (ACDA), pp. 49-62. 2023. (*Equal Contribution) (**Best Student Presentation Award**)
- [4] Quanquan Liu, Jessica Shi, **Shangdi Yu**, Laxman Dhulipala, and Julian Shun, “[Parallel Batch-Dynamic Algorithms for \$k\$ -Core Decomposition and Related Graph Problems](#)”, Proceedings of the ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), pp. 191–204, 2022 (**Best Paper Award**)

- [5] Laxman Dhulipala, Quanquan Liu, Sofya Raskhodnikova, Jessica Shi, Julian Shun, and **Shangdi Yu**, “[Differential Privacy from Locally Adjustable Graph Algorithms: k-Core Decomposition, Low Outdegree Ordering, and Densest Subgraphs](#)”, Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS), pp. 754-765, 2022.
- [6] Yiqiu Wang, Rahul Yesantharao, **Shangdi Yu**, Laxman Dhulipala, Yan Gu, and Julian Shun, “[ParGeo: A Library for Parallel Computational Geometry](#)”, Proceedings of the European Symposium on Algorithms (ESA), 2022.
- [7] **Shangdi Yu**, Yiqiu Wang, Yan Gu, Laxman Dhulipala, Julian Shun, “[ParChain: A Framework for Parallel Hierarchical Agglomerative Clustering using Nearest-Neighbor Chain](#)”, Proceedings of the VLDB Endowment 15, 2, pp. 285–298, 2021
- [8] Yiqiu Wang, **Shangdi Yu**, Yan Gu, and Julian Shun, “[Fast Parallel Algorithms for Euclidean Minimum Spanning Tree and Hierarchical Spatial Clustering](#)”, Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pp. 1982-1995. 2021.
- [9] Yiqiu Wang, **Shangdi Yu**, Yan Gu, and Julian Shun, “[A Parallel Batch-Dynamic Data Structure for the Closest Pair Problem](#)”, Proceedings of the International Symposium on Computational Geometry (SoCG), pp. 60:1-60:16, 2021.
- [10] Laxman Dhulipala, Quanquan Liu, Julian Shun, and **Shangdi Yu**, “[Parallel Batch-Dynamic k-Clique Counting](#)”, Proceedings of the SIAM Symposium on Algorithmic Principles of Computer Systems (APOCS), pp.129–143, 2021.

WORKSHOP PAPERS AND POSTERS

- [11] Zhen Guo, **Shangdi Yu**, “[AuthentiGPT: Detecting Machine-Generated Text via Black-Box Language Models Denoising](#)”, NeurIPS’23 Workshop: Generative AI for Education (GAIED). 2023.
- [12] Zhen Guo, Peiqi Wang, Yanwei Wang, and **Shangdi Yu**, “[Improving Small Language Models in Domain-Specific QA via Generative Data Augmentation](#)”, LLM4AI’23: KDD Workshop on Foundations and Applications in Large-scale AI Models-Pre-training, Fine-tuning, and Prompt-based Learning. 2023.
- [13] Yiqiu Wang, **Shangdi Yu**, Laxman Dhulipala, Yan Gu, and Julian Shun, “[POSTER: ParGeo: A Library for Parallel Computational Geometry](#)”, ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP), 2022

JOURNALS

- [14] Yiqiu Wang, **Shangdi Yu**, Laxman Dhulipala, Yan Gu, and Julian Shun, “[GeoGraph: A Framework for Graph Processing on Geometric Data](#)”, ACM SIGOPS Operating Systems Review (OSR), Vol. 55 Issue 1, pp. 38-46, 2021.
- [15] Xiang Fu*, **Shangdi Yu***, and Austin R. Benson, “[Modelling and Analysis of Tagging Networks in Stack Exchange Communities](#)”, Journal of Complex Networks 8, No.5, 2020. (*Equal Contribution)

WORK EXPERIENCE

Student Researcher, Google

Graph Mining Team

Remote

Nov. 2022 – Present, Part-time

- Developed dynamic hierarchical agglomerative clustering (HAC) algorithms for graphs.

Software Engineer Intern, Meta Inc.

PyTorch Compiler Team

Menlo Park, CA

May 2022 – Aug. 2022, Full-time

- Researched compiler optimizations that are incorporated into PyTorch’s optimization pipeline. [2]

Technology Intern, New York Times Company

Messaging Team

New York, NY

May 2018 – Aug. 2018, Full-time

- Analyzed the engagement levels and subscription habits of more than fifteen million newsletter subscribers.

TEACHING AND MENTORING EXPERIENCE

Teaching Assistant

| | |
|---|--------------------|
| Design and Analysis of Algorithms (6.064), MIT EECS | <i>Spring 2021</i> |
| Introduction to Analysis of Algorithms (CS4820), Cornell CS | <i>Spring 2019</i> |
| Machine Learning for Intelligent Systems (CS4780), Cornell CS | <i>Fall 2018</i> |
| Intro to Computing Using Python (CS1110), Cornell CS | <i>Fall 2017</i> |
| <ul style="list-style-type: none">• Taught weekly recitations and led problem-solving sessions and programming labs.• Developed practice problems, exams, and course evaluation surveys to help students understand course material. | |

Grader

| | |
|---|--------------------|
| Learning and Decision Making (CS5726), Cornell CS | <i>Spring 2019</i> |
| Networks II: Market Design (CS4852), Cornell CS | <i>Spring 2018</i> |
| <ul style="list-style-type: none">• Graded students' assignments and exams. | |

Guest Lecturer

| | |
|---|--------------------|
| Algorithm Engineering (6.827), MIT EECS | <i>Spring 2022</i> |
| <ul style="list-style-type: none">• Gave a guest lecture on parallel hierarchical agglomerative clustering. [7] | |

Mentor

| | |
|--|-----------------------|
| For high school student Yifan Kang | <i>2023 – Present</i> |
| <ul style="list-style-type: none">• Yifan is exploring data structures designed for multi-socket non-uniform memory accesses machines. | |
| For undergraduate student Jamison Meindl | <i>2023 – Present</i> |
| <ul style="list-style-type: none">• Jamison is exploring the performance comparison between different parallel graph clustering algorithms. | |
| For undergraduate students Sabiyah Ali and Pranali Vani | <i>2023</i> |
| <ul style="list-style-type: none">• Sabiyah and Pranali worked with me on using stencil computation to accelerate binomial option pricing. | |
| For M.Eng student Terryn Brunelle | <i>2022 – 2023</i> |
| <ul style="list-style-type: none">• Terryn worked with me on her master's thesis on parallelizing tree traversals for binomial option pricing. | |
| For high school student Micheal Huang | <i>2022 – 2023</i> |
| <ul style="list-style-type: none">• Michael worked with me on developing parallel exact and approximate density peaks clustering algorithms. [3] He received the Gold Computer Science Award in S.-T. Yau High School Science Competition in the USA region and the Silver Computer Science Award in the global region. He also presented this work at the Joint Mathematics Meetings. | |
| For undergraduate student Yiwei Zhao | <i>Summer 2020</i> |
| <ul style="list-style-type: none">• Yiwei worked with me on exploring parallel density-based spatial clustering and its high-dimensional improvements. | |
| MIT EECS Graduate Application Assistance Program (GAAP) | <i>Fall 2020</i> |
| <ul style="list-style-type: none">• Mentored EECS PhD applicants from underrepresented groups, including guidance through the EECS PhD application process and feedback on statements of purpose and resumes. | |

Workshop Participant

| | |
|--|--------------------|
| Evidence-based Teaching Workshop (MIT Teaching + Learning Lab) | <i>Winter 2021</i> |
| <ul style="list-style-type: none">• Explored how to engage students and lead discussions in general teaching settings both in-person and online. | |

Volunteer

| | |
|--|------------------|
| Engaging Youth in Learning (EDUC 4040) at Cornell | <i>Fall 2016</i> |
| <ul style="list-style-type: none">• Coordinated a film club at Dewitt Middle School, guiding students through the process of writing, filming, and editing short films.• Facilitated Mandarin-English translation to assist non-native English-speaking participants. | |

PROFESSIONAL SERVICE

Program Committee Member: SDM24, GRADES-NDA (SIGMOD23 Workshop), DRAGSTERS (PLDI23 Workshop)

Conference Reviewer: ALENEX{24,22}, SOSA24, HPEC23, KDD23, ACDA{23,21}, PPOPP{23,22}, SPAA{22, 21}, SEA22, HiPC21, IPDPS21, ICDCS21, Euro-Par20, ESA20

Journal Reviewer: TKDE, PLOS ONE, JEA, TOPC, TOMS

OTHER ACTIVITIES

Co-founder & Treasurer

MIT Film Makers Association

Cambridge, MA

Jan. 2022- Jan. 2023

- Organized the first AI for Film Making Hackathon, bringing together enthusiasts in film-making to explore new ways of visualization using state-of-the-art AI technology.

Treasurer

MIT EECS Graduate Student Association (GSA)

Cambridge, MA

Jan. 2021- Jan. 2022

- Managed event finances and prepared an annual budget report of ~\$80,000.
- Helped to plan and organize EECS department Visit Day and Orientation, and other social and academic events.

Co-organizer

MIT EECS Theory Lunch Speaker Series

Cambridge, MA

Oct. 2019 – Apr. 2020

- Organized a weekly speaker series at MIT attended by ~50 students, which aims to encourage academic discussions and create an opportunity for students to socialize.

Panel Speaker

Cornell Explore CS Research

Remote

Nov. 2020

- Volunteered at Google's ExploreCSR series at Cornell, which aims to encourage undergraduates, especially women, underrepresented minorities, and first-generation college students to consider pursuing a PhD.

TALKS AND PRESENTATIONS

Parallel Hierarchical Agglomerative Clustering [7]

Data Science Seminar at the University of Utah

Nov. 2023

University of California, Berkeley

Nov. 2023

Stanford Systems Seminar

Nov. 2023

Capital Area Theory Seminar (CATS) at the University of Maryland

Oct. 2023

University of Washington

Jun. 2023

Algorithm Engineering course at MIT

Apr. 2022

International Conference on Very Large Data Bases (VLDB)

Sep. 2022

Faster Parallel Exact Density Peaks Clustering [3]

SIAM Conference on Applied and Computational Discrete Algorithms (ACDA)

May 2023

Parallel Filtered Graphs for Hierarchical Clustering [1]

IEEE International Conference on Data Engineering (ICDE)

Apr. 2023

POSTER: ParGeo: A Library for Parallel Computational Geometry [13]

The Applications Driving Architectures Center (ADA) Symposium

May. 2022

Parallel Batch-Dynamic k-Clique Counting [10]

SIAM Symposium on Algorithmic Principles of Computer Systems (APOCS)

Jan. 2021

REFERENCES

References available upon request.