Sepideh Mahabadi

Email: mahabadi@mit.edu
https://www.mit.edu/~mahabadi

Research Interests

My research is focused on Theoretical Foundations of Massive Data.

- ◇ High-Dimensional Computational Geometry, Similarity Search.
- ◇ Streaming Algorithms, Sub-linear Algorithms, Data Summarization.
- ◇ Social Aspects of Algorithms for Massive Data: Algorithmic Fairness, Diversity Maximization, and Differential Privacy.
- ◇ Graph Algorithms, Metric Embeddings.

Positions

- ◇ Senior Researcher at the **Microsoft Research, Redmond**, September 2021-Present.
- ◇ Research Assistant Professor at the **Toyota Technological Institute at Chicago (TTIC)**, September 2018-September 2021.
- ◇ Postdoctoral Research Scientist at **Simons Collaboration on Algorithms and Geometry** based at the Data Science Institute of **Columbia University**, August 2017-August 2018.

Education

- ◇ Ph.D in Computer Science, **Massachusetts Institute of Technology (MIT)**, Computer Science and Artificial Intelligence Laboratory (CSAIL), Theory of Computation Group (TOC), 2011- 2017.

  Advisor: Prof. Piotr Indyk

  · PhD Thesis: Sub-linear Algorithms for Massive Data Problems

  · M.Sc. Thesis: Approximate Nearest Neighbor and Its Many Variants

- ◇ B.S. in Computer Engineering, **Sharif University of Technology**, 2007-2011.
  Advisor: Prof. Mohammad Ghodsi
  B.Sc. Thesis: Minimum Dominating Set Network Creation Game

Publications

- ◇ Sepideh Mahabadi, Mohammad Roghani, Jakub Tarnawski, Ali Vakilian, *Sublinear Metric Steiner Tree via Improved Bounds for Set Cover*. In the16th Innovations in Theoretical Computer Science (**ITCS**), 2025.
- ◇ Ce Jin, Michael Kapralov, Sepideh Mahabadi, Ali Vakilian, *Streaming Algorithms for Connectivity Augmentation*. In the 51st International Colloquium on Automata, Languages, and Programming (**ICALP**), 2024.
- ◇ Arturs Backurs, Zinan Lin, Sepideh Mahabadi, Sandeep Silwal, Jakub Tarnawski *Efficiently Computing Similarities to Private Datasets*. In the 12th International Conference on Learning Representations (**ICLR**), 2024 .
- ◇ Alexandr Andoni, Piotr Indyk, Sepideh Mahabadi, Shyam Narayanan, *Differentially Private Approximate Near Neighbor Counting in High Dimensions*. In the 37th Annual Conference on Neural Information Processing Systems (**NeurIPS**)(**Spotlight**), 2023 .
- ◇ Sepideh Mahabadi, Stojan Trajanovski, *Core-sets for Fair and Diverse Data Summarization*. In the 37th Annual Conference on Neural Information Processing Systems (**NeurIPS**), 2023.
- ◇ Aditya Bhaskara, Sepideh Mahabadi, Ali Vakilian, *Tight Bounds for Volumetric Spanners and Applications*. In the 37th Annual Conference on Neural Information Processing Systems (**NeurIPS**), 2023.
- ◇ Siddharth Gollapudi, Sepideh Mahabadi, Varun Sivashankar, *Composable Coresets for Determinant Maximization: Greedy is Almost Optimal*. In the 37th Annual Conference on Neural Information Processing Systems (**NeurIPS**), 2023.

- ◇ Sepideh Mahabadi, Shyam Narayanan, *Improved Diversity Maximization Algorithms for Matching and Pseudoforest* In (**APPROX**), 2023.
- ◇ Sedjro Hotegni, Sepideh Mahabadi, Ali Vakilian, *Approximating Algorithms for Fair Range Clustering.* In (**ICML**), 2023.
- ◇ Sepideh Mahabadi, David Woodruff, Samson Zhou, *Adaptive Sketches for Robust Regression with Importance Sampling* . In (**RANDOM**), 2022.
- ◇ Martin Aumüller, Sariel Har-Peled, Sepideh Mahabadi, Rasmus Pagh, Francesco Silvestri, *Sampling a Near Neighbor in High Dimensions–Who is the Fairest of Them All?.* In (**CACM, TODS**, 2022 **SIGMOD Record**), 2021.
- ◇ Arturs Backurs, Sepideh Mahabadi, Konstantin Makarychev, Yury Makarychev *Two-sided Kirszbraun Theorem.* In the 37th International Symposium on Computational Geometry (**SoCG**), 2021.
- ◇ Julia Chuzhoy, Sepideh Mahabadi, Zihan Tan, *Towards Better Approximation of Graph Crossing Number.* In the 61st Annual IEEE Symposium on Foundations of Computer Science (**FOCS**), 2020.
- ◇ Alexandr Andoni, Collin Burns, Yi Li, Sepideh Mahabadi, David Woodruff, *Streaming Complexity of SVMs.* In the 23rd International Conference on Approximation Algorithms for Combinatorial Optimization Problems (**APPROX**), 2020.
- ◇ Sepideh Mahabadi, Ali Vakilian, *Individual Fairness for k-Clustering.* In the 37th International Conference on Machine Learning (**ICML**), 2020.
- ◇ Sepideh Mahabadi, Ilya Razenshteyn, David Woodruff, Samson Zhou, *Non-Adaptive Adaptive Sampling on Turnstile Streams.* In the 52nd ACM Symposium on Theory of Computing (**STOC**), 2020.
- ◇ Piotr Indyk, Sepideh Mahabadi, Shayan Oveis Gharan, Alireza Rezaei, *Composable Core-sets for Determinant Maximization Problems via Spectral Spanners.* In the 31st Annual ACM-SIAM Symposium on Discrete Algorithms (**SODA**), 2020.
- ◇ Sariel Har-Peled, Sepideh Mahabadi, *Near Neighbor: Who is the Fairest of Them All.* In the 33rd Annual Conference on Neural Information Processing Systems (**NeurIPS**), 2019.
- ◇ Piotr Indyk, Sepideh Mahabadi, Shayan Oveis Gharan, Alireza Rezaei, *Composable Core-sets for Determinant Maximization: A simple Near-Optimal Algorithm.* In the 36th International Conference on Machine Learning (**ICML**)(**Long Talk**), 2019.
- ◇ Sariel Har-Peled, Piotr Indyk, Sepideh Mahabadi, *Approximate Sparse Linear Regression.* In the 45th International Colloquium on Automata, Languages, and Programming (**ICALP**), 2018.
- ◇ Sepideh Mahabadi, Konstantin Makarychev, Yury Makarychev, Ilya Razenshteyn, *Nonlinear Dimension Reduction via Outer Bi-Lipschitz Extensions.* In the 50th ACM Symposium on Theory of Computing (**STOC**), 2018.
- ◇ Piotr Indyk, Sepideh Mahabadi, Ronitt Rubinfeld, Ali Vakilian, Anak Yodpinyanee, *Set Cover in Sub-linear Time.* In the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (**SODA**), 2018.
- ◇ Piotr Indyk, Sepideh Mahabadi, Ronitt Rubinfeld, Jonathan Ullman, Ali Vakilian, Anak Yodpinyanee, *Fractional Set Cover in the Streaming Model.* In the 20th International Conference on Approximation Algorithms for Combinatorial Optimization Problems (**APPROX**), 2017.
- ◇ Sariel Har-Peled, Sepideh Mahabadi, *Proximity in the Age of Distraction: Robust Approximate Nearest Neighbor Search.* In the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (**SODA**), 2017.
- ◇ Sariel Har-Peled, Piotr Indyk, Sepideh Mahabadi, Ali Vakilian, *Towards Tight Bounds for the Streaming Set Cover Problem.* In the 35th Symposium on Principles of Database Systems (**PODS**), 2016.

- ◇ Piotr Indyk, Robert Kleinberg, Sepideh Mahabadi, Yang Yuan, *Simultaneous Near Neighbor Search*. In the 32nd International Symposium on Computational Geometry (**SoCG**), 2016.
- ◇ Sepideh Mahabadi, *Approximate Nearest Line Search in High Dimensions*. In the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (**SODA**), 2015.
- ◇ Erik Demaine, Piotr Indyk, Sepideh Mahabadi, Ali Vakilian, *On Streaming and Communication Complexity of the Set Cover Problem*. In the 28th International Symposium on Distributed Computing (**DISC**), 2014.
- ◇ Piotr Indyk, Sepideh Mahabadi, Mohammad Mahdian, Vahab Mirrokni, *Composable Core-sets for Diversity and Coverage Maximization*. In the 33rd Symposium on Principles of Database Systems (**PODS**), 2014.
- ◇ Sofiane Abbar, Sihem Amer-Yahia, Piotr Indyk, Sepideh Mahabadi, Kasturi Varadarajan, *Diverse Near Neighbor Problem*. In the 29th International Symposium on Computational Geometry (**SoCG**), 2013.
- ◇ Sofiane Abbar, Sihem Amer-Yahia, Piotr Indyk, Sepideh Mahabadi, *Efficient Computation of Diverse News*. In the 22nd International World Wide Web Conference (**WWW**), 2013.
- ◇ M. Sadeghi, H. Pezeshk, C. Eslahchi, S. Ahmadian, S. Mahabadi, *Construction of Random Perfect Phylogeny Matrix*. Advances and Applications in Bioinformatics and Chemistry 2010, 3: 89-96.
- ◇ H. Mirzaei, S. Ahmadian, S. Mahabadi, M. Sadeghi M., C. Eslahchi, H. Pezeshk, *An Algorithm for Construction of all Perfect Phylogeny Matrices*. Communications in Mathematical and in Computer Chemistry (MATCH), 2009, 62, 2: 251-259.

Honors and Awards

- ◇ **ACM PODS Alberto O. Mendelzon Test-of-Time Award, 2024** for the paper *Composable Core-sets for Diversity and Coverage Maximization*.
- ◇ **Rising Stars**
  - · Selected to give a Rising Stars Talk in the TCS Women Spotlight workshop, 2020.
  - · Selected to participate in Rising Stars in EECS workshop, hosted at UIUC, 2019.
- ◇ **Gold Medal** in the International Olympiad in Informatics (IOI) 2007, Croatia.
  - · The only female contestant to win a gold medal at IOI 2007.
  - · The first Iranian female contestant to win a gold medal at IOI.
- ◇ Ranked 14^{th} and 13^{th} respectively in the **ACM-ICPC World Finals** 2013, St. Petersburg; and 2011, Florida.
- ◇ Awarded as *Outstanding Student* by university president, Sharif University of Technology, 2007.
- ◇ Recipient of the grant for undergraduate studies from the Iranian National Elites Foundation, for outstanding academic success, 2007 - 2011.

Service

- ◇ **Reviewing Services**
  - · **Program Committees:** FOCS 2025, SODA 2025, ICALP 2023, STOC 2023, WOLA 2022, SoCG 2022, HALG 2022, STOC 2021, SODA 2021, RANDOM 2020, ITCS 2019.
  - · Reviewer in NeurIPS 2023, ICML 2018 and NeurIPS 2017.
  - · External Reviewer for the following conferences: STOC, FOCS, SODA, SoCG, ICALP, ITCS, DISC, APPROX, RANDOM, ESA, SPAA; and the following journals: SICOMP, ALGO, TALG.
- ◇ **Workshop Organization**
  - · "Workshop on High-Dimensional and Complex Data Algorithms'', Venice, May 2025.
  - · "Extroverted Sublinear Algorithms'', Simons Institute, June 2024.

- "Sketching and Algorithm Design Workshop'', Simons Institute, October 2023.
- "Workshop on Local Algorithms (WOLA)'', Virtual, July 2020.
- "Recent Trends in Theoretical CS'', TTIC, January 2020.

◇ **Member of the *Diversity, Equity, and Inclusion Committee* at TTIC.**

◇ **Member of the International Olympiad in Informatics (IOI) 2017 Host Scientific Committee**

◇ **Member of the Student Scientific Chapter (SSC)** of Computer Engineering Department, 2009 - 2010. SSC is the student committee concerned with directing the department extra-curriculum activities.

Teaching Experience

◇ **Instructor** for the *Algorithms for Massive Data* course at TTIC, Spring 2021.

◇ **Guest Lecturer** for *Approximation Algorithms* taught by Julia Chuzhoy, TTIC, Fall 2019.

◇ **Teaching Assistant**:

- Introduction to Algorithms (Spring 2017, MIT),
- Randomized Algorithms (Spring 2013, MIT),
- Design and Analysis of Algorithms (Spring 2010, Sharif UT),
- Theory of Languages and Machines (Fall 2010, Sharif UT),
- Introduction to Programming (Fall 2010, Sharif UT)

◇ **Member of the Scientific Committee and Contest Designer** at Young Scholars Club (YSC), 2007-2009. YSC is the sole regulating body for scientific Olympiads in Iran.

- Teaching Special Topics in Computer Science such as Algorithms and Data Structures, Graph Theory, Computational Geometry and Linear Algebra in training campus of Iranian National Olympiad in Informatics (INOI), YSC, 2008 - 2009.
- Instructor and Scientific Observer of the Iranian team in IOI, Plovdiv, Bulgaria, August 2009.

Mentorship

◇ Mentor/Co-Mentor of the following Summer Interns at MSR Redmond

- Mohammad Roghani
- Sherry Sarkar
- Ce Jin
- Sandeep Silwal
- Shyam Narayanan
- Thuy Duong Vuong

Research Visits

◇ **Long-term participant in Sublinear Algorithms program** at Simons Institute, Summer 2024

◇ **Long-term participant in Data Structures and Optimization for Fast Algorithms program** at Simons Institute, Fall 2023

◇ **Research internship at Toyota Technological Institute, Chicago, 2016**
Mentors: Julia Chuzhoy and Yury Makarychev
Projects: Prioritized embeddings, and algorithms for finding an approximately optimal graph drawing.

◇ **Visit at the University of Illinois Urbana-Champaign, July-August 2015**
Mentor: Sariel Har-Peled
Project: Robust nearest neighbor search.

- **Research internship at Yahoo! Labs NYC, August 2014**
  Mentors: Howard Karloff and Edo Liberty
  Project: k-nearest neighbor classification.
- **Research internship at Google, Mountain View, 2013**
  Mentor: Mohammad Mahdian
  Project: Core-sets for diversity and coverage maximization problems under different notions of diversity.
- **Research internship at ADSC**, joint research center between UIUC and A*STAR, 2010.
  Mentors: Marianne Winslett and Jiangbo Lu.
  Projects: Pill identification and indoor navigation.

Talks

- **Industry Workshop at FOCS**, October 2024
  Diverse Graph-Based Nearest Neighbor Search
- **Michigan AI seminar series and the Women in Computing Seminar (WiCs) series**, October 2024
  Recent Directions in Approximate Nearest Neighbor Search
- **Massive Data Models and Computational Geometry Workshop**, Bonn, Germany, September 2024
  Recent Advances in Diversity Maximization in the Offline and Composable Coreset Models
- **ACM PODS Alberto O. Mendelzon Test-of-Time Award**, Santiago, Chile, June 2024
  Composable Core-sets for Diversity and Coverage Maximization
- **Sublinear Algorithms Bootcamp Workshop at Simons Institute**, May 2024
  Coresets and Their Applications to Design Sublinear Algorithms
- **TTIC Anniversary Workshop**, November 2023
  Fair and Diverse Data Summarization
- **Dynamic Graphs and Algorithm Design Workshop at Simons Institute**, September 2023
  Recent Advances in Diversity Maximization in the Offline and Composable Coreset Models
- **RAISE (Responsible AI and Software Experiences) at UW**, November 2022
  Sampling a Near Neighbor in High Dimensions — Who is the Fairest of Them All?
- **The Statistics and Data Sciences (SDS) Seminar Series**, UT Austin, October 2022
  Diversity Maximization over Large Data Sets
- **Sublinear Algorithms Workshop, MIT**, August 2022
  Sampling a Near Neighbor in High Dimensions — Who is the Fairest of Them All?
- **IDEAL Workshop on High-Dimensional Geometry and Analysis**, May 2022
  Non-Adaptive Adaptive Sampling in Turnstile Streams
- **Theory Colloquium**, UCSB, April 2022
  Diversity Maximization over Large Data Sets
- **Workshop on Algorithms for Large Data**, August 2021
  Two-sided Kirszbraun Theorem
- **ICPCU Alumni Lecture Series**, May 2021
  Diversity and Fairness in Data Summarization Algorithms
- **Michigan-Purdue Online Theory Seminar**, September 2020
  Non-Adaptive Adaptive Sampling in Turnstile Streams
- **MIT Sub-linear Reading Group**, September 2020
  Determinant Maximization over Large Data Sets
- **Rising Stars Talk, TCS Women Spotlight Workshop**, June 2020
  Composable Core-sets for the Determinant Maximization Problem

- ◇ **TCS+ Talk**, April 2020
  Non-Adaptive Adaptive Sampling on Turnstile Streams
- ◇ **CMU**, November 2019
  Composable Core-sets for Determinant Maximization Problems via Spectral Spanners
- ◇ **MSR Redmond**, October 2019
  Diversity Maximization over Large Data Sets
- ◇ **Metric Embeddings and Dimensionality Reduction workshop**, May 2019
  Nonlinear Dimension Reduction via Outer Bi-Lipschitz Extension
- ◇ **Optimization Methods in Computer Vision and Image Processing workshop**, ICERM, May 2019
  Diversity Maximization over Large Data Sets
- ◇ **Midwest Theory Day**, Purdue University, April 2019
  Diversity Maximization over Large Data Sets
- ◇ **UIUC**, February 2019
  Composable Core-sets for Determinant Maximization Problems via Spectral Spanners
- ◇ **Junior Theorists Workshop**, Northwestern Univeristy, November 2018
  Composable Core-sets for Determinant Maximization Problems via Spectral Spanners
- ◇ **University of Chicago**, October 2018
  Nonlinear Dimension Reduction via Outer Bi-Lipschitz Extension
- ◇ **Google New York**, July 2018
  Nonlinear Dimension Reduction via Outer Bi-Lipschitz Extension
- ◇ **Workshop on Local Algorithms (WOLA)**, MIT, June 2018
  Set Cover in Sub-linear Time
- ◇ **Rutgers University/DIMCAS**, May 2018
  Set Cover in Sub-linear Time
- ◇ **Midwest Theory Day**, TTIC, April 2018
  Set Cover in Sub-linear Time
- ◇ **Simons Collaboration on Algorithms and Geometry**, February 2018
  Fractional Set Cover in the Streaming Model
- ◇ **New York Area Theory Day**, NYU, December 2017
  Set Cover in Sub-linear Time
- ◇ **University of Washington**, April 2017
  Approximate Nearest Neighbor and Its Many Faces
- ◇ **IBM Watson**, December 2016
  Approximate Nearest Neighbor and Its Many Variants
- ◇ **University of Toronto**, November 2015
  Towards Tight Bounds for the Streaming Set Cover Problem
- ◇ **Brown University**, November 2014
  Approximate Nearest Line Search in High Dimensions