Neha Makhija

nehamakhija.github.io

Email : makhija.n@northeastern.edu

LinkedIn : linkedin.com/in/neha-makhija/

RESEARCH INTERESTS

Can we design algorithms for modern data science tasks that are performant and easy to implement but still maintain theoretical best-case guarantees, thus achieving the best of theory and practice? I study this question in the domain of explainable data management, provenance, and knowledge representation.

EDUCATION

- PhD in Computer Science   Sept 2019 - 2025 (Expected)
  DataLab at Northeastern University, Khoury College of Computer Sciences   GPA: 3.97 / 4
  Advised by: Dr. Wolfgang Gatterbauer

- Bachelor of Engineering (Computer Engineering)   2015 - 2019
  Thadomal Shahani Engineering College, University of Mumbai   GPA: 9.15 / 10

PUBLICATIONS AND PREPRINTS

(Papers with alphabetically-decided ordering of authors are indicated with [AO])

- Resilience for Regular Path Queries: Towards a Complexity Classification **(Under Review)**
  [AO] *Antoine Amarilli, Wolfgang Gatterbauer, Neha Makhija, Mikaël Monet*
  [ArXiv Preprint] [Code: Computational Verification of Hardness Gadgets]

- A Unified and Practical Approach for Generalized Deletion Propagation **(Under Review)**
  Neha Makhija, Wolfgang Gatterbauer
  [ArXiv Preprint] [OSF Repository with Code]

- A Unified Approach for Resilience and Causal Responsibility with Integer Linear Programming (ILP) and LP Relaxations **[SIGMOD 2024]**
  Neha Makhija, Wolfgang Gatterbauer
  [ACM] [ArXiv (long version)] [Webpage] [Slides] [Video]

- Minimally Factorizing the Provenance of Self-Join Free Conjunctive Queries **[PODS 2024]**
  Neha Makhija, Wolfgang Gatterbauer
  [ACM] [ArXiv (long version)] [Slides] [Video]

- Discovering Dichotomies for Problems in Database Theory **[VLDB PhD Workshop 2023]**
  Neha Makhija, Wolfgang Gatterbauer
  [VLDB] [ArXiv] [Webpage]

- Loch Prospector: Metadata Visualization for Lakes of Open Data
  Neha Makhija, Mansi Jain, Nikolaos Tziavelis, Laura Di Rocco, Sara Di Bartolomeo, Cody Dunne
  [Short Paper at IEEE VIS 2020]
  [IEEE VIS] [OSF] [Video] [Teaser] [Webpage]

EXPERIENCE

- **Simons Institute, University of California, Berkeley,** Visiting Graduate Student   Aug - Dec 2023
  Workshop on "Logic and Algorithms in Database Theory and AI" [Workshop Webpage]
  Invited Talk: "A Unified Approach for Reverse Data Management" [Talk Webpage] [Video]

- **RelationalAI,** Summer Research Intern   May - Aug 2023

- **RelationalAI,** Summer Research Intern   May - Aug 2022

- **IBM Research, Yorktown Heights,** Summer Research Intern   May - Aug 2021

- **Ushva Clean Technology,** Machine Learning Intern   Nov 2017 - Aug 2018

Honours & Awards

- VLDB Travel Award   2023
- Graduate Fellowship, Northeastern University   2019
- Principal's Award of Excellence, Thadomal Shahani Engineering College   2019
- Leadership Award, Thadomal Shahani Engineering College   2019

Teaching And Mentoring

- Research Co-op Mentor for Undergraduate student Yihan Luo for the project 'Building an Intuitive Reverse Data Management System'   June 2024 - Dec 2024
- Teaching Assistant, CS7240: Principles of Scalable Data Management: Theory, Algorithms and Database Systems   Spring 2023
- Head Teaching Assistant, CS3200: Database Design   Fall 2022

Service

- Reviewer, VLDB 2025   April 2024-2025
- Organizer, DataLab Seminar   Fall 2024, Spring 2023
- Microsoft TEALs: Volunteer at Match Charter Public High School Boston   Sept 2021-2022
- MIT Spark Lecture: 'Introduction to Graph Theory (With Puzzles!)' [Slides]   July 2022
- Organizer, Northeastern PhD Open House   2022
- Co-founder/Lead, TSEC CodeCell   2017-2019

Talks and Posters

- A Unified Approach for Reverse Data Management

    - Everything Data Seminar, Duke University   Nov 2024
    - Database Seminar, Columbia University   Nov 2024
    - Data Management and Mining Laboratory Seminar, University of British Columbia   Oct 2024
    - Data Labs Seminars, University of Illinois Chicago   Oct 2024
    - Databases and Information Systems Seminar, Humboldt-Universität zu Berlin   Oct 2024
    - Northeast Database Day 2024 [Slides]   May 2024
    - Database Seminar, University of California San Diego   Feb 2024
    - Northeastern University Theory Lunch   Feb 2024
    - UtahDB Lab Seminar, University of Utah   Nov 2023
    - Cornell Database Seminar   Sept 2023
    - VLDB PhD Workshop   Aug 2023
    - Data Systems Group at MIT CSAIL   March 2023

- A Unified Approach for Resilience and Causal Responsibility

    - SIGMOD 2024 [Slides] [Video]   June 2024
    - Institute of Algebra, TU Dresden   March 2024

- Minimally Factorizing the Provenance of Self-Join Free Conjunctive Queries

- – PODS 2024 [Slides] [Video]   June 2024
- – Boston Computation Club   August 2021

- Loch Prospector: Meta Data Visualization for Lakes of Open Data

  - – IEEE VIS 2020 [Video] [Teaser]   October 2020

GRADUATE COURSEWORK

Advanced Algorithms, Large Scale and Parallel Data Processing, Scalable Database Management, Information Visualization, Distributed Systems, Graph Theory, Effective Scientific Communication

REFERENCES

- Wolfgang Gatterbauer
  Associate Professor, Khoury College of Computer Sciences, Northeastern University
  Email address : w.gatterbauer@northeastern.edu

- Mirek Riedewald
  Professor, Khoury College of Computer Sciences, Northeastern University
  Email address : m.riedewald@northeastern.edu

- Antoine Amarilli
  Advanced Researcher in the LINKS team at Inria Lille (on leave from Telecom Paris)
  Email address: a3nm@a3nm.net