Hanchao Ma

Computer & Data Science Department
Case Western Reserve University, Cleveland, OH, 44106

Phone: (412) 905-9232
Email: hxm382@case.edu

Research Interests

Knowledge, Data Management, and AI-Powered Data Systems

AI-Powered Domain-Specific Aata and Workflow Management (e.g., Material Science, Power Grid, Cyber Security)

Advanced Retrieval Augmented Generation (RAG) systems for Large Language Models

Advanced Data Systems towards Empowering Large Language Models

Education

Ph.D. Computer Science
Case Western Reserve University, OH 09/2019-12/2025
Advisor: Dr. Yinghui Wu
Thesis: " Query Suggestions for Knowledge Exploration"

M.S. in Information Science
University of Pittsburgh, PA 09/2012/-05/2014

B.S. in Information System
Nanjing University of Posts and Telecommunications, Nanjing, China 09/2008-07/2012

Professional Appointment

Academia

1. **Postdoctoral Associate** 09/2025-present
   Computer & Data Sciences Department
   Case Western Reserve University
   Cleveland, OH
   Supervisor: Dr. Yinghui Wu

2. **Research Assistant** 09/2019-09/2025
   Computer & Data Sciences Department
   Case Western Reserve University
   Cleveland, OH
   Supervisor: Dr. Yinghui Wu

3. **Research Assistant** 09/2017-07/2019
   Computer Sciences Department
   Washington State University
   Pullman, WA
   Supervisor: Dr. Yinghui Wu

4. **Research Intern** 06/2018-09/2018
   Pacific Northwest National Laboratory
   Richland, WA
   Mentor: Dr. Sutanay Choudhury

5. **Research Intern** 06/2017-09/2017
   Microsoft Research Asia
   Beijing, China
   Mentor: Dr. Yuanchao Shu

Industry

1. **Applied Scientist Intern** 06/2023-09/2023
   Shopping Video Team, Amazon
   Seattle, WA Mentor: Daniel Blackburn

2. **Software Engineer** 05/2015-05/2016
   Chem Advisor, Inc.
   Pittsburgh, PA

Professional Honors and Awards

1. **Outstanding Dissertation Award**, Computer and Data Science Department at CWRU 2025
2. ICDE Student Travel Award 2022,2023
3. CIKM Student Travel Award 2022
4. WSDM Student Travel Award 2022
5. **Best Paper Award**, IEEE International Conf. on Big Data 2020
6. VLDB Student Travel Award 2019

Publications

Citations: 245, source: Google Scholar

$^{+}$equal contribution, $^{\#}$undergraduate/graduate student mentored

Refereed Conference Papers

1. *"Generating Skyline Datasets for Data Science Models"*, Mengying Wang$^{\#+}$, **Hanchao Ma**$^{+}$, Yiyang Bian, Yangxin Fan, and Yinghui Wu in EDBT, 2025.
2. *"ModsNet: Performance-aware Top-k Model Search using Exemplar Datasets"*, Mengying Wang$^{\#}$, **Hanchao Ma**, Sheng Guan, Yiyang Bian, Haolai Che, Abhishek Daundkar, Alp Sehirlioglu, and Yinghui Wu in VLDB, 2024.
3. *"Selecting Top-k Data Science Models by Example Dataset"*, Mengying Wang$^{\#}$, Sheng Guan, **Hanchao Ma**, Yiyang Bian, Haolai Che, Alp Sehirlioglu, and Yinghui Wu in CIKM, 2023.
4. *"Fair Group Summarization with Graph Patterns"*, **Hanchao Ma**, Sheng Guan, Mengying Wang$^{\#}$, Qi Song, and Yinghui Wu in ICDE, 2023.
5. *"GALE: Active Adversarial Learning for Erroneous Node Detection in Graphs"*, Sheng Guan, **Hanchao Ma**$^{+}$, Mengying Wang, and Yinghui Wu in ICDE, 2023.

6. *"Crowdsourced Materials Science Resource and Workflow Exploration"*, Mengying Wang$^{+\#}$, **Hanchao Ma**$^{+}$, Sheng Guan, and Yinghui Wu in CIKM, 2022.

7. *"RoboGNN: Robustifying Node Classification under Link Perturbation"*, Sheng Guan, **Hanchao Ma**+, and Yinghui Wu in IJCAI, 2022.

8. *"Subgraph Query Generation with Fairness and Diversity Constraints"*, **Hanchao Ma**, Sheng Guan, Mengying Wang$^{\#}$, Yen-shuo Chang$^{\#}$, and Yinghui Wu in ICDE, 2022.

9. *"Diversified Subgraph Query Generation with Group Fairness"*, **Hanchao Ma**, Sheng Guan, Christopher Toomey$^{\#}$, and Yinghui Wu in WSDM, 2022.

10. *"GEDet: detecting erroneous nodes with a few examples"*, Sheng Guan, **Hanchao Ma**$^{+}$, Sutanay Choudhury, and Yinghui Wu in VLDB, Demo Track, 2021.

11. *"Explaining Missing Data in Graphs: A Constraint-based Approach"*, Qi Song, Peng Lin, **Hanchao Ma**, and Yinghui Wu in ICDE, 2021

12. "GRIP: Constraint-based Explanation of Missing Entities in Graph Search", Qi Song, Peng Lin, **Hanchao** $^{+}$, and Yinghui Wu in SIGMOD, Demo Track, 2021

13. *"GEDet: Adversarially Learned Few-shot Detection of Erroneous Nodes in Graphs"*, Sheng Guan, Peng Lin, **Hanchao Ma**$^{+}$, and Yinghui Wu in IEEE BigData, 2020. (**Best Paper Award**)

14. *"Attribute-Driven Backbone Discovery"* Sheng Guan, **Hanchao Ma**, and Yinghui Wu in KDD, 2019.

15. *"Ontology-based Entity Matching in Attributed Graphs"* **Hanchao Ma**, Morteza Alipourlangouri, Yinghui Wu, Fei Chiang, Jiaxing Pi in VLDB, 2019.

Refereed Journal Articles

1. *"An efficient ride-sharing framework for maximizing shared route"* Na Ta, Guoliang Li, Tianyu Zhao, Jianhua Feng, **Hanchao Ma**, Zhiguo Gong in TKDE Journal, 2017.

Proposals:

1. Crowdsourced Materials Data Engine for Unpublished XRD Results, NSF:2104007, 2021-2025

Students Research Mentoring

*Computer and Data Science Department at **Case Western Reserve University***

Research Projects

1. **Christopher Toomey**, CS Undergraduate Student
(1 research paper in WSDM, 2022, Next Role: Senior Data Scientist@Avison Young)

2. **Yen-shuo Chang**, CS MS Student
(One Research Paper in ICDE 2022)

3. **Yiyang Bian**, CS MS Student
(1 Research Paper in CIKM 2022, Next Role: Ph.D. Student @UCR)

4. **Mengying Wang**, CS Ph.D. Student
(Software Development in CRUX Team, 2 papers in ICDE 2022,2023. Next Role: Applied Scientist Intern@Amazon)

CRUX: A Crowdsourced Material Science XRD Data Assets Management System.[1]

1. **Nikki D'Costa**, DS Undergraduate Student
   (Software Development in the CRUX Team)
2. **My Le**, CS Undergraduate Student
   (Software Development in the CRUX Team)
3. **Khanh Khuat**, CS Undergraduate Student
   (Software Development in the CRUX Team)
4. **Shrinidhi Hegde**, CS MS Student
   (Software Development in CRUX Project, Next Role: Software Developer@MultiCASE, Inc.)
5. **Abhishek Daundkar**, Material Science Ph.D. Student
   (Software Development in CRUX Team)

Teaching Experiences

Case Western Reserve University

1. CSDS 395 *"Senior Project"*, Teaching Assistant, serving approx. 30 students Spring 2021
2. CSDS 234 "Structured and Unstructured Data", TA&Lecture, serving approx. 30 students Fall 2020
3. CSDS 234 "Database Systems", Teaching Assistant, serving approx. 30 students Spring 2020
4. CSDS 234 "Structured and Unstructured Data", TA&Lecture, serving approx. 30 students Fall 2019

Washington State University

1. CPTS 317 "Formal Language and Automata", TA&Lecture, serving approx. 80 students Spring 2019
2. CPTS 317 "Formal Language and Automata", TA&Lecture, serving approx. 80 students Spring 2018
3. CPTS 516,350 "Algorithm", Teaching Assistant, serving approx. 90 students Spring 2018
4. CPTS 415 "Big Data", Teaching Assistant, serving approx. 40 students Fall 2018
5. CPTS 415 "Big Data", Teaching Assistant, serving approx. 40 students Fall 2017

Professional Service

Conference Reviewer/External Reviewer

SIGMOD 2020; VLDB 2020; ICDE 2020, 2021, 2022; KDD 2019; SIGIR 2023

Invited Journal Reviewer

Journal of Data and Information Quality 2025

Journal of Big Data 2025

Conference Volunteer Service

ICDE 2022, 2023; CIKM 2022; WSDM 2020; VLDB 2019

[1] https://cruxproject.org/