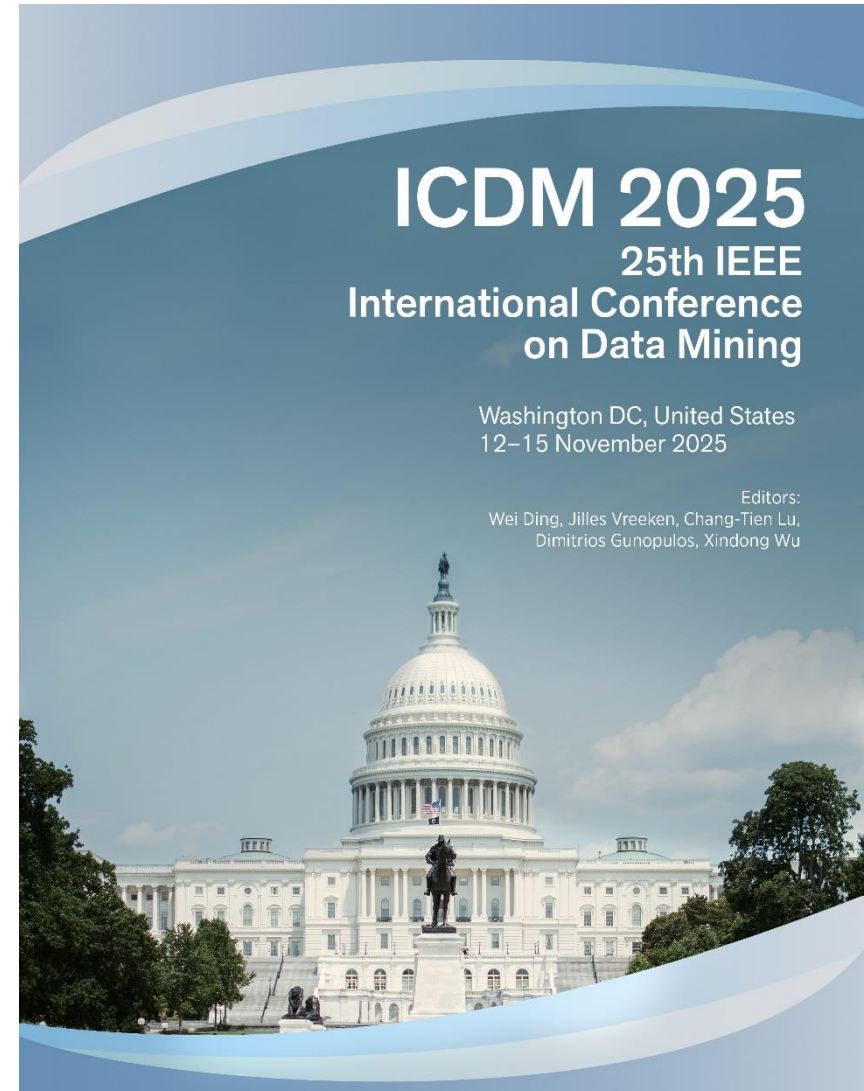




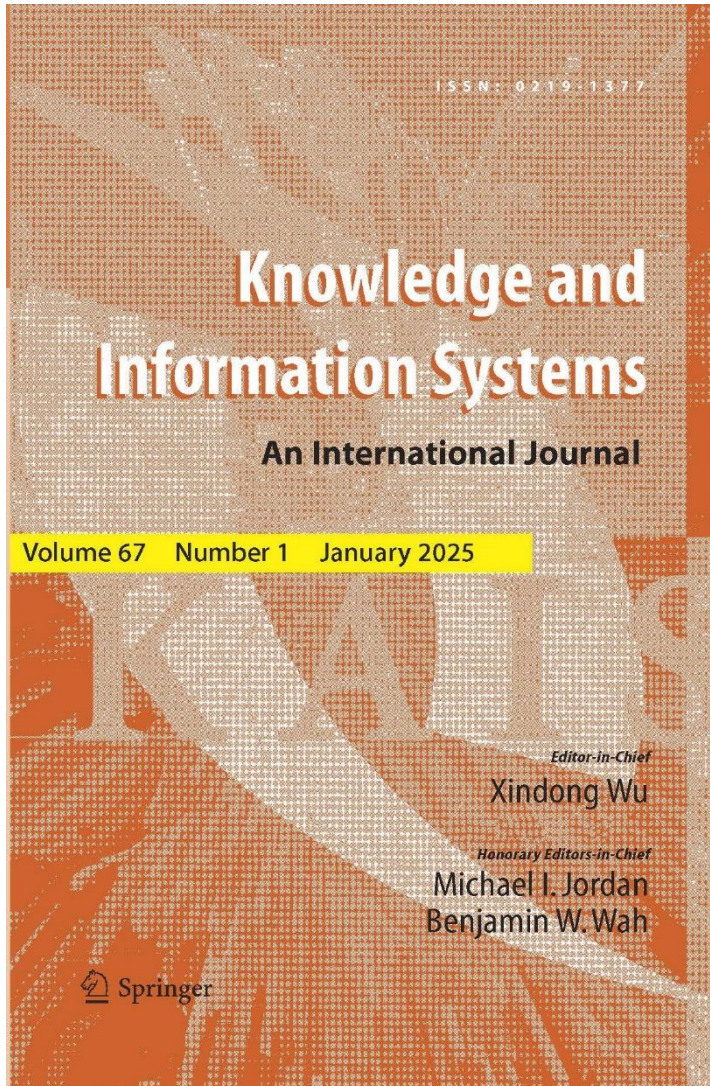
ICDM '25 Award Ceremony

Steering Committee Chair:
Xindong Wu

1. ICDM Service Award
2. ICDM Research Award
3. Best Paper Awards (5)
4. 10-Year ICDM Impact Award
5. Tao Li Award
6. KAIS Journal Best Paper Award
7. ICDM Female Scholar Awards



Major ICDM Activities in the Past 25 Years



- ① KAIS journal of the best papers every year since 2001 (**the 2024 Impact Factor: 3.1; h-5: 73 by Google Scholar; both top 12!**)
- ② IEEE ICDM Research Contributions Award and IEEE ICDM Outstanding Service Award since 2001(**Nobel Prizes in Data Mining**)
- ③ Data mining on ICDM submissions 2003 ~ 2015 (**ICDM rules**)
- ④ A panel discussion on "top 10 algorithms in data mining" in 2006 (**Google citations: 8169** as of 11/14/25)
- ⑤ **Workshops proceedings** by IEEE CPS since 2006
- ⑥ Double blind reviewing since 2007 (**and triple blind since 2011**)
- ⑦ IEEE ICDM **10-Year Highest-Impact Paper Award** since 2010
- ⑧ Tao Li Award since 2018
- ⑨ ICDM Female Scholar Awards since 2023



2025
ICDM Steering Committee Awards
(each with a check of \$2,500)

- **IEEE ICDM Outstanding Service Award**
 - highest recognition for service achievements in Data Mining.
- **IEEE ICDM Research Contributions Award**
 - the highest recognition for research achievements in Data Mining.



2025 IEEE ICDM

Nomination and Evaluation Committees

- ◆ Takashi Washio (Co-Chair), Kansai University, Japan
- ◆ Xindong Wu (Co-Chair), Hefei University of Technology, China
- ◆ Diane Cook, Washington State University, USA
- ◆ Peter Flach, University of Bristol, UK
- ◆ Vipin Kumar, University of Minnesota, USA
- ◆ Jian Pei, Simon Fraser University, Canada
- ◆ Claudia Plant, University of Vienna, Austria
- ◆ Jie Tang, Tsinghua University, China
- ◆ Geoff Webb, Monash University, Australia
- ◆ Xingquan Hill Zhu, Florida Atlantic University, USA



2025 IEEE ICDM Outstanding Service Award

Dr. Charu Aggarwal

IBM T.J. Watson Research Center,
USA

- ◆ IEEE ICDM Steering Committee
- ◆ Program Co-Chair for ICDM 2015
- ◆ Editor-in-Chief, TKDD (2017-2024),
ACM SIGKDD Explorations (2014-2018)

A talk by Charu at the banquet...





IBM Research AI

IEEE ICDM Outstanding Service Award Presentation

Charu C. Aggarwal

Acknowledgements

- A lot of people have contributed to my award directly or indirectly
- I would like to thank...
 - **Research collaborators:** Jiawei Han, Philip Yu, Jiliang Tang, Suhang Wang, Karthik Subbian, Chandan Reddy, Mahsa Salehi, and many more
 - **Service collaborators:** Xindong Wu (TKDD, ICDM, ICBK), Zhi-Hua Zhou (ICDM), Hanghang Tong (Explorations), Haixun wang (Explorations), Balaji Krishnapuram (KDD), Alex Smola (KDD), Jian Pei (KDD) and many more
 - **Family:** My wife and daughter for their support

My Significant Service Memories

- First organizing committee participation: Sponsorship Chair at KDD 2001
- Memorable conferences chaired: IEEE ICDM, IEEE ICBK, IEEE Big Data, KDD, CIKM
- Memorable journal editorships: SIGKDD Explorations, ACM TKDD, IEEE TKDE, DMKD, IEEE Trans. Big Data
- Committees: IEEE ICDM Steering Committee, SIAM Industry Committee, Vice Chair of SIAG on Data Mining

Historical Perspective

- The importance of services has increased significantly with the expansion of data-centric conferences
 - In 2001 (first ICDM), the number of submissions were a small fraction of what they are today
 - Same is true for most data-centric conferences
 - Conferences have difficulty in recruiting researchers in service roles
 - Some conferences even conscript authors as reviewers on pain of conference rejection

Benefits of Voluntary Participation

- Voluntary participation has benefits both for the participant and the conference/journal
- The quality of services is higher with eager reviewers, editors, and organizing committee members
- Staying involved in the review process helps provide perspective on which areas are gradually becoming more important
- Services roles also help junior researchers getting plugged into the research community

Thank You!





2025 IEEE ICDM Research Contributions Award

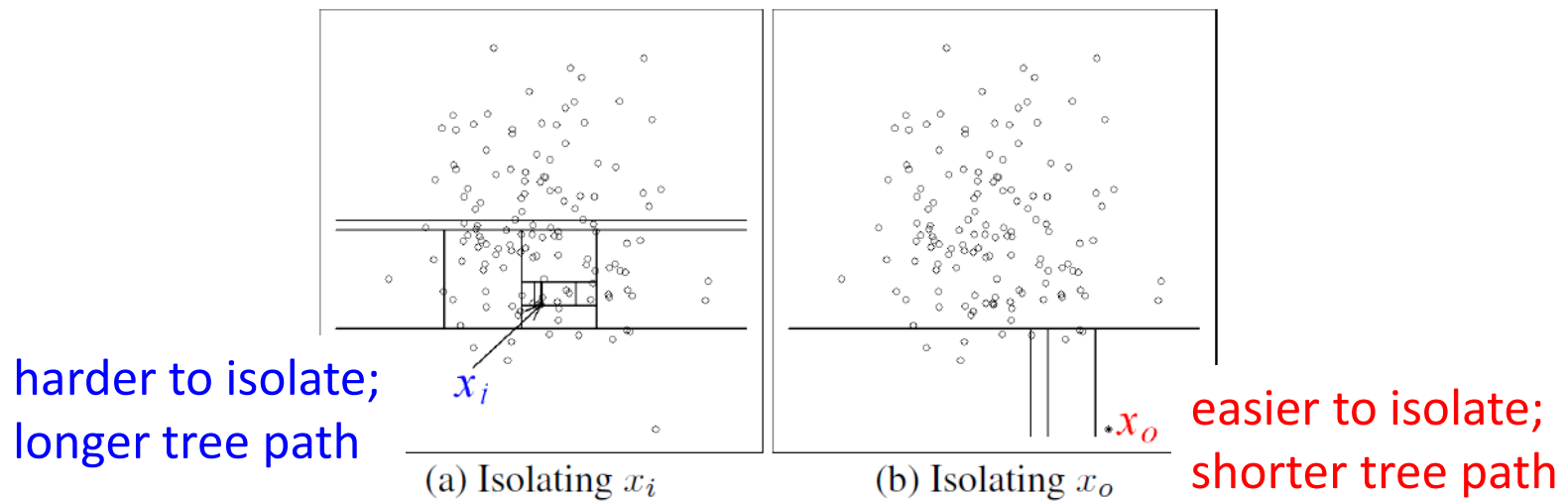
Professor Zhi-Hua Zhou,
F'IEEE, F'ACM, F'AAAI, F'AAAS
Nanjing University, China

- Isolation Forest (ICDM, 2018), 8800+ Google Scholar citations, Journal version (TKDD, 2012), another 2500+ citations.
- Seminal and foundational contributions to ensemble learning. His contributions have helped to define the research scope of ensemble learning and deepened our understanding of the field.
- Pioneering contribution to what is now known as "knowledge distillation," a technique that is fundamental to the development of large language models and other big models.



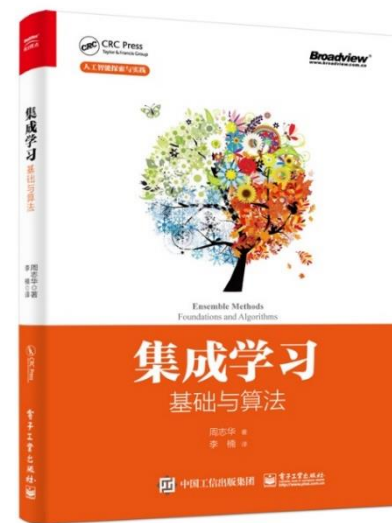
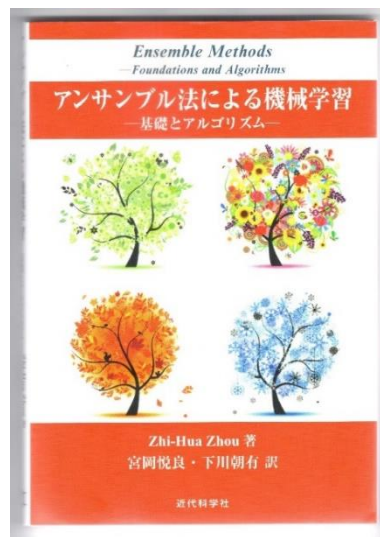
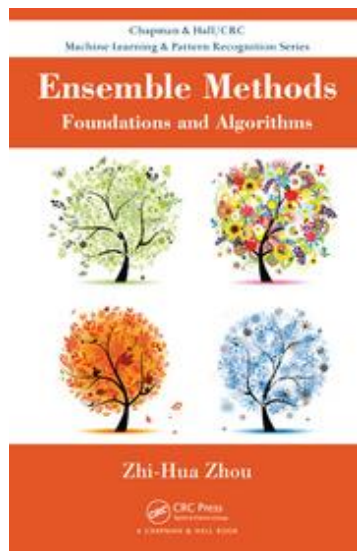
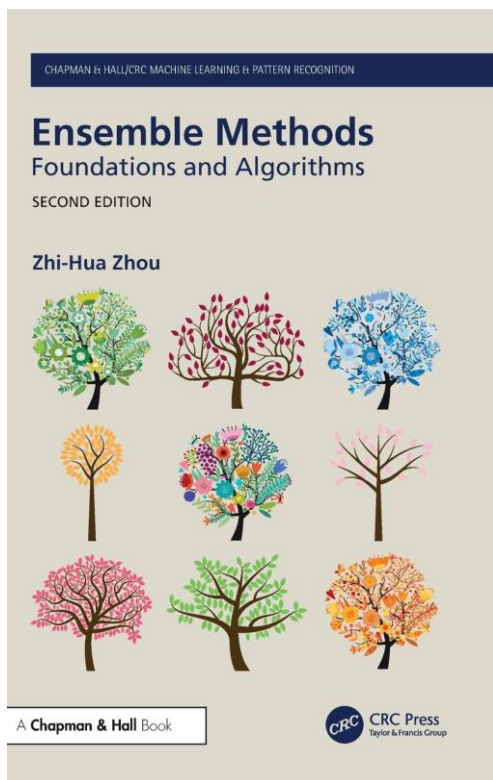
In his **2008 IEEE ICDM** paper, **Isolation Forest**, he proposed the "iForest" algorithm, which is now widely used for anomaly detection in industry. The impact of this work is demonstrated by its over 11,300 citations: the original conference paper has been cited more than 8,800 times, and its subsequent journal version in ACM TKDD (2012) has received over 2,500 citations.

- Very efficient
- Based on **ensemble of completely-random trees**
- Based on **isolation** rather than distance/density



[Liu, Ting, Zhou, ICDM'08; TKDD 2012]

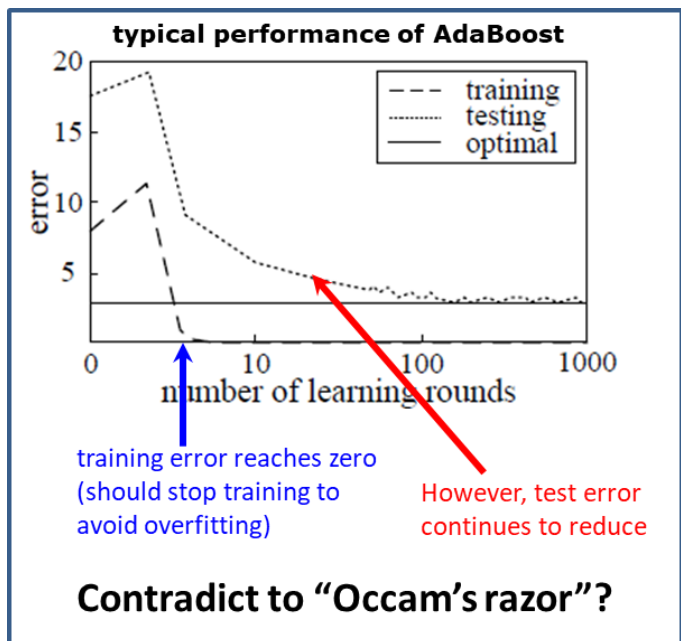
He has made seminal and foundational contributions to ensemble learning, often called the "key to winning" in data mining tasks. His influential work spans virtually all aspects of the field, encompassing theory, algorithms, and applications, as comprehensively detailed in his well-known book, Ensemble Methods: Foundations and Algorithms (1st ed., 2012; 2nd ed., 2025).



1st edition and its Japanese and Chinese versions

His contributions have helped to define the research scope of ensemble learning and deepened our understanding of the field. For example, his 2013 paper in the AI journal, On the Doubt about Margin Explanation of Boosting, solved the long-standing mystery of why the "top-ten" algorithm **AdaBoost** appears resistant to overfitting. This work inspired the development of new algorithms that optimize margin distributions rather than single margins.

The most fascinating fundamental theoretical issue about AdaBoost:
Why AdaBoost seems resistant to overfitting ?



His work solved this holy grail problem. By rigorously proving a new theory, he disclosed that, behind the mystery is the "**margin distribution**"; that is, even after the training error of AdaBoost reaches zero, it is still able to improve the margin distribution via further training, leading to even stronger generalization ability. His theory also revealed that AdaBoost will finally overfit when margin distribution cannot be improved further, though very late. This inspired the born of a new family of learning algorithms that optimize margin distributions rather than single margins.

As early as 2004, he published a paper in IEEE TKDE titled NeC4.5: Neural Ensemble based C4.5. In this work, he proposed an approach where one model is used to generate pseudo-data to train a second, simpler, yet more powerful model. This represents a **pioneering contribution to what is now known as "knowledge distillation"**, a technique that is fundamental to the development of large language models and other big models today.

In this paper, he proposed an approach:

Generate input-output pairs from a complicated learned model A to obtain a set of pseudo-data, and then use the pseudo-data to train a model B; the B model can be simpler, more explainable, and even with stronger generalization ability than model A

We also provided theoretical explanation to why it works



ICDM '25 Awards Committee

- **Committee Chair: Hai-Ning Liang, Chair**
<hainingliang@hkust-gz.edu.cn>
 - 2024 Impact Paper Award Winner
- **Committee Members:**
 - Diane Cook <djcook@wsu.edu>
 - Wei Ding <Wei.Ding@umb.edu>
 - Jilles Vreeken <vreeken@cispa.de>
 - Geoff Webb <geoff.webb@monash.edu>
 - Xindong Wu <xwu@hfut.edu.cn>
 - Lina Yao <lina.yao@unsw.edu.au>



ICDM '25 Best Paper Awards

- **ICDM '25 Best Paper Award (\$1K from TCII):**
 - Attributed Hypergraph Generation with Realistic Interplay Between Structure and Attributes
 - by Jaewan Chun <jjwpalace@kaist.ac.kr>, Seokbum Yoon, Minyoung Choe, Geon Lee, Kijung Shin
- **Best Paper Runner-Up Award (each with \$500 from TCII):**
 - DP-FedLoRA: Privacy-Enhanced Federated Fine-Tuning for On-Device Large Language Models
 - by Honghui Xu <hxu10@kennesaw.edu>, Shiva Shrestha, Wei Chen, Zhiyuan Li, Zhipeng Cai
 - Test-Time Graph Rebirth For GNN Generalization Under Distribution Shifts
 - by Xin Zheng, Bo Li, Yu, Qin Zhang, Haishuai Wang, Yuxuan Liang, Alan Wee-Chung Liew, Shirui Pan <s.pan@griffith.edu.au>
- **Best Student Paper Award (\$1K from TCII):**
 - HyHG: A Temporal Hypergraph Contrastive Learning Framework for Biomedical Hypothesis Generation
 - by Amir Shariatmadari <ahs5ce@virginia.edu>, Sikun Guo, Nathan Sheffield, Aidong Zhang, Kishlay Jha
- **Best Student Paper Runner-Up Award (\$500 from TCII):**
 - Zero-Shot Cross-City Trajectory Prediction Using Hypernetworks
 - by Jonas Gunkel <jonas.gunkel@dlr.de>, Andrea Tundis, Max Mühlhäuser



10 Year High Impact Award

(with a check of \$1,000)

- Selected by the Awards Committee
 - Considered some of the most cited papers that appeared in ICDM 2016.
 - A short list of papers were selected and ranked by members of the awards committee.
- The paper that stood out:
 - **Fusing Similarity Models with Markov Chains for Sparse Sequential Recommendation**
 - by Ruining He and Julian McAuley <jmcauley@ucsd.edu>
 - ICDM 2016, pp. 191-200

A talk in another session ...

10-Year Highest-Impact Paper Award -- IEEE ICDM 2025

Fusing Similarity Models with Markov Chains for **Sparse Sequential** Recommendation

Ruining He (now Research Engineer@Google DeepMind)
Julian McAuley (Prof. @UCSD)



UC San Diego

Thank you for this incredible honor!



Awards Committee



Prof. Hai-Ning Liang, Chair



Prof. Julian McAuley

Co-author and my amazing PhD advisor



Back to ~10 years ago ...



- Many thanks to Julian, who introduced me to the data mining & recommender systems field
- Worked on some **novel problems** like personalized *fashion* preference & *fashion trends* modeling
- How about **novel methods** for existing problems?

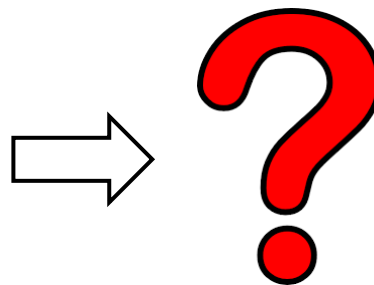
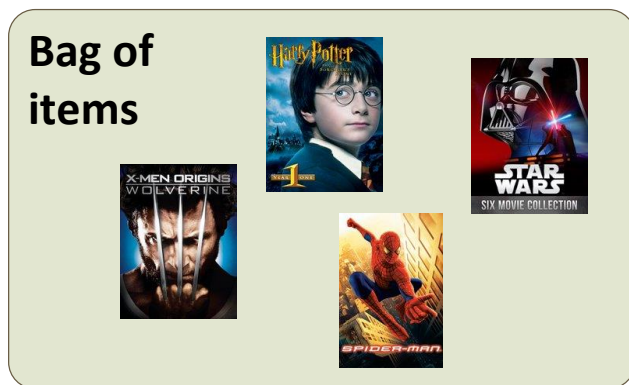
Back to ~10 years ago ...

- **Background**
 - *Seq.-to-seq. learning with neural networks* was relatively new
 - Recommender Systems was not a hot topic (yet)
 - Deep Learning for Recommender Systems on the horizon
- **Problem:** Personalized **sparse sequential** recommendation (cold-start)
- **How:** Long-term P13N + Short-term seq. Dynamics



Back to ~10 years ago ...

- Existing Methods
 - Collaborative Filtering Models (e.g., Matrix Factorization)

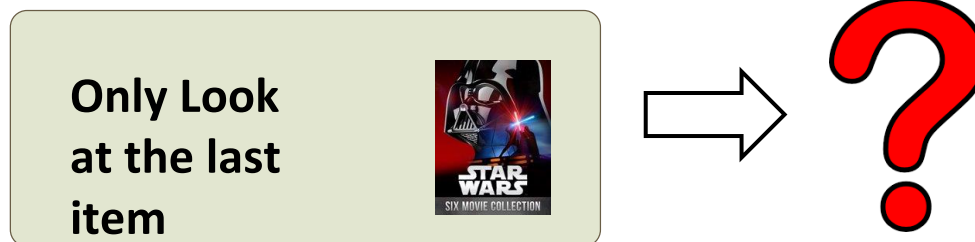


- Great at capturing a user's general, **long-term** taste
- But **sequential dynamics** (local contexts) are lost!



Back to ~10 years ago ...

- Existing Methods
 - Sequential (Markov) Models



- Captures local, **item-to-item** dynamics
- But user's **long-term taste** / P13N is lost!



- Existing Methods

-
- Diagram illustrating a matrix factorization process. The matrix is partitioned into three blocks: a green block (top-left), an orange block (top-right), and a green block (bottom-right). The orange block contains the values u_1 and u_2 (rows) and items 1, 2, 3, 4 (columns). The green blocks contain 0s and 1s. The text "from item" is on the left and "to item" is at the bottom.

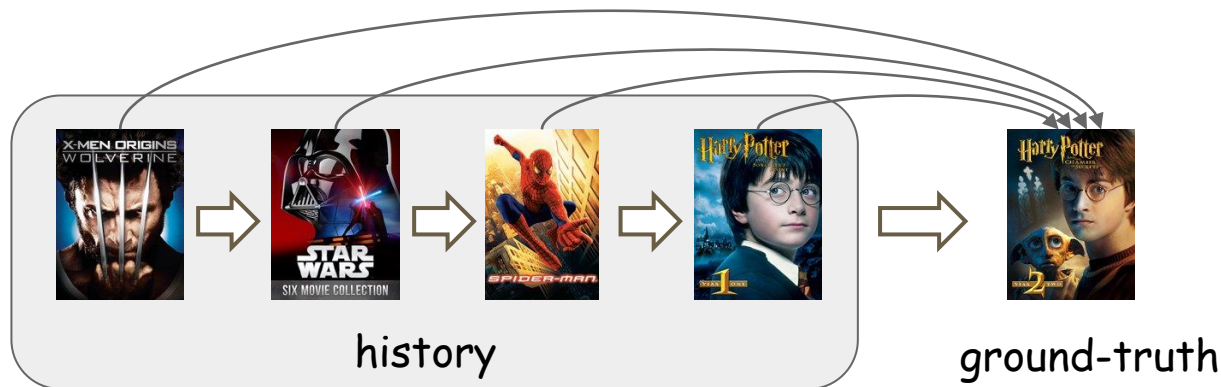
user embedding

- 

Back to ~10 years ago ...

- Our Insights

- A. Model each user as a sequence of items (*sequence modeling*)

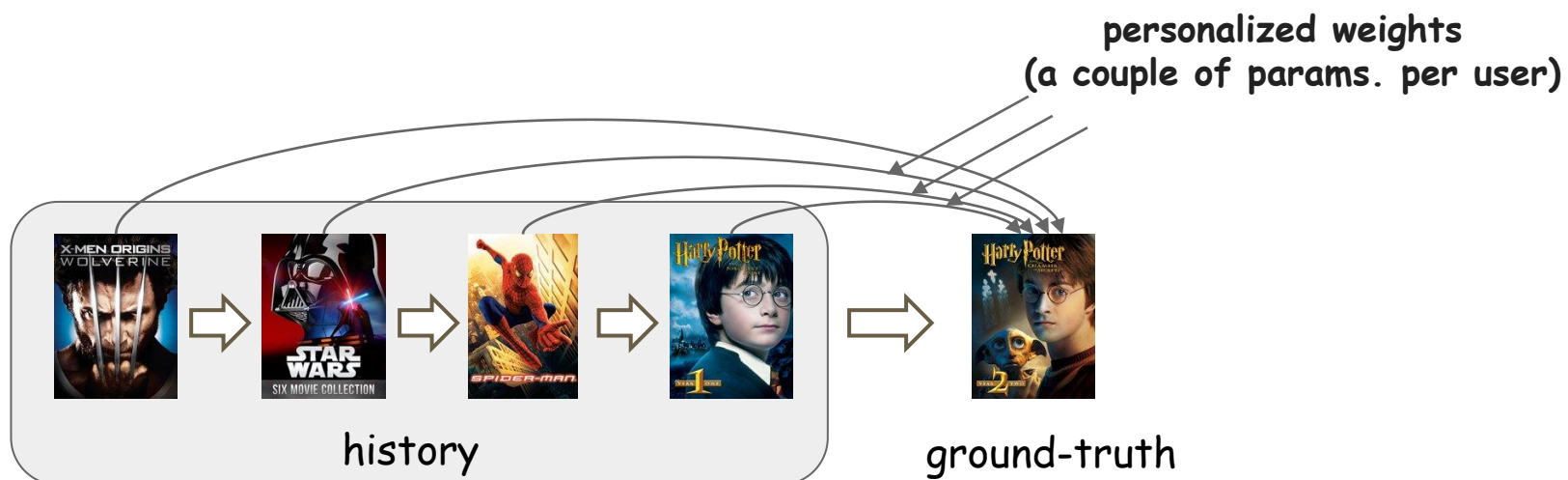


- $\text{Score}(\text{candidate}) = \sum \text{Pairwise item similarity}$ (subjected to normalization)
- Avoid learning user embedding, which doesn't handle sparsity well

Back to ~10 years ago ...

- Our Insights

- A. Model each user as a sequence of items (*sequence modeling*)
- B. **Personalized**, **high-order** Markov Chain modeling



Back to ~10 years ago ...



- **Our Key Results**

- State-of-the-art results on 9 sparse datasets (at publication time)



- A case study on MovieLens dataset



the sparser data → the larger improvement!

Impact over time

- Based on Gemini Deep Research (sampled 100 citations, Oct. 2025)

Time Period	%Direct Baseline	%Indirect Baseline	% Related Work	#Citations (in 100 samples)
Early Wave (2017-2019)	62%	8%	30%	29
Peak Influence (2020-2022)	33%	26%	41%	46
Neural Era (2023-)	4%	24%	72%	25

"Neural
Wave" →



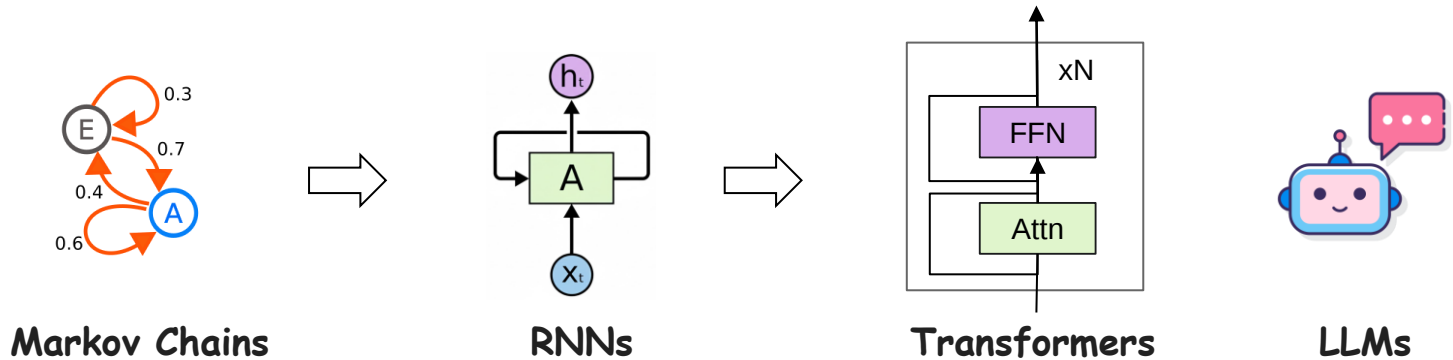
UC San Diego

Reflection: what we did right

- **Important Problem**
 - Cold-start, still a key challenge for modern Recommender Systems
- **Method on the right direction**
 - Model user as a sequence of items ("handcrafted" sequence modeling)
- **Due Diligence**
 - Reproducibility: open sourced data & code
 - Readability

Looking Ahead

- Tools have changed



- Fundamental problems still remains ...
 - Cold-start
 - Balancing long-term preference with short-term intent

Thanks again!

*Fusing Similarity Models with Markov Chains
for **Sparse Sequential** Recommendation, ICDM' 16.*



Long-term P13N



Short-term Seq.
dynamics

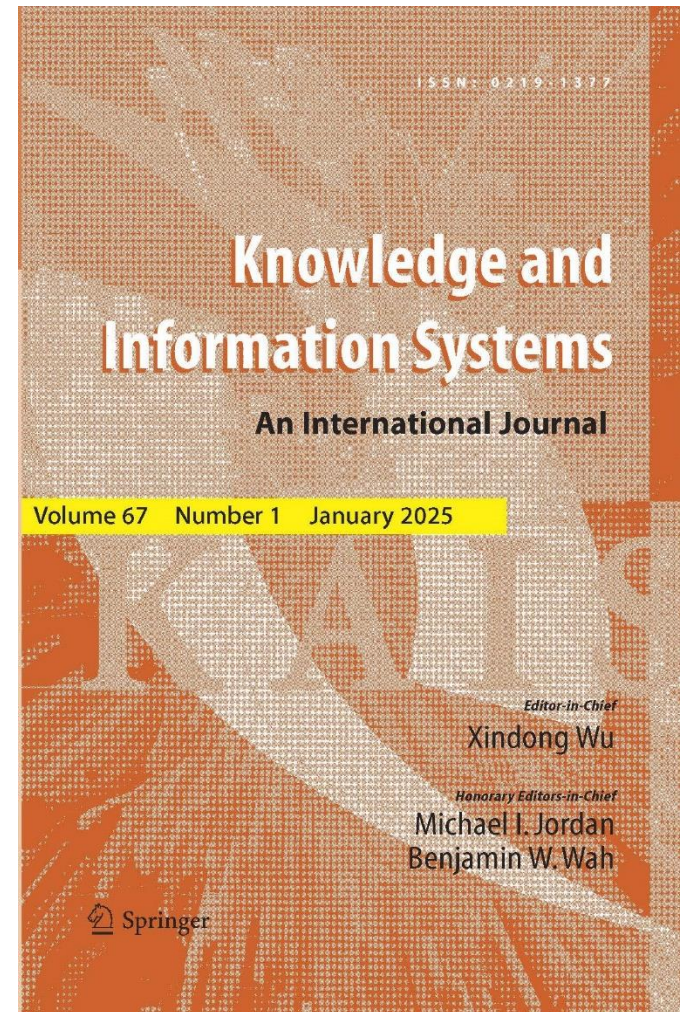


2025 KAIS Journal Best Paper Award

(A check of \$1K from Springer)

Best Paper published in
2022:

- Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond
- by Xuhong Li, Haoyi Xiong, Xingjian Li, Xuanyu Wu, Xiao Zhang, Ji Liu, Jiang Bian, Dejing Dou <dou@cs.uoregon.edu>
- *Knowledge and Information Systems*, Volume 64, Issue 12, Pages 3197-3234





2025 Tao Li Award

(a \$2K check from Tao Li Foundation)

Was selected by a committee

- Tao Li Award Committee Chair:
Xingquan Zhu

2025 Tao Li Award winner:

- Jundong Li jl6qk@virginia.edu

A short talk in another session ...



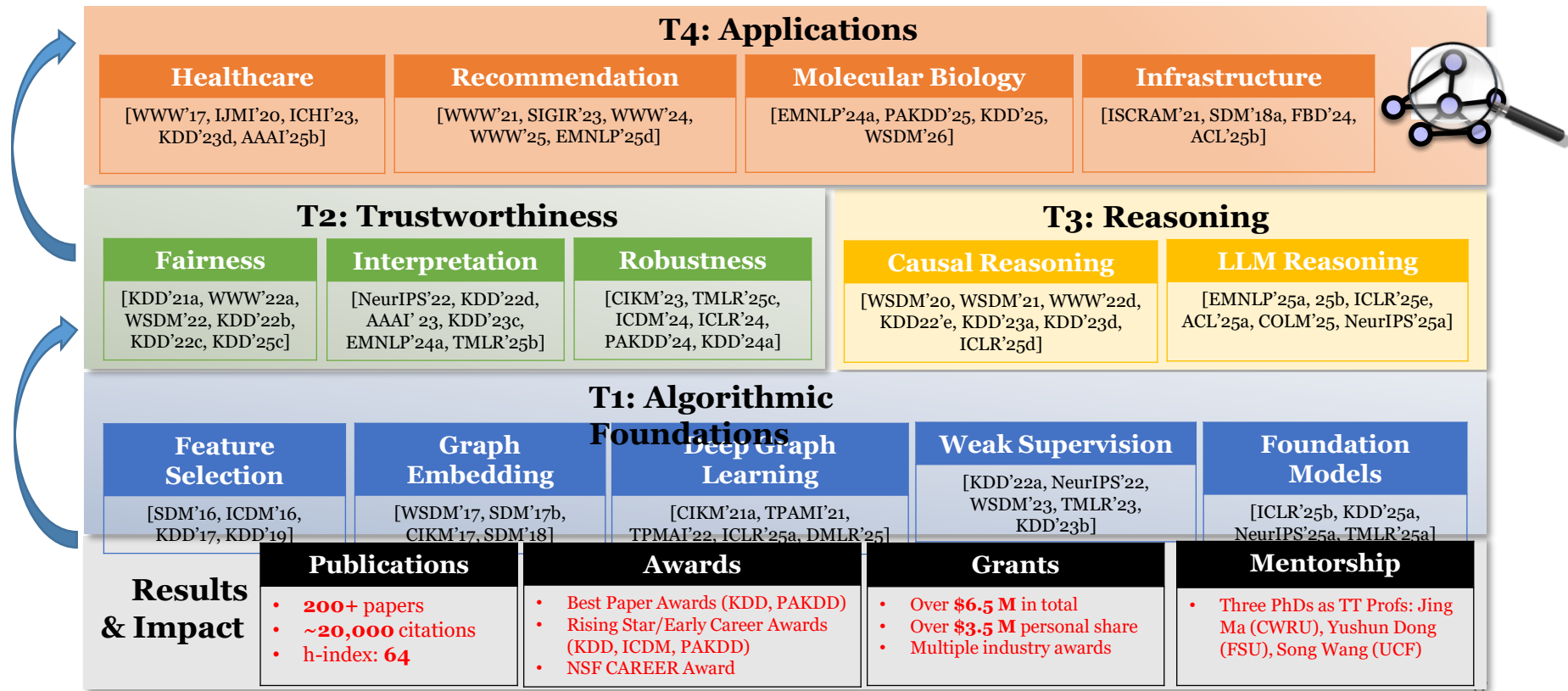


ICDM 2025 Tao Li Award Presentation

Jundong Li, Associate Professor

University of Virginia
<https://jundongli.github.io>
jundong@virginia.edu

Research Theme: Learning from Structured Data



Learning from Structured Data – Foundations (T1)

1. Well-cited CSUR paper (**over 4,000 citations**)
2. *Scikit-feature* (over 1.5k Github stars)
3. “Five ML projects you can no longer overlook” – featured by KDnuggets

- Well-cited papers:
1. LANE (WSDM’17)
~600 citations
 2. DANE (CIKM’17)
~500 citations
 3. AANE (SDM’17)
~400 citations

1. NSF III Small award (as PI)
2. JP. Morgan Chase Faculty Research Award
3. Cisco Faculty Research Award

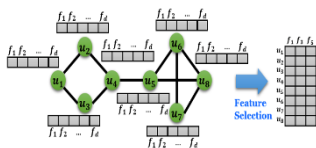
1. Tutorials in KDD’23, SDM’23, and WWW’25
2. Workshops in KDD’23, KDD’24, and WWW’25

Highlights

Feature Selection

Unsupervised FS
FS on graphs
Personalized FS

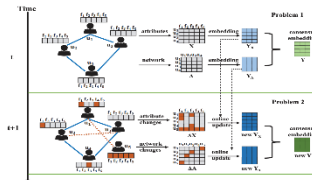
2017



2019

Graph Embedding

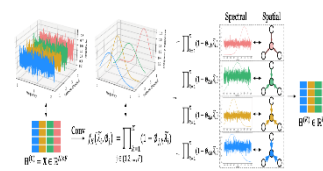
Attributed graphs
Dynamic graphs
Multi-layered graphs



Deep Graph Learning

Graph operators
Over-smoothing
Robust GNNs

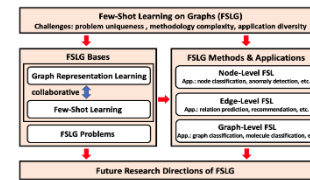
2021



2023

Weak Supervision

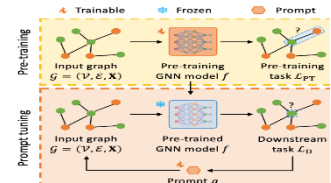
Few-shot learning
Self-supervised
Extreme-weak signals



Foundation Models

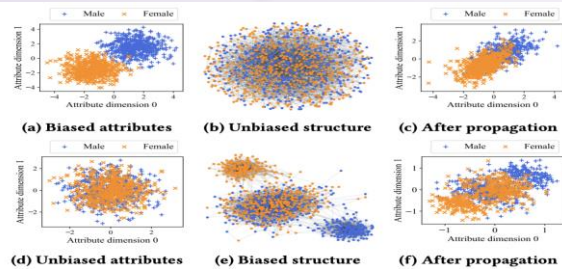
Graph prompting
Efficient Fine-tuning
Adaptation

2025



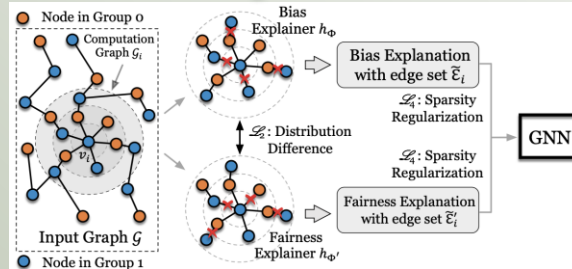
Learning from Structured Data – Trustworthiness (T2)

Fairness and Bias Mitigation



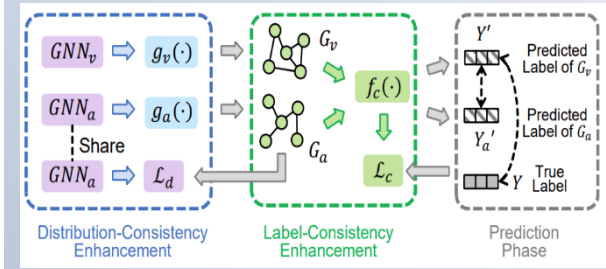
Mitigating Graph Data Bias
Improving Fairness
Understanding Bias

Explanation and Interpretation



Concept-based Explanation
Counterfactual Explanation
Explanation Usability

Robustness and Generalization



Adversarial Attacks
Certified Robustness
OOD Generalization

Highlights:

Open-source package: *PyGDebias*

PyGDebias

STEERING ELECTRONS OUT OF THE DRIFT WITH DEEP LEARNING



The beamline on Jefferson Lab's Continuous Electron Beam Accelerator Facility is monitored in the Machine Control Center during a Jefferson Lab campus tour. (Jefferson Lab photo/Aileen Devlin)

Machine learning project aims to improve accelerator performance by streamlining beam-tuning tasks.

NEWPORT NEWS, VA – When racing through the streets of Tokyo – on a sanctioned road course, that is – a little drift can be a good thing. But inside the primary particle accelerator at the U.S. Department of Energy's Thomas Jefferson National Accelerator Facility, even the tiniest drift of the electron beam can bring scientific research to a screeching halt. It also can make operators in the accelerator's control room furious. Fast.

"Beam tuning can be a maddening process," said Chris Townsend, a Jefferson Lab staff scientist in the Center for Advanced Studies of Accelerators. "It's very iterative. It is brute force and a lot of trial and error. It is turning knobs and seeing what effect that has."

The Continuous Electron Beam Accelerator Facility, a world-class DOE User Facility, is a 16-mile particle "racetrack" that relies upon more than

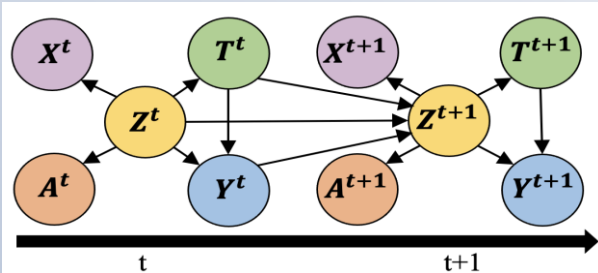
Deployed by Jefferson Lab in the operation of particle accelerators

Best Paper Award on "Concept-based Explanation"

May 7-10 2024
PAKDD

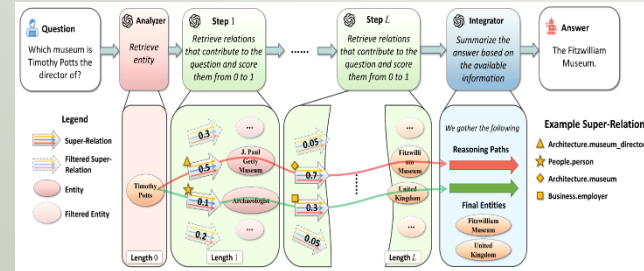
Learning from Structured Data – Reasoning (T3)

Causal Reasoning



Hidden Confounders
Interference Effects
Treatment Entanglement

LLM Reasoning



KG-enhanced Reasoning
Efficient Reasoning
Multi-agent Communication

Highlights:

Best Research Paper Award on
"High-order Interference"



AAAI New Faculty
Highlights



NSF CAREER Award



Deployed in domains of
public policy, healthcare,
and finance

Learning from Structured Data – Applications (T4)

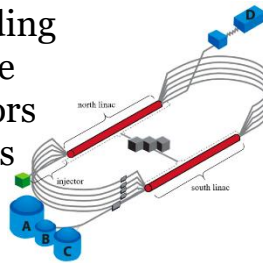
- **Healthcare:** chronic kidney disease (UVA Medicine), mental disorders (UVA SDS), hospital-acquired infections (UVA Biocomplexity)
- **Recommendation:** job marketplace (LinkedIn), social recommendation (Snap), conversational recommendation (Netflix)
- **Molecular Biology:** gene-gene interactions (UTHealth Houston), protein mutation effects/molecular dynamics (UVA Chemical Engineering)
- **Infrastructure Systems:** particle accelerators (Jefferson Lab), electrical buses (GMU), well water (Northeastern)

Highlights:

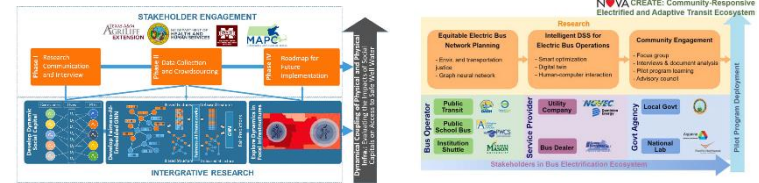
Multiple industry research awards



DOE funding on particle accelerators operations

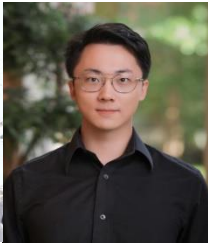


NSF SAI and SCC awards on infrastructure systems



Acknowledgements

- Thanks to the ICDM 2025 Tao Li Award selection committee
- Thanks to my nominator (Prof. Aidong Zhang), PhD advisor (Prof. Huan Liu), mentors, students, collaborators, and family
- Thanks to the funding support from UVA, NSF, ONR, DOE and industry partners





ICDM 2025 Women Award Committee

Selected by the Award Committee

- ◆ Senjuti Basu Roy (Chair), New Jersey Institute of Technology
- ◆ Wei Ding, U. of Massachusetts Boston
- ◆ Fusheng Wang, Stony Brook University
- ◆ Chang-Tien Lu, Virginia Tech
- ◆ Xindong Wu, Hefei University of Technology, China
- ◆ Linsey Pang, PayPal, USA
- ◆ Ling Chen, University of Technology Sydney



2025 ICDM Young Female Scholar Award in Data Science (a check of \$1K each)

- ICDM Young Female Scholar Award in Data Science
 - Xuan Wang xuanw@vt.edu
 - Li Zhang li.zhang@utrgv.edu
- ICDM Female Student Award in Data Science
 - Jinghan Zhang jinghaz@clemson.edu
 - Xinwen Zhang ellenz@temple.edu



Thanks!

IEEE International Conference on Data Mining

The world's premier research conference in data mining