

JINYANG LI

Computer Science and Engineering, University of Michigan, Michigan, USA
jinyli@umich.edu | <https://JinyangLi01.github.io>

EDUCATION

University of Michigan, Ann Arbor

Ph.D. candidate, Computer Science and Engineering

Advisor: [H. V. Jagadish](#)

Michigan, USA

Sep. 2020 – Present

Peking University

Bachelor of Science, Computer Science and Technology

Beijing, China

Sep. 2016 – Jun. 2020

RESEARCH INTEREST

- Query processing and optimization.
- Responsible data management: fairness, bias, and diversity issues.

PUBLICATIONS

1. **Jinyang Li**, Yuval Moskovitch, Julia Stoyanovich, H. V. Jagadish
Query Refinement for Diversity Constraint Satisfaction
Under review
2. **Jinyang Li**, Alon Silberstein, Yuval Moskovitch, Julia Stoyanovich, H. V. Jagadish
ERICA: Query Refinement for Diversity Constraint Satisfaction
VLDB demo 2023
3. Yuval Moskovitch, **Jinyang Li**, H. V. Jagadish
Dexer: Detecting and Explaining Biased Representation in Ranking
SIGMOD Demo 2023
4. **Jinyang Li**, Yuval Moskovitch, H. V. Jagadish
Detection of Groups with Biased Representation in Ranking
ICDE 2023
5. Yuval Moskovitch, **Jinyang Li**, H. V. Jagadish
Bias analysis and mitigation in data-driven tools using provenance
Proceedings of the 14th International Workshop on the Theory and Practice of Provenance, 2022
[PDF](#)
6. **Jinyang Li**, Yuval Moskovitch, H. V. Jagadish
DENOUNCER: detection of unfairness in classifiers
VLDB Demo 2021
[PDF](#)
7. Yinda Zhang, **Jinyang Li**, Yutian Lei, Tong Yang, Zhetao Li, Gong Zhang, Bin Cui
On-Off Sketch: A Fast and Accurate Sketch on Persistence
VLDB 2021
[PDF](#)
8. Tong Yang, Haowei Zhang, **Jinyang Li**, Junzhi Gong, Steve Uhlig, Shigang Chen, Xiaoming Li,
HeavyKeeper: An Accurate Algorithm for Finding Top-k Elephant Flows
IEEE/ACM Transactions on Networking (ToN), 2019
[PDF](#)

RESEARCH EXPERIENCE

Research Assistant, Database Group, University of Michigan

Advisor: Professor H. V. Jagadish

Michigan, USA

Sep. 2020 – Present

- We study the problem of modifying relational queries to have the result satisfy constraints on the sizes of multiple subgroups in it, to improve diversity and group representation in query results. With the help of a

provenance model, we develop an efficient query refinement algorithm. Experimental results demonstrate the effectiveness and efficiency of our solution on a variety of datasets and constraints. (Pub. 1, 2)

- We study the problem of detecting groups with biased representation in the top- k ranked items, eliminating the need to pre-define protected groups. We propose efficient search algorithms for two different fairness measures: global bound representation, and proportional representation. We also propose a method to explain the bias in the representations of groups utilizing the notion of Shapley values. (Pub. 3, 4, 5, 6)

Research Assistant, FORWARD Data Lab, UIUC

Illinois, USA

Advisor: Professor Kevin Chen-Chuan Chang

Jun. 2019 – Jul. 2020

- We studied ordered access for relational data. We propose a theoretical optimality that takes the number of updates into consideration and theoretically proved the lower bound of change-aware order indexing.
- We designed a novel index structure realizing the theoretical optimality in an update-aware manner to support ordered access to RDBMS.

Research Assistant, Network Big Data Lab, Peking University

Beijing, China

Advisor: Professor Tong Yang

Jan. 2019 – Jun. 2019

- We worked on algorithms and data structures to find top- k elephant flows in network traffic measurement. The proposed algorithm incurs a small, constant processing overhead per packet. (pub. 8)
- We study the problem of the persistence of items -- whether an item appears recurrently in many time windows of a data stream. For persistence estimation, using the characteristic that the persistence of an item is increased periodically, we compress increments during the hashing process. To find persistent items, we separate persistent and non-persistent items, further improving the accuracy. (Pub. 7)

TECHNICAL SKILLS

Programming languages: Python, C, C++, GO, SQL

Data platforms: MongoDB, Apache Hadoop