Keming Li

 Personal Website | ✉ likm2020@mail.sustech.edu.cn | 📞 +86 181 3808 9583 ⬡ ghlkm

---

EDUCATION

Master of Computer Science and Engineering　　　　Aug. 2020 – July 2023 (expected)

Southern University of Science and Technology (SUSTech), GPA: 3.32/4.0　　　*Supervisor: Prof. Bo Tang*

Bachelor of Computer Science and Engineering　　　　Aug. 2016 – June 2020

Southern University of Science and Technology (SUSTech), GAP: 3.69/4.0　　　*Supervisor: Prof. Bo Tang*

- Key courses: Calculus (98), Linear Algebra (92), Discrete Mathematics (95), Probability and Statistics (98), Data Structures and Algorithms (87), Advanced Algorithms (92), Advanced Artificial Intelligence (93), Principles of Database Systems (92), Operating Systems (86)

- Bachelor thesis: Cover Ratio Maximization

RESEARCH INTEREST

Algorithms and systems for data analytics. Specifically, algorithms for information retrieval, computational geometry, and graph processing, and systems for data mining and graph data.

PUBLICATIONS

- Marrying Top-k with Skyline Queries: Relaxing the Preference Input while Producing Output of Controllable Size. *In SIGMOD'21.* Kyriakos Mouratidis, **Keming Li**, and Bo Tang.

- τ-LevelIndex: Towards Efficient Query Processing in Continuous Preference Space. *In SIGMOD'22.* Jiahao Zhang, Bo Tang, Man Lung Yiu, Xiao Yan, and **Keming Li**.

- Quantifying the Competitiveness of a Dataset in Relation to General Preferences. *Major Revision of VLDBJ.* Kyriakos Mouratidis, **Keming Li**, and Bo Tang.

- Dynamic Graph Storage: An Experimental Survey. *In submission to VLDB'23.* Xiangyu Zhi, Xiao Yan, **Keming Li**, Bo Tang, Yanchao Zhu, and Minqi Zhou.

- An Efficient System for Dynamic Path Pattern Matching with Shared Execution and Parallel Computation. *Targeted for SIGMOD'24.* **Keming Li**, and Bo Tang.

RESEARCH PROJECTS

Dynamic Pattern Matching System　　　　**Leader**　　　　Feb. 2022 – Present

- Given a data graph and many path patterns, the project aims to efficiently return positive matchings (new ones that become valid) and negative matchings (old ones that become invalid) when the data graph is updated with node/edge deletions/insertions. Targeted for SIGMOD'24.
- To speed up pattern matching, I designed a dynamic matching framework and a data structure called partial matching graph, which marks partial matching via dynamic programming.
- To allow different paths to share computation when updating partial matching and searching complete matching, I proposed algorithms to merge the paths into a shared path and store their partial matching in a shared data structure. The algorithms are proven to reduce complexity.
- To parallize the algorithms with multiple threads, I conducted a multi-threading system using AtomicCAS, memory order, thread fence, and volatile for correctness and efficiency.

A Survey for Graph Storage Systems　　　　**Core Member**　　　　Feb. 2022 – Present

- The project investigates 21 graph storage systems, tests their performance in various aspects, and summarizes guidelines for the design of graph storage systems. In submission for VLDB'23.
- I proposed to classify existing graph storage systems into 3 categories according to the basic data structures they use, and evaluated the performance of 4 systems for transaction management, concurrency control, multi-threading, persistency strategy, and space consumption.
- I helped summarizing key design considerations of graph storage systems and how existing techniques (e.g., hybrid storage and indexes) can be composed to meet different requirements.

τ-LevelIndex for Top-k Retrieval　　　　**Core Member**　　　　Feb. 2021 – Feb. 2022

- This project designs a general index for the efficient processing of continuous preference space top-k queries (a subject of computational geometry). <u>Accepted by SIGMOD'22</u> ([code](), [paper]()).
- For efficient index building, I proposed a partition-based approach, which progressively partitions the preference space and prunes possible top-k options, and conducted analysis to show that it reduces both computation and memory compared with our initial insertion-based approach.
- I implemented a prototype system adapting our index and gave examples of answering 3 representative top-k queries via our system, implemented these algorithms in C++, and achieved a speedup of 2 to 3 orders of magnitude over SOTA.
- An related work dubbed Spatial Heatmap, which divides the preference space into equal-sized boxes, benefits various top-k queries and provides insightful visualizations. I handled part of the algorithm design and conducted all implementations. This work received major revision of VLDBJ.

Marrying Top-k with Skyline   **Core Member**   **July 2020 – Feb. 2021**

- This project develops a dominance-based query operator ORD and a utility-based query operator ORU, which has desirable properties including personalization, controllable output size, and flexibility, to bridge top-k queries and skyline queries. <u>Accepted by SIGMOD'21</u> ([code](), [paper]()).
- I proposed the ρ-domination concept to prune input constraints to speed up the quadratic programming (QP) solver. To assert the ρ-domination relation between options, I used Gram-Schmidt Orthogonalization to generate the bounding box with distance ρ from a preference vector.
- I wrote 7K+ lines of C++ code to implement all algorithms, tested 3 QP solvers for use in our project, designed and run all experiments, and conducted the case studies in 2 months.
- I implemented a prototype system, which had a user-friendly front-end, answering ORD and ORU queries. It had been used on some universities' Big Data courses.

SKILLS

Programming Languages: proficient in C/C++ and Python, familiar with Java, Matlab, and SQL
Languages: Mandarin (native), Cantonese (native), and English (fluent)
Tools: Jetbrains (Clion, PyCharm, IntelliJ IDEA), Jupyter Notebooks, Git, and Linux

HONORS/AWARDS

- SUSTech CSE Outstanding Academic Achievement Award for Undergraduates (2020)
- SUSTech Zhiren College Outstanding Undergraduate Award (2020)
- SUSTech Artificial Intelligence CARP contest, fifth place (2019)
- SUSTech Artificial Intelligence IMP contest, third place (2019)
- Freshman Coding Award of Huawei Data Communication Product Line (2019)
- SUSTech Excellent Teaching Assistant Award (2018, 2019)
- University Merit Student Scholarship (2017, 2018, 2019)
- National Endeavor Scholarship (2017)

MISCELLANEOUS

SUSTech Thesis LaTeX Template   **Initiator and Maintainer**   **Sept. 2021 – present**

- This project provides official LaTeX thesis template for bachelor, master, and Ph.D. It uses Github Actions' CI and CD to automatically execute the software development workflows ([link]()). I made videos and slides as user guidelines, and keeps improving the template based on feedback.

SUSTech DBGroup   **Mentor of Undergrads**   **Sept. 2021 – June. 2022**

- I co-supervised a undergrad named [Lan Lu]() on her thesis on subgraph matching, who won Best 10 Graduates (2020) in SUSTech and became a Ph.D. student in University of Pennsylvania.
- I co-supervised a undergrad named [Wendi Zhou]() on her thesis on defining self-driving complexity, who became a master's student in New York University.

HQoS for Differentiated Service   **Intern at Huawei**   **July 2019 – Oct. 2019**

- The project maintains a routing system, which provides differentiated bandwidth for users according to their categories in both access and core networks. I led the code reconstruction with regular expressions, committed about 40K+ lines of C++ code, and won the coding award.

SUSTech CSE Courses   **Teaching Assistant**   **March 2018 – July 2019**

- Assisted in 4 department core courses including Computer Programming Fundamentals, C/C++ Programming, Data Structures and Algorithms, and Operating System.
- Maintained a course website by utilizing Stanfor Moss, DSL, vue and flask for these course.