Wonseok Shin

(+82) 10-5301-8475 | wsshin@theory.snu.ac.kr | gratus907.github.io | gratus907

Research Interests

- Randomized Algorithms, Mathematical and Numerical Optimization
- Scalable Algorithms and Methods for Data Mining and Data Management
- Graph Neural Networks, Multimodal Learning, Self and Semi-supervised Learning, Representation Learning
- Optimization, Quantization and Pruning of Neural Networks

Education

Seoul National University, Computer Theory and Application Lab — Seoul, Korea

M.S in Computer Science and Engineering — Sep.2022 - Aug.2024 (expected)

- Advisor : Prof. Kunsoo Park
- Research Topic: Practical Algorithms for Large Scale Graph Data
- Published 1 paper in VLDB as first author on cardinality estimation of subgraph matching
- Current GPA: 4.24 / 4.3

Seoul National University — Seoul, Korea

B.S. in Computer Science and Engineering, B.S in Mathematical Science (Double Major) — Mar.2018 - Aug.2022

- Graduated with **Summa Cum Laude** (3.91 / 4.3 GPA)
- Thesis: Adaptive Matching Order for Subgraph Matching Problem (Advisor : Prof. Kunsoo Park)
- Relevant Courseworks (CS) : Algorithms, Theory of Computation, Machine learning for Bioinformatics
- Relevant Courseworks (Math) : Mathematical and Numerical Optimization, Infinitely Large Neural Networks

Experience

AlgenDrug. Co. Ltd — Seoul, Korea

Research Internship — Mar.2022 - Jul.2022

- Conducted research on graph pattern mining for prediction of drug toxicity. [J2]
- Developed multimodal contrastive learning method to train graph neural networks on molecular property prediction [P1]

Seoul National University, Computer Theory and Application Lab — Seoul, Korea

Undergraduate Research Opportunity Program — Aug.2020 - Apr.2021

- Conducted research on matching orders for SOTA subgraph matching algorithm DAF [J1, C1]

Projects

Framework of Practical Algorithms for NP-hard Graph Problems — Seoul, Korea

SW Star Lab Project by IITP (Participated during Master's Study) — Sep.2022 - Aug.2024

- Developed algorithm for approximate subgraph counting in large graphs, outperforming existing sampling and GNN based methods by up to two orders of magnitude in terms of accuracy. Accepted in VLDB 2024 [C2, First author]
- Research on developing efficient algorithm for subhypergraph matching (In progress)
- Research on developing efficient algorithm for graph similarity search (In progress)

Efficient Subgraph Matching for Drug Hepatotoxicity Prediction — Seoul, Korea

Capstone project with AlgenDrug Co., Ltd. — Sep.2021 - Dec.2021

- Implementation and development of efficient subgraph isomorphism algorithm for chemical graphs
- Gained 10x performance boost in substructure search on PubChem and ZINC molecule graph dataset.

Publications

Conference Publications

C2 **Cardinality Estimation of Subgraph Matching: A Filtering-Sampling Approach** — VLDB 2024
Wonseok Shin, Siwoo Song, Kunsoo Park, Wook-Shin Han — (Accepted)

C1 **Improved adaptive matching order for subgraph matching problem** — KCC 2021
Seunghwan Min, **Wonseok Shin**, Chaewon Kim, Kunsoo Park

Journal Publications

J2 **Supervised Chemical Graph Mining Improves Drug-Induced Liver Injury (DILI) Prediction** — iScience 2023
Sangsoo Lim, Youngkuk Kim, Jeonghyeon Gu, Sunho Lee, **Wonseok Shin**, Sun Kim

J1 **New Adaptive Matching Order and Performance Comparison for Subgraph Matching Problem** — J. of KIISE 2022
Seunghwan Min, **Wonseok Shin**, Chaewon Kim, Kunsoo Park — (Ext. ver. of C1)

Preprints / Works in Progress

P1 **Triangular Constrastive Learning on Molecular Graphs** — MoML 2023
MinGyu Choi, **Wonseok Shin**, Yijingxiu Lu, Sun Kim

Honors & Awards

2022 **Bachelor's Thesis Poster Presentation Award**, Dept. of Computer Science and Engineering, Seoul National University — Seoul, Korea
2021 **Best Paper Award (Computer Theory)**, Korea Computer Congress 2021 — Jeju, Korea
2020 **National Scholarship For Science and Engineering**, Korea Student Aid Foundation — Korea

Programming Competitions

International

2022 **106th Place (Top 1% among ~10,000 teams)**, Google Hash Code 2022 — Online
2021 **504th Place (Top 1.5% among ~37,000 contestants)**, Google Codejam 2021 — Online
2021 **211th Place (Top 2% among ~10,000 teams)**, Google Hash Code 2021 — Online
2020 **468th Place (Top 5% among ~10,000 teams)**, Google Hash Code 2020 — Online

Domestic

2023 **Finalist**, Samsung Collegiate Programming Contest — Online
2022 **Finalist**, Samsung Collegiate Programming Contest — Online
2021 **18th Place**, ICPC Korea First Round — Online

Skills

- **Programming**: C++, Python, JAVA
- **Machine Learning**: PyTorch, scikit-learn
- **Scientific Computing**: NumPy, SciPy, Pandas
- **Data Visualization**: Seaborn, Matplotlib
- **ML Experiment**: TensorBoard
- **Bio/Cheminformatics**: RDKit, DeepChem
- **Tools/Environments**: Git, LaTeX, Linux
- **Languages**: Korean (Native), English (Fluent)

Other Participations

- **Teaching Assistant**: Engineering Mathematics, Algorithms, Automata Theory
- **Tutor**: Basic Computing-First Adventures in Computing (Mentored student teams for data visualization projects using python)
- **Problem Tester for Programming Competitions**: Sogang University, Chungang University, ICPC Sinchon training camp

Portfolio (Selected Projects)

Cardinality Estimation of Subgraph Matching

Duration	October 2022 - March 2024
Results	• Developed algorithm for approximate subgraph counting in large graphs, outperforming existing sampling and GNN based methods by up to two orders of magnitude in terms of accuracy. • Accepted in VLDB 2024.
Role	• Proposed the filtering-sampling idea and developed algorithm • Implemented the proposed algorithm C++ • Conducted experiments to analyze the performance of the proposed algorithm • Wrote the paper as first author
Skills	Understanding of probabilistic algorithms and scalable algorithms for large graph data Implementation and experimental analysis of algorithms on graph in C++
Links	Paper(VLDB) Code

TriCL: Triangular Contrastive Learning on Molecular Graphs

Duration	March 2022 - July 2022
Results	• Achieved state-of-the-art performance on molecular property prediction dataset (MoleculeNet)
Role	• Mathematical formulation of the proposed multimodal contrastive loss function • Implemented the proposed loss function in PyTorch
Skills	PyTorch, PyTorch Geometric, NumPy, RDKit Multimodal learning, Contrastive learning, Graph Neural Networks Communication and collaboration with domain experts
Links	Paper (ArXiv) Press (TriCL)

Efficient Subgraph Matching for Drug Hepatotoxicity Prediction

Duration	Sep 2021 - Dec 2021
Results	• Achieved 10x performance compared to existing library on PubChem and ZINC dataset
Role	• Developed efficient subgraph matching algorithm on molecular graphs • Implemented the proposed algorithm in C++ • Conducted experiments to analyze the performance of the proposed algorithm
Skills	C++, Boost, RDKit, Python