

## I. Personal Information

Semih Salihoglu

Associate Professor

David R. Cheriton Faculty Fellow

David R. Cheriton School of Computer Science, University of Waterloo

200 University Avenue West Waterloo, ON, Canada N2L 3G1

Email: semih.salihoglu@uwaterloo.ca

## II. Education

Degree	Discipline	Institution	Date of Convocation
Doctorate <i>Massive-scale Processing of Record-Oriented and Graph Data</i>	Computer Science	Stanford University	09/2015
Bachelor's	Computer Science and Economics	Yale University	05/2006

## III. Employment History

Position	Institution	Start Date	End Date
Associate Professor	University of Waterloo, School of Computer Science	July 2022	Present
Assistant Professor	University of Waterloo, School of Computer Science	Jan 2016	July 2022
Research Intern	Microsoft Research Redmond	Jun 2013	Aug 2013
Software Engineer	Google, Inc	Oct 2006	Jul 2009

## IV. Recognitions

- David R. Cheriton Faculty Fellow (2022-2025), An award to support the work of leading faculty in the School of Computer Science with an emphasis on supporting research that addresses problems associated with designing and implementing efficient and reliable computing systems.
- ACM SIGMOD Research Highlights Award 2022 for the paper titled “Accurate Summary-based Cardinality Estimation Through the Lens of Cardinality Estimation Graphs”.
- Best Experiments, Analysis, and Benchmark Paper Award, International Conference on Very Large Data Bases (VLDB) 2022 for the paper titled “Accurate Summary-based

Cardinality Estimation Through the Lens of Cardinality Estimation Graphs”.

- Best Paper Award, International Conference on Very Large Data Bases (VLDB) 2018 for the paper titled “The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing”.
- Best Paper Runner-up Award, International Conference on Scientific and Statistical Database Management 2013, for the paper titled “GPS: A Graph Processing System”.
- Distinguished Associate Editor, SIGMOD 2023
- Distinguished PC/Review Board Member, PVLDB 2018, 2020, 2022, SIGMOD 2018, 2021

## V. Publications

### Journal Papers

1. Jeremy Chen, Yuqing Huang, Mushi Wang, Semih Salihoglu, Ken Salem. Accurate Summary-based Cardinality Estimation Through the Lens of Cardinality Estimation Graphs.(SIGMOD Research Highlight Award version for broader audience) SIGMOD Record 52(1), 2023,
2. Khaled Ammar, Siddhartha Sahu, Semih Salihoglu, Tamer Ozsü. Optimizing Differentially-Maintained Recursive Queries on Dynamic Graphs. PVLDB 15(11), 2022,
3. Jeremy Chen, Yuqing Huang, Mushi Wang, Semih Salihoglu, Ken Salem. Accurate Summary-based Cardinality Estimation Through the Lens of Cardinality Estimation Graphs. PVLDB 15(8), 2022,
4. Guodong Jin, Semih Salihoglu. Making RDBMSs Efficient on Graph Workloads Through Predefined Joins. PVLDB 15(5), 2022,
5. Pranjal Gupta, Amine Mhedhbi, Semih Salihoglu. Integrating Column-Oriented Storage and Query Processing Techniques Into Graph Database Management Systems. PVLDB, 14(11), 2021
6. Amine Mhedhbi, Chathura Kankanamge, Semih Salihoglu. Optimizing One-time and Continuous Subgraph Queries using Worst-Case Optimal Joins. ACM TODS 46(2), 2021: 1-45
7. Amine Mhedhbi, Semih Salihoglu. Optimizing Subgraph Queries by Combining Binary and Worst-Case Optimal Joins. PVLDB, 12(11), 2019: 1692-1704
8. Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin., Tamer Ozsü. The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing: Extended Survey. VLDBJ, 29(2-3), 2019: 595-618

9. Khaled Ammar, Frank McSherry, Semih Salihoglu. Distributed Evaluation of Subgraph Queries Using Worstcase Optimal Low-Memory Dataflows. PVLDB, 11(6), 2018: 691-704
10. Paraschos Koutris, Semih Salihoglu, Dan Suciu (equal contribution). Algorithmic Aspects of Parallel Data Processing. Foundations and Trends in Databases, 8(4), 2018: 239-370
11. Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin, Tamer Ozsu. The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing. PVLDB, 11(4), 2018: 420-431 (*Best Paper Award*)
12. Semih Salihoglu, Jennifer Widom. Optimizing Graph Algorithms on Pregel-like Systems. PVLDB, 7(7), 2014: 577-588
13. Foto Afrati, Anish Das Sarma, Semih Salihoglu, Jeffrey D. Ullman. Upper and Lower Bounds on the Cost of a MapReduce Computation. PVLDB, 6(4), 2013: 277-288
14. Ibrahim Abdelaziz, Razen Al-Harbi , Semih Salihoglu, Panos Kalnis. Combining Vertex-centric Graph Processing with SPARQL for Large-scale RDF Data Analytics. IEEE Transactions on Parallel and Distributed Systems, 28(12), 2017: 3374-3388

## Book Chapters

1. Semih Salihoglu, Nikolay Yakovets. Graph Querying. Sherif Sakr, Albert Y. Zomaya. Encyclopedia of Big Data Technologies, 2019

## Conference Papers

1. Arif Usta, Chang Liu, Semih Salihoglu. Analysis of Open Government Datasets From a Data Design and Integration Perspective International Conference on Extending Database Technology (EDBT), 2024
2. Chang Liu, Arif Usta, Semih Salihoglu, Jian Zhao. Governor: Turning Open Government Data Portals into Interactive Databases. ACM Conference on Human Factors in Computing Systems (SIGCHI), 2023
3. Xiyang Feng, Guodong Jin, Ziyi Chen, Chang Liu, Semih Salihoglu. KÙZU Graph Database Management System, CIDR 2023
4. Steven Xia, Nafisa Anzum, Semih Salihoglu, Jian Zhao. KTabulator: Interactive Ad hoc Table Creation using Knowledge Graphs. ACM Conference on Human Factors in Computing Systems (SIGCHI), 2021
5. Kaleb Alway, Eric Blais, Semih Salihoglu. Box Covers and Domain Orderings for Beyond Worst-Case Join Processing. ICDT, 2021: 3:1-3:23

6. Siddhartha Sahu, Semih Salihoglu. Graphsurge: Graph Analytics on View Collections Using Differential Computation. SIGMOD, 2021: 1518-1530
7. Amine Mhedhbi, Pranjal Gupta, Shahid Khaliq, Semih Salihoglu. A+ Indexes: Lightweight and Highly Flexible Adjacency Lists for Graph Database Management Systems. ICDE, 2021: (1-12)
8. Luca Castelli Aleardi, Semih Salihoglu, Gurprit Singh, Maks Ovsjanikov. Spectral Measures of Distortion for Change Detection in Dynamic Graphs. Complex Networks, 2018: 54–66
9. Mustafa Korkmaz, Martin Karsten, Ken Salem, Semih Salihoglu. Workload-Aware CPU Performance Scaling for Transactional Database Systems, SIGMOD, 2018: 291-306
10. Foto Afrati, Manas Joglekar, Chris Re, Semih Salihoglu, Jeffrey D. Ullman. GYM: A Multiround Join Algorithm In MapReduce. ICDT, 2017: 4:1-4:18
11. Foto Afrati, Anish Das Sarma, Anand Rajaraman, Pokey Rule, Semih Salihoglu, Jeffrey D. Ullman, Anchor Points Algorithms for Hamming and Edit Distance. ICDT, 2014: 4-14
12. Sungpack Hong, Semih Salihoglu, Jennifer Widom, Kunle Olukotun. Simplifying Scalable Graph Processing with a Domain-Specific Language. International Symposium on Code Generation and Optimization (CGO), 2014: 208-218
13. Semih Salihoglu, Jennifer Widom. GPS: A Graph Processing System. International Conference on Scientific and Statistical Database Management (SSDBM), 2013: 1-12 (*Best Paper Runner-up Award*)
14. Robert Ikeda, Semih Salihoglu, Jennifer Widom. Provenance-Based Refresh in Data-Oriented Workflows. Conference on Information and Knowledge Management (CIKM), 2011: 1659-1668

## Demonstration Papers

1. Guodong Jin, Nafisa Anzum, Semih Salihoglu. GRainDB: A Relational-core Graph-Relational DBMS. CIDR, 2022
2. Nafisa Anzum, Semih Salihoglu, Daniel Vogel. GraphWrangler: An Interactive Graph View on Relational Data. SIGMOD, 2019: 1865-1868
3. Chathura Kankanamge, Siddhartha Sahu, Amine Mhedhbi, Jeremy Chen, Semih Salihoglu. Graphflow: An Active Graph Database. SIGMOD, 2017: 1695-1698
4. Ibrahim Abdelaziz, Razen Harbi, Semih Salihoglu. Panos Kalnis, and Mamoullis. SPARTex: A Vertex-Centric Framework for RDF Data Analytics. VLDB, 2015: 1880-1883

5. Semih Salihoglu, Jaeho Shin, Vikesh Khanna, Ba Quan Truong, Jennifer Widom. Graft: A Debugging Tool For Apache Graph. SIGMOD, 2015: 1403-1408
6. Robert Ikeda, Junsang Cho, Charlie Fang, Semih Salihoglu, Satoshi Torikai, Jennifer Widom. Provenance-Based Debugging and Drill-Down in Data-Oriented Workflows. ICDE, 2012: 1249-1252

## Workshop Papers

1. Nafisa Anzum, Semih Salihoglu. R2GSync and Edge Views: Practical RDBMS to GDBMS Synchronization. GRADES-NDA Workshop on Graph Data Management Experiences & Systems and Network Data Analytics, 2021: 1-9
2. Semih Salihoglu, Jennifer Widom. HelP: High-level Primitives for Large-Scale Graph Processing. Graph Data-management Experiences and Systems Workshop, 2014: 1-6
3. Foto Afrati, Anish Das Sarma, Semih Salihoglu, Jeffrey D. Ullman. Towards an Understanding of the Limits of MapReduce Computation. Cloud Futures Workshop, 2012

## VI. Presentations and Tutorials

1. Modern Architectures for Graph Database Management Systems, University of Zurich Colloquium, 2023
2. Kuzu Graph Database Management System, Technical University of Munich, Toronto AICamp, Royal Bank of Canada, BASF Knowledge Graph Group (Online), Invited Talks, 2023
3. Cardinality Estimation Graphs, CWI and Eindhoven University of Technology, Netherlands, Invited Talks, 2023
4. Tutorial on Modern Techniques for Querying Graph-Structured Relations: Foundations, Systems Implementations and Open Challenges (w/ Amine Mhedhbi), VLDB, 2022
5. Round Table on Graph Mining Algorithms, Invited Panelist, ICDE, 2022
6. Integrating Column-Oriented Storage and Processing Techniques into GraphflowDB, Concordia University, Invited Talk, 2022
7. Round Table on Graph Data Management and Analytics, VLDB, 2021
8. GraphflowDB and Modern Query Processing Techniques for Graph Database Management Systems. Google Tech Talk (Online), 2021
9. GraphflowDB and Modern Query Processing Techniques for Graph Database Management Systems. Pinterest Tech Talk (Online), 2021

10. Integrating Column-Oriented Storage and Processing Techniques into GraphflowDB. Dutch Seminar on Data Systems Design (Online), 2021
11. Query Processor of GraphflowDB and Techniques for Graph Databases of 2020s, Seminars at LinkedIn (Online), 2020, Carleton University, Canada (Online), 2020, The Global Graph Summit, USA, Seminar 2020
12. Worst-case Optimal Join-Style Subgraph Query Evaluation in GraphflowDB, Seminar at RelationalAI (Online), 2019
13. Optimizing Subgraph Queries With a Mix of Tradition and Modernity, Seminar at Technical User Community Meeting LDBC, Netherlands, 2019
14. How can Worst-case Optimal Join Algorithms be So Interesting? Seminars at METU, Turkey, 2019, Bilkent University, Turkey, 2019, Bogazici University, Turkey, 2019, Istanbul Technical University, Turkey 2019, University at Buffalo, USA, 2019, Purdue University, USA, 2018
15. Tutorial on Algorithmic Aspects of Parallel Query Processing (w/ Paraschos Koutris and Dan Suciu), SIGMOD, 2018
16. Graph Query Processing using Worst-case Optimal Join Algorithms and Differential Computation, Presentation at Shonan Workshop on Graph Database Systems, Japan, 2018
17. Applications of New Join and Spectral Algorithms in Dynamic Graph Processing, Seminar at Ecole Polytechnique, France, 2017
18. Surprises in a Survey of the Graphs, Computations, and Software Used in Practice, Presentation at Workshop on Big Graph Analysis Systems, Denmark, 2017
19. Graphflow Graph Database, Seminar at SAP Hana Tech Days, Germany, 2017
20. Graph Databases and The Graphflow Active Graph Database: Invited Talk at KW Intersect Tech Meetup, Canada, 2017
21. Tutorial on Designing Good Algorithms for MapReduce and Beyond (w/ Foto Afrati, Magda Balazinska, Anish Das Sarma, Bill Howe, and Jeffrey D. Ullman, ACM Symposium on Cloud Computing (SoCC), 2012

## VII. Research Funding History

Program	Role	Total Funding	Share	Start Date	End Date
NSERC CREATE (under review): Sustainable Data Systems for Data Science	co-PI	\$1650K	\$275K	June-24	June-30

NSERC Discovery (under review): Architectural Principles of Modern Datalog Database Management Systems	PI	\$540K	\$540K	April-24	April-29
Waterloo-Huawei Joint Innovation Lab (Industry Grant): Making Graph AI Practical Through DBMS-Graph Embedding Integration	PI	\$210K	\$210K	Sep-22	Sep-25
NSERC CRD: Continuous Graph Querying and Graph OLAP Using Differential Computation	PI	\$270K	\$270K	Oct-19	Aug-24
Waterloo-Huawei Joint Innovation Lab (Industry Grant): Continuous Graph Querying and Graph OLAP Using Differential Computation	PI	\$213,000	\$213,000	Oct-19	Oct-23
NSERC Discovery: Systems and Foundations For Massive-Scale Data Management	PI	\$217K	\$217K	Apr-16	Apr-23
Startup funds from the Department of Computer Science at the University of Waterloo	PI	\$90K	\$90K	Jan-16	Oct-21
TOTAL (excluding under reviews)		\$1,000K	\$1000K		

## VIII. Student Supervision

### Details<sup>1</sup>

Trainee	Role	Degree	Period	Status
Gaurav Sehgal	Supervisor	Master's	June 2023 - present	In Progress
Anurag Chakraborty	Supervisor	Master's	Sep 2022 - present	In Progress
Guodong Jin	Supervisor	Postdoc	Sep 2022 - Sep 2025	In Progress
Jiamin Hou (Visiting Student)	Mentor	Doctorate	Dec 2023 - present	In Progress
Arif Usta	Supervisor	Postdoc	Sep 2021 - Sep 2023	Completed
Chang Liu	Supervisor	Master's	Sep 2020 - Sep-2022	Completed

<sup>1</sup>The details of undergraduate students are omitted.

Amine Mhedhbi	Supervisor	Doctorate	Jan 2017 - May 2023	Completed (Prof. at Polytech. Montréal)
Siddhartha Sahu	Supervisor	Doctorate	Sep 2016 - present	In Progress
Khaled Ammar	Co-supervisor	Doctorate	Sep 2014-May 2023	Completed (at Borealis)
Tamal Adhikary	Co-supervisor	Master's	Sep 2019-May 2022	Completed (at Huawei)
Guodong Jin (Visiting Student)	Mentor	Doctorate	Sep 2018 - present	Completed (in Postdoc)
Xiyang Feng	Supervisor	Master's	Sep 2019 - Dec 2021	Completed
Tamal Adhikary	Co-supervisor	Master's	Sep 2019 - May 2022	Completed
Nafisa Anzum	Supervisor	Master's	Sep 2019 - Sep 2020	Completed (in PhD)
Pranjal Gupta	Supervisor	Master's	Sep 2019 - Sep 2020	Completed (SAP)
Jeremy Chen	Supervisor	Master's	Jan 2019 - Sep 2020	Completed (Snowflake)
Kaleb Alway	Co-supervisor	Master's	Sep 2017 - Sep 2019	Completed (SAP)
Shahid Khaliq	Supervisor	Master's	Sep 2017 - Sep 2019	Completed (Top Hat)
Chathura Kankanamge	Supervisor	Master's	Sep 2017 - Sep 2019	Completed (Amazon)
Arman Naeimian	Co-supervisor	Master's	Sep 2017 - Sep 2019	Completed (Sandvine)
Azin Nazari	Co-supervisor	Master's	Sep 2017 - Sep 2019	Completed (Amazon)

## IX. Teaching

1. CS 848: Knowledge Graphs (UWaterloo - Graduate level course) – Instructor (Fall



- 2022 - 14 students)
2. CS 348: Introduction to Database Systems (UWaterloo - 3rd year Undergraduate level course) – Instructor (Fall 2021 - 200 students, Winter 2022 - 200 students)
  3. CS 341: Algorithms (UWaterloo - 3rd year Undergraduate level course) – Instructor (Summer 2016 - 70 students, Winter 2017 - 150 students, Winter 2018 - 160 students, Winter 2019 - 160 students, Spring 2020 - 200 students)
  4. CS 848: Graph Analytics and Data Management (UWaterloo - Graduate level course) – Instructor (Winter 2020 - 11 students)
  5. CS 848: Graph Data Management t (UWaterloo - Graduate level course) – Instructor (Fall 2018 - 10 students)
  6. CS 848/858: Modern Data Processing Systems (UWaterloo - Graduate level course) – Instructor (Fall 2016 - 20 students)

## **X. Service**

### **External Service**

#### **A. Program Co-Chair**

1. SIGMOD 2022 Demonstration Track
2. GRADES-NDA Workshop on Graph Data Management Experiences & Systems and Network Data Analytics 2020, 2022

#### **B. Program Committee**

1. ACM Special Interest Group on Management of Data (SIGMOD) 2023 (Associate Editor)
2. ACM Special Interest Group on Management of Data (SIGMOD) 2017, 2018, 2019, 2020, 2021
3. ACM Special Interest Group on Management of Data (SIGMOD) Demonstration Track 2019
4. IEE International Conference on Data Engineering (ICDE) 2017, 2018, 2021
5. International Conference on Extending Database Technology 2020 (EDBT) (Area Chair)
6. ACM Symposium on Cloud Computing (SoCC) 2019
7. Workshop on Graph Algorithms Building Blocks (GABBS), 2017

8. Workshop on Algorithms for MapReduce and Beyond (BeyondMR), 2014, 2015, 2016, 2017
9. High Performance Graph Data Management and Processing Workshop (HPGDMP), 2016

### **C. Journal and External Reviewer Service**

1. IEEE Transactions on Knowledge and Data Engineering (TKDE), Associate Editor, 2023 January-present
2. Proceedings of Very Large Data Bases Endowment (PVLDB) Review Board, 2017, 2019, 2020, 2021, 2022
3. The International Journal on Very Large Data Bases (VLDBJ), 2017, 2019-2022
4. TODS, 2018, 2019, 2020
5. IEEE Transactions on Knowledge and Data Engineering (TKDE) 2013, 2017, 2018, 2019
6. IEEE Transactions on Parallel and Distributed Systems (TPDS), 2017, 2018
7. IEEE Transactions on Big Data, 2017
8. Communications of the ACM 2015
9. USENIX ATC 2017, EDBT 2014, ICDE 2012 (external)

### **Internal Service**

1. 2018 – 2019: School Advisory Committees on Appointments (SACA) - Faculty Hiring Committee
2. 2016–2018, 2019–2022: Grad Committee - Admissions and scholarship review
3. 2016, 2017: Undergraduate Research Opportunities Conference - Mini-project instructor for attending students from Canada