Lukasz Golab

Associate Professor and Canada Research Chair
University of Waterloo
Department of Management Sciences, Faculty of Engineering
Cross-appointed to the School of Computer Science, Faculty of Mathematics

Bio: I joined Waterloo in 2011 and was awarded a Tier-2 Canada Research Chair in 2015. From 2006 to 2011 I was a Senior Member of Research Staff at AT&T Labs. I have a BSc in Computer Science from the University of Toronto (2001; with High Distinction) and a PhD in Computer Science from the University of Waterloo (2006; with Alumni Gold Medal).

Research interests (also see my Data Science Lab)

- Data for good (sustainability, clean energy and transportation, health, education, equality)
- Data profiling and cleaning (discovering column dependencies and business rules from data, detecting data errors)
- Data systems (blockchains, data stream systems and real-time analytics, graph and social network mining, distributed data management)

Recent professional service: Associate Editor for Information Systems; PC co-chair for DEBS 2020, Review Board Member for PVLDB 2018; PC member for SSDBM 2020, AMW 2020, ICDE 2020, EDM 2019, ICDE 2019, CIKM 2018, SIGMOD 2018, SSDBM 2018; Publicity co-chair for SIGMOD 2020 and ICDE 2019

Recent teaching: MSCI446 (Data Mining), MSCI623 (Big Data Analytics).

Media

- Researchers use blockchain to drive electric vehicle infrastructure
- Speeding up blockchains to meet real-world needs
- Gender differences in engineering applicants (Globe and Mail)
- Are smart meters delivering on their promise?, The Weather Wild Card: Assessing Time-of-Use Electricity Pricing, CTV news

Tutorials

- Z. Abedjan, L. Golab, F. Naumann, *Data Profiling*, **SIGMOD** 2017, 1747-1751 and **ICDE** 2016, 1432-1435. Slides here
- L. Golab, T. Johnson, *Data Stream Warehousing*, **ICDE** 2014, 1290-1293 and **SIGMOD** 2013, 949-952. Slides here

Publications

2020

- L. Chen, L. Golab, *Micro-Journal Mining to Understand Mood Triggers*, **Computing**, special issue on Data analytics for engineering, science, and society, to appear

2019

- H. Saxena, L. Golab, I. Ilyas, *Distributed Implementations of Dependency Discovery Algorithms*, **PVLDB** 12(11): 1624-1636, 2019
- A. Khan, L. Golab, M. Kargar, J. Szlichta, M. Zihayat, *Compact Group Discovery in Attributed Graphs and Social Networks*, **Journal of Information Processing and Management**
- M. Vollmer, L. Golab, K. Boehm, D. Srivastava, *Informative Summarization of Numeric Data*, **SSDBM** 2019, 97-108
- C. Gorenflo, L. Golab, S. Keshav, *Mitigating Trust Issues in Electric Vehicle Charging using a Blockchain*, **e-Energy** 2019, 160-164
- S. Chopra, A. Khan, M. Mirsafian, L. Golab, *Gender Differences in Work-Integrated Learning Assessments*, **EDM** 2019, 524-527
- H. Davoudi, P. Godfrey, L. Golab, M. Kargar, D. Srivastava, J. Szlichta, *Bringing Order to Data*, **AMW** 2019, 11:1-11:5
- D. Kumar, R. Cohen, L. Golab, *Online abuse detection: the value of preprocessing and neural attention models*, NAACL workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (**WASSA**), 16-24
- S. Chopra, M. Mirsafian, A. Khan, L. Golab, *Gender Differences in Science and Engineering: A Data Mining Approach*, EDBT workshop on **Data Analytics Solutions for Real-Life Applications**, 7:1-7:8
- C. Gorenflo, S. Lee, L. Golab, S. Keshav, *FastFabric: Scaling Hyperledger Fabric to 20,000 Transactions per Second*, IEEE Int. Conf. on Blockchain and Cryptocurrency (**ICBC**) 2019, 455-463
- H. Saxena, L. Golab, I. Ilyas, *Distributed Discovery of Functional Dependencies*, **ICDE** 2019, 1590-1593
- L. Golab, *Types of Stream Processing Algorithms*, Encyclopedia of Big Data Technologies, Springer, 2019

2018

- S. Chopra, L. Golab, T. J. Pretti, A. Toulis, *Using data mining methods for research in co-operative education*, **Int. Journal of Work Integrated Learning**, 19(3): 297-310, 2018
- Z. Abedjan, L. Golab, F. Naumann, T. Papenbrock, *Data Profiling*, Morgan & Claypool Publishers, 2018
- K. El Gebaly, G. Feng, L. Golab, F. Korn, D. Srivastava, *Explanation Tables*, IEEE **Data Engineering Bulletin**, 41(3):43-51, 2018
- J. Szlichta, P. Godfrey, L. Golab, M. Kargar, D. Srivastava, *Effective and Complete Discovery of Bidirectional Order Dependencies via Set-based Axioms*, **VLDB Journal**, 27(4): 573-591
- S. Chopra, H. Gautreau, A. Khan, M. Mirsafian, L. Golab, *Gender Differences in Undergraduate Engineering Applicants: A Text Mining Approach*, **EDM** 2018, 44-54
- S. Chopra, L. Golab, *Job Description Mining to Understand Work-Integrated Learning*, **EDM** 2018, 32-43
- L. Gao, L. Golab, T. Ozsu, G. Aluc, *Stream WatDiv - A Streaming RDF Benchmark*, SIGMOD workshop on **Semantic Big Data**, 3:1-3:6
- M. Zihayat, A. An, L. Golab, M. Kargar, J. Szlichta, *Effective Team Formation in Expert Networks*, **AMW** 2018, 4:1-4:4
- M. Langouri, Z. Zheng, F. Chiang, L. Golab, J. Szlichta, *Contextual Data Cleaning*, ICDE workshop on **Context in Analytics**, 21-24

- G. Tang, S. Keshav, L. Golab, K. Wu, *Bikeshare Pool Sizing for Bike-And-Ride Multimodal Transit*, **Trans. on Intelligent Transportation Systems**, 19(7): 2279-2289
- S. Chopra, Y. Jiang, A. Toulis, L. Golab, *Data Analytics to Improve Co-Operative Education*, EDBT workshop on **Data Analytics Solutions for Real-Life Applications**, 16-21
- A. Mihaylov, P. Godfrey, L. Golab, M. Kargar, D. Srivastava, J. Szlichta, *FastOD: Bringing Order to Data*, **ICDE** 2018, 1561-1564, demo paper
- Z. Zheng, M. Langouri, Z. Qu, I. Currie, F. Chiang, L. Golab, J. Szlichta, *FastOFD: Contextual Data Cleaning with Ontology Functional Dependencies*, **EDBT** 2018, 694-697, demo paper
- A. Andrade, S. Chopra, B. Nurlybayev, L. Golab, *Quantifying the Impact of Entrepreneurship on Cooperative Job Creation*, **Int. Journal of Work Integrated Learning**, 19(1): 51-68

2017

- X. Meng, L. Golab, *Optimal Reducer Placement to Minimize Data Transfer in MapReduce-Style Processing*, IEEE **BigData** 2017, 339-346
- S. Baskaran, A. Keller, F. Chiang, L. Golab, J. Szlichta, *Efficient Discovery of Ontology Functional Dependencies*, **CIKM** 2017, 1847-1856
- C. Gorenflo, I. Rios, L. Golab, S. Keshav, *Usage Patterns of Electric Bicycles: An Analysis of the WeBike Project*, **J. of Advanced Transportation**, 2017, Article ID 3739505
- Y. Yang, L. Golab, M. T. Ozsu, ViewDF: *Declarative Incremental View Maintenance for Streaming Data*, **Information Systems**, 71 (2017) 55-67
- A. Toulis, L. Golab, *Social Media Mining to Understand Public Mental Health*, **DMAH** workshop at VLDB 2017, 55-70
- K. El Gebaly, L. Golab, J. Lin, *Portable In-Browser Data Cube Exploration*, **IDEA** workshop at KDD 2017, 35-39
- C. Gorenflo, L. Golab, S. Keshav, *Managing Sensor Data Streams: Lessons Learned from the WeBike Project*, **SSDBM** 2017, 1:1-1:11
- S. Fink, L. Golab, S. Keshav, H. de Meer, *How Similar is the Usage of Electric Cars and Electric Bicycles?*, **EV-Sys** 2017, 334-340
- A. Toulis, L. Golab, *Graph Mining to Characterize Competition for Employment*, **Network Data Analytics** workshop at SIGMOD 2017, 3:1-3:7
- R. Miller, L. Golab, C. Rosenberg, *Modelling Weather Effects for Impact Analysis of Residential Time-of-Use Electricity Pricing*, **Energy Policy** 105 (2017) 534-546
- J. Szlichta, P. Godfrey, L. Golab, M. Kargar, D. Srivastava, *Effective and Complete Discovery of Order Dependencies via Set-based Axiomatization*. **PVLDB** 10(7): 721-732, 2017
- G. Feng, L. Golab, D. Srivastava, *Scalable Informative Rule Mining*, **ICDE** 2017, 48
- M. Zihayat, A. An, L. Golab, M. Kargar, J. Szlichta, *Authority-Based Team Discovery in Social Networks*, **EDBT** 2017, 498-501
- X. Liu, L. Golab, W. Golab, I. Ilyas, S. Jin, *Smart Meter Data Analytics: Systems, Algorithms and Benchmarking*, **TODS** 42(1): 2:1-2:39, 2017

2016

- M. Kargar, L. Golab, J. Szlichta, *eGraphSearch: Effective Keyword Search in Graphs*, **CIKM** 2016, 2461-2464, demo paper. Tech report here
- I. Rios, L. Golab, S. Keshav, *Analyzing the Usage Patterns of Electric Bicycles*, **EV-Sys** 2016, 2
- A. Baer, P. Casas, A. D'Alconzo, P. Fiadino, L. Golab, M. Mellia, E. Schikuta, *DBStream: A Holistic Approach to Large-Scale Network Traffic Monitoring and Analysis*, **Computer Networks** 107 (2016) 5-19

- L. Gebhard, L. Golab, S. Keshav, H. de Meer, *Range prediction for electric bicycles*, **e-Energy** 2016, 224-234
- Y. Jiang, L. Golab, *On Competition for Undergraduate Co-op Placements: A Graph Mining Approach*, **EDM** 2016, 394-399
- Y. Jiang, R. Levman, L. Golab, J. Nathwani, *Analyzing the Impact of the 5CP Ontario Peak Reduction Program on Large Consumers*, **Energy Policy** 93 (2016) 96-100. Short version here
- Y. Jiang, S. J. Syed, L. Golab, *Data mining of undergraduate course evaluations*, **Informatics in Education** 15(1): 85-102

2015

- Y. Yang, L. Golab, M. T. Ozsu, *ViewDF: declarative incremental view maintenance for streaming data*, VLDB Workshop on Business Intelligence for the Real Time Enterprise (**BIRTE**) 2015. Early version here
- C. Ge, M. Kaufmann, L. Golab, P. M. Fischer, A. Goel, *Indexing bi-temporal windows*, **SSDBM** 2015, 19
- Z. Abedjan, L. Golab, F. Naumann, *Profiling relational data - a survey*, **VLDB Journal** 24(4): 557-581
- J. Szlichta, L. Golab, D. Srivastava, *On Axiomatization and Inference Complexity over a Hierarchy of Functional Dependencies*, **AMW** 2015
- Y. Jiang, S. Lee, L. Golab, *Analyzing student and employer satisfaction with cooperative education through multiple data sources*, **Asia-Pacific Journal of Cooperative Education**, 16(4):225-240, 2015
- X. Gao, L. Golab, S. Keshav, *What's wrong with my solar panels: a data-driven approach*, **EnDM** 2015, 86-93
- X. Liu, L. Golab, W. Golab, I. Ilyas, *Benchmarking Smart Meter Data Analytics*, **EDBT** 2015, 385-396
- A. Baer, L. Golab, S. Ruehrup, M. Schiavone, P. Casas, *Cache-Oblivious Scheduling of Shared Workloads*, **ICDE** 2015, 855-866
- L. Golab, F. Korn, F. Li, B. Saha, D. Srivastava, *Size-Constrained Weighted Set Cover*, **ICDE** 2015, 879-890
- X. Liu, L. Golab, I. Ilyas, *SMAS: A Smart Meter Data Analytics System*, **ICDE** 2015, 1476-1479, demo paper

2014

- A. Baer, A. Finamore, P. Casas, L. Golab, M. Mellia, *Large-Scale Network Traffic Monitoring with DBStream, a System for Rolling Big Data Analysis*, **IEEE BigData** 2014, 165-170
- K. El Gebaly, P. Agrawal, L. Golab, F. Korn, D. Srivastava, *Interpretable and Informative Explanations of Outcomes*, **PVLDB** 8(1):61-72, 2014
- L. Golab, M. Hadjieleftheriou, H. Karloff, B. Saha, *Distributed Data Placement to Minimize Communication Costs via Graph Partitioning*, **SSDBM** 2014, 20-31. Tech report here: CoRR abs/1312.0285
- A. Baer, P. Casas, L. Golab, A. Finamore, *DBStream: an Online Aggregation, Filtering and Processing System for Network Traffic Monitoring*, Int. Workshop on Traffic Analysis and Characterization (**TRAC**) 2014, 611-616
- S. J. Syed, Y. Jiang, L. Golab, *Data Mining of Undergraduate Course Evaluations*, **EDM** 2014, 347-348

- T. Carpenter, L. Golab, S. J. Syed, *Is the grass greener? Mining electric vehicle opinions*, **e-Energy** 2014, 241-252
- Y. Jiang R. Levman, L. Golab, J. Nathwani, *Predicting peak-demand days in the Ontario peak reduction program for large consumers*, **e-Energy** 2014, 221-222
- O. Ardakanian, N. Koochakzadeh, R. P. Singh, L. Golab, S. Keshav, *Computing Electricity Consumption Profiles from Household Smart Meter Data*, **EnDM** 2014, 140-147. Slides here
- L. Golab, H. Karloff, F. Korn, B. Saha, D. Srivastava, *Discovering Conservation Rules*, **TKDE**, 26(6):1332-1348, 2014
- G. Beskales, I. Ilyas, L. Golab, A. Galuillin, *Sampling from Repairs of Conditional Functional Dependency Violations*, **VLDB Journal**, 23(1):103-128, 2014

2013

- C. Ge, L. Golab, *Lazy data structure maintenance for main-memory analytics over sliding windows*, **DOLAP** 2013, 33-38
- M. Deziel, D. Olawo, L. Truchon, L. Golab, *Analyzing the mental health of Engineering students using classification and regression*, **EDM** 2013, 228-231
- G. Beskales, I. Ilyas, L. Golab, A. Galiullin, *On the Relative Trust between Inconsistent Data and Inaccurate Constraints*, **ICDE** 2013, 541-552
- L. Golab, *Data Warehouse Quality: Summary and Outlook*, in S. Sadiq (ed.), Handbook of Data Quality Research and Practice, Springer-Verlag Berlin Heidelberg 2013

2012

- A. Baer, L. Golab, *Towards Benchmarking Stream Data Warehouses*, **DOLAP** 2012, 105-112
- L. Golab, T. Johnson, S. Sen, J. Yates, *A Sequence-Oriented Stream Warehouse Paradigm for Network Monitoring Applications*, **PAM** 2012, 53-63
- L. Golab, H. Karloff, F. Korn, B. Saha, D. Srivastava, *Discovering Conservation Rules*, **ICDE** 2012, 738-749
- L. Golab, T. Johnson, V. Shkapenyuk, *Scalable Scheduling of Updates in Streaming Data Warehouses*, **TKDE**, 24(6): 1092-1105, 2012

2011

- M. Bateni, L. Golab, M. Hajiaghayi, H. Karloff, *Scheduling to Minimize Staleness and Stretch in Real-Time Data Warehouses*, **Theory of Computing Systems,** 49(4):757-780, 2011
- L. Golab, F. Korn, D. Srivastava, *Efficient and Effective Analysis of Data Quality using Pattern Tableaux*, **IEEE Data Engineering Bulletin**, 34(3):26-33, 2011
- L. Golab, F. Korn, D. Srivastava, *Discovering Pattern Tableaux for Data Quality Analysis: a Case Study*, **QDB** 2011, 47-53, 2011
- L. Golab, T. Johnson, *Consistency in a Stream Warehouse*, **CIDR** 2011, 114-122

2010

- L. Golab, M. T. Ozsu, *Data Stream Management*, Morgan & Claypool Publishers, 2010
- G. Beskales, I. Ilyas, L. Golab, *Sampling the Repairs of Functional Dependency Violations under Hard Constraints*, **PVLDB** 3(1):197-207, 2010
- L. Golab, H. Karloff, F. Korn, D. Srivastava, *Data Auditor: Exploring Data Quality and Semantics using Pattern Tableaux*, **PVLDB** 3(2):1641-1644, 2010, demo paper

- D. Srivastava, L. Golab, R. Greer, T. Johnson, J. Seidel, V. Shkapenyuk, O. Spatscheck, J. Yates, *Enabling Real Time Data Analysis*, **PVLDB** 3(1): 1-2, 2010

2009

- L. Golab, H. Karloff, F. Korn, A. Saha, D. Srivastava, *Sequential Dependencies*, **PVLDB** 2(1):574-585, 2009
- L. Golab, T. Johnson, J. S. Seidel, V. Shkapenyuk, *Stream Warehousing with DataDepot*, **SIGMOD** 2009, 847-854
- G. Cormode, L. Golab, F. Korn, A. McGregor, D. Srivastava, X. Zhang, *Estimating the Confidence of Conditional Functional Dependencies*, **SIGMOD** 2009, 469-482
- L. Golab, T. Johnson, V. Shkapenyuk, *Scheduling Updates in a Real-Time Stream Warehouse*, **ICDE** 2009, 1207-1210
- M. Bateni, L. Golab, M. Hajiaghayi, H. Karloff, *Scheduling to Minimize Staleness and Stretch in Real-Time Data Warehouses*, **SPAA** 2009, 29-38
- L. Golab, *Stream Models*, Encyclopedia of Database Systems, 2009, 2834-2836
- L. Golab, *Data Stream*, Encyclopedia of Database Systems, 2009, 638

2008

- L. Golab, H. Karloff, F. Korn, D. Srivastava, B. Yu, *On Generating Near-Optimal Tableaux for Conditional Functional Dependencies*, **PVLDB** 1(1):376-390, 2008
- L. Golab, T. Johnson, O. Spatscheck, *Prefilter: Predicate Pushdown at Streaming Speeds*, **SSPS** 2008, 29-37
- L. Golab, T. Johnson, N. Koudas, D. Srivastava, D. Toman, *Optimizing Away Joins on Data Streams*, **SSPS** 2008, 48-57

2006

- L. Golab, K. G. Bijay, M. T. Ozsu, *Multi-Query Optimization of Sliding Window Aggregates by Schedule Synchronization*, **CIKM** 2006, 844-845
- L. Golab, P. Prahladka, M. T. Ozsu, *Indexing Time-Evolving Data with Variable Lifetimes*, **SSDBM** 2006, 265-274
- L. Golab, K. G. Bijay, M. T. Ozsu, *On Concurrency Control in Sliding Window Queries over Data Streams*, **EDBT** 2006, 608-626

2005

- L. Golab, M. T. Ozsu, *Update-Pattern-Aware Modeling and Processing of Continuous Queries*, **SIGMOD** 2005, 658-669

2004

- L. Golab, S. Garg, M. T. Ozsu, *On Indexing Sliding Windows over On-Line Data Streams*, **EDBT** 2004, 712-729
- L. Golab, D. DeHaan, A. Lopez-Ortiz, E. Demaine, *Finding Frequent Items in Sliding Windows with Multinomially-Distributed Item Frequencies*, **SSDBM** 2004, 425-426
- L. Golab, *Querying Sliding Windows Over Online Data Streams*, **EDBT Ph.D. Workshop** 2004, 1-11

2003

- L. Golab, D. DeHaan, E. Demaine, A. Lopez-Ortiz, J. I. Munro, *Identifying Frequent Items in Sliding Windows over On-Line Packet Streams*, **IMC** 2003, 173-178
- L. Golab, M. T. Ozsu, *Processing Sliding Window Multi-Joins in Continuous Queries over Data Streams*, **VLDB** 2003, 500-511
- L. Golab, M. T. Ozsu, *Issues in Data Stream Management*, **SIGMOD Record**, 32(2):5-14, 2003