Da Yan

Assistant Professor
Department of Computer Science
College of Arts and Science
The University of Alabama at Birmingham

UH 4151, 1402 10th Avenue South
Birmingham, AL 35294-1241
(205) 207-1962
yanda@uab.edu
http://bit.ly/danielYan

Research Interests

Core Computer Science

- Database and data mining
- Parallel and distributed computing
- Machine learning, especially deep learning
- Algorithm design

Applications & Interdisciplinary Research

- Geospatial data management, GIS, Transportation Engineering
- Bioinformatics, Clinical Informatics
- Network Science
- Analysis of images, videos and natural language

Employment

Assistant Professor — Department of Computer Science, The University of Alabama at Birmingham, Birmingham, AL — 2016 — now

Postdoctoral Fellow — Department of Computer Science & Engineering, The Chinese University of Hong Kong, Sha Tin, New Territories, Hong Kong — 2014 — 2016

Education

Ph.D. in Computer Science — The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong — 2009 — 2014

B.S. in Computer Science — Fudan University, Shanghai, China — 2005 — 2009

Awards

HKIS-Towngas 2015 Young Scientist Award in Physical/Mathematical Science (Sole Winner)	2015
DASFAA 2011 Best Paper Award	2011
Senior Member of ACM	Since 2021
Senior Member of IEEE	Since 2021
UAB Kevin and Joann Reilly Faculty Award ($2,000)	2021 — 2022
South BD Hub Azure for Research Award ($25,000 Cloud Credits)	2017 — 2018
Google Cloud COVID-19 Research Credit ($5,000 Cloud Credits)	2020 — 2021
Dagstuhl Seminar 19051 - NSF Support Grant for Junior Researchers	2019
ICDE 2013 Student Travel Award	2013
Fudan Excellent Graduate	2009
SCSK Corporation Scholarship	2007 — 2008
Fudan Excellent Undergraduate Scholarship	2005 — 2009

Funded Research Grants

Sole PI, National Science Foundation (NSF), OAC-1755464 **6/1/2018 — 5/31/2022**

Title: CRII: OAC: Scalable Cyberinfrastructure for Big Graph and Matrix/Tensor Analytics
Amount: $170,941

PI, National Science Foundation (NSF), OAC-2106461 **10/1/2021 — 9/30/2024**

Title: Collaborative Research: OAC Core: Large-Scale Spatial Machine Learning for 3D Surface Topology in Hydrological Applications
Amount: $238,838
Note: with Zhe Jiang from the University of Florida as the PI from the leading institution

PI, Alabama Research and Development Enhancement Fund **1/15/2021 — 1/15/2023**
(ARDEF), Grant Number: 1ARDEF21 03

Title: Comprehensive Data Science Software Toolkit to Improve Alabama's Mobility Planning for Serving Businesses and Vulnerable Populations
Amount: $391,955
Note: with Virginia Sisiopiku as the Co-PI

Co-PI, National Science Foundation (NSF), DGE-1723250 **5/1/2017 — 7/31/2022**

Title: SaTC: EDU: Captivology-Stimuli-Based Learning (CAPITAL) for Big Data Security (BigSec): Towards a Science/Engineering, Career-Oriented Training

Amount: $43,848
Note: with Fei Hu (PI) and Debra Mccallum (Co-PI) from the University of Alabama

PI, UAB CAS Interdisciplinary Innovation Award **3/1/2020 — 2/28/2021**
Title: Planning Ahead for the Future Urban Transportation Challenges: A Data Science Solution
Amount: $30,000
Note: with Virginia Sisiopiku as the Co-PI

Co-PI, The Southeastern Transportation Research, Innovation, Development and Education (STRIDE) Center
Title: STRIDE I2: Mitigating Network Congestion by Integrating Transportation Network Companies and Urban Transit
Amount: $4,000 plus benefits (Summer 2020), $5,507 plus benefits (Summer 2019)

Invited Talks

Some Insights on Curriculum Design in Data Science and Big Data Related Courses, Data Science Education and Workforce Working Group Webinar Series, South Big Data Hub.	10/1/2021
Keynote II: Parallel Mining of Frequent Subtree Patterns, The 2nd International Workshop on Large Scale Graph Data Analytics In conjunction with VLDB 2020	9/4/2020
User-Friendly Distributed Frameworks for Processing Big Data, Stevens Institute of Technology, Hoboken, NJ.	12/23/2019
A Task-Centric Approach to Revolutionize Big Graph Systems Research, Dagstuhl Seminar 19491: Big Graph Processing Systems, Schloss Dagstuhl, Germany.	12/2/2019
Big Data Systems: Combining Ease of Programming with Performance, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN.	8/20/2019
Big Data Systems: Combining Ease of Programming with Performance, East China Normal University (ECNU), Shanghai, China.	5/22/2019
Expert Panel, The 16th Annual MidSouth Computational Biology and Bioinformatics Society Conference (MCBIOS 2019), Birmingham, AL.	3/30/2019
Mining Subgraphs from a Big Graph: Solution and Challenges, Dagstuhl Seminar 19051: Data Structures for the Cloud and External Memory Data, Schloss Dagstuhl, Germany.	1/30/2019

User-Friendly Distributed Frameworks for Processing Big Data, Queen's University Belfast, Belfast, Northern Ireland.	8/17/2018
T-thinker: A Task-Centric Framework to Revolutionize Big Data Systems Research, Computing Community Consortium (CCC) Early Career Researcher Symposium, Washington, D.C.	8/1/2018
User-Friendly Distributed Frameworks for Processing Big Data, International Forum on Artificial Intelligence and Trustworthy Software, East China Normal University (ECNU), Shanghai, China.	7/13/2018
User-Friendly Distributed Frameworks for Processing Big Data, Auburn University, Auburn, AL.	5/22/2018
Featured Panel: Future of Convergence: An International Perspective, Anniversary of the Society for Design and Process Science (SDPS), 2017.	11/7/2017
Poster: Bridging High Performance of HPC Community with Programming Friendliness of Data Science Community, The Computing Community Consortium (CCC) Symposium on Computing Research 2017, Washington, D.C.	10/23/2017
Big Data Frameworks: Bridging High Performance of HPC Community with Programming Friendliness of Data Science Community, Massachusetts Institute of Technology (MIT), Cambridge, MA.	9/21/2017
Systems for Big Graph Analytics (Alabama IEEE Computer Society Talk), The University of Alabama, Tuscaloosa, AL.	10/21/2016
Systems for Big Graph Analytics, City University of Hong Kong, Kowloon Tong, Hong Kong.	6/2/2016
Systems and Algorithms for Big Data Analytics, University of California, River- side (UCR), CA.	2/29/2016
Systems and Algorithms for Big Data Analytics, Michigan Technological University, Houghton, MI.	2/22/2016
Systems and Algorithms for Big Data Analytics, University of Notre Dame, Notre Dame, IN.	2/9/2016
Systems and Algorithms for Big Data Analytics, Hong Kong Baptist University (HKBU), Kowloon Tong, Hong Kong.	1/21/2016
Distributed Graph Computing: Algorithms and Systems, Sun Yat-Sen University (SYSU), Guangzhou, China.	11/14/2014

Distributed Graph Computing Beyond Pregel, Alibaba Xixi Park (Taobao City), Hangzhou, China.	5/23/2014
Distributed Graph Computing: Algorithms and Systems, National Institute of Informatics (NII), Tokyo, Japan.	5/8/2014
Query Processing in Spatial Databases, East China Normal University (ECNU), Shanghai, China.	12/2012

Conference Program Committee

ACM International Conference on Management of Data (SIGMOD)	2019, 2020, 2021
International Conference on Very Large Data Bases (VLDB)	2018 (Research), 2021 (Demo)
SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)	2020, 2021
IEEE International Conference on Data Engineering (ICDE)	2020, 2021, 2022
AAAI Conference on Artificial Intelligence (AAAI)	2021
International Joint Conference on Artificial Intelligence (IJCAI)	2017, 2021 (Senior PC), 2022 (IJCAI-ECAI)
International Symposium on Spatial and Temporal Databases (SSTD)	2021 (Webmaster)
International Workshop on Data Mining in Bioinformatics (BIOKDD) in Conjunction with KDD, Leading Program Co-Chair	2018, 2019, 2020, 2021
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)	2020, 2021
IEEE International Conference on Big Data (IEEE BigData)	2021
ACM International Conference on Information and Knowledge Management (CIKM)	2021 (Applied)
IEEE International Conference on Tools with Artificial Intelligence	2020
Annual MidSouth Computational Biology and Bioinformatics Society Conference (MCBIOS), Poster Chair	2019
Workshop on Data Mining in Biomedical Informatics and Healthcare (DMBIH) in Conjunction with ICDM	2019 (Co-organizer), 2020
International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD) in Conjunction with IEEE BigData	2019, 2020
IEEE Big Data Congress	2018, 2019
International Conference on Parallel Processing (ICPP)	2018

IEEE International Conference on Information Reuse and Integration for Data Science (IRI)	2018, 2019, 2020, 2021
International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP)	2017
4th International Workshop on Big Data Management and Service (BDMS 2017) with DASFAA	2017
2nd Annual International Conference on Information System and Artificial Intelligence (ISAI 2017)	2017
8th International Conference on Database Management Systems (DMS 2017)	2017
IEEE International Conference on Parallel and Distributed Systems (ICPADS)	2016
1st International Workshop on Graph Analytics and Query Processing (GAP 2016) in Conjunction with APWeb	2016

Journal Review and Guest Editor

BMC Supplements, Guest Co-Editor	2019 — 2020
IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), Guest Editor	2018 — 2022
ACM Transactions on Database Systems (TODS)	2017 — 2020
The VLDB Journal (VLDBJ)	2017 — 2020
IEEE Transactions on Parallel and Distributed Systems (TPDS)	2016 — 2020
IEEE Transactions on Knowledge and Data Engineering (TKDE)	2014 — 2020
ACM Computing Surveys (CSUR)	2020
ACM Transactions on Architecture and Code Optimization (TACO)	2020
Journal of Parallel and Distributed Computing (JPDC)	2020
IEEE Transactions on Big Data (TBD)	2020
ACM/IEEE Transactions on Networking (TON)	2020
IEEE Transactions on Network Science and Engineering (TNSE)	2018, 2019
World Wide Web Journal (WWWJ)	2015, 2016
IEEE Transactions on Cloud Computing (TCC)	2018
Knowledge and Information Systems (KAIS)	2015 — 2017
Information Systems (IS)	2018
Information Sciences	2018 — 2021

Future Generation Computer Systems (FGCS)	2017
Frontiers of Computer Science (FCS)	2016, 2017
Distributed and Parallel Databases (DAPD)	2016
Computational Intelligence	2016

Some journal review services have been tracked by Publons at:
https://publons.com/researcher/1442652/da-yan/

Grant Review

Since 2019, I serve in multiple NSF CISE panels each year. Per NSF requirement, the concrete panel information is not disclosed.

Event Chairing

- **Backup Session Chair of Demo Track in SSTD 2021**
- **Session Chair of Session SPT 4 in ICDE 2021**
- **Session Chair of Session 24 in SIGMOD 2020**
- **Session Chair of Session Co-Chair of SIGMOD 2020 Research Zoomtable on Spatial Data Management**
- **Session Chair of Session 09D Graph Algorithms 5 in VLDB 2020**
- **Session Chair of Session 14D Graph Algorithms 8 in VLDB 2020**
- **Co-organizer of Dagstuhl Seminar 22031 Bringing Graph Databases and Network Visualization Together**

Teaching

CS 467/667/767: Machine Learning	Fall 2017, Fall 2018, Fall 2019, Fall 2020, Fall 2021
CS 485/685/785: Foundations of Data Science	Spring 2017, Spring 2018, Spring 2019, Spring 2020, Spring 2021
CS 665/765: Deep Learning	Fall 2018
CS 334/534: Networking	Fall 2016, Fall 2017, Spring 2021

TA Duties at HKUST

COMP104: Programming Fundamentals and Methodology	Fall 2009

COMP152: Object-Oriented Programming and Data Structures	Spring 2010
COMP570: Introduction to Advanced Algorithmic Techniques	Fall 2010
COMP231: Database Management Systems	Spring 2011
COMP4431: Multimedia Computing	Fall 2012

Services at UAB

- **Computer Science judge for UAB-CORD Central Alabama Regional Science and Engineering Fair in 2018 — 2020**
- **Hosting summer intern students, Tianyi Miao (2018) and Le'Dederic Zellander**
- **Computer Science faculty representative on the UABTeach Steering Committee**
- **Department website update coordinator in 2016 — 2017 (during department name change from CIS to CS)**
- **Coordinator of visitors to the department for seminars and faculty candidate talks in 2017 — 2018**
- **Attending commencements and various on-campus and in-state events for introducing the CS programs**
- **Graduate student admissions committee where I help Dr. Nitesh Saxena and Dr. Ragib Hasan review PhD student applications, and help Dr. Chengcui Zhang review MSDS student applications**
- **Program committee of Informatics Institute's COVID-19 Data Science Hackathon**
- **Mentor of Team 6 in Informatics Institute's COVID-19 Data Science Hackathon (awarded second prize)**
- **PhD thesis committee for Xinpeng Liao from UAB Computer Science**
- **PhD thesis committee for Hadia M.R.M. Ahmed from UAB Computer Science**
- **PhD thesis committee for Ramin Goudarzi Karim from UAB Math**
- **PhD thesis committee for Sandeep Chowdary Vejandla from UAB Math**
- **PhD thesis committee for Chia-Min Lin from UAB Physics**
- **Master's thesis committee for Taniya Sultana from UAB Civil Engineering**
- **Master's thesis committee for Sahila Sarjana from UAB Civil Engineering**
- **Master's thesis committee for Orhun Vural from UAB Computer Science**
- **Special topic exam committee for David Hoxie from UAB Physics**
- **Participating in the Grand Challenge team organized by Dr. Janelle M. Chiasera**
- **Participating in planning meetings for NSF EPSCoR RII Track-1 proposal organized by Dr. Yogesh Vohra**

- **Participating in SoE Strategic Planning and New Business Development Committee organized by Dr. Lee Moradi**

Students

- **Guimu Guo**, PhD candidate (Fall 2017 — now), Alabama EPSCoR Graduate Research Scholar Program (GRSP) Awardee
- **Jalal Khalil**, PhD student (Fall 2019 — now)
- **Mirza Tanzim Sami**, PhD student (Spring 2020 — now)
- **Abdullateef Almudaifer**, PhD student (Spring 2020 — now)
- **Saugat Adhikari**, PhD student (Fall 2021 — now), UAB Blazer Fellowship Awardee
- **Bhadhan Joy**, MSDS student (thesis track)

Alumni

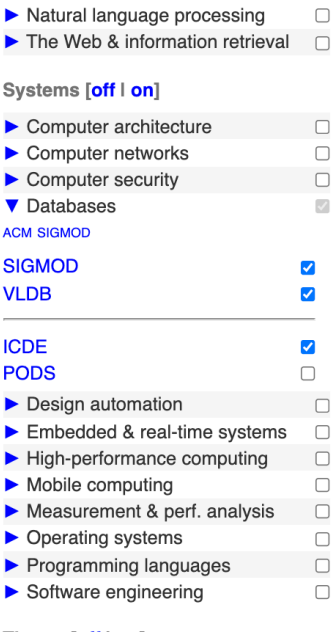
- **Wenwen Qu**, Summer Intern in 2019 from ECNU
- **Md Mashiur Rahman Chowdhury,** MSCS (Fall 2017 — Fall 2019), UAB Blazer Fellowship Awardee
- **Tianyi Miao**, Summer Intern in 2018 from Indian Springs School, now at UPenn for undergraduate study
- **Weida Tan**, **MSCS** (RA in 2016-2017), now at Fledging (as the founder)

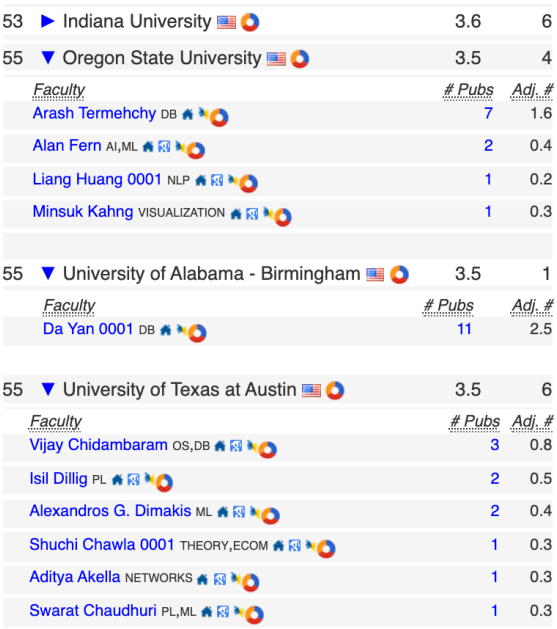
Accepted Publications

Publication Highlights

"In computer science, your preference should be for conference publication" as well explained by Prof. Michael Ernst from the University of Washington on his webpage. I frequently publish in the best conferences in my field, including VLDB (6 papers), ICDE (7 papers), KDD (4 papers), and others such as WWW, ICML, AAAI. These conferences are used to rank Computer Science departments all over the U.S. by CSRankings (http://csrankings.org/), which is an important point of reference by students when they choose schools to apply for graduate studies, especially the best students worldwide looking for Ph.D. study.

By publishing frequently in the best conferences in my major research area "Databases," I helped UAB Computer Science secure a nationwide area rank of #55 in Databases according to CSRankings (last visited on 08/31/2021):





I also frequently publish in other top conferences such as ICDM, SDM, SIGMOD, CIKM, EDBT, EMNLP, PPoPP, EuroSys, SoCC, etc., and publish in the best journals in my field including VLDB Journal, IEEE TKDE, IEEE TPDS, Communications of the ACM (CACM), ACM TKDD, ACM TIST, IEEE/ACM TCBB, etc.

Besides the following list of publications, you can also refer to my Google Scholar page or DBLP page for my publications.

Conference Full Papers

1. Zeru Zhang, Zijie Zhang, Yang Zhou, Lingfei Wu, Sixing Wu, Xiaoying Han, Dejing Dou, Tianshi Che, Da Yan, **"Adversarial Attack against Cross-lingual Knowledge Graph Alignment**," The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021), accepted and to appear.
2. Jahandad Pirayesh, Haiquan Chen, Xiao Qin, Wei-Shinn Ku, Da Yan, **"MentalSpot: Effective Early Screening for Depression Based on Social Contagion**," The 30th ACM International Conference on Information and Knowledge Management (CIKM 2021), accepted and to appear.
3. Bo Hui, Da Yan, Haiquan Chen, Wei-Shinn Ku, **"TrajNet: A Trajectory-Based Deep Learning Model for Traffic Prediction**," The 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2021), Pages 716-724.
4. Zhe Jiang, Wenchong He, Marcus Stephen Kirby, Sultan Asiri, Da Yan, **"Weakly Supervised Spatial Deep Learning based on Imperfect Vector Labels with Registration Errors**," The 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2021), Pages 767-775.
5. Xin Zhao, Zeru Zhang, Zijie Zhang, Lingfei Wu, Jiayin Jin, Yang Zhou, Ruoming Jin, Dejing Dou, Da Yan, **"Expressive 1-Lipschitz Neural Networks for Robust Multiple Graph Learning against Adversarial Attacks**," The 38th International Conference on Machine Learning (ICML 2021), Pages 12719-12735.
6. Bo Hui, Haiquan Chen, Da Yan, Wei-Shinn Ku, **"EDGE: Entity-Diffusion Gaussian Ensemble for Interpretable Tweet Geolocation Prediction**," The 37th IEEE International Conference on Data Engineering (ICDE 2021), Pages 1092-1103.
7. Wenchong He, Arpan Man Sainju, Zhe Jiang, Da Yan, **"Deep Neural Network for 3D Surface Segmentation based on Contour Tree Hierarchy**," The 21th SIAM International Conference on Data Mining (SDM 2021), Pages 253-261.
8. Yueen Ma, Da Yan, Cheng Long, D. Rangaprakash and Gopikrishna Deshpande, **"Predicting Autism Spectrum Disorder from Brain Imaging Data by Graph Convolutional Network**," The 2021 International Joint Conference on Neural Networks (IJCNN 2021), accepted and to appear.
9. Xingkun Yin, Da Yan, Abdullateef Almudaifer, Sibo Yan, Yang Zhou, **"Forecasting Stock Prices Using Stock Correlation Graph: A Graph Convolutional Network Approach**," The 2021 International Joint Conference on Neural Networks (IJCNN 2021), accepted and to appear.
10. Da Yan, Guimu Guo, Md Mashiur Rahman Chowdhury, M. Tamer Özsu, Wei-Shinn Ku, John C.S. Lui, **"G-thinker: A Distributed Framework for Mining Subgraphs in a Big Graph**,"

The 36th IEEE International Conference on Data Engineering (ICDE 2020), Pages 1369-1380.

11. Arpan Man Sainju, Wenchong He, Zhe Jiang, Da Yan, **"Spatial Classification with Limited Observations Based on Physics-Aware Structural Constraint**," The 34th AAAI Conference on Artificial Intelligence (AAAI 2020), Pages 898-905.
12. Bo Hui, Da Yan, Wei-Shinn Ku and Wenlu Wang, **"Predicting Economic Growth by Region Embedding: A Multigraph Convolutional Network Approach**," The 29th ACM International Conference on Information and Knowledge Management (CIKM 2020), Pages 555-564.
13. Yuechun Gu, Da Yan, Sibo Yan, Zhe Jiang., **"Price Forecast with High-Frequency Finance Data: An Autoregressive Recurrent Neural Network Model with Technical Indicators**," The 29th ACM International Conference on Information and Knowledge Management (CIKM 2020), Pages 2485-2492.
14. Yingzhe Dong, Da Yan, Abdullateef Almudaifer, Sibo Yan, Zhe Jiang, Yang Zhou, **"BELT: A Pipeline for Stock Price Prediction Using News**," 2020 IEEE International Conference on Big Data (IEEE BigData 2020), Pages 1137-1146.
15. Yang Zhou, Jiaxiang Ren, Ruoming Jin, Zijie Zhang, Dejing Dou, Da Yan, **"Unsupervised Multiple Network Alignment with Multinominal GAN and Variational Inference**," 2020 IEEE International Conference on Big Data (IEEE BigData 2020), Pages 868-877.
16. Xingyao Wang, Da Yan, Ke Chen, Yancong Deng, Cheng Long, Kunlin Zhang, Sibo Yan, **"Lane Extraction and Quality Evaluation: A Hough Transform Based Approach**," The 3rd IEEE Conference on Multimedia Information Processing and Retrieval (MIPR 2020 2020), Pages 7-12.
17. Ramin Goudarzi Karim, Guimu Guo, Da Yan, Carmeliza Navasca, **"Accurate Tensor Decomposition with Simultaneous Rank Approximation for Surveillance Videos**," The 54th Asilomar Conference on Signals, Systems, and Computers (ACSCC 2020), Pages 842-846.
18. Da Yan, James Cheng, Hongzhi Chen, Cheng Long, Purushotham Bangalore, **"Lightweight Fault Tolerance in Pregel-Like Systems**," The 48th International Conference on Parallel Processing (ICPP 2019), Article No. 69 Pages 1–10.
19. Yi Yang, Da Yan, Shuigeng Zhou, Guimu Guo, **"Parallel Clique-Like Subgraph Counting and Listing**," The 38th International Conference on Conceptual Modeling (ER 2019), Pages 484-497.
20. Ji Cheng, Da Yan, Xiaotian Hao, Wilfred Ng, **"Mining Order-Preserving Submatrices Under Data Uncertainty: A Possible-World Approach**," The 35th IEEE International Conference on Data Engineering (ICDE 2019), Pages 1154-1165.

21. Harry Kai-Ho Chan, Cheng Long, Da Yan, Raymond Chi-Wing Wong, **"Fraction-Score: A New Support Measure for Co-location Pattern Mining**," The 35th IEEE International Conference on Data Engineering (ICDE 2019), Pages 1514-1525.
22. Hongzhi Chen, Miao Liu, Yunjian Zhao, Xiao Yan, Da Yan, James Cheng, **"G-Miner: an efficient task-oriented graph mining system**," The 13th European Conference on Computer Systems (EuroSys 2018), Article No. 32 Pages 1–12.
23. Qizhen Zhang, Hongzhi Chen, Da Yan, James Cheng, Boon Thau Loo, Purushotham V. Bangalore, **"Architectural implications on the performance and cost of graph analytics systems**," The ACM Symposium on Cloud Computing (SoCC 2017), Pages 40-51.
24. Huanhuan Wu, Yunjian Zhao, James Cheng, **Da Yan**, **"Efficient Processing of Growing Temporal Graphs**," The 22nd International Conference on Database Systems for Advanced Applications (DASFAA 2017), Pages 387-403.
25. Yi Yang, Da Yan, Huanhuan Wu, James Cheng, Shuigeng Zhou, John C. S. Lui, **"Diversified Temporal Subgraph Pattern Mining**," The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016), Pages 1965-1974.
26. Cheng Chen, Hejun Wu, Dyce Jing Zhao, Da Yan, James Cheng, **"SGraph: A Distributed Streaming System for Processing Big Graphs**," The 2nd International Conference on Big Data Computing and Communications (BIGCOM 2016), Pages 285-294.
27. Da Yan, James Cheng, Yi Lu, Wilfred Ng, **"Effective Techniques for Message Reduction and Load Balancing in Distributed Graph Computation**," The 24th International Conference on World Wide Web (WWW 2015), Pages 1307-1317.
28. Huanhuan Wu, James Cheng, Yi Lu, Yiping Ke, Yuzhen Huang, Da Yan, Hejun Wu, **"Core decomposition in large temporal graphs**," 2015 IEEE International Conference on Big Data (IEEE BigData 2015), Pages 649-658.
29. Da Yan, James Cheng, Wilfred Ng, Steven Liu, **"Finding distance-preserving subgraphs in large road networks**," The 29th IEEE International Conference on Data Engineering (ICDE 2013), Pages 625-636.
30. Da Yan, Zhou Zhao, Wilfred Ng, **"Leveraging read rates of passive RFID tags for real-time indoor location tracking**," The 21st ACM International Conference on Information and Knowledge Management (CIKM 2012), Pages 375-384.
31. Da Yan, Zhou Zhao, Wilfred Ng, **"Monochromatic and bichromatic reverse nearest neighbor queries on land surfaces**," The 21st ACM International Conference on Information and Knowledge Management (CIKM 2012), Pages 942-951.
32. Zhou Zhao, Da Yan, Wilfred Ng, **"Mining probabilistically frequent sequential patterns in uncertain databases**," The 15th International Conference on Extending Database Technology (EDBT 2012), Pages 74-85.

33. Zhou Zhao, Da Yan, Wilfred Ng, **"A probabilistic convex hull query tool,"** The 15th International Conference on Extending Database Technology (EDBT 2012), Pages 570-573.
34. Da Yan, Raymond Chi-Wing Wong, Wilfred Ng, **"Efficient methods for finding influential locations with adaptive grids,"** The 20th ACM International Conference on Information and Knowledge Management (CIKM 2011), Pages 1475-1484.
35. Da Yan, Wilfred Ng, **"Robust Ranking of Uncertain Data,"** The 16th International Conference on Database Systems for Advanced Applications (DASFAA 2011), Pages 254-268. **Best Paper Award (Sole Winner).**

Conference Short Papers

1. Bo Hui, Da Yan, Haiquan Chen, Wei-Shinn Ku, **"Trajectory WaveNet: A Trajectory-Based Model for Traffic Forecasting,"** The 21st IEEE International Conference on Data Mining (ICDM 2021), accepted and to appear.
2. Da Yan, Wenwen Qu, Guimu Guo, Xiaoling Wang, **"PrefixFPM: A Parallel Framework for General-Purpose Frequent Pattern Mining,"** The 36th IEEE International Conference on Data Engineering (ICDE 2020), Pages 1938-1941.
3. Da Yan, Hongzhi Chen, James Cheng, Zhenkun Cai, Bin Shao, **"Scalable De Novo Genome Assembly Using Pregel,"** The 34th IEEE International Conference on Data Engineering (ICDE 2018), Pages 1216-1219.

Journal Papers

1. Wenchong He, Arpan Man Sainju, Zhe Jiang, Da Yan, Yang Zhou, **"Earth Imagery Segmentation on Terrain Surface with Limited Training Labels: A Semi-supervised Approach based on Physics-Guided Graph Co-Training,"** ACM Transactions on Intelligent Systems and Technology (TIST), accepted in 2021 and to appear.
2. Da Yan, Guimu Guo, Jalal Khalil, M. Tamer Ozsu, Wei-Shinn Ku, John C.S. Lui, **"G-thinker: A General Distributed Framework for Finding Qualified Subgraphs in a Big Graph with Load Balancing,"** VLDB Journal (VLDBJ), accepted in 2021 and to appear.
3. Da Yan, Wenwen Qu, Guimu Guo, Xiaoling Wang, Yang Zhou, **"PrefixFPM: A Parallel Framework for General-Purpose Mining of Frequent and Closed Patterns,"** VLDB Journal (VLDBJ), accepted in 2021 and to appear.
4. Wei-Chih Chen, Joanna N. Schmidt, Da Yan, Yogesh K. Vohra, Cheng-Chien Chen, "Machine learning and evolutionary prediction of superhard B-C-N compounds. npj Computational Materials 7, 114 (2021). https://doi.org/10.1038/s41524-021-00585-7

5. Guimu Guo, Hongzhi Chen, Da Yan, James Cheng, Jake Y. Chen, Zechen Chong, **"Scalable De Novo Genome Assembly Using a Pregel-Like Graph-Parallel System,"** IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 18(2): 731-744 (2021).
6. Yang Zhou, Jiaxiang Ren, Ruoming Jin, Zijie Zhang, Jingyi Zheng, Zhe Jiang, Da Yan, Dejing Dou, **"Unsupervised Adversarial Network Alignment with Reinforcement Learning,"** ACM Transactions on Knowledge Discovery from Data (TKDD), accepted in 2021 and to appear.
7. Arpan Man Sainju, Wenchong He, Zhe Jiang, Da Yan, Haiquan Chen, **"Flood Inundation Mapping with Limited Observations Based on Physics-Aware Topography Constraint,"** Frontiers in Big Data 4: 707951 (2021).
8. Sherif Sakr, Angela Bonifati, Hannes Voigt, Alexandru Iosup, Khaled Ammar, Renzo Angles, Walid G. Aref, Marcelo Arenas, Maciej Besta, Peter A. Boncz, Khuzaima Daudjee, Emanuele Della Valle, Stefania Dumbrava, Olaf Hartig, Bernhard Haslhofer, Tim Hegeman, Jan Hidders, Katja Hose, Adriana Iamnitchi, Vasiliki Kalavri, Hugo Kapp, Wim Martens, M. Tamer Özsu, Eric Peukert, Stefan Plantikow, Mohamed Ragab, Matei Ripeanu, Semih Salihoglu, Christian Schulz, Petra Selmer, Juan F. Sequeda, Joshua Shinavier, Gábor Szárnyas, Riccardo Tommasini, Antonino Tumeo, Alexandru Uta, Ana Lucia Varbanescu, Hsiang-Yun Wu, Nikolay Yakovets, Da Yan, Eiko Yoneki, **"The Future is Big Graphs! A Community View on Graph Processing Systems,"** Communications of the ACM (CACM), September 2021, Vol. 64 No. 9, Pages 62-71.
9. Guimu Guo, Da Yan, M. Tamer Özsu, Zhe Jiang, Jalal Khalil, **"Scalable Mining of Maximal Quasi-Cliques: An Algorithm-System Codesign Approach,"** Proceedings of the VLDB Endowment (PVLDB) 14(4): 573-585 (2020).
10. Da Yan, Yuzhen Huang, Miao Liu, Hongzhi Chen, James Cheng, Huanhuan Wu, Chengcui Zhang, **"GraphD: Distributed Vertex-Centric Graph Processing Beyond the Memory Limit,"** IEEE Transactions on Parallel and Distributed Systems (TPDS) 29(1): 99-114 (2018).
11. Fanhua Shang, Yuanyuan Liu, James Cheng, Da Yan, **"Fuzzy Double Trace Norm Minimization for Recommendation Systems,"** IEEE Transactions on Fuzzy Systems 26(4): 2039-2049 (2018).
12. Da Yan, James Cheng, M. Tamer Özsu, Fan Yang, Yi Lu, John C. S. Lui, Qizhen Zhang, Wilfred Ng, **"A General-Purpose Query-Centric Framework for Querying Big Graphs,"** Proceedings of the VLDB Endowment (PVLDB) 9(7): 564-575 (2016).

13. Da Yan, Zhou Zhao, Wilfred Ng, Steven Liu, **"Probabilistic Convex Hull Queries over Uncertain Data**," IEEE Transactions on Knowledge and Data Engineering (TKDE) 27(3): 852-865 (2015).
14. Da Yan, Zhou Zhao, Wilfred Ng, **"Efficient processing of optimal meeting point queries in Euclidean space and road networks**," Knowledge and Information Systems (KAIS) 42(2): 319-351 (2015).
15. Da Yan, James Cheng, Zhou Zhao, Wilfred Ng, **"Efficient location-based search of trajectories with location importance**," Knowledge and Information Systems (KAIS) 45(1): 215-245 (2015).
16. Da Yan, James Cheng, Kai Xing, Yi Lu, Wilfred Ng, Yingyi Bu, **"Pregel Algorithms for Graph Connectivity Problems with Performance Guarantees**," Proceedings of the VLDB Endowment (PVLDB) 7(14): 1821-1832 (2014).
17. Da Yan, James Cheng, Yi Lu, Wilfred Ng: Blogel, **"A Block-Centric Framework for Distributed Computation on Real-World Graphs**," Proceedings of the VLDB Endowment (PVLDB) 7(14): 1981-1992 (2014)
18. Yi Lu, James Cheng, Da Yan, Huanhuan Wu, **"Large-Scale Distributed Graph Computing Systems: An Experimental Evaluation**," Proceedings of the VLDB Endowment (PVLDB) 8(3): 281-292 (2014).
19. Zhou Zhao, Da Yan, Wilfred Ng, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases," IEEE Transactions on Knowledge and Data Engineering (TKDE) 26(5): 1171-1184 (2014).
20. Da Yan, Zhou Zhao, Wilfred Ng, "Efficient Algorithms for Finding Optimal Meeting Point on Road Networks," Proceedings of the VLDB Endowment (PVLDB) 4(11): 968-979 (2011).

Conference Poster Papers

1. Ji Cheng, Guimu Guo, Da Yan, Xiaotian Hao, Wilfred Ng, **"EasyRain: A User-Friendly Platform for Comparing Precipitation Nowcasting Models**," 2019 IEEE International Conference on Big Data (IEEE BigData 2019), Pages 6019-6021.
2. Guimu Guo, Jalal Khalil, Da Yan, Virginia P. Sisiopiku, **"Realistic Transport Simulation with Open Data**," 2019 IEEE International Conference on Big Data (IEEE BigData 2019), pages 6066-6068.
3. Da Yan, Guimu Guo, Md Mashiur Rahman Chowdhury, M. Tamer Özsu, John C. S. Lui, Weida Tan, **"T-thinker: a task-centric distributed framework for compute-intensive divide-and-conquer algorithms**," The 24th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP 2019), Pages 411-412.

Conference Demo Papers

1. Qizhen Zhang, Da Yan, James Cheng, **"Quegel: A General-Purpose System for Querying Big Graphs**," The 2016 International Conference on Management of Data (SIGMOD 2016), Pages 2189-2192.
2. Zhou Zhao, Da Yan, Wilfred Ng, Shi Gao, **"A transfer learning based framework of crowd-selection on Twitter**," The 19nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013), Pages 1514-1517.

Workshop Papers

1. Wenwen Qu, Da Yan, Guimu Guo, Xiaoling Wang, Lei Zou, Yang Zhou, **"Parallel Mining of Frequent Subtree Patterns**," The 2nd International Workshop on Large Scale Graph Data Analytics (LSGDA) held in Conjunction with VLDB 2020, published in Communications in Computer and Information Science 1281, Springer 2020, ISBN 978-3-030-61132-3.
2. Guimu Guo, Jalal Khalil, Da Yan, Virginia P. Sisiopiku, **"Realistic Transport Simulation: Tackling the Small Data Challenge with Open Data**," The 1st International Workshop on Big Data Tools, Methods, and Use Cases for Innovative Scientific Discovery (BTSD) held in Conjunction with IEEE BigData 2019, pages 4512-4519.

Conference Tutorials

1. Guimu Guo, Da Yan, **"Systems and Algorithms for Massively Parallel Graph Mining**," Tutorial 3, 2020 IEEE International Conference on Big Data (IEEE BigData 2020).
2. Da Yan, Yingyi Bu, Yuanyuan Tian, Amol Deshpande, James Cheng, **"Big Graph Analytics Systems**," The 2016 International Conference on Management of Data (SIGMOD 2016), Pages 2241-2243.

Books

1. Da Yan, Yingyi Bu, Yuanyuan Tian, Amol Deshpande, **"Big Graph Analytics Platforms**," Foundations and Trends in Databases 7(1-2): 1-195 (2017).
2. Da Yan, Yuanyuan Tian, James Cheng, **"Systems for Big Graph Analytics**," Springer Briefs in Computer Science, Springer 2017, ISBN 978-3-319-58216-0, pp. 1-92.

Book Chapters

1. Zongliang Yue, Da Yan, Guimu Guo, Jake Chen, **"Biological Network Mining,"** Modeling Transcriptional Regulation - Methods and Protocols, Methods in Molecular Biology – Springer, 2021.
2. Da Yan, Hang Liu, **"Parallel Graph Processing,"** Encyclopedia of Big Data Technologies 2019.
3. Sibo Yan, Da Yan, **"Volatility Estimation in the Era of High-Frequency Finance,"** FinTech as a Disruptive Technology for Financial Institutions, IGI Global, 2019.

Editorials

1. Da Yan, Steve Qin, Debswapna Bhattacharya, Jake Y. Chen, Mohammed J. Zaki, **"20th International Workshop on Data Mining in Bioinformatics (BIOKDD 2021),"** The 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD 2021), Pages 4175-4176.
2. Da Yan, Sharma V. Thankachan, Jake Y. Chen, **"Guest Editorial for Selected Papers from BIOKDD 2019,"** IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 18(3): 809-810 (2021).
3. Jonathan D. Wren, Yongsheng Bai, Zhaohui S. Qin, Da Yan, Ramin Homayouni, **"Proceedings of the 2019 MidSouth Computational Biology and Bioinformatics Society (MCBIOS) Conference,"** BMC Bioinformatics 21-S(4): 254 (2020).
4. Da Yan, Xin Gao, Samah J. Fodeh, Jake Y. Chen, **"Guest Editorial for Selected Papers from BIOKDD 2018 and DMBIH 2018,"** IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 17(6): 1832-1834 (2020).