

Curriculum Vitae

Jian Pei, Ph.D.

Arthur S. Pearce Distinguished Professor of Computer Science

Chair, Department of Computer Science, Trinity College of Arts & Sciences
Professor, Department of Biostatistics and Bioinformatics, School of Medicine
Professor, Department of Electric and Computer Engineering, Pratt School of Engineering

Duke University

E-mail: j.pei@duke.edu
(As of on November 20, 2023)

Contents

1	Education	3
2	Work Experience	3
2.1	Courtesy appointments	4
3	Awards and Distinctions	4
3.1	Major Awards	4
3.2	Best Paper Awards	5
3.3	Others	5
4	Research Interests	6
5	Teaching Experience	6
5.1	Regular Courses	6
5.2	Special Courses	7
6	Grants	8
7	Publications	12
7.1	Textbook	12
7.2	Monographs	12
7.3	Research Articles in Refereed Journals	12
7.4	Research Articles in Refereed Conference Proceedings	19
7.5	Research Articles in Refereed Workshop Proceedings	36
7.6	Tutorials	37
7.7	Book Chapters	39

7.8	Demonstrations	39
7.9	Patents	40
7.10	Edited Books, Proceedings, and Special Issues in Journals and Magazines	40
8	Keynote Speeches in Conferences and Invited Talks (Since September 1, 2022)	41
9	Training of Highly Qualified Personnel (HQP)	42
9.1	Current graduate students	42
9.2	Former graduate students	43
10	Professional Memberships and Activities	45
10.1	Membership of professional associations	45
10.2	Membership on the editorial boards	45
10.3	Guest-editorship	46
10.4	Leadership in professional organizations, membership of advisory boards, steering committees, councils, and special committees	46
10.5	Members of organizing committees	47
10.6	Member of international conference program committees	53
10.7	Internal Services	59

1 Education

- **Ph.D. on Computing Science.** School of Computing Science, Simon Fraser University, Burnaby, Canada. (Ph.D. thesis defended on May 22, 2002).
 - Thesis title: Pattern-growth Methods for Frequent Pattern Mining
 - Senior supervisor: Dr. Jiawei Han, Professor, ACM Fellow, IEEE Fellow, the recipient of ACM SIGKDD 2004 Innovation Award and IEEE ICDM 2002 Outstanding Contribution Award
- September 1997 - April 1999, Ph.D. student in Department of Computer Science and Technology, Peking University, Beijing, China. Research directions: data mining, knowledge discovery, and data warehousing
- **Master of Engineering in Computer Science,** Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, August 1993.
- **Bachelor of Engineering in Computer Science,** Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, August 1991.

2 Work Experience

- July 1, 2023 - present, Chair, Department of Computer Science, Duke University.
- July 1, 2023 - present, Arthur S. Pearse Distinguished Professor of Computer Science (with tenure), Trinity College of Arts & Sciences; Professor of Department of Biostatistics and Bioinformatics, School of Medicine; Professor of Department of Electric and Computer Engineering, Pratt School of Engineering, Duke University.
- January 1, 2023 - June 30, 2023, Professor (with tenure), Department of Computer Science, Trinity College of Arts & Sciences; Department of Biostatistics and Bioinformatics, School of Medicine; Department of Electric and Computer Engineering, Pratt School of Engineering, Duke University.
- July 10, 2022 - December 31, 2022, Research professor, Department of Computer Science, Trinity College of Arts & Sciences; Department of Biostatistics and Bioinformatics, School of Medicine; Department of Electric and Computer Engineering, Pratt School of Engineering, Duke University.
- September 1, 2011 - present, Professor (with tenure), School of Computing Science, Simon Fraser University.
 - July 1, 2022 - now, no pay leave.
 - January 1, 2015 - August 31, 2022, Associate Member, Department of Statistics and Actuarial Science, Simon Fraser University.
 - June 1, 2016 - May 31, 2018, Associate Member, Faculty of Health Sciences, Simon Fraser University.
- May 1, 2014 - August 31, 2017, Canada Research Chair (Tier 1) in Big Data Science. (The position was terminated due to my extended no-pay leave from the university following the regulation of NSERC.)
- October 2015 - August 31, 2017, Founding Director, Pacific Blue Cross Health Informatics Laboratory, Simon Fraser University.
- September 1, 2008 - August 31, 2011, Associate Professor (with tenure), School of Computing Science, Simon Fraser University.
 - September 1, 2009 - August 31, 2011, Associate Director, Research and Industrial Relations, School of Computing Science, Simon Fraser University.
 - January 1, 2009 - August 31, 2011, Director, Collaborative Research and Industry Relations.
- April 16, 2008 - August 31, 2008, Assistant Professor (with tenure), School of Computing Science, Simon Fraser University.

- August 1, 2004 - April 15, 2008, Assistant Professor (tenure track), School of Computing Science, Simon Fraser University.
- August 1, 2002 - July 31, 2004, Assistant Professor (tenure track), Department of Computer Science and Engineering, University at Buffalo, the State University of New York.
 - November, 2003 - July, 2004, participating faculty, the Center of Unified Biometrics and Sensors (CUBS), University at Buffalo, the State University of New York.
- September, 1999 - August, 2002, research assistant in School of Computing Science, Simon Fraser University.
- September, 2001 - April, 2002, sessional instructor in School of Computing Science, Simon Fraser University.
- April 1999, - August, 1999, research associate in School of Computing Science, Simon Fraser University.
- September, 1993 - August, 1997, staff, manager, project manager, China Construction Bank.

2.1 Courtesy appointments

- September 1, 2022 - August 31, 2027, Adjunct Professor, Department of Statistics and Actuarial Science, Simon Fraser University.
- October 2015 - September 2018, Distinguished Visiting Professor, Tsinghua University, China.
- July 1, 2015 - June 30, 2020, Adjunct Professor, Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada
- April 1, 2015 - December 31, 2016, Distinguished Adjunct Professor, King Abdulaziz University, Saudi Arabia.
- April 1, 2013 - March 31, 2016, Chao Kuang-Piu Chair Professor (professorship by courtesy), School of Computer Science and Technology, Zhejiang University, China.

3 Awards and Distinctions

3.1 Major Awards

- *Fellow of the Canadian Academy of Engineering (CAE)* (Class 2020) for “his seminal contributions to the foundation of data mining, data analytics and applications.”
- *Fellow of the Royal Society of Canada (RSC)* (Class 2019) for “his seminal contributions to the foundation of data mining, data analytics and applications.” Recognition by the RSC is the highest honour an individual can achieve in the Arts, Social Sciences and Sciences in Canada.
- *The 2017 ACM SIGKDD Innovation Award* “for his seminal contributions to the foundation of data mining and applications, particularly in pattern mining and spatial data mining.” The ACM SIGKDD Innovation Award is “the highest award for technical excellence in the field of Knowledge Discovery and Data Mining (KDD)”, and “is conferred on one individual or one group of collaborators whose outstanding technical innovations in the KDD field have had a lasting impact in advancing the theory and practice of the field. The contributions must have significantly influenced the direction of research and development of the field or transferred to practice in significant and innovative ways and/or enabled the development of commercial systems.”
- *Fellow of the Association of Computing Machinery (ACM)* for “contributions to the foundation, methodology and applications of data mining” (elevated on December 8, 2015).
- *The 2015 ACM SIGKDD Service Award* “for his significant technical contributions to the principles, practice and application of data mining and for his outstanding services to society and the data mining community.” The ACM SIGKDD Service Award “is the highest service award in the field of knowledge discovery and data mining (KDD), and is conferred on one individual or one group for their outstanding professional services and contributions to the field of knowledge discovery and data mining”.

- *The 2014 IEEE ICDM Research Contributions Award* “for his work on the core frontiers of data mining, including pattern mining, classification, clustering, anomaly detection and outlier analysis.” The IEEE ICDM Research Contributions Award “is the highest recognition for research achievements in Data Mining, and is given to one individual or one group who has made influential research contributions to the field of Data Mining”.
- *Fellow of the Institute of Electrical and Electronics Engineers (IEEE)* for “contributions to data mining and knowledge discovery” (elevated on January 1, 2014).
- Awarded one of *NSERC’s one hundred 2008 Discovery Accelerator Supplements (DAS)*.
- *2005 British Columbia Innovation Council Young Innovator Award*.

3.2 Best Paper Awards

- *Influential Paper Award*, the Thirty-fourth IEEE International Conference on Data Engineering (ICDE’18), Paris, France, April 16-19, 2018. This award is to recognize the impact of the paper “Preserving Privacy in Social Networks Against Neighborhood Attacks”, by B. Zhou and J. Pei, in Proceedings of the Twenty-fourth IEEE International Conference on Data Engineering (ICDE’08), Cancun, Mexico, April 7-12, 2008.
- *Best Student Paper Award* for the paper “Preference-driven Similarity Join”, the 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI’17), Leipzig, Germany, August 23-26, 2017.
- *Best Paper Award* for the paper “Mining Contrast Subspaces”, the Eighteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’14), Tainan, Taiwan, May 13-16, 2014.
- *The Most Influential Paper Award*, the Thirteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’09), Bangkok, Thailand, April 27-30, 2009. This award is to recognize the impact of the paper “Mining Access Patterns efficiently from Web logs” by J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu, in Proceedings of the 2000 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’00), Kyoto, Japan, April 2000.
- *Best Application Paper Award* for the paper “Context-Aware Query Suggestion by Mining Click-Through and Session Data”, the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), August 24-27, 2008, Las Vegas, NV, USA.
- *IEEE Outstanding Paper Award* of the IEEE Twenty-First International Conference on Advanced Information Networking and Applications (AINA’07) Niagara Falls, Canada, May 21-23, 2007.
- *Best Applications Paper Award Runner-up*, the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’04), August 22-25, 2004, Seattle, Washington, USA.

3.3 Others

- *ACM Distinguished Speaker* (September 2011 - August 2015).
- *The 2013 Outstanding Service Award*, ACM International Conference on Information and Knowledge Management (CIKM), October 30, 2013.
- *The 2010 Faculty of Applied Science Medals of Excellence Award: Superior Performance in Research by an Early Career Faculty Member*, Simon Fraser University.
- *President’s Ph.D. Research Stipend* from Simon Fraser University, in the 2001/2002 academic year for the Spring semester.
- *Graduate Fellowship* from Dean of Graduate Studies, Simon Fraser University, in the 2000/2001 academic year for the Spring semester, and in the 2001/2002 academic year for the Fall semester.
- *ASI Exchange 2002 Excellent Poster Award – 2nd Prize*, March 27th, 2002, Centre for Systems Science, Simon Fraser University.
- *Ebco/Epic Graduate Scholarships in Expert Systems* for academic years 2001/2002 and 2000/2001, Simon Fraser University.

- *Research & Development Award in Data Mining Systems* from Intelligent Database Systems Research Laboratory, Simon Fraser University.
- *Student Travel Award* from the Sixth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'2000), August 20-23, 2000, Boston, MA, USA.
- *Student Travel Scholarship* from the 2001 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'2001), May 20th, 2001, Santa Barbara, CA, USA.
- *ASI Exchange'2000 Communication Award*, March 14th, 2000, British Columbia, Canada.

4 Research Interests

- Data science
- Big data processing and analytics
- Data mining and knowledge discovery
- Data lakes, data warehousing, enterprise business intelligence, and online analytic processing (OLAP)
- Information retrieval and Web search, social networks, social media, and social search
- Health-informatics and bioinformatics
- Database systems
- Applied machine learning and artificial intelligence

5 Teaching Experience

5.1 Regular Courses

- *Instructor of CompSci-590 Introduction to Data Science*, Spring 2023 at Duke University. It is a graduate course.
- *Instructor of ECE-590 Algorithms*, Fall 2022 at Duke University. It is a graduate course.
- *Instructor of CMPT-843 Database and Knowledge-base Systems*, Summer 2007, Fall 2008, Fall 2009, Spring 2013, Spring 2015 and Fall 2019, at Simon Fraser University. It is a graduate course.
- *Instructor of CMPT-741 Foundations of Data Mining*, Fall 2007, Fall 2005, Fall 2015 and Summer 2021, at Simon Fraser University. It is a graduate course.
- *Instructor of CMPT-740 Database Systems*, Summer 2005, at Simon Fraser University. It is a graduate course.
- *Instructor of CMPT-459 Special Topics on Database Systems: Data Mining*, Fall 2007, Fall 2015, Summer 2020, and Summer 2021, at Simon Fraser University. It is an undergraduate course.
- *Instructor of CMPT-456 Information Retrieval and Web Search*, Fall 2008, Spring 2013, Fall 2019, and Summer 2020, at Simon Fraser University, Fall 2010, at the Simon Fraser University - Zhejiang University Joint Dual Degree Program. It is an undergraduate course.
- *Instructor of CMPT-454 Database Systems II*, Spring 2005, Fall 2005, and Spring 2008, at Simon Fraser University. It is an undergraduate course.
- *Instructor of CMPT-354 Database Systems I*, Fall 2021, Fall 2007, Fall 2006 (2 sessions) and Summer 2005, at Simon Fraser University. It is an undergraduate course.
- *Instructor of CMPT-225 Data Structures and Programming*, Fall 2010, at the Simon Fraser University - Zhejiang University Joint Dual Degree Program. It is an undergraduate course.
- *Instructor of CSE-626 Data Mining*, Fall 2002 and Fall 2003, at State University of New York at Buffalo. It is a graduate course.

- *Instructor of CSE-462 Database Concepts*, Fall 2003, at State University of New York at Buffalo. It is an undergraduate course.
- *Instructor of CSE-711 Seminars: Research topics in Data and Information Management*, Spring 2003 and Spring 2004, at State University of New York at Buffalo. It is a graduate course.
- *Instructor of CMPT-300 Operating Systems I*, Fall 2001 and Spring 2002, at Simon Fraser University. It is an undergraduate course.
- *Teaching Assistant of CMPT-882 Machine Learning*, Spring 2002, at Simon Fraser University. It is a graduate course.
- *Teaching Assistant of CMPT-459 Special Topics on Database Systems: Data Mining and Data Warehousing*, Fall 2000, at Simon Fraser University. It is an undergraduate course.

5.2 Special Courses

- *ACM Distinguished Speaker* (September 2011 - August 2015).
- *Instructor of the 2015 Dragon Star Program Course “Big Data Analytics: Statistical and Data Mining Approaches”*, June 29-July 3, 2015, at the University of Science and Technology China, Hefei, China. It is a 30 hours graduate course. The Dragon Star Program (sponsored by the Natural Science Foundation of China and Chinese Academy of Science) invites selected Chinese professors in US and Canada to teach advanced graduate courses in China.
- *Instructor of the 2014 Dragon Star Program Course “Big Data Analytics and Data Mining”*, May 5-9, 2014, at Institute of Computing Technology, Chinese Academy of Science, Beijing, China. It is a 30 hours graduate course. The Dragon Star Program (sponsored by the Natural Science Foundation of China and Chinese Academy of Science) invites selected Chinese professors in US and Canada to teach advanced graduate courses in China.
- *An instructor of the 2012 Dragon Star Program Course “Data Mining”*, May 21-25, 2012, at Renmin University, Beijing, China. It is a 25 hours graduate course. The Dragon Star Program (sponsored by the Natural Science Foundation of China and Chinese Academy of Science) invites selected Chinese professors in US and Canada to teach advanced graduate courses in China. The other two instructor of this course are Professors Jiawei Han and Qiang Yang.
- *Instructor of the 2009 Dragon Star Program Course “Data Mining”*, August 10-14, 2009, at Zhejiang University, Hangzhou, China. It is a 25 hours graduate course. The Dragon Star Program (sponsored by the Natural Science Foundation of China and Chinese Academy of Science) invites selected Chinese professors in US and Canada to teach advanced graduate courses in China.
- *Instructor of Special Graduate Short Course “Modeling, Querying and Mining Uncertain and Probabilistic Data”*, April 7-10, 2009, at Renmin University, China. This is a 24 hours short course invited by Renmin University, China.
- *Instructor of the 2008 VLDB School Course “Query answering techniques on uncertain and probabilistic data”*, July 23, 2008, at Zhangjiajie, China. It is a 6 hours short course invited by the VLDB Endowment.
- *Instructor of the 2006 Dragon Star Program Course “Data Mining”*, March 20-24, 2006, at Zhong Shan University, Guangzhou, China. It is a 25 hours graduate course. The Dragon Star Program (sponsored by the Natural Science Foundation of China and Chinese Academy of Science) invites selected Chinese professors in US and Canada to teach advanced graduate courses in China. More than 200 attendants registered the course.
- *Instructor of the 2003 Dragon Star Program Course “Data Mining”*, December 15-19, 2003, at Shanghai Jiao Tong University, Shanghai, China. It is a 30 hours graduate course. The Dragon Star Program (sponsored by the Natural Science Foundation of China and Chinese Academy of Science) invites selected Chinese professors in US and Canada to teach advanced graduate courses in China. More than 60 attendants registered the course.

6 Grants

Source	Title	Amount	Duration	My role
National Science Foundation (NSF) Early-concept Grants for Exploratory Research (EAGER) Program	End-to-End Learning of Paradoxes and Interpretations for Data Storytelling	US\$125,000	2023-2024	PI (100%)
The Pratt School of Engineering at Duke University the Beyond the Horizon Initiative	A Duke Initiative on Data and AI Model Sharing, Discovery, and Integration in Markets (DAIMS ²)	US\$150,000	2023-2024	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program	Querying and Mining Dynamics in Evolving Graphs and Networks	CA\$320,000	2020-2025	PI (100%)
National Research Council Canada (NRC) New Beginnings Initiative	Methods and Algorithms for Unsupervised Domain Adaptation	CA\$43,300	2021-2022	External collaborator (57%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Canada Research Chair Program	Connecting Big Data with Many People: a Context-aware On-demand Analytics Approach (early termination due to no-pay leave)	CA\$1,400,000	2014-2018	Chair holder (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program	Querying Dynamics in Evolving Graphs and Networks (early termination due to no-pay leave)	CA\$210,000	2017-2018	PI (100%)
DND/NSERC Discovery Grant Supplement	Querying Dynamics in Evolving Graphs and Networks	CA\$120,000	2017-2018	PI (100%)
Natural Sciences and Engineering Research Council (NSERC) Collaborative Research and Development Grants	Entity augmentation and data cleaning for machine learning	CA\$180,000	2017-2018	Co-PI (15%)
Vancity Vancouver City Savings Credit Union	Entity augmentation and data cleaning for machine learning	CA\$90,000	2017-2018	Co-PI (15%)
LifeLabs Medical Laboratories	Health informatics research	CA\$300,000	2016-2018	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Strategic Partnership Grants	Fraud investigation at scale: Methods and tools	CA\$495,000	2016-2018	PI (66.7%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Research Tools and Instruments Grants Program	Computational Infrastructure for Online Big Data Analytics	CA\$148,408	2016-2017	PI

Source	Title	Amount	Duration	My role
Pacific Blue Cross Inc.	Pacific Blue Cross Health Informatics Laboratory	CA\$300,000	2015-2018	Founding director (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development (CRD) Grants	Optimizing Hadoop to Scale to Big Systems and Big-Data	CA\$237,494	2016-2018	co-PI (50%)
Simba Technology Inc.	Optimizing Hadoop to Scale to Big Systems and Big-Data	CA\$90,505	2016-2018	co-PI (50%)
Huawei Technologies Co., Ltd.	Mining Big Sequence Data: from Patterns to Application Solutions	CA\$261,063	2016-2017	PI (100%)
Yahoo! Faculty Research and Engagement Program (FREP) award	Mining contextual patterns for mobile recommendation	US\$15,000	2015	PI (100%)
Simba Technologies Inc.	Optimizing Hadoop to Scale to Big Systems and Big-Data	CA\$50,000	2014-2015	Co-PI (70%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program	Towards Context-aware On-demand Data Mining	CA\$245,000	2011-2016	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) ENGAGE Grants Program	Building a Bridge between SQL-based Applications and Big Data Infrastructure	CA\$25,000	2014	PI (100%)
MITACS Accelerate Internship	Efficient OLAP Queries on In-memory Databases with Frequent Updates	CA\$30,000	2014	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) ENGAGE Grants Program	Complex Data Queries during Local Search Optimization	CA\$25,000	2014	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) ENGAGE Grants Program	Data Mining Applications in Network Security Intelligence	CA\$25,000	2012	PI (100%)
BCFRST Foundation Natural Resources and Applied Sciences (NRAS) Endowment Research Team Program	Interactive Searching and Mining of Large Evolving Graphs	CA\$300,000	2011-2014	PI (75%)
Graphics, Animation and New Media (GRAND), Networks of Centres of Excellence of Canada (NCE)	Access to New Media (NEWS) – Data Mining subproject	CA\$18,000	2012-2013	PI (100%)

Source	Title	Amount	Duration	My role
Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development (CRD) Grants	Aggregate Keyword Search in Relational Data	CA\$40,000	2011-2012	PI (100%)
SAP Business Objects Academic Research Center Project	Aggregate Keyword Search in Relational Data	CA\$20,000	2011-2012	PI (100%)
Graphics, Animation and New Media (GRAND), Networks of Centres of Excellence of Canada (NCE)	Access to New Media (NEWS) – Data Mining subproject	CA\$20,000	2011-2012	PI (100%)
MITACS Globalink Program	Personal Social Network Management	CA\$5,000	2011	PI (100%)
Graphics, Animation and New Media (GRAND), Networks of Centres of Excellence of Canada (NCE)	Access to New Media (NEWS) – Data Mining subproject	CA\$29,750	2010-2011	PI (100%)
NSERC Business Intelligence Strategic Network	Automatic Relationalization and Analytic Exploration of Documents	CA\$18,500	2010-2011	PI (100%)
Australian Research Council Discovery Projects	Next generation data mining techniques for analysing large evolving networks	AU\$697,000	2011-2013	Participating Investigator (10%)
MITACS Globalink Program	Centrality Analysis on Probabilistic Graphs and Social Networks	CA\$5,000	2010	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development (CRD) Grants	Extracting and Tagging Tabular Data on the Web	CA\$14,500	2009-2010	PI (100%)
SAP Business Objects Academic Research Center Project	Extracting and Tagging Tabular Data on the Web	CA\$12,500	2009-2010	PI (100%)
Australian Research Council Discovery Projects	Mining Multiple Information Sources through Collaborative and Comparative Analysis	AU\$335,000	2010-2012	Participating Investigator (10%)
Natural Sciences and Engineering Research Council of Canada (NSERC) 2008 Discovery Accelerator Supplements	Analyzing and Mining Large Uncertain Data Sets: Models, Methods and Applications	CA\$120,000	2008-2011	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program	Analyzing and Mining Large Uncertain Data Sets: Models, Methods and Applications	CA\$160,000	2008-2013	PI (100%)
Microsoft AdCenter	Context-aware Keyword Search and Bidding	US\$20,000	2008-2009	PI (100%)

Source	Title	Amount	Duration	My role
NSERC/MITACS Joint Internship Project (Partner Organization: Business Objects)	Advanced Data Analytics Techniques for On-line Data-Sharing Communities	CA\$30,000	2008-2009	PI (100%)
Michael Smith Foundation for Health Research Team Planning Award	Genomics Data Mining for Personalized Medicine	CA\$50,000	2007-2008	Co-PI (20%)
Simon Fraser University Community Trust Endowment Fund	Bioinformatics for Combating Infectious Diseases: Novel Methods for Drug and Target Identification and Evaluation	CA\$1,750,000	2007-2012	Investigator (5.71%)
2006 IBM Eclipse Innovation Award (IBM Faculty Award)	Collaborative Data Mining and Effective Knowledge Sharing	US\$22,000	2006-2007	PI (100%)
2006 HP Technology for Teaching Grant Award	Learning by Playing Games in the Classroom	US\$70,000	2006-2007	PI (100%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Research and Development Grants	Data Mining in Marketing and Services for CIBC	CA\$40,000	2006-2007	Co-PI (50%)
Canada Foundation for Innovation Foundation New Opportunities Fund Award / British Columbia Knowledge Development Fund	Scientific Data Acquisition, Transportation and Storage	CA\$240,000	2006	Co-PI (16.7%)
Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program	Online Analytic Processing and Interactive Mining Complex Data from Multiple Sources	CA\$81,000	2005-2008	PI (100%)
Simon Fraser University Endowed Research Fellowship Award	Data Mining and Data Warehousing	CA\$5,000	10/1/04-3/31/05	PI (100%)
Simon Fraser University President Research Grant	Data Mining and Data Warehousing	CA\$10,000	10/1/04-9/30/06	PI (100%)
NSF IIS-0308001	A Semantic Summarization Approach to Data Warehousing and Online Analytical Processing	US\$260,519	9/15/03-9/14/06	PI (100%)
UB Interdisciplinary Research and Creative Activities Fund (IRCAF)	Real-Time Data Mining Techniques in Large-Scale Hydro-Informatics Applications	US\$40,000	11/15/03-11/14/04	PI (50%)
NSF EIA-0306762	Epidemiological Analysis and Early Warning of Disease Outbreaks by Automated Reading and Mining of Pre-Hospital Care Reports	US\$50,000	8/1/03-7/31/04	Co-PI

Source	Title	Amount	Duration	My role
CUBRC match	Epidemiological Analysis and Early Warning of Disease Outbreaks by Automated Reading and Mining of Pre-Hospital Care Reports	US\$50,000	8/1/03-7/31/04	Co-PI
NIH Pre-NPEBC	Computational Approaches to Disease Causes and Treatment	US\$1,177,500	8/15/03-7/31/06	Investigator
United University Profession (UUP)	Professional Development Award	US\$510	9/1/02-6/30/03	PI (100%)

7 Publications

My students are marked by “[*student*]”, and my visitors are marked by “[*visitor*]”. To promote my students, I always ask my students and visitors to take the first authorship whenever they make substantial contribution in a work.

7.1 Textbook

1. J. Han, J. Pei, and H. Tong. *Data Mining: Concepts and Techniques*, fourth edition, Elsevier Inc., 2023, ISBN: 978-0-12-811760-6.
2. J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*, third edition, Morgan Kaufmann, 2011, ISBN: 978-0-1238-1479-1.

7.2 Monographs

1. M. Hua^[*student*] and J. Pei, *Ranking Queries on Uncertain Data*, Springer, 2011, ISBN: 978-1-4419-9379-3.
2. G. Dong and J. Pei, *Sequence Data Mining*, Springer, 2007, ISBN: 978-0-3876-9936-3.

7.3 Research Articles in Refereed Journals

1. Z. Cong^[*student*], B. Shi, S. Li, J. Yang, Q. He, and J. Pei. **FairSample: Training Fair and Accurate Graph Convolutional Neural Networks Efficiently**. To appear in *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society.
2. Q. Shen, L. Wu, Y. Zhang, Y. Pang, Z. Wei, F. U, B. Long, and J. Pei. **Multi-Interest Multi-Round Conversational Recommendation System with Fuzzy Feedback based User Simulator**. To appear in *ACM Transactions on Recommender Systems*, ACM Press.
3. H. Jiang, H. Yu, X. Cheng, J. Pei, R. Pless, and J. Yu. **DP2-Pub: Differentially Private High-Dimensional Data Publication with Invariant Post Randomization**. To appear in *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society.
4. H. Shi, M. Tayebi, J. Pei, and J. Cao. **Cost-Sensitive Learning for Medical Insurance Fraud Detection with Temporal Information**. To appear in *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society.

5. L. Xia, C. Huang, Y. Xu, and J. Pei. **Multi-Behavior Sequential Recommendation with Temporal Graph Transformer**. To appear in *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society.
6. Z. Zhang, C. Niu, P. Cui, J. Pei, B. Zhang, and W. Zhu. **Permutation-equivariant and Proximity-aware Graph Neural Networks with Stochastic Message Passing**. To appear in *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society.
7. C. Liu, Z. Zhou, J. Pei, Y. Zhang, and Y. Shi. **Decentralized Composite Optimization in Stochastic Networks: A Dual Averaging Approach with Linear Convergence**. To appear in *IEEE Transactions on Automatic Control*, IEEE Control Systems Society.
8. Z. Zhang, P. Cui, J. Pei, X. Wang, and W. Zhu. **Eigen-GNN: a Graph Structure Preserving Plug-in for GNNs**. *IEEE Transactions on Knowledge and Data Engineering*, Volume 35, Number 3, pages 2544-2555, March 2023, IEEE Computer Society.
9. R. Huang, J. Wang, S. Song, X. Lin, X. Zhu, and J. Pei. **Efficiently Cleaning Structured Event Logs: A Graph Repair Approach**. *ACM Transactions on Database Systems*, Volume 48, Issue 1, Article No. 3, pages 1-44, March 2023, ACM Press.
10. H. Jiang, J. Pei, D. Yu, J. Yu, B. Gong, and X. Cheng. **Applications of Differential Privacy in Social Network Analysis: A Survey**. *IEEE Transactions on Knowledge and Data Engineering*, Volume 35, Number 1, pages 108-127, January 2023, IEEE Computer Society.
11. L. Wu, Y. Chen, K. Shen, X. Guo, H. Gao, S. Li, J. Pei, and B. Long. **Graph Neural Networks for Natural Language Processing: A Survey**. *Foundations and Trends in Machine Learning*, Volume 16, No. 2, pages 119-328, January 2023, NOW Publishers.
12. H. Shi, Y. Yang, L. Wang, D. Ma, M.F. Beg, J. Pei, and J. Cao. **Two-Dimensional Functional Principal Component Analysis for Image Feature Extraction**. *Journal of Computational and Graphical Statistics*, Volume 31, Issue 4, pages 1127-1140, 2022, Taylor & Francis Group.
13. J. Pei. **A Survey on Data Pricing: from Economics to Data Science**. *IEEE Transactions on Knowledge and Data Engineering*, Volume 34, Issue 10, pages 4586-4608, October 2022, IEEE Computer Society.
14. Z. Cong^[student], X. Luo^[student], J. Pei, F. Zhu, and Y. Zhang. **Data Pricing in Machine Learning Pipelines**. *Knowledge and Information Systems*, Volume 64, Issue 6, pages 1417-1455, June 2022, Springer-Verlag.
15. F. Huang, S. Gao, J. Pei, and H. Huang. **Accelerated Zeroth-Order and First-Order Momentum Methods from Mini to Minimax Optimization**. *Journal of Machine Learning Research*, Volume 23, Number 36, pages 1-70, January 2022, JMLR, Inc.
16. X. Hu^[student], L. Chu, J. Pei, W. Liu, and J. Bian. **Model Complexity of Deep Learning: a Survey**. *Knowledge and Information Systems*, Volume 63, Issue 10, pages 2585-2619, August 2021, Springer-Verlag.
17. J. Liu, L. Xiong, J. Pei, J. Luo, H. Zhang, and W. Yu. **Group-based Skyline for Pareto Optimal Groups**. To appear in *IEEE Transactions on Knowledge and Data Engineering*, Volume 33, Issue 7, pages 2914-2929, July 2021, IEEE Computer Society.
18. H. Shi, D. Ma, Y. Nie, M. Beg, J. Pei, and J. Cao. **Early Diagnosis of Alzheimer's Disease on ADNI Data Using Novel Longitudinal Score Based on Functional Principal Component Analysis**. *Journal of Medical Imaging*, Volume 8, Issue 2, April 2021, SPIE.
19. Y. Yang^[student] and J. Pei. **Influence Analysis in Evolving Networks: A Survey**. *IEEE Transactions on Knowledge and Data Engineering*, Volume 33, Issue 3, pages 1045-1063, March 2021, IEEE Computer Society.
20. W. Yu, X. He, J. Pei, X. Chen, L. Xiong, J. Liu, and Z. Qin. **Visually Aware Recommendation with Aesthetic Features**. *The VLDB Journal*, Volume 30, Issue 4, pages 495-513, July 2021, Springer Berlin / Heidelberg.

21. J. Liu, J. Yang, L. Xiong, J. Pei, J. Luo, Y. Guo, S. Ma, and C. Fan. **Skyline Diagram: Efficient Space Partitioning for Skyline Queries.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 33, Issue 1, pages 271-286, January 2021, IEEE Computer Society.
22. D-W. Choi^[student], J. Pei, and X. Lin. **On Spatial Keyword Covering.** *Knowledge and Information Systems: An International Journal*, Volume 62, Issue 7, pages 2577-2612, July 2020, Springer-Verlag.
23. W. Yu, J. Liu, J. Pei, L. Xiong, X. Chen, and Z. Qin. **Efficient Contour Computation of Group-based Skyline.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 32, Issue 7, pages 1317-1332, July 2020, IEEE Computer Society.
24. Y. Yang^[student], X. Mao^[student], J. Pei, and X. He. **Continuous Influence Maximization.** *ACM Transactions on Knowledge Discovery from Data*, Volume 14, No. 3, Article 29, 38 pages, March 2020, ACM Press.
25. M. Lei^[visitor], L. Chu^[student], Z. Wang^[visitor], J. Pei, C. He, and X. Zhang. **Mining Top-k Sequential Patterns in Transaction Database Graphs: A New Challenging Problem and a Sampling-based Approach.** *World Wide Web Journal*, Volume 23, Issue 1, pages 103-130, January 2020, Springer-Verlag.
26. Z. Zhao^[student], L. Chu^[student], D. Tao, and J. Pei. **Classification with Label Noise: A Markov Chain Sampling Framework.** *Data Mining and Knowledge Discovery*, Volume 33, Issue 5, pages 1468-1504, September 2019, Springer-Verlag.
27. J. Liu, J. Yang, L. Xiong, and J. Pei. **Secure and Efficient Skyline Queries on Encrypted Data.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 31, Issue 7, pages 1397-1411, July 2019, IEEE Computer Society.
28. W. Yu, X. Lin, W. Zhang, J. Pei, and J. McCann. **SimRank*: Effective and Scalable Pairwise Similarity Search Based on Graph Topology.** *the VLDB Journal*, Volume 28, Issue 3, pages 401-426, June 2019, Springer Berlin / Heidelberg.
29. P. Cui, X. Wang, J. Pei, and W. Zhu. **A Survey on Network Embedding.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 31, Issue 5, pages 833-852, May 2019, IEEE Computer Society.
30. D. Zhu, P. Cui, Z. Zhang, J. Pei, and W. Zhu. **High-order Proximity Preserved Embedding For Dynamic Networks.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 30, Issue 11, pages 2134-2144, November 2018, IEEE Computer Society.
31. J. Hu^[student] and J. Pei. **Subspace Multi-clustering: A Review.** *Knowledge and Information Systems: An International Journal*, Volume 56, Issue 2, pages 257-284, August 2018, Springer-Verlag.
32. Y. Yang^[student], Z. Wang^[visitor], J. Pei, and E. Chen. **Tracking Influential Individuals in Dynamic Networks.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 29, Issue 11, pages 2615-2628, November 2017, IEEE Computer Society.
33. Z. Wang^[visitor], Y. Yang^[student], J. Pei, L. Chu^[student], and E. Chen. **Activity Maximization by Effective Information Diffusion in Social Networks.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 29, Issue 11, pages 2374-2387, November 2017, IEEE Computer Society.
34. Y. Yang^[student], J. Pei, and A. Al-Barakati. **Measuring In-Network Node Similarity Based on Neighborhoods: A Unified Parametric Approach.** *Knowledge and Information Systems: An International Journal*, Volume 53, Issue 1, pages 43-70, October 2017, Springer-Verlag.
35. J. Hu^[student], Q. Qian, J. Pei, R. Jin, and S. Zhu. **Finding Multiple Stable Clusterings.** *Knowledge and Information Systems: An International Journal*, Volume 51, Issue 3, pages 991-1021, June 2017, Springer-Verlag.
36. A. Campbell^[student], X. Mao^[student], J. Pei, and A. Al-Barakati. **Multidimensional Business Benchmarking Analysis on Data Warehouses.** *Intelligent Data Analysis – An International Journal*, Volume 13, Issue 1, pages 51-75, January 2017, IOS Press.
37. K. Yu^[student], X. Wu, W. Ding, and J. Pei. **Scalable and Accurate Online Feature Selection for Big Data.** *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Volume 11 Issue 2, Article No. 6, December 2016, ACM Press.

38. S. Liu, J. Yin, X. Wang, W. Cui, K. Cao, and J. Pei. **Online Visual Analytics of Text Streams.** *IEEE Transactions on Visualization and Computer Graphics*, Volume 22, Issue 11, pages 2451-2466, November 2016, IEEE Computer Society.
39. N. X. Vinh, J. Chan, S. Romano, J. Bailey, C. Leckie, K. Ramamohanarao, and J. Pei. **Discovering Outlying Aspects in Large Datasets.** *Data Mining and Knowledge Discovery*, Volume 30, Issue 6, pages 1520-1555, November 2016, Springer-Verlag.
40. X. Xu^[student], C. Gao^[student], J. Pei, K. Wang, and A. Al-Barakati. **Continuous Similarity Search for Evolving Queries.** *Knowledge and Information Systems: An International Journal*, Volume 48, Issue 3, pages 649-678, September 2016, Springer-Verlag.
41. L. Duan^[visitor], G. Tang^[student], J. Pei, J. Bailey, G. Dong, V. Nguyen, A. Campbell, and C. Tang. **Efficient Discovery of Contrast Subspaces for Object Explanation and Characterization.** *Knowledge and Information Systems: An International Journal*, Volume 47, Issue 1, pages 99-129, April 2016, Springer-Verlag.
42. L. Duan^[visitor], G. Tang^[student], J. Pei, J. Bailey, A. Campbell^[student], and C. Tang. **Mining Outlying Aspects on Numeric Data.** *Data Mining and Knowledge Discovery*, Volume 29, Issue 5, pages 1116-1151, September 2015, Springer-Verlag.
43. G. Tang^[student], J. Pei, J. Bailey, and G. Dong. **Mining Multidimensional Contextual Outliers from Categorical Relational Data.** *Intelligent Data Analysis – An International Journal*, Volume 19, No. 5, pages 1171-1192, September 2015, IOS Press.
44. X. Zhang, W. Dou, J. Pei, S. Nepal, C. Yang, C. Liu, and J. Chen. **Proximity-aware Local-recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud.** *IEEE Transactions on Computers*, Volume 64, Issue 8, pages 2293-2307, August 2015, IEEE Computer Society.
45. K. Yu^[student], W. Ding, D.A. Simovici, H. Wang, J. Pei, and X. Wu. **Classification with Streaming Features: An Emerging Pattern Mining Approach.** *ACM Transactions on Knowledge Discovery from Data*, Volume 9, Issue 4, Article No. 30, June 2015, ACM Press.
46. Y. Yang, Q. S. Lu, G. Tang^[student], and J. Pei. **The Impact of Market Competition on Search Advertising.** *Journal of Interactive Marketing*, Volume 30, pages 46-55, May 2015, Elsevier.
47. C. Yang, X. Zhang, C. Zhong, C. Liu, J. Pei, K. Ramamohanarao, and J. Chen. **A spatiotemporal compression based approach for efficient big data processing on Cloud.** *Journal of Computer and System Sciences*, Volume 80, Issue 8, pages 1563-1583, December 2014, Elsevier.
48. D. Huang^[student], K. Xu, and J. Pei. **Malicious URL Detection by Dynamically Mining Patterns without Pre-defined Elements.** *World Wide Web Journal*, Volume 17, Issue 6, pages 1375-1394, November 2014, Springer-Verlag.
49. G. Tang^[student], J. Pei, and W-S. Luk. **Email Mining: Tasks, Common Techniques, and Tools.** *Knowledge and Information Systems: An International Journal*, Volume 41, Issue 1, pages 1-31, October 2014, Springer-Verlag.
50. Y. Yang, J. X. Yu, H. Gao, J. Pei, and J. Li. **Mining Most Frequently Changing Component in Evolving Graphs.** *World Wide Web Journal*, Volume 17, Issue 3, pages 351-376, May 2014, Springer-Verlag.
51. Y. Zhang, W. Zhang, J. Pei, X. Lin, Q. Lin, and A. Li. **Consensus-based Ranking of Multi-valued Objects: A Generalized Borda Count Approach.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 26, Issue 1, pages 83-96, January 2014, IEEE Computer Society.
52. Y-C. Lo, J-Y. Li, M-Y. Yeh, S-D. Lin, and J. Pei. **What Distinguish One from Its Peers in Social Networks?.** *Data Mining and Knowledge Discovery*, Volume 27, Issue 3, pages 396-420, November 2013, Springer-Verlag.
53. Z. Liao, D. Jiang, J. Pei, Y. Huang, E. Chen, H. Cao, and H. Li. **A vIHMM Approach to Context-Aware Search.** *ACM Transactions on the Web*, Volume 7, Issue 4, Article No. 2, 38 pages, October 2013, ACM Press.

54. D. Jiang, J. Pei, and H. Li. **Mining Search and Browse Logs for Web Search: A Survey.** *ACM Transactions on Intelligent Systems and Technology*, Volume 4, Issue 4, Article Number 57, 37 pages, September 2013, ACM Press.
55. B. Jiang^[student], J. Pei, Y. Tao, and X. Lin. **Clustering Uncertain Data Based on Probability Distribution Similarity.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 25, No. 4, pages 751-763, April 2013, IEEE Computer Society.
56. Y. Cui^[student], J. Pei, G. Tang^[student], W-S. Luk, D. Jiang, and M. Hua. **Finding Email Correspondents in Online Social Networks.** *World Wide Web Journal*, Volume 16, Issue 2, pages 195-218, March 2013, Springer-Verlag.
57. J. Chen^[visitor], J. Huang^[visitor], B. Jiang^[student], J. Pei, and J. Yin. **Recommendations for Two-Way Selections Using Skyline View Queries.** *Knowledge and Information Systems: An International Journal*, Volume 34, Issue 2, pages 397-424, February 2013, Springer-Verlag.
58. J. Huang^[visitor], B. Jiang^[student], J. Pei, J. Chen^[visitor], and Y. Tang. **Skyline Distance: A Measure of Multidimensional Competence.** *Knowledge and Information Systems: An International Journal*, Volume 34, Issue 2, pages 373-396, February 2013, Springer-Verlag.
59. Y. Liu, Y. Zhao, L. Chen, J. Pei, and J. Han. **Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays.** *IEEE Transactions on Parallel and Distributed Systems*, Volume 23, Issue 11, pages 2138-2149, November 2012, IEEE Computer Society.
60. Q. Jiang^[student], A. Campbell, G. Tang^[student], and J. Pei. **Multi-level Relationship Outlier Detection.** *the International Journal of Business Intelligence and Data Mining (IJBIDM)*, Volume 7, Number 4, pages 253-273, October 2012, InderScience Publishers.
61. L. Li^[student], S. Petschulat, G. Tang^[student], J. Pei, W-S. Luk. **Efficient and Effective Aggregate Keyword Search on Relational Databases.** *the International Journal of Data Warehousing and Mining (IJDWM)*, Volume 7, Issue 4, pages 41-81, October 2012, InderScience Publishers.
62. M. Hua^[student] and J. Pei. **Clustering in Applications with Multiple Data Sources — A Mutual Subspace Clustering Approach.** *Neurocomputing*, Volume 92, pages 133-144, September 2012, Elsevier.
63. Z. Xing^[student], J. Pei, and P.S. Yu. **Early Classification on Time Series.** *Knowledge and Information Systems: An International Journal*, Volume 31, Issue 1, pages 105-127, April 2012, Springer-Verlag.
64. B. Zhou^[student] and J. Pei. **Aggregate Keyword Search on Large Relational Databases.** *Knowledge and Information Systems: An International Journal*, Volume 30, Number 2, pages 283-318, February 2012, Springer-Verlag.
65. B. Jiang^[student], J. Pei, X. Lin, and Y. Yuan. **Probabilistic Skylines on Uncertain Data: Model and Bounding-Pruning-Refining Methods.** *Journal of Intelligent Information Systems*, Volume 38, Number 1, pages 1-39, February 2012, Springer-Verlag.
66. X. Sun, H. Wang, J. Li, and J. Pei. **Publishing Anonymous Survey Rating Data.** *Data Mining and Knowledge Discovery*, Volume 23, pages 379-406, November 2011, Springer-Verlag.
67. Z. Liao, D. Jiang^[student], E. Chen, J. Pei, H. Cao, and H. Li. **Mining Concept Sequences from Large-Scale Search Logs for Context-Aware Query Suggestion.** *ACM Transactions on Intelligent Systems and Technology*, Volume 3, Issue 1, pages 17:1-40, October 2011, ACM Press.
68. R. C.W. Wong, A. W.C. Fu, K. Wang, P. S. Yu, and J. Pei. **Can the Utility of Anonymized Data Be Used for Privacy Breaches?.** *ACM Transactions on Knowledge Discovery in Data*, Volume 5, Issue 3, pages 16:1-24, August 2011, ACM Press.
69. B. Zhou^[student] and J. Pei. **K-Anonymity and L-Diversity Approaches for Privacy Preservation in Social Networks against Neighborhood Attacks.** *Knowledge and Information Systems: An International Journal*, Volume 28, Number 1, pages 47-77, July 2011, Springer-Verlag.
70. Y. Zhang, W. Zhang, X. Lin, B. Jiang^[student], and J. Pei. **Ranking Uncertain Sky: the Probabilistic Top-k Skyline Operator.** *Information System Journal*, Volume 36, Issue 5, pages 898-915, July 2011, Elsevier Ltd.

71. M. Hua^[student], J. Pei, and X. Lin. **Ranking Queries on Uncertain Data.** *The VLDB Journal*, Volume 20, Number 1, pages 129-153, February 2011, Springer Berlin / Heidelberg.
72. Z. Xing^[student], J. Pei, and E. Keogh. **A Brief Survey on Sequence Classification.** *ACM SIGKDD Explorations*, Volume 12, Issue 1, pages 40-48, June 2010, ACM Press.
73. Z. Lin^[student], B. Jiang^[student], and J. Pei. **Mining Discriminative Items in Multiple Data Streams.** *World Wide Web Journal*, Volume 13, Issue 4, pages 497-522, December 2010, Springer-Verlag.
74. Z. Xing^[student] and J. Pei. **Exploring Disease Association from the NHANES Data: Data Mining, Pattern Summarization, and Visual Analytics.** *the International Journal of Data Warehousing and Mining (IJDDM)*, Volume 6, Issue 3, pages 11-27, July-September 2010, InderScience Publishers.
75. E. Loekito, J. Bailey, and J. Pei. **Binary Decision Diagram Based Approach for Mining Frequent Subsequences.** *Knowledge and Information Systems: An International Journal*, Volume 24, Number 2, pages 235-268, August 2010, Springer-Verlag.
76. S. Yuen, Y. Tao, X. Xiao, J. Pei, D. Zhang. **Superseding Nearest Neighbor Search on Uncertain Spatial Databases.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 22, Number 7, pages 1041-1055, July 2010, IEEE Computer Society.
77. X. Cheng, J. Xu^[student], J. Pei, and J. Liu. **Hierarchical Distributed Data Classification in Wireless Sensor Networks.** *Computer Communication*, Volume 33, Issue 12, pages 1404-1413, July 15, 2010, Elsevier.
78. W. Zhang, X. Lin, Y. Zhang, J. Pei, and W. Wang. **Threshold-based Probabilistic Top-k Dominating Queries,** *The VLDB Journal*, Volume 19, Number 2, pages 283-305, April 2010, Springer Berlin / Heidelberg.
79. M. A. Cheema, X. Lin, W. Wang, W. Zhang, and J. Pei. **Probabilistic Reverse Nearest Neighbor Queries on Uncertain Data.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 22, Number 4, pages 550-564, April 2010, IEEE Computer Society.
80. B. Aljaber, N. Stokes, J. Bailey, and J. Pei. **Document Clustering of Scientific Texts Using Citations Contexts.** *Information Retrieval*, Volume 13, Number 2, pages 101-131, April 2010, Springer-Verlag.
81. M. Hua^[student], M. K. Lau^[student], J. Pei, and K. Wu. **Continuous K-Means Monitoring with Low Reporting Cost in Sensor Networks.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 21, Number 12, pages 1679-1691, December 2009, IEEE Computer Society.
82. X. Zeng^[student], J. Pei, K. Wang, and J. Li. **PADS: A Simple Yet Effective Pattern-Aware Dynamic Search Method for Fast Maximal Frequent Pattern.** *Knowledge and Information Systems: An International Journal*, Volume 20, Number 3, pages 375-391, September 2009, Springer-Verlag.
83. M. Hua^[student] and J. Pei. **Continuously Monitoring Top-K Uncertain Data Streams: A Probabilistic Threshold Method.** *Distributed and Parallel Databases: An International Journal*, Volume 26, Number 1, (special issue on ranking in databases), pages 29-65, August, 2009, Springer-Verlag.
84. B. Zhou^[student] and J. Pei. **Link Spam Target Detection Using Page Farms.** *ACM Transactions on Knowledge Discovery in Data*, Volume 3, Number 3, pages 13:1-38, July 2009, ACM Press.
85. R. She, J. S.-C. Chu, K. Wang, J. Pei, and N. Chen. **genBlastA: Enabling BLAST to identify homologous gene sequences.** *Genome Research*, Number 19, pages 143-149, 2009, Cold Spring Harbor Laboratory Press.
86. M. P. Ng, I. A. Vergara, C. Frech, Q. Chen, X. Zeng^[student], J. Pei, and N. Chen. **OrthoClusterDB: a Web Server for Synteny Blocks.** *BMC Bioinformatics*, Volume 10, Article 192, 2009.
87. M. Hua^[student], J. Pei, A. W.-C. Fu, X. Lin, and H-F Leung. **Top-k Typicality Queries and Efficient Query Answering Methods on Large Databases.** *The VLDB Journal*, Volume 18, Number 3, pages 809-835, June 2009, Springer Berlin / Heidelberg.
88. R. C.-W. Wong^[visitor], A. W.-C. Fu, K. Wang, and J. Pei. **Anonymization based Attack in Privacy Preserving Data Publishing.** *ACM Transactions on Database Systems*, Volume 34, Issue 2, pages 8:1-46, June 2009, ACM Press.

89. D. Jiang^[student] and J. Pei. **Mining Frequent Cross-Graph Quasi-Cliques.** *ACM Transactions on Knowledge Discovery in Data*, Volume 2, Number 4, pages 16:1-42, January 2009, ACM Press.
90. R. C.-W. Wong^[visitor], J. Pei, A. W.-C. Fu, and K. Wang. **Online Skyline Analysis with Dynamic Preferences on Nominal Attributes.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 21, Number 1, pages 35-49, January 2009, IEEE Computer Society.
91. B. Zhou^[student], J. Pei, and W.-S. Luk. **A Brief Survey on Anonymization Techniques for Privacy Preserving Publishing of Social Network Data.** *ACM SIGKDD Explorations*, Volume 10, Issue 2, pages 12-22, December 2008, ACM Press.
92. J. Li, R. C.-W. Wong, A. W.-C. Fu, and J. Pei. **Anonymisation by Local Recoding in Data with Attribute Hierarchical Taxonomies.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 20, Number 9, pages 1181-1194, September 2008, IEEE Computer Society.
93. H. Wang and J. Pei. **Clustering by Pattern Similarity.** *Journal of Computer Science and Technology*, Volume 23, Number 4, pages 481-496, July, 2008, Springer.
94. D. Jiang^[student], J. Pei, M. Ramanathan, C. Lin, C. Tang, and A. Zhang. **Mining Gene-Sample-Time Microarray Data: A Coherent Gene Cluster Discovery Approach.** *Knowledge and Information Systems: An International Journal*, Volume 13, Number 3, pages 305-335, November 2007, Springer-Verlag.
95. Y. Tao, X. Xiao, and J. Pei. **Efficient Skyline and Top-k Retrieval in Subspaces.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 19, Number 8, pages 1072-1088, August 2007, IEEE Computer Society.
96. M. Cho^[student], J. Pei, and K. Wang. **Answering Ad Hoc Aggregate Queries from Data Streams Using Prefix Aggregate Trees.** *Knowledge and Information Systems: An International Journal*, Volume 12, Number 3, pages 301-329, August 2007, Springer-Verlag.
97. C. Liu, K. Wu, and J. Pei. **An Energy Efficient Data Collection Framework for Wireless Sensor Networks by Exploiting Spatiotemporal Correlation.** *IEEE Transactions on Parallel and Distributed Systems*, Volume 18, Number 7, pages 1010-1023, July 2007, IEEE Computer Society.
98. J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, and D. Yang. **H-Mine: Fast and space-preserving frequent pattern mining in large databases.** *IIE Transactions*, Volume 39, Issue 6, pages 593-605, June, 2007, Taylor & Francis.
99. J. Pei, J. Han, and W. Wang. **Constraint-Based Sequential Pattern Mining: The Pattern-Growth Methods.** *Journal of Intelligent Information Systems*, Volume 28, Number 2, pages 133-160, April, 2007, Springer-Verlag.
100. J. Pei, Y. Yuan, X. Lin, W. Jin, M. Ester, Q. Liu, W. Wang, Y. Tao, J.X. Yu, and Q. Zhang. **Towards Multidimensional Subspace Skyline Analysis.** *ACM Transactions on Database Systems*, Volume 31, Number 4, pages 1335-1381, December 2006, ACM Press.
101. Y. Chen, G. Dong, J. Han, J. Pei, B. W. Wah, and J. Wang. **Regression Cubes with Lossless Compression and Aggregation.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 18, Number 12, pages 1585-1599, December 2006, IEEE Computer Society.
102. J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. W. Fu. **Utility-Based Anonymization for Privacy Preservation with Less Information Loss.** *ACM SIGKDD Explorations*, Volume 8, Issue 2, pages 21-30, December 2006.
103. J. Pei, H. Wang, J. Liu^[student], K. Wang, J. Wang, and P. S. Yu. **Discovering Frequent Closed Partial Orders from Strings.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 18, Number 11, pages 1467-1481, November 2006, IEEE Computer Society.
104. I. Pekerskaya^[student], J. Pei, and K. Wang. **Mining Changing Regions from Access-Constrained Snapshots: A Cluster-Embedded Decision Tree Approach.** *Journal of Intelligent Information Systems (Special Issue on Mining Spatio-Temporal Data)*, Volume 27, Number 3, pages 215-242, November 2006, Springer-Verlag.
105. Y. Huang, J. Pei, and H. Xiong. **Co-location Mining with Rare Spatial Features.** *GeoInformatica*, Volume 10, Number 3, pages 239-260, September 2006, Springer Netherlands.

106. J. Wang, J. Han, and J. Pei. **Closed Constrained-Gradient Mining in Retail Databases.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 18, Number 6, pages 764-769, June 2006, IEEE Computer Society.
107. J. Han, Y. Chen, G. Dong, J. Pei, B. W. Wah, J. Wang, and Y. D. Cai. **Stream Cube: An Architecture for Multi-Dimensional Analysis of Data Streams.** *Distributed and Parallel Databases*, Volume 18, Number 2, pages 173-197, September 2005, Springer Science + Business Media.
108. D. Jiang^[student], J. Pei, and A. Zhang. **An Interactive Approach to Mining Gene Expression Data.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 17, Number 10, pages 1363-1378, October 2005, IEEE Computer Society.
109. M. Cho^[student], J. Pei, H. Wang, and W. Wang. **Preference-based Frequent Pattern Mining.** *International Journal of Data Warehousing and Mining*, Volume 1, Number 4, pages 56-77, October-December, 2005, InderScience Publishers.
110. J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.C. Hsu. **Mining Sequential Patterns by Pattern-growth: The PrefixSpan Approach.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 16, Number 11, pages 1424-1440, November 2004, IEEE Computer Society.
111. J. Pei, G. Dong, W. Zou, and J. Han. **Mining Condensed Frequent Pattern Bases.** *Knowledge and Information Systems: An International Journal*, Volume 6, Number 5, pages 570-594, September 2004, Springer-Verlag.
112. G. Dong, J. Han, J. Lam, J. Pei, K. Wang, and W. Zou. **Mining Constrained Gradients in Large Databases.** *IEEE Transactions on Knowledge and Data Engineering*, Volume 16, Number 8, pages 922-938, August 2004, IEEE Computer Society.
113. J. Pei, J. Han, and L. V. S. Lakshmanan. **Pushing Convertible Constraints in Frequent Itemset Mining.** *Data Mining and Knowledge Discovery: An International Journal*, Volume 8, Issue 3, pages 227-252, May 2004, Kluwer Academic Publishers.
114. J. Han, J. Pei, and X. Yan. **From Sequential Pattern Mining to Structured Pattern Mining: A Pattern-Growth Approach.** (Invited paper) *Journal of Computer Science and Technology*, Volume 19, No. 3, pages 257-279, May 2004, Allerton Press, Inc.
115. J. Han, J. Pei, Y. Yin, and R. Mao. **Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach.** *Data Mining and Knowledge Discovery: An International Journal*, Volume 8, Issue 1, pages 53-87, January 2004, Kluwer Academic Publishers.
116. D. Jiang^[student], J. Pei, and A. Zhang. **Towards Interactive Exploration of Gene Expression Patterns.** *ACM SIGKDD Explorations (Special Issue on Microarray Data Analysis)*, Volume 5, Issue 2, pages 79-90, December 2003.
117. Z. Chen, C. Li, J. Pei, Y. Tao, H. Wang, W. Wang, J. Yang, J. Yang, and D. Zhang. **Recent Progress on Selected Topics in Database Research: A Report from Nine Young Chinese Researchers Working in the United States.** (Invited paper) *Journal of Computer Science and Technology*, Volume 18, No. 5, September, 2003, pages 538-552, Allerton Press, Inc.
118. J. Pei and J. Han. **Constrained Frequent Pattern Mining: A Pattern-Growth View.** *ACM SIGKDD Explorations (Special Issue on Constraints in Data Mining)*, Volume 4, Issue 1, pages 31-39, June 2002.
119. J. Han and J. Pei. **Mining Frequent Patterns by Pattern-Growth: Methodology and Implications.** *ACM SIGKDD Explorations (Special Issue on Scalable Data Mining Algorithms)*, Volume 2, Issue 2, pages 14-20, December, 2000.

7.4 Research Articles in Refereed Conference Proceedings

1. X. Luo^[student], J. Pei, C. Xu, W. Zhang, and J. Xu. **Fast Shapley Value Computation in Data Assemblage Tasks as Cooperative Simple Games.** In *Proceedings of the 2024 ACM SIGMOD International Conference on Management of Data (SIGMOD'24)*, Santiago, Chile, June 5-15, 2024.

2. N. Wu, M. Gong, L. Shou, J. Pei, and D. Jiang. **RUEL: Retrieval-Augmented User Representation with Edge Browser Logs for Sequential Recommendation.** In *Proceedings of the Thirty-second ACM International Conference on Information and Knowledge Management (CIKM'23)*, University of Birmingham and Eastside Rooms, UK, October 21-25, 2023.
3. X. Wu, Z. Hu, J. Pei, and H. Huang. **Serverless Biased Stochastic Methods for Multi-Party Collaborative Imbalanced Data Mining.** In *Proceedings of the Twenty-ninth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'23)*, Long Beach, CA, USA, August 6-10, 2023.
4. R. Xue, H. Han, M. Torkamani, J. Pei, and X. Liu. **LazyGNN: Large-Scale Graph Neural Networks via Lazy Propagation.** In *Proceedings of the Fortieth International Conference on Machine Learning (ICML'23)*, Honolulu, HI, USA, July 23-29, 2023.
5. N. Chen, L. Shou, J. Pei, M. Gong, B. Cao, J. Chang, J. Li, and D. Jiang. **Alleviating Over-smoothing for Unsupervised Sentence Representation.** In *Proceedings of the Sixty-first Annual Meeting of the Association for Computational Linguistics (ACL'23)*, Toronto, ON, Canada, July 9-14, 2023.
6. N. Chen, L. Shou, T. Song, M. Gong, J. Pei, J. Chang, D. Jiang, and J. Li. **Structural Contrastive Pretraining for Cross-Lingual Comprehension.** In *Findings of the Sixty-first Annual Meeting of the Association for Computational Linguistics (ACL'23)*, Toronto, ON, Canada, July 9-14, 2023.
7. J. Zhang, Q. Sun, J. Liu, L. Xiong, J. Pei, and K. Ren. **Efficient Sampling Approaches to Shapley Value Approximation.** In *Proceedings of the 2023 ACM SIGMOD International Conference on Management of Data (SIGMOD'23)*, Seattle, WA, USA, June 18-23, 2023.
8. J. Peng, H. Zou, J. Liu, S. Li, Y. Jiang, J. Pei, and P. Cui. **Offline Policy Evaluation in Large Action Spaces via Outcome-Oriented Action Grouping.** In *Proceedings of the 2023 ACM Web Conference (WWW'23)*, Austin, TX, USA, April 30 - May 4, 2023.
9. H. Zou, H. Wang, R. Xu, B. Li, J. Pei, Y.J. Jian, and P. Cui. **Factual Observation Based Heterogeneity Learning for Counterfactual Prediction.** In *Proceedings of the Second Conference on Causal Learning and Reasoning (CLear'23)*, Tübingen, German, April 11-14, 2023.
10. J. Zhang, H. Xia, Q. Sun, J. Liu, L. Xiong, J. Pei, and K. Ren. **Dynamic Shapley Value Computation.** In *Proceedings of the Thirty-ninth IEEE International Conference on Data Engineering (ICDE'23)*, Anaheim, CA, USA, April 3-7, 2023.
11. L. Xia, Y. Shao, C. Huang, Y. Xu, H. Xu, and J. Pei. **Disentangled Graph Social Recommendation.** In *Proceedings of the Thirty-ninth IEEE International Conference on Data Engineering (ICDE'23)*, Anaheim, CA, USA, April 3-7, 2023.
12. Z. Xu, L. Shou, J. Pei, M. Gong, Q. Su, X. Quan, and J. Jiang. **A Graph Fusion Approach for Cross-Lingual Machine Reading Comprehension.** In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI'23)*, Washington, DC, USA, February 7-14, 2023.
13. S. Wu, R. Zhao, Y. Zheng, J. Pei, and B. Liu. **Identify Event Causality with Knowledge and Analogy.** In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI'23)*, Washington, DC, USA, February 7-14, 2023.
14. S. Liang, L. Shou, J. Pei, M. Gong, W. Zuo, X. Zuo, and D. Jiang. **Label-aware Multi-level Contrastive Learning for Cross-lingual Spoken Language Understanding.** In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP'22)*, Abu Dhabi, December 7-11, 2022.
15. H. Ren, L. Shou, J. Pei, N. Wu, M. Gong, and D. Jiang. **Lexicon-Enhanced Self-Supervised Training for Multilingual Dense Retrieval.** In *Findings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP'22)*, Abu Dhabi, December 7-11, 2022.
16. N. Liu, X. Wang, D. Bo, C. Shi, and J. Pei. **Revisiting Graph Contrastive Learning from the Perspective of Graph Spectrum.** In *Proceedings of the Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS'22)*, New Orleans, LA, November 28 - December 3, 2022.
17. Y. Zhang, R. Bao, J. Pei, and H. Huang. **Toward Unified Data and Algorithm Fairness via Adversarial Data Augmentation and Adaptive Model Fine-tuning.** In *Proceedings of the Twenty-second*

- IEEE International Conference on Data Mining (ICDM'22)*, Orlando, FL, USA, November 28 - December 1, 2022.
18. X. Luo^[student], J. Pei, Z. Cong^[student], and C. Xu^[student]. **On Shapley Value in Data Assemblage Under Independent Utility**. In *Proceedings of the Forty-eighth International Conference on Very Large Databases (VLDB'22)*, September 5-9, 2022, Sydney, Australia.
 19. J. Li, J. Pei, and H. Huang. **Communication-Efficient Robust Federated Learning with Noisy Labels**. In *the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, Washington DC, USA, August 14-18, 2022.
 20. Y. Zhang, S. Gao, J. Pei, and H. Huang. **Improving Social Network Embedding via New Second-Order Continuous Graph Neural Networks**. In *Proceedings of the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, Washington DC, USA, August 14-18, 2022.
 21. N. Chen, L. Shou, M. Gong, J. Pei, and D. Jiang. **Bridging the Gap between Language Models and Cross-Lingual Sequence Labeling**. In *Proceedings of the 2022 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'22)*, Seattle, Washington, WA, July 10-15, 2022.
 22. Z. Fan, H. Fang, Z. Zhou, J. Pei, M. Friedlander, C. Liu, and Y. Zhang. **Improving Fairness for Data Valuation in Horizontal Federated Learning**. In *Proceedings of the Thirty-eighth IEEE International Conference on Data Engineering (ICDE'22)*, Kuala Lumpur, Malaysia, May 9-12, 2022.
 23. H. Wang, C. Xu^[student], C. Zhang, J. Xu, P. Zhe, and J. Pei. **vChain+: Optimizing Verifiable Blockchain Boolean Range Queries**. In *Proceedings of the Thirty-eighth IEEE International Conference on Data Engineering (ICDE'22)*, Kuala Lumpur, Malaysia, May 9-12, 2022.
 24. Z. Li, C. Huang, L. Xia, Y. Xu, and J. Pei. **Spatial-Temporal Hypergraph Self-Supervised Learning for Urban Crime Prediction**. In *Proceedings of the Thirty-eighth IEEE International Conference on Data Engineering (ICDE'22)*, Kuala Lumpur, Malaysia, May 9-12, 2022.
 25. Y. Zhang, H. Gao, J. Pei, and H. Huang. **Robust Self-Supervised Structural Graph Neural Network for Social Network Prediction**. In *Proceedings of the Thirty-first Web Conference (WWW'22)*, Lyon, France, April 25-29, 2022.
 26. Y. Zhang, L. Wu, Q. Shen, Y. Pang, Z. Wei, F. Xu, B. Long, and J. Pei. **Multiple Choice Questions based Multi-Interest Policy Learning for Conversational Recommendation**. In *Proceedings of the Thirty-first Web Conference (WWW'22)*, Lyon, France, April 25-29, 2022.
 27. G. Singh, L. Chu, L. Wang, J. Pei, and Y. Zhang. **Mining Minority-class Examples with Uncertainty Estimates**. In *Proceedings of the Twenty-eighth International Conference on Multimedia Modeling (MMM'22)*, Qui Nhon, Vietnam, April 5-8, 2022.
 28. L. Charette, L. Chu, Y. Chen, L. Wang, J. Pei, and Y. Zhang. **Cosine Model Watermarking Against Ensemble Distillation**. In *Proceedings of the Thirty-sixth AAAI Conference on Artificial Intelligence (AAAI'22)*, Vancouver, BC, Canada, February 22 - March 1, 2022.
 29. N. Che, L. Shou, M. Gong, and J. Pei. **From Good to Best: Two-Stage Training for Cross-lingual Machine Reading Comprehension**. In *Proceedings of the Thirty-sixth AAAI Conference on Artificial Intelligence (AAAI'22)*, Vancouver, BC, Canada, February 22 - March 1, 2022.
 30. Y. Pang, L. Wu, Q. Shen, Y. Zhang, Z. Wei, F. Xu, E. Chang, B. Long, and J. Pei. **Heterogeneous Global Graph Neural Networks for Personalized Session-based Recommendation**. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM'22)*, Phoenix, AZ, USA, February 21-25, 2022.
 31. M. Bajaj, L. Chu, Z. Y. Xue, J. Pei, L. Wang, P. C.-H. Lam, and Y. Zhang. **Robust Counterfactual Explanations on Graph Neural Networks**. In *Proceedings of the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS'21)*, December 6-14, 2021.
 32. Y. Guo, L. Shou, J. Pei, M. Gong, M. Xu, Z. Wu, and D. Jiang. **Learning from Multiple Noisy Augmented Data Sets for Better Cross-Lingual Spoken Language Understanding**. In *Proceedings*

- of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP'21), Punta Cana, Dominican Republic, November 7-11, 2021.
33. C. H. Lam, L. Chu, M. Torgonskiy, J. Pei, Y. Zhang, and L. Wang. **Finding Representative Interpretations on Convolutional Neural Networks**. In *Proceedings of the 2021 IEEE International Conference on Computer Vision (ICCV'21)*, Virtual, October 11-17, 2021.
 34. Z. Cong^[student], L. Chu, Y. Yang, and J. Pei. **Comprehensible Counterfactual Explanation on Kolmogorov-Smirnov Test**. In *Proceedings of the Forty-seventh International Conference on Very Large Data Bases (VLDB'21)*, Copenhagen, Denmark, August 16-20, 2021.
 35. C. Xu^[student], C. Zhang, J. Xu, and J. Pei. **SlimChain: Scaling Blockchain Transactions through Off-Chain Storage and Parallel Processing**. In *Proceedings of the Forty-seventh International Conference on Very Large Data Bases (VLDB'21)*, Copenhagen, Denmark, August 16-20, 2021.
 36. J. Liu, J. Lou, J. Liu, L. Xiong, J. Pei, and J. Sun. **Dealer: An End-to-End Model Marketplace with Differential Privacy**. In *Proceedings of the Forty-seventh International Conference on Very Large Data Bases (VLDB'21)*, Copenhagen, Denmark, August 16-20, 2021.
 37. S. Liang, M. Gong, J. Pei, L. Shou, W. Zuo, X. Zuo, and D. Jiang. **Reinforced Iterative Knowledge Distillation for Cross-Lingual Named Entity Recognition**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
 38. A. Banitalebi-Dehkordi, N. Vedula, J. Pei, F. Xia, L. Wang, and Y. Zhang. **Auto-Split: A General Framework of Collaborative Edge-Cloud AI**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
 39. Q. Zhang, B. Gu, C. Deng, J. Pei, and H. Huang. **AsySQN: Faster Vertical Federated Learning Algorithms with Better Computation Resource Utilization**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
 40. H. Huang, X. Geng, J. Pei, G. Long, and D. Jiang. **Reasoning over Entity-Action-Location Graph for Procedural Text Understanding**. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)*, Bangkok, Thailand, August 1-6, 2021.
 41. Y. Zhou, X. Geng, T. Shen, J. Pei, W. Zhang, and D. Jiang. **Modeling Event-Pair Relations in External Knowledge Graphs for Script Reasoning**. In *Findings of ACL in the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)*, Bangkok, Thailand, August 1-6, 2021.
 42. P. Wang, J. Wang, J. Pei, and W. Zheng. **Automating Entity Matching Model Development**. In *Proceedings of the Thirty-seventh IEEE International Conference on Data Engineering (ICDE'21)*, Chania, Crete, Greece, April 19-22, 2021.
 43. J. Liu, L. Xiong, Q. Zhang, J. Pei, and J. Luo. **Eclipse: Generalizing kNN and Skyline**. In *Proceedings of the Thirty-seventh IEEE International Conference on Data Engineering (ICDE'21)*, Chania, Crete, Greece, April 19-22, 2021.
 44. S. Liang, L. Shou, J. Pei, M. Gong, W. Zuo, and D. Jiang. **CalibreNet: Calibration Networks for Multilingual Sequence Labeling**. In *Proceedings of the Fourteenth ACM International Conference on Web Search and Data Mining (WSDM'21)*, Jerusalem, Israel, March 8-12, 2021.
 45. F. Yuan, L. Shou, J. Pei, W. Lin, M. Gong, Y. Fu, and D. Jiang. **Reinforced Multi-Teacher Selection for Knowledge Distillation**. In *Proceedings of the Thirty-fifth AAAI Conference on Artificial Intelligence (AAAI'21)*, Online, February 2-9, 2021. (Acceptance rate: 1692/9034)
 46. Y. Huang, L. Chu, Z. Zhou, L. Wang, J. Liu, J. Pei, and Y. Zhang. **Personalized Cross-Silo Federated Learning on Non-IID Data**. In *Proceedings of the Thirty-fifth AAAI Conference on Artificial Intelligence (AAAI'21)*, Online, February 2-9, 2021. (Acceptance rate: 1692/9034)

47. L. Xia, C. Huang, Y. Xu, P. Dai, X. Zhang, H. Yang, J. Pei, and L. Bo. **Knowledge-Enhanced Hierarchical Graph Transformer Network for Multi-Behavior Recommendation**. In *Proceedings of the Thirty-fifth AAAI Conference on Artificial Intelligence (AAAI'21)*, Online, February 2-9, 2021. (Acceptance rate: 1692/9034)
48. J. Liu, L. Shou, J. Pei, M. Gong, M. Yang, and D. Jiang. **Cross-lingual Machine Reading Comprehension with Language Branch Knowledge Distillation**. In *Proceedings of the Twenty-eighth International Conference on Computational Linguistics (COLING'20)*, Online, December 8-13, 2020.
49. X. Zhang, L. Shou, J. Pei, M. Gong, L. Wen, and D. Jiang. **A Graph Representation of Semi-structured Data for Web Question Answering**. In *Proceedings of the Twenty-eighth International Conference on Computational Linguistics (COLING'20)*, Online, December 8-13, 2020.
50. L. Chu^[student], Y. Zhang^[student], Y. Yang^[student], L. Wang, and J. Pei. **Online Density Bursting Subgraph Detection from Temporal Graphs**. In *Proceedings of the Forty-sixth International Conference on Very Large Data Bases (VLDB'20)*, Tokyo, Japan, August 31 - September 4, 2020.
51. X. Hu^[student], W. Liu, J. Bian, and J. Pei. **Measuring Model Complexity of Neural Networks with Curve Activation Functions**. In *Proceedings of the Twenty-sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)*, San Diego, CA, USA, August 23-27, 2020. (Acceptance rate: 216/1279)
52. L. Shou, S. Bo, F. Cheng, M. Gong, J. Pei, and D. Jiang. **Mining Implicit Relevance Feedback from User Behavior for Web Question Answering**. In *Proceedings of the Twenty-sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)* (Applied Data Science Track), San Diego, CA, USA, August 23-27, 2020. (Acceptance rate: 121/756)
53. X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, and J. Pei. **AM-GCN: Adaptive Multi-channel Graph Convolutional Networks**. In *Proceedings of the Twenty-sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)*, San Diego, CA, USA, August 23-27, 2020. (Acceptance rate: 216/1279)
54. F. Huang, S. Gao, J. Pei, and H. Huang. **Momentum-Based Policy Gradient Methods**. In *Proceedings of the Thirty-seventh International Conference on Machine Learning (ICML'20)*, Vienna, Austria, July 13-18, 2020. (Acceptance rate: 1088/4990)
55. L. Luo, J. Pei, and H. Huang. **Sinkhorn Regression**. In *Proceedings of the Twenty-ninth International Joint Conference on Artificial Intelligence and the Seventeenth Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI'20)*, Yokohama, Japan, July 11-17, 2020. (Acceptance rate: 592/4717)
56. S. Gao, F. Huang, J. Pei, and H. Huang. **Discrete Model Compression with Resource Constraint**. In *Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'20)*, Seattle, WA, USA, June 14-18, 2020.
57. Z. Cong^[student], L. Chu^[student], L. Wang, X. Hu^[student], and J. Pei. **Exact and Consistent Interpretation of Piecewise Linear Models Hidden behind APIs: A Closed Form Solution**. In *Proceedings of the Thirty-sixth IEEE International Conference on Data Engineering (ICDE'20)*, Dallas, TX, USA, April 20-24, 2020.
58. Y. Yang^[student], Z. Wang, T. Jin, J. Pei, and E. Chen. **Tracking Top-k Influential Users with Relative Errors**. In *Proceedings of the Twenty-eighth ACM International Conference on Information and Knowledge Management (CIKM 2019)* (long paper with oral presentation), Beijing, China, November 3-7, 2019.
59. Z. Zhao^[student], L. Chu^[student], D. Tao, and J. Pei. **Classification with Label Noise: A Markov Chain Sampling Framework**. In *Proceedings of the 2019 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'19)*, Würzburg, Germany, September 16-20, 2019.
60. L. Chu^[student], Z. Wang^[visitor], J. Pei, Y. Zhang^[student], Y. Yang^[student], and E. Chen. **Finding Theme Communities from Database Networks**. In *Proceedings of the Forty-fifth International Conference on Very Large Data Bases (VLDB'19)*, Los Angeles, CA, USA, August 26-30, 2019.

61. M. Dolatshah, M. Teoh^[student], J. Wang, and J. Pei. **Cleaning Crowdsourced Labels Using Oracles for Statistical Classification**. In *Proceedings of the Forty-fifth International Conference on Very Large Data Bases (VLDB'19)*, Los Angeles, CA, USA, August 26-30, 2019.
62. H. Gao, J. Pei and H. Huang. **ProGAN: Network Embedding via Proximity Generative Adversarial Network** (oral presentation). In *Proceedings of the Twenty-fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*, Anchorage, AL, USA. August 4-8, 2019. (Acceptance rate: 110/1200)
63. H. Gao, J. Pei and H. Huang. **Conditional Random Field Enhanced Graph Convolutional Neural Networks** (oral presentation). In *Proceedings of the Twenty-fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*, Anchorage, AL, USA. August 4-8, 2019. (Acceptance rate: 110/1200)
64. K. Tu, J. Ma, P. Cui, J. Pei, and W. Zhu. **AutoNE: Hyperparameter Optimization for Massive Network Embedding**. In *Proceedings of the Twenty-fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*, Anchorage, AL, USA. August 4-8, 2019. (Acceptance rate: 110/1200)
65. C. Fan, Y. Zhang, Y. Pan, X. Li, C. Zhang, R. Yuan, D. Wu, W. Wang, J. Pei, and H. Huang. **Multi-Horizon Time Series Forecasting with Temporal Attention Learning**. In *Proceedings of the Twenty-fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*, Anchorage, AL, USA. August 4-8, 2019. (Acceptance rate: 110/1200)
66. S. Yu, B. Gu, K. Ning, H. Chen, J. Pei and H. Huang. **Tackle Balancing Constraint for Incremental Semi-Supervised Support Vector Learning**. In *Proceedings of the Twenty-fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*, Anchorage, AL, USA. August 4-8, 2019. (Acceptance rate: 170/1200)
67. H. Gao, J. Pei, and H. Huang. **Demystifying Dropout**. In *Proceedings of the Thirty-sixth International Conference on Machine Learning (ICML'19)*, Long Beach, CA, USA, June 9-15, 2019.
68. W. Yang, L. Tan, C. Lu, A. Cui, H. Li, X. Chen, K. Xiong, M. Wang, M. Li, J. Pei, and J. Lin. **Detecting Customer Complaint Escalation with Recurrent Neural Networks and Manually-Engineered Features**. In *Proceedings of the Seventeenth Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'19)*, Minneapolis, MN, USA, June 2-7, 2019.
69. D-W. Choi^[student], J. Pei, and T. Heinis. **Efficient Mining of Regional Movement Patterns in Semantic Trajectories**. In *Proceedings of the Forty-fourth International Conference on Very Large Data Bases (VLDB'18)*, Rio de Janeiro, Brazil, August 27-31, 2018.
70. L. Chu^[student], X. Hu^[student], J. Hu^[student], L. Wang, and J. Pei. **Exact and Consistent Interpretation for Piecewise Linear Neural Networks: A Closed Form Solution**. In *Proceedings of the Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*, London, United Kingdom, August 19-23, 2018. (Acceptance rate: 107/983)
71. Z. Zhang, P. Cui, X. Wang, J. Pei, X. Yao, and W. Zhu. **Arbitrary-Order Proximity Preserved Network Embedding**. In *Proceedings of the Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*, London, United Kingdom, August 19-23, 2018. (Acceptance rate: 107/983)
72. L. Luo, W. Zhang, Z. Zhang, W. Zhu, T. Zhang, and J. Pei. **Sketched Follow-The-Regularized-Leader for Online Factorization Machine**. In *Proceedings of the Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*, London, United Kingdom, August 19-23, 2018. (Acceptance rate: 181/983)
73. J. Peng, D. Zhang^[student], J. Wang, and J. Pei. **AQP++: Connecting Approximate Query Processing With Aggregate Precomputation for Interactive Analytics**. In *Proceedings of the 2018 ACM SIGMOD International Conference on Management of Data (SIGMOD'18)*, Huston, TX, USA, June 10-15, 2018.

74. X. Lin, W. Zhang, M. Zhang, P. Zhao, W. Zhu, J. Pei and J. Huang. **Online Compact Convexified Factorization Machine**. In *Proceedings of the Twenty-seventh International World Wide Web Conference (WWW'18)*, Lyon, France, April 23-27, 2018.
75. Y. Yang^[student], L. Chu^[student], Y. Zhang^[student], Z. Wang^[visitor], J. Pei, and E. Chen. **Mining Density Contrast Subgraphs**. In *Proceedings of the Thirty-fourth IEEE International Conference on Data Engineering (ICDE'18)*, Paris, France, April 16-19, 2018.
76. J. Liu, J. Yang, L. Xiong, J. Pei, and J. Luo. **Skyline Diagram: Finding the Voronoi Counterpart for Skyline Queries**. In *Proceedings of the Thirty-fourth IEEE International Conference on Data Engineering (ICDE'18)*, Paris, France, April 16-19, 2018.
77. Z. Zhang, P. Cui, J. Pei, X. Wang, and W. Zhu. **TIMERS: Error-Bounded SVD Restart on Dynamic Networks**. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, New Orleans, LA, USA, February 2-7, 2018. (Acceptance rate: 933/3800+)
78. C. Gao^[student], J. Wang, J. Pei, R. Li, and Y. Chang. **Preference-driven Similarity Join**. In *Proceedings of the 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI'17)* (long paper and Best Student Paper Award), Leipzig, Germany, August 23-26, 2017.
79. C. Gao^[student], J. Pei, J. Wang, and Y. Chang. **Schemaless Join for Result Set Preferences**. In *Proceedings of the Eighteenth IEEE International Conference on Information Reuse and Integration (IRI 2017)* (full regular paper for oral presentation), San Diego, CA, USA, August 4-6, 2017.
80. C-Y. Kuo, M-Y. Yeh, and J. Pei. **Principal Pattern Mining on Graphs**. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'17)*, Sydney, Australia, July 31-August 3, 2017.
81. J. Liu, J. Yang, L. Xiong, and J. Pei. **Secure Skyline Queries on Cloud Platform**. In *Proceedings of the Thirty-third IEEE International Conference on Data Engineering (ICDE'17)*, San Diego, CA, USA, April 19-22, 2017.
82. X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, and S. Yang. **Community Preserving Network Embedding**. In *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence (AAAI'17)* (oral presentation), San Francisco, CA, USA, February 4-9, 2017. (Acceptance rate: 638/2590)
83. Z. Zheng, D. Wang, J. Pei, Y. Yuan, C. Fan, and L. Xiao. **Urban Traffic Prediction through the Second Use of Inexpensive Big Data from Buildings**. In *Proceedings of the Twenty-fifth ACM International Conference on Information and Knowledge Management (CIKM 2016)* (Industry track), Indianapolis, IN, USA, October 24-28, 2016. (Acceptance rate: 22/111)
84. J. Liu, L. Xiong, J. Pei, J. Luo, and H. Zhang. **Finding Pareto Optimal Groups: Group-based Skyline**. In *Proceedings of the Forty-second International Conference on Very Large Data Bases (VLDB'16)*, New Delhi, India, September 5-9, 2016.
85. Z. Wang^[visitor], L. Chu^[student], J. Pei, E. Chen, and A. Al-Barakati. **Tradeoffs between Density and Size in Extracting Dense Subgraphs: A Unified Framework** (full paper). In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16)*, San Francisco, CA, USA, August 18-21, 2016. (Acceptance rate: 43/316)
86. L. Chu^[student], Z. Wang^[visitor], J. Pei, J. Wang, Z. Zhao, and E. Chen. **Finding Gangs in War from Signed Networks**(full paper with poster presentation). In *Proceedings of the Twenty-second ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'16)*, San Francisco, CA, USA, August 13-17, 2016. (Acceptance rate: 142/784)
87. M. Ou, P. Cui, J. Pei, and W. Zhu. **Asymmetric Transitivity Preserving Graph Embedding** (full paper with oral presentation). In *Proceedings of the Twenty-second ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'16)*, San Francisco, CA, USA, August 13-17, 2016. (Acceptance rate: 70/784)
88. H-J. Hung, H-H. Shuai, D-N. Yang, L-H. Huang, W-C. Lee, J. Pei, and M-S. Chen. **When Social Influence Meets Item Inference** (full paper with oral presentation). In *Proceedings of the Twenty-second ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'16)*, San Francisco, CA, USA, August 13-17, 2016. (Acceptance rate: 70/784)

89. Y. Yang^[student], X. Mao^[student], J. Pei, and X. He. **Continuous Influence Maximization: What Discounts Should We Offer to Social Network Users?**. In *Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD'16)*, San Francisco, CA, USA, June 26-July 1, 2016.
90. D-W. Choi^[student], J. Pei, and X. Lin. **Finding the Minimum Spatial Keyword Cover**. In *Proceedings of the Thirty-second IEEE International Conference on Data Engineering (ICDE'16)*, Helsinki, Finland, May 16-20, 2016.
91. J. Hu^[student], Q. Qian, J. Pei, R. Jin, and S. Zhu. **Multi-clustering via Clustering Stability**. In *Proceedings of the Fifteenth IEEE International Conference on Data Mining series (ICDM'15)*, Atlantic City, NJ, USA, November 14-17, 2015. (Acceptance rate: 8.4%)
92. L. Duan^[visitor], G. Tang^[student], J. Pei, J. Bailey, A. Campbell^[student], and C. Tang. **Mining Outlying Aspects on Numeric Data**. In *Proceedings of the 2015 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD'15)*, Porto, Portugal, September 7-11, 2015. (Journal track)
93. L. Chu, S. Wang, S. Liu, Q. Huang, and J. Pei. **ALID: Scalable Dominant Cluster Detection**. In *Proceedings of the Forty-first International Conference on Very Large Data Bases (VLDB'15)*, Kohala Coast, HI, USA, August 31-September 4, 2015.
94. Y. Zhang, J. Tang, Z. Yang, J. Pei, and P.S. Yu. **COSNET: Connecting Heterogeneous Social Networks with Local and Global Consistency**. In *Proceedings of the Twenty-first ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'15)*, Sydney, NW, Australia, August 10-13, 2015. (Acceptance rate: 159/819)
95. K. Yu^[student], D. Wang, W. Ding, D. Small, J. Pei, X. Wu, and S. Islam. **Tornado Forecasting with Multiple Markov Boundaries**. In *Proceedings of the Twenty-first ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'15)*, Sydney, NW, Australia, August 10-13, 2015.
96. N. X. Vinh, J. Chan, J. Bailey, C. Leckie, K. Ramamohanarao, and J. Pei. **Scalable Outlying-Inlying Aspects Discovery via Feature Ranking**. In *Proceedings of the Nineteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'15)*, Ho Chi Minh City, Viet Nam, May 19-22, 2015. (Acceptance rate: 90/405)
97. Y.-F. Lin, H.-H. Chen, V. S. Tseng, and J. Pei. **Reliable Early Classification on Multivariate Time Series with Numerical and Categorical Attributes**. In *Proceedings of the Nineteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'15)*, Ho Chi Minh City, Viet Nam, May 19-22, 2015. (Acceptance rate: 90/405)
98. J. Wang, S. Song, X. Lin, X. Zhu, and J. Pei. **Cleaning Structured Event Logs: A Graph Repair Approach**. In *Proceedings of the Thirty-first IEEE International Conference on Data Engineering (ICDE'15)*, Seoul, Korea, April 13-17, 2015. (Acceptance rate: 103/409)
99. Z. Yu, X. Yu, Y. Liu, W. Li, and J. Pei. **Mining Frequent Co-occurrence Patterns across Multiple Data Streams**. In *Proceedings of the Eighteenth International Conference on Extending Database Technology (EDBT'15)*, Brussels, Belgium, March 23-27, 2015. (Acceptance rate: 47/184)
100. L. Chang, X. Lin, L. Qin, J.X. Yu, and J. Pei. **Efficiently Computing Top-K Shortest Path Join**. In *Proceedings of the Eighteenth International Conference on Extending Database Technology (EDBT'15)*, Brussels, Belgium, March 23-27, 2015. (Acceptance rate: 47/184)
101. K. Yu^[student], W. Ding, X. Wu, and J. Pei. **Towards Scalable and Accurate Online Feature Selection for Big Data**. In *Proceedings of the Fourteenth IEEE International Conference on Data Mining (ICDM'14)*, Shenzhen, China, December 14-17, 2014. (Full paper acceptance rate: 71/727)
102. T. Guo, X. Zhu, J. Pei, and C. Zhang. **SNOC: Streaming Network Node Classification**. In *Proceedings of the Fourteenth IEEE International Conference on Data Mining (ICDM'14)*, Shenzhen, China, December 14-17, 2014. (Full paper acceptance rate: 71/727)
103. J. Han, J. Wen, and J. Pei. **Within-Network Classification Using Radius-Constrained Neighborhood Patterns**. In *Proceedings of the Twenty-third ACM International Conference on Information*

- and Knowledge Management (CIKM'14), Shanghai, China, November 3-7, 2014. (Knowledge Management track acceptance rate: 95/457)
104. G. Tang^[student], K. Wu, J. Pei, J. Tang, and J. Lei. **An Appliance-driven Approach to Detection of Corrupted Load Curve Data**. In *Proceedings of the Twenty-third ACM International Conference on Information and Knowledge Management (CIKM'14)*, Shanghai, China, November 3-7, 2014. (Database track acceptance rate: 25/123)
 105. X. Hu^[visitor], J. Pei, and Y. Tao^[visitor]. **Shortest Unique Queries on Strings**. In *Proceedings of the Twenty-first International Symposium on String Processing and Information Retrieval (SPIRE'14)*, Ouro Preto, Brazil, October 20-23, 2014. (Full paper acceptance rate: 20/45)
 106. W. Yu, X. Lin, W. Zhang, L. Chang, and J. Pei. **More is Simpler: Effectively and Efficiently Assessing Node-Pair Similarities Based on Hyperlinks**. In *Proceedings of the Fortieth International Conference on Very Large Data Bases (VLDB'14)*, Hangzhou, China, September 1-5, 2014.
 107. Q. Qian, J. Hu^[student], R. Jin, and J. Pei. **Distance Metric Learning Using Dropout: A Structured Regularization Approach**. In *Proceedings of the Twentieth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'14)*, New York, NY, USA, August 24-27, 2014. (Acceptance rate: 151/1036)
 108. L. Zhang^[visitor], J. Pei, Y. Jia, B. Zhou, and X. Wang. **Do Neighbor Buddies Make a Difference in Reblog Likelihood? An Analysis on SINA Weibo Data** (full paper with long presentation). In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM'14)*, Beijing, China, August 17-20, 2014. (Full paper acceptance rate: 18%)
 109. L. Duan^[visitor], G. Tang^[student], J. Pei, J. Bailey, G. Dong, A. Campbell^[student], and C. Tang. **Mining Contrast Subspaces**. (Regular paper with long presentation) In *Proceedings of the Eighteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'14)*, Tainan, Taiwan, May 13-16, 2014. (Acceptance rate: 40/371)
 110. J. Chan, X.V. Nguyen, W. Liu, J. Bailey, C. Leckie, K. Ramamohanarao, and J. Pei. **Structure-aware Distance Measures for Comparing Clusterings in Graphs**. (Regular paper with short presentation) In *Proceedings of the Eighteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'14)*, Tainan, Taiwan, May 13-16, 2014. (Acceptance rate: 101/371)
 111. Y. Wang, J. Pei, X. Lin, and Q. Zhang. **An Iterative Fusion Approach to Graph-based Semi-supervised Learning from Multiple Views**. (Regular paper with short presentation) In *Proceedings of the Eighteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'14)*, Tainan, Taiwan, May 13-16, 2014. (Acceptance rate: 101/371)
 112. J. Hu^[student], J. Pei, and J. Tang. **How Can I Index My Thousands of Photos Effectively and Automatically? An Unsupervised Feature Selection Approach**. (Full paper) In *Proceedings of the Fourteenth SIAM International Conference on Data Mining (SDM'14)*, Philadelphia, PA, USA, April 24-26, 2014.
 113. Y. Li, J. Bailey, L. Kulik, and J. Pei. **Efficient Matching of Substrings in Uncertain Sequences**. (Poster) In *Proceedings of the Fourteenth SIAM International Conference on Data Mining (SDM'14)*, Philadelphia, PA, USA, April 24-26, 2014.
 114. G. Tang^[student], Y. Yang, and J. Pei. **Price Information Patterns in Web Search Advertising: An Empirical Case Study on Accommodation Industry**. In *Proceedings of the Thirteenth IEEE International Conference on Data Mining (ICDM'13)*, Dallas, TX, USA, December 7-10, 2013. (Acceptance rate: 94/809)
 115. C.L. Kam, C. Raïssi, M. Kaytoue, and J. Pei. **Mining Statistically Significant Sequential Patterns**. In *Proceedings of the Thirteenth IEEE International Conference on Data Mining (ICDM'13)*, Dallas, TX, USA, December 7-10, 2013. (Acceptance rate: 94/809)
 116. Y. Li, J. Bailey, L. Kulik, and J. Pei. **Mining Probabilistic Frequent Spatio-Temporal Sequential Patterns with Gap Constraints from Uncertain Databases**. In *Proceedings of the Thirteenth IEEE International Conference on Data Mining (ICDM'13)*, Dallas, TX, USA, December 7-10, 2013. (Acceptance rate: 94/809)

117. X. Mao^[student], B. Lin, X. He, D. Cai, and J. Pei. **Parallel Field Alignment for Cross Media Retrieval**. In *Proceedings of the Twenty-first ACM International Conference on Multimedia (MM'13)*, Barcelona, Catalunya, Spain, October 21-25, 2013.
118. Y-C. Lo, J-Y. Li, M-Y. Yeh, S-D. Lin, and J. Pei. **What Distinguish One from Its Peers in Social Networks?**. In *Proceedings of the 2013 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD'13)*, Prague, Czech Republic, September 23-27, 2013. (Journal track, acceptance rate: 14/182)
119. Y. Wang, P. Wang, J. Pei, and W. Wang. **A Data-adaptive and Dynamic Segmentation Index for Whole Matching on Time Series**. In *Proceedings of the Thirty-Ninth International Conference on Very Large Data Bases (VLDB'13)*, Riva del Garda, Trento, Italy, August 26-30, 2013.
120. G. Tang^[student], J. Pei, J. Bailey, and G. Dong. **Mining Multidimensional Contextual Outliers from Categorical Relational Data**. In *Proceedings of the Twenty-fifth International Conference on Scientific and Statistical Database Management (SSDBM'13)*, Baltimore, Maryland, USA, July 29-31, 2013.
121. Y. Xiong, Y. Zhu, J. Pei, and P.S. Yu. **Towards Cohesive Anomalies Mining**. In *Proceedings of the Twenty-seventh AAAI Conference on Artificial Intelligence (AAAI'13)*, Bellevue, WA, USA, July 14-18, 2013.
122. J. Pei, W. C-H. Wu, and M-Y. Yeh^[visitor]. **On Shortest Unique Substring Queries**. In *Proceedings of the Twenty-ninth IEEE International Conference on Data Engineering (ICDE'13)*, Brisbane, Queensland, Australia, April 8-12, 2013. (Acceptance rate: 20%)
123. H. Maserrat^[student] and J. Pei. **Community Preserving Lossy Compression of Social Networks**. In *Proceedings of the Twelfth IEEE International Conference on Data Mining (ICDM'12)*, Brussels, Belgium, December 10-13, 2012. (Acceptance rate: 81/756)
124. T. Dwyer, A. Fedorova, S. Blagodurov, M. Roth, F. Gaud, and J. Pei. **A Practical Method for Estimating Performance Degradation on Multicore Processors and its Application to HPC Workloads**. In *Proceedings of the Twenty-fifth International Conference for High Performance Computing, Networking, Storage and Analysis (SC'12)*, Salt Lake City, UT, USA, November 10-16, 2012. (Acceptance rate: 100/472)
125. W. Liu, A. Kan, J. Chan, J. Bailey, C. Leckie, R. Kotagiri, and J. Pei. **On Compressing Weighted Time-evolving Graphs**. In *Proceedings of the Twenty-first ACM International Conference on Information and Knowledge Management (CIKM'12)*, Maui, HI, USA, October 29-November 2, 2012.
126. Y. Qian, H. Li, D. Jiang, Y. Hu, J. Pei, and Q. Zheng. **Mining Query Subtopics from Search Log Data**. In *Proceedings of the Thirty-fifth Annual ACM SIGIR Conference (SIGIR'12)*, Portland, OR, USA, August 12-16, 2012. (Acceptance rate: 98/483)
127. W.C-H. Wu, M-Y. Yeh, and J. Pei. **Random Error Reduction in Similarity Search on Time Series: A Statistical Approach**. In *Proceedings of the Twenty-eighth IEEE International Conference on Data Engineering (ICDE'12)*, Washington, DC, USA, April 1-5, 2012. (Acceptance rate: 17.7%)
128. M. Hua^[student] and J. Pei. **Aggregate Queries on Probabilistic Record Linkages**. In *Proceedings of the Fifteenth International Conference on Extending Database Technology (EDBT'12)*, Berlin, Germany, March 26-30, 2012. (Acceptance rate: 43/193)
129. C. Wang, L.Y. Yuan, J-H. You, O.R. Zaiane, and J. Pei. **On Pruning for Top-k Ranking in Uncertain Databases**. In *Proceedings of the Thirty-seventh International Conference on Very Large Data Bases (VLDB'11)*, Seattle, WA, USA, August 29-September 3, 2011.
130. C. Raïssi and J. Pei. **Towards Bounding Sequential Patterns**. In *Proceedings of the Seventeenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'11)*, San Diego, CA, USA, August 21-24, 2011. (Acceptance rate: 125/714)
131. Y. Tao, C. Sheng^[visitor], and J. Pei. **On k-skip Shortest Paths**. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data (SIGMOD'11)*, Athens, Greece, June 12-16, 2011.

132. Z. Xing^[student], J. Pei, P. S. Yu, and K. Wang. **Extracting Interpretable Features for Early Classification on Time Series**. In *Proceedings of the Eleventh SIAM International Conference on Data Mining (SDM'11)*, Phoenix, AZ, USA, April 28-30, 2011. (Acceptance rate: 86/343)
133. B. Jiang^[student] and J. Pei. **Outlier Detection on Uncertain Data: Objects, Instances, and Inferences**. In *Proceedings of the Twenty-seventh IEEE International Conference on Data Engineering (ICDE'11)*, Hannover, Germany, April 11-16, 2011. (Acceptance rate: 98/494)
134. D. Kang, D. Jiang^[student], J. Pei, Z. Liao, X. Sun, and H-J. Choi. **Multidimensional Mining of Large-Scale Search Logs: A Topic-Concept Cube Approach**. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM'11)*, Hong Kong, China, February 9-12, 2011. (Acceptance rate: 83/372)
135. Q. He, D. Kifer, J. Pei, P. Mitra, and L. Giles. **Citation Recommendation without Author Supervision**. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM'11)*, Hong Kong, China, February 9-12, 2011. (Acceptance rate: 83/372)
136. R. C.-W. Wong, A. W.-C. Fu, K. Wang, Y. Xu, J. Pei, and P.S. Yu. **Probabilistic Inference Protection on Uncertain Data**. In *Proceedings of the Tenth IEEE International Conference on Data Mining (ICDM'10)*, Sydney, Australia, December 14-17, 2010. (Acceptance rate: 155/797)
137. C. Raïssi, J. Pei, and T. Kister. **Computing Closed Skycubes**. In *Proceedings of the Thirty-sixth International Conference on Very Large Data Bases (VLDB'10)*, Singapore, September 13-17, 2010. (Acceptance rate: 16.1%)
138. H. Maserrat^[student] and J. Pei. **Neighbor Query Friendly Compression of Social Networks**. In *Proceedings of the Sixteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'10)*, Washington, DC, USA, July 25-28, 2010. (Acceptance rate: 77/578)
139. B. Xiang, D. Jiang^[student], J. Pei, X. Sun, E. Chen, and H. Li. **Context-Aware Ranking in Web Search**. In *Proceedings of the Thirty-third Annual ACM SIGIR Conference (SIGIR'10)*, Geneva, Switzerland, July 19-23, 2010. (Acceptance rate: 87/520)
140. Y. Tao, K. Yi, C. Sheng, J. Pei, and F. Li. **Logging Every Footstep: Quantile Summaries for the Entire History**. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data (SIGMOD'10)*, Indianapolis, Indiana, USA, June 6-11, 2010. (Acceptance rate: 20%)
141. Q. He, J. Pei, D. Kifer, P. Mitra, and C.L. Giles. **Context-aware Citation Recommendation**. In *Proceedings of the Nineteenth International World Wide Web Conference (WWW'10)*, Raleigh, NC, USA, April 26-30, 2010. (Acceptance rate: 104/743)
142. M. Hua^[student] and J. Pei. **Probabilistic Path Queries in Road Networks: Traffic Uncertainty Aware Path Selection**. In *Proceedings of the Thirteenth International Conference on Extending Database Technology (EDBT'10)*, Lausanne, Switzerland, March 22-26, 2010. (Acceptance rate: 54/307)
143. Y. Tao, J. Pei, J. Li, X. Xiao, K. Yi, and Z. Xing^[student]. **Hiding Correlation by Independence Masking**. In *Proceedings of the Twenty-Sixth International Conference on Data Engineering (ICDE'10)*, Long Beach, California, USA, March 1-6, 2010. (Acceptance rate: 20%)
144. Q. He, B. Chen, J. Pei, B. Qiu, P. Mitra, and C. L. Giles. **Detecting Topic Evolution in Scientific Literature: How Can Citations Help?**. In *Proceedings of the Eighteenth ACM Conference on Information and Knowledge Management (CIKM'09)*, Hong Kong, November 2-6, 2009. (Acceptance rate: 123/847)
145. X. Cheng, J. Xu^[student], J. Pei, and J. Liu. **Hierarchical Distributed Data Classification in Wireless Sensor Networks**. In *Proceedings of the Sixth IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS'09)*, Macau, China, October 12-15, 2009. (Acceptance rate: 62/245, selected for the special issue in *Computer Communication*.)
146. Y. Zhao, H. Zhang, L. Cao, J. Pei, and C. Zhang. **Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns**. In *Proceedings of the 2009 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'09)*, Bled, Slovenia, September 7-11, 2009.

147. Z. Xing^[student], J. Pei, and P. S. Yu. **Early Classification on Time Series: A Nearest Neighbor Approach**. In *Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI'09)*, Pasadena, CA, USA, July 14-17, 2009. (Acceptance rate: 331/1,290)
148. H. Zhong, T. Xie, L. Zhang, J. Pei, and H. Mei. **MAPO: Mining and Recommending API Usage Patterns**. In *Proceedings of the Twenty-third European Conference on Object-Oriented Programming (ECOOP 2009)*, Genova, Italy, July 6-10, 2009. (Acceptance rate: 25/117)
149. B. Zhou^[student], D. Jiang^[student], J. Pei, and H. Li. **OLAP on Search Logs: An Infrastructure Supporting Data-Driven Applications in Search Engines**. In *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, Paris, France, June 28 - July 1, 2009. (Acceptance rate: 12/122)
150. J. Wang, X. He, C. Wang, J. Pei, J. Bu, C. Chen, Z. Guan, W. V. Zhang. **Can We Learn a Template-Independent Wrapper for News Article Extraction from a Single Training Site?**. In *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, Paris, France, June 28 - July 1, 2009.
151. Y. Han^[visitor], B. Zhou^[student], J. Pei and Y. Jia. **Understanding Importance of Collaborations in Co-authorship Networks: A Supportiveness Analysis Approach**. In *Proceedings of 2009 SIAM International Conference on Data Mining (SDM'09)*, April 30 - May 2, 2009, Sparks, Nevada. (Acceptance rate: 54/351)
152. H. Cao, D. Jiang^[student], J. Pei, E. Chen, and H. Li. **Towards Context-Aware Search by Learning A Very Large Variable Length Hidden Markov Model from Search Logs**. In *Proceedings of the Eighteenth International World Wide Web Conference (WWW'09)*, April 20-24, 2009, Madrid, Spain. (Acceptance rate: 104/888)
153. J. Wang, X. He, C. Wang, J. Pei, J. Bu, C. Chen, and Z. Guan. **News Article Extraction with Template-Independent Wrapper**. In *Proceedings of the Eighteenth International World Wide Web Conference (WWW'09)* (poster), April 20-24, 2009, Madrid, Spain. (Acceptance rate: 93/225)
154. B. Jiang^[student] and J. Pei. **Online Interval Skyline Queries on Time Series**. In *Proceedings of the Twenty-fifth IEEE International Conference on Data Engineering (ICDE'09)*, March 29 - April 4, 2009, Shanghai, China. (Acceptance rate: 93/554)
155. Y. Tao, L. Ding, X. Lin, and J. Pei. **Distance-based Representative Skyline**. In *Proceedings of the Twenty-fifth IEEE International Conference on Data Engineering (ICDE'09)*, March 29 - April 4, 2009, Shanghai, China. (Acceptance rate: 93/554)
156. J. Pei, Y. Tao, J. Li, X. Xiao. **Privacy Preserving Publishing on Multiple Quasi-Identifiers**. In *Proceedings of the Twenty-fifth IEEE International Conference on Data Engineering (ICDE'09)*, March 29 - April 4, 2009, Shanghai, China. (Acceptance rate: 150/554)
157. B. Zhou^[student] and J. Pei. **Answering Aggregate Keyword Queries on Relational Databases Using Minimal Group-bys**. In *Proceedings of the Twelfth International Conference on Extending Database Technology (EDBT'09)*, March 23-26, 2009, Saint-Petersburg, Russia. (Acceptance rate: 92/283)
158. Y. Xiao, W. Wu, J. Pei, W. Wang, and Z. He. **Efficiently Indexing Shortest Paths by Exploiting Symmetry in Graphs**. In *Proceedings of the Twelfth International Conference on Extending Database Technology (EDBT'09)*, March 23-26, 2009, Saint-Petersburg, Russia. (Acceptance rate: 92/283)
159. B. Zhou^[student], Y. Han^[visitor], J. Pei, B. Jiang^[student], Y. Tao, and Y. Jia. **Continuous Privacy Preserving Publishing of Data Streams**. In *Proceedings of the Twelfth International Conference on Extending Database Technology (EDBT'09)*, March 23-26, 2009, Saint-Petersburg, Russia. (Acceptance rate: 92/283)
160. K. Tsoukalas^[student], B. Zhou^[student], J. Pei, and D. Cubranic. **PLEDS: A Personalized Entity Detection System Based on Web Log Mining Techniques** (industrial track). In *Proceedings of the Twelfth International Conference on Extending Database Technology (EDBT'09)*, March 23-26, 2009, Saint-Petersburg, Russia.

161. Y. Xu, B. Fang, K. Wang, A. W.-C. Fu, and J. Pei. **Publishing Sensitive Transactions for Item-set Utility**. In *Proceedings of the Eighth IEEE International Conference on Data Mining (ICDM'08)*, December 15-19, 2008, Pisa, Italy. (Acceptance rate: 144/724)
162. B. Jiang^[student], J. Pei, X. Lin, D. W.-L. Cheung, and J. Han. **Mining Preferences from Superior and Inferior Examples**. In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, August 24-27, 2008, Las Vegas, NV, USA. (Acceptance rate: 95/510)
163. H. Cao, D. Jiang^[student], J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. **Context-Aware Query Suggestion by Mining Click-Through and Session Data** (Best Application Paper Award). In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, August 24-27, 2008, Las Vegas, NV, USA. (Acceptance rate: 12/83)
164. R. C.-W. Wong^[visitor], A. W.-C. Fu, J. Pei, Y. S. Ho, T. Wong, and Y. Liu. **Efficient Skyline Querying with Variable User Preferences on Nominal Attributes**. In *Proceedings of the Thirty-fourth International Conference on Very Large Databases (VLDB'08)*, August 24-30, 2008, Auckland, New Zealand. (Acceptance rate: 46/273)
165. K. Tsoukalas^[student], B. Zhou^[student], J. Pei, and D. Cubranic. **PLEDS: A Personalized Entity Detection System Based on Web Log Mining Techniques** (invited paper). In *Proceedings of the Ninth International Conference on Web-Age Information Management (WAIM'08)*, July 20-22, 2008, Zhangjiajie, China.
166. W. Zhang, X. Lin, J. Pei, and Y. Zhang. **Managing Uncertain Data: A Probabilistic Approach** (invited paper). In *Proceedings of the Ninth International Conference on Web-Age Information Management (WAIM'08)*, July 20-22, 2008, Zhangjiajie, China.
167. M. Hua^[student], J. Pei, W. Zhang, and X. Lin. **Ranking Queries on Uncertain Data: A Probabilistic Threshold Approach**. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD'08)*, Vancouver, BC, Canada, June 9-12, 2008. (Acceptance rate: 78/435)
168. E. Soroush, K. Wu, and J. Pei. **Fast and Quality-Guaranteed Data Streaming in Resource-Constrained Sensor Networks**. In *Proceedings of the Ninth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'08)*, Hong Kong, China, May 26-30, 2008. (Acceptance rate: 44/300)
169. B. Zhou^[student], J. Pei, and Z. Tang. **A Spamicity Approach to Unsupervised Web Spam Detection** (full paper). In *Proceedings of the 2008 SIAM International Conference on Data Mining (SDM'08)*, Atlanta, GA, USA, April 24-26, 2008. (Acceptance rate: 40/282)
170. Z. Xing^[student], J. Pei, G. Dong, and P. S. Yu. **Mining Sequence Classifiers for Early Prediction** (poster paper). In *Proceedings of the 2008 SIAM International Conference on Data Mining (SDM'08)*, Atlanta, GA, USA, April 24-26, 2008. (Acceptance rate: 77/282)
171. B. Zhou^[student] and J. Pei. **Preserving Privacy in Social Networks against Neighborhood Attacks** (full presentation paper). In *Proceedings of the Twenty-fourth International Conference on Data Engineering (ICDE'08)*, Cancun, Mexico, April 7-12, 2008. (Acceptance rate: 12.1%)
172. M. Hua^[student], J. Pei, W. Zhang, and X. Lin. **Efficiently Answering Probabilistic Threshold Top-k Queries on Uncertain Data** (poster paper). In *Proceedings of the Twenty-fourth International Conference on Data Engineering (ICDE'08)*, Cancun, Mexico, April 7-12, 2008. (Acceptance rate: 31%)
173. X. Zeng^[student], J. Pei, I. Vergara, M. Nesbitt, K. Wang, and N. Chen. **OrthoCluster: A New Tool for Mining Syntenic Blocks and Applications in Comparative Genomics**. In *Proceedings of the Eleventh International Conferences on Extending Database Technology (EDBT'08)*, Nantes, France, March 25-30, 2008.
174. B. C. M. Fung, K. Wang, A. W.-C. Fu, and J. Pei. **Anonymity for Continuous Data Publishing**. In *Proceedings of the Eleventh International Conferences on Extending Database Technology (EDBT'08)*, Nantes, France, March 25-30, 2008.

175. F. M. Jiang, J. Pei, and A. W.-C. Fu. **IX-Cubes: Iceberg Cubes for Data Warehousing and OLAP on XML Data**. In *Proceedings of the ACM Sixteenth Conference on Information and Knowledge Management (CIKM'07)*, Lisbon, Portugal, November 6-9, 2007.
176. J. Pei, B. Jiang^[student], X. Lin, and Y. Yuan. **Probabilistic Skylines on Uncertain Data**. In *Proceedings of the Thirty-third International Conference on Very Large Data Bases (VLDB'07)*, Vienna, Austria, September 23-28, 2007. (Acceptance rate: 46/263)
177. M. Hua^[student], J. Pei, A. W.-C. Fu, X. Lin, and H-F Leung. **Efficiently Answering Top-k Typicality Queries on Large Databases**. In *Proceedings of the Thirty-third International Conference on Very Large Data Bases (VLDB'07)*, Vienna, Austria, September 23-28, 2007. (Acceptance rate: 46/263)
178. R. C.-W. Wong, A. W.-C. Fu, K. Wang, and J. Pei. **Minimality Attack in Privacy Preserving Data Publishing**. In *Proceedings of the Thirty-third International Conference on Very Large Data Bases (VLDB'07)*, Vienna, Austria, September 23-28, 2007. (Acceptance rate: 46/263)
179. M. Acharya, T. Xie, J. Pei, and J. Xu. **Mining API Patterns as Partial Orders from Source Code: From Usage Scenarios to Specifications**. In *Proceedings of the Sixth Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (ESEC/FSE'07)*, Dubrovnik, Croatia, September 3-7, 2007. (Acceptance rate: 43/251)
180. M. Hua^[student] and J. Pei. **Cleaning Disguised Missing Data: A Heuristic Approach**. In *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*, San Jose, California, USA, August 12-15, 2007.
181. R. C.-W. Wong^[visitor], J. Pei, A. W.-C. Fu, and K. Wang. **Mining Favorable Facets**. In *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*, San Jose, California, USA, August 12-15, 2007.
182. J. Pei, J. Xu, Z. Wang, W. Wang, and K. Wang. **Maintaining K-Anonymity against Incremental Updates**. In *Proceedings of the Nineteenth International Conference on Scientific and Statistical Database Management (SSDBM'07)*, Banff, Canada, July 9-11, 2007.
183. R. C.-W. Wong, Y. Liu, J. Yin, Z. Huang, A. W.-C. Fu, and J. Pei. **(α, k) -anonymity Based Privacy Preservation by Lossy Join**. In *Proceedings of the Ninth Asia-Pacific Web Conference and the Eighth International Conference on Web-Age Information Management (APWEB/WAIM'07)*, Huangshan, China, June 16-18, 2007. (Acceptance rate: 49/554)
184. J. Pei, M. K. Lau^[student], and P. S. Yu. **TS-Trees: A Non-Alterable Search Tree Index for Trustworthy Databases on Write-Once-Read-Many (WORM) Storage** (IEEE Outstanding Paper Award). In *Proceedings of the IEEE Twenty-first International Conference on Advanced Information Networking and Applications (AINA'07)*, Niagara Falls, ON, Canada, May 21-23, 2007. (Acceptance rate: 134/444)
185. B. Zhou^[student] and J. Pei. **Sketching Landscapes of Page Farms**. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM'07)*, Minneapolis, MN, USA, April 26-28, 2007. (Acceptance rate: 76/296)
186. Y. Bu, T.-W. Leung, A. W.-C. Fu, E. Keogh, J. Pei, and S. Meshkin. **WAT: Finding Top-K Discords in Time Series Database**. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM'07)*, Minneapolis, MN, USA, April 26-28, 2007. (Acceptance rate: 76/296)
187. J. Pei, A. W. C. Fu, X. Lin, and H. Wang. **Computing Compressed Skyline Cubes Efficiently**. In *Proceedings of the twenty-third IEEE International Conference on Data Engineering (ICDE'07)*, Istanbul, Turkey, April 16-20, 2007. (Acceptance rate: 122/659)
188. Y. Liu, L. Chen, J. Pei, Q. Chen, and Y. Zhao. **Mining Frequent Trajectory Patterns for Activity Monitoring Using Radio Frequency Tag Arrays** (one of the three papers in the best paper session). In *Proceedings of the Fifth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'07)*, White Plains, NY, USA, March 19-23, 2007. (Acceptance rate: 20/208)
189. B.-W. On, E. Elmacioglu, D. Lee, J. Kang, and J. Pei. **Improving Grouped-Entity Resolution using Quasi-Cliques**. In *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM'06)*, Hong Kong, December 18-22, 2006. (Acceptance rate: less than 20%)

190. Y. Xu, K. Wang, A. W. C. Fu, R. She, and J. Pei. **Classification Spanning Correlated Data Streams**. In *Proceedings of the ACM Fifteenth Conference on Information and Knowledge Management (CIKM'06)*, Arlington, VA, USA, November 6-11, 2006. (Acceptance rate: 15%)
191. W. Zhu, J. Pei, J. Yin, Y. Xie. **Granularity Adaptive Density Estimation and on-Demand Clustering of Concept-Drifting Data Streams**. In *Proceedings of the Eighth International Conference on Data Warehousing and Knowledge Discovery (DaWaK'06)*, Krakow, Poland, September 4-8, 2006. (Acceptance rate: 52/199)
192. J. Li, R. C.-W. Wong, A. W.-C. Fu, and J. Pei. **Achieving k-Anonymity by Clustering in Attribute Hierarchical Structures** (one of the six best papers selected for the journal special issue). In *Proceedings of the Eighth International Conference on Data Warehousing and Knowledge Discovery (DaWaK'06)*, Krakow, Poland, September 4-8, 2006. (Acceptance rate: 52/199)
193. C. Aggarwal, J. Pei, and B. Zhang^[student]. **On Privacy Preservation against Adversarial Data Mining**. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, Philadelphia, PA, USA, August 20-23, 2006. (Acceptance rate: 22%)
194. J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. W.-C. Fu. **Utility-Based Anonymization Using Local Recoding**. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, Philadelphia, PA, USA, August 20-23, 2006. (Acceptance rate: 22%)
195. H. Wang, J. Yin, J. Pei, P. S. Yu, and J. X. Yu. **Suppressing Model Overfitting in Mining Concept-Drifting Data Streams**. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, Philadelphia, PA, USA, August 20-23, 2006. (Acceptance rate: 22%)
196. J. Li, H. Li, L. Wong, J. Pei, and G. Dong. **Minimum Description Length (MDL) Principle: Generators Are Preferable to Closed Patterns**. In *Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI'06)*, Boston, MA, USA, July 16-20, 2006. (Acceptance rate: 171/774)
197. B.-W. On, D. Lee, E. Elmacioglu, J. Kang, and J. Pei. **An Effective Approach to Entity Resolution Problem Using Quasi-Clique and its Application to Digital Libraries** (short paper). In *Proceedings of the ACM/IEEE 2006 Joint Conference on Digital Libraries (JCDL'06)*, Chapel Hill, NC, USA, June 11-15, 2006.
198. Y. Tao, X. Xiao, and J. Pei. **SUBSKY: Efficient Computation of Skylines in Subspaces**. In *Proceedings of the Twenty-second International Conference on Data Engineering (ICDE'06)*, Atlanta, GA, USA, April 3-7, 2006. (Acceptance rate: 89/456)
199. J. Pei, J. Liu^[student], H. Wang, K. Wang, P. S. Yu, and J. Wang. **Efficiently Mining Frequent Closed Partial Orders**. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM'05)*, New Orleans, Louisiana, USA, November 27-30 2005. (Acceptance rate: 141/630)
200. H. Wang and J. Pei. **A Random Method for Quantifying Changing Distributions in Data Streams**. In *Proceedings of the Ninth European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'05)*, Porto, Portugal, October 3-7, 2005. (Acceptance rate: 70/250)
201. J. Ye, X. Zhou, J. Pei, L. Chen, and L. Zhang. **A Stratification-Based Approach to Accurate and Fast Image Annotation**. In *Proceedings of the Sixth International Conference on Web-Age Information Management (WAIM'05)*, Hangzhou, China, October 11-13, 2005. (Acceptance rate: 48/488)
202. C. Liu, K. Wu, and J. Pei. **A Dynamic Clustering and Scheduling Approach to Energy Saving in Data Collection from Wireless Sensor Networks**. In *Proceedings of the Second Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON'05)*, Santa Clara, California, USA, September 26-29, 2005. (Acceptance rate: 55/202)
203. J. Pei, W. Jin, M. Ester, and Y. Tao. **Catching the Best Views in Skyline: A Semantic Approach Based on Decisive Subspaces**. In *Proceedings of the Thirty-first International Conference on Very Large Data Bases (VLDB'05)*, Trondheim, Norway, August 30-September 2, 2005. (Acceptance rate: 16.5%)

204. J. Pei, D. Jiang^[student], and A. Zhang. **On Mining Cross-Graph Quasi-Cliques**. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'05)*, Chicago, IL, USA, August 21-24, 2005. (Acceptance rate: 40/343)
205. H. Wang, J. Pei, and P. S. Yu. **Pattern Based Similarity Search for Microarray Data** (Industrial and Government Track poster paper). In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'05)*, Chicago, IL, USA, August 21-24, 2005. (Acceptance rate: 25/75)
206. H. Yu, J. Pei, S. Tang, and D. Yang. **Mining Most General Multidimensional Summarization of Probable Groups in Data Warehouses**. In *Proceedings of the Seventeenth International Scientific and Statistical Database Management Conference (SSDBM'05)*, Santa Barbara, California, USA, June 27-29, 2005.
207. M. Cho^[student], J. Pei, and D. W.-L. Cheung. **Cross Table Cubing: Mining Iceberg Cubes from Data Warehouses** (poster paper). In *Proceedings of The Fifth SIAM International Conference on Data Mining (SDM'05)*, Newport Beach, CA, USA, April 21-23, 2005.
208. D. Jiang^[student], J. Pei, and A. Zhang. **A General Approach to Mining Quality Pattern-based Clusters from Gene Expression Data**. In *Proceedings of the Tenth International Conference on Database Systems for Advanced Applications (DASFAA'05)*, Beijing, China, April 18-20, 2005. (Acceptance rate: 67/302)
209. G. Dong, C. Jiang, J. Pei, J. Li, and L. Wong. **Mining Succinct Systems of Minimal Generators of Formal Concepts**. In *Proceedings of the Tenth International Conference on Database Systems for Advanced Applications (DASFAA'05)*, Beijing, China, April 18-20, 2005. (Acceptance rate: 67/302)
210. J. Pei, D. Jiang^[student], and A. Zhang. **Multi-Graph Mining: A Cross-Graph Quasi-Clique Approach** (research poster paper). In *Proceedings of the Twenty-first International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 5-8, 2005. (Acceptance rate: 100/521)
211. C. Wang, W. Wang, J. Pei, Y. Zhu and B. Shi. **Scalable Mining Large Disk-based Graph Databases** (research full paper). In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, Seattle, WA, USA, August 22-25, 2004. (Acceptance rate: 40/337)
212. D. Jiang^[student], J. Pei, M. Ramanathan, C. Tang, and A. Zhang. **Mining Coherent Gene Clusters from Three-Dimensional Microarray Data** (industrial full paper). In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, Seattle, WA, USA, August 22-25, 2004. (Acceptance rate: 13/47)
213. L. Deng, J. Pei, J. Ma, and D.L. Lee. **A Rank Sum Test Method for Informative Gene Discovery** (industrial full paper). In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, Seattle, WA, USA, August 22-25, 2004. (Acceptance rate: 13/47)
214. H. Wang, F. Chu, W. Fan, P. S. Yu, and J. Pei. **A Fast Algorithm for Subspace Clustering by Pattern Similarity** (full paper). In *Proceedings of the Sixteenth International Conference on Scientific and Statistical Database Management (SSDBM'04)*, Santorini Island, Greece, 21-23 June 2004.
215. W. Wang, B. Shi, C. Wang, H. Zhou, J. Pei, and M. Hong. **Efficient Pattern-Growth Methods for Frequent Tree Pattern Mining** (full paper). In *Proceedings of the Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'04)*, Sydney, Australia, May 26-28, 2004. (Acceptance rate: 50/235).
216. J. Pei, X. Zhang^[student], M. Cho^[student], H. Wang, and P. S. Yu. **MaPle: A Fast Algorithm for Maximal Pattern-based Clustering** (Regular paper). In *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, Florida, USA, November 19-22, 2003. (Acceptance rate: 58/501)
217. D. Jiang^[student], J. Pei, and A. Zhang. **Interactive Exploration of Coherent Patterns in Time-Series Gene Expression Data**. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, Washington, DC, USA, August 24-27, 2003. (Acceptance rate: 70/258)

218. C. Tang, A. Zhang, and J. Pei. **Mining Phenotypes and Informative Genes from Gene Expression Data**. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, Washington, DC, USA, August 24-27, 2003. (Acceptance rate: 70/258)
219. J. Wang, J. Han, and J. Pei. **CLOSET+: Searching for the Best Strategies for Mining Frequent Closed Itemsets**. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, Washington, DC, USA, August 24-27, 2003. (Acceptance rate: 70/258)
220. L. V. S. Lakshmanan, J. Pei, and Y. Zhao. **QC-Trees: An Efficient Summary Structure for Semantic OLAP**. In *Proceedings of the 2003 ACM-SIGMOD International Conference on Management of Data (SIGMOD'03)*, San Diego, CA, June 9-12, 2003. (Acceptance rate: 53/342)
221. J. Pei. **A General Model for Online Analytical Processing of Complex Data**. In *Proceedings of the Twenty-second International Conference on Conceptual Modeling (ER'03)*, Chicago, IL, October 13-16, 2003. (Acceptance rate: 38/153)
222. H. C. Kum, J. Pei, and W. Wang. **ApproxMAP: Approximate Mining of Consensus Sequential Patterns**. In *Proceedings of the 2003 SIAM International Conference on Data Mining (SIAM DM '03)*, San Francisco, CA, May 1-3, 2003.
223. D. Jiang^[student], J. Pei, and A. Zhang. **DHC: A Density-based Hierarchical Clustering Method for Time Series Gene Expression Data** (Regular paper). In *Proceedings of the Third IEEE Symposium on Bio-informatics and Bio-engineering (BIBE'03)*, Washington D.C., March 10-12, 2003. (Acceptance rate: 45/129)
224. Y. Huang, H. Xiong, S. Shekhar, and J. Pei. **Mining Confident Co-location Rules without A Support Threshold**. In *Proceedings of the Eighteenth Annual ACM Symposium on Applied Computing (SAC'03)*, Melbourne, Florida, March 9-12, 2003.
225. J. Pei, G. Dong, W. Zou, and J. Han. **On Computing Condensed Frequent Pattern Bases** (Regular paper). In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02)*, Maebashi TERRSA, Maebashi City, Japan, December 9-12, 2002.
226. J. Pei, J. Han, and W. Wang. **Mining Sequential Patterns with Constraints in Large Databases** (Regular paper). In *Proceedings of Eleventh International Conference on Information and Knowledge Management (CIKM'02)*, McLean, VA, November 4-9, 2002.
227. L. V. S. Lakshmanan, J. Pei, J. Han. **Quotient Cube: How to Summarize The Semantics of A Data Cube**. In *Proceedings of Twenty-eighth International Conference on Very Large Databases (VLDB'02)*, Hong Kong, China, August 20-23, 2002.
228. J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, and D. Yang. **H-Mine: Hyper-structure Mining of Frequent Patterns in Large Databases**. In *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01)*, San Jose, California, November 29-December 2, 2001.
229. W. Li, J. Han, and J. Pei. **CMAR: Accurate and Efficient Classification Based on Multiple Class-association Rules**. In *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01)*, San Jose, California, November 29-December 2, 2001.
230. H. Pinto, J. Han, J. Pei, K. Wang, Q. Chen, and U. Dayal. **Multi-Dimensional Sequential Pattern Mining**. In *Proceedings of the Tenth ACM International Conference on Information and Knowledge Management (CIKM'01)*, Atlanta, Georgia, November 2001.
231. G. Dong, J. Han, J. Lam, J. Pei, and K. Wang. **Mining Multi-Dimensional Constrained Gradients in Data Cubes**. In *Proceedings of the Twenty-seventh International Conference on Very Large Data Base (VLDB'01)*, Rome, Italy, September 2001.
232. J. Han, J. Pei, G. Dong, and K. Wang. **Efficient Computation of Iceberg Cubes with Complex Measures**. In *Proceedings of the 2001 ACM-SIGMOD International Conference on Management of Data (SIGMOD'01)*, Santa Barbara, CA, May 2001.

233. J. Pei, J. Han, and L. V. S. Lakshmanan. **Mining Frequent Itemsets With Convertible Constraints.** In *Proceedings of the 2001 IEEE International Conference on Data Engineering (ICDE'01)*, Heidelberg, Germany, April 2001.
234. J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M. Hsu. **PrefixSpan: Mining Sequential Patterns by Prefix-Projected Pattern Growth.** In *Proceedings of the 2001 IEEE International Conference on Data Engineering (ICDE'01)*, Heidelberg, Germany, April 2001.
235. J. Pei and J. Han. **Can We Push More Constraints into Frequent Pattern Mining?.** In *Proceedings of the 2000 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2000)*, Boston, MA, August 2000.
236. J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M. Hsu. **FreeSpan: Frequent pattern-projected sequential pattern mining.** In *Proceedings of the 2000 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2000)*, Boston, MA, August 2000.
237. J. Han, J. Pei, and Y. Yin. **Mining Frequent Patterns without Candidate Generation.** In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD'00)*, Dallas, TX, May 2000.
238. J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu. **Mining Access Patterns efficiently from Web logs.** In *Proceedings of the 2000 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'00)*, Kyoto, Japan, April 2000.

7.5 Research Articles in Refereed Workshop Proceedings

1. H. Wang, Q. Zhu, and J. Pei. **Clinical Assessment of Pneumocystosis with MIMIC Data.** In *Proceedings of the IEEE BIBM 2023 workshop: Biomedical Informatics Applications in Translational Research and Rare Diseases*, Istanbul, Turkey, December 5-8, 2023.
2. J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. W.-C. Fu. **Utility-Based Anonymization for Privacy Preservation with Less Information Loss** (invited paper, one of the two best papers in the workshop selected for the special issue in *ACM SIGKDD Explorations*). In *Proceedings of the Second ACM SIGKDD Workshop on Utility-Based Data Mining (UBDM'06)*, in conjunction with the *Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, Philadelphia, PA, August 20, 2006.
3. T. Xie and J. Pei. **MAPO: Mining API Usages from Open Source Repositories** (short paper). In *Proceedings of the Third International Workshop on Mining Software Repositories (MSR 2006)*, Shanghai, China, May 22-23, 2006.
4. G. Dong, J. Han, L. V. S. Lakshmanan, J. Pei, H. Wang, and P. S. Yu. **Online mining of changes from data streams: Research problems and preliminary results.** In *Proceedings of the 2003 ACM SIGMOD Workshop on Management and Processing of Data Streams*, in cooperation with 2003 ACM-SIGMOD International Conference on Management of Data (SIGMOD'03), San Diego, CA, June 8, 2003.
5. Y. Chen, G. Dong, J. Han, J. Pei, B. Wah, J. Wang. **Online Analytical Processing Stream Data: Is It Feasible?.** In *Proceedings of 2002 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'2002)*, Madison, Wisconsin, June 2, 2002.
6. J. Pei, A. K. H. Tung, and J. Han. **Fault-tolerant frequent pattern mining: Problems and Challenges.** In *Proceedings of the 2001 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'01)*, Santa Barbara, CA, May 2001.
7. J. Pei, J. Han, and R. Mao. **CLOSET: An efficient algorithm for mining frequent closed itemsets.** In *Proceedings of the 2000 ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'00)*, Dallas, TX, May, 2000.

7.6 Tutorials

1. L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo. **Graph Neural Networks: Foundation, Frontiers, and Applications**. In *Proceedings of the 2023 ACM Web Conference (WebConf'23)*, Austin, TX, USA, April 30 - May 4, 2023.
2. J. Pei and X. Yu. **Data Acquisition and Valuation in Data Markets**. In *Proceedings of the Thirty-ninth IEEE International Conference on Data Engineering (ICDE'23)*, Anaheim, CA, USA, April 3-7, 2023.
3. P. Cui and J. Pei. **Data-Centric Trustworthy AI**. In *Proceedings of the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, Washington, DC, USA, August 14-18, 2022.
4. L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo. **Graph Neural Networks: Foundation, Frontiers, and Applications**. In *Proceedings of the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, Washington, DC, USA, August 14-18, 2022.
5. L. Wu, P. Cui, J. Pei, L. Zhao, and X. Guo. **Graph Neural Networks: Foundation, Frontiers, and Applications**. In *Proceedings of the Thirty-first International Joint Conference on Artificial Intelligence (IJCAI'22)*, Messe Wien, Vienna, Austria, July 23-29, 2022.
6. J. Pei, Z. Cong^[student], X. Luo^[student], and F. Zhu. **Data and Model Pricing in the Pipeline of Machine Learning**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
7. L. Shou, M. Gong, J. Pei, X. Geng, X. Zhou, and D. Jiang. **Language Scaling: Applications, Challenges and Approaches**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
8. Z. Zhou, L. Chu, L. Wang, J. Pei, and Y. Zhang. **Towards Fair Federated Learning**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
9. F. Zhu, H. Liu, X. Wu, and J. Pei. **Data Asset for Collaborative Intelligence**. In *Proceedings of the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, Singapore, August 14-18, 2021.
10. X. Hu^[student], J. Pei, L. Chu, J. Bian, and W. Liu. **Deep Learning Model Complexity: Concepts and Approaches**. In *Proceedings of the Twenty-first SIAM International Conference on Data Mining (SDM'21)*, April 29 - May 1, 2021, Alexandria, VA, USA.
11. L. Shou, M. Gong, J. Pei, X. Geng, X. Zhou, and D. Jiang. **Scaling NLP Applications to 100+ Languages**. In *Proceedings of the Thirtieth The Web Conference (WWW'21)*, Ljubljana, Slovenia, April 19-23, 2021.
12. F. Zhu and J. Pei. **Data Asset and Governance: Opportunities and Challenges at the Frontier**. In *Proceedings of the Twentieth IEEE International Conference on Data Mining (ICDM'20)*, Sorrento, Italy, November 17-20, 2020.
13. J. Pei. **Data Pricing: From Economics to Data Science**. In *Proceedings of the Twenty-sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)*, San Diego, CA, USA, August 23-27, 2020.
14. F. Wang, P. Cui, J. Pei, Y. Song, and C. Zang. **Recent Advances on Graph Analytics and Its Applications in Healthcare**. In *Proceedings of the Twenty-sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)*, San Diego, CA, USA, August 23-27, 2020.
15. X. Huang, P. Cui, Y. Dong, J. Li, H. Liu, J. Pei, L. Song, J. Tang, F. Wang, H. Yang, and W. Zhu. **Learning From Networks: Algorithms, Theory, and Applications**. In *Proceedings of the Twenty-fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*, Anchorage, AK, USA, August 4-8, 2019.

16. Z.-J. M. Shen, R. Yuan, D. Wu, and J. Pei. **Data Science in Retail-as-a-Service**. In *Proceedings of the Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*, London, United Kingdom, August 19-23, 2018.
17. P. Cui, J. Pei, W. Zhu, T. Berger-Wolf, I. Brugere, and B. Perozzi. **Modeling Data With Networks + Network Embedding: Problems, Methodologies and Frontiers**. In *Proceedings of the Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*, London, United Kingdom, August 19-23, 2018.
18. P. Cui, J. Pei, and W. Zhu. **Network Embedding-Enabling Network Analytics and Inference in Vector Space**. In *Proceedings of the Twenty-third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*, Halifax, NS, Canada, August 13-17, 2017.
19. J. Pei. **Mining Uncertain and Probabilistic Data for Big Data Analytics**. In *Proceedings of the Twelfth IEEE International Conference on Data Mining (ICDM'12)*, Brussels, Belgium, December 10-13, 2012.
20. D. Jiang^[student], J. Pei, and H. Li. **Enhancing Web Search by Mining Search and Browse Logs**. In *Proceedings of the Thirty-fourth Annual ACM SIGIR Conference (SIGIR'11)*, Beijing, China, July 24-28, 2011.
21. M. Hay, K. Liu, G. Miklau, J. Pei, and E. Terzi. **Privacy-aware Data Management in Information Networks**. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data (SIGMOD'11)*, Athens, Greece, June 12-16, 2011.
22. K. Liu, G. Miklau, J. Pei, and E. Terzi. **Privacy-aware Data Mining in Information Networks**. In *Proceedings of the Sixteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'10)*, Washington, DC, USA, July 25-28, 2010.
23. D. Jiang^[student], J. Pei, and H. Li. **Web Search/Browse Log Mining: Challenges, Methods, and Applications**. In *Proceedings of the Sixteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'10)*, Washington, DC, USA, July 25-28, 2010.
24. D. Jiang^[student], J. Pei, and H. Li. **Web Search/Browse Log Mining: Challenges, Methods, and Applications**. In *Proceedings of the 33rd Annual ACM SIGIR Conference (SIGIR'10)*, Geneva, Switzerland, July 19-23, 2010.
25. D. Jiang^[student], J. Pei, and H. Li. **Web Search/Browse Log Mining: Challenges, Methods, and Applications**. In *Proceedings of the Nineteenth International World Wide Web Conference (WWW'10)*, Raleigh, NC, USA, April 26-30, 2010.
26. J. Han and J. Pei. **Preference Queries from OLAP and Data Mining Perspective**. In *Proceedings of the Twenty-fifth IEEE International Conference on Data Engineering (ICDE'09)*, Shanghai, China, March 29 - April 4, 2009.
27. J. Pei, M. Hua^[student], Y. Tao, and X. Lin. **Mining Uncertain and Probabilistic Data: Problems, Challenges, Methods and Applications**. In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, Las Vegas, NV, USA, August 24-27, 2008.
28. J. Pei, M. Hua^[student], Y. Tao, and X. Lin. **Query Answering Techniques on Uncertain and Probabilistic Data**. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD'08)*, Vancouver, Canada, June 9-12, 2008.
29. J. Pei, B. Zhou^[student], Z. Tang, and H. Huang^[student]. **Data Mining Techniques for Web Spam Detection**. In *Proceedings of the Twelfth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'08)*, Osaka, Japan, May 20-23, 2008.
30. T. Xie and J. Pei. **Data Mining for Software Engineering**. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Data Mining (KDD'06)*, Philadelphia, USA, August 20-23, 2006.
31. J. Pei, H. Wang and P. S. Yu. **Online Mining Data Streams: Problems, Applications and Progress**. In *Proceedings of the Sixth International Conference on Web-Age Information Management (WAIM'05)*, Hangzhou, China, October 11-13, 2005.

32. H. Wang, J. Pei and P. S. Yu. **Online Mining Data Streams: Problems, Applications and Progress.** In *Proceedings of the Twenty-first International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 5-8, 2005.
33. J. Pei, H. Wang and P. S. Yu. **Online Mining Data Streams: Problems, Applications and Progress.** In *Proceedings of the Tenth ACM International Conference on Data Mining (KDD'04)*, Seattle, WA, August 22-25, 2004.
34. J. Pei, S. J. Upadhyaya, F. Farooq and V. Govindaraju. **Data Mining for Intrusion Detection: Techniques, Applications and Systems.** In *Proceedings of the Twentieth IEEE International Conference on Data Engineering (ICDE'04)*, Boston, MA, March 30-April 2, 2004.
35. J. Han, L. V. S. Lakshmanan and J. Pei. **Mining Frequent Patterns: Methods and Applications.** In *the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2001)*, San Francisco, California, USA, August 26-29, 2001.
36. J. Pei and J. Han. **Sequential Pattern Mining: From Shopping History Analysis to Weblog Mining and DNA Mining.** In *the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'01)*, Hong Kong, China, April 16-18, 2001.

7.7 Book Chapters

1. M. Hua^[student] and J. Pei, **A Survey of Utility-based Privacy-preserving Data Transformation Methods**, in C. Aggarwal and P. S. Yu (eds.), *Privacy-Preserving Data Mining: Models and Algorithms*, Springer-Verlag, 2007.
2. Y. Xu, K. Wang, A. W.-C. Fu, R. She, and J. Pei. **Privacy-preserving Data Stream Classification**, in C. Aggarwal and P. S. Yu (eds.), *Privacy-Preserving Data Mining: Models and Algorithms*, Springer-Verlag, 2007.
3. J. Han, J. Pei, and X. Yan. **Sequential Pattern Mining by Pattern-Growth: Principles and Extensions**, in W. Chu and T. Lin (eds.), *Foundations and Advances in Data Mining*, Studies in Fuzziness and Soft Computing 180, pages 183-220, Springer-Verlag GmbH, 2005.
4. C. Giannella, J. Han, J. Pei, X. Yan, and P. S. Yu, **Mining Frequent Patterns in Data Streams at Multiple Time Granularities**, in H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (eds.), *Next Generation Data Mining*, AAAI/MIT, 2004.

7.8 Demonstrations

1. J. Liu, Q. Lin, J. Zhang, K. Ren, J. Lou, J. Liu, L. Xiong, J. Pei, and J. Sun. **Demonstration of Dealer: An End-to-End Model Marketplace with Differential Privacy.** In *Proceedings of the Forty-seventh International Conference on Very Large Data Bases (VLDB'21)*, Copenhagen, Denmark, August 16-20, 2021.
2. J. Liu, L. Xiong, J. Pei, J. Luo, H. Zhang, and S. Zhang. **SkyRec: Finding Pareto Optimal Groups.** In *Proceedings of the Twenty-eighth ACM International Conference on Information and Knowledge Management (CIKM'19)*, Beijing, China, November 3-7, 2019.
3. B. Zhou^[student] and J. Pei. **OSD: An Online Web Spam Detection.** In *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, Paris, France, June 28 - July 1, 2009. (Acceptance rate: 10/26)
4. T. Wang, B. Yang, J. Gao, D. Yang, S. Tang, H. Wu, K. Liu, and J. Pei. **MobileMiner: A Real World Case Study of Data Mining in Mobile Communication.** In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data (SIGMOD'09)*, June 29-July 2, 2009, Providence, Rhode Island, USA. (Acceptance rate: 32/86)
5. M. Hua^[student] and J. Pei. **DiMaC: A Disguised Missing Data Cleaning Tool.** In *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, August 24-27, 2008, Las Vegas, NV, USA.

6. M. Hua^[student] and J. Pei. **DiMaC: A System for Cleaning Disguised Missing Data**. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data (SIGMOD'08)*, June 9-12, 2008, Vancouver, Canada. (Acceptance rate: 30/94)
7. X. Zhou, J. Zhang, W. Wang, B. Shi, and J. Pei. **Using High Dimensional Indexes to Support Relevance Feedback Based Interactive Images Retrieval**. In *Proceedings of the Thirty-second International Conference on Very Large Data Bases (VLDB'06)*, Seoul, Korea, September 12-15, 2006.
8. W. Wang, C. Wang, Y. Zhu, B. Shi, J. Pei, X. Yan, and J. Han. **GraphMiner: A Structural Pattern Mining System for Large Disk-based Graph Databases and Its Applications**. In *Proceedings of the Twenty-fourth ACM SIGMOD International Conference on Management of Data (SIGMOD'05)*, Baltimore, Maryland, USA, June 14-16, 2005.
9. D. Jiang^[student], J. Pei and A. Zhang. **GPX: Interactive Mining of Gene Expression Data**. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases (VLDB'04)*, Toronto, ON, Canada, August 30-September 3, 2004. (Acceptance rate: 34/86)
10. L. V. S. Lakshmanan, J. Pei, and Y. Zhao. **Efficacious Data Cube Exploration by Semantic Summarization and Compression**. In *Proceedings of the Twenty-ninth International Conference on Very Large Data Bases (VLDB'03)*, Berlin, Germany, September 9-12, 2003. (Acceptance rate: 23/86)
11. L. V. S. Lakshmanan, J. Pei, and Y. Zhao. **SOCQET: Semantic OLAP with Compressed Cube and Summarization**. In *Proceedings of the 2003 ACM-SIGMOD International Conference on Management of Data (SIGMOD'03)*, San Diego, CA, June 9-12, 2003. (Acceptance rate: 22/73)
12. G. Dong, J. Han, J. Lam, J. Pei, J. Wang, K. Wang. **CubeExplorer: Online Exploration of Data Cubes**. In *Proceedings of 2002 ACM-SIGMOD International Conference on Management of Data (SIGMOD'02)*, Madison, Wisconsin, June 3-6, 2002.
13. T. Wang, S. Tang, D. Yang, J. Gao, Y. Wu, J. Pei. **COMMIX: Towards Effective Web Information Extraction, Integration and Query Answering**. In *Proceedings of 2002 ACM-SIGMOD International Conference on Management of Data (SIGMOD'02)*, Madison, Wisconsin, June 3-6, 2002.
14. J. Han, J. Pei, Y. Liao, and H. Jamil. **DNA-Miner: A System Prototype for Mining DNA Sequences**. In *Proceedings of the 2001 ACM-SIGMOD International Conference on Management of Data (SIGMOD'01)*, Santa Barbara, CA, May 2001.
15. J. Pei, R. Mao, K. Hu, and H. Zhu. **Towards Data Mining Benchmarking: A Testbed for Performance Study of Frequent Pattern Mining**. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD'00)*, Dallas, TX, May 2000.

7.9 Patents

- K. Tsoukalas^[student], J. Pei, and D. Cubranic, *Systems and Methods to Automatically Generate Enhanced Information Associated with a Selected Web Table*, U.S. Patent 20100198823, August 5, 2010.
- J. Han, J. Pei, Y. Yin, and R. Mao, *Methods and System for Mining Frequent Patterns*, U.S. Patent, No. 6665669 and 20030028531, December 16, 2003.

7.10 Edited Books, Proceedings, and Special Issues in Journals and Magazines

1. Q. Yang, D. Agarwal, and J. Pei (Eds.): *the Eighteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'12)*, Beijing, China, August 12-16, 2012. ACM 2012, ISBN 978-1-4503-1462-6.
2. D. J. Cook, J. Pei, W. Wang, O. R. Zaïane, and X. Wu (Eds.): *the Eleventh IEEE International Conference on Data Mining (ICDM'11)*, Vancouver, BC, Canada, December 11-14, 2011. IEEE 2011, ISBN 978-0-7695-4408-3.

3. M. Spiliopoulou, H. Wang, D. J. Cook, J. Pei, W. Wang, O. R. Zaïane, and X. Wu (Eds.): *the Eleventh IEEE International Conference on Data Mining Workshops (ICDMW'11)*, Vancouver, BC, Canada, December 11, 2011. ISBN 978-0-7695-4409-0.
4. X. Chen, T. S. Dillon, H. Ishbuchi, J. Pei, H. Wang, and M. Arif Wani (Eds.): *the Tenth IEEE International Conference on Machine Learning and Applications and Workshops (ICMLA'11)*, Honolulu, Hawaii, USA, December 18-21, 2011. Volume 1: Main Conference. IEEE Computer Society 2011.
5. X. Chen, T. S. Dillon, H. Ishbuchi, J. Pei, H. Wang, and M. Arif Wani (Eds.): *the Tenth IEEE International Conference on Machine Learning and Applications and Workshops (ICMLA'11)*, Honolulu, Hawaii, USA, December 18-21, 2011. Volume 2: Special Sessions and Workshop. IEEE Computer Society 2011.
6. H. T. Shen, J. Pei, M. T. Özsu, L. Zou, J. Lu, T. W. Ling, G. Yu, Y. Zhuang, and J. Shao (Eds.): *Web-Age Information Management - WAIM 2010 International Workshops: IWGD 2010, XMLDM 2010, WCMT 2010*, Jiuzhaigou Valley, China, July 15-17, 2010, Revised Selected Papers. Lecture Notes in Computer Science 6185, Springer 2010, ISBN 978-3-642-16719-5.
7. R. Huang, Q. Yang, J. Pei, J. Gama, X. Meng, and X. Li (Eds.): *Proceedings of the Fifth International Conference on Advanced Data Mining and Applications (ADMA'09)*, Beijing, China, August 17-19, 2009. Lecture Notes in Computer Science 5678, Springer 2009, ISBN 978-3-642-03347-6.
8. J. Pei, L. Getoor, and A. de Keijzer (Eds.): *Proceedings of the 1st ACM SIGKDD Workshop on Knowledge Discovery from Uncertain Data*, Paris, France, June 28, 2009. ACM 2009, ISBN 978-1-60558-675-5.
9. Q. Li, L. Feng, J. Pei, X. S. Wang, X. Zhou, and Q.-M. Zhu (Eds.): *Proceedings of the 2009 Joint International Conferences on Advances in Data and Web Management (APWeb/WAIM'09)*, Suzhou, China, April 2-4, 2009, Lecture Notes in Computer Science 5446, Springer 2009, ISBN 978-3-642-00671-5.
10. R. Ng and J. Pei. *Special Issue on Data Mining for Health Informatics. SIGKDD Explorations*, Volume 9, Issue 1, June, 2007.
11. A. S. Varde and J. Pei (Eds.): *Proceedings of the First Ph.D. Workshop in the Sixteenth ACM Conference on Information and Knowledge Management (PIKM'07)*, Lisbon, Portugal, November 9, 2007. ACM 2007.

8 Keynote Speeches in Conferences and Invited Talks (Since September 1, 2022)

1. **Data Valuation in Federated Learning**, Decision Sciences Seminar Series, Fuqua School of Business, Duke University, November 1, 2023.
2. **Data Valuation in Federated Learning**, 2023 Fall Lectures in Climate Data Science, Columbia University, October 5, 2023.
3. **Data Valuation in Federated Learning**, keynote speech at the International Workshop on Federated Learning for Distributed Data Mining, associated with the 29th ACM SIGKDD Conference (KDD'23), August 7, 2023, Long Beach, California, USA.
4. **Data and AI Markets for E-Commerce**, invited talk at KDD Cup 2023 Workshop, associated with the 29th ACM SIGKDD Conference (KDD'23), August 7, 2023, Long Beach, California, USA.
5. **Data and AI Model Markets: From Data Sharing, Discovery, and Integration to Values and Ecosystems**, Distinguished Seminar Series on Data Science & Artificial Intelligence, Department of Computing, the Hong Kong Polytechnic University, March 14, 2023.
6. **Data and AI Model Markets: From Data Sharing, Discovery, and Integration to Values and Ecosystems**, Charles L. and Ann Lee Brown Distinguished Seminar Series, Department of Electrical and Computer Engineering, University of Virginia, February 24, 2023.
7. **Exact, Concise, and Consistent Data Driven Interpretation**, keynote speech at the Fourth IEEE International Conference on Cognitive Machine Intelligence, December 15, 2022, virtual conference.

8. **Exact, Concise, and Consistent Data Driven Interpretation**, Department of Technology, Operations, and Statistics Seminar, Leonard N. Stern School of Business, New York University, September 23, 2022.

9 Training of Highly Qualified Personnel (HQP)

9.1 Current graduate students

Ph.D. students at Duke University (6)

Name	Department	Starting time
Bryce Bern	Computer Science	Fall 2023
Ja'Nya Breeden	Computer Science	Fall 2023
Zhuofan Jia	Computer Science	Fall 2023
Yi Yang	Computer Science	Fall 2023
Xuan Yang	Computer Science	Fall 2023
Minxing Zhang	Computer Science	Fall 2023

M.Sc. students at Duke University (1)

Name	Department	Starting time
Isabella Huanfei Wang	Electrical and Computer Engineering	July 2023

Undergraduate research projects at Duke University (2)

Name	Research project	Year
Michelle Si	Interpretation in data valuation	2022-2023
Chole Zhu	Sentiment analysis in recommender systems	2022

Ph.D. student committees at Duke University (7)

Name	Department	Advisor	Year
Zhi Chen	Computer Science	Cynthia Rudin	2023
Jonathan Donnelly	Computer Science	Cynthia Rudin	2023
Fangzhu Shen	Computer Science	Sudeepa Roy	2023
Meng Xia	Electrical and Computer Engineering	Yiran Chen	2023
Tunhou Zhang	Electrical and Computer Engineering	Yiran Chen	2023
Yujie Zhang	Electrical and Computer Engineering	Michael Reiter, Neil Gong	2023
Chudi Zhong	Computer Science	Cynthia Rudin	2023

Master's student committees at Duke University (4)

Name	Department	Advisor	Year
Xiaoyin Chen	Computer Science	Samuel Wiseman	2023
Rongze Gui	Computer Science	Bhuwan Dhingra	2023
Danyu Sun	Computer Science	Sudeepa Roy	2023
Greg Kotliar Szumel	Computer Science	Samuel Wiseman	2023

Ph.D. students at Simon Fraser University (3)

Name	Starting time
Xuan Luo	Spring 2021
Xinglu Wang	Fall 2021
Chirong Zhang	Fall 2020

9.2 Former graduate students

Master's students at Duke University (1)

Name	Graduation time	Thesis topic	Department
Xinran Song	April 2023	A Cost-sensitive, Semi-supervised, and Active Learning Approach for Priority Outlier Investigation	Statistical Science

Postdoctoral fellow (4)

Name	Time period	Affiliation immediately after graduation
Dong-Wan Choi	Spring 2015 - Fall 2016	Imperial College London
Lingyang Chu	Fall 2015 - April 2019	Huawei Canada
Kui Yu	Fall 2013 - Summer 2015	The University of Adelaide, Australia
Cheng Xu	Spring 2020 - Spring 2022	Fortinet

Ph.D. students at Simon Fraser University and the State University of New York at Buffalo (13)

Name	Graduation time	Thesis title	First job
Akiko Campbell	April 2016	Multidimensional Benchmarking	LiveLabs
Moonjung Cho	October 2006	Novel Techniques for Data Warehousing and Online Analytical Processing in Emerging Applications	Prism Health Networks and an adjunct professor at Bryant & Stratton College
Zicun Cong	October 2022	Towards Trustworthy Data Analytics: Algorithmic Tools for Interpretability and Fairness	Fortinet
Juhua Hu	December 2017	Subspace Clustering Methods for Understandable Information Organization	Tenure track faculty at University of Washington
Xia Hu	May 2021	Understanding Deep Neural Networks from the Perspective of Piecewise Linear Property	Postdoc
Ming Hua	November 2009	Ranking Queries on Uncertain Data	Research Scientist at Facebook
Bin Jiang	May 2011	Summarizing Certainty in Uncertain Data	Research Scientist at Facebook
Daxin Jiang	May 2005	Mining Coherent Patterns and Clusters from Genomic Data	Tenure track faculty at Nanyang Technology University, Singapore
Hossein Maserrat	August 2012	Compression of Social Networks	Software engineer in a start-up company

Guanting Tang	June 2015	New Methods in Outlier Detection	Entrepreneur
Zhengzheng Xing	December 2010	Early Classification and Feature Selection on Temporal Sequences	Applied Researcher in the Machine Learning Group, Amazon.com
Yu Yang	February 2019	Leveraging the Power of Social Propagations: Algorithm Designs for Social Marketing	Tenure track faculty at the City University of Hong Kong, Hong Kong, China
Bin Zhou	May 2011	Keyword Search on Large-Scale Data: From Relational and Graph Data to OLAP Infrastructure	Tenure track faculty at the University of Maryland at Baltimore County, MD, USA

Master's students at Simon Fraser University and the State University of New York at Buffalo (31)

Name	Graduation time	Thesis topic
Yvonne Ling-Chih Cheng	August 2011	A Tool for OLAP on Text Documents
Zicun Cong	August 2016	Mining Identification Rules for Classifying Mobile Application Traffic
Yi Cui	August 2011	Finding Email Correspondents in Social Networks
Chuancong Gao	August 2019	Data Integration on Complex Data
Da Huang	April 2012	Malicious URL Detection by Dynamically Mining Patterns without Pre-defined Elements
Qiang Jiang	August 2012	Multi-level Relationship Outlier Detection
Thusjanthan Kubendranathan	November 2010	Mining Multidimensional Distinct Patterns
Karlygash Kussainova	October 2023	An Empirical Data Mining Study on Non-fungible Token (NFT) Markets
Mag Man Ki Lau	October 2007	Energy-Preserving Maintenance of K -Centers on Large Wireless Sensor Networks
Luping Li	January 2012	Efficient And Effective Aggregate Keyword Search On Relational Databases
Jiaxing Liang	June 2015	Individual skyline subspace on user annotated tags
Zenhua Lin	June 2010	Mining Discriminative Items in Multiple Data Streams
Jian Liu	May 2005	Mining Frequent Closed Partial Orders from Strings
Ling Liu	June 2015	A new method for outlier detection on time series data
Beier Lu	June 2015	Health query expansion using WordNet and UMLS
Xiao Meng	October 2017	Query Processing of Schema Design Problems for Data-driven Renormalization
Brittany Nielsen	December 2009	Reverse Centrality Queries in Complex Networks
Irene Pekerskaya	May 2006	Mining Changing Regions from Access-Constrained Data Sets: A Cluster-Embedded Decision Tree Approach
Yu Tao	June 2014	Suspicious URL and Device Detection by Log Mining
Kathleen Tsoukalas	October 2009	Extracting and Tagging Tabular Data on the Web
Xiaoqian Wang	June 2016	Efficiently Compressing String Columnar Data Using Frequent Pattern Mining

Li Xiong	June 2015	Exploring the power of frequent neighborhood patterns on edge weight estimation
Ji Xu	October 2009	Hierarchical Distributed Classification in Wireless Sensor Networks
Xiaoning Xu	June 2014	Continuous Similarity Search for Evolving Queries
Xinghuo Zeng	October 2007	Efficient Maximal Frequent Itemset Mining by Pattern-Aware Dynamic Scheduling
Xiaoling Zhang	December 2003	Pattern-Based Clustering for DNA Micro-array Data Analysis
Yanyan Zhang	May 2019	Online Density Bursting Subgraph Detection from Temporal Graphs
Geoff Zenger	November 2009	Trajectory-Based Point of Interest Recommendation
Zijin Zhao	July 2017	Classification in the Presence of Heavy Label Noise: A Markov Chain Sampling Framework
Bin Zhou	May 2007	Mining Page Farms and Its Applications in Link Spam Detection
Yajie Zhou	May 2019	Improving Robustness of Neural Network with Adversarial Training

10 Professional Memberships and Activities

10.1 Membership of professional associations

- Fellow of the *Association for Computing Machinery (ACM)* for “contributions to the foundation, methodology and applications of data mining” (elevated on December 8, 2015).
- Fellow of the *Institute of Electrical and Electronics Engineers (IEEE)* for “contributions to data mining and knowledge discovery” (elevated on January 1, 2014).
- Member of the *Special Interest Group on Knowledge Discovery in Data (SIGKDD)* of the *Association for Computing Machinery (ACM)*
- Member of the *Computer Society of the Institute of Electrical and Electronics Engineers (IEEE)*

10.2 Membership on the editorial boards

- Editor-in-Chief (2013-2016), Associate Editor-in-Chief (2010-2012), Associate editor (2007-2011), *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, IEEE Press.
- Editorial board member, *Computational Intelligence: An International Journal*, Wiley Press, (2015-).
- Associate Editor, *IEEE Transactions on Big Data (TBD)*, IEEE Press, (2015-)
- Associate Editor, *ACM Transactions on Knowledge Discovery from Data (TKDD)*, ACM Press, (2009-)
- Associate Editor, *Knowledge and Information Systems: An International Journal (KAIS)*, Springer, (2011-)
- Editorial board member and action editor, *Data Mining and Knowledge Discovery: An International Journal (DMKD)*, Springer, (2011-)
- Steering Committee member and Associate Editor, *IEEE Transactions on Emerging Topics in Computing (TETC)*, IEEE Press (2013-)
- Editorial board member, *Journal of Intelligent Information Systems (JIIS)*, Springer, (2014-)
- Associate Editor, *Statistical Analysis and Data Mining, (SAM)*, Wiley, (2011-)

- Associate Editor, International Journal of Health Information Science and Systems, Springer, (2011-)
- Editorial board member, Intelligent Data Analysis: An International Journal (IDA), IOS Press, (2008-)
- Leading area editor (data management and mining, 2014-), subject area editor (machine learning and data mining, 2010-2013), Journal of Computer Science and Technology (JCST), Springer, (2010-)
- Member of the ECML/PKDD 2013-2015 Guest Editorial Board
- Member of the Review Board for the Proceedings of the VLDB Endowment (2008-2012, 2015)
- Member of the International Editorial Review Board, the International Journal of Data Warehousing and Mining (IJDWM), IGI Global Press, (2004-)
- Associate editor, ACM SIGMOD Digital Symposium Collection (DiSC) (2002-2006)

10.3 Guest-editorship

1. (Co-guest-editor with Drs. D. Agarwal, R. Caruana, and K. Wang) Special Issue on the Best Papers from the Eighteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'12). To appear in *ACM Transactions on Knowledge Discovery from Data*, 2013.
2. (Co-guest-editor with Drs. S. Abiteboul, C. Koch, and K-L. Tan) Special Issue on the Best Papers from the Twenty-Seventh International Conference on Data Engineering (ICDE'11). *IEEE Transactions on Knowledge and Data Engineering*, Volume 24, Number 10, October 2012.
3. (Co-guest-editor with Drs. Joao Gama, Qiang Yang, Ronghuai Huang, and Xue Li) Special Issue on the Best Papers from the Fifth International Conference on Advanced Data Mining and Applications (ADMA'09). *Knowledge and Information Systems*, Volume 27, Number 2, May 2011.
4. (Co-guest-editor with Dr. Raymond T. Ng) Special Issue on Data Mining for Health Informatics, *ACM SIGKDD Explorations*, Volume 9, Issue 1, June 2007.

10.4 Leadership in professional organizations, membership of advisory boards, steering committees, councils, and special committees

- Member, *2023 IEEE John Von Neumann Medal Committee* (October 1, 2022 - September 30, 2023)
- Member, *Processes and Selection Committees*, the Royal Society of Canada, Division of Applied Sciences and Engineering. (December 2021 - November 2023)
- Chair, *Scientific Advisory Board* of Aicadium Inc. (November 2021 -)
- Chair, Steering committee of *ACM International Conference on Web Search and Data Mining (WSDM)* (February 2020 - January 2022)
- ACM Special Interest Groups Governance Board (SGB) Executive Committee Member at Large (July 1, 2020 - June 30, 2022)
- Steering committee member of *ACM International Conference on Web Search and Data Mining (WSDM)* (July 2017 - March 2022)
- Scientist of *Creative Destruction Lab (CDL)*, a nonprofit organization that delivers an objectives-based program for massively scalable, seed-stage, science- and technology-based companies (September 2020 -)
- Steering committee member of *IEEE International Conference on Data Mining (ICDM)* (January 2016 -)
- Chair of *ACM SIGKDD* (July 2017-June 2021)
- Director of *ACM SIGKDD* (July 2013-June 2025)
- Member of the 2021 IEEE ICDM Awards Nomination Committee.
- Chair, *CCF-IEEE Distinguished Young Computer Scientist Award Committee* (April 2018 - March 2020)
- International Scientific Advisory Committee member of the *Data Science and AI Research (DSAIR) Center*, Nanyang Technological University, Singapore (January 2017 -)

- Advisory board member, Symphony Protocol, a next-generation blockchain-based protocol to empower a data-driven economy.
- Advisory board member, Aida Software Corporation, successfully acquired by Vision Critical Communications Inc. in 2017.
- Advisory board member, adSage Inc.
- Chair of the *ACM SIGKDD Data Science Education Committee* (August 2015-2016)
- Chair of the *Search Committee for the 2017 Editor-in-Chief of IEEE Intelligent Systems*.
- Steering committee member of the *IEEE Technical Committee on Scalable Computing (TCSC)* (2014-)
- Member of the *IEEE Computer Society Big Data Steering Committee* (2013-)
- Member of the *Engineering Panel of the Research Grants Council (RGC)*, Hong Kong (November 2014-October 2016)
- Member of the *steering committee of the Web-Age Information Management (WAIM) conferences* (2009-2015)
- Member of the *steering committee of the IEEE International Conferences on Big Data and Distributed Systems (BDDS)* (2013-)
- Member of the *steering committee of the Pacific Asia Knowledge Discovery and Data Mining (PAKDD) conferences* (2013-)
- Member of the *steering committee of the M.Sc. Program in Computing and Data Analytics*, Saint Mary's University, NS, Canada (2015-)
- Member of the *Academic Advisory Board*, Faculty of Computer Science and Technology, Shandong University, China (2013-)
- Member of the *steering committee of the 2011 Workshop for Ph.D. Students in Information and Knowledge Management (PIKM 2011)*, in conjunction with the Twentieth ACM Conference on Information and Knowledge Management (CIKM 2011), October 24-28, 2011, Glasgow, UK.
- Advisor, project “An Epidemiology of Information: Data Mining the 1918 Influenza Pandemic” of the Digging into Data Challenge (an international grant program funded by eight leading research agencies from the United States, the United Kingdom, Canada, and the Netherlands), 2012-2013.

10.5 Members of organizing committees

1. **Associate editor**, *Proceedings of the VLDB Endowment (PVLDB)*, Volume 17, the Fiftieth International Conference on Very Large Data Bases (VLDB'24), Guangzhou, China.
2. **Area chair**, the Fortieth IEEE International Conference on Data Engineering (ICDE'24), April 16-19, 2024, Utrecht, Netherlands.
3. **Area Chair**, the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI'24), February 20-27, 2024, Vancouver, British Columbia, Canada.
4. **Area chair**, the Twenty-third IEEE International Conference on Data Mining (ICDM'23), December 1-4, 2023, Shanghai, China.
5. **Senior program committee member** of the Thirty-second ACM International Conference on Information and Knowledge Management (CIKM'23), October 21-25, 2023, Birmingham, UK.
6. **Area chair**, the Thirty-fourth European Conference on Machine Learning (ECML'23) and the Twenty-seventh European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'23), September 18-22, 2023, Turin, Italy.
7. **Area chair**, the Twenty-ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'23), August 6-10, 2023, Long Beach, CA, USA.

8. **Senior program committee member** of the Applied Science Track, *the Twenty-ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'23)*, August 6-10, 2023, Long Beach, CA, USA.
9. **Senior program committee member**, *the Twenty-seventh Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'23)*, May 25-28, 2023, Osaka, Japan.
10. **Senior program committee member**, *the Twenty-third SIAM International Conference on Data Mining (SDM'23)*, April 27-29, 2023, Minneapolis, MN, USA.
11. **Associate chair of the technical research program**, *the Thirty-ninth IEEE International Conference on Data Engineering (ICDE'23)*, April 7-10, 2023, Los Angeles, CA, USA.
12. **Area chair**, *the Thirty-seventh AAAI Conference on Artificial Intelligence (AAAI'23)*, February 7-14, 2023, Washington DC, USA.
13. **Area chair**, *the Twenty-second IEEE International Conference on Data Mining (ICDM'22)*, November 30 - December 3, 2022, Orlando, FL, USA.
14. **Senior program committee member**, *the Thirty-first ACM International Conference on Information and Knowledge Management (CIKM'22)*, October 17-22, Atlanta, GA, USA.
15. **Associate editor, the Scalable Data Science Track**, *The Forty-eighth International Conference on Very Large Databases (VLDB'22)*, September 5-9, 2022, Sydney, Australia.
16. **Best paper award committee chair**, *the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, August 14-18, 2022, Washington DC, USA.
17. **Senior program committee member**, the Research track, *the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, August 14-18, Washington DC, USA.
18. **Senior program committee member**, the Applied Data Science track, *the Twenty-eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'22)*, August 14-18, Washington DC, USA.
19. **Senior program committee member**, *the Thirty-first International Joint Conference on Artificial Intelligence (IJCAI'22)*, July 23-29, 2022, Vienna, Austria.
20. **Associate editor**, *the 2022 ACM SIGMOD International Conference on Management of Data (SIGMOD'22)*, June 12-17, Philadelphia, PA, USA.
21. **Senior program committee member**, *the Twenty-second SIAM International Conference on Data Mining (SDM'22)*, April 28-30, 2022, Alexandria, VA, USA.
22. **Area chair**, *the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)*, February 22-March 1, 2022, Vancouver, BC, Canada.
23. **Member of the award committee**, *the Twenty-first IEEE International Conference on Data Mining (ICDM'21)*, November 7-10, 2021, Auckland, New Zealand.
24. **Area chair**, *the Twenty-first IEEE International Conference on Data Mining (ICDM'21)*, November 7-10, 2021, Auckland, New Zealand.
25. **Senior program committee member**, *the Thirtieth ACM International Conference on Information and Knowledge Management (CIKM'21)*, November 1-5, 2021, Gold Coast, Queensland, Australia.
26. **Senior program committee member**, *the Twenty-seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'21)*, August 14-18, Singapore.
27. **Area chair**, *the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI'21)*, August 19-26, 2021, Montreal, QC, Canada.
28. **Senior program committee member**, *the Twenty-fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'21)*, May 11-14, 2021, New Delhi, India.
29. **Award committee co-chair**, *the Twenty-first SIAM International Conference on Data Mining (SDM'21)*, April 29 - May 1, 2021, Alexandria, VA, USA.

30. **Senior program committee member**, *the Twenty-first SIAM International Conference on Data Mining (SDM'21)*, April 29 - May 1, 2021, Alexandria, VA, USA.
31. **Senior program committee member**, *the Fourteenth ACM International Conference on Web Search and Data Mining (WSDM'21)*, March 2021, Jerusalem, Israel.
32. **Area chair**, *the Thirty-fifth AAAI Conference on Artificial Intelligence (AAAI'21)*, February 2-9, 2021, virtual.
33. **Senior program committee member**, *the Twenty-sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'20)*, August 22-27, San Diego, CA, USA.
34. **Senior program committee member**, *the Twenty-fourth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'20)*, May 11-14, 2020, Singapore.
35. **Senior program committee member**, *the Twentieth SIAM International Conference on Data Mining (SDM'20)*, May 7-9, 2020, Cincinnati, OH, USA.
36. **Area chair**, *the Nineteenth IEEE International Conference on Data Mining (ICDM'19)*, November 8-11, 2019, Beijing, China.
37. **Panel co-chair**, *the Nineteenth IEEE International Conference on Data Mining (ICDM'19)*, November 8-11, 2019, Beijing, China.
38. **Senior program committee member**, *the Twenty-eighth ACM International Conference on Information and Knowledge Management (CIKM'19)*, November 3-7, 2019, Beijing, China.
39. **Senior program committee member**, *the Twenty-fifth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'19)*, August 3-7, 2019, Anchorage, AK, USA.
40. **Senior program committee member**, *the Nineteenth SIAM International Conference on Data Mining (SDM'19)*, May 2-4, 2019, Calgary, Alberta, Canada.
41. **Senior program committee member**, *the Twenty-third Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'19)*, April 17-19, 2019, Macau, China.
42. **Area chair**, *the Eighteenth IEEE International Conference on Data Mining (ICDM'18)*, November 17-20, 2018, Singapore.
43. **Program committee co-chair**, *the Forty-fourth International Conference on Very Large Data Bases (VLDB'18)*, August 27-31, 2018, Rio, Brasil.
44. **Senior program committee member**, *the Twenty-fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18)*, August 19-23, 2018, London, United Kingdom.
45. **Senior program committee member**, *the Twenty-second Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'18)*, June 3-6, 2018, Melbourne, Australia.
46. **Program committee co-chair**, *the Twenty-third International Conference on Database Systems for Advanced Applications (DASFAA'18)*, May 21-24, 2018, Gold Coast, Australia.
47. **Award committee member**, *the Thirty-fourth IEEE International Conference on Data Engineering (ICDE'18)*, April 16-19, 2018, Paris, France.
48. **Senior program committee member**, *the Eleventh ACM International Conference on Web Search and Data Mining (WSDM'18)*, February 5-9, 2018, Los Angeles, CA, USA.
49. **Senior program committee member**, *the Fifth IEEE International Conference on Big Data (Big-Data'17)*, December 11-14, 2017, Boston, MA, USA.
50. **Area chair**, *the Seventeenth IEEE International Conference on Data Mining (ICDM'17)*, November 18-21, 2017, New Orleans, LA, USA.
51. **Senior program committee member**, *the Twenty-sixth ACM International Conference on Information and Knowledge Management (CIKM'17)*, November 6-10, 2017, Singapore.
52. **Area chair**, *the Twenty-ninth IEEE International Conference on Tools with Artificial Intelligence (IC-TAI'17)*, November 6-8, 2017, Boston, MA, USA.

53. **Senior program committee member**, *the Seventeenth SIAM International Conference on Data Mining (SDM'17)*, April 27-29, 2017, Huston, TX, USA.
54. **Best paper award committee member**, *the Twenty-sixth World Wide Web Conference (WWW'17)*, April 3-7, 2017, Perth, Australia.
55. **Research track co-chair**, "Web Mining and Content Analysis", *the Twenty-sixth World Wide Web Conference (WWW'17)*, April 3-7, 2017, Perth, Australia.
56. **Best paper and Test of Time Award Committee Chair**, *the Sixteenth IEEE International Conference on Data Mining (ICDM'16)*, December 12-15, 2016, Barcelona, Spain.
57. **Tutorial co-chair**, *the Sixteenth IEEE International Conference on Data Mining (ICDM'16)*, December 12-15, 2016, Barcelona, Spain.
58. **Award committee chair**, *the Twenty-second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, August 13-17, 2016, San Francisco, CA, USA.
59. **Senior program committee member**, *the Twenty-second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, August 13-17, 2016, San Francisco, CA, USA.
60. **Senior program committee member**, *the Sixteenth International Conference on Web Engineering (ICWE'16)*, June 6-9, 2016, Lugano, Switzerland.
61. **Senior program committee member**, *the Twentieth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'16)*, April 19-22, 2016, Auckland, New Zealand.
62. **Panel co-chair**, *the Fifteenth IEEE International Conference on Data Mining (ICDM'15)*, November 14-17, 2015, Atlantic City, NJ, USA.
63. **Tutorial co-chair**, *the Twenty-first ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*, August 10-13, 2015, Sydney, Australia.
64. **Senior program committee member**, *the Twenty-first ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*, August 10-13, 2015, Sydney, Australia.
65. **Program committee co-chair**, *the 2015 IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM'15)*, August 2015, Wroclaw, Poland.
66. **Senior program committee member**, *the Nineteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'15)*, May 19-22, 2015, Ho Chi Minh City, Viet Nam.
67. **Senior program committee member**, *the Fifteenth SIAM International Conference on Data Mining (SDM'15)*, April 30-May 2, 2015, Vancouver, BC, Canada.
68. **Senior program committee member**, *the Twentieth International Conference on Database Systems for Advanced Applications (DASFAA'15)*, April 20-23, 2015, Hanoi, Vietnam.
69. **Senior program committee member**, *the Eighth ACM International Conference on Web Search and Data Mining (WSDM'15)*, February 2-6, 2015, Shanghai, China.
70. **General co-chair**, *the Fourteenth IEEE International Conference on Data Mining (ICDM'14)*, December 14-17, 2014, Shenzhen, Guangdong, China.
71. **Program committee co-chair**, *the Second IEEE International Conference on Big Data (IEEE Big-Data'14)*, October 27-30, Washington DC, USA.
72. **Tutorial co-chair and Senior program committee member**, *the Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*, August 24-27, 2014, New York, NY, USA.
73. **The Industry and government track invited talk advisory committee**, *the Twentieth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14)*, August 24-27, 2014, New York, NY, USA.
74. **Senior program committee member**, *the Eighteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'14)*, May 13-16, 2014, Tainan, Taiwan.

75. **Senior program committee member**, *the Fourteenth SIAM International Conference on Data Mining (SDM'14)*, April 24-16, 2014, Philadelphia, PA, USA.
76. **Program committee vice chair**, *the Thirteenth IEEE International Conference on Data Mining (ICDM'13)*, December 8-11, 2013, Dallas, TX, USA.
77. **Program committee co-chair**, *the First IEEE ICDM Workshop on Causal Discovery (CD'13)*, December 8, 2013, Dallas, TX, USA.
78. **Area chair**, *the Twenty-fifth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'13)*, November 4-6, 2013, Washington DC, USA.
79. **General co-chair**, *the Second IEEE International Conference on Big Data and Distributed Systems (BDDS'13)*, December 3-5, 2013, Sydney, Australia.
80. **Program committee co-chair**, *the Twenty-second ACM International Conference on Information and Knowledge Management (CIKM'13)*, October 28-November 1, 2013, San Francisco, CA, USA.
81. **Workshop program co-chair**, *the Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13)*, Aug 11-14, 2013, Chicago, IL, USA.
82. **Senior program committee member**, *the Nineteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13)*, Aug 11-14, 2013, Chicago, IL, USA.
83. **Senior program committee member**, *the Thirteenth SIAM International Conference on Data Mining (SDM'13)*, May 2-4, 2013, Austin, TX, USA.
84. **Program committee co-chair**, *the Seventeenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'13)*, April 13-17, 2013, Gold Coast, Queensland, Australia.
85. **Program committee track chair (data mining)**, *the Twenty-ninth IEEE International Conference on Data Engineering (ICDE'13)*, April 8-11, 2013, Brisbane, Queensland, Australia.
86. **Program committee vice chair**, *the Twelfth IEEE International Conference on Data Mining (ICDM'12)*, December 10-13, 2012, Brussels, Belgium.
87. **Area chair**, *the Twenty-third European Conference on Machine Learning (ECML'12) and the Sixteenth European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'12)*, September 24-28, 2012, Bristol, United Kingdom.
88. **Program committee co-chair**, *the Eighteenth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'12)*, August 12-16, 2012, Beijing, China.
89. **Senior program committee member**, *the Sixteenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'12)*, May 29-June 1, 2012, Kuala Lumpur, Malaysia.
90. **Program committee co-chair**, *the Tenth IEEE International Conference on Machine Learning and Applications (ICMLA'11)*, December 18-21, 2011, Honolulu, Hawaii, USA.
91. **Program committee co-chair**, *the Eleventh IEEE International Conference on Data Mining (ICDM'11)*, December 11-14, 2011, Vancouver, Canada.
92. **Senior program committee member**, *the Twenty-third IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2011)*, November 7-9, 2011, Boca Raton, FL, USA.
93. **Area chair**, *the Twenty-second European Conference on Machine Learning (ECML'11) and the Fifteenth European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'11)*, September 5-9, Athens, Greece.
94. **Member of the best paper selection committee**, *the Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, Aug 21-24, 2011, San Diego, CA, USA.
95. **Senior program committee member**, *the Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, Aug 21-24, 2011, USA.
96. **Program committee co-chair**, *Data Mining Track, the Twenty-sixth ACM Symposium On Applied Computing (SAC'11)*, March 21-25, 2011, Taichung, Taiwan, China.

97. **Senior member**, *the Fourth ACM International Conference on Web Search and Data Mining (WSDM'11)*, February 9-12, 2011, Hong Kong, China.
98. **Président d'honneur (Honorary President)**, *11^{ème} Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances (the Eleventh Francophone International Conference on Mining and Knowledge Management, EGC'11)*, January 25-28, 2011, Brest, France.
99. **Senior program committee member**, *the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10)*, July 25-28, 2010, Washington, DC, USA.
100. **Program committee co-chair**, *the Tenth SIAM International Conference on Data Mining (SDM'10)*, April 22-May 1, 2010, Columbus, Ohio, USA.
101. **Program committee vice-chair**, *the Ninth IEEE International Conference on Data Mining (ICDM'09)*, December 6-9, 2009, Miami, Florida, USA.
102. **Knowledge Management Track Program committee vice-chair**, *the Eighteenth ACM Conference on Information and Knowledge Management (CIKM'09)*, November 2-6, 2009, Hong Kong, China.
103. **Program committee co-chair**, *the Fifth International Conference on Advanced Data Mining and Applications (ADMA'09)*, August 17-19, 2009, Beijing, China.
104. **Organization committee co-chair**, *the First ACM SIGKDD Workshop on Knowledge Discovery from Uncertain Data (U'09)*, in conjunction with the *Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, June 28, 2009, Paris, France.
105. **Program committee co-chair**, *the Joint International Conferences on Asia-Pacific Web Conference (APWeb) and Web-Age Information Management (WAIM)*, April 6-8, 2009, Suzhou, China.
106. **Workshops co-chair**, *the Twenty-fifth IEEE International Conference on Data Engineering (ICDE'09)*, March 29-April 4, 2009, Shanghai, China.
107. **Program committee vice-chair**, *the Eighth IEEE International Conference on Data Mining (ICDM'08)*, December 15-19, 2008, Pisa, Italy.
108. **Area chair**, *the Nineteenth European Conference on Machine Learning (ECML'08) and the twelfth European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'08)*, September 15-19, 2008, Antwerp, Belgium.
109. **Program committee co-chair**, *the First International Workshop on Multidimensional Data Mining (MMD'08)*, September 19, 2008, Antwerp, Belgium. In conjunction with the *Nineteenth European Conference on Machine Learning (ECML'08) and the Twelfth European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'08)*, September 15-19, 2008, Antwerp, Belgium.
110. **Publicity chair**, *the Fourteenth International Conference on Knowledge Discovery and Data Mining (KDD'08)*, August 24-27, 2008, Las Vegas, NV, USA.
111. **Finance chair**, *the 2008 ACM SIGMOD International Conference on Management of Data and the 2008 International Conference on Principles of Database Systems (SIGMOD/PODS'08)*, June 11-14, 2008, Vancouver, Canada.
112. **Area chair** (area: frequent pattern mining), *the Eighth SIAM International Conference on Data Mining (SDM'08)*, April 24-26, 2008, Atlanta, Georgia, USA.
113. **Area chair**, *the Twelfth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'08)*, May 20-23, 2008, Osaka, Japan.
114. **Tutorial chair**, *the Seventh IEEE International Conference on Data Mining (ICDM'07)*, October 28-31, 2007, Omaha, NE, USA.
115. **Program committee co-chair**, *the First Ph.D. Workshop in CIKM (PIKM'07)*, November 9, 2007, Lisbon, Portugal. In conjunction with the *ACM Sixteenth Conference on Information and Knowledge Management (CIKM'07)*, November 6-8, 2007, Lisbon, Portugal.
116. **Senior member**, *the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*, August 12-15, 2007, San Jose, California, USA.

117. **Tutorial chair**, *the Joint Conference of the Ninth Asia Pacific Web Conference and the Eighth International Conference on Web-Age Information Management (APWeb/WAIM'07)*, June 11-13, 2007, Huangshan, China.
118. **Area chair**, *the Eleventh Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'07)*, May 22-25, 2007, Nanjing, China.
119. **Program committee vice-chair** (area: mining frequent patterns), *the Seventh SIAM International Conference on Data Mining (SDM'07)*, April 26-28, 2007, Minneapolis, Minnesota, USA.
120. **Program committee co-chair**, *the International Workshop on Web-based Massive Data Processing (WMDP'06)*, October 22, 2006, Wuhan, China. In conjunction with the Seventh International Conference on Web Information Systems Engineering (WISE'06).
121. **Publicity chair**, *the Ninth Pacific Rim International Conferences on Artificial Intelligence (PRICAI'06)*, August 7-11, 2006, Guilin, China.
122. **Program committee vice-chair**, *the Fifth IEEE International Conference on Data Mining (ICDM'05)*, November 26-30, 2005, New Orleans, Louisiana, USA.
123. **Industrial co-chair**, *the Twenty-third International Conference on Conceptual Modeling (ER'04)*, November 8-12, 2004, Shanghai, China.
124. **Program committee co-chair**, *the First International Workshop on Conceptual Model-Directed Web Information Integration and Mining (CoMWIM'04)*, November 8, 2004, Shanghai, China. In conjunction with the Twenty-third International Conference on Conceptual Modeling (ER'04).

10.6 Member of international conference program committees

1. Data Engineering Future Technologies (DEFT) track, *the Fortieth IEEE International Conference on Data Engineering (ICDE'24)*, April 16-19, 2024, Utrecht, Netherlands.
2. *The 2023 ACM SIGMOD International Conference on Management of Data (SIGMOD'23)*, June 2023, Seattle, WA, USA.
3. *The Fifteenth ACM International Conference on Web Search and Data Mining (WSDM'23)*, February 27 - March 3, 2023, Singapore.
4. *The Forty-eighth International Conference on Very Large Databases (VLDB'22)*, September 5-9, 2022, Sydney, Australia.
5. *The Forty-fifth International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'22)*, July 11-15, 2022, Madrid, Spain.
6. *The Thirty-eighth IEEE International Conference on Data Engineering (ICDE'22)*, May 9-12, 2022, Kuala Lumpur, Malaysia.
7. *The Thirty-first Web Conference (WWW'22)*, April 25-29, 2022, Lyon, France.
8. *The Forty-fourth International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'21)*, July 11-15, 2021, Montreal, Quebec, Canada.
9. *The Thirty-seven IEEE International Conference on Data Engineering (ICDE'21)*, April 19-23, 2021, Chania, Crete, Greece.
10. *The Thirtieth Web Conference (WWW'21)*, April 19-23, 2021, Ljubljana, Slovenia.
11. *The Forty-third International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20)*, July 25-30, 2020, Xi'an, China.
12. *The 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD'20)*, June 14-19, 2020, Portland, OR, USA.
13. *The Thirty-third Canadian Conference on Artificial Intelligence (AI'20)*, May 12-15, 2020, Ottawa, ON, Canada.

14. *The Thirty-sixth IEEE International Conference on Data Engineering (ICDE'20)*, April 20-24, 2020, Dallas, TX, USA.
15. *The Forty-second International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*, July 21-25, 2019, Paris, France.
16. *The Thirty-fifth IEEE International Conference on Data Engineering (ICDE'19)*, April 8-12, 2019, Macau, China.
17. *The Twenty-second International Conference on Extending Database Technology (EDBT'19)*, March 26-29, 2019, Lisbon, Portugal.
18. *The Forty-first International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'18)*, July 8-12, 2018, Ann Arbor, MI, USA.
19. *The Twenty-seventh International World Wide Web Conference (WWW'18)*, April 23-27, 2018, Lyon, France.
20. *The Twenty-third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*, August 13-17, Halifax, Nova Scotia, Canada.
21. *The Fortieth International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'17)*, August 7-11, 2017, Tokyo, Japan.
22. *The Twentieth International Conference on Extending Database Technology (EDBT'17)*, March 21-24, 2017, Venice, Italy.
23. *The Tenth ACM International Conference on Web Search and Data Mining (WSDM'17)*, February 6-10, 2017, Cambridge, UK.
24. *The Sixteenth IEEE International Conference on Data Mining (ICDM'16)*, December 12-15, 2016, Barcelona, Spain.
25. *The Fourth IEEE International Conference on Big Data (BigData'16)*, November 15-18, 2016, Washington D.C., USA.
26. *The Twenty-first Ibero American Congress on Pattern Recognition (CIARP'16)*, November 8-11, 2016, Lima, Perú.
27. *The Twenty-sixth European Conference on Machine Learning and the Seventeenth European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'16)*, September 19-23, 2016, Riva Del Garda, Italy.
28. *The Forty-second International Conference on Very Large Data Bases (VLDB'16)*, September 5-9, 2016, New Delhi, India.
29. *The Eighth IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM'16)*, August 18-21, 2016, Davis, CA, USA.
30. *The Thirty-ninth ACM International Conference on Research and Development in Information Retrieval (SIGIR'16)*, July 17-21, 2016, Pisa, Italy.
31. *The Twenty-fifth International Joint Conference on Artificial Intelligence (IJCAI'16) Special Track on AI and the Web*, July 9-15, 2016, New York, NY, USA.
32. *The 2016 ACM SIGMOD International Conference on Management of Data (SIGMOD'16)*, June 26-July 1, 2016, San Francisco, CA, USA.
33. *The Twenty-ninth Canadian Conference on Artificial Intelligence (AI'16)*, May 31-June 3, 2016, Victoria, BC, Canada.
34. *The Thirty-second IEEE International Conference on Data Engineering (ICDE'16)*, May 16-20, 2016, Helsinki, Finland.
35. *The Twenty-first International Conference on Database Systems for Advanced Applications (DASFAA'16)*, April 16-19, 2016, Dallas, TX, USA.
36. *The Twenty-fifth International World Wide Web Conference (WWW'16)*, April 11-15, 2016, Montreal, QB, Canada.

37. *The Ninth ACM International Conference on Web Search and Data Mining (WSDM'16)*, February 2016, San Francisco, CA, USA.
38. *The 2015 Asia Information Retrieval Societies Conference (AIRS'15)*, December 2-4, Brisbane, Queensland, Australia.
39. *The Fifteenth IEEE International Conference on Data Mining (ICDM'15)*, November 14-17, Atlantic City, NJ, USA.
40. *The Twenty-fifth European Conference on Machine Learning and the Sixteenth European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'15)*, September 7-11, 2015, Porto, Portugal.
41. *The Forty-first International Conference on Very Large Data Bases (VLDB'15)*, August 31-September 4, 2015, Kohala Coast, HA, USA.
42. *The Thirty-eighth Annual ACM SIGIR Conference (SIGIR'15)*, August 9-13, 2015, Santiago, Chile.
43. *The Twenty-eighth Canadian Conference on Artificial Intelligence (AI'15)*, June 2-5, 2015, Halifax, Nova Scotia, Canada.
44. *The Twenty-fourth International World Wide Web Conference (WWW'15)*, May 18-22, 2015, Florence, Italy.
45. *The Thirty-first IEEE International Conference on Data Engineering (ICDE'15)*, April 13-16, 2015, Seoul, Korea.
46. *The Fourteenth IEEE International Conference on Data Mining (ICDM'14)*, December 14-17, Shenzhen, Guangdong, China.
47. *The Twenty-fourth European Conference on Machine Learning and the Sixteenth European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'14)*, September 15-19, 2014, Nancy, France.
48. *The 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'14)*, August 17-20, 2014, Beijing, China.
49. *The Thirty-seventh ACM International Conference on Research and Development in Information Retrieval (SIGIR'14)*, July 6-11, 2014, Gold Coast, Australia.
50. *The 2014 SIGMOD International Conference on Management of Data (SIGMOD'14)*, June 22-27, 2014, Snowbird, UT, USA.
51. *The Twenty-seventh Canadian Conference on Artificial Intelligence (AI'14)*, May 6-9, 2014, Montreal, Quebec, Canada.
52. *The Nineteenth International Conference on Database Systems for Advanced Applications (DASFAA'14)*, April 21-24, 2014, Bali, Indonesia.
53. *The Thirtieth IEEE International Conference on Data Engineering (ICDE'14)*, April 7-14, 2014, Chicago, IL, USA.
54. *The Twenty-third International World Wide Web Conference (WWW'14)*, the Search Systems and Applications track and the Web Mining track, April 7-11, 2014, Seoul, Korea.
55. *The Seventh ACM International Conference on Web Search and Data Mining (WSDM'14)*, February 24-28, 2014, New York, NY, USA.
56. *The 2013 IEEE International Conference on Big Data (IEEE Big Data 2013)*, October 6-9, 2013, Silicon Valley, CA, USA.
57. *The Twenty-third European Conference on Machine Learning and the Sixteenth European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'13)*, September 23-27, 2013, Prague, Czech Republic.
58. *The 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13)*, August 25-28, 2013, Niagara Falls, Ontario, Canada.

59. *The Thirty-sixth ACM International Conference on Research and Development in Information Retrieval (SIGIR'13)*, July 28-August 1, 2013, Dublin, Ireland.
60. *The Fourteenth International Conference on Web-Age Information Management (WAIM'13)*, June 14-16, 2013, Beidaihe, Hebei, China.
61. *The Twenty-second International World Wide Web Conference (WWW'13), the Search Systems and Applications track, the Web Mining track, and the Content Analysis track*, May 13-17, 2013, Rio de Janeiro, Brazil.
62. *The Sixth International Conference on Web Search and Data Mining (WSDM'13)*, February 5-8, 2013, Rome, Italy.
63. *The Twelfth IEEE International Conference on Data Mining (ICDM'12)*, December 10-13, 2012, Brussels, Belgium.
64. *The Thirty-eighth International Conference on Very Large Data Bases (VLDB'12)*, August 27-31, 2012, Istanbul, Turkey.
65. *The Thirty-fifth ACM International Conference on Research and Development in Information Retrieval (SIGIR'12)*, August 12-16, 2012, Portland, Oregon, USA.
66. *The Twelfth SIAM International Conference on Data Mining (SDM'12)*, April 26-28, 2012, Anaheim, CA, USA.
67. *The Twenty-first International World Wide Web Conference (WWW'12), the Web Mining track and the social networks track*, April 16-20, 2012, Paris, France.
68. *The Seventeenth International Conference on Database Systems for Advanced Applications (DASFAA'12)*, April 15-18, 2012, Busan, South Korea.
69. *The Twenty-eighth IEEE International Conference on Data Engineering (ICDE'12)*, April 1-5, 2012, Washington DC, USA.
70. *The Fifth ACM International Conference on Web Search and Data Mining (WSDM'12)*, February 8-12, 2012, Seattle, WA, USA.
71. *The Thirty-seventh International Conference on Very Large Data Bases (VLDB'11)*, August 2011, Seattle, WA, USA.
72. *The Twenty-seventh IEEE International Conference on Data Engineering (ICDE'11)*, April 11-16, 2011, Hannover, Germany.
73. *The Nineteenth International World Wide Web Conference (WWW'11), the Social Systems and Graph Analysis track*, March 28 - April 1, 2011, Hyderabad, India.
74. *The IEEE Twenty-five International Conference on Advanced Information Networking and Applications (AINA'11)*, March 22-25, 2011, Singapore.
75. *The Tenth IEEE International Conference on Data Mining (ICDM'10)*, December 13-17, 2010, Sydney, Australia.
76. *The Twentieth European Conference on Machine Learning and the Thirteenth European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'10)*, September 20-24, 2010, Barcelona, Catalonia, Spain.
77. *The 2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM'10)*, August 9-11, 2010, Odense, Denmark.
78. *The Eighth Annual Conference on Privacy, Security and Trust (PST'10)*, August 17-19, 2010, Ottawa, ON, Canada.
79. *The 2010 ACM SIGMOD International Conference on Management of Data (SIGMOD'10)*, June 6-11, 2010, USA.
80. *The Eighteenth International World Wide Web Conference (WWW'10) Data Mining Track*, April 26-30, Raleigh, NC, USA.

81. *The IEEE Twenty-fourth International Conference on Advanced Information Networking and Applications (AINA'10)*, April 20-23, 2010, Perth, Australia.
82. *The Twenty-sixth IEEE International Conference on Data Engineering (ICDE'10)*, March 1-6, 2010, Long Beach, California, USA.
83. *The Third ACM International Conference on Web Search and Data Mining (WSDM'10)*, February 4-6, 2010, New York, NY, USA.
84. *The Eleventh International Conference on Data Warehousing and Knowledge Discovery (DaWaK'09)*, August 31-September 4, 2009, Linz, Austria.
85. *The Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)*, June 28-July 1, 2009, Paris, France.
86. *The IEEE Twenty-third International Conference on Advanced Information Networking and Applications (AINA'09)*, May 26-29, 2009, United Kingdom.
87. *The Ninth SIAM International Conference on Data Mining (SDM'09)*, April 30-May 2, 2009, Sparks, Nevada, USA.
88. *The Fourteenth International Conference on Database Systems for Advanced Applications (DASFAA'09)*, April 21-23, 2009, Brisbane, Australia.
89. *The Seventeenth International World Wide Web Conference (WWW'09) Data Mining Track*, April 20-24, Madrid, Spain.
90. *The Twenty-fifth IEEE International Conference on Data Engineering (ICDE'09)*, March 29-April 4, 2009, Shanghai, China.
91. *The Twelfth International Conference on Extending Database Technology (EDBT'09)*, March 23-26, 2009, Saint-Petersburg, Russia.
92. *The Second ACM International Conference on Web Search and Data Mining (WSDM'09)*, February 9-11, 2009, Barcelona, Spain.
93. *The ACM Seventeenth Conference on Information and Knowledge Management (CIKM'08)*, October 26-30, 2008, Napa Valley, California, USA.
94. *The Twenty-seventh International Conference on Conceptual Modeling (ER'08) (demo program)*, October 20-23, Barcelona, Spain.
95. *The Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, August 24-27, 2008, Las Vegas, NV, USA.
96. *The Ninth International Conference on Web-Age Information Management (WAIM'08)*, July 20-22, 2008, Zhangjiajie, Hunan, China.
97. *The Tenth Asia Pacific Web Conference (APWeb'08)*, April 26-28, Shenyang, China.
98. *The Sixteenth International World Wide Web Conference (WWW'08) Data Mining Track*, April 21-25, Beijing, China.
99. *The Twenty-fourth IEEE International Conference on Data Engineering (ICDE'08)*, April 7-12, 2008, Cancun, Mexico.
100. *2008 ACM Symposium of Applied Computing (SAC'08), Special Track on Data Mining*, March 16-20, Fortaleza, Brazil.
101. *The ACM Sixteenth Conference on Information and Knowledge Management (CIKM'07)*, November 6-8, 2007, Lisboa, Portugal.
102. *2007 IEEE International Conference on Bioinformatics and Biomedicine*, November 2-4, 2007, Silicon Valley, CA, USA.
103. *The Sixteenth International World Wide Web Conference (WWW'07) Data Mining Track*, May 8-12, 2007, Banff, Alberta, Canada.

104. *The Twenty-third IEEE International Conference on Data Engineering (ICDE'07)*, April 16-20, 2007, Istanbul, Turkey.
105. *The Sixth IEEE International Conference on Data Mining (ICDM'06)*, December 18-22, 2006, Hong Kong, China.
106. *The ACM Fifteenth Conference on Information and Knowledge Management (CIKM'06)*, November 6-11, 2006, Arlington, VA, USA.
107. *The Seventeenth European Conference on Machine Learning and the Tenth European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'06)*, September 18-22, 2006, Berlin, Germany.
108. *The Eighth International Conference on Data Warehousing and Knowledge Discovery (DaWaK'06)*, September 4-8, 2006, Krakow, Poland.
109. *The Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, August 20-23, 2006, Philadelphia, PA, USA.
110. *The Twenty-third International Conference on Machine Learning*, June 25-29, Pittsburgh, PA, USA.
111. *The Seventh International Conference on Web-Age Information Management (WAIM'06)*, June 17-19, 2006, Hong Kong, China.
112. *2006 ACM Symposium of Applied Computing (SAC'06)*, Special Track on Data Mining, April 23-27, 2006, Dijon, France.
113. *The Sixth SIAM International Conference on Data Mining (SDM'06)*, April 20-22, 2006, Bethesda, Maryland, USA.
114. *The Tenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'06)*, 9-12 April 2006, Singapore.
115. *The Twenty-second IEEE International Conference on Data Engineering (ICDE'06)*, April 3-7, 2006, Atlanta, GA, USA.
116. *The Eighth Asia Pacific Web Conference (APWeb'06)*, January 16-18, 2006, Harbin, China.
117. *The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'05)*, August 21-24, 2005, Chicago, IL, USA.
118. *The Seventh International Conference on Data Warehousing and Knowledge Discovery (DaWaK'05)*, August 22-26, 2005, Copenhagen, Denmark.
119. *2005 IEEE International Conference on Granular Computing (IEEE GrC'05)*, July 25-27, 2005, Beijing, China.
120. *The Twenty-fourth International Conference on Conceptual Modeling (ER'05)*, Klagenfurt, Austria, October 24-48, 2005.
121. *The Fifth SIAM International Conference on Data Mining (SDM'05)*, Newport Beach, CA, USA, April 21-23, 2005.
122. *The Seventh Asia Pacific Web Conference (APWeb'05)*, Shanghai, China, March 29-April 1, 2005.
123. *2005 ACM Symposium of Applied Computing (SAC'05)*, Special Track on Data Mining, March 13-17, 2005, Santa Fe, New Mexico, USA.
124. *The Third ACS/IEEE International Conference on Computer Systems and Applications (AICCSA'05)*, January 3-6, 2005, Cairo, Egypt.
125. *The Twenty-third International Conference on Conceptual Modeling (ER'04)*, November 8-12, 2004, Shanghai, China.
126. *The Fourth IEEE International Conference on Data Mining (ICDM '04)*, November 1-4, 2004, Bournemouth, United Kingdom.
127. *The Sixth International Conference on Data Warehousing and Knowledge Discovery (DaWaK'04)*, September 1-3, 2004, Zaragoza, Spain.

128. *The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, August 22-25, 2004, Seattle, WA, USA.
129. *2004 SIAM International Conference on Data Mining (SDM'04)*, April 22-24, 2004, Orlando, FL, USA.
130. *The Sixth Asia Pacific Web Conference (APWeb'04)*, April 14-17, 2004, Hangzhou, China.
131. *The Twentyth IEEE International Conference on Data Engineering (ICDE'04)*, March 30-April 2, 2004, Boston, MA, USA.
132. *2004 ACM Symposium of Applied Computing (SAC'04), Data Stream Track, Special Track on Data Mining, Technical Track on Novel E-Commerce Applications*, March 14-17, 2004, Nicosia, Cyprus.
133. *The Ninth International Conference on Database Systems for Advanced Applications (DASFAA'04)*, February 24-27, 2004, Jeju island, Korea.
134. *The Third IEEE International Conference on Data Mining (ICDM '03)*, November 19-22, 2003, Melbourne, Florida.
135. *The Fifth International Conference on Data Warehousing and Knowledge Discovery (DaWaK'03)*, September 1-5, 2003, Praha, Czech Republic.
136. *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, August 24-27, 2003, Washington, DC.
137. *The Fourth International Conference on Web-Age Information Management (WAIM'03)*, August 17-19, 2003, Chengdu, Sichuan, China.
138. *The Eighth International Conference on Database Systems for Advanced Applications (DASFAA'03)*, March 26-28, 2003, Kyoto, Japan.
139. *2003 ACM Symposium of Applied Computing (SAC'03), Technical Track on Novel E-Commerce Applications*, March 9-12, 2003, Melbourne, Florida, USA.
140. *The Eleventh ACM International Conference on Information and Knowledge Management (CIKM'02)*, November 4-9, 2002, McLean, VA, USA.
141. *The Eighteenth IEEE International Conference on Data Engineering (ICDE'02)*, February 26-March 1, 2002, San Jose, CA, USA.

10.7 Internal Services

- Academic year 2022-2023
 - Ph.D. Recruiting and Admission Committee and Communications Committee, Department of Computer Science, Duke University.
 - Faculty Search Committee, Department of Biostatistics and Bioinformatics, Duke University.
 - AI and Machine Learning Faculty Search Working Group Coordinator, Department of Electronic and Computer Engineering, Duke University.
 - Strategic Planning Committee, Department of Electronic and Computer Engineering, Duke University.