Message from the Chairs .................................................................................................................................. xxv
Organizing Committee .................................................................................................................................... xxvii
Reviewers.......................................................................................................................................................... xxix

Keynotes
Mosaics: Stratosphere, Flink and Beyond ........................................................................................................... 3
Volker Markl
Leveraging Data and People to Accelerate Data Science .................................................................................... 4
Laura Haas

TKDE Posters
Semantic Bootstrapping: A Theoretical Perspective .......................................................................................... 7
Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q. Zhu
IF-Matching: Towards Accurate Map-Matching with Information Fusion ....................................................... 9
Gang Hu, Jie Shao, Fenglin Liu, Yuan Wang, and Heng Tao Shen
A Mixed Generative-Discriminative Based Hashing Method ............................................................................. 11
Qi Zhang, Yang Wang, Jin Qian, Binbin Deng, and Xuanjing Huang
SPIRIT: A Tree Kernel-Based Method for Topic Person Interaction Detection
(Extended Abstract) ............................................................................................................................................. 13
Yung-Chun Chang, Chien Chin Chen, and Wen-Lian Hsu
Efficient Cache-Supported Path Planning on Roads (Extended Abstract) ....................................................... 15
Ying Zhang, Yu-Ling Hsueh, Wang-Chien Lee, and Yi-Hao Jhang
Personalized Influential Topic Search via Social Network Summarization ..................................................... 17
Jianxin Li, Chengfei Liu, Jeffrey Xu Yu, Yi Chen, Timos Sellis, and J. Shane Culpepper
Recommendation for Repeat Consumption from User Implicit Feedback ......................................................... 19
Jun Chen, Chaokun Wang, Jianmin Wang, and Philip S. Yu
PINOCCHIO: Probabilistic Influence-Based Location Selection over Moving  
Objects ...........................................................................................................................................21  
   Meng Wang, Hui Li, Jiangtao Cui, Ke Deng, Sourav S. Bhowmick,  
   and Zhenhua Dong  
K-Join: Knowledge-Aware Similarity Join .................................................................23  
   Zeyuan Shang, Yaxiao Liu, Guoliang Li, and Jianhua Feng  
Mining Suspicious Tax Evasion Groups in Big Data ................................................................25  
   Feng Tian, Tian Lan, Qinghua Zheng, Kuo-Ming Chao, Nick Godwin,  
   Nazaraf Shah, and Fan Zhang  
Influence Maximization in Trajectory Databases ......................................................................27  
   Long Guo, Dongxiang Zhang, Gao Cong, Wei Wu, and Kian-Lee Tan  
A Generic Method for Accelerating LSH-Based Similarity Join Processing  
(Extended Abstract) .........................................................................................................................29  
   Chenyun Yu, Sarana Nutanong, Hangyu Li, Cong Wang, and Xingliang Yuan  
The Moving K Diversified Nearest Neighbor Query ...............................................................31  
   Yu Gu, Guanli Liu, Jianzhong Qi, Hongfei Xu, Ge Yu, and Rui Zhang  
The Interaction Between Schema Matching and Record Matching in Data  
Integration (Extended Abstract) .....................................................................................................33  
   Binbin Gu, Zhixi Li, Xiangliang Zhang, An Liu, Guanfeng Liu, Kai Zheng,  
   Lei Zhao, and Xiaofang Zhou  
Learning Mixtures of Markov Chains from Aggregate Data with Structural  
Constraints (Extended Abstract) .......................................................................................................35  
   Dixin Luo, Hongteng Xu, Yi Zhen, Bistra Dilkina, Hongyuan Zha,  
   Xiaokang Yang, and Wenjun Zhang  
Patient Flow Prediction via Discriminative Learning of Mutually-Correcting  
Processes (Extended Abstract) .........................................................................................................37  
   Hongteng Xu, Weichang Wu, Shamim Nemati, and Hongyuan Zha  
Crowdsourced Data Management: A Survey ...........................................................................39  
   Guoliang Li, Jiannan Wang, Yudian Zheng, and Michael Franklin  
Conflict-Aware Weighted Bipartite b-Matching and Its Application  
to E-Commerce ...............................................................................................................................41  
   Cheng Chen, Lan Zheng, Venkatesh Srinivasan, Alex Thomo, Kui Wu,  
   and Anthony Sukow  
Mutually Beneficial Confluent Routing .......................................................................................43  
   Xinpeng Zhang, Yasuhiito Asano, and Masatoshi Yoshikawa  
Exploit Every Bit: Effective Caching for High-Dimensional Nearest Neighbor  
Search ...............................................................................................................................................45  
   Bo Tang, Man Lung Yiu, and Kien A. Hua
Bridging Feature Selection and Extraction: Compound Feature Generation
(Extended Abstract) ..................................................................................................................47
  Sreevani and C.A. Murthy

A Novel Cost-Based Model for Data Repairing .................................................................49
  Shuang Hao, Nan Tang, Guoliang Li, Jian He, Na Ta, and Jianhua Feng

Fast Memory Efficient Local Outlier Detection in Data Streams (Extended Abstract) ..........................................................................................................................51
  Mahsa Salehi, Christopher Leckie, James C. Bezdek, Tharshan Vaithianathan,
  and Xuyun Zhang

Finding Causality and Responsibility for Probabilistic Reverse Skyline Query
Non-Answers ......................................................................................................................53
  Yunjun Gao, Qing Liu, Gang Chen, Linlin Zhou, and Baihua Zheng

Bring Order into the Samples: A Novel Scalable Method for Influence
Maximization (Extended Abstract) .....................................................................................55
  Xiaoyang Wang, Ying Zhang, Wenjie Zhang, Xuemin Lin, and Chen Chen

Scalable Temporal Latent Space Inference for Link Prediction in Dynamic
Social Networks (Extended Abstract) ................................................................................57
  Linhong Zhu, Dong Guo, Junming Yin, Greg Ver Steeg, and Aram Galstyan

Collective Travel Planning in Spatial Networks ................................................................59
  Shuo Shang, Lisi Chen, Zhewei Wei, Christian S. Jensen, Ji-Rong Wen,
  and Panos Kalnis

Proxies for Shortest Path and Distance Queries .............................................................61
  Shuai Ma, Kaiyu Feng, Jianxin Li, Haixun Wang, Gao Cong, and Jinpeng Huai

gMark: Schema-Driven Generation of Graphs and Queries ...........................................63
  Guillaume Bagan, Angela Bonifati, Radu Ciucanu, George H. L. Fletcher,
  Aurélien Lemay, and Nicky Advokaat

SEDEX: Scalable Entity Preserving Data Exchange .......................................................65
  Yoones A. Sekhavat and Jeffrey Parsons

Efficient Distributed Density Peaks for Clustering Large Data Sets
in MapReduce ....................................................................................................................67
  Yanfeng Zhang, Shimin Cheny, and Ge Yu

Secure Data Deduplication with Dynamic Ownership Management in Cloud
Storage ...............................................................................................................................69
  Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang

Maximizing Acceptance in Rejection-aware Spatial Crowdsourcing ..........................71
  Libin Zheng and Lei Chen
ICDE Short Paper Posters

Mobi-SAGE: A Sparse Additive Generative Model for Mobile App Recommendation ..............................................................75
Hongzhi Yin, Liang Chen, Weiqing Wang, Xingzhong Du,
Quoc Viet Hung Nguyen, and Xiaofang Zhou

Clustering with Adaptive Manifold Structure Learning ..........................................................79
Zhao Kang, Chong Peng, and Qiang Cheng

Mining Precise-Positioning Episode Rules from Event Sequences ..............................................83
Xiang Ao, Ping Luo, Jin Wang, Fuzhen Zhuang, and Qing He

NetClus: A Scalable Framework for Locating Top-K Sites for Placement of Trajectory-Aware Services ..........................................................................................................................87
Shubhadip Mitra, Priya Saraf, Richa Sharma, Arnab Bhattacharya,
Sayan Ranuy, and Harsh Bhandari

Preserving-Ignoring Transformation Based Index for Approximate k Nearest Neighbor Search ..........................................................................................................................91
Gang Hu, Jie Shao, Dongxiang Zhang, Yang Yang, and Heng Tao Shen

Scalable Top-K Structural Diversity Search .....................................................................................95
Lijun Chang, Chen Zhang, Xuemin Lin, and Lu Qin

K-Dominant Skyline Join Queries: Extending the Join Paradigm to K-Dominant Skylines .................................................................................................................................99
Anuradha Awasthi, Arnab Bhattacharya, Sanchit Gupta, and Ujjwal Kumar Singh

SF-sketch: A Fast, Accurate, and Memory Efficient Data Structure to Store Frequencies of Data Items ..................................................................................................................103
Tong Yang, Lingtong Liu, Yibo Yan, Muhammad Shahzad, Yulong Shen,
Xiaoming Li, Bin Cui, and Gaogang Xie

Direction-Aware Why-Not Spatial Keyword Top-k Queries ......................................................................107
Lei Chen, Yafei Li, Jianliang Xu, and Christian S. Jensen

Searching Time Series with Invariance to Large Amounts of Uniform Scaling ......................................111
Yilin Shen, Yanping Chen, Eamonn Keogh, and Hongxia Jin

Mining Maximal Cliques on Dynamic Graphs Efficiently by Local Strategies ..................................115
Shengli Sun, Yimo Wang, Weilong Liao, and Wei Wang

Parallelizing Skip Lists for In-Memory Multi-Core Database Systems .............................................119
Zhongle Xie, Qingchao Cai, H.V. Jagadish, Beng Chin Ooi, and Weng-Fai Wong

From Raw Footprints to Personal Interests: Bridging the Semantic Gap via Trip Intention Aggregation ..................................................................................................................123
Long Guo, Dongxiang Zhang, Huayu Wu, Bin Cui, and Kian-Lee Tan

CrowdFusion: A Crowdsourced Approach on Data Fusion Refinement ........................................127
Yunfan Chen, Lei Chen, and Chen Jason Zhang
Correcting and Speeding-Up Bounds for Non-Uniform Graph Edit Distance .......................................................... 131
  David B. Blumenthal and Johann Gamper

Distance-Aware Encoding of Numerical Values for Privacy-Preserving Record Linkage ...................................................... 135
  Dimitrios Karapiperis, Aris Gkoulalas-Divanis, and Vassilios S. Verykios

Query Optimizations over Decentralized RDF Graphs ........................................................................................................... 139
  Ibrahim Abdelaziz, Essam Mansour, Mourad Ouzzani, Ashraf Aboulnaga, and Panos Kalnis

LTD-RBM: Robust and Fast Latent Truth Discovery Using Restricted Boltzmann Machines ........................................................................ 143
  Klaus Broelemann, Thomas Gottron, and Gjergji Kasneci

A Comparative Study of Log-Structured Merge-Tree-Based Spatial Indexes for Big Data ........................................................................ 147
  Young-Seok Kim, Taewoo Kim, Michael J. Carey, and Chen Li

Skew-Resistant Graph Partitioning ........................................................................................................................................... 151
  Angen Zheng, Alexandros Labrinidis, and Christos Faloutsos

Efficient Multistream Classification Using Direct Density Ratio Estimation .................................................................................. 155
  Ahsanul Haque, Swarup Chandra, Latifur Khan, Kevin Hamlen, and Charu Aggarwal

A Distance Measure for the Analysis of Polar Opinion Dynamics in Social Networks ........................................................................ 159
  Victor Amelkin, Petko Bogdanov, and Ambuj K. Singh

Cross-Network Clustering and Cluster Ranking for Medical Diagnosis ...................................................................................... 163
  Jingchao Ni, Hongliang Fei, Wei Fan, and Xiang Zhang

Reverse Query-Aware Locality-Sensitive Hashing for High-Dimensional Furthest Neighbor Search ................................................................ 167
  Qiang Huang, Jianlin Feng, and Qiong Fang

Ontology- and Sentiment-Aware Review Summarization .............................................................................................................. 171
  Nhat X.T. Le, Vagelis Hristidis, and Neal Young

ACTS: An Active Learning Method for Time Series Classification .......................................................................................... 175
  Fengchao Peng, Qiong Luo, and Lionel M. Ni

FuseM: Query-Centric Data Fusion on Structured Web Markup .................................................................................................... 179
  Ran Yu, Ujwal Gadiraju, Besnik Fetahu, and Stefan Dietze

Scalable Processing of Massive Uncertain Graph Data: A Simultaneous Processing Approach ................................................................ 183
  Zhaonian Zou, Faming Li, Jianzhong Li, and Yingshu Li

On Edge Classification in Networks with Structure and Content .............................................................................................. 187
  Charu C. Aggarwal, Yao Li, Philip S. Yu, and Yuchen Zhao
Adaptive State Space Partitioning of Markov Decision Processes for Elastic Resource Management .................................................................191
  Konstantinos Lolos, Ioannis Konstantinou, Verena Kantere, and Nectarios Koziris
Sampling and Reconstruction Using Bloom Filters .................................................................195
  Neha Sengupta, Amitabha Bagchi, Srikanta Bedathur, and Maya Ramanath
Top-k Frequent Items and Item Frequency Tracking over Sliding Windows of Any Sizes ........................................................................................................199
  Chunyao Song, Xuanming Liu, and Tingjian Ge
Joint Gaussian Based Measures for Multiple-Instance Learning ........................................203
  Linfei Zhou, Claudia Plant, and Christian Böhm
Answering Location-Aware Graph Reachability Queries on GeoSocial Data ..................207
  Mohamed Sarwat and Yuhan Sun
IRanker: Query-Specific Ranking of Reviewed Items .............................................................211
  Moloud Shahbazi, Matthew Wiley, and Vagelis Hristidis
Analyzing and Visualizing Scalar Fields on Graphs ............................................................215
  Yang Zhang, Yusu Wang, and Srinivasan Parthasarathy
Enterprise Community Detection .........................................................................................219
  Jiawei Zhang, Philip S. Yu, and Yuanhua Lv
Trading Data in Good Faith: Integrating Truthfulness and Privacy Preservation in Data Markets .........................................................................................223
  Chaoyue Niu, Zhenzhe Zheng, Fan Wu, Xiaofeng Gao, and Guihai Chen
Integrative Dynamic Reconfiguration in a Parallel Stream Processing Engine .................227
  Kasper Grud Skat Madsen, Yongluan Zhou, and Jianneng Cao
Secure KNN Queries over Encrypted Data: Dimensionality Is Not Always a Curse ..........231
  Xinyu Lei, Alex X. Liu, and Rui Li

Industry Posters

Improving Predictable Shared-Disk Clusters Performance for Database Clouds ..................237
  Anjan Kumar Amirishetty, Yunrui Li, Tolga Yurek, Mahesh Girkar, Wilson Chan,
  Graham Ivey, Vsevolod Panteleenko, and Ken Wong
DeepSD: Supply-Demand Prediction for Online Car-Hailing Services Using Deep Neural Networks .................................................................................243
  Dong Wang, Wei Cao, Jian Li, and Jieping Ye
Dynamic Statistics Collection in the Teradata Unified Data Architecture ..........................255
  Sung Jin Kim, Mohammed Al-Kateb, Paul Sinclair, Alain Crolotte,
  Chengyang Zhang, and Linda Rose
Hot or Not? Forecasting Cellular Network Hot Spots Using Sector Performance Indicators ..................................................259
  Joan Serrà, Ilias Leontiadis, Alexandros Karatzoglou, and Konstantina Papagiannaki

Joint User-Entity Representation Learning for Event Recommendation in Social Network ..........................................................271
  Lijun Tang and Eric Yi Liu

TencentBoost: A Gradient Boosting Tree System with Parameter Server ...........................................................................281
  Jie Jiang, Jiawei Jiang, Bin Cui, and Ce Zhang

SSD-Assisted Backup and Recovery for Database Systems ........................................................................................................285
  Yongseok Son, Jaeyoon Choi, Jekyeom Jeon, Cheolgi Min, Sunggon Kim, Heon Young Yeom, and Hyuck Han

The Good, the Bad, and the KPIs: How to Combine Performance Metrics to Better Capture Underperforming Sectors in Mobile Networks ..............................................................................................297
  Ilias Leontiadis, Joan Serrà, Alessandro Finamore, Giorgos Dimopoulos, and Konstantina Papagiannaki

Too Big to Eat: Boosting Analytics Data Ingestion from Object Stores with Scoop ........................................................................309
  Yosef Moatti, Eran Rom, Raul Gracia-Tinedo, Dalit Naor, Doron Chen, Josep Sampe, Marc Sanchez-Artigas, Pedro Garcia-Lopez, Filip Gluszak, Eric Deschdt, Francesco Pace, Daniele Venzano, and Pietro Michiardi

Research Track

Session: Graphs

UniWalk: Unidirectional Random Walk Based Scalable SimRank Computation over Large Graph ........................................................................325
  Xiongcai Luo, Jun Gao, Chang Zhou, and Jeffrey Xu Yu

A Fast Order-Based Approach for Core Maintenance ..................................................................................................................337
  Yikai Zhang, Jeffrey Xu Yu, Ying Zhang, and Lu Qin

Scalable and Interactive Graph Clustering Algorithm on Multicore CPUs ..........................................................................................349
  Son T. Mai, Martin Storgaard Dieu, Ira Assent, Jon Jacobsen, Jesper Kristensen, and Mathias Birk

Fast Computation of Dense Temporal Subgraphs .........................................................................................................................361
  Shuai Ma, Renjun Hu, Luoshu Wang, Xuelian Lin, and Jinpeng Huai
Session: Keyword Search, Text and Strings

Reverse Keyword-Based Location Search .................................................................................................................. 375
  Xike Xie, Xin Lin, Jianliang Xu, and Christian S. Jensen

Reverse Top-k Geo-Social Keyword Queries in Road Networks .................................................................................... 387
  Jingwen Zhao, Yunjun Gao, Gang Chen, Christian S. Jensen, Rui Chen, and Deng Cai

BWT Arrays and Mismatching Trees: A New Way for String Matching with k Mismatches ............................................... 399
  Yangjun Chen and Yujia Wu

Source-LDA: Enhancing Probabilistic Topic Models Using Prior Knowledge Sources ......................................................... 411
  Justin Wood, Patrick Tan, Wei Wang, and Corey Arnold

Session: Data Mining

Network Backboning with Noisy Data .......................................................................................................................... 425
  Michele Coscia and Frank M.H. Neffke

Scalable Informative Rule Mining ..................................................................................................................................... 437
  Guoyao Feng, Lukasz Golab, and Divesh Srivastava

Streaming k-Means Clustering with Fast Queries ........................................................................................................ 449
  Yu Zhang, Kanat Tangwongsan, and Srikanta Tirthapura

Density Based Clustering over Location Based Services ................................................................................................ 461
  Md Farhadur Rahman, Weimo Liu, Saad Bin Suhaim, Saravanan Thirumuruganathan, Nan Zhang, and Gautam Das

Session: Query Optimization and Provenance

Provenance-Aware Query Optimization ....................................................................................................................... 473
  Xing Niu, Raghav Kapoor, Boris Glavic, Dieter Gawlick, Zhen Hua Liu, and Venkatesh Radhakrishnan

A SQL-Middleware Unifying Why and Why-Not Provenance for First-Order Queries ................................................................ 485
  Seokki Lee, Sven Köhler, Bertram Ludäscher, and Boris Glavic

Extended Characteristic Sets: Graph Indexing for SPARQL Query Optimization ................................................................ 497
  Marios Meimaris, George Papastefanatos, Nikos Mamoulis, and Ioannis Anagnostopoulos

TT-Join: Efficient Set Containment Join ....................................................................................................................... 509
  Jianye Yang, Wenjie Zhang, Shiyu Yang, Ying Zhang, and Xuemin Lin
Session: Systems for New Analytics

Scalable Linear Algebra on a Relational Database System .................................................................523
   Shangyu Luo, Zekai J. Gao, Michael Gubanov, Luis L. Perez, and Christopher Jermaine

KeystoneML: Optimizing Pipelines for Large-Scale Advanced Analytics ...........................................535
   Evan R. Sparks, Shivaram Venkataraman, Tomer Kaftan, Michael J. Franklin, and Benjamin Recht

Parallel SPARQL Query Optimization .................................................................................................547
   Buwen Wu, Yongluan Zhou, Hai Jin, and Amol Deshpande

Efficient Scalable Accurate Regression Queries in In-DBMS Analytics ..............................................559
   Christos Anagnostopoulos and Peter Triantafillou

Towards Unified Data and Lifecycle Management for Deep Learning ..............................................571
   Hui Miao, Ang Li, Larry S. Davis, and Amol Deshpande

Session: Top-K, KNN and Skyline Querying

Monitoring the Top-m Rank Aggregation of Spatial Objects in Streaming Queries ..........................................................585
   Farhana M. Choudhury, Zhifeng Bao, J. Shane Culpepper, and Timos Sellis

Answering Top-k Exemplar Trajectory Queries ......................................................................................597
   Sheng Wang, Zhifeng Bao, J. Shane Culpepper, Timos Sellis, Mark Sanderson, and Xiaolin Qin

V-Tree: Efficient kNN Search on Moving Objects with Road-Network Constraints ........................................609
   Bilong Shen, Ying Zhao, Guoliang Li, Weimin Zheng, Yue Qin, Bo Yuan, and Yongming Rao

Sweet KNN: An Efficient KNN on GPU through Reconciliation between Redundancy Removal and Regularity ..................................................................................621
   Guoyang Chen, Yufei Ding, and Xipeng Shen

Secure Skyline Queries on Cloud Platform .............................................................................................633
   Jinfei Liu, Juncheng Yang, Li Xiong, and Jian Pei

Session: New Hardware

Accelerating Multi-Column Selection Predicates in Main-Memory - The Elf Approach ..........................................................647
   David Broneske, Veit Köppen, Gunter Saake, and Martin Schäler
Revisiting the Design of Data Stream Processing Systems on Multi-Core Processors
Shuhao Zhang, Bingsheng He, Daniel Dahlmeier, Amelie Chi Zhou, and Thomas Heinze

DIDO: Dynamic Pipelines for In-Memory Key-Value Stores on Coupled CPU-GPU Architectures
Kai Zhang, Jiayu Hu, Bingsheng He, and Bei Hua

On Log-Structured Merge for Solid-State Drives
Risi Thonangi and Jun Yang

Session: Security and Encryption
Adaptively Secure Conjunctive Query Processing over Encrypted Data for Cloud Computing
Rui Li and Alex X. Liu

Towards a Unifying Attribute Based Access Control Approach for NoSQL Datastores
Pietro Colombo and Elena Ferrari

Frequency-Hiding Dependency-Preserving Encryption for Outsourced Databases
Boxiang Dong and Wendy Wang

Secure and Efficient Query Processing over Hybrid Clouds
Kerim Yasin Oktay, Murat Kantarcioğlu, and Sharad Mehrotra

Session: Similarity Search
Capturing the Moment: Lightweight Similarity Computations
Georgios Damaskinos, Rachid Guerraoui, and Rhicheek Patra

An Efficient Framework for Exact Set Similarity Search Using Tree Structure Indexes
Yong Zhang, Xiuxing Li, Jin Wang, Ying Zhang, Chunxiao Xing, and Xiaojie Yuan

Role Discovery in Graphs Using Global Features: Algorithms, Applications and a Novel Evaluation Strategy
Pratik Vinay Gupte, Balaraman Ravindran, and Srinivasan Parthasarathy

Similarity Search in Graph Databases: A Multi-Layered Indexing Approach
Yongjiang Liang and Peixiang Zhao
Session: Potpourri

Posterior Snapshot Isolation ..................................................................................................................797
    Xuan Zhou, Xin Zhou, Zhengtai Yu, and Kian-Lee Tan

PrivSuper: A Superset-First Approach to Frequent Itemset Mining under Differential Privacy ..................................................................................................................................................809
    Ning Wang, Xiaokui Xiao, Yin Yang, Zhenjie Zhang, Yu Gu, and Ge Yu

Quantifying Differential Privacy under Temporal Correlations ..................................................................821
    Yang Cao, Masatoshi Yoshikawa, Yonghui Xiao, and Li Xiong

Tracking Matrix Approximation over Distributed Sliding Windows .........................................................833
    Haida Zhang, Zengfeng Huang, Zhewei Wei, Wenjie Zhang, and Xuemin Lin

Session: Social Networks

Temporal Influence Blocking: Minimizing the Effect of Misinformation in Social Networks ...........................................847
    Chonggang Song, Wynne Hsu, and Mong Li Lee

Complex Event-Participant Planning and Its Incremental Variant ..................................................................859
    Yurong Cheng, Ye Yuan, Lei Chen, Christophe Giraud-Carrier, and Guoren Wang

Most Influential Community Search over Large Social Networks ............................................................871
    Jianxin Li, Xinjue Wang, Ke Deng, Xiaochun Yang, Timos Sellis, and Jeffrey Xu Yu

Boosting Information Spread: An Algorithmic Approach .............................................................................883
    Yishi Lin, Wei Chen, and John C.S. Lui

Session: Data Cleaning

Cleaning Data with Forbidden Itemsets .........................................................................................................897
    Joeri Rammelaere, Floris Geerts, and Bart Goethals

Parallel Progressive Approach to Entity Resolution Using MapReduce ..................................................909
    Yasser Al-towim and Sharad Mehrotra

A Collective, Probabilistic Approach to Schema Mapping ........................................................................921
    Angelika Kimmig, Alex Memory, Ren’ee J. Miller, and Lise Getoor

Cleaning Relations Using Knowledge Bases ..........................................................................................933
    Shuang Hao, Nan Tang, Guoliang Li, and Jian Li

Session: Learning and Outlier Detection

Time Series Classification by Sequence Learning in All-Subsequence Space ............................................947
    Thach Le Nguyen, Severin Gsponer, and Georgiana Ifrim
Multi-Tactic Distance-Based Outlier Detection .................................................................959
  Lei Cao, Yizhou Yan, Caitlin Kuhlman, Qingyang Wang, Elke A. Rundensteiner,
  and Mohamed Eltabakh

Link Prediction across Aligned Networks with Sparse and Low Rank Matrix
Estimation ..................................................................................................................................971
  Jiawei Zhang, Jianhui Chen, Shi Zhi, Yi Chang, Philip S. Yu, and Jiawei Han

LSHiForest: A Generic Framework for Fast Tree Isolation Based Ensemble
Anomaly Analysis .........................................................................................................................983
  Xuyun Zhang, Wanchun Dou, Qiang He, Rui Zhou, Christopher Leckie,
  Ramamohanarao Kotagiri, and Zoran Salcic

Session: Crowdsourcing and Recommender Systems
Prediction-Based Task Assignment in Spatial Crowdsourcing ...............................................997
  Peng Cheng, Xiang Lian, Lei Chen, and Cyrus Shahabi

Trichromatic Online Matching in Real-Time Spatial Crowdsourcing ..................................1009
  Tianshu Song, Yongxin Tong, Libin Wang, Jieying She, Bin Yao, Lei Chen,
  and Ke Xu

Tuning Crowdsourced Human Computation ............................................................................1021
  Chen Cao, Jiayang Tu, Zheng Liu, Lei Chen, and H.V. Jagadish

Scalable and Interpretable Product Recommendations via Overlapping
Co-Clustering ...............................................................................................................................1033
  Reinhard Heckel, Michail Vlachos, Thomas Parnell, and Celestine Duenner

Session: Distributed Processing
In-Memory Distributed Matrix Computation Processing and Optimization ..........................1047
  Yongyang Yu, Mingjie Tang, Walid G. Aref, Qutaibah M. Malluhi,
  Mostafa M. Abbas, and Mourad Ouzzani

Fast and Scalable Distributed Set Similarity Joins for Big Data Analytics ..........................1059
  Chuitian Rong, Chunbin Lin, Yasin N. Silva, Jianguo Wang, Wei Lu,
  and Xiaoyong Du

Fast and Scalable Distributed Boolean Tensor Factorization ..................................................1071
  Namyong Park, Sejoon Oh, and U. Kang

Spinner: Scalable Graph Partitioning in the Cloud ...............................................................1083
  Claudio Martella, Dionysios Logothetis, Andreas Loukas, and Georgos Siganos

Distributed Publish/Subscribe Query Processing on the Spatio-Textual Data
Stream ..........................................................................................................................................1095
  Zhida Chen, Gao Cong, Zhenjie Zhang, Tom Z.J. Fuz, and Lisi Chen
Industry Track

**Session: Predictive Analytics**

Predictive Provisioning: Efficiently Anticipating Usage in Azure SQL Database ..............................................1111  
Lalitha Viswanathan, Bikash Chandra, Willis Lang, Karthik Ramachandra, Jignesh M. Patel, Ajay Kalhan, David J. Dewitt, and Alan Halverson

Xshare-a-Ride: A Search Optimized Dynamic Ride Sharing System  
with Approximation Guarantee .........................................................................................................................1117  
Raja Subramaniam Thangaraj, Koyel Mukherjee, Gurulingesh Raravi, Asmita Metrewar, Narendra Annamaneni, and Koushik Chattopadhyay

Real-Time Novel Event Detection from Social Media ..........................................................................................1129  
Quanzhi Li, Armineh Nourbakhsh, Sameena Shah, and Xiaomo Liu

Anomaly Detection in Large Databases Using Behavioral Patterning ...............................................................1140  
Hanna Mazzawi, Gal Dalaly, David Rozenblatz, Liat Ein-Dorx, Matan Niniox, and Ofer Lavi

The Microsoft Reactive Framework Meets the Internet of Moving Things .......................................................1150  
Abdel-tawab M. Hendawi, Jayant Gupta, Youying Shi, Hossam Fattah, and Mohamed Ali

**Session: New Systems**

Twitter Heron: Towards Extensible Streaming Engines .......................................................................................1165  
Maosong Fu, Ashvin Agrawal, Avrilia Floratou, Bill Graham, Andrew Jorgensen, Mark Li, Neng Lu, Karthik Ramasamy, Sriram Rao, and Cong Wang

Feisu: Fast Query Execution over Heterogeneous Data Sources on Large-Scale Clusters ......................................1173  
An Qin, Yuan Yuan, Dai Tan, Pengyu Sun, Xiang Zhang, Hao Cao, Rubao Lee, and Xiaodong Zhang

DistributedLog: A High Performance Replicated Log Service .............................................................................1183  
Sijie Guo, Robin Dhamankar, and Leigh Stewart

Making Big Data Simple with dashDB Local .....................................................................................................1195  
Sam Lightstone, Russ Ohanian, Michael Haide, James Cho, Michael Springgay, and Torsten Steinbach

**Session: Optimization and Benchmarks**

Optimizing UNION ALL Join Queries in Teradata ...............................................................................................1209  
Mohammed Al-Kateb, Paul Sinclair, Alain Crolotte, Lu Ma, Grace Au, and Sanjay Nair
Multi-Query Optimization for Complex Event Processing in SAP ESP ..................................................1213
   Shuhao Zhang, Hoang Tam Vo, Daniel Dahlmeier, and Bingsheng He

BigBench V2: The New and Improved BigBench ...................................................................................1225
   Ahmad Ghazal, Todor Ivanov, Pekka Kostamaa, Alain Crolotte, Ryan Voong,
   Mohammed Al-Kateb, Waleed Ghazal, and Roberto V. Zicari

Providing Enhanced Functionality for Data Store Clients .................................................................1237
   Arun Iyengar

Modeling Scalability of Distributed Machine Learning .................................................................1249
   Alexander Ulanov, Andrey Simanovsky, and Manish Marwah

Applications Track

Session 1

Automated Personalized Feedback in Introductory Java Programming MOOCs ......................................1259
   Victor J. Marin, Tobin Pereira, Srinivas Sridharan, and Carlos R. Rivero

Vibration Analysis for IoT Enabled Predictive Maintenance ...........................................................1271
   Deokwoo Jung, Zhenjie Zhang, and Marianne Winslett

Mining Spatio-Temporal Reachable Regions over Massive Trajectory Data ......................................1283
   Guojun Wu, Yichen Ding, Yanhua Li, Jie Bao, Yu Zheng, and Jun Luo

Smart Personalized Routing for Smart Cities .......................................................................................1295
   Abdeltawab M. Hendawi, Aqeel Rustum, Amr A. Ahmadain, David Hazel,
   Ankur Teredesai, Dev Oliver, Mohamed Ali, and John A. Stankovic

Session 2

Robust Power Line Outage Detection with Unreliable Phasor Measurements ..................................1309
   Jose Cordova-Garcia and Xin Wang

Database System Support for Personalized Recommendation Applications ........................................1320
   Mohamed Sarwat, Raha Moraffah, Mohamed F. Mokbel, and James L. Avery

Efficient Exploration of Telco Big Data with Compression and Decaying ........................................1332
   Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti,
   and Mohamed F. Mokbel

Privacy Preserving Anonymity for Periodical SRS Data Publishing ...............................................1344
   Jie-Teng Wang and Wen-Yang Lin

Demo Track
Session 1: Cloud, Stream, Query Processing, Provenance

A Learning-Based Service for Cost and Performance Management of Cloud Databases .................................................................1361
Ryan Marcus, Sofiya Semenova, and Olga Papaemmanouil

AdaStorm: Resource Efficient Storm with Adaptive Configuration .................................................................1363
Zujian Weng, Qi Guo, Chunkai Wang, Xiaofeng Meng, and Bingsheng He

AZTEC: A Cloud-based Computational Platform to Integrate Biomedical Resources .........................................................1365
Patrick Tan, Yichao Zhou, Xinxin Huang, Giuseppe M. Mazzeo, and Chelsea Ju

Demonstrating SolveDB: An SQL-Based DBMS for Optimization Applications .........................................................1367
Laurynas Šiknys and Torben Bach Pedersen

HDBExpDetector: Aggregate Sudden-Change Detector over Dynamic Web Databases .........................................................1369
Saad Bin Suhaim, Nan Zhang, Gautam Das, and Ali Jaoua

IS2R: A System for Refining Reverse Top-k Queries .................................................................1371
Qing Liu, Yunjun Gao, Linlin Zhou, and Gang Chen

POLYTICS: Provenance-Based Analytics of Data-Centric Applications .................................................................1373
Pierre Bourhis, Daniel Deutch, and Yuval Moskovitch

Selective In-Place Appends for Real: Reducing Erases on Wear-prone DBMS Storage .................................................................1375
Sergey Hardock, Ilia Petrovy, Robert Gottstein, and Alejandro Buchmann

Understanding the Security Challenges of Oblivious Cloud Storage with Asynchronous Accesses .................................................................1377
Cetin Sahin, Aaron Magat, Victor Zakhary, Amr El Abbadi, Huijia (Rachel) Lin, and Stefano Tessaro

Urd: A Data Summarization Tool for the Acquisition of Meaningful Cardinality Constraints with Probabilistic Intervals .................................................................1379
Tania K. Roblot and Sebastian Link

Session 2: Graph Analytics, Social Networks, Machine Learning

Demonstration of Kite: A Scalable System for Microblogs Data Management .................................................................1383
Amr Magdy and Mohamed F. Mokbel

Adaptive Topic Follow-Up on Twitter .................................................................1385
Abdulrahman Alsaud, Mehdi Sadri, Yasser Altowim, and Sharad Mehrotra

BEAMS: Bounded Event Detection in Graph Streams .................................................................1387
Mohammad Hossein Namaki, Keyvan Sasani, Yinghui Wu, and Tingjian Ge
GQFast: Fast Graph Exploration with Context-Aware Autocompletion ..................................................1389
Chunbin Lin, Jianguo Wang, and Yannis Papakonstantinou

GscalerCloud: A Web-Based Graph Scaling Service .............................................................................1391
J.W. Zhang, Anwesha Mal, and Y.C. Tay

ModelHub: Deep Learning Lifecycle Management .................................................................................1393
Hui Miao, Ang Li, Larry S. Davis, and Amol Deshpande

Privacy Cyborg: Towards Protecting the Privacy of Social Media Users ...............................................1395
Theodore Georgiou, Amr El Abbadi, and Xifeng Yan

SocialLens: Searching and Browsing Communities by Content and Interaction ....................................1397
Hongyun Cai, Vincent W. Zheng, Penghe Chen, Fanwei Zhu, Kevin Chen-Chuan Chang, and Zi Huang

TABOO: Detecting Unstructured Sensitive Information Using Recursive Neural Networks ................1399
Jan Neerbeky, Ira Assentz, and Peter Dolog

Session 3: Applications, Data Visualization, Text Analysis, Data Integration

A Demonstration of TextDB: Declarative and Scalable Text Analytics on Large Data Sets ..........................................................1403
Zuozi Wang, Flavio Bayer, Seungjin Lee, Kishore Narendran, Xuxi Pan, Qing Tang, Jimmy Wang, and Chen Li

A Human-and-Machine Cooperative Framework for Entity Resolution with Quality Guarantees ..........................................................1405
Zhaoqiang Chen, Qun Chen, and Zhanhuai Li

A Scalable Data Integration and Analysis Architecture for Sensor Data of Pediatric Asthma ..........................1407
Dimitris Stripelis, José Luis Ambite, Yao-Yi Chiang, Sandrah P. Eckel, and Rima Habre

DANCE: Data Cleaning with Constraints and Experts ..........................................................1409
Ahmad Assadi, Tova Milo, and Slava Novgorodov

DV8: Interactive Analysis of Aviation Data ..................................................................................1411
Behrooz Omidvar-Tehrani, Amab Nandi, Nicholas Meyer, Dalton Flanagan, and Seth Young

Hippo in Action: Scalable Indexing of a Billion New York City Taxi Trips and Beyond ..........................1413
Jia Yu, Raha Moraffah, and Mohamed Sarwat

Landslide Information Service Based on Composition of Physical and Social Sensors ................................1415
Aibek Musaev and Calton Pu

xx
MAROON+: A System for Profiling Entities over Time
Furong Li, Mong Li Lee, and Wynne Hsu

SPATE: Compacting and Exploring Telco Big Data
Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, and Mohamed F. Mokbel

VisFlow: A Visual Database Integration and Workflow Querying System
Xin Mou, Hasan M. Jamil, and Xiaogang Ma

PhD Symposium Session 1

Keynote: Research with Real Users
Magda Balazinska

Multiple-Query Optimization of Regular Path Queries
Zahid Abul-Basher

RELNA: Ranking Attributes in Social Networks to Detect Overlapping Communities Efficiently
Bella Martínez-Seis

PhD Symposium Session 2

Judgment Analysis Based on Crowdsourced Opinions
Sujoy Chatterjee, Anirban Mukhopadhyay, and Malay Bhattacharyya

Effects of User Interactions on Online Social Recommender Systems
Anahita Davoudi

Tutorials

Community Search over Big Graphs: Models, Algorithms, and Opportunities
Xin Huang, Laks V.S. Lakshmanan, and Jianliang Xu

Bringing Semantics to Spatiotemporal Data Mining: Challenges, Methods, and Applications
Chao Zhang, Quan Yuan, and Jiawei Han

Web-Scale Blocking, Iterative and Progressive Entity Resolution
Kostas Stefanidis, Vassilis Christophides, and Vasilis Efthymiou

The Challenges of Global-Scale Data Management
Faisal Nawab, Divyakant Agrawal, and Amr El Abbadi

Handling Uncertainty in Geo-Spatial Data
Andreas Züfle, Goce Trajcevski, Dieter Pfoser, Matthias Renz, Matthew T. Rice, Timothy Leslie, Paul Delamater, and Tobias Emrich
Panels

Data Science Education: We're Missing the Boat, Again .................................................................1473
   Bill Howe, Michael Franklin, Laura Haas, Tim Kraska, and Jeffrey Ullman

Small Data ..............................................................................................................................................1475
   Oliver Kennedy, D. Richard Hipp, Stratos Idreos, Amélie Marian, Arnab Nandi,
   Carmela Troncoso, and Eugene Wu

HDMM Workshop

Session 1

Fairness in Group Recommendations in the Health Domain ...............................................................1481
   Maria Stratigi, Haridimos Kondylakis, and Kostas Stefanidis

MASC: Automatic Sleep Stage Classification Based on Brain and Myoelectric Signals
.............................................................................................................................................................................1489
   Yuta Suzuki, Makito Sato, Hiroaki Shiokawa, Masashi Yanagisawa,
   and Hiroyuki Kitagawa

Survival Topic Models for Predicting Outcomes for Trauma Patients ....................................................1497
   Yuanyang Zhang, Richard Jiang, and Linda Petzold

Session 2

Case Study: Classification Algorithms Comparison for the Multi-Label Problem
of Chronic Pelvic Pain Diagnosing ....................................................................................................................1507
   Vinicius Oliverio and Omero Bendicto Poli-Neto

Enabling Real-Time Drug Abuse Detection in Tweets ................................................................................1510
   Nhathai Phan, Soon Ae Chun, Manasi Bhole, and James Geller

Mother Smoking During Pregnancy and ADHD in Children ................................................................1515
   Jing (Melody) Yao

Session 3

Integrated Theory-and Data-Driven Feature Selection in Gene Expression
Data Analysis .........................................................................................................................................................1525
   Vineet K. Raghu, Xiaoyu Ge, Panos K. Chrysanthis, and Panayiotis V. Benos

A Mortality Study for ICU Patients Using Bursty Medical Events ...........................................................1533
   Luca Bonomi and Xiaqian Jiang

Characterizing Organ Donation Awareness from Social Media ..............................................................1541
   Diogo F. Pacheco, Diego Pinheiro, Martin Cadeiras, and Ronaldo Menezes
DesWeb Workshop

Session 1
Towards Scalable Non-Monotonic Stream Reasoning via Input Dependency Analysis ......................................................1553
Thu-Le Pham, Alessandra Mileo, and Muhammad Intizar Ali
Distance-Based Triple Reordering for SPARQL Query Optimization .................................................................1559
Marios Meimaris and George Papastefanatos
NOUS: Construction and Querying of Dynamic Knowledge Graphs ..........................................................1563
Sutanay Choudhury, Khushbu Agarwal, Sumit Purohit, Baichuan Zhang, Meg Pirrung, Will Smith, and Mathew Thomas

Session 2
Ontology-Based Approach for Academic Evaluation System .........................................................................................1569
Siti Aminah, lis Afriyanti, and Adila Krisnadhi
PolyFuse: A Large-Scale Hybrid Data Fusion System .........................................................................................1575
Michael Gubanov
On Recommending Evolution Measures: A Human-Aware Approach ..........................................................1579
Kostas Stefanidis, Haridimos Kondylakis, and Georgia Troullinou

Active and HardDB Workshops

HardDB Keynote
The Configurable Cloud - Accelerating Hyperscale Datacenter Services with FPGA ..................................................1587
André Putnam

HardDB Papers
Parallel Selectivity Estimation for Optimizing Multidimensional Spatial Join Processing on GPUs ..................................1591
Jianting Zhang, Simin You, and Le Gruenwald
Are Databases Fit for Hybrid Workloads on GPUs? A Storage Engine's Perspective ...............................................1599
Marcus Pinnecke, David Broneske, Gabriel Campero Durand, and Gunter Saake
Active Invited Talks (Academic)

Data Management Systems on Future Hardware: Challenges and Opportunities .................................................................1609
  Bingsheng He

Processing Declarative Queries through Generating Imperative Code in Managed Runtimes .................................................................1610
  Stratis D. Viglas

Enabling Effective Utilization of GPUs for Data Management Systems ...................................................................................1612
  Xiaodong Zhang

Active Invited Talks (Industry)

Tradeoffs and Considerations in the Design of Accelerators for Database Applications ........................................................................1615
  Roger Moussalli

Hardware Acceleration of Database Analytics .........................................................................................................................1616
  Evangelia Sitaridi

Author Index