



Research Session 1A: Graph Processing

- 1  **Distance-Aware Influence Maximization in Geo-Social Network**
(*Xiaoyang Wang, Ying Zhang, Wenjie Zhang, Xuemin Lin*)
- 13  **Topical Influence Modeling via Topic-Level Interests and Interactions on Social Curation Services**
(*Daehoon Kim, Jae-Gil Lee, Byung Suk Lee*)
- 25  **BlackHole: Robust Community Detection Inspired by Graph Drawing**
(*Sungsu Lim, Junghoon Kim, Jae-Gil Lee*)
- 37  **Revenue Maximization by Viral Marketing: A Social Network Host's Perspective**
(*Arijit Khan, Benjamin Zehnder, Donald Kossmann*)
-

Research Session 1B: Crowdsourcing

- 49  **Online Mobile Micro-Task Allocation in Spatial Crowdsourcing**
(*Yongxin Tong, Jieying She, Bolin Ding, Libin Wang, Lei Chen*)
- 61  **Crowdsourced POI Labelling: Location-Aware Result Inference and Task Assignment**
(*Huiqi Hu, Yudian Zheng, Zhifeng Bao, Guoliang Li, Jianhua Feng, Reynold Cheng*)
- 73  **Mutual Benefit Aware Task Assignment in a Bipartite Labor Market**
(*Liu Zheng, Lei Chen*)
- 85  **Computing Connected Components with Linear Communication Cost in Pregel-Like Systems**
(*Xing Feng, Lijun Chang, Xuemin Lin, Lu Qin, Wenjie Zhang*)



Research Session 2A: Graph Algorithmics

- 97  **VColor: A Practical Vertex-Cut Based Approach for Coloring Large Graphs**
(*Yun Peng, Byron Choi, Bingsheng He, Shuigeng Zhou, Ruzhi Xu, Xiaohui Yu*)
- 109  **Compressing Graphs by Grammars**
(*Sebastian Maneth, Fabian Peternek*)
- 121  **Planar: Parallel Lightweight Architecture-Aware Adaptive Graph Repartitioning**
(*Angen Zheng, Alexandros Labrinidis, Panos K. Chrysanthis*)
- 133  **I/O Efficient Core Graph Decomposition at Web Scale**
(*Dong Wen, Lu Qin, Ying Zhang, Xuemin Lin, Jeffrey Xu Yu*)
-

Research Session 2B: Beyond Relational Query Processing

- 145  **Reachability and Time-Based Path Queries in Temporal Graphs**
(*Huanhuan Wu, Yuzhen Huang, James Cheng, Jinfeng Li, Yiping Ke*)
- 157  **Scalable Supergraph Search in Large Graph Databases**
(*Bingqing Lyu, Lu Qin, Xuemin Lin, Lijun Chang, Jeffrey Xu Yu*)
- 169  **Hobbes3: Dynamic Generation of Variable-Length Signatures for Efficient Approximate Subsequence Mappings**
(*Jongik Kim, Chen Li, Xiaohui Xie*)
- 181  **NoSE: Schema Design for NoSQL Applications**
(*Michael J. Mior, Kenneth Salem, Ashraf Aboulmaga, Rui Liu*)



Research Session 2C: Privacy

- 193  **Towards Virtual Private NoSQL Datastores**
(*Pietro Colombo, Elena Ferrari*)
- 205  **Differentially Private Multi-Party High-Dimensional Data Publishing**
(*Sen Su, Peng Tang, Xiang Cheng, Rui Chen, Zequn Wu*)
- 217  **Optimizing Secure Classification Performance with Privacy-Aware Feature Selection**
(*Erman Pattuk, Murat Kantarcioglu, Huseyin Ulusoy, Bradley Malin*)
- 229  **Differentially Private Frequent Subgraph Mining**
(*Shengzhi Xu, Sen Su, Li Xiong, Xiang Cheng, Ke Xiao*)
-

Research Session 3A: Graph Proximity

- 241  **Being Prepared in a Sparse World: The Case of KNN Graph Construction**
(*Antoine Boutet, Anne-Marie Kermarrec, Nupur Mittal, Francois Taiani*)
- 253  **pSCAN: Fast and Exact Structural Graph Clustering**
(*Lijun Chang, Wei Li, Xuemin Lin, Lu Qin, Wenjie Zhang*)
- 265  **CSI_GED: An Efficient Approach for Graph Edit Similarity Computation**
(*Karam Gouda, Mosab Hassaan*)
- 277  **Semantic Proximity Search on Graphs with Metagraph-Based Learning**
(*Yuan Fang, Wenqing Lin, Vincent W. Zheng, Min Wu, Kevin Chen-Chuan Chang, Xiao-Li Li*)



Research Session 3B: Scalable Query Processing

- 289  **Private Spatial Data Aggregation in the Local Setting**
(*Rui Chen, Haoran Li, A.K. Qin, Shiva Prasad Kasiviswanathan, Hongxia Jin*)
- 301  **A Novel, Low-Latency Algorithm for Multiple Group-By Query Optimization**
(*Duy-Hung Phan, Pietro Michiardi*)
- 313  **Load Balancing and Skew Resilience for Parallel Joins**
(*Aleksandar Vitorovic, Mohammed Elseidy, Christoph Koch*)
- 325  **Platform-Independent Robust Query Processing**
(*Srinivas Karthik, Jayant R. Haritsa, Sreyash Kenkre, Vinayaka Pandit*)
-

Research Session 3C: Preference and Trust

- 337  **Authentication of Function Queries**
(*Guolei Yang, Ying Cai, Zhenbi Hu*)
- 349  **“Told You I Didn’t Like It”: Exploiting Uninteresting Items for Effective Collaborative Filtering**
(*Won-Seok Hwang, Juan Parc, Sang-Wook Kim, Jongwuk Lee, Dongwon Lee*)
- 361  **Practical Private Shortest Path Computation Based on Oblivious Storage**
(*Dong Xie, Guanru Li, Bin Yao, Xuan Wei, Xiaokui Xiao, Yunjun Gao, Minyi Guo*)
- 373  **Practical Privacy-Preserving User Profile Matching in Social Networks**
(*Xun Yi, Elisa Bertino, Fang-Yu Rao, Athman Bouguettaya*)



Research Session 4A: Graph Mining

- 385  **An Embedding Approach to Anomaly Detection**
(*Renjun Hu, Charu Aggarwal, Shuai Ma, Jinpeng Huai*)
- 397  **Cross-Layer Betweenness Centrality in Multiplex Networks with Applications**
(*Tanmoy Chakraborty, Ramasuri Narayanam*)
- 409  **NXgraph: An Efficient Graph Processing System on a Single Machine**
(*Yuze Chi, Guohao Dai, Yu Wang, Guangyu Sun, Guoliang Li, Huazhong Yang*)
- 421  **Mining Social Ties Beyond Homophily**
(*Hongwei Liang, Ke Wang, Feida Zhu*)
-

Research Session 4B: Conscious Big Data Processing

- 433  **Self-Adaptive Linear Hashing for Solid State Drives**
(*Chengcheng Yang, Peiquan Jin, Lihua Yue, Dezhi Zhang*)
- 445  **On Main-Memory Flushing in Microblogs Data Management Systems**
(*Amr Magdy, Rami Alghamdi, Mohamed F. Mokbel*)
- 457  **ICE: Managing Cold State for Big Data Applications**
(*Badrish Chandramouli, Justin Levandoski, Eli Cortez*)
- 469  **HAWK: Hardware Support for Unstructured Log Processing**
(*Prateek Tandon, Faissal M. Sleiman, Michael Cafarella, Thomas F. Wenisch*)



Research Session 4C: Data Streams

- 481  **Efficient Handling of Concept Drift and Concept Evolution Over Stream Data**
(*Ahsanul Haque, Latifur Khan, Michael Baron, Bhavani Thuraisingham, Charu Aggarwal*)
- 493  **Quality-Driven Disorder Handling for M-Way Sliding Window Stream Joins**
(*Yuanzhen Ji, Jun Sun, Anisoara Nica, Zbigniew Jerzak, Gregor Hackenbroich, Christof Fetzer*)
- 505  **Context-Aware Advertisement Recommendation for High-Speed Social News Feeding**
(*Yuchen Li, Dongxiang Zhang, Ziquan Lan, Kian-Lee Tan*)
- 517  **Tolerating Correlated Failures in Massively Parallel Stream Processing Engines**
(*Li Su, Yongluan Zhou*)
-

Research Session 5A: Graph Patterns

- 529  **Spatial Influence — Measuring Followship in the Real World**
(*Huy Pham, Cyrus Shahabi*)
- 541  **Durable Graph Pattern Queries on Historical Graphs**
(*Konstantinos Semertzidis, Evaggelia Pitoura*)
- 553  **Link Prediction in Graph Streams**
(*Peixiang Zhao, Charu Aggarwal, Gewen He*)
- 565  **SimRank Computation on Uncertain Graphs**
(*Rong Zhu, Zhaonian Zou, Jianzhong Li*)



Research Session 5B: Parallel and Distributed Big Data Processing

- 577  **Input Selection for Fast Feature Engineering**
(*Michael R. Anderson, Michael Cafarella*)
- 589  **When Two Choices Are Not Enough: Balancing at Scale in Distributed Stream Processing**
(*Muhammad Anis Uddin Nasir, Gianmarco De Francisci Morales, Nicolas Kourtellis, Marco Serafini*)
- 601  **HadoopViz: A MapReduce Framework for Extensible Visualization of Big Spatial Data**
(*Ahmed Eldawy, Mohamed F. Mokbel, Christopher Jonathan*)
- 613  **Efficient Fault-Tolerance for Iterative Graph Processing on Distributed Dataflow Systems**
(*Chen Xu, Markus Holzemer, Manohar Kaul, Volker Markl*)
-

Research Session 5C: Clustering

- 625  **A Model-Based Approach for Text Clustering with Outlier Detection**
(*Jianhua Yin, Jianyong Wang*)
- 637  **Streaming Spectral Clustering**
(*Shinjae Yoo, Hao Huang, Shiva Prasad Kasiviswanathan*)
- 649  **Accelerating Large Scale Centroid-Based Clustering with Locality Sensitive Hashing**
(*Ryan McConville, Xin Cao, Weiru Liu, Paul Miller*)
- 661  **PurTreeClust: A Purchase Tree Clustering Algorithm for Large-Scale Customer Transaction Data**
(*Xiaojun Chen, Joshua Zhexue Huang, Jun Luo*)



Research Session 6A: Spatial Analytics

- 673  **TRANSFORMERS: Robust Spatial Joins on Non-Uniform Data Distributions**
(*Mirjana Pavlovic, Thomas Heinis, Farhan Tauheed, Panagiotis Karras, Anastasia Ailamaki*)
- 685  **Finding the Minimum Spatial Keyword Cover**
(*Dong-Wan Choi, Jian Pei, Xuemin Lin*)
- 697  **Answering Why-Not Spatial Keyword Top- k Queries via Keyword Adaption**
(*Lei Chen, Jianliang Xu, Xin Lin, Christian S. Jensen, Haibo Hu*)
- 709  **Influence Based Cost Optimization on User Preference**
(*Jianye Yang, Ying Zhang, Wenjie Zhang, Xuemin Lin*)
-

Research Session 6B: Analytics on Big Data

- 721  **The CRO Kernel: Using Concomitant Rank Order Hashes for Sparse High Dimensional Randomized Feature Maps**
(*Kave Eshghi, Mehran Kafai*)
- 731  **MuVE: Efficient Multi-Objective View Recommendation for Visual Data Exploration**
(*Humaira Ehsan, Mohamed A. Sharaf, Panos K. Chrysanthis*)
- 743  **Collaborative Analytics for Data Silos**
(*Jinkyu Kim, Heonseok Ha, Byung-Gon Chun, Sungroh Yoon, Sang K. Cha*)
- 755  **Visualization-Aware Sampling for Very Large Databases**
(*Yongjoo Park, Michael Cafarella, Barzan Mozafari*)



Research Session 6C: Uncertain and Probabilistic Data

- 767  **Answering Why-Not Questions on Metric Probabilistic Range Queries**
(*Lu Chen, Yunjun Gao, Kai Wang, Christian S. Jensen, Gang Chen*)
- 779  **Analyzing Data-Centric Applications: Why, What-If, and How-To**
(*Pierre Bourhis, Daniel Deutch, Yuval Moskovitch*)
- 791  **CLEAR: Clustering based on Locality Embedding And Reconstruction**
(*Zhonglong Zheng, Minqi Mao, Songxia Ma*)
- 799  **OLAP over Probabilistic Data Cubes I: Aggregating, Materializing, and Querying**
(*Xike Xie, Xingjun Hao, Torben Bach Pedersen, Peiquan Jin, Jinchuan Chen*)
-

Research Session 7A: Scalable Matrix-Based Analytics

- 811  **SCouT: Scalable Coupled Matrix-Tensor Factorization — Algorithm and Discoveries**
(*ByungSoo Jeon, Inah Jeon, Lee Sael, U Kang*)
- 823  **Topology-Aware Optimization of Big Sparse Matrices and Matrix Multiplications on Main-Memory Systems**
(*David Kernert, Wolfgang Lehner, Frank Köhler*)
- 835  **2PCP: Two-Phase CP Decomposition for Billion-Scale Dense Tensors**
(*Xinsheng Li, Shengyu Huang, K. Selçuk Candan, Maria Luisa Sapino*)
- 847  **Distributed Low Rank Approximation of Implicit Functions of a Matrix**
(*David P. Woodruff, Peilin Zhong*)



Research Session 7B: Trajectories and Roads

- 859  **Fuzzy Trajectory Linking**
(*Huayu Wu, Mingqiang Xue, Jianneng Cao, Panagiotis Karras, Wee Siong Ng, Kee Kiat Koo*)
- 871  **Keyword-Aware Continuous kNN Query on Road Networks**
(*Bolong Zheng, Kai Zheng, Xiaokui Xiao, Han Su, Hongzhi Yin, Xiaofang Zhou, Guohui Li*)
- 883  **Crowdsourcing-Based Real-Time Urban Traffic Speed Estimation: From Trends to Speeds**
(*Huiqi Hu, Guoliang Li, Zhifeng Bao, Yan Cui, Jianhua Feng*)
-

Research Session 8A: Data Explorations and Event Analytics

- 895  **Learning Abstract Snippet Detectors with Temporal Embedding in Convolutional Neural Networks**
(*Jiajun Liu, Kun Zhao, Brano Kusy, Ji-Rong Wen, Kai Zheng, Raja Jurdak*)
- 906  **Interactive Data Exploration with Smart Drill-Down**
(*Manas Joglekar, Hector Garcia-Molina, Aditya Parameswaran*)
- 918  **ClEver: Clustering Events with High Density of True-to-False Occurrence Ratio**
(*Georgios Theodoridis, Thierry Benoist*)
- 930  **Event Regularity and Irregularity in a Time Unit**
(*Lijian Wan, Tingjian Ge*)



Research Session 8B: Spatial Analytics

- 942  **Discovering Interpretable Geo-Social Communities for User Behavior Prediction**
(*Hongzhi Yin, Zhiting Hu, Xiaofang Zhou, Hao Wang, Kai Zheng, Quoc Viet Hung Nguyen, Shazia Sadiq*)
- 954  **SPORE: A Sequential Personalized Spatial Item Recommender System**
(*Weiying Wang, Hongzhi Yin, Shazia Sadiq, Ling Chen, Min Xie, Xiaofang Zhou*)
- 966  **Reverse Nearest Neighbor Heat Maps: A Tool for Influence Exploration**
(*Yu Sun, Rui Zhang, Andy Yuan Xue, Jianzhong Qi, Xiaoyong Du*)
- 978  **Automatic User Identification Method Across Heterogeneous Mobility Data Sources**
(*Wei Cao, Zhengwei Wu, Dong Wang, Jian Li, Haishan Wu*)
-

Research Session 8C: Web Data Processing

- 990  **Fast Top-K Search in Knowledge Graphs**
(*Shengqi Yang, Fangqiu Han, Yinghui Wu, Xifeng Yan*)
- 1002  **Learning to Query: Focused Web Page Harvesting for Entity Aspects**
(*Yuan Fang, Vincent W. Zheng, Kevin Chen-Chuan Chang*)
- 1014  **Discovering Neighborhood Pattern Queries by Sample Answers in Knowledge Base**
(*Jialong Han, Kai Zheng, Aixin Sun, Shuo Shang, Ji-Rong Wen*)
- 1026  **Incremental Updates on Compressed XML**
(*Stefan Böttcher, Rita Hartel, Thomas Jacobs, Sebastian Maneth*)



Research Session 9A: Visual Analytics in Social Networks

- 1038  **Edge Classification in Networks**
(*Charu Aggarwal, Gewen He, Peixiang Zhao*)
- 1050  **Minfer: A Method of Inferring Motif Statistics from Sampled Edges**
(*Pinghui Wang, John C.S. Lui, Don Towsley, Junzhou Zhao*)
- 1062  **SLR: A Scalable Latent Role Model for Attribute Completion and Tie Prediction in Social Networks**
(*Lizi Liao, Qirong Ho, Jing Jiang, Ee-Peng Lim*)
- 1074  **TOPIC: TOWard Perfect InfluenCe Graph Summarization**
(*Lei Shi, Sibai Sun, Yuan Xuan, Yue Su, Hanghang Tong, Shuai Ma, Yang Chen*)
-

Research Session 9B: Optimization of Temporal, Spatial Data

- 1086  **A GPU-Based Index to Support Interactive Spatio-Temporal Queries Over Historical Data**
(*Harish Doraiswamy, Huy T. Vo, Cláudio T. Silva, Juliana Freire*)
- 1098  **An Interval Join Optimized for Modern Hardware**
(*Danila Piatov, Sven Helmer, Anton Dignös*)
- 1110  **Efficiently Computing Reverse k Furthest Neighbors**
(*Shenlu Wang, Muhammad Aamir Cheema, Xuemin Lin, Ying Zhang, Dongxi Liu*)
- 1122  **Indexing Multi-Metric Data**
(*Maximilian Franzke, Tobias Emrich, Andreas Züfle, Matthias Renz*)



Research Session 9C: Data Integration and Strings

- 1134  **DataXFormer: A Robust Transformation Discovery System**
(*Ziawasch Abedjan, John Morcos, Ihab F. Ilyas, Mourad Ouzzani, Paolo Papotti, Michael Stonebraker*)
- 1146  **Joint Repairs for Web Wrappers**
(*Stefano Ortona, Giorgio Orsi, Tim Furche, Marcello Buoncristiano*)
- 1158  **Fast Motif Discovery in Short Sequences**
(*Honglei Liu, Fangqiu Han, Hongjun Zhou, Xifeng Yan, Kenneth S. Kosik*)
- 1170  **A Novel Fast and Memory Efficient Parallel MLCS Algorithm for Long and Large-Scale Sequences Alignments**
(*Yanni Li, Yuping Wang, Zhensong Zhang, Yaxin Wang, Ding Ma, Jianbin Huang*)



Industrial and Applications 1: Distributed/Parallel Systems

- 1182  **SQL-SA for Big Data Discovery: Polymorphic and Parallelizable SQL User-Defined Scalar and Aggregate Infrastructure in Teradata Aster 6.20**
(*Xin Tang, Robert Wehrmeister, James Shau, Abhirup Chakraborty, Daley Alex, Awny Al Omari, Feven Atnafu, Jeff Davis, Litao Deng, Deepak Jaiswal, Chittaranjan Keswani, Yafeng Lu, Chao Ren, Tom Reyes, Kashif Siddiqui, David Simmen, Devendra Vidhani, Ling Wang, Shuai Yang, Daniel Yu*)
- 1194  **Flow-Join: Adaptive Skew Handling for Distributed Joins Over High-Speed Networks**
(*Wolf Rödiger, Sam Idicula, Alfons Kemper, Thomas Neumann*)
- 1206  **Moolle: Fan-Out Control for Scalable Distributed Data Stores**
(*SungJu Cho, Andrew Carter, Joshua Ehrlich, Jane Alam Jan*)
-

Industrial and Applications 2: Potpourri 1

- 1218  **Personal Recommendation Using Deep Recurrent Neural Networks in NetEase**
(*Sai Wu, Weichao Ren, Chengchao Yu, Gang Chen, Dongxiang Zhang, Jingbo Zhu*)
- 1230  **Recommendations Meet Web Browsing: Enhancing Collaborative Filtering Using Internet Browsing Logs**
(*Royi Ronen, Elad Yom-Tov, Gal Lavee*)
- 1239  **SPDO: High-Throughput Road Distance Computations on Spark Using Distance Oracles**
(*Shangfu Peng, Jagan Sankaranarayanan, Hanan Samet*)



Industrial and Applications 3: Potpourri 2

- 1251  **GARNET: A Holistic System Approach for Trending Queries in Microblogs**
(*Christopher Jonathan, Amr Magdy, Mohamed F. Mokbel, Albert Jonathan*)
- 1263  **Using SSDs to Scale up Google Fusion Tables, a Database-in-the-Cloud**
(*Yingyi Bu, Felix Halim, Changkyu Kim, Hongrae Lee, Jayant Madhavan*)
- 1275  **FastFunction: Replacing a Herd of Lemmings with a Cheetah — A Ruby Framework for Interaction with PostgreSQL Databases**
(*Henrietta Dombrovskaya, Srivathsava Rangarajan, Jonathan Marks*)
-

Industrial and Applications 4: Real Time Analytics

- 1287  **A Column Store Engine for Real-Time Streaming Analytics**
(*Alex Skidanov, Anders J. Papito, Adam Prout*)
- 1298  **Fault-Tolerant Real-Time Analytics with Distributed Oracle Database In-Memory**
(*Niloy Mukherjee, Shasank Chavan, Maria Colgan, Mike Gleeson, Xiaoming He, Allison Holloway, Jesse Kamp, Kartik Kulkarni, Tirthankar Lahiri, Juan Loaiza, Neil Macnaughton, Atrayee Mullick, Sujatha Muthulingam, Vivekanandhan Raja, Raunak Rungta*)
- 1310  **Virtual Lightweight Snapshots for Consistent Analytics in NoSQL Stores**
(*Fernando Chirigati, Jérôme Siméon, Martin Hirzel, Juliana Freire*)



Demo Session 1D

- 1322  **ALEX: Automatic Link Exploration in Linked Data**
(*Ahmed El-Roby, Ashraf Aboulnaga*)
- 1326  **A Query System for Social Media Signals**
(*Dolan Antenucci, Michael R. Anderson, Penghua Zhao, Michael Cafarella*)
- 1330  **Beat the DIvA — Decentralized Identity Validation for Online Social Networks**
(*Leila Bahri, Amira Soliman, Jacopo Squillaci, Barbara Carminati, Elena Ferrari, Sarunas Girdzijauskas*)
- 1334  **OptImatch: Semantic Web System for Query Problem Determination**
(*Guilherme Damasio, Piotr Mierzejewski, Jaroslaw Szlichta, Calisto Zuzarte*)
- 1338  **INSQ: An Influential Neighbor Set Based Moving k NN Query Processing System**
(*Chuanwen Li, Yu Gu, Jianzhong Qi, Ge Yu, Rui Zhang, Qingxu Deng*)
- 1342  **graphVizdb: A Scalable Platform for Interactive Large Graph Visualization**
(*Nikos Bikakis, John Liagouris, Maria Krommyda, George Papastefanatos, Timos Sellis*)



Demo Session 1E

- 1346  **QB2OLAP : Enabling OLAP on Statistical Linked Open Data**
(*Jovan Varga, Lorena Etcheverry, Alejandro A. Vaisman, Oscar Romero, Torben Bach Pedersen, Christian Thomsen*)
- 1350  **SEED: A System for Entity Exploration and Debugging in Large-Scale Knowledge Graphs**
(*Jun Chen, Yueguo Chen, Xiaoyong Du, Xiangling Zhang, Xuan Zhou*)
- 1354  **Ranking Support for Matched Patterns Over Complex Event Streams: The CEP^R System**
(*Jiaqi Gu, Jin Wang, Carlo Zaniolo*)
- 1358  **QPlain: Query By Explanation**
(*Daniel Deutch, Amir Gilad*)
- 1362  **InVerDa — Co-Existing Schema Versions Made Foolproof**
(*Kai Herrmann, Hannes Voigt, Thorsten Seyschab, Wolfgang Lehner*)
- 1366  **ANALOC: Efficient ANALytics Over LOCation Based Services**
(*Md Farhadur Rahman, Saad Bin Suhaim, Weimo Liu, Saravanan Thirumuruganathan, Nan Zhang, Gautam Das*)



Demo Session 2D

- 1370  **A New Privacy-Preserving Solution for Clustering Massively Distributed Personal Times-Series**
(*Tristan Allard, Georges Hébrail, Florent Masegla, Esther Pacitti*)
- 1374  **Mercury: Metro Density Prediction with Recurrent Neural Network on Streaming CDR Data**
(*Victor C. Liang, Richard T.B. Ma, Wee Siong Ng, Li Wang, Marianne Winslett, Huayu Wu, Shanshan Ying, Zhenjie Zhang*)
- 1378  **ORLF: A Flexible Framework for Online Record Linkage and Fusion**
(*El Kindi Rezig, Eduard C. Dragut, Mourad Ouzzani, Ahmed K. Elmagarmid, Walid G. Aref*)
- 1382  **TemProRA: Top-k Temporal-Probabilistic Results Analysis**
(*Katerina Papaioannou, Michael Böhlen*)
- 1386  **Leveraging Non-Volatile Memory for Instant Restarts of In-Memory Database Systems**
(*David Schwalb, Martin Faust, Markus Dreseler, Pedro Flemming, Hasso Plattner*)
- 1390  **Java2SDG: Stateful Big Data Processing for the Masses**
(*Raul Castro Fernandez, Panagiotis Garefalakis, Peter Pietzuch*)



Demo Session 2E

- 1394  **Flexible Hybrid Stores: Constraint-Based Rewriting to the Rescue**
(*Francesca Bugiotti, Damian Bursztyn, Alin Deutsch, Ioana Manolescu, Stamatis Zampetakis*)
- 1398  **WatCA: The Waterloo Consistency Analyzer**
(*Hua Fan, Shankha Chatterjee, Wojciech Golab*)
- 1402  **DebEAQ — Debugging Empty-Answer Queries on Large Data Graphs**
(*Elena Vasilyeva, Thomas Heinze, Maik Thiele, Wolfgang Lehner*)
- 1406  **Cruncher: Distributed In-Memory Processing for Location-Based Services**
(*Ahmed S. Abdelhamid, Mingjie Tang, Ahmed M. Aly, Ahmed R. Mahmood, Thamir Qadah, Walid G. Aref, Saleh Basalamah*)
- 1410  **A Demonstration of GeoSpark: A Cluster Computing Framework for Processing Big Spatial Data**
(*Jia Yu, Jinxuan Wu, Mohamed Sarwat*)



Tutorials

- 1414  **Indoor Data Management**
(*Hua Lu, Muhammad Aamir Cheema*)
- 1418  **Scaling Up Truth Discovery**
(*Laure Berti-Équille*)
- 1420  **Scalable Data Management: NoSQL Data Stores in Research and Practice**
(*Felix Gessert, Norbert Ritter*)
- 1424  **The Era of Big Spatial Data**
(*Ahmed Eldawy, Mohamed F. Mokbel*)
- 1428  **Accelerating Database Workloads by Software-Hardware-System Co-Design**
(*Rajesh R. Bordawekar, Mohammad Sadoghi*)
- 1432  **Data Profiling**
(*Ziawasch Abedjan, Lukasz Golab, Felix Naumann*)
- 1436  **Blocking for Large-Scale Entity Resolution: Challenges, Algorithms, and Practical Examples**
(*George Papadakis, Themis Palpanas*)
- 1440  **Microblogs Data Management and Analysis**
(*Amr Magdy, Mohamed F. Mokbel*)
-

Panels

- 1444  **Dark Data: Are We Solving the Right Problems?**
(*Michael Cafarella, Ihab F. Ilyas, Marcel Kornacker, Tim Kraska, Christopher Ré*)
- 1446  **Big Data Quality — Whose Problem is it?**
(*Shazia Sadiq, Paolo Papotti*)



TKDE Posters

- 1448  **Similarity Group-By Operators for Multi-Dimensional Relational Data**
(*Mingjie Tang, Ruby Y. Tahboub, Walid G. Aref, Mikhail J. Atallah, Qutaibah M. Malluhi, Mourad Ouzzani, Yasin N. Silva*)
- 1450  **Using Memetic Algorithm for Instance Coreference Resolution**
(*Xingsi Xue, Yuping Wang*)
- 1452  **Crowdsourcing for Top-K Query Processing Over Uncertain Data**
(*Eleonora Ciceri, Piero Fraternali, Davide Martinenghi, Marco Tagliasacchi*)
- 1454  **Adaptive Noise Immune Cluster Ensemble Using Affinity Propagation**
(*Zhiwen Yu, Guoqiang Han, Le Li, Jiming Liu, Jun Zhang*)
- 1456  **Efficient Similarity Join Based on Earth Mover's Distance Using MapReduce**
(*Jia Xu, Bin Lei, Yu Gu, Marianne Winslett, Ge Yu, Zhenjie Zhang*)
- 1458  **NATERGM: A Model for Examining the Role of Nodal Attributes in Dynamic Social Media Networks**
(*Shan Jiang, Hsinchun Chen*)
- 1460  **Tensor Canonical Correlation Analysis for Multi-View Dimension Reduction**
(*Yong Luo, Dacheng Tao, Kotagiri Ramamohanarao, Chao Xu, Yonggang Wen*)
- 1462  **TRIP: An Interactive Retrieving-Infering Data Imputation Approach**
(*Zhixu Li, Lu Qin, Hong Cheng, Xiangliang Zhang, Xiaofang Zhou*)
- 1464  **t-Closeness Through Microaggregation: Strict Privacy with Enhanced Utility Preservation**
(*Jordi Soria-Comas, Josep Domingo-Ferrer, David Sánchez, Sergio Martínez*)
- 1466  **Domain-Sensitive Recommendation with User-Item Subgroup Analysis**
(*Jing Liu, Yu Jiang, Zechao Li, Xi Zhang, Hanqing Lu*)

*TKDE Posters continued ...*

- 1468  **Semantic-Aware Blocking for Entity Resolution**
(*Qing Wang, Mingyuan Cui, Huizhi Liang*)
- 1470  **Privacy-Preserving Indoor Localization on Smartphones**
(*Andreas Konstantinidis, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, Paschalis Mpeis, Nikos Pelekis, Yannis Theodoridis*)
- 1472  **CRoM and HuspExt: Improving Efficiency of High Utility Sequential Pattern Extraction**
(*Oznur Kirmemis Alkan, Pinar Karagoz*)
- 1474  **Joint Structure Feature Exploration and Regularization for Multi-Task Graph Classification**
(*Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, Philip S. Yu*)
- 1476  **Efficient Answering of Why-Not Questions in Similar Graph Matching**
(*Md. Saiful Islam, Chengfei Liu, Jianxin Li*)
- 1478  **CloudKeyBank: Privacy and Owner Authorization Enforced Key Management Framework**
(*Xiuxia Tian, Ling Huang, Tony Wu, Xiaoling Wang, Aoying Zhou*)
- 1480  **Association Discovery in Two-View Data**
(*Matthijs van Leeuwen, Esther Galbrun*)
- 1482  **i²MapReduce: Incremental MapReduce for Mining Evolving Big Data**
(*Yanfeng Zhang, Shimin Chen, Qiang Wang, Ge Yu*)
- 1484  **Incremental Semi-Supervised Clustering Ensemble for High Dimensional Data Clustering**
(*Zhiwen Yu, Peinan Luo, Si Wu, Guoqiang Han, Jane You, Hareton Leung, Hau-San Wong, Jun Zhang*)

*TKDE Posters continued ...*

- 1486  **Virtual Denormalization via Array Index Reference for Main Memory OLAP**
([Yansong Zhang](#), [Xuan Zhou](#), [Ying Zhang](#), [Yu Zhang](#), [Mingchuan Su](#), [Shan Wang](#))
- 1488  **Reverse Keyword Search for Spatio-Textual Top-k Queries in Location-Based Services**
([Xin Lin](#), [Jianliang Xu](#), [Haibo Hu](#))
- 1490  **Distributed In-Memory Processing of All k Nearest Neighbor Queries**
([Georgios Chatzimilioudis](#), [Constantinos Costa](#), [Demetrios Zeinalipour-Yazti](#),
[Wang-Chien Lee](#), [Evaggelia Pitoura](#))
- 1492  **Accelerated Continuous Conditional Random Fields for Load Forecasting**
([Hongyu Guo](#))
- 1494  **Querying Knowledge Graphs by Example Entity Tuples**
([Nandish Jayaram](#), [Arijit Khan](#), [Chengkai Li](#), [Xifeng Yan](#), [Ramez Elmasri](#))
- 1496  **Efficient Top-k Retrieval on Massive Data**
([Xixian Han](#), [Jianzhong Li](#), [Hong Gao](#))
- 1498  **Safe Distribution and Parallel Execution of Data-Centric Workflows Over the Publish/Subscribe Abstraction**
([Matin Jergler](#), [Hans-Arno Jacobsen](#), [Mohammad Sadoghi](#), [Richard Hull](#), [Roman Vaculín](#))
- 1500  **Top-k Dominating Queries on Incomplete Data**
([Xiaoye Miao](#), [Yunjun Gao](#), [Baihua Zheng](#), [Gang Chen](#), [Huiyong Cui](#))
- 1502  **Subspace Based Network Community Detection Using Sparse Linear Coding**
([Arif Mahmood](#), [Michael Small](#))
- 1504  **DSP-CC: I/O Efficient Parallel Computation of Connected Components in Billion-Scale Networks**
([Min-Soo Kim](#), [Sangyeon Lee](#), [Wook-Shin Han](#), [Himchan Park](#), [Jeong-Hoon Lee](#))

*TKDE Posters continued ...*

- 1506  **Mining Temporal Patterns in Interval-Based Data**
(*Yi-Cheng Chen, Wen-Chih Peng, Suh-Yin Lee*)
- 1508  **Pattern-Aided Regression Modeling and Prediction Model Analysis**
(*Guozhu Dong, Vahid Taslimitehrani*)
- 1510  **Geo-Social K-Cover Group Queries for Collaborative Spatial Computing**
(*Yafei Li, Rui Chen, Jianliang Xu, Qiao Huang, Haibo Hu, Byron Choi*)
- 1512  **Diversified Hidden Markov Models for Sequential Labeling**
(*Maoying Qiao, Wei Bian, Richard Yi Da Xu, Dacheng Tao*)
- 1514  **Metric All- k -Nearest-Neighbor Search**
(*Lu Chen, Yunjun Gao, Gang Chen, Haida Zhang*)
- 1516  **Efficient Enforcement of Action-Aware Purpose-Based Access Control Within Relational Database Management Systems**
(*Pietro Colombo, Elena Ferrari*)
- 1518  **A Practical and Effective Sampling Selection Strategy for Large Scale Deduplication**
(*Guilherme Dal Bianco, Renata Galante, Carlos A. Heuser, Marcos Gonçalves, Sergio Canuto*)
- 1520  **Anonymizing Collections of Tree-Structured Data**
(*Olga Gkountouna, Manolis Terrovitis*)
- 1522  **Enabling Scalable Geographic Service Sharing with Weighted Imprecise Voronoi Cells**
(*Xike Xie, Peiquan Jin, Man Lung Yiu, Jiang Du, Christian S. Jensen, Mingxuan Yuan*)
- 1524  **Beyond Millisecond Latency kNN Search on Commodity Machine**
(*Bailong Liao, Leong Hou U, Man Lung Yiu, Zhiguo Gong*)
- 1526  **Unsupervised Ranking of Multi-Attribute Objects Based on Principal Curves**
(*Chun-Guo Li, Xing Mei, Bao-Gang Hu*)

*TKDE Posters continued ...*

- 1528  **Scalable Algorithms for Nearest-Neighbor Joins on Big Trajectory Data**
(*Yixiang Fang, Reynold Cheng, Wenbin Tang, Silviu Maniu, Xuan Yang*)
- 1530  **Maximizing a Record's Standing in a Relation**
(*Yu Tang, Yilun Cai, Nikos Mamoulis*)
- 1532  **Structure-Preserving Subgraph Query Services**
(*Zhe Fan, Byron Choi, Qian Chen, Jianliang Xu, Haibo Hu, Sourav S. Bhowmick*)
- 1534  **TrGraph: Cross-Network Transfer Learning via Common Signature Subgraphs**
(*Meng Fang, Jie Yin, Xingquan Zhu, Chengqi Zhang*)
- 1536  **Crawling Hidden Objects with kNN Queries**
(*Hui Yan, Zhiguo Gong, Nan Zhang, Tao Huang, Hua Zhong, Jun Wei*)
- 1538  **Indexing Evolving Events from Tweet Streams**
(*Hongyun Cai, Zi Huang, Divesh Srivastava, Qing Zhang*)
- 1540  **App Relationship Calculation: An Iterative Process**
(*Ming Liu, Chong Wu, Xiang-Nan Zhao, Chin-Yew Lin, Xiao-Long Wang*)
- 1542  **Efficient Probabilistic Supergraph Search**
(*Wenjie Zhang, Xuemin Lin, Ying Zhang, Ke Zhu, Gaoping Zhu*)
- 1544  **Improving Accuracy and Robustness of Self-Tuning Histograms by Subspace Clustering**
(*Andranik Khachatryan, Emmanuel Müller, Klemens Böhm, Christian Stier*)
- 1546  **CrowdOp: Query Optimization for Declarative Crowdsourcing Systems**
(*Ju Fan, Meihui Zhang, Stanley Kok, Meiyu Lu, Beng Chin Ooi*)
- 1548  **Time-Series Classification with COTE: The Collective of Transformation-Based Ensembles**
(*Anthony Bagnall, Jason Lines, Jon Hills, Aaron Bostrom*)

*TKDE Posters continued ...*

- 1550  **A Framework for Enabling User Preference Profiling Through Wi-Fi Logs**
(*Yao-Chung Fan, Yu-Chi Chen, Kuan-Chieh Tung, Kuo-Chen Wu, Arbee L.P. Chen*)
- 1552  **Understanding Short Texts Through Semantic Enrichment and Hashing**
(*Zheng Yu, Haixun Wang, Xuemin Lin, Min Wang*)
- 1554  **Application Sensitive Energy Management Framework for Storage Systems**
(*Norifumi Nishikawa, Miyuki Nakano, Masaru Kitsuregawa*)
- 1556  **Capacity-Constrained Network-Voronoi Diagram**
(*KwangSoo Yang, Apurv Hirsh Shekhar, Dev Oliver, Shashi Shekhar*)
- 1558  **Similar Subtree Search Using Extended Tree Inclusion**
(*Tomoya Mori, Atsuhiko Takasu, Jesper Jansson, Jaewook Hwang, Takeyuki Tamura, Tatsuya Akutsu*)
- 1560  **Change-Point Detection in a Sequence of Bags-of-Data**
(*Kensuke Koshijima, Hideitsu Hino, Noboru Murata*)
- 1562  **Reputation Aggregation in Peer-to-Peer Network Using Differential Gossip Algorithm**
(*Ruchir Gupta, Yatindra Nath Singh*)
- 1564  **Differentially Private Frequent Itemset Mining via Transaction Splitting**
(*Sen Su, Shengzhi Xu, Xiang Cheng, Zhengyi Li, Fangchun Yang*)
- 1566  **Location Aware Keyword Query Suggestion Based on Document Proximity**
(*Shuyao Qi, Dingming Wu, Nikos Mamoulis*)
- 1568  **The Prediction of Venture Capital Co-Investment Based on Structural Balance Theory**
(*Yun Zhou, Zhiyuan Wang, Jie Tang, Jar-Der Luo*)
- 1570  **XACML Policy Evaluation with Dynamic Context Handling**
(*Nariman Ammar, Zaki Malik, Abdelmounaam Rezgui, Elisa Bertino*)



TKDE Posters continued ...

- 1572  **kNNVWC: An Efficient k -Nearest Neighbours Approach Based on Various-Widths Clustering**
(*Abdulmohsen Almalawi, Adil Fahad, Zahir Tari, Muhammad Aamir Cheema, Ibrahim Khalil*)
- 1574  **Fast Best-Effort Search on Graphs with Multiple Attributes**
(*Senjuti Basu Roy, Tina Eliassi-Rad, Spiros Papadimitriou*)
- 1576  **Top-k Spatio-Textual Similarity Join**
(*Huiqi Hu, Guoliang Li, Zhifeng Bao, Jianhua Feng, Yongwei Wu, Zhiguo Gong, Yaoqiang Xu*)
- 1578  **Joint Search by Social and Spatial Proximity**
(*Kyriakos Mouratidis, Jing Li, Yu Tang, Nikos Mamoulis*)