

# **2011 IEEE 27th International Conference on Data Engineering (ICDE 2011)**

**Hannover, Germany  
11 – 16 April 2011**

**Pages 1-684**



**IEEE Catalog Number: CFP11026-PRT  
ISBN: 978-1-4244-8959-6**

## Keynotes

---

- 1        **Embarrassingly Scalable Database Systems**  
*Anastasia Ailamaki, EPFL, Switzerland*
- 2        **Ontological Queries: Rewriting and Optimization**  
*Georg Gottlob, Giorgio Orsi, Andreas Pieris, University of Oxford, UK*
- 14       **Playing Games with Databases**  
*Johannes Gehrke, Cornell University, USA*

## Research Session 1: Social Networks and Personal Information

---

- 15       **Interactive Itinerary Planning**  
*Senjuti Basu Roy<sup>1</sup>, Gautam Das<sup>1</sup>, Sihem Amer-Yahia<sup>2</sup>, Cong Yu<sup>3</sup>*  
*<sup>1</sup>University of Texas at Arlington, USA; <sup>2</sup>Yahoo! Labs, Spain; <sup>3</sup>Google Research, USA*
- 27       **CubeLSI: An Effective and Efficient Method for Searching Resources in Social Tagging Systems**  
*Bin Bi, Sau Dan Lee, Ben Kao, Reynold Cheng, University of Hong Kong, China*
- 39       **Adding Regular Expressions to Graph Reachability and Pattern Queries**  
*Wenfei Fan<sup>1</sup>, Jianzhong Li<sup>2</sup>, Shuai Ma<sup>1</sup>, Nan Tang<sup>1</sup>, Yinghui Wu<sup>1</sup>*  
*<sup>1</sup>University of Edinburgh, UK; <sup>2</sup>Harbin Institute of Technology, China*
- 51       **Efficient Core Decomposition in Massive Networks**  
*James Cheng<sup>1</sup>, Yiping Ke<sup>2</sup>, Shumo Chu<sup>1</sup>, M. Tamer Özsu<sup>3</sup>*  
*<sup>1</sup>Nanyang Technological University, Singapore; <sup>2</sup>Shenzhen Institute of Advanced Technology, China; <sup>3</sup>University of Waterloo, Canada*

## Research Session 2: Web Applications and Cloud Computing

---

- 63       **T-Verifier: Verifying Truthfulness of Fact Statements**  
*Xian Li<sup>1</sup>, Weiyi Meng<sup>1</sup>, Clement Yu<sup>2</sup>*  
*<sup>1</sup>Binghamton University, USA; <sup>2</sup>University of Illinois at Chicago, USA*
- 75       **Flexible Use of Cloud Resources Through Profit Maximization and Price Discrimination**  
*Konstantinos Tsakalozos<sup>1</sup>, Herald Kllapi<sup>1</sup>, Eva Sitaridi<sup>2</sup>, Mema Roussopoulos<sup>1</sup>, Dimitris Paparas<sup>2</sup>, Alex Delis<sup>1</sup>*  
*<sup>1</sup>University of Athens, Greece; <sup>2</sup>Columbia University, USA*
- 87       **Intelligent Management of Virtualized Resources for Database Systems in Cloud Environment**  
*Pengcheng Xiong<sup>1</sup>, Yun Chi<sup>2</sup>, Shenghuo Zhu<sup>2</sup>, Hyun Jin Moon<sup>2</sup>, Calton Pu<sup>1</sup>, Hakan Hacigümüş<sup>2</sup>*  
*<sup>1</sup>Georgia Institute of Technology, USA; <sup>2</sup>NEC Laboratories America, USA*
- 99       **Extensibility and Data Sharing in Evolving Multi-Tenant Databases**  
*Stefan Aulbach<sup>1</sup>, Michael Seibold<sup>1</sup>, Dean Jacobs<sup>2</sup>, Alfons Kemper<sup>1</sup>*  
*<sup>1</sup>Technische Universität München, Germany; <sup>2</sup>SAP AG, Germany*

---

## Research Session 3: Streams and Sensor Networks

---

- 111 **Semantic Stream Query Optimization Exploiting Dynamic Metadata**  
*Luping Ding<sup>1</sup>, Karen Works<sup>2</sup>, Elke A. Rundensteiner<sup>2</sup>*  
<sup>1</sup>Oracle Corporation, USA; <sup>2</sup>Worcester Polytechnic Institute, USA
- 123 **High-Performance Nested CEP Query Processing Over Event Streams**  
*Mo Liu<sup>1</sup>, Elke A. Rundensteiner<sup>1</sup>, Daniel J. Dougherty<sup>1</sup>, Chetan Gupta<sup>2</sup>, Song Wang<sup>2</sup>, Ismail Ari<sup>3</sup>, Abhay Mehta<sup>2</sup>*  
<sup>1</sup>Worcester Polytechnic Institute, USA; <sup>2</sup>HP Labs, USA; <sup>3</sup>Ozyegin University, Turkey
- 135 **Continuous Monitoring of Distance-Based Outliers Over Data Streams**  
*Maria Kontaki, Anastasios Gounaris, Apostolos N. Papadopoulos, Kostas Tsichlas, Yannis Manolopoulos, Aristotle University, Greece*
- 147 **Algorithms for Local Sensor Synchronization**  
*Lixing Wang, Yin Yang, Xin Miao, Dimitris Papadias, Yunhao Liu, Hong Kong University of Science & Technology, China*

---

## Research Session 4: Data Warehousing, OLAP and Data Grids

---

- 159 **Semi-Streamed Index Join for Near-Real Time Execution of ETL Transformations**  
*Mihaela A. Bornea<sup>1</sup>, Antonios Deligiannakis<sup>2</sup>, Yannis Kotidis<sup>1</sup>, Vasilis Vassalos<sup>1</sup>*  
<sup>1</sup>Athens University of Economics & Business, Greece; <sup>2</sup>Technical University of Crete, Greece
- 171 **Similarity Measures for Multidimensional Data**  
*Eftychia Baikousi, Georgios Rogkakos, Panos Vassiliadis, University of Ioannina, Greece*
- 183 **Distributed Cube Materialization on Holistic Measures**  
*Arnab Nandi<sup>1</sup>, Cong Yu<sup>2</sup>, Philip Bohannon<sup>3</sup>, Raghu Ramakrishnan<sup>3</sup>*  
<sup>1</sup>University of Michigan, USA; <sup>2</sup>Google Research, USA; <sup>3</sup>Yahoo! Labs, USA
- 195 **HyPer: A Hybrid OLTP&OLAP Main Memory Database System Based on Virtual Memory Snapshots**  
*Alfons Kemper, Thomas Neumann, Technische Universität München, Germany*

---

## Research Session 5: Data Mining and Knowledge Discovery I

---

- 207 **LTS: Discriminative Subgraph Mining by Learning from Search History**  
*Ning Jin, Wei Wang, University of North Carolina at Chapel Hill, USA*
- 219 **Active Learning Based Frequent Itemset Mining Over the Deep Web**  
*Tantan Liu, Gagan Agrawal, Ohio State University, USA*
- 231 **SystemML: Declarative Machine Learning on MapReduce**  
*Amol Ghoting<sup>1</sup>, Rajasekar Krishnamurthy<sup>2</sup>, Edwin Pednault<sup>1</sup>, Berthold Reinwald<sup>2</sup>, Vikas Sindhwani<sup>1</sup>, Shirish Tatikonda<sup>2</sup>, Yuanyuan Tian<sup>2</sup>, Shivakumar Vaithyanathan<sup>2</sup>*  
<sup>1</sup>IBM T.J. Watson Research Center, USA; <sup>2</sup>IBM Almaden Research Center, USA
- 243 **Mining Large Graphs: Algorithms, Inference, and Discoveries**  
*U Kang, Duen Horng Chau, Christos Faloutsos, Carnegie Mellon University, USA*

## Research Session 6: Distributed and Mobile Systems

---

- 255 **Accurate Latency Estimation in a Distributed Event Processing System**  
*Badrish Chandramouli<sup>1</sup>, Jonathan Goldstein<sup>2</sup>, Roger Barga<sup>1</sup>, Mirek Riedewald<sup>3</sup>, Ivo Santos<sup>4</sup>*  
<sup>1</sup>Microsoft Research, USA; <sup>2</sup>Microsoft Corporation, USA; <sup>3</sup>Northeastern University, USA; <sup>4</sup>Microsoft Research, Germany
- 267 **Subscriber Assignment for Wide-Area Content-Based Publish/Subscribe**  
*Albert Yu, Pankaj K. Agarwal, Jun Yang, Duke University, USA*
- 279 **Collaborative Caching for Spatial Queries in Mobile P2P Networks**  
*Qijun Zhu<sup>1</sup>, Dik Lun Lee<sup>1</sup>, Wang-Chien Lee<sup>2</sup>*  
<sup>1</sup>Hong Kong University of Science & Technology, China; <sup>2</sup>Pennsylvania State University, USA
- 291 **ES<sup>2</sup>: A Cloud Data Storage System for Supporting Both OLTP and OLAP**  
*Yu Cao<sup>1</sup>, Chun Chen<sup>2</sup>, Fei Guo<sup>1</sup>, Dawei Jiang<sup>1</sup>, Yuting Lin<sup>1</sup>, Beng Chin Ooi<sup>1</sup>, Hoang Tam Vo<sup>1</sup>, Sai Wu<sup>1</sup>, Quanqing Xu<sup>1</sup>*  
<sup>1</sup>National University of Singapore, Singapore; <sup>2</sup>Zhejiang University, China

## Research Session 7: Uncertain and Probabilistic Data

---

- 303 **Deriving Probabilistic Databases with Inference Ensembles**  
*Julia Stoyanovich<sup>1</sup>, Susan Davidson<sup>1</sup>, Tova Milo<sup>2</sup>, Val Tannen<sup>1</sup>*  
<sup>1</sup>University of Pennsylvania, USA; <sup>2</sup>Tel Aviv University, Israel
- 315 **Providing Support for Full Relational Algebra in Probabilistic Databases**  
*Robert Fink, Dan Olteanu, Swaroop Rath, University of Oxford, UK*
- 327 **Creating Probabilistic Databases from Imprecise Time-Series Data**  
*Saket Sathé, Hoyoung Jeung, Karl Aberer, EPFL, Switzerland*
- 339 **A Novel Probabilistic Pruning Approach to Speed up Similarity Queries in Uncertain Databases**  
*Thomas Bernecker<sup>1</sup>, Tobias Emrich<sup>1</sup>, Hans-Peter Kriegel<sup>1</sup>, Nikos Mamoulis<sup>2</sup>, Matthias Renz<sup>1</sup>, Andreas Züfle<sup>1</sup>*  
<sup>1</sup>LMU München, Germany; <sup>2</sup>University of Hong Kong, China

## Research Session 8: Query Processing and Optimization I

---

- 351 **Interactive SQL Query Suggestion: Making Databases User-Friendly**  
*Ju Fan, Guoliang Li, Lizhu Zhou, Tsinghua University, China*
- 363 **Computing Structural Statistics by Keywords in Databases**  
*Lu Qin, Jeffrey Xu Yu, Lijun Chang, Chinese University of Hong Kong, China*
- 375 **Program Transformations for Asynchronous Query Submission**  
*Mahendra Chavan<sup>1</sup>, Ravindra Guravannavar<sup>2</sup>, Karthik Ramachandra<sup>1</sup>, S. Sudarshan<sup>1</sup>*  
<sup>1</sup>IIT Bombay, India; <sup>2</sup>IIT Hyderabad, India
- 387 **Representative Skylines Using Threshold-Based Preference Distributions**  
*Atish Das Sarma<sup>1</sup>, Ashwin Lall<sup>2</sup>, Danupon Nanongkai<sup>3</sup>, Richard J. Lipton<sup>3</sup>, Jim Xu<sup>3</sup>*  
<sup>1</sup>Google Research, USA; <sup>2</sup>Denison University, USA; <sup>3</sup>Georgia Institute of Technology, USA

## Research Session 9: Outlier Processing

---

- 399 **Outlier Detection in Graph Streams**  
*Charu C. Aggarwal<sup>1</sup>, Yuchen Zhao<sup>2</sup>, Philip S. Yu<sup>2</sup>*  
<sup>1</sup>IBM T.J. Watson Research Center, USA; <sup>2</sup>University of Illinois at Chicago, USA
- 410 **Locality Sensitive Outlier Detection: A Ranking Driven Approach**  
*Ye Wang, Srinivasan Parthasarathy, Shirish Tatikonda, Ohio State University, USA*
- 422 **Outlier Detection on Uncertain Data: Objects, Instances, and Inferences**  
*Bin Jiang, Jian Pei, Simon Fraser University, Canada*
- 434 **Statistical Selection of Relevant Subspace Projections for Outlier Ranking**  
*Emmanuel Müller<sup>1</sup>, Matthias Schiffer<sup>2</sup>, Thomas Seidl<sup>2</sup>*  
<sup>1</sup>KIT, Germany; <sup>2</sup>RWTH Aachen, Germany

## Research Session 10: Data Integration, Metadata Management and Interoperability

---

- 446 **A Unified Model for Data and Constraint Repair**  
*Fei Chiang, Renée J. Miller, University of Toronto, Canada*
- 458 **Fast-Join: An Efficient Method for Fuzzy Token Matching Based String Similarity Join**  
*Jiannan Wang, Guoliang Li, Jianhua Fe, Tsinghua University, China*
- 470 **On Data Dependencies in Dataspace**  
*Shaoxu Song<sup>1</sup>, Lei Chen<sup>1</sup>, Philip S. Yu<sup>2</sup>*  
<sup>1</sup>Hong Kong University of Science & Technology, China; <sup>2</sup>University of Illinois at Chicago, USA
- 482 **Precisely Serializable Snapshot Isolation (PSSI)**  
*Stephen Revilak, Patrick O'Neil, Elizabeth O'Neil, University of Massachusetts at Boston, USA*

## Research Session 11: Privacy and Security

---

- 494 **MobiMix: Protecting Location Privacy with Mix-zones Over Road Networks**  
*Balaji Palanisamy, Ling Liu, Georgia Institute of Technology, USA*
- 506 **General Secure Sensor Aggregation in the Presence of Malicious Nodes**  
*Keith B. Frikken<sup>1</sup>, Kyle Kauffman<sup>2</sup>, Aaron Steele<sup>1</sup>*  
<sup>1</sup>Miami University, USA; <sup>2</sup>Air Force Institute of Technology, USA
- 517 **Secure and Efficient In-Network Processing of Exact SUM Queries**  
*Stavros Papadopoulos<sup>1</sup>, Aggelos Kiayias<sup>2</sup>, Dimitris Papadias<sup>3</sup>*  
<sup>1</sup>Chinese University of Hong Kong, China; <sup>2</sup>National & Kapodistrian University of Athens, Greece; <sup>3</sup>Hong Kong University of Science & Technology, China
- 529 **Preventing Equivalence Attacks in Updated, Anonymized Data**  
*Yeye He, Siddharth Barman, Jeffrey F. Naughton, University of Wisconsin-Madison, USA*

---

## Research Session 12: Temporal, Spatial and Multimedia Data

---

- 541 **Efficient Continuously Moving Top-K Spatial Keyword Query Processing**  
*Dingming Wu<sup>1</sup>, Man Lung Yiu<sup>2</sup>, Christian S. Jensen<sup>3</sup>, Gao Cong<sup>4</sup>*  
<sup>1</sup>Aalborg University, Denmark; <sup>2</sup>Hong Kong Polytechnic University, China; <sup>3</sup>Aarhus University, Denmark; <sup>4</sup>Nanyang Technological University, Singapore
- 553 **Large Scale Hamming Distance Query Processing**  
*Alex X. Liu, Ke Shen, Eric Torng, Michigan State University, USA*
- 565 **Authentication of Moving kNN Queries**  
*Man Lung Yiu, Eric Lo, Duncan Yung, Hong Kong Polytechnic University, China*
- 577 **Influence Zone: Efficiently Processing Reverse k Nearest Neighbors Queries**  
*Muhammad Aamir Cheema, Xuemin Lin, Wenjie Zhang, Ying Zhang, University of New South Wales, Australia*

---

## Research Session 13: Distributed Systems

---

- 589 **RAFTing MapReduce: Fast Recovery on the Raft**  
*Jorge-Arnulfo Quiané-Ruiz, Christoph Pinkel, Jörg Schad, Jens Dittrich, Universität des Saarlandes, Germany*
- 601 **Processing Private Queries Over Untrusted Data Cloud Through Privacy Homomorphism**  
*Haibo Hu, Jianliang Xu, Chushi Ren, Byron Choi, Hong Kong Baptist University, China*
- 613 **Real-Time Quantification and Classification of Consistency Anomalies in Multi-Tier Architectures**  
*Kamal Zellag, Bettina Kemme, McGill University, Canada*
- 625 **One-Copy Serializability with Snapshot Isolation Under the Hood**  
*Mihaela A. Bornea<sup>1</sup>, Orion Hodson<sup>2</sup>, Sameh Elnikety<sup>3</sup>, Alan Fekete<sup>4</sup>*  
<sup>1</sup>Athens University of Economics & Business, Greece; <sup>2</sup>Microsoft Research, UK; <sup>3</sup>Microsoft Research, USA; <sup>4</sup>University of Sydney, Australia

---

## Research Session 14: Semistructured Data, XML and Web Data Management

---

- 637 **Efficient Maintenance of Common Keys in Archives of Continuous Query Results from Deep Websites**  
*Fajar Ardian, Sourav S. Bhowmick, Nanyang Technological University, Singapore*
- 649 **How Schema Independent Are Schema Free Query Interfaces?**  
*Arash Termehchy, Marianne Winslett, Yodsawalai Chodpathumwan, University of Illinois at Urbana-Champaign, USA*
- 661 **XClean: Providing Valid Spelling Suggestions for XML Keyword Queries**  
*Yifei Lu<sup>1</sup>, Wei Wang<sup>1</sup>, Jianxin Li<sup>2</sup>, Chengfei Liu<sup>2</sup>*  
<sup>1</sup>University of New South Wales, Australia; <sup>2</sup>Swinburne University of Technology, Australia
- 673 **Top-k Keyword Search Over Probabilistic XML Data**  
*Jianxin Li<sup>1</sup>, Chengfei Liu<sup>1</sup>, Rui Zhou<sup>1</sup>, Wei Wang<sup>2</sup>*  
<sup>1</sup>Swinburne University of Technology, Australia; <sup>2</sup>University of New South Wales, Australia

---

## Research Session 15: Text, Uncertain and Probabilistic Data

---

- 685 **Selectivity Estimation for Extraction Operators Over Text Data**  
*Daisy Zhe Wang<sup>1</sup>, Long Wei<sup>1</sup>, Yunyao Li<sup>2</sup>, Frederick Reiss<sup>2</sup>, Shivakumar Vaithyanathan<sup>2</sup>*  
<sup>1</sup>University of California at Berkeley, USA; <sup>2</sup>IBM Almaden Research Center, USA
- 697 **Join Queries on Uncertain Data: Semantics and Efficient Processing**  
*Tingjian Ge, University of Kentucky, USA*
- 709 **Interval-Based Pruning for Top-k Processing Over Compressed Lists**  
*Kaushik Chakrabarti<sup>1</sup>, Surajit Chaudhuri<sup>1</sup>, Venkatesh Ganti<sup>2</sup>*  
<sup>1</sup>Microsoft Research, USA; <sup>2</sup>Google Inc., USA
- 721 **Stochastic Skyline Operator**  
*Xuemin Lin, Ying Zhang, Wenjie Zhang, Muhammad Aamir Cheema, University of New South Wales, Australia*

---

## Research Session 16: Data Mining and Knowledge Discovery II

---

- 733 **Discovery of Complex Glitch Patterns: A Novel Approach to Quantitative Data Cleaning**  
*Laure Berti-Équille<sup>1</sup>, Tamraparni Dasu<sup>2</sup>, Divesh Srivastava<sup>2</sup>*  
<sup>1</sup>University of Rennes 1, France; <sup>2</sup>AT&T Labs Research, USA
- 745 **Towards Exploratory Hypothesis Testing and Analysis**  
*Guimei Liu<sup>1</sup>, Mengling Feng<sup>2</sup>, Yue Wang<sup>1</sup>, Limsoon Wong<sup>1</sup>, See-Kiong Ng<sup>2</sup>, Tzia Liang Mah<sup>2</sup>, Edmund Jon Deoon Lee<sup>1</sup>*  
<sup>1</sup>National University of Singapore, Singapore; <sup>2</sup>Institute for Infocomm Research, Singapore
- 757 **SMM: A Data Stream Management System for Knowledge Discovery**  
*Hetal Thakkar<sup>1</sup>, Nikolay Laptev<sup>2</sup>, Hamid Mousavi<sup>2</sup>, Barzan Mozafari<sup>2</sup>, Vincenzo Russo<sup>3</sup>, Carlo Zaniolo<sup>2</sup>*  
<sup>1</sup>Google Inc., USA; <sup>2</sup>University of California at Los Angeles, USA; <sup>3</sup>CNR-ICAR, Italy
- 769 **Knowledge Transfer with Low-Quality Data: A Feature Extraction Issue**  
*Brian Quanz, Jun Huan, Meenakshi Mishra, University of Kansas, USA*

---

## Research Session 17: Database User Interfaces and Information Visualization

---

- 780 **EdiFlow: Data-Intensive Interactive Workflows for Visual Analytics**  
*Véronique Benzaken<sup>1</sup>, Jean-Daniel Fekete<sup>2</sup>, Pierre-Luc Hémerly<sup>2</sup>, Wael Khemiri<sup>2</sup>, Ioana Manolescu<sup>2</sup>*  
<sup>1</sup>LRI, France; <sup>2</sup>INRIA, France
- 792 **A Continuous Query System for Dynamic Route Planning**  
*Nirmesh Malviya<sup>1</sup>, Samuel Madden<sup>1</sup>, Arnab Bhattacharya<sup>2</sup>*  
<sup>1</sup>MIT, USA; <sup>2</sup>IIT Kanpur, India
- 804 **Optimal Location Queries in Road Network Databases**  
*Xiaokui Xiao<sup>1</sup>, Bin Yao<sup>2</sup>, Feifei Li<sup>2</sup>*  
<sup>1</sup>Nanyang Technological University, Singapore; <sup>2</sup>Florida State University, USA
- 816 **Spatio-Temporal Joins on Symbolic Indoor Tracking Data**  
*Hua Lu<sup>1</sup>, Bin Yang<sup>2</sup>, Christian S. Jensen<sup>3</sup>*  
<sup>1</sup>Aalborg University, Denmark; <sup>2</sup>MPII, Germany; <sup>3</sup>Aarhus University, Denmark

---

## Research Session 18: Query Processing and Optimization II

---

- 828     **MaxFirst for MaxBRkNN**  
*Zenan Zhou<sup>1</sup>, Wei Wu<sup>2</sup>, Xiaohui Li<sup>1</sup>, Mong Li Lee<sup>1</sup>, Wynne Hsu<sup>1</sup>*  
<sup>1</sup>National University of Singapore, Singapore; <sup>2</sup>Institute for Infocomm Research, Singapore
- 840     **SQPR: Stream Query Planning with Reuse**  
*Evangelia Kalyvianaki<sup>1</sup>, Wolfram Wiesemann<sup>1</sup>, Quang Hieu Vu<sup>2</sup>, Daniel Kuhn<sup>1</sup>, Peter Pietzuch<sup>1</sup>*  
<sup>1</sup>Imperial College London, UK; <sup>2</sup>Institute for Infocomm Research, Singapore
- 852     **Memory-Constrained Aggregate Computation Over Data Streams**  
*K.V.M. Naidu<sup>1</sup>, Rajeev Rastogi<sup>1</sup>, Scott Satkin<sup>2</sup>, Anand Srinivasan<sup>3</sup>*  
<sup>1</sup>Yahoo! Labs, India; <sup>2</sup>Carnegie Mellon University, USA; <sup>3</sup>Google Inc., India
- 864     **A New, Highly Efficient, and Easy To Implement Top-Down Join Enumeration Algorithm**  
*Pit Fender, Guido Moerkotte, Universität Mannheim, Germany*

---

## Research Session 19: Storage

---

- 876     **Transactional In-Page Logging for Multiversion Read Consistency and Recovery**  
*Sang-Won Lee<sup>1</sup>, Bongki Moon<sup>2</sup>*  
<sup>1</sup>Sungkyunkwan University, Korea; <sup>2</sup>University of Arizona, USA
- 888     **Answering Approximate String Queries on Large Data Sets Using External Memory**  
*Alexander Behm, Chen Li, Michael J. Carey, University of California at Irvine, USA*
- 900     **Discovering Popular Routes from Trajectories**  
*Zaiben Chen, Heng Tao Shen, Xiaofang Zhou, University of Queensland, Australia*

---

## Research Session 20: Privacy and Scientific

---

- 912     **Spectrum Based Fraud Detection in Social Networks**  
*Xiaowei Ying<sup>1</sup>, Xintao Wu<sup>1</sup>, Daniel Barbará<sup>2</sup>*  
<sup>1</sup>University of North Carolina at Charlotte, USA; <sup>2</sup>George Mason University, USA
- 924     **Identity Obfuscation in Graphs Through the Information Theoretic Lens**  
*Francesco Bonchi<sup>1</sup>, Aristides Gionis<sup>1</sup>, Tamir Tassa<sup>2</sup>*  
<sup>1</sup>Yahoo! Research, Spain; <sup>2</sup>Open University, Israel
- 936     **Monte Carlo Query Processing of Uncertain Multidimensional Array Data**  
*Tingjian Ge<sup>1</sup>, David Grabiner<sup>2</sup>, Stan Zdonik<sup>2</sup>*  
<sup>1</sup>University of Kentucky, USA; <sup>2</sup>Brown University, USA



---

## Research Session 21: XML and RDF

---

- 948 **Massively Parallel XML Twig Filtering Using Dynamic Programming on FPGAs**  
*Roger Moussalli, Mariam Salloum, Walid Najjar, Vassilis J. Tsotras, University of California at Riverside, USA*
- 960 **Selectivity Estimation of Twig Queries on Cyclic Graphs**  
*Yun Peng, Byron Choi, Jianliang Xu, Hong Kong Baptist University, China*
- 972 **Efficient XQuery Rewriting Using Multiple Views**  
*Ioana Manolescu<sup>1</sup>, Konstantinos Karanasos<sup>1</sup>, Vasilis Vassalos<sup>2</sup>, Spyros Zoupanos<sup>1</sup>*  
*<sup>1</sup>INRIA, France; <sup>2</sup>Athens University of Economics & Business, Greece*
- 984 **Characteristic Sets: Accurate Cardinality Estimation for RDF Queries with Multiple Joins**  
*Thomas Neumann<sup>1</sup>, Guido Moerkotte<sup>2</sup>*  
*<sup>1</sup>Technische Universität München, Germany; <sup>2</sup>Universität Mannheim, Germany*

---

## Research Session 22: Query Processing and Optimization III

---

- 995 **PrefJoin: An Efficient Preference-Aware Join Operator**  
*Mohamed E. Khalefa, Mohamed F. Mokbel, Justin J. Levandoski, University of Minnesota, USA*
- 1007 **Decomposing DAGs into Spanning Trees: A New Way to Compress Transitive Closures**  
*Yangjun Chen, Yibin Chen, University of Winnipeg, Canada*
- 1019 **Preference Queries Over Sets**  
*Xi Zhang, Jan Chomicki, University at Buffalo, USA*
- 1031 **A Unified Approach for Computing Top-k Pairs in Multidimensional Space**  
*Muhammad Aamir Cheema<sup>1</sup>, Xuemin Lin<sup>1</sup>, Haixun Wang<sup>2</sup>, Jianmin Wang<sup>3</sup>, Wenjie Zhang<sup>1</sup>*  
*<sup>1</sup>University of New South Wales, Australia; <sup>2</sup>Microsoft Research Asia, China; <sup>3</sup>Tsinghua University, China*

---

## Research Session 23: Data Mining and Knowledge Discovery III

---

- 1043 **Bidirectional Mining of Non-Redundant Recurrent Rules from a Sequence Database**  
*David Lo<sup>1</sup>, Bolin Ding<sup>2</sup>, Lucia<sup>1</sup>, Jiawei Han<sup>2</sup>*  
*<sup>1</sup>Singapore Management University, Singapore; <sup>2</sup>University of Illinois at Urbana-Champaign, USA*
- 1055 **Finding Top-k Profitable Products**  
*Qian Wan, Raymond Chi-Wing Wong, Yu Peng, Hong Kong University of Science & Technology, China*
- 1067 **Efficient Spectral Neighborhood Blocking for Entity Resolution**  
*Liangcai Shu<sup>1</sup>, Aiyou Chen<sup>2</sup>, Ming Xiong<sup>3</sup>, Weiyi Meng<sup>1</sup>*  
*<sup>1</sup>Binghamton University, USA; <sup>2</sup>Bell Labs, USA; <sup>3</sup>Google Inc., USA*
- 1079 **Consensus Spectral Clustering in Near-Linear Time**  
*Dijun Luo, Chris Ding, Heng Huang, Feiping Nie, University of Texas at Arlington, USA*

---

## Research Session 24: Indexing

---

- 1091 **On Dimensionality Reduction of Massive Graphs for Indexing and Retrieval**  
*Charu C. Aggarwal<sup>1</sup>, Haixun Wang<sup>2</sup>*  
<sup>1</sup>IBM T.J. Watson Research Center, USA; <sup>2</sup>Microsoft Research Asia, China
- 1103 **HashFile: An Efficient Index Structure for Multimedia Data**  
*Dongxiang Zhang<sup>1</sup>, Divyakant Agrawal<sup>2</sup>, Gang Chen<sup>3</sup>, Anthony K.H. Tung<sup>1</sup>*  
<sup>1</sup>National University of Singapore, Singapore; <sup>2</sup>University of California at Santa Barbara, USA; <sup>3</sup>Zhejiang University, China
- 1115 **CT-Index: Fingerprint-Based Graph Indexing Combining Cycles and Trees**  
*Karsten Klein, Nils Kriege, Petra Mutzel, Technische Universität Dortmund, Germany*
- 1127 **Partitioning Techniques for Fine-Grained Indexing**  
*Eugene Wu, Samuel Madden, MIT, USA*

---

## Research Session 25: Systems, Experiments, Applications

---

- 1139 **Jackpine: A Benchmark to Evaluate Spatial Database Performance**  
*Suprio Ray, Bogdan Simion, Angela Demke Brown, University of Toronto, Canada*
- 1151 **Hyracks: A Flexible and Extensible Foundation for Data-Intensive Computing**  
*Vinayak Borkar, Michael J. Carey, Raman Grover, Nicola Onose, Rares Vernica, University of California at Irvine, USA*
- 1163 **On Query Result Diversification**  
*Marcos R. Vieira<sup>1</sup>, Humberto L. Razente<sup>2</sup>, Maria C.N. Barioni<sup>2</sup>, Marios Hadjieleftheriou<sup>3</sup>, Divesh Srivastava<sup>3</sup>, Caetano Traina Jr.<sup>4</sup>, Vassilis J. Tsotras<sup>1</sup>*  
<sup>1</sup>University of California at Riverside, USA; <sup>2</sup>Universidade Federal do ABC, Brazil; <sup>3</sup>AT&T Labs Research, USA; <sup>4</sup>Universidade de São Paulo, Brazil
- 1175 **Generating Test Data for Killing SQL Mutants: A Constraint-Based Approach**  
*Shetal Shah, S. Sudarshan, Suhas Kajbaje, Sandeep Patidar, Bhanu Pratap Gupta, Devang Vira, IIT Bombay, India*

---

## Industry Session 1: Web-Information Management

---

- (NA) **Google Product Search: An End-to-End View from Content Acquisition to Serving User Requests**  
*Thomas Hofmann, Google, USA*
- (NA) **Project Voldemort: Reliable Distributed Storage**  
*Alex Feinberg, LinkedIn, USA*

---

## Industry Session 2: Data Warehousing and Business Intelligence

---

- 1187 **Implementing Sentinels in the TARGIT BI Suite**  
*Morten Middelfart<sup>1</sup>, Torben Bach Pedersen<sup>2</sup>*  
<sup>1</sup>TARGIT, Denmark; <sup>2</sup>Aalborg University, Denmark
- 1199 **RCFile: A Fast and Space-Efficient Data Placement Structure in MapReduce-Based Warehouse Systems**  
*Yongqiang He<sup>1</sup>, Rubao Lee<sup>2</sup>, Yin Huai<sup>2</sup>, Zheng Shao<sup>1</sup>, Namit Jain<sup>1</sup>, Xiaodong Zhang<sup>2</sup>, Zhiwei Xu<sup>3</sup>*  
<sup>1</sup>Facebook, USA; <sup>2</sup>Ohio State University, USA; <sup>3</sup>Chinese Academy of Sciences, China
- 1209 **Web-Scale Information Extraction with Vertex**  
*Pankaj Gulhane<sup>1</sup>, Amit Madaan<sup>1</sup>, Rupesh Mehta<sup>2</sup>, Jeyashankher Ramamirtham<sup>1</sup>, Rajeev Rastogi<sup>1</sup>, Sandeep Satpal<sup>2</sup>, Srinivasan H. Sengamedu<sup>1</sup>, Ashwin Tengli<sup>2</sup>, Charu Tiwari<sup>1</sup>*  
<sup>1</sup>Yahoo! Labs, India; <sup>2</sup>Microsoft IDC, India

---

## Industry Session 3: Database Technology

---

- 1221 **High Performance Database Logging Using Storage Class Memory**  
*Ru Fang<sup>1</sup>, Hui-I Hsiao<sup>1</sup>, Bin He<sup>1</sup>, C. Mohan<sup>1</sup>, Yun Wang<sup>2</sup>*  
<sup>1</sup>IBM Almaden Research Center, USA; <sup>2</sup>IBM China Research Laboratory, China
- 1232 **Dynamic Prioritization of Database Queries**  
*Sivaramkrishnan Narayanan, Florian Waas, Greenplum Inc., USA*
- 1242 **The Extensibility Framework in Microsoft StreamInsight**  
*Mohamed Ali<sup>1</sup>, Badrish Chandramouli<sup>2</sup>, Jonathan Goldstein<sup>1</sup>, Roman Schindlauer<sup>1</sup>*  
<sup>1</sup>Microsoft Corporation, USA; <sup>2</sup>Microsoft Research, USA

---

## Industry Session 4: Cloud Computing

---

- 1254 **Relational Databases, Virtualization, and the Cloud**  
*Maximilian Ahrens<sup>1</sup>, Gustavo Alonso<sup>2</sup>*  
<sup>1</sup>Zimory GmbH, Germany; <sup>2</sup>ETH Zürich, Switzerland
- 1255 **Adapting Microsoft SQL Server for Cloud Computing**  
*Philip A. Bernstein, Istvan Cseri, Nishant Dani, Nigel Ellis, Ajay Kalhan, Gopal Kakivaya, David B. Lomet, Ramesh Manne, Lev Novik, Tomas Talius, Microsoft Corporation, USA*
- 1264 **Predicting In-Memory Database Performance for Automating Cluster Management Tasks**  
*Jan Schaffner<sup>1</sup>, Benjamin Eckart<sup>1</sup>, Dean Jacobs<sup>2</sup>, Christian Schwarz<sup>1</sup>, Hasso Plattner<sup>1</sup>, Alexander Zeier<sup>1</sup>*  
<sup>1</sup>Hasso Plattner Institute, Germany; <sup>2</sup>SAP AG, Germany

Demo 1A and 1B

---

- 1276 **ATOM: Automatic Target-Driven Ontology Merging**  
*Salvatore Raunich, Erhard Rahm, Universität Leipzig, Germany*
- 1280 **Automatic Generation of Mediated Schemas Through Reasoning Over Data Dependencies**  
*Xiang Li, Christoph Quix, David Kensche, Sandra Geisler, Lisong Guo, RWTH Aachen, Germany*
- 1284 **DBridge: A Program Rewrite Tool for Set-Oriented Query Execution**  
*Mahendra Chavan<sup>1</sup>, Ravindra Guravannavar<sup>2</sup>, Karthik Ramachandra<sup>1</sup>, S. Sudarshan<sup>1</sup>*  
*<sup>1</sup>IIT Bombay, India; <sup>2</sup>IIT Hyderabad, India*
- 1288 **SmartTrace: Finding Similar Trajectories in Smartphone Networks Without Disclosing the Traces**  
*Costandinos Costa<sup>1</sup>, Christos Laoudias<sup>1</sup>, Demetrios Zeinalipour-Yazti<sup>1</sup>, Dimitrios Gunopoulos<sup>2</sup>*  
*<sup>1</sup>University of Cyprus, Cyprus; <sup>2</sup>University of Athens, Greece*
- 1292 **Real-Time Pattern Matching with FPGAs**  
*Louis Woods, Jens Teubner, Gustavo Alonso, ETH Zürich, Switzerland*
- 1296 **Scientific Workflow Design 2.0: Demonstrating Streaming Data Collections in Kepler**  
*Lei Dou<sup>1</sup>, Daniel Zinn<sup>1</sup>, Timothy McPhillips<sup>1</sup>, Sven Köhler<sup>1</sup>, Sean Riddle<sup>1</sup>, Shawn Bowers<sup>2</sup>, Bertram Ludäscher<sup>1</sup>*  
*<sup>1</sup>University of California at Davis, USA; <sup>2</sup>Gonzaga University, USA*
- 1300 **Social Networking on Top of the WebdamExchange System**  
*Émilien Antoine<sup>1</sup>, Alban Galland<sup>1</sup>, Kristian Lyngbaek<sup>1</sup>, Amélie Marian<sup>2</sup>, Neokis Polyzotis<sup>3</sup>*  
*<sup>1</sup>INRIA, France; <sup>2</sup>Rutgers University, USA; <sup>3</sup>University of California at Santa Cruz, USA*

- 1304 **AMC — A Framework for Modelling and Comparing Matching Systems as Matching Processes**  
*Eric Peukert<sup>1</sup>, Julian Eberius<sup>1</sup>, Erhard Rahm<sup>2</sup>*  
*<sup>1</sup>SAP Research, Germany; <sup>2</sup>Universität Leipzig, Germany*
- 1308 **Using Markov Chain Monte Carlo to Play Trivia**  
*Daniel Deutch, Ohad Greenshpan, Boris Kostenko, Tova Milo, Tel Aviv University, Israel*
- 1312 **DiRec: Diversified Recommendations for Semantic-Less Collaborative Filtering**  
*Rubi Boim, Tova Milo, Slava Novgorodov, Tel Aviv University, Israel*
- 1316 **SkyEngine: Efficient Skyline Search Engine for Continuous Skyline Computations**  
*Yu-Ling Hsueh<sup>1</sup>, Roger Zimmermann<sup>2</sup>, Wei-Shinn Ku<sup>3</sup>, Yifan Jin<sup>4</sup>*  
*<sup>1</sup>Teradata, USA; <sup>2</sup>National University of Singapore, Singapore; <sup>3</sup>Auburn University, USA; <sup>4</sup>University of Hong Kong, China*

## Demo 2A and 2B

---

- 1324 **Ruby on Semantic Web**  
*Vadim Eisenberg, Yaron Kanza, Technion, Israel*
- 1328 **Preference-Based Datacube Analysis with MYOLAP**  
*Paolo Biondi, Matteo Golfarelli, Stefano Rizzi, Università di Bologna, Italy*
- 1332 **Continuous, Online Monitoring and Analysis in Large Water Distribution Networks**  
*Xiuli Ma<sup>1</sup>, Hongmei Xiao<sup>1</sup>, Shuiyuan Xie<sup>1</sup>, Qiong Li<sup>1</sup>, Qiong Luo<sup>2</sup>, Chunhua Tian<sup>3</sup>*  
*<sup>1</sup>Peking University, China; <sup>2</sup>Hong Kong University of Science & Technology, China; <sup>3</sup>IBM Research, China*
- 1336 **Integrating Code Search into the Development Session**  
*Mu-Woong Lee<sup>1</sup>, Seung-won Hwang<sup>1</sup>, Sunghun Kim<sup>2</sup>*  
*<sup>1</sup>POSTECH, Korea; <sup>2</sup>Hong Kong University of Science & Technology, China*
- 1340 **The Proactive Promotion Engine**  
*Karen Works, Elke A. Rundensteiner, Worcester Polytechnic Institute, USA*
- 1344 **NORMS: An Automatic Tool to Perform Schema Label Normalization**  
*Serena Sorrentino, Sonia Bergamaschi, Maciej Gawinecki, Università di Modena e Reggio Emilia, Italy*
- 1348 **GuideMe! — The World of Sights in Your Pocket**  
*Sergej Zerr, Kerstin Bischoff, Sergey Chernov, L3S Research Center, Germany*
- 1352 **CompRec-Trip: A Composite Recommendation System for Travel Planning**  
*Min Xie<sup>1</sup>, Laks V.S. Lakshmanan<sup>1</sup>, Peter T. Wood<sup>2</sup>*  
*<sup>1</sup>University of British Columbia, Canada; <sup>2</sup>Birkbeck University of London, UK*
- 1356 **Advanced Search, Visualization and Tagging of Sensor Metadata**  
*Ioannis Paparrizos, Hoyoung Jeung, Karl Aberer, EPFL, Switzerland*

## Panels

---

- 1360 **Distributed Data Management in 2020?**  
*M. Tamer Özsu<sup>1</sup>, Patrick Valduriez<sup>2</sup>, Serge Abiteboul<sup>2</sup>, Bettina Kemme<sup>3</sup>, Ricardo Jiménez-Péris<sup>4</sup>, Beng Chin Ooi<sup>5</sup>*  
*<sup>1</sup>University of Waterloo, Canada; <sup>2</sup>INRIA, France; <sup>3</sup>McGill University, Canada; <sup>4</sup>Universidad Politécnica de Madrid, Spain; <sup>5</sup>National University of Singapore, Singapore*
- 1361 **Robust Query Processing**  
*Goetz Graefe, HP Labs, USA*

## Seminars

---

- 1362 **Non-Metric Similarity Search Problems in Very Large Collections**  
*Benjamin Bustos<sup>1</sup>, Tomáš Skopal<sup>2</sup>*  
<sup>1</sup>University of Chile, Chile; <sup>2</sup>Charles University in Prague, Czech Republic
- 1366 **Next Generation Data Integration for Life Sciences**  
*Sarah Cohen-Boulakia<sup>1</sup>, Ulf Leser<sup>2</sup>*  
<sup>1</sup>Université Paris-Sud 11, France; <sup>2</sup>Humboldt-Universität zu Berlin, Germany
- 1370 **Modern B-Tree Techniques**  
*Goetz Graefe, Harumi Kuno, HP Labs, USA*
- 1374 **Query Optimizer Plan Diagrams: Production, Reduction and Applications**  
*Jayant R. Haritsa, Indian Institute of Science, India*
- 1378 **Schemas for Safe and Efficient XML Processing**  
*Dario Colazzo<sup>1</sup>, Giorgio Ghelli<sup>2</sup>, Carlo Sartiani<sup>3</sup>*  
<sup>1</sup>Université Paris-Sud 11, France; <sup>2</sup>Università di Pisa, Italy; <sup>3</sup>Università della Basilicata, Italy
- 1380 **Keyword-Based Search and Exploration on Databases**  
*Yi Chen<sup>1</sup>, Wei Wang<sup>2</sup>, Ziyang Liu<sup>1</sup>*  
<sup>1</sup>Arizona State University, USA; <sup>2</sup>University of New South Wales, Australia