

VLDB2014

40th International Conference on Very Large Data Bases, Hangzhou, China



Proceedings of the VLDB Endowment

Volume 7, No. 9 – May 2014

**Proceedings of the 40th International Conference on
Very Large Data Bases, Hangzhou, China**

Program Chairs and Editors-in-Chief:

H. V. Jagadish, Aoying Zhou

Associate Editors – Research and Innovative Systems Tracks:

**Shivnath Babu, Lei Chen, Graham Cormode, Bin Cui, Wynne Hsu, Martin Kersten,
Donald Kossman, Elke Rundensteiner, Kyuseok Shim, Wang-Chiew Tan, Letizia Tanca, Jeffrey Yu**

Associate Editors – Experiments and Analysis Track:

Gao Cong, Jens Dittrich

Associate Editors – Vision Track:

Zachary Ives

Proceedings Chairs:

Li Xiong, Cong Yu

PVLDB – Proceedings of the VLDB Endowment

Volume 7, No. 9, May 2014.

The 40th International Conference on Very Large Data Bases, Hangzhou, China.

Copyright 2014 VLDB Endowment

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>. Obtain permission prior to any use beyond those covered by the license. Contact copyright holder by emailing info@vldb.org.

Volume 7, Number 9, May 2014: VLDB 2014

Pages ii - x and 661 - 784

ISSN 2150-8097

Additional copies only online at: portal.acm.org, arxiv.org/corr, and www.vldb.org

TABLE OF CONTENTS

Front Matter

Copyright Notice	ii
Table of Contents	iii
VLDB 2014 Organization and Review Board	iv

Letters

What to Expect from a Revision	<i>H. V. Jagadish</i>	ix
--------------------------------------	-----------------------	----

Research Papers

PRESS: A Novel Framework of Trajectory Compression in Road Networks.....	661
..... <i>Renchu Song, Weiwei Sun, Baihua Zheng, Yu Zheng</i>	
Finding the Cost-Optimal Path with Time Constraint over Time-Dependent Graphs.....	673
..... <i>Yajun Yang, Hong Gao, Jeffrey Xu Yu, Jianzhong Li</i>	
Optimal Crowd-Powered Rating and Filtering Algorithms.....	685
..... <i>Aditya Parameswaran, Stephen Boyd, Hector Garcia-Molina,</i>	
..... <i>Ashish Gupta, Neoklis Polyzotis, Jennifer Widom</i>	
Incremental Record Linkage.....	697
..... <i>Anja Gruenheid, Xin Luna Dong, Divesh Srivastava</i>	
Low-Latency Handshake Join	709
..... <i>Pratanu Roy, Jens Teubner, Rainer Gemulla</i>	
Path Problems in Temporal Graphs.....	721
..... <i>Huanhuan Wu, James Cheng, Silu Huang, Yiping Ke, Yi Lu, Yanyan Xu</i>	
Retrieving Regions of Interest for User Exploration.....	733
..... <i>Xin Cao, Gao Cong, Christian S. Jensen, Man Lung Yiu</i>	
SK-LSH: An Efficient Index Structure for Approximate Nearest Neighbor Search.....	745
..... <i>Yingfan Liu, Jiangtao Cui, Zi Huang, Hui Li, Heng Tao Shen</i>	
On Arbitrage-free Pricing for General Data Queries.....	757
..... <i>Bing-Rong Lin, Daniel Kifer</i>	
Splitter: Mining Fine-Grained Sequential Patterns in Semantic Trajectories.....	769
..... <i>Chao Zhang, Jiawei Han, Lidan Shou, Jiajun Lu, Thomas La Porta</i>	
Towards Building Wind Tunnels for Data Center Design.....	781
..... <i>Avrilia Floratou, Frank Bertsch, Jignesh M. Patel, Georgios Laskaris</i>	

VLDB 2014 ORGANIZATION AND REVIEW BOARD

Honorary Chair

Yunhe Pan, Chinese Academy of Engineering

General Chairs

Chun Chen, Zhejiang University

Sharad Mehrotra, University of California, Irvine

Program Chairs and Editors-in-Chief of PVLDB 7

H. V. Jagadish, University of Michigan

Aoying Zhou, East Normal University, China

Research and Innovative Systems Tracks Associate Editors

Shivnath Babu, Duke University

Lei Chen, Hong Kong University of Science and Technology

Graham Cormode, University of Warwick

Bin Cui, Peking University, China

Wynne Hsu, NUS

Martin Kersten, CWI

Donald Kossman, ETH

Elke Rundensteiner, WPI

Kyuseok Shim, Seoul National University

Wang-Chiew Tan, University of California, Santa Cruz

Letizia Tanca, Poli Milano

Jeffrey Yu, Chinese University of Hong Kong

Experiments and Analysis Track Associate Editors

Gao Cong, Nanyang Technology University

Jens Dittrich, Saarland

Visions Track Associate Editor

Zachary Ives, University of Pennsylvania

Industrial and Applications Track Associate Editors

Umeshwar Dayal, HP

C. Mohan, IBM

Ge Yu, Northeastern University, China

Demonstration Chairs

Mong-Li Lee, NUS
Feifei Li, University of Utah
Sunil Prabhakar, Purdue

Tutorial Chairs

Xiaoyong Du, Renmin University
Murat Kantarcioglu, University of Texas, Dallas
Divesh Srivastava, AT&T Labs

Research Track Review Board

Sibel Adali, Rensselaer Polytechnic Institute
Foto Afrati, NTU Athens
Yanif Ahmad, JHU
Jose Luis Ambite, ISI - USC
Walid Aref, Purdue University
Claudia Bauzer Medeiros, University of Campinas
Srikanta Bedathur, IIIT Delhi
Michael Benedikt, Oxford University
Sonia Bergamaschi, Universita Modena
Laure Berti-Equille, IRD, France
Leopoldo Bertossi, Carleton University, Ottawa
Subhash Bhalla, University of Aizu, Japan
Peter Boncz, CWI
Angela Bonifati, University of Lille 1
Rajesh Bordawekar, IBM Watson Research Center
Omar Boucelma, Aix-Marseille University
Nico Bruno, Microsoft Research
Andrea Cali, University of London, Birkbeck College
Malu Castellanos, HP Labs
Badrish Chandramouli, Microsoft Research
Adriane Chapman, Mitre
Gang Chen, Zhejiang University
Yi Chen, New Jersey Institute of Technology
James Cheng, CUHK
Reynold Cheng, University of Hong Kong
Brian Cooper, Google, USA

Workshop Chairs

Anastasia Ailamaki, EPFL
Kaushik Chakrabarti, Microsoft

Panel Chairs

Hakan Hacigumus, NEC Labs
Jignesh Patel, University of Wisconsin
Xiaoyang Sean Wang, Fudan University

Philippe Cudré-Mauroux, University of Fribourg
Carlo Curino, MIT
Gautam Das, UT Arlington and QCRI
Sudipto Das, Microsoft Research
Anish Das Sarma
Atish Das Sarma, eBay Research Labs
Khuzaima Daudjee, University of Waterloo
Antonios Deligiannakis, Technical University of Crete
Daniel Deutch, Ben Gurion University
Yanlei Diao, University of Massachusetts Amherst
Xin (Luna) Dong, Google, USA
Sameh Elnikety, Microsoft Research
Mohamed Eltabakh, Worcester Polytechnic Institute
Ihab F. Ilyas, QCRI
Hakan Ferhatosmanoglu, Bilkent University
Ada Wai-Chee Fu, Chinese University of Hong Kong
Minos Garofalakis, Technical University of Crete
Wolfgang Gatterbauer, Carnegie Mellon University
Tingjian Ge, University of Massachusetts, Lowell
Buğra Gedik, Bilkent University
Rainer Gemulla, Max-Planck-Institut Saarbrücken
Gabriel Ghinita, University of Massachusetts Boston
Parke Godfrey, York University
Lukasz Golab, University of Waterloo
Sergio Greco, University of Calabria
Le Gruenwald, University of Oklahoma

Giovanna Guerrini, Universita Genova
 Krishna Gummadi, MPI-SWS
 Rahul Gupta, Google Research
 Rajeev Gupta, IBM Research
 Shyam Gupta, IIT Delhi
 Marios Hadjieleftheriou, AT&T labs
 Wook-Shin Han, KNU, Korea
 Kuno Harumi, HP Labs
 Bingsheng He, NTU Singapore
 Sven Helmer, Free University of Bozen-Bolzano
 Jan Hidders, TUDelft
 Wei Hong, Cisco System Inc.
 Katja Hose, Aalborg University
 Zi Huang, University of Queensland
 Jeong-Hyon Hwang, SUNY - Albany
 Seung-won Hwang, POSTECH, Korea
 Stratos Idreos, CWI
 Yoshiharu Ishikawa, Nagoya University
 Zachary Ives, University of Pennsylvania
 Ricardo Jimenez-Peris, Technical University of Madrid
 Cheqing Jin, East China Normal University
 Ruoming Jin, Kent State University
 Alekh Jindal, Saarland University/MIT
 Ryan Johnson, University of Toronto
 Dmitri V Kalashnikov, UC Irvine
 Panos Kalnis, KAUST, Saudi Arabia
 Ben Kao, Hong Kong University
 Panagiotis Karras, Rutgers University
 Yiping Ke, Institute of High Performance Computing
 Bettina Kemme, McGill University
 Daniel Kifer, PSU
 Benny Kimelfeld, IBM
 Hideaki Kimura, Microsoft Jim Gray Systems Lab
 George Kollios, Boston University
 Christian König, Microsoft Research
 Tim Kraska, Brown University
 Laks V. S. Lakshmanan, University of British Columbia
 Mounia Lalmas, Yahoo Inc.
 Mong-Li Lee, National University of Singapore
 Wolfgang Lehner, Technische University Dresden
 Justin Levandoski, Microsoft Research
 Chengkai Li, The University of Texas at Arlington
 Cuiping Li, Renmin University of China
 Feifei Li, University of Utah
 Guoliang Li, Tsinghua University
 Jianzhong Li, Harbin Institute of Technology
 Yunyao Li, IBM Almaden
 Zhanhuai Li, Northwestern Polytechnical University
 Dan Lin, Missouri S&T, USA
 Xuemin Lin, University of New South Wales
 Bin Liu, NEC Labs America
 Ziyang Liu, NEC Labs America
 Eric Lo, The Hong Kong Polytechnic University
 Qiong Luo, HKUST
 Shuai Ma, Beihang University
 Ashwin Machanavajjhala, Duke University
 Brad Malin, Duke University
 Nikos Mamoulis, University of Hong Kong
 Stefan Manegold, CWI
 Murali Mani, University of Michigan
 Ioana Manolescu, INRIA, France
 Amélie Marian, Rutgers University
 Volker Markl, TU Berlin
 Marta Mattoso, Federal University of Rio de Janeiro
 Frank McSherry, Microsoft
 Alexandra Meliou, Umass Amherst
 Marco Mesiti, University of Milano
 Dan Miranker, The University of Texas at Austin
 Mohamed Mokbel, University of Minnesota
 Bongki Moon, Seoul National University
 Yasuhiko Morimoto, Hiroshima University
 Mirella Moro, Universidade Federal de Minas Gerais
 Kyriakos Mouratidis, SMU, Singapore
 Karin Murthy, IBM India
 Arnab Nandi, Ohio State University
 Wolfgang Nejdl, University of Hannover

Thomas Neumann, Technology University Munchen
Boris Novikov, St Petersburg University
Dan Olteanu, Oxford University
Gultekin Ozsoyoglu, Case Western Reserve University
Tamer Ozsu, University of Waterloo
Esther Pacitti, University of Montpellier
Themis Palpanas, University of Trento
Ippokratis Pandis, IBM Almaden
Stelios Pappas, Microsoft Research
Aditya Parameswaran, Stanford University
Srinivasan Parthasarathy, The Ohio State University
Jignesh Patel, University of Wisconsin
Andrew Pavlo, Brown University
Peter Pietzuch, Imperial College London
Neoklis Polyzotis, University of California - Santa Cruz
Cecilia M. Procopiuc, AT&T Labs
Li Qian, University of Michigan
Jorge Quiané-Ruiz, QCRI
Elisa Quintarelli, Politecnico di Milano
Maya Ramanath, IIT Delhi
Louiqa Raschid, University of Maryland
Vibhar Rastogi, Yahoo
Matthias Renz, University of Munich
Kenneth Ross, Columbia University
Sourav S Bhowmick, NTU, Singapore
Dimitris Sacharidis, IMIS Athena, Greece
Kenneth Salem, University of Waterloo
Maria Sapino, University of Torino
Kai-Uwe Sattler, TU Ilmenau
Monica Scannapieco, ISTAT
Bernhard Seeger, University of Marburg
Lidan Shou, Zhejiang University
Adam Silberstein, Trifacta
Lisa Singh, Georgetown University
Radu Sion, Stony Brook University
Yufei Tao, Chinese University of Hong Kong

Nesime Tatbul, ETH Zurich
Arash Termehchy, Oregon State University
Evimaria Terzi, University of Boston
Martin Theobald, Max Planck Institute, Germany
Srikanta Tirthapura, Iowa State University
Riccardo Torlone, Roma Tre University
Anthony Tung, National University of Singapore
Kostas Tzoumas, Technical University of Berlin
Sergei Vassilvitskii, Google Research
Marcos Vaz Salles, University of Copenhagen (DIKU)
Stratis Viglas, University of Edinburgh
Hoang Tam Vo, National University of Singapore
Daisy Zhe Wang, University of Florida
Haixun Wang, Microsoft Research Asia
Ke Wang, Simon Fraser University
Wei Wang, University of New South Wales
Xiaoling Wang, East China Normal University
Ingmar Weber, Yahoo
Raymond Chi Wing Wong, HKUST
Sai Wu, Zhejiang University
Yuqing Wu, Indiana University
Xiaokui Xiao, NTU
Dong Xin, Google
Jianliang Xu, Hong Kong Baptist University
Jun (Jim) Xu, Georgia Institute of Technology
Xifeng Yan, University of Santa Barbara
Xiaoyan Yang, Advanced Digital Science Center
Ke Yi, HKUST
Ge Yu, Northeastern University, China
Hwanjo Yu, POSTECH, Korea
Meihui Zhang, National University of Singapore
Wenjie Zhang, The University of New South Wales
Ying Zhang, The University of New South Wales
Zhenjie Zhang, Advanced Digital Science Center
Wenzhao Zhou, Georgetown University
Xiaofang Zhou, University of Queensland

PhD Workshop Chairs

Erich Neuhold, University of Vienna
Yun Yao Li, IBM

Sponsorship Chairs

Mike Carey, University of California, Irvine
Lizhu Zhou, Tsinghua University

Local Organization Chair

Lidan Shou, Zhejiang University

Web Management Chair

Sai Wu, Zhejiang University

Conference and Registration Chairs

Ke Chen, Zhejiang University
Cuiping Li, Renmin University

Publicity Chairs

Vasilis Vassalos, AUEB, Greece
Dunlu Peng, USST, China

Proceedings Chairs

Li Xiong, Emory University
Cong Yu, Google Research

Treasury Chair

Li (Eric) Qian, University of Michigan

VLDB Endowment Liaison

Kyu-Young Whang, KAIST

PVLDB Managing Editor

Divesh Srivastava, AT&T Labs

PVLDB Information Director

Gerald Weber, University of Auckland

PVLDB Advisory Committee

Philip Bernstein, Michael Böhlen, Peter Buneman,
Susan Davidson, Z. Meral Ozsoyoglu, S. Sudarshan,
Gerhard Weikum

Logo Design

Guanmin Guo

WHAT TO EXPECT FROM A REVISION

What does the possibility of revision do to the conference review process? This is a serious question, for which we do not yet have community consensus. Let me state here the principles according to which we are trying to operate in PVLDB, with the full recognition that all evaluations are subjective, and therefore suffer from some variance, in spite of our efforts at having a common shared set of metrics.

As a (very loose) lower bound, a revision process improves upon author feedback used in so many conferences. Author feedback does indeed provide a channel to correct egregious reviewing errors. However, authors are incentivized to rebut every criticism, no matter how reasonable, because they have already submitted something, and acknowledging deficiencies simply increases the likelihood of rejection. In turn, reviewers take authors' rebuttals with a big grain of salt. In consequence, author feedback has little influence on most paper decisions, and I have heard many complaints about this process not fulfilling its goals from dissatisfied authors who worked hard to create a rebuttal and then felt that it was not adequately considered by the reviewers.

In contrast, with a revision process, authors can choose to accept criticism and revise their manuscript to address it. Therefore, they attempt to rebut only a few chosen criticisms that they prefer not to address for some reason. The result is both better dialog with referees, and an improved manuscript.

I think everyone agrees up to this point in my argument. Divergence begins after this. The fundamental question is what should be the criterion for a revision request. There are some reviewers who feel this should be driven by the size of the revision request. My view, and one that I have attempted to follow this volume, is that the size of the revision, and the time that it may take, is immaterial. A revision request is appropriate if the deficiency, whether in assumption, in definition, in proof, in experiment, or in presentation, is fixable in a predictable way. (Sometimes, there may be 2 or 3 alternative ways). On the other hand, a revision request is not appropriate if the deficiency is one of insufficient novelty or insufficient technical depth. In short, we want to see papers that are really novel and really deep and really push the envelope in new ways: if the execution is less than perfect, we will help you with a critique and opportunity to fix.

I have made the statement from the previous paragraph in many forums. And an unfortunate consequence is that this has raised author hopes up unreasonably. So let me caution authors that the additional flexibility offered by the revision process requires that the submitted paper be that much more exciting to the reviewers. If reviewers rate a submission WR, it may be true that after revision they would see a manuscript they would rate WA. But that is not enough. I want at least some reviewers to feel truly excited about the manuscript. I want them to be rating it the equivalent of an A or SA after revision.

When the revision is submitted, reviewers get the manuscript again for review. They have access to the original manuscript and the original reviews. They are discouraged from finding new flaws with the manuscript after revision. So the central question is whether the authors adequately addressed the shortcomings raised in the first round. (Presumably the reviewers agreed that if these were well-addressed, the paper would be a good one to accept). Reviewers are forced to rate the manuscript as an accept or a reject after revision. Nothing would give me greater pleasure than to report that we had a 100% acceptance rate after revision. Unfortunately, that is far from being the case. Occasionally, this is because the revision request is about justifying some crucial assumption that the authors genuinely are not able to address. More often, this is because the authors did not take the revision request seriously enough, and submitted a revision that the referees felt addressed their concerns partially, but not enough to accept the paper.

In summary, the revision process can be a boon to authors, and is potentially a great mechanism for avoiding erroneous rejections. However, it comes with higher expectations, and authors should

remember that their burden after revision is not just to meet a minimum bar but rather to generate enthusiasm.

H. V. Jagadish, University of Michigan, USA
Founding Editor-in-Chief, PVLDB