August 12–16, 2012
Beijing, China

Association for
Computing Machinery

Advancing Computing as a Science & Profession

KDD'12

The 18th ACM SIGKDD International Conference on

Knowledge Discovery and DataMining

Sponsored by:

KDD.org and ACM SIGMOD

Supported by:

Huawei, Baidu, Aliyun.com, Facebook, EMC², Elsevier, PayPal of Ebay, SAS, IBM Research, Opera Solutions, Soso, SAP, Microsoft, Google, Greenplum, Nokia Research Center, Deloitte, Adobe, Media6Degrees, NSF, CityGrid.com, HP, CRC Press, Springer, Zementis, Technicolor, and Morgan & Claypool Publishers

Advancing Computing as a Science & Profession

The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York, New York 10121-0701

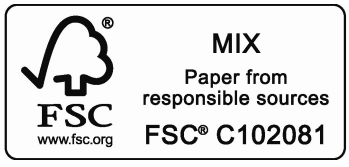
ISBN: 978-1-4503-1462-6

Additional copies may be ordered prepaid from:

ACM Order Department
PO Box 30777
New York, NY 10087-0777, USA

Phone: 1-800-342-6626 (USA and Canada)
+1-212-626-0500 (Global)
Fax: +1-212-944-1318
E-mail: acmhelp@acm.org
Hours of Operation: 8:30 am – 4:30 pm ET

ACM Order Number: 618122



Printed in the USA

General Chairs' Welcome Message

KDD-2012, the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, is being held in Beijing, China, on August 12--16, 2012. KDD is the leading international forum for the exchange of research results and practical experience in the field of knowledge discovery and data mining. As we enter the era of Big Data, there is an increasing need to share our innovative ideas and state-of-the-art solutions on knowledge-discovery and data mining from multiple perspectives, ranging from academic research to industrial applications, from scientific discovery to health services. This volume highlights a snapshot of the developments in this rapidly developing field.

The KDD-2012 technical program features four parallel research tracks, an industrial / government track, and an industry practice expo (IPE), an Asia-pacific Track (APT) and a new KDD Summer School. The APT is specifically designed to take advantage of the fact that this KDD is held in Asia, by inviting a sample of top Asia-pacific researchers to present their data mining projects. We hope this new track will enable the KDD community to learn more on the developments in Asia-pacific region. The KDD Summer School allows students attending the conference to take advantage of availability of the many KDD experts at one location. The courses in the summer school will introduce two of the hot topics in KDD today: `Mining of Complex Data' and `Large Scale Data Mining'.

The KDD 2012 program also features keynotes from leaders of KDD research and practice, 20 workshops, 7 tutorials and one panel. The 2012 KDD Cup competition focuses on microblogging and Web search advertising problems from Tencent, China's Internet and social media giant, which also represent the largest KDD CUP data sets ever. Dozens of technical demonstrations and exhibits from vendors and other organizations underscore the conference's dual role as the leading industry and academic forum to discuss the advances in this field of research and applications.

The call for papers attracted 755 research papers and 113 industrial and government submissions from around the world. Each paper was independently reviewed by three members of the program committee for originality, significance, technical quality, and clarity of presentation. In the end, the program committee accepted 133 papers into the research track, representing an acceptance rate of 17.6%. The industry and government track of the conference accepted 30 papers into the program, representing an acceptance rate of 26.5%.

From the conference, we can see that the KDD community is continuing to be vibrant and growing, and KDD innovations remain very strong and have very high quality. The resulting program is notable for its diversity and vitality. With the exciting program as our backdrop, we certainly hope this conference will become a new starting point, where research and development on data science and data mining, and research and development on Big data analytics, will continue to flourish.

Qiang Yang
KDD2012 General Chair
HKUST, Hong Kong, China

Dou Shen
KDD2012 Associate General Chair
CityGrid Media, Seattle, WA, USA

Research Track Program Chairs' Welcome Message

The KDD conference has seen remarkable growth since its origins as an IJCAI workshop in Detroit in 1989, evolving into a full-fledged research conference in 1995, underscoring the important role data mining as a field has played in extracting knowledge and actionable insights from vast troves of data that is being generated in the digital world around us. This year we received a record 755 submissions to the research program, from which 133 papers were accepted, for an aggregate acceptance rate of 17.6% (quite similar to recent years).

Among the academic conferences, the KDD conference has typically more of an emphasis on research motivated by real-world applications. It is important to keep in mind that it is this synergy of research in areas like algorithms, computational geometry, database, graph theory, machine learning, natural language processing, statistics, visualization and many others when applied to problems arising in diverse fields such as web, medicine, climatology, marketing that drives our field forward, makes it vibrant and fun – who would know that ideas in computational geometry can be adapted to construct fast algorithms to improve online advertising and movie recommendations?

The breadth of topics covered in this year's research program is truly comprehensive, including social networks, privacy, text mining, predictive modeling, time-series forecasting, spatial data analysis, geometry, and more. We are very fortunate to have 4 world-class keynote speakers this year spanning industry and academia, providing inspirational talks on cutting-edge techniques and issues in web mining, information networks, statistical inference for big data, and social computing.

The process of whittling down the initial 734 submissions to the final set of 133 accepted papers required the coordination and time of a large number of willing volunteers. The program committee (PC) consisted of over 350 reviewers (PC members) and 50 senior PC members. In the first phase each submitted paper was automatically assigned to 3 reviewers (after a bidding process). Once the reviews from each of the 3 reviewers were completed, the program chairs rejected papers that did not receive much support from any of the reviewers. We rejected 259 papers at this stage. Special care was taken to minimize the error of rejecting a potentially good paper at this stage. The papers that survived the first phase were assigned to the senior PC members based on their bids, they had the option of initiating a discussion for any of their papers, e.g., if there was significant divergence in scores among reviewers, or if a paper was on the borderline of being accepted. Following the discussion phase, the senior PC members provided a recommendation score and a detailed meta-review for each paper. In the final phase, we (the program chairs) analyzed all of this information, starting with the obvious accept and reject decisions, and then gradually focusing in more detail on the papers near the borderline, seeking additional reviews and input from the PC and senior PC members where appropriate. We also initiated a shepherding phase with 15 papers having the opportunity of fixing mild issues we thought would be possible to address before they can be accepted. 13 of them were accepted after thorough revisions. Finally, it is quite likely that in hindsight some worthy papers may have been rejected as part of this process - these errors are an unfortunate reality of modern computer science conferences, and hard to avoid when a very large number of decisions have to be made over a short time span based on a subjective reviewing process. Nevertheless, we, the PC chairs, are responsible for those unfortunate errors and welcome suggestions on the matter.

We conclude with some well-deserved words of thanks to the large supporting cast. We thank all authors for submitting their research to KDD 2012. We are extremely grateful to the PC members and senior PC members for volunteering their time to help in the reviewing process. Diligent peer review is an essential cornerstone for research progress - we thank you all for the many detailed and conscientious reviews that

were generated. Last year's program chairs, Joydeep Ghosh and Padhraic Symth, were generous in providing us with advice and expertise based on lessons learned in 2011. We would also like to convey our sincere thanks to the staff who run the Microsoft CMT system - they were more than willing to answer questions and help us at all times during the submission, reviewing, and decision process. And finally we would like to thank all of our colleagues on the KDD 2012 organization committee who were a pleasure to work with and without whom we would not have such a wonderful conference.

Let the 18th ACM SIGKDD conference begin - we hope you enjoy it!

Deepak Agarwal and Jian Pei
ACM SIGKDD 2012 Program Chairs

Asia Pacific Track Chairs' Welcome

It is our great honor to introduce the ***Asia Pacific Track*** at the 2012 ACM KDD conference. The Asia Pacific track features invited talks from leading data mining research groups located in the Asia Pacific region. We believe that this opportunity for the KDD community to hear directly from these experts will stimulate new research ideas and cross-continent collaborations.

In the last decade, universities and research laboratories in the Asia-Pacific region have made significant progress and contribution to the KDD research. Taking the advantage of the opportunity that KDD 2012 is held in Beijing, China, the conference features a new program of Asia-Pacific Track. The purpose of the Track is to promote KDD research achievements and to elevate the attention and momentum of the KDD research in this region. The program features invited speeches on KDD research results and projects in universities and industrial laboratories in this region. Recommended by a program committee consisting of Tu-Bao Ho, Jianyong Wang, Hiroshi Motoda, and Zhihua Zhou, it is our delight to invite the following 12 distinguished researchers to speak in the Asia Pacific Track.

1. Ming-Syan Chen, National Taiwan University
2. Paul Compton, University of New South Wales
3. Geoff Holmes, University of Waikato
4. Masaru Kitsuregawa, University of Tokyo
5. Cuiping Li, Renmin University
6. Deyi Li, Chinese Academy of Engineering
7. Jianzhong Li, Harbin Institute of Technology
8. Zhongzhi Shi, Chinese Academy of Sciences
9. Gordon Sun, Tencent
10. Naonori Ueda, NTT Communication Lab
11. Bo Zhang, Tsinghua University
12. Aoying Zhou, Eastern Normal University

We would like to thank all of our speakers for accepting our invitation to present their work at the KDD conference. We warmly welcome you to the Asia Pacific Track and hope that you will enjoy it.

David Cheung and Wei Wang
Asia Pacific Track Co- Chairs

Foreword to the KDD 2012 Industry Practice Expo

We warmly welcome you to the second year of the Industry Practice Expo (IPE) track at the KDD 2012 conference. The IPE features invited talks from leading experts who have developed and deployed large-scale data mining applications and systems in their respective fields that generated measurable and meaningful impact. The IPE aspires to bridge the gap between academia and industrial practice by bringing forth key trends, significant insights, and best practices. It complements the two well-established KDD tracks, Research and Industry/Government, which continue to feature the latest advances in the field via peer-reviewed publications. We trust that this opportunity for the KDD community to hear directly from industry pioneers will inspire new advances and broader inter-disciplinary collaboration between the academic researchers and their industry counterparts.

Last year's inaugural IPE was a huge success. The talks were well received and it was fascinating to hear about the numerous discussions between the speakers and attendees during the breaks. An important feedback from the attendees was that IPE should focus on technical talks and avoid talks that are purely marketing or promotional in nature. We have worked hard with our speakers for 2012 to achieve this right balance between business context and technical depth in the IPE talks.

In this year's IPE, we present to you renowned experts from around the world, from whose talks we hope will you will learn and take key ideas back to your respective professional activities. Featured topics at the IPE include large national scale projects on credit scoring and tax compliance, and real-world predictive analytics applications in decision management systems, e-commerce, semantic search, and telecommunications. Social media sites like LinkedIn, Facebook, and Twitter have taken the world by storm. We will hear about the key lessons learned in developing successful recommendation systems that power many interesting applications on these social media sites. Large-scale machine learning algorithms are an integral part of many successful data mining systems. One of the leading developers of industry strength machine learning software will share his experiences in building robust software that is now being used widely across the world. We are fortunate to have such a diverse panel of speakers be part of the IPE and would like to take this opportunity to thank them for their participation.

It has been a great pleasure to work with the IPE advisory comprising of Chid Apte, Usama Fayyad, Gregory Piatetsky-Shapiro, Ted Senator, Ramasamy Uthurusamy, Qiang Yang, and Michael Zeller. We thank them for their advice and support throughout the search for candidate speakers. We would also like to express our heartfelt gratitude to Michelle Chambers who was one of the speakers at the inaugural IPE in 2011. Michelle recommended several speakers and worked closely with us as we discussed the KDD opportunity with them.

We warmly welcome you to the KDD 2012 Industry Practice Expo and hope that you will enjoy it. Your suggestions and comments are very welcome.

Best regards,

Rajesh Parekh and Ying Li

Industry Practice Expo Track co-chairs

Foreword to the Industry & Government Track

The 2012 Industry & Government Track continued the direction that was implemented in 2010 and 2011 by clearly focusing on three areas of applied KDD work that are the sources of value to industrial and government organizations: (1) deployed applications, (2) knowledge discoveries, and (3) emerging applications. These areas build on the results of KDD research by applying the results of such research to enable the construction of novel and useful applications, the discovery of novel, interesting and useful knowledge, and the extension and application of research results to specific real world problems of interest. Papers were solicited and reviewed in these three areas:

- Deployed KDD systems that are providing real value to industry, government, or other organizations or professions. These deployed systems could support ongoing knowledge discovery or could be applications that employ discovered knowledge, or some combination of the two.
- Discoveries of knowledge with demonstrable value to industry, government, or other users (e.g., scientific or medical professionals). This knowledge must be "externally validated" as interesting and useful; it cannot simply be a model that has better performance on some traditional KDD metric such as accuracy or area under the curve.
- Emerging applications and technology that provide insight relevant to the above value propositions. These emerging applications must have clear user interest and support to distinguish them from KDD research papers, or they must provide insight into issues and factors that affect the successful use of KDD technology and methods. Papers that describe infrastructure that enables the large-scale deployment of KDD techniques also are in this area.

Industry & Government Track papers not only document and recognize significant accomplishments in these three areas, but also provide feedback to KDD researchers about needs, constraints, and applicability of KDD algorithms, methods and techniques in the real world. Our emphasis was on papers that advance the understanding of practical, applied, or pragmatic issues related to the use of KDD technologies in industry and government settings and highlight new research challenges arising from attempts to create such real KDD applications. We encouraged potential authors to address these issues and challenges in the Call for Papers and the accompanying Author Instructions http://www.kdd.org/kdd2012/author_instructions_industry.shtml

Our review process began with a validation by the reviewers that the authors correctly classified their paper's area – to ensure that the paper was reviewed according to the applicable criteria for that area – and a check that all required elements for that area were addressed in the paper. These required elements were:

- Deployed KDD case studies describe deployed projects with measurable benefits that include KDD technology. These papers must clearly describe the industry or government problem that is solved, the overall architecture of the deployed system, the data sources used, the reasons for the choices of particular KDD technologies, how KDD technologies solved the problem, the particular KDD process embodied by the deployed application, the use and payoff of the application, the costs to develop the application, the maintenance plan, and the number and types of users.
- Papers that describe discoveries of knowledge must clearly state what data sources and background knowledge were used, what data mining algorithms were tried, what overall KDD process was used, what the new discovered knowledge is, how the new knowledge was

validated, and what the value to the industry or government is of such newly discovered knowledge. Note that this category is intended for significant discoveries of domain knowledge that were enabled by KDD technologies and processes, and not for reporting of KDD research results.

- Emerging application and technology papers discuss prototype applications, tools for focused domains or tasks, useful techniques or methods, useful system architectures, scalability enablers, tool evaluations, or integration of KDD with other technologies. Such papers must clearly explain the requirements arising from the particular industry or government setting for which the application is being developed and from the particular databases on which the application is based. These papers must also identify how the emerging solution is using KDD technologies to address these requirements, the deployment plan, and the evaluation methodology and metrics for the emerging application. Pragmatic issues and considerations include important practical and research considerations, approaches, and architectures that enable successful applications. This category may include comparative evaluations of different KDD technologies for particular application problems. Preferences will be given to papers whose insights may generalize to other domains or problems. Product advertisements will not be accepted.

Specific review criteria included:

- Value/Payoff: What are the measurable benefits of the deployed application, the value of the newly discovered knowledge, or the generalizable insights from the emerging application?
- Cost: What is the cost/effort of the reported project? (To develop and maintain the application or to discover the new knowledge)
- External Significance: Is there clear evidence that the application problem being addressed and the KDD solution matters to an identified external set of users?
- External Validity: Is the knowledge discovered/used by the application externally validated? (This can be demonstrated by its use by a business, government, or profession; by publication in a peer-reviewed domain journal, or other means)
- Does the paper thoroughly describe the data sources and background knowledge used?
- Does the paper thoroughly describe the selection, use, appropriateness and effectiveness of the KDD technologies and/or data mining algorithms?
- Does the paper describe the KDD process enabled by the application or used to discover knowledge?
- Novelty/Originality: Does the work apply KDD technology to a new problem? Does it develop or apply a new type of KDD technology to an existing problem? Does it apply KDD technology to achieve a better result than with previous technology?
- Technical soundness: Does the paper show an accurate understanding of KDD research and technology? Does it identify relevant past work? Does it appropriately match techniques to problem characteristics?
- Evaluation: Is the reported work carefully evaluated? Is it statistically sound? Are the results compared to appropriate baselines? To other approaches and solutions? To previously reported work? Are its claims supported?
- Deployment/Use: Is there clear evidence that the application is deployed or on a path to deployment, or that the discovered knowledge will be used?

- Clarity: Is the paper written clearly? Is it organized logically? Are there sufficient figures and tables to illustrate its main points? Is it accessible to readers without specialized domain knowledge? Is it accessible to readers from different technical specialties and application domains?
- Does the paper contain insights or results that will generalize or apply to other domains or is it specialized to a particular domain?
- Will the paper be accessible to non-specialists in the particular domain?

Overall, the Program Committee reviewed 113 papers, representing an over 50% increase in submissions as compared to 2011 which clearly indicates the increased interest in data mining across Industry and Government. We accepted 30 papers, of which 8 describe deployed applications, 5 knowledge discoveries, and 17 emerging applications. The application areas cover a broad range, including finance, security, internet, text mining, science, public health, retail, mobile computing, and social networks, showing the many different fields to which KDD is being applied and adding value.

We hope you find these papers interesting and look forward to continued improvements in the use of KDD techniques to provide real value to industry and government.

Michael Zeller — *Zementis, Inc.*

Hui Xiong — *Rutgers, the State University of New Jersey*

KDD 2012 Industry & Government Track Program Chairs

Table of Contents

Organizers List xxv

KDD'12 Senior Program Committee Members xxvii

KDD12 Research Track Program Committee xxviii

KDD'12 Industry & Government Track Program Committee Members xxxi

KDD'12 Research Track External Reviewers xxxii

ACM SIGKDD KDD 2012 Sponsors & Supporters xxxv

Keynote Addresses

- **Nine Real Hard Problems We'd Like You to Solve** 1
  Robin Li *(Baidu, Inc)*
- **Mining Heterogeneous Information Networks: The Next Frontier** 2
  Jiawei Han *(University of Illinois at Urbana-Champaign)*
- **Divide-and-Conquer and Statistical Inference for Big Data** 4
  Michael I. Jordan *(University of California, Berkeley)*
- **Experiments in Social Computation (and the Data They Generate)** 5
  Michael Kearns *(University of Pennsylvania)*

Research Session A1: PageRank and Social Networks

- **Rise and Fall Patterns of Information Diffusion: Model and Implications** 6
  Yasuko Matsubara *(Kyoto University)*, Yasushi Sakurai *(NTT Communication Science Labs)*, B. Aditya Prakash *(Carnegie Mellon University)*, Lei Li *(University of California, Berkeley)*, Christos Faloutsos *(Carnegie Mellon University)*
- **Efficient Personalized PageRank with Accuracy Assurance** 15
  Yasuhiro Fujiwara, Makoto Nakatsuji, Takeshi Yamamuro, Hiroaki Shiokawa, Makoto Onizuka *(NTT Softweare Innovation Center)*
- **PageRank on an Evolving Graph** 24
  Bahman Bahmani *(Stanford University)*, Ravi Kumar *(Yahoo! Research)*, Mohammad Mahdian *(Google, Inc.)*, Eli Upfal *(Brown University)*
- **Information Diffusion and External Influence in Networks** 33
  Seth Myers, Chenguang Zhu, Jure Leskovec *(Stanford University)*
- **The Missing Models: A Data-Driven Approach for Learning How Networks Grow** 42
  Rob Patro, Geet Duggal, Emre Sefer, Hao Wang, Darya Filippova, Carl Kingsford *(University of Maryland)*

Research Session A2: Pattern Mining

- **Finding Minimum Representative Pattern Sets** 51
  Guimei Liu, Haojun Zhang, Limsoon Wong *(National University of Singapore)*
- **Mining Emerging Patterns by Streaming Feature Selection** 60
  Kui Yu *(Hefei University of Technology & University of Massachusetts, Boston)*, Wei Ding, Dan A. Simovici *(University of Massachusetts, Boston)*, Xindong Wu *(Hefei University of Technology & University of Vermont)*
- **Linear Space Direct Pattern Sampling Using Coupling from the Past** 69
  Mario Boley *(Fraunhofer IAIS and University of Bonn)*, Sandy Moens *(University of Antwerp)*, Thomas Gärtner *(Fraunhofer IAIS and University of Bonn)*

- **Mining Top-K High Utility Itemsets** 78
  Cheng Wei Wu, Bai-En Shie *(National Cheng Kung University)*, Philip S. Yu *(University of Illinois at Chicago)*, Vincent S. Tseng *(National Cheng Kung University)*
- **Sampling Minimal Frequent Boolean (DNF) Patterns** 87
  Geng Li, Mohammed J. Zaki *(Rensselaer Polytechnic Institute)*

Research Session A3: Probabilistic Models

- **The Contextual Focused Topic Model** 96
  Xu Chen, Mingyuan Zhou, Lawrence Carin *(Duke University)*
- **Practical Collapsed Variational Bayes Inference for Hierarchical Dirichlet Process** 105
  Issei Sato *(The University of Tokyo)*, Kenichi Kurihara *(Google)*, Hiroshi Nakagawa *(The University of Tokyo)*
- **Overlapping Decomposition for Causal Graphical Modeling** 114
  Lei Han, Guojie Song *(Peking University)*, Gao Cong *(Nanyang Technological University)*, Kunqing Xie *(Peking University)*
- **TM-LDA: Efficient Online Modeling of Latent Topic Transitions in Social Media** 123
  Yu Wang, Eugene Agichtein, Michele Benzi *(Emory University)*
- **Multi-View Clustering Using Mixture Models in Subspace Projections** 132
  Stephan Günnemann, Ines Färber, Thomas Seidl *(RWTH Aachen University)*

Research Session A4: Supervised Learning

- **A Simple Methodology for Soft Cost-Sensitive Classification** 141
  Te-Kang Jan, Da-Wei Wang *(Academia Sinica)*, Chi-Hung Lin, Hsuan-Tien Lin *(National Taiwan University)*
- **Intelligible Models for Classification and Regression** 150
  Yin Lou *(Cornell University)*, Rich Caruana *(Microsoft Corporation)*, Johannes Gehrke *(Cornell University)*
- **NASA: Achieving Lower Regrets and Faster Rates via Adaptive Stepsizes** 159
  Hua Ouyang, Alexander Gray *(Georgia Institute of Technology)*
- **Learning in Non-Stationary Environments with Class Imbalance** 168
  Thomas Ryan Hoens, Nitesh V. Chawla *(University of Notre Dame)*
- **Linear Support Vector Machines Via Dual Cached Loops** 177
  Shin Matsushima *(The University of Tokyo)*, S. V. N. Vishwanathan *(Purdue University)*, Alexander J. Smola *(Yahoo! Research)*

Industry/Govt Track A5: Mobile Computing

Session Chair: Yu Zheng

- **Discovering Regions of Different Functions in a City Using Human Mobility and POIs** ... 186
  Jing Yuan, Yu Zheng, Xing Xie *(Microsoft Research Asia)*
- **Constructing Popular Routes from Uncertain Trajectories** 195
  Ling-Yin Wei *(National Chiao Tung University)*, Yu Zheng *(Microsoft Research Asia)*, Wen-Chih Peng *(National Chiao Tung University)*
- **GetJar Mobile Application Recommendations with Very Sparse Datasets** 204
  Kent Shi, Kamal Ali *(GetJar Inc.)*
- **Differentially Private Transit Data Publication: A Case Study on the Montreal Transportation System** 213
  Rui Chen, Benjamin C. M. Fung, Bipin C. Desai *(Concordia University)*, Nériah M. Sossou *(Société de transport de Montréal)*

Asia-Pacific Track A6: Session 1

- **Interaction and Collective Intelligence in Internet Computing** 222
  Deyi Li *(Chinese Academy of Engineering)*
- **Building an Engine for Big Data** 223
  Masaru Kitsuregawa *(University of Tokyo)*

- **A New Challenge of Information Processing Under the 21st Century** 224
  Bo Zhang *(Tsinghua University)*
- **Developing Data Mining Applications** 225
  Geoff Holmes *(University of Waikato)*

Research Session B1: Social Opinions

- **Learning from Crowds in the Presence of Schools of Thought** 226
  Yuandong Tian *(Carnegie Mellon University)*, Jun Zhu *(Tsinghua University)*
- **Social Sampling** 235
  Anirban Dasgupta, Ravi Kumar, D. Sivakumar *(Yahoo! Research)*
- **From User Comments to On-Line Conversations** 244
  Chunyan Wang *(Stanford University)*, Mao Ye, Bernardo A. Huberman *(HP Labs)*
- **eTrust: Understanding Trust Evolution in an Online World** 253
  Jiliang Tang, Huiji Gao, Huan Liu *(Arizona State University)*, Atish Das Sarma *(eBay Inc.)*

Research Session B2: Time Series

- **Searching and Mining Trillions of Time Series Subsequences Under Dynamic Time Warping** 262
  Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen *(University of California, Riverside)*, Gustavo Batista *(University of São Paulo)*, Brandon Westover *(Brigham and Women's Hospital)*, Qiang Zhu, Jesin Zakaria, Eamonn Keogh *(University of California, Riverside)*
- **Fast Mining and Forecasting of Complex Time-Stamped Events** 271
  Yasuko Matsubara *(Kyoto University)*, Yasushi Sakurai *(NTT Communication Science Labs)*, Christos Faloutsos *(Carnegie Mellon University)*, Tomoharu Iwata *(NTT Communication Science Labs)*, Masatoshi Yoshikawa *(Kyoto University)*
- **Mining Recent Temporal Patterns for Event Detection in Multivariate Time Series Data** 280
  Iyad Batal *(University of Pittsburgh)*, Dmitriy Fradkin *(Siemens Corporate Research)*, James Harrison *(University of Virginia)*, Fabian Moerchen *(Siemens Corporate Research)*, Milos Hauskrecht *(University of Pittsburgh)*
- **A Shapelet Transform for Time Series Classification** 289
  Jason Lines, Luke M. Davis, Jon Hills, Anthony Bagnall *(University of East Anglia)*

Research Session B3: Matrices and Tensors

- **Accelerated Singular Value Thresholding for Matrix Completion** 298
  Yao Hu, Debing Zhang *(Zhejiang University)*, Jun Liu *(Siemens Corporate Research)*, Jieping Ye *(Arizona State University)*, Xiaofei He *(Zhejiang University)*
- **Fast Bregman Divergence NMF Using Taylor Expansion and Coordinate Descent** 307
  Liangda Li, Guy Lebanon, Haesun Park *(Georgia Institute of Technology)*
- **GigaTensor: Scaling Tensor Analysis Up by 100 Times - Algorithms and Discoveries** 316
  U Kang, Evangelos Papalexakis, Abhay Harpale, Christos Faloutsos *(Carnegie Mellon University)*
- **Active Learning for Online Bayesian Matrix Factorization** 325
  Jorge Silva, Lawrence Carin *(Duke University)*

Research Session B4: Unsupervised Learning

- **A Sparsity-Inducing Formulation for Evolutionary Co-Clustering** 334
  Shuiwang Ji, Wenlu Zhang *(Old Dominion University)*, Jun Liu *(Siemens Corporate Research)*
- **Detecting Changes of Clustering Structures Using Normalized Maximum Likelihood Coding** 343
  So Hirai, Kenji Yamanishi *(The University of Tokyo)*

- **Subspace Correlation Clustering: Finding Locally Correlated Dimensions in Subspace Projections of the Data** 352
  Stephan Günnemann, Ines Färber, Kittipat Virochsiri, Thomas Seidl *(RWTH Aachen University)*
- **Dependency Clustering Across Measurement Scales** 361
  Claudia Plant *(Florida State University)*

Industry/Govt Track B5: Social Network Analysis

Session Chair: Amol Ghoting

- **A Framework for Summarizing and Analyzing Twitter Feeds** 370
  Xintian Yang *(The Ohio State University)*, Amol Ghoting *(IBM Thomas J. Watson Research Center)*, Yiye Ruan, Srinivasan Parthasarathy *(The Ohio State University)*
- **Entity-Centric Topic-Oriented Opinion Summarization in Twitter** 379
  Xinfan Meng *(Peking University)*, Furu Wei, Xiaohua Liu, Ming Zhou *(Microsoft Research Asia)*, Sujian Li, Houfeng Wang *(Peking University)*
- **Community Discovery and Profiling with Social Messages** 388
  Wenjun Zhou *(University of Tennessee)*, Hongxia Jin *(IBM Research - Almaden)*, Yan Liu *(University of Southern California)*
- **Finding Trending Local Topics in Search Queries for Personalization of a Recommendation System** 397
  Ziad Al Bawab, George H. Mills, Jean-Francois Crespo *(Yahoo! Labs)*

Industrial Practice Expo B6: Session 1

- **China's National Personal Credit Scoring System: A Real-Life Intelligent Knowledge Application** 406
  Yong Shi *(Chinese Academy of Sciences / China Data Technology)*
- **Maximizing Return and Minimizing Cost with the Right Decision Management Systems** 407
  Rich Holada *(IBM)*

Research Session C1: Social and Web Mining Applications

- **Efficient and Domain-Invariant Competitor Mining** 408
  Theodoros Lappas *(Boston University)*, George Valkanas, Dimitrios Gunopulos *(University of Athens)*
- **Discriminative Clustering for Market Segmentation** 417
  Peter Haider *(University of Potsdam)*, Luca Chiarandini *(Universitat Pompeu Fabra)*, Ulf Brefeld *(University of Bonn)*
- **Interacting Viruses in Networks: Can Both Survive?** 426
  Alex Beutel, B. Aditya Prakash, Roni Rosenfeld, Christos Faloutsos *(Carnegie Mellon University)*
- **Aggregating Web Offers to Determine Product Prices** 435
  Rakesh Agrawal, Samuel Ieong *(Microsoft Research)*

Research Session C2: Event Mining

- **Mining Event Periodicity from Incomplete Observations** 444
  Zhenhui Li, Jingjing Wang, Jiawei Han *(University of Illinois at Urbana-Champaign)*
- **Towards Heterogeneous Temporal Clinical Event Pattern Discovery: A Convolutional Approach** 453
  Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, Shahram Ebadollahi *(IBM T.J. Watson Research Center)*
- **The Long and the Short of It: Summarising Event Sequences with Serial Episodes** 462
  Nikolaj Tatti, Jilles Vreeken *(University of Antwerp)*
- **Efficient Event Pattern Matching with Match Windows** 471
  Bruno Cadonna, Johann Gamper *(Free University of Bozen-Bolzano)*, Michael H. Böhlen *(University of Zurich)*

Research Session C3: Matrix Approximation

- **Optimal Exact Least Squares Rank Minimization** 480
  Shuo Xiang *(Arizona State University)*, Yunzhang Zhu, Xiaotong Shen *(University of Minnesota)*, Jieping Ye *(Arizona State University)*
- **Large-Scale Distributed Non-Negative Sparse Coding and Sparse Dictionary Learning** 489
  Vikas Sindhwani, Amol Ghoting *(IBM T.J. Watson Research Center)*
- **Learning Binary Codes for Collaborative Filtering** 498
  Ke Zhou, Hongyuan Zha *(Georgia Institute of Technology)*
- **Low Rank Modeling of Signed Networks** 507
  Cho-Jui Hsieh, Kai-Yang Chiang, Inderjit S. Dhillon *(University of Texas at Austin)*

Research Session C4: Supervised Learning with Multivariate Data

- **A Structural Cluster Kernel for Learning on Graphs** 516
  Madeleine Seeland *(Technische Universität München)*, Andreas Karwath, Stefan Kramer *(Johannes Gutenberg-Universität Mainz)*
- **Multi-Label Hypothesis Reuse** 525
  Sheng-Jun Huang, Yang Yu, Zhi-Hua Zhou *(Nanjing University)*
- **Rank-Loss Support Instance Machines for MIML Instance Annotation** 534
  Forrest Briggs, Xiaoli Z. Fern, Raviv Raich *(Oregon State Univeristy)*
- **Inductive Multi-Task Learning with Multiple View Data** 543
  Jintao Zhang, Jun Huan *(University of Kansas)*

Industry/Govt Track C5: Web Applications

Session Chair: Enhong Chen

- **Scalable Misbehavior Detection in Online Video Chat Services** 552
  Xinyu Xing *(Georgia Institute of Technology)*, Yu-li Liang *(University of Colorado at Boulder)*, Sui Huang *(Ohio State University)*, Hanqiang Cheng *(McGill University)*, Richard Han, Qin Lv *(University of Colorado at Boulder)*, Xue Liu *(McGill University)*, Shivakant Mishra *(University of Colorado at Boulder)*, Yi Zhu *(Tsinghua University)*
- **Keyword-Propagation-Based Information Enriching and Noise Removal for Web News Videos** 561
  Jun Zhang, Xiaoming Fan, Jianyong Wang, Lizhu Zhou *(Tsinghua University)*
- **Harnessing the Wisdom of the Crowds for Accurate Web Page Clipping** 570
  Lei Zhang *(University of Science and Technology of China & HP Labs China)*, Linpeng Tang *(Shanghai Jiao Tong University & HP Labs China)*, Ping Luo *(HP Labs China)*, Enhong Chen *(University of Science and Technology of China)*, Limei Jiao, Min Wang *(HP Labs China)*, Guiquan Liu *(University of Science and Technology of China)*
- **Bootstrapped Language Identification for Multi-Site Internet Domains** 579
  Uwe F. Mayer *(eBay Research Labs)*

Industrial Practice Expo C6: Session 2

- **Semantic Search and a New Moore's Law Effect in Knowledge Engineering** 586
  Wei-Ying Ma *(Microsoft Research Asia)*
- **Key Lessons Learned Building Recommender Systems for Large-Scale Social Networks** 587
  Christian Posse *(LinkedIn Inc.)*

Research Session A1: Community Mining

- **Magnet Community Identification on Social Networks** 588
  Guan Wang, Yuchen Zhao, Xiaoxiao Shi *(University of Illinois at Chicago)*, Philip S. Yu *(University of Illinois at Chicago & King Abdulaziz University)*

- **Vertex Neighborhoods, Low Conductance Cuts, and Good Seeds for Local Community Methods** 597
  David F. Gleich *(Purdue University)*, C. Seshadhri *(Sandia National Laboratories)*
- **Overlapping Community Detection via Bounded Nonnegative Matrix Tri-Factorization** 606
  Yu Zhang, Dit-Yan Yeung *(Hong Kong University of Science and Technology)*
- **DEMON: A Local-First Discovery Method for Overlapping Communities** 615
  Michele Coscia *(CID - Harvard Kennedy School)*, Giulio Rossetti *(University of Pisa)*, Fosca Giannotti *(KDDLab ISTI-CNR)*, Dino Pedreschi *(University of Pisa)*
- **On the Separability of Structural Classes of Communities** 624
  Bruno Abrahao, Sucheta Soundarajan, John Hopcroft, Robert Kleinberg *(Cornell University)*

Research Session A2: Sequential and Spatio-temporal Patterns

- **Discovering Lag Intervals for Temporal Dependencies** 633
  Liang Tang, Tao Li *(Florida International University)*, Larisa Shwartz *(IBM T.J. Watson Research Center)*
- **Testing the Significance of Spatio-Temporal Teleconnection Patterns** 642
  Jaya Kawale, Snigdhansu Chatterjee, Dominick Ormsby, Karsten Steinhaeuser, Stefan Liess, Vipin Kumar *(University of Minnesota)*
- **SeqiBloc: Mining Multi-time Spanning Blockmodels in Dynamic Graphs** 651
  Jeffrey Chan, Wei Liu, Christopher Leckie, James Bailey, Kotagiri Ramamohanarao *(University of Melbourne)*
- **USpan: An Efficient Algorithm for Mining High Utility Sequential Patterns** 660
  Junfu Yin, Zhigang Zheng, Longbing Cao *(University of Technology, Sydney)*
- **Mining Large-Scale, Sparse GPS Traces for Map Inference: Comparison of Approaches** 669
  Xuemei Liu *(Shanghai Jiao Tong University)*, James Biagioni, Jakob Eriksson *(University of Illinois at Chicago)*, Yin Wang, George Forman *(Hewlett-Packard Labs)*, Yanmin Zhu *(Shanghai Jiao Tong University)*

Research Session A3: Personalization and Recommendation

- **Transparent User Models for Personalization** 678
  Khalid El-Arini *(Carnegie Mellon University)*, Ulrich Paquet *(Microsoft Research)*, Ralf Herbrich *(Facebook, Inc.)*, Jurgen Van Gael *(Rangespan Ltd.)*, Blaise Agüera y Arcas *(Microsoft Corporation)*
- **Estimating Entity Importance via Counting Set Covers** 687
  Aristides Gionis *(Yahoo! Research)*, Theodoros Lappas, Evimaria Terzi *(Boston University)*
- **ComSoc: Adaptive Transfer of User Behaviors Over Composite Social Network** 696
  Erheng Zhong *(Hong Kong University of Science and Technology)*, Wei Fan *(IBM T.J. Watson Research Center)*, Junwei Wang, Lei Xiao, Yong Li *(Tencent Inc.)*
- **Online Learning to Diversify from Implicit Feedback** 705
  Karthik Raman, Pannaga Shivaswamy, Thorsten Joachims *(Cornell University)*
- **Playlist Prediction via Metric Embedding** 714
  Shuo Chen, Joshua L. Moore *(Cornell University)*, Douglas Turnbull *(Ithaca College)*, Thorsten Joachims *(Cornell University)*

Research Session A4: Supervised Learning with Auxilliary Information

- **Parallel Field Ranking** 723
  Ming Ji *(University of Illinois at Urbana-Champaign)*, Binbin Lin, Xiaofei He, Deng Cai *(Zhejiang University)*, Jiawei Han *(University of Illinois at Urbana-Champaign)*
- **Semi-Supervised Learning with Mixed Knowledge Information** 732
  Fanhua Shang, L. C. Jiao *(Xidian University)*, Fei Wang *(IBM T. J. Watson Research Center)*
- **Batch Mode Active Sampling Based on Marginal Probability Distribution Matching** 741
  Rita Chattopadhyay, Zheng Wang *(Arizona State University)*, Wei Fan *(IBM T. J. Watson Research Center)*, Ian Davidson *(University of California, Davis)*, Sethuraman Panchanathan, Jieping Ye *(Arizona State University)*

- **SPG-GMKL: Generalized Multiple Kernel Learning with a Million Kernels** 750
  Ashesh Jain *(Indian Institute of Technology Delhi)*, S. V. N. Vishwanathan *(Purdue University)*, Manik Varma *(Microsoft Research India)*
- **Efficient Evaluation of Large Sequence Kernels** 759
  Pavel P. Kuksa *(NEC Laboratories America, Inc.)*, Vladimir Pavlovic *(Rutgers University)*

Industry/Govt Track A5: Computational Advertising

Session Chair: Ron Kohavi

- **Estimating Conversion Rate in Display Advertising from Past Performance Data** 768
  Kuang-chih Lee, Burkay Orten, Ali Dasdan, Wentong Li *(Turn Inc.)*
- **Multimedia Features for Click Prediction of New Ads in Display Advertising** 777
  Haibin Cheng, Roelof van Zwol *(Yahoo! Labs)*, Javad Azimi *(Oregon State University)*, Eren Manavoglu, Ruofei Zhang, Yang Zhou, Vidhya Navalpakkam *(Yahoo! Labs)*
- **Trustworthy Online Controlled Experiments: Five Puzzling Outcomes Explained** 786
  Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, Ya Xu *(Microsoft)*
- **Position-Normalized Click Prediction in Search Advertising** 795
  Ye Chen, Tak W. Yan *(Microsoft Corporation)*
- **Bid Optimizing and Inventory Scoring in Targeted Online Advertising** 804
  Claudia Perlich, Brian Dalessandro, Rod Hook, Ori Stitelman, Troy Raeder *(Media6Degrees)*, Foster Provost *(NYU Stern School of Business)*

Asia-Pacific Track A6: Session 2

- **Algorithms for Mining Uncertain Graph Data** 813
  Jianzhong Li *(Harbin Institute of Technology)*
- **Experience with Discovering Knowledge by Acquiring It** 814
  Paul Compton *(University of New South Wales)*
- **Bayesian Relational Data Analysis** 815
  Naonori Ueda *(NTT Communication Lab)*
- **Cross-Media Knowledge Discovery** 816
  Zhongzhi Shi *(Chinese Academy of Sciences)*

Author Index 817

Research Session B1: Review, Discussion, and Q & A

- **Review Spam Detection via Temporal Pattern Discovery** 823
  Sihong Xie, Guan Wang, Shuyang Lin *(University of Illinois at Chicago)*, Philip S. Yu *(University of Illinois at Chicago & King Abdulaziz University)*
- **Selecting a Characteristic Set of Reviews** 832
  Theodoros Lappas, Mark Crovella, Evimaria Terzi *(Boston University)*
- **Mining Contentions from Discussions and Debates** 841
  Arjun Mukherjee, Bing Liu *(University of Illinois at Chicago)*
- **Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow** 850
  Ashton Anderson *(Stanford University)*, Daniel Huttenlocher, Jon Kleinberg *(Cornell University)*, Jure Leskovec *(Stanford University)*

Research Session B2: Outlier and Intrusion Detection

- **Integrating Community Matching and Outlier Detection for Mining Evolutionary Community Outliers** 859
  Manish Gupta *(University of Illinois at Urbana-Champaign)*, Jing Gao *(State Universityof New York, Buffalo)*, Yizhou Sun, Jiawei Han *(University of Illinois at Urbana-Champaign)*

- **Different Slopes for Different Folks: Mining for Exceptional Regression Models with Cook's Distance** 868
  Wouter Duivesteijn *(LIACS, Leiden University)*, Ad Feelders *(ICS, Utrecht University)*, Arno Knobbe *(LIACS, Leiden University)*
- **A Near-Linear Time Approximation Algorithm for Angle-based Outlier Detection in High-dimensional Data** 877
  Ninh Pham, Rasmus Pagh *(IT University of Copenhagen)*
- **Intrusion as (Anti)social Communication: Characterization and Detection** 886
  Qi Ding, Natallia Katenka *(Boston University)*, Paul Barford *(University of Wisconsin - Madison)*, Eric Kolaczyk, Mark Crovella *(Boston University)*

Research Session B3: Feature Selection

- **Robust Multi-Task Feature Learning** 895
  Pinghua Gong *(Tsinghua University)*, Jieping Ye *(Arizona State University)*, Changshui Zhang *(Tsinghua University)*
- **Unsupervised Feature Selection for Linked Social Media Data** 904
  Jiliang Tang, Huan Liu *(Arizona State University)*
- **Model Mining for Robust Feature Selection** 913
  Adam Woznica, Phong Nguyen *(University of Geneva)*, Alexandros Kalousis *(University of Applied Sciences)*
- **Feature Grouping and Selection Over an Undirected Graph** 922
  Sen Yang, Lei Yuan, Ying-Cheng Lai, Xiaotong Shen, Peter Wonka, Jieping Ye *(Arizona State University)*

Research Session B4: Nearest Neighbors

- **Maximum Inner-Product Search Using Cone Trees** 931
  Parikshit Ram, Alexander G. Gray *(Georgia Institute of Technology)*
- **A Probabilistic Model for Multimodal Hash Function Learning** 940
  Yi Zhen, Dit-Yan Yeung *(Hong Kong University of Science and Technology)*
- **On Socio-Spatial Group Query for Location-Based Social Networks** 949
  De-Nian Yang *(Academia Sinica)*, Chih-Ya Shen *(National Taiwan University)*, Wang-Chien Lee *(The Pennsylvania State University)*, Ming-Syan Chen *(National Taiwan University)*
- **Random Forests for Metric Learning with Implicit Pairwise Position Dependence** 958
  Caiming Xiong, David Johnson, Ran Xu, Jason J. Corso *(SUNY at Buffalo)*

Industry/Govt Track B5: Business Intelligence

Session Chair: Sreenivas Gollapudi

- **Empowering Authors to Diagnose Comprehension Burden in Textbooks** 967
  Rakesh Agrawal *(Microsoft Research)*, Sunandan Chakraborty *(New York University)*, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi *(Microsoft Research)*
- **Coupled Behavior Analysis for Capturing Coupling Relationships in Group-based Market Manipulations** 976
  Yin Song, Longbing Cao *(University of Technology Sydney)*, Xindong Wu *(Heifei University of Technology & University of Vermont)*, Gang Wei, Wu Ye *(Shanghai Stock Exchange)*, Wei Ding *(University of Massachusetts Boston)*
- **HySAD: A Semi-Supervised Hybrid Shilling Attack Detector for Trustworthy Product Recommendation** 985
  Zhiang Wu *(Nanjing University of Finance and Economics)*, Junjie Wu *(Beihang University)*, Jie Cao *(Nanjing University of Finance and Economics)*, Dacheng Tao *(University of Technology, Sydney)*
- **Following the Electrons: Methods for Power Management in Commercial Buildings** 994
  Gowtham Bellala, Manish Marwah, Martin Arlitt, Geoff Lyon, Cullen Bash *(Hewlett Packard Laboratories)*

Industrial Practice Expo B6: Session 3

- **Ensembles and Model Delivery for Tax Compliance** 1003
  Graham Williams *(Australian Taxation Office)*
- **Leveraging Predictive Modeling to Reduce Signal Theft in a Multi-Service Organization Environment** 1004
  Seymour Douglas *(Deloitte Financial Advisory Services)*

Research Session C1: Team, Trends, and Social Profiling

- **Capacitated Team Formation Problem on Social Networks** 1005
  Samik Datta, Anirban Majumder *(Bell Labs, Bangalore)*, KVM Naidu *(Walmart Labs, Bangalore)*
- **Finding Trendsetters in Information Networks** 1014
  Diego Saez-Trumper *(Universitat Pompeu Fabra)*,
  Giovanni Comarela, Virgílio Almeida *(Universidade Federal de Minas Gerais)*,
  Ricardo Baeza-Yates *(Yahoo! Research)*, Fabrício Benevenuto *(Universidade Federal de Ouro Preto)*
- **Towards Social User Profiling: Unified and Discriminative Influence Model for Inferring Home Locations** 1023
  Rui Li, Shengjie Wang, Hongbo Deng, Rui Wang *(University of Illinois at Urbana-Champaign)*,
  Kevin Chen-Chuan Chang *(University of Illinois at Urbana-Champaign & Illinois at Singapore)*
- **Event-Based Social Networks: Linking the Online and Offline Social Worlds** 1032
  Xingjie Liu *(The Pennsylvania State University)*, Qi He, Yuanyuan Tian *(IBM Almaden Research Center)*,
  Wang-Chien Lee *(The Pennsylvania State University)*, John McPherson *(IBM Almaden Research Center)*,
  Jiawei Han *(University of Illinois at Urbana-Champaign)*

Research Session C2: Privacy

- **Differential Identifiability** 1041
  Jaewoo Lee, Chris Clifton *(Purdue University)*
- **Anonymizing Set-Valued Data by Nonreciprocal Recoding** 1050
  Mingqiang Xue *(National University of Singapore)*, Panagiotis Karras *(Rutgers University)*,
  Chedy Raïssi *(INRIA)*, Jaideep Vaidya *(Rutgers University)*, Kian-Lee Tan *(National University of Singapore)*
- **Adversarial Support Vector Machine Learning** 1059
  Yan Zhou, Murat Kantarcioglu, Bhavani Thuraisingham *(University of Texas at Dallas)*,
  Bowei Xi *(Purdue University)*

Research Session C3: Supervised Learning Applications

- **Web Image Prediction Using Multivariate Point Processes** 1068
  Gunhee Kim *(Carnegie Mellon University)*, Li Fei-Fei *(Stanford University)*,
  Eric P. Xing *(Carnegie Mellon University)*
- **Transductive Multi-label Ensemble Classification for Protein Function Prediction** 1077
  Guoxian Yu *(South China University of Technology)*,
  Carlotta Domeniconi, Huzefa Rangwala *(George Mason University)*,
  Guoji Zhang, Zhiwen Yu *(South China University of Technology)*
- **Multi-Domain Active Learning for Text Classification** 1086
  Lianghao Li, Xiaoming Jin *(Tsinghua University)*, Sinno Jialin Pan *(Institute for Infocomm Research, Singapore)*,
  Jian-Tao Sun *(Microsoft Research Asia)*
- **Modeling Disease Progression via Fused Sparse Group Lasso** 1095
  Jiayu Zhou, Jun Liu *(Arizona State University)*,
  Vaibhav A. Narayan *(Johnson & Johnson Pharmaceutical Research & Development, LLC)*,
  Jieping Ye *(The Biodesign Institute & Arizona State University)*

Research Session C4: Information Extraction

- **Open Domain Event Extraction from Twitter** 1104
  Alan Ritter, Mausam, Oren Etzioni *(University of Washington)*, Sam Clark *(Decide, Inc.)*

- **Stratified K-Means Clustering Over a Deep Web Data Source** 1113
  Tantan Liu, Gagan Agrawal *(The Ohio State University)*
- **Metro Maps of Science** 1122
  Dafna Shahaf, Carlos Guestrin *(Carnegie Mellon University)*, Eric Horvitz *(Microsoft Research)*
- **Active Sampling for Entity Matching** 1131
  Kedar Bellare, Suresh Iyengar *(Yahoo! Labs)*, Aditya Parameswaran *(Stanford University)*, Vibhor Rastogi *(Yahoo! Research)*

Industry/Govt Track C5: Medical Informatics

Session Chair: Jimeng Sun

- **An Integrated Data Mining Approach to Real-time Clinical Monitoring and Deterioration Warning** 1140
  Yi Mao, Wenlin Chen, Yixin Chen, Chenyang Lu, Marin Kollef, Thomas C. Bailey *(Washington University in St. Louis)*
- **Multi-Source Learning for Joint Analysis of Incomplete Multi-Modality Neuroimaging Data** 1149
  Lei Yuan, Yalin Wang *(Arizona State University)*, Paul M. Thompson *(University of California, Los Angeles)*, Vaibhav A. Narayan *(Johnson & Johnson Pharmaceutical Research & Development, LLC)*, Jieping Ye *(Arizona State University)*
- **RainMon: An Integrated Approach to Mining Bursty Timeseries Monitoring Data** 1158
  Ilari Shafer, Kai Ren, Vishnu Naresh Boddeti, Yoshihisa Abe, Gregory R. Ganger, Christos Faloutsos *(Carnegie Mellon University)*
- **SympGraph: A Framework for Mining Clinical Notes Through Symptom Relation Graphs** 1167
  Parikshit Sondhi *(University of Illinois at Urbana Champaign)*, imeng Sun, Hanghang Tong *(IBM T.J. Watson Research Center)*, ChengXiang Zhai *(University of Illinois at Urbana Champaign)*

Industrial Practice Expo C6: Session 4

- **Experiences and Lessons in Developing Industry-Strength Machine Learning and Data Mining Software** 1176
  Chih-Jen Lin *(National Taiwan University & eBay Research Laboratories)*

Research Session A1: Ads and Video Recommendation

- **Joint Optimization of Bid and Budget Allocation in Sponsored Search** 1177
  Weinan Zhang *(Shanghai Jiao Tong University)*, Ying Zhang *(Nankai University)*, Bin Gao *(Microsoft Research Asia)*, Yong Yu *(Shanghai Jiao Tong University)*, Xiaojie Yuan *(Nankai University)*, Tie-Yan Liu *(Microsoft Research Asia)*
- **The Untold Story of the Clones: Content-Agnostic Factors That Impact YouTube Video Popularity** 1186
  Youmna Borghol *(NICTA & University of New South Wales)*, Sebastien Ardon *(NICTA)*, Niklas Carlsson *(Linköping University)*, Derek Eager *(University of Saskatchewan)*, Anirban Mahanti *(NICTA)*
- **SHALE: An Efficient Algorithm for Allocation of Guaranteed Display Advertising** 1195
  Vijay Bharadwaj *(Netflix)*, Peiji Chen, Wenjing Ma, Chandrashekhar Nagarajan *(Yahoo! Labs)*, John Tomlin *(opTomax Solutions)*, Sergei Vassilvitskii *(Google)*, Erik Vee *(Facebook)*, Jian Yang *(Yahoo! Labs)*
- **Factoring Past Exposure in Display Advertising Targeting** 1204
  Neha Gupta *(University of Maryland, College Park)*, Abhimanyu Das, Sandeep Pandey, Vijay K. Narayanan *(Yahoo! Research)*
- **Online Allocation of Display Ads with Smooth Delivery** 1213
  Anand Bhalgat *(University of Pennsylvania)*, Jon Feldman, Vahab Mirrokni *(Google Inc.)*

Research Session A2: Graph Mining

- **Streaming Graph Partitioning for Large Distributed Graphs** 1222
  Isabelle Stanton *(University of California, Berkeley)*, Gabriel Kliot *(Microsoft Research)*

- **RolX: Structural Role Extraction & Mining in Large Graphs** 1231
  Keith Henderson, Brian Gallagher *(Lawrence Livermore National Laboratory)*,
  Tina Eliassi-Rad *(Rutgers University)*, Hanghang Tong *(IBM T.J.Watson Research Center)*,
  Sugato Basu *(Google Research)*,
  Leman Akoglu, Danai Koutra, Christos Faloutsos, Lei Li *(Carnegie Mellon University)*
- **Fast Algorithms for Maximal Clique Enumeration with Limited Memory** 1240
  James Cheng *(Nanyang Technological University)*, Linhong Zhu *(Institute for Infocomm Research)*,
  Yiping Ke *(Institute of High Performance Computing)*, Shumo Chu *(Nanyang Technological University)*
- **Summarization-based Mining Bipartite Graphs** 1249
  Jing Feng, Xiao He, Bettina Konte, Christian Böhm *(University of Munich)*,
  Claudia Plant *(Florida State University)*
- **Mining Coherent Subgraphs in Multi-Layer Graphs with Edge Labels** 1258
  Brigitte Boden, Stephan Günnemann, Holger Hoffmann, Thomas Seidl *(RWTH Aachen University)*

Research Session A3: Recommendation

- **Circle-Based Recommendation in Online Social Networks** 1267
  Xiwang Yang *(Polytechnic Institute of NYU)*, Harald Steck *(Bell Labs, Alcatel-Lucent)*,
  Yong Liu *(Polytechnic Institute of NYU)*
- **Incorporating Heterogeneous Information for Personalized Tag Recommendation in Social Tagging Systems** 1276
  Wei Feng, Jianyong Wang *(Tsinghua University)*
- **Cross-domain Collaboration Recommendation** 1285
  Jie Tang, Sen Wu *(Tsinghua University)*, Jimeng Sun *(IBM TJ Watson Research Center)*,
  Hang Su *(Tsinghua University)*
- **RecMax: Exploiting Recommender Systems for Fun and Profit** 1294
  Amit Goyal, Laks V. S. Lakshmanan *(University of British Columbia)*
- **Learning Personal + Social Latent Factor Model for Social Recommendation** 1303
  Yelong Shen, Ruoming Jin *(Kent State University)*

Research Session A4: Clustering

- **Two Approaches to Understanding When Constraints Help Clustering** 1312
  Ian Davidson *(University of California, Davis)*
- **Chromatic Correlation Clustering** 1321
  Francesco Bonchi, Aristides Gionis, Francesco Gullo, Antti Ukkonen *(Yahoo! Research - Spain)*
- **Locally-Scaled Spectral Clustering Using Empty Region Graphs** 1330
  Carlos D. Correa, Peter Lindstrom *(Lawrence Livermore National Laboratory)*
- **Active Spectral Clustering via Iterative Uncertainty Reduction** 1339
  Fabian L. Wauthier *(University of California, Berkeley)*, Nebojsa Jojic *(Microsoft Research)*,
  Michael I. Jordan *(University of California, Berkeley)*
- **Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks** 1348
  Yizhou Sun, Brandon Norick, Jiawei Han *(University of Illinois at Urbana-Champaign)*,
  Xifeng Yan *(University of California at Santa Barbara)*, Philip S. Yu *(University of Illinois at Chicago)*,
  Xiao Yu *(University of Illinois at Urbana-Champaign)*

Industry/Govt Track A5: Intelligent Systems

Session Chair: Brian Dalessandro

- **Design Principles of Massive, Robust Prediction Systems** 1357
  Troy Raeder, Ori Stitelman, Brian Dalessandro, Claudia Perlich *(M6D Research)*,
  Foster Provost *(New York University and M6D Research)*
- **PatentMiner: Topic-driven Patent Analysis and Mining** 1366
  Jie Tang, Bo Wang, Yang Yang, Po Hu, Yanting Zhao, Xinyu Yan, Bo Gao, Minlie Huang *(Tsinghua University)*,
  Peng Xu, Weichang Li, Adam K. Usadi *(ExxonMobil Research and Engineering Company)*

- **Storytelling in Entity Networks to Support Intelligence Analysts** 1375
  M. Shahriar Hossain, Patrick Butler *(Virginia Tech)*, Arnold P. Boedihardjo *(U. S. Army Corps of Engineers)*, Naren Ramakrishnan *(Virginia Tech)*
- **A Framework for Robust Discovery of Entity Synonyms** 1384
  Kaushik Chakrabarti, Surajit Chaudhuri, Tao Cheng *(Microsoft Research)*, Dong Xin *(Google)*
- **SmartDispatch: Enabling Efficient Ticket Dispatch in an IT Service Environment** 1393
  Shivali Agarwa, Renuka Sindhgatta, Bikram Sengupta *(IBM Research)*

Asia-Pacific Track A6: Session 3

- **Social Media Data Analysis for Revealing Collective Behaviors** 1402
  Aoying Zhou, Weining Qian, Haixin Ma *(Eastern China Normal University)*
- **Information Processing in Social Networks** 1403
  Ming-Syan Chen *(National Taiwan University)*
- **Understanding Users' Satisfaction for Search Engine Evaluation** 1404
  Gordon Sun *(Tencent Technology)*
- **Similarity Search in Real World Networks** 1405
  Cuiping Li *(Renmin University)*

Research Session B1: Keywords and Documents

- **Large-Scale Learning of Word Relatedness with Constraints** 1406
  Guy Halawi *(Tel Aviv University)*, Gideon Dror, Evgeniy Gabrilovich, Yehuda Koren *(Yahoo! Research)*
- **Latent Association Analysis of Document Pairs** 1415
  Gengxin Miao, Ziyu Guan, Louise Moser, Xifeng Yan *(University of California, Santa Barbara)*, Shu Tao, Nikos Anerousis, Jimeng Sun *(IBM T.J. Watson Research Center)*
- **LIEGE: Link Entities in Web Lists with Knowledge Base** 1424
  Wei Shen, Jianyong Wang *(Tsinghua University)*, Ping Luo, Min Wang *(HP Labs China)*
- **Automatic Taxonomy Construction from Keywords** 1433
  Xueqing Liu *(Microsoft Research Asia & Tsinghua University)*, Yangqiu Song, Shixia Liu, Haixun Wang *(Microsoft Research Asia)*

Research Session B2: Patterns

- **An Enhanced Relevance Criterion for More Concise Supervised Pattern Discovery** 1442
  Henrik Großkreutz *(Fraunhofer IAIS)*, Daniel Paurat *(University of Bonn)*, Stefan Rueping *(Fraunhofer IAIS)*
- **Efficient Frequent Item Counting in Multi-Core Hardware** 1451
  Pratanu Roy, Jens Teubner, Gustavo Alonso *(ETH Zurich)*
- **On Nested Palindromes in Clickstream Data** 1460
  Michel Speiser *(IBM Research - Zurich & ETH Zurich)*, Gianluca Antonini, Abderrahim Labbi *(IBM Research - Zurich)*, Juliana Sutanto *(ETH Zurich)*

Research Session B3: Spatial and Pattern Recognition

- **Mining Discriminative Components with Low-Rank and Sparsity Constraints for Face Recognition** 1469
  Qiang Zhang, Baoxin Li *(Arizona State University)*
- **Fast Algorithms for Comprehensive N-point Correlation Estimates** 1478
  William B. March *(Georgia Institute of Technology)*, Andrew J. Connolly *(University of Washington)*, Alexander G. Gray *(Georgia Institute of Technology)*
- **On "One of the Few" Objects** 1487
  You Wu, Pankaj K. Agarwal *(Duke University)*, Chengkai Li *(The University of Texas at Arlington)*, Jun Yang *(Duke University)*, Cong Yu *(Google Research)*

Demonstration Track

- **BC-PDM: Data Mining, Social Network Analysis and Text Mining System Based on Cloud Computing** 1496
  Le Yu, Jian Zheng, Bin Wu, Bai Wang, Chongwei Shen, Long Qian, Renbo Zhang
  (Beijing University of Posts and Telecommunication)
- **Query-Driven Discovery of Semantically Similar Substructures in Heterogeneous Networks** 1500
  Xiao Yu, Yizhou Sun, Peixiang Zhao, Jiawei Han *(University of Illinois at Urbana-Champaign)*
- **DAGger: Clustering Correlated Uncertain Data (to predict asset failure in energy networks)** 1504
  Dan Olteanu, Sebastiaan J. van Schaik *(University of Oxford)*
- **UFIMT: An Uncertain Frequent Itemset Mining Toolbox** 1508
  Yongxin Tong, Lei Chen *(The Hong Kong University of Science and Technology)*,
  Philip S. Yu *(University of Illinois at Chicago)*
- **Visual Exploration of Collaboration Networks Based on Graph Degeneracy** 1512
  Christos Giatsidis, Klaus Berberich *(École Polytechnique)*, Dimitrios M. Thilikos *(University of Athens)*,
  Michalis Vazirgiannis *(Athens University of Economics and Business)*
- **TourViz: Interactive Visualization of Connection Pathways in Large Graphs** 1516
  Duen Horng (Polo) Chau, Leman Akoglu *(Carnegie Mellon University)*, Jilles Vreeken *(University of Antwerp)*,
  Hanghang Tong *(IBM T.J. Watson Research)*, Christos Faloutsos *(Carnegie Mellon University)*
- **D-INDEX: A Web Environment for Analyzing Dependences Among Scientific Collaborators** 1520
  Claudio Schifanella, Luigi Di Caro *(Università di Torino)*,
  Mario Cataldi, Marie-Aude Aufaure *(École Centrale Paris)*
- **Information Propagation Game: A Tool to Acquire Human Playing Data for Multi-Player Influence Maximization on Social Networks** 1524
  Hung-Hsuan Chen *(The Pennsylvania State University)*,
  Yan-Bin Ciou, Shou-De Lin *(National Taiwan University)*
- **MoodLens: an Emoticon-Based Sentiment Analysis System for Chinese Tweets** 1528
  Jichang Zhao, Li Dong, Junjie Wu, Ke Xu *(Beihang University)*
- **Intelligent Advertising Framework for Digital Signage** 1532
  Phil Tian, Addicam V. Sanjay, Kunapareddy Chiranjeevi, Shahzad Malik Malik *(Intel Corporation)*
- **AssocExplorer: An Association Rule Visualization System for Exploratory Data Analysis** 1536
  Guimei Liu, Andre Suchitra, Haojun Zhang *(National University of Singapore)*,
  Mengling Feng, See-Kiong Ng *(Institute for Infocomm Research)*,
  Limsoon Wong *(National University of Singapore)*
- **GeoSearch: Georeferenced Video Retrieval System** 1540
  Youngwoo Kim, Jinha Kim, Hwanjo Yu *(Pohang University of Science and Technology)*
- **Siren: An Interactive Tool for Mining and Visualizing Geospatial Redescriptions** 1544
  Esther Galbrun *(University of Helsinki)*, Pauli Miettinen *(Max Planck Institute for Informatics)*
- **Navigating Information Facets on Twitter (NIF-T)** 1548
  Shamanth Kumar, Fred Morstatter, Grant Marshall, Huan Liu *(Arizona State University)*,
  Ullas Nambiar *(IBM Research Lab India)*
- **HeteRecom: A Semantic-Based Recommendation System in Heterogeneous Networks** 1552
  Chuan Shi, Chong Zhou *(Beijing University of Posts and Telecommunications)*,
  Xiangnan Kong *(University of Illinois at Chicago)*,
  Philip S. Yu *(University of Illinois at Chicago & King Abdulaziz University)*,
  Gang Liu, Bai Wang *(Beijing University of Posts and Telecommunications)*
- **VOXSUP: A Social Engagement Framework** 1556
  Yusheng Xie, Daniel Honbo *(Voxsup, Inc.)*,
  Kunpeng Zhang, Yu Cheng, Ankit Agrawal *(Northwestern University)*, Alok Choudhary *(Voxsup, Inc.)*

- **A System for Extracting Top-K Lists from the Web** 1560
  Zhixian Zhang, Kenny Q. Zhu *(Shanghai Jiao Tong University)*, Haixun Wang *(Microsoft Research Asia)*
- **EventSearch: A System for Event Discovery and Retrieval on Multi-Type Historical Data** 1564
  Dongdong Shan, Wayne Xin Zhao, Rishan Chen, Baihan Shu, Ziqi Wang, Junjie Yao, Hongfei Yan, Xiaoming Li *(Peking University)*
- **EvaPlanner: An Evacuation Planner with Social-Based Flocking Kinetics** 1568
  Cheng-Te Li, Shou-De Lin *(National Taiwan University)*
- **PubMed Search and Exploration with Real-Time Semantic Network Construction** 1572
  Jinoh Oh, Taehoon Kim, Sun Park, Hwanjo Yu *(Pohang University of Science and Technology)*

Author Index 1576

KDD 2012 Conference Organization

General Chair: Qiang Yang *(Hong Kong University of Science and Technology)*

Honorary Conference Chairs: Lu Ruqian *(Institute of Mathematics, CAS)*
Bo Zhang *(Tsinghua University)*

Advisory Committee Chair: Deyi Li *(Chinese Academy of Engineering)*

Program Chairs: Deepak Agarwal *(LinkedIn)*
Jian Pei *(Simon Fraser University)*

Associate General Chair: Dou Shen *(CityGrid Media)*

Industry & Government Track Chairs: Michael Zeller *(Zementis)*
Hui Xiong *(Rutgers)*

Industry Practice Expo Chairs: Ying Li *(Concurix Corporation)*
Rajesh Parekh *(Groupon)*

Asia Pacific Track Chairs: David Wai-Lok Cheung *(University of Hong Kong)*
Wei Wang *(University of North Carolina at Chapel Hill)*

Workshop Chairs: Zhi-Hua Zhou *(Nanjing University)*
Sofus A. Macskassy *(Information Sciences Institute / USC)*

Tutorial Chairs: Jennifer Neville *(Purdue University)*
Yehuda Koren *(Google)*

Local Arrangements Chairs: Jie Tang *(Tsinghua University)*
Lei Zhang *(Web Search & Mining Group Microsoft Research Asia)*
Chao Liu *(Tencent Inc.)*

Exhibit and Demo Chairs: Tie-Yan Liu *(Microsoft Research Asia)*
Jian-Tau Sun *(Machine Learning group, Microsoft Research Asia)*
Natasha Balac *(University of California, San Diego)*

Industrial Liaison Chairs: Haixun Wang *(Microsoft Research Asia)*
Haifeng Wang *(Baidu)*

KDD Best Paper Awards Chairs: Ke Wang *(Simon Fraser University)*
Rich Caruana *(Microsoft Research)*

SIGKDD 2012 Awards Chair: Ramasamy Uthurusamy *(General Motors Corp.)*

KDD Cup Chairs: Brian Dalessandro *(Media6Degrees)*
Claudia Perlich *(Media6Degrees)*
Gordon Sun *(Tencent Inc.)*
Yading Yue *(Tencent Inc.)*

KDD Summer School: Maosong Sun *(Tsinghua University)*
Du Xiaoyong *(Renmin University of China)*
Hang Li *(Microsoft Research Asia)*

Media Chairs: Ming Hua *(Facebook Inc.)*
Jianyong Wang *(Tsinghua University)*
Rayid Ghani *(Accenture Technology Labs)*
Emilia Palaveeva *(Yantra PR)*

Panel Chairs: Raghu Ramakrishnan *(Microsoft)*
Charles X. Ling *(University of Western Ontario)*

Poster Chairs: Yan Liu *(University of Southern California)*
Ping Luo *(HP Labs)*

Publications Chairs: Alexandru Niculescu-Mizil *(NEC Laboratories America)*
Jingrui He *(IBM T.J. Watson Research Center)*

Registration Chair: Xingquan Zhu *(University of Technology Sydney)*

Social Network Chairs: Ron Bekkerman *(LinkedIn)*
Qiaozhu Mei *(University of Michigan)*

Sponsorship Chairs: Balaji Krishnapuram *(Siemens Medical Solutions USA)*
Shiqiang Yang *(Tsinghua University)*
Peng Cui *(Tsinghua University)*

Student Awards Chairs: Jimeng Sun *(IBM T.J. Watson Research Center)*
Jennifer G. Dy *(Northeastern University)*

Student Volunteer Chairs: Juanzi Li *(Tsinghua University)*
Min-Ling Zhang *(Southeast University)*

Treasurer: Raymond Chi-Wing Wong *(Hong Kong University of Science and Technology)*

Video Chairs: Steven C.H. Hoi *(Nanyang Technological University)*
Shou-de Lin *(National Taiwan University)*

Web Masters: Weike Pan *(Hong Kong University of Science and Technology)*
Guang-Tong Zhou *(Simon Fraser University)*
Xiao Meng *(Simon Fraser University)*

KDD'12 Senior Program Committee Members

Naoki Abe, *IBM*
Rakesh Agrawal, *Microsoft Research*
James Bailey, *The University of Melbourne*
Soumen Chakrabarti, *IIT Mumbai*
Bee-Chung Chen, *Yahoo! Research*
Diane Cook, *Washington State University*
Ian Davidson, *University of California, Davis*
Inderjit Dhillon, *University of Texas, Austin*
Charles Elkan, *University of California, San Diego*
Martin Ester, *Simon Fraser University*
Johannes Gehrke, *Cornell University*
Lise Getoor, *University of Maryland*
Joydeep Ghosh, *University of Texas, Austin*
Bart Goethals, *University of Antwerp*
Dimitrios Gunopulos, *University of Athens*
Jiawei Han, *University of Illinois at Urbana-Champaign*
David Jensen, *University of Massachussets, Amherst*
Christopher Jermaine, *Rice University*
Benjamin Kao, *The University of Hong Kong*
George Karypis, *University of Minnesota*
Sathiya Keethi, *Yahoo!*
Eamonn Keogh, *University of California, Riverside*
Ravi Kumar, *Yahoo! Research*
Laks Lakshmanan, *The University of British Columbia*
Jure Leskovec, *Stanford University*
Hang Li, *Microsoft Research*
Michael Mahoney, *Stanford University*
Nikos Mamoulis, *The University of Hong Kong*
David Mimno, *Princeton University*
Raymond Ng, *The University of British Columbia*
Zai-qing Nie, *Microsoft Research*
Srinivasan Parthasarathy, *The Ohio State University*
Daryl Pregibon, *Google Research*
Saharon Rosset, *Tel Aviv University*
Joerg Sander, *The University of Alberta*
Sunita Sarawagi, *IIT Bombay*
Alex Smola, *Yahoo! Research*
Ramakrishnan Srikant, *Google*
Dacheng Tao, *University of Technology, Sydney*
Andrew Tomkins, *Google Research*
Suresh Venkatsubramanian, *University of Utah*
Jianyong Wang, *Tsinghua University*
Wei Wang, *University of North Carolina at Chapel*
Ji-rong Wen, *Microsoft Research*
Xifeng Yan, *University of California, Santa Barbara*
Jeffrey Yu, *The Chinese University of Hong Kong*
Philip Yu, *University of Illinois at Chicago*
Mohammed Zaki, *Rensselaer Polytechnic Institute*
Changshui Zhang, *Tsinghua University*
Zhi-Hua Zhou, *Nanjing University*

KDD'12 Research Track Program Committee

James Abello, *DIMACS/Rutgers*
Foto Afrati, *National Technical University of Athens*
Charu Aggarwal, *IBM Research-Watson*
Eugene Agichtein, *Emory University*
Christoforos Anagnostopoulos, *University of Cambridge*
Arthur Asuncion, *Google*
Sitaram Asur, *HP Research Labs*
Vassilis Athitsos, *University of Texas at Arlington*
Gustavo Batista, *University of Sao Paulo - USP*
Roberto Bayardo, *Google*
Tanya Berger-Wolf, *University of Illinois, Chicago*
Michael Berthold, *KNIME*
Kanishka Bhaduri, *NASA*
Mikhail Bilenko, *Microsoft Research*
Francesco Bonchi, *Yahoo! Research*
Gavin Brown, *University Manchester*
Deng Cai, *Zhengjian University, China*
Toon Calders, *Technische Universiteit Eindhoven*
Ricardo Campello, *University of Sao Paulo*
Ed Chang, *Google*
Jonathan Chang, *Facebook*
Kevin Chang, *University of Illinois at Urbana-Champaign*
Nitesh Chawla, *University of Notre Dame*
Sanjay Chawla, *University of Sydney*
Lei Chen, *Hong Kong University of Science and Technology*
Ming-Syan Chen, *NTU*
Weizhu Chen, *Microsoft*
Xuewen Chen, *University of Kansas*
Ye Chen, *Microsoft*
Yixin Chen, *Washington University in St Louis*
Zheng Chen, *Microsoft Research*
Zhiyuan Chen, *UMBC*
Hong Cheng, *Chinese University of Hong Kong*
James Cheng, *Nanyang Technological University*
David Cheung, *University of Hong Kong*
Wei Chu, *Microsoft*
Gao Cong, *Nanyang Technological University*
Alfredo Cuzzocrea, *University of Calabria*
Gautam Das, *University of Texas at Arlington*
Anirban Dasgupta, *Yahoo Research*
Umesh Dayal, *HP Labs*
Laura Dietz, *Max Planck Institute and University of Massachusetts Amherst*
Carlotta Domeniconi, *George Mason University*
Guozhu Dong, *Wright State University*
Zhicheng Dou, *Microsoft Research Asia*
Artur Dubrawski, *Carnegie Mellon University*
Murat Dundar, *Indiana University-Purdue University Indianapolis*
Saso Dzeroski, *Jozef Stefan Institute*
Tina Eliassi-Rad, *Rutgers University*
Andrew Fast, *Elder Research*
Ada Fu, *Chinese University of Hong Kong*
Johannes Furnkranz, *TU, Darmstadt*
Joao Gama, *University of Porto*
Jing Gao, *Univ. of Buffalo*
Rainer Gemulla, *Max-Planck Institute*
Amol Ghoting, *IBM TJ Watson Research Center*
C. Lee Giles, *Penn State University*
Aristides Gionis, *Yahoo! Research*
Aris Gkoulalas-Divanis, *IBM Research - Zurich*
Sharad Goel, *Yahoo! Research*
Michaela Goetz, *Twitter*
Rahul Gupta, *Google Research*
Qi He, *IBM Almaden Research Center*
Tin Kam Ho, *AT&T Bell Labs*
Ming Hua, *Facebook*
Jun Huan, *University of Kansas*
Jian Huang, *Google*
Jimmy Huang, *York University*
Shuiwang Ji, *Old Dominion University*
Yan Jia, *National University of Defense Technology*
Daxin Jiang, *Microsoft*
Rong Jin, *Michigan State University*
Ruoming Jin, *Kent State University*
Xiaoming Jin, *Tsinghua University*
Panos Kalnis, *KAUST*
Anitha Kannan, *Microsoft Research*
Panagiotis Karras, *Rutgers Business School*
Latifur Khan, *University of Texas at Dallas*
Daniel Kifer, *Penn State University*
Irwin King, *The Chinese University of Hong Kong*
George Kollios, *Boston University*
Nick Koudas, *University of Toronto and Sysomos Inc*
Akhil Kumar, *Penn State University*
Diane Lambert, *Google Research*

Thodoros Lappas, *Boston University*
Dongwon Lee, *Penn State University*
Hongrae Lee, *Google Research*
Mong Li Lee, *NUS*
Ronny Lempel, *Yahoo! Research*
Carson Leung, *University of Manitoba*
Guoliang Li, *Tsinghua University*
Jinyan Li, *University of Technology Sydney*
Jiuyong Li, *University of South Australia*
Lihong Li, *Yahoo! Research*
Ee Peng Lim, *Singapore Management University*
Chih Jen Lin, *National Taiwan University*
Jessica Lin, *George Mason University*
Xuemin Lin, *University of New South Wales*
Charles Ling, *University of Western Ontario*
Bing Liu, *University of Illinois at Chicago*
Chao Liu, *Microsoft Research*
Huan Liu, *Arizona State*
Jun Liu, *Siemens Corporate Research at Princeton*
Tie-Yan Liu, *Microsoft Research*
Yan Liu, *University of Southern California*
Eric Lo, *The Hong Kong Polytechnic University*
Stefano Lonardi, *University of California, Riverside*
Aurelie Lozano, *IBM Research*
Chang-Tien Lu, *Virginia Tech*
Jiebo Luo, *University of Rochester*
Shiqian Ma, *Columbia University*
Mohammad Mahdian, *Google*
Cecilia Mascolo, *University of Cambridge*
Prem Melville, *IBM TJ Watson Research Center*
Aditya Menon, *University of California, San Diego*
Donald Metzler, *University of Southern California*
George Michailidis, *University of Michigan*
Abdullah Mueen, *Microsoft Research*
S. Muthu Muthukrishnan, *Rutgers*
Olfa Nasraoui, *University of Louisville*
Jennifer Neville, *Purdue University*
David Newman, *University of California, Irvine*
Siegfried Nijssen, *Katholieke Universiteit Leuven*
Tim Oates, *University of Maryland Baltimore County*
Sandeep Pandey, *Yahoo! Research*
Spiros Papadimitriou, *IBM*
Panagiotis Papapetrou, *Aalto University*
Laurence Park, *University of Western Sydney*
Stott Parker, *UCLA*
Dino Pedreschi, *KDD Lab, University of Pisa, Italy*
Bernhard Pfahringer, *University of Waikato*
Jeff Phillips, *University of Utah*
Claudia Plant, *Florida State University*
John Platt, *Microsoft Research*
Kunal Punera, *Yahoo! Research*
Chedy Raissi, *INRIA*
Daniel Ramage, *Stanford University*
Naren Ramakrishnan, *Virginia Tech*
Louiqa Raschid, *University of Maryland*
Parisa Rashidi, *University of Florida*
Rajeev Rastogi, *Yahoo! Labs*
Chandan Reddy, *Wayne State*
Matthias Renz, *Ludwig-Maximilians-Universitat Munchen*
Matthew Richardson, *Microsoft Research*
Mirek Riedewald, *Northeastern University*
Venu Satuluri, *Ohio State University*
Thomas Seidl, *RWTH Aachen*
Sundararajan Sellamanickam, *Yahoo! Labs, Bangalore*
Jialie Shen, *Singapore Management University*
Kyuseok Shim, *Seoul National University*
Luo Si, *Purdue*
Vikas Sindhwani, *IBM Research-Watson*
Ambuj Singh, *University of California, Santa Barbara*
D. Sivakumar, *Yahoo! Research*
Myra Spiliopoulou, *Otto-von-Guericke-Universitaet Magdeburg*
Rohini Srihari, *University at Buffalo, State University of New York*
Michael Steinbach, *University of Minnesota*
Nick Street, *University of Iowa*
Jimeng Sun, *IBM TJ Watson Research Center*
Pang-Ning Tan, *Michigan State University*
Jie Tang, *Tsinghua University, China*
Yufei Tao, *Chinese University of Hong Kong*
Shirish Tatikonda, *IBM Almaden Research Center*
Manolis Terrovitis, *IMIS, Athena, Greece*
Evimaria Terzi, *Boston University*
Alex Thomo, *University of Victoria*
Hanghang Tong, *IBM Research*
Vincent Tseng, *National Cheng Kung University*
Leong Hou U, *University of Macau*
Duygu Ucar, *Stanford University*
Antti Ukkonen, *Yahoo! Research*
Lyle Ungar, *University of Pennsylvania*
Jaideep Vaidya, *Rutgers Univ.*

Hamed Valizadegan, *University of Pittsburgh*
Michalis Vlachos, *IBM Research*
Chris Volinsky, *AT&T Labs - Research*
Jilles Vreeken, *University of Antwerp, Belgium*
Haixun Wang, *Microsoft Asia*
Ke Wang, *Simon Fraser University*
Xuanhui Wang, *Yahoo! Research*
Markus Weimer, *Yahoo! Research*
Sholom Weiss, *IBM Research*
Tony Wirth, *University of Melbourne*
Ran Wolff, *University of Haifa*
Raymond Chi-Wing Wong, *Hong Kong University of Science and Technology*
Junjie Wu, *Beihang University*
Xindong Wu, *University of Vermont*
Xintao Wu, *University of North Carolina at Charlotte*
Xiaokui Xiao, *Nanyang Technological University*
Hui Xiong, *Rutgers, the State University of New Jersey*
Xiaowei Xu, *University of Arkansas at Little Rock*
Guirong Xue, *Shanghai Jiaotong University*
Kenji Yamanishi, *University of Tokyo*
Rong Yan, *Facebook*
Jieping Ye, *Arizona State University*
Mi-Yen Yeh, *Academia Sinica*
Kevin Yip, *The Chinese University of Hong Kong*
Hwanjo Yu, *Pohang University of Science and Technology*
Lei Yu, *Binghamton University*
Shipeng Yu, *Siemens Medical Solutions*
Bianca Zadrozny, *Fluminense Federal University*
Osmar Zaiane, *University of Alberta*
Carlo Zaniolo, *University of California, Los Angeles*
Hongyuan Zha, *Georgia Tech*
Aidong Zhang, *University at Buffalo*
Dell Zhang, *University of London*
Liang Zhang, *Yahoo! Research*
Nan Zhang, *George Washington University*
Zhongfei Mark Zhang, *SUNY Binghamton*
Wenjun Zhou, *University of Tennessee - Knoxville*
Xingquan Zhu, *University of Technology, Sydney*
Arthur Zimek, *Ludwig-Maximilians-Universitat Munchen*

KDD'12 Industry & Government Track Program Committee Members

Dean Abbott, *Abbott Analytics*
John Akred, *Accenture*
Michael Berthold, *KNIME*
Kanishka Bhaduri, *NASA*
Eric Bloedorn, *MITRE - USA*
Rainer Burkhardt, *JgenCy*
Longbing Cao, *University of Technology Sydney*
Huanhuan Cao, *Nokia Research Center*
Haifeng Chen, *NEC Research Lab*
Enhong Chen, *University of Science and Technology of China*
James Cox, *SAS*
Kamalika Das, *NASA Ames Research Center*
Chao Deng, *China Mobile*
Wei Ding, *University of Massachusetts Boston*
David Duling, *SAS*
Wei Fan, *IBM Research*
Nancy Grady, *Science Applications International Corporation*
Simon Gyorgy, *Mayo Clinic*
Jianying Hu, *IBM T.J. Watson Research Center*
Ming Hua, *Facebook*
Sergei Israilev, *Jabiru Ventures*
Daxin Jiang, *Microsoft*
Bin Jiang, *Facebook*
Hongxia Jin, *IBM Research*
Yu Kai, *Baidu, Inc.*
Mei Kobayashi, *IBM Research Lab (Tokyo)*
Rui Kuang, *University of Minnesota*
Mohit Kumar, *Accenture*
Rick Lawrence, *IBM Research-Watson*
Aleksandar Lazarevic, *Travelers*
Christoph Lingenfelder, *IBM*
Nathan Nan Liu, *Hong Kong UST*
Fei Liu, *Bosch Research*
Jiebo Luo, *University of Rochester*
Ping Luo, *HP Lab*
Gideon Mann, *Google*
Manish Marwah, *HP Labs*
Gabor Melli, *VigLink*
Theodore Omtzigt, *Stillwater Supercomputing*
Gaurav Pandey, *Mount Sinai School of Medicine*
Mark Plutowski, *Flipboard*
Karl Rexer, *Rexer Analytics*
Mark Richards, *ISO*
Neal Rothleder, *MITRE*
Ramendra Sahoo, *Citi*
D Sculley, *Google*
Ted Senator, *SAIC*
Abhishek Sharma, *NEC Research Lab*
Ali Shaykhian, *NASA*
Manas Somaiya, *eBay*
Jacob Spoelstra, *Opera Solutions*
Jian-Tao Sun, *Microsoft Research*
AC Surendran, *Microsoft*
Arun Swami, *Google*
Shu Tao, *IBM Research*
Wayne Thompson, *SAS*
Hanghang Tong, *IBM Research*
Ramasamy Uthurusamy, *GM*
Peter van der Putten, *Pegasystems & Leiden University*
Ranga Vatsavai, *Oak Ridge National Labs*
Andrew Veal, *Government Communications HQ, UK*
Dave Vennergrund, *Delta Solutions and Technologies, Inc.*
Fei Wang, *IBM Research*
Stefan Weber, *Siemens AG*
Zhen Wen, *IBM Research*
Junjie Wu, *Beihang University*
Xing Xie, *Microsoft Research Asia*
Jiangang Yu, *Microsoft*
Kai Yu, *NEC Laboratories America*
Ruofei Zhang, *Yahoo Labs*
Zheng Zhao, *SAS*
Yu Zheng, *Microsoft*
Zijian Zheng, *Microsoft*
Zhaohui Zheng, *Yahoo!*
Scott Zoldi, *FICO*

KDD'12 Research Track External Reviewers

Mohammad-Ali Abbasi
Jaafar Ben Abdallah
Artur Abdullin
Zubin Abraham
Ferit Akova
Darko Aleksovski
Can Altinigneli
Saeed Amizadeh
Rajul Anand
Sam Anzaroot
Gowtham Atlure
Gowtham Atluri
Anton Bakalov
Lakshika Balasuriya
Honhji Bao
Iyad Batal
Christian Beecks
Gabriel Bender
Thomas Bernecker
Kasturi Bhattacharjee
Jinbo Bi
Hoda Bidkhori
Sam Blasiak
Brigitte Boden
Petko Bogdanov
Ivan Brugere
Fahrettin Cakir
Xin Cao
Lei Cen
Yashu Chamber
Kung-Hua Chang
Xi Chen
Chen Cheng
Si-Chi Chin
Byron Choi
Shumo Chu
Prakash Mandayam Comar
Michele Coscia
Aaron Crandall
Andrew Cron
Bruno Crémilleux
Decheng Dai
Kamalika Das
Mahashweta Das
Ritendra Datta
Elnaz Delpisheh
Sanjoy Dey
Arman Didandeh
Ivica Dimitrovski
Martin Dimkovski
Wenkui Ding
Anton Dries
Lian Duan
Denisa Duma
Narayanan Edakunni
Dora Erdos
Ines Faerber
James Faghmous
S M Faisal
Hadi Fanaee
Lujun Fang
Zheng Fang
Gang Fang
Hongliang Fei
Zhuo Feng
Jing Feng
Carlos Ferreira
Neil Fore
Sergej Fries
Glenn Fung
Stanislav Funiak
Bin Gao
Chuancong Gao
Huiji Gao
Priyanka Garg
Liang Ge
Rossetti Giulio
Rafael Giusti
Olga Gkountouna
Sebastian Goebl
Behzad Golshan
naiyang Guan
Ziyu Guan
Francesco Gullo
Pritam Gundecha
Tias Guns
Fan Guo
Chetan Gupta
Stephan Günnemann
Habiba
MohammadTaghi HajiAghayi
Yue Han
Qian Han
Hyoil Han
Elena Harris
Marwan Hassani
Basheer Hawwash
Xiao He
Di He
Jingrui He
Eric Heim
Minh Hoang
Thanh Lam Hoang
Ryan Hoens
Yuan Hong
Dzung Hong
Xia Hu
Weishu Hu
Yifan Hu
Xin Huang
Dino Ienco
Elena Ikonomovska
Allison Irvine
Anca Ivanescu
Arun Iyer
Ruoyi Jiang
Mehdi Kargar
Shiva Kasiviswanathan
Jaya Kawale
Rohan Khade
Mohammad Khoshneshin
Younghoon Kim
Woo-Cheol Kim
Hyun-Woo Kim
Steffen Kirchhoff
Arno Knobbe
Dragi Kocev
Bettina Konte
Parisa Kordjamshidi
Petr Kosina
NgaiMeng Kou
Philipp Kranen

Hardy Kremer
Georg Krempl
Peer Kröger
Da Kuang
Brian Kulis
Arjun Kumar
Shamanth Kumar
Choudur Lakshminarayan
Diane Lambert
Sean Landman
Nick Larusso
Chao-Hui Lee
Jeonghoon Lee
Victor Lee
Yuhong Lee
Matthijs Van Leeuwen
Jian Li
Ming Li
Wu-Jun Li
Nan Li
Yingming Li
Xiao Li
Fangtao Li
LongJie Li
Baichuan Li
John Liagouris
Shou-de Lin
Guang Ling
Lin Liu
Yi Liu
Haibin Liu
Sitong Liu
Lei Liu
Nathan Liu
Bo Liu
Steven Loscalzo
Grigorios Loukides
Cécile Low-Kam
Eric Hsueh-Chan Lu
Yan Luo
Kathy Macropol
Gjorgji Madjarov
Das Mahashweta
Son T. Mai
Julien Mairal
Charalampos Mavroforakis
Kyle McQuisten
Aranyak Mehta
Iñigo Mendialdua
Jun Miao
Pauli Miettinen
Iris Miliaraki
Darakhshan Mir
Seyed Mirebrahim
Varun Mithal
Samaneh Moghaddam
Anna Monreale
Gianmarco De Francisci Morales
Fred Morstatter
Vinod Nair
Abbas Nayebi
Vincenzo Nicosia
Johannes Niedermayer
Sadegh Nobari
Eirini Ntoutsi
Laurie O'Connor
Omar Odibat
Marcia Oliveira
Jeffrey Junfeng Pan
Pance Panov
Vanja Paunic
Diego Pennacchioli
Marc Plantevit
Michael Plavinski
Yevgeniy Podolyan
Anton Polishko
Miao Qiao
Tao Qin
Lu Qin
Nguyen Quang
Brian Quanz
Alain Rakotomamonjy
Sayan Ranu
Vikas Raykar
Mike Rechenthin
Rabbany K. Reihaneh
Matt Revelle
Elaine Ribeiro
Carlos Rojas
Giulio Rossetti
Senjuti Basu Roy
Yiye Ruan
Brian Ruttenberg
Senay Yasar Saglam
Tanwistha Saha
Ankan Saha
Shaghayegh Sahebi
Esin Saka
Hamman Samuel
Thomas Sandholm
Cicero Nogueira dos Santos
Tamas Sarlos
Christoph Sawade
Hassan Sayyadi
James Schaffer
Leander Schietgat
Johannes Schneider
Erich Schubert
Matthias Schubert
Yu-Keng Shih
Diego Furtado Silva
Michael Simpson
Ayush Singhal
Nick Skapura
Ivica Slavkov
Yangqiu Song
Vinicius Mourao Alves de Souza
Sargur Srihari
Daniela Stojanova
Ja-Hwung Su
Amarnag Subramanya
Fabian Suchanek
Aixin Sun
Khaled Ahsan Talukder
Syed K. Tanbeer
Jiliang Tang
Chayant Tantipathananandh
Vahid Taslimi
Maryam Tavafi
Yasushi Terazono
Saravanan Thirumuruganathan
Brian Thompson
Ryota Tomioka
Aneta Trajanov
Roberto Trasarti
Pucktada Treeratpituk
Honglei Tu
Vladimir Vacic
Thanasis Vergoulis
Davide Verzotto
Krishnamurthy Viswanathan

Peter Wackersreuther
Jiannan Wang
Qifan Wang
Taifeng Wang
Wenjun Wang
Wen Wang
Ye Wang
Xufei Wang
Jia Wang
Guan Wang
Robert Wright
Huanhuan Wu
Shanchan Wu
Leting Wu
shiming Xiang
Guang Xiang
Chenyan Xiong
Yi Xu
Zhiqiang Xu
Jianpeng Xu
Hui Yan
Haiqin Yang
Tao Yang
Sen Yang
Jaewon Yang
Ming Yang
Yiyang Yang
Chen Yang
Wen Yao
Jihang Ye
Zheng Ye
Beta Chi-Lap Yip
Man Lung Yiu
Shengkang Yu
Xiaofeng Yu
Guoxian Yu
Yintao Yu
Xiao Yu
Jun ZHANG
Zhihua ZHang
Reza Zafarani
Jesin Zakaria
Bernard Zenko
Mingyang Zhang
Yuchen Zhang
Zhihua Zhang
Yang Zhang
Wenjie Zhang
Peng Zhang
Chenyi Zhang
Min-Ling Zhang
Jiashu Zhao
Xueyi Zhao
Kang Zhao
Chao Zhou
Qin Zhou
Zhuojie Zhou
Ke Zhou
Zhe Zhu
Morteza Zihayat
Max Zimmermann
Albrecht Zimmermann
Bo Zong
Haitao Zou
Andreas Züfle

ACM SIGKDD KDD 2012 Sponsors & Supporters

Sponsors:

Supporters


CITYGRID.com


CRC Press
Taylor & Francis Group
Springer


ZEMENTIS
Adaptive Decision Technology


technicolor


MORGAN & CLAYPOOL
PUBLISHERS