The purpose of the SIAM Activity Group on Data Mining and Analytics (SIAG/DMA) is to advance the mathematics of data mining, to highlight the importance and benefits of the application of data mining, and to identify and explore the connections between data mining and other applied sciences. The activity group organizes the yearly SIAM International Conference on Data Mining (SDM), organizes minisymposia at the SIAM Annual Meeting, and maintains a membership directory and electronic mailing list.

This conference is held in cooperation with the American Statistical Association.
Table of Contents

Program-At-A-Glance................................... Fold out section
General Information.................................................. 2
Get-togethers............................................................... 7
Invited Plenary Presentations.............................. 10
Tutorials ................................................................. 11
Workshops............................................................... 12
Program Schedule .................................................. 15
Poster Sessions........................................................ 21 & 25
Abstracts ................................................................. 29
Speaker and Organizer Index....................................... 57
Conference Budget…Inside Back Cover
Hotel Meeting Room Map…Back Cover

Organizing Committee

Steering Committee Chair
Srinivasan Parthasarathy
Ohio State University, USA

Steering Committee
Chid Apte
IBM T.J. Watson Research Center, USA
Christos Faloutsos
Carnegie Mellon University, USA
Joydeep Ghosh
The University of Texas at Austin, USA
Jiawei Han
University of Illinois at Urbana-Champaign, USA
Chandrika Kamath
Lawrence Livermore National Laboratory, USA
Vipin Kumar
University of Minnesota, USA
Haesun Park
Georgia Institute of Technology, USA
Srinivasan Parthasarathy
Ohio State University, USA
Qiang Yang
Hong Kong University of Science and Technology, USA
Philip Yu
University of Illinois at Chicago, USA

Conference Co-Chairs
Carlotta Domeniconi
George Mason University, USA

Dimitrios Gunopulos
University of Athens, Greece

Program Co-Chairs
Nitesh Chawla
University of Notre Dame, USA
Wei Wang
University of California, Los Angeles, USA

Workshops Co-Chairs
Leman Akoglu
Carnegie Mellon University, USA
Arthur Zimek
University of Southern Denmark, Denmark

Tutorials Chair
Jilles Vreeken
Saarland University and Max Planck Institute for Informatics, Germany

Doctoral Forum Chair
Shuiwang Ji
Washington State University, USA

Sponsorship Co-Chairs
James Faghmous
Cahn School of Medicine, Mount Sinai, USA
Jiliang Tang
Yahoo Research, USA

Panels Chair
Chris Jermaine
Rice University, USA

Publicity Co-Chairs
Danai Koutra
University of Michigan, USA
Matthijs van Leeuwen
Leiden University, The Netherlands
Panagiotis Papapetrou
Stockholm University, Sweden

Awards Chair
Zoran Obradovic
Temple University, USA

Senior Program Committee
Dino Pedreschi
University of Pisa, Italy
Bianca Zadrozny
IBM-Research, Brazil

Charu Aggarwal
IBM T.J. Watson Research Center, USA
Leman Akoglu
Carnegie Mellon University, USA
Arindam Banerjee
University of Minnesota, USA
Ron Bekkerman
University of Haifa, Israel
Tanya Berger-Wolf
University of Illinois, Chicago, USA
Bettina Bernedt
KU Leuven, Belgium
Francesco Bonchi
Yahoo! Barcelona, Spain
Sanjay Chawla
University of Sydney, Australia
Carlotta Domeniconi
George Mason University, USA
Christos Faloutsos
Carnegie Mellon University, USA
Joao Gama
University of Porto, Portugal
Aristides Gionis
Aalto University, Finland
David Gleich
Purdue University, USA
Jiawei Han
University of Illinois Urbana Champaign, USA
Pallika Kanani
Oracle Labs, USA
Eamonn Keogh
University of California Riverside, USA
Chih Jen Lin
National Taiwan University, Taiwan
Wagner Meira
Universidade Federal de Minas Gerais, Brazil
Ernestina Mensalvas
Universidad Politecnica de Madrid, Spain
Zoran Obradovic
Temple University, USA
Srinivasan Parthasarathy
Ohio State University, USA
Jian Pei
Simon Fraser University, Canada
B. Aditya Prakash
Virginia Tech, USA
Rajeev Rastogi
Amazon, India
<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Celine Robarde</td>
<td>National Institute of Applied Science, France</td>
</tr>
<tr>
<td>Mohak Shah</td>
<td>Bosch, USA</td>
</tr>
<tr>
<td>Kyuseok Shim</td>
<td>Seoul National University, Korea</td>
</tr>
<tr>
<td>Arno Siebes</td>
<td>Universiteit Utrecht, Netherlands</td>
</tr>
<tr>
<td>Masashi Sugiyama</td>
<td>RIKEN/University of Tokyo, Japan</td>
</tr>
<tr>
<td>Yizhou Sun</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>Hui Xiong</td>
<td>Rutgers University, USA</td>
</tr>
<tr>
<td>Jeffrey Yu</td>
<td>Chinese University of Hong Kong, Hong Kong</td>
</tr>
<tr>
<td>Zhi-Hua Zhou</td>
<td>Nanjing University, China</td>
</tr>
<tr>
<td>Feng Chen</td>
<td>University at Albany, SUNY</td>
</tr>
<tr>
<td>Hailfeng Chen</td>
<td>NEC Research Lab, USA</td>
</tr>
<tr>
<td>Jianhui Chen</td>
<td>Yahoo Labs</td>
</tr>
<tr>
<td>Wei Chen</td>
<td>Microsoft Research, Asia</td>
</tr>
<tr>
<td>Hong Cheng</td>
<td>Hong Kong Chinese University, Hong Kong</td>
</tr>
<tr>
<td>Jian Cheng</td>
<td>Chinese Academy of Science, China</td>
</tr>
<tr>
<td>Wei Cheng</td>
<td>NEC Lab, USA</td>
</tr>
<tr>
<td>Yu Cheng</td>
<td>IBM T.J. Watson Research Center, USA</td>
</tr>
<tr>
<td>Malolan Chettur</td>
<td>IBM Research, India</td>
</tr>
<tr>
<td>Diane Cook</td>
<td>Washington State University, USA</td>
</tr>
<tr>
<td>Robson Cordeiro</td>
<td>Universidade de Sao Paulo, Brazil</td>
</tr>
<tr>
<td>Peng Cui</td>
<td>Tsinghua University, China</td>
</tr>
<tr>
<td>Alfred Cuzzocrea</td>
<td>University of Calabria, Italy</td>
</tr>
<tr>
<td>Xuan-Hong Dang</td>
<td>University of California Santa Barbara, USA</td>
</tr>
<tr>
<td>Mahashweta Das</td>
<td>HP Labs, USA</td>
</tr>
<tr>
<td>Hasan Davulcu</td>
<td>Arizona State University, USA</td>
</tr>
<tr>
<td>Yuxiao Dong</td>
<td>University of Notre Dame, USA</td>
</tr>
<tr>
<td>Haimonti Dutta</td>
<td>Columbia University, USA</td>
</tr>
<tr>
<td>William Eberle</td>
<td>Tennessee Technological University, USA</td>
</tr>
<tr>
<td>Dora Erdos</td>
<td>Boston University, USA</td>
</tr>
<tr>
<td>Lisa Friedland</td>
<td>Northeastern University, USA</td>
</tr>
<tr>
<td>Sorelle Friedler</td>
<td>Haverford College, USA</td>
</tr>
<tr>
<td>Yanjie Fu</td>
<td>Missouri University of Science and Technology, USA</td>
</tr>
<tr>
<td>Yun Fu</td>
<td>Northeastern University, USA</td>
</tr>
<tr>
<td>Amita Gajewar</td>
<td>Microsoft Corp, USA</td>
</tr>
<tr>
<td>Jing Gao</td>
<td>University of Buffalo, USA</td>
</tr>
<tr>
<td>Mohamed Ghalwash</td>
<td>Temple University, USA</td>
</tr>
<tr>
<td>Manuel Gomez-Rodriguez</td>
<td>Max Planck Institute, Germany</td>
</tr>
<tr>
<td>Quanquan Gu</td>
<td>University of Virginia, USA</td>
</tr>
<tr>
<td>Ziyu Guan</td>
<td>Northwest University, China</td>
</tr>
<tr>
<td>Jie Gui</td>
<td>Chinese Academy of Science, China</td>
</tr>
<tr>
<td>Francesco Gullo</td>
<td>Yahoo! Research, Spain</td>
</tr>
<tr>
<td>Stephan Günemann</td>
<td>Technical University of Munich, Germany</td>
</tr>
<tr>
<td>Yuan Hao</td>
<td>Google, USA</td>
</tr>
<tr>
<td>Xia Hu</td>
<td>Arizona State University, USA</td>
</tr>
<tr>
<td>Jun (Luke) Huan</td>
<td>University of Kansas, USA</td>
</tr>
<tr>
<td>Heng Huang</td>
<td>University of Texas at Arlington, USA</td>
</tr>
<tr>
<td>Sheng-Jun Huang</td>
<td>Nanjing University of Aeronautics and Astronomy, China</td>
</tr>
<tr>
<td>Xin Huang</td>
<td>University of British Columbia, Canada</td>
</tr>
<tr>
<td>Bo Jin</td>
<td>Dalian University of Technology, China</td>
</tr>
<tr>
<td>Reid Johnson</td>
<td>University of Notre Dame, USA</td>
</tr>
<tr>
<td>Chandrika Kamath</td>
<td>LLNL, USA</td>
</tr>
<tr>
<td>Murat Kantarcio glu</td>
<td>University of Texas, Dallas, USA</td>
</tr>
<tr>
<td>Masayuki Karasuyama</td>
<td>Nagoya Institute of Technology, Japan</td>
</tr>
<tr>
<td>George Karypis</td>
<td>University of Minnesota, USA</td>
</tr>
<tr>
<td>Philip Kegelmeyer</td>
<td>Sandia National Labs, USA</td>
</tr>
<tr>
<td>Daniel Keren</td>
<td>Haifa University, Israel</td>
</tr>
<tr>
<td>Name</td>
<td>Affiliation</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Deguang Kong</td>
<td>Samsung, USA</td>
</tr>
<tr>
<td>Xiangnan Kong</td>
<td>Worcester Polytechnic Institute, USA</td>
</tr>
<tr>
<td>Danai Koutra</td>
<td>University of Michigan, USA</td>
</tr>
<tr>
<td>Georg Krempl</td>
<td>Otto-von-Guericke Univ. Magdeburg, Germany</td>
</tr>
<tr>
<td>Cheng-Te Li</td>
<td>Academic Sinica, Taiwan</td>
</tr>
<tr>
<td>Geng Li</td>
<td>Yahoo!, USA</td>
</tr>
<tr>
<td>Jiuyong Li</td>
<td>University of South Australia, Australia</td>
</tr>
<tr>
<td>Rong-Hua Li</td>
<td>Shenzhen University, China</td>
</tr>
<tr>
<td>Shuai Li</td>
<td>University of Cambridge, United Kingdom</td>
</tr>
<tr>
<td>Yuefeng Li</td>
<td>Queensland University of Technology, Australia</td>
</tr>
<tr>
<td>Zhenhui Li</td>
<td>Penn State University, USA</td>
</tr>
<tr>
<td>Jessica Lin</td>
<td>George Mason University, USA</td>
</tr>
<tr>
<td>Shou-de Lin</td>
<td>National Taiwan University, Taiwan</td>
</tr>
<tr>
<td>Xuemin Lin</td>
<td>New South Wales University, Australia</td>
</tr>
<tr>
<td>Bin Liu</td>
<td>IBM Watson, USA</td>
</tr>
<tr>
<td>Chuanren Liu</td>
<td>Drexel University, USA</td>
</tr>
<tr>
<td>Huan Liu</td>
<td>Arizona State University, USA</td>
</tr>
<tr>
<td>Tongliang Liu</td>
<td>University of Technology Sydney, Australia</td>
</tr>
<tr>
<td>Yan Liu</td>
<td>University of Southern California, USA</td>
</tr>
<tr>
<td>David Lo</td>
<td>Singapore Management University, Singapore</td>
</tr>
<tr>
<td>Aurelie Lozano</td>
<td>IBM Research, USA</td>
</tr>
<tr>
<td>Chang-Tien Lu</td>
<td>Virginia Tech, USA</td>
</tr>
<tr>
<td>Shuai Ma</td>
<td>Beihang University, China</td>
</tr>
<tr>
<td>Arun Maiya</td>
<td>Institute for Defense Analysis, USA</td>
</tr>
<tr>
<td>Julian McAuley</td>
<td>University of California, San Diego, USA</td>
</tr>
<tr>
<td>Luiz Merschmann</td>
<td>Universidade Federal de Ouro Preto (UFOP), Brazil</td>
</tr>
<tr>
<td>Pauli Miettinen</td>
<td>Max-Planck-Institute for Informatics, Germany</td>
</tr>
<tr>
<td>Benjamin Miller</td>
<td>MIT Lincoln Laboratory, USA</td>
</tr>
<tr>
<td>Abdullah Mueen</td>
<td>University of New Mexico, USA</td>
</tr>
<tr>
<td>Vinh Nguyen</td>
<td>University of Melbourne, Australia</td>
</tr>
<tr>
<td>Xia Ning</td>
<td>Indiana University - Purdue University, USA</td>
</tr>
<tr>
<td>Eduardo Ogasawara</td>
<td>Centro Federal de Educação Tecnológica (CEFET-RJ), Brazil</td>
</tr>
<tr>
<td>Nikunj Oza</td>
<td>NASA Ames Research Center, USA</td>
</tr>
<tr>
<td>Aline Paes</td>
<td>Universidade Federal Fluminense (UFF), Brazil</td>
</tr>
<tr>
<td>Evangelos Papalexakis</td>
<td>University of California, Riverside, USA</td>
</tr>
<tr>
<td>Douglass Parker</td>
<td>University of California Los Angeles, USA</td>
</tr>
<tr>
<td>Ruggero Pensa</td>
<td>University of Torino, Italy</td>
</tr>
<tr>
<td>Bryan Perozzi</td>
<td>Google, USA</td>
</tr>
<tr>
<td>Marc Plantefit</td>
<td>University de Lyon, France</td>
</tr>
<tr>
<td>Buyue Qian</td>
<td>IBM T. J. Watson Research Center, USA</td>
</tr>
<tr>
<td>Chao Qian</td>
<td>University of Science and Technology of China</td>
</tr>
<tr>
<td>Mu Qiao</td>
<td>IBM, USA</td>
</tr>
<tr>
<td>Zhiwei Qin</td>
<td>Walmart Labs, USA</td>
</tr>
<tr>
<td>Chedy Raissi</td>
<td>LORIA INRIA, France</td>
</tr>
<tr>
<td>Arti Ramesh</td>
<td>SUNY Binghamton, USA</td>
</tr>
<tr>
<td>Balaraman Ravindran</td>
<td>IIT-Madras, India</td>
</tr>
<tr>
<td>Alejandro Rodriguez</td>
<td>University of Madrid, Spain</td>
</tr>
<tr>
<td>Mahsa Saleh</td>
<td>IBM, Australia</td>
</tr>
<tr>
<td>Saket Sathe</td>
<td>IBM T. J. Watson Research Center, USA</td>
</tr>
<tr>
<td>Lisa Singh</td>
<td>Georgetown University, USA</td>
</tr>
<tr>
<td>Dongjin Song</td>
<td>NEC Labs America, USA</td>
</tr>
<tr>
<td>Le Song</td>
<td>Georgia Tech, USA</td>
</tr>
<tr>
<td>Yangqiu Song</td>
<td>UIUC, USA</td>
</tr>
<tr>
<td>Arnaud Soulet</td>
<td>University François Rabelais des Tours, France</td>
</tr>
<tr>
<td>Sucheta Soundarajan</td>
<td>Syracuse University, USA</td>
</tr>
<tr>
<td>Jerzy Stefanowski</td>
<td>Technical University Poznan, Poland</td>
</tr>
<tr>
<td>Andrea Targarelli</td>
<td>University of Calabria, Italy</td>
</tr>
<tr>
<td>Hana Tai</td>
<td>National Taipei University, Taiwan</td>
</tr>
<tr>
<td>Pang-Ning Tan</td>
<td>Michigan State University, USA</td>
</tr>
<tr>
<td>Nan Tang</td>
<td>QCRI, Qatar</td>
</tr>
<tr>
<td>Dacheng Tao</td>
<td>The University of Technology, Sydney, Australia</td>
</tr>
<tr>
<td>Hanghang Tong</td>
<td>Arizona State University, USA</td>
</tr>
<tr>
<td>Caetano Traina</td>
<td>ICMC-USP, Brazil</td>
</tr>
<tr>
<td>Panayiotis Tsaparas</td>
<td>University of Ioannina, Greece</td>
</tr>
<tr>
<td>Grigoris Tsoumakas</td>
<td>Aristotle University Thessaloniki, Greece</td>
</tr>
<tr>
<td>Matthijs van Leeuwen</td>
<td>Leiden University, Netherlands</td>
</tr>
<tr>
<td>Adriano Veloso</td>
<td>Universidade Federal de Minas Gerais, Brazil</td>
</tr>
<tr>
<td>Suresh, Venkatasubramiam</td>
<td>University of Utah, USA</td>
</tr>
<tr>
<td>Jilles Vreeken</td>
<td>Max Planck Institute, Germany</td>
</tr>
<tr>
<td>Fei Wang</td>
<td>University of Connecticut, USA</td>
</tr>
<tr>
<td>Jianyong Wang</td>
<td>Tsinghua University, China</td>
</tr>
</tbody>
</table>
SIAM International Conference on Data Mining

Ping Zhang
IBM Thomas J. Watson Research Center, USA

Xiang Zhang
Penn State University, USA

Xiangliang Zhang
KAUST, Saudi Arabia

Xiaoming Zhang
Beihang University, China

Yao Zhang
Virginia Tech, USA

Zhao Zhang
Soochow University, China

Yu Zheng
Microsoft, China

Jiayu Zhou
Michigan State University, USA

Yuanchun Zhou
Chinese Academy of Sciences, China

Hengshu Zhu
Baidu, China

Xingquan Zhu
Florida Atlantic University, USA

Arthur Zimek
University of Southern Denmark, Denmark

Albrecht Zimmermann
University de Caen, France

Ke Wang
Simon Fraser University, Canada

Ting Wang
Lehigh University, USA

Xiang Wang
Google, USA

Zheng Wang
University of Michigan, USA

Takahashi Washio
Osaka University, Japan

Ingmar Weber
QCRI, Qatar

Raymond Chi-Wing Wong
Hong Kong University of Science and Technology, Hong Kong

Jia Wu
University of Technology, Australia

Xintao Wu
University of Arkansas, USA

Ren Xiang
University of Illinois Urbana-Champaign, USA

Keli Xiao
Stony Brook University, USA

Yanghua Xiao
Fudan University, China

Jian Xu
University of Notre Dame, USA

Tong Xu
University of Science and Technology of China, China

Makoto Yamada
Kyoto University, Japan

Yang Yang
University of California Los Angeles, USA

De-Chuan Zhan
Nanjing University, China

Chao Zhang
Dalian University of Technology, China

Kun Zhang
Xavier University of Louisiana, USA

Lumin Zhang
Facebook, USA

Min-Ling Zhang
Southeast University, China

Hotel Address
Westin Galleria Houston
5060 West Alabama Street
Houston, Texas 77056 USA
Phone Number: +1-713-960-8100
Toll Free Reservations (USA and Canada): +1-888-627-8514
Fax: +1-713-960-6549
Hotel web address: http://westingalleriahoustonhotel.com

Hotel Telephone Number
To reach an attendee or leave a message, call +1-713-960-8100. If the attendee is a hotel guest, the hotel operator can connect you with the attendee’s room.

Hotel Check-in and Check-out Times
Check-in time is 3:00 PM.
Check-out time is 12:00 PM.

Child Care
The Westin Galleria Houston recommends Mom’s Best Friend (http://momsbestfriend.com/houston/) for attendees interested in child care services. Attendees are responsible for making their own child care arrangements. Their phone number is +1-713-776-2669 ext. 204.

Corporate Members and Affiliates
SIAM corporate members provide their employees with knowledge about, access to, and contacts in the applied mathematics and computational sciences community through their membership benefits. Corporate membership is more than just a bundle of tangible products and services; it is an expression of support for SIAM and its programs. SIAM is pleased to acknowledge its corporate members and sponsors. In recognition of their support, non-member attendees who are employed by the following organizations are entitled to the SIAM member registration rate.

SIAM Registration Desk
The SIAM registration desk is located in the Galleria Foyer except Wednesday night, when it will be in the Chevy Chase Room. It is open during the following hours:

- Wednesday, April 26
  5:00 PM – 7:00 PM
- Thursday, April 27
  7:00 AM – 7:30 PM
- Friday, April 28
  7:15 AM – 3:30 PM
- Saturday, April 29
  7:15 AM – 4:00 PM
Corporate/Institutional Members
The Aerospace Corporation
Air Force Office of Scientific Research
Amazon
Aramco Services Company
Bechtel Marine Propulsion Laboratory
The Boeing Company
CEA/DAM
Department of National Defence (DND/CSEC)
DSTO- Defence Science and Technology Organisation
Hewlett-Packard
Huawei FRC French R&D Center
IBM Corporation
IDA Center for Communications Research, La Jolla
IDA Center for Communications Research, Princeton
Institute for Defense Analyses, Center for Computing Sciences
Lawrence Berkeley National Laboratory
Lawrence Livermore National Labs
Lockheed Martin
Los Alamos National Laboratory
Max-Planck-Institute for Dynamics of Complex Technical Systems
Mentor Graphics
National Institute of Standards and Technology (NIST)
National Security Agency (DIRNSA)
Naval PostGrad
Oak Ridge National Laboratory, managed by UT-Battelle for the Department of Energy
Sandia National Laboratories
Schlumberger-Doll Research
United States Department of Energy
U.S. Army Corps of Engineers, Engineer Research and Development Center
US Naval Research Labs

Funding Agency
SIAM and the conference organizing committee wish to extend their thanks and appreciation to U.S. National Science Foundation for its support of this conference.

Join SIAM and save!
Leading the applied mathematics community . . .
SIAM members save up to $130 on full registration for the 2017 SIAM Conference on SIAM International Conference on Data Mining! Join your peers in supporting the premier professional society for applied mathematicians and computational scientists. SIAM members receive subscriptions to SIAM Review, SIAM News and SIAM Unwrapped, and enjoy substantial discounts on SIAM books, journal subscriptions, and conference registrations.

If you are not a SIAM member and paid the Non-Member rate to attend the conference, you can apply the difference between what you paid and what a member would have paid ($130 for a Non-Member) towards a SIAM membership. Contact SIAM Customer Service for details or join at the conference registration desk.

If you are a SIAM member, it only costs $15 to join the SIAM Activity Group on Data Mining and Analytics (SIAG/DMA). As a SIAG/DMA member, you are eligible for an additional $15 discount on this conference, so if you paid the SIAM member rate to attend the conference, you might be eligible for a free SIAG/DMA membership. Check at the registration desk.

Free Student Memberships are available to students who attend an institution that is an Academic Member of SIAM, are members of Student Chapters of SIAM, or are nominated by a Regular Member of SIAM.

Join onsite at the registration desk, go to www.siam.org/joinsiam to join online or download an application form, or contact SIAM Customer Service:
Telephone: +1-215-382-9800 (worldwide); or 800-447-7426 (USA and Canada only)
Fax: +1-215-386-7999
E-mail: membership@siam.org
Postal mail: Society for Industrial and Applied Mathematics, 3600 Market Street, 6th floor, Philadelphia, PA 19104-2688 USA

Standard Audio/Visual Set-Up in Meeting Rooms
SIAM does not provide computers for any speaker. When giving an electronic presentation, speakers must provide their own computers. SIAM is not responsible for the safety and security of speakers’ computers.

The Plenary Session Room will have two (2) screens, one (1) data projector and one (1) overhead projector. The data projectors support VGA connections only. Presenters requiring an HDMI or alternate connection must provide their own adaptor.

All other concurrent/breakout rooms will have one (1) screen and one (1) data projector. The data projectors support VGA connections only. Presenters requiring an HDMI or alternate connection must provide their own adaptor.

If you have questions regarding availability of equipment in the meeting room of your presentation, please see a SIAM staff member at the registration desk.

List current March 2017.
**Internet Access**
Complimentary wireless Internet access will be available for SIAM attendees in the meeting space and guestrooms in the SIAM block.

In addition, a limited number of computers with Internet access will be available during registration hours.

**Registration Fee Includes**
- Admission to all technical sessions
- Admission to all tutorial sessions
- Admission to workshops
- Business Meeting (open to SIAG/DMA members)
- Coffee breaks daily
- Continental breakfast daily
- Room set-ups and audio/visual equipment
- USB of conference proceedings, workshop and tutorial notes
- Welcome Reception and Poster Session
- Doctoral Forum and Poster Session

**Job Postings**
Please check with the SIAM registration desk regarding the availability of job postings or visit [http://jobs.siam.org](http://jobs.siam.org).

**Important Notice to Poster Presenters**
The poster sessions are scheduled for Thursday, April 27, 7:00 PM – 9:00 PM and Friday, April 28, 7:00 PM -9:00 PM. Presenters are requested to put up their posters no later than 7:00 PM, the official start time of both sessions.

Papers presented on Thursday and Saturday will have their poster slots during the Welcome Reception and Poster Session on Thursday, April 27. Papers presented on Friday will have their poster slots during the Doctoral Forum and Poster Session on Friday, April 28. Boards and push pins will be available to Thursday’s presenters at 12:00 PM on Thursday, April 27 and at 9:00 AM on Friday, April 28 for Friday’s presenters.

For information about preparing a poster, please visit [http://www.siam.org/meetings/guidelines/presenters.php](http://www.siam.org/meetings/guidelines/presenters.php).

**SIAM Books and Journals**
Display copies of books and complimentary copies of journals are available on site. SIAM books are available at a discounted price during the conference. Titles on display forms are available with instructions on how to place a book order.

**Table Top Displays**
SIAM
Springer

**Name Badges**
A space for emergency contact information is provided on the back of your name badge. Help us help you in the event of an emergency!

**Comments?**
Comments about SIAM meetings are encouraged! Please send to: Cynthia Phillips, SIAM Vice President for Programs (vpp@siam.org).

**Get-togethers**
Welcome Reception and Poster Session
Thursday, April 27
7:00 PM – 9:00 PM

Business Meeting (open to SIAG/DMA members)
Friday, April 28
6:15 PM – 7:00 PM

Complimentary beer and wine will be served.

**Statement on Inclusiveness**
As a professional society, SIAM is committed to providing an inclusive climate that encourages the open expression and exchange of ideas, that is free from all forms of discrimination, harassment, and retaliation, and that is welcoming and comfortable to all members and to those who participate in its activities. In pursuit of that commitment, SIAM is dedicated to the philosophy of equality of opportunity and treatment for all participants regardless of gender, gender identity or expression, sexual orientation, race, color, national or ethnic origin, religion or religious belief, age, marital status, disabilities, veteran status, field of expertise, or any other reason not related to scientific merit. This philosophy extends from SIAM conferences, to its publications, and to its governing structures and bodies. We expect all members of SIAM and participants in SIAM activities to work towards this commitment.

**Please Note**
SIAM is not responsible for the safety and security of attendees’ computers. Do not leave your personal electronic devices unattended. Please remember to turn off your cell phones and other devices during sessions.
Recording of Presentations

Audio and video recording of presentations at SIAM meetings is prohibited without the written permission of the presenter and SIAM.

Social Media

SIAM is promoting the use of social media, such as Facebook and Twitter, in order to enhance scientific discussion at its meetings and enable attendees to connect with each other prior to, during and after conferences. If you are tweeting about a conference, please use the designated hashtag to enable other attendees to keep up with the Twitter conversation and to allow better archiving of our conference discussions. The hashtag for this meeting is #SIAMSDM17.

SIAM’s Twitter handle is @TheSIAMNews.
SIAM Activity Group on
Data Mining and Analytics (SIAG/DMA)
www.siam.org/activity/dma/index.php

A GREAT WAY
TO GET INVOLVED!
Collaborate and interact with mathematicians and applied scientists whose work involves data mining.

ACTIVITIES INCLUDE:
• Special sessions at SIAM Annual Meetings
• Annual conference

BENEFITS OF SIAG/DMA MEMBERSHIP:
• Listing in the SIAG’s online membership directory
• Additional $15 discount on registration at SIAM International Conference on Data Mining (excludes student)
• Electronic communications from your peers about recent developments in your specialty
• Eligibility for candidacy for SIAG/DMA office
• Participation in the selection of SIAG/DMA officers

ELIGIBILITY:
• Be a current SIAM member.

COST:
• $15 per year
• Student members can join two activity groups for free!

2016-17 SIAG/DMA OFFICERS
• Chair: Jeremy Kepner, Massachusetts Institute of Technology
• Vice Chair: Jaideep Vaidya, Rutgers University
• Program Director: Wagner Meira, Universidade Federal de Minas Gerais
• Secretary: Danai Koutra, University of Michigan

TO JOIN:
SIAG/DMA: my.siam.org/forms/join_siag.htm
SIAM: www.siam.org/joinsiam
** Invited Plenary Speakers **

** All Invited Plenary Presentations will take place in Galleria Ballroom I **

**Thursday, April 27**

8:15 AM - 9:30 AM

**IP1** Title Not Available at Time of Publication

*William Craig Fugate, Federal Emergency Management Agency, USA*

1:30 PM - 2:45 PM

**IP2** Collective Learning in Society and the Economy

*Cesar Hidalgo, Massachusetts Institute of Technology, USA*

**Friday, April 28**

8:15 AM - 9:30 AM

**IP3** Machine Learning under Resource Constraints

*Katharina Morik, Technische Universität Dortmund, Germany*

1:30 PM - 2:45 PM

**IP4** Repeated Choice, Markov Models, and LAMP

*Andrew Tomkins, Google, Inc., USA*
Tutorials

Thursday, April 27
10:00 AM - 12:00 PM

TS1: Tutorial Session: An Introduction to Redescription Mining
Post Oak

2:45 PM - 4:45 PM

TS2: Tutorial Session: Leveraging Propagations for Data Mining
Post Oak

5:00 PM - 7:00 PM

TS3: Tutorial Session: Opportunities, Challenges and Methods for Higher Education Data Mining
Post Oak

Friday, April 28
10:00 AM - 12:00 PM

TS4: Tutorial Session: IoT Big Data Stream Mining
Post Oak

2:45 PM - 3:45 PM

TS4: Tutorial Session (Continued): IoT Big Data Stream Mining
Post Oak

Saturday, April 29
3:45 PM - 5:45 PM

TS5: Tutorial Session: Summarizing Large-Scale Graph Data
Galleria Ballroom I
Workshops

Saturday, April 29

8:00 AM - 12:15 PM (Half Day Workshop)

**Workshop 2: Data Mining for Medicine and Healthcare**
*Tanglewood*

8:00 AM - 5:45 PM (Full Day Workshops)

**Workshop 1: Machine Learning Methods for Recommender Systems**
*Galleria Ballroom II*

**Workshop 4: Inferring Networks From Non-Network Data**
*Post Oak*

**Workshop 5: Mining Big Data in Climate and Environment**
*Bellaire*

**Workshop 6: Women in Data Science**
*San Felipe*

1:30 PM - 5:45 PM (Half Day Workshop)

**Workshop 3: Workshop on Data Mining for Oil & Gas**
*Tanglewood*
SIAM Presents is an audio-visual archive comprised of more than 2,000 presentations posted in over 40 searchable topics, including:

- algebraic geometry
- atmospheric and oceanographic science
- computational science
- data mining
- geophysical science
- optimization
- uncertainty quantification and more...

The collection, Featured Lectures from our Archives, includes audio and slides from more than 30 conferences since 2008, including talks by invited and prize speakers, select minisymposia, and minitutorials. Presentations from SIAM meetings are being added throughout the year.

In addition you can view short video clips of speaker interviews from sessions at Annual Meetings starting in 2010.

Plans for adding more content are on the horizon. Keep an eye out!

The audio, slide, and video presentations are part of SIAM’s outreach activities to increase the public’s awareness of mathematics and computational science in the real world, and to bring attention to exciting and valuable work being done in the field. Funding from SIAM, the National Science Foundation, and the Department of Energy was used to partially support this project.

New presentations are posted every few months as the program expands with sessions from additional SIAM meetings. Users can search for presentations by category, speaker name, and/or key words.

www.siam.org/meetings/presents.php
2017 SIAM
International Conference on **DATA MINING**

April 27 - 29, 2017

The Westin Galleria Houston
Houston, Texas, USA
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, April 26</td>
<td></td>
<td>Registration</td>
<td>Chevy Chase</td>
</tr>
<tr>
<td></td>
<td>5:00 PM-7:00 PM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Continental Breakfast</td>
<td>Galleria Foyer</td>
</tr>
<tr>
<td></td>
<td>7:30 AM-8:00 AM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Welcome Remarks</td>
<td>Galleria Ballroom I</td>
</tr>
<tr>
<td></td>
<td>8:00 AM-8:15 AM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday, April 27</td>
<td></td>
<td>Registration</td>
<td>Galleria Foyer</td>
</tr>
<tr>
<td></td>
<td>7:00 AM-7:30 PM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IP 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8:15 AM-9:30 AM</td>
<td>Title Not Available at Time of Publication</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chair: Wei Wang, University of California, Los Angeles, USA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abstract not available at time of publication</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>William Craig Fugate, Federal Emergency Management Agency, USA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coffee Break</td>
<td>Galleria Foyer</td>
</tr>
<tr>
<td></td>
<td>9:30 AM-10:00 AM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this tutorial we will give an overview of redescription mining, from the intuition behind the concept and its links to existing data analysis techniques to more recent developments in algorithms and interactive mining techniques. We will also cover five areas where applications for redescription mining have been proposed. The tutorial will give the attendants knowledge of the state-of-the-art techniques in redescription mining and open problems in method development, as well as examples and information on how to apply redescription mining to real-world data analysis problems.

Pauli Miettinen
Max Planck Institute for Informatics, Germany

Esther Galbrun
Inria, France

Yashu Liu, Arizona State University, USA; Shuang Qiu, University of Michigan, Ann Arbor, USA; Ping Zhang, IBM T.J. Watson Research Center, USA; Pinghua Gong, University of Michigan, Ann Arbor, USA; Fei Wang, IBM T.J. Watson Research Center, USA; Guoliang Xue, Arizona State University, USA; Jieping Ye, University of Michigan, Ann Arbor, USA

Ioakeim Perros, Georgia Institute of Technology, USA; Fei Wang, Weill Cornell Medicine, USA; Ping Zhang, IBM T.J. Watson Research Center, USA; Peter Walker, Naval Medical Research Center, USA; Richard Vuduc, Georgia Institute of Technology, USA; Jyotishman Pathak, Weill Cornell Medicine, USA; Jimeng Sun, Georgia Institute of Technology, USA

Honglei Liu, University of California, Santa Barbara, USA; Bian Wu, Washington State University, USA
Thursday, April 27

**CP3**

**Anomaly and Outlier Detection**

10:00 AM-11:40 AM

Room: Tanglewood

Chair: Xia Ben Hu, Texas A&M University, USA

10:00-10:15 Outlier Detection with Autoencoder Ensembles

Jinghui Chen, University of Virginia, USA; Saket Sathe, Charu C. Aggarwal, and Deepak Turaga, IBM T.J. Watson Research Center, USA

10:20-10:35 Gleaning Wisdom from the Past: Early Detection of Emerging Rumors in Social Media

Liang Wu and Jundong Li, Arizona State University, USA; Xia Hu, Texas A&M University, USA; Huan Liu, Arizona State University, USA

10:40-10:55 VolTime: Unsupervised Anomaly Detection on Users’ Online Activity Volume

Daniel Chino, Alceu Costa, and Agma J. Traina, University of Sao Paulo, Brazil; Christos Faloutsos, Carnegie Mellon University, USA

11:00-11:15 Detecting Malicious Behavior in Computer Networks Via Cost-Sensitive and Connectivity Constrained Classification

Houping Xiao and Jing Gao, State University of New York at Buffalo, USA; Long Vu and Deepak Turaga, IBM T.J. Watson Research Center, USA

11:20-11:35 Efficiently Discovering Unexpected Pattern-Co-Occurrences

Arno Siebes and Roel Bertens, Universiteit Utrecht, The Netherlands; Jilles Vreeken, Max Planck Institute for Informatics, Germany

**Coffee Break**

2:30 PM-2:45 PM

Room: Galleria Foyer

**Lunch Break**

12:00 PM-1:15 PM

Attendees on their own

---

Thursday, April 27

**IP2**

**Collective Learning in Society and the Economy**

1:15 PM-2:30 PM

Room: Galleria Ballroom 1

Chair: Nitesh Chawla, University of Notre Dame, USA

How do economies learn to develop new products, industries, skills, and occupations? Where do they get the knowledge they need? And how is that learning affected by technologies, institutions, and culture? In this presentation I will present work describing the key questions defining the field of collective learning, and also, will present novel data visualization tools designed to facilitate collective learning in teams and nations.

Cesar Hidalgo

Massachusetts Institute of Technology, USA

**TS2**

**Tutorial Session: Leveraging Propagations for Data Mining**

2:45 PM-4:45 PM

Room: Post Oak

Chair: Jilles Vreeken, Saarland University and Max Planck Institute for Informatics, Germany

Can we guess if a user is sick from her tweet? How do opinions get formed in online forums? Which people should we immunize, to prevent an epidemic as fast as possible? How to quickly zoom out of a graph? Graphs---also known as networks---are powerful tools for modeling processes and situations of interest in real-life like social-systems, cyber-security, epidemiology and biology. They are ubiquitous, from online social networks, gene-regulatory networks, to router graphs.

This tutorial will cover recent and state-of-the-art research on how propagation-like processes can help big-data mining specifically large networks and time-series, algorithms behind network problems, and their practical applications in various diverse settings. Topics include diffusion and virus propagation in networks, anomaly and outbreak detection, event prediction and connections with work in public health, the web and online media, social sciences/humanities and cyber security.

B. Aditya Prakash

Virginia Tech, USA

Naren Ramakrishnan

Virginia Tech, USA
Thursday, April 27

**CP4 Classification II**

2:45 PM-4:25 PM

*Room: Galleria Ballroom I*

**Chair:** Polo Chau, Georgia Institute of Technology, USA

2:45-3:00 Time-Aware Subscription Prediction Model for User Acquisition in Digital News Media

*Heidar Davoudi, Morteza Zibayat, and Aijun An, York University, Canada*

3:05-3:20 The Power of Certainty: A Dirichlet-Multinomial Model for Belief Propagation

*Divy Dwarakanath, Carnegie Mellon University, USA; Stephan Guennemann, Technical University of Munich, Germany; Christos Faloutsos, Carnegie Mellon University, USA*

3:25-3:40 From Theory to Practice: Efficient Active Cost-Sensitive Classification with Expected Error Reduction

*Yexun Zhang and Yanfeng Wang, Shanghai Jiao Tong University, China; Wenbin Cai, Microsoft Research Asia; Siyuan Zhou and Ya Zhang, Shanghai Jiao Tong University, China*

3:45-4:00 Generalized Inverse Classification

*Micah T. Lash, Qihang Lin, Nick Street, Jennifer Robinson, and Jeffrey Ohlmann, University of Iowa, USA*

4:05-4:20 Predict Land Covers with Transition Modeling and Incremental Learning

*Xiaowei Jia, University of Minnesota, Twin Cities, USA; Ankush Khandelw, Gururuprasad Nayak, and James Gerber, University of Minnesota, USA; Kimberly Carlson, University of Hawaii, Manoa, USA; Paul West and Vipin Kumar, University of Minnesota, USA*

---

Thursday, April 27

**CP5 Health Informatics**

2:45 PM-4:25 PM

*Room: Galleria Ballroom II*

**Chair:** Danai Koutra, University of Michigan, USA

2:45-3:00 Unified and Contrasting Graphical Lasso for Brain Network Discovery

*Xinyue Liu and Xiangnan Kong, Worcester Polytechnic Institute, USA; Ann Ragin, Northwestern University, USA*

3:05-3:20 T-Bne: Tensor-Based Brain Network Embedding

*Bokai Cao, University of Illinois, Chicago, USA; Lifang He, South China University of Technology, China; Xiaokai Wei, University of Illinois at Chicago, USA; Mengqi Xing, Philip Yu, Heide Klumpp, and Alex Leow, University of Illinois, Chicago, USA*

3:25-3:40 An Rnn Architecture with Dynamic Temporal Matching for Personalized Predictions of Parkinson’s Disease

*Chao Che, Dalian University of Technology, China; Cao Xiao, University of Washington and IBM T.J. Watson Research Center, USA; Jian Liang, Tsinghua University, P. R. China; Bo Jin, Dalian University of Technology, China; Jiayu Zhou, Michigan State University, USA; Fei Wang, IBM T.J. Watson Research Center, USA*

3:45-4:00 Clustering with Domain-Specific Usefulness Scores

*Yale Chang and Junxiang Chen, Northeastern University, USA; Michael Cho, Peter Castaldi, and Edwin Silverman, Harvard Medical School, USA; Jennifer Dy, Northwestern University, USA*

4:05-4:20 Brainzoom: High Resolution Reconstruction from Multi-Modal Brain Signals

*Xiao Fu, University of Minnesota, Twin Cities, USA; Kejun Huang, University of Minnesota, USA; Otilia Stretcu and Hyun Ah Song, Carnegie Mellon University, USA; Evangelos Papalexakis, University of California, Riverside, USA; Partha Talukdar, Indian Institute of Science, Bangalore, India; Tom Mitchell, Carnegie Mellon University, USA; Nicholas Sidiropoulos, University of Minnesota, USA; Christos Faloutsos, Carnegie Mellon University, USA*

---

Thursday, April 27

**CP6 Clustering and Recommender Systems**

2:45 PM-4:25 PM

*Room: Tanglewood*

**Chair:** Leman Akoglu, Carnegie Mellon University, USA

2:45-3:00 Uncovering Group Level Insights with Accordant Clustering

*Amit Dhurandhar, IBM Research, USA; Margareta Ackerman, San Jose State University, USA; Xiang Wang, Google, Inc., USA*

3:05-3:20 A Method to Accelerate Human in the Loop Clustering

*Anni Coden, Marina Danilevsky, Daniel Gruhl, Linda Kato, and Meenakshi Nagarajan, IBM Research, USA*

3:25-3:40 Model-Based Von Mises-Fisher Co-Clustering with a Conscience

*Aghiles Salah and Mohamed Nadif, Paris Descartes, France*

3:45-4:00 Targeted Matrix Completion

*Natali Ruchansky, University of Southern California, USA; Evimaria Terzi and Mark Crovella, Boston University, USA*

4:05-4:20 Active Positive-Definite Matrix Completion

*Charalampos Mavroforakis, Dora Erdos, Mark Crovella, and Evimari Terzi, Boston University, USA*

---

**Organizational Break**

4:45 PM-5:00 PM
Tutorial Session: Opportunities, Challenges and Methods for Higher Education Data Mining
5:00 PM-7:00 PM
Room: Post Oak
Chair: Jilles Vreeken, Saarland University and Max Planck Institute for Informatics, Germany

The application of big data approaches, specifically methods inspired from recommender system domain to predict student performance is largely a new area of research. The types of solutions available depend largely on the type of available data, and problem definition. For instance, for the purposes of degree planning, one task is to predict grades for a student in a class in the future (or in the next term). To predict student performance within individual class assessments, either in a typical classroom, an online, or a MOOC setting, students interaction with a learning management system (LMS) provides different ways to address this problem.

In this tutorial we will provide a formal definition of higher educational mining and discuss different opportunities for inter-disciplinary research in this field. Challenges as they relate to data acquisition and ethics will be discussed in detail followed with current practices surrounding the development of degree planning tools and course analytics to assist students and teachers.

Huzefa Rangwala
George Mason University, USA

Aditya Johri
George Mason University, USA

Asmaa El Badrawy
University of Minnesota, USA

George Karypis
University of Minnesota, USA
Thursday, April 27

CP9

Recommender Systems
5:00 PM-6:40 PM

Room:Tanglewood

Chair: Xia Ning, Purdue University, USA

5:00-5:15 Selection of Negative Samples for One-Class Matrix Factorization
Hsiang-Fu Yu, Amazon.com, USA; Mikhail Bilenko, Microsoft Corporation, USA; Chih-Jen Lin, National Taiwan University, Taiwan

5:20-5:35 HBGG: a Hierarchical Bayesian Geographical Model for Group Recommendation
Ziyu Lu, Hui Li, Nikos Mamoulis, and David Cheung, University of Hong Kong, Hong Kong

5:40-5:55 Collaborative User Network Embedding for Social Recommender Systems
Chuxu Zhang, Rutgers University, USA; Lu Yu, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Yan Wang and Chirag Shah, Rutgers University, USA; Xiangliang Zhang, King Abdullah University of Science & Technology (KAUST), Saudi Arabia

6:00-6:15 Redundancies in Data and Their Effect on the Evaluation of Recommendation Systems: A Case Study on the Amazon Reviews Datasets
Daniel Basaran, Ludwig-Maximilians-Universität München, Germany; Eirini C. Ntoutsi, Leibniz University Hannover, Germany; Arthur Zimek, University of Southern Denmark, Denmark

6:20-6:35 Price Recommendation on Vacation Rental Websites
Yang Li, Northwestern Polytechnical University, China; Suhang Wang, Arizona State University, USA; Tao Yang and Quan Pan, Northwestern Polytechnical University, China; Jiliang Tang, Michigan State University, USA

Welcome Reception and Poster Session
7:00 PM-9:00 PM

Room:Galleria Ballroom III and Galleria Foyer

Papers presented on Thursday and Saturday will have their poster slots during this session.

Friday, April 28

IP3

Machine Learning under Resource Constraints
8:15 AM-9:30 AM

Room:Galleria Ballroom I

Chair: Dimitrios Gunopulos, University of Athens, Greece

Big data are produced by various sources. Most often, they are distributedly stored at computing farms or clouds. Analytics on the Hadoop Distributed File System (HDFS) then follows the MapReduce programming model. According to the Lambda architecture of Nathan Marz and James Warren, this is the batch layer. It is complemented by the speed layer, which aggregates and integrates incoming data streams in real time. When considering big data and small devices, obviously, we imagine the small devices being hosts of the speed layer, only. Analytics on the small devices is restricted by memory and computation resources. The interplay of streaming and batch analytics offers a multitude of configurations. In this talk, we discuss opportunities for using sophisticated models for learning spatio-temporal models. In particular, we investigate graphical models, which generate the probabilities for connected (sensor) nodes. We even approximate likelihood estimates such that they can be computed on very restricted devices.

Katharina Morik
Technische Universität Dortmund, Germany

Coffee Break
9:30 AM-10:00 AM

Room:Galleria Foyer
The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. Advanced analysis of big data streams from sensors and devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. Dealing with the evolution over time of such data streams, i.e., with concepts that drift or change completely, is one of the core issues in IoT stream mining. This tutorial is a gentle introduction to mining IoT big data streams. The first part introduces data stream learners for classification, regression, clustering, and frequent pattern mining. The second part deals with scalability issues inherent in IoT applications, and discusses how to mine data streams on distributed engines such as Spark, Flink, Storm, and Samza.

Gianmarco De Francisci Morales  
Qatar Computing Research Institute, Qatar

Albert Bifet  
Télécom ParisTech, France

Abdul R. Khan  
King Fahd University of Petroleum and Minerals, Saudi Arabia

Joao Gama  
University of Porto, Portugal

Wei Fan  
Baidu Research Big Data Lab, China
Friday, April 28

**CP12**

**Feature, Extraction, Selection, and Dimension Reduction**

10:00 AM-11:40 AM

Room: Tanglewood

Chair: Jiayu Zhou, Michigan State University, USA

10:00-10:15 *Roflmao: Robust Oblique Forests with Linear Matrix Operations*

Tyler M. Tomita, Johns Hopkins University, USA

10:20-10:35 *Exploiting Hierarchical Structures for Unsupervised Feature Selection*

Suhang Wang and Yilin Wang, Arizona State University, USA; Jiliang Tang, Michigan State University, USA; Charu C. Aggarwal, IBM T.J. Watson Research Center, USA; Suhas Ranganath and Huan Liu, Arizona State University, USA

10:40-10:55 *Embedded Supervised Feature Selection for Multi-Class Data*

Lin Chen, Arizona State University, USA; Jiliang Tang, Michigan State University, USA; Baoxin Li, Arizona State University, USA

11:00-11:15 *Correlation by Compression*

Kailash Budhathoki and Jilles Vreeken, Max Planck Institute for Informatics, Germany

11:20-11:35 *A Deflation Method for Structured Probabilistic PCA*

Rajiv Khanna and Joydeep Ghosh, University of Texas at Austin, USA; Russell Poldrack, Stanford University, USA; Oluwasanmi Koyejo, University of Illinois, Urbana-Champaign, USA

**Lunch Break**

12:00 PM-1:15 PM

Attendees on their own

Friday, April 28

**IP4**

**Repeated Choice, Markov Models, and LAMP**

1:15 PM-2:30 PM

Room: Galleria Ballroom I

Chair: Carlotta Domeniconi, George Mason University, USA

In this talk I’ll discuss modeling of user consumption data, using the framework of Discrete Choice. I’ll start with some background on choice models, survey the key ideas, then give an example modeling repeated consumption of the same item (for instance a song, restaurant, or web page). I’ll then show how choice theory can be integrated with first-order markov models to develop lightweight but powerful “LAMP” sequence models that efficiently learn how to take history into account. Finally, I’ll show some comparisons between these models and standard deep network sequence models.

Andrew Tomkins

Google, Inc., USA

**Coffee Break**

2:30 PM-2:45 PM

Room: Galleria Foyer

Friday, April 28

**TS4**

**Tutorial Session: IoT Big Data Stream Mining, Continued**

2:45 PM-3:45 PM

Room: Post Oak

Chair: Jilles Vreeken, Saarland University and Max Planck Institute for Informatics, Germany

The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. Advanced analysis of big data streams from sensors and devices is bound to become a key area of data mining research as the number of applications requiring such processing increases. Dealing with the evolution over time of such data streams, i.e., with concepts that drift or change completely, is one of the core issues in IoT stream mining. This tutorial is a gentle introduction to mining IoT big data streams. The first part introduces data stream learners for classification, regression, clustering, and frequent pattern mining. The second part deals with scalability issues inherent in IoT applications, and discusses how to mine data streams on distributed engines such as Spark, Flink, Storm, and Samza.

Gianmarco De Francisci Morales

Qatar Computing Research Institute, Qatar

Albert Bifet

Télécom ParisTech, France

Abdul R. Khan

King Fahd University of Petroleum and Minerals, Saudi Arabia

Joao Gama

University of Porto, Portugal

Wei Fan

Baidu Research Big Data Lab, China
Friday, April 28

CP13

Novel Applications
2:45 PM-4:45 PM
Room: Galleria Ballroom I
Chair: Xuan-Hong Dang, University of California, Santa Barbara, USA

2:45-3:00 Identifying Deep Contrasting Networks from Time Series Data: Application to Brain Network Analysis
John Boaz Lee, Xiangnan Kong, and Yihan Bao, Worcester Polytechnic Institute, USA; Constance Moore, University of Massachusetts Medical School, USA

3:05-3:20 Cumulative Knowledge-based Regression Models for Next-term Grade Prediction
Sara Morsy, University of Minnesota, USA; George Karypis, University of Minnesota, USA

Miguel Araujo, Carnegie Mellon University, USA; Miguel Almeida, Jaime Ferreira, Luis Silva, and Pedro Bizarro, Feedzai, Portugal

4:05-4:20 MultiC2: An Optimization Framework for Learning from Task and Worker Dual Heterogeneity
Yao Zhou, Lei Ying, and Jingrui He, Arizona State University, USA

Friday, April 28

CP14

Mining Graphs II
2:45 PM-4:45 PM
Room: Galleria Ballroom II
Chair: Manuel Gomez Rodriguez, Max Planck Institute for Software Systems, Germany

2:45-3:00 FACETS: Adaptive Local Exploration of Large Graphs
Robert Pienta and Minsuk Kahng, Georgia Institute of Technology, USA; Zhiyuan Lin, Stanford University, USA; Jilles Vreeken, Max Planck Institute for Informatics, Germany; Partha Talukdar, Indian Institute of Science, Bangalore, India; James Abello, Rutgers University, USA; Ganesh Parameswaran, Yahoo! Inc., USA; Duen Horng Chau, Georgia Institute of Technology, USA

3:05-3:20 Absenteeism Detection in Social Media
Fang Jin, Virginia Tech, USA

3:25-3:40 Multimodal Network Alignment
Huda Nassar and David F. Gleich, Purdue University, USA

3:45-4:00 Near-Optimal and Practical Algorithms for Graph Scan Statistics
Jose Cadena and Anil Vullikanti, Virginia Tech, USA; Feng Chen, University of Albany - State University of New York, USA

4:05-4:20 Accelerated Attributed Network Embedding
Xiao Huang, Texas A&M University, USA; Jundong Li, Arizona State University, USA; Xia Hu, Texas A&M University, USA

Friday, April 28

CP15

Novel Learning Methods
2:45 PM-4:45 PM
Room: Tanglewood
Chair: Huan Liu, Arizona State University, USA

2:45-3:00 Statistical Learning Theory Approach for Data Classification with l-Diversity
Koray Mancuhan and Chris Clifton, Purdue University, USA

3:05-3:20 Private and Right-Protected Big Data Publication: An Analysis
Michail Vlachos, IBM Research, USA; Reinhard Heckel, University of California, Berkeley, USA

3:25-3:40 Differentially Private Rank Aggregation
Michael Hay, Colgate University, USA; Liudmila Elagina and Gerome Miklau, University of Massachusetts, Amherst, USA

3:45-4:00 Hash-Based Feature Learning for Incomplete Continuous-Valued Data
Shuai Yuan, Pang-Ning Tan, Kendra Cheruvellil, Emi Last Name, Nicholas Skaff, and Patricia Soranno, Michigan State University, USA

4:05-4:20 Multi-Task Multiple Kernel Relationship Learning
Keerthiram Murugesan and Jaime Carbonell, Carnegie Mellon University, USA

4:25-4:40 Multivariate Confidence Intervals
Jussi Korpela, Emilia Oikarinen, Kai Puolamäki, and Antti Ukkonen, Finnish Institute of Occupational Health, Finland

Organizational Break
4:45 PM-5:00 PM
**Saturday, April 29**

**CP16**

**Optimization Methods**

8:00 AM-9:40 AM

Room: Galleria Ballroom I

Chair: Xiangnan Kong, Worcester Polytechnic Institute, USA

8:00-8:15 A Fast Trust-Region Newton Method for Softmax Logistic Regression
Nayyar Zaidi and Geoff Webb, Monash University, Australia

8:20-8:35 Learning from Multi-Modality Multi-Resolution Data: An Optimization Approach
Yada Zhu, IBM T.J. Watson Research Center, USA; Jianbo Li, Three Bridges Capital, USA; Jingrui He, Arizona State University, USA

8:40-8:55 Sparse Graphical Modeling Via Stochastic Complexity
Kohei Miyaguchi, Shin Matsushima, and Kenji Yamanishi, University of Tokyo, Japan

9:00-9:15 Limited-Memory Common-Directions Method for Distributed Optimization and Its Application on Empirical Risk Minimization
Ching-Pei Lee, University of Wisconsin, Madison, USA; Po-Wei Wang, Carnegie Mellon University, USA; Weizhu Chen, Microsoft, USA; Chih-Jen Lin, National Taiwan University, Taiwan

9:20-9:35 Automatic Frankensteining: Creating Complex Ensembles Autonomously
Martin Wistuba, Nicolas Schilling, and Lars Schmidt-Thieme, University of Hildesheim, Germany

**Registration**

7:15 AM-4:00 PM

Room: Galleria Foyer

**Continental Breakfast**

7:30 AM-8:00 AM

Room: Galleria Foyer
Saturday, April 29

**Workshop 2 (half day): Data Mining for Medicine and Healthcare**
8:00 AM-12:15 PM
Room:Tanglewood

Workshop 1 (full day): Machine Learning Methods for Recommender Systems
8:00 AM-5:45 PM
Room:Galleria Ballroom II

Workshop 4 (full day): Inferring Networks From Non-Network Data
8:00 AM-5:45 PM
Room:Post Oak

Workshop 5 (full day): Mining Big Data in Climate and Environment
8:00 AM-5:45 PM
Room:Bellaire

Workshop 6 (full day): Women in Data Science
8:00 AM-5:45 PM
Room:San Felipe

Coffee Break
10:00 AM-10:15 AM
Room:Galleria Foyer

---

Saturday, April 29

**CP17**

**Data Stream Mining**
10:15 AM-11:55 AM
Room:Galleria Ballroom I

Chair: Kun Zhang, Xavier University, USA

10:15-10:30 Error Metrics for Learning Reliable Manifolds from Streaming Data
Frank Schoeneman, Suchismit Mahapatra, Varun Chandola, Nils Napp, and Jaroslaw Zola, University of Buffalo, USA

10:35-10:50 CSTG: An Effective Framework for Cost-Sensitive Sparse Online Learning
Zhong Chen, Xavier University, USA; Zhide Fang, LSU Health Sciences Center, USA; Wei Fan, Baidu Research Big Data Lab, China; Andrea Edwards and Kun Zhang, Xavier University, USA

10:55-11:10 Concept Drift Detection with Hierarchical Hypothesis Testing
Shujian Yu, University of Florida, USA; Zubin Abraham, Robert Bosch LLC, USA

Rose Yu, Yaguang Li, Cyrus Shahabi, Ugur Demiryurek, and Yan Liu, University of Southern California, USA

11:35-11:50 H-Fuse: Efficient Fusion of Aggregated Historical Data
Zongge Liu and Hyun Ah Song, Carnegie Mellon University, USA; Vladimir Zadorozhny, University of Pittsburgh, USA; Christos Faloutsos, Carnegie Mellon University, USA; Nicholas Sidiropoulos, University of Minnesota, USA

Lunch Break
12:15 PM-1:30 PM
Attendees on their own

Saturday, April 29

**Workshop 3 (half-day): Workshop on Data Mining for Oil & Gas**
1:30 PM-5:45 PM
Room:Tanglewood
CP18
Spatial-Temporal Mining
1:30 PM-3:10 PM
Room: Galleria Ballroom I
Chair: Jilles Vreeken, Max Planck Institute for Informatics, Germany

1:30-1:45 Efficiently Summarising Event Sequences with Rich Interleaving Patterns
Apratim Bhattacharyya, Max Planck Institute for Informatics, Germany and Saarland University, Germany; Jilles Vreeken, Max Planck Institute for Informatics, Germany

1:50-2:05 Discovery of Causal Time Intervals
Zhenhui Li, Guanjie Zheng, Amal Agarwal, Lingzhou Xue, and Thomas Lauvaux, Pennsylvania State University, USA

2:10-2:25 Robust Map Matching for Heterogeneous Data Via Dominance Decompositions
Martin P. Seybold, University of Stuttgart, Germany

2:30-2:45 Uncovering the Spatiotemporal Patterns of Collective Social Activity
Martin Jankowiak, Courant Institute of Mathematical Sciences, New York University, USA; Manuel Gomez-Rodriguez, Max Planck Institute for Software Systems, Germany

2:50-3:05 Discovering Bursts Revisited: Guaranteed Optimization of the Model Parameters
Nikolaj Tatti, Helsinki Institute for Information Technology HIIT, Finland

Coffee Break
3:30 PM-3:45 PM
Room: Galleria Foyer

TS5
Tutorial Session: Summarizing Large-Scale Graph Data
3:45 PM-5:45 PM
Room: Post Oak
Chair: Jilles Vreeken, Saarland University and Max Planck Institute for Informatics, Germany

Recent advances in computing resources have made possible collecting enormous amounts of data, such as social media interactions, web browsing, product and service purchases, autonomous vehicle routing, activities via smart home sensors and health/wellness sensors, and more. Since summarization helps humans find structure and meaning in data, the data mining community has taken a strong interest in the task and accordingly has proposed many methods for a variety of data types. This tutorial aims to provide a comprehensive overview of summarization techniques for large-scale graphs, which are very prevalent due to their intrinsic ability to represent many natural phenomena and encode relationships between entities.

Graph summarization, compression, and pattern discovery are tightly coupled: describing a graph succinctly leads to the discovery of interesting patterns, as well as to the detection of deviations or outliers. The objective of this tutorial is to give a systematic overview of graph summarization methods for static and dynamic networks with main emphasis on the important concepts, intuition, and main objectives, describe successful real-world applications, and present the open research problems in the field. Connections to compression and pattern/outlier discovery will be drawn throughout the tutorial.

Danai Koutra
University of Michigan, USA
2017 SIAM
International Conference on DATA MINING

April 27 - 29, 2017

The Westin Galleria Houston
Houston, Texas, USA

Abstracts are printed as submitted by the author.
2017 SIAM
International Conference on DATA MINING
April 27 - 29, 2017
The Westin Galleria Houston
Houston, Texas, USA
A
Abraham, Zubin, CP17, 10:55 Sat
Adhikari, Bijaya, CP10, 10:20 Fri
Akoglu, Leman, CP10, 10:00 Fri
Araujo, Miguel, CP13, 3:25 Fri

B
Bhattacharyya, Apratim, CP18, 1:30 Sat
Budhathoki, Kailash, CP12, 11:00 Fri

C
Cadena, Jose, CP14, 3:45 Fri
Cao, Bokai, CP5, 3:05 Thu
Chang, Yale, CP5, 3:45 Thu
Che, Chao, CP5, 3:25 Thu
Chen, Jinghui, CP3, 10:00 Thu
Chen, Lin, CP12, 10:40 Fri
Chen, Zhong, CP17, 10:35 Sat
Chino, Daniel, CP3, 10:40 Thu
Curtin, Ryan, CP7, 6:00 Thu

D
Dang, Xuan-Hong, CP8, 6:20 Thu
Danilevsky, Marina, CP6, 3:05 Thu
Davoudi, Heidar, CP4, 2:45 Thu
Dhurandhar, Amit, CP6, 2:45 Thu

E
Eswaran, Dhivya, CP4, 3:05 Thu
Evangelopoulos, Xenophon, CP11, 10:20 Fri

F
Fan, Wei, TS4, 10:00 Fri
Fan, Wei, TS4, 2:45 Fri
Fu, Xiao, CP5, 4:05 Thu
Fugate, William Craig, IP1, 8:15 Thu

G
Galbrun, Esther, CP8, 5:40 Thu
Gomez-Rodriguez, Manuel, CP18, 2:30 Sat

H
Hämäläinen, Wilhelmiina, CP7, 6:20 Thu
Hay, Michael, CP15, 3:25 Fri
He, Jingrui, CP16, 8:20 Sat
Heckel, Reinhard, CP15, 3:05 Fri
Hidalgo, Cesar, IP2, 1:15 Thu
Hu, Qiong, CP2, 10:40 Thu
Huang, Xiao, CP14, 4:05 Fri

I
Jermaine, Chris, PD1, 5:00 Fri
Jia, Xiaowei, CP4, 4:05 Thu
Jin, Fang, CP14, 3:05 Fri

K
Kakkar, Vishal, CP7, 5:40 Thu
Kanann, Ramakrishnan, CP11, 11:20 Fri
Khanna, Rajiv, CP12, 11:20 Fri
Korpela, Jussi, CP15, 4:25 Fri
Koutra, Danai, TS5, 3:45 Sat

L
Lash, Michael T., CP4, 3:45 Thu
Lee, Ching-Pei, CP16, 9:00 Sat
Lee, John Boaz, CP13, 2:45 Fri
Li, Jundong, CP10, 11:20 Fri
Li, Yaguang, CP17, 11:15 Sat
Li, Zhennui, CP18, 1:50 Sat
Liu, Honglei, CP2, 11:20 Thu
Liu, Xinyue, CP5, 2:45 Thu
Liu, Yashu, CP2, 10:00 Thu
Liu, Zongge, CP17, 11:35 Sat
Lu, Ziyu, CP9, 5:20 Thu

M
Malliaros, Fragkiskos D., CP8, 6:00 Thu
Mancuhan, Koray, CP15, 2:45 Fri
Mavroforakis, Charalampos, CP6, 4:05 Thu

N
Nassar, Huda, CP14, 3:25 Fri
Nelakurthi, Arun Reddy, CP11, 10:40 Fri
Nie, Zhi, CP1, 10:20 Thu
Ntouisi, Eirini C., CP9, 6:00 Thu

P
Peel, Leto, CP10, 11:00 Fri
Perdacher, Martin, CP7, 5:00 Thu
Perros, Ioakeim, CP2, 11:00 Thu
Pienta, Robert, CP14, 2:45 Fri
Polina, Rozenshtein, CP10, 10:40 Fri
Prakash, B. Aditya, TS2, 2:45 Thu

R
Rangwala, Huzefa, TS3, 5:00 Thu
Ruchansky, Natalia, CP6, 3:45 Thu
Rugeles, Daniel, CP11, 10:00 Fri

S
Salah, Aghiles, CP6, 3:25 Thu
Schoeneman, Frank, CP17, 10:15 Sat
Seybold, Martin P., CP18, 2:10 Sat
Siebes, Arno, CP3, 11:20 Thu

T
Tan, Chang Wei, CP7, 5:20 Thu
Tang, Jiliang, CP8, 5:20 Thu
Tatti, Nikolaj, CP18, 2:50 Sat
Tomita, Tyler M., CP12, 10:00 Fri
Tomkins, Andrew, IP4, 1:15 Fri

Italicized names indicate session organizers
W
Wang, Suhang, CP1, 10:00 Thu
Wang, Yilin, CP12, 10:20 Fri
Wistuba, Martin, CP16, 9:20 Sat
Wu, Liang, CP3, 10:20 Thu

X
Xiao, Houping, CP3, 11:00 Thu
Xue, Yanbing, CP1, 11:00 Thu

Y
Yang, Tao, CP9, 6:20 Thu
Yu, Hsiang-Fu, CP9, 5:00 Thu
Yuan, Shuai, CP15, 3:45 Fri

Z
Zaidi, Nayyar, CP16, 8:00 Sat
Zhang, Chuxu, CP9, 5:40 Thu
Zhang, Daoqiang, CP2, 10:20 Thu
Zhang, Daoqiang, CP1, 10:40 Thu
Zhang, Si, CP13, 3:45 Fri
Zhang, Yao, CP8, 5:00 Thu
Zhang, Yao, CP14, 4:25 Fri
Zhang, Yexun, CP4, 3:25 Thu
Zhao, Yan, CP13, 4:25 Fri
Zhou, Yao, CP13, 4:05 Fri

*Italicized names indicate session organizers*
## SDM17 Budget

### Conference Budget
SIAM International Conference on Data Mining  
April 27 - 29, 2017  
Houston, TX

### Expected Paid Attendance
225

#### Revenue

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration Income</td>
<td>$109,810.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$109,810.00</strong></td>
</tr>
</tbody>
</table>

#### Expenses

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printing</td>
<td>$1,100.00</td>
</tr>
<tr>
<td>Organizing Committee</td>
<td>$2,900.00</td>
</tr>
<tr>
<td>Invited Speakers</td>
<td>$8,350.00</td>
</tr>
<tr>
<td>Food and Beverage</td>
<td>$37,000.00</td>
</tr>
<tr>
<td>AV Equipment and Telecommunication</td>
<td>$15,500.00</td>
</tr>
<tr>
<td>Advertising</td>
<td>$6,500.00</td>
</tr>
<tr>
<td>Proceedings</td>
<td>$6,000.00</td>
</tr>
<tr>
<td>Conference Labor (including benefits)</td>
<td>$37,095.00</td>
</tr>
<tr>
<td>Other (supplies, staff travel, freight, misc.)</td>
<td>$5,050.00</td>
</tr>
<tr>
<td>Administrative</td>
<td>$12,101.00</td>
</tr>
<tr>
<td>Accounting/Distribution &amp; Shipping</td>
<td>$6,614.00</td>
</tr>
<tr>
<td>Information Systems</td>
<td>$12,161.00</td>
</tr>
<tr>
<td>Customer Service</td>
<td>$4,557.00</td>
</tr>
<tr>
<td>Marketing</td>
<td>$7,033.00</td>
</tr>
<tr>
<td>Office Space (Building)</td>
<td>$4,649.00</td>
</tr>
<tr>
<td>Other SIAM Services</td>
<td>$4,845.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$171,455.00</strong></td>
</tr>
</tbody>
</table>

Net Conference Expense  
- $61,645.00

Support Provided by SIAM  
$61,645.00

Estimated Support for Travel Awards not included above:

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Career and Students</td>
<td>18</td>
</tr>
</tbody>
</table>

**Support Provided by SIAM**  $61,645.00

Estimated Support for Travel Awards not included above:  $0.00