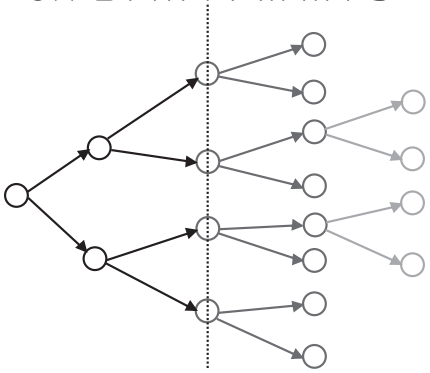
Final Program and Abstracts

2016 SIAM International Conference on DATA MINING



May 5-7, 2016
Hilton Miami Downtown
Miami, Florida, USA

Sponsored by the SIAM Activity Group on Data Mining and Analytics

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.

Society for Industrial and Applied Mathematics
3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Telephone: +1-215-382-9800 Fax: +1-215-386-7999
Conference Email: meetings@siam.org
Conference Web: www.siam.org/meetings/
Membership and Customer Service: (800) 447-7426 (US& Canada) or +1-215-382-9800 (worldwide)

www.siam.org/meetings/sdm16

Table of Contents

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

Qiang Yang
Hong Kong University of Science and Technology, USA

Philip Yu
University of Illinois at Chicago, USA

Qiang Yang
Hong Kong University of Science and Technology, USA

Conference Co-Chairs

Carlotta Domeniconi
George Mason University, USA

Ke Wang
Simon Fraser University, Canada

Program Co-Chairs

Sanjay Chawla
Qatar Computing Research Institute, Qatar and University of Sydney, Australia

Wagner Meira
Universidade Federal de Minas Gerais, Brazil

Workshop Co-Chairs

Tanya Berger-Wolf
University of Illinois, Chicago, USA

David Gleich
Purdue University, USA

Tutorial Chair

Tao Li
Florida International University, USA

Doctoral Forum Chair

Mitsunori Ogihara
University of Miami, USA

Publicity Co-Chairs

B. Aditya Prakash
Virginia Tech, USA

Venu Satuluri
Twitter, USA

Panel Chair

Wei Wang
University of California at Los Angeles, USA

Sponsorship Co-Chairs

Nitesh Chawla
University of Notre Dame, USA

Philip Kegelmeyer
Sandia National Laboratories, USA

Senior Program Committee

Charu Aggarwal
IBM T.J. Watson Research Center, USA

Leman Akoglu
SUNY Stony Brook, USA

Ira Assent
Aarhus University, Denmark

Arindam Banerjee
University of Minnesota, USA

Francesco Bonchi
Yahoo!, Spain

Andre Carvalho
ICMC-USP, Brazil

Nitesh Chawla
University of Notre Dame, USA

Diane Cook
Washington State University, USA

Carlotta Domeniconi
George Mason University, USA

Tina Eliassi-Rad
Rutgers University, USA

Aristides Gionis
Aalto University, Finland

Marta Gonzales
Massachusetts Institute of Technology, USA

Dimitrios Gunopulos
University of Athens, Greece

George Karypis
University of Minnesota, USA

Eamonn Keogh
University of California, Riverside, USA

Tamara Kolda
Sandia National Laboratories, USA

Ee-Peng Lim
Singapore Management University, Singapore

Yan Liu
University of Southern California, USA

Srini Parthasarthy
Ohio State University, USA

Dino Pedreschi
University of Pisa, Italy

Alexandre Plastino
Universidade Federal Fluminense, Brazil

Rajeev Rastogi
Amazon, India

Celine Robardet
INSA, France

Cynthia Rudin
Massachusetts Institute of Technology, USA

Arno Siebes
Utrecth University, Netherlands

Evimaria Terzi
Boston University, USA

Hanghang Tong
Arizona State University, USA

Caetano Traina
ICMC-USP, Brazil

Suresh Venkatasubr
University of Utah, USA

Fei Wang
University of Connecticut, USA

Wei Wang
University of California, Los Angeles, USA

Jieping Ye
University of Michigan, USA

Bianca Zadrozny
Universidade Federal Fluminense, Brazil

Mohammed Zaki
Rensselaer Polytechnic Institute, USA, USA

Yu Zheng
Microsoft, China

Program Committee

Aijun An
York University, Canada

Aris Anagnostopoulos
Sapienza University of Rome, Italy

Annalisa Appice
University of Bari Aldo Moro, Italy

Gowtham Atluri
University of Minnesota, USA

Tony Bagnall
University of East Anglia, United Kingdom

Gustavo Batista
University of São Paulo, Brazil

Christian Bauckhage
Bonn University, Germany

Fabricio Benevenuto
Federal University of Minas Gerais, Brazil

Kanishka Bhaduri
Netflix, USA

Albert Bifet
Huawei Noah's Ark Lab, Hong Kong

K. Selçuk Candan
Arizona State University, USA

Huiping Cao
New Mexico State University, USA

Nan Cao
IBM T. J. Watson Research Center, USA

Loic Cerf
Universidade Federal de Minas Gerais, Brazil

Shiyu Chan
University of Illinois at Urbana–Champaign, USA

Xiangyu Chang
Xi'an Jiaotong University, China

Yi Chang
Yahoo! Labs, USA

Polo Chau
Georgia Institute of Technology, USA

Feng Chen
University at Albany, SUNY, USA

Jianhui Chen
Yahoo! Labs, USA

Wei Chen
Microsoft Research Asia

Jian Cheng
Chinese Academy of Science, China

Yu Cheng
IBM T.J. Watson Research Center, USA

Robson Cordeiro
Universidade de São Paulo, Brazil

Alfredo Cuzzocrea
Italian National Research Council, Italy

Xuan-Hong Dang
University of California Santa Barbara, USA

Kamalika Das
NASA/University Affliated Research Center, USA

Mahashweta Das
Hewlett Packard Labs, USA

Hasan Davulcu
Arizona State University, USA

Sandra de Amo
Universidade Federal de Uberlândia, Brazil

Wei Ding
University of Massachusetts Boston, USA

Haimonti Dutta
Columbia University, USA

Saso Dzeroski
Jožef Stefan Institute, Slovenia

William Eberle
Tennessee Technological University, USA

Dora Erdos
Boston University, USA

Victor Fang
Awake Networks, USA

Sorelle Friedler
Haverford College, USA

Yun Fu
Northeastern University, USA

Amita Gajewar
Microsoft Research, USA

Esther Galbrun
Boston University, USA

Jing Gao
University of Buffalo, USA

Jun Gao
Peking University, China

Yong Ge
University of North Carolina Charlotte, USA

Mohamed Ghalwash
Temple University, USA

Joachim Giesen
Friedrich-Schiller-Universität Jena, Germany

David Gleich
Purdue University, USA

Manuel Gomez-Rodriguez
Max Planck Institute for Software Systems, Germany

Pinghua Gong
Arizona State University, USA

Ziyu Guan
Northwest University, China

Jie Gui
Chinese Academy of Science, China

Francesco Gullo
Yahoo! Research, Italy

Stephan Günnemann
Carnegie Mellon University, USA

Manish Gupta
Microsoft India, India

Yuan Hao
Google, USA

Bing Hu
Samsung Research, USA

Xia Hu
Arizona State University, USA

Xiaohua Hu
Drexel University, USA

Jun (Luke) Huan
University of Kansas, USA

Bo Jin
Dalian University of Technology, China

Chandrika Kamath
Lawrence Livermore National Laboratory, USA

Murat Kantarcioglu
University of Texas at Dallas, USA

Purushottam Kar
Microsoft Research India, India

Jaya Kawale
Adobe, USA

Jeremy Kepner
Massachusetts Institute of Technology, USA

Daniel Keren
Haifa University, Israel

Deguang Kong
Samsung, USA

Xiangnan Kong
Worcester Polytechnic Institute, USA

Georg Krempl
Otto-von-Guericke Universität Magdeburg, Germany

Cheng-Te Li
Academic Sinica, Taiwan

Geng Li
Yahoo!, USA

Jiuyong Li
University of South Australia, Australia

Lei Li
Florida International University, USA

Eduardo Ogasawara
Centro Federal de Educação Tecnológica (CEFET-RJ), Brazil

Nikunj Oza
NASA Ames Research Center, USA

Aline Paes
Universidade Federal Fluminense (UFF), Brazil

Spiros Papadimitriou
Rutgers University, USA

Gisele Papp
Universidade Federal de Minas Gerais, Brazil

Adriana Prado
EMC, Brazil

B. Aditya Prakash
Virginia Tech, USA

Buyue Qian
IBM T. J. Watson Research Center, USA

Mu Qiao
IBM, USA

Zhiwei Qin
Walmart Labs, USA

Xiaojun Quan
Institute for Infocomm Research, Singapore

Huzefa Rangwala
George Mason University, USA

Sayan Ranu
Indian Institute of Technology Madras, Spain

Chandan Reddy
Wayne State University, USA

Saket Sathe
IBM Australia

Fabrizio Silvestri
Yahoo! Labs, United Kingdom

Le Song
Georgia Tech, USA

Yangqiu Song
University of Illinois at Urbana–Champaign, USA

Sucheta Soundarajan
Rutgers University, USA

Elaine Sousa
Universidade de São Paulo, Brazil

Jerzy Stefanowski
Pozna University of Technology, Poland

Masashi Sugiyama
University of Tokyo, Japan

Yizhou Sun
Northeastern University, USA

Andrea Tagarelli
University of Calabria, Italy

Hana Tai
National Taipei University, Taiwan

Chenhao Tan
Cornell University, USA

Pang-Ning Tan
Michigan State University, USA

Jie Tang
Tsinghua University, China

Nikolaj Tatti
Aalto University, Finland

Panayiotis Tsaparas
University of Ioannina, Greece

Grigorios Tsoumakas
Aristotle Univ. Thessaloniki, Greece

Johan Ugander
University of Washington/Microsoft Research, USA

Antti Ukkonen
Helsinki University of Technology, Finland

Matthijs van Leeuwen
Katholieke Universiteit Leuven, Belgium

Adriano Veloso
Universidade Federal de Minas Gerais, Brazil

José Viterbo Filho
Universidade Federal Fluminense (UFF), Brazil

Jilles Vreeken
Max-Planck-Institute, Germany

Jianyong Wang
Tsinghua University, China

Jie Wang
Arizona State University, USA

Ke Wang
Simon Fraser University, Canada

Ting Wang
Lehigh University, USA

Wendy (Hui) Wang
Stevens Institute of Technology, USA

Xiang Wang
IBM T. J. Watson Research Center, USA

Zheng Wang
University of Michigan, USA

Takashi Washio
Osaka University, Japan

Ingmar Weber
Qatar Computing Research Institute, Qatar

Tim Weninger
University of Notre Dame, USA

Shuai Li
Qatar Computing Research Institute, Qatar

Yuefeng Li
Queensland University of Technology, Australia

Qin Lijing
Tsinghua University, China

Jessica Lin
George Mason University, USA

Shou-de Lin
National Taiwan University, Taiwan

Ji Liu
University of Rochester, USA

Qi Liu
University of Science and Technology of China, China

Wei Liu
University of Technology, Sydney, Australia

David Lo
Singapore Management University, Singapore

Chang-Tien Lu
Virginia Tech, USA

Shuai Ma
Beihang University, China

Arun Maiya
Institute for Defense Analysis, USA

Leandro Marinho
Universidade Federal de Campina Grande, Brazil

Julian McAuley
University of California, San Diego, USA

Rosa Meo
University of Turin, Italy

Luiz Merschmann
Universidade Federal de Ouro Preto (UFOP), Brazil

Pauli Miettinen
Max-Planck-Institute for Informatics, Germany

Benjamin Miller
MIT Lincoln Laboratory, USA

Abdullah Mueen
University of New Mexico, USA

Emmanuel Mülle
Karlsruhe Institute of Technology (KIT), Germany

Sriraam Natarajan
Indiana University, Bloomington, USA

Vinh Nguyen
University of Melbourne, Australia

Xia Ning
Indiana University - Purdue University Indianapolis, USA

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

Shuo Xiang
Arizona State University, USA

Keli Xiao
Stony Brook University, USA

Yanghua Xiao
Fudan University, China

Dawei Yin
Yahoo! Labs, USA

Guoxian Yu
Southwest University, China

Lei Yu
Binghamton University, USA

Dayu Yuan
Google, USA

Jing Yuan
Microsoft Research Asia

Lei Yuan
Dow AgroSciences, USA

Kun Zhang
Xavier University of Lousiana, USA

Lumin Zhang
Facebook, USA

Ping Zhang
IBM Thomas J. Watson Research Center, USA

Xiang Zhang
Case Western Reserve University, USA

Xiaoming Zhang
Beihang University, China

Zhao Zhang
Soochow University, China

Jiayu Zhou
Michigan State University, USA

Wenjun Zhou
University of Tennessee Knoxville, USA

Hengshu Zhu
Baidu.com, China

Shiai Zhu
University of Ottawa, Canada

Xingquan Zhu
Florida Atlantic University, USA

Arthur Zimek
Ludwig-Maximilians-Universitaet Muenchen, Germany

SIAM Registration Desk

The SIAM registration desk is located in the Symphony Ballroom Registration Area. It is open during the following hours:

Wednesday, May 4
5:00 PM – 7:00 PM

Thursday, May 5
7:00 AM – 7:00 PM

Friday, May 6
7:30 AM – 3:30 PM

Saturday, May 7
7:30 AM – 4:00 PM

Hotel Address

Hilton Miami Downtown
1601 Biscayne Boulevard
Miami, Florida 33132
Direct Telephone: +1-305-374-0000
Toll Free Reservations (US and Canada): +1-800-Hiltons
Fax: +1-305-714-3654
Hotel website:
http://www.hiltonmiamidowntown.com/

Hotel Telephone Number

To reach an attendee or leave a message, call +1-305-374-0000. 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 Hilton Miami Downtown provided the following list of child care providers for attendees interested in child care services. Attendees are responsible for making their own child care arrangements.

Nanny Poppinz of Miami
(305)767-2422
www.nannypoppinz.com
1825 Ponce De Leon Blvd. Miami

Nurse Core of Miami
(305)418-4005
www.nursecore.com
13468 Biscayne Blvd. Miami

The Babysitting Company
(305)890-7000
1-888-407-7822
www.thebabysittingcompany.com
6338 Collins Ave. Miami Beach

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.

Corporate Institutional Members

The Aerospace Corporation

Air Force Office of Scientific Research

Aramco Services Company

AT&T Laboratories - Research

Bechtel Marine Propulsion Laboratory

The Boeing Company

CEA/DAM

Department of National Defence (DND/ CSEC)

DSTO- Defence Science and Technology Organisation

ExxonMobil Upstream Research

Hewlett-Packard

IBM Corporation

IDA Center for Communications Research, La Jolla

IDA Center for Communications Research, Princeton

Institute for Computational and Experimental Research in Mathematics (ICERM)

Institute for Defense Analyses, Center for Computing Sciences

Lawrence Berkeley National Laboratory

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)

Oak Ridge National Laboratory, managed by UT-Battelle for the Department of Energy

Sandia National Laboratories

Schlumberger-Doll Research

Tech X Corporation

U.S. Army Corps of Engineers, Engineer Research and Development Center

United States Department of Energy

List current March 2016.

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.

Leading the applied mathematics community

Join SIAM and save!

SIAM members save up to $130 on full registration for the 2016 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 $10 to join the SIAM Activity Group on the Data Mining and Analytics (SIAG/ DMA). As a SIAG/DMA member, you are eligible for an additional $10 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 (U.S. 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.

Internet Access

Complimentary wireless Internet access will be available for SIAM attendees in the meeting space, and guestrooms for those attendees who booked within the SIAM room block.

SIAM will also provide a limited number of email stations for attendees 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
- Doctoral Forum and Poster Session
- Room set-ups and audio/visual equipment
- USB of conference proceedings, workshop and tutorial notes
- Welcome Reception and Poster Session

Job Postings

Please check with the SIAM registration desk regarding the availability of job postings or visit http://jobs.siam.org.

Important Notice to Poster Presenters

The poster sessions are scheduled for Thursday, May 5, 6:30 PM – 8:30 PM and Friday, May 6, 7:00 PM -9:00 PM. Papers being presented on Thursday and Saturday will have accompanying poster presentations during the Welcome Reception and Poster Session on Thursday, May 5. Boards and push pins for Thursday's poster session will be available beginning at 5:00 PM on Wednesday, May 4. Presenters are requested to put up their posters no later than 6:30 PM, the official start time of that session.

Papers being presented on Friday will have accompanying poster presentations during the Doctoral Forum and Poster Session on Friday, May 6. Boards and push pins for Friday's poster session will be available beginning at 9:00 AM on Friday, May 6. Presenters are requested to put up their posters no later than 7:00 PM, the official start time of that session.

For information about preparing a poster, please visit *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.

Conference Sponsors

Gold

Supporter

iCeNSA
Interdisciplinary Center for Network Science & Applications
University of Notre Dame

Doctoral Forum Reception Sponsors

UNIVERSITY OF MIAMI
COLLEGE of ARTS & SCIENCES

UNIVERSITY OF MIAMI
CENTER for COMPUTATIONAL SCIENCE

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:

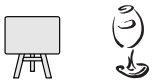
Cynthia Phillips, SIAM Vice President for Programs (*vpp@siam.org*).

Get-togethers

Welcome Reception and Poster Session

Thursday, May 5

6:30 PM – 8:30 PM

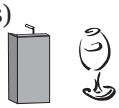


Business Meeting

(open to SIAG/DMA members)

Friday, May 6

6:30 PM – 7:00 PM



Complimentary beer and wine will be served.

Doctoral Forum and Poster Session

Friday, May 6

7:00 PM - 9:00 PM

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 laptop computers unattended. Please remember to turn off your cell phones, pagers, etc. 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 #SIAMSDM16.

Invited Plenary Speakers

All Invited Plenary Presentations will take place in Symphony Ballroom 1.

Thursday, May 5

8:15 AM - 9:30 AM

IP1 Big Data, Small Models, and Extreme Behaviors in the Real World

Neil F. Johnson, *University of Miami, USA*

1:30 PM - 2:45 PM

IP2 Title Not Available at Time of Publication

Virgilio Almeida, *Federal University of Minas Gerais, Brazil*

Friday, May 6

8:15 AM - 9:30 AM

IP3 Data and Algorithmic Bias in the Web

Ricardo Baeza-Yates, *Universitat Pompeu Fabra, Spain & Universidad de Chile, Chile*

1:30 PM - 2:45 PM

IP4 Sum-Product Networks: Deep Models with Tractable Inference

Pedro Domingos, *University of Washington, USA*

Tutorials

Thursday, May 5

10:00 AM - 12:00 PM

TS1: Tutorial Session: Mining Personal Traits in Social Media
Soprano

3:00 PM - 5:00 PM

TS2: Tutorial Session: Biomedical Data Mining with Matrix Models
Soprano

Friday, May 6

10:00 AM - 12:00 PM

TS3: Tutorial Session: Large Scale Hierarchical Classification: Foundations, Algorithms and Applications
Soprano

3:00 PM - 5:00 PM

TS4: Tutorial Session: Optimal Connectivity on Big Graphs: Measures, Algorithms and Applications
Soprano

Saturday, May 7

10:00 AM - 12:00 PM

TS5: Tutorial Session: Problems with Incomplete Networks: Biases, Skewed Results, and Solutions
Picasso

TS6: Tutorial Session: Towards Veracity Challenge in Big Data
Soprano

Workshops

Saturday, May 7
Workshops

8:30 AM - 5:00 PM (Full Day Workshop)

Workshop 1: Data Mining for Medicine and Healthcare

Symphony Ballroom I

Workshop 3: Mining Networks and Graphs:
A Big Data Analytic Challenge (MNG 2016)

Tenor

1:30 PM - 5:00 PM (Half Day Workshop)

Workshop 2: Machine Learning Methods for Recommender Systems (MLRec)

Picasso

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

siam®
Society for Industrial and Applied Mathematics • 3600 Market Street, 6th Floor • Philadelphia, PA 19104-2688 USA
Phone: +1-215-382-9800 • Fax +1-215-386-7999 • *service@siam.org* • *www.siam.org*

SIAM Activity Group on Data Mining (SIAG/DMA)

www.siam.org/activity/dma

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 $10 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:

- $10 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:

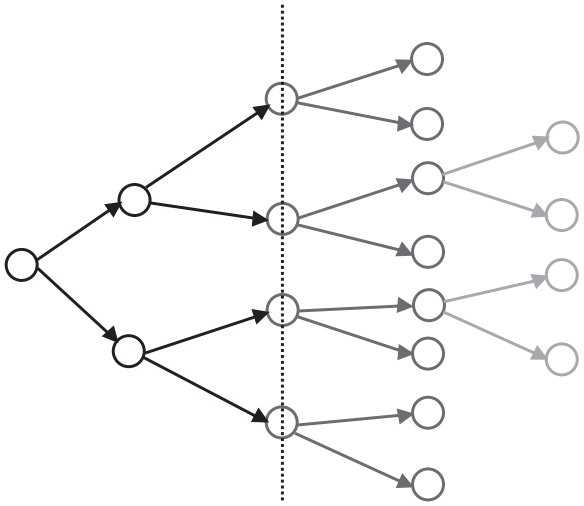
SIAG/DMA: my.siam.org/forms/join_siag.htm

SIAM: www.siam.org/joinsiam

Notes

SDM16 Program

2016 SIAM
International Conference
on DATA MINING



May 5-7, 2016
Hilton Miami Downtown
Miami, Florida, USA

Wednesday, May 4

Registration

5:00 PM-7:00 PM

Room:Symphony Ballroom Registration Area

Thursday, May 5

Registration

7:00 AM-7:00 PM

Room:Symphony Ballroom Registration Area

Continental Breakfast

7:30 AM-8:00 AM

Room:Symphony Ballroom II

Announcements

8:00 AM-8:15 AM

Room:Symphony Ballroom I

IP1

Big Data, Small Models, and Extreme Behaviors in the Real World

8:15 AM-9:30 AM

Room:Symphony Ballroom I

Chair: Carlotta Domeniconi, George Mason University, USA

Nobody would argue that data are getting bigger, not only in terms of the amount available but also in terms of its resolution — in time, in space, and now down toward the level of the individual objects in an otherwise large interconnected population. In the 'dark' ages, there was far too little data available to distinguish between multiple theories of how real-world complex systems work, be they from a biological system such as the brain; a social system such as an insurgency or terrorist campaign; or a financial market. However the tide has turned and available data now far outstrips theoretical understanding in these fields. In this talk I discuss the importance of developing minimal generative models that are consistent with the statistical features of the available big data being produced; I argue that the development of each should go hand in hand; and I argue that extreme real-world behaviors can be used to sort out the good from the bad in terms of what additional data to collect and how to mine it. Big data examples to be discussed range from collective neuronal processes in the brain, through to online global extremist activity and subsecond temporal fractures in global financial exchanges.

Neil F. Johnson

University of Miami, USA

Coffee Break

9:30 AM-10:00 AM

Room:Symphony Ballroom II

Thursday, May 5

TS1

Tutorial Session: Mining Personal Traits in Social Media

10:00 AM-12:00 PM

Room:Soprano

Chair: Tao Li, Florida International University, USA

There has been rapid growth in applications informed by social media data in domains such as public health, marketing, and political science. A critical component in such applications is the need to "understand social users better." In this tutorial, the presenters attempt to present the following

- Introduction of the personal traits mining problem in social media;
- Landscape of recent developments toward the problem;
- Presentation of representative supervised vs. semi/unsupervised solutions;
- iPython based hands-on demonstration or exercise
- Implications and open issues

Aron Culotta
Illinois Institute of Technology, USA;

Dongwon Lee
Pennsylvania State University, USA

Thursday, May 5

CP1

New Methods and Applications

10:00 AM-12:00 PM

Room:Tenor

Chair: To Be Determined

10:00-10:15 Clustering in the Face of Fast Changing Streams
Liudmila Ulanova, Nurjahan Begum, Mohammad Shokoohi-Yekta, and Eamonn Keogh, University of California, Riverside, USA

10:20-10:35 Kernelized Sparse Self-Representation for Clustering and Recommendation
Xiao Bian, GE Global Research, USA; Feng Li, and Xia Ning, Indiana University - Purdue University Indianapolis, USA

10:40-10:55 Multi-Domain Manifold Learning for Drug-Target Interaction Prediction
Ruichu Cai, Guangdong University of Technology, China

11:00-11:15 Finding Surprisingly Frequent Patterns of Variable Lengths in Sequence Data
Reza Sadoddin, Joerg Sander, and Davood Rafiei, University of Alberta, Canada

11:20-11:35 Identifying Connectivity Patterns for Brain Diseases Via Multi-Side-View Guided Deep Architectures
Jingyuan Zhang, Bokai Cao, Sihong Xie, Chun-Ta Lu, and Philip Yu, University of Illinois, Chicago, USA; Ann Ragin, Northwestern University, USA

11:40-11:55 Regularized Weighted Linear Regression for High-Dimensional Censored Data
Yan Li, Bhanukiran Vinzamuri, and Chandan Reddy, Wayne State University, USA

Thursday, May 5

CP2

Network Analysis

10:00 AM-12:00 PM

Room:Symphony Ballroom I

Chair: To Be Determined

10:00-10:15 Node Classification in Signed Social Networks
Jiliang Tang, Yahoo! Labs, USA; Charu C. Aggarwal, IBM T.J. Watson Research Center, USA; Huan Liu, Arizona State University, USA

10:20-10:35 Uncovering Multiple Diffusion Networks Using the First-Hand Sharing Pattern
Pei-Lun Liao, *Chung-Kuang Chou*, and Ming-Syan Chen, National Taiwan University, Taiwan

10:40-10:55 Integrating Community and Role Detection in Information Networks
Ting Chen, Northeastern University, USA; Lu-An Tang, NEC Laboratories America, USA; Yizhou Sun, Northeastern University, USA; Zhengzhang Chen, Haifeng Chen, and Guofei Jiang, NEC Laboratories America, USA

11:00-11:15 Exploiting Emotional Information for Trust/Distrust Prediction
Ghazaleh Beigi, Arizona State University, USA; Jiliang Tang, Yahoo! Labs, USA; Suhang Wang and Huan Liu, Arizona State University, USA

11:20-11:35 Online Prediction of User Actions Through An Ensemble Vote from Vector Representation and Frequency Analysis Models
Changsung Moon, Dakota Medd, Paul Jones, and Steve Harenberg, North Carolina State University, USA; William Oxbury, Heilbronn Institute, England; Nagiza Samatova, North Carolina State University and Oak Ridge National Laboratory, USA

11:40-11:55 FairPlay: Fraud and Malware Detection in Google Play
Bogdan Carbunar, Mahmudur Rahman, and *Mizanur Rahman*, Florida International University, USA; Duen Horng Chau, Georgia Institute of Technology, USA

Thursday, May 5

CP3

Recommendation Systems and Factorization

10:00 AM-12:00 PM

Room:Picasso

Chair: To Be Determined

10:00-10:15 Synergies That Matter: Efficient Interaction Selection Via Sparse Factorization Machine

Jianpeng Xu, Kaixiang Lin, Pang-Ning Tan, and Jiayu Zhou, Michigan State University, USA

10:20-10:35 A Spatial-Temporal Probabilistic Matrix Factorization Model for Point-of-Interest Recommendation

Huayu Li, University of North Carolina, Charlotte, USA; Richang Hong, Hefei University of Technology, China; Zhiang Wu, Nanjing University of Finance and Economics, China; Yong Ge, University of North Carolina, Charlotte, USA

10:40-10:55 Top-N Recommendation with Novel Rank Approximation

Zhao Kang and Qiang Cheng, Southern Illinois University, Carbondale, USA

11:00-11:15 Vocal Competence Based Karaoke Recommendation: A Maximum-Margin Joint Model

Chu Guan, University of Science and Technology of China, China; Yanjie Fu, Rutgers University, USA; Xinjiang Lu, Northwestern Polytechnical University, China; Hui Xiong, Rutgers University, USA; Enhong Chen and Yingling Liu, University of Science and Technology of China, China

11:20-11:35 A Confidence-Based Approach for Balancing Fairness and Accuracy

Benjamin Fish, Jeremy Kun, and *Adam D. Lelkes*, University of Illinois, Chicago, USA

11:40-11:55 Differentially Private Significance Testing on Paired-Sample Data

Christine M. Task, Knexus Research, USA; Chris Clifton, Purdue University, USA

Thursday, May 5

Lunch Break

12:00 PM-1:30 PM

Attendees on their own

IP2

Title Not Available at Time of Publication

1:30 PM-2:45 PM

Room:Symphony Ballroom I

Chair: Wagner Meira, Universidade Federal de Minas Gerais, Brazil

Abstract not available at time of publication.

Virgilio Almeida
Federal University of Minas Gerais, Brazil

Coffee Break

2:45 PM-3:00 PM

Room:Symphony Ballroom II

Thursday, May 5

TS2

Tutorial Session: Biomedical Data Mining with Matrix Models

3:00 PM-5:00 PM

Room:Soprano

Chair: Tao Li, Florida International University, USA

In the last decade, advances in high-throughput technologies, growth of clinical data warehouses, and rapid accumulation of biomedical knowledge provided unprecedented opportunities and challenges to researchers in biomedical informatics. One distinct solution, to efficiently conduct big data analytics for biomedical problems, is the application of matrix computation and factorization methods such as non-negative matrix factorization, joint matrix factorization, tensor factorization. Compared to probabilistic and information theoretic approaches, matrix-based methods are fast, easy to understand and implement. In this tutorial, we provide a review of recent advances in algorithms and methods using matrix and their potential applications in biomedical informatics. We survey various related articles from data mining venues as well as from biomedical informatics venues to share with the audience key problems and trends in matrix computation research, with different novel applications such as drug repositioning, personalized medicine, and electronic phenotyping.

Fei Wang
University of Connecticut, USA

Ping Zhang
IBM T.J. Watson Research Center, USA

Thursday, May 5

CP4

Outlier Detection

3:00 PM-5:00 PM

Room:Tenor

Chair: To Be Determined

3:00-3:15 A General Framework to Increase the Robustness of Model-Based Change Point Detection Algorithms to Outliers and Noise
Xi Chen, University of Minnesota, USA; Yuanshun Yao, University of California, Santa Barbara, USA; Sichao Shi, University of California, San Diego, USA; Snigdhansu Chatterjee and Vipin Kumar, University of Minnesota, USA; James Faghmous, Icahn School of Medicine at Mount Sinai, USA

3:20-3:35 LODES: Local Density Meets Spectral Outlier Detection
Saket Sathe and Charu C. Aggarwal, IBM T.J. Watson Research Center, USA

3:40-3:55 Routine Mining Based Anomaly Detection in Mobile Phone Data
Tian Qin and Guojie Song, Peking University, China

4:00-4:15 A Scalable Approach for Outlier Detection in Edge Streams Using Sketch-Based Approximations
Stephen Ranshous, Steve Harenberg, and Kshitij Sharma, North Carolina State University, USA; Nagiza Samatova, North Carolina State University and Oak Ridge National Laboratory, USA

4:20-4:35 R1stm: One-Class Support Tensor Machine with Randomised Kernel
Sarah M. Erfani, University of Melbourne, Australia

4:40-4:55 Scalable Anomaly Ranking of Attributed Neighborhoods
Bryan Perozzi and Leman Akoglu, Stony Brook University, USA

Thursday, May 5

CP5

Advanced Classification

3:00 PM-5:00 PM

Room:Picasso

Chair: To Be Determined

3:00-3:15 Linear and Kernel Classification: When to Use Which?
Hsin-Yuan Huang, National Taiwan University, Taiwan

3:20-3:35 Pattern Aided Classication
Guozhu Dong and Vahid Taslimitehrani, Wright State University, USA

3:40-3:55 Tersesvm : A Scalable Approach for Learning Compact Models in Large-Scale Classification
Rohit Babbar, Max Plank Institute, Germany; Krikamol Muandet, Max Planck Institute for Intelligent Systems, Germany

4:00-4:15 Discriminative Training of Structured Dictionaries Via Block Orthogonal Matching Pursuit
Wenling Shang, University of Michigan, USA

4:20-4:35 A Unified View of Localized Kernel Learning
John Moeller, *Sarathkrishna Swaminathan*, and Suresh Venkatasubramanian, University of Utah, USA

4:40-4:55 Binary Classifier Calibration Using an Ensemble of Linear Trend Estimation
Mahdi Pakdaman Naeini and Gregory Cooper, University of Pittsburgh, USA

Thursday, May 5

CP6

Clustering and Embedding

3:00 PM-5:00 PM

Room:Symphony Ballroom I

Chair: To Be Determined

3:00-3:15 Power Simultaneous Spectral Data Embedding and Clustering
Kais Allab, Lazhar Labiod, and Mohamed Nadif, Paris Descartes, France

3:20-3:35 Process Trace Clustering: A Heterogeneous Information Network Approach
Phuong Nguyen, University of Illinois at Urbana-Champaign, USA; Aleksander Slominski, Vinod Muthusamy, and Vatche Ishakian, IBM T.J. Watson Research Center, USA; Klara Nahrstedt, University of Illinois at Urbana-Champaign, USA

3:40-3:55 Lagrangian Constrained Clustering
Mohadeseh Ganji, University of Melbourne, Australia

4:00-4:15 Fast Multiplier Methods to Optimize Non-Exhaustive, Overlapping Clustering
Yangyang Hou, Purdue University, USA; Joyce Whang, Sungkyunkwan University, Korea; David F. Gleich, Purdue University, USA; Inderjit S. Dhillon, University of Texas at Austin, USA

4:20-4:35 Stochastic Co-Clustering for Document-Term Data
Aghiles Salah, Nicoleta Rogovschi, and Mohamed Nadif, Paris Descartes, France

4:40-4:55 On Finding the Maximum Edge Biclique in a Bipartite Graph: A Subspace Clustering Approach
Eran Shaham, Honghai Yu, and Xiao-Li Li, Institute for Infocomm Research, Singapore

Organizational Break

5:00 PM-5:15 PM

Thursday, May 5

CP7

Scaling Clustering Algorithms Part I

5:15 PM-6:15 PM

Room:Tenor

Chair: To Be Determined

5:15-5:30 Geometric Methods to Accelerate *k*-Means Algorithms
Petr Ryšavý and Greg Hamerly, Baylor University, USA

5:35-5:50 Pivot-Based K-Means Algorithm for Numerous-Class Data Sets
Takashi Hattori, NTT Corporation, Japan

5:55-6:10 *k*-Means for Streaming and Distributed Big Sparse Data
Artem Barger and Danny Feldman, Haifa University, Israel

Thursday, May 5

CP8

Scaling Clustering Algorithms Part II

5:15 PM-6:15 PM

Room:Picasso

Chair: Sanjay Chawla, Qatar Computing Research Institute, Qatar and University of Sydney, Australia

5:15-5:30 *Halite*$_{ds}$: Fast and Scalable Subspace Clustering for Multidimensional Data Streams
Afonso E. Da Silva, Lucas L. Sanches, Antonio C. Fraideinberze, and *Robson L. F. Cordeiro*, University of Sao Paulo, Brazil

5:35-5:50 Online Clustering of Multivariate Time-Series
Masud Moshtaghi, Christopher Leckie, and James Bezdek, The University of Melbourne, Australia

5:55-6:10 Learning A Task-Specific Deep Architecture For Clustering
Zhangyang Wang and Shiyu Chang, University of Illinois at Urbana-Champaign, USA; Jiayu Zhou, Michigan State University, USA; Meng Wang, Hefei University of Technology, China; Thomas Huang, University of Illinois at Urbana-Champaign, USA

Welcome Reception and Poster Session

6:30 PM-8:30 PM

Room:Symphony Ballroom II

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

Friday, May 6

Continental Breakfast

7:30 AM-8:00 AM

Room:Symphony Ballroom II

Registration

7:30 AM-3:30 PM

Room:Symphony Ballroom Registration Area

Announcements

8:00 AM-8:15 AM

Room:Symphony Ballroom I

Friday, May 6

IP3

Data and Algorithmic Bias in the Web

8:15 AM-9:30 AM

Room:Symphony Ballroom I

Chair: Sanjay Chawla, Qatar Computing Research Institute, Qatar and University of Sydney, Australia

The Web is the largest public big data repository that humankind has created. In this overwhelming data ocean we need to be aware of the quality and in particular, of biases the exist in this data, such as redundancy, spam, etc. These biases affect the algorithms that we design to improve the user experience. This problem is further exacerbated by biases that are added by these algorithms, specially in the context of recommendation systems. We give several examples and their relation to sparsity, novelty, and privacy, stressing the importance of the user context to avoid these biases.

Ricardo Baeza-Yates
Universitat Pompeu Fabra, Spain & Universidad de Chile, Chile

Coffee Break

9:30 AM-10:00 AM

Room:Symphony Ballroom II

Friday, May 6

TS3

Tutorial Session: Large Scale Hierarchical Classification: Foundations, Algorithms and Applications

10:00 AM-12:00 PM

Room:Soprano

Chair: Tao Li, Florida International University, USA

The tutorial will cover technical material related to large scale hierarchical classification. This will be meant for an audience with intermediate expertise in data mining having a background in classification (supervised learning). Formal definitions of hierarchical classification and variants will be discovered, along with a brief discussion on structured learning.

Huzefa Rangwala
George Mason University, USA

Azad Naik
George Mason University, USA

Friday, May 6

CP9

Feature Selection

10:00 AM-12:00 PM

Room:Tenor

Chair: To Be Determined

10:00-10:15 Kernelized Matrix Factorization for Collaborative Filtering
Xinyue Liu, Worcester Polytechnic Institute, USA; Charu C. Aggarwal, IBM T.J. Watson Research Center, USA; Yu-Feng Li, Nankai University, China; Xiangnan Kong and Xinyuan Sun, Worcester Polytechnic Institute, USA; Saket Sathe, IBM T.J. Watson Research Center, USA

10:20-10:35 Robust Unsupervised Feature Selection on Networked Data
Jundong Li, Arizona State University, USA; Xia Hu, Texas A&M University, USA; Liang Wu and Huan Liu, Arizona State University, USA

10:40-10:55 Euclidean Co-Embedding of Ordinal Data for Multi-Type Visualization
Dung D. Le and Hady W. Lauw, Singapore Management University, Singapore

11:00-11:15 K-Nearest Neighbor Search and Outlier Detection Via Minimax Distances
Morteza Haghir Chehreghani, Xerox Research Centre Europe, France

11:20-11:35 Nonlinear Joint Unsupervised Feature Selection
Xiaokai Wei, Bokai Cao, and Philip S. Yu, University of Ilinois at Chicago, USA

11:40-11:55 A Framework to Adjust Dependency Measure Estimates for Chance
Simone Romano, Vinh Nguyen, James Bailey, and Karin Verspoor, University of Melbourne, Australia

Friday, May 6

CP10

Applications

10:00 AM-12:00 PM

Room:Picasso

Chair: To Be Determined

10:00-10:15 Risk Prediction with Electronic Health Records: A Deep Learning Approach
Yu Cheng, IBM T.J. Watson Research Center, USA

10:20-10:35 Predicting the Popularity of News Articles
Yaser Keneshloo, Virginia Tech, USA; *Shuguang Wang* and Eui-Hong (Sam) Han, The Washington Post, Washington, DC, USA; Naren Ramakrishnan, Virginia Tech, USA

10:40-10:55 Uncovering Latent Behaviors in Ant Colonies
Mohamed Kafsi, École Polytechnique Fédérale de Lausanne, Switzerland

11:00-11:15 The Impact of Community Safety on House Ranking
Zijun Yao, Yanjie Fu, Bin Liu, and Hui Xiong, Rutgers University, USA

11:20-11:35 *IPath* Forecasting the Pathway to Impact
Liangyue Li, and Hanghang Tong, Arizona State University, USA; Jie Tang, Tsinghua University, P. R. China; Wei Fan, Baidu, USA

11:40-11:55 Cost-Sensitive Batch Mode Active Learning: Designing Astronomical Observation by Optimizing Telescope Time and Telescope Choice
Xide Xia, Finale Doshi-Velez, and Pavlos Protopapas, Harvard University, USA

Friday, May 6

CP11

Graphs, Networks, Communities

10:00 AM-12:00 PM

Room:Symphony Ballroom I

Chair: To Be Determined

10:00-10:15 A Fast Kernel for Attributed Graphs
Yu Su and Fangqiu Han, University of California, Santa Barbara, USA; Richard Harang, U.S. Army Research Laboratory, USA; Xifeng Yan, University of California, Santa Barbara, USA

10:20-10:35 Birdnest: Bayesian Inference for Ratings-Fraud Detection
Bryan Hooi, Neil Shah, and Alex Beutel, Carnegie Mellon University, USA; Stephan Gunneman, Technical University of Munich, Germany; Leman Akoglu, Stony Brook University, USA; Mohit Kumar and Disha Makhija, Flipkart, India; Christos Faloutsos, Carnegie Mellon University, USA

10:40-10:55 Unstable Communities in Network Ensembles
Aditya Prakash, Virginia Tech, USA

11:00-11:15 CamIp: Confidence-Aware Modulated Label Propagation
Yuto Yamaguchi, University of Tsukuba, Japan; Christos Faloutsos, Carnegie Mellon University, USA; Hiroyuki Kitagawa, University of Tsukuba, Japan

11:20-11:35 Query-Driven Maximum Quasi-Clique Search
Pei Lee and Laks V.S. Lakshmanan, University of British Columbia, Canada

11:40-11:55 Distributed Representations of Expertise
Fangqiu Han, Shulong Tan, and Huan Sun, University of California, Santa Barbara, USA; Mudhakar Srivatsa, IBM T.J. Watson Research Center, USA; Deng Cai, Zhejiang University, China; Xifeng Yan, University of California, Santa Barbara, USA

Lunch Break

12:00 PM-1:30 PM

Attendees on their own

Friday, May 6

IP4

Sum-Product Networks: Deep Models with Tractable Inference

1:30 PM-2:45 PM

Room:Symphony Ballroom I

Chair: : Ke Wang, Simon Fraser University, Canada

Sum-product networks (SPNs) are a new class of deep probabilistic models where inference remains tractable regardless of the number of hidden layers. I will present generative and discriminative algorithms for learning SPN weights, and an algorithm for learning SPN structure. SPNs have achieved impressive results in a wide variety of domains, including object recognition, image completion, activity recognition, language modeling, collaborative filtering, and click prediction. Our algorithms can easily learn SPNs with many layers of latent variables, making them arguably the most powerful type of deep learning to date. (Joint work with Abe Friesen, Rob Gens and Hoifung Poon.)

Pedro Domingos

University of Washington, USA

Coffee Break

2:45 PM-3:00 PM

Room:Symphony Ballroom II

Friday, May 6

TS4

Tutorial Session: Optimal Connectivity on Big Graphs: Measures, Algorithms and Applications

3:00 PM-5:00 PM

Room:Soprano

Chair: Tao Li, Florida International University, USA

Graph mining has been playing a pivotal role in many disciplines. Among others, a common and fundamental property of the graphs arising from these domains is connectivity. The goal of this tutorial is to (1) provide a concise review of the recent advances in optimizing graph connectivity and its applications; and (2) identify the open challenges and future trends. We believe this is an emerging, high-impact topic in graph mining, which will attract both researchers and practitioners in the data mining research community. Our emphasis will be on (1) the recent emerging techniques on addressing graph connectivity optimization problem; and (2) the open challenges/future trends, with a careful balance between the theories, algorithms and applications.

Hanghang Tong
Arizona State University, USA

Friday, May 6

CP12

Temporal, Stream and Pattern Mining

3:00 PM-5:00 PM

Room:Symphony Ballroom I

Chair: To Be Determined

3:00-3:15 Temporal Kernel Descriptors for Learning with Time-Sensitive Patterns
Doyen Sahoo, Abhishek Sharma, and Steven Hoi, Singapore Management University, Singapore; Peilin Zhao, Institute for Infocomm Research, Singapore

3:20-3:35 Modelling Recurrent Events for Improving Online Change Detection
Alexandr Maslov and Mykola Pechenizkiy, Eindhoven University of Technology, Netherlands; Indré Žliobaité, Aalto University, Finland; Tommi K, University of Jyvaskyla, Finland

3:40-3:55 Macfp: Maximal Approximate Consecutive Frequent Pattern Mining under Edit Distance
Jingbo Shang, Jian Peng, and Jiawei Han, University of Illinois at Urbana-Champaign, USA

4:00-4:15 Dpclass: An Effective But Concise Discriminative Patterns-Based Classification Framework
Jingbo Shang, Wenzhu Tong, Jian Peng, and Jiawei Han, University of Illinois at Urbana-Champaign, USA

4:20-4:35 Fast Lossless Frequent Itemset Mining in Data Streams Using Crucial Patterns
Ariyam Das and Carlo Zaniolo, University of California, Los Angeles, USA

4:40-4:55 Flexibly Mining Better Subgroups
Hoang Vu Nguyen, Max Planck Institute for Informatics, Germany; *Jilles Vreeken*, Max-Planck-Institut fuer Informatik, Germany

Friday, May 6

CP13

Inference with Optimization

3:00 PM-5:00 PM

Room:Tenor

Chair: To Be Determined

3:00-3:15 Deterministic Column Sampling for Low-Rank Matrix Approximation: Nystrom vs. Incomplete Cholesky Decomposition
Tom Goldstein, University of Maryland, USA

3:20-3:35 A Polynomial Expansion Line Search for Large-Scale Unconstrained Optimization of Smooth L2-Regularized Loss Functions, with Implementation in Apache Spark
Michael Hynes, University of Waterloo, Canada

3:40-3:55 Structured Regression on Multilayer Networks
Athanasia Polychronopoulou and Zoran Obradovic, Temple University, USA

4:00-4:15 RelSim: Relation Similarity Search in Schema-Rich Heterogeneous Information Networks
Chenguang Wang, Peking University, China

4:20-4:35 Collective Opinion Spam Detection Using Active Inference
Shebuti Rayana, Stony Brook University, USA

4:40-4:55 Discovery of Precursors to Adverse Events Using Time Series Data
Vijay Manikanda Janakiraman, Bryan Matthews, and Nikunj Oza, NASA Ames Research Center, USA

Friday, May 6

CP14

Structured Learning

3:00 PM-5:00 PM

Room:Picasso

Chair: To Be Determined

3:00-3:15 Effective Crowd Expertise Modeling Via Cross Domain Sparsity and Uncertainty Reduction
Sihong Xie, Qingbo Hu, Weixiang Shao, and Jingyuan Zhang, University of Illinois, Chicago, USA; Jing Gao, State University of New York at Buffalo, USA; Wei Fan, IBM T.J. Watson Research Center, USA; Philip S. Yu, University of Ilinois at Chicago, USA

3:20-3:35 Gspartan: a Geospatio-Temporal Multi-Task Learning Framework for Multi-Location Prediction
Jianpeng Xu, Pang-Ning Tan, Lifeng Luo, and Jiayu Zhou, Michigan State University, USA

3:40-3:55 Learning Correlative and Personalized Structure for Online Multi-Task Classification
Peng Yang, Giangxia Li, Peilin Zhao, Xiao-Li Li, and Sujatha Das Gollapalli, Institute for Infocomm Research, Singapore

4:00-4:15 Online Sparse Passive Aggressive Learning with Kernels
Jing Lu, Singapore Management University, Singapore; Peilin Zhao, Institute for Infocomm Research, Singapore; *Steven Hoi*, Singapore Management University, Singapore

4:20-4:35 ADMM for Training Sparse Structural SVMs with Augmented ℓ Regularizers
Balamurugan Palaniappan, Inria, France; *Anusha Posinasetty* and Shirish Shevade, Indian Institute of Science, Bangalore, India

4:40-4:55 Structural Orthogonal Procrustes Regression for Face Recognition with Pose Variations and Misalignment
Ying Tai, Jian Yang, Fanlong Zhang, Yigong Zhang, and Lei Luo, Nanjing University of Science & Technology, China; Jianjun Qian, Nanjing Normal University, China

Friday, May 6

Organizational Break

5:00 PM-5:15 PM

PD1

NSF Panel

5:15 PM-6:30 PM

Room:Symphony Ballroom I

SIAG/DMA Business Meeting

6:30 PM-7:00 PM

Room:Symphony Ballroom I

Complimentary beer and wine will be served.

Doctoral Forum and Poster Session

7:00 PM-9:00 PM

Room:Symphony Ballroom II

Papers presented on Friday will have their poster slots during the Doctoral Forum session.

Saturday, May 7

Registration

7:30 AM-4:00 PM

Room:Symphony Ballroom Registration Area

Continental Breakfast

8:00 AM-8:30 AM

Room:Symphony Ballroom II

Saturday, May 7

CP15

Probalistic Inference Part I

8:30 AM-9:30 AM

Room:Picasso

Chair: To Be Determined

8:30-8:45 Capricorn: An Algorithm for Subtropical Matrix Factorization
Sanjar Karaev, Max-Planck Institute for Informatics, Germany; Pauli Miettinen, Max Planck Institute for Informatics, Germany

8:50-9:05 Automatic Unsupervised Tensor Mining with Quality Assessment
Evangelos E. Papalexakis, Carnegie Mellon University, USA

9:10-9:25 Rank Selection for Non-Negative Matrix Factorization Using Normalized Maximum Likelihood Coding
Yu Ito, University of Tokyo, Japan; Shin-Ichi Oeda, National Institute of Technology, Kisarazu College, Japan; Kenji Yamanishi, University of Tokyo, Japan

Saturday, May 7

CP16

Probalistic Inference Part II

8:30 AM-9:30 AM

Room:Soprano

Chair: To Be Determined

8:30-8:45 Sparse Hybrid Variational-Gibbs Algorithm for Latent Dirichlet Allocation
Ximing Li, Jihong Ouyang, and Xiaotang Zhou, Jilin University, China

8:50-9:05 Scaling Lifted Probabilistic Inference and Learning Via Graph Databases
Mayukh Das, Indiana University Bloomington, USA; Yuqing Wu, Pomona College, USA; Tushar Khot, Allen Institute for Artificial Intelligence, USA; Kristian Kersting, Technical University of Dortmund, Germany; Sriraam Natarajan, Indiana University, USA

9:10-9:25 Estimating Posterior Ratio for Classification: Transfer Learning from Probabilistic Perspective
Song Liu and Kenji Fukumizu, The Institute of Statistical Mathematics, Japan

Saturday, May 7

Workshop 3 (full day): Mining Networks and Graphs: A Big Data Analytic Challenge (MNG 2016)

8:30 AM-5:00 PM

Room:Tenor

Workshop 1 (full day): Data Mining for Medicine and Healthcare

8:30 AM-5:00 PM

Room:Symphony Ballroom I

Coffee Break

9:30 AM-10:00 AM

Room:Symphony Ballroom II

Saturday, May 7

TS5

Tutorial Session: Problems with Incomplete Networks: Biases, Skewed Results, and Solutions

10:00 AM-12:00 PM

Room:Picasso

Chair: Tao Li, Florida International University, USA

Networked representations of physical and social phenomena are often incomplete because the phenomena are partially observed. Working with incomplete networks can skew analyses. Hoping to acquire the full data is often unrealistic, but one may be able to collect data selectively to enrich the incomplete network. For example, suppose a cyber-network administrator has partially observed a network through trace-routes. Which parts of the partially observed network should be more closely examined to give the best (i.e., most complete) view of the entire network? With a limited query budget, how should this further exploration be done? Alternatively, suppose that one has obtained a sample of a Twitter retweet network from a Web site. The sample was collected for some other purpose (unbeknownst to us), and so may not contain the most useful structural information for ones purposes. How should one best supplement this sampled data? This tutorial addresses the aforementioned questions.

Tina Eliassi-Rad
Rutgers University, USA

Sucheta Soundarajan
Syracuse University, USA

Ali Pinar
Sandia National Laboratories, USA

Brian Gallagher
Lawrence Livermore National Laboratory, USA

Saturday, May 7

TS6

Tutorial Session: Towards Veracity Challenge in Big Data

10:00 AM-12:00 PM

Room:Soprano

Chair: Tao Li, Florida International University, USA

Big data leads to big challenges, not only in the volume of data but also in its velocity, variety and veracity. Especially, the veracity issue poses great difficulty to many decision making tasks when the data contains inaccurate or even false information that could mislead the decisions and eventually result in invaluable loss. Unfortunately, we cannot expect real-world data to be clean and accurate. Instead, data inconsistency, ambiguity and uncertainty widely exist. Such ubiquitous veracity problems motivate numerous efforts towards improving the information quality, trustworthiness and reliability. The efforts are taken from different perspectives to identify reliable information sources and trustworthy claims: 1) A series of approaches has recently been developed to estimate source reliability and detect true claims simultaneously by examining the relationship between sources and claims, and 2) some other approaches infer the trustworthiness of claims or the reliability of sources by building analytic models based on claims or sources' features. Due to their important roles in solving the veracity issue, the approaches in both categories have attracted considerable attention, but a combined view of both types of approaches has never been presented. To answer the need of a systematic introduction and comparison of the field, we will present an organized picture towards the veracity issue in this tutorial.

Jing Gao
State University of New York at Buffalo, USA

Qi Li
SUNY College at Buffalo, USA

Bo Zhao
LinkedIn, USA

Wei Fan
IBM T.J. Watson Research Center, USA

Jiawei Han
University of Illinois at Urbana-Champaign, USA

continued in next column

Saturday, May 7

CP17

High-dimensional Analysis

10:00 AM-12:00 PM

Room:Alto

Chair: To Be Determined

10:00-10:15 Constrained Group Testing to Predict Binding Response of Candidate Compounds
Paul Quint, Stephen Scott, N. V. Vinodchandran, and Brad Worley, University of Nebraska, USA

10:20-10:35 Regularized Parametric Regression for High-Dimensional Survival Analysis
Yan Li, Wayne State University, USA; Kevin Xu, University of Toledo, USA; Chandan Reddy, Wayne State University, USA

10:40-10:55 Copula-HDP-HMM: Non-Parametric Modeling of Temporal Multivariate Data for I/O Efficient Bulk Cache Preloading
Lavanya S. Tekumalla, Chiranjib Bhattacharyya, and *Anusha Posinasetty*, Indian Institute of Science, Bangalore, India

11:00-11:15 On Skewed Multi-Dimensional Distributions: the *fusionrp* Model, Algorithms, and Discoveries
Venkata Krishna Pillutla and Zhanpeng Fang, Carnegie Mellon University, USA; Pravallika Devineni, University of California, Riverside, USA; Christos Faloutsos, Carnegie Mellon University, USA; Danai Koutra, University of Michigan, USA; Jie Tang, Tsinghua University, China

11:20-11:35 Universal Dependency Analysis
Hoang Vu Nguyen, Max Planck Institute for Informatics, Germany; Panagiotis Mandros, Max Plank Institute, Germany; *Jilles Vreeken*, Max-Planck-Institut fuer Informatik, Germany

11:40-11:55 High Dimensional Structured Estimation with Noisy Designs
Amir Asiaee T., University of Minnesota, Duluth, USA; Soumyadeep Chaterjee, Yahoo! Inc., USA; Arindam Banerjee, University of Minnesota, USA

Saturday, May 7

CP18

Spatio-temporal Mining

10:00 AM-12:00 PM

Room:Chamber Boardroom

Chair: To Be Determined

10:00-10:15 Linear-Time Detection of Non-Linear Changes in Massively High Dimensional Time Series
Hoang Vu Nguyen and *Jilles Vreeken*, Max-Planck-Institut fuer Informatik, Germany

10:20-10:35 Learning Linear Dynamical Systems from Multivariate Time Series: A Matrix Factorization Based Framework
*Zitao Li*u and Milos Hauskrecht, University of Pittsburgh, USA

10:40-10:55 Spatio-Temporal Tensor Analysis for Whole-Brain Fmri Classification
Guixiang Ma, University of Illinois, Chicago, USA

11:00-11:15 Speeding Up All-Pairwise Dynamic Time Warping Matrix Calculation
Diego Silva and Gustavo Batista, University of Sao Paulo, Brazil

11:20-11:35 Joint Learning of Representation and Structure for Sparse Regression on Graphs
Chao Han, Shanshan Zhang, Mohamed Ghalwash, Slobodan Vucetic, and Zoran Obradovic, Temple University, USA

11:40-11:55 Infusing Geo-Recency Mixture Models for Effective Location Prediction in LBSN
Roland Assam, Subramanyam Sathyanarayana, and Thomas Seidl, RWTH-Aachen, Germany

Lunch Break

12:00 PM-1:30 PM

Attendees on their own

Saturday, May 7

Workshop 2 (half day): Machine Learning Methods for Recommender Systems (MLRec)

1:30 PM-5:00 PM

Room:Picasso

Coffee Break

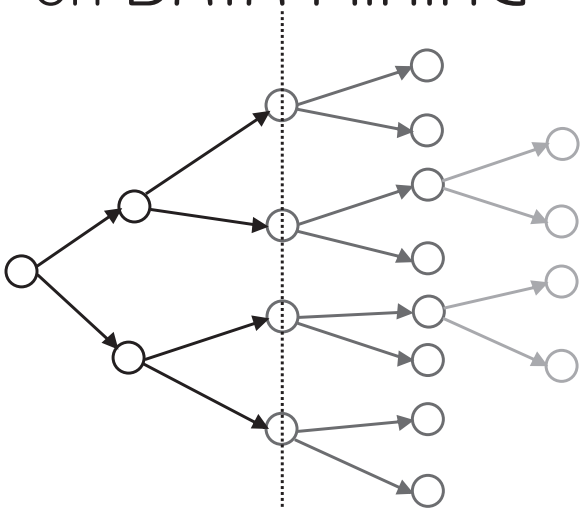
3:00 PM-3:30 PM

Room:Symphony Ballroom II

Notes

Abstracts

2016 SIAM
International Conference
on DATA MINING



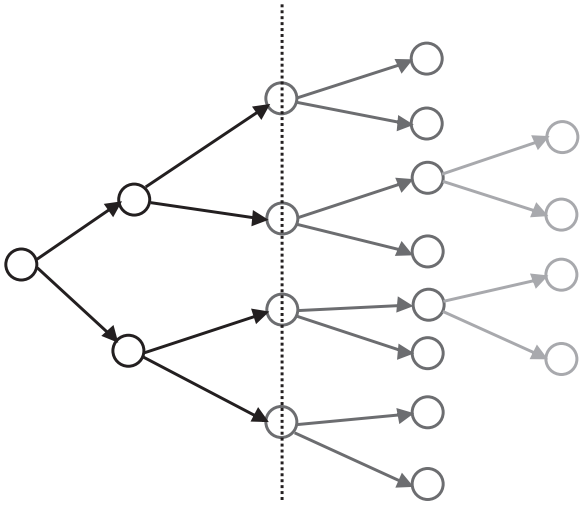
May 5-7, 2016
Hilton Miami Downtown
Miami, Florida, USA

Abstracts are printed as submitted by the authors.

Notes

Speaker and Organizer Index

2016 SIAM
International Conference
on DATA MINING



May 5-7, 2016
Hilton Miami Downtown
Miami, Florida, USA

Italicized names indicate session organizers

Italicized names indicate session organizers

SDM16 Budget

Conference Budget
SIAM International Conference on Data Mining
May 5-7, 2016
Miami, FL

Expected Paid Attendance		260	
Revenue			
Registration Income			$118,025
	Total		$118,025
Expenses			
Printing			$1,100.00
Organizing Committee			$2,900.00
Invited Speakers			$8,350.00
Food and Beverage			$33,300.00
AV Equipment and Telecommunication			$10,500.00
Advertising			$5,700.00
Proceedings			$12,282.00
Conference Labor (including benefits)			$46,288.00
Other (supplies, staff travel, freight, misc.)			$5,750.00
Administrative			$12,302.00
Accounting/Distribution & Shipping			$6,600.00
Information Systems			$12,233.00
Customer Service			$4,453.00
Marketing			$6,957.00
Office Space (Building)			$4,524.00
Other SIAM Services			$4,600.00
	Total		$177,839
Net Conference Expense			($59,814)
Support Provided by SIAM			$59,814
			$0
Estimated Support for Travel Awards not included above:			
Early Career / Students		17	$17,000

Hilton Miami Downtown