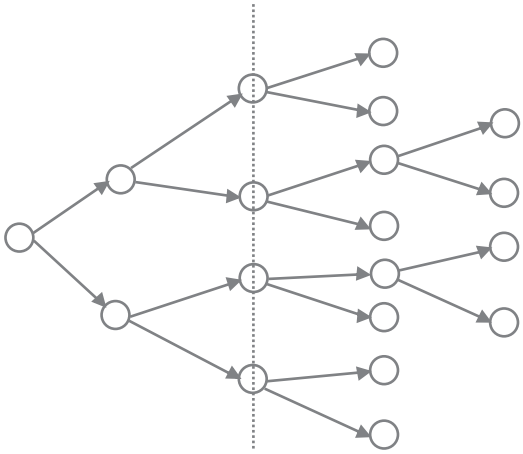
Final Program and Abstracts

2018 SIAM International Conference on DATA MINING



May 3-5, 2018

San Diego Marriott Mission Valley
San Diego, California, 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.

siam®

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 E-mail: meetings@siam.org
Conference Web: www.siam.org/meetings/sdm18
Membership and Customer Service:
(800) 447-7426 (USA & Canada) or +1-215-382-9800 (worldwide)
www.siam.org/meetings/sdm18

Table of Contents

Program-At-A-Glance… ..... Fold out section
General Information..... 2
Get-togethers..... 8
Invited Plenary Presentations..... 10
Tutorials 11
Workshops..... 13
Program Schedule..... 15
Poster Session .....21 & 26
Abstracts 29
Speaker and Organizer Index..... 55
Conference Budget ... Inside Back Cover
Hotel Floor Plan..... Back Cover

Organizing Committee Co-Chairs

Zoran Obradovic
Temple University, USA

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

Tanya Berger-Wolf
University of Illinois, USA

Dimitrios Gunopulos
University of Athens, Greece

Program Co-Chairs

Martin Ester
Simon Fraser University, Canada

Dino Pedreschi
University of Pisa, Italy

Workshops Co-Chairs

Joao Gama
University of Porto-LIAAD, Portugal

Jing Gao
SUNY Buffalo, USA

Tutorials Chair

Yan Liu
University of Southern California, USA

Doctoral Forum Chair

Julian McAuley
University of California, San Diego, USA

Sponsorship Co-Chairs

Matteo Riondato
TwoSigma, USA

Jiliang Tang
Michigan State University, USA

Publicity Co-Chairs

Zhenhui Li
Pennsylvania State University, USA

Gregor Stiglic
University of Maribor, Slovenia

Senior Program Committee

Charu Aggarwal
IBM, USA

Leman Akoglu
Carnegie Mellon University, USA

Francesco Bonchi
ISI Foundation, Turin, Italy

Toon Calders
Université Libre de Bruxelles, Belgium

Nitesh Chawla
Notre Dame, USA

Ian Davidson
University of California, Davis, USA

Carlotta Domeniconi
George Mason University, USA

Joao Gama
INESC TEC – LIAAD, Portugal

Aristides Gionis
Aalto University, Finland

Jiawei Han
University of Illinois at Urbana-Champaign, USA

Chandrika Kamath
Lawrence Livermore National Laboratory, USA

Pallika Kanani
Oracle Labs, USA

George Karypis
University of Minnesota, Minneapolis, USA

Xuemi Lin
University of New South Wales, Australia

Wagner Meira Jr.
Universidade Federal de Minas Gerais, Brazil

Katharina Morik
TU Dortmund, Germany

Mirco Nanni
ISTI-CNR Pisa, Italy

B. Aditya Prakash
Virginia Tech, USA

Manuel Gomez Rodriguez
Max-Planck Institute, Germany

Mohak Shah
Bosch, USA

Thomas Seidl
LMU München, Germany

Joerg Sander
University of Alberta, Canada

Kyuseok Shim
Seoul National University, South Korea

Masashi Sugiyama
The University of Tokyo, Japan

Jimeng Sun
Georgia Institute of Technology, USA

Yizhou Sun
University of California, Los Angeles

Evimari Terzi
Boston University, USA

Hanghang Tong
Arizona State University, USA

Ke Wang
Simon Fraser University, Canada

Wei Wang
University of California, Los Angeles, USA

Jia Wu
Macquarie University, Australia

Hui Xiong
Rutgers University, USA

Philip S Yu
University of Illinois at Chicago, USA

Jeffrey Xu Yu
Chinese University of Hong Kong, Hong Kong

Zhi-Hua Zhou
Nanjing University, China

Arthur Zimek
University of Southern Denmark, Denmark

Program Committee

Zubin Abraham
Robert Bosch, USA

David Anastasiu
SJSU, USA

Renato Assunção
UFMG, Brazil

Maurizio Atzori
University of Cagliari, Italy

Christian Bauckhage
Fraunhofer IAIS, Germany

Michele Berlingerio
IBM Research, Ireland

Michael Berthold
Universitat Konstanz, DE, Germany

Alex Beutel
Google, Inc., USA

Ricardo Campello
James Cook University, Australia

Andre Carvalho
USP, Brazil

Michelangelo Ceci
University of Bari, Italy

Loic Cerf
UFMG, Brazil

Hau Chan,
University of Nebraska-Lincoln, USA

Zhengzhang Chen
NEC Labs, USA

Liangzhe Chen
Virginia Tech, USA

Korris Chung
Hong Kong Polytechnic University, Hong Kong

Michele Coscia
Harvard Kennedy School, Cambridge, USA

Alfredo Cuzzocrea
University of Trieste and ICAR-CNR, Italy

Hasan Davulcu
Arizona State University, USA

Abir De
Indian Institute of Technology Kharagpur, India

Ke Deng
RMIT University, Australia

Sauptik Dhar
Robert Bosch LLC, USA

Josep Domingo-Ferrer
Universitat Rovira i Virgili, Catalonia, Spain

Yuxiao Dong
Microsoft Research, USA

Bo Du
Computer School, Wuhan University, China

Nan Du
Google Research, USA

Wouter Duivesteijn
TU Eindhoven, Netherlands

Jayanta Dutta
Robert Bosch LLC, USA

Saso Dzeroski
Jozef Stefan Institute, Ljubljana, Slovenia

Tapio Elomaa
Tampere University of Technology, Finland

Shobeir Fakhrei
University of Southern California, USA

YaJu Fan
Lawrence Livermore National Laboratory, USA

Elena Ferrari
University of Insubria, Varese, Italy

Carlos Ferreira
INESC TEC, Portugal

Lisa Friedland
Northeastern University, USA

Yanjie Fu
Missouri University of Science and Technology, USA

Yun Fu
Northeastern University, USA

Benjamin C. M. Fung
McGill University, Canada

Johannes Fürnkranz
TU Darmstadt, Germany

Byron Gao
Texas State University, USA

Jing Gao
University at Buffalo, USA

Kiran Garimella
Aalto University, Finland

Minos Garofalakis
Technical University of Crete, Greece

Mohamed Ghalwash
Temple University, USA

Aris Gkoulalas-Divanis
IBM Resarch, Cambridge, USA

Bart Goethals
University of Antwerp, Belgium

Marta Gonzalez
MIT, USA

Quanquan Gu
University of Virginia, USA

Francesco Gullo
UniCredit, R&D Dept., IT, Germany

Stephan Günnemann
Technical University of Munich, Germany

Sara Hajian
Eurecat – Technology Centre of Catalonia, Netherlands

Satoshi Hara
Osaka University, Japan

Marwan Hassani
Eindhoven University of Technology, Netherlands

Lifang He
University of Illinois at Chicago, USA

Jaakko Hollmén
Aalto University, Finland

Michael E. Houle
NII, Brazil

Xia Hu
TAMU, USA

Kejun Huang
University of Minnesota, USA

Szymon Jaroszewicz
Polish Academy of Science, Poland

Shuiwang Ji
Washington State University, USA

Junjun Jiang
National Institute of Informatics, Tokyo, Japan

Meng Jiang
University of Notre Dame, USA

Alipio Jorge
INESC, Portugal

Murat Kantarcioglu
UT Dallas, USA

Amin Karbasi
Yale University, USA

Przemyslaw Kazienko
Wrocław University of Technology, Poland

Xiangnan Kong
Worcester Polytechnic Institute, USA

Deguang Kong
Yahoo Research, USA

Danai Koutra
University of Michigan, USA

Hady Lauw
Singapore Management University, Singapore

Florian Lemmerich
University of Wurzburg, Germany

Bruno Lepri
FBK, Trento, Italy

Jiuyong Li
University of South Australia, Australia

Ming Li
Nanjing University, China

Shuai Li
University of Cambridge, United Kingdom

Liangyue Li
Arizona State University, USA

Zhenhui (Jessie) Li
The Pennsylvania State University, USA

Hongwei Liang
Simon Fraser University, Canada

Jefrey Lijffijt
Ghent University, Belgium

Ee-peng Lim
Singapore Management University, Singapore

Shuyang Lin
Facebook Inc., USA

Jessica Lin
George Mason University, USA

Guannan Liu
Beihang University, China

Bin Liu
IBM Research, USA

Mingsheng Long
Tsinghua University, China

Jiayi Ma
Wuhan University, China

Giuseppe Manco
ICAR-CNR, IT, Italy

Julian McAuley
UCSD, USA

Ryan McBride
Simon Fraser University, Canada

Rosa Meo
University of Torino, Italy

Luiz Merschmann
UFLA, Brazil

Pauli Miettinen
Max Planck Institute for Informatics, Germany

Anna Monreale
University of Pisa, Italy

Joao Moreira
INESC TEC, Portugal

Abdullah Mueen
University of New Mexico, USA

Emmanuel Müller
Hasso Plattner Institute, Germany

Azad Naik
Microsoft, USA

Amedeo Napoli
LORIA, Nancy, France

Franco Maria Nardini
ISTI-CNR, Italy

Athanasios Nikolakopoulos
University of Minnesota, USA

Xia Ning
Indiana University - Purdue University Indianapolis, USA

Eirini Ntoutsi
Leibniz Universität Hannover, Germany

Pedro O.S. Vaz de Melo
Universidade Federal de Minas Gerais, Brazil

Eduardo Ogasawara
CEFET/RJ, Brazil

Nikunj Oza
NASA Ames, USA

Evangelos Papalexakis
UC Riverside, USA

Panagiotis Papapetrou
Stockholm University, Sweden

Mykola Pechenizkiy
TU Eindhoven, Netherlands

Ruggero Pensa
University of Torino, Italy

Bryan Perozzi
Google, Inc., USA

Claudia Plant
University of Vienna, Austria

Marc Plantevit
Université de Lyon, France

Kai Puolamäki
Aalto University, Finland

Milos Radovanovic
U. Novi Sad, Serbia

Chedy Raissi
Inria, FR, France

Jan Ramon
Inria, FR, France

Huzefa Rangwala
George Mason University, USA

Chandan Reddy
Virginia Tech, USA

Celine Robardet
INSA Lyon, France

Alejandro Rodriguez
University of Madrid, Spain

Yu Rong
Tencent AI Lab, China

Polina Rozenshtein
Aalto University, Finland

Salvatore Ruggieri
University of Pisa, Italy

Ahmet Erdem Sariyuce
University at Buffalo, USA

Saket Sathe
IBM Thomas J. Watson, USA

Yücel Saygin
Sabanci University, Turkey

Matthias Schubert
Ludwig-Maximilians-Universität München, Germany

Erich Schubert
University of Heidelberg, Germany

Weixiang Shao
Google, Inc., USA

Chuan Shi
Beijing University of Posts and Telecommunications, China

Anshumali Shrivastava
Rice University, USA

Parkishit Sondhi
Walmart Labs, USA

Dongjin Song
NEC Labs America, USA

Arnaud Soulet
University François Rabelais of Tours, France

Myra Spiliopoulou
Otto-von-Guericke-University Magdeburg, Germany

Giovanni Stilo
Sapienza University, Italy

Einoshin Suzuki
Kyushu University, Japan

Andrea Tagarelli
DIMES, University of Calabria, IT, Italy

Pang-Ning Tan
Michigan State University, USA

Nikolaj Tatti
Aalto University, Finland

Yannis Theodoridis
University of Piraeus, Greece

Vicenç Torra
University of Skövde, Sweden

Panayiotis Tsaparas
University of Ioannina, Greece

Ranga Vatsavai
North Carolina State University, USA

Adriano Veloso
UFMG, Brazil

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

Slobodan Vucetic
Temple University, USA

Anil Vullikanti
Virginia Tech, USA

Monica Wachowicz
University of New Brunswick, Canada

Can Wang
Zhejiang University, China

Beidou Wang
Simon Fraser University, Canada

Jiannan Wang
Simon Fraser University, Canada

Hao Wang
Qihu 360 Inc., China

Jianyong Wang
Tsinghua University, China

Haishuai Wang
Harvard University, USA

Hui (Wendy) Wang
Stevens Institute of Technology, USA

Takashi Washio
The Institute of Scientits, USA

Ingmar Weber
QCRI, Qatar

Xiaokai Wei
Facebook Inc., USA

Xintao Wu
University of Arkansas, USA

Keli Xiao
Stony Brook University, USA

Yusheng Xie
Baidu Inc., USA

Sihong Xie
Lehigh University, USA

Tong Xu
University of Science and Technology of China, China

Yuan Yao
Nanjing University, China

Guoxian Yu
Southwest University, China

Carlo Zaniolo
UCLA, USA

De-Chuan Zhan
Nanjing University, China

Wei Zhang
Macquarie University, Australia

Min-Ling Zhang
School of Computer Science and Engineering, Southeast University, China

Xiangliang Zhang
King Abdullah University of Science and Technology, Saudi Arabia

Xiang Zhang
The Pennsylvania State University, USA

Jiawei Zhang
University of Illinois at Chicago, USA

Chao Zhang
University of Illinois, Urbana Champaign, USA

Yao Zhang
Virginia Tech, USA

Yu Zheng
Microsoft Research Asia, China

Chuan Zhou
Institute of Information Engineering,Chinese Academy of Sciences, China

Jiayu Zhou
Michigan State University, USA

Tianqing Zhu
Deakin University, Australia

Chen Zhu
Baidu Inc., China

Hengshu Zhu
Baidu Inc., China

Xingquan Zhu
Florida Atlantic University, USA

Albrecht Zimmermann
Université de Caen Normandie, France

Conference Themes

Methods and Algorithms

- Classification
- Clustering
- Frequent Pattern Mining
- Probabilistic & Statistical Methods
- Graphical Models
- Spatial & Temporal Mining
- Data Stream Mining
- Anomaly & Outlier Detection
- Feature Extraction, Selection and Dimension Reduction
- Mining with Constraints
- Data Cleaning & Preprocessing
- Computational Learning Theory
- Multi-Task Learning
- Online Algorithms
- Big Data, Scalable & High-Performance Computing Techniques
- Mining with Data Clouds
- Mining Graphs
- Mining Semi Structured Data
- Mining Image Data
- Mining on Emerging Architectures
- Text & Web Mining
- Optimization Methods
- Other Novel Methods

Applications

- Astronomy & Astrophysics
- High Energy Physics
- Recommender Systems
- Climate / Ecological / Environmental Science
- Risk Management
- Supply Chain Management
- Customer Relationship Management
- Finance
- Genomics & Bioinformatics
- Drug Discovery
- Healthcare Management
- Automation & Process Control
- Logistics Management
- Intrusion & Fraud detection
- Bio-surveillance
- Sensor Network Applications
- Social Network Analysis
- Intelligence Analysis
- Other Novel Applications & Case Studies

Human Factors and Social Issues

- Ethics of Data Mining
- Intellectual Ownership
- Privacy Models
- Privacy Preserving Data Mining & Data Publishing
- Risk Analysis
- User Interfaces
- Interestingness & Relevance
- Data & Result Visualization
- Other Human Factors and Social Issues

SIAM Registration Desk

The SIAM registration desk is located in the Rio Vista Ballroom Foyer. It is open during the following hours:

Wednesday, May 2

5:00 PM – 7:00 PM

Thursday, May 3

7:00 AM – 7:30 PM

Friday, May 4

7:30 AM – 4:00 PM

Saturday, May 5

7:30 AM – 4:00 PM

Hotel Address

San Diego Marriott Mission Valley

8757 Rio San Diego Drive

San Diego, California 92108

USA

Phone Number: +1 (619) 692-3800

Hotel web address:

http://www.marriott.com/hotels/travel/sanmv-san-diego-marriott-mission-valley/

Hotel Telephone Number

To reach an attendee or leave a message, call +1 (619) 692-3800. 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 4:00 PM.

Check-out time is 11:00 AM.

Child Care

The San Diego Marriott Mission Valley recommends *www.care.com* for attendees interested in child care services. Attendees are responsible for making their own child care arrangements.

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

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

Exxon Mobil

Hewlett-Packard

Huawei FRC French R&D Center

IBM Corporation

IDA Center for Communications Research, La Jolla

IDA Center for Communications Research, Princeton

IFP Energies nouvelles

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

List current March 2018.

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 $140 on full registration for the 2018 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 ($140 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:

(800) 447-7426 (USA & Canada) or
+1-215-382-9800 (worldwide)
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.

A data (LCD) projector and screen will be provided in all technical session meeting rooms. The data projectors support both VGA and HDMI connections. Presenters requiring an alternate connection must provide their own adaptor.

Internet Access

Attendees booked within the SIAM room block will receive complimentary wireless Internet access in their guest rooms of the hotel. All conference attendees will have complimentary wireless Internet access in the meeting space of the hotel.

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

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 3, 7:00 PM – 9:00 PM and Friday, May 4, 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, May 3. Papers presented on Friday will have their poster slots during the Doctoral Forum and Poster Session on Friday, May 4. Boards and push pins will be available to Thursday's presenters at 7:00 AM on Thursday, May 3 and at 9:00 AM on Friday, May 4 for Friday's presenters.

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 Display

Cambridge University Press

Conference Sponsors

Disney Research

Two Sigma

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*).

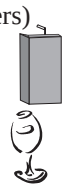
Get-togethers

- Welcome Reception and Poster Session

  Thursday, May 3

  7:00 PM – 9:00 PM

- Business Meeting (open to SIAG/DMA members)

  Friday, May 4

  7:00 PM – 9:00 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 #SIAMSDM18.

SIAM's Twitter handle is @TheSIAMNews.

Changes to the Printed Program

The printed program and abstracts were current at the time of printing, however, please review the online program schedule (*http://meetings.siam.org/program.cfm?CONFCODE=DT18*) for the most up-to-date information.

Invited Plenary Speakers

**** All Invited Plenary Presentations will take place in Salon A-D****

Thursday, May 3

8:15 AM - 9:30 AM

IP1 Title Not Available at Time of Publication

Bernhard Schölkopf, *Max Planck Institute for Intelligent Systems, Germany and Amazon, Germany*

1:15 PM - 2:30 PM

IP2 Learning to Rank Results Optimally in Search and Recommendation

Charles Elkan, *University of California, San Diego USA*

Friday, May 4

8:15 AM - 9:30 AM

IP3 Towards a Structure/Function simulation of a Cancer Cell

Trey G. Ideker, *University of California, San Diego USA*

Saturday, May 5

8:00 AM - 9:15 AM

IP4 From Robots to Biomolecules: Computing Meets the Physical World

Lydia Kavraki, *Rice University, USA*

Tutorials

**** All Tutorials will take place in Sierra 5-6****

Thursday, May 3

10:00 AM - 12:00 PM

TS1: Tutorial Session: A Critical Review of Online Social Data: Biases, Methodological: Pitfalls, and Ethical Boundaries

2:45 PM - 4:45 PM

TS2: Tutorial Session: Data Mining Critical Infrastructure Systems - Models and Tools

5:00 PM - 7:00 PM

TS3: **Tutorial Session:** Knowledge Discovery from Temporal Social Networks

Friday, May 4

10:00 AM - 12:00 PM

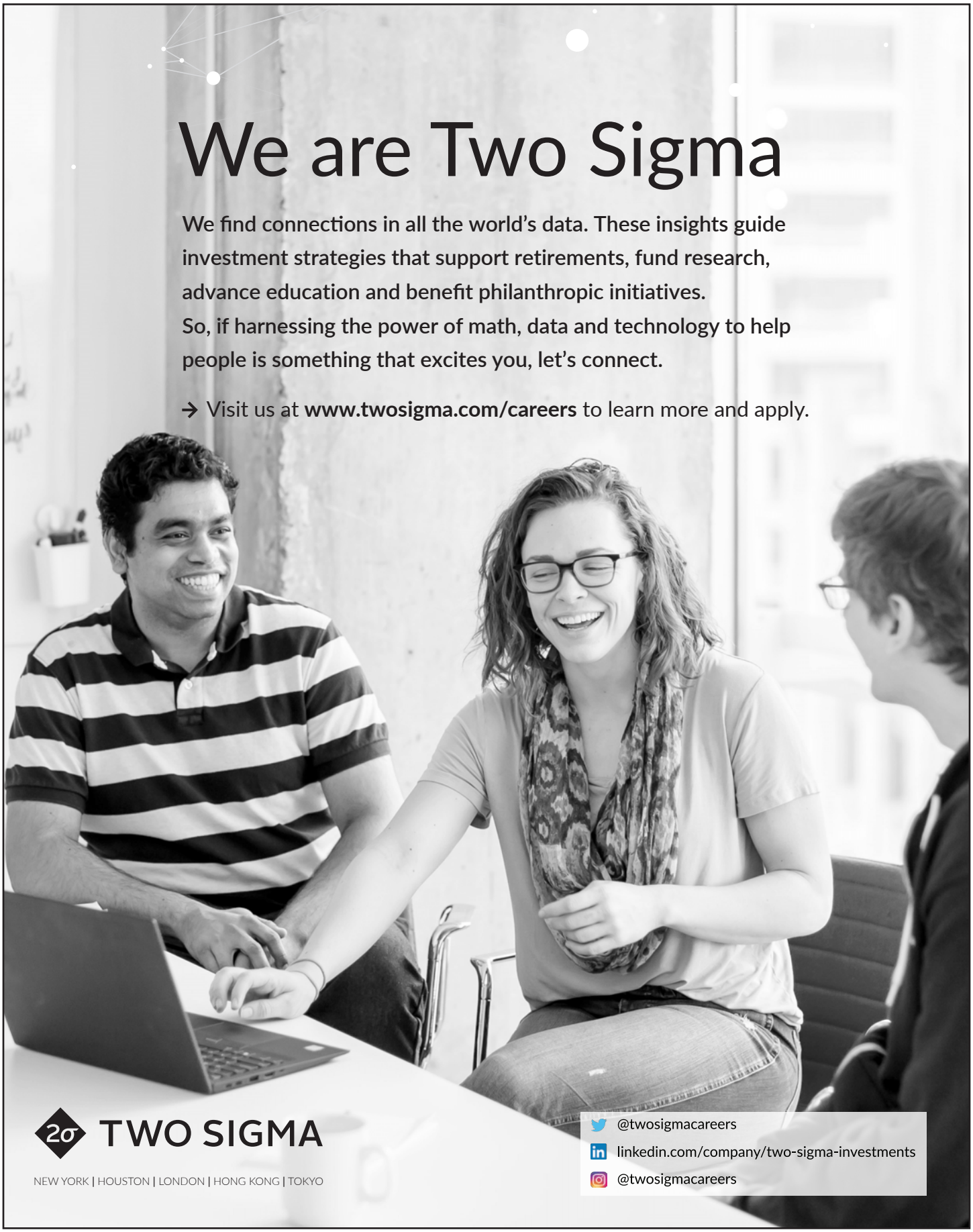
TS4: Tutorial Session: Problems with Partially Observed (Incomplete) Networks: Biases, Skewed Results, and Solutions

1:15 PM - 3:15 PM

TS5: Tutorial Session: The Canonical Polyadic Tensor: Decomposition and Variants for Mining Multi-Dimensional Data

3:30 PM - 5:10 PM

TS5: Tutorial Session, continued: The Canonical Polyadic Tensor: Decomposition and Variants for Mining Multi-Dimensional Data


We are Two Sigma
We find connections in all the world's data. These insights guide investment strategies that support retirements, fund research, advance education and benefit philanthropic initiatives.
So, if harnessing the power of math, data and technology to help people is something that excites you, let's connect.
→ Visit us at www.twosigma.com/careers to learn more and apply.
TWO SIGMA
NEW YORK | HOUSTON | LONDON | HONG KONG | TOKYO
@twosigmacareers
linkedin.com/company/two-sigma-investments
@twosigmacareers

Workshops

Saturday, May 5 Workshops

9:30 AM – 5:45 PM

Workshop 1: Big Traffic Data Analytics
Sierra 6

Workshop 2: Artificial Intelligence in Insurance
Sierra 5

Workshop 3: Machine Learning Methods for Recommender Systems
Salon A-C

Workshop 4: Cost-Sensitive Learning
Balboa 1-2

Workshop 5: Data Mining for Geophysics and Geology
Cabrillo Salon 1

Workshop 6: Data Mining for Medicine and Healthcare
Cabrillo Salon 2

SIAM Activity Group on
Data Mining & Analytics (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 meetings**
- **Biennial conference**

BENEFITS OF SIAG/DMA MEMBERSHIP:

- **Listing in the SIAG's online membership directory**
- **Additional $15 discount on registration at the SIAM Conference on Data Mining & Analytics**
- **Electronic communications 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 2 activity groups for free!**

2017-18 SIAG/DMA OFFICERS

Chair:	Ali Pinar, Sandia National Laboratories
Vice Chair:	Hans De Sterck, Monash University
Program Director:	Danai Koutra, University of Michigan
Secretary:	Jiliang Tang, Arizona State University

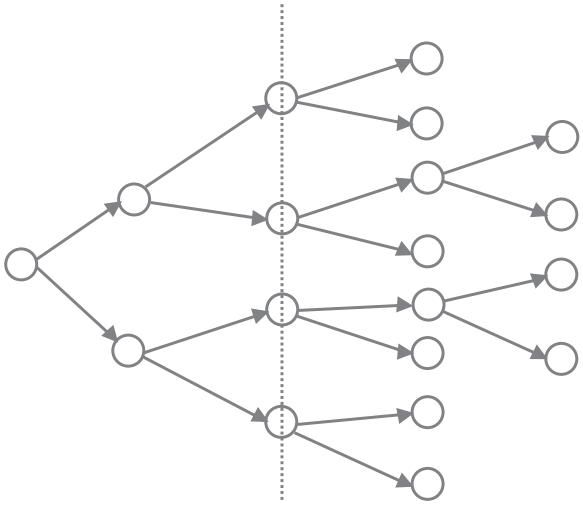
TO JOIN:

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

SIAM: ***www.siam.org/joinsiam***

Program Schedule

2018 SIAM International Conference on DATA MINING

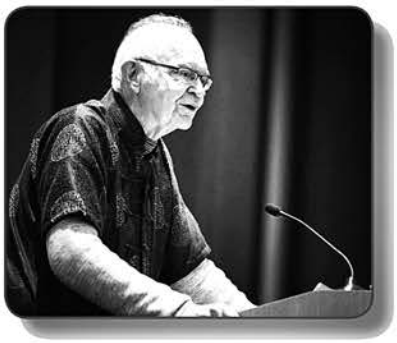


May 3-5, 2018

San Diego Marriott Mission Valley
San Diego, California, USA

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

6/17

Wednesday, May 2

Registration
5:00 PM-7:00 PM

Room:Rio Vista Ballroom Foyer

Thursday, May 3

Registration
7:00 AM-7:30 PM

Room:Rio Vista Ballroom Foyer

Continental Breakfast
7:30 AM-8:00 AM

Room:Salon E-H

Announcements
8:00 AM-8:15 AM

Room:Salon A-D

Thursday, May 3

IP1
Title To Be Announced
8:15 AM-9:30 AM

Room:Salon A-D

Chair: Dino Pedreschi, Universita' di Pisa, Italy

Abstract Not Available At Time Of Publication

Bernhard Schölkopf

Max Planck Institute for Intelligent Systems, Germany and Amazon, Germany

Coffee Break
9:30 AM-10:00 AM

Room:Salon E-H

Thursday, May 3

TS1

Tutorial Session: A Critical Review of Online Social Data: Biases, Methodological: Pitfalls, and Ethical Boundaries

10:00 AM-12:00 PM

Room:Sierra 5-6

Chair: Yan Liu, University of Southern California, USA

To set the context for the issues we review, we first discuss why (purposes and applications), and how (prototypical data processing and analysis pipeline) social data is used, and provide examples of typical limitations, tradeoffs or mistakes.

Alexandra Olteanu
BM Research, USA

Emre Kiciman
Microsoft Research, USA

Carlos Castillo
Universitat Pompeu Fabra, Spain

Thursday, May 3

CP1

Clustering

10:00 AM-12:00 PM

Room:Salon A-C

Chair: Fosca Giannotti, Institute of Information Science and Technologies (ISTI), Italy

10:00-10:15 Many-to-Many Correspondences Between Partitions: Introducing a Cut-Based Approach

Roland Glantz, Karlsruhe Institute of Technology, Germany; *Henning Meyerhenke*, University of Cologne, Germany

10:20-10:35 Graph Sketching-Based Space-Efficient Data Clustering

Anne Morvan, CEA and Université Paris-Dauphine, France; Krzysztof Choromanski, Google, Inc., USA; Cédric Gouy-Pailler, CEA, France; Jamal Atif, Université Paris Dauphine, France

10:40-10:55 Image Constrained Blockmodelling: A Constraint Programming Approach

Mohadeseh Ganji, University of Melbourne, Australia; *Jeffrey Chan*, RMIT University, Australia; Peter. J Stuckey, James Bailey, Christopher Leckie, and Kotagiri Ramamohanarao, University of Melbourne, Australia; Ian Davidson, University of California, Davis, USA

11:00-11:15 NLRR++: Scalable Subspace Clustering Via Non-Convex Block Coordinate Descent

Jun Wang, Harbin Institute of Technology, China; Cho-Jui Hsieh, University of California, Davis, USA; Daming Shi, Harbin Institute of Technology, China

11:20-11:35 Unsupervised Neural Categorization for Scientific Publications

Keqian Li, Hanwen Zha, Yu Su, and Xifeng Yan, University of California, Santa Barbara, USA

11:40-11:55 Mixtures of Block Models for Brain Networks

Zilong Bai, University of California, Davis, USA; Peter Walker, Naval Medical Research Center, USA; Ian Davidson, University of California, Davis, USA

Thursday, May 3

CP2

Sequence Data

10:00 AM-12:00 PM

Room:Salon D

Chair: - Jessie Li, Pennsylvania State University, USA

10:00-10:15 Causal Inference on Event Sequences

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

10:20-10:35 Outlier Detection over Distributed Trajectory Streams

Jiali Mao, Pengda Sun, Cheqing Jin, and Aoying Zhou, East China Normal University, China

10:40-10:55 Jump: A Fast Deterministic Algorithm to Find the Closest Pair of Subsequences

Xingyu Cai, Shanglin Zhou, and Sanguthevar Rajasekaran, University of Connecticut, USA

11:00-11:15 Streaming Tensor Factorization for Infinite Data Sources

Shaden Smith and Kejun Huang, University of Minnesota, USA; Nicholas Sidiropoulos, University of Virginia, USA; George Karypis, University of Minnesota and Army HPC Research Center, USA

11:20-11:35 Mining Top-K Quantile-Based Cohesive Sequential Patterns

Len Feremans, Boris Cule, and Bart Goethals, University of Antwerp, Belgium

11:40-11:55 Staple: Spatio-Temporal Precursor Learning for Event Forecasting

Yue Ning, Virginia Tech, USA

Lunch Break

12:00 PM-1:15 PM

Attendees on their own

Thursday, May 3

IP2

Learning to Rank Results Optimally in Search and Recommendation

1:15 PM - 2:30 PM

Room: Salon A-D

Chair: Martin Ester, Simon Fraser University, Canada

Consider the scenario where an algorithm is given a context, and then it must select a slate of results to display. For example, the context may be a search query, an advertising slot, or an opportunity to show recommendations. We want to compare many alternative ranking functions that select results in different ways. However, online A/B testing with traffic from actual users is expensive. This research provides a method to use traffic that was exposed to a past ranking function to obtain an estimate of the utility of a hypothetical new ranking function. The method is a purely offline computation, and relies on just one assumption that is quite reasonable. We show further how to design a ranking function that is the best possible, given the same assumption. Learning an optimal function for a search engine to apply to rank results given a query is a special case. Experimental results on data logged by a real-world e-commerce web site are positive.

Charles Elkan
University of California, San Diego, USA

Coffee Break

2:30 PM-2:45 PM

Room:Salon E-H

Thursday, May 3

TS2

Tutorial Session: Data Mining Critical Infrastructure Systems - Models and Tools

2:45 PM-4:45 PM

Room:Sierra 5-6

Chair: Yan Liu, University of Southern California, USA

Critical Infrastructure Systems such as transportation, water and power grid systems are vital to our national security, economy, and public safety. These different infrastructures are also interdependent on each other in such a complicated way that failures in one may lead to failures in the others. Recent events, like Hurricanes Harvey and Irma, show how the interdependencies among different CI networks lead to catastrophic failures among the whole system. Hence, analyzing these CI networks becomes a very important problem. Can we understand how different CISs are related to each? Moreover, given a natural disaster like a hurricane or an earthquake, can we analyze and predict their impact on the CISs? How can we help policy makers in such situations?

Different types of approaches (empirical, agent based, system-dynamics based, etc.) have been proposed. In this tutorial, we will cover recent and state-of-the-art research on interesting CIS problems both generally and in specific popular systems (power and transportation). We will also cover tools that support decision making during events. The tutorial will be in 2 hours, with a 5 min break. This tutorial has not been presented before in other venues.

Liangzhe Chen
Virginia Tech, USA

Aditya Prakash
Virginia Tech, USA

Thursday, May 3

CP3

Networks 1

2:45 PM-4:45 PM

Room:Salon A-C

Chair: Shobeir Fakhraei, University of Southern California, USA

2:45-3:00 Near-Optimal Mapping of Network States Using Probes

Bijaya Adhikari, Pavan Rangudu, B. Aditya Prakash, and Anil Vullikanti, Virginia Tech, USA

3:05-3:20 Network Inference from Contrastive Groups Using Discriminative Structural Regularization

Ruihua Cheng and Zhi Wei, New Jersey Institute of Technology, USA; Kai Zhang, Temple University, USA

3:25-3:40 Group Centrality Maximization Via Network Design

Sourav Medya, Arlei Silva, and Ambuj Singh, University of California, Santa Barbara, USA; Prithwish Basu, Raytheon Systems, USA; Ananthram Swami, Army Research Laboratory, USA

3:45-4:00 Robust Road Map Inference Through Network Alignment of Trajectories

Sanjay Chawla, University of Sydney, Australia; *Sofiane Abbar*, Rade Stanojevic, Saravanan Thirumuruganathan, Fethi Filali, and Ahid Aleimat, Qatar Computing Research Institute, Qatar

4:05-4:20 AspEm: Embedding Learning by Aspects in Heterogeneous Information Networks

Yu Shi, University of Illinois, Urbana-Champaign, USA; Huan Gui, Facebook, USA; Qi Zhu, University of Illinois, Urbana-Champaign, USA; Lance Kaplan, U.S. Army Research Laboratory, USA; Jiawei Han, University of Illinois at Urbana-Champaign, USA

4:25-4:40 Semi-Supervised Embedding in Attributed Networks with Outliers

Jiongqian Liang, Peter Jacobs, Jiankai Sun, and Srinivasan Parthasarathy, Ohio State University, USA

Thursday, May 3

CP4

Social Media

2:45 PM-4:45 PM

Room:Salon D

Chair: Hongwei Liang, Simon Fraser University, Canada

2:45-3:00 Online Truth Discovery on Time Series Data

Liuyi Yao and Lu Su, State University of New York, Buffalo, USA; Qi Li, University of Illinois, Urbana-Champaign, USA; Yaliang Li, Baidu Research Big Data Lab, China; Fenglong Ma, Jing Gao, and Aidong Zhang, State University of New York, Buffalo, USA

3:05-3:20 Modeling the Interaction Coupling of Multi-View Spatiotemporal Contexts for Destination Prediction

Kunpeng Liu and Pengyang Wang, Missouri University of Science and Technology, USA; Jiawei Zhang, Florida State University, USA; Guannan Liu, Beihang University, China; *Yanjie Fu* and Sajal Das, Missouri University of Science and Technology, USA

3:25-3:40 Who Will Attend This Event Together? Event Attendance Prediction Via Deep Lstm Networks

Xian Wu, Yuxiao Dong, and Baoxu SHI, University of Notre Dame, USA; Ananthram Swami, Army Research Laboratory, USA; Nitesh Chawla, University of Notre Dame, USA

3:45-4:00 You Are How You Move: Linking Multiple User Identities From Massive Mobility Traces

Huandong Wang and Yong Li, Tsinghua University, P. R. China; Gang Wang, Virginia Tech, USA; Depeng Jin, Tsinghua University, P. R. China

4:05-4:20 Click Versus Share: A Feature-Driven Study of Micro-Video Popularity and Virality in Social Media

Jingtao Ding, Yanghao Li, Yong Li, and Depeng Jin, Tsinghua University, P. R. China

4:25-4:40 A Probabilistic Hough Transform for Opportunistic Crowd-Sensing of Moving Traffic Obstacles

Michiaki Tatsubori, IBM Research - Tokyo, Japan; Aisha Walcott-Bryant and Reginald Bryant, IBM Research, Kenya; John Wamburu, University of Massachusetts, Amherst, USA

Thursday, May 3

Organizational Break

4:45 PM-5:00 PM

TS3

Tutorial Session: Knowledge Discovery from Temporal Social Networks

5:00 PM-7:00 PM

Room:Sierra 5-6

Chair: Yan Liu, University of Southern California, USA

Data is structured in the form of networks. And now? How to analyze them? Extracting knowledge of network data is not a simple task and requires the use of appropriate tools and techniques, especially in scenarios that take into account the volume and evolving aspects of the network. There is a vast literature on how to collect, process, and model social media data in the form of networks, as well as key metrics of centrality. However, there is still much to be discussed in relation to the analysis of the underlying network. In this tutorial we consider that data has already been collected and is already structured as a network. The goal is to discuss techniques to analyze network data, especially considering time perspective. First, concepts related to problem definition, temporal networks and metrics for network analysis will be presented. Next, in a more practical aspect will be shown techniques of visualization and processing of temporal networks. In the end, applications with real data will be discussed, illustrating how network data knowledge extraction works from start to finish.

Fabíola Pereira
Federal University of Uberlandia, Brazil

Joao Gama
University of Porto, Portugal

Thursday, May 3

CP5

Time Series Data 1

5:00 PM-6:40 PM

Room:Salon A-C

Chair: - Abdullah Mueen, University of New Mexico, USA

5:00-5:15 Interpretable Categorization of Heterogeneous Time Series Data

Ritchie Lee, Carnegie Mellon University, USA; Mykel Kochenderfer, Stanford University, USA; Ole Mengshoel, Carnegie Mellon University, USA; Joshua Silbermann, Johns Hopkins University, USA

5:20-5:35 Efficient Search of the Best Warping Window for Dynamic Time Warping

Chang Wei Tan and Matthieu Herrmann, Monash University, Australia; Germain Forestier, University of Haute Alsace, France; Geoff Webb and Francois Petitjean, Monash University, Australia

5:40-5:55 Accelerating Time Series Searching with Large Uniform Scaling

Yilin Shen and Yanping Chen, Samsung Research America, USA; Eamonn Keogh, University of California, Riverside, USA; Hongxia Jin, Samsung Research America, USA

6:00-6:15 Evolving Separating References for Time Series Classification

Xiaosheng Li and Jessica Lin, George Mason University, USA

6:20-6:35 Classifying Multivariate Time Series by Learning Sequence-Level Discriminative Patterns

Guruprasad Nayak, University of Minnesota, USA; Varun Mithal, LinkedIn, USA; Xiaowei Jia and Vipin Kumar, University of Minnesota, USA

Thursday, May 3

CP6

Health Informatics

5:00 PM-6:40 PM

Room:Salon D

Chair: Dino Pedreschi, University of Pisa, Italy

5:00-5:15 Health-Atm: A Deep Architecture for Multifaceted Patient Health Record Representation and Risk Prediction

Tengfei Ma and Cao Xiao, IBM Research, USA; Fei Wang, Cornell University, USA

5:20-5:35 Uncorrelated Patient Similarity Learning

Mengdi Huai, Chenglin Miao, and Qiuling Suo, State University of New York, Buffalo, USA; Yaliang Li, Baidu Research Big Data Lab, China; Jing Gao and Aidong Zhang, State University of New York, Buffalo, USA

5:40-5:55 Eeg-Based Motion Intention Recognition Via Multi-Task Rnns

Weitong Chen, University of Queensland, Australia; Sen Wang, Griffith University, Brisbane, Australia; Xiang Zhang and Lina Yao, University of New South Wales, Australia; Lin Yue, Northeast Normal University, People's Republic of China; Buyue Qian, Xi'an Jiaotong University, P.R. China; Xue Li, University of Queensland, Australia

6:00-6:15 Multi-Task Learning Based Survival Analysis for Predicting Alzheimer's Disease Progression with Multi-Source Block-Wise Missing Data

Yan Li, University of Michigan, USA; Tao Yang, Arizona State University, USA; Jiayu Zhou, Michigan State University, USA; Jieping Ye, University of Michigan, Ann Arbor, USA

6:20-6:35 Deep Attention Model for Triage of Emergency Department Patients

Djordje Gligorijevic, Jelena Stojanovic, Wayne Satz, Ivan Stojkovic, Kraftin Schreyer, Daniel Del Portal, and Zoran Obradovic, Temple University, USA

Thursday, May 3

Welcome Reception and Poster Session Part I

7:00 PM-9:00 PM

Room:Rio Vista Pavilion

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

Friday, May 4

Registration

7:30 AM-4:00 PM

Room:Rio Vista Ballroom Foyer

Continental Breakfast

7:30 AM-8:00 AM

Room:Salon E-H

Announcements

8:00 AM-8:15 AM

Room:Salon A-D

Friday, May 4

IP3

Towards a Structure/ Function Simulation of a Cancer Cell

8:15 AM - 9:30 AM

Chair: Tanya Y. Berger-Wolf, University of Illinois, Chicago, USA

Recently we and other laboratories have launched the Cancer Cell Map Initiative (ccmi.org) and have been building momentum. The goal of the CCMI is to produce a complete map of the gene and protein wiring diagram of a cancer cell. We and others believe this map, currently missing, will be a critical component of any future system to decode a patient's cancer genome. I will describe efforts along several lines: 1. Coalition building. We have made notable progress in building a coalition of institutions to generate the data, as well as to develop the computational methodology required to build and use the maps. 2. Development of technology for mapping gene-gene interactions rapidly using the CRISPR system. 3. Causal network maps connecting DNA mutations (somatic and germline, coding and noncoding) to the cancer events they induce downstream. 4. Development of software and database technology to visualize and store cancer cell maps. 5. A machine learning system for integrating the above data to create multi-scale models of cancer cells. In a recent paper by Ma et al., we have shown how a hierarchical map of cell structure can be embedded with a deep neural network, so that the model is able to accurately simulate the effect of mutations in genotype on the cellular phenotype.

Trey G. Ideker

University of California, San Diego, USA

Coffee Break

9:30 AM-10:00 AM

Room:Salon E-H

Friday, May 4

TS4

Tutorial Session: Problems with Partially Observed (Incomplete) Networks: Biases, Skewed Results, and Solutions

10:00 AM-12:00 PM

Room:Sierra 5-6

Chair: Yan Liu, University of Southern California, USA

Networked representations of physical and social phenomena are ubiquitous. Examples include social and information networks, technological and communication networks, co-purchasing networks, etc. These networks are often incomplete because the phenomena are partially observed. Working with incomplete networks can skew analyses. Acquiring the full data is often unrealistic (e.g., obtaining the Twitter Firehose is not viable), but one may be able to collect data selectively to enrich the incomplete network. With a limited query budget, which parts of a partially observed network should be examined to give the best (i.e., most complete) view of the entire network? 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 one's purposes. How should one best supplement this sampled data? This tutorial addresses the above questions. In particular, it we will focus on multi-armed bandit and reinforcement learning solutions.

Tina Eliassi-Rad

Northeastern University, USA

Sucheta Soundarajan

Syracuse University, USA

Sahely Bhadra

Indian Institute of Information Technology, India

Friday, May 4

CP7

Graphs

10:00 AM-12:00 PM

Room:Salon A-C

Chair: B. Aditya Prakash, Virginia Tech, USA

10:00-10:15 Learning Graph Representation Via Frequent Subgraphs

Dang Nguyen, Wei Luo, Tu Nguyen, Svetha Venkatesh, and Dinh Phung, Deakin University, Australia

10:20-10:35 On2Vec: Embedding-Based Relation Prediction for Ontology Population

Muhao Chen, University of California, Los Angeles, USA

10:40-10:55 On Spectral Graph Embedding: A Non-Backtracking Perspective and Graph Approximation

Fei Jiang, Peking University, China; Lifang He, South China University of Technology, China; Yi Zheng, Peking University, China; Enqiang Zhu, Guangzhou University, China; Jin Xu, Peking University, China; Philip Yu, University of Illinois, Chicago, USA

11:00-11:15 A Family of Tractable Graph Distances

Stratis Ioannidis, Northeastern University, USA; Jose Bento, Boston College, USA

11:20-11:35 Fast Flow-Based Random Walk with Restart in a Multi-Query Setting

Yujun Yan, Mark Heimann, Di Jin, and Danai Koutra, University of Michigan, USA

11:40-11:55 Ensemble-Spotting: Ranking Urban Vibrancy Via Poi Embedding with Multi-View Spatial Graphs

Pengyang Wang, Missouri University of Science and Technology, USA; Jiawei Zhang, Florida State University, USA; Guannan Liu, Beihang University, China; *Yanjie Fu*, Missouri University of Science and Technology, USA; Charu C. Aggarwal, IBM T.J. Watson Research Center, USA

Friday, May 4

CP8

Matrix Factorization and Topic Models

10:00 AM-12:00 PM

Room:Salon D

Chair: Lifang He, Cornell University, USA

10:00-10:15 Latitude: A Model for Mixed Linear–Tropical Matrix Factorization

Sanjar Karaev, Max-Planck Institute for Informatics, Germany; James Hook, University of Bath, United Kingdom; Pauli Miettinen, Max Planck Institute for Informatics, Germany

10:20-10:35 Topic Modeling Based on Keywords and Context

Johannes Schneider, Universitaet Liechtenstein, Switzerland

10:40-10:55 Discovering Hidden Topical Hubs and Authorities in Online Social Networks

Roy Ka-Wei Lee, Singapore Management University, Singapore; Tuan-Anh Hoang, Leibniz University Hannover, Germany; Ee-Peng Lim, Singapore Management University, Singapore

11:00-11:15 SamBaTen: Sampling-Based Batch Incremental Tensor Decomposition

Ekta Gujral, Ravdeep Pasricha, and Evangelos Papalexakis, University of California, Riverside, USA

11:20-11:35 ParaSketch: Parallel Tensor Factorization Via Sketching

Bo Yang and Ahmed Zamzam, University of Minnesota, USA; Nicholas Sidiropoulos, University of Virginia, USA

11:40-11:55 The Trustworthy Pal: Controlling the False Discovery Rate in Boolean Matrix Factorization

Sibylle Hess, Nico Piatkowski, and Katharina Morik, TU Dortmund, Germany

Lunch Break

12:00 PM-1:15 PM

Attendees on their own

Friday, May 4

TS5

Tutorial Session: The Canonical Polyadic Tensor: Decomposition and Variants for Mining Multi-Dimensional Data

1:15 PM-3:15 PM

Room:Sierra 5-6

Chair: Yan Liu, University of Southern California, USA

Multi-dimensional or multi-way datasets are becoming increasingly common in science and engineering applications. Data structures that live in three or more dimensions often exhibit informative hidden structures that can be discovered and understood through tensor decompositions. The purpose of this tutorial is to dive deep into the canonical polyadic tensor decomposition (also known as CANDECOMP, PARAFAC, or just CP), giving attendees the mathematical and algorithmic tools to understand existing methods and have a strong foundation for developing their own tools. The tutorial begin with the basics and build up to very recent developments. It is appropriate for anyone at the graduate school level or higher, with a basic understanding of numerical methods. A unique feature of our proposed tutorial will be hands-on exercises using the Tensor Toolbox for MATLAB to apply tensor decompositions to real-world open source datasets. Through these exercises, we hope to give attendees a glimpse into the application of these methods and the open problems that still exist (like choosing the rank of the tensor decomposition). We expect that most attendees will already have access to MATLAB through their universities, but we also intend to work with Mathworks to get temporary licenses for participants. We will work with one dataset that is nearly 2 GB, so we will invite participants to download the datasets ahead of time.

continued in next column

Tamara G. Kolda
Sandia National Laboratories, USA

Daniel M. Dunlavy
Sandia National Laboratories, USA

Friday, May 4

CP9

Novel Learning Methods 1

1:15 PM-3:15 PM

Room:Salon A-C

Chair: Martin Ester, Simon Fraser University, Canada

1:15-1:30 Exploiting Structure for Fast Kernel Learning
Trefor W. Evans and Prasanth B. Nair, University of Toronto, Canada

1:35-1:50 Global Nonlinear Metric Learning by Gluing Local Linear Metrics
Yaxin Peng, Lingfang Hu, and Shihui Ying, Shanghai University, China; *Chaomin Shen*, East China Normal University, China

1:55-2:10 Co-Regularized Monotone Retargeting for Semi-Supervised Letor
Shalmali Joshi, Rajiv Khanna, and Joydeep Ghosh, University of Texas at Austin, USA

2:15-2:30 Markov Chain Monitoring
Harshal Chaudhari, Boston University, USA; Michael Mathioudakis, University of Helsinki, Finland; Evimaria Terzi, Boston University, USA

2:35-2:50 Multi-view Weak-label Learning Based on Matrix Completion
Qiaoyu Tan and *Guoxian Yu*, Southwest University, China; Carlotta Domeniconi, George Mason University, USA; Jun Wang and Zili Zhang, Southwest University, China

2:55-3:10 Efficient and Effective Accelerated Hierarchical Higher-Order Logistic Regression for Large Data Quantities
Nayyar Zaidi, Francois Petitjean, and Geoffrey Webb, Monash University, Australia

Friday, May 4

CP10

Classification

1:15 PM-3:15 PM

Room:Salon D

Chair: Takashi Washio, Osaka University, Japan

1:15-1:30 Discriminative Prototype Set Learning for Nearest Neighbor Classification
Shin Ando, Tokyo University of Science, Japan

1:35-1:50 ALE: Additive Latent Effect Models for Grade Prediction
Zhiyun Ren, George Mason University, USA; Xia Ning, Indiana University - Purdue University Indianapolis, USA; Huzefa Rangwala, George Mason University, USA

1:55-2:10 A Salient Ensemble of Trees Using Cascaded Linear Classifiers with Feature-Cost Constraints
Chien-Wen Huang, Chung-Kuang Chou, and Ming-Syan Chen, National Taiwan University, Taiwan

2:15-2:30 An Lstm Approach to Patent Classification Based on Fixed Hierarchy Vectors
Matthias Schubert, Ludwig-Maximilians-Universität München, Germany; Marawan Shalaby, Technical University of Munich, Germany; Jan Stutzki, Ludwig Maximilian University of Munich, Germany; Stephan Günnemann, Technical University of Munich, Germany

2:35-2:50 Limited-Memory Common-Directions Method for Distributed L1-Regularized Linear Classification
Wei-Lin Chiang and Yu-Sheng Li, National Taiwan University, Taiwan; Ching-Pei Lee, University of Wisconsin, Madison, USA; Chih-Jen Lin, National Taiwan University, Taiwan

2:55-3:10 A Practitioners' Guide to Transfer Learning for Text Classification Using Convolutional Neural Networks
Tushar Semwal, Indian Institute of Technology Guwahati, India; Gaurav Mathur and Promod Yenigalla, Samsung R&D Institute India, Bangalore; Shivashankar Nair, Indian Institute of Technology, Guwahati, India

Coffee Break

3:15 PM-3:30 PM

Room:Salon E-H

Friday, May 4

TS5

Tutorial Session, continued: The Canonical Polyadic Tensor: Decomposition and Variants for Mining Multi-Dimensional Data

1:15 PM-3:15 PM

Room:Sierra 5-6

Chair: Yan Liu, University of Southern California, USA

Multi-dimensional or multi-way datasets are becoming increasingly common in science and engineering applications. Data structures that live in three or more dimensions often exhibit informative hidden structures that can be discovered and understood through tensor decompositions. The purpose of this tutorial is to dive deep into the canonical polyadic tensor decomposition (also known as CANDECOMP, PARAFAC, or just CP), giving attendees the mathematical and algorithmic tools to understand existing methods and have a strong foundation for developing their own tools. The tutorial begin with the basics and build up to very recent developments. It is appropriate for anyone at the graduate school level or higher, with a basic understanding of numerical methods. A unique feature of our proposed tutorial will be hands-on exercises using the Tensor Toolbox for MATLAB to apply tensor decompositions to real-world open source datasets. Through these exercises, we hope to give attendees a glimpse into the application of these methods and the open problems that still exist (like choosing the rank of the tensor decomposition). We expect that most attendees will already have access to MATLAB through their universities, but we also intend to work with Mathworks to get temporary licenses for participants. We will work with one dataset that is nearly 2 GB, so we will invite participants to download the datasets ahead of time.

continued on next page

Tamara G. Kolda
Sandia National Laboratories, USA

Daniel M. Dunlavy
Sandia National Laboratories, USA

Friday, May 4

CP11

Time Series Data 2

3:30 PM-5:10 PM

Room:Salon A-C

Chair: Joao Gama, University of Porto, Portugal

3:30-3:45 Sparse Decomposition for Time Series Forecasting and Anomaly Detection

Sunav Choudhary, Adobe Systems, USA; Gaurush Hiranandani, UIUC, USA; Shiv Saini, Adobe Systems, USA

3:50-4:05 StreamCast: Fast and Online Mining of Power Grid Time Sequences

Bryan Hooi, Hyun Ah Song, Amritanshu Pandey, Marko Jereminov, Larry Pileggi, and Christos Faloutsos, Carnegie Mellon University, USA

4:10-4:25 Exact Mean Computation in Dynamic Time Warping Spaces

Markus Brill, Technische Universitaet Berlin, Germany; Till Fluschnik, *Vincent Froese*, Brijnesh Jain, Rolf Niedermeier, and David Schultz, TU Berlin, Germany

4:30-4:45 Framework for Inferring Leadership Dynamics of Complex Movement from Time Series

Chainarong Amornbunchornvej and Tanya Y. Berger-Wolf, University of Illinois, Chicago, USA

4:50-5:05 Brain EEG Time Series Selection: A Novel Graph-Based Approach for Classification

Chenglong Dai, Nanjing University of Aeronautics and Astronautics, China; Jia Wu, Macquarie University, Sydney, Australia; *Dechang Pi* and Lin Cui, Nanjing University of Aeronautics and Astronautics, China

Friday, May 4

CP12

Applied Data Science

3:30 PM-5:10 PM

Room:Salon D

Chair: - Jiayu Zhou, Michigan State University, USA

3:30-3:45 A Rare and Critical Condition Search Technique and its Application to Telescope Stray Light Analysis

Keiichi Kisamori, National Institute of Advanced Industrial Science and Technology, Japan; Takashi Washio, Osaka University, Japan; Yoshio Kameda, National Institute of Advanced Industrial Science and Technology, Japan; Ryohei Fujimaki, NEC Global, Japan

3:50-4:05 Revenue Maximization on the Multi-Grade Product

Ya-Wen Teng, National Taiwan University, Taiwan; Chih-Hua Tai, National Taipei University of Technology, Taiwan; Philip Yu, University of Illinois, Chicago, USA; Ming-Syan Chen, National Taiwan University, Taiwan

4:10-4:25 Avoidance Region Discovery: A Summary of Results

Emre Eftelioglu, Shashi Shekhar, and Xun Tang, University of Minnesota, USA

4:30-4:45 Learning Convolutional Text Representations for Visual Question Answering

Zhengyang Wang and Shuiwang Ji, Washington State University, USA

4:50-5:05 Black-Box Expectation Propagation for Bayesian Models

Ximing Li, ; Changchun Li, Jinjin Chi, Jihong Ouyang, and Wenting Wang, Jilin University, China

Organizational Break

5:10 PM-5:15 PM

Friday, May 4

Panel: Broadening Participation in Data Science

5:15 PM-6:30 PM

Room:Salon D

Award Ceremony and SIAG/DMA Business Meeting

6:30 PM-7:00 PM

Room:Salon D

Complimentary beer and wine will be served

Doctoral Forum and Poster Session Part II

7:00 PM-9:00 PM

Room:Rio Vista Pavilion

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

Saturday, May 5

Registration

7:30 AM-4:00 PM

Room:Rio Vista Ballroom Foyer

Continental Breakfast

7:30 AM-8:00 AM

Room:Salon E-H

IP4

From Robots to Biomolecules: Computing Meets the Physical World

8:00 AM - 9:15 AM

Room: Salon A-D

Chair: Dimitrios Gunopulos, University of Athens, Greece

The development of fast and reliable motion planning algorithms has deeply influenced many domains in robotics, such as industrial automation and autonomous exploration, but has also contributed novel methodologies to distant domains such as computational structural biology. This talk will present recent work on the computation of low-level plans from high-level specifications. High-level specifications declare what the robot must do, rather than how this task is to be done. The talk will also discuss robotics-inspired methods for computing the flexibility of proteins and for molecular docking with the ultimate goal of deciphering molecular function and aiding the discovery of new therapeutics.

Lydia Kavraki
Rice University, USA

Coffee Break

9:15 AM-9:30 AM

Room:Salon E-H

Saturday, May 5

CP13

Personalization

9:30 AM-11:30 PM

Room:Salon D

Chair:Beidou Wang, Simon Fraser University, Canada

9:30-9:45 Learning to Interact with Users: A Collaborative-Bandit Approach

Konstantina Christakopoulou and Arindam Banerjee, University of Minnesota, USA

9:50-10:05 Robust Cost-Sensitive Learning for Recommendation with Implicit Feedback

Peng Yang, King Abdullah University of Science & Technology (KAUST), Saudi Arabia; Peilin Zhao, South China University of Technology, China; Yong Liu, NTUC Link, Singapore; Xin Gao, King Abdullah University of Science & Technology (KAUST), Saudi Arabia

10:10-10:25 Modeling Item-Specific Effects for Video Click

Fei Tan, Kuang Du, Zhi Wei, and Haoran Liu, New Jersey Institute of Technology, USA; Chenguang Qin and Ran Zhu, PPLive, Inc, China

10:30-10:45 One-Class Recommendation with Asymmetric Textual Feedback

Mengting Wan and Julian McAuley, University of California, San Diego, USA

10:50-11:05 Online It Ticket Automation Recommendation Using Hierarchical Multi-Armed Bandit Algorithms

Qing Wang, Tao Li, and S.S. Iyengar, Florida International University, USA; Larisa Shwartz, IBM T.J. Watson Research Center, USA; Genady Grabarnik, St. John's University, USA

11:10-11:25 Investigating Deep Reinforcement Learning Techniques in Personalized Dialogue Generation

Min Yang, Chinese Academy of Sciences, China; Qiang Qu, Shenzhen Institute of Advanced Technology, China; Kai Lei, Peking University, China; Jia Zhu, South China Normal University, China; Zhou Zhao, Zhejiang University, China; Xiaojun Chen and Joshua Zhexue Huang, Shenzhen University, China

Saturday, May 5

Workshop 1: Big Traffic Data Analytics

9:30 AM-5:45 PM

Room:Sierra-6

Workshop 2: Artificial Intelligence in Insurance

9:30 AM-5:45 PM

Room:Sierra 5

Workshop 3: Machine Learning Methods for Recommender Systems

9:30 AM-5:45 PM

Room:Salon A-C

Workshop 4: Cost-Sensitive Learning

9:30 AM-5:45 PM

Room:Balboa 1-2

Workshop 5: Data Mining for Geophysics and Geology

9:30 AM-5:45 PM

Room:Cabrillo Salon 1

Workshop 6: Data Mining for Medicine and Healthcare

9:30 AM-5:45 PM

Room:Cabrillo Salon 2

Lunch Break

12:15 PM-1:30 PM

Attendees on their own

Saturday, May 5

CP14

Networks 2

1:30 PM-3:30 PM

Room:Salon D

Chair: Danai Koutra, University of Michigan, USA

1:30-1:45 Reconstructing a Cascade from Temporal Observations

Han Xiao, Polina Rozenshtein, Nikolaj Tatti, and Aristides Gionis, Aalto University, Finland

1:50-2:05 Modeling Co-Evolution Across Multiple Networks

Wenchao Yu, University of California, Los Angeles, USA; Charu C. Aggarwal, IBM T.J. Watson Research Center, USA; Wei Wang, University of California, Los Angeles, USA

2:10-2:25 Multi-Layered Network Embedding

Jundong Li, Chen Chen, Hanghang Tong, and Huan Liu, Arizona State University, USA

2:30-2:45 Maximizing the Effect of Information Adoption: A General Framework

Tianyuan Jin, Tong Xu, Hui Zhong, Enhong Chen, Zhefeng Wang, and Qi Liu, University of Science and Technology of China, China

2:50-3:05 SMACD: Semi-Supervised Multi-Aspect Community Detection

Ekta Gujral and Evangelos Papalexakis, University of California, Riverside, USA

3:10-3:25 Toward Relational Learning with Misinformation

Liang Wu and Jundong Li, Arizona State University, USA; Fred Morstatter, University of Southern California, USA; Huan Liu, Arizona State University, USA

Coffee Break

3:30 PM-3:45 PM

Room:Salon E-H

Saturday, May 5

CP15

Novel Learning Methods 2

3:45 PM-5:25 PM

Room:Salon D

Chair: Jiliang Tang, Michigan State University, USA

3:45-4:00 Personalized Ranking on Poisson Factorization

Li-Yen Kuo, Chung-Kuang Chou, and Ming-Syan Chen, National Taiwan University, Taiwan

4:05-4:20 Strongly Hierarchical Factorization Machines and Anova Kernel Regression

Ruocheng Guo, Hamidreza Alvari, and Paulo Shakarian, Arizona State University, USA

4:25-4:40 A Novel Genetic Algorithm for Feature Selection in Hierarchical Feature Spaces

Pablo Silva, Universidade Federal Fluminense, Brazil; Alexandre Plastino, Fluminense Federal University, Brazil; Alex A. Freitas, University of Kent, United Kingdom

4:45-5:00 Making Kernel Density Estimation Robust Towards Missing Values in Highly Incomplete Multivariate Data Without Imputation

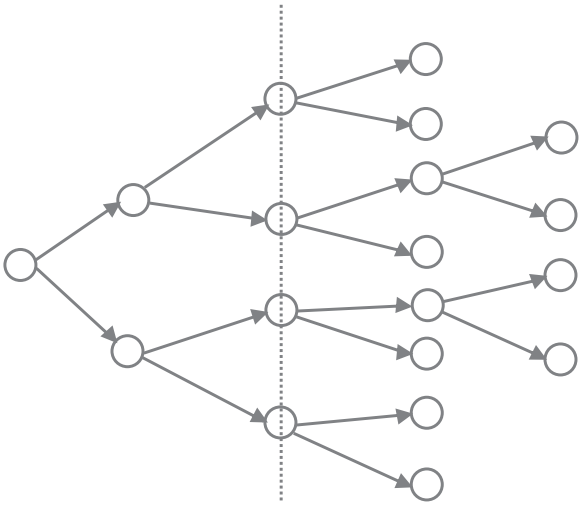
Richard Leibrandt and Stephan Gunneman, Technische Universität München, Germany

5:05-5:20 Dense Neighborhood Pattern Sampling in Numerical Data

Arnaud Giacometti and *Arnaud Soulet*, Universite François Rabelais, France

Abstracts

2018 SIAM International Conference on DATA MINING



May 3-5, 2018

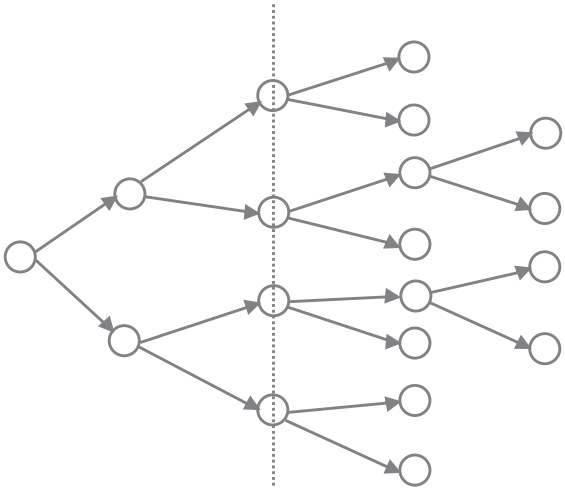
San Diego Marriott Mission Valley
San Diego, California, USA

Abstracts are printed as submitted by the authors.

Notes

Organizer and Speaker Index

2018 SIAM International Conference on DATA MINING



May 3-5, 2018

San Diego Marriott Mission Valley
San Diego, California, USA

A
Abbar, Sofiane, CP3, 3:45 Thu
Adhikari, Bijaya, CP3, 2:45 Thu
Amornbunchornvej, Chainarong, CP11, 4:30 Fri
Ando, Shin, CP10, 1:15 Fri

B
Bai, Zilong, CP1, 11:40 Thu
Budhathoki, Kailash, CP2, 10:00 Thu

C
Cai, Xingyu, CP2, 10:40 Thu
Chan, Jeffrey, CP1, 10:40 Thu
Chaudhari, Harshal, CP9, 2:15 Fri
Chen, Liangzhe, TS2, 2:45 Thu
Chen, Muhao, CP7, 10:20 Fri
Chen, Weitong, CP6, 5:40 Thu
Cheng, Ruihua, CP3, 3:05 Thu
Chiang, Wei-Lin, CP10, 2:35 Fri
Choudhary, Sunav, CP11, 3:30 Fri
Christakopoulou, Konstantina, CP13, 9:30 Sat

D
Ding, Jingtao, CP4, 4:05 Thu

E
Eftelioglu, Emre, CP12, 4:10 Fri
Eliassi-Rad, Tina, TS4, 10:00 Fri
Evans, Trefor W., CP9, 1:15 Fri

F
Feremans, Len, CP2, 11:20 Thu
Froese, Vincent, CP11, 4:10 Fri
Fu, Yanjie, CP4, 3:05 Thu
Fu, Yanjie, CP7, 11:40 Fri

G
Gligorijevic, Djordje, CP6, 6:20 Thu
Gujral, Ekta, CP8, 11:00 Fri
Gujral, Ekta, CP14, 2:50 Sat
Guo, Ruocheng, CP15, 4:05 Sat

H
Hess, Sibylle, CP8, 11:40 Fri
Hooi, Bryan, CP11, 3:50 Fri
Huai, Mengdi, CP6, 5:20 Thu
Huang, Chien-Wen, CP10, 1:55 Fri

I
Ioannidis, Stratis, CP7, 11:00 Fri

J
Jiang, Fei, CP7, 10:40 Fri
Jin, Tianyuan, CP14, 2:30 Sat
Joshi, Shalmali, CP9, 1:55 Fri

K
Karaev, Sanjar, CP8, 10:00 Fri
Kavraki, Lydia, IP4, 8:00 Sat
Kisamori, Keiichi, CP12, 3:30 Fri
Kolda, Tamara G., TS5, 1:15 Fri
Kolda, Tamara G., TS5, 3:30 Fri
Kuo, Li-Yen, CP15, 3:45 Sat

L
Lee, Ritchie, CP5, 5:00 Thu
Lee, Roy Ka-Wei, CP8, 10:40 Fri
Leibrandt, Richard, CP15, 4:45 Sat
Li, Jundong, CP14, 2:10 Sat
Li, Keqian, CP1, 11:20 Thu
Li, Xiaosheng, CP5, 6:00 Thu
Li, Ximing, CP12, 4:50 Fri
Li, Yan, CP6, 6:00 Thu
Liang, Jiongqian, CP3, 4:25 Thu
Liu, Yan, MT1, 10:00 Thu
Liu, Yan, MT2, 2:45 Thu
Liu, Yan, MT3, 5:00 Thu
Liu, Yan, MT4, 10:00 Fri
Liu, Yan, MT5, 1:15 Fri
Liu, Yan, MT0, 3:30 Fri

M
Ma, Tengfei, CP6, 5:00 Thu
Mao, Jiali, CP2, 10:20 Thu
Medya, Sourav, CP3, 3:25 Thu
Meyerhenke, Henning, CP1, 10:00 Thu
Morvan, Anne, CP1, 10:20 Thu

N
Nayak, Guruprasad, CP5, 6:20 Thu
Nguyen, Dang, CP7, 10:00 Fri
Ning, Yue, CP2, 11:40 Thu

O
Olteanu, Alexandra, TS1, 10:00 Thu

P
Pereira, Fabíola, TS3, 5:00 Thu
Pi, Dechang, CP11, 4:50 Fri

R
Ren, Zhiyun, CP10, 1:35 Fri

S
Schneider, Johannes, CP8, 10:20 Fri
Schölkopf, Bernhard, IP1, 8:15 Thu
Schubert, Matthias, CP10, 2:15 Fri
Semwal, Tushar, CP10, 2:55 Fri
Shen, Chaomin, CP9, 1:35 Fri
Shen, Yilin, CP5, 5:40 Thu
Shi, Yu, CP3, 4:05 Thu
Silva, Pablo, CP15, 4:25 Sat
Smith, Shaden, CP2, 11:00 Thu
Soulet, Arnaud, CP15, 5:05 Sat

T
Tan, Chang Wei, CP5, 5:20 Thu
Tan, Fei, CP13, 10:10 Sat
Tatsubori, Michiaki, CP4, 4:25 Thu
Teng, Ya-Wen, CP12, 3:50 Fri

W
Wan, Mengting, CP13, 10:30 Sat
Wang, Huandong, CP4, 3:45 Thu
Wang, Jun, CP1, 11:00 Thu
Wang, Qing, CP13, 10:50 Sat
Wang, Zhengyang, CP12, 4:30 Fri
Wu, Liang, CP14, 3:10 Sat

Italicized names indicate session organizers

Wu, Xian, CP4, 3:25 Thu

X

Xiao, Han, CP14, 1:30 Sat

Y

Yan, Yujun, CP7, 11:20 Fri
Yang, Bo, CP8, 11:20 Fri
Yang, Min, CP13, 11:10 Sat
Yang, Peng, CP13, 9:50 Sat
Yao, Liuyi, CP4, 2:45 Thu
Yu, Guoxian, CP9, 2:35 Fri
Yu, Wenchao, CP14, 1:50 Sat

Z

Zaidi, Nayyar, CP9, 2:55 Fri

Italicized names indicate session organizers

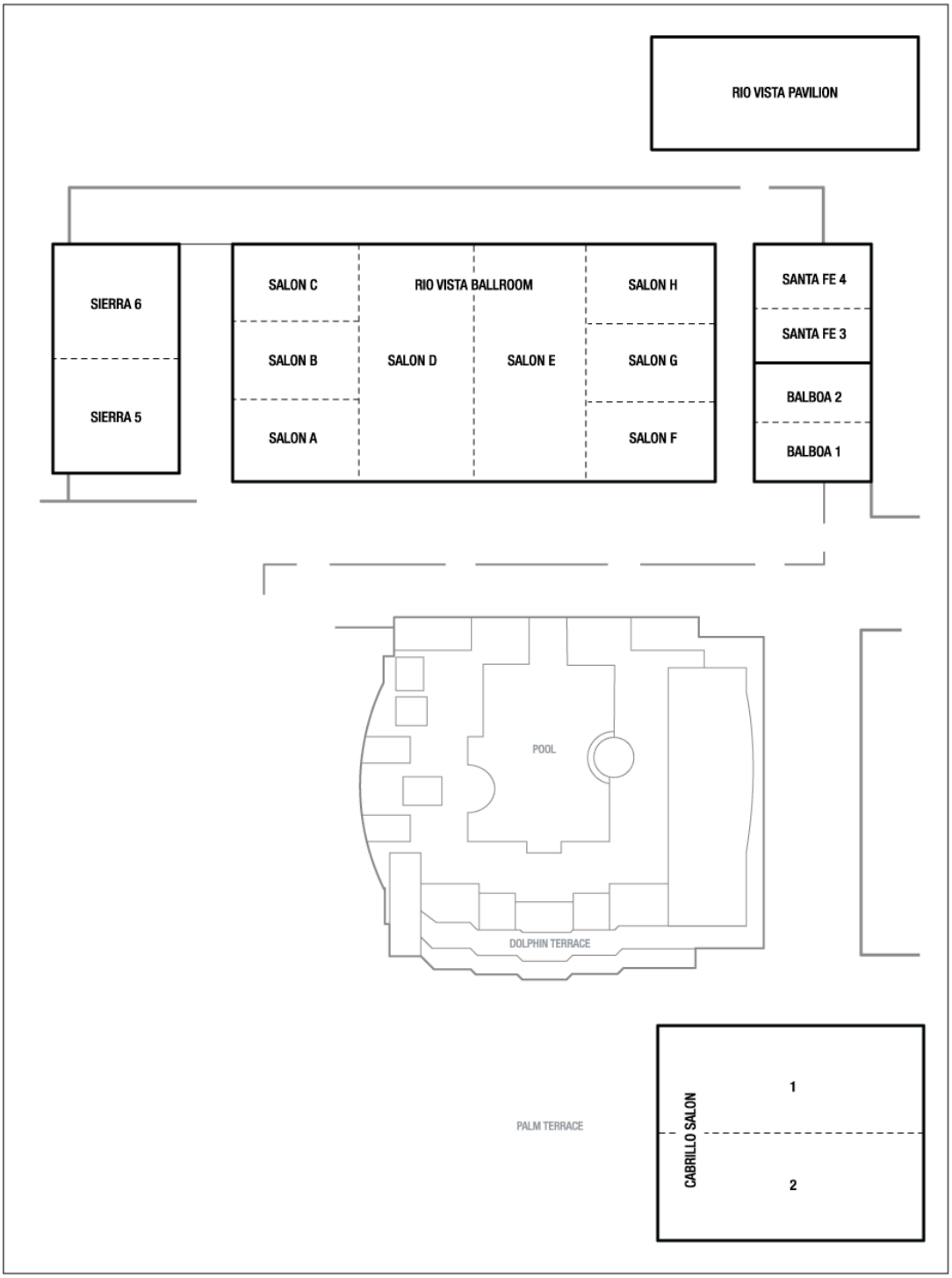
Notes

SDM18 Budget

Conference Budget
SIAM International Conference on Data Mining
May 3-5, 2018
San Diego, California

Expected Paid Attendance		260
Revenue		
Registration Income		$129,555
	Total	$129,555
Expenses		
Printing		$1,200
Organizing Committee		2,700
Invited Speakers		7,480
Food and Beverage		34,000
AV Equipment and Telecommunication		17,100
Advertising		6,900
Proceedings		6,000
Conference Labor (including benefits)		49,906
Other (supplies, staff travel, freight, misc.)		7,550
Administrative		12,598
Accounting/Distribution & Shipping		8,240
Information Systems		13,939
Customer Service		5,376
Marketing		8,797
Office Space (Building)		5,716
Other SIAM Services		7,033
	Total	$194,535
Net Conference Expense		($64,980)
Support Provided by SIAM		$64,980
		$0

Estimated Support for Travel Awards not included above:

Early Career and Students	16	$12,100

San Diego Marriott Mission Valley
Hotel Floor Plan