# KDD 2016 Agenda at a Glance

## Friday, August 12th

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00AM - 5:00PM</td>
<td>BPDM Workshop</td>
<td>Continental 9</td>
</tr>
<tr>
<td>4:00AM - 6:00PM</td>
<td>Registration</td>
<td>East Lounge</td>
</tr>
</tbody>
</table>

## Saturday, August 13th (Tutorials)

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00AM - 5:00PM</td>
<td>Registration</td>
<td>East Lounge</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>BPDM Workshop</td>
<td>Continental 9</td>
</tr>
<tr>
<td>10:00AM-10:30AM</td>
<td>Coffee Break</td>
<td>Continental Foyer</td>
</tr>
<tr>
<td>12:00PM - 1:00PM</td>
<td>Lunch - ON OWN</td>
<td></td>
</tr>
<tr>
<td>3:00PM - 3:30PM</td>
<td>Coffee Break</td>
<td>Continental Foyer</td>
</tr>
</tbody>
</table>

## MORNING TUTORIALS

<table>
<thead>
<tr>
<th>Time</th>
<th>Tutorial</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00AM - 12:00PM</td>
<td>Tutorial 12: Enabling the Discovery of Reliable Information from Passively and Actively Crowdsourced Data</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td></td>
<td>Tutorial 10: Business Applications of Predictive Modeling at Scale</td>
<td>Imperial Ballroom A &amp; B</td>
</tr>
<tr>
<td></td>
<td>Tutorial 3: Collective Sensemaking via Social Sensors</td>
<td>Continental 1-4</td>
</tr>
<tr>
<td></td>
<td>Tutorial 7: Algorithmic Bias: From Discrimination Discovery to Fairness-Aware Data Mining</td>
<td>Continental 6-8</td>
</tr>
<tr>
<td></td>
<td>Tutorial 8: Extracting Optimal Performance From Dynamic Time Warping</td>
<td>Grand Ballroom Salon A</td>
</tr>
</tbody>
</table>

## AFTERNOON TUTORIALS
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Tutorial 11: Leveraging Propagation for Data Mining: Models, Algorithms and Applications</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Tutorial 9: Scalable Learning of Graphical Models</td>
<td>Imperial Ballroom A &amp; B</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Tutorial 4: Lifelong Machine Learning and Computer Reading the Web</td>
<td>Continental 1-4</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Tutorial 6: Healthcare Data Mining with Matrix Models</td>
<td>Continental 5</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Tutorial 5: IoT Big Data Stream Mining</td>
<td>Continental 6-8</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Tutorial 2: Deep Learning</td>
<td>Grand Ballroom Salon A</td>
</tr>
</tbody>
</table>

**Sunday, August 14th (Workshops and Opening)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30AM - 6:00PM</td>
<td>Registration</td>
<td>East Lounge</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Exhibitor Move-In</td>
<td>Continental Ballroom</td>
</tr>
<tr>
<td>10:00AM-10:30AM</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td>12:00PM - 1:00PM</td>
<td>Lunch - ON OWN</td>
<td></td>
</tr>
<tr>
<td>3:00PM - 3:30PM</td>
<td>Coffee Break</td>
<td></td>
</tr>
</tbody>
</table>

**FULL DAY WORKSHOPS**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00AM - 5:00PM</td>
<td>KDD Cup 2016: Towards Measuring the Impact of Research Institutions</td>
<td>Union Square 23/24</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>Workshop 1: Mining and Learning with Graphs</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>Workshop 2: Large-Scale Sports Analytics</td>
<td>Franciscan A/B</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>Workshop 3: Data Science for Food, Energy and Water</td>
<td>Market - Parc 55</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>Workshop 4: Mining and Learning from Time Series</td>
<td>Imperial Ballroom A &amp; B</td>
</tr>
<tr>
<td>Time</td>
<td>Workshop</td>
<td>Venue</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>Workshop 5: The 5th International Workshop on Urban Computing</td>
<td>Union Square 15/16</td>
</tr>
<tr>
<td>8:00AM - 5:00PM</td>
<td>Workshop 6: Interactive Data Exploration and Analytics</td>
<td>Embarcadero - Parc 55</td>
</tr>
<tr>
<td></td>
<td><strong>HALF-DAY MORNING WORKSHOPS</strong></td>
<td></td>
</tr>
<tr>
<td>8:00AM - 12:00PM</td>
<td>Workshop 7: Machine Learning for Large scale Transportation Systems</td>
<td>Franciscan C/D</td>
</tr>
<tr>
<td>8:00AM - 12:00PM</td>
<td>Workshop 8: Large-Scale Deep Learning for Data Mining</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>8:00AM - 12:00PM</td>
<td>Workshop 9: Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>8:00AM - 12:00PM</td>
<td>Workshop 10: Enterprise Intelligence</td>
<td>Union Square 22</td>
</tr>
<tr>
<td>8:00AM - 12:00PM</td>
<td>Workshop 11: Workshop on Issues of Sentiment Discovery and Opinion Mining</td>
<td>Mission - Parc 55</td>
</tr>
<tr>
<td></td>
<td><strong>HALF-DAY AFTERNOON WORKSHOPS</strong></td>
<td></td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Workshop 12: Outlier Definition, Detection, and Description On-Demand</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Workshop 13: Workshop on Causal Discovery</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Workshop 14: Machine Learning Meets Fashion: Data, Algorithms and Analytics for the Fashion Industry</td>
<td>Franciscan C/D</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Workshop 15: Machine Learning for Prognostics and Health Management</td>
<td>Mission - Parc 55</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Workshop 16: 15th International Workshop on Data Mining in Bioinformatics</td>
<td>Union Square 22</td>
</tr>
<tr>
<td>5:30PM - 7:00PM</td>
<td>KDD 2016 Opening Session</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>6:00PM - 7:00PM</td>
<td>Poster Reception Presenter Set-Up</td>
<td>Golden Gate</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>7:00PM -</td>
<td>Poster Reception: Research (oral only)</td>
<td>Golden Gate</td>
</tr>
<tr>
<td>12:00AM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7:30AM - 5:15PM</td>
<td>Registration</td>
<td>East Lounge</td>
</tr>
<tr>
<td>8:00AM - 6:00PM</td>
<td>Press Room</td>
<td>Union Square 24</td>
</tr>
<tr>
<td>9:15AM - 5:30PM</td>
<td>Exhibit Hall</td>
<td>Continental</td>
</tr>
<tr>
<td>8:00AM - 9:30AM</td>
<td>Keynote: Jennifer Chayes</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>9:00AM - 12:00PM</td>
<td>Hands-On Tutorial: Recommender Systems</td>
<td>Market - Parc 55</td>
</tr>
<tr>
<td>9:00AM - 12:00PM</td>
<td>Hands-On Tutorial: Big Natural Language Data Processing</td>
<td>Embarcadero - Parc 55</td>
</tr>
<tr>
<td>9:30AM - 10:00AM</td>
<td>Coffee Break</td>
<td>Exhibit Hall &amp; Foyers</td>
</tr>
<tr>
<td>10:00AM - 12:00PM</td>
<td>Research Session R1: Graphs and Rich Data</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>10:00AM - 12:00PM</td>
<td>Research Session R2: Large-Scale Data Mining</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>10:00AM - 12:00PM</td>
<td>Applied Data Science Invited Talks 1</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>10:00AM - 12:00PM</td>
<td>Applied Data Science Session A1: Social Good - I</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>12:15PM - 1:30PM</td>
<td>Lunch Keynote - Joe Hellerstein (RSVP Required)</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Hands-On Tutorial: MXNet</td>
<td>Market - Parc 55</td>
</tr>
<tr>
<td>1:00PM - 5:00PM</td>
<td>Hands-On Tutorial: Spark 2.0</td>
<td>Embarcadero - Parc 55</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Research Session R3: Graphs and Social Networks - I</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Research Session R4: Streams and Temporal Evolution - I</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Panel Session: Is Deep Learning the New Yosemite A/B/C</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Applied Data Science Session A2: Social Good - II</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>1:45PM - 6:15PM</td>
<td>Networking Session: Data Science in India - India Chapter of ACM SIGKDD</td>
<td>Union Square 15/16</td>
</tr>
<tr>
<td>3:45PM - 4:15PM</td>
<td>Coffee Break</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>Research Session R5: Graphs and Social Networks - II</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>Research Session R6: Streams and Temporal Evolution - II</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>Applied Data Science Invited Talks 2</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>Applied Data Science Session A3: Miscellaneous Topics</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>6:00PM - 7:00PM</td>
<td>Poster Reception Presenter Set-Up</td>
<td>Golden Gate</td>
</tr>
<tr>
<td>7:00PM - 12:00AM</td>
<td>Poster Reception: Applied Data Science</td>
<td>Golden Gate</td>
</tr>
</tbody>
</table>

**Tuesday, August 16th (Main Conference)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:45AM - 5:15PM</td>
<td>Registration</td>
<td>East Lounge</td>
</tr>
<tr>
<td>8:00AM - 6:00PM</td>
<td>Press Room</td>
<td>Union Square 24</td>
</tr>
<tr>
<td>9:15AM - 5:30PM</td>
<td>Exhibit Hall</td>
<td>Continental</td>
</tr>
<tr>
<td>8:00AM - 9:30AM</td>
<td>Keynote: Greg Papadopoulos</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>9:00AM - 12:00PM</td>
<td>Hands-On Tutorial: AWS</td>
<td>Market - Parc 55</td>
</tr>
<tr>
<td>9:00AM - 12:00PM</td>
<td>Hands-On Tutorial: CNTK</td>
<td>Embarcadero - Parc 55</td>
</tr>
<tr>
<td>9:30AM - 10:00AM</td>
<td>Coffee Break</td>
<td>Multi-Level Foyers</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>9:30AM - 12:30PM</td>
<td>Networking Session: Standards for Predictive Analytics in the Era of Big and Fast Data - Data Mining Group</td>
<td>Union Square 15/16</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Research Session R7: Clustering</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Research Session R8: Deep Learning and Embedding</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Applied Data Science Invited Talks 3</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>VC Office Hours</td>
<td>Union Square 17/18</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Applied Data Science Session A4: Recommendations and Ranking</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>12:15PM - 1:30PM</td>
<td>KDD 2016 Business Lunch (RSVP Required)</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Research Session R9: Unsupervised Learning and Anomaly Detection</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Research Session R10: Recommender Systems</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Applied Data Science Invited Talks 4</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>VC Office Hours</td>
<td>Union Square 17/18</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>VC Office Hours</td>
<td>Union Square 19/20</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>Applied Data Science Session A5: Social Networks and Social Media</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>1:45PM - 3:45PM</td>
<td>SIGKDD Dissertation Awards</td>
<td>Embarcadero - Parc 55</td>
</tr>
<tr>
<td>3:45PM - 4:15PM</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>Turing Talk: Whitfield Diffie with Opening Introduction by Alex Wolf</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>VC Office Hours</td>
<td>Union Square 17/18</td>
</tr>
<tr>
<td>4:15PM - 6:15PM</td>
<td>VC Office Hours</td>
<td>Union Square 19/20</td>
</tr>
<tr>
<td>6:00PM - 7:00PM</td>
<td>Poster Reception Presenter Set-Up</td>
<td>Golden Gate</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>7:00PM - 12:00AM</td>
<td>Poster Reception: Research (poster only)</td>
<td>Golden Gate</td>
</tr>
<tr>
<td>8:15AM - 3:00PM</td>
<td>Registration</td>
<td>East Lounge</td>
</tr>
<tr>
<td>8:00AM - 6:00PM</td>
<td>Press Room</td>
<td>Union Square 24</td>
</tr>
<tr>
<td>9:15AM - 1:30PM</td>
<td>Exhibit Hall</td>
<td>Continental</td>
</tr>
<tr>
<td>8:00AM-9:30AM</td>
<td>Keynote: Nando de Freitas</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>9:00AM - 12:00PM</td>
<td>Hands-On Tutorial: Streaming Analytics</td>
<td>Market - Parc 55</td>
</tr>
<tr>
<td>9:00AM - 12:00PM</td>
<td>Hands-On Tutorial: R on Spark</td>
<td>Embarcadero - Parc 55</td>
</tr>
<tr>
<td>9:30AM - 10:00AM</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Research Session R11: Supervised Learning</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Research Session R12: Sequence Mining</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Applied Data Science Invited Talks 5</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>10:00AM-12:00PM</td>
<td>Applied Data Science Session A6: Machine Learning Algorithms</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>12:00PM - 1:00PM</td>
<td>Lunch - ON OWN</td>
<td></td>
</tr>
<tr>
<td>1:30PM - 4:00PM</td>
<td>Exhibitor Move-Out</td>
<td>Continental</td>
</tr>
<tr>
<td>1:00PM - 3:00PM</td>
<td>Research Session R13: User-Behavior Modeling</td>
<td>Plaza Room A/B</td>
</tr>
<tr>
<td>1:00PM - 3:00PM</td>
<td>Research Session R14: Optimization</td>
<td>Imperial Ballroom</td>
</tr>
<tr>
<td>1:00PM - 3:00PM</td>
<td>Applied Data Science Session A7: E-Commerce</td>
<td>Yosemite A/B/C</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Location</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1:00PM - 3:00PM</td>
<td>Applied Data Science Session A8: Systems and Experimentation</td>
<td>Cyril Magnin - Parc 55</td>
</tr>
<tr>
<td>3:00PM - 3:15PM</td>
<td>Coffee Break</td>
<td>Grand Ballroom</td>
</tr>
<tr>
<td>3:15PM - 4:00PM</td>
<td>KDD 2016 Closing Session</td>
<td>Grand Ballroom</td>
</tr>
</tbody>
</table>
KDD 2016 Chairs’ Welcome Message

It is our great pleasure to welcome you to the 2016 ACM Conference on Knowledge Discovery and Data Mining – KDD’16. We hope that the content and the professional network at KDD’16 will help you succeed professionally by enabling you to: identify technology trends early; make new/creative contributions; increase your productivity by using newer/better tools, processes or ways of organizing teams; identify new job opportunities; and hire new team members.

We are living in an exciting time for our profession. On the one hand, we are witnessing the industrialization of data science, and the emergence of the industrial assembly line processes characterized by the division of labor, integrated processes/pipelines of work, standards, automation, and repeatability. Data science practitioners are organizing themselves in more sophisticated ways, embedding themselves in larger teams in many industry verticals, improving their productivity substantially, and achieving a much larger scale of social impact. On the other hand, we are also witnessing astonishing progress from research in algorithms and systems -- for example the field of deep neural networks has revolutionized speech recognition, NLP, computer vision, image recognition, etc. By facilitating interaction between practitioners at large companies & startups on the one hand, and the algorithm development researchers including leading academics on the other, KDD’16 fosters technological and entrepreneurial innovation in the area of data science.

This year’s conference continues its tradition of being the premier forum for presentation of results in the field of data mining, both in the form of cutting edge research, and in the form of insights from the development and deployment of real world applications. Further, the conference continues with its tradition of a strong tutorial and workshop program on leading edge issues of data mining. The mission of this conference has broadened in recent years even as we placed a significant amount of focus on both the research and applied aspects of data mining. As an example of this broadened focus, this year we have introduced a strong hands-on tutorial program during the conference in which participants will learn how to use practical tools for data mining. KDD’16 also gives researchers and practitioners a unique opportunity to form professional networks, and to share their perspectives with others interested in the various aspects of data mining. For example, we have introduced office hours for budding entrepreneurs from our community to meet leading Venture Capitalists investing in this area. We hope that KDD 2016 conference will serve as a meeting ground for researchers, practitioners, funding agencies, and investors to help create new algorithms and commercial products.

The call for papers attracted a significant number of submissions from countries all over the world. In particular, the research track attracted 784 submissions and the applied data science track attracted 331 submissions. Papers were accepted either as full papers or as posters. The overall acceptance rate either as full papers or posters was less than 20%. For full papers in the research track, the acceptance rate was lower than 10%. This is consistent with the fact that the KDD Conference is a premier conference in data mining and the acceptance rates historically tend to be low. It is noteworthy that the applied data science track received a larger number of submissions compared to previous years. We view this as an encouraging sign that research in data mining is increasingly becoming relevant to industrial applications. All papers were reviewed by at least three program committee members and then discussed by the PC members in a discussion moderated by a
meta-reviewer. Borderline papers were thoroughly reviewed by the program chairs before final decisions were made.

In addition, KDD’16 continues to have a strong tutorial and workshop program, as is its tradition since its inception. The program committee reviewed and accepted the following:

<table>
<thead>
<tr>
<th>Venue or Track</th>
<th>Reviewed</th>
<th>Accepted</th>
<th>Acceptance Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Track Papers</td>
<td>784</td>
<td>70+72</td>
<td>8.9%+9.2%</td>
</tr>
<tr>
<td>Applied Data Science Track Papers</td>
<td>331</td>
<td>40+26</td>
<td>12%+7.9%</td>
</tr>
<tr>
<td>Workshops</td>
<td>47</td>
<td>17</td>
<td>36.2%</td>
</tr>
<tr>
<td>Tutorials</td>
<td>37</td>
<td>11</td>
<td>29.7%</td>
</tr>
<tr>
<td>Applied Data Science Talks</td>
<td>Invited</td>
<td>12</td>
<td>Invited</td>
</tr>
<tr>
<td>Hands-on Tutorials</td>
<td>Invited</td>
<td>8</td>
<td>Invited</td>
</tr>
<tr>
<td>Regular Keynotes</td>
<td>Invited</td>
<td>4</td>
<td>Invited</td>
</tr>
<tr>
<td>Turing Award Winner Talk</td>
<td>Special</td>
<td>1</td>
<td>Special</td>
</tr>
</tbody>
</table>

In the table above the research track papers and applied data science track papers are broken up into the full presentations and poster presentations. We have four excellent keynote presentations from Jennifer Chayes, Joe Hellerstein, Greg Papadopoulous, and Nando de Freitas. In addition, we are honored to have a Turing Award winner, Whitfield Diffie, as one of the keynote speakers. We also encourage attendees to attend the keynote and invited talk presentations.

Putting together KDD’16 was a team effort. We first thank the authors for providing the content of the program. We are grateful to the program committee and the senior program committee, who worked very hard in reviewing papers and providing feedback for authors. Finally, we thank the numerous sponsors and ACM SIGKDD for their support in hosting the conference.

We hope that you will find this program interesting and thought-provoking and that the conference will provide you with a valuable opportunity to share ideas with other researchers and practitioners from institutions around the world.

Charu Aggarwal
Alex Smola
KDD’16 Research Track Chairs

Rajeev Rastogi
Dou Shen
KDD’16 Applied Data Science Track Chairs

Balaji Krishnapuram
Mohak Shah
KDD’16 General Chairs
Program Highlights

Keynote Talks

- Joe Hellerstein: Professor / University of California, Berkeley; Trifacta. People, Computers, and The Hot Mess of Real Data.
- Greg Papadopoulos: Venture Partner / NEA. A VC View of Investing in ML.
- Nando de Freitas: Professor / Oxford University; Google DeepMind. Learning to Learn and Compositionality with Deep Recurrent Neural Networks.

Research and Applied Data Science Tracks

- 142 Research Track Papers
- 66 Applied Data Science Track Papers

Applied Data Science Track Invited Talks

- Jonathan Becher: Chief Digital Officer / SAP. Can You Teach The Elephant To Dance? AKA: Culture Eats Data Science for Breakfast
- Ashok Srivastava: Chief Data Scientist / Verizon. Large-Scale Machine Learning at Verizon: Theory and Applications
- Duncan Watts: Principal Researcher / Microsoft. Computational Social Science: Exciting Progress and Future Challenges
- Jeff Schneider: Engineering Lead / Uber Advanced Technology Center. Bayesian Optimization and Embedded Learning Systems
- Oliver Downs: Chief Scientist and CTO / Ampero. How Machine Learning has Finally Solved Wanamaker’s Dilemma
- Danny Shapiro: Senior Director of Automotive / NVIDIA. Accelerating the Race to Autonomous Cars
- Ralf Herbrich: Director of Machine Learning / Amazon. Learning Sparse Models at Scale
- Caitlin Smallwood: Vice President of Science and Algorithms / Netflix. It’s About Time
- Andy Palmer: Co-Founder and CEO / Tamr, Inc.

Applied Data Science Panels

Moderator: Usama Fayyad: Chief Data Officer & Group Managing Director / Barclays. *Can Tools Effectively Unleash the Power of Big Data?*

**Tutorials**

- Enabling the Discovery of Reliable Information from Passively and Actively Crowdsourced Data
- Propagation for Data Mining: Models, Algorithms and Applications
- Business Applications of Predictive Modeling at Scale
- Scalable learning of graphical models
- Extracting Optimal Performance From Dynamic Time Warping
- Algorithmic bias: from discrimination discovery to fairness-aware data mining
- Healthcare Data Mining with Matrix Models
- IoT Big Data Stream Mining
- Lifelong Machine Learning and Computer Reading the Web
- Exploring and Mining Events Using Large Scale Social Media Data: Challenges in Extraction, Modeling and Analyzing
- Deep Learning

**Hands-On Tutorials**

- Introduction to Spark 2.0
- MXNet: Multi-Language Machine Learning Library
- Streaming Analytics
- Amazon Web Services (AWS)
- Computational Network Toolkit (CNTK)
- Scalable R on Spark
- Big Natural Language Data Processing
- Recommender Systems
Workshops

Full-day Workshops:
- KDD Cup 2016: Towards Measuring the Impact of Research Institutions
- Workshop 1: Mining and Learning with Graphs
- Workshop 2: Large-Scale Sports Analytics
- Workshop 3: Data Science for Food, Energy and Water
- Workshop 4: Mining and Learning from Time Series
- Workshop 5: The 5th International Workshop on Urban Computing
- Workshop 6: Interactive Data Exploration and Analytics

Half-day, Morning Workshops:
- Workshop 7: Machine Learning for Large Scale Transportation Systems
- Workshop 8: Large-scale Deep Learning for Data Mining
- Workshop 10: Enterprise Intelligence
- Workshop 11: Workshop on Issues of Sentiment Discovery and Opinion Mining

Half-day, Afternoon Workshops:
- Workshop 12: Outlier Definition, Detection, and Description On-Demand
- Workshop 13: Workshop on Causal Discovery
- Workshop 14: Machine Learning Meets Fashion: Data, Algorithms and Analytics for the Fashion Industry
- Workshop 15: Machine Learning for Prognostics and Health Management
- Workshop 16: 15th International Workshop on Data Mining in Bioinformatics
KDD 2016 Tutorial Program

Saturday August 13, 2016

8:00am-12:00pm Tutorial 12 - Plaza Room A/B
Title: Enabling the Discovery of Reliable Information from Passively and Actively Crowdsourced Data
Instructors: Jing Gao (SUNY Buffalo); Qi Li (SUNY Buffalo); Bo Zhao (LinkedIn); Wei Fan (Baidu); Jiawei Han (UIUC)

8:00am-12:00pm Tutorial 10 - Imperial Ballroom A & B
Title: Business Applications of Predictive Modeling at Scale
Instructors: Qiang Zhu (LinkedIn); Songtao Guo (LinkedIn); Paul Ogilvie (LinkedIn); Yan Liu (LinkedIn)

8:00am-12:00pm Tutorial 3 - Continental 1-4
Title: Exploring and Mining Events Using Large Scale Social Media Data: Challenges in Extraction, Modeling and Analyzing
Instructors: Yuheng Hu (UIC), Yu-Ru Lin (Univ. of Pittsburgh), Jiebo Luo (University of Rochester)

8:00am-12:00pm Tutorial 7 - Continental 6-8
Title: Algorithmic bias: From Discrimination Discovery to Fairness-Aware Data Mining
Instructors: Sara Hajian (Eurecat); Francesco Bonchi (ISI Foundation); Carlos Castillo (Eurecat)

8:00am-12:00pm Tutorial 8 - Grand Ballroom Salon A
Title: Extracting Optimal Performance From Dynamic Time Warping
Instructors: Abdullah Mueen (University of New Mexico); Eamonn Keogh (University of California Riverside)

1:00pm-5:00pm Tutorial 11 - Plaza Room A/B
Title: Leveraging Propagation for Data Mining: Models, Algorithms and Applications
Instructors: B. Aditya Prakash (Virginia Tech); Naren Ramakrishnan (Virginia Tech)

1:00pm-5:00pm Tutorial 9 - Imperial Ballroom A & B
Title: Scalable Learning of Graphical Models
Instructors: Francois Petitjean (Monash University); Geoffrey I. Webb (Monash University)

1:00pm-5:00pm Tutorial 4 - Continental 1-4
Title: Lifelong Machine Learning and Computer Reading the Web
Instructors: Zhiyuan Chen (Google); Estevam R. Hruschka Jr. (University of São Carlos and Carnegie Mellon University); Bing Liu (University of Illinois at Chicago)
1:00pm-5:00pm **Tutorial 6** - Continental 5  
**Title:** Healthcare Data Mining with Matrix Models  
**Instructors:** Fei Wang (University of Connecticut); Ping Zhang (IBM); Joel Dudley (Icahn School of Medicine at Mount Sinai)

1:00pm-5:00pm **Tutorial 5** - Continental 6-8  
**Title:** IoT Big Data Stream Mining  
**Instructors:** Gianmarco De Francisci Morales (OCRI); Albert Bifet (University of Waikato); Latifur Khan (UT Dallas), Joao Gama (University of Porto); Wei Fan (Big Data)

1:00pm-5:00pm **Tutorial 2** - Grand Ballroom Salon A  
**Title:** Deep Learning  
**Instructors:** Ruslan Salakhutdinov (Carnegie Mellon University)
KDD 2016 Workshop Program

Sunday August 14, 2016

Full Day Workshops 8:00am-5:00pm

<table>
<thead>
<tr>
<th>KDD Cup: Towards Measuring the Impact of Research Institutions</th>
<th>Union Square 23/24</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://kddcup2016.azurewebsites.net/">https://kddcup2016.azurewebsites.net/</a></td>
<td></td>
</tr>
<tr>
<td>Organizers:</td>
<td></td>
</tr>
<tr>
<td>Hao Ma (Microsoft)</td>
<td></td>
</tr>
<tr>
<td>Kuansan Wang (Microsoft)</td>
<td></td>
</tr>
<tr>
<td>Jie Tang (Tsinghua University)</td>
<td></td>
</tr>
</tbody>
</table>

Agenda: TBD (Check website for latest agenda).

Workshop 1: Mining and Learning with Graphs

<table>
<thead>
<tr>
<th>Workshop 1: Mining and Learning with Graphs</th>
<th>Plaza Room A/B</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.mlgworkshop.org/2016/">http://www.mlgworkshop.org/2016/</a></td>
<td></td>
</tr>
<tr>
<td>Organizers:</td>
<td></td>
</tr>
<tr>
<td>Shobeir Fakhraei (University of Maryland College Park)</td>
<td></td>
</tr>
<tr>
<td>Lise Getoor (UC Santa Cruz)</td>
<td></td>
</tr>
<tr>
<td>Danai Koutra (University of Michigan)</td>
<td></td>
</tr>
<tr>
<td>Julian McAuley (University of California, San Diego)</td>
<td></td>
</tr>
<tr>
<td>Sean Taylor (Facebook)</td>
<td></td>
</tr>
</tbody>
</table>

Agenda:

- 8:50 - 9:00: Opening remarks
- 9:00 - 9:40: Lars Backstrom
- 9:40 - 10:00: Spotlights 1
- 10:00 - 10:30: Coffee
- 10:30 - 10:50: Spotlights 2
- 10:50 - 11:30: Leman Akoglu
- 11:30 - 12:00: Posters
- 12:00: Lunch (+ posters)
- 13:10 - 13:50: Tamara Kolda
- 13:50 - 14:10: Contributed talks 1
- 14:10 - 14:50: Yizhou Sun
- 14:50 - 15:10: Contributed talks 2
15:10 - 15:40: Coffee
15:40 - 16:20: Jennifer Neville
16:20 - 17:00: SVN Vishwanathan
17:00: Closing remarks

Workshop 2: Large-Scale Sports Analytics
http://www.large-scale-sports-analytics.org/
Organizers:
Patrick Lucey (STATS)
Yisong Yue (Caltech)
Jenna Wiens (University of Michigan)
Stuart Morgan (Australian Institute of Sport)

Agenda:
- 8:15 - 8:20 Opening Remarks
- 8:20 - 9:00 Invited Talk by Peter Carr (Disney Research)
  - Automated Sports Broadcasting
- 9:00 - 9:40 Invited Talk by Sarah Rudd (StatDNA)
  - Making Strides in Quantifying and Understanding Soccer
- 9:40 - 10:00 Poster Spotlight 1
- 10:00 - 10:40 Coffee Break and Poster Session 1
- 10:40 - 11:20 Invited Talk by Irfan Essa (Georgia Tech)
  - Computational Video for Sports: Challenges for Large Scale Data Analysis
- 11:20 - 12:00 Invited Talk by Jesse Paquette (tag.bio)
  - Making an Idea Machine: Modular Architecture for a Scaleable Exploratory Data Analysis Platform in Genomics, Sports and Beyond
- 12:00 - 13:30 Lunch Break
- 13:30 - 14:10 Chaitanya Chemudugunta (Blizzard Entertainment)
  - Overview of Data Science challenges at Blizzard Entertainment
- 14:10 - 14:50 Invited Talk by Joel Brooks (MIT)
  - Recognizing and Analyzing Ball Screen Defense in the NBA
- 14:50 - 15:10 Poster Spotlight 2
- 15:10 - 16:00 Coffee Break and Poster Session 2
- 16:00 - 16:50 Panel Discussion
  - AI in Sport: What does it mean and how will it change sport in the future?
- 16:50 - 17:00: Final Remarks
Workshop 3: Data Science for Food, Energy and Water
https://sites.google.com/site/2016dsfew/

Organizers:
Naoki Abe (IBM Research)
Chid Apte (IBM Research)
Vipin Kumar (University of Minnesota)
Shashi Shekhar (University of Minnesota)
Mitch Tuinstra (Purdue University)
Ranga Raju Vatsavai (NC State University)

Agenda:
● 8:00 - 8:15 Greetings and workshop introduction
● 8:15 - 9:00 Invited Talk 1
● 9:00 - 10:00 Paper presentation session 1
● 10:00 - 10:30 Coffee break & poster session
● 10:30 - 11:15 Paper presentation session 2
● 11:15 - 12:00 Invited Talk 2
● 12:00 - 13:00 Lunch
● 13:00 - 14:00 Paper presentation session 3
● 14:00 - 14:45 Invited Talk 3
● 14:45 - 15:30 Panel discussion
● 15:30 - 16:00 Coffee break and poster session
● 16:00 - 16:45 Paper presentation session 4
● 16:45 - 17:00 Closing

Workshop 4: Mining and Learning from Time Series
http://www-bcf.usc.edu/~liu32/milets16/

Organizers:
Eamonn Keogh (University of California Riverside)
Yan Liu (University of Southern California)
Abdullah Mueen (University of New Mexico)
Dehua Cheng (University of Southern California)

Agenda:
● 8:00 - 9:00 : Keynote 1
● 9:00 - 10:00: Session 1
  ○ The Great Time Series Classification Bake Off: An Experimental Evaluation of Recently Proposed Algorithms Anthony Bagnall, Aaron Bostrom, James Large and Jason Lines
  ○ On the Effect of Endpoints on Dynamic Time Warping Diego Silva, Gustavo Batista and Eamonn Keogh
● Sparse plus low-rank graphical models of time series for functional connectivity in MEG
  Nicholas J. Foti, Rahul Nadkarni, Adrian Kc Lee and Emily B. Fox
● 10:00 - 10:30: Coffee break
● 10:30 - 11:50: Session 2
  ○ Parallel News-Article Traffic Forecasting with ADMM Stratis Ioannidis, Yunjiang Jiang, Saeed Amizadeh and Nikolay Laptev
  ○ Using Time Series Techniques to Forecast and Analyze Wake and Sleep Behavior Jennifer A. Williams and Diane J. Cook
  ○ Scalable Clustering of Correlated Time Series using Expectation Propagation Christopher Aicher and Emily B. Fox
  ○ Time Lag Concerned Dynamic Dependency Network Structure Learning Sizhen Du, Haikun Hong and Guojie Song
● 12:00 - 13:00: Lunch
● 13:00 - 14:00: Keynote 2 (50 minute talk, 10 minutes questions)
● 14:00 - 15:20: Session 3 (20 minutes each)
  ○ Evaluating Improvements to the Shapelet Transform Aaron Bostrom, Anthony Bagnall and Jason Lines
  ○ Space-Time Clustering with Stability Probe while Riding Downhill Xin Huang, Iliyan Iliev, Alexander Brenning and Yulia Gel
  ○ Short-term Time Series Forecasting with Regression Automata Qin Lin, Christian Hammerschmidt, Gaetano Pellegrino and Sicco Verwer
  ○ Granger Causality Networks for Categorical Time Series Alex Tank, Emily Fox and Ali Shojaie
● 15:30 - 16:00: Coffee Break
● 16:00 - 17:00: Panel Discussion and Poster Session

Workshop 5: The 5th International Workshop on Urban Computing
http://www2.cs.uic.edu/~urbcomp2013/urbcomp2016/
Organizers:
Yu Zheng (Microsoft Research, China)
Zhenhui Li (Penn State University)
Alexandre Bayen (U.C. Berkeley)
Ouri E. Wolfson (University of Illinois at Chicago)
Qiang Yang Hong (Kong University of Science and Technology)
Claudio T. Silva (New York University, USA)

Agenda:
● 8:00 - 8:05 Opening and Welcome
● 8:05 - 8:55 Keynote
● 8:55 - 10:00 Session 1: Urban Traffic
Short Byproducts of Urban Infrastructure Interfaces: Evidence from Parking Compliance. Konstantinos Pelechrinis

An Active Learning Framework Incorporating User Input For Mining Urban Data

Short Robustness and Resilience of cities around the world. Sofiane Abbar, Javier Borge Holthoefer, Tahar Zanouda

Short Exploring Foursquare-derived features for crime prediction in New York City, Cristina Kadar

Short Recognizing Cities from Google Street View - Computer Vision Analysis of Urban Visual Identity. Lezhi Li, Kang Qi, Ian Lee, Jingwen Wei

10:00 - 10:30 Coffee break

10:30 - 12:00 Session 2: Urban Development
  Invited Talk
  Reducing Inefficiencies in Taxi Systems, Chenguang Zhu
  Monitoring Manhattan's traffic from 5 cameras? Siheng Chen, Yaoqing Yang, Jelena Kovacevic, Christos Faloutsos

12:00 - 13:30 Lunch

13:30 - 15:30 Session 3: Trajectory Data Mining
  Invited Talk
  Personalised Recommendations for Modes of Transport: A Sequence-based Approach. Gunjan Kumar, Houssem Jerbi, Michael O'Mahony
  Short Estimating Evacuation Hotspots using GPS data: What happened after the large earthquakes in Kumamoto, Japan?
  Short Analyzing Interlinked Urban Functions through Network Motifs Using Smart-Card Data
  On the Dominant Role of Returners’ Human Mobility Networks on Urban Energy Consumption. John E Taylor

15:30 - 16:00 Coffee Break

16:00 - 17:20 Session 4: Urban Energy
  Invited Talk
  CityBES: A Web-based Platform to Support City-Scale Building Energy Efficiency, Tianzhen Hong
  Disaggregating Appliance-Level Energy Consumption: A Probabilistic Framework, Sabina Tomkins

17:20 - 17:30 Closing
Workshop 6: Interactive Data Exploration and Analytics (IDEA)
http://poloclub.gatech.edu/idea2016/

Organizers:
Polo Chau (Georgia Tech)
Jilles Vreeken (Max Planck Institute for Informatics, Saarland University)
Matthijs van Leeuwen (Universiteit Leiden)
Dafna Shahaf (Microsoft Research)
Christos Faloutsos (Carnegie Mellon University)

Embarcadero - Parc 55

Agenda:
● 8:15 - 8:30 Welcome to IDEA’16
● 8:30 - 9:20 Invited Talk by Jerome F. Friedman (Stanford University)
  ○ Regression Location and Scale Estimation with Application to Censoring
● 9:20 - 10:00 Contributed session 1
● 10:00 - 10:30 Coffee break
● 10:30 - 11:10 Contributed session 2
  ○ Akash Srivastava, James Zou and Charles Sutton: Clustering with a Reject Option: Interactive Clustering as Bayesian Prior Elicitation.
● 11:10 - 12:00 Invited Talk by Jeffrey Heer (University of Washington, Trifacta)
  ○ Predictive Interaction
● 12:00 - 13:00 Lunch break
● 13:00 - 13:45 Invited Talk by Saleema Amershi (Microsoft Research)
  ○ Towards Usable Machine Learning
● 13:45 - 14:45 Contributed session 3
● 14:30 - 14:45 Invited Demo by Andriy Kashcha
  ○ Visualization of Large Graphs on the Web
● 14:45 - 15:30 Invited Talk by Eamonn Keogh (UC Riverside)
  ○ At Last! Time Series Joins, Motifs, Discords and Shapelets at Interactive Speeds
● 15:30 - 16:00 Coffee break + poster & demo & networking session
● 16:00 - 17:00 Poster and Demo and Networking session
Half Day Workshops 8:00am-12:00pm

Workshop 7: Machine Learning for Large Scale Transportation Systems
http://www.lstskdd.com/
Organizers:
Kevin Novak (Uber)
George Mohler (Metromile)
Chetan Ramaiah (Metromile)
Allen Tran (Metromile)

Agenda:
● 8:00 - 8:25 Contributed Talk
  ○ Alexander Chen, Hsing-Kuo Pao, and Yuh-Jye Lee: Online Traffic Speed Forecasting with Multi-Periodicity Models
● 8:25 - 8:50 Contributed Talk
  ○ Mogeng Yin, Madeleine Sheehan, Danqing Zhang, Jean-Francois Paiement, and Alexey Pozdnukhov: A Generative Model of Urban Activities from Cellular Data
● 8:50 - 9:15 Contributed Talk
  ○ Hussein Dia, Farid Javanshour, and Jack Hil: Network Impacts of Autonomous Shared Mobility
● 9:15 - 10:00 Invited Talk by Jeff Schneider
  ○ Bayesian optimization and its applications for autonomous vehicles
● 10:00 - 10:30 Coffee Break
● 10:30 - 10:55 Contributed Talk
  ○ Chetan Ramaiah, Allen Tran, Evan Cox and George Mohler: Deep learning for driving detection on mobile phones
● 10:55 - 11:20 Contributed Talk
  ○ Ming Zeng, Tong Yu, Xiao Wang, Vincent Su, Le T. Nguyen, and Ole J. Mengshoel: Improving Demand Prediction in Bike Sharing System by Learning Global Features
● 11:30 - 12:00 Panel

Workshop 8: Workshop on Large-scale Deep Learning for Data Mining
Organizers:
Xuewen Chen (Wayne State University)
Li Deng (Microsoft Research)
Taghi Khoshgoftaar (Florida Atlantic University)

Franciscan C/D

Cyril Magnin - Parc 55
Agenda:

- 8:00 - 8:05 Opening remarks
- 8:05 - 8:50 Keynote by Li Deng (Chief Scientist for AI Microsoft Research and ASG)
  - Question, and Answers Two Types of Big Data and Three Styles of Deep Learning for AI Applications
- 8:50 - 8:55 Talk preparation
- 8:55 - 10:20 Oral presentation session
  - Leveraging Multi-Layer Deep Features for Large-Scale Visual Recognition, Tianyi Zhao, and Jianping Fan
  - Learning Multi-Layer Coarse-to-Fine Representations for Over 10,000 Image Categories, Ji Zhang, Kuizhi Mei, and Jianping Fan
  - Film2Vec --- A Feature-based Film Distributed Representation for Rating Prediction, Xanh Ho, Thuy Vu and Nhung Nguyen
  - Long-term face tracking in the wild using deep learning, Kunlei Zhang, Elaheh Barati, Elaheh Rashedi, and Xuewen Chen
  - Applying Deep Learning to Improve Maritime Situational Awareness, Kathy Tang, and David Crandall
- 10:20 - 10:30 Coffee Break
- 10:30 - 12:00 Oral presentation session
  - Improving Deep Neural Network Design for New Text Data Representations, Joseph Prusa, and Taghi Khoshgoftaar
  - Contextual LSTM (CLSTM) models for Large-scale NLP tasks, Shalini Ghosh, Oriol Vinyals, Brian Strope, Scott Roy, Tom Dean, and Larry Heck
  - Effective Auto-Encoder for Unsupervised Sparse Representation, Faria Mahnaz, Melih Aslan, and Xue-Wen Chen
  - Deep Learning for Chemical Compound Stability Prediction, Ruoqian Liu, Logan Ward, Chris Wolverton, Ankit Agrawal, Wei-Keng Liao, and Alok Choudhary
  - Deep Learning for Financial Sentiment Analysis, Sahar Sohangir, Dingding Wang and Anna Pomeranets
- 12:00 - 13:00 Lunch Break
Workshop 9: 5th International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications
http://bigmine.github.io/bigmine16/

Organizers:
Wei Fan (Baidu Research Big Data Lab)
Albert Bifet (Telecom-ParisTech)
Jesse Read (Telecom-ParisTech)
E-mail: jesse.read at telecom-paristech.fr
Qiang Yang (Hong Kong University of Science and Technology)
Philip Yu (University of Illinois at Chicago)

Agenda:
● 8:00 - 8:05 Opening Remarks
● 8:05 - 8:40 Invited Talk by Hanghang Tong
  ○ Inside the Atoms: Mining a Network of Networks and Beyond
● 8:40 - 8:55 Mudhar Bin Rabieah and Christos-Savvas Bouganis
  ○ FPGASVM: A Framework for Accelerating Kernelized Support Vector Machine Training using FPGAs
● 8:55 - 9:10 Hongxia Yang, Quan Lu, Angus Xianen Qiu and Chun Han
  ○ Large Scale CVR Prediction through Dynamic Transfer Learning of Global and Local Features
● 9:10 - 9:25 Qiang Ma, Musen Wen, Zhen Xia and Datong Chen
  ○ A Sublinear, Massive-scale Look-alike Audience Extension System
● 9:25 - 10:00 Invited Talk by Joseph Bradley
  ○ Foundations for Scaling ML in Apache Spark
● 10:00 - 10:30 Morning Coffee Break
● 10:30 - 11:05 Invited Talk by Charles Elkan
  ○ From Practice to Theory in Learning from Massive Data
● 11:05 - 11:20 Md Abdul Kader, Arnold P. Boedihardjo, Sheikh Motahar Naim and M. Shahrriar Hossain
  ○ Contextual embedding for distributed representations of entities in a text corpus
● 11:20 - 11:35 Manan Shah
  ○ Disease Propagation in Social Networks: A Novel Study of Infection Genesis and Spread on Twitter
● 11:35 - 11:50 Nesreen Ahmed, Ted Willke and Ryan Rossi
  ○ Exact and Estimation of Local Edge-centric Graphlet Counts
● 11:50 - 12:05 Harish S. Bhat, R. W. M. A. Madushani and Shagun Rawat
  ○ Scalable SDE Filtering and Inference with Apache Spark
● 12:05 - 12:10 Concluding Remarks
Workshop 10: Enterprise Intelligence
http://enterpriserelevance.com/kdd2016/

Organizers:
Abhishek Gupta (LinkedIn)
George Karypis (University of Minnesota, Twin Cities)

Agenda:
● 8:00 - 9:00 Invited Talk by Daniel Tunkelang
● 9:00 - 10:00 Panel Discussion
● 10:00 - 10:30 Coffee Break
● 10.30 - 11:00 Invited Talk by Aditya Parmeswaram
● 11:00- 11.30 Five-minute Paper Pitch
● 11.30 - 12:00 Poster Session

Workshop 11: Workshop on Issues of Sentiment Discovery and Opinion Mining
http://sentic.net/wisdom/

Organizers:
Yongzheng Zhang, LinkedIn (USA)
Erik Cambria, Nanyang Technological University (Singapore)
Yunqing Xia, Microsoft Research Asia (China)
Bing Liu, University of Illinois at Chicago (USA)

Agenda:
● 8:00 – 8:10 Opening Remarks
● 8:10 – 9:00 Keynote talk by Jie Tang (Tsinghua University): Social Influence and Sentiment Analysis
● 9:00 – 9:20 Adithya Rao and Nemanja Spasojevic (Lithium Technologies | Klout): Actionable and Political Text Classification using Word Embeddings and LSTM
● 9:20 – 9:40 N. Vedula, S. Parthasarathy, and V. Shalin (Ohio State University, Wright State University): Predicting Trust Relations Among Users in a Social Network: On the role of Influence, Cohesion and Valence
● 9:40 – 10:00 Arindam Paul, Ankit Agrawal, Wei-keng Liao, and Alok Choudhary (Northwestern University): AnonyMine: Mining anonymous social media posts using psycho-lingual and crowd-sourced dictionaries
● 10:00 – 10:30 Coffee Break
● 10:30 – 11:20 Keynote talk by Vita Markman (Linkedin Corporation)
  ○ Obtaining Quality Labeled Data for Opinion Mining in Short Text

• 11:40 – 12:00 Himanshu Shekhar, Shankar Setty, Uma Mudenagudi, and B.V. Bhoomaraddi (College of Engineering and Technology, Hubli-India): Vehicular Traffic Analysis from Social Media Data

Half Day Workshops 1:00pm-5:00pm

<table>
<thead>
<tr>
<th>Workshop 12: Outlier Definition, Detection, and Description on Demand</th>
<th>Cyril Magnin - Parc 55</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://outlier-analytics.org/odd16kdd">http://outlier-analytics.org/odd16kdd</a></td>
<td></td>
</tr>
<tr>
<td><strong>Organizers:</strong></td>
<td></td>
</tr>
<tr>
<td>Leman Akoglu (Stony Brook University)</td>
<td></td>
</tr>
<tr>
<td>Franziska Bell (Uber)</td>
<td></td>
</tr>
<tr>
<td>Emmanuel Müller (Hasso-Plattner-Institute)</td>
<td></td>
</tr>
<tr>
<td>Ted E. Senator</td>
<td></td>
</tr>
</tbody>
</table>

**Agenda:**

• 13:15 - 13:30 Welcome and Remarks
• 13:30 - 14:30 Invited Talk by Jeff Schneider
• 14:30 - 15:30 Contributed Session
  ▪ Charmgil Hong, Rumi Ghosh and Soundar Srinivasan: Dealing with Class Imbalance using Thresholding
  ▪ Timothy La Fond, Jennifer Neville and Brian Gallagher: Generating Local Explanations of Network Anomalies via Score Decomposition
  ▪ Shalmoli Gupta: Fast and Accurate K-means Clustering with Outliers
• 15:30 - 16:00 Coffee break
• 16:00 - 17:00 Contributed Session
  ▪ Andrey Lokhov, Nathan Lemons, Thomas McAndrew, Aric Hagberg and Scott Backhaus: Detection of Cyber-Physical Faults and Intrusions from Physical Correlations
  ▪ Roel Bertens, Jilles Vreeken and Arno Siebes: Beauty and Brains: Detecting Anomalous Pattern Co-Occurrences
• 17:00 Closing
**Workshop 13: Workshop on Causal Discovery**


**Organizers:**
- Jiuyong Li (University of South Australia)
- Kun Zhang (Carnegie Mellon University)
- Elias Bareinboim (Purdue University)
- Lin Liu (University of South Australia)

**Agenda:**
- 13:00 – 13:05 Opening
- 13:05 – 14:05 Invited Talk by Frederick Eberhardt
  - Foundations of Causal Discovery
- 14:05 – 14:25 Dan Garant and David Jensen
  - Evaluating Causal Models by Comparing Interventional Distributions
- 14:25 – 14:45 Elena Sokolova, Daniel von Rhein, Jill Naaijen, Perry Groot, Tom Claassen, Jan Buitelaar and Tom Heskes
  - Handling hybrid and missing data in constraint-based causal discovery to study the etiology of ADHD
- 14:45 – 15:05 Gilles Blondel, Marta Arias and Ricard Gavaldà
  - Identifiability and Transportability in Dynamic Causal Networks
- 15:05 – 15:25 Zhalama, Jiji Zhang, and Wolfgang Mayer
  - Weakening Faithfulness: Some Heuristic Causal Discovery Algorithms
- 15:25 – 15:30 Discussions
- 15:30 – 16:00 Coffee break
- 16:00 – 16:20 Dhanya Sridhar and Lise Getoor
  - Joint Probabilistic Inference of Causal Structure
- 16:20 – 16:40 Alex Tank, Emily Fox and Ali Shojaie
  - Identifiability of Non-Gaussian Structural VAR Models for Subsampled and Mixed Frequency Time Series
- 16:40 – 16:55 Discussions
- 16:55 – 17:00 Closing
### Workshop 14: Machine Learning Meets Fashion: Data, Algorithms and Analytics for the Fashion Industry

**http://kddfashion2016.mybluemix.net/**

**Organizers:**
- Vikas C. Raykar (IBM Research)
- Brad Klingenberg (Director of Styling Algorithms, Stitch Fix)
- Heng Xu (Pennsylvania State University)
- Raghavendra Singh (IBM Research)
- Amrita Saha (IBM Research)

**Agenda:**
- 13:00-13:05 - Introduction and Agenda
- 13:05-13:30 - Invited Talk 1
- 13:30-14:30 - Oral Paper Presentations
- 14:30-15:00 - Invited Talk 2
- 15:00-15:30 - Invited Talk 3
- 15:30-16:10 - Coffee break & Poster Session
- 16:10-16:50 - Oral Paper Presentations
- 16:50-17:00 - Open House (5 min talks/demos)

### Workshop 15: 1st ACM SIGKDD Workshop on Machine Learning for Prognostics and Health Management

**https://sites.google.com/site/mlforphm2016/**

**Organizers:**
- Michael Giering (United Technologies Research Center)
- Kishore Reddy (United Technologies Research Center)
- Soumalya Sarkar (United Technologies Research Center)
- Soumik Sarkar (Iowa State University)
- Madhu Shashanka (Charles Schwab)
- Abhishek Srivastav (GE Global Research)

**Agenda:**
- 13:00 - 13:45 Invited Talk by Ashok Srivastava (Chief Data Scientist, Verizon)
- 13:45 - 14:45 Regular session 1
  - Sequential Hypothesis Tests Using Markov Models of Time-Series Data
  - Prognostics of Combustion Instabilities from Hi-speed Flame Video using A Deep Convolutional Selective Autoencoder
  - A Framework of Combining Deep Learning and Survival Analysis for Asset Health Management
Applying Deep Learning for Prognostic Health Monitoring of Aerospace and Building Systems

- 14:45 - 15:00 Coffee break
- 15:00 - 16:00 Regular Session 2
  - Multi-Sensor Prognostics using an Unsupervised Health Index based on an LSTM Encoder-Decoder
  - A Classifier Development Process for Mechanical Health
  - Temporal Learning in Video Data Using Deep Learning and Gaussian Processes
  - Activity Recognition in Prognostics and Health Monitoring (PHM) Related Service Environment from Electroencephalography (EEG) via Deep Learning

- 16:00 - 16:10 PM Break
- 16:10 - 17:00 Panel with Achalesh Pandey (GE Global Research), Dragos Margineantu (Boeing), and Michael Giering (United Technologies Research Center)

Workshop 16: 15th International Workshop on Data Mining in Bioinformatics
http://home.biokdd.org/biokdd16/

Organizers:
Sara C. Madeira (Universidade de Lisboa)
Chandan K. Reddy (Virginia Tech)
Mohammed J. Zaki (Rensselaer Polytechnic Institute)
Jake Y. Chen (University of Alabama at Birmingham)

| Union Square | 22 |

Agenda: TBD (Check website for latest agenda).
KDD 2016 Hands-On Tutorial Program

Monday August 15, 2016

9:00AM-12:00PM Hands-On Tutorial 1 - Market - Parc 55
Title: Recommender Systems
Instructors: Bee-Chung Chen (LinkedIn); Deepak Agarwal (LinkedIn)

9:00AM-12:00PM Hands-On Tutorial 2 - Embarcadero - Parc 55
Title: Big Natural Language Data Processing
Instructors: Gabor Melli (OpenGov, Inc.); Matt Seal (OpenGov, Inc.)

1:00PM-5:00PM Hands-On Tutorial 3 - Market - Parc 55
Title: MXNet: Multi-Language Machine Learning Library
Instructors: Mu Li (Carnegie Mellon University); Tianqi Chen (UW)

1:00PM-5:00PM Hands-On Tutorial 4 - Embarcadero - Parc 55
Title: Spark 2.0
Instructors: Michael Armbrust (Databricks); Doug Bateman (Databricks); Reynold Xin (Databricks); Matei Zaharia (Databricks)

Tuesday August 16, 2016

9:00AM-12:00PM Hands-On Tutorial 5 - Market - Parc 55
Title: Amazon Web Services (AWS)
Instructors: Manjeet Chayel, Keith Steward (Amazon)

9:00AM-12:00PM Hands-On Tutorial 6 - Embarcadero - Parc 55
Title: Computational Network Toolkit (CNTK)
Instructors: Frank Seide (Microsoft)

Wednesday August 17, 2016

9:00AM-12:00PM Hands-On Tutorial 7 - Market - Parc 55
Title: Streaming Analytics
Instructors: Ashish Gupta (LinkedIn)

9:00AM-12:00PM Hands-On Tutorial 8 - Embarcadero - Parc 55
Title: R on Spark
Instructors: Mario Inchiosa (Microsoft)
KDD 2016 Conference Program

Sunday August 14, 2016 Detailed Program

Sunday August 14, 2016 5:30pm – 7:00pm, KDD 2016 Opening and Awards Ceremony - Grand Ballroom

Sunday August 14, 6:00pm - 7:00pm, Poster Set-Up (Research Track, Oral Papers) – Golden Gate

Sunday August 14, 7:00pm - 12:00am, Poster Session: Research Track (Oral Papers) – Golden Gate

Skinny-dip: Clustering in a Sea of Noise
Samuel Maurus (Helmholtz Zentrum München & Technische Universität München), Claudia Plant (University of Vienna)

Lexis: An Optimization Framework for Discovering the Hierarchical Structure of Sequential Data
Payam Siyari (Georgia Institute of Technology), Bistra Dilkina (Georgia Institute of Technology), Constantine Dovrolis (Georgia Institute of Technology)

NetCycle: Collective Evolution Inference in Heterogeneous Information Networks
Yizhou Zhang (Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University), Yun Xiong (Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University), Xiangnan Kong (Atwater Kent Lab, Department of Computer Science, Worcester Polytechnic Institute), Yangyong Zhu (Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University)

Targeted Topic Modeling for Focused Analysis
Shuai Wang (University of Illinois at Chicago), Zhiyuan Chen (University of Illinois at Chicago), Geli Fei (University of Illinois at Chicago), Bing Liu (University of Illinois at Chicago), Sherry Emery (University of Illinois at Chicago)

Positive-Unlabeled Learning in Streaming Networks
Shiyu Chang (University of Illinois at Urbana-Champaign), Yang Zhang (University of Illinois at Urbana-Champaign), Jiliang Tang (Michigan State University), Dawei Yin (Yahoo! Inc.), Yi Chang (Yahoo! Inc.), Mark A Hasegawa-Johnson (University of Illinois at Urbana-Champaign), Thomas S Huang (University of Illinois at Urbana-Champaign)

Robust Influence Maximization
Wei Chen (Microsoft Research), Tian Lin (Tsinghua University), Zihan Tan (Tsinghua University), Mingfei Zhao (Tsinghua University), Xuren Zhou (The Hong Kong University of Science and Technology)

Towards Conversational Recommender Systems
Konstantina Christakopoulou (University of Minnesota), Filip Radlinski (Microsoft), Katja Hofmann (Microsoft)

FRAUDAR: Bounding Graph Fraud in the Face of Camouflage
Bryan Hooi (Carnegie Mellon University), Hyun Ah Song (Carnegie Mellon University), Alex Beutel (Carnegie Mellon University), Neil Shah (Carnegie Mellon University), Kijung Shin (Carnegie Mellon University), Christos Faloutsos (Carnegie Mellon University)
Dynamic Clustering of Streaming Short Documents
Shangsong Liang (University College London), Emine Yilmaz (University College London), Evangelos Kanoulas (University of Amsterdam)

CatchTartan: Representing and Summarizing Dynamic Multicontextual Behaviors
Meng Jiang (University of Illinois at Urbana-Champaign), Christos Faloutsos (Carnegie Mellon University), Jiawei Han (University of Illinois at Urbana-Champaign)

AnyDBC: An Efficient Anytime Density-based Clustering Algorithm for Very Large Complex Datasets
Son T. Mai (Aarhus University), Ira Assent (Aarhus University), Martin Storgaard (Aarhus University)

Asymmetric Transitivity Preserving Graph Embedding
Mingdong Ou (Tsinghua University), Peng Cui (Tsinghua University), Jian Pei (Simon Fraser University), Ziwei Zhang (Tsinghua University), Wenwu Zhu (Tsinghua University)

Mining Subgroups with Exceptional Transition Behavior
Florian Lemmerich (GESIS - Leibniz Institute for the Social Sciences & University of Koblenz-Landau), Martin Becker (University of Würzburg), Philipp Singer (GESIS - Leibniz Institute for the Social Sciences & University of Koblenz-Landau), Denis Helic (TU Graz), Andreas Hoito (University of Würzburg), Markus Strohmaier (GESIS - Leibniz Institute for the Social Sciences & University of Koblenz-Landau)

Structural Deep Network Embedding
Daixin Wang (Tsinghua University), Peng Cui (Tsinghua University), Wenwu Zhu (Tsinghua University)

node2vec: Scalable Feature Learning for Networks
Aditya Grover (Stanford University), Jure Leskovec (Stanford University)

Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences
Yasuko Matsubara (Kumamoto University), Yasushi Sakurai (Kumamoto University)

Extreme Multi-label Loss Functions for Recommendation, Tagging, Ranking & Other Missing Label Applications
Himanshu Jain (Indian Institute of Technology Delhi), Yashoteja Prabhu (Indian Institute of Technology Delhi), Manik Varma (Microsoft Research)

PTE: Enumerating Trillion Triangles On Distributed Systems
Ha-Myung Park (KAIST), Sung-Hyon Myaeng (KAIST), U Kang (Seoul National University)

When Social Influence Meets Item Inference
Hui-Ju Hung (The Pennsylvania State University), Hong-Han Shuai (Academia Sinica), De-Nian Yang (Academia Sinica), Liang-Hao Huang (Academia Sinica), Wang-Chien Lee (The Pennsylvania State University), Jian Pei (Simon Fraser University), Ming-Syan Chen (National Taiwan University)

DeepIntent: Learning Attentions for Online Advertising with Recurrent Neural Networks
Shuangfei Zhai (Binghamton University), Keng-hao Chang (Microsoft), Ruofei Zhang (Microsoft), Zhongfei Mark Zhang (Microsoft)

Robust Influence Maximization
Xinran He (University of Southern California), David Kempe (University of Southern California)
Keeping it Short and Simple: Summarising Complex Event Sequences with Multivariate Patterns  
Roel Bertens (Utrecht University), Jilles Vreeken (Max Planck Institute for Informatics and Saarland University), Arno Siebes (Utrecht University)

Compact and Scalable Graph Neighborhood Sketching  
Takuya Akiba (National Institute of Informatics), Yosuke Yano (National Institute of Informatics)

Accelerating Online CP Decompositions for Higher Order Tensors  
Shuo Zhou (The University of Melbourne), Nguyen Xuan Vinh (The University of Melbourne), James Bailey (The University of Melbourne), Yunzhe Jia (The University of Melbourne), Ian Davidson (The University of California - Davis)

Predicting Matchups and Preferences in Context  
Shuo Chen (Cornell University), Thorsten Joachims (Cornell University)

Ranking Causal Anomalies via Temporal and Dynamical Analysis on Vanishing Correlations  
Wei Cheng (NEC Laboratories America), Kai Zhang (NEC Laboratories America), Haifeng Chen (NEC Laboratories America), Guofei Jiang (NEC Laboratories America), Zhengzhang Chen (NEC Laboratories America), Wei Wang (UCLA)

FINAL: Fast Attributed Network Alignment  
Si Zhang (Arizona State University), Hanghang Tong (Arizona State University)

Point-of-Interest Recommendations: Learning Potential Check-ins from Friends  
Huayu Li (The University of North Carolina at Charlotte), Yong Ge (University of Arizona), Richang Hong (Hefei University of Technology), Hengshu Zhu (Baidu Research-Big Data)

QUINT: On Query-Specific Optimal Networks  
Liangyue Li (Arizona State University), Yuan Yao (Nanjing University), Jie Tang (Tsinghua University), Wei Fan (Baidu Big Data Lab), Hanghang Tong (Arizona State University)

Annealed Sparsity via Adaptive and Dynamic Shrinking  
Kai Zhang (NEC Laboratories America), Shandian Zhe (Purdue University), Chaoran Cheng (New Jersey Institute of Technology), Zhi Wei (New Jersey Institute of Technology), Zhengzhang Chen (NEC Laboratories America), Haifeng Chen (NEC Laboratories America), Guofei Jiang (NEC Laboratories America), Yuan Qi (Purdue University), Jieping Ye (University of Michigan, Ann Arbor)

ABRA: Approximating Betweenness Centrality in Static and Dynamic Graphs with Rademacher Averages  
Matteo Riondato (Two Sigma Investments), Eli Upfal (Brown University)

TRIÈST: Counting Local and Global Triangles in Fully-Dynamic Streams with Fixed Memory Size  
Lorenzo De Stefani (Brown University), Alessandro Epasto (Google), Matteo Riondato (Two Sigma Investments), Eli Upfal (Brown University)

Towards Optimal Cardinality Estimation of Unions and Intersections with Sketches  
Daniel Ting (Facebook)

Unified Point-of-Interest Recommendation with Temporal Interval Assessment  
Yanchi Liu (Rutgers University), Chuanren Liu (Drexel University), Bin Liu (Rutgers University), Meng Qu (Rutgers University), Hui Xiong (Rutgers University)
Temporal Order-based First-Take-All Hashing for Fast Attention-Deficit-Hyperactive-Disorder Detection
Hao Hu (University of Central Florida), Joey Velez-Ginorio (University of Central Florida), Guo-Jun Qi (University of Central Florida)

A Multiple Test Correction for Streams and Cascades of Statistical Hypothesis Tests
Geoffrey I Webb (Monash University), Francois Petitjean (Monash University)

Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization
Junming Liu (Rutgers University), Leilei Sun (Dalian University of Technology), Weiwei Chen (Rutgers University), Hui Xiong (Rutgers University)

Graph Wavelets via Sparse Cuts
Arlei Silva (University of California, Santa Barbara), Xuan Hong Dang (University of California, Santa Barbara), Prithwish Basu (Raytheon BBN Technologies), Ambuj Singh (University of California, Santa Barbara), Ananthram Swami (Army Research Laboratory)

Why Should I Trust You?" Explaining the Predictions of Any Classifier"
Marco Tulio Ribeiro (University of Washington), Sameer Singh (University of Washington), Carlos Guestrin (University of Washington)

Overcoming Key Weaknesses of Distance-based Neighbourhood Methods using a Data Dependent Dissimilarity Measure
Kai Ming Ting (Federation University), Ye Zhu (Monash University), Mark Carman (Monash University), Yue Zhu (Nanjing University), Zhi-Hua Zhou (Nanjing University)

Continuous Experience-aware Language Model
Subhabrata Mukherjee (Max Planck Institute for Informatics), Stephan Günnemann (Technical University of Munich), Gerhard Weikum (Max Planck Institute for Informatics)

Hesam Amoualian (University of Grenoble Alps / LIG ), Marianne Clausel (University of Grenoble Alps / LJK), Eric Gaussier (University of Grenoble Alps / LIG ), Massih-Reza Amini (University of Grenoble Alps / LIG)

Structural Neighborhood Based Classification of Nodes in a Network
Sharad Nandanwar (Indian Institute of Science), M. N. Murty (Indian Institute of Science)

Fast Memory-efficient Anomaly Detection in Streaming Heterogeneous Graphs
Emaad Manzoor (Stony Brook University), Sadegh M. Milajerdi (University of Illinois at Chicago), Leman Akoglu (Stony Brook University)

FASCINATE: Fast Cross-Layer Dependency Inference on Multi-layered Networks
Chen Chen (Arizona State University), Hanghang Tong (Arizona State University), Lei Xie (City University of New York), Lei Ying (Arizona State University), Qing He (University at Buffalo)

XGBoost: A Scalable Tree Boosting System
Tianqi Chen (University of Washington), Carlos Guestrin (University of Washington)

Generalized Hierarchical Sparse Model for Arbitrary-Order Interactive Antigenic Sites Identification in Flu Virus Data
Lei Han (Department of Statistics, Rutgers University), Yu Zhang (Department of Computer Science and Engineering, Hong Kong University of Science and Technology), Xiu-Feng Wan (College of Veterinary Medicine, Mississippi State University), Tong Zhang (Department of Statistics, Rutgers University)

A Subsequence Interleaving Model for Sequential Pattern Mining
Jaroslav Fowkes (University of Edinburgh), Charles Sutton (University of Edinburgh)

Partial Label Learning via Feature-Aware Disambiguation
Min-Ling Zhang (Southeast University), Bin-Bin Zhou (Southeast University), Xu-Ying Liu (Southeast University)

Semi-Markov Switching Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems
Igor Melnyk (University of Minnesota), Arindam Banerjee (University of Minnesota), Bryan Matthews (NASA Ames Research Center), Nikunj Oza (NASA Ames Research Center)

Robust Large-Scale Machine Learning in the Cloud
Steffen Rendle (Google Inc.), Dennis Fetterly (Google Inc.), Eugene J. Shekita (Google Inc.), Bor-ying Su (Google Inc.)

Inferring Network Effects from Observational Data
David Arbour (UMass Amherst), Dan Garant (UMass Amherst), David Jensen (UMass Amherst)

Just One More: Modeling Binge Watching Behavior
William Trouleau (EPFL), Azin Ashkan (Technicolor Research), Weicong Ding (Technicolor Research), Brian Eriksson (Technicolor Research)

GMove: Group-Level Mobility Modeling Using Geo-Tagged Social Media
Chao Zhang (University of Illinois at Urbana-Champaign), Keyang Zhang (University of Illinois at Urbana-Champaign), Quan Yuan (University of Illinois at Urbana-Champaign), Luming Zhang (Hefei University of Technology), Tim Hanratty (U.S. Army Research Laboratory), Jiawei Han (University of Illinois at Urbana-Champaign)

Revisiting Random Binning Features: Fast Convergence and Strong Parallelizability
Lingfei Wu (College of William and Mary), Ian E.H. Yen (University of Texas at Austin), Jie Chen (IBM Research), Rui Yan (Baidu Inc.)

Come-and-Go Patterns of Group Evolution: A Dynamic Model
Tianyang Zhang (Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University), Peng Cui (Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University), Christos Faloutsos (Computer Science Department, Carnegie Mellon University), Yunfei Lu (Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University), Hao Ye (Tencent Corporation), Wenwu Zhu (Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University), Shiqiang Yang (Tsinghua National Lab for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University)

Communication Efficient Distributed Kernel Principal Component Analysis
Maria Florina Balcan (Carnegie Mellon University), Yingyu Liang (Princeton University), Le Song (Georgia Institute of Technology), David Woodruff (IBM Research), Bo Xie (Georgia Institute of Technology)

The Limits of Popularity-Based Recommendations, and the Role of Social Ties
Marco Bressan (Sapienza University of Rome), Stefano Leucci (Sapienza University of Rome), Alessandro Panconesi (Sapienza University of Rome), Prabhakar Raghavan (Google), Erisa Terolli (Sapienza University of Rome)
Robust Extreme Multi-label Learning
Chang Xu (Peking University; University of Technology Sydney), Dacheng Tao (University of Technology Sydney), Chao Xu (Peking University)

Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective
Tong Xu (University of Science and Technology of China), Hengshu Zhu (Baidu Research-Big Data Lab), Xiangyu Zhao (University of Science and Technology of China), Qi Liu (University of Science and Technology of China), Hao Zhong (Rutgers, the State University of New Jersey), Enhong Chen (University of Science and Technology of China), Hui Xiong (Rutgers, the State University of New Jersey)

Efficient Frequent Directions Algorithm for Sparse Matrices
Mina Ghashami (University of Utah), Edo Liberty (Yahoo Research), Jeff M. Phillips (University of Utah)

Smart Reply: Automated Response Suggestion for Email
Anjuli Kannan (Google), Karol Kurach (Google), Sujith Ravi (Google), Tobias Kaufmann (Google), Andrew Tomkins (Google), Balint Miklos (Google), Greg Corrado (Google), Laszlo Lukacs (Google), Marina Ganea (Google), Peter Young (Google), Vivek Ramavajjala (Google)

Modeling Precursors for Event Forecasting via Nested Multi-Instance Learning
Yue Ning (Discovery Analytics Center, Department of Computer Science, Virginia Tech), Sathappan Muthiah (Discovery Analytics Center, Department of Computer Science, Virginia Tech), Huzefa Rangwala (Department of Computer Science, George Mason University), Naren Ramakrishnan (Discovery Analytics Center, Department of Computer Science, Virginia Tech)

Assessing Human Error Against a Benchmark of Perfection
Ashton Anderson (Microsoft), Jon Kleinberg (Cornell University), Sendhil Mullainathan (Harvard University)

Approximate Personalized PageRank on Dynamic Graphs
Hongyang Zhang (Stanford University), Peter Lofgren (Stanford University), Ashish Goel (Stanford University)

Structured Doubly Stochastic Matrix for Graph Based Clustering
Xiaoqian Wang (University of Texas at Arlington), Feiping Nie (University of Texas at Arlington), Heng Huang (University of Texas at Arlington)

Privacy-preserving Class Ratio Estimation

Joint Community and Structural Hole Spanner Detection via Harmonic Modularity
Lifang He (Shenzhen University), Chun-Ta Lu (University of Illinois at Chicago), Jiaqi Ma (Tsinghua University), Jianping Cao (National University of Defense Technology), Linlin Shen (Shenzhen University), Philip S. Yu (University of Illinois at Chicago)

Sampling of Attributed Networks from Hierarchical Generative Models
Pablo Robles (Purdue University), Sebastian Moreno (Universidad Adolfo Ibañez), Jennifer Neville (Purdue University)

Goal-Directed Inductive Matrix Completion
Si Si (University of Texas at Austin), Kai-Yang Chiang (University of Texas at Austin), Cho-Jui Hsieh (University of California, Davis), Nikhil Rao (Technicolor R&I), Inderjit S Dhillon (University of Texas at Austin)
Monday August 15, 2016 Detailed Program

Monday 7:30am - 5:15pm, Registration, Registration Desk - East Lounge
Monday 9:15am - 5:30pm, Exhibit Hall - Continental

Monday 8:00am - 9:30am, Grand Ballroom
Keynote Session 1:
Graphons and Machine Learning: Modeling and Estimation of Sparse Massive Networks
Speaker: Jennifer Chayes chaired by Charu Aggarwal
Abstract: There are numerous examples of sparse massive networks, in particular the Internet, WWW and online social networks. How do we model and learn these networks? In contrast to conventional learning problems, where we have many independent samples, it is often the case for these networks that we can get only one independent sample. How do we use a single snapshot today to learn a model for the network, and therefore be able to predict a similar, but larger network in the future? In the case of relatively small or moderately sized networks, it’s appropriate to model the network parametrically, and attempt to learn these parameters. For massive networks, a non-parametric representation is more appropriate. In this talk, we first review the theory of graphons, developed over the last decade to describe limits of dense graphs, and the more the recent theory describing sparse graphs of unbounded average degree, including power-law graphs. We then show how to use these graphons as non-parametric models for sparse networks. Finally, we show how to get consistent estimators of these non-parametric models, and moreover how to do this in a way that protects the privacy of individuals on the network.

Monday 9:30am - 10:00am, Coffee Break

Monday 10:00am - 12:00pm
Research Session R1: Graphs and Rich Data, Plaza AB
Chair: Jiawei Han

FRAUDAR: Bounding Graph Fraud in the Face of Camouflage

CatchTartan: Representing and Summarizing Dynamic Multicontextual Behaviors
Meng Jiang, UIUC; Christos Faloutsos, Carnegie Mellon University; Jiawei Han, University of Illinois at Urbana-Champaign

Compact and Scalable Graph Neighborhood Sketching
Takuya Akiba, NII; Yosuke Yano, National Institute of Informatics

QUINT: On Query-Specific Optimal Networks
Liangyue Li, Arizona State University; Yuan Yao, Nanjing University; Jie Tang, Tsinghua University; Wei Fan, Baidu; Hanghang Tong, Arizona State University
Graph Wavelets via Sparse Cuts
Arlei Lopes da Silva, UC, Santa Barbara; Xuan-Hong Dang, UCSB; Prithwish Basu, Raytheon BBN; Ambuj Singh, UCSB; Ananthram Swami, Army Lab

Monday 10:00am - 12:00pm
Research Session R2: Large Scale Data Mining, Imperial Ballroom
Chair: Johannes Gehrke

PTE: Enumerating Trillion Triangles On Distributed Systems
Ha-Myung Park, KAIST; Sung-Hyon Myaeng, KAIST; U Kang, Seoul National University

Towards Optimal Cardinality Estimation of Unions and Intersections with Sketches
Daniel Ting, Facebook

Fast Memory-efficient Anomaly Detection in Streaming Heterogeneous Graphs
Emaad Manzoor, Stony Brook University; Leman Akoglu, SUNY Stony Brook

XGBoost: A Scalable Tree Boosting System
Tianqi Chen, University of Washington; Carlos Guestrin, Dato/Univ of Washington

Communication Efficient Distributed Kernel Principal Component Analysis
Yingyu Liang, Princeton University; Bo Xie, ; David Woodruff, IBM Research; Le Song; Maria-Florina Balcan

Monday 10:00am - 12:00pm
Applied Data Science Invited Talks 1, Yosemite
Chair: Usama Fayyad

Can You Teach The Elephant To Dance? AKA: Culture Eats Data Science for Breakfast
Jonathan Becher: Chief Digital Officer / SAP

Large-Scale Machine Learning at Verizon: Theory and Applications
Ashok Srivastava: Chief Data Scientist / Verizon

Computational Social Science: Exciting Progress and Future Challenges
Duncan Watts: Principal Researcher / Microsoft.

Monday 10:00am - 12:00pm
Applied Data Science A1: Social Good - I, Cyril Magnin (Parc)
Chair: Yu Zheng

Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta
Michael Madaio, Carnegie Mellon University; Shang-Tse Chen, Georgia Institute of Technology; Oliver Haimson, University of California, Irvine; Wenwen Zhang, Georgia Institute of Technology; Xiang Cheng, Emory University; Matthew Hinds-Aldrich, Atlanta Fire Rescue Department; Duen Horng Chau, Georgia Tech; Bistra Dilkina, Georgia Tech
Dynamic and Robust Wildfire Risk Prediction System: An Unsupervised Approach
Mahsa Salehi, IBM Australia; Laura Rusu, IBM Research; Timothy Lynar, IBM Research; Anna Phan, IBM Research

Aircraft Trajectory Prediction made easy with Predictive Analytics
Samet Ayhan, University of Maryland; Hanan Samet, University of Maryland

Catch Me If You Can: Detecting Pickpocket Suspects from Large-Scale Transit Records
Bowen Du, Beihang University; Chuanren Liu, Drexel University; Wenjun Zhou, U of Tennessee; Hui Xiong, Rutgers

Identifying Police Officers at Risk of Adverse Events
Samuel Carton, University of Michigan; Jennifer Helsby, University of Chicago; Kenneth Joseph, Carnegie Mellon University; Ayesha Mahmud, Princeton University; Youngsoo Park, University of Arizona; Joe Walsh, University of Chicago; Crystal Cody, Charlotte-Mecklenburg Police Department; Estella Patterson, Charlotte-Mecklenburg Police Department; Lauren Haynes, University of Chicago; Rayid Ghani, University of Chicago

Monday 12:15pm - 1:30pm, Grand Ballroom
Lunch, Keynote Session 2
People, Computers, and The Hot Mess of Real Data
Speaker: Joe Hellerstein chaired by Dou Shen
Abstract: In practice, end-to-end data analysis is rarely a cleanly engineered process. Acquiring data can be tricky. Data assessment, wrangling and feature extraction are time-consuming and subjective. Models and algorithms used to derive data products are highly contextualized by time-varying properties of data sources, code and application needs. All of these issues would ideally benefit from an organizational view, but are often driven by individual users. Viewed holistically, both agile analytics and the establishment of analytic pipelines involve interactions between people, computation and infrastructure. In this talk I’ll share some anecdotes from our research, user studies, and field experience with companies (Trifacta, Captricity), as well as an emerging open-source project (Ground).

Monday 1:45pm - 3.45pm
Research Session R3: Graphs and Social Networks - I, Plaza AB
Chair: Ambuj Singh

Come-and-Go Patterns of Group Evolution: A Dynamic Model
Tianyang Zhang, Tsinghua University; Peng Cui, Tsinghua University; Christos Faloutsos, Carnegie Mellon University; Wenwu Zhu, Tsinghua University; Shiqiang Yang

Robust Influence Maximization
Wei Chen, Microsoft Research; Tian Lin, Tsinghua University; Zihan Tan, IIIS, Tsinghua University; Mingfei Zhao, IIIS, Tsinghua University; Xuren Zhou, The Hong Kong University of Science and Technology

Robust Influence Maximization
Xinran He, University of Southern California; David Kempe, University of Southern California

Structural Neighborhood based Classification of Nodes in a Network
Sharad Nandanwar, Indian Institute of Science; Musti Narasimha Murty, Indian Institute of Science
Asymmetric Transitivity Preserving Graph Embedding
Mingdong Ou, Tsinghua University; Peng Cui, Tsinghua University; Jian Pei, Simon Fraser University; Wenwu Zhu, Tsinghua University

**Monday 1:45pm - 3.45pm**
**Research Session R4: Streams and Temporal Evolution - I, Imperial Ballroom**
**Chair: Peter Flach**

NetCycle: Collective Evolution Inference in Heterogeneous Information Networks
Yizhou Zhang, Fudan University; Xiong Yun; Xiangnan Kong, Worcester Polytechnic Institute; Yangyong Zhu, Fudan University

Positive-Unlabeled Learning in Streaming Networks
Shiyu Chang, UIUC; Yang Zhang, UIUC; Jiliang Tang, Yahoo Labs; Dawei Yin; Yi Chang, Yahoo! Labs; Mark Hasegawa-Johnson, UIUC; Thomas Huang, UIUC

TRIEST: Counting Local and Global Triangles in Fully-Dynamic Streams with Fixed Memory Size
Lorenzo De Stefani, Brown University; Alessandro Epasto, Brown; Matteo Riondato, Two Sigma Investments; Eli Upfal, Brown University

Temporal Order-based First-Take-All Hashing for Fast Attention-Deficit-Hyperactive-Disorder Detection
Hao Hu, University of Central Florida; Joey Velez-Ginorio, University of Central Florida; Guojun Qi, University of Central Florida

Approximate Personalized PageRank on Dynamic Graphs
Hongyang Zhang, Stanford University; Peter Lofgren, Stanford University

**Monday 1:45pm - 3.45pm**
**Panel Discussion: Is Deep Learning the New 42?, Yosemite**
**Chair: Andrei Broder**

**Monday 1:45pm - 3.45pm**
**Applied Data Science Session A2: Social Good - II, Cyril Magnin (Parc)**
**Chair: Ying Li**

Recruitment Market Trend Analysis with Sequential Latent Variable Models
Chen Zhu, Baidu HR; Hengshu Zhu, Baidu Inc.; Hui Xiong, Rutgers; Ding Pengliang; Xie Fang

Ranking Universities Based on Career Outcomes of Graduates
Navneet Kapur, GoFundMe; Nikita Lytkin, LinkedIn Corporation; Bee-Chung Chen, LinkedIn Corporation; Deepak Agarwal, LinkedIn Corporation; Igor Perisic, LinkedIn Corporation

Designing Policy Recommendations to Reduce Home Abandonment in Mexico
Klaus Ackermann, Monash University; Eduardo Blancas Reyes, The University of Chicago; Sue He, University of Virginia; Thomas Anderson Keller, UC San Diego; Paul van der Boor, Data Science for Social Good; Romana Khan, Data Science for Social Good; Rayid Ghani, University of Chicago
A Non-parametric Approach to Detect Epileptogenic Lesions using Restricted Boltzmann Machines
Yijun Zhao, Tufts University; Bilal Ahmed, Tufts; Carla Brodley, Northeastern University; Jennifer Dy, NEU

Batch model for batched timestamps data analysis with application to the SSA disability program
Qingqi Yue, NIH; Ao Yuan, NIH; Xuan Che, NIH; Elizabeth Rasch, NIH; Minh Huynh, Impaq; Chunxiao Zhou, NIH

Monday 3:45pm - 4:15pm, Coffee Break

Monday 4:15pm - 6:15pm
Research Session R5: Graphs and Social Networks - II, Plaza AB
Chair: Tanya Berger-Wolf

FINAL: Fast Attributed Network Alignment
Si Zhang, Arizona State University; Hanghang Tong, Arizona State University

Inferring Network Effects from Observational Data
David Arbour, University of Massachusetts Am; Dan Garant, University of Massachusetts Amherst; David Jensen, UMass Amherst

Joint Community and Structural Hole Spanner Detection via Harmonic Modularity
Lifang He; Chun-Ta Lu, UIC; Jiaqi Ma, Tsinghua University; Jianping Cao, NUDT; Linlin Shen, ; Philip S. Yu, UI Chicago

Sampling of Attributed Networks From Hierarchical Generative Models
Pablo Robles Granda, Purdue University; Sebastian Moreno; Jennifer Neville, Purdue

FASCINATE: Fast Cross-Layer Dependency Inference on Multi-layered Networks
Chen Chen, Arizona State University; Hanghang Tong, Arizona State University; Lei Xie, City University of New York; Lei YIng, Arizona State University; Qing He, Arizona State University

Monday 4:15pm - 6:15pm
Research Session R6: Streams and Temporal Evolution - II, Imperial Ballroom
Chair: Jing Gao

GMove: Group-Level Mobility Modeling using Geo-Tagged Social Media
Chao Zhang, UIUC; Keyang Zhang, ; Quan Yuan, University of Illinois Urbana-; Luming Zhang, ; Tim Hanratty, ; Jiawei Han, University of Illinois at Urbana-Champaign

A multiple test correction for streams and cascades of statistical hypothesis tests
Geoff Webb, Monash University; Francois Petitjean, Monash

Continuous Experience-aware Language Model
Subhabrata Mukherjee, Max Planck Informatics; Stephan Günnemann, Technical University of Munich; Gerhard Weikum, Max Planck Institute for Informatics

Streaming-LDA: a copula-based approach to modeling topic dependencies in document streams
Hesam Amoualian, University Grenoble Alps; Marianne Clausel, University of Grenoble Alps; Eric Gaussier, University of Grenoble Alps; Massih-Reza Amini, University of Grenoble Alps
A Subsequence Interleaving Model for Sequential Pattern Mining
Jaroslav Fowkes, University of Edinburgh; Charles Sutton, University of Edinburgh

**Monday 4:15pm - 6:15pm**
**Applied Data Science Invited Talks 2, Yosemite**
**Chair: Rajesh Parekh**

Bayesian Optimization and Embedded Learning Systems
Jeff Schneider: Engineering Lead / Uber Advanced Technology Center

Ingo Mierswa: Founder and CTO / RapidMiner

How Machine Learning has Finally Solved Wanamaker’s Dilemma
Oliver Downs: Chief Scientist and CTO / Amplero

**Monday 4:15pm - 6:15pm**
**Applied Data Science Session A3: Miscellaneous Topics, Cyril Magnin (Parc)**
**Chair: Tina Eliassi-Rad**

The Legislative Influence Detector: Finding Text Reuse in State Legislation
Matthew Burgess, University of Michigan; Eugenia Giraudy, YouGov; Julian Katz-Samuels, University of Michigan; Joe Walsh*, University of Chicago; Derek Willis, ProPublica; Lauren Haynes, University of Chicago; Rayid Ghani, University of Chicago

DopeLearning: A Computational Approach to Rap Lyrics Generation
Eric Malmi, Aalto University; Pyry Takala, Aalto University; Hannu Toivonen, University of Helsinki; Tapani Raiko, Aalto University; Aristides Gionis, Aalto University

Days on Market: Measuring Liquidity in Real Estate Markets
Author(s): Hengshu Zhu, Baidu Inc.; Hui Xiong, Rutgers; Fangshuang Tang, University of Science and Technology of China; Yong Ge, ; Qi Liu, University of Science and Technology of China; Enhong Chen, ; Yanjie Fu, Rutgers University

Developing a Data-Driven Player Ranking in Soccer using Predictive Model Weights
Joel Brooks, Massachusetts Institute of Technology; Matthew Kerr, Massachusetts Institute of Technology; John Guttag, Massachusetts Institute of Technology

Identifying Earmarks in Congressional Bills
Vrushank Vora, Data Science for Social Good; Joe Walsh, Data Science for Social Good; Madian Khabasa, Microsoft; Ellery Wulczyn, Wikimedia Foundation; Matthew Heston, Northwestern University; Rayid Ghani, University of Chicago; Chris Berry, University of Chicago

**Monday August 15, 6:00pm - 7:00pm Poster Set-Up (Applied Data Science Track) – Golden Gate**
Monday August 15, 7:00pm - 12:00am Poster Session: Applied Data Science Track – Golden Gate

Collaborative Knowledge Base Embedding for Recommender Systems
Fuzheng Zhang (Microsoft Research), Nicholas Jing Yuan (Microsoft), Defu Lian (Big Data Research Center, University of Electronic Science and Technology of China), Xing Xie (Microsoft Research), Wei-Ying Ma (Microsoft Research)

Repeat Buyer Prediction for E-Commerce
Guimei Liu (Institute for Infocomm Research), Tam T. Nguyen (Institute for Infocomm Research), Gang Zhao (Development Bank of Singapore), Wei Zha (Institute for Infocomm Research), Jianbo Yang (General Electric), Jianneng Cao (Institute for Infocomm Research), Min Wu (Institute for Infocomm Research), Peilin Zhao (Institute for Infocomm Research), Wei Chen (Development Bank of Singapore)

Matrix Computations and Optimization in Apache Spark
Reza Bosagh Zadeh (Stanford and Matroid), Xiangrui Meng (Databricks), Alexander Ulanov (HP Labs), Burak Yavuz (Databricks), Li Pu (Twitter), Shivaram Venkataraman (University of California, Berkeley), Evan Sparks (University of California, Berkeley), Aaron Staple (Databricks), Matei Zaharia (Databricks and Massachusetts Institute of Technology)

Contextual Intent Tracking for Personal Assistants
Yu Sun (University of Melbourne), Nicholas Jing Yuan (Microsoft Corporation), Yingzi Wang (University of Science and Technology of China & Microsoft Research), Xing Xie (Microsoft Research), Kieran McDonald (Microsoft Corporation), Rui Zhang (University of Melbourne)

Ranking Relevance in Yahoo Search
Dawei Yin (Yahoo! Inc.), Yuening Hu (Yahoo! Inc.), Jiliang Tang (Yahoo! Inc.), Tim Daly (Yahoo! Inc.), Mianwei Zhou (Yahoo! Inc.), Hua Ouyang (Yahoo! Inc.), Jianhui Chen (Yahoo! Inc.), Chăngsũng Kang (Yahoo! Inc.), Hongbo Deng (Yahoo! Inc.), Chikashi Nobata (Yahoo! Inc.), Jean-Marc Langlois (Yahoo! Inc.), Yi Chang (Yahoo! Inc.)

Large-Scale Item Categorization in e-Commerce Using Multiple Recurrent Neural Networks
Jung-Woo Ha (NAVER Corp.), Hyuna Pyo (NAVER Corp.), Jeonghee Kim (NAVER Corp.)

DopeLearning: A Computational Approach to Rap Lyrics Generation
Eric Malmi (Aalto University), Pyry Takala (Aalto University), Hannu Toivonen (University of Helsinki), Tapani Raiko (Aalto University), Aristides Gionis (Aalto University)

Audience Expansion for Online Social Network Advertising
Haishan Liu (LinkedIn Corporation), David Pardoe (LinkedIn Corporation), Kun Liu (LinkedIn Corporation), Manoj Thakur (LinkedIn Corporation), Frank Cao (LinkedIn Corporation), Chongzhe Li (LinkedIn Corporation)

Engagement Capacity and Engaging Team Formation for Reach Maximization of Online Social Media Platforms
Alexander Nikolaev (University at Buffalo), Shounak Gore (University at Buffalo), Venu Govindaraju (University at Buffalo)

Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta
Michael Madaio (Carnegie Mellon University), Shang-Tse Chen (Georgia Institute of Technology), Oliver L. Haimson (University of California, Irvine), Wenwen Zhang (Georgia Institute of Technology), Xiang Cheng (Emory University), Matthew Hinds-Aldrich (Atlanta Fire Rescue Department), Duen Horng Chau (Georgia Institute of Technology), Bistra Dilkina (Georgia Institute of Technology)
From Online Behaviors to Offline Retailing
Ping Luo (Institute of Computing Technology, CAS), Su Yan (Institute of Computing Technology, CAS), Zhiqiang Liu (Baidu, Inc.), Zhiyong Shen (Baidu, Inc.), Shengwen Yang (Baidu, Inc.), Qing He (Institute of Computing Technology, CAS)

GLMix: Generalized Linear Mixed Models For Large-Scale Response Prediction
XianXing Zhang (LinkedIn), Yitong Zhou (LinkedIn), Yiming Ma (LinkedIn), Bee-Chung Chen (LinkedIn), Liang Zhang (LinkedIn), Deepak Agarwal (LinkedIn)

Dynamic and Robust Wildfire Risk Prediction System: An Unsupervised Approach
Mahsa Salehi (IBM Research), Laura Irina Rusu (IBM Research), Timothy Lynar (IBM Research), Anna Phan (IBM Research)

Days on Market: Measuring Liquidity in Real Estate Markets
Hengshu Zhu (Baidu Research-Big Data Lab), Hui Xiong (Rutgers University), Fangshuang Tang (University of Science and Technology of China), Qi Liu (University of Science and Technology of China), Yong Ge (University of Arizona), Enhong Chen (University of Science and Technology of China), Yanjie Fu (Missouri University of Science and Technology)

Catch Me If You Can: Detecting Pickpocket Suspects from Large-Scale Transit Records
Bowen Du (Beihang University), Chuanren Liu (Drexel University), Wenjun Zhou (University of Tennessee), Zhenshan Hou (Beihang University), Hui Xiong (Rutgers)

Boosted Decision Tree Regression Adjustment for Variance Reduction in Online Controlled Experiments
Alexey Poyarkov (Yandex), Alexey Drutsa (Yandex), Andrey Khalyavin (Yandex), Gleb Gusev (Yandex), Pavel Serdyukov (Yandex)

Recruitment Market Trend Analysis with Sequential Latent Variable Models
Chen Zhu (BAIDU Inc.), Hengshu Zhu (Baidu Inc. & Baidu Research-Big Data Lab), Hui Xiong (Rutgers University), Pengliang Ding (BAIDU Inc.), Fang Xie (BAIDU Inc.)

An Empirical Study on Recommendation with Multiple Types of Feedback
Liang Tang (LinkedIn), Bo Long (Particle Media Inc), Bee-Chung Chen (LinkedIn), Deepak Agarwal (LinkedIn)

Online Dual Decomposition for Performance and Delivery-Based Distributed Ad Allocation
Jim C Huang (Amazon), Rodolphe Jenatton (Amazon), Cedric Archambeau (Amazon)

Email Volume Optimization at LinkedIn
Rupesh Gupta (LinkedIn), Guanfeng Liang (LinkedIn), Hsiao-Ping Tseng (LinkedIn), Ravi Kiran Holur Vijay (LinkedIn), Xiaoyu Chen (LinkedIn), Romer Rosales (LinkedIn)

An Engagement-Based Customer Lifetime Value System for E-commerce
Ali Vanderveld (Groupon), Addhyan Pandey (Groupon), Angela Han (Google), Rajesh Parekh (Facebook)

Aircraft Trajectory Prediction Made Easy with Predictive Analytics
Samet Ayhan (University of Maryland), Hanan Samet (University of Maryland)

Developing a Data-Driven Player Ranking in Soccer Using Predictive Model Weights
Joel Brooks (Massachusetts Institute of Technology), Matthew Kerr (Massachusetts Institute of Technology), John Guttag (Massachusetts Institute of Technology)
Question Independent Grading using Machine Learning: The Case of Computer Program Grading
Gursimran Singh (Aspiring Minds), Shashank Srikant (Aspiring Minds), Varun Aggarwal (Aspiring Minds)

The Legislative Influence Detector
Matthew Burgess (University of Michigan), Eugenia Giraudy (UC Berkeley), Julian Katz-Samuels (University of Michigan), Joe Walsh (University of Chicago), Derek Willis (ProPublica), Lauren Haynes (University of Chicago), Rayid Ghani (University of Chicago)

Identifying Police Officers at Risk of Adverse Events
Samuel Carton (University of Michigan), Jennifer Helsby (University of Chicago), Kenneth Joseph (Carnegie Mellon University), Ayesha Mahmud (Princeton University), Youngsoo Park (University of Arizona), Joe Walsh (University of Chicago), Crystal Cody (Charlotte-Mecklenburg Police Department), CPT Estella Patterson (Charlotte-Mecklenburg Police Department), Lauren Haynes (University of Chicago), Rayid Ghani (University of Chicago)

Predicting Disk Replacement towards Reliable Data Centers
Mirela Madalina Botezatu (IBM Research), Ioana Giurgiu (IBM Research), Jasmina Bogojeska (IBM Research), Dorothea Wiesmann (IBM Research)

Data-Driven Metric Development for Online Controlled Experiments: Seven Lessons Learned
Alex Deng (Microsoft), Xiaolin Shi (Yahoo Research)

Ranking Universities Based on Career Outcomes of Graduates
Navneet Kapur (GoFundMe), Nikita Lytkin (LinkedIn Corporation), Bee-Chung Chen (LinkedIn Corporation), Deepak Agarwal (LinkedIn Corporation), Igor Perisic (LinkedIn Corporation)

Designing Policy Recommendations to Reduce Home Abandonment in Mexico
Klaus Ackermann (Monash University), Eduardo Blancas Reyes (University of Chicago), Sue He (University of Virginia), Thomas Anderson Keller (UC San Diego), Paul van der Boor (University of Chicago), Romana Khan (Northwestern University), Rayid Ghani (University of Chicago), Jose Carlos Gonzalez (Infonavit)

Evaluating Mobile Apps with A/B and Quasi A/B Tests
Ya Xu (LinkedIn Corporation), Nanyu Chen (LinkedIn Corporation)

Deep Crossing: Web-Scale Modeling without Manually Crafted Combinatorial Features
Ying Shan (Microsoft Corporation), T. Ryan Hoens (Microsoft Corporation), Jian Jiao (Microsoft Corporation), Haijing Wang (Microsoft Corporation), Dong Yu (Microsoft Corporation), JC Mao (Microsoft Corporation)

A Non-parametric Approach to Detect Epileptogenic Lesions using Restricted Boltzmann Machines
Yijun Zhao (Tufts University), Bilal Ahmed (Tufts University), Thomas Thesen (New York University), Karen E Blackmon (New York University), Jennifer G Dy (Northeastern University), Carla E Brodley (Northeastern University), Ruben Kuzniekcy (New York University), Orrin Devinsky (New York University)

Batch Model for Batched Timestamps Data Analysis with Application to the SSA Disability Program
Qingqi Yue (Clinical Center/ National Institutes of Health), Ao Yuan (Clinical Center/ National Institutes of Health), Xuan Che (Clinical Center/ National Institutes of Health), Minh Huynh (Impac International LLC), Chunxiao Zhou (Clinical Center/ National Institutes of Health)

Identifying Decision Makers from Professional Social Networks
Shipeng Yu (LinkedIn), Evangelia Christakopoulou (University of Minnesota), Abhishek Gupta (LinkedIn)
Minimizing Legal Exposure of High-Tech Companies through Collaborative Filtering Methods
Bo Jin (Dalian University of Technology), Chao Che (Dalian University), Kuifei Yu (Zhigu Tech), Yue Qu (Dalian University of Technology), Li Guo (Dalian University of Technology), Cuili Yao (Dalian University of Technology), Ruixun Yu (Northeastern University), Qiang Zhang (Dalian University)

EMBERS at 4 years: Experiences operating an Open Source Indicators Forecasting System
Satthapan Muthiah (Virginia Tech), Patrick Butler (Virginia Tech), Rupinder Paul Khandpur (Virginia Tech), Parang Saraf (Virginia Tech), Nathan Self (Virginia Tech), Alla Rozovskaya (Virginia Tech), Liang Zhao (Virginia Tech), Jose Cadena (Virginia Tech), Chang-Tien Lu (Virginia Tech), Anil Vullikanti (Virginia Tech), Achla Marathe (Virginia Tech), Kristen Summers (IBM Watson Group), Graham Katz (CACI), Andy Doyle (CACI), Jaime Arredondo (University of California at San Diego), Dipak K Gupta (San Diego State University), David Mares (University of California at San Diego), Naren Ramakrishnan (Virginia Tech)

Predictors without Borders: Behavioral Modeling of Product Adoption in Three Developing Countries
Muhammad R Khan (University of Washington), Joshua E Blumenstock (University of Washington)

Identifying Earmarks in Congressional Bills
Ellery Wulczyn (Stanford University), Madian Khabsa (Microsoft ), Ruvshank Vora (University of Chicago), Matthew Heston (Northwestern), Joe Walsh (University of Chicago), Christopher Berry (University of Chicago), Rayid Ghani (University of Chicago)

Anomaly Detection Using Program Control Flow Graph Mining From Execution Logs
Animesh Nandi (IBM Research), Atri Mandal (IBM Research), Shubham Atreja (IIT), Gargi B. Dasgupta (IBM Research), Subhrajit Bhattacharya (IBM Research)

Bid-aware Gradient Descent for Unbiased Learning with Censored Data in Display Advertising
Weinan Zhang (Shanghai Jiao Tong University), Tianxiong Zhou (TukMob Inc.), Jun Wang (University College London), Jian Xu (TouchPal Inc.)

Scalable Time-Decaying Adaptive Prediction Algorithm
Yinyan Tan (Research and Standard Department, Huawei Software Technologies CO. LTD), Zhe Fan (Research and Standard Department, Huawei Software Technologies CO. LTD), Guilin Li (Research and Standard Department, Huawei Software Technologies CO. LTD), Fangshan Wang (Research and Standard Department, Huawei Software Technologies CO. LTD), Zhengbing Li (Research and Standard Department, Huawei Software Technologies CO. LTD), Shiikai Liu (Research and Standard Department, Huawei Software Technologies CO. LTD), Qiuling Pan (Research and Standard Department, Huawei Software Technologies CO. LTD), Eric P. Xing (School of Computer Science, Carnegie Mellon University), Qirong Ho (School of Computer Science, Carnegie Mellon University)

Computational Drug Repositioning Using Continuous Self-Controlled Case Series
Zhaobin Kuang (University of Wisconsin), James Thomson (Morgridge Institute), Michael Caldwell (Marshfield Clinic), Peggy Peissig (Marshfield Clinic), Ron Stewart (Morgridge Institute), David Page (University of Wisconsin)

Domain Adaptation in the Absence of Source Domain Data
Boris Chidlovskii (Xerox Research Centre Europe), Stephane Clinchant (Xerox Research Centre Europe), Gabriela Csurka (Xerox Research Centre Europe)

Compute Job Memory Recommender System Using Machine Learning
Taraneh Taghavi (Qualcomm Technologies Incorporation), Maria Luptetini (Qualcomm Technologies Incorporation), Yaron Kretchmer (Qualcomm Technologies Incorporation)
CaSMoS: A Framework for Learning Candidate Selection Models over Structured Queries and Documents
Fedor Borisyuk (LinkedIn Corporation), Krishnaram Kenthapadi (LinkedIn Corporation), David Stein (LinkedIn Corporation), Bo Zhao (LinkedIn Corporation)

Kam1n0: MapReduce-based Assembly Clone Search for Reverse Engineering
Steven H. H. Ding (McGill University), Benjamin C. M. Fung (McGill University), Philippe Charland (Defence R&D Canada - Valcartier)

Text Mining in Clinical Domain: Dealing with Noise
Hoang Nguyen (Data61 - CSIRO), Jon Patrick (University of Sydney)

How to Get Them a Dream Job?
Jia Li (University of Illinois at Chicago), Dhruv Arya (LinkedIn), Viet Ha-Thuc (LinkedIn), Shakti Sinha (LinkedIn)

Detecting Devastating Diseases in Search Logs
John Paparrizos (Columbia University), Ryan W. White (Microsoft Research), Eric Horvitz (Microsoft Research)

Singapore in Motion: Insights on Public Transport Service Level Through Farecard and Mobile Data Analytics
Hasan Poonawala (IBM research), Vinay Kolar (Cisco systems inc.), Sebastien Blandin (IBM research), Laura Wynter (IBM research), Sambit Sahu (IBM research)

Joint Optimization of Multiple Performance Metrics in Online Video Advertising
Sahin Cem Geyik (Turn Inc.), Sergey Faleev (Turn Inc.), Jianqiang Shen (Turn Inc.), Sean O'Donnell (Turn Inc.), Santanu Kolay (Turn Inc.)

Analyzing Volleyball Match Data from the 2014 World Championships Using Machine Learning Techniques
Jan Van Haaren (KU Leuven), Horesh Ben Shitrit (Second Spectrum), Jesse Davis (KU Leuven), Pascal Fua (Ecole Polytechnique Federale de Lausanne)

MAP: Frequency-Based Maximization of Airline Profits based on an Ensemble Forecasting Approach
Bo An (Nanyang Technological University), Haipeng Chen (Nanyang Technological University), Noseong Park (University of Maryland), V.S. Subrahmanian (University of Maryland)

CompanyDepot: Employer Name Normalization in the Online Recruitment Industry
Qiaoling Liu (CareerBuilder LLC), Faizan Javed (CareerBuilder LLC), Matt Mcnair (CareerBuilder LLC)

Images Don't Lie: Transferring Deep Visual Semantic Features to Large-Scale Multimodal Learning to Rank
Corey Lynch (Etsy), Kamelia Aryafar (Etsy), Josh Attenberg (Etsy)

Understanding Behaviors that Lead to Purchasing: A Case Study of Pinterest
Caroline Lo (Stanford University), Dan Frankowski (Pinterest), Jure Leskovec (Pinterest & Stanford University)

Scalable Fast Rank-1 Dictionary Learning for fMRI Big Data Analysis
Xiang Li (University of Georgia), Milad Makkie (University of Georgia), Binbin Lin (University of Michigan), Mojtaba Sedigh Fazli (University of Georgia), Ian Davidson (University of California, Davis), Jieping Ye (University of Michigan), Tianming Liu (University of Georgia), Shannon Quinn (University of Georgia)

Deploying Analytics with the Portable Format for Analytics (PFA)
Jim Pivarski (Open Data Group), Collin Bennett (Open Data Group), Robert L. Grossman (Open Data Group)
Talent Circle Detection in Job Transition Networks
Huang Xu (Northwestern Polytechnical University), Zhiwen Yu (Northwestern Polytechnical University), Jingyuan Yang (Rutgers University), Hui Xiong (Rutgers University), Hengshu Zhu (Baidu Research-Big Data Lab)

Improving the Sensitivity of Online Controlled Experiments: Case Studies at Netflix
Huizhi Xie (Netflix), Juliette Aurisset (Netflix)

When Recommendation Goes Wrong - Anomalous Link Discovery in Recommendation Networks
Bryan Perozzi (Stony Brook University), Michael Schueppert (Google), Jack Saalweachter (Google), Mayur Thakur (-)

Gemello: Creating a Detailed Energy Breakdown from Just the Monthly Electricity Bill
Nipun Batra (IIIT Delhi), Amarjeet Singh (IIIT Delhi), Kamin Whitehouse (University of Virginia)

Crime Rate Inference with Big Data
Hongjian Wang (Pennsylvania State University), Daniel Kifer (Pennsylvania State University), Corina Graif (Pennsylvania State University), Zhenhui Li (Pennsylvania State University)

EMBERS AutoGSR: Automated Coding of Civil Unrest Events
Parang Saraf (Virginia Tech), Naren Ramakrishnan (Virginia Tech)

Convolutional Neural Networks for Steady Flow Approximation
Xiaoxiao Guo (University of Michigan), Wei Li (Autodesk Research), Francesco Iorio (Autodesk Research)
Tuesday August 16, 2016 Detailed Program

Tuesday 7:45am - 5:15pm, Registration, Registration Desk - East Lounge
Tuesday 9:15am - 5:30pm, Exhibit Hall - Continental

Tuesday 8:00am - 9:30am, Grand Ballroom
Keynote Session 3:
A VC View of Investing in ML
Speaker: Greg Papadopoulos chaired by Rajeev Rastogi
Abstract: We are seeing a remarkable watershed in the application of data science across markets and industries. A trifecta of advances in algorithms, cheap cycles, and the capture of networked data from everywhere are no doubt the catalysts. The results for many are continuous improvements in efficiencies, and for some are a fundamental re-imagination and disruption of just about every industry. This talk will give examples we are seeing (and funding!) for the latter, and then focus on our views of the ecosystem of value-from-data infrastructure and end-application companies. A big question is whether the enormous collective advances in tools, techniques and education are in-fact converting would-be differentiated products into democratized features used everywhere. We'll follow the value and make our own predictions on future as ML as a business.

Tuesday 9:30am - 10:00am, Coffee Break

Tuesday 9:30am - 12:30pm Special Session: Standards for Predictive Analytics in the Era of Big and Fast Data - Data Mining Group

Tuesday 10:00am - 12:00pm
Research Session R7: Clustering, Plaza AB
Chair: Martin Ester

Skinny-dip: Clustering in a Sea of Noise
Samuel Maurus, Helmholtz Zentrum München; Claudia Plant

Targeted Topic Modeling for Focused Analysis
Shuai Wang, University of Illinois at Chicago; Geli Fei, Univ of Illinois at Chicago; Zhiyuan Chen, UIC; Bing Liu, Univ of Illinois at Chicago; Sherry Emery, University of Illinois at Chicago

AnyDBC: An Efficient Anytime Density-based Clustering Algorithm for Very Large Complex Datasets
Son Mai, Aarhus University; Ira Assent, ; Martin Storgaard, Aarhus University

Structured Doubly Stochastic Matrix for Graph Based Clustering
Xiaoqian Wang, Univ. of Texas at Arlington; Feiping Nie, University of Texas at Arlington; Heng Huang, Univ. of Texas at Arlington

Dynamic Clustering of Streaming Short Documents
Shangsong Liang, University College London; Emine Yilmaz, University College London; Evangelos Kanoulas, University of Amsterdam
Tuesday 10:00am - 12:00pm
Research Session RT8: Deep Learning and Embedding, Imperial Ballroom
Chair: Jure Leskovec

Structural Deep Network Embedding
Daixin Wang, Tsinghua University; Peng Cui, Tsinghua University; Wenwu Zhu, Tsinghua University

node2vec: Scalable Feature Learning for Networks
Aditya Grover, Stanford University; Jure Leskovec, Stanford University

Smart Reply: Automated Response Suggestion for Email
Anjuli Kannan, ; Karol Kurach, Google; Sujith Ravi, Google; Tobias Kaufmann, Google, Inc.; Andrew Tomkins, ; Balint Miklos, Google, Inc.; Greg Corrado, ; László Lukács, ; Marina Ganea, ; Peter Young, ; Vivek Ramavajjala,

DeepIntent: Learning Attentions for Online Advertising with Recurrent Neural Networks
Shuangfei Zhai, Binghamton University; Keng-hao Chang, Microsoft; Ruofei Zhang, Microsoft; Zhongfei Zhang,

ABRA: Approximating Betweenness Centrality in Static and Dynamic Graphs with Rademacher Averages
Matteo Riondato, Two Sigma Investments; Eli Upfal, Brown University

Tuesday 10:00am - 12:00pm
Applied Data Science Invited Talks 3, Yosemite
Chair: Evangelos Simoudis

Accelerating the Race to Autonomous Cars
Danny Shapiro: Senior Director of Automotive / NVIDIA

Profiling Users from Online Social Behaviors, with applications in Tencent Social Ads
Ching Law: GM Social Ads / Tencent

LearningSparse Models at Scale
Ralf Herbrich: Director of Machine Learning / Amazon

Tuesday 10:00am - 12:00pm
Applied Data Science Session A4: Recommendations and Ranking, Cyril Magnin (Parc)
Chair: Hema Raghavan

Collaborative Knowledge Base Embedding for Recommender Systems
Fuzheng Zhang, Microsoft; Nicholas Jing Yuan, Microsoft Research; Defu Lian, ; Xing Xie, Microsoft Research; Wei-Ying Ma

Contextual Intent Tracking for Personal Assistants
Yu Sun, University of Melbourne; Nicholas Jing Yuan, Microsoft Research; Yingzi Wang, Microsoft Research; Xing Xie, Microsoft Research; Kieran McDonald, Microsoft Corporation; Rui Zhang, University of Melbourne
Ranking Relevance in Yahoo Search
Dawei Yin, Yahoo Labs; Yuening Hu; Jiliang Tang, Yahoo Labs; Tim Daly, Yahoo; Mianwei Zhou, Yahoo Inc; Hua Ouyang; Jianhui Chen, Yahoo; Changsung Kang, Yahoo Labs; Hongbo Deng, Yahoo; Chikashi Nobata; Jean-Marc Langlois; Yi Chang, Yahoo Labs

An Empirical Study on Recommendation with Multiple Types of Feedback
Liang Tang, LinkedIn; Bo Long, LinkedIn; Bee-Chung Chen, LinkedIn; Deepak Agarwal, LinkedIn

Minimizing Legal Exposure for High-Tech Companies through Collaborative Filtering Methods
Bo Jin, Dalian University of Technology; Chao Che, Dalian University; Kuifei Yu, Zhigu Inc.; Yue Qu, Dalian University of Technology; Li Guo, Dalian University of Technology; Cuili Yao, Dalian University of Technology

Tuesday 12:00pm - 1:30pm Business Lunch

Tuesday 1:45pm - 3:45pm
Research Session R9: Unsupervised Learning and Anomaly Detection, Plaza AB
Chair: Sanjay Chawla

Modeling Precursors for Event Forecasting via Nested Multi-Instance Learning
Yue Ning, Virginia Tech; Sathappan Muthiah, Virginia Tech; Huzefa Rangwala, George Mason University; Naren Ramakrishnan, Virginia Tech

Ranking Causal Anomalies via Temporal and Dynamical Analysis on Vanishing Correlations
Wei Cheng, NEC Labs America; Kai Zhang, NEC Labs America; Haifeng Chen, NEC Research Lab; Guofei Jiang, NEC Labs America; Wei Wang, UC Los Angeles

Overcoming key weaknesses of Distance-based Neighbourhood Methods using a Data Dependent Dissimilarity Measure
Ting Kai Ming, Federation University; Ye Zhu, Monash University; Mark Carman, Monash University; Yue Zhu, Nanjing University

Partial Label Learning via Feature-Aware Disambiguation
Min-Ling Zhang, Southeast University; Binbin Zhou, Southeast University; Xu-Ying Liu, Southeast University

Semi-Markov Switching Vector Autoregressive Model-based Anomaly Detection in Aviation Systems
Igor Melnyk, University of Minnesota; Arindam Banerjee, University of Minnesota; Bryan Matthews, NASA Ames Research Center; Nikunj Oza, NASA Ames Research Center

Tuesday 1:45pm - 3:45pm
Research Session R10: Recommender Systems, Imperial Ballroom
Chair: Xavier Amatriain

Towards Conversational Recommender Systems
Konstantina Christakopoulou, University of Minnesota; Katja Hofmann, Microsoft; Filip Radlinski, Microsoft
When Social Influence Meets Item Inference
Hui-Ju Hung, Pennsylvania State University; Hong-Han Shuai, Academia Sinica; De-Nian Yang, Academic Sinica; Liang-Hao Huang, Academia Sinica; Wang-Chien Lee, The Pennsylvania State University; Jian Pei, Simon Fraser University; Ming-Syan Chen, Academia Sinica

Point-of-Interest Recommendations: Learning Potential Check-ins from Friends
Yong Ge, UNC Charlotte; Huayu Li, University of North Carolina a; Hengshu Zhu, Baidu Inc.

The Limits of Popularity-Based Recommendations, and the Role of Social Ties
Marco Bressan, Sapienza University of Rome; Stefano Leucci, Sapienza University of Rome; Alessandro Panconesi, ; Prabhakar Raghavan, Google; Erisa Terolli, Sapienza University of Rome

Goal-Directed Inductive Matrix Completion
Si Si, UT Austin; Kai-Yang Chiang, UT Austin; Cho-Jui Hsieh, UT Austin; Nikhil Rao, Technicolor Research; Inderjit Dhillon, UTexas

Tuesday 1:45pm - 3:45pm
Applied Data Science Invited Panels, Yosemite
Chair: Usama Fayyad

Big Data Needs Big Dreamers: Lessons from successful Big Data investors
Moderator: Evangelos Simoudis: Managing Partner / Synapse Partners

Can Tools Effectively Unleash the Power of Big Data?
Moderator: Usama Fayyad: Chief Data Officer & Group Managing Director / Barclays

Tuesday 1:45pm - 3:45pm
Applied Data Science Session A5: Social Networks and Social Media, Cyril Magnin (Parc)
Chair: Jeffrey Pan

Audience Expansion for Online Social Network Advertising
Haishan Liu, LinkedIn Corporation; David Pardoe, LinkedIn Corporation; Kun Liu, LinkedIn Corporation.

Engagement Capacity and Engaging Team Formation for Reach Maximization of Online Social Media Platforms
Alexander Nikolaev, University at Buffalo; Shounak Gore, University at Buffalo; Venu Govindaraju, University at Buffalo

Email Volume Optimization at LinkedIn
Rupesh Gupta, LinkedIn; Xiaoyu Chen; Guanfeng Liang; Romer Rosales, LinkedIn; Hsiao-Ping Tseng; Ravi Kiran Holur Vijay

Identifying Decision Makers from Professional Social Networks
Shipeng Yu, LinkedIn; Evangelia Christakopoulou, University of Minnesota; Abhishek Gupta, LinkedIn
EMBERS at 4 years: Experiences operating an Open Source Indicators Forecasting System
Sathappan Muthiah, Virginia Tech; Naren Ramakrishnan, Virginia Tech; Patrick Butler, Virginia Tech; Rupinder Khandpur, Virginia Tech; Parang Saraf, Virginia Tech; Anil Vullikanti, Virginia Tech; Achla Marathe, Virginia Tech; Graham Katz, CACI; Andrew Doyle, CACI; Jaime Arredondo, UCSD; Dipak Gupta, SDSU; David Mares, UCSD; Jose Cadena, Virginia Tech; Liang Zhao, VT; Nathan Self, ; Alla Rozovskaya, Virginia Tech; Kristen Summers, IBM

Tuesday 3:45pm - 4:15pm, Coffee Break

Tuesday 4:15pm - 6:15pm, Grand Ballroom
Turing Talk: The Evolving Meaning of Information Security
Speaker: Whitfield Diffie Opening Introduction by Alex Wolf
Abstract: When you are developing security systems, new penetration techniques seem to appear as responses to new security measures but in general the flow is the other way around: security exists and evolves because of the evolution of threats. Beginning with the rise of radio in the 20th Century attacks on communication networks have shown two forms: those that go for the big kill—such as the breaking of Enigma—and those that assemble small seemingly innocuous leaks of information into a comprehensive understanding of the target’s behavior.
We will analyze the way in which these trends interact with others to create a situation in which what is possible in security and even the meaning of security in communication networks needs reexamination.

Tuesday August 16, 6:00pm - 7:00pm Poster Set-Up (Research Track, Poster Papers) – Golden Gate

Tuesday August 16, 7:00pm - 12:00am Poster Session: Research Track (Poster Papers) – Golden Gate

Bayesian Inference of Arrival Rate and Substitution Behavior from Sales Transaction Data with Stockouts
Benjamin Letham (Massachusetts Institute of Technology), Lydia M. Letham (Massachusetts Institute of Technology), Cynthia Rudin (Duke University)

Collaborative Multi-View Denoising
Lei Zhang (Chinese Academy of Sciences), Shupeng Wang (Chinese Academy of Sciences), Xiaoyu Zhang (Chinese Academy of Sciences), Yong Wang (Chinese Academy of Sciences), Binbin Li (Chinese Academy of Sciences), Dinggang Shen (University of North Carolina), Shuiwang Ji (Washington State University)

Deep Visual-Semantic Hashing for Cross-Modal Retrieval
Yue Cao (Tsinghua University), Mingsheng Long (Tsinghua University), Jianmin Wang (Tsinghua University), Qiang Yang (Hong Kong University of Science and Technology), Philip S. Yu (University of Illinois at Chicago & Tsinghua University)

Infinite Ensemble for Image Clustering
Hongfu Liu (Northeastern University), Ming Shao (Northeastern University), Sheng Li (Northeastern University), Yun Fu (Northeastern University)

Dynamics of Large Multi-View Social Networks: Synergy, Cannibalization and Cross-View Interplay
Yu Shi (University of Illinois at Urbana-Champaign), Myunghwan Kim (LinkedIn Corporation), Shaunak Chatterjee (LinkedIn Corporation), Mitul Tiwari (LinkedIn Corporation), Souvik Ghosh (LinkedIn Corporation), Romer Rosales (LinkedIn Corporation)
Meta Structure: Computing Relevance in Large Heterogeneous Information Networks
Zhipeng Huang (The University of Hong Kong), Yudian Zheng (The University of Hong Kong), Reynold Cheng (The University of Hong Kong), Yizhou Sun (Northeastern University), Nikos Mamoulis (The University of Hong Kong), Xiang Li (The University of Hong Kong)

A Truth Discovery Approach with Theoretical Guarantee
Houping Xiao (The State University of New York at Buffalo), Jing Gao (The State University of New York at Buffalo), Zhaoran Wang (Princeton University), Shiyu Wang (University of Illinois at Urbana-Champaign), Lu Su (The State University of New York at Buffalo), Han Liu (Princeton University)

Predicting Socio-Economic Indicators using News Events
Sunandan Chakraborty (New York University), Ashwin Venkataraman (New York University), Srikanth Jagabathula (New York University), Lakshminarayanan Subramanian (New York University)

Convex Optimization for Linear Query Processing under Approximate Differential Privacy
Ganzhao Yuan (South China University of Technology), Yin Yang (Hamad Bin Khalifa University), Zhenjie Zhang (Illinois at Singapore Pte. Ltd.), Zhifeng Hao (Foshan University)

Accelerated Stochastic Block Coordinate Descent with Optimal Sampling
Aston Zhang (University of Illinois at Urbana-Champaign), Quanquan Gu (University of Virginia)

Lightweight Monitoring of Distributed Streams
Arnon Lazerson (Technion -- Israel Institute of Technology), Daniel Keren (Haifa University), Assaf Schuster (Technion -- Israel Institute of Technology)

Distributing the Stochastic Gradient Sampler for Large-Scale LDA
Yuan Yang (Beihang University), Jianfei Chen (Tsinghua University), Jun Zhu (Tsinghua University)

Label Noise Reduction in Entity Typing by Heterogeneous Partial-Label Embedding
Xiang Ren (Univ. of Illinois at Urbana-Champaign), Wenqi He (Univ. of Illinois at Urbana-Champaign), Meng Qu (Univ. of Illinois at Urbana-Champaign), Clare R. Voss (Army Research Laboratory), Heng Ji (Rensselaer Polytechnic Institute), Jiawei Han (Univ. of Illinois at Urbana-Champaign)

Multi-layer Representation Learning for Medical Concepts
Edward Choi (Georgia Institute of Technology), Mohammad Taha Bahadori (Georgia Institute of Technology), Elizabeth Searles (Children's Healthcare of Atlanta), Catherine Coffey (Children's Healthcare of Atlanta), Michael Thompson (Children's Healthcare of Atlanta), James Bost (Children's Healthcare of Atlanta), Javier Tejedor-Sojo (Children's Healthcare of Atlanta), Jimeng Sun (Georgia Institute of Technology)

Efficient Shift-Invariant Dictionary Learning

Beyond Sigmoids: The NetTide Model for Social Network Growth, and Its Applications
Chengxi Zang (Department of Computer Science and Technology, Tsinghua), Peng Cui (Department of Computer Science and Technology, Tsinghua), Christos Faloutsos (Computer Science Department, Carnegie Mellon University)

Parallel Dual Coordinate Descent Method for Large-scale Linear Classification in Multi-core Environments
Wei-Lin Chiang (National Taiwan University), Mu-Chu Lee (National Taiwan University), Chih-Jen Lin (National Taiwan University)
Absolute Fused Lasso and Its Application to Genome-Wide Association Studies
Tao Yang (Arizona State University), Jun Liu (SAS Institute Inc.), Pinghua Gong (University of Michigan), Ruiwen Zhang (SAS Institute Inc.), Xiaotong Shen (University of Minnesota), Jieping Ye (University of Michigan)

Efficient Processing of Network Proximity Queries via Chebyshev Acceleration
Mustafa Coskun (Case Western Reserve University), Ananth Grama (Purdue University), Mehmet Koyuturk (Case Western Reserve University)

FLASH: Fast Bayesian Optimization for Data Analytic Pipelines
Yuyu Zhang (Georgia Institute of Technology), Mohammad Taha Bahadori (Georgia Institute of Technology), Hang Su (Georgia Institute of Technology), Jimeng Sun (Georgia Institute of Technology)

Transfer Knowledge between Cities
Ying Wei (Hong Kong University of Science and Technology), Yu Zheng (Microsoft Research), Qiang Yang (Hong Kong University of Science and Technology)

Towards Confidence in the Truth: A Bootstrapping based Truth Discovery Approach
Houping Xiao (The State University of New York at Buffalo), Jing Gao (The State University of New York at Buffalo), Qi Li (The State University of New York at Buffalo), Fenglong Ma (The State University of New York at Buffalo), Lu Su (The State University of New York at Buffalo), Yunlong Feng (Katholieke Universiteit Leuven), Aidong Zhang (The State University of New York at Buffalo)

Online Optimization Methods for the Quantification Problem
Purushottam Kar (Indian Institute of Technology Kanpur), Shuai Li (University of Insubria), Harikrishna Narasimhan (Harvard University), Sanjay Chawla (Qatar Computing Research Institute, HBKU), Fabrizio Sebastiani (Qatar Computing Research Institute, HBKU)

City-Scale Map Creation and Updating using GPS Collections
Chen Chen (Stanford University), Cewu Lu (Stanford University), Qixing Huang (Toyota Technological Institute at Chicago), Qiang Yang (Hong Kong University of Science and Technology), Dimitrios Gunopulos (University of Athens), Leonidas Guibas (Stanford University)

Multi-Task Feature Interaction Learning
Kaixiang Lin (Michigan State University), Jianpeng Xu (Michigan State University), Inci M. Baytas (Michigan State University), Shuiwang Ji (Washington State University), Jiayu Zhou (Michigan State University)

Learning Cumulatively to Become More Knowledgeable
Geli Fei (University of Illinois at Chicago), Shuai Wang (University of Illinois at Chicago), Bing Liu (University of Illinois at Chicago)

Fast Unsupervised Online Drift Detection Using Incremental Kolmogorov-Smirnov Test
Denis Moreira dos Reis (Universidade de São Paulo), Peter Flach (University of Bristol), Stan Matwin (Polish Academy of Sciences), Gustavo Batista (ICMC-USP (Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo))

From Truth Discovery to Trustworthy Opinion Discovery: An Uncertainty-Aware Quantitative Modeling Approach
Mengting Wan (University of California, San Diego), Xiangyu Chen (University of Illinois, Urbana-Champaign), Lance Kaplan (U.S. Army Research Laboratory), Jiawei Han (University of Illinois, Urbana-Champaign), Jing Gao (SUNY Buffalo), Bo Zhao (LinkedIn)
Causal Clustering for 1-Factor Measurement Models
Erich Kummerfeld (University of Pittsburgh), Joseph Ramsey (Carnegie Mellon University)

Compressing Convolutional Neural Networks in the Frequency Domain
Wenlin Chen (Washington University in St. Louis), James Wilson (University of Edinburgh), Stephen Tyree (Nvidia), Kilian Q Weinberger (Cornell University), Yixin Chen (Washington University in St. Louis)

Subjectively Interesting Component Analysis: Data Projections that Contrast with Prior Expectations
Bo Kang (Ghent University), Jefrey Lijffijt (Ghent University), Raúl Santos-Rodríguez (University of Bristol), Tijl De Bie (Ghent University)

A Text Clustering Algorithm Using an Online Clustering Scheme for Initialization
Jianhua Yin (Tsinghua University), Jianyong Wang (Tsinghua University)

A Real Linear and Parallel Multiple Longest Common Subsequences (MLCS) Algorithm
Yanni Li (Xidian University), Hui Li (Xidian University), Tihua Duan (Shanghai Finance University), Sheng Wang (Conventry University), Zhi Wang (Xidian University), Yang Cheng (Xidian University)

Probabilistic Robust Route Recovery with Spatio-Temporal Dynamics
Hao Wu (Fudan University), Jiangyun Mao (Fudan University), Weiwei Sun (Fudan University), Bihua Zheng (Singapore Management University), Hanyuan Zhang (Fudan University), Ziyang Chen (Fudan University), Wei Wang (Fudan University)

Safe Pattern Pruning: An Efficient Approach for Predictive Pattern Mining
Kazuya Nakagawa (Nagoya Institute of Technology), Shinya Suzumura (Nagoya Institute of Technology), Masayuki Karasuyama (Nagoya Institute of Technology), Koji Tsuda (University of Tokyo), Ichiro Takeuchi (Nagoya Institute of Technology)

FUSE: Full Spectral Clustering
Wei Ye (Ludwig-Maximilians-Universität München), Sebastian Goebl (Ludwig-Maximilians-Universität München), Claudia Plant (University of Vienna), Christian Böhm (Ludwig-Maximilians-Universität München)

MANTRA: A Scalable Approach to Mining Temporally Anomalous Sub-trajectories
Prithu Banerjee (University of British Columbia), Pranali Yawalkar (IIT Madras), Sayan Ranu (IIT Madras)

Hierarchical Incomplete Multi-source Feature Learning for Spatiotemporal Event Forecasting
Liang Zhao (Virginia Tech), Jieping Ye (University of Michigan), Cheng Chen (University at Albany, SUNY), Chang-Tien Lu (Virginia Tech), Naren Ramakrishnan (Virginia Tech)

Diversified Temporal Subgraph Pattern Mining
Yi Yang (Fudan University), Da Yan (The Chinese University of Hong Kong), Huanhuan Wu (The Chinese University of Hong Kong), James Cheng (The Chinese University of Hong Kong), Shuigeng Zhou (Fudan University), John C.S. Lui (The Chinese University of Hong Kong)

User Identity Linkage by Latent User Space Modelling
Xin Mu (Nanjing University), Feida Zhu (Singapore Management University), Ee-Peng Lim (Singapore Management University), Jing Xiao (Ping An Technology (Shenzhen) Co., Ltd), Jianzong Wang (Ping An Technology (Shenzhen) Co., Ltd), Zhi-Hua Zhou (Nanjing University)

Unbounded Human Learning: Optimal Scheduling for Spaced Repetition
Siddharth Reddy (Cornell University), Igor Labutov (Cornell University), Siddhartha Banerjee (Cornell University), Thorsten Joachims (Cornell University)

Fast Component Pursuit for Large-Scale Inverse Covariance Estimation
Lei Han (Rutgers University), Yu Zhang (Hong Kong University of Science and Technology), Tong Zhang (Rutgers University)

Burstiness Scale: A Parsimonious Model for Characterizing Random Series of Events
Rodrigo Augusto da Silva Alves (CEFET-MG), Renato Martins Assuncao (UFMG), Pedro Olmo Stancioli Vaz de Melo (UFMG)

Robust and Effective Metric Learning Using Capped Trace Norm
Zhouchuan Huo (University of Texas at Arlington), Feiping Nie (University of Texas at Arlington), Heng Huang (University of Texas at Arlington)

Online Asymmetric Active Learning with Imbalanced Data
Xiaoxuan Zhang (The University of Iowa), Tianbao Yang (The University of Iowa), Padmini Srinivasan (The University of Iowa)

Finding Gangs in War from Signed Networks
Lingyang Chu (Simon Fraser University), Zhefeng Wang (University of Science and Technology of China), Jian Pei (Simon Fraser University), Jiannan Wang (Simon Fraser University), Zijin Zhao (Simon Fraser University), Enhong Chen (University of Science and Technology of China)

Scalable Pattern Matching over Compressed Graphs via Dedensification
Antonio Maccioni (Roma Tre University), Daniel J. Abadi (Yale University)

A Multi-Task Learning Formulation for Survival Analysis
Yan Li (Wayne State University), Jie Wang (University of Michigan), Jieping Ye (University of Michigan), Chandan K Reddy (Wayne State University)

Lossless Separation of Web Pages into Layout Code and Data
Adi Omari (Technion), Benny Kimelfeld (Technion), Sharon Shoham (Tel Aviv University), Eran Yahav (Technion)

Parallel Lasso Screening for Big Data Optimization
Qingyang Li (Arizona State University), Shuang Qiu (University of Michigan), Shuiwang Ji (Washington State University), Paul M Thompson (University of Southern California), Jieping Ye (University of Michigan), Jie Wang (University of Michigan)

Latent Space Model for Road Networks to Predict Time-Varying Traffic
Dingxiong Deng (University of Southern California), Cyrus Shahabi (University of Southern California), Ugur Demiryurek (University of Southern California), Lanhong Zhu (Information Sciences Institute), Rose Yu (University of Southern California), Yan Liu (University of Southern California)

Portfolio Selections in P2P Lending: A Multi-Objective Perspective
Hongke Zhao (University of Science and Technology of China), Qi Liu (University of Science and Technology of China), Guifeng Wang (University of Science and Technology of China), Yong Ge (Eller College of Management, University of Arizona), Enhong Chen (University of Science and Technology of China)

Compressing Graphs and Indexes with Recursive Graph Bisection
Laxman Dhulipala (Carnegie Mellon University), Igor Kabiljo (Facebook), Brian Karrer (Facebook), Giuseppe Ottaviano (Facebook), Sergey Pupyrev (Facebook), Alon Shalita (Facebook)
The Million Domain Challenge: Broadcast Email Prioritization by Cross-domain Recommendation
Beidou Wang (Zhejiang University and Simon Fraser University), Martin Ester (Simon Fraser University), Yikang Liao (Zhejiang University), Jiajun Bu (Zhejiang University), Yu Zhu (Zhejiang University), Ziyu Guan (Northwest University of China), Deng Cai (Zhejiang University)

Scalable Partial Least Squares Regression on Grammar-Compressed Data Matrices
Yasuo Tabei (Japan Science and Technology Agency), Hiroto Saigo (Kyushu University), Yoshihiro Yamanishi (Kyushu University), Simon J Puglisi (University of Helsinki)

Reconstructing an Epidemic Over Time
Polina Rozenshtein (Aalto University), Aristides Gionis (Aalto University), B. Aditya Prakash (Virginia Tech), Jilles Vreeken (Saarland University and Max Planck Institute for Informatics)

Data-driven Automatic Treatment Regimen Development and Recommendation
Leilei Sun (Dalian University of Technology), Chuanren Liu (Drexel University), Chonghui Guo (Dalian University of Technology), Hui Xiong (Rutgers University), Yanming Xie (China Academy of Chinese Medical Sciences)

Squish: Near-Optimal Compression for Archival of Relational Datasets
Yihan Gao (University of Illinois at Urbana-Champaign), Aditya Parameswaran (University of Illinois at Urbana-Champaign)

Smart Broadcasting: Do You Want to be Seen?
Mohammad Reza Karimi (Sharif University), Erfan Tavakoli (Sharif University), Mehrdad Farajtabar (Georgia Institute of Technology), Le Song (Georgia Institute of Technology), Manuel Gomez Rodriguez (Max Planck Institute for Software Systems)

Scalable Betweenness Centrality Maximization via Sampling
Ahmad Mahmoody (Brown University), Charalampos E. Tsourakakis (Harvard University), Eli Upfal (Brown University)

Predict Risk of Relapse for Patients with Multiple Stages of Treatment of Depression
Zhi Nie (Arizona State University), Pinghua Gong (University of Michigan, Ann Arbor), Jieping Ye (University of Michigan, Ann Arbor)

From Prediction to Action: A Closed-Loop Approach for Data-Guided Network Resource Allocation
Yanan Bao (University of California, Davis), Huasen Wu (University of California, Davis), Xin Liu (University of California, Davis)

Towards Robust and Versatile Causal Discovery for Business Applications
Giorgos Borboudakis (University of Crete; Gnosis Data Analysis IKE), Ioannis Tsamardinos (University of Crete; Gnosis Data Analysis IKE)

How to Compete Online for News Audience: Modeling Words that Attract Clicks
Joon Hee Kim (Korea Advanced Institute of Science and Technology), Amin Mantrach (Yahoo! Research, Sunnyvale), Alejandro Jaimes (AiCure), Alice Oh (Korea Advanced Institute of Science and Technology)

Interpretable Decision Sets: A Joint Framework for Description and Prediction
Himabindu Lakkaraju (Stanford University), Stephen H. Bach (Stanford University), Jure Leskovec (Stanford University)

Recurrent Marked Temporal Point Processes: Embedding Event History to Vector
Nan Du (Georgia Institute of Technology), Hanjun Dai (Georgia Institute of Technology), Rakshit Trivedi (Georgia Institute of Technology), Utkarsh Upadhyay (Max Planck Institute for Software Systems), Manuel Gomez-Rodriguez (Max Planck Institute for Software Systems), Le Song (Georgia Institute of Technology)

Improving Survey Aggregation with Sparsely Represented Signals
Tianlin Shi (Stanford University), Forest Agostinelli (University of California - Irvine), Matthew Staib (Massachusetts Institute of Technology), David Wipf (Microsoft Research), Thomas Moscibroda (Microsoft Research)

Optimal Reserve Prices in Upstream Auctions: Empirical Application on Online Video Advertising
Miguel Angel Alcobendas Lisbona (Yahoo! Inc.), Sheide Chammas (Yahoo! Inc.), Kuang-chih Lee (Yahoo! Inc.)

Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit
Chunqiu Zeng (Florida International University), Qing Wang (Florida International University), Shekoofeh Mokhtari (Florida International University), Tao Li (Florida International University)

Optimally Discriminative Choice Sets in Discrete Choice Models: Application to Data-Driven Test Design
Igor Labutov (Cornell University), Frans Schalekamp (Cornell University), Kelvin Luu (University of Washington), Hod Lipson (Columbia University), Christoph Studer (Cornell University)

Topic Modeling of Short Texts: A Pseudo-Document View
Yuan Zuo (Beihang University), Junjie Wu (Beihang University), Hui Zhang (Beihang University), Hao Lin (Beihang University), Fei Wang (Beihang University), Ke Xu (Beihang University), Hui Xiong (Rutgers, the State University of New Jersey)

Online Feature Selection: A Limited-Memory Substitution Algorithm and Its Asynchronous Parallel Variation
Haichuan Yang (University of Rochester), Ryohei Fujimaki (NEC), Yukitaka Kusumura (NEC), Ji Liu (University of Rochester)
Wednesday August 17, 2016 Detailed Program

Wednesday 8:15am - 3:00pm, Registration, Registration Desk - East Lounge
Wednesday 9:15am - 1:30pm, Exhibit Hall - Continental

Wednesday 8:00am - 9:30am, Grand Ballroom
Keynote Session 5: Learning to learn and compositionality with deep recurrent neural networks
Speaker: Nando de Freitas chaired by Alex Smola
Abstract: Deep neural network representations play an important role in computer vision, speech, computational linguistics, robotics, reinforcement learning and many other data-rich domains. In this talk I will show that learning-to-learn and compositionality are key ingredients for dealing with knowledge transfer so as to solve a wide range of tasks, for dealing with small-data regimes, and for continual learning. I will demonstrate this with three examples: learning learning algorithms, neural programmers and interpreters, and learning communication.

Wednesday 9:30am-10:00am, Coffee Break

Wednesday 10:00am - 12:00pm
Research Session R11: Supervised Learning, Plaza AB
Chair: Quanquan Gu

Extreme Multi-label Loss Functions for Recommendation, Tagging, Ranking & Other Missing Label Applications
Himanshu Jain, IIT Delhi; Yashoteja Prabhu; Manik Varma

Annealed Sparsity via Adaptive and Dynamic Shrinking
Kai Zhang, NEC labs America; Shandian Shan, Purdue University; Zhengzhang Chen, NEC Lab America; Chaoran Cheng, New Jersey Institute of Technology; Zhi Wei, New Jersey Institute of Technology; Guofei Jiang, NEC Lab America; Jieping Ye

"Why Should I Trust you?" Explaining the Predictions of Any Classifier
Marco Tulio Ribeiro, University of Washington; Sameer Singh, University of Washington, Seattle; Carlos Guestrin, Dato/Univ of Washington

Robust Extreme Multi-label Learning
Chang Xu, Peking University; Dacheng Tao, University of Technology Sydney; Chao Xu, Peking University

Privacy-preserving Class Ratio Estimation
Arun Iyer; Saketh Nath, IIT Bombay; Sunita Sarawagi, IIT Bombay

Wednesday 10:00am - 12:00pm
Research Session R12: Sequence Mining, Imperial Ballroom
Chair: Leman Akoglu
Mining Subgroups with Exceptional Transition Behavior
Florian Lemmerich, Gesis; Martin Becker, University of Würzburg; Philipp Singer, Gesis; Denis Helic, TU Graz; Andreas Hotho, University of Wuerzburg; Markus Strohmaier

Lexis: An Optimization Framework for Discovering the Hierarchical Structure of Sequential Data
Payam Siyari, Georgia Institute of Technology; Bistra Dilkina, Georgia Tech; Constantine Dovrolis, Georgia Institute of Technology

Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences
Yasuko Matsubara, Kumamoto University; Yasushi Sakurai, Kumamoto University

Keeping it Short and Simple: Summarising Complex Event Sequences with Multivariate Patterns
Roel Bertens, Universiteit Utrecht; Jilles Vreeken, Max-Planck Institute for Informatics and Saarland University; Arno Siebes

Unified Point-of-Interest Recommendation with Temporal Interval Assessment
Yanchi Liu, Rutgers University; Chuanren Liu, Drexel University; Bin Liu, Rutgers University; Meng Qu, Rutgers University; Hui Xiong, Rutgers

Wednesday 10:00am - 12:00pm
Applied Data Science Invited Talks 5, Yosemite
Chair: Rajesh Parekh

It’s About Time
Caitlin Smallwood: Vice President of Science and Algorithms / Netflix

TBD
Andy Palmer: Co-Founder and CEO / Tamr, Inc.

Wednesday 10:00am - 12:00pm
Applied Data Science Session A6: Machine Learning Algorithms, Cyril Magnin (Parc)
Chair: Ralf Herbrich

Matrix Computations and Optimization in Apache Spark
Reza Zadeh, Stanford University; Xiangrui Meng; Alexander Ulanov; Burak Yavuz; Li Pu; Shivaram Venkataraman; Evan Sparks; Aaron Staple; Matei Zaharia

GLMix: Generalized Linear Mixed Models For Large-Scale Response Prediction
XianXing Zhang, LinkedIn; Bee-Chung Chen, LinkedIn; Liang Zhang, LinkedIn; Yitong Zhou, LinkedIn Corporation; Yiming Ma, LinkedIn; Deepak Agarwal, LinkedIn

Boosted Decision Tree Regression Adjustment for Variance Reduction of Online Controlled Experiments
Alexey Poyarkov, Yandex; Alexey Drutsa, Yandex; Andrey Khalyavin, Yandex; Gleb Gusev, Yandex; Pavel Serdyukov, Yandex

Online dual decomposition for performance and delivery-based distributed ad allocation
Jim Huang, Amazon; Rodolphe Jenatton, Amazon; Cedric Archambeau, Amazon
Deep Crossing: Web-Scale Modeling without Manually Crafted Combinatorial Features
Ying Shan, Microsoft Corporation; Thomas Hoens, Microsoft; Jian Jiao, Microsoft Corporation; Haijing Wang, Microsoft Corporation; Dong Yu, Microsoft Research; JC Mao, Microsoft Corporation

Wednesday 12:00pm - 1:00pm Lunch -- on your own

Wednesday 1:00pm - 3:00pm
Research Session R13: User Behavior Modeling, Plaza AB
Chair: Hanghang Tong

Predicting Matchups and Preferences in Context
Shuo Chen, Cornell; Thorsten Joachims, Cornell University

Just One More: Modeling Binge Watching Behavior
William Trouleau, EPFL; Azin Ashkan, Technicolor; Weicong Ding, Technicolor Research; Brian Eriksson, Technicolor

Taxi Driving Behavior Analysis in Latent Vehicle-to-Vehicle Networks: A Social Influence Perspective
Tong Xu, USTC; Hengshu Zhu, Baidu Inc.; Xiangyu Zhao, USTC; Hao Zhong, Rutgers University; Qi Liu, University of Science and Technology of China; Enhong Chen, ; Hui Xiong, Rutgers

Assessing Human Error Against a Benchmark of Perfection
Ashton Anderson, Stanford University; Jon Kleinberg, Cornell University; Sendhil Mullainathan, Harvard

Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization
Junming Liu, Rutgers University; Leilei Sun; Hui Xiong, Rutgers; Weiwei Chen

Wednesday 1:00pm - 3:00pm
Research Session R14: Optimization, Imperial Ballroom
Chair: Jiliang Tang

Generalized Hierarchical Sparse Model for Arbitrary-Order Interactive Antigenic Sites Identification in Flu Virus Data
Lei Han, Rutgers University; Yu Zhang, Hong Kong University of Science and Technology; Xiu-Feng Wan, Mississippi State University; Tong Zhang, Rutgers University

Accelerating Online CP Decompositions for Higher Order Tensors
Shuo Zhou, University of Melbourne; Nguyen Vinh, University of Melbourne; James Bailey, ; Yunzhe Jia, University of Melbourne; Ian Davidson, University of California-Davis

Robust Large-Scale Machine Learning in the Cloud
Steffen Rendle, Google; Dennis Fetterly, Google, Inc.; Eugene Shekita, Google, Inc.; Bor-Yiing Su, Google, Inc.

Revisiting Random Binning Feature: Fast Convergence and Strong Parallelizability
Lingfei Wu, College of William and Mary; En-Hsu Yen, University of Texas at Austin; Jie Chen, IBM Research; Rui Yan, Baidu Inc.

Efficient Frequent Directions Algorithm for Sparse Matrices
Wednesday 1:00pm - 3:00pm
Applied Data Science Session A7: E-commerce, Yosemite
Chair: Roberto Bayardo

Repeat Buyer Prediction for E-Commerce
Guimei Liu; Tam T. Nguyen, Institute for Infocomm Research; Gang Zhao, Development Bank of Singapore; Wei Zha, Institute for Infocomm Research; Jianbo Yang, General Electric; Jianneng Cao, Institute for Infocomm Research; Min Wu, Institute for Infocomm Research; Peilin Zhao, Institute for Infocomm Research, A*STAR; Wei Chen, Development Bank of Singapore

Large-Scale Item Categorization in e-Commerce Using Multiple Recurrent Neural Networks
Hyuna Pyo, NAVER LABS; Jung-Woo Ha, NAVER LABS; Jeonghee Kim, NAVER LABS

From Online Behaviors to Offline Retailing
Ping Luo, Chinese Academy of Sciences

An Engagement-Based Customer Lifetime Value System for E-commerce
Ali Vanderveld, Groupon; Angela Han, Groupon; Addhyan Pandey, Groupon; Rajesh Parekh

Predictors without Borders: Behavioral Modeling of Product Adoption in Three Developing Countries
Muhammad Khan, University of Washington; Joshua Blumenstock, University of Washington

Wednesday 1:00pm - 3:00pm
Applied Data Science Session A8: Systems and Experimentation, Cyril Magnin (Parc)
Chair: Gautam Shroff

Question Independent Grading using Machine Learning: The Case of Computer Program Grading
Gursimran Singh, Aspiring Minds; Shashank Srikant; Varun Aggarwal

Predicting Disk Replacement towards Reliable Data Centers
Mirela Botezatu, IBM Research; Ioana Giurgiu, IBM Research; Jasmina Bogojeska, IBM Research; Dorothea Wiesmann, IBM Research

Data-Driven Metric Development for Online Controlled Experiments: Seven Lessons Learned
Xiaolin Shi, Yahoo Labs; Alex Deng, Microsoft

Evaluating Mobile App Release
Ya Xu, LinkedIn Corporation; Nanyu Chen, LinkedIn Corporation

Anomaly Detection Using Program Control Flow Graph Mining from Execution Logs
Animesh Nandi, IBM Research; Atri Mandal, IBM Research; Shubham Atreja, IIT Kanpur; Gargi Dasgupta, IBM Research; Subhrajit Bhattacharya, IBM Research

Wednesday 3:00pm - 3:15pm, Coffee Break

Wednesday 3:15pm - 4:00pm, KDD 2016 Closing, Grand Ballroom