Programme Booklet

Organization

General Co-Chairs:
Xueqi Cheng, ICT, Chinese Academy of Sciences, China
Hang Li, Huawei Technologies, China

Honorary chairs:
Guojie Li, ICT, Chinese Academy of Sciences, China
Hong Mei, Shanghai Jiaotong University, China

Co-Organizer:
Xuanjing Huang, Fudan University, China

Advisor:
Irwin King, Chinese University of Hong Kong, China

PC Co-Chairs:
Evgeniy Gabrilovich, Google, USA
Jie Tang, Tsinghua University, China

Local Chair:
Qi Zhang, Fudan University, China

Winter School Co-Chairs:
Minyi Guo, Shanghai Jiaotong University, China
Maosong Sun, Tsinghua University, China

Best Paper Award Committee chair:
Qiang Yang, Hong Kong University of Science and Technology, China

Workshop Co-Chairs:
Jiafeng Guo, ICT, Chinese Academy of Science, China
Yoelle Maarek, Yahoo, USA

Tutorial Co-Chairs:
Eytan Adar, University of Michigan, USA
Min Zhang, Tsinghua University, China

Doctoral Consortium Co-Chairs:
Tie-Yan Liu, Microsoft Research Asia, China
Cheng Xiang Zhai, UIUC, USA

Practice and Experience Track Co-Chairs:
Vanja Josifovski, Google, USA
Ying Li, EV Analysis Corporation, USA

Finance Chair:
Xiaolong Jin, ICT, Chinese Academy of Science, China

Sponsorship Co-Chairs:
Pavel Serdyukov, Yandex, Russia
Dou Shen, Baidu, China

Publicity Co-Chairs;
Chia-Jung Lee, UMass, USA
Xipeng Qiu, fudan University, China

Proceedings Chair:
Jing Zhang, Tsing Hua, China

Web Master:
Junming Huang, University of Electronic of Science and Technology of China, China

Winter School Organizer:
Long Zheng, Shanghai Jiaotong University, China
Local Arrangements:

WSDM Steering committee:
Ricardo Baeza-Yates, Yahoo, USA
Paolo Boldi, University of Milano, Italy
Andrei Broder, Google, USA
Brian D. Davison, Lehigh University, USA
Nick Koudas, University of Toronto, Canada
Hang Li, Huawei Technologies, China
Bing Liu, University of Illinois at Chicago, USA
Marc Najork, Google, USA

Table of Contents

Preface from General Chairs

On behalf of the organizing committee, we warmly welcome you to the 8th ACM International Conference on Web Search and Data Mining (WSDM 2015), held in Shanghai, China, from Jan. 31 to Feb. 6, 2015.

WSDM is one of the premier conferences covering research on search and data mining on the web. WSDM publishes original and high-quality papers, with an emphasis on practical yet principled novel approaches.

This is the first time WSDM is held in China, which has more than 500 million internet users and many internet companies, and more importantly, a very large number of researchers and engineers working in the field. We believe that WSDM 2015 in Shanghai, the largest city in China, will forge stronger ties between the local community and the international community of web search and data mining.

The technical program of WSDM 2015 includes 3 keynote speeches, 5 practice and experience talks, 39 technical papers, 4 workshops, 6 tutorials, a panel discussion, doctorial consortium, and winter school. The 8th edition of WSDM continues the traditional activities of the conference and it also adds a new line-up to the conference, namely winter school. The goal of the winter school is to promote research on web search and data mining and introduce the fundamentals as well as recent advance of the field to people interested in it particularly students.

We would like to express our sincere gratitude to the PC co-chairs, Evgeniy Gabrilovich and Jie Tang, who did a great job in selecting high quality papers, together with 28 senior program committee members, 220 program committee members, and 41 additional reviewers. We are also very grateful to all the other organizing committee members, including honorary

chairs Guojie Li and Hong Mei, co-organizer Xuanjing Huang, advisor Irwin King, local chair Qi Zhang, winter school chairs Minyi Guo and Maosong Sun, best paper award committee chair Qiang Yang, workshop chairs Jiafeng Guo and Yoelle Maarek, tutorial chairs Eytan Adar and Min Zhang, doctorial consortium chairs Tie-Yan Liu and Cheng Xiang Zhai, Practice and Experience Track chairs Vanja Josifovski and Ying Li, finance chair Xiaolong Jin, sponsor chairs Pavel Serdyukov and Dou Shen, publication chair Jing Zhang, publicity chair Chia-Jung Lee and Xipeng Qiu, web master Junming Huang, and winter school organizer Long Zheng, whose effective work made all the events of the conference organized properly and smoothly. We would also like to thank the local organizing team members and the student volunteers from Fudan University, Shanghai Jiaotong University, and Institute of Computing Technology, who worked effectively to make the conference run smoothly and superbly.

We gratefully acknowledge the sponsorships from a number of companies, including Baidu, Google, Microsoft, Huawei, Alibaba, Sogou, IBM Research, SpeechOcean, A9, Yahoo! Labs, Yandex, Tencent, and Facebook. Special thanks also go to Fudan University, Shanghai Jiaotong University, and Institute of Computing Technology for their generous support.

Finally, we thank all the authors, presenters, and participants of the conference, and hope that you will enjoy the conference as well as the visit to Shanghai.

Xueqi Cheng
Chinese Academy of Sciences

Hang Li
Huawei Technologies

Foreword from the Program Committee Chairs

We are delighted to welcome you to the 8th ACM International Conference on Web Search and Data Mining (WSDM 2015), held in Shanghai, China, from Jan. 31 to Feb. 6, 2015. As in the previous years, WSDM continues to be a leading forum for reporting the latest research developments in the field. We are pleased to present here the proceedings of the conference.

We received a total of 238 submissions, out of which 39 were accepted for publication in the proceedings, thus reaching an acceptance rate of 16.4% (compared to 18% last year). The accepted papers are from 44 institutions in 16 countries, making WSDM a truly international forum. This year, oral presentation slots were allocated to all papers. Yet, in order to maintain the single track model that most attendees prefer, we introduced this year short 10-minute presentations. Out of the 39 accepted papers, 18 were assigned such shorter slots, while 21 were assigned longer 20-minute slots. The type of slot was chosen by the Senior PC members and Program Chairs, based on whether the topic and the content of the paper would appeal to a broad group of conference attendees. In addition to oral presentations, authors of all accepted papers were invited to present their work in an interactive poster session.

We would like to acknowledge the tremendous work of the 28 Senior Program Committee members, 220 Program Committee members, and 41 additional reviewers. The credit for creating a high quality technical program goes to them. To recognize the hard work of the best reviewers, which rarely makes it to the public limelight, this year we created Outstanding Reviewer Awards. 3 members of the Senior Program Committee and 27 members of the Program Committee received these awards.

The technical program this year featured keynotes by prominent researchers from the industry and academia, Lada Adamic (Facebook), Michael Franklin (UC Berkeley), and Thorsten Joachims (Cornell University). We thank the keynote speakers for their illuminating talks and for sharing their insights and wisdom with the conference attendees. The program also featured 5 Practice & Experience talks, which were introduced for the first time at WSDM'14. These were presented by Tushar Chandra (Google), Rong Jin (Alibaba), Jure Leskovec (Stanford), Kaihua Zhu (Baidu), and Juchao Zhuo (Tencent). We thank the P&E speakers for taking the time off their busy schedules to attend the conference and share their experience in applying recent research results in industrial settings.

We would also like to thank the general chairs, Xueqi Cheng and Hang Li, who were wonderful partners in putting the conference program together, and provided us insightful guidance and advice all along the way.

Finally, we are grateful to the authors and attendees, who are the ones who make this conference possible, and have allowed WSDM to become a wonderful forum to advance Web research and become the premier ACM conference in the field.

We hope that you will find this program interesting and thought-provoking. Our goal was not only to give you the opportunity to share ideas with other researchers and practitioners from around the world, but also to foster new research and innovation in this fascinating Web era.

Evgeniy Gabrilovich — *Google*

Jie Tang — *Tsinghua University*

WSDM 2015 Sponsors

Platinum Sponsors

Gold Sponsors

Silver Sponsors

Bronze Sponsors

Other Sponsors

Local Information

The conference will be held in **Crowne Plaza Fudan Shanghai,** located at 199 Handan Road, Shanghai.

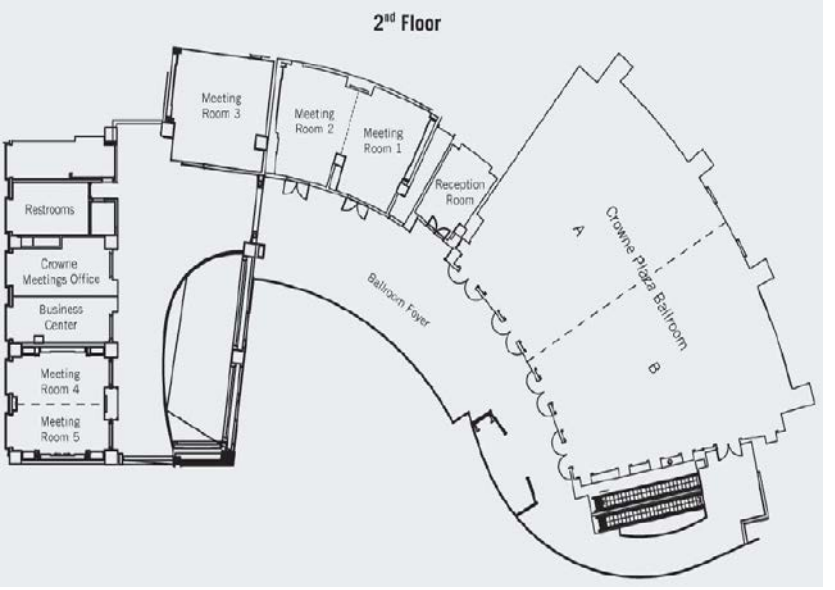
Telephone : 86 – 21 – 5552 9999

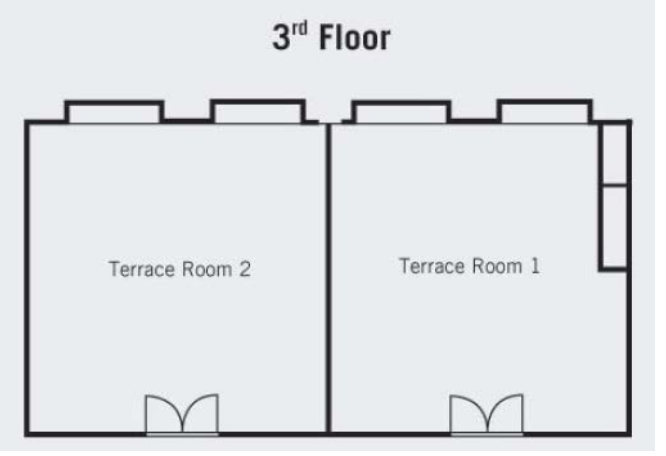
Facsimile : 86 – 21 – 5552 9990

Website : www. crowneplaza.com

About the Crowne Plaza and room numbering

As a reference: a map of meeting room floor





Connecting to the WIFI

Connect with Ease

You may need to perform the following steps from your room on a daily basis for connection in wired&wireless areas within the hotel

1. Plug the cable to your laptop
2. Launch your internet browser. If applicable, please click “Connecting for the First time” if you are the first device access to the internet in your room. For the second and more devices of internet connection in the same room. Please choose “Connecting Additional Devices”, then select the price plan and accept the terms and conditions.
3. If the “Usage Details” page is loaded, click on “Start Surfing” or key in the website directly you wrd like to surf.

For Wireless connection

1. Select the correct hotel’s wireless signal network option using your laptop’s wireless Network Connection Manager
2. Launch your internet browser. If application, select “Hotel Guest login”.
3. Enter username and password. If displayed, and accept the terms and conditions.
4. If the “Usage Details” page is loaded, click on “Start Surfing” to access the Internet or key in the website directly you would like to surf.

Tips:

1. Public area: connecting ‘**Crowneplaza’**, no password
2. Meeting room: connecting ‘Conference’, password:**WSDM2015**.

Restaurant Information

Crowne Plaza Fudan Shanghai is close to Wujiaochang (literally Five-angle District) and Daxue Road (literally University Road), where there are many shops, cafés and wonderful restaurants.

With five roads converging into a star-shaped cluster, Wujiaochang is a prosperous sub-center of the city, full of large shopping malls such as Wanda Plaza, Bai Lian You Yi Cheng and Walmart. There are especially some widely-recommended Chinese food restaurants, including Tsui Wah Restaurant, Wai Po Jia and Wang Xiang Yuan, which must be appealing to those who would like to try authentic Chinese cuisines.

Daxue Road, however, is the right place for foreign visitors seeking international cuisines around Fudan University. In addition to the beautiful street scenes, visitors can enjoy a decent full course meal in a country-featured style, while the average price is usually a little bit higher than that in Chinese food restaurants.

Listed below are some of the restaurants on Daxue Road and Wujiaochang, recommended by TripAdvisor, Shanghai WOW and a local review website Dianping.

Comprehensive Western

- MYLK Italian: Serving decent Italian food in a homelike atmosphere
- AGOGO Kitchen & Bar: An amazing restaurant in western-American style
- Pu & Rosemary: A fantastic restaurant in Mediterranean style.
- West Dyke Steak: A popular restaurant in Wujiaochang

Asian Cuisine

- Tairyo Japanese Restaurant: A traditional Japanese buffet.
- Kadomatsu Japanese Izakaya
- Guangzhou Banana Leaf: Mainly serves Southeast Asian cuisines.
- Ta Ding Thailand Fashion Restaurant

Chinese Cuisine

- Tsui Wah Restaurant, Xin Wang Restaurant: Both serve Cantonese cuisines.
- Wai Po Jia: A famous restaurant for local cuisine lovers.
- Wang Xiang Yuan: A good restaurant providing spicy food of Hunan and Sichuan cuisines.

Schedule at a Glance

	Mon. Feb 2	Tue. Feb 3	Wed. Feb 4	Thurs. Feb 5	Fri. Feb 6
08:45		Opening Session			
09:00	Tutorials & Doctoral Consortium	Keynote By Franklin	Keynote By Adamic	Keynote By Joachims	Workshops
10:00		Coffee Break	Coffee Break	Coffee Break	
10:30	Coffee Break	Session 1:PE Talk	Session 4:Web Mining	Session 7: User Modeling Mobility and Recomm-endation	Coffee Break
11:00	Tutorials & Doctoral Consortium				Workshops
11:10		Session 1: Panel			
12:00					
12:10		Lunch	Lunch	WSDM Business Lunch Meeting	
12:30	Lunch				Lunch
13:30		Session2: Web Search	Session2: Web Search	Session 8: PE Talk	
14:00	Tutorials & Doctoral Consortium				Workshops
14:50			Coffee Break	Coffee Break	

	Feb 2	Feb 3	Feb 4	Feb 5	Feb 6
15:10	Tutorials & Doctoral	Coffee Break	Coffee Break	Coffee Break	Workshops
15:20			Session 6: Crowdsourcing Temporal and Location-based mining	Session 9: Web Mining(2)	
15:30	Coffee Break				Coffee Break
15:40		Session 3: Social Networks			
16:00	Tutorials & Doctoral Consortium				Workshops
17:00				Closing Session	
17:20					
17:30					
18:00		Poster Session	Banquet		
18:30	Reception				
19:30					
20:00					

Schedule

Monday, 2th Feb

<u>Monday, 9:00-10:30</u>

Tutorials

- Dynamic Information Retrieval Modeling, *Hui Yang(Georgetown University), Marc Sloan(University College London),June Wang(University College London)*
  Room: Meeting Room 1+2 (2 F)
- Scalability and Efficiency Challenges in Large-Scale Web Search Engines*, B. Barla Cambazoglu (Yahoo Labs), Ricardo Baeza-Yates(Yahoo Labs)*
  Room: Meeting Room 3 (2 F)
- Offline Evaluation and Optimization for Interactive Systems, *Li-hong Li(Microsoft Research)*
  Room: Terrace Room 1 (3 F)

<u>Monday, 9:00-10:30</u> **Room: Terrace Room 2 (3 F)**

Doctoral consortium

<u>Monday, 10:30-11:00</u> **Room: Foyer**

Coffee Break

<u>Monday, 11:00-12:30</u>

Tutorials:

- Dynamic Information Retrieval Modeling, *Hui Yang*

(Georgetown University), Marc Sloan(University College London),June Wang(University College London)

Room: Meeting Room 1+2 (2 F)

- Scalability and Efficiency Challenges in Large-Scale Web Search Engines*, B. Barla Cambazoglu (Yahoo Labs), Ricardo Baeza-Yates(Yahoo Labs)*

  Room: Meeting Room 3 (2 F)
- Offline Evaluation and Optimization for Interactive Systems, *Lihong Li(Microsoft Research)*

  Room: Terrace Room 1 (3 F)

<u>Monday, 11:00-12:30</u> **Room: Terrace Room 2(3 F)**

Doctoral consortium

<u>Monday, 12:30-14:00</u>

Lunch

<u>Monday, 14:00-15:30</u>

Tutorials

- Real-Time Bidding: A New Frontier of Computational Advertising Research*, Jun Wang (University College London),Shuai Yuan (University College London)*

  Room: Meeting Room 1+2 (2 F)
- Learning about health and medicine from Internet data, *Elad Yom-Tov (Microsoft Research),Ingemar Johansson Cox (University College London),Vasileios Lampos (University College London)*

  Room: Meeting Room 3 (2 F)

- Distributed Graph Algorithmics: Theory and Practice, *Silvio Lattanzi (Google Research), Vahab Mirrokni(Google Research)*
  Room: Terrace Room 1 (3 F)

<u>Monday, 14:00-15:30</u> **Room: Terrace Room 2 (3 F)**

Doctoral consortium

<u>Monday, 15:30-16:00</u> **Room: Foyer**

Coffee Break

<u>Monday, 16:00-17:30</u>

Tutorials

- Real-Time Bidding: A New Frontier of Computational Advertising Research*, Jun Wang (University College London),Shuai Yuan (University College London)*
  Room: Meeting Room 1+2 (2 F)
- Learning about health and medicine from Internet data, *Elad Yom-Tov (Microsoft Research),Ingemar Johansson Cox (University College London),Vasileios Lampos (University College London)*
  Room: Meeting Room 3 (2 F)
- Distributed Graph Algorithmics: Theory and Practice, *Silvio Lattanzi (Google Research), Vahab Mirrokni(Google Research)*
  Room: Terrace Room 1 (3 F)

<u>Monday,16:00-17:30</u> **Room: Terrace Room 2 (3 F)**

Doctoral consortium

<u>Monday,18:30-20:00</u> **Room: Mezzanine Bar (2 F)**

Reception

Tuesday, 3th Feb

Tuesday,8:45-9:00 **Room: Grand Ballroom(2 F)**

Opening Session

Tuesday,9:00-10:00 **Room: Grand Ballroom(2 F)**

Keynote Speech

Chair: Xueqi Cheng (Chinese Academy of Sciences)

- Making Sense of Big Data with the Berkeley Data Analytics Stack, *Michael Franklin (University of California, Berkeley)*

Tuesday,10:00-10:30 **Room: Foyer**

Coffee Break

Tuesday,10:30-11:10 **Room: Grand Ballroom(2 F)**

Session 1: Practice and Experience Talk

Chair: Ying Li (EV Analysis Corporation)

- New Directions in Recommender Systems, *Jure Leskovec (Stanford university)*

Tuesday,11:10-12:10 **Room: Grand Ballroom(2 F)**

Session 1: Panel on Large Scale Data Understanding

- Big Data: New Paradigm or "Sound and Fury, Signifying Nothing"? *Andrei Broder (Google),Lada Adamic (Facebook),Michael Franklin (University of California, Berkeley),Maarten de Rijke (University of Amsterdam),Eric Xing (Carnegie Mellon University),Kai Yu (Baidu)*

Tuesday,12:10-13:30 **Room: Café Mix (1 F)**

Lunch

Tuesday,13:30-15:10 **Room: Grand Ballroom(2 F)**

Session 2: Web Search

Chair: Maarten de Rijke (University of Amsterdam)

- Delayed-Dynamic-Selective(DDS)Prediction for Reducing Extreme Tail Latency in Web Search, *Saehoon Kim (POSTECH), Yuxiong He (Microsoft Research), Seung-won Hwang(POSTECH), Sameh Elnikety (Microsoft Research), Seungjin Choi (POSTECH)*
- MergeRUCB: A Method for Large-Scale Online Ranker Evaluation, *Masrour Zoghi (University of Amsterdam), Shimon Whiteson (University of Amsterdam), Maarten de Rijke (University of Amsterdam)*
- Engagement Periodicity in Search Engine Usage: Analysis and its Application to Search Quality Evaluation, *Alexey Drutsa (Yandex), Gleb Gusev (Yandex), Pavel Serdyukov (Yandex)*
- Toward Predicting the Outcome of an A/B Experiment for Search Relevance, *Lihong Li (Microsoft Corp), Jin Young Kim (Microsoft Corp), Imed Zitouni (Microsoft Corp)*
- Optimal Space-time Tradeoffs for Inverted Indexes, *Giuseppe Ottaviano (National Research Council of Italy), Nicola Tonellotto (National Research Council of Italy), Rossano Venturini (National Research Council of Italy & University of Pisa)*
- Understanding and predicting Graded Search Satisfaction, *Jiepu Jiang (University of Massachusetts Amherst), Ahmed Hassan Awadallah (Microsoft Research Redmond), Xiaolin Shi (Microsoft Research Redmond), Ryen W. White (Microsoft Research Redmond)*

- Robust Tree-based Causal Inference for Complex Ad Effectiveness Analysis, *Pengyuan Wang (Yahoo Labs),Wei Sun (Purdue University),Dawei Yin (Yahoo Labs),Jian Yang (Yahoo Labs),Yi Chang (Yahoo Labs)*

Tuesday,15:10-15:40 **Room: Foyer**

Coffee Break

Tuesday,15:40-17:20 **Room: Grand Ballroom(2 F)**

Session3: Social Networks

Chair: Elad Yom-Tov (Microsoft Research)

- The Power of Random Neighbors in Social Network, *Silvio Lattanzi (Google, Inc.), Yaron Singer (Harvard University)*
- Negative Link Prediction in Social Media, *Jiliang Tang (Arizona State University) ,Shiyu Chang (University of Illinois at Urbana-Champaign), Charu Aggarwal (IBM T.J. Watson Research Center), Huan Liu (Arizona State University)*
- Sarcasm Detection on Twitter: A Behavioral Modeling Approach, *Ashwin Rajadesingan (Arizona State University), Reza Zafarani (Arizona State University), Huan Liu (Arizona State University)*
- Modeling and Predicting Retweeting Dynamics on Microblogging Platforms, *Shuai Gao (Shandong University), Jun Ma (Shandong University), Zhumin Chen (Shandong University)*
- On Integrating Network and Community Discovery, *Jialu Liu (University of Illinois at Urbana-Champaign), Charu Aggarwal (IBM T.J. Watson Research Center), Jiawei Han (University of Illinois at Urbana-Champaign)*

- On the Accuracy of Hyper-local Geotagging of Social Media Content, *David Flatow (Cornell Tech & Stanford University), Mor Naaman (Cornell Tech), Ke Eddie Xie (Cornell Tech & Twitter Inc.), Yana Volkovich (Cornell Tech & Barcelona Media), Yaron Kanza (Cornell Tech & Technion - Israel Institute of Technology)*

Tuesday,18:00-20:00 **Room: Grand Ballroom(2 F)**

Poster Session

Wednesday, Feb 4

Wednesday,9:00-10:00 **Room: Grand Ballroom(2 F)**

Keynote Speech

Chair: Evgeniy Gabrilovich (Google)

- The Information Life of Social Networks, *Lada A. Adamic(Facebook)*

Wednesday,10:00-10:30 **Room: Foyer**

Coffee Break

Wednesday,10:30-12:10 **Room: Grand Ballroom(2 F)**

Session 4: Web Mining

Chair: Huan Liu (Arizona State University)

- Learning to Recommend Related Entities to Search Users, *Bin Bi (University of California, Los Angeles), Hao Ma (Microsoft Research), Bo-June (Paul) Hsu (Microsoft Research), Wei Chu (Microsoft), Kuansan Wang (Microsoft Research), Junghoo Cho (University of California, Los Angeles)*
- Will This Paper Increase Your h-index? Scientific Impact Prediction, *Yuxiao Dong (University of Notre Dame), Reid A. Johnson (University of Notre Dame), Nitesh V. Chawla (University of Notre Dame)*
- Concept Graph Learning from Educational Data, *Yiming Yang (Carnegie Mellon University), Hanxiao Liu (Carnegie Mellon University), Jaime Carbonell (Carnegie Mellon University), Wanli Ma (Carnegie Mellon University)*

- Review Synthesis for Micro-Review Summarization, *Thanh-Son Nguyen (Singapore Management University), Hady W. Lauw (Singapore Management University), Panayiotis Tsaparas (University of Ioannina)*
- Fast and Space-Efficient Entity Linking in Queries, *Roi Blanco (Yahoo Labs), Giuseppe Ottaviano (ISTI-CNR), Edgar Meij (Yahoo Labs)*
- On Tag Recommendation for Expertise Profiling: A Case Study in the Scientific Domain, *Isac S. Ribeiro (Universidade Federal de Minas Gerais), Rodrygo L. T. Santos (Universidade Federal de Minas Gerais), Marcos A. Gonçalves (Universidade Federal de Minas Gerais), Alberto H. F. Laender (Universidade Federal de Minas Gerais)*
- FLAME:A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering, *Yao Wu (Simon Fraser University), Martin Ester (Simon Fraser University)*

Wednesday,12:10-13:30 **Room: Café Mix (1 F)**

Lunch

Wednesday,13:30-14:50 **Room: Grand Ballroom(2 F)**

Session 5: Practice and Experience Talk

Chair: Paul Bennett (Microsoft)

- Semantic Matching in APP Search, *Juchao Zhuo (Tencent Inc.), Zeqian Huang (Tencent Inc.), Yunfeng Liu (Tencent Inc.), Zhanhui Kang (Tencent Inc.), Xun Cao (Tencent Inc.), Mingzhi Li (Tencent Inc.), Long Jin (Tencent Inc.)*

- Boosting Search with Deep Understanding of Contents and Users, *Kaihua Zhu (Baidu)*

Wednesday,14:50-15:20 **Room: Foyer**

Coffee Break

Wednesday,15:20-17:00 **Room: Grand Ballroom(2 F)**

Session 6: Crowdsourcing, Temporal and Location-based mining

Chair: Charlie Clarke (University of Waterloo)

- Driven by Food: Modeling Geographic choice, *Ravi Kumar (Google),Mohammad Mahdian (Google), BoPang (Google) , Andrew Tomkins(Google), Sergei Vassilvitskii (Google)*
- Hiring Behavior Models for Online Labor Markets, *Marios Kokkodis (NYU Stern), Panagiotis Papadimitriou (Elance-oDesk), Panagiotis G. Ipeirotis (NYU Stern)*
- Just in Time Recommendations-Modeling the Dynamics of Boredom in Activity Streams, *Komal Kapoor (University of Minnesota), Karthik Subbian (University of Minnesota), Jaideep Srivastava (University of Minnesota), Paul Schrater (University of Minnesota)*
- Leveraging In-Batch Annotation Bias for Crowdscourced Active Learning, *Honglei Zhuang (LinkedIn Corporation & University of Illinois at Urbana-Champaign), Joel Young (LinkedIn Corporation)*
- Listwise Approach for Rank Aggregation in Crowdsourcing, *Shuzi Niu (Chinese Academy of Sciences), Yanyan Lan (Chinese Academy of Sciences), Jiafeng Guo (Chinese Academy of Sciences), Xueqi Cheng (Chinese Academy of Sciences), Lei Yu (Chinese Academy of Sciences), Guoping Long (Chinese Academy of*

Sciences)

- WorkRank: Using Employer Implicit Judgements to Infer Worker Reputation, *Maria Daltayanni (University of California, Santa Cruz),Luca de Alfaro (University of California, Santa Cruz),Panagiotis Papadimitriou (Elance-oDesk)*

Thursday, Feb 5

Thursday,9:00-10:00 **Room: Grand Ballroom(2 F)**

Keynote Speech

Chair: Jie Tang (Tsinghua University)

- Learning from User Interactions, *Thorsten Joachims (Cornell University)*

Thursday,10:00-10:30 **Room: Foyer**

Coffee Break

Thursday,10:30-12:10 **Room: Grand Ballroom(2 F)**

Session 7: User Modeling, Mobility, and Recommendation

Chair: Grace Hui Yang (Georgetown University)

- User Modeling for a Personal Assistant, *Ramanathan Guha (Google), Vineet Gupta (Google), Vivek Raghunathan (Google), Ramakrishnan Srikant (Google)*
- Predicting The Next App That You Are Going To Use, *Ricardo Baeza-Yates (Yahoo Labs), Di Jiang (HKUST), Fabrizio Silvestri (Yahoo Labs), Beverly Harrison (Yahoo Labs)*
- You Are Where You Go: Inferring Demographic Attributes from Location Check-Ins, *Yuan Zhong (Microsoft Research & Northeastern University),Nicholas Jing Yuan (Microsoft Research), Wen Zhong (Stony Brook University), Fuzheng Zhang (University of Science and Technology of China & Microsoft Research),Xing Xie (Microsoft Research)*
- SimApp: A Framework for Detecting Similar Mobile Applications by Online Kernel Learning, *Ning Chen (Nanyang Technological University),Steven C. H. Hoi (Singapore Management*

University),Shaohua Li (Nanyang Technological University), Xiaokui Xiao (Nanyang Technological University)

- Personalized Mobile App Recommendation: Reconciling App Functionality and User Privacy Preference, *Bin Liu (Rutgers University),Deguang Kong (Samsung Research America), Lei Cen (Purdue University),Neil Zhenqiang Gong (University of California, Berkeley),Hongxia Jin (Samsung Research America), Hui Xiong (Rutgers University)*
- Inferring Movement Trajectories from GPS Snippets, *Mu Li (Carnegie Mellon University),Amr Ahmed (Google Strategic Technologies),Alexander J. Smola (Carnegie Mellon University & Google Strategic Technologies)*

Thursday,12:10-13:30 **Room: Grand Ballroom(2 F)**

WSDM Business Lunch Meeting

Thursday,13:30-14:50 **Room: Grand Ballroom(2 F)**

Session 8: Practice and Experience Talk

Chair: Xuanjing Huang (Fudan University)

- Regressing Towards Simpler Prediction Systems, *Tushar Chandra(Google)*
- Global Optimization for Display Ad, *Rong Jin (Alibaba)*

Thursday,14:50-15:20 **Room: Foyer**

Coffee Break

Thursday,15:20-17:00 **Room: Grand Ballroom(2 F)**

Session 9: Web Mining(2)

Chair: Fabrizio Silvestri (Yahoo!)

- Back to the Past: Supporting Interpretations of Forgotten Stories by Time-aware Re-Contextualization, *Nam Khanh Tran (Leibniz Universität),Andrea Ceroni (Leibniz Universität), Nattiya Kanhabua (Leibniz Universität), Claudia Niederée (Leibniz Universität)*
- Diluted Treatment Effect Estimation for Trigger Analysis in Online Controlled Experiments, *Alex Deng (Microsoft), Victor Hu (Microsoft)*
- Inverting a Steady-State, *Ravi Kumar (Google),Andrew Tomkins (Google),Sergei Vassilvitskii (Google),Erik Vee (Google)*
- Automatic Gloss Finding for a Knowledge Base Using Ontological Constraints, *Bhavana Dalvi (Carnegie Mellon University), Einat Minkov (University of Haifa), Partha P. Talukdar (Indian Institute of Science), William W. Cohen (Carnegie Mellon University)*
- Finding Subgraphs with Maximum Total Density and Limited Overlap, *Oana Denisa Balalau (Telecom Paristech), Francesco Bonchi (Yahoo Labs), T-H. Hubert Chan (The University of Hong Kong), Francesco Gullo (Yahoo Labs), Mauro Sozio (Telecom Paristech)*
- Modeling Website Popularity Competition in the Attention-Activity Marketplace, *Bruno Ribeiro (Carnegie Mellon University), Christos Faloutsos (Carnegie Mellon University)*
- Exploring the Space of Topic Coherence Measures, *Michael Röder (Leipzig University), Andreas Both (Unister GmbH), Alexander Hinneburg (Martin-Luther-University)*

<u>Thursday,17:05-17:20</u> **Room: Grand Ballroom(2 F)**

Closing Session

Friday, Feb 6

Friday, 9:00-10:30

Workshops

- HIA'15: Heterogeneous Information Access Workshop at WSDM 2015, *Ke Zhou (Yahoo Labs), Roger Jie Luo (Yahoo Labs), Djoerd Hiemstra (University of Twente), Joemon M. Jose (University of Glasgow)*
  Room: Meeting Room 1+2 (2 F)
- DL-WSDM'15: Workshop on Deep Learning for Web Search and Data Mining, *Bin Gao (Microsoft Research), Jiang Bian (Microsoft Research)*
  Room: Meeting Room 3 (2 F)
- The 2nd workshop on Vertical Search Relevance at WSDM 2015, *Dawei Yin (Yahoo Labs), Chih-Chieh Hung (Rakuten Inc.), Rui Li (Yahoo Labs), Yi Chang (Yahoo Labs)*
  Room: Terrace Room 1 (3 F)
- WSDM'15 Workshop Summary / Scalable Data Analytics: Theory and Applications, *Kaizhu Huang (Xi'an Jiaotong-Liverpool University), Haiqin Yang (Chinese University of Hong Kong), Irwin King (Chinese University of Hong Kong), Michael R. Lyu (Chinese University of Hong Kong)*
  Room: Terrace Room 2 (3 F)

Friday, 10:30-11:00 **Room: Foyer**

Coffee Break

Friday, 11:00-12:30

Workshops

- HIA'15: Heterogeneous Information Access Workshop at

WSDM 2015, *Ke Zhou (Yahoo Labs), Roger Jie Luo (Yahoo Labs), Djoerd Hiemstra (University of Twente), Joemon M. Jose (University of Glasgow)*

Room: Meeting Room 1+2 (2 F)

- DL-WSDM'15: Workshop on Deep Learning for Web Search and Data Mining, *Bin Gao (Microsoft Research), Jiang Bian (Microsoft Research)*

  Room: Meeting Room 3 (2 F)
- The 2nd workshop on Vertical Search Relevance at WSDM 2015, *Dawei Yin (Yahoo Labs), Chih-Chieh Hung (Rakuten Inc.), Rui Li (Yahoo Labs), Yi Chang (Yahoo Labs)*

  Room: Terrace Room 1 (3 F)
- WSDM'15 Workshop Summary / Scalable Data Analytics: Theory and Applications, *Kaizhu Huang (Xi'an Jiaotong-Liverpool University), Haiqin Yang (Chinese University of Hong Kong), Irwin King (Chinese University of Hong Kong), Michael R. Lyu (Chinese University of Hong Kong)*

  Room: Terrace Room 2 (3 F)

Friday, 12:30-14:00

Lunch

Friday, 14:00-15:30

Workshops

- HIA'15: Heterogeneous Information Access Workshop at WSDM 2015, *Ke Zhou (Yahoo Labs), Roger Jie Luo (Yahoo Labs), Djoerd Hiemstra (University of Twente), Joemon M. Jose (University of Glasgow)*

  Room: Meeting Room 1+2 (2 F)
- DL-WSDM'15: Workshop on Deep Learning for Web Search and

Data Mining, *Bin Gao (Microsoft Research), Jiang Bian (Microsoft Research)*

Room: Meeting Room 3 (2 F)

Friday, 13:30-16:00 **Room: Foyer**

Coffee Break

Friday, 16:00-17:30

Workshops

- HIA'15: Heterogeneous Information Access Workshop at WSDM 2015, *Ke Zhou (Yahoo Labs), Roger Jie Luo (Yahoo Labs), Djoerd Hiemstra (University of Twente), Joemon M. Jose (University of Glasgow)*

  Room: Meeting Room 1+2 (2 F)
- DL-WSDM'15: Workshop on Deep Learning for Web Search and Data Mining, *Bin Gao (Microsoft Research), Jiang Bian (Microsoft Research)*

  Room: Meeting Room 3 (2 F)

Doctoral Consortium

WSDM 2015 Doctoral Consortium will be on Monday February 2. Doctoral students will present research proposals and make discussions with their mentors.

Mentors:

- Paul Hsu (Microsoft Research)
- Huan Liu (Arizona State University)
- Panagiotis de Rijke (University of Amsterdam)
- Kuansan Wang (Microsoft)
- Jirong Wen (Renmin University of China)
- Hui Xiong (Rutgers University)
- Grace Hui Yang (Georgetown University)

Accepted Proposals:

1. An Approach to the Problem of Annotation of Research Publications ,*Ekaterina Chemayak(National Research University)*
2. Incorporating Phrase-level Sentiment Analysis on Textual Reviews for Personalized Recommendation, *Yongfeng Zhang (Tsinghua University)*
3. Mining Groups Stability in Ubiquitous and Social Environments Communities, Classes, Clusters, *Mark Kibanov (University of Kassel)*
4. Sentiment-Specific Representation Learning for Document-Level Sentiment Analysis, *Duyu Tang (Harbin Institute of Technology)*
5. Chronological Scientific Information Recommendation via Supervised Dynamic To Modeling, *Zhuoren Jiang (Dalian Maritime University)*
6. Topics, Task & Beyond: Learning Representations for Personalization, *Rishabh Mehrotra (University College London)*

Tutorials

The following tutorials will be given on February 2, 2015

- Dynamic Information Retrieval Modeling, *Hui Yang (Georgetown University), Marc Sloan (University College London), Jun Wang (University College London)*
- Scalability and Efficiency Challenges in Large-Scale Web Search Engines , *B. Barla Cambazoglu (Yahoo Labs), Ricardo Baeza-Yates (Yahoo Labs)*
- Offline Evaluation and Optimization for Interactive Systems*, Lihong Li (Microsoft Research)*
- Real-Time Bidding: A New Frontier of Computational Advertising Research *, Jun Wang (University College London), Shuai Yuan (University College London)*
- Learning about Health and Medicine from Internet Data, *Elad Yom-Tov (Microsoft Research), Ingemar Johansson Cox (University College London), Vasileios Lampos (University College London)*
- Distributed Graph Algorithmics: Theory and Practice*, Silvio Lattanzi (Google Research), Vahab Mirrokni (Google Research)*

Dynamic Information retrieval Modeling

Grace Hui Yang, Marc Sloan, and Jun Wang

In Dynamic Information Retrieval Modeling we model dynamic systems which change or adapt over time or a sequence of events using a range of techniques from artificial intelligence and reinforcement learning. Many of the open problems in current IR research can be described as dynamic systems, for instance, session search

or computational advertising. State of the art research provides solutions to these problems that are responsive to a changing environment, learning from past interactions and predict future utility. Advances in IR interface, personalization and ad display demand models that can react to users in real time and in an intelligent, contextual way.

The objective of this tutorial is to provide a comprehensive and up-to-date introduction to Dynamic Information Retrieval Modeling. Dynamic IR Modeling is the statistical modeling of IR systems that can adapt to change. It is a natural follow-up to previous statistical IR modeling tutorials with a fresh look on state-of-the-art dynamic retrieval models and their applications including session search and online advertising. The tutorial will cover techniques ranging from classic relevance feedback to the latest applications of partially observable Markov decision processes (POMDPs) and will present to fellow researchers and practitioners a handful of useful algorithms and tools for solving and evaluating IR problems incorporating dynamics.

Scalability and Efficiency Challenges in Large-Scale Web Search Engines

B. Barla Cambazoglu and Ricardo Baeza-Yates

This tutorial aims to provide a fairly comprehensive overview of the scalability and efficiency challenges in large-scale web search engines. In particular, the tutorial provides an in-depth architectural overview of a web search engine, mainly focusing on the web crawling, indexing, and query processing components. The scalability and efficient issues encountered in the above-mentioned components are presented at four different granularities: at the level of a single

computer, a cluster of computers, a single data center, and a multi-center search engine. The tutorial also points at the open research problems and provides recommendations to researchers who are new to the field.

Offline Evaluation and Optimization for Interactive Systems: A Practical Guide
Lihong Li

Evaluating and optimizing an interactive system (like search engine, recommender and advertising systems) from historical data against a predefined online metric is challenging, especially when that metric is computed from user feedback such as clicks and payments. The key challenge is counterfactual in nature: we only observe a user's feedback for the action taken by the system, but do not know what that user would have reacted if the system chose a different action. The standard approach to evaluating such metrics of a user-interacting system is online A/B tests (a.k.a. randomized controlled experiments), which can be expensive for several reasons. Offline Evaluation becomes critical, with the aim of evaluating the same metrics without running too many costly experiments on live users. In recent years, substantial advances have been made to address this problem, resulting in reliable solutions that have proven effective in important real-world problems and that have been used by industry leaders. This tutorial reviews the basic theory as well as representative techniques, and illustrates how to apply them in practice, using several case studies done at Microsoft and Yahoo!

Real-Time Bidding: A New Frontier of Computational Advertising Research

Shuai Yuan and Jun Wang

In display and mobile advertising, the most significant development in recent years is the Real-Time Bidding (RTB), which allows selling and buying in real-time one ad impression at a time. Since then RTB has fundamentally changed the landscape of the digital marketing by scaling the buying process across a large number of available inventories. The demand for automation, integration and optimization in RTB brings new research opportunities in the DM/ML fields. However, despite its rapid growth and huge potential, many aspects of RTB remain unknown to the research community for many reasons. In this tutorial, together with invited distinguished speakers from online advertising industry, we aim to bring the insightful knowledge from the real-world systems to bridge the gaps and provide an overview of the fundamental infrastructure, algorithms, and technical and research challenges of this new frontier of computational advertising. We will also introduce to researchers the datasets, tools, and platforms which are publicly available thus they can get hands-on quickly.

This tutorial aims to provide not only a comprehensive and systematic introduction to RTB and computational advertising in general, but also the emerging research challenges and research tools and datasets in order to facilitate the research. Compared to previous computational advertising tutorials in relevant top-tier conferences, this tutorial takes a fresh, neutral, and the latest look of the field and focused on the fundamental changes brought by RTB. We expect the audience, after attending the tutorial, to understand the real-time online advertising mechanisms and the state of the art

techniques, as well as to grasp the research challenges in this field. Our motivation is to help the audience acquire domain knowledge and obtain relevant datasets, and to promote research activities in RTB and computational advertising in general.

Learning About Health and Medicine from Internet Data
Elad Yom-Tov, Ingemar Cox and Vasileios Lampos

Surveys show that around 80% of US Internet users consult the Internet when they require medical information. People seek this information using both traditional search engines and via social media. The information created using the search process offers an unprecedented opportunity for applications to monitor and improve the quality of life of people with a variety of medical conditions. In recent years research in this area has addressed public-health questions such as the effect of the media on development of anorexia, developed tools for measuring influenza rates and assessing drug safety, and examined the effects of health information on individual wellbeing.
This tutorial will show how Internet data can facilitate medical research, providing an overview of the state-of-the-art in this area. During the tutorial we will discuss the information which can be gleaned from a variety of Internet data sources, including social media, search engines, and specialized medical websites. We will provide an overview of analysis methods used in recent literature, and show how results can be evaluated using publicly-available health information and online experimentation. Finally, we will discuss ethical and privacy issued and possible technological solutions. This tutorial is intended for researchers of user generated content who are interested in applying their knowledge to improve health and

medicine.

Distributed Graph Algorithmics: Theory and Practice
Silvio Lattanzi and Vahab Mirrokni

As a fundamental tool in modeling and analyzing social and information networks, large-scale graph mining is an important component of any tool set for big data analysis. Processing graphs with hundreds of billions of edges is only possible via developing distributed algorithms under distributed graph mining frameworks such as MapReduce, Pregel, Gigraph, and alike. For these distributed algorithms to work well in practice, we need to take into account several metrics such as the number of rounds of computation and the communication complexity of each round. For example, given the popularity and ease-of-use of MapReduce framework, developing practical algorithms with good theoretical guarantees for basic graph algorithms is a problem of great importance.
In this tutorial, we first discuss how to design and implement algorithms based on traditional MapaReduce architecture. In this regard, we discuss various basic graph theoretic problems such as computing connected components, maximum matching, MST, counting triangle and overlapping and describe the sampling, filtering, local random walk, and core-set techniques to develop efficient algorithms in this framework. At the end, we explore the possibility of employing other distributed graph processing frameworks. In particular, we study the effect of augmenting MapReduce with a distributed hash table (DHT) service and also discuss the use of a new graph processing framework called ASYMP based on asynchronous message-passing method. In particular, we will show that us-

ing ASPMP, one can improve the CPU usage, and achieve significantly improved running time.

Practice and Experience Talks

The following practice and experience talks will be given on Feb. 3 – Feb. 5, 2015.

- New Directions in Recommender Systems, Semantic Matching in APP Search, *Juchao Zhuo (Tencent)*
- Boosting Search with Deep Understanding of Contents and Users, *Kaihua Zhu (Baidu)*
- Regressing Towards Simpler Prediction Systems, *Tushar Chandra (Google)*
- Global Optimization for Display Ad, *Rong Jin (Alibaba)*

New Directions in Recommender Systems
Jure Leskovec (Stanford)

Abstract: Recommender systems are an integral part of how we experience the web today and they have become so ubiquitous that we do not even notice them anymore. However, today's recommender systems mostly treat items they recommend as black boxes and primarily focus on extracting correlations and co-counts from user behavior data. In this talk I argue that next generation recommender systems will require deep understanding of items being recommended as well as modeling the relationships between those items. I will present example how auxiliary data about items (descriptions, reviews, product specifications) can be used to improve recommendations.

Semantic Matching in APP Search

Juchao Zhuo (Tencent)

Abstract: Past years, with the growth of smartphones and applications, APP market has become an important mobile internet portal. As important function in application market, APP search gains lots of attentions. However, mismatch between queries and APP is the most critical problem in APP search because of less text with term matching search engine. This talk will describe a sematic matching architecture in APP search, which mining topics and tags in big data. It enriches query and APP representations with topics and tags to achieve semantic matching in search.

Some challenge must be considered:

- How to extract tag-APP relationship from large web text.
- How to use machine learning technologies to process denoising and computing confidence.
- How to hybrid ranking apps retrieved by different matching method.

These will be introduced in some of our related works and as examples to describe how semantic matching is used in Tencent MyApp, an application market which serving hundreds of millions of users.

Boosting Search with Deep Understanding of Contents and Users

Kaihua Zhu (Baidu)

Abstract: Recent years have witnessed dramatic changes in how people interact with search engines. Search engine are expected be more intelligent in understanding user' intention and fulfilling users' needs with direct answers rather than raw information. Furthermore, search engines are expected to be equipped with recommendation and dialogue capabilities, making the interaction with users

more natural and smoother. In this talk, I will introduce Baidu's work on how to make some of them come true through the deep understanding of users, queries and web pages, and discuss challenges behind these technologies.

Regressing Towards Simpler Prediction Systems
Tushar Chandra (Google)

Abstract: This talk will focus on our experience in managing the complexity of Sibyl, a large scale machine learning system that is widely used within Google. We believe that a large fraction of the challenges faced by Sibyl are inherent to large scale machine learning and that other systems are likely to encounter them as well. Thus, these challenges present interesting opportunities for future research.

Global Optimization for Display Ad
Rong Jin (Alibaba)

Abstract: Online display advertisement has been examined by numerous studies. Most online display ad systems take the greedy approach, namely they display, for each user, the set of ads that match best with the user's interests. One shortcoming of the greedy approach is that it does not take into account the budget limitation of each advertiser. As a result, we often observed that some ads are popular and match with the interests of millions of users; but due to the budget restriction, these ads can only be presented by a limited times, leading to a suboptimal performance.

Keynote Speeches

We're delighted to have three distinguished researchers to give keynotes at WSDM 2015.

- Making Sense of Big Data with the Berkeley Data Analytics Stack, *Michael Franklin (UC Berkeley)*
- The Information Life of Social Networks, *Lada Adamic (Facebook)*
- Learning from User Interactions, *Thorsten Joachims (Cornell University)*

Making Sense of Big Data with the Berkeley Data Analytics Stack

Abstract: The Berkeley AMPLab is creating a new approach to data analytics. Launching in early 2011, the lab aims to seamlessly integrate the three main resources available for making sense of data at scale: Algorithms (machine learning and statistical techniques), Machines (in the form of scalable clusters and elastic cloud computing), and People (both individually as analysts and in crowds). The lab is realizing its ideas through the development of a freely-available Open Source software stack called BDAS: the Berkeley Data Analytics Stack. In the four years the lab has been in operation, we've released major components of BDAS. Several of these components have gained significant traction in industry and elsewhere: the Mesos cluster resource manager, the Spark in-memory computation framework, and the Shark query processing system. BDAS features prominently in many industry discussions of the future of the

Big Data analytics ecosystem – a rare degree of impact for an ongoing academic project.

Given this initial success, the lab is continuing on its research path, moving "up the stack" to better integrate and support advanced analytics and to make people a full-fledged resource for making sense of data. In this talk, I'll first outline the motivation and insights behind our research approach and describe how we have organized to address the cross-disciplinary nature of Big Data challenges. I will then describe the current state of BDAS with an emphasis on our newest efforts, including some or all of: the GraphX graph processing system, the Velox and MLBase machine learning platforms, and the SampleClean framework for hybrid human/computer data cleaning. Finally I will present our current views of how all the pieces will fit together to form a system that can adaptively bring the right resources to bear on a given data-driven question to meet time, cost and quality requirements throughout the analytics lifecycle.

Bio: **Michael Franklin** is the Thomas M. Siebel Professor of Computer Science and Chair of the Computer Science Division at the University of California, Berkeley. He has over 30 years of experience in the database, data analytics, and data management fields as a researcher, lab director, faculty member, entrepreneur, and software developer. Prof. Franklin is also the Director of the Algorithms, Machines, and People Laboratory (AMPLab) at UC Berkeley. The AMPLab currently works with 23 industrial sponsors including founding sponsors Amazon Web Services, Google, and SAP, and received a National Science Foundation CISE "Expeditions in Computing" award, announced as part of the White House Big Data research initiative in March 2012. AMPLab is well-known for creating

a number of popular systems in the Open Source Big Data ecosystem including Spark, Mesos, Shark, GraphX and MLlib, all parts of the Berkeley Data Analytics Stack (BDAS). Prof. Franklin is also a co-PI and Executive Committee member for the Berkeley Institute for Data Science, part of a multi-campus initiative to advance Data Science Environments. He is an ACM Fellow, a two-time winner of the ACM SIGMOD "Test of Time" award, and recipient of the outstanding Advisor Award from the Computer Science Graduate Student Association at Berkeley.

The Information Life of Social Networks

Abstract: Vast amounts of information are propagated in online social networks such as Facebook. This talk will describe several studies characterizing how information diffuses over social ties, from the growth of individual cascades to the predictability of their eventual size. It will also characterize the diffusion of specific kinds of information, including rumors, memes, and social movements.

Bio: **Lada Adamic** leads the Product Science group within Facebook's Data Science Team. She is also an adjunct associate professor at the University of Michigan's School of Information and Center for the Study of Complex Systems. Her research interests center on information dynamics in networks: how information diffuses, how it can be found, and how it influences the evolution of a network's structure. Her projects have included identifying expertise in online question and answer forums, studying the dynamics of viral marketing, and characterizing the structural and communication patterns in online social media. She has received an NSF CAREER award, a University of Michigan Henry Russell award, the 2012 Lagrange

Prize in Complex Systems.

Learning from User Interactions

Abstract: The ability to learn from user interactions can give systems access to unprecedented amounts of world knowledge. This is already evident in search engines, recommender systems, and electronic commerce, and other applications are likely to follow in the near future (e.g., education, smart homes). More generally, the ability to learn from user interactions promises pathways for solving knowledge-intensive tasks ranging from natural language understanding to autonomous robotics.

Learning from user interactions, however, means learning from data that does not necessarily fit the assumptions of the standard machine learning models. Since interaction data consists of the choices that humans make, it has to be interpreted with respect to how humans make decisions, which is influenced by the decision context and constraints like human motivation and human abilities. In this talk, I argue that we need learning approaches that explicitly model user-interaction data as the result of human decision making. To this effect, the talk explores how integrating micro-economic models of human behavior into the learning process leads to new learning algorithms that have provable guarantees under verifiable assumptions and to learning systems that perform robustly in practice. These findings imply that the design space of such human-interactive learning systems encompasses not only the machine learning algorithm itself, but also the design of the interaction under an appropriate model of user behavior.

Bio: **Thorsten Joachims** is a Professor in the Department of Computer Science and in the Department of Information Science at Cornell University. His research interests center on a synthesis of theory and system building in machine learning, with applications in information access, language technology, and recommendation. His past research focused on support vector machines, text classification, structured output prediction, convex optimization, learning to rank, learning with preferences, and learning from implicit feedback. In 2001, he finished his dissertation advised by Prof. Katharina Morik at the University of Dortmund. From there he also received his Diplom in Computer Science in 1997. Between 2000 and 2001 he worked as a PostDoc at the GMD Institute for Autonomous Intelligent Systems. From 1994 to 1996 he was a visiting scholar with Prof. Tom Mitchell at Carnegie Mellon University.

Accepted Papers

- WorkerRank: Using Employer Implicit Judgements To Infer Worker Reputation, *Maria Daltayanni, Luca de Alfaro and Panagiotis Papadimitriou.*
- You Are Where You Go: Inferring Demographic Attributes from Location Check-ins, *Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang and Xing Xie.*
- Automatic Gloss Finding for a Knowledge Base using Ontological Constraints, *Bhavana Dalvi, Einat Minkov, Partha Talukdar and William Cohen.*
- DDS Prediction: Reducing Extreme Tail Latency in Web Search, *Saehoon Kim, Yuxiong He, Seung-Won Hwang, Sameh Elnikety and Seungjin Choi*
- MergeRUCB: A Method for Large-Scale Online Ranker Evaluation, *Masrour Zoghi, Shimon Whiteson and Maarten de Rijke.*
- FLAME: A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering, *Yao Wu and Martin Ester*
- Concept Graph Learning from Educational Data, *Yiming Yang, Hanxiao Liu, Jaime Carbonell and Wanli Ma*
- The Power of Random Neighbors in Social Networks, *Silvio Lattanzi and Yaron Singer.*
- Robust Tree-based Causal Inference for Complex Ad Effectiveness Analysis, *Pengyuan Wang, Wei Sun, Dawei Yin, Jian Yang and Yi Chang*
- Negative Link Prediction in Social Media, *Jiliang Tang, Shiyu Chang, Charu Aggarwal and Huan Liu.*

- On the Bounds and Accuracy of Hyper-local Geocoding of Social Media Content, *David Flatow, Mor Naaman, Eddie Xie, Yana Volkovich and Yaron Kanza.*
- Optimal Space-time Tradeoffs for Inverted Indexes, *Giuseppe Ottaviano, Nicola Tonellotto and Rossano Venturini*
- Finding Subgraphs with Maximum Total Density and Limited Overlap, *Oana Denisa Balalau, Francesco Bonchi, T-H. Hubert Chan, Francesco Gullo and Mauro Sozio*
- On Hiring Decisions in Online Labor Markets, *Marios Kokkodis, Panagiotis Papadimitriou and Panos Ipeirotis.*
- Modeling Geographic Choice, *Ravi Kumar, Mohammad Mahdian, Bo Pang, Andrew Tomkins and Sergei Vassilvitskii*
- Leveraging In-Batch Annotation Bias for Crowdsourced Active Learning, *Honglei Zhuang and Joel Young*
- Predicting The Next App That You Are Going To Use, *Ricardo Baeza-Yates, Di Jiang, Fabrizio Silvestri and Beverly Harrison*
- Modeling and Predicting Retweeting Dynamics on Microblogging Platforms, *Shuai Gao, Jun Ma and Zhumin Chen*
- Learning to Recommend Related Entities to Search Users, *Bin Bi, Hao Ma, Paul Hsu, Wei Chu, Kuansan Wang and Junghoo Cho.*
- SimApp: A Framework for Detecting Similar Mobile Applications by Online Kernel Learning, *Ning Chen, Steven C.H. Hoi, Shaohua Li and Xiaokui Xiao.*
- Timing Matters! – Modeling Dynamics of Boredom in Activity Streams, *Komal Kapoor, Karthik Subbian, Jaideep Srivastava and Paul Schrater*.
- Diluted Treatment Effect Estimation for Trigger Analysis in Online Controlled Experiments, *Alex Deng and Victor Hu.*

- Listwise Approach for Rank Aggregation in Crowdsourcing, *Shuzi Niu.*
- User Modeling for a Personal Assistant, *Ramanathan Guha, Vineet Gupta, Vivek Raghunathan and Ramakrishnan Srikant.*
- Inverting a Steady-State, *Ravi Kumar, Andrew Tomkins, Sergei Vassilvitskii and Erik Vee*
- Toward Predicting the Outcome of an A/B Experiment for Search Relevance*, Jin Kim, Lihong Li and Imed Zitouni.*
- Modeling Website Popularity Competition in the Attention-Activity Marketplace, *Bruno Ribeiro and Christos Faloutsos.*
- Inferring Movement Trajectories from GPS Snippets, *Mu Li, Amr Ahmed and Alex Smola.*
- Will This Paper Increase Your h-index? Scientific Impact Prediction, *Yuxiao Dong, Reid Johnson and Nitesh Chawla.*
- Back to the Past: Supporting Interpretations of Forgotten Stories by Time-aware Re-Contextualization, *Nam Khanh Tran, Andrea Ceroni, Nattiya Kanhabua and Claudia Niederée.*
- Sarcasm Detection on Twitter: A Behavioral Modeling Approach, *Ashwin Rajadesingan, Reza Zafarani and Huan Liu.*
- Fast and Space-efficient Entity Linking in Queries, *Roi Blanco, Giuseppe Ottaviano and Edgar Meij.*
- Engagement Periodicity in Search Engine Usage: Analysis and Its Application to Search Quality Evaluation, *Alexey Drutsa, Pavel Serdyukov and Gleb Gusev.*
- Understanding and Predicting Graded Search Satisfaction, *Jiepu Jiang, Ahmed Hassan, Xiaolin Shi and Ryen White.*

- On Tag Recommendation for Expertise Profiling: A Case Study in the Scientific Domain, *Isac Ribeiro, Rodrygo Santos, Alberto Laender and Marcos Gonçalves.*
- Review Synthesis for Micro-Review Summarization, *Thanh-Son Nguyen, Hady Lauw and Panayiotis Tsaparas.*
- Personalized Mobile App Recommendation: Reconcling App Functionality and User Privacy Preference, *Bin Liu, Deguang Kong, Lei Cen, Neil Zhenqiang Gong, Hongxia Jin and Hui Xiong.*
- On Integrating Network and Community Discovery, *Jialu Liu, Charu Aggarwal and Jiawei Han.*
- Exploring the Space of Coherence Measures, *Michael Röder, Andreas Both and Alexander Hinneburg.*

Workshops

There are four workshops hosted at WSDM 2015, covering novel topics and emerging areas in web search and data mining.

- HIA'15: Heterogeneous Information Access Workshop at WSDM 2015(full day)
- DL-WSDM'15:Workshop on Deep Learning for web Search and Data Mining (full day)
- The 2nd workshop on Vertical Search Relevance at WSDM 2015(half day)
- WSDM'15 Workshop Summary / Scalable Data Analytics: Theory and Applications(half day)

HIA'15: Heterogeneous Information Access Workshop at WSDM 2015

Organisers: Ke Zhou *(Yahoo Labs)*, Roger Jie Luo *(Yahoo Labs)*, Djoerd Hiemstra *(University of Twente)*, Joemon M. Jose *(University of Glasgow)*

Information access is becoming increasingly heterogeneous. Especially when the user's information need is for exploratory purpose, returning a set of diverse results from different re-sources could benefit the user. Aggregated search and composite retrieval are two instances of this new heterogeneous information access paradigm. Compared with traditional homogeneous search, optimization and evaluation in the context of heterogeneous information is more challenging and requires taking into account more complex user be-

haviors and interactions. We would like to create a forum to encourage discussion and exchange of ideas on heterogeneous information access in different contexts. To facilitate the discussion, we encourage submissions on ideas and results from different aspects of heterogeneous information access including aggregated search, composite retrieval, personal search, structured search, etc. Another objective of the workshop is to encourage submissions with novel ideas (e.g. new applications) on heterogeneous information access and potential future directions of this area.

DL-WSDM'15: Deep Learning for web Search and Data Mining

Organisers: Bin Gao *(Microsoft Research)*, Jiang Bian *(Microsoft Research)*

Deep learning has been a very hot topic in the machine learning community. It has brought break-through results in image classification and speech recognition. Most recently, researchers have also got many promising results in natural language processing using deep learning techniques. As machine learning techniques are widely used in the Web search and data mining applications, many researchers and practitioners are studying the possibility of applying the recently-developed deep learning techniques into these applications. Some of them have made very promising progress, and thus it is a good time to hold a workshop to discuss and share the problems and progress in using deep learning techniques to improve Web search and data mining tasks. The main objective and goal of this workshop is to bring together researchers and practitioners that are applying machine learning (especially deep learning) techniques in Web search and data mining tasks, and enable them

to share their latest research results, to express their opinions, and to discuss future directions. We expect that the WSDM audience get more familiar with the progress of deep learning and get more willing to try deep learning techniques in their research or application problems.

The 2nd workshop on Vertical Search Relevance at WSDM 2015

Organisers: Dawei Yin *(Yahoo Labs)*, Chih-Chieh Hung *(Rakuten Inc.)*, Rui Li *(Yahoo Labs)*, Yi Chang *(Yahoo Labs)*

As the web information exponentially grows and the needs of users become more specific, traditional general web search engines are not able to perfectly satisfy the nowadays user requirement. Vertical search engines have emerged in various domains, which more focus on specific segments of online content, including local, shopping, medical information, travel search, etc. Vertical search engines start attracting more attention while relevance ranking in different vertical search engines is becoming the key technology. In addition, vertical search results are often slotted into general Web search results. Hence, designing effective ranking functions for vertical search has become practically important to improve users' experience in both web search and vertical search. The workshop bring together researchers from IR, ML, NLP, and other areas of computer and information science, who are working on or interested in this area. It provides a forum for the researchers to identify the issues and the challenges, to share their latest research results, to express a diverse range of opinions about this topic, and to discuss future directions.

WSDM'15 Workshop Summary / Scalable Data Analytics: Theory and Applications

Organisers: Kaizhu Huang *(Xi'an Jiaotong-Liverpool University)*, Haiqin Yang *(Chinese University of Hong Kong)*, Irwin King *(Chinese University of Hong Kong)*, Michael R. Lyu *(Chinese University of Hong Kong)*

With the fast evolving technology for data collection, data transmission, and data analysis, the scientific, biomedical, and engineering research communities are undergoing a profound transformation where discoveries and innovations increasingly rely on massive amounts of data. New prediction techniques, including novel statistical, mathematical, and modeling techniques are enabling a paradigm shift in scientific and biomedical investigation. Data become the fourth pillar of science and engineering, offering complementary insights in addition to theory, experiments, and computer simulation. Advances in machine learning, data mining, and visualization are enabling new ways of extracting useful information from massive data sets. The characteristics of volume, velocity, variety and veracity bring challenges to current data analytics techniques. It is desirable to scale up data analytics techniques for modeling and analyzing big data from various domains. The workshop aims to provide professionals, researchers, and technologists with a single forum where they can discuss and share the state-of-the-art theories and applications of scalable data analytics technologies.