



# The 2016 International Conference on Data Science

Shanghai | December 16-19, 2016



**Host**

School of Data Science and Institute for Big Data, Fudan University

**Sponsor**



# General Program

December 17<sup>th</sup>: Hyatt Regency Shanghai (凯悦酒店)

Time	Event	Venue
11:30-12:15	<b>Keynote Speech</b> (Fudan Forum) <b>Michael Jordan</b> (University of California, Berkeley) <i>On computational thinking, inferential thinking and data science</i>	Regency Ballroom
12:20-13:30	Lunch	Hyatt Regency Shanghai
13:30-14:15	Opening Ceremony Launch Ceremony of Fudan-Zhongzhi Big Data Institute for Finance and Investment	Regency Ballroom II & III
14:20-15:10	<b>Keynote Speech, Chair: Zhaojun Bai</b> <b>Wen Gao</b> (Peking University) <i>Multimedia big data processing</i>	Regency Ballroom II & III
15:10-15:30	Tea Break	
	<b>Session I: Structured Inference with High-dimensional Data</b> <b>Chair: Cun-Hui Zhang, Rutgers University</b>	Events V
15:30-17:30	<ul style="list-style-type: none"> <li><i>Pairwise difference approach attains minimax optimality in high dimensional partially linear model</i>  <b>Fang Han</b> (University of Washington)</li> <li><i>Inference in high-dimensional semi-parametric graphical models</i>  <b>Mladen Kolar</b> (University of Chicago)</li> <li><i>Contraction-expansion approach to simultaneous inference in linear models</i>  <b>Peter Song</b> (University of Michigan)</li> <li><i>Randomizing the simplex: Scalable and accurate inference for clustered categorical data</i>  <b>Wei Lin</b> (Peking University)</li> <li><i>On structure testing for component covariance matrices of a high-dimensional mixture</i>  <b>Jianfeng Yao</b> (University of Hong Kong)</li> </ul>	
	<b>Session II: Industrial Big Data</b> <b>Chair: Tong Zhang, Rutgers University</b>	Regency Ballroom II & III
	<ul style="list-style-type: none"> <li><i>Distributed machine learning platform and its applications in ant financial</i>  <b>Xiaoguang Wang</b> (Alibaba)</li> <li><i>A competitor of random Fourier features</i>  <b>Ping Li</b> (Rutgers University)</li> <li><i>Binary embeddings for big data applications</i>  <b>Wei Liu</b> (Tencent)</li> <li><i>Practices of applying big data in financial risk management and credit scoring in China</i>  <b>Mingjie Zhu</b> (CreditX)</li> </ul>	

Time	Event	Venue
15:30-17:30	<b>Session III: Big Data Business: what's new in the market?</b> <b>Chair: Libo Wu, Fudan University</b>	Events II
	<ul style="list-style-type: none"> <li>• <i>Big data, the fourth industrial revolution to change the world</i>  <b>Sanshou Wang</b> (Jusfour Big Data Group Co., Ltd)</li> <li>• <i>Calculation+, data driving force</i>  <b>Yezhi Sun</b> (Inspur Group Shanghai Branch)</li> <li>• <i>The new financial regulation: big data is promising</i>  <b>Jiong Ding</b> (Shanghai Promotion Center of Cloud Computing Industry)</li> <li>• <i>Big data for ICT applications and thinking</i>  <b>Yao Li</b> (Shanghai Ideal Information Industry Co., Ltd)</li> <li>• <i>The evolution of big data technologies and the trends</i>  <b>Yuanhao Sun</b> (Transwarp Technology)</li> </ul>	
	<b>Session IV: Big Data Application in Industries</b> <b>Chair: Xiaohua Xuan, Hua Data Technology</b>	Events III
	<ul style="list-style-type: none"> <li>• <i>Big data/AI applications and incubation</i>  <b>Xiaohua Xuan</b> (Hua Data Technology)</li> <li>• <i>Data driven business in China: opportunity and challenge</i>  <b>Dondong Ge</b> (Cardinal Operations)</li> <li>• <i>Property data solutions for financial services</i>  <b>Xiaohai Zhang</b> (Shanghai Gujia Network Technology)</li> <li>• <i>Data tomography of online consumer credit</i>  <b>Roger Gu</b> (Wacai)</li> </ul>	
6:30pm	Dinner	Yan Yuan Hotel ( 燕园 )

**December 18<sup>th</sup>: East Sub-building of Guanghua Towers  
( 光华东辅楼 ), Fudan University**

Time	Event	Venue
08:30-09:20	<b>Keynote Speech, Chair: Eric Xing</b> <b>Michael Franklin</b> (University of Chicago) <i>Distributed and streaming database technology</i>	Room 102
09:20-09:40	Tea Break	
09:40-10:20	<b>Distinguished Speech, Chair: Zhengjun Zhang</b> <b>Cun-Hui Zhang</b> (Rutgers University) <i>Beyond Gaussian approximation: bootstrap in large scale</i>	Room 102
	<b>Distinguished Speech, Chair: Tong Zhang</b> <b>Haixun Wang</b> (Facebook) <i>Short text understanding</i>	Room 101
10:30-12:30	<b>Session I: New Statistical Analytic Tools for High-dimensional Data and Related Topics</b> <b>Chair: Nan Zhang, Fudan University</b>	4 <sup>th</sup> Floor, Meeting Room 1
	<ul style="list-style-type: none"> <li>• <i>Embracing blessing of massive scale in big data</i>  <b>Guang Cheng</b> (Purdue University)</li> <li>• <i>Principal sub-manifolds</i>  <b>Zhigang Yao</b> (Nation University of Singapore)</li> <li>• <i>High-dimensional A-learning for dynamic treatment regimes</i>  <b>Rui Song</b> (NC State University)</li> <li>• <i>Rank-consistency of the generalized Bradley-terry model with link misspecification</i>  <b>Yongdai Kim</b> (Seoul National University)</li> </ul>	

Time	Event	Venue
	<b>Session II: Recent Advances in Statistical Methodologies for High-dimensional Data</b> <b>Chair: Yin Xia, Fudan University</b> <ul style="list-style-type: none"> <li><i>Wild residual bootstrap inference for penalized quantile regression with heteroscedastic error</i>  <b>Lan Wang</b> (University of Minnesota)</li> <li><i>Tensor response regression and application in neuroimaging analysis</i>  <b>Lexin Li</b> (University of California, Berkeley)</li> <li><i>On imbalanced multiclass high dimensional low sample size classification</i>  <b>Lingsong Zhang</b> (Purdue University)</li> <li><i>Spatial adaptivity in trend filtering</i>  <b>Adityanand Guntuboyina</b> (University of California, Berkeley)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 2
	<b>Session III: Network Analysis</b> <b>Chair: Mark Wolters, Fudan University</b> <ul style="list-style-type: none"> <li><i>Confidence sets for the source of a diffusion in regular trees</i>  <b>Po-Ling Loh</b> (University of Wisconsin-Madison)</li> <li><i>Statistical and computational limits of recovery of a single community in networks</i>  <b>Yihong Wu</b> (Yale University)</li> <li><i>DNA seriation under planted Hamiltonian path model</i>  <b>Jiaming Xu</b> (Purdue University)</li> <li><i>Community detection in degree-corrected block models</i>  <b>Anderson Zhang</b> (Yale University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 3
10:30-12:30	<b>Session IV: Multi-media Data Processing/Vision</b> <b>Chair: Yu-Gang Jiang, Fudan University</b> <ul style="list-style-type: none"> <li><i>Video content recognition with deep learning</i>  <b>Yu-Gang Jiang</b> (Fudan University)</li> <li><i>Empowering ChatBot with all-seeing eyes</i>  <b>Tao Mei</b> (Microsoft Research Asia)</li> <li><i>Collaborative deep reinforcement learning for joint object search</i>  <b>Yizhou Wang</b> (Peking University)</li> <li><i>Network representation: a revisit in the big data era</i>  <b>Peng Cui</b> (Tsinghua University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 4
	<b>Session V: Natural Language Processing</b> <b>Chair: Xiangyang Xue, Fudan University</b> <ul style="list-style-type: none"> <li><i>Challenges and thinking of machine translation and natural language understanding on big data</i>  <b>Min Zhang</b> (Soochow University)</li> <li><i>The role of information retrieval in big data era</i>  <b>Jianyun Nie</b> (Université de Montréal)</li> <li><i>NLP for microblog summarization</i>  <b>Kam-Fai Wong</b> (The Chinese University of Hong Kong)</li> <li><i>Deep Learning for natural language processing</i>  <b>Xuanjing Huang</b> (Fudan University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 5
	<b>Session VI: Big Graph</b> <b>Chair: Xuemin Lin, University of New South Wales</b> <ul style="list-style-type: none"> <li><i>Towards processing of big graphs, theory, algorithms, and systems</i>  <b>Xuemin Lin</b> (University of New South Wales)</li> <li><i>Community search in massive networks</i>  <b>Hong Cheng</b> (The Chinese University of Hong Kong)</li> <li><i>Towards big graph search: challenges &amp; techniques</i>  <b>Shuai Ma</b> (Bei Hang University)</li> <li><i>Natural language question answering over knowledge graph ---- A data-driven approach</i>  <b>Lei Zou</b> (Peking University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 6

Time	Event	Venue
12:30	Group Photo	East Gate of Guang Hua Towers
12:30-14:00	Lunch	3 <sup>rd</sup> Floor, Dan Yuan Cafeteria ( 旦苑 )
14:00-14:40	<p><b>Distinguished Speech, Chair: Jianqing Fan</b>  <b>Martin Wainwright</b> (University of California, Berkeley)  <i>Statistics meets optimization: challenges at the interface</i></p>	Room 102
14:40-15:00	<p><b>Distinguished Speech, Chair: Yuanyuan Zhou</b>  <b>Eric Xing</b> (Carnegie Mellon University)  <i>Strategies &amp; Principles for Distributed Machine Learning</i></p>	Room 101
14:40-15:00	Tea Break	
	<p><b>Session I: Recent Advances on High-dimensional and Nonparametric Statistics</b>  <b>Chair: Tony Cai, University of Pennsylvania</b></p> <ul style="list-style-type: none"> <li>• <i>Estimating the absolute value of the inner product of two normal mean vectors</i>  <b>Mark Low</b> (University of Pennsylvania)</li> <li>• <i>Composition estimation from sparse count data via a regularized likelihood</i>  <b>Hongzhe Li</b> (University of Pennsylvania)</li> <li>• <i>Multiple testing of submatrices of a precision matrix</i>  <b>Yin Xia</b> (Fudan University)</li> <li>• <i>Statistical inference for model parameters with stochastic gradient descent</i>  <b>Xi Chen</b> (New York University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 1
15:00-17:00	<p><b>Session II: High-dimensional Statistical Inference</b>  <b>Chair: Weidong Liu, Shanghai Jiao Tong University</b></p> <ul style="list-style-type: none"> <li>• <i>FDR control in large scale tests of independence: a first step</i>  <b>Shan Luo</b> (Shanghai Jiao Tong University)</li> <li>• <i>On dimensionality effects in linear discriminant analysis for large dimensional data</i>  <b>Cheng Wang</b> (Shanghai Jiao Tong University)</li> <li>• <i>A lack-of-fit test with screening in sufficient dimension for ultrahigh dimensional data</i>  <b>Liping Zhu</b> (Renmin University)</li> <li>• <i>Variable selection and direction estimation for single-index models via DC-TGDR method</i>  <b>Wei Zhong</b> (Xiamen University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 2
	<p><b>Session III: Financial Econometrics</b>  <b>Chair: Yazhen Wang, University of Wisconsin-Wadison</b></p> <ul style="list-style-type: none"> <li>• <i>Statistical analysis of combined low-frequency and high-frequency financial data</i>  <b>Yazhen Wang</b> (University of Wisconsin-Wadison)</li> <li>• TBA  <b>Xiaohong Chen</b> (Yale University)</li> <li>• <i>Solving the high-dimensional Markowitz optimization problem: when sparse regression meets random matrix theory</i>  <b>Yingying Li</b> (Hong Kong University of Science and Technology)</li> <li>• <i>Testing high-dimensional covariance matrices when heteroscedasticity is present</i>  <b>Xinghua Zheng</b> (Hong Kong University of Science and Technology)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 3

Time	Event	Venue
	<b>Session IV: Nonparametric and High Dimensional Inferences</b> <b>Chair: Ming Yuan, University of Wisconsin-Madison</b> <ul style="list-style-type: none"> <li><i>Distributed estimation and inference with statistical guarantees</i> <b>Junwei Lu</b> (Princeton University)</li> <li><i>Efficient multivariate entropy estimation via <math>k</math>-nearest neighbor distances</i> <b>Richard Samworth</b> (University of Cambridge)</li> <li><i>Estimation of a two-component mixture model with applications to multiple testing</i> <b>Bodhi Sen</b> (Columbia University)</li> <li><i>Convex and non-convex approaches for high-dimensional tensor regression</i> <b>Garvesh Raskutti</b> (University of Wisconsin-Madison)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 4
15:00-17:00	<b>Session V: Algorithm Foundation</b> <b>Chair: Weiguo Gao, Fudan University</b> <ul style="list-style-type: none"> <li><i>Modelling, regularization algorithms and applications</i> <b>Jin Cheng</b> (Fudan University)</li> <li><i>Consistency-driven multiple graph matching</i> <b>Junchi Yan</b> (IBM Research, China)</li> <li><i>Matrix optimization: recent progress on algorithm foundation</i> <b>Chao Ding</b> (Chinese Academy of Sciences)</li> <li><i>Online transfer learning</i> <b>Michael Ng</b> (Hong Kong Baptist University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 5
	<b>Session VI: Knowledge Graph and AI</b> <b>Chair: Deqing Yang, Fudan University</b> <ul style="list-style-type: none"> <li><i>Experiences and challenges to build large scale customer care</i> <b>Junlan Feng</b> (Institute for Big Data, ChinaMobile Lab)</li> <li><i>Knowledge graphs: ontologies in the big data era</i> <b>Jeff Pan</b> (University of Aberdeen)</li> <li><i>Enabling machines to understand human language by knowledge graphs</i> <b>Yanghua Xiao</b> (Fudan University)</li> <li><i>Computational translation of model experiments based on big genomic data</i> <b>Junhee Seok</b> (Korea University)</li> </ul>	4 <sup>th</sup> Floor, Meeting Room 6
18:00	Banquet	Nan Yue Chun ( 南粤春 )

# December 19<sup>th</sup>: East Sub-building of Guanghua Towers (光华东辅楼), Fudan University

Time	Event	Venue
08:30-09:20	<b>Keynote Speech, Chair: Qiwei Yao</b> <b>Jianqing Fan</b> (Princeton University & Fudan University) <i>Principles of robustification for big data with financial applications</i>	<b>Room 102</b>
09:20-09:40	Tea Break	
09:40-10:20	<b>Distinguished Speech, Chair: Xuming He</b> <b>Jianhua Guo</b> (Northeast Normal University of China) <i>Latent structural detection for high dimensional data via latent tree models</i>	<b>Room 101</b>
	<b>Distinguished Speech, Chair: Chengxiang Zhai</b> <b>Yuanyuan Zhou</b> (University of California, San Diego) <i>Applying big data analysis in cloud infrastructure management and beyond</i>	<b>Room 103</b>
	<b>Session I: Big Data in Bio-Sciences</b> <b>Chair: Xuming He, University of Michigan</b>	<b>4<sup>th</sup> Floor, Meeting Room 1</b>
10:30-12:30	<ul style="list-style-type: none"> <li>• <i>Integrating genome wide association signals and genomic annotations in disease risk prediction</i>  <b>Hongyu Zhao</b> (Yale University)</li> <li>• <i>Robust hybrid learning for estimating dynamic treatment regimes in personalized medicine</i>  <b>Donglin Zeng</b> (University of North Carolina at Chapel Hill)</li> <li>• <i>Latent Gaussian mixture models for nationwide Kidney Transplant Center evaluation</i>  <b>Yi Li</b> (University of Michigan)</li> <li>• <i>Feature-based regression models for imaging biomarker detection in current medical research and practice</i>  <b>Hongtu Zhu</b> (University of North Carolina at Chapel Hill)</li> </ul>	
	<b>Session II: Recent Developments in Multiple Testing and Post-Selection Inference</b> <b>Chair: Wenguang Sun, University of Southern California</b>	<b>4<sup>th</sup> Floor, Meeting Room 2</b>
	<ul style="list-style-type: none"> <li>• <i>Matrix completion, saddle points, and gradient descent</i>  <b>Jason Lee</b> (University of Southern California)</li> <li>• <i>Private false discovery rate control and robustness of the Benjamini-Hochberg procedure</i>  <b>Weijie Su</b> (University of Pennsylvania)</li> <li>• <i>CARS: covariate assisted ranking and screening for large-scale two-sample inference</i>  <b>Wenguang Sun</b> (University of Southern California)</li> <li>• <i>False discovery rate control for identifying simultaneous signals</i>  <b>Dave Zhao</b> (University of Illinois at Urbana-Champaign)</li> </ul>	

Time	Event	Venue
10:30-12:30	<b>Session III: High-dimensional Dependent Data</b> <b>Chair: Qiwei Yao, The London School of Economics and Political Science</b>	4 <sup>th</sup> Floor, Meeting Room 3
	<ul style="list-style-type: none"> <li><i>Traffic flow analysis and prediction using functional data analysis</i> <b>Jeng-Min Chiou</b> (Academia Sinica)</li> <li><i>DECO related feature space partitioning for distributed sparse regression</i> <b>Chenlei Leng</b> (University of Warwick)</li> <li><i>Data visualization and modeling evolving dynamics of brain signals</i> <b>Hernando Ombao</b> (University of California, Irvine)</li> <li><i>Identifying genetic variants for learning ability with neuroimaging</i> <b>Heping Zhang</b> (Yale University)</li> </ul>	
	<b>Session IV: Big Data Systems</b> <b>Chair: Feifei Li, University of Utah</b>	4 <sup>th</sup> Floor, Meeting Room 4
	<ul style="list-style-type: none"> <li><i>Simba: Towards building interactive big spatial data systems</i> <b>Feifei Li</b> (University of Utah)</li> <li><i>Scalable in-memory transaction processing using HTM and RDMA</i> <b>Haibo Chen</b> (Shanghai Jiao Tong University)</li> <li><i>Hybrid human-machine big data integration</i> <b>Guoliang Li</b> (Tsinghua University)</li> <li><i>Data-driven approaches for large scale knowledge graph construction</i> <b>Yanhua Xiao</b> (Fudan University)</li> </ul>	
12:30-14:00	<b>Session V: Text Data Retrieval and Mining</b> <b>Chair: Chengxiang Zhai, University of Illinois at Urbana-Champaign</b>	4 <sup>th</sup> Floor, Meeting Room 4
	<ul style="list-style-type: none"> <li><i>Analysis and mining of big text data: Opportunities and challenges</i> <b>Chengxiang Zhai</b> (University of Illinois at Urbana-Champaign)</li> <li><i>Understanding users for search and recommendation</i> <b>Min Zhang</b> (Tsinghua University)</li> <li><i>Deep learning for relevance matching in information retrieval</i> <b>Yanyan Lan</b> (Chinese Academy of Sciences)</li> <li><i>Text summarization in the big data era</i> <b>Xiaojun Wan</b> (Peking University)</li> </ul>	
12:30-14:00	Lunch	Yan Yuan Hotel

## Registration

Date	Time	Venue
Fri, Dec 16	14:00-20:00	Fuxuan Hotel ( 复宣酒店 )
Sat, Dec 17	09:00-17:00	Hyatt Regency Shanghai ( 凯悦酒店 )
Sun, Dec 18	09:00-17:00	East Sub-building of Guanghua Towers ( 光华东辅楼 )
Mon, Dec 19	09:00-12:00	East Sub-building of Guanghua Towers ( 光华东辅楼 )

## Shuttle Bus Schedule

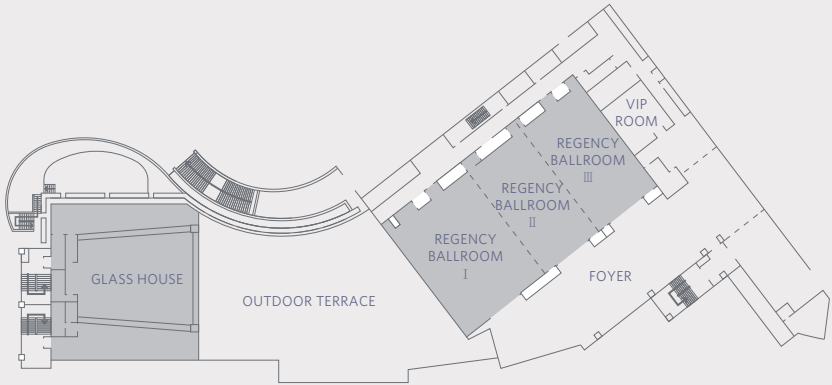
Departure Time	Itinerary	Venue	
Dec 17	08:20	Crowne Plaza Fudan → Hyatt Regency	Gate of Crowne Plaza Fudan
	08:20	Fuxuan Hotel → Hyatt Regency	Gate of Fuxuan Hotel
	18:00	Hyatt Regency → Yan Yuan Hotel	Gate of Hyatt Regency
Dec 18	17:30	East Guanghua Tower → Nan Yue Chun	Gate of East Guanghua Tower
	20:30	Nan Yue Chun → Fuxuan Hotel	Gate of Nan Yue Chun
	20:30	Nan Yue Chun → Crowne Plaza Fudan	Gate of Nan Yue Chun



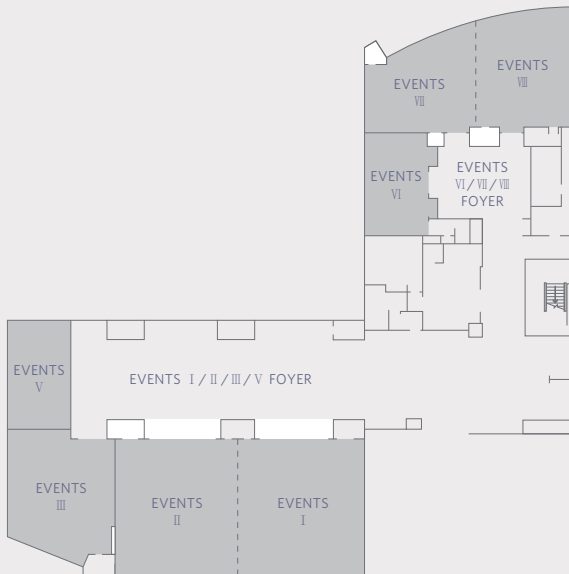
HYATT REGENCY SHANGHAI, WUJIAOCHANG  
88 East Guoding Road, Yangpu District  
Shanghai 200433, China  
T +86 21 2565 1234  
shanghaiwujiaochang.regency.hyatt.com  
shanghaiwujiaochang.regency@hyatt.com



FLOOR PLAN  
Seventh Floor



FLOOR PLAN  
Sixth Floor



# 复旦大学 Fudan University



## Organizing Committee

- Zhaojun Bai
- Tony Cai
- Jianqing Fan
- Weiguo Gao
- Zhengjun Zhang

## Contacts

School of Data Science  
Fudan University  
Tel: 86-21-65648878  
Email: [sds\\_forum@fudan.edu.cn](mailto:sds_forum@fudan.edu.cn)  
<http://www.fbdc.fudan.edu.cn/>