

Manual

**2017 IEEE International Conference
on Big Knowledge**

ICBK 2017

Welcome Message from Conference Organizers

Welcome to IEEE ICBK 2017, the 8th IEEE International Conference on Big Knowledge, and to Hefei, the beautiful capital of Anhui Province of China!

Big Knowledge deals with fragmented knowledge from heterogeneous, autonomous information sources for complex and evolving relationships, in addition to domain expertise. The IEEE International Conference on Big Knowledge (ICBK) is a premier international forum for presentation of original research results in Big Knowledge opportunities and challenges, as well as for exchange and dissemination of innovative, practical development experiences. The conference covers all aspects of Big Knowledge, including algorithms, software, systems, and applications. ICBK draws researchers and application developers from a wide range of Big Knowledge related areas such as statistics, machine learning, pattern recognition, knowledge visualization, expert systems, high performance computing, World Wide Web, and big data analytics. By promoting novel, high quality research findings, and innovative solutions to challenging Big Knowledge problems, the conference seeks to continuously advance the state-of-the-art in Big Knowledge. ICBK 2017 is the first international edition, and the 8th edition in its conference history. The previous 7 editions (annually from 2010 through 2016) were all organized in Hefei, China, always on August 9-10.

This year, the conference received 70 paper submissions from all over the world, and we accepted 19 of them as regular papers in this proceedings, an acceptance rate of 27%. In addition, there are 6 workshops which have attracted 32 papers for presentation.

We are delighted to welcome our prominent keynote speakers: Professor Ming Li from the University of Waterloo, a Fellow of the Royal Society of Canada, ACM, and IEEE., and Professor Jian Pei from Simon Fraser University, a Fellow of the ACM and IEEE.

The organization of a successful conference would not be possible without the dedicated efforts of many individuals. We would like to express our gratitude to all functional chairs on our organizing committee listed on a separate page of this proceedings. We owe special thanks to our conference sponsors: the financial and organizational support of the National Key Research and Development Program of China under grant 2016YFB1000901 and its 15 participating institutions, the IEEE Computer Society, the Anhui Association for Artificial Intelligence and the Anhui Association for Science and Technology. We especially thank the local institutions that have supported the conference, in particular Hefei University of Technology and the University of Science and Technology of China. Last but not least, we would like to thank all authors who submitted research papers to the conference and its workshops, and all participants. We are

encouraged by your scientific contributions, support and participation to explore further this new emerging field of Big Knowledge.

Big Knowledge is still growing. We wish all participants a productive conference with new discoveries, new collaboration opportunities and a pleasant visit to Hefei, China.

Xindong Wu and Tamer Özsu, **ICBK 2017 Program Chairs**

Jim Hendler and Ruqian Lu, **ICBK 2017 Conference Chairs**

August 9-10, 2017

Conference Organization

Honorary Chair

Nanning Zheng, *Xi'an Jiaotong University, China*

Conference Chairs

Jim Hendler, *Rensselaer Polytechnic Institute, USA*

Ruqian Lu, *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

Program Committee Chairs

Xindong Wu, *Hefei University of Technology, China and University of Louisiana at Lafayette, USA*

Tamer Ozsu, *University of Waterloo, Canada*

Panel Chairs

Guoyin Wang, *Chongqing University of Posts and Telecommunications, China*

Witold Pedrycz, *University of Alberta, Canada*

Finance Chair

Lei Li, *Hefei University of Technology, China*

Local Arrangement Chair

Qiping Zhang, *Hefei University of Technology, China*

Publication Chair

Huanhuan Chen, *University of Science and Technology of China*

Award Committee Chairs

Jun Liu, *Xi'an Jiaotong University, China*

Michel Dumontier, *Stanford University, USA*

PhD Forum Chairs

Zhi Jin, *Peking University, China*

Dimitris Karagiannis, *University of Vienna, Austria*

Sponsorship Chairs

Dandan Song, *Beijing Institute of Technology, China*

Angel Rubio, *Ceit-IK4 Research Alliance, Spain*

Publicity Chairs

Xiaolong Jin, *Institute of Computing Technology, Chinese Academy of Sciences*

Francoise Soulie Fogelman, *KXEN, France and Tianjing University, China*

Tutorial Chairs

Songmao Zhang, *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

Martin Wirsing, *Ludwig-Maximilian's University of Munich, Germany*

Workshops Chairs

Enhong Chen, *University of Science and Technology of China*

Alfredo Cuzzocrea, *University of Trieste, Italy*

Workshop Committees

ICBK Workshop on Big Data Partitioning and Mining

Organizing Committee

Xiaofeng Zhu, *University of North Carolina at Chapel Hill, USA*

Jilian Zhang, *Guangxi University of Finance and Economics, China*

Shichao Zhang, *Guangxi Normal University, China*

ICBK Workshop on Linked Data Mining (LDM 2017)

Organizing Committee

Prof. Jun Liu, *Xi'an Jiaotong University, China*

Prof. Zheng Yan, *Aalto University, Finland / Xidian University, China*

ICBK Workshop on Research and Application of Big Knowledge in Urban Governance

Organizing Committee

Kehui Liu, *Beijing Institute of New Technology Application, China*

Yifang Jiang, *Tianjin Institute of Geotechnical Investigation Surveying, China*

Huanhuan Chen, *University of Science and Technology of China, China*

ICBK Workshop on Sparse, Uncertain, and Incomplete Data Modeling and Online Learning (DMOL 2017)

Organizing Committee

Dr. Zhongqiu Zhao, *Hefei University of Technology, China*

Dr. Peipei Li, *Hefei University of Technology, China*

ICBK Workshop on Analyzing and Predicting Interaction Behaviors (APIB 2017)

Organizing Committee

Ming Gao, *East China Normal University, China*

Qi Liu, *University of Science and Technology of China, China*

ICBK Workshop on Dynamic Knowledge Modeling and Evolution (DKME 2017)

Keynote 1

Tracking in Dynamic Networks

Jian Pei, *Simon Fraser University*



Abstract

In many application scenarios ranging from social networks to IoT, we need to process and analyze a huge amount of data, connected, evolving, linkages being more interesting than entities individually. Modeling such temporal data in nature as graphs provides a conceptually convenient way to support novel and meaningful intelligent applications. At the same time, it also posts great challenges in many aspects, such as algorithm design and computing system development. In this talk, I will present some interesting and novel application scenarios where graphs play a central role, as well as the corresponding algorithms. Moreover, I will briefly introduce our on-going effort to build a distributed cloud-based graph computing engine that can query huge graphs and networks in seconds.

Short Bio:

Jian Pei is a Canada Research Chair (Tier 1) in Big Data Science and a Professor in the School of Computing Science at Simon Fraser University. His expertise is on developing effective and efficient data analysis techniques for novel data intensive applications. He is one of the most cited authors in data mining, database systems, and information retrieval. Since 2000, he, with H-index 73, has published one textbook, two monographs and over 200 research papers in refereed journals and conferences, which have been cited by more than 67,000 in literature. His research has generated remarkable impact substantially beyond academia. He is the recipient of several prestigious awards, such as the IEEE ICDM Research Contributions Award and the ACM SIGKDD Service Award. He is an ACM Fellow and an IEEE Fellow.

Keynote 2

Deep Learning in Bioinformatics

Ming Li, *University of Waterloo*



Abstract

I will discuss several applications of deep learning in bioinformatics, and focus on our recent work on de novo peptide sequencing by deep learning. De novo peptide sequencing from tandem mass spectrometry data is the key technology in proteomics for the characterization of proteins, especially for antibodies. During the past 20 years, many traditional computer science algorithms were developed for this problem. These algorithms include dynamic programming, linear programming, HMM, graph algorithms, and statistical methods. Deep learning significantly improves these traditional approaches. This is joint work with H.N. Tran, X. Zhang, L. Xin and P. Shan, to appear in PNAS.

Short Bio:

Dr. Ming Li is a Canada Research Chair in Bioinformatics and a University Professor at the University of Waterloo. He is a fellow of the Royal Society of Canada, ACM, and IEEE. He is a recipient of E.W.R. Steacie Fellowship Award in 1996, the 2001 Killam Fellowship, and the 2010 Killam Prize. Together with Paul Vitanyi they have co-authored the book *An Introduction to Kolmogorov Complexity and Its Applications*.

The 8th IEEE International Conference on Big Knowledge Agenda

Wednesday , August 9, 2017			
7:00 am to 12:00 am	Registration Hotel Lobby		
9:00 am to 10:00 am	Keynote 1: Jian Pei, Simon Fraser University Tracking in Dynamic Networks Chair: Xindong Wu, Room: Function Room1, 多功能1厅 (Second Floor)		
10:00 am to 10:20 am	Coffee Break		
10:20 am to 12:00 pm	Session 1: Statistics Learning Chair: Jilian Zhang Room: Function Room1, 多功能1厅(Second Floor)	Session 2: Recommendation Systems Chair: Guoyin Wang Room: Lu Gu, 泸沽(Second Floor)	Session 3 : Knowledge Engineering Chair: Guozhu Dong Room: Qian Dao, 千岛(Second Floor)
12:00 pm to 14:00 pm	Lunch Break Wan Coffee, 皖咖啡(Ground Floor)		
14:00 pm to 15:40 pm	Session 4: Dynamic Systems Chair: Xuguang Lan Room: Function Room1, 多功能1厅	Workshop 1-1: Big Data Partitioning and Mining Chair: Peng Liu Room: Lu Gu, 泸沽	Workshop 3-1: Research and Application of Big Knowledge in Urban Governance Chair: Yu-Sheng Hu Room: Qian Dao, 千岛
15:40 pm to 16:10 pm	Coffee Break		
16:10 pm to 17:50 pm	Workshop 5-1: Analyzing and Predicting Interaction Behaviors Chair: Xiaofeng He Room: Function Room1, 多功能1厅	Workshop 1-2: Big Data Partitioning and Mining Chair: Shijie Hao Room: Lu Gu, 泸沽	Workshop 3-2: Research and Application of Big Knowledge in Urban Governance Chair: LiYuan Geng Room: Qian Dao, 千岛

18:00 pm to 20:00 pm	Banquet Shangri-La ballroom A, 香格里拉大宴会 A 厅(Second Floor)		
Thursday, August 10, 2017			
9:00 am to 10:00 am	Keynote 2: Ming Li, University of Waterloo Deep Learning in Bioinformatics Chair: Tamer Ozsu Room: Function Room1,多功能 1 厅		
10:00 am to 10:20 am	Coffee Break		
10:20 am to 12:00 pm	Workshop 5-2: Analyzing and Predicting Interaction Behaviors Chair: Ming Gao Room: T Function Room1, 多功能 1 厅	Workshop 2: Unsupervised Learning Chair: Bifan Wei Room: Lu Gu, 泸沽	Workshop 4: Sparse, Uncertain and Incomplete Data Modeling and Online Learning Chair: Hong Yu Room: Qian Dao, 千岛
12:00 pm to 14:00 pm	Lunch Break (Wan Coffee)		
14:00 pm to 15:30 pm	Panel 1: Big Graph Processing and Analytics Chair: Tamer Ozsu Room: Qian Dao, 千岛	Panel 2: Knowledge Engineering with Big Data Chair: Xindong Wu Room: Function Room1,多功能 1 厅	

Session 1: Statistics Learning. Wednesday, August 9, 2017, 10:20 am to 12:00 pm

- Multi-label Classification Using Stacked Hierarchical Dirichlet Processes with Reduced Sampling Complexity
Sophie Burkhardt and Stefan Kramer
- Analyzing Structure of Terrorist Networks by Using Graph Metrics
Muhammet Serkan Cinar, Burkay Genc, Hayri Sever, and Vijay V Raghavan
- A Global Flight Networks Analysis Approach Using Markov Clustering and PageRank
Kecheng Xu, Teng Long, Shaojie Qiao, Yating Zheng, and Nan Han
- Multi-layer Big Knowledge Visualization Scheme for Comprehending Neoplasm Ontology Content
Ling Zheng, Christopher Ochs, James Geller, Hao Liu, Yehoshua Perl, and Sherri De Coronado
- A New Online Feature Selection Method Using Neighborhood Rough Set
Peng Zhou, Xuegang Hu, and Peipei Li

Session 2: Recommendation Systems. Wednesday, August 9, 10:20am to 12:00pm

- Privacy-Preserving Pattern Mining on Online Density Estimates
Michael Geilke and Stefan Kramer
- Exploiting Nonlinear Relationships for Top-N Recommender Systems
Zhao Kang, Chong Peng, Ming Yang, and Qiang Cheng
- Iteratively Multiple Projections Optimization for Product Quantization in Nearest Neighbor Search
Jin Li, Xuguang Lan, and Nanning Zheng
- Factored Proximity Models for Top-N Recommendations
Athanasios N. Nikolakopoulos, Vassilis Kalantzis, Efstratios Gallopoulos, and John D. Garofalakis
- A Novel Three-Way Clustering Algorithm for Mixed-Type Data
Hong Yu, Zhihua Chang, and Bing Zhou

Session 3: Knowledge Engineering. Wednesday, August 9, 10:20 am to 12:00 pm

- Subpopulation-Wise Conditional Correlation Modeling and Analysis
Guozhu Dong and Sanjeev Bhatta
- Self-Learning and Embedding Based Entity Alignment
Saiping Guan, Xiaolong Jin, Yantao Jia, Yuanzhuo Wang, Huawei Shen, and Xueqi Cheng
- Towards Question Improvement on Knowledge Sharing Platforms: A Stack Overflow Case Study
Rishabh Gupta and P. Krishna Reddy
- TDN: Twice-Least-Square Double-Parallel Neural Networks
Guo-Qiang Li and Keith C.C. Chan
- Keyphrase Extraction Using Sequential Pattern Mining and Entropy
Qingren Wang, Victor S. Sheng, and Chenyi Hu

Session 4: Dynamic Systems. Wednesday, August 9, 2017, 14:00 pm to 15:40 pm

- Learning Question Similarity with Recurrent Neural Networks
Borui Ye, Guangyu Feng, Anqi Cui, and Ming Li
- Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information
Yong Zhuang, Matthew Almeida, Melissa Morabito, and Wei Ding
- Granger Causality for Multivariate Time Series Classification
Dandan Yang, Huanhuan Chen, Yinlong Song, and Zhichen Gong
- Revisit Word Embeddings with Semantic Lexicons for Modeling Lexical Contrast
Jiawei Liu, Zhenyu Liu, and Huanhuan Chen

Workshop 1-1 Wednesday, August 9, 2017, 14:00 pm to 15:40 pm

- A Novel Approach Based on Neural Networks and Support Vector Machine for Stock
Fan Yajing and Gan Yujian
- Unsupervised Spectral Feature Selection with Local Structure Learning
Shichao Zhang, Yue Fang, Cong Lei, Yangding Li, Rongyao Hu, and Yonggang Li
- Unsupervised Feature Selection via Local Structure Learning and Self-Representation
Shichao Zhang, Cong Lei, Yue Fang, Yangding Li, Rongyao Hu, and Xiaoyi Hu
- Randomized Perturbation for Privacy-Preserving Social Network Data Publishing
Peng Liu, Li-e Wang, and Xianxian Li
- Robust Features Selection via Structure Learning and Multiple Subspace Learning
Yonghua Zhu, Xuejun Zhang, Rongyao Hu, and Guoqiu Wen

Workshop 1-2: Wednesday, August 9, 2017, 16:10 pm to 17:50 pm

- Low-Light Image Enhancement by Refining Illumination Map with Self-Guided Filtering
Zhuang Feng and Shijie Hao
- A Potential-Based Clustering Method by Fast Search and Find of Cluster Centers
Xin Liu, Yongjian Liu, and Qing Xie
- TLRec: Transfer Learning for Cross-Domain Recommendation
Leihui Chen, Jianbing Zheng, Ming Gao, Aoying Zhou, Wei Zeng and Hui Chen
- Causal Structure Learning Algorithm Based on Streaming Features
Xiaoxue Guo and Jing Yang

Workshop 2: Thursday, August 10, 2017, 10:20 am to 12:00 pm

- A Method Study of Online Publication Time Extraction for Chinese Web News
Liangliang Wang and Gongqing Wu
- Synonymous Entity Recognition Based on Feature Fusion
Desheng Cai, Jingjing He, Gongqing Wu, and Xuegang Hu

Workshop 3-1 Wednesday, August 9, 2017, 14:00 pm to 15:40 pm

- Research on the Emergency Events Monitoring of Minsheng Hotline Based on Text
Xue Bin and Tao Haijun
- Subject Extraction Method of Urban Complaint Data
Zhian Dong and Xueqiang Lv
- Emergency Events Classification Based on Minsheng Hotline Unbalanced Short-Text
Li-Yuan Geng, Wei Jin, and Han-Bing Qu
- Risk Assessment of Long Distance Oil and Gas Pipeline Based on Grey Clustering
Yusheng Hu, Kehui Liu, Dong Xu, Zhengang Zhai, and Haiyun Liu
- Big Data in Urban Construction Archives and Urban Management: Detected Underground Pipes Alignment with Urban Construction Records
BingZhang Chi, Hui Meng, KuiXiu Zhai, Guangyin Zhai, and Yan Liu

Workshop 3-2: Wednesday, August 9, 2017, 16:10 pm to 17:50 pm

- Research on Innovation of Management Concept of Urban Underground Pipelines Based on Big-Knowledge-Thinking in the Big Data Era
Hui Wan and Decheng Kong
- Mining Course Trajectories of Successful and Failure Students: A Case Study
Jin Soung Yoo, Yei-Sol Woo, and Sang Jun Park
- Mining Place Design Knowledge from Multi-source Data in an Informed Design Platform
Linlin You, Bige Tuncer, and Hexu Xing
- An Anomaly Detection Method for Medicare Fraud Detection
WeiJia Zhang and Xiaofeng He

Workshop 4: Thursday, August 10, 2017, 10:20 am to 12:00 pm

- Modeling Topic Evolution in Social Media Short Texts
Wei Li, Xiaohui Cui, Kevin Michael Amaral, Rajani Sadasivam, and Ping Chen
- Incorporating Entity Correlation Knowledge into Topic Modeling
Qilin Wang, Dandan Song, and Xiuquan Li
- EEG-Based Person Recognition Analysis and Criticism
Meriem Romaissa Boubakeur, Guoyin Wang, Chenglin Zhang, and Ke Liu
- Smoking Cessation Recruitment Analysis: A Case Study
Wei Li, Xiaohui Cui, Kevin Michael Amaral, Rajani Sadasivam, and Ping Chen
- Online Streaming Feature Selection Based on Conditional Information Entropy
Huaming Wang, Guoyin Wang, Xianhua Zeng, and Siyuan Peng

Workshop 5-1: Wednesday, August 9, 2017, 16:10 pm to 17:50 pm

- BMNR: Design and Implementation a Benchmark for Metrics of Network Robustness
Jianbing Zheng, Yanbin Li, Yanji Hou, Ming gao, and Aoying Zhou
- A Hybrid Abnormal Advertising Traffic Detection Method
Kun Wang, Guohai Xu, and Xiaofeng He
- Video Captioning with Semantic Information from the Knowledge Base
Dan Wang and Dandan Song
- BiUCB: A Contextual Bandit Algorithm for Cold-Start and Diversified Recommendation
Lu Wang, Chengyu Wang, Keqiang Wang, and Xiaofeng He

Workshop 5-2: Thursday, August 10, 2017, 10:20 am to 12:00 pm

- A Transfer Learning Based Boosting Model for Emotion Analysis
Ruolan Yong, Chengyu Wang, and Xiaofeng He
- Discovering Knowledge by Behavioral Analytics for Elderly Care
Bruce X.B. Yu and Keith C.C. Chan
- Image Retrieval Based on Hierarchical Locally Constrained Diffusion Process
Xianhua Zeng, Meng Hu, and Suwen Zhu

Panel 1: Thursday, August 10, 2017, 14:00 pm to 15:30 pm

- Graph data are of growing importance in many applications, including the semantic web (i.e., RDF), social network analysis, bioinformatics, software engineering, e-commerce, finance and trading, fraud detection, and recommendation systems, because they model complicated structures and relationships well. The size and complexity of these graphs raise significant data management and data analysis challenges. This has led to a number of different algorithms and approaches to graph processing as well as systems that are based on these algorithms. There are two different camps in graph processing, with different techniques and approaches: property graph processing, and RDF graph processing. In this panel, we will give a unified view of both approaches and then discuss the techniques developed for each one.

Panel 2: Thursday, August 10, 2017, 14:00 pm to 15:30 pm

- Knowledge Engineering with Big Data (BigKE) is a research frontier that leads big data analytics to big knowledge. The main objective of BigKE is to learn and fuse large-volume, low-quality, un-ordered fragmented knowledge from heterogeneous and autonomous sources for complex and evolving relationships, and provide personalized, demand-driven knowledge services. BigKE differs from traditional knowledge engineering which relies on domain experts, and fragmented knowledge originates from user-generated contents (UGC). In terms of knowledge acquisition and knowledge re-engineering, BigKE is expected to break through the limitations of traditional knowledge engineering on domain expertise, and has wide potential applications in pervasive medicine, online education and Web based commerce. Currently, BigKE is supported by the National Key Research and Development Program of China, under “Knowledge Engineering with Big Data: Fundamental Theory and Applications” (grant 2016YFB1000900).

LEVEL 2

