Introduction to the *IEEE Transactions on Big Data*

Qiang Yang, *Fellow, IEEE*

It is my great pleasure to present this inaugural issue of the *IEEE Transactions on Big Data* (*IEEE TBDATA*). Big Data is a new field that encompasses multiple disciplines and impacts a wide range of sectors of our society. Its rapid rise in recent years can be attributed to several technological advances. The increasing availability of sensors made data generation and collection easier and cheaper. Advances in telecommunications technologies and services facilitated the massive exchange of data among client devices, data centers and clouds. The fast reduction in data storage and processing costs gives rise to fast growth in increasing computational power. As a result, novel applications are widely found that span across diverse fields as never before.

Big Data can be characterized by its extraordinary characteristics along several dimensions. The first of the dimensions is the size of data. Data sets grow in sizes partly because they are being gathered by cheaper and easier-to operate information sensing mobile devices. This is referred to as Volume by industry leaders [1]. Other dimensions are equally important, including Big Data's Variety (the data types are many and heterogeneous), Velocity (the speed is fast in which the data is generated and processed to meet the demands) and Veracity (the quality of the data being captured can vary greatly). These complexities pose a major challenge as well as new opportunity for today's information technology communities.

The term Big Data goes well beyond the data itself; it is also often used to refer to a new methodology to approach our problems and solutions. As pointed out in [2], our scientific advances fall in different stages, or paradigms, as the human race moves forward. The first paradigm is known as the empirical stage, which happened when scientific discovery was mainly driven by recording empirical observations through tools such as telescopes. The second stage was when theories were introduced to summarize the observations and make predictions. Scientists such as Newton used mathematics and physical laws to build models to explain the empirical observations. The third paradigm came as a result of the arrival of digital computers, when large-scale simulations were used to mimic the dynamics of nature. With the arrival of the Big Data, we are at the beginning of the fourth paradigm of scientific discovery, when knowledge discovery is done through hypothesis testing driven by the availability of the massive digital data. In this fourth-paradigm way of scientific thinking, data becomes a first-class citizen, giving birth to the particular practice of knowledge discovery known as Data Science.

Thus, Big Data is situated at the cross roads of many disciplines, and this new *IEEE Transactions on Big Data* aspires to lead this technological revolution to the next level. The journal will serve as a forum for the Big Data community to exchange ideas and report its successes. In particular, the journal will cover the following broad areas:

Big Data Analytics

Big Data Analytics is aimed at making sense of data by applying efficient and scalable algorithms on Big Data for its analysis, learning, modelling, visualization and understanding. This includes the design of efficient and effective algorithms and systems to integrate the data and uncover the hidden values from data. It also includes methodologies and algorithms for automatic or mixed-initiative knowledge discovery and learning, data transformation and modelling, predictions and explanations of the data. Breakthroughs in this area include new algorithms, methodologies, systems and applications for knowledge discovery, understanding and applications based on the Big Data. New computing paradigms are expected in new areas such as human computation, crowd sourcing, sentiment analysis as well as data visualization technologies.

Big Data Infrastructure and Tools

Big Data Infrastructure deals with new computing architectures and models to enable efficient and scalable high performance, parallel and distributed computation to support all aspects of computation with Big Data. Examples include well-known industrial systems such as HDFS, Hadoop, SPARK and STORM, to name a few. Key innovations are expected in novel algorithms and systems for making increasingly efficient use of computing resources for Big Data computation. Of particular interest are systems that integrate traditionally disparate areas of research, such as data collection and transformation, networking, data management and learning and data analytics. Emerging issues are particularly important, such as the issue of how to limit the energy use of large-scale computation via innovative algorithms and system design.

• *The author is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong. E-mail: qyang@cse.ust.hk.*

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TBDATA.2015.2469391

Big Data Policy, Practice and Standards

Big impact implies big responsibility. Big Data is a broad area where technology meets social and policy issues. History has taught us that technology advancement must go hand in hand with policy and standards development. Practitioners know very well that Big Data research cannot be done without the datasets themselves, thus the issue of data sharing and data openness is critical. If different data sets cannot be aligned easily, then they cannot be fully integrated to result in the scale needed to produce impact. Thus, many governments, companies and organizations are starting to recognize that data sharing is a key for Big Data development. However, how to ensure that data is shared across departments, organizations and societal boundaries remains a complex issue. Closely related are technological issues as well as issues such as how to divide the benefits and responsibilities of data sharing.

An equally important issue is data quality, value and provenance. Data quality has direct impact on analytic results, and yet we do not fully understand how to quantify this impact. We also do not know well how to automatically or semi-automatically transform the data so that the quality is guaranteed. Recent works begin to recognize the importance of this issue. For example, works on identifying the trustworthiness of crowd-sourced data have attracted attention. Likewise, it is also an open issue of how to put a quantitative valuation on the worth of data for data exchange and sharing.

Big Data Privacy and Security

Policies on data availability, data sharing and standards are also closely related to the issue of user privacy and data security. On one hand, data analytics is aimed to uncovering values from data; thus the more data is included the better. Data analytics also requires sharing data and exchange data, which demands openness. On the other hand, revealing more data to the outside world runs the potential risk of jeopardising users' private information and cause damage to the reputation of the analytic services. How to allow analytic systems to freely access the needed data while protecting user privacy is a critical research and practice issue. A related and equally important issue is data security, which deals with the problem of ensuring the intended access control of the data while data is being collected, stored, managed, transported and analysed.

Data Science

The fourth paradigm of scientific knowledge discovery champions the philosophy that with sufficient data, science can be carried out by directly examining and analyzing the data. New laws of nature can be discovered from the data. In each scientific discipline, the data and knowledge will be of specific, domain dependent nature. For example, genetic data in biological science may be rich in the number of features, but small in the number of data samples. In contrast, the data from particle physics may be the opposite. It is thus an open issue how to ensure that domain specific scientific knowledge is seamlessly integrated with general knowledge discovery and data analytic functions. Another important issue is how to explain the discovered knowledge back to the scientists. In many data science areas, black box solutions may not be desirable.

Big Data Theory

Big Data provides renewed motivations for developing new theories that help evaluate and compare the existing approaches and quantify computational boundaries. Many of the fundamental theories of computation are based on the assumption that the data follows ideal distributions, are independently sampled and can be manipulated in a given computing environment. Big Data challenges these assumptions by considering data that extend beyond the traditional limitations. Statistics, computational learning theory and game theory are but a few examples that can help establish a new theory to understand and exploit big data for knowledge discovery.

Big Data Applications

There is no doubt that Big Data can only thrive if its applications grow. Theory, algorithms and systems that enable novel applications of Big Data field are all the necessary elements for the new field. Indeed, novel applications are abound today, ranging from smart cities to Big Data in healthcare and from digital government services to online commerce industry. On one hand, Big Data tools and systems enable these applications to scale to the real world; on the other hand, successes in these application fields also provide new impetus for the tools to improve. Especially important to researchers and practitioners alike are case studies that tell why Big Data applications work or not work, with experience and lessons for others to draw from.

Data Articles: A New Feature

In Big Data research, a critical issue is the availability of realistic, large-scale and complex data. While data are abundant in many government agencies, industrial companies and practicing institutions, they exist mostly in islands and in isolation. There exists a chasm between the data generators and data consumers. The public cannot easily access much of these data. A consequence is that while in scientific disciplines, researchers stress repeatable experiments as a necessary test for new discoveries, in Big Data areas, repeating experiments on the same data by independent research teams is seldom practiced. To make a difference for the better, in *IEEE TBDATA* we introduce a new type of submission known as Data Articles. This type

of submission will require that authors upload original large-scale data sets together with the meta data and a companion article to describe the meaning of attributes and intended usage of data in verifying systems and algorithms. To make the explanation accessible, the journal will accept a three-page article on the data together with preliminary test results to show the utility of the data. These articles will be citable just like any other articles in the journal. By creating this new category of papers, we hope to make an understanding in the Big Data community that contributing data is equally important as other contributions to the field, as new data can help inspire new algorithms, present new challenges, and serve as new baselines.

Big Data is necessarily a multi-disciplinary subject. A Big Data system touches on many aspects of science and engineering, including computer science, engineering, business management, social science, and policies, to name a few. More than one group of researchers is needed to aim for the same objective since we need multiple angles to look at the same problem. In Big Data, multiple views and methodologies will be the most beneficial. Therefore, *IEEE TBDATA* welcomes articles that report on the integration of different disciplines, technologies and systems to accomplish the mission of harnessing the Big Data.

Submission Requirements

The style of this new IEEE journal is similar to many other IEEE Transactions, with submissions received through the IEEE Manuscript Central system (http://www.computer.org/web/tbd). The journal will publish four issues per year, with five to 10 articles in each issue. Five types of papers are considered: Regular papers will be given 14 double column or 30 single column pages, short papers are given eight double column or 15 single column pages, comment papers are given two double column or four single column pages, and survey papers are given 20 double column or 40 single column pages. In addition, the new Data Submission Paper Type includes a publicly downloadable data source and a three double-column page paper (or six single column pages). Extensions of conference publications are welcome, but should include a significant amount of new contribution and clearly state the differences from the conference versions. It is expected that more than 30% of the difference between the published version and the submission should be in the new contribution. This difference should be fully explained in the cover letter.

The new IEEE journal received strong support from many organizations and individuals. The journal is financially sponsored by the IEEE Computer Society, IEEE Communications Society, IEEE Computational Intelligence Society, IEEE Sensors Council, IEEE Consumer Electronics Society, IEEE Signal Processing Society, IEEE Systems, Man and Cybernetics Society, The IEEE Systems Council, The Vehicular Technology Society. In addition, the journal is technically cosponsored by The IEEE Control Systems Society, IEEE Signal Processing Society, IEEE Power and Energy Society and IEEE Biometrics Council. Representatives from these societies and councils formed a steering committee that provided valuable advices throughout the journals birth. I would in particular like to thank the journal's Steering Committee Chair, Professor Steven Crago, and IEEE staff members for consistently guiding the journal through its creation process. The journal's Advisory Committee, consisting of industry and academia heavyweights, has lent a helping hand whenever needed. An editorial board of more than 50 associate editors of the journal will help ensure that authors get fast and informative feedback from review process, lifting the journal to the highest level. With *IEEE TBDATA*, we now have a common platform to exchange ideas and showcase our innovations for the important area of Big Data. We welcome authors and special issue guest editors to contribute their newest results in this journal. We look forward to an exciting journey together!

Qiang Yang
Editor in Chief

References

[1] (2015) [Online]. Available: http://www.gartner.com/it-glossary/big-data
[2] K. M. Tolle, D. S. W. Tansley, and A. J. G. Hey, "AJG: The fourth paradigm: Dataintensive scientific discovery," *Proc. IEEE*, vol. 99, no. 8, pp. 1334–1337, Aug. 2011.

Charu Aggarwal received the BS degree from IIT Kanpur in 1993 and the PhD degree from Massachusetts Institute of Technology in 1996. He is a distinguished research staff member (DRSM) at the IBM T.J. Watson Research Center in Yorktown Heights, New York. He has worked extensively in the field of data mining, with particular interests in data streams, privacy, uncertain data and social network analysis. He has published 14 (three authored and 11 edited) books, over 250 papers in refereed venues, and has applied for or been granted over 80 patents. His h-index is 70. Because of the commercial value of the above-mentioned patents, he has received several invention achievement awards and has thrice been designated a Master Inventor at IBM. He received an IBM Corporate Award (2003) for his work on bioterrorist threat detection in data streams, a recipient of the IBM Outstanding Innovation Award (2008) for his scientific contributions to privacy technology, and received an IBM Research Division Award (2008) for his scientific contributions to data stream research. He has received two Best Paper Awards and an EDBT Test-of-Time Award (2014). He has served as the general or a program co-chair of the IEEE Big Data Conference (2014), the ICDM Conference (2015), the ACM CIKM Conference (2015), and the KDD Conference (2016). He also co-chaired the data mining track at the WWW Conference 2009. He served as an associate editor of the *IEEE Transactions on Knowledge and Data Engineering* from 2004 to 2008. He is an associate editor of the *ACM Transactions on Knowledge Discovery and Data Mining* , an action editor of the *Data Mining and Knowledge Discovery Journal* , an associate editor of *the IEEE Transactions on Big Data*, and an associate editor of the *Knowledge and Information Systems Journal*. He is an editor-in-chief of the *ACM SIGKDD Explorations*. He is a fellow of the IEEE (2010), ACM (2013), and the SIAM (2015) for "contributions to knowledge discovery and data mining algorithms."

Srinivas Aluru received the BTech degree from the Indian Institute of Technology Madras in 1992, and the MS, and PhD degrees in 1991 and 1994, respectively, from Iowa State University, all in computer science. He is a professor in the School of Computational Science and Engineering, College of Computing at Georgia Institute of Technology. He co-leads the Georgia Tech Strategic Initiative in Data Engineering and Science, and directs the Intel Parallel Computing Center in Big Data in Biosciences and Public Health. Earlier, he held faculty positions at Iowa State University, Indian Institute of Technology Bombay, New Mexico State University, and Syracuse University. He conducts research in high-performance computing, bioinformatics and systems biology, combinatorial scientific computing, and applied algorithms. He pioneered the development of parallel methods in computational biology, and contributed to the assembly and analysis of complex plant genomes. His contributions in scientific computing lie in parallel Fast Multipole Method, domain decomposition methods, spatial data structures, and applications in computational electromagnetics and materials informatics. He has been conducting federally funded big data research since the inception of the program, and has participated in the OSTP and NITRD led white house events in big data. He received the US National Science Foundation (NSF) Career Award (1997), IBM Faculty Award (2002), and Swarnajayanti fellowship from the Government of India (2007). He is currently serving as the chair in the ACM Special Interest Group on Bioinformatics, Computational Biology, and Biomedical Informatics (SIGBio). He played a lead organizational role in several conferences in the fields of parallel processing (IPDPS, SC, ICPP, HiPC) and bioinformatics (BCB, ICCABS, WABI). In addition to the *IEEE Transaction on Big Data*, he serves on the editorial boards of the *IEEE Transactions on Parallel and Distributed Systems*, the *Journal of Parallel and Distributed Computing*, and the *International Journal of Data Mining and Bioinformatics*. He is a fellow of the American Association for the Advancement of Science (AAAS) and the IEEE.

Chaitan Baru received the BTech degree in electronics engineering from the Indian Institute of Technology, Madras, and the ME and PhD degrees in electrical engineering from the University of Florida, Gainesville. He is a distinguished scientist and an associate director of Data Initiatives at the San Diego Supercomputer Center (SDSC), UC San Diego, where he works on applied and applications-oriented research problems related to data management and data analytics. He leads the Advanced Cyberinfrastructure Development (ACID) Group and is the Director of the Center for Large-scale Data Systems research (CLDS), at SDSC. He is currently on assignment as a senior advisor for Data Science in the Computer and Information Sciences Directorate at the US National Science Foundation, Arlington, Virginia. In 2012, he launched the workshop series on Big Data Benchmarking, which has engendered new activity in industry standard benchmarks for Big Data, including creation of the TPCx-HS standard; formation of the TPC-BigBench subcommittee, and creation the SPEC Research Group on Big Data, which he co-chairs. He has participated in a number of "data cyberinfrastructure" initiatives at SDSC, including as Principal Investigator (PI) of the OpenTopography project Cyberinfrastructure Lead, Tropical Ecology, Assessment and Monitoring network (TEAM); co-investigator of the Cyberinfrastructure for Comparative Effectiveness Research project (CYCORE); a member of the founding Senior Management Team of the National Ecological Observatory Network (NEON) and Co-PI of the NEON Cyberinfrastructure Testbed; Co-PI of the CUAHSI Hydrologic Information Systems (CUAHSI-HIS); the director, NEES Cyberinfrastructure Center (NEESit); a PI/project director, Geosciences Network (GEON). Prior to joining SDSC in 1996, he was at IBM, where he led one of the development teams for DB2 Parallel Edition Version 1 (released Dec 1995); and at the University of Michigan, where he served on the faculty of the EECS Department.

Albert Bifet is a senior researcher at Huawei Noah's Ark Lab. He is the author of a book on adaptive stream mining and pattern learning and mining from evolving data streams. He is one of the leaders of MOA and Apache SAMOA software environments for implementing algorithms and running experiments for online learning from evolving data streams. He is serving as a co-chair of the Industrial track of ECML PKDD 2015, and as a co-chair of BigMine (2015, 2014, 2013, 2012), and ACM SAC Data Streams Track (2016, 2015, 2014, 2013, 2012).

Francesco Bonchi received the PhD degree in computer science from the University of Pisa in December 2003. He is the director of Research at Yahoo Labs in Barcelona, Spain, where he is leading the Web Mining Research group. His recent research interests include mining query-logs, social networks, and social media, as well as the privacy issues related to mining these kinds of sensible data. In the past he has been interested in data mining query languages, constrained pattern mining, mining spatiotemporal and mobility data, and privacy preserving data mining. He is a member of the ECML PKDD Steering Committee, an associate editor of the newly created *IEEE Transactions on Big Data* (*TBDATA*), of the *IEEE Transactions on Knowledge and Data Engineering* (*TKDE*), the *ACM Transactions on Intelligent Systems and Technology* (*TIST*), *Knowledge and Information Systems* (*KAIS*), and a member of the Editorial Board of *Data Mining and Knowledge Discovery* (*DMKD*). He has been a program co-chair of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010). He has also served as a program co-chair of the first and second ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD (PinKDD 2007 and 2008), the first IEEE International Workshop on Privacy Aspects of Data Mining (PADM 2006), and the fourth International Workshop on Knowledge Discovery in Inductive Databases (KDID 2005). He is a co-editor of the book *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques* published by Chapman & Hall/CRC Press.

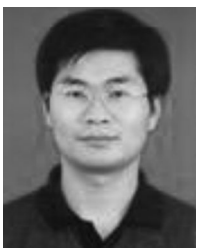
Suren Byna is a computer scientist in the Scientific Data Management Group at Lawrence Berkeley National Laboratory (LBNL). His research interests are in scalable scientific data management. More specifically, he works in the areas of parallel I/O, data management systems for managing scientific data, and heterogeneous computing. He is also interested in energy efficient parallel computing. He has published more than 65 papers. He received the Best Poster Award at Supercomputing '03 (SC03) and the Best Paper award at Cray User Group (CUG) meeting 2013. Papers he co-authored were nominated for best papers at SC08 and HPDC 2015.

Wo Chang is a digital data advisor for the NIST Information Technology Laboratory (ITL). His responsibilities include, but are not limited to, promoting a vital and growing Big Data community at NIST with external stakeholders in commercial, academic, and government sectors. He is currently the convenor for the ISO/IEC JTC 1 Big Data Study Group, co-chairs the NIST Big Data Public Working Group, and chairs the ISO/IEC JTC/1 SC29 WG11 (MPEG) Multimedia Preservation AHG. Prior to joining ITL Office, he was a manager of the Digital Media Group in ITL and his duties included overseeing several key projects including digital data, long-term preservation and management of EHRs, motion image quality, and multimedia standards. In the past, he was the deputy chair for the US National Body for MPEG (INCITS L3.1) and chaired several other key projects for MPEG, including MPQF, MAF, MPEG-7 Profiles and Levels, and co-chaired the JPEG Search project. He was one of the original members of the W3C's SMIL WG and developed one of the SMIL reference software. Furthermore, he also participated in the HL7 and ISO/IEC TC215 for health informatics and IETF for the protocols development of SIP, RTP/RTPC, RTSP, and RSVP. His research interests include digital data preservation, cloud computing, big data analytics, content metadata description, digital file formats, multimedia synchronization, and Internet protocols.

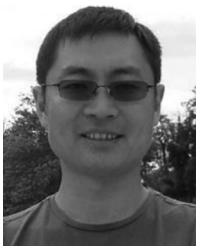
Sanjay Chawla received the PhD degree from the University of Tennessee (USA) in 1995. He is a professor in the School of Information Technologies, University of Sydney. He is currently on extended leave as a principal scientist at Qatar Computing Research Institute (QCRI). His research interests span data mining, machine learning with a specialization in outlier detection, class imbalanced classification, adversarial learning and spatiotemporal data. From 2008 to 2011, he served as the head (department chair) of the School of Information Technologies. Additionally, he was an academic visitor at Yahoo! Research in 2012. His work has been recognized by several best paper awards, including in leading conferences such as SIAM International Conference in Data Mining (2006) and IEEE International Conference in Data Mining (2010). He is a co-author on a popular text in spatial database management systems: *Spatial Databases: A Tour*, which has been translated into Chinese and Russian. He serves on the editorial boards of *IEEE TKDE* and *DMKD Journal*, and served previously as a PC chair of PAKDD (2012).

Jinjun Chen received the PhD degree in information technology from Swinburne University of Technology, Australia. He is an associate professor from Faculty of Engineering and IT, University of Technology Sydney (UTS), Australia. He is the director of Lab for Cloud Computing and Data Intensive Systems at UTS. His research interests include big data, data science, data intensive systems, scalability, cloud computing, software engineering and services, workflow management, privacy and security, and related various research topics. His research results have been published in more than 130 papers in international journals and conferences, including *IEEE Transactions on Service Computing*, *ACM Transactions on Autonomous and Adaptive Systems*, *ACM Transactions on Software Engineering and Methodology* (*TOSEM*), *IEEE Transactions on Software Engineering* (*TSE*), *IEEE Transactions on Parallel and Distributed Systems* (*TPDS*), *IEEE Transactions on Cloud Computing* (*TCC*), and *IEEE Transactions on Computers* (*TC*). He received UTS Vice-Chancellor's Awards for Research Excellence Highly Commended (2014), UTS Vice-Chancellor's Awards for Research Excellence Finalist (2013), Swinburne Vice-Chancellor's Research Award for early career researchers (2008), IEEE Computer Society Outstanding Leadership Award (2008-2009) and (2010-2011), IEEE Computer Society Service Award (2007), Swinburne Faculty of ICT Research Thesis Excellence Award (2007). He is an associate editor for *ACM Computing Surveys*, *IEEE Transactions on Big Data*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Parallel and Distributed Systems*, and *IEEE Transactions on Cloud Computing*. He is the chair of IEEE Computer Society's Technical Committee on Scalable Computing (TCSC), a vice chair of Steering Committee of Australasian Symposium on Parallel and Distributed Computing, the founder and a coordinator of IEEE TCSC Technical Area on Big Data and MapReduce, Technical Area on Workflow Management in Scalable Computing Environments, the founder and a Steering Committee co-chair of IEEE International Conference on Big Data and Cloud Computing, and IEEE International Conference on Big Data Science and Engineering.

Peter Chen received the BS degree in electrical engineering in 1968 from the National Taiwan University, and the PhD degree in computer science/applied mathematics from the Harvard University in 1973. In 1970, he worked one summer at IBM. After graduating from Harvard, he spent one year at Honeywell and a summer at Digital Equipment Corporation. From 1974 to 1978, he was an assistant professor at the MIT Sloan School of Management. From 1978 to 1983, he was an associate professor at the University of California, Los Angeles (UCLA Management School). From 1983 to 2011, he held the position of M.J. Foster Distinguished Chair Professor of computer science at Louisiana State University and, for several years, Adjunct Professor in its Business School and Medical School (Shreveport).[1] During this time period, he was a visiting professor once at Harvard in 1989-1990 and three times at Massachusetts Institute of Technology (EECS Department in 1986-1987, Sloan School in 1990-1991, and Division of Engineering Systems in 2006-2007). Since 2008, he has been a Honorary chair professor in the Institute of Service Science at National Tsing Hua University. Currently, he is a distinguished career scientist and a faculty member at Carnegie Mellon University.



Xueqi Cheng is the director of the Division of the network science and technology in Institute of Computing Technology (ICT), CAS. As a research area manager, he leads a team of passionate researchers to advance the state-of-the-art in web search and mining, network science, information security. He has served on the organizing and program committees of some international and domestic conferences, including WWW, NCIRCS, IEEE ISI, etc. He has published more than 60 papers in fields such as web search, information retrieval, information security, P2P, and network science.



Shuguang Cui received the PhD degree in electrical engineering from Stanford University, CA, in 2005. He is a professor in electrical and computer engineering at the Texas A&M University, College Station, TX. His current research interests focus on data oriented large-scale information analysis and system design, including large-scale distributed estimation and detection, information theoretical approaches for large data set analysis, complex cyber-physical system design, and cognitive network optimization. He was selected as the Thomson Reuters Highly Cited researcher and listed in the World's Most Influential Scientific Minds by ScienceWatch in 2014. He received the IEEE Signal Processing Society 2012 Best Paper Award. He has served as the TPC co-chairs for many IEEE conferences. He has also been serving as the area editor for *IEEE Signal Processing Magazine*, and associate editor for *IEEE Transactions on Big Data*, *IEEE Transactions on Signal Processing*, IEEE JSAC Series on Green Communications and Networking, and *IEEE Transactions on Wireless Communications*. He was an elected member for IEEE Signal Processing Society SPCOM Technical Committee (2009∼2014) and the elected vice chair for IEEE ComSoc Wireless Technical Committee (2015∼2016). He is a member of the Steering Committee for *IEEE Transactions on Big Data* and a member of the IEEE ComSoc Emerging Technology Committee. He is a fellow of the IEEE and an IEEE ComSoc distinguished lecturer (2015∼2016).

Wenfei Fan received the BS and MS degrees from Peking University, and the PhD degree from the University of Pennsylvania. He is the chair in Web Data Management in the School of Informatics, University of Edinburgh, United Kingdom. He is a fellow of the Royal Society of Edinburgh, United Kingdom, a fellow of the ACM, a national professor of the 1000-Talent Program and a Yangtze River Scholar, China. He received the Alberto O. Mendelzon Test-of-Time Award of ACM PODS 2010 and 2015, an ERC Advanced Fellowship in 2015, the Roger Needham Award in 2008, the Outstanding Overseas Young Scholar Award in 2003, the Career Award in 2001 (USA), and several Best Paper Awards (VLDB 2010, ICDE 2007, Computer Networks 2002). His current research interests include database theory and systems, in particular big data, data quality, data integration, distributed query processing, query languages, recommender systems, social networks, and web services.

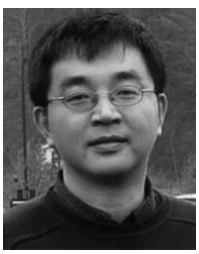
Pascal Frossard (S'96-M'01-SM'04) received the MS and PhD degrees, both in electrical engineering, from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1997 and 2000, respectively. Between 2001 and 2003, he was a member of the research staff at the IBM T.J. Watson Research Center, Yorktown Heights, NY, where he worked on media coding and streaming technologies. Since 2003, he has been a faculty at EPFL, where he heads the Signal Processing Laboratory (LTS4). His research interests include graph signal processing, image representation and coding, visual information analysis, and distributed signal processing and communications. He has been the general chair of IEEE ICME 2002 and Packet Video 2007. He has been the technical program chair of IEEE ICIP 2014 and EUSIPCO 2008, and a member of the organizing or technical program committees of numerous conferences. He has been an associate editor of the *IEEE Transactions on Signal Processing* (2015-), *IEEE Transactions on Big Data* (2015-), *IEEE Transactions on Image Processing* (2010-2013), the *IEEE Transactions on Multimedia* (2004-2012), and the *IEEE Transactions on Circuits and Systems for Video Technology* (2006-2011). He is the chair of the IEEE Image, Video and Multidimensional Signal Processing Technical Committee (2014-2015), and an elected member of the IEEE Visual Signal Processing and Communications Technical Committee (2006-) and of the IEEE Multimedia Systems and Applications Technical Committee (2005-). He has served as a Steering Committee chair (2012-2014) and Vice-Chair (2004-2006) of the IEEE Multimedia Communications Technical Committee and as a member of the IEEE Multimedia Signal Processing Technical Committee (2004-2007). He received the Swiss National Science foundation (NSF) Professorship Award in 2003, the IBM Faculty Award in 2005, the IBM Exploratory Stream Analytics Innovation Award in 2008 and the IEEE Transactions on Multimedia Best Paper Award in 2011.

Joydeep Ghosh received the BTech degree in 1983 and the PhD degree in 1988 from The University of Southern California. He is currently the Schlumberger Centennial chair professor of electrical and computer engineering at the University of Texas, Austin. He joined the UT-Austin faculty in 1988. He is the founder-director in Intelligent Data Exploration and Analysis Lab (IDEAL). His research interests lie primarily in data mining and web mining, predictive modeling / predictive analytics, machine learning approaches such as adaptive multi-learner systems, and their applications to a wide variety of complex real-world problems such as healthcare. He has published more than 400 refereed papers and 50 book chapters, and coedited over 20 books. He has received 14 Best Paper Awards over the years, including the 2005 Best Research Paper Award across UT and the 1992 Darlington Award. He also received the 2015 Technical Achievement Award from IEEE CS for development of multi-learner systems. Dr. Ghosh has been a plenary/keynote speaker on several occasions such as ICDM'13,(Health Informatics workshops at) KDD14, ICML13 and ICHI13; MICAI'12, KDIR'10 and ISIT'08, and has widely lectured on intelligent analysis of large-scale data. He served as the Conference co-chair or Program co-chair for several top data mining oriented conferences, including SDM'13, SDM'12, KDD 2011, CIDM'07, ICPR'08 (Pattern Recognition Track) and SDM'06. He was the Conf. co-chair for Artificial Neural Networks in Engineering (ANNIE) 1993 to 1996 and 1999 to 2003. He has also co-organized workshops on health informatics, high dimensional clustering, web analytics, web mining and parallel/ distributed knowledge discovery. He has served as a cofounder, consultant or advisor to several successful startups in addition to consulting for large corporations. Currently, he is on the advisory board of Cognitive Scale (Healthcare thrust) and Accordion Health. He is a fellow of the IEEE.

Marta González does research focusing on urban computing, particularly in the intersections of people with the built environment as well as with their social networks. A recent leader in this emergent field, She has introduced new tools into transportation research during her time at MIT. Her current research explores human mobility patterns using mobile phone communication, propagation of mobile phone viruses, and urban transportation models. In the same manner, She integrates methods of complex systems with a physics approach, combining computation, geographic information systems, and network theory to model human interaction. Her ultimate goal is to design urban mobility solutions, which will pave the way for sustainable development of smart cities. In the past few years, she has twice received the Best Paper Award in the Association for Computing Machinery's Special Interest Group on Knowledge Discovery and Data Mining International Workshop on Urban Computing. Marta holds a PhD from the University of Stuttgart in Computational Physics of Complex Systems.

Ananth Grama received the BEngg degree in computer science from the Indian Institute of Technology, Roorkee in 1989, the MS degree in computer engineering from the Wayne State University in 1990, and the PhD degree in computer science from the University of Minnesota in 1996. He is the director in the Computational Science and Engineering program and a professor of computer science at Purdue University. He also serves as the associate director in the Center for Science of Information, a Science and Technology Center of the National Science Foundation. He has been at Purdue since 1996. His research interests lie in parallel and distributed systems, numerical methods, large-scale data analysis, and their applications. Ananth is a recipient of the National Science Foundation CAREER award (1998), University Faculty Scholar Award (2002-2007), and is a fellow of the American Association for the Advancement of Sciences (2013).

Yike Guo received the first-class honors degree in computing science from Tsinghua University, China, in 1985 and the PhD degree in computational logic from Imperial College in 1993 under the supervision of Professor John Darlington. He is a professor of computing science in the Department of Computing at Imperial College London. He is the founding director in the Data Science Institute at Imperial College, as well as leading the Discovery Science Group in the department. He also holds the position of CTO of the tranSMART Foundation, a global open source community using and developing data sharing and analytics technology for translational medicine. He founded InforSense, a software company for life science and health care data analysis, and served as a CEO for several years before the company's merger with IDBS, a global advanced R&D software provider, in 2009. He has been working on technology and platforms for scientific data analysis since the mid-1990s, where his research focuses on knowledge discovery, data mining and large-scale data management. He has contributed to numerous major research projects including: the United Kingdom EPSRC platform project, Discovery Net; the Wellcome Trust-funded Biological Atlas of Insulin Resistance (BAIR); and the European Commission U-BIOPRED project. He is currently the principal investigator in the European Innovative Medicines Initiative (IMI) eTRIKS project, a €23M project that is building a cloud-based informatics platform, in which tranSMART is a core component for clinico-genomic medical research, and co-Investigator of Digital City Exchange, a £5.9M research program exploring ways to digitally link utilities and services within smart cities. He has published more than 200 articles, papers, and reports. Projects he has contributed to have been internationally recognized, including winning the "Most Innovative Data Intensive Application Award" at the Supercomputing 2002 conference for Discovery Net, and the Bio-IT World "Best Practices Award" for U-BIOPRED in 2014. He is a senior member of the IEEE and is a fellow of the British Computer Society.



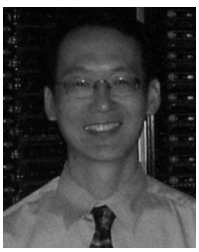
Rong Jin received the PhD degree in computer science from Carnegie Mellon University in 2003 and the US National Science Foundation (NSF) Career Award in 2006. He is a vice president at Alibaba group, and a professor in the Department of Computer Science and Engineering at Michigan State University. His research is focused on statistical machine learning and its application to big data analysis. He has published more than 200 technical papers, most in prestigious conferences (NIPS, ICML, KDD, SIGIR, CVPR, ICCV) and journals (*TPAMI, JMLR, JML*, and *TKDD*). He has served as an area chair for NIPS 2013 and SIGIR 2008, and an associate editor for *TPAMI* and *ACM KDD*.

Saurav Karmakar received the bachelor's of technology degree in electronics & communication engineering from the Kalyani Government Engineering College, University of Kalyani, India, the master's degree in mathematics and statistics from Georgia State University with concentration in bioinformatics. He received another master's degree and the doctorate (December 2011) from the Department of Computer Science, Georgia State University. He accomplished my doctoral dissertation in the field of text document mining and visualization under his advisor, Dr. Ying Zhu. During his last doctoral years, he was a Brain and Behavior fellow and was selected for their spotlight program in March 2011. He is a chief research engineer in a global electronic giant, where he is trying to accomplish the next generation of intelligence implementation. Until last year, he served as a co-principal investigator on a collaborative research project in the field of document co-reference resolution techniques, funded by the National Science Foundation. Alongside, he was a senior architect in the big data oriented data science and analytics practice for a premier company in the east coast in US. Alongside, he is also serving as the director in Data Sciences Engineering division of a medical health monitoring and prognosis-based start-up company based in Atlanta, GA. He was also involved as a senior research engineer at a premier research company in advanced machine learning and big data oriented research and development from early 2012 to the very end of 2013. His research interests are diversified on text mining, natural language processing, information retrieval and visualization, machine intelligence, entity/pattern recognition, recommendation techniques, stochastic implementations and ideas, Bayesian Statistics, etc. He is also involved in other collaborative research and was recently appointed as industrial mentor for the REU program started by the Department of Computer Science, Georgia State University. He has served as chair of the ACM chapter, GSU for two years and founded multiple organizations such as IEEE Student branch, CS Club, Cricket Club, Squash Club, where he served as the founding President and in 2011 he served as the leader in charge of six such organizations. Carrying such work with desperate enthusiasm, pouring new ideas and changing the shape of the structure brought our ACM chapter Chapter Excellence Award for Outstanding Recruitment Program in 2009–2010, which is one of the five categories in which ACM recognizes one chapter from all over the world. Due to such effort he also received the ultimate recognition from GSU: Royal Flame Student Leader of the year 2011, along with the outstanding service award from the Department of Computer Science. He is involved in ASHA Chapter Atlanta, a nonprofit organization to address socio-economic challenges in India through providing education, and started a chapter in GSU of the same as well.

George Karypis is currently pursuing research interests that span the areas of data mining, bioinformatics, parallel processing, CAD, and scientific computing. Over the years, he has developed algorithms to solve a variety of problems including dynamic load balancing of unstructured parallel computations, graph and circuit partitioning, protein remote homology prediction and fold recognition, protein structure prediction, recommender systems, data clustering, document classification and clustering, frequent pattern discovery in diverse datasets (transactions, sequences, graphs), parallel Cholesky factorization, and parallel preconditioners. His research has resulted in the development of software libraries for serial and parallel graph partitioning (METIS and ParMETIS), hypergraph partitioning (hMETIS), for parallel Cholesky factorization (PSPASES), for collaborative filtering-based recommendation algorithms (SUGGEST), clustering high-dimensional datasets (CLUTO), and finding frequent patterns in diverse datasets (PAFI). In addition, he has developed two web-based servers for clustering gene expression data (gCLUTO) and for predicting the secondary structure of proteins (YASSPP).

Vipin Kumar received the BE degree in electronics & communication engineering from Indian Institute of Technology Roorkee (formerly, University of Roorkee), India, in 1977, the ME degree in electronics engineering from Philips International Institute, Eindhoven, Netherlands, in 1979, and the PhD degree in computer science from University of Maryland, College Park, in 1982. He is a regents professor at the University of Minnesota, where he holds the William Norris Endowed chair in the Department of Computer Science and Engineering. His current research interests include data mining, high-performance computing, and their applications in climate/ecosystems and biomedical domains. He is the lead PI of a five-year, $10 Million project, "Understanding Climate Change - A Data Driven Approach," funded by the US National Science Foundation (NSF)'s Expeditions in Computing program that is aimed at pushing the boundaries of computer science research. He also served as the head in the Computer Science and Engineering Department from 2005 to 2015 and the director in Army High Performance Computing Research Center (AHPCRC) from 1998 to 2005. His research has resulted in the development of the concept of isoefficiency metric for evaluating the scalability of parallel algorithms, as well as highly efficient parallel algorithms and software for sparse matrix factorization (PSPASES) and graph partitioning (METIS, ParMetis, hMetis). He has authored more than 300 research articles, and has coedited or coauthored 11 books including widely used text books *Introduction to Parallel Computing* and *Introduction to Data Mining*, both published by Addison Wesley. He has served as a chair/co-chair for many international conferences and workshops in the area of data mining and parallel computing, including 2015 IEEE International Conference on Big Data, IEEE International Conference on Data Mining (2002), and International Parallel and Distributed Processing Symposium (2001). He co-founded SIAM International Conference on Data Mining and served as a founding co-editor-in-chief of *Journal of Statistical Analysis and Data Mining* (an official journal of the *American Statistical Association*). Currently, he serves on the steering committees of the SIAM International Conference on Data Mining and the IEEE International Conference on Data Mining, and is series editor for the Data Mining and Knowledge Discovery Book Series published by CRC Press/Chapman Hall. He received the Distinguished Alumnus Award from the Indian Institute of Technology (IIT) Roorkee (2013), the Distinguished Alumnus Award from the Computer Science Department, University of Maryland College Park (2009), and IEEE Computer Society's Technical Achievement Award (2005). His foundational research in data mining and its applications to scientific data was honored by the ACM SIGKDD 2012 Innovation Award, which is the highest award for technical excellence in the field of Knowledge Discovery and Data Mining (KDD). He is a fellow of the ACM, IEEE, and AAAS.

Adrien Lebre received the PhD degree from Grenoble Institute of Technologies in September 2006 and the MS degree from the University of Grenoble in 2002. He is a full time researcher at INRIA (on leave from an associate professor position at the Ecole des Mines de Nantes). His research interests are distributed/Internet computing. He has taken part in several EU projects dealing with Storage and BigData concerns such as the on-going BigStorage Marie Sklodowska Curie European Training Network. Since 2011, he has been a member of the architect and executive board of the Grid'5000 testbed. He has been involved in several program committees of conferences and workshops dealing with storage challenges (such as CCGRID 2015/2013, SC 2013, . . .).



Tsengdar Lee received the MS degree in civil engineering in 1988 and the PhD degree in atmospheric science in 1992 both from Colorado State University. He manages the High-End Computing Program from NASA Headquarters. He is responsible for maintaining the high-end computing capability to support the agency's aeronautics research, human exploration, scientific discovery, and space operations missions. He is also the manager of the NASA Weather Data Analysis Program, focusing on the transition of research results into the operational forecast centers and the acceleration of operational use of research data. Two major activities include the multi-agency Joint Center for Satellite Data Assimilation and the Short-term Prediction Research and Transition Center. In 2011, he served as an acting chief technology officer (CTO) for Information Technology (IT) in the NASA Office of the chief information officer. In this capacity, he funded agency-wide IT research and advanced prototyping and created NASA's IT Labs. He also chaired the CTO-IT Council. He joined NASA in 2001 as the High-End Computing Program manager for the Earth Science Enterprise. He was responsible for the Earth science computational modeling needs, primarily focusing on weather and climate modeling. Between 2002 and 2006, he also managed the Earth Science Global Modeling Program. He funded research efforts to study the global climate change, weather forecasting, and hurricane prediction problems. Prior to 2001, he held positions as a senior technical advisor with Northrop Grumman Information Technology and Senior Staff engineer with Litton PRC. He worked on the Advanced Weather Information Processing System (AWIPS) project for the National Weather Service. He was responsible for the rapid development, integration, and commercialization of the AWIPS client-server system. He was also a principal engineer on the effort to develop the AWIPS network monitoring and control system. He was a research scientist and worked on the dispersion problem of biochemical agents during his short tenure with the Science Applications International Corporation between 1994 and 1996. Trained as a short-term weather modeler, his work focused on the integration of weather and ancillary geographical information data into weather models to produce reliable forecasts. His research pioneered the modeling of land surface hydrology's impact on weather forecasting.

Ching-Yung Lin received the PhD degree from Columbia University in electrical engineering. He has been with IBM Research since 2000 after receiving the PhD degree. He is an IBM distinguished researcher and chief scientist, graph computing. He is also an adjunct professor in Columbia University since 2010 and in New York University (NYU) since 2014. He was an affiliate assistant/associate professor in the University of Washington from 2003 to 2009, and an adjunct associate professor in Columbia University from 2005 to 2007. His interest is mainly on *fundamental research* of large-scale multimodality signal understanding, network graph computing, and computational social & cognitive sciences, and *applied research* on security, commerce, and collaboration. Since 2011, he has been leading a team of more than 40 PhD researchers in worldwide IBM Research Labs (Watson, Almaden, Cambridge, Austin, India, China, Brazil, Australia, etc.) and more than 20 professors and researchers in nine universities (Northeastern, Northwestern, Columbia, Minnesota, Rutgers, CMU, New Mexico, USC, and UC Berkeley). He is currently the principal investigator of three external funded ($21.4M) Big Data projects: DARPA Anomaly Detection at Multiple Scales (ADAMS), DARPA Social Media in Strategic Communications (SMISC), and ARL Social and Cognitive Network Academic Research Center (SCNARC). He leads an IBM major initiative on graph computing for Linked Big Data called IBM System G. He was elevated to fellow for of the IEEE for "*contributions to network science and multimedia security and retrieval*" in November 2011. He was the first IEEE fellow cited for contribution to network science.

Jimmy Lin received the PhD degree in electrical engineering and computer science from MIT in 2004. He holds the David R. Cheriton chair in the David R. Cheriton School of Computer Science at the University of Waterloo. His research lies at the intersection of information retrieval and natural language processing; his current work focuses on large-scale distributed algorithms and infrastructure for data analytics. From 2010 to 2012, he spent an extended sabbatical at Twitter, where he worked on services designed to surface relevant content to users and analytics infrastructure to support data science.

Jiangchuan Liu (S'01-M'03-SM'08) received the BEng degree (cum laude) from Tsinghua University, Beijing, China, in 1999, and the PhD degree from The Hong Kong University of Science and Technology in 2003. He is a full professor in the School of Computing Science, Simon Fraser University, British Columbia, Canada, and an NSERC E.W.R. Steacie Memorial Fellow. He is an EMC-Endowed visiting chair professor of Tsinghua University, Beijing, China (2013-2016). From 2003 to 2004, he was an assistant professor at The Chinese University of Hong Kong. He co-received the inaugural Test of Time Paper Award of IEEE INFOCOM (2015), ACM TOMCCAP Nicolas D. Georganas Best Paper Award (2013), ACM Multimedia Best Paper Award (2012), IEEE Globecom Best Paper Award (2011), and IEEE Communications Society Best Paper Award on Multimedia Communications (2009). His research interests include multimedia systems and networks, cloud computing, social networking, online gaming, big data computing, wireless sensor networks, and peer-to-peer and overlay networks. He has served on the editorial boards of *IEEE Transactions on Big Data*, *IEEE Transactions on Multimedia*, *IEEE Communications Surveys and Tutorials*, *IEEE Access*, *IEEE Internet of Things Journal*, *Elsevier Computer Communications*, and *Wiley Wireless Communications and Mobile Computing*. He is a senior member of the IEEE.

Nathan Liu received the PhD degree in computer science from Hong Kong University of Science and Technology. He is a senior research scientist at Yahoo! Labs. His research interests include large-scale machine learning and data mining with applications to search and recommendation systems.

T.S. Eugene Ng received the PhD degree in computer science from Carnegie Mellon University. He is an associate professor of computer science at Rice University. He received the US National Science foundation (NSF) CAREER Award and an Alfred P. Sloan Fellowship. His research interest lies in developing new network models, network architectures, and holistic networked systems that enable a robust and manageable network infrastructure.

Raymond Ng is a professor of computer science at the University of British Columbia. His main research area for the past two decades is on data mining, with a specific focus on health informatics and text mining. He has published more than 180 peer-reviewed publications on data clustering, outlier detection, OLAP processing, health informatics and text mining. He received two Best Paper Awards—from 2001 ACM SIGKDD conference, which is the premier data mining conference worldwide, and the 2005 ACM SIGMOD conference, which is one of the top database conferences worldwide. He was one of the program co-chairs of the 2009 International Conference on Data Engineering, and one of the program co-chairs of the 2002 ACM SIGKDD Conference. He was also one of the general co-chairs of the 2008 ACM SIGMOD conference. For the past decade, he has co-led several large scale genomic projects, funded by Genome Canada, Genome BC, and industrial collaborators. The total amount of funding of those projects well exceeded $40 million Canadian dollars. He now holds the chief informatics officer position of the PROOF Centre of Excellence, which focuses on biomarker development for end-stage organ failures. On the text mining side, he has co-led various projects focusing on managing conversations, such as emails, blogs, tweets, and meetings. One primary objective is to create metadata, such as natural language summaries, for conversations, to facilitate access to the otherwise unstructured raw data. Those projects were partially funded by NSERC, Google, and Yahoo.

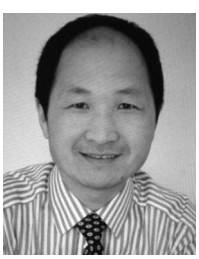
Lionel M. Ni is a chair professor in the Department of Computer and Information Science and Vice Rector, Academic Affairs, University of Macau. He has chaired over 30 professional conferences and has received eight awards for authoring outstanding papers. He served on the editorial boards of the *IEEE Transactions on Computers* and *IEEE Transactions on Parallel and Distributed Systems* and is serving on the editorial boards of *Communications of the ACM* and *ACM Transactions on Sensor Networks*. His current research interests include big data, wireless sensor networks, mobile computing, and pervasive computing. He is a fellow of the IEEE and Hong Kong Academy of Engineering Science.

Zoran Obradovic is a professor of computer and information sciences and the director in the Center for Data Analytics and Biomedical Informatics, Temple University in Philadelphia and is an internationally recognized leader in data mining and bioinformatics. He joined Temple in 2000 from Washington State University, where he was named a researcher of the Year by the College of Engineering and Architecture. At Temple in 2008, he received College of Science and Technology Faculty Research Excellence Award and in 2009, the overall Temple University Faculty Research Award. His research focuses on improving predictive modeling and decision support through data-driven discovery and modeling of hidden patterns in large data sets. He has published about 300 articles addressing data mining challenges in health informatics, the social sciences, environmental management, and other domains. His group's pioneering research on the prediction and functional analysis of intrinsically disordered regions in proteins has provided new insight into how protein structure establishes function and the program his team developed was the best rated predictor of intrinsic disorder at three consecutive international competitions organized by protein structure prediction assessment community (CASP 5-7). He is a general co-chair for 2013 and 2014 SIAM International Conference on Data Mining and was the program and/or a track chair at many data mining and biomedical informatics conferences. He currently serves as an editorial board member for seven journals and is the executive editor for applications at *Statistical Analysis and Data Mining* journal.

Yew-Soon Ong received the PhD degree in artificial intelligence in complex design from the Computational Engineering and Design Center, University of Southampton, United Kingdom in 2003. He is currently an associate professor with the School of Computer Engineering, Nanyang Technological University (NTU), Singapore, a director in the A*Star SIMTECH-NTU Joint Lab on Complex Systems and a Program Principal investigator in the Rolls-Royce@NTU Corporate Lab on Large Scale Data Analytics. His current research interest in computational intelligence spans across memetic computing, evolutionary design, machine learning and Big data. He is the founding technical editor-in-chief of *Memetic Computing Journal*, founding chief editor of the Springer book series on studies in Adaptation, Learning, and Optimization, an associate editor of the *IEEE Transactions on Evolutionary Computation*, the *IEEE Transactions on Neural Networks & Learning Systems*, *IEEE Computational Intelligence Magazine*, *IEEE Transactions on Cybernetics*, and others. He received the 2015 IEEE Computational Intelligence Magazine Outstanding Paper Award and the 2012 IEEE Transactions on Evolutionary Computation Outstanding Paper Award.

Manish Parashar is a distinguished professor of computer science at Rutgers University. He is also the founding director in the Rutgers Discovery Informatics Institute (RDI2). His research interests are in the broad areas of Parallel and Distributed Computing and Computational and Data-Enabled Science and Engineering. He serves on the editorial boards and organizing committees of a large number of journals and international conferences and workshops, and has deployed several software systems that are widely used. For more information please visit http://parashar.rutgers.edu/. He has also received a number of awards and is a fellow of AAAS, a fellow of the IEEE/IEEE Computer Society, and an ACM distinguished scientist.

Jian Pei received the PhD degree in computing science from the Simon Fraser University in 2002 under Dr. Jiawei Han's supervision. He is a Canada research chair (Tier 1) in Big Data Science, and a professor of computing science at the School of Computing Science and the Department of Statistics and Actuarial Science at Simon Fraser University, Canada. His research interests are to develop effective and efficient data analysis techniques for novel data intensive applications. Particularly, he is currently interested in various techniques of data mining, web search, information retrieval, data warehousing, online analytical processing, and database systems, as well as their applications in social networks, healthcare-informatics, business intelligence. His research has been supported in part by government programs and many industry partners. Since 2000, he has published one monograph and over 200 research papers in refereed journals and conferences. His publications have been cited by more than 50,000 times. He has served in the organization committees and the program committees of over 200 international conferences and workshops, and has been a reviewer for the leading academic journals in his fields. He is the editor-in-chief of *IEEE Transactions of Knowledge and Data Engineering* (*TKDE*), and an associate editor or editorial board member of a few premier journals in his areas. He is an ACM distinguished speaker. He received the 2014 IEEE ICDM Research Contributions Award, the British Columbia Innovation Council 2005 Young Innovator Award, an NSERC 2008 Discovery Accelerator Supplements Award, an IBM Faculty Award (2006), a KDD Best Application Paper Award (2008), an IEEE Outstanding Paper Award (2007). He is a fellow of the IEEE and a senior member of the ACM.



Yonghong Peng is a principal investigator in Big Data Science and Technology (BDST) and Bio-Medical Informatics (BMI) at the University of Bradford, United Kingdom. He is the chair for Big Data Task Force (BDTF) of IEEE Computational Intelligence Society (IEEE CIS), and a member of Data Mining and Big Data Analytics Technical Committee of IEEE CIS. He is also a founding member of IEEE Technical SubCommittee on Big Data (TSCBD) of IEEE Communications Society, and a member of Big Data Task Force of China Information Industry Association (CIIA). He is currently acting as an associate editor for *IEEE Transactions on Big Data*, a member of editorial board of the *International Journal of Big Data Intelligence*, and an academic editor of *PeerJ* and *PeerJ Computer Science*.

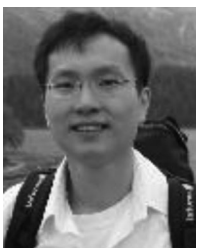
Clifton Pfua received the bachelor's (first class honors) and PhD (on identity crime detection) degrees are from Clayton School of Information Technology, Monash University, Australia. He is at NCS Group, working on Big Data analytics under the Smart and Safe City Centre of Excellence. In his free time, he volunteers professional services to events, conferences, and journals (related to data mining/analytics, security and health informatics). He was also part of teams which won some analytics competitions. His specialization was Big Data analytics in public security (attack and disaster preparation/recovery/response; cyber security; internal security; and predictive policing) and fraud (government, banking, and insurance). His teams have won challenges such as GE Flight Quest Phase 1 Competition 2013 and Fraud Detection in Mobile Advertising Competition 2012. He has published more than 30 technical papers on various industry applications of data analytics, such as fraud detection and activity recognition.

Sudarsan Rachuri is the associate program manager for smart manufacturing design and analysis program at NIST. Prior to joining NIST, he was a research professor at George Washington University. His primary research objectives are to develop and transfer knowledge to industry about information models for sustainable manufacturing, green products, assembly representation, system level analysis, and tolerance representation. Specific focus is on identifying integration and technology issues that promote industry acceptance of information models, and standards that will enable designers to develop products that are sustainable and manufactured in a distributed and collaborative environment. His primary areas of interest are smart and sustainable manufacturing, scientific computing, CAD/CAM/CAE, design for sustainability, data analytics, and ontology. He is an ASME fellow, having been elected in 2012 for his significant contributions in the areas of information and semantic modeling of product life cycle management, and the application of measurement science for sustainable manufacturing.

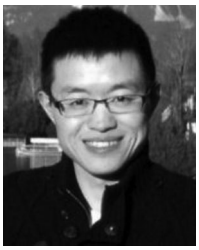
Cynthia Rudin received the undergraduate degree from the University at Buffalo and the PhD degree in applied and computational mathematics from Princeton University. She is an associate professor of statistics at the Massachusetts Institute of Technology associated with the Computer Science and Artificial Intelligence Laboratory and the Sloan School of Management, and directs the Prediction Analysis Lab. Her interests are in machine learning, data mining, applied statistics, and knowledge discovery (Big Data). Her application areas are in energy grid reliability, healthcare, and computational criminology. Previously, she was an associate research scientist at the Center for Computational Learning Systems at Columbia University, and prior to that, the US National Science Foundation (NSF) postdoctoral research fellow at NYU. She received the the College of Arts and Sciences Outstanding Senior Award in Sciences and Mathematics, and three separate outstanding senior awards from the Departments of Physics, Music, and Mathematics all from the University at Buffalo. She is the recipient of the 2013 INFORMS Innovative Applications in Analytics Award, an NSF CAREER award, and was named as one of the "Top 40 Under 40" by Poets and Quants in 2015. Her work has been featured in *Businessweek*, *The Wall Street Journal*, the *New York Times*, the *Boston Globe*, the *Times of London*, Fox News (Fox & Friends), the *Toronto Star*, *WIRED Science*, *U.S. News and World Report*, *Slashdot*, *CIO* magazine, Boston Public Radio, and on the cover of *Computer*. She is currently the chair-elect for the INFORMS Data Mining Section, and currently serves on committees for US Defense Advanced Research Projects Agency (DARPA), the National Academy of Sciences, the US Department of Justice, and the American Statistical Association.

Claudio T. Silva is a professor of computer science and engineering and data science at New York University. His research interests include visualization, visual analytics, reproducibility and provenance, geometric computing, data science/big data, sports analytics, urban computing and computer graphics. He has held positions in academia and industry, including at AT&T, IBM, Lawrence Livermore National Labs, Sandia National Labs, and the University of Utah. He has advised 15 PhD and eight MS students, and mentored six postdoctoral associates. He has published more than 220 journal and conference papers, is an inventor of 12 US patents, and authored 12 papers that have received "Best Paper Awards" (including honorable mentions). He has over 10,000 citations according to Google Scholar. He received the 2014 IEEE Visualization Technical Achievement Award "in recognition of seminal advances in geometric computing for visualization and for contributions to the development of the VisTrails data exploration system." His research has been funded by the US National Science Foundation (NSF), US Department of Energy (DOE), NIH, NASA, DOD, AT&T, IBM, ExxonMobil, McGraw-Hill Education, MLBAM, Moore Foundation, Sloan Foundation, LLNL, Sandia, Los Alamos, State of Utah, University of Utah, Center for Urban Science and Progress, and New York University. He is a fellow of the IEEE.

Malcolm Slaney is a research scientist in the Machine Hearing Group at Google Research. He is a consulting professor at Stanford CCRMA, where he has led the Hearing Seminar for more than 20 years, and an affiliate faculty in the Electrical Engineering Department, University of Washington. He is a coauthor, with A.C. Kak, of the IEEE book *Principles of Computerized Tomographic Imaging*. This book was republished by SIAM in their "Classics in Applied Mathematics" Series. He is a coeditor, with Steven Greenberg, of the book *Computational Models of Auditory Function*. Before joining Google, He was at Bell Laboratory, Schlumberger Palo Alto Research, Apple Computer, Interval Research, IBM's Almaden Research Center, Yahoo! Research, and Microsoft Research. For many years, he has led the auditory group at the Telluride Neuromorphic Cognition Workshop. His recent work is on understanding the role of attention in conversational speech and general audio perception. He is a fellow of the IEEE.



Tsz-Wo Sze received the BEng and MPhil degrees from the Hong Kong University of Science and Technology respectively in 1999 and 2001, and the PhD degree in computer science from the University of Maryland College Park in 2007. He is a member of Technical Staff at Hortonworks and also a member of the Project Management Committee at Apache Hadoop. His interests include distributed computing, algorithms and mathematical analysis. He is one of the earliest contributors of the Hadoop Distributed File System (HDFS) and remains one of the most active contributors. He used Hadoop with Yahoo's clusters to accomplish a world record of Pi computation in 2010.

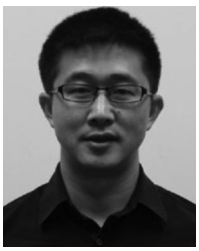


Jie Tang is an associate professor with the Department of Computer Science and Technology, Tsinghua University. His interests include social network analysis, data mining, and machine learning. He published more than 100 journal/conference papers and holds 10 patents. He served as a PC co-chair of WSDM'15, ASONAM'15, ADMA'11, SocInfo'12, KDD-CUP co-chair of KDD'15, Poster co-chair of KDD'14, Workshop co-chair of KDD'13, local chair of KDD'12, Publication co-chair of KDD'11, and as the PC member of more than 50 international conferences. He is the principal investigator of National High-tech R&D Program (863), NSFC project, Chinese Young Faculty Research Funding, National 985 funding, and international collaborative projects with Minnesota University, IBM, Google, Nokia, Sogou, etc. He leads the project Arnetminer.org for academic social network analysis and mining, which has attracted millions of independent IP accesses from 220 countries/regions in the world. He was honored with the Newton Advanced Scholarship Award, CCF Young Scientist Award, NSFC Excellent Young Scholar, and IBM Innovation Faculty Award.

Dacheng Tao (F'15) is a professor of computer science with the Centre for Quantum Computation & Intelligent Systems, and the Faculty of Engineering and Information Technology in the University of Technology, Sydney. He mainly applies statistics and mathematics to data analytics problems and his research interests spread across computer vision, data science, image processing, machine learning, and video surveillance. His research results have expounded in one monograph and 100+ publications at prestigious journals and prominent conferences, such as *IEEE TPAMI*, *T-NNLS, T-IP, JMLR, IJCV*, NIPS, ICML, CVPR, ICCV, ECCV, AISTATS, ICDM; and ACM SIGKDD, with several best paper awards, such as the best theory/algorithm paper runner up award in IEEE ICDM'07, the best student paper award in IEEE ICDM'13, and the 2014 ICDM 10 Year Highest-Impact Paper Award. He is a fellow of the IEEE.

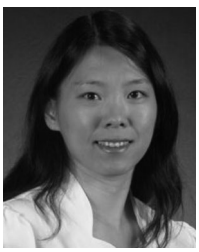
Ankur M. Teredesai is a professor of computer science & systems at the Institute of Technology, University of Washington Tacoma. His research spans data science with its applications for societal impact and social good. Apart from his academic appointments at RIT and the University of Washington Tacoma, he has significant industry experience, having held various positions at C-DAC Pune, Microsoft, IBM T.J. Watson Labs, and a variety of technology startups. He has published more than 70 papers on data mining and machine learning, has managed large teams of data scientists and engineers, and deployed numerous data science solutions from internet advertising, social networks, handwritten zip code recognizers, to his current focus on healthcare analytics. His recent applied research contributions include cost and risk prediction for readmission due to chronic conditions such as heart failure. Other applications of his work have enabled trust-enhanced recommendations, distributed data mining algorithms for big data, novelty detection in video, and link-analysis. He heads the Center for Data Science, and serves as the information officer for ACM SIGKDD (Special Interest Group in Knowledge Discovery and Data Mining). He is currently an associate editor for *ACM SIGKDD Explorations* and *IEEE Transactions on Big Data* and serves on program committees of major international conferences in data mining, machine learning, and related areas. He is an active advocate for women and nontraditional students to consider computing careers. He serves as the graduate program coordinator (GPC) for MS in Computer Science and Systems, and is the co-founder of the Information Technology degree program at University of Washington Tacoma.

Thomas Walsh received the PhD degree in computer science at Rutgers University, and previously held research positions at MIT, the University of Kansas, and the University of Arizona. He is a data scientist with Kronos Incorporated, where he applies machine learning and Big Data techniques to workforce management problems. His core research focuses on efficient learning in sequential decision making problems with rich structure.



Chonggang Wang received the PhD degree from Beijing University of Posts and Telecommunications (BUPT), China, in 2002. He is currently a member technical staff of InterDigital Communications with focuses on Internet of Things (IoT) R&D activities including technology development and standardization. His current research interest includes IoT, Mobile Communication and Computing, and Big Data Management and Analytics. He is the founding editor-in-chief of *IEEE Internet of Things Journal* and an IEEE ComSoc distinguished lecturer (2015-2016).

Honggang Wang received the PhD degree in computer engineering from University of Nebraska-Lincoln in 2009. He is an associate professor at UMass Dartmouth. His research interests include Big Data, Wireless Health, Body Area Networks (BAN), cyber and multimedia security, mobile multimedia and cloud, wireless networks, and cyber-physical system. He serves as an associate editor of the *IEEE Transactions on Big Data*, *IEEE Transactions on Multimedia*, *IEEE Communication Magazine*, *IEEE Internet of Things Journal* and *IEEE Access Journal*.



Wei Wang received the PhD degree in computer science from the University of California at Los Angeles in 1999. She is a professor in the Department of Computer Science, University of California at Los Angeles and the director in the Scalable Analytics Institute (ScAi). She was a professor in computer science and a member of the Carolina Center for Genomic Sciences and Lineberger Comprehensive Cancer Center at the University of North Carolina at Chapel Hill from 2002 to 2012, and was a research staff member at the IBM T. J. Watson Research Center between 1999 and 2002. Her research interests include big data, data mining, bioinformatics and computational biology, and databases. She has filed seven patents, and has published one monograph and more than one hundred research papers in international journals and major peer-reviewed conference proceedings. She received the IBM Invention Achievement Awards in 2000 and 2001. She received a UNC Junior Faculty Development Award in 2003 and the US National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award in 2005. She was named a Microsoft Research New Faculty fellow in 2005. She was honored with the 2007 Phillip and Ruth Hettleman Prize for Artistic and Scholarly Achievement at UNC. She was recognized with an IEEE ICDM Outstanding Service Award in 2012 and an Okawa Foundation Research Award in 2013. She has been an associate editor of the *IEEE Transactions on Knowledge and Data Engineering*, *ACM Transactions on Knowledge Discovery in Data*, *Journal of Knowledge and Information Systems*, *Journal of Data Mining and Knowledge Discovery*, *International Journal of Knowledge Discovery in Bioinformatics*, and an editorial board member of the *International Journal of Data Mining and Bioinformatics* and the *Open Artificial Intelligence Journal*. She serves on the organization and program committees of international conferences including ACM SIGMOD, ACM SIGKDD, ACM BCB, VLDB, ICDE, EDBT, ACM CIKM, IEEE ICDM, SIAM DM, SSDBM, BIBM.

Limsoon Wong received the BSc (Eng) degree in 1988 from Imperial College London and the PhD degree in 1994 from the University of Pennsylvania. He is a KITHCT chair professor of computer science and a professor of pathology at the National University of Singapore. He currently works mostly on knowledge discovery technologies and their application to biomedicine. He has also done significant research in database query language theory and finite model theory, as well as significant development work in broad-scale data integration systems. Other recognitions for his scientific works in these two fields include the 2003 FEER Asian Innovation Gold Award, for his work on treatment optimization of childhood leukemias; and the ICDT 2014 Test of Time Award, for his work on naturally embedded query languages. He is a fellow of the ACM, named in 2013 for his contributions to database theory and computational biology.

Haitao (Tony) Xia received the BE and ME degrees from Southeast University, Nanjing, China, and the PhD degree from the University of Oklahoma in 2004. He is a senior manager of R&D at Avago Technologies, leading the research and development of advanced read channel and Serdes architectures for data storage systems. He is the current president of Chinese American Information Storage Society (CAISS), President of Southeast University Silicon Valley Alumni Association, an associate editor of *IEEE Transactions on Big Data*, and was the chairman of IEEE Data Storage Technical Committee from 2013 to 2014. Before his work at Avago/LSI, he was at Linked-A-Media Devices on signal processing and coding in the area of magnetic recording channels and nonvolatile memories. He has published more than 20 articles in peer-reviewed journals/conferences, and has more than 80 US patent granted to his name. He is a senior member of the IEEE and a member of Sigma Xi.

Hui Xiong received the BE degree in automation from the University of Science and Technology of China (USTC), Hefei, China, the MS degree in computer science from the National University of Singapore (NUS), Singapore, and the PhD degree in computer science from the University of Minnesota-Twin Cities, USA, in 2005. He is currently a professor and vice chair in the Management Science and Information Systems Department, and the director in Rutgers Center for Information Assurance, at Rutgers, the State University of New Jersey, where he received a two-year early promotion/tenure (2009), the Rutgers University Board of Trustees Research Fellowship for Scholarly Excellence (2009), the ICDM-2011 Best Research Paper Award (2011), an IBM ESA Innovation Award (2008), the Junior Faculty Teaching Excellence Award (2007), the Junior Faculty Research Award (2008), and Dean's Award for Meritorious Research (2010, 2011, 2013) at Rutgers Business School. Dr. Xiong's general area of research is data and knowledge engineering, with a focus on developing effective and efficient data analysis techniques for emerging data intensive applications. His research has been supported in part by the National Science Foundation (NSF), IBM Research, SAP Corporation, Panasonic USA Inc., Awarepoint Corp., Citrix Systems Inc., and Rutgers University. He has published prolifically in refereed journals and conference proceedings, such as *IEEE Transactions on Knowledge and Data Engineering*, the *VLDB Journal*, *INFORMS Journal on Computing*, *Machine Learning, the Data Mining and Knowledge Discovery Journal*, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), SIAM International Conference on Data Mining (SDM), IEEE International Conference on Data Mining (ICDM), and ACM International Symposium on Advances in Geographic Information Systems (ACM GIS). He is a co-Editor-in-Chief of *Encyclopedia of GIS* (Springer, 2008) and an Associate Editor of *IEEE Transactions on Data and Knowledge Engineering* (*TKDE*) and *ACM Transactions on Management Information Systems* (*TMIS*). He has served regularly on the organization and program committees of numerous conferences, including as a Program Co-Chair of the Industrial and Government Track for the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), a Program Co-Chair for the IEEE 2013 International Conference on Data Mining (ICDM), and a General Co-Chair for the IEEE 2015 International Conference on Data Mining (ICDM). He is an ACM distinguished scientist and a senior member of the IEEE.

Daqing Zhang received the PhD degree from the University of Rome "La Sapienza" in 1996. He is a full professor at Telecom SudParis, Institut Mines-Telecom, France. His research interests include context-aware computing, urban computing, mobile computing, big data analytics, pervasive elderly care, etc. He has published more than 200 technical papers in leading conferences and journals, where his work on context model is widely accepted by the pervasive computing, mobile computing and service-oriented computing communities. He served as the general or program chair for more than 10 international conferences, giving keynote talks at more than 16 international conferences. He is the associate editor for four journals including *ACM Transactions on Intelligent Systems and Technology*, *IEEE Transactions on Big Data*, etc. He is the winner of the 10-years CoMoRea impact paper award at IEEE PerCom 2013, the Best Paper award at IEEE UIC 2012 and the Best Paper Runner Up award at Mobiquitous 2011.

Tong Zhang received the PhD degree from Stanford Computer Science. After graduation, he was at IBM T.J. Watson Research Center in Yorktown Heights, New York, and Yahoo! Research in New York City. He is currently with the Statistics Department, Rutgers University.

Yu Zheng is a lead researcher in Microsoft Research and a chair professor at Shanghai Jiao Tong University. His research interests include big data analytics, spatiotemporal data mining, applied machine learning, and ubiquitous computing. Currently, he is leading the research on urban computing in Microsoft Research, and is passionate about using big data to tackle urban challenges. He received five Best Paper Awards from ICDE'13 and ACM SIGSPATIAL'10, etc. Zheng currently serves as the editor-in-chief of *ACM Transactions on Intelligent Systems and Technology* and is a member of Editorial Advisory Board of *IEEE Spectrum*. Zheng has served as chair on over 10 prestigious international conferences—most recently, as the program co-chair of ICDE 2014 (Industrial Track). He has been invited to give over 10 keynote speeches at international conferences and forums (e.g., APEC 2014 Smart City Forum). In 2013, he was named one of the Top Innovators under 35 by *MIT Technology Review* (TR35) and featured by *Time Magazine* for his research on urban computing. He is also an adjunct professor at Hong Kong Polytechnic University and an affiliate professor at Southwest Jiaotong University.

Wenwu Zhu received the PhD degree from New York University Polytechnic School of Engineering in 1996 in electrical and computer engineering. He is currently a professor and deputy head in the Computer Science Department, Tsinghua University. Prior to his current post, he was a senior researcher and research manager at Microsoft Research Asia. He was the chief scientist and director at Intel Research China from 2004 to 2008. He was at Bell Labs New Jersey as a member of Technical Staff during 1996-1999. He has published over 200 refereed papers in the areas of multimedia computing, communications and networking. He is a inventor or co-inventor of over 50 patents. His current research interests are in the area of multimedia big data computing, social multimedia computing, multimedia cloud computing, and multimedia communications and networking. He served(s) on various editorial boards, such as guest editors for the *Proceedings of the IEEE*, *IEEE T-CSVT*, and *IEEE JSAC*; an associate editor for *IEEE Transactions on Mobile Computing*, *IEEE Transactions on Multimedia*, *IEEE Transactions on Circuits and Systems for Video Technology,* and *IEEE Transactions on Big Data*. He received the six Best Paper Awards including ACM Multimedia 2012 and *IEEE Transactions on Circuits and Systems for Video Technology* in 2001. He is a fellow of the IEEE, SPIE fellow, and ACM distinguished scientist.