

# DSAA'2015 Keynote Speeches

**Title: Toward personal knowledge bases**

**Serge Abiteboul**  
**INRIA & ENS Cachan**

## *Abstract*

A Web user today has his/her data and information distributed in a number of services that operate in silos. Computer wizards already know how to control their personal data to some extent. It is now becoming possible for everyone to do the same, and there are many advantages to doing so. Everyone should now be in a position to manage his/her personal information. Furthermore, we will argue that we should move towards personal knowledge bases and discuss advantages to do so. We will mention recent works around a datalog dialect, namely Webdamlog.

## *Short Bio*



Serge Abiteboul obtained his Ph.D. from the University of Southern California, and a State Doctoral Thesis from the University of Paris-Sud. He has been a researcher at the Institut National de Recherche en Informatique et Automatique since 1982 and is now Distinguished Affiliated Professor at Ecole Normale Supérieure de Cachan. He was a Lecturer at the École Polytechnique and Visiting Professor at Stanford and Oxford University. He has been Chair Professor at Collège de France in 2011-12 and Francqui Chair Professor at Namur University in 2012-2013. He co-founded the company Xyleme in 2000. Serge Abiteboul has received the ACM SIGMOD Innovation Award in 1998, the EADS Award from the French Academy of sciences in 2007; the Milner Award from the Royal Society in 2013; and a European Research Council Fellowship (2008-2013). He became a member of the French Academy of Sciences in 2008, and a member the Academy of Europe in 2011. He is a member of the Conseil national du numérique and Chairman of the Scientific board of the Société d'Informatique de France. His research work focuses mainly on data, information and knowledge management, particularly on the Web.

## **Title: On information propagation, social influence, and communities**

**Francesco Bonchi**

**Research Leader at ISI Foundation, Turin, Italy, and Eurecat, Technological Center of Catalunya, Barcelona, Spain**

### ***Abstract***

With the success of online social networks and microblogging platforms such as Facebook, Tumblr, and Twitter, the phenomenon of influence-driven propagations, has recently attracted the interest of computer scientists, sociologists, information technologists, and marketing specialists. In this talk we will take a data mining perspective, discussing what (and how) can be learned from a social network and a database of traces of past propagations over the social network. Starting from one of the key problems in this area, i.e. the identification of influential users, we will provide a brief overview of our recent contributions in this area. We will expose the connection between the phenomenon of information propagation and the existence of communities in social network, and we will go deeper in this new research topic arising at the overlap of information propagation analysis and community detection.

### ***Short Bio***



Francesco Bonchi is Research Leader at the ISI Foundation, Turin, Italy, where he leads the "Algorithmic Data Analytics" group. He is also Scientific Director for Data Mining at Eurecat (Technological Center of Catalunya), Barcelona. Before he was Director of Research at Yahoo Labs in Barcelona, Spain, where he was 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 will be PC Chair of the 16th IEEE International Conference on Data Mining (ICDM 2016) to be held in Barcelona in December 2016. He is member of the ECML PKDD Steering Committee, Associate Editor of the newly created IEEE Transactions on Big Data (TBD), 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 member of the Editorial Board of Data Mining and Knowledge Discovery (DMKD). He has been program co-chair of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2010). Dr. Bonchi has also served as program co-chair of the first and second ACM SIGKDD International Workshop on Privacy, Security, and Trust in KDD (PinKDD 2007 and 2008), the 1st IEEE International Workshop on Privacy Aspects of Data Mining (PADM 2006), and the 4th International Workshop on Knowledge Discovery in Inductive Databases (KDID 2005). He is co-editor of the book "Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques" published by Chapman & Hall/CRC Press.

He earned his Ph.D. in computer science from the University of Pisa in December 2003.

## **Title: Towards Data-driven Models of Human Behavior**

**Nuria Oliver**

**Scientific Director at Telefonica R&D**

### ***Abstract***

### ***Short Bio***



Nuria Oliver is currently the Scientific Director and founder of the User, Data and Media Intelligence research team in Telefonica Research (Barcelona, Spain). She is responsible for the HCI, Mobile Computing, Big and Personal Data Mining, User Modeling and Multimedia Research Areas. Nuria received the BSc (honors) and MSc degrees in Electrical Engineering and Computer Science from the ETSIT at the Universidad Politecnica of Madrid (UPM), Spain, in 1992 and 1994 respectively. She received her PhD degree from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in June 2000. From July 2000 until November 2007, she was a researcher at Microsoft Research in Redmond, WA. At the end of 2007, she returned to Spain to co-create the Research organization at Telefonica R&D by creating and leading her research team.

Her research interests include artificial intelligence, health monitoring, mobile computing, personal and big data analysis, statistical machine learning and data mining, smart environments, context awareness, multimedia data analysis, recommender systems, social network analysis, computational social sciences, and human computer interaction. She is currently working on the previous disciplines to build human-centric intelligent systems and improve the world with technology.

Nuria has written over 100 scientific papers in international conferences, journals and book chapters. Her work has been widely recognized by the scientific community with over 8200 citations. According to Google Scholar Nuria is the most cited female computer scientist in Spain. Nuria has over 40 patent applications and granted patents. She is also in the program committee and a reviewer of the top conferences in her research areas (IJCAI, IUI, UMAP, ACM Multimedia, ICMI-MLMI, SocialComp, Interaccion, PervasiveHealth, MIR, LoCA, MMM, CVPR, Ubicomp, MobileHCI, ICCV, AAI, etc...). She was program co-chair of IUI 2009 and of MIR 2010, general conference co-chair of UMAP 2011, industry-day co-chair of IJCAI 2011, track co-chair of ACM WWW 2013 and track founder and co-chair in ACM MM 2014, among others.

She regularly gives invited talks about her research both for scientific and the general public. She has been a keynote speaker at IEEE EUSIPCO 2011, European Wireless Conference 2014, ACM and IEEE MODELS 2014 and NTTS 2015, among others. She has given two TEDx talks, a WIRED talk and is co-organizer of the first ever TEDxBCN event devoted to Education. She has spoken to the Spanish Senate.

She believes in the power of technology to empower and increase the quality of life of people. She has received a number of awards, including a ACM ICMI 2014 10-year Technical Impact Award (2014), a Rising Talent Award by the Women's Forum for the Economy and

Society (2009), MIT's 'TR100 Young Innovators Award' (2004) and the First Spanish Award of EECS graduates (1994). She is a Senior Member of the ACM and IEEE.

Besides her scientific publications, she is very interested in making science available to the general public. She has been a technology writer for Tecno2000 magazine and 'El Pais' newspapers, among others. Her work has been featured on multiple newspapers, magazines, radio and TV stations both in Spain and the US. She has been featured in Glamour magazine as 'one of the top female directors in Spain' (2015), diari ARA, in EL PAIS Sunday magazine as one of a few 'female directors in technology' (2012), named <% Rising Talent by the Women's Forum for Economy & Society (October 2009),%> one of the '13 most influential young women in Spain' (MujerHoy Magazine, 2012), one of '100 leaders of the future' by Capital Magazine (May 2009) and one of the 'Generation XXI: 40 Spanish youngsters that will make news in the Third Millenium ' by EL PAIS (2000).

She is proficient in multiple foreign languages and she studied classical ballet for 13 years. Her hobbies include traveling, cinema, art, classical music, yoga, Formula Dodge racing, dance and swimming.

## **Title: The multi-facets of a data science project to answer: how are organs formed?**

**Bin Yu**

**Chancellor's Professor, Departments of Statistics and EECS, University of California at Berkeley**

### ***Abstract***

Genome wide data reveal an intricate landscape where gene actions and interactions in diverse spatial areas are common both during development and in normal and abnormal tissues. Understanding local gene networks is thus key to developing treatments for human diseases. Given the size and complexity of recently available systematic spatial data, defining the biologically relevant spatial areas and modeling the corresponding local biological networks present an exciting and on-going challenge. It requires the integration of biology, statistics and computer science; that is, it requires data science.

In this talk, I present results from a current project co-led by biologist Erwin Frise from Lawrence Berkeley National Lab (LBNL) to answer the fundamental systems biology question in the talk title. My group (Siqi Wu, Antony Joseph, Karl Kumbier) collaborates with Dr. Erwin and other biologists (Ann Hommands) of Celniker's Lab at LBNL that generate the *Drosophila* spatial expression embryonic image data. We leverage our group's prior research experience from computational neuroscience to use appropriate ideas of statistical machine learning in order to create a novel image representation decomposing spatial data into building blocks (or principal patterns). These principal patterns provide an innovative and biologically meaningful approach for the interpretation and analysis of large complex spatial data. They are the basis for constructing local gene networks, and we have been able to reproduce almost all the links in the Nobel-prize winning (local) gap-gene network. In fact, Celniker's lab is running knock-out experiments to validate our predictions on gene-gene interactions. Moreover, to understand the decomposition algorithm of images, we have derived sufficient and almost necessary conditions for local identifiability of the algorithm in the noiseless and complete case. Finally, we are collaborating with Dr. Wei Xue from Tsinghua Univ to devise a scalable open software package to manage the acquisition and computation of imaged data, designed in a manner that will be usable by biologists and expandable by developers.

### ***Short Bio***



Bin Yu is Chancellor's Professor in the Departments of Statistics and of Electrical Engineering & Computer Science at the University of California at Berkeley. Her current research interests focus on statistics and machine learning theory, methodologies, and algorithms for solving high-dimensional data problems. Her group is engaged in interdisciplinary research with scientists from genomics, neuroscience, and remote sensing.

She obtained her B.S. degree in Mathematics from Peking University in 1984, her M.A. and Ph.D. degrees in Statistics from the University of California at Berkeley in 1987 and 1990, respectively. She held faculty positions at the Univ of Wisconsin-Madison and Yale University and was a Member of Technical Staff at Bell Labs, Lucent. She was Chair of Department of Statistics at UC

Berkeley from 2009 to 2012, and is a founding co-director of the Microsoft Lab on Statistics and Information Technology at Peking University, China.

She is Member of the U.S. National Academy of Sciences and Fellow of the American Academy of Arts and Sciences. She was a Guggenheim Fellow in 2006, an Invited Speaker at ICIAM in 2011, and the Tukey Memorial Lecturer of the Bernoulli Society in 2012. She was President of IMS (Institute of Mathematical Statistics) in 2013-2014.