BIG DATA INSTITUTE

UCL

Theory Big Data

1st UCL Workshop on the Theory of Big Data

7th - 9th January 2015

Conference Programme

Contents

Welcome Message

Dear Colleagues,

Welcome to UCL and to Theory of Big Data 2015! We are delighted to host you at UCL over the coming three days, and hope that if you have any questions or if there is anything we can do to help, you won't hesitate to ask. This Workshop is part of the UCL Big Data Institute activities. This institute organises activities around Big Data at UCL, and is itself part of a broader UCL initiative on Data Science. The aim of the workshop series is to crystalize developments around the theoretical underpinning of Big Data, hoping to encourage future innovations.

Welcome to London!

From the organising committee:
Sofia Olhede & Patrick Wolfe
Arthur Gretton,
Franz Kiraly,
Ioannis Kosmidis,
Richard Samworth,
Ricardo Silva,
Ming Yuan,
Russell Rodrigues

The Theory of Big Data Workshop is organised by the UCL Big Data Institute, with support from its founding partner, Elsevier.

For more information about the Big Data Institute, visit www.ucl.ac.uk/big-data/bdi or follow us on Twitter @uclbdi

Venue location

All workshop activities will take place in the Anatomy Building (entrance via Gower Street) – outlined in red above. [View a Google Street View shot of the Building Entrance.]

Within the Anatomy Building, the conference talks will be in the JZ Young Theatre, posters will be presented in the JBS Haldane Hub (G28), and refreshments (plus lunch for delegates who have purchased a lunch ticket) will be in the Gavin de Beer room – all on the ground floor of the building and accessible from the main atrium. During the poster presentation session (15.10-17.00 on Thursday 8th January), refreshments will also be in the JBS Haldane Hub. Delegates are free to

browse the posters in the Hub at leisure, outside of the poster session, and to avail of the seating and table space in the Hub at any time.

Internet Connection Instructions

Visitors with access to Eduroam should be able to connect to the network automatically at UCL.

Guests without Eduroam access, please connect your device to the UCLGuest Network. Open your web browser and navigate to a page outside of UCL. The browser will automatically redirect to the UCLGuest Welcome page. Click the 'Self Service' link and enter your details, along with **Event Code 10144062** and click 'Generate Account'.

The system will generate login details (which will also be emailed to you). Make a note of them, then click the link to the 'Login' page and enter the details to connect. (It may take up to 60 seconds for your account to become active after it has been generated – if you cannot login, please wait a short while and try again).

Conference Programme:

Day 1: Wednesday 7th January 2015

09.00	Registration opens
10.00-10.10	Welcome and Introduction
10.10-11.10	Maximin effects in inhomogeneous large-scale data – Prof Nicolai Meinshausen
11.10-11.40	Break: Tea & Coffee available for all in Gavin de Beer room
11.40-12.20	Random walk models of graph formation – Dr Peter Orbanz
12.20-13.10	Understanding the Behaviour of Large Networks – Prof Patrick Wolfe
13.10-14.00	Break: Buffet lunch available for delegates who have purchased a lunch ticket
AFTERNOON SESSION: Coping with Big Data – an Analytics and Computational Perspective Organised by the Turing Gateway to Mathematics	
14.00-14.30	Registration. Tea/coffee available for all in Gavin de Beer room
14.30-14.35	Welcome and Introduction – Turing Gateway to Mathematics / University College London Jane Leeks (TGM) and Prof Patrick Wolfe (UCL)
14.35-15.05	Statistical Challenges of Big Data – Prof Niall Adams, Imperial College London
15.05-15.35	Big Data Challenges for Business – Giles Pavey, dunnhumby
15.35-16.05	Computational Limitations & Cloud Area – Prof Derek McAuley, Horizon Digital Economy Research
16.05-16.20	Break – tea/coffee available for all in Gavin de Beer room
16.20-16.50	Practical Challenges of Data Analytics – Leigh Lapworth, Rolls Royce
16.50-17.10	Questions and Chaired Open Discussion
17.10-	Tea / Coffee and Networking in Gavin de Beer room

Day 2: Thursday 8th January 2015

08.30	Registration opens
09.00-09.50	Network comparison – Prof Gesine Reinert
09.50-10.40	Some recent developments in large-scale data analysis – Dr Rajen Shah
10.40-11.10	Break – tea and coffee available for all in Gavin de Beer room
11.10-12.00	Learning Overcomplete Latent Variable Models through Tensor Decompositions – Prof Animashree Anandkumar
12.00-12.50	Multistage Bandits – Dr Philippe Rigollet
12.50-14.00	Break – buffet lunch available for delegates who have purchased a lunch ticket
14.00-15.10	Oral presentations: Sharmodeep Bhattacharhyya Patrick Rubin-Delanchy Clifford Lam Zoltan Szabo
15.10-17.00	Poster presentations and refreshments – both in JBS Haldane Hub

Day 3: Friday 9th January 2015

08.30	Registration opens
09.00-09.50	High-dimensional Ordinary Least-squares Projector for Screening Variables – Prof Chenlei Leng
09.50-10.40	Randomized dual coordinate ascent with arbitrary sampling – Dr Peter Richtarik
10.40-11.10	Break – tea and coffee available for all in Gavin de Beer room
11.10-12.00	General Bayesian updating for complex applications – Prof Chris Holmes
12.00-12.50	Vast time series segmentation – Prof Qiwei Yao
12.50-14.00	Break – buffet lunch available for delegates who have purchased a lunch ticket
14.00-14.45	Oral presentations: Dino Sejdinovic Heiko Strathmann Daniel John Lawson
14.45	Closing remarks

Invited Speakers: Profiles and Abstracts

Prof Nicolai Meinshausen, ETH Zurich

Maximin effects in inhomogeneous large-scale data
Speaking: 7th January, 10.10am

Abstract: Large-scale data are often characterised by some degree of inhomogeneity as data are either recorded in different time regimes or taken from multiple sources. We look at regression models and the effect of randomly changing coefficients, where the change is either smoothly in time or some other dimension or even without any such structure. Fitting varying-coefficient models or mixture models can be appropriate solutions but are computationally very demanding and often try to return more information than necessary. If we just ask for a model estimator that shows good predictive properties for all regimes of the data, then we are aiming for a simple linear model that is reliable for all possible subsets of the data. We propose a maximin effects estimator and look at its prediction accuracy from a theoretical point of view in a mixture model with known or unknown group structure. Under certain circumstances the estimator can be computed orders of magnitudes faster than standard penalised regression estimators, making computations on large-scale data feasible. We also show how a modified version of bagging can accurately estimate maximin effects.

Profile: Prof. Meinshausen is a Professor at the department of Statistics of ETH Zurich. In 2012 he was a Professor of Statistics at the University of Oxford and in 2007 a Post-doctoral fellow at the University of California Berkeley. He obtained his PhD from ETH Zurich and his Msc. in Applied Computational Mathematics from the University of Oxford. He was awarded the Guy Medal from the Royal Statistical Society in 2011 and he will give the IMS Medallion Lecture in 2015. Currently he is associate editor for the Journal of the Royal Statistical Society Series B and the Journal of Machine Learning Research.

Research Interests: Computational Statistics, High-dimensional Data, Regularization, Lasso-type Estimators, Sparsity, Machine Learning, Multiple Testing, Visualizations, Statistics for Astronomy and Climate Science.

Dr Peter Orbanz, Columbia University

Random walk models of graph formation
Speaking: 7th January, 11.40am

Abstract: The best-understood statistical models of graphs and networks are models based on exchangeable graphs. These are tractable by the standards of random graph models, but inherently (and provably) misspecified for sparse network data -- a graph generated from such a model would bear no resemblance to most networks arising in applications. Developing non-exchangeable models is much more difficult, both mathematically and in terms of inference. I will review the exchangeable case, present a particular type of non-exchangeable model that is statistically tractable, and discuss the more general problem of invariance in networks -- roughly, what is the sparse counterpart to an exchangeable graph? -- which remains unsolved.

Profile: Dr. Orbanz is an Assistant Professor in the Department of Statistics at Columbia University. Before going to New York, he was a Research Fellow in the Machine Learning Group of Zoubin Ghahramani at the University of Cambridge, and previously a graduate student of Joachim M. Buhmann at ETH Zurich.

Research Interests: His main research interests are the statistics of discrete objects and structures: permutations, graphs, partitions, binary sequences. Most of his recent work concerns representation problems and latent variable algorithms in Bayesian nonparametrics. More generally, he is interested in all mathematical aspects of machine learning and artificial intelligence.

Prof Patrick Wolfe, University College London

Understanding the Behaviour of Large Networks
Speaking: 7th January, 12.20pm

Abstract: In this talk - which will be accessible to a general audience - we show how the asymptotic behavior of random networks gives rise to universal statistical summaries. These summaries are related to concepts that are well understood in the other contexts outside of Big Data - such as stationarity and ergodicity - but whose extension to networks requires recent developments from the theory of graph limits and the corresponding analog of de Finetti's theorem. We introduce a new tool based on these summaries, which we call a network histogram, obtained by fitting a statistical model called a blockmodel to a large network. Blocks of edges play the role of histogram bins, and so-called network community sizes that of histogram bandwidths or bin sizes. For more details, see recent work in the Proceedings of the National Academy of Sciences (doi:10.1073/pnas.1400374111, with Sofia Olhede) and the Annals of Statistics (doi:10.1214/13-AOS1173, with David Choi).

Profile: Patrick J. Wolfe is Professor of Statistics and Honorary Professor of Computer Science at University College London, where he is a member of the Department's Senior Management Team and a Royal Society and EPSRC Mathematical Sciences Research Fellow. From 2001-2004 he held a Fellowship and College Lectureship in Engineering and Computer Science at Cambridge University, where he completed his PhD in 2003. Prior to joining UCL he was Assistant (2004-2008) and Associate (2008-2011) Professor at Harvard University. He serves on the Research Section Committee of the Royal Statistical Society, and on the editorial board of Applied and Computational Harmonic Analysis.

Research Interests: The mathematics of Big Data. Modelling and inference for graphs and networks; statistical imaging and image processing; time series and time-frequency analysis; audio signal processing and acoustic modelling.

Prof Gesine Reinert, University of Oxford

Network comparison

Speaking: 8th January, 09.00am

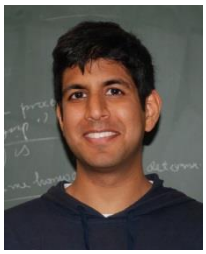
Abstract: Comparing networks is a key question in network analysis, such as networks of protein-protein interactions in different organisms, or trade networks of different commodities between countries. Often the comparison of interest is based on local structures. A typical approach for comparison of networks is alignment - finding matching nodes and aligning the edges between them as well as possible. This approach has as drawback that usually only a small part of the network can be aligned. Moreover it does not scale well with increasing network size.

Instead we shall use our new comparison network, called Netdis, which is not based on alignment, but rather on counts of small sub-graphs in an ensemble of sub-networks derived from the original networks. With this method, we find that a subsampling approach gives good results even when the comparisons are based only on a small fraction of the networks to be compared.

This is joint work with Waqar Ali, Robert Gaunt and Charlotte Deane.

Profile: University Lecturer, Department of Statistics, Oxford, and Fellow at Keble College, Oxford (2000 – present). Senior Research Fellow, King's College, Cambridge (1998 – 2000). Adjunct Assistant Professor, Department of Mathematics, UCLA, Los Angeles (1996 – 1998). Lecturer, Department of Mathematics, USC, Los Angeles (1994 – 1996). Ph.D. in Mathematics, University of Zurich, Title: A weak law of large numbers for empirical measures via Stein's method. Advisor: Prof. A.D. Barbour, D.Phil (1994 – present).

Research Interests: Applied probability, Computational biology and Statistics. In particular, Stein's method, networks, word count statistics.



Dr Rajen Shah, University of Cambridge

Min-wise hashing for large-scale regression

Speaking: 8th January, 09.50am

Abstract: We consider the problem of large-scale regression where both the number of predictors, p, and the number of observations, n, may be in the order of millions or more. Computing a simple OLS or ridge regression estimator for such data, though potentially sensible from a purely statistical perspective (if n is large enough), can be a real computational challenge. One recent approach to tackling this problem in the common situation where the matrix of predictors is sparse, is to first compress the data by mapping it to an n by L matrix with L << p, using a scheme called b-bit min-wise hashing (Li and König, 2011). We study this technique from a theoretical perspective and obtain finite-sample bounds on the prediction error of regression following such data compression, showing how it exploits the sparsity of the data matrix to achieve good statistical performance. Surprisingly, we also find that a main effects model in the compressed data is able to approximate an interaction model in the original data. Fitting interactions requires no modification of the compression scheme, but only a higher-dimensional mapping with a larger L. This is joint work with Nicolai Meinshausen (ETH Zürich).

Profile: Dr Rajen Shah is a Lecturer in Statistics at the Statistical Laboratory, which is part of the Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge.

Research Interests: His research interests include high-dimensional statistics and large-scale data analysis.

Prof Animashree Anandkumar, University of California Irvine

Spectral Methods for Unsupervised and Discriminative Learning with Latent Variables

Speaking: 8th January, 11.10am

Abstract: Incorporating latent or hidden variables is a crucial aspect of statistical modeling. I will present a statistical and a computational framework for guaranteed learning of a wide range of latent variable models under unsupervised and discriminative settings.

It is based on the method of moments, and involves efficient methods for spectral decomposition of low order observed moments (typically up to fourth order). Unsupervised learning of latent variable models such as topic models, hidden Markov models, Gaussian mixtures and network community models is challenging and maximum likelihood estimation of these models is NP-hard. In contrast, we prove that consistent estimation is possible using efficient spectral methods for decomposition of the moment tensors. We establish that the tensor method has low computational and sample complexities. In practice, these methods are fast to implement and embarrassingly parallel, and are thus, scalable to large scale datasets.

Recently, we have provided novel spectral-based approaches for learning discriminative latent variable models such as multi-layer feedforward neural networks and mixtures of classifiers. The moment tensors we construct are based on the label and higher order score functions of the input, in contrast to moments based on the raw input.
The score functions characterize local variation of the probability density function of the input. Thus, incorporating features based on the generative input model is the key ingredient for learning discriminative latent variable models.

Profile: Anima Anandkumar is a faculty member at the EECS Dept. at U.C.Irvine since August 2010. Her research interests are in the area of large-scale machine learning and high-dimensional statistics. She received her B.Tech in Electrical Engineering from IIT Madras in 2004 and her PhD from Cornell University in 2009. She has been a visiting faculty at Microsoft Research New England in 2012 and a postdoctoral researcher at MIT between 2009-2010. She is the recipient of the Alfred.P. Sloan Fellowship, Microsoft Faculty Fellowship, ARO Young Investigator Award, NSF CAREER Award, IBM Fran Allen PhD fellowship, thesis award from ACM SIGMETRICS society, and paper awards from the ACM SIGMETRICS and IEEE Signal Processing societies.

Research Interests: Her research focus is in the area of inference and learning of probabilistic

graphical models and latent variable models. Broadly she is interested in machine learning, high-dimensional statistics, tensor methods, statistical physics, information theory and signal processing.

Dr Philippe Rigollet, Princeton University

Multistage Bandits

Speaking: 8th January, 12:00pm

Abstract: Motivated by practical applications, chiefly clinical trials and web design optimization, we study the regret achievable for stochastic multi-armed bandits under the constraint that the employed policy must function in a small number of stages. Our results show that a very small number of stages gives already close to minimax optimal regret bounds and we also evaluate the number of trials in each stage. [Joint work with: S. Chassang, V. Perchet and E. Snowberg].

Profile: Philippe Rigollet received his Ph.D. in mathematics from University Paris 6 where he studied in the Laboratoire de Probabilites under the supervision of Alexandre Tsybakov. He then moved to Georgia Tech as a Post-Doc working with Vladimir Koltchinskii in the School of Mathematics. In 2008, he joined the faculty in the department of Operations Research and Financial Engineering at Princeton University as an assistant professor. He received a Berkeley-France fund in 2006 and an NSF CAREER award in 2011.

Research Interests: Philippe has developed new tools for the theory of aggregation, which allows a better understanding of finite sample properties of stochastic optimization and sparse prediction procedures for example. This research is at the intersection of Statistics, Machine Learning and Optimization. More recently, Philippe has been interested in understanding the statistical limitations of learning under computational constraints.

Prof Chenlei Leng, Warwick University

High-dimensional Ordinary Least-squares Projector for Screening Variables

Speaking: 9th January, 09.50am

Abstract: Variable selection is a challenging issue in many statistical applications when the number of predictors p far exceeds the number of observations n. In this ultra-high dimensional setting, Fan and Lv (2008) introduced the sure independence screening (SIS) procedure that can significantly reduce the dimensionality while preserving the true model with overwhelming probability, before a refined second stage analysis. However, the aforementioned sure screening property strongly relies on the assumption that the important variables in the model should have large marginal correlations with the response, which rarely holds in reality. Motivated by these concerns, we propose a novel and simple screening technique called the high-dimensional ordinary least-squares projector (HOLP) for high dimensional features. We show that HOLP possesses the sure screening property and gives consistent variable selection without the strong assumption, and has a low computational complexity. Simulation study shows that HOLP performs competitively compared to many other marginal correlation based methods including (iterative) SIS, forward regression and

tilting. An application to a mammalian eye disease data illustrates the attractiveness of HOLP. This is joint work with Xiangyu Wang in Duke.

Profile: Chenlei Leng joined Warwick as a professor of statistics in 2013. He is interested in developing statistical models for analysing small and big datasets. He received his bachelor's degree in mathematics from USTC, China and PhD in statistics from the University of Wisconsin-Madison. He is an elected member of the International Statistical Institute and currently serves as an associate editor of the Journal of the Royal Statistical Society, Series B.

Research Interests: Chenlei's research interests span the areas of high dimensional data analysis, model selection, semi- and non-parametric statistics, longitudinal data analysis, quantile regression, and applied statistics.

Dr Peter Richtarik, University of Edinburgh

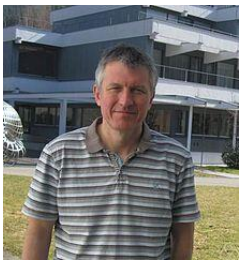
Randomized dual coordinate ascent with arbitrary sampling

Speaking: 9th January, 09.50am

Abstract: We design a novel randomized dual coordinate ascent method (QUARTZ) for minimizing regularized empirical loss; that is, for minimizing the average of a very large number of smooth loss functions (each corresponding to a data vector/example) plus a strongly convex regularizer. The method operates by updating a random set of coordinates of the dual problem at every iteration. We allow for an arbitrary probability law (sampling) to govern the choice of the set of coordinates and show how this enters the complexity bound. For specific choices of the sampling we obtain variants of QUARTZ closely related to SDCA with importance sampling, mini-batch SDCA and distributed SDCA. In a statistically interesting regime for the choice of the regularization parameter, we obtain a linear speedup up to the square root of the number of coordinates (examples). However, our method also enjoys further data-dependent speedup, driven by sparsity and/or spectral properties of the data. Lastly, unlike traditional analysis of SDCA, in our analysis we directly control the decrease of the duality gap. Time permitting, I will comment on other variants of coordinate descent.

Profile: Peter Richtarik is a tenured Assistant Professor of Optimization at the School of Mathematics, University of Edinburgh. He obtained his PhD in Operations Research from Cornell University in 2007 and prior to his current appointment has held a postdoctoral fellowship position (2007-2009) at the Centre for Operations Research and Econometrics, Catholic University of Louvain, Belgium.

Research Interests: Dr. Richtarik's research interests lie in the area of big data optimization, parallel algorithms, gradient methods and machine learning. In a series of recent papers he has developed a general complexity theory of serial and parallel coordinate descent methods. These highly efficient codes are publicly available in the package ACDC. Dr. Richtarik's research is supported by several grants, most notably a $1.1 million grant from the Engineering and Physical Sciences Research Council (EPSRC) for studying optimization algorithms for analysing big digital resources.



Prof Chris Holmes, University of Oxford

General Bayesian updating for complex applications
Speaking: 9th January, 11.10am

Abstract: Data sets are increasing in size and modelling environments are becoming more complex. This presents great opportunities for Bayesian statistics but also major challenges, perhaps the greatest of which is the requirement to define the true sampling distribution, or joint likelihood, for the whole data generator $f_0(x)$, regardless of the study objective. So even if the task is inference for a low-dimensional statistic Bayesian analysis is required to model the complete data distribution and, moreover, assume that the model is ``true''. In this talk we present a coherent procedure for general Bayesian inference which is based on the use of loss functions to connect information in data to parameters of interest. The updating of a prior belief distribution to a posterior then follows from a decision theoretic foundation involving cumulative loss functions and a requirement for coherency. Importantly, the procedure coincides with Bayesian updating when a true likelihood is known, yet provides coherent subjective inference in much more general settings. We demonstrate the approach on examples including model-free general Bayesian co-clustering of time series.

Profile: Prof. Chris Holmes moved to Oxford from Imperial College London in February 2004. At Imperial College he studied for his doctorate in Bayesian statistics, investigating novel nonlinear pattern recognition methods. This was followed by a post-doctoral position and then a lectureship at Imperial. Previous to this he worked in industry for a number of years researching in scientific computing, developing techniques for real-time pattern recognition models in defense and SCADA (Supervisory Control and Data Acquisition) systems. His current research is focused on applications and statistical methods development in the genomic sciences and genetic epidemiology. He holds a programme leaders grant in Statistical Genomics from the Medical Research Council.

Research Interests: Bayesian Statistics, Stochastic Simulation, Markov chain Monte Carlo, Pattern Recognition, Spatial Statistics, Statistical Genetics, Statistical Genomics, Genetic Epidemiology.

Prof. Qiwei Yao, London School of Economics

Large Segmenting Multiple Time Series by Contemporaneous Linear Transformation: PCA for Time Series
Speaking: 9th January, 12.00pm

Abstract: We seek for a contemporaneous linear transformation for a p-variate time series such that the transformed series is segmented into several lower-dimensional subseries, and those subseries are uncorrelated with each other both contemporaneously and serially. The method may be viewed as an extension of principal component analysis (PCA) for multiple time series. Technically it also boils down to an eigenanalysis for a positive definite matrix. When p is large, an additional step is required to perform a permutation in terms of either maximum cross-correlations or FDR based on multiple tests. The asymptotic theory is established for both fixed p and diverging p when the sample size n tends to infinity. Numerical experiments with both simulated and real datasets indicate that the proposed method is an

effective initial step in analysing multiple time series data, which leads to substantial dimension-reduction in modelling and forecasting high-dimensional linear dynamical structures. The method can also be adapted to segment multiple volatility processes.

Profile: Professor of Statistics at London School of Economics. He is a Fellow of both the Institute of Mathematical Statistics and the American Statistical Association, and also a elected member of the International Statistical Institute.

Research Interests: Time series analysis, Nonparametric regression, Dimension reduction and factor modelling, Spatio-temporal modelling, Financial econometrics

Oral Presentations

All oral presentations will take place in the JZ Young Theatre

Click titles to view abstracts online.

Estimating Latent Variable Densities for Exchangeable Network Models

Sharmodeep Bhattacharyya et al.

Presenting on 8th January, 14.00-15.10 session

Distributed Monte Carlo testing

Patrick Rubin-Delanchy et al.

Presenting on 8th January, 14.00-15.10 session

Nonparametric Eigenvalue-Regularized Precision or Covariance Matrix Estimator

Clifford Lam et al.

Presenting on 8th January, 14.00-15.10 session

Consistent Vector-valued Distribution Regression

Zoltán Szabó et al.

Presenting on 8th January, 14.00-15.10 session

Big Hypothesis Testing with Kernels

Dino Sejdinovic et al.

Presenting on 9th January, 14.00-14.45 session

Unbiased Posterior Expectations for Big Data

Heiko Strathmann et al.

Presenting on 9th January, 14.00-14.45 session

Statistical calculations at scale using machine learning algorithms and emulation

Daniel John Lawson & Niall Adams

Presenting on 9th January, 14.00-14.45 session

Poster Presentation Abstracts

Posters will be presented 15.10-17.00 on Thursday 8th January in the JBS Haldane Hub, but can be browsed at leisure throughout the conference.
The poster boards will be numbered as follows:
Click titles to view abstracts online.

1. Community Detection in Massive Populations of Spatial Point Processes with Ecological Applications
James S. Martin et al.

2. Liquidity commonality does not imply liquidity resilience commonality: A functional characterisation for ultra-high frequency cross-sectional LOB data
Efstathios Panayi et al.

3. Testing for Network Community Structure
Beate Franke et al.

4. Network Motif Properties Under the Random Dot-Product Graph Model
P-A. G. Maugis et al.

5. Convulutional Data for Deep Audio Learning
Hazrat Ali et al.

6. Reconstructing Sparse High-Dimensional Signals Using Prior Information
João F.C. Mota et al.

7. Herding in Infinite Dimensional RKHSs
S. Grunewalder et al.

8. Approximating the normalizing constant in large, sparse graphical models
Helen Ogden et al.

9. Optimizing Processes in Visual Data Analysis through Progressive Computations
Cagatay Turkay & Helwig Hauser

10. Estimation of classical defined reliability from large data sample
Kangrui Wang et al.

11. New State-Space Based Method of Estimation of Model Parameter Vector given Large Data Sets
Dalia Chakrabarty et al.

12. How to measure the quality of financial tweets
Paola Cerchiello & Paolo Giudici

13. *Learning Causal Interaction Network of Multivariate Hawkes processes*
Negar Kiyavash et al.

14. *Correlation Clustering Revisited: Fitting the Clusters*
Olgica Milenkovic & Gregory Puleo

15. *Statistical and computational trade-offs in estimation of sparse principal components*
Tengyao Wang et al.

16. *Massively distributed Gaussian Process Classification*
Owen Thomas et al.

17. *Asymptotic behaviour of likelihood-based network community structure analysis*
Yue Hui et al.

18. *Edge-Exchangeable Networks*
Tran Viet Long et al.

19. *Quality Preserving Databases: Statistically Sound and Efficient Use of Public Databases for an Infinite Sequence of Tests*
Saharon Rosset et al.