Report from Dagstuhl Seminar 11451

Data Mining, Networks and Dynamics

Edited by
Lars Eldén[1] and Andreas Frommer[2]

1 **Linköping University, SE,** `lars.elden@liu.se`
2 **Universität Wuppertal, DE,** `Frommer@math.uni-wuppertal.de`

Abstract

In many areas one needs to extract relevant information from signals generated by dynamical systems evolving on networks with a configuration that itself evolves with time. Such problems occur e.g. in surveillance systems for security, in early warning systems of disasters such as earthquakes, hurricanes and forest fires, in collaborative filtering, in search engines for evolving data bases and in reputation systems based on evolving votes. In each of those examples one wants to extract relevant information from an evolving database. Information science is one of the most expansive scientific areas nowadays, much due to the vast amount of information that is available on the Internet, and and the rapid growth of e-business. There are several open problems in these areas of research and even partial answers would have an important impact. There is a pressing need to make progress in analysis tools and in algorithms for such complex tasks. The purpose of the Dagstuhl Seminar 11451 "Data Mining, Networks and Dynamics" was to bring together a diverse community of researchers working in different aspects of this exciting field.

Seminar 06.–11. November, 2011 – www.dagstuhl.de/11451
1998 ACM Subject Classification G 1.3 Numerical linear algebra, G 1.6 Optimization, H 3.3 Information search and retrieval
Keywords and phrases dynamic networks, data mining, optimization on manifolds, iterative methods, graph algorithms, network models in applications
Digital Object Identifier 10.4230/DagRep.1.11.23
Edited in cooperation with Karsten Kahl

1 Executive Summary

Lars Eldén
Andreas Frommer

The main focus of the Dagstuhl Seminar 11451 "Data Mining, Networks and Dynamics" was the theory and the computational aspects of methods for the extraction of information in evolving networks and recent advances in algorithms for related linear algebra problems. The ever increasing size of data sets and its influence on algorithmic progress appeared as a recurrent general theme of the seminar. Some of the participants are experts in the modeling aspects, some are focusing on the theoretical analysis, and some are more directed toward software development and concrete applications. There was a healthy mix of participants of different age and academic status, from several PhD students and post-docs to senior researchers. As the subject has immediate and important applications, the seminar had some attendants with an industrial background (Yahoo and Google). Those participants also contributed greatly by introducing the academic researchers to new applications. The seminar also had several academic participants from application areas, who presented recent advances and new problems. Apart from well-known applications in social networks, search engines and biology approached from a different angle (Alter, Bast, Groh, Harb, Stumme), new applications were presented such as network methods in epidemiology (Poletto), human contact networks (Yoneki), credibility analysis of Twitter postings (Castillo), structure determination in cryo-electron microscopy (Boumal).

On the methodological side, new methods from graph theory and corresponding numerical linear algebra were presented (Bolten, Brannick, Delvenne, Dhillon, Gleich, Ishteva, Kahl, Savas, Szyld). Of particular interest is the extraction of information from extremely large graphs. Given that the class of multigrid methods is standard for solving large sparse matrix problems derived from partial differential equations, it is very natural that these methods should be tried for graph problems. In this direction, new adaptive algebraic multigrid methods for obtaining the stationary distribution of Markov chains were presented.

It has recently been recognized that optimization on manifolds (Boumal, Sepulchre) is a powerful tool for solving problems that occur in information sciences. As real world data are often organized in more than two categories, tensor methods (Alter, De Lathauwer, Eldén, Khoromskij, Sorber) are becoming a hot topic, and the talks showed that the techniques are developing so that now large problems can be treated. Tensor methods have been used for a long time for extremely large problems arising in physics. It has been conjectured that those techniques can be used also for problems in information science. Preliminary discussions along those lines took place. It is also interesting that some tensor computations are based on manifold optimization. Thus there were interesting discussions on the interplay between these areas.

The atmosphere of the meeting was very informal and friendly. During and after the talks lively discussions took place that also continued after dinner. Although it is too early to tell whether the seminar lead to new collaborations between the participants, some preliminary contacts were made. An open problem in spectral partitioning was raised (Eldén) and a preliminary solution was suggested (Gleich, Kahl).

The participants of this seminar had a chance to interact with the Dagstuhl seminar 11542 "Analysis of Dynamic Social and Technological Networks" held at the same time. Indeed, the Thursday morning session was arranged as a common session between both seminars, focusing at introducing the different methodological approaches to all particpants.

2 Table of Contents

3 Overview of Talks

3.1 Discovery of Mechanisms and Prognosis of Cancers from Matrix and Tensor Modeling of Large-Scale Molecular Biological Data

Orly Alter (University of Utah – Salt Lake City, US)

In my Genomic Signal Processing Lab, we believe that future discovery and control in biology and medicine will come from the mathematical modeling of large-scale molecular biological data, such as DNA microarray data, just as Kepler discovered the laws of planetary motion by using mathematics to describe trends in astronomical data [1]. In this talk, I will first describe novel generalizations of the matrix and tensor computations that underlie theoretical physics (e.g., [2, 3]), that we are developing for comparison and integration of multiple genomic datasets recording different aspects of, e.g., cell division and cancer. Second, I will describe our recent experiments [4] that verify a computationally predicted genome-wide mode of regulation [5, 6], and demonstrate that singular value decomposition (SVD) and higher-order SVD (HOSVD) modeling of DNA microarray data can be used to correctly predict previously unknown cellular mechanisms. Third, I will show that mode-1 HOSVD modeling of rRNA sequence alignments suggests a new way of looking at evolution as a composition of changes rather than a hierarchy of changes, and might be used to predict evolutionary mechanisms, i.e., evolutionary pathways and the underlying structural changes that these pathways are correlated, possibly even coordinated with [7]. Last, I will describe the computational prognosis of brain cancers by using generalized SVD to compare global DNA copy numbers in patient-matched normal and tumor samples from the Cancer Genome Atlas [8, 9].

References

1 O. Alter, PNAS 103, 16063–16064 (2006); http://dx.doi.org/10.1073/pnas.0607650103
2 O. Alter, P. Brown, and D. Botstein, PNAS 100, 3351–3356 (2003); http://dx.doi.org/10.1073/pnas.0530258100
3 S. P. Ponnapalli, M. Saunders, C. Van Loan, and O. Alter, PLoS One 6 (12): e28072 (2011); http://dx.doi.org/10.1371/journal.pone.0028072
4 L. Omberg, J. Meyerson, K. Kobayashi, L. Drury, J. Diffley, and O. Alter, MSB 5, 312 (2009); http://dx.doi.org/10.1038/msb.2009.70
5 O. Alter, and G. Golub, PNAS 101, 16577–16582 (2004); http://dx.doi.org/10.1073/pnas.0406767101
6 L. Omberg, G. Golub, and O. Alter, PNAS 104, 18371–18376 (2007); http://dx.doi.org/10.1073/pnas.0709146104
7 C. Muralidhara, A. Gross, R. Gutell, and O. Alter, PLoS One 6 (4), e18768 (2011); http://dx.doi.org/10.1371/journal.pone.0018768
8 C. Lee, and O. Alter, 60th Annual Meeting of the American Society of Human Genetics (Washington DC, November 2–6, 2010).
9 C. Lee, B. Alpert, P. Sankaranarayanan, and O. Alter, PLoS One 7 (1), e30098 (2012); http://dx.doi.org/10.1371/journal.pone.0030098

3.2 Semantic Full-Text Search

Hannah Bast (Universität Freiburg, DE)

Joint work of Bast, Hannah; Bäurle, Florian; Buchhold, Björn; Haußmann, Elmar

Semantic full-text search combines classical full-text search with search in knowledge databases aka ontologies. An example query would be to find all female german professors working in a particular research area. Google will not be of much help for such a query, unless there is a web page with exactly that contents. And a general-purpose ontology is unlikely to contain all the information required to answer this query.

I will explain this in more detail and how it differs from other approaches claiming to be "semantic". I will point out some of the underlying algorithmic problems, especially with respect to scalability, and some of our current solutions. And I will demonstrate our current prototype. All this is fairly recent and, as I like to think, exciting work in progress.

3.3 Multigrid methods for circulant and Toeplitz matrices and their application to queueing networks

Matthias Bolten (Universität Wuppertal, DE)

Toeplitz and circulant preconditioners are used to speed up Krylov subspace methods in queueing applications. While systems involving circulant matrices can be solved in near-optimal, i.e. $O(N \log N)$, time, this is not the case for Toeplitz matrices. Nevertheless, circulant preconditioners often do not provide the required performance. In this talk we will introduce multigrid methods for Toeplitz and circulant matrices that overcome some of these limitations. The convergence theory, including results for aggregation and smoothed aggregation, will be presented and their application to queueing networks will be discussed.

3.4 Synchronization of rotations on a graph

Nicolas Boumal (UC Louvain-la-Neuve, BE)

Joint work of Boumal, Nicolas; Singer, Amit; Absil, Pierre-Antoine
Main reference A. Singer, "Angular synchronization by eigenvectors and semidefinite programming, Applied and Computational Harmonic Analysis," Vol. 30, Issue 1, January 2011, pp. 20–36.
URL http://dx.doi.org/10.1016/j.acha.2010.02.001

I will start by informally describing two application problems. These will motivate the study of an estimation problem in which the quantities to estimate are N rotations and the measurements to help us do so are noisy observations of relative rotations. The measurements induce a graph which consists of N nodes (each associated to one of the unknown rotations) and such that each edge corresponds to a noisy measurement of the relative rotation between the two incident nodes. I will then state a few results we have obtained on the theoretical and algorithmic aspects of this problem, and try to gradually transform the presentation into a group discussion.

3.5 Algebraic Distance based Coarsening with Application to Clustering

James Brannick (Pennsylvania State University, US)

In this talk I introduce an algebraic distance technique for multilevel graph organization. The ideas build on recent work concerning development of algebraic multigrid solvers for linear systems, in particular elliptic partial differential equations.

3.6 Information Credibility on Twitter

Carlos Castillo (Yahoo Research – Barcelona, ES)

Joint work of Castillo, Carlos; Mendoza, Marcelo; Poblete, Bárbara
Main reference C. Castillo, M. Mendoza, B. Poblete, "Information credibility on twitter," in Proc. of the 20th Int'l Conf. on World Wide Web (WWW'11), pp. 675–684, ACM, New York, NY, USA, 2011.
URL http://dx.doi.org/10.1145/1963405.1963500

We analyze the information credibility of news propagated through Twitter, a popular microblogging service. Previous research has shown that most of the messages posted on Twitter are truthful, but the service is also used to spread misinformation and false rumors, often unintentionally.

On this paper we focus on automatic methods for assessing the credibility of a given set of tweets. Specifically, we analyze microblog postings related to "trending" topics, and classify them as credible or not credible, based on features extracted from them. We use features from users' posting and re- posting ("re-tweeting") behavior, from the text of the posts, and from citations to external sources.

We evaluate our methods using a significant number of human assessments about the credibility of items on a recent sample of Twitter postings. Our results shows that there are measurable differences in the way messages propagate, that can be used to classify them automatically as credible or not credible, with precision and recall in the range of 70% to 80%.

3.7 Block Term Decompositions and Block Component Analysis

Lieven De Lathauwer (K.U. Leuven, BE)

The Canonical Polyadic Decomposition (CPD) of a higher-order tensor is rank-revealing; it writes the tensor as a minimal sum of rank-1 terms. The Tucker decomposition finds the column space, row space, ... of a higher-order tensor and is multilinear rank-revealing.

The recently introduced Block Term Decompositions (BTD) generalize these two basic decompositions. They write a higher-order tensor as a sum of terms that have low multilinear rank. BTDs enable a new approach to factor analysis and blind source separation, which we call Block Component Analysis. We discuss links with Canonical Polyadic Analysis (CPA), Independent Component Analysis (ICA), Pareto analysis and compressed sensing. We show several illustrative examples.

3.8 Dynamics on networks for communities and centralities

Jean-Charles Delvenne (UC Louvain, BE)

Joint work of Delvenne, Jean-Charles; Yaliraki, Sophia; Barahona, Mauricio
Main reference J.-C. Delvenne, S. Yaliraki, M. Barahona, "Stability of graph communities across time scales," in Proc. of the National Academy of Sciences, 107:29, pp. 12755–12760.
URL http://dx.doi.org/10.1073/pnas.0903215107

Dynamical systems taking place on networks, such as opinion dynamics, synchronisation, consensus or random walks, reveal a lot about their structure.

In particular we show, through a dynamical reinterpretation of well-known concepts, how centrality measures ('which nodes are most central in the network?') and community detection quality functions ('which group of nodes are tightly related to one another?') are intimately related. The dynamical interpretation allows to design new centrality or community quality measures tailored for every particular application.

This is joint work with Sophia Yaliraki, Mauricio Barahona, Renaud Lambiotte, Anne-Sophie Libert [1, 3, 2].

References

1 J.-Ch. Delvenne, S.N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proc. Natl Acad. Sci. U. S. A.*, 107:12755–12760, 2010.
2 J.C. Delvenne and A.S. Libert. Centrality measures and thermodynamic formalism for complex networks. *Physical Review E*, 83(4):046117, 2011.
3 R. Lambiotte, J.-C. Delvenne, and M. Barahona. Dynamics and modular structure in networks. *arXiv:0812.1770*, 2008.

3.9 Fast and memory-efficient low rank approximation of massive graphs

Inderjit Dhillon (University of Texas – Austin, US)

Joint work of Dhillon, Inderjit; Savas, Berkant

Social network analysis requires us to perform a variety of analysis tasks, including summarization and prediction, on massive graphs. In this talk, I will present a fast and memory-efficient procedure called clustered low rank matrix approximation for massive graphs. The procedure involves a fast clustering of the graph followed by approximation of each cluster separately using existing methods, e.g. the singular value decomposition, or stochastic algorithms. The clusterwise approximations are then extended to approximate the entire graph.

This approach has several benefits: (1) important structure of the graph is preserved due to the clustering; (2) accurate low rank approximations are achieved; (3) the procedure is efficient both in terms of computational speed and memory usage. Further, we generalize stochastic algorithms into the clustered low rank approximation framework and present theoretical bounds for the approximation error. Finally, a set of experiments show that our methods outperform existing dimensionality reduction algorithms on massive social networks.

3.10 Simultaneous clustering of time-evolving graphs

Lars Eldén (Linköping University, SE)

Spectral partitioning is a well-known method for the clustering and partitioning of graphs. The Fiedler vector of the graph Laplacian is used to reorder the vertices of the graph. This makes it possible to find a partitioning that is rather close to being optimal. Alternatively the Fiedler vector can be computed using the adjacency matrix of the graph. We discuss the generalization of this method to tensors, which made up from a graph that evolves with time. Our approach is based on a generalization of the Rayleigh quotient characterization of the Fiedler vector. The solution of the tensor Rayleigh quotient maximization problem is computed using a tensor-Krylov method. A couple of examples are given, with a clustering of a sparse tensor obtained from the Enron email data set.

3.11 Techniques for local and global centrality estimation

David F. Gleich (Purdue University, US)

Joint work of Bonchi, Francesco; Esfandiar, Pooya; Gleich, David F.; Greif, Chen; Lakshmanan, Laks V. S.
Main reference F. Bonchi, P. Esfandiar, D.F. Gleich, C. Greif, L.V.S. Lakshmanan, "Fast matrix computations for pair-wise and column-wise Katz scores and commute times," arXiv:1104.3791v1 [cs.SI]; to appear in J. Internet Mathematics.
URL http://arxiv.org/abs/1104.3791v1

In this talk, I will discuss local algorithms for centrality scores, with a particular focus on the Katz score. I will also discuss recent work on how to utilize overlapping partitions of a graph to reduce the communication in the computation of a PageRank or Katz vector. Both of these discussions will focus on static graphs. In the last third of the talk, I present thoughts on adapting these algorithms to time-dependent problems and show some curious properties of graph spectra I cannot explain.

3.12 Social Situation Detection on the Basis of Interaction Geometries and Contextual Privacy Management as an Application

Georg Groh (TU München, DE)

Joint work of Groh, Georg; Fuchs, Christoph; Lehmann, Alexander; Birnkammerer, Stefan

Social Situation models are special forms of episodic short term social context that can be detected with the help of mobile sensors and used for applications in a (distributed, mobile) Social Networking scenario. The talk centers around methods to detect Social Situations by analyzing geometry of interaction (e.g. shoulder angles, interpersonal distances), audio or other sensor data sources. A special approach for privacy management is introduced as an application using spatio-temporal contexts as well as short term and long term social contexts as building blocks.

References

1 Groh, G., Fuchs, C., and Lehmann, A. (2011d). Combining evidence for social situation detection. *Proc. IEEE Socialcom'11, Boston, USA.*
2 Groh, G. and Lehmann, A. (2011). Deducing evidence for social situations from dynamic geometric interaction data. *Int. J. Social Computing and Cyber-Physical Systems (to appear).*
3 Groh, G. and Birnkammerer, S. (2011). Privacy and information markets: Controlling information flows in decentralized social networking. *Proc. IEEE Socialcom'11, Boston, USA.*

3.13 Extracting, Searching and Discovering the Semantics of Tables on the Web

Boulos Harb (Google – New York, US)

Joint work of Das Sarma, Anish; Halevy, Alon; Harb, Boulos; Lee, Hongrae; Wu, Fei; Yu, Cong
Main reference P. Venetis, A. Halevy, J. Madhavan, M. Pasca, W. Shen, F. Wu, G. Miao, C. Wu, "Recovering semantics of tables on the web," in Proc. VLDB Endow., 4 (2011), pp. 528–538.
URL http://dl.acm.org/citation.cfm?id=2002939

The Web provides a large corpus of over 100 million high-quality tables that offer a rich source of structured information on an extensive variety of topics. However, the semantics of each of these HTML tables are rarely explicit in the table itself, and the meaning is often derived from text surrounding the table. Header rows exist in only a few cases and even when they do, the attribute names are not standardized and typically of not much use. Without understanding the attributes and entities a table represents it is difficult to better leverage its content (in search for example) or join or union it with other tables. We have built a scalable system for automatically recovering the semantics of tables on the Web. Given the scale, breadth and heterogeneity of this corpus, we cannot rely on hand-coded domain knowledge to associate semantics with each table. To recover semantics of tables, we leverage a (noisy) database of class labels and relationships automatically extracted from the Web. We add annotations to a table describing the sets of entities and attributes (represented in the rows and columns respectively) the table is modeling, and the binary relationships between the attribute and the subject columns. Our system is extensible to any corpus of tables, and we leverage it to build a search over public Google Fusion Tables (http://www.google.com/fusiontables).

3.14 Low rank nonnegative approximations

Mariya Ishteva (Georgia Institute of Technology, US)

Joint work of Ishteva, Mariya; De Lathauwer, Lieven; Park, Haesun

Nonnegative matrix factorization (NMF) is an (approximate) factorization of a nonnegative matrix into two nonnegative matrices with low rank. It has proven to be a successful dimensionality reduction and factor analysis technique.

We consider a related problem where we do not impose nonnegativity on each factor but only on their product. This can be thought of as a low rank nonnegative approximation of a given matrix. In many applications, e.g., image compression, the NMF constraints are too strong if we only need a nonnegativity in the product. Our approximation combines advantages of both NMF and singular value decomposition (SVD). The approximation error is theoretically not larger than that of NMF (and not smaller than that of SVD) and in many experiments we observed that it is much closer to that of SVD especially when the reduced dimension is small. On the other hand, because of the nonnegativity constraints, the approximation has potentially better interpretability than the one obtained from SVD.

We propose an active-set based algorithm for the computation of the factorization using the Frobenius norm as a distance measure. We also show how the problem of approximating a higher-order nonnegative tensor by a low multilinear rank nonnegative tensor can be reduced to its matrix counterpart.

3.15 Bootstrap Algebraic Multigrid for Markov Chains

Karsten Kahl (Universität Wuppertal, DE)

This talk is devoted to the development of an adaptive algebraic multilevel method for computing steady state vectors of Markov chains. We present an efficient bootstrap algebraic multigrid method for this task.

In our proposed approach, we employ a multilevel eigensolver, with interpolation built using ideas based on compatible relaxation, algebraic distances, and least squares fitting of test vectors. Our adaptive variational strategy for computation of the steady state vector of a given Markov chain is then a combination of this multilevel eigensolver and an associated additive multilevel preconditioned correction process. We show that the bootstrap algebraic multigrid eigensolver by itself can efficiently compute accurate approximations to the steady state vector. An additional benefit of the bootstrap approach is that it yields an accurate interpolation operator for many other eigenmodes. This in turn allows for the use of the resulting multigrid hierarchy as a preconditioner to accelerate the GMRES iteration for computing an additive correction equation for the approximation to the steady state vector. Unlike other existing multilevel methods for Markov chains, our method does not employ any special processing of the coarse-level systems to ensure that stochastic properties of the fine-level system are maintained there. The proposed approach is applied to a range of test problems, involving non-symmetric M-matrices arising from stochastic matrices, showing promising results.

3.16 Tensor numerical methods for multi-dimensional PDEs: application to parabolic equations

Boris N. Khoromskij (MPI für Mathematik in den Naturwissenschaften, DE)

Modern methods of rank-structured tensor decomposition allow an efficient data-sparse separable approximation of multivariate functions and operators, providing linear complexity scaling in the dimension.

The successful applications of tensor methods include the challenging high-dimensional problems arising in material sciences, bio-science, stochastic modeling, signal processing, machine learning and data mining, quantum computing, many-body dynamics etc.

The recent quantized tensor train (QTT) technique is proved to provide the super-compressed representation for a class of multidimensional data-arrays of size N^d with the log-volume storage complexity $O(d \log N)$.

The approach is based on data quantization to the $2 \times 2 \times ... \times 2$-format in the highest possible dimension $D = d \log N$.

Combined with the efficient multi-linear QTT-algebra, this method opens the way to the profound deterministic numerical simulation of high-dimensional PDEs getting rid of the "curse of dimensionality" and restrictions on the grid size N.

However, in some cases, the approach is limited by the "curse of ranks" characterizing the essential entanglement in the system of interest.

We discuss the asymptotically optimal low QTT-rank representations for a class of multivariate functions and operators (matrices), focusing on multivariate potentials, d-dimensional elliptic Green's function and parabolic solution operator.

Numerical illustrations in electronic structure calculations (the Hartree-Fock equation), quantum molecular dynamics (the molecular Schrödinger equation) and stochastic modeling for dynamics of liquids (the Fokker-Planck equation), are presented.

3.17 Multiscale Networks and the Spatial Spread of Infectious Diseases

Chiara Poletto (ISI Foundation, IT)

Human mobility and interactions represent key ingredients in the spreading dynamics of an infectious disease. The flows of traveling individuals form a network characterized by complex features, such as strong topological and traffic heterogeneities, that unfolds at different temporal and spatial scales, from short ranges to the global scale. Computational models can be developed that integrate detailed network structures based on demographic and mobility data, in order to simulate the spatial evolution of an epidemic. Focusing on the 2009 H1N1 influenza pandemic as a paradigmatic example, these approaches allow quantifying the effects of travel restrictions in delaying and controlling the epidemic spread. In addition, simplified model frameworks can be solved providing the assessment of the interplay between individual mobility and epidemic dynamics, under different conditions of heterogeneities characterizing human mobility.

3.18 Supervised Link Prediction Using Multiple Sources

Berkant Savas (Linköping University, SE)

Joint work of Savas, Berkant; Lu, Zhengdong; Tang, Wei; Dhillon, Inderjit S.
Main reference Z. Lu, B. Savas, W. Tang, I. S. Dhillon, "Link prediction using multiple sources of information," in Proc. of the IEEE Int'l Conf. on Data Mining (ICDM'10), pp. 923–928, 2010.
URL http://dx.doi.org/10.1109/ICDM.2010.112

Link prediction is a fundamental problem in social network analysis and modern-day commercial applications. Most existing research approaches this problem by exploring the topological structure of a social network using only one source of information. However, in many application domains, in addition to the social network of interest, there are a number of auxiliary social networks and/or derived proximity networks available. In this talk we will present: (1) a supervised learning framework that can effectively and efficiently learn the dynamics of social networks in the presence of auxiliary networks; (2) a feature design scheme for constructing a rich variety of path-based features using multiple sources, and an effective feature selection strategy based on structured sparsity. Extensive experiments on three real-world collaboration networks show that our model can effectively learn to predict new links using multiple sources, yielding higher prediction accuracy than unsupervised and single source supervised models.

3.19 Computing with fixed (low) rank matrices: a geometric approach

Rodolphe Sepulchre (University of Liège, BE)

Joint work of Sepulchre, Rodolphe;Bonnabel, Silvère; Meyer, Gilles; Mishra, Bamdev; Absil, PA; Vandereycken, B.
Main reference G. Meyer, S. Bonnabel, R. Sepulchre, "Regression on fixed-rank positive semidefinite matrices: a Riemannian approach," Journal of Machine Learning Research, 12(Feb):593–625, 2011.
URL http://www.montefiore.ulg.ac.be/~meyer/papers.php

The talk is an introduction to a recent computational framework for optimization over the set of fixed rank positive semidefinite matrices. The proposed framework is based on quotient Riemannian geometries that lead to rank-preserving efficient computations in the set of fixed-rank matrices. The field of applications is vast, and the talk will survey recent developments that illustrate the potential of the approach in large-scale computational problems encountered in control, optimization, and machine learning.

3.20 Optimization-based algorithms for the decomposition of a tensor in rank-(Lt,Lt,1) terms

Laurent Sorber (K.U. Leuven, BE)

Joint work of Sorber, Laurent; Van Barel, Marc; De Lathauwer, Lieven

The canonical polyadic and rank-(Lt,Lt,1) block term decomposition (CPD and BTD, respectively) are two closely related tensor decompositions. The CPD is an important tool in psychometrics, chemometrics, neuroscience and data mining, while the rank-(Lt,Lt,1) BTD is an emerging decomposition in signal processing and, recently, blind source separation. We present a decomposition that generalizes these two and develop algorithms for its computation. Among these algorithms are alternating least squares schemes, several general unconstrained optimization techniques, as well as matrix-free nonlinear least squares methods. In the latter we exploit the structure of the Jacobian's Gramian by means of efficient expressions for its matrix-vector product. Combined with an effective preconditioner, numerical experiments confirm that these methods are among the most efficient and robust currently available.

3.21 Data Mining in Online and Offline Social Networks

Gerd Stumme (Universität Kassel, DE)

This talk presents selected results of the analysis of two real-world applications: BibSonomy and Conferator.

BibSonomy (http://www.bibsonomy.org) is a social bookmark and publication sharing system that is run by the Knowledge & Data Engineering Group of the University of Kassel. We will discuss how to obtain a conceptual hierarchical clustering of the folksonomy of the system by means of Formal Concept Analysis, and how an adapted version of PageRank, called FolkRank, will improve the search within the system and allow for the analyis of time series. We will also briefly touch the topics of ontology learning and spam detection.

Conferator (http://conferator.org) gives you the opportunity to personalize the conference schedule, and shows you a history of your social contacts at the conference and provides additional information about participants, such as their homepage, facebook and twitter accounts, contact details (skype, icq, etc.).

After the conference, Conferator will thus enable you to recall your social contacts you had during the conference. We will discuss a variety of approaches for detecting the most central participants of the conference.

3.22 Matrices with Perron-Frobenius Properties

Daniel B. Szyld (Temple University – Philadelphia, US)

Joint work of Szyld, Daniel B.; Elhashash, Abed
Main reference A. Elhashash, D.B. Szyld, "Two characterizations of matrices with the Perron-Frobenius property," Numerical Linear Algebra with Applications, vol. 16, pp. 863–869, 2009.
URL http://www.math.temple.edu/szyld

The Perron-Frobenius theorem guarantees the existence of a nonnegative (or positive) eigenvector corresponding to the dominant eigenvalue of a nonnegative matrix. In this talk, we explore the question: which matrices (not necessarily nonnegative) also have such a vector? We provide combinatorial, algebraic and spectral characterizations of such matrices. We also generalize the concept of M-matrix using these Perron Frobenius properties of matrices.

3.23 Wireless Epidemic in Dynamic Human Contact Networks

Eiko Yoneki (University of Cambridge, GB)

Joint work of Yoneki, Eiko; Hui, P.; Crowcroft, J.
Main reference P. Hui, J. Crowcroft, E. Yoneki, "BUBBLE Rap: Social-based Forwarding in Delay Tolerant Networks," IEEE Transactions on Mobile Computing, vol. 10, no. 11, pp. 1576–1589, 2011.
URL http://dx.doi.org/10.1109/TMC.2010.246

Increasing numbers of mobile computing devices form dynamic time-dependent networks in daily life. We have explored new communication paradigms using dynamic interconnectedness between devices. To reduce the overhead of epidemic routing in such networks, we attempt to uncover a hidden stable network structure such as a social network, consisting of a group of people forming meaningful relationships (community) and centrality nodes, that act as rely nodes for message passing. We have exploited device connectivity traces from the real world for modelling social network structure. By using these traces, we have shown that social network properties are an important aspect of building a routing protocol.

The empirical study of contact networks shares many issues with network-based epidemiology, and our work has been extended towards understanding the epidemic spread of infectious diseases. Capturing human interactions will provide an empirical, quantitative measurement of social mixing patterns to underpin mathematical models of the spread of close-contact diseases. We have extracted multi-spread modes of operation from contact networks.

Participants

- Orly Alter
University of Utah –
Salt Lake City, US
- Hannah Bast
Universität Freiburg, DE
- Matthias Bolten
Universität Wuppertal, DE
- Nicolas Boumal
UC Louvain-la-Neuve, BE
- James Brannick
Pennsylvania State Univ. US
- Carlos Castillo
Yahoo Research – Barcelona, ES
- Lieven De Lathauwer
K.U. Leuven, BE
- Jean-Charles Delvenne
UC Louvain, BE
- Inderjit Dhillon
University of Texas – Austin, US
- Lars Eldén
Linköping University, SE
- Andreas Frommer
Universität Wuppertal, DE
- David F. Gleich
Purdue University, US
- Georg Groh
TU München, DE
- Boulos Harb
Google – New York, US
- Mariya Ishteva
Georgia Inst. of Technology, US
- Karsten Kahl
Universität Wuppertal, DE
- Boris N. Khoromskij
MPI für Mathematik in den
Naturwissenschaften, DE
- Chiara Poletto
ISI Foundation, IT
- Berkant Savas
Linköping University, SE
- Rodolphe Sepulchre
University of Liège, BE
- Alexander J. Smola
Yahoo! Res. – Santa Clara, US
- Laurent Sorber
K.U. Leuven, BE
- Gerd Stumme
Universität Kassel, DE
- Daniel B. Szyld
Temple Univ. – Philadelphia, US
- Eiko Yoneki
University of Cambridge, GB