# The Seventh International Conference on Network Analysis NET 2017

June 22-24, 2017

Laboratory of Algorithms and Technologies for Network Analysis (LATNA), National Research University Higher School of Economics, Nizhny Novgorod, Russia





Internet Access:	
Wi-fi:	HSE
Login:	hseguest
Password:	hsepassword

Thursday, June 22 Room 402 HSE, 136 Rodionova Str.

**09:00 – 09:30** Registration

**09:30 – 09:50** Panos M. Pardalos

Data Mining and Optimization Heuristics for Massive Networks

**09:50 – 10:40** Roberto Battiti

From graphs to communities, cliques, quasi-cliques and related gangs

**10:40 – 11:00** Coffee Break

11:00 – 11:50 Nenad Mladenovic

New community detection criterion

11:50 – 12:30 Session 1 (2 talks)

Yury Malkov

Geometric model of rich-club phenomenon in complex networks

Sergey Makrushin

Developing a model of topological structure formation for power transmission grids based on the analysis of the UNEG

12:30 – 14:00 Lunch Break

14:00 – 14:50 George Michailidis

Regularized estimation and testing for High-Dimensional Multi-Block Vector Autoregressive Models

14:50 – 15:10 Coffee Break

**15:10 – 16:10** Session 2 (3 talks):

Jayeeta Deshmukh

Trade credit Guarantee Networks with Frictionless Benefit

Igor Konnov

Application of market models to network equilibrium problems

Olga Pinyagina, Igor Konnov

Selective bi-coordinate variations for network equilibrium problems with mixed demand

**16:10 – 16:30** Coffee Break

16:30 – 17:50 Session 3 (4 talks):

Alina V. Vladimirova, Alexey L. Ryabinin

Accounting for National Scientific Traditions in Bibliometric Network Analysis:

Mapping Asian and African Studies in Russia

Nataliya Matveeva

Analyzed of co-authorship networks and scientific citation based on Google Scholar Sergei Sidorov

Company Co-Mention Network Analysis

Ivan Grechikhin

Comparison of different methods of graphical models identification

18:00 Reception

Friday, June 23 Room 402 HSE, 136 Rodionova Str.

**09:30 – 10:20** Elena Konstantinova Cayley networks in Computer Science **10:20 – 10:40** Coffee Break 10:40 – 11:30 Ludmila Prokhorenkova Some general results on preferential attachment and clustering coefficient 11:30 – 12:30 Session 4 (4 talks): Alexander Rassadin, Andrey Savchenko Deep neural networks performance optimization via the Distillation the Knowledge **Dmitry Sirotkin** *Polynomial-time solvability of the independent set problem in a certain class of* subcubic planar graphs Dmitry Taletskii On the number of maximal independent sets in complete q-ary trees Sergey Ketkov Evader's models in sequential network interdiction **12:30 – 14:00** Lunch Break 14:00 – 14:50 Andrey Raigorodskii Random graphs and applications 14:50 – 15:10 Coffee Break **15:10 – 16:15** Session 5 (3 talks): Alexander Rubchinskiy Decomposition Complexity of Graphs and Its Applications to Analysis of Socio-Economic Systems Irina Utkina Using modular decomposition technique to solve the maximum clique problem **Dmitry** Ignatov Multimodal Clustering for Community Detection **16:15 – 16:30** Coffee Break **16:30 – 17:45** Session 6 (5 talks) Ilva I. Kurochkin The comparative analysis of big networks with various network topologies Alexander Semenov, Stepan Kochemazov, Dmitry Gorbatenko Computational study of activation dynamics on networks of arbitrary structure Andrey Nelyubin, Vladislav Pidinovski, Mikhail Potapov Methods of criteria importance theory and their software implementation Marina Ananieva Information Propagation Strategies in Online Social Networks Ilia Karpov Weighting Relationship Strength in Online Social Networks

Saturday, June 24 Room 402 HSE, 136 Rodionova Str.

**09:30 – 10:20** Oleg Khamisov Looking for an equilibrium on long-term electricity market under network constraints **10:20 – 10:40** Coffee Break **10:40 – 12:20** Session 4 (5 talks): Alexander Strekalovsky Global search in problem with d.c. constraints Tatiana Gruzdeva Minimization of the average production costs via fractional programming problem Mikhail Batsyn, Alexey Nikolaev Vehicle Assignment in Site-Dependent Vehicle Routing Problems with Split Deliveries Andrei Orlov On a Numerical Solving of Bilevel Pricing Problem in Telecommunication Networks via D.C. Maximization **Dmitry** Ignatov Searching for Maximal Quasi-Bicliques via MIP-Based Models **12:20 – 14:00** Lunch Break **14:00** – **14:50** Oleg Prokopyev Finding Critical Links for Closeness Centrality 14:50 – 15:10 Coffee Break **15:10 – 16:10** Session 5 (3 talks): Mario Rosario Guarracino A distributed supervised learning algorithm for smart grid networks Svetlana Kolesnikova Application of Nonlinear Adaptation on Manifolds for Economic Objects Ivan Kuznetsov Intra-organizational networks in organizations: communication networks as a key to success 16:10 – 16:30 Coffee Break **16:30 – 17:30** Session 6 (5 talks): Sergev Slashchinin The Properties of Optimal Portfolios for Markowitz Problem Anvar Kurmukov Modular structure of brain networks in healthy and diseased brains Nadezhda Kostyakova Analysing Online Recomendation Networks Dmitry Zaytsev, Daria Drozdova Mapping the ontologies, methodological approaches and methods of socio-political

research: application of social network analysis

# Plenary talks

# From graphs to communities, cliques, quasi-cliques and related gangs

#### Roberto Battiti

#### Computer Science Università degli Studi di Trento, LION lab, Italy

The maximum clique (MC) problem in graphs is a paradigmatic combinatorial optimization problem with many relevant applications, including information retrieval, computer vision, and social network analysis. Recent interest includes computational biochemistry, bioinformatics, and genomics. If graphs represent the raw information, clique and quasi-cliques represent efforts to extract human-understandable groupings (clusters) for subsequent processing. The talk reviews the state of the art of the problems, presenting some effective solution methods and open issues.

#### **Cayley networks in Computer Science**

# Elena Konstantinova

#### Sobolev Institute of Mathematics, Russian Federation

In this talk we present recent results on networks based on Cayley graphs which become more and more popular in computer science. The design and analysis of interconnection networks has been one of the main topics of research for the past years. Since 1986, when SIAM International Conference on Parallel Processing was held, Cayley graphs are used as a tool to construct vertex-symmetric interconnection networks. There are many advantages in using Cayley graphs as models for interconnection networks such as vertex-transitivity (the same routing algorithm is used for each vertex), edge-transitivity (every edge in the graph looks the same), hierarchical structure (allows recursive constructions), high fault tolerance (the maximum number of vertices that need to be removed and still have the graph remains connected), small degree and diameter. The progress on Cayley network analysis is shown due to the theory of Cayley graphs including open problems and applications. In particular, we discuss Hamiltonicity problem with relationships to generalized Gray codes, and Diameter problem with applications to computer and biological networks.

# Looking for an equilibrium on long-term electricity market under network constraints

#### Oleg Khamisov

#### Melentiev Energy Systems Institute of the RAS, Russian Federation

We study an oligopolistic model which describes long-term electricity market considering generating capacity expansion under network constraints. Participants of the market are represented by generating companies and we are looking for the Nash-Cournot equilibrium. Each company owns different plants at different nodes of the network. Company strategy is year energy production. Consumers are modeled by a linear year inverse demand function and we explain what ideas were used for constructing this function. Total year generation consists of season working days and weekends generation. If we exclude the network constraints then the model is just an example of the potential game and is reduced to an ordinary convex quadratic optimization problem though of big dimension. Then we check whether it is possible to realize the obtained solution within the previously given network constraints. If not, we suggest to use a bilevel optimization approach, which in general can be reduced to an implicitly given global optimization problem. For solving the latter problem an optimization metodology with nonlinear support functions can be used. Finally we present computational results performed on a part of central energy system of Russia and make some conclusions.

# Regularized estimation and testing for High-Dimensional Multi-Block Vector Autoregressive Models

#### George Michailidis

#### University of Florida, Informatics Institute, USA

Dynamical systems comprising of multiple components originate in many scientific areas. A pertinent example is the interactions between financial assets and macroeconomic indicators, which has been studied at an aggregate level in the macroeconomics literature. A key shortcoming of this approach is that it ignores potential influences from other related components (e.g. Gross Domestic Product) that may exert influence on the system's dynamics and structure and thus produces incorrect results. To mitigate this issue, we consider a multiblock linear dynamic system with Granger-causal ordering between blocks, wherein the blocks temporal dynamics are described by vector autoregressive processes and are influenced by blocks higher in the system hierarchy. We obtain the maximum likelihood estimator for the posited model for Gaussian data in the high-dimensional setting with appropriate regularization schemes. To optimize the non-convex likelihood function, we develop an iterative algorithm with convergence guarantees. We establish theoretical properties of the maximum likelihood estimates, leveraging the decomposability of the regularizers and a careful analysis of the iterates of the proposed algorithm. Finally, we develop testing procedures for the null hypothesis of whether a block "Granger-causes" another block of variables. The performance of the model and the testing procedures are evaluated on synthetic data, and illustrated on a data set involving log-returns of the US S&P100 component stocks and key macroeconomic variables for the 2001–16 period.

#### New community detection criterion

Nenad Mladenovic

#### Mathematical Institute, SANU, Serbia

Complex systems in different areas such as telecommunication, transportation, biology, and social sciences are frequently being presented and modeled at networks. One of the most studied phenomenon in complex networks is community structure, i.e., a division of the network into groups or clusters. Such groups should have a greater density of internal connections and a lower density of external or cut connections. Many criteria have been proposed in the literature to cover obvious fact regarding the homogeneity of entities within cluster and diversity of entities among different clusters at the network. In this talk I will discuss most commonly used criteria, such as modularity and ratio-cut. Moreover, the new criterion for community detection is proposed that, for the first time, recognizes known groups for all wellknown small instances from the literature. A quality of the individual community we define as difference between its benefit and cost, where both, benefit and cost greatly depend on the number of internal edges and the number of external edges, respectively. Comparison of the new criterion with modularity and ratio-cut criteria are performed on usual instances from the literature. For all 3 criteria, efficient heuristics based on Variable neighborhood search metaheuristic are designed.

#### Some general results on preferential attachment and clustering coefficient

#### Ludmila Prokhorenkova

#### Yandex, MIPT, Russian Federation

Many social, biological, and information systems can be represented by networks, whose vertices are items and links are relations between these items. That is why the evolution of complex networks attracted a lot of attention in recent years. In particular, numerous random graph models have been proposed to reflect and predict important quantitative and topological aspects of growing real-world networks. Turns out that many real-world networks of diverse nature have some typical properties: small diameter, power-law degree distribution, high clustering, and others. In this talk, I will focus on general approaches to the analysis of such properties and I will review several recent works in this field. The most well-known approach to the modeling of complex networks is preferential attachment. The main idea of this approach is that at each time step a new vertex is added to a graph and is joined to some vertices already existing in the graph chosen with probabilities proportional to their degrees. Preferential attachment allows to obtain a graph with a power-law degree distribution, and many different models are based on this idea: LCD, Buckley-Osthus, Holme-Kim, RAN, and others. I will present a general framework for analyzing preferential attachment models. In particular, I will discuss the PA-class of models which is defined in terms of constraints that are sufficient for the study of the degree distribution. Also, an additional constraint allows to analyze the behavior of the local and global clustering coefficients and the assortativity property in this class. Finally, I will present a general result on the behavior of the global clustering coefficient in scale-free graphs. Namely, I will show that for any sequence of graphs with a power-law degree distributions with a parameter  $\gamma \in (2, 3)$  the global clustering coefficient tends to zero. This result is quite surprising, since there is a common belief that for many real-world networks both the average local and the global clustering coefficients tend to a non-zero limit as the networks become large.

#### **Finding Critical Links for Closeness Centrality**

#### Oleg Prokopyev

#### University of Pittsburg, USA

Closeness centrality is a class of distance-based measures in the net- work analysis literature to quantify reachability of a given vertex (or a group of vertices) by other network agents. In this talk, we consider a new class of critical edge detection problems, where given a group of vertices that represent an important subset of network elements of interest (e.g., servers that provide an essential service to the network), the decision-maker is interested in identifying a subset of critical edges whose removal maximally degrades the closeness centrality of those vertices. We develop a general optimization framework, where the closeness centrality measure can be based on any non-increasing func- tion of distances between vertices, which, in turn, can be interpreted as communication efficiency between them. Our approach includes three well-known closeness centrality measures as special cases: harmonic centrality, decay centrality and k-step reach centrality. Furthermore, for quantifying the centrality of a group of vertices we consider three different approaches for measuring the reachability of the group from any vertex in the network: minimum distance to a vertex in the group, maximum distance to a vertex in the group, and the average centrality of vertices in the group. We study the theoretical computational complexity of the proposed models and describe the corresponding mixed integer programming formulations. For solving medium- and large- scale instances of the problem, we first develop an exact algorithm that exploits the small-world" property of real-life networks, and then propose two conceptually different heuristic algorithms. Finally, we conduct computational experiments with real-world and synthetic network instances under various settings, which reveal interesting insights and demonstrate the advantages and limitations of the proposed models and algorithms.

# Random graphs and applications

Andrey Raigorodskii

Yandex, MIPT, Russian Federation

I will give a survey of various random graph models and their applications in data analysis.

# Session talks

### **Information Propagation Strategies in Online Social Networks**

#### Marina Ananieva

#### National Research University Higher School of Economics, Russian Federation

Online social networks play a major role in the spread of information at very large scale. A lot of effort has been made in order to understand this phenomenon, ranging from popular topic detection to information diffusion modelling, including influential spreaders identification. The way information spreads through society has changed significantly over the past decade with the advent of online social networking. The purpose of this research is identification of subjects who play the most important role in information spread within the social networks and further evaluation of its propagation. We examine the online information diffusion in large-scale online social networks by applying Independent Cascade Model. The model assumes that each node independently influences its neighboring nodes. We show the results of applying IC model to such social networks and compare the results with the real networks extracted for different topics. We test the hypothesis that the subgraph (mask) of information propagation among friends list do not match with the subgraph of de facto disseminated information among the users of the network.

# Vehicle Assignment in Site-Dependent Vehicle Routing Problems with Split Deliveries

Mikhail Batsyn, Alexey Nikolaev

# National Research University Higher School of Economics, LATNA, Russian Federation

In this talk we consider the problem of vehicle assignment in heterogeneous fleet sitedependent Vehicle Routing Problems (VRP) with split deliveries. In such VRP problems vehicles can have different capacities, fixed and travel costs, and sitedependent constraints limit for every customer a set of vehicles, which can serve it. The Vehicle Assignment Problem (VAP) arises in heuristic and exact algorithms, when a vehicle is selected for the current route or vehicles are assigned to all currently constructed routes.

The VAP objective is to minimize the total assignment cost while the cost of assigning a certain vehicle to a certain customer or a certain route is computed in some heuristic way. Without split deliveries, when a delivery to a customer cannot be split between two vehicles or more, the VAP problem is equivalent to the variable cost and size Bin Packing Problem with additional constraints determining for every item a set of bins to which it can be placed.

We show that allowing split deliveries make the VAP problem polynomial, because in this case it can be reduced to the Transportation Problem. We also consider a special case, which is not rare in practice, when customers could be partitioned into classes, such that all customers in a class has the same set of vehicles, which can serve them, and these vehicle sets for different customer classes form a sequence of nested sets. We show that in this case if the cost per demand unit of assigning a vehicle to a customer depends only on the vehicle and does not depend on the customer, then the VAP problem can be solved by a linear algorithm.

# Trade credit Guarantee Networks with Frictionless Benefit

# Jayeeta Deshmukh

### Presidency University, Kolkata, India

In a world where actors are not anonymous, economic and social networks often play a major role in shaping economic outcomes. The garment manufacturing industry in Kolkata, possibly the largest non-branded readymade garment industry in India, is one such industry where industry participants exchange mutual favour in terms of providing trade credit guarantees to each other. This affects the profit level of industry participants significantly and one industry participant provides trade credit guarantee to another industry participants if and only if the latter belongs to his network. Field survey observations on this readymade garment industry shows that producers form various kinds of symmetric networks including dyads, triads, circular networks, etc. which are stable in nature.

This paper tries to provide a theoretical explanation of this when the benefit comes from direct links only, i.e., frictionless benefit and no network externality and industry participants face fixed and positive cost to form a link. In this paper,

producers meet and decide simultaneously whether they are willing to form bilateral links or not. If they decide to form a bilateral link between them then both need to incur a fixed cost of forming link, f>0. Mutual consent is required to form a link. However deletion of one link can be done unilaterally and there is no cost associated with it.

This paper assumes that producers provide trade credit guarantees only to those producers with whom they are directly linked. A producer will get some trade credit guarantee if at least one producer in his direct network is in a position to provide trade credit guarantee. However, the amount of trade credit guarantee depends on whether his indirect neighbours want trade credit guarantee or not. As producers do not know this a prori, each potential borrower forms an expectation regarding how much trade credit guarantee he can get from the network g.

This paper shows how the cost of forming link a bilateral link influences the architecture of a stable network by influencing the decision to form a bilateral link between any two producers. This paper has identified several cost zones at which different degrees of symmetric networks are strategically stable. This paper shows that for a very high cost of forming a bilateral link, an empty network is strategically stable; dyads or symmetric networks of degree 1 are strategically stable networks for high costs of forming a link; symmetric networks of degree two are the only strategically stable networks for the next cost zone, and so on. Complete networks are the only strategically stable networks for a very low cost of forming links. This paper further finds that small complete components of the same sizes co-exist within a larger network, i.e., small world phenomena exists.

# Comparison of different methods of graphical models identification

### Ivan Grechikhin

#### National Research University Higher School of Economics, LATNA, Russian Federation

Graphical models are an important application for different areas of science. In this work, different procedures of graphical models constructing are analyzed from practical point of view. The experiments use precision matrix as an underlying model, where non-zero elements reflect presence of connection between two variables. The distribution of sample correlation coefficient value is known from literature. Four procedures use different methods of p-value adjustment: Bonferroni adjustment (simple and Holm's step-down procedure) and Sidak adjustment(simple and Holm's step-down procedure) and Error Control in Gaussian Graphical Model Selection, Drton, Perlman, 2007). Another procedure uses apriori assumption about distribution of zero and non-zero elements of precision matrix; based on distribution, the procedure maximizes the likelihood of a network via greedy method. The computational experiments are conducted on 6 different types of matrices: low-density, medium-density, high-density, diagonal, tridiagonal, inverse tridiagonal. The quality of procedures is measured using different assessments, such as Type I and Type II errors, False Detection Rate and False Non-Detection Rate.

# Minimization of the average production costs via fractional programming problem

# Tatiana Gruzdeva

# Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the RAS, Russian Federation

We address the problem of minimizing the average production costs of electric power enterprises as a fractional program. Such problems are, in general, nonconvex (with numerous local extremums) and belong to a class of global optimization problems. We develop efficient global search method for fractional programming problems, which is based on two following approaches.

Generalizing the Dinkelbach's idea, we propose to reduce the fractional program with d.c. functions to solving an equation with the vector parameter that satisfies the nonnegativity assumption. In this case, we need to solve the auxiliary d.c. minimization problems.

We also propose reduction of the sum-of-ratios problem to a minimization of a linear function on the nonconvex feasible set given by d.c. inequality constraints.

Thus, based on the solution of these two classes of d.c. optimization problems we developed two-method technology for solving a general fractional program, based on the Strekalovsky's Global Search Theory for d.c. optimization.

First of all, the method developed was justified and tested on an extended set of problems with linear or quadratic functions in the numerators and denominators of the ratios.

Finally, we demonstrate the effectiveness of the proposed approach by carrying out a case study for real-world problem of minimizing the average production costs of Mongolian electric power enterprises.

This research was supported by the Russian Science Foundation (project no. 15-11-20015).

# A distributed supervised learning algorithm for smart grid networks

Mario Rosario Guarracino

National Research Council of Italy

With the spreading of local, distributed and intermittent renewable energy resources, power grids have been equipped with smart meters integrating bi-directional communications, advanced power measurement and management capabilities. Smart meters make it possible to remotely turn power on or off to a customer, read usage information, detect a service outage and the unauthorized use of electricity. To fully exploit their capabilities, we foresee the usage of distributed supervised classification algorithms. By gathering data streams from meters and other sensors located in network nodes, such algorithms can create local classification models for online monitoring, workload balancing, prediction of energy demand and incoming faults. In this talk we present a decentralized distributed classification algorithm based on proximal support vector machines. The method uses partial knowledge, in form of data streams, to build its local model on each node of the network. We demonstrate the performance of the proposed scheme on synthetic datasets.

### **Multimodal Clustering for Community Detection**

#### Dmitry Ignatov

#### National Research University Higher School of Economics, Russian Federation

Multimodal clustering is an unsupervised technique for mining interesting patterns in n-adic binary relations or n-mode networks. Among different types of such generalized patterns one can find biclusters and formal concepts (maximal bicliques) for 2-mode case, triclusters and triconcepts for 3-mode case, closed n-sets for n-mode case, etc. Object-attribute biclustering (OA-biclustering) for mining large binary datatables (formal contexts or 2-mode networks) arose by the end of the last decade due to intractability of computation problems related to formal concepts; this type of patterns was proposed as a meaningful and scalable approximation of formal concepts. In this paper, our aim is to present recent advance in OA-biclustering and its extensions to mining multi-mode communities in SNA setting. We also discuss connection between clustering coefficients known in SNA community for 1-mode and 2-mode networks and OA-bicluster density, the main quality measure of an OA-bicluster. Our experiments with 2-, 3-, and 4-mode large real-world networks show that this type of patterns is suitable for community detection in multi-mode cases within reasonable time even though the number of corresponding n-cliques is still unknown due to computation difficulties. An interpretation of OA-biclusters for 1-mode networks is provided as well.

# Searching for Maximal Quasi-Bicliques via MIP-Based Models

# Dmitry Ignatov

National Research University Higher School of Economics, Russian Federation

A quasi-bi-clique is an extension of the classical quasi-clique concept in the context of bipartite graphs.

Quasi-cliques and quasi-bi-cliques provide natural mathematical models for representation of dense subgraphs. The problem of finding such large subgraphs routinely arises in a number of important data mining and graph analysis applications. The talk consists of two major parts. The first part first reviews the related literature, formal definitions and mixed integer programming modeling techniques for finding maximum quasi-cliques. The main contribution of the first part consists of extending such techniques for the problem of finding maximum quasi-bi-cliques.

The second part provides computational results, where the proposed mixed integer programming models are solved by a standard solver (CPLEX). The talk also provides comparisons of the obtained results against a greedy heuristic from the literature. The obtained results are reasonable and well-articulated.

# Weighting Relationship Strength in Online Social Networks

# Ilia Karpov

# National Research University Higher School of Economics - International laboratory for Applied Network Research, Russian Federation

Many researchers confirm the importance of weighting Online Social Network (OSN) actor relations with complex models because trivial metrics like friendship or group membership provide only a coarse representation of actor closeness. For instance, one can have best friends, casual friends and accidentally added individuals in the same list. This noisiness is usually caused by the low cost of link formation and mental efforts needed to delete it when the link becomes obsolete.

These proposed weighting models are based on the hypotheses that given two actor profiles as an observed variable, we can describe the actor similarity by modelling hidden interaction variables such as communication or liking. It causes the following limitations: (1) it is difficult, if not impossible, to collect all possible interaction variables from the OSN, which requires selection of the most important variables only, (2) the actor similarity is based on profile information which restricts comparison between actors of different types.

We study the problem of weighting relationship strength between three different actor types (users, groups, public pages) in a social network graph using two metrics. The first one is based on hidden actor interactions; the second one is based on actor's texts topic modelling. The first metric computes various interaction data about two actors. Over 30 different relations, provided by OSN, such as "amount of comments", "joint membership" and "amount of simultaneous likes" are included in our research. Having statistics for two relations {R1,R2} we compute the causation coefficient as the difference between the conditional dependencies, p(R1|R2) - p(R2|R1) to get the initial ties of OSN actors. The second metric is based on the hypotheses that texts, written by two actors can be described by a mixture of document distributions. The more the topics tend to be written by only one user, (all documents of the topic in the worst case) the further actors are from each other.

Our evaluation of Facebook data confirms the fact that friendship is not the strongest relation between Facebook users. According to the proposed metric, the strongest relations between users are mutual post liking and simultaneous commenting the same post. One of the key features of the proposed method is obtaining links between different actor types, e.g. the strongest relation between user and group is group posts commenting. Evaluation of the second metric on three class classification task (friends, best friends, relatives) shows accuracy 0.63 that proves this metric to be useful when we lack interaction data, but suppose some interrelation between two actors.

#### Evader's models in sequential network interdiction

#### Sergey Ketkov

#### National Research University Higher School of Economics, LATNA, Russian Federation

We study a sequential network interdiction process, through which an interdictor and an evader interact. The evader tries to minimize his cumulative loss for T rounds, while the interdictor is greedy and each time period blocks a set of k-most vital arcs in the observed network. We consider the evader's strategies that are deterministic and assume that the evader has complete knowledge of the arc costs and graph's structure. We demonstrate that the evader's problem is NP-hard even in the case of T = 2 and no constant factor polynomial time approximation algorithms exist. Some general properties of optimal evasion policies are discussed and a class of greedy policies is introduced. We describe necessary and sufficient conditions for the greedy policies to be optimal in a particular case. We also construct a heuristic algorithm for strategic evader on the basis of our theoretical observations. The numerical experiments demonstrate, that the proposed heuristic consistently outperforms the greedy shortestpath based policy on many real-life network instances.

# **Application Of Nonlinear Adaptation On Manifolds For Economic Objects**

### Svetlana Kolesnikova

#### National Research Tomsk State University, Russian Federation

An application of the method of analytical design of aggregated regulators for constructing systems of control over integrated economic entities with positive and negative feedback loops is discussed. The controlled objects are presented as a system of ordinary non-linear differential or difference equations with chaotic behavior in the case of certain combinations of parameters. A number of control problem formulations are given and the algorithms of control synthesis are reported and theoretically validated. Some illustrative examples of application of the proposed algorithms are provided along with their numerical simulation data. The results obtained would be useful in designing a smart control system and for real-time decision making in the financial policies of a variety of economic entities.

Keywords. non-linear multi-dimensional object, disturbance compensation, deterministic chaos, analytical design of aggregated regulators.

#### Application of market models to network equilibrium problems

# Igor Konnov

#### Kazan Federal University, Russian Federation

We present a general two-side market model with divisible commodities and price functions of participants. Using its equivalent variational inequality reformulation, we obtain an existence result on unbounded sets. Under certain additional conditions the model can also be reduced to a decomposable optimization problem where the goal function is the sum of two functions and one of them is convex separable, whereas the feasible set is the corresponding Cartesian product. We suggest a new cyclic version of the partial linearization method, which makes component-wise steps satisfying a descent condition and utilizes tolerance parameters. We describe an extension of the network flow equilibrium problem with elastic demands and show it is a particular case of the general market model. Besides, we describe a new equilibrium type model for resource allocation problems in wireless communication networks, which also appears to be a particular case of the market model. This enables us to obtain new existence results for these models as some adjustments of that for the general auction model and to solve these problems with the partial linearization method.

# **Analysing Online Recomendation Networks**

#### Nadezhda Kostyakova

#### National Research University Higher School of Economics, Russian Federation

Social networking services have become highly popular around the world. Most of these services consist of connections between people while there are a lot of recommendation networks which connect people and products. Such networks are widely used. People recommend different products by rating them, writing reviews about them and commenting other reviews. Consequently, recommendation networks have big amount of useful information which can be used for analysis of different patterns. However, since this information is used by consumers to make decision about buying products there is a big motivation to abuse it by advertising or discommend some products. We first annotate social network messages with sentiment attitude using text sentiment analysis. Next, we generate generate the two mode network of brands - actors, using the given sentiment category and brands taxonomy. This work provides a comparative study of networks, extracted from advertising services like IRecommend.ru and different social networks, such as Twitter and VKontakte.

# Modular structure of brain networks in healthy and diseased brains

# Anvar Kurmukov

# National Research University Higher School of Economics, Russian Federation

Complex connectivity structure of a human brain can be represented by a graph and therefore studied using powerful methods of network science. The brain graph is reconstructed based on non-invasive neuroimaging data: cortical brain regions become vertices of the network, and structural connections between regions estimated using tractography algorithms become its edges. These graphs are called connectomes, and the area of neuroscience that analyzes brain graphs using network methods and approaches is called connectomics. An ultimate goal of this field is to understand typical structural properties of healthy brain networks and their changes in normal development or brain disease.

One of the key properties of brain networks described in the literature is their modular structure, which is a tendency of brain regions to group into communities with dense intra-group connections and sparse inter-group links. In our studies, we make the next step and hypothesize that anatomical changes of brain connectivity patterns (e.g. caused by a certain disease) immediately affect brain modular organization thus producing meaningful differences in community structure between networks obtained from healthy and diseased brains. To evaluate this idea, we propose a method for classifying normal versus pathological brain networks based on their underlying modular organization. Importantly, connectomics considers both non-overlapping and overlapping communities of brain regions; the latter approach seems to be even more biologically plausible because many cortical areas are known to be heteromodal and to have a role in multiple subnetworks. To accommodate and compare both approaches, we consider two classes of methods for clustering networks, those that use hard partitions and those that produce soft ones. Soft partitions allow us to take into account possible module's overlappings in evaluating optimal community structure of individual brain graphs.

After obtaining community network structure of individual graphs, we next measure a distance between connectomes as a distance between their optimal partitions and construct a kernel-based classifier which predicts subject's phenotype from its brain graph. We demonstrate the performance of the proposed approach in a task of classifying structural connectomes of healthy subjects and those with a neurodegenerative disease.

# The comparative analysis of big networks with various network topologies

# Ilya I. Kurochkin

Institute for information transmission problems of the RAS, Russian Federation

An attempt of comparative analysis of network topology FatTree, JellyFish and DragonFly is made. The modified technique when make comparison is not only according to network graphs of this topology, but also by results of simulation modeling of functioning of a network is offered.

Simulation mathematical modeling consists in a sequential route of path on the network graph with an opportunity to use different criteria of dynamic routing. Treat the considered parameters of the initial network graph: diameter, order of node, values of arcs capacity, values of the minimum cuts. To the parameters received as a result of simulation modeling: percent of failures, residual resource of a network and etc.

In article the basic principles of network topology developing of FatTree, JellyFish and DragonFly for generation of graphs of networks are discussed. Comparing of three network topology on the basis of results of a numerical experiment on simulation modeling of three sets of networks is given.

# Intra-organizational networks in organizations: communication networks as a key to success

# Ivan Kuznetsov

# National Research University Higher School of Economics - International laboratory for Applied Network Research, Russian Federation

Every organization is a network structure, designed for successful solving of its needs through cooperation of individuals (Kadushin, 2012), especially project-based ones. The only way to cooperate is to communicate. Since communicating is not optional, the opening of communication channels for the information to flow in the right directions is critical for the organization's performance as well as the morale of employees (Sinickas 2001).

With increase of project-based organizations in past years a lot of issues arise too. Project based structure could be a tough thing to manage. Employees should communicate both inside and outside of the project team, sometimes individual cannot differentiate their team co-members, there is huge massive of information flowing, etc. So they form informal communication networks to build more effective way to solve arising issues. Of course, it is the managers' responsibility to handle and administer the communication process and a privilege to inspire more constructive interactions. This could possibly lead to slight improvements, but often cause distraction, overload, poor quality and quantity of information with further negative results for successful task implementation.

This paper is an attempt to untie the knot of nets in organizations at the junction of human and project management literature with help of social network analysis. Its focus lies between communication networks and their effect on successful project implementation. In this paper examined such communicational networks as: workflow; employee voice; advice; problem solving; deficiency of contact; conflict; communication distortion. We past that networks are differ both in structure and in its effect on employees perceived efficiency.

The research is longitudinal and is conducted in mid-sized Moscow organization in 3 waves: 1st – April, 2nd – June, 3rd – November 2017.

# Developing a model of topological structure formation for power transmission grids based on the analysis of the UNEG

#### Sergey Makrushin

#### Financial University under the Government of the Russian Federation

The aim of the current research is to find principles for creating a model of topological structure formation for power transmission grids. In the research, the information about Unified National Electricity Grid (UNEG) is used as a source of empirical data about power grids structure. The UNEG is the power transmission grid of Russia, the major part of which is managed by Federal Grid Company of the Unified Energy System (FGC UES). The research methods are based on complex network theory, which is widely used for power transmission grid analysis and for developing models of topological structure formation for infrastructure networks.

For the analysis the UNEG appears as a network, with electric power stations and substations as nodes and high-voltage power lines as links. For the current research data on 514 nodes and 614 links has been gathered. This data includes network topology, geographical coordinates of the nodes, region binding, voltage levels and other properties. The data for the computer model creation has been taken from official documents, online UNEG map from FGC UES and OpenStreetMap GIS. A UNEG computer model has been created for main operating regions of the UNEG except the unified national energy systems of Siberia, East and Urals (partially).

Researches in the field of the complex network theory have shown that there are several universal network models which are applicable to a wide set of real world networks from different domain areas. The point is that all real world complex systems are not formed instantaneously but are grown in a process which is managed by a set of rules. Similarity of the rules which manage growth of complex networks makes it possible to create universal complex networks models which correctly describe topology of the networks. Thus, when we build a model of topological structure formation for some type of networks, we could get an understanding of the rules which manage the growth of these networks. Analysis of these rules could provide clues for main parameters which define structure of a network and consequently help to correctly compare different networks and build up long term forecasts of the network development.

In the current research two ways are used for developing a model of topological structure formation for power transmission grids: analysis of applicability of universal models from complex network theory and analysis of ad-hoc models of topological structure formation for power transmission grids. Analysis of nodes degree distribution is used in the research as the main criterion of applicability for different network models. This criterion is quite popular in complex network theory, and it is stated in previous papers that nodes degree distribution for power grids is conform to the geometrical distribution law, which is typical for the random growing network model. For the UNEG network, more correct statistical analysis of nodes degree distribution is provided in previous papers. It is based on log-likelihood value instead of regression

analysis. Our analysis has shown that only the tail of nodes degree distribution of the UNEG network is conform to the geometrical distribution law, but the head of the distribution is similar to the head of the Poisson distribution law. However, the Poisson distribution law is typical for another network model: the geometrical graph model. Based on these observations, the negative binomial distribution law was taken into account. Peculiarity of the negative binomial distribution is that it could have the head looking like the head of the Poisson distribution and have the tail looking like the tail of the geometrical distribution. The Akaike information criterion and the Bayesian information criterion have shown that a negative binomial distributions of previously considered network models. But the problem is there is no model for generation of a network with a negative binomial distribution of nodes degrees.

Based on this analysis, we get a conclusion that creation of a compound model of topological structure formation for power transmission grids is promising. For this compound model, we need to use basic principles which determine the occurrence of distribution laws in the geometrical graph model and the random growing network model. In the new model we need to use the historical principle from the random growing network model and the spatial principle from the geometrical graph model. In accordance with the historical principle, most nodes are added to network consequently, and it is more likely that older nodes will have more links than younger ones. In accordance with the spatial principle, all nodes in a network are embedded in metrical space, and existence of a link between nodes depends on spatial distance between them.

It should be noted that, in addition to the historical and the spatial principles, the compound model should also take into account the special role of transit nodes with degree of 2 in power grids. Analysis of the UNEG network demonstrated that the considered models could not correctly describe a high proportion of transit nodes in real power grids. Visual analysis of power grids shows that the high proportion of transit nodes is the consequence of high occurrence of chain pattern in power grids topology. The chain pattern is a chain of transit nodes which form a long power line route. This route usually has a certain spatial direction and connects distant parts of a power grid. Typically, all nodes of a chain are added to a grid within the framework of one project of a grid development. The geometrical graph model and the random growing network model do not form sufficient quantity of chain patterns in a network, thus formation of chain patterns should be used as an additional principle of the compound model.

After the formulation of the principles for the compound model we tried to find an adhoc model of topological structure formation for power transmission grids which had implemented them. As a result, we have found the random growth model (RGM) for power grids, described in paper by P. Schultz et al. This model stands out from other models, because it implements the historical and the spatial principles and even has a mechanism for the formation of chain patterns. As a result, the RGM model could form a topological structure of the network which has a nodes degree distribution appropriate to an empirical distribution for power grids.

The algorithm of the RGM consists of an initialization phase, with the network

extending tree-like for minimum cost, and a growth phase, with an attachment rule giving a trade-off between cost-optimization and redundancy. Furthermore, it implements the feature of some lines being split during the grid's evolution. In general, the two-phase process quite corresponds to the real process of growth of a power grid. At an early stage of power grid development, the structure of a grid is planned as a globally optimal network, but at later stages of development, a significant part of new nodes is added with only local optimization of costs for each specific operation. Nevertheless, besides these two stages, an important role in a power grid development is played by the realization of medium range development projects. A common case of such projects is creation of long power line routes, which forms a chain pattern in a power grid. The logic of nodes addition in these projects is between global optimization (first phase of the RGM) and local optimization for individual nodes and links (second phase of the RGM). As well as the creation of original power line routes, the lines splitting procedure in the RGM leads to an abundancy of transit nodes, but it cannot form the same topology structure. Therefore, we need to modify the RGM algorithm by removing the splitting line procedure and by adding the process of network growth based on medium range projects.

In the current research, the analysis of applicability of universal models to the UNEG network has helped us to identify the key principles of a model of topological structure formation for power transmission grids. Due to these principles, we found a suitable ad-hoc model of network formation and found its weaknesses and strengths. These same principles helped us to propose recommendations for the improvement of the selected ad-hoc model.

# Geometric model of rich-club phenomenon in complex networks

# Yury Malkov

# The Institute of Applied Physics of the RAS, Russian Federation

The rich club organization (the presence of a highly connected hub core) influences many structural and functional characteristics of networks including topology. Despite its major role, the literature contains only a very limited set of models capable of generating networks with realistic rich club structure. One possible reason is that the rich club organization is a divisive property among complex networks which exhibit great diversity, in contrast to other metrics (e.g. diameter, clustering or degree distribution) which seem to behave very similarly across many networks. We discuss a simple yet powerful geometry-based growing model which can generate realistic complex networks with high rich club diversity by controlling a single geometric parameter. The model is validated against the Internet, protein-protein interaction, airport and power grid networks.

# References

Csigi, Máté, Attila Kőrösi, József Bíró, Zalán Heszberger, Yury Malkov, and András Gulyás. "Geometric explanation of the rich-club phenomenon in complex networks." Scientific Reports 7 (2017).

# Analyzed of co-authorship networks and scientific citation based on Google Scholar

Nataliya Matveeva, Oleg Poldin

# National Research University Higher School of Economics, Russian Federation

In this study, we analyze the correlation between the co-authorship network parameters and citation characteristics of scientists in Google Scholar. For network constructing profile of scientists from various countries and scientific fields in GS were used. We ran the count data regression model for a sample of more than 30 thousand authors with the first citation after 2007. There is a positive correlation between scholar's citation counts and number of co-authors, between citations and the author's centrality, and between scholar's citations and the average citation of co-authors. The h-index and i10 index are correlated significantly with the number of co-authors and average citation of co-authors.

Cooperation between researchers is an essential feature of scientific activity. It assumes a work of several researchers on a scientific problem, during that scientists exchange ideas, discuss problems and results, and generate new ideas. This kind of partnership can increase performance of future researches and results of this cooperation generally appear in joint publications. The significance and scientific contribution of scientists and their publications can be evaluated by such parameter as citation - the total number of references on their publications in other works and its derivatives (Hirsch-index, I-index etc.). The cooperation between researchers can be analyzed by co-authorship network, where authors of publications are nodes and their joint publications are links.

The present study is based on bibliographic database Google Scholar (GS). In contrast to Web of Science (WoS) and Scopus, Google Scholar provides the opportunity for the users to organize their scientific profiles and to manage their co-authors lists. GS indexing more scientific sources and has free access and the relative simple information extraction. This allows to analyze large amounts of data about scientists from various scientific fields. Furthermore, because authors can manage their profiles in GS this significantly reduce the misidentification of authors with the same names and their works. As drawbacks of GS it can be mentioned the indexation of not verified scientific sources (Jacsó, 2008a) and the fact that information in authors profiles is not required to be filled correctly. Despite this there is a strong correlation between the authors ratings based on GS and other database (WoS, Scopus) (Franceschet, 2009; Wildgaard, 2015).

Here we evaluate the relation between network characteristics of the scientists and their bibliometric characteristic: total number of citations, h-index, i10 - index. We divided sample on scientific fields since it has various citation indicators and we take into account the time effect of citation because the presence in sample of authors with various scientific tenures affects their bibliometric indicators (Ductor et al., 2014). For this study, 110 thousand profiles of scientists registered in GS from various

countries and scientific fields were used. We collected data about number of publications, citation, h-index, i-index, list of co-authors. After preparing the adjacency matrix, we calculate short paths for each nodes using Dijkstra's algorithm (Dijkstra E.W., 1959). Total number of links (joint publications), number of links that authors have (degree centrality) were also computed and the author's position in relation to other participants in the network, measured by closeness centrality. The scientists where divided into the groups according to the keywords in there profiles corresponding to the fields of: computer science, economics and finance, biology and medicine, physics and chemistry, mathematics, social sciences and humanities. In empirical models were evaluated networks characteristics of 34 thousand author's profiles taking into account links with all network participants. Limitations caused by the fact that information about year of the first citation is available only after 2007 year in GS (for the data that were collected in the Jan 2014) that leads to inclusion tenure of citation in explanatory variables only for profiles with first citation after 2007. Thus in our sample there are only early carrier scientists. There were found that the highest average values of citations, indexes Hirsch and i10 are in the fields of biomedicine, physics and chemistry, and the smallest indexes - in economics and finance. The average number of co-authors for all disciplines is 6, this indicator is above average in the natural sciences, below average is in economics & finance. For full sample closeness centrality is 0.185 it is above average in computer science, below average – in economics & finance and physics & chemistry. To estimate relation between scientist collaboration and their bibliometric characteristic were used regression analysis. In our study, the dependent variables are inherently measure the number of events (citations) and are therefore non-negative integers so we used negative binomial model (Ajiferuke, Famoye, 2015). The results were found for three specifications of the model, which is differing by number of explanatory variables. According to the results, the closeness centrality plays a minor role relatively on the number of co-authors. For the specification that includes all independent variables, closeness centrality is insignificant, but it is significant for specification when the variable of the number of co-authors is excluded. Estimations results for the specifications without closeness centrality and in the complete specification are very close. Average co-authors citation and tenure are significant at the 1% level.

Since the variables vary in scale, we estimated the changes of independent variables (citation, H-index and i-index) relatively its average values if one of the dependent variables change in one standard deviation and other variables are fixed. In the full sample, increase in the average co-authors citation on the standard deviation (5728) entail an increase in the citation at 116.8 units and increase in the number of co-authors from 6 to 13 people entail citation growing at 92.4 units. The greatest changes in citation and its derivatives were observed for authors working in the field of biology and medicine, the lowest is in economics and finance. However, the average value and dispersion of the citation for biology and medicine is much higher than those for the economy and finance.

In addition, we analyze the correlation between Hirsch – index and average citation of co-authors, number of co-authors and the logarithm of tenure. All variables are

statistically significant and the number of co-authors has greater significant impact on the h-index than the citation of co-authors, while for the citation effects of these two parameters are comparable. Hirsch index has a linear dependence on publications experience (for citation it is square) and normalized tenure effect for H-index (0.66) is higher than for citation (0.30). To analyze the i-10 index correlations, we used a hurdle models that consist of two parts and describe the distribution of zero and nonzero values of the dependent variable (McDowell A, 2003). We found that i-10 index significantly correlates with the number of co-authors and average citation of coauthors; the normalized effect of average co-authors citation is lower than it is for number of co-authors.

We determined that there is a positive correlation between the scientists' citation and the number of co-authors, between the citation and closeness centrality. There was found that the scientists' citation positively correlated with average citation of their coauthors. The indicators vary according to the authors scientific fields, but the effects are comparable. Thus, analysis of co-authorships social network based on Google Scholar is showed that the scientists who maintain more contacts and more active than others have better bibliometric indicators on the average.

References

1. Ajiferuke I., Famoye F. (2015). Modelling count response variables in informetric studies: Comparison among count, linear, and lognormal regression models. Journal of Informetrics, 9 (3), 499–513.

2. Dijkstra, E.W. (1959). A note on two problems in connexion with graphs.Numerische Mathematik 1, 269-271.

3. Ductor L., Fafchamps M., Goyal S., van der Leij M. J. (2014). Social networks and research output. Review of Economics and Statistics, 96 (5), 936–948.

4. Franceschet M. (2009). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. Scientometrics, 83 (1), 243–258.

5. McDowell, A. (2003). From the help desk: hurdle models. The Stata Journal, 3(2), 178-184).

1. Jacsó P. (2008a). The pros and cons of computing the h-index using Google Scholar. Online information review, 32 (3), 437–452.

2. Wildgaard L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. Scientometrics 104 (3), 873–906.

# On the equality of triangle-matching and triangle-cover numbers in graphs

# Dmitry Mokeev

#### National Research University Higher School of Economics, LATNA, Russian Federation

A triangle-matching in graph is the set of vertex disjoint subgraphs isomorphic to the cycle of three vertices (triangles). A triangle-cover in graph is the set of vertices such as each triangle contains at least one vertex of this set.

We consider a hereditary class of graphs which have a triangle-matching and a triangle-cover of equal cardinalities. Main properties of such class were proven and some families of minimal forbidden induced subgraphs of such class were found.

#### Methods of criteria importance theory and their software implementation

Andrey Nelyubin, Vladislav Pidinovski, Mikhail Potapov

Mechanical Engeneering Research Institute of the RAS, National Research University Higher School of Economics, Institute of Computer Aided Design of the RAS, Russian Federation

Most of the real decision making problems are inherently multicriterial. The decision maker (DM) needs to assess his/her subjective preferences as accurately as possible to choose the best final alternative. For this purpose he/she can utilize mathematical and computer tools that provide opportunities to use complex mathematical methods for multicriteria analysis and optimization.

Among other approaches to analyzing and solving multicriteria problems, the Criteria Importance Theory (CIT) developed in Russia has a number of special advantages. In this theory, a formal definition of the relative importance of criteria is introduced, which makes it possible to correctly take into account incomplete and inaccurate information about the preferences of DM: the relative importance of criteria and the change in preferences along the scale of criteria. This information can be expressed both qualitatively and quantitatively, in the form of intervals of possible values of preference parameters.

Different types of information about the preferences of the decision maker and their combinations require their own methods of analyzing the problem and algorithms for comparing alternatives by preference. Within the framework of CIT, precise and effective methods and algorithms for solving such problems have been developed. It was necessary to obtain solutions of a number of linear, nonlinear, and discrete problems. For some of these problems analytical solutions have been obtained. The paper provides an overview of the proposed methods and special algorithms. To solve multicriteria choice problems using CIT methods, the authors develop the software system DASS. In this system, the solution of the choice problem may be organized in the form of an iterative process. Initially, the DM introduces only basic information about the problem. He/she does not need to immediately indicate the exact information about his/her preferences. The DM clarifies and specifies this information during the process of solving the problem in a dialog mode. To do this, he/she uses the results of multicriteria analysis of solutions already obtained in the previous stages. Such intermediate results make it possible to more clearly see the relationships between the various goals and constraints that determine the decision making problem.

# On a Numerical Solving of Bilevel Pricing Problem in Telecommunication Networks via D.C. Maximization

# Andrei Orlov

# Institute for System Dynamics and Control Theory of Siberian Branch of the RAS, Russian Federation

Hierarchical optimization problems are one of the urgent paradigms. A development of the new efficient numerical methods for bilevel programming problems (BPPs) now is a challenge for modern theory and methods of Mathematical Optimization. There are a lot of applications of the BPPs in control, economy, traffic, telecommunication networks, etc.

In this work we develop new methods for seeking optimistic solutions to the hierarchical problem of optimal pricing in telecommunication networks. These methods based on a possibility of equivalent representation of a bilevel optimization problem as a nonconvex optimization problem (with the help of Karush-Kuhn-Tucker conditions and penalty approach). Then for solving of the obtained nonconvex problem we apply the Global Search Theory (GST) developed by A.S. Strekalovsky. GST allows to construct efficient numerical methods for several classes of one-level and bilevel problems with nonconvex structures (with the dimension up to 1000). Global Search Algorithms based on GST consist of two main stages: 1) the special Local Search Method (LSM), which takes into account the structure of the problem in question; 2) the procedures, based on Global Optimality Conditions, which allow to escape the point provided the LSM. For a local search in the reduced nonconvex problem we apply the idea of consecutive solving of partial problems with respect to two groups of variables. The procedures of Global Search for problem under scrutiny based on the corresponding methods for problems of d.c. maximization because the objective function in this problem can be represented as a difference of two convex functions.

Computational testing of the elaborated methods on the test problems of optimal pricing in telecommunication networks demonstrated the workability and efficiency of the proposed approach.

This work is carried out under financial support of Russian Science Foundation (project no. 15-11-20015).

# Selective bi-coordinate variations for network equilibrium problems with mixed demand

Olga Pinyagina, Igor Konnov

Kazan Federal University, Institute of Computational Mathematics and Information Technologies, Russian Federation

We propose a modification of bi-coordinate method with a special tolerance control and thresholds for network equilibrium problems with mixed demand. Preliminary numerical experiments confirm advantages of this method in comparison with the ordinary conditional gradient method.

The network equilibrium problem with mixed demand generalizes the network equilibrium problems with fixed and elastic demand; it has been stated in paper [1]. Being based on the equilibrium conditions of this problem, we apply the approach originally proposed in [2], [3] for optimal resource allocation problems with simplex type constraints, formulate a modification of the method of bi-coordinate variations, and prove its convergence properties.

References:

[1] Pinyagina O. On a Network Equilibrium Problem with Mixed Demand // In: Kochetov, Yu. et al (eds.) DOOR-2016. LNCS, vol. 9869, pp. 578-583. Springer, Heidelberg (2016).

[2] Konnov I.V., Selective bi-coordinate variations for resource allocation type problems, Computational Optimization and Applications 64 (3), 1-22 (2016).

[3] Konnov I.V., A method of bi-coordinate variations with tolerances and its convergence, Russian Mathematics 60 (1), 68-72 (2016).

# Emotion recognition in sound Deep neural networks performance optimization via the Distillation the Knowledge

Rassadin A., Savchenko A.

National Research University Higher School of Economics, LATNA, Russian Federation

Nowadays, deep neural networks (DNNs) become a popular powerful tool to tackle the most challenging tasks in image recognition, natural language processing and speech recognition. However, it is known that the modern neural network architectures have very high space complexity. Hence, the DNNs application in autonomous mobile systems is still a challenging problem. In this talk we discuss several known techniques to optimize their performance and memory consumption. In our experimental study we focus on the possibility to architecture-independent distillation the knowledge from large DNN by using an unlabeled image dataset. We propose to train rather small SqueezeNet model using not only the labels predicted by large VGG-like DNN, but the vectors of posterior probabilities of all classes at the output of the softmax layer solving a multi-optimization task where one loss, for the posterior probabilities, defined as the Kullback-Leibler divergence (KLD) and the other is cross-entropy for output labels. It is experimentally demonstrated that such usage of the scores for each label improves the accuracy of the optimized DNN.

# Decomposition Complexity of Graphs and Its Applications to Analysis of Socio-Economic Systems

### Alexandr Rubchinsky

# National Research University Higher School of Economics, The National University of Science and Technology MISiS, Russian Federation

A new approach to decomposition of undirected graphs is suggested. The conventional ap-proaches deal with the construction of a single, "the most correct", decomposition of a conside-red graph. However, in some cases, such a decomposition does not exist or there are several dif-ferrent decompositions of such a type. These cases are not vexatious mistakes. Moreover, it is possible to assert that numerous decompositions naturally arise in the study of many graphs. Be-cause many real systems (including systems, whose functioning is determined by human activi-ty) are naturally modeled by graphs, the domain of applications of graph decomposition is large enough. Hence, it seems expedient to consider – as an important generalization of the conven-tional decomposition problems – construction of a family of decompositions instead of a single one.

As in the conventional case, the constructed family of decompositions characterizes the initi-ally considered system. Moreover, in such situations decompositions themselves, forming the above-mentioned family, are of little interest. It turned out that it is much more expedient to fo-cus our attention on calculation of special numerical indices based on these families. The suggested

indices describe such properties of the initial system that cannot be revealed by the conventional decomposition approach. They have different meaningful interpretations in different situations, but generally they describe complexity, entanglement, intricacy and other similar hardly defined, though important, properties of various real systems. Therefore, these indices are referred to as decomposition complexity indices.

The suggested approach to graph decompositions turns out to be helpful in analysis of various real systems, especially in order to extract some hidden facts from raw data concerning these systems. Two different applications of the approach are considered in the report:

• applications for analysis of voting political bodies;

• applications for analysis of crises in stock markets.

In both cases, the presented results cannot be obtained by other known methods. Particularly, for stock market S&P-500 the simple pattern, preventing about big crises in 5 - 7 days before the event, is elaborated. These examples confirm the advantages of the suggested approach.

#### Computational study of activation dynamics on networks of arbitrary structure

Alexander Semenov, Stepan Kochemazov, Dmitry Gorbatenko

Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the RAS, Russian Federation

In the report there will be considered a number of problems, based on the same mathematical model. In particular, we consider a directed graph G representing some collective. The structure of this graph, in a general case, can be arbitrary (for example, it can be constructed according to one of the known random graph models). Graph vertices represent agents, and graph arcs interpret influence of agents on others. With each agent we associate its weight function. We consider situations, when the values of weight function are elements from the set  $\{0,1\}$ , where 0 stands for inactivity and 1 for activity. There can be three types of agents in the collective: activators (always active), deactivators (always inactive) and simple agents (the activity of which depends on the number of active agents in their neighborhood). We introduce discrete time. At each time moment we consider the set of values of weight functions of all graph vertices. These values are recomputed synchronously. This set of values forms the Boolean vector, to which we refer to as to a state of a system (collective) at the considered time moment. As a result we have a discrete dynamical system of automaton type (it can be considered as a finite automaton without inputs and outputs). In the context of the described model the following problems are studied. We want to position as small number of activators as possible so that after relatively small amount of time steps as many simple agents as possible become active (here, because of an arbitrary structure of G, the activation speed greatly depends on the disposition of activators). Then, for a network, in which the majority of simple agents are active, we need to find such a disposition of small number of deactivators (which replace simple agents) that after small amount of time a network reaches mostly inactive state.

Earlier in the context of the proposed model we considered phenomena of conforming and anticonforming behavior in collectives. The problems of finding dispositions of agents with constraints on the number of steps, the percentage of activators and other parameters, were solved by reducing them to Boolean satisfiability problem (SAT). Currently, by using state-of-the-art SAT solvers we can solve the problems of proposed type for networks with thousands of vertices.

In the report we will present new results on applying the automaton models of activation dynamics, described above, to problems studied in computer security. In particular, we will consider the problems of constructing attacks on computer networks (as activation problems), and also the problems of patching vulnerabilities to prevent attacks (as deactivation problems).

# **Company Co-Mention Network Analysis**

#### Sergei Sidorov

#### Saratov State University, Russian Federation

In network analysis, the importance of an object can be found by using different centrality metrics such that degree, closeness, betweenness and so on. In our research we form a network, which we called company co-mention network. The network is constructed quite similar to social networks or co-citation networks. Each company is a node and news mentioning two companies establishes a link between them. Each company acquires a certain value based on the amount of news which mentioned the company. This research examines the network of companies by using companies comention news data. A matrix containing the number of co-mentioning news between pairs of companies is created for network analysis of companies whose shares are traded on major financial markets, and the networked map of the companies is used to visualize the dependence structure of the economy by identifying groups of companies who are more central than others. News analytics data have been employed to collect the companies co-mentioning news data, and R packages has been used for network analysis as well as network visualization.

# Polynomial-time solvability of the independent set problem in a certain class of subcubic planar graphs

# Dmitry Sirotkin

National Research University Higher School of Economics, LATNA, Russian Federation

The independent set problem for a given simple graph is to compute the size of a maximum subset of its pairwise non-adjacent vertices. In this paper we prove polynomial-time solvability of the problem for subcubic planar graphs not containing an induced tree, obtained by coinciding ends of three paths of lengths 3,3,2, correspondingly.

# The Properties of Optimal Portfolios for Markowitz Problem

#### Sergey Slashchinin

#### National Research University Higher School of Economics, LATNA, Russian Federation

Portfolio selection is the construction of portfolios that maximize level of the expected returns from the investment, but at the same time have low involved risks. One fundamental approach for quantifying the risk-return trade-off of assets is mean-variance analysis. In this case, it is crucial to accurately estimate covariance matrix of stock returns. Different estimation techniques will be examined, but we will put the main emphasis on the one of the current industry-standard benchmarks, so-called shrinkage estimators of covariance matrix, proposed by Ledoit and Wolf. To compare the out-of-sample performance of the estimators we conduct experiments using both the real data from the New York Stock Exchange and generated samples. While conducting experiments a new surprising phenomena is observed: efficiency of obtained optimal portfolio is biased with respect to true optimal portfolio. Different aspects of this phenomenon will be discussed.

# Global search in problem with d.c. constraints

Alexander Strekalovsky

Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the RAS, Russian Federation

Method of exact penalty was invented by I.I. Eremin and W.I. Zangvill independently and almost simultaneously.

This approach has been widely used and nowadays becomes increasingly popular. It is viewed by specialists as a very powerful and effective tool for solving difficult real-life problems, including mathematical models of conflict situations (games, equilibria, bilevel problems, hierarchical control etc.).

On the other hand, almost all real-life optimization problems are explicitly or implicitly nonconvex, with a lot (often a huge number!) of stationary vectors and local pits. For such problems the classical optimization methods provide only the KKT points. On the other hand, the methods of Global Optimization (B&B, cut's method etc.) suffer the so-called "curse of dimension", when the exponential growth of computational efforts corresponds to an increase in dimension of the problem in question.

In this paper we continue to develop the apparatus of the so-called Global Optimality Conditions (GOC) for the nonconvex optimization problems, the goal function and inequality constraints of which are given by d.c. functions (the difference of convex functions).

This time we use popular techniques of the Exact Penalization to reduce the original problem to a nonconvex (penalized) problem without inequality constraints. For this purpose, we apply the well-known existence results for the threshold exact penalty parameter.

The current research develops the new GOCs for the problem in question with the help of the exact penalty approach and reduces the original problem to a problem without inequality constraints . After that, we represent the cost function of the penalized problem as a d.c. function. The latter helps us develop the necessary GOCs, which turn out to be related to the KKT Theorem in the original problem.

Then we identify and explain various features of the GOC developed, in particular, the constructive property. Besides, we use illustrative examples to demonstrate that the GOCs successfully escape local pits. Finally, we give the sufficient GOCs.

This research was supported by the Russian Science Foundation (project no. 15-11-20015).

# On the number of maximal independent sets in complete q-ary trees

# Dmitry Taletskii

Lobachevsky State University of Nizhniy Novgorod, Russian Federation

In a graph G, an independent set is a subset I of V(G) such that no two vertices of I are adjacent. A maximal independent set is an independent set that is not a proper subset of any other independent set. Denote by i(G) (respectively, by mi(G)) the number of all (respectively, maximal) independent sets in G. Denote by  $T_{\{q,n\}}$  the complete q-ary tree of height n.

P. Kirschenhofer, H. Prodinger, and R. Tichy proved the existence of constants  $a_q, a_{\{q,1\}}, a_{\{q,2\}}, b_q, a_{\{q,1\}} \neq a_{\{q,2\}}$ , such that for  $q \in \{2,3,4\}$  the asymptotic equality  $i(T_{\{q,n\}}) \sim a_q * (b_q)^{q^n}$  holds as n tends to infinity, and for any  $q \ge 5$  the asymptotic equalities  $i(T_{\{q,2k\}}) \sim a_{\{q,1\}} * (b_q)^{q^{2k}}, i(T_{\{q,2k+1\}}) \sim a_{\{q,2\}} * (b_q)^{q^{2k+1}}$  hold as k tends to infinity.

We investigate an asymptotic behavior of the parameter  $mi(T_{\{q,n\}})$ . Our main results are the following statements.

Theorem 1. There exist constants  $\alpha_2$ ,  $\beta_2$  such that the asymptotic equality  $mi(T_{\{q,n\}}) \sim \alpha_2 * (\beta_2)^{2^n}$  holds as n tends to infinity.

Theorem 2. For any sufficiently large q, there exist pair-wise distinct constants  $\alpha_{\{q,1\}}, \alpha_{\{q,2\}}, \alpha_{\{q,3\}}$  and a constant  $\beta_q$  such that the asymptotic equalities

$$mi(T_{\{q,3k\}}) \sim \alpha_{\{q,1\}} * (\beta_q)^{q^{3k}}, mi(T_{\{q,3k+1\}}) \sim \alpha_{\{q,2\}} * (\beta_q)^{q^{3k+1}},$$
  
$$mi(T_{\{q,3k+2\}}) \sim \alpha_{\{q,3\}} * (\beta_q)^{q^{3k+2}}$$
  
ends to infinity.

hold as k tends to infinity.

Numerical computations show that Theorem 2 holds for any q>10, and it does not hold for any  $3 \le q \le 10$ .

### Accounting for National Scientific Traditions in Bibliometric Network Analysis: Mapping Asian and African Studies in Russia

Alina V. Vladimirova, Alexey L. Ryabinin

#### Institute of Oriental Studies of the RAS, Russian Federation

Network analysis methods are good tools for exploring and mapping research fields that can help Russian scientific communities in their search for new principles of development and self-assessment. They not only give opportunity to highlight characteristics of network structures and associated problems, they also give insight into possible solutions. To demonstrate these possibilities, we have chosen an Asian and African studies field as our case. We introduce ways to account for national scientific traditions in bibliometric network analysis and thus increase quality of its results. The primary data were taken from the Russian Science Citation Index (RINC) which accumulates bibliographic information and texts of more than 2 million publications of Russian authors and then we focused on the Oriens, one of the most prominent national journals in regional studies. We aim to contribute both empirically and methodologically with this paper, as we see questions connected with reasonable scientometrics methods application in research on Russian science as an extremely important practical issue.

#### Using modular decomposition technique to solve the maximum clique problem

#### Irina Utkina

#### National Research University Higher School of Economics, Russian Federation

Today graphs can be used in different fields, such as biology, chemistry, data analysis, mathematics and others, as a structure of data. Due to enormous expand of information, the size of graphs for analysis is increasing, it can be hundred, thousand and even hundred of thousands vertices. Since the computational time of any algorithm on graphs depends on its size, it became a great problem for community. There are many graph decomposition techniques to reduce a graph problem to its smaller fragments; one of them is modular decomposition. In my work I use the fastest algorithm for constructing modular decomposition proposed by Tedder et. al. It creates modular decomposition tree for any input graph in linear time. Then I use this tree to solve the maximum clique problem on graphs from DIMACS benchmarks using Ostergard's algorithm, and compare computational time with Ostergard's algorithm without modular decomposition technique. I test this approach on DIMACS graphs, which have modules, and, for example, on "c-fat200-5". Ostergard's algorithm has found a solution in 2.74 sec, and my algorithm did it in 0.0013 sec.

#### Mapping the ontologies, methodological approaches and methods of sociopolitical research: application of social network analysis

### Dmitry Zaytsev

#### National Research University Higher School of Economics, Russian Federation

Why researchers that use statistics always propose hypothesis about associations between variables and try to prove cause-effect relations between them? And, vice versa, why political linguist tend to use discourse analysis to disclose cognitive manipulation by power-holders through the imposition of certain beliefs, which define the social behavior?

In this paper we propose the mapping of existed methods in social sciences, of methodological approaches which stand behind these methods and allow formulating research question, and of philosophical foundations or ontologies which explain what social reality is. Social research is driven not only by chosen methods, but also by approaches and ontology shared by researcher. Social researcher usually prefers to use methods related to one group of methodological approaches and associated with certain ontology. These can lead to the biased research results. E.g. if you use statistics and want to find cause-effect relation between the level of economic growth and democratization, you will find it as a lot of political scientists did analyzing comparative statistics by countries. Contrary conclusion, that democratization is the cause for economic growth, was made by other political scientists, who used the same statistical methods and comparative approach. In both examples researchers was driven by the positivist ontology that forced them to reveal regularities or 'social law' from systematize data about 'economic growth' and 'democratization'. Then, S. Huntington came to the conclusion that the democratization is the result of more complicated mechanism when economic growth is innovative, that lead to the raise of middle class, increase the level of citizens' education, that start demanding responsiveness from the government how do it spend state budget consist of the taxes that citizens paid. This conclusion became possible because Huntington based his research on the method of case-study and used historical approach based on the idealism as ontology. These allowed him to focus on the certain countries, not abstract processes of 'economic growth' and 'democratization', and deeply analyze them. He was searching and found 'ideal type' or 'ideal mechanism' when economic growth lead to democracy illustrating it on the examples of concrete countries. It is 'ideal' because operates with abstracts and dichotomies 'good-bad': everything that is good (innovations in economy, education and responsiveness lead to 'good' outcome – democracy, contrary to 'bad' extensive economy, illiteracy and closeness of elites that lead to autocracy).

These two examples showed us how do ontology shared by researcher implicitly defines his/her research focus, determines the choice of methodological approaches that already have standardly formulated research questions, and limit our analytical tools or methods. This 'path dependency' leads to the previously known results. We can break this vicious circle (when ontology defines approach and methods which lead

to certain results that was known before research from philosophical foundations of ontology) by identifying the most different clusters of related ontologies-approachesand-methods in social sciences to propose to the social researcher how to mix them to avoid such methodological trap.

Thus, we identify three levels: ontologies or paradigms (P), approaches (A) and methods (M) of social sciences. All three levels are connected with each other but uncertain and multiple ways. That is why between them appears complicated type of associations, which we propose to research with the use of network analysis.