Modified program

Thursday, May 26
Room 209 HSE, 136 Rodionova Str.

09:00 – 09:30 Registration
09:30 – 09:50 Panos M. Pardalos
The Sixth International Conference on Network Analysis NET 2016
09:50 – 10:40 Rene van Bevern
Exploiting demand structure for efficiently serving arcs and edges in networks
10:40 – 11:00 Coffee Break
11:00 – 11:50 Fuad Aleskerov
Power in network structures
11:50 – 12:30 Session 1 (2 talks)
Dmitriy Malyshev
Polynomial-time Approximation Algorithms for the Coloring Problem in Some Hereditary Classes
Andrey Savchenko
Approximate nearest neighbor search in unconstrained face recognition problem with deep neural network features
12:30 – 14:00 Lunch Break
14:00 – 14:50 Oleg Prokopyev
Learning Theory on Graphs
14:50 – 15:10 Coffee Break
15:10 – 16:10 Session 2 (3 talks):
Sergey Makrushin
Analysis of Structure of the Power Transmission Grid in Russia: Evaluation of the Small World Model Applicability
Andrey Rogatkin
Simulation modeling of conformity social behavior model using Facebook Livejournal and Twitter graphs
Ilya Makarov, Oleg Bulanov, Leonid E. Zhukov
Co-author Recommender System
16:10 – 16:30 Coffee Break
16:30 – 17:50 Session 3 (4 talks):
Yulia Dodonova, Leonid Zhukov
Machine learning application to human brain network studies
Nikolay Bazenkov
Double best response as a network stability concept
Mikhail Sivykh
Talent scheduling problem on GPU
Ilya Kurochkin
The comparative analysis of various criteria of routing in telecommunication networks

Friday, May 27
Room 209 HSE, 136 Rodionova Str.

09:30 – 10:20 Sergey Sevastyanov
Efficient analysis of networks with cycles
10:20 – 10:40 Coffee Break
10:40 – 11:30 Valentina Kuskova
Network methods applied: Newest empirical results from network studies in Russia
11:30 – 12:30 Session 4 (3 talks):
Sergey Shvydun
Network Analysis of International Migration
Daria Maltseva
Network studies in Russia: from articles to the structure of a research community
Kolesnik N.A., Kuskova V.V., Tretyak O.A.
Application of network analysis for FMCG distribution channels
12:30 – 14:00 Lunch Break
14:00 – 14:50 Sergey Nikolenko
Buffer Management with Heterogeneous Processing
14:50 – 15:10 Coffee Break
15:10 – 16:10 Session 5 (3 talks):
Theodore Trafalis
Max-Min Fairness (MMF) Optimization Models for Networks
Alexander Lepskiy
An Analysis of Coherence of Financial Analysts' Recommendations in the Framework of the Belief Function Theory
Koldanov Petr
Testing the symmetry of multivariate distribution of stock returns
16:10 – 16:30 Coffee Break
16:30 – 17:50 Session 6 (4 talks)
Vladisvlav Chesnokov
Overlapping community detection in social networks with partly missing node attributes
Aleksey Katerov
Flow-based bounds for the three-index axial assignment problem
Pavel Chebotarev, Vladimir Ivashkin
Logarithmic proximity measures outperform plain ones in graph nodes clustering
Mikhail Goubko
Generalized Huffman Trees and Wiener Index for Graphs with Weighted Vertices
18:00 Reception (room 102)

Saturday, May 28
Room 209 HSE, 136 Rodionova Str.

09:30 – 10:20 Maxim Jukovsky
Random graphs: models and asymptotic characteristics
10:20 – 10:40 Coffee Break
10:40 – 12:20 Session 7 (5 talks)
Alexandr Rubchinsky
A new approach to network decomposition problems
Mikhail Goubko, Vassily Ginz
Improving Spectral Partitions for Optimal Controlled Islanding of Power Grid
Mikhail Goubko, Dmitry Ignatov, Alexey Neznanov
Bayesian Preference Learning for Residential Demand Response in Power Grids
Dmitriy Ignatov
Towards a lattice-based algorithm for consensus clustering
Dmitry Gribanov, A. Y. Chirkov
Computing of the Width of Simplices With Bounded Minors of the Constraints Matrix
12:20 – 14:00 Lunch Break
14:00 – 15:00 Session 8 (3 talks)
Nikolay Karpov
Comparative Analysis of Two Iterative Methods Using Quantification Library
Alexey Goryachih
One class of smooth modification of Space-Filling Curves
Dmitry Mokeev
On the equality of matching and cover numbers for P4 in split graphs and their extentions
15:00 – 15:20 Coffee Break
15:20 – 16:40 Session 9 (4 talks)
Yury Malkov
Similarity search by using hierarchical navigable small world graphs
Evgeny Kozinov
Accelerating Multicriterial Optimization by the Intensive Exploitation of Accumulated Search Data
Kirill Kniazev
Applying the semantics to Dijkstra's algorithm for parametric routing in Indoor navigation systems
Boris Mirkin, Ivan Rodin
Composite dynamic superclusters for the Analysis of Reference Graphs

Power in Network Structures

Fuad Aleskerov, Natalia Mescheryakova, Sergey Shvydun, alesk@hse.ru

National Research University Higher School of Economics
Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation

We consider an application of power indices, which take into account preferences of agents for coalition formation proposed for an analysis of power distribution in elected bodies to reveal most powerful (central) nodes in networks.

These indices take into account the parameters of the nodes in networks, a possibility of group influence from the subset of nodes to single nodes, and intensity of short and long interactions among the nodes.

Some properties of the indices are discussed.

Various applications are presented for the network of countries – those of religion, migration, foreign claims, conflicts, exchange of students, food export-import, etc.

Double best response as a network stability concept

Nikolay Bazenkov, n.bazenkov@yandex.ru

Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation

In network formation games the key point is what network should be considered as a stable solution. We investigate a concept of network stability called double best response by analogy with conventional best response dynamics used in network formation games. Here we apply double best response to the game called minimal-cost connectivity game where every player wants to be connected to as many players as possible with minimal individual cost. The network formation process consists of a sequence of two-stage local adjustments. On the first stage a randomly selected ``leader'' agent chooses a subset of her neighbors and proposes to everyone of them to form a link. On the second stage each ``follower'' decides to accept or decline the proposal. A situation when no leader can make a proposal that will improve her payoff called equilibrium in double best responses (EDBR). Next we investigate the properties of networks stable according to EDBR concept.

First, we analyze the setting where every player has full information about the current network structure. We show that if players are located on a plane then every EDBR network is a subgraph of the Relative Neighborhood Graph (RNG) which is a well-known structure in wireless networks topology control literature. Second, we analyze players which possess only local information about the network structure and show that the “RNG proeprty” still holds if during the local adjustment the ``leader'' is better informed than the ``followers''. We obtain the bottom and top bounds on the efficiency of EDBR networks and show that EDBR produce networks that are significantly more efficient than conventional Nash equilibrium, pairwise stability or strong stability concepts.

Exploiting demand structure for efficiently serving arcs and edges in networks

René van Bevern, rvb@nsu.ru

Novosibirsk State University, Novosibirsk, Russian Federation

We study the problem of finding minimum-length routes for a fleet of vehicles of equal capacity that are initially located in a vehicle depot and have to serve the demands of all clients located in a transportation network. It is canonical to model the transportation network as a graph whose edges are weighted by distances. Yet there is some freedom in modeling the clients: in the Capacitated Vehicle Routing Problem (CVRP), the clients are a subset of the vertices of the graph, whereas in the Capacitated Arc Routing Problem (CARP), the clients are a subset of the arcs or edges of the graph. CARP models tasks where the roads of the network themselves are the clients or all clients along a road have to be served (e.g., in street sweeping, household waste collection, or meter reading).

Both CVRP and CARP are NP-hard and can easily be translated into each other. Due to this perceived equivalence, research is mostly concentrated around the older CVRP. Nevertheless, solving route planning tasks using CARP rather than CVRP, when applicable, is rewarding: based on a recent survey and recent work [1, 2] on the parameterized and approximation complexity of arc routing problems, this talk shows what makes CARP significantly easier than CVRP and presents recent theoretical and experimental results on fixed-parameter algorithms for efficiently solving NP-hard variants of CARP by exploiting the structure of the subgraph that is induced by the client arcs or edges.

References:

1. René van Bevern, Rolf Niedermeier, Manuel Sorge, and Mathias Weller. Complexity of arc routing problems. In *Arc Routing: Problems, Methods, and Applications*, MOS-SIAM Series on Optimization. SIAM, 2014.
2. René van Bevern, Christian Komusiewicz, and Manuel Sorge. Approximation algorithms for mixed, windy, and capacitated arc routing problems. In *Proceedings of the 15th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'15)*, volume 48 of *OpenAccess Series in Informatics (OASIcs)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2015.

Logarithmic proximity measures outperform plain ones in graph nodes clustering

Pavel Chebotarev, **Vladimir Ivashkin** pavel4e@gmail.com

Institute of Control Sciences of the Russian Academy of Sciences

This is a joint paper with Vladimir Ivashkin (MIPT).
We consider a number of graph kernels and proximity measures: commute time kernel, regularized Laplacian kernel, heat kernel, communicability, etc., and the corresponding distances as applied to clustering nodes in random graphs. The model of generating graphs involves edge probabilities for the pairs of nodes that belong to the same class or different classes. It turns out that in most cases, logarithmic measures (i.e., measures resulting after taking logarithm of the proximities) perform much better in distinguishing classes than the “plain” measures. A direct comparison of inter-class and intra-class distances confirms this conclusion. A possible explanation of this fact is that most kernels have a multiplicative nature, while the nature of distances used in cluster algorithms is an additive one (cf. the triangle inequality). The logarithmic transformation is just a tool to transform one nature to another. Moreover, some distances corresponding to the logarithmic measures possess a meaningful cutpoint additivity property. In our experiments, the leader is the so-called logarithmic communicability measure, which distinctly outperforms the “plain” communicability.

Overlapping community detection in social networks with partly missing node attributes

Vladisvlav Chesnokov, v.o.chesnokov@yandex.ru

Bauman Moscow State Technical University, Moscow, Russian Federation

Community detection is widely used in Data Mining, biology, sociology, telecommunications, transport and other sciences. For example, in social networks, nodes can represent people, while edges can be relationships between them: friendship, family ties, working relations, etc and so communities are groups of people having something in common. One of the main features of social graphs is that nodes have many attributes associated with them like age, gender, favorite movie, workplace, hometown, etc. Only few community detection algorithms use the plethora of this information. But the full information about a person's attributes can be obtained rarely due to some restrictions: privacy issues or data loss.

A fast algorithm for overlapping community detection is proposed. It uses information about node attributes, requires neither training nor knowledge about their nature and is tolerant to their partial absence.

There are three assumptions behind the proposed algorithm. The first states that social graphs are formed by affiliation model [1]. Simply speaking, communities define graph structure and not vice versa. Moreover, each community has an associated attribute and each vertex in community has this attribute. The second assumption is triadic structure of social networks [2], i.e. if there is relation between persons A and B and between B and C, then, with high probability, there is the same relation between A and C. The last hypothesis is related to the first one and states that nodes which are close to each other have similar attribute set [3].

The developed algorithm is based on attribute transfer and consists of five steps. Each node is initialized with empty set of attributes --- they are called key attributes further on. The first step is iterative. At each iteration, all nodes of graph are visited and their key attributes sets are updated by the following rule: if the qualified majority of vertex neighbours has common attribute, then it is added to key attributes set for this vertex. The step is finished when there were no updates at last iteration. Due to small world phenomenon, the number of iterations is small: during experiments it was less than 10 for most of graphs. The same procedure with weakened addition rule is performed at the second step. Nodes which belong to several communities or lie on the fringe are attached to existing communities. After that, nodes are joined into communities by attributes and they are split into connected components. The fourth step is merging: components with the same node set are joined into single community. So, each community has a set of attributes which it was possibly formed by. The last step is slightly modified procedure of first and third steps applied to subgraph of vertices with empty key attribute set. Vertices in this subgraph either belong to some community formed by unspecified attribute or does not belong to any community. Communities from this step are added to set obtained at step four. Summing up, the algorithm has near-linear runtime: each iteration of the first and the second steps is linear in graph size because each edge visited at most twice. The third step is finding connecting components, which can easily be done in linear time or less. The fourth step is also solved in linear time with hashes. Moreover, the algorithm can be easily parallelized for large graphs.

The developed algorithm was tested on Facebook and Twitter datasets with ground-truth hand labeled communities from SNAP [4]. It was compared against five algorithms: modularity maximization, Infomap, AGM-fit, BigCLAM and CESNA. Also, tolerance to partial node attribute absence was tested. Experiments showed that the developed method outperforms all competitors by F1-score and Jaccard index by more that 10%. The algorithm still gives high scores when up to 80% of node attributes values are deleted from graph.

References:

1. Yang, J., Leskovec, J.: Community-affiliation graph model for overlapping network community detection. -- In: 12th {IEEE} International Conference on Data Mining, {ICDM} 2012, Brussels, Belgium, December 10-13, 2012, pp. 1170--1175.
2. Granovetter, M.: The strength of weak ties. American Journal of Sociology -- 78(6), 1973.
3. Dougnon, R.Y., Fournier-Viger, P., Nkambou, R.: Advances in Artificial Intelligence: 28th Canadian Conference on Artificial Intelligence, Canadian AI 2015, Halifax, Nova Scotia, Canada, June 2-5, 2015, Proceedings, chap. -- Inferring User Profiles in Online Social Networks Using a Partial Social Graph, pp. 84--99. -- Springer International Publishing, Cham, 2015.
4. Leskovec, J., Krevl, A.: SNAP Datasets: Stanford large network dataset collection. -- http://snap.stanford.edu/data.

Machine learning application to human brain network studies

Yulia Dodonova, Leonid Zhukov
ya.dodonova@mail.ru, leonid.e.zhukov@gmail.com

National Research University Higher School of Economics, Moscow, Russian Federation
The Institute for Information Transmission Problems Russian Academy of Sciences

Network science is becoming a popular instrument for neuroscience research. At the macroscopic scale, aggregated neural pathways can be modeled by graphs called connectomes. These graphs are usually small and contain about 100 vertices labelled according to brain regions; a set of uniquely labeled vertices is the same across different connectomes. The edges are weighted and represent either structural connections (the number of fiber tracts) or functional relations (co-activation) between brain regions. The networks are spatially embedded: vertices are localized in 3D space, and edges have physical lengths. These latter characteristics are rarely incorporated in analysis, although it is sometimes acknowledged that differences in physical size pose additional requirements on normalization of individual connectomes.

Network brain characteristics are expected to provide insights into associations between brain structure and particular phenotypes. More practical questions are whether connectomes can be useful in discriminating between normal and pathological brain structure (which is a classification task) and predicting the onset of disease or treatment outcomes (which can be considered a regression task). However, a typical study is based on a sample of only about a hundred of connectomes and the above questions are mostly addressed in terms of statistical significance of group differences in network characteristics.

In this paper, we use machine learning algorithms to predict phenotypes from brain structure. We use graph kernel approach and run SVM classifier on precomputed matrices of pairwise distances between connectomes. We also explore whether normalized Laplacian spectra of connectomes can be useful for our classification tasks. We compare our results against the performance of classifiers that use vectors of features, such as “bag-of-edges” (vectorized weighted adjacency matrices), weighted degrees, and a set of global network metrics. Finally, we discuss how information on physical properties of the networks can be incorporated in analysis, in a form of both normalization and construction of the appropriate null networks.

One class of smooth modification of Space-Filling Curves

Alexey Goryachih, ***goryachihalexeysergeevich@gmail.com***

Lobachevsky State University, Nizhny Novgorod, Russian Federation

This work presents one class of smooth modification of Space-Filling Curves. These modifications make approximations of Hilbert Curves differentiable in all points, and save differentiability of an optimized function. Some results of applications the modifications and comparison with unmodified curve are presented in this work.

Generalized Huffman Trees and Wiener Index for Graphs with Weighted Vertices

Mikhail Goubko, mgoubko@mail.ru

Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation

For a simple undirected connected graph G the Wiener index WI(G) is calculated as a half of the sum of elements of its distance matrix D(G). WI(G) is one of the most renowned graph indices widely used in mathematical chemistry, quantitative structure-property relation (QSPR) studies, and network analysis. In 1997 Klavzar and Gutman defined the Wiener index for vertex-weighted graphs as a quadratic form VWWI(G, mu) = 1/2**mu*'*D(G)**mu*, where *mu* is the vector of positive vertex weights. VWWI generalizes WI and is closely connected to spectral properties of the graph.

We minimize VWWI over the set of trees with the given vertex weights' and degrees' sequences and show the optimal tree to be the, so-called, generalized Huffman tree built in a bottom-up manner by sequentially connecting vertices of the least weights to vertices of the least degree. Chemical trees are constructed, which minimize the Wiener index over all chemical trees with given vertex weights. It is also conjectured that the tree with given vertex weights' and degrees' sequences that maximizes VWWI is a “reversed” Huffman tree, which is built by sequentially connecting the heaviest vertices to vertices with the biggest degree.
Several applications are discussed. Firstly, we find isomers of simple alcohols with extremal normal boiling point by minimizing the linear combination of VWWI and vertex-degree-based graph indices. Secondly, we suggest a lower bound of the routing (communication) tree cost.

Bayesian Preference Learning for Residential Demand Response in Power Grids

Mikhail Goubko, Dmitry Ignatov, Alexey Neznanov, mgoubko@mail.ru

Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation
Skoltech Center for Energy Systems,
Skolkovo Institute of Science and Technology, Skolkovo, Russian Federation

In the coming years residential consumers will face real-time electricity tariffs with energy prices varying day to day. They are assumed to provide stimuli for demand response (consumption peak shaving through load shift or curtailment). Nevertheless, efficient electricity saving under real-time price schedules requires enormous efforts and requires advanced automation (a recommender system or an automatic control system). To perform well, these systems have to learn consumer's preferences from her actions. Under the assumption of rational (economic) behavior a consumer chooses a scenario of home appliance use to balance the energy bill and the consumer's comfort level. We propose a Bayesian learning algorithm to estimate the comfort level function of a residential appliance user from the history of her actions in different situations. In numeric experiments with datasets generated from a simulation model of a rational consumer interacting with small home appliances our algorithm outperforms popular contemporary regression analysis tools giving more accurate predictions of consumer's actions. This approach can be extended and used to control an air heating and conditioning system, which is responsible for up to half of a household's energy bill.

Improving Spectral Partitions for Optimal Controlled Islanding of Power Grid

Mikhail Goubko, Vassily Ginz, mgoubko@mail.ru

Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation
Skoltech Center for Energy Systems,
Skolkovo Institute of Science and Technology, Skolkovo, Russian Federation

Controlled islanding is a part of frequency stabilization strategy in power grids. It implies partitioning a grid into independent islands in case of massive failures to minimize accident aftermaths, prevent cascading blackouts and keep critical infrastructure alive. To achieve these goals slow coherence of generators must be avoided in the islands formed. Also, line switch off distortion and load shedding must be minimized. Existing approaches account for just one or two of these factors. We propose a two-step islanding algorithm that balances multiple criteria: slow coherence of generators, minimal line switching-off distortion, island size, and minimum load shedding. The mathematical challenge is high dimension (up to 10^4 nodes in a grid) and dependence of cost function on power flow directions (so, spectral clustering is inapplicable).

Several hierarchical spectral clustering schemes are used at the first step to cut the problem dimension (caring for coherence and distortion only), and CPLEX tools for the mixed integer quadratic problem are employed at the second step to choose a balanced partition of the aggregated grid that minimizes a combination of coherence, distortion and load shed (approximated by power imbalance). The problem dimension on the second step depends only on the desired number of islands K but not on the dimension of the initial grid. A greedy heuristic generates a starting solution for the branch and bound scheme assuring high performance of the second step. The algorithm was compared with the basic one-step hierarchical clustering scheme on example grids with 118, 2383, and 9241 nodes, showing good performance (10% average improvement of the cost function) with modest increase in computation time (<1 second on the average). Constrained AC OPF solution was used to estimate load shed for benchmarking.

Computing of the Width of Simplices With Bounded Minors of the Constraints Matrix

D. V. Gribanov, A. Y. Chirkov, dimitry.gribanov@gmail.com

National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

We will show that the width of simplices defined by systems of linear inequalities can be computed in polynomial time if some minors of their constraint matrices are bounded. Additionally, we present some quasi-polynomial-time and polynomial-time algorithms to solve the integer linear optimization problem defined on simplices minus all their integer vertices assuming that some minors of the constraint matrices of the simplices are bounded.

Towards a lattice-based algorithm for consensus clustering

Dmitry I. Ignatov, dignatov@hse.ru

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

We propose a new lattice-based algorithm for consensus clustering. As the input the algorithm takes n partitions of a certain set of objects obtained by k-means algorithm after its n different executions. The resulting consensus partition is extracted from a (partial) antichain of the concept lattice built on formal context objects×classes, where the classes are the set of all cluster labels from each initial k-means partition. We compare the results of the proposed algorithm in terms of ARI measure with the state-of-the-art algorithms on synthetic datasets and in several cases the best measure values are demonstrated by our approach.

Comparative Analysis of Two Iterative Methods Using Quantification Library

NIkolay Karpov, nkarpov@hse.ru

National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

In many areas, such as social science, politics or market research, people need to track sentiment and their changes over time. For sentiment analysis in this field it is more important to correctly estimate proportions of each sentiment expressed in the set of documents (quantification task) than to accurately estimate sentiment of a particular document (classification). Basically, our study was aimed to analyze the effectiveness of two iterative quantification techniques and to compare their effectiveness with baseline methods. All the techniques are evaluated using a set of synthesized data and the SemEval-2016 Task4 dataset. We made the quantification methods from this paper available as a Python open source library. The results of comparison and possible limitations of the quantification techniques are discussed.

Flow-based bounds for the three-index axial assignment problem

Aleksey Katerov, lexus_690@mail.ru

Lobachevsky State University, Nizhny Novgorod, Russian Federation

We consider the three-index axial assignment problem. These problem belongs to the class of integer multi-index transport type problems known to be NP-hard. One of the approaches to solve these problems is based on reducing them to the network flow problem. This approach made it possible to find new subclasses of the integer multi-index transport type problems that could be solved in polynomial time. In this work we propose algorithms for constructing the upper and lower bounds of the three-index axial assignment problem based on flow methods of multi-index problems solving.

Derived bounds can be used for solving three-index axial assignment problems using the branch and bound algorithm. Computational experiments was carried out and it showed prospects of proposed bounds.

Applying the semantics to Dijkstra's algorithm for parametric routing in Indoor navigation systems

Kirill Kniazev, kk@netbell.ru

National Research University Higher School of Economics / NETBELL LLC

Dijkstra’s algorithm is considered to be one of the most popular algorithms for finding the shortest path in a directed graph with non-negative weights, but what modifications should it undergo to meet the real requirements of navigation services with semantic usage?

This presentation is devoted to the implementation of Dijkstra’s algorithm into the indoor-navigation service with semantic layer and reveals several significant steps in modifying its working logic as well as building the usage environment.

To make the topic of the presentation feasible in practice, the content of the work starts with a brief description of the software architecture environment and identifies a set of criteria for the successful implementation on a live example of existing service altera.me.

The main body of the presentation is devoted to the techniques, applicable to improve the working algorithm, such as local search, zones caching and so on. This part of the work also includes the description of implementation process for every method, some live examples for each of them and the analysis of the overall effect on the enhanced algorithm performance.

Methods mentioned during the presentation can significantly increase the performance of Dijkstra's algorithm due to the semantic layer usage and parametric search in entire system. Connecting semantics to the Dijkstra’s algorithm opens new opportunities given by understanding the meaning of things on a map.

Try to imagine the system, which indicates the best route for you based on your capabilities and needs, based on current lifts’ position in building, enabling you to estimate the time for reaching the destination point or find the nearest free audience, build a route only through the lifts or create a walking route through the street, find out the most visited places or easily get you to the room you should like basing on your interests scope.

All these features can be implemented into the working process of Dijkstra’s algorithm, as well as semantics can affect its effectiveness. Current presentation is based on the existing indoor navigation service Altera.me and covers its future improvements.

Testing the symmetry of multivariate distribution of stock returns

Petr Koldanov, pkoldanov@hse.ru

National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

Models of multivariate distributions of stock returns attract important attention in theoretical and applied sense. In particular multivariate distribution models are necessary for portfolio selection and risk management. One of the popular multivariate model in financial analysis is the class of elliptically contoured distributions. Consistency of real data with elliptically contoured model was studied in paper of Chicheportiche & Bouchaud [2012] where it was shown that the joint distribution of real market stock returns is not in accordance with hypothesis of elliptical distributions. However, as pointed by the authors, their methodology differs from usual hypothesis testing using statistical tools. In the present paper we analyze this problem from multiple hypotheses testing theory point of view see Lehmann [2005]. Let Xi is a random variable, corresponding to the return of stock i (i =1;N) and X = (X1; : : : ;XN) be the random vector of stock returns. To study the consistency of real data with elliptically contoured model for X we use one important property of elliptically contoured distributions, namely, the symmetry of density function of joint distribution with respect to the vector of means. To formulate the individual hypotheses we use the pairwise sign symmetry property of multivariate elliptical distribution. In the paper, we construct multiple test which control FWER (Family-Wise Error Rate) for simultaneous individual hypotheses testing

For individual hypotheses testing, we construct tests of a Neyman structure based on sign statistics. Simultaneous inferences are conducted by Holm procedure, which is known to control FWER.

Resulting distribution free multiple test procedure was applied for USA and UK stock markets for different periods of observations. In most cases pairwise sign symmetry hypotheses are not rejected. For the case where sign symmetry hypotheses are rejected for some pairs of stocks, rejection graph has an surprising structure. Despite number of rejected hypotheses the rejection graph has only few hubs and their removing leads to acceptance of all remaining symmetry hypotheses.

Application of network analysis for FMCG distribution channels

Kolesnik N.A., Kuskova V.V., Tretyak O.A., nkolesnik@mail.ru

National Research University Higher School of Economics
International Laboratory for Applied Network Research & Laboratory of Network Organizational Forms (Faculty of Business and Management), Moscow, Russian Federation

In the age of internet sales and megamarkets, do direct marketing and personal sales still have value? Though the subject of direct sales effectiveness has been on the forefront of marketing research for a long time, with network analytics playing a large role in the development of this research area (e.g., Möller et al., 1995; Palmer & Koenig-Lewis, 2009), this question remains largely unanswered. Some studies (e.g., Grewal et al., 2004) have found that internet has replaced a lot of direct sales, while others (Kiang et al., 2000) are not as conclusive. We believe the reason for this lack of consensus, and the respective answer to the very important research question of direct sales effectiveness, is in the level of analysis. To the best of our knowledge, all studies to date looked at internet vs. personal sales either on the level of the distributor, or on the level of the company's financial results, but nobody has looked at it at the level of tactics employed by each individual distributor, which may or may not include internet as a distribution channel.

Our research is designed to answer this important question by using a combination of different tactics as a level of analysis. Each combination of channels is considered an omni-channel tactic of a salesperson (friends, acquaintances, telephone, home presentations, printed advertisement, internet, e-mail, door-to-door). For example, a tactic with a code 11111111 means the usage of all possible channels by a distributor. With eight possible channels to be used or not used by a distributor, we have a potential of 256 combinations or unique tactics, which serve as one of the modes of a three-mode affiliation network. Distributors and individual channels are the other two modes. We use a visual representation of the resulting network of tactics and formal statistical analysis to investigate how these structures differ across companies and in time. More importantly, we plan to address the relative effectiveness of internet vs. direct sales in a rigorous comparative manner not done previously.

Data for research were collected jointly with the Russian Direct Selling Association (www.rdsa.ru), a member of World Federation Direct Selling Association (WFDSA). The first wave of data was collected in spring 2011, the next one in spring 2014. The sampling frame included 5694 independent sellers from 16 biggest direct selling companies in Russia. The questionnaires were spread among companies with RDSA membership, the sample was allocated proportionately to the number of the distributors. The companies are focused on distribution of fast moving consumer goods (FMCG): personal care products, nutrition, perfumes, cosmetics, hygiene products, jewelry, and accessories.
We use graphs to analyze structure and relationships of tactics within the company. To better understand the structure and effectiveness of the company's distribution network, we examine the effects of the network structure on the performance. About 190 tactics are used by more than one salesperson. In addition, different channels within each tactic can be used by distributors with different frequency. Frequency of channel usage is measured by a four-point scale ranging from 0 to 3, "never use", "rarely", "occasionally", and "frequently." The Cronbach alpha value of the scale is 0.77,

indicating satisfactory internal reliability. In our research, the “channel frequency use” for each node is described by a vector, for example (3;0;1;3;0;2;2;0) or (2;0;1;3;0;1;2;0).

The graph is edge-weighted, and nodes have different attributes (e.g., demographics for distributors, working experience, education). Weight of the edge shows the difference between tactics in channel structure and frequency of channels usage. Weight of the edge is measured by Euclidean norm, which assigns to each vector the length of its arrow. For example, tactics are closer to each other on the graph if they have more similar channels with same intensity.

Company success strongly depends on effective management of distributor’s network. We expect that the above network analysis tools will provide us with a deeper understanding of idiosyncrasies of distribution strategies between different companies, and moreover, answer the important question of relative effectiveness of various distribution channels.

References:

1. Grewal, Dhruv, Gopalkrishnan R. Iyer, and Michael Levy. ""Internet retailing: enablers, limiters and market consequences."" Journal of Business Research 57, no. 7 (2004): 703-713.
Kiang, Melody Y., T. S. Raghu, and Kevin Huei-Min Shang. ""Marketing on the Internet—who can benefit from an online marketing approach?."" Decision Support Systems 27, no. 4 (2000): 383-393.
2. Möller, Kristian K., and David T. Wilson, eds. Business marketing: An interaction and network perspective. Springer Science & Business Media, 1995.
Palmer, Adrian, and Nicole Koenig-Lewis. ""An experiential, social network-based approach to direct marketing."" Direct Marketing: An International Journal 3, no. 3 (2009): 162-176.

Accelerating Multicriterial Optimization by the Intensive Exploitation of Accumulated Search Data

Evgeny Kozinov, evgeniy.kozinov@gmail.com

Lobachevsky State University, Nizhny Novgorod, Russian Federation

The decision-making problems are widespread in the scientific and technological human activities. Such problems as the optimal arrangement of the elements of the integrated circuits, aircraft design, medical drug development, etc. can serve as the examples of the applied problems, which optimal decision-making is necessary. All the problems mentioned above can be represented as the global optimization ones. It is worth noting that the above problems are the multicriterial ones, as a rule.

For solving the multicriterial problems, it is necessary to apply the algorithms utilizing efficiently all the computation resources and accumulated data. The methods for reuse of the accumulated search data when altering the problem statement are considered in the paper. The reuse of the accumulated search data allows managing the search process and increasing the efficiency of solving the problems when altering the search parameters. The proposed approach is illustrated by numerical results.

The comparative analysis of various criteria of routing in telecommunication networks

Ilya Kurochkin, qurochkin@gmail.com

Institute for Information Transmission Problems Russian Academy of Sciences, Moscow, Russian Federation

Rational use of resources of telecommunication networks – one of actual tasks in connection with continuous growth of load of data transmission networks. In this regard the assessment and the analysis of various parameters of a network, and also development of new routing algorithms is necessary.

The mathematical model of functioning of a telecommunication network by means of which it was possible to carry out the comparative analysis of various algorithms of routing has been created. Calculation of route for all these heuristic algorithms is made as definition of the minimum cost path. Distinction between algorithms consisted in different ways of determination of these additional costs of arcs. The following approaches of routing on the network graph have been considered: the shortest path, economy of arcs with a small capacity, economy of arcs, belonging to the minimum cuts, uniform use of various routes.

Results of experiments on comparison of efficiency of heuristic algorithms of the routing directed on the maximum increase in the general traffic on networks are presented. Results of the comparative analysis of the realized heuristic algorithms of routing are given.

An Analysis of Coherence of Financial Analysts' Recommendations in the Framework of the Belief Function Theory

Alexander Lepskiy, alex.lepskiy@gmail.com

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

This report is devoted to the analysis of coherence of financial analysts' recommendations with respect to securities of the Russian companies. The study is based on the analysis of approximately 4000 recommendations and forecasts of 23 investment banks with respect to around forty securities of Russian stock market over the period of 2012-2014 years. The predictive history of each of the investment bank was considered as evidence in the framework of the belief function theory. The coherence of recommendations was evaluated with the help of the so-called conflict measure between the evidence, which determined on the subsets of the set of all evidence. Then the analysis of coherence was reduced to analysis of values of the conflict measure. This analysis was performed with the help of game-theoretic methods (Shapley index, interaction index), network analysis methods (centralities), fuzzy relation methods, hierarchical clustering methods.

Co-author Recommender System

Ilya Makarov, Oleg Bulanov, Leonid E. Zhukov
iamakarov@hse.ru, oleggl500@gmail.com, leonid.e.zhukov@gmail.com

National Research University Higher School of Economics,
School of Data Analysis and Artificial Intelligence, Moscow, Russian Federation

Modern bibliographic databases contain significant amount of information on publication activities of research communities. Selecting a co-author for joint research publication is one of the challenging task that researchers encounter regularly. In this paper we analyse historical publication data and propose a new recommender system for co-author selection. Our system is based on co-author graph derived from the bibliographic database as well as content information obtained from Scopus and other publication ranking systems. We formulate the recommendation problem as a link prediction within the co-author graph.

We started by taking a relational database of all publications written by NRU HSE authors and cleaning it by removing duplicate records, inconsistencies, unifying author identifiers and filling in missing data. We then constructed an undirected co-authorship graph with authors as graph nodes and edges containing lists of jointly published papers. We added all publication attributes from the university portal and journal information from Scopus SJR Ranking including subject areas and categories. Publication quality was taken from its quartile SJR rank in Scopus for the publication year, computed as maximal (or average) over different categories per journal. If a journal is not indexed in Scopus then its quartile is defined as a special rank Q5. We also added information about authors administrative units within the university and declared research interests.

We then used this graph to answer a variety of questions about research collaboration patterns: distribution of number of papers written by an author and distribution of number of collaborators, density of graph communities, dependency on research area and administrative staff subdivisions, evolution of collaborations over time etc.

We also constructed many features needed for the predictive model from this graph: node degrees, local clustering coefficients, node similarities etc. We generated additional features from the sequence of induced subgraphs, constructed by removing edges containing less than `k' publications. This procedure also removed nodes that became disconnected, thus reducing the number of authors by approximately 10 times from 19000 to almost 1600 for the k=5 graph. The number of connected components in this case was also reduced from 1000 to 400, but a clustering coefficient only decreased from 0.63 to 0.54. Then the average number of collaborators changed from 6.8 to 2.3. This procedure also ""shrinked"" the graph, decreasing average shortest path from 7.4 to 3.0. We used these smaller graphs to detect clusters by applying various community detection methods. The discovered clusters then became seed clusters for the original graph. Node assignment to these clusters became one of the strong features in the predictive model.

Since we use machine learning approach and our feature space contains many content based features, the recommendation system can be used for new authors, who do not have any connections to HSE community. In this case, the input will consists of authors research interests, publications and his university department, and the system will provide a rank list of proposed HSE candidates for the co-authorship. The system can also be used to estimate collaboration between university departments.

Analysis of Structure of the Power Transmission Grid in Russia: Evaluation of the Small World Model Applicability

Sergey Makrushin, SVMakrushin@fa.ru

Financial University under the Government of the Russian Federation, Moscow, Russian Federation

The subject of the current research is the United national energy system (UNES) – Russia's power transmission grid, whose major part is managed by the Federal Grid Company of the United energy system (FGC UES Inc.). The aims of the current research are to find key network topology properties and identify the network model allowing to contrast it to other transmission power networks The research is mainly focused on the UNES applicability in small world networks the a and finding appropriate methods for the relevant analysis. The research methods are based here on complex network theory, while the UNES computer model is applied to carry out the analysis. Though network theory methods are widely used for power transmission grid analysis in foreign countries, whereas to date they haven't been employed for Russian UNES.

Within the analysis of the UNES by complex theory methods UNES analysis the power transmission grid is seen as a network with electrical power substations and power stations as network nodes and high-voltage power lines as network links. For the current research the data about 514 nodes and 614 links have been gathered. This information includes network topology, geographic coordinates for nodes, region binding, voltage levels and other properties. The data for computer model creation have been gathered from official papers, online UNES map of the FGC UES, OpenStreetMap GIS. The UNES computer model was created for main operating regions of the UNES except the united national energy systems of Siberia, the East, and Urals (partially).

Computer model processing and algorithm development have been performed with Python programming language and *networkx* library. Basic properties of the network have been found using the computer model of the UNES. In particular, average node degree, average path length, overall clustering have been calculated (see Table 1).

In most papers on the analysis of power transmission grids with complex network theory methods, small world model applicability is considered. The issue is of great interest due to the fact that the small world model is one of the milestones of network theory.
It is seen crucial for capability of fast diffusion of information (or any other process) in a network is among the major properties of small world networks. In relation to electrical network it means that small world networks are fragile to cascade failure effects and it forces importance of that model for power grids network analysis. Although much research on applicability of the small world model to national level transmission grid has been done, there is no consolidated opinion on the subject.

One of the causes for the difficulties is the absence of consolidated opinion is using inappropriate methods for identification of small world topology in networks. The paper considers a new, more accurate method for identification of small world topology, which hasn't been used for electrical network analysis yet. Also, the paper proposes modification of that method which is adapted for a network with geographical binding.

In most cases, for analysis of applicability of small world model to a certain network, the average path length and overall clustering are compared between the current network and its random analogue (with the same distribution of vertex degrees). Random relinkage procedure is used to generate a random analogue of the network. Basic steps of this procedure consist of random selection of two pairs of linked nodes, disruption of old links within the pair and creation of two new links between two nodes from different pairs. When random relinkage steps are repeated many times, it leads to the same node degree for each node, but shuffled links between the nodes.

We use the traditional small world criteria to compare the network with one extreme case - a random network, but ignore another extreme case: the regular lattice. It is crucial to note that the Watts-Strogatz small world model defines the small world network as a hybrid between a regular lattice and a random network. Comparison with a regular analogue as per the basic criteria is commonly ignored because of absence of simple and widely known appropriate algorithm for the generation of a regular analogue of the network.

Telesford's paper (Q. K. Telesford et al. ., 2011) has revealed that unilateral comparison in small world criterium has significant flaws. In particular: traditional criteria does not have any certain value interval and for different size networks one criteria value can mean fundamentally different small world status. Moreover, this very criterium is not single-valued, consequently one criterium value has two different interpretations in terms of the network applicability in the small world.

For solving that problems in Telesford at all. paper offered new criteria for identification of small world structure. It is based on latticization algorithm, which is used for generation of a regularized analogue of a network. The new criterium uses comparison of overall clustering between the current network and its regularized analogue as well as the comparison of the average path length between the current network and its random analogue. That criterium does not have fundamental flaws of its predecessors. It has a certain value interval form -1 to 1 and criteria values almost independent from the size of the network. Near to 0 values of the criteria conforms to the small world state of the network. The new criterium is single-valued and has only one interpretation in terms of applicability in the small world state.

The latticization algorithm is based on the definition of closed sequence ("ring") of nodes and links length metrics induced by that sequence. As the randomization algorithm latticization algorithm is based on random relinkage procedure. In latticization this procedure has only one difference: the relinkage step is rolled back if the total length of new links is greater than the total length of disrupted links. The closed sequence of nodes initialized for latticization has been generated by random. Because of this latticization procedure is repeated with different initial sequences and the result with minimum total links length is accepted as the latticization result. As a result of the latticization algorithm, the network will be transformed to a one-dimensional quasi-regular network.

The latticization algorithm and small world criteria based on it have not been used for transmission grid analysis yet. In the current research realization of latticization algorithm has been made and applied to a UNES computer model. Basic properties of latticizated UNES network are shown in Table 1. The value of the small world criteria is based on latticization for UNES is 0,26. It is near to 0 value which is typical for small world network and it means that the UNES is quite near to a small world structure. The

UNES criteria value is greater than 0 and it means that the UNES has rather a random than regular structure.

The latticization algorithm is universal and not targeted at application to a power transmission grid or any other infrastructure network. But the nature of infrastructure networks is quite specific. First of all, its nodes has geographic binding. Thus, building a quasi-regular network for a random one-dimensional sequence of nodes is not adequate in this case. Infrastructure networks have a well-defined two-dimensional structure of nodes and metric for the links length, based on geographical distance. Realizing that specific properties in the current research we have developed new modification of latticization algorithm and called it geo-latticization. In geo-latticization geographical coordinates of nodes and geodesic distance are used for calculation length of links in relinkage process. For improving its performance geo-latticization algorithm for random selection of node pairs in relinkage procedure was changed to distance dependent algorithm with using geohash technique for the fast searching of near nodes.

Basic properties of geo-latticizated UNES network are presented in Table 1. As expected, the two-dimensional quasi-regular network show sufficient less average path length and overall clustering in comparison with one-dimensional analogue. Obviously, network on two-dimensional lattice is more compact than on one-dimensional ring. It is explained by the fact that nodes in two-dimensional grids have more neighbors than in the ring structure, and it makes creation of clusters less probable than in one-dimensional case. As a result, the value of small world criteria based on geo-latticization for UNES is 0,18 is a more accurate method which takes into account its geographic nature shows that UNES is more close to small world structure.

As a result of the current research, we can make a conclusion that structure of the UNES is close to small world. It is an important conclusion for analysis of different properties of the UNES including reliability to cascade failure effects and network operation efficiency. Small world criteria based on proposed geo-latticization algorithm is a direction for further analysis of power transmission grids and different infrastructure networks with geographic binding of nodes. It could be used for comparative analysis of real networks and generation of artificial analogue networks with the same structure. That artificial analogue networks are very important for correct detection of individual and common characteristics of the analyzed network.

Table 1. Basic properties of UNES network and its analogues

Network	UNES	randomized analogue	latticizated analogue	geo-latticizated analogue
average path length	11,9	7,9	62,8	19,1
overall clustering	0,0807	0,0038	0,2010	0,1696

Similarity search by using hierarchical navigable small world graphs

Yury Malkov, yurymalkov@mail.ru

Institute of Applied Physics of the Russian Academy of Sciences, Nizhny Novgorod, Russian Federation

We present a new approximate K-nearest neighbor search algorithm based on navigable small worlds with controllable hierarchy. The algorithm incrementally builds a layered structure consisting from hierarchical set of proximity graphs (layers) for nested subsets of the stored elements. This allows producing graphs similar to the previously studied Navigable Small World (NSW) structures [Malkov Y. et. al., LNCS 7404, p. 132, 2012; Malkov Y. et. al., Information Systems 45, p. 61, 2014] (also investigated as a possible reason of navigability in real-life networks [arXiv:1507.06529]), while additionally having the links separated by their characteristic distance scales.

Performance evaluation has demonstrated that the proposed method outperforms previous state-of-art rivals, such as FLANN, FALCONN, VP-tree, annoy, kgraph, ordering permutations, by a large margin on many of real world datasets.

Network studies in Russia: from articles to the structure of a research community

Daria Maltseva, d_malceva@mail.ru

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

The development of science discipline in many respects depends on the system of relations between researchers working in it – if they see themselves as ""invisible college"" or competitors, interact with each other or prefer some ""significant others"". Also what is important for establishment of a certain discipline is institutional context - the official approvement of discipline and presence of organizations engaged in certain research.

The exclusion of Russia from the context of social sciences typical for the Soviet period has further led to certain lags in some areas. The direction of network studies – recognized as a discipline from the 1950s in Western science - in Russia is quite new research methodology. However, now in Russia we can see network scientists and some institutionalized forms of their cooperation.

It is important to understand who – what actors and organizations – are the drivers of network research in Russia; how these drivers are related to each other and at what research teams – Russian of foreign – they are focused. We study it by citation (references) analysis of articles on “network” topics published in Russian journals (eLibrary.ru resource).

The presentation at the conference will cover methodological questions on data collection from eLibrary resource and discussion of some results of the research, which is currently proceeded. Analyzing the data, which amount is about 10 000 papers, we can see the structure of a network research community in Russia, including most active drivers form several clusters who mostly do not correspond with each other.

Polynomial-time Approximation Algorithms for the Coloring Problem in Some Hereditary Classes

Dmitriy Malyshev, dsmalyshev@rambler.ru

National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

A graph H is called an *induced subgraph* of G if H is obtained from G by deletion of vertices. A class of graphs is called *hereditary* if it is closed under deletion of vertices. It is well known that any hereditary (and only hereditary) class X can be defined by a set of its forbidden induced subgraphs Y. We write X=Free(Y).

A *coloring* of a graph G is a mapping c: V(G) -> N, such that c(u) != c(v) for any adjacent vertices u,v in V(G). The *coloring problem* is to verify, for a given graph G and a natural number k, whether chromatic number of G = x(G) is less or equal than k or not. It is a classical NP-complete problem.

The computational complexity of the coloring problem in the hereditary classes of graphs defined by small forbidden induced subgraphs was intensively studied. For all but three hereditary classes defined by 4-vertex forbidden induced subgraphs, it was shown either NP-completeness or polynomial-time solvability of the problem. The three exceptional cases are Free(O_4, C_4), Free($K_{1,3}$, O_4), Free($K_{1,3}$, O_4, K_2+2K_1). Moreover, it was known that the coloring problem for Free($K_{1,3}$, O_4) is polynomially equivalent to the same problem for Free($K_{1,3}$, O_4, K_2+2K_1). In this talk, we do not find out the complexity of the problem for the two open cases. We present approximation polynomial-time algorithms for the problem with some “asymptotic” performance guarantees. More specific, for a graph G, we present a polynomial-time algorithm for computing a number p(G), such that x(G) ≤ p(G) ≤ r * x(G) + O(1), where r= 3/2 if G is (O_4, $K_{3,3}$)-free and r = 4/3 if G is ($K_{1,3}$, O_4, K_2+2K_1)-free. As Free(O_4, C_4) ⊆ Free(O_4, $K_{3,3}$), the performance guarantee for (O_4, $K_{3,3}$)-free graphs also holds for (O_4, C_4)-free graphs.

The work is partially supported by RF President grant MK-4819.2016.1.

Composite dynamic superclusters for the Analysis of Reference Graphs

Boris Mirkin*, *Ivan Rodin*, *ivvrodin@gmail.com

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

We present a new method for cluster analysis that finds a composite supercluster consisting of two non-overlapping parts: core and shell. Core of the cluster is a set of elements such that connection between them is much higher than connection between other elements in cluster. Second part of composite supercluster is a shell that consists of all elements that were not included in the core. We expand this approach to data that changes over time. In this case, we find superclusters for each time period and then extract the optimal core of dynamic supercluster such that it does not change over the time, while shell of dynamic supercluster changes at each time period. In this work we provide an approximation model for composite dynamic supercluster and retrieve the optimal parameters for its building. Finally we present and test the algorithm for finding such superclusters.

Our approach was applied to the analysis of dynamic reference graphs. Reference graph is a graph based on key phrases retrieved from a time-indexed natural language text corpus. Key phrases are presented as nodes of the reference graph and weighted oriented edges of the graph show how one term “refers” to other. Referring here means that one term appears in the context of the other, or more simply, edge from node A to node B means that if term A appears in document, term B is likely to appear in that document too. Thus, applying algorithm of composite supercluster extraction to the dynamic reference graphs we obtained the core that corresponds to strongly connected terms and that reveals strong long-lasting trends, and changing shell show the information about key terms changing over the time. In our work we tested this approach on two reference graphs. First graph was built for the corpus of Russian newspaper economic articles published in 2014, with automatically extracted key phrases, and the second graph was built for the fantasy novel “Song of Ice and Fire” where names of characters served as the set of key terms. In the case of newspaper articles we obtained the information about most popular themes presented in mass media, and in the case of novel books we obtained the most important and powerful community of characters and “local” communities at each period of time. We also provide a tool for visualization of dynamic reference graphs and composite superclusters.

On the equality of matching and cover numbers for P4 in split graphs and their extentions

Dmitry Mokeev, MokeevDB@gmail.com

Laboratory of Algorithms and Technologies for Networks Analysis
National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

Matching for P4 in graph is the set of vertex disjoint induced subgraphs isomorphic to the path of four vertex. Cover for P4 in graph is the set of vertices such as each induced subgraphs isomorphic to P4 contains at least one of vertices of this set.

A graph is a cograph (short for complement-reducible graph) if it can be constructed from isolated vertices by disjoint union and join operations.

A graph is a split graph if it can be partitioned in an independent set and a clique.

We consider a hereditary class of graphs obtained from split graphs via replacing vertices with arbitrary cographs, for which exist matching and cover for P4 of equal cardinalities. There found full constructive description and the set of minimal forbidden induced subgraphs for such class.

Buffer Management with Heterogeneous Processing

Sergey Nikolenko, sergey@logic.pdmi.ras.ru

Russian Academy of Science, Saint Petersburg, Russian Federation

Recent interest in geometric properties of networks has triggered research on embedding graphs on certain geometric spaces. Modern network processors (NPs) increasingly deal with packets that require heterogeneous processing. We consider the problem of managing a bounded size input queue buffer where each packet requires several rounds of processing before it can be transmitted out. The goal of admission control policies is to maximize the total number of successfully transmitted packets. We introduce a new class of policies (semi FIFO, with lazy policies being the most important innovation) that provides a new accounting infrastructure that makes it easier to prove upper bounds on the competitive ratio.

Sequential Network Interdiction with Incomplete Information

Oleg Prokopyev, prokopyev@gmail.com

University of Pittsburgh, USA

We study sequential interdiction when the interdictor has incomplete initial information about the network and the evader has complete knowledge of the network, including its structure and arc costs. In each time period, the interdictor blocks at most k arcs from the network observed up to that period, after which the evader travels along a shortest path between two (fixed) nodes in the interdicted network. By observing the evader's actions, the interdictor learns about the network structure and arc costs and adjusts its actions to maximize the cumulative cost incurred by the evader. A salient feature of our work is that the feedback in each period is deterministic and adversarial. In addition to studying the regret minimization problem, we also discuss time stability of a policy, which is the number of time periods until the interdictor's actions match those of an oracle interdictor with prior knowledge of the network. We propose a class of simple interdiction policies that have a finite regret and detect when the instantaneous regret reaches zero in real time. More importantly, we establish that this class of policies belongs to the set of efficient policies. This is a joint work with Juan Borrero (University of Pittsburgh) and Denis Saure (University of Chile).

Simulation modeling of conformity social behavior model using Facebook Livejournal and Twitter graphs

Andrey Rogatkin, andreyrogatkin@gmail.com

Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation

A linear threshold model of conformity social behavior with binary choice in social networks is considered. It is shown that in the case when all network agents are equal (have same conformity threshold) in some probabilistic sense, the model is equivalent to non-network Granovetter's model of mob behavior. In this case, all information concerning network structure, which influences the system dynamics, can be represented as a distribution function that is uniquely defined by the graph degree distribution and the common conformity threshold.
The model equivalence was tested experimentally using simulation modelling on real structures of social networks Facebook, Livejournal and Twitter. The graphs of these social networks are qualitatively different. The resulting dynamics coincides with Granovetter's model dynamics with high precision. It was empirically shown that if the graph degree distribution is approximated by power law distribution than the dynamics may greatly vary from the dynamics of the system with true degree distribution. Therefore, the power law distribution does not describe network structure well for the considered behavior model. It was shown that a piecewise function, which follows power law for high degrees and is constant for low degrees, may be selected to describe system dynamics.

It turned out that distribution functions, which define system dynamics for Facebook, Livejournal and Twitter can well be approximated by a two-parameter family of functions, where the first parameter is the common conformity threshold and the second parameter is a positive real number, which is different for studied social networks. Thus experiment shows that it is possible to describe the whole network structure with a single positive real parameter to describe well the dynamics of the model under consideration. The question how to compute this parameter from degree distribution remains open.

A new approach to network-decomposition problems

Alexandr Rubchinsky, arubchinsky@yahoo.com

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

The most conventional statement of network decomposition problems consists in its partition into several subnetworks. Typically, it is supposed that connections within these subnetworks are significantly strongly than connections between them. The same concerns many automatic classification problems often presented as network decomposition ones. The above-mentioned partition into a small number of clearly distinct subnetworks or clusters is a result of investigation by itself. The partition allows us to formulate reasonable hypotheses about a considered system, to select a few important parameters, and so on − in brief, to understand, «what is the world in this location». Especially it is important for many social-economical systems, whose functioning is determined by human behaviour and, therefore, cannot be adequately presented by a numerical or stochastic dependence.

Numerous successful examples of application decomposition / classification approach for various systems investigation are well known and comprehensively described. The result usually consists in the construction of a single, the most reasonable in some sense, classification of the initial set. However, the experience in investigation of both model and real systems, leads to conclusions revealing other possibilities of the decomposition / classification approach. The essence of the matter is the construction of a family of classifications, which can include as a particular case a single classification but in many complicated and important situations includes several ones. This family determines one numerical index that characterizes the complexity of the considered system as whole rather than the complexity of its any specific classification. The index has different meaningful interpretations in different situations, but generally it describes complexity, entanglement, perplexity and other hardly defined, though important, properties of various systems.

The goal of the report is presentation of the suggested decomposition approach and demonstration of its possibilities in analyzing of some important real systems. The material is structured as follows:

1. Divisive-agglomerative algorithm of classification family construction.
2. Formal definition of classification problem complexity.
3. Analysis of activity of the 2-nd, the 3-rd and the 4-th RF Dumas (Parliaments) in 1996 – 2007.
4. Analysis of S&P500 stock market behaviour during pre-crisis periods in 2001 and 2008.

Materials, related to the three first items, are partly presented in the recent publication (Rubchinsky A. Divisive-Agglomerative Algorithm and Complexity of Automatic Classification Problems: Preprint / NRU HSE / WP7/2015/09. M.: 2015, 44 p. //Available in Internet at the address: www.hse.ru/en/org/hse/wp/wp7en).

Let us briefly consider the above items.

1. Divisive-agglomerative algorithm of decompositions family construction. At the preliminary stage the neighbourhood graph G is constructed, basing on dissimilarity matrix D. It is used as an input at the main stage.

The algorithm of the main stage is determined as a three-level procedure. At the external level, several runs of the algorithm of the intermediate level are completed. At every run, a family of classifications is determined. Output of the external stage is a new family of classifications, selected among the above mentioned families. This new family is considered as a complete solution of the initial classification problem. At the intermediate level, one family of classifications is constructed. It is executed by a special Divisive-Agglomerative Algorithm (DAA). DAA is based on the new algorithm of graph dichotomy. It presents the internal level of the suggested classification algorithm of the general three-level procedure of the main stage.

2. Formal definition of classification problem complexity. Denote by k the number of consecutive dichotomies at DAA and by r the number of DAA runs. It is easy to see that the maximal possible number of different classifications, considered at the external level of the algorithm, is equal to $0.5(k+1)kr$. Among them, the minimal possible number of the different classifications is equal to $k-1$, which is much less than the maximal number. It seems that a reasonable measure of complexity of a classification problem is the ratio between the number of actually existing different classifications and its maximal possible number. Sometimes it is worthy to consider also the not-normalized version of complexity.
The value of complexity depends upon the parameters k and r of the essential algorithm. However, computational experiments demonstrated a reasonable convergence of complexity that justifies the above definition.

3. Analysis of activity of the 2-nd, the 3-rd and the 4-th RF Dumas. For classifications of parliament members according to their vote results, the general notion of complexity is interpreted as consistence or rationality of this parliament policy. For “tossing” deputies or / and whole fractions the corresponding clusters become poorly distinguished and partially perplexing that results in relatively high value of complexity of their classifications. By contrast, under consistent policy, deputy’s clusters are clearly distinguished and the complexity level is low enough (i.e. in a given parliament the level of consistency, accordance, rationality is high).
The mentioned reasoning was applied to analysis of activity of 2-nd, 3-rd and 4-th RF Duma (Russian parliament, 1996-2007). The classifications based upon one-month votes were constructed for every month. The average values of complexity were calculated for every month, every year and, finally, for every Dumas. The latest are as follows:

Duma 2	Duma 3	Duma 4
0.418	0.147	0.235

These number mean that Duma 3 was the essentially more consistent than two other Dumas. Consistence of Duma in 2002 (just after the United Russia appearance at the political stage by the end of 2001) exceeds its consistence at any other year of the12-years period).

4. Analysis of S&P500 stock market behaviour during pre-crisis periods in 2001 and 2008. Introduce some necessary notions and definitions:

1. Let us define the essential minimal period, consisting of l consecutive days. All the data found for the period x, x−1, …, x−l +1 are related to day x.

2. Prices of all the shares at closure time are considered for days x, x−1, …, x−l+1. The matrix r of pairwise correlation coefficients is determined based on these prices. Thereafter the latest matrix determines the dissimilarity matrix d.

3. The found dissimilarity matrix defines an automatic classification problem, whose complexity is related to day x.

4. The complexity chaotically enough depends upon consecutive days. Therefore the smoothed complexity, defined as the sum of complexity in day x and in previous s−1 days.

The suggested computational scheme is based on the general approach described in items 1 and 2. It defines useful indices for every days of any considered period. Let us compare, for instance, two sequences of smoothed complexities. The 1st one is stopped at the 5-th day before so-called «dotcom» crisis 9.3.2001. The 2nd one is stopped at the 5-th day before so-called «hypothec» crisis 29.09.2008.

1949	2103	2058	2020	2533	2527	2663
1951	1846	1817	2035	2334	2411	2409

It is possible to say that these two sequences demonstrate close behavior in the following sense: the absolute difference between the corresponding numbers does not exceed 257, while the average difference is equal to 155 that is less than 10% of the any number. Moreover, close behavior at any seven consecutive days of current data and the above data is encountered very seldom, not at each year. This reasoning means that the suggested approach can be applied for short-term prediction of major crises on stock markets. Of course, serious results require much more work, especially with various data about different stock markets.

Approximate nearest neighbour search in unconstrained face recognition problem with deep neural network features

Andrey Savchenko, avsavchenko@hse.ru

Laboratory of Algorithms and Technologies for Networks Analysis,
National Research University Higher School of Economics, Nizhny Novgorod, Russian Federation

The key issue of fast approximate nearest-neighbor methods applied to improve the performance of image recognition is their heuristic nature. It is practically impossible to prove that particular algorithm is optimal. At the same time, in this talk we discuss an alternative solution on the basis of statistical approach. The size of image feature vector is usually high. Hence, the central limit theorem can be applied and it can be assumed, that the typical dissimilarity measures are approximately normally distributed. Thus, it is possible to estimate the conventional joint density for each class. By maximizing this density we choose the next reference image to calculate the distance with the query image. If this distance does not exceed a certain threshold, the search procedure is terminated. Otherwise, the estimation of conventional joint density is repeated. The importance of this procedure is caused by the development of optimal greedy fast approximate nearest neighbor search. To decrease the space (memory) complexity of this method, the pivot-based techniques are applied.

In this talk, we will present an experimental study of such procedure in the unconstrained face recognition task. Two popular datasets were used, namely, LFW (Labeled Faces in the Wild, 1680 classes, 4500 photos in the training set) and PubFig83 (83 classes, 550 images in the training set). The Caffe framework is used in our experiments. The 4096 non-negative features are extracted with the very deep convolution-al neural network of the Oxford's Visual Geometry Group. The approximate nearest neighbor methods (randomized k-d tree, ordered permutations, metrized small world) were used in our experiments. It was shown, that the proposed approach is 1.4-4-times faster, than these known methods. Moreover, it is practically as accurate, as an exhaustive search.

Efficient analysis of networks with cycles

S. V. Sevastyanov, seva@math.nsc.ru

Sobolev Institute of Mathematics, SB RAS, Novosibirsk State University, Novosibirsk

Project Scheduling is quite a high demand area of Operations Research, with a broad range of possible applications. Probably, this can approve (and explain) the very high publication activity that could be observed in the area: dozens of books, thousands of papers. Meanwhile, studying many of those books leaves a strange impression. On the one hand, a continued progress is being declared (in each monograph, if a network is under investigation then it is for sure “generalized”). But on the other hand, at a detailed consid- eration it becomes clear that what is introduced as “contemporary achieve- ments” are, in fact, long-known results, published 20 or 40 years before. (Just the names of the inventors changed.) At that, more efficient methods elaborated in discrete mathematics and proved to be successfully applicable to Project Scheduling problems remain unknown for most authors of those books. The talk to be presented is an outcome of my frugal bibliography in- vestigation in the area of Project Scheduling. It pursues several targets, as to: (a) reconstruct the real (non-straightforward) history of Project Schedul- ing development; (b) remind the listeners some, maybe not brand-new (and sometimes even old), but still worthy results, undeserving to be neglected (as well as their authors); and lastly, (c) introduce to listeners some “fresh” results in the area.

Network Analysis of International Migration

Sergey Shvydun, shvydun@hse.ru

National Research University Higher School of Economics, Moscow
Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russian Federation

Our study employs the network approach to the problem of international migration. For the last years, migration has been drawn a lot of attention and has been examined from many points of view, however, very few studies considered it from the network perspective. The international migration can be represented as a network (or weighted-directed graph) where the nodes correspond to countries and the edges correspond to migration flows. The main focus of our study is to reveal a set of critical or central elements in the network. To do it, we calculated different existing centrality measures. Unfortunately, these measures do not take into account individual properties of nodes as well as interaction intensities between different nodes. Thus, there have been designed several long-range interaction (LRI) indices that consider these features. LRI indices are based on two main ideas: the first one is the analysis of the influence between nodes through paths between them and the second one is the analysis of the influence with the help of simulations. Another important issue is to analyze the evolution of the network and to examine how a set of critical elements is changed over time. In our research the United Nations International Migration Flows Database (version 2008, 2015) was used. The database provides the annual dyadic estimates of migration flows between countries from 1970 to 2013. As a result, we obtained an information on critical elements of the network and showed how it is changed in dynamic.

Talent scheduling problem on GPU

Mikhail Sivykh, sivykh-mikl@yandex.ru

Sobolev institute of mathematics

A film shooting is divided into m scenes, where n actors are involved. Assume that one scene is shot by one day. An actor get a salary for each day when he is on shooting area. These days are counted from the first to the last shooting day of the actor. It's possible that there are empty days between them, but these are still defrayable.

The goal is to find a scenes sequence which will minimize total actors salary.

This problem is NP-hard. To solve these problem was developed VNS algorithm. Start solution is obtained by deterministic greedy algorithm. To increase the speed of the calculation process, the algorithm was adopted to use GPU. It's developed appropriate special data structures that ensure the acceleration up to 35 times.

Max-Min Fairness (MMF) Optimization Models for Networks

Theodore B. Trafalis, ttrafalis@ou.edu

School of Industrial and Systems Engineering, University of Oklahoma, Norman, USA

In this talk we present fair optimization models and methods applied to networks. In particular we investigate the lexicographic max-min optimization models known as max-min fairness (MMF) models. We also discuss applications of fair optimization methods for resource allocation problems and communication networks.

Random graphs: models and asymptotic characteristics

M.E. Zhukovskii, zhukmax@gmail.com

Moscow Institute of Physics and Technology, Moscow, Russian Federation

For some large scale networks, there are random graphs which asymptotic properties are close to the properties of these networks. However, characteristics of such random graphs are much simpler to analyze. In particular, preferential attachment models are used to analyze Web and can be applied for the problem of Web pages ranking. For example, for some random graphs, estimations of the expected values of PageRank are known. In the talk, we review the results on PageRank distributions in random graph models.

Another application of the theory of random graphs, is the probabilistic method. Solutions of some combinatorial problems follow from the fact that probabilities of random graphs (commonly, in the classic binomial and uniform Erdos-Renyi models) to have the considered properties are positive. In the talk, we give a brief survey of random graph models and focus on asymptotic probabilities of the classic Erdos-Renyi model.

To the best of our knowledge, the class of first-order properties of graphs is the most widely studied class of graphs in a sense of asymptotic probability. The random graph G(n, p) obeys Zero-One Law if for each first-order property its probability tends to 0 or tends to 1. In 1988, S. Shelah and J. Spencer showed that if α is an *irrational* positive number and $(n) = n^{-\alpha+o(1)}$, then $G(n, n^{-\alpha})$, obeys Zero-One Law. The random graph $G(n, p)$obeys *Zero-One k-Law* if for each first-order property, which is expressed by a first-order formula with quantifier depth at most k, its probability tends to 0 or tends to 1.

We present the results on Zero-One Law and Zero-One k-Law for the random graph $G(n, n^{-\alpha})$ (for the most non-trivial case) and the distributions of positive rational numbers $\alpha < 1$ such that Zero-One k-Law for $G(n, n^{-\alpha})$ does not hold (in this case, we say that α is in *k-spectrum*). In particular, we found bounds on minimal and maximal limit points of k-spectrum. Moreover, we prove that the minimal k such that k-spectrum is infinite is either 4 or 5.